



버전 15

기본 분석

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

Marcel Proust

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

15.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020. JMP®15 살펴보기 . Cary, NC: SAS Institute Inc.

JMP® 15 기본 분석

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

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

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

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

September 2019

February 2020

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

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

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

목차

기본 분석

1	JMP 알아보기	13
	설명서 및 추가 리소스	
	서식 규칙	15
	JMP 도움말	16
	JMP 설명서 라이브러리	16
	JMP 학습을 위한 추가 리소스	22
	자습서	22
	샘플 데이터 테이블	22
	통계 및 JSL 용어에 대해 알아보기	23
	JMP 팁 및 힌트 알아보기	23
	툴팁	23
	JMP 사용자 커뮤니티	23
	무료 온라인 Statistical Thinking 교육 과정	24
	새로운 사용자를 위한 입문 키트	24
	Statistics Knowledge 포털	24
	JMP 교육 과정	24
	사용자가 작성한 JMP 설명서	24
	JMP 시작하기 창	24
	기술 지원	25
2	기본 분석 소개	27
	기본 분석 방법 개요	
3	분포	29
	분포 플랫폼 사용	
	분포 플랫폼의 개요	32
	범주형 변수	32
	연속형 변수	32
	분포 플랫폼의 예	33
	분포 플랫폼 시작	35
	분포 보고서	36
	히스토그램	37

빈도 보고서	40
분위수 보고서	41
요약 통계량 보고서	41
분포 플랫폼 옵션	43
범주형 변수에 대한 옵션	44
범주형 변수에 대한 표시 옵션	44
범주형 변수에 대한 히스토그램 옵션	45
범주형 변수에 대한 저장 옵션	45
연속형 변수에 대한 옵션	46
연속형 변수에 대한 표시 옵션	47
연속형 변수에 대한 히스토그램 옵션	47
정규 분위수 그림	48
이상치 상자 그림	49
분위수 상자 그림	50
줄기 - 잎	52
CDF 그림	52
평균 검정	53
표준편차 검정	54
동등성 검정	55
신뢰 구간	56
예측 구간	56
공차 구간	56
공정 능력	57
분포 적합	59
연속형 변수에 대한 저장 옵션	63
분포 플랫폼의 추가 예	65
예: 여러 히스토그램에서 데이터 선택	65
예: 기준 변수 사용	66
검정 확률 옵션의 예	67
예측 구간의 예	69
공차 구간의 예	71
공정 능력의 예	72
분포 플랫폼에 대한 통계 상세 정보	73
표준 오차 막대	74
분위수	74
요약 통계량	75
정규 분위수 그림	76
Wilcoxon 부호 순위 검정	76
표준편차 검정	78
정규 분위수	78

표준화된 데이터 저장	79
예측 구간	79
공차 구간	80
연속형 적합 분포	82
이산형 적합 분포	88
레거시 분포 적합기에 대한 상세 정보	91
분포 적합 옵션 (레거시)	91
연속형 적합 분포에 대한 통계 상세 정보 (레거시)	96
이산형 적합 분포에 대한 통계 상세 정보 (레거시)	101
적합 분위수에 대한 통계 상세 정보 (레거시)	102
분포 적합 옵션에 대한 통계 상세 정보 (레거시)	102
4 X로 Y 적합 소개	105
두 변수 간의 관계 분석	
X로 Y 적합 플랫폼의 개요	107
X로 Y 적합 플랫폼 시작	107
JMP 시작하기 창에서 특정 분석 시작	108
5 이변량 분석	109
두 연속형 변수 간의 관계 분석	
이변량 분석의 예	112
이변량 플랫폼 시작	112
이변량 그림	114
적합 옵션	114
적합 옵션	115
적합 옵션 범주	117
동일한 적합 옵션을 여러 번 적용	117
히스토그램 테두리	118
평균 적합	119
평균 적합 보고서	119
선형 적합 및 다항식 적합	120
선형 적합 및 다항식 적합 보고서	120
특수 적합	125
특수 적합 보고서 및 메뉴	126
유연	127
스플라인 적합	127
커널 평활기	128
각 값 적합	129
직교 적합	130
직교 적합 비율 보고서	130

로버스트	131
로버스트 적합	131
Cauchy 적합	131
밀도 타원	132
상관 보고서	133
비모수 밀도	133
분위수 밀도 등고선 보고서	134
그룹화 기준	134
적합 메뉴	135
적합 메뉴 옵션	135
진단 그림	138
이변량 플랫폼의 추가 예	138
특수 적합 옵션의 예	139
직교 적합 옵션의 예	141
로버스트 적합 명령의 예	142
밀도 타원을 사용한 그룹화의 예	144
회귀선을 사용한 그룹화의 예	145
기준 변수를 사용한 그룹화의 예	146
이변량 플랫폼에 대한 통계 상세 정보	147
선형 적합	148
스플라인 적합	148
직교 적합	148
적합 요약 보고서	149
적합 결여 보고서	149
모수 추정값 보고서	150
평활 적합 보고서	150
상관 보고서	151

6 일원 분석 153

연속형 Y와 범주형 X 변수 간의 관계 분석

일원 분석 개요	156
일원 분석의 예	156
일원 분석 플랫폼 시작	158
데이터 형식	158
일원 분석 그림	159
일원 분석 플랫폼 옵션	160
표시 옵션	163
분위수	164
이상치 상자 그림	164
평균 /ANOVA 및 평균 /ANOVA/ 합동 t	165

적합 요약 보고서	166
t-검정 보고서	166
분산 분석 보고서	167
일원 ANOVA 에 대한 평균 보고서	168
블록 평균 보고서	168
평균 다이아몬드 및 X 축 비례	168
평균 선, 오차 막대 및 표준편차 선	169
평균 분석 방법	170
위치에 대한 평균 분석	170
척도에 대한 평균 분석	171
평균 분석 차트	172
평균 분석 옵션	172
평균 비교	173
비교 원 사용	175
개별 쌍 비교 (스튜던트 t)	176
전체 쌍 비교 (Tukey HSD)	176
최량 비교 (Hsu MCB)	176
대조군 비교 (Dunnett)	178
단계별 개별 쌍 비교 (Newman-Keuls)	178
평균 비교 옵션	179
비모수 검정	179
Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서	181
Kolmogorov-Smirnov 2 표본 검정 보고서	182
비모수 다중 비교	183
이분산	185
분산 동일 여부 검정 보고서	186
동등성 검정	187
로버스트	188
로버스트 적합	188
Cauchy 적합	188
검정력	189
검정력 상세 정보 창 및 보고서	189
정규 분위수 그림	190
CDF 그림	190
밀도	191
매칭 열	191
일원 분석 플랫폼의 추가 예	192
평균 분석 차트의 예	192
분산에 대한 평균 분석 차트의 예	193
개별 쌍 비교 (스튜던트 t) 검정의 예	194

전체 쌍 비교 (Tukey HSD) 검정의 예	196
최량 비교 (Hsu MCB) 검정의 예	198
대조군 비교 (Dunnett) 의 예	199
단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예	200
예 : 4 가지 평균 비교 검정 대조	201
비모수 Wilcoxon 검정의 예	202
이분산 옵션의 예	205
동등성 검정의 예	206
로버스트 적합 옵션의 예	207
검정력 옵션의 예	209
정규 분위수 그림의 예	210
CDF 그림의 예	211
밀도 옵션의 예	212
매칭 열 옵션의 예	213
예 : 일원 분석을 위한 데이터 쌓기	214
일원 분석 플랫폼에 대한 통계 상세 정보	221
비교 원	221
검정력	223
적합 요약 보고서	223
분산 동일 여부 검정	224
비모수 검정 통계량	225

7 분할 분석

두 범주형 변수 간의 관계 분석

분할 분석의 예	231
분할 분석 플랫폼 시작	232
데이터 형식	233
분할 분석 보고서	233
분할 분석 플랫폼 옵션	234
모자이크 그림	235
팝업 메뉴	236
분할표	238
분할표 설명	239
검정	239
검정 보고서 설명	240
Fisher 정확 검정	240
비율에 대한 평균 분석	241
대응 분석	241
대응 분석 그림 이해	241
대응 분석 옵션	242

상세 정보 보고서	242
Cochran-Mantel-Haenszel 검정	242
합치도 통계량	243
상대 위험도	243
비율에 대한 2 표본 검정	244
연관성 측도	244
Cochran Armitage 추세 검정	245
정확 검정	246
분할 분석 플랫폼의 추가 예	246
비율에 대한 평균 분석의 예	247
대응 분석의 예	248
Cochran Mantel Haenszel 검정의 예	250
합치도 통계량 옵션의 예	251
상대 위험도 옵션의 예	252
비율에 대한 2 표본 검정의 예	254
연관성 측도 옵션의 예	255
Cochran Armitage 추세 검정의 예	255
분할 분석 플랫폼에 대한 통계 상세 정보	256
합치도 통계량	256
승산비 옵션	257
검정 보고서	257
대응 분석의 상세 정보 보고서	258
8 로지스틱 분석	259
범주형 Y 변수와 연속형 X 변수 간의 관계 분석	
로지스틱 회귀의 개요	261
명목형 로지스틱 회귀	261
순서형 로지스틱 회귀	261
명목형 로지스틱 회귀의 예	262
로지스틱 플랫폼 시작	263
데이터 구조	264
로지스틱 보고서	265
로지스틱 그림	266
반복	266
전체 모형 검정	266
적합 상세 정보	267
모수 추정값	268
로지스틱 플랫폼 옵션	269
ROC 곡선	270
확률 계산식 저장	271

역추정 예측	271
로지스틱 회귀의 추가 예	271
순서형 로지스틱 회귀의 예	272
로지스틱 그림의 추가 예	273
ROC 곡선의 예	275
십자기호 도구를 사용한 역추정 예측의 예	276
역추정 예측 옵션을 사용한 역추정 예측의 예	277
로지스틱 플랫폼에 대한 통계 상세 정보	279

9 테이블 생성 281

대화식으로 요약 테이블 생성

테이블 생성 플랫폼의 예	283
테이블 생성 플랫폼 시작	287
대화상자 사용	288
통계량 추가	289
테이블 생성 출력	292
분석 열	293
그룹화 열	294
열 및 행 테이블	294
테이블 편집	295
테이블 생성 플랫폼 옵션	296
테스트 빌드 패널 표시	297
마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴	297
테이블 생성 플랫폼의 추가 예	298
예: 여러 테이블 생성 및 내용 재배열	298
예: 여러 열을 단일 테이블에 결합	301
예: 페이지 열 사용	303

10 시뮬레이션 305

모수적 대표집을 통해 난제 해결 방안 모색

시뮬레이션 플랫폼 개요	307
시뮬레이션 사용 예	307
분산 성분에 대한 준모수적 신뢰 구간 구성	308
순열 검정 수행	313
일반화 회귀의 요인 유지 탐색	315
비선형 모형에 대한 전향적 검정력 분석	320
시뮬레이션 창 시작	329
시뮬레이션 창	329
시뮬레이션 결과 테이블	330
시뮬레이션 결과 보고서	331

시물레이션 검정력 보고서	331
11 붓스트랩	333
재표집을 통한 통계량 분포 근사	
붓스트랩 개요	335
붓스트랩의 예	336
붓스트랩 창 옵션	338
누적 결과 테이블	340
비누적 붓스트랩 결과 테이블	341
붓스트랩 결과 분석	342
붓스트랩의 추가 예	343
붓스트랩에 대한 통계 상세 정보	348
부분 가중치 계산	348
편향 수정 백분위수 구간	348
붓스트랩의 가용성	349
12 텍스트 탐색기	351
데이터의 비정형 텍스트 탐구	
텍스트 탐색기 플랫폼 개요	353
텍스트 처리 단계	354
텍스트 탐색기 플랫폼의 예	355
텍스트 탐색기 플랫폼 시작	358
정규 표현식 편집기에서 정규 표현식 사용자 정의	360
텍스트 탐색기 보고서	366
요약 개수 보고서	366
용어 및 구 목록	367
텍스트 탐색기 플랫폼 옵션	370
텍스트 준비 옵션	370
텍스트 분석 옵션	375
저장 옵션	376
보고서 옵션	378
잠재 계층 분석	378
잠재 의미 분석 (SVD)	380
SVD 보고서	381
SVD 보고서 옵션	381
주제 분석	383
주제 분석 보고서	384
주제 분석 보고서 옵션	384
판별 분석	385
판별 분석 보고서	385

판별 분석 보고서 옵션 386

 텍스트 탐색기 플랫폼의 추가 예 387

A 참조 자료 391

B 기술 라이선스 고지 사항 395

1 장

JMP 알아보기

설명서 및 추가 리소스


이 장에는 설명서 서식 규칙, 각 JMP 문서에 대한 설명, 도움말 시스템, 기타 지원 제공 위치 등 JMP 설명서에 대한 상세 정보가 포함되어 있습니다.

목차

서식 규칙.....	15
JMP 도움말.....	16
JMP 설명서 라이브러리	16
JMP 학습을 위한 추가 리소스	22
자습서	22
샘플 데이터 테이블	22
통계 및 JSL 용어에 대해 알아보기.....	23
JMP 팁 및 힌트 알아보기	23
툴팁	23
JMP 사용자 커뮤니티	23
무료 온라인 Statistical Thinking 교육 과정	24
새로운 사용자를 위한 입문 키트	24
Statistics Knowledge 포털	24
JMP 교육 과정	24
사용자가 작성한 JMP 설명서	24
JMP 시작하기 창	24
기술 지원.....	25

서식 규칙

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


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

참고 : 참고 섹션에는 특수한 정보와 제한 사항이 표시됩니다.

팁 : 팁 섹션에는 유용한 정보가 표시됩니다.

JMP 도움말

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

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

참고: JMP 도움말은 인터넷이 연결되어 있어야 사용할 수 있습니다. 인터넷이 연결되어 있지 않으면 **도움말 > JMP 설명서 라이브러리**를 선택하여 단일 PDF 파일에서 모든 설명서를 검색할 수 있습니다. 자세한 내용은 "[JMP 설명서 라이브러리](#)"(16 페이지)에서 확인하십시오.

JMP 설명서 라이브러리

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

다음 표에서는 JMP 라이브러리에 포함된 각 문서의 용도와 내용을 설명합니다.

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

문서 제목	문서 용도	문서 내용
기본 분석	이 문서를 사용하여 기본적인 분석을 수행합니다.	<p>"분석" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 분포 • X로 Y 적합 • 테이블 생성 • 텍스트 탐색기 <p>"분석" > "X로 Y 적합"을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표본 분포를 근사화하는 방법과 시뮬레이션 플랫폼을 사용하여 모수적 재표집을 수행하는 방법도 포함되어 있습니다.</p>
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	<p>"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 <p>이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.</p>
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	<p>"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.</p>

문서 제목	문서 용도	문서 내용
실험 설계 가이드	실험 설계 방법을 알아보고 적절한 표본 크기를 결정합니다.	"DOE" 메뉴의 모든 항목을 다룹니다.
Fitting Linear Models	모형 적합 플랫폼과 이 플랫폼의 다양한 분석법에 대해 알아봅니다.	<p>"분석" 메뉴의 모형 적합 플랫폼에서 사용할 수 있는 다음 분석법에 대해 설명합니다.</p> <ul style="list-style-type: none"> 표준 최소 제곱 단계별 일반화 회귀 혼합 모형 MANOVA 로그 선형 분산 명목형 로지스틱 순서형 로지스틱 일반화 선형 모형

문서 제목	문서 용도	문서 내용
Predictive and Specialized Modeling	추가 예측 모형 기법에 대해 알아봅니다.	<p>"분석">"예측 모형" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 신경망 • 파티션 • 붓스트랩 포레스트 • 부스티드 트리 • K 최근접 이웃 • Naive Bayes • 서포트 벡터 머신 • 모형 비교 • 검증 열 생성 • 계산식 저장소 <p>"분석">"전문 모형" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 곡선 적합 • 비선형 • 함수 데이터 탐색기 • 가우스 과정 • 시계열 • 매칭 쌍 <p>"분석">"선별" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 모형화 유틸리티 • 반응 변수 선별 • 공정 변수 선별 • 예측 변수 선별 • 연관성 분석 • 공정 기록 탐색기

문서 제목	문서 용도	문서 내용
Multivariate Methods	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	<p>"분석">"다변량 방법" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 항목 분석 <p>"분석">"군집화" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화
Quality and Process Methods	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	<p>"분석">"품질 및 공정" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 • 계량형 / 계수형 게이지 차트 • 공정 능력 • 모형 기반 다변량 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리

문서 제목	문서 용도	문서 내용
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	<p>"분석">"신뢰성 및 생존" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	<p>"분석">"소비자 조사" 메뉴의 다음 플랫폼에 대해 설명합니다.</p> <ul style="list-style-type: none"> • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Scripting Guide	강력한 JSL(JMP 스크립트 언어)의 활용 방법을 알아봅니다.	스크립트 작성 / 디버깅, 데이터 테이블 조작, 표시 상자 생성 및 JMP 응용 프로그램 생성 등의 다양한 주제를 다룹니다.
JSL Syntax Reference	다양한 JSL 함수 및 인수, 그리고 개체 및 표시 상자로 보내는 메시지에 대해 알아봅니다.	JSL 명령의 구문, 예제 및 참고 사항이 포함되어 있습니다.

JMP 학습을 위한 추가 리소스

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

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

자습서

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

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

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

샘플 데이터 테이블

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

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

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\15\Samples\Data

macOS: \Library\Application Support\JMP\15\Samples\Data

JMP Pro 의 경우에는 JMP 디렉터리가 아니라 JMPPRO 디렉터리에 샘플 데이터가 설치되어 있습니다.

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

통계 및 JSL 용어에 대해 알아보기

도움말 메뉴에는 다음과 같은 인덱스가 포함되어 있습니다.

통계 분석 인덱스 통계 용어에 대한 정의를 제공합니다.

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

JMP 팁 및 힌트 알아보기

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

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

툴팁

JMP 에서 다음과 같은 항목 위에 커서를 놓으면 설명 툴팁 (또는 *가리키기 라벨*) 이 제공됩니다.

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

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

JMP 사용자 커뮤니티

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

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

웹 사이트에서 JMP 리소스에 액세스하려면 **도움말 > JMP 사용자 커뮤니티**를 선택하거나 <https://community.jmp.com/> 을 방문하십시오.

무료 온라인 Statistical Thinking 교육 과정

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

새로운 사용자를 위한 입문 키트

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

Statistics Knowledge 포털

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

JMP 교육 과정

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

사용자가 작성한 JMP 설명서

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

JMP 시작하기 창

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

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

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

기술 지원

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

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

2 장

기본 분석 소개

기본 분석 방법 개요

기본 분석에서는 JMP 에서 종종 수행하는 다음과 같은 초기 분석 유형에 대해 설명합니다.

- 분포 플랫폼에서는 히스토그램, 추가 그래프 및 보고서를 사용하여 단일 변수의 분포를 보여줍니다. 데이터가 어떻게 분포되어 있는지 알게 되면 적절한 유형의 분석을 계획할 수 있습니다. 자세한 내용은 [3 장, "분포"](#) 에서 확인하십시오.
- X 로 Y 적합 플랫폼에서는 지정된 X 및 Y 변수 쌍을 모델링 유형에 따라 컨텍스트별로 분석합니다. 자세한 내용은 [4 장, "X 로 Y 적합 소개"](#) 에서 확인하십시오. 네 가지 분석 유형으로 는 다음이 포함됩니다.
 - 이변량 플랫폼: 두 연속형 X 변수 간의 관계를 분석합니다. 자세한 내용은 [5 장, "이변량 분석"](#) 에서 확인하십시오.
 - 일원 분석 플랫폼: 범주형 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지를 분석합니다. 자세한 내용은 [6 장, "일원 분석"](#) 에서 확인하십시오.
 - 분할 분석 플랫폼: 범주형 X 요인의 값에 따라 좌우되는 범주형 반응 변수의 분포를 분석합니다. 자세한 내용은 [7 장, "분할 분석"](#) 에서 확인하십시오.
 - 로지스틱 플랫폼: 반응 범주 (Y) 에 대한 확률을 연속형 X 예측 변수에 적합시킵니다. 자세한 내용은 [8 장, "로지스틱 분석"](#) 에서 확인하십시오.
- 테이블 생성 플랫폼: 기술 통계량 테이블을 대화식으로 생성합니다. 자세한 내용은 [9 장, "테이블 생성"](#) 에서 확인하십시오.
- 시뮬레이션 기능으로는 모수적 및 비모수적 시뮬레이션을 수행할 수 있습니다. 자세한 내용은 [10 장, "시뮬레이션"](#) 에서 확인하십시오.
- 붓스트랩 분석은 통계량의 표본 분포에 근사한 값을 산출합니다. 복원 표집 방식으로 데이터가 재표집되어 통계량이 계산됩니다. 통계량의 값 분포가 생성될 때까지 이 과정이 반복됩니다. 자세한 내용은 [11 장, "붓스트랩"](#) 에서 확인하십시오.
- 텍스트 탐색기 플랫폼에서는 형식이 지정되지 않은 텍스트 데이터를 범주화하고 분석할 수 있습니다. 분석을 진행하기 전에 정규 표현식을 사용하여 데이터를 정리할 수 있습니다. 자세한 내용은 [12 장, "텍스트 탐색기"](#) 에서 확인하십시오.

3 장

분포

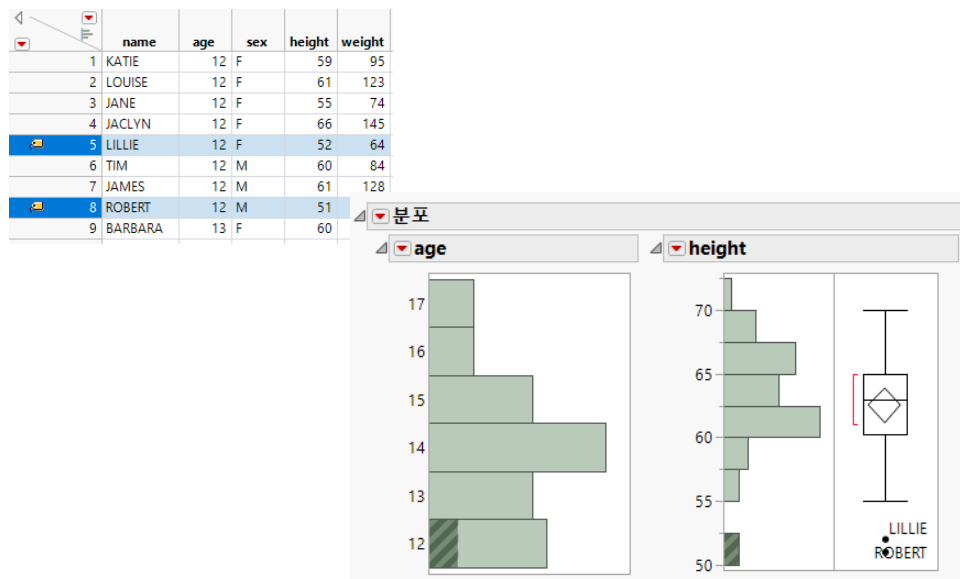
분포 플랫폼 사용

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

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

분포 보고서 창은 대화식입니다. 히스토그램 막대를 클릭하면 다른 모든 히스토그램과 데이터 테이블에서 해당 데이터가 강조 표시됩니다.

그림 3.1 분포 플랫폼의 예



목차

분포 플랫폼의 개요	32
범주형 변수	32
연속형 변수	32
분포 플랫폼의 예	33
분포 플랫폼 시작	35
분포 보고서	36
히스토그램	37
빈도 보고서	40
분위수 보고서	41
요약 통계량 보고서	41
분포 플랫폼 옵션	43
범주형 변수에 대한 옵션	44
범주형 변수에 대한 표시 옵션	44
범주형 변수에 대한 히스토그램 옵션	45
범주형 변수에 대한 저장 옵션	45
연속형 변수에 대한 옵션	46
연속형 변수에 대한 표시 옵션	47
연속형 변수에 대한 히스토그램 옵션	47
정규 분위수 그림	48
이상치 상자 그림	49
분위수 상자 그림	50
줄기-잎	52
CDF 그림	52
평균 검정	53
표준편차 검정	54
동등성 검정	55
신뢰 구간	56
예측 구간	56
공차 구간	56
공정 능력	57
분포 적합	59
연속형 변수에 대한 저장 옵션	63
분포 플랫폼의 추가 예	65
예: 여러 히스토그램에서 데이터 선택	65
예: 기준 변수 사용	66
검정 확률 옵션의 예	67
예측 구간의 예	69
공차 구간의 예	71

공정 능력의 예	72
분포 플랫폼에 대한 통계 상세 정보	73
표준 오차 막대	74
분위수	74
요약 통계량	75
정규 분위수 그림	76
Wilcoxon 부호 순위 검정	76
표준편차 검정	78
정규 분위수	78
표준화된 데이터 저장	79
예측 구간	79
공차 구간	80
연속형 적합 분포	82
이산형 적합 분포	88
레거시 분포 적합기에 대한 상세 정보	91
분포 적합 옵션(레거시)	91
연속형 적합 분포에 대한 통계 상세 정보(레거시)	96
이산형 적합 분포에 대한 통계 상세 정보(레거시)	101
적합 분위수에 대한 통계 상세 정보(레거시)	102
분포 적합 옵션에 대한 통계 상세 정보(레거시)	102

분포 플랫폼의 개요

분포 플랫폼에서는 변수의 모델링 유형, 즉 변수가 범주형 (명목형 또는 순서형) 인지 연속형인지에 따라 변수 처리가 달라집니다.

범주형 변수

범주형 변수의 경우 초기에 표시되는 그래프는 히스토그램입니다. 히스토그램에는 순서형 또는 명목형 변수의 각 수준에 대한 막대가 표시됩니다. 나뉜 (모자이크) 막대 차트를 추가할 수도 있습니다.

빈도 보고서에는 개수와 비율이 표시됩니다. 빨간색 삼각형 메뉴의 옵션을 사용하여 신뢰 구간을 추가하고 확률을 검정할 수 있습니다.

연속형 변수

연속형 숫자 변수의 경우 초기 그래프에는 히스토그램과 이상치 상자 그림이 표시됩니다. 히스토그램에는 연속형 변수의 그룹화된 값에 대한 막대가 표시됩니다. 다음 옵션도 사용할 수 있습니다.

- 정규 분위수 그림
- 분위수 상자 그림
- 줄기 - 잎 그림
- CDF 그림

보고서에는 선택한 사분위수와 요약 통계량이 표시됩니다. 빨간색 삼각형 메뉴에서는 다음 작업을 위한 추가 보고서 옵션을 사용할 수 있습니다.

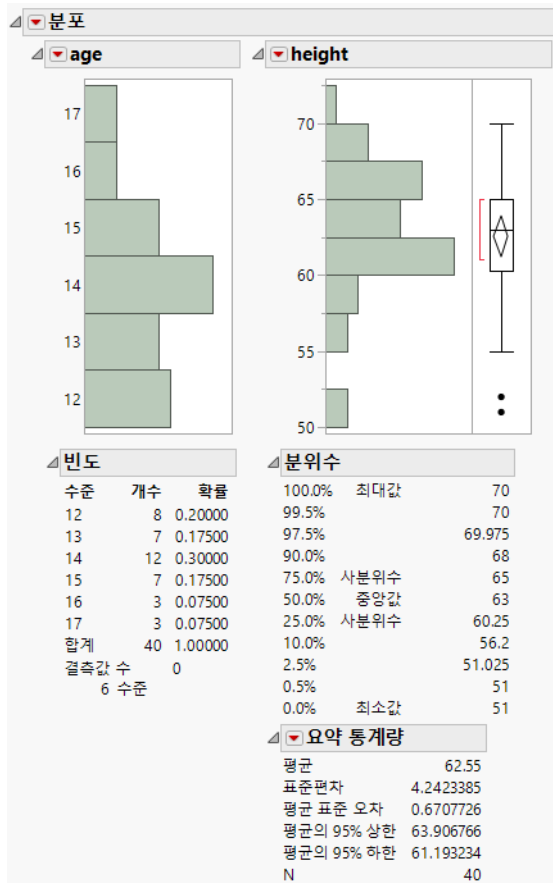
- 순위, 확률 점수, 정규 분위수 값 등을 데이터 테이블에 새 열로 저장 ("저장" 옵션)
- 지정한 상수를 기준으로 열의 평균 및 표준편차 검정 ("평균 검정" 및 "표준편차 검정" 옵션)
- 다양한 분포 및 비모수 평활 곡선 적합 ("연속형 적합" 및 "이산형 적합" 옵션)
- 품질 관리 응용 프로그램에 대한 공정 능력 분석 수행
- 신뢰 구간, 예측 구간 및 공차 구간

분포 플랫폼의 예

40 명의 학생에 대한 데이터가 있고 이 학생들의 연령 및 키 분포를 보려고 한다고 가정해 보겠습니다.

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

그림 3.2 분포 플랫폼의 예



히스토그램에서 다음을 확인할 수 있습니다 .

- 연령은 균일하게 분포되어 있지 않습니다 .

- 키의 경우 극단값 (이상치일 수 있음) 을 갖는 두 개의 점이 있습니다.

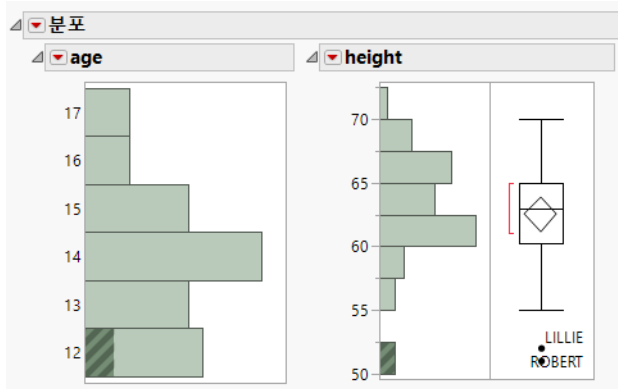
"height" 히스토그램에서 50 에 해당하는 막대를 클릭하여 잠재적 이상치를 보다 자세히 확인합니다.

- "age" 히스토그램에서 해당 연령이 강조 표시됩니다. 잠재적 이상치의 연령은 12 세입니다.
- 데이터 테이블에서 해당 행이 강조 표시됩니다. 잠재적 이상치의 이름은 Lillie 와 Robert 입니다.

"height" 히스토그램에서 잠재적 이상치에 라벨을 추가합니다.

1. 두 개의 이상치를 모두 선택합니다.
2. 이상치 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **행 라벨**을 선택합니다.
데이터 테이블에서 해당 행에 라벨 아이콘이 추가됩니다.
3. 전체 라벨이 표시되도록 상자 그림을 더 넓게 조정합니다.

그림 3.3 라벨이 지정된 잠재적 이상치



분포 플랫폼 시작

분석 > 분포를 선택하여 분포 플랫폼을 시작할 수 있습니다.

그림 3.4 분포 시작 창

각 변수에 대한 히스토그램 및 단변량 통계량을 표시합니다.

열 선택

5개 열

- name
- age
- sex
- height
- weight

☐ 히스토그램만

선택한 열 역할 지정

Y, 열	필수
가중치	선택적 숫자
빈도	선택적 숫자
기준	선택적

작업

확인

취소

제거

재호출

도움말

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 "시작하기"장에서 확인하십시오.

Y, 열 분석하려는 변수를 할당합니다. 각 변수에 대해 히스토그램과 관련 보고서가 표시됩니다.

가중치 연속형 Y의 관측값에 대한 가중치를 지정하는 변수를 할당합니다. 범주형 Y의 경우에는 "가중치" 열이 무시됩니다. 가중치 합계를 기준으로 하는 모든 통계량은 가중치의 영향을 받습니다.

빈도 이 역할에 빈도 변수를 할당합니다. 이 기능은 데이터를 요약할 경우에 유용합니다. 이 경우 Y 값에 대한 열 하나와 Y 값의 발생 빈도에 대한 또 다른 열이 있게 됩니다. 이 변수의 합계는 "요약 통계량" 보고서에 표시되는 전체 개수에 포함됩니다 (N으로 표시). 다른 모든 적률 통계량 (평균, 표준편차 등)도 빈도 변수의 영향을 받습니다.

기준 기준 변수의 각 수준에 대해 개별 보고서를 생성합니다. 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다.

공정 능력 생성 (열에 "규격 한계" 열 특성이 포함된 경우에만 나타남) "규격 한계" 열 특성이 포함된 분석 열에 대한 공정 능력 보고서를 추가합니다.

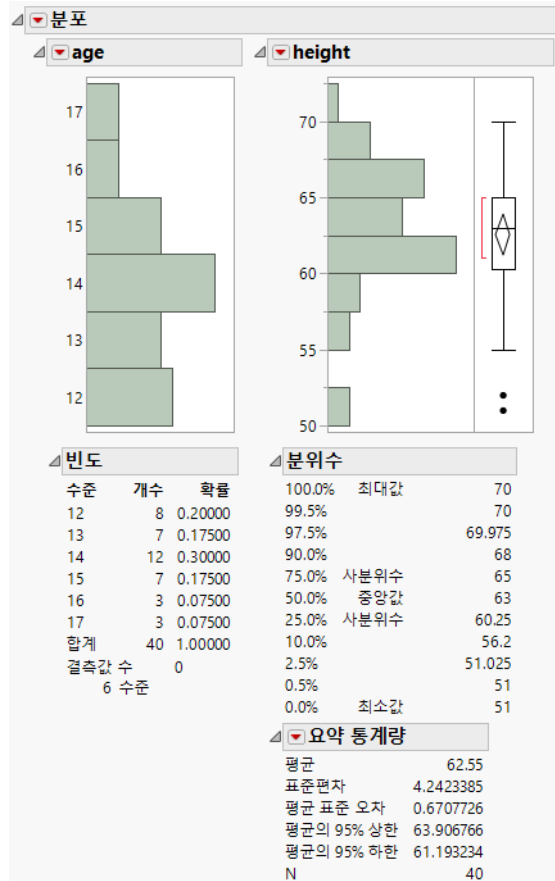
히스토그램만 보고서 창에서 히스토그램을 제외한 모든 내용을 제거합니다.

시작 창에 대한 자세한 내용은 JMP 사용의 "시작하기"장에서 확인하십시오.

분포 보고서

그림 3.5에 표시된 것과 같은 보고서를 생성하려면 "분포 플랫폼의 예"(33 페이지)에 설명된 방법을 따르십시오.

그림 3.5 초기 분포 보고서 창



참고: 데이터 테이블의 행에 "숨김" 행 상태만 적용하면 데이터 점을 표시하는 그림에 해당 점이 표시되지 않습니다. 하지만 히스토그램을 생성하는 데는 숨겨진 행이 사용됩니다. 생성되는 히스토그램과 분석 결과에서 행을 제외하려면 "제외" 행 상태를 적용하십시오. 그런 다음 "분포" 옆의 빨간색 삼각형 메뉴에서 **다시 실행 > 분석 다시 실행**을 선택하십시오. 데이터 테이블에서 제외된 모든 행은 점을 표시하는 그림에서도 숨겨집니다.

초기 "분포" 보고서에는 각 변수에 대한 히스토그램과 보고서가 포함됩니다. 다음 사항에 유의하십시오.

- 보고서의 변수를 바꾸려면 관련 데이터 테이블의 "열" 패널에서 원하는 변수를 히스토그램의 축으로 드래그하여 놓습니다.
- 보고서에 새 변수를 삽입하여 새 히스토그램을 생성하려면 기존 히스토그램의 변수를 히스토그램 외부로 드래그하여 놓습니다. 기존 히스토그램의 앞, 중간 또는 뒤에 새 변수를 삽입할 수 있습니다.

참고 : 변수를 제거하려면 빨간색 삼각형 메뉴에서 **제거**를 선택하십시오.

- "분포" 옆의 빨간색 삼각형 메뉴에는 모든 변수에 영향을 주는 옵션이 포함되어 있습니다. 자세한 내용은 "[분포 플랫폼 옵션](#)"(43 페이지)에서 확인하십시오.
- 각 변수 옆의 빨간색 삼각형 메뉴에는 해당 변수에만 영향을 주는 옵션이 포함되어 있습니다. 자세한 내용은 "[범주형 변수에 대한 옵션](#)"(44 페이지) 또는 "[연속형 변수에 대한 옵션](#)"(46 페이지)에서 확인하십시오. Ctrl 키를 누른 채 변수 옵션을 선택하면 보고서에서 모델링 유형이 동일한 모든 변수에 해당 옵션이 적용됩니다.
- 히스토그램에는 데이터가 시각적으로 표시됩니다. 자세한 내용은 "[히스토그램](#)"(37 페이지)에서 확인하십시오.
- 범주형 변수에 대한 초기 보고서에는 "빈도" 보고서가 포함됩니다. 자세한 내용은 "[빈도 보고서](#)"(40 페이지)에서 확인하십시오.
- 연속형 변수에 대한 초기 보고서에는 "분위수" 및 "요약 통계량" 보고서가 포함됩니다. 자세한 내용은 "[분위수 보고서](#)"(41 페이지) 및 "[요약 통계량 보고서](#)"(41 페이지)에서 확인하십시오.

히스토그램

히스토그램에는 데이터가 시각적으로 표시됩니다. 범주형 (명목형 또는 순서형) 변수의 경우에는 히스토그램에 순서형 또는 명목형 변수의 각 수준에 대한 막대가 표시됩니다. 연속형 변수의 경우에는 히스토그램에 연속형 변수의 그룹화된 값에 대한 막대가 표시됩니다.

데이터 강조 표시 히스토그램 막대 또는 그래프의 이상점을 클릭합니다. 그러면 데이터 테이블에서 해당 행이 강조 표시되고, 다른 히스토그램의 해당 부분도 강조 표시됩니다 (해당되는 경우). 자세한 내용은 "[막대 강조 표시 및 행 선택](#)"(39 페이지)에서 확인하십시오.

부분집합 생성 히스토그램 막대를 두 번 클릭하거나, 히스토그램 막대를 마우스 오른쪽 버튼으로 클릭한 후 **부분집합**을 선택합니다. 그러면 선택한 데이터만 포함된 새 데이터 테이블이 생성됩니다.

전체 히스토그램 크기 조정 커서를 히스토그램 테두리 위에 놓아 양방향 화살표가 표시되도록 합니다. 그런 다음 테두리를 클릭한 채로 드래그합니다.

축 배율 조정 축 배율을 조정하려면 축을 클릭한 채로 드래그합니다.

또는 축 위에 커서를 놓아 손 모양 포인터가 표시되도록 합니다. 그런 다음 축을 두 번 클릭하고 "축 설정" 창에서 파라미터를 설정합니다.

히스토그램 막대 크기 조정 (연속형 변수에만 사용 가능) 여러 가지 옵션으로 히스토그램 막대의 크기를 조정할 수 있습니다. 자세한 내용은 "[연속형 변수의 히스토그램 막대 크기 조정](#)"(38 페이지)에서 확인하십시오.

선택 항목 지정 여러 히스토그램에서 선택한 데이터를 지정할 수 있습니다. 자세한 내용은 "[여러 히스토그램에서 선택 영역 지정](#)"(40 페이지)에서 확인하십시오.

히스토그램 또는 관련 데이터 테이블에 사용 가능한 추가 옵션을 보려면 다음을 수행하십시오.

- 히스토그램을 마우스 오른쪽 버튼으로 클릭합니다. 자세한 내용은 **JMP 사용**에서 확인하십시오.
- 축을 마우스 오른쪽 버튼으로 클릭합니다. 라벨을 추가하거나 축을 수정할 수 있습니다. 자세한 내용은 **JMP 사용**의 "JMP 보고서" 창에서 확인하십시오.
- 변수 옆의 빨간색 삼각형을 클릭하고 **히스토그램 옵션**을 선택합니다. 옵션은 변수의 모델링 유형에 따라 조금씩 다릅니다. 자세한 내용은 "[범주형 변수에 대한 옵션](#)"(44 페이지) 또는 "[연속형 변수에 대한 옵션](#)"(46 페이지)에서 확인하십시오.

연속형 변수의 히스토그램 막대 크기 조정

다음을 사용하여 연속형 변수의 히스토그램 막대 크기를 조정할 수 있습니다.

- 손 도구
- **계급 폭 설정** 옵션
- **증분** 옵션

손 도구 사용

손 도구를 사용하면 데이터를 빠르게 탐색할 수 있습니다.

1. **도구 > 손 도구**를 선택합니다.

참고 : (Windows에만 해당) 메뉴 표시줄을 표시하려면 창 제목 아래의 막대 위에 커서를 놓아야 합니다. **파일 > 환경 설정 > Windows 관련**에서 이 설정을 변경할 수도 있습니다.

2. 손 도구를 히스토그램의 아무 곳이나 놓습니다.
3. 히스토그램 막대를 클릭한 채로 드래그합니다.

각 막대는 여러 관측값을 포함하는 계급으로 생각할 수 있습니다. 세로 히스토그램의 경우는 다음과 같습니다.

- 손 도구를 왼쪽으로 이동하면 계급 폭이 늘어나고 구간이 결합됩니다. 막대 크기가 커지면 막대 수는 줄어듭니다.
- 손 도구를 오른쪽으로 이동하면 계급 폭이 줄어들어 더 많은 막대가 생성됩니다.

- 손 도구를 위 또는 아래로 이동하면 축에서의 계급 위치가 이동되어 각 계급의 내용 및 크기가 변경됩니다.

참고 : 히스토그램 방향을 가로로 변경한 경우에는 위의 방향을 반대로 적용하십시오 . 즉, 손 도구를 아래로 이동하면 계급 폭이 늘어나고, 위로 이동하면 계급 폭이 줄어들며, 왼쪽 또는 오른쪽으로 이동하면 축에서의 계급 위치가 이동됩니다 .

계급 폭 설정 옵션 사용

계급 폭 설정 옵션을 사용하면 히스토그램에 있는 모든 막대의 너비를 보다 정확하게 설정할 수 있습니다 . " 계급 폭 설정 " 옵션을 사용하려면 변수의 빨간색 삼각형 메뉴에서 **히스토그램 옵션 > 계급 폭 설정**을 선택합니다 . 그런 다음 계급 폭 값을 변경합니다 .

증분 옵션 사용

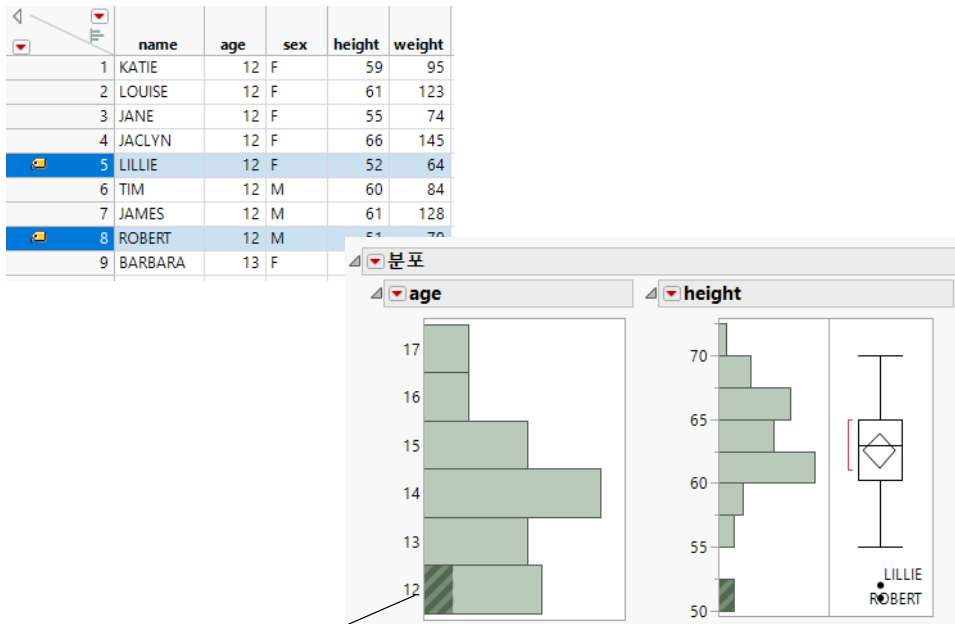
증분 옵션으로도 막대 너비를 정확하게 설정할 수 있습니다 . **증분** 옵션을 사용하려면 축을 두 번 클릭하고 " 증분 " 값을 변경합니다 .

막대 강조 표시 및 행 선택

히스토그램 막대를 클릭하면 해당 막대가 강조 표시되고 데이터 테이블에서 해당하는 행이 선택됩니다 . 다른 모든 그래픽 표시의 해당 부분에서도 선택 항목이 강조 표시됩니다 . 그림 3.6 에서는 **height** 히스토그램의 막대를 강조 표시한 결과를 보여 줍니다 . 데이터 테이블에서는 해당 행이 선택됩니다 .

팁 : 특정 히스토그램 막대를 선택 취소하려면 **Ctrl** 키를 누른 채로 강조 표시된 막대를 클릭하십시오 .

그림 3.6 막대 및 행 강조 표시



막대를 선택하면 다른 출력에서 해당 행 및 요소가 강조 표시됩니다.

여러 히스토그램에서 선택 영역 지정

다음과 같은 방법으로 여러 히스토그램에서 선택 영역을 넓히거나 좁힐 수 있습니다.

- 선택 영역을 넓히려면 Shift 키를 누른 채로 다른 막대를 선택합니다. 이 방법은 *or* 연산자를 사용하는 것과 동등합니다.
- 선택 영역을 좁히려면 Ctrl 키와 Alt 키 (Windows), 또는 Command 키와 Option 키 (macOS)를 누른 채로 다른 막대를 선택합니다. 이 방법은 *and* 연산자를 사용하는 것과 동등합니다.

예는 "예 : 여러 히스토그램에서 데이터 선택"(65 페이지)에서 확인하십시오.

빈도 보고서

명목형 및 순서형 변수의 경우 "빈도" 보고서에 변수의 수준과 함께 관련 발생 빈도 및 확률이 나열됩니다.

"빈도" 보고서에는 범주형 (명목형 또는 순서형) 변수의 각 수준별로 다음 목록에 설명된 정보가 포함됩니다. 결측값은 분석에서 생략됩니다.

팁 : "빈도" 보고서에서 값을 클릭하면 히스토그램과 데이터 테이블에서 해당하는 데이터가 선택됩니다.

수준 반응 변수에 대해 발견된 각 값을 나열합니다.

개수 반응 변수의 각 수준에 대해 발견된 행 수를 나열합니다. "빈도" 변수를 사용할 경우 이 개수는 반응 변수의 각 수준에 대한 빈도 변수의 합계입니다.

확률 반응 변수의 각 수준에 대한 발생 확률 (또는 비율) 을 나열합니다. 이 확률은 각 수준의 개수를 테이블 맨 아래에 표시된 변수의 총 빈도로 나눠서 계산됩니다.

표준 오차 확률 확률의 표준 오차를 나열합니다. 이 열이 숨겨져 있을 수도 있습니다. 이 열을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열 > 표준 오차 확률**을 선택합니다.

누적 확률 확률 열의 누적 합계를 포함합니다. 이 열이 숨겨져 있을 수도 있습니다. 이 열을 표시하려면 테이블을 마우스 오른쪽 버튼으로 클릭하고 **열 > 누적 확률**을 선택합니다.

분위수 보고서

연속형 변수의 경우 "분위수 보고서"에 선택된 분위수 (백분위수라고도 함) 의 값이 나열됩니다. 통계 상세 정보는 "[분위수](#)"(74 페이지)에서 확인하십시오.

요약 통계량 보고서

연속형 변수의 경우 "요약 통계량" 보고서에 평균, 표준편차 및 기타 요약 통계량이 표시됩니다. "요약 통계량" 옆의 빨간색 삼각형 메뉴에서 **요약 통계량 사용자 정의**를 선택하면 이 보고서에 표시되는 통계량을 제어할 수 있습니다.

팁 : 연속형 변수에 대한 분포 분석을 실행할 때마다 보고서에 표시할 요약 통계량을 지정할 수 있습니다. **파일 > 환경 설정 > 플랫폼 > 분포 요약 통계량**을 선택하고 표시하려는 요약 통계량을 선택하십시오.

- "[요약 통계량 보고서 설명](#)"에서는 기본적으로 표시되는 통계량에 대해 설명합니다.
- "[추가 요약 통계량](#)"에서는 **요약 통계량 사용자 정의** 창을 사용하여 보고서에 추가할 수 있는 기타 통계량에 대해 설명합니다.

요약 통계량 보고서 설명

평균 반응 변수에 대해 주어진 분포하에서의 기대값을 추정합니다. 이 값은 열 값의 산술평균입니다. 즉, 비결측값의 합을 비결측값의 개수로 나눈 값입니다.

표준편차 정규 분포는 주로 평균과 표준편차로 정의됩니다. 이러한 모수를 사용하면 표본 크기가 커질 경우 데이터를 손쉽게 요약할 수 있습니다.

- 값의 68%는 평균의 1 표준편차 범위 내에 있습니다.
- 값의 95%는 평균의 2 표준편차 범위 내에 있습니다.
- 값의 99.7%는 평균의 3 표준편차 범위 내에 있습니다.

평균 표준 오차 평균의 표준 오차로, 평균 분포의 표준편차를 추정합니다.

평균의 신뢰 상한 / 하한 평균에 대한 95% 신뢰 한계로, 실제 모집단 평균을 포함할 가능성이 매우 높은 구간을 정의합니다.

N 비결측값의 총 개수입니다.

추가 요약 통계량

가중합 시작 창에서 "가중치" 역할에 할당된 열의 합계입니다. 가중합은 평균 계산 시 N 을 대신해서 분모로 사용됩니다.

합 반응 값의 합입니다.

분산 표본 분산으로, 표본 표준편차의 제곱입니다.

왜도 편중성 또는 대칭성을 측정합니다.

첨도 꼬리의 뾰족함 또는 두꺼움을 측정합니다. 계산식에 대한 자세한 내용은 "[첨도](#)"(76 페이지)에서 확인하십시오.

CV 변동 계수 백분율로, 표준편차를 평균으로 나눈 후 100 을 곱해서 계산됩니다. 변동 계수는 상대적 변동을 평가하는 데 사용할 수 있습니다. 예를 들어 서로 다른 단위 또는 서로 다른 크기로 측정된 데이터의 변동을 비교할 때 이 통계량을 사용할 수 있습니다.

결측값 수 결측 관측값의 개수입니다.

0 개수 0 값의 개수입니다.

고유 값 수 고유 값의 개수입니다.

비수정 SS 비수정 제곱합, 즉 제곱 값의 합입니다.

수정 SS 수정 제곱합, 즉 평균으로부터의 편차를 제곱한 값의 합입니다.

자기상관 ("빈도" 변수를 지정하지 않은 경우에만 표시됨) 잔차가 행 간에 상관되어 있는지 여부를 검정하는 1차 자기상관입니다. 이 검정은 데이터의 비랜덤성을 감지하는 데 유용합니다.

최소값 데이터의 0 번째 백분위수를 나타냅니다.

최대값 데이터의 100 번째 백분위수를 나타냅니다.

중앙값 데이터의 50 번째 백분위수를 나타냅니다.

최빈값 데이터에서 발생 빈도가 가장 큰 값입니다. 최빈값이 여러 개이면 가장 작은 최빈값이 표시됩니다.

절사평균 데이터의 하위 $p\%$ 와 상위 $p\%$ 를 제거한 후 계산된 평균입니다. p 값은 창의 맨 아래에 있는 **절사평균 백분율 입력** 텍스트 상자에 입력합니다. "가중치" 변수를 지정한 경우에는 "절사평균" 옵션을 사용할 수 없습니다.

기하평균 데이터 곱의 n 제곱근입니다. 예를 들어 기하평균은 이자율을 계산하는 데 종종 사용됩니다. 이 통계량은 데이터에 많은 값이 비대칭적 분포로 포함된 경우에도 유용합니다.

참고 : 음수 값이 있으면 결측값이 발생하고, 음수가 아니고 0 인 값이 있으면 결과가 0 이 됩니다.

범위 데이터 최대값과 최소값 사이의 차이입니다.

사분위수 범위 3 사분위수와 1 사분위수 사이의 차이입니다.

중앙 절대 편차 ("가중치" 변수를 지정한 경우에는 표시되지 않음) 중앙값과의 절대 편차에 대한 중앙값입니다.

0 인 비율 0 인 비결측값의 비율입니다.

0 이 아닌 비율 0 이 아닌 비결측값의 비율입니다.

로버스트 평균 로버스트 평균은 Huber M 추정을 사용하여 이상치의 영향을 받지 않는 방식으로 계산됩니다. 자세한 내용은 Huber and Ronchetti 연구 자료 (2009) 에서 확인하십시오.

로버스트 표준편차 로버스트 표준편차는 Huber M 추정을 사용하여 이상치의 영향을 받지 않는 방식으로 계산됩니다. 자세한 내용은 Huber and Ronchetti 연구 자료 (2009) 에서 확인하십시오.

평균 신뢰 구간의 (1- α) 입력 평균 신뢰 구간의 유의 수준을 지정합니다.

절사평균 백분율 입력 절사평균 백분율을 지정합니다. 데이터의 양쪽에서 이 백분율만큼 절사됩니다.

요약 통계량 옵션

"요약 통계량" 옆의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

요약 통계량 사용자 정의 표시하려는 통계량을 목록에서 선택합니다. 모든 요약 통계량을 선택하거나 선택 취소할 수 있습니다.

모든 최빈값 표시 최빈값이 여러 개 있는 경우 최빈값을 모두 표시합니다.

통계 상세 정보는 "[요약 통계량](#)"(75 페이지) 에서 확인하십시오.

분포 플랫폼 옵션

"분포" 옆의 빨간색 삼각형 메뉴에는 분포 플랫폼의 모든 보고서 및 그래프 수에 영향을 주는 옵션이 포함되어 있습니다.

균등 척도 분포를 쉽게 비교할 수 있도록 모든 축 척도의 최소값, 최대값 및 간격을 동일하게 설정합니다.

쌓기 히스토그램과 보고서의 방향을 가로로 변경하고 개별 분포 보고서를 세로로 쌓습니다. 보고서 창을 원래 레이아웃으로 되돌리려면 이 옵션을 선택 취소합니다.

행으로 배열 한 행에 표시할 그림의 개수를 입력합니다. 이 옵션을 사용하면 여러 그림을 하나의 넓은 행으로 표시하지 않고 세로로 표시할 수 있습니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

범주형 변수에 대한 옵션

보고서 창에서 각 변수 옆에 있는 빨간색 삼각형 메뉴에는 해당 변수에 적용되는 추가 옵션이 포함되어 있습니다. 이 섹션에서는 범주형 (명목형 또는 순서형) 변수에 사용할 수 있는 옵션을 설명합니다.

표시 옵션 자세한 내용은 "**범주형 변수에 대한 표시 옵션**"(44 페이지)에서 확인하십시오.

히스토그램 옵션 자세한 내용은 "**범주형 변수에 대한 히스토그램 옵션**"(45 페이지)에서 확인하십시오.

모자이크 그림 각각의 명목형 또는 순서형 반응 변수에 대한 모자이크 막대 차트를 표시합니다. 모자이크 그림은 각 세그먼트가 해당 그룹의 빈도 수에 비례하는 누적 막대 차트입니다.

정렬 기준 히스토그램, 모자이크 그림 및 빈도 보고서를 오름차순 / 내림차순 또는 개수별로 재정렬합니다. 새 순서를 열 특성으로 저장하려면 **저장 > 값 순서화** 옵션을 사용합니다.

검정 확률 가설 확률을 검정하는 보고서를 표시합니다. 자세한 내용은 "**검정 확률 옵션의 예**"(67 페이지)에서 확인하십시오.

신뢰 구간 이 메뉴에는 신뢰 구간이 포함되어 있습니다. 나열된 값 중에서 선택하거나, **기타**를 클릭한 후 값을 직접 입력합니다. 그러면 JMP가 점수 신뢰 구간을 계산합니다.

저장 자세한 내용은 "**범주형 변수에 대한 저장 옵션**"(45 페이지)에서 확인하십시오.

제거 해당 변수와 모든 관련 보고서를 분포 보고서에서 영구적으로 제거합니다.

범주형 변수에 대한 표시 옵션

빈도 "빈도" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "**빈도 보고서**"(40 페이지)에서 확인하십시오.

가로 레이아웃 히스토그램과 보고서의 방향을 세로 또는 가로로 변경합니다.

왼쪽의 축 개수, 확률 및 밀도 축을 오른쪽 대신 왼쪽으로 이동합니다.

이 옵션은 **가로 레이아웃**이 선택된 경우에만 적용할 수 있습니다.

범주형 변수에 대한 히스토그램 옵션

히스토그램 히스토그램을 표시하거나 숨깁니다. 자세한 내용은 "**히스토그램**"(37 페이지)에서 확인하십시오.

세로 히스토그램의 방향을 세로에서 가로 방향으로 변경합니다.

표준 오차 막대 히스토그램의 각 수준에 표준 오차 막대를 그립니다.

막대 분리 히스토그램 막대를 분리합니다.

히스토그램 색상 히스토그램 막대의 색상을 변경합니다.

개수 축 히스토그램 막대가 나타내는 열 값의 빈도를 보여 주는 축을 추가합니다.

확률 축 히스토그램 막대가 나타내는 열 값의 비율을 보여 주는 축을 추가합니다.

밀도 축 히스토그램에 막대 길이를 보여 주는 축을 추가합니다.

개수 및 확률 축은 다음과 같은 계산을 기준으로 합니다.

확률 = (막대 너비) * 밀도

개수 = (막대 너비) * 밀도 * (총 개수)

백분율 표시 각 히스토그램 막대가 나타내는 열 값의 백분율을 라벨로 표시합니다.

참고: 소수 자릿수를 지정하려면 히스토그램을 마우스 오른쪽 버튼으로 클릭한 후 **사용자 정의 > 히스토그램**을 선택하십시오.

개수 표시 각 히스토그램 막대가 나타내는 열 값의 빈도를 라벨로 표시합니다.

범주형 변수에 대한 저장 옵션

수준 수 데이터 테이블에 수준 <열 이름>이라는 새 열을 생성합니다. 각 관측값의 수준 수는 해당 관측값이 포함된 히스토그램 막대에 해당합니다.

값 순서화 (**정렬 기준** 옵션과 함께 사용) 새 순서를 반영하여 데이터 테이블에 새로운 값 순서화 열 특성을 생성합니다.

로그에 스크립트로 현재 보고서를 생성하기 위한 스크립트 명령을 로그 창에 표시합니다. 로그 창을 보려면 **보기 > 로그**를 선택합니다.

연속형 변수에 대한 옵션

보고서 창에서 각 변수 옆에 있는 빨간색 삼각형 메뉴에는 해당 변수에 적용되는 추가 옵션이 포함되어 있습니다. 이 섹션에서는 연속형 변수에 사용할 수 있는 옵션을 설명합니다.

표시 옵션 자세한 내용은 "[연속형 변수에 대한 표시 옵션](#)"(47 페이지)에서 확인하십시오.

히스토그램 옵션 자세한 내용은 "[연속형 변수에 대한 히스토그램 옵션](#)"(47 페이지)에서 확인하십시오.

정규 분위수 그림 변수의 정규 분포 정도를 쉽게 시각화할 수 있습니다. 자세한 내용은 "[정규 분위수 그림](#)"(48 페이지)에서 확인하십시오.

이상치 상자 그림 분포를 표시하며 가능한 이상치를 쉽게 확인할 수 있습니다. 자세한 내용은 "[이상치 상자 그림](#)"(49 페이지)에서 확인하십시오.

분위수 상자 그림 "분위수" 보고서의 특정 분위수를 표시합니다. 자세한 내용은 "[분위수 상자 그림](#)"(50 페이지)에서 확인하십시오.

줄기 - 잎 자세한 내용은 "[줄기 - 잎](#)"(52 페이지)에서 확인하십시오.

CDF 그림 경험적 누적 분포 함수 그림을 생성합니다. 자세한 내용은 "[CDF 그림](#)"(52 페이지)에서 확인하십시오.

평균 검정 평균에 대한 1 표본 검정을 수행합니다. 자세한 내용은 "[평균 검정](#)"(53 페이지)에서 확인하십시오.

표준편차 검정 표준편차에 대한 1 표본 검정을 수행합니다. 자세한 내용은 "[표준편차 검정](#)"(54 페이지)에서 확인하십시오.

동등성 검정 모집단 평균이 가설 값과 동등한지 평가합니다. 자세한 내용은 "[동등성 검정](#)"(55 페이지)에서 확인하십시오.

신뢰 구간 평균 및 표준편차의 신뢰 구간을 선택합니다. 자세한 내용은 "[신뢰 구간](#)"(56 페이지)에서 확인하십시오.

예측 구간 단일 관측값이나 다음 무작위 선택 표본의 평균 및 표준편차에 대한 예측 구간을 선택합니다. 자세한 내용은 "[예측 구간](#)"(56 페이지)에서 확인하십시오.

공차 구간 모집단의 지정된 비율 이상을 포함하기 위한 구간을 계산합니다. 자세한 내용은 "[공차 구간](#)"(56 페이지)에서 확인하십시오.

공정 능력 지정된 규격 한계에 대한 공정 적합성을 측정합니다. 자세한 내용은 "[공정 능력](#)"(57 페이지)에서 확인하십시오.

연속형 적합 연속형 변수에 분포를 적합시킵니다. 자세한 내용은 "[분포 적합](#)"(59 페이지)에서 확인하십시오.

이산형 적합 (모든 데이터 값이 정수인 경우에만 사용 가능) 이산형 변수에 분포를 적합시킵니다. 자세한 내용은 "[분포 적합](#)"(59 페이지)에서 확인하십시오.

저장 연속형 또는 범주형 변수에 대한 정보를 저장합니다. 자세한 내용은 "[예측 구간](#)"(56 페이지)에서 확인하십시오.

제거 해당 변수와 모든 관련 보고서를 분포 보고서에서 영구적으로 제거합니다.

연속형 변수에 대한 표시 옵션

분위수 "분위수" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "[분위수 보고서](#)"(41 페이지)에서 확인하십시오.

분위수 증분 설정 분위수 증분을 변경하거나 기본 분위수 증분으로 되돌립니다.

사용자 분위수 값 또는 증분을 기준으로 사용자 분위수를 설정합니다. 신뢰 수준을 지정하고, 평화된 경험적 우도 분위수를 계산할지 여부를 선택할 수 있습니다. 데이터 집합이 큰 경우에는 다소 시간이 걸릴 수 있습니다.

- 가중 평균 분위수의 추정 방법에 대한 자세한 내용은 "[분위수](#)"(74 페이지)에서 확인하십시오.
- 가중 평균 분위수의 분포 무관 신뢰 한계에 대한 자세한 내용은 Meeker et al. 연구 자료 (2017)의 섹션 5.2에서 확인하십시오.
- 평화된 경험적 우도 분위수는 커널 밀도 추정값을 기반으로 합니다. 이러한 분위수와 해당 신뢰 한계에 대한 자세한 내용은 Chen and Hall 연구 자료 (1993)에서 확인하십시오.
- 소수 빈도를 사용하는 경우에는 신뢰 구간과 평화된 경험적 우도 분위수를 사용할 수 없습니다.

요약 통계량 "요약 통계량" 보고서를 표시하거나 숨깁니다. 자세한 내용은 "[요약 통계량 보고서](#)"(41 페이지)에서 확인하십시오.

요약 통계량 사용자 정의 "요약 통계량" 보고서의 통계량을 추가하거나 제거합니다. 자세한 내용은 "[요약 통계량 보고서](#)"(41 페이지)에서 확인하십시오.

가로 레이아웃 히스토그램과 보고서의 방향을 세로 또는 가로로 변경합니다.

왼쪽의 축 개수, 확률, 밀도 및 정규 분위수 그림 축을 오른쪽 대신 왼쪽으로 이동합니다.

이 옵션은 가로 레이아웃이 선택된 경우에만 적용할 수 있습니다.

연속형 변수에 대한 히스토그램 옵션

히스토그램 히스토그램을 표시하거나 숨깁니다. 자세한 내용은 "[히스토그램](#)"(37 페이지)에서 확인하십시오.

새도그램 히스토그램을 새도그램으로 바꿉니다. 히스토그램은 계급 폭이 변경될 경우 히스토그램 모양이 변경되지만 새도그램은 계급 폭을 달리 한 여러 히스토그램을 중첩 표시한 것입니다. 새도그램에서는 분포의 주요 특징이 덜 뚜렷하게 나타납니다.

새도그램에서는 다음 옵션을 사용할 수 없습니다.

- 표준 오차 막대

- 개수 표시
- 백분율 표시

세로 히스토그램의 방향을 세로에서 가로 방향으로 변경합니다.

표준 오차 막대 표준 오차를 사용하여 히스토그램의 각 수준에 표준 오차 막대를 그립니다. 손 도구로 막대 수를 조정하면 표준 오차 막대가 자동으로 조정됩니다. 자세한 내용은 "[연속형 변수의 히스토그램 막대 크기 조정](#)"(38 페이지) 및 "[표준 오차 막대](#)"(74 페이지)에서 확인하십시오.

계급 폭 설정 히스토그램 막대의 계급 폭을 변경합니다. 자세한 내용은 "[연속형 변수의 히스토그램 막대 크기 조정](#)"(38 페이지)에서 확인하십시오.

히스토그램 색상 히스토그램 막대의 색상을 변경합니다.

개수 축 히스토그램 막대가 나타내는 열 값의 빈도를 보여 주는 축을 추가합니다.

참고 : 히스토그램 막대의 크기를 조정하면 개수 축의 크기도 조정됩니다.

확률 축 히스토그램 막대가 나타내는 열 값의 비율을 보여 주는 축을 추가합니다.

참고 : 히스토그램 막대의 크기를 조정하면 확률 축의 크기도 조정됩니다.

밀도 축 밀도는 히스토그램의 막대 길이입니다. 개수와 확률은 모두 다음과 같은 계산을 기준으로 합니다.

$$\text{확률} = (\text{막대 너비}) * \text{밀도}$$

$$\text{개수} = (\text{막대 너비}) * \text{밀도} * (\text{총 개수})$$

적합 분포에 의해 추가된 밀도 곡선을 살펴볼 때 밀도 축에는 곡선의 점 추정값이 표시됩니다.

참고 : 히스토그램 막대의 크기를 조정하면 밀도 축의 크기도 조정됩니다.

백분율 표시 각 히스토그램 막대가 나타내는 열 값의 비율을 라벨로 표시합니다.

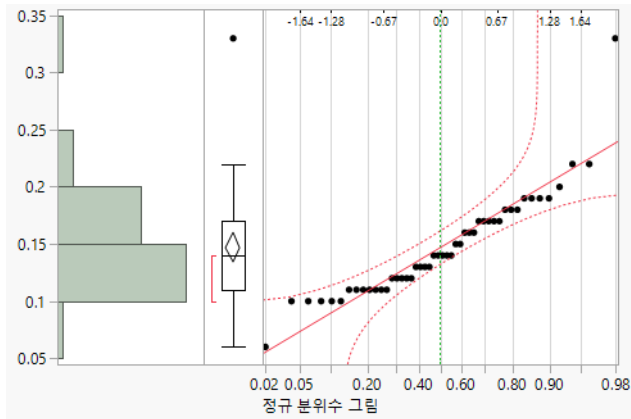
개수 표시 각 히스토그램 막대가 나타내는 열 값의 빈도를 라벨로 표시합니다.

정규 분위수 그림

변수의 정규 분포 정도를 시각화하려면 **정규 분위수 그림** 옵션을 사용합니다. 변수가 정규 분포된 경우 정규 분위수 그림은 대각 직선에 근사합니다. 이 유형의 그림을 분위수 - 분위수 그림 또는 Q-Q 그림이라고도 합니다.

정규 분위수 그림에서는 Lilliefors 신뢰경계 (Conover 1980) 와 확률 및 정규 분위수 척도도 보여 줍니다.

그림 3.7 정규 분위수 그림



다음 사항에 유의하십시오 .

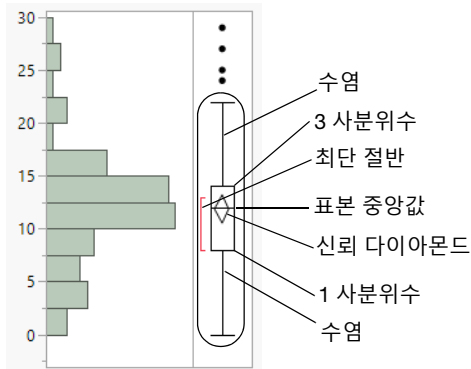
- 세로 축은 열 값을 보여 줍니다 .
- 위쪽 가로 축은 정규 분위수 척도를 보여 줍니다 .
- 아래쪽 가로 축은 각 값의 경험적 누적 확률을 보여 줍니다 .
- 빨간색 파선은 Lilliefors 신뢰경계를 보여 줍니다 .

통계 상세 정보는 " [정규 분위수 그림](#) "(76 페이지) 에서 확인하십시오 .

이상치 상자 그림

분포를 보고 가능한 이상치를 확인하려면 이상치 상자 그림 (Tukey 이상치 상자 그림이라고도 함) 을 사용합니다 . 일반적으로 상자 그림에는 연속형 분포 중 선택된 분위수가 표시됩니다 .

그림 3.8 이상치 상자 그림



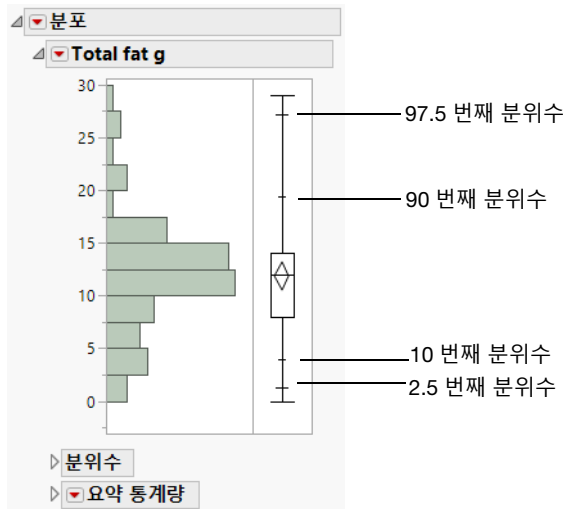
이상치 상자 그림의 경우 다음 사항에 유의하십시오 .

- 상자 내의 가로 선은 표본 중앙값을 나타냅니다 .
- 신뢰 다이아몬드에는 평균과 평균의 95% 상한 및 하한이 포함됩니다 . 다이아몬드의 중앙을 지나는 선을 그리면 평균을 알 수 있습니다 . 다이아몬드의 위쪽 및 아래쪽 점은 평균의 95% 상한 및 하한을 나타냅니다 .
- 상자의 끝은 25 번째 및 75 번째 분위수를 나타내며 각각 1 사분위수 및 3 사분위수라고도 합니다 .
- 1 사분위수와 3 사분위수 사이의 차이를 사분위수 범위 . 라고 합니다 .
- 상자에는 양 끝에서 시작되는 선이 있으며 이 선은 수염이라고도 합니다 . 수염은 상자 끝에서 시작해서 다음과 같이 계산된 거리 내에 있는 가장 바깥쪽 데이터 점으로 이어집니다 .
 $1 \text{ 사분위수} - 1.5 * (\text{사분위수 범위})$
 $3 \text{ 사분위수} + 1.5 * (\text{사분위수 범위})$
 데이터 점이 계산된 범위에 포함되지 않으면 상한 및 하한 데이터 점 값 (이상치 제외) 에 따라 수염이 결정됩니다 .
- 상자 외부의 괄호는 관측값 중 밀도가 가장 높은 50% 에 해당하는 최단 절반을 나타냅니다 (Rousseeuw and Leroy 1987).
- 이상치 상자 그림에서 개체를 제거하는 방법은 "[이상치 또는 분위수 상자 그림에서 개체 제거](#)"(51 페이지) 에서 확인하십시오 .

분위수 상자 그림

분위수 상자 그림에는 " 분위수 " 보고서의 특정 분위수가 표시됩니다 . 분포가 대칭적인 경우 상자 그림의 분위수는 서로 거의 같은 거리로 나타납니다 . 따라서 분포가 대칭적인지 여부를 한눈에 알 수 있습니다 . 예를 들어 분위수 표식이 한 쪽 끝에 밀집해서 모여 있지만 다른 쪽 끝에 더 큰 간격이 있는 경우 분포는 더 큰 간격을 갖는 끝 쪽으로 치우칩니다 .

그림 3.9 분위수 상자 그림



분위수는 값 중 하나로, p 번째 분위수는 $p\%$ 의 값보다 큰 값입니다. 예를 들어 데이터의 10%는 10 번째 분위수 아래에 있으며, 데이터의 90%는 90 번째 분위수 아래에 있습니다.

이상치 또는 분위수 상자 그림에서 개체 제거

이상치 또는 분위수 상자 그림에서 신뢰 다이아몬드와 최단 절반을 제거할 수 있습니다. 단일 그래프에서 제거할 수도 있고, 이후의 모든 그래프에서 제거할 수도 있습니다.

개별 그래프에서 개체를 제거하려면

1. 이상치 상자 그림을 마우스 오른쪽 버튼으로 클릭하고 **사용자 정의**를 선택합니다.
2. **상자 그림**을 클릭합니다.
3. **신뢰 다이아몬드** 또는 **최단 절반** 옆의 체크박스를 선택 취소합니다.

"그래프 사용자 정의" 창에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

이후의 모든 그래프에서 개체를 제거하려면

1. **파일 > 환경 설정 > 플랫폼 > 분포**를 선택합니다.
2. 다음 옵션을 선택 취소합니다.
 - **상자 그림 신뢰 다이아몬드 표시**
 - **이상치 상자 그림 최단 절반 표시**
3. **확인**을 클릭합니다.

이제 분포 플랫폼에서 추가하는 모든 상자 그림에는 신뢰 다이아몬드나 최단 절반이 포함되지 않습니다.

줄기 - 잎

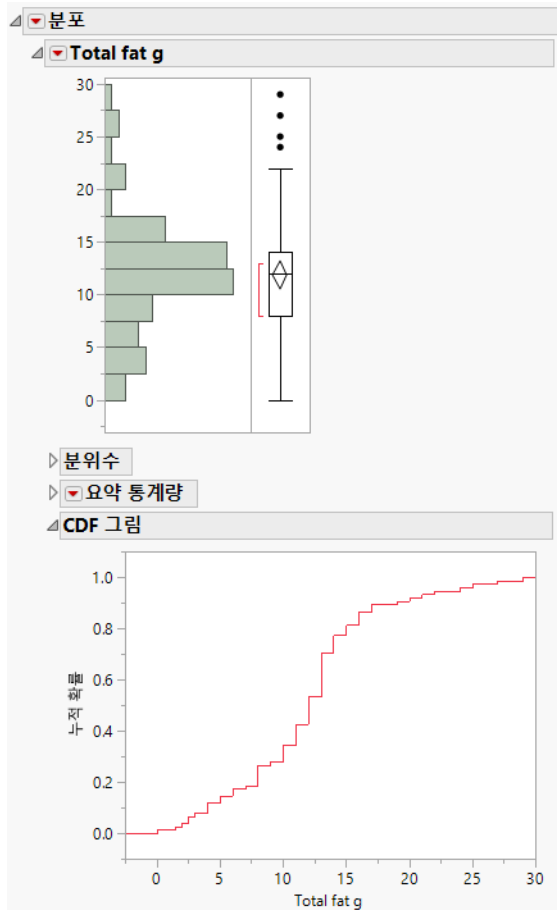
이 그림의 각 행에는 열 값 범위의 선행 숫자를 나타내는 줄기 값이 있습니다. 잎 값은 값의 뒷자리 숫자로 이루어집니다. 줄기와 잎을 결합하면 데이터 점을 확인할 수 있습니다. 일부 경우에 줄기 - 잎 그림의 숫자는 테이블의 실제 데이터에 대한 근사값입니다. 줄기 - 잎 그림은 마우스 버튼을 클릭하거나 브러시 도구를 사용할 때 그에 따라 변경됩니다.

참고: 줄기 - 잎 그림에서는 소수 빈도가 지정된 빈도 이상의 가장 작은 정수로 변환됩니다.

CDF 그림

CDF 그림은 경험적 누적 분포 함수 그림을 생성합니다. 가로 축의 특정 값보다 작거나 같은 데이터의 백분율을 확인하려면 CDF 그림을 사용합니다. 데이터에 분포가 적합되면 적합한 분포에 대한 누적 분포 함수가 CDF 그림에 중첩 표시됩니다.

그림 3.10 CDF 그림



예를 들어 이 CDF 그림에서 총 지방 값이 10g 인 데이터는 약 34% 입니다 .

평균 검정

평균에 대한 1 표본 검정 옵션을 지정하고 검정을 수행하려면 **평균 검정** 창을 사용합니다 . 표준 편차 값을 지정하면 z - 검정이 수행됩니다 . 그렇지 않으면 표본 표준편차를 사용하여 t - 검정이 수행됩니다 . 비모수 Wilcoxon 부호 순위 검정을 요청할 수도 있습니다 .

서로 다른 여러 값을 검정하려면 **평균 검정** 옵션을 반복적으로 사용합니다 . 평균을 검정할 때마다 새 " 평균 검정 " 보고서가 표시됩니다 .

평균 검정 보고서 설명

평균 검정을 위해 계산되는 통계량

t-검정 (또는 z-검정) 양측 및 단측 대립 가설에 대한 검정 통계량 값과 p 값을 나열합니다.

부호 순위 ("Wilcoxon 부호 순위"가 선택된 경우에만 표시됨) 양측 및 단측 대립 가설에 대한 Wilcoxon 부호 순위 통계량과 p 값을 차례로 나열합니다. 이 검정에서는 Pratt 방법을 사용하여 0 값을 처리합니다. 이 검정은 비모수 검정으로서 중앙값이 가정된 값과 같다는 것을 귀무 가설로 하며, 분포가 대칭적이라고 가정합니다. 자세한 내용은 "[Wilcoxon 부호 순위 검정](#)"(76 페이지)에서 확인하십시오.

확률 값

Prob > |t| 모집단 평균이 가설 값과 같을 때 절대 t 값이 관측된 t 값보다 클 확률입니다. 이 값은 양쪽 꼬리 t -검정의 관측된 유의성에 대한 p 값입니다.

Prob > t 모집단 평균이 가설 값과 다르지 않을 때 t 값이 계산된 표본 t 비보다 클 확률입니다. 이 값은 위쪽 꼬리 검정의 p 값입니다.

Prob < t 모집단 평균이 가설 값과 다르지 않을 때 t 값이 계산된 표본 t 비보다 작을 확률입니다. 이 값은 아래쪽 꼬리 검정의 p 값입니다.

평균 검정 옵션 설명

p 값 애니메이션 p 값에 대한 대화식 시각적 표현을 시작합니다. 가설 평균 값을 변경하면서 변경 사항이 p 값에 어떤 영향을 주는지 살펴볼 수 있습니다.

검정력 애니메이션 검정력 및 베타에 대한 대화식 시각적 표현을 시작합니다. 가설 평균과 표본 평균을 변경하면서 변경 사항이 검정력과 베타에 어떤 영향을 주는지 살펴볼 수 있습니다.

검정 제거 평균 검정을 제거합니다.

표준편차 검정

표준편차에 대한 1 표본 검정을 수행하려면 **표준편차 검정** 옵션을 사용합니다 (Neter et al. 1990 참조). 서로 다른 여러 값을 검정하려면 **표준편차 검정** 옵션을 반복적으로 사용합니다. 표준편차를 검정할 때마다 새로운 "표준편차 검정" 보고서가 표시됩니다.

검정 통계량 카이제곱 검정 통계량의 값을 제공합니다. 자세한 내용은 "[표준편차 검정](#)"(78 페이지)에서 확인하십시오.

최소 p 값 모집단 표준편차가 가설 값과 다르지 않을 때 보다 극단적인 카이제곱 값을 얻을 확률입니다. 자세한 내용은 "[표준편차 검정](#)"(78 페이지)에서 확인하십시오.

Prob>ChiSq 모집단 표준편차가 가설 값과 다르지 않을 때 카이제곱 값이 계산된 표본 카이제곱보다 클 확률입니다. 이 값은 한쪽 꼬리 t -검정의 관측된 유의성에 대한 p 값입니다.

Prob<ChiSq 모집단 표준편차가 가설 값과 다르지 않을 때 카이제곱 값이 계산된 표본 카이제곱보다 작을 확률입니다. 이 값은 한쪽 꼬리 t -검정의 관측된 유의성에 대한 p 값입니다.

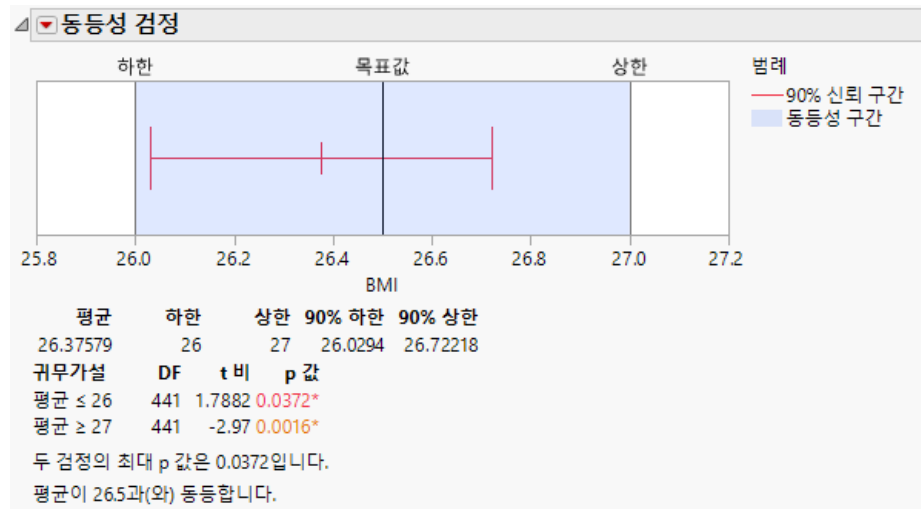
동등성 검정

동등성 검정에서는 모집단 평균이 가설 값과 동등한지 평가합니다. 차이가 없는 것과 동등하게 간주되는 임계 차이를 정의해야 합니다. "동등성 검정" 옵션은 TOST(Two One-Sided Tests: 두 개의 단측 검정) 방법을 사용합니다. 두 개의 단측 t 검정은 실제 평균과 가설 값 사이의 차이가 임계를 초과하는 귀무가설에 대해 구성됩니다. 두 귀무가설이 모두 기각되면 실제 차이가 임계를 초과하지 않는 것입니다. 이 경우 평균이 실제로 가설 값과 동등한 것으로 간주할 수 있다는 결론을 내릴 수 있습니다.

"동등성 검정" 옵션을 선택할 때는 가설 평균, 임계 차이 (실제로 0으로 간주되는 차이) 및 신뢰 수준을 지정합니다. 신뢰 수준은 $1-\alpha$ 이며, 여기서 α 는 각 단측 검정의 유의 수준입니다.

그림 3.11에서는 Diabetes.jmp 샘플 데이터 테이블의 BMI 변수에 대한 "동등성 검정" 보고서를 보여 줍니다. 가설 평균은 26.5이고, 실제로 0으로 간주되는 차이는 0.5로 지정되어 있습니다.

그림 3.11 동등성 검정 보고서



이 보고서에는 다음이 표시됩니다.

- "목표값"과 경계("하한" 및 "상한"이라는 라벨이 표시된 수직선으로 정의됨)를 보여주는 정의된 동등성 구간의 그림
- 계산된 평균의 신뢰 구간. 이 신뢰 구간은 $1-2\alpha$ 유의 수준 구간입니다.
- 계산된 평균, 지정된 하한 및 상한 경계, 평균의 $(1-2\alpha)$ 수준 신뢰 구간을 보여주는 테이블
- 두 개의 단측 검정 결과를 보여주는 테이블
- 결과를 요약하고, 평균을 목표값과 동등한 것으로 간주할 수 있는지 여부를 언급하는 노트

신뢰 구간

연속형 변수의 **신뢰 구간** 옵션은 평균과 표준편차의 신뢰 구간을 표시합니다. **0.90, 0.95** 및 **0.99** 옵션은 평균과 표준편차의 양측 신뢰 구간을 계산합니다. **신뢰 구간 > 기타** 옵션을 사용하여 신뢰 수준을 선택하고 단측 또는 양측 신뢰 구간을 선택할 수 있습니다. 알려진 시그마를 입력할 수도 있습니다. 알려진 시그마를 사용할 경우 평균의 신뢰 구간은 t 값이 아니라 z 값을 기준으로 합니다.

"신뢰 구간" 보고서에서는 평균 및 표준편차 모수 추정값과 함께 $1 - \alpha$ 에 대한 신뢰 상한 및 하한을 보여 줍니다.

예측 구간

예측 구간은 단일 관측값이나 다음 무작위 선택 표본의 평균 및 표준편차와 관련이 있습니다. 계산 시에는 정규 분포에서 특정 표본이 무작위로 선택되었다고 가정합니다. 단측 또는 양측 예측 구간을 선택합니다.

변수의 **예측 구간** 옵션을 선택하면 "예측 구간" 창이 나타납니다. 이 창을 사용하여 신뢰 수준, 미래 표본 수, 그리고 단측 또는 양측 한계를 지정합니다.

관련 정보

- 통계 상세 정보는 "**예측 구간**"(79 페이지)에서 확인하십시오.
- 예는 "**예측 구간의 예**"(69 페이지)에서 확인하십시오.

공차 구간

공차 구간에는 모집단의 지정된 비율 이상이 포함됩니다. 이 구간은 평균이나 표준편차가 아니라 모집단의 지정된 비율에 대한 신뢰 구간입니다. 공차 구간에 대한 자세한 내용은 Meeker et al. 연구 자료 (2017) 및 Tamhane and Dunlop 연구 자료 (2000)에서 확인할 수 있습니다.

변수의 **공차 구간** 옵션을 선택하면 "공차 구간" 창이 나타납니다. 이 창을 사용하여 신뢰 수준, 포함할 비율, 단측 또는 양측 한계, 방법을 지정합니다. 사용할 수 있는 두 가지 방법은 "정규 분포 가정"과 "비모수"입니다. "정규 분포 가정" 옵션은 표본이 정규 분포에서 무작위로 선택되었다는 가정에 기반한 공차 구간을 계산합니다. "비모수" 옵션은 분포 무관 공차 구간을 계산합니다.

관련 정보

- 통계 상세 정보는 "**공차 구간**"(80 페이지)에서 확인하십시오.
- 예는 "**공차 구간의 예**"(71 페이지)에서 확인하십시오.

공정 능력

공정 능력 분석에서는 지정된 규격 한계와 비교하여 공정이 얼마나 잘 수행되고 있는지를 측정합니다. 양호한 공정은 안정적이고 규격 한계 이내인 제품을 일관되게 생산하는 공정입니다. 공정 능력 지수는 공정 중심화 및 변동성별로 요약된 공정 성능을 규격 한계와 관련짓는 측정값입니다.

규격 한계

열에 "규격 한계" 열 특성이 포함되어 있고 시작 창에서 "공정 능력 생성" 옵션이 선택되었으면 "공정 능력" 보고서가 자동으로 생성됩니다. 열에 "분포" 열 특성도 포함되어 있지 않은 한 이 보고서는 정규 분포를 기반으로 합니다. 열에 "분포" 열 특성이 포함되어 있는 경우 "공정 능력" 보고서는 해당 열 특성에 지정된 분포를 기반으로 합니다.

팁 : 한 번에 몇 개의 열에 규격 한계를 추가하는 방법은 **Quality and Process Methods** 의 부록 "Statistical Details" 에서 확인하십시오 .

열에 규격 한계가 포함되어 있지 않으면 분석 변수 이름 옆의 빨간색 삼각형에서 **공정 능력**을 선택하고 "공정 능력 분석" 창에서 규격 한계를 설정합니다.

보고서의 규격 한계를 데이터 테이블에 열 특성으로 저장하려면 "공정 능력" 의 빨간색 삼각형에서 **규격 한계를 열 특성으로 저장**을 선택합니다. 공정 능력 분석을 반복할 때 저장된 규격 한계를 자동으로 가져오게 됩니다.

공정 능력 분석 창

분석 변수 이름 옆의 빨간색 삼각형에서 "공정 능력" 옵션을 선택하면 "공정 능력 분석" 창이 나타납니다.

"공정 능력 분석" 창을 사용하여 규격 한계, 분석을 위한 기본 분포, 시그마 추정 방법 등의 공정 능력 분석 옵션을 지정합니다. 공정 능력을 분석하려면 시그마 추정 방법과 그룹 내 (단기) 변동을 선택해야 합니다. 선택한 공정 능력 옵션에 따라 다른 하위 옵션이 나타납니다.

그림 3.12 공정 능력 분석 창

규격 한계 입력 공정 능력 분석을 위한 규격 하한, 목표값 및 규격 상한을 지정합니다. 이 중 적어도 하나는 비결측값이어야 합니다. "한계 표시" 옵션을 선택하면 분포 플랫폼 보고서의 히스토그램에 규격 한계가 나타납니다.

공정 능력 옵션 선택한 옵션에 따라 다른 추가 옵션이 나타납니다. 다음 옵션 중 하나를 선택합니다.

부분군 크기 = 1 부분군 크기를 1로 설정하고 추가 이동 범위 옵션을 제공합니다. 자세한 내용은 **Quality and Process Methods**의 "Process Capability"장에서 확인하십시오.

부분군 ID 열 사용 부분군 ID 열을 선택하고 추가 부분군 지정 및 이동 범위 옵션을 제공합니다. 자세한 내용은 **Quality and Process Methods**의 "Process Capability"장에서 확인하십시오.

일정한 부분군 크기 사용 일정한 부분군 크기를 설정하고 추가 부분군 지정 및 이동 범위 옵션을 제공합니다. 자세한 내용은 **Quality and Process Methods**의 "Process Capability"장에서 확인하십시오.

과거 시그마 사용 과거에 승인된 시그마 값을 할당합니다. 자세한 내용은 **Quality and Process Methods**의 "Process Capability"장에서 확인하십시오.

비정규 분포 사용 비정규 분포를 선택하고 추가 비정규 분포 옵션을 제공합니다. 자세한 내용은 **Quality and Process Methods**의 "Process Capability"장에서 확인하십시오.

유의 수준 지정 신뢰 한계의 유의 수준을 지정합니다.

공정 능력 분석 보고서

"공정 능력 분석" 창에서 "확인"을 클릭하면 선택한 변수에 대한 공정 능력 보고서가 포함된 공정 능력 보고서가 나타납니다. 이 보고서에 대한 자세한 내용은 **Quality and Process Methods**의 "Process Capability"장에서 확인하십시오.

- 통계 상세 정보는 Quality and Process Methods 의 "Process Capability" 장에서 확인하십시오 .
- 예는 " [공정 능력의 예](#) "(72 페이지) 에서 확인하십시오 .
- 공정 능력 플랫폼에 대한 자세한 내용은 Quality and Process Methods 의 "Process Capability" 장에서 확인하십시오 .

참고 : 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 공정 능력 보고서의 다양한 옵션에 대한 환경 설정을 지정할 수 있습니다 .

분포 적합

" 연속형 적합 " 또는 " 이산형 적합 " 하위 메뉴의 옵션을 사용하여 연속형 변수에 분포를 적합시킬 수 있습니다 . 연속형 변수에 분포를 적합시키면 히스토그램에 곡선이 중첩 표시되고 보고서 창에 분포 비교 보고서와 적합 분포 보고서가 추가됩니다 . 적합 분포 보고서의 빨간색 삼각형 메뉴에는 추가 옵션이 포함되어 있습니다 . 자세한 내용은 " [분포 적합 옵션](#) "(61 페이지) 에서 확인하십시오 .

참고 : 수명 분포 플랫폼에도 분포 적합을 위한 옵션이 포함되어 있으며 여기에는 다른 파라미터화 방법이 사용되고 중도절단된 관측값이 허용될 수 있습니다 . 자세한 내용은 Reliability and Survival Methods 의 "Life Distribution" 장에서 확인하십시오 .

연속형 적합

" 연속형 적합 " 하위 메뉴에는 연속형 분포를 적합시키기 위한 옵션이 포함되어 있습니다 . 이러한 분포의 파라미터화 방법에 대한 자세한 내용은 " [연속형 적합 분포](#) "(82 페이지) 에서 확인하십시오 .

정규 적합 데이터에 정규 분포를 적합시킵니다 . 정규 분포는 대개 대부분의 값이 곡선의 가운데에 있는 대칭적 데이터를 모델링하는 데 사용됩니다 .

Cauchy 적합 데이터에 Cauchy 분포를 적합시킵니다 . Cauchy 분포는 평균과 표준편차가 정의되어 있지 않습니다 . 대부분의 데이터가 본질적으로 Cauchy 분포를 따르는 하지만 데이터에 포함된 이상치의 비율이 높은 경우 (최대 50%) 데이터의 로버스트 위치 및 척도를 추정하는 데 이 분포가 특히 유용할 수 있습니다 .

SHASH 적합 데이터에 SHASH(SinH-ArcSinH) 분포를 적합시킵니다 . SHASH 분포는 정규 분포로 변환된다는 점에서 Johnson 분포와 유사하지만 SHASH 분포에는 정규 분포가 특수한 경우로 포함되어 있습니다 . 이 분포는 대칭적일 수도 있고 비대칭적일 수도 있습니다 .

지수 적합 (모든 관측값이 음수가 아닌 경우에만 사용 가능) 데이터에 지수 분포를 적합시킵니다 . 지수 분포는 오른쪽으로 편중된 분포로 , 대개 수명이나 연속된 이벤트 사이의 시간을 모델링하는 데 사용됩니다 .

감마 적합 (모든 관측값이 양수인 경우에만 사용 가능) 데이터에 감마 분포를 적합시킵니다 . 감마 분포는 양수 값을 모델링하기 위한 유연한 분포입니다 .

로그 정규 적합 (모든 관측값이 양수인 경우에만 사용 가능) 데이터에 로그 정규 분포를 적합시킵니다. 로그 정규 분포는 오른쪽으로 편중된 분포로, 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다.

Weibull 적합 (모든 관측값이 양수인 경우에만 사용 가능) 데이터에 Weibull 분포를 적합시킵니다. Weibull 분포는 유연한 분포로, 대개 수명이나 이벤트까지의 시간을 모델링하는 데 사용됩니다.

정규 2 혼합 적합 두 가지 정규 분포의 혼합 분포를 적합시킵니다. 이 유연한 분포는 이봉 분포 데이터를 적합시킬 수 있습니다.

정규 3 혼합 적합 세 가지 정규 분포의 혼합 분포를 적합시킵니다. 이 유연한 분포는 다봉 분포 데이터를 적합시킵니다.

평활 곡선 적합 비모수 밀도 추정 (커널 밀도 추정) 을 사용하여 평활 곡선을 적합시킵니다. "비모수 밀도" 보고서에 나타나는 슬라이더로 대역폭을 변경하여 평활 정도를 조절할 수 있습니다.

Johnson 적합 데이터에 Johnson 분포를 적합시킵니다. 세 가지 유형의 Johnson 분포 (Su, Sb 및 SI) 중 가장 적절한 분포가 적합되어 보고됩니다. Johnson 분포 계열은 왜도와 첨도의 가능한 모든 조합을 지원하므로 데이터 적합 기능에 유용합니다. Johnson 분포의 선택 프로시저 및 모수 추정에 대한 자세한 내용은 Slifker and Shapiro 연구 자료 (1980) 에서 확인할 수 있습니다.

베타 적합 (모든 관측값이 0 에서 1 사이인 경우에만 사용 가능) 데이터에 베타 분포를 적합시킵니다. 베타 분포는 0 에서 1 사이의 데이터를 모델링하는 데 유용하며, 대개 비율을 모델링하는 데 사용됩니다.

전체 적합 변수에 사용 가능한 모든 연속형 분포를 적합시킵니다. "분포 비교" 보고서에는 각 적합 분포에 대한 통계량이 포함됩니다. 기본적으로는 최량 적합 분포가 선택되어 히스토그램에 표시됩니다. 체크박스를 사용하여 선택한 분포에 대한 적합 보고서를 표시하거나 숨기고 분포 곡선을 중첩 표시할 수 있습니다. 기본적으로 "분포 비교" 목록은 AICc 를 기준으로 오름차순 정렬됩니다.

팁 : "분포" 열에서 분포 이름을 두 번 클릭하면 "분포 비교" 목록의 분포를 빠르게 제거할 수 있습니다. 이렇게 하면 해당하는 적합 분포 보고서도 제거됩니다.

레거시 적합기 사용 "레거시 적합기" 하위 메뉴를 표시하거나 숨깁니다. JMP 15에서는 몇 가지 분포 적합 기능이 업데이트되었습니다. 이 옵션을 활성화하면 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 사용할 수 있습니다. 자세한 내용은 "[레거시 분포 적합기에 대한 상세 정보](#)"(91 페이지)에서 확인하십시오.

이산형 적합

" 이산형 적합 " 하위 메뉴는 모든 데이터 값이 정수일 때 사용할 수 있습니다. " 이산형 적합 " 하위 메뉴에는 이산형 분포를 적합시키기 위한 옵션이 포함되어 있습니다. 이러한 분포의 파라미터화 방법에 대한 자세한 내용은 " **이산형 적합 분포** " (88 페이지) 에서 확인하십시오.

Poisson 적합 데이터에 Poisson 분포를 적합시킵니다. Poisson 분포는 특정 구간 내의 사건 수를 모델링하는 데 유용하며 대개 개수 데이터로 표현됩니다.

음이항 적합 데이터에 음이항 분포를 적합시킵니다. 음이항 분포는 지정된 실패 횟수에 도달하기 전의 성공 횟수를 모델링하는 데 유용합니다. 또한 음이항 분포는 감마 Poisson 분포와 동등합니다.

ZI Poisson 적합 (데이터에 0 값이 있는 경우에만 사용 가능) 데이터에 영과잉 Poisson 분포를 적합시킵니다. 영과잉 Poisson 분포에서는 표준 Poisson 분포에 비해 더 많은 비율의 데이터가 0 값이라고 가정합니다.

ZI 음이항 적합 (데이터에 0 값이 있는 경우에만 사용 가능) 데이터에 영과잉 음이항 분포를 적합시킵니다. 영과잉 음이항 분포에서는 표준 음이항 분포에 비해 더 많은 비율의 데이터가 0 값이라고 가정합니다.

이항 적합 데이터에 이항 분포를 적합시킵니다. 이항 분포는 n 번의 독립 시도 (단, 모든 시도의 성공 확률 p 는 고정됨) 시 총 성공 횟수를 모델링하는 데 사용됩니다. 표본 크기는 모든 관측값에 대해 고정된 표본 크기로 지정할 수도 있고, 각 행의 표본 크기가 포함된 데이터 테이블 내의 다른 열로 지정할 수도 있습니다.

참고: 일정하지 않은 표본 크기를 지정할 경우에는 밀도 곡선, 진단 그림 및 프로파일러를 사용할 수 없습니다.

베타 이항 적합 데이터에 베타 이항 분포를 적합시킵니다. 베타 이항 분포는 이항 분포의 과대산포 버전입니다. 이 분포를 적합시키려면 각 관측값에 대해 표본 크기가 1 보다 커야 합니다. 표본 크기는 모든 관측값에 대해 고정된 표본 크기로 지정할 수도 있고, 각 행의 표본 크기가 포함된 데이터 테이블 내의 다른 열로 지정할 수도 있습니다.

참고: 일정하지 않은 표본 크기를 지정할 경우에는 밀도 곡선, 진단 그림 및 프로파일러를 사용할 수 없습니다.

분포 적합 옵션

각 적합 분포 보고서에는 추가 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다.

밀도 곡선 분포의 추정 모수를 사용하여 히스토그램에 밀도 곡선을 중첩 표시합니다.

진단 그림 다음 옵션이 포함되어 있습니다.

QQ 그림 QQ (분위수 - 분위수) 그림을 표시하거나 숨깁니다. 이 그림은 추정 모수를 사용하여 구한 관측값과 분위수 간의 관계를 보여 줍니다.

PP 그림 PP(백분위수 - 백분위수) 그림을 표시하거나 숨깁니다 . 이 그림은 추정 모수를 사용하여 구한 경험적 CDF(누적 분포 함수) 와 적합 CDF 간의 관계를 보여 줍니다 .

프로파일러 다음 옵션이 포함되어 있습니다 .

분포 프로파일러 CDF(누적 분포 함수) 의 예측 프로파일러를 표시하거나 숨깁니다 .

분위수 프로파일러 분위수 함수의 예측 프로파일러를 표시하거나 숨깁니다 .

열 저장 다음 옵션이 포함되어 있습니다 .

밀도 계산식 저장 추정 모수 값을 사용하여 계산된 밀도 계산식이 포함된 열을 데이터 테이블에 저장합니다 .

분포 계산식 저장 추정 모수 값을 사용하여 계산된 CDF(누적 분포 함수) 가 포함된 열을 데이터 테이블에 저장합니다 .

시뮬레이션 계산식 저장 추정 모수를 사용하여 시뮬레이션된 값을 생성하는 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 이 열은 " 시뮬레이션 " 유틸리티에서 " 스위치 인할 열 " 로 사용할 수 있습니다 . 자세한 내용은 " [시뮬레이션](#) " 장 (305 페이지) 에서 확인하십시오 .

변환 저장 (Johnson 및 SHASH 분포 적합의 경우에만 사용 가능) 변환 계산식이 포함된 열을 데이터 테이블에 저장합니다 . 이 계산식은 적합된 분포를 사용하여 분석 열을 정규 분포로 변환하는 데 사용할 수 있습니다 .

적합도 (Johnson, 평활 곡선, 정규 혼합, 이항 또는 베타 음이항 분포에는 사용 불가능) 적합된 분포에 대한 시뮬레이션 기반 적합도 검정이 포함된 적합도 검정 보고서를 표시하거나 숨깁니다 . 연속형 적합의 경우에는 Anderson-Darling 검정이 포함됩니다 . 이산형 적합의 경우에는 Pearson 카이제곱 검정이 포함됩니다 . 정규 분포의 경우 표본 크기가 2,000 보다 작거나 같고 고정 모수가 없으면 Shapiro-Wilk 정규성 검정도 보고됩니다 .

모수 고정 (Johnson 분포 적합에는 사용 불가능) 모수를 고정하고 고정되지 않은 모수를 다시 추정할 수 있습니다 . 새 모수를 검정하여 해당 모수가 데이터를 적합시키는지 여부를 확인하는 " 적합성 LR 검정 "(LR: 우도비) 보고서도 표시됩니다 .

공정 능력 (Cauchy 또는 이산형 분포 적합에는 사용 불가능) 적합된 분포를 사용하여 공정 능력 분석을 생성할 수 있습니다 . 공정 능력 분석에서는 규격 한계를 기준으로 공정이 얼마나 잘 수행되고 있는지를 측정합니다 . 적합 분포의 빨간색 삼각형 메뉴에서 " 공정 능력 " 옵션을 선택하면 다음 옵션이 포함된 창이 나타납니다 .

규격 한계 입력 규격 한계를 수동으로 입력할 수 있습니다 . 적합된 분포를 사용하여 규격 한계를 계산하려면 이 섹션을 비워 두고 " 분위수 규격 한계 계산 옵션 " 아래의 옵션을 사용하십시오 .

분위수 규격 한계 계산 옵션 적합된 분포를 기준으로 규격 한계를 계산할 수 있습니다. 두 가지 방법을 사용할 수 있습니다.

첫 번째 방법에서는 적합된 분포의 분위수와 연관된 확률을 입력하여 규격 한계를 계산합니다.

두 번째 방법에서는 규격 한계를 계산하는 데 사용되는 K 시그마 승수 값을 입력합니다. 이 방법을 사용할 경우 양측 또는 단측 한계를 생성할 수 있습니다.

확률 또는 시그마 승수 값을 입력한 후 **규격 한계 계산**을 클릭하여 규격 한계를 계산합니다. 이러한 한계는 "규격 한계 입력" 패널에 입력됩니다. **확인**을 클릭하여 입력된 한계를 적용하고 공정 능력 보고서를 생성합니다.

공정 능력 옵션 다음 옵션이 포함되어 있습니다.

"이동 범위 옵션" 개요에는 이동 범위 통계량의 유형을 선택할 수 있는 옵션이 포함되어 있습니다. 자세한 내용은 **Quality and Process Methods**의 "Process Capability" 장에서 확인하십시오.

"비정규 분포 옵션" 개요에는 비정규 공정 능력 계산에 사용되는 방법을 선택할 수 있는 옵션이 포함되어 있습니다. 자세한 내용은 **Quality and Process Methods**의 "Process Capability" 장에서 확인하십시오.

공정 능력 옵션 및 보고서에 대한 자세한 내용은 **Quality and Process Methods**의 "Process Capability" 장에서 확인하십시오.

참고 : 파일 > 환경 설정 > 플랫폼 > 공정 능력에서 공정 능력 보고서의 다양한 옵션에 대한 환경 설정을 지정할 수 있습니다.

적합 제거 보고서 창에서 분포 적합을 제거합니다.

연속형 변수에 대한 저장 옵션

"저장" 메뉴의 옵션을 사용하여 연속형 변수에 대한 정보를 저장할 수 있습니다. 각 저장 옵션은 현재 데이터 테이블에 새 열을 생성합니다. 새 열의 이름은 저장 명령 이름에 변수 이름 (다음 정의에서 <열 이름>으로 표시)을 추가한 형식으로 지정됩니다 (표 3.1).

히스토그램 막대를 결합하기 전과 후처럼 서로 다른 상황에서 동일한 정보를 여러 번 저장하려면 저장 옵션을 반복해서 선택합니다. 저장 옵션을 여러 번 사용하면 열 이름을 고유하게 유지하기 위해 열 이름에 번호가 매겨집니다 (name1, name2 등).

표 3.1 저장 옵션 설명

옵션	데이터 테이블에 추가되는 열	설명
수준 수	수준 < 열 이름 >	<p>각 관측값의 수준 수는 해당 관측값이 포함된 히스토그램 막대에 해당합니다. 히스토그램 막대는 낮은 것부터 높은 것 순서로 1로 시작하는 번호가 매겨집니다.</p> <p>참고 : 소스 정보를 유지하기 위해 새 열에 값 라벨을 추가할 수 있지만 값 라벨은 기본적으로 해제되어 있습니다.</p>
수준 중간점	중간점 < 열 이름 >	<p>각 관측값의 중간점 값은 수준 하한에 수준 너비의 1/2을 더해서 계산됩니다.</p> <p>참고 : 소스 정보를 유지하기 위해 새 열에 값 라벨을 추가할 수 있지만 값 라벨은 기본적으로 해제되어 있습니다.</p>
순위	순위화 < 열 이름 >	<p>해당 열의 각 값에 대한 순위 (1 부터 시작) 를 제공합니다. 중복 반응 값의 경우 데이터 테이블에서의 발생 순서에 따라 연속 순위가 할당됩니다.</p>
순위 평균	평균 순위 < 열 이름 >	<p>값이 고유한 경우에는 평균 순위와 순위가 동일합니다. 값이 k 번 나타나는 경우 평균 순위는 값의 순위 합계를 k 로 나워서 계산됩니다.</p>
확률 점수	확률 < 열 이름 >	<p>비결측 점수가 N 개인 경우 값의 확률 점수는 해당 값의 평균 순위를 $N+1$ 로 나워서 계산됩니다. 이 열은 경험적 누적 분포 함수와 유사합니다.</p>
정규 분위수	N- 분위수 < 열 이름 >	<p>정규 분위수를 저장합니다. 자세한 내용은 "정규 분위수 그림"(76 페이지)에서 확인하십시오.</p>
표준화	표준화 < 열 이름 >	<p>표준화된 값을 저장합니다. 자세한 내용은 "표준화된 데이터 저장"(79 페이지)에서 확인하십시오.</p>
중심화	중심화 < 열 이름 >	<p>0 을 기준으로 중심화한 값을 저장합니다.</p>
로버스트 표준화	로버스트 표준화 < 열 이름 >	<p>로버스트 평균을 중심으로 중심화되고 로버스트 표준편차를 사용하여 표준화된 반응 값을 포함하는 열을 저장합니다.</p>

표 3.1 저장 옵션 설명 (계속)

옵션	데이터 테이블에 추가되는 열	설명
로버스트 중심화	로버스트 중심화 < 열 이름 >	로버스트 평균을 중심으로 중심화된 반응 값을 포함하는 열을 저장합니다.
로그에 스크립트로	(없음)	스크립트를 로그 창에 출력합니다 . 분석을 다시 생성하려면 스크립트를 실행합니다.

