버전 17

기본 분석

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

Marcel Proust

JMP Statistical Discovery LLC
SAS Campus Drive
Cary, North Carolina 27513-2414

17.0

The correct bibliographic citation for this manual is as follows: JMP Statistical Discovery LLC 2022. *JMP® 17 Basic Analysis*. Cary, NC: JMP Statistical Discovery LLC

JMP® 17 Basic Analysis

JMP Statistical Discovery LLC, SAS Campus Drive, Cary, North Carolina 27513-2414.

October 2022

JMP 활용하기

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

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

- JMP 시작 방법에 대한 라이브 및 녹화 웹 캐스트
- 새로운 기능과 고급 기법에 대한 비디오 데모 및 웹 캐스트
- JMP 교육 과정 등록에 대한 상세 정보
- 현지에서 개최되는 세미나 일정
- 다른 사용자들이 JMP 를 사용하는 방법을 보여주는 성공 사례
- JMP 사용자 커뮤니티 , 추가기능 및 스크립트 예를 포함한 사용자용 리소스 , 포럼 , 블로그 , 컨퍼런스 정보 등

jmp.com/getstarted

목차
기본 분석

1 JMP 알아보기 13
설명서 및 추가 리소스
JMP 설명서에 사용되는 서식 규칙 15
JMP 도움말 16
JMP 설명서 라이브러리 16
JMP 학습을 위한 추가 리소스 22
JMP 검색 23
JMP 자습서 23
샘플 데이터 테이블 23
통계 및 JSL 용어에 대해 알아보기 24
JMP 팁 및 힌트 알아보기 24
JMP 툴팁 24
JMP 사용자 커뮤니티 24
무료 온라인 Statistical Thinking 교육 과정 25
JMP 새로운 사용자를 위한 입문 키트 25
Statistics Knowledge 포털 25
JMP 교육 과정 25
사용자가 작성한 JMP 설명서 25
JMP 시작하기 창 25
JMP 기술 지원 26

2 기본 분석 소개 27
기본 분석 방법 개요

3 분포 29
분포 플랫폼 사용
분포 플랫폼 개요 32
분포 플랫폼의 예 32
분포 플랫폼 시작 34
분포 보고서 36
히스토그램 37
빈도 보고서 40

분위수 보고서 41
요약 통계량 보고서 41
분포 플랫폼 옵션 44
범주형 변수에 대한 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 54
신뢰 구간 55
예측 구간 56
공차 구간 56
공정 능력 56
분포 적합 58
연속형 변수에 대한 저장 옵션 63
분포 플랫폼의 추가 예 64
예 : 여러 히스토그램에서 데이터 선택 65
기준 변수의 예 66
두 수준에 대한 검정 확률의 예 67
세 개 이상의 수준에 대한 검정 확률의 예 68
예측 구간의 예 70
공차 구간의 예 72
공정 능력의 예 73
분포 플랫폼에 대한 통계 상세 정보 74
표준 오차 막대에 대한 통계 상세 정보 75
분위수에 대한 통계 상세 정보 75
요약 통계량에 대한 통계 상세 정보 76
정규 분위수 그림에 대한 통계 상세 정보 77
Wilcoxon 부호 순위 검정에 대한 통계 상세 정보 77
표준편차 검정에 대한 통계 상세 정보 79
정규 분위수에 대한 통계 상세 정보 79
표준화된 데이터 저장을 위한 통계 상세 정보 80

예측 구간에 대한 통계 상세 정보 80
공차 구간에 대한 통계 상세 정보 81
연속형 적합 분포에 대한 통계 상세 정보 83
이산형 적합 분포에 대한 통계 상세 정보 89

4 X 로 Y 적합 소개 93
두 변수 간의 관계 분석

5 이변량 분석 95
두 연속형 변수 간의 관계 분석

이변량 플랫폼 개요 97
이변량 분석의 예 97
이변량 플랫폼 시작 99
데이터 형식 99
이변량 보고서 100
이변량 플랫폼 옵션 101
이변량 적합 옵션 103
이변량 적합 보고서 106
평균 적합 보고서 107
선형 적합, 다항식 적합 및 특수 적합 보고서 108
특수 적합 창 113
스플라인 적합 보고서 113
커널 평활기 보고서 114
각 값 적합 보고서 114
직교 적합 보고서 115
로버스트 적합 보고서 115
Passing-Bablok 적합 보고서 116
밀도 타원 보고서 117
비모수 밀도 보고서 119
이변량 플랫폼의 추가 예 120
특수 적합 옵션의 예 120
직교 적합 옵션의 예 121
로버스트 적합 옵션의 예 122
밀도 타원을 사용한 그룹화의 예 124
회귀선을 사용한 그룹화의 예 125
기준 변수를 사용한 그룹화의 예 126
이변량 플랫폼에 대한 통계 상세 정보 128
선형 적합 옵션에 대한 통계 상세 정보 128
다항식 적합 옵션에 대한 통계 상세 정보 129
스플라인 적합 옵션에 대한 통계 상세 정보 129

직교 적합 옵션에 대한 통계 상세 정보 130
로버스트 옵션에 대한 통계 상세 정보 131
Passing-Bablok 적합 옵션에 대한 통계 상세 정보 131
적합 요약 보고서에 대한 통계 상세 정보 132
적합 결여 보고서에 대한 통계 상세 정보 132
모수 추정값 보고서에 대한 통계 상세 정보 133
평활 적합 보고서에 대한 통계 상세 정보 133
이변량 정규 타원 보고서에 대한 통계 상세 정보 134

6 일원 분석 135

연속형 Y 와 범주형 X 변수 간의 관계 분석

일원 분석 플랫폼 개요 138
일원 분석의 예 138
일원 분석 플랫폼 시작 141
데이터 형식 141
일원 분석 보고서 142
일원 분석 플랫폼 옵션 143
일원 분석 보고서 150
분위수 보고서 151
평균 /ANOVA/ 합동 t 보고서 151
평균 및 표준편차 보고서 153
t- 검정 보고서 154
평균 분석 보고서 154
평균 비교 보고서 156
비모수 검정 보고서 159
비모수 다중 비교 보고서 161
이분산 보고서 163
동등성 검정 보고서 165
로버스트 적합 보고서 168
검정력 보고서 168
매칭 열 보고서 170
일원 분석 그림 요소 170
평균 다이아몬드 및 X 축 비례 171
평균 선 , 오차 막대 및 표준편차 선 172
비교 원 172
일원 분석 플랫폼의 추가 예 174
평균 분석의 예 174
분산에 대한 평균 분석의 예 175
개별 쌍 비교 (스튜던트 t) 검정의 예 176
전체 쌍 비교 (Tukey HSD) 검정의 예 178

최량 비교 (Hsu MCB) 검정의 예 179
대조군 비교 (Dunnett) 의 예 181
단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예 182
예 : 4 가지 평균 비교 검정 대조 183
비모수 Wilcoxon 검정의 예 184
이분산 옵션의 예 185
동등성 검정의 예 187
로버스트 적합 옵션의 예 188
검정력 옵션의 예 189
정규 분위수 그림의 예 191
CDF 그림의 예 192
밀도 옵션의 예 193
매칭 열 옵션의 예 194
예 : 일원 분석을 위한 데이터 쌓기 195
일원 분석 플랫폼에 대한 통계 상세 정보 202
비교 원에 대한 통계 상세 정보 202
검정력에 대한 통계 상세 정보 203
ANOM 에 대한 통계 상세 정보 204
적합 요약 보고서에 대한 통계 상세 정보 205
분산 동일 여부 검정에 대한 통계 상세 정보 205
비모수 검정 통계량에 대한 통계 상세 정보 206
로버스트 적합에 대한 통계 상세 정보 209

7 분할 분석 211
두 범주형 변수 간의 관계 분석

분할 분석의 예 213
분할 분석 플랫폼 시작 215
데이터 형식 215
분할 분석 보고서 216
모자이크 그림 217
분할표 219
검정 보고서 221
분할 분석 플랫폼 옵션 222
분할 분석 보고서 225
비율에 대한 평균 분석 보고서 225
대응 분석 보고서 226
Cochran-Mantel-Haenszel 검정 보고서 227
합치도 통계량 보고서 227
상대 위험도 보고서 228
비율에 대한 2 표본 검정 보고서 228

연관성 측도 보고서 229
Cochran Armitage 추세 검정 보고서 230
Fisher 정확 검정 보고서 230
동등성 검정 보고서 230
분할 분석 플랫폼의 추가 예 232
비율에 대한 평균 분석의 예 233
대응 분석의 예 234
Cochran-Mantel-Haenszel 검정의 예 235
합치도 통계량 옵션의 예 236
상대 위험도 옵션의 예 237
비율에 대한 2 표본 검정의 예 239
연관성 측도 옵션의 예 240
Cochran Armitage 추세 검정의 예 241
분할 분석 플랫폼에 대한 통계 상세 정보 242
합치도 통계량에 대한 통계 상세 정보 242
승산비에 대한 통계 상세 정보 243
검정 보고서에 대한 통계 상세 정보 243
대응 분석에 대한 통계 상세 정보 244

8 로지스틱 분석 245

범주형 Y 변수와 연속형 X 변수 간의 관계 분석

로지스틱 플랫폼 개요 247
명목형 로지스틱 회귀의 예 247
로지스틱 플랫폼 시작 249
데이터 형식 250
로지스틱 보고서 250
로지스틱 그림 251
반복 보고서 252
전체 모형 검정 보고서 252
적합 상세 정보 보고서 253
모수 추정값 보고서 253
로지스틱 플랫폼 옵션 254
로지스틱 분석 보고서 256
ROC 곡선 256
역추정 예측 257
로지스틱 회귀의 추가 예 257
순서형 로지스틱 회귀의 예 257
로지스틱 그림의 예 258
ROC 곡선의 예 261
역추정 예측의 예 262

로지스틱 플랫폼에 대한 통계 상세 정보 264

9 테이블 생성 267
대화식으로 요약 테이블 생성

테이블 생성 플랫폼의 예 269
테이블 생성 플랫폼 시작 274
테이블 생성 플랫폼 대화상자 창 275
통계량 추가 276
테이블 생성 플랫폼 보고서 279
분석 열 280
그룹화 열 280
열 및 행 테이블 281
테이블 편집 282
테이블 생성 플랫폼 옵션 283
테스트 빌드 패널 표시 284
마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴 284
테이블 생성 플랫폼의 추가 예 285
예 : 여러 테이블 생성 및 내용 재배열 285
예 : 여러 열을 단일 테이블에 결합 289
예 : 페이지 열 사용 291
그룹화 열 쌓기의 예 292

10 시뮬레이션 295
모수 재표집을 통해 난제 해결 방안 모색

시뮬레이션 기능 개요 297
시뮬레이션 기능의 예 297
시뮬레이션 기능 시작 302
시뮬레이션 창 302
시뮬레이션 결과 테이블 303
시뮬레이션 결과 보고서 304
시뮬레이션 검정력 보고서 304
시뮬레이션 기능의 추가 예 304
순열 검정의 예 305
일반화 회귀에서 요인 유지의 예 307
전향적 검정력 분석의 예 312

11 붓스트랩 323
재표집을 통한 통계량 분포 근사

붓스트랩 기능 개요 325
붓스트랩을 지원하는 JMP 플랫폼 326
붓스트랩의 예 327

붓스트랩 창 옵션 329
누적 붓스트랩 결과 테이블 330
비누적 붓스트랩 결과 테이블 331
붓스트랩 결과 분석 332
붓스트랩의 추가 예 333
붓스트랩에 대한 통계 상세 정보 338
부분 가중치에 대한 통계 상세 정보 338
편향 수정 백분위수 구간에 대한 통계 상세 정보 338

12 텍스트 탐색기 341

데이터의 비정형 텍스트 탐구

텍스트 탐색기 플랫폼 개요 343
텍스트 처리 단계 344
텍스트 탐색기 플랫폼의 예 345
텍스트 탐색기 플랫폼 시작 349
정규 표현식 편집기에서 정규 표현식 사용자 정의 351
텍스트 탐색기 보고서 356
요약 개수 보고서 357
용어 및 구 목록 357
텍스트 탐색기 플랫폼 옵션 360
텍스트 준비 옵션 361
텍스트 분석 옵션 367
저장 옵션 368
텍스트 탐색기의 보고서 옵션 369
잠재 계층 분석 370
잠재 의미 분석 (SVD) 371
SVD 보고서 372
SVD 보고서 옵션 373
주제 분석 375
주제 분석 보고서 376
주제 분석 보고서 옵션 376
판별 분석 377
판별 분석 보고서 377
판별 분석 보고서 옵션 378
용어 선택 379
용어 선택 설정 379
용어 선택 보고서 380
용어 선택 보고서 옵션 382
감정 분석 382
감정 분석 보고서 384

감정 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

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

목차

JMP 설명서에 사용되는 서식 규칙 15
JMP 도움말 16
JMP 설명서 라이브러리 16
JMP 학습을 위한 추가 리소스 22
JMP 검색 23
JMP 자습서 23
샘플 데이터 테이블 23
통계 및 JSL 용어에 대해 알아보기 24
JMP 팁 및 힌트 알아보기 24
JMP 툴팁 24
JMP 사용자 커뮤니티 24
무료 온라인 Statistical Thinking 교육 과정 25
JMP 새로운 사용자를 위한 입문 키트 25
Statistics Knowledge 포털 25
JMP 교육 과정 25
사용자가 작성한 JMP 설명서 25
JMP 시작하기 창 25
JMP 기술 지원 26

JMP 설명서에 사용되는 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 JMP PRO 이 표시됩니다 . JMP Pro 기능의 개요는 jmp.com/software/pro 에서 확인할 수 있습니다 .

참고 : 참고 섹션에는 특수 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 " 에서 확인하십시오 .

JMP 설명서 라이브러리

도움말 시스템 콘텐츠는 JMP 설명서 라이브러리라는 단일 PDF 파일로도 제공됩니다. 이 파일을 열려면 **도움말 > JMP 설명서 라이브러리**를 선택합니다. Documentation PDF Files 추가기능을 다운로드하여 JMP 라이브러리에 있는 각 문서의 개별 PDF 파일을 검색할 수도 있습니다. https://community.jmp.com에서 사용 가능한 추가기능을 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.	"분석" 메뉴의 다음 플랫폼에 대해 설명합니다. • 분포 • X 로 Y 적합 • 테이블 생성 • 텍스트 탐색기 분석 > X로 Y 적합을 통해 이변량 분석, 일원 ANOVA 및 분할 분석을 수행하는 방법을 다룹니다. 붓스트랩을 사용하여 표집 분포에 근사한 값을 산출하는 방법과 시뮬레이션 플랫폼을 사용하여 모수 재표집을 수행하는 방법도 포함되어 있습니다.
Essential Graphing	데이터에 이상적인 그래프를 찾습니다.	"그래프" 메뉴의 다음 플랫폼에 대해 설명합니다. • 그래프 빌더 • 3D 산점도 • 등고선 그림 • 버블 그림 • 평행 그림 • 셀 그림 • 산점도 행렬 • 삼원 그림 • 트리맵 • 차트 • 중첩 그림 이 설명서에서는 배경 맵 및 사용자 맵을 생성하는 방법도 다룹니다.
Profilers	반응 표면의 횡단면을 볼 수 있게 해주는 대화식 프로파일링 도구의 사용 방법을 알아봅니다.	"그래프" 메뉴에 나열된 모든 프로파일러를 다룹니다. 랜덤 입력을 사용한 시뮬레이션 실행과 함께 잡음 요인 분석이 포함됩니다.

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

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

문서 제목	문서 용도	문서 내용
다변량 방법	몇 개의 변수를 동시에 분석하기 위한 기법을 알아봅니다.	분석 > 다변량 방법 메뉴의 다음 플랫폼에 대해 설명합니다. • 다변량 • 주성분 • 판별 • 부분 최소 제곱 • 다중 대응 분석 • 구조 방정식 모형 • 요인 분석 • 다차원 척도법 • 다변량 임베딩 • 항목 분석 분석 > 군집화 메뉴의 다음 플랫폼에 대해 설명합니다. • 계층적 군집화 • K 평균 군집화 • 정규 혼합 • 잠재 계층 분석 • 변수 군집화

문서 제목	문서 용도	문서 내용
품질 및 공정 방법	공정 평가 및 개선을 위한 도구에 대해 알아봅니다.	분석 > 품질 및 공정 메뉴의 다음 플랫폼에 대해 설명합니다. • 관리도 빌더 및 개별 관리도 • 측정 시스템 분석 (EMP 및 유형 1 게이지) • 계량형 / 계수형 게이지 차트 • 공정 변수 선별 • 공정 능력 • 모형 기반 다변량 관리도 • 레거시 관리도 • 파레토도 • 다이어그램 • 규격 한계 관리 • OC 곡선
Reliability and Survival Methods	제품 또는 시스템의 신뢰도 평가 및 향상 방법과 사람 및 제품의 생존 데이터 분석 방법을 알아봅니다.	분석 > 신뢰성 및 생존 메뉴의 다음 플랫폼에 대해 설명합니다. • 수명 분포 • 수명 분포 적합 • 누적 손상 • 재발 분석 • 열화 • 반복 측정 열화 • 파괴 열화 • 신뢰도 예측 • 신뢰도 성장 • 신뢰도 블록 다이어그램 • 수리 가능 시스템 시뮬레이션 • 생존 • 모수 생존 모형 적합 • 비례 위험 모형 적합

문서 제목	문서 용도	문서 내용
Consumer Research	소비자 선호도를 연구하고 해당 정보를 사용하여 보다 나은 제품 및 서비스를 개발하기 위한 방법을 알아봅니다.	분석 > 소비자 조사 메뉴의 다음 플랫폼에 대해 설명합니다. • 범주형 • 선택 • 최대차이 • Uplift • 다중 요인 분석
Genetics	JMP에서 유전자 데이터를 분석하고, 해당 데이터를 사용하여 최적의 유전자 교배를 예측하기 위한 육종 프로그램을 시뮬레이션하는 데 사용할 수 있는 방법을 알아봅니다.	분석 > 유전학 메뉴의 다음 플랫폼에 대해 설명합니다. • 표지자 통계량 • 표지자 시뮬레이션
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 JMP에 대해 배울 수 있습니다.

- "JMP 검색 "
- "JMP 자습서 "
- " 샘플 데이터 테이블 "
- " 통계 및 JSL 용어에 대해 알아보기 "
- "JMP 팁 및 힌트 알아보기 "
- "JMP 툴팁 "
- "JMP 사용자 커뮤니티 "
- " 무료 온라인 Statistical Thinking 교육 과정 "

- "JMP 새로운 사용자를 위한 입문 키트 "
- "Statistics Knowledge 포털 "
- "JMP 교육 과정 "
- " 사용자가 작성한 JMP 설명서 "
- "JMP 시작하기 창 "

JMP 검색

통계 절차의 위치를 잘 모르면 JMP에서 검색을 수행하십시오. 결과는 데이터 테이블 또는 보고서와 같이 검색 시작 창에 맞게 조정됩니다.

1. **도움말** > **JMP 검색**을 클릭합니다 . 또는 Ctrl+ 쉼표를 누릅니다 .
2. 검색 텍스트를 입력합니다 .
3. 원하는 절차가 포함된 결과를 클릭합니다 .

   오른쪽에는 절차에 대한 설명과 위치가 표시됩니다 .
4. 해당 버튼을 클릭하여 결과를 열거나 이동합니다 .

JMP 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.

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

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

샘플 데이터 테이블

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

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

샘플 데이터 테이블은 다음 디렉터리에 설치되어 있습니다 .

Windows: C:\Program Files\SAS\JMP\17\Samples\Data

macOS: \Library\Application Support\JMP\17\Samples\Data

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

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

통계 및 JSL 용어에 대해 알아보기

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

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

JMP를 처음 시작할 때는 "오늘의 유익한 정보" 창이 표시됩니다. 이 창에서는 JMP를 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.

JMP 툴팁

JMP 에서 다음과 같은 항목을 커서로 가리키면 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 JMP에 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 사이트의 JMP 리소스에 액세스하려면 **도움말 > JMP 웹 > 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 jmp.com/statisticalthinking 에서 확인하십시오.

JMP 새로운 사용자를 위한 입문 키트

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

Statistics Knowledge 포털

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

JMP 교육 과정

SAS 에서는 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 jmp.com/training 에서 확인하십시오.

사용자가 작성한 JMP 설명서

JMP 웹 사이트에서는 JMP 사용자가 작성한 추가 JMP 사용 설명서가 제공됩니다 . 자세한 내용은 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 .

JMP 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 jmp.com/support 에서 제공됩니다.

2 장

기본 분석 소개

기본 분석 방법 개요

기본 분석에서는 JMP에서 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 " 분포 " 에서 확인하십시오.
- X 로 Y 적합 플랫폼에서는 지정된 X 및 Y 변수 쌍을 모델링 유형에 따라 컨텍스트별로 분석합니다. 자세한 내용은 "X 로 Y 적합 소개 " 에서 확인하십시오. 네 가지 분석 유형으로는 다음이 포함됩니다.
  - 이변량 플랫폼 : 두 연속형 X 변수 간의 관계를 분석합니다. 자세한 내용은 " 이변량 분석 " 에서 확인하십시오.
  - 일원 분석 플랫폼 : 범주형 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지를 분석합니다. 자세한 내용은 " 일원 분석 " 에서 확인하십시오.
  - 분할 분석 플랫폼 : 범주형 X 요인의 값에 따라 좌우되는 범주형 반응 변수의 분포를 분석합니다. 자세한 내용은 " 분할 분석 " 에서 확인하십시오.
  - 로지스틱 플랫폼 : 반응 범주 (Y) 에 대한 확률을 연속형 X 예측 변수에 적합시킵니다. 자세한 내용은 " 로지스틱 분석 " 에서 확인하십시오.
- 테이블 생성 플랫폼 : 기술 통계량 테이블을 대화식으로 생성합니다. 자세한 내용은 " 테이블 생성 " 에서 확인하십시오.
- 시뮬레이션 기능으로는 모수 및 비모수 시뮬레이션을 수행할 수 있습니다. 자세한 내용은 " 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 " 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 " 텍스트 탐색기 " 에서 확인하십시오.

3 장

분포

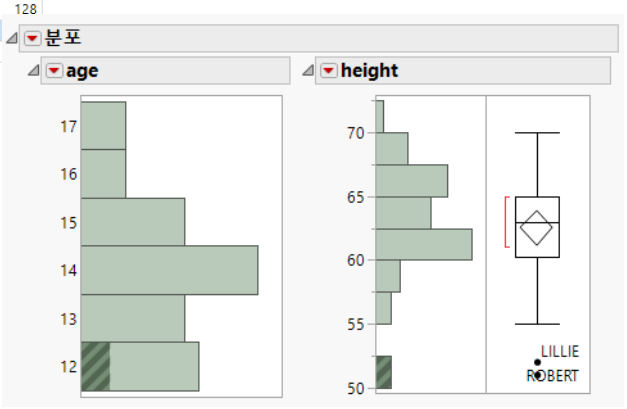
분포 플랫폼 사용

히스토그램, 상자 그림 및 요약 통계량을 사용하여 단일 변수의 분포를 탐색하려면 분포 플랫폼을 사용합니다. *t* 검정, *z* 검정, 카이제곱 검정, 동등성 검정 등을 비롯한 여러 가지 유형의 가설 검정을 수행할 수 있습니다. 신뢰도, 공차 및 예측 구간을 생성하고 공정 능력을 추정할 수도 있습니다. **단변량**이란 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.

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

그림 3.1 분포 플랫폼의 예

	name	age	sex	height	weight
1	KATIE	12	F	59	95
2	LOUISE	12	F	61	123
3	JANE	12	F	55	74
4	JACLYN	12	F	66	145
5	LILLIE	12	F	52	64
6	TIM	12	M	60	84
7	JAMES	12	M	61	128
8	ROBERT	12	M	51	
9	BARBARA	13	F	60	



목차

분포 플랫폼 개요 32
분포 플랫폼의 예 32
분포 플랫폼 시작 34
분포 보고서 36
히스토그램 37
빈도 보고서 40
분위수 보고서 41
요약 통계량 보고서 41
분포 플랫폼 옵션 44
범주형 변수에 대한 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 54
신뢰 구간 55
예측 구간 56
공차 구간 56
공정 능력 56
분포 적합 58
연속형 변수에 대한 저장 옵션 63
분포 플랫폼의 추가 예 64
예: 여러 히스토그램에서 데이터 선택 65
기준 변수의 예 66
두 수준에 대한 검정 확률의 예 67
세 개 이상의 수준에 대한 검정 확률의 예 68
예측 구간의 예 70
공차 구간의 예 72
공정 능력의 예 73
분포 플랫폼에 대한 통계 상세 정보 74

표준 오차 막대에 대한 통계 상세 정보 75
분위수에 대한 통계 상세 정보 75
요약 통계량에 대한 통계 상세 정보 76
정규 분위수 그림에 대한 통계 상세 정보 77
Wilcoxon 부호 순위 검정에 대한 통계 상세 정보 77
표준편차 검정에 대한 통계 상세 정보 79
정규 분위수에 대한 통계 상세 정보 79
표준화된 데이터 저장을 위한 통계 상세 정보 80
예측 구간에 대한 통계 상세 정보 80
공차 구간에 대한 통계 상세 정보 81
연속형 적합 분포에 대한 통계 상세 정보 83
이산형 적합 분포에 대한 통계 상세 정보 89

분포 플랫폼 개요

분포 플랫폼을 사용하여 범주형 변수와 연속형 변수를 모두 분석할 수 있습니다. 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 스코어, 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

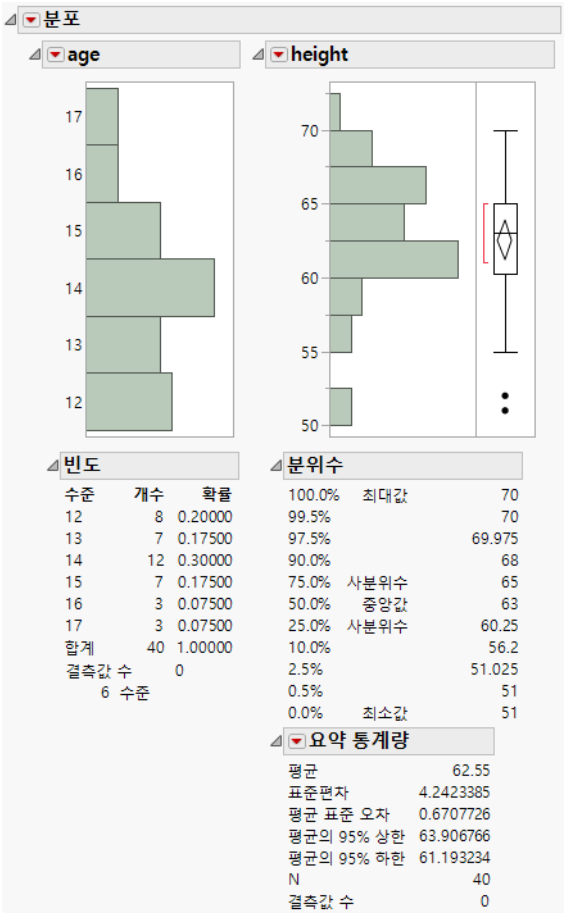
분포 플랫폼의 예

이 예에는 학생 40명의 연령과 키에 대한 데이터가 있습니다. 변수의 분포를 시각화하고 해당 분포에서 이상치를 찾으려고 합니다.

1. **도움말 > 샘플 데이터 폴더**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 .

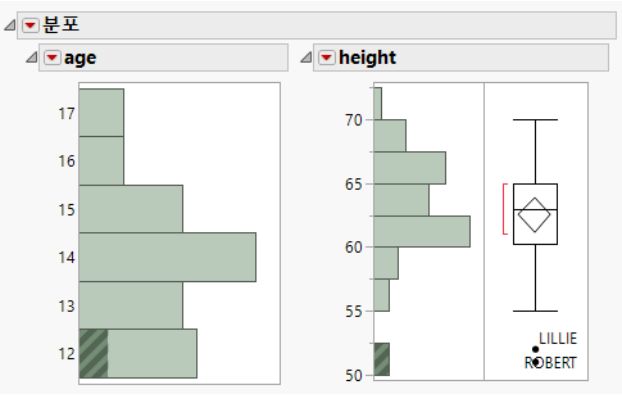
"height" 히스토그램에서 50 에 해당하는 막대를 클릭하여 잠재적 이상치를 보다 자세히 확인합니다 .

- "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.

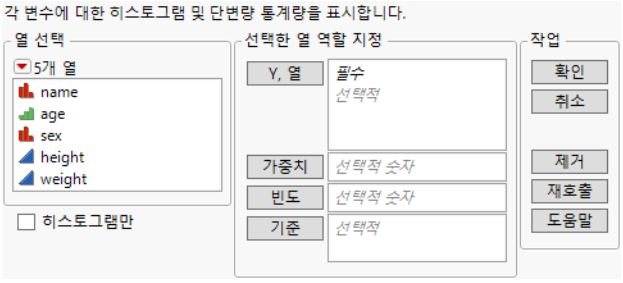
그림 3.3 라벨이 지정된 잠재적 이상치



분포 플랫폼 시작

분석 > 분포를 선택하여 분포 플랫폼을 시작할 수 있습니다.

그림 3.4 분포 시작 창



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

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

가중치 연속형 Y 의 관측값에 대한 가중치를 지정하는 변수를 할당합니다. 범주형 Y 의 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 (N 으로 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 .

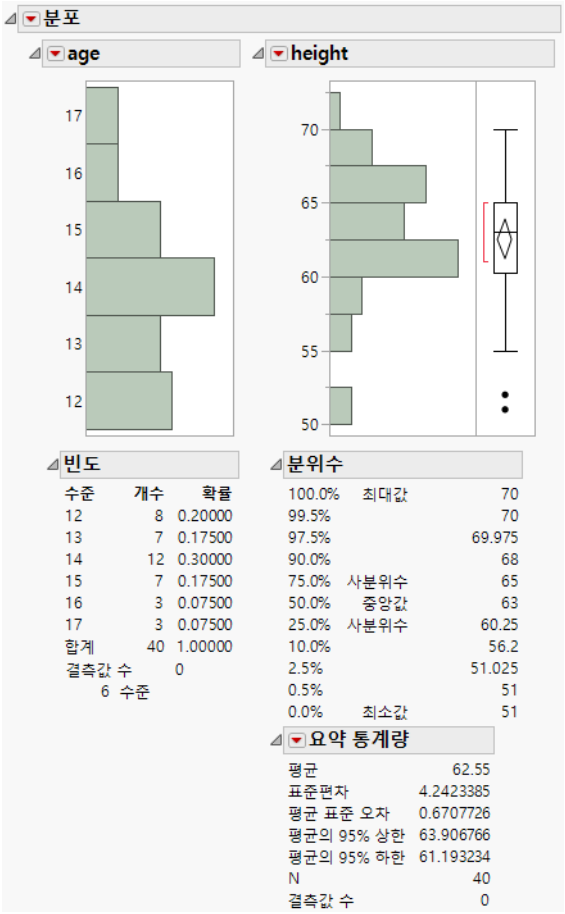
히스토그램만 보고서 창에서 히스토그램을 제외한 모든 내용을 제거합니다 .

시작 창에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

분포 보고서

초기 " 분포 " 보고서에는 각 변수에 대한 히스토그램과 보고서가 포함됩니다 .

그림 3.5 초기 분포 보고서 창



참고 : 그림 3.5 에 표시된 것과 같은 보고서를 생성하려면 " 분포 플랫폼의 예 " 에 설명된 방법을 따르십시오 .

빨간색 삼각형 메뉴

"분포" 옆의 빨간색 삼각형 메뉴에는 모든 변수에 영향을 주는 옵션이 포함되어 있습니다. 자세한 내용은 "분포 플랫폼 옵션"에서 확인하십시오.

각 변수 옆의 빨간색 삼각형 메뉴에는 해당 변수에만 영향을 주는 옵션이 포함되어 있습니다. 자세한 내용은 "범주형 변수에 대한 옵션" 또는 "연속형 변수에 대한 옵션"에서 확인하십시오.

팁 : Ctrl 키를 누른 채 변수 옵션을 선택하면 보고서에서 모델링 유형이 동일한 모든 변수에 해당 옵션이 적용됩니다 .

히스토그램 및 보고서

히스토그램에는 데이터가 시각적으로 표시됩니다. 자세한 내용은 "히스토그램"에서 확인하십시오.

범주형 변수에 대한 초기 보고서에는 "빈도" 보고서가 포함됩니다. 자세한 내용은 "빈도 보고서"에서 확인하십시오.

연속형 변수에 대한 초기 보고서에는 "분위수" 및 "요약 통계량" 보고서가 포함됩니다. 자세한 내용은 "분위수 보고서" 및 "요약 통계량 보고서"에서 확인하십시오.

히스토그램의 변수 바꾸기 , 추가 또는 제거

히스토그램의 변수를 바꾸려면 연결된 데이터 테이블의 "열" 패널에서 원하는 변수를 히스토그램의 축으로 드래그하여 놓습니다.

새 변수를 사용하여 보고서에 새 히스토그램을 생성하려면 기존 히스토그램 외부로 변수를 드래그하여 놓습니다. 새 변수와 히스토그램은 기존 히스토그램 앞, 사이 또는 뒤에 추가할 수 있습니다.

히스토그램에서 변수를 제거하려면 빨간색 삼각형 메뉴에서 **제거**를 선택하십시오.

숨겨진 행 및 제외된 행

데이터 테이블의 행에 "숨김" 행 상태를 적용하면 데이터 점을 표시하는 그림에 해당 점이 표시되지 않습니다. 하지만 히스토그램을 생성하는 데는 여전히 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.

히스토그램

분포 플랫폼에서 히스토그램을 사용하여 데이터를 시각적으로 표시할 수 있습니다 . 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 " 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 히스토그램 테두리 위로 마우스를 이동하여 양방향 화살표가 표시되도록 합니다. 그런 다음 테두리를 클릭한 채로 드래그합니다.

축 배율 조정 축 배율을 조정하려면 축을 클릭한 채로 드래그합니다.

또는 축 위로 마우스를 이동하여 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 " 에서 확인하십시오.

선택 항목 지정 여러 히스토그램에서 선택한 데이터를 지정할 수 있습니다. 자세한 내용은 " 여러 히스토그램에서 선택 영역 지정 " 에서 확인하십시오.

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

- 히스토그램을 마우스 오른쪽 버튼으로 클릭합니다. 자세한 내용은 JMP 사용에서 확인하십시오.
- 축을 마우스 오른쪽 버튼으로 클릭합니다. 라벨을 추가하거나 축을 수정할 수 있습니다. 자세한 내용은 JMP 사용의 에서 확인하십시오.
- 변수 옆의 빨간색 삼각형을 클릭하고 **히스토그램 옵션**을 선택합니다. 옵션은 변수의 모델링 유형에 따라 조금씩 다릅니다. 자세한 내용은 " 범주형 변수에 대한 옵션 " 또는 " 연속형 변수에 대한 옵션 " 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.
2. 손 도구를 히스토그램의 아무 곳에나 놓습니다.
3. 히스토그램 막대를 클릭한 채로 드래그합니다.

히스토그램은 데이터 계급화를 기반으로 합니다 . 각 히스토그램 막대의 높이는 막대로 표시된 계급에 속하는 관측값의 수에 비례합니다 . 기본 세로 방향의 히스토그램에서 다음 사항에 유의하십시오 .

- 손 도구를 왼쪽으로 움직이면 계급 폭이 증가하고 계급 수가 감소합니다 . 계급 폭이 증가하면 막대 수가 감소합니다 .
- 손 도구를 오른쪽으로 움직이면 계급 폭이 감소하고 계급 수가 증가합니다 . 계급 폭이 감소하면 막대 수가 증가합니다 .
- 손 도구를 위쪽 또는 아래쪽으로 움직이면 각 계급의 시작 값이 이동하고 각 계급에 속하는 데이터 점이 이동합니다 .

참고 : 히스토그램 방향을 가로로 변경한 경우에는 위의 방향을 반대로 적용하십시오 . 즉 , 손 도구를 아래쪽으로 움직이면 계급 폭이 증가하고 , 위쪽으로 움직이면 계급 폭이 감소하며 , 왼쪽 또는 오른쪽으로 움직이면 계급 시작 값이 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

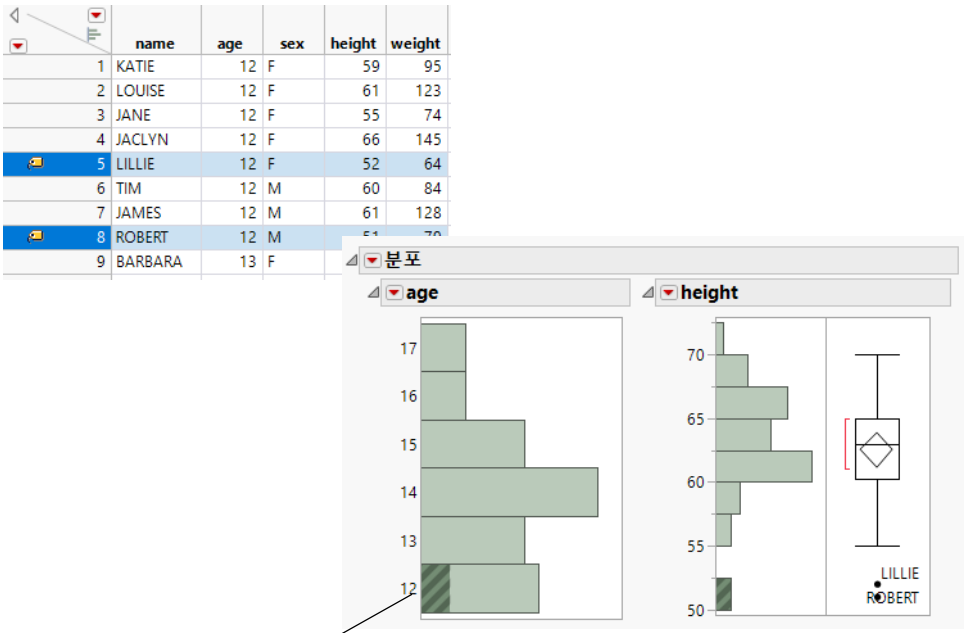
증분 옵션으로도 막대 너비를 정확하게 설정할 수 있습니다 . **증분** 옵션을 사용하려면 축을 두 번 클릭하고 "증분" 값을 변경합니다 .

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 3.6에서는 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



여러 히스토그램에서 선택 영역 지정

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"에서 확인하십시오.

빈도 보고서

명목형 및 순서형 변수의 경우 분포 플랫폼의 "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 분포 플랫폼의 "분위수" 보고서에 선택된 분위수(**백분위수**라고도 함)의 값이 나열됩니다. 통계 상세 정보는 "분위수에 대한 통계 상세 정보"에서 확인하십시오.

요약 통계량 보고서

연속형 변수의 경우 분포 플랫폼의 "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 모평균을 포함할 가능성이 매우 높은 구간을 정의합니다 .

N 비결측값의 총 개수입니다 .

결측값 수 결측 관측값의 개수입니다 .

추가 요약 통계량

가중치 합 시작 창에서 " 가중치 " 역할에 할당된 열의 합계입니다 . 가중치 합은 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 " 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 .

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 .

3* 표준편차 표준편차의 3 배 값입니다 . 환경 설정 > 플랫폼 > 분포에서 K 시그마 승수 플랫폼 환경 설정을 사용하여 승수 값을 설정할 수 있습니다 .

3* 표준편차와 평균의 합 평균에 표준편차의 3 배를 합한 값입니다 . 환경 설정 > 플랫폼 > 분포에서 K 시그마 승수 플랫폼 환경 설정을 사용하여 승수 값을 설정할 수 있습니다 .

3* 표준편차와 평균의 차 평균에서 표준편차의 3 배를 뺀 값입니다 . 환경 설정 > 플랫폼 > 분포에서 K 시그마 승수 플랫폼 환경 설정을 사용하여 승수 값을 설정할 수 있습니다 .

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-?) 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 통계량에 대한 통계 상세 정보"에서 확인하십시오.

분포 플랫폼 옵션

" 분포 " 의 빨간색 삼각형 메뉴에는 분포 플랫폼의 모든 보고서 및 그래프에 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 에서 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 .

플랫폼 환경 설정 현재 플랫폼 환경 설정을 보거나 , 현재 JMP 보고서의 설정과 일치하도록 플랫폼 환경 설정을 업데이트할 수 있는 옵션이 포함되어 있습니다 .

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

" 분포 " 보고서의 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 보고서 표시를 변경하기 위한 다음 옵션이 포함되어 있습니다 .

빈도 " 빈도 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 빈도 보고서 " 에서 확인하십시오 .

가로 레이아웃 히스토그램과 보고서의 방향을 세로 또는 가로로 변경합니다 .

왼쪽의 축 (" 가로 레이아웃 " 옵션을 선택한 경우에만 사용 가능) " 개수 ", " 확률 " 및 " 밀도 " 축을 히스토그램의 오른쪽 대신 왼쪽으로 이동합니다 .

히스토그램 옵션 자세한 내용은 " 범주형 변수에 대한 히스토그램 옵션 " 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 " 두 수준에 대한 검정 확률의 예 " 및 " 세 개 이상의 수준에 대한 검정 확률의 예 " 에서 확인하십시오 .

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

저장 자세한 내용은 " 범주형 변수에 대한 저장 옵션 " 에서 확인하십시오 .

제거 해당 변수와 모든 관련 보고서를 분포 보고서에서 영구적으로 제거합니다 .

범주형 변수에 대한 히스토그램 옵션

히스토그램 히스토그램을 표시하거나 숨깁니다. 자세한 내용은 "히스토그램"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 " 에서 확인하십시오 .

히스토그램 옵션 자세한 내용은 " 연속형 변수에 대한 히스토그램 옵션 " 에서 확인하십시오 .

정규 분위수 그림 변수의 정규 분포 정도를 쉽게 시각화할 수 있습니다 . 자세한 내용은 " 정규 분위수 그림 " 에서 확인하십시오 .

이상치 상자 그림 분포를 표시하며 가능한 이상치를 쉽게 확인할 수 있습니다 . 자세한 내용은 " 이상치 상자 그림 " 에서 확인하십시오 .

분위수 상자 그림 " 분위수 " 보고서의 특정 분위수를 표시합니다 . 자세한 내용은 " 분위수 상자 그림 " 에서 확인하십시오 .

줄기 - 잎 자세한 내용은 " 줄기 - 잎 " 에서 확인하십시오 .

CDF 그림 경험적 누적 분포 함수 그림을 생성합니다 . 자세한 내용은 "CDF 그림 " 에서 확인하십시오 .

평균 검정 평균에 대한 1표본 검정을 수행합니다 . 자세한 내용은 "평균 검정"에서 확인하십시오 .

표준편차 검정 표준편차에 대한 1 표본 검정을 수행합니다 . 자세한 내용은 " 표준편차 검정 " 에서 확인하십시오 .

동등성 검정 모평균이 가설 값과 동등한지 평가합니다 . 자세한 내용은 " 동등성 검정 " 에서 확인하십시오 .

신뢰 구간 평균 및 표준편차의 신뢰 구간을 선택합니다 . 자세한 내용은 " 신뢰 구간 " 에서 확인하십시오 .

예측 구간 단일 관측값이나 다음 무작위 선택 표본의 평균 및 표준편차에 대한 예측 구간을 선택합니다 . 자세한 내용은 " 예측 구간 " 에서 확인하십시오 .

공차 구간 지정된 모비율 이상을 포함하기 위한 구간을 계산합니다 . 자세한 내용은 " 공차 구간 " 에서 확인하십시오 .

공정 능력 (Y 열에 " 감지 한계 " 열 특성이 있는 경우에는 사용 불가능) 지정된 규격 한계에 대한 공정 적합성을 측정합니다 . 자세한 내용은 " 공정 능력 " 에서 확인하십시오 .

> **참고 :** Y 열에 " 감지 한계 " 열 특성이 있는 경우 적합 분포의 빨간색 삼각형 메뉴에서만 공정 능력 분석을 사용할 수 있습니다 . 자세한 내용은 " 공정 능력 " 에서 확인하십시오 .

연속형 적합 연속형 변수에 분포를 적합시킵니다 . Y 열에 " 감지 한계 " 열 특성이 있는 경우 " 연속형 적합 " 옵션은 중도절단 분포를 적합시키고 분포의 일부만 사용할 수 있습니다 . 자세한 내용은 " 분포 적합 " 에서 확인하십시오 .

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

저장 연속형 또는 범주형 변수에 대한 정보를 저장합니다 . 자세한 내용은 " 연속형 변수에 대한 저장 옵션 " 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 " 에서 확인하십시오 .

분위수 증분 설정 분위수 증분을 변경하거나 기본 분위수 증분으로 되돌립니다 .

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

- 가중 평균 분위수의 추정 방법에 대한 자세한 내용은 " 분위수에 대한 통계 상세 정보 " 에서 확인하십시오 .
- 가중 평균 분위수의 분포 무관 신뢰 한계에 대한 자세한 내용은 Meeker et al. 연구 자료의 섹션 5.2 (2017) 에서 확인하십시오 .
- 평활된 경험적 가능도 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 평활된 경험적 가능도 분위수를 사용할 수 없습니다 .

요약 통계량 " 요약 통계량 " 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 요약 통계량 보고서 " 에서 확인하십시오 .

요약 통계량 사용자 정의 " 요약 통계량 " 보고서의 통계량을 추가하거나 제거합니다 . 자세한 내용은 " 요약 통계량 보고서 " 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 레이아웃**이 선택된 경우에만 적용할 수 있습니다 .

연속형 변수에 대한 히스토그램 옵션

히스토그램 히스토그램을 표시하거나 숨깁니다. 자세한 내용은 "히스토그램"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 " 및 " 표준 오차 막대에 대한 통계 상세 정보 " 에서 확인하십시오 .

계급 폭 설정 히스토그램 막대의 계급 폭을 변경합니다 . 자세한 내용은 " 연속형 변수의 히스토그램 막대 크기 조정 " 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 .

확률 = (막대 너비) * 밀도

개수 = (막대 너비) * 밀도 * (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 .

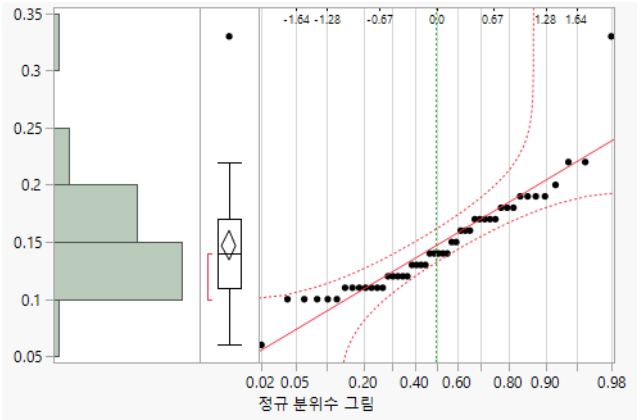
개수 표시 각 히스토그램 막대가 나타내는 열 값의 빈도를 라벨로 표시합니다 .

정규 분위수 그림

분포 플랫폼의 "정규 분위수 그림" 옵션을 사용하여 변수의 정규 분포 범위를 시각화할 수 있습니다. 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 보여줍니다.

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 .

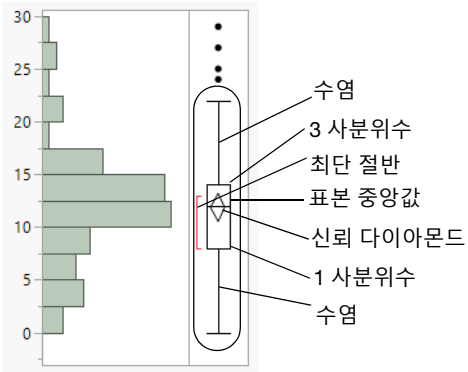
통계 상세 정보는 "정규 분위수 그림에 대한 통계 상세 정보"에서 확인하십시오.

이상치 상자 그림

분포 플랫폼의 "이상치 상자 그림" 옵션을 사용하여 분포를 확인하고 가능한 이상치를 식별할 수 있습니다. 일반적으로 상자 그림에는 연속형 분포 중 선택된 분위수가 표시됩니다.

참고 : 이상치 상자 그림을 Tukey 이상치 상자 그림이라고도 합니다 .

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 사분위수**라고도 합니다 .
- 1 사분위수와 3 사분위수 사이의 차이를 **사분위수 범위** . 라고 합니다 .
- 상자에는 양 끝에서 시작되는 선이 있으며 이 선은 **수염**이라고도 합니다 . 수염은 상자 끝에서 시작해서 다음과 같이 계산된 거리 내에 있는 점들 중 가장 바깥쪽 데이터 점까지 이어집니다 .

  1 사분위수 - 1.5 * (사분위수 범위)

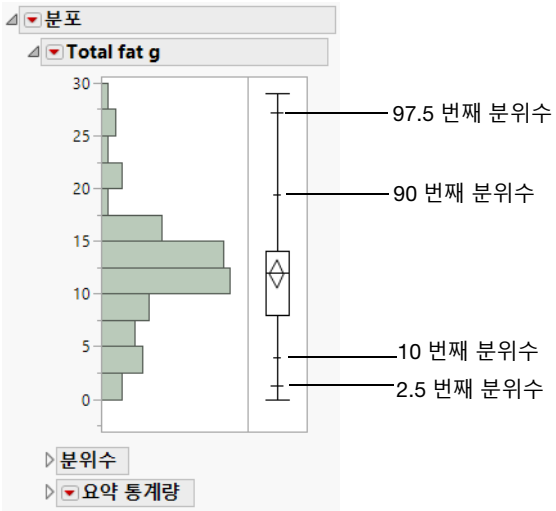
  3 사분위수 + 1.5 * (사분위수 범위)

  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 절반**을 나타냅니다 (Rousseeuw and Leroy 1987).
- 이상치 상자 그림에서 개체를 제거하는 방법은 " 이상치 또는 분위수 상자 그림에서 개체 제거 " 에서 확인하십시오 .

분위수 상자 그림

분포 플랫폼의 "분위수 상자 그림" 옵션을 사용하여 "분위수" 보고서에서 특정 분위수를 표시할 수 있습니다. 따라서 분포가 대칭적인지 여부를 한눈에 알 수 있습니다. 분포가 대칭적인 경우 상자 그림의 분위수는 서로 거의 같은 거리로 나타납니다. 예를 들어 분위수 표식이 한 쪽 끝에 밀접해서 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, p번째 분위수는 p%의 값보다 큰 값입니다. 예를 들어 데이터의 10%는 10번째 분위수 아래에 있으며, 데이터의 90%는 90번째 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 에서 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

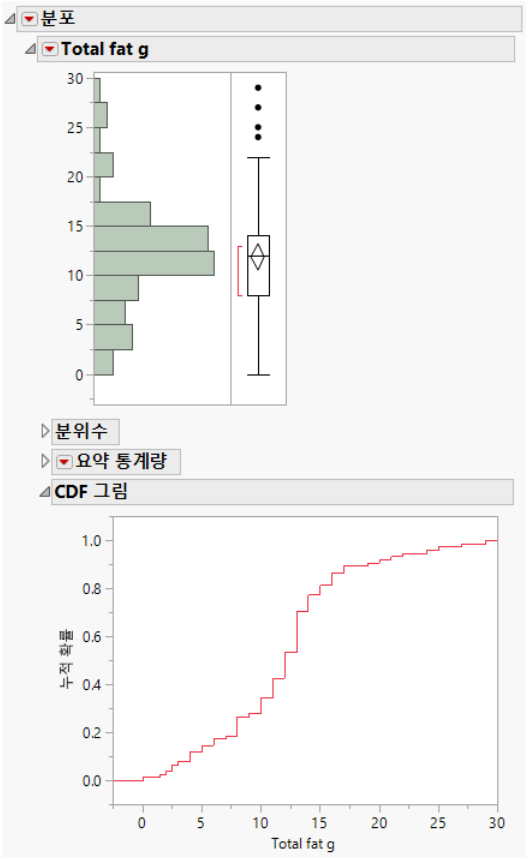
분포 플랫폼의 "줄기-잎" 옵션을 사용하여 데이터를 대화식 텍스트 그림으로 표시할 수 있습니다. 이 그림의 각 행에는 열 값 범위의 선행 숫자를 나타내는 **줄기** 값이 있습니다. **잎** 값은 값의 뒷자리 숫자로 이루어집니다. 줄기와 잎을 결합하면 데이터 점을 확인할 수 있습니다. 일부 경우에 줄기-잎 그림의 숫자는 테이블의 실제 데이터에 대한 근사값입니다. 줄기-잎 그림은 마우스 버튼을 클릭하거나 브러시 도구를 사용할 때 그에 따라 변경됩니다.

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

CDF 그림

분포 플랫폼의 "CDF 그림" 옵션을 사용하여 경험적 누적 분포 함수 그림을 생성할 수 있습니다. 가로 축의 특정 값보다 작거나 같은 데이터의 백분율을 확인하려면 CDF 그림을 사용합니다. 데이터에 분포가 적합되면 적합된 분포에 대한 누적 분포 함수가 CDF 그림에 중첩 표시됩니다.

그림 3.10 CDF 그림



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

평균 검정

분포 플랫폼의 "평균 검정" 옵션을 사용하여 평균에 대한 1표본 검정 옵션을 지정하고 검정을 수행할 수 있습니다. 표준편차 값을 지정하면 *z*-검정이 수행됩니다. 그렇지 않으면 표본 표준편차를 사용하여 *t*-검정이 수행됩니다. 비모수 Wilcoxon 부호 순위 검정을 요청할 수도 있습니다.

"평균 검정" 옵션을 반복적으로 사용하여 다른 여러 값을 검정할 수 있습니다. 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 귀무가설로 하며 , 분포가 대칭적이라고 가정합니다 . 자세한 내용은 "Wilcoxon 부호 순위 검정에 대한 통계 상세 정보 " 에서 확인하십시오 .

확률 값

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

Prob < t 모평균이 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

분포 플랫폼의 " 표준편차 검정 " 옵션을 사용하여 표준편차에 대한 1 표본 검정을 수행할 수 있습니다 . 이 옵션을 반복적으로 사용하여 다른 여러 값을 검정할 수 있습니다 . 표준편차를 검정할 때마다 새로운 " 표준편차 검정 " 보고서가 표시됩니다 .

검정 통계량 카이제곱 검정 통계량의 값을 제공합니다 . 자세한 내용은 " 표준편차 검정에 대한 통계 상세 정보 " 에서 확인하십시오 .

최소 p 값 모표준편차가 가설 값과 다르지 않을 때 보다 극단적인 카이제곱 값을 얻을 확률입니다 . 자세한 내용은 " 표준편차 검정에 대한 통계 상세 정보 " 에서 확인하십시오 .

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

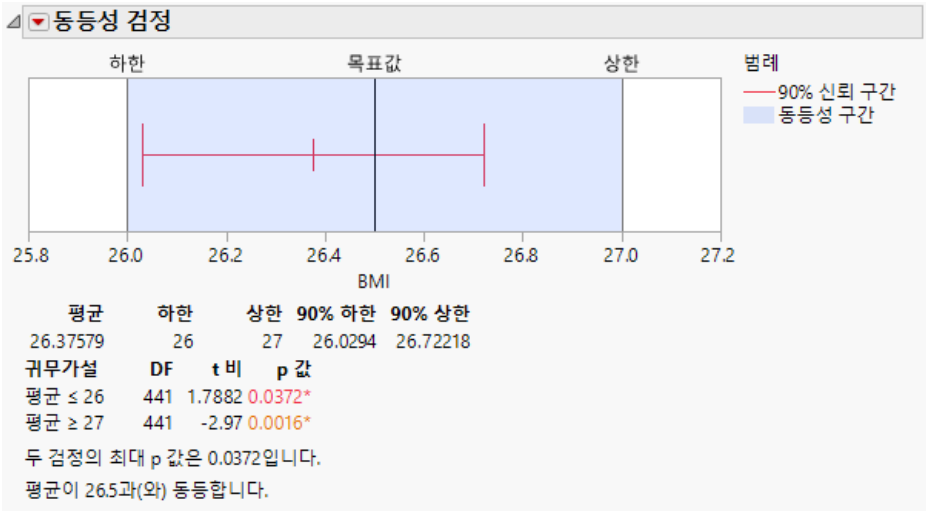
동등성 검정

분포 플랫폼의 "동등성 검정" 옵션을 사용하면 모평균이 가설 값과 동일한지 여부를 평가할 수 있습니다. 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 실제적으로 가설 값과 동등한 것으로 간주할 수 있다는 결론을 내릴 수 있습니다.

"동등성 검정" 옵션을 선택할 때는 가설 평균, 임계 차이(실제적으로 0으로 간주되는 차이) 및 신뢰 수준을 지정합니다. 신뢰 수준은 1-α이며, 여기서 α는 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, 실제적으로 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 보여 주는 정의된 동등성 구간의 그림
- 계산된 평균의 신뢰 구간 . 이 신뢰 구간은 1 - 2 * 유의 수준 구간입니다 .
- 계산된 평균 , 지정된 하한 및 상한 경계 , 평균의 (1 - 2 * 알파) 수준 신뢰 구간을 보여 주는 테이블
- 두 개의 단측 검정 결과를 보여 주는 테이블
- 결과를 요약하고 , 평균을 목표값과 동등한 것으로 간주할 수 있는지 여부를 언급하는 노트

신뢰 구간

분포 플랫폼의 "신뢰 구간" 옵션을 사용하여 평균 및 표준편차에 대한 신뢰 구간을 표시할 수 있습니다. **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

분포 플랫폼의 "예측 구간" 옵션을 사용하여 단일 관측값 또는 무작위 추출된 다음 표본의 평균과 표준편차에 대한 구간을 표시할 수 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 구간에 대한 통계 상세 정보 " 에서 확인하십시오 .
- 예는 " 예측 구간의 예 " 에서 확인하십시오 .

공차 구간

분포 플랫폼의 "공차 구간" 옵션을 사용하여 지정된 모비율 이상을 포함하는 구간을 표시할 수 있습니다. 이것은 지정된 모비율에 대한 구간 추정값입니다. 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.

참고 : " 정규 분포 가정 " 옵션을 지정하면 구간에 사용된 k 값이 보고서 테이블에 포함됩니다 . 값을 보려면 " 공차 구간 " 보고서에서 테이블을 마우스 오른쪽 버튼으로 클릭하고 열 > K 요인을 선택합니다 .

관련 정보

- 통계 상세 정보는 " 공차 구간에 대한 통계 상세 정보 " 에서 확인하십시오 .
- 예는 " 공차 구간의 예 " 에서 확인하십시오 .

공정 능력

분포 플랫폼의 "공정 능력" 옵션을 사용하여 주어진 규격 한계와 비교하여 공정이 얼마나 잘 수행되고 있는지 측정하는 분석을 표시할 수 있습니다. 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 : 여러 열에 대하여 규격 한계를 한 번에 추가하는 방법은 **품질 및 공정 방법**의 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.

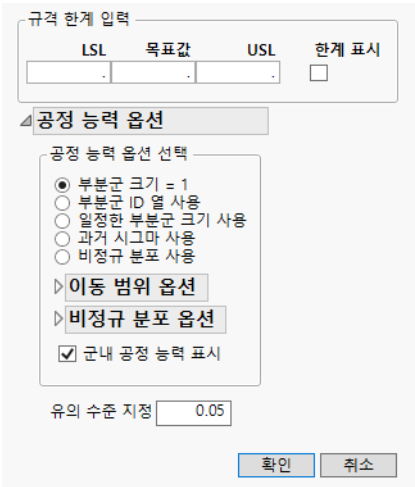
보고서의 규격 한계를 데이터 테이블에 열 특성으로 저장하려면 "공정 능력"의 빨간색 삼각형에서 **규격 한계를 열 특성으로 저장**을 선택합니다. 공정 능력 분석을 반복할 때 저장된 규격 한계를 자동으로 가져오게 됩니다.

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 1 로 설정하고 추가 이동 범위 옵션을 제공합니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

부분군 ID 열 사용 부분군 ID 열을 선택하고 추가 부분군 지정 및 이동 범위 옵션을 제공합니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

일정한 부분군 크기 사용 일정한 부분군 크기를 설정하고 추가 부분군 지정 및 이동 범위 옵션을 제공합니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

과거 시그마 사용 과거에 승인된 시그마 값을 할당합니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

비정규 분포 사용 비정규 분포를 선택하고 추가 비정규 분포 옵션을 제공합니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

군내 공정 능력 표시 (" 부분군 크기 " 가 1 이거나 " 공정 능력 " 옵션에서 " 비정규 분포 사용 " 을 선택한 경우에만 사용 가능합니다 .) 군내 표준편차의 추정값을 보고서에 표시할지 여부를 지정합니다 .

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 **품질 및 공정 방법**의 에서 확인하십시오.

- 통계 상세 정보는 **품질 및 공정 방법**의 에서 확인하십시오 .
- 예는 " 공정 능력의 예 " 에서 확인하십시오 .
- " 공정 능력 " 플랫폼에 대한 자세한 내용은 **품질 및 공정 방법**의 에서 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 " 에서 확인하십시오 .

열에 " 분포 " 열 특성이 포함된 경우 분포 보고서에서 해당 열 특성의 분포는 기본값으로 적합됩니다 .

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 에서 확인하십시오 .

연속형 적합

" 연속형 적합 " 하위 메뉴에는 연속형 분포를 적합시키기 위한 옵션이 포함되어 있습니다 . 이러한 분포의 파라미터화 방법에 대한 자세한 내용은 " 연속형 적합 분포에 대한 통계 상세 정보 " 에서 확인하십시오 .

정규 적합 데이터에 정규 분포를 적합시킵니다 . 정규 분포는 대개 대부분의 값이 곡선의 가운데에 있는 대칭적 데이터를 모델링하는 데 사용됩니다 . 정규 분포에 대한 모수 추정은 비편향 추정값을 사용합니다 .

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

스튜던트 t 적합 스튜던트 *t* 분포를 데이터에 적합시킵니다 . 스튜던트 *t* 분포는 정규 분포와 Cauchy 분포 사이의 공간을 차지하는 로버스트 옵션입니다 . 스튜던트 *t* 분포의 자유도가 무한대에 가까워지면 정규 분포와 동등합니다 . 스튜던트 *t* 분포의 자유도가 1 이면 Cauchy 분포와 동등합니다 . 분포 플랫폼은 자유도 값을 추정합니다 .

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 . 로그 정규 분포에 대한 모수 추정은 최대 가능도 추정값을 사용합니다 .

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 추정을 사용하여 평활 곡선을 적합시킵니다 (" 커널 평활기 보고서 "). " 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 Sl) 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 15 에서 분포 적합의 일부 기능이 업데이트되었습니다 . 이 옵션을 활성화하면 호환성을 위해 유지된 이전 JMP 릴리스의 기능을 사용할 수 있습니다 . 이러한 레거시 적합기에 대한 설명서는 JMP 16.1 도움말의 "Details for the Legacy Distribution Fitters" 섹션에서 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 분포에 대한 통계 상세 정보 " 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 시행 (단 , 모든 시행의 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 " 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 이항 분포에는 사용 불가능) 적합된 분포에 대한 적합도 검정이 포함된 적합도 검정 보고서를 표시하거나 숨깁니다.

연속형 적합의 경우 적합도 검정은 Anderson-Darling 검정입니다. 검정의 *p* 값은 Stephens 연구 자료 (1974) 의 섹션 4.1 에 설명된 절차와 유사하게 모수 붓스트랩을 사용하여 시뮬레이션됩니다. 정규 분포의 경우 표본 크기가 2,000 보다 작고 고정 모수가 없으면 Shapiro-Wilk 정규성 검정도 보고됩니다.

이산형 적합의 경우 적합도 검정은 Pearson 카이제곱 검정입니다.

모수 고정 (Johnson 분포 또는 평활 곡선 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: 가능도비) 보고서도 표시됩니다.

공정 능력 (Cauchy, 스튜던트 *t* 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 **품질 및 공정 방법**의 에서 확인하십시오 .

" 비정규 분포 옵션 " 개요에는 비정규 공정 능력 계산에 사용되는 방법을 선택할 수 있는 옵션이 포함되어 있습니다 . 자세한 내용은 **품질 및 공정 방법**의 에서 확인하십시오 .

" 공정 능력 " 옵션 및 보고서에 대한 자세한 내용은 **품질 및 공정 방법**의 에서 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

분포 플랫폼의 "저장" 메뉴 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>	각 관측값의 수준 수는 해당 관측값이 포함된 히스토그램 막대에 해당합니다. 히스토그램 막대는 낮은 것부터 높은 것 순서로 1로 시작하는 번호가 매겨집니다. **참고 :** 소스 정보를 유지하기 위해 새 열에 값 라벨을 추가할 수 있지만 값 라벨은 기본적으로 해제되어 있습니다 .
수준 중간점	중간점 <열 이름>	각 관측값의 중간점 값은 수준 하한에 수준 너비의 1/2을 더해서 계산됩니다. **참고 :** 소스 정보를 유지하기 위해 새 열에 값 라벨을 추가할 수 있지만 값 라벨은 기본적으로 해제되어 있습니다 .

표 3.1 저장 옵션 설명 (계속)

옵션	데이터 테이블에 추가되는 열	설명
순위	순위화 <열 이름>	해당 열의 각 값에 대한 순위(1부터 시작)를 제공합니다. 중복 반응 값의 경우 데이터 테이블에서의 발생 순서에 따라 연속 순위가 할당됩니다.
순위 평균	평균 순위 <열 이름>	값이 고유한 경우에는 평균 순위와 순위가 동일합니다. 값이 k번 나타나는 경우 평균 순위는 값의 순위 합계를 k로 나눠서 계산됩니다.
확률 스코어	확률 <열 이름>	비결측 스코어가 N개인 경우 값의 확률 스코어는 해당 값의 평균 순위를 $N+1$로 나눠서 계산됩니다. 이 열은 경험적 누적 분포 함수와 유사합니다.
정규 분위수	N-분위수 <열 이름>	정규 분위수를 저장합니다. 자세한 내용은 "정규 분위수 그림에 대한 통계 상세 정보"에서 확인하십시오.
표준화	표준화 <열 이름>	표준화된 값을 저장합니다. 자세한 내용은 "표준화된 데이터 저장을 위한 통계 상세 정보"에서 확인하십시오.
중심화됨	중심화 <열 이름>	0을 기준으로 중심화한 값을 저장합니다.
로버스트 표준화	로버스트 표준화 <열 이름>	로버스트 평균을 중심으로 중심화되고 로버스트 표준편차를 사용하여 표준화된 반응 값을 포함하는 열을 저장합니다.
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

이 섹션에는 분포 플랫폼을 사용한 예가 포함되어 있습니다 .

- " 예 : 여러 히스토그램에서 데이터 선택 "
- " 기준 변수의 예 "
- " 두 수준에 대한 검정 확률의 예 "
- " 세 개 이상의 수준에 대한 검정 확률의 예 "
- " 예측 구간의 예 "
- " 공차 구간의 예 "

- "공정 능력의 예"

예 : 여러 히스토그램에서 데이터 선택

이 예에서는 한 번에 여러 히스토그램에서 데이터를 선택하고 강조 표시하는 방법을 보여 줍니다.

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

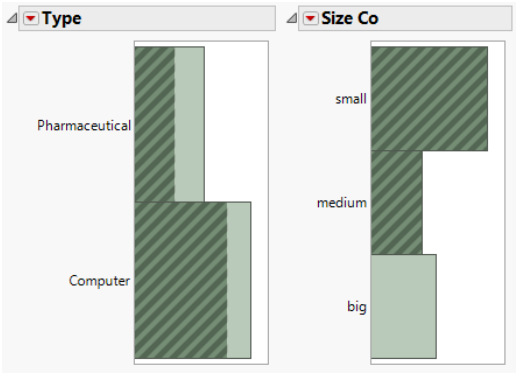
   소규모 회사의 유형 분포를 확인하려고 합니다 .
5. "small" 옆의 막대를 클릭합니다 .

   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 .

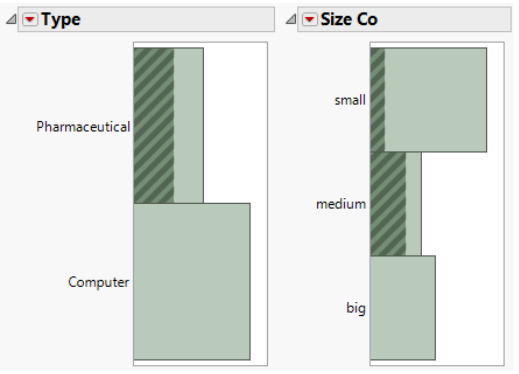
   소규모 및 중간 규모 회사의 유형 분포를 확인할 수 있습니다 . 그림 3.13 의 왼쪽을 참조하십시오 . 선택 범위를 좁히기 위해 소규모 및 중간 규모의 제약 회사만 확인하려고 합니다 .
7. Ctrl 및 Shift 키 (Windows 의 경우) 또는 Command 및 Shift 키 (macOS 의 경우) 를 누른 채로 있습니다 . "Type" 히스토그램에서 "Computer" 막대를 클릭하여 선택 취소합니다 .

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

기준 변수의 예

이 예에서는 기준 변수를 사용하여 범주형 변수의 각 수준에 대해 별도의 분석을 생성하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 폴더**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 .

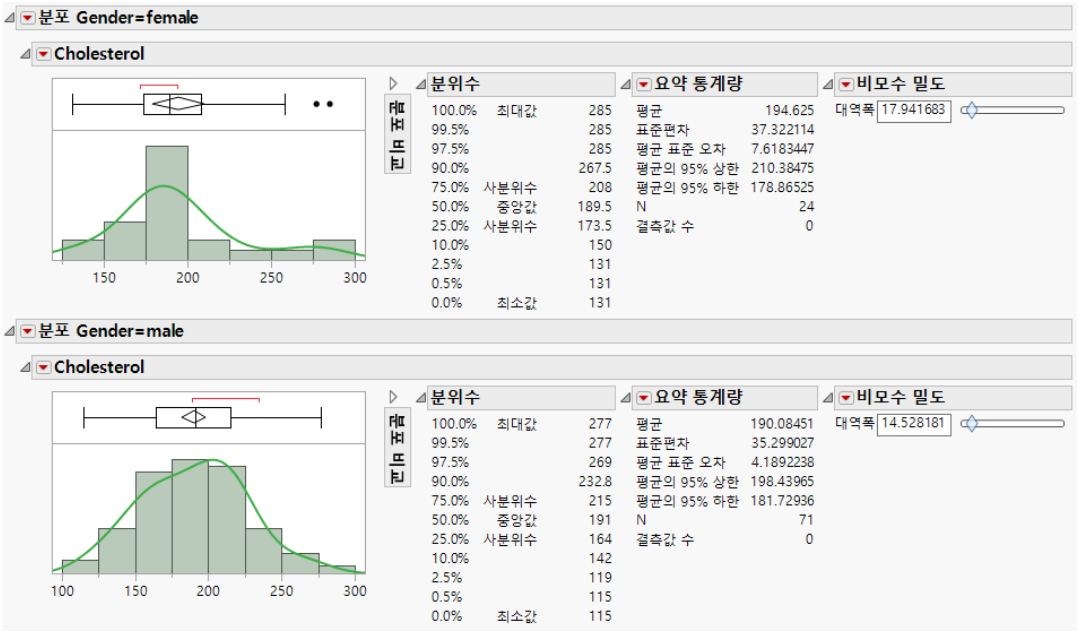
두 히스토그램 모두에 평활 곡선을 추가합니다.

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



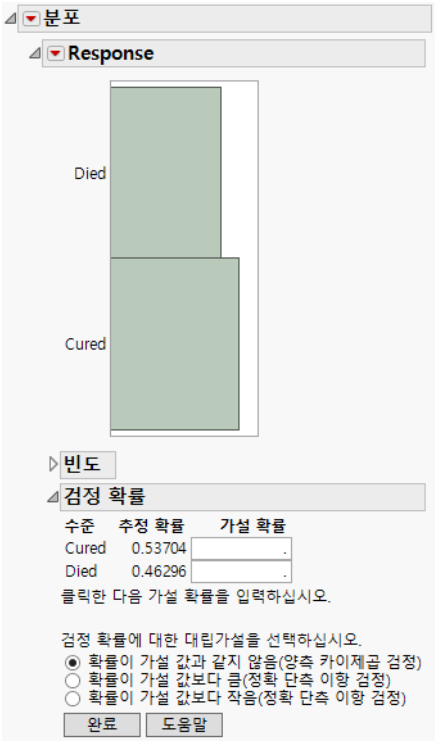
두 수준에 대한 검정 확률의 예

이 예에서는 수준이 정확히 두 개인 변수에 대해 분포 플랫폼에서 " 검정 확률 " 보고서를 생성하는 방법을 보여 줍니다 .

검정 확률 보고서 시작

1. **도움말 > 샘플 데이터 폴더**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 수준이 정확히 두 개인 변수에 대한 검정 확률 보고서 옵션

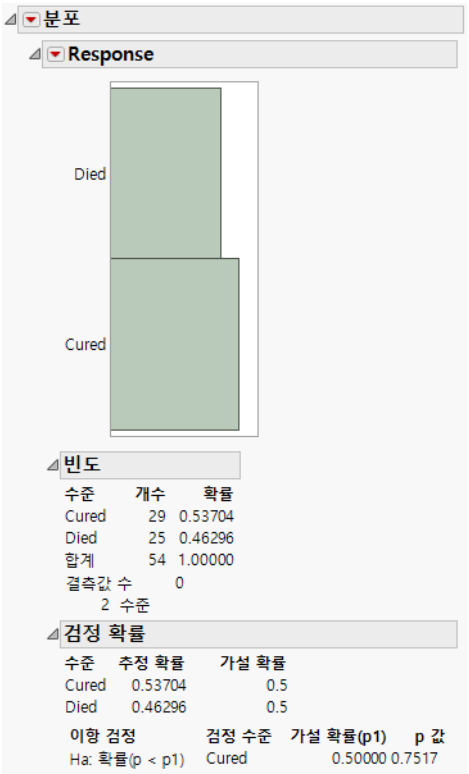


검정 확률 보고서 생성

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

이항 검정의 정확 확률이 계산됩니다 .

그림 3.16 수준이 정확히 두 개인 변수에 대한 검정 확률 보고서의 예



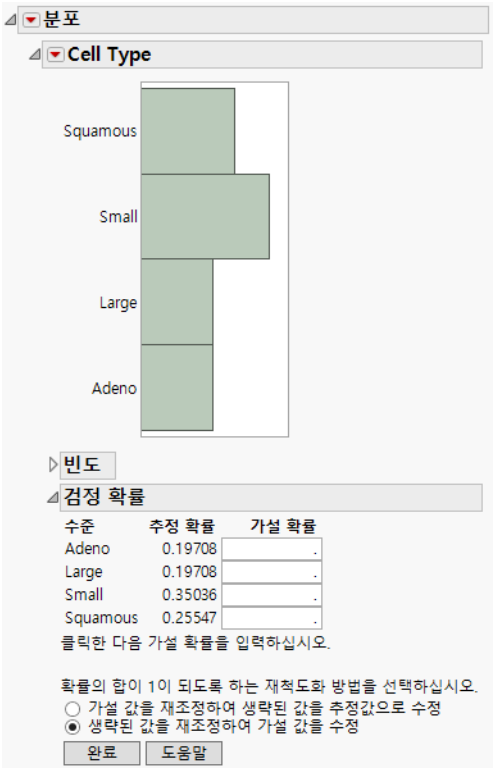
세 개 이상의 수준에 대한 검정 확률의 예

이 예에서는 수준이 세 개 이상인 변수에 대해 분포 플랫폼에서 " 검정 확률 " 보고서를 생성하는 방법을 보여 줍니다 .

검정 확률 보고서 시작

1. **도움말 > 샘플 데이터 폴더**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 3.17 수준이 세 개 이상인 변수에 대한 검정 확률 보고서 옵션



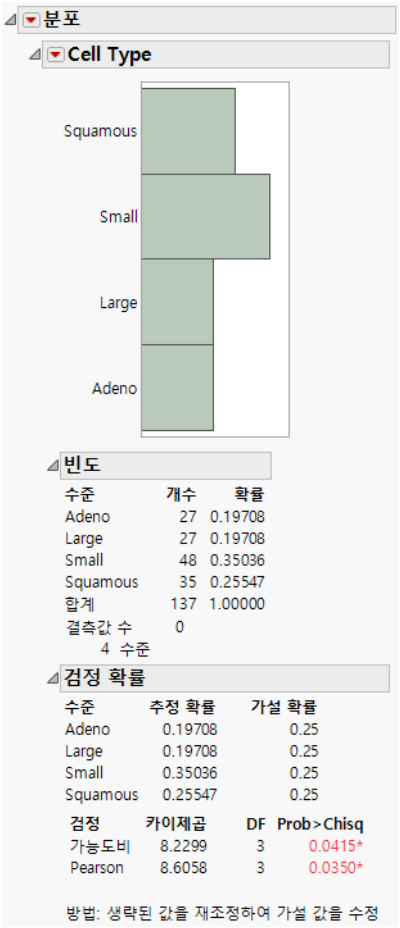
검정 확률 보고서 생성

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

1. 네 개의 " 가설 확률 " 필드 모두에 0.25 를 입력합니다 .
2. **생략된 값을 재조정하여 가설 값을 수정**을 선택합니다 .
3. **완료**를 클릭합니다 .

   가능도비 및 Pearson 카이제곱 검정이 계산됩니다 .

그림 3.18 수준이 세 개 이상인 변수에 대한 검정 확률 보고서

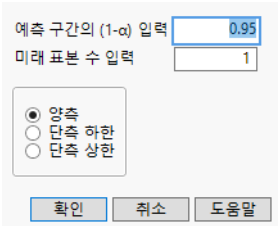


예측 구간의 예

이 예에서는 " 분포 " 보고서에 예측 구간을 추가하는 방법을 보여 줍니다 . 다음 10 개의 오존 수준 관측값에 대한 예측 구간을 계산하는 데 관심이 있다고 가정해 보겠습니다 .

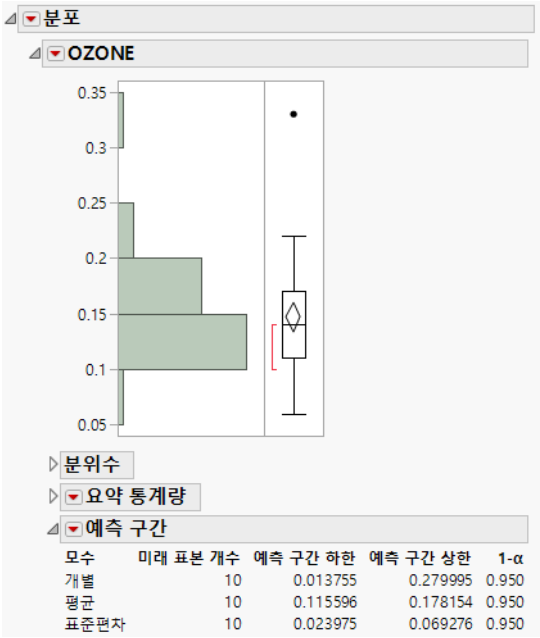
1. **도움말 > 샘플 데이터 폴더**를 선택하고 Cities.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. OZONE 을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "OZONE" 의 빨간색 삼각형을 클릭하고 **예측 구간**을 선택합니다 .

그림 3.19 예측 구간 창



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

그림 3.20 예측 구간 보고서의 예



이 예에서는 다음 사항을 95% 신뢰할 수 있습니다 .

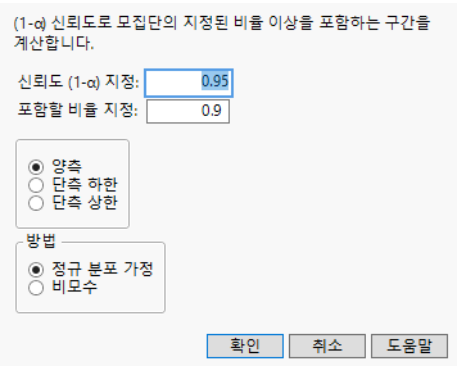
- 다음 10 개의 관측값은 각각 0.013755 에서 0.279995 사이에 있습니다 .
- 다음 10 개 관측값의 평균은 0.115596 에서 0.178154 사이에 있습니다 .
- 다음 10 개 관측값의 표준편차는 0.023975 에서 0.069276 사이에 있습니다 .

공차 구간의 예

이 예에서는 " 분포 " 보고서에 공차 구간을 추가하는 방법을 보여 줍니다 . 오존 수준 측정값의 90% 를 포함하는 구간을 추정하려고 한다고 가정해 보겠습니다 .

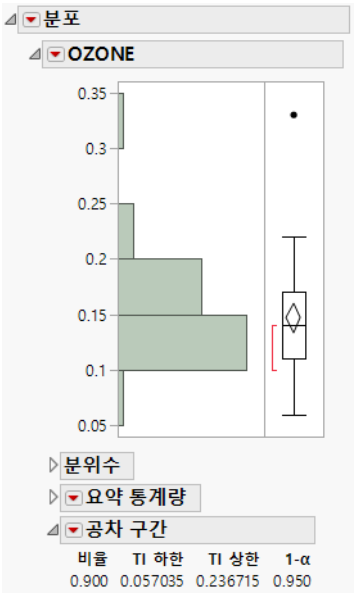
1. **도움말 > 샘플 데이터 폴더**를 선택하고 Cities.jmp 를 엽니다 .
2. **분석 > 분포**를 선택합니다 .
3. OZONE 을 선택하고 **Y, 열**을 클릭합니다 .
4. **확인**을 클릭합니다 .
5. "OZONE" 의 빨간색 삼각형을 클릭하고 **공차 구간**을 선택합니다 .

