IBM SPSS Statistics Base 32

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 267 페이지의 『주의사항』에 있는 정보를 확인하십시오.

제품 정보

이 개정판은 새 개정판에 별도로 명시하지 않는 한 버전 32, 릴리스 0, IBM® SPSS® 통계 수정 0 및 모든 후속 릴리스와 수정에 적용됩니다.

© Copyright International Business Machines Corporation .

목차

제 1 장 핵심 기능 1

거듭제곱 분석 1
- 평균 2
- 비율 13
- 상관관계 22
- 회귀 29
- 검증력 분석: 정밀도 31
- 거듭제곱 분석: 격자 값 31

메타분석 32
- 메타분석 연속형 32
- 메타분석 연속형 효과 크기 40
- 메타분석 이분형 47
- 메타분석 이분형 효과 크기 55
- 메타분석 회귀분석 62
- 메타분석 도표화 옵션 66

코드북 72
- 코드북 결과 탭 73
- 코드북 통계 탭 74

빈도 74
- 빈도분석 통계 75
- 빈도분석 도표 76
- 빈도분석 형식 76

기술통계 77
- 기술통계 옵션 77
- DESCRIPTIVES 명령 추가 기능 78

백분위수 78
- 백분위수 결측값 79

탐색 79
- 데이터 탐색 통계 80
- 데이터 탐색 도표 80
- 데이터 탐색: 옵션 81
- EXAMINE 명령 추가 기능 81

교차분석 81
- 교차 분석표 레이어 82
- 교차 분석표 수평배열 막대도표 82
- 테이블 레이어에 레이어 변수를 표시하는 교차 분석표 82
- 교차 분석표 통계 83
- 교차 분석표 셀 표시 84
- 교차 분석표 형식 85

요약 85
- 요약분석 옵션 85
- 사례 요약 통계 86

평균 87
- 그룹별 평균분석: 옵션 87

OLAP 큐브 88
- OLAP 큐브 통계 89
- OLAP 큐브 차이 90
- OLAP 큐브 제목 90

비율 91
- 비율 소개 91
- 일표본 비율 91

대응표본 비율 ... 93
독립표본 비율 ... 96
T 검정 ... 98
T 검정 ... 98
독립 표본 T 검정 ... 99
대응표본 T 검정 ... 100
일표본 T 검정 ... 101
T TEST 명령 추가 기능 ... 102
일원 분산 분석 ... 103
일원 분산 분석 대비 ... 103
일원 분산 분석 사후 검정 ... 104
일원 분산 분석: 옵션 ... 105
ONEWAY 명령 추가 기능 ... 106
GLM 일변량 분석 ... 106
GLM 모형 ... 108
GLM 대비 ... 109
GLM 프로파일 도표 ... 110
GLM 사후분석 비교 ... 112
GLM 저장 ... 114
GLM 추정 주변 평균 ... 114
GLM 옵션 ... 115
UNIANOVA 명령 추가 기능 ... 116
블랜드 알트만 ... 116
블랜드 알트만: 기준 ... 117
블랜드 알트만: 인쇄 ... 117
블랜드 알트만: 줄거리 ... 117
카이제곱 검정 ... 117
Chi 스퀘어 테스트: 정확 ... 118
제곱 테스트: 셀 ... 118
카이제곱 테스트: 형식 ... 118
이변량 상관계수 ... 118
이변량 상관: 옵션 ... 120
이변량 상관계수 신뢰구간 ... 120
CORRELATIONS 및 NONPAR CORR 명령 추가 기능 ... 120
근접성 맵핑 ... 121
근접 매핑: 데이터 ... 122
근접 매핑: 가중치 ... 123
근접 매핑: 초기 ... 123
근접 매핑: 모델 ... 124
근접 매핑: 제한 사항 ... 125
근접 매핑: 보조 ... 126
근접 매핑: 기준 ... 126
근접 매핑: 출력 ... 127
근접 매핑: 저장 ... 132
편상관계수 ... 132
편상관: 옵션 ... 133
PARTIAL CORR 명령 추가 기능 ... 133
거리 ... 134
거리 측도: 상이성 측도 ... 134
거리 측도: 유사성 측도 ... 134
PROXIMITIES 명령 추가 기능 ... 135
거리 상관 ... 135
거리 상관: 기준 ... 136
거리 상관: 인쇄 ... 136
거리 상관관계: 플롯 ... 137
선형모형 ... 137
선형 모형 구하기 ... 137
목표 ... 138

기본 사항 138
모형 선택 139
앙상블 139
고급 139
모형 옵션 140
모형 요약 140
자동 데이터 준비 140
예측자 중요도 140
관찰에 의한 예측 140
잔차 140
이상값 141
효과 141
계수 141
평균 추정 142
모형 작성 요약 142
선형 회귀 142
선형 회귀분석 변수 선택 방법 143
선형 회귀분석 규칙 설정 143
선형 회귀분석 도표 143
선형 회귀: 새로운 변수로 저장 144
선형 회귀분석 통계 145
선형 회귀분석 옵션 145
REGRESSION 명령 추가 기능 146
순서 회귀분석 146
순서 회귀분석: 옵션 147
순서 회귀분석 결과 147
순서 회귀분석 위치 모형 148
순서 회귀분석 척도 모형 148
PLUM 명령 추가 기능 149
선형 탄력적 순 회귀 149
Linear Elastic Net Regression: 옵션 150
선형 라소 회귀분석 152
선형 올가미 회귀분석: 옵션 153
선형 리지 회귀분석 155
선형 능선 회귀: 옵션 155
매개 분석 157
모형 유형 162
종속변수 163
예측자 164
기준 164
인쇄 165
도표 166
시계열 필터 167
시계열 필터: 옵션 168
곡선 추정 169
곡선 추정 모형 169
곡선 추정: 저장 170
편최소제곱회귀 170
모형 171
옵션 171
최근접 이웃 분석 172
이웃 173
변수 174
파티션 174
저장 175
결과 175
옵션 175
모형 보기 176

판별 분석 ... 178
판별 분석 범위지정 ... 179
판별 분석 케이스 선택 ... 179
판별 분석 통계 ... 179
판별 분석 단계적 방법 ... 180
판별 분석 분류 ... 180
판별 분석 저장 ... 181
DISCRIMINANT 명령 추가 기능 ... 181
요인 분석 ... 181
요인 분석 선택변수 ... 182
요인 분석: 기술통계 ... 182
요인 분석: 요인추출 ... 182
요인 분석: 회전 ... 183
요인 분석: 요인 스코어 ... 184
요인 분석: 옵션 ... 184
FACTOR 명령 추가 기능 ... 184
군집화 프로시저 선택 ... 184
이단계 군집 분석 ... 185
이단계 군집 분석 옵션 ... 186
이단계 군집 분석 결과 ... 187
군집 뷰어 ... 187
계층적 군집 분석 ... 191
계층적 군집 분석 ... 192
계층적 군집 분석: 통계 ... 192
계층적 군집 분석: 도표 ... 193
계층적 군집 분석: 새로운 변수로 저장 ... 193
CLUSTER 명령 구문 추가 기능 ... 193
K-Means 군집 분석 ... 193
K-Means 군집 분석 효율성 ... 194
K-Means 군집 분석: 반복 ... 194
K-Means 군집 분석: 저장 ... 194
K-Means 군집 분석: 옵션 ... 195
QUICK CLUSTER 명령 추가 기능 ... 195
비모수 검정 ... 195
일표본 비모수 검정 ... 195
일표본 Wilcoxon 검정 ... 199
독립 표본 비모수 검정 ... 199
대응표본 비모수 검정 ... 201
NPTESTS 명령 추가 기능 ... 204
레거시 대화 상자 ... 204
다중 응답 분석 ... 213
다중 응답 분석 ... 213
다중 응답 세트 정의 ... 214
다중반응 빈도분석 ... 214
다중반응 교차분석 ... 215
결과 보고 ... 216
결과 보고 ... 216
행별 요약 보고서 ... 216
열별 요약 보고서 ... 218
REPORT 명령 추가 기능 ... 220
신뢰성 분석 ... 220
신뢰도 분석: 통계량 ... 221
RELIABILITY 명령 추가 기능 ... 223
가중 카파 ... 224
가중 카파: 기준 ... 225
가중 카파: 인쇄 ... 225
다차원척도법 ... 225
다차원척도법 데이터의 형태 ... 226

다차원척도법 측도 계산 227
다차원척도법: 모형 227
다차원척도법: 옵션 227
ALSCAL 명령 추가 기능 227
ALSCAL과 PROXMAP 비교 228
비율 통계량 232
비율 통계량: 통계량 233
P-p 도표 234
Q-Q 도표 236
ROC 분석 239
ROC 분석: 옵션 240
ROC 분석: 표시 240
ROC 분석: 집단 정의(문자열) 241
ROC 분석: 집단 정의(숫자) 241
ROC 곡선 241
ROC 곡선: 옵션 242
시뮬레이션 242
모형 파일을 기반으로 시뮬레이션을 계획하는 방법 243
사용자 정의 방정식을 기반으로 시뮬레이션을 계획하는 방법 243
예측 모형 없이 시뮬레이션 계획 244
시뮬레이션 계획에서 시뮬레이션을 실행하는 방법 244
시뮬레이션 작성기 245
시뮬레이션 실행 대화 상자 253
시뮬레이션의 도표 결과로 작업 255
지리 공간 모형화 256
지도 선택 257
데이터 소스 258
지리 공간 연관 규칙 259
시공간적 예측 262
마침 265

주의사항 267
상표 268

색인 269

제 1 장 핵심 기능

IBM SPSS 통계 Base Edition에는 다음 핵심 기능이 포함되어 있습니다.

거듭제곱 분석

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

IBM SPSS 통계 에서는 다음과 같은 거듭제곱 분석 프로시저를 제공합니다.

일표본 T 검정
일표본 분석에서 관측 데이터는 단일 변량 표본으로 수집됩니다. 표본 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 평균 모수에 관한 통계적 추론을 그리는 것으로 가정합니다.

대응표본 T 검정
대응표본 분석에서 관측 데이터에는 두 개의 대응표본과 상관분석 표본이 포함되어 있으며 각 케이스에는 두 개의 측정이 있습니다. 각 표본의 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 두 평균의 차이에 관한 통계적 추론을 그리는 것으로 가정합니다.

독립표본 T 검정
독립표본 분석에서 관측 데이터에는 두 개의 독립표본이 포함되어 있습니다. 각 표본의 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 두 평균의 차이에 관한 통계적 추론을 그리는 것으로 가정합니다.

일원 분산 분석
분산분석(ANOVA)은 종종 정규 분포로 간주되는 여러 모집단의 평균을 추정하는 통계 방법입니다. 분산분석의 일반 유형인 일원배치 분산분석은 두 표본 *t* 검정의 확장입니다.

예. 정확한 대립가설을 발견하는 데 사용하는 가설검정의 거듭제곱은 검정에서 검정 가설을 거부할 확률입니다. 제 II 유형 오차 확률이 대립가설이 참일 때 영가설을 허용할 확률이므로, 거듭제곱은 대립가설이 참일 때 영가설을 거부할 확률인 (1-제 II 유형 오차 확률)로 표시할 수 있습니다.

통계 및 도표. 단측 검정, 양측 검정, 유의수준, 제 1 유형 오차율, 검정 가정, 모집단 표준편차, 검정 중인 모집단 평균, 가설값, 표본크기별 2차원 거듭제곱, 효과 크기별 2차원 거듭제곱, 표본크기별 3차원 거듭제곱, 효과 크기별 3차원 거듭제곱, 요인회전 차수, 집단 대응, Pearson 곱 적률 상관계수, 평균차이, 효과 크기의 도표 범위, 통합 모집단 표준편차, 대비 및 대응별 차이, 대비계수, 대비검정, BONFERRONI, SIDAK, LSD, 전체 표본크기별 거듭제곱, 통합 표준편차별 2차원 거듭제곱, 전체 표본별 3차원 거듭제곱, 전체 표본크기별 3차원 거듭제곱, 통합 표준편차, 스튜던트 T 분포, 비중앙 T-분포,

거듭제곱 분석 데이터 고려사항

데이터
일표본 분석에서 관측 데이터는 단일 변량 표본으로 수집됩니다.

대응표본 분석에서 관측 데이터에는 두 개의 대응표본과 상관분석 표본이 포함되어 있으며 각 케이스에는 두 개의 측정이 있습니다.

독립표본 분석에서 관측 데이터에는 두 개의 독립표본이 포함되어 있습니다.

분산분석(ANOVA)은 종종 정규 분포로 간주되는 여러 모집단의 평균을 추정하는 통계 방법입니다.

가정
일표본 분석에서 표본 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 평균 모수에 관한 통계적 추론을 그리는 것으로 가정합니다.

대응표본 분석에서 각 표본의 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 두 평균의 차이에 관한 통계적 추론을 그리는 것으로 가정합니다.

독립표본 분석에서 각 표본의 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 두 평균의 차이에 관한 통계적 추론을 그리는 것으로 가정합니다.

일원배치 분산분석에서 여러 모집단의 평균을 추정하는 통계 방법은 종종 정규 분포되는 것으로 간주됩니다.

거듭제곱 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 거듭제곱 분석 > 평균 비교 > 일표본 T 검정 또는 **대응표본 T 검정** 또는 **독립표본 T 검정** 또는 **일원배치 분산분석**
2. 필수 검정 가정을 정의합니다.
3. **확인**을 클릭하십시오.

평균

IBM SPSS 통계 Base Edition에는 다음과 같은 통계 기능이 포함되어 있습니다.

일표본 T 검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

일표본 분석에서 관측 데이터는 단일 변량 표본으로 수집됩니다. 표본 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 평균 모수에 관한 통계적 추론을 그리는 것으로 가정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 검정력 분석 > 평균 > 1-샘플 T 검정

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』 의 내용을 참조하십시오.
4. **검정력**이 검정 가정 **추정** 메소드로 선택된 경우, 적절한 **샘플 크기**를 입력하여 검정력 추정 값을 입력하십시오. 값은 1보다 큰 정수여야 합니다.
5. 선택적으로 **지정하기** 필드에서 옵션을 선택하십시오.

   가정된 값
   기본 설정은 **모집단 평균** 및 **널값** 설정을 제공합니다.

   모집단 평균
   검정 중인 모집단 평균을 지정하는 값을 입력하십시오. 값은 단일 숫자여야 합니다.

   널값
   선택적으로 검정할 가정된 값을 지정하는 값을 입력하십시오. 값은 단일 숫자여야 합니다.

   효과 크기
   Cohen의 f는 검정력 또는 샘플 크기의 추정에 대한 입력으로서의 효과 크기를 추정하기 위해 사용됩니다. 정의된 효과 크기 **값**은 프로시저의 중간 단계로 전달되고 원하는 검정력 또는 샘플 크기를 계산합니다.
6. **모집단 표준편차** 값을 입력합니다. 값은 0보다 큰 단일 숫자여야 합니다.
7. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

비방향성(양측) 분석
선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

방향성(단측) 분석
선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.

8. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
9. 선택적으로 **플롯**을 클릭하여 3 페이지의 『일표본 T 검정의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정할 수 있습니다.

   참고: 플롯은 **거듭제곱**이 검정 가정으로 선택된 경우에만 사용 가능합니다.
10. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **가정된 값**이 **지정** 목록에서 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

일표본 *T* 검정의 거듭제곱 분석*:* 도표

플롯 대화 상자는 샘플 및 효과 크기 차트별 2차원 및 3차원 거듭제곱을 설명하기 위해 출력되는 플롯을 제어하는 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표

거듭제곱 추정 대 표본크기
이 선택적 설정을 사용으로 설정할 경우 표본크기별 2차원 거듭제곱 도표를 제어하는 옵션이 제공됩니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

표본크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한
표본크기별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

거듭제곱 추정 대 효과크기
기본적으로 이 선택적 설정은 사용 안함으로 설정됩니다. 사용으로 설정할 경우 출력에 도표가 표시됩니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기의 기본 도표 범위가 사용됩니다.

효과크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다.

하한
효과크기별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 -5.0보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
효과크기별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5.0을 초과하지 않아야 합니다.

3차원 도표

거듭제곱 추정 대
표본크기(X축) 및 효과크기(Y축)별 3차원 거듭제곱 도표, 수평/수직 회전 설정, 표본/효과크기의 사용자 지정 도표 범위를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

효과크기(X축) 및 표본크기(Y축)
이 선택적 설정은 표본크기(X축) 및 효과크기(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

효과크기(Y축) 및 표본크기(X축)
이 선택적 설정은 표본크기(Y축) 및 효과크기(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

표본크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한
표본크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

효과크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기의 기본 도표 범위가 사용됩니다.

하한
효과크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 -5.0보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
효과크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5.0을 초과하지 않아야 합니다.

대응표본 T 검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

대응표본 분석에서 관측 데이터에는 두 개의 대응표본과 상관분석 표본이 포함되어 있으며 각 케이스에는 두 개의 측정이 있습니다. 각 표본의 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 두 평균의 차이에 관한 통계적 추론을 그리는 것으로 가정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 거듭제곱 분석 > 평균 > 대응-표본 T 검정

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』 의 내용을 참조하십시오.

4. **거듭제곱**이 검정 가정 **추정** 메소드로 선택된 경우 적절한 **샘플 크기** 값을 입력하십시오. 값은 2보다 크거나 같은 양의 정수여야 합니다.
5. 선택적으로 **지정하기** 필드에서 옵션을 선택하십시오.

   가정된 값
   기본 설정은 **평균** 및 **표준 편차** 설정을 제공합니다.

   모집단 평균 차이
   단일 모집단 평균이 필요한 경우, **모집단 평균차이** 값을 입력합니다. 단일값이 지정된 경우 모집단 평균차이 μ_d를 나타냅니다.

   참고: 샘플 크기가 선택된 경우 값은 0이 될 수 없습니다.

   그룹 1 및 그룹 2에 대한 모집단 평균
   지정된 그룹 쌍에 여러 모집단이 필요한 경우 **그룹 1에 대한 모집단 평균** 및 **그룹 2에 대한 모집단 평균**에 대한 값을 입력하십시오. 여러 값이 지정된 경우 그룹 차이 μ_1 및 μ_2의 모집단 평균 차이를 나타냅니다.

   참고: 샘플 크기가 선택된 경우 두 값은 동일할 수 없습니다.

   평균 차이에 대한 모집단 표준 편차
   단일 모집단 평균이 지정된 경우 평균 차이 값의 모집단 표준 편차를 입력합니다. 단일값이 지정된 경우 값은 집단 차이 σ_d의 모집단 표준편차를 나타냅니다. 값은 0보다 큰 단일 숫자여야 합니다.

   그룹 1 및 그룹 2에 대한 모집단 표준 편차
   여러 모집단 평균이 지정되면 그룹 1과 그룹 2 값에 대한 모집단 표준 편차를 입력합니다. 다중 값이 지정된 경우 집단 차이 σ_1 및 σ_2의 모집단 표준편차를 나타냅니다. 값은 0보다 큰 단일 숫자여야 합니다.

   Pearson 제품-모멘트 상관 계수
   선택적으로 Pearson 제품-모멘트 상관 계수(ρ)를 지정하는 값을 입력하십시오. 값은 -1과 1사이의 단일 숫자 값이어야 합니다. 값은 0이 될 수 없습니다.

   참고: 단일 **평균차이의 모집단 표준편차** 값이 지정된 경우에는 이 설정이 무시됩니다. 그렇지 않으면 **그룹 1 및 그룹 2에 대한 모집단 표준 편차**의 값이 σ_d를 계산하는 데 사용됩니다.

   효과 크기
   거듭제곱 또는 샘플 크기 추정에 대한 입력으로 효과 크기를 추정합니다. 정의된 효과 크기 **값**은 프로시저의 중간 단계로 전달되고 원하는 검정력 또는 샘플 크기를 계산합니다.
6. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

   비방향성(양측) 분석
   선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

   방향성(단측) 분석
   선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.
7. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
8. 선택적으로 **플롯**을 클릭하여 5 페이지의 『대응표본 T 검정의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정하십시오.

   참고: 플롯은 **거듭제곱**이 검정 가정으로 선택된 경우에만 사용 가능합니다.
9. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **가정된 값**이 **지정하기** 목록에서 선택된 경우에만 사용 가능합니다.

대응표본 *T* 검정의 거듭제곱 분석: 도표

플롯 대화 상자는 샘플 및 효과 크기 차트별 2차원 및 3차원 거듭제곱을 설명하기 위해 출력되는 플롯을 제어하는 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표

거듭제곱 추정 대 표본크기
이 선택적 설정을 사용으로 설정할 경우 표본크기별 2차원 거듭제곱 도표를 제어하는 옵션이 제공됩니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

표본크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한
표본크기별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

거듭제곱 추정 대 효과크기
기본적으로 이 선택적 설정은 사용 안함으로 설정됩니다. 사용으로 설정할 경우 출력에 도표가 표시됩니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기의 기본 도표 범위가 사용됩니다.

효과크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다.

하한
효과크기별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 -5.0보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
효과크기별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5.0을 초과하지 않아야 합니다.

3차원 도표

거듭제곱 추정 대
표본크기(X축) 및 효과크기(Y축)별 3차원 거듭제곱 도표, 수평/수직 회전 설정, 표본/효과크기의 사용자 지정 도표 범위를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

효과크기(X축) 및 표본크기(Y축)
이 선택적 설정은 표본크기(X축) 및 효과크기(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

효과크기(Y축) 및 표본크기(X축)
이 선택적 설정은 표본크기(Y축) 및 효과크기(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

표본크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한
표본크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

효과크기 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기의 기본 도표 범위가 사용됩니다.

하한

효과크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 -5.0보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

효과크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5.0을 초과하지 않아야 합니다.

독립표본 T 검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

독립표본 분석에서 관측 데이터에는 두 개의 독립표본이 포함되어 있습니다. 각 표본의 데이터는 고정 평균 및 분산이 있는 정규 분포를 독립적으로 동등하게 따르며 두 평균의 차이에 관한 통계적 추론을 그리는 것으로 가정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 거듭제곱 분석 > 평균 > 독립-표본 T 검정

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』 의 내용을 참조하십시오.

   샘플 크기의 비율을 지정하기 위한 **그룹 크기 비율** 값을 입력하십시오. 값은 0.01-100 사이의 단일 값이어야 합니다.
4. **거듭제곱**이 검정 가정 **추정** 메소드로 선택된 경우, 비교를 위해 **그룹 1 및 그룹 2의 샘플 크기**의 샘플 크기를 지정하는 값을 입력하십시오. 값은 1보다 큰 정수여야 합니다.
5. 선택적으로 **지정하기** 필드에서 옵션을 선택하십시오.

   가정된 값

   기본 설정은 **모집단 평균** 및 **모집단 표준 편차** 설정을 제공합니다.

   모집단 평균 차이

   단일 모집단 평균이 지정된 경우, 모집단 평균 차이 값을 입력합니다. 단일 값이 지정되면 그룹 평균 차이 σ_d의 모집단이 표시됩니다. 값은 0보다 큰 단일 숫자여야 합니다.

   그룹 1 및 그룹 2에 대한 모집단 평균

   여러 모집단 평균이 지정되면 그룹 1과 그룹 2값에 대한 모집단 평균을 입력합니다. 여러 값이 지정된 경우 그룹 차이 σ_1 및 σ_2의 모집단 평균을 나타냅니다. 값은 0보다 큰 단일 숫자여야 합니다.

   모집단 표준편차:

   모집단 표준편차가 **두 집단에 대해 등등함**인지 아니면 **두 집단에 대해 동등하지 않음**인지 지정합니다.

   - 집단 표준 편차가 두 그룹에 대해 동일한 경우 σ를 나타내는 **풀링된 표준 편차**의 값을 입력하고 두 개의 그룹 분산이 같거나 $\sigma_1 = \sigma_2 = \sigma$인 것으로 가정합니다.
   - 두 그룹의 모집단 표준 편차가 같지 않으면 σ_1 및 σ_2를 나타내는 **그룹 1 및 그룹 2**의 표준 편차 값을 입력합니다.

참고: 그룹 1 및 그룹 2의 표준 편차의 값이 동일한 경우 이 값은 단일 값으로 처리됩니다.

효과 크기
거듭제곱 또는 샘플 크기 추정에 대한 입력으로 효과 크기를 추정합니다. 정의된 효과 크기 **값**은 프로시저의 중간 단계로 전달되고 원하는 검정력 또는 샘플 크기를 계산합니다.

6. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

비방향성(양측) 분석
선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

방향성(단측) 분석
선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.

7. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
8. 선택적으로 **플롯**을 클릭하여 8 페이지의 『독립표본 T 검정의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정하십시오.

참고: 플롯은 **거듭제곱**이 검정 가정으로 선택된 경우에만 사용 가능합니다.

9. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **가정된 값**이 **지정** 목록에서 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

독립표본 *T* 검정의 거듭제곱 분석*:* 도표

플롯 대화 상자는 샘플 비율, 효과 크기 또는 평균 차이 차트별로 2 및 3차원 거듭제곱 을 설명하기 위해 출력되는 도표를 제어하기 위한 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표

거듭제곱 추정 대 표본크기 비율/거듭제곱 추정 대 표본크기
이 선택적 설정을 사용으로 설정할 경우 표본크기 비율별 2차원 거듭제곱 도표를 제어하는 옵션이 제공됩니다. 다중 거듭제곱 값이 지정된 경우(**거듭제곱 추정 대 샘플 크기**) 추가 설정을 사용할 수 없습니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

표본크기 비율 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한
표본크기 비율별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 0.01 및 100 범위에 있어야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기 비율별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 0.01 및 100 범위에 있어야 하며 **하한** 값보다 커야 합니다.

거듭제곱 추정 대 효과 크기(또는 평균차이)
기본적으로 이 선택적 설정은 사용 안함으로 설정됩니다. 사용으로 설정할 경우 출력에 도표가 표시됩니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기(또는 평균차이)의 기본 도표 범위가 사용됩니다.

효과크기(또는 평균차이) 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다.

하한
효과크기별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 -5.0보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
효과크기별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5.0을 초과하지 않아야 합니다.

3차원 도표

거듭제곱 추정 대

표본크기 비율(X축) 및 효과크기(Y축)별 3차원 거듭제곱 도표, 수평/수직 회전 설정, 표본/효과크기의 사용자 지정 도표 범위를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

X축의 효과크기(또는 평균차이) 및 Y축의 표본크기

이 선택적 설정은 표본크기 비율(X축) 및 효과크기(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

Y축의 효과크기(또는 평균차이) 및 X축의 표본크기

이 선택적 설정은 표본크기(Y축) 및 효과크기(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

수직 회전

이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전

이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

표본크기 비율 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한

표본크기 비율별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 0.01 및 100 범위에 있어야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

표본크기 비율별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 0.01 및 100 범위에 있어야 하며 **하한** 값보다 커야 합니다.

효과크기(또는 평균차이) 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기의 기본 도표 범위가 사용됩니다.

하한

효과크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 -5.0보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

효과크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

일원배치 분산분석의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

분산분석(ANOVA)은 종종 정규 분포로 간주되는 여러 모집단의 평균을 추정하는 통계 방법입니다. 분산분석의 일반 유형인 일원배치 분산분석은 두 표본 *t* 검정의 확장입니다. 이 프로시저에서는 다중 집단 평균, 전체 검정 및 지정된 대비를 사용하여 검정을 비교하기 위해 두 가지 유형의 가설을 추정하는 접근 방법을 제공합니다. 과잉 검정에서는 모든 집단 평균이 동일하다는 영가설에 중점을 둡니다. 지정된 대비를 사용하는 검정에서는 전체 분산분석 가설을 더 작지만 더 설명적이고 유용한 평균으로 구분합니다.

1. 메뉴에서 다음을 선택합니다.

분석 > 거듭제곱 분석 > 평균 > 단방향 분산 분석

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』의 내용을 참조하십시오.
4. 선택적으로 **지정하기** 필드에서 옵션을 선택하십시오.

   가정된 값
   기본 설정은 **합동 모집단 표준 편차** 설정을 제공합니다. 0보다 큰 단일 숫자 값을 입력하십시오. 기본값은 1입니다.

   효과 크기: Cohen's f
   Cohen의 f는 검정력 또는 샘플 크기의 추정에 대한 입력으로서의 효과 크기를 추정하기 위해 사용됩니다. 정의된 효과 크기 **값**은 프로시저의 중간 단계로 전달되고 원하는 검정력 또는 샘플 크기를 계산합니다.

   효과 크기: Eta-squared
   Eta-squared(^{2})은 거듭제곱 또는 샘플 크기의 추정에 대한 입력으로 효과 크기를 추정하는 데 사용됩니다. 정의된 효과 크기 **값**은 프로시저의 중간 단계로 전달되고 원하는 검정력 또는 샘플 크기를 계산합니다.
5. **거듭제곱**이 검정 가정 **추정** 설정으로 선택된 경우 **그룹 크기** 및 **그룹 평균** 값을 지정하십시오. 두 개 이상의 그룹 크기 값을 지정해야 하며, 각 그룹 크기 값은 2보다 크거나 같아야 합니다(2,147, 483, 647을 초과할 수 없음). 집단 평균 값도 두 개 이상 지정해야 하며, 지정된 값의 개수는 집단 크기 값과 같아야 합니다.

   거듭제곱이 검정 가정 **추정** 설정으로 선택된 경우 **그룹 크기** 및 **그룹 평균** 값을 지정하십시오. 그룹 가중치는 **거듭제곱**이 선택될 때 그룹 크기 가중치를 지정합니다.

   참고: 집단 크기 값이 지정된 경우에는 **집단 가중값**이 무시됩니다.

   추가 **그룹 크기**, **그룹 가중치** 및 **그룹 평균** 값을 포함하려면 **추가**를 클릭하십시오. 기존 값을 제거하려면 **삭제**를 누르십시오.
6. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
7. 선택적으로 **대비**를 클릭하여 10 페이지의 『일원배치 분산분석의 거듭제곱 분석: 대비』 설정(대비 검정, 쌍별 차이, 효과 크기 및 신뢰구간에 따라 샘플 크기 추정)을 지정하십시오.
8. 선택적으로 **플롯**을 클릭하여 11 페이지의 『일원배치 분산분석의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정하십시오.

   참고: 플롯은 **그룹 크기** 값이 지정되고 **거듭제곱**이 **추정** 설정으로 선택된 경우에만 사용 가능합니다.
9. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택된 경우에만 사용 가능합니다.

일원배치 분산분석의 거듭제곱 분석*:* 대비

대비 대화 상자는일원배치 분산분석의 검정력 분석에 대비, 계수, 효과 크기, 쌍별 차이 및 신뢰구간 반폭 설정을 지정하는 옵션을 제공합니다.

대비검정

선형 대비를 사용한 검정
사용으로 설정할 경우, 대비 및 계수 설정을 사용할 수 있습니다.

검정 방향

비방향성(양측) 분석
선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

방향성(단측) 분석
선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.

계수

표를 사용하여 대비 계수를 지정하고 대비검정을 요청하십시오. 표 값은 선택적입니다. 지정된 값의 개수는 **집단 크기** 및 **집단 평균**에 지정된 값과 같아야 합니다. 지정된 모든 값의 합은 0이어야 합니다. 그렇지 않으면 마지막 값이 자동으로 수정됩니다.

효과 크기

δ로 측정되는 대비 테스트의 효과 크기를 지정하십시오. 단일 숫자 값을 지정해야 합니다. **테스트 방향**이 **방향성 (단측) 분석**으로 설정된 경우 지정된 값은 ≥ 0이어야 합니다. **검정력**이 **추정** 테스트 가정으로 지정된 경우 지정된 값은 0이 될 수 없습니다.

대응별 차이

대응별 차이 검정의 거듭제곱 추정

대응별 차이 검정의 거듭제곱을 추정할지 여부를 제어합니다. 기본적으로 선택적 설정은 사용 안 함으로 설정되므로, 대응별 차이의 결과는 표시되지 않습니다.

Bonferroni 수정

대응별 차이의 거듭제곱을 추정하는 데 Bonferroni 수정을 사용합니다. 기본 설정입니다.

Sidak 수정

대응별 차이의 거듭제곱을 추정하는 데 Sidak 수정을 사용합니다.

최소유의차(LSD)

대응별 차이의 거듭제곱을 추정하는 데 LSD 수정을 사용합니다.

신뢰구간의 절반 너비를 지정하십시오.

신뢰구간 절반 너비 값을 기준으로 표본 크기를 추정합니다. 0 - 1 범위의 값을 입력하십시오. 단일 표본 이항 검정의 경우 값의 범위는 0 - 0.5여야 합니다.

참고: 중복된 값은 무시됩니다.

- 지정된 절반 너비 값을 목록에 추가하려면 **추가**를 클릭하십시오.
- 기존의 절반 너비 값을 강조표시하고 **변경**을 클릭하여 값을 업데이트하십시오.
- 기존의 절반 너비 값을 강조표시하고 **제거**를 클릭하여 목록에서 값을 제거하십시오.

일원배치 분산분석의 거듭제곱 분석*:* 도표

플롯 대화 상자는 샘플 및 효과 크기 도표별 2차원 및 3차원 거듭제곱을 설명하기 위해 출력되는 플롯을 제어합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표

거듭제곱 추정 대 전체 표본크기

이 선택적 설정을 사용으로 설정할 경우 전체 표본크기별 2차원 거듭제곱 차트를 제어하는 옵션이 제공됩니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

전체 표본크기의 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 전체 표본크기의 기본 도표 범위가 사용됩니다.

하한

전체 표본크기별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 다음 값보다 크거나 같아야 합니다.

- 2 x **집단 크기**에 지정된 정수 수
- 2 x **집단 크기**에 지정된 정수의 합 / **집단 크기**의 최소 정수 값

값은 **상한** 값보다 클 수 없습니다.

상한

전체 표본크기별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 다음 값보다 작거나 같아야 합니다.

- 5000 / **집단 크기**에 지정된 최대 정수 값 x **집단 크기**에 지정된 정수의 합

값은 **하한** 값보다 커야 하며 2147483647을 초과하지 않아야 합니다.

거듭제곱 추정 대 통합 표준편차

기본적으로 이 선택적 설정은 사용 안함으로 설정됩니다. 이 설정은 통합 표준편차별 2차원 거듭제곱 도표를 제어합니다. 사용으로 설정할 경우 출력에 도표가 표시됩니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 통합 표준편차의 기본 도표 범위가 사용됩니다.

참고:

지정된 **집단 평균** 값이 모두 같을 경우 도표가 사용 안함으로 설정됩니다.

통합 표준편차의 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다.

하한

통합 표준편차별 2차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

통합 표준편차별 2차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 합니다.

3차원 도표

거듭제곱 추정 대

전체 샘플 크기(x축) 및 효과 크기(y축) 차트, 수직 및 수평 회전 설정, 사용자 지정 샘플 및 효과 크기의 플롯 범위별로 3차원 거듭제곱을 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

참고:

지정된 **집단 평균** 값이 모두 같을 경우 도표가 사용 안함으로 설정됩니다.

통합 표준편차(X축) 및 전체 표본크기(Y축)

이 선택적 설정은 전체 표본크기(X축) 및 통합 표준편차(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

통합 표준편차(Y축) 및 전체 표본크기(X축)

이 선택적 설정은 전체 표본크기(Y축) 및 통합 표준편차(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

전체 표본크기의 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한

전체 표본크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

전체 표본크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 합니다.

통합 표준편차의 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기의 기본 도표 범위가 사용됩니다.

하한

통합 표준편차별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

통합 표준편차별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 합니다.

거듭제곱 추정 대 표본크기

효과 크기와 사용자가 지정한 샘플 및 효과 크기의 플롯 범위에 따라 3차원 거듭제곱을 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

참고:

지정된 **집단 평균** 값이 모두 같을 경우 도표가 사용 안함으로 설정됩니다.

통합 표준편차(X축) 및 전체 표본크기(Y축)
선택적 설정은 총 샘플 크기(x축) 및 효과 크기(y축)에 의한 3차원 거듭제곱을 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

통합 표준편차(Y축) 및 전체 표본크기(X축)
선택적 설정은 총 샘플 크기(y축) 및 효과 크기(x축) 차트의 3차원 거듭제곱을 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

전체 표본크기의 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 표본크기의 기본 도표 범위가 사용됩니다.

하한
전체 표본크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
전체 표본크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 합니다.

효과크기 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 효과크기의 기본 도표 범위가 사용됩니다.

하한
전체 표본크기별 3차원 거듭제곱 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
전체 표본크기별 3차원 거듭제곱 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 합니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

비율

IBM SPSS 통계 Base Edition에는 다음과 같은 통계 기능이 포함되어 있습니다.

일표본 이항검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

이항 분포는 일련의 베르누이 시행을 기반으로 하며, 서로 독립된 것으로 간주되는 고정된 총 시행 수를 포함하여 실험을 모형화하는 데 사용할 수 있습니다. 시행할 때마다 성공 결과에 대해 동일한 확률을 갖는 이분법적 결과가 도출됩니다.

일표본 이항검정은 비율 모수를 가설값과 비교하여 비율 모수에 대한 통계적 추론을 수행합니다. 이러한 검정에 대한 거듭제곱을 추정하는 방법은 정규 근사값 또는 이항 열거입니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > **검정력 분석** > **제안** > **1-샘플 이항 검정**

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』의 내용을 참조하십시오.
4. **검정력**이 검정 가정 **추정** 설정으로 선택된 경우 적절한 **총 시행 횟수** 값을 입력하십시오. 값은 1보다 크거나 같은 정수여야 합니다.
5. **모집단 비율** 필드에 비율 모수의 대립가설 값을 지정하는 값을 입력합니다. 값은 단일 숫자여야 합니다.

   참고: 검정력 값이 지정되면 **모집단 비율** 값은 **널값**과 같을 수 없습니다.
6. 선택적으로 검정할 비율 모수의 영가설 값을 지정하는 값을 **널값** 필드에 입력합니다. 값은 0-1 범위의 단일 숫자여야 합니다. 기본값은 0.50입니다.
7. 거듭제곱 추정 방법을 선택합니다.

   정규 근사값
   정규 근사값을 사용합니다. 기본 설정입니다.

   연속성 수정 적용
   정규 근사값 방법에 연속성 수정을 사용할지 여부를 제어합니다.

   이항 열거
   이항 열거 방법을 사용합니다. 선택적으로 표본크기를 추정하는 데 허용되는 최대 시간(분)을 지정하려면 **시간 한계** 필드를 사용합니다. 시간 한계에 도달하면 분석이 종료되고 경고 메시지가 표시됩니다. 지정할 경우 값은 분 수를 표시하는 한 개의 양의 정수여야 합니다. 기본 설정은 5분입니다.

   참고: 선택한 검정력 추정 가정은 **총 시행 횟수** 값이 500을 초과할 때 영향을 미치지 않습니다.
8. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

   비방향성(양측) 분석
   선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

   방향성(단측) 분석
   선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.
9. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
10. 선택적으로 **플롯**을 클릭하여 14 페이지의 『일표본 이항의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정하십시오.

    참고: 플롯은 **검정력**이 검정 가정 **추정**으로 선택되고 **이항 열거**가 선택되지 않은 경우에만 사용 가능합니다.
11. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

일표본 이항의 거듭제곱 분석: 도표

플롯 대화 상자는 2차원 및 3차원 거듭제곱을 도표로 설명하기 위해 출력되는 도표를 제어하기 위한 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표
2차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 영가설 값
사용으로 설정할 경우 이 선택적 설정은 널값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 대립가설 값
사용으로 설정할 경우 이 선택적 설정은 대체 값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 가설값 간 차이
사용으로 설정할 경우 이 선택적 설정은 가설값 간 차이별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 총 시행 수
사용으로 설정할 경우 이 선택적 설정은 총 시행 수별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

총 시행 수의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
2차원 거듭제곱 추정 대 총 시행 수 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
2차원 거듭제곱 추정 대 총 시행 수 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

3차원 도표
3차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 총 시행 수
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 가설값 간 차(Y축)
이 선택적 설정은 총 시행 수(X축) 및 가설값 간 차이(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 가설값 간 차(X축)
이 선택적 설정은 총 시행 수(Y축) 및 가설값 간 차이(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

총 시행 수의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
3차원 거듭제곱 추정 대 총 시행 수 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
3차원 거듭제곱 추정 대 총 시행 수 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

거듭제곱 추정 대 영가설 값
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 대립가설 값(Y축)
이 선택적 설정은 널값(X축) 및 대체 값(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 대립가설 값(X축)
이 선택적 설정은 널값(Y축) 및 대체 값(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전

이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

대응표본 이항검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

이항 분포는 일련의 베르누이 시행을 기반으로 하며, 서로 독립된 것으로 간주되는 고정된 총 시행 수를 포함하여 실험을 모형화하는 데 사용할 수 있습니다. 시행할 때마다 성공 결과에 대해 동일한 확률을 갖는 이분법적 결과가 도출됩니다.

대응표본 이항은 두 개의 관련된 이항 모집단에서 표본 추출한 매치 대응 개체를 기준으로 두 개의 비율 모수를 비교하기 위해 McNemar 검정의 거듭제곱을 추정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > **거듭제곱 분석** > **제안** > **관련 샘플 이항 검정**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』 의 내용을 참조하십시오.
4. **거듭제곱**이 검정 가정 **추정** 메소드로 선택된 경우 적절한 **총 대응 수** 값을 입력하십시오.
5. **비율** 또는 **빈도**에 대한 검정값을 지정하려면 선택합니다.
   - **비율**을 선택한 경우 **비율 1** 및 **비율 2** 필드에 값을 입력합니다. 값의 범위는 0-1이어야 합니다.
   - **빈도**를 선택한 경우 **빈도 1** 및 **빈도 2** 필드에 값을 입력합니다. 값은 0과 **총 대응 수**로 지정한 값 사이에 있어야 합니다.

   비율 참고사항:

   - **비율**은 **거듭제곱** 값을 지정한 경우에만 사용 가능한 옵션입니다.
   - **검정 값은 한계**가 선택되지 않은 경우: 0 ≤ **비율 1** + **비율 2** ≤ 1
   - **주변 검정값임**을 선택한 경우:
     - **비율 1** * **비율 2** > 0
     - **비율 1** < 1
     - **비율 2** < 1
     - **비율 1**과 **비율 2**의 값은 동일할 수 없습니다.

   빈도 참고사항:

   - **빈도** 설정은 **거듭제곱**이 검정 가정 **추정** 설정으로 선택된 경우에만 사용 가능합니다.
   - **검정 값은 한계**가 선택되지 않은 경우: 0 ≤ **계수 1** + **계수 2** ≤ **총 대응 수**
   - **주변 검정값임**을 선택한 경우:
     - **빈도 1** * **빈도 2** > 0
     - **빈도 1** < **총 대응 수**

- **빈도 2 < 총 대응 수**

6. 선택적으로 **주변 검정값임**을 선택하여 지정된 비율 또는 빈도 값이 주변 검정값인지 여부를 제어할 수 있습니다. **주변 검정값임**이 사용으로 설정된 경우에는 **매치 대응 간 상관관계** 값을 지정해야 합니다. 값은 -1과 1 사이의 단일 값이어야 합니다.
7. 거듭제곱 추정 방법을 선택합니다.

   정규 근사값
   정규 근사값을 사용합니다. 기본 설정입니다.

   이항 열거
   이항 열거 방법을 사용합니다. 선택적으로 표본크기를 추정하는 데 허용되는 최대 시간(분)을 지정하려면 **시간 한계** 필드를 사용합니다. 시간 한계에 도달하면 분석이 종료되고 경고 메시지가 표시됩니다. 지정할 경우 값은 분 수를 표시하는 한 개의 양의 정수여야 합니다. 기본 설정은 5분입니다.
8. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

   비방향성(양측) 분석
   선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

   방향성(단측) 분석
   선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.
9. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
10. 선택적으로 **플롯**을 클릭하여 16 페이지의 『대응표본 이항검정의 거듭제곱 분석』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정하십시오.

    참고: 플롯은 **검정력**이 검정 가정 **추정**으로 선택되고 **이항 열거**가 선택되지 않은 경우에만 사용 가능합니다.
11. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

대응표본 이항의 거듭제곱 분석: 도표

플롯 대화 상자는 2차원 및 3차원 거듭제곱 추정 차트를 설명하기 위해 출력되는 도표를 제어합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표
2차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 총 대응 수
사용으로 설정할 경우 이 선택적 설정은 총 대응 수별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

총 대응 수의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
2차원 거듭제곱 추정 대 총 대응 수 도표의 하한을 제어합니다. 값은 0보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
2차원 거듭제곱 추정 대 총 대응 수 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 2500을 초과하지 않아야 합니다.

거듭제곱 추정 대 위험도 차이
사용으로 설정할 경우 이 선택적 설정은 위험도 차이별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 위험도 비율
사용으로 설정할 경우 이 선택적 설정은 위험도 비율별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

위험도 비율의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
2차원 거듭제곱 추정 대 위험도 비율 도표의 하한을 제어합니다. 값은 **상한** 값보다 클 수 없습니다.

상한
2차원 거듭제곱 추정 대 위험도 비율 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 10을 초과하지 않아야 합니다.

거듭제곱 추정 대 오즈비
사용으로 설정할 경우 이 선택적 설정은 오즈비별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

오즈비의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
2차원 거듭제곱 추정 대 오즈비 도표의 하한을 제어합니다. 값은 **상한** 값보다 클 수 없습니다.

상한
2차원 거듭제곱 추정 대 오즈비 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 10을 초과하지 않아야 합니다.

거듭제곱 추정 대 매치 대응 간 상관관계(C)
일치하는 쌍 도표 간의 상관과 비교하여 추정된 거듭제곱 Ω를 제어합니다. 차트는 한계 비율 또는 계수가 지정되는 경우(불일치 비율이 아니라)에만 작성됩니다.

3차원 도표
3차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 불일치 비율
사용으로 설정할 경우 이 선택적 설정은 불일치 비율별 3차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 주변 비율
사용으로 설정할 경우 이 선택적 설정은 주변 비율별 3차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

참고: 이 설정은 **주변 검정값임**을 선택한 경우에만 사용할 수 있습니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

독립표본 이항검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

이항 분포는 일련의 베르누이 시행을 기반으로 하며, 서로 독립된 것으로 간주되는 고정된 총 시행 수를 포함하여 해당 실험을 모형화하는 데 사용할 수 있습니다. 시행할 때마다 "성공" 결과에 대해 동일한 확률을 갖는 이분법적 결과가 도출됩니다. 독립표본 이항검정은 두 개의 독립된 비율 모수를 검정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 검정력 분석 > 제안 > 독립-샘플 이항 검정

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』의 내용을 참조하십시오.

   선택적으로 **그룹 크기 비율** 값을 지정하십시오. 기본값은 1입니다.
4. **검정력**이 검정 가정 **추정** 설정으로 선택된 경우 **그룹 1 및 그룹 2에 대한 총 시행 수**를 지정하는 값을 입력하십시오. 값은 1보다 큰 정수여야 합니다.
5. 두 집단에 대한 비율 모수를 지정합니다. 두 값 모두 0-1 범위에 있어야 합니다.

   참고: 거듭제곱 값을 지정한 경우에는 두 값이 같을 수 없습니다.
6. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
7. 원하는 검정 메소드를 선택합니다.

   카이제곱 검정
   Pearson 카이제곱 검정을 기반으로 거듭제곱을 추정합니다. 기본 설정입니다.

   통합 표준편차임
   이 선택적 설정은 표준편차 추정이 통합인지 아니면 비통합인지 여부를 제어합니다. 이 설정은 기본적으로 사용으로 설정됩니다.

   연속성 수정 적용
   이 선택적 설정은 연속성 수정을 사용할지 여부를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

   T 검정
   스튜던트 T 검정을 기반으로 거듭제곱을 추정합니다.

   통합 표준편차임
   이 선택적 설정은 표준편차 추정이 통합인지 아니면 비통합인지 여부를 제어합니다. 이 설정은 기본적으로 사용으로 설정됩니다.

   우도비 검정
   우도비 검정을 기반으로 거듭제곱을 추정합니다.

   Fisher의 정확 검정
   Fisher의 정확 검정을 기반으로 거듭제곱을 추정합니다.

   노트:

   - 경우에 따라 Fisher의 정확검정을 완료하는 데 상당한 시간이 걸릴 수 있습니다.
   - Fisher의 정확검정을 선택할 경우 모든 도표가 차단됩니다.
8. 거듭제곱 추정 방법을 선택합니다.

   정규 근사값
   정규 근사값을 사용합니다. 기본 설정입니다.

   이항 열거
   이항 열거 방법을 사용합니다. 선택적으로 표본크기를 추정하는 데 허용되는 최대 시간(분)을 지정하려면 **시간 한계** 필드를 사용합니다. 시간 한계에 도달하면 분석이 종료되고 경고 메시지가 표시됩니다. 지정할 경우 값은 분 수를 표시하는 한 개의 양의 정수여야 합니다. 기본 설정은 5분입니다.

9. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

   비방향성(양측) 분석
   선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

   방향성(단측) 분석
   선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.

10. 선택적으로 **플롯**을 클릭하여 20 페이지의 『독립표본 이항검정의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정할 수 있습니다.

    참고: 플롯은 **검정력**이 검정 가정 **추정**으로 선택되고 **이항 열거**가 선택되지 않은 경우에만 사용 가능합니다.

11. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

독립표본 이항검정의 거듭제곱 분석*:* 도표

플롯 대화 상자는 2차원 및 3차원 거듭제곱을 도표로 설명하기 위해 출력되는 도표를 제어하기 위한 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표
2차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 집단 크기 비율/거듭제곱 추정 대 집단 크기
사용으로 설정할 경우 이 선택적 설정은 집단 크기 비율별 2차원 거듭제곱 도표를 제어합니다. 선택할 경우 이 설정으로 차트가 표시됩니다. 다중 거듭제곱 값이 지정된 경우(**거듭제곱 추정 대 샘플 크기**) 추가 설정을 사용할 수 없습니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

집단 크기 비율의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
2차원 거듭제곱 추정 대 총 대응 수 도표의 하한을 제어합니다. 값은 .01보다 커야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
2차원 거듭제곱 추정 대 총 대응 수 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 100을 초과하지 않아야 합니다.

거듭제곱 추정 대 위험도 차이
사용으로 설정할 경우 이 선택적 설정은 위험도 차이별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 위험도 비율
사용으로 설정할 경우 이 선택적 설정은 위험도 비율별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

위험도 비율의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
2차원 거듭제곱 추정 대 위험도 비율 도표의 하한을 제어합니다. 값은 **상한** 값보다 클 수 없습니다.

상한
2차원 거듭제곱 추정 대 위험도 비율 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 10을 초과하지 않아야 합니다.

거듭제곱 추정 대 오즈비
사용으로 설정할 경우 이 선택적 설정은 오즈비별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

오즈비의 도표 범위

선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한

2차원 거듭제곱 추정 대 오즈비 도표의 하한을 제어합니다. 값은 **상한** 값보다 클 수 없습니다.

상한

2차원 거듭제곱 추정 대 오즈비 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 10을 초과하지 않아야 합니다.

3차원 도표

3차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 비율

선택할 경우 이 선택적 설정은 다음과 같은 비율별 거듭제곱 옵션을 제공합니다.

집단 1의 비율(X축) 및 집단 2의 비율(Y축)

집단 1의 비율(X축) 및 집단 2의 비율(Y축)별 3차원 거듭제곱 차트를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

집단 1의 비율(Y축) 및 집단 2의 비율(X축)

집단 2의 비율(X축) 및 집단 1의 비율(Y축)별 3차원 거듭제곱 차트를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 집단 크기

선택할 경우 이 선택적 설정은 다음과 같은 집단 크기별 거듭제곱 옵션을 제공합니다.

집단 1의 크기(X축) 및 집단 2의 크기(Y축)

집단 1의 시행 수(X축) 및 집단 2의 시행 수(Y축)별 3차원 거듭제곱 차트를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

집단 1의 크기(Y축) 및 집단 2의 크기(X축)

집단 2의 시행 수(X축) 및 집단 1의 시행 수(Y축)별 3차원 거듭제곱 차트를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

집단 1 크기의 도표 범위

선택할 경우 집단 1 도표 범위에 대한 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한

2차원 거듭제곱 추정 대 오즈비 도표의 하한을 제어합니다. 값은 2보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

2차원 거듭제곱 추정 대 오즈비 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 2500을 초과하지 않아야 합니다.

집단 2 크기의 도표 범위

선택할 경우 집단 2 도표 범위에 대한 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한

2차원 거듭제곱 추정 대 오즈비 도표의 하한을 제어합니다. 값은 2보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한

2차원 거듭제곱 추정 대 오즈비 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 2500을 초과하지 않아야 합니다.

수직 회전

이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

상관관계

IBM SPSS 통계 Base Edition에는 다음과 같은 통계 기능이 포함되어 있습니다.

일표본 Pearson 상관 검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

Pearson 적률 상관계수는 이변량 정규 분포를 따르는 것으로 간주되는 두 개의 척도 확률변수 간 선형 여관의 강도를 측정합니다. 관례적으로 이 상관계수는 무차원 수량이며 두 연속형 변수 간의 공분산을 표준화하여 구할 수 있으므로 -1과 1 사이에 있습니다.

이 검정은 Fisher의 점근법을 사용하여 일표본 Pearson 상관에 대한 거듭제곱을 추정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 검정력 분석 > 상관 관계 > Pearson 제품-모멘트

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』의 내용을 참조하십시오.
4. **검정력**이 검정 가정 **추정** 설정으로 선택된 경우 적절한 **샘플 크기(쌍)** 값을 입력하십시오. 값은 3보다 큰 단일 정수여야 합니다.
5. **Pearson 상관 모수** 필드에 상관 모수의 대립가설 값을 지정하는 값을 입력합니다. 값은 -1과 1 사이의 단일 숫자여야 합니다.

   참고: 검증력이 지정된 경우 **Pearson 상관 매개변수** 값은 -1 또는 1이 될 수 없으며 **널값**과 같을 수 없습니다.
6. 선택적으로 검정할 상관 모수의 영가설 값을 지정하는 값을 **널값** 필드에 입력합니다. 값은 -1과 1 사이의 단일 숫자여야 합니다. 기본값은 0입니다.

   참고: 검정력이 지정되면 **널값**은 -1 또는 1이 될 수 없습니다.
7. 선택적으로 **거듭제곱 추정에 편향 수정 수식 사용**을 선택하여 편향 조정을 사용할지 아니면 무시할지 여부를 지정합니다. 이 설정은 기본적으로 사용으로 설정되므로, 거듭제곱 추정에 편향 조정 항이 포함되어 있습니다. 이 설정을 선택하지 않을 경우 편향 조정 항이 무시됩니다.
8. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

   비방향성(양측) 분석
   선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

   방향성(단측) 분석
   선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.
9. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.

10. 선택적으로 **플롯**을 클릭하여 23 페이지의 『일표본 Pearson 상관의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정할 수 있습니다.

    참고: 플롯은 **거듭제곱**이 검정 가정으로 선택된 경우에만 사용 가능합니다.

11. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

일표본 *Pearson* 상관의 거듭제곱 분석*:* 도표

플롯 대화 상자는 2차원 및 3차원 거듭제곱을 도표로 설명하기 위해 출력되는 도표를 제어하기 위한 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표
2차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 영가설 값
사용으로 설정할 경우 이 선택적 설정은 널값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 대립가설 값
사용으로 설정할 경우 이 선택적 설정은 대체 값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 가설값 간 차이
사용으로 설정할 경우 이 선택적 설정은 가설값 간 차이별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 표본크기(대응)
사용으로 설정할 경우 이 선택적 설정은 표본크기별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

표본크기의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 2차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 2차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

3차원 도표
3차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 표본크기
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 가설값 간 차(Y축)
이 선택적 설정은 표본크기(X축) 및 가설값 간 차(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 가설값 간 차(X축)
이 선택적 설정은 표본크기(Y축) 및 가설값 간 차(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

표본크기(대응)의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 3차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 3차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

거듭제곱 추정 대 영가설 값
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 대립가설 값(Y축)
이 선택적 설정은 널값(X축) 및 대체 값(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 대립가설 값(X축)
이 선택적 설정은 널값(Y축) 및 대체 값(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

일표본 Spearman 상관 검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

Spearman 순위-순서 상관계수는 일반적으로 중도절단되고 정규 분포를 따르지 않는 두 변수 간의 단조 관계를 측정하는 순위 기반의 비모수 통계입니다. Spearman 순위-순서 상관은 두 변수들의 순위 값 간의 Pearson 상관계수와 동일하며, -1과 1 사이의 범위이기도 합니다. Spearman 순위 상관 검정의 검증력을 탐지하는 것은 수문학적 시계열 데이터의 분석에서 중요한 주제입니다.

이 검정은 Fisher의 점근법을 사용하여 일표본 Spearman 순위-순서 상관에 대한 거듭제곱을 추정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > **검증력 분석** > **상관 관계** > **Spearman 순위-순서**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』 의 내용을 참조하십시오.
4. **검정력**이 검정 가정 **추정** 설정으로 선택된 경우 적절한 **샘플 크기(쌍)** 값을 입력하십시오. 값은 3보다 큰 단일 정수여야 합니다.
5. **Spearman 상관 모수** 필드에 상관 모수의 대립가설 값을 지정하는 값을 입력합니다. 값은 -1과 1 사이의 단일 숫자여야 합니다.

참고: 검증력이 지정된 경우 **Spearman 상관 매개변수** 값은 -1 또는 1이 될 수 없으며 **널값**과 같을 수 없습니다.

6. 선택적으로 검정할 상관 모수의 영가설 값을 지정하는 값을 **널값** 필드에 입력합니다. 값은 -1과 1 사이의 단일 숫자여야 합니다. 기본값은 0입니다.

   참고: 검정력이 지정되면 **널값**은 -1 또는 1이 될 수 없습니다.

7. 선택적으로 거듭제곱 분석을 위해 근사 분산을 추정하는 방식을 판별하는 옵션을 선택합니다.

   Bonett 및 Wright
   Bonett 및 Wright가 제안한 분산을 추정합니다. 기본 설정입니다.

   Fieller, Hartley 및 Pearson
   Fieller, Hartley 및 Pearson이 제안한 분산을 추정합니다.

   Caruso 및 Cliff
   Caruso 및 Cliff가 제안한 분산을 추정합니다.

8. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

   비방향성(양측) 분석
   선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

   방향성(단측) 분석
   선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.

9. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.

10. 선택적으로 **플롯**을 클릭하여 25 페이지의 『일표본 Spearman 검정의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정하십시오.

    참고: 플롯은 **거듭제곱**이 검정 가정으로 선택된 경우에만 사용 가능합니다.

11. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

일표본 *Spearman* 검정의 거듭제곱 분석*:* 도표

플롯 대화 상자는 2차원 및 3차원 거듭제곱을 도표로 설명하기 위해 출력되는 도표를 제어하기 위한 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표
2차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 영가설 값
사용으로 설정할 경우 이 선택적 설정은 널값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 대립가설 값
사용으로 설정할 경우 이 선택적 설정은 대체 값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 가설값 간 차이
사용으로 설정할 경우 이 선택적 설정은 가설값 간 차이별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 표본크기(대응)
사용으로 설정할 경우 이 선택적 설정은 표본크기별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

표본크기의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 2차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 2차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

3차원 도표
3차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 표본크기
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 가설값 간 차(Y축)
이 선택적 설정은 표본크기(X축) 및 가설값 간 차(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 가설값 간 차(X축)
이 선택적 설정은 표본크기(Y축) 및 가설값 간 차(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

표본크기(대응)의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 3차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 3차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

거듭제곱 추정 대 영가설 값
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 대립가설 값(Y축)
이 선택적 설정은 널값(X축) 및 대체 값(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 대립가설 값(X축)
이 선택적 설정은 널값(Y축) 및 대체 값(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

Pearson 편상관 검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

편상관계수는 다른 또는 여러 다른 변수의 효과를 제거한 후 두 확률변수 간의 연관으로 설명할 수 있습니다. 혼동스러운 상황에서 유용한 측정입니다. Pearson 상관계수와 마찬가지로, 편상관계수도 -1과 1 사이의 무차원 수량입니다.

이 검정은 Fisher의 점근법을 사용하여 일표본 Pearson 상관에 대한 거듭제곱을 추정합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 검정력 분석 > 상관 관계 > 부분

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』의 내용을 참조하십시오.
4. **검정력**이 테스트 가정 **추정** 설정으로 선택된 경우 적절한 **샘플 크기** 값을 입력하십시오. 값은 1보다 큰 하나의 정수여야 합니다.
5. **파티션화되는 것으로 추정되는 변수의 수**를 지정하는 값을 입력하십시오. 값은 0보다 크거나 같은 단일 정수여야 합니다.
6. **부분 상관 매개변수**의 대체 가설 값을 지정하는 값을 입력하십시오. 값은 -1과 1 사이의 단일 숫자여야 합니다.

   참고: 검정력이 지정된 경우 **부분 상관 매개변수** 값은 -1 또는 1이 될 수 없으며 **널값**와 같을 수 없습니다.
7. 선택적으로 검정할 편상관 모수의 영가설 값을 지정하는 값을 **널값** 필드에 입력합니다. 값은 -1과 1 사이의 단일 숫자여야 합니다. 기본값은 0입니다.

   참고: 검정력이 지정되면 **널값**은 -1 또는 1이 될 수 없습니다.
8. 단측 검정 또는 양측 검정인지 여부를 선택합니다.

   비방향성(양측) 분석
   선택할 경우 양측 검정이 사용됩니다. 기본 설정입니다.

   방향성(단측) 분석
   선택할 경우 단측 검정을 위해 거듭제곱이 계산됩니다.
9. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
10. 선택적으로 **플롯**을 클릭하여 27 페이지의 『Pearson 편상관의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정하십시오.

    참고: 플롯은 **거듭제곱**이 검정 가정으로 선택된 경우에만 사용 가능합니다.
11. 선택적으로 **정밀도**를 클릭하여 신뢰구간 반폭의 값을 지정하여 신뢰구간에 따라 표본 크기를 추정하십시오. 추가 정보는 31 페이지의 『검증력 분석: 정밀도』의 내용을 참조하십시오.

참고: 정밀도는 **샘플 크기**가 검정 가정 **추정** 메소드로 선택되고 **비방향성(양면)** 분석이 **검정 방향**으로 선택된 경우에만 사용 가능합니다.

Pearson 편상관의 거듭제곱 분석*:* 도표

플롯 대화 상자는 2차원 및 3차원 거듭제곱을 도표로 설명하기 위해 출력되는 도표를 제어하기 위한 옵션을 제공합니다. 대화 상자는 또한 3차원 도표의 수직 및 수평 회전도를 제어합니다.

2차원 도표
2차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 영가설 값
사용으로 설정할 경우 이 선택적 설정은 널값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 대립가설 값
사용으로 설정할 경우 이 선택적 설정은 대체 값별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 상쇄된 변수의 수(D)*
이 옵션을 사용하는 경우 이 선택적 설정은 파티션된 변수 도표의 수로 2차원 전원을 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 가설값 간 차이
사용으로 설정할 경우 이 선택적 설정은 가설값 간 차이별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 표본크기
사용으로 설정할 경우 이 선택적 설정은 표본크기별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

표본크기의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 2차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 2차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

3차원 도표
3차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 표본크기
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 가설값 간 차(Y축)
이 선택적 설정은 표본크기(X축) 및 가설값 간 차(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 가설값 간 차(X축)
이 선택적 설정은 표본크기(Y축) 및 가설값 간 차(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

표본크기(대응)의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 3차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 3차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

거듭제곱 추정 대 영가설 값
선택할 경우 이 설정으로 다음 옵션이 사용으로 설정됩니다.

(X축) 및 대립가설 값(Y축)
이 선택적 설정은 널값(X축) 및 대체 값(Y축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

(Y축) 및 대립가설 값(X축)
이 선택적 설정은 널값(Y축) 및 대체 값(X축)별 3차원 거듭제곱 도표를 제어합니다. 기본적으로 이 도표는 표시되지 않습니다. 지정할 경우 차트가 표시됩니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

회귀

IBM SPSS 통계 Base Edition에는 다음과 같은 통계 기능이 포함되어 있습니다.

일변량 선형 회귀 검정의 거듭제곱 분석

이 기능을 사용하려면 IBM SPSS 통계 Base Edition이 필요합니다.

거듭제곱 분석은 연구 계획, 디자인 및 전도에 중추적인 역할을 수행합니다. 일반적으로 거듭제곱은 소규모 예비 연구를 제외하고 표본 데이터를 수집하기 전에 계산합니다. 거듭제곱을 정확하게 추정하면 참 대립 가설 하에서 유한한 표본 크기를 기반으로 통계적으로 유의적인 차이가 발견될 가능성을 조사자에게 알려줄 수 있습니다. 거듭제곱이 너무 낮으면 유의적 차이를 발견할 가능성이 거의 없으며 실제 차이가 정말로 존재하는 경우에도 비유의적 결과가 발생할 가능성이 큽니다.

일변량 선형 회귀분석은 조사자가 척도 결과의 값을 설명하거나 예측하기 위해 여러 변수의 값을 사용하는 기본적이고 표준적인 통계적 접근 방식입니다.

일변량 선형 회귀 검정의 거듭제곱 분석은 다중 일변량 선형 회귀 모형에서 제 III 유형 *F* 검정의 거듭제곱을 추정합니다. 효과크기가 다중 (편)상관으로 표시된 경우 고정 예측자 및 변량 예측자 둘 다에 대한 접근 방식이 제공됩니다. 고정 예측자의 경우 거듭제곱 추정은 비중앙 *F*-분포를 기반으로 합니다. 변량 예측자의 경우 목표변수와 예측자가 다변량 정규 분포를 따르는 것으로 가정합니다. 이 경우 거듭제곱 추정은 동일한 다중 상관계수의 분포를 기반으로 합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 거듭제곱 분석 > 회귀분석 > 일변량 선형

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 검정 가정 **추정** 설정(**샘플 크기** 또는 **거듭제곱**)을 선택하십시오.
3. **샘플 크기**가 선택되면 샘플 크기 추정 값에 **단일 거듭제곱 값**을 입력하거나(값은 0과 1사이의 단일 값이어야 함) **그리드 거듭제곱 값**을 선택한 후 **그리드**를 클릭하여 특정 거듭제곱 값의 범위에 대한 투영된 샘플 크기를 보십시오.

   자세한 정보는 31 페이지의 『거듭제곱 분석: 격자 값』 의 내용을 참조하십시오.
4. **예상 거듭제곱**이 선택되면 적절한 **샘플 크기**를 입력하여 거듭제곱 평가 값을 입력하십시오. **모형의 절편 항 포함**이 사용으로 설정된 경우, 값은 총 모형 예측자 수 + 2보다 크거나 같아야 합니다. 그렇지 않은 경우 값은 총 모형 예측자 수 + 1보다 크거나 같은 단일 정수여야 합니다.
5. **모집단 다중 편상관계수** 필드에 다중 편상관계수의 값을 지정합니다. 값은 -1과 1 사이의 단일 값이어야 합니다.

   참고: 거듭제곱 값을 지정한 경우 **모집단 다중 편상관계수** 값은 0일 수 없습니다.

   모집단 다중 편상관계수를 선택하면 다음 설정이 사용 가능합니다.

   모형의 총 예측자 수
   총 예측자 수 또는 완전 모형의 예측자 수(절편 포함 안됨, 해당하는 경우)를 지정합니다. 값은 1보다 크거나 같은 단일 정수여야 합니다.

   검정 예측자 수
   검정 예측자 수 또는 중첩 모형의 예측자 수(절편 포함 안됨, 해당하는 경우)를 지정합니다. 값은 1보다 크거나 같아야 하지만, **모형의 총 예측자 수** 값을 초과하면 안됩니다.

6. **완전 모형** 및 **중첩 모형**에 다중 상관계수에 대한 **R 제곱 값**을 지정합니다. 값은 0과 1 사이의 단일값이어야 합니다.

   참고: 거듭제곱 값을 지정한 경우 **완전 모형** 값은 **중첩 모형** 값보다 커야 합니다.

   R 제곱 값을 선택하면 다음 설정이 사용 가능합니다.

   총 예측자 수 - 완전 모형
   전체 모델의 총 예측변수 수를 지정합니다(적용 가능한 경우 절편을 포함하지 않음). 값은 1보다 크거나 같은 단일 정수여야 합니다.

   총 예측자 수 - 중첩 모형
   중첩 모델의 총 예측변수 수를 지정합니다(적용 가능한 경우 절편을 포함하지 않음). 값은 1이상이어야 하지만 **예측변수의 총 수-전체 모형** 값보다 작아야 합니다.
7. 선택적으로 검정을 위한 제 I 유형 오차율의 유의수준을 **유의수준** 필드에 지정합니다. 값은 0과 1사이의 단일 double 값이어야 합니다. 기본값은 0.05입니다.
8. 선택적으로 **모형의 절편 항 포함** 설정을 선택할 수 있습니다. 이 설정은 기본적으로 사용으로 설정됩니다. 선택하지 않을 경우 절편 항이 거듭제곱 분석에서 제외됩니다.
9. 모형 예측자가 **고정** 또는 **변량**인지 여부를 선택할 수 있습니다. 기본 설정은 **고정**입니다.
10. 선택적으로 **플롯**을 클릭하여 30 페이지의 『일변량 선형 회귀의 거듭제곱 분석: 도표』 설정(차트 출력, 2차원 플롯 설정 및 3차원 플롯 설정)을 지정할 수 있습니다.

    참고: 플롯은 **거듭제곱**이 검정 가정으로 선택된 경우에만 사용 가능합니다.

일변량 선형 회귀의 거듭제곱 분석: 도표

2차원 및 3차원 거듭제곱 도표를 설명하기 위해 출력되는 도표를 제어할 수 있습니다. 3차원 도표의 경우 수직/수평 회전 각도 및 도구팁의 표시도 제어할 수 있습니다.

2차원 도표
2차원 거듭제곱 추정 대 도표를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 다중 편상관계수(T)
사용으로 설정할 경우 이 선택적 설정은 다중 편상관계수별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

거듭제곱 추정 대 표본크기
사용으로 설정할 경우 이 선택적 설정은 표본크기별 2차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

표본크기의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 2차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 2차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

3차원 도표
3차원 거듭제곱 추정 대 도표, 수평/수직 회전 설정 및 표본크기의 사용자 지정 도표 범위를 제어하는 옵션을 제공합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

거듭제곱 추정 대 표본크기
사용으로 설정할 경우 이 선택적 설정은 표본크기별 3차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

(X축) 및 다중 편상관계수(Y축)
이 선택적 설정은 표본크기(X축) 및 다중 편상관계수(Y축)별 3차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

(Y축) 및 다중 편상관계수(X축)
이 선택적 설정은 표본크기(Y축) 및 다중 편상관계수(X축)별 3차원 거듭제곱 도표를 제어합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다. 선택할 경우 이 설정으로 차트가 표시됩니다.

표본크기의 도표 범위
선택할 경우 하한 및 상한 옵션을 사용할 수 있습니다. **하한** 또는 **상한** 필드에 정수 값을 지정하지 않을 경우 기본 도표 범위가 사용됩니다.

하한
표본크기별 2차원 거듭제곱 추정 도표의 하한을 제어합니다. 값은 4보다 크거나 같아야 하며 **상한** 값을 초과하지 않아야 합니다.

상한
표본크기별 2차원 거듭제곱 추정 도표의 상한을 제어합니다. 값은 **하한** 값보다 커야 하며 5000을 초과하지 않아야 합니다.

수직 회전
이 선택적 설정은 3차원 도표의 수직 회전 각도(왼쪽부터 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수직으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 10입니다.

수평 회전
이 선택적 설정은 3차원 도표의 수평 회전 각도(앞쪽에서 시계방향)를 설정합니다. 마우스를 사용하여 차트를 수평으로 회전할 수 있습니다. 이 설정은 3차원 도표가 요청될 때 적용됩니다. 값은 359 이하의 단일 정수 값이어야 합니다. 기본값은 325입니다.

검증력 분석: 정밀도

정밀도 대화 상자는 모든 검정력 분석 프로시저에 사용 가능하며(일변량 선형 회귀분석 제외), 지정된 신뢰구간 절반 너비를 기반으로 표본 크기를 추정하는 옵션을 제공합니다. **단일 검증력 값** 또는 **그리드 검증력 값**이 **표본 크기** 추정값에 대해 지정되고 **검정 방향** 설정이 **비방향성(양면 분석)**으로 설정되는 경우 대화 상자를 사용할 수 있습니다.

참고: 검정 방향 설정은 단방향 ANOVA에 적용되지 않습니다.

표본 크기 추정 및 **그리드 검증력 값** 옵션을 선택한 경우(대화 상자를 표시하려면 **그리드** 제어를 클릭하십시오.)

신뢰구간 유형
적절한 신뢰구간 유형과 연속성 정정이 적용되는지 여부를 선택하십시오.

참고: 신뢰구간 유형은 검증력 분석 비율 프로시저(단일 표본 비율, 관련 표본 이항 검정 및 독립 표본 이항 검정)에만 사용할 수 있습니다.

신뢰구간의 절반 너비를 지정하십시오.
신뢰구간 절반 너비 값을 기준으로 표본 크기를 추정합니다. 0 - 1 범위의 값을 입력하십시오. 단일 표본 이항 검정의 경우 값의 범위는 0 - 0.5여야 합니다.

참고: 중복된 값은 무시됩니다.

- 지정된 절반 너비 값을 목록에 추가하려면 **추가**를 클릭하십시오.
- 기존의 절반 너비 값을 강조표시하고 **변경**을 클릭하여 값을 업데이트하십시오.
- 기존의 절반 너비 값을 강조표시하고 **제거**를 클릭하여 목록에서 값을 제거하십시오.

거듭제곱 분석: 격자 값

격자 값 대화 상자는 지정된 모든 **POWER** 범위 값의 예상된 표본크기를 격자 형식으로 보기 위해 **POWER** 값의 범위를 지정하는 옵션을 제공합니다.

표본 크기 추정 및 **그리드 검증력 값** 옵션이 선택된 경우 모든 검증력 분석 프로시저에 대해 **그리드 값** 대화 상자를 사용할 수 있습니다(대화 상자를 표시하려면 **그리드** 제어를 클릭하십시오).

단일 거듭제곱 지정
선택하면 분석을 실행하는 데 하나 이상의 값이 필요합니다. 여러 값이 허용되며 각 값은 `[0, 1]`에 속해야 합니다. 각 값을 하나 이상의 공백으로 구분하여 여러 값을 지정할 수 있습니다. **추가**, **변경** 및 **제거** 제어를 사용하여 검증력 값 목록의 값에 대해 작업하십시오.

모든 값은 고유해야 합니다(중복 값은 허용되지 않음).

거듭제곱 범위 지정
이 옵션을 선택하면 **시작** 값(value1)에서 **끝** 값(value2)까지의 범위에서 증분(value3) **을** 사용하여 전력 값의 범위를 지정할 수 있습니다. 지정된 경우 유효한 [value1 TO value2 BY value3]의 집합은 하나만 허용됩니다. 0 ≤ value1 ≤ value2 ≤ 1을 충족해야 합니다. value1 = value2 의 경우 value3 에 관계없이 단일 value1 을 지정하는 것과 동일합니다.

참고: 단일 검증력 지정 및 **검증력 범위 지정** 옵션은 독립적입니다. 단일 옵션 또는 두 옵션을 모두 선택할 수 있습니다.

메타분석

메타분석은 유사한 연구 질문에 답하는 연구 콜렉션에서 얻은 데이터의 분석입니다. 이러한 연구를 1 차 연구라고 합니다. 메타분석은 통계적 방법을 사용하여 효과의 전체적인 추정값을 산출하고, 연구 간 이질성을 조사하며, 출판 편향의 영향 또는 보다 일반적으로 최종 결과에 대한 소규모 연구 효과를 조사합니다.

IBM SPSS 통계 은(는) 2진 데이터(예: 로그 오즈비) 및 연속 데이터(예: Hedges의 g) 둘 다에 대한 표준 효과 크기 및 일반(계산된) 효과 크기를 지원합니다. 메타분석 정보(예: 연구별 효과 크기와 해당 표준오차 및 메타분석 모형 및 방법)는 메타분석 선언 단계 중에 지정됩니다. 이 정보는 모든 후속 메타분석에 자동으로 사용됩니다.

변량효과, 공통 효과 및 고정 효과 메타분석 모형이 지원됩니다. 선택한 메타분석 모형에 따라 다양한 추정 방법(예: 역 분산 및 Mantel-Haenszel)을 공통 효과 및 고정 효과 모형에 사용할 수 있습니다. 변량효과 모형의 연구 간 분산 모수에 대해 여러 가지 다른 추정량을 사용할 수도 있습니다.

메타분석 연속형

메타분석 연속형 프로시저는 효과 크기 추정을 위해 활성 데이터 세트에 원시 데이터가 제공된 경우 연속형 결과를 사용하여 메타분석을 수행합니다.

예
제2형 당뇨병 치료에 도움이 되는 변덕스럽지만 논쟁의 여지가 있는 약품을 조사하기 위해 역사상 여러 연구가 수행되었습니다. 경구용 약품이 식사 후 혈당 수치를 낮출 수 있다고 주장되었습니다. 데이터는 1979년부터 1986년까지 여러 연구 사이트에서 수집되었습니다.

수석 연구자가 경구용 약품의 효과에 대한 통계적 추론을 도출하려고 합니다. 데이터가 여러 연구에서 생성되었기 때문에 연구 전반에 걸쳐 결과를 종합하여 효과를 전반적으로 이해하고 결과의 근본적인 변동 원인을 식별하는 아이디어를 제안했습니다.

통계
신뢰 수준, 반복 방법, 단계-정지, 수렴 공차, 샘플 평균, 샘플 분산, 표준 편차, 추정된 효과 크기, Cohen의 *d*, Hedges의 *g*, Glass의 델타, 평균 차, 누적 분석, 추정 방법, 트림 및 필, 회귀 기반 테스트, 랜덤-효과 모델, 고정 효과 모델, 제한된 최대 가능성 추정기, 경험적 Bayes 추정기, Hedges 추정기, Hunter-Schmidt 추정기, DerSimonian-Laird 추정기, Sidik-Jonkman 추정기, Knapp-Hartung 표준 오차 조정, 절단된 Knapp-Hartung 표준 오차 조정, 계수, EGGER의 회귀 기반 검정, 절편, 승법 모형, 승법 분산 매개변수, 이차 추정기, 동질성 검정, 이질성 측정, 예측 간격, 추정된 표준 오차, 추정된 *p*-값, 누적 전체 효과 크기, 추정된 연구 가중치.

메타분석 연속형 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. **처리 그룹** 섹션에서 **연구 크기** 변수를 선택하여 처리 그룹의 샘플 크기를 표시하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
3. 처리 그룹에 대한 샘플 평균을 표시하려면 **평균** 변수를 선택하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.

4. 샘플 표준 편차를 지정하려면 **표준 편차**를 선택하고, 샘플 분산을 지정하려면 **변화**를 선택한 다음 처리 그룹의 표준 편차/분산을 나타내는 변수를 선택하십시오.
5. **제어 그룹** 섹션에서 **연구 크기** 변수를 선택하여 제어 그룹의 샘플 크기를 표시하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
6. 제어 그룹에 대한 샘플 평균을 표시하려면 **평균** 변수를 선택하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
7. 샘플 표준 편차를 지정하려면 **표준 편차**를 선택하고, 샘플 분산을 지정하려면 **변화** 를 선택한 다음 제어 그룹의 표준 편차/분산을 나타내는 변수를 선택하십시오.
8. 선택적으로 **연구 ID** 및/또는 **연구 레이블** 변수를 선택하십시오. 선택한 **연구 ID** 변수는 선택한 **연구 레이블** 변수와 같을 수 없습니다.
9. 선택적으로 **효과 크기** 설정을 선택하십시오.

   Cohen의 d
   기본 설정은 Cohen의 *d*를 추정합니다. **조정된 표준 오차를** 선택하면 이 설정은 대체 공식을 사용하여 코헨의 *d*와 그 분산을 $2(N_{treatment} + N_{control} - 2)$ 으로 나눈 값으로 추정합니다.

   Hedges의 g
   Hedges의 *g* 를 추정합니다. **조정된 표준 오차를** 선택하면 이 설정은 대체 공식을 사용하여 헤지의 *g*와 그 분산을 $2(N_{treatment} + N_{control} - 3.94)$ 으로 나누어 추정합니다.

   Glass 델타
   대조집단을 기준으로 Glass 델타를 추정합니다. **처리군에 기반한 표준화**가 선택된 경우, Glass의 델타는 처리군의 표준 편차를 기반으로 표준화됩니다.

   비표준화 평균차이
   두 모집단 표준편차가 같다고 가정하여 평균차이를 추정합니다. **같지 않은 그룹 분산**이 선택되면 두 집단 표준 편차가 동일하지 않은 것으로 가정하여 평균 차이가 추정됩니다.
10. 선택적으로 **모형** 설정을 선택하십시오. **자르기 및 채움** 설정이 사용 가능한 경우, 설정은 또한 자르기 및 채움 분석에서 풀링에 사용되는 모형을 제어합니다. **편향** 설정이 사용 가능한 경우, 설정은 회귀분석 기반 검정에서 사용하는 모형도 제어합니다.

    변량효과
    기본 설정은 변량효과 모형을 작성합니다.

    고정 효과
    고정 효과 모형을 작성합니다.
11. 선택적으로 다음을 수행할 수 있습니다.
    - **기준을 클릭하세요...** 를 사용하여 일반적인 기준을 지정할 수 있습니다.
    - 하위 그룹 및 누적 분석을 지정하려면 **분석**을 클릭하십시오.
    - 추정 방법을 지정하려면 **추론**을 클릭하십시오.
    - 대비 테스트를 제어하려면 **대비**를 클릭하십시오.
    - EGGER 회귀 기반 테스트를 수행하여 공개 바이어스에 액세스하려면 **바이어스**를 클릭하십시오.
    - 게시 바이어스의 트림 및 채우기 분석을 구현하려면 **트림 및 채우기**를 클릭하십시오.
    - 테이블 출력을 제어하려면 **인쇄**를 클릭하십시오.
    - 활성 데이터 세트에 예상 통계를 저장하려면 **저장**을 클릭하십시오.
    - 출력에 포함할 도표를 지정하려면 **플롯**을 누르십시오.
12. **확인**을 클릭하십시오.

메타분석 연속형: 기준

기준 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에 제공되는 원시 데이터의 연속 결과와 함께 메타 분석에 대한 기준을 지정하기 위한 설정을 제공합니다.

신뢰구간
이 선택적 설정은 신뢰수준을 지정합니다. 값은 0 - 100 범위의 숫자 값이어야 합니다. 기본 설정은 95입니다.

결측 데이터 범위
이 선택적 설정은 프로시저의 결측 데이터 처리 방법을 제어합니다.

분석별 결측값 제외
기본 설정은 특정 분석 각각에 사용된 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

목록별 결측값 제외
프로시저에 지정된 모든 분석에서 사용되는 모든 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

사용자 결측값
이 선택적 설정은 사용자가 결측값을 처리하는 방법을 제어합니다.

제외
기본 설정은 사용자 결측값을 유효한 값으로 처리합니다.

포함(N)
사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 값으로 처리합니다.

반복

최대반복계산
이 선택적 설정은 반복적 방법의 최대반복수를 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 100입니다. 값 0은 반복계산을 수행하지 않음을 의미합니다.

최대 단계 이분
이 선택적 설정은 반복적 방법의 최대 단계 이분을 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 5입니다. 0 값은 단계 이분이 적용되지 않음을 의미합니다.

수렴
이 선택적 설정은 수렴허용을 지정합니다. 값은 단일 양수값이어야 합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 기본값은 1E-6입니다.

메타분석 연속형 기준 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타 분석 연속형** 대화 상자에서 **기준**을 클릭하십시오.
3. 적합한 기준 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 분석

분석 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에 제공되는 원시 데이터의 연속 결과를 사용하여 메타분석에 대한 하위 그룹 및 누적 분석을 지정하기 위한 설정을 제공합니다.

부집단 분석
부집단 분석을 실행할 변수를 선택하십시오. 변수는 **누적 분석**에 정의된 변수와 같을 수 없습니다.

누적 분석
누적 분석을 실행하고 누적 메타분석이 수행될 변수를 선택하십시오. 변수는 **하위 그룹 분석**에 정의된 변수와 같을 수 없습니다. **오름차순**이 선택되면 누적 분석은 지정된 변수를 오름차순 기반으로 합니다. **내림차순**이 선택되면 누적 분석은 지정된 변수를 내림차순 기반으로 합니다.

누적 통계
추정된 누적 전체 효과 크기를 저장하는 옵션을 제공합니다. **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

누적 효과 크기
추정된 누적 전체 효과 크기를 저장합니다.

표준오차
누적 전체 효과 크기의 추정된 표준오차를 저장합니다.

신뢰구간 하한
누적 전체 효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
누적 전체 효과 크기의 추정된 신뢰구간 상한을 저장합니다.

p-값
누적 전체 효과 크기의 예상 *p*-값을 저장합니다.

대상
저장 데이터 세트 또는 데이터 파일을 지정하는 옵션을 제공합니다. **데이터 세트**를 선택하면 새 데이터 세트 이름을 지정할 수 있습니다(기본 데이터 세트 이름을 유지할 수 있음). **데이터 파일을** 선택한 후 **찾아보기...를** 클릭합니다. 을 클릭하여 저장 파일 이름과 위치를 선택합니다.

메타분석 연속형: 분석 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타 분석 연속형** 대화 상자에서 **분석**을 클릭하십시오.
3. 적합한 분석 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 추론

추론 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에 제공되는 원시 데이터에 대한 연속 결과와 함께 메타 분석을 위한 추정 방법을 지정하기 위한 설정을 제공합니다.

참고: 추론 대화 상자는 **랜덤 효과** 모델이 선택된 경우에만 사용 가능합니다.

추정량
추정량을 지정하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 추정량도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 추정량도 제어합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 표준오차 조정도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 표준오차 조정도 제어합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 연속형: 추론 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타-분석 연속형** 대화 상자에서 **추론**을 클릭하십시오.
3. 적합한 추론 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 대비

대조 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에서 제공되는 원시 데이터에 대한 연속형 결과에서의 메타 분석에 대한 대비 검정을 제어하기 위한 설정을 제공합니다.

변수
이 목록에는 사용 가능한 모든 데이터 세트 변수가 표시됩니다. 목록에서 변수를 선택하여 **대조** 목록으로 이동하십시오.

대비
이 목록은 활성 데이터 세트에 변수로 저장된 계수를 식별합니다. 다중 변수가 허용됩니다. 문자변수는 지원되지 않습니다.

사용자 입력 계수 값
사용자 제공 대비의 계수를 지정하는 설정을 제공합니다. 숫자 값만 허용됩니다. 유효한 대비검정을 공식화하려면 지정된 값의 개수가 유효한 연구 수와 매치해야 합니다.

메타분석 연속형: 대비 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타-분석 연속형** 대화 상자에서 **대조**를 클릭하십시오.
3. 적합한 대비 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 편향

편향 대화 상자는 효과 크기를 추정하기 위해 활성 데이터 세트에 제공된 원시 데이터에 대해 연속형 결과를 사용하여 메타 분석에 대해 Eggers 회귀 기반 테스트를 수행하여 공개 편향을 사용 가능하게 하는 설정을 제공합니다.

Egger의 회귀분석 기반 검정
이 설정을 선택하면 Egger의 회귀분석 기반 검정을 수행하여 출판 편향을 사용으로 설정합니다.

변수
이 목록은 사용 가능한 모든 데이터 세트 변수를 제공합니다.

공변량
변수 목록에서 선택된 변수는 공변량으로 처리됩니다. 다중 공변량이 허용됩니다.

요인
변수 목록에서 선택된 변수는 요인으로 취급됩니다. 다중 요인이 허용됩니다.

회귀분석에 절편 포함
회귀분석 기반 검정에서 절편 항을 제어합니다.

고정 효과 모형에 산포도 모수 포함
승법 모형 설정을 제어하고 승법 산포도 모수를 분석에 도입합니다. 이 설정은 고정 효과 모형을 선택한 경우에만 사용할 수 있습니다.

t 분포를 기준으로 통계량 추정
회귀분석 기반 검정에 사용된 분포를 제어합니다. 이 설정은 기본적으로 사용 가능하며, 이는 *t*-분포를 기반으로 하는 통계를 추정합니다. 이 설정을 선택하지 않으면 정규 분포를 기준으로 통계량이 추정됩니다.

메타분석 연속형: 편향 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타 분석 연속형** 대화 상자에서 **편향**을 클릭하십시오.
3. 적합한 편향 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 자르기 및 채우기

자르기 및 채움 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에서 제공되는 원시 데이터에서 연속형 결과에서의 메타 분석에 대한 출판 편향의 자르기 및 채움 분석을 구현하기 위한 설정을 제공합니다.

결측 연구 수 추정
출판 편향의 자르기 및 채우기 분석을 제어합니다. 이 설정을 선택하면 다른 대화 상자 설정이 사용으로 설정됩니다.

대체 연구 측면
결측 연구가 대체된 깔때기 도표의 방향을 지정하는 옵션을 제공합니다.

Egger 검정의 기울기로 결정
기본 설정은 Egger 검정의 추정된 기울기를 기준으로 측면을 결정합니다.

왼쪽
깔때기 도표의 왼쪽을 대체합니다.

오른쪽
깔때기 도표의 오른쪽을 대체합니다.

방법
결측 연구 수 추정 방법을 지정합니다.

선형
기본 설정은 선형 추정량을 계산합니다.

다음을 실행하십시오.
실행 추정량을 계산합니다.

2차
2차 추정량을 계산합니다.

반복 프로세스
반복 추정량 및 표준오차 조정을 지정하는 설정을 제공합니다.

고정 효과 모형
선택할 경우, 고정 효과 모형이 사용되며 반복 추정 및 표준오차 조정 옵션은 사용할 수 없습니다.

변량효과 모형
선택할 경우, 변량효과 모형이 사용되며 다음 설정을 사용할 수 있습니다.

추정량
반복 추정량을 지정하는 설정을 제공합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
자르기 및 채우기 알고리즘의 반복에 Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 연속형: 자르기 및 채우기 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타 분석 연속형** 대화 상자에서 **자르기 및 채움**을 클릭하십시오.
3. 적합한 자르기 및 채우기 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 인쇄

인쇄 대화 상자는 효과 크기를 추정하기 위해 활성 데이터 세트에 제공되는 원시 데이터의 연속 결과를 사용하여 메타 분석을 위한 테이블 출력을 제어하기 위한 설정을 제공합니다.

동질성/이질성
동질성 및 이질성 검정을 제어하는 설정을 제공합니다.

동질성 검정
선택할 경우 해당 동질성 검정이 출력에 제공됩니다.

이질성 측도
선택할 경우 이질성 측도가 출력에 제공됩니다.

효과 크기
다음과 같은 효과 크기 설정을 제공합니다.

개별 연구
개별 연구 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다.

누적 효과 크기
누적 분석 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. **분석** 대화 상자에서 **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

변량효과 모형의 예측구간
예측구간의 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. 이 설정은 변량효과 모형을 지정한 경우에만 사용할 수 있습니다.

메타분석 연속형: 인쇄 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타-분석 연속형** 대화 상자에서 **인쇄**를 클릭하십시오.
3. 적합한 인쇄 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 저장

저장 대화 상자에서는 효과 크기 추정을 위해 활성 데이터 세트에 제공된 원시 데이터의 연속 결과를 사용하여 메타 분석을 위해 활성 데이터 세트에 추정된 통계를 저장하기 위한 설정을 제공합니다.

개별 연구
추정된 효과 크기를 저장하는 옵션을 제공합니다.

개별 효과 크기
추정된 효과 크기를 저장합니다.

표준오차
효과 크기의 추정된 표준오차를 저장합니다.

신뢰구간 하한
효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
효과 크기의 추정된 신뢰구간 상한을 저장합니다.

p-값
효과 크기의 예상 *p*-값을 저장합니다.

연구 가중값
추정된 연구 가중값을 저장합니다.

연구 가중값 백분율
표준화된 연구 가중값을 백분율로 저장합니다.

메타분석 연속형: 저장 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 원시 데이터...
2. **메타-분석 연속형** 대화 상자에서 **저장**을 클릭하십시오.
3. 적합한 추정된 통계 저장 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형: 도표

플롯 대화 상자는 다음 플롯 유형에 대한 설정을 제공합니다.

- 66 페이지의 『포리스트 도표』
- 67 페이지의 『누적 포리스트 도표』
- 68 페이지의 『거품 도표』
- 69 페이지의 『깔때기 도표』
- 70 페이지의 『Galbraith 도표』

메타분석 연속형 효과 크기

메타분석 연속형 크기 프로시저는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공된 경우 연속형 결과를 사용하여 메타분석을 수행합니다.

예

제2형 당뇨병 치료에 도움이 되는 변덕스럽지만 논쟁의 여지가 있는 약품을 조사하기 위해 역사상 여러 연구가 수행되었습니다. 경구용 약품이 식사 후 혈당 수치를 낮출 수 있다고 주장되었습니다. 데이터는 1979년부터 1986년까지 여러 연구 사이트에서 수집되었습니다.

수석 연구자가 경구용 약품의 효과에 대한 통계적 추론을 도출하려고 합니다. 데이터가 여러 연구에서 생성되었기 때문에 연구 전반에 걸쳐 결과를 종합하여 효과를 전반적으로 이해하고 결과의 근본적인 변동 원인을 식별하는 아이디어를 제안했습니다.

통계

신뢰 수준, 반복 방법, 단계-정지, 수렴 공차, 샘플 평균, 샘플 분산, 표준 편차, 추정된 효과 크기, Cohen의 *d*, Hedges의 *g*, Glass의 델타, 평균 차, 누적 분석, 추정 방법, 트림 및 필, 회귀 기반 테스트, 랜덤-효과 모델, 고정 효과 모델, 제한된 최대 가능성 추정기, 경험적 Bayes 추정기, Hedges 추정기, Hunter-Schmidt 추정기, DerSimonian-Laird 추정기, Sidik-Jonkman 추정기, Knapp-Hartung 표준 오차 조정, 절단된 Knapp-Hartung 표준 오차 조정, 계수, EGGER의 회귀 기반 검정, 절편, 승법 모형, 승법 분산 매개변수, 이차 추정기, 동질성 검정, 이질성 측정, 예측 간격, 추정된 표준 오차, 추정된 *p*-값, 누적 전체 효과 크기, 추정된 연구 가중치.

메타분석 연속형 효과 크기 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 효과 크기를 나타내는 **효과 크기** 변수를 선택하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
3. **표준 편차**를 선택하여 효과 크기 표준 편차를 지정하거나 **분산**을 선택하여 효과 크기 분산을 지정한 다음 효과 크기에 대한 표준 편차/분산을 나타내는 변수를 선택합니다.
4. 선택적으로 **연구 ID** 및/또는 **연구 레이블** 변수를 선택하십시오. 선택한 **연구 ID**변수는 선택한 **연구 레이블** 변수와 같을 수 없습니다.
5. 선택적으로 **모형** 설정을 선택하십시오. **자르기 및 채움** 설정이 사용 가능한 경우, 설정은 또한 자르기 및 채움 분석에서 풀링에 사용되는 모형을 제어합니다. **편향** 설정이 사용 가능한 경우, 설정은 회귀분석 기반 검정에서 사용하는 모형도 제어합니다.

   변량효과
   기본 설정은 변량효과 모형을 작성합니다.

   고정 효과
   고정 효과 모형을 작성합니다.
6. 선택적으로 다음을 수행할 수 있습니다.
   - **기준을 클릭하세요...** 를 사용하여 일반적인 기준을 지정할 수 있습니다.
   - 하위 그룹 및 누적 분석을 지정하려면 **분석**을 클릭하십시오.
   - 추정 방법을 지정하려면 **추론**을 클릭하십시오.
   - 대비 테스트를 제어하려면 **대비**를 클릭하십시오.
   - EGGER 회귀 기반 테스트를 수행하여 공개 바이어스에 액세스하려면 **바이어스**를 클릭하십시오.
   - 게시 바이어스의 트림 및 채우기 분석을 구현하려면 **트림 및 채우기**를 클릭하십시오.
   - 테이블 출력을 제어하려면 **인쇄**를 클릭하십시오.
   - 활성 데이터 세트에 예상 통계를 저장하려면 **저장**을 클릭하십시오.
   - 출력에 포함할 도표를 지정하려면 **플롯**을 누르십시오.
7. **확인**을 클릭하십시오.

메타분석 연속형 효과 크기: 기준

기준 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 결과에 대한 메타 분석을 위한 기준방법을 지정하는 설정을 제공합니다.

신뢰구간
이 선택적 설정은 신뢰수준을 지정합니다. 값은 0 - 100 범위의 숫자 값이어야 합니다. 기본 설정은 95입니다.

결측 데이터 범위
이 선택적 설정은 프로시저의 결측 데이터 처리 방법을 제어합니다.

분석별 결측값 제외
기본 설정은 특정 분석 각각에 사용된 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

목록별 결측값 제외
프로시저에 지정된 모든 분석에서 사용되는 모든 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

사용자 결측값
이 선택적 설정은 사용자가 결측값을 처리하는 방법을 제어합니다.

제외
기본 설정은 사용자 결측값을 유효한 값으로 처리합니다.

포함(N)
사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 값으로 처리합니다.

반복

최대반복계산
이 선택적 설정은 반복적 방법의 최대반복수를 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 100입니다. 값 0은 반복계산을 수행하지 않음을 의미합니다.

최대 단계 이분
이 선택적 설정은 반복적 방법의 최대 단계 이분을 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 5입니다. 0 값은 단계 이분이 적용되지 않음을 의미합니다.

수렴
이 선택적 설정은 수렴허용을 지정합니다. 값은 단일 양수값이어야 합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 기본값은 1E-6입니다.

메타분석 연속형 효과 크기 기준 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **기준**을 클릭하십시오.
3. 적합한 기준 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 분석

분석 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 결과에 대한 메타 분석의 하위 그룹 및 누적 분석 지정에 대한 설정을 제공합니다.

부집단 분석
부집단 분석을 실행할 변수를 선택하십시오. 변수는 **누적 분석**에 정의된 변수와 같을 수 없습니다.

누적 분석
누적 분석을 실행하고 누적 메타분석이 수행될 변수를 선택하십시오. 변수는 **하위 그룹 분석**에 정의된 변수와 같을 수 없습니다. **오름차순**이 선택되면 누적 분석은 지정된 변수를 오름차순 기반으로 합니다. **내림차순**이 선택되면 누적 분석은 지정된 변수를 내림차순 기반으로 합니다.

누적 통계
추정된 누적 전체 효과 크기를 저장하는 옵션을 제공합니다. **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

누적 효과 크기
추정된 누적 전체 효과 크기를 저장합니다.

표준오차
누적 전체 효과 크기의 추정된 표준오차를 저장합니다.

신뢰구간 하한
누적 전체 효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
누적 전체 효과 크기의 추정된 신뢰구간 상한을 저장합니다.

p-값
누적 전체 효과 크기의 예상 *p*-값을 저장합니다.

대상
저장 데이터 세트 또는 데이터 파일을 지정하는 옵션을 제공합니다. **데이터 세트**를 선택하면 새 데이터 세트 이름을 지정할 수 있습니다(기본 데이터 세트 이름을 유지할 수 있음). **데이터 파일을** 선택한 후 **찾아보기...를** 클릭합니다. 을 클릭하여 저장 파일 이름과 위치를 선택합니다.

메타분석 연속형 효과 크기 분석 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **분석**을 클릭하십시오.
3. 적합한 분석 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 추론

추론 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 결과에 대한 메타 분석을 위한 추정 방법을 지정하는 설정을 제공합니다.

참고: 추론 대화 상자는 **랜덤 효과** 모델이 선택된 경우에만 사용 가능합니다.

추정량
추정량을 지정하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 추정량도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 추정량도 제어합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 표준오차 조정도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 표준오차 조정도 제어합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 연속형 효과 크기 추론 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **추론**을 클릭하십시오.
3. 적합한 추론 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 대비

대비 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 결과에 대한 메타 분석을 위한 대비 검정을 제어하는 설정을 제공합니다.

변수
이 목록에는 사용 가능한 모든 데이터 세트 변수가 표시됩니다. 목록에서 변수를 선택하여 **대조** 목록으로 이동하십시오.

대비
이 목록은 활성 데이터 세트에 변수로 저장된 계수를 식별합니다. 다중 변수가 허용됩니다. 문자변수는 지원되지 않습니다.

사용자 입력 계수 값
사용자 제공 대비의 계수를 지정하는 설정을 제공합니다. 숫자 값만 허용됩니다. 유효한 대비검정을 공식화하려면 지정된 값의 개수가 유효한 연구 수와 매치해야 합니다.

메타분석 연속형 효과 크기 대비 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **대비**를 클릭하십시오.
3. 적합한 대비 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 편향

편향 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 결과에서의 메타 분석에 대해 Eggers의 회귀분석 기반 검정을 수행하여 출판 편향을 사용하기 위한 설정을 제공합니다.

Egger의 회귀분석 기반 검정
이 설정을 선택하면 Egger의 회귀분석 기반 검정을 수행하여 출판 편향을 사용으로 설정합니다.

변수
이 목록은 사용 가능한 모든 데이터 세트 변수를 제공합니다.

공변량
변수 목록에서 선택된 변수는 공변량으로 처리됩니다. 다중 공변량이 허용됩니다.

요인
변수 목록에서 선택된 변수는 요인으로 취급됩니다. 다중 요인이 허용됩니다.

회귀분석에 절편 포함
회귀분석 기반 검정에서 절편 항을 제어합니다.

고정 효과 모형에 산포도 모수 포함
승법 모형 설정을 제어하고 승법 산포도 모수를 분석에 도입합니다. 이 설정은 고정 효과 모형을 선택한 경우에만 사용할 수 있습니다.

t 분포를 기준으로 통계량 추정
회귀분석 기반 검정에 사용된 분포를 제어합니다. 이 설정은 기본적으로 사용 가능하며, 이는 *t*-분포를 기반으로 하는 통계를 추정합니다. 이 설정을 선택하지 않으면 정규 분포를 기준으로 통계량이 추정됩니다.

메타분석 연속형 효과 크기 편향 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **편향**을 클릭하십시오.
3. 적합한 편향 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 자르기 및 채우기

자르기 및 채움 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공될 때 연속 결과를 사용하여 메타 분석에 대한 공개 편향의 자르기 및 채움 분석을 구현하기 위한 설정을 제공합니다.

결측 연구 수 추정
출판 편향의 자르기 및 채우기 분석을 제어합니다. 이 설정을 선택하면 다른 대화 상자 설정이 사용으로 설정됩니다.

대체 연구 측면
결측 연구가 대체된 깔때기 도표의 방향을 지정하는 옵션을 제공합니다.

Egger 검정의 기울기로 결정
기본 설정은 Egger 검정의 추정된 기울기를 기준으로 측면을 결정합니다.

왼쪽
깔때기 도표의 왼쪽을 대체합니다.

오른쪽
깔때기 도표의 오른쪽을 대체합니다.

방법
결측 연구 수 추정 방법을 지정합니다.

선형
기본 설정은 선형 추정량을 계산합니다.

다음을 실행하십시오.
실행 추정량을 계산합니다.

2차
2차 추정량을 계산합니다.

반복 프로세스
반복 추정량 및 표준오차 조정을 지정하는 설정을 제공합니다.

고정 효과 모형
선택할 경우, 고정 효과 모형이 사용되며 반복 추정 및 표준오차 조정 옵션은 사용할 수 없습니다.

변량효과 모형
선택할 경우, 변량효과 모형이 사용되며 다음 설정을 사용할 수 있습니다.

추정량
반복 추정량을 지정하는 설정을 제공합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
자르기 및 채우기 알고리즘의 반복에 Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 연속형 효과 크기 자르기 및 채우기 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **자르기 및 채움**을 클릭하십시오.
3. 적합한 자르기 및 채우기 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 인쇄

인쇄 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 메타 분석에 대한 테이블 출력을 연속형 결과로 제어하기 위한 설정을 제공합니다.

동질성/이질성
동질성 및 이질성 검정을 제어하는 설정을 제공합니다.

동질성 검정
선택할 경우 해당 동질성 검정이 출력에 제공됩니다.

이질성 측도
선택할 경우 이질성 측도가 출력에 제공됩니다.

효과 크기
다음과 같은 효과 크기 설정을 제공합니다.

개별 연구
개별 연구 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다.

누적 효과 크기
누적 분석 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. **분석** 대화 상자에서 **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

변량효과 모형의 예측구간
예측구간의 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. 이 설정은 변량효과 모형을 지정한 경우에만 사용할 수 있습니다.

메타분석 연속형 효과 크기 인쇄 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **인쇄**를 클릭하십시오.
3. 적합한 인쇄 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 저장

저장 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공될 때 연속형 결과를 갖는 메타 분석을 위해 활성 데이터 세트에 추정된 통계를 저장하기 위한 설정을 제공합니다.

개별 연구
추정된 효과 크기를 저장하는 옵션을 제공합니다.

표준오차
효과 크기의 추정된 표준오차를 저장합니다. 이 설정은 **효과 크기** 변수가 지정되지 않은 경우에만 사용 가능합니다.

신뢰구간 하한
효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
효과 크기의 추정된 신뢰구간 상한을 저장합니다.

p-값
효과 크기의 예상 *p*-값을 저장합니다.

연구 가중값
추정된 연구 가중값을 저장합니다.

연구 가중값 백분율
표준화된 연구 가중값을 백분율로 저장합니다.

메타분석 연속형 효과 크기 저장 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 연속형 결과 > 사전 계산된 효과 크기
2. **메타 분석 연속형 효과 크기** 대화 상자에서 **저장**을 클릭하십시오.
3. 적합한 추정된 통계 저장 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 연속형 효과 크기: 도표

플롯 대화 상자는 다음 플롯 유형에 대한 설정을 제공합니다.

• 66 페이지의 『포리스트 도표』

- 67 페이지의 『누적 포리스트 도표』
- 68 페이지의 『거품 도표』
- 69 페이지의 『깔때기 도표』
- 70 페이지의 『Galbraith 도표』

메타분석 이분형

메타분석 이분형 프로시저는 효과 크기 추정을 위해 활성 데이터 세트에 원시 데이터가 제공된 경우 이분형 결과를 사용하여 메타분석을 수행합니다.

예

제2형 당뇨병 치료에 도움이 되는 변덕스럽지만 논쟁의 여지가 있는 약품을 조사하기 위해 역사상 여러 연구가 수행되었습니다. 경구용 약품이 식사 후 혈당 수치를 낮출 수 있다고 주장되었습니다. 데이터는 1979년부터 1986년까지 여러 연구 사이트에서 수집되었습니다.

수석 연구자가 경구용 약품의 효과에 대한 통계적 추론을 도출하려고 합니다. 데이터가 여러 연구에서 생성되었기 때문에 연구 전반에 걸쳐 결과를 종합하여 효과를 전반적으로 이해하고 결과의 근본적인 변동 원인을 식별하는 아이디어를 제안했습니다.

통계

신뢰구간, 로그 오즈비, 페토의 로그 오즈비, 로그 위험 비율, 위험차, 랜덤 효과, 고정 효과, 역분산, Mantel-Haenszel, 반복, 단계적 이등분, 수렴, 누적 통계, 누적 효과 크기, 제한된 최대 우도, REML, 최대 우도, ML, Empirical Bayes, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Egger 검정, Harbord 검정, Peters 검정, 회귀분석에서의 절편, 분산 모수, 동질성, 이질성, 지수 통계, 표준 오차, *p* 값, 연구 가중치

메타분석 이분형 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 원시 데이터...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. **처리 그룹** 섹션에서 **성공** 변수를 선택하여 처리 그룹의 "성공" 계수를 표시하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
3. 처리 그룹의 "실패" 수를 표시하려면 **실패** 변수를 선택하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
4. **처리 그룹** 섹션에서 **성공** 변수를 선택하여 처리 그룹의 "성공" 계수를 표시하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
5. 제어 그룹의 "실패 수" 를 표시하려면 **실패** 변수를 선택하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
6. 선택적으로 **연구 ID** 및/또는 **연구 레이블** 변수를 선택하십시오. 선택한 **연구 ID**변수는 선택한 **연구 레이블** 변수와 같을 수 없습니다.
7. 선택적으로 **효과 크기** 설정을 선택하십시오. 사용 가능한 옵션은 **로그 오즈비**, **페토 로그 오즈비**, **로그 위험 비율** 및 **위험 차이**입니다.
8. 선택적으로 **모형** 설정을 선택하십시오. **자르기 및 채움** 설정이 사용 가능한 경우, 설정은 또한 자르기 및 채움 분석에서 풀링에 사용되는 모형을 제어합니다. **편향** 설정이 사용 가능한 경우, 설정은 회귀분석 기반 검정에서 사용하는 모형도 제어합니다.

   변량효과
   기본 설정은 변량효과 모형을 작성합니다.

   고정 효과
   고정 효과 모형을 작성합니다. **역분산**은 역분산 가중치를 추정합니다. **Mantel-Haenszel**은 Mantel-Haenszel 가중치를 추정합니다.
9. 선택적으로 다음을 수행할 수 있습니다.
   - **기준을 클릭하세요...** 를 사용하여 일반적인 기준을 지정할 수 있습니다.

- 하위 그룹 및 누적 분석을 지정하려면 **분석**을 클릭하십시오.
- 추정 방법을 지정하려면 **추론**을 클릭하십시오.
- 대비 테스트를 제어하려면 **대비**를 클릭하십시오.
- EGGER 회귀 기반 테스트를 수행하여 공개 바이어스에 액세스하려면 **바이어스**를 클릭하십시오.
- 게시 바이어스의 트림 및 채우기 분석을 구현하려면 **트림 및 채우기**를 클릭하십시오.
- 테이블 출력을 제어하려면 **인쇄**를 클릭하십시오.
- 활성 데이터 세트에 예상 통계를 저장하려면 **저장**을 클릭하십시오.
- 출력에 포함할 도표를 지정하려면 **플롯**을 누르십시오.

10. **확인**을 클릭하십시오.

메타분석 이분형: 기준

기준 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에 제공되는 원시 데이터에서 2진형 결과에서의 메타 분석에 대한 기준을 지정하기 위한 설정을 제공합니다.

신뢰구간
이 선택적 설정은 신뢰수준을 지정합니다. 값은 0 - 100 범위의 숫자 값이어야 합니다. 기본 설정은 95입니다.

결측 데이터 범위
이 선택적 설정은 프로시저의 결측 데이터 처리 방법을 제어합니다.

분석별 결측값 제외
기본 설정은 특정 분석 각각에 사용된 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

목록별 결측값 제외
프로시저에 지정된 모든 분석에서 사용되는 모든 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

사용자 결측값
이 선택적 설정은 사용자가 결측값을 처리하는 방법을 제어합니다.

제외
기본 설정은 사용자 결측값을 유효한 값으로 처리합니다.

포함(N)
사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 값으로 처리합니다.

반복

최대반복계산
이 선택적 설정은 반복적 방법의 최대반복수를 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 100입니다. 값 0은 반복계산을 수행하지 않음을 의미합니다.

최대 단계 이분
이 선택적 설정은 반복적 방법의 최대 단계 이분을 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 5입니다. 0 값은 단계 이분이 적용되지 않음을 의미합니다.

수렴
이 선택적 설정은 수렴허용을 지정합니다. 값은 단일 양수값이어야 합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 기본값은 1E-6입니다.

0값 빈도 수정
이 선택적 설정은 효과 크기 추정에서 0값 빈도 데이터(해당하는 경우)를 수정하는 방법을 제어합니다.

0값이 하나 이상 포함된 연구의 각 값만 조정
기본 설정에서는 0값 빈도 데이터만 조정합니다.

0값을 포함하는 연구가 하나 이상 있는 경우에만 모든 연구의 각 값 조정
연구에 0값 빈도가 한 개 이상 있으면 모든 빈도를 조정합니다.

모든 연구의 각 값 조정
0값 빈도가 있는지 에 상관없이 모든 데이터를 조정합니다.

조정하지 않고 0값 유지
데이터를 조정하지 않습니다.

추가 값
이 선택적 설정은 0값 빈도 데이터에 추가되는 값을 지정합니다. 기본값은 0.5입니다. 지정된 값은 0보다 크고 1보다 작거나 같은 단일 숫자 값이어야 합니다.

메타분석 이분형 기준 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 원시 데이터...
2. **메타 분석 2진형** 대화 상자에서 **기준**을 클릭하십시오.
3. 적합한 기준 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 분석

분석 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에 제공되는 원시 데이터에 대한 이진 결과를 사용하여 메타 분석을 위한 하위 그룹 및 누적 분석을 지정하기 위한 설정을 제공합니다.

부집단 분석
부집단 분석을 실행할 변수를 선택하십시오. 변수는 **누적 분석**에 정의된 변수와 같을 수 없습니다.

누적 분석
누적 분석을 실행하고 누적 메타분석이 수행될 변수를 선택하십시오. 변수는 **하위 그룹 분석**에 정의된 변수와 같을 수 없습니다. **오름차순**이 선택되면 누적 분석은 지정된 변수를 오름차순 기반으로 합니다. **내림차순**이 선택되면 누적 분석은 지정된 변수를 내림차순 기반으로 합니다.

누적 통계
추정된 누적 전체 효과 크기를 저장하는 옵션을 제공합니다. **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

누적 효과 크기
추정된 누적 전체 효과 크기를 저장합니다.

누적 효과 크기(지수 형식)
추정된 누적 전체 효과 크기를 지수 형식으로 저장합니다.

표준오차
누적 전체 효과 크기의 추정된 표준오차를 저장합니다.

신뢰구간 하한
누적 전체 효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
누적 전체 효과 크기의 추정된 신뢰구간 상한을 저장합니다.

신뢰구간 하한(지수 형식)
누적 전체 효과 크기의 추정된 신뢰구간 하한을 지수 형식으로 저장합니다.

신뢰구간 상한(지수 형식)
누적 전체 효과 크기의 추정된 신뢰구간 상한을 지수 형식으로 저장합니다.

p-값
누적 전체 효과 크기의 예상 *p*-값을 저장합니다.

대상
저장 데이터 세트 또는 데이터 파일을 지정하는 옵션을 제공합니다. **데이터 세트**를 선택하면 새 데이터 세트 이름을 지정할 수 있습니다(기본 데이터 세트 이름을 유지할 수 있음). **데이터 파일을** 선택한 후 **찾아보기...를** 클릭합니다. 을 클릭하여 저장 파일 이름과 위치를 선택합니다.

메타분석 이분형: 분석 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 원시 데이터...
2. **메타 분석 2진형** 대화 상자에서 **분석**을 클릭하십시오.
3. 적합한 분석 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 추론

추론 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에 제공되는 원시 데이터에 대한 2진 결과와 함께 메타 분석을 위한 추정 메소드를 지정하기 위한 설정을 제공합니다.

참고: 추론 대화 상자는 **랜덤 효과** 모델이 선택된 경우에만 사용 가능합니다.

추정량
추정량을 지정하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 추정량도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 추정량도 제어합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 표준오차 조정도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 표준오차 조정도 제어합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 이분형: 추론 설정 정의

1. 메뉴에서 다음을 선택합니다.

분석 > 메타 분석 > 2진형 결과 > 원시 데이터...

2. **메타 분석 2진형** 대화 상자에서 **추론**을 클릭하십시오.
3. 적합한 추론 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 대비

대비 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에서 제공되는 원시 데이터에 대한 연속형 결과에서의 메타 분석에 대한 대비 검정을 제어하기 위한 설정을 제공합니다.

변수
이 목록에는 사용 가능한 모든 데이터 세트 변수가 표시됩니다. 목록에서 변수를 선택하여 **대조** 목록으로 이동하십시오.

대비
이 목록은 활성 데이터 세트에 변수로 저장된 계수를 식별합니다. 다중 변수가 허용됩니다. 문자변수는 지원되지 않습니다.

사용자 입력 계수 값
사용자 제공 대비의 계수를 지정하는 설정을 제공합니다. 숫자 값만 허용됩니다. 유효한 대비검정을 공식화하려면 지정된 값의 개수가 유효한 연구 수와 매치해야 합니다.

지수 통계 표시
지수 형식 통계의 포함을 제어합니다. 이 설정을 선택할 경우 지수 효과 크기 및 신뢰구간 한계를 포함한 지수 형식 통계가 결과에 포함됩니다. **효과 크기**가 **로그 오즈비**, **Peto의 로그 오즈비** 또는 **로그 위험 비율**로 지정되면 설정을 사용할 수 있습니다.

메타분석 이분형: 대비 설정 정의

1. 메뉴에서 다음을 선택합니다.

분석 > 메타 분석 > 2진형 결과 > 원시 데이터...

2. **메타-분석 연속형** 대화 상자에서 **대비**를 클릭하십시오.
3. 적합한 대비 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 편향

편향 대화 상자는 효과 크기를 추정하기 위해 활성 데이터 세트에 제공된 원시 데이터에 대해 2진 결과를 사용하여 메타 분석에 대해 회귀 기반 테스트를 수행하여 공개 편향을 사용 가능하게 하는 설정을 제공합니다.

회귀분석 기반 검정
회귀분석 기반 검정을 지정하는 옵션을 제공합니다. 둘 이상의 검정을 선택할 수 있습니다.

Egger 검정
선택할 경우 Egger 검정이 수행됩니다.

Harbord 검정
선택할 경우 Harbord 검정이 수행됩니다. **효과 크기**가 **로그 오즈비** 또는 **로그 위험 비율**로 지정된 경우 테스트를 사용할 수 있습니다.

Peters 검정
선택할 경우 Peters 검정이 수행됩니다. **효과 크기** 가 **로그 오즈비**로 지정된 경우 테스트를 사용할 수 있습니다.

변수
이 목록은 사용 가능한 모든 데이터 세트 변수를 제공합니다.

공변량
변수 목록에서 선택된 변수는 공변량으로 처리됩니다. 다중 공변량이 허용됩니다.

요인
변수 목록에서 선택된 변수는 요인으로 취급됩니다. 다중 요인이 허용됩니다.

회귀분석에 절편 포함
회귀분석 기반 검정에서 절편 항을 제어합니다.

고정 효과 모형에 산포도 모수 포함
승법 모형 설정을 제어하고 승법 산포도 모수를 분석에 도입합니다. 이 설정은 고정 효과 모형을 선택한 경우에만 사용할 수 있습니다.

t 분포를 기준으로 통계량 추정
회귀분석 기반 검정에 사용된 분포를 제어합니다. 이 설정은 기본적으로 사용 가능하며, 이는 *t*-분포를 기반으로 하는 통계를 추정합니다. 이 설정을 선택하지 않으면 정규 분포를 기준으로 통계량이 추정됩니다.

메타분석 이분형: 편향 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 원시 데이터...
2. **메타 분석 2진형** 대화 상자에서 **편향**을 클릭하십시오.
3. 적합한 편향 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 자르기 및 채우기

자르기 및 채움 대화 상자는 효과 크기 추정을 위해 활성 데이터 세트에서 제공되는 원시 데이터에서 2진형 결과에서의 메타 분석에 대한 출판 편향의 자르기 및 채움 분석을 구현하기 위한 설정을 제공합니다.

결측 연구 수 추정
출판 편향의 자르기 및 채우기 분석을 제어합니다. 이 설정을 선택하면 다른 대화 상자 설정이 사용으로 설정됩니다.

대체 연구 측면
결측 연구가 대체된 깔때기 도표의 방향을 지정하는 옵션을 제공합니다.

Egger 검정의 기울기로 결정
기본 설정은 Egger 검정의 추정된 기울기를 기준으로 측면을 결정합니다.

왼쪽
깔때기 도표의 왼쪽을 대체합니다.

오른쪽
깔때기 도표의 오른쪽을 대체합니다.

방법
결측 연구 수 추정 방법을 지정합니다.

선형
기본 설정은 선형 추정량을 계산합니다.

다음을 실행하십시오.
실행 추정량을 계산합니다.

2차
2차 추정량을 계산합니다.

반복 프로세스
반복 추정량 및 표준오차 조정을 지정하는 설정을 제공합니다.

고정 효과 모형
선택할 경우, 고정 효과 모형이 사용되며 반복 추정 및 표준오차 조정 옵션은 사용할 수 없습니다.

변량효과 모형
선택할 경우, 변량효과 모형이 사용되며 다음 설정을 사용할 수 있습니다.

추정량
반복 추정량을 지정하는 설정을 제공합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
자르기 및 채우기 알고리즘의 반복에 Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 이분형: 자르기 및 채우기 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 원시 데이터...
2. **메타 분석 2진형** 대화 상자에서 **자르기 및 채움**을 클릭하십시오.
3. 적합한 자르기 및 채우기 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 인쇄

인쇄 대화 상자는 효과 크기를 추정하기 위해 활성 데이터 세트에 제공되는 원시 데이터에 대한 2진 결과를 사용하여 메타 분석을 위한 테이블 출력을 제어하기 위한 설정을 제공합니다.

동질성/이질성
동질성 및 이질성 검정을 제어하는 설정을 제공합니다.

동질성 검정
선택할 경우 해당 동질성 검정이 출력에 제공됩니다.

이질성 측도
선택할 경우 이질성 측도가 출력에 제공됩니다.

효과 크기
다음과 같은 효과 크기 설정을 제공합니다.

개별 연구
개별 연구 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다.

누적 효과 크기
누적 분석 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. **분석** 대화 상자에서 **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

변량효과 모형의 예측구간
예측구간의 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. 이 설정은 변량효과 모형을 지정한 경우에만 사용할 수 있습니다.

지수 통계 표시
지수 형식 통계의 포함을 제어합니다. 이 설정을 선택할 경우 지수 효과 크기 및 신뢰구간 한계를 포함한 지수 형식 통계가 결과입니다. 프로세스는 필수 및 선택적 효과 크기 추정 및 예측 통계표에 모두 적용됩니다. **효과 크기**가 **로그 오즈비**, **Peto의 로그 오즈비** 또는 **로그 위험 비율**로 지정되면 설정을 사용할 수 있습니다.