분포 플랫폼의 추가 예

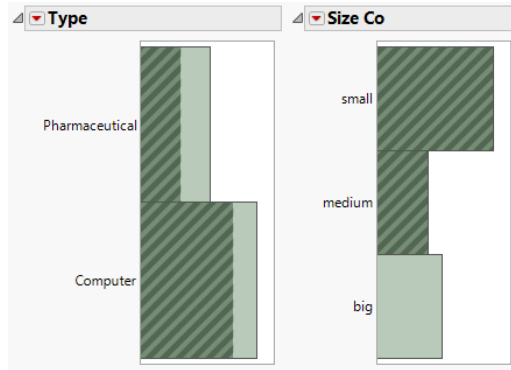
- " 예 : 여러 히스토그램에서 데이터 선택 "
- " 예 : 기준 변수 사용 "
- " 검정 확률 옵션의 예 "
- " 예측 구간의 예 "
- " 공차 구간의 예 "
- " 공정 능력의 예 "

예 : 여러 히스토그램에서 데이터 선택

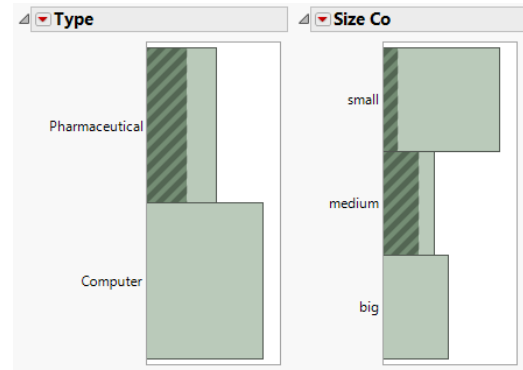
1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Companies.jmp 를 엽니다 .
2. 분석 > 분포를 선택합니다 .
3. Type 및 Size Co 를 선택하고 Y, 열을 클릭합니다 .
4. 확인을 클릭합니다 .
소규모 회사의 유형 분포를 확인하려고 합니다 .
5. "small" 옆의 막대를 클릭합니다 .
제약 회사보다 컴퓨터 회사에 소규모 회사가 더 많음을 확인할 수 있습니다 . 선택 범위를 넓히기 위해 중간 규모 회사를 추가합니다 .
6. Shift 키를 누른 채로 있습니다 . "Size Co" 히스토그램에서 "medium" 옆의 막대를 클릭합니다 .
소규모 및 중간 규모 회사의 유형 분포를 확인할 수 있습니다 . 그림 3.13 의 왼쪽을 참조하십시오 . 선택 범위를 좁히기 위해 소규모 및 중간 규모의 제약 회사만 확인하려고 합니다 .
7. Ctrl 및 Shift 키 (Windows 의 경우) 또는 Command 및 Shift 키 (macOS 의 경우) 를 누른 채로 있습니다 . "Type" 히스토그램에서 "Computer" 막대를 클릭하여 선택 취소합니다 .

소규모 및 중간 규모 회사 중 제약 회사의 수를 확인할 수 있습니다. 그림 3.13의 오른쪽을 참조하십시오.

그림 3.13 여러 히스토그램에서 데이터 선택



Shift 키를 사용하여 선택 범위를 넓힙니다.



Ctrl+Shift(Windows) 또는 Command+Shift(macOS)를 사용하여 선택 범위를 좁힙니다.

예 : 기준 변수 사용

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Lipid Data.jmp 를 엽니다.
 2. **분석 > 분포**를 선택합니다.
 3. Cholesterol 을 선택하고 **Y, 열**을 클릭합니다.
 4. Gender 를 선택하고 **기준**을 클릭합니다.
- 이렇게 하면 Gender 의 각 수준 (female 및 male) 에 대해 개별 분석이 수행됩니다.
5. **확인**을 클릭합니다.

히스토그램과 보고서의 방향을 변경합니다.

6. "분포" 의 빨간색 삼각형을 클릭하고 **쌓기**를 선택합니다.

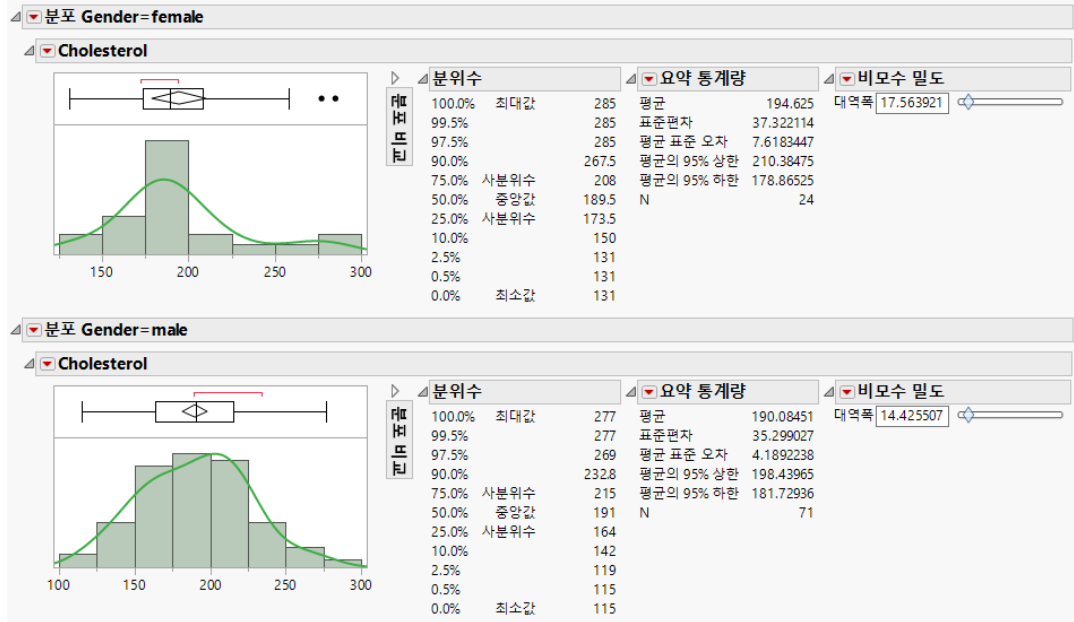
두 히스토그램 모두에 평활 곡선을 추가합니다.

7. Ctrl 키를 누른 채로 있습니다. "Cholesterol" 의 빨간색 삼각형을 클릭하고 **연속형 적합 > 평활 곡선 적합**을 선택합니다.

"분포 비교" 보고서를 숨깁니다.

8. Ctrl 키를 누른 채로 있습니다. **분포 비교** 옆의 회색 표시 아이콘을 클릭합니다.

그림 3.14 성별별 개별 분포



검정 확률 옵션의 예

수준이 세 개 이상인 변수에 대한 검정 확률 보고서를 시작합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 VA Lung Cancer.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. Cell Type 을 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "Cell Type" 의 빨간색 삼각형을 클릭하고 **검정 확률**을 선택합니다.
그림 3.15 의 왼쪽을 참조하십시오.

수준이 정확히 두 개인 변수에 대한 검정 확률 보고서를 시작합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penicillin.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. Response 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "Response" 의 빨간색 삼각형을 클릭하고 **검정 확률**을 선택합니다.
그림 3.15 의 오른쪽을 참조하십시오.

그림 3.15 검정 확률 옵션의 예

분포

Cell Type

빈도

검정 확률

수준	추정 확률	가설 확률
Adeno	0.19708	
Large	0.19708	
Small	0.35036	
Squamous	0.25547	

클릭한 다음 가설 확률을 입력하십시오.

확률의 합이 1이 되도록 하는 재적도화 방법을 선택하십시오.

☐ 가설 값을 재조정하여 생략된 값을 추정값으로 수정
☒ 생략된 값을 재조정하여 가설 값을 수정

완료 **도움말**

분포

Response

빈도

검정 확률

수준	추정 확률	가설 확률
Cured	0.53704	
Died	0.46296	

클릭한 다음 가설 확률을 입력하십시오.

검정 확률에 대한 대립가설을 선택하십시오.

☒ 확률이 가설 값과 같지 않을(양측 카이제곱 검정)
☐ 확률이 가설 값보다 큼(정확 단측 이항 검정)
☐ 확률이 가설 값보다 작음(정확 단측 이항 검정)

완료 **도움말**

수준이 세 개 이상인 변수에 대한 보고서 옵션

수준이 정확히 두 개인 변수에 대한 보고서 옵션

예 : 검정 확률 보고서 생성

수준이 세 개 이상인 변수에 대한 검정 확률 보고서를 생성하려면 다음을 수행하십시오.

1. 그림 3.15의 왼쪽을 참조하십시오. 네 개의 "가설 확률" 필드 모두에 0.25를 입력합니다.
2. **생략된 값을 재조정하여 가설 값을 수정** 버튼을 클릭합니다.
3. **완료**를 클릭합니다.

우도비 및 Pearson 카이제곱 검정이 계산됩니다. 그림 3.16의 왼쪽을 참조하십시오.

수준이 정확히 두 개인 변수에 대한 검정 확률 보고서를 생성하려면 다음을 수행하십시오.

1. 그림 3.15의 오른쪽을 참조하십시오. 두 "가설 확률" 필드 모두에 0.5를 입력합니다.
2. **확률이 가설 값보다 작음** 버튼을 클릭합니다.
3. **완료**를 클릭합니다.

이항 검정의 정확한 확률이 계산됩니다. 그림 3.16의 오른쪽을 참조하십시오.

그림 3.16 검정 확률 보고서의 예



예측 구간의 예

다음 10 개의 오존 수준 관측값에 대한 예측 구간을 계산하는 데 관심이 있다고 가정해 보겠습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Cities.jmp** 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. **OZONE** 을 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "OZONE" 의 빨간색 삼각형을 클릭하고 **예측 구간**을 선택합니다.

그림 3.17 예측 구간 창

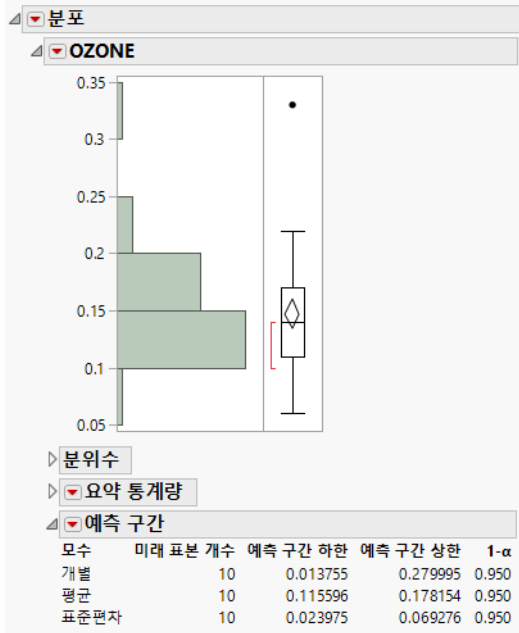
예측 구간의 (1- α) 입력

미래 표본 수 입력

☒ 양측
☐ 단측 하한
☐ 단측 상한

6. " 예측 구간 " 창에서 미래 표본 수 입력 옆에 10 을 입력합니다 .
7. 확인 을 클릭합니다 .

그림 3.18 예측 구간 보고서의 예



이 예에서는 다음 사항을 95% 신뢰할 수 있습니다 .

- 다음 10 개의 관측값은 각각 0.013755 에서 0.279995 사이에 있습니다 .
- 다음 10 개 관측값의 평균은 0.115596 에서 0.178154 사이에 있습니다 .
- 다음 10 개 관측값의 표준편차는 0.023975 에서 0.069276 사이에 있습니다 .

공차 구간의 예

오존 수준 측정값의 90% 를 포함하는 구간을 추정하려고 한다고 가정해 보겠습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Cities.jmp** 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. **OZONE** 을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "OZONE" 의 빨간색 삼각형을 클릭하고 **공차 구간**을 선택합니다 .

그림 3.19 공차 구간 창

(1- α) 신뢰도로 모집단의 지정된 비율 이상을 포함하는 구간을 계산합니다.

신뢰도 (1- α) 지정:

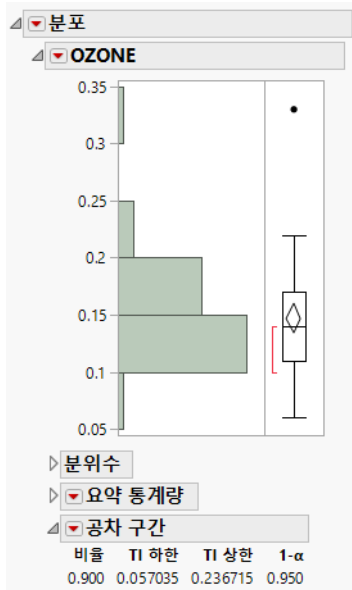
포함할 비율 지정:

☒ 양측
☐ 단측 하한
☐ 단측 상한

방법
☒ 정규 분포 가정
☐ 비모수

6. 기본 선택 사항을 유지하고 **확인**을 클릭합니다 .

그림 3.20 공차 구간 보고서의 예



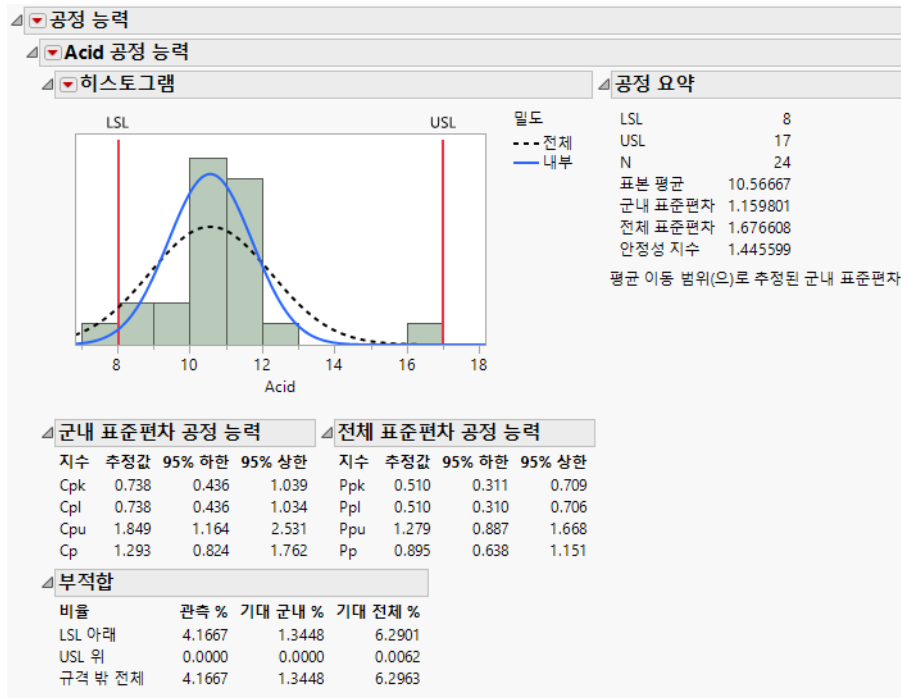
이 예에서는 TI(공차 구간) 하한 및 TI 상한 값에 따라 모집단의 90% 이상이 0.057035 에서 0.236715 사이에 있음을 95% 신뢰할 수 있습니다.

공정 능력의 예

피클의 산도를 특성화하려고 한다고 가정해 보겠습니다. 규격 하한과 규격 상한은 각각 8 과 17 입니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Quality Control/Pickles.jmp** 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. **Acid** 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "Acid" 의 빨간색 삼각형을 클릭하고 **공정 능력**을 선택합니다.
6. **LSL**(규격 하한) 로 8 을 입력합니다.
7. **USL**(규격 상한) 로 17 을 입력합니다.
8. **확인**을 클릭합니다.

그림 3.21 공정 능력 보고서의 예



보고서에는 공정 능력 결과가 추가되었습니다. 공정 능력 보고서의 히스토그램에는 규격 한계가 표시되므로 시각적으로 데이터를 이 한계와 비교할 수 있습니다. 그림에서 알 수 있듯이 일부 산도 수준은 규격 하한보다 낮고 일부는 규격 상한에 매우 가깝습니다. Ppk 값은 0.510 이며, 이는 지정된 규격 한계를 기준으로 할 때 공정 능력이 없음을 나타냅니다.

분포 플랫폼에 대한 통계 상세 정보

- "표준 오차 막대"
- "분위수"
- "요약 통계량"
- "정규 분위수 그림"
- "Wilcoxon 부호 순위 검정"
- "표준편차 검정"
- "정규 분위수"
- "표준화된 데이터 저장"

- "예측 구간"
- "공차 구간"
- "연속형 적합 분포"
- "이산형 적합 분포"

표준 오차 막대

표준 오차 막대는 표준 오차 ($\sqrt{np_i(1-p_i)}$) 를 사용하여 계산됩니다 ($p_i = n_i / n$).

분위수

이 섹션에서는 분위수 계산 방법에 대해 설명합니다.

한 열에 있는 n 개의 비결측값 중 p 번째 분위수를 계산하려면 n 개의 값을 오름차순으로 배열하며 이 열 값을 y_1, y_2, \dots, y_n 이라고 합니다. p 번째 분위수의 순위 번호는 $p / 100(n+1)$ 로 계산합니다.

- 결과가 정수인 경우 p 번째 분위수는 이 순위에 해당하는 값이 됩니다.
- 결과가 정수가 아닌 경우에는 보간법을 사용하여 p 번째 분위수를 구합니다. p 번째 분위수, 즉 q_p 는 다음과 같이 계산됩니다.

$$q_p = (1-f)y_i + fy_{i+1}$$

다음은 각 요소에 대한 설명입니다.

- n 은 변수의 비결측값 수입니다.
- y_1, y_2, \dots, y_n 은 정렬된 변수 값을 나타냅니다.
- y_{n+1} 은 y_n 으로 간주됩니다.
- i 는 $(n+1)p$ 의 정수부이고 f 는 소수부입니다.
- $(n+1)p = i + f$

예를 들어 15 개의 행이 있는 데이터 테이블에서 연속형 열의 75 번째 및 90 번째 분위수 값을 구하려고 한다고 가정해 보겠습니다. 열을 오름차순으로 배열한 후 이러한 분위수가 포함된 순위를 다음과 같이 계산합니다.

$$\frac{75}{100}(15+1) = 12 \quad \text{및} \quad \frac{90}{100}(15+1) = 14.4$$

값 y_{12} 가 75 번째 분위수입니다. 90 번째 분위수는 14 번째 순위 값과 15 번째 순위 값의 가중 평균을 계산하여 보간됩니다 ($y_{90} = 0.6y_{14} + 0.4y_{15}$).

요약 통계량

이 섹션에는 "요약 통계량" 보고서의 특정 통계량에 대한 통계 상세 정보가 포함되어 있습니다.

평균

평균은 비결측값의 합을 비결측값의 개수로 나눈 값입니다. **가중치** 또는 **빈도** 변수를 할당한 경우 JMP 에서 평균은 다음과 같이 계산됩니다.

1. 각 열 값에 해당 가중치 또는 빈도를 곱합니다.
2. 결과 값을 모두 더한 후 가중치 또는 빈도의 합으로 나눕니다.

표준편차

표준편차는 평균을 중심으로 한 분포의 퍼짐 정도를 측정한 것입니다. 종종 s 로 표시하며, 표본 분산 (s^2 으로 표시)의 제곱근과 같습니다.

$$s = \sqrt{s^2}$$

다음은 각 요소에 대한 설명입니다.

$$s^2 = \sum_{i=1}^N \frac{w_i (y_i - \bar{y}_w)^2}{N-1}$$

= 가중 평균

$$\bar{y}_w$$

평균 표준 오차

평균 표준 오차는 표본 표준편차 s 를 N 의 제곱근으로 나눠서 계산됩니다. 시작 창에서 가중치 또는 빈도 열을 지정한 경우에는 가중치 또는 빈도 합을 제곱근이 분모가 됩니다.

왜도

왜도는 평균에 대한 3 차 적률을 기준으로 하며 다음과 같이 계산됩니다.

$$\sum w_i z_i^3 \frac{N}{(N-1)(N-2)} \quad \text{여기서,} \quad z_i = \frac{x_i - \bar{x}}{s}$$

또한 w_i 는 가중치 항 (균일 가중 항목의 경우 1) 입니다.

첨도

첨도는 평균에 대한 3 차 적률을 기준으로 하며 다음과 같이 계산됩니다 .

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n w_i^2 \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

여기서 w_i 는 가중치 항 (균일 가중 항목의 경우 1) 입니다 . 이 계산식을 사용할 경우 정규 분포의 첨도는 0 입니다 . 이 계산식을 종종 초과 첨도라고 합니다 .

정규 분위수 그림

각 값의 경험적 누적 확률은 다음과 같이 계산됩니다 .

$$\frac{r_i}{N+1}$$

여기서 r_i 는 i 번째 관측값의 순위이고 , N 은 비결측이면서 제외되지 않은 관측값의 개수입니다 . 정규 분위수 값은 다음과 같이 계산됩니다 .

$$\Phi^{-1}\left(\frac{r_i}{N+1}\right)$$

여기서 Φ 는 정규 분포의 누적 확률 분포 함수입니다 .

이 정규 분위수 값은 정규 분포에 기대되는 순서 통계량에 대한 Van Der Waerden 근사값입니다 .

Wilcoxon 부호 순위 검정

Wilcoxon 부호 순위 검정을 사용하여 단일 모집단의 중앙값을 검정하거나 일반적인 중앙값의 매칭 쌍 데이터를 검정할 수 있습니다 . 매칭 쌍의 경우에는 검정 범위를 줄여 중앙값 0 에 대한 쌍체 차이로 구성된 단일 모집단을 검정합니다 . 검정 시 기본 모집단은 대칭적인 것으로 가정됩니다 .

Wilcoxon 검정에서는 동일 값이 허용됩니다 . 차이가 0 인 경우에는 Pratt 이 제안한 방법을 사용하여 검정 통계량이 조정됩니다 . 자세한 내용은 Lehmann and D'Abrera 연구 자료 (2006), Pratt 연구 자료 (1959) 및 Cureton 연구 자료 (1967) 에서 확인하십시오 .

단일 모집단의 중앙값 검정

- 다음과 같이 N 개의 관측값이 있습니다 .

$$X_1, X_2, \dots, X_N$$

- 귀무가설은 다음과 같습니다 .

$$H_0: X \text{ 의 분포가 } m \text{ 을 기준으로 대칭적입니다 .}$$

- 관측값과 가설 값 m 사이의 차이는 다음과 같이 계산됩니다.

$$D_j = X_j - m$$

매칭 쌍 데이터를 사용하여 두 모집단 중앙값의 동일 여부 검정

매칭 쌍 데이터에는 특수한 형태의 Wilcoxon 부호 순위 검정이 적용됩니다.

- 다음과 같이 두 개의 모집단에서 얻은 N 쌍의 관측값이 있습니다.

$$X_1, X_2, \dots, X_N \text{ 및 } Y_1, Y_2, \dots, Y_N$$

- 귀무가설은 다음과 같습니다.

$$H_0: X - Y \text{의 분포가 } 0 \text{을 기준으로 대칭적입니다.}$$

- 관측값 쌍 사이의 차이는 다음과 같이 계산됩니다.

$$D_j = X_j - Y_j$$

Wilcoxon 부호 순위 검정 통계량

이 검정 통계량은 부호 순위의 합을 기준으로 합니다. 부호 순위는 다음과 같이 정의됩니다.

- 차이 절대값 $|D_j|$ 는 작은 값부터 큰 값의 순서로 순위화됩니다.
- 차이가 0 인 경우가 있더라도 순위는 값 1 부터 시작합니다.
- 동일한 절대 차이가 있는 경우에는 해당 관측값 순위의 평균 또는 중간 순위가 할당됩니다.

차이 D_j 의 순위 또는 중간 순위는 R_j 로 나타냅니다. D_j 의 부호 순위는 다음과 같이 정의합니다.

- 차이 D_j 가 양수이면 부호 순위는 R_j 입니다.
- 차이 D_j 가 0 이면 부호 순위는 0 입니다.
- 차이 D_j 가 음수이면 부호 순위는 $-R_j$ 입니다.

부호 순위 통계량은 다음과 같이 계산됩니다.

$$S = \frac{1}{2} \sum_{j=1}^N \text{부호 순위}$$

다음을 정의합니다.

d_0 - 0 과 동일한 부호 순위의 개수입니다.

R^+ - 양수 부호 순위의 합입니다.

그런 후 다음을 구합니다.

$$S = R^+ - \frac{1}{4}[N(N+1) - d_0(d_0+1)]$$

Wilcoxon 부호 순위 검정 p 값

$N \leq 20$ 인 경우 정확한 p 값이 계산됩니다.

$N > 20$ 인 경우 아래에 정의된 것과 같이 이 통계량에 대한 스튜던트 t 근사가 사용됩니다. 이때 동일 값에 대해서는 수정이 적용됩니다. 자세한 내용은 Iman 연구 자료 (1974) 및 Lehmann and D'Abrera 연구 자료 (2006) 에서 확인하십시오.

귀무가설하에서 S 의 평균은 0 입니다. S 의 분산은 다음과 같이 구합니다.

$$Var(S) = \frac{1}{24} \left[N(N+1)(2N+1) - d_0(d_0+1)(2d_0+1) - \frac{1}{2} \sum_{i>0} d_i(d_i+1)(d_i-1) \right]$$

$Var(S)$ 표현식의 마지막 합은 동일 값을 수정한 것입니다. $i > 0$ 인 경우에 대한 표기인 d_i 는 0 이 아닌 부호 순위의 i 번째 그룹에 있는 값의 개수를 나타냅니다. 특정 부호 순위에 대해 동일 값이 없으면 $d_i = 1$ 이고 합은 0 입니다.

다음 계산식으로 구한 통계량 t 는 자유도가 $N - 1$ 인 근사 t 분포를 갖습니다.

$$t = \frac{S}{\sqrt{\frac{N \cdot Var(S) - S^2}{N - 1}}}$$

표준편차 검정

다음은 이 검정 통계량을 계산하기 위한 계산식입니다.

$$\frac{(n-1)s^2}{\sigma^2}$$

이 검정 통계량은 모집단이 정규 데이터일 때 자유도가 $n - 1$ 인 카이제곱 변수로 분포됩니다.

최소 p 값은 양쪽 꼬리 검정의 p 값으로, 다음과 같이 계산됩니다.

$$2 * \min(p1, p2)$$

여기서 $p1$ 은 아래쪽 꼬리의 p 값이고 $p2$ 는 위쪽 꼬리의 p 값입니다.

정규 분위수

정규 분위수 값은 다음과 같이 계산됩니다.

$$\Phi^{-1}\left(\frac{r_i}{N+1}\right)$$

다음은 각 요소에 대한 설명입니다.

Φ - 정규 분포의 누적 확률 분포 함수입니다.

r_i - i 번째 관측값의 순위입니다.

N - 비결측 관측값의 개수입니다.

표준화된 데이터 저장

표준화된 값은 다음 계산식을 사용하여 계산됩니다.

$$\frac{X - \bar{X}}{S_X}$$

다음은 각 요소에 대한 설명입니다.

X - 원래 열입니다.

\bar{X} - X 열의 평균입니다.

S_X - X 열의 표준편차입니다.

예측 구간

JMP에서 예측 구간을 계산하는 데 사용되는 계산식은 다음과 같습니다.

- 미래 관측값이 m 개인 경우 :

$$[\underline{y}_m, \tilde{y}_m] = \bar{X} \pm t_{(1-\alpha/2; n-1)} \times \sqrt{1 + \frac{1}{n}} \times s \quad (\text{단, } m \geq 1)$$

- 미래 관측값 m 개의 평균을 사용하는 경우 :

$$[Y_l, Y_u] = \bar{X} \pm t_{(1-\alpha/2; n-1)} \times \sqrt{\frac{1}{m} + \frac{1}{n}} \times s \quad (\text{단, } m \geq 1)$$

- 미래 관측값 m 개의 표준편차를 사용하는 경우 :

$$[s_l, s_u] = \left[s \times \sqrt{\frac{1}{F_{(1-\alpha/2; (n-1), m-1)}}}, s \times \sqrt{F_{(1-\alpha/2; (m-1), n-1)}} \right] \quad (\text{단, } m \geq 2)$$

여기서 m 은 미래 관측값의 개수이고, n 은 현재 분석 표본에 포함된 점의 개수입니다.

- 단측 구간은 분위수 함수에서 $1-\alpha$ 를 사용하여 구합니다.

자세한 내용은 Meeker et al. 연구 자료 (2017, ch. 4)에서 확인하십시오.

공차 구간

이 섹션에는 단측 및 양측 공차 구간에 대한 통계 상세 정보가 포함되어 있습니다.

정규 분포 기반의 구간

단측 구간

단측 구간은 다음과 같이 계산됩니다.

$$\text{하한} = \bar{x} - g's$$

$$\text{상한} = \bar{x} + g's$$

다음은 각 요소에 대한 설명입니다.

$$g' = t(1 - \alpha, n - 1, \Phi^{-1}(p) \cdot \sqrt{n}) / \sqrt{n}$$

s - 표준편차입니다.

t - 비중심 t 분포의 분위수입니다.

Φ^{-1} - 표준 정규 분위수입니다.

양측 구간

양측 구간은 다음과 같이 계산됩니다.

$$[T_{p_L}, T_{p_U}] = [\bar{x} - g(1 - \alpha/2; p, n)s, \bar{x} + g(1 - \alpha/2; p, n)s]$$

여기서 s 는 표준편차이고, $g(1 - \alpha/2; p, n)$ 은 상수입니다.

g 를 확인하려면 공차 구간으로 구한 모집단의 비율을 고려합니다. Tamhane and Dunlop 연구 자료 (2000) 에서는 이 비율을 다음과 같이 구합니다.

$$\Phi\left(\frac{\bar{x} + gs - \mu}{\sigma}\right) - \Phi\left(\frac{\bar{x} - gs - \mu}{\sigma}\right)$$

여기서 Φ 는 표준 정규 CDF(누적 분포 함수) 입니다.

따라서 g 는 다음 방정식을 해결합니다.

$$P\left\{\Phi\left(\frac{\bar{X} + gs - \mu}{\sigma}\right) - \Phi\left(\frac{\bar{X} - gs - \mu}{\sigma}\right) \geq 1 - \gamma\right\} = 1 - \alpha$$

여기서 $1 - \gamma$ 는 공차 구간에 포함된 모든 미래 관측값의 비율입니다.

정규 분포 기반의 공차 구간에 대한 자세한 내용은 Meeker et al. 연구 자료 (2017) 의 표 J.1a, J.1b, J.6a 및 J.6b 에서 확인하십시오 .

비모수적 구간

단측 하한

크기가 n 인 표본에서 비율 β 이상의 표본 분포를 포함하기 위한 $100(1 - \alpha)\%$ 단측 공차 하한은 순서 통계량 $x_{(l)}$ 입니다 . 지수 l 은 다음과 같이 계산됩니다 .

$$l = n - \Phi_{bin}^{-1}(1 - \alpha, n, \beta)$$

여기서 $\Phi_{bin}^{-1}(1 - \alpha, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포의 $(1 - \alpha)$ 번째 분위 수입니다 .

실제 신뢰 수준은 $\Phi_{bin}(n - l, n, \beta)$ 로 계산되며 , 여기서 $\Phi_{bin}(x, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포 확률 변수가 x 보다 작거나 같을 확률입니다 .

단측 분포 무관 공차 하한을 계산하려면 표본 크기 n 이 $(\log \alpha) / (\log \beta)$ 이상이어야 합니다 .

단측 상한

크기가 n 인 표본에서 비율 β 이상의 표본 분포를 포함하기 위한 $100(1 - \alpha)\%$ 단측 공차 상한은 순서 통계량 $x_{(u)}$ 입니다 . 지수 u 는 다음과 같이 계산됩니다 .

$$u = 1 + \Phi_{bin}^{-1}(1 - \alpha, n, \beta)$$

여기서 $\Phi_{bin}^{-1}(1 - \alpha, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포의 $(1 - \alpha)$ 번째 분위 수입니다 .

실제 신뢰 수준은 $\Phi_{bin}(u - 1, n, \beta)$ 로 계산되며 , 여기서 $\Phi_{bin}(x, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포 확률 변수가 x 보다 작거나 같을 확률입니다 .

단측 분포 무관 공차 상한을 계산하려면 표본 크기 n 이 $(\log \alpha) / (\log \beta)$ 이상이어야 합니다 .

양측 구간

크기가 n 인 표본에서 비율 β 이상의 표본 분포를 포함하기 위한 $100(1 - \alpha)\%$ 양측 공차 구간은 다음과 같이 계산됩니다 .

$$[\tilde{T}_{p_L}, \tilde{T}_{p_U}] = [x_{(l)}, x_{(u)}]$$

여기서 $x_{(i)}$ 는 i 번째 순서 통계량이며 , l 및 u 는 다음과 같이 계산됩니다 .

$v = n - \Phi_{bin}^{-1}(1 - \alpha, n, \beta)$ 라고 가정합니다. 여기서 $\Phi_{bin}^{-1}(1 - \alpha, n, \beta)$ 는 시행 횟수가 n 이고 성공 확

률이 β 인 이항 분포의 $(1 - \alpha)$ 번째 분위수입니다. v 가 2 보다 작으면 양측 분포 무관 공차 구간을 계산할 수 없습니다. v 가 2 보다 크거나 같으면 $l = \text{floor}(v/2)$ 이고 $u = \text{floor}(n + 1 - v/2)$ 입니다.

실제 신뢰 수준은 $\Phi_{\text{bin}}(u-l-1, n, \beta)$ 로 계산되며, 여기서 $\Phi_{\text{bin}}(x, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포 확률 변수가 x 보다 작거나 같을 확률입니다.

양측 분포 무관 공차 구간을 계산하려면 표본 크기 n 이 다음 방정식의 이상이어야 합니다.

$$1 - \alpha = 1 - n\beta^{n-1} + (n-1)\beta^n$$

분포 무관 공차 구간에 대한 자세한 내용은 Meeker et al. 연구 자료 (2017, sec. 5.3) 에서 확인하십시오.

연속형 적합 분포

이 섹션에는 "연속형 적합" 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다.

정규 적합

"정규 적합" 옵션은 정규 분포의 모수를 추정합니다. 정규 분포의 모수는 다음과 같습니다.

- μ (평균) - x 축에서의 분포 위치를 정의합니다.
- σ (표준편차) - 분포의 산포 또는 퍼짐 정도를 정의합니다.

표준 정규 분포는 $\mu = 0$ 이고 $\sigma = 1$ 일 때 나타납니다.

$$\text{pdf: } \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (\text{단, } -\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma)$$

$$E(x) = \mu$$

$$\text{Var}(x) = \sigma^2$$

Cauchy 적합

"Cauchy 적합" 옵션은 위치가 μ 이고 척도가 σ 인 Cauchy 분포를 적합시킵니다.

$$\text{pdf: } \left\{ \pi\sigma \left[1 + \left(\frac{x-\mu}{\sigma} \right)^2 \right] \right\}^{-1} \quad (\text{단, } -\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma)$$

$$E(x) = \text{정의되지 않음}$$

$$\text{Var}(x) = \text{정의되지 않음}$$

SHASH 적합

"SHASH 적합" 옵션은 SHASH(SinH-ArcSinH) 분포를 적합시킵니다. SHASH 분포는 정규 분포의 변환을 기반으로 하며 정규 분포를 특수한 경우로 포함합니다. 이 분포는 대칭적일 수도 있고 비대칭적일 수도 있습니다. 형태는 두 개의 형태 모수 γ 및 δ 에 의해 결정됩니다. SHASH 분포에 대한 자세한 내용은 Jones and Pewsey 연구 자료 (2009)에서 확인하십시오.

$$\text{pdf: } f(x) = \frac{\delta \cosh(w)}{\sqrt{\sigma^2 + (x - \theta)^2}} \phi[\sinh(w)] \quad \left(\text{단, } -\infty < \gamma, x, \theta < \infty, 0 < \delta, \sigma \right)$$

다음은 각 요소에 대한 설명입니다.

- 표준 정규 pdf 입니다.

$$\phi(\cdot)$$

$$w = \gamma + \delta \sinh^{-1} \left(\frac{x - \theta}{\sigma} \right)$$

- $\gamma=0$ 이고 $\delta=1$ 인 경우 SHASH 분포는 위치가 θ 이고 척도가 σ 인 정규 분포와 동등합니다.
- 변환 $\sinh(w)$ 는 $\mu=0$ 이고 $\sigma=1$ 인 정규 분포를 따릅니다.

지수 적합

지수 분포는 생존 데이터와 같이 시간 경과에 따라 무작위로 발생하는 사건을 설명하는 데 특히 유용합니다. 지수 분포는 중복되지 않는 사건의 발생 사이 경과 시간을 모델링하는 데도 유용할 수 있습니다. 중복되지 않는 사건의 예로는 사용자의 컴퓨터 쿼리 후 서버에서 응답할 때까지의 시간, 서비스 데스크에서의 고객 내방, 전화 교환대에서의 전화 수신 등이 포함됩니다.

지수 분포는 $\beta=1$ 이고 $\alpha=\sigma$ 일 경우 특수한 형태의 2 모수 Weibull 분포가 되며, $\alpha=1$ 일 경우에는 특수한 형태의 감마 분포가 되기도 합니다.

$$\text{pdf: } \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right) \quad \left(\text{단, } 0 < \sigma, 0 \leq x \right)$$

$$E(x) = \sigma$$

$$\text{Var}(x) = \sigma^2$$

Devore 연구 자료 (1995)에 따르면 지수 분포는 무기억성을 가집니다. 무기억성은 t 시간 후 여전히 유효한 어떤 성분을 확인할 경우 추가 수명의 분포 (t 시간 후까지도 해당 성분이 유지된 경우 추가 수명의 조건부 확률)가 원래 분포와 동일함을 의미합니다.

감마 적합

"감마 적합" 옵션은 감마 분포 모수 $\alpha > 0$ 및 $\sigma > 0$ 을 추정합니다. 적합 감마 보고서에서 모수 α (알파) 는 형태 또는 곡률을 설명합니다. 모수 σ (시그마) 는 분포의 척도 모수입니다. 데이터는 0 보다 커야 합니다.

$$\text{pdf: } \frac{1}{\Gamma(\alpha)\sigma^\alpha} x^{\alpha-1} \exp(-x/\sigma) \quad (\text{단, } 0 < x, 0 < \alpha, \sigma)$$

$$E(x) = \alpha\sigma$$

$$\text{Var}(x) = \alpha\sigma^2$$

- 표준 감마 분포의 경우 $\sigma = 1$ 입니다. 시그마는 값이 1 이 아닐 경우 가로 축을 따라 분포를 늘리거나 줄이므로 척도 모수라고 합니다.
- 카이제곱 $\chi^2_{(v)}$ 분포는 $\sigma = 2$ 이고 $\alpha = v/2$ 일 때 발생합니다.
- 지수 분포는 $\alpha = 1$ 일 때 발생합니다.

표준 감마 밀도 함수는 $\alpha \leq 1$ 일 때는 감소하기만 합니다. $\alpha > 1$ 일 때 밀도 함수는 0 에서 시작하여 최대값까지 증가하다가 감소합니다.

로그 정규 적합

"로그 정규 적합" 옵션은 2 모수 로그 정규 분포의 μ (척도) 및 σ (형태) 모수를 추정합니다. $X = \ln(Y)$ 가 정규 분포인 경우에만 Y 변수가 로그 정규 분포로 나타납니다. 데이터는 0 보다 커야 합니다.

$$\text{pdf: } \frac{1}{\sigma\sqrt{2\pi}} \frac{\exp\left[-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right]}{x} \quad (\text{단, } 0 \leq x, -\infty < \mu < \infty, 0 < \sigma)$$

$$E(x) = \exp(\mu + \sigma^2/2)$$

$$\text{Var}(x) = \exp(2(\mu + \sigma^2)) - \exp(2\mu + \sigma^2)$$

Weibull 적합

Weibull 분포는 α (척도) 및 β (형태) 값에 따라 다른 형태로 나타납니다. 이 함수는 특히 기계 장치 및 생물학 분야에서 수명 길이를 추정하는 데 적절한 모형을 제공하는 경우가 많습니다.

Weibull 분포의 pdf 는 다음과 같습니다.

$$\text{pdf: } \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right] \quad (\text{단, } \alpha, \beta > 0, 0 < x)$$

$$E(x) = \alpha \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$\text{Var}(x) = \alpha^2 \left\{ \Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right\}$$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다.

정규 2 혼합 적합 및 정규 3 혼합 적합

"정규 2 혼합 적합" 및 "정규 3 혼합 적합" 옵션은 두 개 또는 세 개의 정규 분포를 적합시킵니다. 이러한 유연한 분포는 이봉 또는 다봉 데이터를 적합시킬 수 있습니다. 각 그룹에 대해 평균, 표준편차 및 전체 대비 비율이 개별적으로 추정됩니다. 다음 방정식에서 k 는 혼합되는 정규 분포의 수와 같습니다.

$$\text{pdf: } \sum_{i=1}^k \frac{\pi_i}{\sigma_i} \phi\left(\frac{x - \mu_i}{\sigma_i}\right)$$

$$E(x) = \sum_{i=1}^k \pi_i \mu_i$$

$$\text{Var}(x) = \sum_{i=1}^k \pi_i (\mu_i^2 + \sigma_i^2) - \left(\sum_{i=1}^k \pi_i \mu_i \right)^2$$

여기서 μ_i , σ_i 및 π_i 는 각각 i 번째 그룹의 평균, 표준편차 및 비율이며, $\phi(\cdot)$ 는 표준 정규 pdf 입니다.

Johnson 적합

"Johnson 적합" 옵션은 Johnson 분포 체계에서 최적의 적합 분포를 선택하여 적합시키며, Johnson 분포 체계에 포함된 세 개의 분포는 모두 하나의 변환된 정규 분포를 기반으로 합니다. 이 세 가지 분포는 다음과 같습니다.

- Johnson Su - 경계가 없습니다.
- Johnson Sb - 양쪽 꼬리에 경계가 있습니다. 이 경계는 추정 가능한 모수에 의해 정의됩니다.
- Johnson Sl - 한쪽 꼬리에 경계가 있습니다. 이 경계는 추정 가능한 모수에 의해 정의됩니다. Johnson Sl 계열에는 로그 정규 분포 계열이 포함됩니다.

선택된 분포에 대한 적합만 보고됩니다. Johnson 분포의 선택 프로시저 및 모수 추정에 대한 자세한 내용은 Slifker and Shapiro 연구 자료 (1980) 에서 확인할 수 있습니다. 모수 추정에는 최대 우도가 사용되지 않습니다.

Johnson 분포는 유연하기 때문에 널리 사용됩니다. 특히 Johnson 분포 체계는 왜도와 첨도의 가능한 모든 조합을 지원하므로 데이터 적합 기능 면에서 주목됩니다. 하지만 SHASH 분포 또한 매우 유연하므로 Johnson 분포보다 권장됩니다.

Z 가 표준 정규 변량인 경우 이 분포 체계는 다음과 같이 정의됩니다.

$$Z = \gamma + \delta f(Y)$$

다음은 Johnson Su 의 경우에 해당하는 각 요소에 대한 설명입니다.

$$f(Y) = \ln\left(Y + \sqrt{1 + Y^2}\right) = \sinh^{-1}Y$$

$$Y = \frac{X - \theta}{\sigma} \quad -\infty < X < \infty$$

다음은 Johnson Sb 의 경우에 해당하는 각 요소에 대한 설명입니다.

$$f(Y) = \ln\left(\frac{Y}{1 - Y}\right)$$

$$Y = \frac{X - \theta}{\sigma} \quad \theta < X < \theta + \sigma$$

다음은 Johnson SI 의 경우 (단, $\sigma = \pm 1$) 에 해당하는 각 요소에 대한 설명입니다.

$$f(Y) = \ln(Y)$$

$$Y = \frac{X - \theta}{\sigma} \quad \begin{array}{ll} \theta < X < \infty & \text{단, } \sigma = 1 \\ -\infty < X < \theta & \text{단, } \sigma = -1 \end{array}$$

Johnson Su

$$\text{pdf: } \frac{\delta}{\sigma} \left[1 + \left(\frac{x - \theta}{\sigma} \right)^2 \right]^{-1/2} \phi \left[\gamma + \delta \sinh^{-1} \left(\frac{x - \theta}{\sigma} \right) \right] \quad (\text{단, } -\infty < x, \theta, \gamma < \infty, 0 < \theta, \delta)$$

Johnson Sb

$$\text{pdf: } \phi \left[\gamma + \delta \ln \left(\frac{x - \theta}{\sigma - (x - \theta)} \right) \right] \left(\frac{\delta \sigma}{(x - \theta)(\sigma - (x - \theta))} \right) \quad (\text{단, } \theta < x < \theta + \sigma, 0 < \sigma)$$

Johnson SI

$$\text{pdf: } \frac{\delta}{|x - \theta|} \phi \left[\gamma + \delta \ln \left(\frac{x - \theta}{\sigma} \right) \right] \quad (\text{단, } \sigma = 1 \text{ 일 경우 } \theta < x, \sigma = -1 \text{ 일 경우 } \theta > x)$$

여기서 $\phi(\cdot)$ 는 표준 정규 pdf 입니다.

베타 적합

베타 분포는 간격 0,1 사이에 들어가도록 제한된 확률 변수의 동작을 모델링하는 데 유용합니다. 예를 들어 비율은 항상 0 에서 1 사이입니다. " 베타 적합 " 옵션은 두 개의 형태 모수, 즉 $\alpha > 0$ 및 $\beta > 0$ 을 추정합니다. 베타 분포에서는 값이 0,1 구간에만 있습니다.

$$\text{pdf: } \frac{1}{B(\alpha, \beta)\sigma^{\alpha+\beta-1}} x^{\alpha-1} x^{\beta-1} \quad (\text{단, } 0 < x < 1, 0 < \sigma, \alpha, \beta)$$

$$E(x) = \sigma \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(x) = \frac{\sigma^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

여기서 $B(\cdot)$ 는 베타 함수입니다.

전체 적합

" 분포 비교 " 보고서에서는 분포 목록이 AICc 를 기준으로 오름차순 정렬됩니다. 체크박스를 사용하여 선택한 분포에 대한 적합 보고서를 표시하거나 숨기고 분포 곡선을 중첩 표시할 수 있습니다.

AICc 와 BIC 의 계산식은 다음과 같습니다.

$$\text{AICc} = -2\log L + 2k + \frac{2k(k+1)}{n - (k+1)}$$

$$\text{BIC} = -2\log L + k \ln(n)$$

다음은 각 요소에 대한 설명입니다.

- $\log L$ 은 로그 우도입니다.
- n 은 표본 크기입니다.
- k 는 모수의 개수입니다.

"AICc 가중치" 열은 합계가 1 이 되는 정규화된 AICc 값을 보여 줍니다. AICc 가중치는 적합한 분포 중 하나가 true 일 때 특정 분포가 true 분포일 확률로 해석됩니다. 따라서 AICc 가중치가 1 과 가장 가까운 분포가 더 적절한 적합입니다. AICc 가중치는 다음과 같이 비결측 AICc 값만 사용하여 계산됩니다.

$$\text{AICcWeight} = \exp[-0.5(\text{AICc} - \min(\text{AICc}))] / \sum(\exp[-0.5(\text{AICc} - \min(\text{AICc}))])$$

여기서 $\min(\text{AICc})$ 는 적합한 분포 중에서 가장 작은 AICc 값입니다.

" 분포 비교 " 보고서의 측도에 대한 자세한 내용은 **Fitting Linear Models** 의 부록 "Statistical Details" 에서 확인하십시오 .

이산형 적합 분포

이 섹션에는 " 이산형 적합 " 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다 .

Poisson 적합

Poisson 분포는 단일 척도 모수 $\lambda > 0$ 을 갖습니다 .

$$\text{pmf: } \frac{e^{-\lambda} \lambda^x}{x!} \quad (\text{단, } 0 \leq \lambda < \infty, x = 0, 1, 2, \dots)$$

$$E(x) = \lambda$$

$$\text{Var}(x) = \lambda$$

Poisson 분포는 이산형 분포이므로 해당 중첩 곡선은 각 정수에서 도약이 나타나는 단계 함수입니다 .

음이항 적합

음이항 분포는 지정된 실패 횟수에 도달하기 전의 성공 횟수를 모델링하는 데 유용합니다 . 다음 파라미터화 계산식에는 평균 모수 λ 와 산포 모수 σ 가 포함되어 있습니다 .

$$\text{pmf: } \frac{\Gamma[x + (1/\sigma)]}{\Gamma[x + 1]\Gamma[1/\sigma]} \left[\frac{(\lambda\sigma)^x}{(1 + \lambda\sigma)^{x + (1/\sigma)}} \right], x = 0, 1, 2, \dots$$

$$E(x) = \lambda$$

$$\text{Var}(x) = \lambda + \sigma\lambda^2$$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

음이항 분포와 감마 Poisson 분포 간의 관계

음이항 분포는 감마 Poisson 분포와 동등합니다 . 감마 Poisson 분포는 데이터가 몇 가지 Poisson(μ) 분포의 조합이고 각 Poisson(μ) 분포의 μ 이 다른 경우에 유용합니다 .

감마 Poisson 분포는 $x|\mu$ 가 Poisson 분포를 따르고 μ 가 Gamma(α, τ) 분포를 따른다고 가정한 결과입니다 . 감마 Poisson 에는 $\lambda = \alpha\tau$ 및 $\sigma = \tau + 1$ 모수가 있습니다 . σ 모수는 산포 모수입니다 . $\sigma > 1$ 이면 과대산포가 나타납니다 . 즉, x 에서 Poisson 만으로 설명할 경우보다 더 많은 변동이 나타납니다 . $\sigma = 1$ 이면 x 가 Poisson(λ) 으로 축소됩니다 .

$$\text{pmf: } \frac{\Gamma\left(x + \frac{\lambda}{\sigma - 1}\right)}{\Gamma(x + 1)\Gamma\left(\frac{\lambda}{\sigma - 1}\right)} \left(\frac{\sigma - 1}{\sigma}\right)^x \sigma^{-\frac{\lambda}{\sigma - 1}} \quad (\text{단}, 0 < \lambda, 1 \leq \sigma, x = 0, 1, 2, \dots)$$

$$E(x) = \lambda$$

$$\text{Var}(x) = \lambda\sigma$$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

감마 Poisson 분포는 $\sigma_{\text{negbin}} = (\sigma_{\text{gp}} - 1) / \lambda_{\text{gp}}$ 인 음이항 분포와 동등합니다 .

모수 λ 및 σ 를 갖는 감마 Poisson 분포를 모수 λ 를 갖는 Poisson 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoGammaPoisson.jsl 을 실행하십시오 .

ZI Poisson 적합

ZI(영과잉) Poisson 분포는 척도 모수 $\lambda > 0$ 과 영과잉 모수 π 를 갖습니다 .

$$\text{pmf: } \begin{cases} \pi + (1 - \pi)\exp[-\lambda] & \text{단}, x = 0 \\ (1 - \pi)\frac{\lambda^x}{x!}\exp[-\lambda] & \text{단}, x = 1, 2, \dots \end{cases}$$

$$E(x) = (1 - \pi)\lambda$$

$$\text{Var}(x) = \lambda(1 - \pi)(1 + \lambda\pi)$$

ZI 음이항 적합

ZI(영과잉) 음이항 분포는 척도 모수 $\lambda > 0$, 산포 모수 $\sigma > 0$ 및 영과잉 모수 π 를 갖습니다 .

$$\text{pmf: } \begin{cases} \pi + (1 - \pi)(1 + \lambda\sigma)^{-(1/\sigma)} & \text{단}, x = 0 \\ (1 - \pi)\frac{\Gamma[x + (1/\sigma)]}{\Gamma[x + 1]\Gamma[1/\sigma]}\left[\frac{(\lambda\sigma)^x}{(1 + \lambda\sigma)^{x + (1/\sigma)}}\right] & \text{단}, x = 1, 2, \dots \end{cases}$$

$$E(x) = (1 - \pi)\lambda$$

$$\text{Var}(x) = \lambda(1 - \pi)[1 + \lambda(\sigma + \pi)]$$

이항 적합

"이항 적합" 옵션에서는 두 가지 형식의 데이터, 즉 표본 크기 상수나 표본 크기를 포함하는 열이 허용됩니다.

$$\text{pmf: } \binom{n}{x} p^x (1-p)^{n-x} \quad (\text{단, } 0 \leq p \leq 1, x = 0, 1, 2, \dots, n)$$

$$E(x) = np$$

$$\text{Var}(x) = np(1-p)$$

여기서 n 은 독립 시행 횟수입니다.

참고: 이항 분포 모수의 신뢰 구간은 점수 구간입니다. 자세한 내용은 Agresti and Coull 연구 자료 (1998) 에서 확인하십시오.

베타 이항 적합

베타 이항 분포는 데이터가 몇 가지 Binomial(p) 분포의 조합이고 각 Binomial(p) 분포의 p 가 다른 경우에 유용합니다. 한 예로, 제조 라인 간에 평균 결함 수 (p) 가 다른 경우 여러 라인에서 결합된 전체 결함 수를 들 수 있습니다.

베타 이항 분포는 $x | \pi$ 가 Binomial(n, π) 분포를 따르고 π 가 Beta(α, β) 를 따른다고 가정한 결과입니다. 베타 이항 분포는 모수 $p = \alpha / (\alpha + \beta)$ 및 $\delta = 1 / (\alpha + \beta + 1)$ 를 갖습니다. 모수 δ 는 산포 모수입니다. $\delta > 1$ 이면 과대산포가 나타납니다. 즉, x 에서 이항 분포만으로 설명할 경우보다 더 많은 변동이 나타납니다. $\delta < 1$ 이면 과소산포가 나타납니다. $\delta = 1$ 이면 x 가 Binomial(n, p) 분포를 따릅니다. 베타 이항 분포는 $n \geq 2$ 일 때만 존재합니다.

$$\text{pmf: } \binom{n}{x} \frac{\Gamma\left(\frac{1}{\delta} - 1\right) \Gamma\left[x + p\left(\frac{1}{\delta} - 1\right)\right] \Gamma\left[n - x + (1-p)\left(\frac{1}{\delta} - 1\right)\right]}{\Gamma\left[p\left(\frac{1}{\delta} - 1\right)\right] \Gamma\left[(1-p)\left(\frac{1}{\delta} - 1\right)\right] \Gamma\left(n + \frac{1}{\delta} - 1\right)}$$

$$(\text{단, } 0 \leq p \leq 1, \max(-\frac{p}{n-p-1}, -\frac{1-p}{n-2+p}) \leq \delta \leq 1, x = 0, 1, 2, \dots, n)$$

$$E(x) = np$$

$$\text{Var}(x) = np(1-p)[1 + (n-1)\delta]$$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다.

$x | \pi \sim \text{Binomial}(n, \pi)$ 인 반면 $\pi \sim \text{Beta}(\alpha, \beta)$ 임을 기억하십시오. 모수 $p = \alpha / (\alpha + \beta)$ 및 $\delta = 1 / (\alpha + \beta + 1)$ 는 플랫폼에 의해 추정됩니다. α 및 β 의 추정값을 구하려면 다음 계산식을 사용합니다.

$$\hat{\alpha} = \hat{p} \left(\frac{1 - \hat{\delta}}{\hat{\delta}} \right)$$

$$\hat{\beta} = (1 - \hat{p}) \left(\frac{1 - \hat{\delta}}{\hat{\delta}} \right)$$

δ 의 추정값이 0인 경우에는 이 계산식이 적용되지 않습니다. 이 경우에는 베타 이항 분포가 $\text{Binomial}(n, p)$ 로 축소된 것이며 \hat{p} 는 p 의 추정값입니다.

베타 이항 분포 모수의 신뢰 구간은 프로파일 우도 구간입니다.

산포 모수 δ 를 갖는 베타 이항 분포를 모수 p 및 $n=20$ 을 갖는 이항 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 `demoBetaBinomial.jsl`을 실행하십시오.

레거시 분포 적합기에 대한 상세 정보

JMP 15에서는 몇 가지 분포 적합 기능이 업데이트되었습니다. 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다. 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다.

- "분포 적합 옵션 (레거시)"
- "연속형 적합 분포에 대한 통계 상세 정보 (레거시)"
- "이산형 적합 분포에 대한 통계 상세 정보 (레거시)"
- "적합 분위수에 대한 통계 상세 정보 (레거시)"
- "분포 적합 옵션에 대한 통계 상세 정보 (레거시)"

분포 적합 옵션 (레거시)

연속형 변수에 분포를 적합시키려면 연속형 적합 또는 이산형 적합 옵션을 사용합니다.

참고: JMP 15에서는 몇 가지 분포 적합 기능이 업데이트되었습니다. 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다. 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다.

그러면 히스토그램에 곡선이 중첩 표시되고 보고서 창에 "모수 추정값" 보고서가 추가됩니다. 빨간색 삼각형 메뉴에는 추가 옵션이 포함되어 있습니다. 자세한 내용은 "[분포 적합 옵션 \(레거시\)](#)"(93 페이지)에서 확인하십시오.

참고: 수명 분포 플랫폼에도 분포 적합을 위한 옵션이 포함되어 있으며 여기에는 다른 파라미터화 방법이 사용되고 중도절단이 허용될 수 있습니다. 자세한 내용은 **Reliability and Survival Methods**의 "Life Distribution"장에서 확인하십시오.

연속형 적합 (레거시)

이 섹션에서는 " 레거시 적합기 " 하위 메뉴의 분포 중 업데이트된 " 연속형 적합 " 옵션의 해당 분포와는 다른 분포에 대해 설명합니다 .

- Weibull 분포 , 임계가 있는 Weibull 분포 및 극단값 분포는 특히 기계 장치 및 생물학 분야에 서 수명 길이를 추정하는 데 적절한 모형을 제공하는 경우가 많습니다 .
- 감마 분포는 0 으로 제한되며 형태가 유연합니다 .
- 베타 분포는 간격 0,1 사이에 들어가도록 제한된 확률 변수의 동작을 모델링하는 데 유용합니다 . 예를 들어 비율은 항상 0 에서 1 사이입니다 .
- 평활 곡선 분포는 비모수 밀도 추정 (커널 밀도 추정) 을 사용하여 평활 곡선을 적합시킵니다 . 평활 곡선은 히스토그램에 중첩 표시되며 그림 아래에는 슬라이더가 나타납니다 . 슬라이더로 커널 표준편차를 변경하여 평활 정도를 조절할 수 있습니다 . 초기 커널 표준편차 추정 값은 데이터의 표준편차를 사용하여 계산됩니다 .
- Johnson Su, Johnson Sb 및 Johnson Sl 분포는 왜도와 첨도의 가능한 모든 조합을 지원하므로 데이터 적합 기능에 유용합니다 .
- GLog(일반화 로그) 분포는 생물학적 분석 데이터처럼 거의 정규 분포를 따르지 않으며 대개 분산이 일정하지 않은 데이터를 적합시키는 데 유용합니다 .

모든 분포 비교

모두 옵션은 변수에 적용 가능한 모든 연속형 분포를 적합시킵니다 . " 분포 비교 " 보고서에는 각 적합 분포에 대한 통계량이 포함됩니다 . 체크박스를 사용하여 선택한 분포에 대한 적합 보고서를 표시하거나 숨기고 분포 곡선을 중첩 표시할 수 있습니다 . 기본적으로는 최량 적합 분포가 선택됩니다 .

분포 표시 목록은 AICc 를 기준으로 오름차순 정렬됩니다 .

변수에 음수 값이 포함되었으면 양수 값을 갖는 데이터를 필요로 하는 분포는 분포 표시 목록에 포함되지 않습니다 . 이 명령으로는 연속형 분포만 적합됩니다 . 베타 분포와 Johnson Sb 분포처럼 임계 모수가 있는 분포는 가능한 분포 목록에 포함되지 않습니다 .

관련 정보

통계 상세 정보는 다음 섹션에서 확인하십시오 .

- " 연속형 적합 분포에 대한 통계 상세 정보 (레거시)" (96 페이지)
- " 적합 분위수에 대한 통계 상세 정보 (레거시)" (102 페이지)
- " 분포 적합 옵션 (레거시)" (93 페이지)

이산형 적합 (레거시)

" 이산형 적합 " 옵션은 모든 데이터 값이 정수일 때 사용할 수 있습니다 . 이산형 변수에 분포 (예 : Poisson 또는 이항 분포) 를 적합시키려면 " 이산형 적합 " 옵션을 사용합니다 . 사용 가능한 분포는 다음과 같습니다 .

- Poisson
- 감마 Poisson
- 이항
- 베타 이항

관련 정보

통계 상세 정보는 다음 섹션에서 확인하십시오 .

- " 이산형 적합 분포에 대한 통계 상세 정보 (레거시)" (101 페이지)
- " 적합 분위수에 대한 통계 상세 정보 (레거시)" (102 페이지)
- " 분포 적합 옵션 (레거시)" (93 페이지)

분포 적합 옵션 (레거시)

각 적합 분포 보고서에는 추가 옵션이 포함된 빨간색 삼각형 메뉴가 있습니다 .

진단 그림 분위수 그림 또는 확률도를 생성합니다 . 자세한 내용은 " **진단 그림** " (94 페이지) 에서 확인하십시오 .

밀도 곡선 분포의 추정 모수를 사용하여 히스토그램에 밀도 곡선을 중첩 표시합니다 .

적합도 적합 분포의 적합도 검정을 계산합니다 . 자세한 내용은 " **적합도** " (95 페이지) 에서 확인하십시오 .

모수 고정 모수를 고정하고 고정되지 않은 모수를 다시 추정할 수 있습니다 . 새 모수를 검정하여 해당 모수가 데이터를 적합시키는지 여부를 확인하는 " 적합성 LR 검정 "(LR: 우도비) 보고서도 표시됩니다 .

분위수 지정한 특정 확률 하한 값의 비척도화 및 비중심화 분위수를 반환합니다 .

K 시그마 규격 한계 설정 공정의 규격 한계를 모르는 경우 분포를 기준으로 규격 한계를 설정하려면 이 옵션을 사용합니다 .

일반적으로 규격 한계는 엔지니어링 고려 사항을 사용하여 구합니다 . 엔지니어링 고려 사항이 없고 데이터가 잘 작동하는 공정에서 얻은 것일 경우 적합 분포의 분위수를 사용하면 규격 한계를 설정하는 데 도움이 될 수 있습니다 . 자세한 내용은 " **K 시그마 규격 한계 설정** " (104 페이지) 에서 확인하십시오 .

규격 한계 지정한 규격 한계와 목표값을 기준으로 표준 공정 능력 지수의 일반화 값을 계산합니다 . 자세한 내용은 " **규격 한계** " (95 페이지) 에서 확인하십시오 .

적합 분위수 저장 적합 분위수 값을 데이터 테이블에 새 열로 저장합니다. 자세한 내용은 "[적합 분위수에 대한 통계 상세 정보 \(레거시\)](#)"(102 페이지)에서 확인하십시오.

밀도 계산식 저장 현재 데이터 테이블에 밀도 계산식으로 계산된 적합 값을 포함하는 새 열을 생성합니다. 밀도 계산식에서는 추정된 모수 값을 사용합니다.

변환 저장 새 열을 생성하고 계산식을 저장합니다. 이 계산식은 적합 분포를 사용하여 열을 정규 분포로 변환할 수 있습니다. 이 옵션은 Johnson 분포, GLog 분포 또는 SHASH 분포 중 하나를 적합시키는 경우에만 사용할 수 있습니다.

적합 제거 보고서 창에서 분포 적합을 제거합니다.

진단 그림

진단 그림 옵션은 분위수 그림 또는 확률도를 생성합니다. 적합 분포에 따라 다음 네 가지 형식 중 하나의 그림이 표시됩니다.

적합 분위수 대 데이터

- 임계가 있는 Weibull
- 감마
- 베타
- Poisson
- 감마 Poisson
- 이항
- 베타 이항

적합 확률 대 데이터

- 정규
- 정규 혼합
- 지수

적합 확률 대 로그 척도 데이터

- Weibull
- 로그 정규
- 극단값

적합 확률 대 표준 정규 분위수

- SHASH
- Johnson SI

- Johnson Sb
- Johnson Su
- GLog

진단 그림의 빨간색 삼각형 메뉴에서는 다음과 같은 옵션을 사용할 수 있습니다.

회전 x 축과 y 축을 서로 바꿉니다.

신뢰 한계 정규 분위수 그림의 경우 Lilliefors 95% 신뢰 한계를 그리고, 다른 모든 분위수 그림의 경우 $\alpha = 0.001$ 이고 $b = 0.99$ 인 95% 균등 정밀 띠를 그립니다 (Meeker and Escobar 1998).

적합선 직선으로 된 대각 참조선을 그립니다. 변수가 선택된 분포에 적합하면 값이 이 참조선 주위에서 나타납니다.

중앙값 참조선 반응 중앙값에 가로선을 그립니다.

적합도

적합도 옵션은 적합 분포의 적합도 검정을 계산합니다. 적합도 검정은 카이제곱 검정이 아니라 EDF(경험적 분포 함수) 검정입니다. EDF 검정은 카이제곱 검정에 비해 향상된 검정력과 히스토그램 중간점의 불변성을 포함한 장점을 제공합니다.

- 정규 분포의 경우 표본 크기가 2,000 보다 작거나 같으면 Shapiro-Wilk 정규성 검정이 보고됩니다. 2,000 보다 큰 표본의 경우에는 KSL 검정이 계산됩니다.
- 표본 크기가 30 보다 작거나 같은 이산형 분포의 경우 거의 정확한 검정을 구성하기 위해 결합된 두 개의 단측 정확 Kolmogorov 검정을 사용하여 적합도 검정이 구성됩니다. 자세한 내용은 Conover 연구 자료(1972)에서 확인하십시오. 표본 크기가 30 보다 큰 경우에는 Pearson 카이제곱 적합도 검정이 수행됩니다.

관련 정보

- 통계 상세 정보는 "[분포 적합 옵션 \(레거시\)](#)"(93 페이지)에서 확인하십시오.

규격 한계

규격 한계 옵션을 선택하면 규격 한계와 목표값을 입력할 수 있는 창이 열립니다. 규격 한계와 목표값을 입력하고 나면 일반화된 표준 공정 능력 지수가 계산됩니다. 정규 분포의 경우 3σ 는 하위 0.135 백분위수부터 중앙값(또는 평균)까지의 거리이자 중앙값(또는 평균)부터 상위 99.865 백분위수까지의 거리입니다. 이러한 백분위수는 적합 분포에서 추정되며, 표준 계산식에서 3σ 는 적절한 백분위수부터 중앙값까지의 거리로 대체됩니다.

관련 정보

- 통계 상세 정보는 "[분포 적합 옵션 \(레거시\)](#)"(93 페이지)에서 확인하십시오.

연속형 적합 분포에 대한 통계 상세 정보 (레거시)

이 섹션에는 " 연속형 적합 " 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다 .

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

정규

정규 분포 적합에 대한 자세한 내용은 " **정규 적합** "(82 페이지) 에서 확인하십시오 .

로그 정규

로그 정규 분포 적합에 대한 자세한 내용은 " **로그 정규 적합** "(84 페이지) 에서 확인하십시오 .

Weibull, 임계가 있는 Weibull 및 극단값

Weibull 분포는 α (척도) 및 β (형태) 값에 따라 다른 형태로 나타냅니다 . 이 함수는 특히 기계 장치 및 생물학 분야에서 수명 길이를 추정하는 데 적절한 모형을 제공하는 경우가 많습니다 .

Weibull 분포와 임계가 있는 Weibull 분포의 pdf 는 다음과 같습니다 .

$$\text{pdf: } \frac{\beta}{\alpha} \left(\frac{x-\theta}{\alpha} \right)^{\beta-1} \exp \left[- \left(\frac{x-\theta}{\alpha} \right)^{\beta} \right] \quad (\text{단, } \alpha, \beta > 0, \theta < x)$$

$$E(x) = \theta + \alpha \Gamma \left(1 + \frac{1}{\beta} \right)$$

$$\text{Var}(x) = \alpha^2 \left\{ \Gamma \left(1 + \frac{2}{\beta} \right) - \Gamma^2 \left(1 + \frac{1}{\beta} \right) \right\}$$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

Weibull 옵션은 임계 모수 (θ) 를 0 으로 설정합니다 . **임계가 있는 Weibull** 옵션은 최소 관측값을 사용하여 임계 모수 (θ) 를 추정하고 , 나머지 관측값을 사용하여 α 및 β 를 추정합니다 . 필요한 임계를 알고 있으면 **모수 고정** 옵션을 사용하여 임계를 설정합니다 . 자세한 내용은 " **분포 적합 옵션** "(61 페이지) 에서 확인하십시오 .

참고 : 분포 플랫폼에서 임계가 있는 Weibull 분포의 임계 모수에 사용되는 추정 기법은 수명 분포 플랫폼에서 사용되는 추정 기법과는 다릅니다 . 이 분포를 적합시키는 데는 수명 분포에서의 추정 방법을 사용하는 것이 좋습니다 . 자세한 내용은 Reliability and Survival Methods 의 "Life Distribution" 장에서 확인하십시오 .

극단값 분포는 $\delta = 1/\beta$ 및 $\lambda = \ln(\alpha)$ 로 재파라미터화된 2 모수 Weibull(α, β) 분포와 동등합니다.

지수

지수 분포 적합에 대한 자세한 내용은 "[지수 적합](#)"(83 페이지) 에서 확인하십시오.

감마

감마 적합 옵션은 감마 분포 모수 $\alpha > 0$ 및 $\sigma > 0$ 을 추정합니다. 적합 감마 보고서에서 모수 α (알파) 는 형태 또는 곡률을 설명합니다. 모수 σ (시그마) 는 분포의 척도 모수입니다. 세 번째 모수인 θ (임계) 는 하한점 모수입니다. 음수 값이 없는 한 이 모수는 기본적으로 0 으로 설정되어 있습니다. **모수 고정** 옵션을 사용하여 이 값을 직접 설정할 수도 있습니다. 자세한 내용은 "[분포 적합 옵션](#)"(61 페이지) 에서 확인하십시오.

$$\text{pdf: } \frac{1}{\Gamma(\alpha)\sigma^\alpha} (x - \theta)^{\alpha-1} \exp(-(x - \theta)/\sigma) \quad (\text{단, } 0 \leq x, 0 < \alpha, \sigma)$$

$$E(x) = \alpha\sigma + \theta$$

$$\text{Var}(x) = \alpha\sigma^2$$

- 표준 감마 분포의 경우 $\sigma = 1$ 입니다. 시그마는 값이 1 이 아닐 경우 가로 축을 따라 분포를 늘리거나 줄이므로 척도 모수라고 합니다.
- 카이제곱 $\chi^2_{(v)}$ 분포는 $\sigma = 2$ 이고 $\alpha = v/2$ 이며 $\theta = 0$ 일 때 발생합니다.
- 지수 분포는 $\alpha = 1$ 이고 $\theta = 0$ 일 때 발생하는 감마 곡선에 속합니다.

표준 감마 밀도 함수는 $\alpha \leq 1$ 일 때는 감소하기만 합니다. $\alpha > 1$ 일 때 밀도 함수는 0 에서 시작하여 최대값까지 증가하다가 감소합니다.

베타

표준 베타 분포는 간격 0,1 사이에 들어가도록 제한된 확률 변수의 동작을 모델링하는 데 유용합니다. 예를 들어 비율은 항상 0 에서 1 사이입니다. **베타** 적합 옵션은 두 개의 형태 모수 $\alpha > 0$ 및 $\beta > 0$ 과 두 개의 임계 모수 θ 및 σ 를 추정합니다. 임계 하한은 θ 로 나타내고 임계 상한은 $\theta + \sigma$ 로 나타냅니다. 베타 분포는 $\theta \leq x \leq (\theta + \sigma)$ 구간에서만 값을 갖습니다. θ 는 최소값을 사용하여 추정되고 σ 는 범위를 사용하여 추정됩니다. 표준 베타 분포는 $\theta = 0$ 이고 $\sigma = 1$ 일 때 발생합니다.

모수 고정 옵션을 사용하면 모수를 고정 값으로 설정할 수 있습니다. 임계 상한은 최대 데이터 값보다 크거나 같아야 하며, 임계 하한은 최소 데이터 값보다 작거나 같아야 합니다. "모수 고정" 옵션에 대한 자세한 내용은 "[분포 적합 옵션](#)"(61 페이지) 에서 확인하십시오.

$$\text{pdf: } \frac{1}{B(\alpha, \beta)\sigma^{\alpha+\beta-1}} (x - \theta)^{\alpha-1} (\theta + \sigma - x)^{\beta-1} \quad (\text{단, } \theta \leq x \leq \theta + \sigma, 0 < \alpha, \beta)$$

$$E(x) = \theta + \sigma \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(x) = \frac{\sigma^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

여기서 $B(\cdot)$ 는 베타 함수입니다 .

정규 혼합

정규 혼합 분포 적합에 대한 자세한 내용은 "[정규 2 혼합 적합 및 정규 3 혼합 적합](#)"(85 페이지) 에서 확인하십시오 .

평활 곡선

평활 곡선 옵션은 비모수 밀도 추정 (커널 밀도 추정) 을 사용하여 평활 곡선을 적합시킵니다 . 평활 곡선은 히스토그램에 중첩 표시되며 그림 아래에는 슬라이더가 나타납니다 . 슬라이더로 커널 표준편차를 변경하여 평활 정도를 조절할 수 있습니다 . 초기 커널 표준편차 추정값은 데이터의 표준편차를 사용하여 계산됩니다 .

SHASH

SHASH 분포 적합에 대한 자세한 내용은 "[SHASH 적합](#)"(83 페이지) 에서 확인하십시오 .

Johnson Su, Johnson Sb, Johnson Sl

Johnson 분포 체계에는 세 가지 분포가 포함되어 있으며 이러한 분포는 모두 변환된 정규 분포를 기반으로 합니다 . 이 세 가지 분포는 다음과 같습니다 .

- Johnson Su - 경계가 없습니다 .
- Johnson Sb - 양쪽 꼬리에 경계가 있습니다 . 이 경계는 추정 가능한 모수에 의해 정의됩니다 .
- Johnson Sl - 한쪽 꼬리에 경계가 있습니다 . 이 경계는 추정 가능한 모수에 의해 정의됩니다 . Johnson Sl 계열에는 로그 정규 분포 계열이 포함됩니다 .

S 는 체계 , 즉 범위의 서브스크립트를 나타냅니다 . JMP 에는 다른 방법이 구현되기는 했지만 , Slifker and Shapiro 연구 자료(1980)에서 특정 Johnson 체계의 선택 기준을 확인할 수 있습니다 .

Johnson 분포는 유연하기 때문에 널리 사용됩니다 . 특히 Johnson 분포 체계는 왜도와 첨도의 가능한 모든 조합을 지원하므로 데이터 적합 기능 면에서 주목됩니다 .

Z 가 표준 정규 변량인 경우 이 분포 체계는 다음과 같이 정의됩니다 .

$$Z = \gamma + \delta f(Y)$$

다음은 Johnson Su 의 경우에 해당하는 각 요소에 대한 설명입니다 .

$$f(Y) = \ln\left(Y + \sqrt{1 + Y^2}\right) = \sinh^{-1}Y$$

$$Y = \frac{X - \theta}{\sigma} \quad -\infty < X < \infty$$

다음은 Johnson Sb 의 경우에 해당하는 각 요소에 대한 설명입니다 .

$$f(Y) = \ln\left(\frac{Y}{1-Y}\right)$$

$$Y = \frac{X - \theta}{\sigma} \quad \theta < X < \theta + \sigma$$

다음은 Johnson SI 의 경우 (단 , $\sigma = \pm 1$) 에 해당하는 각 요소에 대한 설명입니다 .

$$f(Y) = \ln(Y)$$

$$Y = \frac{X - \theta}{\sigma} \quad \begin{array}{ll} \theta < X < \infty & \text{단, } \sigma = 1 \\ -\infty < X < \theta & \text{단, } \sigma = -1 \end{array}$$

Johnson Su

$$\text{pdf: } \frac{\delta}{\sigma} \left[1 + \left(\frac{x - \theta}{\sigma} \right)^2 \right]^{-1/2} \phi \left[\gamma + \delta \sinh^{-1} \left(\frac{x - \theta}{\sigma} \right) \right] \quad (\text{단, } -\infty < x, \theta, \gamma < \infty, 0 < \theta, \delta)$$

Johnson Sb

$$\text{pdf: } \phi \left[\gamma + \delta \ln \left(\frac{x - \theta}{\sigma - (x - \theta)} \right) \right] \left(\frac{\delta \sigma}{(x - \theta)(\sigma - (x - \theta))} \right) \quad (\text{단, } \theta < x < \theta + \sigma, 0 < \sigma)$$

Johnson SI

$$\text{pdf: } \frac{\delta}{|x - \theta|} \phi \left[\gamma + \delta \ln \left(\frac{x - \theta}{\sigma} \right) \right] \quad (\text{단, } \sigma = 1 \text{ 일 경우 } \theta < x, \sigma = -1 \text{ 일 경우 } \theta > x)$$

여기서 $\phi(\cdot)$ 는 표준 정규 pdf 입니다 .

다음 사항에 유의하십시오 .

- 작업 순서와 컴퓨터 정밀도로 인해 컴퓨터마다 모수 추정값이 다를 수 있습니다 .
- 기본 보고서에는 모수 신뢰 구간이 숨겨져 있습니다 . Johnson 분포는 정규 분포로 변환되므로 이 경우에는 모수 신뢰 구간이 그다지 의미가 없습니다 . 모수 신뢰 구간을 표시하려면 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 하한 및 95% 상한**을 선택합니다 .

GLog(일반화 로그)

이 분포는 생물학적 분석 데이터처럼 거의 정규 분포를 따르지 않으며 대개 분산이 일정하지 않은 데이터를 적합시키는 데 유용합니다. GLog 분포는 μ (위치), σ (척도) 및 λ (형태) 모수로 설명됩니다.

$$\text{pdf: } \phi \left\{ \frac{1}{\sigma} \left[\log \left(\frac{x + \sqrt{x^2 + \lambda^2}}{2} \right) - \mu \right] \right\} \frac{x + \sqrt{x^2 + \lambda^2}}{\sigma(x^2 + \lambda^2 + x\sqrt{x^2 + \lambda^2})}$$

(단, $0 \leq \lambda, 0 < \sigma, -\infty < \mu < \infty$)

GLog 분포는 정규 분포로 변환되며, 다음과 같은 관계에 있습니다.

$$z = \frac{1}{\sigma} \left[\log \left(\frac{x + \sqrt{x^2 + \lambda^2}}{2} \right) - \mu \right] \sim N(0,1) \text{ 이면 } x \sim \text{Glog}(\mu, \sigma, \lambda) \text{ 입니다.}$$

$\lambda = 0$ 이면 GLog 가 LogNormal(μ, σ) 로 축소됩니다.

참고 : 기본 보고서에는 모수 신뢰 구간이 숨겨져 있습니다. GLog 분포는 정규 분포로 변환되므로 이 경우에는 모수 신뢰 구간이 그다지 의미가 없습니다. 모수 신뢰 구간을 표시하려면 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 하한** 및 **95% 상한**을 선택합니다.

모두

" 분포 비교 " 보고서에서는 분포 목록이 AICc 를 기준으로 오름차순 정렬됩니다.

AICc 의 계산식은 다음과 같습니다.

$$\text{AICc} = -2\log L + 2v + \frac{2v(v+1)}{n-(v+1)}$$

다음은 각 요소에 대한 설명입니다.

- $\log L$ 은 로그 우도입니다.
- n 은 표본 크기입니다.
- v 은 모수의 개수입니다.

열에 음수 값이 포함되었으면 양수 값을 갖는 데이터를 필요로 하는 분포는 분포 목록에 포함되지 않습니다. 연속형 분포만 나열됩니다. 베타 분포와 Johnson Sb 분포와 같이 임계 모수가 있는 분포는 가능한 분포 목록에 포함되지 않습니다.

이산형 적합 분포에 대한 통계 상세 정보 (레거시)

이 섹션에는 " 이산형 적합 " 메뉴의 옵션에 대한 통계 상세 정보가 포함되어 있습니다 .

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용**을 선택하면 이러한 기능을 사용할 수 있습니다 .

Poisson

Poisson 분포 적합에 대한 자세한 내용은 "**Poisson 적합**"(88 페이지) 에서 확인하십시오 .

감마 Poisson

이 분포는 데이터가 몇 가지 Poisson(μ) 분포의 조합이고 각 Poisson(μ) 분포의 μ 이 다른 경우에 유용합니다 . 한 예로 , 교차로 간에 평균 사고 발생 횟수 (μ) 가 다른 경우 여러 교차로에서 결합된 전체 사고 발생 횟수를 들 수 있습니다 .

감마 Poisson 분포는 $x|\mu$ 가 Poisson 분포를 따르고 μ 가 Gamma(α, τ) 분포를 따른다고 가정한 결과입니다 . 감마 Poisson 에는 $\lambda = \alpha\tau$ 및 $\sigma = \tau + 1$ 모수가 있습니다 . σ 모수는 산포 모수입니다 . $\sigma > 1$ 이면 과대산포가 나타납니다 . 즉 , x 에서 Poisson 만으로 설명할 경우보다 더 많은 변동이 나타납니다 . $\sigma = 1$ 이면 x 가 Poisson(λ) 으로 축소됩니다 .

$$\text{pmf: } \frac{\Gamma\left(x + \frac{\lambda}{\sigma - 1}\right)}{\Gamma(x + 1)\Gamma\left(\frac{\lambda}{\sigma - 1}\right)} \left(\frac{\sigma - 1}{\sigma}\right)^x \sigma^{-\frac{\lambda}{\sigma - 1}} \quad (\text{단, } 0 < \lambda, 1 \leq \sigma, x = 0, 1, 2, \dots)$$

$$E(x) = \lambda$$

$$\text{Var}(x) = \lambda\sigma$$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

$x|\mu \sim \text{Poisson}(\mu)$ 인 반면 $\mu \sim \text{Gamma}(\alpha, \tau)$ 임을 기억하십시오 . 이 플랫폼은 $\lambda = \alpha\tau$ 이고 $\sigma = \tau + 1$ 이라고 추정합니다 . α 및 τ 의 추정값을 구하려면 다음 계산식을 사용합니다 .

$$\hat{\tau} = \hat{\sigma} - 1$$

$$\hat{\alpha} = \frac{\hat{\lambda}}{\hat{\tau}}$$

σ 의 추정값이 1 인 경우에는 이 계산식이 적용되지 않습니다 . 이 경우에는 감마 Poisson 이 Poisson(λ) 으로 축소된 것이며 $\hat{\lambda}$ 는 λ 의 추정값입니다 .

α 의 추정값이 정수인 경우 감마 Poisson 은 다음 pmf 를 갖는 음이항과 동등합니다 .

$$p(y) = \binom{y+r-1}{y} p^r (1-p)^y \quad (\text{단}, 0 \leq y)$$

이때 $r = \alpha$ 이고 $(1-p)/p = \tau$ 입니다 .

모수 λ 및 σ 를 갖는 감마 Poisson 분포를 모수 λ 를 갖는 Poisson 분포와 비교하려면 JMP Samples/Scripts 폴더에 있는 demoGammaPoisson.jsl 을 실행하십시오 .

이항

이항 분포 적합에 대한 자세한 내용은 "[이항 적합](#)"(90 페이지) 에서 확인하십시오 .

베타 이항

베타 이항 분포 적합에 대한 자세한 내용은 "[베타 이항 적합](#)"(90 페이지) 에서 확인하십시오 .

적합 분위수에 대한 통계 상세 정보 (레거시)

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용** 을 선택하면 이러한 기능을 사용할 수 있습니다 .

진단 그림의 적합 분위수와 **적합 분위수 저장** 명령으로 저장된 적합 분위수는 다음 방법을 사용하여 구합니다 .

1. 데이터를 정렬하고 순위화합니다 . 동일 값에는 서로 다른 순위가 할당됩니다 .
2. $p_{[i]} = \text{rank}_{[i]} / (n+1)$ 를 계산합니다 .
3. $\text{quantile}_{[i]} = \text{Quantile}_d(p_{[i]})$ 를 계산합니다 . 여기서 Quantile_d 는 특정 적합 분포에 대한 분위수 함수이고 $i = 1, 2, \dots, n$ 입니다 .

분포 적합 옵션에 대한 통계 상세 정보 (레거시)

이 섹션에서는 분포 적합을 위한 적합도 검정과 적합된 분포와 관련된 규격 한계의 통계 상세 정보를 설명합니다 .

참고 : JMP 15 에서는 몇 가지 분포 적합 기능이 업데이트되었습니다 . 이 섹션에서는 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 자세히 설명합니다 . 변수의 빨간색 삼각형 메뉴에서 **연속형 적합 > 레거시 적합기 사용** 을 선택하면 이러한 기능을 사용할 수 있습니다 .

적합도

표 3.2 JMP 적합도 검정 설명

분포	모수	적합도 검정
정규 ^a	μ 및 σ 를 알 수 없음	Shapiro-Wilk($n \leq 2000$ 인 경우), Kolmogorov-Smirnov-Lillefors ($n > 2000$ 인 경우)
	μ 및 σ 를 모두 알고 있음	Kolmogorov-Smirnov-Lillefors
	μ 또는 σ 중 하나를 알고 있음	(없음)
로그 정규	μ 및 σ 를 알고 있거나 알 수 없음	Kolmogorov D
Weibull	α 및 β 를 알고 있거나 알 수 없음	Cramér-von Mises W^2
임계가 있는 Weibull	α, β 및 θ 를 알고 있거나 알 수 없음	Cramér-von Mises W^2
극단값	α 및 β 를 알고 있거나 알 수 없음	Cramér-von Mises W^2
지수	σ 를 알고 있거나 알 수 없음	Kolmogorov D
감마	α 및 σ 를 알고 있음	Cramér-von Mises W^2
	α 또는 σ 중 하나를 알 수 없음	(없음)
베타	α 및 β 를 알고 있음	Kolmogorov D
	α 또는 β 중 하나를 알 수 없음	(없음)
이항	p 를 알고 있거나 알 수 없으며, n 을 알고 있음	Kolmogorov D ($n \leq 30$ 인 경우), Pearson χ^2 ($n > 30$ 인 경우)
베타 이항	p 및 δ 를 알고 있거나 알 수 없음	Kolmogorov D ($n \leq 30$ 인 경우), Pearson χ^2 ($n > 30$ 인 경우)
Poisson	λ 를 알고 있거나 알 수 없음	Kolmogorov D ($n \leq 30$ 인 경우), Pearson χ^2 ($n > 30$ 인 경우)
감마 Poisson	λ 또는 σ 를 알고 있거나 알 수 없음	Kolmogorov D ($n \leq 30$ 인 경우), Pearson χ^2 ($n > 30$ 인 경우)

a. 세 가지 Johnson 분포와 GLog 분포의 경우에는 데이터가 정규 데이터로 변환된 후 적절한 정규성 검정이 수행됩니다.

K 시그마 규격 한계 설정

공정 능력 분석을 위해 K 값을 입력하고 단측 또는 양측을 선택할 수 있습니다. K 표준편차에 해당하는 꼬리 확률은 정규 분포에서 계산됩니다. 이 확률은 적합시킨 특정 분포에 맞는 분위수로 변환됩니다. 결과 분위수는 공정 능력 분석에서 규격 한계에 사용됩니다. 이 옵션은 **분위수** 옵션과 유사하지만 확률 대신 K 를 제공해야 합니다. K 는 규격 한계와 평균 사이의 거리를 나타내는 표준편차 수에 해당합니다.

예를 들어 정규 분포에서 $K=3$ 인 경우 평균의 하위 및 상위 3 표준편차는 각각 0.00135 번째 분위수와 0.99865 번째 분위수에 해당합니다. 규격 하한은 적합 분포의 0.00135 번째 분위수에서 설정되고, 규격 상한은 0.99865 번째 분위수에서 설정됩니다. 공정 능력 분석은 이러한 규격 한계를 기준으로 반환됩니다.

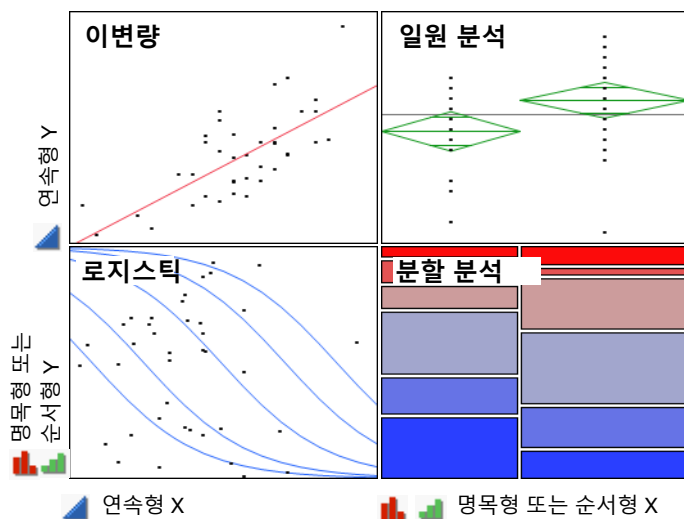
4 장

X 로 Y 적합 소개 두 변수 간의 관계 분석

X로 Y 적합 플랫폼에서는 지정된 X 및 Y 변수 쌍을 모델링 유형에 따라 컨텍스트별로 분석합니다.
다음은 네 가지 분석 유형입니다.

- 이변량 적합
- 일원 분산 분석
- 로지스틱 회귀
- 분할표 분석

그림 4.1 네 가지 분석 유형의 예



목차

X로 Y적합 플랫폼의 개요..... 107

X로 Y적합 플랫폼 시작..... 107

 JMP 시작하기 창에서 특정 분석 시작..... 108

X로 Y 적합 플랫폼의 개요

X로 Y 적합 플랫폼은 네 가지 상세 플랫폼 (또는 분석 유형) 을 모은 것입니다 .

상세 플랫폼	모델링 유형	설명
이변량	연속형 X로 연속형 Y 적합	두 연속형 변수 간의 관계를 분석합니다 . 자세한 내용은 " 이변량 분석 "에서 확인하십시오 .
일원 분석	명목형 또는 순서형 X로 연속형 Y 적합	범주형 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지를 분석합니다 . 자세한 내용은 " 일원 분석 "에서 확인하십시오 .
로지스틱	연속형 X로 명목형 또는 순서형 Y 적합	반응 범주에 대한 확률을 연속형 X 예측 변수에 적합시킵니다 . 자세한 내용은 " 로지스틱 분석 "에서 확인하십시오 .
분할 분석	명목형 또는 순서형 X로 명목형 또는 순서형 Y 적합	범주형 X 요인의 값에 따라 좌우되는 범주형 반응 변수의 분포를 분석합니다 . 자세한 내용은 " 분할 분석 "에서 확인하십시오 .

X로 Y 적합 플랫폼 시작

분석 > X로 Y 적합을 선택하여 X로 Y 적합 플랫폼을 시작할 수 있습니다 .

그림 4.2 X로 Y 적합 시작 창

두 변수 사이의 관계를 모델링합니다.

열 선택

4개 열

- Dose
- Response
- Count
- ln(dose)

선택한 열 역할 지정

Y, 반응

Y, 반응

X, 요인

X, 요인

블록

가중치

빈도

기준

작업

확인

취소

제거

재호출

도움말

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 "시작하기"장에서 확인하십시오.

이변량, 일원 분석, 로지스틱, 분할 분석 이 격자에는 서로 다른 데이터 유형 조합의 분석 결과가 표시됩니다. 열을 할당하면 적용 가능한 플랫폼이 격자 위에 라벨로 표시됩니다.

블록 (선택 사항. 일원 분석 및 분할 분석에만 사용 가능):

- 일원 분석 플랫폼의 경우 블록 변수를 지정하면 교호작용이 없는 이원 분석을 구성하는 두 번째 요인이 식별됩니다. 데이터는 균형을 이루고 있어야 하며 그룹 셀 별로 각 블록 내의 데이터 개수가 같아야 합니다. 블록 변수를 지정할 경우 데이터는 균형을 이루고 있어야 하며 그룹 셀 별로 각 블록 내의 데이터 개수가 같아야 합니다. 그림에서는 Y 변수의 값이 블록 변수에 의해 중심화됩니다.
- 분할 분석 플랫폼의 경우 블록 변수를 지정하면 두 번째 요인이 식별되고 Cochran-Mantel-Haenszel 검정이 수행됩니다.

시작 창에 대한 자세한 내용은 JMP 사용의 "시작하기"장에서 확인하십시오.

JMP 시작하기 창에서 특정 분석 시작

"JMP 시작하기"창에서 특정 분석 (**이변량, 일원 분석, 로지스틱** 또는 **분할 분석**)을 시작할 수 있습니다. 이 옵션을 선택할 경우 분석에 적절한 모델링 유형(Y 및 X 변수)을 지정합니다(표 4.1).

"JMP 시작하기"창에서 특정 분석을 시작하려면 **기본 분석** 범주를 클릭하고 특정 분석을 선택합니다.

플랫폼 시작 옵션은 대부분 동일합니다. 하지만 Y 및 X 플랫폼 버튼 중 일부는 수행하는 특정 분석에 맞게 사용자 정의됩니다.

표 4.1 플랫폼 및 버튼

플랫폼 또는 분석	Y 버튼	X 버튼
X로 Y적합	Y, 반응	X, 요인
이변량	Y, 반응	X, 회귀변수
일원 분석	Y, 반응	X, 그룹화
로지스틱	Y, 범주형 반응	X, 연속형 회귀변수
분할 분석	Y, 반응 범주	X, 그룹화 범주

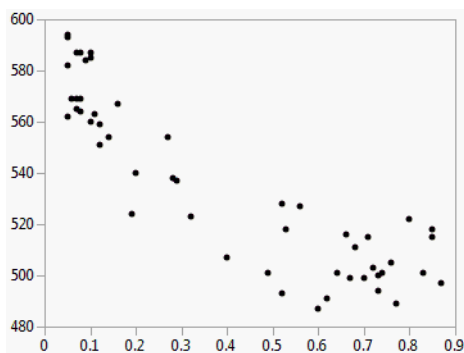
이변량 분석

두 연속형 변수 간의 관계 분석

이변량 플랫폼에서는 두 연속형 변수 사이의 관계를 보여 줍니다. 이 플랫폼은 X 로 Y 적합 플랫폼의 연속형 대 연속형 분석법입니다. 이변량이란 단어는 단순히 관련 변수가 한 개 (단변량) 또는 여러 개 (다변량)가 아니라 두 개임을 의미합니다.

이변량 분석 결과는 산점도로 표시됩니다. 그림의 각 점은 단일 관측값의 X 및 Y 값을 나타냅니다. 즉, 각 점이 두 개의 변수를 나타냅니다. 산점도를 사용하면 두 변수 사이의 관계 정도 및 패턴을 한눈에 확인할 수 있습니다. 대화식으로 단순 선형 회귀, 다항 회귀 등의 다른 적합 유형을 추가할 수 있습니다.

그림 5.1 이변량 분석의 예



목차

이변량 분석의 예	112
이변량 플랫폼 시작	112
이변량 그림	114
적합 옵션	114
적합 옵션	115
적합 옵션 범주	117
동일한 적합 옵션을 여러 번 적용	117
히스토그램 테두리	118
평균 적합	119
평균 적합 보고서	119
선형 적합 및 다항식 적합	120
선형 적합 및 다항식 적합 보고서	120
특수 적합	125
특수 적합 보고서 및 메뉴	126
유연	127
스플라인 적합	127
커널 평활기	128
각 값 적합	129
직교 적합	130
직교 적합 비율 보고서	130
로버스트	131
로버스트 적합	131
Cauchy 적합	131
밀도 타원	132
상관 보고서	133
비모수 밀도	133
분위수 밀도 등고선 보고서	134
그룹화 기준	134
적합 메뉴	135
적합 메뉴 옵션	135
진단 그림	138
이변량 플랫폼의 추가 예	138
특수 적합 옵션의 예	139
직교 적합 옵션의 예	141
로버스트 적합 명령의 예	142
밀도 타원을 사용한 그룹화의 예	144
회귀선을 사용한 그룹화의 예	145
기준 변수를 사용한 그룹화의 예	146

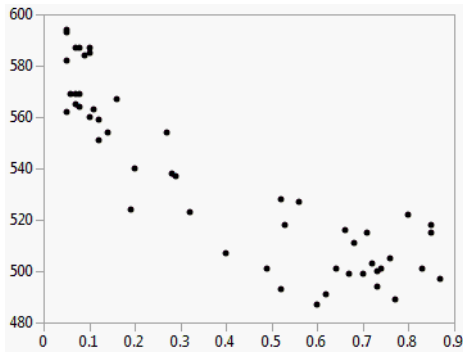
이변량 플랫폼에 대한 통계 상세 정보	147
선형 적합	148
스플라인 적합	148
직교 적합	148
적합 요약 보고서	149
적합 결여 보고서	149
모수 추정값 보고서	150
평활 적합 보고서	150
상관 보고서	151

이변량 분석의 예

이 예에서는 SAT.jmp 샘플 데이터 테이블을 사용합니다. 여기에는 미국 50 개 주와 컬럼비아 특별구에 거주하는 학생의 SAT 시험 점수가 언어 영역과 수학 영역으로 나뉘어 있습니다. 2004 년도 SAT 시험의 학생 응시율과 언어 영역 시험 점수 간에 어떤 관계가 있는지 알아보려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 SAT.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. 2004 Verbal 을 선택하고 **Y, 반응**을 클릭합니다.
4. % Taking (2004) 을 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.

그림 5.2 응시율별 SAT 점수의 예



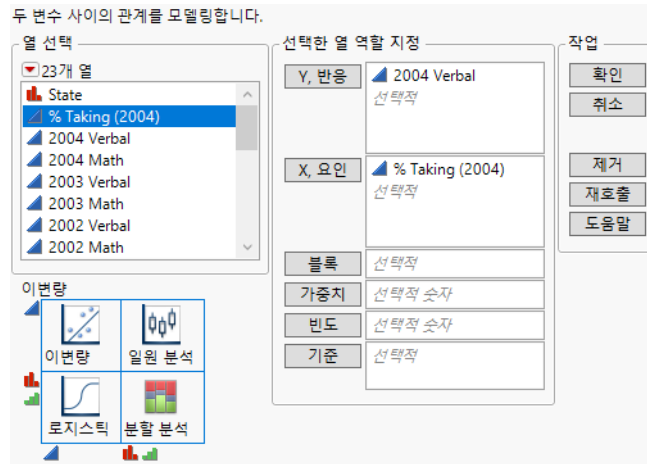
시험에 응시한 모집단의 비율이 낮을수록 언어 영역 점수가 상승했음을 알 수 있습니다.

이변량 플랫폼 시작

이변량 분석을 수행하려면 다음을 수행하십시오.

1. **분석 > X 로 Y 적합**을 선택합니다.
2. **Y, 반응**에 연속형 열을 입력합니다.
3. **X, 요인**에 연속형 열을 입력합니다.

그림 5.3 이변량 시작 창



다이얼로그 위에 "이변량" 이라고 표시되면 이변량 분석을 수행할 것임을 나타냅니다.

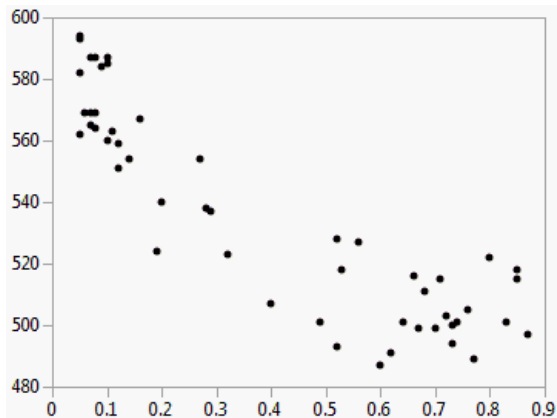
참고 : "JMP 시작하기" 창에서 이변량 분석을 시작할 수도 있습니다. 보기 > **JMP 시작하기** > 기본 분석 > **이변량**을 선택합니다.

이 시작 창에 대한 자세한 내용은 "**X 로 Y 적합 소개**" 장 (105 페이지) 에서 확인하십시오. "열 선택" 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 **JMP 사용**의 "시작하기" 장에서 확인하십시오.

이변량 그림

그림 5.4 에 표시된 것과 같은 그림을 생성하려면 "이변량 분석의 예"(112 페이지)에 설명된 방법을 따르십시오.

그림 5.4 이변량 그림



이변량 보고서는 각 X 및 Y 변수 쌍에 대한 그림으로 시작됩니다. 두 가지 방법 중 하나로 변수를 드래그하여 놓아 그림의 변수를 바꿀 수 있습니다. 한 가지 방법은 한 축의 변수를 다른 축으로 드래그하여 놓아 기존 변수를 바꾸는 것이고, 다른 방법은 관련 데이터 테이블의 "열" 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다.

다른 JMP 그림과 마찬가지로 이 그림에서도 그림 크기를 조정하거나, 화살표 또는 브러시 도구로 점을 강조 표시하거나, 점에 라벨을 지정하는 등의 상호 작용이 가능합니다. 이러한 기능에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

빨간색 삼각형 메뉴 내에 있는 적합 옵션을 사용하여 그림에서 곡선을 적합시키고 통계 보고서 및 추가 메뉴를 볼 수 있습니다. 자세한 내용은 "적합 옵션"(114 페이지)에서 확인하십시오.

적합 옵션

참고: "적합 그룹" 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에 표시됩니다. 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 를 기준으로 정렬할 수 있습니다. 자세한 내용은 Fitting Linear Models 의 "Standard Least Squares Report and Options" 장에서 확인하십시오.

"이변량 적합"의 빨간색 삼각형 메뉴에는 표시 옵션, 적합 옵션 및 제어 옵션이 포함되어 있습니다.

점 표시 산점도에서 점을 숨기거나 표시합니다. 체크 표시는 점이 표시되어 있음을 나타냅니다.

히스토그램 테두리 산점도의 x 및 y 축에 히스토그램을 추가합니다. 체크 표시는 히스토그램 테두리가 설정되어 있음을 나타냅니다. 자세한 내용은 "[히스토그램 테두리](#)"(118 페이지)에서 확인하십시오.

참고: 데이터 테이블의 행에 "숨김" 행 상태만 적용하면 산점도에 해당 점이 표시되지 않습니다. 하지만 히스토그램을 생성하는 데는 숨겨진 행이 사용됩니다. 생성되는 히스토그램과 분석 결과에서 행을 제외하려면 "제외" 행 상태를 적용하고 "이변량 적합"의 빨간색 삼각형 메뉴에서 **다시 실행 > 분석 다시 실행**을 선택하십시오.

요약 통계량 그림과 관련된 상관, 신뢰 구간, 평균, 표준편차 등의 요약 통계량을 표시합니다.

그룹화 기준 분류 (또는 그룹화) 변수를 선택할 수 있습니다. 그러면 그룹화 변수의 각 수준에 대해 개별 분석이 계산되고 산점도에 회귀 곡선 또는 타원이 중첩 표시됩니다. 자세한 내용은 "[그룹화 기준](#)"(134 페이지)에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

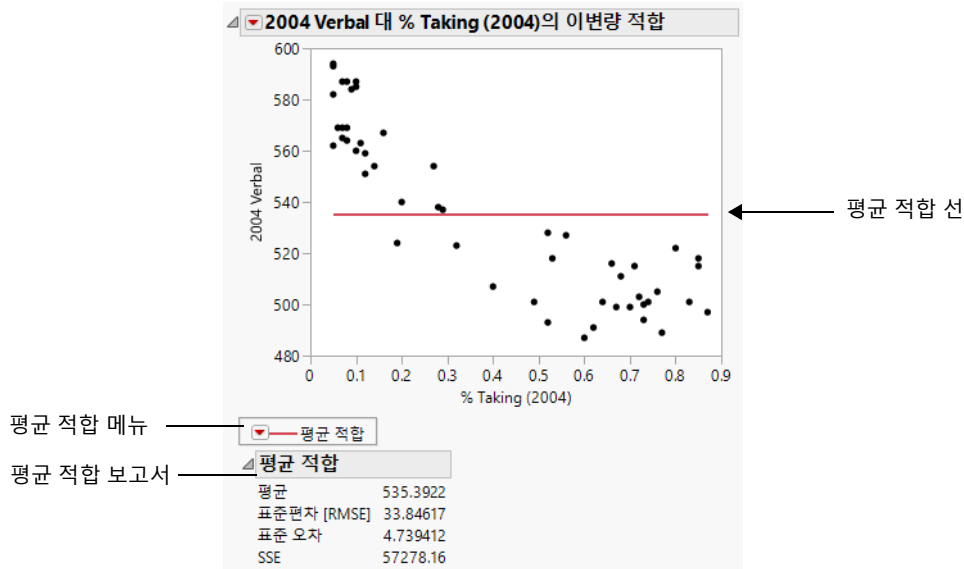
그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

적합 옵션

각 적합 옵션으로 다음을 추가할 수 있습니다.

- 선, 곡선 또는 분포 (산점도)
- 빨간색 삼각형 메뉴 (보고서 창)
- 특정 보고서 (보고서 창)

그림 5.5 평균 적합 옵션의 예



다음과 같은 적합 옵션을 사용할 수 있습니다.

평균 적합 Y 반응 변수의 평균을 나타내는 수평선을 산점도에 추가합니다. 자세한 내용은 "[평균 적합](#)"(119 페이지)에서 확인하십시오.

선형 적합 최소 제곱 회귀를 사용하여 산점도에 적합 직선을 추가합니다. 자세한 내용은 "[선형 적합 및 다항식 적합](#)"(120 페이지)에서 확인하십시오.

다항식 적합 최소 제곱 회귀를 사용하여 특정 차수의 다항식 곡선을 적합시킵니다. 자세한 내용은 "[선형 적합 및 다항식 적합](#)"(120 페이지)에서 확인하십시오.

특수 적합 Y 및 X를 변환합니다. 변환에는 로그, 제곱, 역수 및 지수가 포함됩니다. 중심화 다항식을 해제하거나, 절편 및 기울기를 제약하거나, 다항식 모형을 적합시킬 수도 있습니다. 자세한 내용은 "[특수 적합](#)"(125 페이지)에서 확인하십시오.

유연 추정된 회귀 곡선의 평활도를 제어할 수 있는 옵션을 제공합니다. 자세한 내용은 "[유연](#)"(127 페이지)에서 확인하십시오.

직교 적합 X가 변동하는 것으로 간주되는 경우에 유용한 직교 회귀 적합을 위한 옵션을 제공합니다. 이 옵션의 경우 X 및 Y의 분산에 대한 다양한 가정을 반영하는 하위 옵션이 제공됩니다. 자세한 내용은 "[직교 적합](#)"(130 페이지)에서 확인하십시오.

로버스트 적합 모형에서 데이터 집합에 포함된 이상치의 영향을 줄이는 옵션을 제공합니다. 자세한 내용은 "[로버스트](#)"(131 페이지)에서 확인하십시오.

밀도 타원 X 및 Y 변수에 대한 이변량 정규 분포 적합을 위해 밀도 타원을 그립니다. 자세한 내용은 "[밀도 타원](#)"(132 페이지)에서 확인하십시오.

비모수 밀도 평활된 표면을 기반으로 밀도 등고선을 그립니다. 이 등고선은 데이터 점의 밀도를 설명합니다. 자세한 내용은 "**비모수 밀도**"(133 페이지)에서 확인하십시오.

참고: 적합 제거 옵션을 사용하여 적합을 제거할 수 있습니다. 자세한 내용은 "**적합 메뉴 옵션**"(135 페이지)에서 확인하십시오.

적합 옵션 범주

적합 옵션 범주에는 회귀 적합과 밀도 추정이 포함됩니다.

범주	설명	적합 옵션
회귀 적합	회귀 방법은 관측된 데이터 점에 곡선을 적합시킵니다. 이 적합 방법에는 최소 제곱 적합과 함께 스플라인 적합, 커널 평활, 직교 적합 및 로버스트 적합이 포함됩니다.	평균 적합 선형 적합 다항식 적합 특수 적합 유연 직교 적합 로버스트
밀도 추정	밀도 추정은 점에 이변량 분포를 적합시킵니다. 타원형 등고선이 특징인 이변량 정규 밀도나 일반 비모수 밀도 중 하나를 선택할 수 있습니다.	밀도 타원 비모수 밀도

동일한 적합 옵션을 여러 번 적용

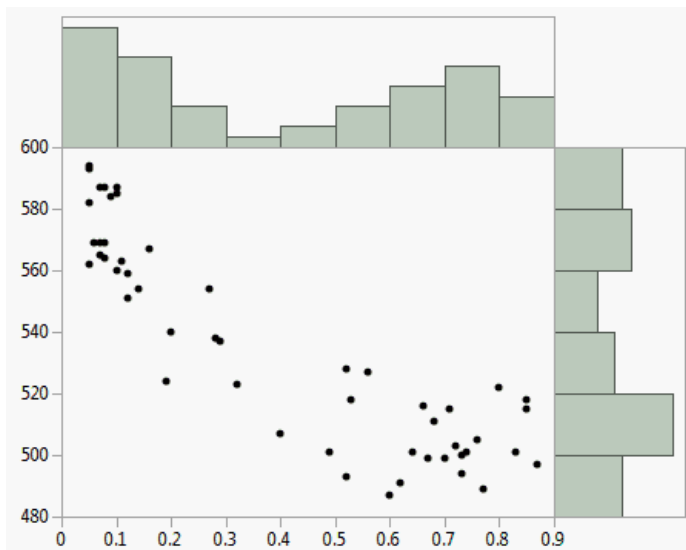
동일한 적합 옵션을 여러 번 선택하여 각각의 새 적합이 산점도에 중첩 표시되도록 할 수 있습니다. 적합을 시도한 후 특정 점을 제외하고 다시 적합시킬 수 있으며, 각 적합을 동일한 산점도에서 비교할 수 있습니다.

보고서 창에서 한 가지 적합 옵션을 여러 분석에 적용하려면 **Ctrl** 키를 누른 채로 적합 옵션을 선택합니다.

히스토그램 테두리

히스토그램 테두리 옵션은 산점도의 x 및 y 축에 히스토그램을 추가합니다. 히스토그램을 사용하여 X 및 Y 변수의 주변 분포를 시각화할 수 있습니다.