그림 3.21 공차 구간 창



6. 기본 선택 사항을 유지하고 **확인**을 클릭합니다 .

그림 3.22 공차 구간 보고서의 예



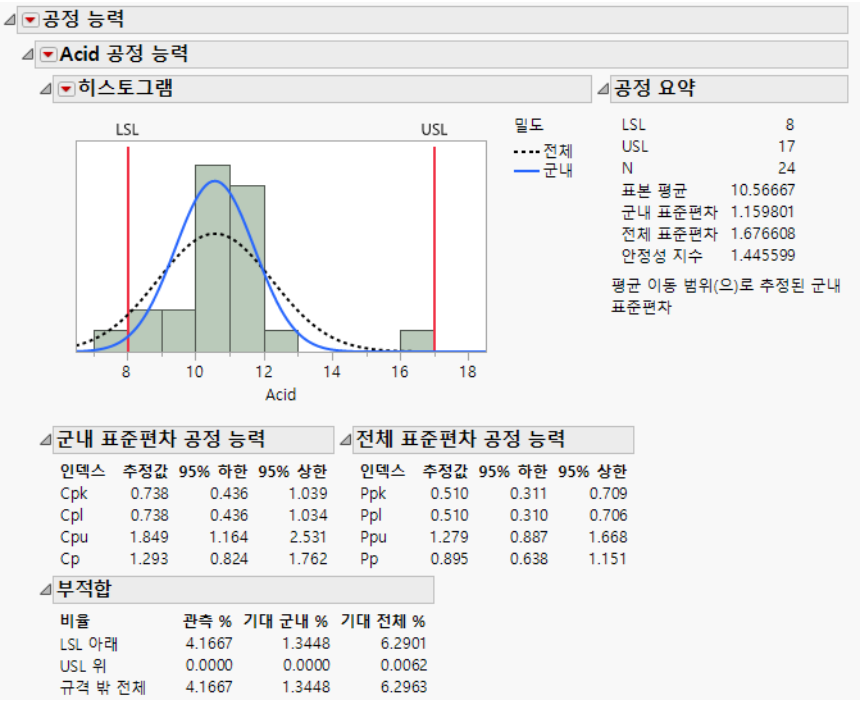
이 예에서는 TI(공차 구간) 하한 및 TI 상한 값에 따라 모집단의 90% 이상이 0.057035 에서 0.236715 사이에 있음을 95% 신뢰할 수 있습니다.

공정 능력의 예

이 예에서는 " 분포 " 보고서에 공정 능력 결과를 추가하는 방법을 보여 줍니다. 피클의 산도를 특성화하려고 한다고 가정해 보겠습니다. 규격 하한과 규격 상한은 각각 8 과 17 입니다.

1. **도움말 > 샘플 데이터 폴더**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 3.23 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

이 섹션에는 분포 플랫폼에 대한 통계 상세 정보가 포함되어 있습니다 .

- " 표준 오차 막대에 대한 통계 상세 정보 "
- " 분위수에 대한 통계 상세 정보 "
- " 요약 통계량에 대한 통계 상세 정보 "
- " 정규 분위수 그림에 대한 통계 상세 정보 "
- "Wilcoxon 부호 순위 검정에 대한 통계 상세 정보 "
- " 표준편차 검정에 대한 통계 상세 정보 "
- " 정규 분위수에 대한 통계 상세 정보 "
- " 표준화된 데이터 저장을 위한 통계 상세 정보 "
- " 예측 구간에 대한 통계 상세 정보 "

- "공차 구간에 대한 통계 상세 정보"
- "연속형 적합 분포에 대한 통계 상세 정보"
- "이산형 적합 분포에 대한 통계 상세 정보"

표준 오차 막대에 대한 통계 상세 정보

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

분위수에 대한 통계 상세 정보

이 섹션에서는 분포 플랫폼의 분위수 계산 방법에 대해 설명합니다.

한 열에 있는 n 개의 비결측값 중 p 번째 분위수를 계산하려면 n 개의 값을 오름차순으로 배열하며 이 열 값을 $y_1, y_2, ..., y_n$ 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 정의됩니다.

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

  다음은 각 요소에 대한 설명입니다.
  - n 은 변수의 비결측값 수입니다.
  - $y_1, y_2, ..., y_n$ 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.

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

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 통계량에 대한 통계 상세 정보

이 섹션에는 " 분포 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 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 .

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

$\bar{y}_w$ = 가중 평균

평균 표준 오차

평균 표준 오차는 표본 표준편차 s를 N의 제곱근으로 나눠서 계산됩니다. 시작 창에서 가중치 또는 빈도 열을 지정한 경우에는 가중치 또는 빈도 합의 제곱근이 분모가 됩니다.

왜도

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

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

또한 w_i는 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_i는 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 그림에 대한 통계 상세 정보

분포 플랫폼에서 각 정규 분위수 그림 값에 대한 경험적 누적 확률은 다음과 같이 계산됩니다 .

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

여기서 r_i는 i번째 관측값의 순위이고 , N은 비결측이면서 제외되지 않은 관측값의 개수입니다 .

정규 분위수 값은 다음과 같이 계산됩니다 .

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

여기서 Φ는 정규 분포의 누적 확률 분포 함수입니다 .

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

Wilcoxon 부호 순위 검정에 대한 통계 상세 정보

분포 플랫폼에서 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 .
  $X_1, X_2, \ldots, X_N$
- 귀무가설은 다음과 같습니다 .
  H_0: X 의 분포가 m 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 .
  $X_1, X_2, \ldots, X_N$ 및 $Y_1, Y_2, \ldots, Y_N$
- 귀무가설은 다음과 같습니다 .
  H_0: $X - Y$ 의 분포가 0 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 순위**가 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 정확 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 > 0인 경우에 대한 표기인 d_i는 0이 아닌 부호 순위의 i번째 그룹에 있는 값의 개수를 나타냅니다. 특정 부호 순위에 대해 동일 값이 없으면 $d_i = 1$이고 합은 0입니다.

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 검정에 대한 통계 상세 정보

분포 플랫폼에서 표준편차 검정의 검정 통계량은 다음과 같이 계산됩니다 .

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

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

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

2 * min(p1,p2)

여기서 p1은 아래쪽 꼬리의 p 값이고 p2는 위쪽 꼬리의 p 값입니다.

정규 분위수에 대한 통계 상세 정보

분포 플랫폼의 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 저장을 위한 통계 상세 정보

분포 플랫폼의 표준화된 값은 다음과 같이 계산됩니다 .

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

다음은 각 요소에 대한 설명입니다 .

X - 원래 열입니다 .
$\overline{X}$ - X 열의 평균입니다 .
S_X - X 열의 표준편차입니다 .

예측 구간에 대한 통계 상세 정보

분포 플랫폼의 예측 구간은 다음과 같이 정의됩니다 .

- 미래 관측값이 m 개인 경우 :

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

- 미래 관측값 m 개의 평균을 사용하는 경우 :

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

- 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-α 를 사용하여 구합니다 .

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

공차 구간에 대한 통계 상세 정보

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

정규 분포 기반의 구간

단측 구간

단측 구간은 다음과 같이 계산됩니다 .

하한 = $\bar{x} - k's$

상한 = $\bar{x} + k's$

다음은 각 요소에 대한 설명입니다 .

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

s - 표준편차입니다 .

t - 비중심 t 분포의 분위수입니다 .

Φ^{-1} - 표준 정규 분위수입니다 .

양측 구간

양측 구간은 다음과 같이 계산됩니다 .

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

여기서 s 는 표준편차이고 $k_{(1-\alpha/2;p,n)}$ 는 상수입니다 .

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

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

여기서 Φ 는 표준 정규 CDF(누적 분포 함수) 입니다 .

따라서 k 는 다음 방정식의 해를 구합니다 .

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

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 .

비모수 구간

단측 하한

크기가 n 인 표본에서 표집된 분포의 최소 비율 β 를 포함하기 위한 100(1 - α)% 단측 공차 하한은 순서 통계량 $x_{(l)}$ 입니다 . 지수 l 은 다음과 같이 계산됩니다 .

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

여기서 $\Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포의 (1 - α) 번째 분위수입니다 .

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

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

단측 상한

크기가 n 인 표본에서 표집된 분포의 최소 비율 β 를 포함하기 위한 100(1 - α)% 단측 공차 상한은 순서 통계량 $x_{(u)}$ 입니다 . 지수 u 는 다음과 같이 계산됩니다 .

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

여기서 $\Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$ 는 시행 횟수가 n 이고 성공 확률이 β 인 이항 분포의 (1 - α) 번째 분위수입니다 .

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

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

양측 구간

크기가 n 인 표본에서 비율 β 이상의 표본 분포를 포함하기 위한 100(1 - α)% 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 .

$\nu = n - \Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$라고 가정합니다. 여기서 $\Phi^{-1}_{bin}(1\text{-}\alpha, n, \beta)$는 시행 횟수가 n이고 성공 확률이 β 인 이항 분포의 (1 - α) 번째 분위수입니다 . ν 가 2 보다 작으면 양측 분포 무관 공차 구간을 계산할 수 없습니다 . ν 가 2 보다 크거나 같으면 $l = \text{floor}(\nu/2)$ 이고 $u = \text{floor}(n + 1 - \nu/2)$ 입니다 .

실제 신뢰 수준은 $\Phi_{bin}(u\text{-}l\text{-}1, n, \beta)$ 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.

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 분포에 대한 통계 상세 정보

이 섹션에는 분포 플랫폼의 " 연속형 적합 " 메뉴 옵션에 대한 통계 상세 정보가 포함되어 있습니다. 별도로 지정한 경우를 제외하고 모수 추정값에 대한 신뢰 구간은 가능도 기반 계산을 사용합니다. Y 열에 " 감지 한계 " 열 특성이 있는 경우 " 연속형 적합 " 옵션은 중도절단 분포를 적합시키고 분포의 일부만 사용할 수 있습니다. 중도절단 데이터에 분포를 적합시키는 방법에 대한 자세한 내용은 Meeker and Escobar(1998) 연구 자료에서 확인하십시오.

정규 적합

"정규 적합" 옵션은 정규 분포의 두 모수를 추정합니다.

- 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.

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

$E(x) = \mu$

$Var(x) = \sigma^2$

참고 : 평균 추정값에 대한 신뢰 구간은 t 분포를 기반으로 합니다. 척도 모수에 대한 신뢰 구간은 χ^2 분포를 기반으로 합니다.

Cauchy 적합

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

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

$E(x)$ = 정의되지 않음

Var(x) = 정의되지 않음

스튜던트 t 적합

"스튜던트 t 적합" 옵션은 위치가 μ 이고, 척도가 σ 이고, 자유도가 ν 인 스튜던트 t 분포를 적합시킵니다.

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

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 형상 모수 γ 및 δ 에 의해 결정됩니다. SHASH 분포에 대한 자세한 내용은 Jones and Pewsey 연구 자료 (2009) 에서 확인하십시오.

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

다음은 각 요소에 대한 설명입니다.

$\phi(\cdot)$ - 표준 정규 pdf 입니다.

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

- $\gamma = 0$ 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.

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

E(x) = σ

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

Devore 연구 자료(1995)에 따르면 지수 분포는 **무기억성**을 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.

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

$$\mathrm{E}(x) = \alpha\sigma$$

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

- **표준** 감마 분포의 경우 $\sigma = 1$ 입니다. 시그마는 값이 1 이 아닐 경우 가로 축을 따라 분포를 늘리거나 줄이므로 척도 모수라고 합니다.
- 카이제곱 $\chi^2_{(\nu)}$ 분포는 $\sigma = 2$ 이고 $\alpha = \nu/2$ 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(형상) 모수를 추정합니다. $X = \ln(Y)$ 가 정규 분포인 경우에만 Y 변수가 로그 정규 분포로 나타납니다. 데이터는 0 보다 커야 합니다.

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

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

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

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 같이 정의됩니다 .

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

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

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

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 .

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

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

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

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

참고 : 정규 혼합 분포 모수 추정값에 대한 신뢰 구간은 Wald 기반 계산을 사용합니다 .

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 가능도가 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 .

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

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

다음은 Johnson Sb 의 경우에 해당하는 각 요소에 대한 설명입니다 .

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

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

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

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

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

Johnson Su

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

Johnson Sb

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

Johnson Sl

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

여기서 $\phi(\cdot)$ 는 표준 정규 pdf 입니다 .

참고 : Johnson 분포 모수 추정값에 대한 신뢰 구간은 Wald 기반 계산을 사용합니다 .

베타 적합

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

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

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

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

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 같이 정의됩니다 .

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

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

다음은 각 요소에 대한 설명입니다 .

- logL 은 로그 가능도입니다 .
- n 은 표본 크기입니다 .
- k 는 모수 수입니다 .

"AICc 가중치" 열은 합계가 1이 되는 정규화된 AICc 값을 보여 줍니다. AICc 가중치는 적합된 분포 중 하나가 true일 때 특정 분포가 true 분포일 확률로 해석됩니다. 따라서 AICc 가중치가 1

과 가장 가까운 분포가 더 적절한 적합입니다. AICc 가중치는 비결측 AICc 값만 사용해서 계산됩니다.

AICcWeight = exp[-0.5(AICc-min(AICc))] / sum(exp[-0.5(AICc-min(AICc))])

여기서 min(AICc) 는 적합된 분포 중에서 가장 작은 AICc 값입니다 .

"분포 비교" 보고서의 측도에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

이산형 적합 분포에 대한 통계 상세 정보

이 섹션에는 분포 플랫폼의 "이산형 적합" 메뉴 옵션에 대한 통계 상세 정보가 포함되어 있습니다.

Poisson 적합

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

pmf: $\frac{e^{-\lambda}\lambda^{x}}{x!}$ (단 , $0 \le \lambda < \infty$, $x = 0,1,2,...$)

$E(x) = \lambda$

$Var(x) = \lambda$

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 λ와 산포 모수 σ가 포함되어 있습니다.

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

$E(x) = \lambda$

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

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 μ이 다른 경우에 유용합니다.

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

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

$E(x)=\lambda$

$Var(x)=\lambda\sigma$

여기서 $\Gamma(\cdot)$ 는 Gamma 함수입니다 .

감마 Poisson 분포는 $\sigma_{negbin}=(\sigma_{gp}-1)/\lambda_{gp}$ 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 .

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

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

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

ZI 음이항 적합

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

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

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

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

이항 적합

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

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

$E(x)=np$

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

여기서 n 은 독립 시행 횟수입니다 .

참고 : 이항 분포 모수의 신뢰 구간은 스코어 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 . $\delta < 0$ 이면 과소산포가 나타납니다 . $\delta = 0$ 이면 x 가 Binomial(n,p) 분포를 따릅니다 . 베타 이항 분포는 $n \geq 2$ 일 때만 존재합니다 .

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

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

$E(x) = np$

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

여기서 Γ(·) 는 Gamma 함수입니다 .

$x|\pi \sim$ Binomial(n,π) 인 반면 $\pi \sim$ Beta(α,β) 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 0 인 경우에는 이 계산식이 적용되지 않습니다. 이 경우에는 베타 이항 분포가 Binomial(n,p) 로 축소된 것이며 $\hat{p}$ 는 p 의 추정값입니다.

베타 이항 분포 모수의 신뢰 구간은 프로파일 가능도 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.

4 장

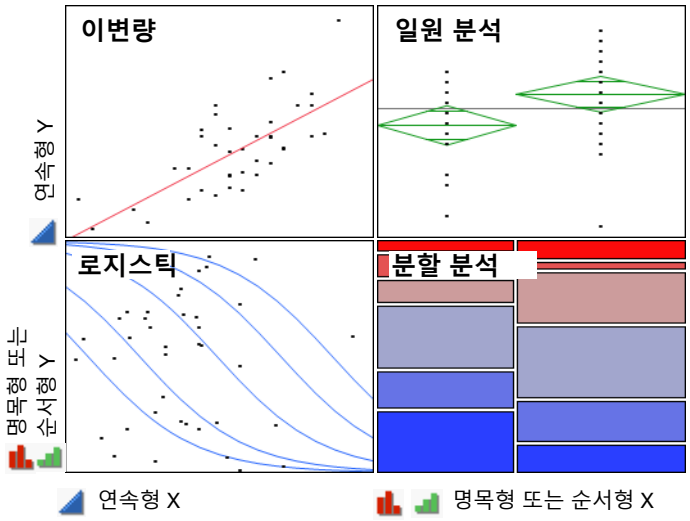
X 로 Y 적합 소개
두 변수 간의 관계 분석

변수 쌍을 분석하려면 "X 로 Y 적합" 플랫폼을 사용합니다. 산점도, 선형 회귀, ANOVA, 다중 비교, 로지스틱 회귀, 분할표 등을 사용하여 이 작업을 수행할 수 있습니다. 특정 분석은 변수의 모델링 유형에 따라 달라집니다.

"X 로 Y 적합"은 다음 네 가지 플랫폼 중 하나를 시작합니다.

- 연속형 Y 및 연속형 X 를 분석하려면 이변량을 사용합니다 .
- 연속형 Y 및 범주형 X 를 분석하려면 ANOVA 를 포함한 일원 분석을 사용합니다 .
- 범주형 Y 및 연속형 X 를 분석하려면 로지스틱 회귀를 사용합니다 .
- 범주형 Y 및 범주형 X 를 분석하려면 분할 분석을 사용합니다 .

그림 4.1 네 가지 분석 유형의 예



X 로 Y 적합 플랫폼은 이변량 , 일원 분석 , 로지스틱 및 분할 분석의 네 가지 플랫폼 (또는 분석 유형) 모음입니다 .

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.
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.
로지스틱	연속형 X로 명목형 또는 순서형 Y 적합	반응 범주에 대한 확률을 연속형 X 예측 변수에 적합시킵니다. 자세한 내용은 "로지스틱 분석"에서 확인하십시오.

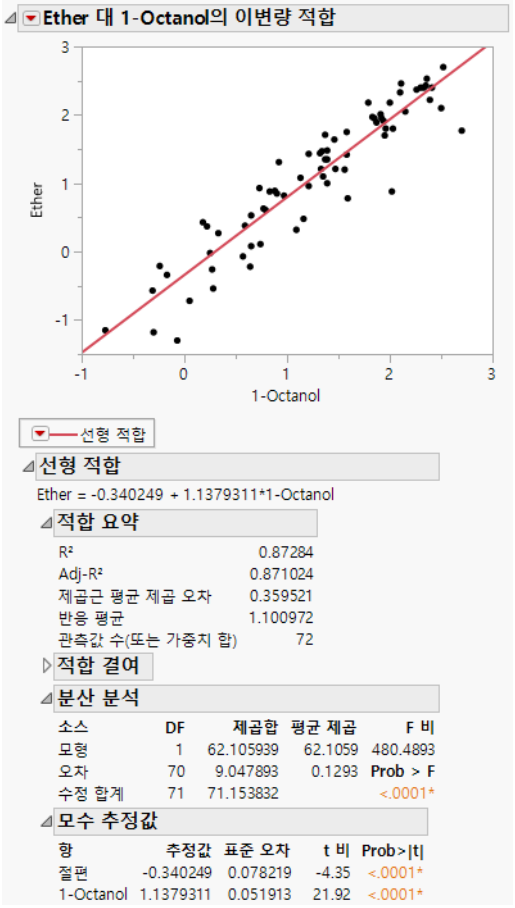
5 장

이변량 분석
두 연속형 변수 간의 관계 분석

이변량 플랫폼을 사용하여 두 연속형 변수 간의 관계를 조사합니다. 산점도에 데이터가 그래픽 형식으로 제공됩니다. 옵션을 사용하면 단순 선형 회귀선, 다항 회귀 곡선 또는 평활기를 데이터에 적합시킬 수 있습니다.

이변량 플랫폼은 X로 Y 적합 플랫폼의 **연속형 대 연속형** 분석법입니다. 이변량이란 단어는 단순히 관련 변수가 한 개(단변량) 또는 여러 개(다변량)가 아니라 두 개임을 의미합니다.

그림 5.1 이변량 분석의 예



목차

이변량 플랫폼 개요 97
이변량 분석의 예 97
이변량 플랫폼 시작 99
데이터 형식 99
이변량 보고서 100
이변량 플랫폼 옵션 101
이변량 적합 옵션 103
이변량 적합 보고서 106
평균 적합 보고서 107
선형 적합, 다항식 적합 및 특수 적합 보고서 108
특수 적합 창 113
스플라인 적합 보고서 113
커널 평활기 보고서 114
각 값 적합 보고서 114
직교 적합 보고서 115
로버스트 적합 보고서 115
Passing-Bablok 적합 보고서 116
밀도 타원 보고서 117
비모수 밀도 보고서 119
이변량 플랫폼의 추가 예 120
특수 적합 옵션의 예 120
직교 적합 옵션의 예 121
로버스트 적합 옵션의 예 122
밀도 타원을 사용한 그룹화의 예 124
회귀선을 사용한 그룹화의 예 125
기준 변수를 사용한 그룹화의 예 126
이변량 플랫폼에 대한 통계 상세 정보 128
선형 적합 옵션에 대한 통계 상세 정보 128
다항식 적합 옵션에 대한 통계 상세 정보 129
스플라인 적합 옵션에 대한 통계 상세 정보 129
직교 적합 옵션에 대한 통계 상세 정보 130
로버스트 옵션에 대한 통계 상세 정보 131
Passing-Bablok 적합 옵션에 대한 통계 상세 정보 131
적합 요약 보고서에 대한 통계 상세 정보 132
적합 결여 보고서에 대한 통계 상세 정보 132
모수 추정값 보고서에 대한 통계 상세 정보 133
평활 적합 보고서에 대한 통계 상세 정보 133
이변량 정규 타원 보고서에 대한 통계 상세 정보 134

이변량 플랫폼 개요

이변량 플랫폼을 사용하면 대화식으로 모형을 적합시키고 산점도에서 해당 적합을 볼 수 있습니다. 동일한 그림에서 여러 모형 적합을 비교할 수 있습니다.

이변량 플랫폼은 하나 이상의 연속형 Y 변수와 하나 이상의 연속형 X 변수가 있을 때 "X로 Y 적합"에서 시작됩니다. 처음에는 이변량 분석 플랫폼에 X 변수와 Y 변수의 각 조합에 대한 산점도가 표시됩니다. 산점도의 빨간색 삼각형 옵션을 사용하여 두 연속형 변수에 대한 모형을 대화식으로 탐색합니다. 데이터에 단순 선형 회귀선을 적합시키거나 더 복잡한 회귀 모형을 적합시킬 수 있습니다. 밀도 추정값을 탐색할 수도 있습니다.

이변량 플랫폼의 적합 옵션 범주에는 회귀 적합 및 밀도 추정이 포함됩니다.

표 5.1 이변량 플랫폼 적합 범주

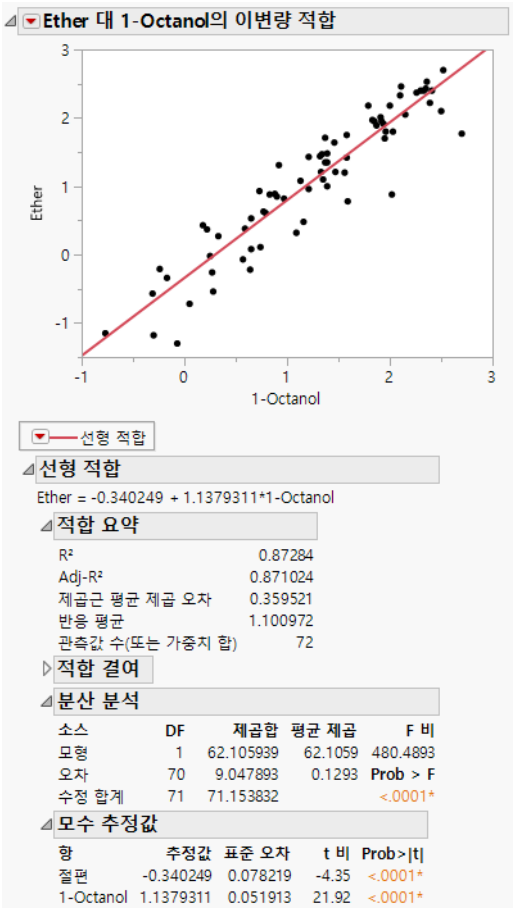
범주	설명	적합 옵션
회귀 적합	회귀 방법은 관측된 데이터 점에 모형을 적합시킵니다. 이 적합 방법에는 최소 제곱 적합과 함께 스플라인 적합, 커널 평활, 직교 적합, 변환 및 로버스트 적합이 포함됩니다.	**평균 적합** **선형 적합** **다항식 적합** **특수 적합** **유연** **직교 적합** **Passing-Bablok 적합** **로버스트**
밀도 추정	밀도 추정은 점에 이변량 분포를 적합시킵니다. 타원형 등고선이 특징인 이변량 정규 밀도나 일반 비모수 밀도 중 하나를 선택할 수 있습니다.	**밀도 타원** **비모수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

이변량 분석의 예

이변량 플랫폼을 사용하여 산점도를 생성하고 두 연속형 변수에 대한 회귀선을 적합시킵니다. 이 예에서는 서로 다른 용매에서 화합물의 용해도 수준이 측정되었습니다. 두 용매의 결과를 비교하려고 합니다.

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Solubility.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. Ether 를 선택하고 **Y, 반응**을 클릭합니다.
4. 1-Octanol 을 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 선형 적합을 선택합니다.

그림 5.2 이변량 적합 보고서

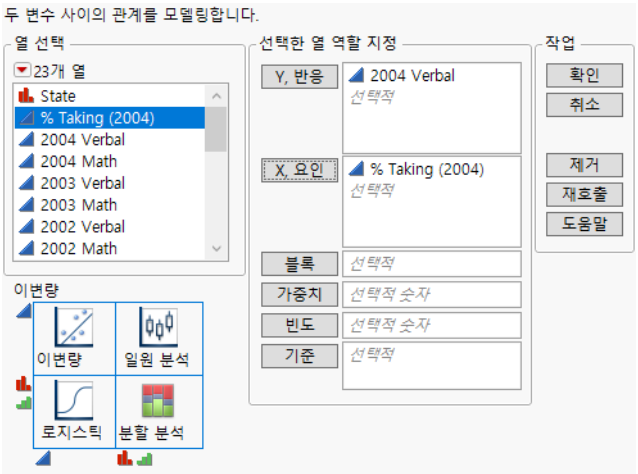


두 용매의 결과가 비슷하다는 것을 알 수 있습니다. 선형 모형의 R^2 값은 0.87이며, 이는 모형이 반응 변동의 87%를 설명하고 있음을 나타냅니다. X 변수와 Y 변수 모두 측정값이므로 직교 적합 또는 Passing-Bablok 적합 옵션을 고려할 수 있습니다.

이변량 플랫폼 시작

분석 > X 로 Y 적합을 선택하여 이변량 플랫폼을 시작할 수 있습니다 . X 로 Y 적합 시작 창은 네 가지 유형의 분석에 사용됩니다 . 연속형 Y 변수와 연속형 X 변수를 입력하면 이변량 플랫폼이 시작됩니다 .

그림 5.3 이변량 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 . 이변량 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 . 반응 변수는 종속 변수라고도 합니다 . 이러한 변수의 모델링 유형은 연속형이어야 합니다 .

X, 요인 분석할 하나 이상의 예측 변수입니다 . 요인 변수는 독립 변수라고도 합니다 . 이러한 변수의 모델링 유형은 연속형이어야 합니다 .

블록 (이변량 분석에는 적용할 수 없음) 블록 변수를 지정하는 열입니다 .

가중치 데이터 테이블의 각 관측값에 대한 가중치를 포함하는 열입니다 . 값이 0 보다 큰 행만 분석에 포함됩니다 .

빈도 분석의 각 행에 빈도를 할당합니다 . 빈도 할당은 데이터를 요약할 때 유용합니다 .

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

데이터 형식

이변량 플랫폼에서 데이터는 요약되지 않거나 요약된 데이터 열로 구성될 수 있습니다 .

요약되지 않은 데이터 각 관측값에 대한 하나의 행과 X 및 Y 값에 대한 열이 있습니다 .

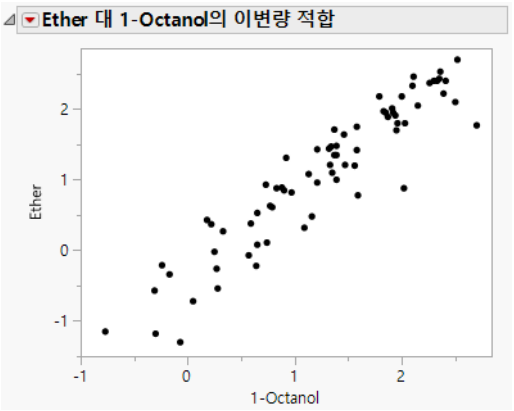
요약된 데이터 각 행이 공통된 X 및 Y 값을 갖는 일련의 관측값을 나타냅니다 . 데이터 테이블에는 각 행에 대한 관측값 수가 포함된 빈도 열이 있어야 합니다 . 시작 창에서는 이 열을 " 빈도 " 열로 입력합니다 .

참고 : X 로 Y 적합 시작 창은 연속형 , 순서형 및 명목형 모델링 유형의 열을 허용합니다 . 연속형 모델링 유형을 사용하는 Y, 반응 열과 X, 요인 열의 모든 쌍에 대해 이변량 플랫폼이 시작됩니다 . X 로 Y 적합 시작 창은 다른 열 유형 조합에 대해 일원 분석 , 분할 분석 또는 로지스틱 플랫폼을 실행합니다 .

이변량 보고서

처음에는 "이변량" 보고서에 각 X 및 Y 변수 쌍에 대한 산점도가 포함됩니다. 모형을 데이터에 적합시키고 빨간색 삼각형 옵션을 사용하여 통계 보고서를 볼 수 있습니다. 자세한 내용은 "이변량 플랫폼 옵션"에서 확인하십시오.

그림 5.4 이변량 그림



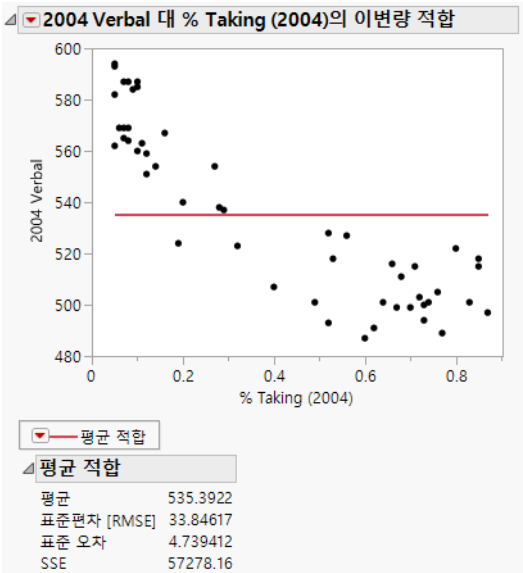
대화식으로 변수 바꾸기

그림에서 한 축의 변수를 다른 축으로 드래그한 후 놓는 방법으로 변수를 대화식으로 바꿀 수 있습니다. 연결된 데이터 테이블의 "열" 패널에서 변수를 선택하여 축으로 드래그하는 방법으로 변수를 바꿀 수도 있습니다.

이변량 플랫폼 옵션

"이변량 적합"의 빨간색 삼각형 메뉴에는 표시, 적합 및 제어 옵션이 포함되어 있습니다. 각 적합 옵션은 산점도에 선, 곡선 또는 분포를 추가하고, 그림 아래에 적합을 위한 빨간색 삼각형 메뉴를 추가하고, 보고서 창에 특정 적합 보고서를 추가합니다.

그림 5.5 평균 적합 옵션의 예



참고 : " 그룹 적합 " 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에만 표시됩니다 . " 그룹 적합 " 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 을 기준으로 정렬할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

" 이변량 적합 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

점 표시 산점도에 점을 표시하거나 숨깁니다 .

히스토그램 테두리 산점도의 가로 축과 세로 축에 히스토그램을 표시하거나 숨깁니다 .

참고 : 숨겨진 행의 데이터 점은 산점도에서 숨겨지지만 히스토그램에서는 숨겨지지 않습니다 . 히스토그램과 분석 결과에서 행을 제외하려면 " 숨기고 제외하기 " 행 상태를 적용하고 " 이변량 적합 " 의 빨간색 삼각형 메뉴에서 **다시 실행 > 분석 다시 실행**을 선택하십시오 .

요약 통계량 그림에 표시되는 변수에 대한 요약 통계량을 표시하거나 숨깁니다 . 측정 기준에는 두 변수 간의 상관 및 공분산과 각 변수에 대한 단변량 평균 및 표준편차가 포함됩니다 .

평균 적합 Y 반응 변수의 평균을 적합시킵니다 . 이것은 기울기가 0 으로 제한된 단순 회귀 모형입니다 . 자세한 내용은 " 평균 적합 보고서 " 에서 확인하십시오 .

선형 적합 최소 제곱 회귀선을 적합시킵니다. 적합이 그림에 표시되고 적합 보고서가 제공됩니다. 자세한 내용은 "선형 적합, 다항식 적합 및 특수 적합 보고서"에서 확인하십시오.

다항식 적합 최소 제곱 회귀를 사용하여 선택한 차수의 다항식 곡선을 적합시킵니다. 자세한 내용은 "선형 적합, 다항식 적합 및 특수 적합 보고서"에서 확인하십시오.

특수 적합 Y 및 X 변수에 대한 변환을 포함하는 회귀 모형을 적합시킬 수 있습니다. 변환에는 로그, 제곱근, 역수 및 지수가 포함됩니다. **중심화 다항식**을 해제하거나, 절편 및 기울기를 제약하거나, 변환된 변수를 사용하여 다항식 모형을 적합시킬 수도 있습니다. 자세한 내용은 "특수 적합 창"에서 확인하십시오.

유연 유연한 모형을 적합시킬 수 있습니다. 모형에는 스플라인, 커널 평활기 및 점별 적합이 포함됩니다.

스플라인 적합 벌점 최소 제곱 모형을 데이터에 적합시킵니다. 평활 모수 λ 를 사용하여 모형 적합의 평활도를 조정할 수 있습니다. 자세한 내용은 "스플라인 적합 보고서"에서 확인하십시오.

커널 평활기 국소 가중 최소 제곱 모형을 데이터에 적합시킵니다. 이 모형을 LOWESS(국소 가중 산점도 평활) 모형이라고도 합니다. 평활 모수 α 를 사용하여 모형의 평활도를 제어할 수 있습니다. 자세한 내용은 "커널 평활기 보고서"에서 확인하십시오.

각 값 적합 각 X 값에 대해 평균 반응을 연결합니다. 자세한 내용은 "각 값 적합 보고서"에서 확인하십시오.

직교 적합 X 변수와 Y 변수 모두 오차를 사용하여 측정될 때 사용할 수 있는 직교 회귀 모형을 적합시킬 수 있습니다. 자세한 내용은 "직교 적합 보고서"에서 확인하십시오.

단변량 분산, 주성분 표준화된 첫 번째 주성분을 데이터에 적합시킵니다.

등분산 X 와 Y 의 오차 분산이 동일하다고 가정하는 분산 비율 1 을 사용하여 직교 회귀 모형을 적합시킵니다. 이 방법을 Deming 회귀라고도 합니다.

X 를 Y 에 적합 Y 에 분산이 없음을 나타내는 분산 비율 0 을 사용하여 직교 회귀 모형을 적합시킵니다.

지정된 분산 비율 직교 회귀 모형에 대해 지정된 분산 비율을 입력할 수 있습니다.

Passing-Bablok 적합 Passing-Bablok 절차를 사용하여 회귀 모형을 적합시킵니다. 이 옵션에는 Bland-Altman 분석 옵션이 포함됩니다 . X 변수와 Y 변수 모두 오차를 사용하여 측정될 때 사용합니다. 자세한 내용은 "Passing-Bablok 적합 보고서" 및 "Passing-Bablok 적합 옵션에 대한 통계 상세 정보"에서 확인하십시오.

로버스트 로버스트 회귀 모형을 적합시킬 수 있습니다. 로버스트 모형을 사용하여 모형 적합에 대한 이상치의 영향을 줄일 수 있습니다. 자세한 내용은 "로버스트 적합 보고서"에서 확인하십시오.

로버스트 적합 Huber M- 추정 방법을 사용하여 로버스트 회귀 모형을 적합시킵니다.

Cauchy 적합 Cauchy 연결 함수를 사용한 최대 가능도로 모수가 추정되는 로버스트 회귀 모형을 적합시킵니다.

밀도 타원 지정된 백분율의 이변량 정규 밀도 타원을 그림에 추가할 수 있습니다 . 등고선에는 데이터 점의 지정된 백분율이 포함됩니다 . 이 옵션을 사용하여 상관을 추정할 수 있습니다 . 자세한 내용은 " 밀도 타원 보고서 " 에서 확인하십시오 .

비모수 밀도 비모수 밀도 등고선을 그림에 추가할 수 있습니다. 이 등고선은 데이터 점의 밀도를 설명합니다 . 자세한 내용은 " 비모수 밀도 보고서 " 에서 확인하십시오 .

그룹화 기준 그룹화 변수를 지정할 수 있습니다. 그룹화 변수를 지정하면 그룹화 변수의 각 수준에 대해 후속 적합이 계산됩니다 . 선 , 곡선 또는 타원은 그룹별로 산점도에 중첩됩니다 . 이렇게 하면 그룹별 적합을 시각적으로 비교할 수 있습니다 . 자세한 내용은 " 밀도 타원을 사용한 그룹화의 예 " 및 " 회귀선을 사용한 그룹화의 예 " 에서 확인하십시오 .

참고 : " 그룹화 기준 " 옵션을 사용하면 하나의 그림에 그룹 적합을 시각화하고 결과를 하나의 보고서에 표시할 수 있습니다 . 또는 시작 창의 " 기준 " 옵션을 사용하면 그룹화된 적합이 각각 개별 보고서에 표시됩니다 .

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 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 .

플랫폼 환경 설정 현재 플랫폼 환경 설정을 보거나 , 현재 JMP 보고서의 설정과 일치하도록 플랫폼 환경 설정을 업데이트할 수 있는 옵션이 포함되어 있습니다 .

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 .

이변량 적합 옵션

이변량 플랫폼 모형 특정 빨간색 삼각형 옵션은 지정된 적합에 따라 달라집니다 . 이러한 메뉴는 산점도 아래의 모형 적합 라벨 옆에 있습니다 .

- " 대부분의 이변량 적합에 적용되는 옵션 "
- " 정규 타원에 적용되는 옵션 "
- " 분위수 밀도 등고선에 적용되는 옵션 "

대부분의 이변량 적합에 적용되는 옵션

이변량 플랫폼 모형 특정 빨간색 삼각형 옵션은 지정된 적합에 따라 달라집니다. 일부 적합에서 사용할 수 없는 옵션도 있습니다.

적합선 모형 적합을 설명하는 선, 곡선 또는 등고선을 표시하거나 숨깁니다. " 분위수 밀도 등고선 " 의 경우에는 사용할 수 없습니다.

적합 신뢰 곡선 기대값 (평균) 에 대한 신뢰 한계 (곡선) 를 표시하거나 숨깁니다.

개별값 신뢰 곡선 개별 예측값의 신뢰 한계를 표시하거나 숨깁니다. 신뢰 한계는 오차의 변동과 모수 추정값의 변동을 반영합니다.

선 색상 적합 곡선의 색상을 선택할 수 있습니다.

선 스타일 각 적합에 대한 선 스타일을 선택할 수 있습니다.

선 너비 각 적합에 대한 선 너비를 선택할 수 있습니다.

보고서 각 적합에 대한 보고서를 표시하거나 숨깁니다.

참고 : 이 옵션은 이변량 그림을 수정하지 않습니다.

프로파일러 X 변수에 대한 예측 추적을 표시하거나 숨깁니다. 하나 이상의 반응에 대한 최적 설정을 찾고 시뮬레이션을 사용하여 반응 분포를 탐색할 수 있습니다. 자세한 내용은 Profilers 의 에서 확인하십시오.

예측값 저장 현재 데이터 테이블에 **예측 < 열 이름 >** 이라는 새 열을 생성합니다. 여기서 **열 이름** 은 Y 변수의 이름입니다. 이 열에는 예측 계산식과 예측값이 포함됩니다.

팁 : 예측 계산식은 사용자가 테이블에 추가하는 새 행에 대해 자동으로 값을 계산합니다.

잔차 저장 현재 데이터 테이블에 **잔차 < 열 이름 >** 이라는 새 열을 생성합니다. 여기서 **열 이름** 은 Y 변수의 이름입니다. 이 열에는 관측된 반응 값에서 예측값을 뺀 값이 포함됩니다.

참고 : 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 현재 데이터 테이블에 **스튜던트화 잔차 < 열 이름 >** 이라는 새 열을 생성합니다. 여기서 **열 이름** 은 Y 변수의 이름입니다. 이 열에는 잔차를 표준 오차로 나눈 값이 포함됩니다.

평균 신뢰 한계 계산식 데이터 테이블에 **평균 < 열 이름 > 의 95% 하한** 및 **평균 < 열 이름 > 의 95% 상한** 이라는 두 개의 새 열을 생성합니다. 여기서 **열 이름** 은 Y 변수의 이름입니다. 이 열은 평균 반응에 대한 95% 신뢰 하한 및 상한의 계산식과 값이 모두 포함됩니다.

개별값 신뢰 한계 계산식 데이터 테이블에 **개별 < 열 이름 > 의 95% 하한** 및 **개별 < 열 이름 > 의 95% 상한** 이라는 두 개의 새 열을 생성합니다. 여기서 **열 이름** 은 Y 변수의 이름입니다. 이 열은 개별 예측에 대한 95% 신뢰 하한 및 상한의 계산식과 값이 모두 포함됩니다.

잔차 그림 (선형 , 다항식 , Passing-Bablok 및 특수 적합에만 사용 가능) 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 표시하거나 숨깁니다 .

Bland Altman 분석 (Passing-Bablok 에만 사용 가능) 매칭 쌍 및 Bland-Altman 분석을 새 보고서 창에 표시합니다 .

α **수준 설정** 신뢰 곡선을 계산하는 데 사용되는 유의 수준을 지정할 수 있습니다 .

적합 신뢰 구간 음영 (일부 적합에는 사용 불가능) 기대 반응 (평균) 에 대한 음영 신뢰 영역을 표시하거나 숨깁니다 .

개별값 신뢰 구간 음영 (일부 적합에는 사용 불가능) 개별 예측에 대한 음영 신뢰 영역을 표시하거나 숨깁니다 .

계수 저장 (유연한 스플라인 적합에만 사용 가능) 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 .

적합 제거 그래프에서 적합을 제거하고 해당 보고서도 제거합니다 .

정규 타원에 적용되는 옵션

" 정규 타원 " 의 빨간색 삼각형 옵션은 이변량 플랫폼의 밀도 타원 적합에 사용할 수 있습니다 .

음영 등고선 밀도 타원의 내부 영역에 음영을 적용하거나 음영을 제거합니다 .

내부 점 선택 타원 내부의 점을 선택합니다 .

외부 점 선택 타원 외부의 점을 선택합니다 .

분위수 밀도 등고선에 적용되는 옵션

" 분위수 밀도 등고선 " 의 빨간색 삼각형 옵션은 이변량 플랫폼의 비모수 타원 적합에 사용할 수 있습니다 .

커널 제어 각 변수의 표준편차를 제어하는 슬라이더를 표시하거나 숨깁니다 . 표준편차는 등고선 밀도를 결정하기 위한 X 및 Y 값 범위를 정의합니다 .

5% 등고선 5% 등고선을 표시하거나 숨깁니다 .

등고선 등고선을 표시하거나 숨깁니다 .

등고선 채우기 채워진 등고선을 표시하거나 숨깁니다 .

색상 테마 등고선의 색상 테마를 변경할 수 있습니다 .

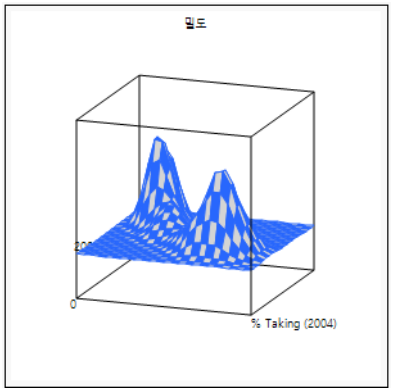
밀도별 점 선택 사용자가 지정한 분위수 범위에 있는 점을 선택할 수 있습니다 .

밀도 분위수별 색상 밀도에 따라 다른 점 색상을 적용합니다 .

밀도 분위수 저장 현재 데이터 테이블에 각 점에 대한 밀도 분위수가 포함된 새 열을 생성합니다 .

그물 그림 두 분석 변수의 격자에 대한 3 차원 밀도 그림을 표시하거나 숨깁니다 .

그림 5.6 그물 그림의 예



최빈 군집화 데이터의 최빈 군집화 결과를 표시하거나 숨깁니다. 현재 데이터 테이블에 각 데이터 쌍에 대한 군집 번호가 포함된 새 열을 생성합니다. 최빈 군집화는 밀도 추정값을 기반으로 합니다. JMP 에서는 10,404 개의 밀도 추정값 격자를 생성합니다. 밀도 분포의 최빈값 수에 따라 군집 수가 결정됩니다. 최빈값은 **군집 중심**입니다. 나머지 점은 거리 및 밀도 추정값을 기반으로 각 최빈값에 반복적으로 군집화됩니다.

참고 : JMP 의 다른 군집화 방법에 대한 자세한 내용은 **다변량 방법**의 에서 확인하십시오.

밀도 격자 저장 밀도 추정값과 관련 분위수를 포함하는 새 데이터 테이블을 생성합니다.

참고 : 최빈 군집화 값을 먼저 저장한 다음 밀도 격자를 저장하면 격자 테이블에 군집 값도 포함됩니다.

이변량 적합 보고서

이변량 플랫폼을 사용하면 단일 X 변수를 기반으로 Y 변수를 예측하기 위해 다양한 모형을 적합시킬 수 있습니다. 이 섹션에는 특정 모형 적합에 대해 생성되는 보고서 관련 정보가 포함되어 있습니다.

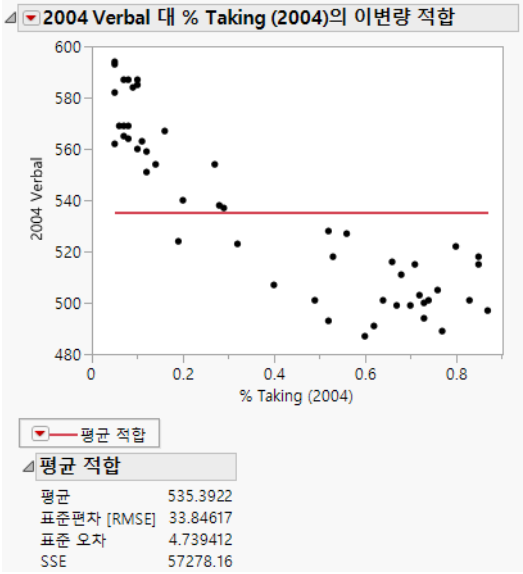
- "평균 적합 보고서"
- "선형 적합, 다항식 적합 및 특수 적합 보고서"
- "특수 적합 창"
- "스플라인 적합 보고서"
- "커널 평활기 보고서"
- "각 값 적합 보고서"
- "직교 적합 보고서"
- "로버스트 적합 보고서"

- "Passing-Bablok 적합 보고서 "
- " 밀도 타원 보고서 "
- " 비모수 밀도 보고서 "

평균 적합 보고서

이변량 플랫폼의 " 평균 적합 " 옵션을 사용하여 Y 반응 변수의 평균을 적합시킵니다 . 평균 선을 기준으로 사용하여 다른 적합과 비교할 수 있습니다 .

그림 5.7 평균 적합의 예



이변량 플랫폼의 " 평균 적합 " 보고서에는 요약 통계량 테이블이 포함됩니다 .

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 . RMSE 를 값 개수의 제곱근으로 나눠서 계산됩니다 .

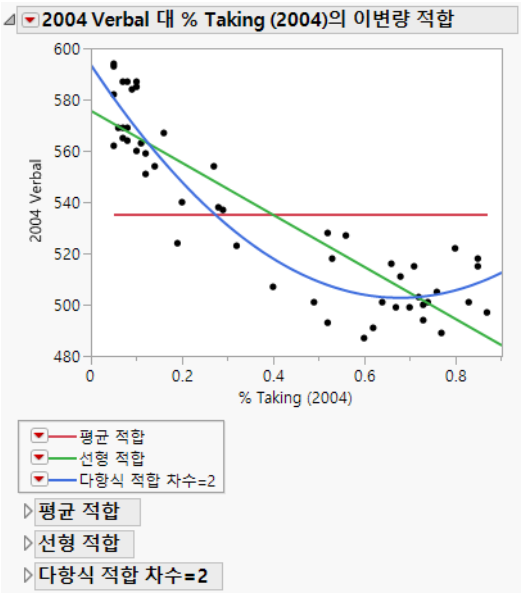
SSE 단순 평균 모형의 오차제곱합입니다 . 각 모형 적합에 대한 분산 분석 테이블에서 **오차**의 제곱합으로 표시됩니다 .

"평균 적합" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오.

선형 적합 , 다항식 적합 및 특수 적합 보고서

이변량 플랫폼에서 " 선형 적합 ", " 다항식 적합 " 또는 " 특수 적합 " 옵션을 사용하여 회귀 모형을 적합시킵니다 . 여러 모형을 적합시킨 후 산점도에서 적합을 비교할 수 있습니다 .

그림 5.8 선형 적합 및 다항식 적합의 예



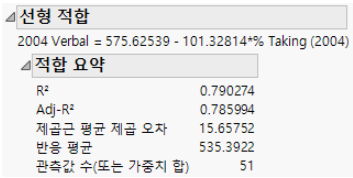
"선형 적합"과 "다항식 적합 차수" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오. 통계 상세 정보는 "선형 적합 옵션에 대한 통계 상세 정보"에서 확인하십시오.

이변량 플랫폼에는 선택한 각 적합에 대한 보고서가 있습니다. "선형", "다항식" 및 "변환된 적합" 보고서에는 각각 적합 방정식이 있는 텍스트 상자가 포함되어 있습니다. 각 적합 보고서에는 적합 요약, ANOVA(분산 분석) 및 모수 추정값 테이블이 포함되어 있습니다. 데이터에 반복 실험이 있는 경우 적합 결여에 대한 네 번째 테이블이 나타납니다. 변환된 Y 변수에 대한 적합에는 원래 척도 테이블의 적합 측도 요약이 포함됩니다.

적합 요약

이변량 플랫폼 적합 보고서의 " 적합 요약 " 테이블에는 모형 적합의 수치 요약이 포함됩니다 . " 적합 요약 " 테이블 위에 적합 방정식이 표시됩니다 .

그림 5.9 적합 요약 테이블


선형 적합
2004 Verbal = 575.62539 - 101.32814*% Taking (2004)
적합 요약

R^2	0.790274
Adj-R^2	0.785994
제곱근 평균 제곱 오차	15.65752
반응 평균	535.3922
관측값 수(또는 가중치 합)	51

" 적합 요약 " 테이블에는 다음 통계량이 포함됩니다 .

R^2 모형에 의해 설명된 변동의 비율입니다 . 나머지 변동은 랜덤 오차에 기인합니다 . 모형이 완벽하게 적합되는 경우에는 R^2 가 1 입니다 . 자세한 내용은 " 적합 요약 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 .

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 .

Adj-R^2 모형의 모수 수에 맞게 조정된 R^2 통계량입니다 . Adj-R^2 을 사용하면 포함된 모수의 수가 다른 모형을 쉽게 비교할 수 있습니다 . 자세한 내용은 " 적합 요약 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 .

제곱근 평균 제곱 오차 랜덤 오차의 표준편차 추정값입니다 . 이 값은 " 분산 분석 " 보고서에 표시된 오차 평균 제곱의 제곱근입니다 (그림 5.11).

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

관측값 수 (또는 가중치 합) 적합을 추정하는 데 사용되는 관측값 수입니다 . 가중치 변수가 있는 경우에는 가중치의 합계입니다 .

적합 결여

이변량 적합 보고서의 " 적합 결여 " 테이블에는 적합 결여 검정의 결과가 포함됩니다. 적합 결여 검정은 반복된 X 값이 있고 모형이 포화되지 않은 경우에만 사용할 수 있습니다. 반복 실험에서 계산된 제곱합을 **순수 오차**라고 합니다. 이 값은 어떤 형태의 모형을 사용하더라도 설명하거나 예측할 수 없는 전체 오차의 비율입니다.

그림 5.10 선형 적합의 적합 결여 테이블

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

모형의 잔차와 순수 오차 사이의 차이를 **적합 결여 오차**라고 합니다. 모형이 잘못 지정된 경우 적합 결여 오차가 순수 오차보다 유의하게 클 수 있습니다. 잘못 지정된 모형은 데이터를 잘 설명하지 못하는 모형입니다. 적합 결여 검정의 귀무가설은 적합 결여 오차가 0이라는 것입니다. 따라서 *p* 값이 작으면 적합 결여가 유의함을 나타냅니다.

" 적합 결여 " 테이블에는 다음 열이 포함됩니다 .

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

DF 각 오차 소스의 DF(**자유도**) 입니다 .

- 총 오차 DF 는 해당 ANOVA(분산 분석) 테이블의 **오차** 행에 나오는 자유도입니다 . 자세한 내용은 " 분산 분석 " 에서 확인하십시오 . 총 오차 DF 값은 ANOVA 테이블에 있는 총 DF 값과 모형 DF 값 사이의 차이입니다 . 오차 DF 는 적합 결여의 자유도와 순수 오차의 자유도로 분할됩니다 .
- 순수 오차 DF 는 반복된 각 관측값 그룹에서 풀링된 것입니다 . 자세한 내용은 " 적합 결여 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 .
- 적합 결여 DF 는 총 오차 DF 와 순수 오차 DF 사이의 차이입니다 .

제곱합 각 오차 소스의 SS(제곱합) 입니다 .

- 총 오차 SS 는 해당 분산 분석 테이블의 오차 행에 나오는 제곱합입니다 . 자세한 내용은 " 분산 분석 " 에서 확인하십시오 .
- 순수 오차 SS 는 반복된 각 관측값 그룹에서 풀링된 것입니다 . 순수 오차 SS 를 DF 로 나눈 값은 지정된 예측 변수 설정에서 반응의 분산을 추정합니다 . 이 추정값은 모형의 영향을 받지 않습니다 . 자세한 내용은 " 적합 결여 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 .
- 적합 결여 SS 는 총 오차 제곱합과 순수 오차 제곱합 사이의 차이입니다 . 적합 결여 SS 가 크다면 해당 모형이 데이터에 적절하지 않은 것일 수 있습니다 .

평균 제곱 소스의 평균 제곱 , 즉 제곱합을 DF 로 나눈 값입니다 . 순수 오차 평균 제곱에 비해 적합 결여 평균 제곱이 크면 모형이 잘 적합되지 않음을 나타냅니다 . F 비는 공식 가설 검정을 수행하는 데 사용할 수 있습니다 .

F 비 순수 오차 평균 제곱에 대한 적합 결여 평균 제곱의 비율입니다 . F 비 값이 클수록 적합 결여 오차가 0 일 가능성이 낮아집니다 .

Prob > F 적합 결여 검정의 *p* 값입니다 . 귀무가설은 적합 결여 오차가 0 이라는 것입니다 . *p* 값이 작으면 적합 결여가 유의함을 나타냅니다 .

최대 R² 모형에서 모형의 변수만으로 얻을 수 있는 최대 R^2 값입니다. 자세한 내용은 " 적합 결여 보고서에 대한 통계 상세 정보 " 에서 확인하십시오.

분산 분석

이변량 적합 보고서의 "분산 분석" 테이블에는 적합 모형을 모든 예측값이 반응 평균과 동일한 모형과 비교하기 위한 계산이 포함됩니다. ANOVA(분산 분석) 테이블의 값은 모형의 유효성을 평가하기 위해 F 비를 계산하는 데 사용됩니다. F 비와 관련된 p 값이 작으면 해당 모형은 반응 평균만 사용할 때보다 데이터에 대한 적합도가 더 높은 것으로 간주됩니다.

그림 5.11 선형 적합의 분산 분석 테이블

선형 적합

2004 Verbal = 575.62539 - 101.32814*% Taking (2004)

▷ 적합 요약

▷ 적합 결여

◢ 분산 분석

소스	DF	제곱합	평균 제곱	F 비
모형	1	45265.424	45265.4	184.6379
오차	49	12012.733	245.2	Prob > F
수정 합계	50	57278.157		<.0001*

" 분산 분석 " 테이블에는 다음 열이 포함됩니다.

소스 세 가지 변동 소스이며 모형 , 오차 및 수정 합계 중 하나입니다.

DF 각 변동 소스에 대한 관련 DF(자유도) 입니다. 수정 합계 DF 는 항상 관측값 수에서 1 을 뺀 값이며 다음과 같이 모형 자유도와 오차 자유도로 분할됩니다.

- 모형 DF 는 모형 적합에 사용되는 모수 (절편 제외) 의 수입니다.
- 오차 DF 는 수정 합계 DF 와 모형 DF 의 차이입니다.

제곱합 각 변동 소스에 대한 관련 SS(제곱합) 입니다.

- 총 (수정 합계) SS 는 반응 값과 표본 평균 간의 차이에 대한 제곱합입니다. 이 값은 반응 값의 총 변동을 나타냅니다.
- 오차 SS 는 적합된 값과 실제값 간의 차이에 대한 제곱합입니다. 이 값은 적합 모형에 의해 설명되지 않은 변동을 나타냅니다.
- 모형 SS 는 수정 합계 SS 와 오차 SS 간의 차이입니다. 이 값은 모형에 의해 설명된 변동을 나타냅니다.

평균 제곱 모형 및 오차 변동 소스에 대한 평균 제곱 통계량입니다. 각 평균 제곱 값은 제곱합을 해당 DF 로 나눈 것입니다.

참고 : 오차에 대한 평균 제곱의 제곱근은 " 적합 요약 " 테이블의 RMSE 와 동일합니다.

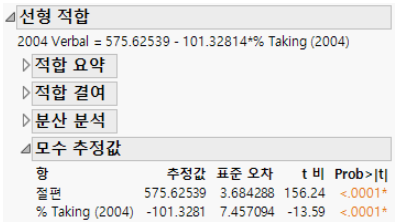
F 비 모형 평균 제곱을 오차 평균 제곱으로 나눈 것입니다 . F 비는 모형이 모든 예측값과 반응 평균이 동일한 모형과 유의하게 다른지 여부를 검정하기 위한 검정 통계량입니다. 적합의 기본 가설은 절편을 제외한 모든 회귀 모수가 0 이라는 것입니다. 이 가설이 참이면 오차의 평균 제곱과 모형의 평균 제곱이 모두 오차 분산을 추정하며 , 해당 비율은 F 분포를 따릅니다.

Prob > F 검정에 대한 관측된 유의 확률 (*p* 값) 입니다 . *p* 값이 작으면 회귀 효과의 증거로 간주됩니다 .

모수 추정값

이변량 적합 보고서의 " 모수 추정값 " 테이블에는 모형 모수 추정값이 포함됩니다 .

그림 5.12 선형 적합의 모수 추정값 테이블


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*

" 모수 추정값 " 테이블에는 다음 열이 포함됩니다 .

항 추정된 모수에 해당하는 모형 항입니다 . 첫 번째 항은 절편입니다 .

추정값 각 항에 대한 모수 추정값입니다 . 이 값은 모형 계수의 추정값입니다 .

표준 오차 모수 추정값의 표준 오차 추정값입니다 .

t 비 각 모수가 0 이라는 가설에 대한 검정 통계량입니다 . 이 값은 표준 오차에 대한 모수 추정값의 비율입니다 . 모형에 대한 일반적인 가정이 주어지면 *t* 비는 스튜던트 *t* 분포를 따릅니다 .

Prob>|t| 실제 모수 값이 0 이라는 검정에 반대되는 양측 대립가설에 대한 *p* 값입니다 .

추가 통계량을 표시하려면 보고서에서 마우스 오른쪽 버튼을 클릭하고 **열** 메뉴를 선택합니다. 다음 통계량은 기본적으로 표시되지 않습니다.

95% 하한 모수 추정값에 대한 95% 신뢰 하한입니다 .

95% 상한 모수 추정값에 대한 95% 신뢰 상한입니다 .

표준화 베타 모든 항이 평균 0, 분산 1 로 표준화된 회귀 모형의 모수 추정값입니다 . 자세한 내용은 " 모수 추정값 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 .

VIF 모형의 각 항에 대한 VIF(분산 팽창 계수) 입니다 . VIF 값이 높으면 모형의 항 사이에 공선성 문제가 있음을 나타냅니다 .

설계 표준 오차 모수 추정값의 상대 분산 제곱근입니다 . 자세한 내용은 " 모수 추정값 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 .

원래 척도에서 적합

이변량 적합 보고서의 "원래 척도에서 적합" 테이블에는 변환되지 않은 척도에서 측정된 모형 적합의 수치 요약이 포함됩니다. 이 테이블은 Y 변수가 변환된 경우에만 사용할 수 있습니다.

특수 적합 창

이변량 플랫폼에서 "특수 적합" 옵션을 사용하여 변환된 변수가 있는 회귀 모형을 적합시킵니다. 기울기와 절편을 제약하거나, 특정 차수의 다항식 모형을 적합시키거나, 다항식을 중심화할 수도 있습니다.

"특수 적합" 옵션을 선택하면 다음 옵션이 포함된 "변환 또는 제약 조건 지정" 창이 열립니다.

Y 변환 또는 X 변환 Y 또는 X 변수에 대한 변환을 지정할 수 있습니다. 자연 로그, 제곱근, 제곱, 역수 및 지수 변환을 사용할 수 있습니다.

차수 지정된 차수의 다항식을 적합시킬 수 있습니다.

중심화 다항식 다항식을 중심화할 수 있습니다. 다항식 중심화를 비활성화 또는 활성화하려면 이 옵션을 선택 취소하거나 선택합니다. 다항식을 중심화하면 회귀 계수가 안정화되고 다중공선성이 감소합니다.

참고 : X 변수의 변환에는 다항식 중심화가 지원되지 않습니다.

절편을 다음 값으로 제약 모형 절편을 지정된 값으로 제한할 수 있습니다.

기울기를 다음 값으로 제약 모형 기울기를 지정된 값으로 제한할 수 있습니다.

팁 : Y = X 선을 그림에 추가하려면 " 특수 적합 " 을 선택한 후 **절편을 다음 값으로 제약 : 0** 및 **기울기를 다음 값으로 제약 : 1** 을 선택합니다.

"변환된 적합" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오. 예는 "특수 적합 옵션의 예"에서 확인하십시오.

스플라인 적합 보고서

이변량 플랫폼에서 **유연 > 스플라인 적합** 옵션을 사용하여 지정된 람다(λ) 값에 따라 평활도(또는 유연성)가 변하는 평활 스플라인을 적합시킵니다. 람다 값은 스플라인 계산식의 조정 모수입니다. λ 값이 작으면 유연한 적합을 위해 모형 오차 항에 높은 가중치를 부여합니다. λ 값이 크면 직선에 근접하는 경직된 적합을 위해 오차 항에 낮은 가중치를 부여합니다.

"스플라인 적합" 메뉴에서 λ를 선택하거나, **스플라인 적합 > 기타**를 선택하여 λ 값을 지정합니다. 또는 "평활 스플라인 적합" 보고서에서 대화식으로 λ를 조정합니다.

참고 : X 변수를 표준화하려면 **스플라인 적합 > 기타**를 선택하고 **X 표준화** 옵션을 선택합니다.

"평활 스플라인 적합" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오. 이 적합에 대한 통계 상세 정보는 "스플라인 적합 옵션에 대한 통계 상세 정보"에서 확인하십시오.

" 평활 스플라인 적합 " 테이블에는 다음 열이 포함됩니다.

R^2 평활 스플라인 모형으로 설명되는 변동의 비율입니다. 자세한 내용은 " 평활 적합 보고서에 대한 통계 상세 정보 " 에서 확인하십시오.

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

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

커널 평활기 보고서

이변량 플랫폼에서 **유연 > 커널 평활기** 옵션을 사용하여 국소 가중 최소 제곱 모형을 데이터에 적합시킵니다. 평활기는 영역의 표집된 지점에서 국소 가중 모형(평균, 선형 또는 2차)을 반복적으로 찾아 구성됩니다. 다수의 로컬 적합(총 512개)을 결합하여 전체 영역에 대한 평활 곡선을 생성합니다. 이 방법을 LOWESS(국소 가중 산점도 평활)라고도 합니다. "커널 평활기" 옵션은 완벽에 가깝게 적합되는 경우에 대한 작은 조정이 있는 Cleveland 방법(1979)을 구현합니다. Cleveland의 Biweight 함수에서 6 * q50 인수는 max(6 * q50, 2 * q90)으로 대체됩니다. 여기서, q50 및 q90은 각각 50번째 및 90번째 백분위수입니다. "로컬 평활기" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오.

" 로컬 평활기 " 보고서에는 다음 열이 포함됩니다.

R^2 평활기 모형으로 설명되는 변동의 비율을 측정합니다. 자세한 내용은 " 평활 적합 보고서에 대한 통계 상세 정보 " 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 적합에 대해 다항식 차수 또는 람다를 지정할 수 있습니다.

가중치 함수 가중치 함수를 지정할 수 있습니다. LOWESS 모형은 트라이큐브 가중치를 사용합니다. 가중치 함수는 각 x_i 및 y_i 가 모형 적합에 미치는 영향을 결정합니다.

평활 (알파) 값을 입력하거나 슬라이더를 이동하여 각 로컬 적합에 대해 고려되는 점 수를 지정할 수 있습니다. 알파는 0 에서 1 사이의 평활 모수입니다. 알파가 증가할수록 곡선이 평활해집니다.

표집 델타 적합 프로세스에 사용되는 표집 비율을 지정할 수 있습니다. 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.

강건성 적합 루틴의 강건성을 지정할 수 있습니다 . 반복할 때마다 적합 곡선에서 멀리 떨어진 점의 영향을 줄이기 위해 점의 가중치를 다시 적용합니다.

각 값 적합 보고서

이변량 플랫폼에서 " 각 값 적합 " 옵션을 사용하여 고유한 각 X 값에 대한 평균 반응을 연결하는 선을 적합시킵니다.

"각 값 적합" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오.

" 각 값 적합 " 테이블에는 모형 적합에 대한 요약 통계량이 포함됩니다 .

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 보고서

이변량 플랫폼에서 " 직교 적합 " 옵션을 사용하여 Deming 회귀 모형을 포함한 직교 회귀 모형을 적합시킵니다 . 직교 모형은 Y 뿐만 아니라 X 의 변동도 설명합니다 .

" 직교 적합 비율 " 테이블에는 직교 회귀 모형에 대한 요약 통계량이 포함됩니다 .

변수 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 사용된 유의 수준입니다 . 유의 수준을 변경하려면 텍스트 상자에 값을 입력합니다 .

"직교 적합 비율" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오. 이 적합에 대한 통계 상세 정보는 "직교 적합 옵션에 대한 통계 상세 정보"에서 확인하십시오. 이 옵션의 예는 "직교 적합 옵션의 예"에서 확인하십시오.

로버스트 적합 보고서

이변량 플랫폼에서 "로버스트" 옵션을 사용하여 로버스트 회귀 모형을 적합시킵니다. 로버스트 방법은 모형에 대한 이상치의 영향을 줄이기 위해 사용됩니다.

"로버스트 적합" 및 "Cauchy 적합" 보고서에는 다음과 같은 요약 통계량이 포함된 두 개의 테이블이 있습니다.

시그마 RMSE(제곱근 평균 제곱 오차) 와 동일한 시그마 값입니다 .

카이제곱 모형이 반응의 평균보다 더 잘 적합된다는 가설에 대한 검정 통계량입니다 .

p 값 기울기가 0 인 카이제곱 검정의 *p* 값입니다 .

LogWorth LogWorth 는 -log$_{10}$(*p* 값) 으로 정의된 *p* 값의 변환입니다 .

모수 모형 모수의 이름입니다 .

로버스트 추정값 각 항에 대한 모수 추정값입니다 . 이 값은 모형 계수의 추정값입니다 .

표준 오차 모수 추정값의 표준 오차 추정값입니다 .

"로버스트 적합" 및 "Cauchy 적합" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오 . 이 옵션의 예는 "로버스트 적합 옵션의 예"에서 확인하십시오 .

Passing-Bablok 적합 보고서

이변량 플랫폼에서 "Passing-Bablok 적합" 옵션을 사용하여 Passing-Bablok 방법으로 회귀 모형을 적합시킵니다 . Passing-Bablok 회귀는 두 가지 다른 분석 방법의 측정값을 비교하기 위해 개발되었습니다 . 이 방법은 강한 상관관계가 있는 두 변수 간의 선형 관계를 가정합니다 . 적합선 및 Y = X 를 나타내는 점선이 산점도에 추가됩니다 . "Passing-Bablok 적합" 의 빨간색 삼각형 메뉴에서 " 매칭 쌍 " 옵션을 사용하여 쌍체 *t*-검정 및 Bland-Altman 분석을 수행합니다 .

"Passing-Bablok 적합 " 보고서에는 세 개의 테이블이 포함되어 있습니다 .

비모수 : Kendall τ

Kendall τ 를 사용하여 X 변수와 Y 변수 간의 상관관계를 평가할 수 있습니다 .

X X 변수입니다 .

Y Y 변수입니다 .

Kendall t X 변수와 Y 변수 간의 비모수 상관 측도입니다 . 값은 -1 에서 1 사이이며 , 값이 0 에 가까우면 X 변수와 Y 변수 간의 독립성을 나타냅니다 .

Prob>|τ| X 변수와 Y 변수 간의 독립성에 대한 가설 검정과 관련된 p 값입니다 . *p* 값이 작으면 변수가 종속적이며 Passing-Bablok 방법이 적절하다는 것을 뒷받침합니다 .

선형성에 대한 CUSUM 검정

" 선형성에 대한 CUSUM 검정 " 테이블에는 선형성 검정 결과가 포함됩니다 . *p* 값이 작으면 선형성에 대한 귀무가설을 기각하여 Passing-Bablok 회귀가 적절하지 않을 수 있음을 나타냅니다 .

최대 CUSUM 잔차의 부호를 기준으로 각 행에 할당되고 각 점에서 Passing-Bablok 선까지의 수직 거리에 따라 정렬되는 $\sqrt{L/I}$ 및 $-\sqrt{I/L}$ 값의 누적합 절대값의 최대값입니다 . *I* 는 양수 잔차를 갖는 관측값 수이고 *L* 은 음수 잔차를 갖는 관측값 수입니다 . 방법 간에 강한 상관관계가 있으면 *I* 가 *L* 과 같습니다 . 따라서 누적합이 +1 과 -1 의 합인 경우가 종종 있습니다 . CUSUM 값이 작으면 Passing-Bablok 선 양쪽에 점이 랜덤 분포되어 있음을 나타내고 이 결과는 선형성 가설을 뒷받침합니다 .

H CUSUM 검정에 대한 검정 통계량입니다 . 이 검정 통계량은 최대 CUSUM 을 음수 잔차 수 + 1 의 제곱근으로 나누어 정의됩니다 . 이 검정 통계량은 Kolmogorov-Smirnov 분포를 따릅니다 .

Prob > H CUSUM 검정의 *p* 값입니다 . *p* 값이 작으면 Passing-Bablok 절차가 적절하지 않을 수 있음을 나타냅니다 .

모수 추정값

"모수 추정값" 테이블에는 95% 신뢰 구간을 갖는 절편 및 기울기의 Passing-Bablok 적합 추정값이 포함됩니다.

매칭 쌍 보고서

"Passing-Bablok 적합 " 의 빨간색 삼각형 메뉴에서 "Bland Altman 분석 " 옵션을 사용하여 쌍체 *t*-검정 및 Bland-Altman 분석을 수행합니다. 매칭 쌍에 대한 자세한 내용은 **예측 및 전문 모델링** 의 에서 확인하십시오 .

Bland-Altman 분석

"Bland-Altman 분석 " 테이블에는 다음 모수에 대한 값 , 표준편차 및 신뢰 한계가 포함됩니다 .

편향 X 변수와 Y 변수 간의 평균 차이입니다 .

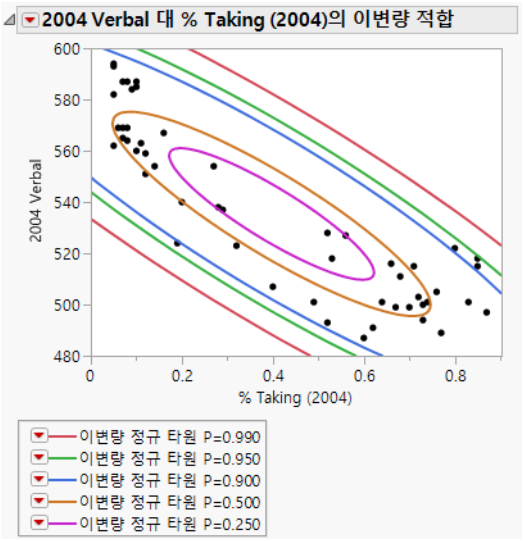
합치도 한계 편향 $\pm z_{1-\alpha/2}$*(편향의 표준편차) 로 설정된 합치도 상한 및 하한입니다 .

밀도 타원 보고서

이변량 플랫폼에서 "밀도 타원" 옵션을 사용하여 지정된 점 집합이 포함된 타원을 하나 이상 그리고 요인 간의 상관관계를 추정 및 검정합니다. 타원 안에 포함되는 점 개수는 "밀도 타원" 메뉴에서 선택한 확률 값에 따라 결정됩니다.

"이변량 정규 타원" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오. 이 옵션의 예는 "밀도 타원을 사용한 그룹화의 예"에서 확인하십시오.

그림 5.13 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.

타원은 이변량 정규 분포를 따른다는 가정하에 지정된 백분율의 데이터가 놓일 것으로 기대되는 영역을 보여 줍니다. 밀도 타원의 모양은 두 변수 사이의 상관관계를 나타내는 그래픽 표시자입니다. 좁은 타원은 강한 상관관계를 나타내고 원형에 더 가까운 타원은 두 변수 간의 상관관계가 낮음을 나타냅니다.

팁 : 여러 변수 쌍에 대한 밀도 타원 및 상관계수를 행렬 형식으로 보려면 **분석 > 다변량 방법** 메뉴에서 **다변량** 플랫폼을 사용합니다 . 자세한 내용은 **다변량 방법**의 에서 확인하십시오 .

"이변량 정규 타원" 보고서에는 밀도 타원에 대한 요약 통계량 테이블이 포함됩니다.

변수 타원을 생성하는 데 사용된 변수의 이름입니다 .

평균 각 변수의 평균입니다 .

표준편차 각 변수의 표준편차입니다 .

상관 Pearson 상관계수입니다 . 자세한 내용은 " 이변량 정규 타원 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 .

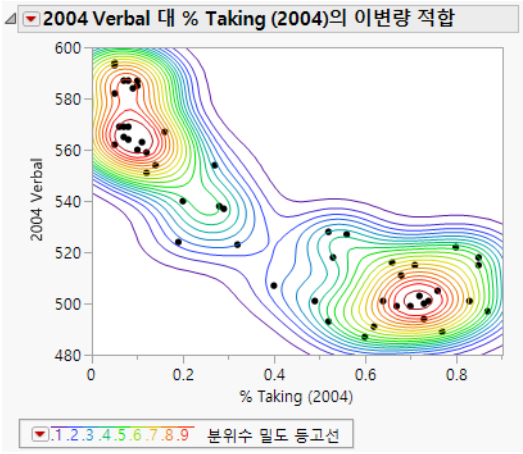
유의 확률 X 변수와 Y 변수 간에 선형 관계가 없는 경우 상관계수의 절대값이 관측값보다 더 클 확률입니다 .

개수 계산에 사용된 관측값의 개수입니다 .

비모수 밀도 보고서

이변량 플랫폼에서 " 비모수 밀도 " 옵션을 사용하여 평활 비모수 이변량 표면을 데이터에 적합시킵니다 . 비모수 밀도 적합은 등고선 집합을 사용하여 시각화됩니다 . 등고선은 5% 구간에 있습니다 . 즉 , 추정된 비모수 분포에서 생성된 점 중 약 5% 가 가장 안쪽 등고선 내에 있고 10% 가 다음 등고선 내에 있는 식입니다 . 가장 바깥쪽 등고선에 점의 약 95% 가 포함됩니다 .

그림 5.14 비모수 밀도의 예



비모수 밀도 등고선 격자의 크기를 변경하려면 "이변량"의 빨간색 삼각형 메뉴에서 Shift 키를 누른 채로 **비모수 밀도**를 선택합니다. 기본값인 102보다 큰 값을 입력합니다.

"분위수 밀도 등고선" 메뉴의 옵션에 대한 자세한 내용은 "이변량 적합 옵션"에서 확인하십시오.

팁 : 데이터 점의 수가 많은 경우 " 비모수 밀도 " 옵션을 사용하여 산점도에서 패턴을 볼 수 있습니다 .

"분위수 밀도 등고선" 보고서에는 비모수 밀도를 생성하는 데 사용된 표준편차 테이블이 포함됩니다.

변수 등고선을 생성하는 데 사용된 변수의 이름입니다 .

커널 표준편차 커널 대역폭입니다 . 초기값은 변수의 표준편차를 기반으로 합니다 . 슬라이더를 조정하여 등고선의 밀도를 높이거나 낮출 수 있습니다 .

이변량 플랫폼의 추가 예

이 섹션에는 이변량 플랫폼을 사용한 예가 포함되어 있습니다.

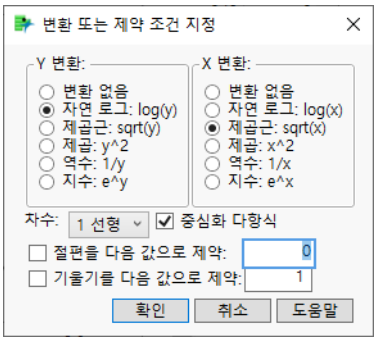
- "특수 적합 옵션의 예"
- "직교 적합 옵션의 예"
- "로버스트 적합 옵션의 예"
- "밀도 타원을 사용한 그룹화의 예"
- "회귀선을 사용한 그룹화의 예"
- "기준 변수를 사용한 그룹화의 예"

특수 적합 옵션의 예

이 예에서는 이변량 플랫폼을 사용하여 회귀 모형을 변환된 변수에 적합시키는 방법을 보여 줍니다.

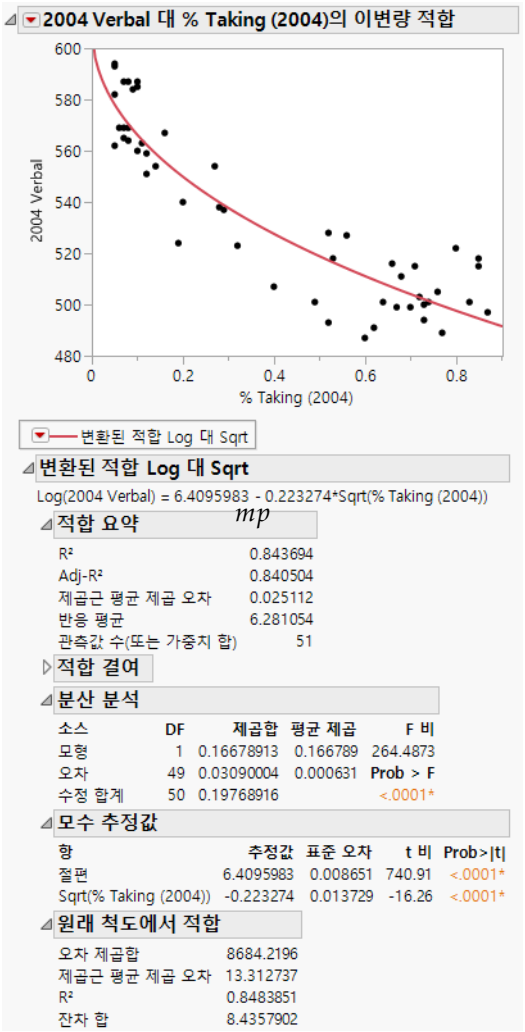
1. **도움말 > 샘플 데이터 폴더**를 선택하고 SAT.jmp 를 엽니다.
2. **분석 > X 로 Y 적합**을 선택합니다.
3. 2004 Verbal 을 선택하고 **Y, 반응**을 클릭합니다.
4. % Taking (2004) 을 선택하고 **X, 요인**을 클릭합니다.
5. **확인**을 클릭합니다.
6. "이변량 적합"의 빨간색 삼각형을 클릭하고 **특수 적합**을 선택합니다. "변환 또는 제약 조건 지정" 창이 나타납니다. 이 창에 대한 자세한 내용은 "특수 적합 창"에서 확인하십시오.
7. "Y 변환"에 대해 **자연 로그 : log(y)** 를 선택합니다.
8. "X 변환"에 대해 **제곱근 : sqrt(x)** 를 선택합니다.

그림 5.15 변환 또는 제약 조건 지정 창



9. **확인**을 클릭합니다.

그림 5.16 특수 적합 보고서의 예



모형이 데이터를 잘 적합시키는 것으로 나타납니다. 그림에 표시된 적합은 원래 척도에서 점 구름을 통과하고 모형 $R^2 = 0.84$ 입니다.