메타분석 이분형: 인쇄 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 원시 데이터...
2. **메타 분석 2진형** 대화 상자에서 **인쇄**를 클릭하십시오.
3. 적합한 인쇄 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 저장

저장 대화 상자에서는 효과 크기 추정을 위해 활성 데이터 세트에 제공된 원시 데이터의 2진형 결과를 사용하여 메타 분석을 위해 활성 데이터 세트에 추정된 통계를 저장하기 위한 설정을 제공합니다.

개별 연구
추정된 효과 크기를 저장하는 옵션을 제공합니다.

개별 효과 크기
추정된 효과 크기를 저장합니다.

개별 효과 크기(지수 형식)
추정된 효과 크기를 지수 형식으로 저장합니다.

표준오차
효과 크기의 추정된 표준오차를 저장합니다.

신뢰구간 하한
효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
효과 크기의 추정된 신뢰구간 상한을 저장합니다.

신뢰구간 하한(지수 형식)
효과 크기의 추정된 신뢰구간 하한을 지수 형식으로 저장합니다.

신뢰구간 상한(지수 형식)
효과 크기의 추정된 신뢰구간 상한을 지수 형식으로 저장합니다.

p-값
효과 크기의 예상 *p*-값을 저장합니다.

연구 가중값
추정된 연구 가중값을 저장합니다.

연구 가중값 백분율
표준화된 연구 가중값을 백분율로 저장합니다.

메타분석 이분형: 저장 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 원시 데이터...
2. **메타 분석 2진형** 대화 상자에서 **저장**을 클릭하십시오.
3. 적합한 추정된 통계 저장 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형: 도표

플롯 대화 상자는 다음 플롯 유형에 대한 설정을 제공합니다.

- 66 페이지의 『포리스트 도표』
- 67 페이지의 『누적 포리스트 도표』
- 68 페이지의 『거품 도표』
- 69 페이지의 『깔때기 도표』
- 70 페이지의 『Galbraith 도표』
- 71 페이지의 『L'Abb´e 도표』

메타분석 이분형 효과 크기

메타분석 이분형 효과 크기 프로시저는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공된 경우 이분형 결과를 사용하여 메타분석을 수행합니다.

예

제2형 당뇨병 치료에 도움이 되는 변덕스럽지만 논쟁의 여지가 있는 약품을 조사하기 위해 역사상 여러 연구가 수행되었습니다. 경구용 약품이 식사 후 혈당 수치를 낮출 수 있다고 주장되었습니다. 데이터는 1979년부터 1986년까지 여러 연구 사이트에서 수집되었습니다.

수석 연구자가 경구용 약품의 효과에 대한 통계적 추론을 도출하려고 합니다. 데이터가 여러 연구에서 생성되었기 때문에 연구 전반에 걸쳐 결과를 종합하여 효과를 전반적으로 이해하고 결과의 근본적인 변동 원인을 식별하는 아이디어를 제안했습니다.

통계

신뢰구간, 로그 오즈비, 페토의 로그 오즈비, 로그 위험 비율, 위험차, 랜덤 효과, 고정 효과, 역분산, Mantel-Haenszel, 반복, 단계적 이등분, 수렴, 누적 통계, 누적 효과 크기, 제한된 최대 우도, REML, 최대 우도, ML, Empirical Bayes, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, Egger 검정, Harbord 검정, Peters 검정, 회귀분석에서의 절편, 분산 모수, 동질성, 이질성, 지수 통계, 표준 오차, *p* 값, 연구 가중치

메타분석 이분형 효과 크기 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 효과 크기를 나타내는 **효과 크기** 변수를 선택하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
3. **표준 편차**를 선택하여 효과 크기 표준 편차를 지정하거나 **분산**을 선택하여 효과 크기 분산을 지정한 다음 효과 크기에 대한 표준 편차/분산을 나타내는 변수를 선택합니다.
4. 선택적으로 **연구 ID** 및/또는 **연구 레이블** 변수를 선택하십시오. 선택한 **연구 ID**변수는 선택한 **연구 레이블** 변수와 같을 수 없습니다.
5. 선택적으로 **효과 크기** 설정을 선택하십시오. 사용 가능한 옵션은 **로그 오즈비**, **페토 로그 오즈비**, **로그 위험 비율** 및 **위험 차이**입니다.
6. 선택적으로 **모형** 설정을 선택하십시오. **자르기 및 채움** 설정이 사용 가능한 경우, 설정은 또한 자르기 및 채움 분석에서 풀링에 사용되는 모형을 제어합니다. **편향** 설정이 사용 가능한 경우, 설정은 회귀분석 기반 검정에서 사용하는 모형도 제어합니다.

   변량효과
   기본 설정은 변량효과 모형을 작성합니다.

   고정 효과
   고정 효과 모형을 작성합니다.
7. 선택적으로 다음을 수행할 수 있습니다.

- **기준을 클릭하세요...** 를 사용하여 일반적인 기준을 지정할 수 있습니다.
- 하위 그룹 및 누적 분석을 지정하려면 **분석**을 클릭하십시오.
- 추정 방법을 지정하려면 **추론**을 클릭하십시오.
- 대비 테스트를 제어하려면 **대비**를 클릭하십시오.
- EGGER 회귀 기반 테스트를 수행하여 공개 바이어스에 액세스하려면 **바이어스**를 클릭하십시오.
- 게시 바이어스의 트림 및 채우기 분석을 구현하려면 **트림 및 채우기**를 클릭하십시오.
- 테이블 출력을 제어하려면 **인쇄**를 클릭하십시오.
- 활성 데이터 세트에 예상 통계를 저장하려면 **저장**을 클릭하십시오.
- 출력에 포함할 도표를 지정하려면 **플롯**을 누르십시오.

8. **확인**을 클릭하십시오.

메타분석 이분형 효과 크기: 기준

기준 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공될 때 메타 분석에 대한 기준을 2진형 결과로 지정하기 위한 설정을 제공합니다.

신뢰구간
이 선택적 설정은 신뢰수준을 지정합니다. 값은 0 - 100 범위의 숫자 값이어야 합니다. 기본 설정은 95입니다.

결측 데이터 범위
이 선택적 설정은 프로시저의 결측 데이터 처리 방법을 제어합니다.

분석별 결측값 제외
기본 설정은 특정 분석 각각에 사용된 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

목록별 결측값 제외
프로시저에 지정된 모든 분석에서 사용되는 모든 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함합니다.

사용자 결측값
이 선택적 설정은 사용자가 결측값을 처리하는 방법을 제어합니다.

제외
기본 설정은 사용자 결측값을 유효한 값으로 처리합니다.

포함(N)
사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 값으로 처리합니다.

반복

최대반복계산
이 선택적 설정은 반복적 방법의 최대반복수를 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 100입니다. 값 0은 반복계산을 수행하지 않음을 의미합니다.

최대 단계 이분
이 선택적 설정은 반복적 방법의 최대 단계 이분을 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 5입니다. 0 값은 단계 이분이 적용되지 않음을 의미합니다.

수렴
이 선택적 설정은 수렴허용을 지정합니다. 값은 단일 양수값이어야 합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 기본값은 1E-6입니다.

메타분석 이분형 효과 크기 기준 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...

2. **메타 분석 2진형 효과 크기** 대화 상자에서 **기준**을 클릭하십시오.

3. 적합한 기준 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 분석

분석 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 2진형 결과에 대한 메타 분석의 누적 분석 및 하위 그룹 지정에 대한 설정을 제공합니다.

부집단 분석
부집단 분석을 실행할 변수를 선택하십시오. 변수는 **누적 분석**에 정의된 변수와 같을 수 없습니다.

누적 분석
누적 분석을 실행하고 누적 메타분석이 수행될 변수를 선택하십시오. 변수는 **하위 그룹 분석**에 정의된 변수와 같을 수 없습니다. **오름차순**이 선택되면 누적 분석은 지정된 변수를 오름차순 기반으로 합니다. **내림차순**이 선택되면 누적 분석은 지정된 변수를 내림차순 기반으로 합니다.

누적 통계
추정된 누적 전체 효과 크기를 저장하는 옵션을 제공합니다. **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

누적 효과 크기
추정된 누적 전체 효과 크기를 저장합니다.

누적 효과 크기(지수 형식)
추정된 누적 전체 효과 크기를 지수 형식으로 저장합니다.

표준오차
누적 전체 효과 크기의 추정된 표준오차를 저장합니다.

신뢰구간 하한
누적 전체 효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
누적 전체 효과 크기의 추정된 신뢰구간 상한을 저장합니다.

신뢰구간 하한(지수 형식)
누적 전체 효과 크기의 추정된 신뢰구간 하한을 지수 형식으로 저장합니다.

신뢰구간 상한(지수 형식)
누적 전체 효과 크기의 추정된 신뢰구간 상한을 지수 형식으로 저장합니다.

p-값
누적 전체 효과 크기의 예상 *p*-값을 저장합니다.

대상
저장 데이터 세트 또는 데이터 파일을 지정하는 옵션을 제공합니다. **데이터 세트**를 선택하면 새 데이터 세트 이름을 지정할 수 있습니다(기본 데이터 세트 이름을 유지할 수 있음). **데이터 파일을** 선택한 후 **찾아보기...를** 클릭합니다. 을 클릭하여 저장 파일 이름과 위치를 선택합니다.

메타분석 이분형 효과 크기 분석 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...
2. **메타 분석 2진형 효과 크기** 대화 상자에서 **분석**을 클릭하십시오.
3. 적합한 분석 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 추론

추론 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공될 때 2진 결과와 함께 메타 분석을 위한 추정 메소드를 지정하기 위한 설정을 제공합니다.

추정량
추정량을 지정하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 추정량도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 추정량도 제어합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

노트:

- 자르기 및 채우기 설정이 지정된 경우 자르기 및 채우기 분석에서 통합에 사용되는 표준오차 조정도 제어합니다.
- 편향 설정이 지정된 경우 회귀분석 기반 검정에 사용되는 표준오차 조정도 제어합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 이분형 효과 크기 추론 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...
2. **메타 분석 2진형 효과 크기** 대화 상자에서 **추론**을 클릭하십시오.
3. 적합한 추론 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 대비

대비 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 2진 결과에 대한 메타 분석을 위한 대비 검정을 제어하는 설정을 제공합니다.

변수
이 목록에는 사용 가능한 모든 데이터 세트 변수가 표시됩니다. 목록에서 변수를 선택하여 **대조** 목록으로 이동하십시오.

대비
이 목록은 활성 데이터 세트에 변수로 저장된 계수를 식별합니다. 다중 변수가 허용됩니다. 문자변수는 지원되지 않습니다.

사용자 입력 계수 값
사용자 제공 대비의 계수를 지정하는 설정을 제공합니다. 숫자 값만 허용됩니다. 유효한 대비검정을 공식화하려면 지정된 값의 개수가 유효한 연구 수와 매치해야 합니다.

지수 통계 표시
지수 형식 통계의 포함을 제어합니다. 이 설정을 선택할 경우 지수 효과 크기 및 신뢰구간 한계를 포함한 지수 형식 통계가 결과에 포함됩니다. **효과 크기**가 **로그 오즈비**, **Peto의 로그 오즈비** 또는 **로그 위험 비율**로 지정되면 설정을 사용할 수 있습니다.

메타분석 이분형 효과 크기 대비 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...
2. **메타-분석 2진형 효과 크기** 대화 상자에서 **대비**를 클릭하십시오.
3. 적합한 대비 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 편향

편향 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 결과에서의 메타 분석에 대해 Eggers의 회귀분석 기반 검정을 수행하여 출판 편향을 사용하기 위한 설정을 제공합니다.

Egger의 회귀분석 기반 검정
이 설정을 선택하면 Egger의 회귀분석 기반 검정을 수행하여 출판 편향을 사용으로 설정합니다.

변수
이 목록은 사용 가능한 모든 데이터 세트 변수를 제공합니다.

공변량
변수 목록에서 선택된 변수는 공변량으로 처리됩니다. 다중 공변량이 허용됩니다.

요인
변수 목록에서 선택된 변수는 요인으로 취급됩니다. 다중 요인이 허용됩니다.

회귀분석에 절편 포함
회귀분석 기반 검정에서 절편 항을 제어합니다.

고정 효과 모형에 산포도 모수 포함
승법 모형 설정을 제어하고 승법 산포도 모수를 분석에 도입합니다. 이 설정은 고정 효과 모형을 선택한 경우에만 사용할 수 있습니다.

t 분포를 기준으로 통계량 추정
회귀분석 기반 검정에 사용된 분포를 제어합니다. 이 설정은 기본적으로 사용 가능하며, 이는 *t*-분포를 기반으로 하는 통계를 추정합니다. 이 설정을 선택하지 않으면 정규 분포를 기준으로 통계량이 추정됩니다.

메타분석 이분형 효과 크기 편향 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...
2. **메타-분석 2진형 효과 크기** 대화 상자에서 **편향**을 클릭하십시오.
3. 적합한 편향 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 자르기 및 채우기

자르기 및 채움 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공될 때 2진 결과를 사용하여 메타 분석에 대한 공개 편향의 자르기 및 채움 분석을 구현하기 위한 설정을 제공합니다.

결측 연구 수 추정
출판 편향의 자르기 및 채우기 분석을 제어합니다. 이 설정을 선택하면 다른 대화 상자 설정이 사용으로 설정됩니다.

대체 연구 측면
결측 연구가 대체된 깔때기 도표의 방향을 지정하는 옵션을 제공합니다.

Egger 검정의 기울기로 결정
기본 설정은 Egger 검정의 추정된 기울기를 기준으로 측면을 결정합니다.

왼쪽
깔때기 도표의 왼쪽을 대체합니다.

오른쪽
깔때기 도표의 오른쪽을 대체합니다.

방법
결측 연구 수 추정 방법을 지정합니다.

선형
기본 설정은 선형 추정량을 계산합니다.

다음을 실행하십시오.
실행 추정량을 계산합니다.

2차
2차 추정량을 계산합니다.

반복 프로세스
반복 추정량 및 표준오차 조정을 지정하는 설정을 제공합니다.

고정 효과 모형
선택할 경우, 고정 효과 모형이 사용되며 반복 추정 및 표준오차 조정 옵션은 사용할 수 없습니다.

변량효과 모형
선택할 경우, 변량효과 모형이 사용되며 다음 설정을 사용할 수 있습니다.

추정량
반복 추정량을 지정하는 설정을 제공합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
자르기 및 채우기 알고리즘의 반복에 Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 이분형 효과 크기 자르기 및 채우기 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...
2. **메타 분석 2진형 효과 크기** 대화 상자에서 **자르기 및 채움**을 클릭하십시오.
3. 적합한 자르기 및 채우기 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 인쇄

인쇄 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 2진 결과에서의 메타 분석에 대한 테이블 출력을 제어하기 위한 설정을 제공합니다.

동질성/이질성
동질성 및 이질성 검정을 제어하는 설정을 제공합니다.

동질성 검정
선택할 경우 해당 동질성 검정이 출력에 제공됩니다.

이질성 측도
선택할 경우 이질성 측도가 출력에 제공됩니다.

효과 크기
다음과 같은 효과 크기 설정을 제공합니다.

개별 연구
개별 연구 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다.

누적 효과 크기
누적 분석 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. **분석** 대화 상자에서 **누적 분석** 변수를 선택한 경우에만 설정을 사용할 수 있습니다.

변량효과 모형의 예측구간
예측구간의 표시를 제어합니다. 선택할 경우 해당 출력이 제공됩니다. 이 설정은 변량효과 모형을 지정한 경우에만 사용할 수 있습니다.

지수 통계 표시
지수 형식 통계의 포함을 제어합니다. 이 설정을 선택할 경우 지수 효과 크기 및 신뢰구간 한계를 포함한 지수 형식 통계가 결과입니다. 프로세스는 필수 및 선택적 효과 크기 추정 및 예측 통계표에 모두 적용됩니다. **효과 크기**가 **로그 오즈비**, **Peto의 로그 오즈비** 또는 **로그 위험 비율**로 지정되면 설정을 사용할 수 있습니다.

메타분석 이분형 효과 크기 인쇄 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...
2. **메타 분석 2진형 효과 크기** 대화 상자에서 **인쇄**를 클릭하십시오.
3. 적합한 인쇄 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 저장

저장 대화 상자는 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 2진 결과를 사용하여 메타 분석을 위해 활성 데이터 세트에 추정된 통계를 저장하기 위한 설정을 제공합니다.

개별 연구
추정된 효과 크기를 저장하는 옵션을 제공합니다.

개별 효과 크기(지수 형식)
추정된 효과 크기를 지수 형식으로 저장합니다.

표준오차
효과 크기의 추정된 표준오차를 저장합니다.

신뢰구간 하한
효과 크기의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
효과 크기의 추정된 신뢰구간 상한을 저장합니다.

신뢰구간 하한(지수 형식)
효과 크기의 추정된 신뢰구간 하한을 지수 형식으로 저장합니다.

신뢰구간 상한(지수 형식)
효과 크기의 추정된 신뢰구간 상한을 지수 형식으로 저장합니다.

p-값
효과 크기의 예상 *p*-값을 저장합니다.

연구 가중값
추정된 연구 가중값을 저장합니다.

연구 가중값 백분율
표준화된 연구 가중값을 백분율로 저장합니다.

메타분석 이분형 효과 크기 저장 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 2진형 결과 > 사전 계산된 효과 크기...
2. **메타-분석 2진형 효과 크기** 대화 상자에서 **저장**을 클릭하십시오.
3. 적합한 추정된 통계 저장 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 이분형 효과 크기: 도표

플롯 대화 상자는 다음 플롯 유형에 대한 설정을 제공합니다.

- 66 페이지의 『포리스트 도표』
- 67 페이지의 『누적 포리스트 도표』
- 68 페이지의 『거품 도표』
- 69 페이지의 『깔때기 도표』
- 70 페이지의 『Galbraith 도표』
- 71 페이지의 『L'Abb´e 도표』

메타분석 회귀분석

메타분석 회귀분석 프로시저는 메타 회귀분석을 수행합니다.

예
제2형 당뇨병 치료에 도움이 되는 변덕스럽지만 논쟁의 여지가 있는 약품을 조사하기 위해 역사상 여러 연구가 수행되었습니다. 경구용 약품이 식사 후 혈당 수치를 낮출 수 있다고 주장되었습니다. 데이터는 1979년부터 1986년까지 여러 연구 사이트에서 수집되었습니다.

수석 연구자가 경구용 약품의 효과에 대한 통계적 추론을 도출하려고 합니다. 데이터가 여러 연구에서 생성되었기 때문에 연구 전반에 걸쳐 결과를 종합하여 효과를 전반적으로 이해하고 결과의 근본적인 변동 원인을 식별하는 아이디어를 제안했습니다.

통계
신뢰수준, 반복적 방법, 단계 이분, 수렴허용, 표본 평균, 표본 분산, 표준편차, 추정된 효과 크기, 추정 방법, 회귀분석 기반 검정, 변량효과 모형, 고정 효과 모형, 산포도 모수, 제한최대우도 추정량, 경험적 베이즈 추정량, Hedges 추정량, Hunter-Schmidt 추정량, DerSimonian-Laird 추정량, Sidik-Jonkman 추정량, Knapp-Hartung 표준오차 조정, 잘린 Knapp-Hartung 표준오차 조정, 계수, 모형 계수 검정, 지수 통계

메타분석 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 메타 회귀분석

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 효과 크기를 나타내는 단일 종속 **효과 크기** 변수를 선택하십시오. 선택한 변수는 숫자여야 합니다. 문자변수는 지원되지 않습니다.
3. 다음 설정 중 하나를 선택한 후 해당하는 단일 숫자 변수를 선택합니다.

   표준오차
   가중값으로 변환될 표준오차를 지정하는 변수를 선택합니다. 기본 설정입니다.

   분산
   가중값으로 변환될 분산을 지정하는 변수를 선택합니다.

   가중치
   가중값을 지정하는 변수를 선택합니다.
4. 선택적으로, 요인 변수를 **요인** 목록에 추가하십시오. 선택된 각 요인 변수의 경우 각 변수에 대해 선택적 값을 지정하여 사용자 정의 **마지막 범주**를 지정할 수 있습니다. **재설정**을 클릭하여 **마지막 범주** 값을 기본 설정으로 복원할 수 있습니다.

   참고: 지정된 **마지막 범주** 값과 일치하는 케이스가 없는 경우, 마지막으로 발생한 값이 마지막 범주로 처리됩니다.
5. 선택적으로 숫자 공변량 변수를 선택합니다.
6. 선택적으로 **모형** 설정을 선택하십시오.

   변량효과
   기본 설정은 변량효과 모형을 작성합니다.

   고정 효과
   고정 효과 모형을 작성합니다. 선택적으로 **분산 모수 포함** 설정을 선택할 수 있습니다.
7. 선택적으로 다음을 수행할 수 있습니다.
   - **기준을 클릭하세요...** 를 사용하여 일반적인 기준을 지정할 수 있습니다.
   - 추정 방법을 지정하려면 **추론**을 클릭하십시오.
   - 테이블 출력을 제어하려면 **인쇄**를 클릭하십시오.
   - 예상 통계를 예측한 후 활성 데이터 세트에 저장하려면 **저장**을 클릭하십시오.
   - 출력에 포함할 도표를 지정하려면 **플롯**을 누르십시오.
8. **확인**을 클릭하십시오.

메타분석 회귀분석: 기준

기준 대화 상자는 메타 분석 회귀분석의 기준을 지정하기 위한 설정을 제공합니다.

신뢰구간
이 선택적 설정은 신뢰수준을 지정합니다. 값은 0 - 100 범위의 숫자 값이어야 합니다. 기본 설정은 95입니다.

사용자 결측값
이 선택적 설정은 사용자가 결측값을 처리하는 방법을 제어합니다.

제외
기본 설정은 사용자 결측값을 유효한 값으로 처리합니다.

포함(N)
사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 값으로 처리합니다.

반복

최대반복계산
이 선택적 설정은 반복적 방법의 최대반복수를 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 100입니다. 값 0은 반복계산을 수행하지 않음을 의미합니다.

최대 단계 이분
이 선택적 설정은 반복적 방법의 최대 단계 이분을 지정합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 값은 단일 양의 정수여야 합니다. 기본값은 5입니다. 0 값은 단계 이분이 적용되지 않음을 의미합니다.

수렴
이 선택적 설정은 수렴허용을 지정합니다. 값은 단일 양수값이어야 합니다. 이 설정은 반복적 방법을 사용할 경우에 사용 가능합니다. 기본값은 1E-6입니다.

메타분석 회귀분석 기준 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 메타 회귀분석
2. **메타 분석 회귀분석** 대화 상자에서 **기준**을 클릭하십시오.
3. 적합한 기준 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 회귀분석: 추론

추론 대화 상자는 메타 분석 회귀분석을 위한 추정 방법을 지정하기 위한 설정을 제공합니다.

회귀분석에 절편 포함
이 설정은 기본적으로 사용으로 설정됩니다.

t 분포를 기준으로 통계량 추정
회귀분석 기반 검정에 사용된 분포를 제어합니다. 이 설정은 기본적으로 사용 가능하며, 이는 *t*-분포를 기반으로 하는 통계를 추정합니다. 이 설정을 선택하지 않으면 정규 분포를 기준으로 통계량이 추정됩니다.

추정량
추정량을 지정하는 설정을 제공합니다.

제한최대우도(REML)
기본 설정은 반복적 방법을 적용하며 제한최대우도 추정량을 계산합니다.

최대우도(ML)
반복적 방법을 적용하며 최대우도 추정량을 계산합니다.

경험적 베이즈
반복적 방법을 적용하며 경험적 베이즈 추정량을 계산합니다.

Hedges
비반복적 방법을 적용하며 Hedges 추정량을 계산합니다.

Hunter-Schmidt
비반복적 방법을 적용하며 Hunter-Schmidt 추정량을 계산합니다.

DerSimonian-Laird
비반복적 방법을 적용하며 DerSimonian-Laird 추정량을 계산합니다.

Sidik-Jonkman
비반복적 방법을 적용하며 Sidik-Jonkman 추정량을 계산합니다.

표준오차 조정
Knapp-Hartung 표준오차 조정을 적용할지 여부를 제어하는 설정을 제공합니다.

수정 없음
기본 설정은 수정을 적용하지 않습니다.

Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용합니다.

잘린 Knapp-Hartung 수정 적용
Knapp-Hartung 수정 방법을 적용하며 분산-공분산 행렬을 추정할 때 1보다 작으면 값을 자릅니다.

메타분석 회귀분석 추론 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 메타 회귀분석
2. **메타 분석 회귀분석** 대화 상자에서 **추론**을 클릭하십시오.
3. 적합한 추론 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 회귀분석: 인쇄

인쇄 대화 상자는 메타 분석 회귀분석에 대한 테이블 출력을 제어하기 위한 설정을 제공합니다.

모형 계수 검정
모형 계수 검정을 제어합니다. 이 설정은 기본적으로 지정되지 않으며, 이 경우 검정이 표시되지 않습니다. 이 설정을 지정할 경우 검정이 출력에 제공됩니다.

지수 통계 표시
모수 추정값을 제어합니다. 이 설정은 기본적으로 지정되지 않으며, 이 경우 모수 추정값이 표시되지 않습니다. 이 설정을 지정할 경우 모수 추정값이 출력에 제공됩니다.

메타분석 회귀분석 인쇄 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 메타 회귀분석
2. **메타 분석 회귀분석** 대화 상자에서 **인쇄**를 클릭하십시오.
3. 적합한 인쇄 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 회귀분석: 저장

저장 대화 상자에서는 예측된 통계를 메타 분석 회귀분석의 활성 데이터 세트에 저장하기 위한 설정을 제공합니다.

개별 연구
추정된 효과 크기를 저장하는 옵션을 제공합니다.

예측값
적합값을 예측하고 저장합니다.

예측값의 표준오차
적합값의 표준오차를 예측하고 저장합니다.

신뢰구간 하한
적합값의 추정된 신뢰구간 하한을 저장합니다.

신뢰구간 상한
적합값의 추정된 신뢰구간 상한을 저장합니다.

잔차
잔차를 예측하고 저장합니다.

잔차의 표준오차
잔차의 표준오차를 예측하고 저장합니다.

레버리지
레버리지를 예측하고 저장합니다.

누적 통계
선형 예측 저장 옵션을 제공합니다.

참고: 고정 효과가 **모델** 설정으로 선택되면 누적 통계 설정이 무시됩니다.

고정 선형 예측
선형 예측을 예측하고 저장합니다.

고정 선형 예측의 표준오차
선형 예측의 표준오차를 예측하고 저장합니다.

최적 선형 비편향 예측
변량효과의 최적 선형 비편향 예측을 예측하고 저장합니다.

BLUP의 표준오차
변량효과의 최적 선형 비편향 예측(BLUP)의 표준오차를 예측하고 저장합니다.

메타분석 회귀분석 저장 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 메타 회귀분석
2. **메타-분석 회귀분석** 대화 상자에서 **저장**을 클릭하십시오.
3. 적합한 저장 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

메타분석 회귀분석: 도표

플롯 대화 상자는 거품 도표를 정의하기 위한 설정을 제공합니다. 추가 정보는 68 페이지의 『거품 도표』의 내용을 참조하십시오.

메타분석 도표화 옵션

다음 플롯 옵션은 메타 분석 프로시저에 사용할 수 있으며 **플롯** 대화 상자에서 액세스할 수 있습니다.

포리스트 도표

포리스트 플롯 탭은 원시 데이터 및 사전 계산된 효과 크기 데이터가 활성 데이터 세트에서 제공되는 경우 연속형 및 2진형 결과와 함께 메타 분석의 결과에 표시되는 포리스트 플롯 차트를 제어하기 위한 설정을 제공합니다.

포리스트 도표
이 선택적 설정은 포리스트 도표 출력을 사용 및 사용 안함으로 설정합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

표시 열
이 선택적 설정은 표로 작성된 추정 통계량을 지정합니다.

효과 크기
추정된 효과 크기가 포함됩니다.

표준오차
추정된 효과 크기의 표준오차가 포함됩니다.

신뢰구간 한계
추정된 효과 크기의 신뢰구간 하한과 상한이 포함됩니다.

p-값
각 개별 연구에 대한 예상 효과 크기의 *p*값을 포함합니다.

가중치
1차 연구의 가중값이 포함됩니다.

지수 형식 표시

참고: 이 설정은 이분형 결과를 사용하는 메타분석에만 적용됩니다.

효과 크기와 신뢰 구간의 표시 및 도표 열을 제어합니다. 이 설정을 선택할 경우 오즈비, Peto의 오즈비 또는 위험도 비율이 도표로 작성되고 지수 형식 통계가 표로 작성됩니다. 이 설정을 선택하지 않을 경우 로그 오즈비, Peto의 로그 오즈비 또는 로그 위험도 비율이 도표로 작성되고 로그 변환 통계가 표로 작성됩니다.

변수
사용 가능한 데이터 세트 변수가 나열됩니다.

추가 열
선택적으로 추가 열을 표시할 변수를 선택합니다. 변수 순서에 따라 열 표시 순서가 결정됩니다.

정렬
선택적으로 포리스트 도표를 정렬할 기준 변수를 지정합니다. 변수가 지정되면 포리스트 플롯이 **오름차순**(기본 설정)으로 정렬됩니다. **내림차순**이 선택되면 포리스트 플롯이 내림차순으로 정렬됩니다.

도표 열의 위치
도표 열의 배치를 제어합니다.

오른쪽
기본 설정은 도표 열을 다른 표 열의 오른쪽에 배치합니다.

왼쪽
도표 열을 다른 표 열의 왼쪽에 배치합니다.

참조선
포리스트 도표에 추가되는 참조선을 제어합니다. **전체 효과 크기**는 예상 전체 효과 크기를 나타내는 행을 추가합니다. **널 효과 크기**는 널 효과 크기를 표시하는 행을 추가합니다.

주석
주석 표시를 제어합니다.

동질성
동질성 검정 통계를 출력합니다.

이질성
이질성 검정 통계를 출력합니다.

테스트
전체 효과 크기 검정과 부집단 간 동질성 검정(부집단 분석이 있을 경우)을 출력합니다.

자르기 범위
자르기 범위를 지정합니다. 이 설정을 선택할 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다.

이분형 결과를 사용하는 메타분석의 경우, **효과 크기** 설정이 **위험도 차이**로 설정되어 있으면 `Upper bound ≤ 1` 및 `Lower bound ≥ −1` 조건도 충족해야 합니다. **지수 형식 표시** 설정을 선택한 경우 `Lower bound ≥ 0`도 만족해야 합니다.

포리스트 도표 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 지속적인 산출 또는 2진형 결과 > 원시 데이터 ... 또는 **미리 계산된 효과 크기...**
2. **메타 분석** 대화 상자에서 **플롯**을 클릭한 후 **포리스트 플롯** 탭을 클릭하십시오.
3. 적합한 포리스트 도표 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

누적 포리스트 도표

누적 포리스트 플롯 탭은 원시 데이터 및 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 및 2진형 결과와 함께 메타 분석의 출력에 표시되는 누적 포리스트 플롯 차트를 제어하기 위한 설정을 제공합니다.

누적 포리스트 도표
이 선택적 설정은 포리스트 도표 출력을 사용 및 사용 안함으로 설정합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

표시 열
이 선택적 설정은 표로 작성된 추정 통계량을 지정합니다.

누적 효과 크기
누적 전체 효과 크기가 포함됩니다.

표준오차
누적 전체 효과 크기의 표준오차가 포함됩니다.

신뢰구간 한계
누적 전체 효과 크기의 신뢰구간 하한과 상한이 포함됩니다.

p-값
각 개별 연구에 대한 누적 전체 효과 크기의 *p*-값을 포함합니다.

지수 형식 표시

참고: 이 설정은 이분형 결과를 사용하는 메타분석에만 적용됩니다.

효과 크기와 신뢰 구간의 표시 및 도표 열을 제어합니다. 이 설정을 선택할 경우 오즈비, Peto의 오즈비 또는 위험도 비율이 도표로 작성되고 지수 형식 통계가 표로 작성됩니다. 이 설정이 사용으로 설정되지 않은 경우 로그 오즈비, Peto의 로그 오즈비 또는 로그 위험도 비율이 도표로 작성되고 로그 변환 통계가 표로 작성됩니다.

변수
사용 가능한 데이터 세트 변수가 나열됩니다.

추가 열
선택적으로 추가 열을 표시할 변수를 선택합니다. 변수 순서에 따라 열 표시 순서가 결정됩니다.

도표 열의 위치
도표 열의 배치를 제어합니다.

오른쪽
기본 설정은 도표 열을 다른 표 열의 오른쪽에 배치합니다.

왼쪽
도표 열을 다른 표 열의 왼쪽에 배치합니다.

자르기 범위
자르기 범위를 지정합니다. 이 설정을 선택할 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다.

이분형 결과를 사용하는 메타분석의 경우, **효과 크기** 설정이 **위험도 차이**로 설정되어 있으면 `Upper bound ≤ 1` 및 `Lower bound ≥ −1` 조건도 충족해야 합니다. **지수 형식 표시** 설정을 선택한 경우 `Lower bound ≥ 0`도 만족해야 합니다.

누적 포리스트 도표 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 지속적인 산출 또는 2진형 결과 > 원시 데이터 ... 또는 **미리 계산된 효과 크기...**
2. **메타 분석** 대화 상자에서 **플롯**을 클릭한 후 **누적 포리스트 플롯** 탭을 클릭하십시오.
3. 적합한 누적 포리스트 도표 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

거품 도표

거품 도표 탭은 원시 데이터 및 사전 계산된 효과 크기 데이터가 활성 데이터 세트에서 제공되는 경우 연속형 및 2진형 결과에 대한 메타 분석 출력에 표시되는 거품 도표 차트를 제어하기 위한 설정을 제공합니다.

참고: 거품 도표는 메타분석 회귀분석에 사용할 수 있는 유일한 도표 유형입니다.

거품 도표
이 선택적 설정은 거품 도표 출력을 사용 및 사용 안함으로 설정합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

참고: 이 설정은 메타분석 회귀분석에 사용할 수 없습니다.

변수
사용 가능한 데이터 세트 변수가 나열됩니다.

예측자
X축에서 연속 예측자로 처리될 변수를 지정합니다. 다중 변수가 허용되며 각각은 별도의 거품 도표를 작성합니다.

참고: 메타 분석 회귀 분석의 경우 이 설정의 레이블이 **거품 도표 예측변수**이며 거품 도표를 작성할 공변량을 선택할 수 있습니다.

평균 중심화 예측자
이 선택적 설정은 X축의 연속 예측자를 제어합니다. 이 설정은 기본적으로 사용 안함으로 설정되며, 이 경우 예측자가 중앙에 배치되지 않습니다. 이 설정이 사용으로 설정된 경우 예측자가 평균 중심화됩니다.

레이블
선택적으로 거품 도표에 레이블을 지정할 변수를 지정합니다. **배치** 목록은 레이블을 자동(**자동**)으로 또는 **오른쪽**, **왼쪽**, **위쪽** 및 **맨 아래**에 배치하기 위한 옵션을 제공합니다.

가중값에 비례하여 거품 그리기
가중값과 관련하여 거품 그리기 방식을 제어합니다. 이 설정은 기본적으로 사용으로 설정되며, 이 경우 가중값에 비례하여 거품을 그립니다. 이 설정이 사용 안함으로 설정된 경우 모든 거품을 동일한 크기로 그립니다.

적합선 표시
적합 회귀선의 표시를 제어합니다. 이 설정은 기본적으로 사용으로 설정되며, 이 경우 적합 회귀선이 거품 도표에 추가됩니다. 이 설정을 사용 안함으로 설정할 경우 적합 회귀선이 거품 도표에 추가되지 않습니다.

신뢰구간 한계 표시
신뢰구간 한계 표시를 제어합니다. 이 설정은 기본적으로 사용으로 설정되며, 이 경우 신뢰구간 한계가 거품 도표에 추가됩니다. 이 설정을 사용 안함으로 설정할 경우 신뢰구간 한계가 거품 도표에 추가되지 않습니다.

X축의 범위
X축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다. 지정된 값은 모든 거품 도표에 적용됩니다.

Y축의 범위
Y축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다. 지정된 값은 모든 거품 도표에 적용됩니다.

이분형 결과를 사용하는 메타분석의 경우, **효과 크기** 설정이 **위험도 차이**로 설정되어 있으면 `Upper bound ≤ 1` 및 `Lower bound ≥ −1` 조건도 충족해야 합니다.

거품 도표 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 지속적인 산출 또는 2진형 결과 > 원시 데이터 ... 또는 **미리 계산된 효과 크기...**

   또는

   분석 > 메타 분석 > 메타 회귀분석
2. **메타 분석** 대화 상자에서 **플롯**을 클릭한 후 **거품 도표** 탭을 클릭하십시오.
3. 적합한 거품 도표 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

깔때기 도표

퍼널 도표 탭은 원시 데이터 및 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속형 및 2진형 결과와 함께 메타 분석의 출력에 표시되는 퍼널 도표 차트를 제어하기 위한 설정을 제공합니다.

깔때기 도표
이 선택적 설정은 깔때기 도표 출력을 사용 및 사용 안함으로 설정합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

Y축 값
Y축 값을 제어합니다.

표준오차
표준오차를 도표로 작성합니다.

역 표준오차
역 표준오차를 도표로 작성합니다.

분산
분산을 도표로 작성합니다.

역 분산
역 분산을 도표로 작성합니다.

X축의 범위
X축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다. 지정된 값은 모든 깔때기 도표에 적용됩니다.

이분형 결과를 사용하는 메타분석의 경우, **효과 크기** 설정이 **위험도 차이**로 설정되어 있으면 `Upper bound ≤ 1` 및 `Lower bound ≥ −1` 조건도 충족해야 합니다.

Y축의 범위
Y축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound ≥ 0` 조건을 충족해야 합니다. 지정된 값은 모든 깔때기 도표에 적용됩니다.

자르기 및 채우기 관련 대체 연구 포함
트림 및 채우기 대화 상자에서 **누락된 연구 수 추정** 설정이 사용으로 설정된 경우 대치된 연구가 퍼널 도표에 포함되는지 여부를 제어합니다. 이 설정은 기본적으로 사용 안함으로 설정되며, 이 경우 대체 연구가 제외됩니다. **관찰된 연구의 전체 효과 크기 표시**가 선택되면, 추정된 전체 효과 크기를 나타내는 수직 참조 라인이 추가됩니다. 이 설정은 관측 연구와 대체 연구를 모두 고려합니다.

변수
사용 가능한 데이터 세트 변수가 나열됩니다.

레이블
선택적으로 깔때기 도표에 레이블을 지정할 변수를 지정합니다. **배치** 목록은 레이블을 자동(**자동**)으로 또는 **오른쪽**, **왼쪽**, **위쪽** 및 **맨 아래**에 배치하기 위한 옵션을 제공합니다.

깔때기 도표 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 지속적인 산출 또는 2진형 결과 > 원시 데이터 ... 또는 **미리 계산된 효과 크기...**
2. **메타 분석** 대화 상자에서 **플롯**을 클릭한 후 **퍼널 도표** 탭을 클릭합니다.
3. 적합한 깔때기 도표 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

Galbraith 도표

Galbraith Plot 탭은 원시 데이터 및 사전 계산된 효과 크기 데이터가 활성 데이터 세트에서 제공되는 경우 연속형 및 2진형 결과와 함께 메타 분석의 출력에 표시되는 Galbraith 플롯 차트를 제어하기 위한 설정을 제공합니다.

Galbraith 도표
이 선택적 설정은 거품 도표 출력을 사용 및 사용 안함으로 설정합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

신뢰구간 한계 표시
신뢰구간 한계 표시를 제어합니다. 이 설정은 기본적으로 사용으로 설정되며, 이 경우 신뢰구간 한계가 거품 도표에 추가됩니다. 이 설정을 사용 안함으로 설정할 경우 신뢰구간 한계가 거품 도표에 추가되지 않습니다.

X축의 범위
X축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound ≥ 0` 조건을 충족해야 합니다. 지정된 값은 모든 Galbraith 도표에 적용됩니다.

Y축의 범위
Y축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다. 지정된 값은 모든 Galbraith 도표에 적용됩니다.

변수
사용 가능한 데이터 세트 변수가 나열됩니다.

레이블
선택적으로 거품 도표에 레이블을 지정할 변수를 지정합니다. **배치** 목록은 레이블을 자동(**자동**)으로 또는 **오른쪽**, **왼쪽**, **위쪽** 및 **맨 아래**에 배치하기 위한 옵션을 제공합니다.

Galbraith 도표 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 메타 분석 > 지속적인 산출 또는 2진형 결과 > 원시 데이터 ... 또는 **미리 계산된 효과 크기...**
2. **메타 분석** 대화 상자에서 **플롯**을 클릭한 후 **Galbraith Plot** 탭을 클릭하십시오.
3. 적합한 Galbraith 도표 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

L'Abb´e 도표

L'Abb'e 플롯 탭은 원시 데이터 및 사전 계산된 효과 크기 데이터가 활성 데이터 세트에 제공되는 경우 연속 및 2진형 결과와 함께 메타 분석의 출력에 표시되는 L'Abb'e 플롯 차트를 제어하기 위한 설정을 제공합니다.

L'Abb´e 도표
이 선택적 설정은 L'Abb´e 도표 출력을 사용 및 사용 안함으로 설정합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

가중값에 비례하여 점 그리기
가중값과 관련하여 점 그리기 방식을 제어합니다. 이 설정은 기본적으로 사용으로 설정되며, 이 경우 가중값에 비례하여 점을 그립니다. 이 설정이 사용 안함으로 설정된 경우 모든 점을 동일한 크기로 그립니다.

참조선

전체 효과 크기
추정된 전체 효과 크기를 나타내는 참조선의 표시를 제어합니다. 이 설정은 기본적으로 사용 안함으로 설정되며, 이 경우 참조선이 표시되지 않습니다. 이 설정이 사용으로 설정된 경우 참조선이 표시됩니다.

영 효과 크기
효과 없음을 나타내는 참조선의 표시를 제어합니다. 이 설정은 기본적으로 사용으로 설정되며, 이 경우 참조선이 표시됩니다. 이 설정이 사용으로 설정되지 않은 경우 참조선이 표시되지 않습니다.

X축의 범위
X축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다. 지정된 값은 모든 Galbraith 도표에 적용됩니다.

이분형 결과를 사용하는 메타분석의 경우, **효과 크기** 설정이 **위험도 차이**로 설정되어 있으면 `Lower bound ≥ 0` 및 `Upper bound ≤ 1` 조건도 충족해야 합니다. **효과 크기** 설정이 **로그 위험도 비율**로 설정된 경우 `Upper bound ≤ 0` 조건을 충족해야 합니다.

Y축의 범위
Y축 도표 범위를 지정합니다. 이 설정이 사용으로 설정된 경우 두 개의 숫자 값을 지정해야 하며 이 값은 `Upper bound > Lower bound` 조건을 충족해야 합니다. 지정된 값은 모든 Galbraith 도표에 적용됩니다.

이분형 결과를 사용하는 메타분석의 경우, **효과 크기** 설정이 **위험도 차이**로 설정되어 있으면 `Lower bound ≥ 0` 및 `Upper bound ≤ 1` 조건도 충족해야 합니다. **효과 크기** 설정이 **로그 위험도 비율**로 설정된 경우 `Upper bound ≤ 0` 조건을 충족해야 합니다.

변수
사용 가능한 데이터 세트 변수가 나열됩니다.

레이블
선택적으로 거품 도표에 레이블을 지정할 변수를 지정합니다. **배치** 목록은 레이블을 자동(**자동**)으로 또는 **오른쪽**, **왼쪽**, **위쪽** 및 **맨 아래**에 배치하기 위한 옵션을 제공합니다.

L'Abb´e 도표 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > **메타 분석** > **지속적인 산출 또는 2진형 결과** > **원시 데이터 ...**
2. **메타 분석** 대화 상자에서 **플롯**을 클릭한 후 **L'Abb´e 플롯** 탭을 클릭하십시오.
3. 적합한 L'Abb´e 도표 설정을 선택하고 정의합니다.
4. **계속**을 클릭합니다.

코드북

코드북은 변수 이름, 변수 레이블, 값 레이블, 결측값과 같은 사전 정보와 모든 또는 지정된 변수의 요약 통계 및 활성 데이터 세트의 다중 응답 세트를 보고합니다. 명목 및 순서 변수 및 다중 응답 세트의 경우 요약 통계에 개수 및 퍼센트가 포함됩니다. 척도변수의 경우 요약 통계에 평균, 표준 편차 및 사분위수가 포함됩니다.

참고: 코드북은 분할 파일 상태를 무시합니다. 여기에는 결측값의 다중대체를 위해 만든 분할 파일 집단이 포함됩니다(결측값 추가 기능 옵션에서 사용 가능).

코드북 가져오기

1. 메뉴에서 다음을 선택합니다.

   분석 > **보고서** > **코드북**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 변수 탭을 클릭합니다.
3. 하나 이상의 변수 및/또는 다중 응답 세트를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 표시된 변수 정보를 제어합니다.
- 표시된 통계를 제어하거나 모든 요약 통계를 제외합니다.
- 변수 및 다중 응답 세트가 표시되는 순서를 제어합니다.
- 표시된 요약 통계를 변경하기 위해 소스 목록에서 변수의 측정 수준을 변경합니다. 자세한 정보는 74 페이지의 『코드북 통계 탭』 주제를 참조하십시오.

측정 수준 변경

일시적으로 변수의 측정 수준을 변경할 수 있습니다. 다중 응답 세트의 측정 수준은 변경할 수 없습니다. 항상 명목변수로 취급됩니다.

1. 소스 목록에서 변수를 마우스 오른쪽 단추로 클릭합니다.
2. 팝업 메뉴에서 측정 수준을 선택합니다.

이렇게 하면 측정 수준이 일시적으로 변경됩니다. 실제 항목에서 측정 수준은 숫자 변수로만 유용합니다. 문자열 변수는 명목 또는 순서로 제한되며, 명목과 순서 모두 코드북 프로시저에 의해 동일하게 처리됩니다.

코드북 결과 탭

결과 탭은 각 변수 및 다중 응답 세트에 대해 포함된 변수 정보, 변수 및 다중 응답 세트가 표시되는 순서 및 선택적 파일 정보 테이블의 내용을 제어합니다.

변수 정보

각 변수에 대해 표시되는 사전 정보를 제어합니다.

위치. 파일 순서대로 변수 위치를 표시하는 정수입니다. 이것은 다중 응답 세트에서는 사용할 수 없습니다.

레이블. 변수 또는 다중 응답 세트와 연결된 설명 레이블입니다.

유형. 기본 데이터 유형입니다. 숫자, 문자 또는 다중 응답 세트 중 하나입니다.

형식. *A4*, *F8.2* 또는 *DATE11*과 같이 변수의 표시 형식입니다. 이것은 다중 응답 세트에서는 사용할 수 없습니다.

측정 수준. 가능한 값은 명목, 순서, 척도 및 알 수 없음입니다. 표시된 값은 사전에 저장된 측정 수준이며 변수 탭의 소스 변수 목록에서 측정 수준을 변경하더라도 지정된 임시 측정 수준 재지정에 의해 영향을 받지 않습니다. 이것은 다중 응답 세트에서는 사용할 수 없습니다.

참고: 숫자변수의 측정 수준은 외부 소스 또는 새롭게 작성한 변수에서 데이터를 읽는 것과 같이 측정 수준을 명시적으로 설정하지 않은 경우 첫 번째 데이터가 전달되기 전에 "알 수 없는" 상태일 수 있습니다. 자세한 정보는 주제를 참조하십시오.

역할. 일부 대화 상자는 정의된 역할을 기준으로 분석에 사용되는 변수를 미리 선택할 수 있는 기능을 제공합니다.

값 레이블. 관련 데이터 값과 연결된 설명 레이블입니다.

- 통계 탭에서 개수 또는 퍼센트를 선택할 경우 여기서 값 레이블을 선택하지 않더라도 정의된 값 레이블이 결과에 포함됩니다.
- 다중 이분 세트의 경우 "값 레이블"은 정의 방법에 따라 변수 세트의 기본 변수에 대한 변수 레이블 또는 계산된 값의 레이블입니다. 자세한 정보는 주제를 참조하십시오.

결측값. 사용자 정의 결측값. 통계 탭에서 개수 또는 퍼센트를 선택할 경우 여기서 결측값을 선택하지 않더라도 정의된 값 레이블이 결과에 포함됩니다. 이것은 다중 응답 세트에서는 사용할 수 없습니다.

사용자 정의 속성. 사용자가 정의한 사용자 정의 변수 속성입니다. 결과에는 각 변수와 연결된 모든 사용자 정의 변수 속성에 대한 이름 및 값이 모두 포함됩니다. 자세한 정보는 주제를 참조하십시오. 이것은 다중 응답 세트에서는 사용할 수 없습니다.

예약된 속성. 예약된 시스템 변수 속성입니다. 시스템 속성을 표시할 수 있지만 변경하지 않아야 합니다. 시스템 속성 이름은 달러 부호($)로 시작합니다. "@" 또는 "$@"로 시작하는 이름을 가진 비표시 속성은 포함되지 않습니다. 결과에는 각 변수와 연결된 모든 시스템 속성에 대한 이름 및 값이 모두 포함됩니다. 이것은 다중 응답 세트에서는 사용할 수 없습니다.

파일 정보

선택적 파일 정보 테이블에는 다음 파일 속성이 포함될 수 있습니다.

파일 이름. IBM SPSS 통계 데이터 파일의 이름입니다. 데이터 세트가 IBM SPSS 통계 형식으로 저장되지 않은 경우 데이터 세트 이름이 없습니다. 데이터 편집기 창의 제목 막대에 파일 이름이 표시되지 않은 경우 활성 데이터 세트의 파일 이름이 없습니다.

위치. IBM SPSS 통계 데이터 파일의 디렉토리(폴더) 위치입니다. 데이터 세트가 IBM SPSS 통계 형식으로 저장되지 않은 경우 위치가 없습니다.

케이스 수. 활성 데이터 세트의 케이스 수입니다. 이 값은 조건 필터링으로 인해 요약 통계에서 제외될 수 있는 케이스를 포함하는 케이스 수의 총계입니다.

레이블. `FILE LABEL` 명령으로 정의된 파일 레이블(있는 경우)입니다.

문서. 데이터 파일 문서 텍스트입니다.

가중치 상태. 가중치를 사용하는 경우 가중 변수 이름이 표시됩니다. 자세한 정보는 주제를 참조하십시오.

사용자 정의 속성. 사용자가 정의한 사용자 정의 데이터 파일 속성입니다. `DATAFILE ATTRIBUTE` 명령을 통해 정의된 데이터 파일 속성입니다.

예약된 속성. 예약된 시스템 데이터 파일 속성입니다. 시스템 속성을 표시할 수 있지만 변경하지 않아야 합니다. 시스템 속성 이름은 달러 부호($)로 시작합니다. "@" 또는 "$@"로 시작하는 이름을 가진 비표시 속성은 포함되지 않습니다. 결과에는 시스템 데이터 파일 속성의 이름 및 값이 모두 포함됩니다.

변수 표시 순서

변수 및 다중 응답 세트가 표시되는 순서 제어를 위해 다음 방법을 사용할 수 있습니다.

문자순. 변수 이름별 문자 순서입니다.

파일. 변수가 데이터 세트에 표시되는 순서입니다(데이터 편집기에 표시되는 순서). 오름차순의 경우 다중 응답 세트는 선택한 모든 변수 다음에 마지막으로 표시됩니다.

측정 수준 측정 수준별로 정렬합니다. 이를 통해 네 개(명목, 순서, 척도 및 알 수 없음)의 정렬 그룹을 만듭니다. 다중 응답 세트는 명목으로 취급됩니다.

참고: 숫자변수의 측정 수준은 외부 소스 또는 새롭게 작성한 변수에서 데이터를 읽는 것과 같이 측정 수준을 명시적으로 설정하지 않은 경우 첫 번째 데이터가 전달되기 전에 "알 수 없는" 상태일 수 있습니다.

변수 목록. 변수 및 다중 응답 세트가 변수 탭의 선택한 변수 목록에 나타나는 순서입니다.

사용자 정의 속성 이름. 정렬 순서 옵션 목록에도 사용자가 정의한 사용자 정의 변수 속성의 이름이 포함됩니다. 오름차순의 경우 속성이 없는 변수가 맨 위에 정렬되고 속성이 있지만 속성에 대해 정의된 값이 없는 변수가 다음에 위치하고 속성에 대해 정의된 값을 가진 변수가 값의 문자순으로 위치합니다.

최대 범주 수

결과에 값 레이블, 개수 또는 각 고유 값에 대한 퍼센트가 포함된 경우 값의 수가 지정된 값을 초과하면 테이블에서 이 정보를 출력하지 않을 수 있습니다. 기본적으로 이 정보는 변수의 고유 값 수가 200을 초과할 경우 출력되지 않습니다.

코드북 통계 탭

통계 탭을 통해 결과에 포함된 요약 통계를 제거하거나 전체적으로 요약 통계 표시를 차단할 수 있습니다.

빈도 및 퍼센트

명목 및 순서 변수, 다중 응답 세트 및 레이블 값 또는 척도변수의 경우 사용 가능한 통계는 다음과 같습니다.

개수 변수의 각 값 또는 값 범위를 갖는 케이스의 개수나 수입니다.

퍼센트*(Percent)*. 특정 값을 갖는 케이스의 백분율입니다.

중심 경향 및 산포

척도변수의 경우 사용 가능한 통계는 다음과 같습니다.

평균. 중심 경향 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

표준 편차. 평균에 대한 산포 측도입니다. 정규 분포에서 케이스의 68%는 평균의 표준 편차 내에 있으며 케이스의 95%는 2배 표준 편차 내에 있습니다. 예를 들어, 평균 연령이 45세이고 표준 편차가 10인 경우 정규 분포 내에서 95% 케이스는 25세와 65세 사이에 있습니다.

Quartiles. 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

참고: 변수 탭의 소스 변수 목록에서 일시적으로 변수와 연결된 측정 수준을 변경할 수 있으며 해당 변수에 대해 표시된 요약 통계를 변경할 수 있습니다.

빈도

빈도분석 프로시저로 여러 종류의 변수를 설명하는 데 유용한 그래픽 표시와 통계를 사용할 수 있습니다. 데이터를 살펴볼 때는 먼저 빈도분석 프로시저부터 시작하는 것이 좋습니다.

빈도분석 보고서와 막대도표에서는 빈도에 따라 내림차순이나 오름차순으로 개별값을 조정하고 범주 순서를 지정합니다. 변수에 개별값이 많을 때는 빈도분석 보고서가 출력되지 않을 수도 있습니다. 도표에 빈도분석(기본값) 또는 퍼센트를 레이블로 붙여넣을 수 있습니다.

예

산업 유형별 회사의 고객 분포는 어떠합니까? 결과에서 정부 기관에는 37.5%, 법인 단체에는 24.9%, 학회에는 28.1%, 의료 기관에는 9.4%의 고객이 분포되어 있음을 알 수 있습니다. 판매 수입과 같은 연속적이고 양적인 데이터에서 평균 제품 판매액은 $3,576이며 표준 편차는 $1,078이라는 것을 알 수 있습니다.

통계량 및 도표

빈도 개수, 퍼센트, 누적 퍼센트, 평균, 중위수, 최빈값, 합계, 표준 편차, 분산, 범위, 최소값 및 최대값, 평균의 표준 오차, 왜도와 첨도(표준 오차 포함), 분위수, 사용자 정의 백분위수, 막대도표, 원도표, 히스토그램 등을 선택할 수 있습니다.

데이터 고려사항

데이터

숫자 코드나 문자를 사용하여 범주형 변수(명목 수준이나 순위 수준 측정)를 코딩합니다.

가정

표와 퍼센트를 통해 분포의 데이터를 설명하고 특히 순서가 지정되었거나 지정되지 않은 범주가 있는 변수를 기술할 수 있습니다. 평균과 표준 편차와 같이 선택적 요약 통계의 대부분은 정규 이론을 따르며 대칭 분포로 되어 있는 양적변수에 적합합니다. 중위수, 분위수, 백분위수 등의 강력한 통계는 정규성 가정에 상관 없이 양적변수에 적합합니다.

빈도표 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **기술통계량** > **빈도...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 범주형 변수와 양적변수를 하나 이상 선택합니다.
3. 선택적으로 APA 유형 지침을 준수하는 출력결과 표를 작성하려면 **APA 유형 표 작성**을 선택합니다.
4. 선택적으로 다음을 수행할 수 있습니다.
   - 양적변수에 대한 기술통계를 보려면 **통계**를 클릭합니다.
   - 막대도표, 원도표, 히스토그램 등을 보려면 **도표**를 클릭합니다.
   - 결과가 출력되는 순서를 보려면 **형식**을 클릭합니다.
   - 특정 조건에 따라 피벗표의 특성을 자동으로 변경하는 조건을 지정하려면 **유형**을 클릭합니다.
   - 평균, 중위수, 비율, 오즈비, 상관계수 또는 회귀계수와 같은 추정에 대한 표준오차 및 신뢰구간의 강력한 추정을 파생시키려면 **붓스트랩**을 클릭합니다. 가설검정을 구축하는 데도 사용할 수 있습니다.

빈도분석 통계

백분위수 값. 특정 퍼센트는 위에 있고 다른 퍼센트는 아래에 있도록 정렬된 데이터를 그룹으로 나누는 양적변수 값을 말합니다. 분위수(제 25, 50, 75 백분위수)를 사용하여 관측값을 동일한 크기의 네 그룹으로 나눕니다. 네 가지 이외에 다른 그룹 수인 경우에는 **절단점 - n 동일 크기 그룹 수**를 선택합니다. 개별 백분위수를 지정할 수도 있습니다. 예를 들어, 제 95 백분위수는 관측값의 95%이하인 값을 나타냅니다.

중심 경향. 분포 위치를 설명하는 통계에는 평균, 중위수, 최빈값, 모든 값 합계 등이 포함됩니다.

- 평균. 중심 경향 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.
- 중앙값*(Median)*. 전체 사례의 절반이 그 값보다 크거나 작거나 큰 값, 즉 50분위수. 케이스 수가 짝수인 경우 중위수는 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중위수는 평균과 달리 중심을 벗어난 값에는 영향을 받지 않는 중심 경향 측도이며, 상한 극단값 또는 하한 극단값에 따라 달라질 수 있습니다.
- 모드. 가장 자주 발생하는 값입니다. 여러 값에서 최대 발생 빈도를 공유하는 경우 각각을 최빈값이라고 합니다. 빈도분석 프로시저는 여러 최빈값 중에서 가장 작은 최빈값만 보고합니다.
- *Sum*. 비결측값을 갖는 전체 케이스 값의 총계입니다.

산포. 데이터에 분포되어 있는 변동의 양을 측정하는 통계에는 표준 편차, 분산, 범위, 최소값, 최대값, 평균의 표준 오차 등이 있습니다.

- 표준 편차. 평균에 대한 산포 측도입니다. 정규 분포에서 케이스의 68%는 평균의 표준 편차 내에 있으며 케이스의 95%는 2배 표준 편차 내에 있습니다. 예를 들어, 평균 연령이 45세이고 표준 편차가 10인 경우 정규 분포 내에서 95% 케이스는 25세와 65세 사이에 있습니다.
- 분산*(Variance)*. 평균에 대한 산포 측도값으로, 평균으로부터의 제곱 편차 합계를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.
- *Range*. 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.
- 최소값*(minimum)*. 숫자변수의 가장 작은 값입니다.
- 최대. 숫자변수의 가장 큰 값입니다.
- *S. E.* 평균. 동일 분포로부터 선택한 표본 간에 발생할 수 있는 평균값의 차이에 대한 측도입니다. 이 값을 사용하여 관측 평균과 가설 값을 간략하게 비교할 수 있습니다. 즉, 표준 오차에 대한 차이 비율이 ?2보다 작거나 +2보다 큰 경우 두 값이 다르다고 판단할 수 있습니다.
- 이력서. 표준 편차 및 평균에 대한 비율입니다. 값은 백분율로 표시됩니다.

분포. 첨도와 왜도는 분포의 형태와 대칭을 설명하는 통계입니다. 표준 오차와 함께 표시됩니다.

- 왜도*(Skewness)*. 분포의 비대칭에 대한 측도입니다. 정규 분포는 대칭적이며 왜도 값이 0입니다. 유의한 양의 왜도가 포함된 분포는 오른쪽 꼬리가 더 깁니다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.
- 첨도*(Kurtosis)*. 이상치가 있는 범위의 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 데이터가 정규 분포보다 극단적인 이상값을 나타냄을 표시합니다. 음수인 첨도는 데이터가 정규 분포보다 극단적인 이상치가 적음을 나타냅니다. 정규 분포에 대해 값이 0인 경우에 사용되는 첨도의 정의를 때때로 초과 첨도라고 합니다. 일부 소프트웨어는 정규 분포에 대해 값이 3인 첨도를 보고할 수 있습니다.

값들이 그룹 중심점임. 데이터의 값이 그룹의 중심점일 경우에는 (예를 들어, 30대인 사람들의 연령을 모두 35로 코딩하는 것을 말합니다) 이 옵션을 선택하여 원래의 그룹화하지 않은 데이터의 중위수와 백분위수를 추정할 수 있습니다.

빈도분석 도표

참고: 붓스트랩 대화 상자에서 **붓스트랩 수행**이 사용으로 설정된 경우에는 출력결과에 도표가 작성되지 않습니다.

차트 유형

원도표는 전체에서 부분이 차지하는 정도를 표시합니다. 원도표의 각 조각은 단일 그룹화 변수로 정의한 그룹에 해당합니다. 막대도표에서 개별 값이나 범주에 대한 개수를 별도의 막대로 구분하여 표시하고 시각적으로 범주를 비교할 수 있습니다. 히스토그램에도 막대는 있으나 동일 구간 척도마다 도표화됩니다. 각 막대의 높이는 구간 내 양적변수 값의 개수를 나타냅니다. 히스토그램은 분포의 형태, 중심, 산포 등을 나타냅니다. 히스토그램에 정규 곡선을 추가하여 데이터가 정규적으로 분포되었는지 확인할 수 있습니다.

차트 값

막대도표의 경우 빈도 또는 퍼센트를 척도축에 레이블로 지정할 수 있습니다.

빈도분석 형식

출력순서. 빈도표를 데이터의 실제 값이나 값의 개수(발생 빈도)에 따라 내림차순이나 오름차순으로 정렬할 수 있습니다. 그러나 히스토그램이나 백분위수가 필요한 경우에는 빈도분석에서 해당 변수를 양적변수로 가정하여 값을 오름차순으로 표시합니다.

다중 변수. 다중 변수에 대한 통계표를 작성할 경우에는 단일 테이블(**변수 비교**)에 모든 변수를 표시하거나 각 변수(**각 변수별로 결과를 나타냄**)에 대해 개별 통계표를 표시할 수 있습니다.

범주가 많은 테이블은 출력하지 않음. 이 옵션으로 지정한 수 이상의 값은 테이블에 출력되지 않도록 할 수 있습니다.

기술통계

기술통계 프로시저를 사용하여 단일 테이블의 여러 변수에 대해 일변량 요약 통계를 표시하고 표준화 값(*z* 스코어)을 계산합니다. 변수 순서는 평균 크기(오름차순이나 내림차순)를 기준으로 하거나 변수를 선택한 순서(기본값)를 기준으로 또는 문자순으로 지정할 수 있습니다.

z 스코어를 저장하면 데이터 편집기의 데이터로 추가되므로 도표, 데이터 목록, 분석 등에 사용할 수 있습니다. 변수를 다른 단위(예: 일인당 국내총생산(GDP)과 문맹률)로 기록하는 경우 *z* 스코어 변환에서는 시각적으로 보다 비교하기 쉬운 공통 척도에 변수를 둡니다.

예. 데이터의 각 케이스에 몇 달 동안의 일일 판매량을 합산한 영업사원별 일일 매출 총액이 있는 경우(예: 갑, 을, 병 등에 한 항목을 입력한 경우) 기술통계 프로시저로 각 영업사원의 일일 매출 평균을 계산할 수 있으며 최고 평균 판매액부터 최저 평균 판매액까지 순서를 지정할 수 있습니다.

통계. 표본 결과, 평균, 최소값, 최대값, 표준 편차, 분산, 범위, 합계, 평균의 표준 오차, 표준 오차에 대한 첨도 및 왜도입니다.

기술통계 데이터 고려 사항

데이터. 오차, 이상값, 이상항목 예외를 기록하기 위해 숫자변수를 그래픽으로 확인한 후 사용합니다. 기술통계 프로시저는 파일이 클 때(예: 수천 케이스) 특히 효과적입니다.

가정. *z* 스코어를 포함한 대부분의 사용 가능한 통계는 정규 이론에 기초하며 대칭 분포로 된 양적변수(구간 수준 또는 비율 수준 척도)에 적합합니다. 이때 순서 없는 범주나 비대칭 분포로 되어 있는 변수는 제외됩니다. *z* 스코어의 분포는 원래 데이터 형태와 같으므로 *z* 스코어를 계산해도 데이터 문제를 해결할 수 없습니다.

기술통계량 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **기술통계량** > **기술통계...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 변수를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- **표준화된 값을 변수로 저장**을 선택하여 *z* 점수를 새 변수로 저장합니다.
- 선택적인 통계와 출력 순서를 선택하려면 **옵션**을 클릭합니다.

기술통계 옵션

평균 및 합계. 평균이나 산술 평균이 기본적으로 표시됩니다.

산포. 데이터의 산포도나 변동을 측정하는 통계에는 표준편차, 분산, 범위, 최소값, 최대값, 평균의 표준오차가 포함됩니다.

- 표준 편차. 평균에 대한 산포 측도입니다. 정규 분포에서 케이스의 68%는 평균의 표준 편차 내에 있으며 케이스의 95%는 2배 표준 편차 내에 있습니다. 예를 들어, 평균 연령이 45세이고 표준 편차가 10인 경우 정규 분포 내에서 95% 케이스는 25세와 65세 사이에 있습니다.
- 분산*(Variance)*. 평균에 대한 산포 측도값으로, 평균으로부터의 제곱 편차 합계를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.
- *Range*. 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.
- 최소값*(minimum)*. 숫자변수의 가장 작은 값입니다.
- 최대. 숫자변수의 가장 큰 값입니다.
- *S.E.* 평균. 동일 분포로부터 선택한 표본 간에 발생할 수 있는 평균값의 차이에 대한 측도입니다. 이 값을 사용하여 관측 평균과 가설 값을 간략하게 비교할 수 있습니다. 즉, 표준 오차에 대한 차이 비율이 ?2보다 작거나 +2보다 큰 경우 두 값이 다르다고 판단할 수 있습니다.
- 이력서. 표준 편차 및 평균에 대한 비율입니다. 값은 백분율로 표시됩니다.

분포. 첨도와 왜도는 분포의 형태와 대칭의 특성을 나타내는 통계로 표준 오차와 함께 표시됩니다.

- 첨도*(Kurtosis)*. 이상치가 있는 범위의 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 데이터가 정규 분포보다 극단적인 이상값을 나타냄을 표시합니다. 음수인 첨도는 데이터가 정규 분포보다 극단적인 이상치가 적음을 나타냅니다. 정규 분포에 대해 값이 0인 경우에 사용되는 첨도의 정의를 때때로 초과 첨도라고 합니다. 일부 소프트웨어는 정규 분포에 대해 값이 3인 첨도를 보고할 수 있습니다.
- 왜도*(Skewness)*. 분포의 비대칭에 대한 측도입니다. 정규 분포는 대칭적이며 왜도 값이 0입니다. 유의한 양의 왜도가 포함된 분포는 오른쪽 꼬리가 더 깁니다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.

출력 순서. 기본적으로 변수는 변수목록에서 선택한 순서대로 표시됩니다. 선택적으로 변수를 문자순 또는 오름차순이나 내림차순으로 출력할 수 있습니다.

DESCRIPTIVES 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- VARIABLES 하위 명령을 사용하여 일부 변수에 대해 표준화 스코어(*z* 스코어)를 저장합니다.
- VARIABLES 하위 명령을 사용하여 표준화 스코어가 포함된 새 변수의 이름을 지정합니다.
- MISSING 하위 명령을 사용하여 변수의 결측값을 가진 분석 케이스에서 제외합니다.
- SORT 하위 명령을 사용하여 출력의 변수를 평균을 제외한 모든 통계의 값별로 정렬합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

백분위수

백분위수 프로시저는 백분위수의 표를 표시합니다.

백분위수 값
백분위값을 선택하면 기본적으로 5%, 10%, 25%, 50%, 75%, 90%, 95% 백분위값이 표시됩니다. 특정 퍼센트는 위에 있고 다른 퍼센트는 아래에 있도록 정렬된 데이터를 그룹으로 나누는 양적변수 값을 말합니다.

사분위수
분위수(제 25, 50, 75 백분위수)를 사용하여 관측값을 동일한 크기의 네 그룹으로 나눕니다.

사용자 정의
사용자 정의를 선택하는 경우 하나 이상의 값을 입력하여 분석을 실행하십시오. 입력 값은 0-100범위의 숫자여야 합니다. **추가**, **변경**및 **제거** 단추를 사용하여 백분위수 값 목록의 값에 대해 작업하십시오.

백분위수 방법
기본적으로 백분위수 계산을 위해 HAVERAGE 방법이 선택되어 있습니다.

백분위수 테이블 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 기술통계 > 백분위수 ..

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자변수를 선택하여 변수 필드로 옮깁니다.
3. 백분위수를 계산할 값을 지정하려면 **백분위수 값** 을 선택하십시오.
4. 백분위수 계산 방법을 선택하려면 **백분위수 방법** 을 선택하십시오.
5. 선택적으로 다음을 수행할 수 있습니다.
   - **결측값** 을 클릭하여 결측값 처리를 제어합니다.
   - 백분위수의 추정값에 대한 표준 오차 및 신뢰구간의 강력한 추정값을 도출하려면 **부트스트랩** 을 클릭하십시오.

백분위수 결측값

결측값
결측값의 처리를 제어합니다.

목록별 결측값 제외

지정된 변수에 대한 결측값이 있는 케이스는 모든 분석에서 제외됩니다. 이는 기본값입니다.

변수별
변수별로 결측값이 있는 케이스는 제외합니다.

탐색

데이터 탐색 프로시저를 사용하여 모든 케이스나 분리된 케이스 그룹에 대한 요약 통계와 그래픽 표시를 만들 수 있습니다. 데이터 탐색 프로시저를 사용하는 것은 데이터 선별, 이상값 식별, 설명, 가정 확인, 부모집단(케이스 그룹) 간 차이 특성화 등의 여러 가지 원인 때문입니다. 데이터 선별로 예외적인 값이나 극단값, 데이터 간 갭, 기타 특성을 볼 수 있습니다. 데이터를 탐색하여 데이터 분석에 사용하려는 통계 기법이 적합한지 여부를 결정할 수 있습니다. 데이터 탐색에서 이 기법을 사용할 때 정규 분포가 필요한 경우 데이터 변환이 필요한지를 알 수 있습니다. 또는 비모수 검정이 필요한지를 알 수 있습니다.

예. 생쥐가 네 번의 다른 강화 계획을 거치면서 미로 찾기를 할 때 시간 분포를 가정해 봅니다. 각 그룹마다 시간 분포가 거의 정규적인지 여부와 네 분산이 동일한지를 알 수 있습니다. 또한 다섯 가지의 최대 시간과 다섯 가지의 최소 시간에 해당하는 케이스를 구별할 수도 있습니다. 상자도표와 줄기와 잎그림 도표는 각 그룹에 대한 학습 시간 분포를 도표로 보여 줍니다.

통계 및 도표. 평균, 중위수, 5% 잘린 평균, 표준 오차, 분산, 표준 편차, 최소값, 최대값, 범위, 사분위수 범위, 첨도와 왜도 및 해당 표준 오차, 평균의 신뢰구간(및 지정한 신뢰수준), 백분위수, Huber의 M-추정량, Andrew의 웨이브 추정량, Hampel의 redescending M 추정량, Tukey의 biweight 추정량, 최대 다섯 개 값 및 최소 다섯 개 값, 정규성 검정에서 Lilliefors 유의수준의 Kolmogorov-Smirnov 통계, Shapiro-Wilk 통계 등을 선택할 수 있습니다. 상자도표, 줄기와 잎그림 도표, 히스토그램, 정규성도표, Levene 검정과 평균-산포 도표, 변환 등을 선택할 수 있습니다.

데이터 탐색 데이터 고려 사항

데이터. 데이터 탐색 프로시저는 양적변수(구간 수준 측도나 비율 수준 측도)에 대해 사용할 수 있습니다. 요인변수(데이터를 케이스 그룹으로 나눌 때 사용)에는 적절한 개별 값의 수(범주)가 포함되어야 합니다. 이들 값은 단문자나 숫자일 수 있습니다. 케이스 레이블 변수는 상자도표에서 이상값 레이블에 사용되며 단문자와 긴문자(처음 15바이트), 숫자 등이 될 수 있습니다.

가정. 데이터 분포가 대칭이거나 정규적일 필요는 없습니다.

데이터 탐색

1. 메뉴에서 다음을 선택합니다.

   분석 > 기술통계량 > 탐색...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 종속변수를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 하나 이상의 요인변수를 선택합니다. 이때 변수 값들은 케이스 그룹을 정의합니다.
- 케이스를 설명할 식별변수를 선택합니다.
- 추정량, 이상값, 백분위수, 빈도표 등을 확인하려면 **통계**를 클릭합니다.
- 히스토그램, 정규 확률 도표와 검정, Levene 통계와 평균-산포 도표를 보려면 **도표**를 클릭합니다.
- 결측값을 처리하려면 **옵션**을 클릭합니다.

데이터 탐색 통계

기술통계. 중심 경향과 산포의 측도를 기본값으로 표시합니다. 중심 경향 측도로 분포 위치를 알 수 있으며 여기에는 평균, 중위수, 5% 잘린 평균 등이 포함됩니다. 산포 측도를 통해 값의 상이성을 알 수 있는데 여기에는 표준오차, 분산, 표준 편차, 최소값, 최대값, 범위, 사분위수 범위 등이 포함됩니다. 기술통계에도 분포 형태에 대한 측도가 있는데 왜도와 첨도가 표준 오차와 함께 표시됩니다. 평균에 대한 95% 수준의 신뢰구간이 표시되며 다른 신뢰수준을 지정할 수도 있습니다.

M-추정량. 위치를 추정하기 위해 표본 평균과 중위수를 대체합니다. 계산되는 추정량은 케이스에 적용되는 가중치에 따라 다릅니다. Huber의 M-추정량, Andrew의 웨이브 추정량, Hampel의 redescending M 추정량, Tukey의 biweight 추정량 등이 표시됩니다.

이상값. 최대 다섯 개 값과 최소 다섯 개 값을 케이스 레이블과 함께 표시합니다.

백분위수

백분위수 프로시저는 백분위수의 표를 표시합니다. **'백분위수'를** 선택하면 기본적으로 5, 10, 25, 50, 75, 90, 95백분위수의 값이 표시됩니다. 특정 퍼센트는 위에 있고 다른 퍼센트는 아래에 있도록 정렬된 데이터를 그룹으로 나누는 양적변수 값을 말합니다.

사분위수

분위수(제 25, 50, 75 백분위수)를 사용하여 관측값을 동일한 크기의 네 그룹으로 나눕니다.

사용자 정의

사용자 정의를 선택하는 경우 하나 이상의 값을 입력하여 분석을 실행하십시오. 입력 값은 0-100범위의 숫자여야 합니다. **추가**, **변경**및 **제거** 단추를 사용하여 백분위수 값 목록의 값에 대해 작업하십시오.

방법

기본적으로 백분위수 계산에 대해 HAVERAGE 방법이 선택되어 있습니다.

데이터 탐색 도표

상자도표. 이러한 옵션을 사용하여 여러 종속변수가 있을 때 상자도표의 출력을 제어할 수 있습니다. **요인 수준들과 함께**로 각 종속변수에 따라 출력 형식을 다르게 만들 수 있습니다. 출력 형식에서 상자도표는 요인변수로 정의한 그룹마다 나타납니다. **종속변수와 함께**로 요인변수를 사용하여 정의한 각 그룹의 출력 형식을 다르게 만들 수 있습니다. 출력 형식에서 상자도표는 각 종속변수에 대해 단계별로 나타납니다. 이러한 출력 형식은 서로 다른 변수가 다른 시간에 측정한 단일 특성을 나타낼 때 특히 유용합니다.

기술통계. 기술통계에서 줄기와 잎그림 도표 및 히스토그램을 선택할 수 있습니다.

검정과 함께 정규성도표. 정규확률과 디트렌딩된 정규 확률 도표를 출력합니다. 정규성 검정에 대해 Lilliefors 유의수준을 가지는 Kolmogorov-Smirnov 통계가 표시됩니다. 정수가 아닌 가중치를 지정하면 가중 표본 결과가 3과 50 사이에 있을 때의 Shapiro-Wilk 통계가 계산됩니다. 가중치를 지정하지 않거나 정수 가중치를 지정하면 가중 표본 결과가 3과 5000 사이에 있을 때의 통계가 계산됩니다.

Levene 검정과 평균-산포 도표. 평균-산포 도표에 대한 데이터 변환을 제어합니다. 모든 평균-산포 도표에서 분산의 동질성에 대한 회귀선의 기울기와 Levene 검정이 표시됩니다. 변환을 선택하면 Levene 검정은 변환된 데이터에 따라 달라집니다. 요인변수를 선택하지 않으면 평균-산포 도표가 작성되지 않습니다. **제곱값 추정**으로 셀에서 등분산에 대한 거듭제곱값 변환을 추정할 수 있을 뿐 아니라 모든 셀에서 중위수의 자연 로그에 대한 사분위수 범위의 자연 로그 도표를 작성할 수 있습니다. 평균-산포 도표를 사용하여 그룹에서 분산을 안정화(등분산화)할 변환에 대한 거듭제곱값을 결정할 수 있습니다. **변환**을 사용하여 거듭제곱값 중 하나를 선택하고 거듭제곱값 추정의 설명을 따라 수행하여 변환 데이터 도표를 만들 수 있습니다. 변환 데이터의 사분위수 범위와 중위수가 도표에 나타납니다. **변환하지 않음**으로 원시 데이터 도표를 만들 수 있습니다. 이 도표는 거듭제곱값 1로 변환하는 것과 같습니다.

데이터 탐색 거듭제곱값 변환

평균-산포 도표에 대한 거듭제곱값 변환을 나타냅니다. 데이터를 변환하려면 변환할 거듭제곱값을 선택해야 합니다. 다음 중 하나를 선택합니다.

- **자연 로그.** 자연 로그 변환으로 이는 기본값입니다.
- **1/제곱근.** 각 데이터 값에서 제곱근의 역수를 계산합니다.
- **역수.** 각 데이터 값의 역수를 계산합니다.

- **제곱근.** 각 데이터 값의 제곱근을 계산합니다.
- **제곱.** 각 데이터 값을 제곱합니다.
- **큐브.** 각 데이터 값을 큐브합니다.

데이터 탐색: 옵션

결측값. 결측값의 처리를 제어합니다.

- **목록별 결측값 제외.** 종속변수나 요인변수에 대한 결측값이 있는 케이스는 모든 분석에서 제외됩니다. 이는 기본값입니다.
- **대응별 결측값 제외.** 그룹(셀)에서 변수에 대한 결측값이 없는 케이스는 해당 그룹의 분석에 포함됩니다. 이러한 케이스에는 다른 그룹에서 사용한 변수에 대한 결측값이 있을 수 있습니다.
- **결측값 출력.** 요인변수에 대한 결측값이 구분된 범주로 처리됩니다. 이 추가 범주에 대해 모든 결과가 작성됩니다. 빈도표에 결측값에 대한 범주가 포함됩니다. 요인변수에 대한 결측값은 포함되나 결측값이라는 레이블이 붙습니다.

EXAMINE 명령 추가 기능

데이터 탐색 프로시저는 EXAMINE 명령 구문을 사용합니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 요인변수에서 정의한 그룹에 대해 결과와 도표 이외의 전체 결과와 도표를 요청합니다(TOTAL 하위 명령 사용).
- 상자도표의 그룹에 대해 공통 척도를 지정합니다(SCALE 하위 명령 사용).
- 요인변수의 상호작용을 지정합니다(VARIABLES 하위 명령 사용).
- 기본값 이외의 백분위수를 지정합니다(PERCENTILES 하위 명령 사용).
- 5가지 방법에 따라 백분위수를 계산합니다(PERCENTILES 하위 명령 사용).
- 평균-산포 도표에 대한 거듭제곱값 변환을 지정합니다(PLOT 하위 명령 사용).
- 표시할 극단값 수를 지정합니다(STATISTICS 하위 명령 사용).
- M-추정량, 위치의 강력한 추정량에 대한 모수를 지정합니다(MESTIMATORS 하위 명령 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

교차분석

교차 분석표 프로시저를 사용하면 이원배치표와 다원배치표를 만들고 이원배치표에 대한 다양한 검정과 연관 측도를 사용할 수 있습니다. 테이블의 구조와 범주 정렬 여부에 따라 사용할 검정이나 측도가 결정됩니다.

부분 감마 계수의 예외로 교차 분석표 통계 및 연관성 측도는 각 이원표에 대해 별도로 계산됩니다. 행, 열, 레이어 요인(제어변수) 등을 지정하면 각 레이어 요인 값(또는 둘 이상의 제어변수에 대한 값 조합)에 대해 결합된 통계와 측도의 패널을 작성하게 됩니다. 예를 들어, 성별이 생활(흥미진진, 평범, 무미건조)에 대해 기혼(예, 아니오) 테이블에 대한 레이어 요인인 경우 여자에 대한 이원배치표 결과가 남자에 대한 이원배치표와는 별도로 분리되어 계산되고 여자에 대한 이원배치표는 남자에 대한 이원배치표 다음에(역으로도 해당) 패널로 인쇄됩니다.

예. 중소기업 고객이 대기업 고객보다 서비스 판매(예: 교육 및 컨설팅)에서 이윤이 많이 남습니까? 교차 분석표에서 대부분의 중소기업(직원 500명 미만)은 높은 서비스 이윤을 거두는 반면 대부분의 대기업(직원 2,500명 이상)은 이보다 낮은 서비스 이윤을 거둔다는 사실을 알 수 있습니다.

통계 및 연관 측도. Pearson 카이제곱, 우도비 카이제곱, 선형별 연관 검정, Fisher의 정확 검정, Yates의 수정 카이제곱, Pearson의 *r*, Spearman의 rho, 분할계수, 파이, Cramér의 *V*, 대칭 및 비대칭 람다, Goodman과 Kruskal의 타우, 불확실성 계수, 감마, Somers의 *d*, Kendall의 타우-*b*, Kendall의 타우-*c*, 에타 계수, Cohen의 kappa, 상대적 위험 추정값, 승산비, McNemar 검정, Cochran 및 Mantel-Haenszel 통계, 열 비율 통계를 선택할 수 있습니다.

교차 분석표 데이터 고려 사항

데이터. 각 테이블 변수의 범주를 정의하려면 숫자 또는 문자열 변수의 값을 사용하십시오. 예를 들어, 성별의 경우 데이터를 1과 2 또는 남자와 여자로 코딩할 수 있습니다.

가정. 일부 통계와 측도는 통계의 구분에 논의된 대로 정렬된 범주(순서 데이터)나 양적인 값(구간 데이터나 비율 데이터)을 가정합니다. 그 밖에 통계와 측도는 테이블 변수에 정렬되지 않은 범주(명목 데이터)가 있을 때 유효합니다. 카이제곱 통계(파이 및 Cramér의 *V*, 분할계수 등)에서는 데이터가 다항분포에서 비롯된 무작위 표본이어야 합니다.

참고: 원래 변수는 범주를 나타내는 숫자 코드(예: 1 = 하한, 2 = 중간, 3 = 상한) 또는 문자열 값입니다. 그러나 문자열 값의 영문자 순서는 범주의 원래 순서를 반영하는 것으로 가정합니다. 예를 들어, 하한, 중간, 상한의 값을 갖는 문자열 변수의 경우 범주 순서는 원래 올바른 순서가 아닌 상한, 하한, 중간 순으로 해석됩니다. 일반적으로 순서 데이터를 나타내려면 숫자 코드를 사용하는 것이 더 정확합니다.

교차 분석표 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 기술 통계 > 교차 분석표...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 행 변수와 열 변수를 각각 하나 이상 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 하나 이상의 제어변수를 선택합니다.
- 이원배치표나 하위 테이블에 대한 검정과 연관 측도를 보려면 **통계**를 클릭합니다.
- 관측값과 기대값, 퍼센트, 잔차 등에 대해서는 **셀**을 클릭합니다.
- 범주 순서를 제어하려면 **형식**을 클릭합니다.

교차 분석표 레이어

하나 이상의 레이어 변수를 선택하면 각 레이어 변수(제어변수)의 범주마다 분리된 교차표가 작성됩니다. 예를 들어, 행 변수와 열 변수가 하나씩 있고 레이어 변수에 범주가 둘 있을 경우 레이어 변수의 범주마다 이원배치표가 작성됩니다. 제어변수의 다른 레이어를 작성하려면 **다음**을 클릭합니다. 첫 번째 레이어 변수 범주에 두 번째 레이어 변수가 조합되고 두 번째 레이어 변수 범주에 세 번째 레이어 변수가 조합되는 등의 각 조합의 경우마다 하위 테이블이 생성됩니다. 통계와 연관 측도가 필요한 경우에는 이원배치 하위 테이블에만 적용됩니다.

교차 분석표 수평배열 막대도표

수평배열 막대도표 출력. 수평배열 막대도표로 케이스 그룹의 데이터를 요약할 수 있습니다. 행에 지정한 각 변수 값에 대해 하나의 수평배열 막대도표가 생성됩니다. 각 군집내의 막대를 정의하는 변수는 열에서 지정한 변수가 됩니다. 이 변수의 각 값마다 다른 색상과 패턴의 막대 세트가 하나 있습니다. 열이나 행에 둘 이상의 변수를 지정하면 두 변수 조합마다 수평배열 막대도표가 작성됩니다.

테이블 레이어에 레이어 변수를 표시하는 교차 분석표

테이블 레이어에 레이어 변수를 표시합니다. 레이어 변수(제어변수)를 교차표의 테이블 레이어로 표시하도록 선택할 수 있습니다. 레이어 변수의 범주를 드릴다운할 수 있을 뿐만 아니라 행 및 열 변수에 대한 전체 통계를 표시하는 보기를 만들 수 있습니다.

demo.sav 데이터 파일(설치 디렉토리의 Samples 디렉토리에 있음)을 사용하는 예제는 아래에 표시되며 다음과 같은 방법으로 수집됩니다.

1. 행 변수로 가구소득범주*(단위: 1000) [*소득범주*]*, 열 변수로 *PDA* 소유 *[PDA* 보유*]*, 레이어 변수로 피교육수준 *[*교육수준*]*을 선택합니다.
2. **표 레이어에 레이어 변수 표시**를 선택합니다.
3. 셀 표시 하위 대화 상자에서 **열**을 선택합니다.
4. 교차표 프로시저를 실행하고 교차표를 두 번 클릭한 다음 교육수준 드롭다운 목록에서 **대학 졸업**을 선택합니다.

교차표의 선택한 보기는 대학 졸업 수준의 반응자에 대한 통계를 표시합니다.

교차 분석표 통계

카이제곱. 행 두 개와 열 두 개가 있는 테이블의 경우 **카이제곱**을 선택하여 Pearson 카이제곱, 우도비 카이제곱, Fisher의 정확 검정 및 Yates의 수정 카이제곱(연속수정)을 계산할 수 있습니다. 2 × 2 테이블의 경우 더 큰 테이블의 행 또는 열이 누락되지 않은 테이블에 5 미만의 예상 빈도가 있는 셀이 있을 때 Fisher의 정확한 테스트가 계산됩니다. Yates의 정정된 카이 제곱은 다른 모든 2 × 2 테이블에 대해 계산됩니다. 임의의 수의 행과 열이 있는 테이블에서는 **카이제곱**을 선택하여 Pearson 카이제곱과 우도비 카이제곱을 계산합니다. 두 테이블 모두 변수가 양적인 경우 **카이제곱**으로 선형 대 선형 연관 검정을 수행할 수 있습니다.

상관. 행과 열에 정렬된 값이 있는 테이블의 경우 **상관**을 통해 Spearman의 상관계수, rho(숫자 데이터만 사용)를 작성할 수 있습니다. Spearman의 rho는 순위 순서 간 연관 측도입니다. 두 테이블 변수(요인)가 양적인 경우 **상관**을 통해 Pearson 상관계수, *r*, 변수 간 선형 연관 측도를 작성합니다.

명목. 명목 데이터(고유한 순서가 없는 데이터 즉, 카톨릭, 개신교, 유대교 등)에서는 **분할계수**, **파이**(계수)와 **Cramér의 V**, **람다**(대칭 및 비대칭 람다와 Goodman과 Kruskal의 타우), **불확실성 계수** 등을 선택할 수 있습니다.

- 비상 계수. 카이제곱을 기반으로 한 연관 측도입니다. 값 범위는 0에서 1 사이이며 0은 행 변수와 열 변수 간에 연관이 없음을 나타내고 1에 가까운 값은 변수 간에 높은 연관이 있음을 나타냅니다. 가능한 최대값은 테이블의 행 수와 열 수에 따라 달라집니다.
- 파이 및 *Cramer*의 *V(Phi and Cramer's V)*. 파이는 카이제곱 통계량을 표본 결과로 나누고 결과의 제곱근을 구하는 연관 측도를 기반으로 하는 카이제곱입니다. Cramer의 V는 카이제곱을 기반으로 한 연관 측도입니다.
- 람다*(Lambda)*. 독립변수의 값을 사용하여 종속변수의 값을 예측할 때 오차의 비례 감소를 반영하는 연관 측도입니다. 값 1은 독립변수가 종속변수를 완전하게 예측한다는 의미이고, 값 0은 독립변수가 종속변수를 예측하는 데 도움이 되지 않는다는 의미입니다.
- 불확실성 계수. 한 변수의 값을 사용하여 다른 변수의 값을 예측할 때 오차의 비례 감소를 나타내는 연관 측도입니다. 예를 들어, 값 0.83은 다른 변수 값을 예측하는 데 하나의 변수 정보가 오차를 83%까지 축소한 것을 나타냅니다. 프로그램에서는 불확실성 계수의 대칭 버전과 비대칭 버전을 모두 계산합니다.

순서. 행과 열이 모두 정렬된 값을 가질 때 해당 테이블에서는 **감마**(2원배치표에 대해 0차 감마, 3원배치표에서 10원배치표까지에 대해 조건 감마), **Kendall의 타우-b**, **Kendall의 타우-c** 등을 선택합니다. 행 범주로부터 열 범주를 예측할 때는 **Somers의 d**를 선택합니다.

- 감마*(Gamma)*. -1과 1 사이에 있는 두 개의 순서 변수 간 연관의 대칭적 측도입니다. 절대값 1에 가까운 값은 두 변수 간의 강한 관계를 나타냅니다. 0에 가까울수록 관계가 적거나 없음을 의미합니다. 2원배치표의 경우 0차 감마가, 3원배치 이상의 테이블에서는 조건부 감마가 표시됩니다.
- *Somers*의 *d(Somers' d)*. 두 순서 변수 간의 연관 측도로 -1에서 1 사이입니다. 절대값 1에 가까운 값은 두 변수 간의 강한 관계를 나타내고 0에 가까운 값은 변수 간의 관계가 거의 또는 전혀 없음을 표시합니다. Somers의 d는 독립변수에 대해 동률이 아닌 대응 수가 포함된 경우에만 달라지는 감마의 비대칭 확장입니다. 이 통계의 대칭 버전도 계산됩니다.
- *Kendall*의 타우-*b(Kendall's tau-b)*. 동률을 고려하는 순서변수나 순위변수에 대한 비모수 상관 측도입니다. 계수의 부호는 관계의 방향을 나타내고 이에 대한 절대값은 강도를 나타냅니다. 이때 절대값이 클수록 관계가 더 강력합니다. 가능한 값의 범위는 -1부터 1까지이지만 -1 또는 +1 값은 제곱표에서만 구할 수 있습니다.
- *Kendall*의 타우-*c(Kendall's tau-c)*. 동률을 무시하는 순서변수에 대한 비모수 연관 측도입니다. 계수의 부호는 관계의 방향을 나타내고 이에 대한 절대값은 강도를 나타냅니다. 이때 절대값이 클수록 관계가 더 강력합니다. 가능한 값의 범위는 -1부터 1까지이지만 -1 또는 +1 값은 제곱표에서만 구할 수 있습니다.

명목 대 구간 척도. 하나의 변수가 범주형이고 다른 변수가 양적일 때 **에타**를 선택합니다. 범주형 변수는 숫자로 코딩되어야 합니다.

- *ETA*. 0-1범위의 연관 측도이며, 0은 행 변수와 열 변수 사이에 연관이 없음을 나타내고 1에 가까울수록 연관 정도가 높음을 나타냅니다. 에타는 구간 척도(예: 수입)에 대해 측정된 종속변수와 범주(예: 성별)의 수가 제한된 독립변수에 적합합니다. 두 에타 값이 계산되는데, 하나는 행 변수를 구간변수로 처리하고 다른 하나는 열 변수를 구간변수로 처리합니다.

Kappa. Cohen의 kappa는 두 평가자가 동일한 오브젝트를 평가할 때 두 평가자의 평가 간 일치를 측정합니다. 값 1은 완전 동의를 나타내고 값 0은 동의의 여지가 없음을 나타냅니다. Kappa는 행 및 열 값이 동일한 척도를 나타내는 제곱표를 기반으로 합니다. 한 변수에 대한 관측값이 있지만 다른 변수에는 0의 계수가 지정되지 않은 셀입

니다. 데이터 스토리지 유형(문자열 또는 숫자)이 두 변수에 대해 동일하지 않은 경우에는 Kappa가 계산되지 않습니다. 문자열 변수의 경우 두 개 변수 모두 정의된 길이가 동일해야 합니다.

리스크. 2 x 2테이블의 경우, 요인의 존재와 이벤트 발생 사이의 연관 강도에 대한 측도입니다. 통계의 신뢰구간에 값 1이 포함되어 있으면 요인이 이벤트와 관련되어 있지 않음을 나타냅니다. 요인 발생이 드문 경우 승산비를 상대적 위험의 추정값으로 사용할 수 있습니다.

McNemar. 관련된 두 이분형 변수에 대한 비모수 검정을 수행합니다. 카이제곱 분포를 사용하여 반응 변화를 검정합니다. 이 검정은 "전후" 계획에서 실험 개입에 따른 반응의 변화를 파악하는 데 유용합니다. 더 큰 제곱표에 대해서는 대칭에 대한 McNemar-Bowker 검정이 보고됩니다.

Cochran 및 *Mantel-Haenszel* 통계*(Cochran's and Mantel-Haenszel statistics)*. Cochran및 Mantel-Haenszel 통계는 하나 이상의 레이어 (제어) 변수에 의해 정의되는 공변량 패턴에 따라 이분형 요인변수와 이분형 반응변수 간의 독립성을 검정하는 데 사용할 수 있습니다. 다른 통계는 레이어별로 계산되지만 Cochran 및 Mantel-Haenszel 통계는 모든 레이어에 대해 한 번만 계산됩니다.

교차 분석표 셀 표시

유의한 카이제곱 검정의 데이터에서 패턴을 확인하기 위해 교차 분석표 프로시저로 관측빈도와 기대빈도 간 차이를 측정하는 잔차의 세 가지 유형과 기대빈도를 표시합니다. 테이블의 각 셀에는 선택한 개수, 퍼센트, 잔차의 조합이 포함되어 있습니다.

개수. 행 변수와 열 변수가 서로 독립된 상태일 때 실제로 관측된 케이스 수와 기대된 케이스 수를 의미합니다. 지정된 정수보다 작은 수를 숨기도록 선택할 수 있습니다. 숨겨진 값은 **<N**로 표시됩니다. 여기서 **N**은 지정된 정수입니다. 0 값이 허용되어 숨겨진 수가 없도록 지정한 경우라도 지정된 정수는 2와 같거나 그 이상이어야 합니다.

열 비율 비교. 이 옵션은 열 비율의 쌍대 비교를 계산하고 지정된 행에 대해 유의적으로 다른 대응 열을 나타냅니다. 유의차가 아래첨자를 사용하여 APA 스타일 형식으로 교차 분석표에 나타나고 0.05 유의 수준에서 계산됩니다. 참고: 관측 개수나 열 퍼센트를 선택하지 않고 이 옵션이 지정된 경우, 관측 개수는 열 비율 검정 결과를 나타내는 APA 스타일의 아래첨자로 교차 분석표에 포함됩니다.

- **P-값 조정(Bonferroni 방법).** 열 비율의 쌍대 비교는 다중 비교가 행해지는 사건에 대한 관측 유의 수준을 조정하는 Bonferroni 수정을 사용합니다.

퍼센트. 퍼센트는 행 가로방향으로나 열 세로방향으로 추가될 수 있습니다. 표(하나의 레이어)에 나타난 총 케이스 수의 퍼센트도 사용할 수 있습니다.

참고: 빈도 그룹에서 **낮은 빈도 숨기기**가 선택된 경우, 숨겨진 빈도와 연관된 퍼센트도 숨겨집니다.

잔차. 원래의 비표준화 잔차로 인해 관측값과 기대값 간의 차이가 나타납니다. 표준화 잔차와 수정된 표준화 잔차도 사용할 수 있습니다.

- 표준화되지 않음. 관측값과 기대값 간 차이입니다. 기대값은 두 변수 간에 관계가 없는 경우 셀에 기대한 케이스 수입니다. 행 변수와 열 변수가 서로 독립적인 경우 잔차가 양수이면 셀에 예상보다 많은 케이스가 있음을 의미합니다.
- 표준화됨. 표준 편차의 추정값으로 나눈 잔차입니다. Pearson 잔차라고도 하는 표준화 잔차는 평균이 0이고 표준 편차가 1이 됩니다.
- 수정된 표준화. 셀의 잔차(관측값에서 기대값을 뺀 값)를 표준 오차의 추정값으로 나눈 값입니다. 결과로 나타나는 표준화 잔차는 평균보다 높거나 낮은 표준 편차 단위로 표현됩니다.

APA 유형 표 작성. APA 유형 지침을 준수하는 출력결과 표를 작성합니다.

참고: APA 유형 표 작성을 선택한 경우에는 **관측**, **기대**, **행**, **열**, **총계** 옵션을 사용할 수 없습니다.

정수가 아닌 가중치. 셀 개수는 각 셀에 있는 케이스의 수를 나타내기 때문에 일반적으로 정수 값입니다. 데이터 파일이 분수 값(예: 1.25)이 있는 가중 변수로 가중되어 있는 경우에는 셀 개수가 분수 값이 될 수도 있습니다. 셀 개수를 계산하기 전이나 후에 자름 또는 반올림하거나 테이블 표시와 통계 계산 모두에 분수 셀 개수를 사용할 수 있습니다.

- 셀 개수 반올림*(Round cell counts)*. 케이스 가중치는 그대로 사용하지만 셀의 누적 가중치는 통계를 계산하기 전에 반올림합니다.
- 셀 개수 자름*(Truncate cell counts)*. 케이스 가중치는 그대로 사용하지만 통계를 계산하기 전에 셀의 누적 가중치를 자릅니다.

- 케이스 가중치 반올림*(Round case weights)*. 케이스 가중치가 사용하기 전에 반올림됩니다.
- 케이스 가중치 자름*(Truncate case weights)*. 케이스 가중치를 사용하기 전에 자릅니다.
- 조정 안 함. 케이스 가중치는 있는 그대로 사용되며 분수 셀 개수가 사용됩니다. 그러나 표본추출 및 검정에서만 사용할 수 있는 정확한 통계를 요청하면 정확한 검정 통계량을 계산하기 전에 셀의 누적 가중치를 자르거나 반올림합니다.

교차 분석표 형식

행을 행 변수 값의 오름차순이나 내림차순으로 정렬할 수 있습니다.

요약

요약 통계 프로시저는 하나 이상의 그룹화 변수의 범주 내에 있는 변수에 대한 하위 그룹 통계를 계산합니다. 그룹화 변수의 모든 수준은 교차 분석표로 작성됩니다. 사용자는 통계의 표시 순서를 선택할 수 있습니다. 모든 범주에서 각 변수에 대한 요약 통계도 함께 표시됩니다. 각 범주의 데이터 값이 나열되거나 표시되지 않을 수 있습니다. 데이터 세트가 큰 경우 처음 *n* 케이스만 나열할 수 있습니다.

예. 지역 및 고객의 업종에 따라 제품의 평균 판매액은 어떻게 나타납니까? 평균 판매액은 다른 지역보다도 서부 지역에서 약간 높았으며 서부 지역의 기업 고객이 가장 높은 평균 판매액을 산출한다는 것을 알 수 있습니다.

통계. 합계, 케이스 수, 평균, 중위수, 그룹 중위수, 평균의 표준 오차, 최소값, 최대값, 범위, 그룹화 변수의 첫 번째 범주에 대한 변수값, 그룹화 변수의 마지막 범주에 대한 변수값, 표준 편차, 분산, 첨도, 첨도의 표준 오차, 왜도, 왜도의 표준 오차, 총 합계 %, 전체 *N* 퍼센트, 합계 %, *N* 퍼센트, 기하평균, 조화 평균을 구할 수 있습니다.

요약 통계 데이터 고려 사항

데이터. 그룹화 변수는 그 값이 숫자이거나 문자가 될 수 있는 범주형 변수입니다. 범주의 수는 적당해야 합니다. 다른 변수는 순위화될 수 있어야 합니다.

가정. 일부 선택적 하위 그룹 통계(예: 평균과 표준 편차)는 정규 이론을 기초로 하며 대칭 분포를 가진 양적변수에 적절합니다. 강력한 통계분석 방법(예: 중위수와 범위)은 정규성 가정을 충족시키거나 충족시키지 않을 수 있는 양적변수에 적절합니다.

사례 요약분석 결과 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 보고서 > 사례 요약...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 변수를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 하나 이상의 그룹화 변수를 선택하여 데이터를 하위 그룹으로 나눕니다.
- **옵션**을 클릭하여 결과 제목을 변경하거나 결과 아래에 캡션을 추가하거나 결측값을 가지는 케이스를 제외시킵니다.
- **통계**를 클릭하여 선택적 통계를 구합니다.
- **케이스 출력**을 선택하여 각 하위 그룹에 케이스를 나열합니다. 기본적으로 시스템은 파일의 처음 100개의 케이스만 나열합니다. **첫 케이스부터의 출력수**에 대한 값 *n*을 높이거나 낮출 수 있으며, 이 항목을 선택 해제하여 모든 케이스를 나열할 수 있습니다.

요약분석 옵션

사례 요약을 사용하여 결과의 제목을 변경하거나 결과표 아래에 나타날 캡션을 추가할 수 있습니다. 텍스트에 줄바꿈을 삽입하려는 위치에 \n을 입력하여 제목과 캡션에서 줄바꿈을 제어할 수 있습니다.

합계에 대한 부제목을 표시하거나 표시하지 않도록 선택할 수 있으며 분석에 사용된 변수에 대한 결측값을 가지는 케이스를 포함시키거나 제외시킬 수도 있습니다. 마침표나 별표를 사용하여 결과에 결측값을 표시하는 것이

바람직한 경우도 있습니다. 값이 없는 경우 표시하고 싶은 문자, 어구 또는 코드를 입력합니다. 그렇지 않으면 결과의 결측 케이스에 특수 처리가 적용되지 않습니다.

사례 요약 통계

각 그룹화 변수의 각 범주 내에 있는 변수에 대해 다음 하위 그룹 통계량인 합계, 케이스 수, 평균, 중위수, 그룹 중위수, 평균의 표준 오차, 최소값, 최대값, 범위, 그룹화 변수의 첫 번째 범주에 대한 변수값, 그룹화 변수의 마지막 범주에 대한 변수값, 표준 편차, 분산, 첨도, 첨도의 표준 오차, 왜도, 왜도의 표준 오차, 총 합계 %, 전체 *N* 퍼센트, 합계 %, *N* 퍼센트, 기하평균, 조화 평균 중에서 하나를 선택할 수 있습니다. 셀 통계 목록에 통계가 나타나는 순서는 결과에 출력되는 순서입니다. 모든 범주의 각 변수에 대한 요약 통계도 함께 표시됩니다.

처음*(First)*. 데이터 파일에서 처음 나타나는 데이터 값을 표시합니다.

기하평균*(Geometric Mean)*. 데이터 값 간의 곱에 대한 n제곱근입니다. 여기서 n은 케이스 수를 나타냅니다.

그룹 중앙값*(Grouped Median)*. 그룹으로 코딩된 데이터에 대해 계산된 중앙값입니다. 예를 들어, 나이 데이터에서 30대의 각 값은 35로 코딩되고 40대의 각 값은 45로 코딩되는 경우 그룹 중위수는 코딩된 데이터에 대해 계산된 중위수입니다.

조화 평균*(harmonic mean)*. 그룹의 표본 결과가 같지 않을 때 평균 그룹 크기를 추정하는 데 사용됩니다. 조화 평균은 전체 표본 수를 표본 결과의 역수의 합으로 나눈 값입니다.

첨도*(Kurtosis)*. 이상치가 있는 범위의 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 데이터가 정규 분포보다 극단적인 이상값을 나타냄을 표시합니다. 음수인 첨도는 데이터가 정규 분포보다 극단적인 이상치가 적음을 나타냅니다. 정규 분포에 대해 값이 0인 경우에 사용되는 첨도의 정의를 때때로 초과 첨도라고 합니다. 일부 소프트웨어는 정규 분포에 대해 값이 3인 첨도를 보고할 수 있습니다.

마지막*(Last)*. 데이터 파일에서 마지막으로 나타나는 데이터 값을 표시합니다.

최대. 숫자변수의 가장 큰 값입니다.

평균. 중심 경향 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

중앙값*(Median)*. 전체 사례의 절반이 그 값보다 크거나 작거나 큰 값, 즉 50분위수. 케이스 수가 짝수인 경우 중위수는 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중위수는 평균과 달리 중심을 벗어난 값에는 영향을 받지 않는 중심 경향 측도이며, 상한 극단값 또는 하한 극단값에 따라 달라질 수 있습니다.

최소값*(minimum)*. 숫자변수의 가장 작은 값입니다.

N. 케이스 수(관측값 또는 레코드)입니다.

전체 케이스 수 %. 각 범주에서 전체 케이스 수의 백분율입니다.

총 합계 백분율. 각 범주에서 총 합계의 백분율입니다.

Range. 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.

왜도*(Skewness)*. 분포의 비대칭에 대한 측도입니다. 정규 분포는 대칭적이며 왜도 값이 0입니다. 유의한 양의 왜도가 포함된 분포는 오른쪽 꼬리가 더 깁니다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.

표준 편차. 평균에 대한 산포 측도입니다. 정규 분포에서 케이스의 68%는 평균의 표준 편차 내에 있으며 케이스의 95%는 2배 표준 편차 내에 있습니다. 예를 들어, 평균 연령이 45세이고 표준 편차가 10인 경우 정규 분포 내에서 95% 케이스는 25세와 65세 사이에 있습니다.

첨도의 표준 오차*(Standard Error of Kurtosis)*. 표준 오차에 대한 첨도의 비율을 정상성 검사로 사용할 수 있습니다(즉, 비율이 -2 미만이거나 +2보다 크면 정상성을 거부할 수 있습니다). 첨도가 높은 양의 값인 경우 분포의 양 끝이 정규 분포의 양끝보다 길어지고 음의 값인 경우 양끝이 짧아집니다(상자 형태 균일 분포와 유사).

평균의 표준 오차*(Standard Error of Mean)*. 동일 분포로부터 선택한 표본 간에 발생할 수 있는 평균값의 차이에 대한 측도입니다. 이 값을 사용하여 관측 평균과 가설 값을 간략하게 비교할 수 있습니다. 즉, 표준 오차에 대한 차이 비율이 ?2보다 작거나 +2보다 큰 경우 두 값이 다르다고 판단할 수 있습니다.

왜도의 표준 오차*(Standard Error of Skewness)*. 표준 오차에 대한 왜곡의 비율을 정상성 검사로 사용할 수 있습니다(즉, 이 비율이 -2 보다 작거나 +2보다 크면 정상성을 거부할 수 있습니다). 왜도가 큰 양의 값인 경우 오른쪽이 길어지고 큰 음의 값인 경우 왼쪽이 길어집니다.

Sum. 비결측값을 갖는 전체 케이스 값의 총계입니다.

분산*(Variance)*. 평균에 대한 산포 측도값으로, 평균으로부터의 제곱 편차 합계를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.

평균

그룹별 평균분석 프로시저로 하나 이상의 독립변수 범주 내에서 종속변수에 대한 하위 그룹 평균과 관련 일변량 통계량을 계산합니다. 선택적으로 일원 분산 분석표 및 에타, 선형성 검정을 사용할 수 있습니다.

예. 세 가지의 서로 다른 식용유에 녹아 있는 지방의 평균량을 계산하고 일원 분산 분석을 수행하여 평균의 차이를 확인할 수 있습니다.

통계. 합계, 케이스 수, 평균, 중위수, 그룹 중위수, 평균의 표준 오차, 최소값, 최대값, 범위, 그룹화 변수의 첫 번째 범주에 대한 변수값, 그룹화 변수의 마지막 범주에 대한 변수값, 표준 편차, 분산, 첨도, 첨도의 표준 오차, 왜도, 왜도의 표준 오차, 총 합계 %, 전체 *N* 퍼센트, 합계 %, *N* 퍼센트, 기하평균, 조화 평균을 구할 수 있습니다. 옵션에는 분산 분석, 에타, 에타제곱, 선형성 검정 *R*과 R^2이 있습니다.

그룹별 평균분석 데이터 고려 사항

데이터. 종속변수는 양적변수이고 독립변수는 범주형 변수입니다. 범주형 변수의 값은 숫자 또는 문자입니다.

가정. 일부 선택적 하위 그룹 통계(예: 평균과 표준 편차)는 정규 이론을 기초로 하며 대칭 분포를 가진 양적변수에 적절합니다. 중위수 등과 같은 강력한 통계는 정규성 가정을 충족시키거나 충족시키지 않을 수 있는 양적변수에 적절합니다. 분산 분석은 정규성으로부터 시작되어야 하나 각 셀의 데이터는 대칭적입니다. 분산 분석은 그룹이 동일한 분산을 가진 모집단으로부터 작성되었다고 가정합니다. 이러한 가정을 검정하려면 일원 분산 분석 프로시저에서 Levene 분산의 동질성 검정을 사용합니다.

하위 그룹 평균 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 평균...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 종속변수를 선택합니다.
3. 다음 방법 중 하나를 사용하여 범주형 독립변수를 선택합니다.
   - 하나 이상의 독립변수를 선택합니다. 각 결과가 종속변수마다 구분되어 표시됩니다.
   - 독립변수의 여러 레이어를 선택합니다. 각 레이어는 표본을 나눕니다. 레이어 1과 레이어 2에 독립변수가 하나씩 있을 때 결과는 각 독립변수에 대한 별도의 테이블과는 달리 하나의 교차 분석표에 표시됩니다.
4. 선택에 따라서는 **옵션**을 클릭하여 각 통계, 분산 분석표 및 에타, 에타제곱, *R*, R^2을 선택합니다.

그룹별 평균분석: 옵션

각 그룹화 변수의 각 범주 내에 있는 변수에 대해 다음 하위 그룹 통계량인 합계, 케이스 수, 평균, 중위수, 그룹 중위수, 평균의 표준 오차, 최소값, 최대값, 범위, 그룹화 변수의 첫 번째 범주에 대한 변수값, 그룹화 변수의 마지막 범주에 대한 변수값, 표준 편차, 분산, 첨도, 첨도의 표준 오차, 왜도, 왜도의 표준 오차, 총 합계 %, 전체 *N* 퍼센트, 합계 %, *N* 퍼센트, 기하평균, 조화 평균 중에서 하나를 선택할 수 있습니다. 하위 그룹 통계의 표시 순서를 변경할 수 있습니다. 셀 통계 목록에 통계가 나타나는 순서는 결과에 표시되는 순서입니다. 모든 범주의 각 변수에 대한 요약 통계도 함께 표시됩니다.

처음*(First)*. 데이터 파일에서 처음 나타나는 데이터 값을 표시합니다.

기하평균*(Geometric Mean)*. 데이터 값 간의 곱에 대한 n제곱근입니다. 여기서 n은 케이스 수를 나타냅니다.

그룹 중앙값*(Grouped Median)*. 그룹으로 코딩된 데이터에 대해 계산된 중앙값입니다. 예를 들어, 나이 데이터에서 30대의 각 값은 35로 코딩되고 40대의 각 값은 45로 코딩되는 경우 그룹 중위수는 코딩된 데이터에 대해 계산된 중위수입니다.

조화 평균*(harmonic mean)*. 그룹의 표본 결과가 같지 않을 때 평균 그룹 크기를 추정하는 데 사용됩니다. 조화 평균은 전체 표본 수를 표본 결과의 역수의 합으로 나눈 값입니다.

첨도*(Kurtosis)*. 이상치가 있는 범위의 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 데이터가 정규 분포보다 극단적인 이상값을 나타냄을 표시합니다. 음수인 첨도는 데이터가 정규 분포보다 극단적인 이상치가 적음을 나타냅니다. 정규 분포에 대해 값이 0인 경우에 사용되는 첨도의 정의를 때때로 초과 첨도라고 합니다. 일부 소프트웨어는 정규 분포에 대해 값이 3인 첨도를 보고할 수 있습니다.

마지막*(Last)*. 데이터 파일에서 마지막으로 나타나는 데이터 값을 표시합니다.

최대. 숫자변수의 가장 큰 값입니다.

평균. 중심 경향 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

중앙값*(Median)*. 전체 사례의 절반이 그 값보다 크거나 작거나 큰 값, 즉 50분위수. 케이스 수가 짝수인 경우 중위수는 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중위수는 평균과 달리 중심을 벗어난 값에는 영향을 받지 않는 중심 경향 측도이며, 상한 극단값 또는 하한 극단값에 따라 달라질 수 있습니다.

최소값*(minimum)*. 숫자변수의 가장 작은 값입니다.

N. 케이스 수(관측값 또는 레코드)입니다.

합계 *N*의 퍼센트*(Percent of total N)*. 각 범주에서 전체 케이스 수의 백분율입니다.

총 합계의 퍼센트*(Percent of total sum)*. 각 범주에서 총 합계의 백분율입니다.

Range. 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.

왜도*(Skewness)*. 분포의 비대칭에 대한 측도입니다. 정규 분포는 대칭적이며 왜도 값이 0입니다. 유의한 양의 왜도가 포함된 분포는 오른쪽 꼬리가 더 깁니다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.

표준 편차. 평균에 대한 산포 측도입니다. 정규 분포에서 케이스의 68%는 평균의 표준 편차 내에 있으며 케이스의 95%는 2배 표준 편차 내에 있습니다. 예를 들어, 평균 연령이 45세이고 표준 편차가 10인 경우 정규 분포 내에서 95% 케이스는 25세와 65세 사이에 있습니다.

첨도의 표준 오차*(Standard Error of Kurtosis)*. 표준 오차에 대한 첨도의 비율을 정상성 검사로 사용할 수 있습니다(즉, 비율이 -2 미만이거나 +2보다 크면 정상성을 거부할 수 있습니다). 첨도가 높은 양의 값인 경우 분포의 양 끝이 정규 분포의 양끝보다 길어지고 음의 값인 경우 양끝이 짧아집니다(상자 형태 균일 분포와 유사).

평균의 표준 오차*(Standard Error of Mean)*. 동일 분포로부터 선택한 표본 간에 발생할 수 있는 평균값의 차이에 대한 측도입니다. 이 값을 사용하여 관측 평균과 가설 값을 간략하게 비교할 수 있습니다. 즉, 표준 오차에 대한 차이 비율이 ?2보다 작거나 +2보다 큰 경우 두 값이 다르다고 판단할 수 있습니다.

왜도의 표준 오차*(Standard Error of Skewness)*. 표준 오차에 대한 왜곡의 비율을 정상성 검사로 사용할 수 있습니다(즉, 이 비율이 -2 보다 작거나 +2보다 크면 정상성을 거부할 수 있습니다). 왜도가 큰 양의 값인 경우 오른쪽이 길어지고 큰 음의 값인 경우 왼쪽이 길어집니다.

Sum. 비결측값을 갖는 전체 케이스 값의 총계입니다.

분산*(Variance)*. 평균에 대한 산포 측도값으로, 평균으로부터의 제곱 편차 합계를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.

첫번째 레이어에 대한 통계

아노바 테이블과 에타. 첫 번째 레이어의 각 독립변수에 대해 일원 분산 분석표를 표시하고 에타 및 에타제곱(연관 측도)을 계산합니다.

선형성 시험. 선형 및 비선형 성분과 관련된 제곱합, 자유도 및 평균 제곱과 F 비율, R및 R 제곱을 계산합니다. 독립변수가 단문자열인 경우 선형성은 계산되지 않습니다.

OLAP 큐브

OLAP(Online Analytical Processing) 큐브 프로시저는 하나 이상의 범주형 그룹화 변수 범주 내에서 연속형 요약변수의 합계, 평균 및 기타 일변량 통계량을 계산합니다. 각 그룹화 변수의 각 범주에 대해 테이블의 별도 레이어가 작성됩니다.

예. 다양한 영역 및 영역 내의 제품군에 대한 합계 및 평균 판매액을 구합니다.

통계. 합계, 케이스 수, 평균, 중위수, 그룹 중위수, 평균의 표준 오차, 최소값, 최대값, 범위, 그룹화 변수의 처음 범주에 대한 변수값, 그룹화 변수의 마지막 범주에 대한 변수값, 표준 편차, 분산, 첨도, 첨도의 표준 오차, 왜도, 왜도의 표준 오차, 전체 케이스 퍼센트, 총 합계 퍼센트, 그룹화 변수내의 전체 케이스 퍼센트, 그룹화 변수 내의 총 합계 퍼센트, 기하평균 및 조화 평균을 선택할 수 있습니다.

OLAP 큐브 데이터 고려 사항

데이터. 요약변수는 양적(구간 또는 비율 척도에서 측정되는 연속형 변수)이며 그룹화 변수는 범주형입니다. 범주형 변수의 값은 숫자 또는 문자입니다.

가정. 일부 선택적 하위 그룹 통계(예: 평균과 표준 편차)는 정규 이론을 기초로 하며 대칭 분포를 가진 양적변수에 적절합니다. 중위수, 범위 등 강력한 통계는 정규성 가정을 충족시키거나 충족시키지 않을 수 있는 양적변수에 적절합니다.

OLAP 큐브 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **보고서** > **OLAP 큐브...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 연속형 요약변수를 선택합니다.
3. 범주형 그룹화 변수를 하나 이상 선택합니다.

또는 다음 작업을 수행합니다.

- 다양한 요약 통계를 선택합니다(**통계** 클릭). 요약 통계를 선택하기 전에 그룹화 변수를 하나 이상 선택해야 합니다.
- 그룹화 변수에 의해 정의되는 대응변수와 대응그룹 사이의 차이를 계산합니다(**차이** 클릭).
- 통계테이블 제목을 작성합니다(**제목** 클릭).
- 지정된 정수보다 작은 수를 숨기십시오. 숨겨진 값은 **<N**로 표시됩니다. 여기서 **N**은 지정된 정수입니다. 지정된 정수는 2보다 크거나 같아야 합니다.

OLAP 큐브 통계

각 그룹화 변수의 각 범주 내에 있는 요약 변수에 대해 다음 하위 그룹 통계량인 합계, 케이스 수, 평균, 중위수, 그룹 중위수, 평균의 표준 오차, 최소값, 최대값, 범위, 그룹화 변수의 첫 번째 범주에 대한 변수값, 그룹화 변수의 마지막 범주에 대한 변수값, 표준 편차, 분산, 첨도, 첨도의 표준 오차, 왜도, 왜도의 표준 오차, 전체 케이스 %, 총 합계 %,그룹화 변수내 전체 케이스 %, 그룹화 변수내 총 합계 수 %, 기하평균 및 조화 평균을 선택할 수 있습니다.

하위 그룹 통계의 표시 순서를 변경할 수 있습니다. 셀 통계 목록에 통계가 나타나는 순서는 결과에 표시되는 순서입니다. 모든 범주의 각 변수에 대한 요약 통계도 함께 표시됩니다.

처음*(First)*. 데이터 파일에서 처음 나타나는 데이터 값을 표시합니다.

기하평균*(Geometric Mean)*. 데이터 값 간의 곱에 대한 n제곱근입니다. 여기서 n은 케이스 수를 나타냅니다.

그룹 중앙값*(Grouped Median)*. 그룹으로 코딩된 데이터에 대해 계산된 중앙값입니다. 예를 들어, 나이 데이터에서 30대의 각 값은 35로 코딩되고 40대의 각 값은 45로 코딩되는 경우 그룹 중위수는 코딩된 데이터에 대해 계산된 중위수입니다.

조화 평균*(harmonic mean)*. 그룹의 표본 결과가 같지 않을 때 평균 그룹 크기를 추정하는 데 사용됩니다. 조화 평균은 전체 표본 수를 표본 결과의 역수의 합으로 나눈 값입니다.

첨도*(Kurtosis)*. 이상치가 있는 범위의 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 데이터가 정규 분포보다 극단적인 이상값을 나타냄을 표시합니다. 음수인 첨도는 데이터가 정규 분포보다 극단적인 이상치가 적음을 나타냅니다. 정규 분포에 대해 값이 0인 경우에 사용되는 첨도의 정의를 때때로 초과 첨도라고 합니다. 일부 소프트웨어는 정규 분포에 대해 값이 3인 첨도를 보고할 수 있습니다.