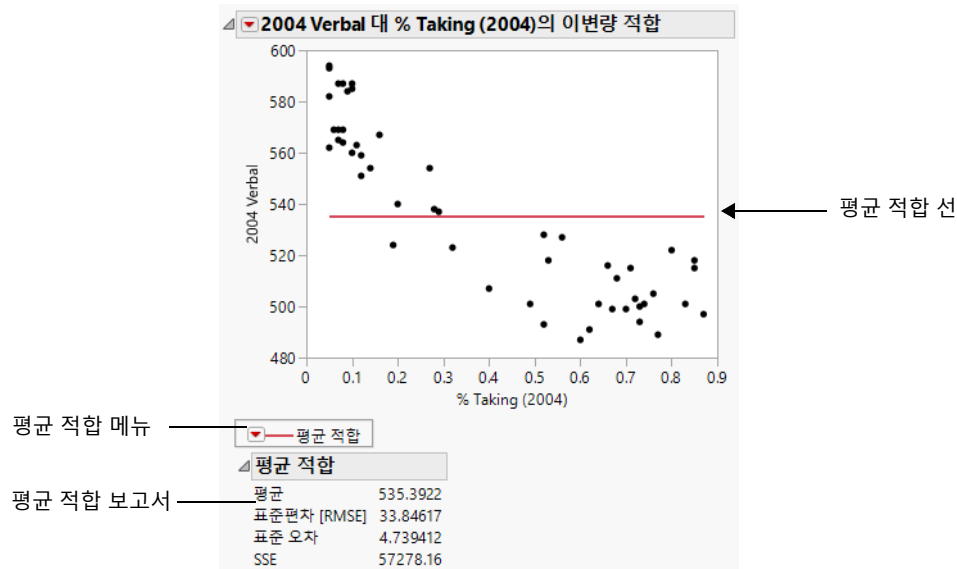
그림 5.6 히스토그램 테두리의 예



평균 적합

평균 적합 옵션을 사용하여 Y 반응 변수의 평균을 나타내는 수평선을 산점도에 추가할 수 있습니다. 먼저 평균을 적합시킨 후 평균 선을 다른 적합 (직선, 신뢰 곡선, 다항식 곡선 등)에 참조로 사용할 수 있습니다.

그림 5.7 평균 적합의 예



평균 적합 보고서

"평균 적합" 보고서에는 평균 적합에 대한 요약 통계량이 표시됩니다.

평균 반응 변수의 평균입니다. 모형에 지정된 효과가 없는 경우에는 예측된 반응입니다.

표준편차 [RMSE] 반응 변수의 표준편차입니다. 평균 제곱 오차의 제곱근으로, RMSE(제곱근 평균 제곱 오차)라고도 합니다.

표준 오차 반응 평균의 표준편차입니다. RMSE를 값 개수의 제곱근으로 나눠서 계산됩니다.

SSE 단순 평균 모형의 오차제곱합입니다. 각 모형 적합에 대한 분산 분석 테이블에서 오차의 제곱합으로 표시됩니다.

"평균 적합" 메뉴의 옵션에 대한 자세한 내용은 "**적합 메뉴**"(135 페이지)에서 확인하십시오.

선형 적합 및 다항식 적합

선형 적합 옵션을 선택하면 최소 제곱 회귀를 사용하여 산점도에 직선 적합을 추가할 수 있습니다. **다항식 적합** 옵션을 선택하면 최소 제곱 회귀를 사용하여 특정 차수의 다항식 곡선을 적합시킬 수 있습니다.

그림 5.8 선형 적합 및 다항식 적합의 예

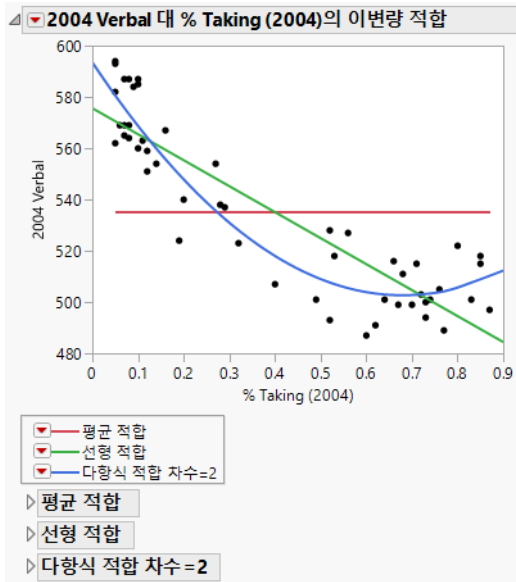


그림 5.8에서는 평균 선에 대한 선형 적합과 2차 다항식 적합을 비교하는 예를 보여 줍니다.

다음 사항에 유의하십시오.

- **선형 적합** 출력은 1차 다항식 적합과 동등합니다.
- **평균 적합** 출력은 0차 다항식 적합과 동등합니다.

"선형 적합"과 "다항식 적합 차수" 메뉴의 옵션에 대한 자세한 내용은 "[적합 메뉴](#)"(135 페이지)에서 확인하십시오. 이 적합에 대한 통계 상세 정보는 "[선형 적합](#)"(148 페이지)에서 확인하십시오.

선형 적합 및 다항식 적합 보고서

"선형 적합" 및 "다항식 적합" 보고서는 적합 방정식으로 시작됩니다.

그림 5.9 적합 방정식의 예

선형 적합
2004 Verbal = 575.62539 - 101.32814*%Taking (2004)
적합 요약
적합 결여
분산 분석
모수 추정값
다항식 적합 차수=2
2004 Verbal = 562.47113 - 111.07875*%Taking (2004) + 196.95725*(%Taking (2004)-0.39706)^2

팁 : 방정식을 클릭하여 편집할 수 있습니다.

각 "선형 적합" 및 "다항식 적합 차수" 보고서에는 세 개 이상의 보고서가 포함됩니다. 네 번째인 "적합 결여" 보고서는 데이터에 X 반복 실험이 있는 경우에만 표시됩니다.

적합 요약 보고서

"적합 요약" 보고서에는 동일한 데이터에 대한 선형 적합 및 2 차 다항식 적합의 반응과 관련된 수치 요약이 표시됩니다. 여러 "적합 요약" 보고서를 비교하여 더 나은 모형을 확인할 수 있습니다. 더 나은 모형일수록 R^2 값이 크고 제공된 평균 제공 오차는 작습니다.

그림 5.10 선형 적합 및 다항식 적합의 적합 요약 보고서

선형 적합 2004 Verbal = 575.62539 - 101.32814*%Taking (2004) 적합 요약 <table> <tr> <td>R^2</td> <td>0.790274</td> </tr> <tr> <td>Adj-R^2</td> <td>0.785994</td> </tr> <tr> <td>제공된 평균 제공 오차</td> <td>15.65752</td> </tr> <tr> <td>반응 평균</td> <td>535.3922</td> </tr> <tr> <td>관측값 수(또는 가중합)</td> <td>51</td> </tr> </table>	R^2	0.790274	Adj- R^2	0.785994	제공된 평균 제공 오차	15.65752	반응 평균	535.3922	관측값 수(또는 가중합)	51	다항식 적합 차수=2 2004 Verbal = 562.47113 - 111.07875*%Taking (2004) + 196.95725*(%Taking (2004)-0.39706)^2 적합 요약 <table> <tr> <td>R^2</td> <td>0.879413</td> </tr> <tr> <td>Adj-R^2</td> <td>0.874389</td> </tr> <tr> <td>제공된 평균 제공 오차</td> <td>11.99566</td> </tr> <tr> <td>반응 평균</td> <td>535.3922</td> </tr> <tr> <td>관측값 수(또는 가중합)</td> <td>51</td> </tr> </table>	R^2	0.879413	Adj- R^2	0.874389	제공된 평균 제공 오차	11.99566	반응 평균	535.3922	관측값 수(또는 가중합)	51
R^2	0.790274																				
Adj- R^2	0.785994																				
제공된 평균 제공 오차	15.65752																				
반응 평균	535.3922																				
관측값 수(또는 가중합)	51																				
R^2	0.879413																				
Adj- R^2	0.874389																				
제공된 평균 제공 오차	11.99566																				
반응 평균	535.3922																				
관측값 수(또는 가중합)	51																				

"적합 요약" 보고서에는 다음 열이 포함됩니다.

R^2 모형에 의해 설명된 변동의 비율을 측정합니다. 나머지 변동은 해당 모형으로 설명되지 않으며 이는 랜덤 오차에 기인합니다. 모형이 완벽하게 적합되는 경우에는 R^2 가 1 입니다.

참고 : R^2 값이 낮으면 설명되지 않은 변동을 설명하는 변수가 모형에 없는 것일 수 있습니다. 하지만 데이터의 내재 변동 범위가 큰 경우에는 유용한 회귀 모형이라도 R^2 값이 낮을 수 있습니다. 일반적인 R^2 값에 대해 알아보려면 연구 영역의 문헌을 읽어 보십시오.

그림 5.10 의 R^2 값은 2 차 다항식 적합이 선형 적합보다 좀더 낫다는 것을 나타냅니다. 자세한 내용은 "적합 요약 보고서"(149 페이지) 에서 확인하십시오.

Adj- R^2 모수의 개수가 서로 다른 모형들을 비교하기 좋도록 계산에 자유도를 사용하여 R^2 값을 조정합니다. 자세한 내용은 "적합 요약 보고서"(149 페이지) 에서 확인하십시오.

제공된 평균 제공 오차 랜덤 오차의 표준편차를 추정합니다. "분산 분석" 보고서에 표시된 오차에 대한 평균 제공의 제공근입니다 (그림 5.12).

반응 평균 반응 변수의 표본 평균 (산술평균) 을 제공합니다. 모형 효과가 지정되지 않은 경우에는 예측된 반응입니다.

관측값 적합을 추정하는 데 사용된 관측값의 개수를 제공합니다. 가중치 변수가 있는 경우에는 가중치의 합계입니다.

적합 결여 보고서

참고 : " 적합 결여 " 보고서는 동일한 x 값을 갖는 행이 여러 개 있는 경우에만 표시됩니다.

" 적합 결여 " 보고서를 사용하면 모형이 적절한지 여부에 관계없이 오차를 추정할 수 있습니다. 이 보고서는 동일한 x 값에서 여러 관측값이 나타나는 경우에 표시됩니다. 이러한 정확 반복 실험에서 측정되는 오차를 **순수 오차**라고 합니다. 이 값은 어떤 형태의 모형을 사용하더라도 설명하거나 예측할 수 없는 표본 오차의 비율입니다. 하지만 반복 실험되는 x 값이 매우 적어서 자유도가 매우 낮은 경우에는 적합 결여 검정이 그다지 유용하지 않을 수 있습니다.

그림 5.11 선형 적합 및 다항식 적합에 대한 적합 요약 보고서의 예

선형 적합				
2004 Verbal = 575.62539 - 101.32814*%Taking (2004)				
▶ 적합 요약				
▶ 적합 결여				
소스	DF	제곱합	평균 제곱	F 비
적합 결여	36	9662.983	268.416	1.4850
순수 오차	13	2349.750	180.750	Prob > F
총 오차	49	12012.733		0.2252
				최대 R ²
				0.9590

다항식 적합 차수=2				
2004 Verbal = 562.47113 - 111.07875*%Taking (2004) + 196.95725*%Taking (2004)^2				
▶ 적합 요약				
▶ 적합 결여				
소스	DF	제곱합	평균 제곱	F 비
적합 결여	35	4557.2473	130.207	0.7204
순수 오차	13	2349.7500	180.750	Prob > F
총 오차	48	6906.9973		0.7862
				최대 R ²
				0.9590

모형의 잔차와 순수 오차 사이의 차이를 **적합 결여 오차**라고 합니다. 회귀변수의 함수 형태가 잘못된 경우에는 적합 결여 오차가 순수 오차보다 매우 클 수 있습니다. 이 경우 다른 유형의 모형 적합을 시도해야 합니다. "적합 결여" 보고서에서는 적합 결여 오차가 0인지 여부를 검정합니다.

" 적합 결여 " 보고서에는 다음 열이 포함됩니다.

소스 변동의 소스로는 적합 결여, 순수 오차 및 총 오차의 세 가지가 있습니다.

DF 각 오차 소스의 DF (자유도) 입니다.

- **총 오차**의 DF 는 분산 분석 테이블의 오차 행 ("분산 분석 보고서" (123 페이지) 참조) 에 나오는 자유도입니다. 이 값은 이 테이블에 나오는 **합계 DF** 와 **모형 DF** 사이의 차이입니다. **오차 DF** 는 적합 결여의 자유도와 순수 오차의 자유도로 분할됩니다.
- **순수 오차 DF** 는 각 효과에 대해 동일한 값을 갖는 여러 행이 있는 경우 각 그룹에서 합동화됩니다. 자세한 내용은 "적합 결여 보고서" (149 페이지) 에서 확인하십시오.
- **적합 결여 DF** 는 **총 오차 DF** 와 **순수 오차 DF** 사이의 차이입니다.

제곱합 각 오차 소스의 제곱합 (줄여서 SS 라고 함) 입니다.

- **총 오차 SS** 는 해당 분산 분석 테이블의 **오차** 행 (" **분산 분석 보고서** "(123 페이지) 참조)에 나오는 제곱합입니다.
- **순수 오차 SS** 는 x 변수에 대해 동일한 값을 갖는 여러 행이 있는 경우 각 그룹에서 합동화됩니다. 이 값은 모형 x 효과로 설명되지 않는 실제 랜덤 오차의 비율을 추정합니다. 자세한 내용은 "**적합 결여 보고서**"(149 페이지)에서 확인하십시오.
- **적합 결여 SS** 는 **총 오차** 제곱합과 **순수 오차** 제곱합 사이의 차이입니다. 적합 결여 SS가 크다면 해당 모형이 데이터에 적절하지 않은 것일 수 있습니다. 아래에 설명된 F 비는 적합 결여로 인한 변동이 순수 오차의 무시 가능한 부분으로서 허용될 만큼 작은지 여부를 결정합니다.

평균 제곱 제곱합을 관련 자유도로 나눈 것입니다. 이 계산에 따라 제곱합이 평균 (평균 제곱)으로 변환됩니다. 통계 검정의 F 비는 평균 제곱의 비율입니다.

F 비 적합 결여의 평균 제곱 대비 순수 오차의 평균 제곱에 대한 비율입니다. 이 값은 적합 결여 오차가 0 이라는 가설을 결정합니다.

Prob > F 적합 결여 분산과 순수 오차 분산으로 인한 변동이 동일한 경우 F 값이 더 클 확률입니다. p 값이 높으면 적합 결여 정도가 크지 않은 것입니다.

최대 R^2 모형에서 모형의 변수만으로 얻을 수 있는 최대 R^2 입니다. 자세한 내용은 "**적합 결여 보고서**"(149 페이지)에서 확인하십시오.

분산 분석 보고서

회귀에 대한 ANOVA(분산 분석)에서는 표본의 총 변동을 여러 성분으로 분할합니다. 이러한 성분은 모형의 유효성을 평가하는 F 비를 계산하는 데 사용됩니다. F 비와 관련된 확률이 작으면 해당 모형은 반응 평균만 사용할 때보다 통계적 적합도가 더 높은 것으로 간주됩니다.

그림 5.12 의 " 분산 분석 " 보고서에서는 선형 적합 (**선형 적합**) 과 2 차 다항식 적합 (**다항식 적합**) 을 비교합니다. 두 적합 모두 평균 가로 선보다 통계적으로 더 낫습니다.

그림 5.12 선형 적합 및 다항식 적합에 대한 분산 분석 보고서의 예

▲ 선형 적합						▲ 다항식 적합 차수=2					
2004 Verbal = 575.62539 - 101.32814*%Taking (2004)						2004 Verbal = 562.47113 - 111.07875*%Taking (2004) + 196.95725*%Taking (2004)-0.39706)^2					
▶ 적합 요약						▶ 적합 요약					
▶ 적합 결여						▶ 적합 결여					
▲ 분산 분석						▲ 분산 분석					
소스	DF	제곱합	평균 제곱	F 비		소스	DF	제곱합	평균 제곱	F 비	
모형	1	45265.424	45265.4	184.6379		모형	2	50371.160	25185.6	175.0265	
오차	49	12012.733	245.2	Prob > F		오차	48	6906.997	143.9	Prob > F	
수정 합계	50	57278.157		<.0001*		수정 합계	50	57278.157		<.0001*	

" 분산 분석 " 보고서에는 다음과 같은 열이 포함됩니다.

소스 변동 소스로, **모형**, **오차** 및 **수정 합계** 중 하나입니다.

DF 각 변동 소스의 DF(자유도)입니다.

- 계산에 사용된 각 모수 추정값에 대해 비결측값의 총 개수 (N) 에서 자유도를 뺍니다. 총 표본 변동의 계산에는 평균 추정값이 사용됩니다. 따라서 합계에서 자유도 1 을 빼서 50 을 유지합니다. 수정된 자유도의 합계는 **모형 및 오차** 항목으로 분할됩니다.
- 선형 적합의 경우 합계 중 자유도 1(**모형** 행에 표시) 이 단일 회귀 모수 (기울기) 를 추정하는 데 사용됩니다. 2 차 다항식 적합의 경우에는 자유도 2 가 모수 (β_1 및 β_2) 를 추정하는 데 사용됩니다.
- **오차** 자유도는 **수정 합계 DF** 와 **모형 DF** 사이의 차이입니다.

제공합 각 변동 소스의 제공합 (줄여서 SS 라고 함) 입니다.

- 이 예에서 각 반응과 표본 평균 사이의 거리에 대한 총 제공합 (**수정 합계**) 은 그림 5.12 에 표시된 것과 같이 57,278.157 입니다. 이 값은 다른 모든 모형과의 비교에 사용되는 기본 모형 (또는 단순 평균 모형) 의 제공합입니다.
- 선형 회귀의 경우 각 점과 적합선 사이의 거리에 대한 제공합이 12,012.733 에서부터 줄어 듭니다. 이 값은 모형 적합 후의 잔차 또는 설명되지 않은 (**오차**) SS 입니다. 2 차 다항식 적합의 잔차 SS 는 6,906.997 로, 선형 적합보다 좀더 많은 변동을 설명합니다. 즉, 모형 SS 는 선형 적합보다 2 차 다항식 적합의 경우에 더 높으므로 이 모형은 더 많은 변동을 설명합니다. **수정 합계 SS** 에서 **오차 SS** 를 뺀 값은 모형에 기인한 제공합입니다.

평균 제공 제공합을 관련 자유도로 나눈 것입니다. 통계 검정의 F 비는 다음과 같은 평균 제공의 비율입니다.

- 선형 적합의 **모형** 평균 제공은 45,265.4 입니다. 이 값은 오차 분산을 추정하지만, 단 모형 모수가 0 이라는 가설에 기반합니다.
- **오차** 평균 제공은 245.2 입니다. 이 값은 오차 분산을 추정합니다.

F 비 모형 평균 제공을 오차 평균 제공으로 나눈 것입니다. 적합의 기본 가설은 절편을 제외한 모든 회귀 모수가 0 이라는 것입니다. 이 가설이 참이면 오차의 평균 제공과 모형의 평균 제공이 모두 오차 분산을 추정하며, 해당 비율은 F 분포를 따릅니다.

Prob > F 지정된 모형이 전체 반응 평균보다 더 잘 적합시키지 않는 경우 F 값이 더 클 것으로 관측된 유의 확률 (p 값) 입니다. 관측된 유의 확률이 0.05 이하이면 회귀 효과의 증거로 간주됩니다.

모수 추정값 보고서

선형 적합에 대한 "모수 추정값" 보고서의 절편 및 단일 x 변수입니다.

k 차 다항식 적합의 경우 모형 절편에 대한 추정값과 X 변수의 k 제곱 각각에 대한 모수 추정값이 있습니다.

그림 5.13 선형 적합 및 다항식 적합에 대한 모수 추정값 보고서의 예

▲ 선형 적합 2004 Verbal = 575.62539 - 101.32814*%Taking (2004) ▶ 적합 요약 ▶ 적합 결여 ▶ 분산 분석 ▲ 모수 추정값				
항	추정값	표준 오차	t 비	Prob> t
절편	575.62539	3.684288	156.24	<.0001*
% Taking (2004)	-101.3281	7.457094	-13.59	<.0001*

▲ 다항식 적합 차수=2 2004 Verbal = 562.47113 - 111.07875*%Taking (2004) + 196.95725*(%Taking (2004)-0.39706)^2 ▶ 적합 요약 ▶ 적합 결여 ▶ 분산 분석 ▲ 모수 추정값				
항	추정값	표준 오차	t 비	Prob> t
절편	562.47113	3.583843	156.95	<.0001*
% Taking (2004)	-111.0788	5.942967	-18.69	<.0001*
(% Taking (2004)-0.39706)^2	196.95725	33.06487	5.96	<.0001*

"모수 추정값" 보고서에는 다음 열이 포함됩니다.

항 요청된 모형의 각 모수 이름을 나열합니다. 절편은 모든 모형에서 상수 항입니다.

추정값 선형 모형의 모수 추정값을 나열합니다. 예측 계산식은 이러한 추정값과 해당 변수 값의 선형 결합입니다.

표준 오차 모수 추정값의 표준 오차에 대한 추정값을 나열합니다. 이러한 값은 검정 및 신뢰 구간을 구성하는 데 사용됩니다.

t 비 각 모수가 0이라는 가설에 대한 검정 통계량을 나열합니다. 이 값은 표준 오차에 대한 모수 추정값의 비율입니다. 가설이 참이면 이 통계량은 스튜던트 t 분포를 따릅니다.

Prob>|t| 각 t 비에서 계산된 관측된 유의 확률을 나열합니다. 이 값은 귀무가설이 참일 때 t 비 (절대값 기준) 가 계산된 값보다 클 확률입니다. 종종 0.05 (일부 경우에는 0.01) 보다 작은 값은 모수가 0 과 유의하게 다르다는 증거로 해석됩니다.

추가 통계량을 표시하려면 보고서에서 마우스 오른쪽 버튼을 클릭하고 **열** 메뉴를 선택합니다. 기본적으로 표시되는 통계량은 다음과 같습니다.

95% 하한 모수 추정값에 대한 95% 신뢰 구간의 하한점입니다.

95% 상한 모수 추정값에 대한 95% 신뢰 구간의 상한점입니다.

표준화 베타 표준화된 모수 추정값입니다. 이 값은 서로 다른 척도로 측정된 X 변수들의 효과를 비교하는 데 유용합니다. 자세한 내용은 "[모수 추정값 보고서](#)"(150 페이지) 에서 확인하십시오.

VIF 분산 팽창 계수입니다.

설계 표준 오차 모수 추정값의 설계 표준 오차입니다. 자세한 내용은 "[모수 추정값 보고서](#)"(150 페이지) 에서 확인하십시오.

특수 적합

참고 : 이 옵션의 예는 "[특수 적합 옵션의 예](#)"(139 페이지) 에서 확인하십시오.

특수 적합 옵션을 사용하여 Y 및 X 를 변환할 수 있습니다. 다양한 이유로 데이터를 변환할 수 있지만, 대개는 데이터가 좀더 충분히 정규 분포를 따르도록 하기 위해 변환을 수행합니다. 이렇게 하면 적절한 검정을 수행할 수 있습니다. 기울기와 절편을 제약하거나, 특정 차수의 다항식을 적합시키거나, 다항식을 중심화할 수도 있습니다.

"변환 또는 제약 조건 지정" 창에는 다음과 같은 옵션이 포함되어 있습니다.

Y 변환 또는 X 변환 Y 또는 X 변수를 변환하려면 다음 옵션 중 하나를 사용합니다.

- 자연 로그
- 제곱근
- 제곱
- 역수
- 지수

차수 지정된 차수의 다항식을 적합시키려면 이 옵션을 사용합니다.

중심화 다항식 다항식 중심화를 해제하려면 **중심화 다항식** 체크박스를 선택 취소합니다 (그림 5.19). X 변수의 변환에 대해서는 다항식 중심화가 수행되지 않습니다. 다항식을 중심화하면 회귀 계수가 안정화되고 다중공선성이 감소합니다.

절편을 다음 값으로 제약 모형 절편을 지정한 값으로 제약하려면 이 체크박스를 선택합니다.

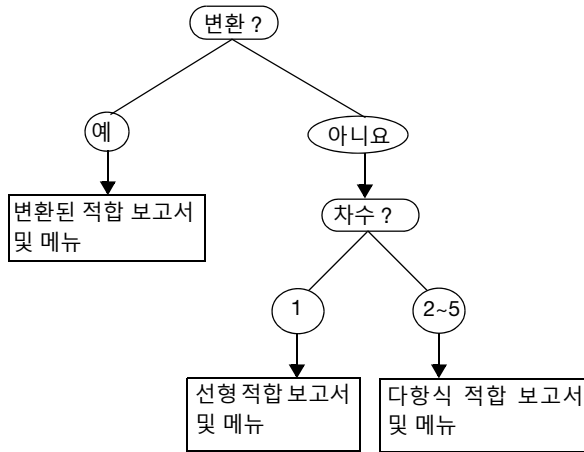
기울기를 다음 값으로 제약 모형 기울기를 지정한 값으로 제약하려면 이 체크박스를 선택합니다.

이 적합의 빨간색 삼각형 옵션에 대한 자세한 내용은 "**적합 메뉴**"(135 페이지)에서 확인하십시오.

특수 적합 보고서 및 메뉴

"특수 적합" 창에서 선택한 항목에 따라 보고서 및 메뉴가 다르게 표시됩니다. 그림 5.14의 순서도에서는 선택 항목에 따라 표시되는 보고서 및 메뉴를 보여 줍니다.

그림 5.14 특수 적합 순서도의 예



변환된 적합 보고서

"변환된 적합" 보고서에는 "선형 적합 및 다항식 적합 보고서"(120 페이지)에 설명된 보고서가 포함됩니다. 하지만 Y 를 변환한 경우에는 "원래 척도에서 적합" 보고서도 표시됩니다. 이 보고서에는 원래 Y 변수를 기준으로 한 적합 측도와 원래 척도로 다시 변환된 적합 모형이 표시됩니다.

유연

추정된 회귀 곡선의 평활도를 제어하려면 "유연" 메뉴의 옵션을 사용합니다.

- "스플라인 적합"에서는 벌점 최소 제공 방법을 사용합니다. 람다 모수를 사용하여 평활도를 조정할 수 있습니다. 자세한 내용은 "스플라인 적합"(127 페이지)에서 확인하십시오.
- "커널 평활기"는 로컬 가중 적합을 기반으로 합니다. 알파 모수를 사용하여 로컬 동작의 영향을 제어할 수 있습니다. 자세한 내용은 "커널 평활기"(128 페이지)에서 확인하십시오.
- "각 값 적합"은 각 X 값에서의 평균 반응을 계산합니다. 자세한 내용은 "각 값 적합"(129 페이지)에서 확인하십시오.

스플라인 적합

스플라인 적합 옵션을 사용하여 람다(λ) 값에 따라 평활도(또는 유연도)가 달라지는 평활 스플라인을 적합시킬 수 있습니다. 람다 값은 스플라인 계산식의 조정 모수입니다. λ 값이 감소하면 스플라인 모형의 오차 항 가중치가 커지며 적합은 더 유연해지고 곡선 형태가 됩니다. λ 값이 증가하면 적합이 덜 유연해져 곡선보다는 직선에 가깝게 됩니다.

다음 사항에 유의하십시오 .

- 적합 스플라인은 X 에 따른 Y 분포의 기대값을 확인하는 데 유용합니다 .
- 적합 곡선의 각 부분에 가장 근접한 점이 해당 부분에 가장 많은 영향을 미칩니다 . λ 값이 낮을수록 이 영향이 커져 매우 유연한 곡선이 생성됩니다 .
- 메뉴에 나열되지 않은 람다 값을 사용하려면 **스플라인 적합 > 기타**를 선택합니다 . X 변수의 척도가 변경되면 적합 모형도 변경됩니다 . 이를 방지하려면 **표준화 X** 옵션을 선택합니다 . 그러면 원래 X 변수나 척도화된 X 변수의 적합 모형이 동일하게 유지됩니다 .
- λ 값을 여러 개 시험해 보는 것이 유용할 수 있습니다 . "평활 스플라인 적합" 보고서 아래의 **람다** 슬라이더를 사용하여 다른 λ 값을 시험해 볼 수 있습니다 . 하지만 λ 는 데이터 척도에 따라 변동됩니다 . 예를 들어 인치 단위로 측정된 X 의 λ 값은 센티미터 단위로 측정된 X 의 λ 값과 같지 않습니다 .

"평활 스플라인 적합" 메뉴의 옵션에 대한 자세한 내용은 "**적합 메뉴**"(135 페이지)에서 확인하십시오 . 이 적합에 대한 통계 상세 정보는 "**스플라인 적합**"(148 페이지)에서 확인하십시오 .

평활 스플라인 적합 보고서

"평활 스플라인 적합" 보고서에는 스플라인 적합의 R^2 와 오차 제곱합이 포함됩니다 . 이러한 값을 사용하여 스플라인 적합을 다른 적합과 비교하거나 여러 스플라인 적합을 서로 비교할 수 있습니다 .

R^2 평활 스플라인 모형으로 설명되는 변동의 비율을 측정합니다 . 자세한 내용은 "**평활 적합 보고서**"(150 페이지)에서 확인하십시오 .

오차 제곱합 각 점에서 적합 스플라인까지의 거리에 대한 제곱합입니다 . 이 값은 스플라인 모형을 적합시킨 후 설명되지 않은 오차 (잔차)입니다 .

람다 변경 숫자를 직접 입력하거나 슬라이더를 이동하는 방법으로 λ 값을 변경할 수 있습니다 .

커널 평활기

커널 평활기 옵션은 영역 내의 표집된 점에서 단순 곡선 (선형 또는 2 차) 의 로컬 가중 적합을 반복적으로 찾아 구성된 곡선을 생성합니다 . 다수의 로컬 적합 (총 128 개) 을 결합하여 전체 영역에 대한 평활 곡선을 생성합니다 . 이 방법을 *Loess* 또는 *Lowess* 라고도 하며 , 이는 *Locally Weighted Scatterplot Smoother*, 즉 로컬 가중 산점도 평활기의 머리글자어입니다 . 자세한 내용은 Cleveland 연구 자료 (1979) 에서 확인하십시오 .

이 방법을 사용하면 변수 간의 관계를 빠르게 확인할 수 있으므로 수행할 분석 또는 적합 유형을 결정하는 데 도움이 됩니다 .

"로컬 평활기" 메뉴의 옵션에 대한 자세한 내용은 "**적합 메뉴**"(135 페이지)에서 확인하십시오 .

로컬 평활기 보고서

"로컬 평활기" 보고서에는 평활기 적합의 R^2 와 오차 제곱합이 포함됩니다. 이러한 값을 사용하여 평활기 적합을 다른 적합과 비교하거나 여러 평활기 적합을 서로 비교할 수 있습니다.

R^2 평활기 모형으로 설명되는 변동의 비율을 측정합니다. 자세한 내용은 "**평활 적합 보고서**"(150 페이지)에서 확인하십시오.

오차 제곱합 각 점에서 적합 평활기까지의 거리에 대한 제곱합입니다. 이 값은 평활기 모형을 적합시킨 후 설명되지 않은 오차 (잔차) 입니다.

로컬 적합 (람다) 각 로컬 적합의 다항식 차수를 선택합니다. 2 차 다항식은 로컬 비평탄 정도를 보다 간편하게 추적할 수 있습니다. 람다는 이 방법으로 적합된 다항식의 차수입니다. 람다는 0, 1 또는 2 일 수 있습니다.

가중치 함수 각 로컬 적합의 근방에 있는 데이터에 가중치를 적용할 방법을 지정합니다. Loess 방법에서는 트라이큐브를 사용합니다. 가중치 함수는 각 x_i 및 y_i 가 선형 적합에 미치는 영향을 결정합니다. 이 영향은 x_i 와 x 의 거리가 멀어질수록 감소하다가 최종적으로 0 이 됩니다.

평활 (알파) 각 로컬 적합의 대상이 되는 점의 수를 제어합니다. 슬라이더를 사용하거나 값을 직접 입력합니다. 알파는 평활 모수로, 임의의 양수일 수 있지만, 일반적인 값은 1/4 ~ 1 입니다. 알파가 증가할수록 곡선이 평활해집니다.

표집 델타 적합 프로세스에 사용되는 표집 양을 제어합니다. 기본 표집 델타는 아무 점도 건너뛰지 않음을 의미하는 0 입니다. 표집 델타를 늘리면 마지막 표본 점을 기준으로 지정된 델타 범위 내의 점들을 적합 프로세스에서 건너뛰게 됩니다. 이 옵션을 사용하여 데이터가 조밀할 때 사용되는 점의 수를 줄일 수 있습니다.

로버스트 적합 곡선에서 멀리 떨어진 점의 영향을 줄이기 위해 점의 가중치를 다시 적용합니다. 이 프로세스를 반복할 횟수 (패스 수) 를 지정합니다. 이 옵션의 목적은 이상치에 작은 가중치를 할당하여 곡선을 수렴시키고 이상치를 자동으로 제외하는 것입니다.

각 값 적합

각 값 적합 옵션은 각각의 고유 X 값에 하나의 값을 적합시킵니다. 적합된 값은 각각의 고유 X 값에 대한 반응 평균입니다.

"각 값 적합" 메뉴의 옵션에 대한 자세한 내용은 "**적합 메뉴**"(135 페이지)에서 확인하십시오.

각 값 적합 보고서

"각 값 적합" 보고서에는 모형 적합에 대한 요약 통계량이 표시됩니다.

관측값 수 관측값의 총 개수입니다.

고유 값 수 고유 X 값의 개수입니다.

자유도 순수 오차 자유도입니다.

제곱합 순수 오차 제곱합입니다.

평균 제곱 순수 오차 평균 제곱입니다.

직교 적합

참고 : 이 옵션의 예는 "**직교 적합 옵션의 예**"(141 페이지)에서 확인하십시오.

직교 적합 옵션은 X 와 Y 모두의 변동성을 설명하는 선형 모델을 적합시킵니다.

직교 적합 옵션

다음 옵션 중 하나를 선택하여 분산 비율을 지정할 수 있습니다.

단변량 분산, 주성분 X 및 Y 의 표본에서 계산된 단변량 분산 추정값을 사용합니다. 이 값은 표준화된 첫 번째 주성분이 됩니다. 측정 시스템 응용 프로그램에서는 오차 분산이 모집단 분산에 비례하지 않을 수 있으므로 이 옵션을 사용하지 않는 것이 좋습니다.

등분산 분산 비율로 1 을 사용합니다. 이 값은 오차 분산이 동일하다고 가정합니다. 등분산을 사용하는 것은 표준화되지 않은 첫 번째 주성분 선을 사용하는 것과 동등합니다. 예를 들어 산점도가 X 및 Y 방향에서 동일하게 척도화된 경우, 정규 밀도 타원을 표시하면 이 선이 타원의 최장축이 됩니다.

X 를 Y 에 적합 분산 비율 0 을 사용합니다. 이 값은 Y 에 사실상 분산이 없음을 나타냅니다.

지정된 분산 비율 비율을 원하는 대로 입력할 수 있으므로 X 의 측정 오차와 Y 의 반응 오차에 대해 알려진 정보를 사용할 수 있습니다.

"직교 적합 비율" 메뉴의 옵션에 대한 자세한 내용은 "**적합 메뉴**"(135 페이지)에서 확인하십시오. 이 적합에 대한 통계 상세 정보는 "**직교 적합**"(148 페이지)에서 확인하십시오.

직교 적합 비율 보고서

"직교 적합 비율" 보고서에는 직교 회귀 모형에 대한 요약 통계량이 표시됩니다.

변수 선을 적합시키는 데 사용되는 변수의 이름입니다.

평균 각 변수의 평균입니다.

표준편차 각 변수의 표준편차입니다.

분산 비율 선을 적합시키는 데 사용되는 분산 비율입니다.

상관 두 변수 간의 상관계수입니다.

절편 적합선의 절편입니다.

기울기 적합선의 기울기입니다.

CL 하한 기울기의 신뢰 하한입니다.

CL 상한 기울기의 신뢰 상한입니다.

알파 신뢰 구간을 계산하는 데 사용되는 유의 수준을 입력합니다.

로버스트

참고 : 이 옵션의 예는 "[로버스트 적합 명령의 예](#)"(142 페이지)에서 확인하십시오.

로버스트 옵션은 데이터 집합에서 이상치의 영향을 줄이기 위한 두 가지 방법을 제공합니다. 이상치는 부정확한 추정 및 결정을 초래할 수 있습니다.

"로버스트 적합" 및 "Cauchy 적합" 메뉴의 옵션에 대한 자세한 내용은 "[적합 메뉴](#)"(135 페이지)에서 확인하십시오.

로버스트 적합

"로버스트 적합" 옵션은 반응 변수에서 이상치의 영향을 줄입니다. Huber M- 추정 방법이 사용됩니다. Huber M- 추정은 다음과 같이 Huber 손실 함수를 최소화하는 모수 추정값을 찾는 방법입니다.

$$l(e) = \sum_i \rho(e_i)$$

다음은 각 요소에 대한 설명입니다.

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & \text{단, } |e| < k \\ k|e| - \frac{1}{2}k^2 & \text{단, } |e| \geq k \end{cases}$$

e_i 는 잔차를 나타냅니다.

Huber 손실 함수는 이상치에 벌점을 부과하며, 작은 오차의 경우에는 2 차 형태로 증가하고 큰 오차의 경우에는 선형적으로 증가합니다. JMP 구현에서는 $k=2$ 입니다. 로버스트 적합에 대한 자세한 내용은 Huber 연구 자료 (1973) 및 Huber and Ronchetti 연구 자료 (2009)에서 확인하십시오.

Cauchy 적합

오차가 Cauchy 분포를 따른다고 가정합니다. Cauchy 분포는 정규 분포보다 꼬리가 더 두꺼우므로 이상치의 영향이 줄어듭니다. 이 옵션은 데이터에 포함된 이상치의 비율이 높은 경우에 유

용할 수 있습니다. 하지만 데이터가 정규 분포에 가깝고 이상치가 일부만 있는 경우에 이 옵션을 사용하면 추론이 잘못될 수 있습니다. Cauchy 옵션은 최대 우도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다.

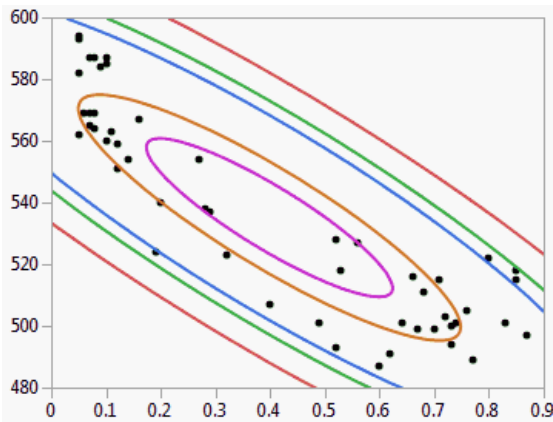
밀도 타원

참고 : 이 옵션의 예는 "[밀도 타원을 사용한 그룹화의 예](#)"(144 페이지)에서 확인하십시오.

밀도 타원 옵션을 사용하면 지정된 점 집합을 포함하는 타원을 그릴 수 있습니다. 점의 개수는 **밀도 타원** 메뉴에서 선택하는 확률에 따라 결정됩니다.

"이변량 정규 타원" 메뉴의 옵션에 대한 자세한 내용은 "[적합 메뉴](#)"(135 페이지)에서 확인하십시오.

그림 5.15 밀도 타원의 예



밀도 타원은 X 및 Y 변수에 대한 이변량 정규 분포 적합을 통해 계산됩니다. 이변량 정규 밀도는 X 및 Y 변수의 평균 및 표준편차와 변수 사이의 상관계수를 사용하는 함수입니다. **기타**를 선택하면 0 보다 크고 1 보다 작거나 같은 범위에서 원하는 확률을 지정할 수 있습니다.

이러한 타원은 밀도 등고선이자 신뢰 곡선입니다. 신뢰 곡선으로 볼 때 밀도 타원은 이변량 정규 분포를 따른다는 가정하에 지정된 백분율의 데이터가 놓일 것으로 기대되는 영역을 보여 줍니다.

밀도 타원은 두 변수 사이의 상관관계를 나타내는 데 유용한 그래프입니다. 두 변수 간의 상관계수가 1 또는 -1 에 가까워질수록 타원이 대각선 방향으로 축소합니다. 두 변수 간에 상관관계가 낮을수록 타원이 원에 가까워지고 대각선 방향의 기울어짐이 덜하게 됩니다.

상관 보고서

각 **밀도 타원** 적합에 대한 상관 보고서에는 X 및 Y 변수의 상관계수가 표시됩니다.

참고: 여러 변수 쌍에 대한 밀도 타원 및 상관계수를 행렬 형식으로 보려면 **분석 > 다변량 방법** 메뉴에서 **다변량** 플랫폼을 사용합니다.

변수 타원을 생성하는 데 사용된 변수의 이름입니다.

평균 X 및 Y 변수 모두의 평균입니다.

표준편차 X 및 Y 변수 모두의 표준편차입니다.

평균과 표준편차에 대한 자세한 내용은 "분포" 장의 "**요약 통계량 보고서**"(41 페이지) 섹션에서 확인하십시오.

상관 Pearson 상관계수입니다. 두 변수가 정확히 선형 관계에 있는 경우 상관계수는 해당 변수 간 관계가 양의 관계인지 음의 관계인지에 따라 1 또는 -1 입니다. 관계가 없으면 상관계수가 0 에 가까워집니다. 자세한 내용은 "**상관 보고서**"(151 페이지)에서 확인하십시오.

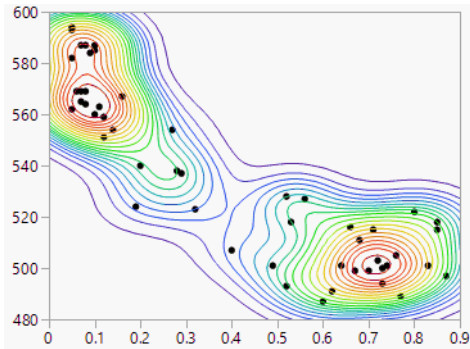
유의 확률 X 변수와 Y 변수 간에 선형 관계가 있지 않은 경우 상관계수의 절대값이 계산된 값보다 더 클 확률입니다.

개수 계산에 사용된 관측값의 개수입니다.

비모수 밀도

그림에 표시되는 점이 수천 개이면 점 집합이 너무 진하게 표시되어 밀도 패턴을 알아보기 어려울 수 있습니다. "비모수 밀도" 옵션을 사용하면 패턴을 확인하기가 쉬워집니다.

"비모수 밀도"는 데이터 점의 밀도를 설명하는 평활 비모수 이변량 표면을 추정합니다. 그림에는 밀도를 보여 주는 일련의 등고선이 추가됩니다(그림 5.16). 이 등고선은 5% 간격의 분위수 등고선입니다. 즉, 추정된 비모수 분포에서 생성된 점 중 약 5%가 최저 등고선 아래에 있고 다음 등고선 아래에 10%가 있는 식입니다. 최고 등고선 아래에는 약 95%의 점이 있습니다.

그림 5.16 비모수 밀도의 예

비모수 밀도 등고선 격자의 크기를 변경하려면 "이변량"의 빨간색 삼각형 메뉴에서 Shift 키를 누른 채로 **비모수 밀도**를 선택합니다. 기본값인 102 보다 큰 값을 입력합니다.

"분위수 밀도 등고선" 메뉴의 옵션에 대한 자세한 내용은 "**적합 메뉴**"(135 페이지)에서 확인하십시오.

분위수 밀도 등고선 보고서

"분위수 밀도 등고선" 보고서에는 비모수 밀도를 생성하는 데 사용된 표준편차가 표시됩니다.

그룹화 기준

참고: 이 옵션의 예는 "**밀도 타원을 사용한 그룹화의 예**"(144 페이지) 및 "**회귀선을 사용한 그룹화의 예**"(145 페이지)에서 확인하십시오.

그룹화 기준 옵션을 사용하면 분류 (그룹화) 변수를 선택할 수 있습니다. 그룹화 변수가 적용되면 이변량 플랫폼에서는 그룹화 변수의 각 수준에 대해 개별 분석을 수행합니다. 산점도에는 각 그룹화 변수에 대한 적합 모형 (선, 곡선, 등고선 또는 타원)이 표시됩니다. 그룹화 변수의 각 수준에 대한 적합은 산점도 아래에서 개별 팝업 메뉴를 사용하여 확인할 수 있습니다.

그룹화 변수가 적용된 경우에는 "이변량 적합"의 빨간색 삼각형 메뉴에서 **그룹화 기준** 옵션이 선택된 상태로 표시됩니다. 그룹화 변수를 변경하려면 먼저 **그룹화 기준** 옵션을 선택하여 기존 변수를 제거 (선택 해제) 합니다. 그런 다음 **그룹화 기준** 옵션을 다시 선택하고 표시되는 창에서 이전과 같이 처리합니다.

그룹화 기준 옵션을 사용하면 그룹화 변수별로 별도의 분석을 생성하되 적합 그림은 동일한 산점도에 표시할 수 있습니다. 이렇게 하면 그룹별 적합을 시각적으로 비교할 수 있습니다.

적합 메뉴

적합 옵션을 선택할 때마다 해당 보고서와 함께 적합 메뉴가 보고서 창에 추가됩니다. 다음 표에서는 각 적합 옵션에 해당하는 적합 메뉴를 보여 줍니다.

적합 옵션	적합 메뉴
평균 적합	평균 적합
선형 적합	선형 적합
다항식 적합	다항식 적합 차수 =X*
특수 적합	선형 적합
	다항식 적합 차수 =X*
	변환된 적합 X*
	제약된 적합
스플라인 적합	평활 스플라인 적합, 람다 =X*
커널 평활기	로컬 평활기
각 값 적합	각 값 적합
직교 적합	직교 적합 비율 =X*
로버스트 적합	로버스트 적합
Cauchy 적합	Cauchy 적합
밀도 타원	이변량 정규 타원 P=X*
비모수 밀도	분위수 밀도 등고선

*X= 변수 (문자 또는 숫자)

적합 메뉴 옵션

적합 메뉴 옵션은 선택한 적합에 따라 다릅니다.

- " 대부분의 적합에 적용되는 옵션 "(136 페이지)
- " 다중 적합에 적용되는 옵션 "(136 페이지)
- " 이변량 정규 타원에 적용되는 옵션 "(137 페이지)
- " 분위수 밀도 등고선에 적용되는 옵션 "(137 페이지)

대부분의 적합에 적용되는 옵션

적합선 모형 적합을 설명하는 선 또는 곡선을 표시하거나 숨깁니다. "이변량 정규 타원" 보고서의 경우 이 옵션은 등고선 테두리를 나타내는 타원을 표시하거나 숨깁니다. "분위수 밀도 등고선"의 경우에는 사용할 수 없습니다.

선 색상 색상 팔레트에서 각 적합에 할당할 색상을 선택할 수 있습니다. "분위수 밀도 등고선"의 경우에는 사용할 수 없습니다.

선 스타일 선 스타일 팔레트에서 각 적합의 선 스타일을 선택할 수 있습니다. "분위수 밀도 등고선"의 경우에는 사용할 수 없습니다.

선 너비 적합선의 선 너비를 변경할 수 있습니다. 기본 선 너비는 가장 가는 선입니다. "분위수 밀도 등고선"의 경우에는 사용할 수 없습니다.

보고서 적합 보고서 표시를 설정하거나 해제합니다. 이변량 그림은 수정하지 않습니다.

적합 제거 그래프에서 적합을 제거하고 해당 보고서도 제거합니다.

다중 적합에 적용되는 옵션

적합 신뢰 곡선 기대값 (평균)에 대한 신뢰 한계를 표시하거나 숨깁니다. 이 옵션은 "스플라인 적합", "밀도 타원", "각 값 적합" 및 "직교 적합"의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다.

개별값 신뢰 곡선 개별 예측값의 신뢰 한계를 표시하거나 숨깁니다. 신뢰 한계는 오차의 변동과 모수 추정값의 변동을 반영합니다. 이 옵션은 "평균 적합", "스플라인 적합", "밀도 타원", "각 값 적합" 및 "직교 적합"의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다.

예측값 저장 현재 데이터 테이블에 예측 < 열 이름 > 이라는 새 열을 생성합니다. 여기서 열 이름은 Y 변수의 이름입니다. 이 열에는 예측 계산식과 계산된 표본 예측값이 포함됩니다. 예측 계산식은 사용자가 테이블에 추가하는 행에 대해 값을 자동으로 계산합니다. 이 옵션은 "각 값 적합" 및 "밀도 타원"의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다.

각 적합에 **예측값 저장** 및 **잔차 저장** 옵션을 사용할 수 있습니다. 이러한 옵션을 여러 번 사용하거나 그룹화 변수와 함께 사용하는 경우에는 데이터 테이블에 생성되는 결과 열의 이름을 각 적합에 맞게 변경하는 것이 좋습니다.

잔차 저장 현재 데이터 테이블에 잔차 < 열 이름 > 이라는 새 열을 생성합니다. 여기서 열 이름은 Y 변수의 이름입니다. 각 값은 실제값(관측값)과 예측값 사이의 차이입니다. **예측값 저장** 옵션과 달리 이 옵션은 새 열에 계산식을 생성하지 않습니다. 이 옵션은 "각 값 적합" 및 "밀도 타원"의 경우에는 사용할 수 없으므로 메뉴에서 흐리게 표시됩니다.

각 적합에 **예측값 저장** 및 **잔차 저장** 옵션을 사용할 수 있습니다. 이러한 옵션을 여러 번 사용하거나 그룹화 변수와 함께 사용하는 경우에는 데이터 테이블에 생성되는 결과 열의 이름을 각 적합에 맞게 변경하는 것이 좋습니다.

스튜던트화 잔차 저장 잔차를 잔차의 표준 오차로 나눈 결과를 포함하는 새 열을 데이터 테이블에 생성합니다.

평균 신뢰 한계 계산식 평균 신뢰 구간의 계산식을 포함하는 새 열을 데이터 테이블에 생성합니다.

개별값 신뢰 한계 계산식 개별 신뢰 구간의 계산식을 포함하는 새 열을 데이터 테이블에 생성합니다.

잔차 그림 (선형, 다항식 및 특수 적합만 해당) 5 가지 진단 그림, 즉 잔차 대 예측값 그림, 실제값 대 예측값 그림, 잔차 대 행 번호 그림, 잔차 대 X 그림, 잔차 정규 분위수 그림을 생성합니다. 자세한 내용은 "**진단 그림**"(138 페이지)에서 확인하십시오.

α 수준 설정 다양한 적합의 신뢰 대역을 계산하는 데 사용되는 유의 수준을 설정할 수 있습니다.

적합 신뢰 구간 음영 적합 신뢰 곡선 옵션과 동일한 곡선을 그리되 곡선 사이의 영역에 음영을 적용합니다.

개별값 신뢰 구간 음영 개별값 신뢰 곡선 옵션과 동일한 곡선을 그리되 곡선 사이의 영역에 음영을 적용합니다.

계수 저장 스플라인 계수를 X, A, B, C 및 D 라는 열을 포함하는 새 데이터 테이블로 저장합니다. X 열에는 매듭 점이 포함됩니다. A, B, C 및 D 는 절편과 3 차 다항식의 1 차, 2 차 및 3 차 계수입니다. 이러한 계수는 X 열의 해당 값에서 다음으로 높은 값까지 확장됩니다.

이변량 정규 타원에 적용되는 옵션

음영 등고선 밀도 타원의 내부 영역에 음영을 적용합니다.

내부 점 선택 타원 내부의 점을 선택합니다.

외부 점 선택 타원 외부의 점을 선택합니다.

분위수 밀도 등고선에 적용되는 옵션

커널 제어 각 변수에 대한 슬라이더를 표시합니다. 슬라이더로 등고선 밀도를 결정하기 위한 X 및 Y 값 범위를 정의하는 표준편차를 변경할 수 있습니다.

5% 등고선 5% 등고선을 표시하거나 숨깁니다.

등고선 10% 등고선을 표시하거나 숨깁니다.

등고선 채우기 등고선 사이의 영역을 채웁니다.

색상 테마 등고선의 색상 테마를 변경합니다.

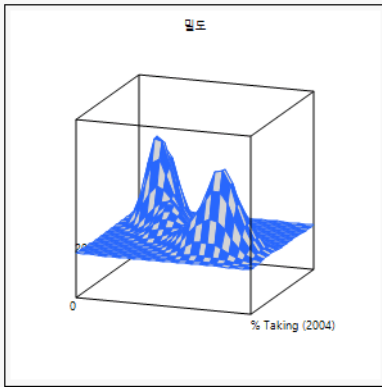
밀도별 점 선택 사용자가 지정한 분위수 범위에 있는 점을 선택합니다.

밀도 분위수별 색상 밀도에 따라 다른 점 색상을 적용합니다.

밀도 분위수 저장 각 점의 밀도 분위수를 포함하는 새 열을 생성합니다.

그물 그림 두 분석 변수의 격자 위에 표시되는 3 차원 밀도 그림입니다.

그림 5.17 그물 그림의 예



최빈 균집화 현재 데이터 테이블에 새 열을 생성하고 해당 열에 균집 값을 채웁니다.

참고 : 최빈 균집화 값을 먼저 저장한 다음 밀도 격자를 저장하면 격자 테이블에 균집 값도 포함됩니다. 균집 값은 그림의 점에 색상 및 표식을 적용하는 데 유용합니다.

밀도 격자 저장 밀도 추정값과 관련 분위수를 새 데이터 테이블에 저장합니다. 이 격자 데이터를 사용하여 밀도를 3D 산점도나 등고선 그림 플랫폼에서처럼 다른 방식으로 시각화할 수 있습니다.

진단 그림

잔차 그림 옵션은 잔차 그림과 모형 적합을 진단하기 위한 다른 그림을 생성합니다. 다음과 같은 그림을 사용할 수 있습니다.

잔차 대 예측값 그림 잔차 대 예측값의 그림입니다. 잔차 히스토그램도 생성됩니다.

실제값 대 예측값 그림 실제값 대 예측값의 그림입니다.

잔차 대 행 번호 그림 잔차 값 대 행 번호의 그림입니다.

잔차 대 X 그림 잔차 값 대 X 변수의 그림입니다.

잔차 정규 분위수 그림 잔차의 정규 분위수 그림입니다.

이변량 플랫폼의 추가 예

- "특수 적합 옵션의 예"
- "직교 적합 옵션의 예"
- "로버스트 적합 명령의 예"
- "밀도 타원을 사용한 그룹화의 예"

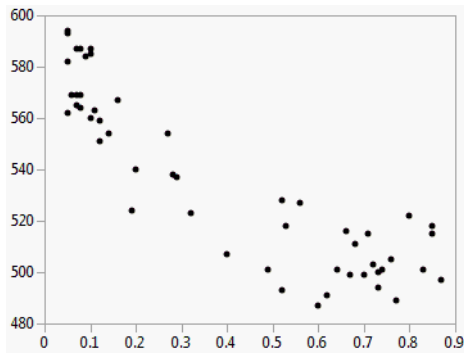
- "회귀선을 사용한 그룹화의 예"
- "기준 변수를 사용한 그룹화의 예"

특수 적합 옵션의 예

Y 를 로그로 변환하고 X 를 제곱근으로 변환하려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 SAT.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. 2004 Verbal 을 선택하고 **Y, 반응**을 클릭합니다 .
4. % Taking (2004) 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

그림 5.18 응시율별 SAT 점수의 예



6. "이변량 적합" 의 빨간색 삼각형을 클릭하고 **특수 적합**을 선택합니다 ." 변환 또는 제약 조건 지정 " 창이 나타납니다 . 이 창에 대한 자세한 내용은 "특수 적합"(125 페이지) 에서 확인하십시오 .

그림 5.19 변환 또는 제약 조건 지정 창

7. "Y 변환" 에서 **자연 로그 : $\log(y)$** 를 선택합니다 .
8. "X 변환" 에서 **제곱근 : \sqrt{x}** 를 선택합니다 .

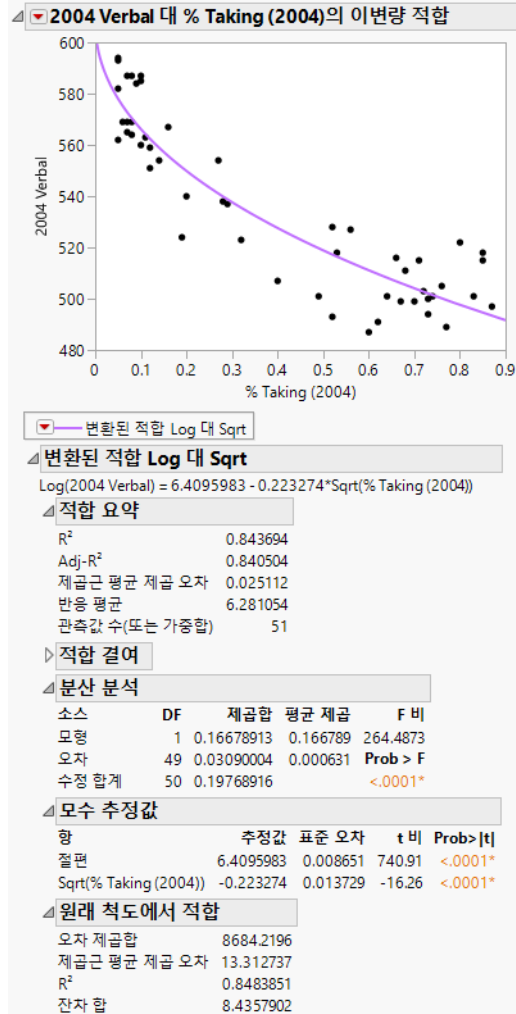
9. **확인**을 클릭합니다.**그림 5.20** 특수 적합 보고서의 예

그림 5.20에서는 원래 척도로 표시된 적합선을 보여 줍니다. 표시된 선은 점들이 몰려 있는 곳을 통과하므로 이 모형은 데이터를 잘 적합시키는 것으로 보입니다.

직교 적합 옵션의 예

이 예는 두 부분으로 구성되어 있습니다. 먼저, 분포 플랫폼을 사용하여 변수를 표준화합니다. 그런 다음 표준화된 변수를 사용하여 직교 모형을 적합시킵니다.

변수 표준화

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > 분포**를 선택합니다.
3. height 및 weight 를 선택하고 **Y, 열**을 클릭합니다.
4. **확인**을 클릭합니다.
5. Ctrl 키를 누른 채로 있습니다. height 의 빨간색 삼각형을 클릭하고 **저장 > 표준화**를 선택합니다.

Ctrl 키를 누른 채로 작업을 수행하면 해당 작업이 보고서 창의 모든 변수에 일괄 적용됩니다. Big Class.jmp 샘플 데이터 테이블에는 두 개의 새 열이 추가되었습니다.

6. "분포" 보고서 창을 닫습니다.

표준화된 변수를 사용하여 직교 모형 적합

1. Big Class.jmp 샘플 데이터 테이블에서 **분석 > X 로 Y 적합**을 선택합니다.
2. 표준화 weight 를 선택하고 **Y, 반응**을 클릭합니다.
3. 표준화 height 를 선택하고 **X, 요인**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "표준화 weight 대 표준화 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.
6. "표준화 weight 대 표준화 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **직교 적합**을 선택합니다. 그런 후 다음을 각각 선택합니다.
 - 등분산
 - X 를 Y 에 적합
 - 지정된 분산 비율 (0.2 입력)
 - 지정된 분산 비율 (5 입력)

그림 5.21 직교 적합 옵션의 예

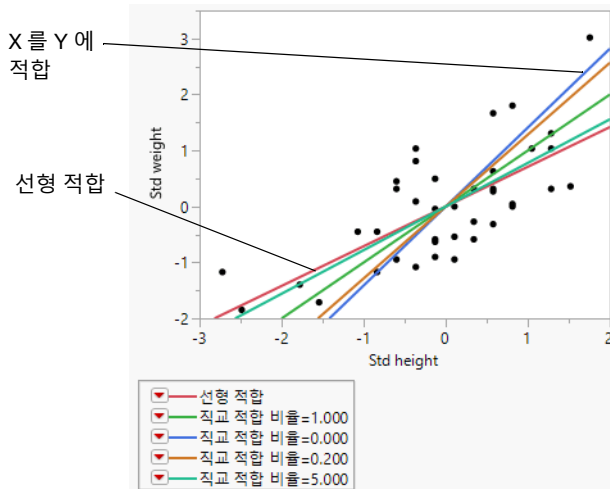


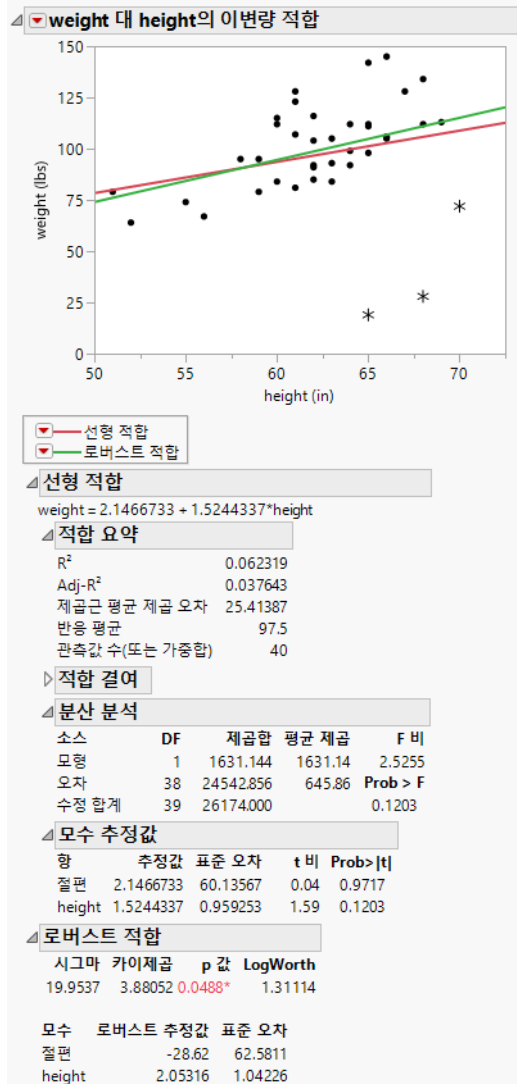
그림 5.21의 산점도에서는 표준화된 height 및 weight 값과 함께 선택한 직교 분산 비율의 동작을 나타내는 다양한 선형 적합을 보여 줍니다. 표준 선형 회귀 (선형 적합)는 X 변수의 분산이 매우 작은 것으로 간주될 때 발생합니다. X를 Y에 적합은 반대로 Y 변수의 분산이 무시될 때의 극단 적합입니다. 다른 모든 선은 이 두 극단 사이에 존재하며 분산 비율이 변경되면 선이 이동합니다. 분산 비율이 높을수록 Y 반응의 변동이 우월하게 되고 적합선의 기울기가 X로 Y 적합에 가까워집니다. 마찬가지로, 비율을 낮추면 선의 기울기가 Y로 X 적합에 가까워집니다.

로버스트 적합 명령의 예

Weight Measurements.jmp 샘플 데이터 테이블의 데이터는 40 명의 학생을 대상으로 한 키 및 체중 측정값을 보여 줍니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Weight Measurements.jmp 를 엽니다.
2. **분석 > X로 Y 적합**을 선택합니다.
3. weight 를 선택하고 **Y, 반응**을 클릭합니다.
4. height 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "weight 대 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.
7. "weight 대 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **로버스트 > 로버스트 적합**을 선택합니다.

그림 5.22 로버스트 적합의 예



표준 분산 분석 보고서를 보면 p 값이 0.1203 이므로 키와 체중이 선형 관계에 있지 않다고 잘못된 결론을 내릴 수 있습니다. 하지만 "로버스트 적합" 보고서를 보면 여기서는 p 값이 0.0489 이므로 키와 체중이 선형 관계에 있다는 결론을 내릴 수 있습니다. 일부 측정값은 비정상적으로 낮은데 이는 사용자 입력이 잘못된 데 기인한 것일 수 있습니다. 이러한 측정값은 분석에 큰 영향을 주지 않았습니다.

밀도 타원을 사용한 그룹화의 예

이 예에서는 Hot Dogs.jmp 샘플 데이터 테이블을 사용합니다. Type 열은 핫도그의 세 가지 종류 (소고기, 돼지고기 또는 가금육) 를 나타냅니다. 이 세 종류의 핫도그를 비용 변수에 따라 그룹화하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hot Dogs.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. **\$/oz** 를 선택하고 **Y, 반응**을 클릭합니다.
4. **\$/lb Protein** 을 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "\$/oz 대 \$/lb Protein 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **그룹화 기준**을 선택합니다.
7. 목록에서 **Type** 을 선택합니다.
8. **확인**을 클릭합니다.
그룹화 기준 옵션을 다시 보면 그 옆에 체크 표시가 표시되어 있습니다.
9. "\$/oz 대 \$/lb Protein 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **밀도 타원 > 0.90** 을 선택합니다.

Type 에 따라 다른 점 색상을 적용하려면 다음 단계를 수행하십시오.

10. 산점도를 마우스 오른쪽 버튼으로 클릭하고 **행 범례**를 선택합니다.
11. 열 목록에서 **Type** 을 선택하고 **확인**을 클릭합니다.

그림 5.23 그룹화의 예

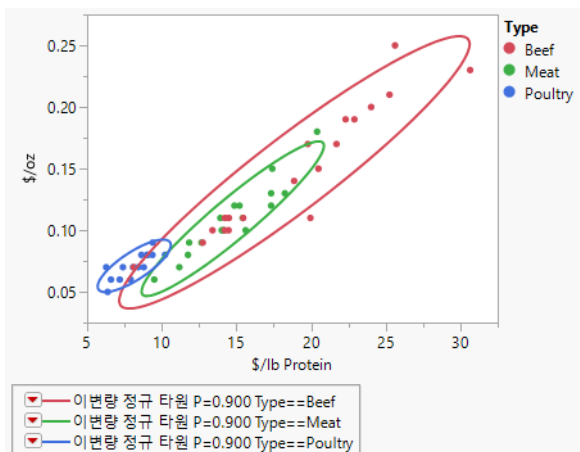


그림 5.23 의 타원은 서로 다른 종류의 핫도그들이 비용 변수에 따라 어떻게 군집화되는지를 명확히 보여 줍니다.

회귀선을 사용한 그룹화의 예

그룹화 변수를 사용하여 회귀선을 중첩 표시하면 서로 다른 그룹의 기울기를 비교할 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. weight 를 선택하고 **Y, 반응**을 클릭합니다.
4. height 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.

그림 5.24 의 왼쪽 예와 같은 그림을 생성하려면 다음을 수행하십시오.

6. "weight 대 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.
- 그림 5.24 의 오른쪽 예와 같은 그림을 생성하려면 다음을 수행하십시오.
7. "선형 적합" 메뉴에서 **적합 제거**를 선택합니다.
8. "weight 대 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **그룹화 기준**을 선택합니다.
9. 목록에서 **sex** 를 선택합니다.
10. **확인**을 클릭합니다.
11. "weight 대 height 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.

그림 5.24 전체 표본 및 그룹화된 표본에 대한 회귀 분석의 예

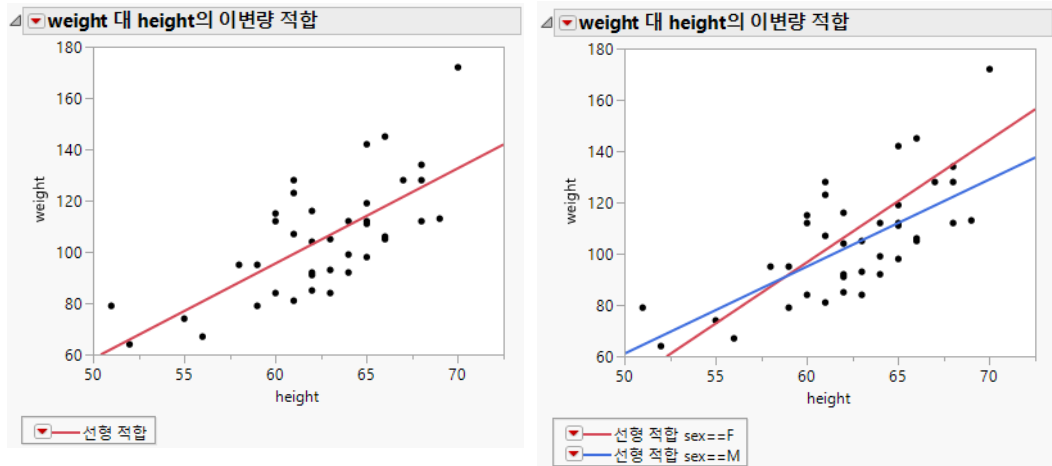


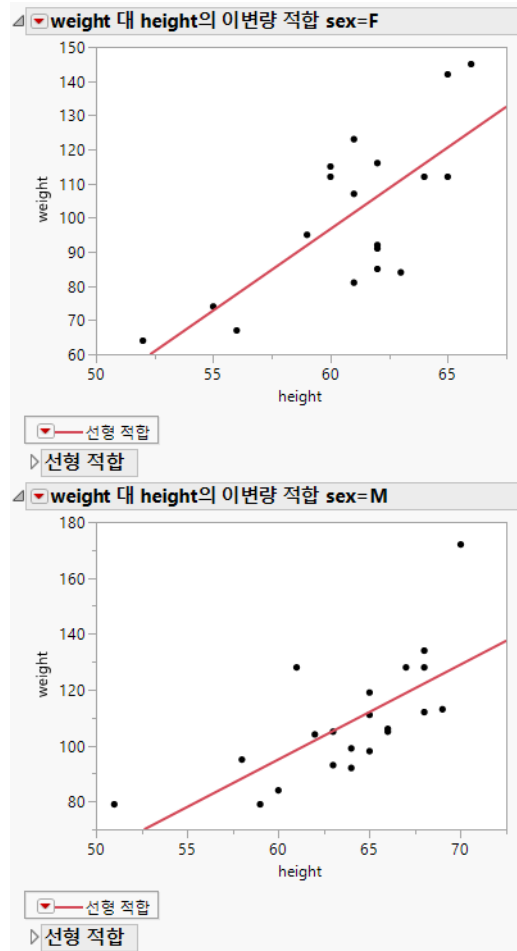
그림 5.24 의 왼쪽 산점도에는 체중과 키의 상관관계를 나타내는 단일 회귀선이 있습니다. 오른쪽 산점도에서는 남학생과 여학생의 회귀선을 개별적으로 보여 줍니다.

기준 변수를 사용한 그룹화의 예

또 다른 그룹화 방법은 시작 창에서 기준 변수를 지정하는 것입니다. 그러면 기준 변수 (또는 여러 기준 변수의 조합) 의 각 수준별로 별도의 보고서와 그래프가 표시됩니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. weight 를 선택하고 **Y, 반응**을 클릭합니다.
4. height 를 선택하고 **X, 요인**을 클릭합니다.
5. sex 를 선택하고 **기준**을 클릭합니다.
6. **확인**을 클릭합니다.
7. Ctrl 키를 누른 채로 "weight 대 height 의 이변량 적합 sex =F" 옆의 빨간색 삼각형 메뉴를 클릭하고 빨간색 삼각형 메뉴에서 **선형 적합**을 선택합니다.

그림 5.25 기준 변수 그림의 예



기준 변수 (sex) 의 각 수준에 대해 개별 분석이 표시됩니다 . 따라서 여학생에 대한 산점도와 남 학생에 대한 산점도가 표시됩니다 .

이변량 플랫폼에 대한 통계 상세 정보

- " 선형 적합 "
- " 스플라인 적합 "
- " 직교 적합 "
- " 적합 요약 보고서 "

- " 적합 결여 보고서 "
- " 모수 추정값 보고서 "
- " 평활 적합 보고서 "
- " 상관 보고서 "

선형 적합

선형 적합 옵션은 잔차 제곱합을 최소화하도록 점을 적합시키는 직선의 모수 β_0 및 β_1 을 찾습니다. i 번째 행에 대한 모형은 $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ 로 기록됩니다.

2 차 다항식은 포물선으로 나타나고 3 차 다항식은 3 차 곡선으로 나타납니다. k 차의 경우 i 번째 관측값에 대한 모형은 다음과 같습니다.

$$y_i = \sum_{j=0}^k \beta_j x_i^j + \epsilon_i$$

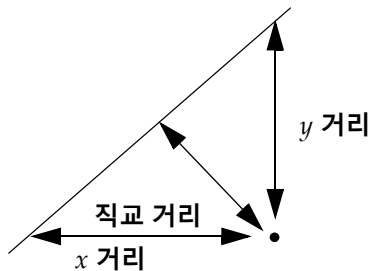
스플라인 적합

3 차 스플라인 방법에서는 결과 곡선이 연속적이고 부드럽게 되도록 매듭지어진 일련의 3 차 다항식이 사용됩니다. 추정은 오차 제곱합과 곡선 범위에 통합된 곡률에 대한 벌점의 조합인 목적 함수를 최소화하는 방법으로 수행됩니다. 이 방법에 대한 자세한 내용은 Reinsch 연구 자료 (1967) 또는 Eubank 학술 자료 (1999) 에서 확인하십시오.

직교 적합

표준 최소 제곱 적합 방법에서는 X 변수가 고정되어 있고 Y 변수는 X + 오차에 따라 달라진다고 가정합니다. X 측정값에 랜덤 변동이 있으면 수직 거리 차의 제곱합을 최소화하는 선을 적합시켜야 합니다 (그림 5.26). 하지만 수직 거리는 X 및 Y 의 척도화 방법에 따라 달라지며, 수직 거리의 척도화는 그래픽 문제가 아니라 통계적 문제로 남아 있습니다.

그림 5.26 적합선에 수직인 선



적합을 수행하려면 X 의 오차 분산에 대한 Y 의 오차 분산의 비율을 지정해야 합니다. 이는 표본 점의 분산이 아니라 오차의 분산이므로 신중하게 선택해야 합니다. 표준 최소 제곱에서는 σ_x^2 이

0 이기 때문에 $(\sigma_y^2)/(\sigma_x^2)$ 비율이 무한대가 됩니다. 오차 비율이 큰 경우에 직교 적합을 수행하면 적합선이 표준 최소 제곱 적합선에 가까워집니다. 비율을 0 으로 지정하면 X 에 대한 Y 가 아니라 Y 에 대한 X 의 회귀와 동등한 적합이 수행됩니다.

이 방법의 가장 일반적인 용도는 동일한 값을 측정하는 데 오차가 있는 두 측정 시스템을 비교하는 것입니다. 따라서 Y 반응 오차와 X 측정 오차가 모두 동일한 유형의 측정 오차입니다. 측정 오차 분산은 어디에서 구할까요? 이변량 데이터에서는 어느 측정 시스템에서 어떤 비율의 오차가 나타나는지 알 수 없으므로 이를 구할 수 없습니다. 따라서 1 과 같은 비율을 임의로 가정하거나, 두 측정 시스템에 의한 동일한 단위의 개별 반복 측정에 의존해야 합니다.

이 방법을 사용하면 Y 와 X 모두의 예측값을 계산할 수 있다는 장점이 있습니다. 예측값은 데이터 점과 가장 가까운 적합선 상의 점이며, 가깝다는 것은 분산 비율을 기준으로 합니다.

신뢰 한계는 Tan and Iglewicz 연구 자료 (1999) 에 설명된 것과 같이 계산됩니다.

적합 요약 보고서

R²

해당 분산 분석 테이블에서 구한 통계량을 사용할 경우 모든 연속형 반응 적합의 R² 는 다음과 같이 계산됩니다.

$$\frac{\text{모형의 제곱합}}{\text{Sum of Squares for C. Total}}$$

Adj-R²

Adj-R² 는 제곱합 대신 사용되는 평균 제곱의 비율로, 다음과 같이 계산됩니다.

$$1 - \frac{\text{오차의 평균 제곱}}{\text{Mean Square for C. Total}}$$

오차의 평균 제곱은 "분산 분석" 보고서에 표시됩니다 (그림 5.12). 수정 합계의 제곱합은 수정 합계의 제곱합을 해당 자유도로 나눠서 계산할 수 있습니다.

적합 결여 보고서

순수 오차 DF

순수 오차 DF 의 경우 키 값이 동일한 관측값이 둘 이상 있는 경우를 고려합니다. 일반적으로 각 효과의 값이 동일한 행이 여러 개 있는 g 개의 그룹이 있는 경우 합동 DF(DF_p)는 다음과 같습니다.

$$DF_p = \sum_{i=1}^g (n_i - 1)$$

여기서 n_i 는 i 번째 그룹에 있는 관측값의 개수입니다.

순수 오차 SS

순수 오차 SS 의 경우, 일반적으로 x 값이 동일한 행이 여러 개 있는 g 개의 그룹이 있다면 합동 $SS(SS_p)$ 는 다음과 같이 기록됩니다.

$$SS_p = \sum_{i=1}^g SS_i$$

여기서 SS_i 는 평균이 수정된 i 번째 그룹의 제곱합입니다.

최대 R^2

순수 오차는 모형 형태에 따라 변동되지 않으며 가능한 최소 분산이므로 최대 R^2 는 다음과 같이 계산됩니다.

$$1 - \frac{SS(\text{순수 오차})}{SS(\text{Total for whole model})}$$

모수 추정값 보고서

표준화 베타

표준화 베타는 다음과 같이 계산됩니다.

$$\hat{\beta}(s_x/s_y)$$

여기서 $\hat{\beta}$ 는 추정된 모수이고, s_x 및 s_y 는 X 및 Y 변수의 표준편차입니다.

설계 표준 오차

설계 표준 오차는 모수 추정값의 표준 오차를 RMSE 로 나눠서 계산됩니다.

평활 적합 보고서

R^2 는 $1 - (SSE / \text{수정 합계 SS})$ 와 같습니다. 여기서 수정 합계 SS 는 선형 적합 ANOVA 보고서에서 확인할 수 있습니다.

상관 보고서

Pearson 상관계수는 r 로 표기되며 다음과 같이 계산됩니다 .

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} \quad \text{여기서,} \quad s_{xy} = \frac{\sum w_i (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{df}$$

여기서 w_i 는 i 번째 관측값의 체중 (체중 열이 지정된 경우) 또는 1(체중 열이 할당되지 않은 경우) 입니다 .

6 장

일원 분석

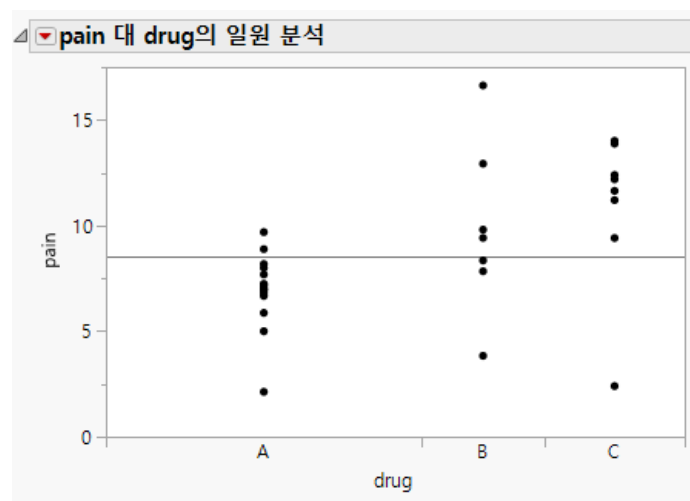
연속형 Y 와 범주형 X 변수 간의 관계 분석

일원 분석 또는 X 로 Y 적합 플랫폼을 사용하면 범주형 단일 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지 살펴볼 수 있습니다. 예를 들어 동일한 약품 유형 (X) 의 서로 다른 범주가 번호 척도로 측정된 환자의 통증 수준 (Y) 에 어떤 영향을 미치는지 확인할 수 있습니다.

일원 분석 플랫폼은 X 로 Y 적합 플랫폼의 연속형 대 명목형 또는 순서형 분석법입니다. 분석 결과는 그림에 표시되며, 다음과 같은 분석을 대화식으로 추가할 수 있습니다.

- 평균을 적합시키고 평균이 동일한지 검정하기 위한 일원 분산 분석
- 비모수 검정
- 분산 동질성 검정
- 평균에 대한 다중 비교 검정 (평균 비교 원 사용)
- 각 그룹에 중첩 표시되는 이상치 상자 그림
- 일원 배치의 검정력 상세 정보

그림 6.1 일원 분석



목차

일원 분석 개요	156
일원 분석의 예	156
일원 분석 플랫폼 시작	158
데이터 형식	158
일원 분석 그림	159
일원 분석 플랫폼 옵션	160
표시 옵션	163
분위수	164
이상치 상자 그림	164
평균 /ANOVA 및 평균 /ANOVA/ 합동 t	165
적합 요약 보고서	166
t-검정 보고서	166
분산 분석 보고서	167
일원 ANOVA에 대한 평균 보고서	168
블록 평균 보고서	168
평균 다이아몬드 및 X 축 비례	168
평균 선, 오차 막대 및 표준편차 선	169
평균 분석 방법	170
위치에 대한 평균 분석	170
척도에 대한 평균 분석	171
평균 분석 차트	172
평균 분석 옵션	172
평균 비교	173
비교 원 사용	175
개별 쌍 비교(스튜던트 t)	176
전체 쌍 비교(Tukey HSD)	176
최량 비교(Hsu MCB)	176
대조군 비교(Dunnett)	178
단계별 개별 쌍 비교(Newman-Keuls)	178
평균 비교 옵션	179
비모수 검정	179
Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서	181
Kolmogorov-Smirnov 2표본 검정 보고서	182
비모수 다중 비교	183
이분산	185
분산 동일 여부 검정 보고서	186
동등성 검정	187
로버스트	188

로버스트 적합	188
Cauchy 적합	188
검정력	189
검정력 상세 정보 창 및 보고서	189
정규 분위수 그림	190
CDF 그림	190
밀도	191
매칭 열	191
일원 분석 플랫폼의 추가 예	192
평균 분석 차트의 예	192
분산에 대한 평균 분석 차트의 예	193
개별 쌍 비교(스튜던트 t) 검정의 예	194
전체 쌍 비교(Tukey HSD) 검정의 예	196
최량 비교(Hsu MCB) 검정의 예	198
대조군 비교(Dunnett)의 예	199
단계별 개별 쌍 비교(Newman-Keuls) 검정의 예	200
예: 4가지 평균 비교 검정 대조	201
비모수 Wilcoxon 검정의 예	202
이분산 옵션의 예	205
동등성 검정의 예	206
로버스트 적합 옵션의 예	207
검정력 옵션의 예	209
정규 분위수 그림의 예	210
CDF 그림의 예	211
밀도 옵션의 예	212
매칭 열 옵션의 예	213
예: 일원 분석을 위한 데이터 쌓기	214
일원 분석 플랫폼에 대한 통계 상세 정보	221
비교 원	221
검정력	223
적합 요약 보고서	223
분산 동일 여부 검정	224
비모수 검정 통계량	225

일원 분석 개요

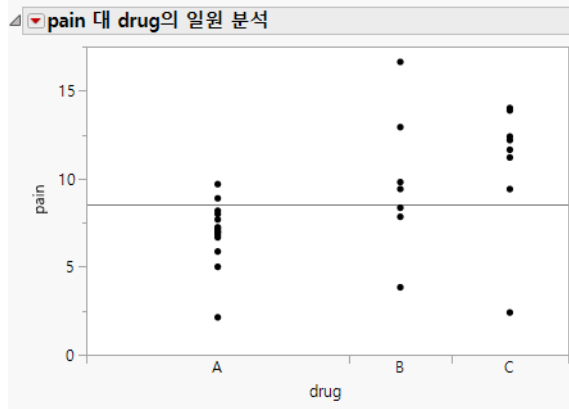
일원 분산 분석에서는 그룹 평균 간에 차이가 있는지 검정합니다. 반응의 전체 변동성은 그룹 내 변동성과 그룹 간 변동성의 두 부분으로 분할됩니다. 그룹 간 변동성이 그룹 내 변동성에 비해 크면 그룹 평균 간의 차이가 유의한 것으로 간주됩니다.

일원 분석의 예

이 예에서는 Analgesics.jmp 샘플 데이터 테이블을 사용합니다. 33 명의 참가자에게 세 가지 유형의 진통제 (A, B 및 C) 를 투여한 후 통증 수준을 슬라이딩 척도로 평가하도록 했습니다. A, B 및 C 에 대한 평균이 유의하게 다른지 여부를 확인하려고 합니다.

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

그림 6.2 일원 분석의 예



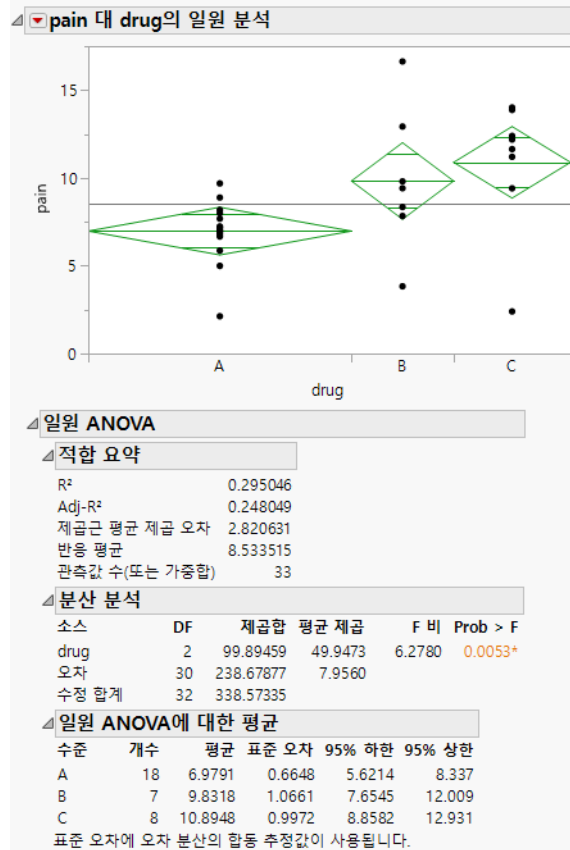
한 가지 약물 (A) 이 다른 약물보다 일관되게 낮은 점수를 받았음을 알 수 있습니다. 또한 x 축의 눈금 간격이 일정하지 않습니다. 눈금 사이의 길이는 각 약물에 대한 점수 (관측값) 의 개수와 비례합니다.

데이터에 대해 분산 분석을 수행합니다.

6. "pain 대 drug 의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다.

참고 : X 요인의 수준이 두 개뿐인 경우 **평균 /ANOVA** 옵션은 **평균 /ANOVA/ 합동 t**로 표시되며 이 옵션을 선택하면 보고서 창에 " 합동 t- 검정 " 보고서가 추가됩니다.

그림 6.3 평균 /ANOVA 옵션의 예



다음 관측값에 유의하십시오.

- 신뢰 구간을 나타내는 평균 다이아몬드가 표시됩니다.
 - 각 다이아몬드의 중심 부근에 있는 선은 그룹 평균을 나타냅니다. 각 약물의 평균에 유의한 차이가 있음을 알 수 있습니다.
 - 각 다이아몬드의 세로 범위는 각 그룹의 평균에 대한 95% 신뢰 구간을 나타냅니다.
 자세한 내용은 "**평균 다이아몬드 및 X축 비례**"(168 페이지)에서 확인하십시오.
- "적합 요약" 테이블에서는 분석에 대한 전체적인 요약 정보를 제공합니다.
- "분산 분석" 보고서에서는 표준 ANOVA 정보를 보여 줍니다. $\text{Prob} > F(p\text{ 값})$ 는 0.0053 으로 나타나는 데 이는 약물 간에 유의한 차이가 있어 보인다는 결론을 뒷받침합니다.

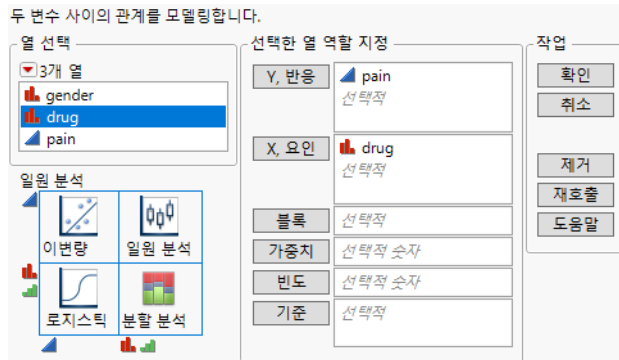
- "일원 ANOVA 에 대한 평균 " 보고서에서는 범주형 요인의 각 수준에 대한 평균 , 표본 크기 및 표준 오차를 보여 줍니다 .

일원 분석 플랫폼 시작

일원 분석을 수행하려면 다음을 수행하십시오 .

1. **분석 > X 로 Y 적합**을 선택합니다 .
2. **Y, 반응**에 연속형 열을 입력합니다 .
3. **X, 요인**에 명목형 또는 순서형 열을 입력합니다 .

그림 6.4 X 로 Y 적합 시작 창



다이아그램 위에 "일원 분석" 이라고 표시되면 일원 분석을 수행할 것임을 나타냅니다 .

참고 : "JMP 시작하기 " 창에서 일원 분석을 시작할 수도 있습니다 . **보기 > JMP 시작하기 > 기본 분석 > 일원 분석**을 선택합니다 .

이 시작 창에 대한 자세한 내용은 "**X 로 Y 적합 소개** " 장 (105 페이지) 에서 확인하십시오 . " 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 **JMP 사용의 " 시작하기 "** 장에서 확인하십시오 .

데이터 형식

일원 분석 플랫폼을 실행하려면 각 행에 X 변수 수준이 동일한 하나 이상의 관측값에 대한 정보가 포함되어 있어야 합니다 . 한 행이 둘 이상의 관측값을 나타내는 경우에는 가중치 또는 빈도 변수를 사용하여 행이 나타내는 관측값의 수를 지정해야 합니다 .

일원 데이터가 JMP 데이터 테이블이 아닌 다른 형식으로 존재하는 경우 데이터 배열이 한 행에 여러 관측값에 대한 정보가 포함되는 방식일 수도 있습니다 . JMP 에서 데이터를 분석하려면 데

이터를 가져온 후 JMP 데이터 테이블의 각 행에 단일 관측값에 대한 정보가 포함되도록 데이터를 재구성해야 합니다. 자세한 내용은 "예: 일원 분석을 위한 데이터 쌓기"(214 페이지)에서 확인하십시오.

일원 분석 그림

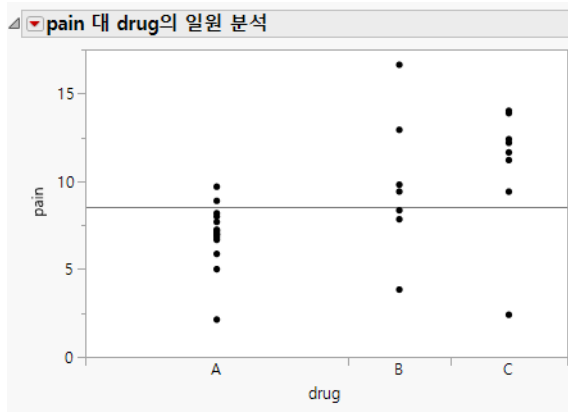
일원 분석 그림에는 각 X 요인 값의 반응 점이 표시됩니다. X 요인의 수준 간에 반응 분포를 비교할 수 있습니다. X의 구분되는 값을 수준이라고도 합니다.

두 가지 방법 중 하나로 그림의 변수를 바꿀 수 있습니다. 한 가지 방법은 한 축의 변수를 다른 축으로 드래그하여 놓아 기존 변수를 바꾸는 것이고, 다른 방법은 관련 데이터 테이블의 "열" 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다.

"일원 분석"의 빨간색 삼각형 메뉴에 있는 옵션을 사용하여 보고서 창에 보고서, 추가 그림 및 검정을 추가할 수 있습니다. 자세한 내용은 "일원 분석 플랫폼 옵션"(160 페이지)에서 확인하십시오.

그림 6.5에 표시된 것과 같은 그림을 생성하려면 "일원 분석의 예"(156 페이지)에 설명된 방법을 따르십시오.

그림 6.5 일원 분석 그림



참고: 데이터 테이블에서 제외된 모든 행은 일원 분석 그림에서도 숨겨집니다.

일원 분석 플랫폼 옵션

참고 : " 적합 그룹 " 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에 표시됩니다 . 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 를 기준으로 정렬할 수 있습니다 . 자세한 내용은 **Fitting Linear Models** 의 "Standard Least Squares Report and Options" 장에서 확인하십시오 .

플랫폼 옵션을 선택하면 그림에 개체가 추가될 수 있으며 보고서 창에는 보고서가 추가됩니다 .

표 6.1 옵션 및 요소의 예

플랫폼 옵션	그림에 추가되는 개체	보고서 창에 추가되는 보고서
분위수	상자 그림	분위수 보고서
평균 /ANOVA	평균 다이아몬드	일원 ANOVA 보고서
평균 및 표준편차	평균 선 , 오차 막대 및 표준편차 선	평균 및 표준편차 보고서
평균 비교	비교 원 (" 단계별 개별 쌍 비교 (Newman-Keuls)" 옵션 제외)	평균 비교 보고서

" 일원 분석 " 의 빨간색 삼각형 메뉴에는 다음과 같은 옵션이 있습니다 . 일부 옵션은 특정 조건이 충족되는 경우에만 표시될 수 있습니다 .

분위수 각 그룹에 대해 다음 분위수를 나열합니다 .

- 0%(최소값)
- 10%
- 25%
- 50%(중앙값)
- 75%
- 90%
- 100%(최대값)

표시 옵션 메뉴에서 **상자 그림**을 활성화합니다 . 자세한 내용은 " 분위수 "(164 페이지) 에서 확인하십시오 .

평균 /ANOVA 각 그룹의 평균을 적합시키고 일원 분산 분석을 수행하여 평균 간에 차이가 있는지 검정합니다 . 자세한 내용은 " 평균 /ANOVA 및 평균 /ANOVA/ 합동 t"(165 페이지) 에서 확인하십시오 .

참고 : X 요인의 수준이 두 개인 경우에는 이 메뉴 옵션이 **평균 /ANOVA/ 합동 t** 로 변경되며 이 옵션을 선택하면 분산이 동일하다는 가정하에 " 합동 t-검정 " 보고서가 생성됩니다.

평균 및 표준편차 각 그룹에 대한 요약 통계량을 제공합니다. 평균에 대한 표준 오차에는 표준편차의 합동 추정값 대신 개별 그룹의 표준편차가 사용됩니다.

그림에는 평균 선, 오차 막대 및 표준편차 선이 포함됩니다. 이러한 요소에 대한 간략한 설명은 " **표시 옵션** "(163 페이지) 에서 확인하십시오. 이러한 요소에 대한 자세한 내용은 " **평균 선, 오차 막대 및 표준편차 선** "(169 페이지) 에서 확인하십시오.

t-검정 (X 요인의 수준이 두 개인 경우에만 사용 가능) 분산이 동일하지 않다는 가정하에 t-검정 보고서를 생성합니다. 자세한 내용은 " **t-검정 보고서** "(166 페이지) 에서 확인하십시오.

평균 분석 방법 ANOM(평균 분석) 절차를 수행하기 위한 5 가지 명령을 제공합니다. 평균, 분산 및 범위를 비교하기 위한 명령이 있습니다. 자세한 내용은 " **평균 분석 방법** "(170 페이지) 에서 확인하십시오.

평균 비교 여러 그룹 평균을 비교하기 위한 다중 비교 방법을 제공합니다. 자세한 내용은 " **평균 비교** "(173 페이지) 에서 확인하십시오.

비모수 그룹 위치에 대한 비모수적 비교 기능을 제공합니다. 자세한 내용은 " **비모수 검정** "(179 페이지) 에서 확인하십시오.

이분산 그룹 분산의 동일 여부에 대한 4 가지 검정을 수행합니다. 또한 그룹 내의 분산이 동일하지 않을 때 평균을 비교하기 위한 ANOVA 검정인 Welch 검정을 제공합니다. 자세한 내용은 " **이분산** "(185 페이지) 에서 확인하십시오.

동등성 검정 차이가 임계값보다 작은지 검정합니다. 자세한 내용은 " **동등성 검정** "(187 페이지) 에서 확인하십시오.

로버스트 이상치가 데이터에 미치는 영향을 줄이기 위한 두 가지 방법을 제공합니다. 자세한 내용은 " **로버스트** "(188 페이지) 에서 확인하십시오.

검정력 통계 검정력에 대한 계산과 지정된 가설 검정에 대한 기타 상세 정보를 제공합니다. 자세한 내용은 " **검정력** "(189 페이지) 에서 확인하십시오.

"모형 적합" 플랫폼 내에 "검정력 상세 정보" 창 및 보고서도 표시됩니다. 검정력 계산에 대한 자세한 내용과 예는 **Fitting Linear Models** 의 부록 "Statistical Details"에서 확인하십시오.

α 수준 설정 가장 일반적인 유의 수준 옵션 중에서 하나를 선택하거나, **기타**를 선택하고 원하는 수준을 직접 지정할 수 있습니다. 유의 수준을 변경하면 다음 작업이 수행됩니다.

- 신뢰 한계가 다시 계산됩니다.
- 그림의 평균 다이아몬드가 조정됩니다 (표시된 경우).
- 보고서의 신뢰 상한 및 신뢰 하한 값이 수정됩니다.
- 모든 " 평균 비교 " 보고서의 임계값 및 비교 원이 변경됩니다.
- 모든 " 비모수 다중 비교 " 보고서의 임계값이 변경됩니다.

정규 분위수 그림 각 그룹의 데이터 분위수를 표시하기 위한 다음 옵션을 제공합니다.

분위수별 실제값 그림 세로 축에 반응 변수가 표시되고 가로 축에 분위수가 표시되는 분위수 그림을 생성합니다. 이 그림에는 범주형 X 요인의 각 수준 내에서 계산된 분위수가 표시됩니다.

실제값별 분위수 그림 x 축과 y 축을 서로 바꿉니다.

적합선 그림에 X 변수의 각 수준에 대한 직선 대각 참조선을 그림니다. 이 옵션은 그림 ("실제값 대 분위" 또는 "분위 대 실제값") 을 생성한 경우에만 사용할 수 있습니다.

CDF 그림 일원 분석 보고서의 모든 그룹에 대한 누적 분포 함수 그림을 표시합니다. 자세한 내용은 "**CDF 그림**"(190 페이지) 에서 확인하십시오.

밀도 그룹 간에 밀도를 비교합니다. 자세한 내용은 "**밀도**"(191 페이지) 에서 확인하십시오.

매칭 열 매칭 모형 분석을 수행하기 위한 매칭 변수를 지정합니다. 서로 다른 그룹의 관측값을 동일한 참가자로부터 얻은 경우처럼 일원 분석의 데이터를 매칭 (쌍체) 데이터로부터 얻은 경우에 이 옵션을 사용합니다.

그러면 매칭되는 점들을 연결하는 매칭 선이 그림에 포함됩니다. 자세한 내용은 "**매칭 열**"(191 페이지) 에서 확인하십시오.

저장 다음과 같은 통계량을 현재 데이터 테이블에 새 열로 저장합니다.

잔차 저장 반응 변수에서 요인 변수의 각 수준 내 반응 변수 평균을 빼서 계산된 값을 저장합니다.

표준화한 데이터 저장 요인 변수의 각 수준 내에서 계산된 표준화된 반응 변수 값을 저장합니다. 이 값은 중심화된 반응을 각 수준 내의 표준편차로 나눈 값입니다.

정규 분위수 저장 범주형 요인 변수의 각 수준 내에서 계산된 정규 분위수 값을 저장합니다.

예측값 저장 요인 변수의 각 수준에 대한 반응 변수의 예측 평균을 저장합니다.

표시 옵션 그림의 요소를 추가하거나 제거합니다. 자세한 내용은 "**표시 옵션**"(163 페이지) 에서 확인하십시오.

다음 옵션에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

표시 옵션

"표시 옵션"을 사용하여 그림의 요소를 추가하거나 제거할 수 있습니다. 일부 옵션은 관련이 없으면 표시되지 않을 수 있습니다.

모든 그래프 모든 그래프를 표시하거나 숨깁니다.

점 그림에 데이터 점을 표시하거나 숨깁니다.

상자 그림 각 그룹에 대한 이상치 상자 그림을 표시하거나 숨깁니다. 예는 "[일원 분석 수행](#)"(219 페이지)에서 확인하십시오.

평균 다이아몬드 가로 축에 비례하여 각 그룹의 평균을 통과하는 수평선을 그림니다. 평균 다이아몬드의 맨 위 점과 맨 아래 점은 각 그룹에 대한 95% 신뢰 구간의 상한 및 하한 점을 보여줍니다. 자세한 내용은 "[평균 다이아몬드 및 X축 비례](#)"(168 페이지)에서 확인하십시오.

평균 선 각 그룹의 평균 위치에 선을 그림니다. 자세한 내용은 "[평균 선, 오차 막대 및 표준편차 선](#)"(169 페이지)에서 확인하십시오.

평균 CI 선 각 그룹의 95% 신뢰 상한 및 하한 위치에 선을 그림니다.

평균 오차 막대 각 그룹의 평균을 식별하고 평균의 1 표준 오차 위와 아래에 오차 막대를 표시합니다. 자세한 내용은 "[평균 선, 오차 막대 및 표준편차 선](#)"(169 페이지)에서 확인하십시오.

총 평균 그림에 Y 변수의 전체 평균을 그림니다.

표준편차 선 각 그룹 평균의 1 표준편차 위와 아래에 선을 표시합니다. 자세한 내용은 "[평균 선, 오차 막대 및 표준편차 선](#)"(169 페이지)에서 확인하십시오.

비교 원 비교 원을 표시하거나 숨깁니다. 이 옵션은 **평균 비교** 옵션 중 하나를 선택한 경우에만 사용할 수 있습니다. 자세한 내용은 "[비교 원](#)"(221 페이지)에서 확인하십시오. 예는 "[일원 분석 수행](#)"(219 페이지)에서 확인하십시오.

여러 평균 연결 그룹 평균을 직선으로 연결합니다.

평균의 평균 그룹 평균의 평균 위치에 선을 그림니다.

X축 비례 각 수준의 표본 크기에 비례하여 x축에 간격을 지정합니다. 자세한 내용은 "[평균 다이아몬드 및 X축 비례](#)"(168 페이지)에서 확인하십시오.

퍼진 점 구간 너비 전체에 점을 분산시킵니다.

지터링된 점 동일한 y 값에 중첩 표시되는 여러 점 사이에 약간의 간격을 추가합니다. 가로로 조정되는 점 간격은 0.375 ~ 0.625이며 $4 * (\text{Uniform}(0,1) - 0.5)^5$ 분포를 따릅니다.

매칭 선 (**매칭 열** 옵션을 선택한 경우에만 표시됨) 매칭되는 점들을 연결합니다.

매칭 점선 (**매칭 열** 옵션을 선택한 경우에만 표시됨) 테이블에 있는 결측 셀의 셀 평균을 연결하는 점선을 그림니다. 선의 끝점으로 사용되는 값은 이원 ANOVA 모형을 사용하여 구합니다.

히스토그램 원래 그림의 오른쪽에 가로 배열 히스토그램을 그림니다 .

로버스트 평균 선 (" 로버스트 " 옵션을 선택한 경우에만 표시됨) 각 그룹의 로버스트 평균 위치에 선을 그림니다 .

범례 " 정규 분위수 그림 ", " CDF 그림 " 및 " 밀도 " 옵션에 대해 범례를 표시합니다 .

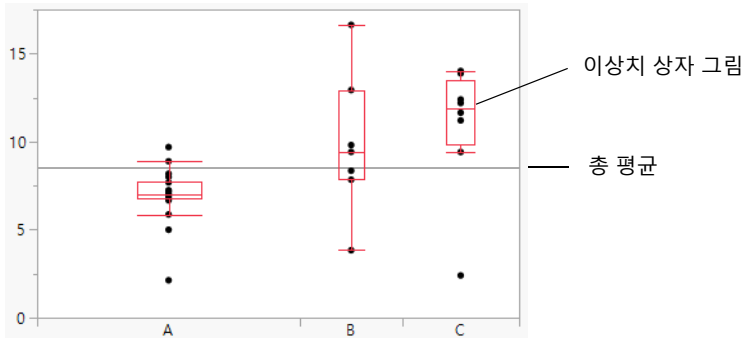
분위수

" 분위수 " 보고서에는 X 요인 변수의 각 수준에 대해 선택된 백분위수가 나열됩니다 . 중앙값은 50 번째 백분위수이고 , 25 번째 및 75 번째 백분위수는 사분위수라고 합니다 .

분위수 옵션은 그림에 다음과 같은 요소를 추가합니다 .

- Y 변수의 전체 평균을 나타내는 총 평균
- 각 요인 수준의 점 분포를 요약한 이상치 상자 그림

그림 6.6 이상치 상자 그림 및 총 평균



참고 : 이러한 요소를 숨기려면 " 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림** 또는 **총 평균**을 선택하십시오 .

이상치 상자 그림

이상치 상자 그림은 데이터의 분포를 그래픽으로 요약한 것입니다 . 이상치 상자 그림의 경우 다음 사항에 유의하십시오 (그림 6.7).

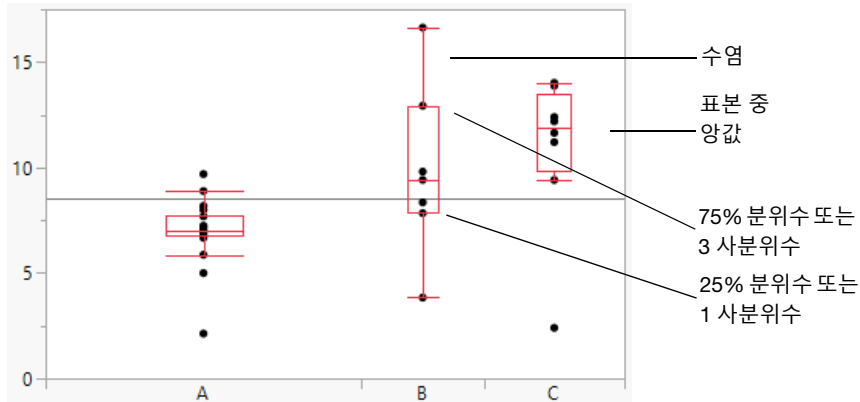
- 상자 내의 가로 선은 표본 중앙값을 나타냅니다 .
- 상자 끝은 75번째 및 25번째 분위수를 나타내며 각각 3사분위수 및 1사분위수라고도 합니다 .
- 1 사분위수와 3 사분위수 사이의 차이를 사분위수 범위 . 라고 합니다 .

- 각 상자에는 양 끝에서 시작되는 선이 있으며 이 선은 수염이라고도 합니다. 수염은 상자 끝에서 시작해서 다음과 같이 계산된 거리 내에 있는 가장 바깥쪽 데이터 점으로 이어집니다.

$$3 \text{ 사분위수} + 1.5 * (\text{사분위수 범위})$$

$$1 \text{ 사분위수} - 1.5 * (\text{사분위수 범위})$$
 데이터 점이 계산된 범위에 포함되지 않으면 상한 및 하한 데이터 점 값 (이상치 제외)에 따라 수염이 결정됩니다.

그림 6.7 이상치 상자 그림의 예



평균 /ANOVA 및 평균 /ANOVA/ 합동 t

평균 /ANOVA 옵션은 분산 분석을 수행합니다. X 요인에 정확히 두 개의 수준만 포함되어 있는 경우 이 옵션은 **평균 /ANOVA/ 합동 t**로 표시됩니다. 이 경우 다른 보고서 외에도 합동 분산 (또는 등분산)을 가정하는 합동 t-검정 보고서가 표시됩니다.

일원 분석 그림에 평균 다이아몬드가 추가됨 자세한 내용은 "**표시 옵션**"(163 페이지) 및 "**평균 다이아몬드 및 X 축 비례**"(168 페이지)에서 확인하십시오.

보고서 자세한 내용은 "**적합 요약 보고서**"(166 페이지), "**분산 분석 보고서**"(167 페이지), "**일원 ANOVA에 대한 평균 보고서**"(168 페이지), "**t-검정 보고서**"(166 페이지) 및 "**블록 평균 보고서**"(168 페이지)에서 확인하십시오.

- "t-검정" 보고서는 **평균 /ANOVA/ 합동 t** 옵션을 선택한 경우에만 표시됩니다.
- "블록 평균" 보고서는 시작 창에서 블록 변수를 지정한 경우에만 표시됩니다.

적합 요약 보고서

" 적합 요약 " 보고서에는 일원 분산 분석에 대한 요약이 표시됩니다 .

R² 각 요인 수준에 평균을 적합시킴으로써 설명되는 변동의 비율을 측정합니다. 나머지 변동은 랜덤 오차에 기인합니다. 그룹 평균을 적합시킨 결과 오차 없이 모든 변동이 설명되면 R² 값이 1 이 됩니다. R² 가 0 이면 해당 적합이 전체 반응 평균보다 더 적절한 예측 모형이 아님을 나타냅니다. 자세한 내용은 " [적합 요약 보고서](#) "(223 페이지) 에 대한 통계 상세 정보에서 확인하십시오 .

R² 를 결정 계수라고도 합니다 .

참고 : R² 값이 낮으면 설명되지 않은 변동을 설명하는 변수가 모형에 없는 것일 수 있습니다. 하지만 데이터의 내재 변동 범위가 큰 경우에는 유용한 ANOVA 모형이라도 R² 값이 낮을 수 있습니다. 일반적인 R² 값에 대해 알아보려면 연구 영역의 문헌을 읽어 보십시오 .

Adj-R² 모수의 개수가 서로 다른 모형들을 비교하기 좋도록 계산에 자유도를 사용하여 R² 값을 조정합니다. 자세한 내용은 " [적합 요약 보고서](#) "(223 페이지) 에 대한 통계 상세 정보에서 확인하십시오 .

제공된 평균 제공 오차 랜덤 오차의 표준편차를 추정합니다. " 분산 분석 " 보고서에 표시된 오차에 대한 평균 제공의 제공근입니다 .

반응 평균 반응 변수의 전체 평균 (산술평균) 입니다 .