직교 적합 옵션의 예

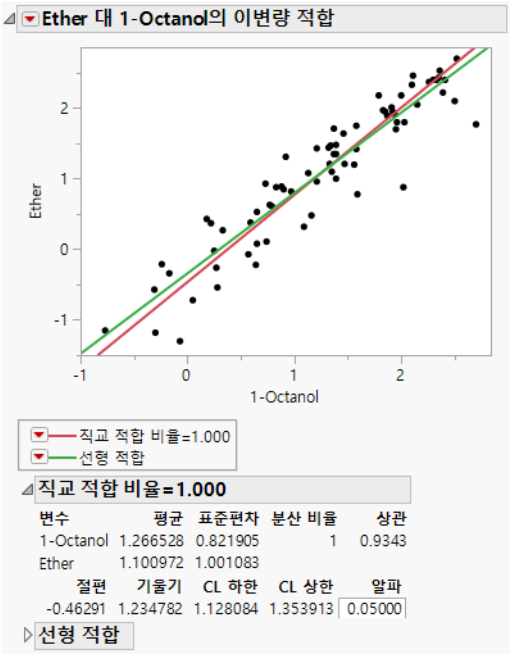
이 예에서는 이변량 플랫폼에서 직교 모형을 사용한 방법 비교를 보여 줍니다 . 방법 비교 데이터에 대해 Passing-Bablok 회귀를 고려할 수도 있습니다 . 자세한 내용은 "Passing-Bablok 적합 보고서 " 에서 확인하십시오 . 서로 다른 용매에서 화합물의 용해도가 측정되었습니다 .

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Solubility.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .

3. Ether 를 선택하고 **Y, 반응**을 클릭합니다 .
4. 1-Octanol 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **직교 적합 > 등분산**을 선택합니다 .
7. 참조를 위해 " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **특수 적합**을 선택합니다 .
8. **절편을 다음 값으로 제약**을 선택하고 0 으로 설정합니다 .
9. **기울기를 다음 값으로 제약**을 선택하고 1 로 설정합니다 .

   이렇게 하면 절편이 0 이고 기울기가 1 인 $Y = X$ 선을 적합시킵니다 .
10. **확인**을 클릭합니다 .

그림 5.17 직교 적합의 예



대부분의 데이터 점에 대해 빨간색 직교 적합선이 녹색 $Y = X$ 선 아래에 있고, 직교 적합의 기울기에 대한 95% 신뢰 구간(CL 하한 및 CL 상한)에 1 이 포함되지 않습니다. Ether 에 대한 용해도를 측정한 화합물은 1-Octanol 에서 측정한 화합물보다 낮게 측정되는 경향이 있습니다.

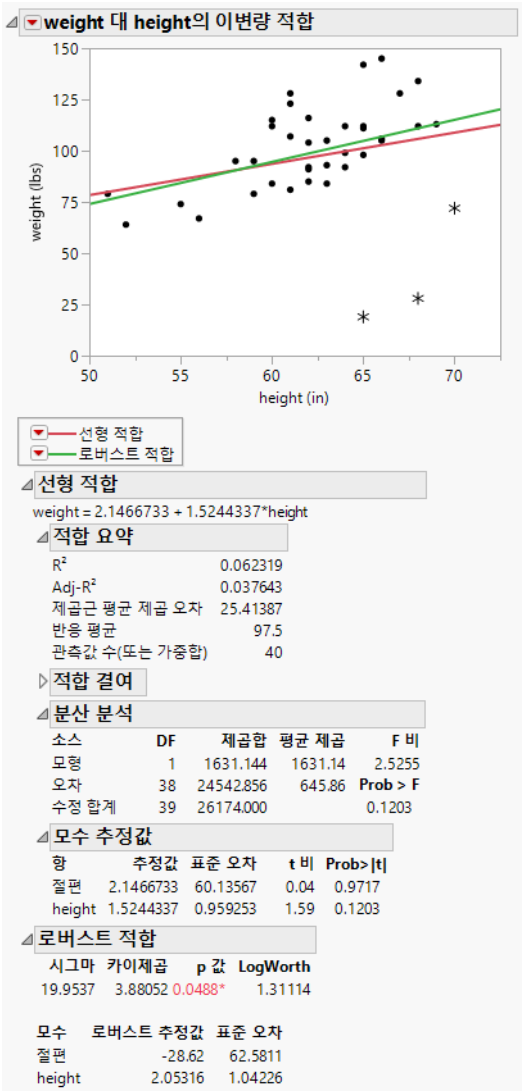
로버스트 적합 옵션의 예

이 예에서는 이변량 플랫폼에서 Huber M- 추정 방법을 사용하여 로버스트 모형을 적합시키는 방법을 보여 줍니다 .

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Weight Measurements.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 .
6. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다 .
7. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **로버스트 > 로버스트 적합**을 선택합니다 .

그림 5.18 로버스트 적합의 예



산점도에서 세 개의 측정값이 별표로 표시되어 있습니다. 이 세 개 관측값의 가중치가 예상보다 낮은 것 같습니다. 이러한 점은 회귀선(빨간색 선으로 표시)을 적합시킬 때 적합에 영향을 미칩니다. "분산 분석" 보고서의 p 값 0.1203은 height와 weight 간에 약한 선형 관계가 있음을 나타냅니다. 그러나 로버스트 적합(녹색 선으로 표시)을 고려해 보면 height와 weight 간의 선형 관계가 표준 적합보다 강합니다. "로버스트 적합"의 p 값 0.0488은 height와 weight 간의 선형 관계에 대한 가설을 뒷받침합니다. 비정상적으로 낮은 가중치로 식별된 측정값이 분석에 영향을 미쳤습니다.

밀도 타원을 사용한 그룹화의 예

이 예에서는 이변량 플랫폼에서 그룹화 (기준) 변수를 사용하고 데이터에 밀도 타원을 추가하는 방법을 보여 줍니다 . 이 예의 데이터 테이블은 세 가지 핫도그 종류를 식별합니다 (beef, meat 또는 poultry). 이 세 종류의 핫도그를 비용 변수에 따라 그룹화하려고 합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 이변량 적합 " 의 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. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **밀도 타원 > 0.90** 을 선택합니다 .

Type에 따라 다른 점 색상을 적용합니다.

10. 산점도를 마우스 오른쪽 버튼으로 클릭하고 **행 범례**를 선택합니다 .
11. 열 목록에서 Type 을 선택하고 **확인**을 클릭합니다 .

그림 5.19 그룹화의 예

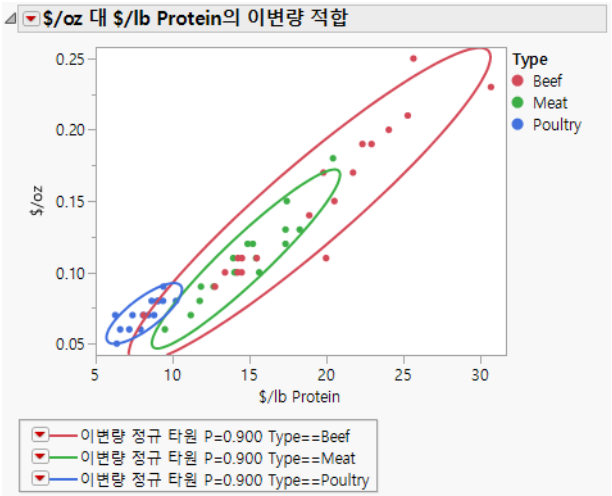


그림 5.19 의 타원은 서로 다른 종류의 핫도그들이 비용 변수에 따라 어떻게 군집화되는지를 명확히 보여 줍니다.

회귀선을 사용한 그룹화의 예

이 예에서는 여러 그룹의 기울기를 비교할 수 있도록 이변량 플랫폼에서 그룹화 변수를 사용하여 회귀선을 중첩하는 방법을 보여 줍니다.

1. **도움말 > 샘플 데이터 폴더**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 5.20 의 왼쪽 예와 같은 그림을 생성하려면 다음을 수행하십시오.

6. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.

그림 5.20 의 오른쪽 예와 같은 그림을 생성하려면 다음을 수행하십시오.

7. " 선형 적합 " 메뉴에서 **적합 제거**를 선택합니다.
8. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **그룹화 기준**을 선택합니다.
9. 목록에서 sex 를 선택합니다.
10. **확인**을 클릭합니다.
11. " 이변량 적합 " 의 빨간색 삼각형을 클릭하고 **선형 적합**을 선택합니다.

그림 5.20 전체 표본 및 그룹화된 표본에 대한 회귀 분석의 예

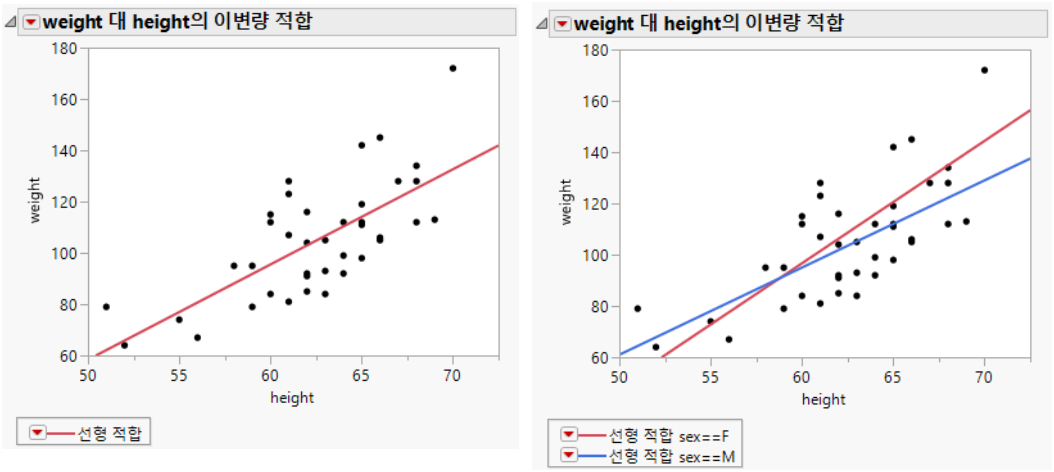


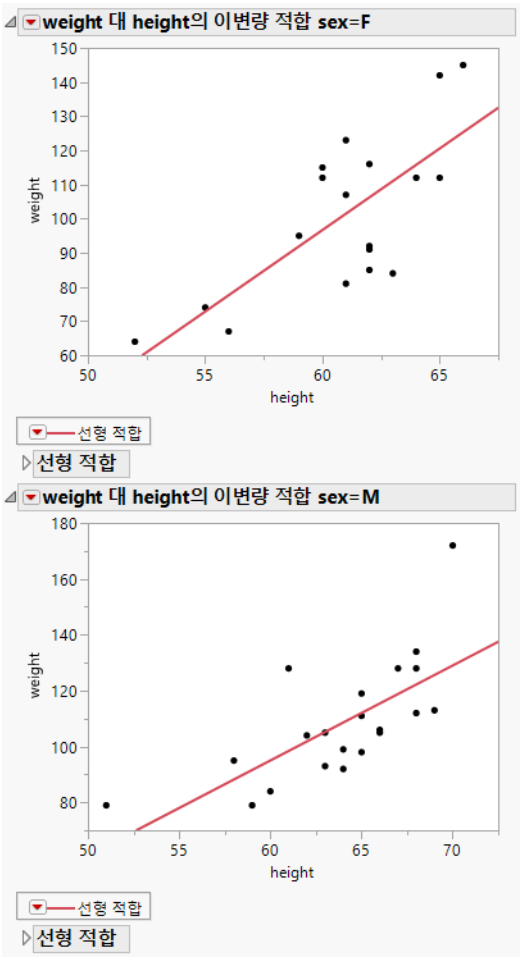
그림 5.20의 왼쪽 산점도에는 체중과 키의 상관관계를 나타내는 단일 회귀선이 있습니다. 오른쪽 산점도에서는 남학생과 여학생의 회귀선을 개별적으로 보여 줍니다.

기준 변수를 사용한 그룹화의 예

이 예에서는 이변량 시작 창에서 기준 변수를 지정하여 그룹화하는 방법을 보여 줍니다. 그러면 기준 변수 (또는 여러 기준 변수의 조합) 의 각 수준별로 별도의 보고서와 그래프가 표시됩니다.

1. **도움말 > 샘플 데이터 폴더**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 누르고 " 이변량 적합 " 의 빨간색 삼각형을 클릭한 후 **선형 적합**을 선택합니다 .

그림 5.21 기준 변수 그림의 예



기준 변수(sex)의 각 수준에 대해 개별 분석이 표시됩니다. 즉, 여학생에 대한 산점도와 남학생에 대한 산점도가 표시됩니다.

이변량 플랫폼에 대한 통계 상세 정보

이 섹션에는 이변량 플랫폼에 대한 통계 상세 정보가 포함되어 있습니다.

- " 선형 적합 옵션에 대한 통계 상세 정보 "
- " 다항식 적합 옵션에 대한 통계 상세 정보 "
- " 스플라인 적합 옵션에 대한 통계 상세 정보 "
- " 직교 적합 옵션에 대한 통계 상세 정보 "
- " 로버스트 옵션에 대한 통계 상세 정보 "
- "Passing-Bablok 적합 옵션에 대한 통계 상세 정보 "
- " 적합 요약 보고서에 대한 통계 상세 정보 "
- " 적합 결여 보고서에 대한 통계 상세 정보 "
- " 모수 추정값 보고서에 대한 통계 상세 정보 "
- " 평활 적합 보고서에 대한 통계 상세 정보 "
- " 이변량 정규 타원 보고서에 대한 통계 상세 정보 "

선형 적합 옵션에 대한 통계 상세 정보

이변량 플랫폼에서 선형 적합 옵션은 잔차 제곱합을 최소화하도록 점을 적합시키는 직선의 β_0 및 β_1 모수를 찾습니다. i번째 행에 대한 모형은 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$로 기록됩니다.

선형 적합에 대한 평균 신뢰 한계는 다음과 같이 정의됩니다 .

$$\hat{y}_i \pm t_{\alpha/2,\, n-2} \left[\sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}} \right] \hat{\sigma}$$

선형 적합에 대한 개별값 신뢰 한계는 다음과 같이 정의됩니다.

$$\hat{y}_i \pm t_{\alpha/2,\, n-2} \left[\sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}} \right] \hat{\sigma}$$

다음은 각 요소에 대한 설명입니다.

$\hat{y}_i$ = i 번째 관측값인 x_i 에서의 예측값

$\bar{x}$ = x 변수의 평균

$$\hat{\sigma} = \sqrt{\frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2}}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$S_{xy} = \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})$$

$t_{\alpha/2,n}$ = 자유도가 n–2 인 중심 t 분포의 (α/2) 번째 분위수

n = 표본 크기

참고 : " 평균 신뢰 한계 계산식 " 및 " 개별값 신뢰 한계 계산식 " 옵션을 사용하여 모형 적합에 대한 한계 계산식을 데이터 테이블의 새 열에 저장할 수 있습니다 . 계산식은 행렬 형식으로 저장됩니다 .

다항식 적합 옵션에 대한 통계 상세 정보

2 차 다항식은 포물선으로 나타나고 3 차 다항식은 3 차 곡선으로 나타납니다 . k 차의 경우 i 번째 행에 대한 모형은 다음과 같습니다 .

$$y_i = \sum_{j=0}^{k} \beta_j x_i^j + \varepsilon_i$$

다음은 각 요소에 대한 설명입니다.

y_i = i 번째 관측값인 x_i 에서의 반응 값

β_j = j 번째 모수

k = 다항식의 차수

ε_i = 모형의 오차 항

스플라인 적합 옵션에 대한 통계 상세 정보

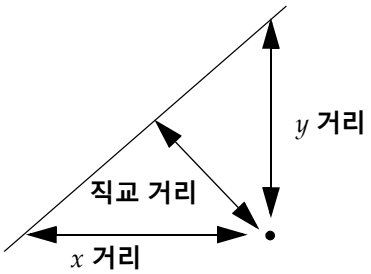
이변량 플랫폼에서 3차 스플라인 방법은 결과 곡선이 매듭 점에서 연속적이고 부드럽게 되도록 매듭지어진 일련의 3차 다항식을 사용합니다. 추정은 오차 제곱합과 곡선 범위에 통합된 곡률에 대한 벌점의 조합인 목적 함수를 최소화하는 방법으로 수행됩니다. 이 방법에 대한 자세한 내용은 Reinsch 연구 자료 (1967) 또는 Eubank 학술 자료 (1999) 에서 확인하십시오.

직교 적합 옵션에 대한 통계 상세 정보

이변량 플랫폼에서 "직교 적합" 옵션을 사용하여 수직 차이의 제곱합을 최소화하는 선을 적합시킬 수 있습니다. X 측정값에 랜덤 변동이 있는 경우 표준 최소 제곱보다 이 방법을 사용하는 것이 좋습니다. 표준 최소 제곱 적합은 X 변수가 고정되어 있고 Y 변수가 X + 오차의 함수라고 가정합니다.

참고 : 수직 거리는 X 및 Y 의 척도화 방법에 따라 달라지며 , 수직 거리의 척도화는 그래픽 문제가 아니라 통계적 문제로 남아 있습니다 .

그림 5.22 적합선에 수직인 선



적합을 수행하려면 X의 오차 분산에 대한 Y의 오차 분산의 비율을 지정해야 합니다. 이는 표본점의 분산이 아니라 오차의 분산이므로 신중하게 선택해야 합니다. 표준 최소 제곱에서는 σ_x^2 이 0이기 때문에 $(\sigma_y^2)/(\sigma_x^2)$ 비율이 무한대가 됩니다. 오차 비율이 큰 직교 적합은 표준 최소 제곱 적합선에 가까워집니다. 비율을 0으로 지정하면 X에 대한 Y가 아니라 Y에 대한 X의 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. 오차 비율은 1(등분산)과 같이 가정된 값이거나, 측정 시스템에 대한 지식을 기반으로 할 수 있습니다.

신뢰 한계는 Tan and Iglewicz 연구 자료(1999)에 설명된 것과 같이 계산됩니다.

로버스트 옵션에 대한 통계 상세 정보

이 섹션에는 이변량 플랫폼의 " 로버스트 적합 " 및 "Cauchy 적합 " 옵션에 대한 상세 정보가 포함되어 있습니다 .

로버스트 적합

이변량 플랫폼에서 " 로버스트 적합 " 옵션은 모형 적합에 대한 반응 변수 이상치의 영향을 줄입니다 . Huber M- 추정 방법이 사용됩니다 . Huber M- 추정은 다음과 같이 Huber 손실 함수를 최소화하는 모수 추정값을 찾는 방법입니다 .

$$l(\varrho) = \sum_i \rho(e_i)$$

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

이변량 플랫폼에서 "Cauchy 적합" 옵션은 최대 가능도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다. 이 방법에서는 오차가 Cauchy 분포를 따른다고 가정합니다. Cauchy 분포는 정규 분포보다 꼬리가 더 두꺼우므로 이상치의 영향이 줄어듭니다.

Passing-Bablok 적합 옵션에 대한 통계 상세 정보

이변량 플랫폼에서 "Passing-Bablok 적합" 옵션은 Passing-Bablok 절차를 사용하여 선형 방정식 $y = \beta_0 + \beta_1 x$ 를 적합시킵니다. 이때 x 와 y 둘 다 오차를 사용하여 측정됩니다. 이 방법은 이상치에 로버스트하며, x 와 y 사이에 선형 관계가 있을 때 적절합니다. 기울기 (β_1) 추정값은 정의되지 않은 0/0 기울기 또는 -1 기울기를 초래하는 희귀 쌍을 제외하고 가능한 모든 데이터 점 쌍에서 구성될 수 있는 모든 기울기의 중앙값으로 계산됩니다. 이러한 기울기의 독립성 결여로 인한 추정 편향을 수정하기 위해 중앙값이 K(-1 보다 작은 기울기의 수) 기반 계수만큼 이동됩니다. 결과는 β_1 에 대한 근사 비편향 추정량입니다. 절편 (β_0) 은 $\{y_i - \beta_1 x_i\}$ 의 중앙값으로 추정됩니다. 자세한 내용은 Passing-Bablok(1983), Passing-Bablok(1984) 및 Bablok et. al. (1988) 연구 자료에서 확인하십시오.

Passing-Bablok 은 방법 비교 연구에 흔히 사용됩니다. 절편은 두 방법 간의 체계적 편향(차이)으로 해석됩니다. 기울기는 두 방법 간의 비례 편향(차이) 크기를 측정합니다.

적합 요약 보고서에 대한 통계 상세 정보

이 섹션에는 이변량 플랫폼의 " 적합 요약 " 보고서에 대한 상세 정보가 포함되어 있습니다 .

R^2

해당 분산 분석 테이블에서 구한 통계량을 사용할 경우 모든 연속형 반응 적합의 R^2 는 다음과 같이 계산됩니다 .

$$\frac{\text{모형의 제곱합}}{\text{Sum of Squares for C. Total}}$$

Adj-R^2

Adj-R^2 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 5.11). **수정 합계**의 제곱합은 **수정 합계**의 **제곱합**을 해당 자유도로 나눠서 계산할 수 있습니다 .

참고 : 절편 또는 기울기에 제약 조건이 있는 경우 R^2 및 Adj-R^2 이 보고되지 않습니다 .

적합 결여 보고서에 대한 통계 상세 정보

이 섹션에는 이변량 플랫폼의 " 적합 결여 " 보고서에 대한 상세 정보가 포함되어 있습니다 .

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 같이 정의됩니다 .

$$DF_p = \sum_{i=1}^{g} (n_i - 1)$$

여기서 n_i 는 i 번째 그룹의 관측값 수입니다 .

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 정의됩니다 .

$$SS_p = \sum_{i=1}^{g} SS_i$$

여기서 SS_i는 평균에 대해 수정된 i번째 그룹의 제곱합입니다.

최대 R²

순수 오차는 모형 형태에 따라 변동되지 않으며 가능한 최소 분산이므로 최대 R² 는 다음과 같이 계산됩니다 .

$$1 - \frac{\text{SS(순수 오차)}}{\text{SS(Total for whole model)}}$$

모수 추정값 보고서에 대한 통계 상세 정보

이 섹션에는 이변량 플랫폼의 " 모수 추정값 " 보고서에 대한 상세 정보가 포함되어 있습니다 .

표준화 베타

표준화 베타는 다음과 같이 계산됩니다 .

$$\hat{\beta}(s_x / s_y)$$

여기서 $\hat{\beta}$ 는 추정된 모수이고, s_x 및 s_y는 X 및 Y 변수의 표준편차입니다.

설계 표준 오차

설계 표준 오차는 모수 추정값의 표준 오차를 RMSE로 나눠서 계산됩니다.

평활 적합 보고서에 대한 통계 상세 정보

이 섹션에는 이변량 플랫폼의 " 평활 적합 " 보고서에 대한 상세 정보가 포함되어 있습니다 .

R²은 1 - (SSE/수정 합계 SS)와 같습니다. 여기서 수정 합계 SS는 선형 적합 ANOVA 보고서에서 확인할 수 있습니다.

이변량 정규 타원 보고서에 대한 통계 상세 정보

이 섹션에는 이변량 플랫폼의 "이변량 정규 타원" 보고서에 대한 상세 정보가 포함되어 있습니다.

Pearson 상관계수는 r 로 표기되며 다음과 같이 계산됩니다 .

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} \text{ 여기서 , } s_{xy} = \frac{\sum w_i (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{df}$$

여기서 w_i는 i번째 관측값의 가중치(가중치 열이 지정된 경우) 또는 1(가중치 열이 할당되지 않은 경우)입니다.

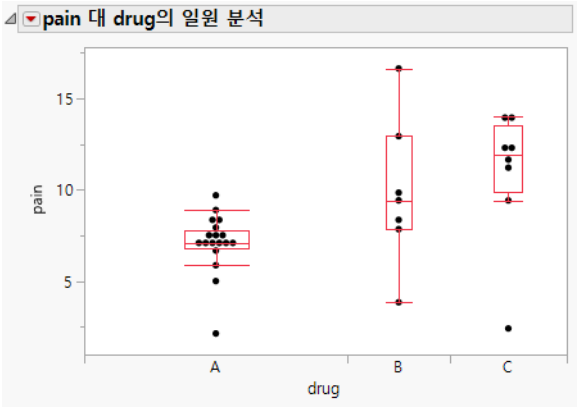
6 장

일원 분석
연속형 Y 와 범주형 X 변수 간의 관계 분석

범주형 X 변수로 정의된 그룹 사이에서 연속형 Y 변수의 분포가 어떻게 달라지는지를 탐색하려면 일원 분석 플랫폼을 사용합니다. ANOVA(분산 분석), ANOM(평균 분석), *t*-검정, 동등성 검정, 비모수 검정, 정확 검정 또는 다중 비교를 사용하여 그룹 간 평균 또는 분산의 차이를 평가할 수 있습니다. 밀도 및 CDF(누적 분포 함수) 그림을 사용하면 X 범주별 반응의 분포를 시각화할 수 있습니다. 예를 들어 동일한 약품 유형(X)의 서로 다른 범주가 수치 척도로 측정된 환자의 통증 수준(Y)에 어떤 영향을 미치는지 확인할 수 있습니다.

일원 분석 플랫폼은 X로 Y 적합 플랫폼의 **연속형 대 명목형 또는 순서형** 분석법입니다.

그림 6.1 일원 분석



목차

일원 분석 플랫폼 개요 . . . 138
일원 분석의 예 . . . 138
일원 분석 플랫폼 시작 . . . 141
데이터 형식 . . . 141
일원 분석 보고서 . . . 142
일원 분석 플랫폼 옵션 . . . 143
일원 분석 보고서 . . . 150
분위수 보고서 . . . 151
평균/ANOVA/합동 t 보고서 . . . 151
평균 및 표준편차 보고서 . . . 153
t-검정 보고서 . . . 154
평균 분석 보고서 . . . 154
평균 비교 보고서 . . . 156
비모수 검정 보고서 . . . 159
비모수 다중 비교 보고서 . . . 161
이분산 보고서 . . . 163
동등성 검정 보고서 . . . 165
로버스트 적합 보고서 . . . 168
검정력 보고서 . . . 168
매칭 열 보고서 . . . 170
일원 분석 그림 요소 . . . 170
평균 다이아몬드 및 X 축 비례 . . . 171
평균 선, 오차 막대 및 표준편차 선 . . . 172
비교 원 . . . 172
일원 분석 플랫폼의 추가 예 . . . 174
평균 분석의 예 . . . 174
분산에 대한 평균 분석의 예 . . . 175
개별 쌍 비교(스튜던트 t) 검정의 예 . . . 176
전체 쌍 비교(Tukey HSD) 검정의 예 . . . 178
최량 비교(Hsu MCB) 검정의 예 . . . 179
대조군 비교(Dunnett)의 예 . . . 181
단계별 개별 쌍 비교(Newman-Keuls) 검정의 예 . . . 182
예: 4가지 평균 비교 검정 대조 . . . 183
비모수 Wilcoxon 검정의 예 . . . 184
이분산 옵션의 예 . . . 185
동등성 검정의 예 . . . 187
로버스트 적합 옵션의 예 . . . 188
검정력 옵션의 예 . . . 189

정규 분위수 그림의 예 191
CDF 그림의 예 192
밀도 옵션의 예 193
매칭 열 옵션의 예 194
예: 일원 분석을 위한 데이터 쌓기 195
일원 분석 플랫폼에 대한 통계 상세 정보 202
비교 원에 대한 통계 상세 정보 202
검정력에 대한 통계 상세 정보 203
ANOM에 대한 통계 상세 정보 204
적합 요약 보고서에 대한 통계 상세 정보 205
분산 동일 여부 검정에 대한 통계 상세 정보 205
비모수 검정 통계량에 대한 통계 상세 정보 206
로버스트 적합에 대한 통계 상세 정보 209

일원 분석 플랫폼 개요

일원 분석 플랫폼을 사용하면 범주형 X 변수의 두 개 이상의 수준에 대해 연속형 Y 변수의 분포를 빠르게 시각화할 수 있습니다. ANOVA(분산 분석) 모형, *t*-검정, ANOM(평균 분석), 동등성 검정, 비모수 검정, 정확 검정 또는 다중 비교를 실행하여 X 변수의 전체 수준에서 평균 또는 분산의 차이를 평가할 수 있습니다. 많은 검정에서 보고서의 일원 분석 그림에 시각화 도구를 추가합니다. 또한 밀도 및 CDF(누적 분포 함수) 그림을 사용하여 X 변수의 전체 수준에 대한 반응 분포를 시각화하는 옵션도 있습니다.

검정은 검정 유형에 따라 "일원 분석"의 빨간색 삼각형 메뉴에서 그룹화됩니다. JMP를 사용하면 다양한 유형의 통계적 검정을 탐색할 수 있습니다. 연구 설계, 데이터 수집, 데이터 분포 및 관심 가설에 따라 최선의 검정이 선택됩니다.

검정 보고서에 대한 자세한 내용은 다음 섹션에서 확인하십시오 .

- " 분위수 보고서 "
- " 평균 /ANOVA/ 합동 t 보고서 "
- " 평균 및 표준편차 보고서 "
- "t- 검정 보고서 "
- " 평균 분석 보고서 "
- " 평균 비교 보고서 "
- " 비모수 검정 보고서 "
- " 비모수 다중 비교 보고서 "
- " 이분산 보고서 "
- " 동등성 검정 보고서 "
- " 로버스트 적합 보고서 "
- " 검정력 보고서 "
- " 매칭 열 보고서 "

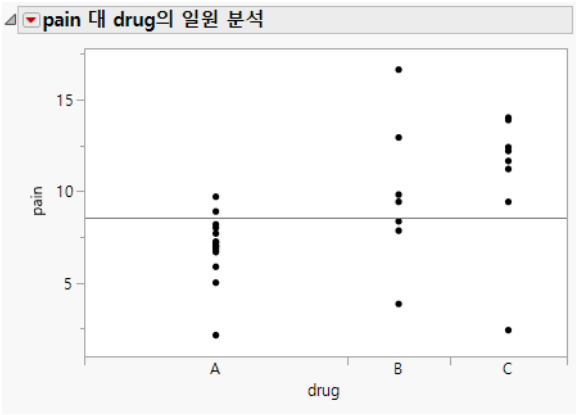
일원 분석의 예

일원 분석 플랫폼을 사용하여 일원 분산 분석을 수행합니다 . 이 분석을 통해 그룹 평균에서 통계적으로 유의한 차이를 검정할 수 있습니다 . 이 예에서는 33 명의 참가자에게 세 가지 유형의 진통제 (A, B, C) 를 투여했습니다 . 참가자에게 통증 수준을 슬라이딩 척도로 평가하게 했습니다 . 처리 A, B 및 C 의 평균 통증 수준이 통계적으로 유의하게 다른지 여부를 확인하려고 합니다 .

1. **도움말 > 샘플 데이터 폴더**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



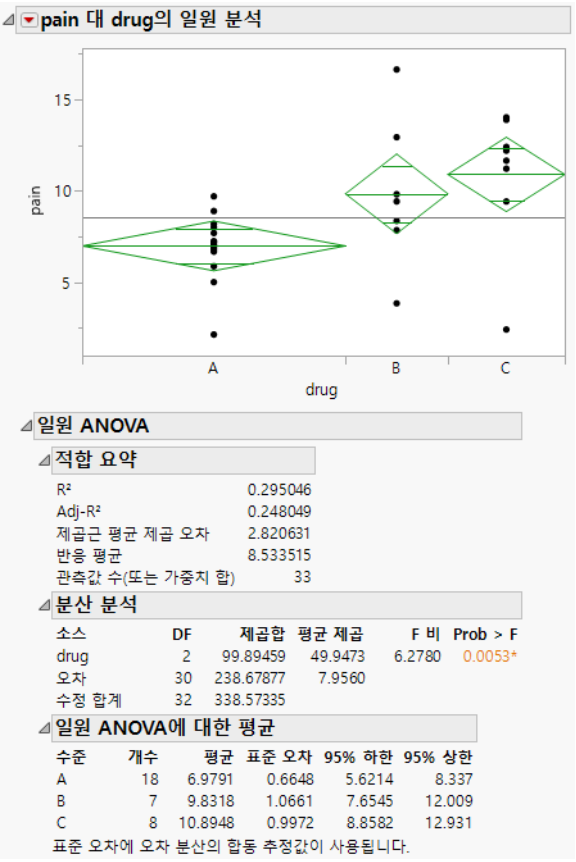
한 가지 약물(A)이 다른 약물보다 일관되게 통증 스코어가 낮음을 알 수 있습니다. 또한 *x* 축의 눈금 간격이 일정하지 않습니다. 눈금 사이의 길이는 각 약물에 대한 스코어(관측값)의 개수와 비례합니다.

데이터에 대해 분산 분석을 수행합니다 .

6. " 일원 분석 " 의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .

참고 : X 변수의 수준이 두 개뿐인 경우 **평균 /ANOVA** 옵션이 **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 .

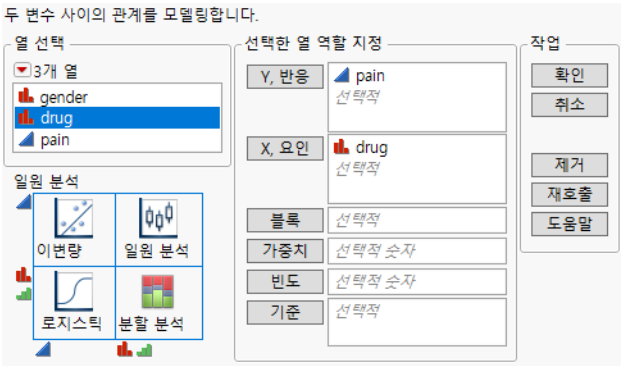
- 신뢰 구간을 나타내는 평균 다이아몬드가 그림에 추가됩니다 .
  - 각 다이아몬드의 중심에 있는 선은 그룹 평균을 나타냅니다 . 약물 A 의 평균이 약물 B 및 C 의 평균보다 낮음을 한 눈에 알 수 있습니다 .
  - 각 다이아몬드의 세로 범위는 각 그룹의 평균에 대한 95% 신뢰 구간을 나타냅니다 .

  자세한 내용은 " 평균 다이아몬드 및 X 축 비례 " 에서 확인하십시오 .
- " 적합 요약 " 테이블에서는 분석에 대한 전체적인 요약 정보를 제공합니다 .
- " 분산 분석 " 보고서에서는 표준 ANOVA 정보를 보여 줍니다 . Prob > F(*p* 값) 는 0.0053 으로 나타나는데 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

분석 > X 로 Y 적합을 선택하여 일원 분석 플랫폼을 시작할 수 있습니다 . X 로 Y 적합 시작 창은 네 가지 유형의 분석에 사용됩니다 . 연속형 Y 변수와 범주형 명목형 또는 순서형 X 변수를 입력하면 일원 분석 플랫폼이 시작됩니다 . X 및 Y 변수의 각 조합에 대한 보고서가 시작됩니다 .

그림 6.4 일원 분석 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 . 일원 분석 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 . 이러한 변수의 데이터 유형은 숫자여야 합니다 .

X, 요인 분석할 하나 이상의 예측 변수입니다 . 이러한 변수의 데이터 유형은 명목형 또는 순서형이어야 합니다 .

블록 블록 변수를 지정하는 열입니다 . 사용할 경우 Y 변수의 값이 블록 변수에 의해 중심화됩니다 . 그룹 셀별로 각 블록에 동일한 개수가 있는 경우 사용 가능한 일원 분석 플랫폼 옵션은 이원 분석과 관련이 있습니다 . 그룹 셀별로 블록에 동일하지 않은 개수가 있는 경우 일원 ANOVA - 블록화 모형이 적합되고 일원 분석 플랫폼 옵션을 사용할 수 없습니다 .

가중치 데이터 테이블의 각 관측값에 대한 가중치를 포함하는 열입니다 . 값이 0 보다 큰 행만 분석에 포함됩니다 .

빈도 분석의 각 행에 빈도를 할당합니다 . 빈도 할당은 데이터를 요약할 때 유용합니다 .

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

데이터 형식

일원 분석 플랫폼에서 데이터는 요약되지 않거나 요약된 데이터 열로 구성될 수 있습니다 .

요약되지 않은 데이터 각 관측값에 대한 하나의 행과 X 및 Y 값에 대한 열이 있습니다 .

요약된 데이터 각 행이 공통된 X 및 Y 값을 갖는 일련의 관측값을 나타냅니다. 데이터 테이블에는 각 행에 대한 관측값 수의 빈도 열이 포함되어야 합니다. 시작 창에서는 이 열을 "빈도" 열로 입력합니다.

일원 데이터가 JMP 데이터 테이블이 아닌 다른 형식으로 존재하는 경우 데이터 배열이 한 행에 여러 관측값에 대한 정보가 포함되는 방식일 수도 있습니다. JMP에서 데이터를 분석하려면 데이터를 가져온 후 JMP 데이터 테이블의 각 행에 단일 관측값에 대한 정보가 포함되도록 데이터를 재구성해야 합니다. 자세한 내용은 "예: 일원 분석을 위한 데이터 쌓기"에서 확인하십시오.

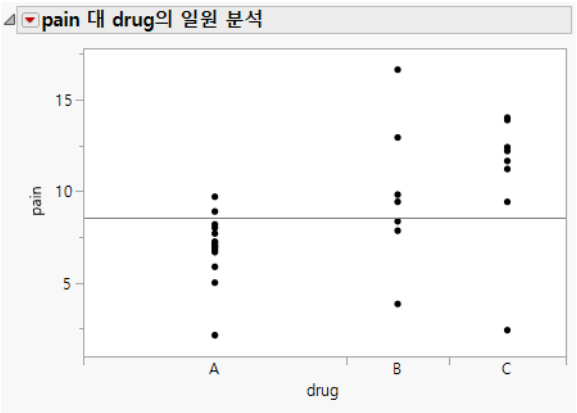
참고 : X 로 Y 적합 시작 창은 연속형, 순서형 및 명목형 데이터 유형의 열을 허용합니다. 연속형 Y, 반응 열과 명목형 또는 순서형 X, 요인 열의 모든 쌍에 대해 일원 분석 플랫폼이 시작됩니다. X 로 Y 적합 시작 창은 다른 열 유형 조합에 대해 이변량, 분할 분석 또는 로지스틱 플랫폼을 실행합니다.

일원 분석 보고서

처음에는 "일원 분석" 보고서에 명목형 또는 순서형 X 변수와 연속형 Y 변수의 각 쌍에 대한 산점도가 포함됩니다. 빨간색 삼각형 옵션을 사용하여 모형 적합, 가설 검정 수행, 동등성 검정 수행, 분포 비교 및 통계 보고서 보기가 가능합니다. 자세한 내용은 "일원 분석 플랫폼 옵션"에서 확인하십시오.

참고 : 시작 창에 블록 변수가 지정된 경우 일원 분석 그림의 Y 변수 값은 블록 변수에 의해 중심화됩니다. 블록이 있는 분석의 경우 사용 가능한 옵션이 다릅니다.

그림 6.5 일원 분석 그림



대화식으로 변수 바꾸기

연결된 데이터 테이블의 "열" 패널에서 변수를 선택하여 축으로 드래그하는 방법으로 변수를 바꿀 수도 있습니다.

일원 분석 플랫폼 옵션

"일원 분석"의 빨간색 삼각형 메뉴에는 검정, 적합 및 표시 옵션이 포함되어 있습니다. 옵션에는 ANOVA(분산 분석), 비모수 검정, 동등성 검정 및 그룹 간 차이를 평가하는 다중 비교 방법이 있습니다. 밀도 및 CDF 그림을 사용하면 X 범주별 반응의 분포를 시각화할 수 있습니다. 옵션은 산점도에 시각화를 추가하고, 통계 보고서를 제공하거나, 추가 분석을 시작합니다.

블록 변수가 사용되면 블록이 있는 일원 ANOVA 모형이 데이터에 적합됩니다. 추가 분석 옵션은 제한됩니다.

참고 : " 그룹 적합 " 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에만 표시됩니다 . " 그룹 적합 " 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 을 기준으로 정렬할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

" 일원 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

분위수 일원 분석 그림에 상자 그림을 표시하거나 숨기고 분위수 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 일원 분석 보고서 " 에서 확인하십시오 .

평균 /ANOVA (X 변수의 수준이 세 개 이상인 경우에만 사용 가능) 일원 분석 그림에 평균 다이아몬드를 표시하거나 숨기고 ANOVA 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 평균 /ANOVA/ 합동 t 보고서 " 에서 확인하십시오 .

평균 /ANOVA/ 합동 t (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 일원 분석 그림에 평균 다이아몬드를 표시하거나 숨기고 ANOVA 보고서를 표시하거나 숨깁니다 . ANOVA 보고서에는 두 그룹의 분산이 같다고 가정하는 합동 *t*- 검정 보고서가 포함됩니다 .

평균 및 표준편차 일원 분석 그림에 평균 선 , 오차 막대 및 표준편차 선을 표시하거나 숨기고 요약 통계량 테이블을 표시하거나 숨깁니다 . 평균의 표준 오차에는 개별 그룹 표준편차가 사용됩니다 . 이러한 그래프 요소에 대한 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 " 에서 확인하십시오 .

t- 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 분산이 같지 않다는 가정하에 *t*- 검정 보고서를 표시하거나 숨깁니다 . 자세한 내용은 "t- 검정 보고서 " 에서 확인하십시오 .

평균 분석 방법 ANOM(평균 분석) 방법을 사용하여 여러 그룹을 비교하기 위한 옵션을 포함합니다 . 자세한 내용은 " 평균 분석 보고서 " 에서 확인하십시오 . ANOM 방법에 대한 자세한 내용은 Nelson et al. 연구 자료 (2005) 에서 확인하십시오 .

참고 : 시작 창에서 블록 변수를 지정하고 블록 및 X 요인 수준의 각 조합에 동일한 개수가 있는 경우 ANOM 차트가 사용 가능한 유일한 " 평균 분석 " 차트입니다 . 시작 창에서 블록 변수를 지정하고 개수가 동일하지 않은 경우 사용 가능한 " 평균 분석 " 옵션이 없습니다 .

ANOM 그룹 평균을 전체 평균과 비교하기 위한 ANOM 의사결정 차트를 표시하거나 숨깁니다 . 이 방법은 데이터가 대략적으로 정규 분포되었다고 가정합니다 . 자세한 내용은 " 평균 분석의 예 " 에서 확인하십시오 .

변환 순위를 사용한 ANOM 그룹 평균의 비모수 비교를 위한 "변환 순위를 사용한 ANOM" 의 사결정 차트를 표시하거나 숨깁니다 . " 변환 순위를 사용한 ANOM" 은 각 그룹 평균의 변환 순위를 전체 평균의 변환 순위와 비교합니다 . ANOM 검정은 변환된 관측값에 일반적인 ANOM 절차와 임계값을 적용합니다 .

팁 : 데이터가 명확하게 비정규 데이터이고 정규 데이터로 변환할 수 없는 경우 이 방법을 사용합니다 .

분산에 대한 ANOM (X 변수의 수준이 세 개 이상이고 각 그룹에 네 개 이상의 관측값이 있는 경우에만 사용 가능) 그룹 표준편차 (또는 분산) 를 제곱근 평균 제곱 오차 (또는 평균 제곱 오차) 와 비교하기 위한 ANOMV 의사결정 차트를 표시하거나 숨깁니다 . 이 방법은 데이터가 대략적으로 정규 분포되었다고 가정합니다 . 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 분석의 예 " 에서 확인하십시오 .

Levene(ADM) 을 사용한 분산에 대한 ANOM 그룹 분산을 비교하기 위해 비정규성에 로버스트한 ANOMV-LEV 의사결정 차트를 표시하거나 숨깁니다 . 이 방법은 ADM(중앙값과의 절대 편차) 의 그룹 평균을 전체 평균 ADM 과 비교합니다 . "Levene(ADM) 을 사용한 분산에 대한 ANOM" 검정에 대한 자세한 내용은 Levene 연구 자료 (1960) 또는 Brown and Forsythe 연구 자료 (1974) 에서 확인하십시오 .

팁 : 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 이상치 및 비정규성에 로버스트합니다 .

범위에 대한 ANOM (균형 설계 및 특정 그룹 크기에만 사용 가능) 그룹 범위를 그룹 범위의 평균과 비교하기 위한 ANOMV-TR 의사결정 차트를 표시하거나 숨깁니다 . 이 검정은 퍼짐에 대한 측도로서 범위를 기준으로 한 척도 차이에 대한 검정입니다 . 자세한 내용은 Wheeler 연구 자료 (2003) 에서 확인하십시오 . 자세한 내용은 " 범위에 대한 ANOM 검정의 제한 사항 " 에서 확인하십시오 .

팁 : 로버스트 ANOMV 방법 중 하나를 선택하려면 ANOMV-TR 이 ANOM-LEV 보다 로버스트하다는 점에 유의하십시오 . 그러나 ANOM-LEV 는 ANOM-TR 보다 더 강력합니다 .

평균 비교 지정된 그룹 평균 집합 간의 다중 비교를 위한 옵션을 포함합니다 . 자세한 내용은 " 평균 비교 보고서 " 에서 확인하십시오 .

개별 쌍 비교 (스튜던트 t) 모든 평균 쌍에 대한 Fisher LSD 다중 비교 보고서를 표시하거나 숨기고 , 비교 원을 일원 분석 그림에 표시하거나 숨깁니다 .

전체 쌍 비교 (Tukey HSD) 모든 평균 쌍에 대한 Tukey-Kramer 다중 비교 보고서를 표시하거나 숨기고 , 비교 원을 일원 분석 그림에 표시하거나 숨깁니다 .

최량 비교 (Hsu MCHB) 최량과의 Hsu 다중 비교 보고서를 표시하거나 숨기고 , 비교 원을 일원 분석 그림에 표시하거나 숨깁니다 .

대조군 비교 (Dunnett) 대조군과의 Dunnett 다중 비교 보고서를 표시하거나 숨기고 , 비교 원을 일원 분석 그림에 표시하거나 숨깁니다 .

팁 : " 대조 수준 " 열 특성을 요인 열에 추가하여 " 대조군 비교 (Dunnett)" 를 선택할 때마다 대조군을 지정하지 않도록 할 수 있습니다 . 자세한 내용은 JMP 사용의 에서 확인하십시오 .

단계별 개별 쌍 비교 (Newman-Keuls) Newman-Keuls 단계별 쌍체 비교 보고서를 표시하거나 숨깁니다 . Newman-Keuls 검정은 모임별 오차율을 제어하지 않습니다 .

비모수 그룹 위치의 비모수 비교를 위한 옵션을 포함합니다 . 자세한 내용은 " 비모수 검정 보고서 " 에서 확인하십시오 .

Wilcoxon/Kruskal-Wallis 검정 Wilcoxon 순위 스코어에 기반한 하나 이상의 검정을 표시하거나 숨깁니다 . Wilcoxon 순위 스코어는 데이터의 단순 순위입니다 . Wilcoxon 검정은 로지스틱 분포를 따르는 오차에 가장 강력한 순위 검정입니다 .

X 변수의 수준이 정확히 두 개인 경우 Wilcoxon 검정은 Mann-Whitney 검정과 동일합니다 . 이 경우 보고서에는 0.5 연속성 수정을 사용하는 Wilcoxon 2 표본 정규 근사 검정 테이블과 연속성 수정을 사용하지 않는 Kruskal-Wallis 카이제곱 근사 검정 테이블이 포함됩니다 .

X 변수의 수준이 세 개 이상인 경우 단순 순위에 기반한 검정을 Kruskal-Wallis 검정이라고 합니다 .

보고서에 대한 자세한 내용은 "Wilcoxon Kruskal-Wallis, 중앙값 , Friedman 순위 및 Van der Waerden 검정 보고서 " 에서 확인하십시오 . 예는 " 비모수 Wilcoxon 검정의 예 " 에서 확인하십시오 .

중앙값 검정 중앙값 순위 스코어에 기반한 검정을 표시하거나 숨깁니다 . 중앙값 순위 스코어는 순위가 중앙값 순위보다 높은지 아니면 낮은지에 따라 0 또는 1 입니다 . 중앙값 검정은 이중 지수 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 보고서에 대한 자세한 내용은 "Wilcoxon Kruskal-Wallis, 중앙값 , Friedman 순위 및 Van der Waerden 검정 보고서 " 에서 확인하십시오 .

Van der Waerden 검정 Van der Waerden 순위 스코어에 기반한 검정을 표시하거나 숨깁니다 . Van der Waerden 순위 스코어는 데이터 순위를 1 + 스코어 값으로 나눈 값입니다 . 스코어 값은 정규 분포 함수의 역을 적용하여 정규 스코어로 변환된 관측값의 수입니다 . Van der Waerden 검정은 정규 분포를 따르는 오차에 가장 강력한 순위 검정입니다 . 보고서에 대한 자세한 내용은 "Wilcoxon Kruskal-Wallis, 중앙값 , Friedman 순위 및 Van der Waerden 검정 보고서 " 에서 확인하십시오 .

Kolmogorov Smirnov 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 경험적 분포 함수를 기반으로 반응 분포가 그룹 간에 동일한지 여부를 판별하는 검정을 표시하거나 숨깁니다 . 보고서에 대한 자세한 내용은 "Kolmogorov-Smirnov 2 표본 검정 보고서 " 에서 확인하십시오 .

Friedman 순위 검정 (시작 창에서 각 블록 내의 관측값 수가 동일한 블록 변수를 지정한 경우에만 사용 가능) Friedman 순위 스코어에 기반한 검정을 표시하거나 숨깁니다 . Friedman 순위 스코어는 블록 변수의 각 수준 내에 있는 데이터의 순위입니다. 이 검정의 모수 버전은 반복 측정 ANOVA 입니다 . 보고서에 대한 자세한 내용은 "Wilcoxon Kruskal-Wallis, 중앙값 , Friedman 순위 및 Van der Waerden 검정 보고서 " 에서 확인하십시오 .

참고 :

- Wilcoxon, 중앙값 , Van der Waerden 및 Friedman 순위 검정의 경우 X 변수의 수준이 세 개 이상이면 일원 검정에 대한 카이제곱 근사가 수행됩니다 .
- 시작 창에서 블록 변수를 지정하고 블록 및 X 요인 수준의 각 조합에 동일한 개수가 있는 경우 Friedman 순위 검정이 사용 가능한 유일한 " 비모수 " 옵션입니다 . 시작 창에서 블록 변수를 지정하고 개수가 동일하지 않은 경우 사용 가능한 " 비모수 " 옵션이 없습니다 .

정확 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 정확 검정을 수행하기 위한 옵션을 포함합니다 . 자세한 내용은 " 정확 검정 보고서 " 에서 확인하십시오 .

Wilcoxon 정확 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 정확 방법을 사용한 Wilcoxon 스코어 검정을 표시하거나 숨깁니다 . 자세한 내용은 " 비모수 Wilcoxon 검정의 예 " 에서 확인하십시오 .

중앙값 정확 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 정확 방법을 사용한 중앙값 스코어 검정을 표시하거나 숨깁니다 .

Van Der Waerden 정확 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 정확 방법을 사용한 Van der Waerden 정규 스코어 검정을 표시하거나 숨깁니다 .

Kolmogorov Smirnov 정확 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 경험적 분포 함수를 기반으로 반응 분포가 그룹 간에 동일한지 여부를 판별하는 검정을 표시하거나 숨깁니다 .

주의 : 정확 검정은 큰 표본 크기를 계산하는 데 시간이 오래 걸릴 수 있습니다 .

비모수 다중 비교 그룹 위치의 비모수 다중 비교를 위한 옵션을 포함합니다 .

Wilcoxon 개별 쌍 비교 각 쌍에 대한 Wilcoxon 검정을 표시하거나 숨깁니다 . 이 절차에서는 전체 유의 수준을 제어하지 않습니다 . 이 값은 " 평균 비교 " 메뉴에 있는 개별 쌍 비교 (스튜던트 t) 옵션의 비모수 버전입니다 . 자세한 내용은 "Wilcoxon 개별 쌍 비교 , Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교 보고서 " 에서 확인하십시오 .

Steel-Dwass 전체 쌍 비교 각 쌍에 대한 Steel-Dwass 검정을 표시하거나 숨깁니다 . 이 검정은 " 평균 비교 " 메뉴에 있는 전체 쌍 비교 (Tukey HSD) 옵션의 비모수 버전입니다 . 자세한 내용은 "Wilcoxon 개별 쌍 비교 , Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교 보고서 " 에서 확인하십시오 .

Steel 대조군 비교 그룹화 변수의 각 수준을 대조 수준과 비교하기 위한 Steel 검정을 표시하거나 숨깁니다. 이 검정은 " 평균 비교 " 메뉴에 있는 대조군 비교 (Dunnett) 옵션의 비모수 버전입니다. 자세한 내용은 "Wilcoxon 개별 쌍 비교, Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교 보고서 " 에서 확인하십시오.

Dunnett 전체 쌍 비교 (결합 순위 기준) Dunn 방법을 사용한 각 쌍의 비교를 표시하거나 숨깁니다 . 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) 보고서 " 에서 확인하십시오.

Dunnett 대조군 비교 (결합 순위 기준) Dunn 방법을 사용한 그룹화 변수의 각 수준과 대조 수준의 비교를 표시하거나 숨깁니다 . 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. " 대조 수준 " 열 특성을 요인 열에 추가하여 "Steel 대조군 비교 " 또는 "Dunnett 대조군 비교 (결합 순위 기준)" 를 선택할 때마다 대조군을 지정하지 않도록 할 수 있습니다. 자세한 내용은 JMP 사용의 에서 확인하십시오. 자세한 내용은 "Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준) 보고서 " 에서 확인하십시오.

참고 : 비모수 검정 선택은 데이터에 따라 달라집니다. 전체 쌍에 대한 가설에 사용할 검정을 선택하는 방법에 대한 지침은 Boos and Duan(2021) 연구 자료에서 확인하십시오.

이분산 그룹 분산 동일성에 대한 네 개의 검정과 표준편차가 같지 않은 동일 평균에 대한 Welch ANOVA 검정을 표시하거나 숨깁니다. 그룹 분산 동일성 검정은 O-Brien, Brown-Forsythe, Levene 및 Bartlett 입니다. 자세한 내용은 " 이분산 보고서 " 에서 확인하십시오.

동등성 검정 동등성, 우월성 또는 비열등성 검정을 위한 방법을 포함합니다. 자세한 내용은 " 동등성 검정 보고서 " 에서 확인하십시오. 이러한 검정은 실제적 또는 임상적으로 유의한 유사성을 감지하려는 경우에 유용합니다.

평균 평균에 대한 동등성, 우월성 또는 비열등성 검정 옵션이 있는 창을 시작합니다. 분산 가정 및 임계 차이를 지정합니다.

표준편차 표준편차에 대한 동등성, 우월성 또는 비열등성 검정 옵션이 있는 창을 시작합니다. 임계 비율을 지정합니다.

로버스트 로버스트 평균 추정값에 대한 옵션을 포함합니다. 로버스트 방법을 사용하면 이상치가 분석에 미치는 영향을 줄일 수 있습니다. 로버스트 추정값은 두꺼운 꼬리 분포에 대해 최소 제곱 추정값보다 더 효율적입니다. 자세한 내용은 " 로버스트 적합 보고서 " 에서 확인하십시오.

로버스트 적합 일원 분석 그림에 로버스트 평균 선을 표시하거나 숨기고, 그룹 평균의 Huber 추정값을 포함하는 " 로버스트 적합 " 보고서를 표시하거나 숨깁니다.

Cauchy 적합 일원 분석 그림에 로버스트 평균 선을 표시하거나 숨기고, Cauchy 오차 분포를 가정하여 그룹 평균의 추정값을 포함하는 "Cauchy 적합 " 보고서를 표시하거나 숨깁니다. 이 로버스트 방법은 극단 이상치에 적절합니다.

검정력 ANOVA 분석을 위한 검정력 계산을 실행할 수 있습니다. 자세한 내용은 " 검정력 보고서 " 에서 확인하십시오. 검정력 계산 및 예에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

α 수준 설정 α 수준을 지정할 수 있습니다. 일반 유의 수준 목록에서 선택하거나, "기타"를 선택하여 수준을 지정할 수 있습니다.

참고 : 유의 수준은 분석과 보고서 전체에 적용됩니다. 여기에는 신뢰 한계, 평균 다이아몬드, 비교 원 및 다중 비교 분석이 포함됩니다. 동등성 검정을 위한 유의 수준은 별도로 설정됩니다.

정규 분위수 그림 각 그룹의 데이터 분위수를 표시하기 위한 옵션을 포함합니다. 자세한 내용은 "정규 분위수 그림의 예"에서 확인하십시오.

분위수별 실제값 그림 일원 분석 그림 오른쪽에 분위수 그림을 표시하거나 숨깁니다. 분위수 그림은 Y 변수에 대한 일원 분석 세로 축을 공유합니다. 각 그룹의 누적 확률은 가로 축에 있습니다. 이 그림에는 X 변수의 각 수준 내에서 계산된 분위수가 표시됩니다.

실제값별 분위수 그림 가로 축에 Y 변수가 있고 세로 축에 누적 확률이 있는 분위수 그림을 표시하거나 숨깁니다. 이 그림에는 범주형 X 변수의 각 수준 내에서 계산된 분위수가 표시됩니다.

적합선 (분위수 그림이 열려 있는 경우에만 사용 가능) 열려 있는 각 분위수 그림에서 X 변수의 각 수준에 대한 데이터에 적합된 참조선을 표시하거나 숨깁니다.

정규 분위수 라벨 (분위수 그림이 열려 있는 경우에만 사용 가능) 열려 있는 각 분위수 그림에 정규 분위수 척도를 표시하거나 숨깁니다.

CDF 그림 일원 분석 보고서에서 모든 그룹에 대한 누적 분포 함수를 표시하거나 숨깁니다. 자세한 내용은 "CDF 그림의 예 " 에서 확인하십시오.

밀도 그룹 간 밀도를 시각화하기 위한 옵션을 포함합니다. 자세한 내용은 "밀도 옵션의 예"에서 확인하십시오.

밀도 비교 각 그룹에 대한 확률 밀도 함수가 중첩된 그림을 표시하거나 숨깁니다.

밀도의 합성 각 그룹의 개수에 따라 가중치가 부여된 합산 밀도 그림을 표시하거나 숨깁니다. " 밀도의 합성 " 그림은 X 변수의 범위에서 각 그룹이 총 밀도에 어떻게 기여하는지 보여 줍니다.

밀도의 비율 X 변수의 각 수준별로 밀도에 대한 기여도를 보여 주는 그림을 표시하거나 숨깁니다. 기여도는 X 변수의 전체 범위에서 총 밀도의 비율로 표시됩니다.

매칭 열 지정된 매칭 변수를 기반으로 일원 분석 그림에서 매칭 적합선 및 해당 적합선을 표시하거나 숨깁니다. 서로 다른 그룹의 관측값을 동일한 참가자로부터 얻은 경우처럼 분석의 데

이터를 매칭 (쌍체) 데이터로부터 얻은 경우에 이 옵션을 사용합니다 . 자세한 내용은 " 매칭 열 보고서 " 에서 확인하십시오 .

저장 다음과 같은 통계량을 현재 데이터 테이블에 새 열로 저장합니다 .

잔차 저장 Y 변수에서 X 변수의 각 수준에 대한 Y 변수의 평균을 뺀 값을 저장합니다 .

표준화한 데이터 저장 X 변수의 각 수준에 대한 Y 변수의 표준화된 값을 저장합니다 . 표준화된 값은 중심화된 반응을 각 수준 내의 표준편차로 나눈 값입니다 .

정규 분위수 저장 X 변수의 각 수준 내에서 계산된 정규 분위수 값을 저장합니다 .

예측값 저장 X 변수의 각 수준에 대한 Y 변수의 예측 평균을 저장합니다 .

표시 옵션 그림의 요소를 추가하거나 제거합니다 . 관련되지 않은 일부 옵션은 표시되지 않을 수 있습니다 .

모든 그래프 일원 분석 그림을 표시하거나 숨깁니다 .

점 일원 분석 그림에 데이터 점을 표시하거나 숨깁니다 .

상자 그림 각 그룹에 대한 이상치 상자 그림을 표시하거나 숨깁니다 . 자세한 내용은 " 이상치 상자 그림 " 에서 확인하십시오 . 예는 " 일원 분석 수행 " 에서 확인하십시오 .

평균 다이아몬드 일원 분석 그림에 평균 다이아몬드를 표시하거나 숨깁니다 . 각 평균 다이아몬드는 해당 그룹 평균의 95% 신뢰 구간에 걸쳐 있으며 평균 위치에 가로선이 있습니다 . 95% 신뢰 구간은 합동 표준편차를 사용하여 계산됩니다 . 자세한 내용은 " 평균 다이아몬드 및 X 축 비례 " 에서 확인하십시오 .

평균 선 각 그룹의 평균에 선을 표시하거나 숨깁니다 . 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 " 에서 확인하십시오 .

평균 CI 선 각 그룹의 95% 신뢰 상한 및 하한 위치에 선을 표시하거나 숨깁니다 . 95% 신뢰 수준은 합동 표준편차를 사용하여 계산됩니다 .

평균 오차 막대 평균으로부터 1 표준 오차 위와 아래에 해당하는 오차 막대와 함께 각 그룹의 평균을 표시하거나 숨깁니다 . 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 " 에서 확인하십시오 .

총 평균 Y 변수의 전체 평균을 표시하거나 숨깁니다 .

표준편차 선 각 그룹 평균의 1 표준편차 위와 아래에 선을 표시하거나 숨깁니다 . 자세한 내용은 " 평균 선 , 오차 막대 및 표준편차 선 " 에서 확인하십시오 .

비교 원 (다중 비교 보고서가 열려 있는 경우에만 사용 가능) 비교 원을 표시하거나 숨깁니다 . 자세한 내용은 " 비교 원에 대한 통계 상세 정보 " 에서 확인하십시오 . 예는 " 일원 분석 수행 " 에서 확인하십시오 .

여러 평균 연결 그룹 평균을 연결하는 직선을 표시하거나 숨깁니다 .

평균의 평균 그룹 평균의 평균을 표시하거나 숨깁니다 .

X 축 비례 (" 매칭 열 " 옵션을 선택한 경우에는 사용 불가능) 가로 축에 간격을 지정합니다 . 이 옵션을 선택하면 간격이 각 수준의 관측값 수에 비례합니다 . 자세한 내용은 " 평균 다이아몬드 및 X 축 비례 " 에서 확인하십시오 .

퍼진 점 데이터 점의 퍼짐을 지정합니다 . 이 옵션을 선택하면 데이터 점이 구간 너비 전체에 퍼집니다 .

지터링된 점 데이터 점의 퍼짐을 지정합니다 . 이 옵션을 선택하면 표식이 중첩되지 않도록 데이터 점이 지터링됩니다 . 특히 이 옵션은 그래프 빌더의 지터에 대한 " 중심화 격자 " 옵션과 동일합니다 . 자세한 내용은 Essential Graphing 의 에서 확인하십시오 .

매칭 선 (" 매칭 열 " 옵션을 선택한 경우에만 사용 가능) 매칭 변수의 각 수준 평균을 연결하는 선을 표시하거나 숨깁니다 .

매칭 점선 (" 매칭 열 " 옵션을 선택하고 매칭 변수의 값이 X 변수의 수준에 대해 모두 결측인 경우에만 사용 가능) 매칭 변수의 결측 수준을 통과하여 평균을 연결하는 점선을 표시하거나 숨깁니다 . 결측 셀 평균 대신 사용되는 값은 이원 ANOVA 모형을 사용하여 구합니다 .

히스토그램 원래 그림 오른쪽에 가로 배열 히스토그램을 표시하거나 숨깁니다 .

로버스트 평균 선 (" 로버스트 " 옵션을 선택한 경우에만 사용 가능) 각 그룹의 로버스트 평균에 선을 표시하거나 숨깁니다 .

범례 정규 분위수, CDF(누적 분포 함수) 및 밀도 그림에 대한 범례를 표시하거나 숨깁니다.

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 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 .

플랫폼 환경 설정 현재 플랫폼 환경 설정을 보거나 , 현재 JMP 보고서의 설정과 일치하도록 플랫폼 환경 설정을 업데이트할 수 있는 옵션이 포함되어 있습니다 .

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 .

일원 분석 보고서

일원 분석 플랫폼을 사용하면 범주형 X 변수와 연속형 Y 변수 사이의 관계를 탐색하기 위해 시각화를 생성하고, 가설 검정을 실행하고, 모형을 적합시킬 수 있습니다. 이 섹션에는 특정 분석 옵션에 대해 생성되는 보고서 관련 정보가 포함되어 있습니다.

- " 분위수 보고서 "
- " 평균 /ANOVA/ 합동 t 보고서 "
- " 평균 및 표준편차 보고서 "
- "t- 검정 보고서 "
- " 평균 분석 보고서 "
- " 평균 비교 보고서 "
- " 비모수 검정 보고서 "
- " 비모수 다중 비교 보고서 "
- " 이분산 보고서 "
- " 동등성 검정 보고서 "
- " 로버스트 적합 보고서 "
- " 검정력 보고서 "
- " 매칭 열 보고서 "

분위수 보고서

일원 분석 플랫폼의 "분위수" 보고서에는 각 그룹의 최소값과 최대값이 나열됩니다. 또한 10번째, 25번째, 50번째(중앙값), 75번째 및 90번째 백분위수도 나열됩니다. 중앙값은 50번째 백분위수이고, 25번째 및 75번째 백분위수는 **사분위수**라고도 합니다.

평균 /ANOVA/ 합동 t 보고서

일원 분석 플랫폼에서 " 평균 /ANOVA" 옵션을 사용하여 분산 분석을 수행합니다 . X 변수의 수준이 두 개뿐인 경우 이 옵션은 **평균 /ANOVA/ 합동 t** 로 표시됩니다 . 보고서에는 각 그룹의 적합 요약 , ANOVA(분산 분석) 및 요약 통계량에 대한 테이블이 포함되어 있습니다 . X 변수의 수준이 두 개뿐인 경우 " 합동 *t*- 검정 " 테이블이 보고서에 포함됩니다 . 시작 창에서 블록 변수를 지정하고 블록 및 X 변수 수준의 각 조합에 동일한 개수가 있는 경우 " 블록 평균 " 테이블이 보고서에 포함됩니다 . 다른 블록 구성을 사용하면 " 일원 ANOVA - 블록화 " 보고서가 생성됩니다 .

적합 요약

일원 분석 플랫폼의 " 적합 요약 " 테이블에는 다음 통계량이 포함됩니다 .

R^2 모형에 의해 설명된 변동의 비율입니다 . 나머지 변동은 랜덤 오차에 기인합니다 . 모형이 완벽하게 적합되는 경우에는 R^2 가 1 입니다 . 자세한 내용은 " 적합 요약 보고서에 대한 통계 상세 정보 " 에서 확인하십시오 . R^2 를 **결정 계수**라고도 합니다 .

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

Adj-R^2 모형의 모수 수에 맞게 조정된 R^2 통계량입니다. Adj-R^2 을 사용하면 포함된 모수의 수가 다른 모형을 쉽게 비교할 수 있습니다. 자세한 내용은 " 적합 요약 보고서에 대한 통계 상세 정보 " 에서 확인하십시오.

제곱근 평균 제곱 오차 랜덤 오차의 표준편차 추정값입니다. 이 값은 " 분산 분석 " 보고서에 표시된 **오차** 평균 제곱의 제곱근입니다.

반응 평균 Y 변수의 전체 평균 (산술평균) 입니다.

관측값 수 (또는 가중치 합) 적합 추정에 사용된 관측값의 개수입니다. 가중치가 사용된 경우에는 가중치의 합입니다. 자세한 내용은 " 적합 요약 보고서에 대한 통계 상세 정보 " 에서 확인하십시오.

합동 t- 검정

일원 분석 플랫폼의 "합동 t-검정" 테이블에는 그룹 간 등분산을 가정하고 두 그룹 평균을 비교하는 *t*-검정의 결과가 요약되어 있습니다. 이 테이블은 X 변수의 수준이 정확히 두 개인 경우에만 사용할 수 있습니다. 자세한 내용은 "t-검정 보고서"에서 확인하십시오.

분산 분석

일원 분석 플랫폼의 "분산 분석" 테이블에는 ANOVA 분석 결과가 요약되어 있습니다. ANOVA는 총 변동을 성분으로 분할합니다.

참고 : 블록 열을 지정한 경우에는 " 분산 분석 " 보고서에 **블록** 변수가 포함됩니다.

소스 변동 소스입니다. 모형 소스, **오차** 및 **수정 합계**가 이러한 소스에 해당합니다.

DF 각 변동 소스의 자유도 (줄여서 DF 라고 함) 입니다.

- **수정 합계**의 자유도는 $N - 1$ 이며, 여기서 N 은 분석에 사용된 총 관측값 수입니다.
- 모형의 자유도는 $k - 1$ 이며, 여기서 k 는 X 변수의 수준 수입니다.

오차 자유도는 **수정 합계** 자유도와 **모형** 자유도의 차이 (즉, $N - k$) 입니다.

제곱합 각 변동 소스의 제곱합 (줄여서 SS 라고 함) 입니다.

- 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

일원 분석 플랫폼의 " 일원 ANOVA 에 대한 평균 " 테이블에는 명목형 또는 순서형 요인의 각 수준에 대한 반응 정보가 요약되어 있습니다.

수준 X 변수의 수준입니다.

개수 각 그룹의 관측값 수입니다.

평균 각 그룹의 평균입니다.

표준 오차 그룹 평균에 대한 표준편차 추정값입니다. 이 표준 오차는 각 수준에서 반응의 분산이 동일하다는 가정하에 추정됩니다. 이 값은 " 적합 요약 " 보고서에 나오는 제곱근 평균 제곱 오차를 그룹 평균 계산에 사용된 값 개수의 제곱근으로 나눈 값입니다.

95% 하한 및 95% 상한 오차 분산의 합동 추정값과 합동 자유도를 기반으로 한 그룹 평균의 95% 신뢰 구간 하한 및 상한입니다.

블록 평균

일원 분석 플랫폼의 "블록 평균" 보고서는 시작 창에서 블록 변수를 지정하고 블록 및 X 변수 수준의 각 조합에 동일한 개수가 있는 경우에만 나타납니다.

평균 및 표준편차 보고서

일원 분석 플랫폼의 " 평균 및 표준편차 " 보고서에는 X 변수의 각 수준에 대한 요약 통계량이 포함됩니다. X 변수의 수준은 그룹 또는 처리를 정의합니다.

수준 X 변수의 수준입니다.

개수 각 그룹의 관측값 수입니다.

평균 각 그룹의 평균입니다.

표준편차 각 그룹의 표준편차입니다.

평균 표준 오차 그룹 평균에 대한 표준 오차 추정값입니다. 이 값은 표준편차를 그룹 평균을 계산하는 데 사용된 값 수의 제곱근으로 나눈 값입니다.

95% 하한 및 95% 상한 개별 그룹 평균, 표준편차 및 표본 크기를 기반으로 한 그룹 평균의 95% 신뢰 구간 하한 및 상한입니다.

참고 : 이러한 신뢰 구간은 일반적으로 ANOVA 신뢰 구간보다 좁습니다.

t- 검정 보고서

일원 분석 플랫폼의 "t- 검정 " 보고서에는 그룹 간 등분산 또는 이분산을 가정하고 두 그룹 평균을 비교하는 *t*- 검정의 결과가 요약되어 있습니다 . 이 보고서는 X 변수의 수준이 정확히 두 개인 경우에만 사용할 수 있습니다 .

"*t*- 검정 " 또는 " 합동 t- 검정 " 테이블에는 사용된 차이 및 분산 가정이 나열됩니다 . 다음 통계량도 보고서에 포함됩니다 .

차이 두 X 수준 간의 추정 차이입니다 .

팁 : 차이 순서를 변경하려면 " 값 순서 " 열 특성을 사용하십시오 .

차이 표준 오차 차이의 표준 오차입니다 .

차이 CL 상한 차이의 신뢰 상한입니다 .

차이 CL 하한 차이의 신뢰 하한입니다 .

신뢰도 신뢰 수준 (1-α) 입니다 . 신뢰 수준을 변경하려면 플랫폼의 빨간색 삼각형 메뉴에 있는 α **수준 설정** 옵션에서 새 유의 수준을 선택합니다 .

t 비 *t* 통계량입니다 .

DF *t*- 검정에 사용된 자유도입니다 .

Prob > |t| 양쪽 꼬리 검정과 관련된 *p* 값입니다 .

Prob > t 단측 상한 꼬리 검정과 관련된 *p* 값입니다 .

Prob < t 단측 하한 꼬리 검정과 관련된 *p* 값입니다 .

t- 검정 그림 귀무가설이 참이라는 가정하에 평균의 차이에 대한 표본 분포를 보여 줍니다 . 빨간색 세로 선은 관측된 평균 차이입니다 . 음영 영역은 *p* 값에 해당합니다 .

평균 분석 보고서

일원 분석 플랫폼의 ANOM(평균 분석) 은 평균 또는 분산에 대한 다중 비교 방법입니다 . 각 ANOM 보고서에는 다음 요소가 있는 의사결정 차트가 포함됩니다 .

- UDL(결정 상한)
- LDL(결정 하한)
- 가로 중심선 . 중심선 위치는 차트 유형에 의해 결정됩니다 .
  - ANOM: 전체 평균
  - 변환 순위를 사용한 ANOM: 변환 순위의 전체 평균
  - 분산에 대한 ANOM: 제곱근 평균 제곱 오차 (분산 척도의 경우에는 평균 제곱 오차)
  - Levene(ADM) 을 사용한 분산에 대한 ANOM: 평균과의 전체 절대 편차
  - 범위에 대한 ANOM: 그룹 범위의 평균

그림에 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 차트 옵션

일원 분석 플랫폼에서 각 "평균 분석"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

유의 수준 설정 목록에서 유의 수준을 선택하거나, **기타**를 선택하여 수준을 지정할 수 있습니다. 유의 수준이 변경되면 결정 상한 및 결정 하한이 업데이트됩니다.

참고 : " 범위에 대한 ANOM" 의 경우 0.10, 0.05 및 0.01 유의 수준만 사용할 수 있습니다.

요약 보고서 표시 평균 분석 방법을 기반으로 하는 요약 보고서를 표시하거나 숨깁니다.

- 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) 의사결정 차트의 세로 축 척도를 표준편차 단위 또는 분산 사이에서 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.

범위에 대한 ANOM 검정의 제한 사항

다른 ANOM 결정 한계와 달리 "범위에 대한 ANOM" 차트의 결정 한계에는 제공된 임계값이 사용됩니다. 따라서 " 범위에 대한 ANOM" 은 다음 조건에서만 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 비교 보고서

일원 분석 플랫폼에는 평균을 비교하는 여러 옵션이 있습니다. 이 섹션에서는 평균 비교 보고서의 옵션과 각 비교 방법에 대한 상세 정보를 다룹니다.

평균 비교 보고서 옵션

일원 분석 플랫폼에서 " 평균 비교 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

차이 행렬 그룹 평균 간의 전체 쌍별 차이 행렬을 표시하거나 숨깁니다.

신뢰 분위수 평균 비교 절차에 사용되는 임계값 및 유의 수준 (α) 을 표시하거나 숨깁니다.

팁 : " 일원 분석 " 의 빨간색 삼각형 메뉴에 있는 "α 수준 설정 " 옵션을 사용하여 유의 수준을 변경합니다.

LSD 임계 행렬 (" 단계별 개별 쌍 비교 " 옵션에는 사용 불가능) 평균의 쌍별 차이에서 이러한 평균에 대한 최소 유의차를 뺀 행렬을 표시하거나 숨깁니다. 양수 값은 유의차가 있는 평균의 쌍을 나타냅니다. Hsu MCB 검정의 경우 두 개의 LSD 행렬이 있습니다. 하나는 최소값과 비교하고 다른 하나는 최대값과 비교합니다.

연결 문자 보고서 (" 개별 쌍 비교 ", " 전체 쌍 비교 " 및 " 단계별 개별 쌍 비교 " 옵션에만 사용 가능) 기존의 문자로 코딩된 보고서를 표시하거나 숨깁니다. 이 보고서에서 문자를 공유하지 않는 평균에는 유의차가 있습니다.

정렬된 차이 보고서 (" 개별 쌍 비교 " 및 " 전체 쌍 비교 " 옵션에만 사용 가능) 전체 쌍별 양수 차이, 차이의 표준 오차, 신뢰 구간, *p* 값 및 신뢰 구간이 있는 차이 크기 그림을 표시하거나 숨깁니다. *p* 값은 평균이 같다는 가설에 해당합니다.

상세 비교 보고서 (" 개별 쌍 비교 " 옵션에만 사용 가능) 각 비교에 대한 상세 보고서를 표시하거나 숨깁니다. 각 비교에서는 수준 간 차이, 차이의 표준 오차, 신뢰 구간, *t* 비, *p* 값 및 자유도를 보여 줍니다. 각 보고서의 오른쪽에는 비교를 보여 주는 그림이 표시됩니다.

참고 : 차이의 표준 오차는 각 쌍의 표본 크기와 MSE 에 기반한 합동 표준 오차입니다.

개별 쌍 비교 (스튜던트 t)

일원 분석 플랫폼에서 " 개별 쌍 비교 (스튜던트 t)" 옵션을 사용하여 각 그룹 수준 쌍 간의 차이를 검정하는 스튜던트 *t*- 검정에 기반한 Fisher LSD(최소 유의차) 검정을 표시합니다. 이 검정의 예는 " 개별 쌍 비교 (스튜던트 t) 검정의 예 "에서 확인하십시오.

전체 쌍 비교 (Tukey HSD)

일원 분석 플랫폼에서 " 전체 쌍 비교 (Tukey HSD)" 옵션을 사용하여 전체 차이 쌍을 검정합니다. 이 방법은 모든 쌍 조합의 검정에 대한 전체 유의 수준을 보호합니다. HSD 구간은 스튜던트 *t* 쌍별 LSD 구간보다 더 넓습니다. 전체 쌍 비교의 비교 원과 비교하면 Tukey HSD 비교의 원이 더 크고 평균 쌍 간의 차이가 덜 유의합니다.

"신뢰 분위수" 테이블의 q^* 통계량은 $q^* = (1/\text{sqrt}(2)) * q$로 계산됩니다. 여기서 q는 스튜던트화 범위 분포의 알파 백분위수입니다. 이 검정의 예는 "전체 쌍 비교(Tukey HSD) 검정의 예"에서 확인하십시오.

"전체 쌍 비교(Tukey HSD)" 옵션은 Tukey 또는 Tukey-Kramer HSD(Honestly Significant Difference) 검정을 수행합니다(Tukey 1953, Kramer 1956). 이 검정은 표본 크기가 동일한 경우에는 정확 유의 수준 검정이고, 표본 크기가 서로 다른 경우에는 보수적 검정입니다(Hayter 1984).

최량 비교 (Hsu MCB)

일원 분석 플랫폼에서 "최량 비교(Hsu MCB)" 옵션을 사용하여 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"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"에서 확인하십시오.

"최량 비교(Hsu MCB)" 옵션은 Hsu MCB 검정을 수행합니다(Hsu 1996, Hsu 1981).

참고 : MCB 에 의해 최대값 또는 최소값으로 간주되지 않는 평균은 잠재적 최대 또는 최소 평균에 대한 Gupta(1965) 의 선택된 부분집합에도 포함되지 않습니다 .

최대값 및 최소값을 사용한 비교

Hsu MCB 보고서에는 사용 가능한 모든 평균 비교 방법에 공통된 테이블 외에 " 최대값 및 최소값을 사용한 비교 " 테이블도 있습니다 . 이 테이블에는 다음 값이 포함됩니다 .

수준 범주형 변수의 수준입니다 .

최대 p 값 사용 특정 수준의 평균이 나머지 수준의 최대 평균을 초과한다는 검정에 대한 p 값입니다 . 평균이 최대 실제 평균 (알 수 없음) 보다 작거나 같은 수준을 선별해서 제외하려면 이 열의 검정을 사용합니다 .

최소 p 값 사용 특정 수준의 평균이 나머지 수준의 최소 평균보다 작다는 검정에 대한 p 값입니다 . 평균이 최소 실제 평균 (알 수 없음) 보다 크거나 같은 수준을 선별해서 제외하려면 이 열의 검정을 사용합니다 .

대조군 비교 (Dunnett)

일원 분석 플랫폼에서 "대조군 비교(Dunnett)" 옵션을 사용하여 그룹 평균을 대조군과 비교합니다. Dunnett LSD 는 중간 비교 개수에서 크기가 조정되므로 스튜던트 t LSD 와 Tukey-Kramer LSD 사이에 있습니다. 이 검정의 예는 "대조군 비교(Dunnett) 의 예"에서 확인하십시오. Dunnett 검정에 대한 자세한 내용은 Dunnett 연구 자료(1955)에서 확인하십시오.

단계별 개별 쌍 비교 (Newman-Keuls)

일원 분석 플랫폼에서 " 단계별 개별 쌍 비교 (Newman-Keuls)" 옵션을 사용하여 반복된 단계별 절차를 통해 그룹 평균을 비교합니다 . 각 반복에서는 두 그룹 평균 사이의 차이를 검정하기 위해 Tukey HSD 검정이 사용됩니다 . 이 검정의 예는 " 단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예 " 에서 확인하십시오 .

" 단계별 개별 쌍 비교(Newman-Keuls)" 옵션은 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.

주의 : Newman-Keuls 검정은 모임별 오차율을 제어하지 않습니다 . 이 절차의 결과를 해석할 때 주의하십시오 .

J 그룹 평균 검정에 다음 절차가 사용됩니다.

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 = 절차 중의 모든 비교에서 *j* 의 최소값

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 1 만큼 늘리고 $k = \max(i, j) + 1$, $j = J$ 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 1 만큼 늘리고 $k = j + 1$, $j = J$ 로 설정한 다음 2 단계로 진행합니다 .
2. *k* 값에 따라 절차를 계속할지 중지할지 결정합니다 .
   - *k* 가 i 보다 크고 k 가 J 보다 작거나 같은 경우 1 단계를 반복합니다 .
   - *k* 가 *i* 보다 작거나 같거나 *k* 가 J 보다 크면 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. "ewman-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.