마지막*(Last)*. 데이터 파일에서 마지막으로 나타나는 데이터 값을 표시합니다.

최대. 숫자변수의 가장 큰 값입니다.

평균. 중심 경향 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

중앙값*(Median)*. 전체 사례의 절반이 그 값보다 크거나 작거나 큰 값, 즉 50분위수. 케이스 수가 짝수인 경우 중위수는 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중위수는 평균과 달리 중심을 벗어난 값에는 영향을 받지 않는 중심 경향 측도이며, 상한 극단값 또는 하한 극단값에 따라 달라질 수 있습니다.

최소값*(minimum)*. 숫자변수의 가장 작은 값입니다.

N. 케이스 수(관측값 또는 레코드)입니다.

범주내 케이스 수 *%(Percent of N in)*. 다른 그룹화 변수의 범주 내에서 지정한 그룹화 변수의 케이스 수 퍼센트입니다. 그룹화 변수가 하나만 있는 경우 이 값은 전체 케이스 수 퍼센트와 동일합니다.

합계 %.다른 그룹화 변수의 범주 내에서 지정한 그룹화 변수의 합계 백분율입니다. 그룹화 변수가 하나만 있는 경우 이 값은 총 합계 퍼센트와 동일합니다.

전체 케이스 수 %. 각 범주에서 전체 케이스 수의 백분율입니다.

총 합계 백분율. 각 범주에서 총 합계의 백분율입니다.

Range. 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.

왜도*(Skewness)*. 분포의 비대칭에 대한 측도입니다. 정규 분포는 대칭적이며 왜도 값이 0입니다. 유의한 양의 왜도가 포함된 분포는 오른쪽 꼬리가 더 깁니다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.

표준 편차. 평균에 대한 산포 측도입니다. 정규 분포에서 케이스의 68%는 평균의 표준 편차 내에 있으며 케이스의 95%는 2배 표준 편차 내에 있습니다. 예를 들어, 평균 연령이 45세이고 표준 편차가 10인 경우 정규 분포 내에서 95% 케이스는 25세와 65세 사이에 있습니다.

첨도의 표준 오차*(Standard Error of Kurtosis)*. 표준 오차에 대한 첨도의 비율을 정상성 검사로 사용할 수 있습니다(즉, 비율이 -2 미만이거나 +2보다 크면 정상성을 거부할 수 있습니다). 첨도가 높은 양의 값인 경우 분포의 양끝이 정규 분포의 양끝보다 길어지고 음의 값인 경우 양끝이 짧아집니다(상자 형태 균일 분포와 유사).

평균의 표준 오차*(Standard Error of Mean)*. 동일 분포로부터 선택한 표본 간에 발생할 수 있는 평균값의 차이에 대한 측도입니다. 이 값을 사용하여 관측 평균과 가설 값을 간략하게 비교할 수 있습니다. 즉, 표준 오차에 대한 차이 비율이 ?2보다 작거나 +2보다 큰 경우 두 값이 다르다고 판단할 수 있습니다.

왜도의 표준 오차*(Standard Error of Skewness)*. 표준 오차에 대한 왜곡의 비율을 정상성 검사로 사용할 수 있습니다(즉, 이 비율이 -2 보다 작거나 +2보다 크면 정상성을 거부할 수 있습니다). 왜도가 큰 양의 값인 경우 오른쪽이 길어지고 큰 음의 값인 경우 왼쪽이 길어집니다.

Sum. 비결측값을 갖는 전체 케이스 값의 총계입니다.

분산*(Variance)*. 평균에 대한 산포 측도값으로, 평균으로부터의 제곱 편차 합계를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.

OLAP 큐브 차이

이 대화 상자에서는 그룹화 변수에 의해 정의된 그룹 간 또는 요약변수 간 산술 차이 및 퍼센트를 계산할 수 있습니다. OLAP 큐브 통계 대화 상자에서 선택한 모든 측도에 대해 차이가 계산됩니다.

변수 간의 차이. 대응변수 간 차이를 계산합니다. 각 대응에 있는 두 번째 변수의 요약 통계 값을 대응의 첫 번째 변수의 요약 통계 값에서 뺍니다. 퍼센트 차분의 경우 변수 빼기의 요약변수 값이 분모로 사용됩니다. 변수 간의 차이를 지정하려면 먼저 기본 대화 상자에서 요약변수를 두 개 이상 선택해야 합니다.

케이스 그룹 간의 차이. 그룹화 변수에서 정의한 대응 그룹 사이의 차이를 계산합니다. 각 대응에 있는 두 번째 범주(범주 빼기)의 요약 통계 값을 대응의 첫 번째 범주에 대한 요약 통계 값에서 뺍니다. 퍼센트 차분은 범주 빼기의 요약 통계 값을 분모로 사용합니다. 그룹 간의 차이를 지정하려면 먼저 기본 대화 상자에서 그룹화 변수를 하나 이상 선택해야 합니다.

OLAP 큐브 제목

결과의 제목을 변경하거나 결과표 아래에 나타날 캡션을 추가할 수 있습니다. 텍스트에 줄 바꿈 삽입 위치에 \n을 입력하여 제목과 캡션에서 줄바꿈을 제어할 수 있습니다.

비율

비율 소개

비율은 이항 비율 또는 비율 차분에 대한 검정 및 신뢰구간을 계산합니다. 일표본 비율(지정된 값에 대해 검정됨), 대응표본(다른 변수) 또는 독립표본(다른 케이스 집단)에 대한 통계량을 사용할 수 있습니다. 검정 통계량 및 신뢰구간의 유형에 대한 다양한 옵션이 제공됩니다. 동일한 기능 중 일부를 제공하는 다른 절차로는 **CROSSTABS**, **NPAR TESTS**, **NPTESTS** 등이 있습니다.

일표본 비율
일표본 검정 및 신뢰구간입니다. 결과에는 관측 비율, 모집단 비율 및 가설 모집단 비율의 차분 추정값, 영가설 및 대립가설 하에서 근사 표준오차, 양측 확률이 있는 지정된 검정 통계량, 비율에 대한 지정된 신뢰구간이 포함됩니다.

대응표본 비율
비율 간 차분에 대한 대응표본 검정 및 신뢰구간입니다. 결과에는 관측 비율, 모집단 비율 차분 추정값, 영가설 및 대립가설 하에서 모집단 차분의 근사 표준오차, 양측 확률이 있는 지정된 검정 통계량, 비율 차분에 대한 지정된 신뢰구간이 포함됩니다.

독립표본 비율
독립표본 검정 및 신뢰구간입니다. 결과에는 관측 비율, 모집단 비율 차분 추정값, 영가설 및 대립가설 하에서 모집단 차분의 근사 표준오차, 양측 확률이 있는 지정된 검정 통계량, 비율 차분에 대한 지정된 신뢰구간이 포함됩니다.

일표본 비율

일표본 비율 프로시저는 개별 이항 비율에 대한 검정 및 신뢰구간을 제공합니다. 데이터는 단순 무작위 표본에서 추출된 것으로 가정하고 각 가설검정 또는 신뢰구간은 이항 비율을 기준으로 하는 별도의 검정 또는 개별 구간입니다. 결과에는 관측 비율, 모집단 비율 및 가설 모집단 비율의 차분 추정값, 영가설 및 대립가설 하에서 근사 표준오차, 양측 확률이 있는 지정된 검정 통계량, 비율에 대한 지정된 신뢰구간이 포함됩니다.

예

통계
Agresti-Coull, Anscombe, Clopper-Pearson(정확), 제프리스, 로짓, Wald, Wald(연속수정), Wilson Score, Wilson Score(연속수정), 정확검정 이항, 중간 확률 수정된 이항, 점수, 점수(연속수정)

데이터 고려사항

데이터
이 프로시저는 요청된 검정 통계량과 양측 확률, 비율 차분에 대한 신뢰구간, 각 집단 또는 변수에 대한 비율, 표준오차 및 개수를 표시합니다. 이 프로시저는 최대 한 개의 검정값으로 제한됩니다.

가정

일표본 비율 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 일표본 비율...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 양적 검정변수를 하나 이상 선택합니다.
3. 선택적으로 다음을 수행할 수 있습니다.
   - **성공 정의** 섹션 아래에서 성공 기준 설정을 선택합니다.

     마지막 값
     데이터의 정렬된 유일값 중에서 마지막 값 또는 가장 높은 값이 사용됩니다. 이 설정은 숫자 또는 문자변수에 적용됩니다. 기본 설정입니다.

첫 번째 값
데이터의 정렬된 유일값 중에서 첫 번째 값 또는 가장 낮은 값이 사용됩니다. 이 설정은 숫자 또는 문자 변수에 적용됩니다.

값
괄호로 묶인 하나 이상의 특정 값입니다. 다중 값은 공백으로 구분해야 합니다. 이 설정은 숫자 또는 문자변수에 적용됩니다. 문자변수 값은 작은따옴표로 묶어야 합니다.

중심점
데이터 관측값 범위의 중간 이상에 있는 값입니다. 이 설정은 숫자 데이터에만 적용됩니다.

절단점
지정된 값 이상에 있는 값입니다. 이 설정은 숫자 데이터에만 적용됩니다.

- **신뢰 구간을 클릭하세요...** 를 사용하여 표시할 신뢰 구간 유형을 지정하거나 모든 신뢰 구간을 표시하지 않도록 설정할 수 있습니다.
- 클릭 **테스트...** 를 사용하여 표시할 테스트 통계 유형을 지정하거나 모든 테스트를 표시하지 않도록 설정할 수 있습니다.
- **누락된 값...을** 클릭하세요. 를 사용하여 누락된 데이터의 처리를 제어할 수 있습니다.
- **부트스트랩을 클릭하세요...** 를 사용하여 평균, 중앙값, 비율, 확률 비율, 상관 계수 또는 회귀 계수와 같은 추정치에 대한 표준 오차 및 신뢰 구간의 견고한 추정치를 도출할 수 있습니다.

4. **확인**을 클릭하십시오.

일표본 비율: 신뢰구간

신뢰구간 대화 상자는 적용 범위 수준을 지정하고 표시할 신뢰구간의 유형을 선택하는 옵션을 제공합니다.

적용 범위 수준
신뢰구간 퍼센트를 지정합니다. 범위(0,100) 내에 있는 숫자 값을 지정해야 합니다. 기본 설정은 95입니다.

구간 유형
표시할 신뢰구간의 유형을 지정하는 옵션을 제공합니다. 사용가능한 옵션은 다음과 같습니다.

- Agresti-Coull
- Anscombe
- Clopper-Pearson(정확)
- 제프리스
- 로짓
- Wald
- Wald(연속수정)
- Wilson Score
- Wilson Score(연속수정)

일표본 비율에 대한 신뢰구간 지정

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 일표본 비율...

2. 표시할 신뢰구간의 유형을 지정하거나 모든 신뢰구간을 표시하지 않으려면 **신뢰구간**을 클릭합니다.

일표본 비율: 검정

검정 대화 상자는 표시할 검정 통계량의 유형을 지정하는 옵션을 제공합니다.

모두
모든 검정 통계량이 결과에 표시됩니다.

없음
검정 통계량이 결과에 표시되지 않습니다.

정확검정 이항
정확검정 이항 확률을 표시합니다.

중간 확률 수정된 이항
중간 확률 수정된 이항 확률을 표시합니다. 기본 설정입니다.

점수
점수 Z 검정 통계량을 표시합니다. 기본 설정입니다.

점수(연속수정)
연속수정 점수 Z 검정 통계량을 표시합니다.

Wald
Wald Z 검정 통계량을 표시합니다.

Wald(연속수정)
연속수정 Wald Z 검정 통계량을 표시합니다.

검정값
0과 1사이의 검정 값을 지정합니다. 기본값은 0.5입니다.

일표본 비율 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 일표본 비율...
2. 일표본 비율 대화 상자에서 **검정**을 클릭합니다.
3. 사용가능한 검정을 한 개 이상 선택합니다.

일표본 비율: 결측값

결측값 대화 상자는 결측값 처리 옵션을 제공합니다.

결측 데이터 범위

분석별 결측값 제외
특정 분석 각각에 사용된 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함함을 나타냅니다. 기본 설정입니다.

목록별 결측값 제외
모든 분석에 사용된 모든 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함함을 나타냅니다.

사용자 결측값
제외는 사용자 결측값을 결측 데이터로 처리합니다. **포함**은 사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 데이터로 처리합니다.

일표본 비율에 대한 결측값 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 일표본 비율...
2. 일표본 비율 대화 상자에서 **결측값**을 클릭합니다.
3. 원하는 결측값 설정을 선택합니다.

대응표본 비율

대응표본 비율 프로시저는 두 개의 관련 또는 대응 이항 비율 차분에 대한 검정 및 신뢰구간을 제공합니다. 데이터는 단순 무작위 표본에서 추출된 것으로 가정하고 각 가설검정 또는 신뢰구간은 별도의 검정 또는 개별 구간입니다. 결과에는 관측 비율, 모집단 비율 차분 추정값, 영가설 및 대립가설 하에서 모집단 차분의 근사 표준오차, 양측 확률이 있는 지정된 검정 통계량, 비율 차분에 대한 지정된 신뢰구간이 포함됩니다.

예

통계

Agresti-Min, Bonett-Price, Newcombe, Wald, Wald(연속수정), 정확검정 이항, 중간 확률 수정된 이항, McNemar, McNemar(연속수정)

데이터 고려 사항

데이터

- 최소 두 개의 변수가 포함된 변수 목록이 필요합니다.
- 단일 변수 목록을 지정한 경우 각 멤버가 목록의 다른 모든 멤버와 쌍을 이룹니다.

가정

- 두 개의 변수 목록이 (**PAIRED**) 키워드 없이 **WITH**로 구분되는 경우, 첫 번째 목록의 각 멤버는 두 번째 목록의 각 멤버와 쌍을 이룹니다.
- 두 개의 변수 목록이 **WITH** 로 구분되고 두 번째 목록이 다음에(**PAIRED**) 오는 경우, 순서대로 두 목록의 멤버가 쌍을 이루게 됩니다. 첫 번째 목록의 첫 번째 멤버가 두 번째 목록의 첫 번째 멤버와 쌍을 이루고 각 목록의 두 번째 멤버들이 쌍으로 구성됩니다. 일치하지 않는 변수는 무시되고 경고 메시지가 발행됩니다.

대응표본 비율 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 대응표본 비율...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 양적 검정변수를 하나 이상 선택합니다.
3. 선택적으로 다음을 수행할 수 있습니다.
   - **성공 정의** 섹션 아래에서 성공 기준 설정을 선택합니다.

     마지막 값
     데이터의 정렬된 유일값 중에서 마지막 값 또는 가장 높은 값이 사용됩니다. 이 설정은 숫자 또는 문자변수에 적용됩니다. 기본 설정입니다.

     첫 번째 값
     데이터의 정렬된 유일값 중에서 첫 번째 값 또는 가장 낮은 값이 사용됩니다. 이 설정은 숫자 또는 문자변수에 적용됩니다.

     값
     괄호로 묶인 하나 이상의 특정 값입니다. 다중 값은 공백으로 구분해야 합니다. 이 설정은 숫자 또는 문자변수에 적용됩니다. 문자변수 값은 작은따옴표로 묶어야 합니다.

     중심점
     데이터 관측값 범위의 중간 이상에 있는 값입니다. 이 설정은 숫자 데이터에만 적용됩니다.

     절단점
     지정된 값 이상에 있는 값입니다. 이 설정은 숫자 데이터에만 적용됩니다.
   - **신뢰 구간을 클릭하세요...** 를 사용하여 표시할 신뢰 구간 유형을 지정하거나 모든 신뢰 구간을 표시하지 않도록 설정할 수 있습니다.
   - 클릭 **테스트...** 를 사용하여 표시할 테스트 통계 유형을 지정하거나 모든 테스트를 표시하지 않도록 설정할 수 있습니다.
   - **누락된 값...을** 클릭하세요. 를 사용하여 누락된 데이터의 처리를 제어할 수 있습니다.
   - **부트스트랩을 클릭하세요...** 를 사용하여 평균, 중앙값, 비율, 확률 비율, 상관 계수 또는 회귀 계수와 같은 추정치에 대한 표준 오차 및 신뢰 구간의 견고한 추정치를 도출할 수 있습니다.
4. **확인**을 클릭하십시오.

대응표본 비율: 신뢰구간

신뢰구간 대화 상자는 적용 범위 수준을 지정하고 표시할 신뢰구간의 유형을 선택하는 옵션을 제공합니다.

적용 범위 수준
신뢰구간 퍼센트를 지정합니다. 범위(0,100) 내에 있는 숫자 값을 지정해야 합니다. 기본 설정은 95입니다.

구간 유형
표시할 신뢰구간의 유형을 지정하는 옵션을 제공합니다. 사용가능한 옵션은 다음과 같습니다.

- Agresti-Min
- Bonett-Price
- Newcombe
- Wald
- Wald(연속수정)

대응표본 비율에 대한 신뢰구간 지정

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 대응표본 비율...
2. 표시할 신뢰구간의 유형을 지정하거나 모든 신뢰구간을 표시하지 않으려면 **신뢰구간**을 클릭합니다.

대응표본 비율: 검정

검정 대화 상자는 표시할 검정 통계량의 유형을 지정하는 옵션을 제공합니다.

모두
모든 검정 통계량이 결과에 표시됩니다.

없음
검정 통계량이 결과에 표시되지 않습니다.

정확검정 이항
정확검정 이항 확률을 표시합니다.

중간 확률 수정된 이항
중간 확률 수정된 이항 확률을 표시합니다. 기본 설정입니다.

McNemar
McNemar Z 검정 통계량을 표시합니다. 기본 설정입니다.

McNemar(연속수정)
연속수정 McNemar Z 검정 통계량을 표시합니다.

Wald
Wald Z 검정 통계량을 표시합니다.

Wald(연속수정)
연속수정 Wald Z 검정 통계량을 표시합니다.

대응표본 비율 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 대응표본 비율...
2. 대응표본 비율 대화 상자에서 **검정**을 클릭합니다.
3. 사용가능한 검정을 한 개 이상 선택합니다.

대응표본 비율: 결측값

결측값 대화 상자는 결측값 처리 옵션을 제공합니다.

결측 데이터 범위

분석별 결측값 제외
특정 분석 각각에 사용된 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함함을 나타냅니다. 기본 설정입니다.

목록별 결측값 제외
모든 분석에 사용된 모든 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함함을 나타냅니다.

사용자 결측값
제외는 사용자 결측값을 결측 데이터로 처리합니다. **포함**은 사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 데이터로 처리합니다.

대응표본 비율에 대한 결측값 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 대응표본 비율...
2. 대응표본 비율 대화 상자에서 **결측값**을 클릭합니다.
3. 원하는 결측값 설정을 선택합니다.

독립표본 비율

독립표본 비율 프로시저는 두 개의 독립 이항 비율 차분에 대한 검정 및 신뢰구간을 제공합니다. 데이터는 단순 무작위 표본에서 추출된 것으로 가정하고 각 가설검정 또는 신뢰구간은 별도의 검정 또는 개별 구간입니다. 결과에는 관측 비율, 모집단 비율 차분 추정값, 영가설 및 대립가설 하에서 모집단 차분의 근사 표준오차, 양측 확률이 있는 지정된 검정 통계량, 비율 차분에 대한 지정된 신뢰구간이 포함됩니다.

예

통계
Agresti-Min, Bonett-Price, Newcombe, Wald, Wald(연속수정), 정확검정 이항, 중간 확률 수정된 이항, McNemar, McNemar(연속수정)

데이터 고려 사항

데이터

- 비교할 두 집단을 식별하기 위한 하나 이상의 종속변수와 하나의 변수가 필요합니다.
- 집단변수는 숫자 또는 문자일 수 있습니다.

가정

독립표본 비율 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 독립 표본 비율...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 양적 검정변수를 하나 이상 선택합니다.
3. 비교할 두 집단을 식별하는 **집단변수**를 한 개 선택합니다.
4. 선택적으로 선택한 **집단변수**에 대한 설정을 지정합니다.
   - **값**을 선택한 경우, 비교할 값으로 두 개의 숫자 또는 문자 값을 괄호 안에 지정할 수 있습니다. 문자 값은 작은따옴표로 묶어야 합니다. 다른 값을 갖는 케이스는 무시됩니다.
   - **중심점**은 숫자변수에만 적용됩니다. 집단변수 분포의 중심점 위치 또는 그 위에 있는 케이스는 두 번째 집단에 지정되고 중심점 아래에 있는 케이스는 첫 번째 집단에 지정됩니다.
   - **절단점**은 숫자변수에만 적용되고 단일 숫자 값을 괄호로 묶어 지정할 수 있습니다. 집단변수의 절단점 위치 또는 그 위에 있는 케이스는 두 번째 집단에 지정되고 절단점 아래에 있는 케이스는 첫 번째 집단에 지정됩니다.
5. 선택적으로 다음을 수행할 수 있습니다.
   - **성공 정의** 섹션 아래에서 성공 기준 설정을 선택합니다.

마지막 값
데이터의 정렬된 유일값 중에서 마지막 값 또는 가장 높은 값이 사용됩니다. 이 설정은 숫자 또는 문자변수에 적용됩니다. 기본 설정입니다.

첫 번째 값
데이터의 정렬된 유일값 중에서 첫 번째 값 또는 가장 낮은 값이 사용됩니다. 이 설정은 숫자 또는 문자변수에 적용됩니다.

값
괄호로 묶인 하나 이상의 특정 값입니다. 다중 값은 공백으로 구분해야 합니다. 이 설정은 숫자 또는 문자변수에 적용됩니다. 문자변수 값은 작은따옴표로 묶어야 합니다.

중심점
데이터 관측값 범위의 중간 이상에 있는 값입니다. 이 설정은 숫자 데이터에만 적용됩니다.

절단점
지정된 값 이상에 있는 값입니다. 이 설정은 숫자 데이터에만 적용됩니다.

- **신뢰 구간을 클릭하세요...** 를 사용하여 표시할 신뢰 구간 유형을 지정하거나 모든 신뢰 구간을 표시하지 않도록 설정할 수 있습니다.
- 클릭 **테스트...** 를 사용하여 표시할 테스트 통계 유형을 지정하거나 모든 테스트를 표시하지 않도록 설정할 수 있습니다.
- **누락된 값...을** 클릭하세요. 를 사용하여 누락된 데이터의 처리를 제어할 수 있습니다.
- **부트스트랩을 클릭하세요...** 를 사용하여 평균, 중앙값, 비율, 확률 비율, 상관 계수 또는 회귀 계수와 같은 추정치에 대한 표준 오차 및 신뢰 구간의 견고한 추정치를 도출할 수 있습니다.

6. **확인**을 클릭하십시오.

독립표본 비율: 신뢰구간

신뢰구간 대화 상자는 적용 범위 수준을 지정하고 표시할 신뢰구간의 유형을 선택하는 옵션을 제공합니다.

적용 범위 수준
신뢰구간 퍼센트를 지정합니다. 범위(0,100) 내에 있는 숫자 값을 지정해야 합니다. 기본 설정은 95입니다.

구간 유형
표시할 신뢰구간의 유형을 지정하는 옵션을 제공합니다. 사용가능한 옵션은 다음과 같습니다.

- Agresti-Caffo
- Brown-Li-Jeffreys
- Hauck-Anderson
- Newcombe
- Newcombe (Continuity Corrected)
- Wald
- Wald(연속수정)

독립표본 비율에 대한 신뢰구간 지정

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 독립 표본 비율...
2. 표시할 신뢰구간의 유형을 지정하거나 모든 신뢰구간을 표시하지 않으려면 **신뢰구간**을 클릭합니다.

독립표본 비율: 검정

검정 대화 상자는 표시할 검정 통계량의 유형을 지정하는 옵션을 제공합니다.

모두
모든 검정 통계량이 결과에 표시됩니다.

없음
검정 통계량이 결과에 표시되지 않습니다.

Hauck-Anderson
Hauck-Anderson Z 검정 통계량을 표시합니다.

Wald
Wald Z 검정 통계량을 표시합니다.

Wald(연속수정)
연속수정 Wald Z 검정 통계량을 표시합니다.

Wald H0
H_0에서 분산추정값을 사용하여 Wald Z 검정 통계량을 표시합니다.

Wald H0(연속수정)
H_0에서 분산추정값을 사용하여 연속수정 Wald Z 검정 통계량을 표시합니다.

독립표본 비율 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 독립 표본 비율...
2. 독립표본 비율 대화 상자에서 **검정**을 클릭합니다.
3. 사용가능한 검정을 한 개 이상 선택합니다.

독립표본 비율: 결측값

결측값 대화 상자는 결측값 처리 옵션을 제공합니다.

결측 데이터 범위

분석별 결측값 제외
특정 분석 각각에 사용된 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함함을 나타냅니다. 기본 설정입니다.

목록별 결측값 제외
모든 분석에 사용된 모든 변수에 대한 데이터가 충분히 있는 모든 케이스를 포함함을 나타냅니다.

사용자 결측값
제외는 사용자 결측값을 결측 데이터로 처리합니다. **포함**은 사용자 결측값 지정을 무시하고 사용자 결측값을 유효한 데이터로 처리합니다.

독립표본 비율에 대한 결측값 설정 정의

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 독립 표본 비율...
2. 독립표본 비율 대화 상자에서 **결측값**을 클릭합니다.
3. 원하는 결측값 설정을 선택합니다.

T 검정

T 검정

세 가지 유형의 *t* 검정을 사용할 수 있습니다.

독립 표본 T 검정(2-표본 T 검정). 두 케이스 그룹에 대해 한 변수의 평균을 비교합니다. 변수의 등식에 대한 각 그룹과 Levene 검정의 기술통계는 물론 평균 차이에 대한 등분산과 비-등분산 *T* 값과 95% 신뢰구간이 제공됩니다.

대응표본 T 검정(독립 T 검정). 단일 그룹에 대해 두 변수의 평균을 비교합니다. 이 검정은 대응표집 쌍이나 사례 통제 연구 계획용입니다. 결과에는 검정 변수에 대한 기술통계, 변수 사이의 상관, 쌍 차이대응에 대한 기술통계, *T* 검정 및 95% 신뢰구간이 포함됩니다.

일표본 T 검정. 한 변수의 평균을 알려진 값이나 가설 값과 비교합니다. 검정 변수의 기술통계가 *T* 검정과 함께 표시됩니다. 검정 변수의 평균과 가설 검정 값 사이의 차이에 대한 95% 신뢰구간이 기본값 결과의 쌍입니다.

독립 표본 T 검정

독립표본 T 검정 프로시저는 두 개의 케이스 집단에 대한 평균을 비교하고 *t* 검정 효과크기 계산을 자동화합니다. 이 테스트의 경우 피험자를 두 그룹에 무작위로 배정하여 반응의 차이가 다른 요인이 아닌 치료(또는 치료 부족)로 인한 것임을 확인하는 것이 가장 이상적입니다. 이것은 남자와 여자의 평균 수입을 비교하는 경우에 해당되지 않습니다. 사람을 무작위로 남자나 여자로 할당할 수 없기 때문입니다. 이러한 상황에서는 다른 요인의 차이가 평균에서 큰 차이가 나지 않게 하거나 차이를 크게 하는지 확인해야 합니다. 평균 수입의 차이는 성별의 영향만 받는 것이 아니라 교육수준과 같은 요인의 영향을 받을 수도 있습니다.

예

고혈압 환자를 무작위로 거짓 처방 그룹(위안을 주기 위해 효과 없는 약을 제공)과 처리 그룹으로 분류하였습니다. 거짓 처방 개체의 대상은 효과 없는 약을 복용하고 치료 개체의 대상은 혈압을 낮추는 것으로 기대되는 새로운 약을 복용하게 됩니다. 대상들을 두 달 동안 치료한 후 거짓 처방 개체와 처리 그룹의 평균 혈압을 비교하기 위해 2-표본 *t* 검정이 사용됩니다. 각 환자는 한 번만 측정하고 한 그룹에 속해 있습니다.

통계

각 변수에 대해 표본크기, 평균, 표준편차, 평균의 표준오차 및 *t* 검정에 대한 효과크기 추정을 선택할 수 있습니다. 평균의 차이에 대해 평균, 표준 오차, 신뢰구간(사용자가 신뢰수준을 지정할 수 있음)을 선택할 수 있습니다. 분산의 등식에 대한 Levene 검정, 평균의 등식에 대한 풀링-분산 및 개별-분산 *t* 검정을 선택할 수 있습니다.

데이터 고려 사항

데이터

중요한 양적변수의 값은 데이터 파일의 단일 열에 있습니다. 프로시저는 두 개의 값을 갖는 그룹화 변수를 사용하여 케이스를 두 그룹으로 분리합니다. 그룹화 변수는 숫자(예: 1과 2 또는 6.25와 12.5)이거나 단문자(예: *yes*와 *no*)가 될 수 있습니다. 다른 방법으로, 연령과 같은 양적변수를 사용할 수 있는데 이 경우 절단점을 지정하여 케이스를 두 그룹으로 분리할 수 있습니다(절단점이 21이면 연령이 21세 미만인 그룹과 21세 이상인 그룹으로 분리됨).

가정

등분산 *t* 검정의 경우 관측값은 동일한 모집단 분산을 갖는 정규 분포로부터 독립된 무작위 표본이어야 합니다. 비-등분산 *t* 검정의 경우 관측값은 정규 분포로부터 독립된 무작위 표본이어야 합니다. 2-표본 *t* 검정은 정규성을 상당히 벗어나 있습니다. 분포를 그래프상에서 확인하면 두 표본이 대칭적이며 이상값이 없음을 볼 수 있습니다.

독립표본 T 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 독립 표본 T 검정...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 양적 검정변수를 하나 이상 선택합니다. 각 변수에 대해 별도의 *t* 검정이 계산됩니다.
3. 단일 그룹화 변수를 선택한 후 **그룹 정의**를 클릭하여 비교하려는 그룹의 두 코드를 지정합니다.
4. 다음 설정은 선택 사항입니다.
   - **효과 크기 추정**을 선택하여 *t* 검정 효과 크기의 추정값을 제어하십시오.
   - **분산 동질성 테스트를** 선택하여 레빈 통계를 계산하여 그룹 분산이 동등한지 테스트합니다. 결과는 별도의 표로 확인할 수 있습니다.
   - 결측 데이터의 처리와 신뢰구간의 수준을 제어하려면 **옵션**을 클릭합니다.
   - 평균, 중위수, 비율, 오즈비, 상관계수 또는 회귀계수와 같은 추정에 대한 표준오차 및 신뢰구간의 강력한 추정을 파생시키려면 **붓스트랩**을 클릭합니다.

독립표본 T 검정 그룹 정의

숫자 집단변수의 경우, 두 개의 값 또는 하나의 절단점을 지정하여 *t* 검정에 대한 두 집단을 정의합니다.

- **지정값 사용.** 그룹 1의 값과 그룹 2의 다른 값을 입력하십시오. 기타 값이 있는 케이스는 분석에서 제외됩니다. 숫자가 정수일 필요는 없습니다(예: 6.25와 12.5 유효).
- **절단점.** 그룹화 변수의 값을 두 변수 세트로 분리하는 수를 입력합니다. 절단점보다 작은 값을 갖는 모든 케이스가 한 그룹을 이루며 절단점보다 큰 값을 갖는 케이스가 다른 그룹을 형성합니다.

문자열 그룹화 변수의 경우, 그룹 1의 문자열과 그룹 2의 다른 값(예: 예와 아니오)을 입력하십시오. 기타 문자열이 있는 케이스는 분석에서 제외됩니다.

독립 표본 T 검정: 옵션

신뢰구간. 기본적으로 평균의 차이에 대한 95% 신뢰구간이 표시됩니다. 다른 신뢰수준을 원하면 1부터 99 사이의 값을 입력합니다.

결측값. 여러 변수를 검정하고 한 개 이상의 변수에 대한 데이터가 없을 경우 어떤 케이스를 포함할 것인지(또는 제외할 것인지) 프로시저에 알려줍니다.

- **분석별 결측값 제외.** 각 *t* 검정에서는 검정된 변수의 데이터가 유효한 모든 케이스를 사용합니다. 표본 결과는 검정마다 다를 수 있습니다.
- **목록별 결측값 제외.** 각 *t* 검정에서는 요청된 *t* 검정에 사용한 모든 변수의 데이터가 유효한 케이스만 사용합니다. 표본 결과는 여러 검정에서 일정합니다.

대응표본 T 검정

대응표본 T 검정 프로시저는 한 그룹에 있는 두 변수의 평균을 비교합니다. 이 프로시저에서는 각 케이스에 나타나는 두 변수 값 사이의 차이를 계산하고 그 평균이 0이 아닌지를 검정합니다. 이 프로시저는 *t* 검정 효과 크기 계산도 자동화합니다.

예

고혈압 연구에서는 환자의 혈압을 연구 시작 단계에서 한 번 측정하여 처방을 내리고 그 후 혈압을 다시 측정하게 됩니다. 이 경우 각 개체는 사전 및 사후 측도값이라는 두 가지 측도값을 갖게 됩니다. 이 검정이 사용되는 또 다른 계획으로는 대응 연구 또는 사례 통제 연구가 있습니다. 이러한 연구의 경우 데이터 파일의 각 레코드에는 환자의 반응과 그 대응 통제 개체의 반응이 함께 포함됩니다. 고혈압 연구에서 환자와 통제 피실험자는 연령별로(75세 대조군 멤버와 75세 환자) 대응될 수 있습니다.

통계

각 변수에 대해 평균, 표본 결과, 표준 편차, 평균의 표준 오차를 선택할 수 있습니다. 각 대응변수에 대해 상관, 평균값의 평균차이, *t* 검정, 평균차이에 대한 신뢰구간(신뢰수준을 지정할 수 있음), *t* 검정에 대한 효과크기 추정을 선택할 수 있으며, 평균차의 표준 편차와 표준 오차를 선택할 수 있습니다.

데이터 고려사항

데이터

각 대응 검정에 대해 두 양적변수(측정의 구간 수준 또는 측정의 비율 수준)를 지정합니다. 대응 연구나 사례 통제 연구의 경우 각 검정 개체와 그 대응 통제 개체의 반응은 데이터 파일의 동일한 케이스에 나타나야 합니다.

가정

각 검정 피실험자와 통제 피실험자에 대한 관찰은 동일한 조건에서 이루어져야 합니다. 그 평균차는 정규 분포이어야 합니다. 각 변수의 분산은 같을 수도 있고, 서로 같지 않을 수 있습니다.

대응표본 T 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 대응표본 T 검정...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

2. 하나 이상의 대응변수를 선택합니다.
3. 선택적으로 **효과크기 추정** 옵션을 선택/변경합니다. 이 설정은 각 변수 쌍에 대해 Cohen의 *d* 및 Hedges 수정을 추정할 때 Standardizer의 계산 방법을 제어합니다.

   차이의 표준편차
   효과크기를 추정하는 데 사용되는 분모입니다. Cohen의 *d*에는 평균차이의 표본 표준편차가 사용됩니다. Hedges 수정에는 수정 요인으로 수정된 평균차이의 표본 표준편차가 사용됩니다.

   차이의 수정된 표준편차
   효과크기를 추정하는 데 사용되는 분모입니다. Cohen의 *d*에는 측도 간 상관계수로 수정된 평균차이의 표본 표준편차가 사용됩니다. Hedges 수정에는 측도 간 상관계수로 수정된 평균차이의 표본 표준편차와 수정 요인이 사용됩니다.

   분산의 평균
   효과크기를 추정하는 데 사용되는 분모입니다. Cohen의 *d*에는 측도 평균 분산의 제곱근이 사용됩니다. Hedges 수정에는 측도 평균 분산의 제곱근과 수정 요인이 사용됩니다.
4. 선택적으로 다음을 수행할 수 있습니다.
   - **효과 크기 추정**을 선택하여 *t* 검정 효과 크기의 추정값을 제어하십시오. 이 설정을 선택할 경우 각 변수 쌍에 대해 Cohen의 *d* 및 Hedges 수정을 추정할 때 Standardizer의 계산 방법을 추가로 제어할 수 있습니다.
   - 결측 데이터의 처리와 신뢰구간의 수준을 제어하려면 **옵션**을 클릭합니다.
   - 평균, 중위수, 비율, 오즈비, 상관계수 또는 회귀계수와 같은 추정에 대한 표준오차 및 신뢰구간의 강력한 추정을 파생시키려면 **붓스트랩**을 클릭합니다.

대응표본 T 검정: 옵션

신뢰구간. 기본적으로 평균의 차이에 대한 95% 신뢰구간이 표시됩니다. 다른 신뢰수준을 원하면 1부터 99 사이의 값을 입력합니다.

결측값. 여러 변수를 검정하고 한 개 이상의 변수에 대한 데이터가 없을 경우 어떤 케이스를 포함할 것인지(또는 제외할 것인지) 프로시저에 알려줍니다.

- **분석별 결측값 제외.** 각 *t* 검정에서는 검정된 대응변수의 데이터가 유효한 모든 케이스를 사용합니다. 표본 결과는 검정마다 다를 수 있습니다.
- **목록별 결측값 제외.** 각 *t* 검정에서는 검정된 모든 대응변수의 데이터가 유효한 케이스만 사용합니다. 표본 결과는 여러 검정에서 일정합니다.

T TEST 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 명령을 실행하여 일표본 및 독립 표본 검정을 생성합니다.
- PAIRS 하위 명령을 사용하여 대응 t 검정의 목록에서 각 변수에 대해 변수를 검정합니다.
- *t* 검정 효과크기의 추정을 제어합니다(ES 부명령문 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

일표본 T 검정

일표본 T 검정 프로시저는 단일 변수의 평균과 지정된 상수와의 차이를 검정하고 *t* 검정 효과크기 계산을 자동화합니다.

예
한 연구원은 학생 그룹의 평균 IQ 점수가 100점과 다른지 여부를 시험해 볼 수 있습니다. 또는 시리얼 제조업자가 생산 라인에서 박스들의 샘플을 가져와 샘플들의 평균 중량이 95% 신뢰 수준에서 1.3 파운드와 상이한지 여부를 확인할 수 있습니다.

통계
각 검정변수에 대해 평균, 표준편차, 평균의 표준오차 및 *t* 검정에 대한 효과크기 추정을 선택할 수 있습니다. 각 데이터 값과 가정된 검정값 사이의 평균 차이, 이 차이가 0인지를 검정하는 *t* 검정, 이 차이에 대한 신뢰구간(사용자가 신뢰수준 지정 가능)을 선택할 수 있습니다.

데이터 고려 사항

데이터
가정된 검정값과 비교하여 양적변수 값을 검정하려면 양적변수를 선택한 후 가정된 검정값을 입력합니다.

가정
이 검정은 데이터가 정규 분포라고 가정하지만 정규성을 상당히 벗어나 있습니다.

일표본 T 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 평균 비교 > 일표본 T 검정...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 동일한 가정값과 비교하여 검정변수를 하나 이상 선택합니다.
3. 각 표본 평균과 비교할 숫자 검정값을 입력합니다.
4. 선택적으로 다음을 수행할 수 있습니다.
   - **효과 크기 추정**을 선택하여 *t* 검정 효과 크기의 추정값을 제어하십시오.
   - 결측 데이터의 처리와 신뢰구간의 수준을 제어하려면 **옵션**을 클릭합니다.

일표본 T 검정: 옵션

신뢰구간. 기본적으로 평균과 가정된 검정값과의 차이에 대해 95% 신뢰구간이 표시됩니다. 다른 신뢰수준을 원하면 1부터 99 사이의 값을 입력합니다.

결측값. 여러 변수를 검정하고 한 개 이상의 변수에 대한 데이터가 없을 경우 어떤 케이스를 포함할 것인지(또는 제외할 것인지) 프로시저에 알려줍니다.

- **분석별 결측값 제외.** 각 *t* 검정에서는 검정된 변수의 데이터가 유효한 모든 케이스를 사용합니다. 표본 결과는 검정마다 다를 수 있습니다.
- **목록별 결측값 제외.** 각 *t* 검정은 요청된 *t* 검정에 사용된 모든 변수의 데이터가 유효한 케이스만 사용합니다. 표본 결과는 여러 검정에서 일정합니다.

T TEST 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 명령을 실행하여 일표본 및 독립 표본 검정을 생성합니다.
- PAIRS 하위 명령을 사용하여 대응 t 검정의 목록에서 각 변수에 대해 변수를 검정합니다.
- *t* 검정 효과크기의 추정을 제어합니다(ES 부명령문 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

T TEST 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 명령을 실행하여 일표본 및 독립 표본 검정을 생성합니다.
- PAIRS 하위 명령을 사용하여 대응 t 검정의 목록에서 각 변수에 대해 변수를 검정합니다.
- *t* 검정 효과크기의 추정을 제어합니다(ES 부명령문 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

일원 분산 분석

일원배치 분산분석 프로시저는 단일 요인(독립) 변수로 양적 종속변수에 대한 일원배치 분산분석을 작성하고 일원배치 분산분석으로 효과크기를 추정합니다. 분산 분석은 여러 평균이 동일하다는 가정을 검정하는 데 사용됩니다. 이 기법은 2-표본 *t* 검정을 확장한 것입니다.

평균 간에 차이가 있는지 판별할 수 있을 뿐 아니라 어떤 평균이 다른지를 확인할 수 있습니다. 평균을 비교하는 방법으로는 사전 대비와 사후 검정의 두 가지 유형의 검정이 있습니다. 대비법은 실험하기 전에 설정하는 검정이고 사후 검정은 실험 후에 실행하는 검정입니다. 범주 전반에서의 추세를 검정할 수도 있습니다.

예

도우넛을 만들 때 지방이 상당량 흡수됩니다. 실험은 세 가지 종류의 기름(땅콩 기름, 콩기름, 라드)을 포함하여 설정됩니다. 땅콩 기름과 콩기름은 불포화 지방이고 라드는 포화 지방입니다. 사용 기름의 종류에 따라 어떤 지방 성분이 흡수되는지를 확인하여 포화 지방과 불포화 지방 간의 흡수 정도를 보려면 사전 대비를 사용합니다.

통계

각 집단: 케이스 수, 평균, 표준 편차, 평균의 표준 오차, 최소값, 최대값, 평균에 대한 95% 신뢰구간, 일원배치 분산분석으로 효과크기 추정. 분산의 동질성에 대한 Levene 검정, 분산 분석 표 및 각 종속 변수에 대한 평균 등식의 로버스트 검정, 사용자가 지정한 사전 대비, 그리고 사후 범위 검정 및 다중 비교: Bonferroni, Sidak, Tukey의 정직 유의차, Hochberg의 GT2, Gabriel, Dunnett, Ryan-Einot-Gabriel-Welsch *F* 검정(R-E-G-W *F*), Ryan-Einot-Gabriel-Welsch 범위 검정(R-E-G-W *Q*), Tamhane의 T2, Dunnett의 T3, Games-Howell, Dunnett의 *C*, Duncan의 다중 범위 검정, Student-Newman-Keuls(S-N-K), Tukey의 *b*, Waller-Duncan, Scheffé 및 최소 유의차.

데이터 고려사항

데이터

종속변수는 양적변수여야 합니다(측정 구간 수준).

가정

각 그룹은 정규 모집단으로부터 독립 무작위 표본입니다. 분산 분석은 데이터가 대칭적인 경우에도 정규성에서 이탈해도 영향을 받지 않습니다. 그룹은 동분산으로 모집단에서 추출합니다. 이러한 가정을 검정하려면 Levene의 분산의 동질성 검정을 사용합니다.

일원배치 분산분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석하기 > 평균 비교 > 일원 분산 분석...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 종속변수를 선택합니다.
3. 독립 요인변수를 하나만 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 전체 검정에 대한 효과크기 계산을 제어하려면 **전체 검정에 대한 효과크기 추정**을 선택합니다. 선택할 경우 출력결과에 "ANOVA 효과크기" 표가 표시됩니다.
- 그룹 간 제곱합을 추세 성분으로 분할하거나 사전 대비를 지정하려면 **대비**를 클릭합니다.
- 사후분석 범위 검정과 대응별 다중 비교를 사용하여 차이가 나는 평균을 판별하려면 **사후분석**을 클릭합니다.
- 결측 데이터의 처리와 신뢰구간의 수준을 제어하려면 **옵션**을 클릭합니다.
- 평균, 중위수, 비율, 오즈비, 상관계수 또는 회귀계수와 같은 추정에 대한 표준오차 및 신뢰구간의 강력한 추정을 파생시키려면 **붓스트랩**을 클릭합니다.

일원 분산 분석 대비

그룹 간 제곱합을 추세 요소로 분할하거나 사전 대비를 지정할 수 있습니다.

다항

그룹 간 제곱합을 추세 요소로 나눕니다. 요인변수의 순서 수준을 통해 종속변수의 추세에 대해 검정할 수 있습니다. 예를 들어, 정렬된 최고차수 수입 수준에서 급여의 증가 및 감소의 선형 추세를 확인할 수 있습니다.

- **차수.** 1차, 2차, 3차, 4차, 5차 다항을 선택할 수 있습니다.

계수

t 통계로 검정할 사전 대비를 사용자가 지정합니다. 요인변수의 각 그룹(범주)에 대한 계수를 입력하고 하나 입력할 때마다 **추가**를 클릭합니다. 각각의 새로운 값은 계수 목록의 하단에 추가됩니다. 추가 대비 변수 세트를 지정하려면 **다음**을 클릭합니다. **다음**과 **이전**을 사용하여 대비 변수 세트끼리 이동할 수 있습니다.

대비에 대한 효과크기 추정

전체 검정에 대한 효과크기 계산을 제어합니다. 이 설정을 사용으로 설정한 경우 다음 중 적어도 하나의 옵션을 선택하여 효과크기를 계산해야 합니다. 이 설정은 대비가 한 개 이상 지정되고 출력결과의 ANOVA 효과크기 표에 표시된 경우에 사용으로 설정됩니다.

모든 집단의 통합 표준편차를 Standardizer로 사용

효과크기를 추정할 때 모든 집단의 통합 표준편차를 Standardizer로 사용합니다. 이 설정은 기본 설정으로, **대비에 대한 효과크기 추정**을 선택한 경우에 사용 가능합니다.

대비에 관련된 집단의 통합 표준편차를 Standardizer로 사용

대비에 관련된 집단의 통합 표준편차를 Standardizer로 사용합니다. 이 설정은 **대비에 대한 효과크기 추정**을 선택한 경우에 사용 가능합니다.

계수의 순서는 요인변수의 범주 값에 대한 오름차순에 해당하므로 중요합니다. 목록에서 첫 번째 계수는 요인변수의 가장 낮은 그룹 값에 해당하고 마지막 계수는 가장 높은 값에 해당합니다. 예를 들어, 요인변수가 범주 6에 있으면 계수 -1, 0, 0, 0, 0.5, 0.5는 다섯 번째와 여섯 번째 그룹을 첫 번째 그룹과 대비합니다. 대부분의 애플리케이션에서 계수는 합계가 0이어야 합니다. 합계가 0이 되지 않는 세트도 사용할 수 있지만 경고 메시지가 표시됩니다.

일원 분산 분석 사후 검정

평균들 간에 차이가 있다고 판단되면 사후분석 범위 검정과 대응별 다중 비교를 수행하여 차이가 나는 평균을 판별할 수 있습니다. 범위 검정은 차이가 없는 평균의 동일 서브세트를 식별하며, 대응별 다중 비교는 각 대응 평균 간 차이를 검정하고 유의수준 0.05에서 다른 그룹 평균이 유의하다는 것을 행렬에 별표로 나타냅니다.

등분산을 가정함

Tukey의 정직 유의차 검정, Hochberg의 GT2, Gabriel 및 Scheffé 검정 등은 다중 비교 검정과 범위 검정입니다. 또한 Tukey의 *b*, S-N-K(Student-Newman-Keuls), Duncan, R-E-G-W의 *F* (Ryan-Einot-Gabriel-Welsch *F* 검정), R-E-G-W의 *Q*(Ryan-Einot-Gabriel-Welsch 범위 검정), Waller-Duncan 등의 범위 검정을 사용할 수도 있습니다. 다중 비교 검정으로는 Bonferroni, Tukey의 정직 유의차 검정, Sidak, Gabriel, Hochberg, Dunnett, Scheffé LSD(최소 유의차) 등을 사용할 수 있습니다.

- *LSD*. T 검정을 사용하여 그룹 평균 간 모든 쌍대 비교를 수행합니다. 다중 비교에 대한 오차 비율을 수정하지 않습니다.
- *Bonferroni*. *t* 검정을 사용하여 그룹 평균 간 쌍대 비교를 수행하지만 각 검정에 대한 오차율을 전체 검정 수로 나눈 실험별 오차율로 설정하여 전체 오차율을 제어합니다. 따라서 관측 유의수준은 다중 비교 작업에 따라 조정됩니다.
- *Sidak*. t 통계를 기준으로 대응별 다중 비교 검정을 수행합니다. Sidak은 다중 비교에 대한 유의 수준을 조정하고 Bonferroni의 경우보다 더 엄격한 한계를 제공합니다.
- *Scheffe*. 가능한 모든 대응별 평균 조합에 대해 동시 결합 쌍대 비교를 수행합니다. F 표본 분포를 사용합니다. 단순한 쌍대 비교만이 아니라 그룹 평균의 가능한 모든 선형 조합을 검토하는 데 사용될 수 있습니다.
- *R-E-G-W*의 *F*. F 검정을 기준으로 하는 Ryan-Einot-Gabriel-Welsch 다중 단계감소 프로시저입니다.
- *R-E-G-W*의 *Q*. 스튜던트화 범위를 기준으로 하는 Ryan-Einot-Gabriel-Welsch 다중 단계감소 프로시저입니다.
- *S-N-K*. 스튜던트화 범위 분포를 사용하여 평균 간 모든 쌍대 비교를 수행합니다. 표본 결과가 같으면 단계별 프로시저를 사용하여 동일서브세트 내의 대응 평균을 비교할 수도 있습니다. 평균은 최고에서 최하까지 정렬되며 극단값차가 먼저 검정됩니다.

- *Tukey*. 스튜던트화 범위 통계를 사용하여 그룹 간 모든 쌍대 비교를 수행합니다. 실험별 오차 비율을 모든 쌍대 비교 콜렉션의 오차 비율로 설정합니다.
- *Tukey*의 *b*. 스튜던트화 범위 분포를 사용하여 그룹 간 쌍대 비교를 수행합니다. 임계값은 Tukey의 정직유의차 검정과 Student-Newman-Keuls에 대한 해당 값의 평균이 됩니다.
- *Duncan*. Student-Newman-Keuls 검정에서 사용한 순서와 동일한 단계별 비교 순서를 사용하여 쌍대 비교를 수행하지만 개별 검정의 오차율이 아닌 검정 콜렉션의 오차율에 대한 보호 수준을 설정합니다. 스튜던트화 범위 통계량을 사용합니다.
- *Hochberg's GT2*. 스튜던트화 최대 계수를 사용하는 다중 비교 및 범위 검정입니다. Tukey의 정직유의차 검정과 유사합니다.
- *Gabriel*. 스튜던트화 최대 계수를 사용했으며 셀 크기가 동일하지 않을 때 일반적으로 Hochberg의 GT2 보다 더 강력한 쌍대 비교 검정을 수행합니다. Gabriel의 검정은 셀 크기가 상당히 다양하게 분포할 경우 광범위하게 사용될 수 있습니다.
- *Waller-Duncan*. t 통계에 기반한 다중 비교 검정으로, Bayesian 접근법을 사용합니다.
- *Dunnett*. 단일 통제 평균에 대해 처리 세트를 비교하는 대응별 다중 비교 *t* 검정을 수행합니다. 마지막 범주가 기본 제어 범주가 됩니다. 처음 범주를 선택할 수도 있습니다. 요인 수준(통제 범주 제외)의 평균이 통제 범주의 평균과 같지 않음을 검정합니다. **<** 요인 **수준에서** 평균이 통제 범주보다 작으면 통제 검정, 요인의 모든 수준에서 평균이 통제 범주보다 크면 대조 검정. **>** 요인의 모든 수준에서 평균이 대조 범주의 평균보다 큰지 테스트합니다.

등분산을 가정하지 않음

분산이 동일하다고 가정하지 않는 다중 비교 테스트는 Tamhane의 T2, Dunnett의 T3, Games-Howell 및 Dunnett의 *C*입니다.

- *Tamhane's T2*. t 검정을 기반으로 하는 보수적 쌍대 비교 검정입니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.
- *Dunnett's T3*. 스튜던트화 최대 계수를 기준으로 한 쌍대 비교 검정을 수행합니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.
- *Games-Howell*. 때때로 자유로운 대응별 비교 검정입니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.
- *Dunnett's C*. 스튜던트화 범위를 기반으로 하는 대응별 비교 검정을 수행합니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.

참고: 활성화된 피벗표에서 형식 메뉴의 **표 특성**을 선택하여 나타난 표 특성 대화 상자에서 **비어있는 행과 열 숨김**을 선택 해제하면 사후검정 결과를 더 쉽게 해석할 수 있습니다.

영가설 검정

사후분석 검정에 대한 유의수준(알파)의 처리 방법을 지정합니다.

동일한 유의수준(알파)을 옵션 설정으로 사용
선택할 경우 옵션 대화 상자에 지정된 것과 동일한 설정이 사용됩니다.

사후분석 검정에 대한 유의수준(알파) 지정
선택할 경우 **수준** 필드에 유의수준(알파)을 지정할 수 있습니다.

일원 분산 분석에 대한 사후 검정 구하기

일원 분산 분석: 옵션

통계
다음 중 하나 이상을 선택합니다.

기술통계
케이스 수, 평균, 표준 편차, 평균의 표준 오차, 최소값, 최대값, 각 그룹의 종속변수에 대한 95% 신뢰구간 등을 계산합니다.

모수 및 변량효과
고정 효과 모형에 대한 표준 편차, 표준 오차 및 95% 신뢰구간과 랜덤 효과 모형에 대한 표준 오차, 95% 신뢰구간 및 그룹-간 성분분산 추정값을 표시합니다.

분산 동질성 검정
Levene 통계를 계산하여 그룹 분산의 등식을 검정합니다. 이러한 검정은 정규성 가정과 별개입니다.

Brown-Forsythe
Brown-Forsythe 통계량을 계산하여 집단 평균의 등식을 검정합니다. 등분산을 가정하지 않는 경우 *F* 통계보다 이 통계를 사용하는 것이 좋습니다.

Welch
Welch 통계량을 계산하여 집단 평균의 등식을 검정합니다. 등분산을 가정하지 않는 경우 *F* 통계보다 이 통계를 사용하는 것이 좋습니다.

결측값
결측값의 처리를 제어합니다.

분석별 결측값 제외
제공된 분석의 종속변수 또는 요인변수에 대해 결측값이 있는 케이스가 해당 분석에 사용되지 않습니다. 요인변수에 지정한 범위 밖의 케이스도 사용하지 않습니다.

목록별 결측값 제외
기본 대화 상자의 종속변수 목록에 포함된 종속변수 또는 요인변수에 대해 결측값이 있는 케이스가 모든 분석에서 제외됩니다. 다중 종속변수를 지정하지 않은 경우에는 아무런 영향이 없습니다.

신뢰구간
기본적으로 평균과 가정된 검정값과의 차이에 대해 95% 신뢰구간이 표시됩니다. 다른 신뢰수준을 원하면 1부터 99 사이의 값을 입력합니다.

평균 도표
부집단 평균을 도표화하는 도표를 표시합니다. 각 집단의 평균은 요인변수 값으로 정의됩니다.

일원 분산 분석 옵션 지정

ONEWAY 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 고정효과 및 랜덤 효과 통계를 구합니다. 표준 편차, 평균의 표준 오차 및 고정효과 모형의 95% 신뢰구간. 분량효과 모형의 표준 오차, 95% 신뢰구간 및 그룹-간 성분분산 추정값을 표시합니다(STATISTICS=EFFECTS 사용).
- 최소유의차, Bonferroni, Duncan 및 Scheffé 다중 비교 검정에 대한 알파 수준을 지정합니다(RANGES 하위 명령 사용).
- 평균 행렬, 표준 편차 및 빈도를 쓰거나 평균 행렬, 빈도, 풀링분산 및 풀링분산의 자유도를 읽습니다. 이러한 행렬은 일원 분산 분석을 구하기 위해 원시 데이터 대신 사용할 수 있습니다(MATRIX 하위 명령 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

GLM 일변량 분석

GLM 일변량 프로시저는 하나 이상의 요인 또는 변수별로 하나의 종속변수에 대한 회귀분석과 분산 분석을 제공합니다. 요인변수는 모집단을 여러 그룹으로 나눕니다. 이러한 일반선형모형 프로시저를 사용하여 다양한 그룹의 단일 종속변수의 평균에 미치는 기타 변수의 효과에 관한 귀무가설을 검정할 수 있습니다. 개별 요인들의 효과뿐 아니라 무작위로 선출된 요인간의 상호작용을 조사할 수 있습니다. 더불어 요인에 미치는 공변량의 효과와 공변량 상호작용을 함께 검토할 수 있습니다. 회귀분석의 경우 독립(예측자)변수가 공변량으로 지정됩니다.

균형 모형과 비균형 모형을 모두 검정할 수 있습니다. 모형의 각 셀이 같은 수의 케이스를 포함하는 경우에는 계획을 수행하는 것이 적절합니다. GLM 일변량은 가설 검정과 더불어 모수 추정값을 생성합니다.

일반적으로 사용되는 사전 대비는 가설 검정에 유용합니다. 또한 전체 *F* 검정에서 유의수준이 표시되면 사후 검정을 사용하여 특정 평균들 간의 차이를 평가할 수 있습니다. 주변 평균 추정을 통해 모형의 셀에 대한 예측평균

값의 추정값을 구할 수 있으며 이들 평균의 프로파일 도표(상호작용 도표)를 사용하여 일부 관계를 더 쉽게 시각화할 수 있습니다.

잔차, 예측값, Cook의 거리, 레버리지 값은 가정을 진단하기 위해 데이터 파일에 새로운 변수로 저장할 수 있습니다.

WLS 가중치를 사용하면 서로 다른 측정 정밀도를 보완하기 위해 여러 관측값에 가중 최소제곱(WLS) 분석에 대한 다른 가중치를 부여하는데 사용되는 변수를 지정할 수 있습니다.

예. 수년간 시카고 마라톤에 참가한 개별 주자에 관한 데이터를 수집하였습니다. 각 주자가 완주한 시간을 종속변수로 두고 날씨(추운, 상쾌한 또는 더운 날씨), 연습 기간, 이전의 마라톤 참가 횟수 및 성별을 기타 요인으로, 연령을 공변량으로 둡니다. 성별이 중요한 효과를 가지며 성별과 날씨의 상호작용이 유의하다는 것을 알 수 있습니다.

방법. 제 I 유형, 제 II 유형, 제 III 유형, 제 IV 유형 제곱합을 사용하여 여러 다른 가설을 평가할 수 있습니다. 제 III 유형이 기본값입니다.

통계. 사후 범위 검정 및 다중 비교에 대해 최소 유의차, Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch 다중 *F*, Ryan-Einot-Gabriel-Welsch 다중 범위, Student-Newman-Keuls, Tukey의 정직유의차, Tukey의 *b*, Duncan, Hochberg의 GT2, Gabriel, Waller-Duncan *t* 검정, Dunnett(단측 및 양측), Tamhane의 T2, Dunnett의 T3, Games-Howell, Dunnett의 *C*를 선택할 수 있습니다. 기술통계량에 대해 모든 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타낼 수 있습니다 분산 동질성에 대한 Levene 검정.

도표. 평균-산포 도표, 잔차 및 프로파일(상호작용)을 선택할 수 있습니다.

GLM 일변량 데이터 고려 사항

데이터. 종속변수는 양적변수입니다. 요인은 범주형입니다. 최대 8자까지 숫자나 문자열 값을 가질 수 있습니다. 공변량은 종속변수와 관련된 양적변수입니다.

가정. 데이터는 정규 모집단에서 추출된 무작위 표본입니다. 즉, 모집단에서는 모든 셀에 대한 분산이 동일합니다. 분산 분석은 데이터가 대칭적인 경우에도 정규성에서 이탈해도 영향을 받지 않습니다. 가정을 확인하려면 분산의 동질성 검정과 평균-산포 도표를 사용할 수 있습니다. 또한 잔차와 잔차 도표를 검토할 수 있습니다.

GLM 일변량 테이블 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 일반 선형 모형 > 일변량...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 종속변수를 한 개 선택합니다.
3. 현재 데이터에 맞게 모수요인, 변량요인, 공변량에 대한 변수를 선택합니다.
4. 선택적으로 WLS 가중치를 사용하여 가중 최소제곱 분석에 대한 가중 변수를 지정할 수 있습니다. 가중 변수 값이 0, 음수 또는 결측이면 해당 케이스는 분석에서 제외됩니다. 모형에 이미 사용된 변수는 가중 변수로 사용할 수 없습니다.

GLM 모형

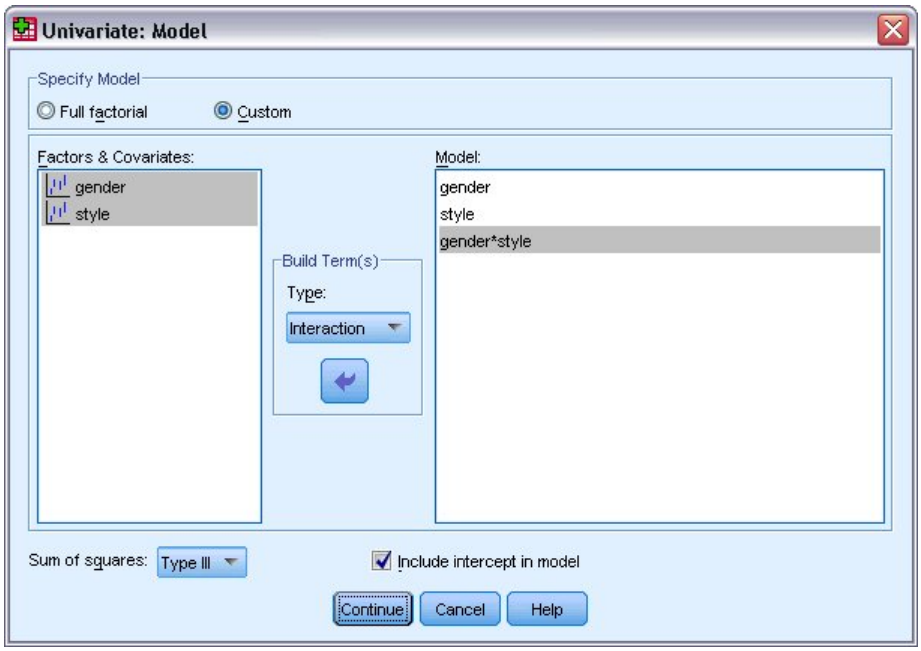


그림 *1.* 일변량 모형 대화 상자

모형 지정. 완전 요인 모형에는 모든 요인 주효과, 모든 공분산 주효과, 모든 요인-대-요인 상호작용이 포함됩니다. 공분산 상호작용은 포함되지 않습니다. 상호작용의 부분집합만 지정하거나 요인 대 공변량 상호작용을 지정하려면 **사용자 정의**를 선택합니다. 모형에 포함할 모든 항을 지정해야 합니다.

요인 및 공변량. 요인 및 공변량이 나열됩니다.

모형. 모형은 데이터의 특성에 따라 다릅니다. **사용자 정의**를 선택한 다음 분석시 중요한 주효과와 상호작용을 선택할 수 있습니다.

제곱합. 제곱합을 계산하는 방법입니다. 결측 셀이 없는 균형 모형이나 비균형 모형의 경우 제 III 유형 제곱합 방법이 가장 일반적으로 사용됩니다.

모형에 절편 포함. 이 모형에는 대개 절편이 포함됩니다. 데이터가 원점을 전달한다고 가정할 경우에는 절편을 제외시킬 수 있습니다.

항 설정 및 사용자 정의 항

항 설정
선택된 요인과 공변량 세트의 모든 조합에 대해 특정 유형(예: 주효과)의 비중첩 항을 포함하려면 이 선택사항을 사용합니다.

사용자 정의 항 설정
중첩 항을 포함하고자 하거나 변수별로 항 변수를 명시적으로 설정하고자 할 때 이 선택사항을 사용합니다. 중첩 항 설정에는 다음 단계가 포함됩니다.

제곱합

모형의 경우 한 가지 유형의 제곱합을 선택할 수 있습니다. 제 III 유형이 가장 일반적으로 사용되며 기본값입니다.

제 I 유형. 이 방법은 또한 제곱합의 계층적 분해 방법으로 알려져 있습니다. 각 항은 모형에서 해당 항의 앞에 오는 항에 따라서만 수정됩니다. 제 I 유형 제곱합은 일반적으로 다음 경우에 사용됩니다.

- 임의의 1차 상호작용 효과 이전에 주효과가 지정되고 2차 상호작용 효과 이전에 1차 상호작용 효과가 지정되는 등의 균형 분산 분석 모형
- 저차항이 고차항 이전에 지정되는 다항 회귀 모형

- 1차 지정 효과가 2차 지정 효과 내에 중첩되고 2차 지정 효과가 3차 지정 효과 내에 중첩되는 등의 순수 지분 모형 (이 중첩 형식은 구문을 통해서만 지정될 수 있습니다.)

제 II 유형. 이 방법은 다른 모든 적합한 효과에 맞게 수정된 모형내의 효과의 제곱합을 계산합니다. 적합한 효과는 검토 중인 효과를 포함하지 않는 모든 효과입니다. 제 II 유형 제곱합 방법은 일반적으로 다음 경우에 사용됩니다.

- 균형 분산 분석 모형
- 주요인 효과만 있는 모형
- 모든 회귀 모형
- 순수 중첩 계획. (이 중첩 형식은 구문을 통해 지정될 수 있습니다.)

제 III 유형. 기본값입니다. 이 방법은 계획에 있는 한 효과의 제곱합을 계산할 때 이 효과를 포함하지 않는 다른 효과에 맞게 수정되며 이 효과를 포함하는 효과(있는 경우)에 직교하는 제곱합 방법을 사용합니다. 제 III 유형 제곱합은 일반 추정 형식이 일정하게 유지되는 한 셀 빈도에 관해 변하지 않는 이점이 있습니다. 따라서 이 제곱합 유형은 보통 결측 셀이 없는 비균형 모형에 유용한 것으로 간주됩니다. 이 방법은 결측 셀이 없는 요인 계획에서 Yates의 가중 평균 제곱 기법과 동등합니다. 제 III 유형 제곱합 방법은 일반적으로 다음 경우에 사용됩니다.

- 제 I 유형과 제 II 유형으로 나열된 모든 모형
- 빈 셀이 없는 모든 균형 모형이나 비균형 모형

제 IV 유형. 이 방법은 결측 셀이 있는 상황에 대해 계획됩니다. 계획내의 효과 *F*의 경우 *F*가 다른 효과에 포함되지 않는다면 제 IV 유형 = 제 III 유형 = 제 II 유형이 됩니다. *F*가 다른 효과에 포함되는 경우 제 IV 유형은 *F*의 모수 사이에서 수행되는 대비를 모든 최고 수준 효과에 동등하게 분포시킵니다. 제 IV 유형 제곱합 방법은 일반적으로 다음 경우에 사용됩니다.

- 제 I 유형과 제 II 유형으로 나열된 모든 모형
- 빈 셀이 있는 모든 균형 모형이나 비균형 모형

GLM 대비

대비는 요인의 수준 차이를 검정하는 데 사용됩니다. 각 개체 간 요인에 대한 반복 측도 모형에서 모형의 각 요인에 대해 대비를 지정할 수 있습니다. 대비는 모수의 선형 조합을 나타냅니다.

GLM 일변량. 가설 검정은 귀무가설 **LB**=0을 기준으로 하며 여기에서 **L**은 대비계수(L) 행렬이고 **B**는 모수 벡터입니다. 대비가 지정되면 **L** 행렬이 작성됩니다. 요인에 해당하는 **L** 행렬의 열이 대비와 일치합니다. 나머지 열은 **L** 행렬 추정이 가능하도록 수정됩니다.

결과에는 각 대비군에 대한 *F* 통계가 포함되어 있습니다. 또한 대비 차이에 대한 Student의 *t* 분포에 따른 Bonferroni 유형 동시 신뢰구간이 표시됩니다.

사용 가능한 대비

사용할 수 있는 대비는 편차, 단순, 차이, Helmert, 반복 및 다항입니다. 편차 대비와 단순 대비의 경우 참조 범주가 마지막 범주인지 또는 처음 범주인지 여부를 선택할 수 있습니다.

대비 유형

편차. 각 수준의 평균(참조 범주 제외)을 모든 수준의 평균(총 평균)과 비교합니다. 요인의 수준들은 원하는 순서대로 지정할 수 있습니다.

단순. 각 수준의 평균을 지정한 수준의 평균과 비교합니다. 이 대비 유형은 대조군이 있는 경우 유용합니다. 처음이나 마지막 범주를 참조로 선택할 수 있습니다.

차이. 각 수준의 평균(처음 제외)을 이전 수준의 평균과 비교합니다. 역 Helmert 대비라고도 합니다.

Helmert. 요인의 각 수준 평균(마지막 제외)을 후속 수준의 평균과 비교합니다.

반복. 각 수준의 평균(마지막 제외)을 후속 수준의 평균과 비교합니다.

다항. 선형 효과, 2차 효과, 3차 효과 등을 비교합니다. 처음 자유도에는 모든 범주에서의 선형 효과가 포함됩니다. 그리고 두 번째 자유도는 2차 효과, 세 번째는 3차 효과의 순으로 계속 적용됩니다. 이들 대비는 다항 추세를 추정하는 데 사용됩니다.

GLM 프로파일 도표

프로파일 도표(상호작용 도표)는 모형의 주변 평균을 비교하는 데 유용합니다. 프로파일 도표는 선도표이며 도표 내의 각 점은 한 요인 수준에서 측정한 종속변수의 추정된 주변 평균(공변량에 따라 수정됨)을 나타냅니다. 개별 선을 만드는데 또 다른 요인의 수준이 사용될 수 있습니다. 또한 또 다른 세 번째 요인의 각 수준이 개별 도표 작성에 사용될 수 있습니다. 모든 모수요인과 변량요인은 어떤 것이라도 도표에 사용될 수 있습니다. 다변량 분석에서는 각 종속변수에 대해 프로파일 도표를 만듭니다. 반복 측도 분석에서 개체 간 요인과 개체-내 요인은 모두 프로파일 도표에 사용될 수 있습니다. GLM 다변량 및 GLM 반복측도는 Advanced Statistics 옵션이 설치된 경우에만 사용할 수 있습니다.

한 요인에 대한 프로파일 도표는 추정된 주변 평균이 여러 수준에서 증가하는지 또는 감소하는지 여부를 보여 줍니다. 둘 이상의 요인의 경우 평행선은 두 요인 간 상호작용이 없음을 나타내는데 이것은 한 요인의 수준만 조사할 수 있음을 의미합니다. 평행하지 않는 선들은 상호작용을 나타냅니다.

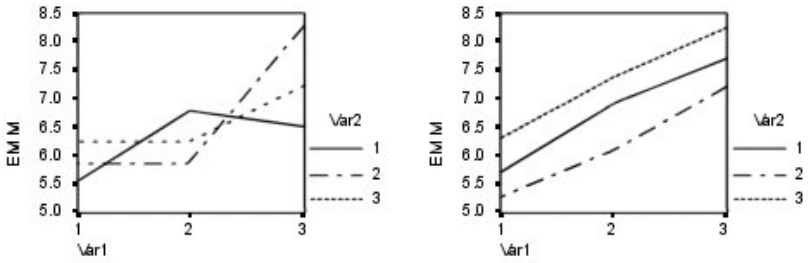


그림 *2.* 평행하지 않은 도표*(*왼쪽*)* 및 평행한 도표*(*오른쪽*)*

수평축에 대한 요인을 선택하고 필요에 따라 개별 선과 개별 도표에 대한 요인을 선택하여 도표를 지정하면 도표가 도표 목록에 추가되어야 합니다.

차트 유형
도표는 선도표 또는 막대도표일 수 있습니다.

오차 막대
신뢰구간 또는 다수의 표준 오차를 표시하는 오차 막대를 포함할 수 있습니다. 신뢰구간은 옵션 대화 상자에서 지정된 유의 수준을 기반으로 합니다.

총평균의 참조선 포함(G)
전체 총 평균을 표시하는 참조선을 포함합니다.

Y축은 0에서 시작됨
모든 양수 또는 모든 음수 값이 있는 선형 차트의 경우 Y축이 0으로 시작하도록 강제 적용합니다. 막대형 차트는 항상 0으로 시작(또는 포함)합니다.

GLM 옵션

이 대화 상자에서는 선택적 통계를 사용할 수 있습니다. 통계는 모수효과 모형을 사용하여 계산됩니다.

표시. 전체 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타내려면 **기술통계**를 선택합니다. **효과크기 추정값**은 각 효과와 각 모수 추정값에 대한 부분 에타제곱을 제공합니다. 에타제곱 통계는 요인에서 기인하는 전체 변동의 비율에 대해 설명합니다. 대립가설을 관측값 기준으로 설정한 경우에 검정력을 구하려면 **관측 검정력**을 선택합니다. **모수 추정값**을 선택하여 각 검정에 대한 모수 추정값, 표준 오차, *t* 검정, 신뢰구간, 관측 검정력을 생성합니다. **대비계수(L) 행렬**을 선택하여 **L** 행렬을 구합니다.

동질성 검정은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서 각 종속변수마다 분산의 동질성에 대한 Levene의 검정을 생성합니다. 평균-산포 도표와 잔차 도표 옵션은 데이터에 대한 가정을 진단하는 데 유용합니다. 이 항목은 요인이 없을 경우에는 사용할 수 없습니다. **잔차 도표**를 선택하여 각 종속변수에 대한 관측-예측-표준화 잔차 도표를 작성합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. **적합결여 검정**을 선택하여 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인합니다. **일반 추정가능 함수**를 사용하여 일반 추정가능 함수를 기준으로 사용자 정의 가설검정을 작성할 수 있습니다. 대비 계수 행렬의 행은 일반 추정가능 함수의 선형결합입니다.

이분산성 검정은 오차의 분산(각 종속 변수에 대한)이 독립 변수의 값에 의존하는지 여부를 검정하는 데 사용될 수 있습니다. **Breusch-Pagan 검정**, **수정된 Breusch-Pagan 검정** 및 **F 검정**의 경우에는 검정의 기반이 되는 모

형을 지정할 수 있습니다. 기본적으로, 모형은 상수 항, 예측값에서 선형인 항, 예측값에서 2차인 항, 그리고 오류 항으로 구성되어 있습니다.

강력한 표준 오차의 모수 추정값은 강력한 또는 이분산성 일관적(HC) 표준 오차, 그리고 *t* 통계, 유의수준 및 강력한 표준 오차를 사용하는 신뢰구간과 함께 모수 추정값의 표를 표시합니다. 다섯 가지의 서로 다른 방법이 강력한 공분산 행렬 추정에 사용 가능합니다.

HC0
모수 추정값의 공분산 행렬에 대해 원래의 점근적 또는 대형 샘플 로버스트, 경험적 또는 "샌드위치" 추정량을 기반으로 합니다. 샌드위치의 중간 부분에는 제곱한 OLS(Ordinary Least Squares) 또는 제곱한 가중치 WLS(Weighted Least Squares) 잔차가 포함됩니다.

HC1
HC0의 수정의 유한 샘플을 N/(N-p)로 곱합니다. 여기서 N은 샘플 크기이며, p는 모형에서 중복되지 않는 모수의 수입니다.

HC2
제곱한 잔차를 1-h로 나눈 값을 포함하는 HC0의 수정입니다. 여기서 h는 케이스의 레버리지입니다.

HC3
잭나이프 추정량의 근사치를 계산하는 HC0의 수정입니다. 제곱한 잔차를 1-h의 제곱으로 나눕니다.

HC4
제곱한 잔차를 1-h의 제곱으로 나누는 HC0의 수정입니다. 이는 h, N 및 p에 따라 변하며 상단 한계는 4입니다.

유의 수준. 사후 검정에 사용되는 유의 수준과 신뢰구간 형성에 사용되는 신뢰수준을 조정할 수 있습니다. 지정된 값은 해당 검정에 대한 관측 검정력을 계산하는 데에도 사용할 수 있습니다. 유의수준을 지정할 때 신뢰구간의 관련된 수준이 대화 상자에 출력됩니다.

UNIANOVA 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- DESIGN 하위 명령을 사용하여 계획에 중첩 효과를 지정합니다.
- TEST 하위 명령을 사용하여 효과에 대한 효과-대 선형 조합에 대한 검정이나 값을 지정합니다.
- CONTRAST 하위 명령을 사용하여 다중 대비를 지정합니다.
- MISSING 하위 명령을 사용하여 사용자 결측값을 포함시킵니다.
- CRITERIA 하위 명령을 사용하여 EPS 기준을 지정합니다.
- LMATRIX, MMATRIX 및 KMATRIX 하위 명령을 사용하여 사용자 정의 **L** 행렬, **M** 행렬 또는 **K** 행렬을 구성합니다.
- CONTRAST 하위 명령을 사용하여 편차나 단순 대비의 경우 중간 참조 범주를 지정합니다.
- CONTRAST 하위 명령을 사용하여 다항 대비에 대한 메트릭 방법을 지정합니다.
- POSTHOC 하위 명령을 사용하여 사후분석 비교에 대한 오차항을 지정합니다.
- EMMEANS 하위 명령을 사용하여 요인 목록에서 요인에 대해 추정된 주변 평균이나 여러 요인 간 요인 상호작용을 계산합니다.
- SAVE 하위 명령을 사용하여 임시 변수 이름을 지정합니다.
- OUTFILE 하위 명령을 사용하여 상관행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 개체 간 분산 분석 테이블로부터 통계를 포함하는 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 계획 행렬을 새 데이터 파일에 저장합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

GLM 사후분석 비교

사후분석 다중 비교 검정. 평균들 간에 차이가 있다고 판단되면 사후분석 범위 검정과 대응별 다중 비교를 수행하여 차이가 나는 평균을 판별할 수 있습니다. 비교는 수정되지 않은 값에 대해 수행됩니다. 이 검정은 개체 간 모수 요인에 대해서만 사용됩니다. GLM 반복 측도에서 개체 간 요인이 없으면 이 검정을 사용할 수 없으며 사후분석 다중 비교 검정은 개체-내 요인 수준 간의 평균에 대해 수행됩니다. GLM 다변량에서 사후 검정은 각 종속변수에 대해 개별적으로 수행됩니다. GLM 다변량 및 GLM 반복측도는 Advanced Statistics 옵션이 설치된 경우에만 사용할 수 있습니다.

Bonferroni와 Tukey의 정직유의차 검정은 보통 다중 비교 검정에 사용됩니다. 스튜던트의 *t* 통계를 기반으로 하는 **Bonferroni 검정**은 다중 비교가 수행되는 사실에 대해 관찰된 유의 수준을 조정합니다. **Sidak의 T 검정**을 사용하는 경우에도 유의 수준이 수정되며 Bonferroni 검정보다 더 강력한 한계를 제공합니다. **Tukey의 정직유의차 검정**은 스튜던트화 범위 통계를 사용하여 모든 그룹 간 쌍대 비교를 수행하며 실험별 오차 비율을 모든 쌍대 비교 콜렉션의 오차 비율로 설정합니다. 다수의 대응평균을 검정할 경우 Tukey의 정직유의차 검정이 Bonferroni 검정보다 효과적인 결과를 나타냅니다. 소수의 대응평균에는 Bonferroni가 보다 효과적입니다.

Hochberg의 GT2는 Tukey의 정직유의차 검정과 유사하나 스튜던트화 최대 계수가 사용됩니다. 대개 Tukey의 검정이 더 효과적입니다. **Gabriel의 쌍대 비교 검정** 또한 스튜던트화 최대 계수를 사용하며 셀 크기가 동일하지 않을 경우 일반적으로 Hochberg의 GT2가 더 효과적인 방법입니다. Gabriel의 검정은 셀 크기가 상당히 다양하게 분포할 경우 광범위하게 사용될 수 있습니다.

Dunnett의 대응별 다중 비교 T 검정은 단일 통제 평균에 대해 하나의 처리 세트를 비교합니다. 마지막 범주가 기본 제어 범주가 됩니다. 처음 범주를 선택할 수도 있습니다. 또한 양쪽검정이나 한쪽검정을 선택할 수 있습니다. 한 요인 수준(제어 범주 제외)에서의 평균이 제어 범주에서의 평균과 동일하지 않은지 검정하려면 양쪽검정을 사용합니다. 임의의 요인 수준의 평균이 제어 범주의 평균보다 작은지 여부를 테스트하려면 **< 제어**를 선택하십시오. 마찬가지로, 임의의 요인 레벨에서의 평균이 제어 범주의 평균보다 큰지 여부를 테스트하려면 **> 제어**를 선택하십시오.

Ryan, Einot, Gabriel 및 Welsch(R-E-G-W)는 두 가지 다중 하향 범위 검정을 개발하였습니다. 다중 하향 프로시저는 우선 모든 평균이 동일한지 검정합니다. 모든 평균이 동일하지 않으면 평균 서브세트에 대한 등식을 검정합니다. **R-E-G-W F**는 *F* 검정을 기준으로 하며 **R-E-G-W Q**는 스튜던트화 범위를 기준으로 합니다. 이들 검정은 Duncan의 다중 범위 검정과 Student-Newman-Keuls(역시 다중 하향 프로시저임)보다 더 효과적인 방법이지만 셀 크기가 다른 경우에는 적합하지 않습니다.

분산이 동일하지 않을 경우 **Tamhane의 T2**(*t* 검정을 기준으로 하는 보존성 쌍대 비교 검정), **Dunnett의 T3**(스튜던트화 최대 계수를 기준으로 하는 쌍대 비교 검정), **Games-Howell 쌍대 비교 검정**(경우에 따라 광범위하게 사용됨) 또는 **Dunnett의 C**(스튜던트화 범위를 기준으로 하는 쌍대 비교 검정)를 사용합니다. 이러한 검정은 유효하지 않으며 모형에 여러 요인이 있는 경우 생성되지 않습니다.

Duncan의 다중 범위 검정, Student-Newman-Keuls(**S-N-K**), **Tukey의 B**는 그룹 평균의 순위를 정하고 범위 값을 계산하는 범위 검정입니다. 이들 검정은 앞서 언급한 다른 검정만큼 자주 사용되지는 않습니다.

Waller-Duncan T 검정은 Bayesian 접근법을 사용합니다. 이 범위 검정은 표본의 크기가 동일하지 않을 경우 표본 결과의 조화 평균을 사용합니다.

Scheffé 검정의 유의 수준은 이 기능에 사용할 수 있는 쌍대 비교 뿐 아니라 검정할 그룹 평균들의 가능한 모든 선형 조합을 허용합니다. 그 결과 Scheffé 검정은 다른 검정 방법보다 좀더 보수적 특성을 갖는데, 이는 유의수준을 위해 평균 차가 더 커야 한다는 것을 의미합니다.

최소유의차(**LSD**) 대응별 다중 비교 검정은 모든 대응 그룹 간 다중 개별 *T* 검정과 동일합니다. 이 검정 방법은 다중 비교에 대한 관측유의 수준을 수정할 수 없다는 단점이 있습니다.

출력된 검정. 쌍대 비교는 LSD, Sidak, Bonferroni, Games-Howell, Tamhane의 T2 및 T3, Dunnett의 *C*, Dunnett의 T3에 제공됩니다. 범위 검정을 위한 동일 서브세트는 S-N-K, Tukey의 *B*, Duncan, R-E-G-W *F*, R-E-G-W *Q* 및 Waller에 제공됩니다. Tukey의 정직유의차 검정, Hochberg의 GT2, Gabriel의 검정, Scheffé의 검정 등은 다중 비교 검정과 범위 검정입니다.

GLM 옵션

이 대화 상자에서는 선택적 통계를 사용할 수 있습니다. 통계는 모수효과 모형을 사용하여 계산됩니다.

표시. 전체 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타내려면 **기술통계**를 선택합니다. **효과크기 추정값**은 각 효과와 각 모수 추정값에 대한 부분 에타제곱을 제공합니다. 에타제곱 통계는 요인에서 기인하는

전체 변동의 비율에 대해 설명합니다. 대립가설을 관측값 기준으로 설정한 경우에 검정력을 구하려면 **관측 검정력**을 선택합니다. **모수 추정값**을 선택하여 각 검정에 대한 모수 추정값, 표준 오차, *t* 검정, 신뢰구간, 관측 검정력을 생성합니다. **대비계수(L) 행렬**을 선택하여 **L** 행렬을 구합니다.

동질성 검정은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서 각 종속변수마다 분산의 동질성에 대한 Levene의 검정을 생성합니다. 평균-산포 도표와 잔차 도표 옵션은 데이터에 대한 가정을 진단하는 데 유용합니다. 이 항목은 요인이 없을 경우에는 사용할 수 없습니다. **잔차 도표**를 선택하여 각 종속변수에 대한 관측-예측-표준화 잔차 도표를 작성합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. **적합결여 검정**을 선택하여 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인합니다. **일반 추정가능 함수**를 사용하여 일반 추정가능 함수를 기준으로 사용자 정의 가설검정을 작성할 수 있습니다. 대비 계수 행렬의 행은 일반 추정가능 함수의 선형결합입니다.

이분산성 검정은 오차의 분산(각 종속 변수에 대한)이 독립 변수의 값에 의존하는지 여부를 검정하는 데 사용될 수 있습니다. **Breusch-Pagan 검정**, **수정된 Breusch-Pagan 검정** 및 **F 검정**의 경우에는 검정의 기반이 되는 모형을 지정할 수 있습니다. 기본적으로, 모형은 상수 항, 예측값에서 선형인 항, 예측값에서 2차인 항, 그리고 오류 항으로 구성되어 있습니다.

강력한 표준 오차의 모수 추정값은 강력한 또는 이분산성 일관적(HC) 표준 오차, 그리고 *t* 통계, 유의수준 및 강력한 표준 오차를 사용하는 신뢰구간과 함께 모수 추정값의 표를 표시합니다. 다섯 가지의 서로 다른 방법이 강력한 공분산 행렬 추정에 사용 가능합니다.

HC0
: 모수 추정값의 공분산 행렬에 대해 원래의 점근적 또는 대형 샘플 로버스트, 경험적 또는 "샌드위치" 추정량을 기반으로 합니다. 샌드위치의 중간 부분에는 제곱한 OLS(Ordinary Least Squares) 또는 제곱한 가중치 WLS(Weighted Least Squares) 잔차가 포함됩니다.

HC1
: HC0의 수정의 유한 샘플을 N/(N-p)로 곱합니다. 여기서 N은 샘플 크기이며, p는 모형에서 중복되지 않는 모수의 수입니다.

HC2
: 제곱한 잔차를 1-h로 나눈 값을 포함하는 HC0의 수정입니다. 여기서 h는 케이스의 레버리지입니다.

HC3
: 잭나이프 추정량의 근사치를 계산하는 HC0의 수정입니다. 제곱한 잔차를 1-h의 제곱으로 나눕니다.

HC4
: 제곱한 잔차를 1-h의 제곱으로 나누는 HC0의 수정입니다. 이는 h, N 및 p에 따라 변하며 상단 한계는 4입니다.

유의 수준. 사후 검정에 사용되는 유의 수준과 신뢰구간 형성에 사용되는 신뢰수준을 조정할 수 있습니다. 지정된 값은 해당 검정에 대한 관측 검정력을 계산하는 데에도 사용할 수 있습니다. 유의수준을 지정할 때 신뢰구간의 관련된 수준이 대화 상자에 출력됩니다.

UNIANOVA 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- DESIGN 하위 명령을 사용하여 계획에 중첩 효과를 지정합니다.
- TEST 하위 명령을 사용하여 효과에 대한 효과-대 선형 조합에 대한 검정이나 값을 지정합니다.
- CONTRAST 하위 명령을 사용하여 다중 대비를 지정합니다.
- MISSING 하위 명령을 사용하여 사용자 결측값을 포함시킵니다.
- CRITERIA 하위 명령을 사용하여 EPS 기준을 지정합니다.
- LMATRIX, MMATRIX 및 KMATRIX 하위 명령을 사용하여 사용자 정의 **L** 행렬, **M** 행렬 또는 **K** 행렬을 구성합니다.
- CONTRAST 하위 명령을 사용하여 편차나 단순 대비의 경우 중간 참조 범주를 지정합니다.
- CONTRAST 하위 명령을 사용하여 다항 대비에 대한 메트릭 방법을 지정합니다.
- POSTHOC 하위 명령을 사용하여 사후분석 비교에 대한 오차항을 지정합니다.
- EMMEANS 하위 명령을 사용하여 요인 목록에서 요인에 대해 추정된 주변 평균이나 여러 요인 간 요인 상호작용을 계산합니다.

- SAVE 하위 명령을 사용하여 임시 변수 이름을 지정합니다.
- OUTFILE 하위 명령을 사용하여 상관행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 개체 간 분산 분석 테이블로부터 통계를 포함하는 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 계획 행렬을 새 데이터 파일에 저장합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

GLM 저장

모형, 잔차, 관련 측도에 의해 예측된 값을 데이터 편집기에 새로운 변수로 저장할 수 있습니다. 이들 변수 중 다수는 데이터에 관한 가정을 검토하는 데 사용할 수 있습니다. 다른 IBM SPSS 통계 세션에서 사용할 수 있도록 값을 저장하려면 현재 데이터 파일을 저장해야 합니다.

예측값. 각 케이스에 대한 모형 예측값입니다.

- 표준화되지 않음. 모형이 예측하는 종속변수의 값입니다.
- 가중됨. 표준화되지 않은 가중 예측값입니다. WLS 변수가 먼저 선택되어야만 이 항목을 사용할 수 있습니다.
- 표준 오차. 독립변수와 같은 값을 갖는 케이스에 대한 종속변수 평균값의 표준 편차 추정값입니다.

진단. 독립변수 값의 특수 조합을 포함하는 케이스와 모형에 큰 영향력을 미칠 수 있는 케이스를 구분하기 위한 측도입니다.

- *Cook*의 거리. 특정 케이스를 회귀계수 계산에서 제외할 때 모든 케이스의 잔차가 얼마나 변경될 수 있는지에 대한 측정값입니다. Cook의 거리가 크면 회귀분석 통계 계산에서 케이스를 제외할 때 계수가 크게 변경됩니다.
- 값 활용. 중심화되지 않은 레버리지 값. 각 관측값이 모형의 적합도에 미치는 상대적인 영향을 나타냅니다.

잔차. 비표준화 잔차는 종속변수의 실제 값에서 모형에 의해 예측된 값을 뺀 것입니다. 표준화, 스튜던트화, 삭제된 잔차도 사용할 수 있습니다. WLS 변수가 선택된 경우 표준화하지 않은 가중치 잔차를 사용할 수 있습니다.

- 표준화되지 않음. 관측값과 모형에 의해 예측된 값 사이 차이입니다.
- 가중됨. 표준화되지 않은 가중치 잔차. WLS 변수가 먼저 선택되어야만 이 항목을 사용할 수 있습니다.
- 표준화됨. 표준 편차의 추정값으로 나눈 잔차입니다. Pearson 잔차라고도 하는 표준화 잔차는 평균이 0이고 표준 편차가 1이 됩니다.
- 스튜던트화됨. 독립변수의 평균으로부터 독립변수에 대한 각 케이스 값의 거리에 따라 케이스마다 다른 표준 편차의 추정값으로 나눈 잔차입니다. 내부적으로 스튜던트화 잔차라고도 합니다.
- 삭제됨. 해당 케이스가 회귀계수 계산에서 제외될 때 케이스에 대한 잔차입니다. 종속변수 값과 조정된 예측값 간 차이입니다.

계수 통계. 모형에서 모수 추정값의 분산-공분산행렬을 현재 세션의 새 데이터 세트나 외부 IBM SPSS 통계 데이터 파일에 씁니다. 또한 각 종속변수에 대해 모수 추정값의 행, 모수 추정값의 표준오차의 행, 모수 추정값에 해당하는 *t* 통계에 대한 유의수준의 행, 그리고 잔차 자유도의 행이 있습니다. 다변량 모형의 경우 각 종속변수에 대해 이와 유사한 행들이 존재합니다. 이분산성 일관적 통계가 선택된 경우(일변량 모형에만 사용 가능함), 분산-공분산행렬이 강력한 추정량을 사용하여 계산되고 표준 오차의 행은 강력한 표준 오차를 표시하며 유의수준은 강력한 오차를 반영합니다. 이 행렬 파일은 행렬 파일을 읽는 다른 프로시저에도 사용할 수 있습니다.

GLM 추정 주변 평균

셀 내의 모집단 주변 평균들의 추정값이 필요한 경우 해당 요인과 상호작용을 선택합니다. 이들 평균은 공변량에 따라 수정됩니다.

주효과 비교
개체 간 요인과 개체-내 요인 모두에 모형 내 주효과의 추정된 주변 평균 간 수정되지 않은 쌍대 비교를 제공합니다. 이 항목은 평균 출력 기준 목록에서 주효과를 선택한 경우에만 사용할 수 있습니다.

단순 주효과 비교
이 설정은 대상 목록에 하나 이상의 제품 또는 상호 작용 효과가 포함될 때마다 활성화됩니다(예: A*B, A*B*C). 이 설정은 다른 요소의 수준 내에 중첩된 주 효과인 단순 주 효과 간의 비교를 지정할 수 있도록 지원합니다.

신뢰구간 수정
신뢰구간과 유의수준에 최소유의차(LSD), Bonferroni, Sidak 조정을 선택합니다. 이 항목은 **주효과 비교** 및/또는 **단순 주효과 비교**가 선택된 경우에만 사용 가능합니다.

추정 주변 평균 지정

1. 메뉴의 > **분석** > **일반 선형 모형**에서 사용 가능한 프로시저 중 하나를 선택합니다.
2. 기본 대화 상자에서 **EM 평균**을 클릭합니다.

GLM 옵션

이 대화 상자에서는 선택적 통계를 사용할 수 있습니다. 통계는 모수효과 모형을 사용하여 계산됩니다.

표시. 전체 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타내려면 **기술통계**를 선택합니다. **효과크기 추정값**은 각 효과와 각 모수 추정값에 대한 부분 에타제곱을 제공합니다. 에타제곱 통계는 요인에서 기인하는 전체 변동의 비율에 대해 설명합니다. 대립가설을 관측값 기준으로 설정한 경우에 검정력을 구하려면 **관측 검정력**을 선택합니다. **모수 추정값**을 선택하여 각 검정에 대한 모수 추정값, 표준 오차, *t* 검정, 신뢰구간, 관측 검정력을 생성합니다. **대비계수(L) 행렬**을 선택하여 **L** 행렬을 구합니다.

동질성 검정은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서 각 종속변수마다 분산의 동질성에 대한 Levene의 검정을 생성합니다. 평균-산포 도표와 잔차 도표 옵션은 데이터에 대한 가정을 진단하는 데 유용합니다. 이 항목은 요인이 없을 경우에는 사용할 수 없습니다. **잔차 도표**를 선택하여 각 종속변수에 대한 관측-예측-표준화 잔차 도표를 작성합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. **적합결여 검정**을 선택하여 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인합니다. **일반 추정가능 함수**를 사용하여 일반 추정가능 함수를 기준으로 사용자 정의 가설검정을 작성할 수 있습니다. 대비 계수 행렬의 행은 일반 추정가능 함수의 선형결합입니다.

이분산성 검정은 오차의 분산(각 종속 변수에 대한)이 독립 변수의 값에 의존하는지 여부를 검정하는 데 사용될 수 있습니다. **Breusch-Pagan 검정**, **수정된 Breusch-Pagan 검정** 및 **F 검정**의 경우에는 검정의 기반이 되는 모형을 지정할 수 있습니다. 기본적으로, 모형은 상수 항, 예측값에서 선형인 항, 예측값에서 2차인 항, 그리고 오류 항으로 구성되어 있습니다.

강력한 표준 오차의 모수 추정값은 강력한 또는 이분산성 일관적(HC) 표준 오차, 그리고 *t* 통계, 유의수준 및 강력한 표준 오차를 사용하는 신뢰구간과 함께 모수 추정값의 표를 표시합니다. 다섯 가지의 서로 다른 방법이 강력한 공분산 행렬 추정에 사용 가능합니다.

HC0
모수 추정값의 공분산 행렬에 대해 원래의 점근적 또는 대형 샘플 로버스트, 경험적 또는 "샌드위치" 추정량을 기반으로 합니다. 샌드위치의 중간 부분에는 제곱한 OLS(Ordinary Least Squares) 또는 제곱한 가중치 WLS(Weighted Least Squares) 잔차가 포함됩니다.

HC1
HC0의 수정의 유한 샘플을 N/(N-p)로 곱합니다. 여기서 N은 샘플 크기이며, p는 모형에서 중복되지 않는 모수의 수입니다.

HC2
제곱한 잔차를 1-h로 나눈 값을 포함하는 HC0의 수정입니다. 여기서 h는 케이스의 레버리지입니다.

HC3
잭나이프 추정량의 근사치를 계산하는 HC0의 수정입니다. 제곱한 잔차를 1-h의 제곱으로 나눕니다.

HC4
제곱한 잔차를 1-h의 제곱으로 나누는 HC0의 수정입니다. 이는 h, N 및 p에 따라 변하며 상단 한계는 4입니다.

유의 수준. 사후 검정에 사용되는 유의 수준과 신뢰구간 형성에 사용되는 신뢰수준을 조정할 수 있습니다. 지정된 값은 해당 검정에 대한 관측 검정력을 계산하는 데에도 사용할 수 있습니다. 유의수준을 지정할 때 신뢰구간의 관련된 수준이 대화 상자에 출력됩니다.

GLM 보조 회귀 모형

보조 회귀 모형 대화 상자는 이분산성 검정에 사용되는 모형을 지정합니다.

예측값 사용
상수 항, 예측값에서 선형인 항, 예측값에서 2차인 항, 그리고 오차항으로 구성된 모형을 사용합니다.

일변량 모형 사용
모형 하위 대화 상자에 지정된 모형을 사용합니다. 지정된 모형에 절편 항이 포함되지 않으면 하나가 포함됩니다.

사용자 정의 모형
사용자가 명시적으로 지정하는 모형을 사용합니다.

항 설정
선택된 요인과 공변량 세트의 모든 조합에 대해 특정 유형(예: 주효과)의 비중첩 항을 포함하려면 이 선택사항을 사용합니다.

사용자 정의 항 설정
중첩 항을 포함하고자 하거나 변수별로 항 변수를 명시적으로 설정하고자 할 때 이 선택사항을 사용합니다. 중첩 항 설정에는 다음 단계가 포함됩니다.

UNIANOVA 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- DESIGN 하위 명령을 사용하여 계획에 중첩 효과를 지정합니다.
- TEST 하위 명령을 사용하여 효과에 대한 효과-대 선형 조합에 대한 검정이나 값을 지정합니다.
- CONTRAST 하위 명령을 사용하여 다중 대비를 지정합니다.
- MISSING 하위 명령을 사용하여 사용자 결측값을 포함시킵니다.
- CRITERIA 하위 명령을 사용하여 EPS 기준을 지정합니다.
- LMATRIX, MMATRIX 및 KMATRIX 하위 명령을 사용하여 사용자 정의 **L** 행렬, **M** 행렬 또는 **K** 행렬을 구성합니다.
- CONTRAST 하위 명령을 사용하여 편차나 단순 대비의 경우 중간 참조 범주를 지정합니다.
- CONTRAST 하위 명령을 사용하여 다항 대비에 대한 메트릭 방법을 지정합니다.
- POSTHOC 하위 명령을 사용하여 사후분석 비교에 대한 오차항을 지정합니다.
- EMMEANS 하위 명령을 사용하여 요인 목록에서 요인에 대해 추정된 주변 평균이나 여러 요인 간 요인 상호작용을 계산합니다.
- SAVE 하위 명령을 사용하여 임시 변수 이름을 지정합니다.
- OUTFILE 하위 명령을 사용하여 상관행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 개체 간 분산 분석 테이블로부터 통계를 포함하는 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 계획 행렬을 새 데이터 파일에 저장합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

블랜드 알트만

블랜드 알트만 분석은 평균 차이 간의 편향성을 평가하는 그래픽 기법입니다. 알트만과 블랜드는 1983년에 이 분석을 제안했습니다. 블랜드 알트만 분석은 평균 차이를 탐색하고 일치의 한계를 구성하여 두 정량적 측정값 간의 일치도를 정량화하는 것을 기반으로 합니다.

예제

의료 실험실에서는 종종 두 가지 측정 방법 간의 일치도를 평가하거나, 두 가지 방법 중에서 선택하거나, 새로운 방법 또는 대체 방법을 평가해야 하는 경우가 있습니다. 또한 두 기기 사이에 정렬 문제가 있는 경우, 그 차이와 원인을 측정하고 평가할 수 있는 도구가 필요합니다.

이 절차에는 통계 기본 버전이 필요합니다.

메뉴에서 선택합니다: **분석** > **기술적 통계** > **블랜드 알트만 분석**...

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

데이터 고려 사항: 변수는 숫자로 표시됩니다.

1. 변수를 선택하고 **쌍을 이룬 변수:**에 추가합니다.
2. **OK**를 선택하면 기본 설정을 고려하여 절차가 실행됩니다.
3. **로그 변환 적용** 또는 **페어링만 적용** 또는 둘 다 선택할 수 있습니다.
4. **확인**을 클릭하십시오.

로그 변환 적용: 차이가 정규 분포가 아닌 경우 **로그 변환**을 사용하는 것이 좋습니다.

쌍으로만: 이 옵션은 첫 번째 열의 첫 번째 변수를 두 번째 열의 첫 번째 변수와 비교할 때 지정합니다.

참고: 쌍을 이루는 변수 중 가장 적은 수의 값을 포함하는 변수 집합의 마지막 값까지만 분석이 수행됩니다. 분석은 동일한 수의 값을 포함하는 쌍을 이루는 변수 세트에 대해 완전히 수행됩니다.

블랜드 알트만: 기준

블랜드 아트만: 기준 탭에서 기본 설정을 변경할 수 있습니다.

1. 기본적으로 비율을 백분율로 표시하는 옵션은 선택되어 있지 않습니다. 추가하려면 **비율을 백분율로 표시**를 선택합니다.
2. **범위** 및 **계약 한도**의 설정을 확인하고 필요한 경우 변경합니다.
3. **계속**을 클릭합니다.

블랜드 알트만: 인쇄

1. 출력에 포함할 옵션을 선택합니다.
2. **분석 세부 정보 인쇄를** 선택하면 표시할 최대 행 수에 대한 기본 설정이 10으로 설정됩니다. 디스플레이에 표시할 행 수를 늘리거나 줄이려면 숫자를 변경합니다. 표시할 수 있는 최대 데이터 행은 1000개입니다. **정렬**은 **기준** 네 가지 목록 옵션과 오름차순 또는 내림차순 **주문**이 될 수 있습니다.

   기본적으로 **Sort**는 첫 번째 측정값을 기준으로 내림차순으로 설정됩니다.
3. **계속**을 클릭합니다.

블랜드 알트만: 줄거리

블랜드 아트만 분석에서 기본 설정을 변경할 수 있습니다: 플롯 탭에서 변경할 수 있습니다.

1. **비율 플롯, 차이 0의 기준선, 합의 한계, 평균 차이의 신뢰 구간** 등 원하는 설정을 선택합니다.
2. **계속**을 클릭합니다.

카이제곱 검정

카이제곱 테스트 절차는 변수를 카테고리별로 표로 만들고 카이제곱 통계를 계산합니다. 카이제곱 검정은 두 범주형 변수 간에 유의미한 연관성이 존재하는지 여부를 확인하는 데 사용되는 통계적 가설 검사입니다. 이 테스트는 변수 간에 연관성이 없다고 가정하는 귀무가설 하에서 데이터에서 관찰된 주파수가 예상 주파수와 다른지 여부를 평가합니다.

이 절차에는 통계 베이스 에디션이 필요합니다.

예

시장 조사에서 분석가는 소비자 선호도를 이해해야 합니다. 이를 위해 관찰된 고객 행동과 예상되는 패턴을 비교하여 마케팅 전략을 효과적으로 수립합니다.

통계

피어슨 카이제곱, 가능성 비율 카이제곱, 피셔의 정확한 검정, 예이츠의 보정 카이제곱(연속성 보정)이 있습니다.

데이터 고려 사항 및 가정

정렬되었거나 정렬되지 않은 숫자 범주형 변수(측정의 순서 수준이나 명목 수준)를 사용합니다. 문자열 변수를 숫자변수로 변환하려면 변환 메뉴의 자동 코딩변경 프로시저를 사용합니다.

비모수 검정은 기본 분포의 형태에 대한 가정이 필요하지 않습니다. 데이터는 무작위 표본이라고 가정합니다. 각 범주의 예상 빈도는 1 이상이어야 합니다. 범주의 20% 이하가 5미만의 예상 빈도를 가지고 있어야 합니다.

메뉴에서 **기술 통계 카이** > **제곱** > **분석을** 클릭합니다.

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

1. 변수를 선택하고 **행과 열에** 변수를 추가합니다.
2. **확인**을 클릭하십시오. 절차는 기본 설정을 고려하여 실행됩니다.

Chi 스퀘어 테스트: 정확

- 점근. 검정 통계량의 점근 분포를 기반으로 하는 유의수준입니다. 일반적으로 0.05보다 작은 값은 유의한 것으로 간주됩니다. 점근 유의수준은 데이터 세트가 크다는 가정을 기반으로 합니다. 데이터 세트가 작거나 분포가 고르지 못한 경우에는 유의수준 표시가 부적합할 수 있습니다.
- *Monte Carlo* 추정값*(Monte Carlo Estimate)*. 차원 및 행과 열 여백이 관측 테이블과 동일한 테이블의 참조 세트에서 반복적으로 표본추출하여 계산되는 정확한 유의 수준에 대한 편향되지 않은 추정값입니다. Monte Carlo 방법을 사용하면 점근법에는 꼭 필요한 가정을 사용하지 않고도 정확한 유의수준을 추정할 수 있습니다. 이 방법은 데이터 세트가 너무 커서 정확한 유의수준을 계산할 수 없을 때 유용하게 사용되지만 이러한 데이터는 점근법의 가정을 만족시키지 않습니다.
- 정확. 관측 결과 또는 더 극단적인 결과의 확률이 정확하게 계산됩니다. 일반적으로 0.05보다 작은 유의 수준은 유의 수준이 있는 것으로 간주되며 행 변수와 열 변수 간에 관계가 있다는 의미입니다.

제곱 테스트: 셀

Chi Square 절차는 예상 주파수를 표시합니다. 표의 각 셀에는 개수와 백분율의 모든 조합을 포함할 수 있습니다.

빈도

행과 열 변수가 서로 독립적인 경우 실제로 관찰되는 경우의 수와 예상되는 경우의 수입니다. 지정된 정수보다 작은 수를 숨기도록 선택할 수 있습니다. 숨겨진 값은 **<N**로 표시됩니다. 여기서 **N**은 지정된 정수입니다. 0 값이 허용되어 숨겨진 수가 없도록 지정한 경우라도 지정된 정수는 2와 같거나 그 이상이어야 합니다.

퍼센트

퍼센트는 행 가로방향으로나 열 세로방향으로 추가될 수 있습니다. 표(하나의 레이어)에 나타난 총 케이스 수의 퍼센트도 사용할 수 있습니다.

참고: 빈도 그룹에서 **낮은 빈도 숨기기**가 선택된 경우, 숨겨진 빈도와 연관된 퍼센트도 숨겨집니다.

APA 유형 표 작성

APA 유형 지침을 준수하는 출력결과 표를 작성합니다.

참고: APA 유형 표 작성을 선택한 경우에는 **관측**, **기대**, **행**, **열**, **총계** 옵션을 사용할 수 없습니다.

카이제곱 테스트: 형식

형식 옵션을 사용하면 테스트 결과에 표시되는 변수의 순서를 지정할 수 있습니다.

기술 통계 > **카이제곱** > **분석** 으로 이동합니다. **Chi 스퀘어 테스트** 대화 상자에서 **포맷을** 클릭합니다.

행을 행 변수 값의 오름차순이나 내림차순으로 정렬할 수 있습니다. **행 순서에서 오름차순** 또는 **내림차순을** 선택합니다.

이변량 상관계수

이변량 상관계수 프로시저는 Pearson 상관계수, Spearman의 Rho 및 Kendall의 타우-*b*를 계산하고 각 유의수준과 함께 출력합니다. 상관은 변수나 순위 순서의 관련성을 측정합니다. 상관계수를 계산하기 전에 잘못된 결과

를 나타낼 수 있는 이상값의 유무와 데이터의 선형성 여부를 확인합니다. Pearson의 상관은 선형 연관의 측도입니다. 두 변수가 완벽하게 관련되어 있을 수는 있으나 선형 관계가 아닌 경우 Pearson의 상관은 그 연관 측정에 적합한 통계가 아닙니다.

신뢰구간 설정은 Pearson과 Spearman에 사용할 수 있습니다.

예
한 농구 팀이 이긴 경기 수는 게임당 평균 스코어와 관련이 있습니까? 산점도를 통해 선형 관계를 파악할 수 있습니다. 1994 – 1995년 NBA 시즌의 데이터를 분석해 보면 Pearson의 상관계수(0.581)는 0.01 수준에서 유의합니다. 시즌마다 경기에서 많이 이길수록 상대편 팀의 스코어는 낮아진다고 생각할 수 있습니다. 이러한 변수는 음(-0.401)으로 상관되며 0.05 수준에서 유의합니다.

통계
각 변수에 대해서는 경우 평균, 표준 편차, 비결측값이 있는 케이스 수를 선택할 수 있습니다. 각 대응변수에 대해서는 Pearson의 상관계수, Spearman의 Rho, Kendall의 타우-*b*, 편차의 교차곱, 공분산을 선택할 수 있습니다.

데이터 고려사항

데이터
Pearson의 상관계수에 대해 대칭인 양적변수를 사용하고 Spearman의 Rho와 Kendall의 타우-*b*에 대해 양적변수나 범주의 순서가 지정된 변수를 사용합니다.

가정
Pearson의 상관계수에서는 각 대응변수를 이변량 정규 분포로 가정합니다.

이변량 상관계수 구하기

메뉴에서 다음을 선택합니다.

분석 > **상관분석** > **이변량...**

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

1. 둘 이상의 숫자변수를 선택합니다.

   다음 옵션을 사용할 수도 있습니다.

   상관계수
   정규 분포된 양적 변수의 경우 **Pearson** 상관계수를 선택합니다. 데이터가 정규 분포를 따르지 않거나 범주 순서가 지정되어 있지 않은 경우 순위 순서 간 연관을 측정하는 **Kendall의 타우-b**나 **Spearman**을 선택합니다. 상관계수 범위는 -1(완전 음의 관계)부터 +1(완전 양의 관계)까지입니다. 0의 값은 선형 관계가 아님을 나타냅니다. 결과 해석시 유의한 상관으로 기인한 인과 관계를 도출하지 않도록 합니다.

   유의성 검정
   양측 확률 또는 단측 확률을 선택할 수 있습니다. 연관 방향을 이미 알고 있는 경우에는 **한쪽**을 선택합니다. 모를 때는 **양쪽**을 선택합니다.

   유의한 상관계수 플래그
   상관계수의 유의수준이 0.05이면 별표 하나로 표시되고 0.01이면 별표 두 개로 표시됩니다.

   하한 삼각형만 표시
   선택할 경우 상관행렬 표의 하한 삼각형만 출력결과에 표시됩니다. 선택하지 않을 경우 전체 상관행렬 표가 출력결과에 표시됩니다. 이 설정을 사용하면 표 출력결과가 APA 양식 지침을 준수하게 됩니다.

   대각 표시
   선택할 경우 상관행렬 표의 아래쪽 삼각형이 대각 값과 함께 출력결과에 표시됩니다. 이 설정을 사용하면 표 출력결과가 APA 양식 지침을 준수하게 됩니다.
2. 선택적으로 다음을 선택할 수 있습니다.
   - **옵션을 클릭하세요...** 을 클릭하여 피어슨 상관관계 통계 및 결측치 설정을 지정합니다.
   - **스타일을 클릭하세요...** 를 사용하여 특정 조건에 따라 피벗 테이블의 속성을 자동으로 변경하는 조건을 지정할 수 있습니다.

- **부트스트랩을 클릭하세요...** 를 사용하여 평균, 중앙값, 비율, 확률 비율, 상관 계수 또는 회귀 계수와 같은 추정치에 대한 표준 오차 및 신뢰 구간의 견고한 추정치를 도출할 수 있습니다.
- **신뢰 구간을 클릭하세요...** 을 클릭하여 신뢰 구간 추정 옵션을 설정합니다.

이변량 상관: 옵션

통계
Pearson 상관의 경우, 다음 중 하나 또는 모두를 선택할 수 있습니다.

평균과 표준편차
각 변수에 대해 표시됩니다. 비결측값이 있는 케이스 수도 함께 표시됩니다. 결측값의 설정에 관계 없이 변수별로 결측값을 처리할 수 있습니다.

교차곱 편차와 공분산
각 대응변수가 표시됩니다. 편차의 교차곱은 수정된 평균 변수 곱의 합과 같습니다. 이 값은 Pearson 상관계수의 분자가 됩니다. 공분산은 두 변수 간 관계의 비표준화 측도이며 교차곱 편차를 *N*–1로 나눈 값과 같습니다.

결측값
다음 중 하나를 선택할 수 있습니다.

대응별 케이스 제외
한 상관계수의 대응변수 하나 또는 모두에 대한 결측값이 있는 케이스는 분석시 제외됩니다. 각 계수는 특정 대응변수의 유효 코드가 들어 있는 모든 케이스에 기초하므로 각 계산에서 정보를 최대한 사용할 수 있습니다. 이로 인해 다양한 케이스에 기초하는 계수 세트가 생성될 수 있습니다.

목록별 결측값 제외
변수에 대한 결측값을 포함하는 케이스는 모든 상관에서 제외됩니다.

이변량 상관계수 신뢰구간

신뢰구간 대화 상자는 신뢰구간 추정 옵션을 제공합니다. 이 대화 상자는 이변량 상관계수 대화 상자에서 **Pearson**, **Kendall의 타우-b** 또는 **Spearman**을 선택한 경우에 사용가능합니다.

이변량 상관계수 모수의 신뢰구간 추정
이변량 상관계수 모수의 신뢰구간 추정을 제어합니다. 선택할 경우 신뢰구간 추정이 발생합니다.

신뢰구간(%)
생성된 모든 신뢰구간의 신뢰수준을 지정합니다. 0-100 사이의 숫자 값을 지정하십시오. 기본값은 95입니다.

Pearson 상관
편향 조정 적용 설정은 편향 조정을 적용할지 여부를 제어합니다. 기본적으로 이 설정은 선택되어 있지 않으므로, 편향 항을 고려하지 않습니다. 선택할 경우 신뢰한계 추정에 대한 편향 수정이 적용됩니다. 이 설정은 이변량 상관계수 대화 상자에서 **Pearson**을 선택한 경우에 사용가능합니다.

Spearman 상관
이 설정은 이변량 상관계수 대화 상자에서 **Spearman**을 선택한 경우에 사용가능하며, 다음 방법을 통해 Spearman 상관 분산을 추정하는 옵션을 제공합니다.

- **Fieller, Hartley 및 Pearson**
- **Bonett 및 Wright**
- **Caruso 및 Cliff**

CORRELATIONS 및 NONPAR CORR 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 요인 분석(MATRIX 하위 명령 사용)과 같이 다른 분석을 얻기 위해 원시 데이터 대신 사용할 수 있는 Pearson 상관계수에 대한 상관행렬을 기록합니다.
- 특정 목록의 각 변수 및 두 번째 목록의 각 변수의 상관을 구합니다(VARIABLES 하위 명령에서 키워드 WITH 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

근접성 맵핑

이 문서는 입력 데이터의 사양, 모델 선택, 변환 함수, 차원 설정, 근접 매핑의 속성 및 속성 포함에 대해 안내합니다.

근접 매핑은 다변량 데이터의 차원을 줄이고 공간 구성에서 개체(사례, 항목 또는 기타 엔티티) 간의 관계를 표시하는 데 사용되는 시각화 기법입니다. 원래 심리학에서 다차원 스케일링(MDS)이라는 이름으로 개발된 이 방법은 근사치(유사성, 유사성 또는 거리)를 사용하여 객체의 공간 배열이 데이터의 구조를 반영하도록 객체의 위치를 지정합니다. 비슷한 물체는 서로 가깝게 표시되고 서로 다른 물체는 더 멀리 떨어져 배치됩니다. 근접 매핑은 복잡한 관계형 데이터를 직관적인 2차원 또는 3차원 도표로 변환하여 해석 및 패턴 인식을 지원합니다.

IBM SPSS 통계 에서 근접 매핑은 PROXMAP 절차를 통해 구현됩니다. 근접 행렬 분석에 국한된 PROXSCAL과 달리 PROXMAP은 다양한 형태의 입력을 허용하고 더 큰 유연성을 제공합니다. 여러 근접 소스를 지원하고, 추가 변수(속성 및 속성)를 통합하며, 다양한 변환 및 제한 옵션을 제공합니다. 이러한 기능 덕분에 PROXMAP은 다변량 구조를 탐색하고 다양한 유형의 데이터를 통합된 공간 표현으로 통합하는 강력한 도구가 되었습니다.

직접 지정되거나 다변량 데이터에서 파생된 여러 소스에서 근접성 정보를 사용할 수 있는 경우, PROXMAP은 기본 신원 모델 외에도 여러 개인차 모델을 적용할 수 있습니다. 사용 가능한 모델에는 가중 아이덴티티(확장) 모델, 가중 유클리드(대각선) 모델, 일반화된 유클리드 모델(선택적 등급 제한 포함)이 있습니다.

주성분 분석(PCA) 및 고전적 MDS와 같은 기존의 차원 축소 기법에 비해 PROXMAP은 보다 직접적이고 유연한 근접성 표현을 제공합니다. 기존 MDS를 사용하여 구성을 초기화할 수 있습니다. 다변량 데이터에 적용할 경우, 클래식 MDS는 PCA의 변형인 주좌표 분석과 동일합니다.

PROXMAP은 다음 유형의 입력을 허용합니다.

입력 유형	설명
원시 근접성 및 파생 근접성	파생된 근사값은 숫자, 서수 또는 명목 변수에서 계산됩니다. 두 가지 유형의 근사값은 선택적으로 단조 함수(예: 서수 또는 스플라인 기반)를 사용하여 변환한 후 거리로 근사값을 구합니다.
속성	구성을 제약하고 조인트 공간에서 방향으로 표시되는 변수(감독 매핑)입니다.
특성	해석에 사용되는 보조 변수로, 공간에 표시되지만 구성에 영향을 미치지 않습니다(비지도 매핑).

PROXMAP의 고유한 기능은 다양한 역할에서 동일한 변수 집합을 사용할 수 있다는 점입니다:

- 근접성 도출의 기초로서
- 구성을 형성하는 속성으로
- 해석을 지원하는 속성으로

또는 각 역할에 별도의 변수 집합을 할당할 수도 있습니다.

선형 변환 외에도 PROXMAP은 속성과 속성 모두에 대해 다양한 변환 옵션(서수, 단조로운 스플라인, 비단조로운 스플라인, 공칭 변환 포함)을 제공합니다. 근사값의 경우 변환은 단조롭고 전력 변환도 포함해야 합니다. 출력에는 스트레스 플롯, 최소 스패닝 트리, 이웃 그래프와 같은 그래픽 진단이 포함되어 해석과 평가에 도움이 됩니다.

예제

근접 매핑은 심리학 및 행동 과학(지각, 기억, 선호도), 마케팅(브랜드 매핑, 세분화), 사회학(이념 매핑, 소셜 네트워크), 교육 및 심리 측정(테스트 진단, 기술 프로파일링), 언어학(의미 유사성, 방언 매핑) 등 여러 분야에 폭넓게 적용될 수 있습니다, 지리학(지각된 거리, 생태 공간 분석), 유전체학 및 생물정보학(발현 패턴, 인구 구조, 단일 세포 발현 이질성 탐색), 화학정보학(분자 유사성), 컴퓨터 과학(임베딩 시각화), 인류학 및 고고학(유물 유사성, 두개골 데이터).

데이터

분석할 데이터에는 근접 행렬(또는 행렬)을 나타내거나 근접 행렬 또는 행렬로 변환된 다변량 데이터를 나타내는 변수가 포함됩니다.

다변량 데이터의 경우 변수는 척도, 서수 및 공칭 측정 수준 변수가 혼합되어 있을 수 있습니다. 필요한 최소 변수 수는 3개입니다

근접 데이터의 경우 사례 수와 동일한 수의 변수를 입력합니다.

근접 데이터와 다변량 데이터 모두에 필요한 최소 사례 수는 3개입니다.

출력에는 공간 구성, 다양한 진단 플롯 및 적합도 통계가 포함됩니다. 바이플롯은 공통 공간에 변수, 속성 및 프로퍼티와 함께 개체를 표시합니다.

PROXMAP 분석을 시작하려면 **매핑** > **근접 매핑** > **분석을** 클릭하세요.

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

근접 매핑: 데이터

데이터 탭을 사용하면 데이터 세트의 변수를 기반으로 하나 이상의 근접 소스를 정의할 수 있습니다. 각 소스는 활성 데이터 집합에서 변수를 선택하고 근접도 행렬을 나타내는지 아니면 근접도를 도출할 다변량 데이터를 나타내는지 지정하여 생성됩니다.

변수

변수 목록에는 활성 데이터 집합의 모든 변수가 표시됩니다. 변수에는 근접도 계산에 사용할 수 있는 숫자, 서수, 명목 변수가 포함됩니다. 하나 이상의 변수를 선택하고 **소스의 변수로** 이동합니다.