관측값 수 (또는 가중합) 적합 추정에 사용된 관측값의 개수입니다 . 가중치가 사용된 경우에는 가중치의 합입니다. 자세한 내용은 " [적합 요약 보고서](#) "(223 페이지) 에 대한 통계 상세 정보에서 확인하십시오 .

t- 검정 보고서

t- 검정에는 다음과 같은 두 가지 유형이 있습니다 .

- 등분산 : **평균 /ANOVA/ 합동 t** 옵션을 선택하면 " 합동 t- 검정 " 보고서가 표시됩니다. 이 t- 검정에서는 등분산을 가정합니다 .
- 이분산 : 빨간색 삼각형 메뉴에서 **t- 검정** 옵션을 선택하면 " t- 검정 " 보고서가 표시됩니다. 이 t- 검정에서는 등분산을 가정합니다 .

" t- 검정 " 보고서에는 다음 열이 포함됩니다 .

t- 검정 그림 귀무가설이 참이라는 가정하에 평균의 차이에 대한 표본 분포를 보여 줍니다. 빨간색 세로 선은 평균의 실제 차이입니다. 음영 영역은 p 값에 해당합니다 .

차이 두 X 수준 간의 추정 차이를 보여 줍니다. 이 그림에서 " 차이 " 값은 두 수준을 비교하는 빨간색 선으로 표시됩니다 .

차이 표준 오차 차이의 표준 오차를 보여 줍니다 .

차이 CL 상한 차이의 신뢰 상한을 보여 줍니다 .

차이 CL 하한 차이의 신뢰 하한을 보여 줍니다.

신뢰도 신뢰 수준 ($1-\alpha$) 을 보여 줍니다. 신뢰 수준을 변경하려면 플랫폼의 빨간색 삼각형 메뉴에 있는 **α 수준 설정** 명령에서 새 유의 수준을 선택합니다.

t 비 t 통계량의 값입니다.

DF t-검정에 사용된 자유도입니다.

Prob > |t| 양쪽 꼬리 검정과 관련된 p 값입니다.

Prob > t 위쪽 꼬리 검정과 관련된 p 값입니다.

Prob < t 아래쪽 꼬리 검정과 관련된 p 값입니다.

분산 분석 보고서

"분산 분석" 보고서에서는 표본의 총 변동을 두 성분으로 분할합니다. 두 평균 제곱의 비율은 F 비가 됩니다. F 비와 관련된 확률이 낮으면 해당 모형은 전체 반응 평균보다 통계적으로 더 적절한 적합한 것입니다.

참고: 블록 열을 지정한 경우에는 "분산 분석" 보고서에 **블록** 변수가 포함됩니다.

소스 세 가지 변동 소스를 나열합니다. 모형 소스, **오차** 및 **수정 합계**가 이러한 소스에 해당합니다.

DF 각 변동 소스의 관련 DF(자유도)를 기록합니다.

- **수정 합계**의 자유도는 $N - 1$ 이며, 여기서 N 은 분석에 사용된 관측값의 총 개수입니다.
- X 요인의 수준이 k 개이이면 모형의 자유도는 $k - 1$ 입니다.

오차 자유도는 **수정 합계** 자유도와 **모형** 자유도의 차이 (즉, $N - k$) 입니다.

제곱합 각 변동 소스의 제곱합 (줄여서 SS 라고 함) 을 기록합니다.

- 전체 반응 평균과 각 반응의 차이에 대한 총 제곱합 (**수정 합계**). **수정 합계** 제곱합은 다른 모든 모형과의 비교에 사용되는 기본 모형입니다.
- 각 점으로부터 해당 그룹 평균까지의 거리에 대한 제곱합. 이 값은 분산 분석 모형을 적합시킨 후 설명되지 않은 나머지 **오차** (잔차) 의 SS 입니다.

총 SS 에서 오차 SS 를 뺀 값은 모형에 기인한 제곱합입니다. 이를 통해 총 변동 중 모형으로 설명되는 변동의 비율을 알 수 있습니다.

평균 제곱 제곱합을 관련 자유도로 나눈 것입니다.

- **모형** 평균 제곱은 그룹 평균이 동일하다는 가설을 전제로 오차의 분산을 추정합니다.
- **오차** 평균 제곱은 모형 평균 제곱과는 독립적으로 오차 항의 분산을 추정하며 모형 가설이 전제되지 않습니다.

F 비 모형 평균 제곱을 오차 평균 제곱으로 나눈 것입니다. 그룹 평균이 동일하다는 가설, 즉 그룹 평균 간에 실질적인 차이가 없다는 가설이 참인 경우에는 오차의 평균 제곱과 모형의 평균 제곱이 모두 오차 분산을 추정합니다. 이 둘의 비율은 F 분포를 따릅니다. 분산 분석 모형의 결과가 합계보다 유의하게 낮은 변동을 보이면 F 비가 기대값보다 높게 나타납니다.

Prob>F 모집단 그룹 평균에 차이가 없는 경우 F 값이 계산된 값보다 클 확률입니다. 관측된 유의 확률이 0.05 이하인 것은 종종 그룹 평균에 차이가 있는 증거로 간주됩니다.

일원 ANOVA 에 대한 평균 보고서

"일원 ANOVA 에 대한 평균" 보고서에는 명목형 또는 순서형 요인의 각 수준에 대한 반응 정보가 요약됩니다.

수준 X 변수의 수준을 나열합니다.

개수 각 그룹의 관측값 수를 나열합니다.

평균 각 그룹의 평균을 나열합니다.

표준 오차 그룹 평균에 대한 표준편차 추정값을 나열합니다. 이 표준 오차는 각 수준에서 반응의 분산이 동일하다는 가정하에 추정됩니다. 이 값은 "적합 요약" 보고서에 나오는 제공된 평균 제공 오차를 그룹 평균 계산에 사용된 값 개수의 제곱근으로 나눈 값입니다.

95% 하한 및 95% 상한 그룹 평균에 대한 95% 신뢰 하한 및 상한을 나열합니다.

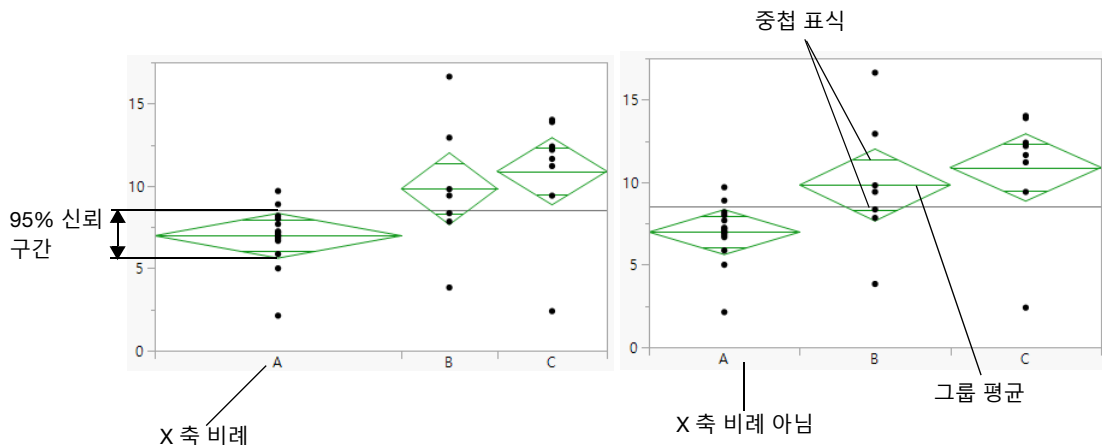
블록 평균 보고서

시작 창에서 블록 변수를 지정한 경우 **평균 /ANOVA** 및 **평균 /ANOVA/ 합동 t** 명령을 선택하면 "블록 평균" 보고서가 생성됩니다. 이 보고서에는 각 블록의 평균과 각 블록의 관측값 수가 표시됩니다.

평균 다이아몬드 및 X 축 비례

평균 다이아몬드는 표본 평균과 신뢰 구간을 나타낸 것입니다.

그림 6.8 평균 다이아몬드 및 X 축 비례 옵션의 예



다음 관측값에 유의하십시오 .

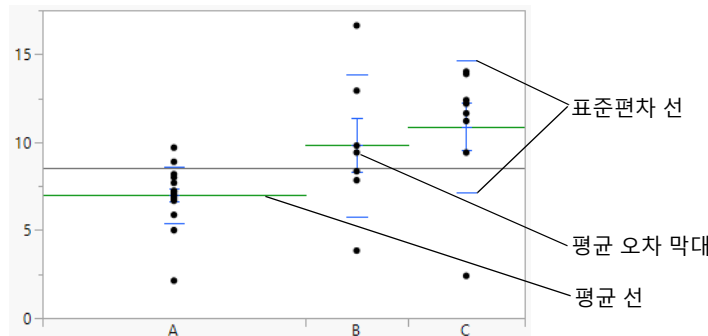
- 각 다이아몬드의 맨 위와 맨 아래는 각 그룹에 대한 $(1 - \alpha) \times 100$ 신뢰 구간을 나타냅니다 . 신뢰 구간은 관측값 간에 분산이 동일하다는 가정하에 계산됩니다 . 따라서 다이아몬드의 높이는 그룹 내의 관측값 수에 대한 제곱근의 역수에 비례합니다 .
- **X 축 비례** 옵션이 선택된 경우 가로 축에서 각 그룹이 차지하는 가로 범위 (다이아몬드의 가로 크기) 는 X 변수의 각 수준에 대한 표본 크기에 비례합니다 . 따라서 일반적으로 다이아몬드의 가로가 좁을수록 세로가 길어지는데, 이는 데이터 점이 적을수록 신뢰 구간이 넓어지기 때문입니다 .
- 각 다이아몬드의 가운데를 가로지르는 평균 선은 그룹 평균을 나타냅니다 .
- 중첩 표식은 그룹 평균의 위쪽과 아래쪽에 선으로 표시됩니다 . 표본 크기가 동일한 그룹의 경우 중첩 표식은 해당 신뢰 수준에서 두 그룹 평균이 유의하게 다르지 않음을 나타냅니다 . 중첩 표식은 그룹 평균 $\pm (\sqrt{2})/2 \times CI/2$ 로 계산됩니다 . 한 다이아몬드의 중첩 표식이 해당 다이아몬드의 중첩 표식보다 다른 다이아몬드의 평균에 더 가까우면 해당 신뢰 수준에서 관련된 두 그룹이 다르지 않음을 나타냅니다 .
- 플랫폼 메뉴에서 **평균 /ANOVA/ 합동 t** 또는 **평균 /ANOVA** 옵션을 선택하면 자동으로 평균 다이아몬드가 표시됩니다 . 하지만 언제든지 빨간색 삼각형 메뉴에서 **표시 옵션 > 평균 다이아몬드**를 선택하여 이를 표시하거나 숨길 수도 있습니다 .

평균 선, 오차 막대 및 표준편차 선

표시 옵션 > 평균 선을 선택하여 평균 선을 표시할 수 있습니다 . 평균 선은 X 변수의 각 수준에 대한 반응 평균을 나타냅니다 .

빨간색 삼각형 메뉴에서 **평균 및 표준편차** 옵션을 선택하면 평균 오차 막대와 표준편차 선이 표시됩니다 . 각 옵션을 개별적으로 설정하거나 해제하려면 **표시 옵션 > 평균 오차 막대** 또는 **표준편차 선**을 선택합니다 .

그림 6.9 평균 선, 평균 오차 막대 및 표준편차 선



평균 분석 방법

ANOM(평균 분석) 방법은 여러 그룹의 평균 및 분산과 그 밖의 위치 및 척도 측도를 비교합니다. 다음과 같은 경우에 이러한 방법을 사용할 수 있습니다.

- 전체 (표본) 평균과 통계적으로 다른 그룹 평균이 있는지 여부를 검정하려는 경우
- RMSE(제곱근 평균 제곱 오차) 와 통계적으로 다른 그룹 표준편차가 있는지 여부를 검정하려는 경우
- 전체 범위 평균과 통계적으로 다른 그룹 범위가 있는지 여부를 검정하려는 경우

참고: 분할 분석 플랫폼 내에서는 반응에 두 가지 범주가 있는 경우에 **비율에 대한 평균 분석**을 사용할 수 있습니다. 자세한 내용은 "[분할 분석](#)" 장 (229 페이지) 에서 확인하십시오.

ANOM 방법과 JMP에서의 ANOM 구현 방식에 대한 자세한 내용은 Nelson et al. 의 저서 (2005) 에서 확인하십시오.

위치에 대한 평균 분석

다음 옵션을 사용하여 그룹의 평균 또는 중심 값이 공통적인지 여부를 검정할 수 있습니다.

- ANOM
- 변환 순위를 사용한 ANOM

ANOM

그룹 평균을 전체 평균과 비교하려면 ANOM 을 사용합니다. 이 방법은 데이터가 대략적으로 정규 분포되었다고 가정합니다. 자세한 내용은 "[평균 분석 차트의 예](#)" (192 페이지) 에서 확인하십시오.

변환 순위를 사용한 ANOM

이 방법은 **ANOM** 분석의 비모수적 버전입니다. 데이터가 명확하게 비정규 데이터이고 정규 데이터로 변환할 수 없는 경우 이 방법을 사용합니다. "변환 순위를 사용한 ANOM" 은 각 그룹 평균의 변환 순위를 전체 평균의 변환 순위와 비교합니다. ANOM 검정을 위해서는 변환된 관측값에 일반적인 ANOM 절차와 임계값을 적용해야 합니다.

변환 순위

n 개의 관측값이 있다고 가정해 보겠습니다. 변환된 관측값은 다음과 같이 계산됩니다.

- 각각의 동일 값을 포함한 모든 관측값에 최소값에서 최대값의 순서로 순위를 매깁니다. 동일한 관측값의 경우 해당 관측값들이 공유하는 순위 블록의 평균을 각 관측값에 할당합니다.
- 순위는 R_1, R_2, \dots, R_n 으로 나타냅니다.

- i 번째 관측값에 해당하는 변환 순위는 다음과 같이 결정됩니다.

$$\text{변환된 } R_i = \text{Normal Quantile} \left[\left(\frac{R_i}{2n+1} \right) + 0.5 \right]$$

ANOM 절차는 변환된 R_i 값에 적용됩니다. 순위는 균등 분포를 따르므로 변환 순위는 접힌 정규 분포를 따릅니다. 자세한 내용은 Nelson et al. 연구 자료 (2005) 에서 확인하십시오.

척도에 대한 평균 분석

다음 옵션을 사용하여 그룹 내의 변동 동질성을 검정할 수 있습니다.

- 분산에 대한 ANOM
- Levene(ADM) 을 사용한 분산에 대한 ANOM
- 범위에 대한 ANOM

분산에 대한 ANOM

그룹 표준편차 (또는 분산) 를 제공된 평균 제공 오차 (또는 평균 제공 오차) 와 비교하려면 이 방법을 사용합니다. 이 방법은 데이터가 대략적으로 정규 분포되었다고 가정합니다. 이 방법을 사용하려면 각 그룹마다 4 개 이상의 관측값이 있어야 합니다. 분산에 대한 ANOM 검정과 관련된 자세한 내용은 Wludyka and Nelson 연구 자료 (1997) 및 Nelson et al. 연구 자료 (2005) 에서 확인하십시오. 예는 "분산에 대한 평균 분석 차트의 예"(193 페이지) 에서 확인하십시오.

Levene(ADM) 을 사용한 분산에 대한 ANOM

이 방법은 ADM(중앙값과의 절대 편차) 의 그룹 평균을 전체 평균 ADM 과 비교하는 로버스트 검정을 제공합니다. 데이터가 비정규 데이터이고 정규 데이터로 변환할 수 없는 것으로 의심되는 경우에 "Levene(ADM) 을 사용한 분산에 대한 ANOM" 을 사용합니다. "Levene(ADM) 을 사용한 분산에 대한 ANOM" 은 "분산에 대한 ANOM" 분석과 유사한 비모수적 분석 방법입니다. "Levene(ADM) 을 사용한 분산에 대한 ANOM" 검정에 대한 자세한 내용은 Levene 연구 자료 (1960) 또는 Brown and Forsythe 연구 자료 (1974) 에서 확인하십시오.

범위에 대한 ANOM

그룹 범위를 그룹 범위의 평균과 비교하려면 이 검정을 사용합니다. 이 검정은 퍼짐에 대한 측도로서 범위를 기준으로 한 척도 차이에 대한 검정입니다. 자세한 내용은 Wheeler 연구 자료(2003) 에서 확인하십시오.

참고 : "범위에 대한 ANOM" 은 균형 설계와 특정 그룹 크기에만 사용할 수 있습니다. 자세한 내용은 "범위에 대한 ANOM 검정의 제한 사항"(172 페이지) 에서 확인하십시오.

범위에 대한 ANOM 검정의 제한 사항

다른 ANOM 결정 한계와 달리 "범위에 대한 ANOM" 차트의 결정 한계에는 제공된 임계값만 사용됩니다. 따라서 "범위에 대한 ANOM"은 다음의 경우에만 사용할 수 있습니다.

- 크기가 동일한 그룹
- 특정 크기 (2~10, 12, 15 및 20) 의 그룹
- 2~30 개의 그룹
- 유의 수준 0.10, 0.05 및 0.01

평균 분석 차트

"평균 분석 방법"의 각 옵션은 보고서 창에 다음을 표시하는 차트를 추가합니다.

- UDL(결정 상한)
- LDL(결정 하한)
- 결정 한계 사이에 다음 항목으로 배치되는 가로 중심선
 - ANOM: 전체 평균
 - 변환 순위를 사용한 ANOM: 변환 순위의 전체 평균
 - 분산에 대한 ANOM: 제공된 평균 제공 오차 (분산 척도의 경우에는 평균 제공 오차)
 - Levene(ADM) 을 사용한 분산에 대한 ANOM: 평균과의 전체 절대 편차
 - 범위에 대한 ANOM: 그룹 범위의 평균

그룹에 표시된 그룹의 통계량이 결정 한계를 벗어나는 경우 이 검정은 해당 그룹의 통계량과 모든 그룹에 대한 전체 평균 통계량 사이에 통계적인 차이가 있음을 나타냅니다.

평균 분석 옵션

"평균 분석 방법"의 각 옵션은 보고서 창의 "평균 분석" 옆에 빨간색 삼각형 메뉴를 추가합니다.

유의 수준 설정 가장 일반적인 유의 수준 옵션 중에서 하나를 선택하거나, **기타**를 선택하고 원하는 수준을 직접 지정합니다. 유의 수준을 변경하면 결정 상한 및 결정 하한이 수정됩니다.

참고 : " 범위에 대한 ANOM" 의 경우에는 0.10, 0.05 및 0.01 만 선택할 수 있습니다.

요약 보고서 표시 보고서는 다음과 같이 평균 분석 방법에 따라 달라집니다.

- ANOM: 그룹 평균 및 결정 한계를 보여 주는 보고서가 생성됩니다.
- 변환 순위를 사용한 ANOM: 그룹 평균 변환 순위 및 결정 한계를 보여 주는 보고서가 생성됩니다.
- 분산에 대한 ANOM: 그룹 표준편차 (또는 분산) 및 결정 한계를 보여 주는 보고서가 생성됩니다.

- Levene(ADM) 을 사용한 분산에 대한 ANOM: 그룹 평균 ADM 및 결정 한계를 보여 주는 보고서가 생성됩니다.
- 범위에 대한 ANOM: 그룹 범위 및 결정 한계를 보여 주는 보고서가 생성됩니다.

분산 척도의 그래프 (분산에 대한 ANOM 에만 사용 가능) 세로 축의 척도를 표준편차에서 분산으로 변경합니다.

표시 옵션 표시 모양을 사용자 정의하기 위한 다음 옵션이 포함되어 있습니다.

결정 한계 표시 결정 한계선을 표시하거나 숨깁니다.

결정 한계 음영 표시 결정 한계 음영을 표시하거나 숨깁니다.

중심선 표시 중심선 통계량을 표시하거나 숨깁니다.

점 옵션: 바늘 표시 바늘을 표시합니다. 이 옵션은 기본 옵션입니다. **연결된 점 표시**는 각 그룹의 평균을 연결하는 선을 표시합니다. **점만 표시**는 각 그룹의 평균을 나타내는 점만 표시합니다.

평균 비교

참고: 평균을 비교하는 또 다른 방법은 ANOM 입니다. 자세한 내용은 "[평균 분석 방법](#)"(170 페이지)에서 확인하십시오.

그룹 평균에 대한 다중 비교를 수행하려면 "평균 비교" 옵션을 사용합니다. 모든 평균 비교 방법에는 평균에 대한 합동 분산 추정값이 사용됩니다. "단계별 개별 쌍 비교 (Newman-Keuls)" 방법을 제외한 각 "평균 비교" 옵션은 그림 옆에 비교 원을 추가하고 보고서 창에 특정 보고서를 추가합니다. 비교 원에 대한 자세한 내용은 "[비교 원 사용](#)"(175 페이지)에서 확인하십시오.

옵션	설명	비모수 메뉴 옵션
개별 쌍 비교 (스튜던트 t)	스튜던트 t 검정을 사용하여 개별 쌍별 비교를 수행합니다. 다중 쌍별 검정을 수행하는 경우 추론 간에 보호가 적용되지 않습니다. 따라서 가설 검정의 유의 수준(제 1 종 오류 비율)이 개별 검정의 유의 수준보다 높습니다. 자세한 내용은 " 개별 쌍 비교 (스튜던트 t) "(176 페이지)에서 확인하십시오.	비모수 > 비모수 다중 비교 > Wilcoxon 개별 쌍 비교

옵션	설명	비모수 메뉴 옵션
전체 쌍 비교 (Tukey HSD)	평균 간의 모든 차이에 대해 크기 조정된 검정을 표시합니다 . 이는 <i>Tukey</i> 또는 <i>Tukey-Kramer HSD</i> (Honestly Significant Difference) 검정 (Tukey 1953, Kramer 1956)입니다. 이 검정은 표본 크기가 동일한 경우에는 정확한 유의 수준 검정이고 , 표본 크기가 서로 다른 경우에는 보수적인 검정입니다 (Hayter 1984). 자세한 내용은 " 전체 쌍 비교 (Tukey HSD)"(176 페이지) 에서 확인하십시오 .	비모수 > 비모수 다중 비교 > Steel-Dwass 전체 쌍 비교
최량 비교 (Hsu MCB)	평균이 알 수 없는 최대값보다 작거나 알 수 없는 최소값보다 큰지 여부를 검정합니다 . 이는 Hsu MCB 검정 (Hsu 1996, Hsu 1981) 입니다 . 자세한 내용은 " 최량 비교 (Hsu MCB)"(176 페이지) 에서 확인하십시오 .	없음
대조군 비교 (Dunnett)	평균이 대조군의 평균과 다른지 여부를 검정합니다. 이는 Dunnett 검정 (Dunnett 1955) 입니다 . 자세한 내용은 " 대조군 비교 (Dunnett)"(178 페이지) 에서 확인하십시오 .	비모수 > 비모수 다중 비교 > Steel 대조군 비교
단계별 개별 쌍 비교 (Newman-Keuls)	단계별 절차에서 스튜던트화 범위 검정을 사용하여 평균 간에 차이가 있는지 여부를 검정합니다. 이는 Newman-Keuls 또는 Student-Newman-Keuls 방법 (Keuls, 1952) 입니다 . 이 검정은 Tukey HSD 검정보다 덜 보수적입니다 . 자세한 내용은 " 단계별 개별 쌍 비교 (Newman-Keuls)"(178 페이지) 에서 확인하십시오 .	없음

참고 : 블록 열을 지정한 경우에는 블록 평균에 맞게 조정된 데이터에 대해 다중 비교 방법이 수행됩니다 .

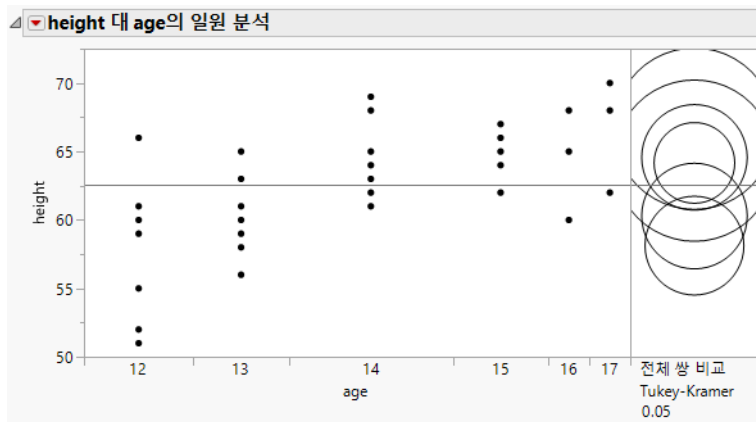
이러한 모든 검정을 보여 주는 예는 " 예 : 4 가지 평균 비교 검정 대조 "(201 페이지) 에서 확인하십시오 .

비교 원 사용

참고 : 비교 원 그림을 영구적으로 숨기려면 **파일 > 환경 설정 > 플랫폼 > 일원 분석**을 선택하고 **비교 원** 옵션을 선택 취소하십시오.

단계별 개별 쌍 비교 (Newman-Keuls)" 방법을 제외한 각 다중 비교 검정은 그룹 평균 비교를 시각적으로 나타내는 **비교 원** 그림으로 시작됩니다. 그림 6.10에서는 "전체 쌍 비교 (Tukey HSD)" 방법에 사용되는 비교 원을 보여 줍니다. 다른 비교 검정의 경우에는 원의 반지름이 이보다 길거나 짧습니다.

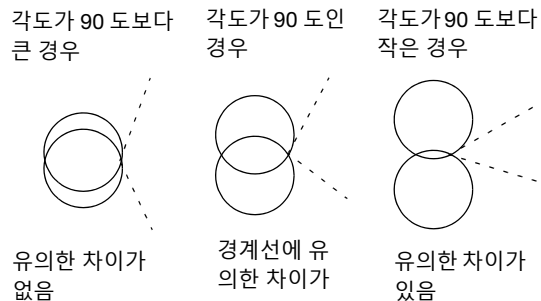
그림 6.10 그룹 평균의 시각적 비교



비교 원의 교차 영역을 검토하여 각 그룹 평균 쌍을 비교할 수 있습니다. 교차점의 외각은 해당 그룹 평균 간에 유의한 차이가 있는지 여부를 알려 줍니다 (그림 6.11).

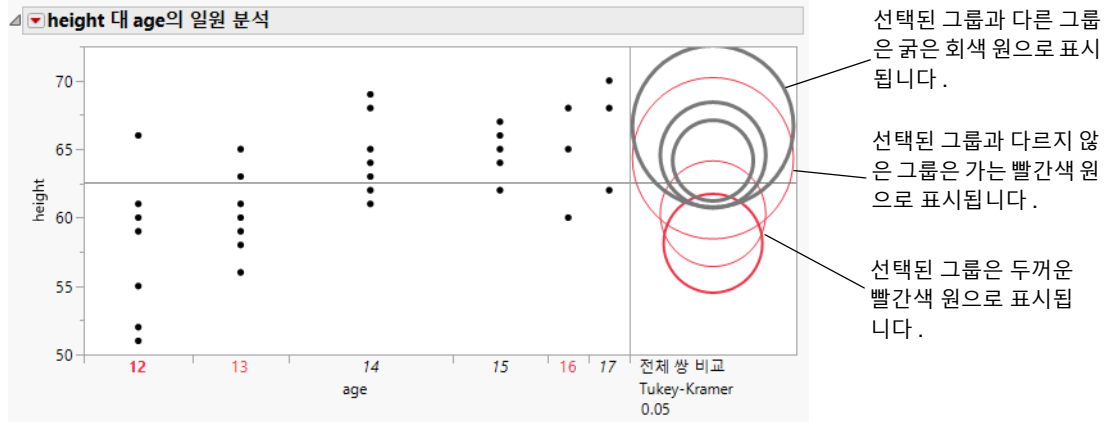
- 유의한 차이가 있는 평균의 경우 비교 원이 교차하지 않거나 약간 교차하므로 교차점의 외각이 90도 미만입니다.
- 원의 교차 각도가 90도를 초과하거나 원이 내포되는 경우에는 평균 간에 유의한 차이가 없는 것입니다.

그림 6.11 교차 각도 및 유의성



교차 각도가 90 도에 가까운 경우 비교 원을 클릭하여 선택하면 평균 간에 유의한 차이가 있는지 여부를 확인할 수 있습니다 (그림 6.12). 원을 선택 취소하려면 원 외부의 공백을 클릭합니다 .

그림 6.12 비교 원 강조 표시



관련 정보

- " 비교 원 "(221 페이지)

개별 쌍 비교 (스튜던트 t)

개별 쌍 비교 (스튜던트 t) 검정에서는 각 그룹 수준 쌍에 대한 스튜던트 t 검정을 보여 주고 개별 비교만 검정합니다 . 자세한 내용은 " 개별 쌍 비교 (스튜던트 t) 검정의 예 "(194 페이지) 에서 확인하십시오 .

전체 쌍 비교 (Tukey HSD)

전체 쌍 비교 (Tukey HSD) 검정 (Tukey-Kramer 라고도 함)에서는 모든 쌍 조합의 유의성 검정을 보호하므로 HSD 구간이 스튜던트 t 쌍별 LSD 보다 커집니다 . 그래픽에서는 비교 원이 더 크고 차이가 덜 유의하게 나타납니다 .

q 통계량은 $q^* = (1/\sqrt{2}) * q$ 로 계산됩니다 . 여기서 q 는 스튜던트화 범위 분포에 필요한 백분위수입니다 . T 통계량에 대한 자세한 내용은 Neter et al. 연구 자료 (1990) 에서 확인하십시오 . " 전체 쌍 비교 (Tukey HSD) 검정의 예 "(196 페이지) 도 확인하십시오 .

최량 비교 (Hsu MCB)

최량 비교 (Hsu MCB) 검정에서는 특정 수준의 평균이 나머지 수준의 최대 평균보다 크거나 나머지 수준의 최소 평균보다 작은지 여부를 검정합니다 . 자세한 내용은 Hsu 연구 자료 (1996) 에서 확인하십시오 . 이 검정의 예는 "최량 비교(Hsu MCB) 검정의 예"(198페이지)에서 확인하십시오.

Hsu MCB 검정의 분위수는 범주형 변수의 수준에 따라 달라집니다. 각 수준의 표본 크기가 동일하지 않으면 비교 원 기법이 정확하지 않습니다. 비교 원의 반지름은 수준의 표준 오차에 최대 분위수 값을 곱해서 구합니다. 유의한 차이가 있는지 정확히 평가하려면 p 값을 사용합니다. 자세한 내용은 "**최대값 및 최소값을 사용한 비교**"(177 페이지)에서 확인하십시오.

참고 : MCB에 의해 최대값 또는 최소값으로 간주되지 않는 평균은 잠재적 최대 또는 최소 평균에 대한 Gupta(1965)의 선택된 부분집합에도 포함되지 않습니다.

신뢰 분위수

이 보고서에서는 범주형 변수의 각 수준에 대한 분위수를 제공합니다. 이러한 분위수는 지정된 알파 값에 대응합니다.

최대값 및 최소값을 사용한 비교

이 보고서에는 단측 Dunnett 검정의 p 값이 표시됩니다. 최량 수준이 아닌 각 수준에 대해 주어진 p 값은 표본 최량 수준의 평균을 대조 수준 (잠재적 최량 수준)으로 처리되는 나머지 각 수준의 평균과 비교하는 검정을 위한 것입니다. 표본 최량 수준의 p 값은 두 번째 표본 최량 수준의 평균을 대조 수준으로 처리되는 표본 최량 수준의 평균과 비교해서 구합니다.

이 보고서에는 세 개의 열이 표시됩니다.

수준 범주형 변수의 수준입니다.

최대 p 값 사용 이 열에서는 범주형 변수의 각 수준에 대해 해당 수준의 평균이 나머지 수준의 최대 평균을 초과하는지 검정하기 위한 p 값을 제공합니다. 평균이 최대 실제 평균 (알 수 없음)보다 작거나 같은 수준을 선별해서 제외하려면 이 열의 검정을 사용합니다.

최소 p 값 사용 이 열에서는 범주형 변수의 각 수준에 대해 해당 수준의 평균이 나머지 수준의 최소 평균보다 작은지 검정하기 위한 p 값을 제공합니다. 평균이 최소 실제 평균 (알 수 없음)보다 크거나 같은 수준을 선별해서 제외하려면 이 열의 검정을 사용합니다.

LSD 임계 행렬

첫 번째로 표시되는 보고서는 최대값에 대한 것이고 두 번째는 최소값에 대한 것입니다.

최대값 보고서의 경우 행 평균에서 열 평균을 빼고 LSD를 뺀 값을 포함하는 열이 표시됩니다. 값이 양수이면 해당 행 평균이 열 평균보다 유의하게 높으며 열 평균이 최대값이 아님을 나타냅니다.

최소값 보고서의 경우 행 평균에서 열 평균을 빼고 LSD를 더한 값을 포함하는 열이 표시됩니다. 값이 음수이면 해당 행 평균이 열 평균보다 유의하게 낮으며 열 평균이 최소값이 아님을 나타냅니다.

대조군 비교 (Dunnett)

대조군 비교 (Dunnett)에서는 일련의 평균을 대조군 평균과 비교합니다. 산출되는 LSD 는 중간 비교 개수를 제한할 수 있도록 크기가 조정되므로 스튜던트 t 와 Tukey-Kramer LSD 사이에 있습니다. 이 검정의 예는 "**대조군 비교 (Dunnett)의 예**"(199 페이지)에서 확인하십시오.

Dunnett 검정 보고서에는 $|t|$ 분위수가 표시되며 이를 스튜던트 t 통계량과 유사한 방식으로 사용할 수 있습니다. LSD 임계 행렬에는 차이 - LSD 의 절대값이 표시됩니다. 값이 양수이면 해당 평균은 대조군 평균에서 멀리 떨어져 LSD 보다 크므로 유의한 차이가 있는 것입니다.

단계별 개별 쌍 비교 (Newman-Keuls)

단계별 개별 쌍 비교(Newman-Keuls) 검정에서는 반복적인 단계별 절차를 사용하여 표본 평균을 비교합니다. 각 반복에서는 두 그룹 평균을 검정하기 위해 Tukey HSD 검정이 사용됩니다. 이 검정의 예는 "**단계별 개별 쌍 비교(Newman-Keuls) 검정의 예**"(200 페이지)에서 확인하십시오.

JMP 에서 J 그룹 평균을 검정하는 데 사용되는 절차는 다음과 같습니다.

다음을 정의합니다.

J = 그룹 수 (그룹 평균을 기준으로 오름차순 정렬됨)

N = 관측값 수

d = 자유도. $N - J$ 로 계산

i = 비교 대상 중 가장 작은 그룹 평균의 인덱스입니다.

j = 비교 대상 중 가장 큰 그룹 평균의 인덱스입니다.

k = 비교 대상 중 가장 큰 그룹 평균의 최소 그룹 인덱스입니다.

절차를 시작할 때 $i = 1, j = J$ 및 $k = 2$ 로 설정합니다.

1. 그룹 i 및 j 에 대해 Tukey HSD 검정을 수행합니다. 여기서 적절한 분위수를 찾기 위한 그룹 수는 $j - i + 1$ 입니다.
 - 검정이 유의하면 그룹 i 및 j 에 유의한 차이가 있는 것으로 결정됩니다. j 를 1만큼 줄입니다. 그 결과 j 가 k 보다 작아지면 i 를 1만큼 늘리고 $k = i + 1, j = J$ 로 설정한 다음 2 단계로 진행합니다.
 - 그 결과 j 가 k 보다 크거나 같아지면 2 단계로 진행합니다.
 - 검정이 유의하지 않으면 그룹 i 및 j 에 유의한 차이가 있는 것으로 결정되지 않습니다. i 를 1만큼 늘리고 $k = j - 1, j = J$ 로 설정한 다음 2 단계로 진행합니다.
2. k 값에 따라 절차를 계속할지 중지할지 결정합니다.
 - k 가 i 보다 크면 1 단계를 반복합니다.
 - k 가 i 보다 작거나 같으면 절차를 중지합니다. 검정되지 않은 나머지 모든 범위는 유의한 차이가 없는 것으로 간주됩니다.

Tukey HSD에 사용되는 분위수는 검정마다 다르며, 검정 대상 평균 사이에 있는 그룹 평균의 개수에 따라 달라집니다. "Newman-Keuls" 보고서에서 "고려되는 최소 분위수" ("최소 q^* "라는 라벨이 지정됨)는 위의 절차에 사용된 최소 스튜던트화 범위 분위수를 2의 제곱근으로 나눈 것입니다.

검정 결과는 "연결 문자 보고서"에 보고됩니다.

참고: 단계별 개별 쌍 비교 (Newman-Keuls) 검정을 사용할 때는 "비교 원" 그래프에 평균 원이 추가되지 않습니다. 이는 검정되는 두 평균 사이에 있는 평균의 개수에 따라 각 비교 방법마다 경계가 다르기 때문입니다. 이로 인해 각 원의 크기가 다르게 됩니다.

평균 비교 옵션

"평균 비교" 보고서에는 다음과 같은 옵션을 제공하는 빨간색 삼각형 메뉴가 포함됩니다.

차이 행렬 모든 평균 차이가 포함된 테이블을 표시하거나 숨깁니다.

신뢰 분위수 평균 비교 절차에 사용되는 임계값 및 유의 수준 (α)을 표시하거나 숨깁니다.

LSD 임계 행렬 (Newman-Keuls 검정에는 사용할 수 없음) 평균의 쌍별 차이에서 이러한 평균에 대한 최소 유의차를 뺀 행렬을 표시하거나 숨깁니다. 양수 값은 유의차가 있는 평균의 쌍을 나타냅니다.

연결 문자 보고서 (스튜던트 t , Tukey HSD 및 Newman-Keuls 검정에만 사용 가능) 기존의 문자로 코딩된 보고서를 표시하거나 숨깁니다. 이 보고서에서 문자를 공유하지 않는 평균에는 유의차가 있습니다.

정렬된 차이 보고서 (스튜던트 t 및 Tukey HSD 검정에만 사용 가능) 모든 쌍의 양수 쪽 차이, 차이의 표준 오차, 신뢰 구간 및 p 값과 함께 신뢰 구간이 중첩 표시된 차이 크기 그림을 표시하거나 숨깁니다. 해당 차이 막대가 완전히 포함되지 않은 신뢰 구간은 서로 유의한 차이가 없는 평균 쌍을 나타냅니다.

상세 비교 보고서 (스튜던트 t 검정에만 사용 가능) 각 비교에 대한 상세 보고서를 표시하거나 숨깁니다. 각 섹션에서는 수준, 표준 오차 및 신뢰 구간 간의 차이와 t 비, p 값 및 자유도를 보여 줍니다. 각 보고서의 오른쪽에는 비교를 보여 주는 그림이 표시됩니다.

비모수 검정

비모수 검정은 일반적인 분산 분석의 정규성 가정이 실용적이지 않을 때 유용합니다. "비모수" 옵션은 그룹 간에 평균 또는 중앙값이 동일하다는 가설을 검정하기 위한 몇 가지 방법을 제공합니다. "비모수 다중 비교" 절차는 쌍별 비교의 전체 오차율을 제어하는 데도 사용할 수 있습니다. 비모수 검정에는 순위 점수라고 하는 반응 순위의 함수가 사용됩니다. 자세한 내용은 Hajek 연구 자료 (1969) 및 SAS Institute Inc. 연구 자료 (2018a)에서 확인하십시오.

다음 사항에 유의하십시오 .

- Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정의 경우 X 요인의 수준이 세 개 이상이면 일원 검정에 대한 카이제곱 근사가 수행됩니다 .
- 블록 열을 지정하면 블록 평균을 사용하여 중심화된 데이터 값에 대해 비모수 검정 (Friedman 순위 검정 제외) 이 수행됩니다 .

Wilcoxon 검정 Wilcoxon 순위 점수를 기반으로 검정을 수행합니다 . Wilcoxon 순위 점수는 데이터의 단순 순위입니다 . Wilcoxon 검정은 로지스틱 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 요인의 수준이 세 개 이상이면 Kruskal-Wallis 검정이 수행됩니다 . 보고서에 대한 자세한 내용은 "[Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서](#)"(181 페이지)에서 확인하십시오 . 예는 "[비모수 Wilcoxon 검정의 예](#)"(202 페이지)에서 확인하십시오 .

Wilcoxon 검정은 Mann-Whitney 검정이라고도 합니다 .

중앙값 검정 중앙값 순위 점수를 기반으로 검정을 수행합니다 . 중앙값 순위 점수는 순위가 중앙값 순위보다 위인지 아래인지에 따라 0 또는 1 이 됩니다 . 중앙값 검정은 이중 지수 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 보고서에 대한 자세한 내용은 "[Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서](#)"(181 페이지)에서 확인하십시오 .

Van der Waerden 검정 Van der Waerden 순위 점수를 기반으로 검정을 수행합니다 . Van der Waerden 순위 점수는 정규 분포 함수의 역함수를 적용하여 정규 점수로 변환된 관측값의 개수에 1 을 더한 값으로 데이터의 순위를 나눈 값입니다 . Van der Waerden 검정은 정규 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 보고서에 대한 자세한 내용은 "[Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서](#)"(181 페이지)에서 확인하십시오 .

Kolmogorov Smirnov 검정 (X 요인의 수준이 두 개인 경우에만 사용 가능) 반응의 분포가 그룹 간에 동일한지 여부를 검정하는 경험적 분포 함수를 기반으로 검정을 수행합니다 . 근사 검정과 정확 검정이 모두 제공됩니다 . 보고서에 대한 자세한 내용은 "[Kolmogorov-Smirnov 2 표본 검정 보고서](#)"(182 페이지)에서 확인하십시오 .

Friedman 순위 검정 (시작 창에서 블록 변수를 지정한 경우에만 사용 가능) Friedman 순위 점수를 기반으로 검정을 수행합니다 . Friedman 순위 점수는 블록 변수의 각 수준 내에 있는 데이터의 순위입니다 . 이 검정의 모수적 버전은 반복 측정 ANOVA 입니다 . 보고서에 대한 자세한 내용은 "[Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서](#)"(181 페이지)에서 확인하십시오 .

참고 : 각 블록 내의 관측값 수는 동일해야 합니다 .

JMP PRO 정확 검정 정확한 Wilcoxon, 중앙값, van der Waerden 및 Kolmogorov-Smirnov 검정을 수행하기 위한 옵션을 제공합니다 . 이러한 옵션은 X 요인의 수준이 두 개인 경우에만 사용할 수 있습니다 . 근사 검정과 정확 검정 모두의 결과가 제공됩니다 .

보고서에 대한 자세한 내용은 "[2 표본: 정확 검정](#)"(181 페이지)에서 확인하십시오 . Wilcoxon 정확 검정의 관련 예는 "[비모수 Wilcoxon 검정의 예](#)"(202 페이지)에서 확인하십시오 .

Wilcoxon, 중앙값, Van der Waerden 및 Friedman 순위 검정 보고서

각 검정의 보고서에는 기술 통계량과 검정 결과가 표시됩니다. 검정 결과는 "일원 검정, 카이제곱 근사" 보고서에 표시되며, X 변수의 수준이 정확히 두 개뿐이면 "2 표본 검정, 정규 근사" 보고서도 표시됩니다. 표시되는 기술 통계량은 다음과 같습니다.

수준 X 의 수준입니다.

개수 각 수준의 빈도입니다.

점수 합 각 수준의 순위 점수 합입니다.

기대 점수 클래스 수준 간에 차이가 없다는 귀무가설하의 기대 점수입니다.

점수 평균 각 수준의 평균 순위 점수입니다.

(평균 - 평균 0)/ 표준 0 표준화된 점수입니다. 평균 0은 귀무가설하의 기대 평균 점수입니다. 표준 0은 귀무가설하의 기대 점수 합에 대한 표준편차입니다. 귀무가설은 그룹 간에 그룹 평균 또는 중앙값이 동일한 위치에 있다는 것입니다.

2 표본 검정, 정규 근사

X 의 수준이 정확히 두 개뿐이면 "2 표본 검정, 정규 근사" 보고서가 표시됩니다. 이 보고서에서는 다음을 제공합니다.

S 더 적은 수가 관측값이 있는 수준의 순위 점수 합을 제공합니다.

Z 정규 근사 검정에 대한 검정 통계량을 제공합니다. 자세한 내용은 "2 표본 정규 근사"(226 페이지)에서 확인하십시오.

Prob>|Z| 표준 정규 분포를 기준으로 정규 근사 검정의 p 값을 제공합니다.

일원 검정, 카이제곱 근사

이 보고서에서는 위치에 대한 카이제곱 검정의 결과를 제공합니다. 자세한 내용은 Conover 연구 자료 (1999)에서 확인하십시오.

카이제곱 카이제곱 검정 통계량의 값을 제공합니다. 자세한 내용은 "일원 카이제곱 근사"(227 페이지)에서 확인하십시오.

DF 검정의 자유도를 제공합니다.

Prob>ChiSq 검정의 p 값을 제공합니다. p 값은 자유도가 X 의 수준 수에서 1을 뺀 값에 해당하는 카이제곱 분포를 기준으로 합니다.

2 표본 : 정확 검정

데이터가 희소 데이터이거나 비대칭적이거나 동일 값이 많은 경우에는 점근적 동작을 기반으로 한 근사보다 정확 검정이 더 적합할 수 있습니다. X 의 수준이 정확히 두 개뿐인 경우 JMP Pro에서는 정확 검정에 대한 검정 통계량이 계산됩니다. **비모수 > 정확 검정**을 선택하고 원하는 검정

을 선택하십시오. 그러면 "2 표본: 정확 검정" 보고서가 표시됩니다. 이 보고서에서는 다음을 제공합니다.

S 더 작은 그룹의 관측값에 대한 순위 점수 합을 제공합니다. X 의 두 수준에 동일한 수의 관측값이 있으면 S 값은 값 순서화 시 X 의 마지막 수준에 대응합니다.

Prob $\leq S$ 검정의 단측 p 값을 제공합니다.

Prob $\geq |S - \text{Mean}|$ 검정의 양측 p 값을 제공합니다.

Kolmogorov-Smirnov 2 표본 검정 보고서

Kolmogorov-Smirnov 검정은 X 의 수준이 정확히 두 개뿐인 경우에만 사용할 수 있습니다. 이 보고서에는 기술 통계량과 검정 결과가 표시됩니다. 표시되는 기술 통계량은 다음과 같습니다.

수준 X 의 두 수준입니다.

개수 각 수준의 빈도입니다.

최대값에서의 EDF X 의 수준별로 두 EDF 간의 차이가 최대인 X 값 위치에서 해당 수준에 대한 EDF(경험적 누적 분포 함수)의 값을 제공합니다. "합계" 행에서는 두 EDF 간의 차이가 최대인 X 값 위치에서의 합동 EDF(전체 데이터 집합에 대한 EDF) 값을 제공합니다.

최대값에서 평균과의 편차 각 수준에 대해 다음과 같이 구한 값을 제공합니다.

- 해당 수준의 최대값에서의 EDF와 합동 데이터 집합(합계)의 최대값에서의 EDF 사이의 차이를 계산합니다.
- 이 차이를 해당 수준에 포함된 관측값 수("개수"로 제공됨)의 제곱근으로 곱합니다.

Kolmogorov-Smirnov 점근적 검정

이 보고서에서는 검정에 대한 상세 정보를 제공합니다.

KS 다음과 같이 계산된 Kolmogorov-Smirnov 통계량입니다.

$$KS = \max_j \sqrt{\frac{1}{n} \sum_i n_i (F_i(x_j) - F(x_j))^2}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

- $x_j(j=1, \dots, n)$ 은 관측값입니다.
- n_i 는 X 의 i 번째 수준에 포함된 관측값의 개수입니다.
- F 는 합동 경험적 누적 분포 함수입니다.
- F_i 는 X 의 i 번째 수준에 대한 경험적 누적 분포 함수입니다.

Kolmogorov-Smirnov 통계량의 이 버전은 X 의 수준이 세 개 이상인 경우에도 적용됩니다. 하지만 JMP에서는 X 의 수준이 두 개뿐일 때만 Kolmogorov-Smirnov 분석을 수행합니다.

KSa $KS\sqrt{n}$ 으로 계산된 점근적 Kolmogorov-Smirnov 통계량입니다. 여기서 n 은 총 관측값 수입니다.

$D = \max|F_1 - F_2|$ 두 수준의 EDF 간 최대 절대 편차입니다. 이 값은 두 표본을 비교하는 데 일반적으로 사용되는 버전의 Kolmogorov-Smirnov 통계량입니다.

Prob > D 검정의 p 값입니다. 이 값은 수준 간에 차이가 없다는 귀무가설하에 D 가 계산된 값을 초과할 확률입니다.

$D+ = \max(F_1 - F_2)$ 첫 번째 그룹의 수준이 두 번째 그룹의 수준을 초과한다는 대립가설에 대한 단측 검정 통계량입니다.

Prob > D+ $D+$ 검정의 p 값입니다.

$D- = \max(F_2 - F_1)$ 두 번째 그룹의 수준이 첫 번째 그룹의 수준을 초과한다는 대립가설에 대한 단측 검정 통계량입니다.

Prob > D- $D-$ 검정의 p 값입니다.

Kolmogorov-Smirnov 정확 검정

Kolmogorov-Smirnov 정확 검정에 대한 보고서에서는 점근적 검정과 동일한 통계량을 제공하지만 p 값은 정확하게 계산됩니다.

비모수 다중 비교

이 옵션은 비모수 다중 비교를 수행하기 위한 몇 가지 방법을 제공합니다. "Wilcoxon 개별 쌍 비교 검정" 을 제외하고 이러한 검정은 모두 순위를 기반으로 하며 전체 실험 오차율을 제어합니다. 이러한 검정에 대한 자세한 내용은 Dunn 연구 자료 (1964) 및 Hsu 연구 자료 (1996) 에서 확인하십시오. 보고서에 대한 자세한 내용은 "[비모수 다중 비교 절차](#)"(183 페이지) 에서 확인하십시오.

비모수 다중 비교 절차

Wilcoxon 개별 쌍 비교 각 쌍에 대한 Wilcoxon 검정을 수행합니다. 이 절차에서는 전체 유의 수준을 제어하지 않습니다. 이 값은 "평균 비교" 메뉴에 있는 **개별 쌍 비교 (스튜던트 t)** 옵션의 비모수적 버전입니다. 자세한 내용은 "[Wilcoxon 개별 쌍 비교](#), [Steel-Dwass 전체 쌍 비교](#) 및 [Steel 대조군 비교](#)"(184 페이지) 에서 확인하십시오.

Steel-Dwass 전체 쌍 비교 개별 쌍에 대해 Steel-Dwass 검정을 수행합니다. 이 검정은 "평균 비교" 메뉴에 있는 **전체 쌍 비교 (Tukey HSD)** 옵션의 비모수적 버전입니다. 자세한 내용은 "[Wilcoxon 개별 쌍 비교](#), [Steel-Dwass 전체 쌍 비교](#) 및 [Steel 대조군 비교](#)"(184 페이지) 에서 확인하십시오.

Steel 대조군 비교 각 수준을 대조 수준과 비교합니다. 이 검정은 "평균 비교" 메뉴에 있는 **대조군 비교 (Dunnett)** 옵션의 비모수적 버전입니다. 자세한 내용은 "[Wilcoxon 개별 쌍 비교](#), [Steel-Dwass 전체 쌍 비교](#) 및 [Steel 대조군 비교](#)"(184 페이지) 에서 확인하십시오.

Dunnett 전체 쌍 비교 (결합 순위 기준) "Steel-Dwass 전체 쌍 비교" 옵션과 유사하게 각 쌍에 대한 비교를 수행합니다. Dunnett 방법은 비교 대상인 쌍뿐만 아니라 모든 데이터에 대한 순위를 계산합니다. 보고되는 p 값은 Bonferroni 조정을 반영합니다. 이 값은 조정되지 않은 p 값에 비교 개수를 곱한 것입니다. 조정된 p 값이 1 을 초과할 경우 1 로 보고됩니다. 자세한 내

용은 "Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준)"(185 페이지)에서 확인하십시오.

Dunnett 대조군 비교 (결합 순위 기준) "Steel 대조군 비교" 옵션과 유사하게 각 수준을 대조 수준과 비교합니다. Dunnett 방법은 비교 대상인 쌍뿐만 아니라 모든 데이터에 대한 순위를 계산합니다. 보고되는 p 값은 Bonferroni 조정을 반영합니다. 이 값은 조정되지 않은 p 값에 비교 개수를 곱한 것입니다. 조정된 p 값이 1 을 초과할 경우 1 로 보고됩니다. 자세한 내용은 "Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준)"(185 페이지)에서 확인하십시오.

Wilcoxon 개별 쌍 비교, Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교

이러한 다중 비교 절차에 대한 보고서에서는 검정 결과와 신뢰 구간을 제공합니다. 이러한 검정의 경우 해당 비교에 사용된 두 수준만 결합하여 구한 표본 내에서 관측값의 순위가 매겨집니다.

q* 신뢰 구간을 계산하는 데 사용된 분위수입니다.

알파 신뢰 구간을 계산하는 데 사용된 유의 수준입니다. "일원 분석" 메뉴에서 " α 수준 설정" 옵션을 선택하여 신뢰 수준을 변경할 수 있습니다.

수준 쌍별 비교에서 사용된 X 변수의 첫 번째 수준입니다.

- 수준 쌍별 비교에서 사용된 X 변수의 두 번째 수준입니다.

점수 평균 차이 첫 번째 수준 ("수준")에 포함된 관측값의 순위 점수 평균에서 두 번째 수준 ("-수준")에 포함된 관측값의 순위 점수 평균을 뺀 값입니다. 이때 연속성 수정이 적용됩니다.

첫 번째 수준의 관측값 수는 n_1 로 나타내고 두 번째 수준의 관측값 수는 n_2 로 나타냅니다. 이러한 두 수준으로 구성된 표본 내에서 관측값의 순위가 매겨집니다. 동일 순위에 대해서는 평균을 구합니다. 첫 번째 수준의 순위 합은 ScoreSum_1 으로 나타내고 두 번째 수준의 순위 합은 ScoreSum_2 로 나타냅니다.

평균 점수의 차이가 양수인 경우 "점수 평균 차이"는 다음과 같이 제공됩니다.

$$\text{점수 평균 차이} = (\text{ScoreSum}_1 - 0.5) / n_1 - (\text{ScoreSum}_2 + 0.5) / n_2$$

평균 점수의 차이가 음수인 경우 "점수 평균 차이"는 다음과 같이 제공됩니다.

$$\text{점수 평균 차이} = (\text{ScoreSum}_1 + 0.5) / n_1 - (\text{ScoreSum}_2 - 0.5) / n_2$$

차이 표준 오차 점수 평균 차이의 표준 오차입니다.

Z 평균에 차이가 없다는 귀무가설하에서 점근적 표준 정규 분포를 따르는 표준화된 검정 통계량입니다.

p 값 Z를 기반으로 한 점근적 검정의 p 값입니다.

Hodges-Lehmann 위치 이동에 대한 Hodges-Lehmann 추정량입니다. 첫 번째 수준에 포함된 관측값에서 두 번째 수준에 포함된 관측값을 빼서 구한 모든 쌍체 차이가 생성됩니다. Hodges-Lehmann 추정량은 이러한 차이의 중앙값입니다. 차이 그림 막대 차트에는 Hodges-Lehmann 추정값의 크기가 표시됩니다.

CL 하한 Hodges-Lehmann 통계량의 신뢰 하한입니다.

참고 : 그룹 표본 크기가 메모리 문제를 일으킬 만큼 큰 경우에는 계산되지 않습니다.

CL 상한 Hodges-Lehmann 통계량의 신뢰 상한입니다.

참고 : 그룹 표본 크기가 메모리 문제를 일으킬 만큼 큰 경우에는 계산되지 않습니다.

Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준)

이러한 비교는 전체 데이터 집합의 관측값 순위를 기반으로 합니다. "Dunnett 대조군 비교 (결합 순위 기준)" 검정의 경우 대조 수준을 선택해야 합니다.

수준 쌍별 비교에서 사용된 X 변수의 첫 번째 수준입니다.

- 수준 쌍별 비교에서 사용된 X 변수의 두 번째 수준입니다.

점수 평균 차이 첫 번째 수준 ("수준")에 포함된 관측값의 순위 점수 평균에서 두 번째 수준 ("-수준")에 포함된 관측값의 순위 점수 평균을 뺀 값입니다. 이때 연속성 수정이 적용됩니다. 순위는 전체 표본 내의 관측값에 순위를 매기는 방식으로 결정됩니다. 동일 순위에 대해서는 평균을 구합니다. 연속성 수정에 대한 자세한 내용은 "[점수 평균 차이](#)"(184 페이지)에서 확인하십시오.

차이 표준 오차 점수 평균 차이의 표준 오차입니다.

Z 평균에 차이가 없다는 귀무가설하에서 점근적 표준 정규 분포를 따르는 표준화된 검정 통계량입니다.

p 값 Z를 기반으로 한 점근적 검정의 p 값입니다.

이분산

그룹 간에 분산이 동일하지 않으면 분산 분석의 일반적인 가정이 충족되지 않습니다. 따라서 ANOVA F-검정이 유효하지 않습니다. JMP에서는 그룹 분산의 동일 여부에 대한 4가지 검정과 그룹 모집단 분산이 동일하지 않을 때 유효한 ANOVA를 제공합니다. 처음 세 가지 등분산 검정의 기본 개념은 각 그룹의 퍼짐 정도를 측정하기 위해 구성된 새 반응 변수에 대해 분산 분석을 수행하는 것입니다. 네 번째 검정은 정규 분포하에서 우도비 검정과 유사한 Bartlett 검정입니다.

참고 : 이분산을 검정하는 또 다른 방법은 ANOMV입니다. 자세한 내용은 "[평균 분석 방법](#)"(170 페이지)에서 확인하십시오.

다음과 같은 등분산 검정을 사용할 수 있습니다.

O'Brien 새 변수의 그룹 평균이 원래 반응의 그룹 표본 분산과 동일하도록 종속 변수를 생성합니다. O'Brien 변수에 대한 ANOVA는 실제로는 그룹 표본 분산에 대한 ANOVA입니다(O'Brien 1979, Olejnik and Algina 1987).

Brown-Forsythe 반응이 각 관측값과 그룹 중앙값의 차이에 대한 절대값인 경우 ANOVA 에서 구한 F -검정을 표시합니다 (Brown and Forsythe 1974).

Levene 반응이 각 관측값과 그룹 중앙값의 차이에 대한 절대값인 경우 ANOVA 에서 구한 F -검정을 표시합니다 (Levene 1960). 퍼짐 정도는 SAS 기본 값인 $z_{ij}^2 = (y_{ij} - \bar{y}_i)^2$ 대신 $z_{ij} = |y_{ij} - \bar{y}_i|$ 으로 측정됩니다.

Bartlett 표본 분산의 가중 산술평균을 표본 분산의 가중 기하평균과 비교합니다. 기하평균은 항상 산술평균보다 작거나 같으며 모든 표본 분산이 동일한 경우에만 이 둘이 동일하게 유지됩니다. 그룹 분산 간의 변동이 클수록 이 두 평균의 차이가 커집니다. 이 두 평균으로부터 χ^2 분포 (사실상 특정 공식에서는 F 분포) 에 근사한 값을 산출하는 함수가 생성됩니다. 값이 크면 산술 또는 기하 비율의 값이 큰 것이므로 그룹 분산의 변동 폭이 커집니다. Bartlett 카이 제곱 검정 통계량을 자유도로 나누면 테이블에 표시된 F 값이 됩니다. Bartlett 검정은 정규성 가정 위배에 대해 그다지 로버스트하지 않습니다 (Bartlett and Kendall 1946).

F-검정 양측 (X 변수의 수준이 두 개인 경우에만 사용 가능) 검정 대상 그룹이 두 개뿐이면 이분산에 대한 표준 F -검정도 수행됩니다. F -검정은 큰 분산 추정값 대 작은 분산 추정값의 비율입니다. F 분포의 p 값은 양측 검정을 구성하기 위해 두 배가 됩니다.

참고 : 블록 열을 지정한 경우에는 블록 평균에 맞게 데이터가 조정된 후에 데이터에 대해 분산 검정이 수행됩니다.

자세한 내용은 "[이분산 옵션의 예](#)" (205 페이지) 에서 확인하십시오.

분산 동일 여부 검정 보고서

"분산 동일 여부 검정" 보고서에는 그룹 평균과 총 평균 및 중앙값 사이의 차이가 표시되고 검정 절차에 대한 요약이 제공됩니다.

등분산 검정 결과 그룹 분산에 유의한 차이가 있는 것으로 나타나면 일반적인 ANOVA 검정 대신 Welch 검정을 사용하십시오. Welch 통계량은 일반적인 ANOVA F -검정을 기반으로 합니다. 하지만 그룹 평균 분산의 역수로 평균에 가중치가 적용된다는 차이점이 있습니다 (Welch 1951, Brown and Forsythe 1974, Asiribo and Gurland 1990). 수준이 두 개인 경우 Welch ANOVA 는 이분산 t -검정과 동등합니다.

분산 동일 여부 검정 보고서 설명

수준 데이터에서 나타나는 요인 수준을 나열합니다.

개수 각 수준의 빈도를 기록합니다.

표준편차 각 요인 수준의 반응 표준편차를 기록합니다. 이 표준편차는 O'Brien 변수의 평균과 동일합니다. 데이터에서 한 수준이 한 번만 나타나는 경우에는 표준편차가 계산되지 않습니다.

평균에 대한 평균 절대 차이 반응 및 그룹 평균의 평균 절대 차이를 기록합니다. 이 평균 절대 차이는 Levene 변수의 그룹 평균과 동일합니다.

중앙값에 대한 평균 절대 차이 반응 및 그룹 중앙값의 절대 차이를 기록합니다. 이 평균 절대 차이는 Brown-Forsythe 변수의 그룹 평균과 동일합니다.

검정 수행된 검정의 이름을 나열합니다.

F 비 각 검정에 대해 계산된 F 통계량을 기록합니다. 자세한 내용은 "[분산 동일 여부 검정](#)"(224 페이지)에서 확인하십시오.

DFNum 각 검정의 분자 자유도를 기록합니다. 요인에 k 개의 수준이 있는 경우 분자의 자유도는 $k - 1$ 입니다. 데이터에 한 번만 나타나는 수준은 O'Brien, Brown-Forsythe 또는 Levene의 검정 통계량을 계산하는 데 사용되지 않습니다. 이 경우 분자 자유도는 계산에 사용된 수준 수에서 1을 뺀 값입니다.

DFDen 각 검정의 분모에 사용된 자유도를 기록합니다. O'Brien, Brown-Forsythe 및 Levene의 경우 검정 통계량을 계산하는 데 사용된 각 요인 수준에 대해 자유도를 뺍니다. 요인에 k 개의 수준이 있는 경우 분모 자유도는 $n - k$ 입니다.

p 값 모든 수준의 분산이 동일한 경우 F 비 값이 계산된 값보다 클 확률입니다.

참고 : X 변수에 관측값 수가 5 개 미만인 수준이 있으면 경고가 나타납니다. 표본 크기가 작을 때 위 검정의 성능에 대한 자세한 내용은 Brown and Forsythe 연구 자료(1974) 및 Miller 연구 자료(1972)에서 확인하십시오.

Welch 검정 보고서 설명

F 비 평균 동등성 검정에 대한 F -검정 통계량을 표시합니다.

DFNum 검정의 분자 자유도를 기록합니다. 요인에 k 개의 수준이 있는 경우 분자의 자유도는 $k - 1$ 입니다. 데이터에 한 번만 나타나는 수준은 Welch ANOVA를 계산하는 데 사용되지 않습니다. 이 경우 분자 자유도는 계산에 사용된 수준 수에서 1을 뺀 값입니다.

DFDen 검정의 분모 자유도를 기록합니다. 자세한 내용은 "[분산 동일 여부 검정](#)"(224 페이지)에서 확인하십시오.

Prob>F 모든 수준의 평균이 동일한 경우 F 값이 계산된 값보다 클 확률입니다. 관측된 유의 확률이 0.05 이하이면 수준 간에 평균이 동일하지 않다는 증거로 간주됩니다.

t-검정 F 비와 t -검정 사이의 관계를 보여 줍니다. F 비의 제곱근으로 계산됩니다. X 요인의 수준이 두 개인 경우에만 표시됩니다.

동등성 검정

동등성 검정에서는 평균에 실제적 차이가 있는지 여부를 평가합니다. 임계 차이를 선택해야 하며 그보다 작은 차이는 실제적으로 동등한 것으로 간주됩니다. 가장 간단하고 직접적인 검정에서는 차이 구간의 양쪽에서 두 개의 단측 t -검정이 사용됩니다. 두 검정이 모두 기각되거나 평균의 차이가 임계와 유의하게 다른 것으로 결론나면 해당 그룹은 실제적으로 동등한 것입니다. **동**

등성 검정 옵션은 TOST(Two One-Sided Tests: 두 개의 단측 검정) 방법을 사용합니다. 자세한 내용은 "[동등성 검정의 예](#)"(206 페이지)에서 확인하십시오.

로버스트

이상치는 부정확한 추정 및 결정을 초래할 수 있습니다. **로버스트** 옵션은 데이터 집합에서 이상치의 영향을 줄이기 위한 두 가지 방법으로 "로버스트 적합"과 "Cauchy 적합"을 제공합니다.

로버스트 적합

"로버스트 적합" 옵션은 반응 변수에서 이상치의 영향을 줄입니다. Huber M- 추정 방법이 사용됩니다. Huber M- 추정은 다음과 같이 Huber 손실 함수를 최소화하는 모수 추정값을 찾는 방법입니다.

$$l(\boldsymbol{\ell}) = \sum_i \rho(e_i)$$

다음은 각 요소에 대한 설명입니다.

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & \text{단, } |e| < k \\ k|e| - \frac{1}{2}k^2 & \text{단, } |e| \geq k \end{cases}$$

e_i 는 잔차를 나타냅니다.

Huber 손실 함수는 이상치에 벌점을 부과하며, 작은 오차의 경우에는 2 차 형태로 증가하고 큰 오차의 경우에는 선형적으로 증가합니다. 로버스트 적합에 대한 자세한 내용은 Huber 연구 자료 (1973) 및 Huber and Ronchetti 연구 자료 (2009)에서 확인하십시오. 자세한 내용은 "[로버스트 적합 옵션의 예](#)"(207 페이지)에서 확인하십시오.

Cauchy 적합

Cauchy 적합 옵션은 오차가 Cauchy 분포를 따른다고 가정합니다. Cauchy 분포는 정규 분포보다 꼬리가 더 두꺼우므로 이상치의 영향이 줄어듭니다. 이 옵션은 데이터에 포함된 이상치의 비율이 높은 경우에 유용할 수 있습니다. 하지만 데이터가 정규 분포에 가깝고 이상치가 일부만 있는 경우에 이 옵션을 사용하면 추론이 잘못될 수 있습니다. Cauchy 옵션은 최대 우도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다.

검정력

검정력 옵션은 통계 검정력을 계산하고 지정된 가설 검정에 대한 기타 상세 정보를 제공합니다. 자세한 내용은 "**검정력 옵션의 예**"(209 페이지)에서 확인하십시오. 통계 상세 정보는 "**검정력**"(223 페이지)에서 확인하십시오.

- **LSV**(최소 유의값)는 특정 p 값이 알파가 되는 모수 또는 모수 함수의 값입니다. 즉, 어떤 p 값이 알파가 될 때 효과의 유의성이 얼마나 작게 선언되는지 알 수 있습니다. LSV는 확률 척도가 아니라 모수 척도의 유의성에 대한 측정 수단을 제공합니다. 이를 통해 설계와 데이터의 민감도를 보여 줍니다.
- **LSN**(최소 유의 개수)은 데이터의 형식이 동일한 경우 지정된 p 값 알파를 산출할 관측값의 총 개수입니다. LSN은 알파, 시그마 및 델타 값(유의 수준, 오차의 표준편차 및 효과 크기)이 주어진 상태에서 유의한 결과를 얻을 수 있을 정도로 추정값의 분산을 줄이는 데 필요한 관측값의 개수로 정의됩니다. 유의성을 달성하기 위해 더 많은 데이터가 필요한 경우 LSN을 통해 필요한 데이터 양을 알 수 있습니다. LSN은 검정력이 약 50%가 되도록 하는 총 관측값 수입니다.
- 검정력은 그룹 간에 실제 차이가 존재할 때 유의성을 달성(p 값 < 알파) 할 확률입니다. 이는 표본 크기, 효과 크기, 오차의 표준편차 및 유의 수준에 따라 달라집니다. 검정력은 특정 유의 수준에서 실험을 통해 차이(효과 크기)를 발견할 가능성을 나타냅니다.

참고: 일원 배치의 그룹이 두 개뿐인 경우 검정력에 의해 계산된 LSV는 다중 비교 테이블에 표시된 LSD(최소 유의차)와 동일합니다.

검정력 상세 정보 창 및 보고서

검정력 상세 정보 창 및 보고서는 모형 적합 플랫폼을 통해 시작된 일반 적합 플랫폼에서와 동일합니다. 검정력 계산에 대한 자세한 내용은 **Fitting Linear Models**의 부록 "Statistical Details"에서 확인하십시오.

알파, 시그마, 델타 및 개수를 나타내는 4개의 각 열에 단일 값, 두 개의 값, 또는 값 시퀀스의 시작, 끝 및 증분 값을 입력해야 합니다(그림 6.32). 지정한 값의 가능한 모든 조합에 대해 검정력 계산이 수행됩니다.

알파 (α) 유의 수준으로, 0에서 1 사이입니다(일반적으로 0.05, 0.01 또는 0.10). 기본적으로 표시되는 값은 0.05입니다.

시그마 (σ) 모형의 잔차에 대한 표준 오차입니다. 기본적으로 제공되는 값은 RMSE(제공된 평균 제공 오차)에서 구한 추정값입니다.

델타 (δ) 원시 효과 크기입니다. 효과 크기 계산에 대한 자세한 내용은 **Fitting Linear Models**의 "Standard Least Squares Report and Options"장에서 확인하십시오. 첫 번째 위치는 기본적으로 가설의 제공된 제공함을 n 으로 나눈 값(즉, $\delta = \sqrt{SS/n}$)으로 설정되어 있습니다.

개수 (n) 모든 그룹의 총 표본 크기입니다. 첫 번째 위치에는 기본적으로 실제 표본 크기가 입력되어 있습니다.

검정력 계산 네 개의 값, 즉 α , σ , δ 및 n 을 모두 사용하여 검정력 (유의한 결과를 얻을 확률) 을 계산합니다.

최소 유의 개수 계산 주어진 α , σ 및 δ 로 약 50% 의 검정력을 달성하는 데 필요한 관측값 수를 계산합니다.

최소 유의값 계산 p 값이 α 가 되는 모수 또는 선형 적합의 값을 계산합니다. 계산에는 α , σ , n , 그리고 추정값의 표준 오차가 사용됩니다. 이 기능은 X 요인의 수준이 두 개일 때만 사용할 수 있으며 일반적으로 개별 모수에 사용됩니다.

조정된 검정력 및 신뢰 구간 후향적으로 검정력을 분석할 경우에는 표준 오차와 검정 모수의 추정값을 사용합니다.

- 조정된 검정력은 비중심성 모수의 보다 비편향적인 추정값으로부터 계산된 검정력입니다.
- 조정된 검정력의 신뢰 구간은 비중심성 추정값의 신뢰 구간을 기반으로 합니다.

랜덤 변동은 원래 델타에서 발생하므로 원래 델타에 대해서만 조정된 검정력 및 신뢰 한계가 계산됩니다.

정규 분위수 그림

다음과 같은 두 가지 유형의 정규 분위수 그림을 생성할 수 있습니다.

- **분위수별 실제값 그림**: 반응 값 대 정규 분위수 값 그림을 생성합니다. X 변수의 각 수준에 대해 개별적으로 분위수가 계산되어 그림에 표시됩니다.
- **실제값별 분위수 그림**: 정규 분위수 값 대 반응 값 그림을 생성합니다. X 변수의 각 수준에 대해 개별적으로 분위수가 계산되어 그림에 표시됩니다.

적합선 옵션은 분위수 그림에 적합선을 표시하거나 숨깁니다.

자세한 내용은 "[정규 분위수 그림의 예](#)"(210 페이지) 에서 확인하십시오.

CDF 그림

CDF 그림에서는 일원 분석 보고서의 모든 그룹에 대한 누적 분포 함수를 보여 줍니다. CDF 그림은 X 요인의 여러 수준 간에 반응 분포를 비교하려는 경우에 유용합니다. 자세한 내용은 "[CDF 그림의 예](#)"(211 페이지) 에서 확인하십시오.

밀도

밀도 옵션은 X 요인의 여러 수준 간에 반응의 분포 및 합성을 비교하는 몇 가지 방법을 제공합니다. 다음과 같은 세 가지 밀도 옵션이 있습니다.

- **밀도 비교**: 각 그룹의 밀도 추정값을 나타내는 평활 곡선을 표시합니다. 평활 곡선은 각 그룹의 밀도 추정값입니다.
- **밀도의 합성**: 각 그룹의 반응 개수에 따라 가중치가 적용된 밀도 요약 표시합니다. "밀도의 합성" 그림에서는 X 값별로 각 그룹이 전체에 얼마나 기여하는지를 보여 줍니다.
- **밀도의 비율**: 각 X 수준에서 그룹이 합계에 기여하는 정도를 전체에 대한 비율로 보여 줍니다.

자세한 내용은 "[밀도 옵션의 예](#)"(212 페이지)에서 확인하십시오.

매칭 열

매칭 모형 분석을 위해 매칭 (ID) 변수를 지정하려면 **매칭 열** 옵션을 사용합니다. **매칭 열** 옵션은 일원 분석에 사용되는 데이터가 매칭 (쌍체) 데이터인 경우를 처리합니다. 동일한 참가자로부터 서로 다른 그룹의 관측값을 얻은 경우에 매칭 데이터가 발생할 수 있습니다. 자세한 내용은 "[매칭 열 옵션의 예](#)"(213 페이지)에서 확인하십시오.

참고: 매칭의 특수한 경우에는 쌍체 t -검정을 수행합니다. **매칭 쌍** 플랫폼에서는 이 유형의 데이터를 처리하지만 데이터가 서로 다른 행이 아니라 서로 다른 열에 쌍으로 구성되어 있어야 합니다.

매칭 열 옵션은 다음과 같은 두 가지 기본 작업을 수행합니다.

- 반복적 비례 적합 알고리즘을 사용하여 그룹화 변수 (X로 Y 적합 분석의 X 변수)와 선택한 매칭 변수를 모두 포함하는 가법 모형을 적합시킵니다. 이와 동등한 선형 모형은 매우 느리고 방대한 메모리 리소스를 필요로 하므로 참가자가 수백 명인 경우에는 반복적 비례 적합 알고리즘이 효과를 발휘합니다.
- 그룹 간에 매칭되는 점 사이에 선을 그립니다. 매칭 ID 값이 동일한 관측값이 여러 개이면 관측값 그룹의 평균에서부터 선이 그려집니다.

매칭 열 옵션은 매칭되는 점들을 연결하는 **매칭 선** 옵션을 자동으로 활성화합니다. 이 선을 해제하려면 **표시 옵션 > 매칭 선**을 선택합니다.

"매칭 적합" 보고서에는 F -검정을 사용한 효과가 표시됩니다. 이는 교호작용 항이 있는 모형과 없는 모형 두 개를 실행하는 경우 모형 적합 플랫폼에서 얻을 수 있는 검정과 동등합니다. 수준이 두 개뿐인 경우 F -검정은 쌍체 t -검정과 동등합니다.

참고: 모형 적합 플랫폼에 대한 자세한 내용은 Fitting Linear Models의 "Model Specification" 장에서 확인하십시오.

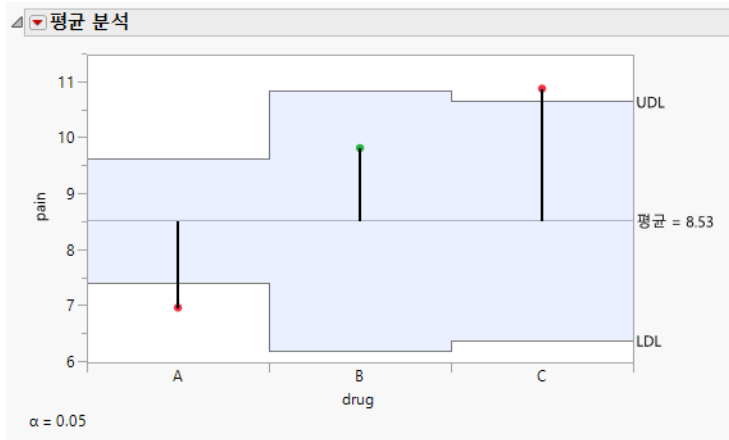
일원 분석 플랫폼의 추가 예

- " 평균 분석 차트의 예 "
- " 분산에 대한 평균 분석 차트의 예 "
- " 개별 쌍 비교 (스튜던트 t) 검정의 예 "
- " 전체 쌍 비교 (Tukey HSD) 검정의 예 "
- " 최량 비교 (Hsu MCB) 검정의 예 "
- " 대조군 비교 (Dunnett) 의 예 "
- " 단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예 "
- " 예 : 4 가지 평균 비교 검정 대조 "
- " 비모수 Wilcoxon 검정의 예 "
- " 이분산 옵션의 예 "
- " 동등성 검정의 예 "
- " 로버스트 적합 옵션의 예 "
- " 검정력 옵션의 예 "
- " 정규 분위수 그림의 예 "
- " CDF 그림의 예 "
- " 밀도 옵션의 예 "
- " 매칭 열 옵션의 예 "
- " 예 : 일원 분석을 위한 데이터 쌓기 "

평균 분석 차트의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Analgesics.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. pain 을 선택하고 **Y, 반응**을 클릭합니다 .
4. drug 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 평균 분석 " 의 빨간색 삼각형을 클릭하고 **평균 분석 방법 > ANOM** 을 선택합니다 .

그림 6.13 평균 분석 차트의 예



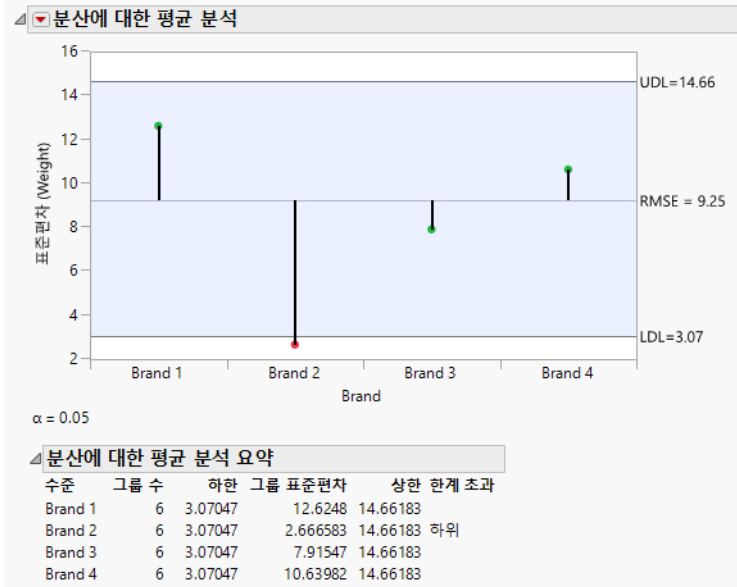
예를 들어 약물 A와 C의 평균은 전체 평균과 유의하게 다릅니다. 약물 A의 평균은 더 낮고 약물 C의 평균은 더 높습니다. 표본 크기가 다르기 때문에 약물 유형에 대한 결정 한계는 동일하지 않습니다.

분산에 대한 평균 분석 차트의 예

이 예에서는 Spring Data.jmp 샘플 데이터 테이블을 사용합니다. 스프링 길이를 0.10인치 늘리기 위해 필요한 중량을 확인하기 위해 4가지 브랜드의 스프링을 테스트했습니다. 각 브랜드에서 6개의 스프링을 테스트했습니다. 또한 ANOMV 검정은 비정규성에 대해 로버스트하지 않으므로 데이터의 정규성을 확인했습니다. 이제 각 브랜드를 검토하여 브랜드 간에 변동성이 유의하게 다른지 여부를 확인하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Spring Data.jmp를 엽니다.
2. **분석 > X로 Y 적합**을 선택합니다.
3. Weight를 선택하고 **Y, 반응**을 클릭합니다.
4. Brand를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "Weight 대 Brand의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 분석 방법 > 분산에 대한 ANOM**을 선택합니다.
7. "분산에 대한 평균 분석" 옆의 빨간색 삼각형을 클릭하고 **요약 보고서 표시**를 선택합니다.

그림 6.14 분산에 대한 평균 분석 차트의 예



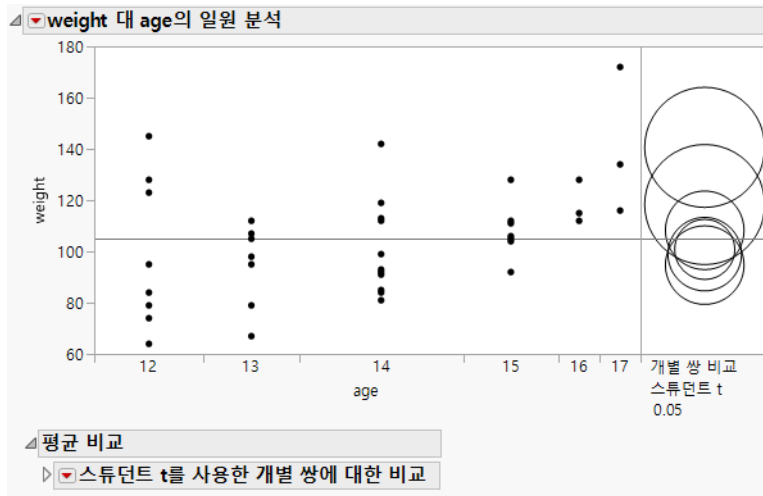
Brand 2의 표준편차는 결정 하한을 넘습니다. 따라서 Brand 2는 다른 브랜드보다 분산이 유의하게 낮습니다.

개별 쌍 비교 (스튜던트 t) 검정의 예

이 예에서는 Big Class.jmp 샘플 데이터 테이블을 사용합니다. 연령별 체중의 일원 배치를 표시하고, 가능한 모든 t -검정을 나타내는 비교 원을 사용하여 그룹 비교를 표시하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp를 엽니다.
2. **분석 > X로 Y 적합**을 선택합니다.
3. weight를 선택하고 **Y, 반응**을 클릭합니다.
4. age를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "weight 대 age의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 개별 쌍 비교 (스튜던트 t)**를 선택합니다.

그림 6.15 개별 쌍 비교 (스튜던트 t) 비교 원의 예



이 평균 비교 방법은 평균의 실제 차이가 유의한 것으로 간주될 차이보다 큰지 확인하는 방법이라고 할 수 있습니다. 이 차이를 **LSD**(최소 유의차)라고 합니다. **LSD**라는 용어는 스튜던트 t 구간과 다른 검정의 구간 관련 문맥에서 사용됩니다. 비교 원 그래프에서 원 중심 간의 거리는 실제 차이를 나타냅니다. **LSD**는 원이 직각으로 교차한 경우의 거리입니다.

그림 6.16 개별 쌍 비교 (스튜던트 t)에 대한 평균 비교 보고서의 예

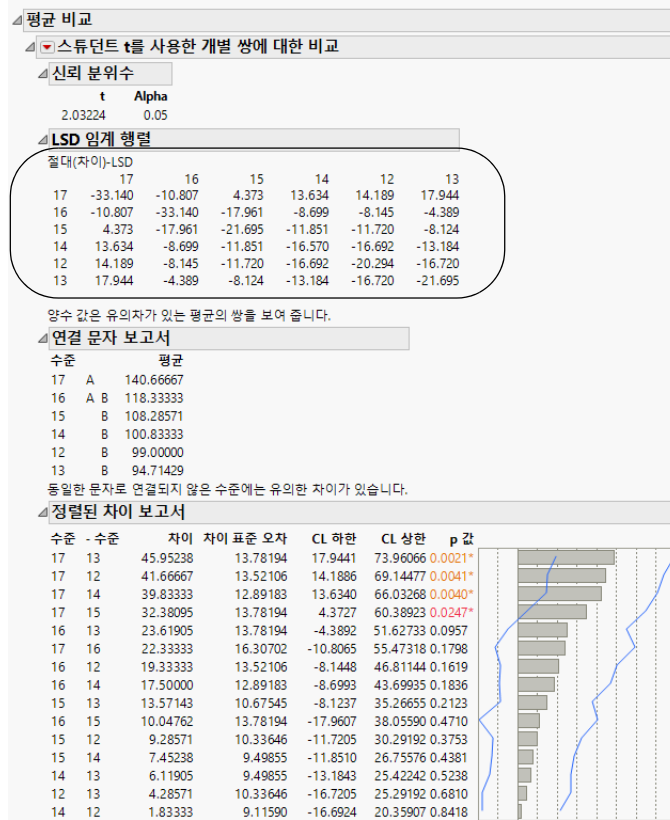


그림 6.16의 LSD 임계 테이블에서는 평균의 절대 차이와 LSD(최소 유의차) 간의 차이를 보여 줍니다. 값이 양수이면 두 평균의 차이가 LSD보다 크고 두 그룹이 유의하게 다른 것입니다.

전체 쌍 비교 (Tukey HSD) 검정의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Big Class.jmp를 엽니다.
2. 분석 > X로 Y 적합을 선택합니다.
3. weight를 선택하고 Y, 반응을 클릭합니다.
4. age를 선택하고 X, 요인을 클릭합니다.
5. 확인을 클릭합니다.
6. "weight 대 age의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 평균 비교 > 전체 쌍 비교(Tukey HSD)를 선택합니다.

그림 6.17 전체 쌍 비교 (Tukey HSD) 비교 원의 예

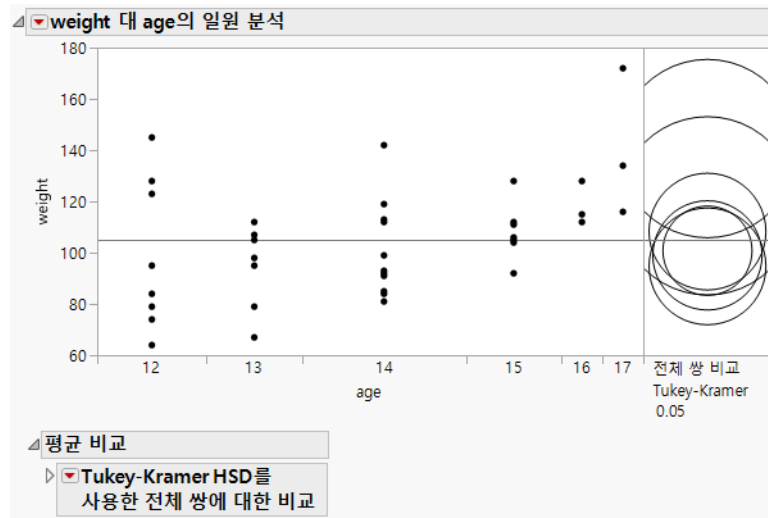


그림 6.18 전체 쌍 비교 (Tukey HSD) 에 대한 평균 비교 보고서의 예

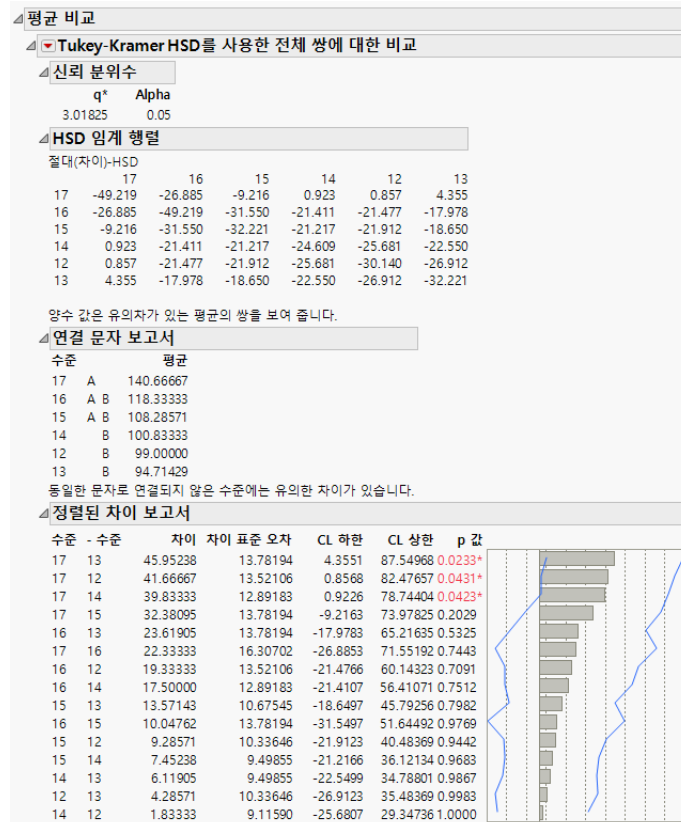
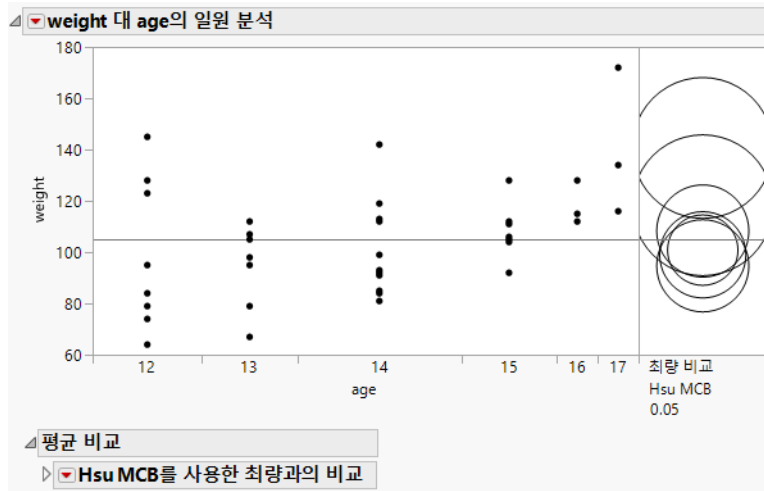


그림 6.18의 Tukey-Kramer HSD 임계 행렬에서는 평균의 실제 절대 차이에서 HSD를 뺀 값을 보여 줍니다. 이 값은 유의하게 간주될 차이를 나타냅니다. 양수 값을 가진 쌍은 유의하게 다른 것입니다. $q^*(\text{HSD 임계 행렬 테이블 위에 표시})$ 는 HSD를 척도화하는 데 사용된 분위수입니다. 이 값은 스튜던트 t 와 유사한 계산 역할을 합니다.

최량 비교 (Hsu MCB) 검정의 예

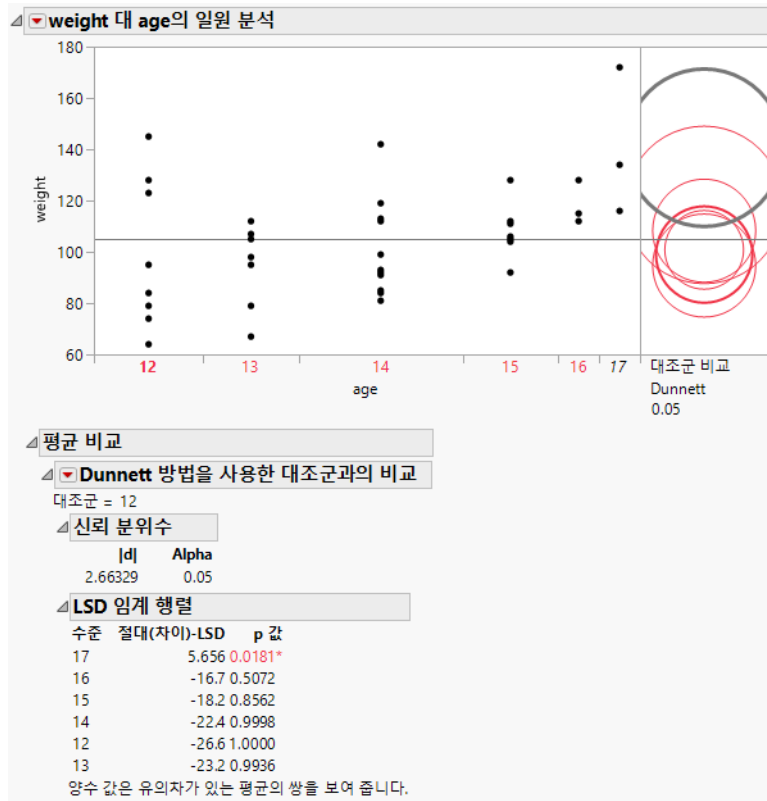
1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Big Class.jmp를 엽니다.
2. 분석 > X로 Y 적합을 선택합니다.
3. weight를 선택하고 Y, 반응을 클릭합니다.
4. age를 선택하고 X, 요인을 클릭합니다.
5. 확인을 클릭합니다.
6. "weight 대 age의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 평균 비교 > 최량 비교 (Hsu MCB)를 선택합니다.

그림 6.19 최량 비교 (Hsu MCB) 비교 원의 예



6. "weight 대 age의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 대조군 비교 (Dunnett)** 를 선택합니다.
7. 대조군으로 사용할 그룹을 선택합니다. 이 예에서는 12 세를 선택합니다.
또는 산점도에서 행을 클릭하여 강조 표시한 후 **평균 비교 > 대조군 비교 (Dunnett)** 옵션을 선택합니다. 이 검정에서는 선택된 행이 대조군으로 사용됩니다.
8. **확인**을 클릭합니다.

그림 6.21 대조군 비교 (Dunnett) 비교 원의 예



이 비교 원을 사용하면 수준 "17" 이 대조 수준 "12" 와 유의하게 다른 유일한 수준이라고 결론 내릴 수 있습니다.

단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. weight 를 선택하고 **Y, 반응**을 클릭합니다.

4. age 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "weight 대 age 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 단계별 개별 쌍 비교 (Newman-Keuls)** 를 선택합니다 .

그림 6.22 단계별 개별 쌍 비교 (Newman-Keuls) 에 대한 평균 비교 보고서의 예

△ 평균 비교		
△ Newman-Keuls를 사용한 단계별 개별 쌍 비교		
경고: Newman-Keuls 검정은 가족군별 오차율을 제어하지 않습니다. 결과를 해석할 때는 주의하십시오.		
△ 최소 q*		
최소 q*	Alpha	
2.45044	0.05	
△ 연결 문자 보고서		
수준		평균
17	A	140.66667
16	A B	118.33333
15	A B	108.28571
14	B	100.83333
12	B	99.00000
13	B	94.71429
동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다.		

" 연결 문자 보고서 " 에는 수준 "17" 이 "16" 및 "15" 를 제외한 다른 모든 수준과 유의하게 다른 것으로 나타납니다 .

예 : 4 가지 평균 비교 검정 대조

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .
 2. **분석 > X 로 Y 적합**을 선택합니다 .
 3. weight 를 선택하고 **Y, 반응**을 클릭합니다 .
 4. age 를 선택하고 **X, 요인**을 클릭합니다 .
 5. **확인**을 클릭합니다 .
 6. "weight 대 age 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 비교**의 각 옵션을 하나씩 선택합니다 . " 대조군 비교 (Dunnett)" 옵션의 경우 대조군으로 17 세를 선택합니다 .
- 4 가지 방법은 모두 그룹 평균 간의 차이를 검정합니다 . 각 검정은 특정 가설에 사용되며 서로 다른 사실을 확인할 수 있습니다 .

그림 6.23 4 가지 다중 비교 검정의 비교 원

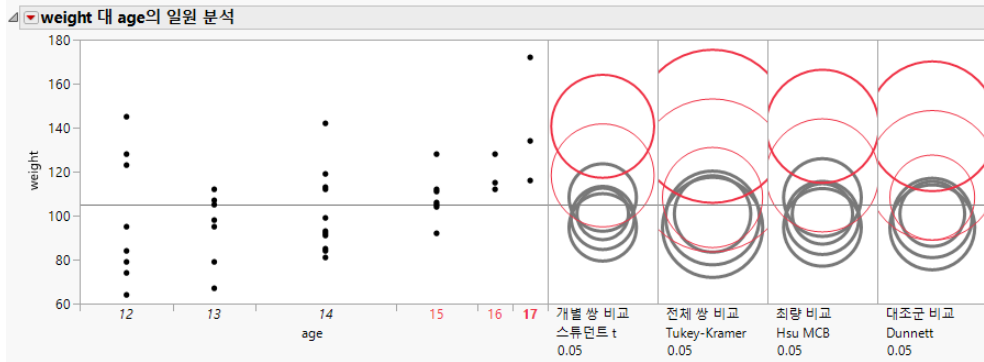


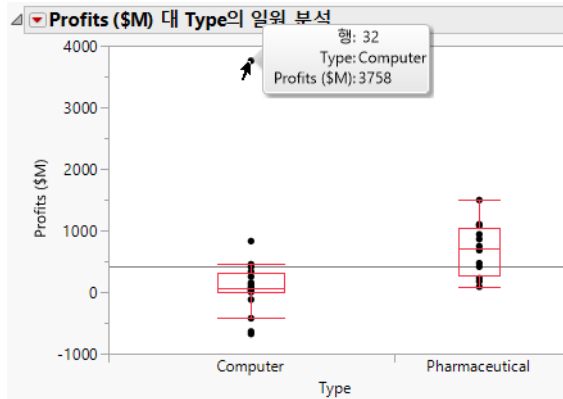
그림 6.23 에서는 17 세 그룹이 강조 표시되어 있습니다 . 다른 대조군 원은 17 세 그룹과의 관계에 따라 색상이 적용되어 있습니다 . 스튜던트 t 및 Hsu 방법의 경우에는 15 세 그룹 (위에서 세 번째 원)이 회색입니다. 이는 이 그룹이 17세 그룹과 유의하게 다름을 나타냅니다. 하지만 Tukey 및 Dunnett 방법의 경우에는 15 세 그룹이 빨간색으로 , 17 세 그룹과 유의하게 다르지 않음을 나타냅니다 .

비모수 Wilcoxon 검정의 예

회사의 평균 수익이 회사 유형에 따라 다른지 여부를 검정하려고 한다고 가정해 보겠습니다 . Companies.jmp 의 데이터는 제약 회사 (12 개) 와 컴퓨터 회사 (20 개) 라는 두 가지 회사 유형에 대한 다양한 측정 기준으로 구성되어 있습니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Companies.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Profits (\$M) 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Type 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Profits (\$M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림**을 선택합니다 .

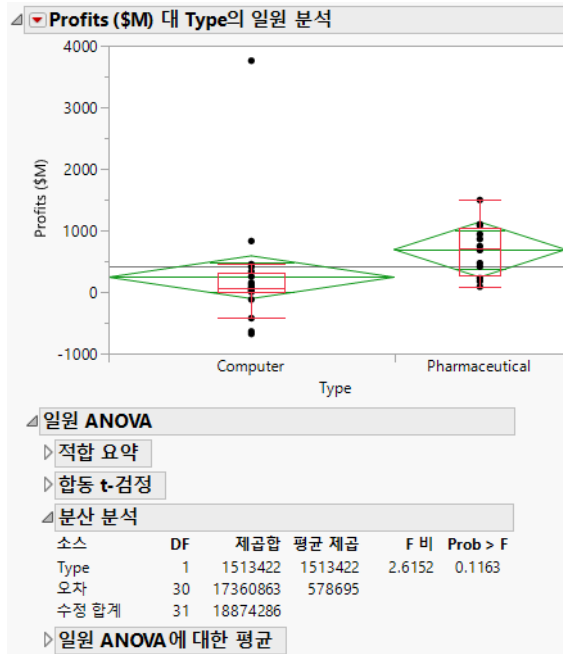
그림 6.24 컴퓨터 회사의 수익 분포



이 상자 그림에서는 데이터가 정규 분포 또는 대칭 분포가 아닌 것으로 나타납니다. 32 행 회사의 경우에는 모수 검정에 영향을 줄 수 있는 매우 큰 값이 있습니다.

7. "Profits (\$M) 대 Type 의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA/ 합동 t** 를 선택합니다.

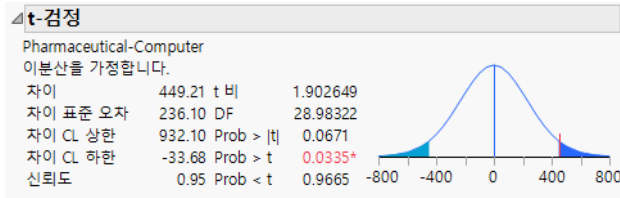
그림 6.25 회사 분산 분석



F-검정에서는 p 값이 크기 때문에 ($p = 0.1163$) 유의성이 없음을 보여 줍니다. 이는 32 행의 값이 커서 정규성 가정이 위배되었기 때문일 수 있습니다.

8. "Profits (\$M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **t-검정**을 선택합니다 .

그림 6.26 t-검정 결과



양측 검정의 "Prob > |t|" 는 0.0671 입니다 . t-검정에서는 등분산을 가정하지 않지만 이분산 t-검정도 모수 검정입니다 .

9. "Profits (\$M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **비모수 > Wilcoxon 검정**을 선택합니다 .

그림 6.27 Wilcoxon 검정 결과



Wilcoxon 검정은 비모수 검정입니다 . 이 검정은 순위를 기반으로 하므로 이상치의 영향을 받지 않습니다 . 또한 정규성이 필요하지 않습니다 .

Wilcoxon 검정 통계량에 대한 정규 근사와 카이제곱 근사는 모두 p 값이 0.0010 일 때 유의성을 나타냅니다 . 따라서 분포의 이 위치에서 유의한 차이가 있으며 회사 유형에 따라 평균 수익에 차이가 있다는 결론을 내릴 수 있습니다 .

정규 검정과 카이제곱 검정은 검정 통계량의 점근적 분포를 기반으로 합니다 . JMP Pro 를 사용하면 정확한 검정을 수행할 수 있습니다 .

10. **JMP PRO** "Profits (\$M) 대 Type 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **비모수 > 정확 검정 > Wilcoxon 정확 검정**을 선택합니다 .

그림 6.28 Wilcoxon 정확 검정 결과

Wilcoxon/Kruskal-Wallis 검정(순위 합)					
수준	개수	점수 합	기대 점수	점수 평균	(평균-평균0)/표준0
Computer	20	245.000	330.000	12.2500	-3.289
Pharmaceutical	12	283.000	198.000	23.5833	3.289
2표본 검정, 정규 근사					
S	Z	Prob> Z			
283	3.28916	0.0010*			
일원 검정, 카이제곱 근사					
카이제곱	DF	Prob>ChiSq			
10.9470	1	0.0009*			
2표본: 정확 검정					
S	Prob≥S		Prob≥ S-Mean		
283	0.0003*		0.0005*		

관측된 검정 통계량 값은 $S = 283$ 입니다. 이는 표본 크기가 더 작은 Type 수준(Pharmaceutical)의 순위 합입니다. 평균 중간 순위와의 절대 차이가 S 의 절대값에서 중간 순위의 평균을 뺀 값을 초과하는 것으로 관측될 확률은 0.0005입니다. 이는 위치의 차이에 대한 양측 검정으로, 회사 유형별로 수익이 다르지 않다는 가설을 기각할 것을 지지합니다.

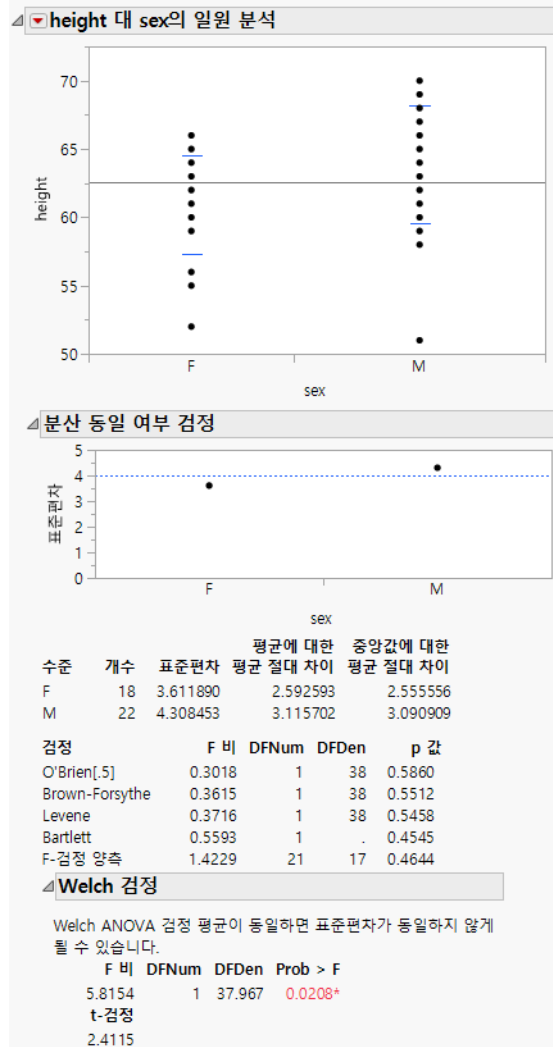
이 예에서는 비모수 검정이 정규성 기반 분산 분석 ANOVA 검정과 이분산 t -검정보다 더 적절합니다. 비모수 검정은 32 행에 있는 큰 값의 영향을 받지 않으며 정규성 가정이 필요하지 않습니다.

이분산 옵션의 예

두 평균 대신 두 분산(남학생 및 여학생)이 동일한지 검정하려고 한다고 가정해 보겠습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp를 엽니다.
2. **분석 > X로 Y적합**을 선택합니다.
3. height를 선택하고 **Y, 반응**을 클릭합니다.
4. sex를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "height 대 sex의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **이분산**을 선택합니다.

그림 6.29 이분산 보고서의 예



양측 F -검정의 p 값이 크기 때문에 분산이 동일하지 않다는 결론을 내릴 수 없습니다.

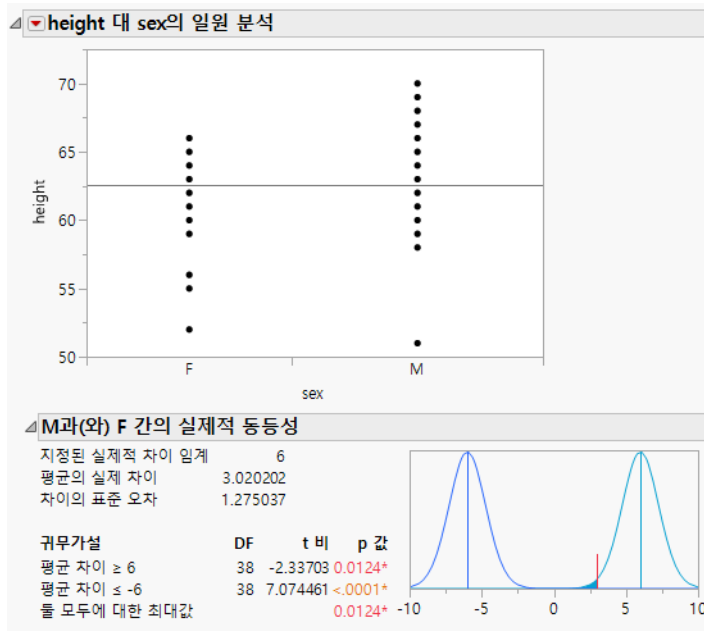
동등성 검정의 예

이 예에서는 Big Class.jmp 샘플 데이터 테이블을 사용합니다. 남학생과 여학생 사이의 키 차이가 6 인치 미만인지 확인하려고 합니다.

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Big Class.jmp 를 엽니다.
2. 분석 > X 로 Y 적합을 선택합니다.
3. height 를 선택하고 Y, 반응을 클릭합니다.

4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "height 대 sex 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **동등성 검정**을 선택합니다 .
7. " 실제로 0 으로 간주되는 차이 " 로 6 을 입력합니다 .
8. **확인**을 클릭합니다 .

그림 6.30 동등성 검정의 예



두 개의 단측 검정을 사용하면 둘 모두에서 p 값이 작다는 것을 확인할 수 있습니다 . 따라서 모집단 평균의 유의한 차이가 -6 에서 6 사이의 어딘가에 있다고 결론 내릴 수 있습니다 . 이 예에서는 평균이 실제로 동등하다고 선언할 수 있습니다 .

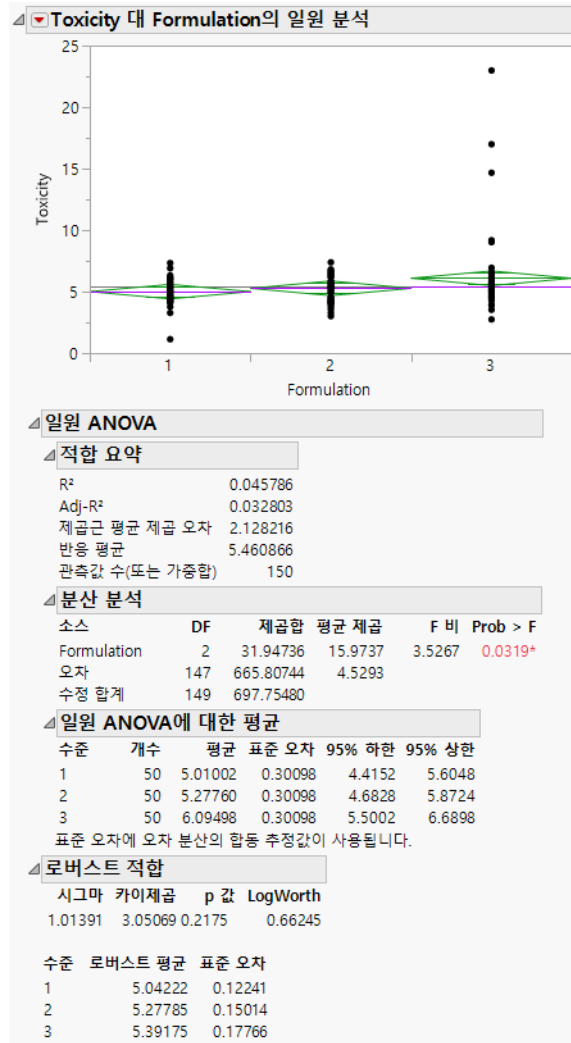
로버스트 적합 옵션의 예

Drug Toxicity.jmp 샘플 데이터 테이블의 데이터는 세 가지 약물 제제의 독성 수준을 보여 줍니다 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Drug Toxicity.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Toxicity 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Formulation 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .

6. "Toxicity 대 Formulation 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .
7. "Toxicity 대 Formulation 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **로버스트 > 로버스트 적합** 을 선택합니다 .

그림 6.31 로버스트 적합의 예

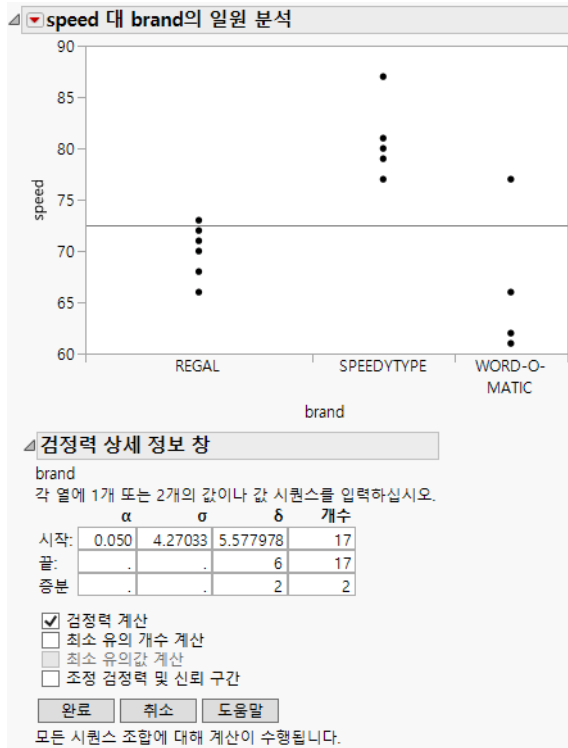


표준 "분산 분석" 보고서를 보면 p 값이 0.0319 이므로 세 가지 제제 사이에 차이가 있다는 잘못된 결론을 내리게 될 수 있습니다 . 하지만 " 로버스트 적합 " 보고서를 보면 여기서는 p 값이 0.21755 이므로 세 가지 제제가 유의하게 다르다는 결론을 내리지 않을 것입니다. 일부 관측값의 독성은 비정상적으로 높아서 데이터에 과도한 영향을 주는 것으로 보입니다 .

검정력 옵션의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Typing Data.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. **speed** 를 선택하고 **Y, 반응**을 클릭합니다 .
4. **brand** 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "speed 대 brand 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **검정력**을 선택합니다 .
7. " 시작 " 행에서 델타 (세 번째 상자) 값으로 2 를 입력하고 " 개수 " 에 11 을 입력합니다 .
8. " 끝 " 행에서 델타 값으로 6 을 입력하고 " 개수 " 상자에 17 을 입력합니다 .
9. " 증분 " 행에서 델타 값과 " 개수 " 모두에 2 를 입력합니다 .
10. **검정력 계산** 체크박스를 선택합니다 .

그림 6.32 검정력 상세 정보 창의 예



11. **완료**를 클릭합니다 .

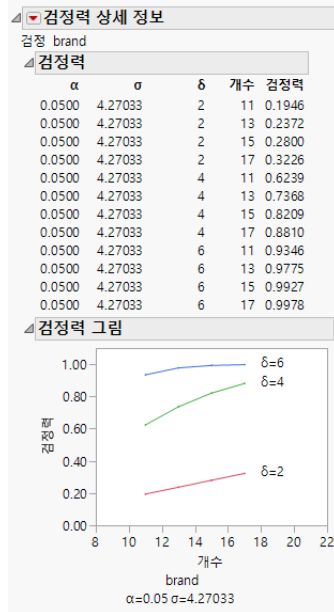
참고 : 필요한 옵션이 모두 적용되기 전까지는 **완료** 버튼이 흐리게 표시됩니다 .

델타 값과 개수의 각 조합에 대해 검정력이 계산되어 "검정력" 보고서에 표시됩니다.

검정력 값을 그림으로 표시하려면 다음을 수행하십시오.

12. "검정력 상세 정보"의 빨간색 삼각형을 클릭하고 **검정력 그림**을 선택합니다.

그림 6.33 검정력 보고서의 예



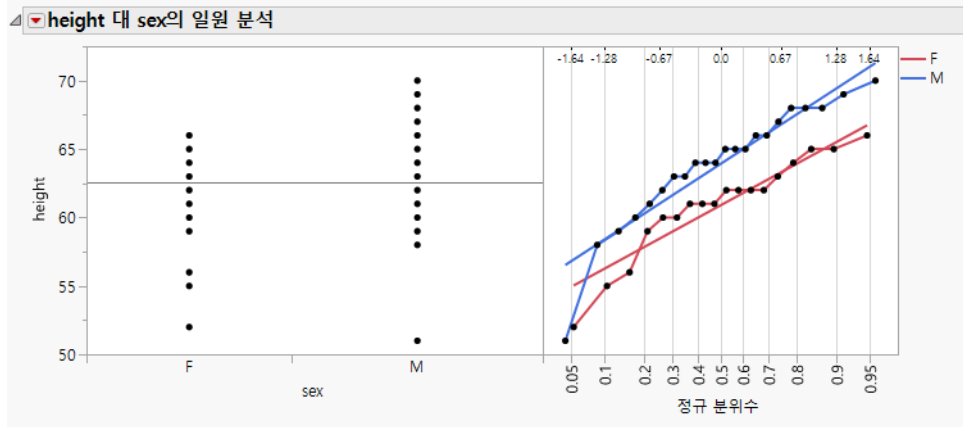
13. 그림의 데이터를 모두 표시하려면 "검정력" 축을 클릭한 채로 세로로 드래그해야 할 수 있습니다.

델타 값과 개수의 각 조합에 대해 검정력이 표시됩니다. 예상한 대로 "개수"(표본 크기) 값과 델타 값(평균 차이)이 클수록 검정력이 높아집니다.

정규 분위수 그림의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. height 를 선택하고 **Y, 반응**을 클릭합니다.
4. sex 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "height 대 sex 의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **정규 분위수 그림 > 분위수별 실제값 그림**을 선택합니다.

그림 6.34 정규 분위수 그림의 예



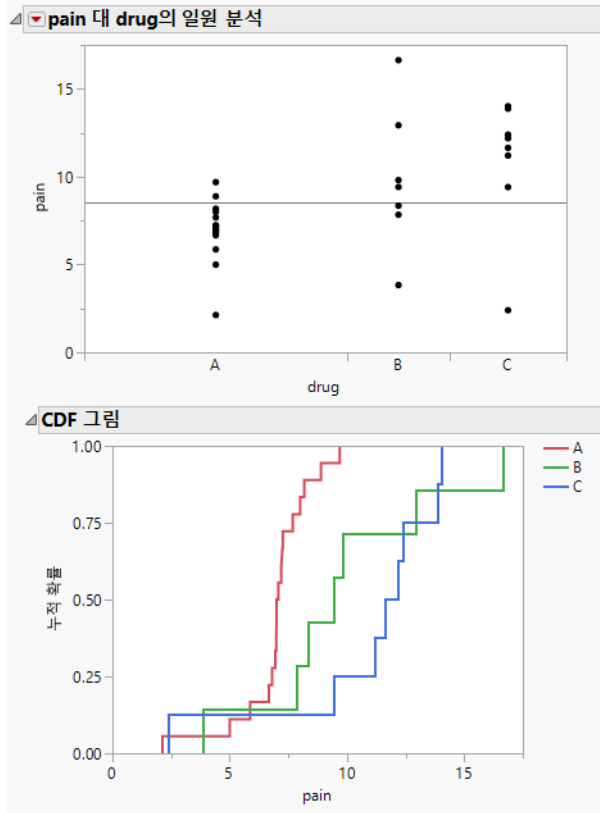
다음 사항에 유의하십시오 .

- 기본적으로 적합선이 표시됩니다 .
- 데이터 점은 적합선에 매우 가깝게 늘어서 있어 정규 분포를 나타냅니다 .

CDF 그림의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Analgesics.jmp 를 엽니다 .
2. 분석 > X 로 Y 적합을 선택합니다 .
3. pain 을 선택하고 Y, 반응을 클릭합니다 .
4. drug 를 선택하고 X, 요인을 클릭합니다 .
5. 확인을 클릭합니다 .
6. "pain 대 drug 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **CDF 그림**을 선택합니다 .

그림 6.35 CDF 그림의 예

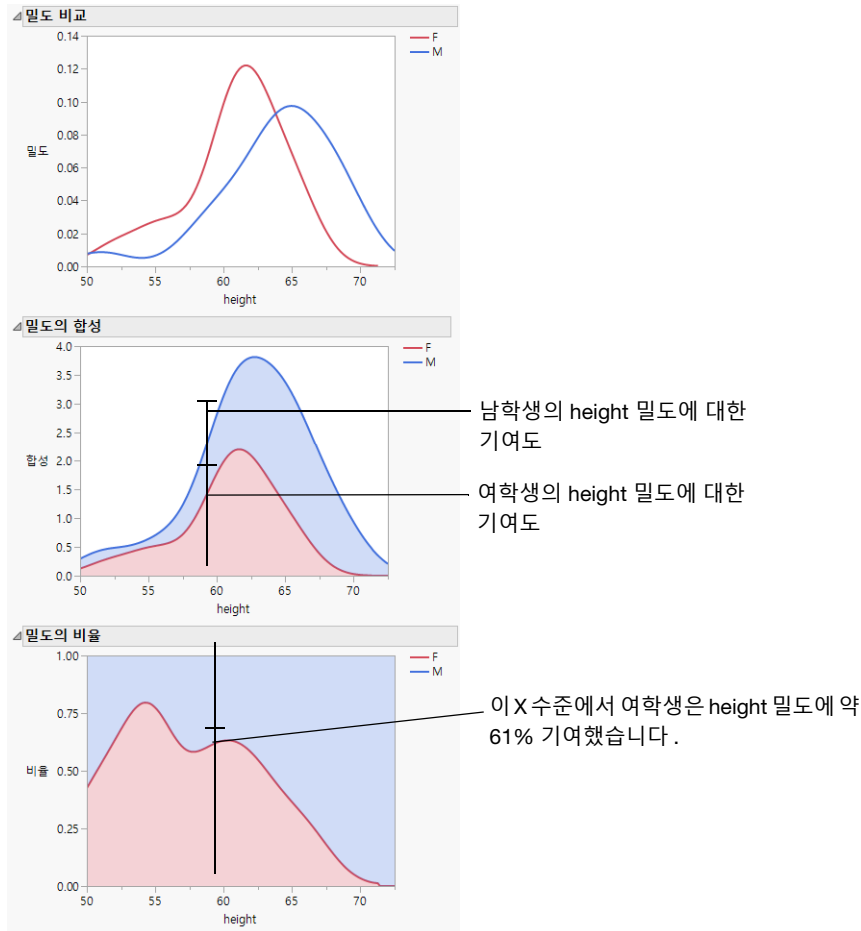


초기 일원 분석의 X 변수 수준은 CDF 그림에 서로 다른 곡선으로 표시됩니다. CDF 그림의 가로 축에는 초기 일원 분석의 y 값이 사용됩니다.

밀도 옵션의 예

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Big Class.jmp 를 엽니다.
2. 분석 > X 로 Y 적합을 선택합니다.
3. height 를 선택하고 Y, 반응을 클릭합니다.
4. sex 를 선택하고 X, 요인을 클릭합니다.
5. 확인을 클릭합니다.
6. "height 대 sex 의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 세 개의 옵션, 즉 밀도 > 밀도 비교, 밀도 > 밀도의 합성 및 밀도 > 밀도의 비율을 모두 선택합니다.

그림 6.36 밀도 옵션의 예



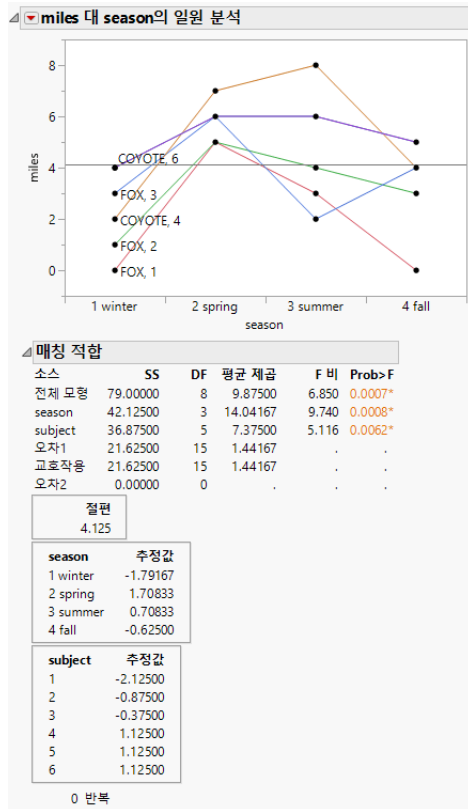
매칭 열 옵션의 예

이 예에서는 6 종의 동물과 이 동물들이 계절별로 이동한 거리 (마일) 가 포함되어 있는 Matching.jmp 샘플 데이터 테이블을 사용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Matching.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. miles 를 선택하고 **Y, 반응**을 클릭합니다 .
4. season 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "miles 대 season 의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **매칭 열**을 선택합니다 .
7. 매칭 열로 **subject** 를 선택합니다 .

8. **확인**을 클릭합니다.

그림 6.37 매칭 열 보고서의 예



이 그림에서는 개체를 매칭 변수로 하여 계절별 이동 거리 (마일) 를 그래프로 보여 줍니다. 그래프에서 각 개체의 첫 번째 추정값 옆에 표시되는 라벨은 **species** 및 **subject** 변수에 따라 결정되었습니다.

"매칭 적합" 보고서에는 F -검정을 사용한 **season** 및 **subject** 효과가 표시됩니다. 이는 교호작용항이 있는 모형과 없는 모형 두 개를 실행하는 경우 모형 적합 플랫폼에서 얻을 수 있는 검정과 동등합니다. 수준이 두 개뿐인 경우 F -검정은 쌍체 t -검정과 동등합니다.

참고: 모형 적합 플랫폼에 대한 자세한 내용은 **Fitting Linear Models**의 "Model Specification" 장에서 확인하십시오.

예 : 일원 분석을 위한 데이터 쌓기

데이터가 JMP 데이터 테이블이 아닌 다른 형식으로 존재하는 경우 데이터 배열이 한 행에 여러 관측값에 대한 정보가 포함되는 방식일 수도 있습니다. JMP에서 데이터를 분석하려면 데이터

를 가져온 후 JMP 데이터 테이블의 각 행에 단일 관측값에 대한 정보가 포함되도록 데이터를 재구성해야 합니다. 예를 들어 데이터가 스프레드시트에 들어 있다고 가정해 보겠습니다. 세 개의 생산 라인에서 생산된 제품에 대한 데이터가 세 개의 열 집합에 배열되어 있습니다. JMP 데이터 테이블에서는 세 개 생산라인에서 얻은 데이터를 단일 열 집합에 쌓아 각 행이 단일 제품의 데이터를 나타내도록 해야 합니다.

설명 및 목표

이 예에서는 서로 다른 세 개의 생산 라인에서 무작위로 표집된 시리얼 상자의 중량이 포함된 Fill Weights.xlsx 파일을 사용합니다. 그림 6.38에서는 이 데이터의 형식을 보여 줍니다.

- "ID" 열에는 측정된 각 시리얼 상자의 식별자가 포함되어 있습니다.
- "Line" 열에는 해당 생산 라인에서 표집된 상자의 중량 (온스) 이 포함되어 있습니다.

그림 6.38 데이터 형식

Weights					
ID	Line A	ID	Line B	ID	Line C
215	12.42	705	13.63	254	11.73
287	12.49	670	12.56	282	11.40
381	12.80	715	12.87	938	12.78
		683	13.09	597	12.19
		514	13.31	179	12.25
		517	12.64		
		946	12.75		

상자의 목표 충전 중량은 12.5 온스입니다. 세 개의 생산 라인이 목표를 충족하고 있는지 여부에 관심이 있지만 우선은 세 개의 라인이 동일한 평균 충전 비율을 달성하고 있는지 확인하려고 합니다. 일원 분석을 사용하여 평균 충전 중량 간에 차이가 있는지 검정할 수 있습니다.

일원 분석 플랫폼을 사용하려면 다음을 수행해야 합니다.

1. 데이터를 JMP 로 가져옵니다. 자세한 내용은 "[데이터 가져오기](#)"(215 페이지)에서 확인하십시오.
2. 데이터를 재구성하여 JMP 데이터 테이블에서 각 행이 단일 관측값만 반영하도록 합니다. 데이터를 재구성하려면 시리얼 상자 ID, 라인 식별자 및 중량을 해당 열에 쌓아야 합니다. 자세한 내용은 "[데이터 쌓기](#)"(217 페이지)에서 확인하십시오.

데이터 가져오기

이 예에서는 Microsoft Excel 의 데이터를 JMP 로 가져오는 두 가지 방법을 보여 줍니다. 한 가지 방법을 선택하거나 두 방법을 모두 사용해 보십시오.

- **파일 > 열기** 옵션을 사용하여 Excel 가져오기 마법사로 Microsoft Excel 파일의 데이터를 가져옵니다. 자세한 내용은 "[Excel 가져오기 마법사를 사용하여 데이터 가져오기](#)"(216 페이지)에서 확인하십시오. 이 방법은 모든 Excel 파일에 편리한 방법입니다.

- Microsoft Excel 의 데이터를 새 JMP 데이터 테이블에 복사하여 붙여넣습니다 . 자세한 내용은 "[Excel 에서 데이터 복사 및 붙여넣기](#)" (216 페이지) 에서 확인하십시오 . 데이터 파일이 작은 경우에 이 방법을 사용할 수 있습니다 .

Microsoft Excel 에서 데이터를 가져오는 방법에 대한 자세한 내용은 JMP 사용의 " 데이터 가져오기 " 장에서 확인하십시오 .

Excel 가져오기 마법사를 사용하여 데이터 가져오기

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Samples/Import Data** 폴더에 있는 **Fill Weights.xlsx** 를 엽니다 .
파일이 Excel 가져오기 마법사에서 열립니다 .
2. **열 머리글이 시작되는 행** 옆에 3 을 입력합니다 .
Excel 파일의 1 행에는 테이블에 대한 정보가 포함되어 있고 2 행은 비어 있습니다 . 열 머리글 정보는 3 행에서 시작됩니다 .
3. **열 머리글이 포함된 행 수**에 2 를 입력합니다 .
Excel 파일의 3 행과 4 행에는 열 머리글 정보가 포함되어 있습니다 .
4. **가져오기**를 클릭합니다 .

그림 6.39 Excel 가져오기 마법사를 사용하여 생성된 JMP 테이블

	Weights-ID	Weights-Line A	Weights-ID 2	Weights-Line B	Weights-ID 3	Weights-Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.40
3	381	12.80	715	12.87	938	12.78
4	•	•	683	13.09	597	12.19
5	•	•	514	13.31	179	12.25
6	•	•	517	12.64	•	•
7	•	•	946	12.75	•	•

7 개의 행에 데이터가 있고 각 행마다 여러 개의 ID 가 나타납니다 . 세 개 라인 각각에 대해 ID 및 중량 열이 있으며 총 열 수는 6 개입니다 .

ID 열 이름의 "Weights" 부분은 불필요한 데다 혼동을 일으킬 수 있습니다 . 지금 열 이름을 바꿀 수도 있지만 데이터를 쌓은 후 열 이름을 바꾸는 것이 더 효율적입니다 .

5. "[데이터 쌓기](#)" (217 페이지) 로 진행합니다 .

Excel 에서 데이터 복사 및 붙여넣기

1. Microsoft Excel 에서 **Fill Weights.xlsx** 를 엽니다 .
2. 테이블 내의 데이터를 선택하되 불필요한 "Weights" 머리글은 제외합니다 .
3. 마우스 오른쪽 버튼을 클릭하고 **복사**를 선택합니다 .
4. JMP 에서 **파일 > 새로 만들기 > 데이터 테이블**을 선택합니다 .

5. 편집 > 열 이름과 함께 붙여넣기를 선택합니다.

편집 > 열 이름과 함께 붙여넣기 옵션은 클립보드의 선택 내용에 열 이름이 포함된 경우에 사용됩니다.

그림 6.40 " 열 이름과 함께 붙여넣기 " 를 사용하여 생성된 JMP 테이블

	ID	Line A	ID 2	Line B	ID 3	Line C
1	215	12.42	705	13.63	254	11.73
2	287	12.49	670	12.56	282	11.40
3	381	12.80	715	12.87	938	12.78
4	•	•	683	13.09	597	12.19
5	•	•	514	13.31	179	12.25
6	•	•	517	12.64	•	•
7	•	•	946	12.75	•	•

6. " 데이터 쌓기 "(217 페이지) 로 진행합니다.

데이터 쌓기

" 쌓기 " 옵션을 사용하여 새 데이터 테이블의 각 행에 하나의 관측값이 포함되도록 할 수 있습니다. " 쌓기 " 옵션에 대한 자세한 내용은 JMP 사용의 " 데이터 재구성 " 장에서 확인하십시오.

1. JMP 데이터 테이블에서 테이블 > 쌓기를 선택합니다.
2. 6 개의 열을 모두 선택하고 열 쌓기를 클릭합니다.
3. 다중 계열 쌓기를 선택합니다.

두 개의 계열, 즉 ID 와 라인을 쌓을 것이므로 " 계열 수 " 는 변경하지 않고 기본값 2 로 설정된 상태로 둡니다. 계열이 포함된 열은 인접해 있지 않고 한 열씩 번갈아 표시되어 있습니다 ("ID", "Line A", "ID", "Line B", "ID", "Line C"). 따라서 " 인접 " 은 선택하지 않습니다.

4. 행별로 쌓기를 선택 취소합니다.
5. 결측 행 제거를 선택합니다.
6. 출력 테이블 이름 옆에 Stacked 를 입력합니다.
7. 확인을 클릭합니다.

새 데이터 테이블에서 데이터 및 데이터 2 는 ID 및 중량 데이터가 포함된 열입니다.

8. 라벨 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 열 삭제 를 선택합니다.

라벨 열 내의 항목은 가져온 데이터 테이블에 있는 상자 ID 열의 머리글입니다. 이러한 항목은 필요하지 않습니다.

9. 각 열의 열 머리글을 두 번 클릭하여 이름을 바꿉니다. 열 이름을 다음과 같이 변경합니다.
 - 데이터 -> ID
 - 라벨 2 -> Line
 - 데이터 2 -> Weight

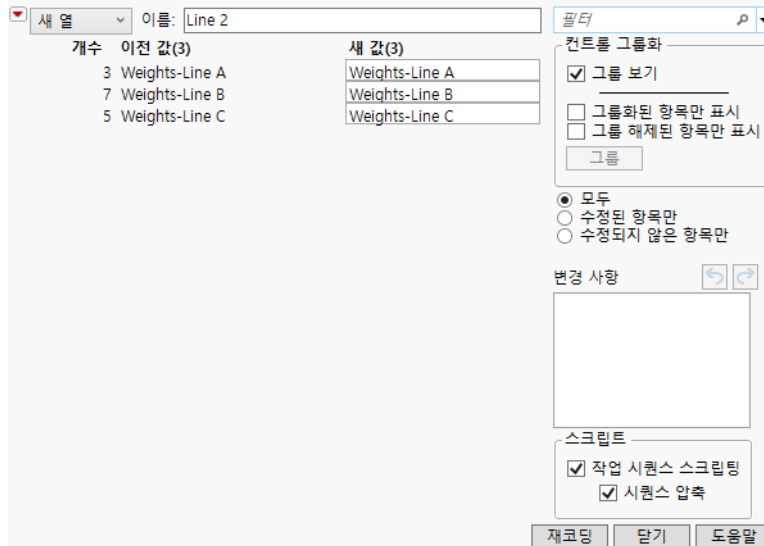
10. "열" 패널에서 ID 왼쪽의 아이콘을 클릭하고 **명목형**을 선택합니다.

ID 는 숫자로 지정되기는 했지만 식별자이므로 모델링 유형을 명목형으로 처리해야 합니다. 이 예에서는 이 단계가 중요하지 않지만 항상 열에 적절한 모델링 유형을 할당하는 것이 좋습니다.

11. (**파일 > 열기**를 사용하여 Excel 에서 데이터를 가져온 경우에만 해당) 다음을 수행합니다.

1. Line 열 머리글을 클릭하여 해당 열을 선택하고 **열 > 재코딩**을 선택합니다.
2. **새 값** 열의 값을 아래의 그림 6.41 에 표시된 것과 매칭되도록 변경합니다.

그림 6.41 열 값 재코딩



3. **완료 > 현재 위치**를 클릭합니다.

이제 JMP 분석에 적절하게 새 데이터 테이블이 구성되었습니다. 각 행에는 단일 시리얼 상자에 대한 데이터가 포함되어 있습니다. 첫 번째 열에는 상자 ID 가 있고, 두 번째 열에는 생산 라인이 있으며, 세 번째 열에는 상자의 중량이 있습니다 (그림 6.42).

그림 6.42 재코딩된 데이터 테이블

	ID	Line	Weight
1	215	Line A	12.42
2	287	Line A	12.49
3	381	Line A	12.8
4	705	Line B	13.63
5	670	Line B	12.56
6	715	Line B	12.87
7	683	Line B	13.09
8	514	Line B	13.31
9	517	Line B	12.64
10	946	Line B	12.75
11	254	Line C	11.73
12	282	Line C	11.4
13	938	Line C	12.78
14	597	Line C	12.19
15	179	Line C	12.25

일원 분석 수행

이 부분에는 다음과 같은 작업이 포함되어 있습니다.

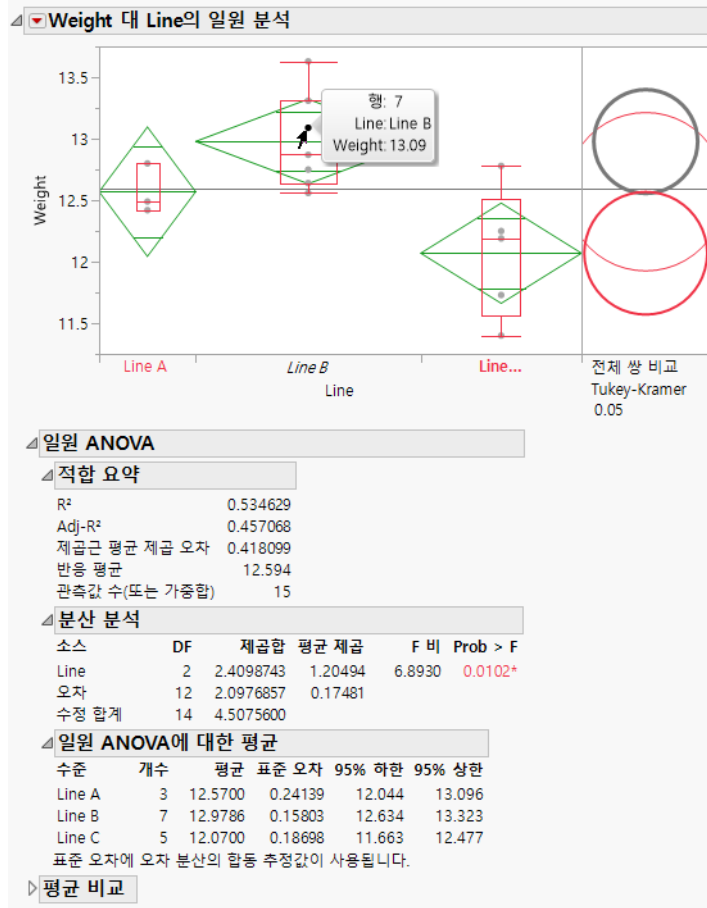
- 일원 분산 분석을 수행하여 세 개의 생산 라인 간에 평균 충전 중량의 차이가 있는지 검정합니다.
- 비교 원을 생성하여 차이가 있을 수 있는 라인을 확인합니다.
- 가중치를 조정하거나 상자를 보다 상세히 검토하려는 경우 ID 로 점에 라벨을 지정합니다.

시작하기 전에 **Stacked** 데이터 테이블을 사용하고 있는지 확인하십시오.

1. **분석 > X 로 Y 적합**을 선택합니다.
2. **Weight** 를 선택하고 **Y, 반응**을 클릭합니다.
3. **Line** 을 선택하고 **X, 요인**을 클릭합니다.
4. **확인**을 클릭합니다.
5. "Weight 대 Line 의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다.
 그림의 평균 다이아몬드는 생산 라인 평균에 대한 95% 신뢰 구간을 보여 줍니다. 평균 다이아몬드 외부의 점은 이상치인 것처럼 보일 수도 있습니다. 하지만 그렇지 않습니다. 이를 확인하기 위해 그림에 상자 그림을 추가합니다.
6. "Weight 대 Line 의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림**을 선택합니다.
 모든 점이 상자 그림 경계 내에 있습니다. 따라서 모두 이상치가 아닙니다.
7. 데이터 테이블의 "열" 패널에서 ID 를 마우스 오른쪽 버튼으로 클릭하고 **라벨 / 라벨 해제**를 선택합니다.

8. 그림의 점 위에 커서를 놓으면 해당 점의 ID 값과 함께 Line 및 Weight 데이터가 표시됩니다(그림 6.43).
9. "Weight 대 Line의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 비교 > 전체 쌍 비교(Tukey HSD)**를 선택합니다.
그림 오른쪽의 패널에 비교 원이 표시됩니다.
10. 맨 아래 비교 원을 클릭합니다.

그림 6.43 Weight 대 Line 의 일원 분석



"분산 분석" 보고서에서 p 값이 0.0102 인 것은 평균이 모두 동일하지 않다는 증거입니다. 이 그림에는 라인 C에 대한 비교 원이 선택되어서 빨간색으로 표시되어 있습니다. 라인 B에 대한 원은 굵은 회색으로 표시되므로 라인 C의 평균은 유의 수준 0.05에서 라인 B의 평균과 차이가 있습니다. 라인 A 및 B의 평균은 통계적으로 유의한 차이를 나타내지 않습니다.

그림에 표시된 평균 다이아몬드는 평균에 대한 95% 신뢰 구간 범위에 있습니다. 95% 신뢰 구간의 수치 한계는 "일원 ANOVA에 대한 평균" 보고서에서 제공됩니다. 이 두 가지는 모두 라인

B 및 C에 대한 신뢰 구간에 목표 충전 중량인 12.5가 포함되지 않았음을 나타냅니다. 라인 B는 중량 초과로 보이고 라인 C는 중량 미달로 보입니다. 이 두 생산 라인의 경우 목표 충전 중량 미충족을 초래한 근본적인 원인을 해결해야 합니다.

일원 분석 플랫폼에 대한 통계 상세 정보

- "비교 원"
- "검정력"
- "적합 요약 보고서"
- "분산 동일 여부 검정"
- "비모수 검정 통계량"

비교 원

두 개의 평균을 비교하는 한 가지 방법은 실제 평균 차이가 LSD(최소 유의차)보다 큰지 확인하는 것입니다. 이 최소 유의차는 스튜던트 t 통계량에 두 평균의 차이에 대한 표준 오차를 곱한 것으로, 계산식은 다음과 같습니다.

$$\text{LSD} = t_{\alpha/2} \text{std}(\hat{\mu}_1 - \hat{\mu}_2)$$

두 독립 평균의 차이에 대한 표준 오차는 다음 관계식으로 계산됩니다.

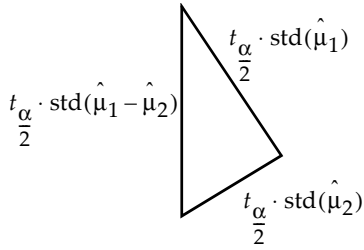
$$[\text{std}(\hat{\mu}_1 - \hat{\mu}_2)]^2 = [\text{std}(\hat{\mu}_1)]^2 + [\text{std}(\hat{\mu}_2)]^2$$

평균 간에 상관관계가 없는 경우 이러한 통계량은 다음과 같은 관계를 갖습니다.

$$\text{LSD}^2 = [t_{\alpha/2} \text{std}((\hat{\mu}_1 - \hat{\mu}_2))]^2 = [t_{\alpha/2} \text{std}(\hat{\mu}_1)]^2 + [t_{\alpha/2} \text{std}(\hat{\mu}_2)]^2$$

이러한 제곱 값 간에는 그림 6.44에 그래픽으로 표시된 직각 삼각형의 경우처럼 피타고라스 관계식이 성립됩니다.

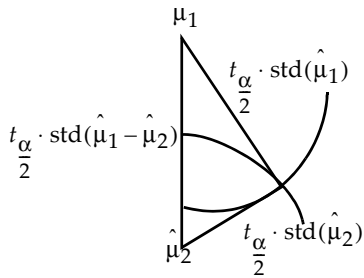
그림 6.44 두 평균 간의 차이 관계



이 삼각형의 빗변은 평균을 비교하기 위한 척도가 됩니다. 실제 차이가 빗변 (LSD) 보다 큰 경우에만 평균이 유의하게 다른 것으로 간주됩니다.

두 평균이 정확히 경계선에 있고 실제 차이는 최소 유의차와 동일하다고 가정해 보겠습니다. 세로 척도에서 측정된 평균 값을 꼭지점으로 하는 삼각형을 그려 봅니다. 또한 각 평균을 중심으로 하고 해당 평균의 신뢰 구간을 지름으로 하는 원을 그려 봅니다.

그림 6.45 t -검정 통계량의 기하학적 관계



각 원의 반지름은 삼각형의 해당 변 길이, 즉 $t_{\alpha/2} \text{std}(\hat{\mu}_i)$ 입니다.

원은 삼각형의 변과 동일하게 직각으로 교차해야 하며 그 관계는 다음과 같습니다.

- 평균이 정확히 최소 유의차만큼 다르면 각 평균을 중심으로 한 신뢰 구간 원이 직각으로 교차합니다. 즉, 탄젠트 각도가 직각입니다.

이제 평균 간에 최소 유의차보다 크거나 작은 차이가 있을 때 원이 교차하는 방식을 살펴보겠습니다.

- 원의 교차 외각이 직각보다 큰 경우에는 평균이 유의하게 다르지 않은 것입니다. 원의 교차 외각이 직각보다 작은 경우에는 평균이 유의하게 다른 것입니다. 즉, 외각이 90도 미만이면 평균이 최소 유의차에서 멀리 떨어져 있는 것입니다.
- 원이 교차하지 않으면 평균이 유의하게 다른 것입니다. 원이 내포 관계에 있으면 평균이 유의하게 다르지 않은 것입니다 (그림 6.11).

스튜던트 t 대신 다른 확률 분위수 값을 사용하여 다양한 다중 비교 검정에 이와 동일한 그래픽 기법을 적용할 수 있습니다.

검정력

검정력을 계산하려면 비중심 F 분포를 사용합니다. 계산식 (O'Brien and Lohr 1984) 은 다음과 같습니다.

$$\text{검정력} = \text{Prob}(F > F_{crit} \mid v_1, v_2, nc)$$

다음은 각 요소에 대한 설명입니다.

F : 비중심 $F(nc, v_1, v_2)$ 분포를 따르며, $F_{crit} = F_{(1-\alpha, v_1, v_2)}$ 는 자유도가 v_1 및 v_2 인 F 분포의 $1 - \alpha$ 분위수입니다.

$v_1 = r - 1$: 분자 DF 입니다.

$v_2 = r(n - 1)$: 분모 DF 입니다.

n : 그룹당 개수입니다.

r : 그룹 수입니다.

$nc = n(CSS)/\sigma^2$: 비중심성 모수입니다.

$$CSS = \sum_{g=1}^r (\mu_g - \mu)^2 : \text{수정 제곱합입니다.}$$

μ_g : g 번째 그룹의 평균입니다.

μ : 전체 평균입니다.

σ^2 : MSE(평균 제곱 오차) 를 사용하여 추정됩니다.

적합 요약 보고서

R^2

모형에 대한 분산 분석 보고서의 통계량을 사용하는 경우 연속형 반응 적합에 대한 R^2 는 항상 다음과 같이 계산됩니다.

$$\frac{\text{Sum of Squares (Model)}}{\text{Sum of Squares (C Total)}}$$

Adj- R^2

Adj- R^2 는 제곱합 대신 사용되는 평균 제곱의 비율로, 다음과 같이 계산됩니다.

$$1 - \frac{\text{Mean Square (Error)}}{\text{Mean Square (C Total)}}$$

오차의 평균 제곱은 "분산 분석" 보고서에서 구하며, 수정 합계의 평균 제곱은 수정 합계 제곱합을 해당 자유도로 나눠서 계산할 수 있습니다. 자세한 내용은 "분산 분석 보고서"(167 페이지)에서 확인하십시오.

분산 동일 여부 검정

F 비

O'Brien 검정에서는 새 변수의 그룹 평균이 원래 반응의 그룹 표본 분산과 동일하도록 종속 변수를 생성합니다. O'Brien 변수는 다음과 같이 계산됩니다.

$$r_{ijk} = \frac{(n_{ij} - 1.5)n_{ij}(y_{ijk} - \bar{y}_{ij})^2 - 0.5s_{ij}^2(n_{ij} - 1)}{(n_{ij} - 1)(n_{ij} - 2)}$$

여기서 n 은 y_{ijk} 관측값의 개수를 나타냅니다.

Brown-Forsythe 는 $z_{ij} = |y_{ij} - \tilde{y}_i|$ 에 대한 ANOVA 에서 구한 모형 F 통계량이며, 여기서 \tilde{y}_i 는 i 번째 수준의 중앙값 반응입니다.

Levene F 는 $z_{ij} = |y_{ij} - \bar{y}_i|$ 에 대한 ANOVA 에서 구한 모형 F 통계량이며, 여기서 \bar{y}_i 는 i 번째 수준의 평균 반응입니다.

Bartlett 검정은 다음과 같이 계산됩니다.

$$T = \frac{v \log \left(\sum_i \frac{v_i}{v} s_i^2 \right) - \sum_i v_i \log(s_i^2)}{1 + \left(\frac{\sum_i \frac{1}{v_i} - \frac{1}{v}}{3(k-1)} \right)} \quad \text{단,} \quad v_i = n_i - 1 \quad \text{및} \quad v = \sum_i v_i$$

여기서 n_i 는 i 번째 수준의 관측값 수이고, s_i^2 은 i 번째 수준의 반응 표본 분산입니다. Bartlett 통계량은 χ^2 분포를 따릅니다. 카이제곱 검정 통계량을 자유도로 나누면 보고된 F 값이 됩니다.

Welch 검정 F 비

Welch 검정 F 비는 다음과 같이 계산됩니다.

$$F = \frac{\left[\frac{\sum_i w_i (\bar{y}_i - \bar{y}_{..})^2}{k-1} \right]}{\left\{ 1 + \frac{2(k-2)}{k^2-1} \left[\sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i-1} \right] \right\}} \quad \text{단, } w_i = \frac{n_i}{s_i^2}, u = \sum_i w_i, \bar{y}_{..} = \frac{\sum_i w_i \bar{y}_i}{u}$$

여기서 n_i 는 i 번째 수준의 관측값 수이고, \bar{y}_i 는 i 번째 수준의 평균 반응이며, s_i^2 은 i 번째 수준의 반응 표본 분산입니다.

Welch 검정 DF Den

분모 자유도에 대한 Welch 근사 계산식은 다음과 같습니다.

$$df = \frac{1}{\left(\frac{3}{k^2-1} \right) \left[\sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i-1} \right]}$$

여기서 w_i , n_i 및 u 는 F 비 계산식에서의 정의와 같습니다.

비모수 검정 통계량

이 섹션에서는 Wilcoxon, 중앙값, van der Waerden 및 Friedman Rank 검정에 사용되는 검정 통계량의 계산식을 제공합니다.

표기

언급된 검정은 점수를 기반으로 하며 다음과 같은 표기를 사용합니다.

$j = 1, \dots, n$ 전체 표본의 관측값입니다.

$i = 1, \dots, k$ X의 수준입니다. 여기서 k 는 총 수준 수입니다.

n_1, n_2, \dots, n_k X의 k 개 수준 각각에 포함된 관측값의 개수입니다.

R_j j 번째 관측값의 중간 순위입니다. 중간 순위는 해당 관측값의 순위(동일 점수가 없는 경우)이거나 평균 순위(동일 점수가 있는 경우)입니다.

α 다양한 검정에서 점수를 정의하는 데 사용되는 중간 순위의 함수입니다.

시작 창에서 블록 변수를 지정한 경우에는 다음 표기가 사용됩니다.

$b = 1, \dots, B$ 블록 변수의 수준입니다. 여기서 B 는 총 블록 수입니다.

R_{bi} 블록 b 내에 있는 X 의 i 번째 수준에 해당되는 중간 순위입니다.

함수 α 는 점수를 다음과 같이 정의합니다.

Wilcoxon 점수

$$\alpha(R_j) = R_j$$

중앙값 점수

$$\alpha(R_j) = \begin{cases} 1 & \text{단, } R_j > \text{중앙값} \\ 0 & \text{단, } R_j < \text{중앙값} \\ t & \text{단, } R_j = \text{중앙값} \end{cases}$$

n_t 가 중앙값 위치에 있는 동일 관측값의 개수이고, n_u 이 중앙값보다 큰 관측값의 개수라고 할 때, t 는 다음과 같이 구합니다.

$$t = \frac{\text{floor}(n/2) - n_u}{n_t}$$

van der Waerden 점수

$$\alpha(R_j) = \text{표준 정규 분위수}(R_j/(n+1))$$

Friedman 순위 점수

$$\alpha(R_{bi}) = R_{bi}$$

2 표본 정규 근사

정규 근사를 기반으로 하는 검정은 X 의 수준이 정확히 두 개뿐인 경우에만 가능합니다. 이 섹션에서 사용되는 표기는 "표기"(225 페이지)에 정의되어 있습니다. 2 표본 정규 근사 보고서에 표시되는 통계량은 아래와 같이 정의됩니다.

S 통계량 S 는 더 작은 그룹에 있는 관측값에 대한 $\alpha(R_j)$ 값의 합입니다. X 의 두 수준에 동일한 수의 관측값이 있으면 S 값은 값 순서화 시 X 의 마지막 수준에 대응합니다.

Z Z 값은 다음과 같이 구합니다.

$$Z = (S - E(S)) / \sqrt{\text{Var}(S)}$$

참고 : Wilcoxon 검정에서는 연속성 수정이 추가됩니다. $(S - E(S))$ 가 0 보다 크면 분자에서 0.5 를 뺍니다. $(S - E(S))$ 가 0 보다 작으면 분자에 0.5 를 더합니다.

$E(S)$ 귀무가설하의 S 기대값입니다. 더 작은 수준 또는 값 순서화 시 마지막 수준 (두 그룹의 관측값 수가 동일한 경우) 의 관측값 수는 n_l 로 나타냅니다.

$$E(S) = \frac{n_l}{n} \sum_{j=1}^n \alpha(R_j)$$

$\text{Var}(S)$ 모든 관측값의 평균 점수를 ave 라고 정의할 경우 S 의 분산은 다음과 같이 구합니다.

$$\text{Var}(S) = \frac{n_1 n_2}{n(n-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

Friedman 순위 검정의 2 표본 정규 근사

Friedman 순위 검정을 사용하는 경우 2 표본 정규 근사에 대한 계산은 위와 동일하되 S 의 분산만 다릅니다. S 의 분산에 대한 계산식은 다음과 같습니다.

$$\text{Var}(S) = \frac{B}{(n-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

일원 카이제곱 근사

참고 : Wilcoxon 점수를 기반으로 하는 카이제곱 검정은 Kruskal-Wallis 검정이라고 합니다.

이 섹션에서 사용되는 표기는 "표기"(225 페이지)에 정의되어 있습니다. 카이제곱 통계량을 계산하는 데는 다음과 같은 통계량이 사용됩니다.

T_i X 의 i 번째 수준에 대한 총 점수입니다.

$E(T_i)$ 수준 간에 차이가 없다는 귀무가설하에 수준 i 의 총 점수에 대한 기대값으로, 다음과 같이 구합니다.

$$E(T_i) = \frac{n_i}{n} \sum_{j=1}^n \alpha(R_j)$$

$\text{Var}(T)$ 모든 관측값의 평균 점수를 ave 라고 정의할 경우 T 의 분산은 다음과 같이 구합니다.

$$Var(T) = \frac{1}{(n-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2$$

이 검정 통계량의 값은 아래와 같이 구합니다. 이 통계량은 $k-1$ 자유도에 대한 점근적 카이제곱입니다.

$$C = \left(\sum_{i=1}^k (T_i - E(T_i))^2 / n_i \right) / Var(T)$$

Friedman 순위 검정에 대한 일원 카이제곱 근사

Friedman 순위 검정에 대한 카이제곱 검정 통계량은 다음과 같이 계산됩니다.

$$C = \frac{\sum_{i=1}^k (T_i - E(T_i))^2 / n_i}{\frac{1}{(k-1)} \sum_{j=1}^n (\alpha(R_j) - ave)^2 / n_i}$$

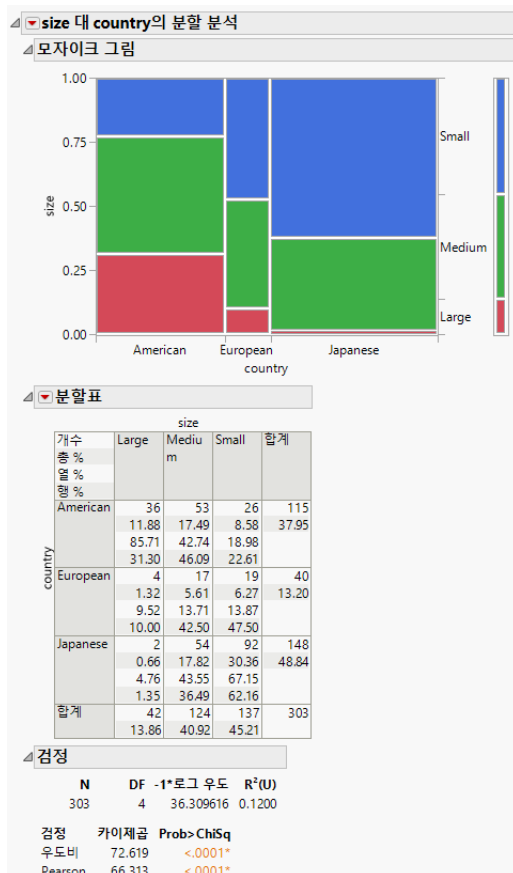
7 장

분할 분석

두 범주형 변수 간의 관계 분석

분할 분석 또는 X로 Y 적합 플랫폼에서는 범주형 (명목형 또는 순서형) 변수 Y의 분포가 두 번째 범주형 변수 X의 수준 간에 어떻게 달라지는지 살펴볼 수 있습니다. 분할 분석 플랫폼은 X로 Y 적합 플랫폼의 범주형 대 범주형 분석법입니다. 분석 결과에는 모자이크 그림, 빈도 수 및 비율이 포함됩니다. 대화식으로 비율에 대한 평균 분석, 대응 분석 그림 등의 추가적인 데이터 분석 및 검정을 수행할 수 있습니다.

그림 7.1 분할 분석의 예



목차

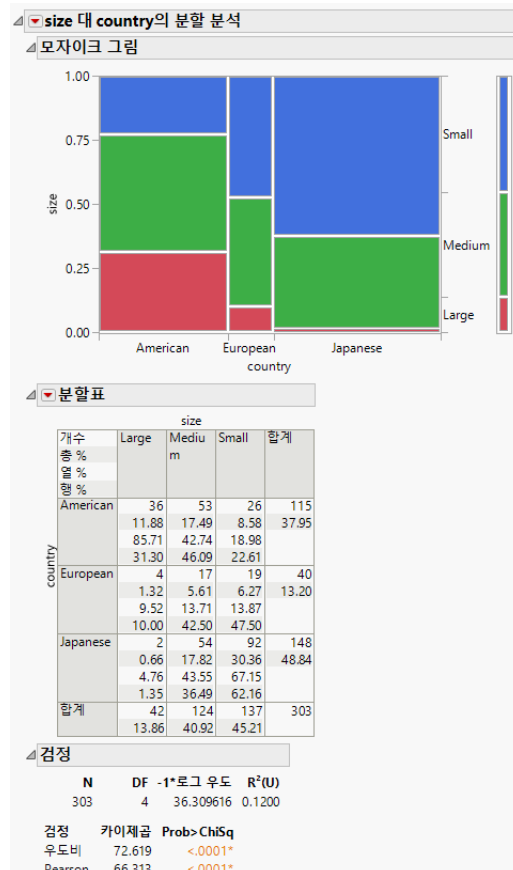
분할 분석의 예	231
분할 분석 플랫폼 시작	232
데이터 형식	233
분할 분석 보고서	233
분할 분석 플랫폼 옵션	234
모자이크 그림	235
팝업 메뉴	236
분할표	238
분할표 설명	239
검정	239
검정 보고서 설명	240
Fisher 정확 검정	240
비율에 대한 평균 분석	241
대응 분석	241
대응 분석 그림 이해	241
대응 분석 옵션	242
상세 정보 보고서	242
Cochran-Mantel-Haenszel 검정	242
합치도 통계량	243
상대 위험도	243
비율에 대한 2 표본 검정	244
연관성 측도	244
Cochran Armitage 추세 검정	245
정확 검정	246
분할 분석 플랫폼의 추가 예	246
비율에 대한 평균 분석의 예	247
대응 분석의 예	248
Cochran Mantel Haenszel 검정의 예	250
합치도 통계량 옵션의 예	251
상대 위험도 옵션의 예	252
비율에 대한 2표본 검정의 예	254
연관성 측도 옵션의 예	255
Cochran Armitage 추세 검정의 예	255
분할 분석 플랫폼에 대한 통계 상세 정보	256
합치도 통계량	256
승산비 옵션	257
검정 보고서	257
대응 분석의 상세 정보 보고서	258

분할 분석의 예

이 예에서는 자동차에 대한 설문조사에서 수집된 데이터를 사용합니다. 이 데이터에는 응답자 속성, 즉 성별, 결혼 여부 및 연령이 포함되어 있습니다. 또한 응답자 소유 자동차의 속성, 즉 생산 국가, 크기 및 유형도 포함되어 있습니다. 자동차 크기 (소형, 중형 및 대형) 와 자동차 생산 국가 사이의 관계를 검토하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. **size** 를 선택하고 **Y, 반응**을 클릭합니다.
4. **country** 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.

그림 7.2 분할 분석의 예



모자이크 그림 및 범례에서 다음을 확인할 수 있습니다.

- 대형 크기 범주에 속하는 일본 자동차는 매우 적습니다.
- 유럽 자동차의 대부분은 소형 및 중형 크기 범주에 속합니다.
- 미국 자동차의 대부분은 대형 및 중형 크기 범주에 속합니다.

분할 분석 플랫폼 시작

분할 분석을 수행하려면 다음을 수행하십시오.

1. **분석 > X 로 Y 적합**을 선택합니다.
2. **Y, 반응**에 명목형 또는 순서형 열을 입력합니다.
3. **X, 요인**에 명목형 또는 순서형 열을 입력합니다.

그림 7.3 X 로 Y 적합 시작 창

두 변수 사이의 관계를 모델링합니다.

다이어그램 위에 "분할 분석" 이라고 표시되면 분할 분석을 수행할 것임을 나타냅니다.

참고 : "JMP 시작하기" 창에서 분할 분석을 시작할 수도 있습니다. **보기 > JMP 시작하기 > 기본 분석 > 분할 분석**을 선택합니다.

이 시작 창에 대한 자세한 내용은 "**X 로 Y 적합 소개**" 장 (105 페이지) 에서 확인하십시오. "열 선택" 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 "시작하기" 장에서 확인하십시오.

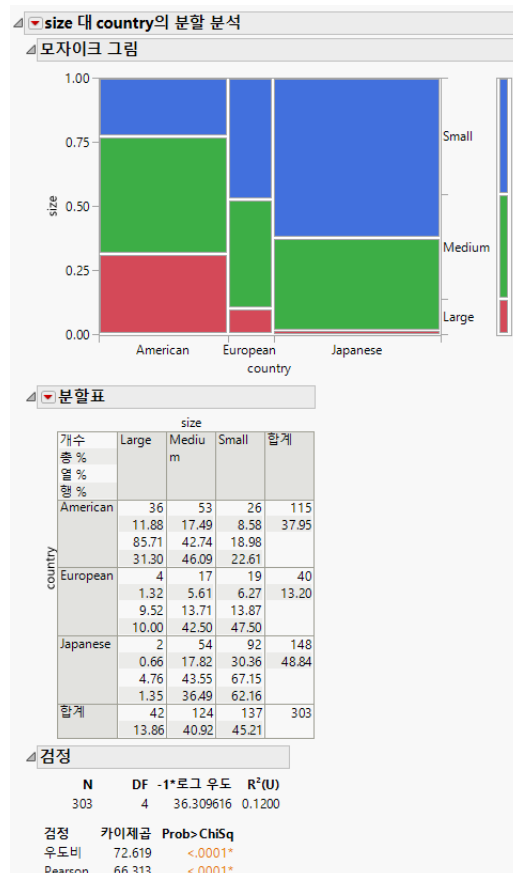
데이터 형식

범주형 데이터는 종종 Y 및 X 변수의 각 수준 조합마다 하나씩의 행만 데이터 테이블에 포함된 요약 형식으로 제공됩니다. 이 경우 빈도 또는 가중치 변수를 사용하여 각 행이 나타내는 관측값의 개수를 나타낼 수 있습니다. 요약된 범주형 데이터의 예는 "비율에 대한 평균 분석의 예"(247 페이지)에서 확인하십시오.

분할 분석 보고서

그림 7.4에 표시된 것과 같은 그림을 생성하려면 "분할 분석의 예"(231 페이지)에 설명된 방법을 따르십시오.

그림 7.4 분할 분석 보고서의 예



참고 : 데이터 테이블에서 제외된 모든 행은 모자이크 그림에서도 숨겨집니다 .

분할 분석 보고서에는 기본적으로 모자이크 그림 , 분할표 및 검정 보고서가 표시됩니다 . 빨간색 삼각형 메뉴 내에 있는 옵션을 사용하여 다른 분석 및 검정을 추가할 수 있습니다 . 이러한 모든 보고서 및 옵션에 대한 자세한 내용은 "[분할 분석 플랫폼 옵션](#)"(234 페이지)에서 확인하십시오 .

분할 분석 플랫폼 옵션

참고 : " 적합 그룹 " 메뉴는 Y 변수를 여러 개 지정한 경우에 표시됩니다 . 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 를 기준으로 정렬할 수 있습니다 . 자세한 내용은 [Fitting Linear Models](#) 의 "Standard Least Squares Report and Options" 장에서 확인하십시오 .

" 분할 분석 " 옆의 빨간색 삼각형 메뉴에 있는 플랫폼 옵션을 사용하여 데이터에 대한 추가 분석 및 검정을 수행할 수 있습니다 .

모자이크 그림 분할표의 데이터를 그래픽으로 나타낸 것입니다. 자세한 내용은 "[모자이크 그림](#)"(235 페이지)에서 확인하십시오 .

분할표 이원 빈도 테이블입니다 . 각 요인 수준마다 하나씩의 행과 각 반응 수준마다 하나씩의 열이 있습니다 . 자세한 내용은 "[분할표](#)"(238 페이지)에서 확인하십시오 .

검정 연속형 데이터에 대한 분산 분석 테이블과 유사합니다 . X 수준 간에 반응 수준 비율이 동일할지 검정한 결과를 보여 줍니다 . 자세한 내용은 "[검정](#)"(239 페이지)에서 확인하십시오 .

α 수준 설정 신뢰 구간에 사용된 유의 수준을 변경합니다 . 일반적인 값 (0.10, 0.05, 0.01) 중 하나를 선택하거나 , **기타** 옵션을 사용하여 특정 값을 선택합니다 .

비율에 대한 평균 분석 (반응의 수준이 정확히 두 개뿐인 경우에만 표시됨) X 수준에 대한 반응 비율을 전체 반응 비율과 비교합니다 . 자세한 내용은 "[비율에 대한 평균 분석](#)"(241 페이지)에서 확인하십시오 .

대응 분석 빈도 테이블에서 유사한 개수 패턴을 가진 행 또는 열을 표시합니다 . 대응 분석 그림에는 분할표의 각 행과 각 열에 해당하는 점이 하나씩 있습니다 . 자세한 내용은 "[대응 분석](#)"(241 페이지)에서 확인하십시오 .

Cochran Mantel Haenszel 세 번째 분류 기준에 따라 블록화한 후 두 범주형 변수 사이에 관계가 있는지 검정합니다 . 자세한 내용은 "[Cochran-Mantel-Haenszel 검정](#)"(242 페이지)에서 확인하십시오 .

합치도 통계량 (X 및 Y 변수의 수준이 동일한 경우에만 표시됨) 카과 통계량 (Agresti 1990), 해당 표준 오차, 신뢰 구간, 가설 검정 및 Bowker 대칭성 검정 (McNemar 검정이라고도 함) 을 표시합니다 . 자세한 내용은 "[합치도 통계량](#)"(243 페이지)에서 확인하십시오 .

상대 위험도 (X 및 Y 변수의 수준이 두 개인 경우에만 표시됨) 위험 비율을 계산합니다 . 자세한 내용은 "[상대 위험도](#)"(243 페이지)에서 확인하십시오 .

승산비 (X 및 Y 변수의 수준이 두 개인 경우에만 표시됨) 승산비 보고서를 생성합니다. 자세한 내용은 "[승산비 옵션](#)"(257 페이지)에서 확인하십시오.

보고서에는 이 비율에 대한 신뢰 구간도 표시됩니다. α 수준 설정 옵션을 사용하여 유의 수준을 변경할 수 있습니다.

비율에 대한 2 표본 검정 (X 및 Y 변수의 수준이 두 개인 경우에만 표시됨) 비율에 대한 2 표본 검정을 수행합니다. 이 검정에서는 X 변수의 두 수준 간에 Y 변수의 비율을 비교합니다. 자세한 내용은 "[비율에 대한 2 표본 검정](#)"(244 페이지)에서 확인하십시오.

연관성 측도 분할표의 변수 간 연관성을 설명합니다. 자세한 내용은 "[연관성 측도](#)"(244 페이지)에서 확인하십시오.

Cochran Armitage 추세 검정 (한 변수의 수준이 두 개이고 다른 변수는 순서형인 경우에만 표시됨) 단일 변수의 수준 간에 이항 비율의 추세를 검정합니다. 자세한 내용은 "[Cochran Armitage 추세 검정](#)"(245 페이지)에서 확인하십시오.

JMP PRO 정확 검정 다음 검정의 정확 검정 버전을 제공합니다.

- Fisher 검정
- Cochran Armitage 추세 검정
- 합치도 통계량

자세한 내용은 "[정확 검정](#)"(246 페이지)에서 확인하십시오.

표시 옵션 > 가로 모자이크 모자이크 그림을 가로 또는 세로로 회전합니다.

데이터 테이블로 만들기 보고서 테이블을 사용하여 JMP 데이터 테이블을 생성합니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

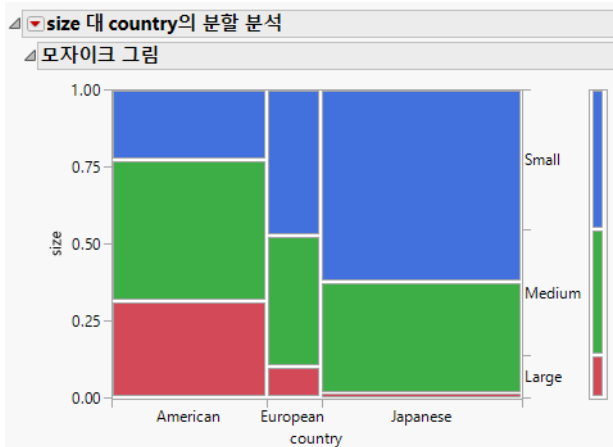
스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

모자이크 그림

모자이크 그림은 이원 빈도 테이블 또는 분할표를 그래픽으로 나타낸 것입니다. 모자이크 그림은 여러 개의 직사각형으로 나뉘며, 각 직사각형의 세로 길이는 X 변수의 각 수준에서 Y 변수가 나타나는 비율에 비례합니다. 모자이크 그림은 Hartigan and Kleiner 연구 자료 (1981) 에서 처음 소개되었으며 Friendly 연구 자료 (1994) 에서 구체화되었습니다.

그림 7.5 에 표시된 것과 같은 그림을 생성하려면 " [분할 분석의 예](#) "(231 페이지) 에 설명된 방법을 따르십시오 .

그림 7.5 모자이크 그림의 예



이 모자이크 그림의 경우 다음 사항에 유의하십시오 .

- 가로 축에서의 비율은 X 변수 (country) 의 수준별 관측값 수를 나타냅니다 .
- 오른쪽 세로 축에서의 비율은 결합된 수준(American, European 및 Japanese) 전체의 자동차 크기 (Small, Medium 및 Large) 별 비율을 나타냅니다 .
- 왼쪽 세로 축의 척도는 반응 확률을 보여 줍니다 . 전체 축은 전체 표본을 나타내는 확률 1 과 동등합니다 .

모자이크 그림의 직사각형을 클릭하면 선택한 직사각형이 강조 표시되고 관련 데이터 테이블의 해당 데이터도 강조 표시됩니다 .

두 가지 방법 중 하나로 변수를 드래그하여 놓아 모자이크 그림의 변수를 바꿀 수 있습니다 . 한 가지 방법은 한 축의 변수를 다른 축으로 드래그하여 놓아 기존 변수를 바꾸는 것이고 , 다른 방법은 관련 데이터 테이블의 " 열 " 패널에서 변수를 클릭한 후 축으로 드래그하는 것입니다 .

팝업 메뉴

셀의 색상을 변경하고 라벨을 지정하려면 모자이크 그림을 마우스 오른쪽 버튼으로 클릭합니다 .

색상 설정 수준에 현재 할당된 색상을 표시합니다 . 자세한 내용은 " [색상 설정](#) "(237 페이지) 에 서 확인하십시오 .

셀 라벨 지정 모자이크 그림에 표시할 라벨을 지정합니다 . 다음 옵션 중 하나를 선택합니다 .

라벨 해제 라벨을 표시하지 않고 , 다른 옵션을 모두 제거합니다 .

개수 표시 각 셀의 관측값 수를 표시합니다 .

백분율 표시 각 셀의 관측값 백분율을 표시합니다.

라벨 표시 각 셀에 해당하는 Y 변수의 수준을 표시합니다.

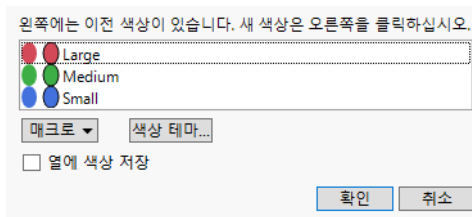
행 라벨 표시 셀이 나타내는 모든 행의 행 라벨을 표시합니다.

참고 : 나머지 마우스 오른쪽 버튼 클릭 옵션에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

색상 설정

색상 설정 옵션을 선택하면 "값에 대한 색상 선택" 창이 나타납니다.

그림 7.6 값에 대한 색상 선택 창



기본 모자이크 색상은 반응 열이 순서형인지 명목형인지, 그리고 기존의 "값 색상" 열 특성이 있는지에 따라 다릅니다. 수준 색상을 변경하려면 두 번째 색상 열의 타원을 클릭하고 새 색상을 선택합니다.

값에 대한 색상 선택 창 설명

매크로 다음과 같이 두 수준 간의 색상 그래디언트를 계산합니다.

- **끝 사이의 그래디언트** 옵션은 변수의 모든 수준에 그래디언트를 적용합니다.
- 수준 범위를 선택한 경우 **선택한 점 사이의 그래디언트** 옵션은 선택된 수준에 색상 그래디언트를 적용합니다. 선택하려는 수준 위에서 포인터를 드래그하거나 Shift 키를 누른 채 첫 번째 수준과 마지막 수준을 클릭하면 수준 범위를 선택할 수 있습니다.
- **색상 반전**을 선택하면 색상 순서가 반전됩니다.
- **이전 색상으로 되돌리기**를 선택하면 변경 사항이 실행 취소됩니다.

색상 테마 색상 테마를 기반으로 각 값의 색상을 변경합니다.

열에 색상 저장 이 체크박스를 선택하면 관련 데이터 테이블의 열에 새 열 특성 (**값 색상**) 이 추가됩니다. 데이터 테이블에서 이 특성을 편집하려면 **열 > 열 정보**를 선택합니다.

분할표

분할표는 이원 빈도 테이블입니다. 각 요인 수준마다 하나씩의 행과 각 반응 수준마다 하나씩의 열이 있습니다.

그림 7.7에 표시된 것과 같은 그림을 생성하려면 "분할 분석의 예"(231 페이지)에 설명된 방법을 따르십시오.

그림 7.7 분할표의 예

분할표				
country	size			
	개수	Large	Mediu m	Small
	총 %			
	열 %			
	행 %			
American		36	53	26
		11.88	17.49	8.58
		85.71	42.74	18.98
		31.30	46.09	22.61
European		4	17	19
		1.32	5.61	6.27
		9.52	13.71	13.87
		10.00	42.50	47.50
Japanese		2	54	92
		0.66	17.82	30.36
		4.76	43.55	67.15
		1.35	36.49	62.16
합계		42	124	137
		13.86	40.92	45.21

분할표의 경우 다음 사항에 유의하십시오.

- "개수", "총 %", "열 %" 및 "행 %"는 행 및 열 머리글이 있는 각 셀(예: "American" 및 "Large" 아래의 셀) 내의 데이터에 해당합니다.
- 마지막 열에는 각 행의 총 개수와 각 행의 총 백분율이 포함됩니다.
- 맨 아래 행에는 각 열의 총 개수와 각 열의 총 백분율이 포함됩니다.

그림 7.7에서 대형이면서 미국에서 생산된 자동차를 집중적으로 살펴보겠습니다. 다음 표에서는 분할표를 사용하여 이러한 자동차에 대해 내릴 수 있는 결론을 설명합니다.

표 7.1 분할표 예를 기반으로 한 결론

수치	설명	표에 사용된 라벨
36	대형이면서 미국에서 생산된 자동차의 수	개수
11.88%	대형이면서 미국에서 생산된 모든 자동차의 백분율 ^a	총 %
85.71%	미국에서 생산된 대형 자동차의 백분율 (36/42) ^b	열 %
31.30%	대형인 미국 자동차의 백분율 (36/115) ^c	행 %

표 7.1 분할표 예를 기반으로 한 결론 (계속)

수치	설명	표에 사용된 라벨
37.95%	미국에서 생산된 모든 자동차의 백분율 (115/303)	(없음)
13.86%	모든 대형 자동차의 백분율 (42/303)	(없음)

- a. 303 은 설문조사 대상 자동차의 총 대수입니다.
- b. 42 는 설문조사 대상 대형 자동차의 총 대수입니다.
- c. 115 는 설문조사 대상 미국 자동차의 총 대수입니다.

팁 : 분할표의 데이터를 표시하거나 숨기려면 " 분할표 " 옆의 빨간색 삼각형 메뉴에서 표시하거나 숨기려는 옵션을 선택하십시오 .

분할표 설명

개수 셀 빈도 , 빈도 소계 및 총 합계 (총 표본 크기) 입니다 .

총 % 총 합계 대비 셀 개수 및 소계의 백분율입니다 .

행 % 행 합계 대비 각 셀 개수의 백분율입니다 .

열 % 열 합계 대비 각 셀 개수의 백분율입니다 .

기대값 독립성을 가정한 상태에서 각 셀의 기대 빈도 (E) 입니다 . 해당 행 합계와 열 합계의 곱을 총 합계로 나뉘서 계산됩니다 .

편차 관측된 셀 빈도 (O) 에서 기대 셀 빈도 (E) 를 뺀 값입니다 .

셀 카이제곱 각 셀에 대해 $(O - E)^2 / E$ 로 계산된 카이제곱 값입니다 .

열 누적 누적 열 합계입니다 .

열 누적 % 누적 열 백분율입니다 .

행 누적 누적 행 합계입니다 .

행 누적 % 누적 행 백분율입니다 .

검정

" 검정 " 보고서에는 X 수준 간에 반응 수준 비율이 동일한지 여부를 확인하기 위한 두 가지 검정의 결과가 표시됩니다 .

그림 7.8 에 표시된 것과 같은 보고서를 생성하려면 " 분할 분석의 예 "(231 페이지) 에 설명된 방법을 따르십시오 .

그림 7.8 검정 보고서의 예

검정			
N	DF	-1*로그 우도	R ² (U)
303	4	36.309616	0.1200
검정	카이제곱	Prob>ChiSq	
우도비	72.619	<.0001*	
Pearson	66.313	<.0001*	

카이제곱 통계량의 경우 다음 사항에 유의하십시오 .

- 두 범주형 변수가 모두 반응 변수 (Y 변수) 인 경우 카이제곱 통계량은 두 변수가 독립적인지를 검증합니다 .
- Y 변수가 하나 있고 고정된 X 변수가 하나 있을 수도 있습니다 . 이 경우 카이제곱 통계량은 각 X 수준 간에 Y 변수의 분포가 동일한지를 검증합니다 .

검정 보고서 설명

N 관측값의 총 개수입니다 .

DF 검정과 관련된 자유도를 기록합니다 . 자유도는 $(c-1)(r-1)$ 와 같습니다 . 여기서 c 는 열 수이고 r 은 행 수입니다 .

-1* 로그 우도 적합 및 불확도를 측정하는 음의 로그 우도로 , 연속형 반응에서의 제곱합과 매우 유사합니다 .

R²(U) 모형 적합에 기인한 전체 불확실성의 비율입니다 .

- R^2 가 1 이면 요인이 범주형 반응을 완전히 예측함을 나타냅니다 .
- R^2 가 0 이면 고정된 기준 반응 비율 대신 모형을 사용함으로써 얻는 이득이 없음을 의미합니다 .

자세한 내용은 " [검정 보고서](#) "(257 페이지) 에서 확인하십시오 .

검정 각 표본 범주에서 반응 비율이 동일하다는 가설에 대한 카이제곱 통계 검정을 나열합니다 . 자세한 내용은 " [검정 보고서](#) "(257 페이지) 에서 확인하십시오 .

Prob>ChiSq 반응과 요인 사이에 아무런 관계가 없는 경우 카이제곱 값이 계산된 값보다 클 확률을 나열합니다 . 두 변수의 수준이 각각 두 개인 경우에는 한쪽 꼬리 검정과 양쪽 꼬리 검정에 대한 Fisher 정확 확률도 표시됩니다 .

Fisher 정확 검정

이 보고서에서는 2x2 테이블에 대한 Fisher 정확 검정 결과를 제공합니다 . 결과는 자동으로 2x2 테이블에 대해 표시됩니다 . Fisher 정확 검정과 $r \times c$ 테이블의 검정에 대한 자세한 내용은 " [정확 검정](#) "(246 페이지) 에서 확인하십시오 .

비율에 대한 평균 분석

참고 : 평균 분석 방법에 대한 자세한 내용은 Nelson et al. 연구 자료 (2005) 에서 확인하십시오 .
" 비율에 대한 평균 분석의 예 "(247 페이지) 도 확인하십시오 .

반응의 수준이 두 개인 경우 이 옵션을 사용하여 X 수준의 반응 비율을 전체 반응 비율과 비교할 수 있습니다 . 이 방법에서는 이항 분포에 대한 정규 근사를 사용합니다 . 따라서 표본 크기가 너무 작으면 결과에 경고가 표시됩니다 .

빨간색 삼각형 메뉴에는 다음과 같은 옵션이 표시됩니다 .

유의 수준 설정 분석에 사용되는 유의 수준을 선택합니다 .

요약 보고서 표시 반응 비율과 X 변수의 각 수준에 대한 결정 한계를 보여 주는 보고서를 생성합니다 . 이 보고서에서는 한계가 초과되었는지 여부를 나타냅니다 .

비율에 대한 반응 수준 전환 분석에 사용된 반응 범주를 변경합니다 .

표시 옵션 결정 한계 , 결정 한계 음영 , 중심선 및 점 옵션을 표시하거나 숨깁니다 .

대응 분석

참고 : " 대응 분석의 예 "(248 페이지) 도 확인하십시오 .

대응 분석은 빈도 테이블에서 유사한 개수 패턴을 가진 행 또는 열을 표시하기 위한 그래픽 기법입니다 . 대응 분석 그림에는 각 행과 각 열에 해당하는 점이 하나씩 있습니다 . 수준 수가 많아서 모자이크 그림으로는 유용한 정보를 도출하기 어려운 경우에 대응 분석을 사용합니다 .

대응 분석 그림 이해

행 프로파일은 행별 비율의 집합 , 즉 한 행의 개수들을 해당 행의 총 개수로 나눈 값들로 정의할 수 있습니다 . 두 행의 행 프로파일이 매우 유사한 경우에는 대응 분석 그림에서 해당 점들이 서로 가깝게 나타납니다 . 행 점 간의 거리를 제공한 값은 행 쌍 간의 동질성을 검정하는 카이제곱 거리에 대략적으로 비례합니다 .

문제는 대칭적으로 정의되므로 열 및 행 프로파일은 서로 비슷합니다 . 행 점과 열 점 간의 거리는 아무 의미도 없습니다 . 하지만 원점으로부터의 열 및 행 방향은 의미가 있으며 이 관계는 그림을 해석하는 데 도움이 됩니다 .

대응 분석 옵션

"대응 분석" 옆의 빨간색 삼각형 메뉴에 있는 옵션을 사용하여 3D 산점도를 생성하고 데이터 테이블에 열 특성을 추가할 수 있습니다.

3D 대응 분석 3D 산점도를 생성합니다.

값 순서화 저장 첫 번째 대응 점수 계수를 기준으로 정렬된 수준 순서를 X 및 Y 열 모두의 열 특성으로 저장합니다.

상세 정보 보고서

"상세 정보" 보고서에는 대응 분석에 대한 통계 정보가 포함되고 그림에 사용된 값이 표시됩니다.

특이값 분할표에 대한 특이값 분해를 제공합니다. 계산식은 "[대응 분석의 상세 정보 보고서](#)"(258 페이지)에서 확인하십시오.

관성 정준 차원의 측면을 고려한 상대적인 변동을 반영하는 특이값 제곱을 나열합니다.

비율 총 관성에 대한 관성 비율입니다.

누적 관성의 누적 비율을 보여 줍니다. 처음 두 개의 특이값에서 큰 관성이 발견되는 경우에는 2D 대응 분석 그림으로 테이블에서의 관계를 충분히 나타낼 수 있습니다.

X 변수 c1, c2, c3 대응 분석 그림에 표시된 값입니다 (그림 7.11).

Y 변수 c1, c2, c3 대응 분석 그림에 표시된 값입니다 (그림 7.11).

Cochran-Mantel-Haenszel 검정

참고 : "[Cochran Mantel Haenszel 검정의 예](#)"(250 페이지)도 확인하십시오.

Cochran-Mantel-Haenszel 검정에서는 세 번째 분류 기준에 따라 블록화한 후 두 범주형 변수 사이에 관계가 있는지 검정합니다.

점수 상관 Y 또는 X가 모두 순서형이거나 구간을 나타낼 때 사용할 수 있습니다. 대립가설은 블록화 변수의 최소 한 개 수준에서 Y와 X 사이에 선형 연관성이 있다는 것입니다.

열 범주별 행 점수 Y가 순서형이거나 구간을 나타낼 때 사용할 수 있습니다. 대립가설은 블록화 변수의 최소 한 개 수준에서 r 개 행의 평균 점수가 동일하지 않다는 것입니다.

행 범주별 열 점수 X가 순서형이거나 구간을 나타낼 때 사용할 수 있습니다. 대립가설은 블록화 변수의 최소 한 개 수준에서 c 개 열의 평균 점수가 동일하지 않다는 것입니다.

범주의 일반 연관성 블록화 변수의 최소 한 개 수준에서 X와 Y 사이에 일종의 연관성이 있는지를 검정합니다.

합치도 통계량

참고 : 통계 상세 정보는 "[합치도 통계량](#)"(256 페이지)에서 확인하십시오. "[합치도 통계량 옵션의 예](#)"(251 페이지)도 확인하십시오.

두 변수의 수준이 동일하면 **합치도 통계량** 옵션을 사용할 수 있습니다. 이 옵션은 카파 통계량 (Agresti 1990), 해당 표준 오차, 신뢰 구간, 가설 검정 및 Bowker 대칭성 검정을 표시합니다.

이 섹션에서 제공되는 카파 통계량과 관련 p 값은 근사값입니다. 정확한 합치도 통계량도 사용할 수 있습니다. 자세한 내용은 "[정확 검정](#)"(246 페이지)에서 확인하십시오.

카파 카파 통계량을 보여 줍니다.

표준 오차 카파 통계량의 표준 오차를 보여 줍니다.

95% 하한 카파에 대한 신뢰 구간의 하한점을 보여 줍니다.

95% 상한 카파에 대한 신뢰 구간의 상한점을 보여 줍니다.

Prob>Z 카파에 대한 단측 검정의 p 값을 보여 줍니다. 귀무가설은 카파가 0 인지를 검정합니다.

Prob>|Z| 카파에 대한 양측 검정의 p 값을 보여 줍니다.

카이제곱 Bowker 검정의 검정 통계량을 보여 줍니다. Bowker 대칭성 검정에서 귀무가설은 정방 테이블의 확률이 대칭성을 만족한다는 것, 즉 테이블 셀의 모든 쌍에 대해 $p_{ij}=p_{ji}$ 가 성립한다는 것입니다. X와 Y 모두 두 개의 수준이 있는 경우 이 검정은 McNemar 검정과 동일합니다.

Prob>ChiSq Bowker 검정의 p 값을 보여 줍니다.

상대 위험도

참고 : "[상대 위험도 옵션의 예](#)"(252 페이지)도 확인하십시오.

상대 위험도 옵션을 사용하여 2 x 2 분할표의 위험 비율을 계산할 수 있습니다. 보고서에는 신뢰 구간도 표시됩니다. 이 방법에 대한 자세한 내용은 Agresti 연구 자료 (1990, sect. 3.4.2)에서 확인하십시오.

상대 위험도 옵션을 선택하면 "상대 위험도 범주 선택" 창이 나타납니다. 단일 반응 및 요인 조합을 선택하거나, 반응 및 요인 수준의 모든 조합에 대한 위험 비율을 계산할 수 있습니다.

비율에 대한 2 표본 검정

참고 : "[비율에 대한 2 표본 검정의 예](#)"(254 페이지) 도 확인하십시오 .

X 및 Y 변수 모두의 수준이 두 개인 경우 두 비율 간의 차이에 대한 신뢰 구간을 확인할 수 있습니다 . 이 경우 신뢰 구간에 해당하는 검정도 계산됩니다 .

설명 수행되는 검정을 보여 줍니다 .

비율 차이 X 변수의 수준 간 비율 차이를 보여 줍니다 .

95% 하한 차이에 대한 신뢰 구간의 하한점을 보여 줍니다 . 조정 Wald 신뢰 구간을 기반으로 합니다 .

95% 상한 차이에 대한 신뢰 구간의 상한점을 보여 줍니다 . 조정 Wald 신뢰 구간을 기반으로 합니다 .

조정 Wald 검정 양쪽 꼬리 및 한쪽 꼬리 검정을 보여 줍니다 .

확률 검정의 p 값을 보여 줍니다 .

< 변수 > 반응의 관심 범주 검정에 사용할 반응 수준을 선택합니다 .

연관성 측도

참고 : "[연관성 측도 옵션의 예](#)"(255 페이지) 도 확인하십시오 .

연관성 측도 옵션을 선택하여 분할표의 변수 간 연관성을 설명하는 몇 가지 통계량을 확인할 수 있습니다 .

감마 부합 / 비부합 쌍의 개수를 기반으로 하며 동일 값 쌍은 무시합니다 . -1 ~ 1 범위의 값을 취합니다 .

Kendall Tau-b 감마와 유사하되, 동일 값에 대한 수정을 사용합니다 . -1 ~ 1 범위의 값을 취합니다 .

Stuart Tau-c 감마와 유사하되, 테이블 크기에 대한 조정과 동일 값에 대한 수정을 사용합니다 . -1 ~ 1 범위의 값을 취합니다 .

Somers D Tau-b 에 대한 비대칭 수정입니다 .

- CIR은 행 변수 X가 독립 변수로 간주되고 열 변수 Y가 종속 변수로 간주됨을 나타냅니다 .
- 마찬가지로 RIC 는 열 변수 Y가 독립 변수로 간주되고 행 변수 X가 종속 변수로 간주됨을 나타냅니다 .

Somers D 는 독립 변수에 동일 값을 가진 쌍이 있을 때만 동일 값에 대한 수정을 사용한다는 점에서 Tau-b 와 다릅니다 . -1 ~ 1 범위의 값을 취합니다 .

람다 비대칭 CIR의 경우와 RIC의 경우에 차이가 있습니다.

- CIR의 경우, 행 변수 X에 대해 열 변수 Y를 예측하는 데 있어서 가능한 항상 정도로 해석됩니다.
- RIC의 경우, 열 변수 Y에 대해 행 변수 X를 예측하는 데 있어서 가능한 항상 정도로 해석됩니다.

0~1 범위의 값을 취합니다.

람다 대칭 대략적으로 두 람다 비대칭 측도의 평균으로 해석됩니다. 0~1 범위의 값을 취합니다.

불확도 계수

- CIR의 경우, 행 변수 X로 설명되는 열 변수 Y의 불확실성 비율입니다.
- RIC의 경우, 열 변수 Y로 설명되는 행 변수 X의 불확실성 비율로 해석됩니다.

0~1 범위의 값을 취합니다.

불확도 계수 대칭 두 불확도 계수 측도의 대칭 버전입니다. 0~1 범위의 값을 취합니다.

참고:

- 각 통계량은 표준 오차 및 신뢰 구간과 함께 표시됩니다.
- 감마, Kendall Tau-b, Stuart Tau-c 및 Somers D는 X가 증가함에 따라 변수 Y가 증가하는 경향이 있는지 여부를 평가하는 순서형 연관성 측도입니다. 평가에 따라 관측값 쌍이 부합/비부합 쌍으로 분류됩니다. 관측값의 X 값이 커질수록 Y 값도 커지는 쌍은 부합 쌍입니다. 관측값의 X 값이 커질수록 Y 값이 작아지는 쌍은 비부합 쌍입니다. 이러한 측도는 두 변수가 모두 순서형일 경우에만 적합합니다.
- 람다 및 불확도 측도는 순서형 및 명목형 변수에 적합합니다.

연관성 측도 통계량에 대한 계산 상세 정보는 SAS Institute Inc. 가이드 (2018b)의 "FREQ Procedure" 장에서 확인하십시오. 다음 참조 자료에도 추가 정보가 포함되어 있습니다.

- Brown and Benedetti(1977)
- Goodman and Kruskal(1979)
- Kendall and Stuart(1979)
- Snedecor and Cochran(1980)
- Somers(1962)

Cochran Armitage 추세 검정

참고: "Cochran Armitage 추세 검정의 예"(255 페이지)도 확인하십시오.

이 Cochran Armitage 추세 검정에서는 단일 변수의 수준 간에 이항 비율의 추세를 검정합니다. 이 검정은 한 변수의 수준이 두 개이고 다른 변수는 순서형인 경우에만 적합합니다. 2수준 변수

는 반응을 나타내고 다른 변수는 순서형 수준이 있는 설명 변수를 나타냅니다. 귀무가설은 추세가 없다는 것으로, 설명 변수의 모든 수준에 대해 이항 비율이 동일함을 의미합니다.

이 검정에서 제공되는 검정 통계량과 p 값은 근사값입니다. 정확한 추세 검정도 사용할 수 있습니다. 자세한 내용은 "[정확 검정](#)"(246 페이지)에서 확인하십시오.

정확 검정

JMP^{PRO} 분할 분석 플랫폼에서는 다음과 같은 정확 검정을 사용할 수 있습니다.

Fisher 정확 검정 $r \times c$ 테이블에 대해 Fisher 정확 검정을 수행합니다. 이 검정은 두 변수 간의 연관성에 대한 검정입니다. Fisher 정확 검정에서는 행 및 열 합계가 고정되어 있다고 가정하고 초기화 분포를 사용하여 확률을 계산합니다.

이 검정은 대표본 분포 가정에 의존하지 않습니다. 따라서 표본 크기가 작은 경우나 희소 테이블의 경우처럼 우도비와 Pearson 검정의 신뢰도가 떨어지는 경우에 적합합니다.

이 보고서에는 다음과 같은 정보가 포함됩니다.

테이블 확률 (P) 관측된 테이블의 확률을 제공합니다. 이 값은 검정의 p 값이 아닙니다.

양측 확률 $\leq P$ 양측 검정의 p 값을 제공합니다.

2×2 테이블의 경우 한 행 또는 열에 포함된 값이 모두 0 인 경우 (이 경우 검정을 계산할 수 없음) 만 제외하고 자동으로 Fisher 정확 검정이 수행됩니다. 자세한 내용은 "[검정](#)"(239 페이지)에서 확인하십시오.

정확 Cochran Armitage 추세 검정 정확한 Cochran Armitage 추세 검정을 수행합니다. 이 검정은 한 변수의 수준이 두 개인 경우에만 사용할 수 있습니다. 추세 검정에 대한 자세한 내용은 "[Cochran Armitage 추세 검정](#)"(245 페이지)에서 확인하십시오.

정확 합치도 통계량 변수 간의 합치도를 검정하기 위한 정확 검정을 수행합니다. 이는 카파 통계량에 대한 정확 검정입니다. 이 검정은 두 변수의 수준이 동일한 경우에만 사용할 수 있습니다. 합치도 검정에 대한 자세한 내용은 "[합치도 통계량](#)"(243 페이지)에서 확인하십시오.

분할 분석 플랫폼의 추가 예

- "[비율에 대한 평균 분석의 예](#)"
- "[대응 분석의 예](#)"
- "[Cochran Mantel Haenszel 검정의 예](#)"
- "[합치도 통계량 옵션의 예](#)"
- "[상대 위험도 옵션의 예](#)"
- "[비율에 대한 2 표본 검정의 예](#)"

- "연관성 측도 옵션의 예"
- "Cochran Armitage 추세 검정의 예"

비율에 대한 평균 분석의 예

이 예에서는 한 지역에 있는 6 개 병원의 진료 예약 시간 준수 현황을 기록한 Office Visits.jmp 샘플 데이터 테이블을 사용합니다. 6 개 병원 각각에서 1 주간의 기록 중 60 건의 예약이 무작위로 선택되었습니다. 정시 내원으로 간주되기 위해서는 환자가 예약 시간으로부터 5 분 이내에 진료실로 안내되어야 합니다. 예약 시간 정시에 내원한 환자의 비율을 검토하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Office Visits.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. On Time 을 선택하고 **Y, 반응**을 클릭합니다.
4. Clinic 을 선택하고 **X, 요인**을 클릭합니다.
5. Frequency 를 선택하고 **빈도**를 클릭합니다.
6. **확인**을 클릭합니다.
7. "On Time 대 Clinic 의 분할 분석" 옆의 빨간색 삼각형을 클릭하고 **비율에 대한 평균 분석**을 선택합니다.
8. "비율에 대한 평균 분석" 옆의 빨간색 삼각형을 클릭하고 **요약 보고서 표시 및 비율에 대한 반응 수준 전환**을 선택합니다.

그림 7.9 비율에 대한 평균 분석의 예

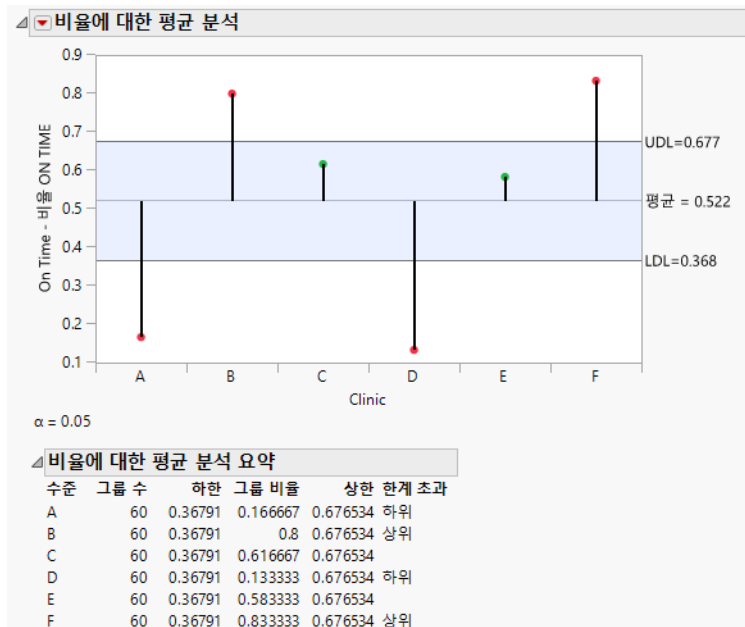


그림 7.9에서는 각 병원의 예약 시간 정시에 내원한 환자의 비율을 보여 줍니다. 그림 7.9에서는 다음을 확인할 수 있습니다.

- 정시 내원 비율은 F 병원의 경우가 가장 높고 그 다음은 B 병원입니다.
- D 병원은 정시 내원 비율이 가장 낮고 그 다음은 A 병원입니다.
- E 병원과 C 병원은 평균에 가까우며 결정 한계를 초과하지 않습니다.

대응 분석의 예

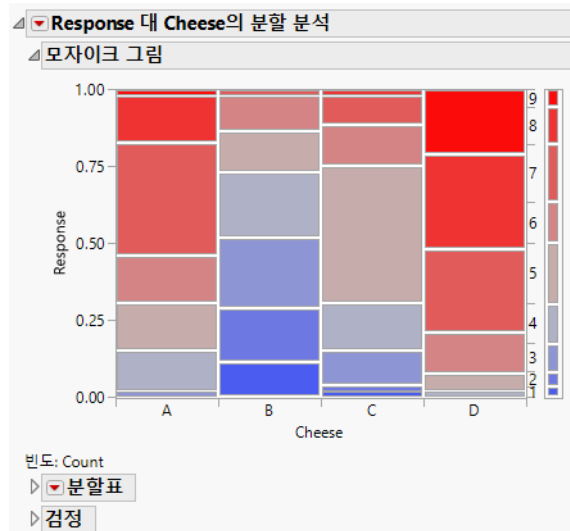
이 예에서는 Cheese.jmp 샘플 데이터 테이블을 사용합니다. 이 테이블에는 McCullagh and Nelder 연구 자료 (1989) 에서 보고된 Newell 치즈의 시식 평가 데이터가 포함되어 있습니다. 이 평가에서는 네 가지 치즈 첨가물에 대해 9 개 이상의 반응 수준 수를 기록했습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Cheese.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. Response 를 선택하고 **Y, 반응**을 클릭합니다.

Response 값의 범위는 1 에서 9 까지이며, 이 중 1 은 호감도가 가장 낮은 것이고 9 는 가장 높은 것입니다.

4. Cheese 를 선택하고 **X, 요인**을 클릭합니다.
A, B, C 및 D 는 4 가지 치즈 첨가물을 나타냅니다.
5. Count 를 선택하고 **빈도**를 클릭합니다.
6. **확인**을 클릭합니다.

그림 7.10 치즈 데이터에 대한 모자이크 그림



이 모자이크 그림에서는 분포가 서로 비슷하지 않음을 알 수 있습니다. 하지만 9 개의 수준에 대한 모자이크 그림을 제대로 파악하기는 어렵습니다. 대응 분석을 수행하면 이러한 경우의 관계를 쉽게 정의할 수 있습니다.

7. 대응 분석 그림을 표시하려면 "Response 대 Cheese 의 분할 분석" 옆의 빨간색 삼각형을 클릭하고 **대응 분석**을 선택합니다.

그림 7.11 대응 분석 그림의 예

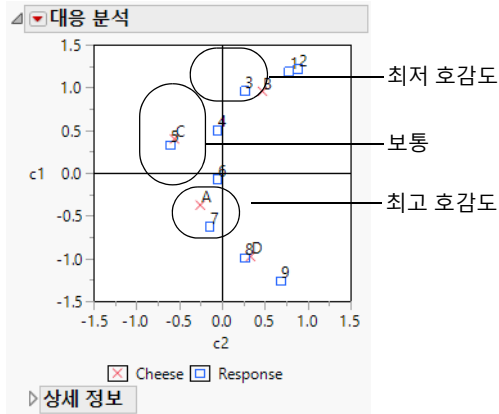
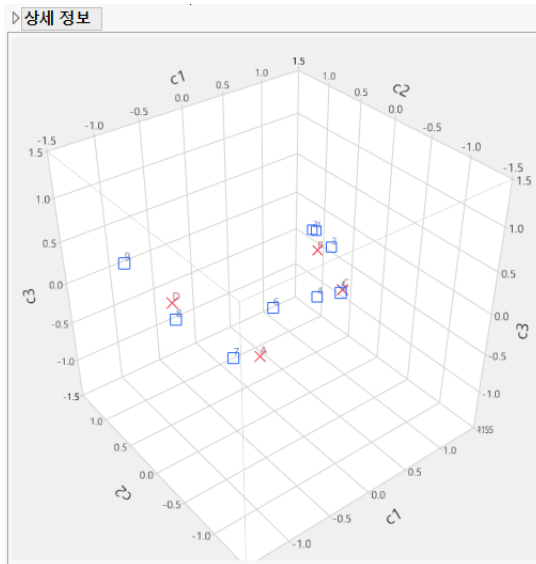


그림 7.11에서는 대응 분석을 그래프로 보여 줍니다. 이 그림의 축에는 "c1" 및 "c2" 라벨이 지정되어 있습니다. 다음을 확인할 수 있습니다.

- c1 은 일반적인 만족도와 일치하는 것으로 보입니다. c1 축의 치즈는 맨 위의 최저 호감도에 서 맨 아래의 최고 호감도로 변화합니다.
 - 치즈 D 는 반응이 8 및 9 로, 호감도가 가장 높습니다.
 - 치즈 B 는 반응이 1, 2 및 3 으로, 호감도가 가장 낮습니다.
 - 치즈 C 와 A 는 반응이 4, 5, 6 및 7 로 보통 호감도를 보입니다.
8. "대응 분석" 옆의 빨간색 삼각형을 클릭하고 **3D 대응 분석**을 선택합니다.

그림 7.12 3D 산점도의 예



다음을 확인할 수 있습니다.

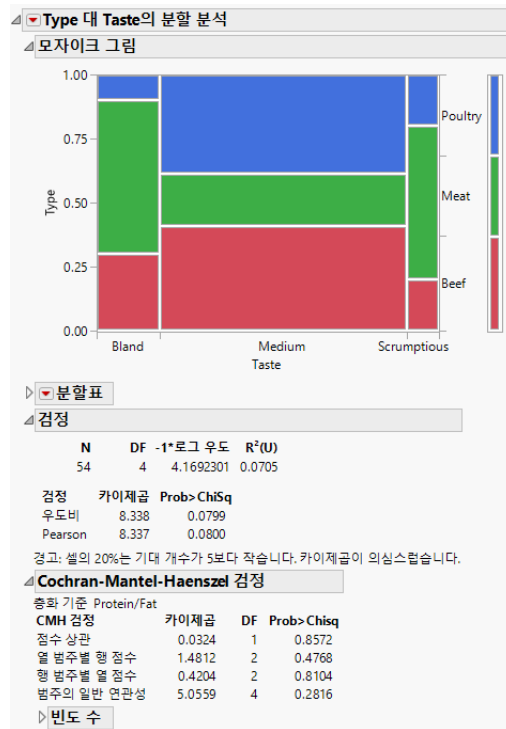
- c1 축을 보면 반응 1 부터 5 까지는 0 의 오른쪽 (양수) 에 나타납니다 . 반응 6 부터 9 까지는 0 의 왼쪽 (음수) 에 나타납니다 .
- c2 축을 보면 A 와 C 가 0 의 오른쪽 (양수) 에 나타납니다 . B 와 D 는 0 의 왼쪽 (음수) 에 나타납니다 .
- c1 은 일반적인 만족도 (최저 호감도부터 최고 호감도까지) 와 일치한다고 결론 내릴 수 있습니다 .

Cochran Mantel Haenszel 검정의 예

이 예에서는 Hot Dogs.jmp 샘플 데이터 테이블을 사용합니다 . 핫도그 유형과 맛 사이의 관계를 검토하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Hot Dogs.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Type 을 선택하고 **Y, 반응**을 클릭합니다 .
4. Taste 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Type 대 Taste 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **Cochran Mantel Haenszel** 을 선택합니다 .
7. 그룹화 변수로 Protein/Fat 을 선택하고 **확인**을 클릭합니다 .

그림 7.13 Cochran-Mantel-Haenszel 검정의 예



다음을 확인할 수 있습니다.

- "검정" 보고서에는 약 0.0799로 다소 유의한 카이제곱 확률이 표시됩니다. 이는 핫도그의 맛과 유형 간 관계에 어느 정도 유의성이 있음을 나타냅니다.
- "Cochran-Mantel-Haenszel 검정" 보고서에는 범주의 일반 연관성에 대한 p 값이 0.2816으로, 5%보다 훨씬 큰 것으로 표시됩니다.

합치도 통계량 옵션의 예

이 예에서는 Attribute Gauge.jmp 샘플 데이터 테이블을 사용합니다. 이 데이터는 세 명의 평가자가 50 개의 부품을 각각 세 번씩 평가한 결과를 제공합니다. 평가자 A와 B 사이의 관계를 검토하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Attribute Gauge.jmp를 엽니다.
2. **분석 > X로 Y 적합**을 선택합니다.
3. A를 선택하고 **Y, 반응**을 클릭합니다.
4. B를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.

그림 7.15 상대 위험도 범주 선택 창

위험 비율이 $P(Y=yj | X=xi) / P(Y=yj | X=xk)$ 입니다.

marital status 반응의 관심 범주: $Y=yj$

☒ Married
☐ Single

본자에 대한 표본 sex 범주: $X=xi$

☒ Female
☐ Male

☐ 모든 조합 계산

" 상대 위험도 범주 선택 " 창에서 다음 사항을 알 수 있습니다 .

- 단일 반응 및 요인 조합에만 관심이 있는 경우 여기에서 해당 조합을 선택할 수 있습니다 . 예를 들어 그림 7.15 의 창에서 **확인**을 클릭할 경우의 계산은 다음과 같습니다 .

$$\frac{P(Y = \text{Married} | X = \text{Female})}{P(Y = \text{Married} | X = \text{Male})}$$

- 모든 ($2 \times 2 = 4$) 반응 및 요인 수준 조합에 대해 위험 비율을 계산하려면 **모든 조합 계산** 체크박스를 선택합니다 (그림 7.16).
7. **모든 조합 계산** 체크박스를 선택하여 모든 조합을 요청합니다 . 다른 항목은 모두 기본 선택 상태대로 둡니다 .

그림 7.16 위험 비율 보고서의 예

marital status 대 sex의 분할 분석

모자이크 그림

분할표

		marital status		
		Married	Single	합계
sex	Female	95 31.35 48.47 68.84	43 14.19 40.19 31.16	138 45.54
	Male	101 33.33 51.53 61.21	64 21.12 59.81 38.79	165 54.46
합계		196 64.69	107 35.31	303

검정

상대 위험도

설명	상대 위험도	95% 하한	95% 상한
$P(\text{Married} \text{Female}) / P(\text{Married} \text{Male})$	1.124623	0.953193	1.326885
$P(\text{Married} \text{Male}) / P(\text{Married} \text{Female})$	0.889187	0.753645	1.049105
$P(\text{Single} \text{Female}) / P(\text{Single} \text{Male})$	0.803329	0.587179	1.099046
$P(\text{Single} \text{Male}) / P(\text{Single} \text{Female})$	1.24482	0.90988	1.703057

상대 위험도의 계산 방식을 보려면 다음 단계를 수행하십시오 .

1. " 상대 위험도 " 보고서의 첫 번째 항목 " $P(\text{Married} | \text{Female}) / P(\text{Married} | \text{Male})$ " 을 검토합니다 .
2. 해당 확률은 분할표에서 확인할 수 있습니다 . 이 확률은 테이블의 행별로 달라지는 두 개의 성별 수준을 기반으로 계산되므로 이를 판독하려면 다음과 같은 " 행 %" 를 사용합니다 .

$$P(\text{Married} | \text{Female}) = 0.6884$$

$$P(\text{Married} | \text{Male}) = 0.6121$$

따라서 계산은 다음과 같습니다 .

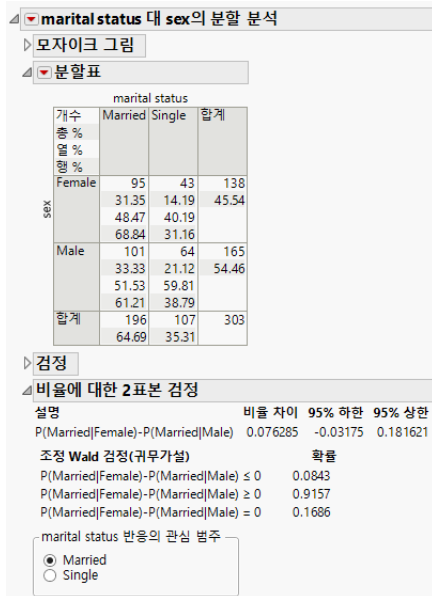
$$P(\text{Married} | \text{Female}) / P(\text{Married} | \text{Male}) = \frac{0.6884}{0.6121} = 1.1247$$

비율에 대한 2 표본 검정의 예

이 예에서는 Car Poll.jmp 샘플 데이터 테이블을 사용합니다 . 남성과 여성의 기혼 확률을 검토하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. marital status 를 선택하고 **Y, 반응**을 클릭합니다 .
4. sex 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "marital status 대 sex 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **비율에 대한 2 표본 검정**을 선택합니다 .

그림 7.17 비율에 대한 2 표본 검정 보고서의 예



이 예에서는 여성과 남성의 기혼 확률을 비교하려고 합니다 . " 분할표 " 의 " 행 % " 를 참조하여 다음을 구합니다 .

$$P(\text{Married} | \text{Female}) = 0.6884$$

$$P(\text{Married} | \text{Male}) = 0.6121$$

이 두 수치의 차이인 0.0763 이 보고서에 표시된 "비율 차이"입니다. 양측 신뢰 구간은 [-0.03175, 0.181621] 입니다. 조정 Wald 방법으로 구한 이 신뢰 구간의 p 값은 0.1686 으로, Pearson 카이제곱 검정으로 구한 p 값 (0.1665) 에 가깝습니다. 일반적으로 두 비율의 차이를 검정하는 데는 조정 Wald 검정보다 Pearson 카이제곱 검정이 더 널리 사용됩니다.

연관성 측도 옵션의 예

이 예에서는 Car Poll.jmp 샘플 데이터 테이블을 사용합니다. 남성과 여성의 기혼 확률을 검토하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. marital status 를 선택하고 **Y, 반응**을 클릭합니다.
4. sex 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "marital status 대 sex 의 분할 분석" 옆의 빨간색 삼각형을 클릭하고 **연관성 측도**를 선택합니다.

그림 7.18 연관성 측도 보고서의 예

측도	값	표준 오차	95% 하한	95% 상한
감마	0.1667	0.1184	-0.0654	0.3987
Kendall Tau-b	0.0795	0.0570	-0.0321	0.1911
Stuart Tau-c	0.0757	0.0543	-0.0307	0.1821
Somers D C R	0.0763	0.0547	-0.0309	0.1835
Somers D R C	0.0828	0.0593	-0.0335	0.1991
람다 비대칭 C R	0.0000	0.0000	0.0000	0.0000
람다 비대칭 R C	0.0000	0.0000	0.0000	0.0000
람다 비대칭	0.0000	0.0000	0.0000	0.0000
불확도 계수 C R	0.0049	0.0070	0.0000	0.0186
불확도 계수 R C	0.0046	0.0066	0.0000	0.0176
불확도 계수 대칭	0.0047	0.0068	0.0000	0.0181

검토하려는 변수 ("sex" 및 "marital status") 가 명목형이므로 람다 및 불확도 측도를 사용합니다. 이러한 측도가 모두 작으므로 성별과 결혼 여부 간에는 약한 연관성이 있는 것으로 보입니다.

Cochran Armitage 추세 검정의 예

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

이 검정을 위해 다음과 같이 "size" 를 순서형 변수로 변경하십시오.

2. "열" 패널에서 "size" 옆의 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **순서형**을 선택합니다.

3. **분석 > X 로 Y 적합**을 선택합니다.
4. **sex** 를 선택하고 **Y, 반응**을 클릭합니다.
5. **size** 를 선택하고 **X, 요인**을 클릭합니다.
6. **확인**을 클릭합니다.
7. "sex 대 size 의 분할 분석 " 옆의 빨간색 삼각형을 클릭하고 **Cochran Armitage 추세 검정**을 선택합니다.

그림 7.19 Cochran Armitage 추세 검정 보고서의 예

점근적 검정	Z	Prob<Z	Prob> Z
-0.37271	0.3547	0.7094	0.7094

양측 p 값 (0.7094) 이 크게 나타납니다. 따라서 서로 다른 크기의 자동차를 구매하는 남성과 여성의 비율에 관계가 있다고 결론 내릴 수 없습니다.

분할 분석 플랫폼에 대한 통계 상세 정보

- " 합치도 통계량 "
- " 승산비 옵션 "
- " 검정 보고서 "
- " 대응 분석의 상세 정보 보고서 "

합치도 통계량

두 개의 반응 변수를 n 개의 개체에 대한 두 개의 독립적 등급으로 볼 경우 평가자의 평가가 완전히 합치되면 카파 계수가 +1 이 됩니다. 관측된 합치도가 우연에 의한 합치 확률을 초과하면 카파 계수가 양수가 되며 그 크기는 합치 강도를 반영합니다. 실제로는 혼치 않지만 관측된 합치도가 우연에 의한 합치 확률보다 작을 경우에는 카파가 음수가 됩니다. 카파의 최소값은 주변 비율에 따라 -1 에서 0 사이입니다.

카파 계수는 다음과 같이 계산됩니다.

$$\hat{\kappa} = \frac{P_0 - P_c}{1 - P_c} \text{ 단, } P_0 = \sum_i p_{ii} \text{ 및 } P_c = \sum_i p_{i+} p_{+i}$$

p_{ij} 는 (i, j) 번째 셀에 있는 개체의 비율이므로 $\sum_i \sum_j p_{ij} = 1$ 입니다.

단순 카과 계수의 점근 분산은 다음과 같이 추정됩니다.

$$\text{var} = \frac{A+B-C}{(1-P_c)^2 n} \quad \text{단, } A = \sum_i p_{ii}[1-(p_{i+}+p_{+i})(1-\hat{\kappa})]^2, B = (1-\hat{\kappa})^2 \sum_{i \neq j} p_{ij}(p_{+i}+p_{j+})^2 \quad \text{및}$$

$$C = [\hat{\kappa} - P_c(1-\hat{\kappa})]^2$$

자세한 내용은 Cohen 연구 자료 (1960) 및 Fleiss et al. 연구 자료 (1969) 에서 확인하십시오.

Bowker 대칭성 검정에서 귀무가설은 2×2 테이블의 확률이 대칭성을 만족 ($p_{ij}=p_{ji}$) 한다는 것입니다.

승산비 옵션

승산비는 다음과 같이 계산됩니다.

$$\frac{p_{11} \times p_{22}}{p_{12} \times p_{21}}$$

여기서 p_{ij} 는 2×2 테이블의 i 번째 행, j 번째 열의 개수입니다.

검정 보고서

$R^2(U)$

$R^2(U)$ 는 다음과 같이 계산됩니다.

$$\frac{-\log \text{likelihood for Model}}{-\log \text{likelihood for Corrected Total}}$$

음의 로드 우도 합계는 전체 표본에서 고정된 반응 비율을 적합시켜 구합니다.

검정

두 개의 카이제곱 검정은 다음과 같습니다.

우도비 카이제곱 검정 통계량은 "검정" 테이블의 모형에 대한 음의 로그 우도의 2 배로 계산됩니다. 일부 책에서는 이 통계량에 G^2 표기를 사용합니다. 두 개의 음의 로그 우도 (전체 모집단

반응 확률을 사용한 로그 우도와 각 모집단 반응 비율을 사용한 로그 우도) 사이의 차이는 다음과 같습니다.

$$G^2 = 2 \left[\sum_{ij} (-n_{ij}) \ln(p_j) - \sum_{ij} -n_{ij} \ln(p_{ij}) \right] \text{ 단, } p_{ij} = \frac{n_{ij}}{N} \text{ 및 } p_j = \frac{N_j}{N}$$

이 계산식은 간단히 다음과 같이 쓸 수 있습니다.

$$G^2 = 2 \sum_i \sum_j n_{ij} \ln \left(\frac{n_{ij}}{e_{ij}} \right)$$

Pearson 카이제곱 검정 통계량은 관측된 셀 개수와 기대 셀 개수 사이의 차이를 제공한 후 합해서 계산합니다. Pearson 카이제곱 검정은 표본 크기가 매우 클 경우 빈도 수는 정규 분포를 따르는 경향이 있다는 특성을 이용합니다. 이 카이제곱 통계량의 친숙한 형태는 다음과 같습니다.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

여기서 O는 관측된 셀 개수이고 E는 기대 셀 개수입니다. 합계에는 모든 셀이 포함됩니다. 여기서는 2x2 테이블에서 종종 수행되는 연속성 수정이 수행되지 않습니다.