Newman-Keuls 검정에 대한 자세한 내용은 Howell 연구 자료(2013)에서 확인하십시오.

참고 : 단계별 개별 쌍 비교 (Newman-Keuls) 검정을 사용할 때는 " 비교 원 " 그래프에 평균 원이 추가되지 않습니다 . 절차의 각 단계에서 비교 원의 크기가 다르기 때문입니다 .

비모수 검정 보고서

일원 분석 플랫폼에는 평균을 비교하는 비모수 옵션이 있습니다 . 이 섹션에서는 이러한 방법에 대한 보고서를 다룹니다 .

방법에는 그룹 간에 동일한 평균 또는 중앙값 가설에 대한 검정이 포함됩니다. 비모수 검정에는 순위 스코어라고 하는 반응 순위의 함수가 사용됩니다. 자세한 내용은 Hajek 연구 자료(1969) 및 SAS Institute Inc. (2020a)에서 확인하십시오. "비모수 다중 비교" 절차는 쌍별 비교의 전체 오차율을 제어하는 데도 사용할 수 있습니다. 자세한 내용은 "비모수 다중 비교 보고서"에서 확인하십시오.

Wilcoxon Kruskal-Wallis, 중앙값 , Friedman 순위 및 Van der Waerden 검정 보고서

일원 분석 플랫폼의 Wilcoxon Kruskal-Wallis, 중앙값 , Friedman 순위 또는 Van der Waerden 검정은 두 개 또는 세 개의 테이블에 요약 통계량과 검정 결과가 제공됩니다 . 요약 통계량 테이블에는 다음 열이 포함되어 있습니다 .

수준 X 의 수준입니다 .

개수 각 수준의 빈도입니다 .

스코어 합 각 수준의 순위 스코어 합입니다 .

기대 스코어 클래스 수준 간에 차이가 없다는 귀무가설하의 기대 스코어입니다 .

스코어 평균 각 수준의 평균 순위 스코어입니다 .

(평균 - 평균 0)/ 표준 0 표준화된 스코어입니다 . 평균 0 은 귀무가설하의 기대 평균 스코어입니다 . 표준 0 은 귀무가설하의 기대 스코어 합에 대한 표준편차입니다 . 귀무가설은 그룹 간에 그룹 평균 또는 중앙값이 동일하다는 것입니다 .

Wilcoxon 2 표본 검정 , 정규 근사 테이블

X 변수의 수준이 정확히 두 개인 경우 "2 표본 검정 , 정규 근사 " 테이블에 다음 열이 포함됩니다 .

S 관측값 수가 더 적은 수준에 대한 순위 스코어의 합입니다 .

Z 0.5 연속성 수정을 사용하는 정규 근사 검정에 대한 검정 통계량입니다 . 자세한 내용은 "2 표본 정규 근사 " 에서 확인하십시오 .

Prob>|Z| 0.5 연속성 수정을 사용하는 정규 근사 검정에 대한 *p* 값입니다. 이 *p* 값은 표준 정규 분포를 기반으로 합니다.

Kruskal-Wallis 검정 , 카이제곱 근사

"Kruskal-Wallis 검정 " 테이블에는 위치에 대한 카이제곱 검정 결과가 포함됩니다. 자세한 내용은 Conover 연구 자료 (1999) 에서 확인하십시오. 그룹 수가 두 개인 경우 Kruskal-Wallis 검정은 Wilcoxon 검정과 동일합니다.

카이제곱 카이제곱 검정 통계량의 값입니다. 자세한 내용은 " 일원 카이제곱 근사 " 에서 확인하십시오.

DF 검정의 자유도입니다.

Prob>ChiSq 검정의 *p* 값입니다. *p* 값은 자유도가 X 의 수준 수에서 1 을 뺀 값에 해당하는 카이제곱 분포를 기준으로 합니다. 그룹 수가 두 개인 경우 이 *p* 값은 0.5 연속성 수정을 사용하지 않는 정규 근사 검정의 *p* 값과 같습니다.

Kolmogorov-Smirnov 2 표본 검정 보고서

일원 분석 플랫폼의 "Kolmogorov-Smirnov 검정 " 보고서에는 요약 테이블과 점근적 검정 테이블이라는 두 개의 테이블이 있습니다. 요약 테이블에는 다음 열이 포함됩니다.

수준 X 변수의 두 수준입니다.

개수 각 수준의 빈도입니다.

최대값에서의 EDF 두 EDF 간의 차이가 최대인 X 변수의 해당 수준에 대한 EDF(경험적 누적 분포 함수) 값입니다. " 합계 " 의 경우 이 값은 두 EDF 간의 차이가 최대인 X 변수의 값에서 합동 EDF(전체 데이터 집합에 대한 EDF) 입니다.

최대값에서 평균과의 편차 각 수준에 대해 다음 단계에서 구한 값입니다.

- 해당 수준의 최대값에서의 EDF 와 합동 데이터 집합 (합계) 의 최대값에서의 EDF 사이의 차이를 계산합니다.
- 이 차이에 해당 수준에 대한 관측값 수의 제곱근을 곱합니다.

점근적 검정 테이블에는 다음 열이 포함됩니다.

KS 다음과 같이 계산된 Kolmogorov-Smirnov 통계량입니다.

$$\mathrm{KS} = \max_j \sqrt{\frac{1}{n}\sum_i n_i (F_i(x_j) - F(x_j))^2}$$

이 계산식에서는 다음과 같은 표기를 사용합니다.

- $x_j (j = 1, \ldots, n)$ 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.

참고 : 이 버전의 Kolmogorov-Smirnov 통계량은 X 변수의 수준이 세 개 이상인 경우에도 적용되지만 JMP 에서 Kolmogorov-Smirnov 옵션은 X 변수의 수준이 정확히 두 개인 경우에만 사용할 수 있습니다 .

KSa $KS\sqrt{n}$ 으로 계산된 점근적 Kolmogorov-Smirnov 통계량입니다 . 여기서 n 은 총 관측값 수입니다 .

D=max|F1-F2| 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(F1-F2) 첫 번째 그룹의 수준이 두 번째 그룹의 수준을 초과한다는 대립가설에 대한 단측 검정 통계량입니다 .

Prob > D+ D+ 검정의 p 값입니다 .

D- = max(F2-F1) 두 번째 그룹의 수준이 첫 번째 그룹의 수준을 초과한다는 대립가설에 대한 단측 검정 통계량입니다 .

Prob > D- D- 검정의 p 값입니다 .

정확 검정 보고서

X 변수의 수준이 정확히 두 개인 경우 일원 분석 플랫폼에서 각 비모수 검정 유형에 대해 정확 검정을 수행할 수 있습니다. Wilcoxon Kruskal-Wallis, 중앙값 및 Van der Waerden 정확 검정의 경우 "2 표본 : 정확 검정 " 테이블에 다음 열이 포함됩니다 .

S 더 작은 그룹의 관측값에 대한 순위 스코어의 합입니다 . X 의 두 수준에 동일한 수의 관측값이 있으면 S 값은 값 순서화 시 X 의 마지막 수준에 대응합니다 .

Prob ≤ S 검정의 단측 p 값입니다 .

Prob ≥ |S-Mean| 검정의 양측 p 값입니다 .

참고: Kolmogorov-Smirnov 정확 검정의 경우 테이블에 점근적 검정과 동일한 통계량이 제공됩니다 . 그러나 p 값은 정확하게 계산됩니다 . 자세한 내용은 "Kolmogorov-Smirnov 2 표본 검정 보고서 " 에서 확인하십시오 .

비모수 다중 비교 보고서

일원 분석 플랫폼의 "비모수 다중 비교" 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. 이 섹션에서는 이러한 방법에 대한 보고서를 다룹니다.

Wilcoxon 개별 쌍 비교 , Steel-Dwass 전체 쌍 비교 및 Steel 대조군 비교 보고서

이러한 다중 비교 절차에 대한 보고서에는 검정 결과와 신뢰 구간이 포함됩니다 . 이러한 검정의 경우 해당 비교에 사용된 두 수준만 고려하여 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 .

스코어 평균 차이 첫 번째 수준 (" 수준 ") 에 포함된 관측값의 순위 스코어 평균에서 두 번째 수준 ("- 수준 ") 에 포함된 관측값의 순위 스코어 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 스코어의 차이가 양수인 경우 " 스코어 평균 차이 " 는 다음과 같이 정의됩니다 .

$$\text{스코어 평균 차이} = (\text{ScoreSum}_1 - 0.5) / n_1 - (\text{ScoreSum}_2 + 0.5) / n_2$$

평균 스코어의 차이가 음수인 경우 " 스코어 평균 차이 " 는 다음과 같이 정의됩니다 .

$$\text{스코어 평균 차이} = (\text{ScoreSum}_1 + 0.5) / n_1 - (\text{ScoreSum}_2 - 0.5) / n_2$$

차이 표준 오차 스코어 평균 차이의 표준 오차입니다 .

Z 평균에 차이가 없다는 귀무가설하에서 점근적 표준 정규 분포를 따르는 표준화된 검정 통계량입니다 .

p 값 Z 를 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 .

차이 그림 스코어 평균 차이의 막대 차트입니다 .

Dunnett 전체 쌍 비교 (결합 순위 기준) 및 Dunnett 대조군 비교 (결합 순위 기준) 보고서

Dunnett 비교 절차는 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 .

스코어 평균 차이 첫 번째 수준 (" 수준 ") 에 포함된 관측값의 순위 스코어 평균에서 두 번째 수준 ("- 수준 ") 에 포함된 관측값의 순위 스코어 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 " 스코어 평균 차이 " 에서 확인하십시오 .

차이 표준 오차 스코어 평균 차이의 표준 오차입니다 .

Z 평균에 차이가 없다는 귀무가설하에서 점근적 표준 정규 분포를 따르는 표준화된 검정 통계량입니다 .

p 값 *Z* 를 기반으로 한 점근적 검정의 *p* 값입니다 .

이분산 보고서

일원 분석 플랫폼은 그룹 분산의 동일성에 대한 4가지 검정을 제공합니다. 처음 세 가지 이분산 검정의 기본 개념은 각 그룹의 퍼짐 정도를 측정하기 위해 구성된 새 반응 변수에 대해 분산 분석을 수행하는 것입니다. 네 번째 검정은 정규 분포하에서 가능도비 검정과 유사한 Bartlett 검정입니다. 블록 변수가 시작 창에 지정된 경우 이분산 옵션을 사용할 수 없습니다.

참고 : 이분산을 검정하는 또 다른 방법은 ANOMV 입니다 . 자세한 내용은 " 평균 분석 보고서 " 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 ANOVA 는 실제로는 그룹 표본 분산에 대한 ANOVA 입니다 (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). 퍼짐은 $z_{ij} = |y_{ij} - \tilde{y}_i|$ 로 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"에서 확인하십시오.

팁 : 이분산 검정을 통해 그룹 분산에 유의한 차이가 있는 것으로 나타나면 표준 ANOVA 검정 대신 Welch 검정을 고려하십시오 . 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 개뿐인 경우 Welch ANOVA 는 이분산 *t*- 검정과 동등합니다 .

분산 동일 여부 검정 보고서

일원 분석 플랫폼의 "분산 동일 여부 검정" 보고서에는 표준편차 그림이 표시되고 요약 테이블이 제공됩니다. 첫 번째 테이블에는 다음 열이 포함되어 있습니다.

수준 데이터에서 나타나는 요인 수준입니다 .

개수 각 수준의 빈도입니다 .

표준편차 각 요인 수준의 반응 표준편차입니다 . 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 차이입니다 . 이 평균 절대 차이는 Levene 변수의 그룹 평균과 동일합니다 .

중앙값에 대한 평균 절대 차이 반응 및 그룹 중앙값의 절대 차이입니다 . 이 평균 절대 차이는 Brown-Forsythe 변수의 그룹 평균과 동일합니다 .

두 번째 테이블에는 등분산 검정이 요약되어 있고 다음 열이 포함됩니다 .

검정 각 검정의 이름입니다 .

F 비 계산된 *F* 통계량입니다 . 자세한 내용은 " 분산 동일 여부 검정에 대한 통계 상세 정보 " 에서 확인하십시오 .

DFNum 분자의 자유도입니다 . 요인에 *k* 개의 수준이 있는 경우 분자의 자유도는 *k* - 1 입니다 . 데이터에 한 번만 나타나는 수준은 O'Brien, Brown-Forsythe 또는 Levene 의 검정 통계량을 계산하는 데 사용되지 않습니다 . 이 경우 분자 자유도는 계산에 사용된 수준 수에서 1 을 뺀 값입니다 .

DFDen 분모에 사용된 자유도입니다. O'Brien, Brown-Forsythe 및 Levene 의 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

F 비 평균 동일성 검정에 대한 *F*- 검정 통계량입니다.

DFNum 검정의 분자 자유도입니다. 요인에 *k* 개의 수준이 있는 경우 분자의 자유도는 *k* - 1 입니다. 데이터에 한 번만 나타나는 수준은 Welch ANOVA 를 계산하는 데 사용되지 않습니다. 이 경우 분자 자유도는 계산에 사용된 수준 수에서 1 을 뺀 값입니다.

DFDen 검정의 분모 자유도입니다. 자세한 내용은 " 분산 동일 여부 검정에 대한 통계 상세 정보 " 에서 확인하십시오.

Prob>F 모든 수준의 평균이 동일한 경우 *F* 값이 계산된 값보다 클 확률입니다.

t- 검정 (X 변수의 수준이 정확히 두 개인 경우에만 사용 가능) *F* 비와 *t*- 검정 사이의 관계를 보여 줍니다. *F* 비의 제곱근으로 계산됩니다.

동등성 검정 보고서

일원 분석 플랫폼에서 "동등성 검정" 하위 메뉴의 옵션을 사용하면 평균 또는 표준편차에 대한 동등성, 우월성 또는 비열등성 검정을 수행할 수 있습니다.

동등성 검정의 경우 TOST(Two One-Sided Tests: 두 개의 단측 검정) 방법을 사용하여 평균 또는 두 표준편차의 비율 간 실제적 동등성을 검정합니다. 평균의 동등성 검정에 대한 내용은 Schuirmann 연구 자료 (1987) 에서 확인하십시오. 실제 차이 또는 실제 비율이 임계값을 초과한다는 귀무가설에 대해 두 개의 단측 *t*- 검정이 구성됩니다. 두 검정이 모두 기각되면 평균 또는 비율의 차이가 통계적으로 두 임계값 중 어느 하나도 초과하지 않습니다. 따라서 그룹은 실제적으로 동등하다고 간주됩니다. 자세한 내용은 "동등성 검정의 예" 에서 확인하십시오.

동등성 검정 보고서에는 그림과 요약 테이블이 포함되어 있습니다. "동등성 검정" 하위 메뉴에서 동등성 검정 옵션을 선택하는 경우 "동등성 검정 규격" 창에서 검정 특성을 지정해야 합니다.

동등성 검정 규격 창

" 동등성 검정 " 하위 메뉴에서 동등성 검정 옵션을 선택하면 검정을 정의할 수 있는 창이 시작됩니다 .

대립가설 동등성 검정의 구조를 정의합니다 .

동등성 (양측) 동등성 검정을 지정합니다 . 그룹 차이가 동등성 한계보다 크지 않음을 보여 주는 것이 목표인 경우 이 옵션을 사용합니다 .

우월성 (단측) 우월성 검정을 지정합니다 . 한 그룹이 다른 그룹보다 우월하거나 더 우수함을 보여 주는 것이 목표인 경우 이 옵션을 사용합니다 .

비열등성 (단측) 비열등성 검정을 지정합니다. 한 그룹이 다른 그룹보다 열등하지 않음을 보여 주는 것이 목표인 경우 이 옵션을 사용합니다 .

대립가설 부분 (단측 검정에만 사용 가능) 대립가설의 방향 (목표) 을 지정합니다 .

가설 그림 가설 검정의 그래픽 표현을 제공합니다 .

비교 유형 수행할 비교 유형을 정의합니다 .

전체 쌍별 모든 그룹 수준 쌍 간의 비교를 지정합니다 .

대조군 비교 각 수준과 지정된 대조 수준의 비교를 지정합니다 .

분산 가정 검정 계산에 사용되는 분산 가정을 정의합니다 .

합동 분산 합동 분산을 기반으로 계산하도록 지정합니다 . 그룹 간 분산이 동일하다고 가정할 때 이 옵션을 사용합니다 .

이분산 동일하지 않은 그룹 분산을 기반으로 계산하도록 지정합니다 .

마진 및 유의 수준 검정의 유의 수준을 정의합니다 .

차이 (평균 검정에만 사용 가능) 동등성 , 우월성 또는 비열등성 마진을 지정합니다 . 이 마진 (델타) 은 실제적 유의성을 가진 차이입니다 . 동등성 검정의 경우 차이가 0 보다 커야 합니다 .

비율 (표준편차 검정에만 사용 가능) 동등성 , 우월성 또는 비열등성 마진을 표준편차의 비율로 지정합니다 . 이 마진은 실제적 유의성을 가진 표준편차 비율을 정의합니다 . 값 범위는 (비율 , 1/ 비율) 로 정의됩니다 . 동등성 검정의 경우 비율이 1 과 달라야 합니다 .

유의 수준 검정의 유의 수준을 지정합니다 .

동등성 검정 보고서

검정 보고서는 대립가설에 대한 설명으로 시작됩니다 . 각 비교에 대해 검정 보고서에는 다음 열이 포함됩니다 .

차이 (평균 검정에만 사용 가능) 추정된 평균 차이입니다 .

차이의 표준 오차 (평균 검정에만 사용 가능) 평균 차이의 추정된 표준 오차입니다 .

비율 (표준편차 검정에만 사용 가능) 추정된 표준편차 비율입니다 .

하한 t 비 , 상한 t 비 (평균 검정에만 사용 가능 . 단측 검정의 경우 한계가 하나만 표시됨) 단측 유의성 검정의 하한 또는 상한 *t* 비입니다 .

하한 F 값 , 상한 F 값 (표준편차 검정에만 사용 가능 . 단측 검정의 경우 한계가 하나만 표시됨) 단측 유의성 검정의 하한 또는 상한 *t* 비입니다 .

하한 p 값 , 상한 p 값 (단측 검정의 경우 *p* 값이 하나만 표시됨) 하한 또는 상한 *t* 비에 해당하는 유의 확률 (*p* 값) 입니다 .

최대 p 값 (동등성 검정의 경우에만 표시됨) 하한 및 상한 *p* 값의 최대값입니다 .

90% 하한 , 90% 상한 평균 차이 또는 표준편차 비율에 대한 1–2α 신뢰 구간의 한계입니다 .

평가 지정된 유의 수준에 대한 가설 검정 평가입니다 .

동등성 검정 옵션

" 동등성 검정 ", " 우월성 검정 " 또는 " 비열등성 검정 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

검정 보고서 동등성 검정을 요약하는 보고서를 표시하거나 숨깁니다. 자세한 내용은 "동등성 검정 보고서 " 에서 확인하십시오 .

산점도 (평균 검정에만 사용 가능) 산점도를 표시하거나 숨깁니다. 채택 영역에 따라 색상이 지정된 쌍별 비교 신뢰 구간이 산점도에 표시됩니다. 구간은 동등 , 우월 또는 비열등 영역을 나타내는 음영을 사용하여 평균 - 평균 척도로 표시됩니다. 이 그림을 **차이 그림** (diffogram) 또는 평균 - 평균 산점도라고도 합니다 .

팁 : 점을 커서로 가리키면 비교되는 그룹과 추정된 차이가 표시됩니다 .

" 산점도 " 에는 다음 옵션이 있습니다 .

참조선 표시 산점도에 점에 대한 참조선을 표시하거나 숨깁니다 . 산점도에 점이 많이 있는 경우 권장되지 않습니다 . 점이 많은 경우에는 점을 커서로 가리켜 라벨을 보는 것이 좋습니다 .

포레스트 그림 포레스트 그림을 표시하거나 숨깁니다. 비교 신뢰 구간 대 평균 차이 또는 표준편차 비율이 표시됩니다 . 구간은 평균 차이 또는 표준편차 비율 척도로 표시됩니다 . 음영은 동등 , 우월 또는 비열등 영역을 나타냅니다 .

팁 : 점을 커서로 가리키면 비교되는 그룹과 추정된 차이 또는 비율이 표시됩니다 .

쌍별 비교 (평균 동등성 검정에만 사용 가능) 전체 쌍별 비교에 대한 실제적 동등성 보고서를 표시하거나 숨깁니다 . 이 보고서에는 " 동등성 검정 보고서 " 에 제공된 값 외에도 각 쌍체 검정에 대한 시각적 표현이 포함됩니다 .

제거 " 일원 분석 " 보고서 창에서 검정 보고서를 제거합니다 .

로버스트 적합 보고서

일원 분석 플랫폼의 로버스트 옵션은 데이터 집합에서 이상치 또는 극단 데이터 점의 영향을 줄이기 위한 두 가지 방법으로 " 로버스트 적합 " 과 "Cauchy 적합 " 을 제공합니다 . 각 방법은 일원 분석 그림의 로버스트 평균 위치에 선을 추가합니다 .

- " 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 . 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 " 에서 확인하십시오 . Huber 손실 함수에 대한 자세한 내용은 " 로버스트 적합에 대한 통계 상세 정보 " 에서 확인하십시오 .
- 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 가능도와 Cauchy 연결 함수를 사용하여 모수를 추정합니다 .

참고 : 블록 변수가 시작 창에 지정된 경우 로버스트 옵션을 사용할 수 없습니다 .

로버스트 옵션에 대해 생성된 보고서에는 두 개의 테이블이 있습니다 . 이 테이블에는 다음 열이 포함됩니다 .

시그마 RMSE(제곱근 평균 제곱 오차) 와 동일한 시그마 값입니다 .

카이제곱 모형이 반응의 평균보다 더 잘 적합된다는 가설에 대한 검정 통계량입니다 .

p 값 기울기가 0 인 카이제곱 검정의 p 값입니다 .

LogWorth LogWorth 는 $-\log_{10}(p$ 값) 으로 정의된 p 값의 변환입니다 .

수준 X 변수의 수준입니다 .

로버스트 평균 Huber 또는 Cauchy 방법을 기반으로 추정된 로버스트 평균입니다 .

표준 오차 모수 추정값의 표준 오차 추정값입니다 .

검정력 보고서

일원 분석 플랫폼의 " 검정력 " 옵션은 통계 검정력을 계산하고 지정된 가설 검정에 대한 기타 상세 정보를 제공합니다 . 자세한 내용은 " 검정력 옵션의 예 " 에서 확인하십시오 . 통계 상세 정보는 " 검정력에 대한 통계 상세 정보 " 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 . LSV 는 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 . LSN 은 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 LSN 을 통해 필요한 데이터 양을 알 수 있습니다 . LSN 은 검정력이 약 50% 가 되도록 하는 총 관측값 수입니다 .
- **검정력**은 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 LSV 는 다중 비교 테이블에 표시된 LSD(최소 유의차) 와 동일합니다 .

검정력 상세 정보

일원 분석 플랫폼의 "검정력 상세 정보" 창 및 테이블은 모형 적합 플랫폼에서와 동일합니다. 검정력 계산에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

알파 , 시그마 , 델타 및 개수를 나타내는 4 개의 각 열에 단일 값 , 두 개의 값 , 또는 값 시퀀스의 시작 , 끝 및 증분 값을 입력해야 합니다 (그림 6.29). 지정한 값의 가능한 모든 조합에 대해 검정력 계산이 수행됩니다 .

알파 (α) 유의 수준으로 , 0 에서 1 사이입니다 (일반적으로 0.05, 0.01 또는 0.10). 기본적으로 표시되는 값은 0.05 입니다 .

시그마 (σ) 모형의 잔차에 대한 표준 오차입니다 . 기본적으로 제공되는 값은 RMSE(제곱근 평균 제곱 오차) 에서 구한 추정값입니다 .

델타 (δ) 원시 효과 크기입니다 . 효과 크기 계산에 대한 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 . 첫 번째 위치는 기본적으로 가설의 제곱근 제곱합을 *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 변수의 수준이 정확히 두 개일 때만 사용할 수 있으며 일반적으로 개별 모수에 사용됩니다 .

조정 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 **델타**에서 발생하므로 원래 델타에 대해서만 조정된 검정력 및 신뢰 한계가 계산됩니다 .

매칭 열 보고서

일원 분석 플랫폼에서 " 매칭 열 " 옵션을 사용하여 매칭 모형 분석을 위한 매칭 (ID) 변수를 지정합니다 . 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 . 블록 변수가 시작 창에 지정된 경우 매칭 열 옵션을 사용할 수 없습니다 . 자세한 내용은 " 매칭 열 옵션의 예 " 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 (X 로 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 사이의 연결선을 일원 분석 그림에 추가합니다 . 매칭 ID 값이 같은 관측값이 여러 개 있는 경우에는 선이 그룹 평균을 연결합니다 . 일원 분석 그림에서 선을 제거하려면 **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 **선형 모형 적합**의 에서 확인하십시오 .

일원 분석 그림 요소

일원 분석 플랫폼에는 연속형 Y 변수 대 범주형 X 변수 그림이 포함됩니다 . 이 그림을 사용하면 X 변수의 각 수준에 대한 Y 변수의 분포를 시각화할 수 있습니다 . 일원 분석 플랫폼의 많은 옵션은 시각화 요소를 그림에 추가합니다 . " 일원 분석 " 의 빨간색 삼각형 메뉴에는 표시 요소를 추가하거나 제거할 수 있는 " 표시 옵션 " 하위 메뉴가 있습니다 . 이 섹션에는 다음 표시 요소에 대한 상세 정보가 포함되어 있습니다 .

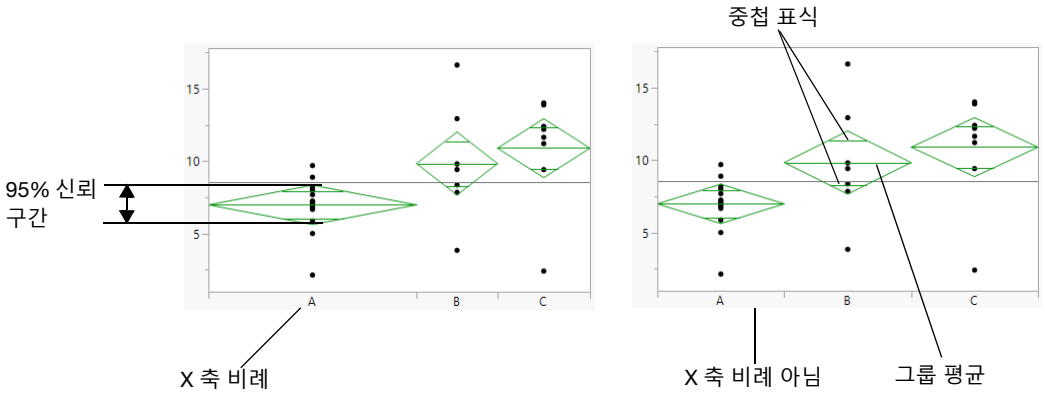
- " 평균 다이아몬드 및 X 축 비례 "
- " 평균 선 , 오차 막대 및 표준편차 선 "

- " 비교 원 "

평균 다이아몬드 및 X 축 비례

일원 분석 플랫폼에서 가로 축은 기본적으로 비례 축입니다. 즉, 그룹 수준의 간격은 각 그룹의 데이터 점 수에 비례합니다. 평균 다이아몬드 표시 옵션 또는 ANOVA 검정은 일원 분석 그림에 평균 다이아몬드를 추가합니다. 평균 다이아몬드에는 X 변수의 수준에 대한 표본 평균과 신뢰 구간이 표시됩니다.

그림 6.6 평균 다이아몬드 및 X 축 비례 옵션



평균 다이아몬드 상세 정보

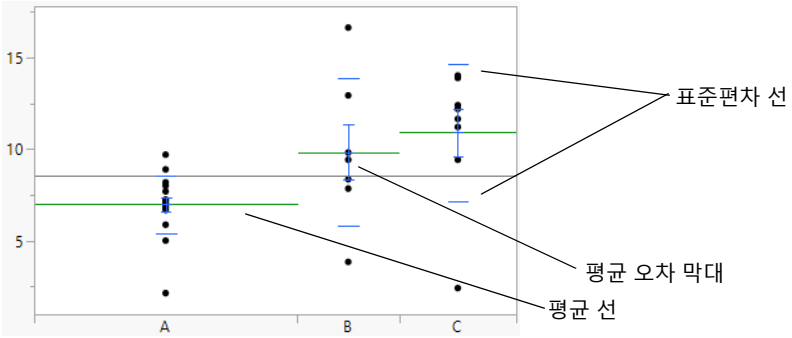
- 각 다이아몬드의 맨 위와 맨 아래는 각 그룹의 평균에 대한 $(1 - \alpha) \times 100$ 신뢰 구간에 걸쳐 있습니다 . 신뢰 구간은 관측값 간에 분산이 동일하다는 가정하에 계산됩니다 . 따라서 다이아몬드의 높이는 그룹 내의 관측값 수에 대한 제곱근의 역수에 비례합니다 .
- **X 축 비례** 옵션이 선택된 경우 가로 축에서 각 그룹이 차지하는 가로 범위 (다이아몬드의 가로 크기) 는 X 변수의 각 수준에 대한 표본 크기에 비례합니다 . 따라서 보통은 다이아몬드의 가로가 좁을수록 세로가 길어지는데, 이는 데이터 점이 적을수록 신뢰 구간 추정값이 넓어지기 때문입니다 .
- 각 다이아몬드의 가운데를 가로지르는 평균 선이 그룹 평균 위치에 표시됩니다 .
- 다이아몬드의 맨 위와 맨 아래 근처의 선은 중첩 표식입니다 . 표본 크기가 동일한 그룹의 경우 중첩 표식 내에서 중첩되는 다이아몬드는 주어진 신뢰 수준에서 두 그룹 평균이 유의하게 다르지 않음을 나타냅니다 . 중첩 표식은 그룹 평균 $\pm (\sqrt{2})/2 \times \mathrm{CI}/2$ 로 계산됩니다 .
- 플랫폼 메뉴에서 **평균 /ANOVA/ 합동 t** 또는 **평균 /ANOVA** 옵션을 선택하면 일원 분석 그림에 평균 다이아몬드가 추가됩니다 . 하지만 언제든지 빨간색 삼각형 메뉴에서 **표시 옵션 > 평균 다이아몬드**를 선택하여 이를 표시하거나 숨길 수도 있습니다 .

평균 선 , 오차 막대 및 표준편차 선

일원 분석 플랫폼에서 **표시 옵션 > 평균 선** 옵션을 선택하여 일원 분석 그림에 평균 선을 표시합니다. 평균 선은 X 변수의 각 수준에 대한 반응 평균 위치에 표시됩니다.

빨간색 삼각형 메뉴에서 **평균 및 표준편차** 옵션을 선택하여 평균 오차 막대와 표준편차 선을 그림에 추가합니다. 각 옵션을 개별적으로 설정하거나 해제하려면 **표시 옵션 > 평균 오차 막대** 또는 **표준편차 선**을 선택합니다.

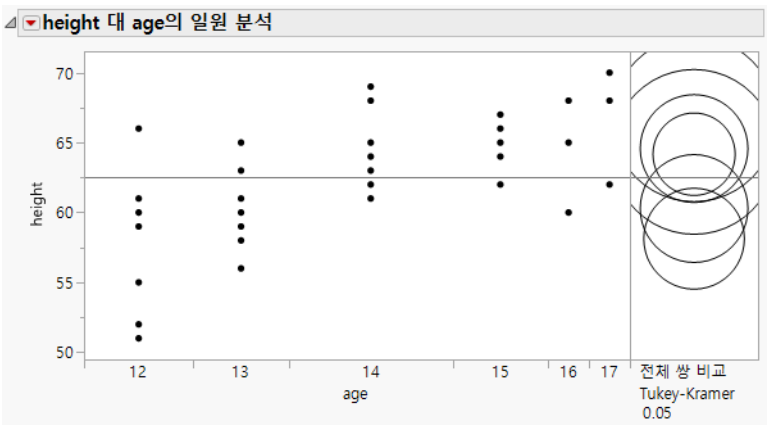
그림 6.7 평균 선 , 평균 오차 막대 및 표준편차 선



비교 원

일원 분석 플랫폼에서 단계별 개별 쌍 비교(Newman-Keuls) 방법을 제외한 각 다중 비교 검정은 일원 분석 그림에 **비교 원**을 추가합니다. 비교 원은 그룹 평균 비교의 대화식 시각적 표현을 제공합니다. 그림 6.8에서는 전체 쌍 비교(Tukey HSD) 방법에 대한 비교 원을 보여 줍니다. 원의 반지름은 검정 및 유의 수준을 기반으로 합니다. 자세한 내용은 "비교 원에 대한 통계 상세 정보"에서 확인하십시오.

그림 6.8 그룹 평균의 시각적 비교



비교 원의 교차 영역을 검토하여 각 그룹 평균 쌍을 비교할 수 있습니다 . 교차점의 외각은 그룹 평균 간에 유의한 차이가 있는지 여부를 나타냅니다 (그림 6.9).

- 유의한 차이가 있는 평균의 경우 비교 원이 교차하지 않거나 약간 교차하므로 교차점의 외각이 90 도 미만입니다 .
- 원의 교차 각도가 90 도를 초과하거나 원이 내포된 경우에는 해당 그룹 평균 간에 유의한 차이가 없는 것입니다 .
- 원을 클릭하여 해당 그룹을 강조 표시합니다 . 회색 원은 강조 표시된 그룹과 유의한 차이가 있는 평균을 가진 그룹에 해당합니다 (그림 6.10). 원을 선택 취소하려면 원 외부의 공백을 클릭합니다 .

그림 6.9 교차 각도 및 유의성

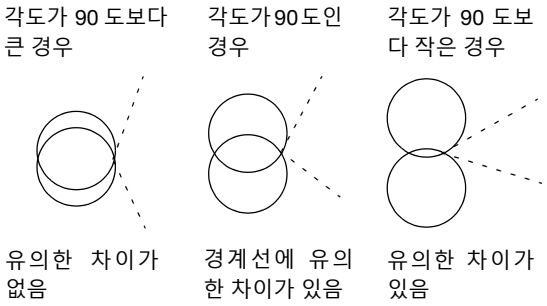
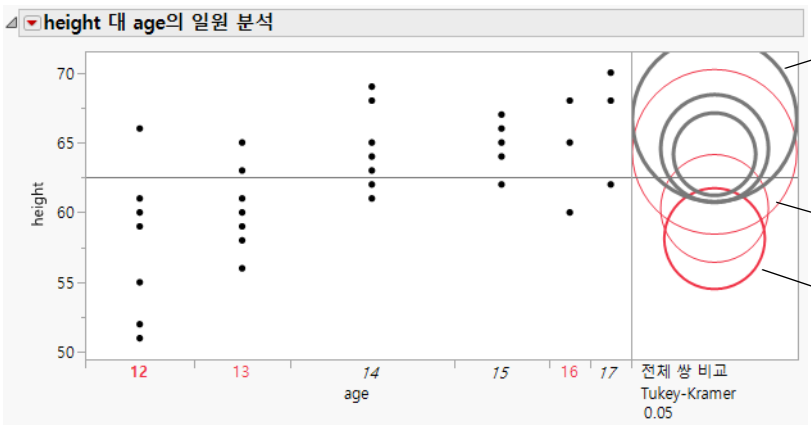


그림 6.10 비교 원 강조 표시



일원 분석 플랫폼의 추가 예

이 섹션에는 일원 분석 플랫폼을 사용한 예가 포함되어 있습니다 .

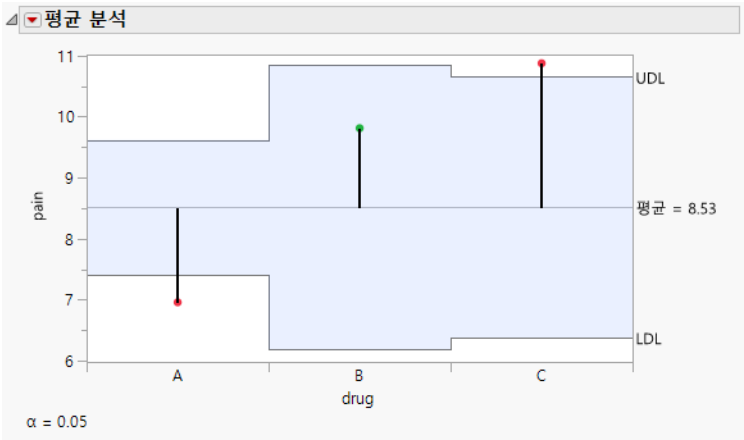
- " 평균 분석의 예 "
- " 분산에 대한 평균 분석의 예 "
- " 개별 쌍 비교 (스튜던트 t) 검정의 예 "
- " 전체 쌍 비교 (Tukey HSD) 검정의 예 "
- " 최량 비교 (Hsu MCB) 검정의 예 "
- " 대조군 비교 (Dunnett) 의 예 "
- " 단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예 "
- " 예 : 4 가지 평균 비교 검정 대조 "
- " 비모수 Wilcoxon 검정의 예 "
- " 이분산 옵션의 예 "
- " 동등성 검정의 예 "
- " 로버스트 적합 옵션의 예 "
- " 검정력 옵션의 예 "
- " 정규 분위수 그림의 예 "
- "CDF 그림의 예 "
- " 밀도 옵션의 예 "
- " 매칭 열 옵션의 예 "
- " 예 : 일원 분석을 위한 데이터 쌓기 "

평균 분석의 예

일원 분석 플랫폼에서 ANOM(다중 비교 평균 분석) 방법을 사용하여 평균의 차이를 검정합니다.

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형을 클릭하고 **평균 분석 방법 > ANOM** 을 선택합니다 .

그림 6.11 평균 분석 의사결정 차트



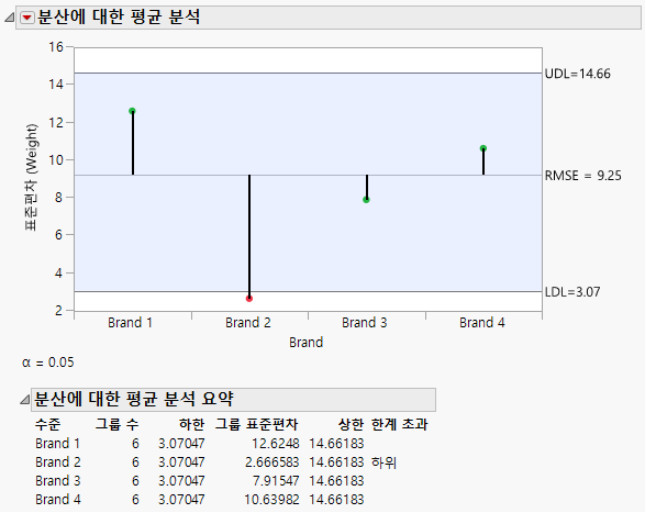
ANOM 의사결정 차트를 보면 약물 A 와 C 의 평균이 전체 평균과 유의하게 다릅니다. 약물 A 의 평균은 더 낮고 약물 C 의 평균은 더 높습니다. 표본 크기가 다르기 때문에 약물 유형에 대한 결정 한계는 동일하지 않습니다.

분산에 대한 평균 분석의 예

일원 분석 플랫폼에서 ANOMV(분산에 대한 평균 분석) 를 사용하여 분산의 차이를 검정합니다 . 스프링 길이를 0.10 인치 늘리기 위해 필요한 중량을 확인하기 위해 4 가지 브랜드의 스프링을 테스트했습니다 . 각 브랜드에서 6 개의 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 폴더**를 선택하고 Spring Data.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Weight 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Brand 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 분석 방법 > 분산에 대한 ANOM** 을 선택합니다 .
7. " 분산에 대한 평균 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **요약 보고서 표시**를 선택합니다 .

그림 6.12 분산에 대한 평균 분석 차트



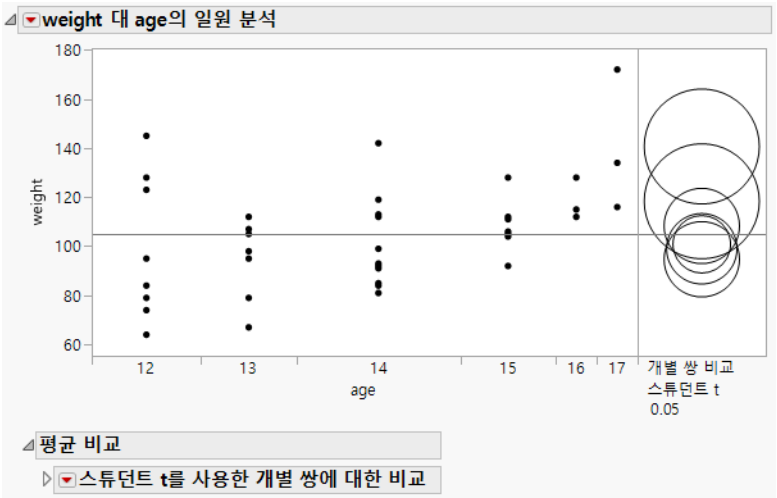
Brand 2의 표준편차는 결정 하한을 넘습니다. 따라서 Brand 2는 다른 브랜드보다 분산이 유의하게 낮습니다.

개별 쌍 비교 (스튜던트 t) 검정의 예

이 예에서는 일원 분석 플랫폼을 사용하여 가능한 모든 *t*- 검정을 사용하는 것을 보여 줍니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 비교 > 개별 쌍 비교 (스튜던트 t)** 를 선택합니다 .

그림 6.13 개별 쌍 비교 (스튜던트 t) 비교 원의 예



평균 비교 방법은 Fisher LSD(최소 유의차) 방법을 사용하여 개별 평균 쌍 간의 차이를 평가합니다. 이 방법은 동시 비교를 위한 오차율을 유지하지 않습니다. 그룹 수가 많은 경우 다른 비교 방법을 고려하십시오.

그림 6.14 개별 쌍 비교 (스튜던트 *t*) 에 대한 평균 비교 보고서의 예

평균 비교

스튜던트 t를 사용한 개별 쌍에 대한 비교

신뢰 분위수

t	Alpha
2.03224	0.05

LSD 임계 행렬

절대(차이)-LSD

	17	16	15	14	12	13
17	-33.140	-10.807	4.373	13.634	14.189	17.944
16	-10.807	-33.140	-17.961	-8.699	-8.145	-4.389
15	4.373	-17.961	-21.695	-11.851	-11.720	-8.124
14	13.634	-8.699	-11.851	-16.570	-16.692	-13.184
12	14.189	-8.145	-11.720	-16.692	-20.294	-16.720
13	17.944	-4.389	-8.124	-13.184	-16.720	-21.695

양수 값은 유의차가 있는 평균의 쌍을 보여 줍니다.

연결 문자 보고서

수준		평균
17	A	140.66667
16	A B	118.33333
15	B	108.28571
14	B	100.83333
12	B	99.00000
13	B	94.71429

동일한 문자로 연결되지 않은 수준에는 유의한 차이가 있습니다.

정렬된 차이 보고서

수준	- 수준	차이	차이 표준 오차	CL 하한	CL 상한	p 값
17	13	45.95238	13.78194	17.9441	73.96066	0.0021*
17	12	41.66667	13.52106	14.1886	69.14477	0.0041*
17	14	39.83333	12.89183	13.6340	66.03268	0.0040*
17	15	32.38095	13.78194	4.3727	60.38923	0.0247*
16	13	23.61905	13.78194	-4.3892	51.62733	0.0957
17	16	22.33333	16.30702	-10.8065	55.47318	0.1798
16	12	19.33333	13.52106	-8.1448	46.81144	0.1619
16	14	17.50000	12.89183	-8.6993	43.69935	0.1836
15	13	13.57143	10.67545	-8.1237	35.26655	0.2123
16	15	10.04762	13.78194	-17.9607	38.05590	0.4710
15	12	9.28571	10.33646	-11.7205	30.29192	0.3753
15	14	7.45238	9.49855	-11.8510	26.75576	0.4381
14	13	6.11905	9.49855	-13.1843	25.42242	0.5238
12	13	4.28571	10.33646	-16.7205	25.29192	0.6810
14	12	1.83333	9.11590	-16.6924	20.35907	0.8418

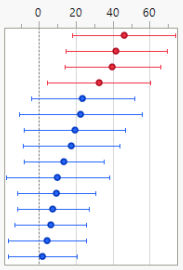


그림 6.14의 LSD 임계 테이블에서는 평균의 절대 차이와 LSD(최소 유의차) 간의 차이를 보여 줍니다. 값이 양수이면 두 평균의 차이가 LSD 보다 크고 두 그룹이 통계적으로 유의하게 다른 것입니다.

전체 쌍 비교 (Tukey HSD) 검정의 예

이 예에서는 일원 분석 플랫폼을 사용하여 Tukey HSD 검정을 보여 줍니다.

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 비교 > 전체 쌍 비교 (Tukey HSD)** 를 선택합니다.

그림 6.15 전체 쌍 비교 (Tukey HSD) 의 비교 원

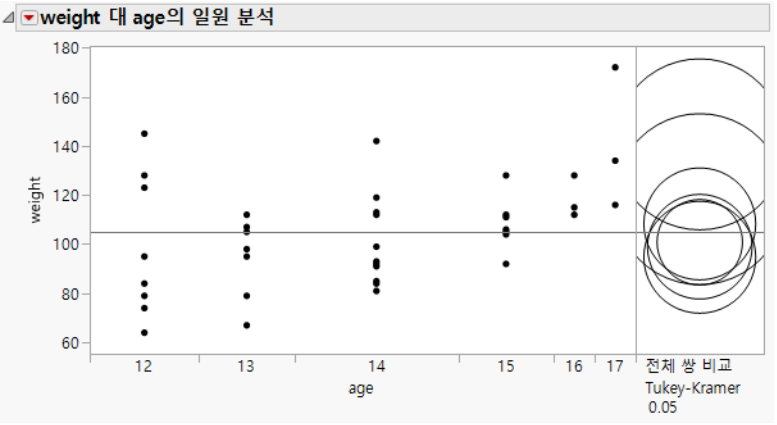
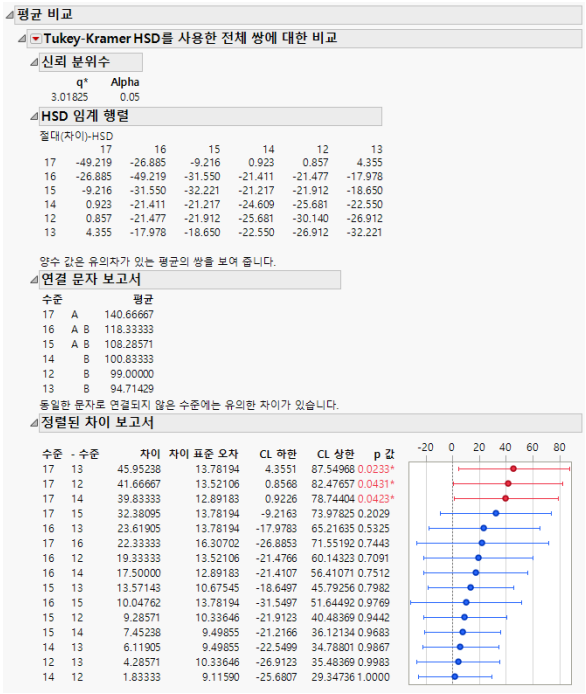


그림 6.16 전체 쌍 비교 (Tukey HSD) 에 대한 평균 비교 보고서



평균 비교

Tukey-Kramer HSD를 사용한 전체 쌍에 대한 비교

신뢰 분위수

q*	Alpha
3.01825	0.05

HSD 임계 행렬

절대(차이)-HSD

	17	16	15	14	12	13
17	-49.219	-26.885	-9.216	0.923	0.857	4.355
16	-26.885	-49.219	-31.550	-21.411	-21.477	-17.978
15	-9.216	-31.550	-32.221	-21.217	-21.912	-18.650
14	0.923	-21.411	-21.217	-24.609	-25.681	-22.550
12	0.857	-21.477	-21.912	-25.681	-30.140	-26.912
13	4.355	-17.978	-18.650	-22.550	-26.912	-32.221

양수 값은 유의차가 있는 평균의 쌍을 보여 줍니다.

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.

정렬된 차이 보고서

수준	- 수준	차이	차이 표준 오차	CL 하한	CL 상한	p 값
17	13	45.95238	13.78194	4.3551	87.54968	0.0233*
17	12	41.66667	13.52106	0.8568	82.47657	0.0431*
17	14	39.83333	12.89183	0.9226	78.74404	0.0423*
17	15	32.38095	13.78194	-9.2163	73.97825	0.2029
16	13	23.61905	13.78194	-17.9783	65.21635	0.5325
17	16	22.33333	16.30702	-26.8853	71.55192	0.7443
16	12	19.33333	13.52106	-21.4766	60.14323	0.7091
16	14	17.50000	12.89183	-21.4107	56.41071	0.7512
15	13	13.57143	10.67545	-18.6497	45.79256	0.7982
16	15	10.04762	13.78194	-31.5497	51.64492	0.9769
15	12	9.28571	10.33646	-21.9123	40.48369	0.9442
15	14	7.45238	9.49855	-21.2166	36.12134	0.9683
14	13	6.11905	9.49855	-22.5499	34.78801	0.9867
12	13	4.28571	10.33646	-26.9123	35.48369	0.9983
14	12	1.83333	9.11590	-25.6807	29.34736	1.0000

그림 6.16의 Tukey-Kramer HSD 임계 행렬에서는 평균의 실제 절대 차이에서 HSD를 뺀 값을 보여 줍니다. 양수 값을 가진 쌍은 지정된 유의 수준에서 통계적으로 유의하게 다른 것입니다. 신뢰 분위수 테이블의 q^* 값은 HSD를 척도화하는 데 사용되는 분위수를 나타냅니다. 이 값은 스튜던트 t-검정의 분위수와 유사한 계산 역할을 합니다.

최량 비교 (Hsu MCB) 검정의 예

이 예에서는 일원 분석 플랫폼을 사용하여 Hsu MCB(최량과의 다중 비교) 검정을 보여 줍니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 비교 > 최량 비교 (Hsu MCB)** 를 선택합니다 .

그림 6.17 최량 비교 (Hsu MCB) 비교 원의 예

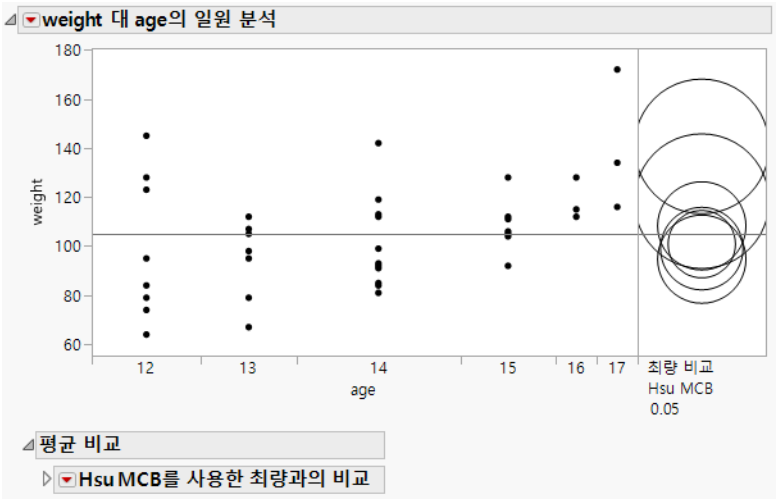


그림 6.18 최량 비교 (Hsu MCB) 에 대한 평균 비교 보고서의 예

평균 비교

Hsu MCB를 사용한 최량과의 비교

신뢰 분위수

d	Alpha
2.23157	0.05
2.23157	
2.34056	
2.38154	
2.35270	
2.34056	

최대값 및 최소값을 사용한 비교

수준	최대 p 값 사용	최소 p 값 사용
17	0.9863	0.0039*
16	0.2219	0.1304
15	0.0491*	0.3172
14	0.0093*	0.6599
12	0.0091*	0.7216
13	0.0046*	0.9408

LSD 임계 행렬

평균[i]-평균[j]-LSD

	17	16	15	14	12	13
17	-36.390	-14.057	0.124	9.131	9.856	13.695
16	-58.724	-36.390	-22.210	-13.202	-12.478	-8.638
15	-63.136	-40.803	-24.987	-15.169	-15.033	-11.415
14	-68.602	-46.269	-29.684	-19.418	-19.614	-16.113
12	-71.840	-49.507	-33.479	-23.543	-23.494	-19.907
13	-76.708	-54.374	-38.558	-28.740	-28.604	-24.987

열에 양수 값이 있는 경우 평균이 최대값보다 상당히 작습니다.

평균[i]-평균[j]+LSD

	17	16	15	14	12	13
17	36.390	58.724	64.638	70.536	73.478	78.210
16	14.057	36.390	42.305	48.202	51.144	55.876
15	-1.626	20.708	24.987	30.074	33.604	38.558
14	-11.064	11.269	14.780	19.418	23.280	28.351
12	-11.493	10.840	14.907	19.877	23.494	28.479
13	-15.197	7.136	11.415	16.502	20.033	24.987

열에 음수 값이 있는 경우 평균이 최소값보다 상당히 큽니다.

"최대값 및 최소값을 사용한 비교" 보고서에서는 각 수준의 평균을 나머지 수준의 최대 및 최소 평균과 비교합니다. 예를 들어 15세의 평균은 나머지 수준의 최대 평균과 유의하게 다릅니다. 17

세의 평균은 나머지 수준의 최소 평균과 유의하게 다릅니다. 최대 평균은 16 세 또는 17 세에서 나타날 수 있습니다. 두 평균 모두 최대 평균과 유의하게 다르지 않기 때문입니다. 동일한 이유로 최소 평균은 17 세를 제외한 다른 연령에 해당할 수 있습니다.

대조군 비교 (Dunnett) 의 예

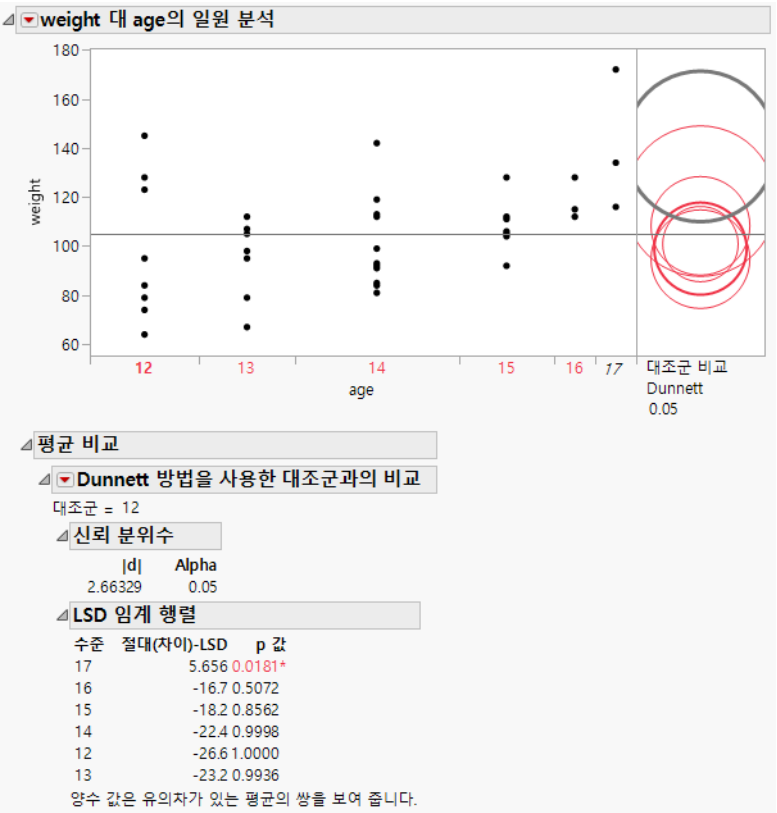
이 예에서는 일원 분석 플랫폼을 사용하여 대조군 비교 (Dunnett) 검정을 보여 줍니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 비교 > 대조군 비교 (Dunnett)** 를 선택합니다 .
7. 대조군으로 사용할 연령을 12 로 선택합니다 .

   팁 : 대조군의 점을 클릭하여 산점도에서 강조 표시한 후 **평균 비교 > 대조군 비교 (Dunnett)** 옵션을 선택합니다 . 이 검정에서는 선택된 그룹이 대조군으로 사용됩니다 .

8. **확인**을 클릭합니다 .

그림 6.19 대조군 비교 (Dunnett) 비교 원의 예



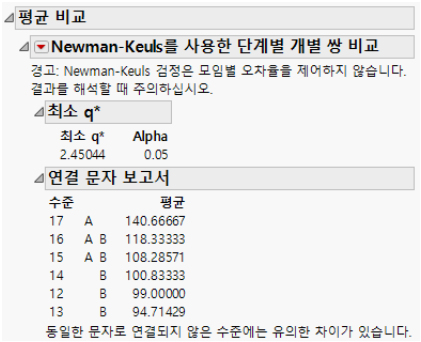
비교 원 또는 "LSD 임계 행렬"의 결과를 사용하면 수준 "17"이 대조 수준 "12"와 유의하게 다른 유일한 수준이라는 결론을 내릴 수 있습니다.

단계별 개별 쌍 비교 (Newman-Keuls) 검정의 예

이 예에서는 일원 분석 플랫폼을 사용하여 Newman-Keuls 검정을 보여 줍니다.

1. **도움말 > 샘플 데이터 폴더**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. "일원 분석"의 빨간색 삼각형 메뉴를 클릭하고 **평균 비교 > 단계별 개별 쌍 비교(Newman-Keuls)**를 선택합니다.

그림 6.20 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 모임별 오차율을 제어하지 않습니다. 결과를 해석할 때 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

이 예에서는 일원 분석 검정을 사용하여 서로 다른 네 가지 평균 비교 검정을 비교합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 비교** 옵션을 각각 선택합니다 . " 대조군 비교 (Dunnett)" 옵션의 경우 대조군으로 17 세를 선택합니다 .

4가지 방법은 모두 그룹 평균 간의 차이를 검정합니다. 각 검정은 특정 가설에 사용되며 서로 다른 사실을 확인할 수 있습니다.

그림 6.21 4 가지 다중 비교 검정의 비교 원

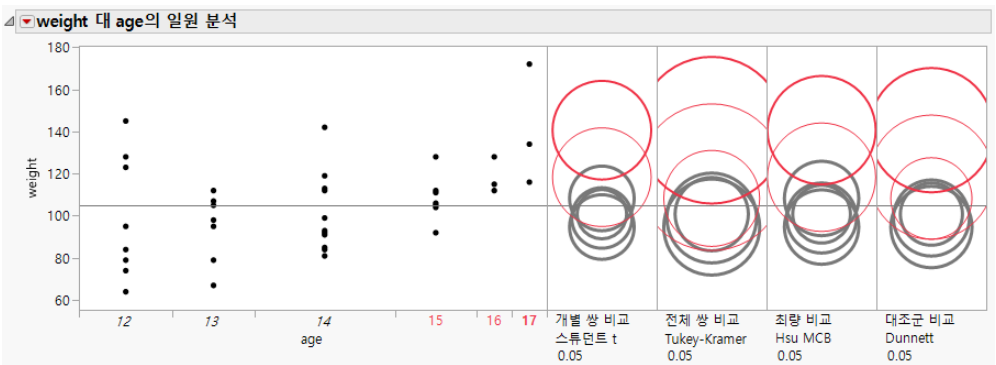


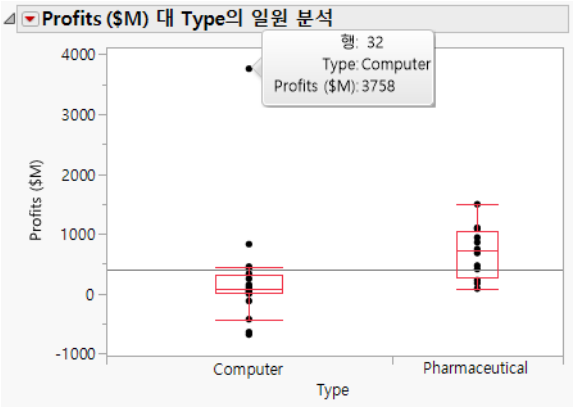
그림 6.21 에서는 17 세 그룹이 강조 표시되어 있습니다. 다른 대조군 원은 17 세 그룹과의 관계에 따라 색상이 적용되어 있습니다. 스튜던트 t 및 Hsu 방법의 경우에는 15 세 그룹 (위에서 세 번째 원) 이 회색입니다. 이는 이 그룹이 17 세 그룹과 유의하게 다름을 나타냅니다. 하지만 Tukey 및 Dunnett 방법의 경우에는 15 세 그룹이 빨간색으로, 17 세 그룹과 유의하게 다르지 않음을 나타냅니다.

비모수 Wilcoxon 검정의 예

일원 분석 플랫폼에서 Wilcoxon 검정을 사용하여 회사의 평균 수익이 회사 유형에 따라 다른지 여부를 확인합니다. 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 폴더**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. " 일원 분석 " 의 빨간색 삼각형을 클릭하고 **표시 옵션 > 상자 그림**을 선택합니다.

그림 6.22 컴퓨터 회사의 수익 분포



상자 그림을 보면 그룹 내의 수익 분포가 정규 분포를 따르거나 대칭이 아님을 알 수 있습니다. 32 행 회사의 경우에는 모수 검정에 영향을 줄 수 있는 매우 큰 값이 있습니다.

7. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **비모수 > Wilcoxon/Kruskal-Wallis 검정**을 선택합니다.

그림 6.23 Wilcoxon 검정 결과

Wilcoxon/Kruskal-Wallis 검정(순위 합)

수준	개수	스코어 합	기대 스코어	스코어 평균	(평균-평균0)/표준0
Computer	20	245.000	330.000	12.2500	-3.289
Pharmaceutical	12	283.000	198.000	23.5833	3.289

Wilcoxon 2표본 검정, 정규 근사

S	Z	Prob>|Z|
283	3.28916	0.0010*

Kruskal-Wallis 검정, 카이제곱 근사

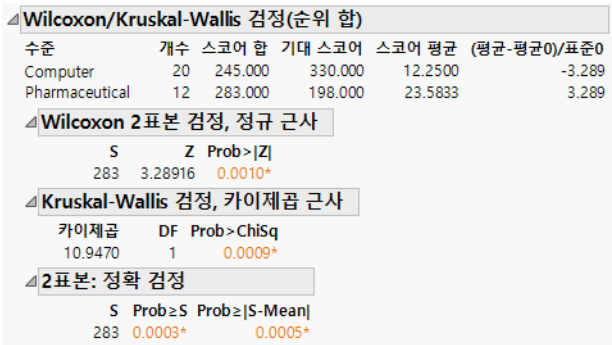
카이제곱	DF	Prob>ChiSq
10.9470	1	0.0009*

0.5 연속성 수정을 사용하는 정규 근사와 Wilcoxon 검정 통계량에 대한 카이제곱 근사는 모두 p 값 0.0010 에서 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 . 정확 검정도 수행할 수 있습니다 .

8. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **비모수 > 정확 검정 > Wilcoxon 정확 검정**을 선택합니다 .

그림 6.24 Wilcoxon 정확 검정 결과



Wilcoxon/Kruskal-Wallis 검정(순위 합)

수준	개수	스코어 합	기대 스코어	스코어 평균	(평균-평균0)/표준0
Computer	20	245.000	330.000	12.2500	-3.289
Pharmaceutical	12	283.000	198.000	23.5833	3.289

Wilcoxon 2표본 검정, 정규 근사

S	Z	Prob>|Z|
283	3.28916	0.0010*

Kruskal-Wallis 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 0.0005 입니다 . 이는 위치의 차이에 대한 양측 검정으로 , 회사 유형별로 수익이 다르지 않다는 가설을 기각할 것을 지지합니다 .

이 예에서는 비모수 검정이 정규성 기반 ANOVA 검정 또는 이분산 t- 검정보다 더 적절합니다 . 비모수 검정은 32 행에 있는 큰 값의 영향을 받지 않으며 정규성 가정이 필요하지 않습니다 .

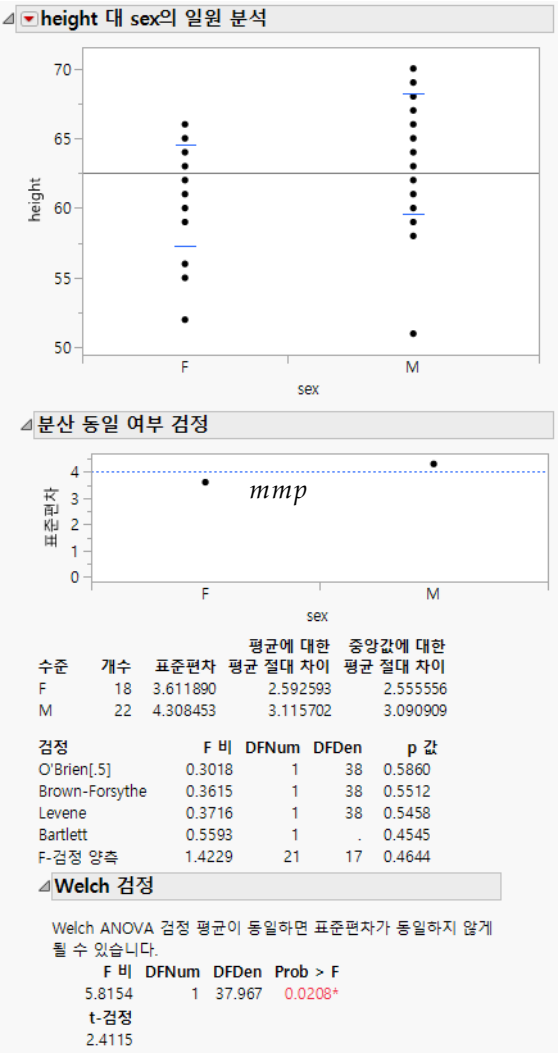
이분산 옵션의 예

일원 분석 플랫폼을 사용하여 두 그룹의 분산이 동일한지 여부를 검정합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **이분산**을 선택합니다 .

그림 6.25 이분산 보고서의 예



수준	개수	표준편차	평균에 대한 평균 절대 차이	중앙값에 대한 평균 절대 차이
F	18	3.611890	2.592593	2.555556
M	22	4.308453	3.115702	3.090909

검정	F 비	DFNum	DFDen	p 값
O'Brien[.5]	0.3018	1	38	0.5860
Brown-Forsythe	0.3615	1	38	0.5512
Levene	0.3716	1	38	0.5458
Bartlett	0.5593	1	.	0.4545
F-검정 양측	1.4229	21	17	0.4644

Welch 검정

Welch ANOVA 검정 평균이 동일하면 표준편차가 동일하지 않게 될 수 있습니다.

F 비	DFNum	DFDen	Prob > F
5.8154	1	37.967	0.0208*

t-검정
2.4115

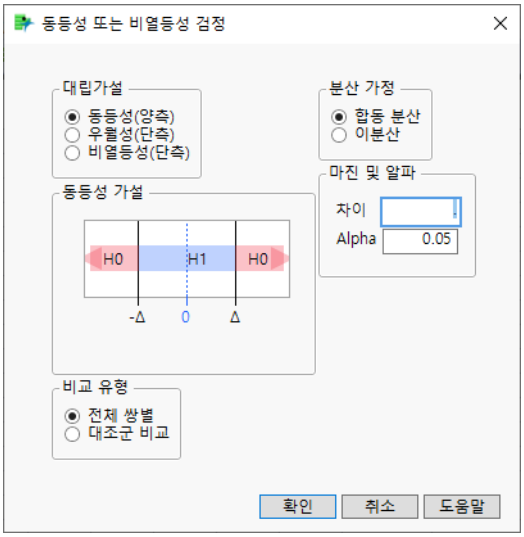
양측 *F*-검정의 *p* 값이 크기 때문에 분산이 동일하지 않다는 결론을 내릴 수 없습니다.

동등성 검정의 예

일원 분석 플랫폼을 사용하여 개 시술에 대해 대체 약물이 모르핀보다 열등하지 않은지 여부를 조사합니다 . 처음 두 시점의 차이를 비교합니다 .

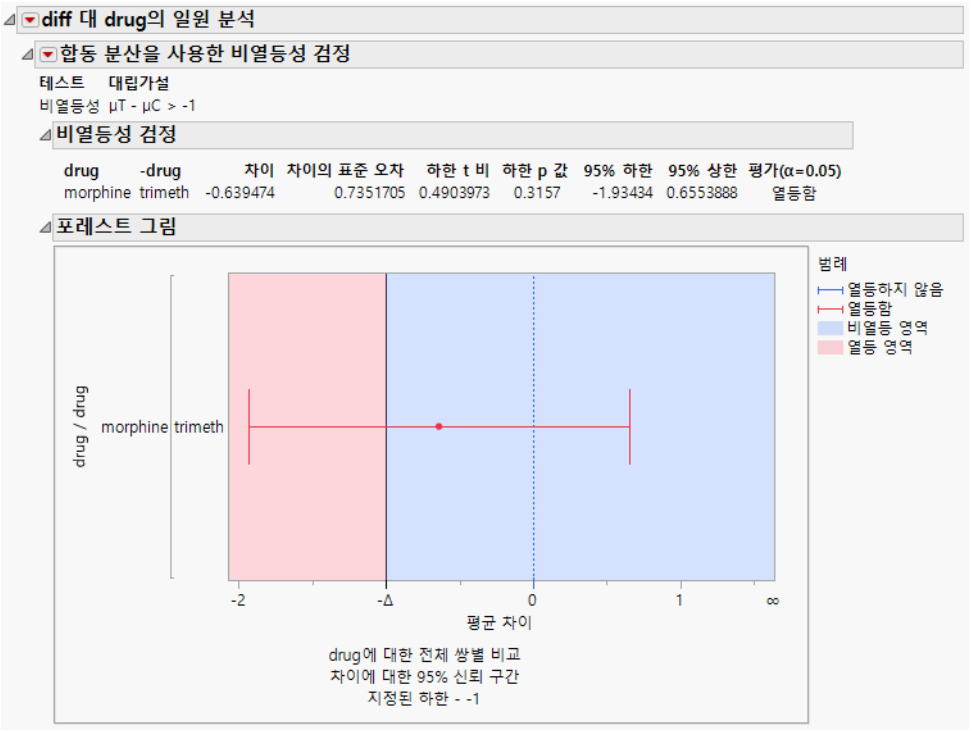
1. **도움말 > 샘플 데이터 폴더**를 선택하고 Dogs.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. diff 를 선택하고 **Y, 반응**을 클릭합니다 .
4. drug 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **동등성 검정 > 평균**을 선택합니다 .

그림 6.26 동등성 또는 비열등성 검정 대화상자



7. **비열등성 (단측)** 을 선택합니다 .
8. 분산 가정을 **합동 분산**으로 설정된 상태 그대로 둡니다 .
9. **차이**를 1 로 설정합니다 . 이 값은 비열등성 한계입니다 .
10. **확인**을 클릭합니다 .

그림 6.27 비열등성 검정의 예



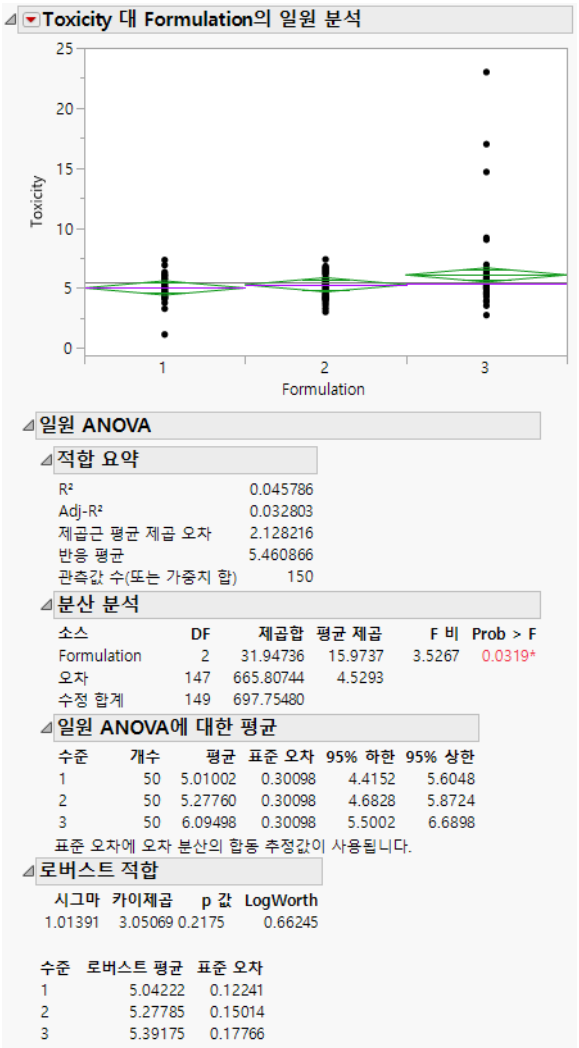
두 약물에 대해 관측된 반응 차이는 -0.639 단위였습니다. 신뢰 구간의 하한 -1.93은 비열등성 하한 -1.0보다 낮습니다. 따라서 대체 약물이 표준 약물보다 열등하지 않다는 결론을 내릴 수 없습니다.

로버스트 적합 옵션의 예

세 그룹 중 하나에 이상치가 포함된 이 예에서는 일원 분석 플랫폼의 로버스트 적합을 사용합니다. 데이터에는 세 가지 약물 제제의 독성 수준이 포함되어 있습니다.

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 /ANOVA** 를 선택합니다.
7. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **로버스트 > 로버스트 적합**을 선택합니다.

그림 6.28 로버스트 적합의 예



표준 "분산 분석" 보고서를 보면 p 값이 0.0319이므로 세 가지 제제 사이에 차이가 있다는 결론을 내릴 수 있습니다. 하지만 "로버스트 적합" 보고서를 보면 p 값이 0.2175이므로 세 가지 제제가 유의하게 다르다는 결론을 내리지 않습니다. 일부 관측값의 독성은 비정상적으로 높아서 데이터에 과도한 영향을 주는 것으로 보입니다.

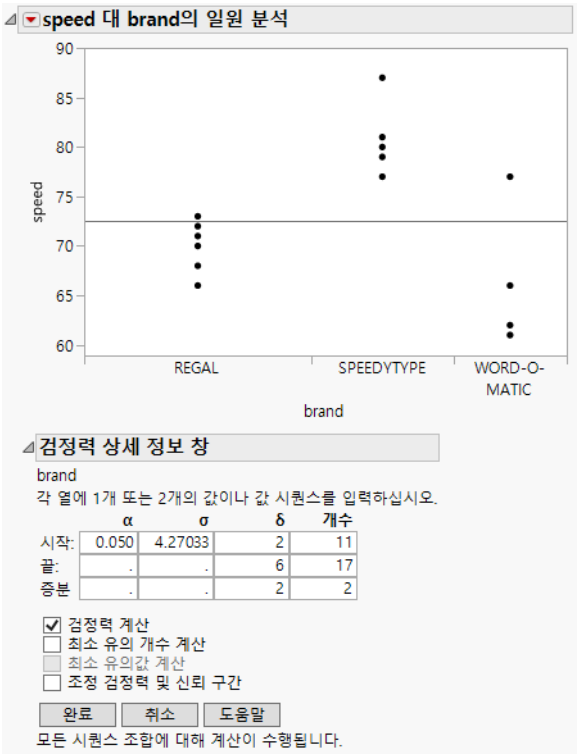
검정력 옵션의 예

이 예에서는 일원 분석 플랫폼의 " 검정력 " 옵션을 보여 줍니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 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 6.29 검정력 상세 정보 창의 예



11. **완료**를 클릭합니다 .

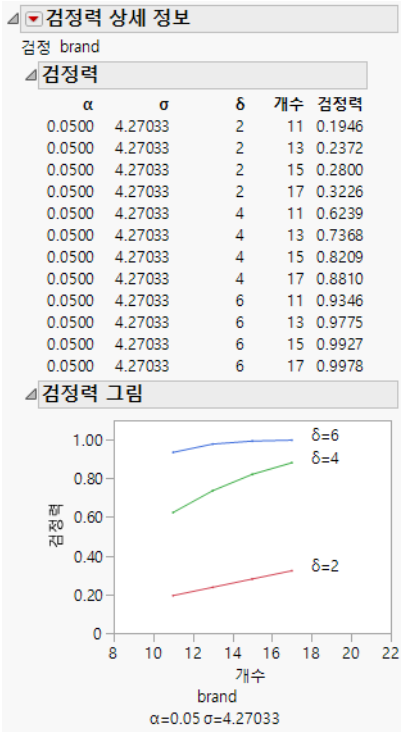
참고 : 필요한 옵션이 모두 적용되기 전까지는 **완료** 버튼이 흐리게 표시됩니다 .

델타 값과 개수의 각 조합에 대해 검정력이 계산되어 "검정력" 보고서에 표시됩니다.

검정력 값을 그림으로 표시하려면 다음을 수행하십시오 .

12. 테이블 아래의 빨간색 삼각형을 클릭하고 **검정력 그림**을 선택합니다 .

그림 6.30 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.

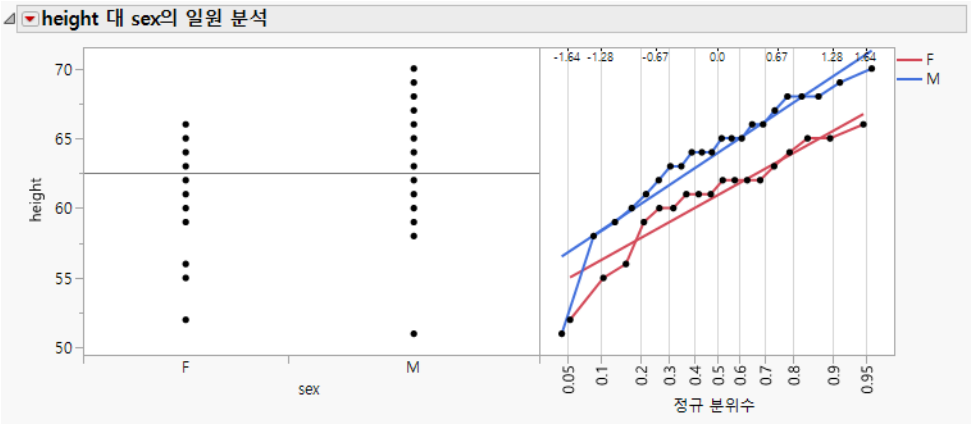
정규 분위수 그림의 예

일원 분석 플랫폼을 사용하여 정규 분위수 그림을 생성합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **정규 분위수 그림 > 분위수별 실제값 그림**을 선택합니다 .

그림 6.31 정규 분위수 그림의 예



다음 사항을 알 수 있습니다 .

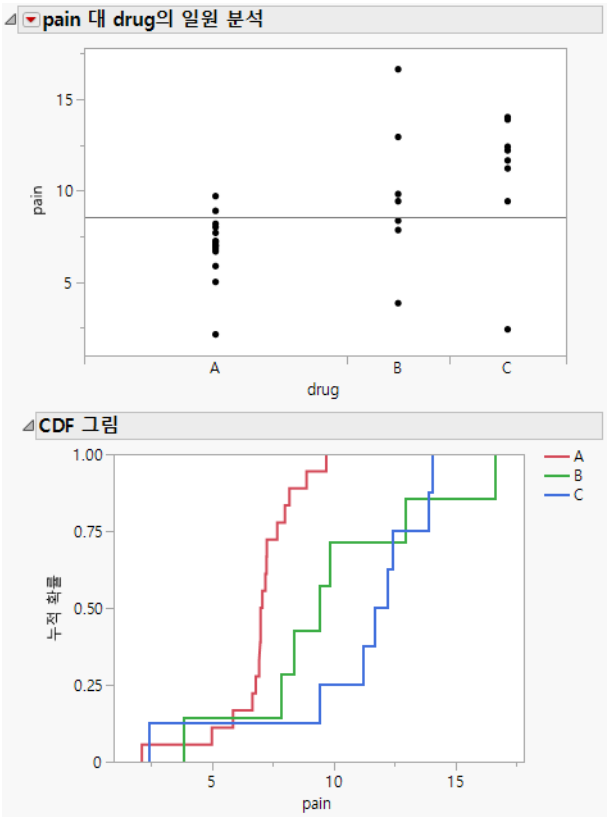
- 기본적으로 적합선이 표시됩니다 .
- 데이터 점은 적합선에 매우 가깝게 늘어서 있어 정규 분포를 나타냅니다 .

CDF 그림의 예

일원 분석 플랫폼을 사용하여 CDF(누적 분포 함수) 그림을 생성합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **CDF 그림**을 선택합니다 .

그림 6.32 CDF 그림의 예



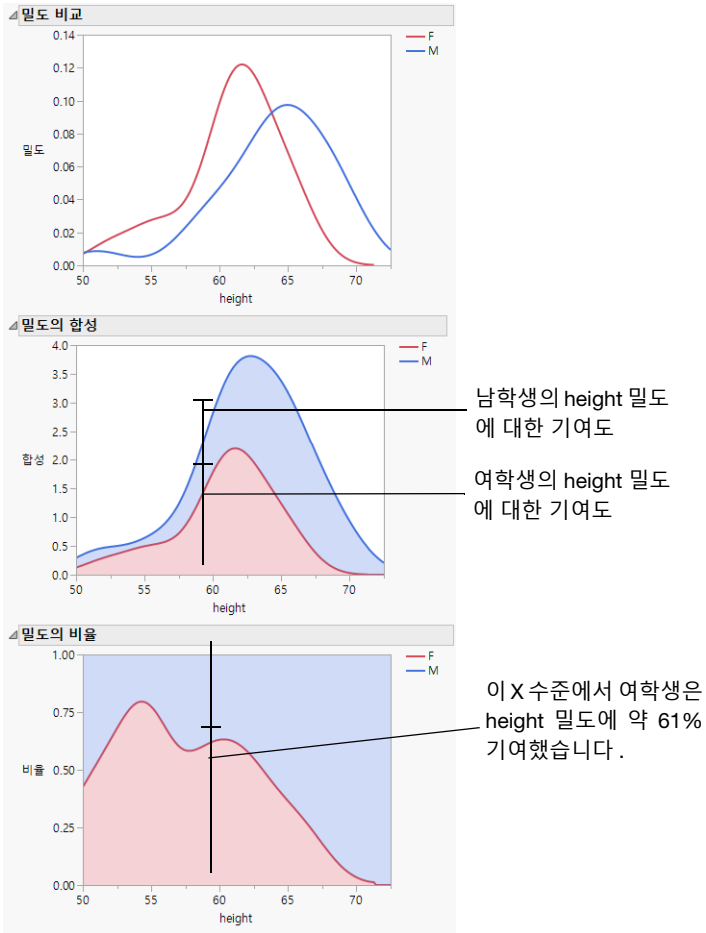
CDF 그림에 X 변수의 각 수준에 대한 곡선이 있습니다.