소스의 변수

소스의 변수에는 현재 활성 근접 소스에 대해 선택된 변수가 나열됩니다. 이러한 변수는 소스의 구조를 정의하며, 다변량 데이터의 경우 근사값을 도출하거나 (근사값이 직접 제공되는 경우) 근접 행렬을 지정하는 데 사용됩니다. 위쪽 또는 아래쪽 화살표 버튼을 사용하여 변수의 순서를 바꾸거나 필요에 따라 변수를 제거할 수 있습니다.

오브젝트 레이블

개체 레이블을 사용하면 출력 플롯과 테이블에 표시할 개체(케이스)에 레이블을 제공하는 변수를 지정할 수 있습니다. 목록에서 변수를 선택하고 **개체 레이블로** 이동하여 할당합니다.

소스 정의

소스의 **변수** 오른쪽에 있는 소스 이름과 소스 레이블 필드는 현재 근접성에 대한 내부 식별자 및 설명 레이블을 정의합니다 source.The **소스 이름은** 앞에 *$*가 오는 고유 식별자이어야 합니다.

다음 버튼을 사용하여 소스를 관리합니다.

- 소스 **추가는** 소스의 변수 창에 있는 변수에서 새 근접 소스를 만듭니다.
- **소스 변경은** 업데이트된 변수 선택을 사용하여 현재 선택된 소스를 수정합니다.
- **소스 제거는** 현재 선택된 소스를 목록에서 삭제합니다.

데이터 유형

데이터 유형 선택기를 사용하여 선택한 변수가 해석되는 방식을 지정합니다. **데이터 유형** 그룹을 사용하면 선택한 변수에서 근사값을 도출하는 방법을 표시할 수 있습니다:

- **근접 데이터** 입력이 이미 근접 행렬인 경우 이 옵션을 사용합니다.
- **다변량 데이터** (기본값)는 선택한 변수를 개체의 특징으로 취급합니다. PROXMAP은 이러한 변수를 기반으로 객체 간의 거리를 계산하여 근접성을 도출합니다.
  - **유클리드** : 근거리: 표준화되지 않은 유클리드 거리를 기준으로 합니다.
  - **표준화된 유클리드** : 거리를 계산하기 전에 변수를 표준화합니다(기본값).

가중값

데이터 유형을 근접 데이터로 설정하면 선택 사항인 **근접 가중치** 목록을 사용할 수 있습니다. 여기에서 근접도 값과 연관시킬 추가 가중치 변수를 지정할 수 있습니다. 이러한 가중치는 최적화 프로세스에서 각 근접성의 영향을 조정합니다.

근접 매핑: 가중치

가중치 탭을 사용하면 각 소스에 대한 근접성(셀) 가중치와 근접성 데이터에 대한 개체(대소문자) 가중치를 지정할 수 있습니다. 이러한 가중치는 근접성이 최적화 프로세스에 기여하는 방식에 영향을 미칩니다. 이 탭은 **데이터** 탭에서 **데이터 유형인 근접 데이터를** 선택한 경우에만 해당됩니다.

변수

변수 목록에는 활성 데이터 집합의 모든 변수가 표시됩니다. 이러한 변수에는 근접 가중치가 포함되어야 합니다. 하나 이상의 변수를 선택하고 **소스의 가중치로** 이동합니다. 이러한 변수에는 특정 소스에 대한 근접성 가중치가 포함되어야 합니다.

소스의 가중치

소스의 가중치에는 현재 활성화된 소스와 연결된 선택된 가중치 변수가 표시됩니다. 이러한 변수는 해당 근접성 행렬의 구조와 일치하는 근접성 가중치 행렬을 정의합니다. 가중치 변수의 수는 소스 행렬의 변수 수와 같아야 합니다.

가중치 **추가**, 가중치 **변경** 또는 가중치 **제거를** 사용하여 **소스에서** 선택한 소스에 대한 가중치 정의를 관리합니다.

출처(들)

소스 목록에는 가중치를 할당할 수 있는 정의된 모든 근접 소스가 표시됩니다. 소스를 선택하여 관련 가중치를 보거나 수정합니다.

오브젝트(케이스) 가중치

선택 창 아래의 개체 레이블 필드에서 개별 사례(개체)에 대한 가중치를 포함하는 변수를 정의할 수 있습니다. 이러한 가중치는 근접 데이터를 사용할 때 손실 함수에서 각 객체의 영향을 조정합니다. 다변량 데이터에는 대/소문자 가중치를 적용할 수 없습니다.

가중치 함수

가중치 함수는 근접도 값의 함수인 특수 근접도 가중치 체계에 대한 세 가지 옵션을 제공합니다:

- **없음** : 없음: 특별한 가중치가 적용되지 않습니다. 사용자가 지정한 근접도 및 객체 가중치(있는 경우)만 사용됩니다.
- **샘몬** : 객체 *i*와 *j* 사이의 근접도에 가중치를 적용합니다. 이 가중치는 샘몬 매핑에서와 같이 더 작은 근접성을 강조합니다.
- **에너지** : 더 작은 근접성에 더 중점을 두는 가중치를 적용합니다.

각 근접도에 적용되는 총 가중치는 사용자가 정의한 근접도 가중치, 케이스 가중치, 특수 가중치 함수의 곱입니다. 이렇게 도출된 가중치는 PROXMAP에서 최소화하는 손실 함수에서 각 근접도의 상대적 중요도를 결정합니다.

근접 매핑: 초기

초기 탭은 PROXMAP에서 사용되는 반복적 최적화 프로세스의 시작 구성을 지정합니다. 초기 구성을 잘하면 컨버전스를 가속화하고 솔루션 품질을 개선할 수 있습니다. 원시 근사값에 대한 클래식 스케일링을 사용한 자동 초기화 또는 순위 순서에 따른 자동 초기화 중에서 선택할 수 있습니다. 데이터 세트에서 사용자 지정 구성을 제공할 수도 있습니다.

클래식 스케일링

기본 옵션입니다. 고전적인 스케일링 솔루션(주좌표 분석)은 근사값을 기반으로 계산됩니다. 지정된 수의 차원으로 분석적으로 도출된 시작 구성을 제공합니다.

두 가지 추가 옵션은 클래식 스케일링 중에 근사값을 처리하는 방법을 제어합니다:

- 상수 **값 추가를** 사용하면 스케일링하기 전에 모든 근사치에 추가할 고정 상수를 지정할 수 있습니다. 이렇게 하면 메트릭 속성을 개선하고 음의 고유값을 피할 수 있습니다. 이 값은 음의 근사값이 되지 않아야 합니다.
- **추가 상수 추정** 옵션을 사용하면 PROXMAP에서 추가할 최적의 상수를 자동으로 추정합니다. 이렇게 하면 결과 거리 행렬이 양의 반정확한 양수 행렬이 됩니다.
- 근사값에 **순위 번호 사용을** 사용하면 원시 값이 아닌 근사값의 순위 순서에 고전적인 배율을 적용할 수 있습니다. 이렇게 하면 `ordinal` 시작 구성이 생성되고 이상값에 대한 민감도가 감소합니다.

사용자 정의 설정

사용자 지정 구성을 사용하면 직접 초기 좌표를 제공할 수 있습니다. 데이터 집합의 변수가 시작 구성의 차원으로 선택됩니다. 이러한 변수는 누락된 데이터가 없는 유효한 숫자 값을 포함해야 합니다.

변수 목록에는 사용 가능한 모든 변수가 포함되어 있습니다. 사용자 지정 구성이 포함된 변수를 선택하고 **차원으로** 이동합니다.

차원 목록에는 초기 좌표로 사용하도록 선택된 변수가 표시됩니다. 선택한 변수의 수는 차원 수와 일치해야 합니다. 더 많은 변수가 제공되면 값이 처음 *p*개의 주성분으로 투영됩니다(여기서 *p*는 차원 수).

현재 차원 수에는 최종 솔루션에 지정된 차원이 표시됩니다(**기준** 탭에 지정됨).

근접 매핑: 모델

모델 탭에서는 여러 근접 소스를 분석할 때 적용할 개인차 모델을 지정할 수 있습니다. 또한 소스 간 근접 변환이 균일하게(무조건) 또는 독립적으로(매트릭스 조건부) 처리되는 방식도 정의합니다.

3원 모델

3방향 모델의 설정은 공간 가중치를 사용하여 여러 소스의 근접성을 공통 공간 표현으로 매핑하는 방법을 결정합니다. 모델은 공용 공간에 공간 가중치를 적용하는 방법을 결정합니다. 일부 모델의 경우 공간 가중치는 공간 가중치와 동일합니다. 단일 소스만 정의된 경우 **ID를** 제외한 모든 옵션이 비활성화됩니다.

옵션	설명
항등	모든 소스가 소스별 스케일링 없이 동일한 구성을 공유한다고 가정합니다. 단일 소스를 지정한 경우에만 사용할 수 있습니다.
확장 (가중 신원)	각 소스는 모든 차원에 균일하게 적용되는 단일 스케일링 계수를 사용하여 모델링됩니다. 이 옵션을 사용하면 각 소스가 공유 공간을 전체적으로 다른 규모로 인식할 수 있습니다.
가중 유클리드 (대각선)	각 소스는 공통 공간의 치수에 서로 다른 가중치를 할당합니다. 이를 통해 소스는 다양한 공간적 특징을 강조할 수 있습니다.
일반화 유클리디안	각 소스에는 공통 공간의 회전 및 가중치 조정(스케일링) 버전이 허용됩니다. 이 모델은 가장 유연한 모델이며 다양한 개인차를 수용할 수 있습니다. • **전체 랭크** : 각 소스는 공유 공간의 전체 차원을 사용합니다. • **감소된 등급** : 축소된 순위 제약 조건을 적용하면 각 소스의 구성(개별 공간, 개별 공간과 함께순위 = 차원 수 - 1)는 공용 공간의 저차원 하위 공간에 표시됩니다. 이 접근 방식은 공통 구성의 직교 회전과 하위 공간에 대한 투영을 포함합니다. 여기서 회전과 투영이 함께 소위 공간 가중치를 형성합니다.

조건부 / 변환

이 창에서는 구성에서 거리를 맞추기 전에 근거리 변환 방식을 제어합니다.

- **무조건** 옵션을 사용하면 모든 소스에 단일 변환 기능을 적용할 수 있습니다. 이 옵션은 소스가 공통 구조를 반영한다고 가정할 때 적합합니다.

  기본 변환은 **변환** 아래에 표시됩니다(동점을 풀 수 있는 서수).

  변환 정의를 클릭하면 모든 변환 유형(선형, 파워, 단조, 스플라인, 서수), 동점 처리 및 기타 매개변수 중에서 선택할 수 있는 대화 상자가 열립니다.

- **매트릭스-조건부** 옵션을 사용하면 다양한 소스 또는 소스 그룹에 대해 서로 다른 변환 함수를 선택할 수 있습니다. 선택하면 각 소스가 기본 변환 기능과 함께 나열됩니다.

  변환 정의를 클릭하여 선택한 소스에 적용할 다른 기능을 선택합니다. 이 옵션은 둘 이상의 소스가 정의된 경우에만 활성화됩니다.

변환 정의에서 공간 구성의 거리와 일치하기 전에 근사값을 최적으로 변환하는 방법을 선택할 수 있습니다. 다음 옵션을 사용할 수 있습니다.

옵션	설명
없음	변환이 적용되지 않습니다. 근사치는 그대로 사용됩니다.
선형	근사값에 직선 변환을 적용합니다. **인터셉트** (체크박스)는 변환에 상수를 추가하여 간격 스케일 모델링을 허용합니다. 이 옵션을 선택하지 않으면 비율 스케일 변환이 적용됩니다(상수 추가 없음).
거듭제곱	전력 변환을 적용합니다. **인터셉트** (체크박스)를 활성화하면 파워 함수에 상수 시프트를 추가할 수 있습니다.
스플라인	근사값에 부드러운 부분 다항식(스플라인) 함수를 맞춥니다. **인터셉트에는** 스플라인 기준의 상수 항이 포함됩니다. **정도는** 다항식 세그먼트의 복잡도를 설정합니다(예: 1 = 선형, 2 = 이차). **내부 매듭은** 스플라인 모양을 정의하는 데 사용되는 내부 점의 수를 제어합니다.
마디 순서	동일한 간격: 매듭점이 근접 범위를 따라 균등한 간격으로 배치됩니다. 백분위수: 매듭점은 근접 분포의 경험적 사분위수에 배치됩니다.
순서형	순위 보존 변환을 적용합니다(단조로운). **동점을 풀 수 있도록 허용합니다** : 묶인 근사값을 다른 값으로 변환할 수 있습니다. **동점 유지** : 묶인 근사값은 변환 후에도 묶인 상태로 유지됩니다.

근접 매핑: 제한 사항

제한 탭을 사용하면 속성 변수를 지정하여 솔루션 공간을 제한할 수 있습니다. 이러한 변수는 공통 객체 공간을 속성의 선형 조합으로 제한하여 구성을 능동적으로 안내합니다. 이를 통해 의미 있는 외부 변수에 차원을 고정하여 해석 가능성을 개선하는 감독 매핑이 가능합니다.

다음 옵션 중 하나를 선택하십시오.

없음
제한은 없습니다. 구성은 (파생된) 근접성에 의해서만 결정됩니다.

Attributes
하나 이상의 외부 변수를 속성 제약 조건으로 사용하여 감독 매핑을 활성화합니다. 이 옵션을 선택하면 어떤 변수가 어트리뷰트로 작동하는지 정의하고 각 변수가 변환되는 방식을 지정할 수 있습니다.

변수

왼쪽의 **변수** 목록에는 데이터 집합에서 사용 가능한 모든 변수가 표시됩니다. 변수를 하나 이상 선택하고 화살표 버튼을 클릭하여 **속성으로** 이동합니다.

Attributes

오른쪽 창에는 선택한 모든 속성 변수가 기본 변환 함수와 함께 표시됩니다. 이러한 변환은 제한된 구성에서 사용하기 위해 속성을 확장하거나 변환하는 방법을 제어합니다.

변환 정의를 클릭하여 선택한 속성에 변환 함수를 할당합니다. 다음 옵션을 사용할 수 있습니다.

- **리니어는** 표준화된 척도로 속성을 유지합니다.
- **스플라인은** 조각별 다항식 함수를 사용하여 부드러운 단조 또는 비단조 변환을 적용합니다.
- **서수는** 순위 순서만 유지합니다.
- **Nominal은** 카테고리를 정렬되지 않은 것으로 취급하여 최적으로 맞춥니다. 단일 변환을 제공합니다(프로퍼티의 명목 변환과 달리).

속성 필드에서 여러 행을 선택하여 여러 속성에 대해 동일한 변환을 정의할 수 있습니다.

공통 공간은 객체 좌표가 변환된 속성에 의해 스팬된 하위 공간 내에 있다는 제약 조건 하에 추정됩니다. **현재 차원 수** 필드에는 **기준** 탭에서 설정된 솔루션의 차원이 표시됩니다. 항상 차원 수보다 많은 수의 속성을 선택합니다.

근접 매핑: 보조

보조 탭을 사용하면 공간 구성의 해석에 도움이 되지만 추정에는 영향을 주지 않는 속성 변수를 정의할 수 있습니다(비지도 매핑).

변수
왼쪽의 **변수** 목록에는 활성 데이터 집합에서 사용 가능한 모든 변수가 표시됩니다. 속성으로 처리할 변수를 하나 이상 선택하고 **속성** 목록으로 이동합니다. 원하는 수의 속성을 선택할 수 있습니다.

특성

속성 창에는 속성 변수로 지정된 모든 변수가 각 변수에 할당된 변환과 함께 나열됩니다.

변환 정의를 클릭하여 선택한 하나 이상의 속성 변수에 변환 함수를 할당합니다. 사용 가능한 옵션은 다음과 같습니다:

- **리니어는** 표준화된 스케일로 속성을 유지합니다.
- **스플라인은** 조각별 다항식 함수를 사용하여 부드러운 단조로운 또는 비단조로운 변환을 적용합니다.
- **서수는** 순위 순서만 유지합니다.
- **공칭** 적합 카테고리는 구성에서 연결된 객체 포인트의 중심점으로 사용됩니다.

근접 매핑: 기준

기준 탭에서는 솔루션의 차원, 최적화 프로세스의 수렴 동작, 솔루션 후 표준화 및 결과 구성의 방향을 제어하는 알고리즘 설정을 지정합니다.

차원(Dimensionality)
공용 공간의 치수 수를 설정합니다. 이 값은 2 이상이어야 하고개체 수 - 1. 치수 수를 늘리면 맞춤도가 향상될 수 있지만 해석 가능성이 떨어질 수 있습니다.

알고리즘 수렴
이 설정은 반복 최적화가 중지되는 시점을 결정합니다.

필드	설명	기본값
함수 최소값	솔루션이 허용되는 손실 함수 값의 임계값을 설정합니다. 손실 함수의 값이 이 최소값 아래로 떨어지면 반복이 중지됩니다. 값은 0 이상이어야 합니다.	1.0×10^{-6}
함수 차이	반복 사이의 손실 함수 변화에 따라 수렴 기준을 설정합니다. 함수 값의 절대 차이가 이 값 아래로 떨어지면 알고리즘이 수렴한 것으로 간주합니다. 값은 0 이상이어야 합니다.	1.0×10^{-12}
구성 차이	반복 간 구성 변경에 대한 수렴 임계값을 설정합니다. 객체 위치의 총 이동이 이 값 아래로 떨어지면 알고리즘이 중지됩니다. 값은 0 이상이어야 합니다.	1.0×10^{-6}
최대 반복	알고리즘이 수행할 수 있는 최대 반복 횟수를 지정합니다. 이 수까지 수렴에 도달하지 못하면 반복이 중단됩니다. 값은 0 이상이어야 합니다.	1000

표준화

거리 계산 전에 다변량 변수(근사값을 도출하는 데 사용되는 경우)를 표준화하는 방법을 지정합니다. 이 옵션은 속성 및 속성 변수에도 적용됩니다.

- 기본적으로 **제곱합 = N이** 선택되어 있습니다.
- 다른 통계 절차에 일관성을 유지하려면 **제곱합 = N−1** 을 선택합니다.

차원

좌표 공간에서 공통 공간 구성의 방향을 지정합니다.

라디오 단추	설명
없음/있는 그대로	제약 조건이 적용되지 않습니다. 방향은 주축 방향이며, 치수에 대한 부호는 임의입니다.
차원 간 양의 상관	차원과 데이터 간의 양의 상관 관계를 최대화하도록 구성의 반영을 설정합니다.
최대 절대 좌표에 대한 양의 사분점	치수를 반영하여 가장 큰 좌표를 가진 객체가 공간의 양의 사분면에 나타나도록 합니다.
차원별	각 차원에 대해 개별적으로 반사를 설정할 수 있습니다. 특정 축에 반사를 적용하려면 차원 옆의 확인란을 선택합니다.

근접 매핑: 출력

출력 하위 탭을 사용하면 PROXMAP 분석의 데이터와 결과를 요약한 표와 그래프를 요청할 수 있습니다. 여기에는 각 주요 분석 탭 (**데이터**, **초기**, **모델**, **제한**, **보충** 및 **기준**)에 대한 하위 탭이 포함되어 있습니다

분석의 모든 측면과 관련된 출력을 요청할 수 있습니다. 각 하위 탭에는 특정 테이블 또는 플롯을 요청하는 확인란이 있습니다.

데이터

데이터 탭에서는 다변량 데이터, 근사치, 가중치 등 데이터와 관련된 출력을 요청할 수 있습니다.

테이블(데이터)

옵션	설명
원래 다변량 데이터	근사값을 도출하는 데 사용되는 변수의 원시 값을 표시합니다.
원래 근접성	사용자가 직접 제공한 근거리(해당되는 경우)를 표시합니다.
파생된 근접성	모든 소스에 대한 다변량 데이터에서 계산된 근사값을 표시합니다.
결합된 근접성	여러 소스가 결합된 경우 집계된 근접성을 표시합니다.

테이블(가중치)

옵션	설명
원래 가중치	각 소스에 사용되는 원시(사용자 제공) 근접 가중치를 나열합니다.
파생된 가중치	원시 가중치, 객체 대소문자 가중치, 누락된 데이터 및 근사값에 적용된 가중치 함수에서 파생된 가중치 목록을 제공합니다.
개체(케이스) 가중치	개체별로 적용된 가중치를 표시합니다(지정된 경우).

그래프(데이터)
근접도 **히스토그램은** 모든 소스에 걸쳐 근접도 값의 분포를 표시합니다.

그래프(가중치)
가중치 **히스토그램은** 근접 가중치(가중치가 제공되거나 파생된 경우)의 분포를 표시합니다.

초기

초기 하위 탭에서는 초기 구성, 적합도 분해, 초기 추정 후 진단 측정과 관련된 출력을 요청할 수 있습니다.

테이블

옵션	설명
좌표	초기 구성에 있는 객체의 좌표를 표시합니다.
변수 방향	구성 공간에서 원래 변수의 벡터 방향을 표시합니다(다변량 데이터의 경우). 명목 변수에 대한 여러 좌표를 제공합니다(다변량 데이터에 있는 경우).
특성 방향	프로퍼티 변수의 방향을 표시합니다(프로퍼티가 정의된 경우에만 활성화됩니다). 명목 변수에 대한 여러 좌표를 제공합니다(다변량 데이터에 있는 경우).
스트레스 분해	전체 손실(스트레스) 값에 대한 각 개체의 기여도를 보고합니다.
진단	초기 솔루션 평가를 위한 적합성 통계 및 진단이 포함되어 있습니다. 여기에는 고전적인 스케일링 시작을 위한 고유값과 고유 벡터가 포함됩니다.

그래프

옵션	설명
구성	개체 공간의 그래픽 표현을 표시합니다.
다변량 Bi-플롯	다변량 데이터에 사용할 수 있는 조인트 플롯에서 객체 포인트와 가변 벡터 방향을 결합합니다.
특성 Bi-플롯	객체와 속성 변수 간의 관계를 시각화합니다(속성이 정의된 경우 사용 가능).
스크리 도표	기존 스케일링 솔루션의 고유값과 나머지 차원을 그래프로 표시합니다.
스트레스 플롯	객체 구성 공간에 겹쳐진 버블 플롯입니다. 각 개체는 원으로 둘러싸인 점으로 표시됩니다. 각 원의 크기는 구성의 전체 노멀라이즈드 스트레스에 대한 해당 오브젝트의 상대적 기여도를 반영합니다.

모델

모델 하위 탭을 사용하면 공통 및 개별 공간, 소스별 변환, 모델 적합도 측정과 관련된 그래픽 및 표 형식의 출력을 요청할 수 있습니다.

테이블

옵션	설명
좌표	최종 공통 공간에 있는 객체의 좌표를 나열합니다.
변수 방향	원래 변수의 방향을 나열합니다(다변량 데이터의 경우). 명목 변수에 대한 여러 좌표를 제공합니다(다변량 데이터에 있는 경우).
공간 가중값	개인차를 허용하는 모델에서 각 소스에 대해 추정된 공간 가중치를 표시합니다.

그래프(공용 공간)

옵션	설명
구성 (기본 선택됨)	개체의 일반적인 공간 구성을 표시합니다. **구성(Configuration**)은 다른 옵션을 선택하지 않는 한 출력에 포함되는 기본 설정입니다.
다변량 Bi-플롯	구성 및 길찾기 데이터가 모두 다변량임을 표시합니다.
감소된 순위 Bi-플롯	모델에 랭크가 1인 축소된 개별 공간이 포함된 경우 2차원 공통 공간에 소스별 투영을 표시합니다. 이 그래프는 공통 공간과 개별 공간을 하나의 그래프에 결합한 것입니다.
최소 범위 트리	구성의 모든 오브젝트를 가능한 가장 짧은 총 에지 길이로 연결합니다. MST는 구성의 기본 구조를 시각화하여 근접성 기반 그룹화 및 전환을 강조하는 데 도움이 됩니다.
이웃	이웃 그래프는 구성 공간의 거리에 따라 구성의 각 개체를 가장 가까운 이웃 k 개체에 연결합니다. 이 그래프는 로컬 구조를 강조하고 구성에서 클러스터나

옵션	설명
	이상값을 감지하는 데 도움이 됩니다. 가장 가까운 이웃의 수 값을 선택할 수 있습니다.
임계값	**구성 공간에서** 원본 또는 변환된 근사값이 지정된 임계값보다 작거나 같은 오브젝트 쌍을 연결합니다. 임계값을 선택할 수 있습니다.

그래프(개별 공간)

옵션	설명
구성	선택한 소스에 대한 개별 공간 구성을 플롯합니다(개인차 모델과 함께 사용).
공간 가중값	확장 및 가중 유클리드 모델의 치수 가중치를 표시합니다.

그래프(변환 및 맞춤 측정값)

옵션	설명
변환	선택한 소스에 대해 변환된 근거리와 원래 근거리를 비교하여 표시합니다.
잔차	변환된 근거리와 적합 거리의 차이를 변환된 근거리와 비교하여 플롯합니다.
적합	장착된 거리와 소스별 변환된 거리를 표시합니다.
셰퍼드	선택한 소스의 근거리와 적합 거리를 원래 근거리와 비교하는 셰퍼드 다이어그램을 만듭니다.

제한사항

제한 하위 탭은 속성 변수를 구성 제약 조건(감독 매핑)으로 사용하는 것과 관련된 출력을 제공합니다. 이러한 속성의 원본, 변환 및 맞춤 버전에 대한 테이블 및 시각화를 요청할 수 있습니다.

테이블 속성

옵션	설명
원래 속성	선택한 속성 변수의 원시 값을 표시합니다.
변환된 속성	최적의 변환이 적용된 후의 속성 값을 나열합니다.
속성 가중치	공용 공간의 제한을 위해 속성에 대해 계산된 가중치를 표시합니다.
속성 방향	조인트 공간의 각 속성과 연관된 벡터 좌표를 나열합니다.

그래프

다음 표에서 하나 이상의 속성을 선택하여 다음 그래프를 생성합니다:

옵션	설명
Bi-플롯	선택한 속성 변수에 대한 벡터로 개체 구성을 표시합니다.
변환	선택한 각 속성에 적용되는 변환 함수를 표시합니다.

옵션	설명
잔차	선택한 속성에 대해 변환된 맞춤 값과 변환된 값의 차이를 표시합니다.

투영 플롯은 공통 공간에 있는 객체의 1차원 투영을 각 속성에 개별적으로 표시합니다.

보조

보조 하위 탭은 구성 해석에 사용되지만 최적화(비지도 매핑)에는 영향을 주지 않는 속성 변수와 관련된 출력을 제공합니다.

테이블 속성

옵션	설명
원래 특성	속성 변수의 원시 값을 표시합니다.
변환된 특성	최적의 변환을 적용한 후의 값을 표시합니다.
특성 방향	구성 공간에 있는 프로퍼티의 방향 벡터를 나열합니다(해당되는 경우).

그래프 속성

목록에서 하나 이상의 속성을 선택하여 다음 그래프를 생성합니다:

옵션	설명
Bi-플롯	선택한 프로퍼티 변수에 대한 벡터로 오브젝트 구성을 표시합니다.
변환	선택한 각 프로퍼티의 변환을 표시합니다.
잔차	선택한 속성에 대해 변환된 값과 변환된 값의 차이를 플롯합니다.

투영 플롯은 공통 공간에 있는 객체의 1차원 투영을 각 속성에 개별적으로 표시합니다.

기준

기준 하위 탭에서는 반복 추정 프로세스 및 모델 수렴과 관련된 진단 정보 및 시각화를 제공합니다.

테이블

옵션	설명
요약 (기본 선택됨)	최종 스트레스 값, 반복 횟수, 경과 시간 및 모든 오류 메시지를 보고합니다. **요약은** 다른 옵션을 선택하지 않는 한 출력에 포함되는 기본 설정입니다.
진단	크러스칼의 스트레스 공식 1, 소외 계수, 분산 설명, 터커의 일치 계수, 피어슨 및 스피어만 상관관계를 포함한 추가 적합도 통계를 나열합니다.
반복 히스토리	각 반복에 대한 정규화된 응력, 반복 간 응력 차이, 반복에 걸친 구성 변경(좌표 차이)을 표시합니다. 테이블의 크기를 제한하기 위해 반복 기록 테이블은 2의 거듭제곱에 따른 반복의 하위 집합만 보고합니다. 모든 반복을 나열하지 않고 전체 수렴 패턴을 캡처합니다
스트레스 분해	전체 스트레스 값을 개별 오브젝트의 기여도로 세분화합니다.

그래프

옵션	설명
반복 히스토리	2의 거듭제곱에 따른 반복 값의 하위 집합을 플롯으로 표시합니다. 정규화된 스트레스 값과 반복 횟수입니다.
스트레스 플롯	정규화된 응력에 대한 기여도에 비례하는 객체 포인트 주위에 원을 표시하여 공통 공간에서 객체의 기여도를 보여주는 버블 플롯입니다.

근접 매핑: 저장

저장 탭을 사용하면 선택한 결과를 새 데이터 집합으로 내보낼 수 있습니다.

다음 항목을 저장하도록 선택할 수 있습니다:

초기좌표
초기 구성의 좌표(초기 공간)입니다.

최종좌표
최종 구성의 좌표(공용 공간)입니다.

모델 좌표
개별 구성(개별 공간)의 좌표입니다. 이 옵션은 소스 수가 하나보다 많은 경우에만 활성화됩니다.

파생된 근접성
다변량 데이터에서 파생된 근거리 또는 직접 제공된 경우 원본 근거리입니다.

결합된 근접성
모든 소스에 대해 집계된 근접성(여러 소스가 정의된 경우).

변환된 근접도
선택한 변환 함수를 적용한 후의 근거리입니다.

변환된 속성
변환된 속성입니다. 이 옵션은 **제한** 탭에서 속성을 지정한 경우에만 활성화됩니다.

적합한 속성
구성에 맞는 속성 값입니다. 이 옵션은 속성이 선택된 경우에만 활성화됩니다.

변환된 특성
보조 프로퍼티의 변환된 값입니다. 이 옵션은 **부가** 탭에 속성이 정의되어 있는 경우에만 활성화됩니다.

적합한 특성
구성에 투영된 프로퍼티 값입니다. 이 옵션은 **속성을** 선택한 경우에만 활성화됩니다.

데이터 세트 선택

선택한 항목이 새 데이터 집합에 기록됩니다. 줄임표 버튼(...)을 클릭합니다 를 클릭하여 다른 **이름으로 데이터 저장** 대화 상자를 열고 새 데이터 집합의 파일 이름과 대상 폴더를 지정할 수 있습니다.

편상관계수

편상관계수 프로시저는 한 개 이상의 추가 변수의 효과를 제어하는 동시에 두 변수간 선형 관계를 설명하는 편상관을 계산합니다. 상관은 선형 연관의 측도입니다. 두 변수가 완벽하게 연관될 수는 있지만 두 변수의 관계가 선형이 아니면 상관계수가 연관 측정에 적합한 통계가 될 수 없습니다.

예

의료보험료와 질병률 사이에 관계가 있습니까? 이 두 요소가 서로 음수 관계일 것으로 예상하지만 연구 자료에서는 상당한 양의 상관이 있다고 보고합니다. 의료 자금이 증가할수록 질병 비율이 증가하는 것으로 나타납니다. 그러나 의료 기관을 방문하는 비율을 제어하면 이와 같이 관측된 양의 상관이 사실상 없어집니다. 이는 보험료가 증가할 때 더 많은 사람들이 의료 기관을 찾으므로 결국 의사와 병원을 통해 보고되는 질병이 더 증가하여 의료 보험료와 질병률이 서로 양의 상관관계를 갖는 것으로 나타나기 때문입니다.

통계
각 변수에 대해서는 경우 평균, 표준 편차, 비결측값이 있는 케이스 수를 선택할 수 있습니다. 자유도와 유의수준이 있는 편상관행렬과 0차 상관행렬을 선택할 수 있습니다.

데이터 고려사항

데이터
대칭이고 양적변수를 사용합니다.

가정
편상관 프로시저는 각 대응변수가 정규 이변량인 것으로 가정합니다.

편상관계수 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 상관 > 부분...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 편상관을 계산할 숫자변수를 둘 이상 선택합니다.
3. 숫자 제어변수를 한 개 이상 선택합니다.

다음 옵션을 사용할 수도 있습니다.

유의성 검정
양측 확률 또는 단측 확률을 선택할 수 있습니다. 연관 방향을 이미 알고 있는 경우에는 **한쪽**을 선택합니다. 모를 때는 **양쪽**을 선택합니다.

관측 유의수준 표시
기본적으로 각 상관계수에 대한 확률과 자유도가 표시됩니다. 이 항목을 선택하지 않으면 유의수준이 0.05인 계수는 별 하나로 표시되고 수준이 0.01인 계수는 별 두 개로 표시되며 자유도는 표시되지 않습니다. 이 설정은 편상관행렬과 0차 상관행렬 모두에 영향을 줍니다.

편상관: 옵션

통계. 다음 중 하나 또는 모두를 선택할 수 있습니다.

- **평균과 표준 편차.** 각 변수에 대해 표시됩니다. 비결측값이 있는 케이스 수도 함께 표시됩니다.
- **0차 상관.** 제어변수를 비롯한 모든 변수 간에 단순 상관행렬을 표시합니다.

결측값. 다음 중 하나를 선택합니다.

- **목록별 결측값 제외.** 제어변수를 포함하여 변수에 대한 결측값이 있는 케이스는 모든 계산에서 제외됩니다.
- **대응별 결측값 제외.** 편상관의 기초가 되는 0차 상관을 계산하는 경우, 둘 다 또는 둘 중 한 대응변수에 대한 결측값이 있는 케이스는 사용되지 않습니다. 쌍별 삭제는 가능한 한 많은 데이터를 사용합니다. 그러나 케이스 수는 계수와는 다릅니다. 쌍별 삭제가 유효하면 특정 편상관에 대한 자유도는 0차 상관의 계산에서 사용한 최소 케이스 수에 따라 달라집니다.

PARTIAL CORR 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 0차 상관행렬을 읽거나 편상관행렬을 작성합니다(MATRIX 하위 명령 사용).
- 두 변수 목록 사이에서 편상관을 얻습니다(VARIABLES 하위 명령에서 키워드 WITH 사용).
- 여러 분석을 얻습니다(여러 VARIABLES 하위 명령 사용).
- 두 제어변수가 있을 때 요청에 대한 순서 값을 지정합니다(예: 1차 및 2차 순서 편상관)(VARIABLES 하위 명령 사용).
- 중복 계수를 출력하지 않습니다(FORMAT 하위 명령 사용).
- 일부 계수를 계산할 수 없을 때 단순 상관행렬을 표시합니다(STATISTICS 하위 명령 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

거리

이 프로시저는 대응 케이스나 대응변수 간에 유사성이나 상이성(거리)을 측정하는 여러 통계를 계산합니다. 이러한 유사성이나 거리 측도는 요인 분석, 군집 분석, 다차원척도법 등과 마찬가지로 복잡한 데이터 세트를 분석하는 데 사용할 수 있습니다.

예. 엔진 크기, MPG, 마력 등의 특정한 특성에 따라 두 자동차 간의 유사성을 측정할 수 있습니까? 자동차 간의 유사성을 계산하여 서로 비슷한 점과 다른 점을 검토할 수 있습니다. 더 형식적인 분석에 대해서는 계층적 군집 분석이나 다차원척도법을 유사성에 적용하여 기본 구조를 탐색할 수 있습니다.

통계. 구간 데이터에 대한 상이성(거리) 측도는 유클리드 거리, 유클리드 거리 제곱값, Chebychev, 블럭, Minkowski, 사용자 정의 등이며 개수 데이터에 대해서는 카이제곱이나 파이-제곱이 있습니다. 또한 이분형 데이터에 대한 상이성 측도로는 유클리드 거리, 유클리드 거리 제곱값, 크기 차이, 패턴 차이, 분산, 형태, Lance와 Williams 등이 있습니다. 이에 반해 구간 데이터에 대한 유사성 측도로는 Pearson 상관계수나 코사인이 있습니다. 이분형 데이터에 대해서는 Russel과 Rao, 단순 일치, Jaccard, dice, Rogers와 Tanimoto, Sokal과 Sneath 1, Sokal과 Sneath 2, Sokal과 Sneath 3, Kulczynski 1, Kulczynski 2, Sokal과 Sneath 4, Hamann, 람다, Anderberg의 *D*, Yule의 *Y*, Yule의 *Q*, Ochiai, Sokal과 Sneath 5, 파이 4-점 상관, 산포 등이 있습니다.

거리 행렬 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **상관분석** > **거리...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 숫자변수를 하나 이상 선택하여 케이스 간 거리를 계산하거나 둘 이상의 숫자변수를 선택하여 변수 간 거리를 계산합니다.
3. 거리 계산에서 옵션을 선택하여 케이스나 변수 간 근접성을 계산합니다.

거리 측도: 상이성 측도

측도에서 사용하고 있는 데이터 유형(구간, 개수, 이분형 등)에 해당하는 사항을 선택한 다음 드롭다운 목록에서 데이터 유형에 해당하는 측도 중 하나를 선택합니다. 데이터 유형에 따라 다음 측도를 사용할 수 있습니다.

- **구간 데이터.** 유클리드 거리, 유클리드 거리 제곱값, Chebychev, 블록, Minkowski 또는 사용자 정의 간격을 선택할 수 있습니다.
- **개수 데이터.** 카이제곱 측도 또는 파이제곱 측도를 선택할 수 있습니다.
- **이분형 데이터.** 유클리드 거리, 유클리드 거리 제곱값, 크기 차이, 패턴 차이, 분산, 형태, Lance와 Williams를 선택할 수 있습니다. 유와 무에 값을 입력하여 의미 있는 두 개의 값을 지정하면 거리 측도는 다른 값을 모두 무시합니다.

값 변환으로 근접성을 계산하기 전에 케이스나 변수에 대한 값을 표준화할 수 있습니다. 이러한 변환은 이분형 데이터에는 적용할 수 없습니다. 표준화 방법으로는 *z* 스코어, −1부터 1까지 범위, 0부터 1까지 범위, 최대값 1, 평균 1, 표준 편차 1 등을 사용할 수 있습니다.

측도변환으로 거리 측도를 통해 생성한 값을 변환할 수 있습니다. 이 값은 거리 측도를 계산한 다음 적용됩니다. 옵션으로는 절대값, 부호 바꾸기, 0–1 범위로 척도 조정 등을 사용할 수 있습니다.

거리 측도: 유사성 측도

측도에서 사용 중인 데이터 유형(구간이나 이분형)에 맞는 사항을 선택한 다음 드롭다운 목록에서 데이터 유형에 해당하는 측도를 선택합니다. 데이터 유형에 따라 다음 측도를 사용할 수 있습니다.

- **구간 데이터.** Pearson 상관계수 또는 코사인을 나타냅니다.
- **이분형 데이터.** Russell과 Rao, 단순 일치, Jaccard, Dice, Rogers와 Tanimoto, Sokal과 Sneath 1, Sokal과 Sneath 2, Sokal과 Sneath 3, Kulczynski 1, Kulczynski 2, Sokal과 Sneath 4, Hamann, 람다, Anderberg의 *D*, Yule의 *Y*, Yule의 *Q*, Ochiai, Sokal과 Sneath 5, 파이 4-점 상관, 산포 등을 나타냅니다. 유와 무에 값을 입력하여 의미 있는 두 개의 값을 지정하면 거리 측도는 다른 값을 모두 무시합니다.

값 변환으로 근접성을 계산하기 전에 케이스나 변수에 대한 데이터 값을 표준화할 수 있습니다. 이러한 변환은 이분형 데이터에는 적용할 수 없습니다. 표준화 방법으로는 *z* 스코어, -1부터 1까지 범위, 0부터 1까지 범위, 최대값 1, 평균 1, 표준 편차 1 등을 사용할 수 있습니다.

측도변환으로 거리 측도를 통해 생성한 값을 변환할 수 있습니다. 이 값은 거리 측도를 계산한 다음 적용됩니다. 옵션으로는 절대값, 부호 바꾸기, 0–1 범위로 척도 조정 등을 사용할 수 있습니다.

PROXIMITIES 명령 추가 기능

거리 측도 절차에서는 PROXIMITIES 명령 구문을 사용합니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- Minkowski 거리 측도에 대한 거듭제곱으로 정수를 지정합니다.
- 사용자 정의된 거리 측도에 대한 거듭제곱 및 근으로 정수를 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

거리 상관

이 기능을 사용하려면 Statistics Base 옵션이 필요합니다.

거리 상관관계 절차는 두 임의 변수 또는 다변량 데이터 간의 의존성을 계산합니다. 기존의 상관관계 방법과 달리 변수의 비선형 관계를 평가합니다.

예

소매업체는 고객의 인구통계학적 특성(연령, 소득, 교육, 결혼 여부)이 구매 행동에 영향을 미치는지 파악하고자 합니다. 관계가 비선형적이거나 여러 요인 간의 상호작용이 있는 경우 기존의 상관관계 분석 방법은 실패할 수 있습니다.

거리 상관 관계 절차는 다음과 같은 이점이 있는 시나리오에 적용할 수 있습니다:

- 비선형 종속성을 캡처합니다. 피어슨 상관관계와 달리 거리 상관관계는 선형이 아니더라도 관계를 감지합니다.
- 다차원 관계를 처리합니다. 인구통계학적 요소의 조합이 구매 패턴과 어떤 관련이 있는지 평가할 수 있습니다.
- 포괄적인 조치를 제공합니다. 유의미한 거리 상관관계는 고객 인구통계가 구매 행동에 큰 영향을 미치며 타겟 마케팅 전략의 지침이 된다는 것을 시사합니다.

관련 통계

- 거리 상관관계 (*dCor(X,Y))*: 고객 인구 통계 *(X)* 와 구매 행동 *(Y)* 간의 의존성을 측정합니다.
- 순열 기반 p-값: 종속성의 통계적 유의성을 평가합니다. P값이 작을수록(예: p<0.05)은 강한 관계를 나타냅니다.
- 거리 공분산 (*dCov(X,Y))*: 변수 간 의존성의 강도를 정량화합니다.
- 거리 분산(*dVar(X), dVar(Y)*): 상관관계를 정규화하기 위해 개별 변수 내의 변동성을 측정합니다.

데이터 고려사항

분석에는 연속형 변수만 사용할 수 있습니다. 거리 상관관계를 계산하려면 최소 2개의 연속 변수가 필요합니다. 최대 30개의 변수가 허용됩니다.

거리 상관관계 구하기

1. 메뉴에서 **상관관계** > **분석,** > **거리 상관관계를** 클릭합니다.

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 거리 상관관계를 계산할 두 개 이상의 숫자 변수를 선택합니다. 소스 데이터 집합에서 최대 30개의 연속형 변수를 선택하여 분석할 수 있습니다.
3. 식별자 변수를 하나 선택합니다. **ID 변수를** 선택하면 결과에 ID가 포함된 거리 매트릭스 테이블이 표시됩니다.

4. **확인을** 클릭하여 지정된 설정으로 프로시저를 실행하고 출력을 생성합니다.

거리 상관: 기준

기준 하위 대화상자에는 정규화를 위한 제어 매개변수와 거리 상관관계 절차를 사용자 지정하기 위한 기타 주요 옵션이 있습니다.

정규화

데이터 집합에 하나의 정규화 유형을 할당하는 옵션을 선택합니다. 정규화는 변수 척도를 표준화하지만, 거리 상관관계를 정확하게 평가하는 데 중요한 데이터의 매우 비선형적인 관계를 의도치 않게 왜곡할 수 있습니다. 따라서 선형적인 관계를 용이하게 하기 위해 이 기능이 포함되어 있습니다.

- **최소-최대** 정규화는 각 피처의 크기를 고정 범위(일반적으로 [0, 1])로 조정합니다. 이 변환은 데이터 집합에 단위나 범위가 다른 변수가 포함된 경우에 유용합니다. 이 정규화를 통해 변수 간 공정한 비교가 가능하며, 단일 변수가 규모에 따라 거리 계산을 지배하지 않도록 합니다. 큰 숫자 범위의 영향을 완화하는 데 도움이 됩니다.
- **Z점수** 정규화는 표준 편차가 1인 평균을 중심으로 데이터를 정렬합니다. 이 접근 방식은 일반적으로 분포된 데이터 집합에 유용합니다. 이 정규화는 각 변수에 표준화된 척도를 부여하므로 이상값의 영향을 줄여줍니다. 모든 변수가 분포에 따라 동일하게 취급되므로 특히 서로 다른 변수를 비교할 때 결과를 더 쉽게 해석할 수 있습니다.
- **강력한 스케일링은** 중앙값과 사분위수 범위(IQR)를 사용하여 데이터를 스케일링하므로 이상값에 탄력적으로 대응할 수 있습니다. 이상값의 영향이 최소화되어 극단적인 값이 존재할 때 보다 신뢰할 수 있는 거리 상관관계 계산이 가능합니다. 이 기술은 데이터 집합에 결과를 왜곡할 수 있는 이상값이 포함되어 있을 때 유용합니다.
- **로그 변환은** 데이터의 왜곡, 특히 양으로 치우친 분포의 왜곡을 줄이도록 설계되었습니다. 이는 데이터 세트 전반의 분산 안정화에 도움이 될 수 있습니다. 로그 변환은 큰 값을 압축하고 작은 값을 확장하여 분포를 정규화하는 데 도움이 됩니다. 이는 특히 왜곡이 심한 데이터 집합을 다룰 때 거리 상관관계 분석의 정확도를 향상시킬 수 있습니다.

신뢰 구간 백분율

통계 분석에 사용할 신뢰 구간 비율을 지정합니다.

신뢰 구간 백분율은 통계적 추정치와 관련된 확실성의 수준을 나타냅니다. 백분율이 높을수록 간격의 정확도에 대한 신뢰도가 높지만 간격이 넓어지고, 백분율이 낮을수록 간격이 좁아지지만 확실성이 떨어집니다

거리 상관관계의 순열

거리 상관관계에서 순열 테스트는 두 변수 간에 관찰된 거리 상관관계의 통계적 유의성을 평가하는 데 사용됩니다. 목표는 관찰된 상관관계가 우연히 발생할 수 있는 상관관계와 크게 다른지 평가하는 것입니다. 데이터 세트 중 하나(또는 둘 다)가 무작위로 순열되어 변수 간의 기존 연관성을 깨뜨립니다. 그런 다음 각 순열에 대해 거리 상관관계를 다시 계산하여 연관성이 없다는 가설 하에 거리 상관관계 값의 분포를 생성합니다.

사용자 지정 시드 설정을 선택하여 순열 테스트를 활성화하고 반복 횟수 및 유의 수준을 지정합니다. **시드에서** 값을 설정합니다. 순열 테스트에 사용되는 난수 생성기의 시드를 선택적으로 설정할 수 있습니다. 시드는 양의 정수 또는 0이어야 합니다. 기본값은 2000000입니다.

최대 반복 횟수를 설정합니다. 이 값은 양의 정수이어야 합니다. 기본값은 100으로 설정되어 있으며 음수 값은 허용되지 않습니다.

관측된 거리 상관관계를 허용된 데이터 세트의 거리 상관관계 분포와 비교하여 p-값을 계산합니다. P값은 관측된 거리 상관관계보다 크거나 같은 허용된 거리 상관관계의 비율을 나타냅니다. 유의 수준은 통계적 유의성을 결정하기 위한 임계값으로 사용됩니다. P값이 유의 수준보다 낮으면 관찰된 상관관계가 통계적으로 유의미한 것으로 간주됩니다. **중요도 레벨의** 값을 0에서 1 사이로 설정합니다. 기본값은 0.05입니다.

거리 상관: 인쇄

상세 정보 인쇄

세부 정보 인쇄를 선택하여 다음과 같은 필수 측정값이 포함된 표를 표시합니다: 관찰 수(N), 평균, 표준 편차(Std). 편차), 각 변수의 최소값과 최대값을 확인할 수 있습니다. 이러한 통계는 데이터 집합에 대한 포괄적인 개요를 제공하며 비교되는 변수의 분포와 특성을 더 잘 이해할 수 있도록 도와줍니다.

거리 상관계수

거리 보정 계수를 선택하여 분석을 위해 계산된 거리 상관 계수를 보여주는 표를 생성합니다. dCor Value는 변수 간의 거리 상관 계수를 보여줍니다. **기준** 하위 대화 상자에서 **사용자 지정 시드 설정을** 선택하고 유의 수준을 설정하면 계산된 상관 계수와 관련된 유의 수준(p-값)을 나타내는 Sig. 열을 볼 수 있습니다.

Number of Replicates (R) 는 분석에 사용된 복제본(또는 선택된 경우 순열)의 수를 나타냅니다.

거리 공분산 추정

사용자가 신뢰 구간(CI) 수준을 선택하면 표에 신뢰 구간의 상한과 하한이 포함되어 추정치에 대한 보다 포괄적인 보기를 제공합니다.

거리 메트릭과 유의수준 추정

이 추정치는 분석되는 변수 간의 관계와 종속성을 평가하는 데 도움이 됩니다.

거리 행렬

거리 행렬을 선택하여 단일 차원 데이터의 거리 행렬을 표시하는 테이블을 생성합니다(예. 변수 "나이"). 각 관측 쌍 사이의 거리가 표시됩니다. 대각선 요소는 0 거리(자체까지의 거리)를 표시합니다. 이 매트릭스는 변수 간의 관계를 종합적으로 분석할 수 있도록 쌍별 거리를 명확하게 보여줍니다. 거리 행렬을 구성하는 옵션을 선택합니다. 거리 행렬 **구성** 대화 상자의 목록에서 거리 행렬을 구성할 변수를 선택합니다.

거리 매트릭스를 파일로 저장

거리 행렬을 파일에 저장하는 옵션을 선택합니다. **아웃파일 파일 디렉토리에** 파일 이름을 입력합니다. 지원되는 파일 확장자는 .sav 입니다.

거리 상관관계: 플롯

플롯을 선택하여 거리 상관관계에 대한 플롯을 생성합니다. **분산형** 차트를 선택하면 선택한 변수 간의 쌍별 거리 또는 상관관계를 보여주는 분산형 차트를 생성할 수 있습니다.

목록에서 X축에 플롯할 변수를 선택합니다. 이 변수는 분산형 차트에 선택된 다른 변수와 상호 배타적이어야 합니다.

목록에서 다른 변수를 선택하여 Y축에 플롯합니다. 이 변수는 스캐터플롯에 선택된 다른 변수와 상호 배타적이어야 합니다.

선형모형

선형모형은 대상과 하나 이상의 예측변수 간의 선형 관계를 기준으로 연속형 대상을 예측합니다.

선형모형은 비교적 단순하며 스코어링에 대해 쉽게 해석되는 수학 공식을 제공합니다. 이러한 모형의 속성은 같은 데이터 세트의 다른 모형 유형(예: 신경망 또는 의사결정 나무)과 비교하여 잘 이해되며 일반적으로 아주 빨리 작성될 수 있습니다.

예. 자택 소유자의 보험 청구에 대해 조사하기 위한 리소스가 제한된 보험 회사가 보험 청구액을 추정하기 위한 모형을 만들고자 합니다. 이 모델을 서비스 센터에 배포하면 영업 담당자는 고객과 통화하면서 청구 정보를 입력하여 지난 데이터를 토대로 '예상' 청구 비용을 즉시 산출할 수 있습니다.

필드 요구사항. 대상과 하나 이상의 입력이 있어야 합니다. 기본적으로 모두 또는 없음의 사전 정의된 역할이 있는 필드는 사용되지 않습니다.

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

대상이 연속형(척도)이어야 합니다. 예측자(입력)에 측정 수준 제한 사항이 없습니다. 범주형(명목 및 순서) 필드는 모형에서 요인으로 사용되며 연속형 필드는 공변량으로 사용됩니다.

참고: 범주형 필드에 범주가 1,000개보다 많을 경우 프로시저가 실행하지 않고 모형이 작성되지 않습니다.

선형 모형 구하기

이 기능을 사용하려면 Statistics Base 옵션이 필요합니다.

메뉴에서 다음을 선택합니다.

분석 > 회귀분석 > 자동 선형 모형...

1. 대상과 입력이 하나 이상씩 있는지 확인합니다.
2. **작성 옵션**을 클릭하여 선택적 작성 및 모형 설정을 지정합니다.
3. **모형 옵션**을 클릭하여 활성 데이터 세트에 스코어를 저장하고 외부 파일로 모형을 내보냅니다.
4. **실행**을 클릭하여 프로시저를 실행하고 모형 오브젝트를 만듭니다.

목표

원하는 주요 목표 해당하는 목표를 선택합니다.

- **표준 모형 만들기.** 이 방법은 예측자를 사용하여 대상을 예측하는 단일 모형을 설정합니다. 일반적으로 표준 모형은 부스팅되었거나 배깅되었거나 큰 데이터 세트 앙상블보다 해석하기 쉽고 점수화이 빠릅니다.
- **모형 정확도 향상(부스팅).** 이 방법은 더 정확한 예측을 하기 위해 모형의 시퀀스를 생성하는 부스팅을 사용하여 앙상블 모형을 작성합니다. 앙상블은 표준 모형보다 작성 및 스코어링하는 데 오래 걸릴 수 있습니다.

  부스팅은 "구성요소 모형"의 성공을 만들어 주며 각 구성요소 모형은 전체 데이터 세트에 빌드됩니다. 성공한 각 구성요소 모형을 빌드하기 전에 이전 구성요소 모형의 잔차를 기반으로 레코드가 가중치가 부여됩니다. 잔차가 큰 케이스가 상대적으로 더 높은 분석 가중치이면 다음 구성요소 모형은 이 레코드를 예상하는 것에 초점을 둡니다. 이러한 구성요소 모형은 모두 앙상블 모형을 형성합니다. 앙상블 모형은 통합 규칙을 사용하여 새 레코드를 점수화합니다. 사용 가능한 규칙은 대상의 측정 수준에 따라 다릅니다.
- **모형 안정성 향상(배깅).** 이 방법은 더 신뢰할 만한 예측을 하기 위해 여러 모형을 생성하는 배깅(부트스트랩 통합)을 사용하여 앙상블 모형을 작성합니다. 앙상블은 표준 모형보다 작성 및 스코어링하는 데 오래 걸릴 수 있습니다.

  부트스트랩 통합에서는 원래 데이터 세트에서 교체를 통해 샘플링을 함으로써 트레이닝 데이터 세트의 복제를 만듭니다. 이는 원래 데이터 세트 크기와 동일한 크기의 부트스트랩 샘플을 작성합니다. 그런 다음 "구성요소 모형"이 각 복제에 빌드됩니다. 이러한 구성요소 모형은 모두 앙상블 모형을 형성합니다. 앙상블 모형은 통합 규칙을 사용하여 새 레코드를 점수화합니다. 사용 가능한 규칙은 대상의 측정 수준에 따라 다릅니다.
- **매우 큰 데이터 세트의 모형을 작성하십시오(IBM SPSS 통계 서버 필요).** 이 방법은 데이터 세트를 별도의 데이터 블록으로 분할하여 앙상블 모형을 작성합니다. 데이터 세트가 너무 커서 위의 모형을 만들 수 없거나 증분 모형설정의 경우 이 옵션을 선택하십시오. 이 옵션은 표준 모형보다 작성하는 데 시간이 덜 걸릴 수 있지만 점수화하는 데는 더 오래 걸릴 수 있습니다. 이 옵션에는 IBM SPSS 통계 Server 연결이 필요합니다.

부스팅, 배깅 및 매우 큰 데이터 세트부스팅 및 배깅과 관련된 설정은 139 페이지의 『앙상블』를 참조하십시오.

기본 사항

자동으로 데이터 준비. 이 옵션은 모형의 예측력을 최대화하기 위해 프로시저에서 대상 및 예측자를 내부적으로 변환할 수 있습니다. 모든 변형은 모형과 함께 저장되며 점수화를 위해 새 데이터에 적용됩니다. 변환된 필드의 원래 버전은 모형에서 제외됩니다. 기본적으로 다음 자동 데이터 준비가 수행됩니다.

- **날짜 및 시간 처리.** 각 날짜 예측자가 참조 날짜(1970-01-01) 이후의 경과 시간이 포함된 새 연속형 예측자로 변환됩니다. 각 시간 예측자가 참조 시간(00:00:00) 이후의 경과 시간이 포함된 새 연속형 예측자로 변환됩니다.
- **측정 수준 조정.** 고유 값이 5개 미만인 연속형 예측자가 순서 예측자로 다시 캐스팅됩니다. 고유 값이 10개보다 많은 순서 예측자가 연속형 예측자로 다시 캐스팅됩니다.
- **이상값 처리.** 절사 값을 넘는 연속형 예측자 값(평균에서 3배 표준 편차)이 절사 값으로 설정됩니다.
- **결측값 처리.** 명목 예측자의 결측값이 훈련 파티션의 최빈값으로 대체됩니다. 순서 예측자의 결측값이 훈련 파티션의 중위수로 대체됩니다. 연속형 예측자의 결측값이 훈련 파티션의 평균으로 대체됩니다.
- **지도되는 합치기.** 대상과 연관하여 처리되는 필드 수를 줄여 더욱 경제적인 모형을 만들 수 있습니다. 입력과 대상 간의 관계를 기반으로 유사한 범주가 식별됩니다. 유의적으로 다르지 않은 범주(즉 0.1보다 큰 P-값을 가지는 범주)가 합쳐집니다. 모든 범주가 하나로 합쳐지는 경우, 필드의 원래 버전과 파생된 버전이 예측자로서 값이 없기 때문에 모형에서 제외됩니다.

신뢰수준. 계수 보기에서 모형 계수의 구간 추정값을 계산하는 데 사용되는 신뢰수준입니다. 0보다 크고 100보다 작은 값을 지정하십시오. 기본값은 95입니다.

모형 선택

모형 선택 방법. 모형 선택 방법(자세한 내용은 아래 참조) 중 하나를 선택하거나 주효과 모형 항으로 사용 가능한 예측자를 모두 입력하는 **모든 예측자 포함**을 선택하십시오. 기본적으로 **단계적 전진**이 사용됩니다.

단계적 전진 선택. 모형에 아무 효과 없이 시작하여 단계별 기준에 따라 더는 추가하거나 제거할 수 없을 때까지 한번에 한 단계에서 효과를 추가 및 제거합니다.

- **입력/제거 기준.** 모형에 효과를 추가해야 할지 제거해야 할지 결정하는 데 사용되는 통계입니다. **정보 기준(AICC)**은 모형에 제공된 훈련 세트의 우도를 기준으로 하며 너무 복잡한 모형에 페널티를 부여하도록 조정됩니다. **F 통계**는 모형 오차에서 향상도의 통계 검정을 기준으로 합니다. **수정된 R-제곱**은 훈련 세트의 적합도를 기준으로 하며 너무 복잡한 모형에 페널티를 부여하도록 조정됩니다. **과적합 방지 기준(ASE)**은 과적합 방지 세트의 적합도(평균 제곱 오차, ASE)를 기준으로 합니다. 과적합 방지 세트는 모형을 훈련하는 데 사용되지 않는 원래 데이터 세트의 약 30%의 무작위 부표본입니다.

  F 통계 외의 다른 기준을 선택한 경우, 각 단계의 기준에서 최대 양의 증가에 해당하는 효과가 모형에 추가됩니다. 기준에서 감소에 해당하는 모형의 효과는 제거됩니다.

  F 통계를 기준으로 선택한 경우, 각 단계에서 지정한 임계값보다 작은 최소 *p*-값이 있는 효과가 모형에 추가됩니다.**다음보다 작은 P-값이 있는 효과 포함** 기본값은 0.05입니다. 지정한 임계값보다 큰 *p*-값이 있는 모형의 효과는 제거됩니다.**다음보다 큰 P-값이 있는 효과 제거** 기본값은 0.10입니다.
- **최종 모형에서 최대 효과 수를 사용자 정의하십시오.** 기본적으로 사용 가능한 모든 효과를 모형에 입력할 수 있습니다. 또는 단계적 알고리즘이 지정된 최대 효과 수가 있는 단계로 끝나는 경우, 알고리즘이 현재 효과 세트로 중지됩니다.
- **최대 단계 수를 사용자 정의하십시오.** 단계적 알고리즘이 특정 단계 수 이후 중지됩니다. 기본적으로 사용 가능한 효과 수는 3회입니다. 또는 양의 정수로 최대 단계 수를 지정하십시오.

최적 서브세트 선택. "가능한 모든" 모형을 확인하거나 최소한 단계적 전진보다 큰 가능한 모형 서브세트를 확인하여 최적 서브세트 기준에 따라 최적 서브세트를 선택합니다. **정보 기준(AICC)**은 모형에 제공된 훈련 세트의 가능성을 기반으로 하며 지나치게 복잡한 모형에 페널티를 주기 위해 조정됩니다. **수정된 R-제곱**은 훈련 세트의 적합도를 기준으로 하며 너무 복잡한 모형에 페널티를 부여하도록 조정됩니다. **과적합 방지 기준(ASE)**은 과적합 방지 세트의 적합도(평균 제곱 오차, ASE)를 기준으로 합니다. 과적합 방지 세트는 모형을 훈련하는 데 사용되지 않는 원래 데이터 세트의 약 30%의 무작위 부표본입니다.

기준의 최대값이 있는 모형이 최적 모형으로 선택됩니다.

참고: 최적 서브세트 선택이 단계적 전진 선택보다 계산이 더 집중됩니다. 최적 서브세트가 부스팅, 배깅 또는 아주 큰 데이터 세트와 함께 수행되는 경우, 단계적 전진 선택을 사용하여 작성되는 표준 모형보다 작성하는 데 상당히 많은 시간이 걸릴 수 있습니다.

앙상블

이 설정은 부스팅, 배깅 또는 아주 큰 데이터 세트가 목표에서 요청될 때 발생하는 앙상블 동작을 결정합니다. 선택한 목표에 해당하지 않는 옵션은 무시됩니다.

배깅 및 아주 큰 데이터 세트. 앙상블을 점수화할 때 앙상블 스코어값을 계산하기 위해 기본 모형에서 예측값을 조합하는 데 사용되는 규칙입니다.

- **연속형 대상의 기본 조합 규칙.** 연속형 대상에 대한 앙상블 예측값은 기본 모형의 예측값 평균 또는 중위수를 사용하여 조합될 수 있습니다.

모형 정확도를 향상시키는 것이 목표인 경우 조합 규칙 선택이 무시됩니다. 부스팅은 항상 가중 다수 투표를 사용하여 범주형 대상을 점수화하고 가중 중위수를 사용하여 연속형 대상을 점수화합니다.

부스팅 및 배깅. 모형 정확도 또는 안정성을 향상시키는 것이 목표일 때 작성할 기본 모형 수를 지정하십시오. 배깅의 경우, 부트스트랩 표본의 수입니다. 양의 정수이어야 합니다.

고급

결과 복제. 난수 시드를 설정하면 분석을 복제할 수 있습니다. 과적합 방지 세트에 있는 레코드를 선택하는 데 난수 생성기가 사용됩니다. 정수를 지정하거나 1과 2147483647 사이(1과 2147483647 포함)의 유사 난수 정수를 만드는 **생성**을 클릭하십시오. 기본값은 54752075입니다.

모형 옵션

데이터 세트에 예측값 저장. 기본 변수 이름은 *PredictedValue*입니다.

모형 내보내기. 이렇게 하면 모델이 외부 `.zip` 파일에 기록됩니다. 스코어링 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다. 고유하고 유효한 파일 이름을 지정하십시오. 기존 파일을 지정하는 경우 파일을 덮어씁니다.

모형 요약

모형 요약 보기는 모형과 그 적합성을 한 눈에 파악할 수 있도록 요약한 스냅샷입니다.

테이블. 다음을 포함한 최고 수준 모형 설정을 열거합니다.

- 필드 탭에 지정된 대상의 이름,
- 자동 데이터 준비가 기본 설정에서 지정한 대로 수행되었는지 여부,
- 모형 선택 설정에 지정된 모형 선택 방법 및 선택 기준. 최종 모형에 대한 선택 기준 값도 더 작고 개선된 형식으로 표현됩니다.

도표. 도표는 최종 모형의 정확도를 표시하며 더 크게 표시된 것이 더 나은 형식입니다. 최종 모형의 경우 값은 100 × 수정된 R^2입니다.

자동 데이터 준비

이 보기에는 제외되는 필드 및 변환된 필드가 어떻게 자동 데이터 준비(ADP) 단계에서 유도되었는지에 대한 정보가 표시됩니다. 변환되었거나 제외된 각 필드에 대해 테이블에 필드 이름, 분석에서 역할, ADP 단계에서 실행한 작업이 나열됩니다. 필드는 필드 이름의 알파벳 오름차순으로 정렬됩니다. 각 필드에서 할 수 있는 작업은 다음과 같습니다.

- **기간 파생: 월** 날짜를 포함한 필드의 값에서 현재 시스템 날짜까지 경과된 시간을 월 단위로 계산합니다.
- **기간 파생: 시간** 시간을 포함한 필드의 값에서 현재 시스템 시간까지 경과된 시간을 시간 단위로 계산합니다.
- **측정 수준을 연속형에서 순서로 변경** 고유 값이 5개 미만인 연속형 필드를 순서 필드로 다시 캐스팅합니다.
- **측정 수준을 순서에서 연속형으로 변경** 고유 값이 10개 이상인 순서 필드를 연속형 필드로 다시 캐스팅합니다.
- **이상값 자름** 절사 값을 넘는 연속형 예측자 값(평균에서 3배 표준 편차)을 절사 값으로 설정합니다.
- **결측값 바꾸기** 명목 필드의 결측값을 최빈값으로, 순서 필드를 중위수로, 연속형 필드를 평균으로 바꿉니다.
- **범주를 합쳐 대상과의 연관 최대화** 입력과 대상 간의 관계를 기반으로 '유사한' 예측자 범주를 식별합니다. 유의적으로 다르지 않은 범주(즉 0.05보다 큰 *p*-값을 가지는 범주)가 합쳐집니다.
- **상수 예측자 제외 / 이상값 처리 후 / 범주 합치기 후** 다른 ADP 작업을 마친 후 단일 값을 가진 예측자를 제거합니다.

예측자 중요도

일반적으로 가장 중요한 예측자 필드에 모형화 노력을 집중하고 가장 쓸모 없는 예측자를 삭제하거나 무시하기를 원합니다. 예측자 중요도 도표를 사용하면 모형 추정 시 각 예측자의 상대적인 중요도를 표시하여 원하는 작업을 수행할 수 있습니다. 값이 상대적이므로 표시된 모든 예측자에 대한 값의 합은 1.0이 됩니다. 예측자 중요도는 모형 정확도와는 관련이 없습니다. 예측 시 각 예측자의 중요도와 관련이 있으며 예측이 정확한지, 정확하지 않은지 여부와는 관련이 없습니다.

관찰에 의한 예측

수직축에 예측값의 구간화된 산점도를 표시하고 수평축에 관측값을 표시합니다. 이상적으로 점이 45도 줄에 있어야 합니다. 이 보기는 레코드가 모형에 의해 특히 잘못 예측되었는지 여부를 알려 줄 수 있습니다.

잔차

모형 잔차의 진단 도표를 표시합니다.

도표 유형. 여러 가지 표시 유형이 있으며 **유형** 드롭다운 목록에서 액세스할 수 있습니다.

- **히스토그램.** 정규 분포의 오버레이가 있는 스튜던트화 잔차의 구간화된 히스토그램입니다. 선형모형은 잔차에 정규 분포가 있다고 가정하므로 히스토그램은 원칙적으로 평활선과 비슷해야 합니다.
- **P-p 도표.** 스튜던트화 잔차를 정규 분포와 비교하는 구간화된 확률대확률 도표입니다. 도표화된 점의 기울기가 정규선보다 덜 가파른 경우 잔차는 정규 분포보다 큰 변동을 표시하며, 기울기가 더 가파른 경우 잔차는 정규 분포보다 적은 변동을 표시합니다. 도표화된 점에 S 형태 곡선이 있으면 잔차 분포가 비대칭됩니다.

이상값

이 테이블에는 모형에 대해 불필요한 영향력을 발휘하는 레코드가 나열되고 레코드 ID(필드 탭에 지정된 경우), 대상값 및 Cook의 거리가 표시됩니다. Cook의 거리는 특정 레코드를 모형 계수 계산에서 제외할 때 모든 레코드의 잔차가 얼마나 변경될 수 있는지에 대한 측도입니다. 큰 Cook의 거리는 레코드를 제외하면 계수가 상당히 변경됨을 나타내므로 영향력이 큰 것으로 고려되어야 합니다.

영향력이 큰 레코드는 주의 깊게 관찰하여 모형 추정에서 덜 중요하게 고려할 수 있을지, 허용 가능한 임계값에 대해 이상값을 자를지, 또는 영향력이 큰 레코드를 완전히 제거할지 결정해야 합니다.

효과

이 보기는 모형에서 각 효과의 크기를 표시합니다.

유형. 여러 가지 표시 유형이 있으며 **유형** 드롭다운 목록에서 액세스할 수 있습니다.

- **다이어그램.** 예측자 중요도를 줄여 효과가 위에서 아래로 정렬되는 도표입니다. 다이어그램의 연결선은 효과 유의수준을 기준으로 가중되며 선이 굵을수록 더 유의한 효과(더 작은*p*-값)입니다. 마우스 커서를 연결선 위에 놓으면 *p*-값 및 효과의 중요도를 알려주는 도구 팁이 표시됩니다. 이는 기본값입니다.
- **표.** 전체 모형과 개별 모형 효과에 대한 ANOVA 테이블입니다. 예측자 중요도를 줄여 개별 효과가 위에서 아래로 정렬됩니다. 기본적으로는 테이블이 축소되어 전체 모형의 결과만 보여줍니다. 개별 모형 효과의 결과를 보려면 테이블에서 **수정된 모형** 셀을 클릭합니다.

예측자 중요도. 보기에 표시되는 예측자를 제어하는 예측자 중요도 슬라이더가 있습니다. 모형을 바꾸지는 않지만 가장 중요한 예측자에 집중할 수 있습니다. 기본적으로 상위 10개 효과가 표시됩니다.

유의수준. 예측자 중요도를 기준으로 표시되는 것 외에 보기에 표시되는 효과를 더욱 제어하는 유의수준 슬라이더가 있습니다. 슬라이더 값보다 큰 유의수준이 있는 효과는 숨겨집니다. 모형을 바꾸지는 않지만 가장 중요한 효과에 집중할 수 있습니다. 기본적으로 이 값은 1.00이므로, 유의수준을 기준으로 필터링된 효과가 없습니다.

계수

이 보기는 모형에서 각 계수의 값을 표시합니다. 요인(범주형 예측자)이 모형 내에서 코딩된 지표이므로 요인을 포함하는 **효과**에는 일반적으로 여러 관련 **계수**가 있으며, 중복(참조) 모수에 해당하는 범주를 제외하고 각 범주에 하나씩 있습니다.

유형. 여러 가지 표시 유형이 있으며 **유형** 드롭다운 목록에서 액세스할 수 있습니다.

- **다이어그램.** 먼저 절편을 표시한 다음 예측자 중요도를 줄여 위에서 아래로 효과를 정렬하는 도표입니다. 요인을 포함하는 효과 내에서 계수가 데이터 값의 오름차순으로 정렬됩니다. 다이어그램의 연결선은 계수의 부호에 따라 색상이 지정되고(다이어그램 키 참조) 계수 유의수준을 기준으로 가중되며 선이 굵을수록 더 유의한 계수(더 작은*p*-값)입니다. 마우스 커서를 연결선 위에 놓으면 계수 값, *p*-값, 모수와 연결된 효과의 중요도를 보여주는 도구 팁이 표시됩니다. 기본값 유형입니다.
- **표.** 개별 모형 계수의 값, 유의수준 검정 및 신뢰구간을 표시합니다. 절편 이후, 예측자 중요도를 줄여 효과가 위에서 아래로 정렬됩니다. 요인을 포함하는 효과 내에서 계수가 데이터 값의 오름차순으로 정렬됩니다. 기본적으로는 테이블이 축소되어 각 모형 모수의 계수, 유의수준 및 중요도만 표시됩니다. 표준 오차, *t* 통계 및 신뢰구간을 보려면 테이블에서 **계수** 셀을 클릭합니다. 테이블에서 모형 모수의 이름 위에 마우스 커서를 놓으면 모수의 이름, 모수와 연결된 효과, (범주형 예측자의 경우) 모형 모수와 연결된 값 레이블을 보여주는 도구 팁이 표시됩니다. 이것은 자동 데이터 준비에서 유사한 범주의 범주형 예측자를 합칠 때 만들어지는 새 범주를 보려는 경우에 특히 유용합니다.

예측자 중요도. 보기에 표시되는 예측자를 제어하는 예측자 중요도 슬라이더가 있습니다. 모형을 바꾸지는 않지만 가장 중요한 예측자에 집중할 수 있습니다. 기본적으로 상위 10개 효과가 표시됩니다.

유의수준. 예측자 중요도를 기준으로 표시되는 것 외에 보기에 표시되는 계수를 더욱 제어하는 유의수준 슬라이더가 있습니다. 슬라이더 값보다 큰 유의수준이 있는 계수는 숨겨집니다. 모형을 바꾸지는 않지만 가장 중요한 계수에 집중할 수 있습니다. 기본적으로 이 값은 1.00이므로, 유의수준을 기준으로 필터링된 계수가 없습니다.

평균 추정

유의한 예측자에 대해 표시되는 도표입니다. 도표는 다른 모든 예측자를 상수로 유지하면서 수직축에 대상의 모형 추정값을 표시하고 수평축에 예측자의 각 값을 표시합니다. 대상에서 각 예측자 계수의 효과를 시각화하는 데 유용합니다.

참고: 유의한 예측자가 없는 경우 평균 추정이 생성되지 않습니다.

모형 작성 요약

모형 선택 알고리즘으로 모형 선택 설정에서 **없음** 외에 다른 항목을 선택한 경우 모형설정 프로세스에 대한 세부정보를 제공합니다.

단계적 전진. 단계적 전진이 선택 알고리즘인 경우, 테이블에 단계적 알고리즘의 마지막 10단계가 표시됩니다. 각 단계에 대해 선택 기준값 및 해당 단계에서 모형의 효과가 표시됩니다. 각 단계가 모형에 얼마나 기여하는지 알 수 있습니다. 지정된 단계에서 모형에 어떤 효과가 있는지 쉽게 알 수 있도록 각 열에서 행을 정렬할 수 있습니다.

최적 서브세트. 최적 서브세트가 선택 알고리즘인 경우, 테이블에 상위 10개 모형이 표시됩니다. 각 모형에 대해 선택 기준값 및 모형의 효과가 표시됩니다. 상위 모형의 안정성을 알 수 있으며, 차이가 별로 없는 유사한 효과가 많은 경우 "상위" 모형에서 상당히 신뢰할 수 있습니다. 아주 다른 효과가 있는 경우 몇몇 효과가 매우 비슷하므로 결합하거나 하나를 제거해야 합니다. 지정된 단계에서 모형에 어떤 효과가 있는지 쉽게 알 수 있도록 각 열에서 행을 정렬할 수 있습니다.

선형 회귀

선형 회귀분석으로 선형 방정식의 계수를 추정합니다. 이 선형 방정식에는 종속변수 값을 적절히 예측하는 독립변수가 한 개 이상 포함됩니다. 예를 들어, 판매원의 연간 총매출(종속변수)을 연령, 학력, 경력 등의 독립변수로부터 예측할 수 있습니다.

예. 농구 시즌에서 한 농구 팀이 이긴 경기 수와 경기 당 해당 팀의 평균 스코어 수가 서로 관련이 있습니까? 산점도는 이러한 변수가 선형 관계가 있음을 보여 줍니다. 이긴 경기 수와 상대편의 평균 스코어 수는 선형 관계가 있습니다. 이러한 변수는 음의 관계를 이룹니다. 이긴 경기 수가 늘어날수록 상대편의 평균 스코어 수는 줄어듭니다. 선형 회귀분석에서 이러한 변수 간의 관계를 모형으로 나타낼 수 있습니다. 적절한 모형을 사용하여 팀이 얼마나 많은 경기에서 승리할 것인지 예측할 수 있습니다.

통계. 각 변수에 대해 유효한 케이스 수, 평균 및 표준 편차를 선택할 수 있습니다. 각 모형에 대해 회귀계수, 상관행렬, 부분상관 및 편상관, 다중 R, R^2, 수정된 R^2, R^2 변화량, 추정값의 표준 오차, 분산 분석표, 예측값 및 잔차를 선택할 수 있습니다. 또한 각 회귀계수에 대한 95% 신뢰구간, 분산-공분산행렬, 분산팽창요인, 허용 오차, Durbin-Watson 검정, 거리 측도(Mahalanobis, Cook, 레버리지 값), DFBETA, DFFIT, 예측 구간, 케이스별 진단정보 등을 선택할 수 있습니다. 도표에 대해 산점도, 편회귀잔차 도표, 히스토그램 및 정규 확률 도표를 선택할 수 있습니다.

선형 회귀분석 데이터 고려 사항

데이터. 종속변수와 독립변수는 양적변수이어야 합니다. 범주형 변수(예: 종교, 전공, 거주 지역)는 이분형(더미) 변수나 다른 유형의 대비변수로 기록되어야 합니다.

가정. 각 독립변수 값에서 종속변수의 분포는 정규 분포이어야 합니다. 종속변수 분포의 분산은 모든 독립변수 값에 대해 항상 일정해야 합니다. 종속변수와 각 독립변수 간 관계는 선형적이어야 하며 모든 관측값은 독립적이어야 합니다.

선형 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 회귀분석 > 선형...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

2. 선형 회귀분석 대화 상자에서 숫자 종속변수를 선택합니다.
3. 숫자 독립변수를 한 개 이상 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 독립변수를 블록으로 분류하고 서로 다른 변수 세트에 대해 입력 방법을 달리 지정합니다.
- 선택 변수를 선택하여 분석을 이 변수에 대한 특정 값을 갖는 케이스 변수 세트로 제한합니다.
- 케이스 식별변수를 선택하여 도표의 점을 식별합니다.
- 가중최소제곱 분석에 대해 WLS 가중치 숫자변수를 선택합니다.

WLS. 가중 최소제곱 모형을 구할 수 있습니다. 데이터 점은 분산의 역수에 의해 가중됩니다. 즉, 분산이 큰 관측값은 분산이 작은 관측값보다 분석에 영향을 적게 준다는 의미입니다. 가중 변수 값이 0, 음수 또는 결측이면 해당 케이스는 분석에서 제외됩니다.

선형 회귀분석 변수 선택 방법

방법을 선택하면 독립변수를 분석에 입력하는 방법을 지정할 수 있습니다. 다른 방법을 사용하여 같은 변수 세트에서 여러 회귀 모형을 작성할 수 있습니다.

- 진입*(회귀분석)(Enter (Regression))*. 한 블록에 있는 모든 변수가 단일 단계에 입력되는 변수 선택 프로시저입니다.
- *Stepwise*. 각 단계에서 F의 확률이 가장 작은 방정식에 없는 독립변수가 입력됩니다 (확률이 충분히 작은 경우). 확률 F가 충분히 커지면 회귀 방정식에 이미 있는 변수가 제거됩니다. 이 방법은 포함하거나 제거할 변수가 더 이상 없으면 종료됩니다.
- 제거*(Remove)*. 블록의 모든 변수가 단일 단계에서 제거되는 변수 선택 프로시저입니다.
- 후진제거법*(Backward Elimination)*. 모든 변수를 방정식에 입력한 다음 순차적으로 제거하는 변수 선택 프로시저입니다. 종속변수와 함께 편상관이 가장 작은 변수가 가장 먼저 제거됩니다. 변수가 제거 기준을 충족하면 제거되고 첫 번째 변수가 제거되면 방정식에 남아 있는 변수 중 편상관이 가장 작은 변수가 그 다음으로 제거됩니다. 방정식에 제거 기준을 만족하는 변수가 없으면 프로시저가 중단됩니다.
- 전진 선택법*(Forward Selection)*. 변수가 순차적으로 모델에 입력되는 순차적 변수 선택 절차. 방정식에 입력할 첫 번째 변수는 종속변수와 최대 양의 상관 또는 음의 상관에 있는 변수입니다. 이 변수는 진입 기준을 만족시켰을 때만 방정식에 입력됩니다. 첫 번째 변수를 입력하면 방정식에 없고 편상관이 가장 큰 독립변수가 다음에 입력되도록 고려됩니다. 진입 기준에 맞는 변수가 없으면 프로시저가 중단됩니다.

결과의 유의수준은 단일 모형 적합을 기준으로 합니다. 따라서 유의수준은 단계적 방법(단계선택, 전진, 후진)을 사용하는 경우에는 일반적으로 유효하지 않습니다.

모든 변수는 지정된 입력 방법과 상관 없이 방정식에 입력되도록 허용 오차 기준을 넘어야 합니다. 기본값 허용 오차 수준은 0.0001입니다. 또한, 기본값 허용 오차 수준 때문에 해당 모형에서 다른 변수의 허용 오차가 허용 오차 기준보다 떨어지는 경우에는 변수가 입력되지 않습니다.

선택한 모든 독립변수는 단일 회귀 모형에 추가됩니다. 단, 다른 변수 서브세트에 대해 다른 입력 방법을 지정할 수 있습니다. 예를 들어, 단계별 선택을 사용하여 회귀 모형에 변수 블록을 입력하거나 전진 선택법을 사용하여 두 번째 블록을 입력할 수 있습니다. 회귀 모형에 두 번째 변수 블록을 추가하려면 **다음**을 클릭합니다.

선형 회귀분석 규칙 설정

선택 규칙으로 정의한 케이스는 분석에 포함됩니다. 예를 들어, 변수를 선택하고 **같음**을 선택한 다음 값으로 5를 입력한 경우에는 선택한 변수의 값이 5인 케이스만 분석에 포함됩니다. 문자열 값도 사용할 수 있습니다.

선형 회귀분석 도표

도표를 사용하여 분산의 정규성, 선형성, 등식 등을 유효하게 가정할 수 있습니다. 도표는 이상값, 특별 관측값, 영향 케이스 등을 탐색하는 데 사용됩니다. 새로운 변수로 저장한 다음에 데이터 편집기에서 독립변수로 도표를 작성하는 데 예측값, 잔차, 대각선 등을 사용할 수 있습니다. 다음 도표를 사용할 수 있습니다.

산점도. 표준화 잔차, 삭제된 잔차, 수정된 예측값, 스튜던트화 잔차, 삭제된 스튜던트화 잔차 중 두 가지를 도표로 사용할 수 있습니다. 표준화 예측값과 비교하여 표준화 잔차를 도표화하고 분산의 선형성과 등식을 확인할 수 있습니다.

소스 변수 목록. 표준화 예측값(*ZPRED), 표준화 잔차(*ZRESID), 삭제된 잔차(*DRESID), 수정된 예측값(*ADJPRED), 스튜던트화 잔차(*SRESID), 삭제된 스튜던트화 잔차(*SDRESID) 등의 예측변수와 잔차 변수 및 독립변수(DEPENDNT)를 나열합니다.

편회귀잔차 도표 모두 출력. 두 변수가 나머지 독립변수에서 따로 회귀될 때 각 독립변수의 잔차와 종속변수의 잔차에 대한 산점도를 표시합니다. 둘 이상의 독립변수의 경우 작성할 편회귀잔차 도표에 대한 방정식에 있어야 합니다.

표준화 잔차 도표. 표준화 잔차의 분포를 정규 분포와 비교하여 표준화 잔차와 정규 확률 도표의 히스토그램을 볼 수 있습니다.

도표를 요청하면 요약 통계가 표준화 예측값과 표준화 잔차(***ZPRED**와 ***ZRESID**)에 표시됩니다.

선형 회귀: 새로운 변수로 저장

예측값, 잔차, 기타 진단 정보에 유용한 통계를 저장할 수 있습니다. 각각을 선택하면 활성 데이터 파일에 새로운 변수가 여러 개 추가됩니다.

예측값. 각 케이스를 예측하는 회귀 모형 값을 말합니다.

- 표준화되지 않음. 모형이 예측하는 종속변수의 값입니다.
- 표준화됨. 각 예측값을 표준화된 형식으로 변환하는 것입니다. 즉, 평균 예측값을 예측값에서 빼고 그 차이를 예측값의 표준 편차로 나눕니다. 표준화 예측값은 평균이 0이고 표준 편차가 1이 됩니다.
- 조정됨. 회귀계수 계산에서 제외되는 케이스에 대한 예측값입니다.
- 평균 예측의 *S.E.*. 예측값의 표준 오차입니다. 독립변수와 같은 값을 갖는 케이스에 대한 종속변수 평균값의 표준 편차 추정값입니다.

거리. 회귀 모형에 큰 영향을 줄 수도 있는 독립변수나 케이스에 대한 특수 값 조합으로 케이스를 식별하는 측도가 됩니다.

- *Mahalanobis*. 독립변수의 케이스 값이 모든 케이스의 평균과 얼마나 다른지에 대한 측도입니다. Mahalanobis 거리가 크면 케이스가 독립변수 하나 이상에 대해 극단값을 갖는 것으로 식별합니다.
- *Cook's*. 특정 케이스를 회귀계수 계산에서 제외할 때 모든 케이스의 잔차가 얼마나 변경될 수 있는지에 대한 측정값입니다. Cook의 거리가 크면 회귀분석 통계 계산에서 케이스를 제외할 때 계수가 크게 변경됩니다.
- 값 활용. 회귀분석의 적합에 대한 점의 영향을 측정합니다. 중심화된 레버리지 값의 범위는 0(적합도에 아무 영향도 미치지 않음)부터 (N-1)/N까지입니다.

예측 구간. 평균과 개별 예측 구간에 대한 상한 및 하한 경계를 말합니다.

- 평균. 평균 예측반응의 예측 구간에 대한 하한 및 상한(두 변수)입니다.
- 개별. 단일 케이스에 대한 종속변수 예측 구간의 하한 및 상한(두 변수)입니다.
- 신뢰구간. 1- 99.99 사이의 값을 입력하여 두 예측 구간의 신뢰수준을 지정하십시오. 이 값을 입력하려면 먼저 평균 또는 개별을 선택해야 합니다. 일반적인 신뢰구간 값은 90, 95, 99입니다.

잔차. 종속변수의 실제 값에서 회귀 방정식을 통해 예측한 값을 뺀 것을 말합니다.

- 표준화되지 않음. 관측값과 모형에 의해 예측된 값 사이 차이입니다.
- 표준화됨. 표준 편차의 추정값으로 나눈 잔차입니다. Pearson 잔차라고도 하는 표준화 잔차는 평균이 0이고 표준 편차가 1이 됩니다.
- 스튜던트화됨. 독립변수의 평균으로부터 독립변수에 대한 각 케이스 값의 거리에 따라 케이스마다 다른 표준 편차의 추정값으로 나눈 잔차입니다. 내부적으로 스튜던트화 잔차라고도 합니다.
- 삭제됨. 해당 케이스가 회귀계수 계산에서 제외될 때 케이스에 대한 잔차입니다. 종속변수 값과 조정된 예측값 간 차이입니다.
- 스튜던터화 삭제됨. 표준 오차로 나눈 케이스의 삭제된 잔차입니다. 스튜던트화 삭제된 잔차와 연관된 스튜던트화 잔차 사이의 차이는 케이스가 자체 예측에서 작성하는 차이를 표시합니다. 외부 스튜던트화 잔차라고도 합니다.

영향력 통계. 회귀계수(DFBETA)의 변경 사항과 특수 케이스를 제외함으로써 생기는 예측값(DFFIT)입니다. 표준화된 DFBETA 및 DFFIT 값 또한 공분산 비율과 함께 사용할 수 있습니다.

- *DfBetas*. 베타 값의 차이는 특정 케이스를 제외하여 발생하는 회귀계수의 변화량입니다. 값은 모형의 각 항에 대해 계산되며 상수를 포함합니다.
- 표준화 *DFBETA(Standardized DfBeta)*. 베타 값의 표준화 차이. 특정 케이스를 제외하여 나타난 회귀계수의 변화량을 나타냅니다. 절대값이 2보다 큰 케이스를 N 제곱근(N은 케이스 수)으로 나누어 결과 값을 확인합니다. 값은 모형의 각 항에 대해 계산되며 상수를 포함합니다.
- *DfFit*. 적합값의 차이는 특정 케이스를 제외하여 나타난 예측값의 변화량입니다.
- 표준화 *DFFIT(Standardized DfFit)*. 적합값의 표준화 차이. 특정 케이스를 제외하여 나타난 예측값의 변화량을 나타냅니다. 절대값이 2보다 큰 표준화 값을 p/N 제곱근(p는 모형에서 모수의 수이고 N은 케이스 수)으로 나누어 결과 값을 확인합니다.
- 공분산 비율. 회귀계수 계산에서 모든 케이스가 포함된 공분산 행렬식에 대한 특정 케이스가 제외된 공분산 행렬식의 비율입니다. 이 비율이 1에 가까우면 케이스로 인해 공분산행렬이 크게 달라지지 않습니다.

계수 통계. 데이터 세트 또는 데이터 파일에 회귀계수를 저장합니다. 데이터 세트는 같은 세션에서 다시 사용할 수 있으나 세션을 마치기 전에 명시적으로 저장하지 않는 한 파일로 저장되지 않습니다. 데이터 세트 이름은 변수 이름 규칙을 준수해야 합니다.

XML 파일에 모형정보 내보내기. 모수 추정값과 필요에 따라 해당 공분산을 XML(PMML) 형식의 지정된 파일로 내보냅니다. 스코어링 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.

선형 회귀분석 통계

다음의 통계를 사용할 수 있습니다.

회귀계수- 추정값 은 회귀계수 *B*, 표준 오차 *B*, 표준화 계수 베타, *B*에 대한 *t* 값 및 *t*의 양측 유의 수준을 표시합니다. **신뢰구간**은 각 회귀분석 계수 또는 공분산 행렬에 대해 지정된 신뢰도 레벨로 신뢰구간을 표시합니다. **공분산 행렬**은 대각선이 있으면 분산을 사용하고 대각선이 없으면 공분산을 사용하여 회귀계수의 분산-공분산행렬을 표시합니다. 상관행렬도 표시됩니다.

모형 적합- 모델에서 입력되고 제거된 변수가 나열되며 다중 *R*, R^2 및 조정된 R^2, 추정값의 표준 오차 및 분산 분석 표와 같은 '적합도 통계량' 이 표시됩니다.

R 제곱 변화량- 독립변수를 추가하거나 삭제하여 생성되는 R^2 통계의 변화량입니다. 하나의 변수와 연결된 R^2 변화량이 큰 경우 해당 변수는 종속변수의 예측자로 적합합니다.

설명- 분석의 각 변수에 대한 유효 케이스 수, 평균 및 표준 편차를 제공합니다. 한쪽 방향 유의수준이 있는 상관행렬과 각 상관에 대한 케이스 수도 표시됩니다.

부분상관*(Part Correlation)*. 모델에 있는 다른 독립변수의 선형 효과가 독립변수에서 제거된 경우 종속변수와 독립변수 간 상관입니다. 변수가 방정식에 추가될 때의 R 제곱 변화량과 관련이 있습니다. 준 편상관이라고도 합니다.

편상관*(Partial Correlation)*. 다른 변수와의 상호 연관으로 인한 상관을 제거한 후 두 변수 사이에 남아 있는 상관입니다. 또한 종속변수와 독립변수 모두에서 모형에 있는 다른 독립변수의 선형 효과를 제거할 때 발생하는 종속변수와 독립변수 간 상관을 나타냅니다.

공선성 진단- 공선성 (또는 다중 공선성) 은 하나의 독립변수가 다른 독립변수의 선형 함수일 때 바람직하지 않은 상황입니다. 척도와 비중심화 교차곱 행렬, 조건 지수, 분산-분해 비율 등의 고유 값이 분산팽창요인(VIF)과 개별 변수에 대한 허용 오차와 함께 표시됩니다.

선택 기준- Akaike 정보 기준 (AIC), Ameniya의 예측 기준 (PC), 예측 기준의 조건부 평균 제곱 오차 (Cp) 및 Schwarz Bayesian 기준 (SBC) 을 포함합니다. 통계는 모델 요약 테이블에 표시됩니다.

잔차- 다른 모델을 비교하기 위해 교차 검증 통계로 사용할 **'PRESS 통계'** 를 선택할 수 있습니다. 잔차의 연속 상관에 대한 **'더반-Watson'** 테스트도 표시합니다. 선택 기준 (*n* 표준 편차보다 큰 이상치) 을 충족하는 케이스에 대한 **'케이스별 진단'** 정보를 선택합니다.

선형 회귀분석 옵션

다음 옵션을 사용할 수 있습니다.

선택법 기준. 이러한 옵션은 앞으로, 뒤로 또는 단계적 변수 선택 방법이 지정된 경우에 적용됩니다. 모델에서 변수를 입력하거나 제거할 수 있는 방법은 *F* 값의 유의성 (확률) 또는 *F* 값 자체에 따라 달라집니다.

- *F 확률 사용(Use Probability of F)*. F값의 유의 수준이 진입값보다 작으면 변수가 모델에 입력되고 유의 수준이 제거값보다 크면 제거됩니다. 진입값은 제거값보다 작아야 하며 두 값 모두 양수이어야 합니다. 모형에 변수를 더 많이 입력하려면 진입값을 높이고 모형에서 변수를 더 많이 제거하려면 제거값을 낮춥니다.
- *F 값 사용(Use F Value)*. F값이 진입값보다 크면 변수가 모형에 입력되고 F값이 제거값보다 작으면 제거됩니다. 진입값은 제거값보다 커야 하고 두 값 모두 양수이어야 합니다. 모형에 더 많은 변수를 입력하려면 진입값을 낮추고 모형에서 변수를 더 많이 제거하려면 제거값을 높입니다.

허용치. 기본적으로 값은 .0001입니다. 공차는 방정식에서 다른 독립변수에 의해 설명되지 않는 변수의 분산 비율입니다. 고려 중인 변수가 분석에 포함된 경우 방정식의 변수에 대한 최소 허용 오차는 방정식에 포함되지 않은 변수의 최소 허용 오차입니다. 회귀분석 방정식에 들어가고 유지하려면 변수가 허용 오차 및 최소 허용 오차 검정을 통과해야 합니다. 변수가 허용 오차 기준을 통과하는 경우 적용되는 방법에 따라 포함될 수 있습니다.

방정식에 상수항 포함. 기본적으로 회귀 모형에는 상수항이 들어 있습니다. 이 옵션을 선택하지 않으면 회귀 작업이 원점을 통과하게 됩니다. 원점을 통과하는 일부 회귀의 결과는 상수항이 있는 회귀의 결과와 다릅니다. 예를 들어, R^2은 일반적인 방법으로 해석할 수 없습니다.

결측값. 다음 중 하나를 선택할 수 있습니다.

- **목록별 결측값 제외.** 모든 변수에 대한 유효한 값이 있는 케이스만 분석에 포함됩니다.
- **대응별 결측값 제외.** 상관된 대응변수의 전체 데이터를 포함하고 있는 케이스는 회귀 분석을 기초로 하는 상관계수를 계산할 때 사용됩니다. 자유도는 최소 대응별 *N*에 따라 다릅니다.
- **평균으로 바꾸기.** 계산에는 모든 케이스가 사용되며 변수의 평균이 결측 관측값을 대체합니다.

REGRESSION 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 원시 데이터 자리에서 상관행렬을 쓰거나 행렬을 읽어 회귀분석을 얻습니다(MATRIX 하위 명령 사용).
- CRITERIA 하위 명령을 사용하여 허용 오차 수준을 지정합니다.
- 같거나 다른 종속 변수에 대한 여러 모형을 얻습니다(METHOD 및 DEPENDENT 하위 명령 사용).
- 추가 통계를 얻습니다(DESCRIPTIVES 및 STATISTICS 하위 명령 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

순서 회귀분석

순서 회귀분석을 사용하여 예측자 세트(요인 또는 공변량)에 대한 다중 순서 반응의 종속성을 모형화할 수 있습니다. 순서 회귀분석 계획은 McCullagh(1980, 1998) 방법을 기초로 하며 이 순서 회귀분석 프로시저는 구문에서 PLUM이라고도 합니다.

표준 선형 회귀분석은 반응(종속) 변수와 예측(독립) 변수 가중 결합간의 차이 제곱값 합계를 최소화하는 것입니다. 추정된 계수는 예측변수의 변경이 반응에 미치는 영향을 반영합니다. 반응수준의 변화량은 반응범위 전체에 걸쳐 동등하다는 의미에서 반응은 숫자로 가정합니다. 예를 들어, 키가 150cm인 사람과 140cm인 사람의 키 차이는 10cm이고, 이 차이는 키가 210cm인 사람과 200cm인 사람의 키 차이가 10cm인 것과 동일한 의미를 갖습니다. 이러한 관계는 순서변수에 대해 반드시 유지할 필요는 없으며 반응 범주를 임의로 선택하고 범주 수도 임의로 결정할 수 있습니다.

예. 순서 회귀분석은 약 복용량에 대한 환자의 반응을 연구하는 데 사용할 수 있습니다. 가능한 반응은 없음, 약함, 보통, 심각함 등으로 분류할 수 있습니다. 약함과 보통 반응간의 차이는 양적으로 명시하기가 어렵거나 불가능하므로 직관에 의존합니다. 또한 약함과 보통 반응간의 차이는 보통과 심각함 반응간의 차이보다 더 크거나 작을 수도 있습니다.

통계 및 도표. 관측/기대빈도 및 누적빈도, 빈도 및 누적빈도에 대한 Pearson 잔차, 관측/기대확률, 공변량 패턴에 따른 각 반응의 관측/기대 누적 확률, 모수 추정값의 근사 상관 및 공분산행렬, Pearson의 카이제곱 및 우도비 카이제곱, 적합도 통계량, 반복 히스토리, 평행선 검정 가정, 모수 추정값, 표준 오차, 신뢰구간, Cox 및 Snell, Nagelkerke, McFadden의 R^2 통계를 구합니다.

순서 회귀분석 데이터 고려 사항

데이터. 종속변수는 순서로 가정하며 숫자 또는 문자일 수 있습니다. 순서는 종속변수 값의 오름차순 정렬로 결정됩니다. 가장 낮은 값을 첫 번째 범주로 정의합니다. 요인변수는 범주형으로 가정됩니다. 공변량 변수는 숫자여야 합니다. 둘 이상의 연속 공변량을 사용하면 매우 큰 셀 확률표가 생성될 수 있습니다.

가정. 반응변수는 하나만 사용할 수 있고 반드시 지정해야 합니다. 또한 전체 독립변수 값에 대한 고유 패턴별 반응은 독립 다항 변수로 가정됩니다.

관련 프로시저. 명목 로지스틱 회귀분석은 명목 종속변수에 대해 유사한 모형을 사용합니다.

순서 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **회귀분석** > **순서...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 종속변수를 하나 선택합니다.
3. **확인**을 클릭하십시오.

순서 회귀분석: 옵션

옵션 대화 상자를 사용하면 반복 추정 알고리즘에 사용된 모수를 조정하여 모수 추정값에 대한 신뢰수준을 선택하고 연결 함수를 선택할 수 있습니다.

반복. 반복 알고리즘을 사용자 정의할 수 있습니다.

- **최대반복수.** 음수가 아닌 정수를 지정합니다. 0을 지정하면 프로시저는 초기 추정값을 반환합니다.
- **최대 단계-반분.** 양의 정수를 지정합니다.
- **로그-우도 수렴.** 로그-우도의 절대 변화량 또는 상대 변화량이 이 값보다 작으면 알고리즘 수행이 중지됩니다. 기준은 0이 지정된 경우 사용되지 않습니다.
- **모수 수렴.** 각 모수 추정값의 절대 변화량 또는 상대 변화량이 이 값보다 작으면 알고리즘 수행이 중지됩니다. 기준은 0이 지정된 경우 사용되지 않습니다.

신뢰구간. 0보다 크거나 같고 100보다 작은 값을 지정합니다.

델타. 영(0) 셀 빈도에 추가된 값입니다. 1보다 작은 음수가 아닌 정수를 지정합니다.

비정칙성 허용 오차. 종속도가 높은 예측자 확인에 사용됩니다. 옵션 목록에서 값을 선택합니다.

연결 함수. 연결 함수는 누적 확률의 변환으로 모형을 추정할 수 있습니다. 다음 다섯 개의 연결 함수를 사용할 수 있습니다.

- **로짓.** $f(x)=\log(x/(1-x))$. 일반적으로 균등하게 분포된 카테고리에 사용됩니다.
- **보 로그-로그.** $f(x)=\log(-\log(1-x))$. 일반적으로 더 높은 범주가 더 개연성이 있는 경우에 사용됩니다.
- **음수 로그-로그.** $f(x)=-\log(-\log(x))$. 일반적으로 더 낮은 범주가 더 개연성이 있는 경우에 사용됩니다.
- **프로빗.** $f(x)=\Phi^{-1}(x)$. 일반적으로 잠복 변수가 정규 분포일 경우 사용됩니다.
- **Cauchit(역 Cauchy).** $f(x)=\tan(\pi(x-0.5))$. 일반적으로 잠복 변수에 많은 극단값이 있는 경우 사용됩니다.

순서 회귀분석 결과

결과 대화 상자에서 뷰어에 표시할 테이블을 작성하고 변수를 작업 파일에 저장할 수 있습니다.

표시. 다음에 대한 테이블을 작성합니다.

- **n단계마다 반복계산과정 출력.** 로그-우도 및 모수 추정값이 반복 인쇄에 지정된 빈도에 따라 인쇄됩니다. 첫 번째 및 마지막 반복은 항상 인쇄됩니다.
- **적합도 통계량.** Pearson 및 우도비 카이제곱 통계로 변수 목록에 지정된 분류기준에 따라 계산됩니다.
- **요약 통계.** Cox 및 Snell, Nagelkerke, McFadden의 R^2 통계입니다.
- **모수 추정값.** 모수 추정값, 표준 오차 및 신뢰구간입니다.
- **모수 추정값의 근사 상관.** 모수 추정값 상관행렬입니다.

- **모수 추정값의 근사 공분산.** 모수 추정값의 공분산행렬입니다.
- **셀 정보.** 관측/기대빈도 및 누적빈도, 빈도 및 누적빈도에 대한 Pearson 잔차, 관측/기대확률, 공변량 패턴에 따른 각 반응 범주의 관측/기대 누적 확률입니다. 여러 공변량 패턴을 사용하는 모형(예: 연속 공변량을 사용하는 모형)의 경우 이 옵션으로 매우 큰 테이블이 생성될 수 있습니다.
- **라인 평행성 검정.** 위치 모수가 종속변수 수준 전체에 걸쳐 동등하다는 가설을 검정합니다. 이 검정은 위치 전용 모형인 경우에만 사용할 수 있습니다.

저장 변수. 다음 변수를 작업 파일에 저장합니다.

- **반응확률 추정.** 요인/공변량 패턴을 반응 범주로 분류하는 모형 추정 확률입니다. 확률은 반응 범주 수만큼 다양합니다.
- **예측 범주.** 요인/공변량 패턴에 대한 최대 추정 확률을 포함하는 반응 범주입니다.
- **예측 범주 확률.** 요인/공변량 패턴을 예측 범주로 분류하는 추정 확률입니다. 또한 이 확률은 요인/공변량 패턴에 대한 최대 추정 확률입니다.
- **실제 범주 확률.** 요인/공변량 패턴을 실제 범주로 분류하는 추정 확률입니다.

로그 우도 출력. 로그-우도의 출력을 제어합니다. **다항 상수항 포함**을 선택하면 우도의 전체 값이 출력됩니다. 결과를 상수가 포함되지 않은 결과와 비교하려면 다항 상수항을 제외하도록 선택할 수 있습니다.

순서 회귀분석 위치 모형

위치 대화 상자에서 분석에 대한 위치 모형을 지정할 수 있습니다.

모형 지정. 주효과 모형에 공분산 주효과와 요인 주효과는 포함되지만 상호작용 효과는 포함되지 않습니다. 사용자 정의 모형을 작성하여 요인 상호작용이나 공변량 상호작용 서브세트를 지정할 수 있습니다.

요인/공변량. 요인 및 공분산이 나열됩니다.

위치 모형. 모형은 선택한 주효과와 상호작용 효과에 따라 다릅니다.

선택된 요인과 공변량에 대해 다음과 같습니다.

상호작용
선택한 모든 변수의 최고-수준 상호작용 항을 생성합니다. 이는 기본값입니다.

주효과
선택한 각 변수에 대한 주효과 항을 생성합니다.

모든 2원 효과
선택한 변수에 대해 가능한 이원 상호작용을 모두 생성합니다.

모든 3원
선택한 변수에 대해 가능한 3원배치 상호작용을 모두 생성합니다.

모든 4원
선택한 변수에 대해 가능한 4원배치 상호작용을 모두 생성합니다.

모든 5원
선택한 변수에 대해 가능한 5원배치 상호작용을 모두 생성합니다.

항 설정 및 사용자 정의 항

항 설정
선택된 요인과 공변량 세트의 모든 조합에 대해 특정 유형(예: 주효과)의 비중첩 항을 포함하려면 이 선택사항을 사용합니다.

사용자 정의 항 설정
중첩 항을 포함하고자 하거나 변수별로 항 변수를 명시적으로 설정하고자 할 때 이 선택사항을 사용합니다. 중첩 항 설정에는 다음 단계가 포함됩니다.

순서 회귀분석 척도 모형

척도 대화 상자에서 분석에 대한 척도 모형을 지정할 수 있습니다.

요인/공변량. 요인 및 공분산이 나열됩니다.

척도 모형. 모형은 선택한 주효과와 상호작용 효과에 따라 다릅니다.

선택된 요인과 공변량에 대해 다음과 같습니다.

상호작용
선택한 모든 변수의 최고-수준 상호작용 항을 생성합니다. 이는 기본값입니다.

주효과
선택한 각 변수에 대한 주효과 항을 생성합니다.

모든 2원 효과
선택한 변수에 대해 가능한 이원 상호작용을 모두 생성합니다.

모든 3원
선택한 변수에 대해 가능한 3원배치 상호작용을 모두 생성합니다.

모든 4원
선택한 변수에 대해 가능한 4원배치 상호작용을 모두 생성합니다.

모든 5원
선택한 변수에 대해 가능한 5원배치 상호작용을 모두 생성합니다.

항 설정 및 사용자 정의 항

항 설정
선택된 요인과 공변량 세트의 모든 조합에 대해 특정 유형(예: 주효과)의 비중첩 항을 포함하려면 이 선택사항을 사용합니다.

사용자 정의 항 설정
중첩 항을 포함하고자 하거나 변수별로 항 변수를 명시적으로 설정하고자 할 때 이 선택사항을 사용합니다. 중첩 항 설정에는 다음 단계가 포함됩니다.

PLUM 명령 추가 기능

선택을 구문 창으로 붙여넣고 결과로 나타나는 PLUM 명령 구문을 편집하는 경우 순서 회귀분석을 사용자 정의할 수 있습니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 귀무가설을 모수의 선형결합으로 지정하여 가설 검정을 사용자 정의합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

선형 탄력적 순 회귀

Linear Elastic Net은 Python sklearn.linear_model.ElasticNet 클래스를 사용하여 하나 이상의 독립변수에서 종속변수에 대한 정규화된 선형 회귀 모델을 추정합니다. 정규화는 L1 (Lasso) 및 L2 (Ridge) 페널티를 결합합니다. 확장에는 지정된 L1 비율에 대한 알파의 다른 값에 대한 추적 도표를 표시하고 교차 검증을 기반으로 L1 비율 및 알파 하이퍼 매개변수 값을 선택하기 위한 선택적 모드가 포함됩니다. 단일 모델이 적합하거나 교차 검증을 사용하여 페널티 비율 및/또는 알파를 선택하는 경우, 검증 데이터의 파티션을 사용하여 표본 외 성능을 추정할 수 있습니다.

L1 페널티 및 알파 정규화 모수의 비율의 지정된 값으로 모델을 맞추는 것 외에도 선형 탄성 넷은 지정된 비율의 알파 값 범위에 대한 계수 값의 추적 도표를 표시하거나 지정된 값 격자에서 k-fold 교차 검증을 통해 하이퍼 모수 값의 선택을 용이하게 할 수 있습니다. 교차 검증을 통해 단일 모델이 적합하거나 비율 및/또는 알파 선택이 수행되는 경우, 최종 모델은 모델의 샘플 외 성능의 유효한 추정치를 얻기 위해 입력 데이터의 파티션에 의해 생성되는 보유 데이터에 적용될 수 있습니다.

Linear Elastic Net Regression 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 회귀분석 > 선형 OLS 대체 > 탄력적 네트워크

   대화 상자를 사용하여 활성 데이터 세트의 각 케이스를 훈련 또는 검증용 표본에 할당하는 변수를 지정할 수 있습니다.

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

2. 숫자 목표변수를 선택합니다. 분석을 실행하는 데 목표 변수가 한 개만 있으면 됩니다.
3. 숫자 종속을 지정하십시오.
4. 하나 이상의 범주형 요인변수 또는 숫자 공변량 변수를 지정하십시오.

선택적으로 **파티션** 은 지정되거나 선택된 모델의 표본 외 성능 추정을 위해 입력 데이터의 검증용 또는 검정 서브세트를 작성하는 방법을 제공합니다. 모든 파티셔닝은 프로시저에서 사용되는 변수에 대해 유효하지 않은 데이터가 있는 케이스를 목록별로 삭제한 후에 수행됩니다. 교차 검증의 경우 훈련 데이터의 중첩 또는 파티션이 Python에서 작성됩니다. 파티션에 의해 작성되는 검증용 데이터는 유효한 모드에 관계없이 추정에 사용되지 않습니다.

훈련 및 검증용 파티션에서 각 표본에 무작위로 할당된 케이스의 비율을 지정하거나 훈련 또는 검증용 표본에 각 케이스를 할당하는 변수를 사용하여 파티션을 정의할 수 있습니다. 훈련 및 변수를 둘 다 지정할 수는 없습니다. 파티션이 지정되지 않은 경우 입력 데이터의 약 30%가 작성되는 홀드아웃 샘플이 작성됩니다.

훈련% 는 훈련 표본에 무작위로 할당할 활성 데이터 세트에 있는 케이스의 상대적 수를 지정합니다. 기본 교육은 70%입니다.

Linear Elastic Net Regression: 옵션

옵션 탭은 다음에 대한 옵션을 제공합니다.

최빈값(D)

이 선택사항은 다음 모드 중 하나를 지정하기 위한 옵션을 제공합니다.

지정된 L1 비율 및 알파로 맞춤
이 모드를 선택하면 지정된 L1 비율 및 알파 정규화 값을 사용하여 단일 모델이 훈련 데이터에 맞춰집니다. 이는 기본값입니다. 파티션이 지정된 경우, 적합한 단일 또는 최종 모델이 샘플 외 성능을 추정하기 위해 보유된 테스트 데이터에 적용됩니다.

도표에서 관측값 및/또는 잔차 대 예측값 도표를 선택할 수 있습니다.

저장아래에서 저장할 예측값 및 잔차를 지정할 수 있습니다.

트레이스 플롯
이 모드를 선택하면 훈련 데이터에 대한 세 개의 도표가 지정된 알파 값 세트에 대한 알파 함수로 표시됩니다.

- 회귀계수의 추적 도표입니다.
- R^2의 도표입니다.
- 평균 제곱 오차 (MSE) 의 도표입니다.

파티션이 유지되지만 이 모드의 최종 모델 결과가 없으므로 보류된 테스트 데이터에 대한 결과가 제공되지 않습니다.

교차 검증을 통한 L1 비율 및/또는 알파 선택
이 모드를 선택하면 모델을 평가하기 위해 교차 검증을 사용하는 격자 검색이 수행되고 최적의 비율 및 알파 값이 검증 중첩에 대한 최상의 평균 R^2 를 기반으로 선택됩니다. **교차 검증 중첩 수** 필드를 사용하여 교차 검증에 대한 다섯 개의 분할 또는 중첩의 기본값을 변경할 수 있습니다. 파티션이 지정된 경우, 적합한 단일 또는 최종 모델이 샘플 외 성능을 추정하기 위해 보유된 테스트 데이터에 적용됩니다.

표시에서 비율 및 알파 값 (**최적**) 이 선택된 모델에 대한 기본 정보, 비교된 모든 모델에 대한 기본 정보 (**모델 비교**) 또는 모든 모델의 모든 분할 또는 중첩에 대한 전체 정보 (**모델 및 중첩 비교**) 만 표시하도록 선택할 수 있습니다. **Best** 가 기본값입니다.

도표에서 검증 중첩에 대한 평균 R^2 및/또는 MSE 도표를 선택할 수 있습니다. 관측값 및/또는 잔차 대 예측값의 도표도 선택할 수 있습니다.

저장아래에서 저장할 예측값 및 잔차를 지정할 수 있습니다.

개별 L1 비율 지정
지정된 L1 비율 및 알파 또는 **추적 도표** 모드에 대해 이 모드를 선택하면 단일 L1 페널티 비율 값을 지정할 수 있습니다. **L1 비율 및/또는 교차 검증을 통한 알파 선택** 모드에 대해 선택한 경우 다중 값을 지정할 수 있습니다.

알파 값의 눈금 지정
L1 비율 및/또는 교차 검증을 통한 알파 선택 모드에 대해 이 모드를 선택하면 **시작** 값 (value1) 에서 **종료** 값 (value2) 까지 **By** (value3) 의 증분으로 고유한 알파 값의 눈금을 지정할 수 있습니다. 지정된 경우 하나의 유효한 [value1 TO value2 BY value3] 세트만 허용됩니다. $0 \leq value1 \leq value2 \leq 1$을 충족해야 합니다. value1 = value2이면 value3에 관계없이 단일 value1을 지정하는 것과 같습니다.

도표는 다양한 알파 값의 수평 X축에 대해 지정된 메트릭을 사용하여 표시됩니다.

개별 알파 지정
지정된 L1 비율 및 알파에 맞춤 모드에 대해 이 모드를 선택하면 단일 알파 정규화 값을 지정할 수 있습니다. **교차 검증을 통한 추적 도표** 또는 **L1 비율 및/또는 알파 선택** 모드에 대해 선택한 경우 다중 값을 지정할 수 있습니다.

알파 값의 눈금 지정
교차 검증을 통한 추적 도표 또는 **L1 비율 및/또는 알파 선택** 모드에 대해 이 모드를 선택하는 경우, 고유한 알파 값의 격자는 **시작** 값 (value1) 에서 증분이 **By** (value3) 인 **종료** 값 (value2) 까지 지정할 수 있습니다. 지정된 경우 하나의 유효한 [value1 TO value2 BY value3] 세트만 허용됩니다. 이는 해당 $0 \leq value1 \leq value2 \leq 1$를 충족해야 합니다. value1 = value2인 경우에는 value3에 관계없이 단일 value1를 지정하는 것과 동일합니다.

값의 범위에 대한 **알파 메트릭** 은 **선형** 또는 **기본 10대수** (10은 지정된 값의 제곱으로 올림) 일 수 있습니다.

도표는 다양한 알파 값의 수평 X축에 대해 지정된 메트릭을 사용하여 표시됩니다.

기준

분석을 제어합니다.

절편 포함
이 기준은 적합 모델에 절편을 포함합니다. 확장 프로시저는 종속변수를 중심으로 하거나 표준화하지 않으며 추정 중에 절편에 페널티를 주지 않습니다.

예측 변수 표준화
모든 독립변수를 표준화합니다.

교차 검증 중첩 수
모델의 교차 검증 평가를 위한 분할 또는 중첩 수입니다. 1보다 큰 양의 정수 값이어야 합니다. 기본값은 5입니다.

Python 랜덤 상태
모델의 교차 검증 평가를 수행하는 동안 사용되는 Python 의 random_state 설정 값입니다. 의사 난수와 관련된 결과의 복제를 허용합니다. 값은 0-2^{32}-1 범위의 정수여야 합니다. 기본값은 0입니다.

시간 제한 (분)
모델 계산을 실행하는 데 허용되는 시간 (분) 입니다. 0을 지정하면 타이머가 꺼집니다. 기본값은 5입니다.

표시

이 옵션은 **L1 비율 및/또는 교차 검증을 통한 알파 선택** 모드에 대해 표시할 출력 양을 지정합니다.

최고
선택한 최적 모델에 대한 기본 결과만 표시합니다. 이는 기본적으로 설정됩니다.

모델 비교
평가된 모든 모형의 기본 결과를 표시합니다.

모형과 중첩 비교
평가된 각 모델의 각 분할 또는 접기에 대한 전체 상세 결과를 표시합니다.

도표

이 옵션은 교차 검증을 사용하여 관측값 또는 잔차 값 대 예측값의 도표를 지정하고 교차 검증 중첩 대 알파 값에 대한 평균 제곱 오차 (MSE) 및/또는 평균 R^2 의 도표를 지정합니다.

평균 교차 검증 평균 제곱 오차 (MSE) 대 알파
교차 검증을 통한L1 비율 및/또는 알파 선택 모드의 경우 교차 검증 중첩에 대한 평균 MSE대 지정되거나 선택된 최상의 L1 비율 값에 대한 알파의 선도표를 표시합니다. **추적 도표** 모드의 경우 전체 훈련 데이터를 기반으로 유사한 도표가 자동으로 생성됩니다.

평균 교차 검증 R 제곱 대 알파
교차 검증을 통한L1 비율 및/또는 알파 선택 모드의 경우, 지정되거나 선택된 최상의 L1 비율 값에 대해 교차 검증 중첩 대 알파에 대한 평균 R^2 의 선도표를 표시합니다. **추적 도표** 모드의 경우 전체 훈련 데이터를 기반으로 유사한 도표가 자동으로 생성됩니다.

관측 대 예측
지정된 모형 또는 최고의 모형에 대해 관측값 대 예측값의 산포도를 표시합니다.

잔차 대 예측
지정된 모형 또는 최고 모형에 대한 잔차값 대 예측값의 산포도를 표시합니다.

저장

활성 데이터 세트에 저장할 변수를 지정합니다.

예측값
지정된 모델 또는 최적 모델의 예측값을 활성 데이터 세트에 저장합니다. **사용자 정의 변수 이름**을 지정할 수도 있습니다.

잔차
지정된 모델 예측 또는 최상의 모델 예측의 잔차를 활성 데이터 세트에 저장합니다. **사용자 정의 변수 이름**을 지정할 수도 있습니다.

선형 라소 회귀분석

Linear Lasso는 Python sklearn.linear_model.Lasso 클래스를 사용하여 하나 이상의 독립변수에서 종속변수에 대한 L1 손실 정규화 선형 회귀 모형을 추정하고, 추적 도표를 표시하고 교차 검증을 기반으로 알파 하이퍼 매개변수 값을 선택하는 선택적 모드를 포함합니다. 단일 모델이 적합하거나 교차 검증을 사용하여 알파를 선택하는 경우 검증 데이터의 파티션을 사용하여 표본 외 성능을 추정할 수 있습니다.

알파 정규화 매개변수의 지정된 값으로 모델을 맞추는 것 외에도, 선형 lasso는 알파 값 범위에 대한 계수 값의 추적 도표를 표시하거나 지정된 값 그리드에서 k-fold 교차 검증을 통해 하이퍼매개변수 값의 선택을 용이하게 할 수 있습니다. 교차 검증을 통해 단일 모델이 적합하거나 알파 선택이 수행되는 경우, 최종 모델은 모델의 샘플 외 성능의 유효한 추정치를 얻기 위해 입력 데이터의 파티션에 의해 작성되는 보류된 데이터에 적용될 수 있습니다.

선형 Lasso 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 회귀분석 > 선형 OLS 대체 > 라소

   대화 상자를 사용하여 활성 데이터 세트의 각 케이스를 훈련 또는 검증용 표본에 할당하는 변수를 지정할 수 있습니다.

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 숫자 목표변수를 선택합니다. 분석을 실행하는 데 목표 변수가 한 개만 있으면 됩니다.
3. 숫자 종속을 지정하십시오.
4. 하나 이상의 범주형 요인변수 또는 숫자 공변량 변수를 지정하십시오.

선택적으로 **파티션** 은 지정되거나 선택된 모델의 표본 외 성능 추정을 위해 입력 데이터의 검증용 또는 검정 서브세트를 작성하는 방법을 제공합니다. 모든 파티셔닝은 프로시저에서 사용되는 변수에 대해 유효하지 않은 데이터가 있는 케이스를 목록별로 삭제한 후에 수행됩니다. 교차 검증의 경우 훈련 데이터의 중첩 또는 파티션이

Python에서 작성됩니다. 파티션에서 작성된 검증용 데이터는 유효한 모드에 관계없이 추정에 사용되지 않습니다.

훈련 및 검증용 파티션에서 각 표본에 무작위로 할당된 케이스의 비율을 지정하거나 훈련 또는 검증용 표본에 각 케이스를 할당하는 변수를 사용하여 파티션을 정의할 수 있습니다. 훈련 및 변수를 둘 다 지정할 수는 없습니다. 파티션이 지정되지 않은 경우 입력 데이터의 약 30%가 작성되는 홀드아웃 샘플이 작성됩니다.

훈련% 는 훈련 표본에 무작위로 할당할 활성 데이터 세트에 있는 케이스의 상대적 수를 지정합니다. 기본 교육은 70%입니다.

선형 올가미 회귀분석: 옵션

옵션 탭은 다음에 대한 옵션을 제공합니다.

최빈값(D)

이 선택사항은 다음 모드 중 하나를 지정하기 위한 옵션을 제공합니다.

지정된 알파에 맞춤
이 모드를 선택하면 하나의 알파 정규화 값만 사용하여 단일 모델이 훈련 데이터에 맞춰집니다. 이는 기본적으로 설정됩니다. 파티션이 지정된 경우, 적합한 단일 또는 최종 모델이 샘플 외 성능을 추정하기 위해 보유된 테스트 데이터에 적용됩니다.

도표에서 관측값 및/또는 잔차 대 예측값 도표를 선택할 수 있습니다.