대응 분석의 상세 정보 보고서

다음 방정식의 특이값을 나열합니다.

$$D_r^{-0.5} (P - rc') D_c^{-0.5}$$

다음은 각 요소에 대한 설명입니다.

- P - 개수를 총 빈도로 나눈 값의 행렬입니다.
- r 및 c - P의 행 및 열 합계입니다.
- D - r 및 c 값의 주대각 행렬입니다.

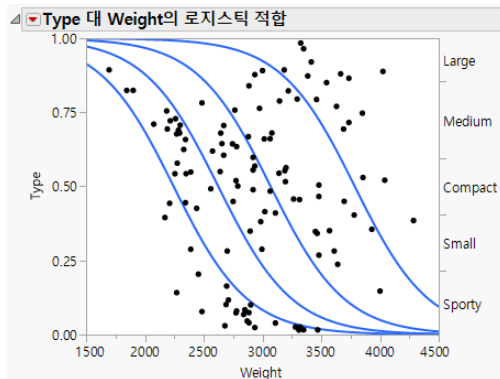
로지스틱 분석

범주형 Y 변수와 연속형 X 변수 간의 관계 분석

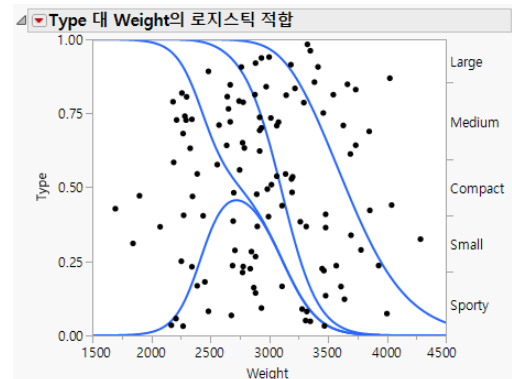
로지스틱 플랫폼에서는 반응 범주에 대한 확률을 연속형 x 예측 변수에 적합시킵니다. 적합된 모형은 각 x 값의 확률을 추정합니다. 로지스틱 플랫폼은 X로 Y 적합 플랫폼의 **명목형** 또는 **순서형 대 연속형** 분석법입니다. 이 플랫폼에서 명목형 반응과 순서형 반응은 다음과 같이 구분됩니다.

- 명목형 로지스틱 회귀는 반응 간에 확률을 분할하기 위한 일련의 곡선을 추정합니다.
- 순서형 로지스틱 회귀는 특정 반응보다 작거나 같은 확률을 모델링합니다. 이는 정렬된 범주에 대한 확률을 산출하기 위해 가로로 이동된 단일 로지스틱 곡선을 추정하는 효과가 있습니다. 이 모형은 덜 복잡하며 순서형 반응에 권장됩니다.

그림 8.1 로지스틱 회귀의 예



순서형 로지스틱 회귀



명목형 로지스틱 회귀

목차

로지스틱 회귀의 개요	261
명목형 로지스틱 회귀	261
순서형 로지스틱 회귀	261
명목형 로지스틱 회귀의 예	262
로지스틱 플랫폼 시작	263
데이터 구조	264
로지스틱 보고서	265
로지스틱 그림	266
반복	266
전체 모형 검정	266
적합 상세 정보	267
모수 추정값	268
로지스틱 플랫폼 옵션	269
ROC 곡선	270
확률 계산식 저장	271
역추정 예측	271
로지스틱 회귀의 추가 예	271
순서형 로지스틱 회귀의 예	272
로지스틱 그림의 추가 예	273
ROC 곡선의 예	275
십자기호 도구를 사용한 역추정 예측의 예	276
역추정 예측 옵션을 사용한 역추정 예측의 예	277
로지스틱 플랫폼에 대한 통계 상세 정보	279

로지스틱 회귀의 개요

로지스틱 회귀는 오랜 전통을 갖고 있으며 투약 반응 데이터나 구매 선택 데이터를 모델링하는 등 매우 다양하게 적용되고 있습니다. 하지만 많은 통계 입문 과정에서 이 매우 단순한 방법을 다루지 않고 있습니다. 로지스틱 회귀 분야의 교재 (Hosmer and Lemeshow 1989) 와 범주형 통계 분야의 많은 교재 (Agresti 1990) 에서 이를 다루고 있습니다. 일부 분석가는 이 방법을 다른 분포 함수, 즉 정규 분포 함수와 함께 사용합니다. 이러한 분석을 프로빗 분석이라고 합니다. 일부 분석가는 연속형 변수를 Y 로 간주하고 범주를 X 로 간주해서 반대로 작업하는 것을 선호하므로 로지스틱 회귀 대신 판별 분석을 사용합니다. 하지만 판별 분석에서는 연속형 데이터가 고정된 회귀변수가 아니라 정규 분포를 따르는 확률 반응이라고 가정합니다.

단순 로지스틱 회귀는 모형 적합 플랫폼에서 사용되는 범주형 반응에 대한 일반적인 기능을 단순화하고 그래픽 정보를 더 포함한 것입니다. 보다 복잡한 로지스틱 회귀 모형의 예는 Fitting Linear Models 의 "Logistic Regression Models" 장에서 확인하십시오.

명목형 로지스틱 회귀

명목형 로지스틱 회귀에서는 반응 수준 중 하나를 선택할 확률을 x 요인에 대한 평활 함수로 추정합니다. 적합한 확률은 0 에서 1 사이여야 하며, 특정 요인 값에 대한 반응 수준 간의 합계는 1 이 되어야 합니다.

로지스틱 확률도에서 세로 축은 확률을 나타냅니다. 반응 수준이 k 개인 경우 $k-1$ 개의 평활 곡선으로 전체 확률 (=1) 이 반응 수준 간에 분할됩니다. 로지스틱 회귀의 적합 원칙은 발생하는 반응 사건에 적합한 확률 (즉, 최대 우도) 의 음의 자연 로그 합을 최소화하는 것입니다.

순서형 로지스틱 회귀

Y 가 순서형인 경우 적합에는 수정된 로지스틱 회귀가 사용됩니다. 각 반응 수준과 같은 수준 또는 그 아래 수준의 반응이 나타날 누적 확률은 곡선으로 모델링됩니다. 이 곡선들은 오른쪽 또는 왼쪽으로 이동된다는 점만 제외하고는 각 수준마다 동일합니다.

순서형 로지스틱 모형은 $r-1$ 개 (여기서 r 은 반응 수준의 개수) 의 누적 로지스틱 비교에 대해 각각 다른 절편과 동일한 기울기를 적합시킵니다. 각 모수 추정값을 개별적으로 검토하고 검토할 수 있지만 이는 거의 필요하지 않습니다.

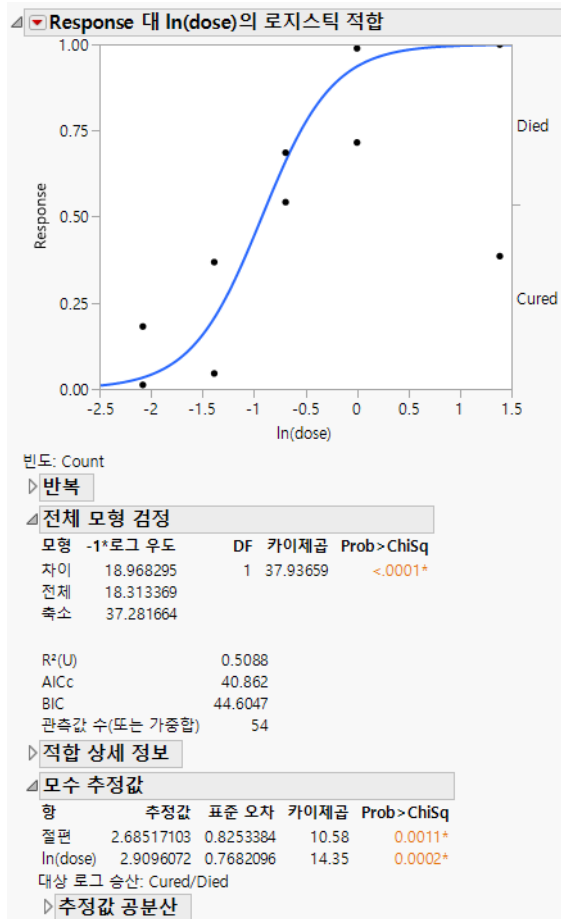
명목형 모형이라도 추정할 모수가 적어서 순서형 모형이 적합한 경우에는 순서형 모형을 사용하는 것이 좋습니다. 사실상 반응 수준이 수백 개인 경우에는 순서형 반응을 적합시키는 것이 실용적입니다.

명목형 로지스틱 회귀의 예

이 예에서는 Penicillin.jmp 샘플 데이터 테이블을 사용합니다. 이 예의 데이터는 각각 12 마리의 토끼를 포함하는 5 개 그룹을 대상으로 연쇄상구균 박테리아를 주입한 실험에서 얻은 결과입니다. 토끼의 신체 조직이 박테리아에 감염된 것으로 확인된 후에는 페니실린을 용량을 달리 해서 투여했습니다. 투여량의 자연 로그 ($\ln(\text{dose})$) 가 토끼의 치료 여부에 영향을 미치는지 확인하려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Penicillin.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. Response 를 선택하고 **Y, 반응**을 클릭합니다.
4. $\ln(\text{dose})$ 을 선택하고 **X, 요인**을 클릭합니다.
빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.
5. **확인**을 클릭합니다.

그림 8.2 명목형 로지스틱 보고서의 예



그림에는 적합된 모형이 $\ln(\text{dose})$ 의 함수로 표시됩니다. 적합된 모형은 예측 치료 확률입니다. p 값이 매우 크며, 이는 투여량이 토끼의 치료 여부에 상당한 영향을 미쳤음을 나타냅니다.

팁: 분석 대상 반응 수준을 변경하려면 시작 창에서 "목표 수준"을 지정하거나, "값 순서" 열 특성을 사용하십시오.

로지스틱 플랫폼 시작

로지스틱 분석을 수행하려면 다음을 수행하십시오.

1. 분석 > X로 Y 적합을 선택합니다.

2. **Y, 반응**에 명목형 또는 순서형 열을 입력합니다.
3. **X, 요인**에 연속형 열을 입력합니다.

그림 8.3 X로 Y 적합 시작 창

두 변수 사이의 관계를 모델링합니다.

다이아그램 위에 "로지스틱" 이라고 표시되면 로지스틱 분석을 수행할 것임을 나타냅니다.

참고 : "JMP 시작하기" 창에서 로지스틱 분석을 시작할 수도 있습니다. **보기 > JMP 시작하기 > 기본 분석 > 로지스틱**을 선택합니다.

반응이 이항 형식이고 명목형 모델링 유형이면 시작 창에 "목표 수준" 메뉴가 나타납니다. 확률을 모델링할 반응 수준을 지정하려면 이 메뉴를 사용합니다.

X로 Y 적합 시작 창에 대한 자세한 내용은 "[X로 Y 적합 소개](#)" 장 (105 페이지) 에서 확인하십시오. "열 선택" 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 "시작하기" 장에서 확인하십시오.

데이터 구조

데이터는 요약되지 않은 데이터나 요약된 데이터로 구성될 수 있습니다.

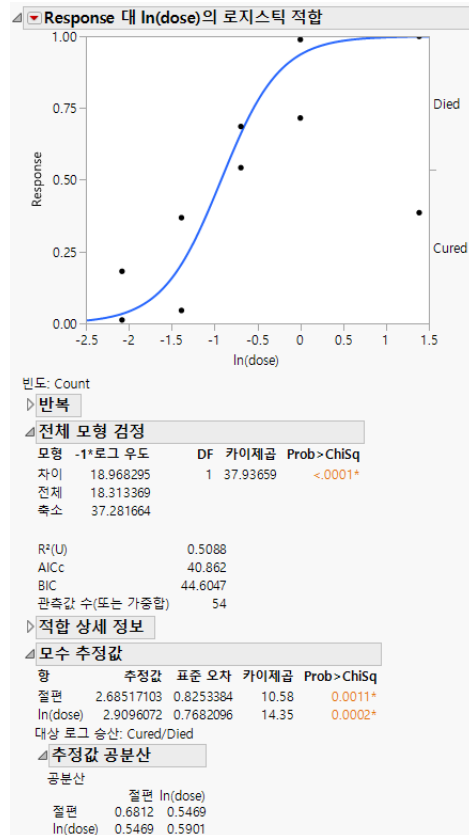
요약되지 않은 데이터 각 관측값마다 X 값과 해당하는 Y 값이 들어 있는 행이 하나씩 있습니다.

요약된 데이터 각 행이 공통된 X 및 Y 값을 갖는 일련의 관측값을 나타냅니다. 데이터 테이블에는 각 행의 개수를 제공하는 열이 포함되어야 합니다. 시작 창에서는 이 열을 "빈도" 열로 입력합니다.

로지스틱 보고서

그림 8.4 에 표시된 것과 같은 그림을 생성하려면 "명목형 로지스틱 회귀의 예"(262 페이지)에 설명된 방법을 따르십시오.

그림 8.4 로지스틱 보고서의 예



로지스틱 보고서 창에는 로지스틱 그림, "반복" 보고서, "전체 모형 검정" 보고서, "적합 상세 정보" 보고서 및 "모수 추정값" 보고서가 포함됩니다.

참고: 빨간색 삼각형 메뉴에서는 초기 보고서 창에 추가할 수 있는 추가 옵션을 제공합니다. 자세한 내용은 "로지스틱 플랫폼 옵션"(269 페이지)에서 확인하십시오.

로지스틱 그림

참고 : " 로지스틱 그림의 추가 예 "(273 페이지) 도 확인하십시오 .

로지스틱 확률도에서는 로지스틱 모형의 적합 상태를 보여 줍니다 . 가로 축의 각 값에서 세로 방향의 확률 척도가 각 반응 범주의 확률로 분할됩니다 . 이 확률은 곡선 사이의 세로 거리로 측정되며 모든 반응 범주 확률의 합계는 1 이 됩니다 .

로지스틱 그림에서 점은 데이터 테이블의 관측값을 나타냅니다 . 각 점의 가로 위치는 연속형 요인의 값에 따라 결정됩니다 . 각 점의 세로 위치는 반응 범주의 값에 해당하는 곡선 사이에 오도록 무작위로 선택됩니다 . 이렇게 점을 지터링하면 점들이 가장 조밀한 곳을 보다 쉽게 확인할 수 있지만 이 경우 세로 위치는 세로 축의 값과 일치하지 않습니다 . 고정된 난수 시드값이 사용되므로 세로 위치는 동일한 모형을 여러 번 적합시켜도 달라지지 않습니다 .

관련 데이터 테이블의 " 열 " 패널에서 변수를 클릭하여 축으로 드래그하는 방법으로 그림의 변수를 바꿀 수 있습니다 .

반복

" 반복 " 보고서에서는 각 반복과 모형 수렴 여부를 결정하는 평가 기준을 보여 줍니다 . " 반복 " 보고서는 명목형 로지스틱 회귀의 경우에만 표시됩니다 .

전체 모형 검정

" 전체 모형 검정 " 보고서에서는 모형이 상수 반응 확률보다 더 잘 적합시키는지 여부를 보여 줍니다 . 이 보고서는 연속형 반응 모형에 대한 분산 분석 보고서와 유사합니다 . 범주형 모형이 데이터를 얼마나 잘 적합시키는지 평가하는 것은 특정 우도비 카이제곱 검정입니다 .

관측된 확률의 음의 자연 로그 합을 음의 로그 우도 ($-1 \times \text{로그 우도}$) 라고 합니다 . 범주형 데이터에 대한 음의 로그 우도는 연속형 데이터에서의 제곱합과 같은 역할을 합니다 . 즉, 데이터에 의해 적합된 모형과 동일 확률의 모형에서 얻은 음의 로그 우도의 차이에 2 를 곱하면 카이제곱 통계량이 됩니다 . 이 검정 통계량은 x 변수가 반응에 아무 영향도 미치지 않는다는 가설을 검정합니다 .

$R^2(U)$ (일부 경우에는 R^2 로 표기) 값은 0 에서 1 사이입니다 . R^2 값이 높다는 것은 모형 적합이 적절하다는 것을 나타내며 , 범주형 모델에서는 이 값이 높은 경우가 거의 없습니다 .

" 전체 모형 검정 " 보고서에는 다음 열이 포함됩니다 .

모형 일부 경우에는 " 소스 " 라고 합니다 .

- **축소** 모형에는 절편만 포함됩니다 .
- **전체** 모형에는 절편뿐만 아니라 모든 효과도 포함됩니다 .
- **차이**는 전체 모형과 축소 모형의 로그 우도 차이입니다 .

DF 모형과 관련된 자유도를 기록합니다 .

-1* 로그 우도 표본 내의 변동 (일부 경우에는 불확도라고 함) 을 측정합니다.

전체 (전체 모형) 는 모형 적합 후에 계산된 음의 로그 우도 (또는 불확도) 입니다. 적합 프로세스에서는 선형 모형과 로지스틱 반응 함수로 반응 비율을 예측합니다. 이 값은 적합 프로세스에 의해 최소화됩니다.

축소 (축소 모형) 는 고정된 기준 비율에 의해 확률이 추정될 경우에 대한 음의 로그 우도 (또는 불확도) 입니다. 이는 모형에 효과가 없는 경우의 기준 불확도입니다.

이 두 개의 음의 로그 우도의 차이는 모형 적합으로 인한 감소분입니다. 이 값의 두 배는 우도비 카이제곱 검정 통계량입니다.

자세한 내용은 **Fitting Linear Models** 의 부록 "Statistical Details" 에서 확인하십시오.

카이제곱 모형이 전체 표본에서 고정된 반응 비율보다 더 잘 적합된다는 가설에 대한 우도비 카이제곱 검정입니다. 이 값은 차이 모형에 대한 -1* 로그 우도의 두 배입니다. 즉, 두 개의 음의 로그 우도 (전체 모 집단 반응 확률을 사용한 로그 우도와 각 모 집단 반응 비율을 사용한 로그 우도) 사이의 차이에 2 를 곱한 것입니다. 자세한 내용은 "[로지스틱 플랫폼에 대한 통계 상세 정보](#)" (279 페이지) 에서 확인하십시오.

Prob>ChiSq 카이제곱 검정의 관측된 유의 확률 (p 값이라고도 함) 입니다. 이 값은 카이제곱 값이 계산된 값보다 클 확률입니다. 대개 이 확률이 0.05 미만이면 모형이 유의한 것으로 판단됩니다.

R²(U) 모형 적합에 기인한 전체 불확실성의 비율로, **차이**의 음의 로그 우도 값을 **축소**의 음의 로그 우도 값으로 나눈 값으로 정의됩니다. R²(U) 값이 1 이면 발생하는 사건의 예측 확률이 1 과 같음을 나타냅니다. 즉, 예측 확률에 불확실성이 없는 것입니다. 로짓 모형의 경우 예측 확률은 거의 확실하지 않으므로 R²(U) 가 작은 경향이 있습니다. 자세한 내용은 "[로지스틱 플랫폼에 대한 통계 상세 정보](#)" (279 페이지) 에서 확인하십시오.

참고 : R²(U) 를 McFadden 유사 R² 라고도 합니다.

AICc 수정된 Akaike 정보 기준입니다. 자세한 내용은 **Fitting Linear Models** 의 부록 "Statistical Details" 에서 확인하십시오.

BIC 베이시안 정보 기준입니다. 자세한 내용은 **Fitting Linear Models** 의 부록 "Statistical Details" 에서 확인하십시오.

관측값 일부 경우에는 가중합이라고 합니다. 계산에 사용된 전체 표본 크기입니다. **가중치** 변수를 지정한 경우에는 가중치의 합계입니다.

적합 상세 정보

"적합 상세 정보" 보고서에는 다음과 같은 통계량이 포함됩니다.

측도 다음과 같은 적합 측도가 포함됩니다.

엔트로피 R² 적합된 모형과 상수 확률 모형의 로그 우도를 비교합니다. R²(U) 와 동일합니다. 자세한 내용은 "[로지스틱 플랫폼에 대한 통계 상세 정보](#)" (279 페이지) 에서 확인하십시오.

일반화 R^2 일반 회귀 모형에 적용할 수 있는 척도입니다. 이 값은 우도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다. 일반화 R^2 척도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2 로 단순화됩니다. 일반화 R^2 를 Nagelkerke 또는 Craig and Uhler R^2 라고도 하는데, 이는 Cox-Snell 유사 R^2 를 정규화한 것입니다. 자세한 내용은 Nagelkerke 연구 자료 (1991) 에서 확인하십시오.

평균 -Log p $-\log(p)$ 의 평균입니다. 여기서 p 는 발생한 사건과 연관된 적합 확률입니다.

RMSE 제공된 평균 제곱 오차입니다. 여기서 차이는 반응과 p (실제로 발생한 사건의 적합 확률) 사이의 차이입니다.

평균 절대 편차 반응과 p (실제로 발생한 사건의 적합 확률) 사이의 차이에 대한 절대값을 평균한 것입니다.

오분류 비율 적합 확률이 가장 높은 반응 범주가 관측된 범주가 아닌 비율입니다.

엔트로피 R^2 및 일반화 R^2 의 경우 값이 1 에 가까울수록 더 나은 적합을 나타냅니다. 평균 -Log p, RMSE, 평균 절대 편차 및 오분류 비율의 경우 값이 작을수록 더 나은 적합을 나타냅니다.

훈련 적합 척도의 값입니다.

정의 적합 척도의 대수 정의입니다.

모수 추정값

명목형 로지스틱 모형은 $k-1$ 개의 로지스틱 비교 각각에 대해 절편 및 기울기 모수를 적합시킵니다. 여기서 k 는 반응 수준의 개수입니다. " 모수 추정값 " 보고서에는 이러한 추정값이 나열됩니다. 각 모수 추정값을 개별적으로 검토하고 검증할 수 있지만 이는 거의 필요하지 않습니다.

항 로지스틱 모형의 각 모수를 나열합니다. 반응 변수의 마지막 수준을 제외하고 각 수준별로 요인에 대한 절편 항과 기울기 항이 하나씩 있습니다.

추정값 로지스틱 모형으로 구한 모수 추정값을 나열합니다.

표준 오차 각 모수 추정값의 표준 오차를 나열합니다. 이 값은 각 항을 0 과 비교하는 통계 검정을 계산하는 데 사용됩니다.

카이제곱 각 모수가 0 이라는 가설에 대한 Wald 검정을 나열합니다. Wald 카이제곱은 (추정값 / 표준 오차)² 으로 계산됩니다.

Prob>ChiSq 카이제곱 검정의 관측된 유의 확률을 나열합니다.

추정값 공분산

모수 추정값의 추정 분산과 모수 추정값 간의 추정 공분산을 보고합니다. 분산 추정값의 제공된 값은 **표준 오차** 섹션에 제공된 값과 동일합니다.

로지스틱 플랫폼 옵션

참고 : " 적합 그룹 " 메뉴는 Y 변수를 여러 개 지정한 경우에 표시됩니다 . 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 를 기준으로 정렬할 수 있습니다 . 자세한 내용은 **Fitting Linear Models** 의 "Standard Least Squares Report and Options" 장에서 확인하십시오 .

승산비 " 모수 추정값 " 보고서에 승산비를 추가합니다 . 자세한 내용은 **Fitting Linear Models** 의 "Logistic Regression Models" 장에서 확인하십시오 .

이 옵션은 수준이 두 개인 반응에만 사용할 수 있습니다 .

역추정 예측 특정 y 값으로부터 x 값을 예측합니다 . 자세한 내용은 " **역추정 예측** "(271 페이지) 에서 확인하십시오 .

이 옵션은 2 수준 명목형 반응에만 사용할 수 있습니다 .

로지스틱 그림 로지스틱 그림을 표시하거나 숨깁니다 .

그림 옵션 다음 옵션이 포함되어 있습니다 .

점 표시 로지스틱 그림에 점을 표시하거나 숨깁니다 .

비율 곡선 표시 각 x 값마다 점이 여러 개 있는 경우에만 유용합니다 . 이 경우 각 값의 적절한 비율 추정값을 구하고 이 비율을 적합한 로지스틱 곡선과 비교할 수 있습니다 . 퇴화 점 (일반적으로 0 또는 1 일 때) 이 너무 많아지지 않도록 JMP에서는 x 값에 점이 세 개 이상 있는 경우에 비율 값만 표시합니다 .

선 색상 그림의 곡선 색상을 선택할 수 있습니다 .

ROC 곡선 ROC(Receiver Operating Characteristic) 곡선을 생성합니다 . ROC 곡선은 각 x 값에 대한 민감도 대 (1 - 특이도) 의 그림입니다 . 자세한 내용은 " **ROC 곡선** "(270 페이지) 에서 확인하십시오 .

Lift 곡선 모형에 대한 Lift 곡선을 생성합니다 . Lift 곡선은 ROC 곡선과 동일한 정보를 보여 주되 , 시작점에서 정렬의 풍부도를 극화하는 방법을 사용합니다 . 세로 축은 모집단 중 해당 부분이 해당 반응 수준 전체의 비율에 비해 선택된 반응 수준에 얼마나 풍부하게 포함되었는지를 보여 줍니다 . Lift 곡선에 대한 자세한 내용은 **Fitting Linear Models** 의 "Logistic Regression Models" 장에서 확인하십시오 .

확률 계산식 저장 계산식을 포함하는 새 데이터 테이블 열을 생성합니다 . 자세한 내용은 " **확률 계산식 저장** "(271 페이지) 에서 확인하십시오 .

다음 옵션에 대한 자세한 내용은 의 "JMP 보고서 " 장에서 확인하십시오 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다. 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다.

ROC 곡선

참고: "[ROC 곡선의 예](#)"(275 페이지)도 확인하십시오.

예를 들어 진단 측도인 x 값이 있고 다음을 나타내는 x 임계값을 결정하려고 한다고 가정해 보겠습니다.

- x 값이 임계보다 크면 조건이 존재하는 것입니다.
- x 값이 임계보다 작으면 조건이 존재하지 않는 것입니다.

예를 들어 암 유형을 예측하기 위한 진단 검사로 혈액 성분 수준을 측정할 수 있습니다. 임계가 다양하고 그로 인해 False Positive(가양성) 및 False Negative(가음성)가 늘어나거나 줄어들 수 있기 때문에 진단 검사를 검토합니다. 그런 다음 해당 비율을 그림으로 표시합니다. 이상적인 상태는 True Negative(진음성)와 True Positive(진양성)를 가장 잘 구분하는 x 기준 값의 범위가 매우 좁은 것입니다. ROC(Receiver Operating Characteristic) 곡선은 이 전환이 얼마나 빠르게 일어나는지를 보여 줍니다. ROC 곡선의 목적은 곡선 아래 면적을 최대화하는 진단 기준을 얻는 것입니다.

의학에서 사용되는 두 가지 표준 정의는 다음과 같습니다.

- **민감도** - 특정 x 값(검사 또는 측정)이 존재하는 조건을 올바르게 예측할 확률입니다. 특정 x 가 조건의 존재를 잘못 예측할 확률은 $1 - \text{민감도}$ 입니다.
- **특이도** - 검사를 통해 조건이 존재하지 않음을 올바르게 예측할 확률입니다.

ROC 곡선은 각 x 값에 대한 민감도 대 ($1 - \text{특이도}$)의 그림입니다. ROC 곡선 아래의 면적은 곡선에 포함된 정보를 요약하는 데 사용되는 일반적인 지표입니다.

이항 결과에 대해 단순 로지스틱 회귀를 수행할 경우 해당 분석에 대한 ROC 곡선을 요청하는 플랫폼 옵션을 사용할 수 있습니다. **ROC 곡선** 옵션을 선택한 후에는 *positive*(양성) 반응을 지정해야 합니다.

검사를 통한 예측이 완벽했다면 전체 비정상 모집단의 아래쪽, 모든 정상 값의 위쪽에 위치하는 하나의 값을 얻게 됩니다. 민감도가 완벽하게 되면 격자 위의 점 (0,1)을 지닙니다. ROC 곡선이 이 이상적인 점에 가까울수록 판별 능력이 높습니다. 예측 능력이 없는 검사의 경우 격자의 대각선과 일치하는 곡선이 생성됩니다 (DeLong et al. 1988).

ROC 곡선은 False Positive 비율과 True Positive 비율의 관계를 그래픽으로 나타낸 것입니다. 관계를 평가하는 표준 방법은 보고서의 그림 아래에 표시된 곡선 아래 면적을 사용하는 것입니다.

그림에는 ROC 곡선에 45 도로 접하는 노란색 선이 그려집니다 . 이는 False Negative 및 False Positive 의 비용이 유사하다는 가정하에 적절한 경계 점을 나타냅니다 .

확률 계산식 저장

확률 계산식 저장 옵션은 새 데이터 테이블 열을 생성합니다 . 이 데이터 테이블 열에는 다음 항목이 저장됩니다 .

- x 요인의 선형 결합 (일반적으로 로짓이라고 함) 계산식
- 반응 수준 확률에 대한 예측 계산식
- 가장 가능성이 높은 반응을 구하는 예측 계산식

역추정 예측

참고 : " 십자기호 도구를 사용한 역추정 예측의 예 "(276 페이지) 및 " 역추정 예측 옵션을 사용한 역추정 예측의 예 "(277 페이지) 도 확인하십시오 .

역추정 예측은 예측을 반대로 하는 것입니다 . 즉 , 특정 y 값으로부터 x 값을 예측합니다 . 하지만 로지스틱 회귀에서는 y 값 대신 Y 수준 중 하나에 기인한 확률을 사용합니다 . 이 기능은 2 수준 명목형 반응에만 사용할 수 있습니다 .

모형 적합 플랫폼에도 신뢰 한계를 사용하여 역추정 예측을 수행하는 옵션이 있습니다 . 역추정 예측에 대한 자세한 내용은 **Fitting Linear Models** 의 "Standard Least Squares Report and Options" 장에서 확인하십시오 .

로지스틱 회귀의 추가 예

이 섹션에는 로지스틱 회귀를 사용하는 추가 예가 포함되어 있습니다 .

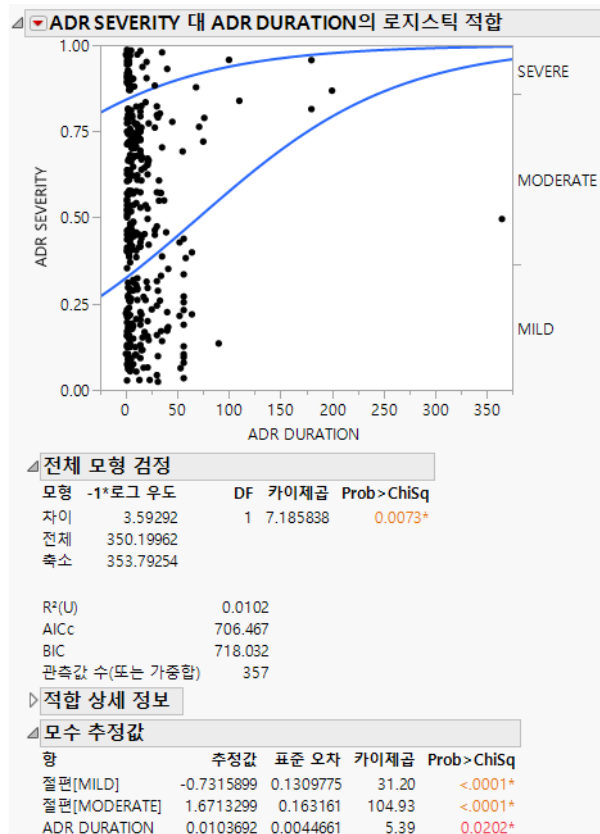
- " 순서형 로지스틱 회귀의 예 "
- " 로지스틱 그림의 추가 예 "
- "ROC 곡선의 예 "
- " 십자기호 도구를 사용한 역추정 예측의 예 "
- " 역추정 예측 옵션을 사용한 역추정 예측의 예 "

순서형 로지스틱 회귀의 예

이 예에서는 AdverseR.jmp 샘플 데이터 테이블을 사용하여 순서형 로지스틱 회귀를 보여 줍니다. 이상 사례의 심각도를 처리 기간 값의 함수로 모델링하려고 한다고 가정해 보겠습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 AdverseR.jmp 를 엽니다.
2. ADR SEVERITY 왼쪽의 아이콘을 마우스 오른쪽 버튼으로 클릭하고 모델링 유형을 순서형으로 변경합니다.
3. **분석 > X 로 Y 적합**을 선택합니다.
4. ADR SEVERITY 를 선택하고 **Y, 반응**을 클릭합니다.
5. ADR DURATION 을 선택하고 **X, 요인**을 클릭합니다.
6. **확인**을 클릭합니다.

그림 8.5 순서형 로지스틱 보고서의 예



이 보고서의 해석 방법은 명목형 보고서의 경우와 동일합니다. 자세한 내용은 "로지스틱 보고서"(265 페이지)에서 확인하십시오.

그림에서 데이터 표식은 x 좌표에 그려 집니다. 여러 개의 데이터 점이 동일한 y 위치에 나타나는 경우에는 점이 지터링됩니다. 즉, 각 점을 더 명확하게 볼 수 있도록 데이터 점 사이에 약간의 간격이 나타납니다.

점이 많은 경우에는 곡선들이 서로 멀어 집니다. 점이 거의 없거나 전혀 없으면 곡선이 서로 가까워 집니다. 데이터가 곡선을 이러한 방식으로 밀어내는 이유는 최대화되는 기준이 모형으로 적합한 확률의 곱이기 때문입니다. 적합한 확률이 작은 데 기인한 점, 즉 적합한 곡선에 의해 밀리는 점을 방지하려고 합니다. 계산 상세 정보에 대한 자세한 내용은 **Fitting Linear Models** 에서 확인 하십시오.

"전체 모형 검정" 보고서와 "모수 추정값" 보고서에 대한 자세한 내용은 "**로지스틱 보고서**"(265 페이지) 에서 확인하십시오. "모수 추정값" 보고서에서 절편 모수는 마지막 반응 수준을 제외한 각 반응 수준에 대해 추정된 값이지만 기울기 모수는 하나만 있습니다. 절편 모수는 반응 수준의 간격을 보여 줍니다. 이들은 항상 단순 증가합니다.

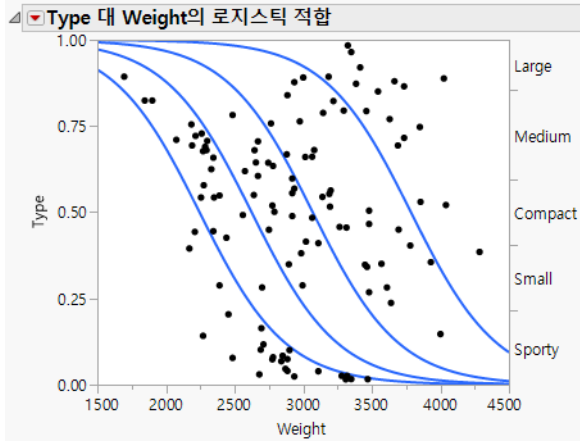
로지스틱 그림의 추가 예

이 예에서는 **Car Physical Data.jmp** 샘플 데이터 테이블을 사용하여 로지스틱 그림의 추가 예를 보여 줍니다. 116 대의 자동차에 대해 중량을 사용하여 자동차 크기 (Type) 를 예측하려고 한다고 가정해 보겠습니다. 자동차 크기는 작은 것부터 큰 것 순서로 "Sporty", "Small", "Compact", "Medium" 또는 "Large" 중 하나일 수 있습니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Car Physical Data.jmp** 를 엽니다.
2. 열 패널에서 Type 왼쪽의 아이콘을 마우스 오른쪽 버튼으로 클릭하고 **순서형**을 선택합니다.
3. Type 을 마우스 오른쪽 버튼으로 클릭하고 **열 정보**를 선택합니다.
4. "열 특성" 메뉴에서 **값 순서**를 선택합니다.
5. 데이터가 하향식 순서, 즉 "Sporty", "Small", "Compact", "Medium", "Large" 의 순서로 되어 있는지 확인합니다.
6. **확인**을 클릭합니다.
7. **분석 > X 로 Y 적합**을 선택합니다.
8. Type 을 선택하고 **Y, 반응**을 클릭합니다.
9. Weight 를 선택하고 **X, 요인**을 클릭합니다.
10. **확인**을 클릭합니다.

보고서 창이 나타납니다.

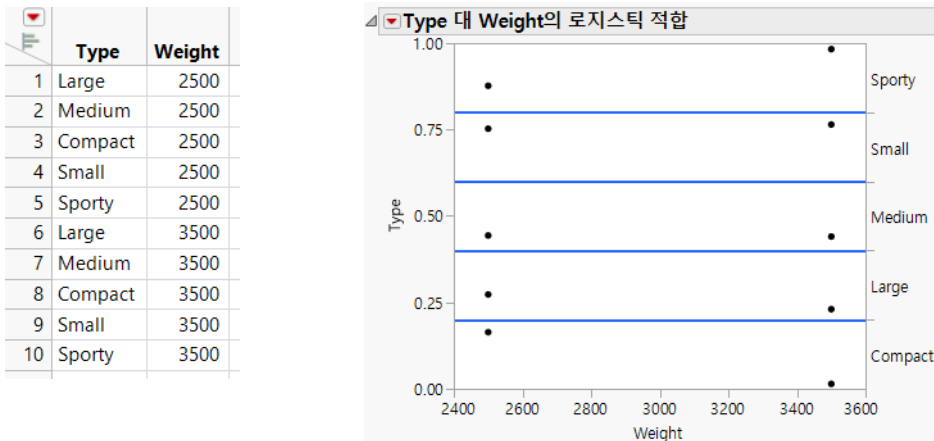
그림 8.6 Type 대 Weight 에 대한 로지스틱 그림의 예



다음 관측값에 유의하십시오 .

- 첫 번째 (맨 아래) 곡선은 주어진 중량의 자동차가 Sporty 일 확률을 나타냅니다 .
- 두 번째 곡선은 자동차가 Small 또는 Sporty 일 확률을 나타냅니다 . 첫 번째 곡선과 두 번째 곡선 사이의 거리만 보면 Small 일 확률에 부합합니다 .
- 예상할 수 있듯이 더 무거운 자동차는 Large 일 가능성이 높습니다 .
- 데이터 표식은 x 좌표에 그려집니다 . y 위치는 해당 행의 반응 범주에 해당하는 범위 내에서 무작위로 지터링됩니다 .

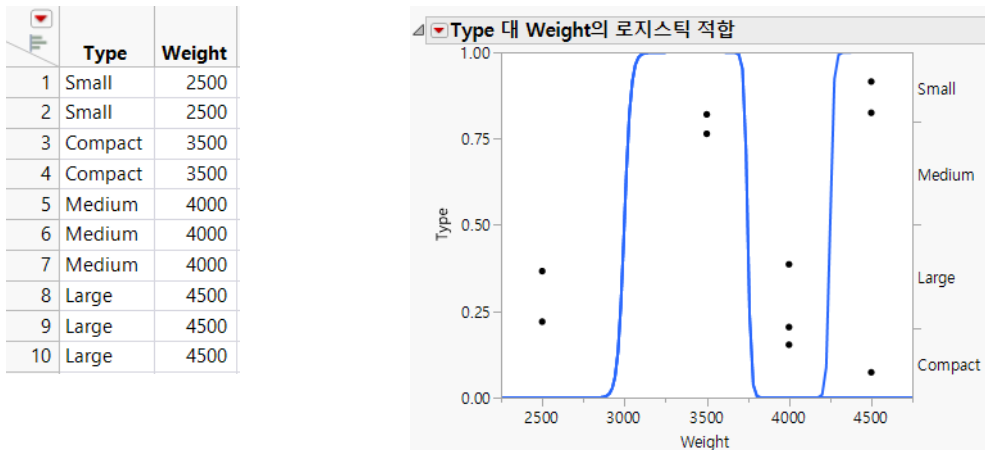
x 변수가 반응에 아무런 영향도 미치지 않는다면 적합선은 수평선이 되고 연속형 요인 범위에서 각 반응의 확률은 상수가 됩니다 . 그림 8.7 에서는 Weight 가 Type 을 예측하는 데 유용하지 않은 경우의 로지스틱 그림을 보여 줍니다 .

그림 8.7 y 대 x 관계가 없음을 보여 주는 샘플 데이터 테이블 및 로지스틱 그림의 예

참고 : 그림 8.7 및 그림 8.8의 그림을 다시 생성하려면 먼저 여기에 표시된 데이터 테이블을 생성한 다음 이 섹션의 처음에 나오는 7~10 단계를 수행해야 합니다.

요인 값으로 반응이 완전히 예측되는 경우에는 로지스틱 곡선이 거의 수직이 됩니다. 각 요인 수준에서 반응 예측은 거의 확실 (확률이 거의 1) 합니다. 그림 8.8에서는 **Weight**가 **Type**을 거의 완벽하게 예측하는 경우의 로지스틱 그림을 보여 줍니다.

그림 8.8 거의 완벽한 y 대 x 관계를 보여 주는 샘플 데이터 테이블 및 로지스틱 그림의 예



이 경우에는 모수 추정값이 매우 커지고 회귀 보고서에 불안정이라는 라벨이 표시됩니다. 이러한 경우에는 Firth 편향 조정 추정값을 사용하는 일반화 선형 모형 분석법을 사용할 수 있습니다. 자세한 내용은 Fitting Linear Models의 "Generalized Linear Models" 장에서 확인하십시오.

ROC 곡선의 예

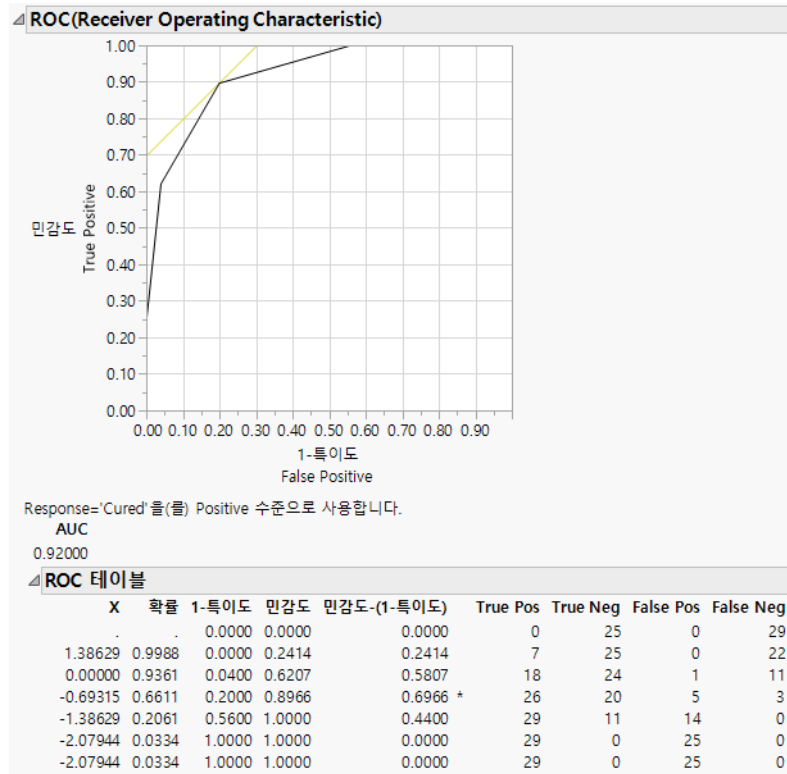
ROC 곡선을 생성해 보려면 다음 단계를 수행하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Penicillin.jmp**를 엽니다.
2. **분석 > X로 Y 적합**을 선택합니다.
3. **Response**를 선택하고 **Y, 반응**을 클릭합니다.
4. **ln(dose)**을 선택하고 **X, 요인**을 클릭합니다.
빈도에는 **Count**가 자동으로 채워집니다. 이미 **Count**에 빈도 역할이 할당되었기 때문입니다.
5. **확인**을 클릭합니다.
6. "Response 대 ln(dose)의 로지스틱 적합" 옆의 빨간색 삼각형을 클릭하고 **ROC 곡선**을 선택합니다.
7. **Positive** 수준으로 **Cured**를 선택하고 **확인**을 클릭합니다.

참고 : 이 예에서는 명목형 반응의 ROC 곡선을 보여 줍니다. 순서형 ROC 곡선에 대한 자세한 내용은 Predictive and Specialized Modeling 의 "Partition Models" 장에서 확인하십시오 .

response 대 ln(dose) 예의 결과는 다음과 같습니다 . ROC 곡선에는 반응 예측을 위해 위에서 설명한 확률이 표시됩니다 . "ROC 테이블 " 에서 " 민감도 -(1- 특이도)" 가 가장 높은 행에는 별표가 표시됩니다 .

그림 8.9 ROC 곡선 및 테이블의 예



이 ROC 곡선은 대각선의 매우 위쪽에 있으므로 모형의 예측 능력이 좋다는 결론을 내릴 수 있습니다 .

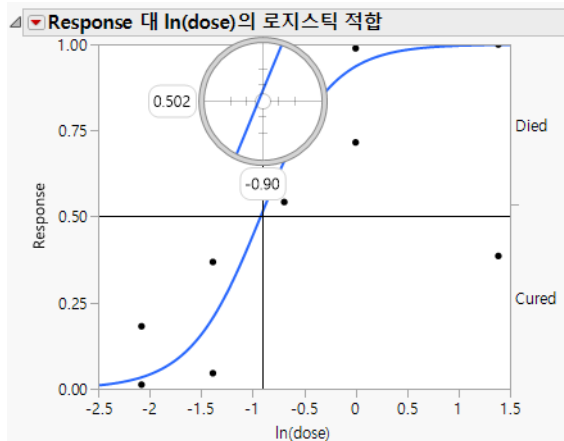
십자기호 도구를 사용한 역추정 예측의 예

페니실린이 투여된 토끼에 대한 연구에서 토끼의 치료 확률이 0.5 가 되는 페니실린 투여량을 알아보려고 합니다 . 이 경우 0.5 에 대한 역추정 예측을 *ED50* 이라고 하며 , 이는 생존률 50% 에 해당하는 유효 투여량을 나타냅니다 . 역추정 예측의 근사값을 시각적으로 구하려면 십자기호 도구를 사용합니다 .

치료 또는 치사 가능성이 동일한 투여량을 알아보려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Penicillin.jmp** 를 엽니다 .
 2. **분석 > X 로 Y 적합**을 선택합니다 .
 3. **Response** 를 선택하고 **Y, 반응**을 클릭합니다 .
 4. **ln(dose)** 을 선택하고 **X, 요인**을 클릭합니다 .
빈도에는 **Count**가 자동으로 채워집니다. 이미 **Count**에 빈도 역할이 할당되었기 때문입니다.
 5. **확인**을 클릭합니다 .
 6. 십자기호 도구를 클릭합니다 .
 7. 십자기호의 가로 선이 세로 (**Response**) 확률 축의 약 0.5 지점에 오도록 합니다 .
 8. 십자기호의 교차점을 예측선 쪽으로 이동하고 가로 축에 표시되는 **ln(dose)** 값을 읽습니다 .
- 이 예에서 **ln(dose)** 이 약 -0.9 인 토끼는 치료될 가능성과 죽을 가능성이 동일합니다 .

그림 8.10 로지스틱 그림에서 사용되는 십자기호 도구의 예



역추정 예측 옵션을 사용한 역추정 예측의 예

반응 수준이 정확히 두 개뿐인 경우 **역추정 예측** 옵션을 사용하여 정확한 역추정 예측을 요청할 수 있습니다 . 하위 반응 범주의 특정 확률에 해당하는 x 값과 해당 x 값의 신뢰 구간을 알 수 있습니다 .

역추정 예측 옵션을 사용하려면 다음 단계를 수행하십시오 .

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Penicillin.jmp** 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. **Response** 를 선택하고 **Y, 반응**을 클릭합니다 .
4. **ln(dose)** 을 선택하고 **X, 요인**을 클릭합니다 .

빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.

5. **확인**을 클릭합니다.
6. "Response 대 ln(dose)의 로지스틱 적합" 옆의 빨간색 삼각형을 클릭하고 **역추정 예측**을 선택합니다 (그림 8.11).
7. **신뢰 수준**에 0.95를 입력합니다.
8. 신뢰 구간으로 **양측**을 선택합니다.
9. 관심 있는 반응 확률을 요청합니다. 이 예에서는 0.5와 0.9를 입력합니다. 즉, 치료 확률이 0.5와 0.9인 ln(dose)의 값을 요청합니다.
10. **확인**을 클릭합니다.

역추정 예측 그림이 나타납니다.

그림 8.11 역추정 예측 창

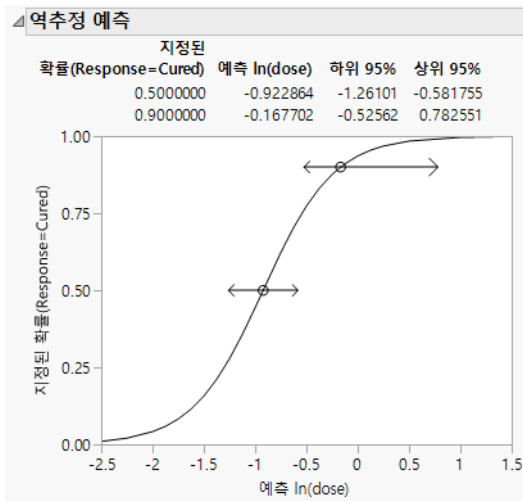
역추정 예측할 하나 이상의 확률 값을 지정하십시오.

ln(dose)	(예측할)	신뢰 수준	확률(Response=Cured)
		0.95	0.5
			0.9
			.
			.
			.
			.
			.
			.

양측

확인 취소 도움말

그림 8.12 역추정 예측 그림의 예



x 값의 추정값과 신뢰 구간이 보고서와 확률도에 표시됩니다. 예를 들어 치료 확률이 90%에 이르는 ln(dose)의 값은 -0.526 ~ 0.783으로 추정됩니다.

로지스틱 플랫폼에 대한 통계 상세 정보

이 섹션에는 "전체 모형 검정" 보고서에 대한 통계 상세 정보가 포함되어 있습니다.

카이제곱

카이제곱 통계량은 때때로 G^2 으로 표기되며, 계산식은 다음과 같습니다.

$$G^2 = 2(\sum -\ln p(\text{background}) - \sum -\ln p(\text{모형}))$$

여기서 합계는 모든 셀이 아니라 모든 관측값에 대한 값입니다.

$R^2(U)$

기준 로그 우도에 대한 이 검정 통계량의 비율을 1 에서 빼서 R^2 를 계산합니다. $R^2(U)$ 를 더 간단하게 계산하는 방법은 다음과 같습니다.

$$\frac{\text{차이 모형에 대한 음의 로그 우도}}{\text{축소 모형에 대한 음의 로그 우도}}$$

이 계산에는 "전체 모형 검정" 보고서의 통계량이 사용됩니다.

참고 : $R^2(U)$ 를 McFadden 유사 R^2 라고도 합니다.

테이블 생성

대화식으로 요약 테이블 생성

테이블 생성 플랫폼을 사용하여 기술 통계량의 테이블을 대화식으로 생성할 수 있습니다. 테이블 생성 플랫폼은 요약 데이터를 테이블 형식으로 표시하는 쉽고 유연한 방법입니다. 테이블은 그룹화 열, 분석 열 및 통계량 키워드를 사용하여 작성됩니다.

그림 9.1 테이블 생성 예

	sex											
	F						M					
	age						age					
	12	13	14	15	16	17	12	13	14	15	16	17
	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight
최소값	64	67	81	92	112	116	79	79	92	104	128	134
평균	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
최대값	145	112	142	112	115	116	128	105	119	128	128	172

		sex							
		Female				Male			
		marital status				marital status			
		Married		Single		Married		Single	
		평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
country	size	age	age	age	age	age	age	age	age
American	Large	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	Medium	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	Small	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
European	Large	34.0	7.071	28.0	.	.	.	26.0	.
	Medium	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	Small	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
Japanese	Large	25.0	.	.	.	32.0	.	.	.
	Medium	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	Small	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
country									
American		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
European		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
Japanese		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

		평균		
Type	Size Co	Profits (\$M)	Sales (\$M)	profit/emp
Computer	biq	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
Pharmaceutical	모두	240.87	5652.02	6159.015
	biq	894.42	7474.04	17140.70
	medium	698.98	4261.06	24035.11
모두	small	156.95	1083.75	38337.19
	모두	690.08	5070.25	23546.12
	모두	409.32	5433.86	12679.18
Type				
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
모두		409.32	5433.86	12679.18

목차

테이블 생성 플랫폼의 예.....	283
테이블 생성 플랫폼 시작.....	287
대화상자 사용.....	288
통계량 추가.....	289
테이블 생성 출력.....	292
분석 열.....	293
그룹화 열.....	294
열 및 행 테이블.....	294
테이블 편집.....	295
테이블 생성 플랫폼 옵션.....	296
테스트 빌드 패널 표시.....	297
마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴.....	297
테이블 생성 플랫폼의 추가 예.....	298
예: 여러 테이블 생성 및 내용 재배열.....	298
예: 여러 열을 단일 테이블에 결합.....	301
예: 페이지 열 사용.....	303

테이블 생성 플랫폼의 예

남학생과 여학생의 키 측정값이 포함된 데이터가 있습니다. 남학생 및 여학생의 평균 키와 두 성별 모두의 키 집계를 보여 주는 테이블을 생성하려고 합니다.

그림 9.2 평균 키를 보여 주는 테이블

	height
sex	평균
F	60.9
M	63.9
모두	62.6

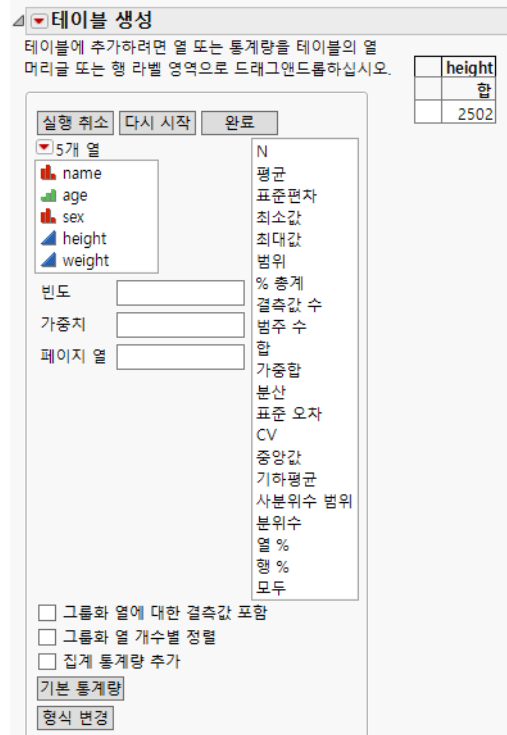
1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Big Class.jmp 를 엽니다 .

2. **분석 > 테이블 생성**을 선택합니다 .

"height" 변수를 검토할 것이므로 테이블의 맨 위에 이 변수를 표시하려고 합니다 .

3. height 를 클릭하고 " 열에 대한 놓기 영역 " 으로 드래그합니다 .

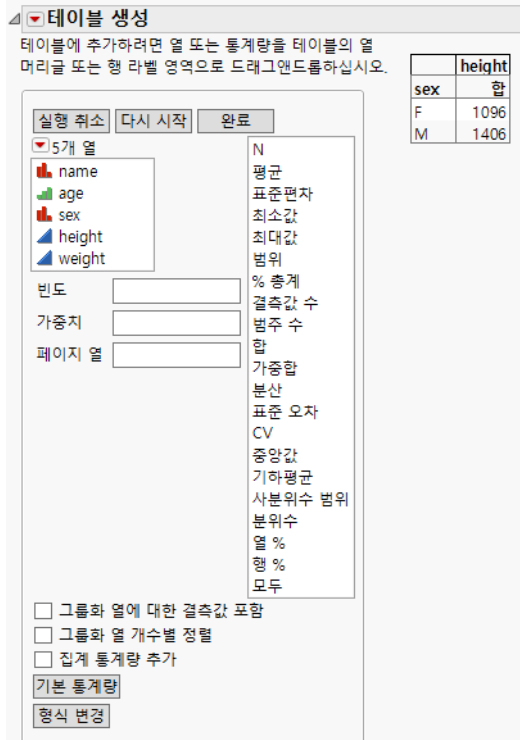
그림 9.3 height 변수 추가



성별별로 통계량을 살펴볼 것이므로 sex 변수를 나란히 표시하려고 합니다.

4. sex 를 클릭하고 숫자 "2502" 옆의 빈 셀로 드래그합니다.

그림 9.4 sex 변수 추가



합 대신 평균을 표시하려고 합니다.

5. 평균을 클릭하고 합 위로 드래그합니다.

그림 9.5 평균 통계량 추가

테이블 생성

테이블에 추가하려면 열 또는 통계량을 테이블의 열 머리글 또는 행 라벨 영역으로 드래그앤드롭하십시오.

sex	height
F	60.9
M	63.9

실행 취소

다시 시작

완료

5개 열

name

age

sex

height

weight

빈도

가중치

페이지 열

☐ 그룹화 열에 대한 결측값 포함

☐ 그룹화 열 개수별 정렬

☐ 집계 통계량 추가

기본 통계량

형식 변경

N

평균

표준편차

최소값

최대값

범위

% 총계

결측값 수

범주 수

합

가중합

분산

표준 오차

CV

중앙값

기하평균

사분위수 범위

분위수

열 %

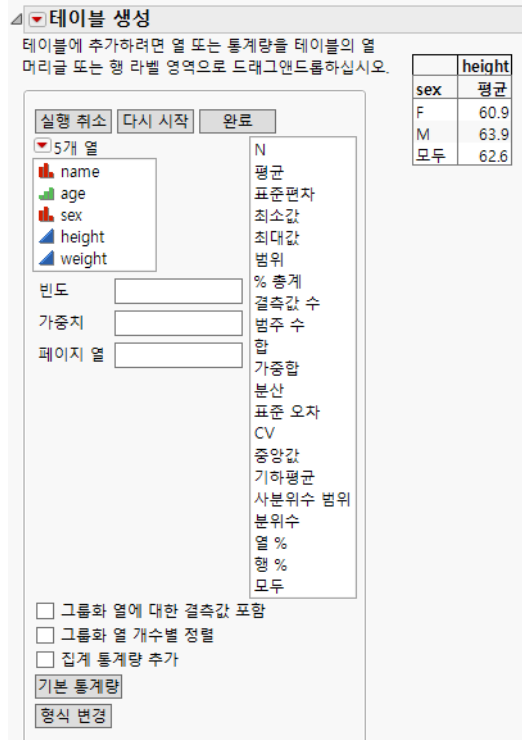
행 %

모두

또한 남학생과 여학생의 결합 평균을 표시하려고 합니다.

- 모두를 클릭하고 **sex** 위로 드래그합니다 . 또는 단순히 **집계 통계량 추가** 체크박스를 선택해도 됩니다 .

그림 9.6 전체 (" 모두 ") 통계량 추가



7. (선택 사항) **완료**를 클릭합니다.

완료된 테이블에 여학생 및 남학생의 평균 키와 둘 모두의 결합 평균 키가 표시됩니다.

테이블 생성 플랫폼 시작

테이블 생성 플랫폼을 시작하려면 **분석 > 테이블 생성**을 선택합니다.

그림 9.7 대화식 테이블 생성

참고 : 빨간색 삼각형 옵션에 대한 자세한 내용은 "[테이블 생성 플랫폼 옵션](#)"(296 페이지)에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 "시작하기"장에서 확인하십시오.

"테이블 생성" 창에는 다음과 같은 옵션이 포함되어 있습니다.

대화식 테이블 / 대화상자 두 모드 사이를 전환합니다. 항목을 드래그하여 놓는 방법으로 사용자 테이블을 생성하려면 대화식 테이블 모드를 사용합니다. 고정 형식을 사용하여 간단한 테이블을 생성하려면 대화상자 모드를 사용합니다. 자세한 내용은 "[대화상자 사용](#)"(288 페이지)에서 확인하십시오.

통계량 옵션 표준 통계량을 나열합니다. 목록의 통계량을 테이블로 드래그하여 테이블에 포함할 수 있습니다. 자세한 내용은 "[통계량 추가](#)"(289 페이지)에서 확인하십시오.

열에 대한 놓기 영역 열 또는 통계량을 드래그하여 여기에 놓으면 열이 생성됩니다.

참고: 데이터 테이블에 "통계량" 옵션의 통계량과 이름이 같은 열이 포함되어 있으면 열 목록의 열 이름을 드래그하여 놓아야 합니다. 그렇지 않으면 JMP에서는 테이블의 열이 동일한 이름의 통계량으로 대체됩니다.

행에 대한 놓기 영역 열 또는 통계량을 드래그하여 여기에 놓으면 행이 생성됩니다.

팁: 열 목록에서 열을 하나 이상 선택하고 통계량을 하나 이상 선택한 다음 놓기 영역을 Alt 키를 누른 채 클릭 (macOS의 경우 Option 키를 누른 채 클릭) 하여 테이블의 행 또는 열을 생성할 수도 있습니다.

결과 셀 드래그하여 놓은 열 또는 통계량을 기준으로 결과 셀을 표시합니다.

빈도 각 행에 빈도를 할당하는 데 사용할 값이 있는 데이터 테이블 열을 식별합니다. 이 옵션은 요약된 데이터에서 각 행에 빈도가 할당된 경우에 유용합니다.

가중치 데이터에 가중치 (중요도 또는 영향력)를 할당하는 데 사용할 변수가 있는 데이터 테이블 열을 식별합니다.

페이지 열 명목형 또는 순서형 열의 각 범주에 대한 별도의 테이블을 생성합니다. 자세한 내용은 "[예: 페이지 열 사용](#)"(303 페이지)에서 확인하십시오.

그룹화 열에 대한 결측값 포함 그룹화 열의 결측값에 대한 별도의 그룹을 생성합니다. 이 옵션을 선택 해제하면 결측값이 테이블에 포함되지 않습니다. 열 특성으로 정의한 모든 결측값 코드가 고려됩니다.

그룹화 열 개수별 정렬 테이블의 정렬 순서를 그룹화 열의 값에 대한 내림차순으로 변경합니다.

집계 통계량 추가 모든 행 및 열에 대한 집계 통계량을 추가합니다.

기본 통계량 분석 열 또는 분석 열이 아닌 열 (예: 그룹화 열)을 드래그하여 놓을 때 표시되는 기본 통계량을 변경할 수 있습니다.

형식 변경 특정 통계량을 표시하기 위한 숫자 형식을 변경할 수 있습니다. 자세한 내용은 "[숫자 형식 변경](#)"(291 페이지)에서 확인하십시오.

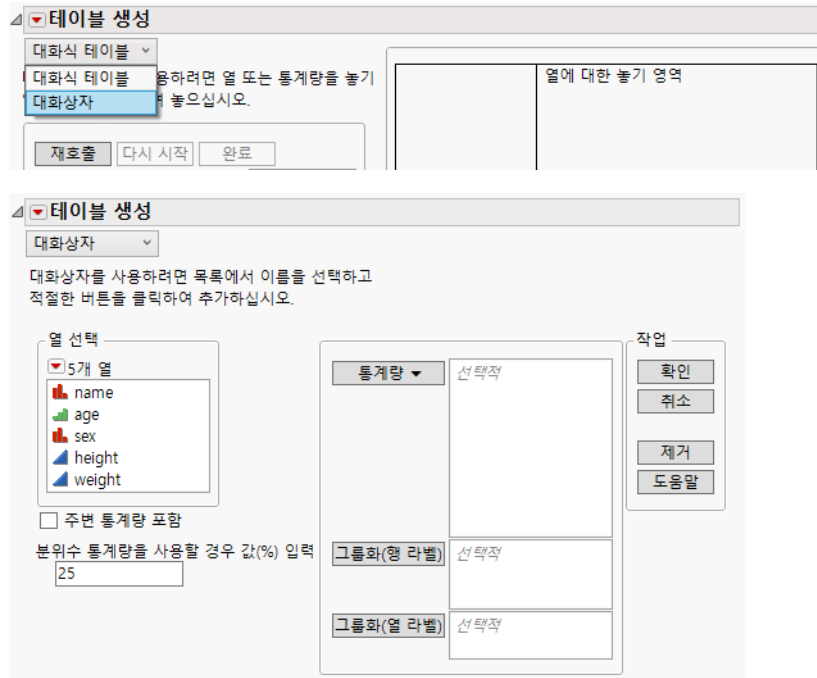
그림 척도 변경 (빨간색 삼각형 메뉴에서 **차트 표시**를 선택한 경우에만 표시됨) 균등 사용자 척도를 지정할 수 있습니다.

균등 그림 척도 (빨간색 삼각형 메뉴에서 **차트 표시**를 선택한 경우에만 표시됨) 각 막대 열에 대해 표시되는 각 열의 데이터와는 별도로 결정된 척도를 사용하려면 이 상자를 선택 취소합니다.

대화상자 사용

드래그하여 놓는 방법 대신 대화식 방법으로 테이블을 작성하려는 경우 "대화상자" 인터페이스를 사용하여 간단한 테이블을 생성할 수 있습니다. **분석 > 테이블 생성**을 선택한 후 그림 9.8에 표시된 것과 같이 메뉴에서 **대화상자**를 선택합니다. 빨간색 삼각형 메뉴에서 **제어판 표시**를 선택한 다음 새 항목을 테이블로 드래그하여 놓는 방법으로 테이블을 변경할 수 있습니다.

그림 9.8 대화상자 사용



대화상자에는 다음 옵션이 포함되어 있습니다.

주변 통계량 포함 그룹화 열의 범주에 대한 요약 정보를 집계합니다.

분위수 통계량을 사용할 경우 값 (%) 입력 특정 비율 지점을 나타내는 값을 입력합니다. 해당 비율 만큼의 인수가 해당 값보다 작거나 같습니다. 예를 들어 75%의 데이터는 75 번째 분위수보다 작습니다. 이는 모든 그룹화 열에 적용됩니다.

통계량 열을 선택한 후 해당 열에 적용할 표준 통계량을 선택합니다. 자세한 내용은 "[통계량 추가](#)"(289 페이지)에서 확인하십시오.

그룹화 (행 라벨) 행 라벨로 사용할 열을 선택합니다.

그룹화 (열 라벨) 열 라벨로 사용할 열을 선택합니다.

통계량 추가

팁: 열과 통계량을 모두 동시에 선택하고 테이블로 드래그할 수 있습니다.

테이블 생성 기능은 표준 통계량 목록을 지원합니다. 이 목록은 제어판에 표시됩니다. 열을 추가할 때와 마찬가지로 이 목록의 키워드를 테이블로 드래그할 수 있습니다. 다음 사항에 유의하십시오.

- 각 셀과 연관된 통계량은 그룹화 열로 정의된 해당 범주의 모든 관측값에서 분석 열의 값에 대해 계산됩니다.
- 요청된 통계량은 모두 행 테이블이나 열 테이블에서 동일한 차원으로 있어야 합니다.
- 연속형 열을 데이터 영역으로 드래그하면 해당 열은 분석 열로 처리됩니다.

참고: 분석 열은 통계량을 계산하려는 연속형 숫자 열입니다. 자세한 내용은 "**분석 열**"(293 페이지)에서 확인하십시오.

테이블 생성에는 다음 키워드가 사용됩니다.

N 열에 있는 비결측값의 수를 제공합니다. 이는 분석 열이 없을 때의 기본 통계량입니다.

평균 열 값의 산술평균을 제공합니다. 비결측값 (정의된 경우 **가중치** 변수를 곱한 값)의 합을 **가중합**으로 나눈 결과입니다.

표준편차 비결측값에 대해 계산된 표본 표준편차를 제공합니다. 이 값은 표본 분산의 제곱근입니다.

최소값 열에 있는 가장 작은 비결측값을 제공합니다.

최대값 열에 있는 가장 큰 비결측값을 제공합니다.

범위 **최대값**과 **최소값**의 차이를 제공합니다.

% 총계 전체 모집단의 총계에 대한 백분율을 계산합니다. 계산에 사용되는 분모는 포함된 모든 관측값의 합계이고, 분자는 해당 범주의 합계입니다. 분석 열이 없는 경우 "% 총계"는 총 개수에 대한 백분율입니다. 분석 열이 있는 경우 "% 총계"는 분석 열의 총 합에 대한 백분율입니다. 따라서 분모는 포함된 모든 관측값에 대한 분석 열의 합이고, 분자는 해당 범주에 대한 분석 열의 합입니다. 키워드를 테이블로 드래그하여 다른 백분율을 요청할 수 있습니다.

- 테이블의 그룹화 변수 중 하나 이상을 "% 총계" 머리글에 놓으면 분모 정의가 변경됩니다. 이 경우 테이블 생성 기능은 해당 그룹화 열의 합을 분모에 사용합니다.
- 열 합계의 백분율을 구하려면 행 테이블의 모든 그룹화 열을 **% 총계** 머리글 ("열 %"와 동일)로 드래그하여 놓습니다. 행 합계의 백분율을 구하려면 열 테이블의 모든 그룹화 열을 **% 총계** 머리글 ("행 %"와 동일)로 드래그하여 놓습니다.

결측값 수 결측값의 개수를 제공합니다.

범주 수 분석 열에 있는 개별 범주의 개수를 제공합니다.

합 해당 열에 있는 모든 값의 합을 제공합니다. 이 통계량은 테이블에 대한 다른 통계량이 없는 경우 분석 열의 기본 통계량입니다.

가중합 열에 있는 모든 가중치 값의 합을 제공합니다. 가중치 역할에 할당된 열이 없는 경우, **가중합**은 비결측값의 총 수입니다.

분산 비결측값에 대해 계산된 표본 분산을 제공합니다. 평균으로부터의 편차에 대한 제곱합을 비결측값의 수에서 1 을 뺀 값으로 나눈 것입니다.

표준 오차 평균의 표준 오차를 제공합니다. 표준편차를 **N** 의 제곱근으로 나눈 것입니다. 열에 가중치 역할이 할당된 경우에는 분모가 가중치 합의 제곱근입니다.

CV (변동 계수) 표준편차를 평균으로 나누고 100 을 곱한 산포 측도입니다.

중앙값 데이터의 위쪽 절반과 아래쪽 절반을 나누는 기준이 되는 50 번째 백분위수, 또는 50 번째 분위수 (중앙값) 를 제공합니다.

기하평균 데이터 곱의 n 제곱근입니다. 예를 들어 기하평균은 이자율을 계산하는 데 종종 사용됩니다. 이 통계량은 데이터에 많은 값이 비대칭적 분포로 포함된 경우에도 유용합니다.

참고 : 음수 값이 있으면 결측값이 발생하고, 음수가 아니고 0 인 값이 있으면 결과가 0 이 됩니다.

사분위수 범위 3 사분위수와 1 사분위수 사이의 차이를 제공합니다.

분위수 특정 비율 지점을 나타내는 값을 제공합니다. 해당 비율만큼의 인수가 해당 값보다 작거나 같습니다. 예를 들어 75% 의 데이터는 75 번째 분위수보다 작습니다. **분위수** 키워드를 클릭하여 테이블로 드래그한 다음 표시되는 상자에 분위수를 입력하는 방법으로 다른 분위수를 요청할 수 있습니다.

열 % 분석 열이 없는 경우 열 합계에 대한 각 셀 개수의 백분율을 제공합니다. 분석 열이 있는 경우 "열 %" 는 분석 열의 열 총 합에 대한 백분율입니다. 맨 위에 통계량이 있는 테이블의 경우 여러 행 테이블과 함께 테이블에 "열 %" 를 추가할 수 있습니다 (세로로 누적).

행 % 분석 열이 없는 경우 행 합계에 대한 각 셀 개수의 백분율을 제공합니다. 분석 열이 있는 경우 "행 %" 는 분석 열의 행 총 합에 대한 백분율입니다. 통계량이 나란히 표시되는 테이블의 경우 여러 열 테이블과 함께 테이블에 "행 %" 를 추가할 수 있습니다 (나란히 표시되는 테이블).

모두 그룹화 열의 범주에 대한 요약 정보를 집계합니다.

숫자 형식 변경

각 셀의 계산식은 분석 열 및 통계량에 따라 달라집니다. 개수의 경우 기본 형식에는 소수점 이하 자릿수가 없습니다. 일부 통계량으로 정의된 각 셀의 경우, JMP에서는 요청된 분석 열과 통계량의 형식을 사용하여 적절한 형식을 결정하려고 합니다. 기본 형식을 재정의하려면 다음을 수행하십시오.

1. "테이블 생성" 창의 맨 아래에 있는 **형식 변경** 버튼을 클릭합니다.
2. 표시되는 패널에서 필드 너비, 쉼표, 테이블에 표시할 소수 자릿수를 차례로 입력합니다 (그림 9.9).
3. 셀 값을 백분율 형식으로 표시하려면 소수 자릿수 다음에 쉼표를 입력하고 **백분율**이라는 단어를 추가합니다.

4. (선택 사항) 사용할 최적의 형식이 JMP 에서 자동으로 결정되도록 하려면 텍스트 상자에 **최적**이라는 단어를 입력합니다.
- JMP 가 각 셀 값의 정밀도를 고려하여 가장 적절한 값 표시 방법을 선택합니다.
5. 변경 사항을 구현하고 " 형식 " 섹션을 닫으려면 **확인**을 클릭하고 , 구현된 변경 사항을 확인하고 " 형식 " 섹션을 닫지는 않으려면 **형식 설정**을 클릭합니다 .

그림 9.9 숫자 형식 변경

형식

☐ 동일한 십진수 형식 사용

특정 통계량을 표시하기 위해 숫자 형식을 변경할 수 있습니다. 각 형식은 필드 너비와 소수 자릿수를 나타내는 두 개의 정수로 구성됩니다. '최적 형식'을 사용하려면 두 번째 정수 대신 '최적'을 사용하십시오. '백분율 형식'을 지정하려면 두 번째 정수 뒤에 '백분율'이라는 단어를 추가하십시오.

N

9, 최적

형식 설정

취소

확인

테이블 생성 출력

테이블 생성 기능의 출력은 나란히 연결된 하나 이상의 열 테이블과 위에서 아래로 연결된 하나 이상의 행 테이블로 구성됩니다 . 출력에 열 테이블 또는 행 테이블이 하나만 포함될 수도 있습니다 .

그림 9.10 테이블 생성 출력

	sex											
	F						M					
	age						age					
	12	13	14	15	16	17	12	13	14	15	16	17
	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight	weight
최소값	64	67	81	92	112	116	79	79	92	104	128	134
평균	100.2	95.3	96.6	102.0	113.5	116.0	97.0	94.3	103.9	110.8	128.0	153.0
최대값	145	112	142	112	115	116	128	105	119	128	128	172

		sex							
		Female				Male			
		marital status				marital status			
		Married		Single		Married		Single	
		평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
country	size	age	age	age	age	age	age	age	age
American	Large	33.6	8.107	41.0	.	34.7	3.931	32.0	6.265
	Medium	31.4	5.827	29.0	9.258	31.3	5.413	32.1	11.05
	Small	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
	모두	31.0	5.657	29.0	9.539	31.8	4.813	26.5	6.455
European	Large	34.0	7.071	28.0	.	.	.	26.0	.
	Medium	31.0	5.06	28.7	5.508	32.3	5.62	31.0	10.13
	Small	29.8	6.611	28.0	1.414	33.8	4.381	25.7	2.517
	모두	31.0	5.657	28.7	5.508	32.3	5.62	31.0	10.13
Japanese	Large	25.0	.	.	.	32.0	.	.	.
	Medium	30.5	4.993	28.0	3.071	32.3	3.878	27.4	5.016
	Small	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
	모두	29.6	4.251	31.1	9.562	29.8	5.357	28.7	4.739
country	size	age	age	age	age	age	age	age	age
American		31.9	6.452	30.0	9.115	32.6	4.919	31.0	8.179
European		31.0	5.612	28.3	3.559	33.3	4.608	28.4	7.328
Japanese		29.8	4.54	30.1	8.113	30.9	4.822	28.3	4.781

		평균		
Type	Size Co	Profits (\$M)	Sales (\$M)	profit/emp
Computer	big	1089.9	20597.48	4530.478
	medium	-85.75	3018.85	-3462.51
	small	44.94	1758.06	7998.815
	모두	240.87	5652.02	6159.015
Pharmaceutical	big	894.42	7474.04	17140.70
	medium	698.98	4261.06	24035.11
	small	156.95	1083.75	38337.19
	모두	690.08	5070.25	23546.12
모두	모두	409.32	5433.86	12679.18
Type	Size Co	Profits (\$M)	Sales (\$M)	profit/emp
Computer		240.87	5652.02	6159.015
Pharmaceutical		690.08	5070.25	23546.12
모두		409.32	5433.86	12679.18

대화식으로 테이블을 생성하려면 다음 과정을 반복해야 합니다.

- 해당 목록에서 항목 (열 또는 통계량) 을 클릭하고 행 또는 열에 대한 놓기 영역으로 드래그합니다. 자세한 내용은 " [테이블 편집](#) "(295 페이지) 및 " [열 및 행 테이블](#) "(294 페이지) 에서 확인하십시오.
- 드래그하여 놓기 과정을 반복하여 테이블에 항목을 추가합니다. 테이블이 업데이트되어 마지막 추가 항목이 반영됩니다. 열 머리글 또는 행 라벨이 이미 있는 경우에는 기존 항목을 기준으로 추가 항목의 위치를 결정할 수 있습니다.

클릭 및 드래그 작업에 대한 다음 사항에 유의하십시오.

- JMP에서는 모델링 유형을 사용하여 열의 역할이 결정됩니다. 연속형 열은 분석 열로 할당됩니다. 자세한 내용은 " [분석 열](#) "(293 페이지) 에서 확인하십시오. 순서형 또는 명목형 열은 그룹화 열로 간주됩니다. 자세한 내용은 " [그룹화 열](#) "(294 페이지) 에서 확인하십시오.
- 초기 테이블에 여러 개의 열을 드래그하여 놓으면 다음 작업이 수행됩니다.
 - 해당 열이 일련의 공통된 값을 공유하는 경우에는 열이 단일 테이블에 결합됩니다. 해당 열 이름과 해당 열에서 수집된 범주의 교차표가 생성됩니다. 각 셀은 하나의 열과 하나의 범주로 정의됩니다.
 - 해당 열이 공통된 값을 공유하지 않는 경우에는 각 열이 개별 테이블에 포함됩니다.
 - 언제든지 열을 마우스 오른쪽 버튼으로 클릭하고 **테이블 결합** 또는 **별도의 테이블** 을 선택하여 기본 작업을 변경할 수 있습니다. 자세한 내용은 " [마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴](#) "(297 페이지) 에서 확인하십시오.
- 열을 내포하려면 첫 번째 열을 포함하는 테이블을 생성한 다음 추가 열을 첫 번째 열로 드래그합니다.
- 올바르게 생성된 테이블에서는 모든 그룹화 열, 모든 분석 열 및 모든 통계량이 각각 함께 표시됩니다. 따라서 JMP에서는 분석 열 목록 내에서 통계량 키워드가 변형되지 않습니다. 또한 그룹화 열 목록 내에 분석 열이 삽입되지도 않습니다.
- "테이블 생성" 제어판을 사용하는 대신 데이터 테이블에서 "테이블" 패널의 열을 "테이블 생성" 테이블로 드래그할 수 있습니다.

참고: 열려 있는 데이터 테이블에서 데이터를 추가하거나 행을 삭제하거나 데이터를 재코딩하면 "테이블 생성" 테이블이 업데이트됩니다.

분석 열

분석 열은 통계량을 계산하려는 숫자 열입니다. 이 열은 연속형 열입니다. 테이블 생성 기능은 그룹화 열로 구성된 각 범주에 대해 분석 열의 통계량을 계산합니다.

모든 분석 열은 행 테이블이나 열 테이블에서 동일한 차원으로 있어야 합니다.

그룹화 열

그룹화 열은 데이터를 여러 정보 범주로 분류하는 데 사용할 열입니다. 이 열에는 문자, 정수 또는 십진수가 포함될 수 있지만 고유 값의 개수는 제한됩니다. 그룹화 열은 명목형이거나 순서형입니다.

다음 사항에 유의하십시오.

- 그룹화 열이 내포된 경우 테이블 생성 기능은 열 값의 계층적 내포 구조로부터 개별 범주를 생성합니다. 예를 들어 F 값 및 M 값을 갖는 Sex 열과 Married 및 Single 값을 갖는 Marital Status 열이 그룹화 열인 경우 테이블 생성 기능은 "F 및 Married", "F 및 Single", "M 및 Married", "M 및 Single" 이라는 네 개의 개별 범주를 생성합니다.
- 열 테이블과 행 테이블 모두에 그룹화 열을 지정할 수 있습니다. 두 경우 모두 각 테이블 셀을 정의하는 범주가 생성됩니다.
- 테이블 생성 기능은 하나 이상의 그룹화 열에 대해 결측값이 있는 관측값을 기본적으로 포함하지 않습니다. **그룹화 열에 대한 결측값 포함** 옵션을 선택하면 이러한 관측값을 포함할 수 있습니다.
- 결측값으로 간주해야 할 코드 또는 값을 지정하려면 "결측값 코드" 열 특성을 사용합니다. **그룹화 열에 대한 결측값 포함** 옵션을 선택하면 이러한 관측값을 포함할 수 있습니다. 결측값 코드에 대한 자세한 내용은 JMP 사용의 "열 정보 창" 장에서 확인하십시오.

열 및 행 테이블

테이블 생성 기능에서 테이블은 열 머리글과 행 라벨로 정의됩니다. 이러한 하위 테이블을 행 테이블 및 열 테이블이라고 합니다 (그림 9.11).

행 및 열 테이블의 예

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다.
2. **분석 > 테이블 생성**을 선택합니다.
3. size 를 " 행에 대한 놓기 영역 " 으로 드래그합니다.
4. country 를 size 머리글의 왼쪽으로 드래그합니다.
5. 평균을 N 머리글 위로 드래그합니다.
6. 표준편차를 평균 머리글 아래로 드래그합니다.
7. age 를 평균 머리글 위로 드래그합니다.
8. type 을 테이블의 맨 오른쪽으로 드래그합니다.
9. sex 를 테이블 아래로 드래그합니다.

그림 9.11 행 및 열 테이블

두 개의 열 테이블

두 개의 행 테이블

country	size	age		type		
		평균	표준편차	Family	Sporty	Work
American	Large	33.8	6.167	28	1	7
	Medium	31.1	6.976	35	14	4
	Small	30.4	5.846	11	8	7
European	Large	30.5	5.802	1	0	3
	Medium	30.9	6.194	9	8	0
	Small	30.8	5.388	5	13	1
Japanese	Large	28.5	4.95	1	0	1
	Medium	30.2	4.668	32	15	7
	Small	29.7	5.928	33	41	18
sex						
Female		30.6	6.386	76	41	21
Male		30.8	5.643	79	59	27

열 테이블이 여러 개인 경우 옆쪽의 라벨이 열 테이블 간에 공유됩니다. 이 경우에는 "country" 및 "sex" 가 테이블 간에 공유됩니다. 마찬가지로, 행 테이블이 여러 개이면 위쪽의 머리글이 행 테이블 간에 공유됩니다. 이 경우에는 "age" 및 "type" 이 테이블 간에 공유됩니다.

테이블 편집

몇 가지 방법으로 테이블에 추가하는 항목을 편집할 수 있습니다.

항목 삭제

테이블에 항목을 추가한 후 다음 중 한 가지 방법으로 항목을 제거할 수 있습니다.

- 항목을 테이블 외부로 드래그합니다.
- 마지막 항목을 제거하려면 **실행 취소**를 클릭합니다.
- 항목을 마우스 오른쪽 버튼으로 클릭하고 **삭제**를 선택합니다.

열 라벨 제거

그룹화 열에서는 해당 열과 연관된 범주 위에 열 이름이 표시됩니다. 일부 열의 경우에는 열 이름이 중복될 수 있습니다. 열 이름을 마우스 오른쪽 버튼으로 클릭하고 **열 라벨 제거**를 선택하면 열 테이블에서 열 이름을 제거할 수 있습니다. 열 라벨을 다시 삽입하려면 연관된 범주 중 하나를 마우스 오른쪽 버튼으로 클릭하고 **열 라벨 복원**을 선택합니다.

통계 키워드 및 라벨 편집

통계 키워드 또는 통계 라벨을 편집할 수 있습니다. 예를 들어 "평균" 대신 "평균값" 이라는 단어를 사용할 수 있습니다. 편집하려는 단어를 마우스 오른쪽 버튼으로 클릭하고 **항목 라벨 변경**

을 선택합니다. 그런 다음 표시되는 상자에 새 라벨을 입력합니다. 또는 편집 상자에 라벨을 직접 입력해도 됩니다.

한 통계 키워드를 다른 통계 키워드로 변경하면 JMP에서는 단지 라벨이 아니라 통계량을 실제로 변경하려는 것으로 간주됩니다. 즉, 첫 번째 기존 통계량을 테이블에서 삭제한 후 두 번째 통계량을 추가하는 것으로 간주됩니다.

테이블 생성 플랫폼 옵션

"테이블 생성" 옆의 빨간색 삼각형 메뉴에서 사용할 수 있는 옵션은 다음과 같습니다.

제어판 표시 이후 상호 작용을 위해 제어판을 표시합니다.

테이블 표시 요약된 데이터를 테이블 형식으로 표시합니다.

차트 표시 요약된 데이터를 요약 통계량 테이블을 반영하는 막대 차트로 표시합니다. 간단한 막대 차트를 사용하여 요약 통계량의 상대적 크기를 시각적으로 비교할 수 있습니다. 기본적으로 모든 막대 열이 동일한 척도를 공유합니다. **균등 그림 척도** 체크박스를 선택 해제하면 각 막대 열이 각각의 표시된 열에 있는 데이터에서 개별적으로 결정된 척도를 사용하도록 할 수 있습니다. **그림 척도 변경** 버튼을 사용하면 균등 사용자 척도를 지정할 수 있습니다. 차트는 0부터 시작하거나 0을 중심으로 표시될 수 있습니다. 데이터가 모두 음수가 아니거나 모두 양수가 아닐 경우 차트 기준선은 0입니다. 그렇지 않으면 차트가 0을 중심으로 표시됩니다.

음영 표시 행이 여러 개인 경우 테이블에 회색 음영 상자를 표시합니다.

툴팁 표시 테이블 영역 위에 커서를 놓을 때 나타나는 토폴을 표시합니다.

테스트 빌드 패널 표시 원래 테이블에서 표집한 랜덤 표본을 사용하여 테스트 빌드를 생성할 수 있는 제어 영역을 표시합니다. 이 기능은 데이터가 많을 때 특히 유용합니다. 자세한 내용은 "**테스트 빌드 패널 표시**"(297 페이지)에서 확인하십시오.

데이터 테이블로 만들기 보고서에서 데이터 테이블을 만듭니다. 서로 다른 행 테이블의 라벨은 동일한 구조로 매핑되지 않을 수 있으므로 각 행 테이블마다 데이터 테이블이 하나씩 생성됩니다.

전체 경로 열 이름 생성된 데이터 테이블의 열 이름에 그룹화 열의 정규화된 열 이름을 사용합니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 "JMP 보고서" 장에서 확인하십시오.

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다.

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다. 이 기능을 지원하는 플랫폼에서 "자동 재계산" 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다.

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다.

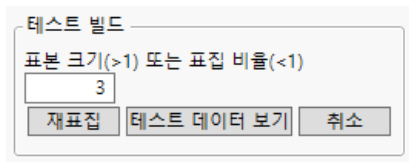
"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 "시작하기"장에서 확인하십시오.

테스트 빌드 패널 표시

데이터 테이블이 매우 큰 경우에는 데이터 테이블의 작은 부분집합을 사용하여 여러 가지 테이블 레이아웃을 시도해 본 후 요약 정보를 가장 잘 보여 주는 테이블을 찾을 수 있습니다. 이 경우 JMP에서는 지정된 크기의 랜덤 부분집합을 생성한 후, 테이블을 작성할 때 이 부분집합을 사용합니다. 테스트 빌드 기능을 사용하려면 다음을 수행하십시오.

1. "테이블 생성"의 빨간색 삼각형을 클릭하고 **테스트 빌드 패널 표시**를 선택합니다.
2. 그림 9.12에 표시된 것과 같이 **표본 크기 (>1) 또는 표집 비율 (<1)** 아래의 상자에 원하는 표본 크기를 입력합니다. 표본 크기는 활성 테이블의 입력한 비율이거나 활성 테이블의 일부 행 수일 수 있습니다.

그림 9.12 테스트 빌드 패널



3. **재표집**을 클릭합니다.
4. JMP 데이터 테이블에서 표집된 데이터를 보려면 **테스트 데이터 보기** 버튼을 클릭합니다. 테스트 빌드 패널을 종료하면 JMP에서는 전체 데이터 테이블을 사용하여 지정된 대로 테이블을 다시 생성합니다.

마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴

"테이블 생성"에서 열을 마우스 오른쪽 버튼으로 클릭하면 다음 옵션이 표시됩니다.

삭제 선택한 열을 삭제합니다.

그룹화 열로 사용 분석 열을 그룹화 열로 변경합니다.

분석 열로 사용 그룹화 열을 분석 열로 변경합니다.

항목 라벨 변경 (개별 열 또는 내포된 열의 경우에만 표시됨) 새 라벨을 입력합니다.

테이블 결합 (범주별 열) (개별 열 또는 내포된 열의 경우에만 표시됨) 개별 열 또는 내포된 열을 결합합니다. 자세한 내용은 "[예: 여러 열을 단일 테이블에 결합](#)"(301 페이지)에서 확인하십시오.

별도의 테이블 (결합된 테이블의 경우에만 표시됨) 각 열에 대해 별도의 테이블을 생성합니다.

그룹화 열 내포 그룹화 열을 세로 또는 가로로 내포합니다.

테이블 생성 플랫폼의 추가 예

- "예 : 여러 테이블 생성 및 내용 재배열"
- "예 : 여러 열을 단일 테이블에 결합"
- "예 : 페이지 열 사용"

예 : 여러 테이블 생성 및 내용 재배열

이 예에는 다음과 같은 단계가 포함되어 있습니다.

1. "개수 테이블 생성"
2. "통계량을 보여 주는 테이블 생성"
3. "테이블 내용 재배열"

개수 테이블 생성

설문 조사 대상자 중 일본, 유럽 및 미국 자동차의 소유자 수가 포함된 테이블을 생성하려고 한다고 가정해 보겠습니다. 또한 이 인원 수를 자동차 크기별로 세분화려고 합니다.

그림 9.13 자동차 소유자 수를 보여 주는 테이블

country	size	N
American	Large	36
	Medium	53
	Small	26
European	Large	4
	Medium	17
	Small	19
Japanese	Large	2
	Medium	54
	Small	92

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Poll.jmp 를 엽니다.
2. **분석 > 테이블 생성**을 선택합니다.
3. country 를 클릭하고 "행에 대한 놓기 영역" 으로 드래그합니다.
4. size 를 클릭하고 country 머리글의 오른쪽으로 드래그합니다.

그림 9.14 테이블에 추가된 country 및 size

테이블 생성

테이블에 추가하려면 열 또는 통계량을 테이블의 열 머리글 또는 행 라벨 영역으로 드래그앤드롭하십시오.

실행 취소

다시 시작

완료

6개 열

sex

marital status

age

country

size

type

빈도

가중치

페이지 열

그룹화 열에 대한 결측값 포함

그룹화 열 개수별 정렬

집계 통계량 추가

기본 통계량

형식 변경

N

평균

표준편차

최소값

최대값

범위

% 총계

결측값 수

범주 수

합

가중합

분산

표준 오차

CV

중앙값

기하평균

사분위수 범위

분위수

열 %

행 %

모두

country	size	N
American	Large	36
	Medium	53
	Small	26
European	Large	4
	Medium	17
	Small	19
Japanese	Large	2
	Medium	54
	Small	92

통계량을 보여 주는 테이블 생성

각 크기의 자동차를 소유한 사람의 나이 평균 및 표준편차를 확인하려고 한다고 가정해 보겠습니다.

그림 9.15 나이의 평균 및 표준편차를 보여 주는 테이블

country	size			
American	Large	age	평균	33.8
			표준편차	6.167
	Medium	age	평균	31.1
			표준편차	6.976
	Small	age	평균	30.4
			표준편차	5.846
European	Large	age	평균	30.5
			표준편차	5.802
	Medium	age	평균	30.9
			표준편차	6.194
	Small	age	평균	30.8
			표준편차	5.388
Japanese	Large	age	평균	28.5
			표준편차	4.95
	Medium	age	평균	30.2
			표준편차	4.668
	Small	age	평균	29.7
			표준편차	5.928

1. 그림 9.14 에서 시작합니다 . age 를 클릭하고 size 머리글의 오른쪽으로 드래그합니다 .
2. 평균을 클릭하고 " 합 " 위로 드래그합니다 .
3. 표준편차를 클릭하고 " 평균 " 아래로 드래그합니다 .

테이블의 " 평균 " 아래에 " 표준편차 " 가 배치됩니다 . " 평균 " 위에 " 표준편차 " 를 놓으면 테이블의 " 평균 " 위에 " 표준편차 " 가 배치됩니다 .

그림 9.16 테이블에 추가된 Age, 평균 및 표준편차

테이블 생성

테이블에 추가하려면 열 또는 통계량을 테이블의 열 머리글 또는 행 라벨 영역으로 드래그앤드롭하십시오.

실행 취소

다시 시작

완료

6개 열

sex

marital status

age

country

size

type

빈도

가중치

페이지 열

N

평균

표준편차

최소값

최대값

범위

% 총계

결측값 수

범주 수

합

가중합

분산

표준 오차

CV

중앙값

기하평균

사분위수 범위

분위수

열 %

행 %

모두

☐ 그룹화 열에 대한 결측값 포함

☐ 그룹화 열 개수별 정렬

☐ 집계 통계량 추가

기본 통계량

형식 변경

country	size			
American	Large	age	평균	33.8
			표준편차	6.167
	Medium	age	평균	31.1
European	Small	age	표준편차	6.976
			평균	30.4
	Large	age	표준편차	5.846
Japanese	Medium	age	평균	30.5
			표준편차	5.802
	Small	age	평균	30.9
	Large	age	표준편차	6.194
			평균	30.8
	Medium	age	표준편차	5.388
	Small	age	평균	28.5
			표준편차	4.95
	Large	age	평균	30.2
	Medium	age	표준편차	4.668
			평균	29.7
	Small	age	표준편차	5.928

테이블 내용 재배열

크기를 맨 위에 배치하여 교차표 레이아웃을 표시하려고 한다고 가정해 보겠습니다 .

그림 9.17 맨 위에 size 표시

	size					
	Large		Medium		Small	
	평균	표준편차	평균	표준편차	평균	표준편차
country	age	age	age	age	age	age
American	33.8	6.167	31.1	6.976	30.4	5.846
European	30.5	5.802	30.9	6.194	30.8	5.388
Japanese	28.5	4.95	30.2	4.668	29.7	5.928

테이블 내용을 재배열하려면 다음 단계를 수행하십시오 .

1. 그림 9.16 에서 시작합니다 . size 머리글을 클릭하고 테이블 머리글의 오른쪽으로 드래그합니다 .

그림 9.18 size 이동

country	size			
American	Large	age	평균	33.8
			표준편차	6.167
	Medium	age	평균	31.1
			표준편차	6.976
	Small	age	평균	30.4
			표준편차	5.846
	Large	age	평균	30.5
European			표준편차	5.802
	Medium	age	평균	30.9
			표준편차	6.194
	Small	age	평균	30.8
			표준편차	5.388
	Large	age	평균	28.5
Japanese			표준편차	4.95
	Medium	age	평균	30.2
			표준편차	4.668
	Small	age	평균	29.7
			표준편차	5.928

		size			
country			Large	Medium	Small
American	age	평균	33.8	31.1	30.4
		표준편차	6.167	6.976	5.846
European	age	평균	30.5	30.9	30.8
		표준편차	5.802	6.194	5.388
Japanese	age	평균	28.5	30.2	29.7
		표준편차	4.95	4.668	5.928

2. "age" 를 클릭하고 Large Medium Small 머리글 아래로 드래그합니다 .
3. " 평균 " 과 " 표준편차 " 를 모두 선택하고 Large 머리글 아래로 드래그합니다 .

이제 테이블의 데이터가 명확하게 표시됩니다 . 따라서 자동차 소유자 나이의 평균과 표준편차를 자동차 크기 및 국가별로 세분해서 확인하기가 쉬워졌습니다 .

예 : 여러 열을 단일 테이블에 결합

아동의 인기도에 대한 자가 보고 요인 (성적 , 스포츠 , 외모 , 경제 여건) 의 중요도를 나타내는 학생 데이터가 있습니다 . 이러한 모든 요인을 추가 통계량 및 요인을 포함하는 결합된 단일 테이블에서 확인하려고 한다고 가정해 보겠습니다 .

그림 9.19 인구 통계학 데이터 추가

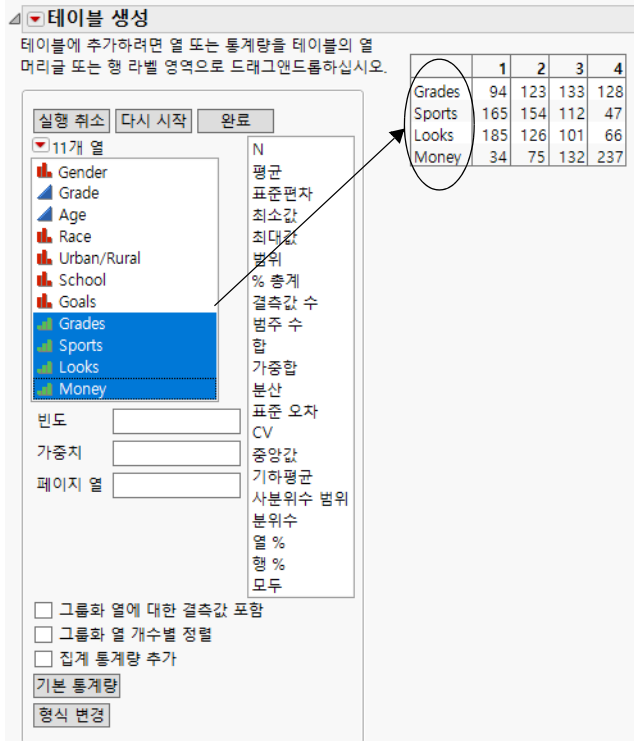
		% 총계															
		Urban/Rural															
Gender		Rural					Suburban					Urban					
		1	2	3	4	모두	1	2	3	4	모두	1	2	3	4	모두	
boy	Grades	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%	
	Sports	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%	
	Looks	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%	
	Money	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%	
girl	Grades	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%	
	Sports	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%	
	Looks	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%	
	Money	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%	

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Children's Popularity.jmp 를 엽니다 .

2. 분석 > 테이블 생성을 선택합니다.

3. Grades, Sports, Looks 및 Money 를 선택하고 " 행에 대한 농기 영역 " 으로 드래그합니다.

그림 9.20 범주별 열



결합된 단일 테이블이 나타납니다.

각 범주의 등급 (1 ~ 4) 백분율에 대한 테이블을 생성하십시오.

4. Gender 를 왼쪽의 빈 머리글로 드래그합니다.

5. % 총계를 번호가 매겨진 머리글 위로 드래그합니다.

6. 모두를 번호 4 옆으로 드래그합니다.

그림 9.21 테이블에 추가된 "Gender", "% 총계 " 및 " 모두 "

		% 총계				
Gender		1	2	3	4	모두
boy	Grades	8.16%	12.76%	14.02%	12.55%	47.49%
	Sports	26.57%	12.55%	6.49%	1.88%	47.49%
	Looks	9.21%	15.48%	12.34%	10.46%	47.49%
	Money	3.56%	6.69%	14.64%	22.59%	47.49%
girl	Grades	11.51%	12.97%	13.81%	14.23%	52.51%
	Sports	7.95%	19.67%	16.95%	7.95%	52.51%
	Looks	29.50%	10.88%	8.79%	3.35%	52.51%
	Money	3.56%	9.00%	12.97%	26.99%	52.51%

인구 통계학 데이터를 추가하여 테이블을 더 세분합니다.

7. Urban/Rural 을 % 총계 머리글 아래로 드래그합니다.

그림 9.22 테이블에 추가된 "Urban/Rural"

		% 총계														
		Urban/Rural														
		Rural					Suburban					Urban				
Gender		1	2	3	4	모두	1	2	3	4	모두	1	2	3	4	모두
boy	Grades	2.30%	2.93%	3.56%	5.02%	13.81%	2.72%	5.86%	5.02%	5.02%	18.62%	3.14%	3.97%	5.44%	2.51%	15.06%
	Sports	6.49%	3.97%	2.72%	0.63%	13.81%	11.30%	4.60%	2.30%	0.42%	18.62%	8.79%	3.97%	1.46%	0.84%	15.06%
	Looks	3.14%	4.60%	2.72%	3.35%	13.81%	3.77%	5.86%	5.44%	3.56%	18.62%	2.30%	5.02%	4.18%	3.56%	15.06%
	Money	1.88%	2.30%	4.81%	4.81%	13.81%	0.84%	2.30%	5.86%	9.62%	18.62%	0.84%	2.09%	3.97%	8.16%	15.06%
girl	Grades	4.39%	4.39%	4.39%	4.18%	17.36%	2.51%	3.35%	2.93%	4.18%	12.97%	4.60%	5.23%	6.49%	5.86%	22.18%
	Sports	2.30%	6.69%	5.44%	2.93%	17.36%	1.67%	3.56%	5.65%	2.09%	12.97%	3.97%	9.41%	5.86%	2.93%	22.18%
	Looks	9.62%	3.35%	3.35%	1.05%	17.36%	7.95%	2.93%	1.26%	0.84%	12.97%	11.92%	4.60%	4.18%	1.46%	22.18%
	Money	1.05%	2.93%	4.18%	9.21%	17.36%	0.84%	3.14%	3.14%	5.86%	12.97%	1.67%	2.93%	5.65%	11.92%	22.18%

지방, 교외 및 도시 지역에 사는 소년의 경우 스포츠가 인기에 가장 중요한 요인임을 알 수 있습니다. 지방, 교외 및 도시 지역에 사는 소녀의 경우에는 외모가 인기에 가장 중요한 요인입니다.

예 : 페이지 열 사용

남학생과 여학생의 키 측정값이 포함된 데이터가 있습니다. 학생의 나이별 키 합을 보여 주는 테이블을 생성하려고 합니다. 그런 다음 데이터를 성별별로 서로 다른 테이블에 층화하려고 합니다. 이렇게 하려면 층화 열을 페이지 열로 추가하여 각 그룹에 대한 페이지를 작성합니다.

그림 9.23 학생의 성별별 키 평균

sex = F		sex = M	
age	평균	age	평균
12	58.6	12	57.3
13	59.0	13	61.3
14	62.6	14	65.3
15	63.0	15	65.2
16	62.5	16	68.0
17	62.0	17	69.0

여학생

남학생

- 도움말 > 샘플 데이터 라이브러리를 선택하고 Big Class.jmp 를 엽니다.
- 분석 > 테이블 생성을 선택합니다.
height 변수를 검토할 것이므로 테이블의 맨 위에 이 변수를 표시하려고 합니다.
- height 를 클릭하고 "열에 대한 놓기 영역" 으로 드래그합니다.
나이별로 통계량을 살펴볼 것이므로 age 변수를 나란히 표시하려고 합니다.
- age 를 클릭하고 숫자 "2502" 옆의 빈 셀로 드래그합니다.
- "평균" 을 클릭하고 "합" 이라는 셀로 드래그합니다.

6. sex 를 클릭하고 **페이지 열**로 드래그합니다 .
7. " 페이지 열 " 목록에서 "F" 를 선택하여 여학생만의 평균 키를 표시합니다 .
8. " 페이지 열 " 목록에서 "M" 을 선택하여 남학생만의 평균 키를 표시합니다 . **선택된 항목 없음** 을 선택하여 모든 값을 표시할 수도 있습니다 .

그림 9.24 페이지 열 사용

테이블 생성

테이블에 추가하려면 열 또는 통계량을 테이블의 열 머리글 또는 행 라벨 영역으로 드래그앤드롭하십시오. **sex = M**

실행 취소 **다시 시작** **완료**

▼ 5개 열

- name
- age
- sex
- height
- weight

빈도

가중치

페이지 열

N

평균

표준편차

최소값

최대값

범위

% 총계

결측값 수

범주 수

합

가중합

분산

표준 오차

CV

중앙값

기하평균

사분위수 범위

분위수

열 %

행 %

모두

☐ 그룹화 열에 대한 결측값 포함
☐ 그룹화 열 개수별 정렬
☐ 집계 통계량 추가

height	
age	평균
12	57.3
13	61.3
14	65.3
15	65.2
16	68.0
17	69.0

시뮬레이션 플랫폼은 JMP Pro 에서만 사용할 수 있습니다.

시뮬레이션 기능으로는 강력한 모수적 및 비모수적 시뮬레이션을 수행할 수 있습니다. 시뮬레이션 기능을 사용하여 수행할 수 있는 작업은 다음과 같습니다.

- 붓스트랩을 확장하여 모수적 붓스트랩을 제공합니다.
- 비표준 상황에서 검정력을 계산합니다.
- 비표준 상황에서 예측값 같은 통계량의 분포와 신뢰 구간에 대해 근사값을 산출합니다.
- 순열 검정을 수행합니다.
- 모형의 예측 변수에 대한 가정의 효과를 탐색합니다.
- 모형을 기준으로 다양한 "what if" 시나리오를 탐색합니다.
- 새로운 통계적 방법 또는 기존 통계적 방법을 평가합니다.

시뮬레이션 옵션은 붓스트랩을 지원하는 모든 보고서를 포함하여 다양한 보고서에서 사용할 수 있습니다. 시뮬레이션 옵션에 액세스하려면 보고서를 마우스 오른쪽 버튼으로 클릭합니다.

그림 10.1 시뮬레이션을 통한 검정력 분석



목차

시뮬레이션 플랫폼 개요	307
시뮬레이션 사용 예	307
분산 성분에 대한 준모수적 신뢰 구간 구성	308
순열 검정 수행	313
일반화 회귀의 요인 유지 탐색	315
비선형 모형에 대한 전향적 검정력 분석	320
시뮬레이션 창 시작	329
시뮬레이션 창	329
시뮬레이션 결과 테이블	330
시뮬레이션 결과 보고서	331
시뮬레이션 검정력 보고서	331

JMP PRO 시뮬레이션 플랫폼 개요

시뮬레이션 플랫폼에서는 보고서의 통계량 열에 대해 시뮬레이션된 결과를 제공합니다. 보고서의 통계량 열을 마우스 오른쪽 버튼으로 클릭하고 "시뮬레이션"을 선택합니다. 그런 다음 "시뮬레이션" 창에서 시뮬레이션의 기초로 사용할 데이터 테이블 내의 열을 지정합니다. 이 열은 스위치 아웃할 열입니다. 이 열은 분석에서 어떤 역할이든 할 수 있습니다. 특히 모형에서는 반응 또는 예측 변수가 될 수 있습니다. 다음으로는 데이터 테이블에서 시뮬레이션에 사용할 계산식이 포함된 열을 지정합니다. 이 열은 스위치 인할 열입니다. 이 열은 스위치 아웃한 열의 대리 역할을 합니다.

참고: 데이터 테이블에는 랜덤 성분이 있는 열이 포함되어 있어야 합니다.

이 방법의 작동 방식은 다음과 같습니다. 스위치 인한 계산식 열의 계산식을 기반으로 시뮬레이션된 값을 포함하는 열이 생성됩니다. 관심 있는 통계량이 포함된 보고서를 생성했던 전체 분석이 시뮬레이션된 값을 포함하는 이 새 열을 사용하여 다시 실행되어 스위치 아웃한 열을 대체합니다. 이 과정이 N 번 반복됩니다. 여기서 N 은 사용자가 지정한 총 표본 수입니다.

시뮬레이션 분석에서는 분석 요약에 보여 주는 출력 데이터 테이블을 생성합니다.

- 데이터 테이블의 각 행은 시뮬레이션된 값을 포함하는 한 열에 대한 분석 결과를 나타냅니다.
- 시뮬레이션에 관련된 보고서 테이블의 각 행마다 열이 하나씩 있습니다.
- 분석을 용이하게 하는 스크립트가 있습니다.

팁: 시뮬레이션 플랫폼에서는 시뮬레이션 기능이 호출된 플랫폼 보고서에 표시된 전체 분석을 다시 실행합니다. 따라서 보고서에 포함된 관련 없는 분석으로 인해 선택한 열에 대한 시뮬레이션이 느리게 실행될 수 있습니다. 시뮬레이션이 오래 걸릴 경우에는 플랫폼 보고서에서 관련 없는 옵션을 제거한 후 시뮬레이션을 실행하십시오.

시뮬레이션 기능은 연관성 분석, 다이어그램, 다차원 척도법, 다중 요인 분석, 신뢰도 블록 다이어그램, 신뢰도 예측, 수리 가능 시스템 시뮬레이션, 반응 변수 선별, 텍스트 탐색기를 제외한 모든 통계 플랫폼에서 사용할 수 있습니다.

JMP PRO 시뮬레이션 사용 예

이 섹션에서는 몇 가지 시뮬레이션 사용 예를 제공합니다. 다른 설명서에서 추가 예도 제공됩니다(아래 참조).