밀도 옵션의 예

이 예에서는 일원 분석 플랫폼의 " 밀도 " 옵션을 보여 줍니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 세 개의 옵션 , 즉 **밀도 > 밀도 비교** , **밀도 > 밀도의 합성** 및 **밀도 > 밀도의 비율**을 모두 선택합니다 .

그림 6.33 밀도 옵션의 예



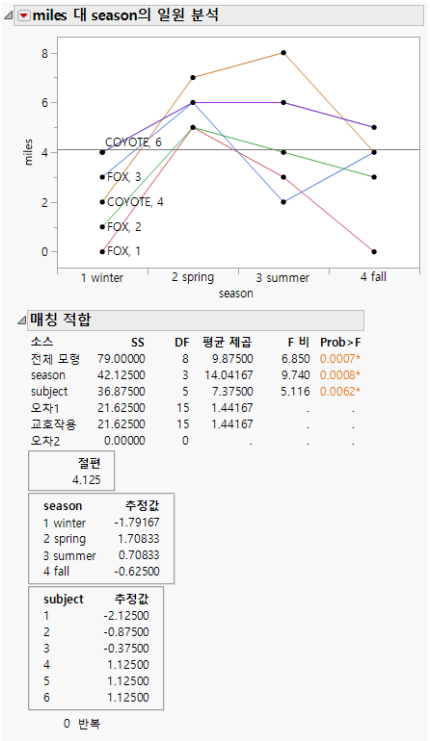
매칭 열 옵션의 예

일원 분석 플랫폼을 사용하여 6 종의 동물에 대한 데이터와 이 동물들이 계절별로 이동한 거리 (마일) 를 분석합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **매칭 열**을 선택합니다 .
7. 매칭 열로 subject 를 선택합니다 .

8. **확인**을 클릭합니다 .

그림 6.34 매칭 열 보고서의 예



이 그림에서는 개체를 매칭 변수로 하여 계절별 이동 거리(마일)를 그래프로 보여 줍니다. 그래프에서 각 개체의 첫 번째 측정값 옆에 표시되는 라벨은 species 및 subject 변수에 따라 결정되었습니다.

"매칭 적합" 보고서에는 *F*-검정을 사용한 season 및 subject 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 **선형 모형 적합**의 에서 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 데이터를 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 테

이블에서는 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 6.35 에서는 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 6.35 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 충진 중량은 12.5온스입니다. 세 개의 생산 라인이 목표를 충족하고 있는지 여부에 관심이 있지만 우선은 세 개의 라인이 동일한 평균 충진 비율을 달성하고 있는지 확인하려고 합니다. 일원 분석을 사용하여 평균 충진 중량 간에 차이가 있는지 검정할 수 있습니다.

일원 분석 플랫폼을 사용하려면 다음을 수행해야 합니다.

1. 데이터를 JMP 로 가져옵니다. 자세한 내용은 " 데이터 가져오기 " 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 " 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 " 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 " 에서 확인하십시오. 데이터 파일이 작은 경우에 이 방법을 사용할 수 있습니다.

Microsoft Excel에서 데이터를 가져오는 방법에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

Excel 가져오기 마법사를 사용하여 데이터 가져오기

1. **도움말 > 샘플 데이터 폴더**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 정보가 3 행에서 시작됩니다 .
3. **열 머리글이 포함된 행 수**에 2 를 입력합니다 .

   Excel 파일의 3 행과 4 행에 모두 열 머리글 정보가 포함됩니다 .
4. **가져오기**를 클릭합니다 .

그림 6.36 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 " 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 붙여넣기**를 선택합니다 .

   편집 > 열 이름과 함께 붙여넣기 옵션은 클립보드의 선택 내용에 열 이름이 포함된 경우에 사용됩니다 .

그림 6.37 " 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 " 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 에서 확인하십시오 .

1. JMP 데이터 테이블에서 **테이블 > 쌓기**를 선택합니다 .
2. 6 개의 열을 모두 선택하고 **열 쌓기**를 클릭합니다 .
3. **다중 계열 쌓기**를 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 쌓기**를 선택 취소합니다 .
5. **결측 행 제거**를 선택합니다 .
6. **출력 테이블 이름** 옆에 Stacked 를 입력합니다 .
7. **확인**을 클릭합니다 .

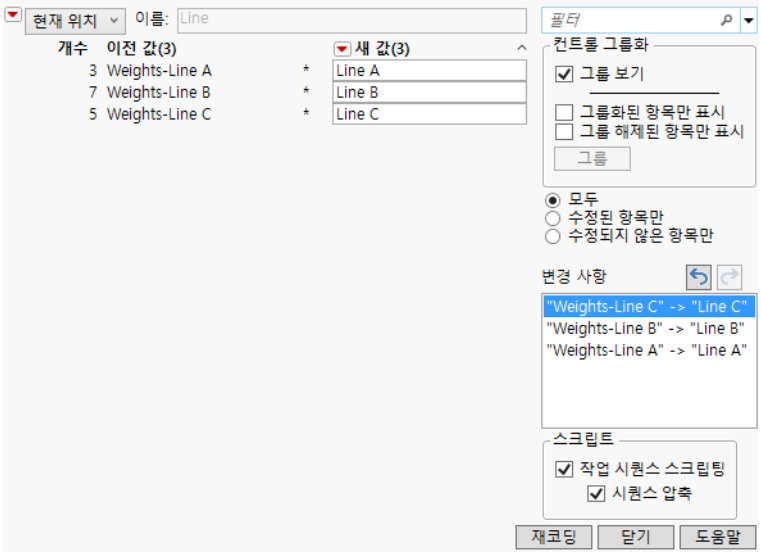
   새 데이터 테이블에서 데이터 및 데이터 2 는 ID 및 중량 데이터가 포함된 열입니다 .
8. 라벨 열 머리글을 마우스 오른쪽 버튼으로 클릭하고 **열 삭제**를 선택합니다 .

   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 .
   - 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 열을 클릭하고 현재 위치를 선택합니다** .
    3. **새 값** 열의 값을 아래의 그림 6.38 에 표시된 것과 매칭되도록 변경합니다 .

그림 6.38 열 값 재코딩



    4. **재코딩**을 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 6.39).

그림 6.39 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 충진 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. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 /ANOVA** 를 선택합니다 .

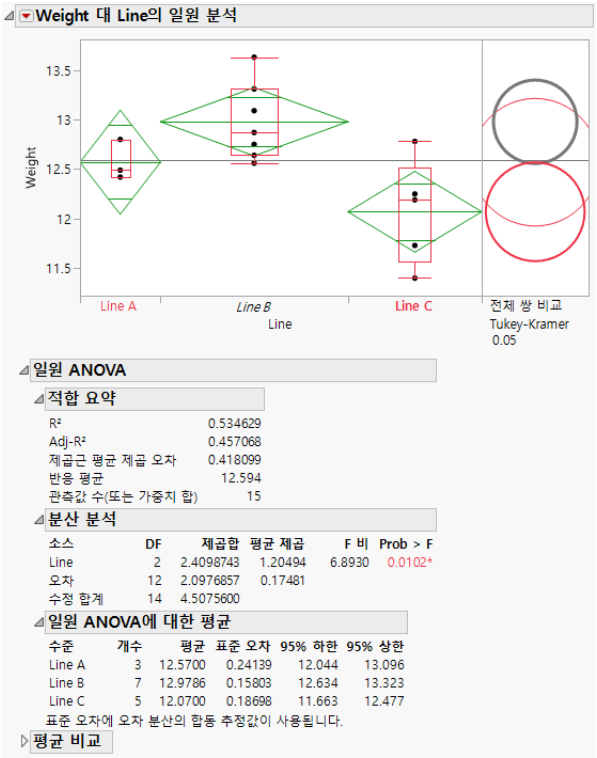
   그림의 평균 다이아몬드는 생산 라인 평균에 대한 95% 신뢰 구간을 보여 줍니다 . 평균 다이아몬드 외부의 점은 이상치인 것처럼 보일 수도 있습니다 . 하지만 그렇지 않습니다 . 이를 확인하기 위해 그림에 상자 그림을 추가합니다 .
6. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **표시 옵션 > 상자 그림**을 선택합니다 .

   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. 그림에서 점을 마우스로 가리키면 해당 점의 ID 값과 함께 Line 및 Weight 데이터가 표시됩니다 (그림 6.40).
9. " 일원 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **평균 비교 > 전체 쌍 비교 (Tukey HSD)** 를 선택합니다 .

   그림 오른쪽의 패널에 비교 원이 표시됩니다 .
10. 맨 아래 비교 원을 클릭합니다 .

그림 6.40 Weight 대 Line 의 일원 분석



"분산 분석" 보고서에서 p 값이 0.0102인 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 충진 중량인 12.5가 포함되지 않았음을 나타냅니다. 라인 B는 중량 초과로 보이고 라인 C는 중량 미달로 보입니다. 이 두 생산 라인의 경우 목표 충진 중량 미충족을 초래한 근본적인 원인을 해결해야 합니다.

일원 분석 플랫폼에 대한 통계 상세 정보

이 섹션에는 일원 분석 플랫폼에 대한 통계 상세 정보가 포함되어 있습니다.

- "비교 원에 대한 통계 상세 정보"
- "검정력에 대한 통계 상세 정보"
- "ANOM 에 대한 통계 상세 정보"
- "적합 요약 보고서에 대한 통계 상세 정보"
- "분산 동일 여부 검정에 대한 통계 상세 정보"
- "비모수 검정 통계량에 대한 통계 상세 정보"
- "로버스트 적합에 대한 통계 상세 정보"

비교 원에 대한 통계 상세 정보

일원 분석 플랫폼에서 비교 원은 다중 비교 검정의 LSD(최소 유의차)를 그래픽으로 표현한 것입니다. 이 최소 유의차는 확률 분위수에 두 평균의 차이에 대한 표준 오차를 곱한 값입니다. "개별 쌍 비교" 옵션의 경우 Fisher LSD가 사용되며 확률 분위수는 스튜던트 *t* 통계량입니다. 이 경우에 대한 비교 원 계산을 보여 줍니다. LSD는 다음과 같이 정의됩니다.

$$\text{LSD} = t_{\alpha/2}\text{std}(\hat{\mu}_1 - \hat{\mu}_2)$$

두 독립 평균의 차이에 대한 표준 오차는 다음 관계식으로 계산됩니다.

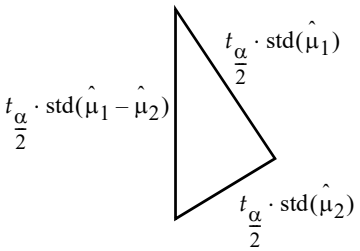
$$[\text{std}(\hat{\mu}_1 - \hat{\mu}_2)]^2 = [\text{std}(\hat{\mu}_1)]^2 + [\text{std}(\hat{\mu}_2)]^2$$

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 6.41에 그래픽으로 표시된 직각 삼각형의 경우처럼 피타고라스 관계식이 성립됩니다.

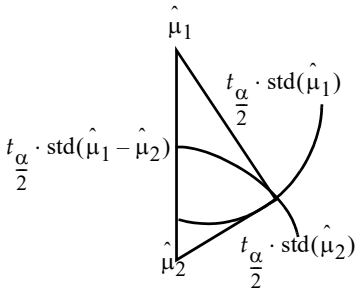
그림 6.41 두 평균 간의 차이 관계



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 6.42 *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 90 도 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 6.9).

스튜던트 *t* 대신 다른 확률 분위수 값을 사용하여 다양한 다중 비교 검정에 이와 동일한 그래픽 기법을 적용할 수 있습니다.

검정력에 대한 통계 상세 정보

일원 분석 플랫폼에서는 비중심 *F* 분포를 사용하여 검정력을 계산합니다. 계산식 (O'Brien and Lohr 1984) 은 다음과 같이 정의됩니다.

검정력 = $\text{Prob}(F > F_{crit}, \nu_1, \nu_2, nc)$

다음은 각 요소에 대한 설명입니다.

F : 비중심 $F(nc, \nu_1, \nu_2)$ 분포를 따르며, $F_{crit} = F_{(1-\alpha, \nu1, \nu2)}$ 는 자유도가 ν_1 및 ν_2 인 *F* 분포의 1 - α

분위수입니다 .

$\nu_1 = r - 1$ 은 분자 DF 입니다 .

$\nu_2 = r(n - 1)$ 은 분모 DF 입니다 .

n : 그룹당 개수입니다 .

r : 그룹 수입니다 .

$nc = n(CSS)/\sigma^2$: 비중심성 모수입니다 .

$CSS = \sum_{g=1}^{r} (\mu_g - \mu)^2$: 수정 제곱합입니다 .

μ_g : g 번째 그룹의 평균입니다 .

μ : 전체 평균입니다 .

σ^2 : MSE(평균 제곱 오차) 를 사용하여 추정됩니다 .

ANOM 에 대한 통계 상세 정보

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 $R_1, R_2, \ldots, R_n$ 으로 나타냅니다 .
- i 번째 관측값에 해당하는 변환 순위는 다음과 같이 결정됩니다 .

$$\text{변환된 } R_i = Normal\ Quantile\left[\left(\frac{R_i}{2n+1}\right) + 0.5\right]$$

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 .

적합 요약 보고서에 대한 통계 상세 정보

이 섹션에는 일원 분석 플랫폼의 " 적합 요약 " 보고서에 대한 상세 정보가 포함되어 있습니다 .

R^2

모형에 대한 분산 분석 보고서의 통계량을 사용하는 경우 연속형 반응 적합에 대한 R^2 는 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 합계**의 평균 제곱은 **수정 합계** 제곱합을 해당 자유도로 나눠서 계산할 수 있습니다 . 자세한 내용은 "분산 분석"에서 확인하십시오.

분산 동일 여부 검정에 대한 통계 상세 정보

이 섹션에는 일원 분석 플랫폼의 " 분산 동일 여부 검정 " 보고서에 대한 상세 정보가 포함되어 있습니다 .

F 비

O′Brien 검정에서는 새 변수의 그룹 평균이 원래 반응의 그룹 표본 분산과 동일하도록 종속 변수를 생성합니다 . O′Brien 변수는 다음과 같이 계산됩니다 .

$$r_{ijk} = \frac{(n_{ij} - 1.5)n_{ij}(y_{ijk} - \overline{y_{ij}})^2 - 0.5s_{ij}^2(n_{ij} - 1)}{(n_{ij} - 1)(n_{ij} - 2)}$$

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 $z_{ij} = |y_{ij} - \overline{y_{i.}}|$ 에 대한 ANOVA 에서 구한 모형 F 통계량이며 , 여기서 $\overline{y_{i.}}$ 는 i 번째 수준의 평균 반응입니다 .

Bartlett 검정은 다음과 같이 계산됩니다 .

$$T = \frac{v\log\left(\sum_i \frac{v_i}{v} s_i^2\right) - \sum_i v_i \log(s_i^2)}{1 + \left(\frac{\sum_i \frac{1}{v_i} - \frac{1}{v}}{3(k-1)}\right)} \text{ 단 , } v_i = n_i - 1 \text{ 및 } v = \sum_i v_i$$

여기서 n_i는 i번째 수준의 관측값 수이고, s_i^2은 i번째 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 .

$$F = \frac{\left[\frac{\sum_i w_i(\bar{y}_i - \tilde{y}_{..})^2}{k-1}\right]}{\left\{1 + \frac{2(k-2)}{k^2-1}\left[\sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i - 1}\right]\right\}} \text{ 단 , } w_i = \frac{n_i}{s_i^2},\ u = \sum_i w_i,\ \tilde{y}_{..} = \sum_i \frac{w_i \bar{y}_{i.}}{u},$$

여기서 n_i는 i번째 수준의 관측값 수이고, $\bar{y}_{i.}$ 는 i번째 수준의 평균 반응이며, s_i^2은 i번째 수준의 반응 표본 분산입니다.

Welch 검정 DF Den

분모 자유도에 대한 Welch 근사 계산식은 다음과 같이 정의됩니다 .

$$df = \frac{1}{\left(\frac{3}{k^2-1}\right)\left[\sum_i \frac{\left(1 - \frac{w_i}{u}\right)^2}{n_i - 1}\right]}$$

여기서 w_i, n_i 및 u는 F 비 계산식에서의 정의와 같습니다.

비모수 검정 통계량에 대한 통계 상세 정보

이 섹션에서는 일원 분석 플랫폼의 Wilcoxon, 중앙값, van der Waerden 및 Friedman 순위 검정에 사용되는 검정 통계량의 계산식을 제공합니다.

표기

언급된 검정은 스코어를 기반으로 하며 다음과 같은 표기를 사용합니다 .

$j = 1,\ldots, n$ 전체 표본의 관측값입니다 .

$i = 1,\ldots, k$ X 의 수준입니다 . 여기서 k 는 총 수준 수입니다 .

$n_1, n_2,\ldots, n_k$ X 의 k 개 수준 각각에 포함된 관측값의 개수입니다 .

R_j j 번째 관측값의 중간 순위입니다 . **중간 순위**는 해당 관측값의 순위 (동일 점수가 없는 경우) 이거나 평균 순위 (동일 점수가 있는 경우) 입니다 .

α 다양한 검정에서 스코어를 정의하는 데 사용되는 중간 순위의 함수입니다 .

시작 창에서 블록 변수를 지정한 경우에는 다음 표기가 사용됩니다 .

$b = 1,\ldots, B$ 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 스코어를 다음과 같이 정의합니다 .

Wilcoxon 스코어

$$\alpha(R_j) = R_j$$

중앙값 스코어

$$\alpha(R_j) = \begin{cases} 1 \text{ 단, } R_j > \text{중앙값} \\ 0 \text{ if } R_j < \text{중앙값} \\ t \text{ 단, } R_j = \text{중앙값} \end{cases}$$

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 스코어

$$\alpha(R_j) = \text{표준 정규 분위수}(R_j/(n+1))$$

Friedman 순위 스코어

$$\alpha(R_{bi}) = R_{bi}$$

2 표본 정규 근사

정규 근사를 기반으로 하는 검정은 X 의 수준이 정확히 두 개뿐인 경우에만 가능합니다 . 이 섹션에서 사용되는 표기는 " 표기 " 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 정의됩니다 .

$$Z = (S - E(S))/\sqrt{Var(S)}$$

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 .

$$E(S) = \frac{n_l}{n}\sum_{j=1}^{n}\alpha(R_j)$$

***Var*(*S*)** 모든 관측값의 평균 스코어를 *ave* 라고 정의할 경우 S 의 분산은 다음과 같이 정의됩니다 .

$$Var(S) = \frac{n_1 n_2}{n(n-1)}\sum_{j=1}^{n}(\alpha(R_j) - ave)^2$$

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 분산은 다음과 같이 계산됩니다 .

$$Var(S) = \frac{B}{(n-1)}\sum_{j=1}^{n}(\alpha(R_j) - ave)^2$$

일원 카이제곱 근사

참고 : Wilcoxon 스코어를 기반으로 하는 카이제곱 검정은 Kruskal-Wallis 검정이라고 합니다 .

이 섹션에서 사용되는 표기는 " 표기 " 에 정의되어 있습니다 . 카이제곱 통계량을 계산하는 데는 다음과 같은 통계량이 사용됩니다 .

T_i X 의 *i* 번째 수준에 대한 총 스코어입니다 .

$E(T_i)$ 수준 간에 차이가 없다는 귀무가설하에 수준 i 의 총 스코어에 대한 기대값으로 , 다음과 같이 정의됩니다 .

$$E(T_i) = \frac{n_i}{n} \sum_{j=1}^{n} \alpha(R_j)$$

Var(T) 모든 관측값의 평균 스코어를 *ave* 라고 정의할 경우 T 의 분산은 다음과 같이 정의됩니다.

$$Var(T) = \frac{1}{(n-1)} \sum_{j=1}^{n} (\alpha(R_j) - ave)^2$$

이 검정 통계량의 값은 아래와 같이 구합니다 . 이 통계량은 자유도가 $k-1$ 인 점근적 카이제곱 분포를 따릅니다 .

$$C = \left(\sum_{i=1}^{k} (T_i - E(T_i))^2 / n_i \right) / Var(T)$$

Friedman 순위 검정에 대한 일원 카이제곱 근사

Friedman 순위 검정에 대한 카이제곱 검정 통계량은 다음과 같이 계산됩니다 .

$$C = \frac{\sum_{i=1}^{k} (T_i - E(T_i))^2 / n_i}{\frac{1}{(k-1)} \sum_{j=1}^{n} (\alpha(R_j) - ave)^2 / n_i}$$

로버스트 적합에 대한 통계 상세 정보

일원 분석 플랫폼의 " 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 .

$$l(\underline{e}) = \sum_{i} \rho(e_i)$$

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 2차 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"에서 확인하십시오.

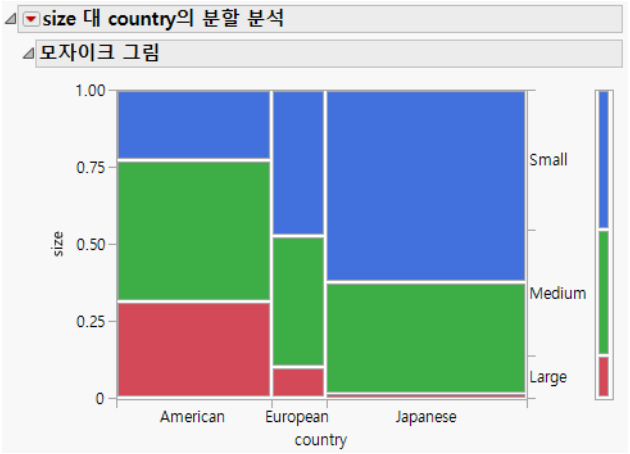
7 장

분할 분석
두 범주형 변수 간의 관계 분석

분할 분석 플랫폼을 사용하여 두 범주형 변수 사이의 관계를 조사합니다. 범주형 변수는 순서형 또는 명목형일 수 있습니다. 분석 결과에는 모자이크 그림, 빈도 수 및 비율의 분할표, 유의성 카이제곱 검정이 포함됩니다. 비율에 대한 평균 분석, 대응 분석, 연관성 측도와 같은 추가 데이터 분석 및 검정을 대화식으로 수행할 수 있습니다.

분할 분석 플랫폼은 X로 Y 적합 플랫폼의 **범주형 대 범주형** 분석법입니다.

그림 7.1 분할 분석의 예



목차

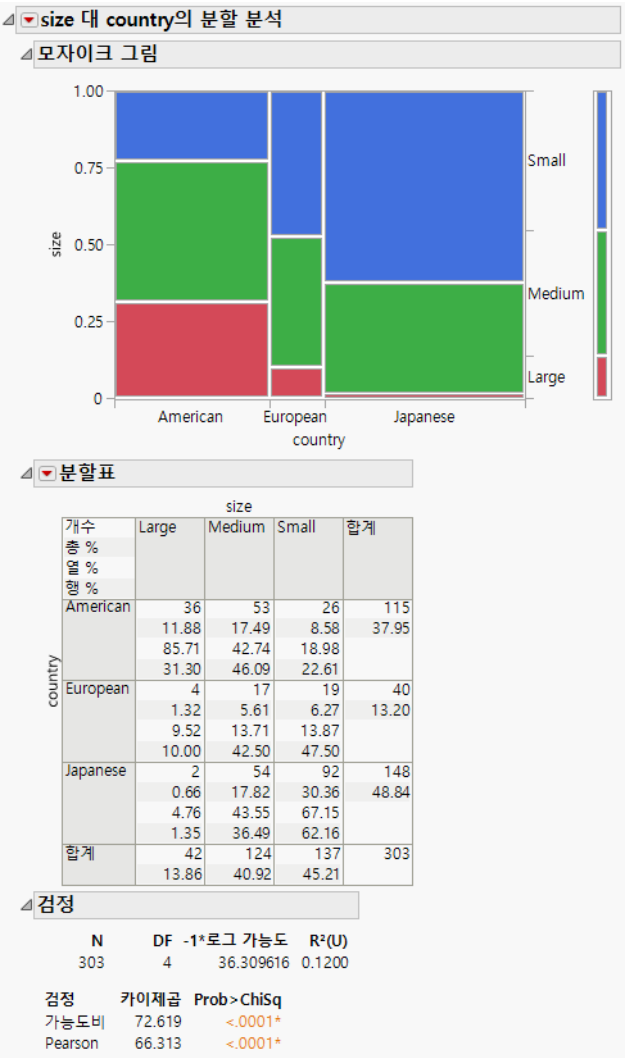
분할 분석의 예 213
분할 분석 플랫폼 시작 215
데이터 형식 215
분할 분석 보고서 216
모자이크 그림 217
분할표 219
검정 보고서 221
분할 분석 플랫폼 옵션 222
분할 분석 보고서 225
비율에 대한 평균 분석 보고서 225
대응 분석 보고서 226
Cochran-Mantel-Haenszel 검정 보고서 227
합치도 통계량 보고서 227
상대 위험도 보고서 228
비율에 대한 2표본 검정 보고서 228
연관성 측도 보고서 229
Cochran Armitage 추세 검정 보고서 230
Fisher 정확 검정 보고서 230
동등성 검정 보고서 230
분할 분석 플랫폼의 추가 예 232
비율에 대한 평균 분석의 예 233
대응 분석의 예 234
Cochran-Mantel-Haenszel 검정의 예 235
합치도 통계량 옵션의 예 236
상대 위험도 옵션의 예 237
비율에 대한 2표본 검정의 예 239
연관성 측도 옵션의 예 240
Cochran Armitage 추세 검정의 예 241
분할 분석 플랫폼에 대한 통계 상세 정보 242
합치도 통계량에 대한 통계 상세 정보 242
승산비에 대한 통계 상세 정보 243
검정 보고서에 대한 통계 상세 정보 243
대응 분석에 대한 통계 상세 정보 244

분할 분석의 예

분할 분석 플랫폼을 사용하여 두 범주형 변수 사이의 관계를 조사합니다 . 이 예에서는 자동차 소유권에 대한 설문 조사에서 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 폴더**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



country	개수 / 총 % / 열 % / 행 %	Large	Medium	Small	합계
American		36	53	26	115
		11.88	17.49	8.58	37.95
		85.71	42.74	18.98	
		31.30	46.09	22.61	
European		4	17	19	40
		1.32	5.61	6.27	13.20
		9.52	13.71	13.87	
		10.00	42.50	47.50	
Japanese		2	54	92	148
		0.66	17.82	30.36	48.84
		4.76	43.55	67.15	
		1.35	36.49	62.16	
합계		42	124	137	303
		13.86	40.92	45.21	

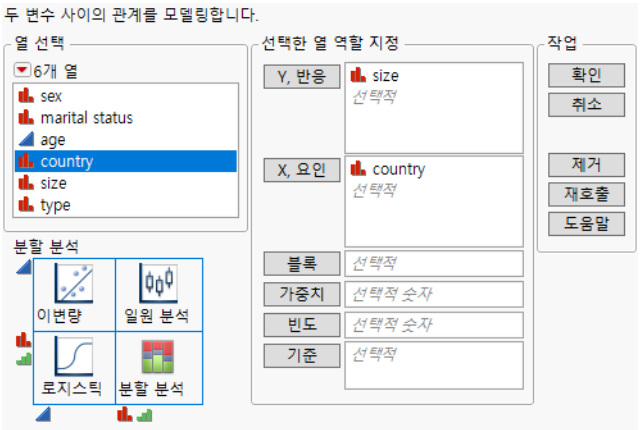
모자이크 그림 및 범례에서 다음을 확인할 수 있습니다.

- 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

분석 > X 로 Y 적합을 선택하여 분할 분석 플랫폼을 시작할 수 있습니다 . X 로 Y 적합 시작 창은 네 가지 유형의 분석에 사용됩니다 . 순서형 또는 명목형 Y 변수와 순서형 또는 명목형 X 변수를 입력하면 분할 분석 플랫폼이 시작됩니다 .

그림 7.3 분할 분석 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 . 분할 분석 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 . 이러한 변수의 모델링 유형은 순서형 또는 명목형이어야 합니다 .

X, 요인 분석할 하나 이상의 예측 변수입니다 . 이러한 변수의 모델링 유형은 순서형 또는 명목형이어야 합니다 .

블록 블록 변수를 지정하는 열입니다 . 블록 변수를 지정하면 두 번째 요인이 식별되고 Cochran-Mantel-Haenszel 검정이 수행됩니다. 추가 분류 변수의 블록화를 층화라고도 합니다 .

가중치 데이터 테이블의 각 관측값에 대한 가중치를 포함하는 열입니다 . 값이 0 보다 큰 행만 분석에 포함됩니다 .

빈도 분석의 각 행에 빈도를 할당합니다 . 빈도 할당은 데이터를 요약할 때 유용합니다 .

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

데이터 형식

분할 분석 플랫폼에서 데이터는 요약되지 않거나 요약된 데이터로 구성될 수 있습니다 .

요약되지 않은 데이터 각 관측값에 대한 하나의 행과 X 및 Y 값에 대한 열이 있습니다 .

요약된 데이터 각 행이 공통된 X 및 Y 값을 갖는 일련의 관측값을 나타냅니다 . 데이터 테이블에는 각 행에 대한 관측값 수가 포함된 빈도 열이 있어야 합니다 . 시작 창에서는 이 열을 " 빈도 " 열로 입력합니다 .

요약된 범주형 데이터의 예는 "비율에 대한 평균 분석의 예"에서 확인하십시오.

참고 : X 로 Y 적합 시작 창은 연속형 , 순서형 및 명목형 모델링 유형의 열을 허용합니다 . 순서형 또는 명목형 모델링 유형을 사용하는 Y, 반응 열과 X, 요인 열의 모든 쌍에 대해 분할 분석 플랫폼이 시작됩니다 . X 로 Y 적합 시작 창은 다른 열 유형 조합에 대해 이변량 , 일원 분석 또는 로지스틱 플랫폼을 실행합니다 .

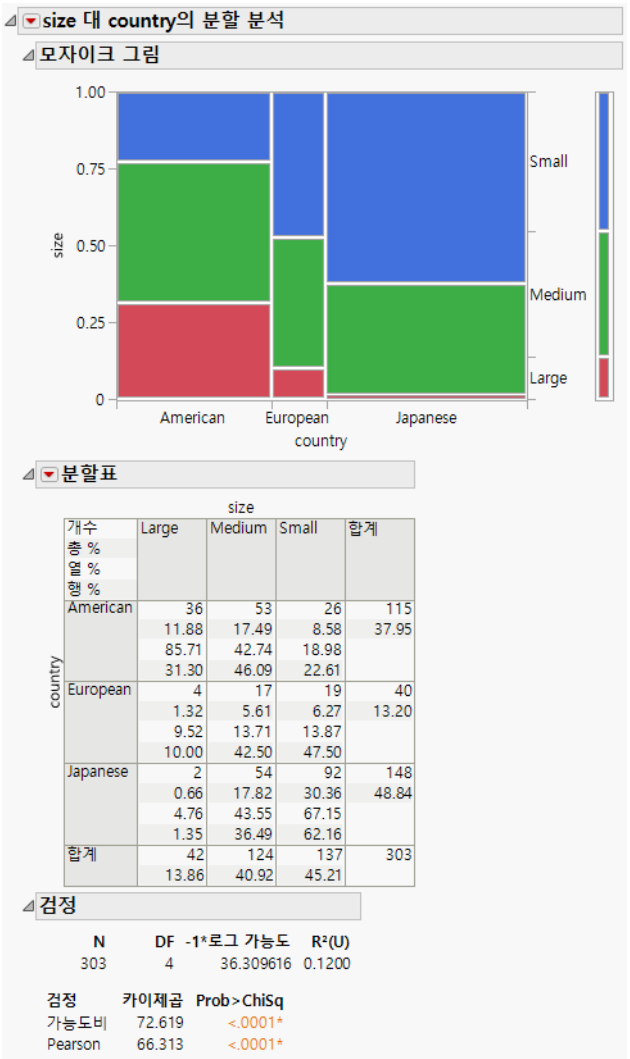
분할 분석 보고서

처음에는 " 분할 분석 " 보고서에 모자이크 그림 , 분할표 및 Y 반응 변수의 수준이 X 요인 변수의 수준에 따라 다른지 여부를 판별하는 검정이 포함됩니다 . 빨간색 삼각형 메뉴 옵션을 사용하여 추가 분석 및 검정을 실행할 수 있습니다 . 자세한 내용은 "분할 분석 플랫폼 옵션"에서 확인하십시오.

이 섹션에는 " 분할 분석 " 보고서의 다음 섹션에 대한 정보가 포함되어 있습니다 .

- " 모자이크 그림 "
- " 분할표 "
- " 검정 보고서 "

그림 7.4 분할 분석 보고서의 예



대화식으로 변수 바꾸기

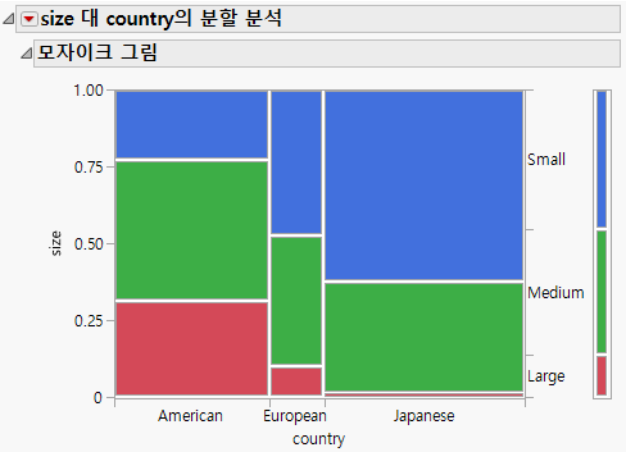
그림에서 한 축의 변수를 다른 축으로 드래그한 후 놓는 방법으로 변수를 대화식으로 바꿀 수 있습니다. 연결된 데이터 테이블의 "열" 패널에서 변수를 선택하여 축으로 드래그하는 방법으로 변수를 바꿀 수도 있습니다.

모자이크 그림

분할 분석 플랫폼에서 모자이크 그림은 분할표라고도 하는 이원 빈도 테이블을 그래픽으로 표현한 것입니다. 모자이크 그림은 다양한 크기의 직사각형으로 나뉘며, 각 직사각형의 세로 길이

는 X 변수의 각 수준 내에서 Y 변수의 비율에 비례합니다 . 모자이크 그림은 Hartigan and Kleiner 연구 자료 (1981) 에서 처음 소개되었으며 Friendly 연구 자료 (1994) 에서 구체화되었습니다 .

그림 7.5 모자이크 그림의 예



이 모자이크 그림의 경우 다음 사항에 유의하십시오 .

- 가로 축의 분할 너비는 X 변수의 수준별 관측값 수를 나타냅니다 .
- 그림 오른쪽에 있는 세로 축의 비율은 X 변수의 결합 수준에 대한 Y 변수 수준의 전체 비율을 나타냅니다 . 이러한 비율은 X 변수와 Y 변수 간에 연관성이 없다는 귀무가설을 나타냅니다 .
- 그림 왼쪽에 있는 세로 축의 척도는 반응 확률을 나타냅니다 . 전체 축은 총 표본을 나타내는 확률 1 과 동일합니다 .

팁 : 모자이크 그림의 직사각형을 클릭하여 선택 영역을 강조 표시하고 연결된 데이터 테이블의 해당 행을 강조 표시할 수 있습니다 . 직사각형을 마우스 오른쪽 버튼으로 클릭하여 그림에 라벨을 추가할 수 있습니다 .

모자이크 그림 팝업 메뉴

" 분할 분석 " 보고서에서 모자이크 그림을 마우스 오른쪽 버튼으로 클릭하여 셀의 색상을 변경하고 라벨을 지정할 수 있습니다 . 그림의 각 부분마다 독자적 메뉴가 있습니다 .

색상 설정 수준에 현재 할당된 색상을 표시하고 할당을 업데이트할 수 있습니다 . 자세한 내용은 " 색상 설정 " 에서 확인하십시오 .

셀 라벨 지정 모자이크 그림의 각 셀에 표시할 라벨을 지정합니다 .

라벨 없음 라벨을 표시하지 않고 이전에 추가한 라벨을 제거합니다 .

개수별 라벨 각 셀의 관측값 수를 표시합니다 .

백분율별 라벨 각 셀의 관측값 백분율을 표시합니다 .

값별 라벨 각 셀에 해당하는 Y 변수의 수준을 표시합니다 .

행별 라벨 셀이 나타내는 모든 행의 행 라벨을 표시합니다 .

선 색상 각 셀 주위의 선에 대한 색상을 지정합니다 .

선 스타일 각 셀 주위의 선에 대한 스타일을 지정합니다 .

선 너비 각 셀 주위의 선에 대한 너비를 지정합니다 .

투명도 셀 색상의 투명도를 지정합니다 .

참고 : 팝업 메뉴의 나머지 옵션에 대한 설명은 JMP 사용의 에서 확인하십시오 .

색상 설정

모자이크 그림 팝업 메뉴에서 " 색상 설정 " 옵션을 선택하면 " 값에 대한 색상 선택 " 창이 나타납니다 . 이 창이 처음 열리면 수준에 할당된 현재 색상이 표시됩니다 .

기본 모자이크 색상은 Y 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 " 창에는 다음 옵션이 포함되어 있습니다 .

매크로 다음 방법 중 하나를 사용하여 색상을 업데이트합니다 .

끝 사이의 그래디언트 변수의 모든 수준에 색상 그래디언트를 적용합니다 .

선택한 점 사이의 그래디언트 선택한 수준에만 색상 그래디언트를 적용합니다 . 선택하려는 수준 위에서 커서를 드래그하거나 Shift 키를 누른 채 첫 번째 수준과 마지막 수준을 클릭하면 수준 범위를 선택할 수 있습니다 .

색상 반전 색상의 순서를 역순으로 바꿉니다 .

이전 색상으로 되돌리기 " 값에 대한 색상 선택 " 창이 열린 후 변경된 내용을 되돌립니다 .

색상 테마 색상 테마를 기반으로 각 값의 색상을 업데이트합니다 .

열에 색상 저장 업데이트된 색상을 데이터 테이블에 저장할 수 있습니다 . 색상 테마를 변경한 다음 이 체크박스를 선택하면 연결된 데이터 테이블의 열에 " 값 색상 " 열 특성이 추가됩니다 . 데이터 테이블에서 이 특성을 편집하려면 **열 > 열 정보**를 선택합니다 .

분할표

분할표는 이원 빈도 테이블입니다 . X 변수의 각 수준에 대한 행과 Y 변수의 각 수준에 대한 열이 있습니다 .

그림 7.6 분할표의 예

분할표

country	개수 / 총 % / 열 % / 행 %	size: Large	Medium	Small	합계
American		36	53	26	115
		11.88	17.49	8.58	37.95
		85.71	42.74	18.98	
		31.30	46.09	22.61	
European		4	17	19	40
		1.32	5.61	6.27	13.20
		9.52	13.71	13.87	
		10.00	42.50	47.50	
Japanese		2	54	92	148
		0.66	17.82	30.36	48.84
		4.76	43.55	67.15	
		1.35	36.49	62.16	
합계		42	124	137	303
		13.86	40.92	45.21	

분할표의 경우 다음 사항에 유의하십시오 .

- " 개수 ", " 총 %", " 열 %" 및 " 행 %" 는 행 머리글과 열 머리글이 있는 각 셀 내의 데이터에 해당합니다 .
- 마지막 열에는 각 행의 총 개수와 백분율이 포함됩니다 .
- 맨 아래 행에는 각 열의 총 개수와 백분율이 포함됩니다 .

그림 7.6에서 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

개수	설명	표에 사용된 라벨
36	대형이면서 미국에서 생산된 자동차의 수	개수
11.88%	총 303 대 중 대형이면서 미국에서 생산된 자동차의 백분율 (36/303)	총 %
85.71%	대형 자동차 42 대 중 미국에서 생산된 자동차의 백분율 (36/42)	열 %
31.30%	미국에서 생산된 115 대 자동차 중 대형 자동차의 백분율 (36/115)	행 %
37.95%	총 303 대 중 미국에서 생산된 자동차의 백분율 (115/303)	총 %
13.86%	총 303 대 중 대형 자동차의 백분율 (42/303)	총 %

팁 :

- 분할표에 통계량을 표시하거나 숨기려면 " 분할표 " 의 빨간색 삼각형 메뉴에서 표시하거나 숨길 통계량을 선택합니다 .
- 분할표의 통계량 형식을 변경하려면 표에서 값을 마우스 오른쪽 버튼으로 클릭하고 " 형식 " 을 선택합니다 .

분할표의 통계량에 대한 설명

분할 분석 플랫폼의 " 분할표 " 에는 다음 통계량이 포함됩니다 .

개수 셀 빈도 , 빈도 소계 및 총 합계입니다 . 총 합계는 총 표본 크기이기도 합니다 .

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 나눠서 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 표 맨 위에서 맨 아래까지 계산된 누적 열 합계입니다 .

열 누적 % 표 맨 위에서 맨 아래까지 계산된 누적 열 백분율입니다 .

행 누적 표 왼쪽에서 오른쪽까지 계산된 누적 행 합계입니다 .

행 누적 % 표 왼쪽에서 오른쪽까지 계산된 누적 행 백분율입니다 .

데이터 테이블로 만들기 분할표에 현재 나타나는 통계량을 포함하는 새 데이터 테이블을 생성합니다 .

형식 분할표의 통계량 형식을 변경할 수 있는 창을 엽니다 .

검정 보고서

분할 분석 플랫폼의 " 검정 " 보고서에는 반응 수준 비율이 X 변수의 수준과 독립적인지 여부를 평가하는 데 사용할 수 있는 두 검정의 결과가 표시됩니다 .

" 검정 " 보고서에는 다음 통계량이 포함됩니다 .

N 관측값의 총 개수입니다 .

DF 검정과 관련된 자유도입니다 . 자유도는 $(c - 1)(r - 1)$ 와 같습니다 . 여기서 c 는 열 수이고 r 은 행 수입니다 .

-1* 로그 가능도 적합 및 불확도를 측정하는 음의 로그 가능도로 , 연속형 반응에서의 제곱합과 매우 유사합니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

$R^2(U)$ 모형 적합에 기인한 전체 불확실성의 비율입니다 .

- R^2 이 1 이면 X 변수의 수준이 Y 변수의 수준을 완전히 예측한다는 의미입니다 .
- R^2 이 0 이면 모형이 고정 기준 반응률을 사용하는 것보다 더 잘 예측하지 못한다는 의미입니다 .

검정 각 표본 범주에서 반응률이 동일하다는 가설에 대한 카이제곱 검정의 이름입니다 .

카이제곱 카이제곱 검정의 검정 통계량입니다 .

Prob>ChiSq 반응과 요인 사이에 아무런 관계가 없는 경우 카이제곱 값이 계산된 값보다 클 확률입니다 . 두 변수의 수준이 각각 두 개인 경우에는 한쪽 꼬리 검정과 양쪽 꼬리 검정에 대한 Fisher 정확 확률도 표시됩니다 .

이 보고서의 통계량에 대한 자세한 내용은 "검정 보고서에 대한 통계 상세 정보"에서 확인하십시오.

Fisher 정확 검정

이 보고서에서는 2 x 2 테이블에 대한 Fisher 정확 검정 결과를 제공합니다. 결과는 2 x 2 테이블에 대해 자동으로 나타납니다. Fisher 정확 검정과 r x c 테이블의 검정에 대한 자세한 내용은 "Fisher 정확 검정 보고서"에서 확인하십시오.

분할 분석 플랫폼 옵션

"분할 분석"의 빨간색 삼각형 메뉴에는 추가 분석을 수행하는 옵션이 포함되어 있습니다.

참고 : " 그룹 적합 " 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에만 표시됩니다 . " 그룹 적합 " 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 을 기준으로 정렬할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

" 분할 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

모자이크 그림 분할표의 그래픽 표현을 표시하거나 숨깁니다 . 자세한 내용은 " 모자이크 그림 " 에서 확인하십시오 .

분할표 이원 빈도 테이블을 표시하거나 숨깁니다 . 이 테이블에는 X 변수의 각 수준에 대한 행과 Y 변수의 각 수준에 대한 열이 있습니다 . 자세한 내용은 " 분할표 " 에서 확인하십시오 .

검정 반응 수준 비율이 X 변수의 수준 간에 동일한지 여부를 측정하는 검정을 표시하거나 숨깁니다 . 이러한 검정은 연속형 데이터에 대한 분산 분석 테이블과 유사합니다 . 자세한 내용은 " 검정 보고서 " 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 (반응 수준이 정확히 두 개인 경우에만 사용 가능) 그룹 비율을 비교하기 위한 ANOMP(비율에 대한 평균 분석) 의사결정 차트를 표시하거나 숨깁니다 . ANOMP 는 X 변수의 수준에 대한 반응 비율을 전체 반응 비율과 비교하는 다중 비교 절차입니다 . 자세한 내용은 " 비율에 대한 평균 분석 보고서 " 에서 확인하십시오 .

대응 분석 빈도 테이블에서 개수 패턴이 유사한 행 또는 열을 식별하는 대응 분석을 표시하거나 숨깁니다. 대응 분석 그림에는 분할표의 각 행과 각 열에 해당하는 점이 하나씩 있습니다. 자세한 내용은 " 대응 분석 보고서 " 에서 확인하십시오.

Cochran Mantel Haenszel 세 번째 분류 변수를 블록화한 후 두 범주형 변수 간에 관계가 있는지 여부를 판별하는 검정을 표시하거나 숨깁니다. 이 옵션을 선택하면 검정에 대한 그룹화 열을 지정할 수 있는 창이 나타납니다. 자세한 내용은 "Cochran-Mantel-Haenszel 검정 보고서 " 에서 확인하십시오.

합치도 통계량 (X 및 Y 변수의 수준이 동일한 경우에만 사용 가능) 수준 간 합치도를 측정하는 통계량이 포함된 보고서를 표시하거나 숨깁니다. 이 보고서에는 카파 통계량 (Agresti 1990) 과 통계량에 대한 표준 오차, 신뢰 구간 및 가설 검정이 포함됩니다. 또한 McNemar 검정이라고도 하는 Bowker 대칭성 검정도 포함됩니다. 자세한 내용은 " 합치도 통계량 보고서 " 에서 확인하십시오.

상대 위험도 (X 변수와 Y 변수의 수준이 각각 정확히 두 개인 경우에만 사용 가능) 반응 수준 간의 상대 위험도를 표시하거나 숨깁니다. 자세한 내용은 " 상대 위험도 보고서 " 에서 확인하십시오.

" 상대 위험도 " 보고서에는 이 비율에 대한 신뢰 구간도 제공됩니다. α 수준 설정 옵션을 사용하여 유의 수준을 변경할 수 있습니다.

위험도 차이 (X 변수와 Y 변수의 수준이 각각 정확히 두 개인 경우에만 사용 가능) 반응 수준 간의 위험도 차이를 표시하거나 숨깁니다.

" 위험도 차이 " 보고서에는 이 비율에 대한 신뢰 구간도 제공됩니다. α 수준 설정 옵션을 사용하여 유의 수준을 변경할 수 있습니다.

승산비 (X 변수와 Y 변수의 수준이 각각 정확히 두 개인 경우에만 사용 가능) 승산비 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 승산비에 대한 통계 상세 정보 " 에서 확인하십시오.

" 승산비 " 보고서에는 이 비율에 대한 신뢰 구간도 제공됩니다. α 수준 설정 옵션을 사용하여 유의 수준을 변경할 수 있습니다.

비율에 대한 2 표본 검정 (X 변수와 Y 변수의 수준이 각각 정확히 두 개인 경우에만 사용 가능) 비율에 대한 2 표본 검정을 표시하거나 숨깁니다. 이 검정에서는 X 변수의 두 수준 간에 Y 변수의 비율을 비교합니다. 자세한 내용은 " 비율에 대한 2 표본 검정 보고서 " 에서 확인하십시오.

연관성 측도 분할표에 있는 변수 간의 연관성 측도가 포함된 보고서를 표시하거나 숨깁니다. 자세한 내용은 " 연관성 측도 보고서 " 에서 확인하십시오.

Cochran Armitage 추세 검정 (한 변수의 수준이 정확히 두 개이고 다른 변수는 순서형인 경우에만 사용 가능) 단일 변수의 수준 간에 이항 비율 추세에 대한 검정을 표시하거나 숨깁니다. 자세한 내용은 "Cochran Armitage 추세 검정 보고서 " 에서 확인하십시오.

정확 검정 다음과 같은 정확 검정에 대한 옵션을 포함합니다.

Fisher 검정 두 범주형 변수 간의 연관성을 검정하는 Fisher 정확 검정을 표시하거나 숨깁니다. 이 검정은 대표본 분포 가정에 의존하지 않습니다. 자세한 내용은 "Fisher 정확 검정 보고서 " 에서 확인하십시오.

Cochran Armitage 추세 검정 (한 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 정확 Cochran-Armitage 추세 검정을 표시하거나 숨깁니다 . 자세한 내용은 "Cochran Armitage 추세 검정 보고서 " 에서 확인하십시오 .

합치도 통계량 (한 변수의 수준이 정확히 두 개인 경우에만 사용 가능) 정확 합치도 통계량 카파를 표시하거나 숨깁니다. 자세한 내용은 "합치도 통계량 보고서"에서 확인하십시오.

참고 : 빈도 변수에 정수가 아닌 값이 있으면 정확 검정을 사용할 수 없습니다 . 또한 전체 표본 크기가 32767보다 크고 분할표가 2 x 2보다 큰 경우 정확 검정 옵션을 사용할 수 없습니다.

동등성 검정 (X 변수와 Y 변수의 수준이 각각 정확히 두 개인 경우에만 사용 가능) 반응 범주 간의 동등성 , 우월성 또는 비열등성을 검정하기 위한 다음 옵션이 포함되어 있습니다 . 자세한 내용은 " 동등성 검정 보고서 " 에서 확인하십시오 .

위험도 차이 반응 수준 간의 위험도 차이에 대한 동등성 , 우월성 또는 비열등성 검정 옵션이 있는 창을 시작합니다 .

상대 위험도 반응 수준 간의 상대 위험도에 대한 동등성 , 우월성 또는 비열등성 검정 옵션이 있는 창을 시작합니다 .

표시 옵션 모자이크 그림을 수정하기 위한 다음 옵션이 포함되어 있습니다 .

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 에서 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 .

플랫폼 환경 설정 현재 플랫폼 환경 설정을 보거나 , 현재 JMP 보고서의 설정과 일치하도록 플랫폼 환경 설정을 업데이트할 수 있는 옵션이 포함되어 있습니다 .

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 .

분할 분석 보고서

분할 분석 플랫폼을 사용하면 시각화 요소를 생성하고 두 범주형 변수 간의 관계를 조사할 수 있습니다. 이 섹션에는 특정 분석 옵션에 대해 생성되는 보고서 관련 정보가 포함되어 있습니다.

- "비율에 대한 평균 분석 보고서"
- "대응 분석 보고서"
- "Cochran-Mantel-Haenszel 검정 보고서"
- "합치도 통계량 보고서"
- "상대 위험도 보고서"
- "비율에 대한 2 표본 검정 보고서"
- "연관성 측도 보고서"
- "Cochran Armitage 추세 검정 보고서"
- "Fisher 정확 검정 보고서"
- "동등성 검정 보고서"

비율에 대한 평균 분석 보고서

분할 분석 플랫폼에서 ANOMP(비율에 대한 평균 분석)는 개별 그룹 비율이 전체 비율과 다른지 여부를 검정하기 위한 다중 비교 방법입니다. 평균 분석 방법에 대한 자세한 내용은 Nelson et al. 연구 자료(2005)에서 확인하십시오. "비율에 대한 평균 분석의 예"의 내용도 참조하십시오.

Y 변수의 수준이 두 개인 경우 이 옵션을 사용하여 X 변수의 수준에 대한 반응 비율을 전체 반응 비율과 비교할 수 있습니다. 이 방법은 이항 분포에 대한 정규 근사를 사용합니다. 따라서 표본 크기가 너무 작으면 결과에 경고가 표시됩니다.

"비율에 대한 평균 분석"의 빨간색 삼각형 메뉴에는 다음과 같은 옵션이 표시됩니다.

유의 수준 설정 "비율에 대한 평균 분석" 차트의 결정 한계를 결정하는 데 사용되는 유의 수준을 지정합니다.

요약 보고서 표시 X 변수의 각 수준에 대한 반응 비율 및 결정 한계가 포함된 보고서를 표시하거나 숨깁니다. 한계 초과 여부도 보고서에 표시됩니다.

비율에 대한 반응 수준 전환 분석에 사용되는 반응 범주를 변경합니다.

표시 옵션 비율에 대한 평균 분석 차트를 수정하기 위한 다음 옵션이 포함되어 있습니다.

결정 한계 표시 비율에 대한 평균 분석 차트에 결정 한계선을 표시하거나 숨깁니다.

결정 한계 음영 표시 비율에 대한 평균 분석 차트에 결정 한계 음영을 표시하거나 숨깁니다.

중심선 표시 비율에 대한 평균 분석 차트에 중심선을 표시하거나 숨깁니다.

점 옵션 " 비율에 대한 평균 분석 " 차트의 점 그리기 스타일을 지정합니다 . 세로 바늘 그리기 , 점 연결 및 점만 표시 중에서 선택할 수 있습니다 . 기본적으로 평균 위치에 그려진 가로선에 점을 연결하는 바늘로 차트를 그립니다 .

대응 분석 보고서

분할 분석 플랫폼의 " 대응 분석 " 옵션은 빈도 테이블에서 유사한 개수 패턴을 가진 행 또는 열을 표시하기 위한 그래픽 기법을 제공합니다. 대응 분석 그림에는 각 행과 각 열에 해당하는 점이 하나씩 있습니다. 수준 수가 많아서 모자이크 그림으로는 유용한 정보를 도출하기 어려운 경우에 대응 분석을 사용합니다. 자세한 내용은 " 대응 분석의 예 " 에서 확인하십시오 .

" 대응 분석 " 그림에는 행 및 열 프로파일이 포함됩니다 . **행 프로파일**은 행별 비율의 집합 , 즉 한 행의 개수를 해당 행의 총 개수로 나눈 값으로 정의할 수 있습니다 . 두 행의 행 프로파일이 매우 유사한 경우에는 대응 분석 그림에서 해당 점들이 서로 가깝게 나타납니다. 행 점 간의 거리를 제곱한 값은 행 쌍 간의 동질성을 검정하는 카이제곱 검정 통계량에 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 이러한 점 간의 관계는 그림을 해석하는 데 도움이 됩니다 .

" 대응 분석 " 의 빨간색 삼각형 메뉴에 있는 옵션을 사용하여 " 분할 분석 " 보고서에 3 차원 산점도를 추가하고 열 특성을 데이터 테이블에 추가합니다 .

3D 대응 분석 3 차원 산점도를 표시하거나 숨깁니다 .

값 순서 저장 " 값 순서화 " 열 특성을 데이터 테이블의 X 변수 열과 Y 변수 열에 모두 저장합니다 . 이 열 특성은 첫 번째 대응 스코어 계수를 기준으로 정렬된 수준의 순서를 지정합니다 .

상세 정보 보고서

분할 분석 플랫폼의 " 상세 정보 " 보고서에는 대응 분석에 대한 통계 정보가 포함되고 그림에 사용된 값이 표시됩니다 .

특이값 분할표의 특이값 분해입니다 . 자세한 내용은 " 대응 분석에 대한 통계 상세 정보 " 에서 확인하십시오 .

관성 정준 차원의 측면을 고려한 상대적인 변동을 반영하는 특이값 제곱입니다 .

비율 총 관성에 대한 관성 비율입니다 .

누적 관성의 누적 비율입니다 .

팁 : 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 .

Y 변수 c1, c2, c3 " 대응 분석 " 그림에 표시된 값입니다 .

Cochran-Mantel-Haenszel 검정 보고서

분할 분석 플랫폼에서 Cochran-Mantel-Haenszel 검정은 세 번째 분류 기준에 따라 블록화한 후 두 범주형 변수 사이의 관계를 평가합니다. 세 번째 분류 변수의 블록화를 층화라고도 합니다.

참고 : 자세한 내용은 "Cochran-Mantel-Haenszel 검정의 예 " 에서 확인하십시오.

"Cochran-Mantel-Haenszel 검정 " 보고서에는 다양한 검정 결과에 대한 테이블이 포함됩니다. 각 검정에는카이제곱 통계량 (카이제곱), 관련 자유도 (DF) 및 유의 확률 (Prob>Chisq) 이 있습니다. 다음과 같은 검정이 보고됩니다.

스코어 상관 (Y 변수와 X 변수가 모두 순서형 또는 구간일 때 적용 가능) 블록 변수의 최소 한 개 수준에서 Y 변수와 X 변수 사이에 선형 연관성이 있다는 대립가설을 검정합니다.

열 범주별 행 스코어 (Y 변수가 순서형 또는 구간일 때 적용 가능) 블록 변수의 최소 한 개 수준에서 행의 평균 스코어가 동일하지 않다는 대립가설을 검정합니다.

행 범주별 열 스코어 (X 가 순서형 또는 구간일 때 적용 가능) 블록 변수의 최소 한 개 수준에서 열의 평균 스코어가 동일하지 않다는 대립가설을 검정합니다.

범주의 일반 연관성 블록 변수의 최소 한 개 수준에서 Y 변수와 X 변수 사이에 일종의 연관성이 있는지 검정합니다.

팁 : Cochran-Mantel-Haenszel 검정의 블록 변수로 상수 값이 포함된 열을 지정하여 행 평균 스코어 검정을 수행할 수 있습니다.

합치도 통계량 보고서

두 범주형 변수의 수준이 동일한 경우 분할 분석 플랫폼의 " 합치도 통계량 " 옵션을 사용하여 카파 통계량을 계산합니다 (Agresti 1990). 이 옵션은 카파 통계량 표준 오차, 신뢰 구간, 가설 검정 및 Bowker 대칭성 검정도 계산합니다. 자세한 내용은 " 합치도 통계량 옵션의 예 " 에서 확인하십시오.

이 보고서의 카파 통계량 및 관련 *p* 값은 근사값입니다. 자세한 내용은 " 합치도 통계량에 대한 통계 상세 정보 " 에서 확인하십시오. 정확 합치도 통계량도 사용할 수 있습니다. 자세한 내용은 "Fisher 정확 검정 보고서 " 에서 확인하십시오.

" 카파 계수 " 보고서에는 다음 통계량이 포함됩니다.

카파 카파 통계량입니다.

표준 오차 카파 통계량의 표준 오차입니다.

95% 하한 카파에 대한 신뢰 구간의 하한점입니다.

95% 상한 카파에 대한 신뢰 구간의 상한점입니다.

Prob>Z 카파에 대한 단측 점근적 검정의 유의 확률 (*p* 값) 입니다. 귀무가설은 카파가 0 인지를 검정합니다.

Prob>|Z| 카파에 대한 양측 점근적 검정의 유의 확률 (*p* 값) 입니다 .

"Bowker 검정 " 보고서에는 다음 통계량이 포함됩니다 .

카이제곱 Bowker 검정의 검정 통계량입니다 . Bowker 대칭성 검정에서 귀무가설은 정방 테이블의 확률이 대칭성을 만족한다는 것 , 즉 테이블 셀의 모든 쌍에 대해 $p_{ij}=p_{ji}$ 가 성립한다는 것입니다 . Y 변수와 X 변수의 수준이 모두 정확히 두 개인 경우 Bowker 검정은 McNemar 검정과 동일합니다 .

Prob>ChiSq Bowker 검정 통계량의 유의 확률 (*p* 값) 입니다 . Bowker 검정의 귀무가설은 정방 테이블의 확률이 대칭성을 만족한다는 것입니다 .

상대 위험도 보고서

분할 분석 플랫폼에서 "상대 위험도" 옵션을 사용하여 2 x 2 분할표에 대한 위험비를 계산합니다. 보고서에는 신뢰 구간도 표시됩니다. 이 방법에 대한 자세한 내용은 Agresti 연구 자료(1990, sect. 3.4.2)에서 확인하십시오. 자세한 내용은 "상대 위험도 옵션의 예"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 위험비를 계산할 수 있습니다.

비율에 대한 2 표본 검정 보고서

분할 분석 플랫폼에서 신뢰 구간을 생성하고 두 비율 간의 차이에 대한 가설 검정을 수행할 수 있습니다 . 이 분석은 X 변수와 Y 변수의 수준이 모두 정확히 두 개일 때 사용할 수 있습니다 . 자세한 내용은 " 비율에 대한 2 표본 검정의 예 " 에서 확인하십시오 .

설명 수행 중인 검정에 대한 설명입니다 .

비율 차이 X 변수의 수준 간 비율 차이입니다 .

95% 하한 차이에 대한 신뢰 구간의 하한점입니다 . 이 값은 조정 Wald 신뢰 구간을 기반으로 합니다 .

95% 상한 차이에 대한 신뢰 구간의 상한점입니다 . 이 값은 조정 Wald 신뢰 구간을 기반으로 합니다 .

조정 Wald 검정 (귀무가설) 한쪽 꼬리 및 양쪽 꼬리 검정의 귀무가설에 대한 설명입니다 .

확률 검정에 대한 유의 확률 (*p* 값) 입니다 .

팁 : 테이블 아래의 라디오 버튼을 사용하여 검정의 관심 반응 수준을 변경할 수 있습니다 .

연관성 측도 보고서

분할 분석 플랫폼에서 " 연관성 측도 " 옵션은 연관성 통계량을 제공합니다 . 자세한 내용은 " 연관성 측도 옵션의 예 " 에서 확인하십시오 .

" 연관성 측도 " 보고서에는 다음 통계량에 대한 값 , 표준 오차 및 신뢰 구간이 포함됩니다 .

감마 순서형 연관성 측도입니다 . 동점 쌍을 무시하고 부합 / 비부합 쌍의 확률 차이로 정의됩니다 . -1 ~ 1 범위의 값을 취합니다 .

Kendall Tau-b 감마와 유사하되, 동점 수에 대한 수정을 사용합니다. -1 ~ 1 범위의 값을 취합니다.

Stuart Tau-c 감마와 유사하되 , 테이블 크기에 대한 조정과 동점 값에 대한 수정을 사용합니다 . -1 ~ 1 범위의 값을 취합니다 .

Somers D Tau-b 에 대한 비대칭 수정입니다 . Somers D 는 독립 변수에 동점 값을 가진 쌍이 있을 때만 동점 값에 대한 수정을 사용합니다 . -1 ~ 1 범위의 값을 취합니다 .

- C|R은 행 변수 X가 독립 변수로 간주되고 열 변수 Y가 종속 변수로 간주됨을 나타냅니다.
- 마찬가지로 R|C 는 열 변수 Y 가 독립 변수로 간주되고 행 변수 X 가 종속 변수로 간주됨을 나타냅니다 .

참고 : 감마 , Kendall Tau-b, Stuart Tau-c 및 Somers D 는 X 가 증가함에 따라 변수 Y 가 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 비대칭 C|R 의 경우와 R|C 의 경우에 차이가 있습니다 . 0 ~ 1 범위의 값을 취합니다 .

- C|R 의 경우 , 이 측도는 행 변수 X 에 대한 정보가 주어졌을 때 열 변수 Y 를 예측하는 데 있어서 가능한 향상 정도로 해석됩니다 .
- R|C 의 경우 , 이 측도는 열 변수 Y 에 대한 정보가 주어졌을 때 행 변수 X 를 예측하는 데 있어서 가능한 향상 정도로 해석됩니다 .

람다 대칭 대략적으로 두 람다 비대칭 측도의 평균으로 해석됩니다. 0 ~ 1 범위의 값을 취합니다.

불확도 계수 C|R 의 경우와 R|C 의 경우에 차이가 있습니다 . 0 ~ 1 범위의 값을 취합니다 .

- C|R 의 경우 , 이 측도는 행 변수 X 로 설명되는 열 변수 Y 의 불확실성 비율입니다 .
- R|C 의 경우, 이 측도는 열 변수 Y 로 설명되는 행 변수 X 의 불확실성 비율로 해석됩니다.

불확도 계수 대칭 두 불확도 계수 측도의 대칭 버전입니다 . 0 ~ 1 범위의 값을 취합니다 .

참고 : 람다 및 불확도 측도는 순서형 변수와 명목형 변수 둘 다에 적합합니다 .

연관성 측도 통계량에 대한 계산 상세 정보는 SAS Institute Inc. 가이드 (2020b) 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 보고서

분할 분석 플랫폼에서 "Cochran Armitage 추세 검정" 옵션은 단일 변수의 수준 간에 이항 비율의 추세를 검정합니다. 이 검정은 한 변수의 수준이 두 개이고 다른 변수는 순서형인 경우에만 적합합니다. 2수준 변수는 반응을 나타내고 다른 변수는 순서형 수준이 있는 설명 변수를 나타냅니다. 귀무가설은 추세가 없다는 것으로, 설명 변수의 모든 수준에 대해 이항 비율이 동일함을 의미합니다. 자세한 내용은 "Cochran Armitage 추세 검정의 예"에서 확인하십시오.

참고 : 이 검정에서 제공되는 검정 통계량 및 유의 확률 (*p* 값) 은 근사값입니다 . 정확 추세 검정도 사용할 수 있습니다 .

Fisher 정확 검정 보고서

r x *c* 테이블에 대한 Fisher 정확 검정은 두 변수 간의 연관성을 검정합니다. *r* x *c* 테이블은 행에 할당된 변수의 수준이 *r* 개이고 열에 할당된 변수의 수준이 *c* 개인 테이블입니다. Fisher 정확 검정에서는 행 및 열 합계가 고정되어 있다고 가정하고 초기하 분포를 사용하여 확률을 계산합니다.

이 검정은 대표본 분포 가정에 의존하지 않습니다. 따라서 작은 표본 크기 또는 희소 테이블과 같이 가능도비와 Pearson 검정의 신뢰도가 떨어지는 경우에 적절합니다.

"Fisher 정확 검정 " 보고서에는 다음과 같은 정보가 포함됩니다 .

테이블 확률 (P) 관측된 테이블의 확률입니다 . 이 값은 검정의 *p* 값이 아닙니다 .

양측 확률 ≤ P 양측 검정에 대한 유의 확률 (*p* 값) 입니다 .

2 x 2 테이블의 경우 한 행 또는 열에 포함된 값이 모두 0인 경우(이 경우 검정을 계산할 수 없음)만 제외하고 자동으로 Fisher 정확 검정이 수행됩니다. 자세한 내용은 "검정 보고서"에서 확인하십시오.

동등성 검정 보고서

분할 분석 플랫폼에서 "동등성 검정" 하위 메뉴의 옵션을 사용하면 위험도 차이나 상대 위험도에 대한 동등성, 우월성 또는 비열등성 검정을 수행할 수 있습니다.

동등성 검정 보고서에는 그림과 요약 테이블이 포함되어 있습니다. "동등성 검정" 하위 메뉴에서 옵션을 선택하는 경우 "동등성 검정 규격" 창에서 검정 특성을 지정해야 합니다.

동등성 검정 규격 창

" 동등성 검정 " 하위 메뉴에서 동등성 검정 옵션을 선택하면 검정을 정의할 수 있는 창이 시작됩니다 .

대립가설 동등성 검정의 구조를 정의합니다 .

동등성 (양측) 동등성 검정을 지정합니다 . 그룹 차이가 동등성 한계보다 크지 않음을 보여 주는 것이 목표인 경우 이 옵션을 사용합니다 .

우월성 (단측) 우월성 검정을 지정합니다 . 한 그룹이 다른 그룹보다 우월하거나 더 우수함을 보여 주는 것이 목표인 경우 이 옵션을 사용합니다 .

비열등성 (단측) 비열등성 검정을 지정합니다. 한 그룹이 다른 그룹보다 열등하지 않음을 보여 주는 것이 목표인 경우 이 옵션을 사용합니다 .

대립가설 부분 (단측 검정에만 사용 가능) 대립가설의 방향 (목표) 을 지정합니다 .

가설 그림 가설 검정의 그래픽 표현을 제공합니다 .

Y 의 관심 범주 비율의 분자에 해당하는 Y 변수 수준을 정의합니다 .

X 의 관심 범주 대조군 비율의 분모에 해당하는 X 변수 수준을 정의합니다 .

마진 및 유의 수준 검정의 유의 수준을 정의합니다 .

차이 (위험도 차이 검정에만 사용 가능) 동등성 , 우월성 또는 비열등성 마진을 지정합니다 . 이 마진 (델타) 은 실제적 유의성을 가진 차이입니다 . 동등성 검정의 경우 차이가 0 보다 커야 합니다 .

비율 (상대 위험도 검정에만 사용 가능) 동등성 , 우월성 또는 비열등성 마진을 위험비로 지정합니다 . 이 마진 (델타) 은 실제적 유의성을 가진 위험비입니다 . 값 범위는 (비율 , 1/ 비율) 로 정의됩니다 . 동등성 검정의 경우 비율이 1 과 달라야 합니다 .

유의 수준 검정의 유의 수준을 지정합니다 .

동등성 검정 보고서

검정 보고서는 대립가설에 대한 설명으로 시작됩니다 . 각 비교에 대해 검정 보고서에는 다음 열이 포함됩니다 .

차이 (위험도 차이 검정에만 사용 가능) 추정된 위험도 차이입니다 .

비율 (상대 위험도 검정에만 사용 가능) 추정된 위험비입니다 .

차이의 표준 오차 (위험도 차이 검정에만 사용 가능) 위험도 차이의 표준 오차 추정값입니다 .

하한 표준 오차 , 상한 표준 오차 (단측 검정의 경우 한계가 하나만 표시됨) 하한 또는 상한 가설값에서 추정된 표준 오차입니다 .

하한 z 비 , 상한 z 비 (단측 검정의 경우 한계가 하나만 표시됨) 단측 유의성 검정의 하한 또는 상한 *z* 비입니다 .

하한 p 값 , 상한 p 값 (단측 검정의 경우 *p* 값이 하나만 표시됨) 하한 또는 상한 *z* 비에 해당하는 유의 확률 (*p* 값) 입니다 .

최대 p 값 (양측 검정의 경우에만 표시됨) 하한 및 상한 *p* 값의 최대값입니다 .

90% 하한 , 90% 상한 위험도 차이 또는 위험비에 대한 1–2α 신뢰 구간의 한계입니다 .

평가 지정된 유의 수준에 대한 가설 검정 평가입니다 .

동등성 검정 옵션

" 동등성 검정 ", " 우월성 검정 " 또는 " 비열등성 검정 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

검정 보고서 위험도 차이 또는 위험비에 대한 동등성 검정 , 우월성 검정 또는 비열등성 검정을 요약하는 보고서를 표시하거나 숨깁니다 . 자세한 내용은 " 동등성 검정 보고서 " 에서 확인하십시오 .

포레스트 그림 포레스트 그림을 표시하거나 숨깁니다 . 비교 신뢰 구간 대 위험도 차이 또는 상대 위험도가 표시됩니다 . 구간은 위험도 차이 또는 상대 위험도 척도로 표시됩니다 . 음영은 동등 , 우월 또는 비열등 영역을 나타냅니다 .

팁 : 점을 커서로 가리키면 비교되는 그룹과 추정된 위험도 차이 또는 상대 위험도가 표시됩니다 .

제거 " 분할 분석 " 보고서 창에서 검정 보고서를 제거합니다 .

분할 분석 플랫폼의 추가 예

이 섹션에는 분할 분석 플랫폼을 사용한 예가 포함되어 있습니다 .

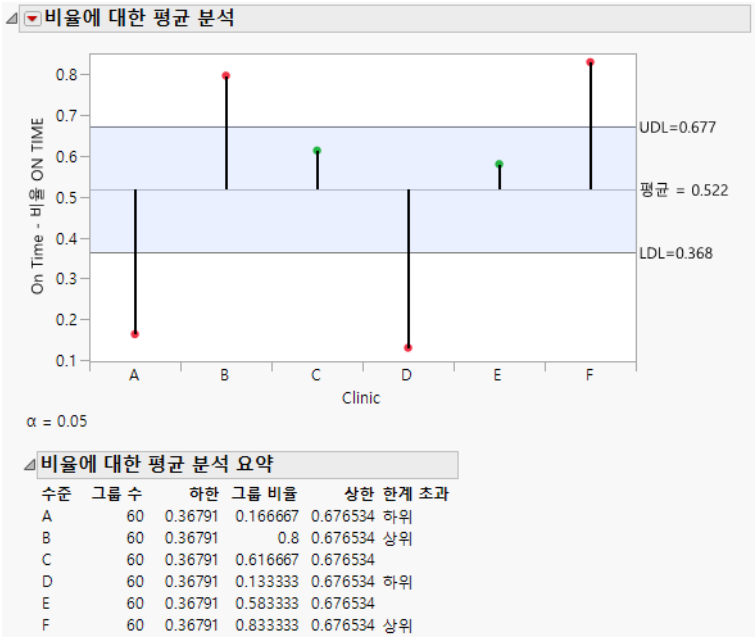
- " 비율에 대한 평균 분석의 예 "
- " 대응 분석의 예 "
- "Cochran-Mantel-Haenszel 검정의 예 "
- " 합치도 통계량 옵션의 예 "
- " 상대 위험도 옵션의 예 "
- " 비율에 대한 2 표본 검정의 예 "
- " 연관성 측도 옵션의 예 "
- "Cochran Armitage 추세 검정의 예 "

비율에 대한 평균 분석의 예

분할 분석 플랫폼을 사용하여 지역 내 6 개 병원에서 예약 시간에 정시 도착한 환자의 비율을 조사합니다 . 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 .

1. **도움말 > 샘플 데이터 폴더**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. " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **비율에 대한 평균 분석**을 선택합니다 .
8. " 비율에 대한 평균 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **요약 보고서 표시** 및 **비율에 대한 반응 수준 전환**을 선택합니다 .

그림 7.7 비율에 대한 평균 분석의 예



비율에 대한 평균 분석 요약

수준	그룹 수	하한	그룹 비율	상한	한계 초과
A	60	0.36791	0.166667	0.676534	하위
B	60	0.36791	0.8	0.676534	상위
C	60	0.36791	0.616667	0.676534	
D	60	0.36791	0.133333	0.676534	하위
E	60	0.36791	0.583333	0.676534	
F	60	0.36791	0.833333	0.676534	상위

평균 분석 그림에서는 각 병원의 예약 시간에 정시 내원한 환자의 비율을 보여 줍니다 . 다음 사항에 유의하십시오 .

- 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

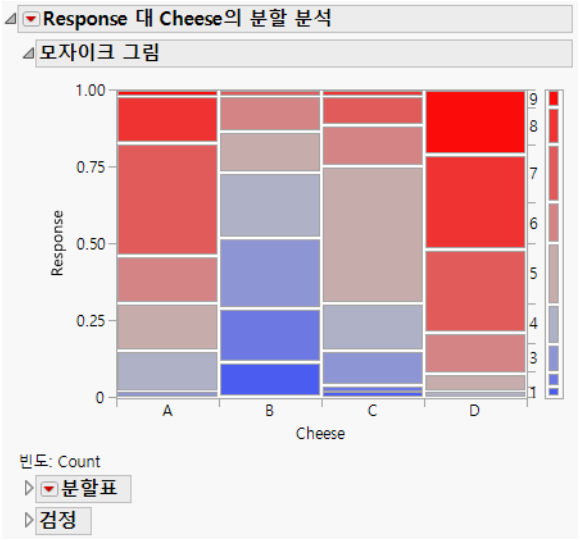
분할 분석 플랫폼을 사용하여 치즈 시식 실험에 대한 대응 분석을 수행합니다 . 이 실험에서는 네 가지 치즈 첨가물에 대해 서로 다른 9 개 반응 수준의 수를 기록했습니다 . 대응 분석은 변수 간의 연관성이 설정될 때 전체 수준에 대한 연관성을 시각화합니다 .

1. **도움말 > 샘플 데이터 폴더**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 7.8 치즈 데이터에 대한 모자이크 그림



모자이크 그림에서 치즈 종류에 따라 순위가 다르다는 것을 알 수 있습니다. 특히 치즈 B는 다른 치즈보다 일관되게 순위가 낮습니다. 다음 단계로 대응 분석을 수행할 수 있습니다.

7. 대응 분석 그림을 표시하려면 " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **대응 분석**을 선택합니다 .

그림 7.9 대응 분석 그림의 예

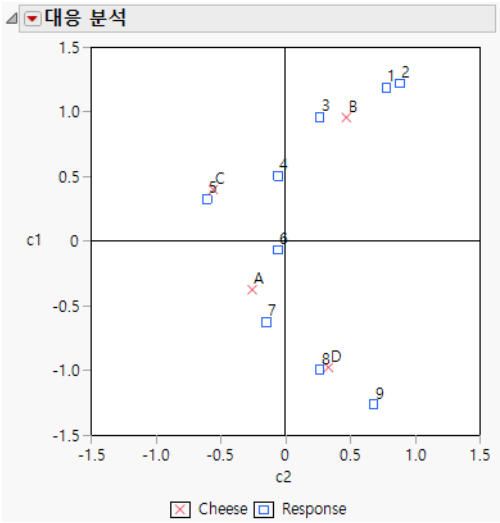


그림 7.9 에서는 대응 분석을 그래프로 보여 줍니다 . 이 그림의 축에는 "c1" 및 "c2" 라벨이 지정되어 있습니다 . 다음 사항에 유의하십시오 .

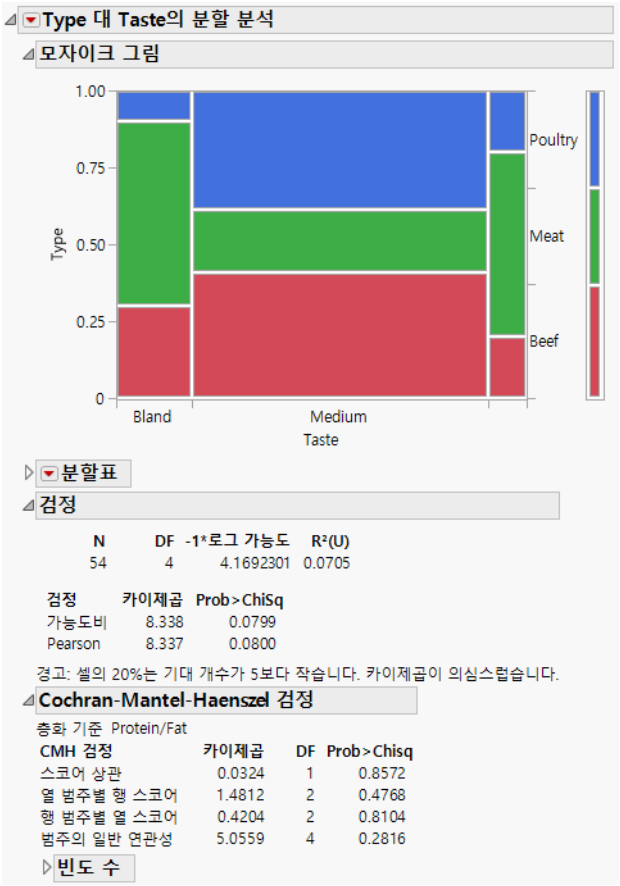
- c1 축은 일반적인 만족도를 설명하는 것으로 보입니다 . c1 축의 치즈는 맨 위의 최저 호감도에서 맨 아래의 최고 호감도로 변화합니다 .
- 치즈 D 는 반응이 8 및 9 로 , 호감도가 가장 높습니다 .
- 치즈 B 는 반응이 1, 2 및 3 으로 , 호감도가 가장 낮습니다 .
- 치즈 C 와 A 는 반응이 4, 5, 6 및 7 로 , 보통 호감도를 보입니다 .

Cochran-Mantel-Haenszel 검정의 예

분할 분석 플랫폼을 사용하여 핫도그 종류와 맛 사이의 관계를 조사합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **Cochran Mantel Haenszel** 을 선택합니다 .
7. 그룹화 변수로 Protein/Fat 을 선택하고 **확인**을 클릭합니다 .

그림 7.10 Cochran-Mantel-Haenszel 검정의 예



다음 사항에 유의하십시오 .

- " 검정 " 보고서에는 약 0.0799 로 미미하게 유의한 카이제곱 확률이 표시됩니다 . 이는 핫도그의 맛과 종류 간 관계에 어느 정도 유의성이 있음을 나타냅니다 .
- "Cochran-Mantel-Haenszel" 보고서에서는 범주의 일반 연관성에 대한 p 값이 0.2816 임을 보여 줍니다 . 단백질 / 지방 함량에 의해 비교가 제어될 때 핫도그 맛과 종류 사이에 관계가 없는 것으로 보입니다 .