저장아래에서 저장할 예측값 및 잔차를 지정할 수 있습니다.

트레이스 플롯
이 모드를 선택하면 훈련 데이터에 대한 세 개의 도표가 지정된 알파 값 세트에 대한 알파 함수로 표시됩니다.

- 회귀계수의 추적 도표입니다.
- R^2의 도표입니다.
- 평균 제곱 오차 (MSE) 의 도표입니다.

파티션이 유지되지만 이 모드의 최종 모델 결과가 없으므로 보류된 테스트 데이터에 대한 결과가 제공되지 않습니다.

교차 검증을 통한 알파 선택

교차 검증을 사용하는 격자 검색을 선택하여 모델을 평가하고 최적 평균을 기반으로 최적 알파를 선택하는 경우 검증 중첩에 대해 R^2 를 선택합니다. **교차 검증 중첩 수** 필드를 사용하여 교차 검증에 대한 다섯 개의 분할 또는 중첩의 기본값을 변경할 수 있습니다. 파티션이 지정된 경우, 적합한 단일 또는 최종 모델이 샘플 외 성능을 추정하기 위해 보유된 테스트 데이터에 적용됩니다.

표시에서 선택한 알파 값 (**최적**) 이 있는 모델에 대한 기본 정보, 비교된 모든 모델에 대한 기본 정보 (**모델 비교**) 또는 모든 모델에 대한 모든 분할 또는 중첩에 대한 전체 정보 (**모델 및 중첩 비교**) 만 표시하도록 선택할 수 있습니다. **Best** 가 기본값입니다.

도표에서 검증 중첩에 대한 평균 R^2 및/또는 MSE 도표를 선택할 수 있습니다. 관측값 및/또는 잔차 대 예측값의 도표도 선택할 수 있습니다.

저장아래에서 저장할 예측값 및 잔차를 지정할 수 있습니다.

개별 알파 지정
지정된 알파에 맞춤 모드를 선택하면 단일 알파 정규화 값을 지정할 수 있습니다. **추적 도표** 또는 **교차 검증을 통한 알파 선택** 모드를 선택할 때 다중 값을 지정할 수 있습니다.

값
하나 이상의 양의 알파 정규화 값을 지정하십시오. 여러 값을 개별적으로 또는 범위로 지정할 수 있습니다. 기본값은 1입니다.

알파 값의 눈금 지정
추적 도표 또는 **교차 검증을 통한 알파 선택** 모드를 선택하면 **시작** 값 (value1) 에서 **종료** 값 (value2) 까지 **By** (value3) 의 증분을 사용하여 고유 알파 값의 격자를 지정할 수 있습니다. 지정된 경우 하나의 유효한

[value1 TO value2 BY value3] 세트만 허용됩니다. 0 ≤ value1 ≤ value2 ≤ 1을 충족해야 합니다. value1 = value2이면 value3에 관계없이 단일 value1을 지정하는 것과 같습니다.

값의 범위에 대한 **알파 메트릭** 은 **선형** 또는 **기본 10대수** (10은 지정된 값의 제곱으로 올림) 일 수 있습니다.

도표는 다양한 알파 값의 수평 X축에 대해 지정된 메트릭을 사용하여 표시됩니다.

기준

분석을 제어합니다.

절편 포함
적합 모델에 절편을 포함합니다. 확장 프로시저는 종속변수를 중심으로 하거나 표준화하지 않으며 추정 중에 절편에 페널티를 주지 않습니다.

예측 변수 표준화
모든 독립변수를 표준화합니다.

교차 검증 중첩 수
모델의 교차 검증 평가를 위한 분할 또는 중첩 수입니다. 1보다 큰 양의 정수 값이어야 합니다. 기본값은 5입니다.

Python 랜덤 상태
모델의 교차 검증 평가를 수행할 때 사용되는 Python 의 random_state 설정 값입니다. 의사 난수와 관련된 결과의 복제를 허용합니다. 0-2^{32}-1 범위의 정수여야 합니다. 기본값은 0입니다.

시간 제한 (분)
모델 계산을 실행하는 데 허용되는 시간 (분) 입니다. 0을 지정하면 타이머가 꺼집니다. 기본값은 5입니다.

표시

교차 검증을 통한 알파 선택 모드에 대해 표시할 출력의 양을 지정합니다.

최고
선택한 최적 모델에 대한 기본 결과만 표시합니다. 이는 기본값입니다.

모델 비교
평가된 모든 모형의 기본 결과를 표시합니다.

모형과 중첩 비교
평가된 각 모델의 각 분할 또는 접기에 대한 전체 상세 결과를 표시합니다.

도표

교차 검증 중첩 대 예측값에 대한 관측값 또는 잔차 값의 도표를 지정하고 교차 검증 중첩 대 알파 값에 대한 평균 평균 제곱 오차 (MSE) 및/또는 평균 R^2 의 도표를 지정합니다.

평균 교차 검증 평균 제곱 오차 (MSE) 대 알파
교차 검증을 통한 알파 선택 모드의 경우 교차 검증 중첩 대 알파 평균 MSE의 선도표를 표시합니다. **추적 도표** 모드의 경우 전체 훈련 데이터를 기반으로 유사한 도표가 자동으로 생성됩니다.

평균 교차 검증 R 제곱 대 알파
교차 검증을 통한 알파 선택 모드의 경우 교차 검증 중첩 대 알파에 대한 평균 R^2 의 선도표를 표시합니다. **추적 도표** 모드의 경우 전체 훈련 데이터를 기반으로 유사한 도표가 자동으로 생성됩니다.

관측 대 예측
지정된 모형 또는 최고의 모형에 대해 관측값 대 예측값의 산포도를 표시합니다.

잔차 대 예측
지정된 모형 또는 최고 모형에 대한 잔차값 대 예측값의 산포도를 표시합니다.

저장

활성 데이터 세트에 저장할 변수를 지정합니다.

예측값
지정된 모델 또는 최적 모델의 예측값을 활성 데이터 세트에 저장합니다. **사용자 정의 변수 이름**을 지정할 수도 있습니다.

잔차
지정된 모델 예측 또는 최상의 모델 예측의 잔차를 활성 데이터 세트에 저장합니다. **사용자 정의 변수 이름**을 지정할 수도 있습니다.

선형 리지 회귀분석

Linear Ridge는 Python sklearn.linear_model.Ridge 클래스를 사용하여 하나 이상의 독립변수에서 종속변수에 대한 L2 또는 제곱 손실 정규화된 선형 회귀 모형을 추정하고 교차 검증을 기반으로 추적 도표를 표시하고 알파 하이퍼 매개변수 값을 선택하는 선택적 모드를 포함합니다. 단일 모델이 적합하거나 교차 검증을 사용하여 알파를 선택하는 경우 검증 데이터의 파티션을 사용하여 표본 외 성능을 추정할 수 있습니다.

알파 정규화 파라미터의 특정 값을 갖는 모델을 피팅하는 것 외에, 선형 리지는 알파 값들의 범위에 대한 계수 값들의 리지 추적 플롯을 디스플레이할 수 있거나, 또는 값들의 특정 그리드들 상에서 k-배 교차 검증을 통해 하이퍼파라미터 값의 선택을 용이하게 할 수 있다. 교차 검증을 통해 단일 모델이 적합하거나 알파 선택이 수행되는 경우, 최종 모델은 모델의 샘플 외 성능의 유효한 추정치를 얻기 위해 입력 데이터의 파티션에 의해 작성되는 보류된 데이터에 적용될 수 있습니다.

선형 능선 회귀 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 회귀분석 > 선형 OLS 대체 > 리지

   대화 상자를 사용하여 활성 데이터 세트의 각 케이스를 훈련 또는 검증용 표본에 할당하는 변수를 지정할 수 있습니다.
2. 숫자 목표변수를 선택합니다. 분석을 실행하는 데 목표 변수가 한 개만 있으면 됩니다.
3. 숫자 종속을 지정하십시오.
4. 하나 이상의 범주형 요인변수 또는 숫자 공변량 변수를 지정하십시오.

선택적으로 **파티션** 은 지정되거나 선택된 모델의 표본 외 성능 추정을 위해 입력 데이터의 검증용 또는 검정 서브세트를 작성하는 방법을 제공합니다. 모든 파티셔닝은 프로시저에서 사용되는 변수에 대해 유효하지 않은 데이터가 있는 케이스를 목록별로 삭제한 후에 수행됩니다. 교차 검증의 경우 훈련 데이터의 중첩 또는 파티션이 Python에서 작성됩니다. 파티션에 의해 작성되는 검증용 데이터는 유효한 모드에 관계없이 추정에 사용되지 않습니다.

훈련 및 검증용 파티션에서 각 표본에 무작위로 할당된 케이스의 비율을 지정하거나 훈련 또는 검증용 표본에 각 케이스를 할당하는 변수를 사용하여 파티션을 정의할 수 있습니다. 훈련 및 변수를 둘 다 지정할 수는 없습니다. 파티션이 지정되지 않은 경우 입력 데이터의 약 30%가 작성되는 홀드아웃 샘플이 작성됩니다.

훈련% 는 훈련 표본에 무작위로 할당할 활성 데이터 세트에 있는 케이스의 상대적 수를 지정합니다. 기본 교육은 70%입니다.

선형 능선 회귀: 옵션

옵션 탭은 다음에 대한 옵션을 제공합니다.

최빈값(D)

이 선택사항은 다음 모드 중 하나를 지정하기 위한 옵션을 제공합니다.

지정된 알파에 맞춤
이 옵션을 선택하면 하나의 알파 정규화 값만 사용하는 훈련 데이터에 단일 모델이 맞춰집니다. 이는 기본적으로 설정됩니다. 파티션이 지정된 경우, 적합한 단일 또는 최종 모델이 샘플 외 성능을 추정하기 위해 보유된 테스트 데이터에 적용됩니다.

도표에서 관측값 및/또는 잔차 대 예측값 도표를 선택할 수 있습니다.

저장아래에서 저장할 예측값 및 잔차를 지정할 수 있습니다.

트레이스 플롯
이 옵션을 선택하면 훈련 데이터에 대한 세 개의 도표가 지정된 알파 값 세트에 대한 알파 함수로 표시됩니다.

- 회귀계수의 능선 추적 도표입니다.
- R^2의 도표입니다.
- 평균 제곱 오차 (MSE) 의 도표입니다.

파티션이 유지되지만 이 모드의 최종 모델 결과가 없으므로 보류된 테스트 데이터에 대한 결과가 제공되지 않습니다.

교차 검증을 통한 알파 선택

이 옵션을 선택하면 모델을 평가하기 위해 교차 검증을 사용하는 격자 검색이 수행되고 검증 중첩에 대한 최상의 평균 R^2 를 기반으로 최상의 알파가 선택됩니다. **교차 검증 중첩 수** 필드를 사용하여 교차 검증에 대한 다섯 개의 분할 또는 중첩의 기본값을 변경할 수 있습니다. 파티션이 지정된 경우, 적합한 단일 또는 최종 모델이 샘플 외 성능을 추정하기 위해 보유된 테스트 데이터에 적용됩니다.

표시에서 선택한 알파 값 (**최적**) 이 있는 모델에 대한 기본 정보, 비교된 모든 모델에 대한 기본 정보 (**모델 비교**) 또는 모든 모델에 대한 모든 분할 또는 중첩에 대한 전체 정보 (**모델 및 중첩 비교**) 만 표시하도록 선택할 수 있습니다. **Best** 가 기본값입니다.

도표에서 검증 중첩에 대한 평균 R^2 및/또는 MSE 도표를 선택할 수 있습니다. 관측값 및/또는 잔차 대 예측값의 도표도 선택할 수 있습니다.

저장아래에서 저장할 예측값 및 잔차를 지정할 수 있습니다.

개별 알파 지정
지정된 알파에 맞춤 모드를 선택하면 단일 알파 정규화 값을 지정할 수 있습니다. **추적 도표** 또는 **교차 검증을 통한 알파 선택** 모드를 선택할 때 다중 값을 지정할 수 있습니다.

값
하나 이상의 양의 알파 정규화 값을 지정하십시오. 여러 값을 개별적으로 또는 범위로 지정할 수 있습니다. 기본값은 1입니다.

알파 값의 눈금 지정
추적 도표 또는 **교차 검증을 통한 알파 선택** 모드를 선택하면 **시작** 값 (value1) 에서 **종료** 값 (value2) 까지 **By** (value3) 의 증분을 사용하여 고유 알파 값의 격자를 지정할 수 있습니다. 지정된 경우 하나의 유효한 [value1 TO value2 BY value3] 세트만 허용됩니다. $0 \leq$ value1 $\leq$ value2 ≤ 1을 충족해야 합니다. value1 = value2이면 value3에 관계없이 단일 value1을 지정하는 것과 같습니다.

값의 범위에 대한 **알파 메트릭** 은 **선형** 또는 **기본 10대수** (10은 지정된 값의 제곱으로 올림) 일 수 있습니다.

도표는 다양한 알파 값의 수평 X축에 대해 지정된 메트릭을 사용하여 표시됩니다.

기준

분석을 제어합니다.

절편 포함
이 기준은 하나 이상의 적합 모델에 절편을 포함합니다. 확장 프로시저는 종속변수를 중심으로 하거나 표준화하지 않으며 추정 중에 절편에 페널티를 주지 않습니다.

예측 변수 표준화
이 기준은 모든 독립변수를 표준화합니다.

교차 검증 중첩 수
모델의 교차 검증 평가에 대한 분할 또는 중첩 수를 설정하려면 이 기준을 사용하십시오. 숫자는 1보다 큰 양의 정수 값이어야 합니다. 기본 설정은 5입니다.

Python 랜덤 상태
Python 의 random_state 설정 값은 모델의 교차 검증 평가를 수행하는 동안 사용됩니다. 의사 난수와 관련된 결과의 복제를 허용합니다. 0-2^{32}-1 범위의 정수여야 합니다. 기본값은 0입니다.

시간 제한 (분)
모델 계산을 실행하는 데 허용되는 시간 (분) 입니다. 0을 지정하면 타이머가 꺼집니다. 기본값은 5입니다.

표시

이 섹션은 **교차 검증을 통한 알파 선택** 모드에 대해 표시할 출력의 양을 지정합니다.

최고
선택한 최적 모델에 대한 기본 결과만 표시합니다. 이는 기본적으로 설정됩니다.

모델 비교
또한 평가된 모든 모델에 대한 기본 결과를 표시합니다.

모형과 중첩 비교
마지막으로 평가된 각 모델의 각 분할 또는 접기에 대한 전체 상세 결과를 표시합니다.

도표

교차 검증 중첩 대 알파 값에 대한 평균 평균 제곱 오차 (MSE) 및/또는 평균 R^2 의 지정 사항을 사용하여 관측 또는 잔차 값 대 예측값의 도표를 지정합니다.

평균 교차 검증 평균 제곱 오차 (MSE) 대 알파
교차 검증을 통한 알파 선택 모드는 교차 검증 중첩 대 알파에 대한 평균 MSE의 선도표를 표시합니다. **추적 도표** 모드의 경우 전체 훈련 데이터를 기반으로 유사한 도표가 자동으로 생성됩니다.

평균 교차 검증 R 제곱 대 알파
교차 검증을 통한 알파 선택 모드는 교차 검증 중첩 대 알파에 대한 평균 R^2 의 선도표를 표시합니다. **추적 도표** 모드의 경우 전체 훈련 데이터를 기반으로 유사한 도표가 자동으로 생성됩니다.

관측 대 예측
지정된 모형 또는 최적 모형에 대한 관측값 대 예측값의 산점도를 표시합니다.

잔차 대 예측
또한 지정된 모델 또는 최적 모델에 대한 잔차 대 예측값의 산점도를 표시합니다.

저장

활성 데이터 세트에 저장할 변수를 지정합니다.

예측값
지정된 모델 또는 최적 모델의 예측값을 활성 데이터 세트에 저장합니다. 그렇지 않으면 **사용자 정의 변수 이름**을 지정할 수 있습니다.

잔차
지정된 모델 예측 또는 최상의 모델 예측의 잔차를 활성 데이터 세트에 저장합니다. 그렇지 않으면 **사용자 정의 변수 이름**을 지정할 수 있습니다.

매개 분석

매개 분석은 독립 변수가 하나 이상의 매개 변수를 통해 종속 변수에 영향을 미치는 메커니즘을 이해하기 위해 사용되는 통계적 방법입니다. 이 절차는 간접 효과에 대한 가설을 검증하고 데이터 내의 영향 경로를 파악하는 데 도움이 됩니다.

의 중재 분석 절차는 다음과 IBM SPSS 통계 같은 기능을 제공합니다:

- 예측 변수, 매개 변수 및 결과 변수 간의 직접적 및 간접적 효과를 분석한다
- 전체적 영향을 구성 요소로 분해하여 인과 관계를 파악한다
- 연속형, 이분형, 계수형 데이터를 포함한 다양한 변수 유형을 처리합니다
- 공변량 조정을 통한 교란 변수 통제
- 효과 이질성을 검토하기 위해 하위군 분석을 수행한다
- 경로 다이어그램 및 기타 그래프를 통해 매개 변수 간의 관계를 시각화합니다

주요 개념

- **예측 변수 (A)** : 그 효과를 분석하고자 하는 독립 변수(예: 치료 노출, 중재 수준).
- **매개 변수(M)** : 예측 변수의 효과를 결과 변수에 전달하는 중간 변수(예: 심리적 상태, 생물학적 지표).

- **결과 변수(Y)** : 예측 변수와 매개 변수의 영향을 모두 받는 종속 변수(예: 건강 결과, 성과 지표).
- **직접 효과** : 매개변수를 거치지 않고 종속변수에 미치는 예측변인의 영향.
- **간접 효과** : 매개변수를 통해 결과 변수에 미치는 예측 변수의 영향.
- **총 효과** : 직접적 경로와 간접적 경로를 모두 종합하여, 예측 변수가 결과에 미치는 전반적인 영향.

애플리케이션

중재 분석은 다양한 학문 분야에서 널리 활용되고 있습니다:

- **보건과학** : 행동적 또는 생물학적 메커니즘을 통해 중재가 건강 결과에 어떤 영향을 미치는지 이해하기
- **심리학** : 치료가 인지적 또는 정서적 과정을 통해 정신 건강에 어떤 영향을 미치는지 살펴보기
- **사회과학** : 정책이 중간 변수를 통해 사회적 결과에 어떤 영향을 미치는지 분석하기
- **분야** : 마케팅 전략이 고객의 태도를 통해 매출에 미치는 영향 분석
- **학력** : 학생의 참여도가 학습 성과에 미치는 영향을 연구

이 절차는 특히 다음과 같은 경우에 유용합니다:

- 인과적 메커니즘에 관한 이론적 모델을 검증한다
- 매개 경로를 파악하여 중재 대상을 선정한다
- 서로 다른 인과 경로의 상대적 중요성을 평가한다
- 하위군 간에 효과가 다른지 평가한다
- 관찰 연구에서 교란 변수의 통제

분석 유형

이 절차는 두 가지 분석 방식을 지원합니다:

- **인과 매개 분석** : 통제된 직접 효과(CDE), 자연적 직접 효과(NDE), 자연적 간접 효과(NIE) 및 분해 구성 요소를 포함한 포괄적인 인과 효과를 계산합니다. 이 접근법은 예측변수와 매개변수 간의 상호작용을 고려할 수 있게 하며, 상세한 효과 분해를 제공합니다.
- **전통적 매개 분석** : 상호작용 항이나 상세한 분해 없이 보다 단순한 매개 효과를 계산합니다. 이 접근법은 예측변수와 매개변수 간의 상호작용이 없을 것으로 예상될 때 적합합니다.

지원되는 변수 조합

이 절차는 다양한 변수 유형의 조합을 지원합니다:

- 연속형 예측변수, 연속형 매개변수, 연속형 결과변수
- 연속형 예측 변수, 이분형 매개 변수, 연속형 결과 변수
- 연속형 예측변수, 연속형 매개변수, 이분형 결과변수
- 연속형 예측 변수, 이분형 매개 변수, 이분형 결과 변수
- 연속형 예측변수, 연속형 매개변수, 계수형 종속변수

통계 방법론

이 절차는 회귀 분석 기반의 접근법을 사용하여 매개 효과를 추정합니다:

- 연속형 변수에 대한 선형 회귀 분석
- 이분형 변수에 대한 로지스틱 회귀 분석
- 카운트 변수에 대한 푸아송 회귀 또는 음의 이항 회귀

표준 오차와 신뢰 구간은 델타법(기본값)이나 부트스트랩 재표본 추출을 사용하여 계산할 수 있습니다. 델타법은 점근적 이론을 바탕으로 빠른 계산을 가능하게 하는 반면, 부트스트래핑은 비정규 분포나 표본 크기가 작은 경우 더 견고한 추론을 제공합니다.

통계

이 절차는 여러 범주로 분류된 포괄적인 통계 결과를 제공합니다:

회귀 모델 요약

중재 회귀 모델:

- 계수 (β_0, β_1, β_2) 및 표준 오차와 p-값
- 연속형 매개변수의 경우: $E(M|a,c) = \beta_0 + \beta_1 a + \beta_2 \boxtimes c$
- 이분형 매개변수의 경우: 로지스틱 회귀 모델

결과 회귀 모델:

- 계수 (θ_0, θ_1, θ_2, θ_3, θ_4) 및 표준 오차와 p-값
- 연속형 결과 변수의 경우: $E(Y|a,m,c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 (a \times m) + \theta_4 c$
- 이분형 결과의 경우: 로지스틱 회귀 모델
- 계수형 결과에 대해서는: 포아송 회귀 또는 음의 이항 회귀

효과 측정 지표

노출 수준 비교 통계:

- **CDE(통제된 직접 효과)** : 매개변수가 일정하게 유지된 상태에서 예측변인이 결과에 미치는 영향
- **NDE(자연적 직접 효과)** : 매개변수가 자연스럽게 변동할 수 있도록 하는 직접 효과
- **NIE(자연적 간접 효과)** : 매개변수를 통해 매개되는 효과
- 연속형 결과에 대해서는 차이(difference)로, 이분형 결과에 대해서는 오즈비(OR) 또는 초과 상대 위험도 (excRR)로 보고됨

분해 분석 결과:

- **TE(총 효과)** : 예측 변수가 결과에 미치는 전체적인 효과
- **PIE(순간접효과)** : 간접효과 구성 요소
- **TDE(총 직접 효과)** : 직접 효과 구성 요소
- **PAI(순 상호작용)** : 예측 변수에 기인하는 상호작용
- **INTmed (중재된 상호작용)** : 중재자를 통한 상호작용
- **INTref (참조 상호작용)** : 기준 상호작용 효과

비율 통계 (이분법적 결과에 대한):

- **PM(매개 비율)** : 전체 효과 중 매개된 비율
- **Percent(INT)** : 상호작용에 의한 비율
- **(ELIM) 비율** : 제거된 비율

추론 방법

델타 방법 (기본값):

- 공분산 행렬($\Sigma\beta$, $\Sigma\theta$)을 바탕으로 계산된 표준 오차
- 정규 분포에 기반한 신뢰 구간
- 대부분의 분석에 적합한 빠른 계산

부트스트랩 방법:

- 재표본화를 기반으로 한 표준오차와 신뢰구간
- 기본값: 95% 신뢰 수준에서 1000개의 부트스트랩 표본
- 백분위수 또는 BCa(편향 보정 및 가속화) 구간
- 비정규 분포에 대해 더 견고함

공분산 행렬 옵션:

- **모델** : 가정된 모델 분산을 사용합니다(데이터가 모델에 잘 부합할 때 가장 적합함)
- **ROBUST** : 이상치와 이분산성을 보정합니다(이분형 변수에 권장됨)

시각화 옵션

경로 다이어그램:

- 중재 경로의 시각적 표현

- 효과 크기와 함께 직접적 및 간접적 효과를 보여준다
- 공변량 효과의 선택적 표시

산점도:

- 예측변인, 매개변인 및 결과변인 간의 관계
- 신뢰 구간을 포함한 추세선 (선형 또는 로지스틱)

막대 그래프:

- 범주형 변수의 개수 또는 비율
- 오차 막대와 신뢰 구간

포레스트 플롯:

- 효과 측정값의 요약 및 신뢰 구간
- 경로 계수 또는 효과 추정치를 표시합니다

데이터 고려 사항

표 *1.* 다양한 요구 사항

가변	요구사항
결과 변수	• 연속형, 이분형 또는 계수형 데이터 • 이분법적 결과의 경우: 기준 범주(LAST, FIRST 또는 특정 값)를 지정하십시오 • 계수형 결과의 경우: 포아송 분포와 음의 이항 분포 중 하나를 선택하십시오
중재 변수	• 연속형 데이터 또는 이진형 데이터 • 연속형 매개변수의 경우: 분석에 사용할 수준(평균, 중앙값 또는 특정 값)을 지정하십시오 • 이분법적 중재 변수의 경우: 기준 범주를 지정하십시오
예측 변수	• 연속형 또는 범주형 데이터 • 효과 비교를 위해 기준 수준(BLEVEL)과 노출 수준(ELEVEL)을 지정하십시오 • 선택적 예측변수-매개변수 상호작용 항 (INTERACTION = TRUE/FALSE)

모형 지정 사항

이 절차는 일반화 선형 모델(GLM) 프레임워크를 사용하며, 분포와 연결 함수를 모두 지정해야 합니다.

중재자 모델의 경우	결과 모델을 위해
• **정체 함수를 사용하는 정규 분포는** 연속형 매개변수에 대해 선형 회귀를 의미한다 • **로짓 연결 함수를 사용하는 이항 분포는** 이분형 매개변수에 대한 로지스틱 회귀를 의미한다	• **신원 링크(Identity link)를 적용한 정규 분포는** 연속형 결과에 대해 선형 회귀를 의미한다 • **로짓 연결 함수를 사용하는 이항 분포는** 이분형 결과에 대한 로지스틱 회귀를 의미한다 • **로그 링크 함수를 갖는 포아송 분포는** 계수형 데이터에 대한 포아송 회귀를 의미합니다(분산이 평균과 같다고 가정함)

중재자 모델의 경우	결과 모델을 위해
	• 과분산 계수 데이터(분산이 평균을 초과하는 경우)를 위한 **로그 연결 함수를 사용한 음의 이항 분포** – **NEGBIN(1)** : 분산 매개변수 α = 1로 고정 – **NEGBIN(MLE)** : 데이터로부터 분산 매개변수를 추정합니다

분산 매개변수 α는 과분산을 조절합니다: $Var(Y) = \mu + \alpha\mu^2$. α = 0일 때, 이 모델은 푸아송 분포로 환원됩니다. α 값이 클수록 평균에 대한 변동성이 더 크다는 것을 의미합니다.

표본 크기 및 수렴

표 *2.* 알고리즘 모수

매개변수	설명	범위
최대 반복 횟수	최대반복계산	10에서 1000 복잡한 모델의 경우 200~500으로 늘리십시오
PCONVERGE	수렴기준 공차한계	1E-8 1E-3 으로 연구 논문에는 1E-8 를, 탐색적 분석에는 1E-4 를 사용하세요
비정칙	특이점 감지 임계값	1E-14 과 1E-6 사이의 숫자 값
맥스스테팔빙	단계적 감축 시도	1~20
리지 페널티	리지 정규화	1E-8 1E-3 으로

데이터 품질 고려사항

데이터 누락

- 분석 변수에 결측값이 있는 사례는 제외됩니다
- 사건 처리 요약 보고서: 유효한 사례 및 제외된 사례

가중값

- 주파수 가중치는 'Scale Weight' 변수를 사용하여 지원됩니다
- 부트스트랩 재표본화는 가중치를 가장 가까운 정수로 반올림합니다

공변량

- 혼란 요인을 통제하기 위해 공변량을 포함한다
- 중재변수와 결과변수 모두에 영향을 미칠 수 있는 요인들을 조정한다

부집단 분석

- BY 절을 사용하여 요인(예: 성별, 연령대)별로 별도의 분석을 수행합니다
- 집단 간 효과의 이질성을 분석하는 데 유용함

가정

모형 가정

- 예측변수-매개변수, 예측변수-결과변수, 매개변수-결과변수 간의 관계에서 측정되지 않은 혼란 요인이 없음
- 분포 함수와 연결 함수의 올바른 정의

- 인과적 해석을 위한 시간적 순서 (예측변수 → 매개변수 → 결과변수)
- 변수에 측정 오차가 없음

통계적 가정

- 안정적인 매개변수 추정을 위한 적절한 표본 크기
- 부트스트랩의 경우: 충분한 수의 표본(최소 1000개 권장)
- 이분법적 결과의 경우: 변수당 충분한 사건 수 (최소 10~15건 권장)
- 상호작용 모델의 경우: 상호작용 효과를 검출하기에 충분한 검정력

중재 분석 결과 확보

1. 메뉴에서 다음을 선택합니다.

   분석 > 중재 분석

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

모형 유형

'모델 유형' 탭에서 **매개변수** 유형과 **결과 변수 유형** 을 선택하여 매개 분석의 통계적 특성을 정의합니다.

분석할 중재 모델의 유형을 선택하십시오. **'중재 변수** 유형' 및 '결과 변수 유형'의 설정은 기본 분포, 연결 함수, 그리고 이후 탭에서 사용할 수 있는 변수를 결정합니다. 이 탭에서 선택한 항목에 따라 **'종속 변수'** 및 **'기준'** 탭에서 활성화되는 옵션이 결정됩니다.

'중재자 유형'과 '결과 유형'은 필수 항목입니다.

중개자 유형

중재 변수를 어떻게 모델링할지 선택하십시오.

연속형
동일성 연결 함수를 갖는 정규 분포를 가정한다. 중재변수로는 연속형 변수만 허용됩니다.

중재 변수가 연속형 척도로 측정되는 경우 이 옵션을 사용하십시오.

2진

LOGIT 연결 함수를 갖는 이항 분포를 가정합니다. 중재 변수로는 이분형 변수만 허용됩니다.

중재자가 두 가지 가능한 값을 가질 때 이 옵션을 사용하십시오.

결과 유형

결과 변수를 모델링할 방식을 선택하십시오. 결과 변수의 유병률이 최소 0.30 이상일 때 출력 테이블이 생성됩니다.

연속형
동일성 연결 함수를 갖는 정규 분포를 가정한다.

'연속형 결과 변수' 옵션을 선택하십시오. 이는 데이터가 정규 분포를 따르고 연결 함수가 IDENTITY임을 의미합니다. 이 옵션을 선택하면 **'종속 변수** ' 탭에서 **결과** 변수를 연속형 변수만 허용하도록 설정할 수 있습니다.

2진

이항 분포를 가정한다. **LOGIT** 함수나 **LOG** 함수 중 하나를 선택할 수 있습니다.

이진 결과 유형을 이항 분포와 로짓 연결 함수를 사용하여 설정하는 옵션을 선택하십시오. 이 옵션을 선택하면 **'종속 변** 수' 탭에서 '결과 변수' 필드를 구성하고 '이분형 결과'를 활성화할 수 있습니다.

'연결 함수' 드롭다운 목록에서 **'Logit'** (기본값) 또는 '이분 결과용 **로그** 연결 함수'를 선택할 수 있습니다. 기본적으로 **'Binary'** 옵션은 이항 분포와 로짓 연결 함수를 설정합니다. **'종속 변수** ' 탭에서 결과 변수를 **'이분형 (이분형** 값 중에서 선택)'으로 지정하여 이분형 값 중에서 선택할 수 있습니다.

빈도

'종속 변수 ' 탭에서 counts 변수를 **결과** 변수로 설정하려면 이 옵션을 선택하십시오. 로그 연결 함수를 사용하여 **푸아송** 분포나 **음의 이항** 분포를 추가할 수 있습니다. **'부정 이항'** 옵션에 대해 추가적인 매개변수 설정을 사용할 수 있습니다. 모델에서 과분산 정도를 조절하는 분산 매개변수 α를 설정할 수 있습니다. α 값이 클수록 평균에 대한 변동성이 더 크다는 것을 의미합니다.

- **값 지정** : 고정된 양의 정수를 입력하십시오.
- **추정값** : 최대우도법을 사용하여 분산 매개변수 α를 추정하려면 이 옵션을 선택하십시오.

종속변수

'종속 변수 ' 탭에서 매개 분석의 결과를 정의하는 결과 변수를 지정할 수 있습니다.

반응(결과) 변수를 지정하십시오. 연속형, 이분형(기준 범주 포함) 및 계수형 변수에 대한 사용자 지정 옵션을 사용할 수 있습니다. 또한 연속형 및 이분형(기준 범주 포함) 변수에 대한 옵션을 포함하여 mediator 변수를 정의합니다. 효과 분석 및 '규모 가중치' 변수 포함에 대한 제어 수준을 설정할 수 있습니다.

또한 ' 중재 변수' 필드에 중재 변수를 지정할 수 있으며, 이는 중재 분석을 위한 중재 변수 구성 요소를 정의합니다. 모든 변수는 종속 변수, 독립 변수 및 척도 가중치 변수 내에서뿐만 아니라 이들 간에도 고유해야 합니다.

결과 변수

중재 분석을 위한 결과를 정의합니다. 연속형, 이진형 및 카운트형 변수를 제공하십시오. 이는 **'모델** 유형' 탭에서 지정한 결과 유형에 따라 결정되며, 기본값은 연속형 변수입니다. 이분법적 결과를 선택한다면, 이분법 변수를 선택할 수 있습니다. 포아송 분포나 음의 이항 분포를 선택한 경우, 계수 변수를 선택할 수도 있습니다.

이분형 결과 변수

결과 회귀 분석에서 결과를 '이분형'으로 선택할 수 있습니다. 이분법적 결과를 선택한 경우, 선택한 이분법 변수에 대한 기준값을 지정할 수 있습니다. 기본적으로 이항 분포와 로짓/로그 연결 함수가 사용됩니다.

결과 변수의 유병률이 최소 0.30 이상일 때 출력 테이블이 생성됩니다.

참조범주

이 옵션을 사용하면 **'모델** 유형' 탭에서 선택한 이분법 결과 유형에 대한 기준 범주를 정의할 수 있습니다. 다음 옵션 중 하나를 선택하십시오:

- **마지막(최고값)** : 마지막 값 또는 최고값을 기준값으로 설정합니다.
- **첫 번째(최소값)** : 첫 번째 또는 최소값을 기준값으로 설정합니다.
- **사용자 지정** : 사용자 지정 참조 선택 항목을 설정하려면 이 옵션을 선택하십시오. '결과 변수'에서 선택한 이진 변수에 대한 이진 변수 수준을 선택하거나, 해당 값이 변수 목록에 포함되어 있는 경우 직접 값을 입력할 수 있습니다.

중재 변수

이 옵션을 사용하면 매개 분석을 위한 **매개 변수를** 정의할 수 있습니다. 데이터 유형은 연속형과 이진형이어야 합니다.

이 옵션은 [**모델** 유형] 탭에서 지정한 **매개변수 유형에** 따라 결정되며, 기본값은 연속형 변수, 정규 분포, 그리고 IDENTITY 링크 함수입니다. 이분형 매개변수를 선택한 경우, 이항 분포와 LOGIT 연결 함수를 사용하여 이분형 변수를 선택할 수 있습니다.

중재자 수준 설정

제어된 직접 효과 계산에 적용할 중재 변수 수준을 선택할 수 있습니다. 다음 옵션 중 하나를 선택하십시오:

- **평균** : 이 옵션을 선택하면 매개변수의 평균값을 매개변수 수준으로 설정합니다.
- **중앙값** : 이 옵션을 선택하면 매개변수 수준을 선택한 매개변수의 중앙값으로 설정합니다.
- **사용자 지정** : 사용자 지정 레벨을 지정하려면 이 옵션을 선택하세요. 이 옵션은 이진 변수에 대해서만 선택할 수 있습니다. 중재자의 등급에 해당하는 숫자를 입력하세요. 해당 값은 바이너리 메디에이터 관측값에 존재해야 합니다. 중재자가 이진형인 경우, 사용자는 이진 값(0 또는 1)을 선택할 수 있습니다.

척도 가중값
척도 가중치 변수를 할당할 옵션을 선택하십시오. 이 필드에는 숫자 데이터 유형의 변수만 입력할 수 있습니다.

예측자

'예측 변수' 탭에서 **'주 예측** 변수' 필드에 예측 변수를 지정할 수 있으며, 이는 매개 분석에서 *A*로 표시되는 **예측 변수** 또는 **노출** 구성 요소를 정의합니다.

주요 예측 변수와 추가 요인 및 공변량을 지정하고, 효과 분석을 용이하게 하기 위해 예측 변수의 기준치 및 노출 값을 설정할 수 있습니다.

데이터셋에서 예측 변수의 관측값은 벡터 a로 표현된다.

주요 예측 변수
매개 분석을 위한 주요 예측 변수 또는 노출 변수 A를 정의합니다.

연속형 변수만 허용됩니다(노출 효과의 정의를 명확히 하기 위해 범주형 변수는 제외됩니다).

기준선 레벨 설정
이 옵션을 사용하여 기준값을 지정할 수 있습니다. **'기준 수준 설정'** 및 **'새로운 노출 수준 설정** ' 필드에 값을 입력하십시오. 예측 변수로 선택된 변수에 대해 두 값 모두 데이터셋에 포함되어 있어야 합니다.

- 기준값 설정: 예측 변수의 자연스러운 순서에서 가장 낮은 수준(또는 첫 번째 값).
- 새로운 노출 수준 설정: 동일한 순서에서 예측 변수의 바로 다음 수준(또는 두 번째 값).

요인
이 목록은 모델 내 범주형 변수를 나타내는 요인 F를 정의합니다. 각 요인은 행렬 C에 포함될 가상 변수 집합으로 변환됩니다. 이러한 더미 변수들은 집단 간 차이를 고려하기 위해 결과 변수 회귀 분석에 포함됩니다.

범주형 변수를 사용할 수 있습니다(예: 치료군, 성별 등).

요인은 선택 사항이지만, 분석에 범주형 그룹이 포함되거나 상호작용 효과가 평가되는 경우에는 이러한 변수가 필수입니다.

공변량
이 목록은 모델 내 잠재적인 교란 요인을 통제하기 위한 공변량 *C*를 정의합니다. 각 공변량은 결과 변수 회귀 모델 Y에 포함됩니다.

공변량이 범주형인 경우, 중복을 피하기 위해 모델에 포함하기 전에 더미 변수로 변환해야 합니다.

각 공변량은 결과 회귀 모델 Y에 주효과로 포함됩니다.

기준

'기준' 탭에서는 공변량 행렬에 대한 추정 방법 선택, 매개변수 추정치 산출, 추론 유형 및 분석 유형 지정, 그리고 알고리즘 조정 옵션을 설정할 수 있습니다.

공분산 행렬(V)
매개 분석에서 매개변수 추정치를 위한 공분산 행렬을 추정하는 데 사용되는 방법을 지정합니다.

- **모델 기반 추정기** : 지정된 통계적 모델 가정을 바탕으로 공분산 행렬을 계산합니다.
- **강건 추정량** : 이방분산성이나 비정규성과 같은 모델 위반 요인을 보정하여 강건한 공분산 행렬 추정치를 제공합니다.

추론 유형
이 옵션은 매개 분석에서 추론에 사용되는 방법을 지정하기 위한 것입니다.

- **델타** : 이 방법은 추정량의 점근적 분포를 이용하여 표준오차와 신뢰구간을 계산합니다.
- **부트스트랩** : 이 방법은 재표본 추출을 활용하여 신뢰 구간을 포함한 추정을 위한 추정량의 분포를 생성합니다.

  옵션:

  - **표본 수** : 산출된 백분위수 및 BCa 구간을 위해 최소 1,000개의 부트스트랩 표본을 사용하는 것이 권장됩니다. 양의 정수를 지정합니다. 기본값은 500입니다.

- **메르센 트위스터의 시드 설정** : 시드를 설정하면 분석 결과를 재현할 수 있습니다. 이 제어를 사용하는 것은 Mersenne Twister를 활성 생성기로 설정하고 난수 생성기 대화 상자에 고정 시작점을 지정하는 것과 비슷합니다. 단, 이 대화 상자에 시드를을 설정하면 난수 생성기의 현재 상태가 유지되고 분석이 완료되면 해당 상태로 복원된다는 차이점이 있습니다.
- **신뢰 구간** : 신뢰 수준을 50보다 크고 100보다 작게 지정하십시오. 백분위수 구간은 단순히 신뢰구간 백분위수에 따라 나열된 붓스트랩 값을 사용합니다. 예를 들어, 95% 백분위수 신뢰구간은 붓스트랩 값의 2.5번째 및 97.5번째 백분위수를 구간의 하한 및 상한으로 사용합니다(필요에 따라 붓스트랩 값을 보간함). BCa(Bias corrected and accelerated) 구간은 더 정확하지만 계산하는 데 더 많은 시간이 걸리는 조정 구간입니다.

분석 유형

- **Causal** : CDE, NDE, NIE, 총 효과, 오즈비, 초과 상대 위험도 및 상호작용 분해를 포함한 전체 인과적 매개 효과를 계산합니다.
- **전통적 방법** : 상호작용 항이나 완전한 인과 분해 없이 더 단순한 매개 효과를 계산합니다.

알고리즘 옵션

'모델 유형 ' 탭에서 결과 유형이 **'이진'** 또는 **'개수'로** 설정된 경우 옵션이 활성화됩니다.

- **최대 반복** 횟수: 알고리즘이 중지되기 전까지 수행할 IRWLS 반복의 최대 횟수를 설정합니다. 10 and1000 사이의 정수 값을 입력할 수 있습니다. 기본값은 100입니다.
- **수렴 허용 오차** : 매개변수 수렴을 위한 임계값(계수 벡터 노름의 절대 변화량)을 설정합니다.

  매개변수 수렴 기준값. 기본값은 1E-6 입니다. 계수의 절대 변화량이 이 값보다 작아지면 알고리즘이 중단됩니다. 출판 수준의 정밀도를 원하시면 1E-8 를, 더 빠른 탐색적 분석을 원하시면 1E-4 를 사용하세요.

  1E-8 과 1E-3 사이의 숫자 값을 입력하세요.

  이 필드는 과학적 표기법 형식을 지원합니다: 1E-6, 1e-6, 0.000001
- **특이점 허용치** : 특이 헤세이안 행렬을 감지하기 위한 임계값(행렬식 임계값)을 설정합니다. 특이 헤시안 행렬을 탐지하기 위한 임계값. 기본값은 1E-12 입니다. 헤세이안 행렬의 행렬식이 이 값보다 낮아지면 알고리즘은 준특이점을 감지합니다. 1E-10 를 사용하여 R/Stata 결과와 일치시키세요. 싱귤래리티 내성에 대한 다음 전제 조건을 확인하십시오:
  - 값이 작을수록 더 보수적인 설정입니다(허위 특이점 탐지 건수가 줄어듭니다).
  - 1E-14 과 1E-6 사이의 숫자를 입력하세요
  - 이 필드는 과학적 표기법 형식을 지원합니다: 1E-12, 1e-12, 0.000000000001
- **최대 단계 절반** :

  이 필드는 반복당 최대 단계 반감 시도 횟수를 나타냅니다. 기본값은 5입니다. 단계적 반감법은 각 반복 단계마다 로그우도 함수가 개선되도록 보장합니다. 단계 절반 조정 실패가 발생하면 이 값을 10~20으로 늘리십시오. 값은 1에서 20 사이의 정수여야 합니다.
- **Ridge 페널티** : 특이 헤시안 행렬에 대한 Ridge 정규화 강도를 설정합니다.

  기본값은 1E-6 입니다. 헤시안이 거의 특이 상태일 때(행렬식의 값이 특이성 허용 오차보다 작을 때), 행렬 역행렬 구성을 안정화하기 위해 이 페널티가 적용됩니다. 상관성이 매우 높은 설명변수에는 더 큰 값 (1E-4, 1E-3)을 사용하십시오.

  1E-8 과 1E-3 사이의 숫자 값을 입력하세요. 이 입력란은 과학적 표기법 형식도 지원합니다: 1E-6, 1e-6, 0.000001

인쇄

'인쇄' 탭에서 옵션을 설정하여, 사용자가 선택한 관련 구성 요소만 포함되도록 출력을 맞춤 설정할 수 있습니다.

분석 결과의 일부로 해당 출력 섹션을 생성하려면 다음 옵션 중 하나를 선택하십시오.

중개 회귀 모델 요약

이 옵션은 기본 제공 옵션으로 설정되어 있으며 변경할 수 없습니다. 출력 결과에는 매개 회귀 모델의 결과를 요약한 표가 생성됩니다. 여기에는 추정 계수(β0, β1, β2), 각 계수의 표준 오차 및 통계적 유의 수준이 포함되어 있습니다. 이러한 결과는 예측 변수(*A*), 매개 변수(*M*), 공변량(*C*) 간의 관계를 이해하는 데 도움이 됩니다. 이 옵션은 기본적으로 선택됩니다.

결과 회귀 모델 요약
이 옵션은 기본 제공 옵션으로 설정되어 있으며 변경할 수 없습니다. 출력 결과에는 결과 회귀 모델의 결과를 요약한 표가 생성됩니다. 이 결과는 추정 계수(θ0, θ1, θ2, θ3, θ4), 각 계수의 표준 오차 및 유의 수준을 제공합니다. 이 결과는 예측변수, 매개변수 및 공변인이 종속변수(*Y*)에 미치는 영향을 평가합니다. 이 옵션은 기본적으로 선택됩니다.

노출 레벨 비교 통계
이 옵션은 기본 제공 옵션으로 설정되어 있으며 변경할 수 없습니다. 출력 결과에는 예측 변수의 지정된 노출 수준에 따른 예측 결과의 차이를 보여주는 섹션이 생성됩니다. 이 통계에는 통제된 직접 효과(*CDE*), 자연적 직접 효과(*NDE*), 자연적 간접 효과(*NIE*)와 같은 추정치가 포함되어 있어, 노출 변수의 변화가 매개변수와 결과에 어떤 영향을 미치는지 파악할 수 있게 해줍니다.

분해 분석 결과
이 옵션을 선택하면 총 효과(*TE*)를 직접 효과, 간접 효과, 공변량 효과(*CDE*), 비직접 효과(*NDE*), 비간접 효과 (*NIE*) 등으로 분해하여 보여주는 표를 생성할 수 있습니다. 또한 매개된 비율(*PM*) 및 상호작용 효과와 같은 추가적인 구성 요소들도 보고합니다.

상호작용 효과 분석
매개변수-결과 변수 간 상호작용 및 예측변수-매개변수 간 상호작용의 기여도를 측정하는 *INTmed* 및 *INTref* 와 같은 상호작용 효과를 분석하는 섹션을 생성하려면 이 옵션을 선택하십시오. 이러한 결과는 상호작용이 결과에 어떤 영향을 미치는지 이해하는 데 도움이 된다.

비율 통계량
r(CDE), *r(NDE)*, *r(NIE)*, *r(INTmed)* 와 같은 비율 통계를 생성하려면 이 옵션을 선택하십시오. 이는 각 구성 요소에 기인하는 전체 효과(*TE*)의 비율을 백분율로 나타낸 것입니다.

모수 추정값에 대한 공분산 행렬
이 옵션은 기본적으로 선택됩니다. 매개변수 추정치 간의 공분산을 나타내는 행렬을 생성할 수 있습니다. 예를 들어, β (매개변수 회귀 분석에서 도출된) 값과 θ (결과 변수 회귀 분석에서 도출된) 값 간의 공분산 등이 있습니다. 델타법을 사용하여 유도된 통계량의 표준오차를 계산하는 데 있어 이는 매우 중요합니다. 이 옵션은 기본적으로 선택됩니다.

케이스 처리 요약
이 옵션은 기본적으로 선택됩니다. 분석에 사용된 데이터에 대한 개요를 제공하는 요약 보고서를 생성할 수 있으며, 여기에는 사례 수, 제외된 사례 수 및 적용된 가중치 등이 포함됩니다. 이는 데이터 세트의 완전성과 품질을 평가하는 데 도움이 됩니다. 이 옵션은 기본적으로 선택됩니다.

기술통계량(D)
분석에 포함된 모든 변수에 대한 기술 통계량(평균, 표준편차, 최소값, 최대값 등)을 보고하는 섹션을 생성하려면 이 옵션을 선택하십시오. 이는 데이터의 분포와 변동성을 한눈에 보여줍니다.

도표

'그래프' 탭에서는 매개변수, 예측변수 및 결과변수 간의 관계를 설명하는 데 도움이 되는 시각화 자료를 생성하기 위한 옵션을 설정할 수 있습니다.

경로 다이어그램, 산점도, 막대 그래프, 포레스트 플롯 등 네 가지 유형의 시각화 옵션을 정의할 수 있습니다. 각 플롯 유형은 매개 분석 모델에 대한 서로 다른 관점을 제공하며, 분석 목적에 맞게 조정할 수 있습니다.

효과 크기, 신뢰 구간 또는 하위 그룹별 분석 결과를 추가하여 각 그래프를 사용자 지정할 수 있습니다.

경로 다이어그램
중재 변수, 예측 변수 및 결과 변수 간의 관계를 시각화하십시오. 경로도는 직접적 효과와 간접적 효과를 포함한 전체 매개 모델을 제시하는 데 유용합니다.

- **공변량 효과 숨기기** : 공변량과 관련된 모든 에지를 숨깁니다. 예측변수-매개변수, 매개변수-결과변수, 예측변수-결과변수 경로는 계속 표시됩니다.
- **가장자리 색상** : 이 옵션을 선택하면 가장자리가 효과 유형(예: 직접 효과, 간접 효과, 공변량 효과)을 나타내는 색상으로 구분됩니다. 녹색은 효과가 더 강한 것을 나타내고, 회색은 효과가 더 약한 것을 나타냅니다. 이 옵션을 선택하지 않으면 기본 테두리 색상이 균일하게 적용됩니다.

산점도
산점도는 변수 간의 관계를 보여주고, 상관관계와 분포 패턴을 파악하는 데 도움을 줍니다.

중재 분석에서 산점도는 예측 변수(X), 중재 변수(M), 결과 변수(Y) 간의 관계를 시각화하는 데 도움이 됩니다. 이러한 관계는 다음과 같은 매개 경로와 일치합니다:

- 경로 a: 예측변수(X)가 매개변수(M)에 미치는 영향
- 경로 b: 매변인(M)이 결과변수(Y)에 미치는 영향
- 경로 c′: 예측 변수(X)가 결과 변수(Y)에 미치는 직접적 영향

3변수 산점도는 예측 변수가 매개 변수와 결과 변수 모두에 어떻게 영향을 미치는지를 보여주며, 이를 통해 예측 변수와 결과 변수 간의 관계에서 매개 변수가 수행하는 역할을 설명해 줍니다. 일반적으로 시각화에는 다음이 포함됩니다:

- 예측 변수와 매개 변수 간의 관계를 나타내는 데이터 점 (경로 a)
- 중재변과 결과변 간의 관계를 보여주는 데이터 점들 (경로 b)
- 예측 변수와 결과 변수 간의 전반적인 관계로, 직접 효과와 간접 효과(경로 c′)를 모두 포함한다

산점도를 사용하면 이러한 관계의 강도, 방향 및 유의성을 시각적으로 파악할 수 있습니다. 일반적으로 축은 매개변수와 결과 변수를 나타내며, 예측 변수는 색상 그라데이션이나 마커를 사용하여 표시됩니다.

막대 플롯

범주형 변수의 비율이나 개수를 표시하며, 매개변수와 결과 변수의 수준을 시각화하는 데 유용합니다. 이 옵션은 선택한 매개변수 모델에 막대 그래프가 필요한 경우에만 사용할 수 있습니다(예: 이분형 매개변수, 이분형 결과 변수, 또는 이분형 결과 변수를 갖는 연속형 매개변수).

이 옵션은 이분형 매개변수, 이분형 결과 변수 또는 이분형 결과 변수를 갖는 연속형 매개변수 변형에 대해 활성화됩니다.

이 옵션은 연속형 매개변수 및 연속형 결과 변수에 대해서는 비활성화되며, 대신 산점도가 사용됩니다.

- **오차 막대** : 막대 차트에 오차 막대를 추가합니다. 그룹화된 막대 그래프에서 오차 막대는 각 매개변수 수준에 대한 비율(Y = 1 또는 Y = 0)의 불확실성을 나타냅니다.
- **신뢰 구간** : 오차 막대의 신뢰 구간을 표시합니다. 이는 **[기준]** 탭의 [**신뢰 구간 수준(%**)] 필드에 입력한 값에 따라 결정됩니다. 신뢰구간의 신뢰수치는 기본적으로 95%입니다.

포리스트 도표

매개 효과의 크기와 불확실성을 평가하는 데 도움이 되는 경로 계수와 그 신뢰 구간을 표시하십시오.

포레스트 플롯은 매개 경로(예: a, b, c′)에 대한 효과 크기와 관련 신뢰 구간을 함께 표시합니다.

옵션:

- **경로 계수** : 각 매개 경로에 대한 경로 계수(a, b, c′)를 표시합니다.
- **신뢰 구간** : 각 경로 계수에 대한 신뢰 구간을 표시합니다. 이는 **[기준]** 탭의 [**신뢰 구간 수준(%**)] 필드에 입력한 값에 따라 결정됩니다. 신뢰구간의 신뢰수치는 기본적으로 95%입니다.

시계열 필터

시계열 분석에서는 정해진 범위를 벗어나 시계열에서 비정상적인 변동을 일으키는 값을 제거하는 것이 필수적입니다. 좋은 필터는 단위 루트와 주기적 구성 요소를 제거할 수 있거나 특정 주파수에서 데이터의 변동을 분리할 수 있어야 합니다.

시계열 필터링은 계량경제학 및 시계열 분석에서 중요한 기술로, 시계열을 추세 및 주기적 구성 요소로 분해하는 데 자주 사용됩니다. 이는 비즈니스 사이클의 변동이나 장기적인 경제 동향을 파악하는 것이 필수적인 거시경제 및 금융 분야에서 특히 중요합니다.

널리 사용되는 시계열 필터는 호드릭-프레스콧(HP) 필터, 박스터-킹(BK) 필터, 크리스티아노-피츠제럴드(CF) 필터입니다.

시계열 필터 가져오기

시계열 필터는 Python 플러그인을 통해 확장 절차로 사용할 수 있습니다. 따라서 절차를 실행하려면 확장 모듈을 설치하세요.

시계열 필터에 대한 절차를 실행하려면 다음 단계를 완료하세요.

1. **데이터 정의** > **날짜 및 시간으로** 이동하여 날짜 변수를 생성합니다.
2. 날짜와 시간을 정의한 후 **예측 시계열** > **필터(TSF)** > **분석을** 클릭합니다.
3. 데이터 집합에 적용할 시계열 필터를 선택합니다.

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
4. 변수를 선택하고 적절한 시계열 필터를 위한 변수 목록으로 이동합니다.
5. 호드릭-프레스콧(HP) 필터의 경우 **람다** 값을 설정하거나 박스터-킹(BK) 필터 및 크리스티아노-피츠제럴드(CF) 필터의 경우 **낮음**, **높음** 및 **K** 값을 설정합니다.

   선택적으로 크리스티아노 피츠제럴드(CF) 필터의 **드리프트를** 선택할 수 있습니다.
6. **실행**을 클릭합니다.

시계열 필터: 옵션

시계열 필터 대화 상자에는 다음과 같은 옵션이 있습니다.

중요사항: 옵션을 설정하기 전에 **데이터 정의** > **날짜 및 시간으로** 이동하여 날짜 변수를 생성합니다.

Hodrick-Prescott(HP)

호드릭-프레스콧(HP) 필터는 거시경제학에서 시계열에 부드러운 추세를 맞추기 위해 사용하는 도구입니다.

호드릭-프레스콧 필터는 시계열 $_{yt를}$ 추세 성분 $_{\tau t와}$ 주기적 성분 $_{ct로}$ 분리합니다.

$y_t = \tau_t + c_t + \boxtimes_t$여기서, $_{\boxtimes는}$ 오류(또는 노이즈) 구성 요소입니다.

구성 요소는 이차 손실 함수를 최소화하여 결정되며 HP 필터는 공식에 의해 명시적으로 제공됩니다,

$HP = [\ ^{\lambda L2-}\ 4\lambda L+ (1 + 6\lambda)-\ ^{4\lambda L-1}\ +\ ^{\lambda L-2}\ ]^{-1}$

여기서 λ는 평활화 매개변수(양수 값)를 나타내고 L은 최소화 문제의 1차 조건에서 도출되는 지연 연산자를 나타냅니다.

변수 목록에서 HP 필터에 사용할 변수를 하나 선택하고 **람다** 값을 설정합니다.

스무딩 파라미터 λ는 핏과 부드러움 사이의 균형을 결정합니다. Λ 값이 클수록 페널티가 높아집니다. 분기별 데이터를 분석하는 경우 λ에 기본값인 1600을 사용하는 것이 좋습니다. 또한 λ는 연간 데이터의 경우 $^{1600/44}$ = 6.25, 월간 데이터의 경우 1600 * ^{34} = 129,600이 되어야 합니다. Λ= 0이면 HP 필터 계열의 출력이 입력 계열과 정확히 일치한다는 의미입니다. 반대로 λ 값이 높을수록 출력 시리즈가 더 부드러워집니다.

Baxter-King(BK)

박스터-킹(BK) 은 거시경제 연구에서 데이터의 주기적 구성 요소를 분리하기 위해 시계열 분석에 일반적으로 사용되는 대역 통과 필터입니다. 목표는 일반적으로 1.5 ~ 8년 사이의 중기 관심 주기에 초점을 맞춰 고주파(단기) 노이즈와 저주파(장기 추세) 성분을 제거하는 것입니다. BK 필터는 기본 시계열의 안정성을 가정합니다. 변수 목록에서 BK 필터에 사용할 변수를 1개 또는 2개 선택합니다.

낮음 (최소 주기), **높음** (최대 주기) 및 **K** (리드-래그 길이)에서 적절한 값을 설정합니다.

Christiano-Fitzgerald(CF)

크리스티아노 피츠제럴드(CF) 필터는 시계열을 추세 및 주기적 구성 요소로 분리하는 데 사용되는 또 다른 대역 통과 필터입니다. 이 필터는 이상적인 밴드패스 필터를 기반으로 작동하는 박스터 킹(BK) 필터와 유사하지만, BK 필터는 밴드패스 필터를 근사화하기 위해 대칭 이동 평균을 가정하지만 CF 필터는 시계열이 드리프트 없이 무작위성을 따른다고 가정합니다. CF 필터링은 데이터가 랜덤 워크 또는 ARIMA 유형의 프로세스에 의해 생성된다는 가정 하에 최적의 필터를 추정하여 구현됩니다.

변수 목록에서 CF 필터에 사용할 변수를 1개 또는 2개 선택합니다. **낮음**, **높음**, **K에서** 적절한 값을 설정하고 CF 필터에 필요한 경우 **드리프트를** 선택합니다.

람다, **낮음**, **높음은** 모든 플로트 값을 사용할 수 있습니다. **K의** 값은 정수여야 합니다. 다음 값을 참조하세요:

낮음	1.5 (연간) 및 6(분기별)
높음	8(연간) 및 32(분기별)
k	3(연간) 및 12(분기별)

곡선 추정

곡선 추정 프로시저로 11가지의 서로 다른 곡선 추정 회귀 모형에 대해 곡선 추정 회귀 통계와 관련 도표를 작성할 수 있습니다. 각 종속변수마다 별도의 모형이 작성됩니다. 예측값, 잔차, 예측 구간을 새로운 변수로 저장할 수 있습니다.

예. 인터넷 서비스 공급자가 해당 네트워크에서 바이러스에 감염된 채로 전달되는 이메일의 트래픽 퍼센트를 시간대별로 추적하고 있습니다. 산점도를 사용하여 그 관계가 비선형임을 밝혀냈습니다. 데이터에 2차 회귀 또는 3차 회귀 모형을 적용하고 모형의 가정과 적합도를 확인할 수 있습니다.

통계. 각 모형에 대해서는 회귀계수, 다중 R, R^2, 수정된 R^2, 추정값의 표준 오차, 분산 분석표, 예측값, 잔차, 예측 구간을 선택할 수 있습니다. 모형에 대해 선형, 로그, 역, 2차, 3차, 거듭제곱, 복합, S-곡선, 로지스틱, 성장, 지수를 선택할 수 있습니다.

곡선 추정 데이터 고려 사항

데이터. 종속변수와 독립변수는 양적변수이어야 합니다. 변수를 선택하는 대신 독립변수로 활성 데이터 세트의 **시간**을 선택하면 곡선 추정 프로시저를 통해 케이스 간 시간 길이가 균일한 시간변수가 작성됩니다. **시간**을 선택하면 종속변수는 시계열 측도가 됩니다. 시계열 분석에서는 데이터 파일 구조가 있어야 하는데 이 구조에서는 각 케이스(행)가 다른 시간의 관측값 세트를 나타내며 케이스 간 시간 길이가 균일합니다.

가정. 데이터를 면밀히 검토하여 독립변수와 종속변수가 관련된 방식(선형, 지수형 등)을 판별합니다. 적합한 모형의 잔차는 무작위로 분포되어있으며 정규적입니다. 선형 모형이 사용된 경우 다음의 가정이 충족되어야 합니다. 각 독립변수 값에서 종속변수의 분포는 정규 분포이어야 합니다. 종속변수 분포의 분산은 모든 독립변수 값에 대해 항상 일정해야 합니다. 종속변수와 독립변수 간 관계는 선형이어야 하고 관측값은 모두 독립적이어야 합니다.

곡선 추정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 회귀분석 > 곡선 추정...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 종속변수를 선택합니다. 각 종속변수마다 별도의 모형이 작성됩니다.
3. 독립변수(활성 데이터 세트의 변수 또는 **시간**)를 선택합니다.
4. 또는 다음 작업을 수행합니다.

- 산점도에서 케이스 레이블에 대한 변수를 선택합니다. 산점도의 각 점에서 점 선택 도구를 사용하여 케이스 레이블 변수 값을 표시할 수 있습니다.
- 예측값, 잔차, 예측 구간을 새로운 변수로 저장하려면 **저장**을 클릭합니다.

다음 옵션을 사용할 수도 있습니다.

- **방정식에 상수항 포함.** 회귀 방정식에서 상수항을 추정합니다. 상수가 기본값으로 포함됩니다.
- **모형 적합 도표화.** 종속변수 값과 독립변수에 대해 선택한 각 모형을 도표화합니다. 각 종속변수마다 별도의 도표가 작성됩니다.
- **분산 분석표 출력.** 선택한 각 모형에 대한 요약 분산 분석표를 나타냅니다.

곡선 추정 모형

하나 이상의 곡선 추정 회귀 모형을 선택할 수 있습니다. 사용할 모형을 결정하려면 데이터를 도표화해야 합니다. 변수가 선형으로 연결되어 있을 때는 단순 선형회귀 모형을 사용합니다. 변수가 선형으로 연결되어 있지 않을 때는 데이터를 변환하도록 합니다. 변환 작업으로도 원하는 결과를 얻을 수 없을 때는 더 복잡한 모형이 필요합니다. 데이터의 산점도를 보고 도표가 수학적 함수 모양일 때는 데이터를 해당 모형의 유형에 맞춥니다. 예를 들어, 데이터가 지수 함수와 비슷하면 지수모형을 사용합니다.

선형. 모형식은 Y = b0 +(b1 * t) 입니다. 계열 값은 시간의 선형 함수로 모형화됩니다.

로그 기록. 방정식이 Y = b0 + (b1 * ln(t))인 모델입니다.

역. 모형식은 Y = b0 + (b1 / t)입니다.

*2*차. 모형식은 Y = b0 +(b1 * t)+(b2 * t**2) 입니다. 이 2차 모형은 "줄어드는" 계열 또는 감소하는 계열을 모형화하는 데 사용할 수 있습니다.

큐빅. 모형식은 Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3)입니다.

전원. 모형식은 Y = b0 * (t**b1) 또는 ln(Y) = ln(b0) + (b1 * ln(t))입니다.

복합. 모형식은 Y = b0 * (b1**t) 또는 ln(Y) = ln(b0) + (ln(b1) * t)입니다.

S 곡선. 모형식은 Y = e**(b0 + (b1/t)) 또는 ln(Y) = b0 + (b1/t)입니다.

로지스틱. 모델식은 Y = 1/(1/u +(b0 * (b1**t))) 또는 ln (1/y-1/u) = ln (b0)+(ln (b1) * t) 입니다. 여기서 u는 상한 값입니다. 로지스틱을 선택한 후 상한값을 지정하여 회귀 방정식에서 사용합니다. 이 값은 양수여야 하며 최대 종속변수 값보다 커야 합니다.

증가. 모형식은 Y = e**(b0 + (b1 * t)) 또는 ln(Y) = b0 + (b1 * t)입니다.

지수. 모형식은 Y = b0 * (e**(b1 * t)) 또는 ln(Y) = ln(b0) + (b1 * t)입니다.

곡선 추정: 저장

저장할 변수. 선택한 각 모형에 대해 예측값과 잔차(종속변수의 예측값에서 모형 예측값을 뺀 값), 예측 구간(상한과 하한) 등을 저장할 수 있습니다. 새로운 변수 이름과 레이블 레이블이 결과 창에 테이블로 나타납니다.

예측할 케이스. 활성 데이터 세트에서 변수 대신 **시간**을 독립 변수로 선택하면 시계열의 마지막을 초과하는 예측 기간을 지정할 수 있습니다. 다음 중 하나를 선택합니다.

- **추정 기간부터 마지막 케이스까지 예측.** 추정 기간의 케이스에 따라 파일에서 모든 케이스의 값을 예측합니다. 추정 기간은 대화 상자의 아래에 표시되는데 데이터 메뉴의 케이스 선택 옵션을 클릭하여 나타나는 범위 하위 대화 상자에서 지정할 수 있습니다. 추정 기간을 지정하지 않으면 값 예측에 모든 케이스를 사용합니다.
- **예측 케이스.** 추정 기간의 케이스에 따라 지정한 날짜, 시간, 관측값을 예측합니다. 이 기능은 시계열의 마지막 케이스 이후의 값을 예측할 때 사용되며 현재 정의된 데이터 변수는 예측 기간의 종료를 지정하는 데 어떤 입력란을 사용할 수 있는지 결정합니다. 날짜 변수가 정의되어 있지 않으면 마지막 관측(케이스) 번호를 지정할 수 있습니다.

데이터 메뉴에서 날짜 정의 옵션을 사용하여 날짜 변수를 작성할 수 있습니다.

편최소제곱회귀

PLS(Partial Least Squares) 회귀분석 프로시저는 부분 최소제곱(PLS, "projection to latent structure"라고도 함) 회귀 모형을 예측합니다. PLS는 보통최소제곱(OLS) 회귀, 정준상관 또는 구조적 방정식 모형의 대안인 예측 기법이며 예측자가 서로 밀접하게 상호 관련되어 있거나 예측자 변수 수가 케이스 수를 초과할 때 특히 유용합니다.

PLS는 주성분분석과 다중 회귀분석의 기능을 결합합니다. 먼저 독립변수와 종속변수 사이에 가능한 많은 공분산으로 설명되는 잠재 요인을 추출합니다. 그런 다음 회귀분석 단계는 독립변수의 분해를 사용하여 종속변수의 값을 예측합니다.

테이블
(잠재 요인으로)설명되는 분산의 비율, 잠재 요인 가중치, 잠재 요인 로딩, 투영에서 독립변수 중요도(VIP) 및 (종속변수에 의한) 회귀 모수 추정값은 모두 기본적으로 생성됩니다.

차트
투영에서 변수 중요도(VIP), 요인 스코어, 처음 세 잠재 요인의 요인 가중치 및 모형까지의 거리는 모두 옵션 탭에서 생성됩니다.

데이터 고려사항

측정 수준
종속 및 독립(예측)변수는 척도, 명목 또는 순서가 될 수 있습니다. 소스 변수 목록에서 변수를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 측정 수준을 선택하여 변수에 대한 측정 수준을 임시로 변경할 수 있지만 프로시저는 적절한 측정 수준이 모든 변수에 지정되었다고 가정합니다. 범주형(명목 또는 순서) 변수는 프로시저에 의해 동일하게 취급됩니다.

범주형 변수 코딩
프로시저는 임시로 프로시저 기간 동안 *c* 코딩 중 하나를 사용하여 범주형 종속변수를 다시 코딩합니다. 변수의 *c* 범주가 있는 경우 변수는 *c* 벡터로 저장되며 첫 번째 범주는 (1,0, ..., 0), 다음 범주는 (0,1,0, ..., 0), ..., 및 최종 범주 (0,0, ..., 0, 1). 범주형 종속변수는 더미 코딩을 사용하여 나타냅니다. 즉, 참조 범주에 해당하는 지표를 생략합니다.

빈도 가중값
가중치는 사용하기 전에 가장 가까운 정수로 반올림됩니다. 결측 가중치나 0.5 미만의 가중치를 갖는 케이스는 분석에 사용되지 않습니다.

누락 값
사용자 및 시스템-결측값을 유효하지 않은 값으로 처리함

재척도
모든 모형 변수는 범주형 변수를 나타내는 지시변수를 포함하여 가운데 위치하고 표준화됩니다.

편최소제곱회귀 구하기

메뉴에서 다음을 선택합니다.

분석 > **회귀분석** > **PLS(Partial Least Squares)...**

1. 종속 변수를 하나 이상 선택합니다.
2. 독립변수를 한 개 이상 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 범주형(명목 또는 순서) 종속변수의 참조 범주를 지정합니다.
- 케이스별 결과와 저장된 데이터 세트를 위한 고유 식별자로 사용할 변수를 지정합니다.
- 추출할 잠재 요인 수의 상한을 지정합니다.

모형

모형 효과 지정. 주효과 모형은 모든 요인과 공변량 주효과를 포함합니다. **사용자 정의**를 선택하여 상호작용을 지정합니다. 모형에 포함할 모든 항을 지정해야 합니다.

요인 및 공변량. 요인 및 공변량이 나열됩니다.

모형. 모형은 데이터의 특성에 따라 다릅니다. **사용자 정의**를 선택한 다음 분석시 중요한 주효과와 상호작용을 선택할 수 있습니다.

항 설정

선택된 요인과 공변량에 대해 다음과 같습니다.

상호작용. 선택한 모든 변수의 최고-수준 상호작용 항을 생성합니다. 이는 기본값입니다.

주효과. 선택한 각 변수에 대한 주효과 항을 생성합니다.

모든 2원배치. 선택한 변수에 대해 가능한 이원 상호작용을 모두 생성합니다.

모든 3원배치. 선택한 변수에 대해 가능한 3원배치 상호작용을 모두 생성합니다.

모든 4원배치. 선택한 변수에 대해 가능한 4원배치 상호작용을 모두 생성합니다.

모든 5원배치 선택한 변수에 대해 가능한 5원배치 상호작용을 모두 생성합니다.

옵션

옵션 탭을 사용하면 개별 케이스, 잠재 요인 및 예측자에 대한 모형 추정값을 저장하고 도표화할 수 있습니다.

각 데이터 유형에 대해 데이터 세트 이름을 지정합니다. 데이터 세트 이름은 유일해야 합니다. 기존 데이터 세트 이름을 지정하는 경우 내용이 바뀌고, 그렇지 않으면 새로운 데이터 세트가 작성됩니다.

- **개별 케이스에 대한 추정값을 저장합니다.** 예측값, 잔차, 잠재 요인 모형에 대한 거리, 잠재 요인 스코어와 같은 케이스별 모형 추정값을 저장합니다. 잠재 요인점수도 도표화합니다.

- **잠재 요인에 대한 추정값을 저장합니다.** 잠재 요인 로딩과 잠재 요인 가중치를 저장합니다. 잠재 요인 가중값도 도표화합니다.
- **독립변수에 대한 추정값을 저장합니다.** 회귀 모수 추정값과 투영에서 변수 중요도(VIP)를 저장합니다. 잠재 요인별로 VIP도 도표화합니다.

최근접 이웃 분석

최근접 이웃 분석은 다른 케이스와의 유사성을 기준으로 케이스를 분류하는 방법입니다. 기계 학습에서 이 분석 방법은 저장된 모든 패턴이나 케이스와 정확히 일치할 필요가 없는 데이터 패턴을 인식하는 방법으로 개발되었습니다. 유사한 케이스들은 서로 가깝게 있으며 유사하지 않은 케이스들은 서로 멀리 떨어져 있습니다. 따라서 두 케이스 사이의 거리는 두 케이스의 상이성 측도가 됩니다.

서로 가깝게 있는 케이스를 "이웃"이라고 합니다. 새 케이스(검증용)가 있는 경우, 해당 모형에서 각 케이스와의 거리가 계산됩니다. 가장 유사한 케이스 분류인 (최근접 이웃)이 기록되고 가장 많은 수의 최근접 이웃이 포함되어 있는 범주에 새 케이스가 놓입니다.

조사할 최근접 항목의 수를 지정할 수 있습니다. 이 값은 *k*입니다.

또한 최근접 이웃 분석은 연속적인 대상 값을 계산하는 데 사용할 수 있습니다. 이 경우, 최근접 이웃의 평균 또는 중위수 대상 값이 사용되어 새 케이스의 예측값을 가져옵니다.

최근접 이웃 분석 고려 사항

대상 및 기능. 대상 및 기능은 다음과 같습니다.

- 명목*(Nominal)*. 변수의 값이 고유한 순위가 없는 범주 (예: 직원이 근무하는 회사의 부서) 를 나타내는 경우 변수를 명목으로 처리할 수 있습니다. 종교, 우편번호 또는 종교 단체 등이 명목변수에 해당합니다.
- 순서*(Ordinal)*. 변수의 값이 고유한 순위가 있는 범주를 나타낼 때 변수를 순서로 처리할 수 있습니다 (예: 매우 불만족에서 매우 만족까지의 서비스 만족도 수준). 순서변수의 예로는 만족도나 신뢰도를 나타내는 태도 스코어 및 선호도 등급 스코어가 있습니다.
- 척도*(scale)*. 변수의 값이 의미 있는 메트릭으로 정렬된 범주를 나타내는 경우 값 간 거리 비교가 적합하도록 변수를 척도 (연속형) 로 처리할 수 있습니다. 척도변수의 예로는 연령과 수입이 있습니다.

  명목 및 순서변수는 최근접 이웃 분석에 의해 동일하게 취급됩니다. 소스 변수 목록에서 변수를 마우스 오른쪽 단추로 클릭하고 팝업 메뉴에서 측정 수준을 선택하여 변수에 대한 측정 수준을 임시로 변경할 수 있지만 프로시저는 적절한 측정 수준이 각 변수에 지정되었다고 가정합니다.

변수 목록의 각 변수 옆에 있는 아이콘은 측정 수준 및 데이터 유형을 식별합니다.

표 *3.* 측정 수준 아이콘

	숫자	문자열	날짜	시간
척도(연속형)		n/a		
순서형				
명목형				

범주형 변수 코딩. 이 프로시저는 프로시저 기간 동안 one-of-*c* 코딩을 사용하여 범주형 예측 변수 및 종속 변수를 임시로 코딩변경합니다. 변수의 카테고리가 *C*인 경우 변수는 첫 번째 카테고리는 (1,0,...,0), 다음 카테고리는 (0,1,0,...,0),.....으로 표시된 C 벡터로 저장됩니다, 그리고 최종 카테고리(0,0,...,0,1).

이 코딩 체계는 기능 공간의 차원을 증가시킵니다. 특히 총 차원 수는 모든 범주형 예측자에 대한 척도 예측자와 수 범주의 수를 합한 값입니다. 그 결과 이 코딩 체계로 인해 훈련이 늦어지게 됩니다. 최근접 이웃 훈련이 매우 느리게 진행될 경우 프로시저를 진행하기 전에 유사한 범주를 조합하거나 거의 사용하지 않는 범주를 가진 케이스를 제외하여 범주형 예측자에서 범주의 수를 줄일 수 있습니다.

검증용 표본이 정의되어 있다 해도 모든 *c* 코딩 중 하나는 훈련 데이터를 기준으로 합니다(174 페이지의 『파티션』 참조). 따라서 검증용 표본에는 훈련 데이터에 나타나지 않는 예측자 범주를 가진 케이스가 포함되어 있는 경우 이러한 케이스에 스코어가 주어지지 않습니다. 검증용 표본에 훈련 데이터에 나타나지 않는 종속변수 범주를 가진 케이스가 포함되어 있는 경우에는 이러한 케이스에 스코어가 주어집니다.

척도 조정. 기본값으로 정규화됩니다. 검증용 표본이 정의되어 있다 해도 모든 척도 조정은 훈련 데이터를 기준으로 합니다(174 페이지의 『파티션』 참조). 특정 변수를 지정하여 파티션을 정의할 경우 기능에는 훈련 및 검증용.표본 전체에서 유사한 분포가 있어야 합니다. 예를 들어, 탐색 프로시저를 사용하여 파티션 전체의 분포를 조사합니다.

빈도 가중치. 빈도 가중치는 이 프로시저에서 무시됩니다.

결과 재현. 이 프로시저에서는 파티션의 임의 할당 및 교차 검증 중첩 과정에서 난수 생성을 사용합니다. 결과를 정확하게 복제하고 동일한 프로시저 설정을 사용하려면 Mersenne Twister에 대한 시드(174 페이지의 『파티션』 참조)를 설정하거나 변수를 사용하여 파티션 및 교차 검증 중첩을 정의하십시오.

최근접 이웃 분석을 얻으려면

메뉴에서 다음을 선택합니다.

분석 > 분류 > 최근접 이웃 항목...

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

1. 대상이 있을 경우 독립변수 또는 예측자로 간주할 수 있는 하나 이상의 기능을 지정하십시오.

   대상(선택적). 대상(종속변수 또는 반응변수)이 지정되지 않은 경우 프로시저는 최근접 이웃 *k*만 찾고 분류나 예측은 이루어지지 않습니다.

   척도 기능 정규화 정규화된 기능에는 추정 알고리즘 성능을 개선할 수 있는 동일한 범위의 값이 있습니다. 조정된 정규화, [2*(x최소)/(최대최소)]−1이 사용됩니다. 조정 정규화된 값은 -1과 1 사이의 값입니다.

   중심 케이스 식별자(선택적). 초점 케이스 식별자를 사용하면 원하는 케이스를 표시할 수 있습니다. 예를 들어, 조사자는 한 학교 구역의 시험 점수인 중심 케이스를 유사한 학교 구역과 비교할 수 있는지 여부를 결정하고 싶어 합니다. 그는 최근접 이웃 분석을 사용하여 주어진 기능군과 관련하여 가장 유사한 학교 구역을 찾습니다. 그런 다음 초점 학교 구역의 시험 점수를 최근접 이웃의 학교 구역과 비교합니다.

   중심 케이스는 또한 임상 연구에도 사용되어 임상 케이스와 유사한 제어 케이스를 선택할 수 있습니다. 중심 케이스는 최근접 이웃 (*k*) 및 거리 표, 기능 공간 도표, 피어 도표 및 4분위 맵에 표시됩니다. 중심 케이스에 대한 정보는 결과 탭에 지정된 파일에 저장됩니다.

   지정된 변수에 대한 양의 값을 가진 케이스는 중심 케이스로 처리됩니다. 이는 양의 정수가 없는 변수를 지정하는 데 유효한 값입니다.

케이스 레이블(선택적). 케이스는 기능 공간 도표, 피어 도표 및 4분위 맵에 있는 이러한 값을 사용하여 설명됩니다.

측정 수준을 알 수 없는 필드

측정 수준 경고는 데이터 세트에서 하나 이상의 변수(필드)에 대해 측정 수준을 알 수 없을 때 표시됩니다. 측정 수준은 이 프로시저의 계산 결과에 영향을 미치기 때문에 모든 변수에 정의된 측정 수준이 있어야 합니다.

데이터 스캔. 활성 데이터 세트의 데이터를 읽고 현재 알 수 없는 측정 수준이 있는 필드에 기본 측정 수준을 할당합니다. 데이터 세트가 큰 경우 시간이 걸릴 수 있습니다.

수동으로 할당. 알 수 없는 측정 수준이 있는 필드를 모두 나열하는 대화 상자를 엽니다. 이 대화 상자에서 해당 필드에 측정 수준을 할당할 수 있습니다. 데이터 편집기의 변수 보기에서도 측정 수준을 할당할 수 있습니다.

이 프로시저에 대해 측정 수준이 중요하기 때문에 모든 필드에 정의된 측정 수준이 있을 때까지는 대화 상자에 액세스하여 이 프로시저를 실행할 수 없습니다.

이웃

최근접 이웃(k) 수. 최근접 이웃 수를 지정합니다. 많은 수의 이웃을 사용한다고 해서 반드시 더 정확한 결과를 얻을 수 있는 것은 아닙니다.

변수 탭에 대상이 지정된 경우 다양한 범위의 값을 번갈아 지정할 수 있으며 프로시저는 해당 범위 내에서 "최적의" 항목 수를 선택할 수 있습니다. 최근접 이웃 수를 결정하는 방법은 기능 탭에 필드선택이 필요한지 여부에 따라 달라집니다.

- 필드선택이 유효한 경우 요청된 범위 내에서 각각의 *k* 값에 대해 필드선택이 수행되며 *k* 및 가장 낮은 오차 비율(또는 대상이 척도일 경우 가장 낮은 제곱합 오차)로 설정된 관련 기능이 선택됩니다.
- 필드선택이 유효하지 않은 경우 *V*-중첩 교차 검증을 사용하여 “최적의” 이웃 수를 선택합니다. 중첩 할당 제어용 파티션 탭을 참조하십시오.

거리 대입. 이는 케이스의 유사성 측도에 사용되는 거리 메트릭을 지정하는 데 사용되는 메트릭입니다.

- **Euclidean 메트릭.** 두 케이스(x와 y) 간 거리는 케이스 값 간 차이 제곱값 합(모든 차원에 대한)의 제곱근입니다.
- **도시 블록 메트릭.** 두 케이스 간 거리는 케이스 값 간 절대 차의 합(모든 차원에 대한)입니다. 이를 Manhattan 거리라고도 합니다.

선택적으로 변수 탭에 대상이 지정된 경우 거리 계산 시 정규화 중요도에 따라 가중치 기능을 선택할 수 있습니다. 예측자에 대한 기능 중요도는 오차 비율 또는 모형에서 오차 비율을 없앤 예측자를 가진 모형의 제곱합 또는 전체 모형에 대한 오차 제곱합으로 계산합니다. 정규화 중요도는 합이 1이 되도록 기능 중요도 값을 다시 부여하여 계산합니다.

척도 대상 예측 변수 탭에 척도 대상이 지정되면 척도 대상은 예측값을 평균을 기준으로 계산하는지, 최근접 이웃의 중위수를 기준으로 계산하는지 여부를 지정합니다.

변수

기능 탭을 사용하면 변수에 대상을 지정할 때 필드선택 옵션을 요청 및 지정할 수 있습니다. 기본적으로 모든 기능은 필드선택을 위해 준비되어 있지만 선택적으로 서브세트 기능군을 선택하여 모형에 적용할 수 있습니다.

중지 기준. 각 단계에서 가장 작은 오차(범주형 대상에 대한 오차 비율과 척도 대상에 대한 오차 제곱합으로 계산)를 보인 모형에 추가된 기능은 모형 세트에 포함됩니다. 전진 선택법은 지정 조건이 충족될 때까지 지속됩니다.

- **지정된 기능 수.** 알고리즘은 고정된 기능 수는 물론 모형에 적용된 기능 수를 추가합니다. 양의 정수를 지정합니다. 선택할 기능 수에 대한 값을 줄이면 경제적인 모형을 만들 수 있지만 중요한 기능이 빠질 위험이 있습니다. 선택할 기능 수에 대한 값을 증가시키면 중요한 기능을 모두 얻을 수 있지만 실제로 모형 오차를 증가시키는 기능을 추가하게 될 위험이 있습니다.
- **절대 오차 비율의 최소 변경.** 절대 오차 비율 변경이 기능을 더 추가해도 모형을 개선할 수 없음을 의미할 경우 알고리즘이 중지됩니다. 양수를 지정합니다. 최소 변경 값을 줄이면 기능을 추가할 수는 있지만 모형에 유용하지 않은 기능이 포함될 위험이 있습니다. 최소 변경 값을 증가시키면 많은 기능을 채택하지 않을 수 있지만 모형에 중요한 기능을 잃을 수 있는 위험이 있습니다. “최적의” 최소 변경은 사용자의 데이터 및 애플리케이션에 따라 달라질 수 있습니다. 어떤 기능이 가장 중요한 기능인지 판단하려면 결과의 필드선택 오차 로그를 참조하십시오. 자세한 정보는 178 페이지의 『변수 선택 오차 로그』 주제를 참조하십시오.

파티션

파티션 탭을 사용하여 데이터 세트를 훈련군과 검증군으로 나누고 케이스를 교차 검증 중첩에 할당(적용 가능할 경우)할 수 있습니다.

훈련 및 검증 파티션 이 그룹에서는 활성 데이터 세트를 훈련 및 검증용 표본으로 분할하는 방법을 지정합니다. **훈련 표본**은 최근접 이웃 모형을 훈련하는 데 사용되는 데이터 레코드로 구성되어 있으며 데이터 세트의 일부 케이스 퍼센트는 모형을 얻기 위해 훈련 표본에 할당해야 합니다. **검증용 표본**은 최종 모형 평가에 사용되는 독립적인 데이터 레코드 세트로, 모형 작성에 검증 케이스가 사용되지 않았기 때문에 검증용 표본 오차는 모형의 예측 기능에 대한 "진정한" 추정값을 제공합니다.

- **케이스를 파티션에 무작위로 할당합니다.** 훈련 표본에 할당할 케이스 퍼센트를 지정합니다. 나머지는 검증용 표본에 할당됩니다.
- **변수를 사용하여 케이스를 할당합니다.** 활성 데이터 세트에 있는 각 케이스를 훈련 및 검증용 표본에 할당하는 숫자 값을 지정합니다. 변수가 양의 값인 케이스는 훈련 표본에 할당되며 값이 0 또는 음의 값인 케이스는 검증용 표본에 할당됩니다. 시스템 결측값이 있는 케이스는 분석에서 제외됩니다. 분할 변수에 대한 모든 사용자 결측값은 항상 유효한 것으로 처리됩니다.

교차 검증 중첩. *V*-중첩 교차 검증은 "최적"의 이웃 수를 결정하는 데 사용됩니다. 교차 검증 중첩은 성능 문제로 인해 필드선택과 함께 사용할 수 없습니다.

교차 검증은 표본을 다수의 부표본 또는 중첩으로 나눕니다. 그런 다음 가장 가까운 모형을 생성하고 각 부표본에서 차례대로 데이터를 제외합니다. 첫 번째 모형은 첫 번째 표본 중첩의 케이스를 제외한 모든 케이스를 기준으로 하며, 두 번째 모형은 두 번째 표본 중첩의 케이스를 제외한 모든 케이스를 기준으로 하는 방식입니다. 각 모형의 경우 모형 생성에서 제외된 부표본에 적용하여 해당 모형의 오차를 추정합니다. "최적"의 가장 가까운 모형 수는 중첩 전체에서 가장 낮은 오차를 생성하는 수입니다.

- **케이스를 중첩에 무작위로 할당합니다.** 교차 검증에 사용할 중첩 수를 지정합니다. 프로시저에서는 케이스를 무작위로 중첩에 할당합니다. 중첩에는 1부터 *V*까지 번호가 매겨집니다.
- **변수를 사용하여 케이스를 할당합니다.** 활성 데이터 세트에 있는 각 케이스를 중첩에 할당하는 숫자 값을 지정합니다. 변수는 숫자여야 하며 1에서 *V*까지의 값을 사용해야 합니다. 이 범위의 값이 누락되고 분할 파일이 적용되는 경우 임의의 분할에서 오류가 발생합니다.

Mersenne Twister 시드를 설정합니다. 시드를 설정하면 분석을 복제할 수 있습니다. 이 제어를 사용하는 것은 Mersenne Twister를 활성 생성기로 설정하고 난수 생성기 대화 상자에 고정 시작점을 지정하는 것과 비슷합니다. 단, 이 대화 상자에 시드를을 설정하면 난수 생성기의 현재 상태가 유지되고 분석이 완료되면 해당 상태로 복원된다는 차이점이 있습니다.

저장

저장된 변수의 이름. 자동 이름 생성을 사용하면 모든 작업이 보존됩니다. 사용자 정의 이름을 사용하면 데이터 편집기에 저장된 변수를 먼저 삭제하지 않고 이전 실행의 결과를 취소/대체할 수 있습니다.

저장할 변수

- **예측값 또는 범주.** 척도 대상에 대한 예측값 또는 범주형 대상에 대한 예측 범주를 저장합니다.
- **예측확률.** 범주형 대상에 대한 예측확률을 저장합니다. 개별 변수는 각 범주의 첫 번째 *n* 범주에 각각 저장됩니다. *n* 범주는 **범주형 대상** 제어를 위해 저장할 최대 범주에 지정되어 있습니다.
- **훈련/검증용 분할변수.** 케이스가 파티션 탭의 훈련 및 검증용 표본에 무작위로 할당되면 파티션값(훈련 또는 검증용)이 케이스가 할당된 훈련 및 검증용 표본에 저장됩니다.
- **교차 검증 중첩 변수.** 케이스가 교차 검증 중첩에 무작위로 할당되면 중첩 값을 케이스가 할당된 교차 검증 중첩에 저장합니다.

결과

뷰어 출력결과

- **케이스 처리 요약.** 케이스 진행 요약 테이블을 표시합니다. 케이스 진행 요약 테이블에는 분석에서 포함한 케이스 수와 제외한 케이스 수가 합계 및 훈련 및 검증용 표본별로 요약되어 표시됩니다.
- **도표 및 표.** 테이블과 도표를 포함한 모형 관련 결과를 표시합니다. 모형 보기에 있는 테이블에는 중심 케이스, 범주형 반응변수 및 오차 요약에 대한 최근접 이웃(*k*)과 거리가 포함됩니다. 모형 보기의 그래픽 결과에는 선택 오차 로그, 기능 중요도 도표, 기능 공간 도표, 피어 도표 및 4분위 맵이 포함됩니다. 자세한 정보는 176 페이지의 『모형 보기』 주제를 참조하십시오.

파일

- **XML에 모형 내보내기.** 스코어링 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다. 분할 파일이 정의된 경우에는 이 옵션을 사용할 수 없습니다.
- **중심 케이스와 최근접 이웃(k) 간의 거리 내보내기.** 각 중심 케이스의 경우 각 중심 케이스의 최근접 이웃 *k*(훈련 표본에서)와 대응하는 가장 가까운 거리 *k*에 대한 개별 변수가 생성됩니다.

옵션

사용자 결측값. 케이스가 분석에 포함되려면 범주형 변수는 유효한 값을 가져야 합니다. 이 제어를 사용하면 범주형 변수 사이에서 사용자 결측값을 유효한 데이터로 취급할지 여부를 결정할 수 있습니다.

시스템 결측값 및 척도변수에 대한 결측값은 항상 유효한 것으로 처리됩니다.

모형 보기

결과 탭에서 **도표 및 표**를 선택하면 프로시저는 뷰어에서 최근접 이웃 모형 오브젝트를 생성합니다. 이 오브젝트를 활성화(두 번 클릭)하면 모형에 대한 대화형 보기를 얻을 수 있습니다. 모형 보기에는 2패널 창이 사용됩니다.

- 첫 번째 패널에서는 기본 보기라고 불리는 모형 개요가 표시됩니다.
- 두 번째 패널에서는 두 가지 보기 유형 중 하나가 표시됩니다.

  보조 모형 보기는 모형에 대한 자세한 정보를 보여줍니다. 단 모형 자체에 초점을 맞추지는 않습니다.

  연결된 보기는 사용자가 기본 보기 부분에서 드릴다운할 때 해당 모형의 특정 기능에 대한 자세한 내용을 보여줍니다.

기본적으로 첫 번째 패널은 기능 공간, 두 번째 패널은 변수 중요도 도표를 보여줍니다. 변수 중요도 도표를 사용할 수 없는 경우, 즉 기능 탭에서 **중요도별 가중치 기능**이 선택되지 않은 경우 보기 드롭다운에서 가장 먼저 볼 수 있는 보기가 표시됩니다.

보기에 사용 가능한 정보가 없을 경우 보기 드롭다운에 있는 해당 항목 텍스트가 비활성화됩니다.

기능 공간

기능 공간 도표는 기능 공간(또는 기능이 3개 이상 있을 경우, 하위 공간)의 대화형 그래프입니다. 각각의 축은 모형의 기능, 훈련 및 검증용 파티션의 케이스에 대한 기능값을 보여주는 도표 상의 점의 위치를 나타냅니다.

키. 도표에 있는 점은 기능값은 물론 또 다른 정보를 제공합니다.

- 형태는 점이 속해 있는 파티션(훈련 또는 검증)을 나타냅니다.
- 점의 색상/음영은 해당 케이스에 대한 대상값을 나타냅니다. 이 경우 개별 색상값은 범주형 대상의 범주를 나타내며 음영은 연속적인 대상의 범위값을 나타냅니다. 훈련 파티션에 대해 표시된 값은 관측값입니다. 검증용 파티션의 경우에는 예측값이 됩니다. 대상이 지정되지 않은 경우 이 기준이 표시되지 않습니다.
- 윤곽이 많다는 것은 케이스가 초점이라는 것을 의미합니다. 중심 케이스는 연결된 최근접 이웃 *k*에 나타납니다.

제어 및 상호작용성 도표에 있는 제어 수를 통해 기능 공간을 탐색할 수 있습니다.

- 도표에 나타나는 서브세트 기능군에서 특정 기능을 선택할 수 있으며 차원에 표시된 기능을 변경할 수 있습니다.
- “중심 케이스”는 기능 공간 도표에서 선택한 간단한 점이 됩니다. 중심 케이스 변수를 지정한 경우 중심 케이스를 나타내는 점이 처음으로 선택됩니다. 그러나 어떠한 점도 사용자가 선택하면 임시로 중심 케이스가 될 수 있습니다. 점 선택 시“일반적인l” 제어가 적용됩니다. 점을 클릭하면 해당 점을 선택하며 모든 다른 모든 점을 선택 취소합니다. 점을 컨트롤-클릭하면 해당 점을 선택한 점 세트에 추가합니다. 피어 도표와 같은 연결된 보기는 기능 공간에서 선택한 케이스를 기준으로 자동으로 업데이트합니다.
- 최근접 이웃 수(*k*)를 변경하여 중심 케이스에 표시할 수 있습니다.
- 도표에서 마우스를 점 위에 두면 케이스 레이블값 또는 케이스 레이블이 정의되지 않은 경우 케이스 번호, 그리고 관측 및 예측 대상 값을 포함한 도구 팁이 표시됩니다.
- “재설정” 단추는 기능 공간을 원래 상태로 되돌릴 때 사용합니다.

필드/변수 추가 및 제거

기능 공간에 새 필드/변수를 추가하거나 현재 표시된 필드/변수를 제거할 수 있습니다.

변수 팔레트

변수 팔레트를 표시해야 변수를 추가하고 제거할 수 있습니다. 변수 팔레트를 표시하려면 모형 뷰어가 편집 모드 상태이고 기능 공간에서 케이스가 선택되어 있어야 합니다.

1. 모형 뷰어를 편집 모드로 바꾸려면 메뉴에서 다음을 선택하십시오.

   보기 > 편집 모드
2. 편집 모드로 전환되면 기능 공간에서 케이스를 클릭합니다.
3. 변수 팔레트를 표시하려면 메뉴에서 다음을 선택하십시오.

보기 > 팔레트 > 변수

기능 공간에 있는 모든 변수가 변수 팔레트에 나열됩니다. 변수 이름 옆에 있는 아이콘은 변수의 측정 수준을 나타냅니다.

4. 변수의 측정 수준을 임시로 변경하려면 변수 팔레트에서 변수를 마우스 오른쪽 단추로 클릭하고 옵션을 선택합니다.

변수 영역

변수가 기능 공간의 "영역"에 추가됩니다. 영역을 표시하려면 변수 팔레트에서 변수를 끌거나 **영역 표시**를 선택합니다.

기능 공간에는 *x*, *y* 및 *z*축 영역이 있습니다.

변수를 영역으로 이동

다음은 변수를 영역으로 이동하는 일반적인 규칙과 팁입니다.

- 변수를 영역으로 이동하려면 변수 팔레트에서 변수를 클릭하고 영역에 끌어 옵니다. **영역 표시**를 선택하는 경우 영역을 마우스 오른쪽 단추로 클릭하고 영역에 추가할 변수를 선택할 수도 있습니다.
- 변수 팔레트에서 변수를 끌어서 다른 변수가 이미 있는 영역에 놓으면 이전 변수가 새 변수로 대체됩니다.
- 특정 영역에서 다른 변수가 이미 있는 영역에 변수를 끌어 놓으면 변수의 위치가 서로 바뀝니다.
- 영역에서 X를 클릭하면 영역에서 해당 변수가 제거됩니다.
- 시각화 템플리트에 여러 개의 그래픽 요소가 있는 경우, 각 그래픽 요소에 연관된 해당 변수 영역을 사용할 수 있습니다. 먼저 그래픽 요소를 선택하십시오.

변수 중요도

일반적으로 모형 작성 노력을 가장 중요한 변수에 집중하고 가장 쓸모 없는 변수를 삭제하거나 무시하기를 원합니다. 변수 중요도 도표를 사용하면 모형 추정 시 각 변수의 상대적인 중요도를 표시하여 원하는 작업을 수행할 수 있습니다. 값이 상대적이므로 표시된 모든 변수에 대한 값의 합은 1.0이 됩니다. 변수 중요도는 모형 정확도와는 관련이 없습니다. 단지 예측 시 각 변수의 중요도와 관련이 있으며 예측이 정확한지, 정확하지 않은지 여부와는 관련이 없습니다.

동위

이 도표에서는 중심 케이스와 각 기능 및 대상에 대한 최근접 이웃(*k*)이 표시됩니다. 기능 공간에서 중심 케이스를 선택한 경우 이 도표를 사용할 수 있습니다.

링크 동작. 피어 도표는 두 가지 방법으로 기능 공간에 연결됩니다.

- 기능 공간에서 선택한 케이스(초점)은 최근접 이웃(*k*)과 함께 피어 도표에 표시됩니다.
- 기능 공간에서 선택한 *k* 값은 피어 도표에 사용됩니다.

최근접 이웃 거리

이 테이블에는 최근접 이웃(*k*)과 중심 케이스의 거리만 표시됩니다. 변수 탭에 중심 케이스 식별자가 지정된 경우 이 테이블을 사용할 수 있으며 이 변수를 기준으로 식별한 중심 케이스만 표시합니다.

각 행:

- **중심 케이스** 열에는 중심 케이스에 대한 케이스 레이블 변수의 값이 포함됩니다. 케이스 레이블이 정의되지 않은 경우 이 열에는 중심 케이스의 케이스 번호가 포함됩니다.
- 최근접 이웃 그룹 아래에 있는 *i*번째 열에는 중심 케이스의 *i*번째로 최근접 이웃에 대한 케이스 레이블 변수의 값이 포함됩니다. 케이스 레이블이 정의되지 않은 경우 이 열에는 중심 케이스의 *i*번째로 최근접 이웃의 케이스 번호가 포함됩니다.
- 최근접 이웃 그룹 아래에 있는 *i*번째 열에는 중심 케이스에 대해 *i*번째로 최근접 이웃이 포함됩니다

사분 지도

이 도표에는 *y*축에 대상 및 *x*축에 척도 기능이 있는 산점도(또는 대상의 측정 수준에 따라 다른 점도표)에서 중심 케이스와 최근접 이웃(*k*)이 표시됩니다. 기능 공간에서 대상을 선택하고 기능 공간에서 중심 케이스를 선택할 경우 이 도표를 사용할 수 있습니다.

- 훈련 파티션의 변수 평균에서 연속적인 변수에 대해 참조선이 그려집니다.

변수 선택 오차 로그

도표에 있는 점은 *x*축에 나열된 기능(*x*축의 왼쪽에 모든 기능이 추가됨)을 가진 모형의 *y*축에 오차(대상의 측정 수준에 따라 오차 비율 또는 오차 제곱합)을 표시합니다 . 대상과 필드선택이 유효하면 이 도표를 사용할 수 있습니다.

k 선택 오차 로그

도표에 있는 점은 *x*축에 최근접 이웃 수(*k*축)를 가진 모형의 *y*축에 오차(대상의 측정 수준에 따라 오차 비율 또는 오차 제곱합)을 표시합니다 . 대상과 *k* 선택이 유효하면 이 도표를 사용할 수 있습니다.

k 및 변수 선택 오차 로그

이들은 *k*로 표시되는 기능 선택 차트입니다 (178 페이지의 『변수 선택 오차 로그』 참조). 이 차트는 대상이 있고 *k* 및 기능 선택사항이 모두 적용되는 경우에 사용할 수 있습니다.

분류표

이 테이블에는 파티션 기준으로 대상의 관측값 대 예측값의 교차 분류가 표시됩니다. 대상이 있고 그 대상이 범주형인 경우 이 테이블을 사용할 수 있습니다.

- 검증용 파티션의 **(결측)** 행에는 대상에 대한 결측값을 가진 검증 케이스가 포함됩니다. 이러한 케이스는 검증용 표본에 대해 전체 퍼센트값에 기여하지만 퍼센트 수정값에는 기여하지 않습니다.

오류 요약

대상값이 있는 경우 이 테이블을 사용할 수 있습니다. 이 테이블에는 모형 관련 오차, 즉 연속적인 대상에 대한 제곱합 및 범주형 대상에 대한 오차 비율(100% - 전체 퍼센트 수정)이 표시됩니다.

판별 분석

판별 분석은 그룹소속에 대한 예측 모형을 작성합니다. 모형은 프로시저로 그룹 간에 가장 큰 차별을 나타내는 예측자 변수의 선형 조합을 기본으로 하는 판별 함수(그룹이 셋 이상인 경우 판별 함수 세트)로 구성되어 있습니다. 이러한 함수는 해당 소속그룹이 알려진 케이스 표본으로부터 작성되며 해당 소속그룹은 알 수 없으나 예측자 변수 측정을 통해 새로운 케이스에 적용될 수는 있습니다.

참고: 그룹화 변수에는 셋 이상의 값이 있습니다. 그룹화 변수에 대한 코드는 정수이어야 하며 최소값과 최대값을 지정해야 합니다. 이러한 한계를 넘는 값이 있는 케이스는 분석에서 제외됩니다.

예. 일반적으로 온대 기후 지역에 살고 있는 사람들은 열대 기후 지역의 사람들보다 일일 칼로리 소비량이 많고 온대 기후 지역 사람들의 상당 부분이 도시에 거주하고 있습니다. 어떤 조사자는 이러한 정보를 함수로 조합하여 두 가지 유형의 국가에 거주하는 개인의 차이가 얼마나 큰지 판별하려고 합니다. 이 조사자는 인구 규모와 경제적 정보도 중요하다고 생각합니다. 판별 분석을 사용하여 다중 선형회귀 방정식의 오른쪽과 같은 모양으로 된 선형 판별 함수의 계수를 추정할 수 있습니다. 즉, *a*, *b*, *c*, *d*라는 계수를 사용하는 경우 함수는 다음과 같습니다.

```
D = a * climate + b * urban + c * population + d * gross domestic product per capita
```

이러한 변수는 두 기후 지역 사이를 판별하는 데 유용하며 *D* 값은 온대 국가와 열대 국가에 따라 달라집니다. 단계별 변수 선택법을 사용하면 이 함수에 네 가지 변수를 모두 포함시키지 않아도 됩니다.

통계. 각 변수에 대해 평균, 표준 편차, 일변량분산 분석 통계를 구합니다. 각 분석에 대해 Box의 *M*, 그룹-내 상관행렬, 그룹-내 공분산행렬, 개별-그룹 공분산행렬, 전체 공분산행렬 통계를 구합니다. 각 정준 판별 함수에 대해

고유 값, 분산 비율, 정준상관, Wilks의 람다, 카이제곱 통계를 구합니다. 각 단계에 대해 사전 확률, Fisher의 함수 계수, 표준화하지 않은 함수 계수, 각 정준 함수에 대한 Wilk의 람다 통계를 구합니다.

판별 분석 데이터 고려 사항

데이터. 그룹화 변수에는 정수로 코딩된 제한된 수의 개별 범주가 포함되어야 합니다. 명목인 독립변수는 더미변수나 대비 변수로 코딩이 변경되어야 합니다.

가정. 케이스는 독립적이어야 합니다. 예측자 변수에는 다변량 정규 분포가 있어야 하고 그룹-내 분산 공분산행렬은 여러 그룹에서 동일해야 합니다. 소속그룹은 상호 배타적(케이스가 둘 이상의 그룹에 소속되지 않음)이며 완전히 그룹화되어야(모든 케이스가 그룹의 소속이 됨) 합니다. 이 프로시저는 소속그룹이 진정한 범주형 변수일 때 효과적이므로 소속그룹이 연속형 변수값(예: 높은 IQ 대 낮은 IQ)에 따라 지정되면 선형 회귀분석을 사용하여 연속형 변수 자체에서 제공하는 다양한 정보를 활용해야 합니다.

판별 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 분류 > 판별...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 정수값으로 된 그룹화 변수를 선택하고 **범위지정**을 클릭하여 특정 범주를 지정합니다.
3. 독립변수나 예측자를 선택합니다. 그룹화 변수에 정수값이 없을 때는 변환 메뉴에서 자동 코딩변경으로 변수를 새로 작성합니다.
4. 독립변수 입력 방법을 선택합니다.
   - **독립변수 모두 진입.** 허용 오차 기준에 맞는 독립변수가 모두 동시에 입력됩니다.
   - **단계적 방법 사용.** 단계선택 분석을 사용하여 변수 입력 및 제거를 제어합니다.
5. 선택적으로 선택변수를 사용하여 케이스를 선택할 수 있습니다.

판별 분석 범위지정

분석할 그룹화 변수의 최소값과 최대값을 지정합니다. 이 범위를 벗어나는 값이 있는 케이스는 판별 분석에서 사용되지 않으나 분석 결과에 따라 기존의 그룹 중 하나로 분류됩니다. 최소값과 최대값은 정수이어야 합니다.

판별 분석 케이스 선택

분석할 케이스를 선택하려면

1. 판별 분석 대화 상자에서 선택변수를 선택합니다.
2. **값**을 클릭하고 선택 값으로 정수를 입력합니다.

선택변수에 대해 지정된 값이 있는 케이스만 판별 함수 작성에 사용됩니다. 통계와 분류결과는 선택한 케이스와 선택하지 않은 케이스에 대해 모두 생성됩니다. 이를 통해 이전 데이터를 기준으로 새로운 케이스를 분류하는 방법이나 생성된 모형에 대한 유효성을 검증하기 위해 서브세트를 조절하고 검정하도록 데이터를 분할하는 방법이 제공됩니다.

판별 분석 통계

기술통계. 사용할 수 있는 옵션은 평균(표준 편차 포함), 일변량분산 분석, Box의 *M* 검정입니다.

- 평균*(Means)*. 전체 평균 및 그룹 평균, 독립변수에 대한 표준 편차를 표시합니다.
- 일변량분산 분석*(Univariate ANOVAs)*. 각 독립변수에 대해 그룹 평균의 등식을 검정하는 일원 분산 분석을 수행합니다.
- *Box*의 *M* . 그룹 공분산 행렬의 방정식에 대한 검정입니다. 표본이 충분히 큰 경우 p 값에 유의수준이 없으면 행렬이 다르고 판단하기 어렵습니다. 이 검정은 다변량 정규성에서 벗어나는 경우 영향을 많이 받습니다.

함수의 계수. 사용할 수 있는 옵션은 Fisher의 분류 계수 및 비표준화 계수입니다.

- *Fisher*의 방법*(Fisher's)*. 분류에 직접 사용할 수 있는 Fisher의 분류 함수 계수를 표시합니다. 각 그룹에 대해 개별적인 일련의 분류 함수 계수가 작성되고 케이스는 판별 스코어(분류 함수 값)가 가장 큰 그룹에 할당됩니다.

- 비표준화*(Unstandardized)*. 표준화하지 않은 판별 함수 계수를 표시합니다.

행렬. 사용할 수 있는 독립변수에 대한 계수의 행렬은 그룹-내 상관행렬, 그룹-내 공분산행렬, 개별-그룹 공분산 행렬, 전체 공분산행렬입니다.

- 그룹 내 상관 관계. 상관을 계산하기 전에 모든 그룹에 대한 개별 공분산 교차표의 평균을 구하여 그룹 내 풀링 상관 행렬을 표시합니다.
- 그룹 내 공분산. 총 공분산 행렬과 다를 수 있는 그룹 내 풀링 공분산 행렬을 표시합니다. 이 행렬은 모든 그룹에 대해 개별 공분산행렬을 평균하여 구합니다.
- 개별 그룹 공분산. 각 그룹에 대해 개별 공분산 교차표를 표시합니다.
- 총 공분산. 단일 표본으로 작성한 것처럼 모든 케이스로부터 공분산 교차표를 표시합니다.

판별 분석 단계적 방법

메소드. 새로운 변수 입력 및 제거에 사용할 통계를 선택합니다. 사용 가능한 대안은 Wilks' lambda, 설명되지 않은 분산, 마하라노비스 거리, 가장 작은 *F* 비율 및 Rao의 *V*입니다. Rao의 *V*를 사용하면 입력할 변수에 대해 *V*의 최소 증가를 지정할 수 있습니다.

- *Wilks' lambda*. Wilks의 람다를 낮추는 정도에 따라 방정식에 입력할 변수를 선택하는 단계별 판별 분석을 위한 변수 선택 방법입니다. 각 단계에서 전체 Wilks의 람다를 최소화하는 변수가 입력됩니다.
- 설명되지 않은 분산. 각 단계에서 그룹 간 설명되지 않은 변동 합계를 최소화하는 변수를 입력합니다.
- 마하라노비스 거리. 독립변수의 케이스 값이 모든 케이스의 평균과 얼마나 다른지에 대한 측도입니다. Mahalanobis 거리가 크면 케이스가 독립변수 하나 이상에 대해 극단값을 갖는 것으로 식별합니다.
- 가장 작은 *F* 비율. 그룹 간 Mahalanobis 거리로부터 계산한 F-비를 최대화하는 단계별 분석의 변수 선택 방법입니다.
- *Rao*의 *V*. 그룹 평균 간 차이의 측도입니다. Lawley-Hotelling 트레이스라고도 하며 각 단계에서 Rao의 V의 증가를 최대화하는 변수가 입력됩니다. 이 옵션을 선택한 다음 변수가 가져야 하는 최소값을 입력하여 분석에 사용합니다.

기준. 사용 가능한 대안은 **F 값 사용** 및 **F의 확률 사용**입니다. 변수를 입력 및 제거하기 위한 값을 입력하십시오.

- *F* 값 사용. F값이 진입값보다 크면 변수가 모형에 입력되고 F값이 제거값보다 작으면 제거됩니다. 진입값은 제거값보다 커야 하고 두 값 모두 양수이어야 합니다. 모형에 더 많은 변수를 입력하려면 진입값을 낮추고 모형에서 변수를 더 많이 제거하려면 제거값을 높입니다.
- *F*-확률 사용. F값의 유의 수준이 진입값보다 작으면 변수가 모델에 입력되고 유의 수준이 제거값보다 크면 제거됩니다. 진입값은 제거값보다 작아야 하며 두 값 모두 양수이어야 합니다. 모형에 변수를 더 많이 입력하려면 진입값을 높이고 모형에서 변수를 더 많이 제거하려면 제거값을 낮춥니다.

표시. 단계 요약은 각 단계 이후의 모든 변수에 대한 통계를 표시하고, **쌍별 거리에 대한 F**는 각 그룹 쌍에 대한 쌍별 *F* 비율의 행렬을 표시합니다.

판별 분석 분류

사전 확률. 이 옵션은 그룹소속의 사전 지식에 대해 분류 계수를 조정할지 여부를 결정합니다.

- **모든 그룹이 동일.** 동일한 사전 확률이 모든 그룹에 가정되며 계수에는 아무런 영향이 없습니다.
- **그룹표본크기로 계산.** 표본에서 관측된 그룹 크기는 그룹소속의 사전 확률을 결정합니다. 예를 들어, 관측의 50%가 첫 번째 그룹, 25%가 두 번째 그룹, 25%가 세 번째 그룹에 속하는 분석에 포함된 경우 분류 계수는 다른 두 그룹에 상대적으로 첫 번째 그룹의 소속 가능성을 증가시키도록 조정됩니다.

표시. 사용 가능한 출력 옵션에는 각 케이스에 대한 결과, 요약표, 순차제거복원 분류 등이 있습니다.

- 케이스별 결과. 각 케이스마다 실제 그룹, 예측 그룹, 사후 확률, 판별 스코어 등에 대한 코드가 표시됩니다.
- 요약 테이블. 판별 분석을 기반으로 각 그룹에 올바르게 지정된 케이스와 올바르지 않게 지정된 케이스의 수입니다. "혼동행렬"이라고도 합니다.
- 순차제거복원 분류. 분석의 각 케이스는 해당 케이스가 아닌 모든 케이스에서 파생된 함수로 분류됩니다. 이 방법을 "U-방법"이라고도 합니다.

결측값을 평균으로 바꾸기. 이 옵션을 선택하면 분류 단계에서만 결측값에 대한 독립변수의 평균을 바꿀 수 있습니다.

공분산행렬 사용. 이 옵션을 선택하여 그룹-내 공분산행렬이나 개별-그룹 공분산행렬을 사용하여 케이스를 분류할 수 있습니다.

- 그룹 내. 그룹 내 풀링 공분산 교차표가 케이스 분류에 사용됩니다.
- 개별 그룹. 개별-집단 공분산행렬이 분류에 사용됩니다. 분류가 판별 함수에 기초하고 원래 변수에 따라 달라지지는 않으므로 이 옵션이 2차 판별과 항상 같지는 않습니다.

도표. 사용할 수 있는 도표 옵션은 결합-그룹, 개별-그룹, 영역 맵입니다.

- 결합-그룹. 처음 두 판별 함수 값의 모든 집단 산점도를 작성합니다. 함수가 하나만 있는 경우에는 산점도 대신 히스토그램이 표시됩니다.
- 개별 그룹. 처음 두 판별 함수 값의 개별-집단 산점도를 작성합니다. 함수가 하나만 있는 경우에는 산점도 대신 히스토그램이 표시됩니다.
- 영역 맵. 함수 값에 따라 케이스를 집단으로 분류하는 데 사용하는 경계의 도표입니다. 숫자는 케이스가 분류된 그룹에 해당합니다. 각 그룹의 평균은 경계 내에서 별표로 표시됩니다. 판별 함수가 하나만 있는 경우에는 맵이 표시되지 않습니다.

판별 분석 저장

새로운 변수를 활성 데이터 파일에 추가할 수 있습니다. 사용할 수 있는 옵션은 예측 소속그룹(단일-변수), 판별 스코어(해법에서의 각 판별 함수에 대한 하나의 변수), 판별 스코어에 지정된 그룹소속 확률(각 그룹에 대한 하나의 변수)입니다.

모형 정보를 XML 형식의 지정된 파일로 내보낼 수 있습니다. 스코어링 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.

DISCRIMINANT 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 하나의 명령으로 여러 판별 분석을 수행하고 ANALYSIS 하위 명령을 사용하여 변수가 입력되는 순서를 제어합니다.
- PRIORS 하위 명령을 사용하여 분류에 대한 사전 확률을 지정합니다.
- ROTATE 하위 명령을 사용하여 회전 패턴 행렬과 구조 행렬을 표시합니다.
- FUNCTIONS 하위 명령을 사용하여 추출된 판별 분석 함수의 수를 제한합니다.
- SELECT 하위 명령을 사용하여 분석을 위해 선택(또는 선택 취소)한 케이스로 분류를 제한합니다.
- MATRIX 하위 명령을 사용하여 상관행렬을 읽고 분석합니다.
- MATRIX 하위 명령을 사용하여 나중에 분석하기 위해 상관행렬을 기록합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

요인 분석

요인 분석을 사용하여 기본 변수 또는 **요인**을 식별하여 관측변수 세트에서 상관 패턴을 설명할 수 있습니다. 요인 분석은 많은 수의 명백한 변수에서 관측된 대부분의 분산을 설명할 수 있는 소수의 요인을 식별함으로써 데이터 축소에 사용됩니다. 요인 분석을 사용하여 일반적인 메커니즘에 관한 가설을 작성하거나 후속 분석(예: 선형 회귀분석에 앞서 공선성을 식별)에 대한 변수를 구분할 수 있습니다.

요인 분석 프로시저에서는 다음과 같은 상당한 유동성을 제공합니다.

- 일곱 가지 요인 추출 방법을 사용할 수 있습니다.
- 비직교 회전에 대한 직접 오블리민과 프로멕스 방법을 포함하여 다섯 가지 회전 방법을 사용할 수 있습니다.
- 요인 스코어를 계산하는 데 세 가지 방법이 있으며 스코어는 이후의 분석에 대한 변수로 저장됩니다.

예. 정치적 여론 설문조사에서 반응자가 질문에 반응하는 기본 항목은 무엇입니까? 설문조사 항목 간의 상관을 살펴보면 항목의 여러 하위 그룹에서 중요한 교차 관계를 볼 수 있는데 조세 관련 질문과 군사 문제 등은 서로 상관이 있음을 알게 됩니다. 요인 분석을 통해 기본 요인의 수를 조사하고 여러 케이스에서 요인이 개념적으로 무엇을 나타내는지 식별할 수 있습니다. 또한 각 반응자에 대한 요인 스코어를 계산하여 후속 분석에서 사용할 수 있습니다. 예를 들어, 로지스틱 회귀 모형을 작성하여 요인 스코어에 따라 투표 행태를 예측할 수 있습니다.

통계. 각 변수에 대해 유효한 케이스 수, 평균 및 표준 편차를 선택할 수 있습니다. 각 요인 분석에 대해 변수의 상관행렬(유의수준, 판별계수, 역 상관행렬 포함), 재연된 상관행렬(역-이미지 행렬 포함), 초기해법(공통성, 고유 값, 설명한 분산 퍼센트), Kaiser-Meyer-Olkin 표본 타당성 측도와 Bartlett의 단위행렬 검정, 비회전 해법(요인 로딩, 공통성, 고유 값 등 포함), 회전 해법(회전 패턴 행렬과 변환 행렬 포함)을 선택할 수 있습니다. 각 사각 회전에 대해 회전 패턴과 구조 행렬, 요인 스코어 계수행렬과 요인 공분산행렬을 선택할 수 있습니다. 도표에 대해 고유 값의 스크리 도표와 처음 둘이나 세 요인의 로딩 도표를 선택할 수 있습니다.

요인 분석 데이터 고려 사항

데이터. 변수는 구간 수준이나 비율 수준에서 양적변수이어야 합니다. 범주형 데이터(예: 출생 지역이나 국가)는 요인 분석에 적합하지 않습니다. 요인 분석에는 Pearson 상관계수를 확실하게 계산할 수 있는 데이터가 적합합니다.

가정. 데이터에는 각 대응변수에 대한 이변량 정규 분포가 있어야 하며 관측값은 독립이어야 합니다. 요인 분석 모형을 사용하여 공통 요인(모형으로 추정된 요인)과 고유 요인(관측 변수간에 겹쳐지지 않는 요인)으로 결정될 변수를 지정합니다. 계산된 추정값은 모든 고유 요인이 서로 또는 공통 요인과 상관이 없다는 가정에 기준합니다.

요인분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 차원 축소 > 요인...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 요인 분석에 대한 변수를 선택합니다.

요인 분석 선택변수

분석할 케이스를 선택하려면

1. 선택변수를 선택합니다.
2. **값**을 클릭하고 선택 값으로 정수를 입력합니다.

요인 분석에서는 선택변수 값이 있는 케이스만을 사용합니다.

요인 분석: 기술통계

통계 단변량 기술에는 평균, 표준 편차 및 각 변수에 대한 유효한 케이스 수가 포함됩니다. **초기해법**에서는 공통성, 고유 값, 설명된 분산 퍼센트 등을 표시합니다.

상관행렬. 사용할 수 있는 옵션은 계수, 유의수준, 행렬식, 역 모형, 재연된 상관행렬, 역-이미지, KMO와 Bartlett의 단위행렬 검정입니다.

- *KMO*와 *Bartlett*의 단위행렬 검정. 표본추출 타당성의 Kaiser-Meyer-Olkin 측도는 변수 간 편상관이 작은지 여부를 검정합니다. Bartlett의 단위행렬 검정은 상관행렬이 요인 모형이 부적합함을 나타내는 항등 행렬인지 여부를 검정합니다.
- 재생산. 요인 해법에서 추정된 상관 행렬입니다. 잔차(추정된 상관과 기대된 상관 간 차이)도 표시됩니다.
- 안티이미지. 역 이미지 상관 행렬에는 편상관계수의 음수가 포함되고 역 이미지 공분산 행렬에는 편공분산의 음수가 포함됩니다. 적합한 요인 모형에서는 대부분 비-대각 원소가 작습니다. 변수의 표본추출 타당성 측도는 역 이미지 상관행렬의 대각선에 표시됩니다.

요인 분석: 요인추출

메소드. 요인추출 방법을 지정할 수 있습니다. 사용할 수 있는 방법은 주성분, 가중치를 부여하지 않은 최소제곱법, 일반화 최소제곱법, 최대 우도, 주축 요인 추출, 알파 요인 추출, 이미지 요인 추출 등입니다.