- "분산 성분에 대한 준모수적 신뢰 구간 구성"
- "순열 검정 수행"
- "일반화 회귀의 요인 유지 탐색"

- "비선형 모형에 대한 전향적 검정력 분석"
- 비정규 변수의 Ppk 및 부적합률에 대한 신뢰 구간 시뮬레이션 방법을 보여 주는 예는 Quality and Process Methods 의 "Process Capability" 장에서 확인하십시오 .

JMP PRO 분산 성분에 대한 준모수적 신뢰 구간 구성

이 예에서는 온도, 시간 및 촉매의 양이 반응에 미치는 효과를 알아보려고 합니다. 온도는 매우 변경하기 힘든 변수 (주요 요인) 이고, 시간은 변경하기 힘든 변수 (하위요인) 이며, 촉매 양은 변경하기 쉬운 변수입니다. 주요 및 하위요인에 대한 자세한 내용은 실험 설계 가이드의 "사용자 설계" 장에서 확인하십시오 .

이러한 주요 및 하위요인 분산 성분에 대한 준모수적 신뢰 구간을 구하려고 합니다. 이전 연구에서는 주요 표준편차가 오차 표준편차의 약 2 배이고, 하위요인 오차는 오차 표준편차의 약 1.5 배인 것으로 나타났습니다. 분산 성분이 점근적으로 정규 분포를 따른다고 가정하는 REML 보고서에서 제공된 Wald 구간은 포함 범위가 매우 작은 특성이 있습니다 . 따라서 시뮬레이션된 분산 성분 분포의 백분위수를 사용하여 신뢰 구간을 구합니다 .

이 예에는 다음과 같은 작업이 포함되어 있습니다 .

- 분할 - 분할구 실험을 위한 사용자 설계를 구성합니다 . 자세한 내용은 "설계 구성" (308 페이지) 에서 확인하십시오 .
- REML 방법을 사용하여 모형을 적합시킵니다 . 자세한 내용은 "모형 적합" (310 페이지) 에서 확인하십시오 .
- 분산 성분의 백분위수 신뢰 구간을 구하기 위해 분산 성분 추정값을 시뮬레이션합니다. 자세한 내용은 "신뢰 구간 생성" (311 페이지) 에서 확인하십시오 .

JMP PRO 설계 구성

이 섹션의 단계를 건너뛰려면 **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Catalyst Design.jmp 를 여십시오 . Catalyst Design.jmp 데이터 테이블에서 **DOE Simulate** 스크립트 옆의 녹색 삼각형을 클릭하십시오 . 그런 다음 "모형 적합" (310 페이지) 절차로 이동하십시오 .

1. **DOE > 사용자 설계**를 선택합니다 .
2. "요인" 개요에서 **N 개 요인 추가** 옆에 3 을 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. 추가된 요인을 두 번 클릭하여 이름을 Temperature, Time 및 Catalyst 로 바꿉니다 .
이러한 요인의 "값" 은 기본값 -1 및 1 을 유지합니다 .
5. Temperature 에서 **쉬움**을 클릭하고 **매우 어려움**을 선택합니다 .
이렇게 하면 Temperature 가 주요 요인으로 정의됩니다 .
6. Time 에서 **쉬움**을 클릭하고 **어려움**을 선택합니다 .
이렇게 하면 Time 이 하위요인으로 정의됩니다 .

7. **계속**을 클릭합니다.

8. " 모형 " 개요에서 **교호작용 > 2 차**를 선택합니다.

이렇게 하면 모형에 모든 2 원 교호작용이 추가됩니다.

9. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **반응 시뮬레이션**을 선택합니다.

이렇게 하면 " 테이블 생성 " 을 선택하여 설계 테이블을 구성한 후에 " 반응 시뮬레이션 " 창이 열립니다.

참고 : 10 단계에서 " 난수 시드값 " 을 설정하고 11 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 설계가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

10. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 . "12345" 를 입력하고 **확인**을 클릭합니다 .

11. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **시작 수**를 선택합니다 . "1000" 을 입력하고 **확인**을 클릭합니다 .

12. **설계 생성**을 클릭합니다 .

13. **테이블 생성**을 클릭합니다 .

참고 : Y 및 Y 시뮬레이션 열의 항목은 그림 10.2 에 표시된 것과 다를 수 있습니다 .

그림 10.2 설계 테이블

설계 기준	주구	하위구	Temperature	Time	Catalyst	Y	Y 시뮬레이션
모형	1	1	1	-1	1	1.1665488716	1.16654887
설계 평가	2	1	1	-1	1	1.8146975238	1.81469752
혼합 조합	3	1	1	1	-1	0.5705597437	0.57055974
DOE 시뮬레이션	4	1	2	1	1	0.205020624	0.20502062
DOE 대화상자	5	1	2	1	-1	1.2521972858	1.25219729
	6	1	2	1	-1	0.5239669952	0.523967
	7	2	3	1	-1	-0.458214576	-0.4582146
	8	2	3	1	-1	-1.534981844	-1.5349818
	9	2	3	1	-1	-0.011730996	-0.011731
	10	2	4	1	1	7.407179478	7.40717948
	11	2	4	1	1	9.7482802841	9.74828028
	12	2	4	1	1	7.8122598279	7.81225983
	13	3	5	-1	-1	0.4107869363	0.41078694
	14	3	5	-1	-1	0.8135478898	0.81354789
	15	3	5	-1	-1	0.9465253996	0.9465254
	16	3	6	-1	1	0.5042748955	0.5042749
	17	3	6	-1	1	3.008172744	3.00817274
	18	3	6	-1	1	-0.552204825	-0.5522048
	19	4	7	-1	-1	-2.823204554	-2.8232046
	20	4	7	-1	-1	1.0770933813	1.07709338
	21	4	7	-1	-1	-2.959587735	-2.9595877
	22	4	8	-1	1	-0.923826938	-0.9238269
	23	4	8	-1	1	-4.336208768	-4.3362088
	24	4	8	-1	1	-3.559568906	-3.5595689

그림 10.3 반응 시뮬레이션 창

효과	Y
절편	1
Temperature	1
Time	1
Catalyst	1
Temperature*Time	1
Temperature*Catalyst	1
Time*Catalyst	1

계수 재설정

☒ 정규 오차 σ : 1
 주구 σ : 1
 하위구 σ : 1

☐ 이항
☐ Poisson

적용

설계 테이블과 "반응 시뮬레이션" 창이 나타납니다. 설계 테이블에는 **DOE 시뮬레이션** 스크립트가 포함되어 있습니다. 언제든지 이 스크립트를 실행하여 다른 모수 값을 지정할 수 있습니다.

다음 섹션으로 진행하여 주구 및 하위구 오차의 표준편차를 지정하고 첫 번째 시뮬레이션 값 집합에 REML 모형을 적합시킬 수 있습니다.

JMP PRO 모형 적합

주구 및 하위구 오차가 정규 분포를 따른다고 가정하십시오. 표준편차 추정값에 따르면 오차 표준편차가 약 1 단위일 경우 주구 표준편차는 약 2 단위이고 하위구 표준편차는 약 1.5 단위입니다. 주구 및 하위구 변동에만 관심이 있으므로 "반응 시뮬레이션" 개요의 "효과"에 할당된 값은 변경하지 않아도 됩니다.

1. "분포" 패널 (그림 10.3)에서 **주구 σ** 옆에 2를 입력합니다.

기본적으로 "정규" 분포가 선택되어 있습니다. 따라서 계산식에 정규 오차가 추가됩니다.

2. **하위구 σ** 옆에 1.5를 입력합니다.

3. **적용**을 클릭합니다.

데이터 테이블에서 Y 시뮬레이션의 계산식이 업데이트되어 지정한 내용이 반영됩니다. 계산식을 보려면 "열" 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다.

4. 데이터 테이블에서 **모형** 스크립트 옆의 녹색 삼각형을 클릭합니다.

5. **Y** 버튼 옆의 Y 변수를 클릭하고 **제거**를 클릭합니다.

6. Y 시뮬레이션을 클릭하고 **Y** 버튼을 클릭합니다.

이렇게 하면 Y가 시뮬레이션 계산식이 포함된 열로 대체됩니다.

7. **실행**을 클릭합니다.

적합된 모형은 시뮬레이션된 반응의 단일 집합을 기반으로 합니다.

참고 : Y 시뮬레이션의 값은 무작위로 생성되므로 보고서의 항목은 그림 10.4 에 표시된 것과 다를 수 있습니다.

그림 10.4 Wald 신뢰 구간을 보여 주는 REML 보고서

REML 분산 성분 추정값							
임의 효과	분산 비율	분산 성분	표준 오차	95% 하한	95% 상한	Wald p 값	총계 백분율
Whole Plots	3.8107856	1.556364	1.8007472	-1.973036	5.0857636	0.3874	68.445
Subplots	0.7568606	0.3091097	0.4661521	-0.604532	1.222751	0.5073	13.594
잔차		0.4084103	0.160228	0.2146172	1.060295		17.961
합계		2.2738839	1.8035082	0.7468087	28.115033		100.000
-2*로그 우도 = 63.890021519							
참고: 합계는 양의 분산 성분의 합입니다.							
음수 추정값을 포함한 합계 = 2.2738839							

JMP PRO 신뢰 구간 생성

다음으로는 분산 성분의 값을 시뮬레이션하고 이를 사용하여 시뮬레이션된 백분위수 신뢰 구간을 구성합니다.

1. "REML 분산 성분 추정값" 개요에서 **분산 성분** 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다.

그림 10.5 시뮬레이션 창

스위치 아웃할 열: Y Simulated, Temperature, Time, Catalyst, Whole Plots, Subplots

스위치 인할 열: Y Simulated

표본 수: 2500

난수 시드값: .

확인 취소 도움말

시뮬레이션에서는 모형을 실행하는 데 사용된 Y 시뮬레이션 열이 각 시뮬레이션에 대해 새로운 시뮬레이션 값 열을 생성하는 Y 시뮬레이션 열의 새 인스턴스로 대체됩니다. 마우스 오른쪽 버튼으로 클릭하여 선택 상태로 표시된 **분산 성분** 열이 "모수 추정값" 테이블에 나열된 각 효과에 대해 시뮬레이션됩니다.

2. **표본 수** 옆에 **200** 을 입력합니다.
3. (선택 사항) **난수 시드값** 옆에 456 을 입력합니다.

이렇게 하면 그림 10.6 에 표시된 값이 1 행의 값만 제외하고 모두 재현됩니다.

4. **확인**을 클릭합니다.

1 행의 항목은 그림 10.6 에 표시된 것과 다를 수 있습니다.

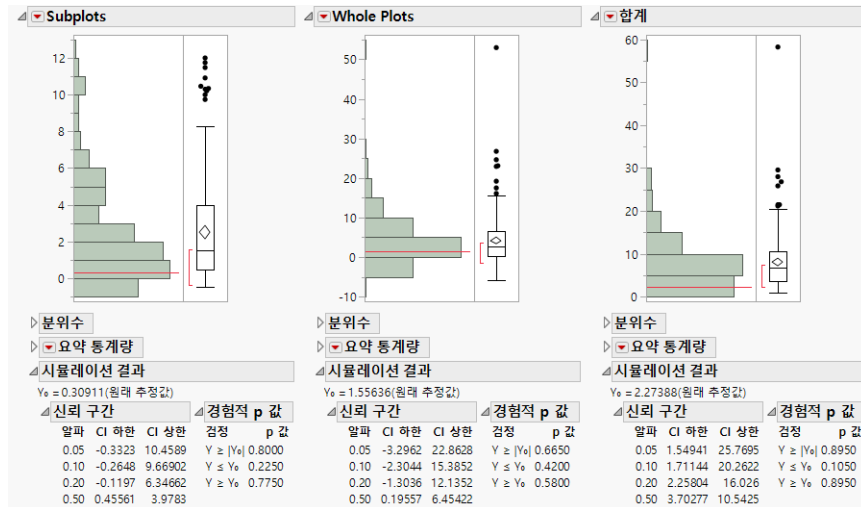
그림 10.6 분산 성분에 대한 시뮬레이션 결과 테이블 (일부분)

최소 제곱 적합 시뮬레이션 ...		Y	SimID	Subplots	Whole Plots	잔차	합계	
난수 시드값 456		1	Y Simulated	0	0.3091096508	1.5563639995	0.4084102799	2.2738839302
결합 데이터 테이블 생성		2	Y Simulated	1	9.9978320675	0.1160090855	1.7118437836	11.825684937
분포		3	Y Simulated	2	4.5617762065	3.3652848741	0.7259348275	8.6529959081
		4	Y Simulated	3	3.0383907272	22.95631251	0.8416572537	26.836360491
월(6/0)		5	Y Simulated	4	-0.176636748	12.156574946	0.9872034755	13.143778421
Y *		6	Y Simulated	5	-0.153702674	-0.060658009	1.3514412755	1.3514412755
SimID		7	Y Simulated	6	1.1063669307	11.804365807	0.9837455205	13.894478258
Subplots		8	Y Simulated	7	0.1630909822	2.6824797817	0.8559296173	3.7015003812
Whole Plots		9	Y Simulated	8	0.2154635485	26.765100099	1.0422128078	28.022776455
잔차		10	Y Simulated	9	4.0615625437	12.75877132	1.0302645404	17.850598404
합계		11	Y Simulated	10	1.8169128804	-0.433818317	1.2569403028	3.0738531832
		12	Y Simulated	11	6.9696566519	2.4634884406	1.67195118802	11.105096973
행		13	Y Simulated	12	1.9758939136	4.2166345958	1.0326091926	7.2251377019
모든 행 201		14	Y Simulated	13	1.1584152266	2.8027318323	0.7050109246	4.6661579836
선택됨 0		15	Y Simulated	14	5.2888052789	0.3194564512	0.9281404099	6.53640214
제외 1		16	Y Simulated	15	1.4827905134	-0.270350632	1.6586718874	3.1414624009
숨김 0		17	Y Simulated	16	0.7636963822	-0.192580268	0.9471174937	1.7108138759
라벨 지정 0		18	Y Simulated	17	1.1919618966	0.756815441	0.9782751744	2.927052512

"최소 제곱 적합 시뮬레이션 결과 (분산 성분)" 데이터 테이블의 첫 번째 행은 분산 성분의 초기값을 포함하므로 제외됩니다. 나머지 행에는 시뮬레이션된 값이 포함되어 있습니다.

5. 분포 스크립트를 실행합니다.

그림 10.7 분산 성분에 대한 분포 그림 (일부분)



"시뮬레이션 결과" 보고서에는 각 분산 성분에 대해 다양한 신뢰 수준의 신뢰 구간이 표시됩니다. 각 테이블의 알파 = 0.05 행에 표시된 95% 구간을 REML 보고서에 제공된 구간 (그림 10.4) 과 비교하십시오.

- 주구 분산 성분에 대해 시뮬레이션된 95% 신뢰 구간은 -3.296 ~ 22.863 입니다. REML 보고서에 제공된 Wald 구간은 -1.973 ~ 5.086 입니다.

- 하위구 분산 성분에 대해 시뮬레이션된 95% 신뢰 구간은 -0.332 ~ 10.459 입니다 . REML 보고서에 제공된 Wald 구간은 -0.605 ~ 1.223 입니다 .

시뮬레이션을 사용하여 구한 구간은 단일 값 집합으로부터 계산된 REML 구간보다 상당히 더 넓습니다 . 보다 정확한 구간을 구하려면 더 많은 수의 시뮬레이션을 실행하는 것이 좋습니다 .

JMP PRO 순열 검정 수행

이 예에서는 세 가지 약물이 통증에 미치는 효과를 연구하려고 합니다 . 관심 사항은 약물의 효과가 다른지 여부입니다 . 표본 크기가 매우 작고 일반적인 ANOVA 가정에 다소 위배되기 때문에 시뮬레이션 기능을 사용하여 순열 검정을 수행할 수 있습니다 .

먼저 , 세 가지 약물의 투여 시 통증 측정값을 무작위로 뒤섞는 계산식을 구성합니다 . 효과가 없다는 귀무가설하에 이러한 배분은 모두 동일한 가능성을 갖습니다 . 따라서 이 방식으로 얻은 F 비는 귀무가설하의 F 비에 근사한 분포를 따릅니다 . 마지막으로 , 관측된 F 비 값을 시뮬레이션으로 구한 귀무 분포와 비교합니다 .

JMP PRO 시뮬레이션 계산식 정의

1. 도움말 > 샘플 데이터 라이브러리를 선택하고 Analgesics.jmp 를 엽니다 .
2. 열 > 새 열 을 선택합니다 .
3. " 열 이름 " 에 Pain Shuffled 를 입력합니다 .
4. " 열 특성 " 목록에서 계산식 을 선택합니다 .
5. 함수 목록에서 행 > Col Stored Value 를 선택합니다 .
6. 열 목록에서 pain 을 두 번 클릭합니다 .
7. 편집기 패널의 위쪽에 있는 기호 목록에서 삼입 키 (^) 를 클릭합니다 .
8. 함수 목록에서 난수 > Col Shuffle 을 선택합니다 .

그림 10.8 완료된 계산식

Col Stored Value (pain , Col Shuffle ())

이 계산식은 pain 열의 항목을 무작위로 뒤섞습니다 .

9. 계산식 편집기 창에서 확인 을 클릭합니다 .
10. 열 정보 창에서 확인 을 클릭합니다 .

순열 검정 수행

1. 분석 > X 로 Y 적합 을 선택합니다 .
2. pain 을 선택하고 Y, 반응을 클릭합니다 .

3. drug 를 선택하고 **X, 요인**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. " 일원 분석 " 의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .

그림 10.9 분산 분석 보고서

분산 분석					
소스	DF	제곱합	평균 제곱	F 비	Prob > F
drug	2	99.89459	49.9473	6.2780	0.0053*
오차	30	238.67877	7.9560		
수정 합계	32	338.57335			

F 비가 6.2780 입니다 .

6. " 분산 분석 " 개요에서 "F 비 " 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .
7. " 스위치 아웃할 열 " 목록에서 **pain** 을 클릭합니다 .
8. " 스위치 인할 열 " 목록에서 **Pain Shuffled** 를 클릭합니다 .
9. **표본 수** 옆에 1000 을 입력합니다 .
10. (선택 사항) **난수 시드값** 옆에 456 을 입력합니다 .

이렇게 하면 이 예의 값이 재현됩니다 .

그림 10.10 완료된 시뮬레이션 창

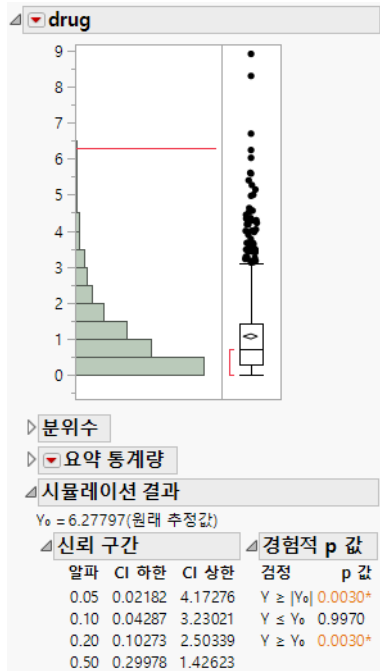
스위치 아웃할 열	스위치 인할 열
<div>pain</div> <div>drug</div>	<div>Pain Shuffled</div>
표본 수	1000
난수 시드값	456
<div>확인</div> <div>취소</div> <div>도움말</div>	

11. **확인**을 클릭합니다 .

시뮬레이션 결과 테이블에서 " 수정 합계 " 및 " 오차 " 열은 비어 있습니다 . 이는 " 분산 분석 " 테이블의 F 비 값이 **drug** 에만 적용되기 때문입니다 .

12. 시뮬레이션된 값 테이블에서 **분포** 스크립트를 실행합니다 .

그림 10.11 귀무 분포하에서 시뮬레이션된 F 비 분포



관측된 F 비 값 6.2780 은 히스토그램에 빨간색 선으로 표시됩니다 . 이 값은 시뮬레이션된 F 비 귀무 분포의 상위 0.5% 에 속합니다 . 이는 세 가지 약물이 **pain** 에 미치는 효과가 다르다는 강력한 증거를 제시합니다 .

JMP PRO 일반화 회귀의 요인 유지 탐색

이 예에서는 한 제약 회사가 정제의 체내 용해도와 용해도에 영향을 미칠 수 있는 다양한 요인에 관한 과거 정보를 갖고 있다고 가정합니다 . 용해도가 70 미만인 정제는 결함이 있는 것으로 간주됩니다 . 따라서 용해도에 영향을 미치는 요인을 파악하려고 합니다 .

이 예에는 다음과 같은 작업이 포함되어 있습니다 .

- 일반화 회귀 모형을 구성합니다 .
- 0 이 아닌 항을 사용하여 축소 모형을 적합시킵니다 .
- 축소 모형에 따라 시뮬레이션을 사용하여 요인 중 하나가 모형에 포함될 우도를 구합니다 .

JMP PRO 모형 적합

이 섹션에서는 일반화 회귀를 사용하여 모형을 적합시킵니다. 이 섹션의 단계를 거치지 않으려면 Tablet Production.jmp 데이터 테이블의 **Generalized Regression** 스크립트 옆에 있는 녹색 삼각형을 클릭하여 모형을 구하십시오.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Tablet Production.jmp 를 엽니다.
2. **분석 > 모형 적합**을 선택합니다.
3. **Dissolution** 을 클릭하고 **Y** 를 클릭합니다.
4. **Mill Time** 부터 **Atomizer Pressure** 까지 선택하고 **추가**를 클릭합니다.
5. "분석법" 목록에서 **일반화 회귀**를 선택합니다.
6. **실행**을 클릭합니다.
7. "모형 시작" 패널에서 **적응형** 상자를 선택합니다.
8. "모형 시작" 패널에서 **시작**을 클릭합니다.

그림 10.12 적응형 Lasso 회귀 기반의 모형

원래 예측 변수에 대한 모수 추정값							
항	추정값	표준 오차	Wald	카이제곱	Prob >		
					ChiSquare	95% 하한	95% 상한
절편	108.6406	24.332975	19.933977		<.0001*	60.948851	156.33236
Mill Time	0.130278	0.028185	21.365139		<.0001*	0.0750364	0.1855195
Screen Size[3-5]	4.1877616	0.541493	59.8106		<.0001*	3.1264548	5.2490685
Screen Size[4-5]	2.3907729	0.567217	17.765534		<.0001*	1.2790481	3.5024977
Mag. Stearate Supplier[Jones Inc-Smith Ind]	0	0	0		1.0000	0	0
Lactose Supplier[Bond Inc-James Ind]	0	0	0		1.0000	0	0
Sugar Supplier[Sour-Sweet]	0	0	0		1.0000	0	0
Talc Supplier[Rough-Smooth]	0	0	0		1.0000	0	0
Blend Time	0.7223423	0.1381848	27.325338		<.0001*	0.4515051	0.9931796
Blend Speed	0.2130888	0.2389205	0.7954524		0.3725	-0.255187	0.6813644
Compressor[Compress1-Compress2]	-0.538112	0.3993673	1.8155148		0.1778	-1.320857	0.2446338
Force	0	0	0		1.0000	0	0
Coating Supplier[Coat-Mac]	0	0	0		1.0000	0	0
Coating Supplier[Down-Mac]	0	0	0		1.0000	0	0
Coating Viscosity	0.1834341	0.0500838	13.414242		0.0002*	0.0852717	0.2815965
Inlet Temp	0	0	0		1.0000	0	0
Exhaust Temp	0	0	0		1.0000	0	0
Spray Rate	-0.204356	0.0420142	23.658267		<.0001*	-0.286702	-0.12201
Atomizer Pressure	0	0	0		1.0000	0	0
Prob >							
정규 분포 모수	추정값	표준 오차	Wald	카이제곱	ChiSquare	95% 하한	95% 상한
척도	2.0224259	0.1641142	151.86339		<.0001*	1.700768	2.3440838

"AICc 검증을 사용한 정규 적응형 Lasso 회귀" 보고서에 표시된 모수 추정값에 관심이 있습니다. 0 이 아닌 모수 추정값에 따라 이 모형은 Mill Time, Screen Size, Blend Time, Blend Speed, Compressor, Coating Viscosity 및 Spray Rate 가 Dissolution 과 관련이 있음을 나타냅니다.

JMP PRO 모형 축소

모형을 축소하기 전에 Tablet Production.jmp 데이터 테이블에 선택된 열이 없는지 확인하십시오. 선택된 열은 아래의 첫 번째 단계에서 선택 취소되지 않습니다. 선택된 열이 없도록 하면 0 인 항이 있는 열이 실수로 포함되는 경우를 방지할 수 있습니다.

이 섹션의 단계를 거치지 않으려면 Tablet Production.jmp 데이터 테이블의 **Generalized Regression Reduced Model** 스크립트 옆에 있는 녹색 삼각형을 클릭하여 축소 모형을 구하십시오.

1. "AICc 검증을 사용한 정규 적응형 Lasso 회귀" 옆의 빨간색 삼각형을 클릭하고 **활성 집합 다시 시작 > 활성 효과를 사용하여 다시 시작**을 선택합니다.

이렇게 하면 "모수 추정값" 보고서에서 0 이 아닌 계수 추정값이 있는 항이 "모형 효과 생성" 목록에 포함됩니다. 반응은 Y 로 입력되며, "일반화 회귀" 분석법이 선택됩니다.

2. **실행**을 클릭합니다.
3. "모형 시작" 패널에서 **적응형** 상자를 선택합니다.
4. "모형 시작" 패널에서 **시작**을 클릭합니다.

그림 10.13 적응형 Lasso 회귀를 사용한 축소 모형

원래 예측 변수에 대한 모수 추정값							
항	추정값	표준 오차	Wald	카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	95.142391	23.430178	16.489099		<.0001*	49.220085	141.0647
Mill Time	0.1394103	0.0275609	25.585971		<.0001*	0.0853919	0.1934288
Screen Size[3-5]	4.3323833	0.534237	65.763638		<.0001*	3.285298	5.3794685
Screen Size[4-5]	2.6331283	0.5457852	23.275584		<.0001*	1.563409	3.7028476
Blend Time	0.7583048	0.1385246	29.966364		<.0001*	0.4868017	1.029808
Blend Speed	0.4575802	0.2289168	3.9955744		0.0456*	0.0089116	0.9062488
Compressor[Compress1-Compress2]	-0.877986	0.4127286	4.5252866		0.0334*	-1.686919	-0.069053
Coating Viscosity	0.198673	0.0486798	16.656386		<.0001*	0.1032624	0.2940836
Spray Rate	-0.212753	0.0410929	26.805238		<.0001*	-0.293294	-0.132213
정규 분포 모수							
척도	추정값	표준 오차	Wald	카이제곱	Prob > ChiSquare	95% 하한	95% 상한
척도	1.9982636	0.1574951	160.97992		<.0001*	1.689579	2.3069483

Blend Speed 추정값의 신뢰 구간 (95% 하한) 은 0 을 포함하는 것과 매우 가깝습니다. 다음으로 는 시뮬레이션 연구를 수행하여 용해도 분포에서 다른 데이터 값이 관측될 때 Blend Speed 가 얼마나 자주 모형에 포함되는지 알아봅니다.

JMP PRO 모형의 Blend Speed 포함 여부 탐색

아래 단계에서는 축소 모형에 대한 보고서 (그림 10.13) 를 사용합니다.

1. "AICc 검증을 사용한 정규 적응형 Lasso 회귀" 옆의 빨간색 삼각형을 클릭하고 **열 저장 > 시뮬레이션 계산식 저장**을 선택합니다.

이렇게 하면 Dissolution 시뮬레이션 계산식이라는 새 열이 Tablet Production.jmp 데이터 테이블에 추가됩니다.

2. (선택 사항) 데이터 테이블의 "열" 패널에서 Dissolution 시뮬레이션 계산식의 오른쪽에 있는 더하기 기호를 클릭합니다.

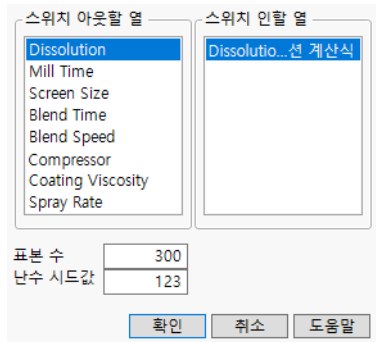
그림 10.14 시뮬레이션 계산식

$$\begin{aligned}
 & 95.142391047 \\
 & + 0.1394103417 \cdot \text{Mill Time} \\
 & + \text{Match}(\text{Screen Size}) \begin{pmatrix} \text{"3"} \Rightarrow 4.3323832797 \\ \text{"4"} \Rightarrow 2.6331282831 \\ \text{"5"} \Rightarrow 0 \\ \text{else} \Rightarrow . \end{pmatrix} \\
 & \text{Random Normal} \\
 & + 0.7583048404 \cdot \text{Blend Time} \\
 & + 0.4575801859 \cdot \text{Blend Speed} \\
 & + \text{Match}(\text{Compressor}) \begin{pmatrix} \text{"Compress1"} \Rightarrow -0.877985955 \\ \text{"Compress2"} \Rightarrow 0 \\ \text{else} \Rightarrow . \end{pmatrix} \\
 & + 0.1986730026 \cdot \text{Coating Viscosity} \\
 & + -0.212753206 \cdot \text{Spray Rate}
 \end{aligned}$$

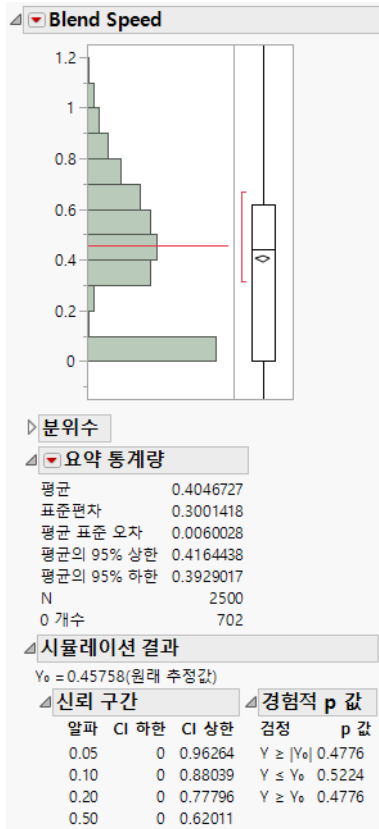
이 계산식은 각 행에 대해서 주어진 모형과 **Dissolution**의 분포 (표준편차가 약 1.998인 정규 분포로 추정)로 구할 수 있는 값을 시뮬레이션합니다.

3. **취소**를 클릭합니다.
4. 축소 모형 보고서 창으로 돌아갑니다. "원래 예측 변수에 대한 모수 추정값" 보고서에서 "추정값" 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다.
"스위치 아웃할 열" 목록에 **Dissolution**이 선택되어 있는지 확인합니다.
5. **표본 수** 옆에 300을 입력합니다.
이 시뮬레이션에서는 300개의 각 분석값에서 **Dissolution** 열을 **Dissolution** 시뮬레이션 계산식 열을 사용하여 시뮬레이션된 값으로 대체하도록 요청합니다.
6. (선택 사항) **난수 시드값**을 123으로 설정합니다.
이렇게 하면 이 예의 값이 재현됩니다.

그림 10.15 완료된 시뮬레이션 창



7. **확인**을 클릭합니다.
- 테이블의 첫 번째 행은 추정값의 초기 값을 포함하므로 제외됩니다. 나머지 행에는 시뮬레이션된 값이 포함되어 있습니다.
8. **Distribution** 스크립트를 실행합니다.
9. Ctrl 키를 누른 채 절편의 빨간색 삼각형을 클릭하고 **표시 옵션 > 요약 통계량 사용자 정의**를 선택합니다.
10. **0 개수**를 선택합니다.
11. **확인**을 클릭합니다.
12. Blend Speed 에 대한 분포 보고서로 스크롤합니다.

그림 10.16 시뮬레이션된 Blend Speed 계수 추정값의 히스토그램

이 " 요약 통계량 " 보고서에서는 시뮬레이션의 $103/300 = 34.3\%$ 에 대해 Blend Speed 추정값이 0 임을 보여 줍니다 .

JMP PRO 비선형 모형에 대한 전향적 검정력 분석

이 예에서는 6 가지 연속형 요인이 부품의 검사 합격 또는 불합격에 미치는 주효과를 알아보려고 합니다 . 반응은 이항이며 총 60 회 런이 가능합니다 .

이 예에는 다음과 같은 작업이 포함되어 있습니다 .

1. 실험을 위한 사용자 설계를 구성합니다 . 자세한 내용은 " 설계 구성 "(322 페이지) 에서 확인하십시오 .

참고 : 사용자 설계는 비선형 상황에는 최적이지 않지만 이 예에서는 간소화를 위해 비선형 설계 플랫폼 대신 사용자 설계 플랫폼을 사용할 수 있습니다 . 비선형 설계 플랫폼을 사용하여 구성한 설계가 직교 설계보다 나은 이유를 보여 주는 예는 실험 설계 가이드의 " 비선형 설계 " 장에서 확인하십시오 .

2. 일반화 선형 모형 분석법을 사용하여 로지스틱 모형을 적합시킵니다. 자세한 내용은 "일반화 선형 모형 적합"(325 페이지)에서 확인하십시오.
3. 우도비 검정 p 값을 시뮬레이션하여 선형 예측 변수에 의해 결정된 확률 값의 범위에서 차이를 검출하는 검정력을 구합니다. 자세한 내용은 "검정력 분석"(325 페이지)에서 확인하십시오.

예제용 계획

로짓을 연결 함수로 사용한 일반화 선형 모형을 사용하여 성공 확률을 모델링합니다. 로짓 연결 함수는 로지스틱 모형을 적합시킵니다.

$$\pi(\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_6 X_6)}}$$

여기서 $\pi(\mathbf{X})$ 는 주어진 설계 설정 $\mathbf{X} = (X_1, X_2, \dots, X_6)$ 에서 부품이 검사에 합격할 확률을 나타냅니다.

선형 예측 변수는 $L(\mathbf{X})$ 로 나타냅니다.

$$L(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \dots + \beta_6 X_6$$

다음으로는 선형 예측 변수의 다음 계수 값에 대한 검정력을 구합니다.

계수	값
β_0	0
β_1	1
β_2	0.9
β_3	0.8
β_4	0.7
β_5	0.6
β_6	0.5

선형 예측 변수의 절편이 0 이므로 모든 요인이 0 으로 설정될 경우 부품의 합격 확률은 50% 가 됩니다. 다른 모든 요인이 0 으로 설정될 경우 i^{th} 번째 요인의 수준과 연관된 확률은 아래와 같습니다.

요인	$X_i = 1$ 일 때의 합격률	$X_i = -1$ 일 때의 합격률	차이
X_1	73.11%	26.89%	46.2%
X_2	71.09%	28.91%	42.2%
X_3	69.00%	31.00%	38.0%
X_4	66.82%	33.18%	33.6%
X_5	64.56%	35.43%	29.1%
X_6	62.25%	37.75%	24.5%

예를 들어 X_1 을 제외한 모든 요인이 0 으로 설정될 경우에 검출하려는 합격률은 46.2% 입니다 . 검출하려는 합격률의 최소 차이는 X_6 을 제외한 모든 요인이 0 으로 설정되고 해당 차이가 24.5% 일 때 발생합니다 .

JMP^{PRO} 설계 구성

참고 : 이 섹션의 단계를 건너뛰려면 **도움말 > 샘플 데이터 라이브러리**를 선택하고 Design Experiment/Binomial Experiment.jmp 를 여십시오 . **DOE Simulate** 스크립트 옆의 녹색 삼각형을 클릭한 다음 "**시뮬레이션된 반응 정의**"(324 페이지) 절차로 이동하십시오 .

1. **DOE > 사용자 설계**를 선택합니다 .
2. " 요인 " 개요에서 **N 개 요인 추가** 옆에 6 을 입력합니다 .
3. **요인 추가 > 연속형**을 클릭합니다 .
4. **계속**을 클릭합니다 .

주효과 설계를 구성 중이므로 " 모형 " 개요의 다른 사항은 변경하지 마십시오 .

5. " 런 수 " 아래에서 **사용자 지정** 옆에 60 을 입력합니다 .
6. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 **반응 시뮬레이션**을 선택합니다 .

이렇게 하면 " 테이블 생성 " 을 선택하여 설계 테이블을 구성한 후에 " 반응 시뮬레이션 " 창이 열립니다 .

참고 : 7 단계에서 " 난수 시드값 " 을 설정하고 8 단계에서 " 시작 수 " 를 설정하면 이 예에 표시된 것과 동일한 설계가 재현됩니다 . 설계를 직접 구성할 때는 이러한 단계가 필요하지 않습니다 .

7. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **난수 시드값 설정**을 선택합니다 . "12345" 를 입력하고 **확인**을 클릭합니다 .
8. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 **시작 수**를 선택합니다 . 1 을 입력하고 **확인**을 클릭합니다 .

9. 설계 생성을 클릭합니다 .
10. 테이블 생성을 클릭합니다 .

참고 : Y 및 Y 시뮬레이션 열의 항목은 그림 10.17 에 표시된 것과 다를 수 있습니다 .

그림 10.17 설계 테이블 (일부분)

사용자 설계			X1	X2	X3	X4	X5	X6	Y	Y 시뮬레이션
설계 기준	사용자 설계 D 최적	1	-1	1	-1	1	1	-1	-0.21014392	1.26423349
모형		2	1	1	1	1	-1	1	5.8763992107	6.07472686
설계 평가		3	-1	-1	-1	-1	-1	-1	-3.199937085	-4.1820759
일반화 회귀		4	-1	-1	1	-1	1	-1	-0.641850224	-1.5527745
DOE 시뮬레이션		5	1	-1	-1	-1	-1	1	-1.491025808	0.54014488
DOE 대화상자		6	1	-1	-1	1	-1	-1	-2.205592155	-2.233822
		7	-1	1	1	1	1	-1	4.04410241	2.85846462
		8	-1	1	-1	-1	1	-1	-2.429659436	0.04200364
		9	-1	1	-1	-1	-1	-1	-4.149024207	-2.9342678
		10	1	1	-1	1	1	1	5.0415703877	6.22525903
		11	-1	-1	1	1	-1	-1	-2.409467688	-1.1483041
		12	-1	1	-1	1	-1	1	1.0090675275	1.29012328
		13	1	1	-1	1	-1	-1	0.7878449979	-0.1493944
		14	-1	-1	-1	-1	1	1	-1.079523708	-1.4825482
		15	1	1	-1	1	1	-1	2.6407125697	2.5470259
		16	1	-1	-1	-1	-1	-1	-4.348634305	-2.7261325
		17	-1	1	1	-1	1	-1	1.7254515483	0.77586708
		18	-1	1	1	1	-1	-1	2.3122582433	1.21855301
		19	-1	-1	-1	1	1	-1	-1.147796108	-1.4200152
		20	-1	-1	-1	-1	1	1	-1.707369788	-0.753644
		21	1	-1	-1	1	1	1	2.6562732086	3.90794303

그림 10.18 반응 시뮬레이션 창

반응 시뮬레이션

효과 Y

잘편

1

X1

1

X2

1

X3

1

X4

1

X5

1

X6

1

계수 재설정

분포

정규

오차 σ: 1

이항

Poisson

적용

설계 테이블과 " 반응 시뮬레이션 " 창이 나타납니다 . 설계 테이블에 두 개의 열이 추가되었습니다 .

- Y 에는 " 반응 시뮬레이션 " 창에서 지정한 대로 시뮬레이션된 값 집합이 포함되어 있습니다 .

- Y 시뮬레이션에는 " 반응 시뮬레이션 " 창에서 지정한 모형 계산식을 사용하여 값을 계산하는 계산식이 포함되어 있습니다. 계산식을 보려면 " 열 " 패널에서 열 이름 오른쪽의 더하기 기호를 클릭합니다.

다음 섹션으로 진행하여 이항 반응을 시뮬레이션하고 시뮬레이션된 반응에 일반화 선형 모형을 적합시키십시오.

JMP^{PRO} 시뮬레이션된 반응 정의

성공 확률이 로지스틱 모형으로 주어지는 이항 반응 데이터를 시뮬레이션할 계획입니다. 반응 시뮬레이션에 대한 자세한 내용은 실험 설계 가이드의 " 사용자 설계 " 장에서 확인하십시오.

참고 : 이 섹션의 단계를 건너뛰려면 **Simulate Model Responses** 스크립트 옆의 녹색 삼각형을 클릭하십시오. 그런 다음 "**일반화 선형 모형 적합**"(325 페이지) 절차로 이동하십시오.

1. " 반응 시뮬레이션 " 창 (그림 10.18) 에서 Y 아래에 다음 값을 입력합니다.
 - " 절편 " 옆에 0 을 입력합니다.
 - X1 옆에는 기본적으로 1 이 입력되어 있습니다. 이 값을 유지합니다.
 - X2 옆에 0.9 를 입력합니다.
 - X3 옆에 0.8 을 입력합니다.
 - X4 옆에 0.7 을 입력합니다.
 - X5 옆에 0.6 을 입력합니다.
 - X6 옆에 0.5 를 입력합니다.
2. " 분포 " 개요에서 **이항**을 선택합니다.

N 값은 1 로 설정된 상태로 둡니다. 이 값은 시도당 1 단위만 있음을 나타냅니다.

그림 10.19 완료된 반응 시뮬레이션 창

효과	Y
절편	0
X1	1
X2	0.9
X3	0.8
X4	0.7
X5	0.6
X6	0.5

계수 재설정

분포

☐ 정규

☒ 이항 N

☐ Poisson

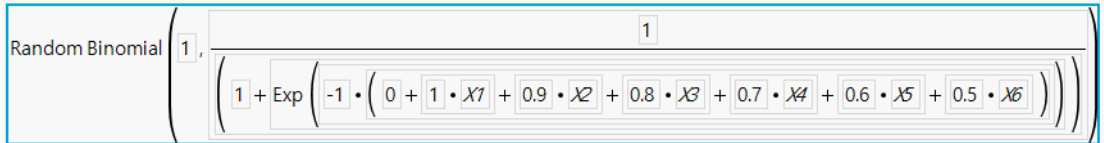
적용

3. **적용**을 클릭합니다.

설계 데이터 테이블에서 Y 시뮬레이션 열이 이항 값을 생성하는 계산식 열로 대체됩니다 .
Y 시도 수라는 열은 각 런의 시도 횟수를 나타냅니다.

4. (선택 사항)" 열 " 패널에서 Y 시뮬레이션 오른쪽의 더하기 기호를 클릭합니다.

그림 10.20 Y 시뮬레이션에 대한 랜덤 이항 계산식



5. **취소**를 클릭합니다.

JMP PRO 일반화 선형 모형 적합

1. 데이터 테이블에서 **모형** 스크립트 옆의 녹색 삼각형을 클릭합니다.

2. **Y** 버튼 옆의 Y 변수를 클릭하고 **제거**를 클릭합니다.

3. Y 시뮬레이션을 클릭하고 **Y** 버튼을 클릭합니다.

이렇게 하면 Y가 무작위로 생성된 이항 값을 포함하는 열로 대체됩니다.

4. " 분석법 " 목록에서 **일반화 선형 모형**을 선택합니다.

5. " 분포 " 목록에서 **이항**을 선택합니다.

" 연결 함수 " 메뉴에 " 로짓 " 함수가 나타납니다.

6. **실행**을 클릭합니다.

적합된 모형은 시뮬레이션된 이항 반응의 단일 집합을 기반으로 합니다.

JMP PRO 검정력 분석

다음으로는 검정의 검정력을 분석하여 " **예제용 계획** "(321 페이지) 에서 구한 계수 값과 선형 예측 변수로 확인된 확률 값 범위의 차이를 검출합니다.

1. " 효과 검정 " 개요에서 **Prob>ChiSq** 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다.

그림 10.21 시뮬레이션 창

"스위치 아웃 열" 목록에서 Y 시뮬레이션 열이 선택되어 있는지 확인하십시오. 이 열에는 모형 적합에 사용된 값이 포함되어 있습니다. 각 시뮬레이션에 대해 "스위치 인할 열"에서 Y 시뮬레이션 열을 선택하면 Y 시뮬레이션의 값이 Y 시뮬레이션 열의 계산식을 사용하여 시뮬레이션된 값을 포함하는 새 열로 대체됩니다.

보고서에서 선택한 **Prob>ChiSq** 열은 연관된 주효과가 0인지 여부에 대한 우도비 검정의 p 값입니다. "효과 검정" 테이블에 나열된 각 효과에 대해 "Prob>ChiSq" 값이 시뮬레이션됩니다.

2. 표본 수 옆에 500 을 입력합니다.
3. 확인을 클릭합니다.

"일반화 선형 모형 시뮬레이션 결과" 데이터 테이블이 나타납니다.

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션된 p 값은 그림 10.22 에 표시된 것과 다를 수 있습니다.

그림 10.22 시뮬레이션 결과 표 (일부분)

일반화 선형 모형 시뮬레이션...										
난수 시드값		123		SimID	X1	X2	X3	X4	X5	X6
결과 데이터 테이블 생성			1	0	0.0103	0.0001	0.0582	0.2303	0.0646	0.0105
분포			2	1	0.6296	0.0071	0.0008	0.0013	0.5624	0.0430
검정력 분석			3	2	0.0013	0.0001	0.0122	0.0185	0.0468	0.3185
			4	3	0.0003	0.0003	0.1105	0.0543	0.0778	0.1251
열 (7/0)			5	4	<.0001	0.0013	0.0073	0.1529	0.0763	0.0098
SimID			6	5	0.0288	0.1204	0.1273	0.0311	0.1197	0.8565
X1			7	6	0.0309	0.0063	0.0069	0.1449	0.1259	0.7667
X2			8	7	0.0016	0.4502	0.0002	0.0027	0.3643	0.3306
X3			9	8	0.0047	0.0308	0.0216	0.0110	0.1086	0.0331
X4			10	9	0.0041	0.0266	0.0053	0.0047	0.0262	0.9868
X5			11	10	0.0113	0.0002	0.0012	0.0220	0.6553	0.0609
X6			12	11	0.0829	<.0001	0.1211	0.0269	0.0012	0.1277
			13	12	0.0022	0.0017	0.4689	0.0179	0.0017	0.9394
			14	13	0.0002	0.0003	0.0133	0.0729	0.8009	0.0158
모든 행		501	15	14	<.0001	0.0102	0.0155	0.0002	0.3178	0.0343
선택됨		0	16	15	0.0026	0.0794	0.0136	0.0035	0.6162	0.3227
제외		1	17	16	0.0014	<.0001	0.0109	0.0191	0.0088	0.0016
숨김		0	18	17	0.0009	0.0008	0.6379	0.0091	0.1335	0.7542
라벨 지정		0	19	18	0.0219	0.0043	0.0190	0.0312	0.6488	0.3548

테이블의 첫 번째 행은 **Prob>ChiSq** 의 초기 값을 포함하므로 제외됩니다. 나머지 500 개 행에는 시뮬레이션된 값이 포함되어 있습니다.

4. **검정력 분석** 스크립트를 실행합니다.

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 그림 10.23 에 표시된 것과 다를 수 있습니다.

그림 10.23 처음 세 개의 효과에 대한 분포 그림



히스토그램에는 각 주효과에 대해 시뮬레이션된 500 개의 **Prob>ChiSq** 값이 표시됩니다. "시뮬레이션 검정력" 개요에는 500 회의 시뮬레이션에서 시뮬레이션된 기각률이 표시됩니다. 좀 더 보기 쉽게 다음과 같이 보고서를 세로로 쌓고 그림을 선택 취소할 수 있습니다.

- "분포"의 빨간색 삼각형을 클릭하고 **쌓기**를 선택합니다.
- Ctrl 키를 누른 채 X1의 빨간색 삼각형을 클릭하고 **이상치 상자 그림**을 선택 취소합니다.
- Ctrl 키를 누른 채 X1의 빨간색 삼각형을 클릭하고 **히스토그램 옵션**을 선택한 다음 **히스토그램**을 선택 취소합니다.

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 그림 10.24 에 표시된 것과 다를 수 있습니다.

그림 10.24 처음 세 개의 효과에 대한 검정력 결과

분포										
X1										
시뮬레이션 결과						시뮬레이션 검정력				
$Y_0 = 0.01029$ (원래 추정값)						알파	기각 수	기각률	95% 하한	95% 상한
신뢰 구간			경험적 p 값			0.01	362	0.724	0.68322	0.76136
알파	CI 하한	CI 상한	검정	p 값		0.05	434	0.868	0.83551	0.89488
0.05	1.51e-7	0.36019	$Y \geq Y_0 $	0.2740		0.10	460	0.92	0.89289	0.9407
0.10	7.23e-7	0.19148	$Y \leq Y_0$	0.7260		0.20	477	0.954	0.93192	0.96915
0.20	6.57e-6	0.08284	$Y \geq Y_0$	0.2740						
0.50	0.00013	0.01229								
X2										
시뮬레이션 결과						시뮬레이션 검정력				
$Y_0 = 0.00014$ (원래 추정값)						알파	기각 수	기각률	95% 하한	95% 상한
신뢰 구간			경험적 p 값			0.01	316	0.632	0.58887	0.67312
알파	CI 하한	CI 상한	검정	p 값		0.05	407	0.814	0.77755	0.84567
0.05	1.08e-6	0.44811	$Y \geq Y_0 $	0.8280		0.10	440	0.88	0.84858	0.90563
0.10	4.39e-6	0.24295	$Y \leq Y_0$	0.1720		0.20	467	0.934	0.90876	0.95262
0.20	2.82e-5	0.12448	$Y \geq Y_0$	0.8280						
0.50	0.00031	0.02695								
X3										
시뮬레이션 결과						시뮬레이션 검정력				
$Y_0 = 0.05823$ (원래 추정값)						알파	기각 수	기각률	95% 하한	95% 상한
신뢰 구간			경험적 p 값			0.01	209	0.418	0.37555	0.4617
알파	CI 하한	CI 상한	검정	p 값		0.05	320	0.64	0.59701	0.68086
0.05	0.00001	0.67078	$Y \geq Y_0 $	0.3400		0.10	364	0.728	0.68737	0.76516
0.10	0.00004	0.53067	$Y \leq Y_0$	0.6600		0.20	417	0.834	0.79886	0.86404
0.20	0.0002	0.36561	$Y \geq Y_0$	0.3400						
0.50	0.00203	0.11806								

"시뮬레이션 검정력" 개요에서 각 행의 "기각률"은 해당 알파 값보다 작은 p 값의 비율을 나타냅니다. 예를 들어 계수 값 0.8, 확률 차이 38%에 해당하는 X3의 경우 유의 수준 0.05에 대한 시뮬레이션된 검정력은 $379/500 = 0.758$ 입니다. 표 10.1에는 유의 수준 0.05에서 모든 효과에 대해 추정된 검정력이 요약되어 있습니다. 이를 통해 "감지할 차이"가 감소할 때 검정력이 얼마나 감소하는지 알 수 있습니다. 또한 24.5%의 효과를 감지할 수 있는 검정력(X6)은 약 0.37에 불과함을 알 수 있습니다.

참고: 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 표 10.1에 표시된 것과 다를 수 있습니다.

표 10.1 유의 수준 0.05에서의 시뮬레이션 검정력

요인	$X_i = 1$ 일 때의 합 격률	$X_i = -1$ 일 때의 합 격률	감지할 차이	알파 = 0.05 일 때의 시뮬레이 션 검정력 (기각률)
X_1	73.11%	26.89%	46.2%	0.852
X_2	71.09%	28.91%	42.2%	0.828
X_3	69.00%	31.00%	38.0%	0.758

표 10.1 유의 수준 0.05 에서의 시뮬레이션 검정력 (계속)

요인	$X_i = 1$ 일 때의 합 격률	$X_i = -1$ 일 때의 합 격률	감지할 차이	알파 = 0.05 일 때의 시뮬레이 션 검정력 (기각률)
X_4	66.82%	33.18%	33.6%	0.654
X_5	64.56%	35.43%	29.1%	0.488
X_6	62.25%	37.75%	24.5%	0.372

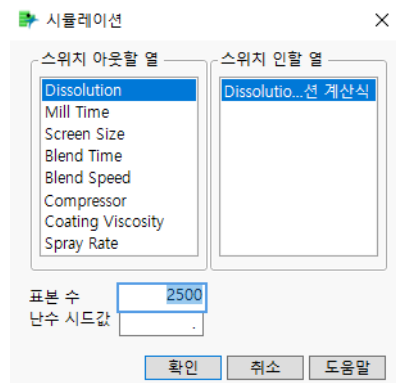
JMP PRO 시뮬레이션 창 시작

시뮬레이션 창을 시작하려면 보고서 창에서 계산된 값이 포함된 열을 마우스 오른쪽 버튼으로 클릭하고 " 시뮬레이션 " 을 선택합니다 . 시뮬레이션 기능은 붓스트랩을 지원하는 모든 보고서를 포함하여 다양한 보고서에서 사용할 수 있습니다 . 시뮬레이션 기능을 사용하려면 데이터를 시뮬레이션하는 랜덤 성분이 있는 계산식이 데이터 테이블에 포함되어 있어야 합니다 .

참고 : 기준 변수를 사용하는 보고서에서는 시뮬레이션 옵션을 사용할 수 없습니다 .

JMP PRO 시뮬레이션 창

그림 10.25 Tablet Production.jmp 에 대한 시뮬레이션 창



" 시뮬레이션 " 창에는 다음과 같은 패널 및 옵션이 포함되어 있습니다 .

스위치 아웃할 열 " 스위치 인할 열 " 로 대체되는 열입니다 .

스위치 인할 열 "스위치 아웃할 열"을 대체하는 열입니다. "스위치 인할 열"의 계산식에 따라 시뮬레이션된 값을 사용하여 분석이 반복됩니다. "스위치 인할 열" 패널에는 계산식이 있는 열만 나열됩니다.

표본 수 시뮬레이션된 데이터 집합에 대해 보고서를 다시 실행할 횟수입니다. 기본값은 2500입니다.

난수 시드값 시뮬레이션된 결과를 제어하는 값입니다. 난수 시드값은 결과를 재현 가능하게 합니다.

"시뮬레이션" 창에서 "확인"을 클릭하면 진행률 표시줄과 "조기 중지" 버튼이 표시된 창이 나타납니다. 진행률 표시줄 위에는 시뮬레이션 중인 표본 수가 표시됩니다. "조기 중지"를 클릭하면 해당 시점까지 계산된 시뮬레이션 값이 "시뮬레이션 결과" 테이블에 표시됩니다. 이 창에는 특정 시점에 실행 중인 분석도 표시됩니다.



시뮬레이션 결과 테이블

시뮬레이션 결과는 표로 나타납니다. 다음 사항에 유의하십시오.

- 테이블의 첫 번째 행에는 보고서에 표시된 테이블 항목에 대한 값이 포함됩니다. 따라서 첫 번째 행은 항상 제외됩니다.
- 나머지 행에서는 시뮬레이션 결과를 제공합니다. 나머지 행의 개수는 시뮬레이션 시작 창에서 지정한 표본 수와 같습니다.
- 보고서의 행은 선택된 계산 값 열이 포함된 보고서 테이블의 첫 번째 열로 식별됩니다. 이 첫 번째 열에 있는 각 항목마다 하나씩의 열이 시뮬레이션 결과 테이블에 표시됩니다.
- 이 테이블에는 "분포" 보고서를 생성하는 **분포** 스크립트가 포함되어 있습니다. 이 보고서에는 시뮬레이션 결과 데이터 테이블의 각 열에 대한 히스토그램, 분위수, 요약 통계량 및 시뮬레이션 결과가 포함됩니다. 표준 "분포" 보고서뿐만 아니라 다음 항목도 이 보고서에 포함됩니다.
 - 히스토그램에 원래 추정값을 나타내는 빨간색 선이 표시됩니다.
 - 원래 추정값과 함께 시뮬레이션에 대한 신뢰 구간 및 경험적 p 값이 포함된 "시뮬레이션 결과" 보고서. 자세한 내용은 "**시뮬레이션 결과 보고서**"(331 페이지)에서 확인하십시오.
 - 시뮬레이션 결과 데이터 테이블의 값이 p 값 형식인 경우에는 "시뮬레이션 검정력" 보고서도 제공됩니다. 자세한 내용은 "**시뮬레이션 검정력 보고서**"(331 페이지)에서 확인하십시오.
- p 값 열을 시뮬레이션한 경우에만 **검정력 분석** 스크립트가 테이블에 포함됩니다. 이 스크립트는 p 값의 히스토그램을 보여 주는 "분포" 보고서를 생성하고 "시뮬레이션 검정력" 보고서를 제공합니다. 자세한 내용은 "**시뮬레이션 검정력 보고서**"(331 페이지)에서 확인하십시오.

시뮬레이션 결과 보고서

원래 추정값 원래 추정값의 값으로, 시뮬레이션 결과 데이터 테이블의 첫 번째 행에도 표시됩니다. 이 추정값에는 Y_0 이라는 라벨이 지정됩니다.

신뢰 구간 유의 수준이 0.05, 0.10, 0.20 및 0.50 일 때 분위수 기반 신뢰 구간의 하한 및 상한입니다.

경험적 p 값 시뮬레이션된 값을 원래 추정값과 비교하는 하나의 양측 검정과 두 개의 단측 검정에 대한 경험적 p 값입니다. 이러한 p 값은 보고서의 "검정" 열에 지정된 범위에 속하는 시뮬레이션된 값의 비율로 계산됩니다.

시뮬레이션 검정력 보고서

알파 유의 수준으로, 0.01, 0.05, 0.10 및 0.20 입니다.

기각 수 해당 유의 수준에서 검정이 기각되는 시뮬레이션의 수입니다.

기각률 해당 유의 수준에서 검정이 기각되는 시뮬레이션의 비율입니다.

95% 하한 및 95% 상한 시뮬레이션된 기각률에 대한 95% 신뢰 구간의 하한 및 상한입니다. 신뢰 구간은 Wilson 점수 방법을 사용하여 계산됩니다. 자세한 내용은 Wilson 연구 자료([1927](#))에서 확인하십시오.

팁: 신뢰 구간을 좁히려면 표본 수를 늘리십시오.

11 장

JMP Pro 붓스트랩

재표집을 통한 통계량 분포 근사

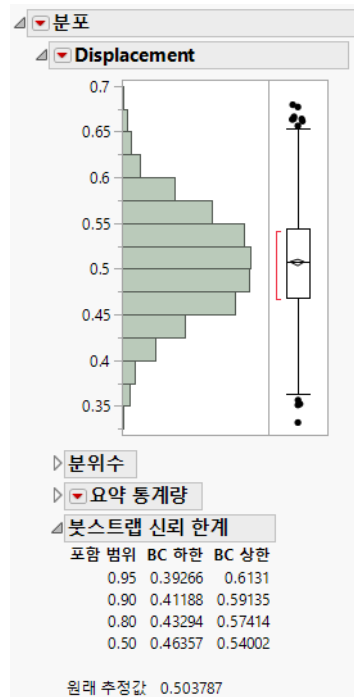
붓스트랩은 JMP Pro 에서만 사용할 수 있습니다.

붓스트랩은 통계량의 표본 분포에 근사한 값을 산출하기 위한 재표집 방법입니다. 붓스트랩을 사용하여 평균, 편향, 표준 오차 및 신뢰 구간 같은 통계량의 분포와 해당 특성을 추정할 수 있습니다. 붓스트랩은 다음과 같은 경우에 특히 유용합니다.

- 통계량의 이론상 분포가 복잡하거나 알 수 없는 경우
- 가정 위배로 인해 모수적 방법을 사용한 추론이 가능하지 않은 경우

참고: 붓스트랩은 보고서에서 마우스 오른쪽 버튼을 클릭하는 방법으로만 사용할 수 있습니다. 플랫폼 명령으로는 제공되지 않습니다.

그림 11.1 기울기 모수에 대한 붓스트랩 결과



목차

붓스트랩 개요.....	335
붓스트랩의 예.....	336
붓스트랩 창 옵션.....	338
누적 결과 테이블.....	340
비누적 붓스트랩 결과 테이블.....	341
붓스트랩 결과 분석.....	342
붓스트랩의 추가 예.....	343
붓스트랩에 대한 통계 상세 정보.....	348
부분 가중치 계산.....	348
편향 수정 백분위수 구간.....	348
붓스트랩의 가용성.....	349

JMP[®] PRO 붓스트랩 개요

붓스트랩은 보고서에 사용되는 관측값을 반복적으로 재표집하여 통계량의 분포를 추정하는 기법입니다. 이때 관측값은 독립적인 것으로 가정합니다.

단순 붓스트랩에서는 복원 표집 방법으로 n 개의 관측값을 재표집하여 크기가 n 인 붓스트랩 표본을 생성합니다. 일부 관측값은 붓스트랩 표본에 나타나지 않을 수 있으며, 다른 관측값은 여러 번 나타날 수 있습니다. 하나의 관측값이 붓스트랩 표본에 나타나는 횟수를 **붓스트랩 가중치**라고 합니다. 붓스트랩 반복 시마다 관심 있는 통계량을 산출한 전체 분석이 다음과 같이 변경되어 다시 실행됩니다.

- n 개의 관측값을 포함하는 붓스트랩 표본이 데이터 집합이 됩니다.
- 붓스트랩 가중치는 분석 플랫폼의 빈도 변수입니다.

관심 있는 통계량의 값 분포가 생성될 때까지 이 과정이 반복됩니다.

하지만 일부 경우에는 단순 붓스트랩이 부적절할 수 있습니다. 예를 들어 데이터 집합이 작거나 로지스틱 회귀를 사용하기 때문에 분리 문제가 발생할 수 있는 경우가 있습니다. 이러한 경우 JMP에서는 부분 가중치를 사용하여 베이지안 붓스트랩을 수행할 수 있습니다. 부분 가중치가 사용되는 경우에는 각 관측값에 부분 가중치가 연관됩니다. 부분 가중치의 합은 n 입니다. 관심 있는 통계량을 계산하려면 분석 플랫폼에서 부분 가중치를 빈도 변수로 처리합니다. 부분 가중치에 대한 자세한 내용은 "**부분 가중치**"(339 페이지) 및 "**부분 가중치 계산**"(348 페이지)에서 확인하십시오.

보고서에서 붓스트랩 분석을 실행하려면 붓스트랩을 수행할 통계량이 포함된 테이블 열을 마우스 오른쪽 버튼으로 클릭하고 "붓스트랩"을 선택합니다.

참고: 붓스트랩은 보고서에서 마우스 오른쪽 버튼을 클릭하는 방법으로만 사용할 수 있습니다. 플랫폼 명령으로는 제공되지 않습니다.

JMP는 많은 통계 플랫폼에서 붓스트랩 기능을 제공합니다. 전체 목록은 "**붓스트랩의 가용성** 349"(334 페이지)에서 확인하십시오. 표본을 구성하는 관측값은 관심 있는 통계량을 계산하는데 사용된 모든 관측값입니다. 보고서에 빈도 열이 사용된 경우 해당 열의 관측값은 빈도 변수로 지정된 횟수만큼 반복된 것처럼 처리됩니다. 보고서에 가중치 변수가 사용된 경우 붓스트랩을 사용하면 해당 변수는 보고서의 계산에서 처리되던 방식대로 처리됩니다.

팁: 붓스트랩 기능은 붓스트랩이 호출된 플랫폼 보고서에 표시된 전체 분석을 다시 실행합니다. 따라서 보고서에 포함된 관련 없는 분석으로 인해 선택한 열에 대한 붓스트랩이 느리게 실행될 수 있습니다. 붓스트랩이 느리게 실행되는 경우에는 플랫폼 보고서에서 관련 없는 옵션을 제거한 후 붓스트랩을 실행하십시오.

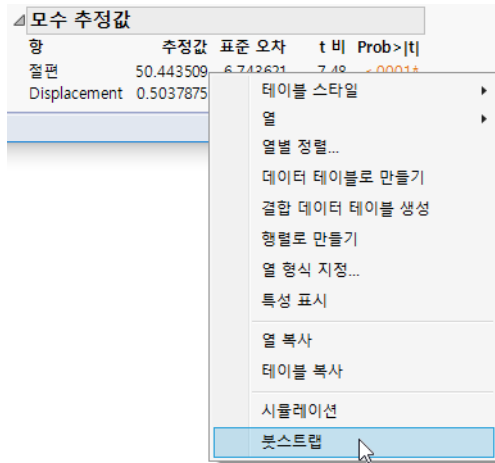
JMP[®] PRO 붓스트랩의 예

이 예에서는 Car Physical Data.jmp 샘플 데이터 테이블을 사용합니다. 한 타이어 제조업체에서 엔진 배기량 (in³) 으로부터 엔진 출력 (마력) 을 예측하려고 합니다. 이 회사는 두 변수 간 관계의 기울기를 추정하는 데 가장 관심이 있습니다. 기울기 값은 배기량이 변화할 때 이에 대응하는 출력의 변화를 예측하는 데 유용합니다.

이 예에서는 분산 동질성에 대한 회귀 가정이 위배되므로 기울기에 대한 회귀 분석의 신뢰 한계가 혼동을 일으킬 수 있습니다. 따라서 이 회사에서는 기울기의 신뢰 구간에 대한 붓스트랩 추정값을 사용합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Car Physical Data.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. Horsepower 를 선택하고 **Y, 반응**을 클릭합니다.
4. Displacement 를 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "Horsepower 대 Displacement 의 이변량 적합" 옆의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.
기울기 추정값은 0.503787, 대략적으로는 0.504 입니다.
7. (선택 사항) **모수 추정값** 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 하한**을 선택합니다.
8. (선택 사항) **모수 추정값** 보고서를 마우스 오른쪽 버튼으로 클릭하고 **열 > 95% 상한**을 선택합니다.
기울기에 대한 회귀 분석으로 구한 신뢰 한계는 0.4249038 과 0.5826711 입니다.
9. **모수 추정값** 보고서의 **추정값** 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다.

그림 11.2 붓스트랩 옵션



마우스 오른쪽 버튼으로 클릭한 열은 **선택한 열 분할** 옵션을 선택할 때도 관련됩니다. 자세한 내용은 "[붓스트랩 창 옵션](#)"(338 페이지)에서 확인하십시오.

10. **붓스트랩 표본 수**에 1000 을 입력합니다 .
11. (선택 사항) 그림 11.3 의 결과와 매칭되는 결과를 얻으려면 **난수 시드값**에 12345 를 입력합니다 .
12. **확인**을 클릭합니다 .
 붓스트랩 프로세스가 실행된 후 붓스트랩 결과 데이터 테이블이 생성됩니다. 이때 기울기 및 절편에 대한 결과는 쌓이지 않은 상태입니다 .
 다음으로 , 붓스트랩 기울기를 분석합니다 .
13. 붓스트랩 결과 테이블에서 **분포** 스크립트를 실행합니다 .
 " 분포 " 보고서에는 " 붓스트랩 신뢰 한계 " 보고서가 포함됩니다 .

그림 11.3 붓스트랩 보고서

분포
Displacement
분위수
100.0% 최대값 0.6800107545
99.5% 0.6483289879
97.5% 0.6212501769
90.0% 0.5781520893
75.0% 사분위수 0.5423786474
50.0% 중앙값 0.5051381996
25.0% 사분위수 0.4692123772
10.0% 0.4374744231
2.5% 0.4035615438
0.5% 0.3686834278
0.0% 최소값 0.3494461047
요약 통계량
붓스트랩 신뢰 한계
포함 범위 BC 하한 BC 상한
0.95 0.40028 0.61892
0.90 0.41646 0.59849
0.80 0.43405 0.57489
0.50 0.46765 0.54007
원래 추정값 0.503787

기울기 (6 단계) 의 추정값은 0.504 입니다. 95% 포함 범위에 대한 붓스트랩 결과에 따르면 이 회사에서는 기울기가 0.40028 에서 0.61892 사이일 것이라고 추정할 수 있습니다. 즉, 배기량이 한 단위 변화할 때 출력이 0.40028 ~ 0.61892 정도로 변화함을 95% 신뢰도로 추정할 수 있습니다. 기울기에 대한 붓스트랩 신뢰 구간 (0.400 ~ 0.619) 은 7 단계 및 8 단계의 일반적인 회귀 가정을 사용하여 얻은 신뢰 구간 (0.425 ~ 0.583) 보다 다소 넓습니다.

참고: " 붓스트랩 신뢰 한계 " 보고서의 "BC 하한" 및 "BC 상한" 열은 편향 수정 구간을 나타냅니다. 자세한 내용은 "편향 수정 백분위수 구간"(348 페이지) 에서 확인하십시오.

JMP PRO 붓스트랩 창 옵션

붓스트랩 분석을 수행하려면 보고서 창의 테이블에서 숫자로 된 표본 통계량 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다. 그러면 선택한 열이 강조 표시되고 "붓스트랩" 창이 나타납니다. "붓스트랩" 창에서 옵션을 선택하고 **확인**을 클릭하면 해당 열의 모든 통계량에 대한 붓스트랩 결과가 기본 결과 테이블에 나타납니다.

참고: 기준 변수를 사용하는 보고서에서는 붓스트랩 옵션을 사용할 수 없습니다.

"붓스트랩" 창에는 다음 옵션이 포함되어 있습니다.

붓스트랩 표본 수 데이터를 재표집하여 통계량을 계산할 횟수를 설정합니다. 값이 클수록 해당 통계량 특성을 더 정확하게 추정할 수 있습니다. 기본적으로 붓스트랩 표본 수는 2500 으로 설정됩니다.

난수 시드값 현재 결과를 복제하기 위해 붓스트랩 분석의 이후 런에서 다시 입력할 수 있는 난수 시드값을 설정합니다. 기본적으로 시드값은 설정되어 있지 않습니다.

부분 가중치 베이지안 붓스트랩 분석을 수행합니다. 붓스트랩 반복 시마다 "부분 가중치 계산"(348 페이지)에 설명된 대로 계산된 가중치가 각 관측값에 할당됩니다. 가중치가 적용된 관측값은 관심 있는 통계량을 계산하는 데 사용됩니다. 기본적으로는 "부분 가중치" 옵션이 선택되어 있지 않으며 단순 붓스트랩 분석이 수행됩니다.

팁: 분석에 사용되는 관측값의 개수가 적거나 로지스틱 회귀 설정 시 분리가 우려되는 경우 에 "부분 가중치" 옵션을 사용하십시오.

"부분 가중치" 옵션이 선택된 경우에는 붓스트랩 반복 시마다 보고서에 사용되는 각 관측값에 0 이 아닌 가중치가 할당됩니다. 이러한 가중치의 합은 n , 즉 관심 있는 통계량의 계산에 사용된 관측값 수가 됩니다. 가중치의 계산 및 사용 방법에 대한 자세한 내용은 "부분 가중치 계산"(348 페이지)에서 확인하십시오.

선택한 열 분할 붓스트랩을 위해 선택한 열의 각 통계량에 대한 붓스트랩 결과를 붓스트랩 결과 테이블에 개별 열로 배치합니다. 붓스트랩 결과 테이블의 각 행 (첫 번째 행 제외) 은 단일 붓스트랩 표본에 해당합니다.

이 옵션을 선택 취소하면 누적 붓스트랩 결과 테이블이 나타납니다. 붓스트랩 반복 시마다 붓스트랩을 위해 선택한 열이 포함된 전체 보고서 테이블에 대한 결과가 이 테이블에 포함됩니다. 보고서 테이블의 각 행에 대한 결과는 누적 붓스트랩 결과 테이블에 행으로 나타납니다. 보고서 테이블의 각 열은 누적 붓스트랩 결과 테이블의 열 하나를 정의합니다. 예는 "누적 결과 테이블"(340 페이지)에서 확인하십시오.

분할 적용 시 쌓인 테이블 삭제 (선택한 열 분할 옵션이 선택된 경우에만 사용 가능) 붓스트랩을 통해 생성되는 결과 테이블의 수를 결정합니다.

"분할 적용 시 쌓인 테이블 삭제" 옵션이 선택되어 있지 않으면 다음과 같은 두 개의 붓스트랩 테이블이 표시됩니다.

- 누적 붓스트랩 결과 테이블 - 붓스트랩을 위해 선택한 열이 포함된 테이블의 각 행에 대한 붓스트랩 결과가 포함됩니다. 이 테이블에서는 보고서의 모든 통계량에 대한 붓스트랩 결과를 제공하며 각 열은 하나의 통계량에 의해 정의됩니다.
- 비누적 붓스트랩 결과 테이블 - 누적 테이블을 분할하여 생성한 테이블입니다. 이 테이블에서는 원래 보고서에서 선택한 열에 대한 결과만 제공합니다.

"분할 적용 시 쌓인 테이블 삭제" 옵션이 선택되어 있고 선택한 열 분할 작업에 성공한 경우에는 누적 붓스트랩 결과 테이블이 표시되지 않습니다.

JMP PRO 누적 결과 테이블

붓스트랩 분석의 초기 결과는 누적 결과 테이블로 나타납니다 (그림 11.4). **분할 적용 시 쌓인 테이블 삭제** 옵션을 선택한 경우에는 이 테이블이 나타나지 않습니다. 그림 11.4에서는 Car Physical Data.jmp에 "X로 Y 적합"의 이변량 모형을 적합시켜서 생성된 "모수 추정값" 보고서를 기반으로 한 붓스트랩 테이블을 보여 줍니다. 자세한 내용은 "**붓스트랩 개요**"(335페이지)에서 확인하십시오.

그림 11.4 누적 붓스트랩 결과 테이블

		X	Y	항	~편향	추정값	표준 오차	t 비	Prob> t	BootID*
X	1	Displacement	Horsepower	절편		50.443509471	6.7436209999	7.48	<.0001	0
X	2	Displacement	Horsepower	Displacement		0.5037874592	0.0398202611	12.65	<.0001	0
	3	Displacement	Horsepower	절편		45.385604758	5.4724273046	8.29	<.0001	1
	4	Displacement	Horsepower	Displacement		0.5451674092	0.0306151772	17.81	<.0001	1
	5	Displacement	Horsepower	절편		40.843862813	7.0508187326	5.79	<.0001	2
	6	Displacement	Horsepower	Displacement		0.5854988173	0.0457041828	12.81	<.0001	2
	7	Displacement	Horsepower	절편		47.765642104	5.4677610087	8.74	<.0001	3
	8	Displacement	Horsepower	Displacement		0.4943459579	0.0305765677	16.17	<.0001	3
	9	Displacement	Horsepower	절편		59.758060908	7.0436720036	8.48	<.0001	4
	10	Displacement	Horsepower	Displacement		0.4385540785	0.042053201	10.43	<.0001	4
	11	Displacement	Horsepower	절편		66.341062413	6.4663383596	10.26	<.0001	5
	12	Displacement	Horsepower	Displacement		0.3969640071	0.0393100058	10.10	<.0001	5
	13	Displacement	Horsepower	절편		41.234989734	5.8543901826	7.04	<.0001	6
	14	Displacement	Horsepower	Displacement		0.5723548142	0.034172131	16.75	<.0001	6
	15	Displacement	Horsepower	절편		43.876867815	7.8085946052	5.62	<.0001	7
	16	Displacement	Horsepower	Displacement		0.5731377467	0.0439823115	13.03	<.0001	7

누적 결과 테이블의 특징은 다음과 같습니다.

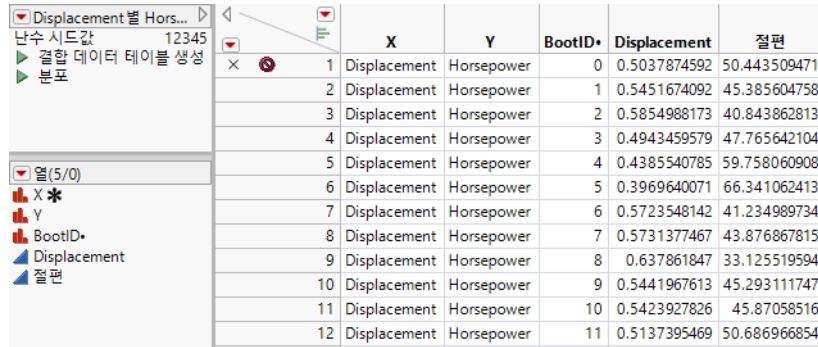
- 각 붓스트랩 표본마다 보고서 테이블의 첫 번째 열에 주어진 각 값에 대한 행이 하나씩 있습니다. 이러한 값은 보고서 테이블의 첫 번째 열 이름을 사용하는 열에 표시됩니다. 이 예에서는 각 붓스트랩 표본마다 항 열에 나타나는 각 항, 즉 "절편" 및 "Displacement"에 대한 결과를 포함하는 행이 하나씩 있습니다.
- 분석에 사용된 데이터 테이블 열은 이 테이블에 나타납니다. 이 예에서는 X가 Displacement이고 Y가 Horsepower입니다.
- 붓스트랩을 수행하는 보고서 테이블의 각 열마다 하나씩의 열이 있습니다. 이 예에서는 ~편향, 추정값, 표준 오차, t 비 및 Prob>|t| 열이 그러한 열에 해당합니다. ~편향은 모수 추정값 중 하나가 편향되지 않은 한 "X로 Y 적합" 보고서에서는 숨겨지는 열입니다.
- BootID 열은 붓스트랩 표본을 식별합니다. BootID=0인 행은 원래 추정값에 해당합니다. 이러한 행은 X로 표시되며 '제외됨' 행 상태를 갖습니다. 이 예에서는 각 붓스트랩 표본이 두 개의 행에 대한 결과, 즉 절편에 대한 결과와 Displacement에 대한 결과를 계산하는 데 사용되었습니다.
- 데이터 테이블 이름에는 "누적 붓스트랩 결과"가 포함됩니다.

선택한 열 분할 옵션을 선택한 경우에는 비누적 결과 테이블도 나타날 수 있습니다. 자세한 내용은 "**비누적 붓스트랩 결과 테이블**"(341 페이지)에서 확인하십시오.

JMP PRO 비누적 붓스트랩 결과 테이블

선택한 보고서 열에 대한 개별 열이 포함된 붓스트랩 테이블을 생성하려면 **선택한 열 분할**을 선택합니다. 각 열은 보고서 테이블의 행 하나에 해당합니다. 예를 들어 그림 11.5에서는 그림 11.4의 추정값 열이 항의 각 수준에 해당하는 두 개의 열 (Displacement 및 절편)로 분할됩니다.

그림 11.5 비누적 붓스트랩 결과 테이블



	X	Y	BootID	Displacement	절편
1	Displacement	Horsepower	0	0.5037874592	50.443509471
2	Displacement	Horsepower	1	0.5451674092	45.385604758
3	Displacement	Horsepower	2	0.5854988173	40.843862813
4	Displacement	Horsepower	3	0.4943459579	47.765642104
5	Displacement	Horsepower	4	0.4385540785	59.758060908
6	Displacement	Horsepower	5	0.3969640071	66.341062413
7	Displacement	Horsepower	6	0.5723548142	41.234989734
8	Displacement	Horsepower	7	0.5731377467	43.876867815
9	Displacement	Horsepower	8	0.637861847	33.125519594
10	Displacement	Horsepower	9	0.5441967613	45.293111747
11	Displacement	Horsepower	10	0.5423927826	45.87058516
12	Displacement	Horsepower	11	0.5137395469	50.686966854

비누적 결과 테이블의 특징은 다음과 같습니다.

- 각 붓스트랩 표본마다 하나의 행이 있습니다.
- 분석에 사용된 데이터 테이블 열은 이 테이블에 나타납니다. 이 예에서는 X가 Displacement이고 Y가 Horsepower입니다.
- 붓스트랩이 수행된 보고서의 각 행마다 하나의 열이 있습니다.
- "붓스트랩" 창에서 "난수 시드값"을 지정한 경우에는 해당 값을 제공하는 "난수 시드값"이라는 테이블 변수가 붓스트랩 결과 테이블에 포함됩니다.
- 비누적 붓스트랩 결과 테이블에는 "소스" 테이블 스크립트와 "분포" 테이블 스크립트가 포함됩니다. "분포" 테이블 스크립트를 사용하면 붓스트랩 표본을 기반으로 한 붓스트랩 신뢰 구간 등의 통계량을 빠르게 구할 수 있습니다.
- BootID 열은 붓스트랩 표본을 식별합니다. BootID=0인 행은 원래 추정값에 해당합니다. 이 행은 X로 표시되며 '제외됨' 행 상태를 갖습니다. 비누적 붓스트랩 테이블에서는 각 행이 단일 붓스트랩 표본으로부터 계산됩니다.
- 이 데이터 테이블의 이름은 "붓스트랩 결과 (<열 이름>)"로 끝납니다. 여기서 <열 이름>은 붓스트랩을 수행한 보고서의 열을 식별합니다.

JMP PRO 붓스트랩 결과 분석

분포 플랫폼을 사용하여 붓스트랩 결과를 분석할 수 있습니다.

- 분석을 통해 비누적 붓스트랩 결과 테이블이 생성된 경우에는 테이블에서 "분포" 스크립트를 실행합니다.
- 분석을 통해 누적 붓스트랩 결과 테이블이 생성된 경우에는 **분석 > 분포**를 선택하고 관심 있는 열을 적절한 역할에 할당합니다. 대부분의 경우 보고서 테이블의 첫 번째 열에 해당하는 열을 기준 역할에 할당하는 것이 좋습니다.

분포 플랫폼에서는 붓스트랩 결과에 대한 요약 통계량을 제공합니다. 또한 BootID 열이 포함된 테이블 (그림 11.6)에 대한 "붓스트랩 신뢰 한계" 보고서도 생성됩니다.

"분포" 보고서를 사용하여 다음과 같은 두 가지 유형의 붓스트랩 신뢰 구간을 구할 수 있습니다.

- "분위수" 보고서에서는 백분위수 구간을 제공합니다. 예를 들어 백분위수 방법을 사용하여 95% 신뢰 구간을 구성하려면 2.5% 및 97.5% 분위수를 구간 경계로 사용합니다.
- "붓스트랩 신뢰 한계" 보고서에서는 편향 수정 백분위수 구간을 제공합니다. 이 보고서에는 포함 수준이 95%, 90%, 80% 및 50% 인 구간이 표시됩니다. "BC 하한" 및 "BC 상한" 열에는 각각 하한점과 상한점이 표시됩니다. 편향 수정 백분위수 구간의 계산에 대한 자세한 내용은 "**편향 수정 백분위수 구간**" (348 페이지)에서 확인하십시오.

그림 11.6 붓스트랩 신뢰 한계 보고서

▼ 분포		
▼ Displacement		
▼ 분위수		
100.0%	최대값	0.6800107545
99.5%		0.6483289879
97.5%		0.6212501769
90.0%		0.5781520893
75.0%	사분위수	0.5423786474
50.0%	중앙값	0.5051381996
25.0%	사분위수	0.4692123772
10.0%		0.4374744231
2.5%		0.4035615438
0.5%		0.3686834278
0.0%	최소값	0.3494461047
▶ 요약 통계량		
▼ 붓스트랩 신뢰 한계		
포함 범위	BC 하한	BC 상한
0.95	0.40028	0.61892
0.90	0.41646	0.59849
0.80	0.43405	0.57489
0.50	0.46765	0.54007
원래 추정값 0.503787		

"붓스트랩 신뢰 한계" 보고서의 맨 아래에 있는 **원래 추정값**은 원래 데이터를 사용한 통계량의 추정값입니다.

"붓스트랩 신뢰 한계" 보고서의 해석에 대한 자세한 내용은 "[붓스트랩 개요](#)"(335 페이지)에서 확인하십시오. Efron 연구 자료 ([1981](#))에는 백분위수 구간과 편향 수정 백분위수 구간 모두의 방법이 설명되어 있습니다.

붓스트랩의 추가 예

이 예에서는 작은 데이터 테이블에 부분 가중치 (베이지안 붓스트랩) 옵션을 사용할 경우의 이점을 보여 줍니다. 데이터는 7가지 토양 유형을 각각 포함하는 세 개의 표본에서 측정된 반응(Y)으로 구성되어 있습니다. 한 과학자가 **wabash** 토양 유형의 평균 반응에 대한 신뢰 구간을 구하려고 합니다.

각 토양 유형마다 관측값이 세 개씩만 있으므로 단순 붓스트랩을 수행하면 **wabash**의 관측값 세 개가 모두 붓스트랩 표본에서 제외될 가능성이 있습니다. 부분 가중치 옵션을 사용하면 모든 붓스트랩 표본에서 각 토양 유형의 모든 관측값을 나타낼 수 있습니다.

이 과학자는 다음과 같은 두 가지 붓스트랩 방법을 모두 사용하여 **wabash** 표본 평균의 분포를 확인합니다.

- "[단순 붓스트랩 분석](#)"(343 페이지)
- "[베이지안 붓스트랩 분석](#)"(346 페이지)

단순 붓스트랩 분석

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Snapdragon.jmp**를 엽니다.
2. **분석 > X로 Y 적합**을 선택합니다.
3. Y를 선택하고 **Y, 반응**을 클릭합니다.
4. **Soil**을 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "Y 대 Soil의 일원 분석" 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA**를 선택합니다.
7. **일원 ANOVA에 대한 평균** 보고서에서 **평균** 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다.
8. **붓스트랩 표본 수**에 1000을 입력합니다.
9. (선택 사항) 그림 11.7의 결과와 매칭되는 결과를 얻으려면 **난수 시드값**에 12345를 입력합니다.
10. **확인**을 클릭합니다.

그림 11.7 단순 붓스트랩에 대한 붓스트랩 결과

	X	Y	BootID•	clarion	clinton	compost	knox	o'neill	wabash	webster
1	Soil	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	Soil	Y	1	32.2333	30.9000	28.0000	35.7000	34.3500	31.9000	31.1000
3	Soil	Y	2	32.4333	30.3000	30.2000	33.9667	31.2000	38.0000	•
4	Soil	Y	3	32.5000	30.7500	29.2000	34.0333	32.8000	38.0000	31.1000
5	Soil	Y	4	31.5000	29.4000	29.9000	35.7500	35.0400	34.9500	•
6	Soil	Y	5	32.5000	30.6000	31.8000	35.7000	34.3000	35.0500	31.5667
7	Soil	Y	6	32.4000	30.1000	29.8500	34.4500	36.0000	38.2000	30.4000
8	Soil	Y	7	31.9000	29.3400	•	34.4000	33.6000	35.9667	31.4500
9	Soil	Y	8	32.1400	31.3500	29.6667	33.1000	35.1000	37.9333	30.6333
10	Soil	Y	9	32.7000	30.8000	30.2000	35.2600	31.2000	34.8500	32.5000
11	Soil	Y	10	32.1000	32.1000	28.0000	33.1000	34.1143	34.0000	31.8000
12	Soil	Y	11	31.5000	30.3000	•	33.1000	34.4000	37.2125	30.6333
13	Soil	Y	12	32.7000	30.4200	31.8000	34.0333	35.1000	35.0500	•
14	Soil	Y	13	32.1667	•	30.1000	34.9000	33.9000	38.2000	31.8000
15	Soil	Y	14	32.3000	•	29.2000	33.9667	34.4000	35.6000	31.9400
16	Soil	Y	15	32.2500	29.1000	29.6667	33.1000	35.1000	31.9000	•
17	Soil	Y	16	31.5000	30.4200	30.0000	•	35.2800	37.8000	31.1000
18	Soil	Y	17	32.4333	29.7000	29.2000	33.1000	34.4000	34.0000	31.1000
19	Soil	Y	18	32.3800	32.1000	29.9000	35.8000	34.6800	35.0500	31.8000
20	Soil	Y	19	32.0600	30.1000	30.2800	•	33.1500	31.9000	31.1000
21	Soil	Y	20	31.9000	29.1000	28.4000	35.9000	34.8000	34.8500	30.8200

그림 11.7의 결측값은 해당 붓스트랩 표본에 주어진 토양 유형의 관측값이 모두 선택되지 않은 붓스트랩 반복을 나타냅니다.

11. **분석 > 분포**를 선택합니다.
12. wabash를 선택하고 **Y, 열**을 클릭합니다.
13. **확인**을 클릭합니다.

그림 11.8 단순 붓스트랩으로 구한 wabash 평균의 분포

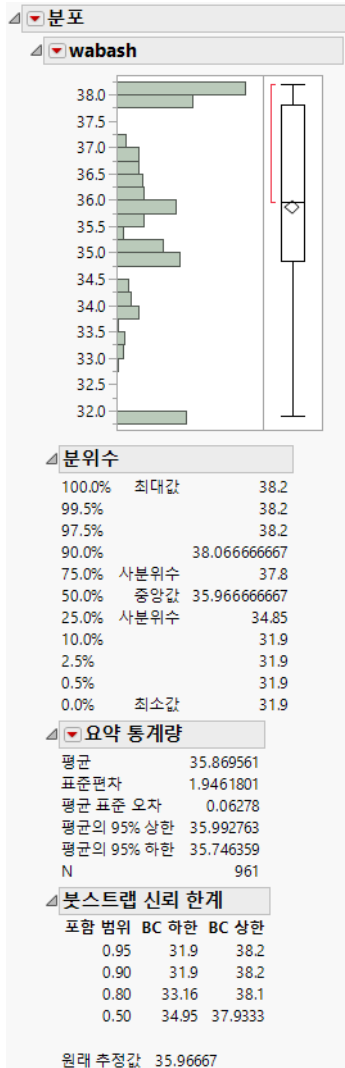


그림 11.8에서는 단순 붓스트랩 분석으로 구한 wabash 평균의 분포를 보여 줍니다. 다음을 확인할 수 있습니다.

- " 요약 통계량 " 보고서는 wabash 에 대한 붓스트랩 평균을 포함하는 행의 개수가 N=961 임을 나타냅니다. 1,000 회 반복을 수행했지만 39 개의 붓스트랩 표본에는 wabash 에 대한 세 개의 관측값이 전혀 포함되지 않았습니다.
- 표본 평균 히스토그램은 평활하지 않고 두 개의 극단값에서 정상점을 보여 줍니다. wabash 에 대한 세 개의 값은 38.2, 37.8 및 31.9 입니다. 분포의 낮은 쪽 정상점은 붓스트랩 표본에 값 31.9 만 포함되어서 발생한 경우이고, 높은 쪽 정상점은 붓스트랩 표본에 값 38.2 와 37.8 이 하나 또는 모두 포함되어서 발생한 경우입니다.

다음으로는 부분 가중치 (베이지안 붓스트랩) 옵션을 사용하여 붓스트랩 표본에서 결측값이 발생하지 않도록 하고 붓스트랩 평균의 분포를 평활하게 합니다 .

베이지안 붓스트랩 분석

1. " 일원 분석 " 보고서에서 **일원 ANOVA 에 대한 평균** 보고서의 **평균** 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다 .
2. **붓스트랩 표본 수**에 1000 을 입력합니다 .
3. (선택 사항) 그림 11.9 의 결과와 매칭되는 결과를 얻으려면 **난수 시드값**에 12345 를 입력합니다 .
4. **부분 가중치** 옵션을 선택합니다 .
5. **확인**을 클릭합니다 .

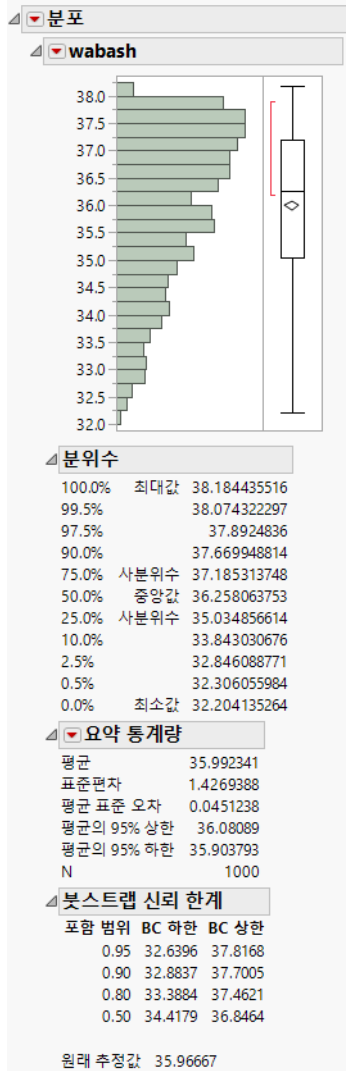
그림 11.9 베이지안 붓스트랩에 대한 붓스트랩 결과

	X	Y	BootID•	clarion	clinton	compost	knox	o'neill	wabash	webster
1	Soil	Y	0	32.1667	30.3000	29.6667	34.9000	33.8000	35.9667	31.1000
2	Soil	Y	1	31.9365	31.3493	29.3234	35.4497	34.7105	33.7270	31.4071
3	Soil	Y	2	32.4189	30.3474	29.5143	35.6281	32.7006	34.1027	32.1212
4	Soil	Y	3	32.2339	30.2001	31.4102	34.8758	33.6674	38.0389	31.5428
5	Soil	Y	4	32.4054	30.3242	30.6227	33.8495	32.9495	36.8344	31.9607
6	Soil	Y	5	32.2262	30.8672	29.7999	33.6759	32.2022	35.5792	31.9058
7	Soil	Y	6	32.3222	31.9732	28.8823	35.6307	34.8863	35.3014	30.1325
8	Soil	Y	7	31.9948	30.8828	29.2516	35.3424	33.2094	36.3367	30.8386
9	Soil	Y	8	31.6254	29.7677	28.6390	34.4697	33.0662	36.6183	31.4667
10	Soil	Y	9	32.3499	29.9416	29.5732	35.2564	31.9583	35.8246	29.9302
11	Soil	Y	10	32.5228	30.4506	28.4859	34.9088	35.8126	34.6317	31.6100
12	Soil	Y	11	31.9057	30.0711	29.0693	35.4018	34.4654	33.2086	31.0309
13	Soil	Y	12	31.7275	29.6189	29.1609	34.3984	33.6840	35.0815	31.5446
14	Soil	Y	13	32.5893	30.5210	28.6054	33.4594	34.0958	33.7692	31.7129
15	Soil	Y	14	32.2473	30.5199	31.6080	35.5617	34.1706	35.7146	31.8023
16	Soil	Y	15	32.2329	29.7275	30.1770	35.3286	32.7820	37.4866	30.9706
17	Soil	Y	16	31.5831	29.7508	28.7122	33.5304	34.5348	37.1092	31.2752
18	Soil	Y	17	32.3545	31.3237	29.1542	35.5890	32.2606	37.2005	30.8430
19	Soil	Y	18	32.3811	29.6241	30.6138	35.4308	33.2024	33.0787	31.2926
20	Soil	Y	19	31.7488	29.7763	28.7327	34.7007	33.8910	34.0573	29.9064

베이지안 붓스트랩 결과 테이블에는 결측값이 없습니다 . Snapdragon.jmp 데이터 테이블의 21 개 행이 모두 포함되었으며 각 붓스트랩 표본에서 붓스트랩 가중치만 달라졌습니다 .

6. **분석 > 분포**를 선택합니다 .
7. **wabash** 를 선택하고 **Y, 열**을 클릭합니다 .
8. **확인**을 클릭합니다 .

그림 11.10 베이지안 붓스트랩으로 구한 wabash 평균의 분포



베이지안 붓스트랩은 wabash 표본 평균에 대해 훨씬 더 평활한 분포를 생성합니다. 1,000 개의 붓스트랩 표본 모두에 wabash 에 대한 세 개의 관측값이 포함되어 있습니다. 각 반복 시마다 서로 다른 부분 가중치를 사용하여 wabash 표본 평균이 계산됩니다.

"붓스트랩 신뢰 한계" 보고서에는 평균에 대한 95% 신뢰 구간이 32.6396 ~ 37.8168 로 나타납니다.

JMP PRO 붓스트랩에 대한 통계 상세 정보

JMP PRO 부분 가중치 계산

부분 가중치 옵션은 베이지안 붓스트랩 (Rubin 1981) 을 기반으로 합니다. 하나의 관측값이 주어진 붓스트랩 표본에 나타나는 횟수를 붓스트랩 가중치라고 합니다. 단순 붓스트랩에서는 각 붓스트랩 표본에 대한 붓스트랩 가중치가 단순 랜덤 복원 표집 방식으로 결정됩니다.

베이지안 방법에서는 표집 확률이 알 수 없는 모수로 처리되며 무정보 사전 분포를 사용하여 확률의 사후 분포를 구합니다. 확률 추정값은 이 사후 분포에서 표집을 통해 구합니다. 이러한 추정값은 다음과 같이 붓스트랩 가중치를 생성하는 데 사용됩니다.

- 형태 모수가 $(n-1)/n$ 이고 척도 모수가 1 인 감마 분포의 n 개 값을 포함하는 벡터를 무작위로 생성합니다.

참고: Rubin 연구 자료 (1981) 에서는 감마 형태 모수로 1 을 사용합니다. JMP Pro 에서 사용되는 형태 모수는 부분 가중치의 평균 및 분산이 단순 붓스트랩 가중치의 평균 및 분산과 같도록 합니다.

- n 개 값의 합인 S 를 계산합니다.
- n 개 값의 벡터에 N/S 를 곱해서 부분 가중치를 계산합니다. 여기서 N 은 행 수 또는 빈도 합 (빈도 변수가 지정된 경우) 과 같습니다.

참고: 분서에 빈도 변수가 지정된 경우에는 행별로 감마 분포의 형태 모수에 빈도 값을 곱하십시오. 빈도 변수 값의 합은 1 보다 커야 합니다. 그러면 형태 모수가 $f_i(N-1)/N$ 과 같아집니다. 여기서 f_i 는 i 번째 행의 빈도 값이며 N 은 빈도 값의 합과 같습니다.

이 절차에서는 붓스트랩 표집의 평균 및 분산이 단순 붓스트랩 가중치의 평균 및 분산과 같도록 각 행의 부분 가중치를 척도화합니다. 각 붓스트랩 표본의 부분 붓스트랩 가중치는 양수이고, 합은 n 이 되며, 평균은 1 이 됩니다.

JMP PRO 편향 수정 백분위수 구간

이 섹션에서는 붓스트랩 결과 테이블에서 "분포" 스크립트를 실행하면 "붓스트랩 신뢰 한계" 보고서에 표시되는 BC (편향 수정) 신뢰 구간의 계산 방법을 설명합니다. 편향 수정 백분위수 구간은 붓스트랩 분포에서 백분위수 구간의 비대칭성 설명 능력을 향상시킵니다. 자세한 내용은 Efron 연구 자료 (1981) 에서 확인하십시오.

표기

- p^* - 관심 있는 통계량의 추정값이 원래 추정값보다 작거나 같은 붓스트랩 표본의 비율입니다.
- z_0 - 표준 정규 분포의 p^* 분위수입니다.

- z_α - 표준 정규 분포의 α 분위수입니다.

편향 수정 신뢰 구간의 끝점

$(1 - \alpha)$ 편향 수정 신뢰 구간의 끝점은 붓스트랩 분포의 분위수로 주어집니다.

- 하한점은 다음 분위수입니다.

$$\Phi\left(2z_0 + z_{\frac{\alpha}{2}}\right)$$

- 상한점은 다음 분위수입니다.

$$\Phi\left(2z_0 + z_{1 - \frac{\alpha}{2}}\right)$$

붓스트랩의 가용성

붓스트랩은 다음 통계 플랫폼에서 사용할 수 있습니다.

- 부스티드 트리
- 붓스트랩 포레스트
- 범주형
- 파괴 열화
- 판별
- 분포
- 곡선 적합
- 수명 분포 적합
- 모수 생존 모형 적합
- 비례 위험 모형 적합
- X 로 Y 적합
- 일반화 선형 모형
- 일반화 회귀
- 수명 분포
- 로지스틱
- 로그 선형 분산
- 다중 대응 분석
- 다변량

- 신경망
- 비선형
- 모수 생존
- 부분 최소 제곱
- 파티션
- 주성분
- 비례 위험
- 표준 최소 제곱
- 생존
- Uplift

12 장

텍스트 탐색기

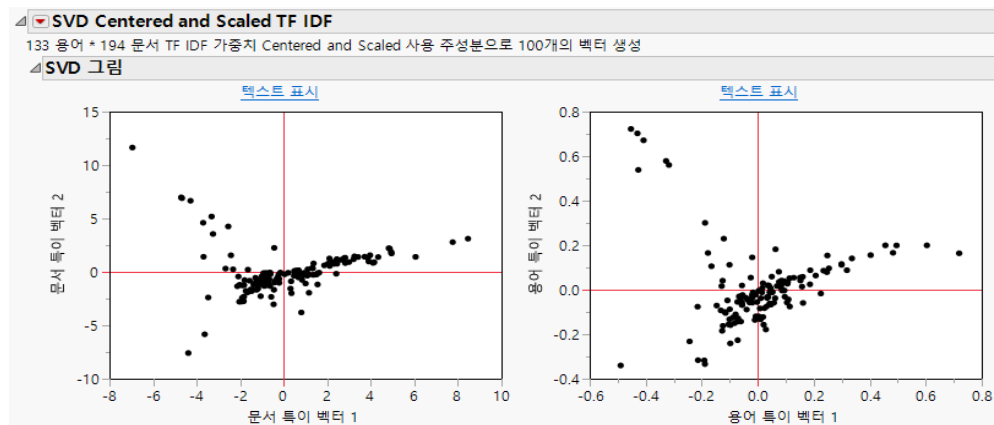
데이터의 비정형 텍스트 탐구

JMP PRO 이 플랫폼의 많은 기능은 JMPPro 에서만 사용할 수 있으며 그러한 경우 이 아이콘이 표시됩니다.

텍스트 탐색기 플랫폼을 사용하면 설문 조사 또는 사고 보고서의 주석 필드와 같은 비정형 텍스트를 분석할 수 있습니다. 도구를 사용하여 유사한 용어를 결합하고 잘못 지정된 용어를 재코딩하고 텍스트 데이터의 기본 패턴을 이해함으로써 텍스트 데이터와 상호 작용할 수 있습니다.

JMP PRO JMP Pro 버전의 플랫폼에는 SVD(특이값 분해)를 사용하여 유사한 문서를 주제별로 그룹화하는 분석 도구도 포함되어 있습니다. 텍스트 문서를 군집화하거나 문서 모음에 있는 용어를 군집화할 수 있습니다. 잠재 계층 분석을 사용하여 문서를 군집화할 수도 있습니다.

그림 12.1 텍스트 탐색기의 SVD 그림



목차

텍스트 탐색기 플랫폼 개요.....	353
텍스트 처리 단계	354
텍스트 탐색기 플랫폼의 예.....	355
텍스트 탐색기 플랫폼 시작.....	358
정규 표현식 편집기에서 정규 표현식 사용자 정의	360
텍스트 탐색기 보고서	366
요약 개수 보고서	366
용어 및 구 목록.....	367
텍스트 탐색기 플랫폼 옵션.....	370
텍스트 준비 옵션	370
텍스트 분석 옵션	375
저장 옵션.....	376
보고서 옵션.....	378
잠재 계층 분석	378
잠재 의미 분석 (SVD)	380
SVD 보고서.....	381
SVD 보고서 옵션	381
주제 분석.....	383
주제 분석 보고서	384
주제 분석 보고서 옵션.....	384
판별 분석.....	385
판별 분석 보고서	385
판별 분석 보고서 옵션.....	386
텍스트 탐색기 플랫폼의 추가 예	387

텍스트 탐색기 플랫폼 개요

비정형 텍스트 데이터는 일반적으로 볼 수 있습니다. 예를 들어 설문 조사, 제품 평가 의견 또는 사고 보고서의 자유 응답 필드에서 비정형 텍스트 데이터가 발생할 수 있습니다. 텍스트 탐색기 플랫폼을 사용하면 비정형 텍스트를 탐색하여 텍스트의 의미를 더 잘 이해할 수 있습니다. 텍스트 분석은 대개 반복적인 프로세스이므로 용어 목록의 큐레이팅과 분석을 번갈아가며 수행할 수 있습니다.

용어 목록 큐레이팅

텍스트 분석에는 몇 가지 고유한 용어가 사용됩니다. 용어 또는 토큰은 가장 작은 텍스트 조각으로, 문장 내의 단어와 유사합니다. 하지만 용어는 정규 표현식을 사용하는 등의 여러 가지 방법으로 정의할 수 있습니다. 텍스트를 용어로 분해하는 프로세스를 토큰화라고 합니다.

- 구는 짧은 용어 모음입니다. 텍스트 탐색기 플랫폼에는 자체적으로 용어로 지정된 구를 관리하는 옵션이 있습니다.
- 문서는 단어 모음을 나타냅니다. JMP 데이터 테이블에서는 텍스트 열의 각 행에 있는 비정형 텍스트가 문서에 해당합니다.
- 코퍼스는 문서 모음을 나타냅니다.

때로는 분석에서 몇 가지 일반적인 단어를 제외하는 것이 좋습니다. 이 제외된 단어를 중지 단어라고 합니다. 플랫폼에 기본 중지 단어 목록이 있지만 사용자가 특정 단어를 중지 단어로 추가할 수도 있습니다. 중지 단어는 용어가 될 수는 없지만 구에 사용될 수는 있습니다.

용어를 재코딩할 수도 있습니다. 그러면 동의어를 하나의 공통 용어로 결합하는 데 유용합니다.

어간 추출 (Stemming) 은 서로 다른 어미를 제거함으로써 시작 부분이 동일한 단어 (어간) 를 결합하는 프로세스입니다. 즉, "jump", "jumped" 및 "jumping" 이 모두 "jump" 라는 용어로 취급됩니다. 어간 추출 절차는 Snowball 문자열 처리 언어에서 사용되는 절차와 유사합니다. 구를 어간 추출할 때는 구의 각 단어가 독립된 용어일 때와 마찬가지로 방식으로 어간 추출됩니다.

용어 목록 분석

텍스트 탐색기 플랫폼의 텍스트 분석에는 BoW(*bag of words*) 방법이 사용됩니다. 따라서 구를 형성할 때 외에는 용어의 순서가 무시됩니다. 이 분석은 용어의 개수를 기준으로 합니다.

정규 표현식, 중지 단어, 재코딩 및 어간 추출을 사용하여 용어 목록을 큐레이팅한 후에는 큐레이팅된 용어 목록에 대한 분석을 수행할 수 있습니다. 플랫폼의 분석 옵션은 DTM(문서 용어 행렬) 을 기반으로 합니다. DTM 의 각 행은 하나의 문서 (JMP 데이터 테이블의 텍스트 열에 있는 셀 하나) 에 해당합니다. DTM 의 각 열은 큐레이팅된 용어 목록의 용어 하나에 해당합니다. 이 방법은 단어 순서를 무시하기 때문에 BoW 방법을 구현합니다. 가장 단순한 형태일 때 DTM 의 각 셀에는 행의 문서에 있는 열의 용어 빈도 (발생 횟수) 가 포함됩니다. DTM 을 위한 다른 가중치 체계가 다양하게 존재하며, 이는 " 저장 옵션 "(376 페이지) 에 설명되어 있습니다.

JMP PRO 플랫폼에서 사용할 수 있는 분석 옵션은 먼저 문서 용어 행렬에 대해 SVD(특이값 분해)를 수행합니다. 이를 통해 데이터의 용어 정보를 나타내는 데 필요한 열 수가 크게 줄어들 수 있습니다. 특이값 분해에 대한 자세한 내용은 **Multivariate Methods**의 부록 "Statistical Details"에서 확인하십시오. 용어 군집화 및 문서 군집화에는 "계층적 군집화" 옵션을 사용할 수 있습니다. 이 옵션을 사용하면 유사한 용어 또는 문서끼리 그룹화할 수 있습니다.

플랫폼 워크플로우

텍스트 탐색기 플랫폼을 사용하기 위해 수행해야 하는 단계는 다음과 같습니다.

1. 토큰화 방법 (기본 제공 또는 사용자 정의 정규 표현식) 을 지정합니다.
2. 보고서를 사용하여 추가 중지 단어를 지정하고, 용어 목록에 구를 추가하고, 용어의 재코딩을 수행하고, 어간 추출 규칙에 대한 예외를 지정합니다.
3. 어간 추출에 대한 환경 설정을 지정합니다.
4. 단어 및 문 개수, SVD 및 군집화 방법을 사용하여 중요한 용어 및 구를 식별합니다.

참고 : **JMP PRO** SVD 및 군집화 옵션은 JMP Pro에서만 사용할 수 있습니다.

5. 결과 (용어 테이블, DTM, 특이 벡터 또는 기타 결과) 를 추가 분석에 사용할 수 있도록 저장합니다.

참고 : **JMP PRO** 특이 벡터를 저장하는 옵션은 JMP Pro에서만 사용할 수 있습니다.

6. 구, 재코딩 및 중지 단어 특성을 유사한 텍스트 데이터의 향후 분석에 사용할 수 있도록 저장합니다.

텍스트 처리 단계

텍스트는 토큰화, 구 추출 및 용어 추출의 세 단계로 처리됩니다.

토큰화 단계

토큰화 단계에서는 다음 작업을 수행합니다.

1. 텍스트를 소문자로 변환합니다.
2. 토큰화 방법 (기본 단어 또는 정규 표현식) 을 적용하여 문자를 토큰으로 그룹화합니다.
3. 지정된 재코딩 정의에 따라 토큰을 재코딩합니다. 재코딩은 어간 추출 (Stemming) 전에 수행됩니다.

구 추출 단계

구 추출 단계에서는 코퍼스 (문서 모음) 에서 나타나는 구를 수집하며, 이를 통해 개별 구를 용어로 처리하도록 지정할 수 있습니다. 구는 중지 단어로 시작하거나 끝날 수 없지만 중지 단어를 포함할 수는 있습니다.

용어 추출 단계

용어 추출 단계에서는 이전 단계에서 추출된 토큰 및 구로부터 용어 목록을 생성합니다.

용어 추출 단계에서는 각 토큰에 대해 다음 작업을 수행합니다.