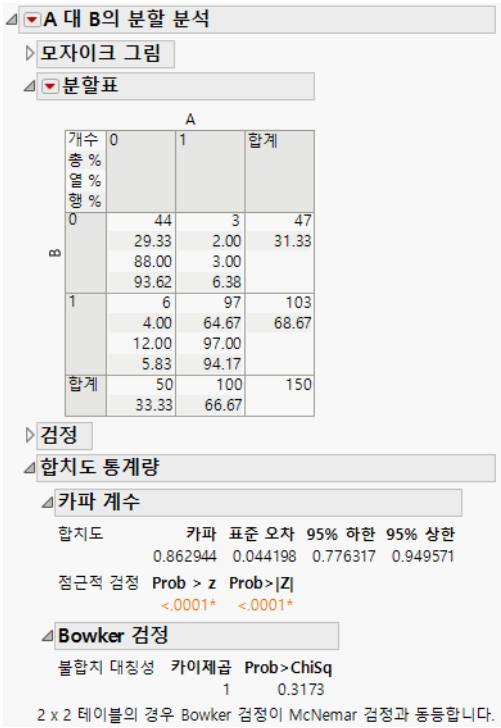
합치도 통계량 옵션의 예

분할 분석 플랫폼을 사용하여 두 평가자 사이의 관계를 조사합니다 . 데이터 테이블에는 세 명의 평가자가 50 개 부품을 각각 세 번씩 평가한 결과가 포함되어 있습니다 . 평가자 A 와 B 사이의 관계를 검토하십시오 .

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Attribute Gauge.jmp 를 엽니다 .

2. **분석 > X 로 Y 적합**을 선택합니다 .
3. A 를 선택하고 **Y, 반응**을 클릭합니다 .
4. B 를 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **합치도 통계량**을 선택합니다 .

그림 7.11 합치도 통계량 보고서의 예



합치도 통계량이 1 에 가까운 0.86 으로 높게 나타나고 p 값이 <.0001 로 작게 나타납니다. 이는 분할표를 대각선 방향으로 보면 알 수 있는 높은 합치도를 더욱 뒷받침합니다. 평가자 간의 합치도는 두 평가자가 모두 등급 0 을 주거나 등급 1 을 준 경우에 나타납니다.

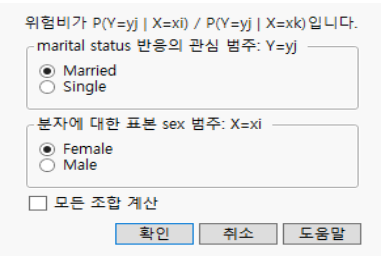
상대 위험도 옵션의 예

분할 분석 플랫폼을 사용하여 설문 조사 참가자의 상대적 기혼 / 비혼 확률을 검토합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **상대 위험도**를 선택합니다 .
   " 상대 위험도 범주 선택 " 창이 나타납니다 .

그림 7.12 상대 위험도 범주 선택 창



" 상대 위험도 범주 선택 " 창에서 다음 사항을 알 수 있습니다 .

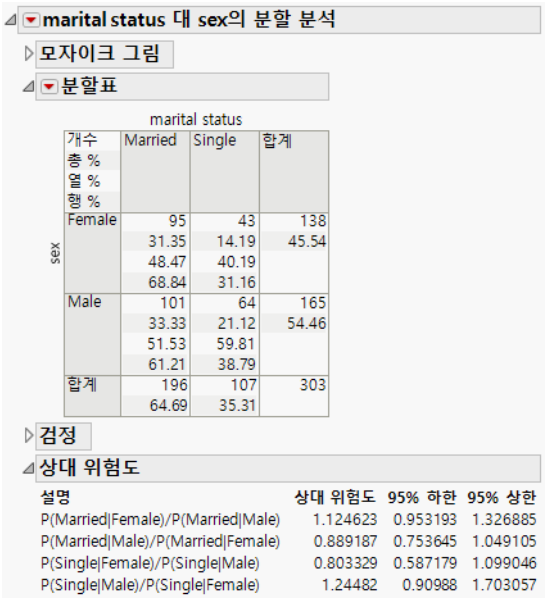
- 단일 반응 및 요인 조합에만 관심이 있는 경우 여기에서 해당 조합을 선택할 수 있습니다 . 예를 들어 그림 7.12 의 창에서 **확인**을 클릭할 경우의 계산은 다음과 같습니다 .

$$\frac{P(Y = \text{Married} | X = \text{Female})}{P(Y = \text{Married} | X = \text{Male})}$$

- 모든 $(2 \times 2 = 4)$ 반응 및 요인 수준 조합에 대해 위험비를 계산하려면 **모든 조합 계산** 체크박스를 선택합니다 (그림 7.13).

7. **모든 조합 계산** 체크박스를 선택하여 모든 조합을 요청합니다 . 다른 항목은 모두 기본 선택 상태대로 둡니다 .

그림 7.13 위험비 보고서의 예



marital status 대 sex의 분할 분석

모자이크 그림

분할표

marital status

개수 총 % 열 % 행 %	Married	Single	합계
Female	95	43	138
	31.35	14.19	45.54
	48.47	40.19	
	68.84	31.16	
Male	101	64	165
	33.33	21.12	54.46
	51.53	59.81	
	61.21	38.79	
합계	196	107	303
	64.69	35.31	

sex

검정

상대 위험도

설명	상대 위험도	95% 하한	95% 상한
P(Married|Female)/P(Married|Male)	1.124623	0.953193	1.326885
P(Married|Male)/P(Married|Female)	0.889187	0.753645	1.049105
P(Single|Female)/P(Single|Male)	0.803329	0.587179	1.099046
P(Single|Male)/P(Single|Female)	1.24482	0.90988	1.703057

상대 위험도는 다음과 같이 계산됩니다 .

1. "상대 위험도" 보고서의 첫 번째 항목 "P(Married|Female)/P(Married|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 상대 위험도는 다음과 같이 계산됩니다 .

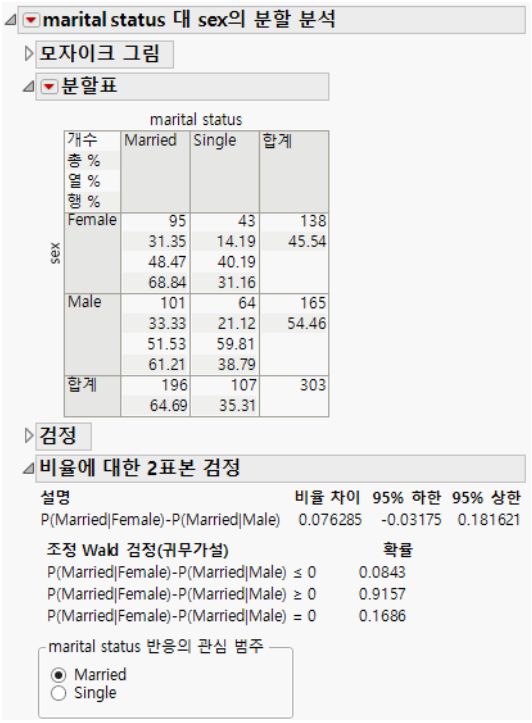
   $$P(\text{Married}|\text{Female})/P(\text{Married}|\text{Male}) = \frac{0.6884}{0.6121} = 1.1247$$

비율에 대한 2 표본 검정의 예

분할 분석 플랫폼을 사용하여 남성과 여성의 기혼 확률을 검토합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **비율에 대한 2 표본 검정**을 선택합니다 .

그림 7.14 비율에 대한 2 표본 검정 보고서의 예


marital status 대 sex의 분할 분석

모자이크 그림

분할표

marital status

개수 총 % 열 % 행 %	Married	Single	합계
Female	95	43	138
	31.35	14.19	45.54
	48.47	40.19	
	68.84	31.16	
Male	101	64	165
	33.33	21.12	54.46
	51.53	59.81	
	61.21	38.79	
합계	196	107	303
	64.69	35.31	

sex

검정

비율에 대한 2표본 검정

설명	비율 차이	95% 하한	95% 상한
P(Married|Female)-P(Married|Male)	0.076285	-0.03175	0.181621

조정 Wald 검정(귀무가설)	확률
P(Married|Female)-P(Married|Male) ≤ 0	0.0843
P(Married|Female)-P(Married|Male) ≥ 0	0.9157
P(Married|Female)-P(Married|Male) = 0	0.1686

marital status 반응의 관심 범주

(•) Married
( ) Single

이 예에서는 여성과 남성의 기혼 확률을 비교하려고 합니다 . " 분할표 " 의 " 행 %" 를 참조하여 다음을 구합니다 .

$$P(\text{Married} \mid \text{Female}) = 0.6884$$

$$P(\text{Married} \mid \text{Male}) = 0.6121$$

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. 두 비율의 차이를 검정하는 데는 조정 Wald 검정보다 Pearson 카이제곱 검정이 더 일반적으로 사용됩니다.

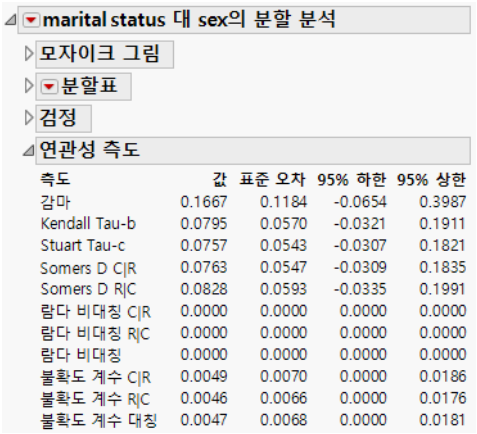
연관성 측도 옵션의 예

분할 분석 플랫폼을 사용하여 결혼과 성별의 연관성을 검토합니다 .

1. **도움말 > 샘플 데이터 폴더**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. " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **연관성 측도**를 선택합니다 .

그림 7.15 연관성 측도 보고서의 예



marital status 대 sex의 분할 분석

모자이크 그림

분할표

검정

연관성 측도

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

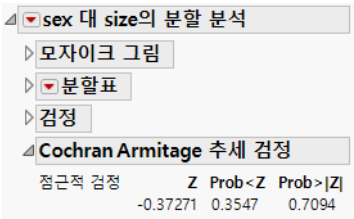
분할 분석 플랫폼을 사용하여 구매한 차량의 크기와 남성과 여성의 비율 사이에 관계가 있는지 여부를 조사합니다.

1. **도움말 > 샘플 데이터 폴더**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. " 분할 분석 " 의 빨간색 삼각형 메뉴를 클릭하고 **Cochran Armitage 추세 검정**을 선택합니다 .

그림 7.16 Cochran Armitage 추세 검정 보고서의 예



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

이 섹션에는 분할 분석 플랫폼에 대한 통계 상세 정보가 포함되어 있습니다.

- "합치도 통계량에 대한 통계 상세 정보"
- "승산비에 대한 통계 상세 정보"
- "검정 보고서에 대한 통계 상세 정보"
- "대응 분석에 대한 통계 상세 정보"

합치도 통계량에 대한 통계 상세 정보

이 섹션에는 분할 분석 플랫폼의 합치도 통계량에 대한 상세 정보가 포함되어 있습니다. 두 개의 반응 변수를 n개 개체에 대한 두 개의 독립된 평가로 볼 경우 평가자의 의견이 완전히 합치되면 카파 계수가 +1이 됩니다. 관측된 합치도가 우연에 의해 기대되는 합치도 크기를 초과하면 카파 계수가 양수가 되며 그 크기는 합치 강도를 반영합니다. 실제로는 흔치 않지만 관측된 합치도가 우연에 의해 기대되는 합치도 크기보다 작으면 카파가 음수가 됩니다. 카파의 최소값은 주변 비율에 따라 다르지만 항상 -1과 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 카파 계수의 점근 분산은 다음과 같이 추정됩니다.

$$\text{var} = \frac{A + B - C}{(1 - P_c)^2 n} \text{ 단, } A = \sum_i p_{ii}[1 - (p_{i+} + p_{+i})(1 - \hat{\kappa})]^2, \; B = (1 - \hat{\kappa})^2 \sum_{i \neq j} \sum p_{ij}(p_{+i} + p_{j+})^2 \text{ 및}$$

$$C = [\hat{\kappa} - P_c(1 - \hat{\kappa})]^2$$

자세한 내용은 Cohen 연구 자료(1960) 및 Fleiss et al. 연구 자료(1969)에서 확인하십시오.

Bowker 대칭성 검정에서 귀무가설은 2 x 2 테이블의 확률이 대칭성을 만족($p_{ij}=p_{ji}$)한다는 것입니다.

승산비에 대한 통계 상세 정보

분할 분석 플랫폼에서 승산비는 다음과 같이 계산됩니다.

$$\frac{p_{11} \times p_{22}}{p_{12} \times p_{21}}$$

여기서 p_{ij}는 2 x 2 테이블의 i번째 행, j번째 열의 개수입니다.

검정 보고서에 대한 통계 상세 정보

이 섹션에는 분할 분석 플랫폼의 검정 보고서에 대한 상세 정보가 포함되어 있습니다.

$R^2(U)$

$R^2(U)$ 는 다음과 같이 계산됩니다.

$$\frac{\text{negative log-likelihood for Model}}{\text{수정 합계에 대한 음의 로그 가능도}}$$

음의 로드 가능도 합계는 전체 표본에서 고정된 반응 비율을 적합시켜 구합니다.

검정

이 섹션에서는 두 카이제곱 검정에 대한 계산을 설명합니다.

가능도비 카이제곱 검정 통계량은 " 검정 " 테이블의 모형에 대한 음의 로그 가능도의 2 배로 계산됩니다. 일부 책에서는 이 통계량에 G^2 표기를 사용합니다. 두 개의 음의 로그 가능도 (**전체 모집단** 반응 확률을 사용한 로그 가능도와 **각 모집단** 반응 비율을 사용한 로그 가능도) 사이의 차이는 다음과 같이 정의됩니다.

$$G^2 = 2\left[\sum_{ij}(-n_{ij})\ln(p_j) - \sum_{ij} -n_{ij}\ln(p_{ij})\right] \text{ 단, } p_{ij} = \frac{n_{ij}}{N} \text{ 및 } p_j = \frac{N_j}{N}$$

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 제곱한 후 합해서 계산합니다 . Pearson 카이제곱 검정은 표본 크기가 매우 클 경우 빈도 수가 정규 분포를 따르는 경향이 있다는 특성을 이용합니다 . Pearson 카이제곱 통계량은 다음과 같이 정의됩니다 .

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

여기서 O는 관측된 셀 개수이고 E는 기대 셀 개수입니다. 합계에는 모든 셀이 포함됩니다. 여기서는 2 x 2 테이블에서 종종 수행되는 연속성 수정이 수행되지 않습니다.

대응 분석에 대한 통계 상세 정보

이 섹션에는 분할 분석 플랫폼의 대응 분석에 대한 상세 정보가 포함되어 있습니다 .

다음과 같은 특이값 분해 (SVD) 의 값을 나열합니다 .

$$\mathbf{D}_r^{-0.5}(\mathbf{P} - rc')\mathbf{D}_c^{-0.5} = \mathbf{U}Diag(\Lambda)\mathbf{V}'$$

다음은 각 요소에 대한 설명입니다 .

P = 개수를 총 빈도로 나눈 값의 행렬

r, *c* = **P** 의 행 및 열 합계

$\mathbf{D}_r$, $\mathbf{D}_c$ = 각각 *r* 및 *c* 값의 대각 행렬

Λ = 상세 정보 보고서에 보고된 특이값의 열 벡터

특이값 분해에 대한 자세한 내용은 **다변량 방법**의 에서 확인하십시오.

상세 정보 보고서의 행 좌표 (*rc*) 와 열 좌표 (*cc*) 는 다음과 같이 계산됩니다 .

$$rc = \mathbf{D}_r^{-0.5}\mathbf{U}Diag(\Lambda)$$

$$cc = \mathbf{D}_c^{-0.5}\mathbf{V}Diag(\Lambda)$$

8 장

로지스틱 분석

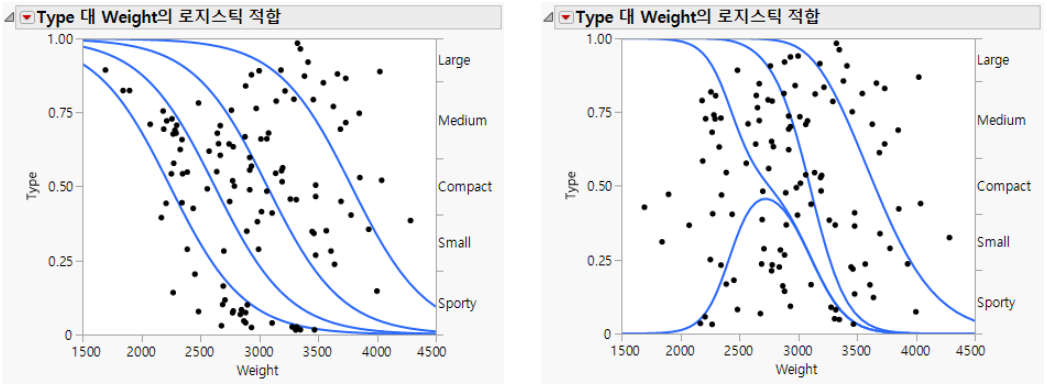
범주형 Y 변수와 연속형 X 변수 간의 관계 분석

로지스틱 플랫폼을 사용하여 로지스틱 회귀 모형을 연속형 X 변수가 있는 범주형 Y 변수에 적합시킵니다. ROC 곡선, 향상도 곡선 및 승산비 추정값을 볼 수 있습니다. 적합 모형은 X 변수의 각 값에 대해 추정 확률을 제공합니다. Y 변수의 특정 확률 값에 대한 X 값을 예측할 수 있는 역추정 예측을 수행할 수도 있습니다.

로지스틱 플랫폼은 X 로 Y 적합 플랫폼의 **명목형 또는 순서형 대 연속형** 분석법입니다. 이 플랫폼에서 명목형 반응과 순서형 반응은 다음과 같이 구분됩니다.

- 명목형 로지스틱 회귀 모형은 명목형 반응 변수의 수준 간에 확률을 분할하는 곡선 집합을 추정합니다. 명목형 로지스틱 회귀 모형의 예가 그림 8.1 의 오른쪽에 나와 있습니다.
- 순서형 로지스틱 회귀 모형은 순서형 반응 변수의 목표 수준보다 작거나 같을 확률을 추정합니다. 이 모형은 정렬된 범주에 대한 확률을 산출하기 위해 가로로 이동된 단일 로지스틱 곡선을 추정합니다. 이 모형은 덜 복잡하며 순서형 반응에 권장됩니다. 순서형 로지스틱 회귀 모형의 예가 그림 8.1 의 왼쪽에 나와 있습니다.

그림 8.1 순서형 및 명목형 로지스틱 회귀의 예



목차

로지스틱 플랫폼 개요 247
명목형 로지스틱 회귀의 예 247
로지스틱 플랫폼 시작 249
데이터 형식 250
로지스틱 보고서 250
로지스틱 그림 251
반복 보고서 252
전체 모형 검정 보고서 252
적합 상세 정보 보고서 253
모수 추정값 보고서 253
로지스틱 플랫폼 옵션 254
로지스틱 분석 보고서 256
ROC 곡선 256
역추정 예측 257
로지스틱 회귀의 추가 예 257
순서형 로지스틱 회귀의 예 257
로지스틱 그림의 예 258
ROC 곡선의 예 261
역추정 예측의 예 262
로지스틱 플랫폼에 대한 통계 상세 정보 264

로지스틱 플랫폼 개요

로지스틱 회귀를 사용하면 연속형 X 변수의 값을 기반으로 범주형 Y 변수 수준의 확률을 모델링할 수 있습니다. 로지스틱 회귀는 투약 반응 데이터나 구매 선택 데이터를 모델링하는 등 다양한 응용 분야에서 오랫동안 사용되어 왔습니다. 로지스틱 회귀 전문 교재(Hosmer and Lemeshow 1989) 외에도 범주형 통계 분야의 많은 교재(Agresti 1990)에서 로지스틱 회귀를 다루고 있습니다.

많은 로지스틱 회귀 설정에서, 특히 연속형 변수를 X 변수로 간주하고 범주형 변수를 Y 변수로 간주해서 거꾸로 작업하는 것을 선호하는 경우 판별 분석을 사용할 수도 있습니다. 하지만 판별 분석에서는 연속형 데이터가 고정된 회귀변수가 아니라 정규 분포를 따르는 확률 반응이라고 가정합니다. 자세한 내용은 **다변량 방법**의 에서 확인하십시오.

X로 Y 적합 플랫폼의 단순 로지스틱 회귀는 모형 적합 플랫폼의 범주형 반응에 대한 일반 모형을 단순화하고 그래픽 정보를 더 포함한 것입니다. 보다 복잡한 로지스틱 회귀 모형의 예는 **선형 모형 적합**의 에서 확인하십시오. **프로빗 분석**이라고도 하는 정규 분포 함수를 사용한 로지스틱 회귀에 대한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오.

명목형 로지스틱 회귀

명목형 로지스틱 회귀 모형은 Y 변수의 수준 중 하나를 연속형 X 변수의 평활 함수로 선택할 확률을 추정합니다. 적합 확률은 0에서 1 사이여야 하며, 주어진 X 변수 값에 대해 Y 변수의 전체 수준에서 합계가 1이 되어야 합니다.

로지스틱 확률도에서 세로 축은 확률을 나타냅니다. Y 변수의 수준이 k개인 경우 k - 1개의 평활 곡선이 Y 변수의 수준 간에 총 확률(1)을 분할합니다. 로지스틱 회귀의 적합 원칙은 발생하는 반응 사건에 적합된 확률(즉, 최대 가능도)의 음의 자연 로그 합을 최소화하는 것입니다.

순서형 로지스틱 회귀

Y 변수가 순서형인 경우 수정된 로지스틱 회귀가 적합에 사용됩니다. Y 변수의 각 수준과 같은 수준 또는 그 아래에서 반응이 나타날 누적 확률은 곡선으로 모델링됩니다. 각 수준의 곡선은 모양이 동일하지만 오른쪽 또는 왼쪽으로 이동됩니다.

순서형 로지스틱 모형은 r - 1개(여기서 r은 Y 변수의 수준 수)의 누적 로지스틱 비교에 대해 각각 다른 절편과 동일한 기울기를 적합시킵니다. 명목형 모형이라도 추정할 모수가 적어서 순서형 모형이 적합한 경우에는 순서형 모형을 사용하는 것이 좋습니다.

명목형 로지스틱 회귀의 예

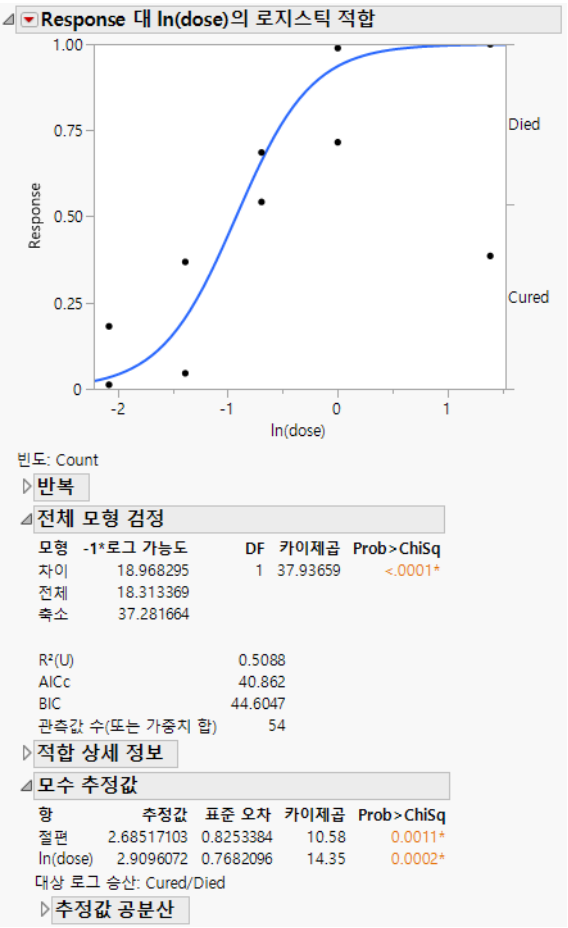
로지스틱 플랫폼을 사용하여 명목형 Y 반응 변수와 연속형 X 요인 변수 사이의 관계를 조사합니다 . 이 예의 데이터는 각각 12 마리의 토끼를 포함하는 5 개 그룹을 대상으로 연쇄상구균 박테리아를 주입한 실험에서 얻은 결과입니다 . 토끼의 신체 조직이 박테리아에 감염된 것으로 확인된

후에는 페니실린을 용량을 달리 해서 투여했습니다 . 투여량의 자연 로그가 토끼의 치료 여부에 영향을 미치는지 확인하려고 합니다 .

1. **도움말 > 샘플 데이터 폴더**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 .

   이전에 Count 열에 " 빈도 " 역할이 할당되었으므로 **빈도**에 Count 가 자동으로 입력됩니다 .
5. " 목표 수준 " 목록에서 Cured 를 선택합니다 .
6. **확인**을 클릭합니다 .

그림 8.2 명목형 로지스틱 보고서의 예



그림에는 적합된 모형이 ln(dose) 의 함수로 표시됩니다 . 적합된 모형은 예측 치료 확률입니다 . p 값이 매우 크며 , 이는 투여량이 토끼의 치료 여부에 상당한 영향을 미쳤음을 나타냅니다 . 오른쪽

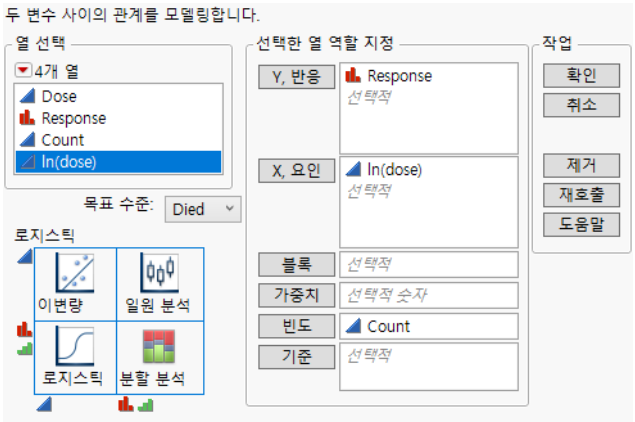
세로 축에 표시된 주변 분포는 X 변수와 Y 변수 간에 연관성이 없는 경우 Y 변수의 수준에 대한 확률을 나타냅니다.

팁 : 분석 대상 반응 수준을 변경하려면 시작 창의 " 목표 수준 " 옵션을 사용하거나 , " 값 순서 " 열 특성을 사용하십시오 .

로지스틱 플랫폼 시작

분석 > X 로 Y 적합을 선택하여 로지스틱 플랫폼을 시작할 수 있습니다 . X 로 Y 적합 시작 창은 네 가지 유형의 분석에 사용됩니다 . 순서형 또는 명목형 Y 변수와 연속형 X 변수를 입력하면 로지스틱 플랫폼이 시작됩니다 .

그림 8.3 로지스틱 시작 창



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 . 로지스틱 시작 창에는 다음 옵션이 포함되어 있습니다 .

Y, 반응 분석할 하나 이상의 반응 변수입니다 . 반응 변수는 종속 변수라고도 합니다 . 이러한 변수의 모델링 유형은 순서형 또는 명목형이어야 합니다 .

X, 요인 분석할 하나 이상의 예측 변수입니다 . 요인 변수는 독립 변수라고도 합니다 . 이러한 변수의 모델링 유형은 연속형이어야 합니다 .

블록 (로지스틱 분석에는 적용할 수 없음) 블록 변수를 지정하는 열입니다 .

가중치 데이터 테이블의 각 관측값에 대한 가중치를 포함하는 열입니다 . 값이 0 보다 큰 행만 분석에 포함됩니다 .

빈도 분석의 각 행에 빈도를 할당합니다 . 빈도 할당은 데이터를 요약할 때 유용합니다 .

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

목표 수준 (명목형 모델링 유형을 사용하는 이항 반응이 하나인 경우에만 사용 가능) 확률을 모델링할 반응 수준을 지정할 수 있습니다 .

팁 : 로지스틱 모형에서 목표값으로 사용되는 기본 수준은 수준의 순서를 기반으로 합니다 . 기본 목표값을 변경하려면 " 값 순서 " 열 특성을 사용하십시오 .

데이터 형식

로지스틱 플랫폼에서 데이터는 요약되지 않거나 요약된 데이터로 구성될 수 있습니다 .

요약되지 않은 데이터 각 관측값에 대한 하나의 행과 X 및 Y 값에 대한 열이 있습니다 .

요약된 데이터 각 행이 공통된 X 및 Y 값을 갖는 일련의 관측값을 나타냅니다 . 데이터 테이블에는 각 행에 대한 관측값 수가 포함된 빈도 열이 있어야 합니다 . 시작 창에서는 이 열을 " 빈도 " 열로 입력합니다 .

로지스틱 플랫폼의 요약된 데이터에 대한 예는 "명목형 로지스틱 회귀의 예"에서 확인하십시오.

참고 : X 로 Y 적합 시작 창은 연속형 , 순서형 및 명목형 모델링 유형의 열을 허용합니다 . 순서형 또는 명목형 Y, 반응 열과 연속형 X, 요인 열의 모든 쌍에 대해 로지스틱 플랫폼이 시작됩니다 . X 로 Y 적합 시작 창은 다른 열 유형 조합에 대해 이변량 , 일원 분석 또는 분할 분석 플랫폼을 실행합니다 .

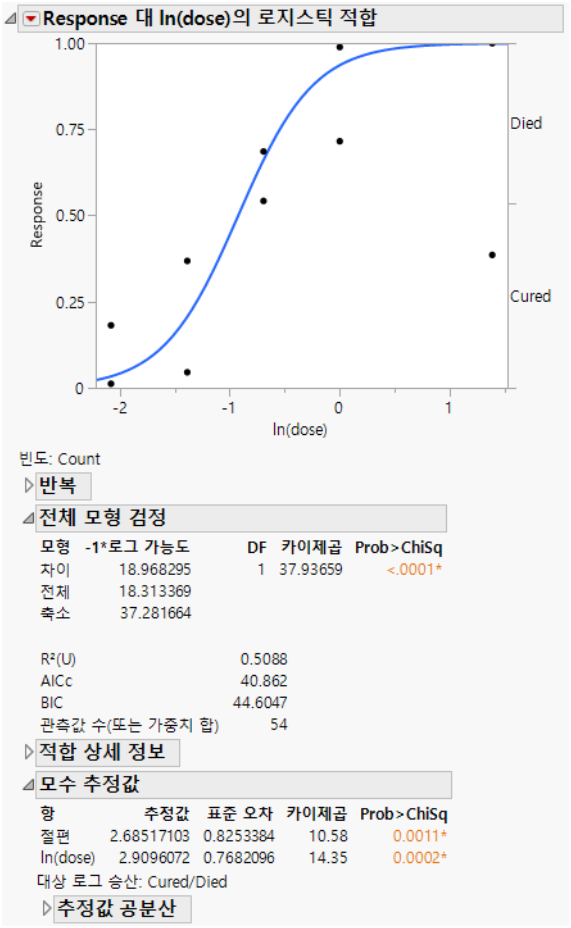
로지스틱 보고서

처음에는 "로지스틱" 보고서에 로지스틱 그림과 "반복", "전체 모형 검정", "적합 상세 정보" 및 "모수 추정값" 보고서가 포함됩니다. 빨간색 삼각형 메뉴 옵션을 사용하여 로지스틱 그림의 모양을 변경하고 다른 그림을 요청할 수 있습니다. 자세한 내용은 "로지스틱 플랫폼 옵션"에서 확인하십시오.

이 섹션에는 " 로지스틱 " 보고서의 다음 섹션에 대한 정보가 포함되어 있습니다 .

- " 로지스틱 그림 "
- " 반복 보고서 "
- " 전체 모형 검정 보고서 "
- " 적합 상세 정보 보고서 "
- " 모수 추정값 보고서 "

그림 8.4 로지스틱 보고서의 예



대화식으로 변수 바꾸기

연결된 데이터 테이블의 "열" 패널에서 변수를 선택하여 축으로 드래그하는 방법으로 변수를 바꿀 수도 있습니다.

로지스틱 그림

로지스틱 플랫폼에서 로지스틱 확률도는 모형 적합을 나타냅니다. 그림 오른쪽의 세로 축은 Y 변수의 각 수준에 대한 관측값 수에 비례하여 분할됩니다. 그림에 Y 변수의 한 수준을 제외한 모든 수준에 대한 곡선이 있습니다. 곡선은 Y 변수의 해당 수준까지 누적된 예측 확률입니다. 그림 왼쪽의 세로 축은 확률 척도를 사용합니다. 특정 수준에 대한 예측 확률은 해당 수준의 곡선과 수준 아래 곡선 사이의 세로 거리로 측정됩니다.

로지스틱 그림에서 점은 데이터 테이블의 관측값을 나타냅니다. 각 점의 가로 위치는 해당 점에 대한 연속형 X 변수의 값에 따라 결정됩니다. 각 점의 세로 위치는 해당 점에 대한 범주형 Y 변수

의 값에 따라 결정됩니다. 각 점은 해당 수준에 대한 곡선과 다음 하위 곡선 사이의 세로 위치에 무작위로 배치됩니다. 이렇게 점을 세로로 지터링하면 점들이 가장 조밀한 곳을 보다 쉽게 확인할 수 있지만 이 경우 세로 위치가 세로 축의 값과 똑바로 일치하지 않습니다. 고정된 난수 시드 값이 사용되므로 세로 위치는 동일한 모형을 여러 번 적합시켜도 달라지지 않습니다.

반복 보고서

로지스틱 플랫폼의 "반복" 보고서에는 각 반복과 로지스틱 모형 수렴 여부를 결정하는 평가 기준이 표시됩니다. "반복" 보고서는 명목형 로지스틱 회귀 모형의 경우에만 표시됩니다.

전체 모형 검정 보고서

로지스틱 플랫폼의 "전체 모형 검정" 보고서에는 모든 반응 확률에 대해 단순히 상수를 사용하는 것보다 모형이 더 잘 적합되는지 여부가 표시됩니다. 이 보고서는 연속형 반응 모형에 대한 분산 분석 보고서와 유사합니다. 이 보고서에 표시된 검정은 로지스틱 회귀 모형이 데이터를 얼마나 잘 적합시키는지 평가하는 가능도비 카이제곱 검정입니다.

관측된 확률의 음의 자연 로그 합을 음의 로그 가능도(**–1*로그 가능도**)라고 합니다. 범주형 데이터에 대한 음의 로그 가능도는 연속형 데이터에 대한 제곱합과 유사합니다. 데이터에 적합된 모형과 각 수준에 대해 확률이 동일한 모형 간의 음의 로그 가능도 차이의 두 배가 카이제곱 통계량입니다. 이 검정 통계량은 X 변수의 값이 Y 변수의 수준과 연관이 없다는 가설을 위한 값입니다.

R^2(U)(일부 경우에는 R^2으로 표기) 값은 0에서 1 사이입니다. R^2 값이 높을수록 모형이 더 적합하다는 것을 나타냅니다. 범주형 모형에서는 R^2 값이 높은 경우가 드뭅니다.

"전체 모형 검정" 보고서에는 다음 열이 포함됩니다.

모형 변동 소스의 라벨입니다.

- **차이** **전체** 모형과 **축소** 모형 간의 차이입니다. 이 모형은 X 변수가 적합에 기여하는 유의성을 측정하는 데 사용됩니다.
- **전체** 절편과 X 변수를 포함하는 완전 모형입니다.
- **축소** 절편 모수만 포함하는 모형입니다.

-1* 로그 가능도 각 모형에 대한 변동을 측정하는 음의 로그 가능도입니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

DF 전체 모형과 축소 모형 간의 차이에 대한 DF(자유도)입니다.

카이제곱 모형이 전체 표본에서 고정된 반응 비율보다 더 잘 적합되지 않는다는 가설에 대한 가능도비 카이제곱 검정 통계량입니다. 이 검정 통계량은 적합 모형과 절편만 있는 축소 모형 간의 음의 로그 가능도 차이의 두 배로 계산됩니다. 자세한 내용은 "로지스틱 플랫폼에 대한 통계 상세 정보"에서 확인하십시오.

Prob>ChiSq 지정된 모형이 절편만 포함된 모형보다 더 잘 적합되지 않을 경우 카이제곱 값이 더 클 확률입니다. 대개 이 확률이 0.05 미만이면 모형이 유의한 것으로 판단됩니다.

R^2(U) 모형 적합에 기인한 전체 불확실성의 비율로, 차이의 음의 로그 가능도 값을 축소의 음의 로그 가능도 값으로 나눈 값으로 정의됩니다. R^2(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 R^2(U) 가 작은 경향이 있습니다. 자세한 내용은 "로지스틱 플랫폼에 대한 통계 상세 정보" 에서 확인하십시오.

R^2(U) 는 불확도 계수인 U 또는 McFadden 유사 R^2 이라고도 합니다.

AICc 수정된 Akaike 정보 기준입니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

BIC Bayes 정보 기준입니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

관측값 수 (또는 가중치 합) 표본의 총 관측값 수입니다. " 모형 적합 " 창에서 " 빈도 " 또는 " 가중치 " 열이 지정되는 경우 이 값은 빈도 또는 가중치 역할에 할당된 열 값의 합입니다.

적합 상세 정보 보고서

로지스틱 플랫폼의 " 적합 상세 정보 " 보고서에는 각 적합 측도의 대수적 정의를 포함하여 다음 통계량의 값이 포함됩니다.

엔트로피 R^2 적합된 모형과 상수 확률 모형의 로그 가능도를 비교합니다. 이 값은 R^2(U) 와 동일합니다. 자세한 내용은 " 로지스틱 플랫폼에 대한 통계 상세 정보 " 에서 확인하십시오.

일반화 R^2 일반 회귀 모형에 적용할 수 있는 측도입니다. 이 값은 가능도 함수 L 을 기반으로 하며 최대값이 1 이 되도록 척도화됩니다. 일반화 R^2 측도는 표준 최소 제곱 설정 시 연속형 정규 반응에 대한 기존 R^2 로 단순화됩니다. 일반화 R^2 을 Nagelkerke 또는 Craig 와 Uhler R^2 이라고도 하는데, 이는 Cox-Snell 유사 R^2 을 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.

RASE 제곱근 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.

N 관측값 수입니다.

엔트로피 R^2 및 일반화 R^2 의 경우 값이 1 에 가까울수록 더 나은 적합을 나타냅니다. 평균 -Log p, RASE, 평균 절대 편차 및 오분류 비율의 경우 값이 작을수록 더 나은 적합을 나타냅니다.

모수 추정값 보고서

명목형 로지스틱 모형은 $k - 1$ 개의 로지스틱 비교 각각에 대해 절편 및 기울기 모수를 적합시킵니다. 여기서 k 는 반응 수준의 개수입니다. 순서형 로지스틱 모형은 각 비교에 대해 단일 기울

기를 적합시킵니다 . 로지스틱 플랫폼의 " 모수 추정값 " 보고서에 이러한 추정값이 나열됩니다 . 각 모수 추정값을 개별적으로 검토하고 검정할 수 있지만 이는 거의 필요하지 않습니다 .

항 로지스틱 모형의 각 모수입니다 . 반응 변수의 마지막 수준을 제외하고 각 수준별로 요인에 대한 절편 항과 기울기 항이 하나씩 있습니다 . 순서형 로지스틱 모형에는 기울기 항이 하나만 포함됩니다 .

추정값 로지스틱 모형으로 구한 모수 추정값입니다 .

표준 오차 각 모수 추정값의 표준 오차입니다 . 표준 오차는 각 항을 0 과 비교하는 통계적 검정을 계산하는 데 사용됩니다 .

카이제곱 각 모수가 0 이라는 가설에 대한 Wald 검정입니다 . Wald 카이제곱은 (추정값 / 표준오차)^2 으로 계산됩니다 .

Prob>ChiSq 카이제곱 검정에 대한 관측된 유의 확률 (*p* 값) 입니다 .

추정값 공분산

모수 추정값의 추정 분산과 모수 추정값 간의 추정 공분산을 보고합니다. 분산 추정값의 제곱근은 "모수 추정값" 테이블의 "표준 오차" 열에 표시된 값과 동일합니다.

참고 : " 추정값 공분산 " 보고서는 명목형 반응 변수에 대해서만 표시되고 순서형 반응 변수에 대해서는 표시되지 않습니다 .

로지스틱 플랫폼 옵션

" 로지스틱 적합 " 의 빨간색 삼각형 메뉴에는 로지스틱 그림 및 로지스틱 적합 보고서에 대한 옵션이 포함되어 있습니다 .

참고 : " 그룹 적합 " 메뉴는 Y 또는 X 변수를 여러 개 지정한 경우에만 표시됩니다 . " 그룹 적합 " 메뉴 옵션을 사용하여 보고서를 배열하거나 R^2 을 기준으로 정렬할 수 있습니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

" 로지스틱 적합 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

승산비 (수준이 두 개인 반응에만 사용 가능) 모수 추정값 보고서에서 승산비가 포함된 열을 추가하거나 제거합니다 . 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오 .

역추정 예측 (2 수준 명목형 반응에만 사용 가능) 하나 이상의 반응 변수 값에 대해 예측 변수 값을 예측할 수 있습니다 . 자세한 내용은 " 역추정 예측 " 에서 확인하십시오 .

로지스틱 그림 로지스틱 그림을 표시하거나 숨깁니다 . 자세한 내용은 " 로지스틱 그림 " 에서 확인하십시오 .

그림 옵션 로지스틱 그림에 영향을 주는 다음 옵션이 포함되어 있습니다 .

점 표시 로지스틱 그림에 점을 표시하거나 숨깁니다 .

비율 곡선 표시 로지스틱 그림에 비율 곡선을 표시하거나 숨깁니다 . 비율 곡선은 X 변수의 각 값에 대해 여러 개의 점이 있는 경우에만 유용합니다 . 이 경우 각 값의 적절한 비율 추정값을 구하고 이 비율을 적합된 로지스틱 곡선과 비교할 수 있습니다 . 퇴화 점 (일반적으로 0 또는 1 일 때) 이 너무 많아지지 않도록 JMP 에서는 *x* 값에 점이 세 개 이상 있는 경우에 비율 값만 표시합니다 .

선 색상 그림의 곡선 색상을 선택할 수 있습니다 .

ROC 곡선 모형에 대한 ROC(Receiver Operating Characteristic) 곡선을 표시하거나 숨깁니다 . ROC 곡선은 X 변수의 각 값에 대한 민감도 대 (1 - 특이도) 를 보여 주는 그림입니다 . 자세한 내용은 "ROC 곡선 " 에서 확인하십시오 .

향상도 곡선 모형에 대한 향상도 곡선을 표시하거나 숨깁니다 . 향상도 곡선은 모형의 예측 능력을 보여 줍니다 . 향상도 곡선은 향상도 대 관측값 비율을 표시합니다 . 향상도 곡선에는 고유한 각 예측 확률 값에 대한 점이 포함됩니다 . 반응 수준의 각 예측 확률은 예측 확률이 고유 예측 확률 값보다 크거나 같은 관측값의 비율을 정의합니다 . 특정 반응 수준에 대한 **향상도** 값은 관측 반응의 전체 비율 대비 해당 부분의 관측 반응 비율을 나타내는 값입니다 . 향상도 곡선에 대한 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오 .

확률 계산식 저장 새 열을 데이터 테이블에 저장합니다 . 새 열에는 모형에서 예측하는 확률의 계산식이 포함됩니다 . 다음 항목이 새 열에 포함됩니다 .

- 요인 변수의 선형 결합 (일반적으로 로짓이라고 함) 계산식
- 반응 수준 확률에 대한 예측 계산식
- 최대 확률 분류 반응에 대한 예측 계산식

다음 옵션에 대한 자세한 내용은 **JMP 사용**의 에서 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 .

플랫폼 환경 설정 현재 플랫폼 환경 설정을 보거나 , 현재 JMP 보고서의 설정과 일치하도록 플랫폼 환경 설정을 업데이트할 수 있는 옵션이 포함되어 있습니다 .

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 .

로지스틱 분석 보고서

로지스틱 플랫폼을 사용하면 연속형 X 변수와 범주형 Y 변수 사이의 관계를 탐색하기 위해 시각화를 생성하고 모형을 적합시킬 수 있습니다. 이 섹션에는 특정 분석 옵션에 대해 생성되는 보고서 관련 정보가 포함되어 있습니다.

- "ROC 곡선"
- "역추정 예측"

ROC 곡선

로지스틱 플랫폼에는 로지스틱 회귀 모형에 대해 ROC(Receiver Operating Characteristic) 곡선을 적합시키는 옵션이 포함되어 있습니다. 로지스틱 플랫폼의 "ROC 곡선" 옵션은 플랫폼 시작 창의 "목표 수준"을 ROC 곡선의 Positive 반응 수준으로 사용합니다. 자세한 내용은 "ROC 곡선의 예"에서 확인하십시오.

예를 들어 진단 측도인 X 변수의 값이 있고 다음을 나타내는 X 변수의 임계값을 결정하려고 한다고 가정해 보겠습니다.

- X 변수의 값이 임계보다 크면 조건이 존재하는 것입니다.
- X 변수의 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 검정을 검토합니다. 그런 다음 해당 비율을 그림으로 표시합니다. 이상적인 상태는 True Negative(진음성)와 True Positive(진양성)를 가장 잘 구분하는 X 변수 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.

두 가지 표준 정의가 의학에서 사용됩니다.

- **민감도**는 X 변수의 특정 값이 존재하는 조건을 올바르게 예측할 확률입니다. 특정 *x* 가 조건의 존재를 잘못 예측할 확률은 1 – 민감도입니다.
- **특이도**는 검정을 통해 조건이 존재하지 않음을 올바르게 예측할 확률입니다.

ROC 곡선은 각 X 변수 값에 대한 민감도 대 (1 - 특이도)의 그림입니다. ROC 곡선 아래의 면적은 곡선에 포함된 정보를 요약하는 데 사용되는 일반적인 지표입니다.

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 지납니다. 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 민감도와 특이도의 합을 최대화하는 경계 점을 나타냅니다.

역추정 예측

로지스틱 플랫폼을 사용하면 평소와 반대 방향으로 예측하는 역추정 예측을 수행할 수 있습니다. 역추정 예측은 X 요인 변수의 값에서 Y 반응 변수의 값을 예측하는 대신 Y 반응 변수의 값에서 X 요인 변수의 값을 예측합니다. 로지스틱 회귀에서는 Y 반응 변수의 값을 예측하는 대신 Y 반응 변수의 각 수준에 기인하는 확률을 예측합니다. 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 **선형 모형 적합**의 에서 확인하십시오.

참고 : 자세한 내용은 " 역추정 예측의 예 " 에서 확인하십시오 .

로지스틱 회귀의 추가 예

이 섹션에는 로지스틱 플랫폼을 사용한 예가 포함되어 있습니다 .

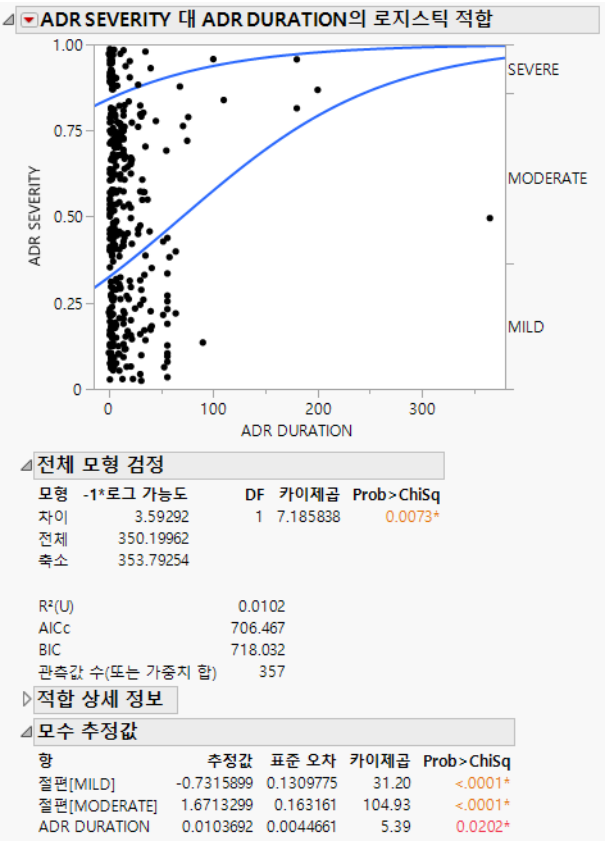
- " 순서형 로지스틱 회귀의 예 "
- " 로지스틱 그림의 예 "
- "ROC 곡선의 예 "
- " 역추정 예측의 예 "

순서형 로지스틱 회귀의 예

이 예에서는 순서형 반응과 연속형 요인 사이의 관계를 검토하는 방법을 보여 줍니다 . 이 예에서 이상 사례의 심각도를 이상 사례 지속 기간의 함수로 모델링하려고 합니다 .

1. **도움말 > 샘플 데이터 폴더**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



그림에서 데이터 표식은 가로 축의 X 좌표 값 및 해당 Y 값에 대한 곡선 아래 임의의 세로 위치에 표시됩니다.

"전체 모형 검정" 보고서와 "모수 추정값" 보고서에 대한 자세한 내용은 "로지스틱 보고서"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 보여줍니다. 이들은 항상 단순 증가합니다.

로지스틱 그림의 예

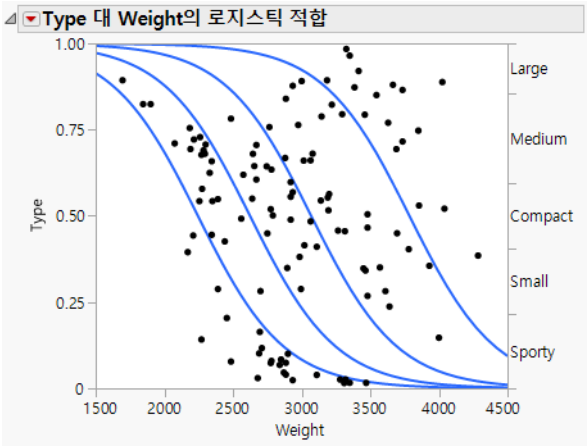
이 예에서는 116 대의 자동차에 대해 중량을 사용하여 자동차 종류를 예측하려고 한다고 가정해 보겠습니다 . 자동차 종류는 작은 것부터 큰 것 순서로 "Sporty", "Small", "Compact", "Medium" 또는 "Large" 중 하나일 수 있습니다 .

1. **도움말 > 샘플 데이터 폴더**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



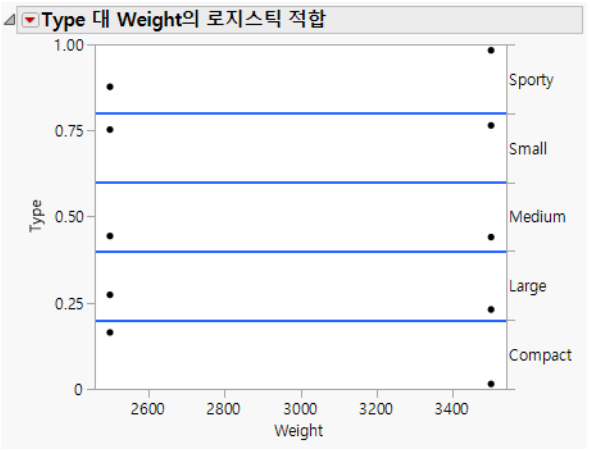
다음 관측값에 유의하십시오 .

- 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

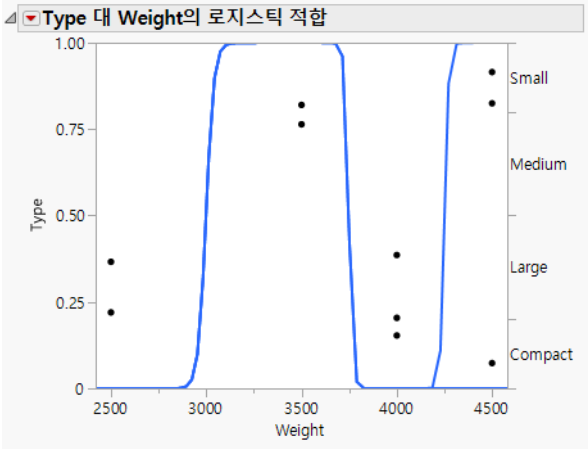
	Type	Weight
1	Large	2500
2	Medium	2500
3	Compact	2500
4	Small	2500
5	Sporty	2500
6	Large	3500
7	Medium	3500
8	Compact	3500
9	Small	3500
10	Sporty	3500



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 Weight가 Type을 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

	Type	Weight
1	Small	2500
2	Small	2500
3	Compact	3500
4	Compact	3500
5	Medium	4000
6	Medium	4000
7	Medium	4000
8	Large	4500
9	Large	4500
10	Large	4500



이 경우에는 모수 추정값이 매우 커지고 회귀 보고서에 **불안정**이라는 라벨이 표시됩니다. 이러한 경우에는 Firth 편향 조정 추정값을 사용하는 일반화 선형 모형 분석법을 사용할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

참고 : 그림 8.7 및 그림 8.8 의 그림을 다시 생성하려면 먼저 여기에 표시된 데이터 테이블을 생성한 다음 이 섹션의 처음에 나오는 7~10 단계를 수행해야 합니다 .

ROC 곡선의 예

로지스틱 플랫폼을 사용하여 ROC 곡선을 적합시킵니다 .

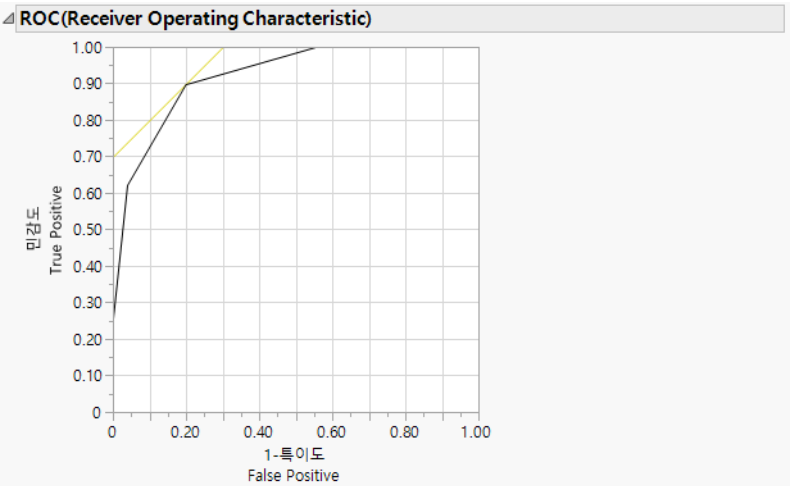
1. **도움말 > 샘플 데이터 폴더**를 선택하고 Penicillin.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .
4. ln(dose) 을 선택하고 **X, 요인**을 클릭합니다 .

   빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.
5. " 목표 수준 " 목록에서 Cured 를 선택합니다 .
6. **확인**을 클릭합니다 .
7. " 로지스틱 적합 " 의 빨간색 삼각형 메뉴를 클릭하고 **ROC 곡선**을 선택합니다 .

참고 : 이 예에서는 명목형 반응의 ROC 곡선을 보여 줍니다. 순서형 ROC 곡선에 대한 자세한 내용은 **예측 및 전문 모델링**의 에서 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. 이 점에 해당하는 X 값은 민감도와 특이도의 합을 최대화하는 임계입니다. ROC 곡선 그림의 노란색 선은 이 점에 대한 접선입니다.

그림 8.9 ROC 곡선 및 테이블의 예



ROC 테이블

X	확률	1-특이도	민감도	민감도-(1-특이도)		True Pos	True Neg	False Pos	False Neg
.	.	0.0000	0.0000	0.0000		0	25	0	29
1.38629	0.9988	0.0000	0.2414	0.2414		7	25	0	22
0.00000	0.9361	0.0400	0.6207	0.5807		18	24	1	11
-0.69315	0.6611	0.2000	0.8966	0.6966	*	26	20	5	3
-1.38629	0.2061	0.5600	1.0000	0.4400		29	11	14	0
-2.07944	0.0334	1.0000	1.0000	0.0000		29	0	25	0
-2.07944	0.0334	1.0000	1.0000	0.0000		29	0	25	0

ROC 곡선이 그림의 왼쪽 위 사분면에 속하고 AUC가 0.5보다 크므로 모형에 예측 능력이 있다는 결론을 내릴 수 있습니다.

역추정 예측의 예

로지스틱 플랫폼에서 십자기호 도구를 사용하여 역추정 예측의 근사값을 시각적으로 산출할 수 있습니다. 반응 수준이 정확히 두 개뿐인 경우 역추정 예측 옵션을 사용하여 정확한 역추정 예측을 요청할 수 있습니다. 역추정 예측은 Y 변수의 목표 수준에 대해 주어진 확률에 해당하는 X 변수의 추정값과 해당 추정값에 대한 신뢰 구간을 제공합니다.

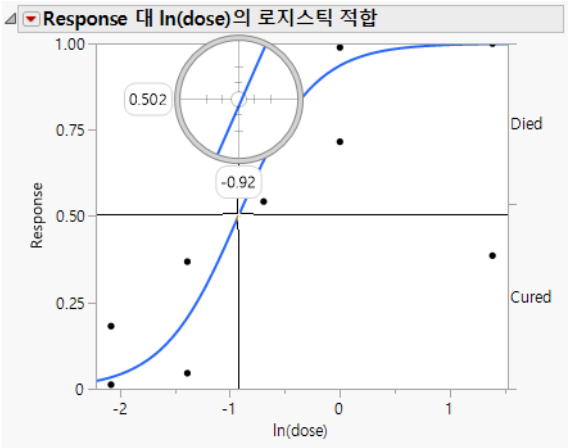
1. **도움말 > 샘플 데이터 폴더**를 선택하고 Penicillin.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Response 를 선택하고 **Y, 반응**을 클릭합니다 .
4. ln(dose) 을 선택하고 **X, 요인**을 클릭합니다 .

   빈도에는 Count가 자동으로 채워집니다. 이미 Count에 빈도 역할이 할당되었기 때문입니다.
5. " 목표 수준 " 목록에서 Cured 를 선택합니다 .
6. **확인**을 클릭합니다 .

십자기호 도구

7. 십자기호 도구를 클릭합니다 .
8. 십자기호가 그림 왼쪽에 있는 세로 (Response) 확률 축의 약 0.5 지점에 오도록 합니다 .
9. 십자기호의 교차점을 예측선 쪽으로 이동하고 십자기호 아래에 표시되는 ln(dose) 값을 읽습니다 .

그림 8.10 로지스틱 그림에서 사용되는 십자기호 도구의 예

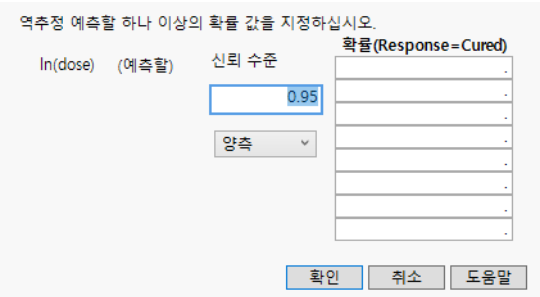


이 예에서 ln(dose)이 약 -0.9인 토끼는 치료될 가능성과 죽을 가능성이 동일합니다. 정확한 역추정 예측을 구하려면 "역추정 예측" 옵션을 사용하십시오.

역추정 예측 옵션

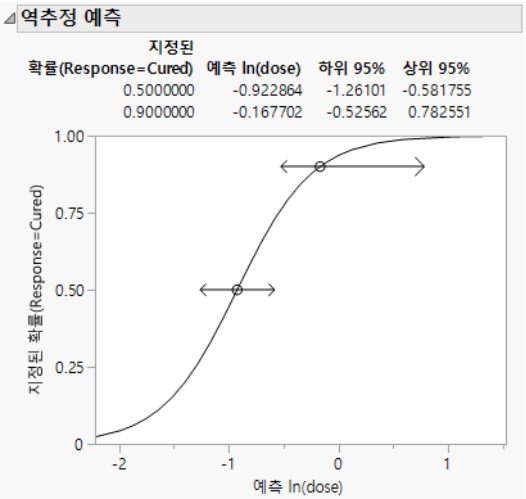
10. " 로지스틱 적합 " 의 빨간색 삼각형 메뉴를 클릭하고 **역추정 예측**을 선택합니다 (그림 8.11).
11. **신뢰 수준**에 0.95 를 입력합니다 .
12. 신뢰 구간으로 **양측**을 선택합니다 .
13. 관심 있는 반응 확률을 요청합니다 . 이 예에서는 0.5 와 0.9 를 입력합니다 . 즉 , 치료 확률이 0.5 와 0.9 인 ln(dose) 의 값을 요청합니다 .

그림 8.11 역추정 예측 창



14. **확인**을 클릭합니다.

그림 8.12 역추정 예측 그림의 예



"역추정 예측" 보고서에 X 변수 값의 추정값과 해당 신뢰 구간이 테이블 형식 및 그래픽 형식으로 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^2으로 표기되며 다음과 같이 정의됩니다.

$$G^2 = 2(\sum -\ln p(\text{background}) - \sum -\ln p(\text{모형}))$$

합계는 모든 셀이 아니라 모든 관측값에 대한 값입니다.

$R^2(U)$

$R^2(U)$ 는 다음과 같이 계산됩니다 .

$$\frac{\text{차이 모형에 대한 음의 로그 가능도}}{\text{축소 모형에 대한 음의 로그 가능도}}$$

이러한 통계량은 "전체 모형 검정" 보고서에 나타납니다.

참고 : $R^2(U)$ 를 McFadden 유사 R^2 라고도 합니다 .

9 장

테이블 생성
대화식으로 요약 테이블 생성

테이블 생성 플랫폼을 사용하여 기술 통계량의 요약 테이블이나 피벗 테이블을 대화식으로 생성할 수 있습니다. 테이블 생성 플랫폼은 요약 데이터를 테이블 형식으로 표시하는 쉽고 유연한 방법입니다. 테이블은 데이터 테이블 열을 테이블 생성의 행 또는 열로 할당한 다음 원하는 요약 통계량을 할당하여 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	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				
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 269
테이블 생성 플랫폼 시작 274
테이블 생성 플랫폼 대화상자 창 275
통계량 추가 276
테이블 생성 플랫폼 보고서 279
분석 열 280
그룹화 열 280
열 및 행 테이블 281
테이블 편집 282
테이블 생성 플랫폼 옵션 283
테스트 빌드 패널 표시 284
마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴 284
테이블 생성 플랫폼의 추가 예 285
예: 여러 테이블 생성 및 내용 재배열 285
예: 여러 열을 단일 테이블에 결합 289
예: 페이지 열 사용 291
그룹화 열 쌓기의 예 292

테이블 생성 플랫폼의 예

데이터를 테이블 형식으로 요약하는 방법을 알아봅니다. 이 예에는 남학생과 여학생의 키 측정값이 포함된 데이터가 있습니다. 남학생 및 여학생의 평균 키와 두 성별 모두의 키 집계를 보여주는 테이블을 생성하려고 합니다.

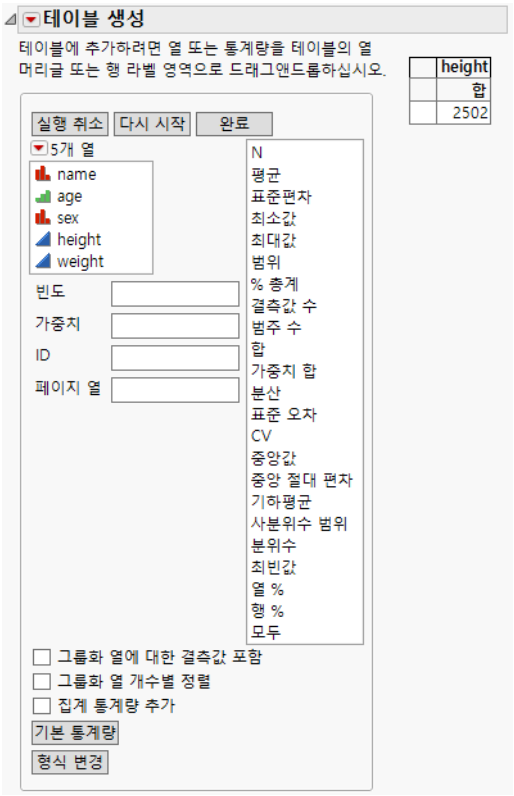
그림 9.2 평균 키를 보여 주는 테이블

	height
sex	평균
F	60.9
M	63.9
모두	62.6

1. **도움말 > 샘플 데이터 폴더**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 .

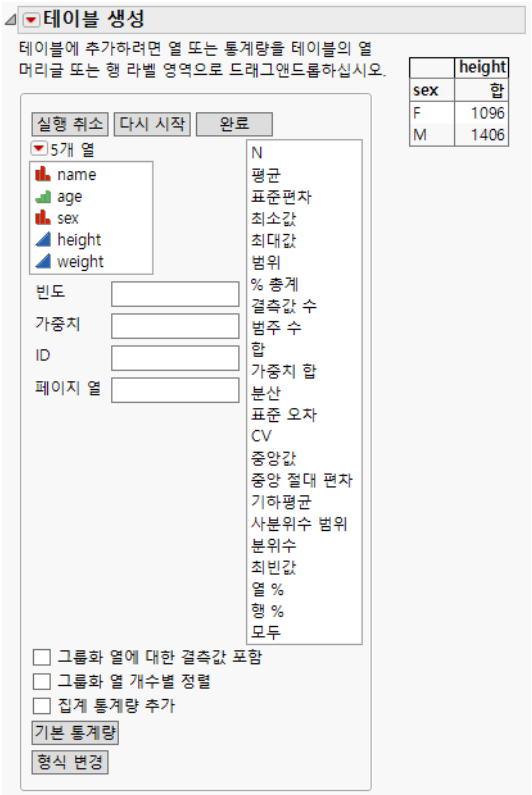
그림 9.3 height 변수 추가



성별별로 통계량을 살펴볼 것이므로 sex 변수를 나란히 표시하려고 합니다.

4. sex 를 클릭하고 숫자 "2502" 옆의 빈 셀로 드래그합니다 .

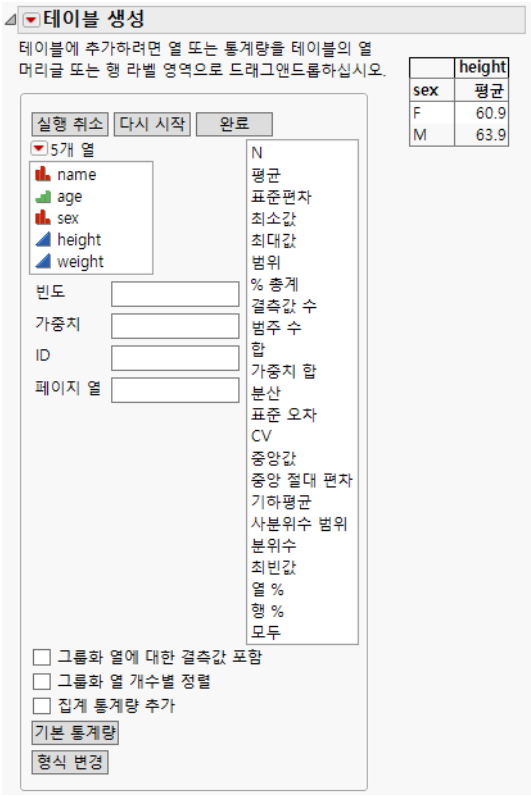
그림 9.4 sex 변수 추가



합 대신 평균을 표시하려고 합니다.

5. **평균**을 클릭하고 **합** 위로 드래그합니다.

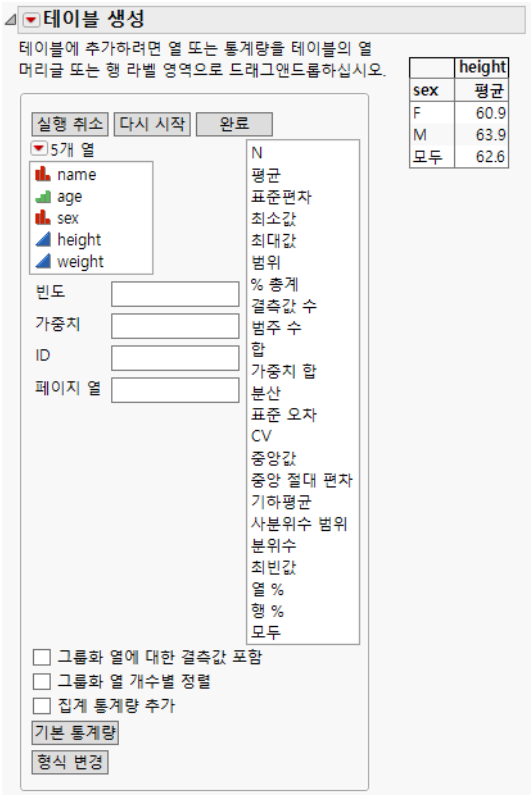
그림 9.5 평균 통계량 추가



또한 남학생과 여학생의 결합 평균을 표시하려고 합니다.

6. **모두**를 클릭하고 sex 위로 드래그합니다 . 또는 단순히 **집계 통계량 추가** 체크박스를 선택해도 됩니다 .

그림 9.6 전체 (" 모두 ") 통계량 추가



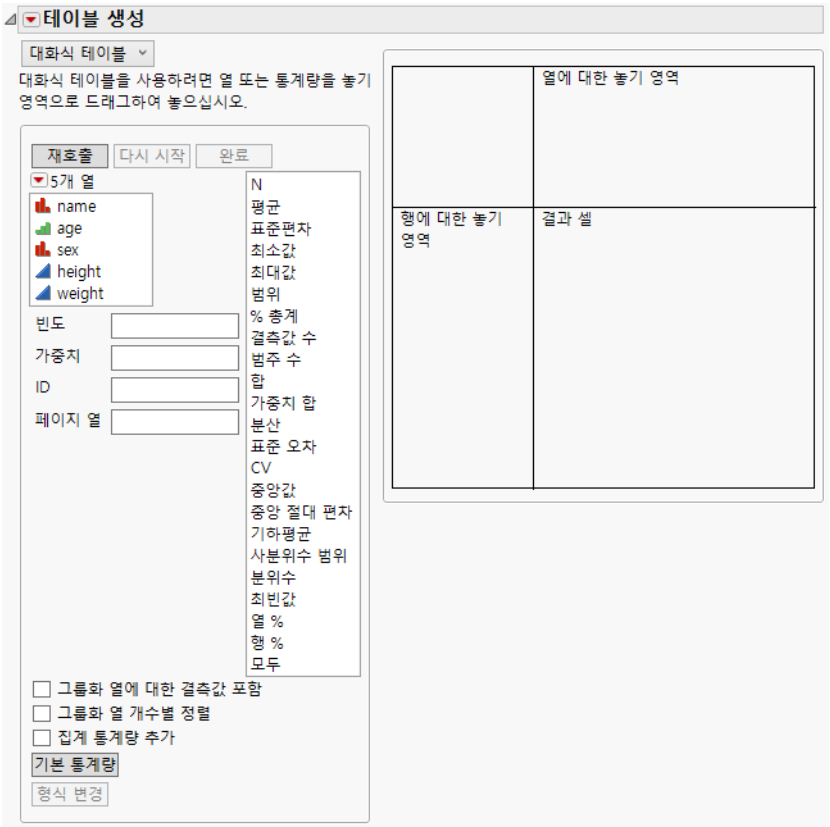
7. (선택 사항) **완료**를 클릭합니다 .

완료된 테이블에 여학생 및 남학생의 평균 키와 둘 모두의 결합 평균 키가 표시됩니다.

테이블 생성 플랫폼 시작

분석 > 테이블 생성을 선택하여 테이블 생성 플랫폼을 시작할 수 있습니다 .

그림 9.7 대화식 테이블 생성



빨간색 삼각형 옵션에 대한 자세한 내용은 "테이블 생성 플랫폼 옵션"에서 확인하십시오. "열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 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 " 테이블 생성 플랫폼 대화상자 창 " 에서 확인하십시오 .

통계량 옵션 표준 통계량을 나열합니다 . 목록의 통계량을 테이블로 드래그하여 테이블에 포함할 수 있습니다 . 자세한 내용은 " 통계량 추가 " 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 . 그러지 않으면 JMP 에서는 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 (macOS 의 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 값이 있는 데이터 테이블 열을 식별합니다 .

ID 고유 발생을 식별하는 데 사용할 값이 있는 데이터 테이블 열을 식별합니다 .

페이지 열 명목형 또는 순서형 열의 각 범주에 대한 별도의 테이블을 생성합니다 . 자세한 내용은 " 예 : 페이지 열 사용 " 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 " 에서 확인하십시오 .

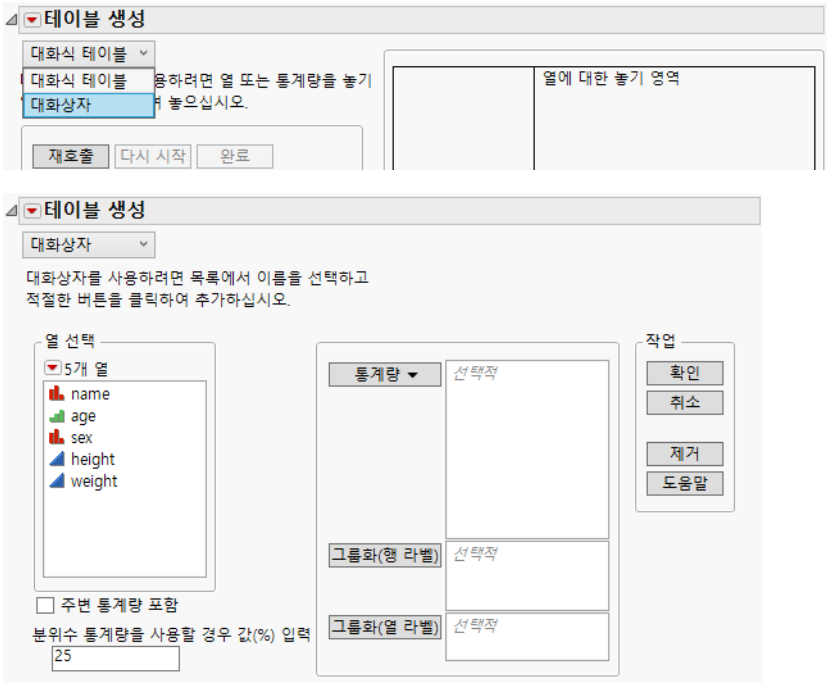
그림 척도 변경 (빨간색 삼각형 메뉴에서 **차트 표시**를 선택한 경우에만 표시됨) 균등 사용자 척도를 지정할 수 있습니다 .

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 .

테이블 생성 플랫폼 대화상자 창

드래그하여 놓기 작업을 사용하여 대화식으로 테이블을 생성하지 않으려면 테이블 생성 플랫폼의 " 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 비율만큼의 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" 에서 확인하십시오.

그룹화 (행 라벨) 행 라벨로 사용할 열을 선택합니다.

그룹화 (열 라벨) 열 라벨로 사용할 열을 선택합니다.

통계량 추가

테이블 생성 플랫폼은 제어판에 나타나는 표준 통계량 목록을 지원합니다. 열을 추가할 때와 마찬가지로 이 목록의 키워드를 테이블로 드래그할 수 있습니다. 다음 사항에 유의하십시오.

팁 : 열과 통계량을 모두 동시에 선택하고 테이블로 드래그할 수 있습니다.

- 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 " 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 **가중치 합**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 , 문자는 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 .

가중치 합 열에 있는 모든 **가중치** 값의 합을 제공합니다 . **가중치** 역할에 할당된 열이 없는 경우 , **가중치 합**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 데이터의 75% 가 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 .

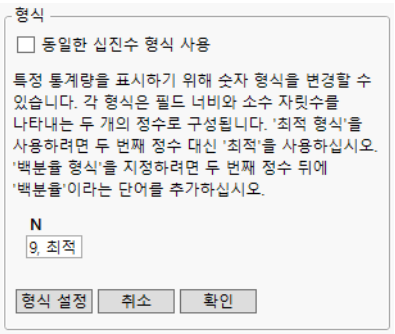
숫자 형식 변경

각 셀의 계산식은 분석 열 및 통계량에 따라 달라집니다 . 개수의 경우 기본 형식에는 소수점 이하 자릿수가 없습니다 . 일부 통계량으로 정의된 각 셀의 경우 , JMP 에서는 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



테이블 생성 플랫폼 보고서

테이블 생성 플랫폼 보고서는 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
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	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				
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 " 및 " 열 및 행 테이블 " 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.

- JMP 에서는 모델링 유형을 사용하여 열의 역할이 결정됩니다. 연속형 열은 분석 열로 할당됩니다. 자세한 내용은 "분석 열"에서 확인하십시오. 순서형 또는 명목형 열은 그룹화 열로 간주됩니다. 자세한 내용은 "그룹화 열"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"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 JMP 에서는 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

테이블 생성 플랫폼의 분석 열은 통계량을 계산할 숫자(연속형) 열입니다. 테이블 생성 기능은 그룹화 열로 구성된 각 범주에 대해 분석 열의 통계량을 계산합니다.

참고 : 모든 분석 열은 행 테이블이나 열 테이블에서 동일한 차원으로 있어야 합니다.

그룹화 열

테이블 생성 플랫폼의 그룹화 열은 데이터를 여러 정보 범주로 분류하는 데 사용할 열(명목형 또는 순서형)입니다. 이 열에는 문자, 정수 또는 십진수가 포함될 수 있지만 고유 값의 개수는 제한됩니다.

다음 사항에 유의하십시오 .

- 그룹화 열이 내포된 경우 테이블 생성 플랫폼은 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 .
- 그룹화 열이 내포된 경우 내포된 값을 쌓아서 테이블의 가로 공간을 절약할 수 있습니다 . 그룹화 열을 쌓으려면 그룹화 열을 마우스 오른쪽 버튼으로 클릭하고 " 그룹화 열 쌓기 " 옵션을 선택합니다 . 쌓인 그룹화 열은 열 값의 계층적 내포를 기반으로 통합됩니다 .
- 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 에서 확인하십시오 .

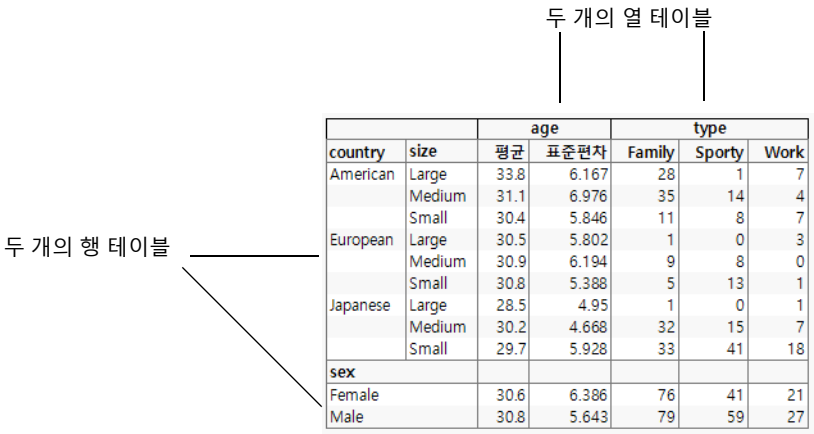
열 및 행 테이블

테이블 생성 플랫폼에서 테이블은 열 머리글과 행 라벨로 정의됩니다. 이러한 하위 테이블을 **행 테이블** 및 **열 테이블**이라고 합니다(그림 9.11).

테이블 생성 플랫폼의 행 및 열 테이블 예

1. **도움말 > 샘플 데이터 폴더**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



		age		type		
country	size	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

테이블 생성 플랫폼에는 테이블에 추가하는 항목을 편집할 수 있는 몇 가지 방법이 있습니다. 항목을 삭제하거나, 열 라벨을 제거하거나, 통계 키워드 또는 라벨을 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 " 의 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 0 부터 시작하거나 0 을 중심으로 표시될 수 있습니다 . 데이터가 모두 음수가 아니거나 모두 양수가 아닐 경우 차트 기준선은 0 입니다 . 그렇지 않으면 차트가 0 을 중심으로 표시됩니다 .

음영 표시 행이 여러 개인 경우 테이블에 회색 음영 상자를 표시합니다 .

툴팁 표시 테이블 영역을 마우스로 가리킬 때 나타나는 팁을 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 " 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 에서 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 .

플랫폼 환경 설정 현재 플랫폼 환경 설정을 보거나 , 현재 JMP 보고서의 설정과 일치하도록 플랫폼 환경 설정을 업데이트할 수 있는 옵션이 포함되어 있습니다 .

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

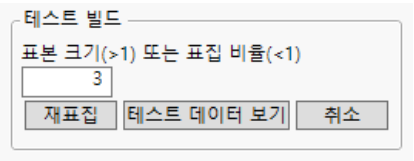
"열 선택"의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오.

테스트 빌드 패널 표시

테이블 생성 플랫폼에서 매우 큰 데이터 테이블을 요약하는 경우 데이터 테이블의 작은 부분집합을 사용하여 여러 가지 테이블 레이아웃을 시도해 본 후 가장 적합한 것을 찾을 수 있습니다. 이 경우 JMP 에서는 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 9.12 에 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 JMP 에서는 전체 데이터 테이블을 사용하여 지정된 대로 테이블을 다시 생성합니다.

마우스 오른쪽 버튼으로 열 클릭 시 표시되는 메뉴

" 테이블 생성 " 보고서 테이블에서 열을 마우스 오른쪽 버튼으로 클릭하면 다음 옵션이 표시됩니다.

삭제 선택한 열을 삭제합니다.

열 묶음 (통계량 열을 마우스 오른쪽 버튼으로 클릭한 경우에만 사용 가능) 더 간편한 보기를 표시하도록 테이블의 통계 열을 통합하는 옵션을 포함합니다.