- 주성분 분석*(Principal Components Analysis)*. 관측변수의 비상관 선형 조합을 형성하는 데 사용되는 요인 추출 방법입니다. 처음 성분이 최대 분산을 가지게 됩니다. 성분이 연속될수록 점진적으로 더 작아지는 분산을 나타내며 각 성분은 서로 상관되지 않습니다. 주성분 분석으로 초기 요인 해법을 구할 수 있으며 상관행렬이 단순할 때 사용됩니다.
- 가중치를 부여하지 않은 최소제곱법*(Unweighted Least-Squares Method)*. 관측된 상관 행렬과 재연된 상관 행렬 간의 차이 제곱값 합을 최소화하는 요인 추출 방법입니다(대각선 무시).
- 일반화 최소제곱법*(Generalized Least-Squares Method)*. 관측된 상관 행렬과 재연된 상관 행렬 간 차이 제곱값 합을 최소화하는 요인 추출 방법입니다. 상관은 역으로 가중되므로 고유성이 높은 변수는 고유성이 낮은 변수보다 가중치가 적습니다.
- 최대-우도법*(Maximum-Likelihood Method)*. 표본이 다변량 정규 분포에서 추출된 경우 관측 상관 행렬을 생성할 가능성이 가장 높은 모수 추정값을 생성하는 요인 추출 방법입니다. 상관은 변수의 역고유성에 의해 가중되고 반복 알고리즘이 사용됩니다.
- 주축 요인 추출*(Principal Axis Factoring)*. 공통성의 초기 추정값으로 대각선으로 배치된 제곱 다중 상관계수를 사용하여 원래 상관 행렬에서 요인을 추출하는 방법입니다. 이러한 요인 로딩은 대각선으로 위치하는 기존의 공통성 추정값을 대신하는 새로운 공통성을 추정하는 데 사용됩니다. 한 반복에서 다음 반복까지 공통성에 대한 변화량이 추출에 대한 수렴 기준을 만족할 때까지 반복은 계속됩니다.
- 알파 요인 추출*(Alpha Factoring)*. 분석의 변수를 잠재적 변수의 우주의 표본으로 간주하는 요인 추출 방법입니다. 이 방법은 요인의 알파 신뢰도를 최대화합니다.
- 이미지 요인 추출*(Image Factoring)*. Guttman이 개발하고 이미지 이론을 기반으로 하는 요인 추출 방법입니다. 편 이미지라고 하는 변수의 공통 부분은 가설 요인의 함수보다는 남아 있는 변수에서 선형 회귀분석으로 정의됩니다.

분석. 상관행렬이나 공분산행렬을 지정할 수 있습니다.

- **상관행렬.** 분석의 변수가 다양한 척도에서 측정된 경우 유용합니다.
- **공분산행렬.** 각 변수의 다양한 분산을 사용하여 여러 그룹에 요인 분석을 적용할 때 유용합니다.

추출. 고유 값이 지정한 값보다 큰 요인을 모두 포함하거나 지정한 요인의 수만 포함할 수 있습니다.

표시. 회전하지 않은 요인해법과 고유 값의 스크리 도표를 요청할 수 있습니다.

- 회전하지 않은 요인 해법*(Unrotated Factor Solution)*. 요인 해법에 대한 회전하지 않은 요인 로딩(요인 패턴 행렬), 공통성 및 고유값을 표시합니다.
- 화면 플롯. 각 요인과 관련된 분산 도표입니다. 그대로 유지할 요인의 수를 결정하는 데 사용됩니다. 일반적으로 도표는 큰 요인들의 가파른 기울기와 나머지 요인들의 점진적 꼬리 부분 (스크리) 간의 뚜렷한 구분을 보여줍니다.

수렴에 대한 최대반복수. 해법을 추정할 때 알고리즘에서 사용할 최대 단계수를 지정할 수 있습니다.

요인 분석: 회전

메소드. 요인 회전 방법 및 Kaiser 정규화 적용 여부를 선택할 수 있습니다. 사용할 수 있는 방법은 베리멕스, 이쿼멕스, 쿼티멕스, 직접 오블리민, 프로멕스입니다.

- 베리멕스 방법*(Varimax Method)*. 각 요인에 대해 로딩이 높은 변수의 수를 최소화하는 직교 회전 방법입니다. 이 방법을 사용하면 요인 해석을 단순화할 수 있습니다.
- 직접 오블리민 방법. 사각 (비직교) 회전 방법입니다. 델타가 0(기본값)일 때 해법은 가장 기울어지는 형태를 나타냅니다. 델타가 음수에 가까워질수록 요인의 기울기가 평평해집니다. 기본값 델타 0을 바꾸려면 0.8 이하의 수를 입력합니다.
- 쿼티멕스 방법*(Quartimax Method)*. 각 변수를 설명하는 데 필요한 요인 수를 최소화하는 회전 방법입니다. 이 방법을 사용하면 관측된 변수의 해석을 단순화할 수 있습니다.
- 이쿼멕스 방법*(Equamax Method)*. 요인을 단순화하는 베리멕스 방법과 변수를 단순화하는 쿼티멕스 방법의 조합인 회전 방법입니다. 요인에 읽어들인 변수의 수와 변수 설명에 사용할 요인 수는 최소화됩니다.
- 프로멕스 회전*(Promax Rotation)*. 요인을 상관시킬 수 있는 사각 회전입니다. 이 회전은 직접 오블리민 회전보다 좀더 빨리 계산될 수 있으므로 큰 데이터 세트에 유용합니다.
- 카이저 정규화 적용. 기본적으로 회전이 지정되면 Kaiser 정규화를 적용할 수 있습니다.

표시. 처음 둘이나 세 요인에 대한 로딩 도표뿐 아니라 회전 해법에 대한 결과를 포함시킬 수 있습니다.

- 회전 해법*(Rotated Solution)*. 회전 해법을 구하려면 회전 방법을 선택해야 합니다. 직교회전에 대해 회전 패턴 행렬과 요인 변환 행렬이 표시됩니다. 오블리크 회전의 경우에는 패턴, 구조, 요인 상관행렬이 표시됩니다.
- 요인로딩 도표*(Factor Loading Plot)*. 처음 세 요인의 3차원 요인 로딩 도표입니다. 두 요인 해법의 경우 2차원 도표가 표시됩니다. 요인이 하나만 추출되면 도표가 표시되지 않습니다. 회전을 요청하면 도표에 회전 해법이 표시됩니다.

수렴에 대한 최대반복수. 회전을 수행할 때 알고리즘에서 사용할 최대 단계 수를 지정할 수 있습니다.

요인 분석: 요인 스코어

변수로 저장. 마지막 해법의 각 요인에 대해 새로운 변수를 하나씩 작성합니다.

메소드. 요인 스코어를 계산하는 다른 방법은 회귀분석, Bartlett 및 Anderson-Rubin입니다.

- 회귀분석방법*(Regression Method)*. 요인 스코어 계수를 추정하는 방법입니다. 생성된 스코어는 평균이 0이며 추정된 요인 스코어와 요인 값 간의 제곱 다중 상관과 동일한 분산을 가집니다. 스코어는 요인이 직교될 경우에도 상관될 수 있습니다.
- *Bartlett* 점수. 요인 스코어 계수를 추정하는 방법입니다. 생성되는 점수는 평균이 0입니다. 변수 범위에 대한 고유 요소의 제곱합은 최소화됩니다.
- *Anderson-Rubin* 방법*(Anderson-Rubin Method)*. 요인 스코어 계수 추정 방법입니다. 추정된 요인의 직교성을 보장하는 Bartlett 방법을 수정한 것입니다. 생성된 스코어의 평균은 0이고 표준 편차는 1이며 스코어는 상호 관련되지 않습니다.

요인 스코어 계수행렬 출력. 요인 스코어를 구하기 위해 변수를 곱하는 계수를 나타냅니다. 요인 스코어간 상관도 볼 수 있습니다.

요인 분석: 옵션

결측값. 결측값 처리 방법을 지정할 수 있습니다. 목록별 결측값 제외, 대응별 결측값 제외, 평균으로 바꾸기 등을 선택할 수 있습니다.

계수표시 형식. 출력되는 행렬을 제어할 수 있습니다. 계수를 크기별로 정렬하고 지정한 값보다 절대값이 작은 계수를 출력하지 않도록 합니다.

FACTOR 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 요인추출과 회전하는 동안 반복을 위한 수렴 기준을 지정합니다.
- 개별 회전된 요인 도표를 지정합니다.
- 요인 스코어를 몇 개나 저장할지 지정합니다.
- 주축 요인 방법에 사용할 대각선 값을 지정합니다.
- 상관행렬 또는 요인로딩 행렬을 나중에 분석하기 위해 디스크에 저장합니다.
- 상관행렬이나 요인로딩 행렬을 읽고 분석합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

군집화 프로시저 선택

군집 분석은 이단계, 계층적 또는 K-Means 군집 분석 프로시저를 사용하여 수행될 수 있습니다. 각 프로시저는 서로 다른 알고리즘을 사용하여 군집을 작성하므로 다른 프로시저에서는 사용할 수 없는 옵션이 있습니다.

이단계 군집 분석. 많은 애플리케이션에서 이단계 군집 분석 프로시저를 방법으로 선택합니다. 이단계 군집 분석 프로시저에는 다음과 같은 고유한 기능이 있습니다.

- 군집 모형 간 선택을 위한 측도 이외의 최적의 군집 수 자동 선택
- 범주형 변수와 연속형 변수를 기준으로 동시에 군집 모형을 작성하는 기능

- 외부 XML 파일에 군집 모형을 저장한 다음 해당 파일을 읽고 최신 데이터를 사용하여 군집 모형을 업데이트하는 기능

이단계 군집 분석 프로시저에서는 대용량 데이터 파일도 분석할 수 있습니다.

계층적 군집 분석. 계층적 군집 분석 프로시저는 소용량 데이터 파일로 제한되지만(수백개의 오브젝트 군집화) 다음과 같은 고유한 기능이 있습니다.

- 케이스 또는 변수 군집 기능
- 가능한 해법범위를 계산하고 각 해법에 대한 소속군집을 저장하는 기능
- 군집 형성, 변수 변환 및 군집 간 상이성 측정을 위한 몇 가지 방법

모든 변수가 같은 유형일 경우 계층적 군집 분석 프로시저에서는 구간(연속), 개수 또는 이분형 변수를 분석할 수 있습니다.

K-Means 군집 분석. K-Means 군집 분석 프로시저는 연속 데이터로 제한되고 군집의 수를 미리 지정해야 하지만 다음과 같은 고유한 기능이 있습니다.

- 각 오브젝트에 대한 군집중심으로부터의 거리 저장 기능
- 초기 군집중심을 읽고 최종 군집중심을 외부 IBM SPSS 통계 파일로 저장하는 기능.

K-Means 군집 분석 프로시저에서는 대용량 데이터 파일도 분석할 수 있습니다.

이단계 군집 분석

이단계 군집 분석 프로시저는 다른 방법으로는 잘 판단할 수 없는 데이터 세트의 기본 그룹(또는 군집) 특성을 분석하는 새로운 기능입니다. 이 프로시저에서 사용하는 알고리즘에는 일반적인 군집 기법과는 다른 몇 가지 기능이 있습니다.

- **범주형 변수와 연속형 변수 처리.** 변수를 독립변수로 가정하여 범주형 변수 및 연속형 변수를 결합 다항 정규 분포로 표시할 수 있습니다.
- **군집 수 자동 선택.** 군집 솔루션별로 모형 선택 기준 값을 비교하여 최적의 군집 수를 자동으로 결정할 수 있습니다.
- **척도.** 이단계 알고리즘에서 레코드를 요약하는 CF(군집 기능)를 구성하여 큰 데이터 파일을 분석할 수 있습니다.

예. 소매 및 소비재 회사는 고객의 구매 습관, 성별, 나이, 소득 수준 등을 설명하는 데이터에 클러스터링 기법을 정기적으로 적용합니다. 이 회사는 판매를 증가시키고 브랜드 로열티를 구축하기 위해 해당 마케팅 및 제품 개발 전략을 각 고객 그룹에 맞게 조정합니다.

거리 측도. 이 기능을 선택하면 두 군집 간 유사성이 계산되는 방식을 결정할 수 있습니다.

- **로그-우도.** 로그-우도 측도를 선택하면 변수에 대한 확률 분포가 계산됩니다. 연속형 변수는 정규 분포로, 범주형 변수는 다항분포로 계산된다고 가정하며 모든 변수를 독립변수로 가정합니다.
- **유클리디안.** 유클리디안 측도를 선택하면 두 군집 간 "직선" 거리가 계산됩니다. 모든 변수가 연속형 변수인 경우에만 이 옵션을 사용할 수 있습니다.

군집 수. 이 기능을 선택하면 군집 수가 결정되는 방식을 지정할 수 있습니다.

- **자동 결정.** 이 프로시저에서는 군집 기준에서 지정한 기준을 사용하여 최적의 군집 수를 자동으로 결정합니다. 사용자가 선택적으로 이 프로시저에 대한 고려 사항으로 최대 군집 수를 지정하는 양의 정수를 입력할 수도 있습니다.
- **고정 수 지정.** 이 옵션에서 양의 정수를 입력하여 해법의 군집 수를 양의 정수를 입력하십시오.

연속형 변수의 개수. 이 옵션을 사용하여 옵션 대화 상자에서 만든 연속형 변수 표준화 지정 사항에 대한 요약을 제공합니다. 자세한 정보는 186 페이지의 『이단계 군집 분석 옵션』 주제를 참조하십시오.

군집 기준. 이 기능을 선택하면 자동 군집화 알고리즘이 군집 수를 결정하는 방식을 지정할 수 있습니다. 군집 기능으로 Schwartz 베이지안 기준(BIC) 또는 Akaike 정보 기준(AIC)을 지정할 수 있습니다.

이단계 군집 분석 데이터 고려 사항

데이터. 이 프로시저는 연속형 변수와 범주형 변수에 모두 적용됩니다. 케이스는 군집화할 오브젝트, 변수는 군집 기준이 되는 속성을 나타냅니다.

케이스 순서. 군집 기능 나무와 최종 해법은 케이스 순서에 따라 달라질 수 있습니다. 순서가 미치는 영향을 최소화하려면 케이스 순서를 무작위로 설정해야 합니다. 주어진 해법의 안정성을 확인하기 위해 케이스를 각기 다른 무작위 순서로 정렬하여 서로 다른 여러 가지 해법을 구할 수도 있습니다. 파일 크기가 너무 커서 이를 수행하기가 어려운 경우에는 케이스 표본을 각기 다른 무작위 순서로 정렬하여 여러 번 실행하는 방법을 대신 사용할 수도 있습니다.

가정. 로그-우도 거리 측도는 군집 모형의 변수를 독립변수로 가정합니다. 또한 연속형 변수마다 개별 정규 분포(가우시안)를 가지며 범주형 변수마다 다항분포 특성을 가진다고 가정합니다. 실제 내부 검정을 통해 이 프로시저가 독립성과 분산에 대한 가정에 그리 큰 영향을 받지 않는다는 결론을 얻었지만 되도록이면 이러한 가정을 충족하는 것이 좋습니다.

두 연속형 변수의 독립성을 검정하려면 이변량 상관계수 프로시저를 사용합니다. 두 범주형 변수의 독립성을 검정하려면 교차 분석표 프로시저를 사용합니다. 연속형 변수와 범주형 변수 간의 독립성을 검정하려면 평균 프로시저를 사용합니다. 연속형 변수의 정규성을 검정하려면 탐색 프로시저를 사용합니다. 범주형 변수에 다항분포가 지정되어 있는지를 검정하려면 카이제곱 검정 프로시저를 사용합니다.

이단계 군집 분석을 수행하려면 다음과 같이 하십시오.

1. 메뉴에서 다음을 선택합니다.

   분석 > 분류 > 이단계 군집...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 범주형 변수와 연속형 변수를 하나 이상 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 군집 구성 기준을 조정합니다.
- 잡음 처리, 메모리 할당, 변수 표준화, 군집 모형 입력에 대한 설정을 선택합니다.
- 모형 뷰어 결과를 요청합니다.
- 모형 결과를 작업 파일이나 외부 XML 파일에 저장합니다.

이단계 군집 분석 옵션

이상값 처리. 이 옵션을 사용하면 CF(군집 기능) 나무가 채워진 경우 군집화를 수행하는 동안 이상값을 처리할 수 있습니다. CF 나무 잎 노드에 케이스를 더 이상 추가할 수 없거나 잎 노드를 더 이상 분할할 수 없으면 CF 나무가 채워진 것입니다.

- 잡음 처리를 선택하고 CF 나무가 채워진 경우 산재되어 있는 잎의 케이스를 "잡음" 잎에 추가하면 나무가 다시 작성됩니다. 잎은 최대 잎 크기 케이스의 지정된 퍼센트보다 더 적은 퍼센트의 케이스를 포함하고 있을 때 산재하고 있는 것으로 간주됩니다. CF 나무가 다시 작성되면 이 나무에 이상값이 적용됩니다. 그렇지 않으면 이상값은 무시됩니다.
- 잡음 처리를 선택하지 않고 CF 나무가 채워진 경우 거리 변경 임계값을 늘려 나무를 다시 작성할 수 있습니다. 최종 군집화 후에 군집에 할당될 수 없는 값은 이상값으로 설명됩니다. –1이라는 식별 번호가 지정된 이상값 군집은 군집 수의 개수에 포함되지 않습니다.

메모리 할당. 이 옵션을 사용하면 군집화 알고리즘이 사용할 최대 메모리 양(MB)를 지정할 수 있습니다. 프로시저에서 이 최대값을 초과하는 경우 시스템에 저장하지 못한 정보를 디스크에 별도로 저장합니다. 4보다 크거나 같은 수를 지정합니다.

- 시스템에 지정할 수 있는 최대값은 시스템 관리자에게 문의하십시오.
- 이 값이 너무 작으면 알고리즘에서 올바른 군집 수나 지정된 군집 수를 찾지 못할 수도 있습니다.

변수 표준화. 표준화된 연속형 변수에 군집화 알고리즘을 수행합니다. 표준화되지 않은 모든 연속형 변수는 표준화될 예정임 목록의 변수로 지정해야 합니다. 표준화를 가정함 목록의 변수로 이미 표준화한 연속형 변수를 선택하면 더 신속하고 간단하게 군집화 알고리즘을 수행할 수 있습니다.

고급 옵션

CF 나무 튜닝 기준. 다음 군집화 알고리즘 설정은 CF(군집 기능) 나무에 적용되기 때문에 변경 시 주의해야 합니다.

- **초기 거리 변경 임계값.** 이 옵션은 CF 나무를 작성할 때 사용되는 초기 임계값입니다. 특정 케이스를 CF 나무 잎에 삽입하면 임계값보다 나무가 좁게 작성되므로 잎이 분리되지 않습니다. 임계값보다 나무가 넓게 작성되면 잎이 분리됩니다.
- **최대 분기(잎 노드당).** 잎 노드가 가질 수 있는 최대 하위 노드 수입니다.
- **최대 나무 깊이** CF 나무가 가질 수 있는 최대 레벨 수입니다.
- **가능한 최대 노드 수.** 함수(b^{d+1} -1)/(b -1)를 기반으로 프로시저에 의해 잠재적으로 생성될 수 있는 CF 트리 노드의 최대 수를 나타냅니다. 여기서 b는 최대 분기이고 d는 최대 트리 깊이입니다. CF 나무 크기가 너무 크면 시스템 자원이 낭비되기 때문에 프로시저 성능을 저하시킬 수 있습니다. 각 노드의 최소 저장 공간은 16바이트입니다.

군집 모형 업데이트. 이전 분석에서 생성한 군집 모형을 가져와 업데이트할 수 있습니다. 입력 파일에는 XML 형식의 CF 나무가 있습니다. 모형은 활성 파일 데이터로 업데이트됩니다. 이전 분석에서 지정한 순서대로 기본 대화 상자의 변수 이름을 선택해야 합니다. 같은 파일 이름에 새로운 모형 정보를 쓰지 않는 한 XML 파일은 변경되지 않습니다. 자세한 정보는 187 페이지의 『이단계 군집 분석 결과』 주제를 참조하십시오.

군집 모형 업데이트를 지정하면 원래 모형에 대해 지정된 CF 나무의 생성과 관련된 옵션이 사용됩니다. 즉, 저장된 모형에 대한 거리 측도, 잡음 처리, 메모리 할당 또는 CF 나무 튜닝 기준 설정이 사용되고 대화 상자에 있는 해당 옵션 설정은 모두 무시됩니다.

참고: 군집 모형 업데이트를 수행할 때 프로시저에서는 활성 데이터 세트에서 선택한 케이스가 원래 군집 모형을 만드는 데 사용되지 않았다고 가정합니다. 또한 모형 업데이트에 사용하는 케이스를 원래 모형을 만드는 데 사용한 케이스와 같은 모집단에서 가져온 것으로 가정합니다. 즉, 연속형 변수의 평균 및 분산, 범주형 변수의 수준이 두 개의 케이스 세트에서 같은 것으로 가정합니다. "새로운" 케이스 세트와 "기존" 케이스 세트를 이질적인 모집단에서 가져온 경우 최상의 결과를 위해서는 케이스 결합 세트에서 이단계 군집 분석 프로시저를 실행해야 합니다.

이단계 군집 분석 결과

결과. 이 그룹은 군집화 결과를 표시할 수 있는 옵션을 제공합니다.

- **피벗 테이블.** 결과는 피벗 테이블에 표시됩니다.
- **모형 뷰어의 도표 및 테이블.** 결과는 모형 뷰어에 표시됩니다.
- **평가 필드.** 이 필드에서는 군집 작성에 사용되지 않은 변수의 군집 데이터를 계산합니다. 모형 뷰어의 입력 기능을 표시 하위 대화 상자에서 선택하여 평가 필드와 함께 표시할 수 있습니다. 결측값이 있는 필드는 무시됩니다.

작업 데이터 파일. 이 그룹을 사용하면 활성 데이터 세트에 변수를 저장할 수 있습니다.

- **소속군집 변수 작성.** 이 변수에는 케이스별 군집 식별 번호가 지정되어 있습니다. 이 변수 이름은 *tsc_n*인데, 여기에서 *n*은 주어진 세션에서 이 프로시저로 활성 데이터 세트를 저장한 순서를 나타내는 양의 정수입니다.

XML 파일. 최종 군집 모형과 CF 나무는 XML 형식으로 내보낼 수 있는 두 가지 유형의 결과 파일입니다.

- **최종 모형 내보내기.** 최종 군집 모형을 XML(PMML) 형식의 지정된 파일로 내보냅니다. 스코어링 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.
- **CF 나무 내보내기.** 이 옵션을 사용하면 클러스터 트리의 현재 상태를 저장하고 나중에 새 데이터를 사용하여 업데이트할 수 있습니다.

군집 뷰어

군집 모형은 일반적으로 검토한 변수를 기준으로 레코드와 유사한 그룹 또는 군집을 찾는 데 사용됩니다. 여기서 동일 그룹 멤버 간의 유사성이 높으며 다른 그룹 멤버 간의 유사성은 낮습니다. 결과를 사용하여 명확하지 않을 수 있는 연관성을 식별할 수 있습니다. 예를 들어 고객 선호도, 수입 수준 및 구매 습관의 군집 분석을 통해 특정 마케팅 캠페인에 반응할 가능성이 높은 고객의 유형을 식별할 수 있습니다.

두 가지 방법으로 군집 출력결과의 결과를 해석할 수 있습니다.

- 군집을 조사하여 해당 군집의 고유한 특성을 파악합니다. 군집 하나에 고수익 차용자가 모두 들어 있습니까? 다른 군집보다 이 군집에 있는 레코드 수가 많습니까?
- 군집에서 필드를 조사하여 군집 내에 값이 분포되는 방법을 파악합니다. 개인의 교육 수준이 군집에서 소속군집을 결정합니까? 높은 신용 스코어가 군집 간의 소속군집 차이를 나타냅니까?

기본 보기와 군집 뷰어의 연결된 다양한 보기를 사용하여 이러한 질문에 대답할 수 있는 정보를 얻을 수 있습니다.

군집 모형에 대한 정보를 보려면 뷰어에서 모형 뷰어 오브젝트를 활성화(두 번 클릭)합니다.

군집 뷰어

군집 뷰어는 2개 패널 즉, 왼쪽에 기본 보기와 오른쪽에 링크 또는 보조 보기로 구성되어 있습니다. 2개의 기본 보기가 있습니다.

- 모형 요약(기본값). 자세한 정보는 188 페이지의 『모형 요약 보기』 주제를 참조하십시오.
- 군집. 자세한 정보는 188 페이지의 『군집 보기』 주제를 참조하십시오.

4개의 링크/보조 보기가 있습니다.

- 예측자 중요도. 자세한 정보는 190 페이지의 『군집 예측변수 중요도 보기』 주제를 참조하십시오.
- 군집 크기(기본값). 자세한 정보는 190 페이지의 『군집 크기 보기』 주제를 참조하십시오.
- 셀 분포. 자세한 정보는 190 페이지의 『셀 분포 보기』 주제를 참조하십시오.
- 군집 비교. 자세한 정보는 190 페이지의 『군집 비교 보기』 주제를 참조하십시오.

모형 요약 보기

모형 요약 보기에는 음영 처리되어 결과를 불량, 양호 또는 우수로 나타내는 군집 결합 및 분리의 실루엣 측도를 비롯하여 군집 모형의 스냅샷 또는 요약이 표시됩니다. 이 스냅샷을 사용하면 결과가 불량한지 신속하게 확인할 수 있습니다. 결과가 불량한 경우 모형화 노드로 돌아가서 군집 모형 설정을 수정하여 향상된 결과를 구할 수 있습니다.

불량, 우수 및 양호 결과는 군집 구조 해석에 관한 Kaufman 및 Rousseeuw(1990) 방법을 기반으로 결정됩니다. 모형 요약 보기에서, 양호 결과는 군집 구조에 대해 적합한 증거 또는 강력한 증거가 있다는 Kaufman 및 Rousseeuw의 평가를 나타내는 데이터, 우수는 약한 증거가 있다는 평가를 나타내는 데이터, 불량은 충분한 증거가 없다는 평가를 나타내는 데이터에 해당합니다.

실루엣 측도는 (B−A) / 최대(A,B)에 따라 전체적인 레코드의 평균을 계산합니다. 여기서 A는 레코드와 군집중심 간의 거리를 나타내며 B는 해당 레코드가 속하지 않는 가장 가까운 군집중심과 레코드 간의 거리를 나타냅니다. 실루엣 계수 1은 모든 케이스가 군집중심에 직접 위치해 있음을 나타냅니다. -1 값은 모든 케이스가 몇몇 다른 군집의 군집중심에 있음을 나타냅니다. 0 값은 평균적으로 케이스가 해당 군집중심과 가장 가까운 다른 군집 간에 등거리에 있음을 나타냅니다.

이 요약에는 다음과 같은 정보가 들어있는 테이블이 있습니다.

- **알고리즘.** 사용된 군집화 알고리즘입니다(예: "이단계").
- **입력 기능.** 여러 개의 필드로서 **입력** 또는 **예측자**라고도 합니다.
- **군집** 해법에 있는 여러 개의 군집입니다.

군집 보기

군집 보기에는 각 군집의 군집 이름, 크기 및 프로파일이 포함된 기능별 군집 격자가 들어 있습니다.

격자의 열에는 다음과 같은 정보가 있습니다.

- **군집.** 알고리즘에 의해 작성된 군집 번호입니다.
- **레이블.** 각 군집에 적용된 모든 레이블입니다(기본적으로 비어 있음). 셀을 두 번 클릭하여 군집 내용을 설명하는 레이블을 입력합니다(예: "고급 승용차 구매자").
- **설명.** 군집중심에 대한 설명입니다(기본적으로 비어 있음). 셀을 두 번 클릭하여 군집에 대한 설명을 입력합니다(예: "$100,000 이상 수입의 55세 이상 전문직 종사자").

- **크기.** 각 군집의 크기입니다(전체 군집 표본에 대한 퍼센트로 표시됨). 격자에 있는 각 크기 셀에는 해당 군집 내에 크기 퍼센트를 표시하는 세로 막대, 숫자 형식의 크기 퍼센트, 군집 케이스 개수가 표시됩니다.
- **기능.** 개별 입력 또는 예측자로서 기본적으로 전체 중요도를 기준으로 정렬됩니다. 크기가 같은 열이 있으면 군집 번호의 오름차순으로 정렬됩니다.

  전체 기능 중요도는 셀 배경 음영 색상으로 표시됩니다. 중요도가 가장 높은 기능은 가장 어둡게 표시되고, 중요도가 가장 낮은 기능은 음영 처리되지 않습니다. 테이블 위에 있는 지침은 각 기능 셀 색상과 연결된 중요도를 나타냅니다.

셀 위에 마우스를 올려 놓으면 기능의 전체 이름/레이블과 셀의 중요도 값이 표시됩니다. 보기 및 기능 유형에 따라 추가 정보가 표시될 수 있습니다. 군집중심 보기에서는 셀 통계 및 셀 값(예:“Mean: 4.32”). 이 추가 정보로 표시됩니다. 범주형 기능의 경우 셀에 빈도가 가장 높은(모달) 범주의 이름과 퍼센트가 표시됩니다.

군집 보기 내에서 다양한 방법을 선택하여 군집 정보를 표시할 수 있습니다.

- 군집 및 기능 전치. 자세한 정보는 189 페이지의 『군집 및 기능 전치』 주제를 참조하십시오.
- 기능 정렬. 자세한 정보는 189 페이지의 『기능 정렬』 주제를 참조하십시오.
- 군집 정렬. 자세한 정보는 189 페이지의 『군집 정렬』 주제를 참조하십시오.
- 셀 내용 선택. 자세한 정보는 189 페이지의 『셀 내용』 주제를 참조하십시오.

군집 및 기능 전치

기본적으로 군집은 열로 표시되고 기능은 행으로 표시됩니다. 이 출력결과를 역으로 바꾸려면 **기능 정렬 기준** 단추 왼족에 있는 **군집 및 기능 전치** 단추를 클릭하십시오. 예를 들어, 표시된 군집 수가 많아 데이터를 보기 위해 가로로 스크롤해야 할 양을 줄이는 경우 이 작업을 수행할 수 있습니다.

기능 정렬

기능 정렬 기준 단추를 사용하여 기능 셀이 표시되는 방법을 선택할 수 있습니다.

- **전체 중요도.** 기본값 정렬 순서입니다. 전체 중요도의 내림차순으로 기능이 정렬되며 정렬 순서는 군집 전체에서 동일합니다. 중요도 값이 같은 기능이 있으면, 기능 이름의 오름차순으로 중요도 값이 같은 기능이 나열됩니다.
- **군집 내 중요도.** 각 군집의 중요도에 대해 기능이 정렬됩니다. 중요도 값이 같은 기능이 있으면, 기능 이름의 오름차순으로 중요도 값이 같은 기능이 나열됩니다. 이 옵션을 선택하면 일반적으로 정렬 순서가 군집 전체에서 달라집니다.
- **이름.** 문자 순의 이름별로 기능이 나열됩니다.
- **데이터 순서.** 데이터 세트에 나열되는 순서대로 기능이 정렬됩니다.

군집 정렬

기본적으로 군집은 크기의 내림차순으로 정렬됩니다. **군집 정렬 기준** 단추를 사용하여 문자순의 이름별로 정렬할 수 있습니다. 고유한 레이블을 작성한 경우 문자숫자형 레이블 순서로 정렬됩니다.

동일한 레이블이 있는 기능은 군집 이름별로 정렬됩니다. 군집이 레이블별로 정렬되어 있고 군집의 레이블을 편집하는 경우 정렬 순서가 자동으로 업데이트됩니다.

셀 내용

셀 단추를 사용하여 기능 및 평가 필드에 대한 셀 내용의 출력결과를 변경할 수 있습니다.

- **군집중심.** 기본적으로, 셀에는 기능 이름/레이블과 각 군집/기능 조합에 대한 중심 경향이 표시됩니다. 연속형 필드에는 평균이 표시되며 범주형 필드에는 범주 퍼센트와 함께 최빈값(가장 자주 발생하는 범주)이 표시됩니다.
- **절대분포.** 기능 이름/레이블과 각 군집 내에 있는 기능의 절대 분포가 표시됩니다. 범주형 기능의 경우 데이터 값의 오름차순으로 정렬된 범주가 겹쳐진 막대도표가 표시됩니다. 연속형 기능의 경우 각 군집에 동일한 엔드포인트와 구간을 사용하는 매끄러운 평활 밀도 도표가 표시됩니다.

  진한 빨간색 출력결과에는 군집 분포가 표시되고 연한 빨간색 출력결과에는 전체 데이터가 표시됩니다.

- **상대분포.** 기능 이름/레이블과 상대분포가 셀에 표시됩니다. 일반적으로 이 출력결과에는 상대분포가 표시된다는 점을 제외하고 절대분포의 출력결과와 비슷합니다.

진한 빨간색 출력결과에는 군집 분포가 표시되고 연한 빨간색 출력결과에는 전체 데이터가 표시됩니다.

- **기본 보기.** 군집 수가 많은 경우 스크롤을 하지 않고 모든 세부사항을 보기 어려울 수 있습니다. 스크롤하는 양을 줄이려면 이 보기를 선택하여 보다 간결한 형태의 테이블로 출력결과를 변경하십시오.

군집 예측변수 중요도 보기

예측자 중요도 보기는 모형 추정에서의 각 필드의 상대적 중요도를 보여줍니다.

군집 크기 보기

군집 크기 보기에는 각 군집이 들어 있는 원도표가 표시됩니다. 원도표의 조각마다 각 군집의 퍼센트 크기가 표시됩니다. 각 조각에 마우스를 올려 놓으면 해당 조각 내의 개수가 표시됩니다.

도표 아래에 다음과 같은 크기 정보가 나열된 테이블이 있습니다.

- 가장 작은 군집의 크기(전체 개수 및 퍼센트)
- 가장 큰 군집의 크기(전체 개수 및 퍼센트)
- 가장 큰 군집의 크기 대 가장 작은 군집의 크기의 비율

셀 분포 보기

셀 분포에는 군집 기본 패널의 테이블에서 선택하는 모든 기능 셀에 대해 확장되어 보다 자세한 데이터 분포의 도표가 표시됩니다.

군집 비교 보기

군집 비교 보기는 격자 스타일의 레이아웃으로 구성되며 행에 기능이 있고 열에서 군집이 선택되어 있습니다. 이 보기를 사용하면 군집을 구성하는 요인을 보다 잘 이해할 수 있습니다. 또한 전체 데이터는 물론 개별 데이터를 비교하여 군집 간의 차이점을 확인할 수 있습니다.

표시할 군집을 선택하려면 군집 기본 패널에서 군집 열 맨 위를 클릭합니다. Ctrl 키 또는 Shift 키를 누른 채 마우스 단추를 클릭하여 비교할 군집을 둘 이상 선택하거나 선택 해제합니다.

참고: 표시할 군집을 최대 5개까지 선택할 수 있습니다.

선택한 순서대로 군집이 표시됩니다. 필드 순서는 **기능 정렬 기준** 옵션에 의해 결정됩니다. **군집 내 중요도**를 선택하면 필드가 항상 전체 중요도별로 정렬됩니다.

배경 도표에 각 기능의 전체 분포가 표시됩니다.

- 범주형 기능은 점도표로 표시됩니다. 점 크기는 각 군집에 대해 빈도가 가장 높은/모달 범주를 나타냅니다(기능별).
- 연속형 기능은 상자도표로 표시됩니다. 이 도표에는 전체 중위수 및 사분위수 범위가 표시됩니다.

이러한 배경 보기에 겹쳐그려진 항목은 선택된 군집의 상자도표입니다.

- 연속형 기능의 경우, 사격형 마커와 가로 선은 각 군집의 중위수 및 사분위수 범위를 나타냅니다.
- 각 군집엔 보기 맨 위에 다른 색상으로 표시됩니다.

군집 뷰어 탐색

군집 뷰어는 대화식 출력결과입니다. 다음을 수행할 수 있습니다.

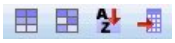
- 세부사항을 보려면 필드 또는 클러스터를 선택하십시오.
- 원하는 항목을 선택하여 군집을 비교할 수 있습니다.
- 출력결과를 변경할 수 있습니다.
- 축을 이동합니다.

도구 모음 사용

도구 모음 옵션을 사용하여 왼쪽 및 오른쪽 패널에 표시되는 정보를 제어할 수 있습니다. 도구 모음 제어를 사용하여 출력결과의 방향(위에서 아래, 왼쪽에서 오른쪽, 오른쪽에서 왼쪽)을 바꿀 수 있습니다. 또한 뷰어를 기본값 설정으로 재설정하고 대화 상자를 열어 기본 패널에 군집 보기 내용을 지정할 수 있습니다.

기능 정렬 기준, **군집 정렬 기준**, **셀** 및 **출력결과** 옵션은 기본 패널에서 **군집** 보기를 선택한 경우에만 사용할 수 있습니다. 자세한 정보는 188 페이지의 『군집 보기』 주제를 참조하십시오.

표 *4.* 도구 모음 아이콘

아이콘	**토픽**
	군집 및 기능 전치 참조
	기능 정렬 기준 참조
	S군집 정렬 기준 참조
	셀 참조

군집 보기 출력결과 제어

기본 패널에서 군집 보기에 표시되는 내용을 제어하려면 **출력결과** 단추를 클릭합니다. 그러면 대화 상자가 열립니다.

기능. 기본적으로 선택됩니다. 모든 입력 기능을 숨기려면 확인 상자를 선택 해제하십시오.

평가 필드. 표시할 평가 필드(군집 모형을 만드는 데 사용되지 않았지만 군집을 평가하기 위해 모형 뷰어에 보낸 필드)를 선택하십시오. 기본적으로 아무 것도 표시되지 않습니다. 참고 평가 필드는 두 개 이상의 값이 포함된 문자열이어야 합니다. 사용 가능한 평가 필드가 없는 경우 이 선택란을 사용할 수 없습니다.

군집 설명. 기본적으로 선택됩니다. 모든 군집 설명을 숨기려면 확인 상자를 선택 해제하십시오.

군집 크기. 기본적으로 선택됩니다. 모든 군집 크기 셀을 숨기려면 확인 상자를 선택 해제하십시오.

최대 범주 수. 범주형 기능의 도표에 표시할 최대 범주 수를 지정합니다. 기본값은 20입니다.

레코드 필터링

특정 군집 또는 군집 그룹의 케이스에 대해 더 알고 싶은 경우 선택된 군집을 기준으로 추가 분석을 위한 레코드 서브세트를 선택할 수 있습니다.

1. 군집 뷰어의 군집 보기에서 군집을 선택하십시오. 다중 군집을 선택하려면 Ctrl 키와 함께 클릭합니다.
2. 메뉴에서 다음을 선택합니다.

   생성 > 레코드 필터링...
3. 필터변수 이름을 입력합니다. 선택된 군집의 레코드는 이 필드에 대한 값으로 1을 받습니다. 필터 상태를 변경할 때까지 다른 모든 레코드는 값으로 0을 받으며 이어지는 분석에서 제외됩니다.
4. **확인**을 클릭하십시오.

계층적 군집 분석

이 프로시저를 사용하면 선택한 특성에 따라 비교적 동일한 케이스(또는 변수) 그룹을 식별하며 별도 군집의 각 케이스(또는 변수)에서 시작하여 하나의 케이스가 남을 때까지 군집을 조합하는 알고리즘을 사용합니다. 원시변수를 분석하거나 다양한 표준 변환에서 선택할 수 있습니다. 근접성 프로시저로 거리 측도나 유사성 측도를 작성합니다. 각 단계마다 통계가 표시되므로 가장 적절한 선택을 할 수 있습니다.

예. 각 그룹에서 비슷한 시청률을 나타내는 텔레비전 쇼 그룹을 구분할 수 있습니까? 계층적 군집 분석을 통해 시청자의 특성에 따라 텔레비전 쇼(케이스)를 동일한 그룹으로 묶을 수 있습니다. 이러한 분석은 마케팅을 위한 대상 선정에도 사용됩니다. 다양한 마케팅 전략을 검정하는데 비교 도시들을 선택할 수 있도록 도시(케이스)를 동일 그룹으로 모을 수 있습니다.

통계. 군집화 스케줄, 거리 행렬(또는 유사성 행렬), 단일 해법이나 해법범위에 대한 소속군집을 선택할 수 있습니다. 도표: 덴드로그램 및 고드름 도표.

계층적 군집 분석 데이터 고려 사항

데이터. 변수는 양적, 이분형, 개수 데이터가 될 수 있습니다. 척도의 차이가 군집 솔루션에 영향을 줄 수 있으므로 변수 척도화는 중요한 문제가 됩니다. 변수의 척도 차이가 큰 경우(예를 들어, 한 변수는 달러로 측정되고 다른 변수는 연도로 측정되는 경우) 표준화를 고려하게 됩니다. 이러한 표준화는 계층적 군집 분석 프로시저에서 자동으로 실행됩니다.

케이스 순서. 입력 데이터에 동률 거리 또는 유사성이 있거나 결합 과정에서 업데이트된 군집 중에 동률 거리 또는 유사성이 발생하는 경우 군집 솔루션의 결과는 파일에 있는 케이스 순서에 따라 달라질 수 있습니다. 주어진 해법의 안정성을 확인하기 위해 케이스를 각기 다른 무작위 순서로 정렬하여 서로 다른 여러 가지 해법을 구할 수도 있습니다.

가정. 거리 측도나 유사성 측도를 적절하게 사용하여 데이터를 분석합니다. 거리 측도 및 유사성 측도에 대한 자세한 내용은 근접성 프로시저를 참조하십시오. 또한 분석 시 관련 변수를 모두 사용해야 합니다. 영향력 있는 변수가 빠지면 해법이 틀릴 수 있습니다. 계층적 군집 분석은 실험적인 방법이므로 그에 따른 결과는 독립 표본에서 확인될 때까지 잠정적인 결과로 다뤄야 합니다.

계층적 군집 분석 결과 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 분류 > 계층적 군집...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 케이스를 묶을 때는 하나 이상의 숫자변수를 선택합니다. 변수를 묶을 때는 셋 이상의 숫자변수를 선택합니다.

선택적으로 케이스를 설명할 식별 기준변수를 선택합니다.

계층적 군집 분석

군집방법. 그룹-간 링크, 그룹-내 링크, 가장 가까운 항목, 가장 먼 항목, 중심값 군집화, 중위수 군집화, Ward의 방법을 사용할 수 있습니다.

측도. 거리 측도나 유사성 측도를 지정하여 군집화에 사용할 수 있습니다. 다음 데이터 유형과 적합한 거리 측도 또는 유사성 측도를 선택합니다.

- **구간척도.** 유클리드 거리, 유클리드 거리 제곱값, 코사인, Pearson 상관계수, Chebychev, 블록, Minkowski, 사용자 정의를 사용할 수 있습니다.
- **개수.** 카이제곱 측도와 파이제곱 측도를 사용할 수 있습니다.
- **이분형.** 유클리드 거리, 유클리드 거리 제곱값, 크기 차이, 패턴 차이, 분산, 산포, 형태, 단순 일치, 파이 4-점 상관, 람다, Anderberg의 *D*, Dice, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance와 Williams, Ochiai, Rogers와 Tanimoto, Russel과 Rao, Sokal과 Sneath 1, Sokal과 Sneath 2, Sokal과 Sneath 3, Sokal과 Sneath 4, Sokal과 Sneath 5, Yule의 *Y*, Yule의 *Q*를 사용할 수 있습니다.

값 변환. 근접성을 계산하기 전에 각 케이스나 값에 대한 데이터 값을 표준화할 수 있습니다(이분형 데이터에는 사용할 수 없음). 표준화 방법으로는 *z* 스코어, -1부터 1까지 범위, 0부터 1까지 범위, 최대값 1, 평균 1, 표준 편차 1 등을 사용할 수 있습니다.

측도변환. 거리 측도에 의해 생성된 값을 변환할 수 있습니다. 이 값은 거리 측도를 계산한 다음 적용됩니다. 절대값, 부호 바꾸기, 0–1 범위로 척도 조정을 사용할 수 있습니다.

계층적 군집 분석: 통계

군집화 스케줄. 각 단계마다 결합된 케이스나 군집, 결합되고 있는 케이스나 군집 간 거리, 케이스나 변수가 군집을 형성하는 마지막 군집 수준을 표시합니다.

근접 행렬. 항목 간 거리나 유사성을 적용합니다.

소속군집. 군집 조합에서 각 케이스가 하나 이상의 단계에 할당된 군집을 표시합니다. 사용할 수 있는 옵션은 단일 해법과 해법범위입니다.

계층적 군집 분석: 도표

덴드로그램. 덴드로그램을 표시합니다. 덴드로그램을 사용하여 군집의 결속을 강화하고 적절한 군집 수를 유지하기 위한 정보를 얻을 수 있습니다.

고드름. 모든 군집이나 지정한 범위 내 군집을 포함하는 고드름 도표를 표시합니다. 고드름 도표는 분석의 각 반복마다 케이스가 군집으로 조합되는 방법에 대한 정보를 표시합니다. 출력방향에서 수직 도표나 수평 도표를 선택할 수 있습니다.

계층적 군집 분석: 새로운 변수로 저장

소속군집. 단일 해법이나 해법범위에 대한 소속군집을 저장할 수 있습니다. 후속 분석에서 저장된 변수를 사용하여 그룹 간 차이를 탐색할 수 있습니다.

CLUSTER 명령 구문 추가 기능

계층적 군집 분석 절차에서는 CLUSTER 명령 구문을 사용합니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 여러가지 군집 분석 방법이 하나의 분석에서 사용됩니다.
- 근접 행렬을 읽고 분석합니다.
- 나중에 분석하기 위해 근접 행렬을 디스크에 기록합니다.
- 사용자 정의(거듭제곱) 거리 측도에서 거듭제곱 및 루트 값을 지정합니다.
- 저장된 변수에 대한 이름을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

K-Means 군집 분석

이 프로시저를 사용하면 여러 개의 케이스를 다루는 알고리즘을 통해 선택한 특성을 기준하여 상대적으로 동질적인 케이스 그룹을 구별할 수 있습니다. 그러나 이러한 알고리즘을 사용하려면 군집의 수를 지정해야 합니다. 이러한 정보를 알 경우에는 군집중심초기값을 지정할 수 있습니다. 케이스를 분류하려면 반복적으로 군집중심을 업데이트하거나 분류만 하는 두 가지 방법 중 하나를 선택할 수 있습니다. 소속군집, 거리 정보, 마지막 군집중심 등을 저장할 수 있습니다. 선택적으로 케이스별 결과를 설명하는 데 사용되는 값을 갖는 변수를 지정할 수 있습니다. 분산 분석 *F* 통계가 필요할 수도 있습니다. 이러한 통계는 필요에 따라 사용할 수 있으며(이 프로시저는 서로 다른 그룹을 구성하려고 함) 통계의 상대적 크기로 그룹 분리에 각 변수가 미치는 영향을 알 수 있습니다.

예. 각 그룹에서 비슷한 시청자 유형을 갖는 텔레비전 쇼를 구별하는 그룹은 무엇입니까? *K*-평균 군집 분석을 사용하면 텔레비전 쇼(케이스)를 시청자의 특성에 따라 *K* 동일성 그룹으로 모을 수 있습니다. 이 방법은 마케팅을 위한 대상 선정에도 사용됩니다. 다양한 마케팅 전략을 검정하는데 비교 도시들을 선택할 수 있도록 도시(케이스)를 동일 그룹으로 모을 수 있습니다.

통계. 완료 해법에 대해 군집중심초기값, 분산 분석표를 선택할 수 있습니다. 각 케이스에 대해 군집 정보, 군집중심으로부터의 거리를 선택할 수 있습니다.

K-Means 군집 분석 데이터 고려 사항

데이터. 변수는 구간 수준이나 비율 수준에서 양적변수이어야 합니다. 사용 변수가 이분형이나 개수일 경우 계층적 군집 분석 프로시저를 사용합니다.

케이스 및 군집중심 초기값 순서. 군집중심 초기값을 선택하는 기본값 알고리즘은 케이스 순서에 따라 달라질 수 있습니다. 반복 대화 상자의 **유동계산 평균 사용** 옵션을 사용하면 군집중심 초기값의 선택 방법과 상관없이 케이스 순서에 따라 결과 해법이 달라질 수 있습니다. 이러한 방법 중 하나를 사용하는 경우 주어진 해법의 안정성을 확인하기 위해 케이스를 각기 다른 무작위 순서로 정렬하여 서로 다른 여러 가지 해법을 구할 수도 있습니다. 군집중심 초기값을 지정하고 **유동계산 평균 사용** 옵션을 사용하지 않으면 케이스 순서와 관련된 문제를 방지할 수 있습니다. 그러나 케이스에서 군집중심까지의 거리가 동률인 경우 군집중심 초기값의 순서가 해법에 영향을 줄

수 있습니다. 주어진 해법의 안정성을 평가하기 위해 중심 초기값의 순열을 각기 달리하여 실행한 분석 결과를 비교할 수 있습니다.

가정. 거리는 단순 유클리드 거리를 사용하여 계산합니다. 다른 거리나 유사성 측도를 사용하려면 계층적 군집 분석 프로시저를 사용합니다. 변수 척도화는 특히 중요한 고려 사항입니다. 한 변수는 달러로 표시되고 다른 변수는 연도로 표시되는 경우와 같이 변수가 서로 다른 척도로 측정되면 결과가 잘못될 수 있습니다. 이러한 경우에는 변수를 표준화한 후 *K*-평균 군집 분석을 수행하도록 합니다. 이 작업은 기술통계 프로시저에서 수행할 수 있습니다. 이 프로시저에서는 적합한 군집의 수를 선택했으며 관련 변수가 모두 포함되었다고 가정합니다. 선택한 군집의 수가 적합하지 않거나 중요 변수가 빠졌을 때는 결과가 잘못될 수 있습니다.

K-Means 군집 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 분류 > K-평균 군집...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 군집 분석에 사용할 변수를 선택합니다.
3. 군집의 수를 지정합니다. 군집 수는 최소 두 개이어야 하고 데이터 파일에 있는 케이스 수보다 작거나 같아야 합니다.
4. **반복 및 분류하기** 또는 **분류만 하기** 중 하나를 선택합니다.
5. 필요한 경우 케이스를 설명할 식별변수를 선택합니다.

K-Means 군집 분석 효율성

K-평균 군집 분석 명령은 계층적 군집 명령으로 사용한 군집화 알고리즘과 같이 모든 대응 케이스 사이의 거리를 계산하지 않으므로 효율적입니다.

최대 효율성을 위해 표본 케이스를 선택한 후 **반복 및 분류하기** 방법을 선택하여 군집중심을 결정합니다. **최종값 저장**을 선택합니다. 그런 다음 전체 데이터 파일을 복구하고 **분류만 하기** 방법을 선택한 후 **초기값 읽기**를 선택하여 표본에서 추정한 중심을 사용하여 전체 파일을 분류합니다. 파일이나 데이터 세트에서 읽고 쓸 수 있습니다. 데이터 세트는 같은 세션에서 다시 사용할 수 있으나 세션을 마치기 전에 명시적으로 저장하지 않는 한 파일로 저장되지 않습니다. 데이터 세트 이름은 변수 명명 규칙을 따라야 합니다. 자세한 정보는 주제를 참조하십시오.

K-Means 군집 분석: 반복

참고: 이 옵션은 K-Means 군집 분석 대화 상자에서 **반복 및 분류하기**를 선택한 경우에만 사용할 수 있습니다.

최대반복수. *K*-평균 알고리즘에서 반복수를 제한합니다. 이 수만큼 반복한 다음에는 수렴 기준이 만족되지 않더라도 중지됩니다. 1부터 999 사이의 값이어야 합니다.

5.0 이전 버전에서 Quick Cluster 명령으로 알고리즘을 다시 생성하려면 **최대반복수**를 1로 설정합니다.

수렴 기준. 반복을 중단할 시기를 결정합니다. 초기 군집 중심 간 최소 거리의 비율을 나타내므로 0보다 크지만 1보다 크지 않아야 합니다. 기준이 0.02와 같으면, 예를 들어, 완전한 반복이 임의의 초기 군집 중심 사이의 최소 거리의 2%를 초과하는 거리만큼 군집 중심을 이동하지 않을 때 반복이 중단됩니다.

유동계산 평균 사용. 군집중심을 각 케이스를 할당한 다음 업데이트하도록 합니다. 이 옵션을 선택하지 않으면 케이스를 모두 할당한 다음 새 군집중심을 계산하게 됩니다.

K-Means 군집 분석: 저장

이후 분석에서 사용하도록 해법에 대한 정보를 새로운 변수로 저장할 수 있습니다.

소속군집. 각 케이스의 마지막 소속군집을 가리키는 새로운 변수를 작성합니다. 새로운 변수 값의 범위는 1부터 군집 수까지입니다.

군집중심으로부터의 거리. 각 케이스와 분류 중심 사이의 유클리드 거리를 가리키는 새로운 변수를 작성합니다.

K-Means 군집 분석: 옵션

통계. 각 케이스에 대해 군집중심초기값, 분산 분석표 및 군집 정보를 선택할 수 있습니다.

- 초기 군집 중심. 각 군집에 대한 변수 평균의 첫 번째 추정값입니다. 기본적으로 적절한 간격으로 떨어져 있는 케이스를 군집 수와 동일하게 데이터에서 선택합니다. 군집중심초기값은 분류의 첫 번째 반올림에 사용된 다음 업데이트됩니다.
- 분산 분석표*(ANOVA table)*. 각 군집변수에 대한 일변량 F 검정을 포함하는 분산 분석표를 표시합니다. F 검정은 기술통계일 뿐이므로 결과로 나타나는 확률을 해석할 수 없습니다. 전체 케이스가 단일 군집에 할당되면 분산 분석표가 표시되지 않습니다.
- 각 케이스에 대한 군집 정보. 각 케이스에 대해 최종 군집 할당 및 케이스를 분류하는 데 사용되는 군집 중심과 케이스 간 유클리디안 거리를 표시합니다. 최종 군집 중심 간 유클리드 거리도 표시합니다.

결측값. 사용 가능한 옵션은 **목록별 결측값 제외** 또는 **대응별 결측값 제외**입니다.

- **목록별 결측값 제외.** 분석에서 군집 변수에 대한 결측값이 있는 케이스를 제외합니다.
- **대응별 결측값 제외.** 비결측값을 가지는 모든 변수에서 계산된 거리에 따라 케이스를 군집에 할당합니다.

QUICK CLUSTER 명령 추가 기능

K-평균 클러스터 프로시저는 QUICK CLUSTER 명령 구문을 사용합니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 첫 번째 *k* 케이스를 초기 군집중심으로 사용하므로 일반적으로 추정하는 데 사용되는 데이터 전달을 피할 수 있습니다.
- 초기 군집중심을 명령 구문의 일부로 직접 지정합니다.
- 저장된 변수에 대한 이름을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

비모수 검정

비모수 검정은 데이터의 기본 분포에 대한 최소한의 가정을 생성합니다. 이 대화 상자에 있는 검정은 데이터가 구성되어 있는 방법에 따라 세 가지 광범위한 범주(일표본 검정, 대응표본 검정, 독립 표본 검정)로

- 일표본 검정은 1개 필드를 분석합니다.
- 대응표본 검정은 동일한 케이스 세트에 대해 2개 이상의 필드를 비교합니다.
- 독립 표본 검정은 다른 필드의 범주로 그룹화되어 있는 필드 1개를 분석합니다.

일표본 비모수 검정

일표본 비모수 검정은 하나 이상의 비모수 검정을 사용하여 단일 필드에서 차이를 식별합니다. 비모수 검정에서는 데이터가 정규 분포를 따른다고 가정하지 않습니다.

사용자의 목표. 목표를 사용하면 다르지만 일반적으로 사용되는 검정 설정을 신속하게 지정할 수 있습니다.

- **관측 데이터와 가설 값 자동으로 비교.** 이 목표는 범주가 2개만 있는 범주형 필드에 이항검정을 적용하고, 기타 모든 범주형 필드에 카이제곱 검정을 적용하며, 연속형 필드에 콜모고로프-스미르노프 검정을 적용합니다.
- **무작위 시퀀스로 검정.** 이 목표에서는 런 검정을 사용하여 무작위로 관측된 데이터 값 시퀀스를 검정합니다.
- **사용자 정의 분석.** 설정 탭에서 검정 설정을 수동으로 변경하려면 이 옵션을 선택하십시오. 이후에 설정 탭에서 옵션을 변경한 후 이 변경 사항을 현재 선택된 목표와 함께 사용할 수 없는 경우 이 설정이 자동으로 선택됩니다.

일표본 비모수 검정 구하기

메뉴에서 다음을 선택합니다.

분석 > 비모수 검정 > 일표본...

1. **실행**을 클릭합니다.

선택적으로 다음을 수행할 수 있습니다.

- 목적 탭에서 목적을 지정합니다.
- 필드 탭에서 필드 할당을 지정합니다.
- 설정 탭에서 전문가 설정을 지정하십시오.

필드 탭

이 필드 탭에서는 검정할 필드를 지정합니다.

사전 정의된 역할 사용. 이 옵션을 기존 필드 정보를 사용합니다. 입력, 대상 또는 모두로 역할이 사전 정의된 모든 필드가 검정 필드로 사용됩니다. 하나 이상의 검정 필드가 필요합니다.

사용자 정의 필드 할당 사용. 이 옵션을 사용하여 필드 역할을 바꿀 수 있습니다. 이 옵션을 선택한 후 아래 필드를 지정합니다.

- **검정 필드.** 하나 이상의 필드를 선택합니다.

설정 탭

설정 탭은 몇 가지 다른 설정 그룹으로 구성되어 있으며 설정을 수정하여 알고리즘이 데이터를 처리하는 방법을 미세 튜닝할 수 있습니다. 기본값 설정을 현재 선택된 목표와 함께 사용할 수 없는 설정으로 변경하면, 목표 탭이 자동으로 업데이트되어 **분석 사용자 정의** 옵션이 선택됩니다.

검정 선택

이 설정은 필드 탭에 지정된 필드에서 수행되는 검정을 지정합니다.

데이터를 기준으로 검정 자동 선택. 이 설정은 유효한(결측되지 않은) 범주가 2개만 있는 범주형 필드에 이항검정을 적용하고, 기타 모든 범주형 필드에 카이제곱 검정을 적용하며, 연속형 필드에 콜모고로프-스미르노프 검정을 적용합니다.

검정 사용자 정의. 이 설정을 사용하면 수행되는 특정 검정을 선택할 수 있습니다.

- **관측 이분형 확률과 가설 값 비교(이항검정).** 이항검정은 모든 필드에 적용할 수 있습니다. 이 옵션은 플래그 필드(범주가 2개만 있는 범주형 필드)의 관측 분포가 지정된 이항 분포에서 기대한 것과 동일한지 검정하는 일표본 검정을 생성합니다. 또한 신뢰구간을 요청할 수 있습니다. 검정 설정에 대한 자세한 내용은 196 페이지의 『이항검정 옵션』의 내용을 참조하십시오.
- **관측 확률과 가설 값 비교(카이제곱 검정).** 카이제곱 검정은 명목 및 순서 필드에 적용됩니다. 이 옵션은 필드 범주의 관측 빈도와 기대 빈도 간의 차이를 기준으로 카이제곱 통계를 계산하는 일표본 검정을 생성합니다. 검정 설정에 대한 자세한 내용은 197 페이지의 『카이제곱 검정 옵션』의 내용을 참조하십시오.
- **관측 분포를 가설 값에 검정(콜모고로프-스미르노프 검정).** 콜모고로프-스미르노프 검정은 연속형 및 순서 필드에 적용됩니다. 이 옵션은 필드의 표본 누적 분포 함수가 정규, 균일, 포아송 또는 지수 분포와 동질적인지를 검정하는 일표본 검정을 생성합니다. 검정 설정에 대한 자세한 내용은 197 페이지의 『Kolmogorov-Smirnov 옵션』의 내용을 참조하십시오.
- **중위수와 가설 값 비교(윌콕슨 부호순위 검정).** Wilcoxon 부호 순위 검정은 연속형 및 순서 필드에 적용됩니다. 이 옵션은 필드의 중위수에 대한 일표본 검정을 생성합니다. 숫자로 가설 중위수를 지정합니다.
- **무작위 시퀀스로 검정(런 검정).** 런 검정은 모든 필드에 적용됩니다. 이 옵션은 이분화 필드의 값 시퀀스가 무작위인지 검정하는 일표본 검정을 생성합니다. 검정 설정에 대한 자세한 내용은 197 페이지의 『런 검정 옵션』의 내용을 참조하십시오.

이항검정 옵션

이항검정은 플래그 필드(범주가 2개만 있는 범주형 필드)에 적용되지만 "성공" 정의 규칙을 사용하면 모든 필드에 적용할 수 있습니다.

가설 비율. "성공" 또는 *p*로 정의된 레코드의 예상 비율을 지정합니다. 0보다 크고 1보다 작은 값을 지정하십시오. 기본값은 0.5입니다.

신뢰구간. 다음과 같은 방법을 사용하여 이분형 데이터의 신뢰구간을 계산할 수 있습니다.

- **클로퍼-피어슨(정확).** 누적 이항 분포를 기준으로 한 정확한 구간입니다.

- **제프리스.** 이전의 제프리스를 사용하는 *p*의 사후 분포를 기준으로 한 베이지언 구간입니다.
- **우도비.** *p*의 우도 함수를 기준으로 한 구간입니다.

범주형 필드의 성공 정의. 이 옵션은 가설 비율에 대해 검정된 데이터 값인 "성공"이 범주형 필드에 대해 정의되는 방법을 지정합니다.

- **데이터의 첫 번째 범주 사용**은 표본에 있는 첫 번째 값으로 이항검정을 수행하여 "성공"을 정의합니다. 이 옵션은 값이 2개만 있는 명목 또는 순서 필드에만 적용됩니다. 이 옵션이 사용되는 필드 탭에 지정된 기타 모든 범주형 필드는 검정되지 않습니다. 이는 기본값입니다.
- **성공값 지정**은 지정된 값 목록으로 이항검정을 수행하여 "성공"을 정의합니다. 문자 또는 숫자 값 목록을 지정합니다. 목록에 있는 값이 표본에 없어도 됩니다.

연속형 필드의 성공 정의. 이 옵션은 검정 값에 대해 검정된 데이터 값인 "성공"이 연속형 필드에 대해 정의되는 방법을 지정합니다. 성공은 절단점 이하 값으로 정의됩니다.

- **표본 중심점**은 최소값 및 최대값의 평균에서 절단점을 설정합니다.
- **사용자 정의 절단점**을 사용하여 절단점의 값을 지정할 수 있습니다.

카이제곱 검정 옵션

모든 범주 확률 동일. 이 옵션은 표본에 있는 모든 범주에 같은 빈도를 생성합니다. 이는 기본값입니다.

기대 확률 사용자 정의. 이 옵션을 사용하여 지정된 범주 목록에 같지 않은 빈도를 지정할 수 있습니다. 문자 또는 숫자 값 목록을 지정합니다. 목록에 있는 값이 표본에 없어도 됩니다. **범주** 열에서 범주 값을 지정합니다. **상대적 빈도** 열에서 각 범주에 0보다 큰 값을 지정합니다. 사용자 정의 빈도는 비율로 취급되므로 예를 들어 빈도 1, 2, 3을 지정하는 것은 빈도 10, 20, 및 30을 지정하는 것과 같으며, 두 경우 모두 레코드의 1/6은 첫 번째 범주에, 1/3은 두 번째 범주에, 1/2은 세 번째 범주에 포함되도록 지정합니다. 사용자 정의 기대 확률이 지정되면, 사용자 정의 범주 값에는 해당 데이터의 모든 필드 값이 포함되어야 합니다. 그렇지 않으면 해당 필드에 검정이 수행되지 않습니다.

Kolmogorov-Smirnov 옵션

이 대화 상자에서는 검정할 분포와 가설 분포의 모수를 지정합니다.

분포의 특정 모수를 표본에서 추정해야 할 경우 Kolmogorov-Smirnov 검정이 더 이상 적용되지 않습니다. 이 경우, Lilliefors 검정 통계량을 사용하여 *p*-값을 추정할 수 있는데, 이를 위해서는 Monte Carlo 표본추출을 사용하여 알 수 없는 평균과 분산을 갖는 정규성을 검정하십시오. Lilliefors 검정은 세 개의 연속 분포(**정규**, **지수**, **균일**)에 적용됩니다. 기본 검정이 이산형(**포아송**)일 경우에는 이 검정이 적용되지 않습니다. 해당하는 분포모수가 지정되지 않은 경우에는 일표본 추론에 대해서만 검정이 정의됩니다.

정규
: **표본 데이터 사용**에 관측 평균과 표준편차가 사용되고 기존 **근사 검정** 결과를 선택하는 옵션이 제공되거나, **Monte Carlo 표본추출을 기반으로 하는 Lilliefors 검정**이 사용됩니다. **사용자 정의**를 사용하여 값을 지정할 수 있습니다.

균일
: **표본 데이터 사용**에 관측 최소값 및 최대값이 사용되고, Monte Carlo 표본추출을 기반으로 하는 Lilliefors 검정이 사용됩니다. **사용자 정의**를 사용하여 최소값과 최대값을 지정할 수 있습니다.

지수
: **표본 평균**에 관측 평균이 사용되고 Monte Carlo 표본추출을 기반으로 하는 Lilliefors 검정이 사용됩니다. **사용자 정의**를 사용하여 관측 평균 값을 지정할 수 있습니다.

포아송
: **평균**을 사용하여 관측 평균 값을 지정할 수 있습니다.

런 검정 옵션

런 검정은 플래그 필드(범주가 2개만 있는 범주형 필드)에 적용되지만 그룹 정의 규칙을 사용하면 모든 필드에 적용할 수 있습니다.

범주형 필드의 그룹 정의. 다음 옵션을 사용할 수 있습니다.

- **표본에 범주 2개만 있음**은 표본에 있는 값으로 런 검정을 수행하여 그룹을 정의합니다. 이 옵션은 값이 2개만 있는 명목 또는 순서 필드에만 적용됩니다. 이 옵션이 사용되는 필드 탭에 지정된 기타 모든 범주형 필드는 검정되지 않습니다.
- **2개 범주로 데이터 코딩변경**은 지정된 값 목록으로 런 검정을 수행하여 이러한 그룹 중 하나를 정의합니다. 표본에 있는 기타 모든 값은 다른 그룹을 정의합니다. 목록에 있는 값이 모두 표본에 표시될 필요는 없지만, 하나 이상의 레코드가 각 그룹에 있어야 합니다.

연속형 필드의 절단점 정의. 이 옵션은 연속형 필드에 그룹이 정의되는 방법을 지정합니다. 첫 번째 그룹은 절단점 이하 값으로 정의됩니다.

- **표본 중위수**은 표본 중위수에서 절단점을 설정합니다.
- **표본 평균**은 표본 평균에서 절단점을 설정합니다.
- **사용자 정의**를 사용하여 절단점의 값을 지정할 수 있습니다.

검정 옵션

유의수준
이 옵션은 모든 검정의 유의수준(알파)을 지정합니다. 0과 1 사이의 숫자 값을 지정합니다. 기본값은 0.05입니다.

신뢰구간(%)
이 옵션은 생성된 모든 신뢰구간의 신뢰수준을 지정합니다. 0-100 사이의 숫자 값을 지정하십시오. 기본값은 95입니다.

제외된 케이스
이 옵션은 검정의 케이스 기준을 결정하는 방법을 지정합니다.

검정별 결측값 제외
특정 검정에 사용되는 필드에 대해 결측값이 있는 레코드가 해당 검정에서 생략됩니다. 분석에 여러 개의 검정이 지정되어 있는 경우 각 검정이 별도로 평가됩니다.

목록별 결측값 제외
필드 탭에 이름이 지정된 필드에 대해 결측값이 있는 레코드가 모든 분석에서 제외됩니다.

Monte Carlo 표본추출

분포의 특정 모수를 표본에서 추정해야 할 경우 Kolmogorov-Smirnov 검정이 더 이상 적용되지 않습니다. 이 경우, Lilliefors 검정 통계량을 사용하여 *p*-값을 추정할 수 있는데, 이를 위해서는 Monte Carlo 표본추출을 사용하여 알 수 없는 평균과 분산을 갖는 정규성을 검정하십시오. Lilliefors 검정은 세 개의 연속 분포(**정규**, **지수**, **균일**)에 적용됩니다. 기본 검정이 이산형(**포아송**)일 경우에는 이 검정이 적용되지 않습니다. 해당하는 분포모수가 지정되지 않은 경우에는 일표본 추론에 대해서만 검정이 정의됩니다.

사용자 정의 난수 시작값 설정
사용으로 설정할 경우, 이 설정은 Monte Carlo 표본추출에 사용되는 난수 **시작값**을 재설정하는 옵션을 제공합니다. 값은 1 - 2,147,483,647 사이의 단일 정수여야 합니다(기본값은 2,000,000). 기본적으로 설정은 해제되어 있으며, 난수 시작값을 생성합니다.

표본 수
Lilliefors 검정에 사용되는 Monte Carlo 표본추출 복제 수를 재설정합니다. 값은 100 - 최대 정수 범위에 있는 단일 정수여야 합니다. 기본 값은 10,000입니다.

시뮬레이션 신뢰수준(%)
Kolmogorov-Smirnov 검정의 추정된 신뢰구간 수준을 재설정합니다. 값은 0 - 100 범위의 단일값이어야 합니다. 기본값은 99입니다.

사용자 결측값

범주형 필드의 사용자 결측값. 범주형 필드를 분석에 포함하려면 레코드에 대해 유효한 값이 있어야 합니다. 이러한 제어를 사용하면 범주형 필드 사이에서 사용자 결측값을 유효한 데이터로 취급할지 여부를 결정할 수 있습니다. 시스템 결측값 및 연속형 필드에 대한 결측값은 항상 유효한 것으로 처리됩니다.

NPTESTS 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 프로시저 런에 일표본, 독립 표본 및 대응표본 검정을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

일표본 Wilcoxon 검정

Wilcoxon 부호순위 검정은 표본 데이터를 기반으로 모집단의 위치를 검정하거나 두 모집단의 위치를 비교하는 데 사용되는 비모수 통계 검정입니다. 일표본 Wilcoxon 부호순위 검정은 데이터가 정규 분포를 따르지 않을 때 표본 t-검정에 대한 비모수적 대안입니다. 표본의 중앙값과 가설 중앙값 사이에 유의적인 차이가 있는지 여부를 판별하는 데 도움이 됩니다. 여기서 검정 통계량은 관측값과 가설 값 사이의 차이 순위를 기준으로 계산됩니다.

한 표본 Wilcoxon 부호순위 검정은 모집단이 중위수에 대해 대칭이라는 가정을 작성합니다. 무작위 표본은 연속형 누적 분포 함수를 사용하는 모집단에서 비롯되며 귀무가설 하에서 중위수에 대해 대칭이므로 관측값과 가설 값 사이의 차이는 0에 대해 대칭적으로 분포되므로 동일한 절대 크기의 양의 차이와 음의 차이는 동일한 발생 확률을 갖습니다.

예제

새 스타트업 회사가 소프트웨어 엔지니어를 고용하는 데 관심이 있지만 경쟁력 있는 급여를 제공하고 있는지 확인하려고 한다고 가정하십시오. 현재 동일한 산업의 다양한 회사에 고용되어 있는 30개의 소프트웨어 엔지니어 샘플에서 급여 데이터를 수집했다고 가정합니다. 회사에서 잠재적인 직원에게 제공하는 중간 급여가 업계에서 소프트웨어 엔지니어의 중간 급여와 상당히 다른지 여부를 판별하려는 경우 이러한 용도로 하나의 샘플 Wilcoxon 기호 순위를 사용할 수 있습니다. 여기서는 시작 시 제공되는 중간 급여가 업계의 중간 급여와 크게 다르지 않다는 귀무가설 즉, 중간 급여 차이가 0이라고 가정합니다.

일표본 Wilcoxon 검정을 실행하는 방법

(이 기능을 사용하려면 Statistics Base 옵션이 필요합니다.)

메뉴에서 다음을 선택하십시오.

1. **분석 > 비모수 검정 > 일표본 ...**
2. **'설정'**을 클릭하십시오.
3. **'항목 선택'**에서 **'테스트 선택'**을 선택하십시오.
4. **'사용자 정의 테스트'**를 선택하십시오.
5. **'중앙값과 가설 값 비교 (Wilcoxon 부호순위 검정)'**를 선택합니다.
6. **'Hypothesized median'** 의 값을 입력하고 **'Run'**을 누르십시오.

자세한 정보는 https://www.ibm.com/support/pages/can-spss-statistics-produce-one-sample-wilcoxon-test 의 내용을 참조하십시오.

독립 표본 비모수 검정

독립 표본 비모수 검정에서는 하나 이상의 비모수 검정을 사용하여 둘 이상의 그룹 간의 차이점을 식별합니다. 비모수 검정에서는 데이터가 정규 분포를 따른다고 가정하지 않습니다.

사용자의 목표. 목표를 사용하면 다르지만 일반적으로 사용되는 검정 설정을 신속하게 지정할 수 있습니다.

- **그룹 전체에서 분포 자동 비교.** 이 목표는 Mann-Whitney U 검정을 2개 그룹이 있는 데이터에 적용하거나, 크루스칼-왈리스 1원 ANOVA를 *k* 그룹이 있는 데이터에 적용합니다.
- **그룹 전체에서 중위수 비교.** 이 목표는 중위수 검정을 사용하여 그룹 전체에서 관측된 중위수를 비교합니다.
- **사용자 정의 분석.** 설정 탭에서 검정 설정을 수동으로 변경하려면 이 옵션을 선택하십시오. 이후에 설정 탭에서 옵션을 변경한 후 이 변경 사항을 현재 선택된 목표와 함께 사용할 수 없는 경우 이 설정이 자동으로 선택됩니다.

독립 표본 비모수 검정을 수행하려면

메뉴에서 다음을 선택합니다.

분석 > 비모수 검정 > 독립 표본...

1. **실행**을 클릭합니다.

선택적으로 다음을 수행할 수 있습니다.

- 목적 탭에서 목적을 지정합니다.
- 필드 탭에서 필드 할당을 지정합니다.
- 설정 탭에서 전문가 설정을 지정합니다.

필드 탭

필드 탭은 검정할 필드 및 그룹을 정의하는 데 사용되는 필드를 지정합니다.

사전 정의된 역할 사용. 이 옵션을 기존 필드 정보를 사용합니다. 대상 또는 둘 다로 역할이 사전 정의된 모든 연속형 및 순서 필드를 검정 필드로 사용합니다. 역할이 입력으로 사전 정의된 단일 범주형 필드는 그룹 필드로 사용됩니다. 그렇지 않으면 기본적으로 그룹 필드가 사용되지 않으며 사용자 정의 필드 할당을 사용해야 합니다. 하나 이상의 검정 필드와 그룹 필드가 필요합니다.

사용자 정의 필드 할당 사용. 이 옵션을 사용하여 필드 역할을 바꿀 수 있습니다. 이 옵션을 선택한 후 아래 필드를 지정합니다.

- **검정 필드.** 하나 이상의 연속형 또는 순서 필드를 선택하십시오.
- **그룹.** 범주형 필드를 선택합니다.

설정 탭

설정 탭은 몇 가지 다른 설정 그룹으로 구성되어 있으며 설정을 수정하여 알고리즘이 데이터를 처리하는 방법을 미세 튜닝할 수 있습니다. 기본값 설정을 현재 선택된 목표와 함께 사용할 수 없는 설정으로 변경하면, 목표 탭이 자동으로 업데이트되어 **분석 사용자 정의** 옵션이 선택됩니다.

검정 선택

이 설정은 필드 탭에 지정된 필드에서 수행되는 검정을 지정합니다.

데이터를 기준으로 검정 자동 선택. 이 설정은 Mann-Whitney U 검정을 2개 그룹이 있는 데이터에 적용하거나, 크루스칼-왈리스 1원 ANOVA를 *k* 그룹이 있는 데이터에 적용합니다.

검정 사용자 정의. 이 설정을 사용하면 수행되는 특정 검정을 선택할 수 있습니다.

- **그룹 전체에서 분포 비교.** 이 옵션은 표본이 동일한 모집단에서 비롯되는지 확인하는 독립 표본 검정을 생성합니다.

  Mann-Whitney U(2개 표본)에서는 각 케이스의 순위를 사용하여 그룹이 동일한 모집단에서 비롯되는지 여부를 검정합니다. 그룹 필드의 오름차순에 있는 첫 번째 값은 첫 번째 그룹을 정의하고 두 번째 값은 두 번째 그룹을 정의합니다. 그룹 필드에 값이 3개 이상 있으면 이 검정이 생성되지 않습니다.

  콜모고로프-스미르노프(2개 표본)는 중위수, 산포, 왜도 등과 같은 두 개의 분포 간의 모든 차이의 영향을 받습니다. 그룹 필드에 값이 3개 이상 있으면 이 검정이 생성되지 않습니다.

  무작위 시퀀스로 검정(2개 표본용 Wald-Wolfowitz)은 그룹소속을 기준으로 사용하여 런 검정을 생성합니다. 그룹 필드에 값이 3개 이상 있으면 이 검정이 생성되지 않습니다.

  크루스칼-왈리스 1원 ANOVA(k 표본)는 Mann-Whitney U 검정의 확장이며 일원 분산 분석의 비모수 동일 분석입니다. 선택적으로 *k* 샘플의 다중 비교(**모두 쌍별** 다중 비교 또는 **단계적 하향식 비교**)를 요청할 수 있습니다.

  순서 지정된 선택 사항 검정(k 표본용 Jonckheere-Terpstra)은 *k* 표본에 자연 순위가 있는 경우 크루스칼-왈리스보다 강력한 방법입니다. 예를 들어, *k* 모집단은 증가 온도 *k*를 나타낼 수도 있습니다. 서로 다른 온도가 동일한 반응 분포를 생성한다는 가설이 온도가 증가할수록 반응의 크기도 증가한다는 가설에 반대되지 않는지 검정합니다. 여기에는 대체 가설에 순위가 있으므로 Jonckheere-Terpstra가 사용하기에 가장 적절한 검정입니다. **가장 작은 값에서 가장 큰 값으로**는 첫 번째 그룹의 위치 모수가 두 번째보다 작거나 같고 두 번째는 세 번째보다 작거나 같은 방식의 대체 가설을 지정합니다. **가장 큰 값에서 가장 작은 값으로**는 첫 번째 그룹의 위치 모수가 두 번째보다 크거나 같고 두 번째는 세 번째보다 크거나 같은 방식의 대체 가설을 지정합니다. 또한 두 옵션 모두에서 대체 가설은 위치가 모두 같지 않다고 가정합니다. 선택적으로 *k* 표본의 다중 비교를 요청할 수 있습니다. **모든 대응별** 다중 비교 또는 **단계적 하향식** 비교를 선택할 수 있습니다.

- **그룹 전체에서 범위 비교.** 이 옵션은 표본의 범위가 동일한지 확인하는 독립 표본 검정을 생성합니다. **모제스의 극단값반동(2개 표본)**은 대조군대 비교군을 검정합니다. 그룹 필드의 오름차순에 있는 첫 번째 값은 대조군을 정의하고 두 번째 값은 비교군을 정의합니다. 그룹 필드에 값이 3개 이상 있으면 이 검정이 생성되지 않습니다.
- **그룹 전체에서 중위수 비교.** 이 옵션은 표본의 중위수가 동일한지 확인하는 독립 표본 검정을 생성합니다. **중위수 검정(k 표본)**에서는 풀링된 표본 중위수(데이터 세트의 모든 레코드에서 계산됨) 또는 사용자 정의 값을 가설 중위수로 사용할 수 있습니다. 선택적으로 *k* 표본의 다중 비교를 요청할 수 있습니다. **모든 대응별** 다중 비교 또는 **단계적 하향식** 비교를 선택할 수 있습니다.
- **그룹 전체에서 신뢰구간 추정. 핫지스-레만 추정(2개 표본)**은 두 그룹의 중위수 차이에 대해 독립 표본 추정과 신뢰구간을 생성합니다. 그룹 필드에 값이 3개 이상 있으면 이 검정이 생성되지 않습니다.

검정 옵션

유의수준. 이 옵션은 모든 검정의 유의수준(알파)을 지정합니다. 0과 1 사이의 숫자 값을 지정합니다. 기본값은 0.05입니다.

신뢰구간(%). 생성된 모든 신뢰구간의 신뢰수준을 지정합니다. 0-100 사이의 숫자 값을 지정하십시오. 기본값은 95입니다.

제외된 케이스. 이 옵션은 검정의 케이스 기준을 결정하는 방법을 지정합니다. **목록별 결측값 제외**는 하위 명령에 이름 지정된 모든 필드에 대해 결측값이 있는 레코드가 모든 분석에서 제외됨을 의미합니다. **검정별 결측값 제외**는 특정 검정에 사용되는 필드에 대해 결측값이 있는 레코드를 해당 검정에서 생략한다는 의미입니다. 분석에 여러 개의 검정이 지정되어 있는 경우 각 검정이 별도로 평가됩니다.

사용자 결측값

범주형 필드의 사용자 결측값. 범주형 필드를 분석에 포함하려면 레코드에 대해 유효한 값이 있어야 합니다. 이러한 제어를 사용하면 범주형 필드 사이에서 사용자 결측값을 유효한 데이터로 취급할지 여부를 결정할 수 있습니다. 시스템 결측값 및 연속형 필드에 대한 결측값은 항상 유효한 것으로 처리됩니다.

NPTESTS 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 프로시저 런에 일표본, 독립 표본 및 대응표본 검정을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

대응표본 비모수 검정

하나 이상의 비모수 검정을 사용하여 2개 이상의 관련 필드 간의 차이를 식별합니다. 비모수 검정에서는 데이터가 정규 분포를 따른다고 가정하지 않습니다.

데이터 고려 사항. 각 레코드는 지정된 개체에 해당하며 둘 이상의 대응 측정이 데이터 세트에 있는 별도의 필드에 저장됩니다. 예를 들어, 다이어트 계획의 효과에 대한 연구의 경우 각 개체의 가중치가 일정한 구간으로 측정되어 다이어트 전 체중, 다이어트 중 체중, 다이어트 후 체중등과 같은 필드에 저장되면 대응표본 비모수 검정을 사용하여 이 연구를 분석할 수 있습니다. 이러한 필드를 "대응"이라고 합니다.

사용자의 목표. 목표를 사용하면 다르지만 일반적으로 사용되는 검정 설정을 신속하게 지정할 수 있습니다.

- **관측 데이터와 가설 데이터 자동으로 비교.** 이 목표는 필드가 2개 지정되면 범주형 데이터에 맥니머의 검정을 적용하고, 필드가 3개 이상 지정되면 범주형 데이터에 Cochran의 Q를 적용합니다. 또한 필드가 2개 지정되면 연속형 데이터에 윌콕슨 대응표집 쌍 부호순위 검정을 적용하고, 필드가 3개 이상 지정되면 연속형 데이터에 프리드만의 순위별 2원 ANOVA를 적용합니다.
- **사용자 정의 분석.** 설정 탭에서 검정 설정을 수동으로 변경하려면 이 옵션을 선택하십시오. 이후에 설정 탭에서 옵션을 변경한 후 이 변경 사항을 현재 선택된 목표와 함께 사용할 수 없는 경우 이 설정이 자동으로 선택됩니다.

측정 수준이 다를 필드가 지정된 경우, 먼저 측정 수준별로 분리된 다음 각 그룹에 적절한 검정이 적용됩니다. 예를 들어, **관측 데이터와 가설 데이터 자동으로 비교**를 목표로 선택하고 3개의 연속형 필드와 2개의 명목 필드를 지정하는 경우 프리드만 검정이 연속형 필드에 적용되고 맥니머 검정이 명목 필드에 적용됩니다.

대응표본 비모수 검정을 수행하려면

메뉴에서 다음을 선택합니다.

분석 > 비모수 검정 > 대응 표본...

1. **실행**을 클릭합니다.

선택적으로 다음을 수행할 수 있습니다.

- 목적 탭에서 목적을 지정합니다.
- 필드 탭에서 필드 할당을 지정합니다.
- 설정 탭에서 전문가 설정을 지정합니다.

필드 탭

이 필드 탭에서는 검정할 필드를 지정합니다.

사전 정의된 역할 사용. 이 옵션을 기존 필드 정보를 사용합니다. 대상 또는 모두로 역할이 사전 정의된 모든 필드가 검정 필드로 사용됩니다. 2개 이상의 검정 필드가 필요합니다.

사용자 정의 필드 할당 사용. 이 옵션을 사용하여 필드 역할을 바꿀 수 있습니다. 이 옵션을 선택한 후 아래 필드를 지정합니다.

- **검정 필드.** 필드를 두 개 이상 선택하십시오. 각 필드는 별도의 대응표본에 해당합니다.

설정 탭

설정 탭은 몇 가지 다른 설정 그룹으로 구성되어 있으며 설정을 수정하여 프로시저가 데이터를 처리하는 방법을 미세 튜닝할 수 있습니다. 기본값 설정을 다른 목적과 함께 사용할 수 없는 사항으로 변경하면, 목적 탭이 자동으로 업데이트되어 **분석 사용자 정의** 옵션이 선택됩니다.

검정 선택

이 설정은 필드 탭에 지정된 필드에서 수행되는 검정을 지정합니다.

데이터를 기준으로 검정 자동 선택. 이 설정은 필드가 2개 지정되면 범주형 데이터에 맥니머의 검정을 적용하고, 필드가 3개 이상 지정되면 범주형 데이터에 Cochran의 Q를 적용합니다. 또한 필드가 2개 지정되면 연속형 데이터에 윌콕슨 대응표집 쌍 부호순위 검정을 적용하고, 필드가 3개 이상 지정되면 연속형 데이터에 프리드만의 순위별 2원 ANOVA를 적용합니다.

검정 사용자 정의. 이 설정을 사용하면 수행되는 특정 검정을 선택할 수 있습니다.

- **2진 데이터의 변경에 대한 검정. McNemar의 검정(2개 표본)**은 범주형 필드에 적용할 수 있습니다. 이 옵션은 2개의 플래그 필드(값이 2개만 있는 범주형 필드) 간의 값 조합이 동일할 가능성이 있는지 여부를 확인하는 대응표본 검정을 생성합니다. 필드 탭에 3개 이상의 필드가 지정되어 있으면 이 검정이 수행되지 않습니다. 검정 설정에 대한 자세한 내용은 203 페이지의 『McNemar의 검정: 성공 정의』의 내용을 참조하십시오. **Cochran의 Q(k 표본)**을 범주형 필드에 적용할 수 있습니다. 이 옵션은 *k* 플래그 필드(값이 2개만 있는 범주형 필드) 간의 값 조합이 동일할 가능성이 있는지 여부를 확인하는 대응표본 검정을 생성합니다. 선택적으로 *k* 표본의 다중 비교를 요청할 수 있습니다. **모든 대응별** 다중 비교 또는 **단계적 하향식** 비교를 선택할 수 있습니다. 검정 설정에 대한 자세한 내용은 203 페이지의 『Cochran의 Q: 성공 정의』의 내용을 참조하십시오.
- **다항 데이터의 변경에 대한 검정. 주변 동질성 검정(2개 표본)**은 쌍을 이룬 두 순서형 필드 사이의 값 조합이 동일한지 여부에 대한 관련 표본 검정을 생성합니다. 주변 동질성 검정은 일반적으로 반복 측도 상황에 사용됩니다. 이 검정은 McNemar 검정을 이분형 반응에서 다항 반응으로 확장한 검정 방법입니다. 필드 탭에 3개 이상의 필드가 지정되어 있으면 이 검정이 수행되지 않습니다.
- **중위수 차이와 가설 값 비교.** 이러한 검정에서는 두 개의 필드 사이의 중위수 차이가 0과 다른지 여부에 대한 관련 표본 검정을 각각 생성합니다. 검정은 연속형 및 순서 필드에 적용됩니다. 필드 탭에 세 개 이상의 필드가 지정되어 있으면 이 검정이 수행되지 않습니다.
- **신뢰구간 추정.** 이 옵션은 쌍을 이루는 두 개의 필드 사이이 중위수 차이에 대한 대응표본 추정 및 신뢰구간을 생성합니다. 검정은 연속형 및 순서 필드에 적용됩니다. 필드 탭에 세 개 이상의 필드가 지정된 경우 이 검정이 수행되지 않습니다.

- **연관 수량화. 켄달의 일치계수(W)(k 표본)**는 감정 평가나 등급 평가자 간의 동의 측도를 생성하며 여기서 각 레코드는 여러 항목(필드)에 대한 한 명의 감정 평가 등급입니다. 선택적으로 *k* 표본의 다중 비교를 요청할 수 있습니다. **모든 대응별** 다중 비교 또는 **단계적 하향식** 비교를 선택할 수 있습니다.
- **분포 비교. Friedman의 순위별 이원 ANOVA(k 표본)**는 *k* 관련 표본이 동일한 모집단에서 추출되었는지 여부에 대한 관련 표본 검정을 생성합니다. 선택적으로 *k* 표본의 다중 비교를 요청할 수 있습니다. **모든 대응별** 다중 비교 또는 **단계적 하향식** 비교를 선택할 수 있습니다.

*McNemar*의 검정*:* 성공 정의

맥니머의 검정은 플래그 필드(범주가 2개만 있는 범주형 필드)에 적용되지만 "성공" 정의 규칙을 사용하면 모든 필드에 적용할 수 있습니다.

범주형 필드의 성공 정의. 이 옵션은 범주형 필드에 대해 "성공"이 정의되는 방법을 지정합니다.

- **데이터의 첫 번째 범주 사용**은 표본에 있는 첫 번째 값으로 검정을 수행하여 "성공"을 정의합니다. 이 옵션은 값이 2개만 있는 명목 또는 순서 필드에만 적용됩니다. 이 옵션이 사용되는 필드 탭에 지정된 기타 모든 범주형 필드는 검정되지 않습니다. 이는 기본값입니다.
- **성공값 지정**은 지정된 값 목록으로 검정을 수행하여 "성공"을 정의합니다. 문자 또는 숫자 값 목록을 지정합니다. 목록에 있는 값이 표본에 없어도 됩니다.

*Cochran*의 *Q:* 성공 정의

Cochran의 Q 검정은 플래그 필드(범주가 2개만 있는 범주형 필드)에 적용되지만 "성공" 정의 규칙을 사용하면 모든 필드에 적용할 수 있습니다.

범주형 필드의 성공 정의. 이 옵션은 범주형 필드에 대해 "성공"이 정의되는 방법을 지정합니다.

- **데이터의 첫 번째 범주 사용**은 표본에 있는 첫 번째 값으로 검정을 수행하여 "성공"을 정의합니다. 이 옵션은 값이 2개만 있는 명목 또는 순서 필드에만 적용됩니다. 이 옵션이 사용되는 필드 탭에 지정된 기타 모든 범주형 필드는 검정되지 않습니다. 이는 기본값입니다.
- **성공값 지정**은 지정된 값 목록으로 검정을 수행하여 "성공"을 정의합니다. 문자 또는 숫자 값 목록을 지정합니다. 목록에 있는 값이 표본에 없어도 됩니다.

검정 옵션

유의수준. 이 옵션은 모든 검정의 유의수준(알파)을 지정합니다. 0과 1 사이의 숫자 값을 지정합니다. 기본값은 0.05입니다.

신뢰구간(%). 생성된 모든 신뢰구간의 신뢰수준을 지정합니다. 0-100 사이의 숫자 값을 지정하십시오. 기본값은 95입니다.

제외된 케이스. 이 옵션은 검정의 케이스 기준을 결정하는 방법을 지정합니다.

- **목록별 결측값 제외**는 하위 명령에 이름 지정된 모든 필드에 대해 결측값이 있는 레코드가 모든 분석에서 제외됨을 의미합니다.
- **검정별 결측값 제외**는 특정 검정에 사용되는 필드에 대해 결측값이 있는 레코드를 해당 검정에서 생략한다는 의미입니다. 분석에 여러 개의 검정이 지정되어 있는 경우 각 검정이 별도로 평가됩니다.

사용자 결측값

범주형 필드의 사용자 결측값. 범주형 필드를 분석에 포함하려면 레코드에 대해 유효한 값이 있어야 합니다. 이러한 제어를 사용하면 범주형 필드 사이에서 사용자 결측값을 유효한 데이터로 취급할지 여부를 결정할 수 있습니다. 시스템 결측값 및 연속형 필드에 대한 결측값은 항상 유효한 것으로 처리됩니다.

NPTESTS 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 프로시저 런에 일표본, 독립 표본 및 대응표본 검정을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

NPTESTS 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 프로시저 런에 일표본, 독립 표본 및 대응표본 검정을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

레거시 대화 상자

다양한 "레거시" 대화 상자에서 비모수 검정도 수행할 수 있습니다. 이 대화 상자는 정확검정 옵션에서 제공하는 기능을 지원합니다.

카이제곱 검정. 변수를 범주로 묶어 표로 작성하고 관측빈도와 기대빈도 간 차이를 기초로 카이제곱 통계를 계산합니다.

이항검정. 이분형 변수의 각 범주에서 관측빈도를 이항 분산에서의 기대빈도와 비교합니다.

런 검정. 한 변수의 두 값이 발생하는 순서가 무작위인지 검정합니다.

일표본 Kolmogorov-Smirnov 검정. 한 변수에 대해 관측된 누적 분포 함수를 지정된 이론적 분포와 비교합니다. 이때 이론적 분포는 정규 분포, 균일분포, 지수분포, 포아송분포가 됩니다.

독립 2-표본 비모수 검정. 하나의 변수에 대한 두 개의 케이스 그룹을 비교합니다. Mann-Whitney의 *U* 검정, 2-표본 Kolmogorov-Smirnov 검정, Moses의 극단값반동 검정, Wald-Wolfowitz run 검정을 사용할 수 있습니다.

대응 2-표본 비모수 검정. 두 변수의 분포를 비교합니다. Wilcoxon 부호순위 검정, 부호 검정, McNemar 검정을 사용할 수 있습니다.

독립 K-표본 비모수 검정. 하나의 변수에 대한 둘 이상의 케이스 그룹을 비교합니다. Kruskal-Wallis 검정, 중위수 검정, Jonckheere-Terpstra 검정을 사용할 수 있습니다.

대응 K-표본 비모수 검정. 둘 이상의 변수 분포를 비교합니다. Friedman 검정, Kendall의 *W*, Cochran의 *Q* 검정을 사용할 수 있습니다.

위의 모든 검정에서 사분위수, 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수 등을 사용할 수 있습니다.

카이제곱 검정

카이제곱 검정 프로시저를 사용하여 변수를 범주로 묶어 표로 작성하고 카이제곱 통계를 계산합니다. 이러한 적합도 검정은 각 범주에서 관측빈도와 기대빈도를 비교하여 모든 범주에 값이 같은 비율로 있는지 또는 각 범주마다 값이 사용자가 지정한 비율로 들어 있는지 확인합니다.

예. 카이제곱 검정을 사용하여 젤리빈 한 상자에 파랑색, 갈색, 녹색, 오렌지색, 빨강색, 노랑색 사탕을 같은 비율로 넣을 수 있습니다. 또한 젤리빈 한 상자에 파랑색 캔디가 5%, 갈색이 30%, 녹색이 10%, 오렌지색이 20%, 빨강색이 15%, 노랑색 캔디가 15% 있는지 검정할 수도 있습니다.

통계. 평균, 표준 편차, 최소값, 최대값, 사분위수를 선택할 수 있습니다. 누락되지 않은 케이스 수와 결측 케이스 수, 각 범주에 대한 관측 케이스 수와 기대 케이스 수, 카이제곱 통계를 구합니다.

카이제곱 검정 데이터 고려 사항

데이터. 정렬되었거나 정렬되지 않은 숫자 범주형 변수(측정의 순서 수준이나 명목 수준)를 사용합니다. 문자열 변수를 숫자변수로 변환하려면 변환 메뉴의 자동 코딩변경 프로시저를 사용합니다.

가정. 비모수 검정은 기본 분포의 형태에 대한 가정이 필요하지 않습니다. 데이터는 무작위 표본이라고 가정합니다. 각 범주의 예상 빈도는 1 이상이어야 합니다. 범주의 20% 이하가 5미만의 예상 빈도를 가지고 있어야 합니다.

카이제곱 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 비모수 검정 > 레거시 대화 상자 > 카이제곱...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

2. 하나 이상의 검정변수를 선택합니다. 각 변수로 서로 다른 검정을 만들 수 있습니다.
3. 선택적으로 기술통계, 사분위수, 결측데이터 처리 방법을 보려면 **옵션**을 클릭합니다.

카이제곱 검정 기대범위와 기대값

기대범위. 기본적으로 변수의 고유 값은 하나의 범주로 정의됩니다. 특정 범위 내에 범주들을 설정하려면 **지정한 범위 사용**을 선택하고 하한과 상한에 해당하는 정수 값을 입력합니다. 범주는 포함 범위 내 정수 값마다 지정되고 경계 이외의 값을 가지는 케이스는 제외됩니다. 예를 들어, 하한값에 1을 지정하고 상한값에 4를 지정하면 카이제곱 검정에 1부터 4까지의 정수값만 사용됩니다.

기대값. 기본적으로 모든 범주의 기대값은 동일합니다. 범주에는 사용자가 지정한 기대비율을 포함할 수 있습니다. **값**을 선택하고 검정변수의 각 범주에 0보다 큰 값을 입력한 다음 **추가**를 클릭합니다. 값을 추가할 때마다 값 목록의 맨 아래에 나타납니다. 이러한 값의 순서는 검정변수에서 범주값의 오름차순에 해당하므로 매우 중요합니다. 목록에서 첫 번째 값은 검정변수의 가장 낮은 그룹 값에 해당하고 목록의 마지막 값은 가장 높은 값에 해당합니다. 값 목록의 요소를 모두 합하여 각 값을 이 합계로 나누어 해당 범주에서 기대되는 케이스 비율을 계산합니다. 예를 들어, 3, 4, 5, 4라는 값 목록은 기대비율을 3/16, 4/16, 5/16, 4/16로 지정합니다.

카이제곱 검정*:* 옵션

통계. 요약 통계 중 하나 또는 모두를 선택할 수 있습니다.

- **기술통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수를 출력합니다.
- **사분위수.** 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

결측값. 결측값의 처리를 제어합니다.

- **검정별 결측값 제외.** 여러 검정을 지정할 때 각 검정은 결측값에 대해 별도로 평가됩니다.
- **목록별 결측값 제외.** 변수에 대한 결측값을 포함하는 케이스는 모든 분석에서 제외됩니다.

NPAR TESTS 명령 추가 기능*(*카이제곱 검정*)*

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- CHISQUARE 하위 명령을 사용하여 서로 다른 변수에 대해 서로 다른 최소값 및 최대값 또는 기대빈도를 지정합니다.
- EXPECTED 하위 명령을 사용하여 서로 다른 기대빈도에 대해 동일한 변수를 검정하거나 다른 범위를 사용합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

이항검정

이항검정 프로시저는 한 이분형 변수의 두 범주에 대한 관측 빈도가 지정된 확률 모수를 가지는 이항 분산의 기대빈도와 비교할 수 있습니다. 기본적으로 두 그룹 모두 확률 모수는 0.5입니다. 확률을 바꾸려면 첫 번째 그룹에 대한 검정 비율을 입력합니다. 두 번째 그룹에 대한 검정 비율은 1에서 첫 번째 그룹에 지정된 확률을 뺀 값이 됩니다.

예. 동전을 던질 때 앞면이 나올 확률은 1/2입니다. 이러한 가설에 기초하여, 다임을 40번 던지고, 결과를 기록합니다(앞 또는 뒤). 이항검정을 사용한 결과 던진 횟수의 3/4이 앞면이고 관측된 유의수준이 적다(0.0027)는 사실을 알게 되었습니다. 이러한 결과는 앞면이 나올 확률이 1/2이 되기 어렵다는 것을 의미합니다.

통계. 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수, 사분위수를 선택할 수 있습니다.

이항검정 데이터 고려 사항

데이터. 검정할 변수는 숫자이거나 이분형이어야 합니다. 문자열 변수를 숫자변수로 변환하려면 변환 메뉴의 자동 코딩변경 프로시저를 사용합니다. **이분형 변수**는 *yes* 또는 *no*, *true* 또는 *false*, 0 또는 1 등과 같이 두 개의 값만 취할 수 있는 변수입니다. 데이터 세트에 있는 첫 번째 값은 첫 번째 그룹을 정의하고 다른 값은 두 번째 그룹을 정의합니다. 변수가 이분형이 아닌 경우에는 절단점을 지정해야 합니다. 절단점은 절단점보다 작거나 같은 값을 가진 케이스를 첫 번째 그룹에 할당하고 나머지 케이스를 두 번째 그룹에 할당합니다.

가정. 비모수 검정은 기본 분포의 형태에 대한 가정이 필요하지 않습니다. 데이터는 무작위 표본이라고 가정합니다.

이항검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 비모수 검정 > 레거시 대화 상자 > 이항...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자 검정변수를 선택합니다.
3. 선택적으로 기술통계, 사분위수, 결측데이터 처리 방법을 보려면 **옵션**을 클릭합니다.

이항검정*:* 옵션

통계. 요약 통계 중 하나 또는 모두를 선택할 수 있습니다.

- **기술통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수를 출력합니다.
- **사분위수.** 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

결측값. 결측값의 처리를 제어합니다.

- **검정별 결측값 제외.** 여러 검정을 지정할 때 각 검정은 결측값에 대해 별도로 평가됩니다.
- **목록별 결측값 제외.** 검정된 변수에 대한 결측값이 있는 케이스는 모든 분석에서 제외됩니다.

NPAR TESTS 명령 추가 기능*(*이항검정*)*

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 변수에 세 개 이상의 범주가 포함된 경우 BINOMIAL 하위 명령을 사용하여 관련 그룹을 선택하고 다른 그룹은 제외합니다.
- BINOMIAL 하위 명령을 사용하여 서로 다른 변수에는 다른 절단점 또는 확률을 지정합니다.
- EXPECTED 하위 명령을 사용하여 서로 다른 절단점 또는 확률에 대해 동일한 변수를 검정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

런 검정

런 검정 프로시저는 두 변수값이 나타나는 순서가 무작위인지 여부를 검정합니다. 런은 일련의 유사한 관측값을 나타냅니다. 런이 너무 많거나 너무 적은 표본은 무작위 상태가 아닐 수 있습니다.

예. 20명이 한 제품을 구매할지 여부를 투표로 알아낸다고 가정합니다. 무작위로 추출한 20명의 표본이 동일한 성별로 구성되었는지 여부를 심각하게 고려해 봐야 합니다. 런 검정은 해당 표본을 무작위로 추출하였는지 판별하는 데 사용됩니다.

통계. 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수, 사분위수를 선택할 수 있습니다.

런 검정 데이터 고려 사항

데이터. 변수는 숫자이어야 합니다. 문자열 변수를 숫자변수로 변환하려면 변환 메뉴의 자동 코딩변경 프로시저를 사용합니다.

가정. 비모수 검정은 기본 분포의 형태에 대한 가정이 필요하지 않습니다. 연속확률 분포의 표본을 사용합니다.

런 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 비모수 검정 > 레거시 대화 상자 > 실행...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자 검정변수를 선택합니다.
3. 선택적으로 기술통계, 사분위수, 결측데이터 처리 방법을 보려면 **옵션**을 클릭합니다.

런 검정 절단점

절단점. 선택한 변수를 이분화하는 절단점을 지정합니다. 관측평균, 중위수 또는 최빈값을 사용하거나 지정된 값을 절단점으로 사용할 수 있습니다. 값이 절단점보다 작으면 케이스는 한 그룹으로 지정되고, 절단점보다 크거나 같으면 다른 그룹으로 지정됩니다. 선택한 각 절단점에 대해 한 개의 검정이 수행됩니다.

런 검정 옵션

통계. 요약 통계 중 하나 또는 모두를 선택할 수 있습니다.

- **기술통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수를 출력합니다.
- **사분위수.** 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

결측값. 결측값의 처리를 제어합니다.

- **검정별 결측값 제외.** 여러 검정을 지정할 때 각 검정은 결측값에 대해 별도로 평가됩니다.
- **목록별 결측값 제외.** 변수에 대한 결측값을 포함하는 케이스는 모든 분석에서 제외됩니다.

NPAR TESTS 명령 추가 기능(런 검정)

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- RUNS 하위 명령 사용하여 서로 다른 변수에는 다른 절단점을 지정하십시오.
- RUNS 하위 명령을 사용하여 서로 다른 사용자 정의 절단점에 대해 동일한 변수를 검정하십시오.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

일표본 Kolmogorov-Smirnov 검정

일표본 Kolmogorov-Smirnov 검정 프로시저를 통해 지정한 이론 분포(정규, 균일, 포아송, 지수모형)로 변수에 대한 관측 누적 분포 함수를 비교할 수 있습니다. Kolmogorov-Smirnov의 *Z*는 관측 누적 분포 함수와 이론 누적 분포 함수 간의 최대 차이(절대값)에서 계산됩니다. 이러한 적합도 검정으로 지정 분포로부터 관측값이 제대로 작성되었는지 확인할 수 있습니다.

버전 27.0부터 Lilliefors 검정 통계량을 사용하여 *p*-값을 추정할 수 있는데, 이를 위해서는 Monte Carlo 표본추출을 사용하여 추정 모수로 정규 분포에 대해 검정하십시오. 이 기능은 이전에는 데이터 탐색 프로시저를 통해서만 가능했습니다.

예
여러 모수 검정에는 정규적으로 분포된 변수가 있어야 합니다. 일표본 Kolmogorov-Smirnov 검정은 변수(예: 가구소득)가 정규적으로 분포되었음을 검정할 때 사용됩니다.

통계
평균, 표준편차, 최소값, 최대값, 비결측 케이스 수, 사분위수, Lilliefors 검정, Monte Carlo 시뮬레이션

일표본 Kolmogorov-Smirnov 검정 데이터 고려사항

데이터
양적변수(측도 구간 또는 비율 수준)를 사용합니다.

가정
Kolmogorov-Smirnov 검정에서는 검정 분포의 모수가 미리 지정되어 있는 것으로 가정합니다. 이 프로시저로 표본에서 모수를 추정합니다. 표본 평균과 표본 표준 편차는 정규 분포에 대한 모수가 되고 표본 최소값과 최대값은 균일 분포의 범위를 정의하며 표본 평균은 포아송 분포의 모수가 되며 표본 평균은 지수 분포에 대한 모수가 됩니다. 가설 분포로부터의 이탈을 발견하기 위한 검정력은 크게 감소할 수 있습니다.

분포의 특정 모수를 표본에서 추정해야 할 경우 Kolmogorov-Smirnov 검정이 더 이상 적용되지 않습니다. 이 경우, Lilliefors 검정 통계량을 사용하여 *p*-값을 추정할 수 있는데, 이를 위해서는 Monte Carlo 표본추출을 사용하여 알 수 없는 평균과 분산을 갖는 정규성을 검정하십시오. Lilliefors 검정은 세 개의 연속 분포(**정규**, **지수**, **균일**)에 적용됩니다. 기본 검정이 이산형(**포아송**)일 경우에는 이 검정이 적용되지 않습니다. 해당하는 분포모수가 지정되지 않은 경우에는 일표본 추론에 대해서만 검정이 정의됩니다.

일표본 Kolmogorov-Smirnov 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **비모수 검정** > **레거시 대화 상자** > **1-표본 K-S...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자 검정변수를 선택합니다. 각 변수로 서로 다른 검정을 만들 수 있습니다.
3. 선택적으로 검정분포 방법을 선택합니다.

   정규
   선택할 경우 분포모수를 표본 데이터(기본 설정) 또는 사용자 정의 설정에서 추정할지 여부를 지정할 수 있습니다. **표본 데이터 사용**을 선택한 경우, 기존 근사 결과와 Monte Carlo 표본추출을 기반으로 하는 Lilliefors 유의확률 수정이 둘 다 사용됩니다. **사용자 정의**를 선택한 경우, **평균** 및 **표준편차**에 대한 값을 제공하십시오.

   균일
   선택할 경우 분포모수를 표본 데이터(기본 설정) 또는 사용자 정의 설정에서 추정할지 여부를 지정할 수 있습니다. **표본 데이터 사용**을 선택한 경우 Lilliefors 검정이 사용됩니다. **사용자 정의**를 선택한 경우, **최소값** 및 **최대값**에 대한 값을 제공하십시오.

   포아송
   선택한 경우 **평균** 모수값을 지정하십시오.

   지수
   선택할 경우 분포모수를 표본 평균(기본 설정) 또는 사용자 정의 설정에서 추정할지 여부를 지정할 수 있습니다. **표본 데이터 사용**을 선택한 경우 Lilliefors 검정이 사용됩니다. **사용자 정의**를 선택한 경우, **평균** 모수값을 제공하십시오.
4. 선택적으로 Monte Carlo 시뮬레이션 모수를 지정하려면 **시뮬레이션**을 클릭하고, 정확검정 모수를 지정하려면 **정확**을 클릭하며, 기술통계, 사분위수, 결측 데이터 처리 방법을 보려면 **옵션**을 클릭합니다.

일표본 *Kolmogorov-Smirnov* 검정*:* 시뮬레이션

분포의 특정 모수를 표본에서 추정해야 할 경우 Kolmogorov-Smirnov 검정이 더 이상 적용되지 않습니다. 이 경우, Lilliefors 검정 통계량을 사용하여 *p*-값을 추정할 수 있는데, 이를 위해서는 Monte Carlo 표본추출을 사용하여 알 수 없는 평균과 분산을 갖는 정규성을 검정하십시오. Lilliefors 검정은 세 개의 연속 분포(**정규**, **지수**, **균일**)에 적용됩니다. 기본 검정이 이산형(**포아송**)일 경우에는 이 검정이 적용되지 않습니다. 해당하는 분포모수가 지정되지 않은 경우에는 일표본 추론에 대해서만 검정이 정의됩니다.

Monte Carlo 시뮬레이션 모수

신뢰수준
이 선택적 설정은 Monte Carlo 시뮬레이션을 사용할 때 Kolmogorov-Smirnov 검정으로 추정되는 신뢰구간 수준을 재설정합니다. 값은 0 - 100 범위여야 합니다. 기본값은 99입니다.

표본 수
이 선택적 설정은 Lilliefors 검정 시 Monte Carlo 표본추출에 사용되는 복제 수를 재설정합니다. 값은 10000 - 최대 표본 수 값 범위에 있는 단일 정수여야 합니다. 기본값은 10000입니다.

정규 분포의 경우 Monte Carlo 결과 표시 안함
이 선택적 설정은 정규 분포 결과에 대해 Monte Carlo 표본추출을 표시하지 않습니다. 기본적으로 이 설정은 선택되어 있지 않습니다. 즉, 기존 근사 결과와 Monte Carlo 표본추출을 기반으로 하는 Lilliefors 검정 결과가 둘 다 표시됩니다.

일표본 *Kolmogorov-Smirnov* 검정*:* 옵션

통계
요약 통계 중 하나 또는 모두를 선택할 수 있습니다.

기술통계
평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수를 출력합니다.

사분위수
제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

결측값
결측값의 처리를 제어합니다.

검정별 결측값 제외
여러 검정을 지정할 때 각 검정은 결측값에 대해 별도로 평가됩니다.

목록별 결측값 제외
변수에 대한 결측값을 포함하는 케이스는 모든 분석에서 제외됩니다.

NPAR TESTS 명령 추가 기능(일표본 *Kolmogorov-Smirnov* 검정)

명령 구문을 사용하여 검정 분포의 모수를 지정할 수도 있습니다(K-S 하위 명령 포함).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

독립 2-표본 비모수 검정

독립 2-표본 비모수 검정 프로시저는 하나의 변수에 대해 두 개의 케이스 그룹을 비교합니다.

예. 더 편안하고 보기에도 좋으며 빠르게 치아를 교정할 수 있도록 새로운 치열 교정기가 개발되었습니다. 새 치열 교정기를 착용할 수 있는 기간을 이전 교정기와 비교하기 위해 이전 교정기를 착용할 어린이 10명과 새 교정기를 착용할 어린이 10명을 임의로 선택합니다. Mann-Whitney의 *U* 검정을 통해 새 교정기를 착용한 어린이의 경우 대체로 이전의 교정기를 착용한 어린이보다 짧은 기간에 같은 효과를 얻을 수 있음을 알 수 있습니다.

통계. 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수, 사분위수를 선택할 수 있습니다. Mann-Whitney의 *U*, Moses의 극단값반동, Kolmogorov-Smirnov의 *Z*, Wald-Wolfowitz run 검정을 선택할 수 있습니다.

독립 2-표본 비모수 검정 데이터 고려 사항

데이터. 순서를 지정할 수 있는 숫자변수를 사용합니다.

가정. 독립 표본과 무작위 표본을 사용합니다. Mann-Whitney *U* 검정은 두 개 분포가 동일한지 검정합니다. 두 분포 사이의 위치에서 차이를 검정하는 데 사용하려면 두 분포의 형태가 같다고 가정해야 합니다.

독립 2-표본 비모수 검정 결과 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 비모수 검정 > 레거시 대화 상자 > 2 독립 표본...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자변수를 선택합니다.
3. 그룹화 변수를 선택한 후 **그룹 정의**를 클릭하여 해당 파일을 두 개의 그룹이나 표본으로 분할합니다.

독립 *2*-표본 비모수 검정 유형

검정 유형. 독립 2-표본(그룹)이 같은 모집단으로부터 비롯된 것인지 확인하는 검정 방법은 다음 네 가지입니다.

Mann-Whitney U 검정은 가장 널리 사용되는 독립 2-표본 검정 방법입니다. 두 그룹에 대한 Wilcoxon 순위 합계 검정과 Kruskal-Wallis 검정은 동일합니다. Mann-Whitney는 2표본 모집단의 위치가 같은지 검정합니다. 두 그룹의 관측값은 서로 조합되어 동률시 할당되는 평균 순위로 순위가 지정됩니다. 동률의 수는 전체 관측값 수보다 작아야 합니다. 모집단의 위치가 똑같을 때는 두 표본 간의 순위가 무작위로 혼합됩니다. 검정을 통해 그룹 1의 스코어가 그룹 2의 스코어보다 앞선 횟수와 그룹 2의 스코어가 그룹 1의 스코어보다 앞선 횟수를 계산합니다. Mann-Whitney *U* 통계는 이 두 숫자 중 작은 통계입니다. 윌콕슨 순위 합계 *W* 통계도 표시됩니다. *W*는 그룹이 동일한 평균 순위를 갖지 않는 한 더 작은 평균 순위가 있는 그룹의 순위 합계입니다. 이 경우, 그룹에서 두 개의 독립 표본 정의 대화 상자에서 마지막으로 이름 지정된 그룹의 순위 합계입니다.

Kolmogorov-Smirnov의 Z 검정과 **Wald-Wolfowitz run 검정**은 더 일반적인 검정 방법으로 분포의 위치와 형태에 있어서 차이가 있는지 확인합니다. Kolmogorov-Smirnov 검정은 두 표본에 대해 관측된 누적 분포 함수 간의 최대 절대차를 기초로 합니다. 이 차이가 매우 클 경우 두 분포에 차이가 있음을 나타냅니다. Wald-Wolfowitz run 검정은 두 그룹의 관측값을 조합하고 순위를 지정합니다. 두 표본이 동일한 모집단의 표본이면 두 그룹은 순위 전체에 무작위로 분산되어야 합니다.

Moses의 극단값반동 검정은 실험 변수가 한 방향의 일부 개체와 반대 방향의 다른 개체에 영향을 준다고 가정합니다. 따라서 극단값반동을 대조군과 비교하여 검정합니다. 이러한 검정은 대조군의 범위에 중점을 두고 실험그룹이 대조군과 조합될 때 해당 실험그룹의 극단값이 그 범위에 얼마나 큰 영향을 주는지를 측정하도록 합니다. 대조군은 독립 2-표본 비모수 검정 그룹 정의 대화 상자의 그룹 1 값으로 정의됩니다. 두 그룹의 관측값이 조합되고 순위가 지정됩니다. 제어 그룹의 범위는 제어 그룹의 가장 큰 값과 가장 작은 값의 순위와 1을 더한 값 사이의 차이로 계산됩니다. 기회 이상값이 범위의 범위를 쉽게 왜곡할 수 있으므로 제어 케이스의 5%가 각 끝에서 자동으로 잘립니다.

독립 *2*-표본 비모수 검정*:* 그룹 정의

파일을 두 개의 그룹 또는 표본으로 분할하려면, 그룹 1의 정수 값을 입력하고 그룹 2에 대한 다른 값을 입력하십시오. 그 외 값이 있는 케이스는 분석에서 제외됩니다.

독립 *2*-표본 비모수 검정*:* 옵션

통계. 요약 통계 중 하나 또는 모두를 선택할 수 있습니다.

- **기술통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스의 수를 출력합니다.
- **사분위수.** 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

결측값. 결측값의 처리를 제어합니다.

- **검정별 결측값 제외.** 여러 검정을 지정할 때 각 검정은 결측값에 대해 별도로 평가됩니다.
- **목록별 결측값 제외.** 변수에 대한 결측값을 포함하는 케이스는 모든 분석에서 제외됩니다.

NPAR TESTS 명령 추가 기능*(*독립 *2*-표본 비모수 검정*)*

또한 명령 구문을 사용하여 Moses 검정에 대해 자를 케이스 수를 지정할 수 있습니다(MOSES 하위 명령 함께 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

대응 2-표본 비모수 검정

대응 2-표본 검정 프로시저는 두 변수의 분포를 비교합니다.

예. 일반적으로 한 가구가 집을 팔려고 할 때 부르는 대로 값을 받을 수 있습니까? Wilcoxon 부호순위 검정을 열 개의 가구에 적용하면 일곱 가구는 청구한 값보다 적게 받고 한 가구만 더 많이 받으며 두 가구는 청구한 대로 값을 받게 됩니다.

통계. 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수, 사분위수를 선택할 수 있습니다. Wilcoxon 부호순위, 부호, McNemar 검정을 선택할 수 있습니다. 정확검정 옵션이 설치된 경우(Windows 운영 체제에서만 사용 가능), 주변 동질성 검정도 사용할 수 있습니다.

대응 2-표본 비모수 검정 데이터 고려 사항

데이터. 순서를 지정할 수 있는 숫자변수를 사용합니다.

가정. 두 변수에 대해 특별한 분포가 가정되어 있지 않더라도 두 차이에 대한 해당 모집단 분포는 대칭인 것으로 가정합니다.

대응 2-표본 비모수 검정 결과 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 비모수 검정 > 레거시 대화 상자 > 2 대응 표본...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 대응변수를 선택합니다.

대응 *2*-표본 비모수 검정 유형

이 섹션의 검정에서는 두 관련 변수의 분포를 비교합니다. 사용할 검정은 데이터 유형에 따라 다릅니다.

데이터가 연속형인 경우 부호 검정이나 Wilcoxon 부호순위 검정을 사용합니다. **부호 검정**은 모든 케이스에 대해 두 변수 간 차이를 계산하고 그 차이를 양수, 음수 또는 동률인 수 등으로 분류합니다. 두 변수가 매우 유사하게 분포된 경우 양수나 음수 차이가 크게 구별되지 않을 수 있습니다. **Wilcoxon 부호순위 검정**은 이러한 변수 부호 차이와 대응 간 차이의 크기에 대한 정보를 모두 고려합니다. Wilcoxon 부호순위 검정은 데이터에 대한 정보를 구체화할 수 있으므로 부호 검정보다 강력한 검정 방법입니다.

데이터가 이분형이면 **McNemar 검정**을 사용합니다. 이 검정은 반복 측도 상황에서 주로 사용되므로 각 개체의 반응이 지정한 이벤트 발생 전후로 한 번씩 나타날 때 사용됩니다. McNemar 검정은 초기 반응률(이벤트 전)이 마지막 반응률(이벤트 후)과 같은지 확인합니다. 이러한 검정은 전후 계획에서 실험 개입에 따라 반응에 다르게 나타나는 변화를 파악하는 데 유용합니다.

데이터가 범주형이면 **주변 동질성 검정**을 사용합니다. 이 검정은 McNemar 검정을 이분형 반응에서 다항 반응으로 확장한 검정 방법입니다. 카이제곱 분포를 사용하여 반응의 변화를 확인하고 전후 계획에서 실험 개입에 따라 반응에 나타나는 변화를 파악하는데 유용합니다. 주변 동질성 검정은 정확한 검정을 설치한 경우에만 사용할 수 있습니다.

대응 *2*-표본 비모수 검정 옵션

통계. 요약 통계 중 하나 또는 모두를 선택할 수 있습니다.

- **기술통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스의 수를 출력합니다.
- **사분위수.** 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

결측값. 결측값의 처리를 제어합니다.

- **검정별 결측값 제외.** 여러 검정을 지정할 때 각 검정은 결측값에 대해 별도로 평가됩니다.
- **목록별 결측값 제외.** 변수에 대한 결측값을 포함하는 케이스는 모든 분석에서 제외됩니다.

NPAR TESTS 명령 추가 기능(대응 *2*-표본 비모수 검정)

또한 명령 구문을 사용하여 목록의 각 변수를 사용하여 특정 변수를 검정할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

K-독립 표본 비모수검정

독립 K-표본 비모수 검정 프로시저는 한 변수에 대해 둘 이상의 케이스 그룹을 비교합니다.

예. 세 가지 상표의 100-와트 전구는 전구가 타는 평균 시간에 차이가 있습니까? Kruskal-Wallis 일원 분산 분석을 통해 세 가지 상표의 전구가 평균 수명에서 차이가 있음을 알 수 있습니다.

통계. 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수, 사분위수를 선택할 수 있습니다. Kruskal-Wallis의 *H*, 중위수 검정을 선택할 수 있습니다.

독립 K-표본 비모수 검정 데이터 고려 사항

데이터. 순서를 지정할 수 있는 숫자변수를 사용합니다.

가정. 독립 표본과 무작위 표본을 사용합니다. Kruskal-Wallis의 *H* 검정에서는 검정할 표본의 형태가 유사해야 합니다.

독립 K-표본 비모수 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 비모수 검정 > 레거시 대화 상자 > K 독립 표본...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자변수를 선택합니다.
3. 그룹화 변수를 하나 선택하고 **범위지정**을 클릭하여 이 그룹화 변수에 대한 최소값 및 최대값을 정수로 지정합니다.