1. 시작 창에 지정된 최소 및 최대 길이 요구 사항이 충족되는지 확인합니다. 숫자만 포함된 토큰은 이 작업에서 제외됩니다.
2. 토큰이 용어가 될 수 있는 자격이 있는지 확인합니다. 기본 단어 토큰화 방법으로 파싱된 토큰에는 알파벳 문자 또는 유니코드 문자가 하나 이상 포함되어야 합니다. 숫자만 포함된 토큰은 이 작업에서 제외됩니다. 정규 표현식 토큰화 방법에서는 정규 표현식을 사용하여 토큰의 일부인 문자를 확인합니다.
3. 토큰이 중지 단어가 아닌지 확인합니다.
4. 어간 추출 및 어간 예외를 적용합니다.

용어 추출 단계에서는 사용자가 추가하는 각 구에 대해 다음 작업을 수행합니다.

1. 용어 목록에 구를 추가합니다. 구는 용어 목록에 어간 추출된 구의 각 단어에 어간 추출을 적용해야 합니다. 원시 토큰은 다르지만 어간이 동일한 구는 용어 목록에서 결합됩니다.
2. 구에 나타나는 토큰 용어 항목을 제거합니다.

텍스트 탐색기 플랫폼의 예

이 예에서는 반려 동물에 대한 설문 조사에서 얻은 텍스트 응답을 살펴보려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 Pet Survey.jsp를 엽니다.
2. **분석 > 텍스트 탐색기**를 선택합니다.
3. Survey Response 를 선택하고 **텍스트 열**을 클릭합니다.
4. "언어" 목록에서 **영어**를 선택합니다.
5. **확인**을 클릭합니다.

그림 12.2 초기 텍스트 탐색기 보고서의 예



194 개 문서에 372 개의 고유 용어가 있음을 한눈에 알 수 있습니다 . 총 1921 개의 토큰화된 용어가 있습니다 . 가장 자주 나타나는 용어는 55 회 나타나는 "cat" 입니다 .

- "Survey Response 에 대한 텍스트 탐색기 " 옆의 빨간색 삼각형을 클릭하고 **용어 옵션 > 어간 추출 (Stemming) > 모든 용어의 어간 추출**을 선택합니다 .
- 구 목록 테이블에서 **cat food** 및 **dog food** 를 선택하고 선택 항목을 마우스 오른쪽 버튼으로 클릭한 다음 **구 추가**를 선택합니다 .
"cat food" 와 "dog food" 라는 용어가 용어 목록에 포함됩니다 .
- 용어 목록을 아래로 스크롤하여 "cat food" 및 "dog food" 항목을 찾습니다 .
각 구가 네 번씩 나타남을 알 수 있습니다 .

그림 12.3 수정 및 스크롤 후 용어 목록

용어	개수
cat- food-	4
couch-	4
dog- food-	4
door-	4
duck-	4
get-	4
great-	4
guard-	4
job-	4
know-	4
made-	4
make-	4
now-	4

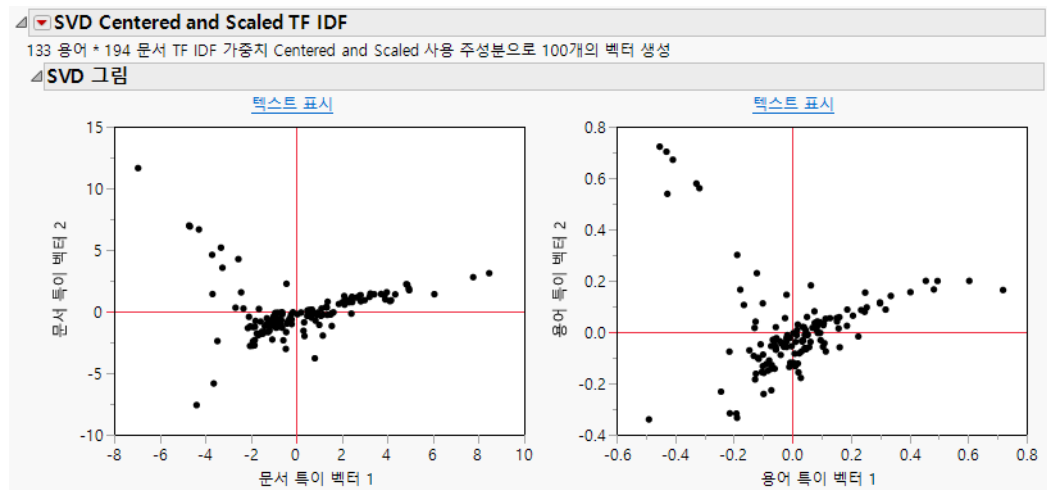
"cat food" 와 "dog food" 는 이제 이 텍스트 탐색기 보고서에 한해 용어로 처리되므로 구 목록에서는 이 두 항목이 회색으로 표시됩니다.

JMP PRO 이 예의 나머지 단계는 JMP Pro 에서만 수행할 수 있습니다.

9. **JMP PRO** "Survey Response 에 대한 텍스트 탐색기" 옆의 빨간색 삼각형을 클릭하고 **잠재 의미 분석, SVD** 를 선택합니다.
10. **JMP PRO** **확인**을 클릭하여 기본값을 적용합니다.

보고서에 두 개의 SVD 그림이 나타납니다. 왼쪽 그림에서는 문서 공간에 있는 처음 두 개의 특이 벡터를 보여 줍니다. 오른쪽 그림에서는 용어 공간에 있는 처음 두 개의 특이 벡터를 보여 줍니다.

그림 12.4 SVD 그림



11. **JMP PRO** 왼쪽 SVD 그림에서 맨 오른쪽에 있는 점 세 개를 선택합니다.
이 세 개의 점은 나머지 점과는 동떨어져서 군집화된 설문 조사 응답을 나타냅니다. 이 군집을 보다 자세히 조사하려면 해당 응답의 텍스트를 읽어 보십시오.
12. **JMP PRO** 왼쪽 SVD 그림 위에 있는 **텍스트 표시** 버튼을 클릭합니다.

그림 12.5 선택한 문서의 텍스트

There was this funny video of a cat trying to jump into someones lap, but fell into the pool instead. [56]
 The funny cat video where the cats jumped through the window right into the bathtub was hilarious. [142]
 We made this funny video of the cat trying to climb the wall to chase a laser pointer. [153]

선택한 점이 나타내는 세 문서의 텍스트가 포함된 창이 나타납니다. 이러한 설문 조사 응답은 모두 "funny", "cat" 및 "video"의 몇 가지 조합을 언급한다는 점에서 유사합니다. 이러한 문서는 첫 번째 특이 벡터에 대해 나머지 문서보다 큰 양의 값을 가집니다. 이렇게 큰 값은 이러한 문서가 해당 차원의 나머지 문서와는 다를 수 있음을 나타냅니다.

특이 벡터의 차원을 더 조사해 보면 차원이 나타내는 것을 해석할 수 있습니다. 예를 들어 그림의 맨 오른쪽에 있는 많은 문서들은 고양이와 관한 응답입니다. 맨 왼쪽의 많은 응답은 개에 관한 것입니다. 따라서 첫 번째 특이 벡터에서는 응답이 고양이와 관한 것인지 개에 관한 것인지에 따라 차이를 찾아냅니다.

텍스트 탐색기 플랫폼 시작

분석 > 텍스트 탐색기를 선택하여 텍스트 탐색기 플랫폼을 시작할 수 있습니다.

그림 12.6 텍스트 탐색기 시작 창

프리 형식 텍스트를 분석합니다.

열 선택

▼ 1개 열

Survey Response

언어

영어 ▼

구당 최대 단어 수: 4

최대 구 수: 5000

단어당 최소 문자 수: 1

단어당 최대 문자 수: 50

어간 추출: 어간 추출 안 함 ▼

토론화: 정규 표현식 ▼

☐ 정규 표현식 사용자 정의

☐ 숫자를 단어로 처리

선택한 열 역할 지정

텍스트 열	끝수 문자
검증	선택적 숫자
ID	선택적
기준	선택적

작업

확인

취소

제거

재호출

도움말

"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 "시작하기"장에서 확인하십시오. 텍스트 탐색기 시작 창에는 다음과 같은 옵션이 포함되어 있습니다.

텍스트 열 텍스트 데이터가 포함된 열을 할당합니다. 여러 열을 지정하면 각 열에 대해 별도의 분석이 생성됩니다.

JMP Pro 검증 JMP Pro에서는 검증 열을 입력할 수 있습니다. "열 선택" 목록에서 아무 열도 선택하지 않은 상태로 "검증" 버튼을 클릭하면 데이터 테이블에 검증 열을 추가할 수 있습니다.

" 검증 열 생성 " 유틸리티에 대한 자세한 내용은 Predictive and Specialized Modeling 의 "Make Validation Column" 장에서 확인하십시오 .

검증 열을 지정해도 문서 용어 행렬의 계산에는 영향이 없습니다 . 하지만 검증 열이 지정된 경우에는 " 잠재 계층 분석 ", " 잠재 의미 분석 ", " 주제 분석 " 및 " 관별 분석 " 옵션에 훈련 데이터 집합만 사용됩니다 .

ID " 연관성을 위해 쌓인 형식의 DTM 저장 " 출력 데이터 테이블에서 개별 응답자를 식별하는 데 사용되는 열을 할당합니다 . 이 출력 데이터 테이블은 연관성 분석에 적합합니다 . 이 열은 " 잠재 계층 분석 " 보고서에서 개별 응답자를 식별하는 데도 사용됩니다 .

기준 변수의 각 수준에 대한 개별 분석으로 구성된 보고서를 생성하는 열을 식별합니다 . 기준 변수가 둘 이상 할당되면 기준 변수의 가능한 각 수준 조합에 대해 개별 보고서가 생성됩니다 .

참고 : 기준 변수를 지정하는 경우 기준 변수의 모든 수준에 " 정규 표현식 사용자 정의 " 옵션 및 설정이 적용됩니다 .

언어 텍스트 처리에 사용되는 언어를 지정합니다 . 이는 어간 추출 (Stemming) 과 중지 단어 , 재 코딩 및 구의 기본 제공 목록에 영향을 줍니다 . 이 옵션은 JMP 실행 언어와는 별개입니다 . " 언어 " 플랫폼 환경 설정이 지정되지 않은 경우 이 " 언어 " 옵션은 JMP " 표시 언어 " 환경 설정에 따라 설정됩니다 . 하지만 텍스트 탐색기의 " 언어 " 옵션은 한국어를 지원하지 않습니다 . JMP 표시 언어가 한국어인 경우 이 옵션은 기본적으로 영어로 설정됩니다 .

구당 최대 단어 수 구가 분석에 구로 포함되기 위해 최대로 포함할 수 있는 단어 수를 지정합니다 .

최대 구 수 구 목록에 나타나는 구의 최대 개수를 지정합니다 .

단어당 최소 문자 수 단어가 분석에 용어로 포함되기 위해 반드시 포함해야 하는 문자 수를 지정합니다 .

단어당 최대 문자 수 단어가 분석에 용어로 포함되기 위해 최대로 포함할 수 있는 문자 수 (최대 2000) 를 지정합니다 .

어간 추출 (Stemming) (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 시작 문자는 유사하지만 종료 문자는 다른 용어를 결합하는 방법을 지정합니다 . 다음 옵션을 사용할 수 있습니다 .

어간 추출 (Stemming) 안 함 용어를 결합하지 않습니다 .

결합 가능한 경우 어간 추출 둘 이상의 용어에서 추출된 어간이 동일한 용어일 경우에만 용어의 어간을 추출합니다 .

모든 용어의 어간 추출 모든 용어의 어간을 추출합니다 .

참고 : " 어간 추출 (Stemming)" 옵션을 사용할 경우 용어 목록에 추가된 구에도 영향을 주게 됩니다 . 구 내의 용어가 어간 추출된 후에는 구 식별이 발생합니다 . 예를 들어 "dogs bark" 와 "dog barks" 는 모두 "dog·bark " 라는 특정 구와 매칭됩니다 .

토큰화 (" 언어 " 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 텍스트를 용어 또는 토큰으로 파싱하기 위한 방법을 지정합니다. 사용 가능한 토큰화 옵션은 다음과 같습니다.

정규 표현식 기본 제공 정규 표현식을 사용하여 텍스트를 파싱합니다. 텍스트를 파싱하는데 사용되는 정규 표현식을 추가, 제거 또는 편집하려면 **정규 표현식 사용자 정의** 옵션을 선택합니다. 자세한 내용은 "[정규 표현식 편집기에서 정규 표현식 사용자 정의](#)"(360 페이지)에서 확인하십시오.

기본 단어 텍스트는 일반적으로 단어를 구분하는 문자를 기준으로 파싱됩니다. 이러한 문자로는 공백, 탭, 줄바꿈 및 대부분의 구두점 표시가 포함됩니다. 분석을 위해 숫자를 용어로 파싱하려면 **숫자를 단어로 처리** 옵션을 선택합니다. 이 옵션을 선택하지 않을 경우 구분자 사이에서 숫자만 포함하는 텍스트는 토큰화 단계에서 무시됩니다.

팁: 기본 단어 토큰화 방법을 사용하는 텍스트 탐색기 보고서에서 **표시 옵션 > 구분자 표시** 옵션을 사용하여 기본 구분자 집합을 볼 수 있습니다.

정규 표현식 사용자 정의 (정규 표현식 토큰화 방법을 사용하는 경우에만 사용 가능) 텍스트 탐색기 정규 표현식 편집기 창을 사용하여 정규 표현식 설정을 수정할 수 있습니다. 일반적이지 않은 단어를 수용하려면 이 옵션을 사용합니다. 예를 들어 전화 번호나 문자와 숫자의 조합으로 구성된 단어를 들 수 있습니다. "정규 표현식 사용자 정의" 옵션은 기본 정규 표현식 방법으로 필요한 결과를 얻을 수 없는 경우에만 사용하는 것이 좋습니다. 이러한 경우는 텍스트에 기본 정규 표현식 방법으로 인식되지 않는 구조가 포함된 경우에 발생할 수 있습니다. 자세한 내용은 "[정규 표현식 편집기에서 정규 표현식 사용자 정의](#)"(360 페이지)에서 확인하십시오.

숫자를 단어로 처리 (기본 단어 토큰화 방법과 함께만 사용 가능) 분석에서 숫자를 용어로 토큰화할 수 있도록 합니다. 이 옵션이 선택되어 있으면 숫자가 포함된 용어에 대해 "단어당 최소 문자 수" 설정이 무시됩니다.

시작 창에서 **정규 표현식 사용자 정의**를 선택한 경우 시작 창에서 **확인**을 클릭하면 텍스트 탐색기 정규 표현식 편집기 창이 나타납니다. 그렇지 않은 경우에는 텍스트 탐색기 보고서가 나타납니다.

참고: 텍스트 입력 처리는 대/소문자를 구분하지 않습니다. 모든 텍스트는 토큰화 단계와 모든 분석 단계 이전에 내부적으로 소문자로 변환됩니다. 이 변환은 텍스트 탐색기 출력에서 정규 표현식 처리와 용어 집계에 영향을 미칩니다.

정규 표현식 편집기에서 정규 표현식 사용자 정의

정규 표현식 사용자 정의 옵션을 선택하면 텍스트 탐색기 정규 표현식 편집기가 나타납니다. 이 창에서 전화 번호, 시간 또는 통화 값과 같은 다양한 기본 제공 정규 표현식을 사용하여 텍스트 문서를 파싱할 수 있습니다. 고유한 정규 표현식 정의를 생성할 수도 있습니다.

Tip: 시작 창에서 "언어" 옵션으로 일본어, 중국어(간체) 또는 중국어(번체)가 지정된 경우에는 정규 표현식 패턴 목록에 지정된 언어의 단일 정규 표현식이 포함됩니다. 다른 정규 표현식 패턴을 추가하려면 단일 정규 표현식 패턴 뒤에 추가하는 것이 좋습니다. 단어 패턴은 긴 일련의 아시아 언어 문자를 하나의 단어로 취합할 수 있기 때문에 언어별 정규 표현식 패턴 앞에 단어 패턴을 사용하지 않는 것이 좋습니다.

1 1.234 spanned text 1.2 -3.14E+0 foot 3" 4' 5gm 10 cm
brown-fox at 919-555-1234 or (919)-555-1234, \$30.00! 3:24 a.m., 5am 6PM 100°F https
://www.jsp.com/community.sas.com "It's time!" <a rel="nofollow" href="only capture the
link" target="_blank"> super!!!!!! EARTH, NA,US,NC, WAKE, CARY, SAS

기본 샘플 텍스트

행

이전

다음

행이 선택되지 않았습니다.

열에 저장

정규 표현식 편집기

단어 구분 기호 목록

금액

단어

HTML 링크 손 도구

시간

번호

나머지

↑

↓

+

-

제목

금액

Regex

(((\e{w,w\$}|r{w\$}ws{0,2}|w{,J}?[0-9]|w{,0-9}+)(\w{,J}?[0-9]|w{,0-9}+ws{0,2}){(\e{w,w\$}|r{w\$}ws{0,2}|w{,J}?[0-9]|w{,0-9}+))

결과

₩1

예제

\$5.24 \$1000 \$.25 €5.55 5.55€ ¥5.55 ₩5.55 ₪5.55 R\$5.55

주석

다양한 통화 기호를 섭주사 또는 섭미사로 사용하는 숫자와 매칭됩니다.

확인

취소

도움말

- 스크립트 편집기 상자에 사용자 데이터의 텍스트를 채우려면 **이전** 및 **다음** 행 단추를 클릭합니다. 이렇게 하면 지정된 텍스트 데이터 행이 파싱되는 방식을 볼 수 있습니다.
- 정규 표현식 토큰화 결과가 포함된 새 열을 데이터 테이블에 저장하려면 **열에 저장** 버튼을 클릭합니다. 정규 표현식의 결과를 지정하는 방법에 대한 자세한 내용은 "[정규 표현식 편집](#)"(362 페이지)에서 확인하십시오.

참고 : 열에 저장 버튼을 클릭할 경우 텍스트 매칭에 정규 표현식만 사용됩니다. 정규 표현식의 출력을 수정하는 데 중지 단어, 재코딩, 어간 추출 (Stemming), 구, 또는 단어당 최소 / 최대 문자 수 등의 설정은 사용되지 않습니다.

정규 표현식 추가

토큰화에 사용할 정규 표현식을 추가하려면 목록 아래의 더하기 기호를 클릭합니다. 그러면 "정규 표현식 라이브러리 선택" 창이 나타납니다. 이 창에는 모든 기본 제공 정규 표현식뿐만 아니라 이전에 정규 표현식 편집기에서 생성한 후 최근에 수정한 정규 표현식도 포함되어 있습니다. 기본 제공 정규 표현식에는 라벨이 지정되어 있습니다. 라이브러리에 저장된 사용자 정규 표현식에는 사용자가 지정한 이름으로 라벨이 지정됩니다. 지정된 이름의 가장 최근 표현식만 정규 표현식 라이브러리에 저장됩니다.

선택한 정규 표현식을 토큰화에 사용할 정규 표현식으로 추가하려면 목록에서 정규 표현식을 하나 이상 선택하고 **확인**을 클릭합니다. 정규 표현식 라이브러리에서 하나 이상의 사용자 정규 표현식을 제거하려면 **선택 항목 삭제** 버튼을 사용합니다. 각 사용자의 정규 표현식 라이브러리는 **TextExplorer** 라는 디렉터리에 JSL 파일로 저장됩니다. 이 디렉터리의 위치는 다음과 같이 컴퓨터 운영 체제에 따라 다릅니다.

- Windows: "C:/Users/< 사용자 이름 >/AppData/Roaming/SAS/JMP/TextExplorer/"
- macOS: "/Users/< 사용자 이름 >/Library/Application Support/JMP/TextExplorer/"

이러한 파일을 다른 사용자와 공유할 수는 있지만 파일을 직접 편집해서는 안 됩니다. 대신 정규 표현식 편집기를 사용하십시오.

정규 표현식 편집

정규 표현식 편집기 패널에 지정된 순서대로 정규 표현식을 처리하여 용어를 토큰화할 수 있습니다. 정규 표현식의 순서를 변경하려면 목록에서 정규 표현식을 선택하고 목록 아래의 위쪽 또는 아래쪽 화살표 버튼을 클릭합니다. 정규 표현식 목록의 항목을 드래그하여 놓는 방법으로 실행 순서를 변경할 수도 있습니다. 파란색 삼각형은 현재 선택된 정규 표현식을 나타냅니다. 정규 표현식을 제거하고 토큰화에서 제외하려면 목록에서 정규 표현식을 선택하고 목록 아래의 빼기 기호를 클릭합니다. "나머지" 정규 표현식은 제거할 수 없으며 정규 표현식 시퀀스에서 마지막으로 나타나야 합니다.

목록에서 정규 표현식을 선택하면 정규 표현식 편집기 패널의 편집 가능한 필드는 선택한 정규 표현식을 참조합니다. 필드를 편집하려면 해당 필드를 클릭하고 내용을 입력합니다.

각 정규 표현식에는 다음과 같은 속성이 있습니다.

제목 현재 창과 이후 정규 표현식 라이브러리에서 정규 표현식을 식별하는 데 사용할 이름을 지정합니다.

정규 표현식 정규 표현식 정의를 지정합니다. 정규 표현식 캡처를 지정하려면 정규 표현식에 하나 이상의 괄호 쌍이 있어야 합니다.

결과 정규 표현식과 매칭되는 텍스트를 대체할 항목을 지정합니다. 이 값은 정적 텍스트, 공백 또는 정규 표현식 캡처의 값일 수 있습니다. 정규 표현식 캡처는 정규 표현식 정의의 결과로 정의됩니다.

- 매칭되는 텍스트를 정적 텍스트로 바꾸려면 "결과" 필드에 정적 텍스트를 지정합니다.
- 매칭되는 텍스트를 무시하려면 "결과" 필드를 비워 둡니다.
- 정규 표현식의 가장 바깥쪽 괄호에서 나온 텍스트를 유지하려면 "결과" 필드에 "\1"(따옴표 제외) 을 사용합니다.
- 정규 표현식의 전체 결과를 유지하려면 "결과" 필드에 "\0"(따옴표 제외) 을 사용합니다.

예제 (선택 사항) 정규 표현식의 동작을 나타내는 색상이 포함된 예제 텍스트 문자열을 지정합니다.

주석 (선택 사항) 정규 표현식과 해당 동작을 설명하는 주석을 지정합니다.

색상 스크립트 편집기 상자와 " 예제 " 필드의 텍스트에서 정규 표현식의 매칭 항목을 식별하는데 사용되는 색상을 지정합니다. 색상을 변경하려면 화살표 버튼을 사용합니다.

참고 : "정규 표현식" 필드의 정규 표현식 정의가 올바르지 않은 경우에는 정규 표현식 목록에서 해당 정규 표현식 이름 옆에 빨간색 X가 표시됩니다.

사용자 정규 표현식 생성

사용자 정규 표현식을 생성하려면 다음 단계를 수행하십시오.

1. 목록 아래의 더하기 기호를 클릭합니다.
2. "정규 표현식 라이브러리 선택" 창에서 "공백" 정규 표현식이 선택되어 있는지 확인합니다.
3. **확인**을 클릭합니다.
4. "정규 표현식 편집기" 패널에서 정규 표현식 정의를 편집합니다.
5. "제목" 필드에서 사용자 정규 표현식의 고유 이름을 지정합니다.

팁 : 정규 표현식 정의 필드를 편집할 때 로그 창을 열어 표시 상태로 두면 유용합니다. 일부 오류 메시지는 로그 창에만 나타나기 때문입니다. 로그 창을 열려면 **보기 > 로그**를 선택하십시오. 인터넷에서 정규 표현식 문제를 해결하는 데 사용할 수 있는 다양한 리소스를 참조할 수 있습니다 (예 : <https://regexr.com/>).

단어 구분 기호 목록

단어 구분 기호 목록 버튼을 사용하면 토큰화 프로세스에서 단어 사이에 나타나는 문자의 목록을 지정할 수 있습니다. 단어 사이 문자는 단어의 시작 문자가 될 수는 없지만 정규 표현식 중 하나에서 허용될 경우 단어 내에는 나타날 수 있습니다. 이 버튼을 클릭하면 나타나는 창의 목록에서 문자를 추가하거나 제거할 수 있습니다. 기본적으로 이 목록에 포함된 문자는 공백 문자뿐입니다. "구분 기호 문자" 창에서 **재설정** 버튼을 클릭하면 구분 기호 문자 목록의 수정 사항이 모두 실행 취소됩니다. 구분 기호 문자 목록의 수정 사항은 현재 정규 표현식 토큰화에만 적용됩니다.

지정된 정규 표현식과 필요한 "나머지" 정규 표현식의 처리는 다음과 같이 진행됩니다.

1. 텍스트 스트림의 현재 문자를 구분 기호 문자 목록과 비교합니다.
 - 구분 기호 문자 목록에 있는 문자일 경우 해당 문자를 무시하고 "나머지" 임시 문자열에서 취합된 모든 문자를 처리한 후 다음 문자로 이동해서 1 단계를 반복합니다.
 - 구분 기호 문자 목록에 없는 문자일 경우 2 단계로 이동합니다.
2. 현재 문자부터 시작되는 문자열을 "나머지" 정규 표현식 바로 앞까지의 각 정규 표현식과 한 번에 하나씩 비교합니다.
 - 현재 문자부터 시작되는 문자열이 정규 표현식 중 하나와 매칭되면 후속 작업이 수행됩니다. 즉, "나머지" 임시 문자열에 취합된 모든 문자가 처리되고, "결과" 필드의 값이 용어로 저장됩니다. 텍스트 스트림의 현재 문자는 매칭되는 문자열 다음의 문자가 됩니다. 그런 다음 1 단계 처리로 돌아갑니다.
 - 현재 문자부터 시작되는 문자열이 "나머지" 정규 표현식까지의 모든 정규 표현식과 매칭되지 않을 경우에는 3 단계로 이동합니다.
3. 현재 문자를 추가하고 현재 문자를 텍스트 스트림의 다음 문자로 설정하여 "나머지" 임시 문자열에 문자를 수집합니다. 그런 다음 1 단계로 돌아갑니다.
 - "나머지" 임시 문자열은 다른 정규 표현식 중 하나가 매칭될 때까지 한 번에 한 문자씩 취합됩니다.
 - "나머지" 정규 표현식의 기본 결과는 취합된 "나머지" 임시 문자열을 삭제하는 것입니다.

팁:

- "나머지" 정규 표현식의 결과를 \1로 설정하면 구두점 표시와 같은 구분 기호 문자를 더 많이 추가할 수 있습니다. 이렇게 하면 지정된 구두점 표시가 결과에 포함되지 않습니다.
- "나머지" 정규 표현식의 결과를 \1로 변경하는 대신 다음 중 하나 이상의 작업을 통해 관심 있는 용어를 캡처할 수 있습니다.
 - 정규 표현식 라이브러리에서 더 많은 정규 표현식을 추가합니다.
 - 사용자 정규 표현식을 생성합니다.

데이터 테이블의 각 행마다 텍스트 문자열의 끝에 도달할 때까지 위의 단계대로 처리가 진행됩니다.

결과를 데이터 테이블에 열로 저장

정규 표현식 토큰화 결과가 포함된 새 열을 데이터 테이블에 저장하려면 **열에 저장** 버튼을 클릭합니다. 새 열은 텍스트 탐색기 시작 창에서 지정된 텍스트 열과 이름이 동일한 문자 열입니다. 단, 열 이름이 고유하도록 이름에 번호가 추가됩니다.

참고: 사용자 정규 표현식 토큰화의 결과를 데이터 테이블의 열에 저장하면 데이터 테이블의 각 행에 있는 원래 텍스트에 대해 정규 표현식 프로세스가 실행됩니다. 소문자로 변환된 텍스트 문자열에는 이 프로세스가 실행되지 않습니다.

텍스트 탐색기 정규 표현식 편집기 닫기

텍스트 탐색기 정규 표현식 편집기 창에서 **확인**을 클릭하면 다음과 같은 작업이 수행됩니다.

1. 텍스트 탐색기 정규 표현식 편집기 창에서 정의된 사용자 정규 표현식이 정규 표현식 라이브러리에 저장됩니다.

주의 : 사용자 정규 표현식이 있는 경우 사용자 정규 표현식 라이브러리는 사용자가 **확인**을 클릭할 때만 저장됩니다. 가장 최근에 저장된 정규 표현식은 다음 번에 사용할 수 있습니다. 고유한 이름을 사용하여 정규 표현식 라이브러리에 추가 정규 표현식을 보관하십시오. 나중에 정규 표현식을 사용할 수 있도록 텍스트 탐색기 보고서 창에서 스크립트를 저장할 수 있습니다.

2. 텍스트 탐색기 보고서가 나타납니다. 이 보고서에서는 지정된 정규 표현식 설정을 사용하여 텍스트를 토큰화한 결과를 보여 줍니다.

텍스트 탐색기 보고서

텍스트 탐색기 보고서 창에는 요약 개수 보고서와 "용어 및 구 목록" 보고서가 포함됩니다.

그림 12.8 텍스트 탐색기 보고서의 예



요약 개수 보고서

텍스트 탐색기 보고서의 첫 번째 테이블에는 다음과 같은 요약 통계량이 포함됩니다.

용어 수 용어 목록의 용어 수입니다.

사례 수 코퍼스의 문서 수입니다.

총 토큰 수 코퍼스의 총 용어 수입니다.

사례별 토큰 수 토큰 수를 사례 수로 나눈 것입니다.

비어 있지 않은 사례 수 한 개 이상의 용어가 포함된 코퍼스 내 문서의 개수입니다.

비어 있지 않은 사례의 비율 한 개 이상의 용어가 포함된 코퍼스 내 문서의 비율입니다.

용어 및 구 목록

"용어 및 구 목록" 보고서에는 토큰화 수행 후 텍스트에 있는 용어 및 구의 테이블이 포함됩니다. "용어 및 구 목록" 보고서의 예는 그림 12.8 에서 확인하십시오. 용어 목록의 "개수" 열은 코퍼스에서 각 용어가 나타난 횟수를 나타냅니다. 구 목록의 "개수" 열은 코퍼스에서 각 구가 나타난 횟수를 나타냅니다. "N" 열은 구 내의 단어 수를 나타냅니다.

기본적으로 용어 목록은 개수를 기준으로 내림차순 정렬되며, 개수가 동일한 용어들은 사전순으로 정렬됩니다. 구 목록은 개수를 기준으로 내림차순 정렬되며, 개수가 동일한 구들은 길이("N")를 기준으로 내림차순 정렬됩니다. 구 목록에서 길이도 동일한 구들은 사전순으로 정렬됩니다. 각 목록의 옵션을 사용하여 각 목록의 정렬 순서를 사전순으로 변경할 수 있습니다.

구 목록에 나타나는 구는 시작 창에 있는 **구당 최대 단어 수** 및 **최대 구 수** 옵션의 설정에 따라 결정됩니다. 데이터 테이블에서 한 번만 나타나는 구는 구 목록에 표시되지 않습니다.

구는 다양한 범위에서 용어로 지정될 수 있습니다. 구 목록에서 용어로 지정된 구는 구의 규격 범위를 기준으로 색상이 지정됩니다 (표 12.1). 다양한 범위에서 구를 지정하는 방법에 대한 자세한 내용은 "[용어 옵션 관리 창](#)"(373 페이지)에서 확인하십시오.

표 12.1 지정된 구의 색상

범위	색상
기본 제공	빨간색
사용자 라이브러리	녹색
프로젝트	파란색
열 특성	주황색
지역	회색

용어 및 구에 대한 작업

각 테이블의 맨 왼쪽 열에서 항목을 선택한 다음 마우스 오른쪽 버튼을 클릭하여 용어 목록 및 구 목록 테이블의 옵션에 액세스할 수 있습니다. 각 테이블의 "개수" 열을 마우스 오른쪽 버튼으로 클릭하고 "데이터 테이블로 만들기"를 선택하여 각 테이블을 데이터 테이블로 저장할 수 있습니다.

용어 목록 팝업 메뉴 옵션

용어 목록 테이블의 "용어" 열에서 마우스 오른쪽 버튼을 클릭하면 다음 옵션이 포함된 팝업 메뉴가 나타납니다.

행 선택하기 데이터 테이블에서 선택된 용어가 포함된 행을 선택합니다.

텍스트 표시 선택된 용어가 포함된 문서를 표시합니다.

참고: 기본적으로 처음 10,000 개의 문서만 표시됩니다. 선택된 용어가 포함된 문서가 10,000 개를 초과하면 이 한계를 늘릴 수 있는 창이 나타납니다.

사전순 용어 목록의 정렬 순서를 사전순과 개수 기준 내림차순 사이에서 전환합니다.

복사 선택된 용어를 클립보드에 추가합니다.

색상 선택된 용어에 색상을 할당할 수 있습니다.

라벨 선택된 용어에 대한 용어 SVD 그림의 해당 점에 라벨을 표시합니다.

포함하는 구 구 목록 테이블에서 선택된 용어가 포함된 구를 선택합니다.

표시자 저장 용어 목록에서 선택된 각 용어에 대한 표시자 열을 데이터 테이블에 저장합니다. 각 행의 표시자 열 값은 해당 행의 문서에 선택된 용어가 포함되어 있으면 1 이고, 그렇지 않으면 0 입니다.

계산식 저장 용어 목록에서 선택된 각 용어에 대한 열 계산식을 데이터 테이블에 저장합니다. 각 행의 열 계산식은 해당 행의 문서에 선택된 용어가 포함되어 있으면 1 로 계산되고, 그렇지 않으면 0 으로 계산됩니다. 이 옵션은 새 문서에 유용합니다.

재코딩 하나 이상의 용어에 대한 값을 변경할 수 있습니다. 이 옵션을 선택하기 전에 목록에서 용어를 선택해야 합니다. 이 옵션을 선택하면 " 재코딩 " 창이 나타납니다. 자세한 내용은 JMP 사용의 " 데이터 입력 및 편집 " 장에서 확인하십시오.

중지 단어 추가 선택된 용어를 중지 단어 목록에 추가하고 용어 목록에서 해당 용어를 제거합니다. 이 작업으로 구 목록도 업데이트됩니다.

어간 예외 추가 (" 언어 " 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 선택된 용어를 어간 추출 (Stemming) 에서 제외된 용어의 목록에 추가합니다.

구 제거 (지정된 구가 용어 목록에서 선택된 경우에만 사용 가능) 지정된 구 집합에서 선택된 구를 제거하고 그에 따라 용어 개수를 업데이트합니다.

필터 표시 용어 목록 위에 검색 필터를 표시하거나 숨깁니다. 자세한 내용은 "검색 필터 옵션"(369 페이지) 에서 확인하십시오.

데이터 테이블로 만들기 보고서 테이블을 사용하여 JMP 데이터 테이블을 생성합니다.

결합 데이터 테이블 생성 보고서에서 선택한 테이블과 유사한 다른 테이블을 검색하여 단일 JMP 데이터 테이블에 결합합니다.

구 목록 팝업 메뉴 옵션

구 목록 테이블의 " 구 " 열에서 마우스 오른쪽 버튼을 클릭하면 다음 옵션이 포함된 팝업 메뉴가 나타납니다.

행 선택하기 데이터 테이블에서 선택된 구가 포함된 행을 선택합니다.

텍스트 표시 선택된 구가 포함된 문서를 표시합니다.

표시자 저장 구 목록에서 선택된 각 구에 대한 표시자 열을 데이터 테이블에 저장합니다. 각 행의 표시자 열 값은 해당 행의 문서에 선택된 구가 포함되어 있으면 1 이고, 그렇지 않으면 0 입니다.

사전순 구 목록의 정렬 순서를 사전순과 개수 기준 내림차순 사이에서 전환합니다.

복사 선택된 구를 클립보드에 추가합니다.

포함하는 항목 선택 구 목록에서 선택된 구를 포함하는 더 큰 구를 선택합니다.

포함된 항목 선택 구 목록 또는 용어 목록에서 선택된 구에 포함된 더 작은 구 또는 용어를 선택합니다.

구 추가 선택된 구를 용어 목록에 추가하고 그에 따라 용어 개수를 업데이트합니다.

중지 단어 추가 선택된 구를 중지 단어 목록에 추가합니다. 이 작업으로 용어 목록도 업데이트됩니다.

필터 표시 구 목록 위에 검색 필터를 표시하거나 숨깁니다. 자세한 내용은 "[검색 필터 옵션](#)"(369 페이지)에서 확인하십시오.

데이터 테이블로 만들기 보고서 테이블을 사용하여 JMP 데이터 테이블을 생성합니다.

결합 데이터 테이블 생성 보고서에서 선택한 테이블과 유사한 다른 테이블을 검색하여 단일 JMP 데이터 테이블에 결합합니다.

검색 필터 옵션

검색 상자 옆의 아래쪽 화살표 버튼을 클릭하여 검색을 구체화할 수 있습니다.

용어 포함 검색 기준의 일부가 포함된 항목을 반환합니다. "ease oom" 을 검색하면 "Release Zoom" 과 같은 메시지가 반환됩니다.

구 포함 검색 기준이 정확히 포함된 항목을 반환합니다. "text box" 를 검색하면 "text" 와 "box" 가 바로 연이어 포함된 항목 (예 : "Context Box" 및 "Text Box") 이 반환됩니다.

구로 시작 검색 기준으로 시작하는 항목을 반환합니다.

구로 끝남 검색 기준으로 끝나는 항목을 반환합니다.

전체 구 전체 문자열로 구성된 항목을 반환합니다. "text box" 를 검색하면 "text box" 만 포함된 항목이 반환됩니다.

정규 표현식 검색 상자에서 와일드카드 (*) 와 마침표 (.) 를 사용할 수 있습니다. "get.*name" 을 검색하면 "get" 다음에 하나 이상의 단어가 포함된 항목을 찾을 수 있습니다. 즉, "Get Color Theme Names", "Get Name Info" 및 "Get Effect Names" 등이 반환됩니다.

결과 반전 검색 기준과 매칭되지 않는 항목을 반환합니다.

모든 용어 일치 문자열이 모두 포함된 항목을 반환합니다. "t test" 를 검색하면 검색 문자열 중 하나 또는 둘 모두가 포함된 요소 (예 : "Pat Test", "Shortest Edit Script" 및 "Paired t test") 가 반환됩니다.

대 / 소문자 무시 검색 기준의 대 / 소문자를 무시합니다.

전체 단어 일치 "모든 용어 일치" 설정에 따라 문자열의 각 단어가 포함된 항목을 반환합니다. "모든 용어 일치" 옵션이 선택되어 있는 경우 "data filter" 를 검색하면 "data" 와 "filter" 가 반환됩니다.

텍스트 탐색기 플랫폼 옵션

이 섹션에서는 텍스트 탐색기 플랫폼에서 사용할 수 있는 옵션에 대해 설명합니다.

- " [텍스트 준비 옵션](#) "
- " [텍스트 분석 옵션](#) "
- " [저장 옵션](#) "
- " [보고서 옵션](#) "

텍스트 준비 옵션

텍스트 탐색기의 빨간색 삼각형 메뉴에는 텍스트 준비를 위한 다음 옵션이 포함되어 있습니다.

표시 옵션 보고서 표시를 제어하는 옵션의 하위 메뉴를 표시합니다.

단어 클라우드 표시 " 단어 클라우드 " 보고서를 표시하거나 숨깁니다. " 단어 클라우드 "의 빨간색 삼각형 메뉴를 사용하여 단어 클라우드의 레이아웃 및 글꼴을 변경할 수 있습니다. 자세한 내용은 " [단어 클라우드 옵션](#) "(372 페이지) 에서 확인하십시오.

너비를 변경하는 방법으로 단어 클라우드의 크기를 대화식으로 조정할 수 있습니다. 이때 높이는 자동으로 결정됩니다. 용어 목록의 행은 단어 클라우드의 용어에 연결됩니다.

용어 목록 표시 용어 목록을 표시하거나 숨깁니다.

구 목록 표시 구 목록을 표시하거나 숨깁니다.

용어 및 구 옵션 표시 각 목록의 팝업 메뉴에서 사용할 수 있는 옵션에 해당하는 버튼을 " 용어 및 구 목록 " 보고서에 표시합니다. 자세한 내용은 " [용어 및 구 목록](#) "(367 페이지) 에서 확인하십시오.

요약 개수 표시 요약 개수 테이블을 표시하거나 숨깁니다. 자세한 내용은 " [요약 개수 보고서](#) "(366 페이지) 에서 확인하십시오.

중지 단어 표시 분석에 사용된 중지 단어의 목록을 표시하거나 숨깁니다. 처음에는 기본 제공 중지 단어 목록이 사용됩니다. 중지 단어를 추가하려면 용어 목록을 마우스 오른쪽 버튼으로 클릭한 후 팝업 메뉴에서 **중지 단어 추가**를 선택합니다. 자세한 내용은 " [용어 옵션 관리 창](#) "(373 페이지) 에서 확인하십시오.

재코딩 표시 재코딩된 용어의 목록을 표시하거나 숨깁니다. 자세한 내용은 " [용어 옵션 관리 창](#) "(373 페이지) 에서 확인하십시오.

지정한 구 표시 사용자가 용어로 처리되도록 지정한 구의 목록을 표시하거나 숨깁니다. 자세한 내용은 "[용어 옵션 관리 창](#)"(373 페이지)에서 확인하십시오.

어간 예외 표시 ("언어" 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출 (stemming)에서 제외된 용어를 표시하거나 숨깁니다. 자세한 내용은 "[용어 옵션 관리 창](#)"(373 페이지)에서 확인하십시오.

구분자 표시 ("언어" 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정되고 선택된 토큰화 방법이 "기본 단어"인 경우에만 사용 가능) 기본 단어 토큰화 방법에 사용된 구분자를 표시하거나 숨깁니다 사용된 구분자 집합을 수정하려면 JSL에서 Add Delimiters() 또는 Set Delimiters() 메시지를 사용해야 합니다.

어간 보고서 표시 ("언어" 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정되고 선택된 어간 추출 방법이 "어간 추출 (Stemming) 안 함"인 경우에만 사용 가능) 두 개의 어간 추출 결과 테이블이 포함된 "어간 추출 (Stemming)" 보고서를 표시하거나 숨깁니다. 왼쪽의 테이블은 각 어간을 해당 용어에 매핑합니다. 오른쪽의 테이블은 각 용어를 해당 어간에 매핑합니다.

선택된 행 표시 현재 선택된 행에 있는 문서의 텍스트가 포함된 창을 엽니다.

모든 테이블에 대해 필터 표시 보고서의 테이블을 검색하는 데 사용할 수 있는 필터를 표시하거나 숨깁니다. 이 옵션은 중지 단어, 지정된 구, 어간 예외, 용어 목록, 구 목록 및 어간 보고서에 적용됩니다. 필터 도구에 대한 자세한 내용은 "[검색 필터 옵션](#)"(369 페이지)에서 확인하십시오.

용어 옵션 용어 목록에 적용되는 옵션의 하위 메뉴를 표시합니다.

어간 추출 (Stemming) ("언어" 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출 옵션에 대한 자세한 내용은 "[텍스트 탐색기 플랫폼 시작](#)"(358 페이지)에서 확인하십시오.

기본 제공 중지 단어 포함 토큰화 프로세스에 사용되는 중지 단어에 기본 제공 중지 단어를 포함할지 여부를 지정합니다.

기본 제공 구 포함 토큰화 프로세스에 사용되는 구에 기본 제공 구를 포함할지 여부를 지정합니다.

중지 단어 관리 중지 단어를 추가하거나 제거할 수 있는 창을 표시합니다. 변경 사항은 사용자, 열 및 지역 수준에서 적용될 수 있습니다. 다른 수준에 지정된 중지 단어를 제외하는 지역 예외를 지정할 수도 있습니다. 자세한 내용은 "[용어 옵션 관리 창](#)"(373 페이지)에서 확인하십시오.

재코딩 관리 재코딩을 추가하거나 제거할 수 있는 창을 표시합니다. 변경 사항은 사용자, 열 및 지역 수준에서 적용될 수 있습니다. 다른 수준에 지정된 재코딩을 제외하는 지역 예외를 지정할 수도 있습니다. 자세한 내용은 "[용어 옵션 관리 창](#)"(373 페이지)에서 확인하십시오.

구 관리 용어로 처리되는 구를 추가하거나 제거할 수 있는 창을 표시합니다. 변경 사항은 사용자, 열 및 지역 수준에서 적용될 수 있습니다. 다른 수준에 지정된 구를 제외하는 지역 예외를 지정할 수도 있습니다. 자세한 내용은 "[용어 옵션 관리 창](#)"(373 페이지)에서 확인하십시오.

어간 예외 관리 ("언어" 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출 예외를 추가하거나 제거할 수 있는 창을 표시합니다. 변경 사항은 사용자, 열 및 지역 수준에서 적용될 수 있습니다. 다른 수준에 지정된 어간 예외를 제외하는 지역 예외를 지정할 수도 있습니다. 자세한 내용은 "[용어 옵션 관리 창](#)"(373 페이지)에서 확인하십시오.

파싱 옵션 파싱 및 토큰화에 적용되는 옵션의 하위 메뉴를 표시합니다.

토큰화 ("언어" 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 토큰화 옵션에 대한 자세한 내용은 "[텍스트 탐색기 플랫폼 시작](#)"(358 페이지)에서 확인하십시오.

정규 표현식 사용자 정의 (정규 표현식 토큰화 방법을 사용하는 경우에만 사용 가능) "정규 표현식 사용자 정의" 창을 표시합니다. 이 옵션을 사용하여 현재 텍스트 탐색기 보고서에 대한 정규 표현식 설정을 수정할 수 있습니다.

참고 : 플랫폼 시작 창에서 기준 변수를 지정한 경우에는 "정규 표현식 사용자 정의" 옵션이 기준 변수의 모든 수준에 일괄 적용됩니다.

숫자를 단어로 처리 ("언어" 옵션이 영어, 독일어, 스페인어, 프랑스어 또는 이탈리아어로 설정되고 선택된 토큰화 방법이 "기본 단어"인 경우에만 사용 가능) 분석에서 숫자를 용어로 토큰화할 수 있도록 합니다. 이 옵션은 단어당 최소 문자 수에 대한 설정에 따라 영향을 받습니다.

단어 클라우드 옵션

"단어 클라우드"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

레이아웃 단어 클라우드의 용어 배열 방식을 지정합니다. 기본 레이아웃은 "정렬됨"으로 설정되어 있습니다.

정렬됨 용어를 빈도가 높은 것부터 낮은 것 순서로 가로 줄로 표시합니다.

사전순 용어를 오름차순 사전순으로 정렬하여 가로 줄로 표시합니다.

중심화됨 용어를 빈도에 따른 크기로 클라우드에 표시합니다.

색상 적용 단어 클라우드의 용어 색상을 지정합니다. 기본 색상은 "없음"으로 설정되어 있습니다.

없음 각 용어를 용어 목록의 색상과 동일한 색상으로 표시합니다.

균등 색상 각 용어를 동일한 색상으로 표시합니다. 범례에서 이 색상을 변경할 수 있습니다.

임의 회색 각 용어를 다양한 회색 음영으로 표시합니다.

임의 색상 각 용어를 다양한 색상으로 표시합니다. 범례에서 색상을 조정할 수 있습니다.

열 값별 각 용어를 그래디언트 색상 척도로 표시합니다. 척도는 "열을 기준으로 용어 점수 매기기" 옵션으로 생성된 용어 점수를 기준으로 합니다. 범례에서 색상과 그래디언트를 조정할 수 있습니다.

글꼴 단어 클라우드의 용어에 적용할 글꼴, 스타일 및 크기를 지정합니다.

범례 표시 단어 클라우드의 범례를 표시하거나 숨깁니다.

용어 옵션 관리 창

다양한 범위에 대해 구, 중지 단어, 재코딩 및 어간 예외 정보를 지정할 수 있습니다. 이러한 정보는 텍스트 탐색기 사용자 라이브러리 (사용자 범위), 현재 프로젝트, 분석 열에 대한 열 특성 (열 범위) 또는 플랫폼 스크립트 (지역 범위)에 저장할 수 있습니다. 텍스트 탐색기 보고서의 스크립트를 저장하여 특정 텍스트 탐색기 인스턴스에 대한 지역 규격 및 지역 예외를 저장할 수 있습니다.

용어 옵션 관리 창은 중지 단어, 재코딩, 구 및 어간 예외의 모음을 관리할 수 있는 네 가지 유사한 창입니다. 그림 12.9에서는 "중지 단어 관리" 창을 보여 줍니다. "구 관리" 및 "어간 예외 관리" 창은 "중지 단어 관리" 창과 동일합니다. "재코딩 관리" 창은 약간 다릅니다. 자세한 내용은 "[재코딩 관리](#)"(375 페이지)에서 확인하십시오.

그림 12.9 중지 단어 관리 창

The image shows the 'Stop Word Management' window. At the top, there is a dropdown menu for '언어' (Language) set to '영어' (English). Below this, the window is divided into five main sections: '기본 제공(잠금)' (Predefined (locked)), '사용자' (User), '열' (Column), '지역' (Region), and '지역 예외' (Region exception). The '기본 제공' section contains a list of common stop words like '&', 'a', 'about', 'above', 'after', 'again', 'against', 'all', 'am', 'an', 'and', 'any'. Each of the five sections has a '항목 추가' (Add item) button with a plus sign. On the right side, there are four buttons: '확인' (OK), '취소' (Cancel), '도움말' (Help), and '가져오기...' (Import...), followed by '내보내기...' (Export...). At the bottom, there are navigation buttons: '<<', '>>', and a close button 'X'.

중지 단어 관리

"중지 단어 관리" 창에는 지정된 중지 단어의 다양한 범위 (또는 위치)를 나타내는 중지 단어 목록이 여러 개 포함되어 있습니다. 각 목록 아래에는 텍스트 편집 상자와 추가 버튼이 있습니다. 이러한 컨트롤을 사용하여 각 범위에 사용자 중지 단어를 추가할 수 있습니다. 중지 단어를 드래그하여 한 범위에서 다른 범위로 이동할 수 있습니다. 한 목록에서 다른 목록으로 항목을 복사하여

붙여넣을 수도 있습니다. 창 아래에 있는 두 개의 버튼을 클릭하면 선택된 항목이 한 범위에서 왼쪽 또는 오른쪽의 다음 범위로 이동합니다. X 버튼을 클릭하면 선택된 항목이 현재 범위에서 제거됩니다. 또한 항목을 두 번 클릭하고 텍스트를 변경하여 목록의 기존 항목을 편집할 수 있습니다.

언어 기본 제공 중지 단어 목록과 사용자 라이브러리의 선택 항목을 저장할 언어를 지정합니다. "언어"에서 "항목 적용"을 선택하면 변경 사항이 마스터 사용자 라이브러리에 저장됩니다. 언어 설정은 기본 제공 범위와 사용자 및 프로젝트 범위에만 적용됩니다.

기본 제공 (잠김) 지정된 언어의 중지 단어에 대한 기본 제공 목록입니다. 기본 제공 중지 단어를 "지역 예외" 목록에 포함하면 중지 단어 목록에서 제외할 수 있습니다.

사용자 지정된 언어의 사용자 라이브러리에 있는 중지 단어를 나열합니다.

프로젝트 (텍스트 탐색기가 "TextExplorer"라는 폴더가 포함된 프로젝트 내에서 시작된 경우에만 사용 가능) 지정된 언어의 현재 프로젝트에 있는 중지 단어를 나열합니다.

열 텍스트 열의 "중지 단어" 열 특성에 지정된 중지 단어를 나열합니다.

지역 지역 범위의 중지 단어를 나열합니다. JSL 을 통해 텍스트 탐색기를 시작한 경우에 이러한 중지 단어를 지정할 수 있습니다. 이러한 중지 단어는 현재 텍스트 탐색기 플랫폼 보고서에서만 사용됩니다.

지역 예외 현재 텍스트 탐색기 플랫폼에서 중지 단어로 처리되지 않는 단어를 나열합니다. JSL 을 통해 텍스트 탐색기를 시작한 경우에 이러한 중지 단어를 지정할 수 있습니다. "지역 예외"에 나열된 단어는 다른 모든 범위에 나열된 단어를 재정의합니다.

가져오기 텍스트 파일에서 중지 단어를 가져올 수 있습니다. 중지 단어는 클립보드에 복사됩니다. 이를 기본 제공 목록 이외의 목록에 붙여넣을 수 있습니다.

내보내기 중지 단어를 클립보드 또는 텍스트 파일로 내보낼 수 있습니다. 중지 단어를 내보낼 범위와 내보낼 위치를 선택할 수 있는 "내보내기" 창이 나타납니다.

사용자 라이브러리 파일은 TextExplorer 디렉터리에 있습니다. 이 디렉터리의 위치는 다음과 같이 컴퓨터 운영 체제에 따라 다릅니다.

- Windows: "C:/Users/< 사용자 이름 >/AppData/Roaming/SAS/JMP/TextExplorer/< 언어 >/"
- macOS: "/Users/< 사용자 이름 >/Library/Application Support/JMP/TextExplorer/< 언어 >/"

마스터 사용자 라이브러리 파일은 TextExplorer 디렉터리 자체에 있습니다. 이러한 파일은 언어 별로 다르지 않습니다.

프로젝트 파일은 프로젝트의 TextExplorer 폴더에 있습니다.

확인을 클릭하면 "사용자", "프로젝트" 및 "열" 목록의 변경 사항이 각각 사용자 라이브러리, 프로젝트 및 열 특성에 저장됩니다. "지역" 및 "지역 예외" 목록에 지정된 모든 항목은 텍스트 탐색기 보고서 스크립트를 저장할 때만 저장됩니다.

중지 단어를 사용자 라이브러리에 저장할 경우 파일 이름은 stopwords.txt 로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "중지 단어"라고 합니다.

재코딩 관리

" 재코딩 관리 " 창은 " 중지 단어 관리 " 창과는 약간 다릅니다 . 각 목록 아래에는 텍스트 편집 상자가 하나가 아니라 두 개 있습니다 . 이전 값 (위쪽 상자에서 지정) 은 새 값 (아래쪽 상자에서 지정) 으로 재코딩됩니다 .

사용자 라이브러리에 재코딩을 저장하는 경우 파일 이름은 `recodes.txt` 로 지정됩니다 . 열 특성에 저장할 경우에는 해당 특성을 " 재코딩 " 이라고 합니다 .

구 관리

사용자 라이브러리에 구를 저장하는 경우 파일 이름은 `phrases.txt` 로 지정됩니다 . 열 특성에 저장할 경우에는 해당 특성을 " 구 " 라고 합니다 .

어간 예외 관리

사용자 라이브러리에 어간 예외를 저장하는 경우 파일 이름은 `stemExceptions.txt` 로 지정됩니다 . 열 특성에 저장할 경우에는 해당 특성을 " 어간 예외 " 라고 합니다 .

참고 : " 어간 예외 관리 " 창의 " 로컬 예외 " 목록에는 어간 예외 목록에서 제외되는 어간 예외가 나열됩니다 . 이 목록에 있는 단어는 어간 추출 작업에 관련되지 않습니다 .

텍스트 분석 옵션

텍스트 탐색기의 빨간색 삼각형 메뉴에는 분석을 위한 다음 옵션이 포함되어 있습니다 .

잠재 계층 분석 최소 행렬 루틴을 사용하여 이진 가중 문서 용어 행렬에 대해 잠재 계층 분석을 수행합니다 . 자세한 내용은 " **잠재 계층 분석** "(378 페이지) 에서 확인하십시오 .

텍스트 탐색기의 빨간색 삼각형 메뉴에서 " 잠재 계층 분석 " 을 선택하면 다음 옵션이 포함된 규격 창이 나타납니다 .

최대 용어 수 잠재 계층 분석에 포함되는 용어의 최대 개수입니다 .

최소 용어 빈도 한 용어가 잠재 계층 분석에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다 .

군집 수 잠재 계층 분석의 군집 수입니다 .

잠재 의미 분석, SVD 문서 용어 행렬의 부분 특이값 분해를 수행합니다 . 자세한 내용은 " **잠재 의미 분석 (SVD)** "(380 페이지) 에서 확인하십시오 .

판별 분석 문서 용어 행렬을 기준으로 그룹 또는 범주에서 각 문서의 소속을 예측합니다 . 자세한 내용은 " **판별 분석** "(385 페이지) 에서 확인하십시오 .

JMP PRO 특이값 분해 규격 창

텍스트 탐색기의 분석 옵션은 DTM(문서 용어 행렬) 을 기반으로 합니다 . DTM 은 용어 목록에 있는 각 용어마다 열을 하나씩 생성함으로써 구성됩니다 (지정된 최대 용어 수까지) . 각 텍스트 문서 (데이터 테이블의 행과 동등) 는 DTM 의 한 행에 해당합니다 . DTM 의 셀 값은 " 규격 " 창에서 사용자가 지정한 가중 유형에 따라 달라집니다 .

그림 12.10 에서는 특이값 분해를 위한 " 규격 " 창을 보여 줍니다 . 텍스트 탐색기의 빨간색 삼각형 메뉴에서 문서 용어 행렬에 대해 특이값 분해를 수행하는 옵션을 선택하면 다음 옵션이 포함된 " 규격 " 창이 나타납니다 .

최대 용어 수 특이값 분해에 포함되는 용어의 최대 개수입니다 .

최소 용어 빈도 한 용어가 특이값 분해에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다 .

가중치 문서 용어 행렬의 셀에 들어갈 값을 결정하는 가중치 체계입니다 . 가중치 체계 옵션에 대한 자세한 내용은 " 문서 용어 행렬 규격 창 "(377 페이지) 에서 확인하십시오 .

특이 벡터 수 특이값 분해의 특이 벡터 수입니다 . 기본값은 문서 수 , 용어 수 또는 100 중 최소값입니다 .

중심화 및 척도화 문서 용어 행렬을 중심화하고 척도화하기 위한 옵션입니다 . **중심화 및 척도화** , **중심화** 및 **비중심화** 중에서 선택할 수 있습니다 . 기본적으로 문서 용어 행렬은 중심화되고 척도화됩니다 .

그림 12.10 SVD 규격 창

저장 옵션

텍스트 탐색기의 빨간색 삼각형 메뉴에는 데이터 테이블 , 테이블 열 및 열 특성에 정보를 저장하기 위한 다음 옵션이 포함되어 있습니다 .

문서 용어 행렬 저장 문서 용어 행렬의 각 열에 해당하는 열을 데이터 테이블에 저장합니다 (지정된 최대 용어 수까지) .



연관성을 위해 쌓인 형식의 DTM 저장 쌓인 형식의 문서 용어 행렬을 JMP 데이터 테이블에 저장합니다 . 쌓인 형식은 연관성 분석 플랫폼의 분석에 적합합니다 . 자세한 내용은 Predictive and Specialized Modeling 의 "Association Analysis" 장에서 확인하십시오 . 텍스트

트 탐색기 시작 창에서 ID 변수를 지정하면 원래 텍스트 데이터 테이블에서 각 용어가 있던 행을 식별하는 데 ID 변수가 사용됩니다. 쌓인 테이블에는 연관성 분석을 시작하기 위한 테이블 스크립트도 포함됩니다.

DTM 계산식 저장 벡터 모델링 유형의 계산식 열을 데이터 테이블에 저장합니다. 벡터의 길이는 최대 용어 수, 최소 용어 빈도 및 가중치에 대해 사용자가 지정한 옵션에 따라 달라집니다. 결과 열에는 Text Score() JSL 함수가 사용됩니다. 이 함수에 대한 자세한 내용은 "도움말">"스크립트 인덱스"에서 확인하십시오.

용어 테이블 저장 용어 목록의 각 용어, 발생 횟수 및 각 용어가 포함된 문서 수가 포함된 JMP 데이터 테이블을 생성합니다. "용어 테이블 저장"을 선택한 후 "열을 기준으로 용어 점수 매기기" 옵션을 선택하면 용어별 점수가 포함된 열이 "용어 테이블 저장" 옵션으로 생성한 데이터 테이블에 추가됩니다.

열을 기준으로 용어 점수 매기기 지정된 열의 값을 기준으로 한 점수를 "용어 테이블 저장" 옵션으로 생성된 JMP 데이터 테이블에 저장합니다. 각 용어의 점수는 지정된 열에 대한 평균 값으로, 이 값은 각 행에서 해당 용어가 나타나는 횟수를 기준으로 가중치를 부여한 값입니다. "용어 테이블 저장" 옵션을 이미 선택한 경우 "열을 기준으로 용어 점수 매기기" 옵션은 "용어 테이블 저장" 옵션으로 생성된 데이터 테이블에 점수를 포함하는 열을 추가합니다. 그렇지 않은 경우에는 용어 테이블에 대한 JMP 데이터 테이블이 생성됩니다. 지정된 열이 연속형이 아니면 지정된 열의 각 수준에 대한 점수를 포함하는 열이 생성됩니다.

문서 용어 행렬 규격 창

텍스트 탐색기의 빨간색 삼각형 메뉴에서 "문서 용어 행렬 저장" 및 "DTM 계산식 저장" 옵션을 선택하면 다음 옵션이 포함된 문서 용어 행렬 "규격" 창이 나타납니다.

최대 용어 수 문서 용어 행렬에 포함되는 용어의 최대 개수입니다.

최소 용어 빈도 한 용어가 문서 용어 행렬에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다.

가중치 문서 용어 행렬의 셀에 들어갈 값을 결정하는 가중치 체계입니다.

"가중치"에는 다음 옵션을 사용할 수 있습니다.

이진 각 문서에서 용어가 나오면 1을 할당하고, 그렇지 않으면 0을 할당합니다. SVD 분석이 이전에 실행되었던 경우를 제외하고는 이 가중 방법이 기본값입니다.

삼진 각 문서에서 용어가 두 번 이상 나오면 2를 할당하고, 한 번만 나오면 1을 할당하며, 나오지 않으면 0을 할당합니다.

빈도 각 문서에서의 용어 발생 횟수를 할당합니다.

로그 빈도 $\log_{10}(1+x)$ 를 할당합니다. 여기서 x 는 각 문서에서의 용어 발생 횟수입니다.

TF IDF $TF * \log_{10}(nDoc / nDocTerm)$ 을 할당합니다. 용어 빈도 - 문서 빈도 역수 (term frequency - inverse document frequency)의 약어입니다. 이 옵션이 기본 가중치 옵션입니다. 계산식의 항은 다음과 같이 정의됩니다.

TF = 문서 내 용어 빈도

$nDoc$ = 코퍼스 내 문서 수

$nDocTerm$ = 해당 용어를 포함하는 문서의 수

참고 : SVD 분석을 실행한 후 " 문서 용어 행렬 저장 " 또는 "DTM 계산식 저장 " 을 선택하면 " 규격 " 창에 가장 최근의 SVD 분석 규격이 포함됩니다.

보고서 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 "JMP 보고서 " 장에서 확인하십시오 .

로컬 데이터 필터 특정 보고서에서 사용되는 데이터를 필터링할 수 있는 로컬 데이터 필터를 표시하거나 숨깁니다 .

다시 실행 분석을 반복하거나 다시 시작할 수 있는 옵션이 포함되어 있습니다 . 이 기능을 지원하는 플랫폼에서 " 자동 재계산 " 옵션은 해당하는 보고서 창에서 데이터 테이블에 대한 변경 사항을 즉시 반영합니다 .

스크립트 저장 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 .

그룹별 스크립트 저장 기준 변수의 모든 수준에 대한 플랫폼 보고서를 재생성하는 스크립트를 여러 대상에 저장할 수 있는 옵션이 포함되어 있습니다 . 시작 창에서 기준 변수를 지정한 경우에만 사용할 수 있습니다 .

JMP PRO 잠재 계층 분석

잠재 계층 분석을 통해 코퍼스의 문서를 유사 문서의 군집으로 그룹화할 수 있습니다 . 잠재 계층 분석 보고서에는 모형 규격 , 모형에 대한 BIC(베이지안 정보 기준) 값 , 그리고 " 텍스트 표시 " 버튼이 포함됩니다 . " 군집 혼합 확률 " 테이블에서 하나 이상의 군집을 선택하고 " 텍스트 표시 " 버튼을 클릭하면 선택한 군집에 속할 가능성이 가장 높은 문서의 텍스트가 포함된 창이 열립니다 .

" 잠재 계층 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

표시 옵션 " 잠재 계층 분석 " 보고서의 내용을 지정합니다 . 기본적으로는 각 군집의 단어 클라우드를 제외한 모든 보고서 옵션이 표시됩니다 .

군집 혼합 확률 관측값이 각 군집에 속할 확률을 보여 주는 테이블을 표시하거나 숨깁니다 .

팁 : " 군집별 혼합 확률 " 테이블에서 하나 이상의 행을 선택하여 해당 군집에 할당된 관측값을 선택할 수 있습니다 .

군집별 용어 확률 문서가 특정 군집에 속할 경우 군집별로 문서에 용어가 포함될 조건부 확률의 추정값이 들어 있는 용어 테이블을 표시하거나 숨깁니다. 기본적으로 이 테이블의 용어는 코퍼스에서의 빈도를 기준으로 내림차순 정렬됩니다.

"가장 특징적인 군집" 열에는 각 용어가 가장 높은 비율로 발생하는 군집이 표시됩니다.

"가장 확률이 높은 군집" 열에는 각 용어가 포함된 문서 중에서 무작위로 선택된 문서가 발견될 확률이 가장 높은 군집이 표시됩니다.

군집별 상위 용어 각 군집에서 점수가 가장 높은 10 개 용어의 테이블을 표시하거나 숨깁니다. 군집 c 의 용어 t 에 대한 점수 $S_{t,c}$ 는 다음과 같이 계산됩니다.

$$S_{t,c} = 100 \bullet \text{mean}(p_t) \bullet \log_{10} \left(\frac{p_{t,c}}{\text{mean}(p_t)} \right)$$

여기서 $\text{mean}(p_t)$ 는 용어 t 의 군집별 평균 용어 확률이고, $p_{t,c}$ 는 군집 c 의 용어 t 에 대한 군집별 용어 확률입니다.

MDS 그림 군집의 근접성을 2 차원으로 표현한 다차원 척도 그림을 표시하거나 숨깁니다. MDS 그림에 대한 자세한 내용은 **Multivariate Methods** 의 "Multidimensional Scaling" 장에서 확인하십시오. "텍스트 표시" 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다.

행별 군집 확률 각 행의 군집 소속 확률을 표시하는 혼합 확률 테이블을 표시하거나 숨깁니다. "가장 가능성이 큰 군집" 은 각 행의 소속 확률이 가장 높은 군집을 나타냅니다.

군집별 단어 클라우드 단어 클라우드를 군집당 하나씩 포함하는 행렬을 표시하거나 숨깁니다.

군집 이름 바꾸기 하나 이상의 군집에 대한 이름을 추가할 수 있습니다.

확률 저장 "혼합 확률" 테이블의 값을 데이터 테이블의 해당 행에 저장합니다.

확률 계산식 저장 각 군집에 대한 계산식 열과 가장 가능성이 높은 군집에 대한 계산식 열을 데이터 테이블에 저장합니다.

저장되는 점수 계산식에서는 가중치 인수를 "LCA" 로 설정하고 `Text Score()` JSL 함수를 사용합니다.

군집별 색상 데이터 테이블의 각 행을 가장 가능성이 큰 군집에 따라 색상을 적용합니다.

제거 텍스트 탐색기 보고서에서 "잠재 계층 분석" 보고서를 제거합니다.

잠재 계층 분석에 대한 자세한 내용은 **Multivariate Methods** 의 "Latent Class Analysis" 장에서 확인하십시오.

참고: 텍스트 탐색기 플랫폼에서 사용되는 LCA 알고리즘은 문서 용어 행렬의 희소성을 활용합니다. 이러한 이유로 텍스트 탐색기 플랫폼의 LCA 결과는 잠재 계층 분석 플랫폼의 결과와 정확하게 매칭되지 않습니다.

JMP^{PRO} 잠재 의미 분석 (SVD)

잠재 의미 분석은 DTM(문서 용어 행렬) 의 부분 SVD(특이값 분해) 를 계산하는 것을 중심으로 이루어집니다 . 이 분해 방법에서는 분석을 위해 텍스트 데이터를 관리 가능한 차원 수로 축소합니다 . 잠재 의미 분석은 PCA(주성분 분석) 를 수행하는 것과 동등합니다 .

부분 특이값 분해는 세 개의 행렬 \mathbf{U} , \mathbf{S} 및 \mathbf{V}' 를 사용하여 DTM 에 근사한 값을 산출합니다 . 이러한 행렬 간의 관계는 다음과 같이 정의됩니다 .

$$DTM \approx \mathbf{U} * \mathbf{S} * \mathbf{V}'$$

$nDoc$ 는 DTM의 문서(행) 수로 정의되고, $nTerm$ 은 DTM의 용어(열) 수로 정의되며, $nVec$ 은 지정된 특이값 수로 정의됩니다 . $nVec$ 은 $\min(nDoc, nTerm)$ 보다 작거나 같아야 합니다 . 따라서 \mathbf{U} 는 DTM 의 왼쪽 특이 벡터를 포함하는 $nDoc \times nVec$ 행렬입니다 . \mathbf{S} 는 $nVec$ 차 주대각 행렬입니다 . \mathbf{S} 의 주대각 항목은 DTM의 특이값입니다 . \mathbf{V}' 는 $nVec \times nTerm$ 행렬입니다 . \mathbf{V}' 의 행(또는 \mathbf{V} 의 열) 은 오른쪽 특이 벡터입니다 .

오른쪽 특이 벡터는 의미 또는 주제 영역이 유사한 서로 다른 용어 간의 연결성을 포착합니다 . 세 개의 용어가 동일한 문서에서 나타나는 경향이 있는 경우 SVD에서는 \mathbf{V}' 에 이 세 개의 용어에 대해 큰 값을 갖는 특이 벡터를 생성할 가능성이 높습니다 . \mathbf{U} 특이 벡터는 이 새로운 용어 공간에 투영된 문서를 나타냅니다 .

잠재 의미 분석에서는 간접적 연결성도 포착합니다 . 두 개의 단어가 동일한 문서에 함께 나타나는 않지만 일반적으로 다른 세 번째 단어가 있는 문서에서 나타나는 경우 SVD 는 해당 연결성의 일부를 포착할 수 있습니다 . 두 개의 문서에 공통된 단어는 없지만 차원 감소 공간에서 연결되는 단어가 포함된 경우 이 두 문서는 SVD 출력에서 유사 벡터에 매핑됩니다 .

SVD 는 텍스트 데이터를 고정 차원 벡터 공간으로 변환하여 모든 유형의 군집화 , 분류 및 회귀 기법에 적용할 수 있도록 합니다 . 저장 옵션을 사용하면 이 벡터 공간을 다른 JMP 플랫폼에서 분석할 수 있는 형태로 내보낼 수 있습니다 .

기본적으로는 특이값 분해가 수행되기 전에 DTM 을 중심화하고 척도화한 후 $nDoc - 1$ 로 나눕니다 . 이 분석은 DTM 상관 행렬의 PCA 와 동등합니다 .

"규격" 창에서 "중심화" 또는 "비중심화" 를 지정할 수도 있습니다 .

- "중심화" 를 지정할 경우 특이값 분해 전에 DTM 이 중심화된 후 $nDoc - 1$ 로 나눕니다 . 이 분석은 DTM 공분산 행렬의 PCA 와 동등합니다 .
- "비중심화" 를 지정할 경우에는 특이값 분해 전에 DTM 이 $nDoc$ 로 나눕니다 . 이 분석은 비척도화 DTM 의 PCA 와 동등합니다 .

SVD 구현에서는 DTM 이 중심화된 경우에도 DTM 의 회소성을 활용합니다 .

JMP^{PRO} SVD 보고서

잠재 의미 분석 옵션은 특이값 분해를 통해 두 개의 SVD 그림과 한 개의 특이값 테이블을 생성합니다.

JMP^{PRO} SVD 그림

첫 번째 그림에는 각 문서에 대한 점이 포함되어 있습니다. 특정 문서에 대해 표시되는 점은 처음 두 개의 특이 벡터 (**U** 행렬의 처음 두 열)에 있는 문서 값을 주대각 특이값 행렬 (**S**)로 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 점수 그림과 동등합니다. 이 그림의 각 점은 문서 (데이터 테이블의 행)를 나타냅니다. 이 그림에서 점을 선택하여 데이터 테이블의 해당 행을 선택할 수 있습니다.

두 번째 그림에는 각 용어에 대한 점이 포함되어 있습니다. 특정 용어에 대해 표시되는 점은 처음 두 개의 특이 벡터 (**V'** 행렬의 처음 두 행)에 있는 용어 값을 주대각 특이값 행렬 (**S**)로 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 적재 그림과 동등합니다. 이 그림의 점은 용어 목록 테이블의 행에 해당합니다.

각 SVD 그림 위에 있는 "텍스트 표시" 버튼을 클릭하면 그림에서 선택된 점의 텍스트가 포함된 창을 열 수 있습니다.

JMP^{PRO} 특이값

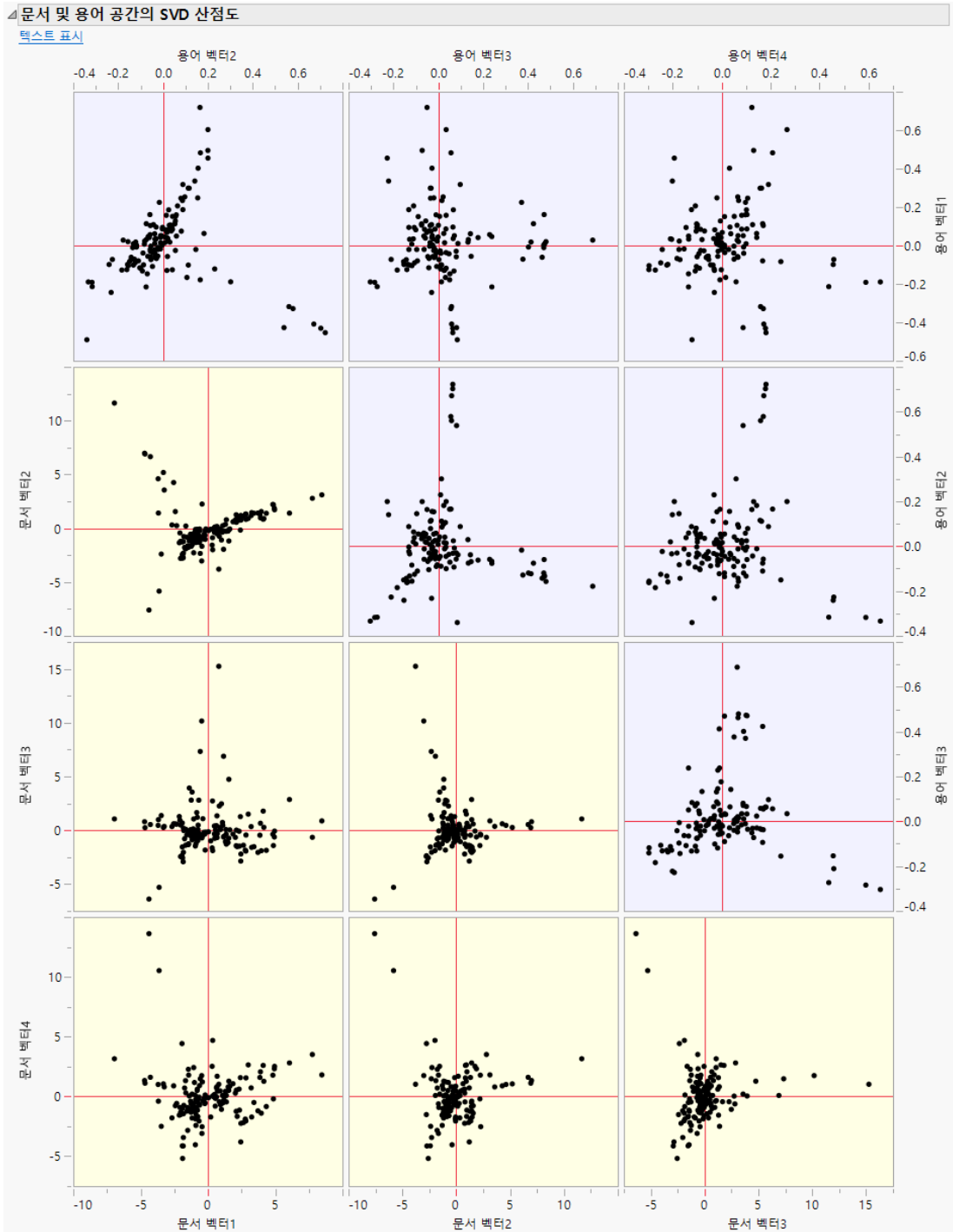
문서 및 용어 SVD 그림 아래에는 특이값 테이블이 표시됩니다. 이 특이값들은 문서 용어 행렬의 특이값 분해에서 **S** 행렬의 주대각 항목입니다. 특이값 테이블에는 동등한 주성분 분석의 해당 고유값이 들어 있는 열이 포함되어 있습니다. 주성분 플랫폼에서와 마찬가지로 각 고유값 (또는 특이값)이 설명하는 변동의 백분율 및 누적 백분율을 나타내는 열이 있습니다. "누적 백분율" 열을 사용하여 DTM에서 유지할 분산의 백분율을 결정한 다음 해당하는 수의 특이 벡터를 사용할 수 있습니다.

JMP^{PRO} SVD 보고서 옵션

SVD의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

SVD 산점도 행렬 용어 및 문서 특이값 분해 벡터에 대한 산점도 행렬을 표시하거나 숨깁니다. 이 옵션을 선택할 경우 산점도 행렬의 크기를 선택해야 합니다. 이 산점도 행렬을 사용하면 특이값 분해의 처음 두 차원보다 더 많은 차원을 시각화할 수 있습니다. "텍스트 표시" 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다.

그림 12.11 문서 및 용어 공간의 SVD 산점도



주제 분석, 회전된 SVD 문서 용어 행렬의 Varimax 회전 부분 특이값 분해를 수행하여 용어 그룹, 즉 주제를 추출합니다. 이 옵션을 여러 번 선택하여 서로 다른 개수의 주제를 찾을 수 있습니다. 자세한 내용은 "[주제 분석](#)"(383 페이지)에서 확인하십시오.

군집 용어 데이터에 있는 용어의 계층적 군집화 분석을 표시하거나 숨깁니다. 덴드로그램 오른쪽에는 군집 수를 설정하고 군집을 데이터 테이블에 저장하기 위한 옵션이 있습니다. 이 데이터 테이블에는 각 용어의 빈도, 해당 용어가 포함된 문서의 개수 및 할당된 군집이 포함됩니다. 계층적 군집화 및 덴드로그램에 대한 자세한 내용은 **Multivariate Methods** 의 "Hierarchical Cluster" 장에서 확인하십시오.

군집 문서 데이터에 있는 문서의 계층적 군집화 분석을 표시하거나 숨깁니다. 덴드로그램 오른쪽에는 군집 수를 설정하고, 군집을 데이터 테이블의 열에 저장하고, 덴드로그램 그림의 선택된 분기에 문서를 표시하기 위한 옵션이 있습니다.

문서 특이 벡터 저장 문서 특이값 분해에서 구한 특이 벡터 중 사용자가 지정한 수만큼을 데이터 테이블에 열로 저장합니다. 처음 두 개의 저장된 열은 문서 SVD 그림에 표시된 점을 나타냅니다. 자세한 내용은 "[잠재 의미 분석 \(SVD\)](#)"(380 페이지)에서 확인하십시오.

특이 벡터 계산식 저장 문서 특이값 분해를 포함하는 벡터 모델링 유형의 계산식 열을 데이터 테이블에 저장합니다. 결과 열에는 Text Score() JSL 함수가 사용됩니다. 이 함수에 대한 자세한 내용은 "도움말">"스크립트 인덱스"에서 확인하십시오.

용어 특이 벡터 저장 용어 특이값 분해에서 구한 특이 벡터 중 사용자가 지정한 수만큼을 새 데이터 테이블에 열로 저장합니다. 이 데이터 테이블에서 각 행은 하나의 용어에 해당합니다. 용어 테이블 데이터 테이블이 이미 열려 있는 경우 이 옵션을 사용하면 열이 해당 데이터 테이블에 저장됩니다. 처음 두 개의 저장된 열은 용어 SVD 그림에 표시된 점을 나타냅니다. 자세한 내용은 "[잠재 의미 분석 \(SVD\)](#)"(380 페이지)에서 확인하십시오.

제거 텍스트 탐색기 보고서 창에서 SVD 보고서를 제거합니다.

JMP PRO 주제 분석

"주제 분석, 회전된 SVD" 옵션은 DTM(문서 용어 행렬)의 부분 SVD(특이값 분해)에 대해 Varimax 회전을 수행합니다. DTM에서 유지할 주제의 개수에 해당하는 회전 특이 벡터의 수를 지정해야 합니다. 주제 수를 지정하면 주제 분석 보고서가 나타납니다.

주제 분석은 회전 PCA(주성분 분석)와 동등합니다. Varimax 회전은 여러 개의 특이 벡터를 구하고 이를 회전시켜 더 직접적으로 좌표 방향(용어 방향)을 향하게 합니다. 이렇게 회전하면 각 회전 벡터가 용어 집합을 향하므로 텍스트를 설명하는 데 도움이 됩니다. 음수 값은 반발력을 나타냅니다. 음수 값을 갖는 용어는 양수 값을 갖는 용어에 비해 주제에 나타나는 빈도가 낮습니다.

JMP^{PRO} 주제 분석 보고서

주제 분석 보고서에서는 회전 후 각 주제에서 적재량이 가장 큰 용어를 보여 줍니다. 추가 보고서에서는 회전된 특이값 분해의 성분을 보여 줍니다.

"주제별 상위 적재" 보고서에서는 각 주제의 용어 테이블을 보여 줍니다. 각 테이블의 용어는 각 주제에 대한 적재량 절대값이 가장 큰 용어입니다. 각 테이블은 적재량 절대값을 기준으로 내림차순 정렬되어 있습니다. 이러한 테이블을 사용하여 각 주제에 해당하는 개념 테마를 판별할 수 있습니다.

주제 분석 보고서에는 다음 보고서도 포함됩니다.

주제 적재 각 용어의 주제 간 적재량 행렬이 포함됩니다. 이 행렬은 회전 PCA의 요인 적재 행렬과 동등합니다.

주제별 단어 클라우드 각 주제당 하나씩의 단어 클라우드가 포함됩니다.

주제 점수 각 주제의 문서 점수 행렬이 포함됩니다. 주제 점수가 높은 문서는 해당 주제와 연관성이 있을 가능성이 높습니다.

주제 점수 그림 "텍스트 표시" 버튼과 각 문서의 주제 점수 그림이 포함됩니다. "텍스트 표시" 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다.

"주제 점수 그림" 보고서는 주제 점수 보고서의 행렬을 시각적으로 나타낸 것입니다. 그림의 각 패널은 주제 중 하나, 또는 주제 점수 행렬의 열 중 하나에 해당합니다. 각 패널 내에서 각 점은 코퍼스의 문서 중 하나, 또는 주제 점수 행렬의 행 중 하나에 해당합니다.

각 주제에 의해 설명된 분산 각 주제에 의해 설명되는 분산의 테이블이 포함됩니다. 이 테이블에는 각 주제에 의해 설명되는 변동이 백분율 또는 누적 백분율에 대한 열도 포함됩니다.

회전 행렬 Varimax 회전에 대한 회전 행렬이 포함됩니다.

JMP^{PRO} 주제 분석 보고서 옵션

"주제 분석"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

주제 산점도 행렬 회전된 특이값 분해 벡터에 대한 산점도 행렬을 표시하거나 숨깁니다. "텍스트 표시" 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다.

표시 옵션 주제 분석 보고서에 나타나는 내용을 표시하거나 숨기기 위한 옵션이 포함되어 있습니다. 자세한 내용은 "**주제 분석 보고서**"(384 페이지)에서 확인하십시오.

주제 이름 바꾸기 하나 이상의 주제에 대한 이름을 추가할 수 있습니다.

문서 주제 벡터 저장 회전 특이값 분해에서 구한 특이 벡터 중 사용자가 지정한 수만큼을 데이터 테이블에 열로 저장합니다.

주제 벡터 계산식 저장 회전 특이값 분해를 포함하는 벡터 모델링 유형의 계산식 열을 데이터 테이블에 저장합니다. 결과 열에는 Text Score() JSL 함수가 사용됩니다. 이 함수에 대한 자세한 내용은 "도움말">"스크립트 인덱스"에서 확인하십시오.

용어 주제 벡터 저장 주제 벡터를 "용어 테이블 저장" 옵션으로 생성된 데이터 테이블에 열로 저장합니다.

제거 SVD 보고서에서 주제 분석 보고서를 제거합니다.

JMP PRO 판별 분석

판별 분석에서는 DTM(문서 용어 행렬)의 열을 기준으로 그룹 또는 범주에서 각 문서의 소속을 예측합니다. 특히 판별 분석에서는 각 문서가 반응 열 범주로 분류되는지를 예측합니다. 판별 분석 옵션을 선택할 때는 범주 또는 그룹이 포함된 반응 열을 선택해야 합니다. 소속 그룹은 DTM의 열을 통해 예측됩니다. 판별 분석에 대한 자세한 내용은 **Multivariate Methods**의 "Discriminant Analysis" 장에서 확인하십시오.

텍스트 탐색기 플랫폼의 판별 분석 방법은 중심화된 DTM의 특이값 분해를 기반으로 합니다. 반응 열의 각 그룹에는 DTM을 중심화하는 데 사용되는 고유한 그룹 평균이 있습니다. 텍스트 탐색기 플랫폼의 판별 분석 방법은 DTM의 희소성을 활용하기 때문에 판별 분석 플랫폼보다 빠릅니다.

JMP PRO 판별 분석 규격 창

텍스트 탐색기의 판별 분석 옵션은 DTM(문서 용어 행렬)을 기반으로 합니다. DTM은 용어 목록에 있는 각 용어마다 열을 하나씩 생성함으로써 구성됩니다(지정된 최대 용어 수까지). 각 텍스트 문서(데이터 테이블의 행과 동등)는 DTM의 한 행에 해당합니다. DTM의 셀 값은 "규격" 창에서 사용자가 지정한 가중 유형에 따라 달라집니다.

텍스트 탐색기의 빨간색 삼각형 메뉴에서 "판별 분석"을 선택하면 다음 옵션이 포함된 "규격" 창이 나타납니다.

최대 용어 수 판별 분석에 포함되는 용어의 최대 개수입니다.

최소 용어 빈도 한 용어가 판별 분석에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다.

가중치 문서 용어 행렬의 셀에 들어갈 값을 결정하는 가중치 체계입니다. 가중치 체계 옵션에 대한 자세한 내용은 "**문서 용어 행렬 규격 창**"(377 페이지)에서 확인하십시오.

특이 벡터 수 판별 분석의 특이 벡터 수입니다. 기본값은 문서 수, 용어 수 또는 100 중 최소값입니다.

JMP PRO 판별 분석 보고서

텍스트 탐색기 플랫폼의 판별 분석 보고서에는 기본적으로 두 개의 열린 보고서, 즉 분류 요약 보고서와 판별 점수 보고서가 포함됩니다. 다른 보고서는 처음에는 닫혀 있습니다.

판별 분석 보고서에는 다음 보고서도 포함됩니다.

용어 평균 판별 분석에 사용된 용어의 테이블을 제공합니다. 이러한 용어는 DTM의 열에 해당합니다. 이 테이블에는 각 용어에 대한 각 그룹 내의 평균과 각 용어의 전체 평균 및 가중 표준편차가 포함됩니다.

각 그룹에 대한 제곱 거리 각 문서에 대해 각 그룹까지의 Mahalanobis 거리를 제공한 결과를 포함하는 테이블을 제공합니다. Mahalanobis 거리에 대한 자세한 내용은 **Multivariate Methods**의 "Correlations and Multivariate Techniques" 장에서 확인하십시오.

각 그룹에 대한 확률 문서가 각 그룹에 속할 확률이 포함된 테이블을 제공합니다.

분류 요약 판별 점수를 요약한 보고서를 제공합니다. 이 보고서는 판별 분석 플랫폼 보고서의 "점수 요약" 보고서에 해당합니다.

판별 점수 각 문서 및 기타 지원 정보의 예측 분류 테이블을 제공합니다. 이 테이블은 판별 분석 플랫폼 보고서의 "판별 점수" 테이블에 해당합니다.

JMP[®] 판별 분석 보고서 옵션

"판별 분석"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

정준 그림 정준 공간의 문서 및 그룹 평균 그림을 숨기거나 표시합니다. 정준 공간은 그룹을 가장 많이 나누는 공간입니다. 반응 변수의 수준이 세 개 이상인 경우에는 정준 좌표의 수를 지정해야 합니다. 정준 좌표를 세 개 이상 지정할 경우 이 옵션은 정준 그림 행렬을 생성합니다.

확률 저장 각 반응 수준에 대한 확률 열 및 가장 확률이 높은 반응이 포함된 열을 데이터 테이블에 저장합니다. 가장 확률이 높은 반응 열에는 모형에 따른 확률이 가장 높은 수준이 포함됩니다.

각 확률 열에서는 관측값이 해당 반응 수준에 속할 사후 확률을 제공합니다. "반응 확률" 열 특성은 각 확률 열에 저장됩니다. 반응 확률 열 특성에 대한 자세한 내용은 JMP 사용의 "열 정보 창" 장에서 확인하십시오.

확률 계산식 저장 가장 확률이 높은 반응을 예측하기 위한 계산식 열을 데이터 테이블에 저장합니다. 첫 번째로 저장되는 열에는 `Text Score()` 함수를 사용하여 각 반응 수준의 확률을 계산하는 계산식이 포함됩니다. 각 반응 수준에 대한 확률을 포함하는 열뿐만 아니라 예측 반응을 포함하는 열도 있습니다.

정준 점수 저장 각 관측값에 대해 정준 공간의 점수가 포함된 열을 데이터 테이블에 저장합니다. 정준 공간은 그룹을 가장 많이 나누는 공간입니다. k 번째 정준 점수의 열에는 정준 <k> 라는 이름이 지정됩니다.

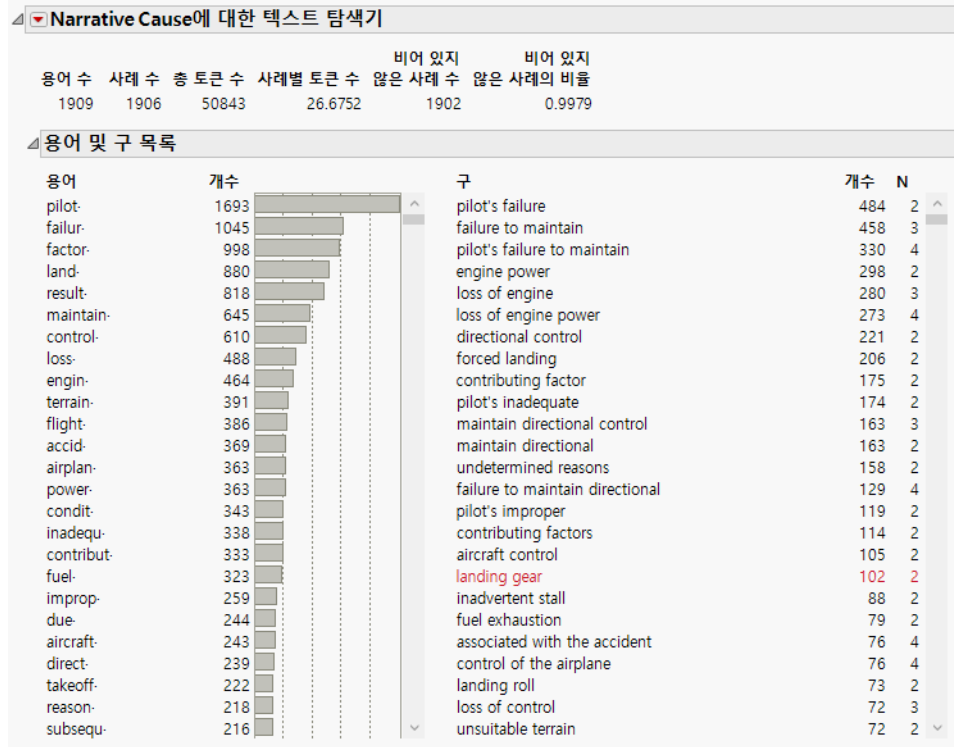
제거 텍스트 탐색기 보고서 창에서 판별 분석 보고서를 제거합니다.

JMP[®] PRO 텍스트 탐색기 플랫폼의 추가 예

이 예에서는 미국에서 2001 년에 발생한 사고에 대한 미국 연방 교통 안전 위원회의 비행기 사고 보고서를 살펴봅니다. 조사 결과에 대한 설명이 포함된 텍스트를 탐색해서 각 사고의 원인을 밝히려고 합니다. 또한 사고 보고서 모음에서 테마를 찾으려고 합니다.

1. **도움말 > 샘플 데이터 라이브러리**를 선택하고 **Aircraft Incidents.jmp** 를 엽니다.
2. **행 > 열 값에 따른 색상 또는 표식**을 선택합니다.
3. 열 목록에서 **Fatal** 을 선택하고 **확인**을 클릭합니다.
심각한 피해를 가져온 사고가 포함되어 있는 행이 빨간색으로 표시됩니다.
4. **분석 > 텍스트 탐색기**를 선택합니다.
5. "열 선택" 목록에서 **Narrative Cause** 를 선택하고 **텍스트 열**을 클릭합니다.
6. "언어" 목록에서 **영어**를 선택합니다.
7. "어간 추출 (Stemming)" 목록에서 **모든 용어의 어간 추출**을 선택합니다.
8. "토큰화" 목록에서 **기본 단어**를 선택합니다.
9. **확인**을 클릭합니다.

그림 12.12 Narrative Cause 에 대한 텍스트 탐색기



보고서에서 약 51,000 개의 토큰과 약 1,900 개의 고유 용어가 있음을 알 수 있습니다.

10. 용어 목록에서 "pilot" 을 마우스 오른쪽 버튼으로 클릭하고 **행 선택하기**를 선택합니다.

데이터 테이블에서 선택된 행의 개수를 통해 1,300 건 이상의 사고 보고서에서 "pilot" 이라는 단어의 몇 가지 형태가 나타남을 알 수 있습니다.

11. "pilot" 을 마우스 오른쪽 버튼으로 클릭하고 **중지 단어 추가**를 선택합니다.

단어 "pilot"의 일부 형태가 다른 용어들에 비해 빈도 높게 나타나므로 이러한 용어는 문서 간의 차이를 확인하는 데 충분한 정보를 제공하지 못합니다. 따라서 어간이 "plot" 인 모든 용어를 중지 단어 목록에 추가합니다.

JMP PRO 이 예의 나머지 단계는 JMP Pro 에서만 수행할 수 있습니다.

12. **JMP PRO** "Narrative Cause 에 대한 텍스트 탐색기" 옆의 빨간색 삼각형을 클릭하고 **잠재 의미 분석, SVD** 를 선택합니다.

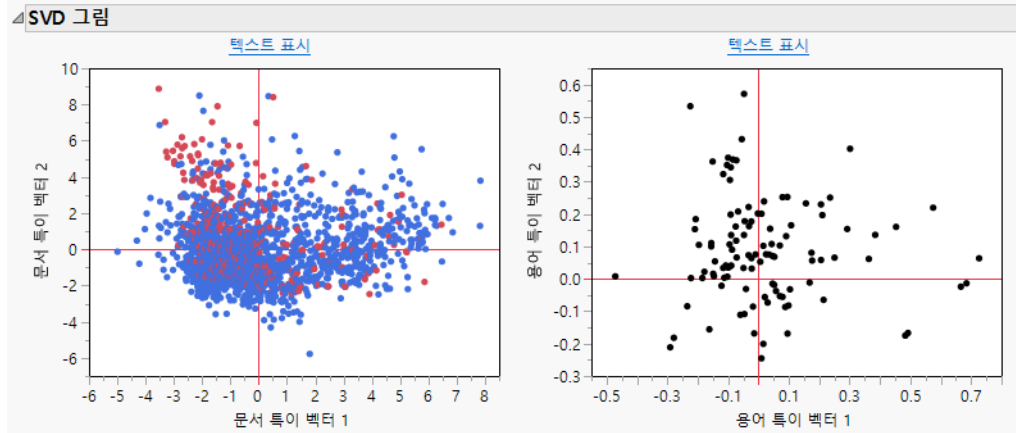
이 분석은 주제 분석을 위한 첫 번째 분석 단계로서, SVD 회전을 수행합니다.

13. **JMP PRO** "규격" 창에서 "최소 용어 빈도" 에 50 을 입력합니다.

약 51,000 개의 토큰이 있으므로 이 빈도는 모든 용어의 0.1% 이상을 나타내는 용어와 동등합니다.

14. **JMP PRO** **확인**을 클릭합니다.

그림 12.13 Narrative Cause 에 대한 SVD 그림



문서 SVD 그림에서는 fatal한 사건과 non-fatal한 사건 사이에 큰 차이가 나타나지 않습니다.

15. **JMP PRO** "SVD 중심화 및 척도화 TF IDF" 옆의 빨간색 삼각형을 클릭하고 **주제 분석, 회전된 SVD**를 선택합니다.

주제를 이루는 용어 그룹을 살펴보려고 합니다.

16. **JMP PRO** "주제 수"에 5를 입력합니다.
17. **JMP PRO** 확인을 클릭합니다.

그림 12.14 Narrative Cause 에 대한 주제별 상위 적재

주제별 상위 적재									
주제 1		주제 2		주제 3		주제 4		주제 5	
용어	적재	용어	적재	용어	적재	용어	적재	용어	적재
power-	0.67567	altitud-	0.48052	factor-	0.5093	control-	0.5125	fuel-	0.4864
loss-	0.66539	low-	0.45137	condit-	0.4677	direct-	0.4957	personnel-	0.4630
forc-	0.62046	dark-	0.42289	unsuit-	0.3984	experi-	0.4382	mainten-	0.4546
engin-	0.61866	night-	0.40408	accid-	0.3909	student-	0.4273	result-	0.4153
suitabl-	0.58926	maintain-	0.39283	select-	0.3842	lack-	0.3625	preflight-	0.3930
lack-	0.53292	instrument-	0.39239	associ-	0.3806	maintain-	0.3616	exhaust-	0.3663
reason-	0.47828	clearanc-	0.37881	area-	0.3628	instructor-	0.3559	inspect-	0.3640
undetermin-	0.46599	airspe-	0.33612	compens-	0.3352	supervis-	0.3282	plan-	0.3474
terrain-	0.37013	condit-	0.33473	failur-	-0.3207	power-	-0.3269	reason-	-0.3468
total-	0.31932	stall-	0.33010	wind-	0.3202	failur-	0.3171	undetermin-	-0.3438
land-	0.29402	flight-	0.32838	inadequ-	0.3080	reason-	-0.3142	improp-	0.3389
		contin-	0.31729	result-	-0.2845	undetermin-	-0.3101	inadequ-	0.3353
		maneu-	0.30931	terrain-	0.2663	crosswind-	0.3085	subsequ-	0.3246
		weather-	0.29484	airspe-	-0.2542	aircraft-	0.3007	maintain-	-0.3117
		adequ-	0.28082			factor-	0.2787	due-	0.2882

적재량이 가장 높은 각 주제에 대한 용어를 사용하면 해당 주제가 사고 보고서의 테마를 포착하는지 여부를 해석할 수 있습니다.

예를 들어 주제 1은 "power", "loss" 및 "engine"에 대한 적재량이 높으므로 사고 원인으로 엔진 동력 손실이라는 테마를 나타냅니다. 이는 전체 사고 보고서에서 273회 나타나는 "엔진 동력 손실"이라는 구와 일치합니다.

주제 2에서 적재량이 높은 단어를 살펴보면, 사고가 어둡거나 낮은 고도와 관련이 있다고 설명할 수 있습니다.

이 텍스트 분석 단계에서는 진행 방법을 다양하게 선택할 수 있습니다. 텍스트 분석은 반복적인 프로세스이므로 주제 정보를 사용하여 중지 단어를 추가하거나 구를 지정하는 방법으로 용어 목록을 세부적으로 큐레이팅할 수 있습니다. 가중치가 부여된 문서 용어 행렬, SVD 또는 회전된 SVD의 벡터를 데이터 테이블에 숫자 열로 저장하여 다른 JMP 분석 플랫폼에서 사용할 수도 있습니다. 다른 플랫폼에서 이러한 열을 사용하는 경우 추가 분석에 데이터 테이블의 다른 열을 포함할 수도 있습니다.

-
- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons.
- Agresti, A., and Coull, B. A. (1998). "Approximate is Better Than 'Exact' for Interval Estimation of Binomial Proportions." *American Statistician* 52:119–126.
- Asiribo, O., and Gurland, J. (1990). "Coping with Variance Heterogeneity." *Communication in Statistics: Theory and Methods* 19:4029–4048.
- Bartlett, M. S., and Kendall, D. G. (1946). "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation." *Supplement to the Journal of the Royal Statistical Society* 8:128–138.
- Bissell, A. F. (1990). "How Reliable is Your Capability Index?" *Applied Statistics* 30:331–340.
- Brown, M. B., and Benedetti, J. K. (1977). "Sampling Behavior of Tests for Correlation in Two-Way Contingency Tables." *Journal of the American Statistical Association* 72:305–315.
- Brown, M. B., and Forsythe, A. B. (1974). "Robust Tests for Equality of Variances." *Journal of the American Statistical Association* 69:364–367.
- Chen, S.-X., and Hall, P. (1993). "Empirical Likelihood Confidence Intervals for Quantiles." *The Annals of Statistics* 21:1166–1181.
- Chou, Y.-M., Owen, D. B., and Borrego, S. A. (1990). "Lower Confidence Limits on Process Capability Indices." *Journal of Quality Technology* 22:223–229.
- Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74:829–836.
- Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Education Psychological Measurement* 20:37–46.
- Conover, W. J. (1972). "A Kolmogorov Goodness-of-fit Test for Discontinuous Distributions." *Journal of the American Statistical Association* 67:591–596.
- Conover, W. J. (1980). *Practical Nonparametric Statistics*. New York: John Wiley & Sons.
- Conover, W. J. (1999). *Practical Nonparametric Statistics*. 3rd ed. New York: John Wiley & Sons.
- Cureton, E. E. (1967). "The Normal Approximation to the Signed-Rank Sampling Distribution when Zero Differences are Present." *Journal of the American Statistical Association* 62:1068–1069.
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics* 44:837–845.
- Devore, J. L. (1995). *Probability and Statistics for Engineering and the Sciences*. Pacific Grove, CA: Duxbury Press.

- Dunn, O. J. (1964). "Multiple Comparisons Using Rank Sums." *Technometrics* 6:241–252.
- Dunnett, C. W. (1955). "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50:1096–1121.
- Efron, B. (1981). "Nonparametric Standard Errors and Confidence Intervals." *The Canadian Journal of Statistics* 9:139–158.
- Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. 2nd ed. Boca Raton, Florida: CRC.
- Fleiss, J. L., Cohen, J., and Everitt, B. S. (1969). "Large-Sample Standard Errors of Kappa and Weighted Kappa." *Psychological Bulletin* 72:323–327.
- Friendly, M. (1994). "Mosaic Displays for Multi-Way Contingency Tables." *Journal of the American Statistical Association* 89:190–200.
- Goodman, L. A., and Kruskal, W. H. (1979). *Measures of Association for Cross Classification*. New York: Springer-Verlag.
- Gupta, S. S. (1965). "On Some Multiple Decision (Selection and Ranking) Rules." *Technometrics* 7:225–245.
- Hajek, J. (1969). *A Course in Nonparametric Statistics*. San Francisco: Holden-Day.
- Hartigan, J. A., and Kleiner, B. (1981). "Mosaics for Contingency Tables." In *Computer Science and Statistics: Proceedings of the Thirteenth Symposium on the Interface*, edited by W. F. Eddy, 268–273. New York: Springer-Verlag.
- Hayter, A. J. (1984). "A Proof of the Conjecture That the Tukey-Kramer Method Is Conservative." *Annals of Mathematical Statistics* 12: 61–75.
- Hosmer, D. W., and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley & Sons.
- Hsu, J. (1981). "Simultaneous Confidence Intervals for All Distances from the 'Best'." *Annals of Statistics* 9:1026–1034.
- Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall.
- Huber, P. J. (1973). "Robust Regression: Asymptotics, Conjecture, and Monte Carlo." *Annals of Statistics* 1:799–821.
- Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics*. 2nd ed. New York: John Wiley & Sons.
- Iman, R. L. (1974). "Use of a t-statistic as an Approximation to the Exact Distribution of Wilcoxon Signed Ranks Test Statistic." *Communications in Statistics—Simulation and Computation* 3:795–806.
- Jones, M. C., and Pewsey, A. (2009). "Sinh-Arcsinh Distributions." *Biometrika* 96:761–780.
- Kendall, M., and Stuart, A. (1979). *The Advanced Theory of Statistics*. 4th ed. Vol. 2. New York: Macmillan.
- Keuls, M. (1952). "The Use of the 'Studentized Range' in Connection with an Analysis of Variance." *Euphytica* 1.2:112–122.
- Kramer, C. Y. (1956). "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." *Biometrics* 12:307–310.

- Lehmann, E. L., and D'Abrera, H. J. M. (2006). *Nonparametrics: Statistical Methods Based on Ranks*. Rev. ed. San Francisco: Holden-Day.
- Levene, H. (1960). "Robust Tests for the Equality of Variance." In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann. Palo Alto, CA: Stanford University Press.
- McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- Meeker, W. Q., and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.
- Meeker, W. Q., Hahn, G. J., and Escobar, L. A. (2017). *Statistical Intervals: A Guide for Practitioners and Researchers*. 2nd ed. New York: John Wiley & Sons.
- Miller, A. J. (1972). "Letter to the Editor." *Technometrics* 14:507.
- Nagelkerke, N. J. D. (1991). "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78:691-692.
- Nelson, P. R., Wludyka, P. S., and Copeland, K. A. F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Philadelphia: Society for Industrial and Applied Mathematics.
- Neter, J., Wasserman, W., and Kutner, M. H. (1990). *Applied Linear Statistical Models*. 3rd ed. Boston: Irwin.
- O'Brien, R. G. (1979). "A General ANOVA Method for Robust Tests of Additive Models for Variances." *Journal of the American Statistical Association* 74:877-880.
- O'Brien, R., and Lohr, V. (1984). "Power Analysis For Linear Models: The Time Has Come." *Proceedings of the Ninth Annual SAS User's Group International Conference*, 840-846. Cary, NC: SAS Institute Inc.
- Olejnik, S. F., and Algina, J. (1987). "Type I Error Rates and Power Estimates of Selected Parametric and Nonparametric Tests of Scale." *Journal of Educational Statistics* 12:45-61.
- Pratt, J. W. (1959). "Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures." *Journal of the American Statistical Association* 54:655-667.
- Reinsch, C. H. (1967). "Smoothing by Spline Functions." *Numerische Mathematik* 10:177-183.
- Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.
- Rubin, D. (1981). "The Bayesian Bootstrap." *The Annals of Statistics* 9:130-134.
- SAS Institute Inc. (2018a). "Introduction to Nonparametric Analysis." In *SAS/STAT 15.1 User's Guide*. Cary, NC: SAS Institute Inc.
<https://support.sas.com/documentation/onlinedoc/stat/151/intronpar.pdf>.
- SAS Institute Inc. (2018b). "The FREQ Procedure." In *SAS/STAT 15.1 User's Guide*. Cary, NC: SAS Institute Inc. <https://support.sas.com/documentation/onlinedoc/stat/151/freq.pdf>.
- Slifker, J. F., and Shapiro, S. S. (1980). "The Johnson System: Selection and Parameter Estimation." *Technometrics* 22:239-246.
- Snedecor, G. W., and Cochran, W. G. (1980). *Statistical Methods*. 7th ed. Ames, Iowa: Iowa State University Press.

- Somers, R. H. (1962). "A New Asymmetric Measure of Association for Ordinal Variables." *American Sociological Review* 27:799–811.
- Tan, C. Y., and Iglewicz, B. (1999). "Measurement-Methods Comparisons and Linear Statistical Relationship." *Technometrics* 41:192–201.
- Tamhane, A. C., and Dunlop, D. D. (2000). *Statistics and Data Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Tukey, J. W. (1953). "The Problem of Multiple Comparisons." In *Multiple Comparisons, 1948–1983*, edited by H. I. Braun, vol. 8 of *The Collected Works of John W. Tukey* (published 1994), 1–300. London: Chapman & Hall. Unpublished manuscript.
- Welch, B. L. (1951). "On the Comparison of Several Mean Values: An Alternative Approach." *Biometrika* 38:330–336.
- Wheeler, D. J. (2003). *Range Based Analysis of Means*. Knoxville, TN: SPC Press.
- Wilson, E. B. (1927). "Probable Inference, the Law of Succession, and Statistical Inference." *Journal of the American Statistical Association* 22:209–212.
- Wludyka, P. S., and Nelson, P. R. (1997). "An Analysis-of-Means-Type Test for Variances From Normal Populations." *Technometrics* 39:274–285.

기술 라이선스 고지 사항

- Scintilla - Copyright © 1998-2017 by Neil Hodgson <neilh@scintilla.org>.

All Rights Reserved.

Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

- Progress® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library - Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages - Copyright © 2009-2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES

(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software - Copyright © 1995-2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- bzip2, the associated library “libbzip2”, and all documentation, are Copyright © 1996-2010, Julian R Seward. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- R software is Copyright © 1999-2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
- Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

Copyright © 1998 - 2003 Daniel Veillard. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

Copyright © 1985, 1986, 1992, 1993

The Regents of the University of California. All rights reserved.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball - Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.
All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako - Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND

NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 2006 –2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS “AS IS” WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