묶음 선택한 통계 열을 단일 열로 통합합니다.

분해 선택한 통계량을 별도의 열로 분리합니다.

템플릿 통합된 통계량의 표시 형식을 정의할 수 있습니다. " 묶음 " 을 선택할 때 맨 왼쪽에 선택된 열이 ^FIRST 로 지정되고 나머지 선택된 열이 ^OTHERS 로 지정됩니다. 구분자를 선택하여 괄호 안에 표시되는 통계량의 목록을 구분할 수도 있습니다.

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 " 에서 확인하십시오 .

그룹화 열 내포 그룹화 열을 세로 또는 가로로 내포합니다 .

별도의 테이블 (결합된 테이블의 경우에만 표시됨) 각 열에 대해 별도의 테이블을 생성합니다 .

열 라벨 제거 열 테이블에서 열 이름을 제거합니다 .

열 라벨 복원 숨겨진 열 이름을 열 테이블에 복원합니다 .

개수별 정렬 그룹화 열의 수준을 각 수준 개수에 따라 높은 것부터 낮은 것 순서로 정렬합니다 . 이 옵션은 특정 그룹화 열에 대한 " 그룹화 열 개수별 정렬 " 옵션의 설정을 재정의하는 데에도 사용할 수 있습니다 .

그룹화 열 쌓기 (그룹화 열이 여러 개 있는 경우에만 사용 가능) 테이블의 가로 공간을 절약하기 위해 내포된 그룹화 열의 값이 통합 형식으로 표시되도록 지정합니다 . 내포된 그룹화 열의 값은 그룹화 열의 계층을 표시하기 위해 들여쓰기가 적용됩니다 . 자세한 내용은 " 그룹화 열 쌓기의 예 " 에서 확인하십시오 .

테이블 생성 플랫폼의 추가 예

이 섹션에는 테이블 생성 플랫폼을 사용한 예가 포함되어 있습니다 .

- " 예 : 여러 테이블 생성 및 내용 재배열 "
- " 예 : 여러 열을 단일 테이블에 결합 "
- " 예 : 페이지 열 사용 "
- " 그룹화 열 쌓기의 예 "

예 : 여러 테이블 생성 및 내용 재배열

이 예에는 테이블 생성 플랫폼을 사용하여 테이블을 생성하고 수정하는 다음 단계가 포함되어 있습니다.

1. " 개수 테이블 생성 "
2. " 통계량을 보여 주는 테이블 생성 "
3. " 테이블 내용 재배열 "

개수 테이블 생성

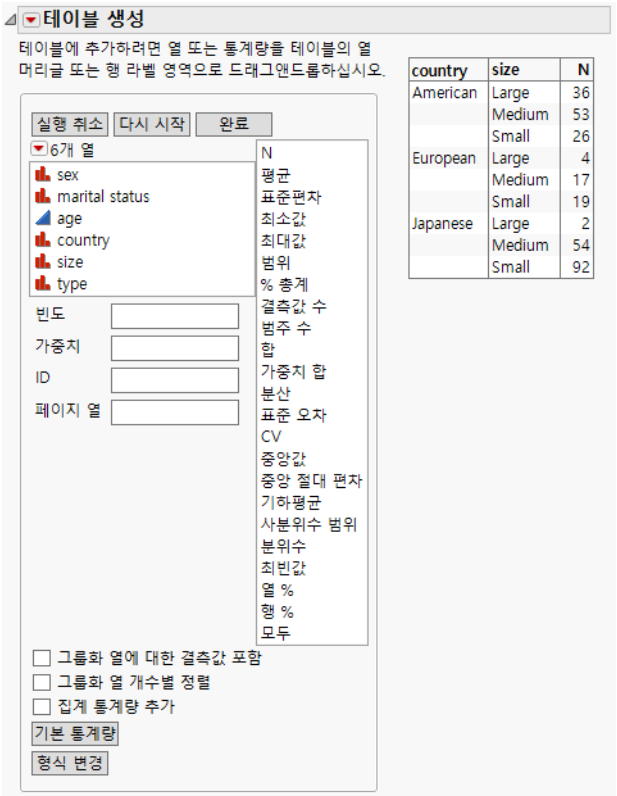
설문 조사 대상자 중 일본, 유럽 및 미국 자동차의 소유자 수가 포함된 테이블을 생성하려고 한다고 가정해 보겠습니다. 또한 이 인원 수를 자동차 크기별로 세분하려고 합니다.

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 폴더**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



통계량을 보여 주는 테이블 생성

각 크기의 자동차를 소유한 사람의 나이 평균 및 표준편차를 확인하려고 한다고 가정해 보겠습니다.

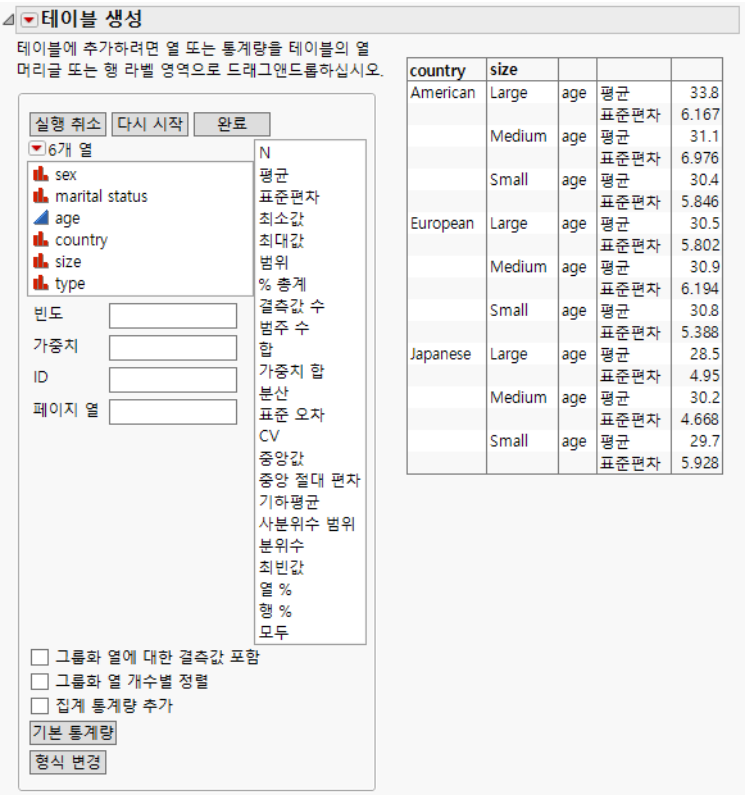
그림 9.15 나이의 평균 및 표준편차를 보여 주는 테이블

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. **평균**을 클릭하고 " 합 " 위로 드래그합니다 .
3. **표준편차**를 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



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

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 재배열합니다 .

1. 그림 9.16 에서 시작합니다 . size 머리글을 클릭하고 테이블 머리글의 오른쪽으로 드래그합니다 .

그림 9.18 size 이동

country	size			Tabula
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

→

			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

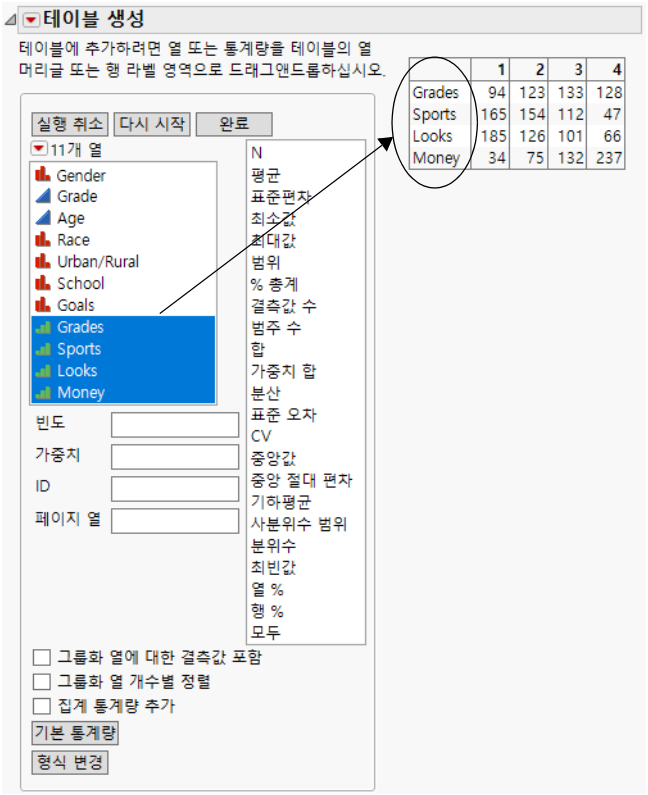
이 예에는 아동의 인기도에 대한 자가 보고 요인(성적, 스포츠, 외모, 경제 여건)의 중요도를 나타내는 학생 데이터가 있습니다. 테이블 생성 플랫폼을 사용하여 이러한 모든 요인을 추가 통계량 및 요인과 함께 결합된 단일 테이블에서 확인하려고 합니다.

그림 9.19 인구 통계학 데이터 추가

		% 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%

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Children's Popularity.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .
3. Grades, Sports, Looks 및 Money 를 선택하고 " 행에 대한 놓기 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. **% 총계**를 번호가 매겨진 머리글 위로 드래그합니다 .
6. **모두**를 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

이 예에는 남학생과 여학생의 키 측정값이 포함된 데이터가 있습니다. 테이블 생성 플랫폼을 사용하여 학생의 나이별 키 평균을 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

age	height 평균
12	58.6
13	59.0
14	62.6
15	63.0
16	62.5
17	62.0

여학생

sex = M

age	height 평균
12	57.3
13	61.3
14	65.3
15	65.2
16	68.0
17	69.0

남학생

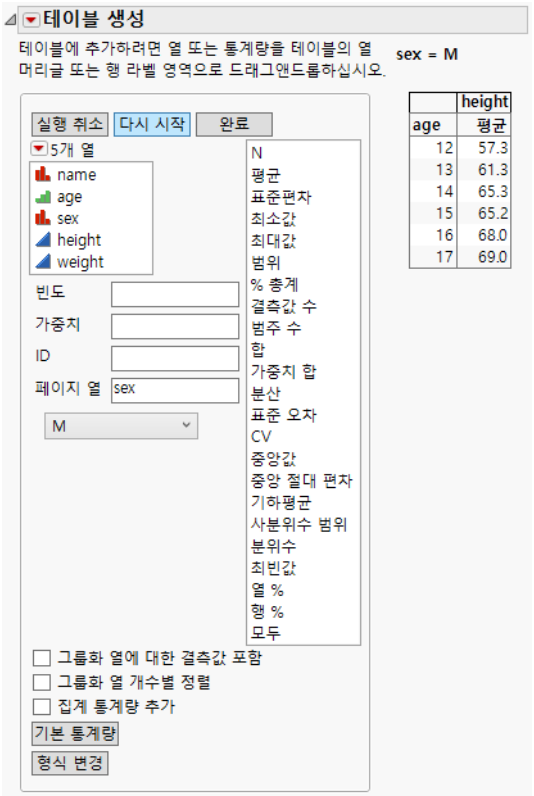
1. **도움말 > 샘플 데이터 폴더**를 선택하고 Big Class.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .

   height 변수를 검토할 것이므로 테이블의 맨 위에 이 변수를 표시하려고 합니다 .
3. height 를 클릭하고 " 열에 대한 놓기 영역 " 으로 드래그합니다 .

   나이별로 통계량을 살펴볼 것이므로 age 변수를 나란히 표시하려고 합니다 .
4. age 를 클릭하고 숫자 "2502" 옆의 빈 셀로 드래그합니다 .
5. " 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



그룹화 열 쌓기의 예

이 예에서는 그룹화 변수 집합을 기반으로 설문 조사 응답자의 고용 정보를 조사하려고 합니다 . 일부 변수 이름과 값이 길기 때문에 보고서 테이블의 가로 공간을 절약하기 위해 그룹화 열을 쌓으려고 합니다 .

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Consumer Preferences.jmp 를 엽니다 .
2. **분석 > 테이블 생성**을 선택합니다 .
3. Years at Current Employer 와 Salary 를 선택하고 " 열에 대한 놓기 영역 " 으로 드래그합니다 .
4. "N" 과 " 평균 " 을 선택하고 "Years at Current Employer" 아래에 " 합 " 이라고 표시된 셀로 드래그합니다 .

5. "Single Status", "School Age Children" 및 "Job Satisfaction" 을 선택하고 테이블 왼쪽의 빈 셀로 드래그합니다 .
6. "Single Status" 머리글을 마우스 오른쪽 버튼으로 클릭하고 **그룹화 열 쌓기**를 선택합니다 .
7. **집계 통계량 추가** 체크박스를 선택합니다 .

그림 9.25 쌓인 그룹화 열

	Years at Current Employer		Salary	
Single Status/School Age Children/Job Satisfaction	평균	N	평균	N
Single	8	177	$54,983	177
Yes	7.918367	49	$60,594	49
Not at all satisfied	5.666667	3	$45,667	3
Somewhat satisfied	7.769231	26	$49,465	26
Extremely satisfied	8.45	20	$77,300	20
No	8.03125	128	$52,835	128
Not at all satisfied	7	13	$46,346	13
Somewhat satisfied	8.352113	71	$48,241	71
Extremely satisfied	7.818182	44	$62,166	44
Not Single	9.608856	271	$59,944	271
Yes	9.676471	136	$63,935	136
Not at all satisfied	7.333333	9	$59,167	9
Somewhat satisfied	8.8	70	$59,725	70
Extremely satisfied	11.12281	57	$69,859	57
No	9.540741	135	$55,923	135
Not at all satisfied	11.57143	7	$48,714	7
Somewhat satisfied	8.706667	75	$51,134	75
Extremely satisfied	10.45283	53	$63,651	53
모두	8.973214	448	$57,984	448

이 테이블에는 평균 급여, 평균 근속 기간 및 그룹화 변수의 각 조합에 대한 응답자 수를 분석한 결과가 표시됩니다. 예를 들어 취학 연령 자녀가 있는 독신 중 직업 만족도가 매우 큰 응답자는 20 명입니다. 이 20 명의 응답자는 현재 직장에서 평균 8.45 년 동안 근무했고 평균 급여는 $77,300 입니다.

Single Status 및 School Age Children 변수의 집계 값도 테이블에 나타납니다. 예를 들어 직업 만족도를 고려하지 않고 취학 연령 자녀가 없는 독신 응답자는 128 명입니다. 이 128 명의 응답자는 현재 직장에서 평균 8.03 년 동안 근무했고 평균 급여는 $52,835 입니다.

10 장

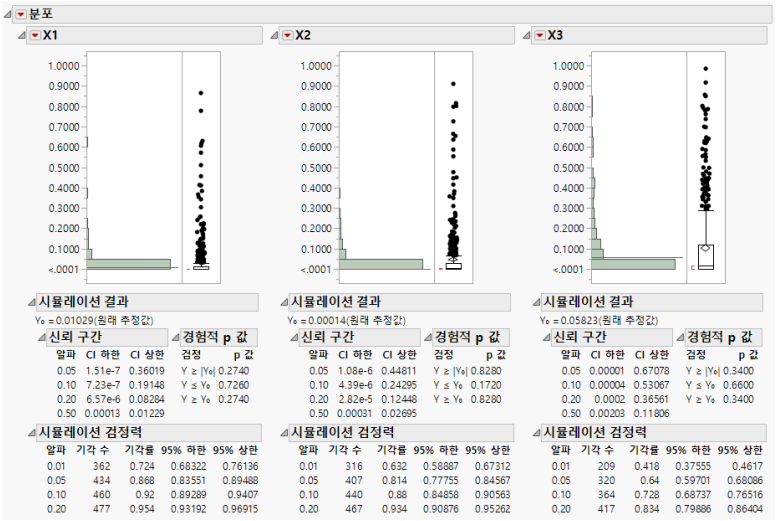
시뮬레이션
모수 재표집을 통해 난제 해결 방안 모색

시뮬레이션 기능으로는 강력한 모수 및 비모수 시뮬레이션을 수행할 수 있습니다. 시뮬레이션 기능을 사용하여 수행할 수 있는 작업은 다음과 같습니다.

- 붓스트랩을 확장하여 모수 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 기능은 붓스트랩을 지원하는 모든 보고서를 포함하여 다양한 보고서에서 사용할 수 있습니다. "시뮬레이션"에 액세스하려면 보고서에서 마우스 오른쪽 버튼을 클릭합니다.

그림 10.1 시뮬레이션을 통한 검정력 분석



목차

시뮬레이션 기능 개요 297
시뮬레이션 기능의 예 297
시뮬레이션 기능 시작 302
    시뮬레이션 창 302
시뮬레이션 결과 테이블 303
    시뮬레이션 결과 보고서 304
    시뮬레이션 검정력 보고서 304
시뮬레이션 기능의 추가 예 304
    순열 검정의 예 305
    일반화 회귀에서 요인 유지의 예 307
    전향적 검정력 분석의 예 312

시뮬레이션 기능 개요

시뮬레이션 기능은 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 N번 반복됩니다. 여기서 N은 사용자가 지정한 총 표본 수입니다.

시뮬레이션 분석에서는 분석 요약을 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 기능은 시뮬레이션이 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 및 텍스트 탐색기를 제외한 모든 통계 플랫폼에서 사용할 수 있습니다.

시뮬레이션 기능의 예

이 예에서는 변경하기 힘든 변수의 분산 성분에 대한 준모수 신뢰 구간을 구하는 것이 목표입니다. 이 데이터에 대해 온도, 시간 및 촉매의 양이 반응에 미치는 효과를 알아보려고 합니다. 온도는 매우 변경하기 힘든 변수(주구 요인)이고, 시간은 변경하기 힘든 변수(하위구 요인)이며, 촉매 양은 변경하기 쉬운 변수입니다. 주구 및 하위구 요인에 대한 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

이전 연구에서는 주구 표준편차가 오차 표준편차의 약 2배이고, 하위구 오차는 오차 표준편차의 약 1.5배인 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.

설계 구성

이 섹션에서는 분할 - 분할구 실험을 위한 사용자 설계를 구성합니다.

팁 : 이 섹션의 단계를 건너뛰려면 **도움말 > 샘플 데이터 폴더**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 " 절차로 이동하십시오 .

1. DOE > 사용자 설계를 선택합니다 .
2. " 요인 " 개요에서 N 개 요인 추가 옆에 3 을 입력합니다 .
3. 요인 추가 > 연속형을 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 쉬움을 클릭하고 매우 어려움을 선택합니다 .

   이렇게 하면 Temperature 가 주구 요인으로 정의됩니다 .
6. Time 에서 쉬움을 클릭하고 어려움을 선택합니다 .

   이렇게 하면 Time 이 하위구 요인으로 정의됩니다 .
7. 계속을 클릭합니다 .
8. " 모형 " 개요에서 교호작용 > 2 차를 선택합니다 .

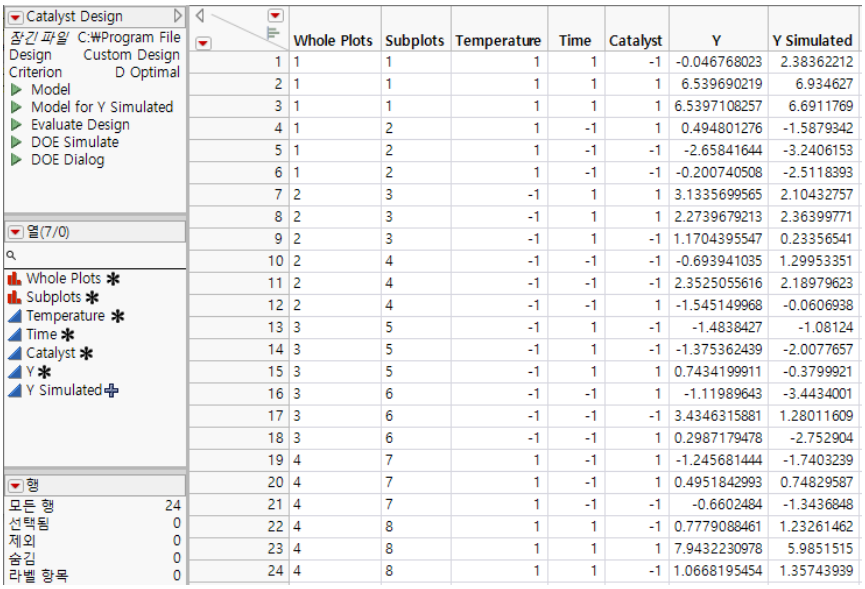
   이렇게 하면 모형에 모든 2 원 교호작용이 추가됩니다 .
9. " 사용자 설계 " 의 빨간색 삼각형을 클릭하고 반응 시뮬레이션을 선택합니다 .

   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 설정을 선택합니다 . "12345" 를 입력하고 확인을 클릭합니다 .
11. (선택 사항) " 사용자 설계 " 옆의 빨간색 삼각형을 클릭하고 시작 수를 선택합니다 . "1000" 을 입력하고 확인을 클릭합니다 .
12. 설계 생성을 클릭합니다 .
13. 테이블 생성을 클릭합니다 .

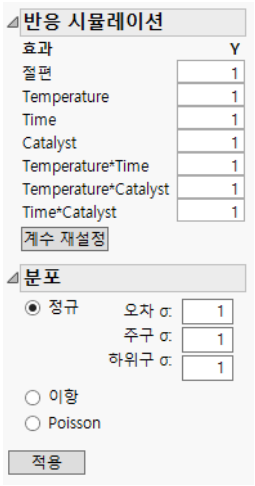
    참고 : Y 및 Y 시뮬레이션 열의 항목은 그림 10.2 에 표시된 것과 다를 수 있습니다 .

그림 10.2 설계 테이블



	Whole Plots	Subplots	Temperature	Time	Catalyst	Y	Y Simulated
1	1	1	1	1	-1	-0.046768023	2.38362212
2	1	1	1	1	1	6.539690219	6.934627
3	1	1	1	1	1	6.5397108257	6.6911769
4	1	2	1	-1	1	0.494801276	-1.5879342
5	1	2	1	-1	-1	-2.65841644	-3.2406153
6	1	2	1	-1	-1	-0.200740508	-2.5118393
7	2	3	-1	1	1	3.1335699565	2.10432757
8	2	3	-1	1	1	2.2739679213	2.36399771
9	2	3	-1	1	-1	1.1704395547	0.23356541
10	2	4	-1	-1	-1	-0.693941035	1.29953351
11	2	4	-1	-1	-1	2.3525055616	2.18979623
12	2	4	-1	-1	1	-1.545149968	-0.0606938
13	3	5	-1	1	-1	-1.4838427	-1.08124
14	3	5	-1	1	-1	-1.375362439	-2.0077657
15	3	5	-1	1	1	0.7434199911	-0.3799921
16	3	6	-1	-1	1	-1.11989643	-3.4434001
17	3	6	-1	-1	-1	3.4346315881	1.28011609
18	3	6	-1	-1	1	0.2987179478	-2.752904
19	4	7	1	-1	1	-1.245681444	-1.7403239
20	4	7	1	-1	1	0.4951842993	0.74829587
21	4	7	1	-1	-1	-0.6602484	-1.3436848
22	4	8	1	1	-1	0.7779088461	1.23261462
23	4	8	1	1	1	7.9432230978	5.9851515
24	4	8	1	1	-1	1.0668195454	1.35743939

그림 10.3 반응 시뮬레이션 창



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.

모형 적합

이 섹션에서는 REML 방법을 사용하여 모형을 적합시킵니다. 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 2 를 입력합니다.

   기본적으로 " 정규 " 분포가 선택되어 있습니다. 따라서 계산식에 정규 오차가 추가됩니다.
2. **하위구 σ** 옆에 1.5 를 입력합니다.
3. **적용**을 클릭합니다.

   데이터 테이블에서 **Y 시뮬레이션**의 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 시뮬레이션**을 클릭하고 Y 버튼을 클릭합니다.

   이렇게 하면 Y 가 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.

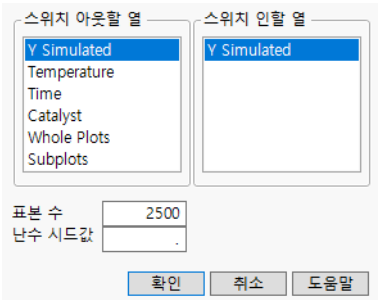
음수 추정값을 포함한 합계 = 2.2738839

신뢰 구간 생성

이 섹션에서는 분산 성분 추정값을 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



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 시뮬레이...
난수 시드값 456
결합 데이터 테이블 생성
분포

열(6/0)
Y *
SimID•
Subplots
Whole Plots
잔차
합계

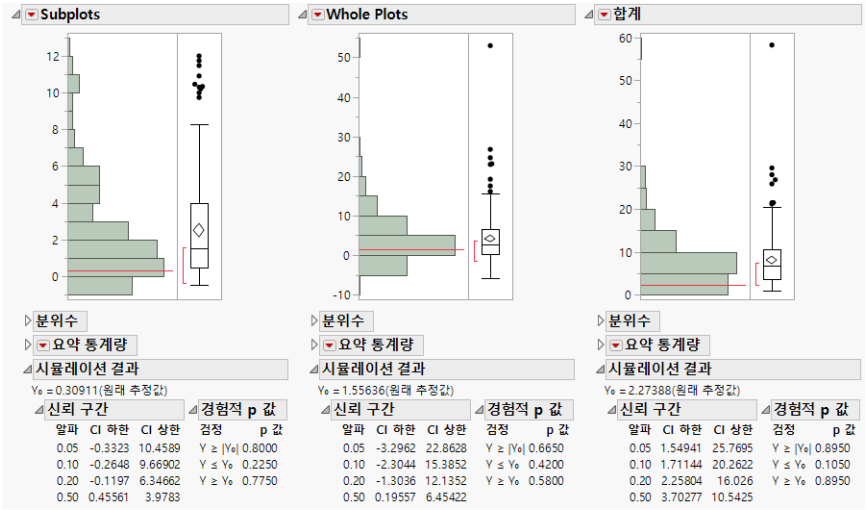
행
모든 행 201
선택됨 0
제외 1
숨김 0
라벨 항목 0

	Y	SimID•	Subplots	Whole Plots	잔차	합계
1	Y Simulated	0	0.3091096508	1.5563639995	0.4084102799	2.2738839302
2	Y Simulated	1	9.9978320675	0.1160090855	1.7118437836	11.825684937
3	Y Simulated	2	4.5617762065	3.3652848741	0.7259348275	8.6529959081
4	Y Simulated	3	3.0383907272	22.95631251	0.8416572537	26.836360491
5	Y Simulated	4	-0.176636748	12.156574946	0.9872034755	13.143778421
6	Y Simulated	5	-0.153702674	-0.060658009	1.3514412755	1.3514412755
7	Y Simulated	6	1.1063669307	11.804365807	0.9837455205	13.894478258
8	Y Simulated	7	0.1630909822	2.6824797817	0.8559296173	3.7015003812
9	Y Simulated	8	0.2154635485	26.765100099	1.0422128078	28.022776455
10	Y Simulated	9	4.0615625437	12.75877132	1.0302645404	17.850598404
11	Y Simulated	10	1.8169128804	-0.433818317	1.2569403028	3.0738531832
12	Y Simulated	11	6.9696566519	2.4634884406	1.6719518802	11.105096973
13	Y Simulated	12	1.9758939136	4.2166345958	1.0326091926	7.2251377019
14	Y Simulated	13	1.1584152266	2.8027318323	0.7050109246	4.6661579836
15	Y Simulated	14	5.2888052789	0.3194564512	0.9281404099	6.53640214
16	Y Simulated	15	1.4827905134	-0.270350632	1.6586718874	3.1414624009
17	Y Simulated	16	0.7636963822	-0.192580268	0.9471174937	1.7108138759
18	Y Simulated	17	1.1919618966	0.756815441	0.9782751744	2.927052512
19	Y Simulated	18	1.0391874862	4.8135329401	1.5524955876	7.4052160139

" 최소 제곱 적합 시뮬레이션 결과 (분산 성분)" 데이터 테이블의 첫 번째 행은 **분산 성분**의 초기값을 포함하므로 제외됩니다. 나머지 행에는 시뮬레이션된 값이 포함되어 있습니다.

5. Distribution 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 .

시뮬레이션 기능 시작

시뮬레이션 기능을 시작하려면 JMP 보고서 창에서 계산된 값이 포함된 열을 마우스 오른쪽 버튼으로 클릭하고 "시뮬레이션"을 선택합니다. 시뮬레이션 기능을 사용하려면 데이터를 시뮬레이션하는 랜덤 성분이 있는 계산식이 데이터 테이블에 포함되어 있어야 합니다. 시뮬레이션 옵션은 붓스트랩을 지원하는 모든 보고서를 포함하여 다양한 보고서에서 사용할 수 있습니다.

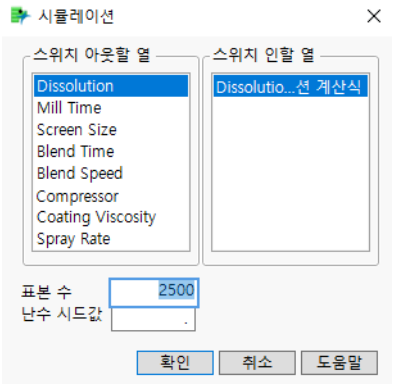
참고 : 기준 변수를 사용하는 보고서에서는 시뮬레이션 기능을 사용할 수 없습니다 .

시뮬레이션 창

"시뮬레이션" 창에서 열과 표본 수를 지정한 후 "확인"을 클릭하여 시뮬레이션을 실행합니다. 진행률 표시줄과 "조기 중지" 옵션이 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.

그림 10.8 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 2500 입니다 .

난수 시드값 시뮬레이션된 결과를 제어하는 값입니다 . 난수 시드값은 결과를 재현 가능하게 합니다 .

시뮬레이션 결과 테이블

시뮬레이션을 실행하면 결과가 데이터 테이블에 나타납니다 . 다음 사항을 알 수 있습니다 .

- 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 시뮬

레이션 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 " 에서 확인하십시오 .
- 시뮬레이션 결과 데이터 테이블의 값이 p 값 형식인 경우에는 " 시뮬레이션 검정력 " 보고서도 제공됩니다 . 자세한 내용은 " 시뮬레이션 검정력 보고서 " 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 " 에서 확인하십시오 .

시뮬레이션 결과 보고서

" 시뮬레이션 결과 " 보고서에는 다음 정보가 포함되어 있습니다 .

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

" 시뮬레이션 검정력 " 보고서에는 다음 정보가 포함되어 있습니다 .

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 스코어 방법을 사용하여 계산됩니다 . 자세한 내용은 Wilson 연구 자료 (1927) 에서 확인하십시오 .

팁 : 신뢰 구간을 좁히려면 표본 수를 늘리십시오 .

시뮬레이션 기능의 추가 예

이 섹션에는 시뮬레이션 기능을 사용한 추가 예가 포함되어 있습니다 .

- " 순열 검정의 예 "

- " 일반화 회귀에서 요인 유지의 예 "
- " 전향적 검정력 분석의 예 "

비정규 변수의 Ppk 및 부적합률에 대한 신뢰 구간 시뮬레이션 방법을 보여 주는 예는 **품질 및 공정 방법**의 에서 확인하십시오.

순열 검정의 예

이 예에서는 시뮬레이션 기능을 사용하여 난수화 또는 순열 검정을 수행합니다. 세 가지 약물이 통증에 미치는 효과를 연구하려고 하며 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.

시뮬레이션 계산식 정의

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Analgesics.jmp 를 엽니다 .
2. 열 > 새 열을 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 삽입 키 (^) 를 클릭합니다 .
8. 함수 목록에서 **난수** > Col Shuffle 을 선택합니다 .

그림 10.9 완료된 계산식

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 적합을 선택합니다 .
2. pain 을 선택하고 Y, 반응을 클릭합니다 .

3. drug 를 선택하고 X, 요인을 클릭합니다 .
4. 확인을 클릭합니다 .
5. " 일원 분석 " 의 빨간색 삼각형을 클릭하고 평균 /ANOVA 를 선택합니다 .

그림 10.10 분산 분석 보고서

분산 분석

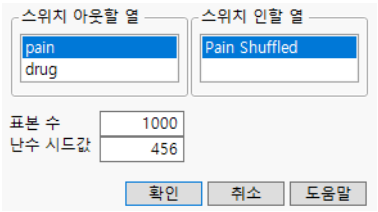
소스	DF	제곱합	평균 제곱	F 비	Prob > F
drug	2	99.89459	49.9473	6.2780	0.0053*
오차	30	238.67877	7.9560		
수정 합계	32	338.57335			

F 비가 6.2780 입니다 .

6. " 분산 분석 " 개요에서 "F 비 " 열을 마우스 오른쪽 버튼으로 클릭하고 시뮬레이션을 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 .

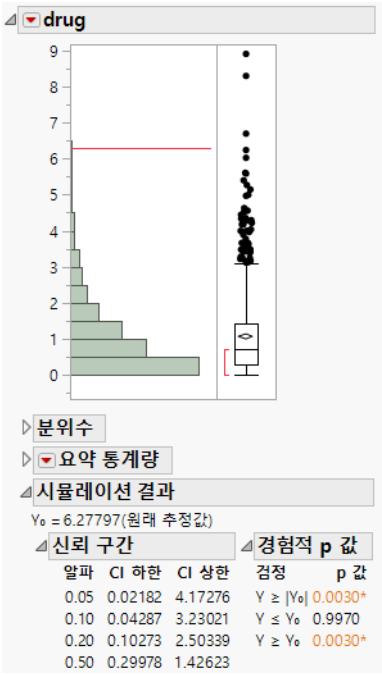
그림 10.11 완료된 시뮬레이션 창



11. 확인을 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 10.12 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 회귀에서 요인 유지의 예

이 예에서는 일반화 회귀 모형을 구성하고 활성 모수를 사용하여 축소 모형을 적합시킵니다. 이 축소 모형을 기반으로 시뮬레이션 기능을 사용하여 요인 중 하나가 모형에 포함될 가능성을 탐색합니다.

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.

모형 적합

이 섹션에서는 일반화 회귀를 사용하여 모형을 적합시킵니다 .

팁 : 이 섹션의 단계를 거치지 않으려면 Tablet Production.jmp 데이터 테이블의 Generalized Regression 스크립트 옆에 있는 녹색 삼각형을 클릭하여 모형을 구하십시오 .

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Tablet Production.jmp 를 엽니다 .

2. 분석 > 모형 적합을 선택합니다 .
3. Dissolution 을 클릭하고 Y 를 클릭합니다 .
4. Mill Time 부터 Atomizer Pressure 까지 선택하고 추가를 클릭합니다 .
5. " 분석법 " 목록에서 일반화 회귀를 선택합니다 .
6. **실행**을 클릭합니다 .
7. " 모형 시작 " 패널에서 적응형 상자를 선택합니다 .
8. " 모형 시작 " 패널에서 시작을 클릭합니다 .

그림 10.13 적응형 Lasso 회귀 기반의 모형

원래 예측 변수에 대한 모수 추정값

항	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
절편	108.64064	24.332978	19.933984	<.0001*	60.94888	156.3324
Mill Time	0.1302779	0.028185	21.365104	<.0001*	0.0750363	0.1855195
Screen Size[3-5]	4.1877689	0.5414927	59.810874	<.0001*	3.1264626	5.2490751
Screen Size[4-5]	2.3907767	0.5672167	17.765609	<.0001*	1.2790525	3.5025009
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	0.7223431	0.1381847	27.325418	<.0001*	0.4515059	0.9931802
Blend Speed	0.2130883	0.2389206	0.7954486	0.3725	-0.255187	0.681364
Compressor[Compress1-Compress2]	-0.538111	0.3993672	1.8155137	0.1778	-1.320857	0.244634
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	13.414234	0.0002*	0.0852717	0.2815964
Inlet Temp	0	0	0	1.0000	0	0
Exhaust Temp	0	0	0	1.0000	0	0
Spray Rate	-0.204356	0.0420142	23.658261	<.0001*	-0.286703	-0.12201
Atomizer Pressure	0	0	0	1.0000	0	0

정규 분포 모수	추정값	표준 오차	Wald 카이제곱	Prob > ChiSquare	95% 하한	95% 상한
척도	2.0224258	0.1641142	151.86341	<.0001*	1.700768	2.3440837

"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 .

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 0인 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 시작을 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 10.14 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	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 0을 포함하는 것과 매우 가깝습니다. 다음으로는 시뮬레이션 연구를 수행하여 용해도 분포에서 다른 데이터 값이 관측될 때 Blend Speed가 얼마나 자주 모형에 포함되는지 알아봅니다.

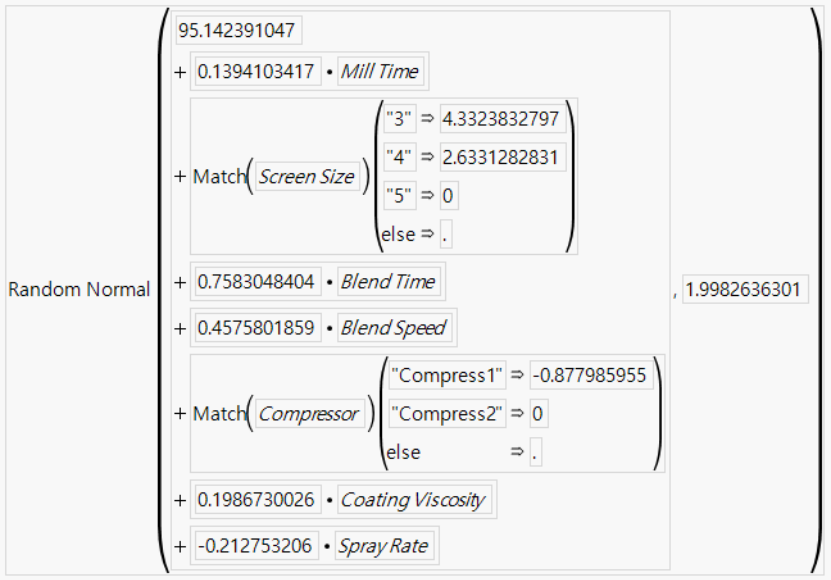
모형의 Blend Speed 포함 여부 탐색

아래 단계에서는 축소 모형에 대한 보고서 (그림 10.14) 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 10.15 시뮬레이션 계산식



이 계산식은 각 행에 대해서 주어진 모형과 Dissolution 의 분포 (표준편차가 약 1.998 인 정규 분포로 추정) 로 구할 수 있는 값을 시뮬레이션합니다 .

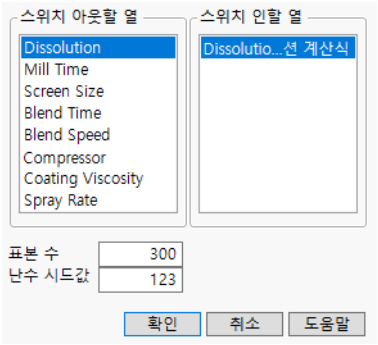
3. **취소**를 클릭합니다 .
4. 축소 모형 보고서 창으로 돌아갑니다 . " 원래 예측 변수에 대한 모수 추정값 " 보고서에서 " 추정값 " 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .

   " 스위치 아웃할 열 " 목록에 Dissolution 이 선택되어 있는지 확인합니다 .
5. **표본 수** 옆에 300 을 입력합니다 .

   이 시뮬레이션에서는 300 개의 각 분석값에서 Dissolution 열을 Dissolution 시뮬레이션 계산식 열을 사용하여 시뮬레이션된 값으로 대체하도록 요청합니다 .
6. (선택 사항) **난수 시드값**을 123 으로 설정합니다 .

   이렇게 하면 이 예의 값이 재현됩니다 .

그림 10.16 완료된 시뮬레이션 창

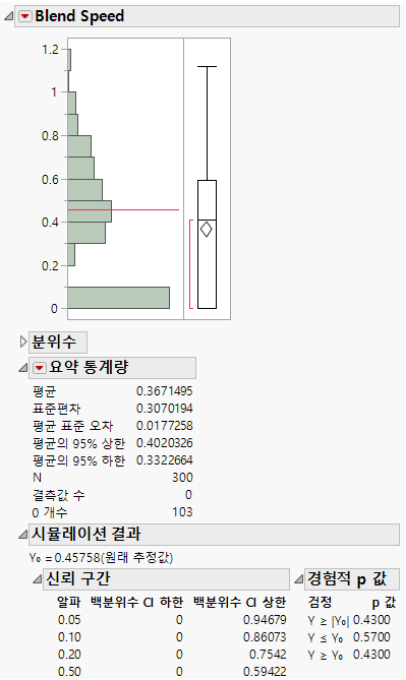


7. 확인을 클릭합니다 .

   테이블의 첫 번째 행은 추정값의 초기 값을 포함하므로 제외됩니다. 나머지 행에는 시뮬레이션된 값이 포함되어 있습니다 .

8. Distribution 스크립트를 실행합니다 .
9. Ctrl 키를 누른 채 Blend Speed 빨간색 삼각형을 클릭하고 **표시 옵션 > 요약 통계량 사용자 정의를 선택합니다** .
10. 0 개수를 선택합니다 .
11. 확인을 클릭합니다 .
12. Blend Speed 에 대한 분포 보고서로 스크롤합니다 .

그림 10.17 시뮬레이션된 Blend Speed 계수 추정값의 히스토그램



이 " 요약 통계량 " 보고서에서는 시뮬레이션의 103/300 = 34.3% 에 대해 Blend Speed 추정값이 0 임을 보여 줍니다 .

전향적 검정력 분석의 예

이 예에서는 시뮬레이션 기능을 사용하여 비선형 모형에 대한 전향적 검정력 분석을 수행합니다. 부품의 검사 합격 또는 불합격에 영향을 미치는 6가지 연속형 요인의 주효과를 알아보려고 합니다. 반응은 이항이며 총 60회 런이 가능합니다.

로짓을 연결 함수로 사용한 일반화 선형 모형을 사용하여 성공 확률을 모델링합니다 . 로짓 연결 함수는 로지스틱 모형을 적합시킵니다 .

$$\pi(\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \ldots + \beta_6 X_6)}}$$

여기서 $\pi(\mathbf{X})$는 주어진 설계 설정 $\mathbf{X} = (X_1, X_2, \ldots, X_6)$에서 부품이 검사에 합격할 확률을 나타냅니다.

선형 예측 변수는 $L(\mathbf{X})$ 로 나타냅니다 .

$$L(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \ldots + \beta_6 X_6$$

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* 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.

설계 구성

이 섹션에서는 실험을 위한 사용자 설계를 구성합니다.

참고: 이 섹션의 단계를 건너뛰려면 **도움말 > 샘플 데이터 폴더**를 선택하고 Design Experiment/Binomial Experiment.jmp 를 여십시오 . DOE Simulate 스크립트 옆의 녹색 삼각형을 클릭한 다음 " 시뮬레이션된 반응 정의 " 절차로 이동하십시오 .

1. DOE > **사용자 설계**를 선택합니다 .

   참고 : 사용자 설계는 비선형 상황에는 최적이 아니지만 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 가이드**의 에서 확인하십시오 .

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 .

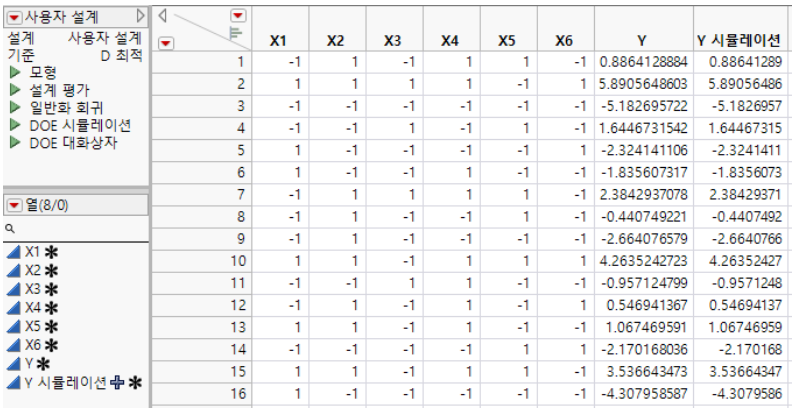
   이렇게 하면 " 테이블 생성 " 을 선택하여 설계 테이블을 구성한 후에 " 반응 시뮬레이션 " 창이 열립니다 .

   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 생성**을 클릭합니다 .
10. **테이블 생성**을 클릭합니다 .

    참고 : Y 및 Y **시뮬레이션** 열의 항목은 그림 10.18 에 표시된 것과 다를 수 있습니다 .

그림 10.18 설계 테이블 (일부분)



사용자 설계

설계 사용자 설계
기준 D 최적

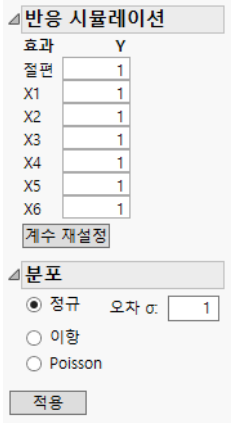
- 모형
- 설계 평가
- 일반화 회귀
- DOE 시뮬레이션
- DOE 대화상자

열(8/0)

X1 ✱
X2 ✱
X3 ✱
X4 ✱
X5 ✱
X6 ✱
Y ✱
Y 시뮬레이션 ✱

	X1	X2	X3	X4	X5	X6	Y	Y 시뮬레이션
1	-1	1	-1	1	1	-1	0.8864128884	0.88641289
2	1	1	1	1	-1	1	5.8905648603	5.89056486
3	-1	-1	-1	-1	-1	-1	-5.182695722	-5.1826957
4	-1	-1	1	-1	1	-1	1.6446731542	1.64467315
5	1	-1	-1	-1	-1	1	-2.324141106	-2.3241411
6	1	-1	-1	1	-1	-1	-1.835607317	-1.8356073
7	-1	1	1	1	1	-1	2.3842937078	2.38429371
8	-1	1	-1	-1	1	-1	-0.440749221	-0.4407492
9	-1	1	-1	-1	-1	-1	-2.664076579	-2.6640766
10	1	1	-1	1	1	1	4.2635242723	4.26352427
11	-1	-1	1	1	-1	-1	-0.957124799	-0.9571248
12	-1	1	-1	1	-1	1	0.546941367	0.54694137
13	1	1	-1	1	-1	-1	1.067469591	1.06746959
14	-1	-1	-1	-1	1	1	-2.170168036	-2.170168
15	1	1	-1	1	1	-1	3.536643473	3.53664347
16	1	-1	-1	-1	-1	-1	-4.307958587	-4.3079586

그림 10.19 반응 시뮬레이션 창



반응 시뮬레이션

효과	Y
절편	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 .

- Y에는 " 반응 시뮬레이션 " 창에서 지정한 대로 시뮬레이션된 값 집합이 포함되어 있습니다.
- **Y 시뮬레이션**에는 " 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.

시뮬레이션된 반응 정의

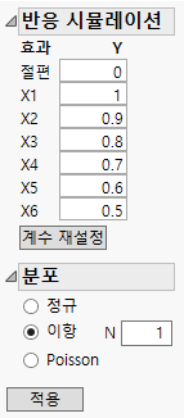
성공 확률이 로지스틱 모형으로 주어지는 이항 반응 데이터를 시뮬레이션할 계획입니다. 반응 시뮬레이션에 대한 자세한 내용은 **실험 설계 가이드**의 에서 확인하십시오.

참고 : 이 섹션의 단계를 건너뛰려면 Simulate Model Responses 스크립트 옆의 녹색 삼각형을 클릭하십시오 . 그런 다음 " 일반화 선형 모형 적합 " 절차로 이동하십시오 .

1. " 반응 시뮬레이션 " 창 (그림 10.19) 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 이항을 선택합니다 .

   N 값은 1 로 설정된 상태로 둡니다 . 이 값은 시행당 1 단위만 있음을 나타냅니다 .

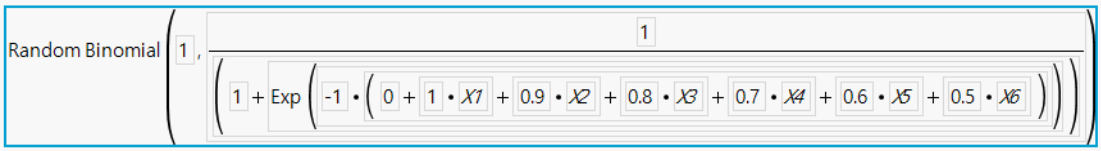
그림 10.20 완료된 반응 시뮬레이션 창



3. 적용을 클릭합니다 .

   설계 데이터 테이블에서 Y 시뮬레이션 열이 이항 값을 생성하는 계산식 열로 대체됩니다 . Y 시행 횟수라는 열은 각 런의 시행 횟수를 나타냅니다 .
4. (선택 사항) " 열 " 패널에서 Y 시뮬레이션 오른쪽의 더하기 기호를 클릭합니다 .

그림 10.21 Y 시뮬레이션에 대한 랜덤 이항 계산식



5. 취소를 클릭합니다 .

일반화 선형 모형 적합

이 섹션에서는 일반화 선형 모형 분석법을 사용하여 로지스틱 모형을 적합시킵니다.

1. 데이터 테이블에서 **모형** 스크립트 옆의 녹색 삼각형을 클릭합니다 .
2. Y 버튼 옆의 Y 변수를 클릭하고 **제거**를 클릭합니다 .
3. **Y 시뮬레이션**을 클릭하고 Y 버튼을 클릭합니다 .

   이렇게 하면 Y 가 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 .

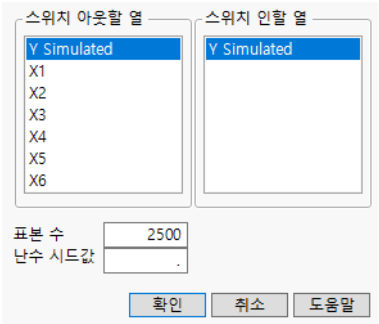
   적합된 모형은 시뮬레이션된 이항 반응의 단일 집합을 기반으로 합니다 .

검정력 분석

이 섹션에서는 가능도비 검정 p 값을 시뮬레이션하여 예제 소개 부분에 제공된 계수 값을 사용한 선형 예측 변수에 의해 결정되는 확률 값 범위에서 차이를 감지할 검정력을 탐색합니다.

1. " 효과 검정 " 개요에서 Prob>ChiSq 열을 마우스 오른쪽 버튼으로 클릭하고 **시뮬레이션**을 선택합니다 .

그림 10.22 시뮬레이션 창



" 스위치 아웃할 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 시뮬레이션**의 값이 **Y 시뮬레이션** 열의 계산식을 사용하여 시뮬레이션된 값을 포함하는 새 열로 대체됩니다 .

보고서에서 선택한 Prob>ChiSq 열은 연관된 주효과가 0 인지 여부에 대한 가능도비 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 10.23 에 표시된 것과 다를 수 있습니다 .

그림 10.23 시뮬레이션 결과 표 (일부분)

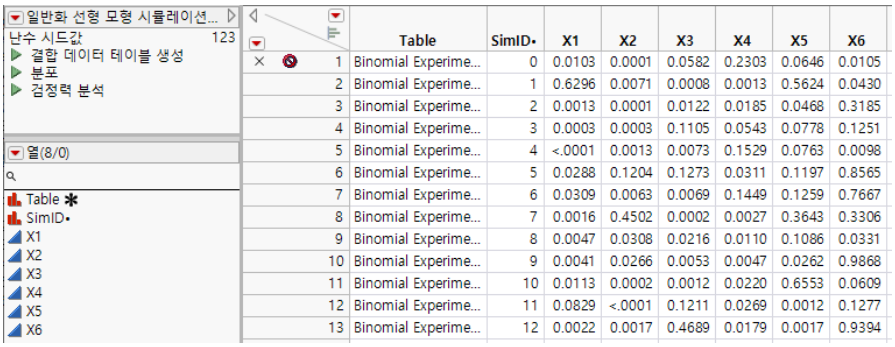


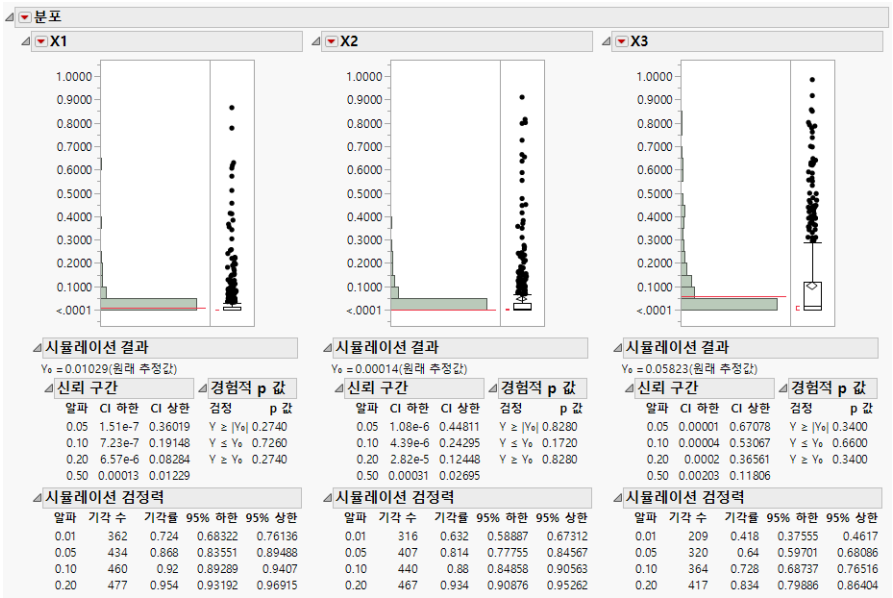
	Table	SimID•	X1	X2	X3	X4	X5	X6
1	Binomial Experime...	0	0.0103	0.0001	0.0582	0.2303	0.0646	0.0105
2	Binomial Experime...	1	0.6296	0.0071	0.0008	0.0013	0.5624	0.0430
3	Binomial Experime...	2	0.0013	0.0001	0.0122	0.0185	0.0468	0.3185
4	Binomial Experime...	3	0.0003	0.0003	0.1105	0.0543	0.0778	0.1251
5	Binomial Experime...	4	<.0001	0.0013	0.0073	0.1529	0.0763	0.0098
6	Binomial Experime...	5	0.0288	0.1204	0.1273	0.0311	0.1197	0.8565
7	Binomial Experime...	6	0.0309	0.0063	0.0069	0.1449	0.1259	0.7667
8	Binomial Experime...	7	0.0016	0.4502	0.0002	0.0027	0.3643	0.3306
9	Binomial Experime...	8	0.0047	0.0308	0.0216	0.0110	0.1086	0.0331
10	Binomial Experime...	9	0.0041	0.0266	0.0053	0.0047	0.0262	0.9868
11	Binomial Experime...	10	0.0113	0.0002	0.0012	0.0220	0.6553	0.0609
12	Binomial Experime...	11	0.0829	<.0001	0.1211	0.0269	0.0012	0.1277
13	Binomial Experime...	12	0.0022	0.0017	0.4689	0.0179	0.0017	0.9394

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 10.24 에 표시된 것과 다를 수 있습니다 .

그림 10.24 처음 두 개의 효과에 대한 분포 그림

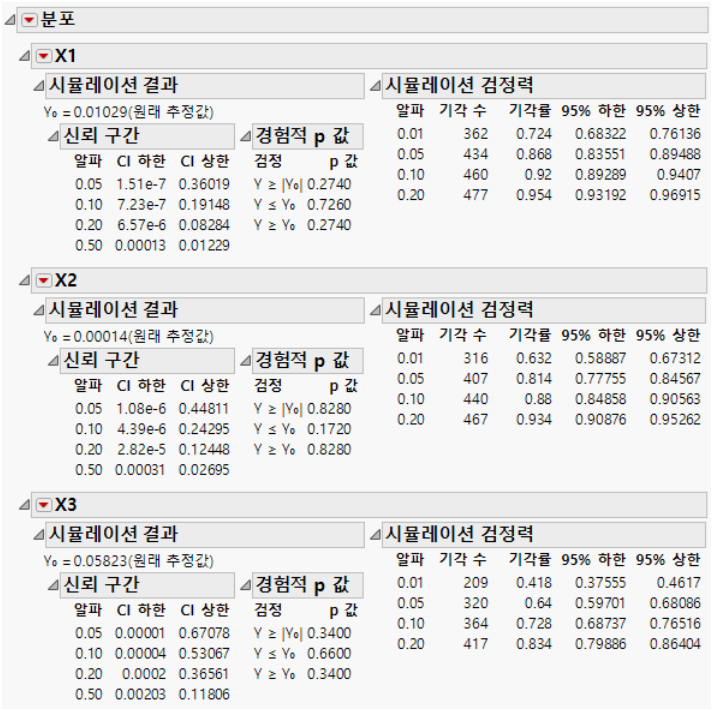


히스토그램에는 각 주효과에 대해 시뮬레이션된 500 개의 Prob>ChiSq 값이 표시됩니다. "시뮬레이션 검정력 " 개요에는 500 회의 시뮬레이션에서 시뮬레이션된 기각률이 표시됩니다. 좀 더 보기 쉽게 보고서를 세로로 쌓고 그림을 선택 취소할 수 있습니다.

5. " 분포 " 의 빨간색 삼각형을 클릭하고 **쌓기**를 선택합니다.
6. Ctrl 키를 누른 채 X1 의 빨간색 삼각형을 클릭하고 **이상치 상자 그림**을 선택 취소합니다.
7. Ctrl 키를 누른 채 X1 의 빨간색 삼각형을 클릭하고 **히스토그램 옵션**을 선택한 다음 **히스토그램**을 선택 취소합니다.

참고 : 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 그림 10.25 에 표시된 것과 다를 수 있습니다.

그림 10.25 처음 세 개의 효과에 대한 검정력 결과



분포

X1

시뮬레이션 결과

Y_0 = 0.01029(원래 추정값)

신뢰 구간

알파	CI 하한	CI 상한
0.05	1.51e-7	0.36019
0.10	7.23e-7	0.19148
0.20	6.57e-6	0.08284
0.50	0.00013	0.01229

경험적 p 값

검정	p 값
Y ≥ |Y_0|	0.2740
Y ≤ Y_0	0.7260
Y ≥ Y_0	0.2740

시뮬레이션 검정력

알파	기각 수	기각률	95% 하한	95% 상한
0.01	362	0.724	0.68322	0.76136
0.05	434	0.868	0.83551	0.89488
0.10	460	0.92	0.89289	0.9407
0.20	477	0.954	0.93192	0.96915

X2

시뮬레이션 결과

Y_0 = 0.00014(원래 추정값)

신뢰 구간

알파	CI 하한	CI 상한
0.05	1.08e-6	0.44811
0.10	4.39e-6	0.24295
0.20	2.82e-5	0.12448
0.50	0.00031	0.02695

경험적 p 값

검정	p 값
Y ≥ |Y_0|	0.8280
Y ≤ Y_0	0.1720
Y ≥ Y_0	0.8280

시뮬레이션 검정력

알파	기각 수	기각률	95% 하한	95% 상한
0.01	316	0.632	0.58887	0.67312
0.05	407	0.814	0.77755	0.84567
0.10	440	0.88	0.84858	0.90563
0.20	467	0.934	0.90876	0.95262

X3

시뮬레이션 결과

Y_0 = 0.05823(원래 추정값)

신뢰 구간

알파	CI 하한	CI 상한
0.05	0.00001	0.67078
0.10	0.00004	0.53067
0.20	0.0002	0.36561
0.50	0.00203	0.11806

경험적 p 값

검정	p 값
Y ≥ |Y_0|	0.3400
Y ≤ Y_0	0.6600
Y ≥ Y_0	0.3400

시뮬레이션 검정력

알파	기각 수	기각률	95% 하한	95% 상한
0.01	209	0.418	0.37555	0.4617
0.05	320	0.64	0.59701	0.68086
0.10	364	0.728	0.68737	0.76516
0.20	417	0.834	0.79886	0.86404

" 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 X3 의 경우 유의 수준 0.05 에 대한 시뮬레이션된 검정력은 379/500 = 0.758 입니다 . 표 10.1 에는 유의 수준 0.05 에서 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 0.37 에 불과함을 알 수 있습니다 .

참고: 반응 값이 시뮬레이션되었으므로 시뮬레이션 검정력 결과는 표 10.1 에 표시된 것과 다를 수 있습니다 .

표 10.1 유의 수준 0.05 에서의 시뮬레이션 검정력

요인	X_i = 1 일 때의 합격률	X_i = -1 일 때의 합격률	감지할 차이	알파 =0.05 일 때의 시뮬레이션 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
X_4	66.82%	33.18%	33.6%	0.654

표 10.1 유의 수준 0.05 에서의 시뮬레이션 검정력 (계속)

요인	$X_i = 1$ 일 때의 합격률	$X_i = -1$ 일 때의 합격률	감지할 차이	알파 =0.05 일 때의 시뮬레이션 검정력 (기각률)
X_5	64.56%	35.43%	29.1%	0.488
X_6	62.25%	37.75%	24.5%	0.372

11 장

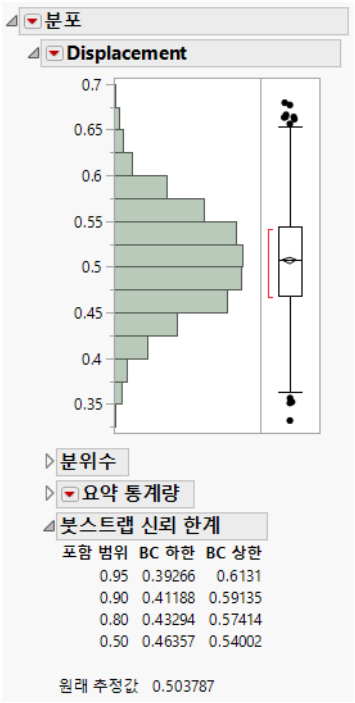
붓스트랩
재표집을 통한 통계량 분포 근사

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 모수 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 기능 개요 325
붓스트랩을 지원하는 JMP 플랫폼 326
붓스트랩의 예 327
붓스트랩 창 옵션 329
누적 붓스트랩 결과 테이블 330
비누적 붓스트랩 결과 테이블 331
붓스트랩 결과 분석 332
붓스트랩의 추가 예 333
붓스트랩에 대한 통계 상세 정보 338
부분 가중치에 대한 통계 상세 정보 338
편향 수정 백분위수 구간에 대한 통계 상세 정보 338

부스트랩 기능 개요

부스트랩은 보고서에 사용되는 관측값을 반복적으로 재표집하여 통계량의 분포를 추정하는 기법입니다. 이때 관측값은 독립적인 것으로 가정합니다.

단순 부스트랩에서는 복원 표집 방법으로 *n* 개의 관측값을 재표집하여 크기가 *n* 인 부스트랩 표본을 생성합니다 . 일부 관측값은 부스트랩 표본에 나타나지 않을 수 있으며 , 다른 관측값은 여러 번 나타날 수 있습니다 . 하나의 관측값이 부스트랩 표본에 나타나는 횟수를 **부스트랩 가중치** 라고 합니다 . 부스트랩 반복 시마다 관심 있는 통계량을 산출한 전체 분석이 다음과 같이 변경되어 다시 실행됩니다 .

- *n* 개의 관측값을 포함하는 부스트랩 표본이 데이터 집합이 됩니다 .
- 부스트랩 가중치는 분석 플랫폼의 빈도 변수입니다 .

관심 있는 통계량의 값 분포가 생성될 때까지 이 과정이 반복됩니다.

하지만 일부 경우에는 단순 부스트랩이 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 Bayes 부스트랩을 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" 및 "부분 가중치에 대한 통계 상세 정보"에서 확인하십시오.

보고서에서 부스트랩 분석을 실행하려면 부스트랩을 수행할 통계량이 포함된 테이블 열을 마우스 오른쪽 버튼으로 클릭하고 "부스트랩"을 선택합니다.

참고 : 부스트랩은 보고서에서 마우스 오른쪽 버튼을 클릭하는 방법으로만 사용할 수 있습니다 . 플랫폼 명령으로는 제공되지 않습니다 .

JMP는 많은 통계 플랫폼에서 부스트랩 기능을 제공합니다. 전체 목록은 "부스트랩을 지원하는 JMP 플랫폼"에서 확인하십시오. 표본을 구성하는 관측값은 관심 있는 통계량을 계산하는 데 사용된 모든 관측값입니다. 보고서에 빈도 열이 사용된 경우 해당 열의 관측값은 빈도 변수로 지정된 횟수만큼 반복된 것처럼 처리됩니다. 보고서에 가중치 변수가 사용된 경우 부스트랩을 사용하면 해당 변수는 보고서의 계산에서 처리되던 방식대로 처리됩니다.

팁 : 부스트랩 기능은 부스트랩이 호출된 플랫폼 보고서에 표시된 전체 분석을 다시 실행합니다 . 따라서 보고서에 포함된 관련 없는 분석으로 인해 선택한 열에 대한 부스트랩이 느리게 실행될 수 있습니다 . 부스트랩이 느리게 실행되는 경우에는 플랫폼 보고서에서 관련 없는 옵션을 제거한 후 부스트랩을 실행하십시오 .

붓스트랩을 지원하는 JMP 플랫폼

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

부스트랩의 예

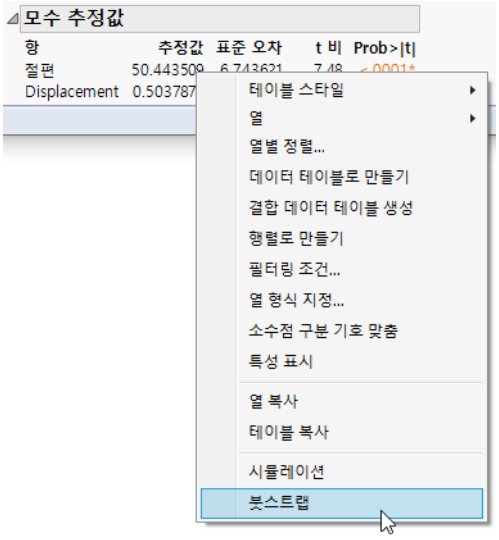
이 예에서는 분산 동질성에 대한 회귀 가정이 위배되므로 기울기에 대한 회귀 분석의 신뢰 한계가 혼동을 일으킬 수 있습니다. 이러한 이유로 기울기에 대한 신뢰 구간의 붓스트랩 추정값이 대신 사용됩니다.

1. **도움말 > 샘플 데이터 폴더**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 " 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 .

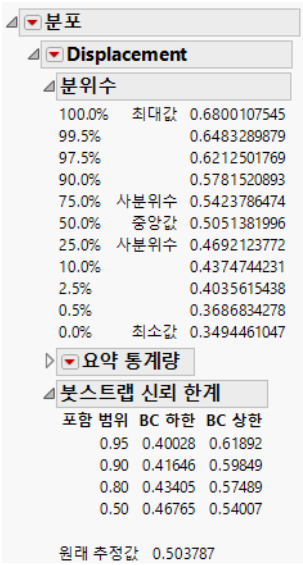
붓스트랩 프로세스가 실행된 후 붓스트랩 결과 데이터 테이블이 생성됩니다. 이때 기울기 및 절편에 대한 결과는 쌓이지 않은 상태입니다 .

다음으로 , 붓스트랩 기울기를 분석합니다 .

13. 붓스트랩 결과 테이블에서 **분포** 스크립트를 실행합니다 .

" 분포 " 보고서에는 " 붓스트랩 신뢰 한계 " 보고서가 포함됩니다 .

그림 11.3 붓스트랩 보고서



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 구간**을 나타냅니다 . 자세한 내용은 " 편향 수정 백분위수 구간에 대한 통계 상세 정보 " 에서 확인하십시오 .

부스트랩 창 옵션

붓스트랩 분석을 수행하려면 JMP 보고서의 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 Bayes 붓스트랩 분석을 수행합니다. 붓스트랩 반복 시마다 " 부분 가중치에 대한 통계 상세 정보 " 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 경우에 " 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 가중치에 대한 통계 상세 정보 " 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 붓스트랩 결과 테이블 " 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 결과를 제공하는 테이블입니다. 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.

누적 붓스트랩 결과 테이블

누적 붓스트랩 결과 테이블에는 각 통계량 및 붓스트랩 반복 조합에 대한 행과 붓스트랩 결과에 대한 단일 열이 포함됩니다. 기본적으로 붓스트랩 분석의 초기 결과는 누적 결과 테이블로 나타납니다(그림 11.4). "분할 적용 시 쌓인 테이블 삭제" 옵션을 선택한 경우 이 테이블이 표시되지 않을 수 있습니다.

그림 11.4에서는 "X로 Y 적합"의 이변량 모형을 Car Physical Data.jmp에 적합시켜 생성된 "모수 추정값" 보고서를 기반으로 한 붓스트랩 테이블을 보여 줍니다. 자세한 내용은 "붓스트랩의 예"에서 확인하십시오.

그림 11.4 누적 붓스트랩 결과 테이블

	X	Y	항	~편향	추정값	표준 오차	t 비	Prob>|t|	BootID•
1	Displacement	Horsepower	절편		50.443509471	6.7436209999	7.48	<.0001	0
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 "X 로 Y 적합 " 보고서에서는 숨겨지는 열입니다 .
- BootID 열은 붓스트랩 표본을 식별합니다 . BootID = 0 인 행은 원래 추정값에 해당합니다 . 이러한 행은 X 로 표시되며 ' 제외됨 ' 행 상태를 갖습니다 . 이 예에서는 각 붓스트랩 표본이 두 개의 행에 대한 결과 , 즉 절편에 대한 결과와 Displacement 에 대한 결과를 계산하는 데 사용되었습니다 .
- 데이터 테이블 이름에는 " 누적 붓스트랩 결과 " 가 포함됩니다 .

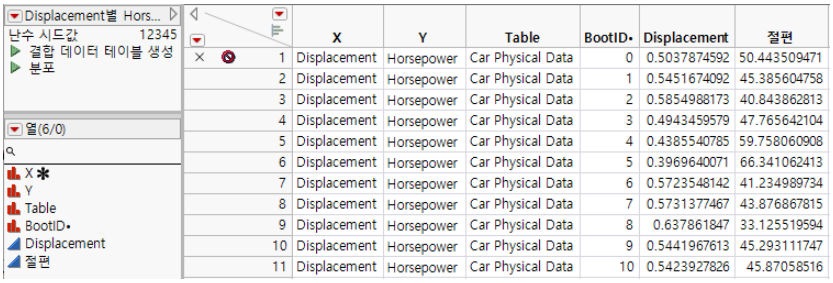
선택한 열 분할 옵션을 선택한 경우에는 비누적 결과 테이블도 나타날 수 있습니다. 자세한 내용은 "비누적 붓스트랩 결과 테이블"에서 확인하십시오.

비누적 붓스트랩 결과 테이블

비누적 붓스트랩 결과 테이블에는 각 붓스트랩 반복에 대한 단일 행과 각 붓스트랩 결과에 대한 별도의 열이 포함됩니다. 붓스트랩을 실행할 때 **선택한 열 분할**을 선택하여 비누적 붓스트랩 결과 테이블을 생성합니다.

이 예에서 그림 11.4(누적)의 항 추정값은 그림 11.5(비누적)의 Displacement 및 Intercept 라는 두 개의 열로 분할됩니다.

그림 11.5 비누적 붓스트랩 결과 테이블


Displacement별 Hors...
난수 시드값 12345
결합 데이터 테이블 생성
분포
열(6/0)
X ✱
Y
Table
BootID•
Displacement
절편

	X	Y	Table	BootID•	Displacement	절편
1	Displacement	Horsepower	Car Physical Data	0	0.5037874592	50.443509471
2	Displacement	Horsepower	Car Physical Data	1	0.5451674092	45.385604758
3	Displacement	Horsepower	Car Physical Data	2	0.5854988173	40.843862813
4	Displacement	Horsepower	Car Physical Data	3	0.4943459579	47.765642104
5	Displacement	Horsepower	Car Physical Data	4	0.4385540785	59.758060908
6	Displacement	Horsepower	Car Physical Data	5	0.3969640071	66.341062413
7	Displacement	Horsepower	Car Physical Data	6	0.5723548142	41.234989734
8	Displacement	Horsepower	Car Physical Data	7	0.5731377467	43.876867815
9	Displacement	Horsepower	Car Physical Data	8	0.637861847	33.125519594
10	Displacement	Horsepower	Car Physical Data	9	0.5441967613	45.293111747
11	Displacement	Horsepower	Car Physical Data	10	0.5423927826	45.87058516

비누적 결과 테이블의 특징은 다음과 같습니다 .

- 각 붓스트랩 표본마다 하나의 행이 있습니다 .
- 분석에 사용된 데이터 테이블 열은 이 테이블에 나타납니다 . 이 예에서는 X 가 Displacement 이고 Y 가 Horsepower 입니다 .
- 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 = 0 인 행은 원래 추정값에 해당합니다 . 이 행은 X 로 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 .

붓스트랩 결과 분석

분포 플랫폼을 사용하여 붓스트랩 결과를 분석할 수 있습니다 .

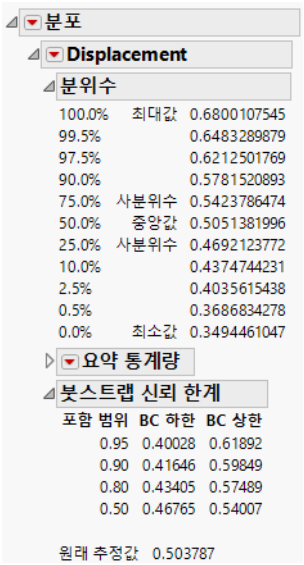
- 분석을 통해 비누적 붓스트랩 결과 테이블이 생성된 경우에는 테이블에서 " 분포 " 스크립트를 실행합니다 .
- 분석을 통해 누적 붓스트랩 결과 테이블이 생성된 경우에는 **분석 > 분포**를 선택하고 관심 있는 열을 적절한 역할에 할당합니다 . 대부분의 경우 보고서 테이블의 첫 번째 열에 해당하는 열을 기준 역할에 할당하는 것이 좋습니다 .

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 구간**을 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 구간**을 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 구간에 대한 통계 상세 정보 " 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 기능 개요"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 (Bayes 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 wabash의 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"
- "Bayes 붓스트랩 분석"

단순 붓스트랩 분석

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Snapdragon.jmp 를 엽니다 .
2. **분석 > X 로 Y 적합**을 선택합니다 .
3. Y 를 선택하고 **Y, 반응**을 클릭합니다 .
4. Soil 을 선택하고 **X, 요인**을 클릭합니다 .
5. **확인**을 클릭합니다 .
6. "Y 대 Soil 의 일원 분석 " 옆의 빨간색 삼각형을 클릭하고 **평균 /ANOVA** 를 선택합니다 .
7. **일원 ANOVA 에 대한 평균** 보고서에서 **평균** 열을 마우스 오른쪽 버튼으로 클릭하고 **붓스트랩**을 선택합니다 .
8. **붓스트랩 표본 수**에 1000 을 입력합니다 .
9. (선택 사항) 그림 11.7 의 결과와 매칭되는 결과를 얻으려면 **난수 시드값**에 12345 를 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 11.7 의 결측값은 해당 붓스트랩 표본에 주어진 토양 유형의 관측값이 모두 선택되지 않은 붓스트랩 반복을 나타냅니다 .

11. **분석 > 분포**를 선택합니다 .
12. wabash 를 선택하고 **Y, 열**을 클릭합니다 .
13. **확인**을 클릭합니다 .

그림 11.8 단순 붓스트랩으로 구한 wabash 평균의 분포

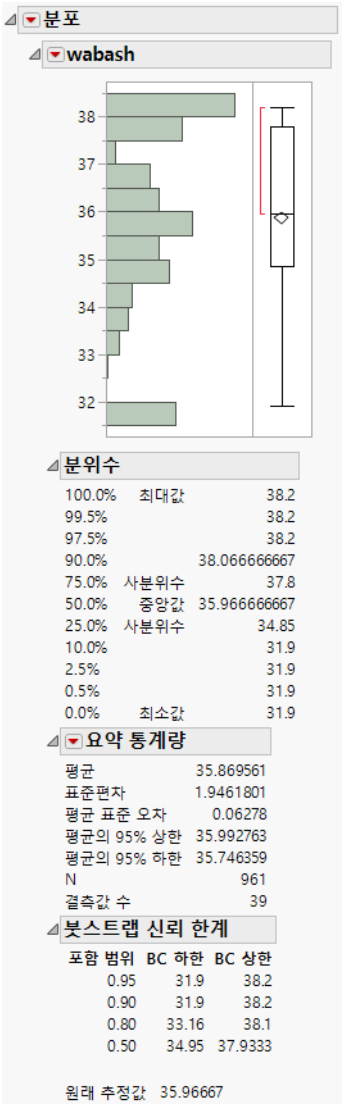


그림 11.8 에서는 단순 붓스트랩 분석으로 구한 wabash 평균의 분포를 보여 줍니다 . 다음 사항에 유의하십시오 .

- " 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 붓스트

랩 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 (Bayes 붓스트랩) 옵션을 사용하여 붓스트랩 표본에서 결측값이 발생하지 않도록 하고 붓스트랩 평균의 분포를 평활하게 합니다 .

Bayes 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 .

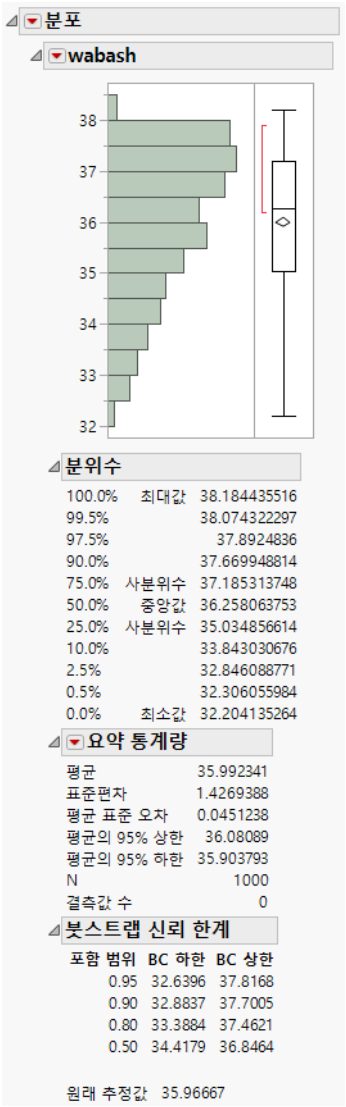
그림 11.9 Bayes 붓스트랩에 대한 붓스트랩 결과

	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

Bayes 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 Bayes 붓스트랩으로 구한 wabash 평균의 분포



Bayes 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 .

붓스트랩에 대한 통계 상세 정보

이 섹션에는 붓스트랩에 대한 통계 상세 정보가 포함되어 있습니다 .

- " 부분 가중치에 대한 통계 상세 정보 "
- " 편향 수정 백분위수 구간에 대한 통계 상세 정보 "

부분 가중치에 대한 통계 상세 정보

이 섹션에서는 붓스트랩 분석에서 부분 가중치를 계산하는 방법을 설명합니다. 부분 가중치 옵션은 Bayes 붓스트랩 (Rubin 1981) 을 기반으로 합니다. 하나의 관측값이 주어진 붓스트랩 표본에 나타나는 횟수를 **붓스트랩 가중치**라고 합니다. 단순 붓스트랩에서는 각 붓스트랩 표본에 대한 붓스트랩 가중치가 단순 랜덤 복원 표집 방식으로 결정됩니다.

Bayes 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 붓스트랩 가중치를 생성하는 데 사용됩니다 .

- 형상 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 형상 모수로 1 을 사용합니다 . JMP 에서 사용되는 형상 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: 분석에 빈도 변수가 지정된 경우에는 행별로 감마 분포의 형상 모수에 빈도 값을 곱하십시오 . 빈도 변수 값의 합은 1 보다 커야 합니다 . 그러면 형상 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.

편향 수정 백분위수 구간에 대한 통계 상세 정보

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 - α) 편향 수정 신뢰 구간의 끝점은 붓스트랩 분포의 분위수로 주어집니다 .

- 하한점은 다음 분위수입니다 .

$$\Phi\left(2z_0 + z_{\frac{\alpha}{2}}\right)$$

- 상한점은 다음 분위수입니다 .

$$\Phi\left(2z_0 + z_{1-\frac{\alpha}{2}}\right)$$

12 장

텍스트 탐색기
데이터의 비정형 텍스트 탐구

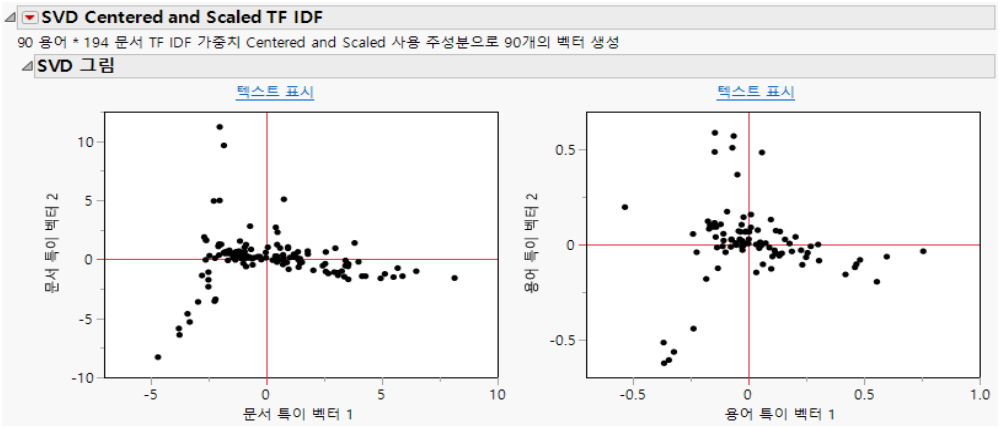
JMP PRO 이 플랫폼의 많은 기능은 JMPPro 에서만 사용할 수 있으며 그러한 경우 이 아이콘이 표시됩니다 .