독립 *K*-표본 비모수 검정 유형

일부 독립 표본이 동일한 모집단에서 비롯된 것인지를 판별하는 데 사용할 수 있는 검정 방법에는 세 가지가 있습니다. Kruskal-Wallis의 *H* 검정, 중위수 검정, Jonckheere-Terpstra 검정은 모두 여러 개의 독립 표본이 동일한 모집단에서 비롯된 것인지를 검정합니다.

Mann-Whitney *U* 검정의 확장인 **Kruskal-Wallis H 검정**은 단방향 분산 분석의 비모수 아날로그이며 분포 위치의 차이를 감지합니다. 강력하지는 않지만 더 일반적인 **중위수 검정**은 위치와 형태에서 분포 차이를 찾아냅니다. Kruskal-Wallis의 *H* 검정과 중위수 검정은 해당 표본을 가져온 *k* 모집단의 연역 순위가 없다고 가정합니다.

k 모집단에 원래부터 연역 순위(오름차순이나 내림차순)가 있는 경우 **Jonckheere-Terpstra 검정**이 더 강력한 방법입니다. 예를 들어, *k* 모집단은 증가 온도 *k*를 나타낼 수도 있습니다. 서로 다른 온도가 동일한 반응 분포를 생성한다는 가설이 온도가 증가할수록 반응의 크기도 증가한다는 가설에 반대되지 않는지 검정합니다. 여기에는 대체 가설에 순위가 있으므로 Jonckheere-Terpstra가 사용하기에 가장 적절한 검정입니다. Jonckheere-Terpstra 검정은 Exact Tests 추가 기능 모듈을 설치한 경우에만 사용할 수 있습니다.

독립 *K*-표본 비모수 검정*:* 범위지정

범위를 지정하려면 그룹화 변수의 최하 범주와 최고 범주에 해당하는 **최소값**과 **최대값**으로 정수값을 입력합니다. 경계를 벗어난 값을 포함하는 케이스는 제외됩니다. 예를 들어, 최소값 1과 최대값 3을 지정하면 1에서 3까지의 정수값만 사용됩니다. 최소값은 최대값보다 작아야 하며 두 값을 모두 지정해야 합니다.

독립 *K*-표본 비모수 검정*:* 옵션

통계. 요약 통계 중 하나 또는 모두를 선택할 수 있습니다.

- **기술통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스의 수를 출력합니다.
- **사분위수.** 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

결측값. 결측값의 처리를 제어합니다.

- **검정별 결측값 제외.** 여러 검정을 지정할 때 각 검정은 결측값에 대해 별도로 평가됩니다.
- **목록별 결측값 제외.** 변수에 대한 결측값을 포함하는 케이스는 모든 분석에서 제외됩니다.

NPAR TESTS 명령 추가 기능*(*독립 *K*-표본 비모수 검정*)*

또한 명령 구문을 사용하여 중위수 검정에 대해 관측된 중위수 이외의 값을 지정할 수 있습니다(MEDIAN 하위 명령 함께 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

K-대응 표본 비모수검정

대응 K-표본 비모수 검정 프로시저는 둘 이상의 변수 분포를 비교합니다.

예. 의사, 변호사, 경찰 공무원, 선생님이 어느 정도의 특권을 가지고 있다고 대중이 생각하는지 조사하기 위해 열 명에게 이러한 네 가지 직업을 특권 순위에 따라 정렬해 줄 것을 요청했습니다. Friedman의 검정으로 사실상 대중이 이들 네 가지 직업이 서로 다른 특권을 가지고 있는 것으로 인식한다는 결과를 알 수 있습니다.

통계. 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수, 사분위수를 선택할 수 있습니다. Friedman, Kendall의 *W*, Cochran의 *Q* 검정을 선택할 수 있습니다.

대응 K-표본 비모수 검정 데이터 고려 사항

데이터. 순서를 지정할 수 있는 숫자변수를 사용합니다.

가정. 비모수 검정은 기본 분포의 형태에 대한 가정이 필요하지 않습니다. 종속표본과 무작위 표본을 사용합니다.

대응 K-표본 비모수 검정 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 비모수 검정 > 레거시 대화 상자 > K 대응 표본...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

2. 둘 이상의 숫자 검정변수를 선택합니다.

대응 *K*-표본 비모수 검정 유형

여러 대응변수의 분포를 비교하는 데는 세 가지 검정을 사용할 수 있습니다.

Friedman 검정은 일표본 반복 측도 계획의 비모수 분석이거나 셀당 하나의 관측값을 갖는 2원배치 분산 분석을 나타냅니다. Friedman은 대응변수 *k*가 동일한 모집단에서 비롯되었음을 나타내는 귀무가설을 검정합니다. 각 케이스의 경우, *k* 변수의 순위는 1에서 *k*까지입니다. 검정 통계는 이러한 순위를 기반으로 합니다.

Kendall의 W는 Friedman 통계의 정규성을 나타냅니다. Kendall의 *W*는 등급 평가자 간의 동의 측도인 일치계수로 해석할 수 있습니다. 각 케이스는 감정 평가나 등급 평가자이며 각 변수는 평가되는 품목이나 사람을 나타냅니다. 각 변수의 경우 순위합이 계산됩니다. Kendall의 *W*는 0(동의 안함)과 1(완전 동의) 사이의 순위를 차지합니다.

Cochran의 Q는 Friedman 검정과 동일하나 모든 반응이 이분형인 경우 적용할 수 있습니다. 이 검정은 McNemar 검정을 *k*-표본 상황으로 확장한 것입니다. Cochran의 *Q*는 여러 개의 이분형 대응변수가 동일한 평균을 가진다는 가설을 검정합니다. 동일한 개인이나 일치된 개인에 따라 변수가 측정됩니다.

대응 *K*-표본 비모수 검정 통계량

통계를 선택할 수 있습니다.

- **기술통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스의 수를 출력합니다.
- **사분위수.** 제 25, 50, 75 백분위수에 해당하는 값을 출력합니다.

NPAR TESTS 명령 추가 기능(대응 *K*-표본 비모수 검정)

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

다중 응답 분석

다중 응답 분석

다중 이분형 및 다중 범주 세트를 분석하는 데 두 프로시저를 사용할 수 있습니다. 다중반응 빈도분석 프로시저는 빈도표를 표시합니다. 다중반응 교차표 프로시저는 2-3차원 교차표를 표시합니다. 프로시저를 사용하기 전에 다중 응답 세트를 정의해야 합니다.

예. 이 예제는 시장 여론 설문조사에서 다중반응 항목의 사용을 보여줍니다. 데이터는 가상이며 실제로 해석해서는 안 됩니다. 한 항공사가 경쟁 항공사를 평가하기 위해 특정 항로를 여행하는 승객들을 설문조사할 수 있습니다. 이 예제에서 American Airlines는 시카고-뉴욕 항로에서 다른 항공사를 이용하는 승객과 항공사를 선택할 때 일정과 서비스의 상대적인 중요성에 대해 알고 싶어합니다. 비행 승무원이 탑승하는 승객들에게 간단한 질문지를 나누어 줍니다. 첫 번째 질문은 American, United, TWA, USAir, 기타 중에서 이 항로에서 최근 6개월 간 한 번 이상 이용한 적이 있는 항공사를 모두 선택하라는 것입니다. 이것은 승객이 둘 이상의 답을 선택할 수 있으므로 다중반응 질문입니다. 그러나 이 질문은 변수가 각 케이스에 하나의 값만 가질 수 있기 때문에 직접 코딩할 수 없습니다. 여러 변수를 사용하여 각 질문에 반응을 매핑해야 합니다. 이를 수행하는 방법은 두 가지입니다. 하나는 각 선택(예: American, United, TWA, USAir 및 기타)에 해당하는 변수를 정의하는 것입니다. 승객이 United를 선택하면 *united* 변수에 코드 1이 지정되며 그렇지 않은 경우 0이 지정됩니다. 이것이 맵핑 변수의 **다중 이분형 방법**입니다. 반응을 매핑하는 다른 방법은 질문에 가능한 최대 반응 수를 평가하고 항공사를 지정하는 데 사용되는 코드를 사용하여 같은 수의 변수를 설정하는 **다중 범주형 방법**입니다. 질문지 표본을 읽으면 사용자가 최근 6개월 동안 이 항로에서 세 곳 이상 다른 항공사를 이용한 적이 없다는 것을 발견할 수 있습니다. 또한 항공사의 자유화로 인해 10개의 다른 항공사가 기타 범주에 이름을 올렸습니다. 다중반응 방법을 사용하면 3개의 변수를 정의합니다. 각 코드는 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta* 등입니다. 해당 승객이 American과 TWA를 선택하는 경우 첫째 변수는 코드 1, 둘째 변수는 코드 3, 셋째 변수에는 결측값 코드를 갖습니다. 또 다른 승객은 American을 선택하고 Delta를 입력할 수 있습니다. 따라서 첫째 변수는 코드 1, 둘째 변수는 코드 5, 셋째 변수는 결측값 코드를 갖습니다. 반면에 다중 이분형 방법을 사용하는 경우 14개의 다른 변수를 이용합니다. 이 설문조사에 어느 매핑 방법이나 사용할 수 있지만 선택하는 방법은 반응 분포에 따라 다릅니다.

다중 응답 세트 정의

다중 응답 세트 정의 프로시저로 기본 변수를 다중 이분 세트나 다중 범주 세트로 그룹화하여 빈도표와 교차표를 구할 수 있습니다. 다중 응답 세트를 20개까지 정의할 수 있습니다. 각 변수 세트의 이름은 고유한 이름이어야 합니다. 변수 세트를 삭제하려면 다중 응답 세트 목록에서 삭제할 변수 세트를 선택하고 **제거**를 클릭합니다. 변수 세트를 바꾸려면 목록에서 선택한 다음 변수 세트 정의 특성을 수정하고 **바꾸기**를 클릭합니다.

기본 변수를 이분형이나 범주형으로 코딩할 수 있습니다. 이분형 변수를 사용하려면 **이분형**을 선택하여 다중 이분 세트를 작성합니다. 개수화 값에 대해 정수 값을 입력합니다. 개수화 값의 발생이 한 번 이상인 변수는 다중 이분 세트의 범주가 됩니다. **범주형**을 선택하여 성분 변수와 값 범위가 같은 다중 범주 세트를 작성합니다. 다중 범주 세트의 범주에 대한 최소 및 최대 범위 값에 해당하는 정수 값을 입력합니다. 프로시저에서는 모든 성분 변수의 포함 범위에 서로 다른 정수 값의 총계를 냅니다. 빈 범주는 표에 포함되지 않습니다.

각 다중 응답 세트에 고유 이름을 7자까지 할당할 수 있습니다. 프로시저에는 할당한 변수 이름 앞에 달러 부호($)가 붙습니다. *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* 및 *width*는 예약된 이름으로 사용할 수 없습니다. 다중 응답 세트의 이름은 다중반응 프로시저에서만 사용되므로 다른 프로시저에서는 참조할 수 없습니다. 선택적으로 다중 응답 세트에 대한 변수 설명 레이블을 입력할 수 있습니다. 레이블은 최대 40자까지 사용할 수 있습니다.

다중 응답 세트 정의 방법

1. 메뉴에서 다음을 선택합니다.

   분석 > 다중 응답 > 변수 세트 정의...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 변수를 두 개 이상 선택합니다.
3. 변수가 이분형으로 코딩되었을 때는 계산하려는 값을 지정합니다. 변수가 범주형으로 코딩되었을 때는 범주 범위를 정의합니다.
4. 각 다중 응답 세트에 대한 고유 이름을 입력합니다.
5. 다중 응답 세트를 정의한 변수 세트 목록에 추가하려면 **추가**를 클릭합니다.

다중반응 빈도분석

다중반응 빈도분석 프로시저로 다중 응답 세트에 대한 빈도표를 작성합니다. 먼저 하나 이상의 다중 응답 세트를 정의해야 합니다. "다중반응변수 세트 정의"를 참조하십시오.

다중 이분 세트에서는 결과에 나타난 범주 이름이 그룹의 기본 변수로 정의된 변수 레이블에서 비롯됩니다. 변수 레이블이 정의되지 않은 경우에는 변수 이름을 레이블로 사용합니다. 다중 범주 세트에서는 범주 레이블이 그룹의 첫 번째 값 레이블에서 비롯됩니다. 첫 번째 변수에 대해 누락된 범주가 그룹의 다른 변수에 대해 표시될 때는 결측 범주에 대한 값 레이블을 정의합니다.

결측값. 결측값이 있는 케이스는 표-대-표 기준에서는 제외됩니다. 선택적으로 다음 중 하나 또는 모두를 선택할 수 있습니다.

- **이분형 결측데이터의 목록별 제외.** 다중 이분 세트의 표 작성으로부터 임의의 변수에 대한 결측값이 있는 케이스는 제외합니다. 이분형 변수 세트로 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다중 이분 세트의 성분 변수에 개수화 값이 없을 때 결측된 것으로 간주됩니다. 일부 변수에 대한 결측값이 있는 케이스는 하나 이상의 변수에 개수화 값이 있을 경우 그룹의 표 작성 작업에 포함됩니다.
- **범주형 결측데이터의 목록별 제외.** 다중 범주 세트의 표 작성에서 임의의 변수에 대한 결측값이 있는 케이스는 제외됩니다. 범주형 변수 세트로 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다중 범주 세트의 성분 변수에 지정한 범위 내 유효한 값이 없을 때 결측된 것으로 간주됩니다.

예. 설문조사 질의로부터 작성한 각 변수는 기본 변수가 됩니다. 다중반응 항목을 분석하려면 이러한 변수를 다중 이분 세트나 다중 범주 세트인 다중 응답 세트 중 하나로 변수를 결합해야 합니다. 예를 들어, 한 항공사가 고객에게 최근 6개월 동안 세 항공사(American, United, TWA) 중 어떤 항공사를 이용했는지를 설문조사하고 이분형 변수를 사용하여 **다중 이분 세트**를 정의할 경우 해당 변수 세트의 세 변수는 각각 그룹화 변수의 범주가 됩니다. 세 항공사의 개수와 퍼센트가 하나의 빈도표에 나타납니다. 세 항공사를 모두 선택한 반응자가 없을 경우 두 변수를 작성하여 각각 세 항공사에 하나씩 코드를 부여할 수 있습니다. **다중 범주 세트**를 정의할 경우에는 기본 변수에 같은 코드를 추가하여 값을 표로 작성할 수 있습니다. 결과로 나타나는 값의 변수 세트는 각 기본 변수에 대한 값

과 같습니다. 예를 들어, 유나이티드에 대한 30개의 응답은 항공사 1에 대한 5개의 응답과 항공사 2에 대한 25개의 유나이티드 응답의 합계입니다. 세 항공사의 개수와 퍼센트가 하나의 빈도표에 나타납니다.

통계. 개수가 표시된 빈도표, 반응 퍼센트, 케이스 퍼센트, 유효 케이스 수, 결측 케이스 수를 선택할 수 있습니다.

다중반응 빈도분석 데이터 고려 사항

데이터. 다중 응답 세트를 사용합니다.

가정. 개수와 퍼센트를 통해 모든 분포에서 데이터에 대한 유용한 설명을 볼 수 있습니다.

관련 프로시저. 다중 응답 세트 정의 프로시저로 다중 응답 세트를 정의할 수 있습니다.

다중반응 빈도분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 다중 응답 > 빈도...
2. 다중 응답 세트를 하나 이상 선택합니다.

다중반응 교차분석

다중반응 교차 분석표 프로시저로 정의한 다중 응답 세트, 기본 변수, 조합 등을 교차 분석표로 만들 수 있습니다. 또한, 케이스나 반응을 기초로 셀 퍼센트를 구하고 결측값 처리를 수정하거나 대응 교차표를 만들 수 있습니다. 먼저 하나 이상의 다중 응답 세트를 정의해야 합니다. "다중 응답 세트 정의"를 참조하십시오.

다중 이분 세트에서는 결과에 나타난 범주 이름이 그룹의 기본 변수로 정의된 변수 레이블에서 비롯됩니다. 변수 레이블이 정의되지 않은 경우에는 변수 이름을 레이블로 사용합니다. 다중 범주 세트에서는 범주 레이블이 그룹의 첫 번째 값 레이블에서 비롯됩니다. 첫 번째 변수에 대해 누락된 범주가 그룹의 다른 변수에 대해 표시될 때는 결측 범주에 대한 값 레이블을 정의합니다. 프로시저에서는 열 범주 레이블을 한 줄마다 최대 8자씩 세 줄로 표시합니다. 단어가 잘리지 않도록 하려면 행과 열 항목을 뒤바꾸고 레이블을 다시 정의합니다.

예. 다중 이분 세트와 다중 범주 세트는 모두 이 프로시저의 다른 변수로 상호 교차되어 작성될 수 있습니다. 항공기 승객 설문조사에서는 승객에게 American, United, TWA 중에서 최근 6개월 간 한 번 이상 이용한 적이 있는 항공사를 모두 선택하고, 항공사 선택 시 비행 스케줄과 서비스 중 중요하게 고려하는 요인을 하나 선택하라는 질문이 있을 수 있습니다. 데이터를 이분형이나 다중 범주형으로 입력한 다음 하나의 변수 세트로 조합하면 서비스와 스케줄이 포함된 질문으로 항공사 선택을 교차 분석표화할 수 있습니다.

통계. 셀, 행, 열, 총계와 셀, 행, 열, 전체 퍼센트를 사용한 교차표를 선택할 수 있습니다. 셀 퍼센트는 케이스나 반응에 따라 달라집니다.

다중반응 교차 분석표 데이터 고려 사항

데이터. 다중 응답 세트나 숫자 범주형 변수를 사용합니다.

가정. 개수와 퍼센트를 통해 분포에서 데이터에 대한 유용한 설명을 볼 수 있습니다.

관련 프로시저. 다중 응답 세트 정의 프로시저로 다중 응답 세트를 정의할 수 있습니다.

다중반응 교차 분석표 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 다중 응답 > 교차분석도...
2. 각 교차표 차원에 대해 여러 숫자변수나 다중 응답 세트를 선택합니다.
3. 각 기본 변수의 범위를 지정합니다.

선택적으로 제어변수나 다중 응답 세트의 범주마다 2차원 교차표를 사용할 수 있습니다. 레이어 목록에 대한 항목을 하나 이상 선택합니다.

다중반응 교차 분석표 범위지정

값 범위는 교차표의 기본 변수에 대해 정의해야 합니다. 표로 만들 범주의 최소값과 최대값을 정수로 입력합니다. 범위 밖의 범주는 분석 시 제외됩니다. 범위 내 값은 정수로 가정되며 정수가 아닌 수는 잘립니다.

다중반응 교차 분석표: 옵션

셀 퍼센트. 셀 개수는 항상 표시됩니다. 행 퍼센트, 열 퍼센트, 두 가지 방법의 테이블(전체) 퍼센트 등을 선택하여 표시할 수 있습니다.

퍼센트 계산 기준. 셀 퍼센트를 케이스(또는 반응자)에 대한 기준으로 삼을 수 있습니다. 다중 범주 세트에 대해 변수 일치를 선택할 때는 이렇게 할 수 없습니다. 셀 퍼센트를 반응에 대한 기준으로 삼을 수 있습니다. 다중 이분 세트에서는 반응 수가 케이스에 대한 개수화 값 수와 같습니다. 다중 범주 세트에서는 반응 수가 지정 범위 내 값의 수와 같습니다.

결측값. 다음 중 하나 또는 모두를 선택할 수 있습니다.

- **이분형 결측데이터의 목록별 제외.** 다중 이분 세트의 표 작성으로부터 임의의 변수에 대한 결측값이 있는 케이스는 제외합니다. 이분형 변수 세트로 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다중 이분 세트의 성분 변수에 개수화 값이 없을 때 결측된 것으로 간주됩니다. 일부 변수에 대한 결측값이 있는 케이스는 하나 이상의 변수에 개수화 값이 있을 때 그룹 표에 포함됩니다.
- **범주형 결측데이터의 목록별 제외.** 다중 범주 세트의 표 작성에서 임의의 변수에 대한 결측값이 있는 케이스는 제외됩니다. 범주형 변수 세트로 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다중 범주 세트의 성분 변수에 지정한 범위 내 유효한 값이 없을 때 결측된 것으로 간주됩니다.

기본적으로 두 다중 범주 세트로 교차 분석표를 수행할 때 프로시저에서는 두 번째 그룹의 각 변수와 함께 첫 번째 그룹의 각 변수를 교차분석하고 각 셀에 대한 개수를 합합니다. 따라서 일부 반응은 테이블에 두 번 이상 나타날 수 있습니다. 다음 옵션을 선택할 수 있습니다.

변수 세트 간의 반응변수순으로 통합. 첫 번째 그룹의 첫 번째 변수와 두 번째 그룹의 첫 번째 변수를 대응시키고 첫 번째 그룹의 두 번째 변수와 두 번째 그룹의 두 번째 변수, 이런 식으로 대응시킵니다. 이 옵션을 선택한 경우, 프로시저에서는 반응자보다는 반응에 대한 셀 퍼센트를 기준으로 합니다. 다중 이분 세트 또는 기본 변수에 대해서는 대응을 사용할 수 없습니다.

MULT RESPONSE 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 최대 5 차원까지 교차 분석표를 작성하십시오(BY 하위 명령 사용).
- 값 레이블 삭제 등 결과 형식 지정 옵션을 변경하십시오(FORMAT 하위 명령 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

결과 보고

결과 보고

케이스 목록 및 기술통계는 데이터를 연구하고 설명하는 기본 도구입니다. 데이터 편집기나 요약 프로시저를 사용하여 케이스 목록을, 빈도 프로시저를 사용하여 빈도 개수와 기술통계를, 평균 프로시저를 사용하여 부모집단 통계를 얻을 수 있습니다. 각각 정보를 쉽게 파악할 수 있도록 계획된 형식을 사용합니다. 다른 형식으로 정보를 출력하고 싶은 경우 행별 요약 보고서 및 열별 요약 보고서를 사용하면 데이터 프레젠테이션을 원하는 대로 제어할 수 있습니다.

행별 요약 보고서

행별 요약 보고서로 행에 서로 다른 요약 통계가 나타나 있는 보고서를 작성할 수 있습니다. 요약 통계가 표시되거나 표시되지 않은 케이스 목록도 사용할 수 있습니다.

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

예. 소매점을 체인으로 운영하는 회사에서 급여, 직위, 직원이 근무하는 점포 및 부서 등의 직원 정보 레코드를 보유하고 있다고 가정합니다. 각 점포와 부서, 점포 내 부서에 대한 요약 통계(예: 평균 급여)와 함께 점포와 부서(구분변수)별로 구분된 개별 직원에 대한 정보(목록)가 있는 보고서를 작성할 수 있습니다.

데이터 열. 케이스 목록이나 요약 통계에 대한 보고서 변수의 목록을 작성하고 데이터 열의 표시 형식을 제어합니다.

구분 열. 보고서를 그룹으로 나누는 구분변수를 목록으로 만들고 구분 열의 요약 통계와 표시 형식을 제어할 수 있습니다. 다중 구분변수의 경우 목록에서 선행하는 구분변수의 범주 내에 각 구분변수의 범주마다 서로 다른 그룹이 존재합니다. 구분변수는 케이스를 제한된 범주 수로 나누는 이산형 범주변수이어야 합니다. 각 구분변수의 개별 값은 전체 데이터 열의 왼쪽에 별도의 열에 표시 및 정렬됩니다.

보고서. 전체 요약 통계, 결측값 표시, 쪽 매김, 제목 등의 전체 보고서 특성을 제어합니다.

케이스 출력. 케이스마다 데이터 열 변수의 실제 값(또는 값 레이블)을 표시합니다. 요약 보고서보다 긴 목록 보고서를 만들 수 있습니다.

미리 보기. 보고서의 첫 번째 쪽만 표시됩니다. 이 옵션으로 전체 보고서 작업을 하지 않고도 형식을 미리 볼 수 있습니다.

데이터가 이미 정렬되어 있음. 구분변수가 표시된 보고서에서 데이터 파일을 구분변수값으로 정렬한 후에 보고서를 작성합니다. 데이터 파일이 이미 구분변수값으로 정렬되었을 경우에는 이 옵션을 선택하면 처리 시간을 줄일 수 있습니다. 이 옵션은 보고서 미리보기를 실행한 다음에 특히 유용합니다.

요약 보고서 구하기: 행별 요약

1. 메뉴에서 다음을 선택합니다.

   분석 > 보고서 > 행으로 보고서 요약...
2. 데이터 열에 대한 변수를 하나 이상 선택합니다. 보고서에서 선택한 각 변수마다 열이 하나씩 작성됩니다.
3. 하위 그룹별로 정렬하여 표시한 보고서의 경우 구분 열 변수를 하나 이상 선택합니다.
4. 구분변수로 정의한 하위 그룹의 요약 통계가 표시된 보고서의 경우, 구분 열 변수 목록에서 구분변수를 선택하고 구분 열에서 **요약**을 클릭하여 요약 측도를 지정합니다.
5. 전체 요약 통계가 표시된 보고서에 대해 **요약**을 클릭하여 요약 측도를 지정합니다.

보고서 데이터 열/구분 형식

형식 대화 상자에서 열 제목, 열 자리수, 텍스트 맞춤, 데이터 값이나 값 레이블 표시 등을 제어할 수 있습니다. 데이터 열 형식은 보고서 오른쪽에 있는 데이터 열의 형식을 제어합니다. 구분 형식은 왼쪽에 있는 구분 열의 형식을 제어합니다.

열 제목. 선택한 변수에 대해 열 제목을 제어합니다. 제목이 길면 자동으로 열 내에서 줄바꿈이 됩니다. Enter 키를 사용하여 원하는 위치에서 제목의 줄이 바뀌도록 수동으로 줄 바꿈을 삽입할 수 있습니다.

열 내부에서 값의 위치. 선택한 변수에 대해 데이터 값이나 값 레이블 맞춤을 열 내에서 제어합니다. 값이나 레이블의 맞춤은 열 머리말의 맞춤에 영향을 주지 않습니다. 문자 수를 지정하여 열 내용을 들여쓰거나 내용을 가운데로 맞출 수 있습니다.

열 내용. 선택한 변수값에 대해 데이터 값이나 정의한 값 레이블의 표시를 제어합니다. 정의된 값 레이블이 없는 값에 대한 데이터 값은 항상 표시됩니다. 이 옵션은 열 요약 보고서에서 데이터 열에 대해 사용할 수 없습니다.

보고서 요약 행/최종 요약 행

두 가지의 요약 행 대화 상자에서 구분 그룹과 전체 보고서에 대한 요약 통계를 표시할 수 있습니다. 요약 행은 구분변수로 정의한 각 범주에 대한 하위 그룹 통계를 제어합니다. 최종 요약 행은 보고서의 끝에 표시된 전체 통계를 제어합니다.

사용 가능한 요약 통계는 합계, 평균, 최소값, 최대값, 케이스 수, 지정 값을 초과하거나 미달하는 케이스 비율, 값의 지정 범위 내 케이스 비율, 표준 편차, 분산, 첨도, 왜도 등입니다.

보고서 구분: 옵션

구분 옵션으로 구분 범주 정보의 간격과 쪽 매김을 제어합니다.

쪽 제어. 선택한 구분변수의 범주에 대해 간격이나 쪽 매김을 제어할 수 있습니다. 구분 범주 사이에 공백줄의 번호를 지정하거나 새 쪽에서 각 구분 범주로 시작할 수 있습니다.

요약 전에 공백줄. 구분 범주 레이블이나 데이터와 요약 통계 사이에 공백줄 수를 제어합니다. 이 옵션은 구분 범주에 대해 두 가지의 개별 케이스 목록과 요약 통계를 포함하는 결합된 보고서에 유용하며 케이스 목록과 요약 통계 사이에 공백을 삽입할 수도 있습니다.

보고서: 옵션

보고서 옵션은 결측값과 보고서 쪽 매김의 표시 및 처리 방식을 제어합니다.

목록별 결측값을 가지는 케이스 제외. 보고서 변수에 대한 결측값이 있는 케이스를 보고서에서 제거합니다.

결측값 나타내기. 데이터 파일에서 결측값을 나타내는 기호를 지정할 수 있습니다. 기호는 단일 문자로서 시스템-결측값과 사용자 결측값을 모두 나타낼 수 있습니다.

시작 쪽번호. 보고서의 첫 번째 쪽에 대한 쪽 번호를 지정할 수 있습니다.

보고서: 레이아웃

보고서 레이아웃에서 각 보고서 쪽의 너비와 길이, 쪽에서 보고서 위치, 공백줄과 레이블 삽입 등을 제어합니다.

쪽 레이아웃. 줄(위/아래)과 문자(왼쪽과 오른쪽)로 표시되는 쪽 여백과 여백 내에 정렬할 보고서 맞춤 등을 제어합니다.

쪽 제목 및 꼬리말. 보고서 본문에서 쪽 제목과 꼬리말을 구분하는 줄 수를 제어합니다.

구분 열. 구분 열 표시를 제어합니다. 다중 구분변수가 지정되어 있으면 구분 열은 별도의 열이나 첫 번째 열에 위치할 수 있습니다. 모든 구분변수를 첫 번째 열에 두면 보고서 너비를 줄일 수 있습니다.

열 제목. 밑줄 제목, 보고서 제목과 본문 사이의 간격, 열 제목의 수직 맞춤 등으로 열 제목 표시를 제어합니다.

데이터 열 행 및 구분레이블. 데이터 열 정보(데이터 값 또는 요약 통계)의 위치를 각 구분 범주의 시작 부분에서 구분 레이블과 관련하여 제어합니다. 데이터 열 정보의 첫 번째 행은 구분 범주 레이블과 같은 줄이나 구분 범주 레이블 뒤에 줄 번호를 지정하여 시작할 수 있습니다. 이 옵션은 열 요약 보고서에 대해서는 사용할 수 없습니다.

보고서: 제목

보고서: 제목 대화 상자에서 보고서 제목과 꼬리말의 내용과 위치를 제어합니다. 쪽 제목과 꼬리말을 각각 10줄까지 지정할 수 있고 왼쪽 맞춤, 가운데 맞춤, 오른쪽 맞춤 등을 각 줄마다 지정할 수도 있습니다.

변수를 제목이나 꼬리말에 삽입하면 현재 값 레이블이나 변수값이 제목이나 꼬리말에 표시됩니다. 제목에 쪽 시작 부분에 있는 변수값에 해당하는 값 레이블이 표시됩니다. 꼬리말에 쪽의 마지막 부분에 있는 변수값에 해당하는 값 레이블이 표시됩니다. 값 레이블이 없을 경우에는 실제 값이 표시됩니다.

특수변수. 특수변수인 *DATE*와 *PAGE*로 현재 날짜와 쪽 번호를 보고서 머리말이나 꼬리말에 삽입할 수 있습니다. 데이터 파일에 *DATE*나 *PAGE*라는 이름의 변수가 있을 경우 이러한 변수를 보고서 제목이나 꼬리말에 사용할 수 없습니다.

열별 요약 보고서

열별 요약 보고서로 서로 다른 요약 통계가 별도의 열에 표시되는 요약 보고서를 만들 수 있습니다.

참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.

예. 소매점을 체인으로 운영하는 회사에서 급여, 직위, 부서 등의 직원 정보 레코드를 보유하고 있다고 가정합니다. 각 부서별로 요약 급여 통계(예: 평균, 최소값, 최대값 등)를 나타내는 보고서를 작성할 수 있습니다.

데이터 열. 요약 통계가 있는 보고서 변수를 목록으로 만들어 각 변수에 대한 표시 형식과 요약 통계를 제어할 수 있습니다.

구분 열. 보고서를 그룹으로 나누는 구분변수를 나열하고 구분 열의 표시 형식을 제어할 수 있습니다. 다중 구분변수의 경우 목록에서 선행하는 구분변수의 범주 내에 각 구분변수의 범주마다 서로 다른 그룹이 존재합니다. 구분변수는 케이스를 제한된 범주 수로 나누는 이산형 범주변수이어야 합니다.

보고서. 결측값 표시, 쪽 매김, 제목 등의 전체 보고서 특성을 제어합니다.

미리 보기. 보고서의 첫 번째 쪽만 표시됩니다. 이 옵션으로 전체 보고서 작업을 하지 않고도 형식을 미리 볼 수 있습니다.

데이터가 이미 정렬되어 있음. 구분변수가 표시된 보고서에서 데이터 파일을 구분변수값으로 정렬한 후에 보고서를 작성합니다. 데이터 파일이 이미 구분변수값으로 정렬되었을 경우에는 이 옵션을 선택하면 처리 시간을 줄일 수 있습니다. 이 옵션은 보고서 미리보기를 실행한 다음에 특히 유용합니다.

요약 보고서 구하기: 열별 요약

1. 메뉴에서 다음을 선택합니다.

   분석 > 보고서 > 열로 보고서 요약...
2. 데이터 열에 대한 변수를 하나 이상 선택합니다. 보고서에서 선택한 각 변수마다 열이 하나씩 작성됩니다.
3. 변수에 대한 요약 측도를 바꾸려면 데이터 열 변수 목록에서 변수를 선택하고 **요약**을 클릭합니다.
4. 하나의 변수에 대해 요약 측도를 둘 이상 구하려면 원시 목록에서 원하는 요약 측도마다 변수를 선택하여 데이터 열 변수 목록으로 여러 번 옮깁니다.
5. 합계, 평균, 비율 등 기존 열의 다른 함수를 포함하는 열을 표시하려면 **총계 삽입**을 클릭합니다. 데이터 열 목록에 총계라는 변수가 표시됩니다.
6. 하위 그룹별로 정렬하여 표시한 보고서의 경우 구분 열 변수를 하나 이상 선택합니다.

데이터 열 요약 함수

요약 행으로 선택한 데이터 열 변수에 대해 표시된 요약 통계를 제어할 수 있습니다.

사용 가능한 요약 통계는 합계, 평균, 최소값, 최대값, 케이스 수, 지정 값을 초과하거나 미달하는 케이스 비율, 값의 지정 범위 내 케이스 비율, 표준 편차, 분산, 첨도, 왜도 등입니다.

총계 열에 대한 데이터 열 요약

열 요약으로 둘 이상의 데이터 열을 요약하는 총계 요약 통계를 제어합니다.

사용 가능한 총계 요약 통계는 열 총계, 열 평균, 최대값, 최소값, 두 열 값의 차, 한 열 값을 다른 열의 값으로 나눈 몫, 열 값을 모두 곱한 곱 등입니다.

열 총계. 총계 열은 요약 열 목록에 있는 열의 합입니다.

열 평균. 총계 열은 요약 열 목록에 있는 열의 평균입니다.

열 최소값. 총계 열은 요약 열 목록에 있는 열의 최소값입니다.

열 최대값. 총계 열은 요약 열 목록에 있는 열의 최대값입니다.

첫 번째 열 – 두 번째 열. 총계 열은 요약 열 목록에 있는 열의 차입니다. 요약 열 목록에는 두 개의 열이 있어야 합니다.

첫 번째 열/ 두 번째 열. 총계 열은 요약 열 목록에 있는 열의 몫입니다. 요약 열 목록에는 두 개의 열이 있어야 합니다.

% 첫 번째 열/ 두 번째 열. 총계 열은 요약 열 목록에서 두 번째 열에 대한 첫 번째 열의 비율입니다. 요약 열 목록에는 두 개의 열이 있어야 합니다.

열 작업 결과. 총계 열은 요약 열 목록에 있는 열의 결과입니다.

보고서 열 형식

열별 요약 보고서의 데이터 및 열 구분 형식 옵션은 행별 요약 보고서에서 설명한 내용과 동일합니다.

열별 요약 보고서 구분 옵션

구분 옵션은 구분 범주에 대한 소계 출력, 간격, 쪽 매김 등을 제어합니다.

소계. 구분 범주에 대한 소계 출력을 제어합니다.

쪽 제어. 선택한 구분변수의 범주에 대해 간격이나 쪽 매김을 제어할 수 있습니다. 구분 범주 사이에 공백줄의 번호를 지정하거나 새 쪽에서 각 구분 범주로 시작할 수 있습니다.

소계 앞에 공백줄. 구분 범주 데이터와 소계 사이의 공백줄 수를 제어합니다.

열별 요약 보고서 옵션

옵션은 총계 출력이나 결측값 표시, 열 요약 보고서의 쪽 매김 등을 제어할 수 있습니다.

총계. 각 열에 대한 총계를 표시합니다. 열의 맨 아래에 나타납니다.

결측값. 보고서에서 결측값을 제외하거나 보고서에서 결측값을 단일 문자로 표시하도록 할 수 있습니다.

열별 요약을 위한 보고서 레이아웃

열별 요약 보고서를 위한 보고서 레이아웃은 행별 요약 보고서에서 설명한 내용과 동일합니다.

REPORT 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 단일 요약 행의 열에 여러 요약 함수를 표시합니다.
- 데이터 열 변수와 다른 변수 또는 요약 함수의 다양한 조합(복합 함수)을 위해 요약 행을 데이터 열에 삽입합니다.
- 중위수, 모드, 빈도 및 퍼센트를 요약 함수로 사용합니다.
- 요약 통계의 출력 형식을 보다 정밀하게 제어합니다.
- 보고서의 여러 지점에 빈 행을 삽입합니다.
- 목록 보고서의 *n*번째 케이스마다 빈행을 삽입합니다.

REPORT 구문의 복잡성으로 인해 구문을 사용하여 대화 상자에서 생성된 보고서를 접근시키고 해당 구문을 복사하고 붙여 넣으며 해당 구문을 세분화하여 원하는 정확한 보고서를 생성하기 위한 새로운 보고서를 작성할 때 유효하다는 것을 발견할 수 있습니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

신뢰성 분석

신뢰성 분석을 사용하여 측정 척도의 특성과 척도를 구성하는 항목을 연구할 수 있습니다. 신뢰성 분석 프로시저에서는 공통적으로 사용된 척도 신뢰도의 측도 수를 계산하고 척도의 개별 항목 간 관계에 대한 정보를 제공합니다. 급내 상관계수를 사용하여 급내 신뢰도 추정값을 계산할 수 있습니다.

신뢰도 분석은 또한 평가자 간 일치도를 평가하여 다양한 평가자 간 신뢰도를 판별하는 Fleiss의 다중 평가자 카파 통계도 제공합니다. 일치도가 높을수록 실제 환경을 반영하는 등급의 신뢰도가 높아집니다. Fleiss의 다중 평가자 카파 옵션은 221 페이지의 『신뢰도 분석: 통계량』 대화 상자에서 사용할 수 있습니다.

예

현재 질문지의 측도 방법이 고객이 만족할 만큼 유용합니까? 신뢰성 분석을 사용하면 질문지의 항목들이 서로 관련되어 있는 정도를 판별할 수 있으며 척도의 전반적 반복도나 내부 일관성에 대한 지수를 구할 수 있습니다. 또한 척도에서 제외해야 하는 문제 항목을 구분할 수 있습니다.

통계

각 변수와 척도에 대한 기술통계, 여러 항목의 요약 통계, 항목간 상관관계 및 공분산, 신뢰도 추정값, 분산분석표, 급내 상관계수, Hotelling의 T^2, Tukey의 가법성 검정 및 Fleiss의 다중 평가자 카파

모형

다음의 신뢰도 모형을 사용할 수 있습니다.

알파(Cronbach)

이 모형은 평균 항목간 상관관계를 기준으로 하는 내적일치(internal consistency) 측도입니다.

오메가(McDonald)

이 모형은 로컬 항목 종속성이 없는 단일 요인을 오차 공분산 형식으로 포함하는 일차원적 모형이라고 가정합니다. 이 모형은 서로 다른 두 항목의 공분산이 적재량의 곱임을 의미합니다.

반분
이 모형에서는 해당 척도를 두 부분으로 분리하고 부분간 상관을 검사합니다.

Guttman
이 모형에서는 실제 신뢰도에 대한 Guttman의 하한값을 계산합니다.

평형
이 모형에서는 여러 복제의 모든 항목에는 등분산과 동일한 오차 분산이 있는 것으로 가정합니다.

절대 평형
이 모형에서는 동형 모형을 가정하며 여러 항목에서 평균은 동일한 것으로 가정합니다.

신뢰도 분석 데이터 고려사항

데이터
데이터는 이분형, 순서 또는 구간이 될 수 있으나 숫자로 코딩되어야 합니다.

가정
관측값은 독립적이어야 하며 항목 간의 오차는 서로 상관되지 않아야 합니다. 각 대응항목은 이변량 정규 분포를 나타내야 합니다. 척도는 가법적이어야 하므로 각 항목은 전체 스코어와 선형으로 관련되어야 합니다. Fleiss의 다중 평가자 카파 통계량에는 다음과 같은 가정이 적용됩니다.

- 신뢰도 통계량을 실행하려면 항목 변수를 두 개 이상 지정해야 합니다.
- 등급 변수를 두 개 이상 선택하면 Fleiss의 다중 평가자 카파 명령문이 붙여넣어집니다.
- 평가자가 서로 연결되어 있지 않습니다.
- 평가자 수는 상수입니다.
- 각 개체는 단일 평가자만 포함하는 동일한 집단에서 평가됩니다.
- 다양한 불일치에 가중값을 지정할 수 없습니다.

관련 프로시저
척도 항목들의 차원을 탐색하려면(항목 스코어의 패턴을 설명하는 데 둘 이상의 구성이 필요한지 여부를 알기 위해) 요인 분석이나 다차원척도법을 사용합니다. 변수의 동일그룹군을 구분하려면 계층적 군집 분석을 사용하여 변수를 모으십시오.

신뢰도 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **축척** > **신뢰성 분석 ...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 둘 이상의 변수를 가법 척도의 잠재적 성분으로 선택합니다.
3. **모형** 드롭 다운 목록에서 모형을 선택합니다.
4. 선택적으로 척도 항목 또는 평가자 간 일치도를 설명하는 다양한 통계를 선택하려면 **통계량**을 클릭합니다.

신뢰도 분석: 통계량

다양한 평가자 간 신뢰도를 판별하기 위해 척도, 항목, 평가자 간 일치도를 설명하는 다양한 통계를 선택할 수 있습니다. 기본적으로 보고되는 통계에는 케이스의 수, 문항의 수가 있으며 다음과 같은 신뢰도 추정값도 포함됩니다.

알파 모형
알파계수; 이분형 데이터의 경우 이 모형은 Kuder-Richardson 20(KR20) 계수와 동일합니다.

Omega 모형
신뢰도분석을 평가하기 위한 McDonald의 오메가 추정입니다.

반분 모형
형식 간 상관, Guttman 반분 신뢰도, Spearman-Brown 신뢰도(같은 길이와 같지 않은 길이), 각 절반에 대한 알파계수입니다.

Guttman 모형
람다 1에서 람다 6까지의 신뢰도 계수입니다.

평형 및 절대 평형 모형
모형 적합도에 대한 검정, 오차 분산의 추정값, 공통 분산, 실제 분산, 추정된 항목 간 공통 상관, 추정된 신뢰도, 신뢰도에 대한 비편향 추정값입니다.

다음에 대한 기술통계량
여러 케이스의 척도나 항목에 대한 기술통계를 생성합니다.

항목
여러 케이스의 항목에 대한 기술통계량을 생성합니다.

척도
척도에 대한 기술통계량을 생성합니다.

항목제거시 척도
다른 항목으로 구성된 척도에 각 항목을 비교하여 요약 통계를 보여 줍니다. 통계에는 척도에서 항목이 삭제된 경우의 척도 평균 및 분산, 삭제된 항목과 다른 항목으로 구성된 척도 간 상관 및 항목이 삭제된 경우의 Cronbach 알파가 포함됩니다.

요약
척도내 모든 항목에서의 항목 분포의 기술통계를 제공합니다.

평균*(Means)*
항목 평균에 대한 요약 통계입니다. 최소, 최대 및 평균 항목 평균, 항목 평균의 범위 및 분산, 최소 항목 평균에 대한 최대 항목 평균의 비율이 표시됩니다.

분산
항목 분산에 대한 요약 통계입니다. 최소, 최대 및 평균 항목 분산, 항목 분산의 범위 및 분산, 최소 항목 분산에 대한 최대 항목 분산의 비율이 표시됩니다.

상관관계
항목 간 상관에 대한 요약 통계입니다. 최소, 최대, 평균 항목 간 상관관계, 항목 간 상관관계의 범위 및 분산, 최소 항목 간 상관관계에 대한 최대 항목 간 상관의 비율 등이 표시됩니다.

공분산*(Covariances)*
항목 간 공분산에 대한 요약 통계입니다. 최소, 최대, 평균 항목 간 공분산, 항목 간 공분산의 범위 및 분산, 최소 항목 간 공분산에 대한 최대 항목 간 공분산의 비율이 표시됩니다.

항목간
항목 간 상관행렬이나 공분산행렬을 생성합니다.

분산분석표
동일한 평균에 대한 검정을 생성합니다.

F 검정(F Test)
반복 측도 분산 분석표를 출력합니다.

Friedman 카이제곱(Friedman Chi-Square)
Friedman의 카이제곱 및 Kendall의 일치계수를 표시합니다. 이 옵션은 순위 형식의 데이터에 적합합니다. 카이제곱 검정은 분산 분석표의 일반적인 F 검정을 대체합니다.

Cochran 카이제곱(Cochran Chi-Square)
Cochran의 Q를 표시합니다. 이 옵션은 이분형 데이터에 적합합니다. Q 통계는 분산 분석표에서 일반적인 F 통계를 대체합니다.

평가자 간 일치도: Fleiss의 카파
평가자 간 일치도를 평가하여 다양한 평가자 간 신뢰도를 판별합니다. 일치도가 높을수록 실제 환경을 반영하는 등급의 신뢰도가 높아집니다. 일반화된 가중되지 않은 카파 통계는 다음과 같은 가정 하에서 일치도를 측정합니다.

- 신뢰도 통계량을 실행하려면 항목 변수를 두 개 이상 지정해야 합니다.
- 등급 변수를 두 개 이상 지정해야 합니다.

- 항목으로 선택한 변수도 등급으로 선택할 수 있습니다.
- 평가자가 서로 연결되어 있지 않습니다.
- 평가자 수는 상수입니다.
- 각 개체는 단일 평가자만 포함하는 동일한 집단에서 평가됩니다.
- 다양한 불일치에 가중값을 지정할 수 없습니다.

개별 범주에 대한 합치도 표시
개별 범주에 대한 합치도를 표시할지 여부를 지정합니다. 기본적으로 출력결과에 개별 범주에 대한 추정값이 표시되지 않습니다. 사용으로 설정할 경우 출력결과에 여러 표가 표시됩니다.

문자열 대소문자 무시
문자변수가 대소문자를 구분하는지 여부를 제어합니다. 기본적으로 문자 등급 값은 대소문자를 구분합니다.

문자열 범주 레이블이 대문자로 표시됨
출력결과 표의 범주 레이블이 대문자로 표시되는지 아니면 소문자로 표시되는지 여부를 제어합니다. 이 설정은 기본적으로 사용으로 설정되므로, 문자열 범주 레이블이 대문자로 표시됩니다.

근사 유의수준(%)
근사 신뢰구간의 유의수준을 지정합니다. 기본 설정은 95입니다.

결측값

사용자 결측값과 시스템 결측값 제외
사용자 결측값과 시스템 결측값의 제외를 제어합니다. 기본적으로 사용자 결측값과 시스템 결측값은 제외됩니다.

사용자 결측값을 유효로 처리함
사용으로 설정할 경우 사용자 결측값과 시스템 결측값을 유효한 데이터로 처리합니다. 이 설정은 기본적으로 사용 안 함으로 설정됩니다.

Hotelling의 T 제곱
척도의 모든 항목의 평균은 0이라는 귀무가설에 대해 다변량 검정을 생성합니다.

Tukey의 가법성 검정
항목 간에 승법 상호작용이 없다는 가정에 대한 검정을 생성합니다.

급내 상관계수
케이스 내에서 측도의 일치 또는 값의 일치를 생성합니다.

모델
급내 상관계수를 계산하는 모형을 선택합니다. 사용할 수 있는 모형은 이차원 혼합, 이차원 변량, 일차원 변량입니다. 사람 효과가 랜덤 효과이고 항목 효과가 고정 효과인 경우 **이차원 혼합** 모형을 선택하고, 사람 효과와 항목 효과가 모두 랜덤 효과인 경우 **이차원 변량** 모형을 선택하거나 사람 효과가 랜덤 효과인 경우 **일차원 변량**을 선택합니다.

유형
지수의 유형을 선택합니다. 사용할 수 있는 유형은 일치와 절대 동의입니다.

신뢰구간(%)
신뢰구간의 수준을 지정합니다. 기본값은 95%입니다.

검정값
가설 검정에 사용하는 가설의 계수값을 지정합니다. 이 값은 관측값의 비교 대상이 되는 값입니다. 기본값은 0입니다.

RELIABILITY 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 상관행렬을 읽고 분석합니다.
- 나중에 분석하기 위해 상관행렬을 씁니다.
- 반분 방법으로 동일반분 이외의 분할 방법을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

가중 카파

Cohen의 가중 카파는 교차 분류에서 관측된 평가자 간 일치 측도로 널리 사용되며, 등급이 순서 구조가 없는 명목 척도일 경우에 적합한 일치 지수입니다. 분할표의 일부 할당이 다른 할당보다 큰 중대성을 갖는다는 점에 자극을 받아 Cohen의 가중 카파가 개발되었습니다. 통계는 일치 또는 불일치를 반영하는 사전 정의된 셀 가중값에 의존합니다.

가중 카파 프로시저는 동일한 범주에 있는 두 순서 개체의 일치도를 측정하는 카파 통계의 중요한 일반화인 Cohen의 가중 카파를 추정하는 옵션을 제공합니다.

참고: 가중 카파 프로시저는 이전에 `STATS WEIGHTED KAPPA.spe` 확장이 제공하던 기능을 대체합니다.

예

평가자 간의 차이를 동등하게 중요한 것으로 처리하면 안되는 상황이 있습니다. 여러 사람이 연구 또는 임상 데이터를 수집하는 의료 산업이 그 예입니다. 이 경우, 데이터를 수집하는 사람들 사이의 변동성을 고려할 때 데이터의 신뢰성에 의문이 생길 수 있습니다.

통계

Cohen의 가중 카파, 선형 척도, 2차 척도, 근사 신뢰구간

가중 카파 데이터 고려사항

데이터

Cohen의 가중 카파 통계를 추정하려면 활성 데이터 세트를 기반으로 하는 이원배치표가 필요합니다.

등급 변수는 같은 유형(모두 문자열 또는 숫자)이어야 합니다.

Cohen의 가중 카파 추정은 표의 행과 열로 표시되는 두 등급 변수의 범주가 적절하게 정렬된 경우에만 적용됩니다(숫자변수 쌍의 경우 숫자 순서가 적용됨, 문자변수 쌍의 경우 문자순으로 적용됩니다).

가정

혼합 변수 쌍을 선택할 경우 Cohen의 가중 카파가 추정되지 않습니다.

등급 변수는 동일한 범주 집합을 공유하는 것으로 가정합니다.

가중 카파 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **축척** > **가중 카파...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. **대응별 평가자**로 지정할 문자변수 또는 숫자변수를 두 개 이상 선택합니다.

   참고: 모두 문자변수를 선택하거나 모두 숫자변수를 선택해야 합니다.
3. 선택적으로 대응별 평가자 또는 행/열 평가자의 표시를 제어하려면 **행 및 열 평가자 지정** 설정을 사용으로 설정합니다.
   - 사용으로 설정할 경우, 대응별 평가자가 표시되지 않고 행/열 평가자가 표시됩니다. **행 평가자** 및 **열 평가자** 필드를 제공하기 위해 사용자 인터페이스가 업데이트됩니다. 결과적으로 **대응별 평가자** 필드가 대체됩니다.
   - 사용 안 함으로 설정할 경우, 행/열 평가자가 표시되지 않고 대응별 평가자가 표시됩니다(기본 설정).

   행 및 열 평가자 지정을 사용으로 설정한 경우, **행 평가자** 및 **열 평가자** 둘 다에 대한 변수를 한 개 이상 지정합니다.

   참고: 행 평가자 및 **열 평가자**에 변수가 한 개만 포함되어 있으면 선택한 변수가 둘 다에 대해 같을 수 없습니다.
4. 선택적으로 가중 척도와 결측값 설정을 지정하려면 **기준**을 클릭하고 표시 형식과 교차표 설정을 지정하려면 **인쇄**를 지정합니다.

가중 카파: 기준

기준 대화 상자는 Cohen의 가중 카파 통계의 추정을 지정하는 옵션을 제공합니다.

가중 척도
일치에 대한 선형 또는 2차 가중값을 지정하는 옵션을 제공합니다. 기본 설정은 선형 (Cicchetti-Allison이라고도 함) 가중치를 사용하는 것입니다.

결측값
대응별 기준에서 결측값이 있는 케이스를 제거하고 사용자 결측값을 유효한 데이터로 처리하는 옵션을 제공합니다.

문자 등급 변수가 대소문자를 구분함
선택할 경우 문자변수가 대소문자 구분으로 처리됩니다.

근사 신뢰구간(%)
이 선택적 설정은 근사 신뢰구간 추정을 위한 신뢰수준을 지정합니다. 0과 100 사이에 있는 단일 이중 값이어야 하며, 기본 설정은 95입니다.

가중 카파: 인쇄

인쇄 대화 상자는 교차표를 제어하는 옵션을 제공합니다.

표시 및 형식
교차표 표시 및 형식을 제어하는 옵션을 제공합니다.

등급 범주가 오름차순으로 표시됨
선택할 경우 교차표의 등급 범주가 오름차순으로 표시됩니다. 이 설정을 선택하지 않을 경우 교차표의 등급 범주가 내림차순으로 표시됩니다. 이 설정은 기본적으로 사용으로 설정됩니다.

문자열 범주 레이블이 대문자로 표시됨
선택할 경우 교차표가 대문자로 표시됩니다. 이 설정을 선택하지 않을 경우 교차표가 소문자로 표시됩니다. 이 설정은 기본적으로 사용으로 설정됩니다.

교차 분석표
교차표에 사용되는 등급 변수를 지정하는 옵션을 제공합니다. 기본적으로 교차표 설정은 사용으로 설정되지 않으므로, 등급 변수의 교차표가 표시되지 않습니다.

등급 변수 교차표 표시
선택할 경우 이 설정으로 모든 또는 사용자 지정 등급 변수의 교차표를 사용할 수 있습니다.

모든 등급 변수 포함
선택할 경우 정의된 모든 등급 변수 쌍에 대해 교차표가 인쇄됩니다.

사용자 정의 등급 변수 포함
선택할 경우 **사용 가능한 변수**, **행 평가자** 및 **열 평가자** 필드를 사용하여 교차표에 포함될 등급 변수를 선택하십시오.

다차원척도법

다차원척도법은 오브젝트간 또는 케이스간 거리 측도 세트에서 구조를 찾으려고 합니다. 이 작업은 공간의 점간 거리가 제공된 상이성에 가능한 한 가까워지도록 관측값을 개념적 공간(보통 2차원이나 3차원)의 특정 위치로 할당함으로써 수행됩니다. 많은 경우 이 개념적 공간의 차원들은 데이터를 좀더 잘 이해할 수 있도록 해석되고 사용될 수 있습니다.

변수를 객관적으로 측정한 경우 다차원척도법을 데이터 축소 기법으로 사용할 수 있습니다. 필요한 경우 다차원척도법 프로시저는 다변량 데이터로부터의 거리를 계산합니다. 다차원척도법은 오브젝트간 또는 개념간 상이성에 대한 주관적 등급에 적용될 수도 있습니다. 또한 다차원척도법 프로시저는 여러 평가자나 질문지 반응자가 있는 경우처럼 다중 소스로부터의 상이성 데이터를 처리할 수 있습니다.

예. 다른 자동차 간의 관계를 사람들은 어떻게 인식합니까? 다른 자동차 제조업체와 모형 간의 유사성 등급을 나타내는 반응자의 데이터가 있는 경우 소비자의 인식을 설명하는 차원을 구분하는데 다차원척도법을 사용할 수 있습니다. 예를 들어, 한 차량의 가격과 크기가 2차원 공간을 정의하며 이 결과에 반응자들이 보고한 유사성이 나타남을 알 수 있습니다.

통계. 각 모형의 경우 데이터 행렬, 최적으로 측정된 데이터 행렬, S-스트레스(Young의 스트레스), 스트레스(Kruskal의 스트레스), RSQ, 자극 좌표, 각 자극에 대한 평균 스트레스와 RSQ(RMDS 모형)를 선택할 수 있습니다. 개별 차이(INDSCAL) 모형의 경우 각 개체에 대한 개체 가중치와 불운지수를 선택할 수 있습니다. 반복된 다차원척도법 모형의 각 행렬의 경우 각 자극에 대한 스트레스와 RSQ를 선택할 수 있습니다. 도표의 경우 자극 좌표(2차원이나 3차원), 상이성-대-거리의 산점도를 선택할 수 있습니다.

다차원척도법 데이터 고려 사항

데이터. 데이터가 상이성 데이터인 경우 모든 상이성은 양적이며 동일한 메트릭으로 측정되어야 합니다. 데이터가 다변량 데이터인 경우 변수는 양적, 이분형 또는 개수 데이터일 수 있습니다. 척도가 달라지면 해법이 달라지므로 변수 척도법은 중요한 문제가 됩니다. 변수 척도에 큰 차이가 있으면(예: 한 변수는 달러로 측정되고 다른 변수는 연도로 측정될 경우) 척도를 표준화해야 합니다. 이 과정은 다차원척도법 프로시저를 통해 자동으로 수행할 수 있습니다.

가정. 다차원척도법 프로시저는 분포성 가정에 상대적으로 적용되지 않습니다. 결과가 정확히 계산되도록 다차원척도법 옵션 대화 상자에서 적절한 측정 수준(순서척도, 구간척도, 비율척도)을 선택합니다.

관련 프로시저. 데이터 축소를 하려면 요인 분석을 사용합니다. 이것은 변수가 양적인 경우 더욱 바람직합니다. 유사한 케이스를 가진 그룹을 식별하려면 계층적 군집 분석이나 *K*-평균 군집 분석을 다차원척도법 분석과 함께 사용합니다.

다차원척도법 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 척도 > 다차원척도법...
2. 분석할 숫자변수를 4개 이상 선택합니다.
3. 거리 그룹에서 **데이터 자체가 거리 행렬**이나 **데이터로부터 거리 행렬 계산하기**를 선택합니다.
4. **데이터로부터 거리 행렬 계산하기**를 선택하면 개별 행렬의 그룹화 변수도 선택할 수 있습니다. 그룹화 변수는 숫자 또는 문자열일 수 있습니다.

선택적으로 다음을 수행할 수 있습니다.

- 데이터 자체가 거리 행렬일 경우 거리 행렬의 형태를 지정합니다.
- 데이터로부터 거리 행렬을 계산할 경우 사용할 거리 측도를 지정합니다.

근접 매핑을 이용한 다차원 척도화

ALSCAL 외에도, 다차원 척도법과 관련된 접근법으로 IBM SPSS 통계 **근접 매핑(PROXMAP)** 을 제공합니다. ALSCAL과 PROXMAP은 모델 적합도를 계산하는 방식이 다릅니다. ALSCAL은 제곱된 거리를 제곱된 비유사도에 맞추는 반면, PROXMAP은 거리를 비유사도에 직접 맞춥니다. 결과적으로, 두 절차는 결과적인 공간적 구성이 비슷해 보일지라도 서로 다른 수치적 결과를 산출할 수 있다. 이러한 차이는 거의 완벽한 해법이 존재하지 않는 데이터를 다룰 때 더욱 두드러질 수 있습니다.

ALSCAL은 제한된 텍스트 기반 출력과 최소한의 진단 정보만을 제공하는 기본적인 다차원 척도 분석을 수행하는 반면, PROXMAP은 비교적 정확한 시각화 자료(셰퍼드 다이어그램, 변환 플롯, 잔차 분석 등), 향상된 통계 출력(다양한 스트레스 측정값, R^2 값, 상세한 적합도 지수 등) 및 속성 적합과 같은 고급 기능을 갖춘 현대적이고 포괄적인 솔루션을 제공합니다.

근접 매핑을 시작하려면 다음 링크를 참조하세요:

- 228 페이지의 『ALSCAL과 PROXMAP 비교』
- 근접성 맵핑
-
-
-

다차원척도법 데이터의 형태

활성 데이터 세트가 한 오브젝트 세트간 거리나 두 오브젝트 세트간 거리를 나타낼 경우 정확한 결과를 얻기 위해서는 데이터 행렬의 형태를 지정해야 합니다.

참고: 모형 대화 상자에서 행 조건부를 지정하는 경우에는 **정방대칭형**을 선택할 수 없습니다.

다차원척도법 측도 계산

다차원척도법은 상이성 데이터를 사용하여 척도화 해법을 작성합니다. 데이터가 다변량 데이터(측정된 변수의 여러 값)인 경우 다차원척도화 해법을 계산하려면 상이성 데이터를 작성해야 합니다. 데이터로부터 상이성 측도를 작성하기 위한 세부사항을 지정할 수 있습니다.

측도. 분석에 필요한 상이성 측도를 지정할 수 있습니다. 해당 데이터 유형에 해당하는 측도에서 하나를 선택한 후 측도 유형을 나타내는 드롭다운 목록에서 하나를 선택합니다. 사용 가능한 선택 사항은 다음과 같습니다.

- **구간**. 유클리드 거리, 유클리드 거리 제곱값, Chebychev, 블록, Minkowski 또는 사용자 정의 간격을 선택할 수 있습니다.
- **개수**. 카이제곱 측도 또는 파이제곱 측도를 선택할 수 있습니다.
- **이분형**. 유클리드 거리, 유클리드 거리 제곱값, 크기 차이, 패턴 차이, 분산, Lance와 Williams를 선택할 수 있습니다.

거리 행렬 만들기. 분석 단위를 선택할 수 있습니다. 변수 사이의 거리 또는 케이스들 사이의 거리를 선택합니다.

값 변환. 특정 경우에(예: 변수가 다른 척도로 측도될 경우) 근접성(이분형 데이터에는 적용할 수 없음)을 계산하기 전에 값을 표준화해야 할 경우가 있습니다. 표준화 드롭다운 목록에서 표준화 방법을 선택합니다. 표준화할 필요가 없으면 **없음**을 선택합니다.

다차원척도법: 모형

다차원척도법 모형의 정확한 추정은 데이터의 특성과 모형 그 자체에 따라 달라집니다.

측정 수준. 데이터의 수준을 지정할 수 있습니다. 선택 사항으로는 순서, 구간척도 또는 비율척도가 있습니다. 변수가 순서인 경우 **동률 관측값 비동률화**를 선택하면 해당 변수는 연속변수로 취급되므로 동률값(다른 케이스에 대한 동일한 값)은 최적 상태로 무효화됩니다.

조건부. 어떤 비교 방법이 의미 있는지 지정할 수 있습니다. 선택 사항으로는 행렬, 행, 조건없음 등이 있습니다.

차원. 척도화 해법의 차원을 지정할 수 있습니다. 하나의 해법이 범위 내의 각 수에 대해 계산됩니다. 1과 6 사이의 정수를 지정해야 하며 척도화 모형으로 **유클리드 거리**를 선택한 경우에만 최소값 1이 허용됩니다. 단일 해법의 경우 최소값과 최대값에 대해 서로 같은 수를 지정합니다.

척도화 모형. 수행되는 척도를 기준으로 가정을 지정할 수 있습니다. 사용 가능한 선택 사항에는 유클리드 거리나 개인차 유클리드 거리(INDSCAL이라고도 함)가 있습니다. 개인차 유클리드 거리 모형의 경우 데이터에 적절하다면 **음의 개체 가중치 허용**을 선택할 수 있습니다.

다차원척도법: 옵션

다차원척도법 분석을 위해 다음 옵션을 지정할 수 있습니다.

표시. 다양한 유형의 결과를 선택할 수 있습니다. 사용 가능한 옵션에는 그룹 도표, 개별 대상도표, 데이터 행렬, 모형 및 옵션 요약이 있습니다.

기준. 반복을 중지할 시기를 결정할 수 있습니다. 기본값을 변경하려면 **S-스트레스 수렴**, **최소 S-스트레스 값**, **최대반복수**에 대한 값을 입력합니다.

다음 값 이하의 거리는 결측값으로 처리. 이 값보다 작은 거리는 분석에서 제외됩니다.

ALSCAL 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 다차원척도법에 대한 문헌에서 ASCAL, AINDS 및 GEMSCAL로 알려진 세 가지 추가 모형 유형을 사용합니다.
- 구간 및 비율 데이터로 다항 변환을 수행합니다.
- 순서 데이터를 사용하여 유사성(거리가 아님)을 분석합니다.
- 명목 데이터를 분석합니다.
- 다양한 좌표와 가중치 행렬을 파일에 저장하고 분석을 위해 다시 읽습니다.

• 다차원 확장법을 제한합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

ALSCAL과 PROXMAP 비교

ALSCAL과 PROXMAP 모두 다차원 척도법(MDS)을 수행하여 근접성 데이터를 지정된 차원으로 표현합니다. PROXMAP에서는 이러한 결과를 최신 용어와 진단 플롯을 사용하여 제시하며, 이를 통해 ALSCAL 출력 결과와 직접 비교할 수 있습니다.

ALSCAL과 PROXMAP은 근접성 데이터를 바탕으로 물체의 공간적 배치를 추정합니다. PROXMAP은 최신 용어와 통합된 진단 플롯을 사용하여 ALSCAL과 동일한 개념적 결과를 지원하는 대화형 프레임워크를 제공합니다.

개념도

다음 표는 ALSCAL의 주요 용어와 산출물이 PROXMAP 용어 체계와 어떻게 대응되는지를 보여줍니다. 결과에 담긴 분석적 의미는 어떤 절차를 적용하더라도 동일합니다.

표 *5. ALSCAL* 및 *PROXMAP* 용어 매핑

ALSCAL 용어 또는 출력	**PROXMAP와 동등한**
유도된 자극의 위치	공통 오브젝트 공간
선형적합 산점도 상이성 거리	**핏 소스 <소스 라벨>** 변환된 근접도 거리
비선형적합 산점도 관측값 거리	**Shepard 도표[1]출처 <출처 라벨>** 원래 근접성 거리 + 변형된 근접성
변환 산점도 관측값 상이성	**변환 소스 <소스 레이블>** 원래 근접성 변환된 근접도

ALSCAL에서 PROXMAP으로의 전환

다음 튜토리얼에서는 PROXMAP을 사용하여 ALSCAL 다차원 척도법(MDS) 분석을 재현하는 방법을 보여주고, 분석 결과에서 나타나는 유사점과 차이점을 해석하는 방법을 설명합니다.

개요

이 튜토리얼에서는 PROXMAP을 설정하여 ALSCAL 분석 결과를 재현하는 방법과, 그 결과로 생성된 설정들을 비교하는 방법을 보여줍니다.

이 예제는 단일 근접성 데이터셋에 초점을 맞추고 있으며, 기본 설정 또는 이에 상응하는 설정을 사용하므로 ALSCAL과 PROXMAP 간의 차이는 결과의 실질적인 변화보다는 크기 조정과 방향에 국한됩니다.

분석에 사용된 데이터

이 분석은 미국 내 12개 도시 간의 비행 거리를 나타내는 근접성 데이터셋인 Dist 기반으로 합니다. 데이터는 하삼각 행렬 형태의 대칭 근접 행렬로 저장됩니다.

[1] 셰퍼드 플롯(Shepard Plot)은 비측도 다차원 척도법(nonmetric Multidimensional Scaling)의 창시자 중 한 명인 로저 셰퍼드(Roger Shepard)를 가리킨다. 이 기능은 ALSCAL의 "비선형 적합" 플롯과 "변환" 플롯을 결합한 것입니다.

데이터 뷰에서 데이터셋은 하삼각 행렬 형식으로 저장된 대칭 근접 행렬(12x12)과 구성에 레이블을 지정하는 데 사용되는 명목형 객체 식별자 변수를 나타냅니다.

변수 보기에서 데이터셋은 근접성 행렬을 포함하는 12개의 척도형 변수와 객체 레이블을 정의하는 1개의 명목형 변수로 구성됩니다.

CITY 변수가 객체 레이블 변수로 지정됩니다.

근접 행렬 외에도, 이 데이터셋에는 구성 내 객체들에 대한 레이블을 제공하는 명목형 객체 식별자 변수(*CITY*)가 포함되어 있습니다.

C ITY의 값은 1에서 12 사이입니다. 이러한 정수에는 값 라벨이 할당되며, 여기서:

- 1 = 애틀랜타
- 2 = 보스턴
- 3 = 신시내티
- 4 = 콜럼버스
- 5 = 댈러스
- 6 = INDNPLIS
- 7 = 리트록
- 8 = 로스앤젤레스
- 9 = 멤피스
- 10 = 세인트루이스
- 11 = 스포캔
- 12 = 탬파

PROXMAP은 이러한 값 레이블을 사용하여 테이블과 그래픽 출력에서 객체를 식별함으로써, 숫자 코드 대신 객체 이름이 표시되도록 합니다.

PROXMAP 분석 설정

‘ > **An** > **alyzeMappingProximity**’ **매핑** 으로 이동

근접 행렬 변수는 ‘근접 매핑’ 대화 상자에서 단일 근접 소스로 지정됩니다. PROXMAP 함수는 대각선을 포함한 하삼각 행렬 형식이나 완전 대칭 행렬 형식으로 저장된 근접성 데이터를 모두 허용하며, 단 행렬이 올바르게 지정된 경우에 한합니다.

데이터 탭 (근접 행렬 및 객체 레이블 지정)

근접 소스를 정의하십시오:

1. 변수 목록에서 *ATLANTA*부터 *TAMPA* 까지의 변수를 선택하고, 이를 **‘소스’의 ‘변수** ’로 이동하십시오.
2. **‘소스 레이블**’ 필드에 이름을 입력하세요(예: 도시 간 비행 거리).
3. **‘소스 추가’를** 클릭하여 소스를 생성합니다(예: *$SRC001*).
4. **‘데이터 유형** ’에서 **‘근접 데이터’를** 선택합니다.
5. *CITY* 변수를 **‘객체** 레이블’로 이동

‘현재 객체 수 = 12’는 근접성 행렬에 의해 정의된 객체의 수를 나타내며, 이 구성에 12개 도시가 포함될 것임을 확인해 줍니다.

모델 탭 (PROXMAP 모델 지정)

이 분석은 단일 근접 소스($SRC001)를 기반으로 하므로, 다음 옵션을 선택하십시오:

1. **‘3-Way** 모델’에서 **‘Identity**’ 옵션을 선택하십시오.
2. **‘조건부/변환** ’에서 **‘무조건’을** 선택하세요
3. **‘변환 정의’를** 클릭한 다음 **‘서수**’ 및 **‘동점 허용**’ 옵션을 선택합니다.

   이는 근접도가 서수적 차이로서 취급됨을 의미합니다. 근접도의 순위는 유지되며, 데이터 내의 동일한 값들은 단조 변환 후에도 여전히 동등한 순위를 유지합니다.

필요한 경우 변환 수준(예: 선형, 지수, 스플라인 등)을 변경할 수 있습니다.

현재 차원 수 = 2는 2차원 구성을 추정할 것임을 나타냅니다.

이 설정은 ALSCAL과의 비교에 사용된 사양과 일치합니다.

'기준' 탭 (차원, 수렴성 및 표준화 지정)

1. **차원 수** : 차원 수는 *2* 로 설정되어 있습니다

   이는 2차원 구성을 추정할 것임을 명시합니다.
2. **알고리즘의 수렴** : 수렴 기준은 반복 최적화 과정을 제어한다. 다음과 같은 설정이 지정되어 있습니다:
   - **함수 minimum:** 0.000001
   - **기능 차이:** 0.000000000001
   - **구성 차이:** 0.000001
   - **최대 반복 횟수:** 1000

   이러한 값들은 반복 알고리즘이 언제 중단되는지를 결정합니다. 수렴은 손실 함수의 개선 폭과 구성의 변화가 지정된 허용 오차 범위 아래로 떨어지거나, 최대 반복 횟수에 도달했을 때 이루어집니다.
3. **표준화** : **제곱합을 N으로** 설정합니다. 이 옵션은 다변량 데이터의 표준화에 특히 유용합니다.
4. **차원** : **차원별로** 선택. 이 옵션을 사용하면 개별 치수를 반영할 수 있습니다. '차원 2'의 **'반영'** 아래에 있는 확인란을 선택하십시오.

   반사는 축의 방향(부호)에만 영향을 미치며, 거리나 적합도는 변하지 않습니다. 이는 일반적으로 해를 기준 구성과 일치시키기 위해 사용됩니다(예를 들어, ALSCAL 해와 비교할 때).

'출력' 탭: '모델' 하위 탭 (표 및 그래프 요청)

1. '테이블'에서 **'좌표'를** 선택하면, 출력에 공통 좌표계의 객체 좌표가 포함된 테이블이 생성됩니다.

   이 분석은 단일 근접 소스를 기반으로 하며 추가 소스나 다변량 변수가 포함되지 않으므로, **'변수 방향 (Variable Directions)'** 및 **'공간 가중치(Space Weights)'** 옵션은 선택되지 않았습니다.
2. '그래프' 아래의 '공용 공간' 섹션에서 **'구성'을** 선택합니다.

   이렇게 하면 소스 $SRC001 (예: 도시 간 비행 거리)에 대한 객체 구성의 그래프가 생성되며, **'데이터'** 탭에서 지정한 변수(예: *CITY*)의 레이블이 표시됩니다.
3. **추가 작업** : **'소스(Sources)** ' 목록에서 ' *$SRC001* '를 **'변환** (Transformation)', **'적합(Fit**)', **'셰퍼드 (Shepard)** '로 이동합니다.

   다음 그래프가 생성됩니다:
   - 수직축에는 **변환된 근접도를**, 수평축에는 원래의 근접도를 표시한 그래프.
   - 구성 내 **추정 거리를 변환된 근접도와** 대비하여 보여주는 산점도.
   - 비측도 다차원 척도법(근접도 매핑)에서 적합도를 시각화하기 위해, 추정된 거리와 변환된 근접도를 원본 근접도와 함께 표시한 진단 산점도.

ALSCAL 분석에서 사용된 차원과 일치하는 2차원 해를 구해야 합니다.

명령 구문

또한, 및 을 ALSCAL 사용하여 분석에 활용되는 PROXMAP다음 명령어 구문을 참조하십시오.

ALSCAL을 이용한 분석에 사용되는 명령어 구문

```
ALSCAL
  VARIABLES=ATLANTA BOSTON CINCNATI COLUMBUS DALLAS INDNPLIS LITTROCK LOSANGEL MEMPHIS
STLOUIS
    SPOKANE TAMPA
  /SHAPE=SYMMETRIC
  /LEVEL=ORDINAL
  /CONDITION=MATRIX
  /MODEL=EUCLID
  /CRITERIA=CONVERGE(0.001) STRESSMIN(0.005) ITER(30) CUTOFF(0) DIMENS(2,2)
  /PLOT.
```

PROXMAP을 이용한 분석에 사용되는 명령어 구문

```
PROXMAP
  /SOURCEID NAME=$SRC001 LABEL='Cities flying distances'  DATA=ATLANTA BOSTON CINCNATI
COLUMBUS DALLAS INDNPLIS
    LITTROCK LOSANGEL MEMPHIS STLOUIS SPOKANE TAMPA
  /DATA PROXIMITIES
  /OBJECTID LABELS=CITY
  /CONDITION UNCONDITIONAL
  /TRANSFORMATIONS SOURCES=$SRC001 FUNCTION=ORDINAL APPROACH=KEEPTIES
  /MODEL IDENTITY
  /CRITERIA DIMENSIONS=2 MAXIMUM=1000 MINIMUM=0.000001 RELATIVE=0.000000001
ABSOLUTE=0.000001
    STANDARDIZATION='N' REFLECTION=DIMENSIONS NUMBERS=1 -1
  /PRINT DATA(RAW) MODEL (COORDINATES) CRITERIA (SUMMARY ITERATIONS)
  /PLOT MODEL (CONFIGURATION TRANSFORMATION ($SRC001) FIT ($SRC001) SHEPARD ($SRC001)).
```

출력 및 진단 그래프

출력 결과에는 공통 객체 공간 내 객체 좌표 표와 도시의 공간적 분포를 보여주는 구성도가 포함됩니다.

변환 플롯, 적합도 플롯, 셰퍼드 플롯 등을 포함한 추가 진단 플롯이 요청됩니다. 이 그래프들은 원래 근접도, 변환된 근접도, 그리고 추정된 거리 간의 관계를 보여주고, 해당 해법의 적합도에 대한 시각적 정보를 제공합니다.

PROXMAP의 용어는 ALSCAL의 용어와 다릅니다. 예를 들어, PROXMAP에서는 ALSCAL 자극을 '객체'라고 부르며, ALSCAL의 불일치를 '변환된 근접성'이라고 부릅니다. PROXMAP의 셰퍼드(Shepard) 플롯은 ALSCAL의 비선형 적합 및 변환 플롯에 표시된 정보를 결합한 것입니다.

결과 및 해석

PROXMAP을 ALSCAL과 유사한 설정으로 실행할 경우, 그 결과로 얻어지는 구성은 매우 유사합니다. 두 해법 간의 차이는 전체적인 배율과 축 대칭에 국한된다.

좌표 값은 수치상으로는 다를 수 있지만, 축척을 고려하면 동일한 기본 구성을 나타냅니다. PROXMAP은 근접 거리의 원래 비율을 기준으로 좌표를 보고합니다.

구성 플롯을 시각적으로 비교해 보면 객체들의 상대적 위치가 유지되는 것을 확인할 수 있으며, 이는 PROXMAP이 ALSCAL의 해법을 성공적으로 재현했음을 입증한다.

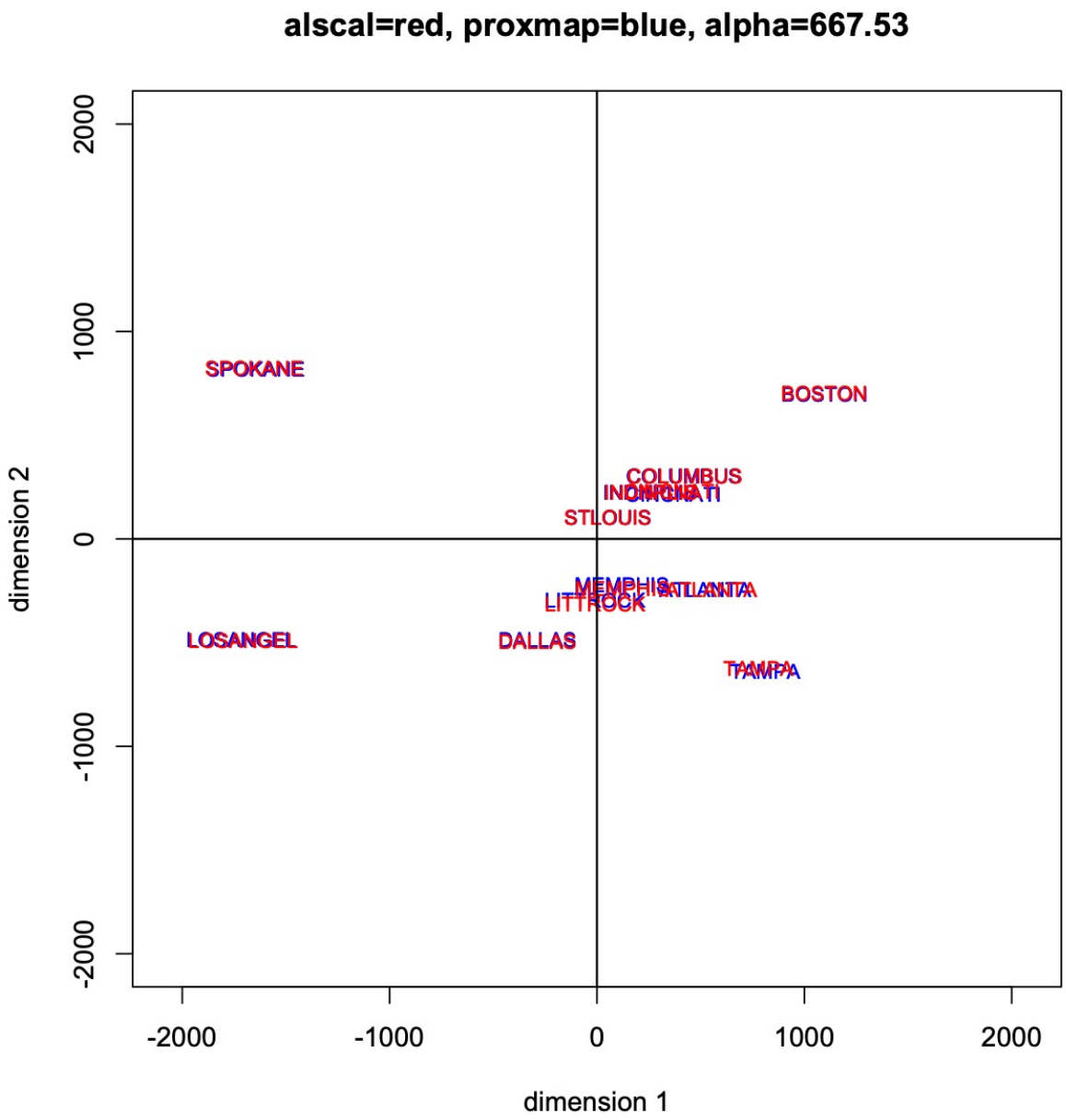


그림 *3. ALSCAL* 자극 구성 *PROXMAP* 좌표 공용 공간

ALSCAL 사용자를 위한 주요 내용

- ALSCAL에 대한 기존의 근접성 데이터는 하삼각 행렬 형태로 저장된 데이터를 포함하여 PROXMAP에서 재사용할 수 있습니다.
- PROXMAP 출력 결과와 그래프에서 객체를 식별하려면 추가적인 객체 라벨 변수가 필요합니다.
- PROXMAP 결과와 ALSCAL 출력을 비교할 때 좌표 방향에 차이가 발생할 수 있습니다. 반사는 축의 방향에만 영향을 미치며, 거리나 적합도는 변경하지 않습니다.
- PROXMAP은 ALSCAL과의 개념적 연속성을 유지하면서도 최신 용어와 현대적인 다차원 척도화 기법을 사용합니다.

비율 통계량

비율 통계 프로시저는 두 척도변수 간의 비율을 설명하기 위한 포괄적인 요약 통계 목록을 제공합니다.

그룹화 변수의 값을 기준으로 하여 오름차순 또는 내림차순으로 결과를 정렬할 수 있습니다. 비율 통계 보고서를 결과로 출력하지 않고 외부 파일에 저장할 수 있습니다.

예
다섯 지역에서 주택의 견적가와 판매가 간의 비율이 균일합니까? 결과를 통해 비율의 분포가 각 지역마다 매우 다르다는 것을 알 수 있을 것입니다.

통계
중위수, 평균, 가중 평균, 신뢰구간, 분산 계수(COD), 변동 계수(COV), 중위수 중심 변동 계수, 평균 중심 변동 계수, 가격 관련 편향(PRB), 가격 관련 차등(PRD), 표준 편차, 평균 절대 편차(AAD), 범위, 최소값과 최대값 및 사용자 지정 범위 또는 중위 비율 내의 백분율에 대해 계산된 농도 지수.

데이터 고려사항

데이터
숫자 코드나 문자를 사용하여 그룹화 변수(명목 수준이나 순위 수준 측정)를 코딩합니다.

가정
비율의 분자와 분모를 정의하는 변수는 양수값을 가지는 척도변수이어야 합니다.

비율 통계량 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > 기술통계량 > 비율...

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 분자변수를 선택합니다.
3. 분모변수를 선택합니다.

또는 다음 작업을 수행합니다.

- 그룹화 변수를 선택하고 결과에서 그룹의 순서를 지정합니다.
- 결과를 뷰어에 표시할지 여부를 선택합니다.
- **N** 키워드를 **PRINT** 부명령문에 추가하도록 선택합니다. 이 설정은 결과의 통계표에 표본 크기를 추가합니다.
- 나중에 다시 사용할 수 있도록 결과를 외부 파일에 저장할지 여부를 선택하고 결과를 저장할 파일의 이름을 지정합니다.

비율 통계량: 통계량

중심경향
중심경향 측도는 비율의 분포를 보여주는 통계량입니다.

중위수
이 값보다 작은 비율의 수와 이 값보다 큰 비율의 수가 같은 값입니다.

평균
비율을 합산하고 그 결과를 총 비율 수로 나눈 결과입니다.

가중 평균(W)
분자의 평균을 분모의 평균으로 나눈 결과입니다. 가중 평균은 분모에 의해 가중치를 부여한 비율의 평균이기도 합니다.

PRB에 대한 신뢰구간 및 중심경향 측도
PRB, 평균, 중위수 및 가중 평균에 대한 신뢰구간을 표시합니다(요청하는 경우). 신뢰수준으로 0 이상, 100 미만의 값을 지정합니다. 기본 설정은 95%입니다.

산포도
관측값의 변동 또는 산포 정도를 측정하는 통계량입니다.

평균절대편차
평균 절대 편차는 중위수에 대한 비율의 절대 편차를 합하여 비율의 전체 수로 나눈 결과입니다.

산포계수
산포계수는 평균 절대 편차를 중위수 퍼센트로 표시한 결과입니다.

CV
CV는 변동 계수를 나타냅니다. 평균에 대한 표준 편차의 비율입니다. 값은 백분율로 표시됩니다.

PRB
가격 관련 편향입니다. 가격이 높은 부동산에 대해 가격 대비 평가 비율이 체계적으로 높거나 낮은지 여부에 대한 지수입니다. 가치 대리 측도의 밑이 2인 로그에 대한 중위수 비율에서 평가 비율의 퍼센트 차이를 회귀분석하며, 판매 가격의 평균 및 중위수 비율 대비 평가 값의 비율로 계산됩니다. 값의 100% 변화량에 대한 평가 비율의 퍼센트 변화량을 제공합니다.

가격 관련 차(R)
가격 관련 차는 회귀 지수로도 알려져 있으며 평균을 가중 평균으로 나눈 결과입니다.

표준편차
표준 편차는 평균에 대한 비율의 제곱 편차를 합하여 비율의 전체 수에서 1을 뺀 수로 나누어 양의 제곱근을 구한 결과입니다.

범위
범위는 최대 비율에서 최소 비율을 뺀 결과입니다.

최소값(N)
최소값은 가장 작은 비율입니다.

최대
최대값은 가장 큰 비율입니다.

집중 지수
집중 계수는 구간 내 해당하는 비율 퍼센트를 측정합니다. 다음 두 가지 방법으로 계산할 수 있습니다.

비율간
여기에서 구간은 구간의 상한값과 하한값을 지정하여 정확하게 정의합니다. 비율 하한값 및 비율 상한값을 입력하고 **추가**를 클릭하여 구간을 정합니다.

중위수 퍼센트 내
여기에서 구간은 중위수 퍼센트를 지정하여 함축적으로 정의합니다. 0과 100 사이의 값을 입력하고 **추가**를 클릭합니다. 구간에서 낮은 값은 (1 − 0.01 × 값) × 중위수의 값과 같으며 높은 값은 (1 + 0.01 × 값) × 중위수의 값과 같습니다.

P-p 도표

P-P 도표 절차는 하나 이상의 순차 또는 시계열 변수의 확률 도표를 생성합니다. 변수들은 플롯팅 전에 표준화되고, 다르게 되고, 변환될 수 있습니다.

사용 가능한 검정 분포에는 베타, 카이제곱, 지수, 감마, 절반정규, 라플라스, 로지스틱, 로그정규, 정규, 파레토, 스튜던트 t, 와이블 및 균일이 있습니다. 선택한 분포에 따라 자유도와 기타 모수를 지정할 수 있습니다.

- 변환된 값에 대한 확률 도표를 작성합니다. 변환 옵션에는 자연 로그 변환, 표준화 값, 차이, 계절 차이가 있습니다.
- 예상 분포 계산 및 "동률" 구분 방법을 지정하거나 값이 같은 여러 관측값을 지정할 수 있습니다.

분포 검정
데이터의 분포 유형을 지정합니다. 드롭 다운 목록에서는 다음 옵션을 제공합니다.

베타
베타 분포. 형태1 및 형태2 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다. 모든 관측값은 0 - 1 범위(경계값 포함)에 있어야 합니다.

카이제곱
카이제곱 분포. 자유도(*df*)를 지정해야 합니다. 음수 관측값은 허용되지 않습니다.

지수
지수 분포. 척도 모수 *a*는 양수여야 합니다. 이 모수를 지정하지 않을 경우 DISTRIBUTION이 표본 평균에서 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

감마
감마 분포. 형태 및 척도 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

절반정규
절반정규 분포. 데이터는 위치가 없거나 중앙 집중화된 것으로 가정합니다 (위치 모수=0). 척도 모수 *a*를 지정하거나 DISTRIBUTION이 최대우도 방법을 사용하여 추정하도록 설정할 수 있습니다.

라플라스
라플라스 또는 이중 지수 분포. LAPLACE는 위치 및 척도 모수(*a* 및 *b)*를 사용합니다. 척도 모수*(b*)는 양수여야 합니다. 모수를 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다.

로지스틱
로지스틱 분포. LOGISTIC는 위치 및 척도 모수(*a* 및 *b)*를 사용합니다. 척도 모수*(b*)는 양수여야 합니다. 모수를 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다.

로그정규
로그정규 분포. 척도 및 형태 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 데이터 자연로그의 평균 및 표준편차에서 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

정규
정규 분포. 위치 모수 *a*는 숫자 값일 수 있지만, 척도 모수 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다.

파레토
파레토 분포. 임계값 및 형태 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 *a*는 최소 관측값과 같다고 가정하고 *b*는 최대우도 방법을 사용하여 추정합니다. 음수 관측값은 허용되지 않습니다.

스튜던트 T
스튜던트 T 분포. 자유도(*df)*를 지정해야 합니다.

균일
균일 분포. UNIFORM은 최소값 및 최대값 모수(*a* 및 *b)*를 사용합니다. *a* 매개변수는 *b* 이상이어야 합니다. 매개변수가 지정되지 않은 경우, DISTRIBUTION은 샘플 데이터에서 이를 가정합니다.

Weibull
Weibull 분포. 척도 및 형태 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 최소 제곱 방법을 사용하여 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

분포모수
분포 처리 방법 및 모수 옵션을 제공합니다.

데이터로부터 추정(E)
선택할 경우 이 설정은 데이터 및 선택한 분포 유형을 기준으로 분포모수를 추정합니다.

지정
선택하는 경우, 선택한 분포 유형에 대한 분포모수를 지정할 수 있습니다.

참고: 사용가능한 모수는 선택한 분포 유형에 따라 다릅니다.

변환
제공된 옵션은 변환 및 주기성 설정을 설정합니다.

자연로그변환(N)
자연 로그(밑 *e*)로 다양한 진폭을 제거하여 데이터를 변환합니다.

표준화 값(Z)
평균이 0이고 표준 편차가 1인 표본으로 시퀀스 또는 시계열 변수를 변환합니다.

차이
비정상 변수를 평균과 분산이 일정한 정상 변수로 변환하기 위해 도표화 전에 사용되는 차이의 차수를 지정합니다. 필드에 적절한 값을 입력하십시오.

계절 차분
변수가 계절 또는 주기 패턴을 나타낼 경우, 이 설정을 사용하여 도표화 전에 변수에 계절 차이를 적용할 수 있습니다.

참고: 이 설정은 주기성이 정의된 순차 또는 시계열 변수를 양적 변수 중 하나로 선택한 경우에만 사용으로 설정됩니다.

비율추정 공식
제공된 옵션은 비율 추정에 사용되는 공식을 설정합니다.

Blom
공식 (r-3/8) / (w+1/4)을 사용하는 비율 추정값에 따라 새 순위변수를 작성합니다. 여기서 w는 케이스 가중치의 합계이고 r은 순위입니다.

Rankit
공식 (r-1/2) / w를 사용합니다. 여기서 w는 관측값 수이고 r은 1부터 w까지의 순위입니다.

Tukey
공식 (r-1/3) / (w+1/3)을 사용합니다. 여기서 r은 순위이고 w는 케이스 가중치의 합계입니다.

Van der Waerden
Van der Waerden의 변환으로 공식 r/(w+1)로 정의됩니다. 여기서 w는 케이스 가중치의 합계이고 r은 범위가 1부터 w까지의 순위입니다.

등순위에 부여되는 순위
제공된 옵션은 등순위 값의 처리 방법을 판별하는 방법을 제어합니다. 다음 표에는 동률값에 순위를 지정하는 서로 다른 방법이 나타나 있습니다.

표 *6.* 메소드 및 결과 순위

값	평균	낮음(L)	높음	임의로 등순위를 떨어뜨림(K)
1,000만	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

P-P 확률 도표 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **기술 통계량** > **P-P 플롯...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자변수를 선택하여 변수 필드로 옮깁니다.
3. 검정 분포를 선택합니다.

선택적으로 변환 옵션을 선택하여 변환된 값의 확률 도표를 작성할 수 있으며 기대분포 계산 방법도 지정할 수 있습니다.

Q-Q 도표

Q-Q 도표 프로시저는 변환된 값에 대한 확률 도표를 생성합니다. 사용 가능한 검정 분포에는 베타, 카이제곱, 지수, 감마, 절반정규, 라플라스, 로지스틱, 로그정규, 정규, 파레토, 스튜던트 t, 와이블 및 균일이 있습니다. 선택한 분포에 따라 자유도와 기타 모수를 지정할 수 있습니다.

- 변환된 값에 대한 확률 도표를 작성합니다. 변환 옵션에는 자연 로그 변환, 표준화 값, 차이, 계절 차이가 있습니다.

- 예상 분포 계산 및 "동률" 구분 방법을 지정하거나 값이 같은 여러 관측값을 지정할 수 있습니다.

분포 검정
데이터의 분포 유형을 지정합니다. 드롭 다운 목록에서는 다음 옵션을 제공합니다.

베타
베타 분포. 형태1 및 형태2 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다. 모든 관측값은 0 - 1 범위(경계값 포함)에 있어야 합니다.

카이제곱
카이제곱 분포. 자유도(*df)*를 지정해야 합니다. 음수 관측값은 허용되지 않습니다.

지수
지수 분포. 척도 모수 *a*는 양수여야 합니다. 이 모수를 지정하지 않을 경우 DISTRIBUTION이 표본 평균에서 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

감마
감마 분포. 형태 및 척도 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

절반정규
절반정규 분포. 데이터는 위치가 없거나 중앙 집중화된 것으로 가정합니다 (위치 모수=0). 척도 모수 *a*를 지정하거나 DISTRIBUTION이 최대우도 방법을 사용하여 추정하도록 설정할 수 있습니다.

라플라스
라플라스 또는 이중 지수 분포. LAPLACE는 위치 및 척도 모수(*a* 및 *b)*를 사용합니다. 척도 모수*(b*)는 양수여야 합니다. 모수를 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다.

로지스틱
로지스틱 분포. LOGISTIC는 위치 및 척도 모수(*a* 및 *b)*를 사용합니다. 척도 모수*(b*)는 양수여야 합니다. 모수를 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다.

로그정규
로그정규 분포. 척도 및 형태 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 데이터 자연로그의 평균 및 표준편차에서 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

정규
정규 분포. 위치 모수 *a*는 숫자 값일 수 있지만, 척도 모수 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 표본 평균 및 표본 표준편차에서 해당 모수를 추정합니다.

파레토
파레토 분포. 임계값 및 형태 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 *a*는 최소 관측값과 같다고 가정하고 *b*는 최대우도 방법을 사용하여 추정합니다. 음수 관측값은 허용되지 않습니다.

스튜던트 T
스튜던트 T 분포. 자유도(*df)*를 지정해야 합니다.

균일
균일 분포. UNIFORM은 최소값 및 최대값 모수(*a* 및 *b)*를 사용합니다. *a* 매개변수는 *b* 이상이어야 합니다. 매개변수가 지정되지 않은 경우, DISTRIBUTION은 샘플 데이터에서 이를 가정합니다.

Weibull
Weibull 분포. 척도 및 형태 모수 *a*와 *b*는 양수여야 합니다. 지정하지 않을 경우 DISTRIBUTION이 최소 제곱 방법을 사용하여 해당 모수를 추정합니다. 음수 관측값은 허용되지 않습니다.

분포모수
분포 처리 방법 및 모수 옵션을 제공합니다.

데이터로부터 추정(E)
선택할 경우 이 설정은 데이터 및 선택한 분포 유형을 기준으로 분포모수를 추정합니다.

지정
선택하는 경우, 선택한 분포 유형에 대한 분포모수를 지정할 수 있습니다.

참고: 사용가능한 모수는 선택한 분포 유형에 따라 다릅니다.

변환
제공된 옵션은 변환 및 주기성 설정을 설정합니다.

자연로그변환(N)
자연 로그(밑 *e*)로 다양한 진폭을 제거하여 데이터를 변환합니다.

표준화 값(Z)
평균이 0이고 표준 편차가 1인 표본으로 시퀀스 또는 시계열 변수를 변환합니다.

차이
비정상 변수를 평균과 분산이 일정한 정상 변수로 변환하기 위해 도표화 전에 사용되는 차이의 차수를 지정합니다. 필드에 적절한 값을 입력하십시오.

계절 차분
변수가 계절 또는 주기 패턴을 나타낼 경우, 이 설정을 사용하여 도표화 전에 변수에 계절 차이를 적용할 수 있습니다.

참고: 이 설정은 주기성이 정의된 순차 또는 시계열 변수를 양적 변수 중 하나로 선택한 경우에만 사용으로 설정됩니다.

비율추정 공식
제공된 옵션은 비율 추정에 사용되는 공식을 설정합니다.

Blom
공식 (r-3/8) / (w+1/4)을 사용하는 비율 추정값에 따라 새 순위변수를 작성합니다. 여기서 w는 케이스 가중치의 합계이고 r은 순위입니다.

Rankit
공식 (r-1/2) / w를 사용합니다. 여기서 w는 관측값 수이고 r은 1부터 w까지의 순위입니다.

Tukey
공식 (r-1/3) / (w+1/3)을 사용합니다. 여기서 r은 순위이고 w는 케이스 가중치의 합계입니다.

Van der Waerden
Van der Waerden의 변환으로 공식 r/(w+1)로 정의됩니다. 여기서 w는 케이스 가중치의 합계이고 r은 범위가 1부터 w까지의 순위입니다.

등순위에 부여되는 순위
제공된 옵션은 등순위 값의 처리 방법을 판별하는 방법을 제어합니다. 다음 표에는 동률값에 순위를 지정하는 서로 다른 방법이 나타나 있습니다.

표 7. 메소드 및 결과 순위

값	평균	낮음(L)	높음	임의로 등순위를 떨어뜨림(K)
1,000만	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Q-Q 확률 도표 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **기술 통계량** > **Q-Q 플롯...**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상의 숫자변수를 선택하여 변수 필드로 옮깁니다.

3. 검정 분포를 선택합니다.

선택적으로 변환 옵션을 선택하여 변환된 값의 확률 도표를 작성할 수 있으며 기대분포 계산 방법도 지정할 수 있습니다.

ROC 분석

ROC(Receiver operating characteristic) 분석은 (임계값이 진단 검정 결과의 전체 범위에서 다양하므로) 분류 검정의 민감도 대 (1 - 특이도)를 도표화하여 모형 예측의 정확도를 평가하기 위한 유용한 방법입니다. 주어진 ROC 곡선 아래의 전체 영역 (AUC) 은 검정 변수가 관측될 때 예측이 올바른 순서가 될 확률을 나타내는 중요한 통계를 공식화합니다 (케이스 그룹에서 무작위로 선택된 한 개체와 대조군에서 무작위로 선택된 다른 개체에 대해). ROC 분석은 단일 AUC, 정밀도-재현율(PR) 곡선 관련 추론을 지원하며, 독립 집단 또는 대응 개체에서 생성되는 두 개의 ROC 곡선을 비교하는 옵션을 제공합니다.

이전 ROC 곡선 프로시저는 단일 ROC 곡선에 대한 통계적 추론을 지원합니다. 이는 새 ROC 분석 프로시저에 의해 복구될 수도 있습니다. 또한 새 ROC 분석 프로시저는 독립 집단 또는 대응 개체에서 생성된 두 개의 ROC 곡선을 비교할 수 있습니다.

PR 곡선은 정밀도 대 재현율을 도표화하고, 관측 데이터 표본이 상당히 왜곡되었을 때 더 많은 정보를 제공하는 경향이 있으며 계층 분포에서 왜곡이 심한 데이터의 경우 ROC 곡선에 대한 대안을 제공합니다.

예
채무를 변제할 고객과 변제하지 않을 고객으로 고객을 올바로 분류하는 것은 은행의 관심사이므로, 이러한 의사결정을 위해 특수 모형을 개발합니다. ROC 분석은 모형 예측의 정확도를 평가할 때 사용할 수 있습니다.

통계
AUC, 음수 집단, 결측값, 양의 분류, 분리점 값, 확신 강도, 양측 근사 신뢰구간, 분포, 표준오차, 독립 집단 계획, 대응 표본 계획, 비모수 가정, 이분형-음 지수 분포가정, 중심점, 절단점, PR 곡선, 단계 선택 보간법, 근사 유의확률(양측), 민감도 및 (1 - 특이도), 정밀도 및 재현율

방법
독립 집단 또는 대응 개체에서 생성된 두 ROC 곡선 아래의 영역을 비교합니다. 두 ROC 곡선을 비교하면 두 가지 비교 진단 접근법에서 정확도에 대한 더 많은 정보를 제공할 수 있습니다.

ROC 분석 데이터 고려사항

데이터
PR 곡선은 정밀도 대 재현율을 도표화하고, 관측 데이터 표본이 상당히 왜곡되었을 때 더 많은 정보를 제공하는 경향이 있습니다. 단순 선형 보간법에서는 PR 곡선의 지나친 낙관적 추정값을 실수로 생성할 수 있습니다.

가정
케이스 집단에서 무작위로 선택된 한 개체 및 대조집단에서 무작위로 선택된 다른 개체에 대해 검정 변수를 관측할 때 예측 순서가 올바를 것으로 가정합니다. 정의된 각 집단에는 유효한 관측값이 적어도 한 개 포함됩니다. 하나의 집단변수만 하나의 프로시저에 사용됩니다.

ROC 분석 구하기

1. 메뉴에서 다음을 선택합니다.

   분석 > **분류** > **ROC 분석**

   참고: 빨간색으로 강조 표시된 필드는 필수 입력 항목입니다. 모든 필수 입력란에 유효한 값을 입력한 후 **붙여넣기** 및 **확인** 버튼이 활성화됩니다.
2. 하나 이상 선택 **테스트 변수** 확률 변수를 테스트합니다.
3. 하나를 고르세요 **상태변수** .
4. 식별 긍정적인 **상태 변수의 값** .
5. 선택적으로 **대응 표본 계획** 옵션을 선택하거나, 단일 집단변수를 선택하십시오. 두 옵션을 모두 선택할 수는 없습니다.
   - 검정 변수에 대한 대응 표본 계획을 요청하려면 **대응 표본 계획** 설정을 사용하십시오. 대응 표본 계획은 상태변수와 연관된 동일한 개체에서 검정 값이 여러 개 측정된 경우 대응 표본 시나리오에서 두 ROC 곡선을 비교합니다.

참고: 대응 표본 계획을 선택하면 **집단변수** 및 **분포가정**(옵션 대화 상자에 있음) 옵션이 사용 안 함으로 설정됩니다.

- 숫자 그룹화 변수를 선택하면 **그룹 정의 ...** 를 클릭할 수 있습니다. 검정 변수에 대한 독립 그룹 디자인을 요청하고 두 개의 값 (중심점 또는 절단점) 을 지정합니다.

6. 선택적으로 분류, 검정 방향, 표준오차 모수 및 결측값 설정을 정의하려면 옵션을 클릭하십시오.
7. 선택적으로 도표화 및 인쇄 설정(ROC 곡선, 정밀도-재현율 곡선 및 모형 품질 설정 포함)을 정의하려면 표시를 클릭하십시오.
8. **확인**을 클릭하십시오.

ROC 분석: 옵션

다음 옵션을 지정하여 ROC 곡선 분석을 수행할 수 있습니다.

분류
양의 값으로 분류할 때 절사 값을 포함할지 여부를 지정할 수 있습니다. 이 설정은 현재의 결과에 영향을 주지 않습니다.

검정 방향
개체가 양성 반응일 것이라는 확신이 증가함을 나타내는 검정 결과 변수의 방향을 지정하는 옵션을 제공합니다.

영역의 표준오차에 대한 모수
곡선 아래 영역의 표준 오차 추정 방법을 지정할 수 있습니다. 사용 가능한 방법은 비모수 및 이분형-음 지수입니다. 기본 **비모수** 설정은 비모수 분포 가정 하에서 추정값을 제공합니다. **이분형-음 지수** 설정은 이분형-음 지수 분포가정 하에서 추정값을 제공합니다.

이 섹션에서는 AUC의 양측 근사 신뢰구간에 대한 신뢰수준을 지정할 수 있습니다. 사용 가능한 범위는 0.0% - 100.0%이고 기본값은 95%입니다.

참고: 이 설정은 독립 집단 계획에만 적용되며 대응 표본 계획에는 적용되지 않습니다.

결측값
결측값 처리 방법을 지정할 수 있습니다. 이 설정을 선택하지 않을 경우 사용자 결측값 및 시스템 결측값이 모두 제외됩니다. 이 설정을 선택할 경우 사용자 결측값이 유효한 값으로 처리되고, 시스템 결측값은 제외됩니다. 검정 변수에서든 또는 상태변수에서든 시스템 결측값이 있는 케이스는 항상 분석에서 제외됩니다.

ROC 분석: 표시

ROC 분석을 위해 다음 표시 설정을 지정할 수 있습니다.

도표
ROC 및 정밀도-재현율 곡선을 도표화하는 옵션을 제공합니다.

ROC 곡선
선택할 경우 ROC 곡선 도표가 출력에 표시됩니다. ROC 곡선 도표에 대각 참조선을 표시하려면 **대각 참조선 있음**을 선택하십시오.

정밀도-재현율 곡선
선택할 경우 정밀도-재현율 곡선 도표가 출력에 표시됩니다. 정밀도-재현율 곡선은 관측 데이터 표본이 상당히 왜곡되었을 때 더 많은 정보를 제공하는 경향이 있으며 계층 분포에서 왜곡이 심한 데이터의 경우 ROC 곡선에 대한 대안을 제공합니다. 기본값 **진양성을 따라 보간** 설정은 진양성을 따라 단계별로 보간법을 수행합니다. **위양성을 따라 보간** 설정은 위양성을 따라 단계별로 보간법을 수행합니다.

전체 모형 품질
이 설정은 추정된 AUC의 신뢰구간에 대한 하한값을 표시하기 위해 막대도표를 작성할지 여부를 제어합니다. 기본적으로 이 설정은 선택되지 않으므로 막대도표가 표시되지 않습니다.

인쇄
해당 통계에 대한 출력을 정의하는 옵션을 제공합니다.

표준오차 및 신뢰구간
이 설정은 "곡선 아래 영역" 표에 표시할 통계를 제어합니다. 이 설정을 선택하지 않으면 추정된 AUC만 표시됩니다. 이 설정을 선택하면 추가 통계가 표시되는데, 여기에는 AUC의 표준 오차, 근사 유의확률(양측), 영가설 아래에서의 근사 신뢰구간 한계가 포함됩니다.

ROC 곡선의 좌표점
이 설정은 분리점 값과 함께 ROC 곡선의 좌표점을 제어합니다. 이 설정을 선택하지 않으면 좌표점이 출력되지 않습니다. 이 설정을 선택하면 각 ROC 곡선의 분리점 값과 함께 민감도 및 (1-특이도) 대응 값이 제공됩니다.

유덴 지수

ROC 곡선의 각 절사 값에 대한 Youden의 지수 값을 표시합니다.

정밀도-재현율 곡선의 좌표점
이 설정은 분리점 값과 함께 정밀도-재현율 곡선의 좌표점을 제어합니다. 이 설정을 선택하지 않으면 좌표점이 출력되지 않습니다. 이 설정을 선택하면 각 정밀도-재현율 곡선의 분리점 값과 함께 정밀도와 재현율 대응 값이 제공됩니다.

분류자 평가 지표
이 설정은 결과에서 분류자 평가 지표 통계표의 표시를 제어합니다. 이 표는 분류 모형이 무작위 할당에 비해 데이터에 얼마나 적합한지를 보여주며 다음 정보를 제공합니다.

- 사용자 지정 검정 변수
- 집단 정보
- 지니(Gini) 지수. 지니 지수는 2*AUC - 1입니다. 여기서 AUC는 ROC 곡선 아래 영역입니다.
- 최대 K-S 및 분리점 값

ROC 분석: 집단 정의(문자열)

문자열 그룹화 변수의 경우, 그룹 1의 문자열과 그룹 2의 다른 값(예: 예와 아니오)을 입력하십시오. 기타 문자열이 있는 케이스는 분석에서 제외됩니다.

참고: 지정된 값은 변수에 있어야 합니다. 그렇지 않으면 적어도 한 개의 그룹이 비어 있음을 나타내는 오류 메시지가 표시됩니다.

ROC 분석: 집단 정의(숫자)

숫자 집단변수의 경우 두 개의 값, 하나의 중심점 또는 하나의 절단점을 지정하여 *t* 검정에 대한 두 그룹을 정의하십시오.

참고: 지정된 값은 변수에 있어야 합니다. 그렇지 않으면 적어도 한 개의 그룹이 비어 있음을 나타내는 오류 메시지가 표시됩니다.

- **지정값 사용.** 그룹 1의 값과 그룹 2의 다른 값을 입력하십시오. 기타 값이 있는 케이스는 분석에서 제외됩니다. 숫자가 정수일 필요는 없습니다(예: 6.25와 12.5 유효).
- **중심점 값 사용.** 선택할 경우 그룹이 < 및 ≥ 중심점 값으로 구분됩니다.
- **절단점 사용.**
  - **절단점.** 그룹화 변수의 값을 두 변수 세트로 분리하는 수를 입력합니다. 절단점보다 작은 값을 갖는 모든 케이스가 한 그룹을 이루며 절단점보다 큰 값을 갖는 케이스가 다른 그룹을 형성합니다.

ROC 곡선

이 프로시저는 대상을 분류하는 두 개의 범주를 갖는 변수가 있는 분류 방법의 성능을 평가하는 유용한 방법입니다.

참고: 버전29.0.1 이후에는 ROC 명령에 대한 메뉴와 대화 상자가 제거되고 이 명령은 구문으로만 사용할 수 있습니다.

예. 채무를 변제하는 고객과 채무를 불이행하는 고객을 정확하게 분류하는 것이 은행의 관심사입니다. 이러한 고객을 분류할 수 있는 특별한 방법이 개발되었습니다. ROC 곡선을 사용하여 이러한 방법이 제대로 수행되는지 평가할 수 있습니다.

통계. ROC 곡선 아래의 신뢰구간과 ROC 곡선의 좌표점을 가지고 있는 영역을 선택할 수 있습니다. 도표에 대해 ROC 곡선을 선택할 수 있습니다.

방법. ROC 곡선 아래의 영역에 대한 추정값을 비모수적으로 또는 이변량 음지수 모형을 사용하여 모수적으로 계산할 수 있습니다.

ROC 곡선 데이터 고려 사항

이전 ROC 곡선 프로시저는 단일 ROC 곡선에 대한 통계적 추론을 지원합니다. 이는 새 ROC 분석 프로시저에 의해 복구될 수도 있습니다. 또한 새 ROC 분석 프로시저는 독립 집단 또는 대응 개체에서 생성된 두 개의 ROC 곡선을 비교할 수 있습니다.

데이터. 검정변수는 양적변수입니다. 검정변수는 일반적으로 판별 분석이나 로지스틱 회귀분석에서 나오는 확률 또는 대상이 특정한 범주에 속한다고 평가자가 "확신하는 정도"를 나타내는 임의의 척도에 대한 스코어로 구성됩니다. 상태변수는 임의의 유형이 될 수 있으며 대상이 속하는 참 범주를 나타냅니다. 상태변수의 값은 양의 값으로 간주해야 할 범주를 나타냅니다.

가정. 평가자 척도에서 숫자가 증가하면 대상이 범주에 속한다는 신뢰도가 증가하는 것을 나타내는 반면에 평가자 척도에서 숫자가 감소하면 대상이 다른 범주에 속한다는 신뢰도가 증가하는 것을 나타낸다고 가정합니다. 사용자는 어떤 방향이 양의 값인지를 선택해야 합니다. 각 대상이 속하는 참 범주가 알려져 있다고 가정합니다.

ROC 곡선 구하기

1. 메뉴에서 다음을 선택합니다. **분석하다** > **나누다** > **ROC 분석...**
2. 하나 이상 선택 **테스트 변수** .
3. 하나를 고르세요 **상태변수** .
4. 상태변수에 대한 양의 값을 확인합니다.
5. 딸깍 하는 소리 **표시하다** , 선택하다 **ROC 곡선** 그리고 **대각선 기준선 포함** , 필요한 경우.
6. **계속**을 클릭하십시오.

ROC 곡선: 옵션

다음 옵션을 지정하여 ROC 곡선 분석을 수행할 수 있습니다.

분류
양의 값으로 분류할 때 절사 값을 포함할지 여부를 지정할 수 있습니다. 이 설정은 현재의 결과에 영향을 주지 않습니다.

검정 방향
양의 범주와 관련된 척도의 방향을 지정할 수 있습니다.

영역의 표준오차에 대한 모수
곡선 아래 영역의 표준 오차 추정 방법을 지정할 수 있습니다. 사용 가능한 방법은 '비모수' 및 '양음수' 지수입니다. 또한 신뢰구간의 수준을 설정할 수 있습니다. 사용할 수 있는 범위는 50.1에서 99.9% 까지 해당됩니다.

결측값
결측값 처리 방법을 지정할 수 있습니다.

시뮬레이션

선형 회귀분석 같은 예측 모형으로 결과 또는 대상 값을 예측하려면 알려진 입력 세트가 필요합니다. 하지만 실제 세계의 많은 응용 분야에서 입력 값은 불확실합니다. 시뮬레이션을 사용하여 예측 모형의 입력에 대한 불확실성을 설명할 수 있고 불확실성이 존재하는 모형의 다양한 결과의 우도를 평가할 수 있습니다. 예를 들어 입력으로 재료비를 포함하는 이윤 모형이 있지만 시장 변동성으로 인해 비용에 불확실성이 있습니다. 불확실성이 있는 모형에 시뮬레이션을 사용하여 이윤에 대한 효과를 확인할 수 있습니다.

IBM SPSS 통계 의 시뮬레이션은 Monte Carlo 방법을 사용합니다. 불확실한 입력은 삼각분포와 같은 확률 분포로 모형화되며 그러한 입력에 대한 시뮬레이션 값은 이러한 분포에서 생성됩니다. 값이 알려진 입력은 알려진 값

에 고정됩니다. 예측 모형은 각각의 불확실한 입력에 대한 시뮬레이션 값 및 알려진 입력에 대한 고정값을 사용하여 모형의 대상을 계산하기 위해 평가됩니다. 프로세스는 여러 번(일반적으로 수만 또는 수십만 번) 반복되어 확률특성의 질문에 대답하는 데 사용될 수 있는 대상 값의 분포를 가져옵니다. IBM SPSS 통계 에서 프로세스가 반복될 때마다 불확실한 입력에 대한 시뮬레이션 값, 고정 입력 값 및 모형의 예측 대상(또는 여러 예측 대상)의 세트로 구성된 별도의 데이터 케이스(레코드)가 생성됩니다.

시뮬레이션할 변수의 확률 분포를 지정하여 예측 모형이 없이도 데이터를 시뮬레이션할 수 있습니다. 생성된 각 데이터 케이스는 지정된 변수의 시뮬레이션된 값 세트로 구성됩니다.

시뮬레이션을 실행하려면 예측 모형, 불확실한 입력에 대한 확률 분포, 이러한 입력 사이의 상관 및 고정 입력 값과 같은 세부 사항을 지정해야 합니다. 시뮬레이션에 대한 모든 세부사항을 지정하고 나면 시뮬레이션을 실행하고, 지정 사항을 **시뮬레이션 계획** 파일에 선택적으로 저장할 수 있습니다. 시뮬레이션 계획은 다른 사용자와 공유할 수 있어 다른 사용자는 어떻게 만들었는지에 대한 세부 사항을 이해할 필요없이 시뮬레이션을 실행할 수 있습니다.

시뮬레이션 작업에는 두 가지 인터페이스를 사용할 수 있습니다. 시뮬레이션 작성기는 시뮬레이션을 계획하고 실행할 사용자를 위한 고급 인터페이스입니다. 시뮬레이션 작성기는 시뮬레이션 계획, 지정 사항을 시뮬레이션 계획 파일에 저장, 결과 지정 및 시뮬레이션 실행에 대한 전체 공정능력을 제공합니다. 시뮬레이션은 IBM SPSS 모형 파일 또는 시뮬레이션 작성기에서 정의한 사용자 정의 방정식 세트를 기반으로 작성할 수 있습니다. 또한 기존 시뮬레이션 계획을 시뮬레이션 작성기에 로드하고 설정 수정 및 시뮬레이션을 실행할 수 있으며 업데이트한 계획을 선택적으로 저장할 수 있습니다. 시뮬레이션 계획이 있으며 시뮬레이션을 주로 실행하려는 사용자를 위해 더 간단한 인터페이스가 준비되어 있습니다. 이 인터페이스를 통해 시뮬레이션을 다른 환경에서 실행할 수 있게 해주는 설정을 수정할 수 있지만, 시뮬레이션을 계획하는 시뮬레이션 작성기의 전체 공정능력은 제공하지 않습니다.

모형 파일을 기반으로 시뮬레이션을 계획하는 방법

1. 메뉴에서 다음을 선택합니다.

   분석 > **시뮬레이션...**
2. **SPSS 모형 파일 선택**과 **계속**을 차례로 클릭합니다.
3. 모형 파일을 엽니다.

   모형 파일은 IBM SPSS 통계 또는 IBM SPSS Modeler에서 작성된 모형 PMML을 포함하는 XML 파일입니다. 자세한 정보는 245 페이지의 『모형 탭』 주제를 참조하십시오.
4. 시뮬레이션 작성기의 시뮬레이션 탭에서 시뮬레이션한 입력에 대한 확률 분포와 고정 입력 값을 지정합니다. 활성 데이터 세트에 시뮬레이션한 입력의 과거 데이터가 들어 있을 경우 **모두 맞춤**을 클릭하여 각 입력에 가장 근접한 분포 뿐만 아니라 이들 간의 상관도 자동으로 결정합니다. 히스토리 데이터에 맞지 않는 각각의 시뮬레이션된 입력의 경우, 분포 유형을 선택하고 필수 모수를 입력하여 분포를 명시적으로 지정해야 합니다.
5. **실행**을 클릭하여 시뮬레이션을 실행합니다. 기본적으로 시뮬레이션의 세부 사항을 지정하는 시뮬레이션 계획은 저장 설정에서 지정한 위치에 저장됩니다.

다음 옵션을 사용할 수 있습니다.

- 저장된 시뮬레이션 계획의 위치를 수정합니다.
- 시뮬레이션한 입력 사이에 나타난 상관을 지정합니다.
- 범주 입력 사이의 연관 분할표를 자동으로 계산하고 해당 입력에 대한 데이터를 생성할 때 이러한 연관을 사용하십시오.
- 민감도 분석을 지정하여 고정 입력 값이 다르거나 시뮬레이션한 입력의 분포모수가 다른 효과를 조사합니다.
- 생성할 최대 케이스 수를 설정하거나 꼬리 표본추출을 요청하는 등 고급 옵션을 지정합니다.
- 결과를 사용자 정의합니다.
- 시뮬레이션한 데이터를 데이터 파일에 저장합니다.

사용자 정의 방정식을 기반으로 시뮬레이션을 계획하는 방법

1. 메뉴에서 다음을 선택합니다.

   분석 > **시뮬레이션...**

2. **방정식으로 입력** 및 **계속**을 차례로 클릭합니다.
3. 시뮬레이션 작성기의 모형 탭에 있는 **새 방정식**을 클릭하여 예측 모형의 각 방정식을 정의합니다.
4. 시뮬레이션 탭을 클릭하여 시뮬레이션한 입력에 대한 확률 분포와 고정 입력 값을 지정합니다. 활성 데이터 세트에 시뮬레이션한 입력의 과거 데이터가 들어 있을 경우 **모두 맞춤**을 클릭하여 각 입력에 가장 근접한 분포 뿐만 아니라 이들 간의 상관도 자동으로 결정합니다. 히스토리 데이터에 맞지 않는 각각의 시뮬레이션된 입력의 경우, 분포 유형을 선택하고 필수 모수를 입력하여 분포를 명시적으로 지정해야 합니다.
5. **실행**을 클릭하여 시뮬레이션을 실행합니다. 기본적으로 시뮬레이션의 세부 사항을 지정하는 시뮬레이션 계획은 저장 설정에서 지정한 위치에 저장됩니다.

다음 옵션을 사용할 수 있습니다.

- 저장된 시뮬레이션 계획의 위치를 수정합니다.
- 시뮬레이션한 입력 사이에 나타난 상관을 지정합니다.
- 범주 입력 사이의 연관 분할표를 자동으로 계산하고 해당 입력에 대한 데이터를 생성할 때 이러한 연관을 사용하십시오.
- 민감도 분석을 지정하여 고정 입력 값이 다르거나 시뮬레이션한 입력의 분포모수가 다른 효과를 조사합니다.
- 생성할 최대 케이스 수를 설정하거나 꼬리 표본추출을 요청하는 등 고급 옵션을 지정합니다.
- 결과를 사용자 정의합니다.
- 시뮬레이션한 데이터를 데이터 파일에 저장합니다.