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 도구가 포함되어 있습니다. 텍스트 문서를 군집화하거나 문서 모음에 있는 용어를 군집화할 수 있습니다. 잠재 계층 분석을 사용하여 문서를 군집화할 수도 있습니다.

JMP PRO JMP Pro 버전의 플랫폼에는 문서의 중요 용어와 감정을 식별하기 위한 도구도 포함되어 있습니다. 용어 선택을 사용하면 서로 다른 응답을 가장 잘 설명하는 용어를 식별할 수 있습니다. 감정 분석을 사용하면 어휘 분석을 사용하여 문서에서 감정 용어를 식별하고 긍정, 부정 및 전반적 감정에 대해 문서를 스코어링할 수 있습니다. 감정 분석 기능에는 기본적인 NLP(자연어 처리) 지원이 포함됩니다.

그림 12.1 텍스트 탐색기의 SVD 그림



목차

텍스트 탐색기 플랫폼 개요 343
텍스트 처리 단계 344
텍스트 탐색기 플랫폼의 예 345
텍스트 탐색기 플랫폼 시작 349
정규 표현식 편집기에서 정규 표현식 사용자 정의 351
텍스트 탐색기 보고서 356
요약 개수 보고서 357
용어 및 구 목록 357
텍스트 탐색기 플랫폼 옵션 360
텍스트 준비 옵션 361
텍스트 분석 옵션 367
저장 옵션 368
텍스트 탐색기의 보고서 옵션 369
잠재 계층 분석 370
잠재 의미 분석 (SVD) 371
SVD 보고서 372
SVD 보고서 옵션 373
주제 분석 375
주제 분석 보고서 376
주제 분석 보고서 옵션 376
판별 분석 377
판별 분석 보고서 377
판별 분석 보고서 옵션 378
용어 선택 379
용어 선택 설정 379
용어 선택 보고서 380
용어 선택 보고서 옵션 382
감정 분석 382
감정 분석 보고서 384
감정 분석 보고서 옵션 386
텍스트 탐색기 플랫폼의 추가 예 387

텍스트 탐색기 플랫폼 개요

텍스트 탐색기 플랫폼을 사용하면 비정형 텍스트를 탐색하여 텍스트의 의미를 더 잘 이해할 수 있습니다. 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 분석은 대개 반복적인 프로세스이므로 용어 목록의 큐레이팅과 분석을 번갈아가며 수행할 수 있습니다.

용어 목록 큐레이팅

텍스트 분석에는 몇 가지 고유한 용어가 사용됩니다. **용어** 또는 **토큰**은 가장 작은 텍스트 조각으로, 문장 내의 단어와 유사합니다. 하지만 용어는 정규 표현식을 사용하는 등의 여러 가지 방법으로 정의할 수 있습니다. 텍스트를 용어로 분해하는 프로세스를 **토큰화**라고 합니다.

- **구**는 짧은 용어 모음입니다. 텍스트 탐색기 플랫폼에는 자체적으로 용어로 지정된 구를 관리하는 옵션이 있습니다.
- **문서**는 단어 모음을 나타냅니다. JMP 데이터 테이블에서는 텍스트 열의 각 행에 있는 비정형 텍스트가 문서에 해당합니다.
- **말뭉치**는 문서 모음을 나타냅니다.

때로는 분석에서 몇 가지 일반적인 단어를 제외하는 것이 좋습니다. 이 제외된 단어를 **중지 단어**라고 합니다. 플랫폼에 기본 중지 단어 목록이 있지만 사용자가 특정 단어를 중지 단어로 추가할 수도 있습니다. 중지 단어는 용어가 될 수 없지만 구에 사용될 수 있습니다. 그러나 구는 중지 단어로 시작하거나 끝날 수 없습니다.

용어를 재코딩할 수도 있습니다. 그러면 동의어를 하나의 공통 용어로 결합하는 데 유용합니다.

어간 추출은 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 마찬가지 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. DTM의 각 행은 하나의 문서(JMP 데이터 테이블의 텍스트 열에 있는 셀 하나)에 해당합니다. DTM의 각 열은 큐레이팅된 용어 목록의 용어 하나에 해당합니다. 이 방법은 단어 순서를 무시하기 때문에 BoW 방법을 구현합니다. 가장 단순한 형태일 때 DTM의 각 셀에는 행의 문서에 있는 열의 용어 빈도(발생 횟수)가 포함됩니다. DTM을 위한 다른 가중치 체계가 다양하게 존재하며, 이는 "저장 옵션"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 **다변량 방법**의 에서 확인하십시오. 용어 군집화 및 문서 군집화에는 "계층적 군집화" 옵션을 사용할 수 있습니다. 이 옵션을 사용하면 유사한 용어 또는 문서끼리 그룹화할 수 있습니다.

텍스트 탐색기 플랫폼 워크플로우

텍스트 탐색기 플랫폼을 사용하기 위해 필요한 단계가 있습니다 .

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 Pro 에서만 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 Pro 에서만 사용할 수 있습니다 .

6. 구 , 재코딩 및 중지 단어 특성을 유사한 텍스트 데이터의 향후 분석에 사용할 수 있도록 저장합니다 .

텍스트 처리 단계

텍스트 탐색기 플랫폼에서 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 전에 수행됩니다 .

참고 : 재코딩 작업은 보고서 창에서 지정된 순서와 관계없이 단일 패스로 내부적으로 처리됩니다 .

구 추출 단계

구 추출 단계에서는 말뭉치(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.

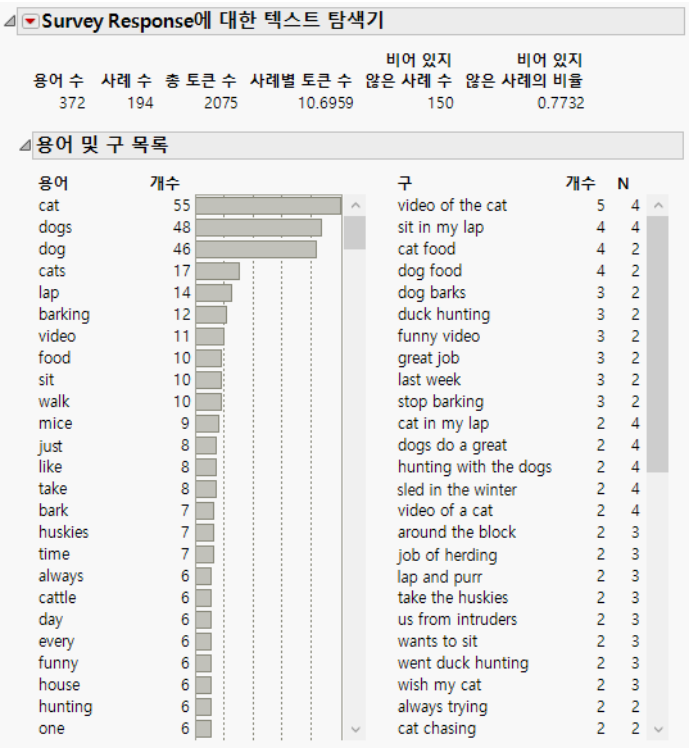
1. 용어 목록에 구를 추가합니다. 구는 용어 목록에서 어간 추출된 구의 각 단어에 어간 추출을 적용해야 합니다. 원시 토큰은 다르지만 어간이 동일한 구는 용어 목록에서 결합됩니다.
2. 구에 나타나는 토큰 용어 항목을 제거합니다.

텍스트 탐색기 플랫폼의 예

JMP 에서 텍스트 응답을 탐색하는 방법을 알아봅니다. 이 예에서는 반려 동물에 대한 설문 조사에서 얻은 텍스트 응답을 살펴보려고 합니다.

1. **도움말 > 샘플 데이터 폴더**를 선택하고 Pet Survey.jmp 를 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



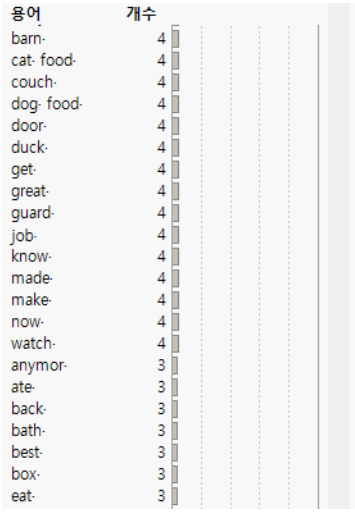
194 개 문서에 372 개의 고유 용어가 있음을 한눈에 알 수 있습니다 . 총 1921 개의 토큰화된 용어가 있습니다 . 가장 자주 나타나는 용어는 55 회 나타나는 "cat" 입니다 .

6. "Survey Response 에 대한 텍스트 탐색기 " 옆의 빨간색 삼각형을 클릭하고 **용어 옵션 > 어간 추출 > 모든 용어의 어간 추출**을 선택합니다 .
7. 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 .
8. 용어 목록 보고서를 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
barn	4
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
watch	4
anymor	3
ate	3
back	3
bath	3
best	3
box	3
eat	3

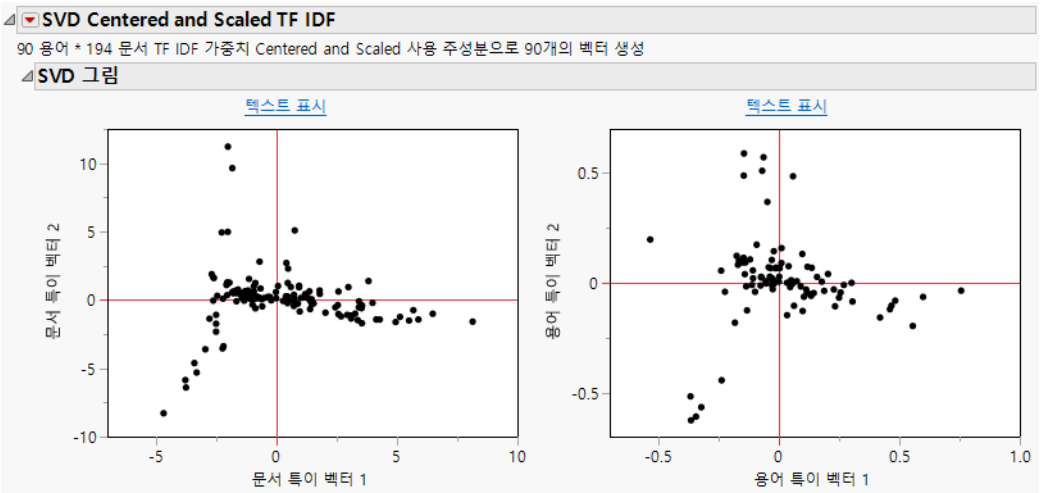
"cat food" 와 "dog food" 는 이제 이 " 텍스트 탐색기 " 보고서에 한해 용어로 처리되므로 구 목록에서 회색으로 표시됩니다 .

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

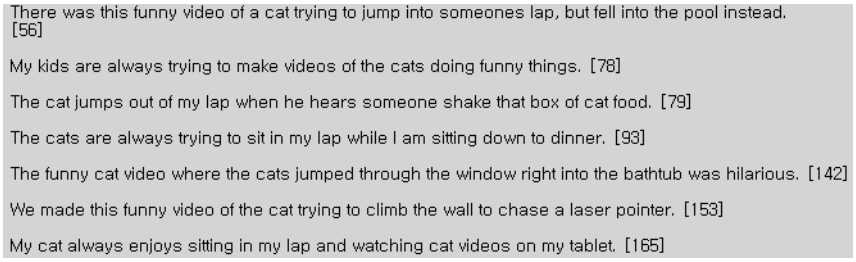


11. JMP PRO 왼쪽 SVD 그림에서 맨 오른쪽에 있는 점 7 개를 선택합니다.

    이 7 개 점은 나머지 점과 떨어져서 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



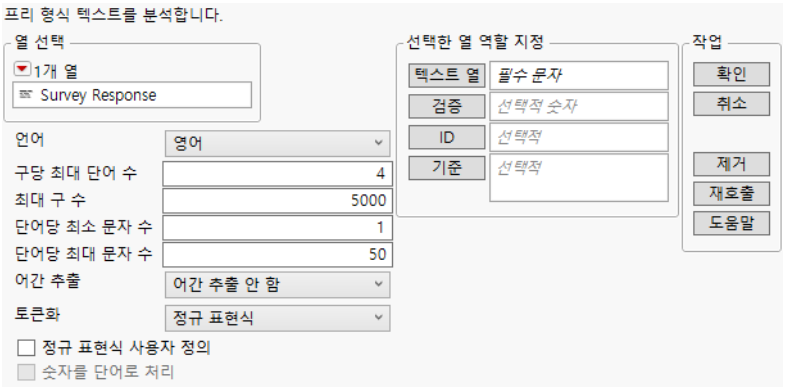
선택한 점이 나타내는 7 개 문서의 텍스트가 포함된 창이 나타납니다. 이러한 설문 조사 응답은 모두 "funny", "cat" 및 "video" 의 몇 가지 조합을 언급한다는 점에서 유사합니다. 이러한 문서는 첫 번째 특이 벡터에 대해 나머지 문서보다 큰 양의 값을 가집니다. 이렇게 큰 값은 이러한 문서가 해당 차원의 나머지 문서와는 다름을 나타냅니다.

특이 벡터의 차원을 더 조사해 보면 차원이 나타내는 것을 해석할 수 있습니다. 예를 들어 그림의 맨 오른쪽에 있는 많은 문서들은 고양이에 관한 응답입니다. 맨 왼쪽의 많은 응답은 개에 관한 것입니다. 따라서 첫번째 특이 벡터에서는 응답이 고양이에 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



" 열 선택 " 의 빨간색 삼각형 메뉴에 포함된 옵션에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 . 텍스트 탐색기 시작 창에는 다음 옵션이 포함되어 있습니다 .

텍스트 열 텍스트 데이터가 포함된 열을 할당합니다 . 여러 열을 지정하면 각 열에 대해 별도의 분석이 생성됩니다 .

JMP PRO 검증 JMP Pro 에서는 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 **예측 및 전문 모델링**의 에서 확인하십시오 .

검증 열을 지정해도 문서 용어 행렬의 계산에는 영향이 없습니다 . 하지만 검증 열이 지정된 경우에는 " 잠재 계층 분석 ", " 잠재 의미 분석 ", " 주제 분석 " 및 " 판별 분석 " 옵션에 훈련 데이터 집합만 사용됩니다 . 검증 열은 " 용어 선택 " 옵션에 대한 일반화 회귀 검증 방법으로 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 추출과 중지 단어 , 재코딩 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 .

구당 최대 단어 수 구가 분석에 구로 포함되기 위해 최대로 포함할 수 있는 단어 수를 지정합니다.

최대 구 수 구 목록에 나타나는 최대 구 수를 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 (" 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 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 " 옵션을 사용하면 용어 목록에 추가된 구에도 영향을 줍니다. 구 내의 용어가 어간 추출된 후에는 구 식별이 발생합니다. 예를 들어 "dogs bark" 와 "dog barks" 는 모두 "dog· bark " 라는 특정 구와 매칭됩니다. " 어간 추출 " 옵션을 선택한 경우 용어 목록에서 구를 제거할 수 없습니다.

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 파싱하는 데 사용되는 정규 표현식을 추가, 제거 또는 편집하려면 **정규 표현식 사용자 정의** 옵션을 선택합니다. 자세한 내용은 " 정규 표현식 편집기에서 정규 표현식 사용자 정의 " 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 " 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 .

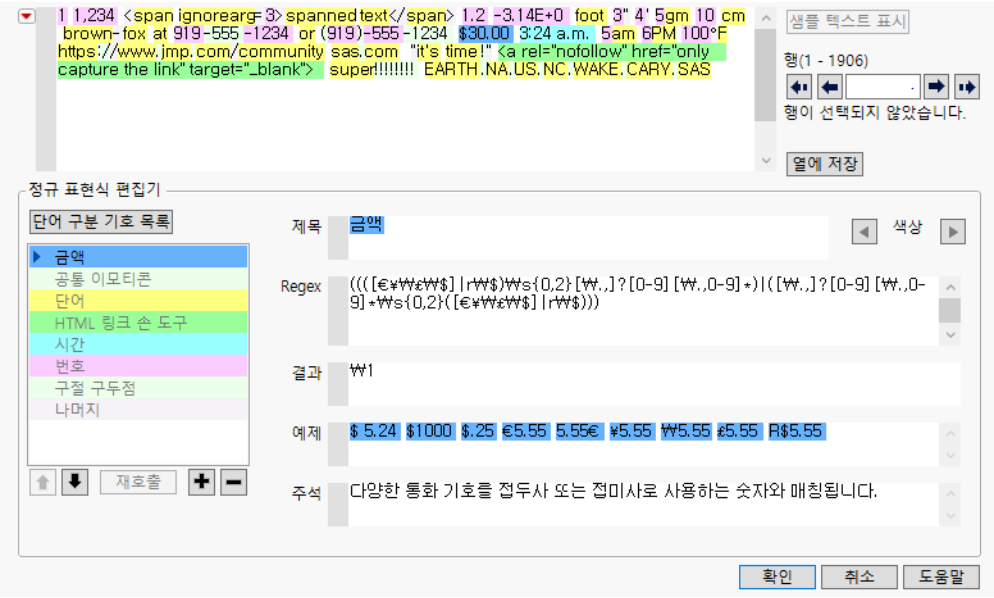
정규 표현식 편집기에서 정규 표현식 사용자 정의

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.

참고 : " 정규 표현식 사용자 정의 " 옵션은 기본 정규 표현식 방법으로 원하는 결과를 얻을 수 없는 경우에만 사용하는 것이 좋습니다 . 이러한 경우는 텍스트에 기본 정규 표현식 방법으로 인식되지 않는 구조가 포함된 경우에 발생할 수 있습니다 .

팁 : 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 .

그림 12.7 텍스트 탐색기 정규 표현식 편집기



스크립트 편집기 상자를 사용한 파싱

창 맨 위의 스크립트 편집기 상자에서는 샘플 텍스트에 대한 파싱 진행 방식을 보여 줍니다. 정규 표현식 편집기 목록의 정규 표현식을 파싱한 결과는 정규 표현식 편집기 목록의 색상에 해당하는 색상으로 강조 표시됩니다.

- 스크립트 편집기 상자에 사용자 데이터의 텍스트를 채우려면 **첫 번째**, **이전**, **다음** 및 **마지막** 행 버튼을 클릭합니다. 이렇게 하면 지정된 텍스트 데이터 행이 파싱되는 방식을 볼 수 있습니다. 편집 상자에 행 번호를 입력하여 스크립트 편집기 상자를 데이터 테이블의 특정 행에 있는 텍스트로 채울 수도 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" 에서 확인하십시오. **열 > 유틸리티 > 텍스트 매칭으로 새 열 생성에서** 정규 표현식 편집기에 액세스하면 **열에 저장** 버튼이 나타나지 않습니다.

참고 : 열에 저장 버튼을 클릭할 경우 텍스트 매칭에 정규 표현식만 사용됩니다. 정규 표현식의 출력을 수정하는 데 중지 단어, 재코딩, 어간 추출, 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.

재호출 버튼을 클릭하여 정규 표현식 목록을 정규 표현식 편집기의 가장 최근 인스턴스에서 사용한 정규 표현식으로 채웁니다. 재호출된 정규 표현식은 편집기의 이전 인스턴스에서 **열에 저장** 버튼이나 **확인** 버튼을 클릭할 때 존재한 정규 표현식입니다.

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 식별하는 데 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 X 가 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 문자**는 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.

다음 단계에서는 지정된 정규 표현식과 필요한 "나머지" 정규 표현식의 처리에 대해 설명합니다.

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 \1 로 설정하면 구두점 표시와 같은 구분 기호 문자를 더 많이 추가할 수 있습니다 . 이렇게 하면 지정된 구두점 표시가 결과에 포함되지 않습니다 .
- " 나머지 " 정규 표현식의 결과를 \1 로 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. 열 > 유틸리티 > 텍스트 매칭으로 새 열 생성에서 독립 실행형 정규 표현식 유틸리티를 사용할 수도 있습니다. 자세한 내용은 JMP 사용의에서 확인하십시오.

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 " 보고서에는 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 말뭉치의 문서 수입니다 .

총 토큰 수 말뭉치의 총 용어 수입니다 .

사례별 토큰 수 토큰 수를 사례 수로 나눈 것입니다 .

비어 있지 않은 사례 수 한 개 이상의 용어가 포함된 말뭉치 내 문서의 개수입니다 .

비어 있지 않은 사례의 비율 한 개 이상의 용어가 포함된 말뭉치 내 문서의 비율입니다 .

용어 및 구 목록

"텍스트 탐색기" 보고서의 "용어 및 구 목록" 섹션에는 토큰화 수행 후 텍스트에 있는 용어 및 구의 테이블이 포함됩니다. "용어 및 구 목록" 보고서의 예는 그림 12.8에서 확인하십시오. 용어 목록의 "개수" 열은 말뭉치에서 각 용어가 나타난 횟수를 나타냅니다. 구 목록의 "개수" 열은 말뭉치에서 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"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 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 용어가 사전순으로 정렬됩니다 . 이 옵션을 선택하지 않으면 용어가 개수 기준 내림차순으로 정렬됩니다 .

수치 순서 (" 사전순 " 옵션을 선택한 경우에만 사용할 수 있습니다 .) 용어 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 항목이 문자열 및 숫자 세그먼트로 분할되고 숫자 세그먼트가 수치 순서로 정렬됩니다 . " 수치 순서 " 옵션에서 사용되는 정렬 규칙에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

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 에서 확인하십시오 .

중지 단어 추가 선택된 용어를 중지 단어 목록에 추가하고 용어 목록에서 해당 용어를 제거합니다 . 이 작업으로 구 목록도 업데이트됩니다 .

참고 : 어간 추출된 단어를 중지 단어로 추가하면 해당 어간과 일치하는 모든 토큰이 중지 단어로 추가됩니다 .

어간 예외 추가 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 선택된 용어를 어간 추출에서 제외된 용어의 목록에 추가합니다 .

구 제거 (지정된 구가 용어 목록에서 선택되고 " 어간 추출 " 방법이 " 어간 추출 안 함 " 으로 선택된 경우에만 사용 가능) 지정된 구 집합에서 선택된 구를 제거하고 그에 따라 용어 개수를 업데이트합니다 .

참고 : 구가 감정 구로 추가된 경우 " 구 제거 " 옵션은 현재 감정 분석 보고서의 감정 용어 목록에서도 구를 제거합니다 .

JMP PRO **감정 추가** (현재 보고서 창에 감정 분석 보고서가 열려 있는 경우에만 사용할 수 있습니다 .) 선택한 용어를 현재 감정 분석 보고서의 감정 용어 목록에 추가합니다 .

참고 : 어간 추출된 단어를 감정 용어로 추가하면 해당 어간과 일치하는 모든 토큰이 감정 용어로 추가됩니다 .

필터 표시 용어 목록 위에 검색 필터를 표시하거나 숨깁니다 . 자세한 내용은 " 검색 필터 옵션 " 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 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 용어가 사전순으로 정렬됩니다 . 이 옵션을 선택하지 않으면 용어가 개수 기준 내림차순으로 정렬됩니다 .

수치 순서 (" 사전순 " 옵션을 선택한 경우에만 사용할 수 있습니다 .) 구 목록 정렬 순서를 지정합니다 . 이 옵션을 선택하면 항목이 문자열 및 숫자 세그먼트로 분할되고 숫자 세그먼트가 수치 순서로 정렬됩니다 . " 수치 순서 " 옵션에서 사용되는 정렬 규칙에 대한 자세한 내용은 JMP 사용의 에서 확인하십시오 .

복사 선택된 구를 클립보드에 추가합니다 .

포함하는 항목 선택 구 목록에서 선택된 구를 포함하는 더 큰 구를 선택합니다 .

포함된 항목 선택 구 목록 및 용어 목록에서 선택된 구에 포함된 더 작은 구 및 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.

JMP PRO 감정 구 추가 (현재 보고서 창에 감정 분석 보고서가 열려 있는 경우에만 사용할 수 있습니다.) 선택한 구를 용어 목록 및 현재 "감정 분석" 보고서의 감정 용어 목록에 추가합니다.

필터 표시 구 목록 위에 검색 필터를 표시하거나 숨깁니다. 자세한 내용은 "검색 필터 옵션"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 반환됩니다. "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 "
- " 텍스트 탐색기의 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 " 에서 확인하십시오 .

너비를 변경하는 방법으로 단어 클라우드의 크기를 대화식으로 조정할 수 있습니다 . 이 때 높이는 자동으로 결정됩니다 . 용어 목록의 행은 단어 클라우드의 용어에 연결됩니다 .

용어 목록 표시 용어 목록 보고서를 표시하거나 숨깁니다 .

구 목록 표시 구 목록 보고서를 표시하거나 숨깁니다 .

용어 및 구 옵션 표시 " 용어 및 구 목록 " 보고서에서 각 목록의 팝업 메뉴에서 사용할 수 있는 옵션에 해당하는 버튼을 표시하거나 숨깁니다 . 자세한 내용은 " 용어 및 구 목록 " 에서 확인하십시오 .

요약 개수 표시 요약 개수 테이블을 표시하거나 숨깁니다 . 자세한 내용은 " 요약 개수 보고서 " 에서 확인하십시오 .

중지 단어 표시 분석에 사용된 중지 단어의 목록을 표시하거나 숨깁니다 . 처음에는 기본 제공 중지 단어 목록이 사용됩니다 . 중지 단어를 추가하려면 용어 목록 보고서를 마우스 오른쪽 버튼으로 클릭한 후 팝업 메뉴에서 **중지 단어 추가**를 선택합니다 . 자세한 내용은 " 용어 옵션 관리 창 " 에서 확인하십시오 .

재코딩 표시 재코딩된 용어의 목록을 표시하거나 숨깁니다 . 자세한 내용은 " 용어 옵션 관리 창 " 에서 확인하십시오 .

지정한 구 표시 사용자가 용어로 처리되도록 지정한 구의 목록을 표시하거나 숨깁니다 . 자세한 내용은 " 용어 옵션 관리 창 " 에서 확인하십시오 .

어간 예외 표시 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출에서 제외된 용어를 표시하거나 숨깁니다 . 자세한 내용은 " 용어 옵션 관리 창 " 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 JSL 에서 `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 안 함 " 인 경우에만 사용 가능) 두 개의 어간 추출 결과 테이블이 포함된 " 어간 추출 " 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 " 에서 확인하십시오 .

용어 옵션 용어 목록 보고서에 적용되는 옵션 하위 메뉴를 표시합니다 .

어간 추출 (" 언어 " 옵션이 영어 , 독일어 , 스페인어 , 프랑스어 또는 이탈리아어로 설정된 경우에만 사용 가능) 어간 추출 옵션에 대한 자세한 내용은 " 텍스트 탐색기 플랫폼 시작 " 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 " 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 " 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 " 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 " 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 " 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 클라우드에서 용어의 배열을 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 클라우드에서 용어의 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 스코어링 " 옵션으로 생성된 용어 스코어를 기준으로 합니다 . 범례에서 색상과 그래디언트를 조정할 수 있습니다 .

글꼴 단어 클라우드에서 용어의 글꼴 , 스타일 및 크기를 지정합니다 .

범례 표시 단어 클라우드에 대한 범례를 표시하거나 숨깁니다 .

용어 옵션 관리 창

다양한 범위에 대해 중지 단어, 재코딩, 구 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 스크립

트를 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"에서 확인하십시오.

그림 12.9 중지 단어 관리 창

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 주 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 프로젝트 내에서 시작된 경우에만 사용할 수 있습니다.) 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 시작할 때 이러한 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 시작할 때 이러한 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 >/"

주 사용자 라이브러리 파일은 TextExplorer 디렉터리 자체에 있습니다. 이러한 파일은 언어별로 다르지 않습니다.

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.txt로 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.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "재코딩"이라고 합니다.

구 관리

사용자 라이브러리에 구를 저장하는 경우 파일 이름은 phrases.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "구"라고 합니다.

어간 예외 관리

사용자 라이브러리에 어간 예외를 저장하는 경우 파일 이름은 stemExceptions.txt로 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 .

JMP PRO 부정 용어 관리

"부정 용어 관리" 창은 감정 분석 보고서에서 사용할 수 있습니다. 자세한 내용은 "감정 분석"에서 확인하십시오.

사용자 라이브러리에 부정 용어를 저장하는 경우 파일 이름은 negations.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "부정 용어"라고 합니다.

참고: "지역" 또는 "지역 예외" 목록에 나타나는 용어는 현재 감정 분석 보고서에만 적용됩니다.

JMP PRO 강조사 용어 관리

감정 분석 보고서에서 사용할 수 있는 "강조사 용어 관리" 창은 "중지 단어 관리" 창과는 약간 다릅니다. 각 목록 아래에 텍스트 편집 상자가 있을 뿐만 아니라 승수 컨트롤도 있습니다. 승수 컨트롤을 사용하면 강조사 용어 집합에 추가할 때 용어의 강조 승수를 지정할 수 있습니다. 자세한 내용은 "감정 분석"에서 확인하십시오.

사용자 라이브러리에 강조사 용어를 저장하는 경우 파일 이름은 intensifiers.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "강조사 용어"라고 합니다.

참고: "지역" 또는 "지역 예외" 목록에 나타나는 용어는 현재 감정 분석 보고서에만 적용됩니다.

JMP PRO 감정 용어 관리

감정 분석 보고서에서 사용할 수 있는 "감정 용어 관리" 창은 "중지 단어 관리" 창과는 약간 다릅니다. 각 목록 아래에 텍스트 편집 상자가 있을 뿐만 아니라 "스코어" 컨트롤도 있습니다. 스코어 컨트롤을 사용하면 감정 용어 집합에 추가할 때 용어의 감정 스코어를 지정할 수 있습니다. 자세한 내용은 "감정 분석"에서 확인하십시오.

사용자 라이브러리에 감정 용어를 저장하는 경우 파일 이름은 sentiments.txt로 지정됩니다. 열 특성에 저장할 경우에는 해당 특성을 "감정 용어"라고 합니다.

참고: "지역" 또는 "지역 예외" 목록에 나타나는 용어는 현재 감정 분석 보고서에만 적용됩니다.

JMP PRO 텍스트 분석 옵션

텍스트 탐색기의 빨간색 삼각형 메뉴에는 분석을 위한 다음 옵션이 포함되어 있습니다 .

잠재 계층 분석 희소 행렬 루틴을 사용하여 이진 가중 문서 용어 행렬에 대해 잠재 계층 분석을 수행합니다 . 자세한 내용은 " 잠재 계층 분석 " 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)" 에서 확인하십시오 .

판별 분석 문서 용어 행렬을 기준으로 그룹 또는 범주에서 각 문서의 소속을 예측합니다 . 자세한 내용은 " 판별 분석 " 에서 확인하십시오 .

용어 선택 서로 다른 응답을 가장 잘 설명하는 용어를 분석합니다 . 용어 선택은 응답이 평가인 경우 감정 분석에도 유용할 수 있습니다 . 자세한 내용은 " 용어 선택 " 에서 확인하십시오 .

감정 분석 (" 언어 " 옵션이 영어로 설정된 경우에만 사용할 수 있습니다 .) 어휘 분석을 사용하여 문서에서 감정 용어를 식별하고 긍정 , 부정 및 전반적 감정에 대해 문서를 스코어링합니다 . 자세한 내용은 " 감정 분석 " 에서 확인하십시오 .

JMP PRO 특이값 분해 규격 창

텍스트 탐색기 플랫폼의 분석 옵션은 DTM(문서 용어 행렬)을 기반으로 합니다. DTM은 용어 목록에 있는 각 용어에 대한 열을 생성하여 구성됩니다("최대 용어 수"에 지정된 개수까지). 각 텍스트 문서(데이터 테이블의 행과 동등)는 DTM의 한 행에 해당합니다. DTM의 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 .

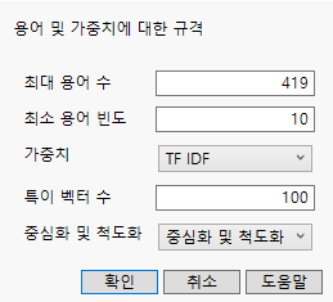
최소 용어 빈도 한 용어가 특이값 분해에 포함되기 위해 충족해야 하는 최소 발생 횟수입니다 .

가중치 문서 용어 행렬의 셀에 들어갈 값을 결정하는 가중치 체계입니다 . 가중치 체계 옵션에 대한 자세한 내용은 " 문서 용어 행렬 규격 창 " 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).

JMP PRO **연관성을 위해 쌓인 형식의 DTM 저장** 쌓인 형식의 문서 용어 행렬을 JMP 데이터 테이블에 저장합니다. 쌓인 형식은 연관성 분석 플랫폼의 분석에 적합합니다. 자세한 내용은 **예측 및 전문 모델링**의 에서 확인하십시오. 텍스트 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 인덱스에서 확인하십시오.

용어 테이블 저장 용어 목록의 각 용어, 발생 횟수 및 각 용어가 포함된 문서 수를 보여 주는 데이터 테이블을 생성합니다. " 용어 테이블 저장 " 을 선택한 후 " 열을 기준으로 용어 스코어링 " 옵션을 선택하면 용어별 스코어가 포함된 열이 " 용어 테이블 저장 " 옵션으로 생성한 데이터 테이블에 추가됩니다.

열을 기준으로 용어 스코어링 지정된 열의 값을 기준으로 한 스코어를 " 용어 테이블 저장 " 옵션으로 생성된 JMP 데이터 테이블에 저장합니다. 각 용어의 스코어는 지정된 열에 대한 평균값으로, 이 값은 각 행에서 해당 용어가 나타나는 횟수를 기준으로 가중치를 부여한 값입니다. " 용어 테이블 저장 " 옵션을 이미 선택한 경우 " 열을 기준으로 용어 스코어링 " 옵션은 " 용어 테이블 저장 " 옵션으로 생성된 데이터 테이블에 스코어를 포함하는 열을 추가합니다. 그렇지 않은 경우에는 용어 테이블에 대한 JMP 데이터 테이블이 생성됩니다 . 지정된 열이 연속형이 아니면 지정된 열의 각 수준에 대한 스코어를 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 1 을 할당하고 , 그렇지 않으면 0 을 할당합니다 . SVD 분석이 이전에 실행되었던 경우를 제외하고는 이 가중 방법이 기본값입니다 .

삼진 각 문서에서 용어가 두 번 이상 나오면 2 를 할당하고 , 한 번만 나오면 1 을 할당하며 , 나오지 않으면 0 을 할당합니다 .

빈도 각 문서에서의 용어 발생 횟수를 할당합니다 .

로그 빈도 *log10(1 + x)* 를 할당합니다 . 여기서 *x* 는 각 문서에서의 용어 발생 횟수입니다 .

TF IDF *TF * log10(nDoc / nDocTerm)* 을 할당합니다 . 용어 빈도 - 문서 빈도 역수 (term frequency - inverse document frequency) 의 약어입니다 . 이 옵션이 SVD 분석의 기본 가중치 옵션입니다 . 계산식의 항은 다음과 같이 정의됩니다 .

TF = 문서 내 용어 빈도

nDoc = 말뭉치 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 .

텍스트 탐색기의 보고서 옵션

다음 옵션에 대한 자세한 내용은 JMP 사용의 에서 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 .

플랫폼 환경 설정 현재 플랫폼 환경 설정을 보거나 , 현재 JMP 보고서의 설정과 일치하도록 플랫폼 환경 설정을 업데이트할 수 있는 옵션이 포함되어 있습니다 .

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

텍스트 탐색기 플랫폼에서 잠재 계층 분석을 통해 말뭉치의 문서를 유사 문서의 군집으로 그룹화할 수 있습니다 . 잠재 계층 분석 보고서에는 모형 규격 , 모형에 대한 BIC(Bayes 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 말뭉치에서의 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 스코어가 가장 높은 10개 용어의 테이블을 표시하거나 숨깁니다 . 군집 c 의 용어 t 에 대한 스코어 $S_{t,c}$ 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 **다변량 방법**의 에서 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 . " 최대 확률 분류 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 최대 확률 분류 군집에 대한 계산식 열을 데이터 테이블에 저장합니다.

저장되는 스코어 계산식에서는 가중치 인수를 `"LCA"` 로 설정하고 `Text Score()` JSL 함수를 사용합니다.

군집별 색상 데이터 테이블의 각 행을 최대 확률 분류 군집에 따라 색상을 적용합니다.

제거 "텍스트 탐색기" 보고서에서 "잠재 계층 분석" 보고서를 제거합니다.

잠재 계층 분석에 대한 자세한 내용은 **다변량 방법**의 에서 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)

텍스트 탐색기 플랫폼에서 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 **U**, **S** 및 **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 **U**는 DTM의 왼쪽 특이 벡터를 포함하는 *nDoc* x *nVec* 행렬입니다. **S**는 *nVec*차 대각 행렬입니다. **S**의 대각 항목은 DTM의 특이값입니다. **V**′는 *nVec* x *nTerm* 행렬입니다. **V**′의 행(또는 **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 **V**′에 이 세 개의 용어에 대해 큰 값을 갖는 특이 벡터를 생성할 가능성이 높습니다. **U** 특이 벡터는 이 새로운 용어 공간에 투영된 문서를 나타냅니다.

잠재 의미 분석에서는 간접적 연결성도 포착합니다. 두 개의 단어가 동일한 문서에 함께 나타나지는 않지만 일반적으로 다른 세 번째 단어가 있는 문서에서 나타나는 경우 SVD는 해당 연결성

의 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.

SVD는 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 DTM을 중심화하고 척도화한 후 *nDoc* - 1로 나눕니다. 이 분석은 DTM 상관 행렬의 PCA와 동등합니다.

" 규격 " 창에서 " 중심화 " 또는 " 비중심화 " 를 지정할 수도 있습니다 .

- " 중심화 " 를 지정할 경우 특이값 분해 전에 DTM 이 중심화된 후 *nDoc* - 1 로 나뉩니다 . 이 분석은 DTM 공분산 행렬의 PCA 와 동등합니다 .
- " 비중심화 " 를 지정할 경우에는 특이값 분해 전에 DTM 이 *nDoc* 로 나뉩니다 . 이 분석은 비척도화 DTM 의 PCA 와 동등합니다 .

SVD 구현에서는 DTM이 중심화된 경우에도 DTM의 희소성을 활용합니다.

JMP PRO SVD 보고서

텍스트 탐색기 플랫폼의 " 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 대각 특이값 행렬(**S**)로 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 스코어 그림과 동일합니다. 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 대각 특이값 행렬(**S**)로 곱한 값으로 정의됩니다. 이 그림은 주성분 플랫폼의 적재 그림과 동일합니다. 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. 이러한 특이값은 문서 용어 행렬의 특이값 분해에서 **S** 행렬의 대각 항목입니다. 특이값 테이블에는 동등한 주성분 분석의 해당 고유값이 들어 있는 열이 포함되어 있습니다. 주성분 플랫폼에서와 마찬가지로 각 고유값(또는

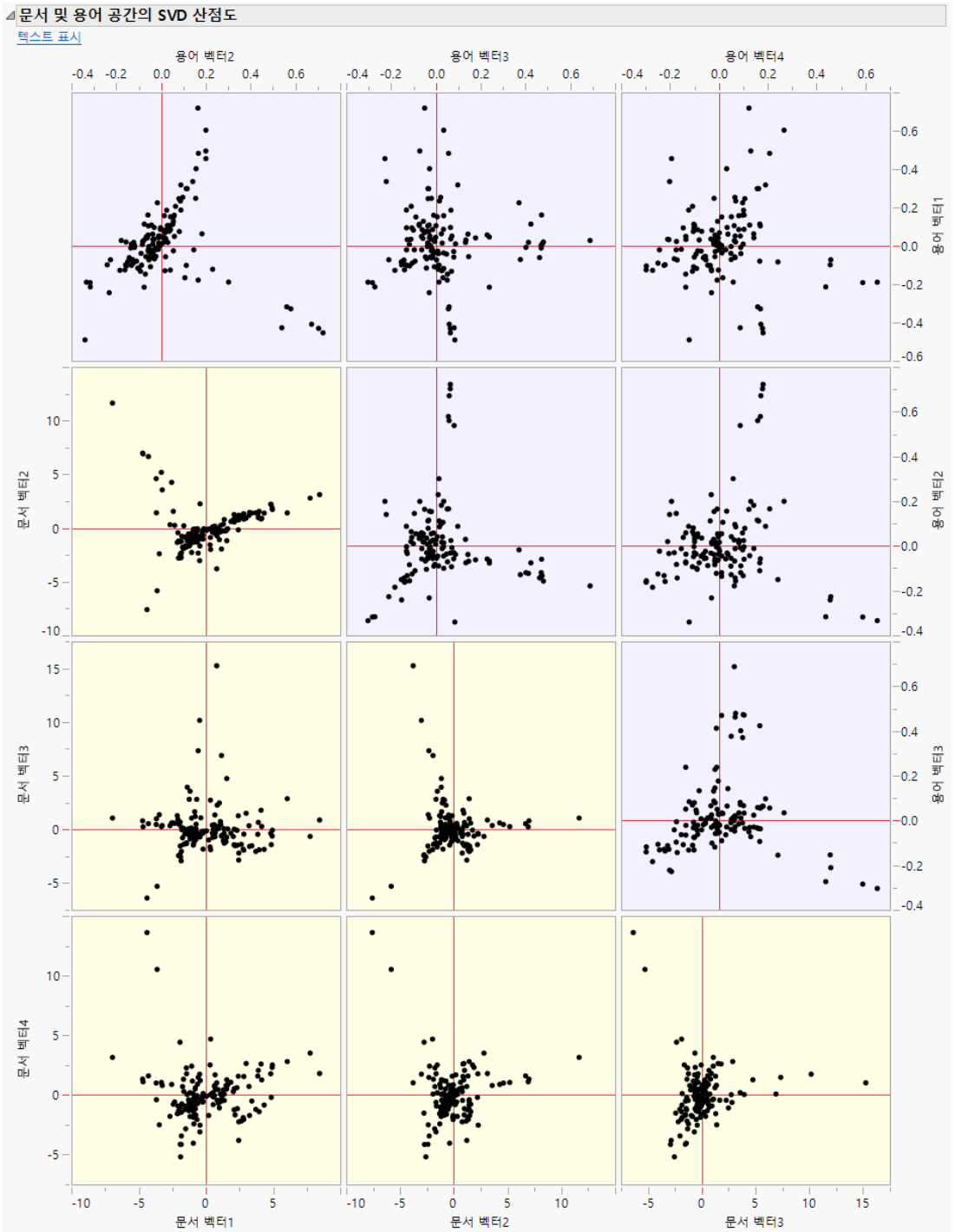
특이값)이 설명하는 변동의 백분율 및 누적 백분율을 나타내는 열이 있습니다. "누적 백분율" 열을 사용하여 DTM에서 유지할 분산의 백분율을 결정한 다음 해당하는 수의 특이 벡터를 사용할 수 있습니다.

JMP PRO SVD 보고서 옵션

텍스트 탐색기 플랫폼에서 "SVD" 의 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 " 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 **다변량 방법**의 에서 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 .

근접 이웃 선택 문서 SVD 그림에서 선택한 점의 최근접 이웃을 찾아 선택합니다 . 이 알고리즘은 특이값 분해의 **U** 행렬을 사용하여 최근접 이웃을 찾습니다 . 이 옵션을 선택할 때 선택할 최근접 이웃 수를 지정해야 합니다 . 기본적으로 이 옵션은 10 개의 최근접 이웃을 선택합니다 .

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)" 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 인덱스에서 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)" 에서 확인하십시오 .

제거 텍스트 탐색기 보고서 창에서 SVD 보고서를 제거합니다 .

JMP PRO 주제 분석

텍스트 탐색기 플랫폼의 " 주제 분석 , 회전된 SVD" 옵션은 DTM(문서 용어 행렬) 의 부분 SVD (특이값 분해) 에 대해 Varimax 회전을 수행합니다 . DTM 에서 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

텍스트 탐색기 플랫폼의 "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 PCA 의 요인 적재 행렬과 동등합니다.

주제별 단어 클라우드 각 주제당 하나씩의 단어 클라우드가 포함됩니다.

주제 스코어 각 주제의 문서 스코어 행렬이 포함됩니다. 주제 스코어가 높은 문서는 해당 주제와 연관성이 있을 가능성이 높습니다.

주제 스코어 그림 " 텍스트 표시 " 버튼과 각 문서의 주제 스코어 그림이 포함됩니다. " 텍스트 표시 " 버튼을 클릭하면 선택한 문서의 텍스트가 포함된 창이 열립니다.

" 주제 스코어 그림 " 보고서는 주제 스코어 보고서의 행렬을 시각적으로 나타낸 것입니다. 그림의 각 패널은 주제 중 하나, 또는 주제 스코어 행렬의 열 중 하나에 해당합니다. 각 패널 내에서 각 점은 말뭉치의 문서 중 하나, 또는 주제 스코어 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

텍스트 탐색기 플랫폼에서 " 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 " 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 회전된 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 인덱스에서 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

텍스트 탐색기 플랫폼의 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 DTM 의 열을 통해 예측됩니다 . 판별 분석에 대한 자세한 내용은 **다변량 방법** 의 에서 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 탐색기 플랫폼의 "판별 분석" 옵션은 DTM(문서 용어 행렬)을 기반으로 합니다. DTM은 용어 목록에 있는 각 용어에 대한 열을 생성하여 구성됩니다("최대 용어 수"에 지정된 개수까지). 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 " 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 스코어 보고서가 포함됩니다. 다른 보고서는 처음에는 닫혀 있습니다.

판별 분석 보고서에는 다음 보고서가 포함됩니다 .

용어 평균 판별 분석에 사용된 용어의 테이블을 제공합니다 . 이러한 용어는 DTM 의 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 제곱한 결과를 포함하는 테이블을 제공합니다 . Mahalanobis 거리에 대한 자세한 내용은 **다변량 방법**의 에서 확인하십시오 .

각 그룹에 대한 확률 문서가 각 그룹에 속할 확률이 포함된 테이블을 제공합니다 .

분류 요약 판별 스코어를 요약한 보고서를 제공합니다 . 이 보고서는 판별 분석 플랫폼 보고서의 " 스코어 요약 " 보고서에 해당합니다 .

판별 스코어 각 문서 및 기타 지원 정보의 예측 분류 테이블을 제공합니다 . 이 테이블은 판별 분석 플랫폼 보고서의 " 판별 스코어 " 테이블에 해당합니다 .

JMP PRO 판별 분석 보고서 옵션

텍스트 탐색기 플랫폼에서 " 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 최대 확률 분류 반응이 포함된 열을 데이터 테이블에 저장합니다 . 최대 확률 분류 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 사용**의 에서 확인하십시오 .

확률 계산식 저장 최대 확률 분류 반응을 예측하기 위한 계산식 열을 데이터 테이블에 저장합니다 . 첫번째로 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 스코어 저장 각 관측값에 대해 정준 공간의 스코어가 포함된 열을 데이터 테이블에 저장합니다 정준 공간은 그룹을 가장 많이 나누는 공간입니다 . *k* 번째 정준 스코어의 열에는 **정준 <k>** 라는 이름이 지정됩니다 .

제거 텍스트 탐색기 보고서 창에서 판별 분석 보고서를 제거합니다 .

JMP PRO 용어 선택

텍스트 탐색기 플랫폼에서 용어 선택은 서로 다른 반응을 가장 잘 설명하는 용어를 식별합니다. 이 분석은 일반화 회귀 플랫폼을 사용하여 DTM(문서 용어 행렬)에서 변수를 선택하고 반응에 가장 큰 영향을 미치는 용어를 식별합니다. 용어 선택은 감정 분석 및 다른 유형의 응답과 유사한 이항 반응과 함께 사용할 수 있습니다. 적합 모형은 지정된 반응 열에 대해 적절한 반응 분포를 사용합니다.

팁 : 용어 선택의 예를 보려면 **도움말 > 샘플 데이터 폴더**를 선택하고 Chips.jmp 를 연 다음 "Text Explorer - Term Selection" 테이블 스크립트를 실행하십시오 .

JMP PRO 용어 선택 설정

"용어 선택" 보고서의 "설정" 섹션을 사용하면 반응 열을 선택하고, 반응의 목표 수준을 지정하고, 모형 설정을 조정할 수 있습니다. 모형 설정을 지정했으면 "실행" 버튼을 클릭하여 모형을 실행합니다. 그러면 요약 보고서에 적합 모형이 나타납니다. 자세한 내용은 "용어 선택 요약 보고서"에서 확인하십시오.

목표 수준

반응 열을 선택하면 목표 수준 개요가 나타납니다.

- 명목형 반응의 경우 로지스틱 회귀 모형에서 목표 수준이 될 반응의 한 수준을 선택합니다 . 로지스틱 회귀 모형의 반응은 목표 수준과 다른 모든 수준 간의 결합입니다 .
- 순서형 반응의 경우 처음에는 모든 반응 수준이 모형에 포함됩니다 . 로컬 데이터 필터를 사용하면 모형에서 제외할 반응 수준을 선택할 수 있습니다 . 포함된 수준의 기본 숫자 값은 정규 반응 분포로 모델링됩니다 .

  참고 : 순서형 반응의 경우 용어 선택 모형은 반응 열의 데이터 유형이 숫자인 경우에만 적합시킬 수 있습니다 .

- 연속형 반응의 경우 로컬 데이터 필터 히스토그램을 사용하여 모형에서 제외할 반응 값을 선택합니다 . 포함된 값은 정규 반응 분포로 모델링됩니다 .
- 다중 반응 모델링 유형이 있는 반응 열의 경우 이항 로지스틱 회귀 모형에서 목표 수준이 될 반응 수준을 하나 이상 선택합니다 . 두 개 이상의 수준을 선택하는 경우 문서의 반응 열에 선택한 수준이 하나라도 있으면 해당 문서는 목표 수준에 속합니다 . 선택한 모든 수준이 문서의 반응 열에 있어야 해당 문서가 목표 수준에 포함되게 하려면 **AND 와 결합** 옵션을 선택합니다 .

모형 설정

기본적으로 일반화 회귀 모델은 조기 중지가 있는 Elastic Net 추정 방법과 AICc 검증 방법을 사용합니다. 모형 설정 개요에서 이러한 설정을 변경할 수 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오.

참고 : 텍스트 탐색기 시작 창에 검증 열이 지정된 경우 용어 선택 보고서의 일반화 회귀 플랫폼은 이 검증 열을 검증 방법으로 사용합니다 .

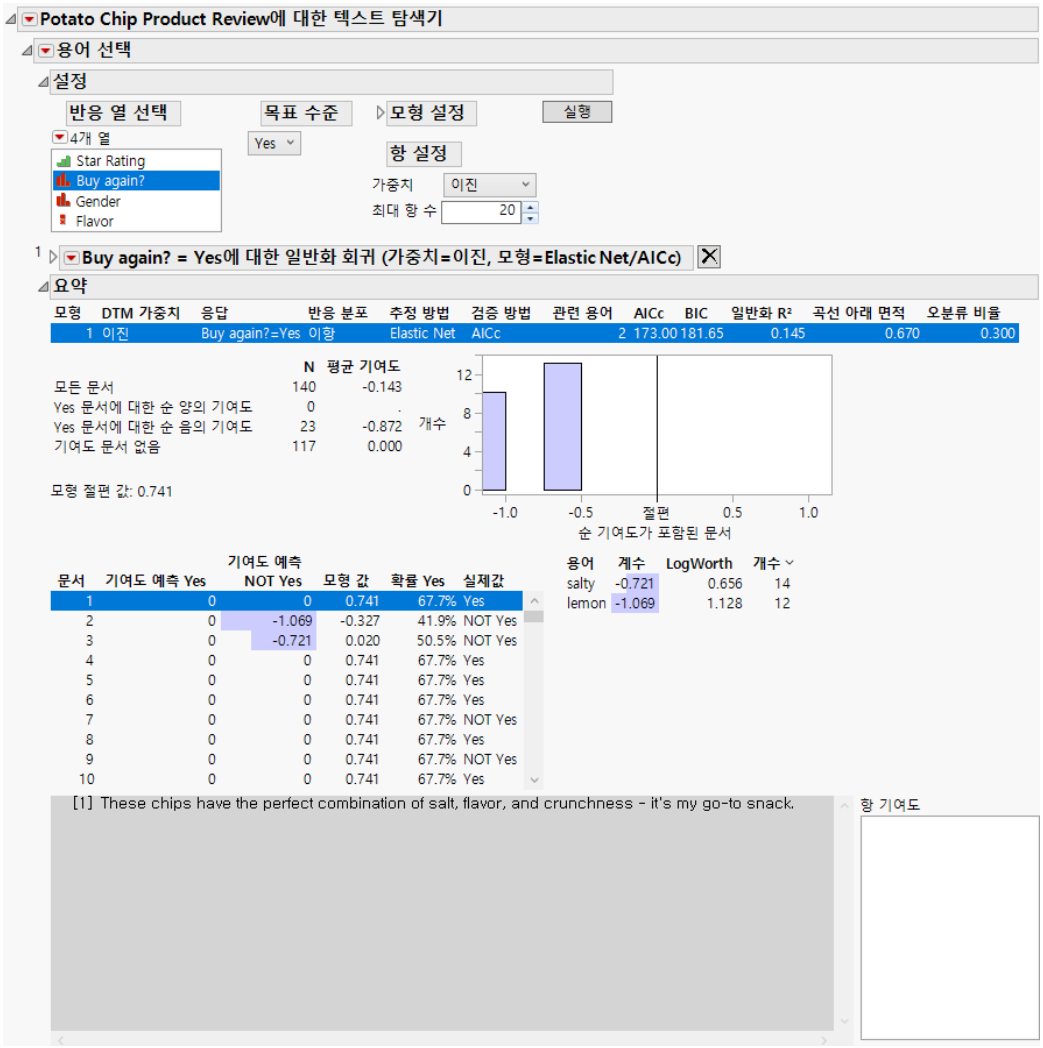
항 설정

항 설정은 회귀 모형에 사용되는 DTM(문서 용어 행렬)을 정의합니다. 가중치 기법과 DTM에 포함된 최대 항 수를 변경할 수 있습니다. 각 항은 DTM의 열에 해당합니다. 말뭉치에 10개 미만의 항목이 있는 항은 모형에서 사용하는 DTM에 포함되지 않습니다. DTM 옵션에 대한 자세한 내용은 "문서 용어 행렬 규격 창"에서 확인하십시오.

JMP PRO 용어 선택 보고서

텍스트 탐색기 플랫폼에서 분석을 실행한 후 "용어 선택" 보고서는 세 개의 섹션으로 구성됩니다. 설정 보고서에는 분석을 지정하기 위한 컨트롤이 포함되어 있습니다. 자세한 내용은 "용어 선택 설정"에서 확인하십시오. 설정 보고서 아래에는 실행한 각 분석에 대해 초기에 닫힌 일반화 회귀 보고서가 있습니다. 자세한 내용은 **선형 모형 적합**의 에서 확인하십시오. 보고서의 마지막 섹션은 요약 보고서입니다.

그림 12.12 용어 선택 보고서



용어 선택 요약 보고서

요약 보고서에는 모형 비교 테이블, 요약 테이블 및 히스토그램, 문서 스코어 테이블, 용어 스코어 테이블 및 텍스트 상자가 포함됩니다.

모형 비교 테이블에는 각 적합 모형에 대한 행이 포함됩니다. 요약 보고서의 나머지 부분에서는 이 테이블에서 현재 선택된 모형의 결과를 보여 줍니다.

요약 테이블에는 문서의 개수와 평균 스코어가 전체 및 모형의 반응 예측값별로 표시됩니다. 평균 기여도는 문서 스코어 테이블의 기여도 값 평균입니다. 요약 히스토그램에는 문서의 전체 기여도 값 분포가 표시됩니다. 히스토그램은 대화식이므로 막대를 클릭하면 문서 스코어 테이블에서 해당하는 문서가 강조 표시됩니다.

문서 스코어 테이블에는 각 문서의 양 및 음의 기여도 값과 각 문서의 예측값 및 실제값이 표시됩니다. 이항 반응 모형의 경우 예측값은 문서가 목표 수준에 속할 확률이고 정규 반응 모형의 경우 예측값은 각 문서에 대한 적합 모형의 예측입니다. 테이블의 행을 선택하면 해당 문서의 텍스트가 테이블 아래의 텍스트 상자에 나타납니다.

"용어 스코어" 테이블에는 적합 모형에 의해 선택된 각 용어, 용어의 해당 계수, 용어의 LogWorth 및 말뭉치에서 용어가 나타나는 횟수가 나열됩니다. 테이블의 행을 선택하면 해당 문서의 텍스트가 테이블 아래의 텍스트 상자에 나타납니다.

텍스트 상자에는 문서 스코어 테이블에서 선택된 문서의 텍스트 또는 용어 스코어 테이블에서 선택된 용어의 컨텍스트가 표시됩니다.

JMP PRO 용어 선택 보고서 옵션

텍스트 탐색기 플랫폼에서 "용어 선택"의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다.

문서 스코어 저장 (요약 테이블에서 분석을 선택한 경우에만 사용할 수 있습니다.) 문서 스코어 테이블의 열을 데이터 테이블의 새 열에 저장합니다. 새 열에는 각 문서에 대한 예측값과 함께 양 및 음의 기여도가 포함됩니다.

용어 스코어 DTM 저장 (요약 테이블에서 분석을 선택한 경우에만 사용할 수 있습니다.) 현재 선택한 분석의 각 관련 용어에 대한 데이터 테이블에 열을 저장합니다. 열에는 "용어 선택"의 "항 설정"에 지정된 가중치를 사용하여 각 문서에 대한 용어 스코어가 포함됩니다.

예측 계산식 저장 (요약 테이블에서 분석을 선택한 경우에만 사용할 수 있습니다.) 현재 선택한 분석에 대한 예측 계산식이 포함된 데이터 테이블에 열을 저장합니다.

용어 클라우드 표시 요약 보고서에서 단어 클라우드를 표시하거나 숨깁니다. 단어 클라우드에는 현재 선택된 분석의 계수 항이 표시됩니다. 단어는 해당 계수의 절대값에 따라 크기가 지정되고 계수의 부호에 따라 색상이 지정됩니다.

제거 텍스트 탐색기 보고서 창에서 용어 선택 보고서를 제거합니다.

JMP PRO 감정 분석

텍스트 탐색기 플랫폼의 감정 분석에서는 어휘 분석을 사용하여 문서에서 감정 용어를 식별하고 긍정적, 부정적 및 전반적 감정에 대해 문서의 스코어를 매깁니다. 이 분석에서는 각 문서가 단일 주제에 대한 이항 감정이 있는 프리 텍스트라고 가정합니다. 감정 분석에서는 기본적인 NLP(자연어 처리)를 결과에 통합합니다. 자연어 처리에 대한 자세한 내용은 https://opennlp.apache.org/에서 확인하십시오. NLP를 사용하지 않으려면 "문서 파싱" 옵션을 선택 취소하십시오.

팁 : 감정 분석의 예를 보려면 **도움말 > 샘플 데이터 폴더**를 선택하고 Chips.jmp 를 연 다음 "Text Explorer - Sentiment Analysis" 테이블 스크립트를 실행하십시오.

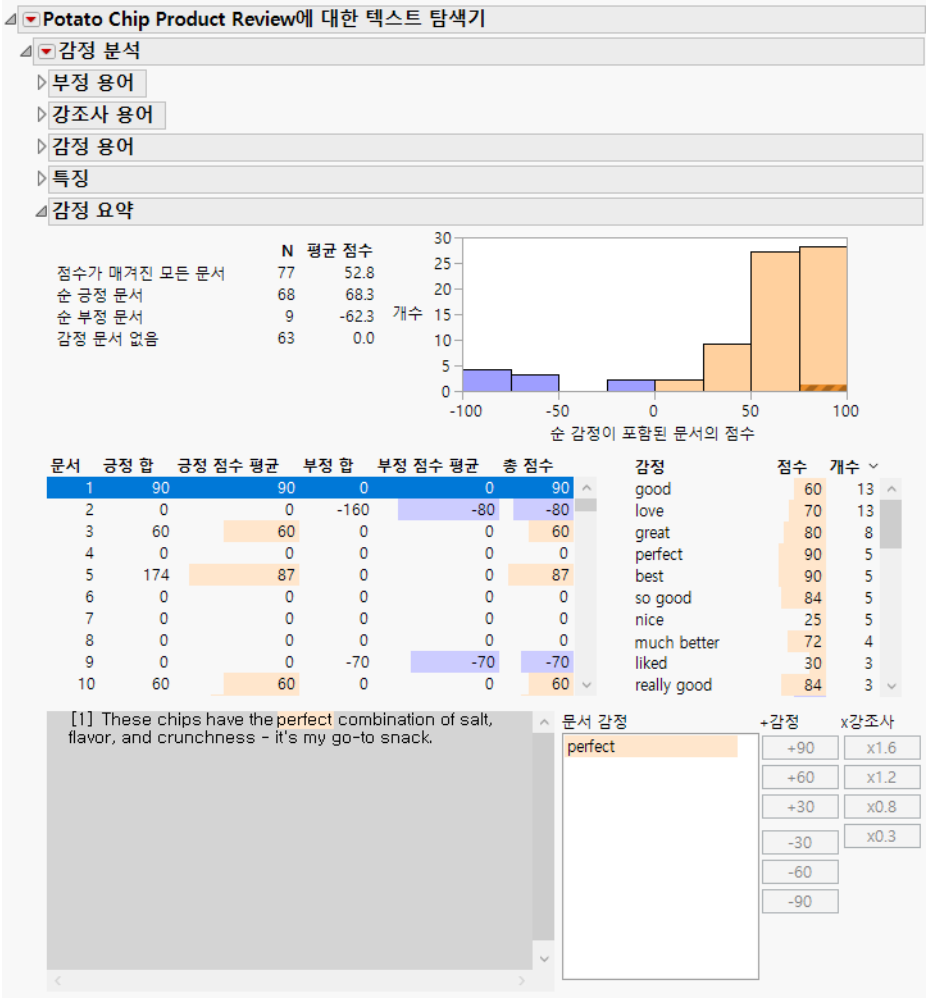
참고 :

- 감정 분석에서 단어는 부정 , 강조사 또는 감정 중 한 용어 클래스에만 적용될 수 있습니다 .
- 감정 분석은 일부 이모티콘 또는 단일 단위로 취급되는 일련의 문자를 인식합니다 . " 감정 용어 " 보고서 또는 " 감정 용어 관리 " 창에서 기본 제공 이모티콘과 해당 기본 감정 스코어를 확인할 수 있습니다 .
- 단어를 부정 , 강조사 또는 감정 용어로 지정할 때 해당 단어가 이미 중지 단어로 지정된 경우 감정 분석 보고서가 열려 있는 동안 해당 단어는 일시적으로 중지 단어에서 제거됩니다 . 이 임시 제거는 전체 텍스트 탐색기 보고서에 영향을 미칩니다 . 감정 분석 보고서가 닫히면 중지 단어로 복원됩니다 .

JMP PRO 감정 분석 보고서

기본적으로 텍스트 탐색기 플랫폼의 감정 분석 보고서에는 " 감정 요약 " 이라는 보고서 하나가 열린 상태로 포함되어 있습니다 . 다른 보고서는 처음에는 닫혀 있습니다 .

그림 12.13 감정 분석 보고서



감정 분석 보고서에는 다음 보고서가 포함됩니다 .

부정 용어

현재 감정 분석의 부정 용어 목록을 포함합니다. 추가 옵션 메뉴를 보려면 목록을 마우스 오른쪽 버튼으로 클릭하십시오. 목록에서 용어를 선택하여 제거할 수 있습니다.

강조사 용어

강조사 용어와 해당하는 승수 값의 목록을 포함합니다. 추가 옵션 메뉴를 보려면 목록을 마우스 오른쪽 버튼으로 클릭하십시오. 목록에서 용어를 선택하여 제거할 수 있습니다.

감정 용어

감정 용어와 해당하는 스코어 값의 목록을 포함합니다. 이 보고서를 사용하여 새 감정 용어를 추가할 수도 있습니다. 가능한 감정 테이블에는 감정 용어로 추가할 수 있는 용어의 개수가 포함됩니다. 용어를 감정 용어로 추가하려면 "가능한 감정" 테이블에서 해당 용어를 선택하고 "+감정" 아래의 버튼 중 하나를 클릭하십시오. 나열되지 않은 감정 스코어 값을 선택하려면 감정 용어 목록에 스코어 값을 추가한 후 편집할 수 있습니다.

"가능한 감정" 테이블에서 용어를 선택하면 해당 용어가 포함된 문서가 감정 용어 보고서의 오른쪽에 나타납니다. 이것은 말뭉치에서 용어가 사용되는 방식에 대한 컨텍스트를 제공합니다.

특징

말뭉치의 특징을 스코어링하는 옵션이 포함되어 있습니다. 특징은 감정 용어로 설명되는 것입니다. 가능한 특징 용어의 목록을 생성하려면 **검색** 버튼을 클릭하십시오. "가능한 특징" 테이블에서 하나 이상의 용어를 선택하면 해당 용어가 포함된 문서의 발췌문이 테이블 오른쪽의 텍스트 상자에 나타납니다. **선택한 특징 스코어링** 버튼을 클릭하여 선택한 특징 용어의 스코어링 결과를 표시하도록 감정 요약 보고서를 업데이트하십시오.

참고 : " 문서 파싱 " 옵션을 선택하면 단어가 한 감정과 동일한 우산 절 내에서 발생할 때 특징 보고서가 단어를 스코어링합니다 .

감정 요약

현재 설정을 기반으로 한 감정 분석의 결과를 포함합니다. 이 보고서에는 요약 테이블 및 히스토그램, 문서 스코어 테이블, 감정 용어 테이블, 텍스트 상자 및 더 많은 감정 용어와 강조사 용어를 추가할 수 있는 제어판이 포함되어 있습니다.

요약 테이블에는 문서를 스코어링한 방법에 따라 세분화된 문서의 개수와 평균 스코어가 표시됩니다. 평균 스코어는 스코어링 옵션의 설정에 따라 결정됩니다. 자세한 내용은 "감정 분석 보고서 옵션"에서 확인하십시오. 요약 히스토그램에는 문서의 전체 감정 분포가 표시됩니다. 히스토그램은 대화식이므로 막대를 클릭하면 문서 스코어 테이블에서 해당하는 문서가 강조 표시됩니다.

문서 스코어 테이블에는 각 문서에 대한 전체 감정 스코어뿐만 아니라 긍정 및 부정 감정 스코어 합계와 평균이 표시됩니다. 테이블의 행을 선택하면 해당 문서의 텍스트가 테이블 아래의 텍스트 상자에 나타납니다. "열 스코어링"을 지정하면 테이블에 스코어링 열의 값이 포함됩니다.

팁: 문서 스코어 테이블의 셀을 마우스로 가리켜 테이블 결과를 생성하는 데 사용된 스코어링 계산을 확인할 수 있습니다 .

감정 용어 테이블에는 각 감정 용어, 해당 스코어 값 및 말뭉치에서 해당 용어가 발생하는 횟수가 나열됩니다.

팁: 여러 단어가 포함된 감정 용어의 경우 스코어 열의 셀을 마우스로 가리키면 스코어를 생성하는 데 사용된 계산을 확인할 수 있습니다.

텍스트 상자에는 문서 스코어 테이블에서 선택된 문서의 텍스트 또는 감정 용어 테이블에서 선택된 용어의 컨텍스트가 표시됩니다. 문서 스코어 테이블에서 문서를 선택하면 해당 문서의 감정 목록이 텍스트 상자 오른쪽에 나타납니다.

팁 : 부정 , 강조사 또는 감정 용어로 분류된 텍스트 상자의 용어를 마우스로 가리키면 분류를 표시하고 제거 버튼이 포함된 상자가 나타납니다 . 부정 , 강조사 또는 감정 용어 목록에서 해당 용어를 빠르게 제거하려면 **제거** 버튼을 클릭하십시오 .

제어판은 텍스트 상자에서 용어를 선택하면 활성화됩니다. 용어를 감정 용어로 추가하려면 텍스트 상자에서 해당 용어를 선택하고 "+감정" 아래의 버튼 중 하나를 클릭하십시오. 용어를 강조사 용어로 추가하려면 텍스트 상자에서 해당 용어를 선택하고 "×강조사" 아래의 버튼 중 하나를 클릭하십시오.

JMP PRO 감정 분석 보고서 옵션

텍스트 탐색기 플랫폼에서 " 감정 분석 " 의 빨간색 삼각형 메뉴에는 다음 옵션이 포함되어 있습니다 .

스코어링 문서의 총 스코어를 계산하기 위한 다음 옵션이 포함되어 있습니다 .

척도화 긍정적 구와 부정적 구의 스코어가 합산됩니다 . 그런 다음 합계를 문서의 구 수로 나누어 총 스코어를 결정합니다 .

최소값 최대값 총 스코어는 긍정 스코어 최대값과 부정 스코어 최소값의 합으로 계산됩니다 .

열 스코어링 계산된 감정과 비교할 수 있는 알려진 정보가 포함된 데이터 테이블 열을 지정합니다 . 스코어 열이 문서 스코어 테이블에 추가됩니다 .

팁 : 총 스코어 열과 스코어 열을 시각적으로 비교하여 감정 스코어링을 평가할 수 있습니다 .

문서 파싱 문서 구문 분석에 NLP(자연어 처리) 가 사용되는지 여부를 지정합니다 . 자연어 처리에 대한 자세한 내용은 https://opennlp.apache.org/ 에서 확인하십시오 .

문서 스코어 저장 문서 스코어 테이블의 열을 데이터 테이블의 새 열에 저장합니다 . 새 열에는 각 문서에 대한 전체 감정 스코어뿐만 아니라 양 및 음의 감정 스코어 합계와 평균이 포함됩니다 .

문서별 감정 스코어 계산 저장 각 감정 용어에 대한 열을 데이터 테이블에 저장합니다 . 각 열에는 각 문서에서 각 감정 용어가 발생한 횟수가 포함됩니다 .

부정 용어 표시 부정 용어 보고서를 표시하거나 숨깁니다 .

강조사 용어 표시 강조사 용어 보고서를 표시하거나 숨깁니다.

감정 용어 표시 감정 용어 보고서를 표시하거나 숨깁니다.

특징 찾기 표시 특징 보고서를 표시하거나 숨깁니다.

감정 클라우드 표시 감정 요약 보고서에서 감정 용어의 단어 클라우드를 표시하거나 숨깁니다.

기본 제공 부정 용어 포함 감정 분석에 사용된 부정 용어에 기본 제공 부정 용어가 포함되는지 여부를 지정합니다.

기본 제공 강조사 용어 포함 감정 분석에 사용된 강조사 용어에 기본 제공 강조사 용어가 포함되는지 여부를 지정합니다.

기본 제공 감정 용어 포함 감정 분석에 사용된 감정 용어에 기본 제공 감정 용어가 포함되는지 여부를 지정합니다.

부정 용어 관리 부정 용어를 추가하거나 제거할 수 있는 창을 표시합니다. 변경 사항은 사용자, 열 및 지역 수준에서 적용될 수 있습니다. 다른 수준에 지정된 부정 용어를 제외하는 지역 예외를 지정할 수도 있습니다. 자세한 내용은 "용어 옵션 관리 창"에서 확인하십시오.

강조사 용어 관리 강조사 용어를 추가하거나 제거할 수 있는 창을 표시합니다. 변경 사항은 사용자, 열 및 지역 수준에서 적용될 수 있습니다. 다른 수준에 지정된 강조사 용어를 제외하는 지역 예외를 지정할 수도 있습니다. 자세한 내용은 "용어 옵션 관리 창"에서 확인하십시오.

감정 용어 관리 감정 용어를 추가하거나 제거할 수 있는 창을 표시합니다. 변경 사항은 사용자, 열 및 지역 수준에서 적용될 수 있습니다. 다른 수준에 지정된 감정 용어를 제외하는 지역 예외를 지정할 수도 있습니다. 자세한 내용은 "용어 옵션 관리 창"에서 확인하십시오.

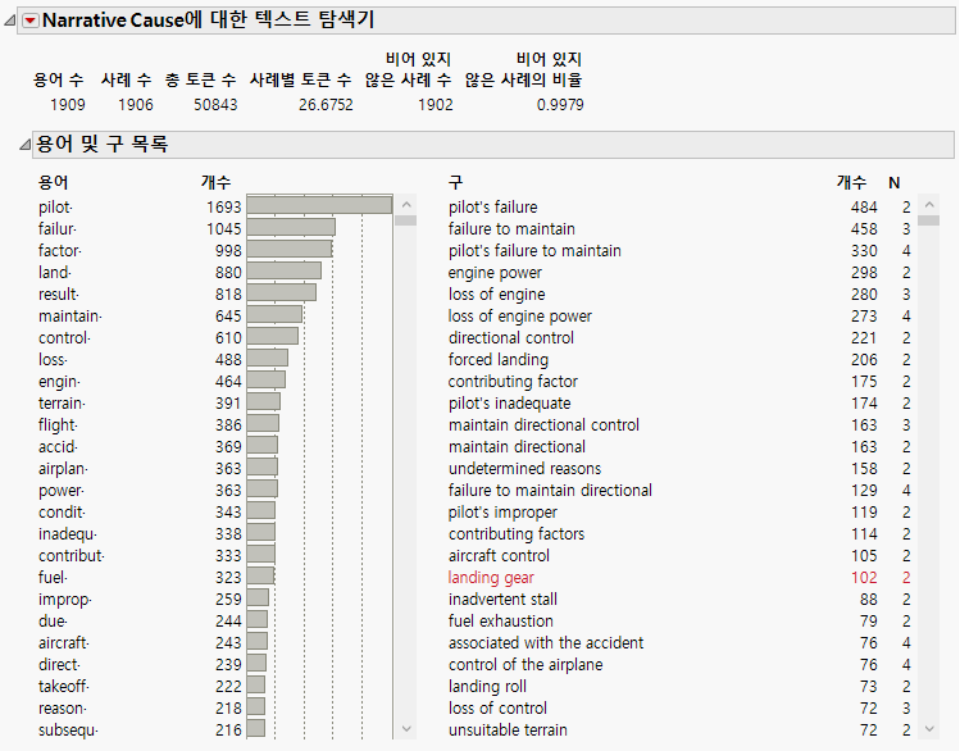
JMP PRO 텍스트 탐색기 플랫폼의 추가 예

이 예에서는 2001 년에 미국에서 발생한 사고에 대한 미국 연방 교통 안전 위원회의 비행기 사고 보고서를 조사합니다. 조사 결과에 대한 설명이 포함된 텍스트를 탐색해서 각 사고의 원인을 밝히려고 합니다. 또한 사고 보고서 모음에서 테마를 찾으려고 합니다.

1. **도움말 > 샘플 데이터 폴더**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" 목록에서 **모든 용어의 어간 추출**을 선택합니다.
8. "토큰화" 목록에서 **기본 단어**를 선택합니다.
9. **확인**을 클릭합니다.

그림 12.14 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 .

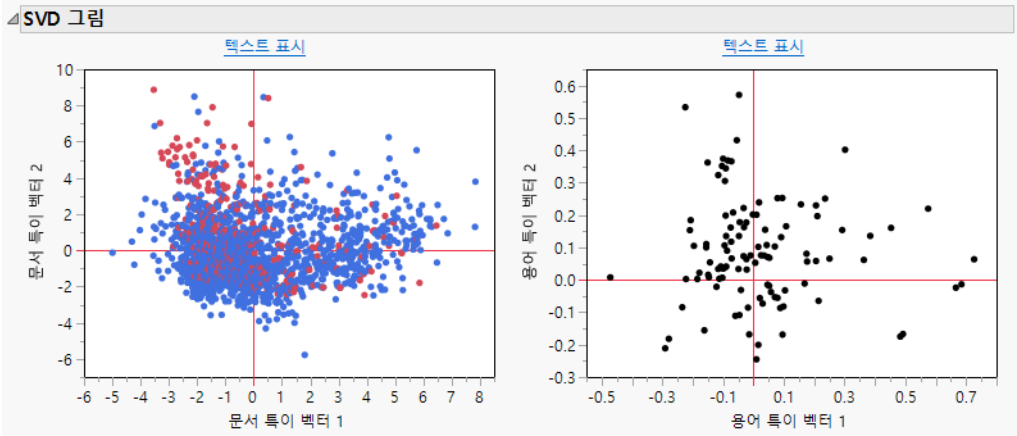
    이 분석은 주제 분석을 위한 첫 번째 분석 단계로서 , SVD 회전을 수행합니다 .

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 12.15 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 5 를 입력합니다 .
17. JMP PRO **확인**을 클릭합니다 .

그림 12.16 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
		continu·	0.31729	result·	-0.2845	undetermin·	-0.3101	inadequ·	0.3353
		maneuv·	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

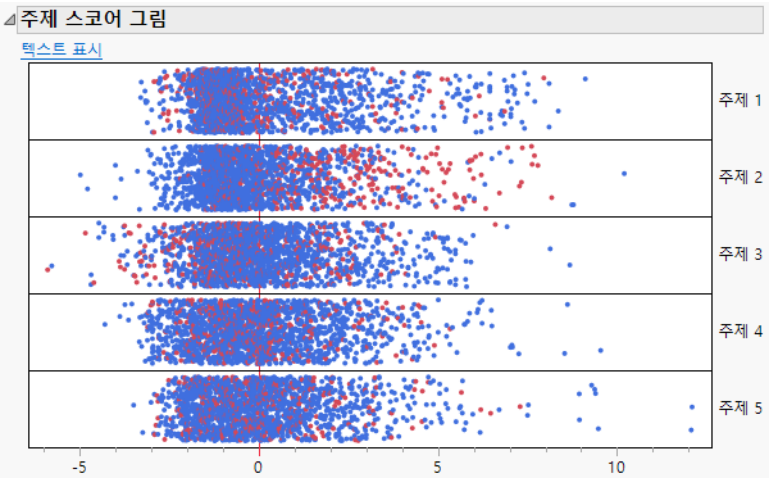
적재량이 가장 높은 각 주제에 대한 용어를 사용하면 해당 주제가 사고 보고서의 테마를 포착하는지 여부를 해석할 수 있습니다 .

예를 들어 주제 1 은 "power", "loss" 및 "engine" 에 대한 적재량이 높으므로 사고 원인으로 엔진 동력 손실이라는 테마를 나타냅니다 . 이는 전체 사고 보고서에서 273 회 나타나는 " 엔진 동력 손실 " 이라는 구와 일치합니다 .

주제 2 에서 적재량이 높은 단어를 살펴보면 , 사고가 어둠이나 저고도와 관련이 있다고 설명할 수 있습니다 .

18. JMP PRO " 주제 스코어 그림 " 옆의 회색 표시 아이콘을 클릭합니다 .

그림 12.17 Narrative Cause 에 대한 주제 스코어 그림



각 주제 스코어 그림에는 말뭉치의 각 문서에 대한 점이 포함됩니다 . 이 그림에서 점을 선택하여 특정 문서의 텍스트를 더 자세히 조사할 수 있습니다 .

주제 2 의 주제를 더 자세히 탐색하려고 합니다 .

19. JMP PRO " 주제 2" 그림에서 맨 오른쪽에 있는 점 세 개를 선택하고 그래프 왼쪽 위의 **텍스트 표시** 버튼을 클릭합니다 .

주제 2 에서 스코어가 가장 높은 세 문서의 텍스트가 새 창에 나타납니다 . 이를 통해 주제 2 가 저고도와 관련 있음을 확인할 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 SVD 의 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.

A 부록

참조 자료

The following sources are referenced in 기본 분석 .

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.

Bablok, W., Passing, H., Bender, R., and Schneider, B. (1988). "A General Regression Procedure for Method Transformation." *Journal of Clinical Chemistry Clinical Biochemistry* 26:783–90.

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.

Boos, D. D., and Duan, S. (2021). "Pairwise Comparisons Using Ranks in the One-Way Model."*American Statistician* 75:414–423.

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.

Howell, D. C. (2013). *Statistical Methods for Psychology.* 8th ed. Belmont, California: Wadsworth.

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* 38:307.

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. (1996). *Applied Linear Statistical Models*. 4th 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.

Passing, H., Bablok, W. (1983). “A New Biometrical Procedure for Testing the Equality of Measurements from Two Different Analytical Methods”. *Journal of Clinical Chemistry Clinical Biochemistry,* 21:709–20.

Passing, H., Bablok, W. (1984). “Comparison of Several Regression Procedures for Method Comparison Studies and Determination of Sample Sizes”. *Journal of Clinical Chemistry Clinical Biochemistry,* 22:431–45.

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. (2020a). “Introduction to Nonparametric Analysis.” In *SAS/STAT 15.2 User’s Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/intronpar.pdf.

SAS Institute Inc. (2020b). “The FREQ Procedure.” In *SAS/STAT 15.2 User’s Guide*. Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/api/docsets/statug/15.2/content/freq.pdf.

Schuirmann, D. J. (1987). “A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability.” *Journal of Pharmacokinetics and Biopharmaceutics* 15:657–680.

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.

Stephens, M. A. (1974). “EDF Statistics for Goodness of Fit and Some Comparisons.” *Journal of the American Statistical Association* 69:730–737.

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.

B 부록

기술 라이선스 고지 사항

- Scintilla is 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.
- Progressë® 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 is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is 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 is 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.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 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:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. 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.

3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. 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.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- 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 is 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 is 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 are 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.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  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. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``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 AUTHORS 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.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  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.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  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”, 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.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  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”, 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.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  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 OF THIRD PARTY RIGHTS. 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.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

- On the Windows operating system, JMP utilizes the OpenBLAS library. OpenBLAS is licensed under the 3-clause BSD license. Full license text follows: Copyright © 2011-2015, The OpenBLAS Project

  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 OpenBLAS project 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 OPENBLAS PROJECT 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.