예측 모형 없이 시뮬레이션 계획

1. 메뉴에서 다음을 선택합니다.

   분석 > **시뮬레이션...**
2. **시뮬레이션된 데이터 작성**을 클릭하고 **계속**을 클릭하십시오.
3. 시뮬레이션 작성기의 모형 탭에서 시뮬레이션할 필드를 선택하십시오. 활성 데이터 세트에서 필드를 선택하거나 **새로 작성**을 클릭하여 새 필드를 정의할 수 있습니다.
4. 시뮬레이션 탭을 클릭하고 시뮬레이션할 필드의 확률 분포를 지정하십시오. 활성 데이터 세트에 해당 필드에 대한 히스토리 데이터가 포함되어 있는 경우 **모두 맞춤**을 클릭하여 자동으로 데이터에 가장 근접한 분포를 결정하고 필드 사이의 상관도 결정하십시오. 히스토리 데이터에 적합하지 않은 필드의 경우 분포 유형을 선택하고 필수 모수를 입력하여 분포를 명시적으로 지정해야 합니다.
5. **실행**을 클릭하여 시뮬레이션을 실행합니다. 기본적으로 시뮬레이션된 데이터는 저장 설정에 지정된 새 데이터 세트에 저장됩니다. 또한 시뮬레이션의 세부사항을 지정하는 시뮬레이션 계획은 저장 설정에 지정된 위치에 저장합니다.

다음 옵션을 사용할 수 있습니다.

- 시뮬레이션된 데이터나 저장된 시뮬레이션 계획의 위치를 수정하십시오.
- 시뮬레이션된 필드 사이의 알려진 상관을 지정하십시오.
- 범주 필드 사이의 연관 분할표를 자동으로 계산하고 해당 필드에 대한 데이터를 생성할 때 이러한 연관을 사용하십시오.
- 민감도 분석을 지정하여 시뮬레이션된 필드의 분포 모수가 다른 경우의 영향을 조사합니다.
- 생성할 케이스 수 설정과 같은 고급 옵션을 지정하십시오.

시뮬레이션 계획에서 시뮬레이션을 실행하는 방법

시뮬레이션 계획에서 시뮬레이션을 실행할 수 있는 두 가지 옵션이 있습니다. 시뮬레이션 계획에서 실행하도록 주로 계획된 시뮬레이션 실행 대화 상자를 사용하거나 시뮬레이션 작성기를 사용할 수 있습니다.

시뮬레이션 실행 대화 상자를 사용하려면 다음을 수행합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > **시뮬레이션...**
2. **기존 시뮬레이션 계획 열기**를 클릭합니다.

3. **시뮬레이션 작성기에서 열기** 확인 상자가 선택되지 않았는지 확인하고 **계속**을 클릭합니다.
4. 시뮬레이션 계획을 엽니다.
5. 시뮬레이션 실행 대화 상자에서 **실행**을 클릭합니다.

시뮬레이션 작성기에서 시뮬레이션을 실행하려면 다음을 수행합니다.

1. 메뉴에서 다음을 선택합니다.

   분석 > 시뮬레이션...
2. **기존 시뮬레이션 계획 열기**를 클릭합니다.
3. **시뮬레이션 작성기에서 열기** 확인 상자를 선택하고 **계속**을 클릭합니다.
4. 시뮬레이션 계획을 엽니다.
5. 시뮬레이션 탭에서 수정하려는 설정을 수정합니다.
6. **실행**을 클릭하여 시뮬레이션을 실행합니다.

선택적으로 다음을 수행할 수 있습니다.

- 민감도 분석을 설정 또는 수정하여 고정 입력 값이 다르거나 시뮬레이션한 입력의 분포모수가 다른 효과를 조사합니다.
- 시뮬레이션한 입력의 분포 및 상관을 새 데이터에 다시 맞춥니다.
- 시뮬레이션한 입력의 분포를 변경합니다.
- 결과를 사용자 정의합니다.
- 시뮬레이션한 데이터를 데이터 파일에 저장합니다.

시뮬레이션 작성기

시뮬레이션 작성기는 시뮬레이션을 계획하고 실행하기 위한 전체적인 공정능력을 제공합니다. 시뮬레이션 작성기를 사용하여 다음과 같은 일반 작업을 수행할 수 있습니다.

- PMML 모형 파일에서 정의한 IBM SPSS 모형에 대한 시뮬레이션을 계획하고 실행합니다.
- 지정한 사용자 정의 방정식 세트로 정의한 예측 모형에 대한 시뮬레이션을 계획하고 실행합니다.
- 예측 모형이 없이 데이터를 생성하는 시뮬레이션을 계획하고 실행합니다.
- 기존 시뮬레이션 계획을 기반으로 시뮬레이션을 실행하고 필요에 따라 계획 설정을 수정합니다.

모형 탭

예측 모형을 기반으로 하는 시뮬레이션의 경우 모형 탭은 모형의 소스를 지정합니다. 예측 모형을 포함하지 않는 시뮬레이션의 경우 모형 탭은 시뮬레이션될 필드를 지정합니다.

SPSS 모형 파일 선택을 선택합니다. 이 옵션은 예측 모형이 IBM SPSS 모형 파일에서 정의된다고 지정합니다. IBM SPSS 모형 파일은 IBM SPSS 통계 또는 IBM SPSS Modeler에서 작성된 모형 PMML을 포함하는 XML 파일이거나 압축된 파일 아카이브(.zip 파일)입니다. 예측 모형은 IBM SPSS 통계 내의 선형 회귀분석 및 의사결정 나무와 같은 프로시저로 만들어지며 모형 파일로 내보낼 수 있습니다. **찾아보기**를 클릭하고 원하는 파일을 탐색하여 다른 모형 파일을 사용할 수 있습니다.

시뮬레이션이 지원하는 PMML 모형

- 선형 회귀
- 자동 선형 모형
- 일반화 선형 모형
- 일반화 선형 혼합 모형
- 일반선형모형
- 이분형 로지스틱 회귀분석
- 다항 로지스틱 회귀분석
- 순서 다항 회귀분석

- Cox 회귀분석
- 나무
- 부스팅된 나무(C5)
- 판별
- 두 단계 군집
- K-평균 군집
- 신경망
- 규칙 세트(의사결정 목록)

참고:

- 대상 필드(변수) 또는 분할이 여러 개인 PMML 모형은 시뮬레이션에서의 사용이 지원되지 않습니다.
- 2진 로지스틱 회귀분석 모형에 대한 문자열 입력 값은 모형에서 8바이트로 제한됩니다. 해당 문자열 입력을 활성 데이터 세트로 맞추는 경우, 데이터의 값이 8바이트를 넘지 않도록 하십시오. 8바이트를 넘는 데이터 값은 입력에 대한 관련 카테고리 분포에서 제외되며 일치하지 않는 카테고리 결과 테이블에서 일치하지 않음으로 표시됩니다.

모형에 대한 방정식으로 입력합니다. 이 옵션은 예측 모형이 사용자가 만들 하나 이상의 사용자 정의 방정식으로 구성되도록 지정합니다. **새 방정식**을 클릭하여 방정식을 만듭니다. 이렇게 하면 방정식 편집기가 열립니다. 기존 방정식을 수정하고 이를 복사하여 새 방정식의 템플리트로 사용하며 순서를 변경하거나 삭제할 수 있습니다.

- 시뮬레이션 작성기는 동시 방정식 시스템 또는 대상 변수에서 선형이 아닌 방정식을 지원하지 않습니다.
- 사용자 정의 방정식은 지정된 순서대로 평가됩니다. 주어진 대상의 방정식이 다른 대상에 좌우될 경우, 다른 대상은 이전 방정식에 의해 정의되어야 합니다.

  예를 들어, 아래에 주어진 세 방정식 세트에서 이윤에 대한 방정식은 수입 및 지출 값에 달려 있으므로 수입 및 지출에 대한 방정식은 이윤에 대한 방정식 앞에 있어야 합니다.

```
revenue = price*volume

expenses = fixed + volume*(unit_cost_materials + unit_cost_labor)

profit = revenue - expenses
```

모형 없이 시뮬레이션된 데이터 작성. 예측 모형 없이 데이터를 시뮬레이션하려면 이 옵션을 선택하십시오. 활성 데이터 세트에서 필드를 선택하거나 새 필드를 정의하도록 **새로 작성**을 클릭하여 시뮬레이션할 필드를 지정하십시오.

방정식 편집기

방정식 편집기를 사용하여 예측 모형에 대한 사용자 정의 방정식을 작성하거나 수정할 수 있습니다.

- 방정식의 표현식에는 활성 데이터 세트의 필드 또는 방정식 편집기에서 정의한 새 입력 필드가 포함될 수 있습니다.
- 측정 수준, 값 레이블 및 목표에 대해 결과를 생성할지 여부과 같은 목표의 특성을 지정할 수 있습니다.
- 이전에 정의된 방정식의 목표를 현재 방정식의 입력으로 사용할 수 있으며 이를 통해 커플 방정식을 작성할 수 있습니다.
- 기술통계 주석을 방정식에 첨부할 수 있습니다. 주석은 모형 탭에서 방정식과 함께 표시됩니다.

1. 대상 이름을 입력합니다. 선택적으로 대상 입력란에서 **편집**을 클릭하여 정의한 입력값 대화 상자를 여십시오. 이를 통해 목표의 기본 특성을 변경할 수 있습니다.
2. 표현식을 작성하려면 숫자표현식 필드에 성분을 붙여넣거나 직접 입력합니다.

- 활성 데이터 세트의 필드를 사용하여 표현식을 작성하거나 **새로 작성** 단추를 클릭하여 새 입력을 정의할 수 있습니다. 이는 정의한 입력값 대화 상자를 엽니다.
- 함수 그룹 목록에서 그룹을 선택하고 함수 목록의 함수를 두 번 클릭하여 함수를 붙여넣거나 함수를 선택한 다음, 함수 그룹 목록 옆에 있는 화살표를 클릭할 수 있습니다. 물음표로 표시된 모수를 입력하십시오. **모두**라는 레이블이 있는 함수 그룹에서는 사용 가능한 모든 함수 목록을 제공합니다. 대화 상자의 지정된 영역에는 현재 선택한 함수에 대한 간략한 설명이 표시됩니다.

- 문자 상수는 따옴표로 묶어야 합니다.
- 값에 소수점이하자리수가 포함된 경우 소수점 구분자로는 마침표(.)를 사용해야 합니다.

참고: 시뮬레이션에서는 문자열 목표가 있는 사용자 정의 방정식을 지원하지 않습니다.

정의된 입력

정의된 입력 대화 상자를 사용하여 새 입력을 정의하고 대상의 특성을 설정할 수 있습니다.

- 방정식에서 사용할 입력이 활성 데이터 세트에 없는 경우 먼저 정의해야 방정식에서 사용할 수 있습니다.
- 예측 모형을 사용하지 않고 데이터를 시뮬레이션하는 경우 활성 데이터 세트에 없는 시뮬레이션된 모든 입력을 정의해야 합니다.

이름. 대상 또는 입력의 이름을 지정하십시오.

목표. 대상의 측정 수준을 지정할 수 있습니다. 기본 측정 수준은 연속형입니다. 이 대상에 대한 결과를 만들 것인지도 지정할 수 있습니다. 예를 들어 결합된 방정식 세트의 경우 최종 방정식에 대한 대상의 결과에만 관심이 있을 수도 있으므로 다른 대상의 결과를 표시하지 않을 수 있습니다.

시뮬레이션할 입력. 입력 값이 지정한 확률 분포에 따라 시뮬레이션되도록 지정합니다(확률 분포는 시뮬레이션 탭에서 지정함). 측정 수준은 입력에 대한 데이터에 가장 맞는 분포를 찾을 때 (시뮬레이션 탭에서 **맞춤** 또는 **모두 맞춤**을 클릭함) 고려되는 기본 분포 세트를 결정합니다. 예를 들어 측정 수준이 연속형일 경우 연속형 데이터에 적합한 정규 분포는 고려하지만 이항 분포는 고려하지 않습니다.

참고: 문자열 입력을 위한 문자열 측정 수준을 선택하십시오. 시뮬레이션할 문자열 입력은 범주형 분포로 제한됩니다.

정 값 입력. 입력 값이 알려져 있으며 알려진 값에 고정되도록 지정합니다. 고정 입력은 숫자나 문자열일 수 있습니다. 고정 입력에 대한 값을 지정합니다. 문자열 값은 따옴표로 묶을 수 없습니다.

값 레이블. 대상, 시뮬레이션한 입력 및 고정 입력에 대한 값 레이블을 지정할 수 있습니다. 값 레이블은 결과 도표 및 표에 사용됩니다.

시뮬레이션 탭

시뮬레이션 탭은 예측 모형 외의 모든 시뮬레이션 특성을 지정합니다. 시뮬레이션 탭에서 실행할 수 있는 일반 작업은 다음과 같습니다.

- 시뮬레이션한 입력에 대한 확률 분포와 고정 입력 값을 지정합니다.
- 시뮬레이션한 입력 사이의 상관을 지정합니다. 범주형 입력의 경우 해당 입력에 대한 데이터를 생성할 때 활성 데이터 세트의 해당 입력 사이에 있는 연관을 사용하도록 지정할 수 있습니다.
- 꼬리 표본추출 및 과거 데이터에 분포를 맞추기 위한 기준과 같은 고급 옵션을 지정합니다.
- 결과를 사용자 정의합니다.
- 시뮬레이션 계획을 저장할 위치를 지정하고 필요에 따라 시뮬레이션한 데이터를 저장합니다.

시뮬레이션한 필드

시뮬레이션을 실행하려면 각 입력 필드를 고정 또는 시뮬레이션으로 지정해야 합니다. 시뮬레이션한 입력은 값이 불확실한 입력이며 지정한 확률 분포에서 생성됩니다. 히스토리 데이터가 시뮬레이션될 입력에 대해 사용 가능한 경우, 해당 입력 간의 상관과 함께 데이터에 가장 적합한 분포가 자동으로 판별될 수 있습니다. 과거 데이터를 사용할 수 없거나 특정 분포 또는 상관이 필요할 경우 분포나 상관을 수동으로 지정할 수도 있습니다.

고정 입력은 값이 알려진 입력이며 시뮬레이션에서 생성된 각 케이스에 일정하게 유지됩니다. 예를 들어 판매에 대한 선형 회귀 모형을 가격을 포함한 입력 수의 함수로서 가지고 있고, 가격을 현재 시장 가격에 고정된 상태를 유지하고 싶다고 가정해 봅시다. 이 경우 가격을 고정 입력으로 지정하면 됩니다.

예측 모형을 기반으로 시뮬레이션하는 경우 모형의 각 예측자가 해당 시뮬레이션의 입력 필드입니다. 예측 모형을 포함하지 않는 시뮬레이션의 경우 모형 탭에 지정된 필드는 시뮬레이션할 입력입니다.

시뮬레이션한 입력에 대해 자동으로 분포를 맞추고 상관을 계산. 활성 데이터 세트에 시뮬레이션하려는 입력의 과거 데이터가 들어 있을 경우 그러한 입력 데이터에 가장 근접한 분포를 자동으로 찾을 수 있을 뿐만 아니라 이들 사이의 상관도 결정할 수 있습니다. 단계는 다음과 같습니다.

1. 시뮬레이션할 각 입력이 활성 데이터 세트의 올바른 필드와 일치하는지 확인하십시오. 입력은 입력 열에 나열되며 활성 데이터 세트의 일치하는 필드가 다음에 맞춤 열에 표시됩니다. 다음에 맞춤 드롭다운 목록에서 다른 항목을 선택하여 입력을 활성 데이터 세트에 있는 다른 필드와 일치시킬 수 있습니다.

   다음에 맞춤 열에 있는 -없음- 값은 입력이 활성 데이터 세트의 필드와 자동으로 일치할 수 없음을 나타냅니다. 기본적으로 입력은 이름, 측정 수준 및 유형(숫자 또는 문자열)이 데이터 세트 필드와 일치합니다. 활성 데이터 세트에 입력 값의 과거 데이터가 들어 있지 않은 경우 입력에 대한 분포를 수동으로 지정하거나 아래의 설명대로 입력을 고정 입력으로 지정합니다.

2. **모두 맞춤**을 클릭합니다.

가장 근접한 분포 및 연관된 모수가 과거 데이터의 히스토그램(또는 막대도표)에 겹친 분포 도표와 함께 분포 열에 표시됩니다. 시뮬레이션한 입력 사이의 상관은 상관 설정에 표시됩니다. 입력에 대한 행을 선택하고 **세부 사항 맞춤**을 클릭하여 맞춤 결과를 검사하고 특정 입력에 대한 자동 분포 맞춤을 사용자 정의할 수 있습니다. 자세한 정보는 249 페이지의 『세부사항 맞춤』 주제를 참조하십시오.

입력에 대한 행을 선택하고 **맞춤**을 클릭하여 특정 입력에 대한 자동 분포 맞춤을 실행할 수 있습니다. 활성 데이터 세트의 필드와 일치하는 시뮬레이션한 모든 필드에 대한 상관도 자동으로 계산됩니다.

참고:

- 시뮬레이션한 입력에 대한 결측값이 있는 케이스는 분포 적합, 상관관계 계산 및 선택적 분할표에서 제외됩니다(범주형 분포를 사용하는 입력의 경우). 선택적으로 범주형 분포를 사용하는 입력의 사용자 결측값이 유효한 값으로 처리되는지 여부를 지정할 수 있습니다. 기본적으로 결측으로 처리됩니다. 자세한 정보는 250 페이지의 『고급 옵션』 주제를 참조하십시오.
- 연속형 및 순서 입력의 경우 검정된 분포에 허용되는 적합을 찾을 수 없으면 경험적 분포가 가장 근접한 적합으로 제안됩니다. 연속형 입력의 경우 경험적 분포는 과거 데이터의 누적 분포 함수입니다. 순서 입력의 경우 경험적 분포는 과거 데이터의 범주형 분포입니다.

수동으로 분포 지정. 유형 드롭다운 목록에서 분포를 선택하고 모수 격자에서 분포모수를 입력하여 시뮬레이션한 입력에 대한 확률 분포를 수동으로 지정할 수 있습니다. 분포의 모수를 입력하고 나면 지정한 모수에 근거한 분포의 도표 예가 모수 격자 옆에 표시됩니다. 다음은 특정 분포에 대한 몇 가지 참고 사항입니다.

- **범주형.** 범주형 분포는 범주라고 하는 고정된 수의 값이 있는 입력 필드에 대해 설명합니다. 각 범주는 모든 범주에 대한 확률의 합계가 1과 같다는 연관된 확률을 가집니다. 범주를 입력하려면 모수 격자의 왼쪽 열을 클릭하고 범주 값을 지정합니다. 오른쪽 열에 범주와 연관된 확률을 입력합니다.

  참고: PMML 모형의 범주형 입력에는 이 모형을 통해 결정되며 수정될 수 없는 범주가 있습니다.
- **음수 이항 - 실패.** 지정한 성공 수가 관측되기 전에 일련의 시행에서 실패 수의 분포를 설명합니다. 모수 *thresh*는 지정한 성공 수이며 모수 *prob*는 주어진 시행의 성공 확률입니다.
- **음수 이항 - 시행.** 지정한 성공 수가 관측되기 전에 필요한 시행 수의 분포를 설명합니다. 모수 *thresh*는 지정한 성공 수이며 모수 *prob*는 주어진 시행의 성공 확률입니다.
- **범위.** 이 분포는 모든 간격에 대한 확률의 합이 1이 되도록 각 간격에 지정된 확률의 간격 세트로 구성됩니다. 주어진 간격 내의 값은 해당 간격에 정의된 균일 분포에서 가져옵니다. 구간은 최소값, 최대값 및 연관된 확률을 입력하여 지정합니다.

  예를 들어 원자재비가 단위당 $10 ~ $15의 범위에 속할 가능성이 40%이고 단위당 $15 ~ $20의 범위에 속할 가능성이 60%라고 생각하는 경우입니다. [10 - 15] 및 [15 - 20], 두 구간으로 이루어진 범위 분포로 비용을 모형화하여 첫 구간과 관련된 확률을 0.4로 설정하고 두 번째 구간과 관련된 확률을 0.6으로 설정할 수 있습니다. 구간은 연속형일 필요가 없으며 겹칠 수도 있습니다. 예를 들면 구간을 $10 - $15와 $20 - $25 또는 $10 - $15와 $13 - $16로 지정할 수도 있습니다.
- **와이블.** 모수 *c*는 분포의 원점 위치를 지정하는 선택적 위치 모수입니다.

다음과 같은 분포의 모수는 변수 계산 대화 상자에서 사용 가능한 연관된 확률변수 함수와 같은 의미를 지닙니다(Bernoulli, 베타, 이항, 지수, 감마, 로그정규, 음수 이항(시행 및 실패), 정규, normal, 포아송 및 균등 분포).

고정 입력 값 지정. 분포 열의 **유형** 드롭다운 목록에서 고정을 선택하고 고정 값을 입력하여 고정 입력 값을 지정합니다. 입력 값이 숫자냐 문자열이냐에 따라 값은 숫자 또는 문자열이 될 수 있습니다. 문자열 값은 따옴표로 묶을 수 없습니다.

시뮬레이션한 값에 상한/하한 지정. 대부분의 분포는 시뮬레이션한 값에 상한 및 하한을 지정하는 것을 지원합니다. **최소** 입력란에 값을 입력하여 하한을, **최대** 입력란에 값을 입력하여 상한을 지정할 수 있습니다.

입력 잠그기. 잠금 아이콘이 있는 열에서 확인 상자를 선택하여 입력을 잠그면 자동 분포 맞춤에서 해당 입력이 제외됩니다. 이 방법은 수동으로 분포 또는 고정 값을 지정하고 이 값이 자동 분포 맞춤의 영향을 받지 않도록 할 때 가장 유용합니다. 시뮬레이션 실행 대화 상자에서 시뮬레이션 계획을 실행할 사용자와 계획을 공유하려는 경우 및 특정 입력이 변경되지 않도록 하려는 경우에도 잠금이 유용합니다. 따라서 잠긴 입력의 스펙은 시뮬레이션 실행 대화 상자에서 수정할 수 없습니다.

민감도 분석. 민감도 분석을 사용하면 지정된 각 값에 대해 독립적인 시뮬레이션 사례 세트(사실상 별도의 시뮬레이션)를 생성하여 고정 입력 또는 시뮬레이션된 입력에 대한 분포 매개변수의 체계적인 변화의 영향을 조사할 수 있습니다. 민감도 분석을 지정하려면 시뮬레이션한 입력을 선택하고 **민감도 분석**을 클릭합니다. 민감도 분석은 단일 고정 입력 또는 시뮬레이션한 입력 값에 대한 단일 분포모수로 제한됩니다. 자세한 정보는 250 페이지의 『민감도 분석』 주제를 참조하십시오.

맞춤 상태 아이콘

다음에 맞춤 열에 있는 아이콘은 각 입력 필드의 맞춤 상태를 나타냅니다.

표 *8.* 상태 아이콘

아이콘	설명
	입력에 대해 지정된 분포가 없으며 입력이 고정으로 지정되지 않았습니다. 시뮬레이션을 실행하려면 이 입력에 대한 분포를 지정하거나 고정으로 정의하고 고정값을 지정해야 합니다.
	이전에 필드에 맞았던 입력이 활성 데이터 세트에 존재하지 않습니다. 입력에 대한 분포를 활성 데이터 세트에 다시 맞추려고 하지 않는 한 해야 할 일은 없습니다.
	가장 근접한 분포를 세부 사항 맞춤 대화 상자에서 다른 분포로 바꾸었습니다.
	입력이 가장 근접한 분포에 설정되었습니다.
	분포를 수동으로 지정했거나 민감도 분석 반복계산이 이 입력에 대해 지정되었습니다.

세부사항 맞춤

세부 사항 맞춤 대화 상자는 특정 입력에 대한 자동 분포 맞춤 결과를 표시합니다. 분포는 가장 근접한 분포를 처음 나열하며 적합도에 따라 순서가 지정됩니다. 사용 열에서 원하는 분포에 대한 단일 선택 단추를 선택하여 가장 근접한 분포를 덮어쓸 수 있습니다. 사용 열에서 단일 선택 단추를 선택하면 해당 입력에 대한 과거 데이터의 히스토그램(또는 막대도표)에 겹친 분포 도표도 표시됩니다.

적합도 통계량. 기본적으로 연속형 필드에 대해 Anderson-Darling 검정을 사용하여 적합도를 결정합니다. 또는, 연속형 필드의 경우에만 고급 옵션 설정에서 선택하여 적합도에 Kolmogorov-Smirnoff 검정을 지정할 수 있습니다. 연속형 입력의 경우 두 검정의 결과는 분포를 순서화하는 데 사용된 선택한 검정과 함께 적합도 통계량 열(A는 Anderson-Darling, K는 Kolmogorov-Smirnoff)에 표시됩니다. 순서 및 명목 입력에는 카이제곱 검정이 사용됩니다. 검정과 연관된 P-값도 표시됩니다.

매개변수. 맞추어진 각 분포와 연관된 분포모수는 모수 열에 표시됩니다. 다음과 같은 분포의 모수는 변수 계산 대화 상자에서 사용 가능한 연관된 확률변수 함수와 같은 의미를 지닙니다(Bernoulli, 베타, 이항, 지수, 감마, 로그정규, 음수 이항(시행 및 실패), 정규, normal, 포아송 및 균등 분포). 범주형 분포에서 모수 이름은 범주이며 모수 값은 연관된 확률입니다.

사용자 정의된 분포 세트로 다시 맞춤. 기본적으로 입력 값의 측정 수준은 자동 분포 맞춤을 고려한 분포 세트를 결정하는 데 사용됩니다. 예를 들어, 연속형 입력은 맞추지만 포아송과 이항 같은 이산형 분포는 맞추지 않을 경우 로그정규 및 감마 같은 연속형 분포를 고려합니다. 다시 맞춤 열에서 분포를 선택하여 기본 분포의 서브세트를 선택할 수 있습니다. 또한 **다음으로 처리(측도)** 드롭다운 목록에서 다른 측정 수준을 선택하고 다시 맞춤 열에서 분포를 선택하여 분포의 기본 세트를 바꿀 수 있습니다. 사용자 정의 분포 세트를 사용해 다시 맞추려면 **다시 맞춤 실행**을 클릭하십시오.

참고:

- 시뮬레이션한 입력에 대한 결측값이 있는 케이스는 분포 적합, 상관관계 계산 및 선택적 분할표에서 제외됩니다(범주형 분포를 사용하는 입력의 경우). 선택적으로 범주형 분포를 사용하는 입력의 사용자 결측값이 유효한 값으로 처리되는지 여부를 지정할 수 있습니다. 기본적으로 결측으로 처리됩니다. 자세한 정보는 250 페이지의 『고급 옵션』 주제를 참조하십시오.
- 연속형 및 순서 입력의 경우 검정된 분포에 허용되는 적합을 찾을 수 없으면 경험적 분포가 가장 근접한 적합으로 제안됩니다. 연속형 입력의 경우 경험적 분포는 과거 데이터의 누적 분포 함수입니다. 순서 입력의 경우 경험적 분포는 과거 데이터의 범주형 분포입니다.

민감도 분석

민감도 분석을 사용하여 고정 입력이 다른 효과나 지정한 값 세트에 대해 시뮬레이션한 입력의 분포모수를 조사할 수 있습니다. 사실상 별도의 시뮬레이션인 시뮬레이션한 케이스의 독립 세트가 지정한 각 값에 대해 생성되어 입력이 다른 효과를 조사할 수 있습니다. 시뮬레이션한 케이스의 각 세트를 **반복**이라고 합니다.

반복. 이 옵션을 선택하면 다양해질 입력에 대해 값의 세트를 지정할 수 있습니다.

- 분포 모수의 값을 변경하는 경우 드롭 다운 목록에서 모수를 선택하십시오. 반복 격자로 모수 값에 값의 세트를 입력합니다. **계속**을 클릭하면 연관된 입력의 모수 격자에 값의 반복 수를 지정하는 지수와 함께 지정된 값이 추가됩니다.
- 범주형 및 범위 분포의 경우 범주의 확률이나 구간은 각각 다양할 수 있지만 범주 값과 구간 엔드포인트는 다양할 수 없습니다. 드롭 다운 목록에서 범주 또는 간격을 선택하고 반복 격자로 모수 값에서 확률 세트를 지정하십시오. 다른 범주 또는 구간의 확률이 이에 따라 자동으로 조정됩니다.

반복하지 않음. 이 옵션을 사용하여 입력에 대한 반복을 취소합니다. **계속**을 클릭하면 반복이 제거됩니다.

상관

시뮬레이션할 입력 필드는 상관되는 경우가 많습니다(예: 키와 몸무게). 시뮬레이션할 입력 값 사이의 상관은 시뮬레이션한 값이 그러한 상관을 유지할 수 있도록 설명되어야 합니다.

맞춤 시 상관 다시 계산. 이 선택은 시뮬레이션한 필드 설정의 **모두 맞춤** 또는 **맞춤** 작업을 통해 분포를 활성 데이터 세트에 맞출 때 시뮬레이션한 입력 사이의 상관이 자동으로 계산되도록 지정합니다.

맞춤 시 상관을 다시 계산하지 않음. 수동으로 상관을 지정하고 분포를 활성 데이터 세트에 자동으로 맞출 때 덮어써지는 것을 방지하려면 이 옵션을 선택합니다. 상관 그리드에 입력된 값은 -1 과 1사이여야 합니다. 값 0은 연관된 입력 쌍 간에 상관이 없음을 지정합니다.

재설정. 모든 상관을 0으로 재설정합니다.

범주형 분포를 사용하는 입력에 맞춘 다방향 분할표 사용. 범주형 분포를 사용하는 입력의 경우 해당 입력 사이의 연관을 설명하는 활성 데이터 세트에서 자동으로 다방향 분할표를 계산할 수 있습니다. 해당 입력에 대한 데이터를 생성할 때 분할표를 사용합니다. 시뮬레이션 계획을 저장하도록 선택하는 경우 분할표를 계획 파일에 저장하여 계획을 실행할 때 사용합니다.

- **활성 데이터 집합에서 컨틴전시 테이블을 계산합니다.** 분할표가 포함된 기존 시뮬레이션 계획에 대해 작업하는 경우 활성 데이터 세트의 분할표를 다시 계산할 수 있습니다. 이 조치를 수행하면 로드된 계획 파일의 분할표를 대체합니다.
- **로드된 시뮬레이션 계획의 우발 상황 테이블을 사용합니다.** 기본적으로 분할표가 포함된 시뮬레이션 계획을 로드할 때 계획의 테이블을 사용합니다. **활성 데이터 세트의 분할표 계산**을 선택하여 활성 데이터 세트의 분할표를 다시 계산할 수 있습니다.

고급 옵션

최대 케이스 수. 시뮬레이션한 데이터의 최대 케이스 수와 생성할 연관 대상값을 지정합니다. 민감도 분석을 지정한 경우 각 반복에 대한 케이스의 최대 수입니다.

중지 기준 대상. 예측 모형이 둘 이상의 대상을 포함한 경우 중지 기준을 적용할 대상을 선택할 수 있습니다.

중지 기준. 이 선택은 허용 가능한 케이스의 최대 수가 생성되기 전에 시뮬레이션을 중지할 기준을 지정합니다.

- **최대에 도달할 때까지 계속.** 케이스의 최대 수에 도달할 때까지 시뮬레이션한 케이스를 생성하도록 지정합니다.

- **꼬리가 표본추출되었을 경우 중지.** 지정한 대상 분포의 꼬리 중 하나가 적절하게 표본추출되었는지 확인하고 싶을 때 이 옵션을 사용합니다. 지정한 꼬리 표본추출이 완료되거나 케이스의 최대 수에 도달할 때까지 시뮬레이션한 케이스가 생성됩니다. 예측 모형에 여러 대상이 포함되어 있을 경우 **중지 기준 대상** 드롭다운 목록에서 이 기준을 적용할 대상을 선택하십시오.

  유형. 10,000,000과 같은 대상 값이나 제 99 백분위수 같은 백분위수를 지정하여 꼬리 영역의 경계를 정의할 수 있습니다. **유형** 드롭다운 목록에서 값을 선택한 경우 값 입력란에 경계값을 입력하고 **방향** 드롭다운 목록을 사용하여 이 값이 왼쪽 꼬리 영역의 경계인지 오른쪽 꼬리 영역의 경계인지를 지정합니다. **유형** 드롭다운 목록에서 백분위수를 선택한 경우 백분위수 입력란에 값을 입력합니다.

  빈도. 꼬리가 적절하게 표본추출될 수 있도록 꼬리 영역에 있어야 하는 대상 값의 수를 지정합니다. 이 수에 도달하면 케이스 생성이 중지됩니다.

- **평균의 신뢰구간이 지정한 임계값 내에 있을 때 중지.** 주어진 대상의 평균이 지정한 정확도와 함께 알려졌는지 확인하고 싶을 때 이 옵션을 사용합니다. 지정한 정확도를 달성했거나 케이스의 최대 수에 도달할 때까지 시뮬레이션한 케이스가 생성됩니다. 이 옵션을 사용하려면 신뢰수준과 임계값을 지정해야 합니다. 시뮬레이션한 케이스는 지정한 수준과 연관된 신뢰구간이 임계값 내에 있을 때까지 생성됩니다. 예를 들면, 이 옵션을 사용하여 평균의 신뢰구간이 95%, 신뢰수준이 평균 값의 5% 이내에 있을 때까지 케이스가 생성되도록 지정할 수 있습니다. 예측 모형에 여러 대상이 포함되어 있을 경우 **중지 기준 대상** 드롭다운 목록에서 이 기준을 적용할 대상을 선택하십시오.

  임계값 유형. 임계값을 숫자값이나 평균의 퍼센트로 지정할 수 있습니다. **임계값 유형** 드롭다운 목록에서 값을 선택한 경우 임계값(값) 입력란에 임계값을 입력합니다. **임계값 유형** 드롭다운 목록에서 퍼센트를 선택한 경우 임계값(퍼센트) 입력란에 값을 입력합니다.

표본추출할 케이스 수. 시뮬레이션된 입력에 대한 분포를 자동으로 활성 데이터 세트에 맞출 때 사용할 케이스 수를 지정합니다. 데이터 세트가 아주 클 경우 분포 맞춤에 사용될 케이스 수를 제한하는 것을 고려해볼 수도 있습니다. **N 케이스로 제한**을 선택한 경우 첫 번째 N 케이스가 사용됩니다.

적합도 기준(연속형). 연속형 입력의 경우 시뮬레이션된 입력에 대한 분포를 활성 데이터 세트에 맞출 때 Anderson-Darling 검정 또는 Kolmogorov-Smirnoff 검정의 적합도를 사용하여 분포를 순위화할 수 있습니다. 기본적으로 Anderson-Darling 검정이 선택되며 꼬리 영역에 가장 적합한지 확인하고 싶을 때 특히 권장합니다.

경험적 분포. 연속형 입력의 경우 경험적 분포는 과거 데이터의 누적 분포 함수입니다. 연속형 입력의 경험적 분포를 계산하기 위해 사용되는 구간 수를 지정할 수 있습니다. 기본값은 100이고 최대값은 1000입니다.

결과 복제. 난수 시드를 설정하면 시뮬레이션을 복제할 수 있습니다. 정수를 지정하거나 1과 2147483647 사이(1과 2147483647 포함)의 유사 난수 정수를 만드는 **생성**을 클릭하십시오. 기본값은 629111597입니다.

참고: 특정 난수 시작값의 경우, 스레드의 수가 변경되지 않는 한 결과는 복제됩니다. 특정 컴퓨터에서, 스레드의 수는 SET THREADS 명령문을 실행하여 변경하지 않는 한 상수입니다. 다른 컴퓨터에서 시뮬레이션을 실행하는 경우, 내부 알고리즘은 각 컴퓨터에 대한 스레드의 수를 결정하는데 사용되므로 스레드의 슈가 변경됩니다.

범주형 분포를 사용하는 입력의 사용자 결측값. 이러한 제어는 범주형 분포를 사용하는 입력의 사용자 결측값이 올바른 값으로 처리되는지 지정합니다. 기타 모든 유형의 입력에 대한 시스템 결측값 및 사용자 결측값은 항상 올바르지 않은 것으로 처리됩니다. 분포 맞춤, 상관 계산 및 선택적 분할표에 케이스가 포함되려면 모든 입력 값이 유효해야 합니다.

밀도 함수

이런 설정으로 연속형 대상의 경우 확률 밀도 함수와 누적 분포 함수에 대한 결과뿐만 아니라 범주형 대상에 대한 예측 값의 막대도표도 사용자 정의할 수 있습니다.

확률 밀도 함수(PDF). 확률 밀도 함수는 대상 값의 분포를 표시합니다. 연속형 대상의 경우 이를 통해 대상이 주어진 영역 이내에 있는 확률을 결정할 수 있습니다. 범주형 대상(명목 또는 순서의 측정 수준을 가진 대상)의 경우 대상의 각 범주에 속하는 케이스의 퍼센트를 표시하는 막대도표를 생성합니다. 아래 설명된 설정을 보고하기 위한 범주 값으로 PMML 모형의 범주형 대상에 대한 추가 옵션을 사용할 수 있습니다.

두 단계 군집 모형 및 K-평균 군집 모형의 경우 소속군집의 막대도표가 생성됩니다.

누적 분포 함수(CDF). 누적 분포 함수는 지정한 값보다 적거나 같은 대상 값의 확률을 표시합니다. 연속형 대상에만 사용할 수 있습니다.

슬라이더 위치. PDF 및 CDF 도표에서 이동할 수 있는 참조선의 초기 위치를 지정할 수 있습니다. 하단 및 상단 선에 지정된 값은 백분위수가 아니라 가로축의 위치를 나타냅니다. **-무한**을 선택하여 하단 선을 제거하거나 **무한**을

선택하여 상단 선을 제거할 수 있습니다. 기본적으로 선은 5번째 및 95번째 백분위수에 있습니다. 다중 대상 또는 민감도 분석 반복계산의 결과로 인해 다중 분포 함수가 단일 도표에 표시되는 경우 기본값은 첫 번째 반복이나 첫 번째 대상의 분포를 나타냅니다.

참조선(연속형). 연속형 대상의 확률 밀도 함수 및 누적 분포 함수에 추가할 다양한 수직 참조선을 요청할 수 있습니다.

- **시그마.** 대상의 평균에서 지정한 수의 표준 편차를 더하고 뺀 값에 참조선을 추가할 수 있습니다.
- **백분위수.** 하단 및 상단 입력란에 값을 입력하여 대상에 대한 하나 또는 두 개의 분포 백분위수 값에 참조선을 추가할 수 있습니다. 예를 들어 상단 입력란의 값 95는 관측값의 95% 미만인 제 95 백분위수를 나타냅니다. 마찬가지로 하단 입력란의 값 5는 관측값의 5% 미만인 제 5 백분위수를 나타냅니다.
- **사용자 정의 참조선.** 대상의 지정값에 참조선을 추가할 수 있습니다.

참고: 다중 대상 또는 민감도 분석 반복계산의 결과로 인해 다중 분포 함수가 단일 도표에 표시되는 경우 참조선은 첫 번째 반복이나 첫 번째 대상의 분포에만 적용됩니다. PDF나 CDF 도표에서 액세스하는 도표 옵션 대화 상자에서 기타 분포에 참조선을 추가할 수 있습니다.

별도의 연속형 대상의 결과로부터 오버레이. 여러 개의 연속형 대상의 경우, 이 옵션은 여러 개의 모든 연속형 대상에 대한 분포 함수를 확률 밀도 함수에 대한 도표 한 개 및 누적 분포 함수에 대한 다른 도표로써 단일 도표에 표시합니다. 이 옵션을 선택하지 않을 경우 각 대상의 결과는 별도의 도표에 표시됩니다.

보고할 범주 값. 범주형 대상이 있는 PMML 모형의 경우 모형 결과는 각 범주에 속하는 대상 값인 예측 확률(각 범주당 하나)의 세트입니다. 확률이 가장 높은 범주는 예측 범주에서 가져오며 위의 **확률 밀도 함수** 설정에서 설명한 막대도표를 생성하는 데 사용됩니다. **예측 범주**를 선택하면 막대도표가 생성됩니다. **예측 확률**을 선택하면 대상의 각 범주에 대한 예측 확률의 분포 히스토그램이 생성됩니다.

민감도 분석의 그룹화. 민감도 분석을 포함한 시뮬레이션은 분석으로 정의한 각 반복(달라질 각 입력 값에 하나의 반복)에 대한 예측된 대상값의 독립 세트를 생성합니다. 반복이 있을 경우 범주형 대상에 대해 예측된 범주의 막대도표가 모든 반복의 결과를 포함한 수평배열 막대도표로 표시됩니다. 범주를 함께 그룹화하도록 선택하거나 반복을 함께 그룹화할 수 있습니다.

결과

토네이도 도표. 토네이도 도표는 다양한 메트릭을 사용하여 대상과 시뮬레이션한 입력 사이의 관계를 표시하는 막대도표입니다.

- **입력 값이 있는 대상의 상관.** 이 옵션은 주어진 대상과 시뮬레이션한 각 입력 사이에 대한 상관계수의 토네이도 도표를 만듭니다. 이 유형의 토네이도 도표는 명목 측정 수준 또는 순서 측정 수준의 대상이나 범주형 분포로 시뮬레이션된 입력을 지원하지 않습니다.
- **분산에 대한 기여도.** 이 옵션은 시뮬레이션한 각 입력으로부터 대상의 분산에 대한 기여도를 표시하는 토네이도 도표를 만들어 사용자가 각 입력이 대상의 전체 불확실성에 기여하는 정도를 평가할 수 있게 해줍니다. 이 토네이도 도표 유형은 측정 수준이 순서 또는 명목인 대상이나 시뮬레이션한 입력(범주형, Bernoulli, 이항, 포아송 또는 음수 이항)을 지원하지 않습니다.
- **변경할 대상의 민감도.** 이 옵션은 입력과 연관된 분포 표준 편차의 지정한 개수를 더하거나 뺌으로써 각 시뮬레이션한 입력 조절이 대상에 미치는 효과를 표시하는 토네이도 도표를 만듭니다. 이 토네이도 도표 유형은 측정 수준이 순서 또는 명목인 대상이나 시뮬레이션한 입력(범주형, Bernoulli, 이항, 포아송 또는 음수 이항)을 지원하지 않습니다.

대상 분포의 상자도표. 상자도표는 연속형 대상에 사용할 수 있습니다. 예측 모형에 연속형 대상이 여러 개이면서 단일 도표에 모든 대상의 상자도표를 표시하려는 경우 **별도 대상의 결과로부터 오버레이**를 선택합니다.

대상 대 입력의 산점도. 대상 대 시뮬레이션한 입력의 산점도는 연속형 및 범주형 대상에 모두 사용할 수 있으며 연속형 및 범주형이 모두 있는 대상의 산점도를 포함합니다. 범주형 대상 또는 범주형 입력에 관련된 산포는 히트맵으로 표시됩니다.

백분위수 값의 표 만들기. 연속형 대상의 경우 대상 분포의 지정한 백분위수 테이블을 만들 수 있습니다. 분위수(제 25, 50, 75 백분위수)를 사용하여 관측값을 동일한 크기의 네 그룹으로 나눕니다. 네 그룹이 아닌 동일한 수의 그룹을 원할 경우 **구간**을 선택하고 수를 지정합니다. **사용자 정의 백분위수**를 선택하여 개별 백분위수를 지정합니다(예: 제 99 백분위수).

대상 분포의 기술통계. 연속형 및 범주형 대상 뿐만 아니라 연속형 입력에 대한 기술통계 테이블을 만드는 옵션입니다. 연속형 대상의 경우 테이블에는 평균, 표준 편차, 중위수, 최소값 및 최대값, 지정한 수준의 평균의 신뢰구

간, 대상 분포의 제 5와 제 95 백분위수가 포함됩니다. 범주형 대상의 경우 테이블에는 대상의 각 범주에 속하는 케이스의 퍼센트가 포함됩니다. PMML 모형의 범주형 대상의 경우 테이블에는 대상의 각 범주에 대한 평균 확률도 포함됩니다. 연속형 입력의 경우 테이블에는 평균, 표준 편차, 최소값 및 최대값이 포함됩니다.

입력의 상관 및 분할표. 이 옵션은 시뮬레이션된 입력 사이의 상관계수 테이블을 표시합니다. 범주형 분포를 사용하는 입력을 분할표에서 생성하는 경우 해당 입력을 위해 생성된 데이터 분할표도 표시됩니다.

결과에 포함할 시뮬레이션한 입력. 기본적으로 시뮬레이션한 모든 입력이 결과에 포함됩니다. 선택한 시뮬레이션한 입력을 결과에서 제외할 수 있습니다. 이렇게 하면 선택한 값이 토네이도 도표, 산점도 및 표 형식 결과에서 제외됩니다.

연속형 대상의 범위 제한. 하나 이상의 연속형 대상에 대해 올바른 값 범위를 지정할 수 있습니다. 지정된 범위 밖에 있는 값은 모든 결과 및 대상과 연관된 분석에서 제외됩니다. 하한을 설정하려면 한계 열에서 **하한**을 선택하고 최소값 열에 값을 입력하십시오. 상한을 설정하려면 한계 열에서 **상한**을 선택하고 최대값 열에 값을 입력하십시오. 하한 및 상한을 모두 설정하려면 한계 열에서 **둘 다**를 선택하고 최소값 및 최대값 열에 값을 입력하십시오.

표시 형식. 대상값 및 입력 값(고정 입력과 시뮬레이션한 입력)을 표시할 때 사용할 형식을 설정할 수 있습니다.

저장

이 시뮬레이션의 계획 저장. 시뮬레이션의 현재 지정 사항을 시뮬레이션 계획 파일에 저장할 수 있습니다. 시뮬레이션 계획 파일의 확장은 *.splan*입니다. 계획은 시뮬레이션 작성기에서 다시 열 수 있으며 경우에 따라 수정하여 시뮬레이션을 실행할 수 있습니다. 시뮬레이션 계획은 다른 사용자와 공유할 수 있으며, 다른 사용자는 시뮬레이션 실행 대화 상자에서 계획을 실행할 수 있습니다. 시뮬레이션 계획은 밀도 함수에 대한 설정, 도표 및 표에 대한 결과 설정, 맞춤에 대한 고급 옵션 설정, 경험적 분포 및 난수 시드를 제외한 모든 사양을 포함합니다.

시뮬레이션한 데이터를 새 데이터 파일로 저장. 시뮬레이션한 입력 값, 고정 입력 값 및 예측 대상 값을 SPSS Statistics 데이터 파일, 현재 세션의 새 데이터 세트 또는 Excel 파일에 저장할 수 있습니다. 데이터 파일의 각 케이스(또는 행)는 대상 값을 생성하는 시뮬레이션한 입력 및 고정 입력과 함께 대상의 예측값으로 구성됩니다. 민감도 분석이 지정되면 각 반복에서는 반복 수로 레이블되는 연속형 케이스 세트가 발생합니다.

시뮬레이션 실행 대화 상자

시뮬레이션 실행 대화 상자는 시뮬레이션 계획이 있으며 시뮬레이션을 주로 실행하려는 사용자를 위해 계획되었습니다. 또한 다른 상황에서 시뮬레이션을 실행하기 위해 필요한 기능도 제공합니다. 시뮬레이션 작성기를 사용하여 다음과 같은 일반 작업을 수행할 수 있습니다.

- 민감도 분석을 설정 또는 수정하여 고정 입력 값이 다르거나 시뮬레이션한 입력의 분포모수가 다른 효과를 조사합니다.
- 불확실한 입력의 확률 분포(및 이러한 입력 값 사이의 상관)를 새 데이터에 다시 맞춥니다.
- 시뮬레이션한 입력의 분포를 수정합니다.
- 결과를 사용자 정의합니다.
- 시뮬레이션을 실행합니다.

시뮬레이션 탭

시뮬레이션 탭을 사용하여 민감도 분석을 지정하며 시뮬레이션한 입력의 확률 분포 및 시뮬레이션한 입력 사이의 상관을 새 데이터에 다시 맞추고 시뮬레이션한 입력과 연관된 확률 분포를 수정할 수 있습니다.

시뮬레이션된 입력 격자는 시뮬레이션 계획에 정의된 각 입력 필드의 항목을 포함합니다. 각 항목은 연관된 분포 곡선의 표본 도표와 함께 입력 이름 및 입력과 연관된 확률 분포 유형을 표시합니다. 각 입력 값에는 분포를 새 데이터에 다시 맞출 때 유용한 연관된 상태 아이콘(체크 표시가 있는 색상 원)도 있습니다. 또한 입력 값은 입력이 잠겨 있으며 시뮬레이션 실행 대화 상자에서 수정하거나 새 데이터에 다시 맞출 수 없음을 나타내는 잠금 아이콘을 포함할 수도 있습니다. 잠겨 있는 입력 값을 수정하려면 시뮬레이션 계획을 시뮬레이션 작성기에서 열어야 합니다.

각 입력은 시뮬레이션 또는 고정입니다. 시뮬레이션한 입력은 값이 불확실한 입력이며 지정한 확률 분포에서 생성됩니다. 고정 입력은 값이 알려진 입력이며 시뮬레이션에서 생성된 각 케이스에 일정하게 유지됩니다. 특정 입력으로 작업하려면 시뮬레이션한 입력 격자에서 입력에 대한 항목을 선택하십시오.

민감도 분석 지정

민감도 분석을 사용하면 지정된 각 값에 대해 독립적인 시뮬레이션 사례 세트(사실상 별도의 시뮬레이션)를 생성하여 고정 입력 또는 시뮬레이션된 입력에 대한 분포 매개변수의 체계적인 변화의 영향을 조사할 수 있습니다. 민감도 분석을 지정하려면 시뮬레이션한 입력을 선택하고 **민감도 분석**을 클릭합니다. 민감도 분석은 단일 고정 입력 또는 시뮬레이션한 입력 값에 대한 단일 분포모수로 제한됩니다. 자세한 정보는 250 페이지의 『민감도 분석』 주제를 참조하십시오.

분포를 새 데이터에 다시 맞춤

시뮬레이션한 입력(및 시뮬레이션한 입력 사이의 상관)에 대한 확률 분포를 활성 데이터 세트에 있는 데이터에 자동으로 다시 맞추려면 다음을 수행하십시오.

1. 모형의 각 입력 값이 활성 데이터 세트의 올바른 필드와 일치하는지 확인합니다. 각각의 시뮬레이션한 입력은 해당 입력 값과 연관된 **필드** 드롭다운 목록에 지정된 활성 데이터 세트의 필드에 맞춰집니다. 아래 표시된 것처럼 물음표가 있는 체크 표시가 된 상태 아이콘이 있는 입력 값을 찾아 일치하지 않는 입력 값을 쉽게 식별할 수 있습니다.

2. **데이터 세트의 필드에 맞춤**을 선택한 다음 목록에서 필드를 선택하여 일치하는 필요한 필드를 수정합니다.
3. **모두 맞춤**을 클릭합니다.

맞춰진 각 입력의 경우 데이터에 가장 근접한 분포가 과거 데이터의 히스토그램(또는 막대도표)에 겹친 분포 도표와 함께 표시됩니다. 허용 가능한 맞춤을 찾을 수 없는 경우 경험적 분포가 사용됩니다. 경험적 분포에 맞춰진 입력의 경우 경험적 분포가 해당 히스토그램으로 나타나므로 과거 데이터의 히스토그램만 볼 수 있습니다.

참고: 상태 아이콘의 전체 목록은 247 페이지의 『시뮬레이션한 필드』 주제를 참조하십시오.

확률 분포 수정

시뮬레이션한 입력의 확률 분포를 수정하고 필요한 경우 시뮬레이션한 입력을 고정 입력으로 변경할 수 있습니다(그 반대도 가능).

1. 입력을 선택하고 **수동으로 분포 설정**을 선택합니다.
2. 분포 유형을 선택하고 분포모수를 지정합니다. 시뮬레이션한 입력을 고정 입력으로 변경하려면 **유형** 드롭다운 목록에서 고정을 선택합니다.

분포의 모수를 입력하고 나면 분포의 표본 도표(입력 값의 항목에 표시됨)가 업데이트되어 변경 사항이 반영됩니다. 수동으로 확률 분포를 지정하는 것에 대한 자세한 내용은 247 페이지의 『시뮬레이션한 필드』 주제를 참조하십시오.

맞춤 시 범주형 입력의 사용자 결측값 포함. 활성 데이터 세트의 데이터에 다시 맞출 때 범주형 분포를 사용하는 입력의 사용자 결측값이 올바른 값으로 처리되는지 지정합니다. 기타 모든 유형의 입력에 대한 시스템 결측값 및 사용자 결측값은 항상 올바르지 않은 것으로 처리됩니다. 상관 계산 및 분포 맞춤에 케이스가 포함되려면 모든 입력 값이 유효해야 합니다.

결과 탭

결과 탭을 사용하여 시뮬레이션으로 생성된 결과를 사용자 정의할 수 있습니다.

밀도 함수. 밀도 함수는 시뮬레이션의 결과 세트를 살피는 주요 수단입니다.

- **확률 밀도 함수.** 확률 밀도 함수는 대상값의 분포를 표시하여 대상이 주어진 영역 이내에 있는 확률을 결정할 수 있습니다. "열악한 서비스", "보통 서비스", "좋은 서비스", "뛰어난 서비스"와 같은 고정 세트 결과를 가진 대상의 경우 대상의 각 범주에 속하는 케이스의 퍼센트를 표시하는 막대도표가 생성됩니다.
- **누적 분포 함수.** 누적 분포 함수는 지정한 값보다 적거나 같은 대상 값의 확률을 표시합니다.

토네이도 도표. 토네이도 도표는 다양한 메트릭을 사용하여 대상과 시뮬레이션한 입력 사이의 관계를 표시하는 막대도표입니다.

- **입력 값이 있는 대상의 상관.** 이 옵션은 주어진 대상과 시뮬레이션한 각 입력 사이에 대한 상관계수의 토네이도 도표를 만듭니다.

- **분산에 대한 기여도.** 이 옵션은 시뮬레이션한 각 입력으로부터 대상의 분산에 대한 기여도를 표시하는 토네이도 도표를 만들어 사용자가 각 입력이 대상의 전체 불확실성에 기여하는 정도를 평가할 수 있게 해줍니다.
- **변경할 대상의 민감도.** 이 옵션은 입력과 연관된 분포 표준 편차 한 개를 더하거나 뺌으로써 각 시뮬레이션한 입력 조절이 대상에 미치는 효과를 표시하는 토네이도 도표를 만듭니다.

대상 대 입력의 산점도. 이 옵션은 대상 대 시뮬레이션한 입력의 산점도를 생성합니다.

대상 분포의 상자도표. 이 옵션은 대상 분포의 상자도표를 생성합니다.

사분위수 테이블. 이 옵션은 대상 분포의 사분위수 테이블을 생성합니다. 분포의 사분위수는 분포의 제 25, 50, 75 백분위수이며, 관측값을 동일한 크기의 네 그룹으로 나눕니다.

입력의 상관 및 분할표. 이 옵션은 시뮬레이션된 입력 사이의 상관계수 테이블을 표시합니다. 시뮬레이션 계획이 분할표에서 범주형 데이터를 생성하도록 지정하면 범주형 분포를 사용하는 입력 사이의 연관 분할표가 표시됩니다.

별도의 대상의 결과로부터 오버레이. 시뮬레이션하는 예측 모형에 대상이 여러 개 들어 있을 경우 별도의 대상 결과를 단일 도표에 표시할 것인지를 지정할 수 있습니다. 이 설정은 확률 밀도 함수, 누적 분포 함수 및 상자도표의 도표에 적용됩니다. 예를 들어 이 옵션을 선택한 경우 모든 대상의 확률 밀도 함수는 단일 도표에 표시됩니다.

이 시뮬레이션의 계획 저장. 시뮬레이션의 모든 수정 사항을 시뮬레이션 계획 파일에 저장할 수 있습니다. 시뮬레이션 계획 파일의 확장은 *.splan*입니다. 계획은 시뮬레이션 실행 대화 상자나 시뮬레이션 작성기에서 다시 열 수 있습니다. 시뮬레이션 계획은 결과 설정을 제외한 모든 지정 사항을 포함합니다.

시뮬레이션한 데이터를 새 데이터 파일로 저장. 시뮬레이션한 입력 값, 고정 입력 값 및 예측 대상 값을 SPSS Statistics 데이터 파일, 현재 세션의 새 데이터 세트 또는 Excel 파일에 저장할 수 있습니다. 데이터 파일의 각 케이스(또는 행)는 대상 값을 생성하는 시뮬레이션한 입력 및 고정 입력과 함께 대상의 예측값으로 구성됩니다. 민감도 분석이 지정되면 각 반복에서는 반복 수로 레이블되는 연속형 케이스 세트가 발생합니다.

여기에서 사용할 수 있는 결과 이외에 더 많은 사용자 정의 결과가 필요할 경우 시뮬레이션 작성기에서 시뮬레이션을 실행해 보십시오. 자세한 정보는 244 페이지의 『시뮬레이션 계획에서 시뮬레이션을 실행하는 방법』 주제를 참조하십시오.

시뮬레이션의 도표 결과로 작업

시뮬레이션에서 생성된 여러 도표는 대화형 기능이 있어 표시를 사용자 정의할 수 있습니다. 대화형 기능은 결과 뷰어에서 도표 오브젝트를 활성화(두 번 클릭)하여 사용할 수 있습니다. 모든 시뮬레이션 도표는 그래프보드 시각화입니다.

연속형 대상의 확률 밀도 함수 도표. 이 도표에는 도표를 별도의 영역으로 나누는 두 개의 이동 수직 참조선이 있습니다. 도표 아래의 테이블은 대상이 각 영역에 있을 확률을 표시합니다. 여러 개의 밀도 함수가 같은 도표에 표시될 경우 테이블은 각 밀도 함수와 연관된 확률에 대한 별도의 행을 가집니다. 각 참조선에는 슬라이더(역삼각형)가 있어 선을 쉽게 이동할 수 있습니다. 여러 가지 추가 기능은 도표의 **도표 옵션** 단추를 클릭하여 사용할 수 있습니다. 특히 슬라이더 위치를 명시적으로 설정하고 고정 참조선을 추가하며 도표 보기를 연속 곡선에서 히스토그램으로(그 반대도 가능) 이동할 수 있습니다. 자세한 정보는 256 페이지의 『도표 옵션』 주제를 참조하십시오.

연속형 대상의 누적 분포 함수 도표. 이 도표에는 동일한 두 개의 이동 가능한 수직 참조선과 위의 확률 밀도 함수 도표에서 설명한 연관된 테이블이 있습니다. 또한 도표 옵션 대화 상자로 액세스를 제공하여 슬라이더 위치를 명시적으로 설정하고 고정 참조선을 추가하며 누적 분포 함수를 증가하는 함수(기본값)로 표시할지 감소하는 함수로 표시지를 지정할 수 있습니다. 자세한 정보는 256 페이지의 『도표 옵션』 주제를 참조하십시오.

민감도 분석 반복이 있는 범주형 대상의 막대도표. 민감도 분석 반복이 있는 범주형 대상의 경우 예측 대상 범주의 결과는 모든 반복의 결과를 포함하는 수평배열 막대도표로 표시됩니다. 도표에는 드롭다운 목록이 있어 범주 또는 반복을 군집화할 수 있습니다. 두 단계 군집 모형 및 K-평균 군집 모형의 경우 군집 수나 반복을 군집화하도록 선택할 수 있습니다.

민감도 분석 반복이 있는 다중 대상의 상자도표. 여러 개의 연속형 대상과 민감도 분석 반복이 있는 예측 모형의 경우 단일 도표에서 모든 대상의 상자도표를 표시하도록 선택하면 군집 상자도표가 생성됩니다. 도표에는 드롭다운 목록이 있어 대상 또는 반복을 군집화할 수 있습니다.

도표 옵션

도표 옵션 대화 상자를 통해 시뮬레이션에서 생성된 확률 밀도 함수 및 누적 분포 함수의 활성화된 도표 표시를 사용자 정의할 수 있습니다.

보기. 보기 드롭다운 목록은 확률 밀도 함수 도표에만 적용됩니다. 보기를 사용하여 도표 보기를 연속 곡선에서 히스토그램으로 전환할 수 있습니다. 이 기능은 여러 개의 밀도 함수가 같은 도표에 표시되는 경우 사용할 수 없습니다. 이 경우 밀도 함수는 연속 곡선으로만 보입니다.

순서. 순서 드롭다운 목록은 누적 분포 함수 도표에만 적용됩니다. 이 목록은 누적 분포 함수를 증가하는 함수(기본값)로 표시할지 감소하는 함수로 표시할지를 지정합니다. 감소하는 함수로 표시할 경우 수평축에 주어진 점에 있는 함수 값은 대상이 해당 점의 오른쪽에 있을 확률입니다.

슬라이더 위치. 상한 및 하한 입력란에 값을 입력하여 이동 참조선의 위치를 명시적으로 설정할 수 있습니다. **-무한**을 선택하여 왼쪽 선을 제거할 수 있으며(음의 무한대로 위치 설정) **무한**을 선택하여 오른쪽 선을 제거할 수 있습니다(무한대로 위치 설정).

참조선. 다양한 고정된 수직 참조선을 확률 밀도 함수 및 누적 분포 함수에 추가할 수 있습니다. 다중 대상 또는 민감도 분석 반복계산의 결과로 인해 여러 함수가 단일 도표에 표시되는 경우 선이 적용될 특정 함수를 지정할 수 있습니다.

- **시그마.** 대상의 평균에서 지정한 수의 표준 편차를 더하고 뺀 값에 참조선을 추가할 수 있습니다.
- **백분위수.** 하단 및 상단 입력란에 값을 입력하여 대상에 대한 하나 또는 두 개의 분포 백분위수 값에 참조선을 추가할 수 있습니다. 예를 들어 상단 입력란의 값 95는 관측값의 95% 미만인 제 95 백분위수를 나타냅니다. 마찬가지로 하단 입력란의 값 5는 관측값의 5% 미만인 제 5 백분위수를 나타냅니다.
- **사용자 정의 위치.** 수평축을 따라 지정한 값에 참조선을 추가할 수 있습니다.

레이블 참조선. 이 옵션은 레이블이 선택된 참조선에 적용되는지를 제어합니다.

도표 옵션 대화 상자에서 연관된 옵션을 선택 취소하고 **계속**을 클릭하면 참조선이 제거됩니다.

지리 공간 모형화

지리 공간 모형화 기술은 지리 공간(지도) 성분을 포함하는 데이터에서 패턴을 검색하도록 설계되었습니다. 지리 공간 모형화 마법사는 시간 성분을 사용하거나 사용하지 않고 지리 공간 데이터를 분석하는 방법을 제공합니다.

이벤트와 지리 공간 데이터를 기준으로 연관 찾기(지리 공간 연관 규칙)
지리 공간 연관 규칙을 사용하면 공간 및 비공간 특성 모두를 기준으로 데이터에서 패턴을 찾을 수 있습니다. 예를 들어, 위치 및 인구 통계학적 속성에 따라 범죄 데이터의 패턴을 식별할 수 있습니다. 이러한 패턴에서 특정 유형 범죄가 발생할 가능성이 높은 지역을 예측하는 규칙을 작성할 수 있습니다.

시계열 및 지리 공간 데이터를 사용하여 예측 작성(시공간적 예측)
공간적 시간적 예측은 위치 데이터, 예측을 위한 입력 필드(예측자), 하나 이상의 시간 필드 및 목표 필드를 포함하는 데이터를 사용합니다. 각 위치에는 각 예측자 값과 각 시간 간격의 목표를 나타내는 많은 데이터 행이 있습니다.

지리 공간 모형화 마법사 사용

1. 메뉴에서 다음을 선택합니다.

   분석 > 공간 및 시간 모델링 > 공간 모델링
2. 마법사의 단계를 따릅니다.

예

도움말 시스템에서 세부 예제를 사용할 수 있습니다.

- 지리 공간 연관 규칙: **도움말 > 주제 > 사례 연구 > Statistics Base > 공간 연관 규칙**
- 공간적 시간적 예측: **도움말 > 주제 > 사례 연구 > Statistics Base > 공간적 시간적 예측**

지도 선택

지리 공간 모형화는 하나 이상의 지도 데이터 소스를 사용할 수 있습니다. 지도 데이터 소스에는 지리 영역 및 기타 지형지물(예: 길 또는 강)을 정의하는 정보가 포함되어 있습니다. 많은 지도 소스에는 인구 통계학적 또는 기타 기술통계 데이터와 범죄 보고서 또는 실업률과 같은 이벤트 데이터도 포함되어 있습니다. 이전에 정의된 지도 지정 사항 파일을 사용하거나 여기에 지도 지정 사항을 정의한 후 나중에 사용하도록 해당 지정 사항을 저장할 수 있습니다.

지도 지정 사항 로드
이전에 정의된 지도 지정 사항(.mplan) 파일을 로드합니다. 여기에 정의하는 지도 데이터 소스를 지도 지정 사항 파일에 저장할 수 있습니다. 공간적 시간적 예측의 경우 두 개 이상의 지도를 식별하는 지도 지정 사항 파일을 선택하면 파일에서 하나의 지도를 선택하도록 요구하는 프롬프트가 표시됩니다.

지도 파일 추가
ESRI 형태(.shp) 파일 또는 ESRI 형태 파일을 포함하는 .zip 아카이브를 추가합니다.

- 해당 .dbf 파일이 .shp 파일과 동일한 위치에 있어야 하고 .shp 파일과 동일한 루트 이름을 가져야 합니다.
- 파일이 .zip 아카이브인 경우 .shp 및 .dbf와 파일이 .zip 아카이브와 동일한 루트 이름을 가져야 합니다.
- 해당 투영(.prj) 파일이 없는 경우 투영 시스템을 선택하도록 프롬프트가 표시됩니다.

관계
지리 공간 연관 규칙의 경우 이 열은 이벤트가 지도의 기능과 연관되는 방식을 정의합니다. 이 설정은 시공간적 예측에 사용할 수 없습니다.

위로 이동, 아래로 이동
지도 요소의 레이어 순서는 목록에 표시된 순서에 따라 결정됩니다. 목록의 첫 번째 지도가 맨 아래 레이어입니다.

지도 선택

시공간적 예측의 경우 두 개 이상의 지도를 식별하는 지도 지정 사항 파일을 선택하면 파일에서 하나의 지도를 선택하도록 요구하는 프롬프트가 표시됩니다. 시공간적 예측은 다중 지도를 지원하지 않습니다.

지리 공간 관계

지리 공간 연관 규칙의 경우 지리 공간 관계 대화 상자는 이벤트가 지도의 기능과 연관되는 방식을 정의합니다.

- 이 설정은 지리 공간 연관 규칙에만 적용됩니다.
- 이 설정은 데이터 소스 선택 단계에서 컨텍스트 데이터로 지정되는 지도와 연관된 데이터 소스에만 영향을 줍니다.

관계

닫기
지도의 지정된 포인트 또는 영역 가까이에서 이벤트가 발생합니다.

범위 내
지도의 지정된 영역 내에서 이벤트가 발생합니다.

포함(O)
이벤트 영역에는 지도 컨텍스트 개체가 포함되어 있습니다.

교차
다른 지도의 선 또는 지역이 서로 교차하는 위치입니다.

십자
다중 지도의 경우 다른 선의 선(도로, 강, 철도)이 서로 교차하는 위치입니다.

대상의 북쪽, 대상의 남쪽, 대상의 동쪽, 대상의 서쪽
지도에 있는 지정된 포인트의 북쪽, 남쪽, 동쪽 또는 서쪽 영역 내에서 이벤트가 발생합니다.

좌표계 설정

지도에 대한 투영(.prj) 파일이 없거나 데이터 소스의 두 개 필드를 좌표 세트로 지정하는 경우 좌표계를 설정해야 합니다.

기본 지역(경도와 위도)(G)
좌표계가 경도와 위도입니다.

단순 카테시안(X 및 Y)
좌표계가 단순 X 및 Y 좌표입니다.

잘 알려진 ID 사용
공통 투영에 대한 "잘 알려진 ID"입니다.

좌표계 이름 사용
좌표계는 이름 지정된 투영을 기준으로 합니다. 이름은 소괄호로 묶습니다.

투영 설정

지도와 함께 제공된 정보를 통해 투영 시스템을 결정할 수 없는 경우 투영 시스템을 지정해야 합니다. 이 조건은 일반적으로 지도와 연관된 투영(.prj) 파일이 없거나 사용할 수 없는 투영 파일인 경우에 발생합니다.

- **구/군/시, 지역 또는 국가(메르카토르 도법)**
- **면적이 큰 국가, 여러 국가 또는 대륙(빈켈 트리펠 도법)**
- **적도와 매우 근접한 영역(메르카토르 도법)**
- **두 극지방 중 하나와 근접한 영역(평사도법)**

메르카토르 도법 투영은 여러 지도에 사용되는 공통 투영입니다. 이 투영은 지구를 평면 표면에 롤아웃된 실린더로 처리합니다. 메르카토르 도법 투영은 큰 개체의 크기와 모양을 왜곡합니다. 이 왜곡은 적도에서 더 멀리 이동하여 극지방에 더 가까워질수록 증가합니다. 빈켈 트리펠 도법 및 평사도법 투영은 지도가 두 개의 차원에 표시된 3차원 구형의 일부분을 나타낸다는 사실에 따라 조정합니다.

투영 및 좌표계

두 개 이상의 지도를 선택하고 지도에 여러 투영 및 좌표계가 있는 경우 사용할 투영 시스템이 포함된 지도를 선택해야 합니다. 투영 시스템은 지도가 결과에 함께 결합될 때 모든 지도에 사용됩니다.

데이터 소스

데이터 소스는 모양 파일과 함께 제공되는 dBase 파일, IBM SPSS 통계 데이터 파일 또는 현재 세션에 열려 있는 데이터 세트일 수 있습니다.

컨텍스트 데이터. 컨텍스트 데이터는 지도의 변수를 식별합니다. 또한 컨텍스트 데이터에는 모형에 입력으로 사용될 수 있는 필드가 포함되어 있습니다. 지도 모양(.shp) 파일과 연관된 컨텍스트 dBase(.dbf) 파일을 사용하려면 컨텍스트 dBase 파일이 모양 파일과 동일한 위치에 있어야 하고 동일한 루트 이름을 가져야 합니다. 예를 들어, 모양 파일이 `geodata.shp`인 경우 dBase 파일의 이름은 `geodata.dbf`여야 합니다.

이벤트 데이터. 이벤트 데이터에는 범죄 또는 사고와 같은 발생하는 이벤트에 대한 정보가 포함됩니다. 이 옵션은 지리 공간 연관 규칙에만 사용할 수 있습니다.

포인트 밀도. 커널 밀도 추정값에 대한 시간 간격 및 좌표 데이터입니다. 이 옵션은 공간적 시간적 예측에만 사용할 수 있습니다.

추가. 데이터 소스를 추가할 수 있는 대화 상자를 엽니다. 데이터 소스는 모양 파일과 함께 제공되는 dBase 파일, IBM SPSS 통계 데이터 파일 또는 현재 세션에 열려 있는 데이터 세트일 수 있습니다.

연관. 데이터를 지도와 연관시키는 데 사용되는 식별자(좌표 또는 키)를 지정할 수 있는 대화 상자를 엽니다. 각 데이터 소스에는 데이터를 지도와 연관시키는 하나 이상의 식별자가 포함되어야 합니다. 일반적으로 모양 파일과 함께 제공되는 dBase 파일에 자동으로 기본 식별자로 사용되는 필드가 포함됩니다. 기타 데이터 소스의 경우에는 식별자로 지정되는 필드를 지정해야 합니다.

키 검증. 지도와 데이터 소스 사이의 키 매치를 검증하기 위한 대화 상자를 엽니다.

지리 공간 연관 규칙

- 하나 이상의 데이터 소스가 이벤트 데이터 소스여야 합니다.
- 모든 이벤트 데이터 소스가 동일한 지도 연관 식별자 양식(좌표 또는 키 값)을 사용해야 합니다.

- 이벤트 데이터 소스가 키 값을 사용하여 지도와 연관된 경우 모든 이벤트 소스는 동일한 지도 기능 유형(예: 다각형, 포인트, 선)을 사용해야 합니다.

시공간적 예측

- 컨텍스트 데이터 소스가 있어야 합니다.
- 데이터 소스가 하나만 있는 경우(연관된 지도가 없는 데이터 파일) 좌표 값을 포함해야 합니다.
- 두 개의 데이터 소스가 있는 경우 하나의 데이터 소스는 컨텍스트 데이터여야 하고 다른 데이터 소스는 포인트 밀도 데이터여야 합니다.
- 세 개 이상의 데이터 소스를 포함할 수 없습니다.

데이터 소스 추가

데이터 소스는 모양 파일 및 컨텍스트 파일과 함께 제공되는 dBase 파일, IBM SPSS 통계 데이터 파일 또는 현재 세션에 열려 있는 데이터 세트일 수 있습니다.

각각에 다른 공간 연관을 사용하려는 경우 동일한 데이터 소스를 여러 번 추가할 수 있습니다.

데이터 및 지도 연관

각 데이터 소스에는 데이터를 지도와 연관시키는 하나 이상의 식별자가 포함되어야 합니다.

좌표
데이터 소스에는 카테시안 좌표를 나타내는 필드가 포함되어 있습니다. X 및 Y 좌표를 나타내는 필드를 선택하십시오. 지리 공간 연관 규칙의 경우 Z 좌표가 있을 수도 있습니다.

키 값
데이터 소스 필드의 키 값은 선택된 지도 키에 해당합니다. 예를 들어, 지역의 지도에는 각 지역의 레이블을 지정하는 이름 식별자(지도 키)가 있을 수 있습니다. 해당 식별자는 지역의 이름(데이터 키)을 포함하는 데이터의 필드에 해당합니다. 필드는 두 목록에 표시되는 순서에 따라 지도 키와 매치됩니다.

키 검증

키 검증 대화 상자는 선택된 식별자 키를 기준으로 지도와 데이터 소스 사이에 매치하는 레코드의 요약을 제공합니다. 매치하지 않는 데이터 키 값이 있는 경우 수동으로 지도 키 값에 매치할 수 있습니다.

지리 공간 연관 규칙

지리 공간 연관 규칙에 대한 지도와 데이터 소스를 정의한 후 마법사의 나머지 단계는 다음과 같습니다.

- 다중 이벤트 데이터 소스가 있는 경우 이벤트 데이터 소스를 합치는 방법을 정의합니다.
- 분석에서 조건과 예측으로 사용할 필드를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 다른 결과 옵션을 선택합니다.
- 점수화 모형 파일을 저장합니다.
- 모형에 사용된 데이터 소스의 예측값과 규칙에 대한 새 필드를 작성합니다.
- 연관 규칙을 작성하기 위한 설정을 사용자 정의합니다.
- 구간화 및 데이터 통합 설정을 사용자 정의합니다.

이벤트 데이터 필드 정의

지리 공간 연관 규칙의 경우 두 개 이상의 이벤트 데이터 소스가 있으면 이벤트 데이터 소스를 합칩니다.

- 기본적으로 모든 이벤트 데이터 소스에 공통인 필드만 포함됩니다.
- 공통 필드, 특정 데이터 소스용 필드 또는 모든 데이터 소스의 필드 목록을 표시하고 포함시킬 필드를 선택할 수 있습니다.

- 공통 필드의 경우 **유형** 및 **측정**이 모든 데이터 소스에 대해 동일해야 합니다. 충돌이 있는 경우 각 공통 필드에 사용할 유형 및 측정 수준을 지정할 수 있습니다.

필드 선택

사용가능 필드 목록에는 이벤트 데이터 소스의 필드와 컨텍스트 데이터 소스의 필드가 포함됩니다.

- **데이터 소스** 목록에서 데이터 소스를 선택하여 표시되는 필드의 목록을 제어할 수 있습니다.
- 두 개 이상의 필드를 선택해야 합니다. 하나 이상이 조건이어야 하고 하나 이상이 예측이어야 합니다. **모두(조건과 예측)** 목록에 대해 두 필드를 선택하는 것을 포함하여 이 요구사항을 충족할 수 있는 여러 방법이 있습니다.
- 연관 규칙은 조건 필드의 값을 기준으로 하는 예측 필드의 값을 예측합니다. 예를 들어, "x=1이고 y=2이면 z=3" 규칙에서는 x 및 y의 값이 조건이고 z의 값이 예측입니다.

결과

규칙 표
각 규칙 표는 신뢰도, 규칙 지지도, 리프트, 조건 지지도 및 전개능력에 대한 상위 규칙 및 값을 표시합니다. 각 표는 선택된 기준의 값에 따라 정렬됩니다. 선택된 기준에 따라 모든 규칙 또는 상위 규칙 **수**를 표시할 수 있습니다.

정렬 가능한 단어 클라우드
선택된 기준의 값에 따른 상위 규칙의 목록입니다. 텍스트의 크기는 규칙의 상대적 중요도를 나타냅니다. 대화형 결과 개체에는 신뢰도, 규칙 지지도, 리프트, 조건 지지도 및 전개능력에 대한 상위 규칙이 포함됩니다. 선택된 기준은 기본적으로 표시되는 규칙 목록을 결정합니다. 결과에서 대화형으로 다른 기준을 선택할 수 있습니다. **표시할 최대 규칙 수**는 결과에 표시되는 규칙 수를 결정합니다.

맵
선택된 기준에 따른 상위 규칙의 대화형 막대도표 및 지도입니다. 각 대화형 결과 개체에는 신뢰도, 규칙 지지도, 리프트, 조건 지지도 및 전개능력에 대한 상위 규칙이 포함됩니다. 선택된 기준은 기본적으로 표시되는 규칙 목록을 결정합니다. 결과에서 대화형으로 다른 기준을 선택할 수 있습니다. **표시할 최대 규칙 수**는 결과에 표시되는 규칙 수를 결정합니다.

모형정보 표

필드 변환
분석에 사용되는 필드에 적용되는 변환에 대해 설명합니다.

레코드 요약
포함되고 제외된 레코드의 수와 퍼센트입니다.

규칙 통계량
조건 지지도, 신뢰도, 규칙 지지도, 리프트 및 전개능력에 대한 요약 통계량입니다. 통계량에는 평균, 최소값, 최대값 및 표준편차가 포함됩니다.

최대 빈도 항목
가장 자주 발생하는 항목입니다. 항목은 규칙의 조건 또는 예측에 포함됩니다. 예를 들어, 나이 < 18 또는 성별=여성입니다.

최대 빈도 필드
규칙에서 가장 자주 발생하는 필드입니다.

제외된 입력
분석에서 제외된 필드와 각 필드가 제외된 원인입니다.

규칙 표, 단어 클라우드 및 지도의 기준

신뢰도
올바른 규칙 예측의 퍼센트입니다.

규칙 지지도
규칙이 참인 케이스의 퍼센트입니다. 예를 들어, 규칙이 "x=1이고 y=2이면 z=3"인 경우 규칙 지지도는 x=1, y=2이고 z=3인 데이터의 실제 케이스 퍼센트입니다.

리프트
리프트는 규칙이 임의 가능성에 비해 예측을 얼마나 향상시키는지에 대한 측도입니다. 이 값은 올바른 예측과 전체 예측값 발생의 비율입니다. 값은 1보다 커야 합니다. 예를 들어, 예측값이 시간의 20%로 발생하고 예측의 신뢰도가 80%인 경우 리프트 값은 4입니다.

조건 지지도
규칙 조건이 존재하는 케이스의 퍼센트입니다. 예를 들어, 규칙이 "x=1이고 y=2이면 z=3"인 경우 조건 지지도는 x=1이고 y=2인 데이터의 케이스 비율입니다.

전개능력
조건이 참일 때 올바르지 않은 예측의 퍼센트입니다. 전개능력은 (1-신뢰도)에 조건 지지도 또는 조건 지지도에서 규칙 지지도를 뺀 값을 곱한 것과 같습니다.

저장

지도와 컨텍스트 데이터를 지도 지정 사항으로 저장(M)
지도 지정 사항을 외부 파일(.mplan)에 저장합니다. 이후 분석을 위해 이 지도 지정 사항 파일을 마법사에 로드할 수 있습니다. 또한 SPATIAL ASSOCIATION RULES 명령을 통해 지도 지정 사항 파일을 사용할 수도 있습니다.

지도와 데이터 파일을 지정 사항에 복사
지도 지정 사항에 사용된 지도 모양 파일, 외부 데이터 파일 및 데이터 세트의 데이터가 지도 지정 사항 파일에 저장됩니다.

점수화
최선의 규칙 값, 규칙에 대한 신뢰도 값, 규칙에 대한 숫자 ID 값을 지정된 데이터 소스에 새 필드로 저장합니다.

스코어링할 데이터 소스
새 필드가 작성된 데이터 소스입니다. 데이터 소스가 현재 세션에서 열려 있지 않은 경우 현재 세션에서 열립니다. 새 필드를 저장하려면 수정된 파일을 명시적으로 저장해야 합니다.

목표 값
선택된 목표(예측) 필드에 대한 새 필드를 작성합니다.

- 각 목표 필드마다 두 개의 새 필드(예측값 및 신뢰도 값)가 작성됩니다.
- 연속형(척도) 목표 필드의 경우 예측값은 값 범위에 대해 설명하는 문자열입니다. "(value1, value2]" 양식의 값은 "value1보다 크고 value2보다 작거나 같음"을 의미합니다.

최선의 규칙 수
지정된 최선의 규칙 수에 대한 새 필드를 작성합니다. 각 규칙마다 세 개의 새 필드(규칙 값, 신뢰도 값 및 규칙에 대한 숫자 ID 값)가 작성됩니다.

이름 접두어
새 필드 이름에 사용할 접두문자입니다.

규칙 작성

규칙 작성 모수는 생성된 연관 규칙에 대한 기준을 설정합니다.

규칙당 항목 수
규칙 조건 및 예측에 포함될 수 있는 필드 값 수입니다. 총 항목 수가 10을 초과할 수 없습니다. 예를 들어, "x=1이고 y=2이면 z=3" 규칙에는 두 개의 조건 항목과 하나의 예측 항목이 있습니다.

최대 예측 수
규칙의 예측에서 발생할 수 있는 필드 값의 최대 수입니다.

최대 조건 수
규칙의 조건에서 발생할 수 있는 필드 값의 최대 수입니다.

대응 제외
지정된 필드 대응이 동일한 규칙에 포함되지 않도록 제외합니다.

규칙 기준

신뢰도
규칙이 결과에 포함되어야 하는 최소 신뢰도입니다. 신뢰도는 올바른 예측의 퍼센트입니다.

규칙 지지도
규칙이 결과에 포함되어야 하는 최소 규칙 지지도입니다. 이 값은 관측된 데이터에서 규칙이 참인 케이스의 퍼센트를 나타냅니다. 예를 들어, 규칙이 "x=1이고 y=2이면 z=3"인 경우 규칙 지지도는 x=1, y=2이고 z=3인 데이터의 실제 케이스 퍼센트입니다.

조건 지지도
규칙이 결과에 포함되어야 하는 최소 조건 지지도입니다. 이 값은 조건이 존재하는 케이스의 퍼센트를 나타냅니다. 예를 들어, 규칙이 "x=1이고 y=2이면 z=3"인 경우 조건 지지도는 x=1이고 y=2인 데이터의 케이스 퍼센트입니다.

리프트
규칙이 결과에 포함되어야 하는 최소 리프트입니다. 리프트는 규칙이 임의 가능성에 비해 예측을 얼마나 향상시키는지에 대한 측도입니다. 이 값은 올바른 예측과 전체 예측값 발생의 비율입니다. 예를 들어, 예측값이 시간의 20%로 발생하고 예측의 신뢰도가 80%인 경우 리프트 값은 4입니다.

동일한 것으로 처리
동일한 필드로 처리되어야 하는 대응 필드를 식별합니다.

구간화와 데이터 통합

- 지도에 있는 기능보다 데이터에 더 많은 레코드가 있는 경우 데이터 통합이 필요합니다. 예를 들어, 개별 구/군에 대한 데이터 레코드가 있지만 시/도/주에 대한 지도가 있습니다.
- 연속형 및 순서형 필드에 대한 데이터 통합 요약 측도 방법을 지정할 수 있습니다. 명목형 필드는 모달 값을 기준으로 데이터 통합됩니다.

연속형
연속형(척도) 필드의 경우 요약 측도는 평균, 중위수 또는 합입니다.

순서형
순서형 필드의 경우 요약 측도는 중위수, 최빈값, 최대값 또는 최소값입니다.

구간 수
연속형(척도) 필드의 최대 구간 수를 설정합니다. 연속형 필드는 항상 값의 범위로 그룹화되거나 "집단화"됩니다. 예를 들어, 5보다 작거나 같음, 5보다 크고 10보다 작거나 같음 또는 10보다 큼입니다.

지도 데이터 통합
데이터와 지도 모두에 데이터 통합을 적용합니다.

특정 필드의 사용자 정의 설정
특정 필드의 기본 요약 측도 및 구간 수를 대체할 수 있습니다.

- 아이콘을 클릭하여 **필드 선택기** 대화 상자를 열고 목록에 추가할 필드를 선택합니다.
- **데이터 통합** 열에서 요약 측도를 선택합니다.
- 연속형 필드에서 **구간** 열의 단추를 클릭하여 **구간** 대화 상자에 필드의 사용자 정의 구간 수를 지정합니다.

시공간적 예측

시공간적 예측의 경우 지도와 데이터 소스를 정의한 후 마법사의 나머지 단계는 다음과 같습니다.

- 목표 필드, 시간 필드 및 선택적 예측자를 지정합니다.
- 시간 필드에 대한 순환 주기 또는 시간 간격을 정의합니다.

선택적으로 다음을 수행할 수 있습니다.

- 다른 결과 옵션을 선택합니다.
- 모형설정 모수를 사용자 정의합니다.
- 통합 설정을 사용자 정의합니다.
- 예측값을 현재 세션의 데이터 세트 또는 IBM SPSS 통계 형식 데이터 파일에 저장합니다.

필드 선택

사용가능 필드 목록에는 선택된 데이터 소스의 필드가 포함됩니다. **데이터 소스** 목록에서 데이터 소스를 선택하여 표시되는 필드의 목록을 제어할 수 있습니다.

대상
목표 필드는 필수입니다. 목표는 값이 예측되는 필드입니다.

- 목표 필드는 연속형(척도) 숫자 필드여야 합니다.
- 두 개의 데이터 소스가 있는 경우 목표는 커널 밀도 추정값이고 "밀도"는 목표 이름으로 표시됩니다. 이 선택을 변경할 수 없습니다.

예측자
하나 이상의 예측자 필드를 지정할 수 있습니다. 이 설정은 선택사항입니다.

시간 필드
시간 필드를 나타내는 하나 이상의 필드를 선택하거나 **순환 주기**를 선택해야 합니다.

- 두 개의 데이터 소스가 있는 경우 두 데이터 소스 모두에서 시간 필드를 선택해야 합니다. 두 시간 필드가 동일한 구간을 나타내야 합니다.
- 순환 주기의 경우 마법사의 **시간 간격 패널**에서 주기를 정의하는 필드를 지정해야 합니다.

시간 간격

이 패널의 옵션은 필드를 선택하는 단계에서 **시간 필드** 또는 **순환 주기**를 선택하는지에 따라 결정됩니다.

시간 필드

선택된 시간 필드. 필드를 선택하는 단계에서 하나 이상의 시간 필드를 선택하면 해당 필드가 이 목록에 표시됩니다.

시간 간격. 목록에서 적절한 시간 간격을 선택하십시오. 시간 간격에 따라 관측값(증가) 및 시작값 사이의 구간과 같은 다른 설정을 지정할 수도 있습니다. 이 시간 간격은 선택된 모든 시간 필드에 사용됩니다.

- 프로시저에서는 모든 케이스(레코드)가 동일한 간격의 구간을 나타낸다고 가정합니다.
- 프로시저는 선택된 시간 간격에 따라 함께 통합되어야 하는 동일한 시간 간격에서 결측된 관측값 또는 다중 관측값을 발견할 수 있습니다. 예를 들어, 시간 간격이 일이고 날짜 2014-10-27 다음에 2014-10-29가 오는 경우 2014-10-28에 대한 결측 관측값이 있습니다. 시간 간격이 월이면 동일한 월의 여러 날짜가 함께 통합됩니다.
- 일부 시간 간격의 경우 추가 설정으로 동일한 간격의 구간에 구분을 정의할 수 있습니다. 예를 들어, 시간 간격이 일이지만 평일만 유효한 경우 한 주에 5일이 있으며 주는 월요일에 시작한다고 지정할 수 있습니다.
- 선택된 시간 필드가 날짜 형식 또는 시간 형식 필드가 아니면 자동으로 시간 간격이 **점**으로 설정되고 변경할 수 없습니다.

순환 필드

필드를 선택하는 단계에서 **순환 주기**를 선택하는 경우 순환 주기를 정의하는 필드를 지정해야 합니다. 순환 주기는 반복적인 순환 변동을 식별합니다(예: 1년의 개월 수 또는 1주의 일 수).

- 순환 주기를 정의하는 최대 3개의 필드를 지정할 수 있습니다.
- 첫 번째 순환 필드는 가장 높은 수준의 순환을 나타냅니다. 예를 들어, 연도, 분기 및 월별 순환 변동이 있는 경우 연도를 나타내는 필드가 첫 번째 순환 필드입니다.
- 첫 번째 및 두 번째 순환 필드의 순환 길이는 후속 수준의 주기입니다. 예를 들어, 순환 필드가 연도, 분기 및 월인 경우 첫 번째 순환 길이는 4이고 두 번째 순환 길이는 3입니다.
- 두 번째 및 세 번째 순환 필드의 시작값은 각 순환 주기의 첫 번째 값입니다.
- 순환 길이와 시작값은 양의 정수여야 합니다.

데이터 통합

- 필드를 선택하는 단계에서 **예측자**를 선택하는 경우 예측자에 대한 데이터 통합 요약 방법을 선택할 수 있습니다.
- 정의된 시간 간격에 두 개 이상의 레코드가 있는 경우 데이터 통합이 필요합니다. 예를 들어, 시간 간격이 월이면 동일한 월의 여러 날짜가 함께 통합됩니다.

- 연속형 및 순서형 필드에 대한 데이터 통합 요약 측도 방법을 지정할 수 있습니다. 명목형 필드는 모달 값을 기준으로 데이터 통합됩니다.

연속형
연속형(척도) 필드의 경우 요약 측도는 평균, 중위수 또는 합입니다.

순서형
순서형 필드의 경우 요약 측도는 중위수, 최빈값, 최대값 또는 최소값입니다.

특정 필드의 사용자 정의 설정
특정 예측자의 기본 데이터 통합 요약 측도를 대체할 수 있습니다.

- 아이콘을 클릭하여 **필드 선택기** 대화 상자를 열고 목록에 추가할 필드를 선택합니다.
- **데이터 통합** 열에서 요약 측도를 선택합니다.

결과

맵

목표 값
선택된 목표 필드에 대한 값의 맵입니다.

상관관계
상관관계 맵입니다.

군집
서로 유사한 위치의 군집을 강조표시하는 맵입니다. 군집 맵은 경험적 모형에만 사용할 수 있습니다.

위치 유사도 임계값
군집을 생성하는 데 필요한 유사도입니다. 값은 0보다 크고 1보다 작은 숫자여야 합니다.

최대 군집 수 지정
표시할 최대 군집 수입니다.

모형 평가 표

모형 지정 사항
목표, 입력 및 위치 필드를 포함하여 분석을 실행하는 데 사용되는 지정 사항의 요약입니다.

시간 정보 요약
모형에 사용되는 시간 필드 및 시간 간격을 식별합니다.

평균 순위 구조의 효과 검정
결과에는 모형과 각 효과에 대한 검정 통계량 값, 자유도 및 유의수준이 포함됩니다.

모형 계수의 평균 구조
결과에는 각 모형 항에 대한 계수 값, 표준오차, 검정 통계량 값, 유의수준 및 신뢰구간이 포함됩니다.

자기회귀 계수
결과에는 각 시차에 대한 계수 값, 표준오차, 검정 통계량 값, 유의수준 및 신뢰구간이 포함됩니다.

공간 공분산 검정
변동도 기반 모수 모형의 경우 공간 공분산 구조에 대한 적합도 검정 결과를 표시합니다. 검정 결과에서 공간 공분산 구조를 모수적으로 모형화할지 비모수 모형을 사용할지를 결정할 수 있습니다.

모수 공간 공분산
변동도 기반 모수 모형의 경우 모수 공간 공분산에 대한 모수 추정값을 표시합니다.

모형 옵션

모형 설정

절편 자동 포함
모형에 절편을 포함합니다.

최대 자동 선형회귀 시차
최대 자동 선형회귀 시차입니다. 이 값은 1과 5 사이의 정수여야 합니다.

공간 공분산
공간 공분산에 대한 추정 방법을 지정합니다.

모수
추정 방법이 모수입니다. 이 방법은 **가우시안**, **지수** 또는 **거듭제곱 지수**일 수 있습니다. 거듭제곱 지수의 경우 **거듭제곱** 값을 지정할 수 있습니다.

비모수
추정 방법이 비모수입니다.

저장

지도와 컨텍스트 데이터를 지도 지정 사항으로 저장(M)
지도 지정 사항을 외부 파일(.mplan)에 저장합니다. 이후 분석을 위해 이 지도 지정 사항 파일을 마법사에 로드할 수 있습니다. 또한 SPATIAL TEMPORAL PREDICTION 명령을 통해 지도 지정 사항 파일을 사용할 수도 있습니다.

지도와 데이터 파일을 지정 사항에 복사
지도 지정 사항에 사용된 지도 모양 파일, 외부 데이터 파일 및 데이터 세트의 데이터가 지도 지정 사항 파일에 저장됩니다.

점수화
목표 필드에 대한 예측값, 분산, 상한 및 하한 신뢰도를 선택된 데이터 파일에 저장합니다.

- 예측값을 현재 세션의 열려 있는 데이터 세트 또는 IBM SPSS 통계 형식 데이터 파일에 저장할 수 있습니다.
- 데이터 파일은 모형에 사용되는 데이터 소스일 수 없습니다.
- 데이터 파일에는 모형에 사용되는 모든 시간 필드와 예측자가 포함되어야 합니다.
- 시간 값은 모형에 사용된 시간 값보다 커야 합니다.

고급

결측값이 있는 케이스의 최대값(%)
결측값이 있는 케이스의 최대 퍼센트입니다.

유의수준
변동도 기반 모수 모형이 적합한지 여부를 결정하는 유의수준입니다. 값은 0보다 크고 1보다 작아야 합니다. 기본값은 0.05입니다. 유의수준은 공간 공분산 구조에 대한 적합도 검정에서 사용됩니다. 적합도 통계량은 모수 또는 비모수 모형의 사용 여부를 결정하는 데 사용됩니다.

불확실성 요인(%)
불확실성 요인은 미래 예측값의 불확실성 증가를 나타내는 퍼센트 값입니다. 예측값 불확실성의 상한과 하한은 미래의 각 단계에서 지정된 퍼센트만큼 증가합니다.

마침

지리 공간 모형화 마법사의 마지막 단계에서는 모형을 실행하거나 생성된 명령문을 명령문 창에 붙여넣을 수 있습니다. 나중에 사용할 수 있도록 생성된 명령문을 수정하고 저장할 수 있습니다.

주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다. 이 자료는 IBM에서 다른 언어로 제공할 수 있습니다. 그러나 자료에 접근하기 위해서는 해당 언어로 된 제품 또는 제품 버전의 사본이 필요할 수 있습니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않는 한, 기능상으로 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수도 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운영에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 책에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 책을 제공한다고 해서 특허에 대한 라이센스까지 부여하는 것은 아닙니다. 라이센스에 대한 의문사항은 다음으로 문의하십시오.

07326
서울특별시 영등포구
국제금융로 *10, 3IFC*
한국 아이.비.엠 주식회사
대표전화서비스*: 02-3781-7114*

2바이트(DBCS) 정보에 관한 라이센스 문의는 한국 IBM에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

Intellectual Property Licensing
2-31 Roppongi 3-chome, Minato-Ku
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 책을 "현상태대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및/또는 프로그램을 사전 통지 없이 언제든지 개선 및/또는 변경할 수 있습니다.

이 정보에서 언급되는 비IBM 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

(i) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함) 간의 정보 교환 및 (ii) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이센스 사용자는 다음 주소로 문의하십시오.

07326
서울특별시 영등포구
국제금융로 *10, 3IFC*
한국 아이.비.엠 주식회사
대표전화서비스*: 02-3781-7114*

이러한 정보는 해당 조건(예를 들면, 사용료 지불 등)하에서 사용될 수 있습니다.

이 정보에 기술된 라이센스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이센스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 프로그램 라이센스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

인용된 성능 데이터와 고객 예제는 예시 용도로만 제공됩니다. 실제 성능 결과는 특정 구성과 운영 조건에 따라 다를 수 있습니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 제품들을 테스트하지 않았으므로, 비IBM 제품과 관련된 성능의 정확성, 호환성 또는 기타 청구에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 인물 또는 기업의 이름과 유사하더라도 이는 전적으로 우연입니다.

저작권 라이센스:

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 이러한 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 애플리케이션 프로그래밍 인터페이스(API)에 부합하는 애플리케이션을 개발, 사용, 판매 또는 배포할 목적으로 IBM에 추가 비용을 지불하지 않고 이들 샘플 프로그램을 어떠한 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한 샘플 프로그램은 모든 조건하에서 완전히 테스트된 것은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 진술하지 않습니다. 본 샘플 프로그램은 일체의 보증 없이 "현상태대로" 제공됩니다. IBM은 귀하의 샘플 프로그램 사용과 관련되는 손해에 대해 책임을 지지 않습니다.

이러한 샘플 프로그램 또는 파생 제품의 각 사본이나 그 일부에는 반드시 다음과 같은 저작권 표시가 포함되어야 합니다.

© Copyright IBM Corp. 2021. 이 코드의 일부는 IBM Corp. 샘플 프로그램에서 파생되었습니다. 파생되었습니다.

© Copyright IBM Corp. 1989-2021. All rights reserved.

상표

IBM, IBM 로고 및 ibm.com은 International Business Machines Corp.의 상표 또는 등록상표입니다. 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 "저작권 및 상표 정보"(www.ibm.com/legal/copytrade.shtml)에 있습니다.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 및/또는 기타 국가에서 사용되는 Adobe Systems Incorporated의 등록상표 또는 상표입니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Intel Centrino, Intel Centrino 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation 또는 그 계열사의 상표 또는 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.

색인

숫자

0차 상관
  편상관 133
2-표본 T 검정
  독립 표본 T 검정 99
2차 모형
  곡선 추정 169
3차모형
  곡선 추정 169
4분위 맵
  최근접 이웃 분석에서 178

A

AAD(평균 절대 편차)
  비율 통계 233
Agresti-Caffo
  독립표본 비율 97
Agresti-Coull
  일표본 비율 92
Agresti-Min
  대응표본 비율 94
Akaike 정보 기준
  선형모형에서 139
Anderson-Rubin 요인 스코어 184
Andrews의 웨이브 추정량
  데이터 탐색 80
Anscombe
  일표본 비율 92

B

Bartlett 요인 스코어 184
Bartlett의 단위행렬 검정
  요인 분석 182
Bonett-Price
  대응표본 비율 94
Bonferroni
  일원 분산 분석 104
  GLM 112
Box의 M 검정
  판별 분석 179
Brown-Forsythe 통계
  일원 분산 분석 105
Brown-Li-Jeffreys
  독립표본 비율 97

C

Chebychev 거리
  거리 측도 134
ChiSquare 117, 118
Clopper-Pearson(정확)
  일표본 비율 92
Cochran의 통계
Cochran의 통계 (계속)
  교차 분석표 83
Cochran의 Q
  대응 K-표본 비모수 검정 213
Cochran의 Q 검정
  대응표본 비모수 검정 202, 203
COD(산포계수)
  비율 통계 233
Cohen의 가중 카파 224, 225
Cohen의 kappa
  교차 분석표 83
Cook의 거리
  선형 회귀분석 144
  GLM 114
COV(변동 계수)
  비율 통계 233
Cox 및 Snell R2
  순서 회귀분석 147
Cramér의 V
  교차 분석표 83
Cronbach의 알파
  신뢰성 분석 220, 221

D

d
  교차 분석표 83
DfBeta
  선형 회귀분석 144
DfFit
  선형 회귀분석 144
Duncan의 다중 범위 검정
  일원 분산 분석 104
  GLM 112
Dunnett의 C
  일원 분산 분석 104
  GLM 112
Dunnett의 T 검정
  일원 분산 분석 104
  GLM 112
Dunnett의 T3
  일원 분산 분석 104
  GLM 112
Durbin-Watson 통계
  선형 회귀분석 145

F

F 통계
  선형모형에서 139
Fisher의 정확 검정
  교차 분석표 83
Fisher의 LSD
  GLM 112
Fleiss의 다중 평가자 카파 220, 221
Friedman 검정
  대응 K-표본 비모수 검정 213

Friedman 검정 (계속)
대응표본 비모수 검정 202

G

Gabriel의 쌍대 비교 검정
일원 분산 분석 104
GLM 112
Galbraith 도표 66
Galbraith 도표 탭
메타분석 연속형 70
메타분석 연속형 효과 크기 70
메타분석 이분형 70
메타분석 이분형 효과 크기 70
Games-Howell의 대응 비교 검정
일원 분산 분석 104
GLM 112
GLM
모델 108
변수 저장 114
사후 검정 112
제곱합 108
프로파일 도표 110
행렬 저장 114
GLM 일변량
대비 109
Goodman과 Kruskal의 감마
교차 분석표 83
Goodman과 Kruskal의 람다
교차 분석표 83
Goodman과 Kruskal의 타우
교차 분석표 83
Guttman 모형
신뢰성 분석 220, 221

H

Hampel의 redescending M 추정량
데이터 탐색 80
Hauck-Anderson
독립표본 비율 97
Helmert 대비
GLM 109
Hochberg의 GT2
일원 분산 분석 104
GLM 112
Hotelling의 T 2
신뢰성 분석 220, 221
Huber의 M-추정량
데이터 탐색 80

I

ICC. 급내 상관계수 참조 221
ICC(급내 상관계수)
신뢰성 분석 221

K

k 및 필드선택
최근접 이웃 분석에서 178
k 선택
최근접 이웃 분석에서 178
K-대응 표본 비모수검정
검정 유형 213
명령 추가 기능 213
통계 213
K-독립 표본 비모수검정
검정 유형 212
결측값 212
그룹화 변수 212
명령 추가 기능 212
범위지정 212
옵션 212
통계 212
K-Means 군집 분석
개요 193
결측값 195
군집 거리 194
군집 정보 저장 194
명령 추가 기능 195
반복 194
방법 193
소속군집 194
수렴 기준 194
예제 193
통계 193, 195
효율 194
kappa
교차 분석표 83
Kendall의 타우-b
교차 분석표 83
이변량 상관 118
Kendall의 타우-c
교차 분석표 83
Kendall의 W
대응 K-표본 비모수 검정 213
Kolmogorov-Smirnov의 Z
독립 2-표본 비모수 검정 209
일표본 Kolmogorov-Smirnov 검정 207
KR20
신뢰성 분석 221
Kruskal-Wallis의 H
독립 2-표본 비모수 검정 211
Kruskal의 타우
교차 분석표 83
Kuder-Richardson 20(KR20)
신뢰성 분석 221

L

L'Abb´e 도표 66
L'Abb´e 도표 탭
메타분석 연속형 71
메타분석 연속형 효과 크기 71
메타분석 이분형 71
메타분석 이분형 효과 크기 71
Lance와 Williams 상이성 측도
거리 측도 134
Levene 검정
데이터 탐색 80
일원 분산 분석 105
Lilliefors 검정
데이터 탐색 80

M

M-추정량
  데이터 탐색 80
Mahalanobis의 거리
  선형 회귀분석 144
  판별 분석 180
Manhattan 거리
  최근접 이웃 분석에서 173
Mann-Whitney의 U
  독립 2-표본 비모수 검정 209
Mantel-Haenszel 통계
  교차 분석표 83
McDonald의 오메가
  신뢰성 분석 220, 221
McFadden R2
  순서 회귀분석 147
McNemar
  대응표본 비율 95
McNemar 검정
  교차 분석표 83
  대응 2-표본 비모수 검정 210
  대응표본 비모수 검정 202, 203
McNemar(연속수정)
  대응표본 비율 95
Minkowski 거리
  거리 측도 134
Monte Carlo 시뮬레이션 242
Moses의 극단값반동 검정
  독립 2-표본 비모수 검정 209

N

Nagelkerke R2
  순서 회귀분석 147
Newcombe
  대응표본 비율 94
  독립표본 비율 97
Newcombe(연속수정)
  독립표본 비율 97
Newman-Keuls
  GLM 112

O

OLAP 큐브
  제목 90
  통계 89

P

P-P 확률 도표 234
Pearson 상관
  교차 분석표 83
  이변량 상관 118
Pearson 잔차
  순서 회귀분석 147
Pearson 카이제곱
  교차 분석표 83
  순서 회귀분석 147
PLUM
  순서 회귀분석 146
PRD(가격 관련 차)
PRD(가격 관련 차) *(계속)*
  비율 통계 233
PROXMAP 분석 121

Q

Q-Q 확률 도표 236

R

R 2
  선형 회귀분석 145
  평균 87
  R 2 변화량 145
r 상관계수
  교차 분석표 83
  이변량 상관 118
R 통계
  선형 회귀분석 145
  평균 87
R-제곱
  선형모형에서 140
R-E-G-W의 F
  일원 분산 분석 104
  GLM 112
R-E-G-W의 Q
  일원 분산 분석 104
  GLM 112
Rao의 V
  판별 분석 180
rho
  교차 분석표 83
  이변량 상관 118
ROC 곡선
  통계량 및 도표 242
ROC 분석
  통계량 및 도표 240
Ryan-Einot-Gabriel-Welsch 다중 범위
  일원 분산 분석 104
  GLM 112
Ryan-Einot-Gabriel-Welsch 다중 F
  일원 분산 분석 104
  GLM 112

S

S-모형
  곡선 추정 169
S-스트레스
  다차원척도법 225
Scheffé 검정
  일원 분산 분석 104
  GLM 112
Shapiro-Wilk의 검정
  데이터 탐색 80
Sidak의 T 검정
  일원 분산 분석 104
  GLM 112
Somers의 d
  교차 분석표 83
Spearman 상관계수
  교차 분석표 83
  이변량 상관 118

Spearman-Brown 신뢰도
신뢰성 분석 221
Student-Newman-Keuls
일원 분산 분석 104
GLM 112

T

T 검정
대응표본 T 검정 100
독립 표본 T 검정 99
일표본 T 검정 101
Tamhane의 T2
일원 분산 분석 104
GLM 112
TSF 167, 168
Tukey의 가법성 검정
신뢰성 분석 220, 221
Tukey의 정직유의차
일원 분산 분석 104
GLM 112
Tukey의 B 검정
일원 분산 분석 104
GLM 112
Tukey의 Biweight 추정량
데이터 탐색 80

V

V
교차 분석표 83

W

Wald
대응표본 비율 94, 95
독립표본 비율 97
일표본 비율 92
Wald H0
독립표본 비율 97
Wald H0(연속수정)
독립표본 비율 97
Wald-Wolfowitz 런검정
독립 2-표본 비모수 검정 209
Wald(연속수정)
대응표본 비율 94, 95
독립표본 비율 97
일표본 비율 92
Waller-Duncan T 검정
일원 분산 분석 104
GLM 112
Welch 통계
일원 분산 분석 105
what-if 분석
시뮬레이션에서 250
Wilcoxon 부호순위 검정
대응 2-표본 비모수 검정 210
대응표본 비모수 검정 202
일표본 비모수 검정 196
Wilks의 람다
판별 분석 180
Wilson Score
일표본 비율 92
Wilson Score(연속수정)
일표본 비율 92

X

X^2 117, 118

Y

Yates의 연속성 보정
교차 분석표 83

Z

z 점수
기술통계 77
변수로 저장 77

가

가중 예측값
GLM 114
가중 최소제곱법
선형 회귀분석 142
가중 카파
교차표 225
기준 225
예 224
인쇄 225
통계 224, 225
가중 평균
비율 통계 233
가중치를 부여하지 않은 최소제곱법
요인 분석 182
감마
교차 분석표 83
거듭제곱 모형
곡선 추정 169
거듭제곱 분석
통계 1
거리
값 변환 134
명령 추가 기능 135
변수 간 거리 계산 134
상이성 측도 134
예 134
유사성 측도 134
측도변환 134
케이스 간 거리 계산 134
통계 134
거리 상관 135
거리 측도
거리 측도 134
계층적 군집 분석 192
최근접 이웃 분석에서 173
거품 도표
메타분석 회귀분석 68
거품 도표 탭
메타분석 연속형 68
메타분석 연속형 효과 크기 68
메타분석 이분형 68
메타분석 이분형 효과 크기 68
검증용 표본

검증용 표본 *(계속)*
최근접 이웃 분석에서 174
결측값
다중반응 교차 분석표 216
다중반응 빈도분석 214
대응 2-표본 비모수 검정 211
대응표본 비율 95
대응표본 T 검정 101
데이터 탐색 81
독립 2-표본 비모수 검정 210
독립 표본 T 검정 100
독립 K-표본 비모수 검정 212
독립표본 비율 98
런 검정 207
백분위수 79
선형 회귀분석 145
열별 요약 보고서 220
요인 분석 184
이변량 상관 120
이항검정 206
일원 분산 분석 105
일표본 비율 93
일표본 Kolmogorov-Smirnov 검정 208
일표본 T 검정 102
최근접 이웃 분석에서 175
카이제곱 검정 205
편상관 133
행별 요약 보고서 218
ROC 곡선 242
ROC 분석에서 240
계층적 군집 분석
값 변환 192
거리 측도 192
거리 행렬 192
고드름 도표 193
군집 변수 191
군집 케이스 191
군집방법 192
군집화 스케줄 192
덴드로그램 193
도표 출력방향 193
명령 추가 기능 193
새 변수 저장 193
소속군집 192, 193
예 191
유사성 측도 192
측도변환 192
통계 191, 192
계층적 분해 108
고드름 도표
계층적 군집 분석 193
고유 값
선형 회귀분석 145
요인 분석 182
곡선 추정
모델 169
분산 분석 169
상수 포함 169
예측 170
예측 구간 저장 170
예측값 저장 170
잔차 저장 170
곱셈
보고서 열 곱하기 219
공간 관계 121
공간 모형화 256
공분산 비율
선형 회귀분석 144
공분산 행렬
선형 회귀분석 145
순서 회귀분석 147
판별 분석 179, 180
GLM 114
공선성 진단 정보
선형 회귀분석 145
공차
선형 회귀 145
과적합 방지 기준
선형모형에서 139
관측개수
교차 분석표 84
관측빈도
순서 회귀분석 147
교차분석
교차 분석표 출력않음 81
레이어 82
셀 표시 84
수평배열 막대도표 82
제어변수 82
통계 83
형식 85
교차표
교차 분석표 81
다중 응답 215
군집 분석
계층적 군집 분석 191
효율 194
K-Means 군집 분석 193
군집 뷰어
개요 188
군집 모형 정보 187
군집 및 기능 이동 189
군집 및 기능 전치 189
군집 보기 188
군집 비교 보기 190
군집 예측자 중요도 보기 190
군집 정렬 189
군집 출력결과 정렬 189
군집 크기 보기 190
군집의 비교 190
군집의 크기 190
군집중심 보기 188
기능 정렬 189
기능 출력결과 정렬 189
기본 보기 189
레코드 필터링 191
모형 요약 188
사용 190
셀 내용 정렬 189
셀 내용 출력결과 189
셀 분포 보기 190
셀의 분포 190
예측변수 중요도 190
요약 보기 188
군집 빈도
이단계 군집 분석 187
군집화
군집 보기 188

군집화 (계속)
전체 출력결과 188
프로시저 선택 184
그룹 간의 차이
OLAP 큐브 90
그룹 비교
OLAP 큐브 90
그룹 중위수
요약 86
평균 87
OLAP 큐브 89
그룹 평균 87, 88
극단값
데이터 탐색 80
근사 유의수준 221
근접 121
근접 매핑 121, 123
근접도
계층적 군집 분석 191
기능 공간 도표
최근접 이웃 분석에서 176
기대개수
교차 분석표 84
기대빈도
순서 회귀분석 147
기술통계
명령 추가 기능 78
출력 순서 77
통계 77
z 점수 저장 77
기술통계량
기술통계 77
데이터 탐색 80
비율 통계 233
빈도분석 75
요약 86
이단계 군집 분석 187
기준
in 블랜드 알트만 117
기준 대화 상자
메타분석 연속형 33, 41
메타분석 이분형 48
메타분석 이분형 효과 크기 56
메타분석 회귀분석 63
기하평균
요약 86
평균 87
OLAP 큐브 89
깔때기 도표 66
깔때기 도표 탭
메타분석 연속형 69
메타분석 연속형 효과 크기 69
메타분석 이분형 69
메타분석 이분형 효과 크기 69

나

나누기
보고서 열 나누기 219
나무 깊이
이단계 군집 분석 186
누적 분포 함수
시뮬레이션에서 251
누적 빈도
누적 빈도 (계속)
순서 회귀분석 147
누적 포리스트 도표 66
누적 포리스트 도표 탭
메타분석 연속형 67
메타분석 연속형 효과 크기 67
메타분석 이분형 67
메타분석 이분형 효과 크기 67

다

다중 비교
일원 분산 분석 104
다중 응답
명령 추가 기능 216
다중 응답 분석
교차표 215
다중반응 교차분석 215
다중반응 빈도분석 214
빈도표 214
다중 응답 세트
코드북 72
다중 응답 세트 정의
범주 214
변수 세트 레이블 214
변수 세트 이름 214
이분형 214
다중 회귀분석
선형 회귀분석 142
다중 R
선형 회귀분석 145
다중반응 교차분석
값 범위지정 215
결측값 216
반응 퍼센트 계산 기준 216
반응변수 세트의 일치 변수 216
셀 퍼센트 216
응답 퍼센트 계산 기준 216
다중반응 빈도분석
결측값 214
다차원척도법
값 변환 227
거리 측도 227
거리 행렬 만들기 227
기준 227
데이터 형태 정의 226
명령 추가 기능 227
예 225
조건부 227
차원 227
척도화 모형 227
출력 옵션 227
측정 수준 227
통계 225
다항 대비
일원 분산 분석 103
GLM 109
단계 선택
선형 회귀분석 143
단계적 전진
선형모형에서 139
단순 대비
GLM 109
대비

대비 (계속)
일원 분산 분석 103
GLM 109
대비 대화 상자
메타분석 연속형 36
메타분석 연속형 효과 크기 43
메타분석 이분형 51
메타분석 이분형 효과 크기 58
대응 2-표본 비모수 검정
검정 유형 210
결측값 211
명령 추가 기능 211
옵션 211
통계 211
대응 연구
대응표본 T 검정 100
대응 표본 210, 212
대응표본 비모수 검정
필드 202
Cochran의 Q 검정 203
McNemar 검정 203
대응표본 비율 93
대응표본 *t* 검정 100
대응표본 T 검정
결측값 101
대응변수 선택 100
옵션 101
덴드로그램
계층적 군집 분석 193
도시-블록 거리
최근접 이웃 분석에서 173
도표
메타분석 66
케이스 레이블 169
확률 도표 234, 236
ROC 곡선 241
ROC 분석에서 239
도표 대화 상자
메타분석 연속형 39
메타분석 연속형 효과 크기 46
메타분석 이분형 55
메타분석 이분형 효과 크기 62
메타분석 회귀분석 66
독립 2-표본 비모수 검정
검정 유형 209
결측값 210
그룹 정의 210
그룹화 변수 210
명령 추가 기능 210
옵션 210
통계 210
독립 표본 비모수 검정
필드 탭 200
독립 표본 T 검정
결측값 100
신뢰구간 100
옵션 100
독립 T 검정
대응표본 T 검정 100
독립성 검정
카이제곱 83
독립표본 비율 96, 97
독립표본 T 검정
그룹 정의 100
독립표본 T 검정 (계속)
그룹화 변수 100
문자열 변수 100
동형 모형
신뢰성 분석 220, 221
디트렌딩된 정규도표
데이터 탐색 80

라

람다
교차 분석표 83
런 검정
결측값 207
명령 추가 기능 207
옵션 207
일표본 비모수 검정 196, 197
절단점 206, 207
통계 207
레버리지 값
선형 회귀분석 144
GLM 114
레이어
교차 분석표 82
로그 모형
곡선 추정 169
로딩 도표
요인 분석 183
로지스틱모형
곡선 추정 169
로짓
일표본 비율 92
링크
순서 회귀분석 147

마

마지막
요약 86
평균 87
OLAP 큐브 89
막대도표
빈도분석 76
메모리 할당
이단계 군집 분석 186
메서드 비교
in 블랜드 알트만 116
메타분석
도표 66
연속형 결과 32
이분형 결과 32
회귀분석 32
메타분석 연속형
거품 도표 탭 68
기준 대화 상자 33, 41
깔때기 도표 탭 69
누적 포리스트 도표 탭 67
대비 대화 상자 36
도표 대화 상자 39
분석 대화 상자 34
인쇄 대화 상자 38
자르기 및 채우기 대화 상자 37
저장 대화 상자 39

메타분석 연속형 (계속)
추론 대화 상자 35
편향 대화 상자 36
포리스트 도표 탭 66
Galbraith 도표 탭 70
L'Abb´e 도표 탭 71
메타분석 연속형 효과 크기
거품 도표 탭 68
깔때기 도표 탭 69
누적 포리스트 도표 탭 67
대비 대화 상자 43
도표 대화 상자 46
분석 대화 상자 41
인쇄 대화 상자 45
자르기 및 채우기 대화 상자 44
저장 대화 상자 46
추론 대화 상자 42
편향 대화 상자 43
포리스트 도표 탭 66
Galbraith 도표 탭 70
L'Abb´e 도표 탭 71
메타분석 이분형
거품 도표 탭 68
기준 대화 상자 48
깔때기 도표 탭 69
누적 포리스트 도표 탭 67
대비 대화 상자 51
도표 대화 상자 55
분석 대화 상자 49
인쇄 대화 상자 53
자르기 및 채우기 대화 상자 52
저장 대화 상자 54
추론 대화 상자 50
편향 대화 상자 51
포리스트 도표 탭 66
Galbraith 도표 탭 70
L'Abb´e 도표 탭 71
메타분석 이분형 효과 크기
거품 도표 탭 68
기준 대화 상자 56
깔때기 도표 탭 69
누적 포리스트 도표 탭 67
대비 대화 상자 58
도표 대화 상자 62
분석 대화 상자 57
인쇄 대화 상자 61
자르기 및 채우기 대화 상자 59
저장 대화 상자 61
추론 대화 상자 57
편향 대화 상자 59
포리스트 도표 탭 66
Galbraith 도표 탭 70
L'Abb´e 도표 탭 71
메타분석 회귀분석
거품 도표 68
기준 대화 상자 63
도표 대화 상자 66
인쇄 대화 상자 65
저장 대화 상자 65
추론 대화 상자 64
모드
빈도분석 75
모수 추정값
순서 회귀분석 147
모형 보기
최근접 이웃 분석에서 176
민감도 분석
시뮬레이션에서 250

바

반복
요인 분석 182, 183
K-Means 군집 분석 194
반복 대비
GLM 109
반복 히스토리
순서 회귀분석 147
반분계수 신뢰도
신뢰성 분석 220, 221
배깅
선형모형에서 138
백분위수
데이터 탐색 80
빈도분석 75
시뮬레이션에서 252
범위
기술통계 77
비율 통계 233
빈도분석 75
요약 86
평균 87
OLAP 큐브 89
베리멕스 회전
요인 분석 183
베타 계수
선형 회귀분석 145
변수 간의 차이
OLAP 큐브 90
변수 비교
OLAP 큐브 90
변수 중요도
최근접 이웃 분석에서 177
변환 행렬
요인 분석 181
보고서
열 값 곱하기 219
열 값 나누기 219
열 비교 219
열별 요약 보고서 218
총계 열 219
행별 요약 보고서 216
혼합 전체 219
보조 회귀 모형
GLM 115
복합모형
곡선 추정 169
부스팅
선형모형에서 138
부호 검정
대응 2-표본 비모수 검정 210
대응표본 비모수 검정 202
분류
ROC 곡선 241
ROC 분석에서 239
분류표
최근접 이웃 분석에서 178
분산

분산 (계속)
기술통계 77
데이터 탐색 80
빈도분석 75
열별 요약 보고서 219
요약 86
평균 87
행별 요약 보고서 217
OLAP 큐브 89
분산 동질성 검정
일원 분산 분석 105
분산 분석
곡선 추정 169
모델 108
선형 회귀분석 145
선형모형에서 141
일원 분산 분석 103
평균 87
GLM 일변량 106
분산팽창요인
선형 회귀분석 145
분석 대화 상자
메타분석 연속형 34
메타분석 연속형 효과 크기 41
메타분석 이분형 49
메타분석 이분형 효과 크기 57
분포 맞춤
시뮬레이션에서 247
분포 측도
기술통계 77
빈도분석 75
분할계수
교차 분석표 83
분할표 81
불확실성 계수
교차 분석표 83
블랜드 알트만
계약 분석 116, 117
블록 거리
거리 측도 134
비모수 검정
대응 2-표본 비모수 검정 210
독립 2-표본 비모수 검정 209
런 검정 206
일표본 Kolmogorov-Smirnov 검정 207
카이제곱 204
K-대응 표본 비모수검정 212
K-독립 표본 비모수검정 211
비선형 135
비율 통계량
통계 233
빈도
교차 분석표 출력않음 76
도표 76
출력 순서 76
통계 75
형식 76
빈도표
데이터 탐색 80
빈도분석 74

사

사례 통제 연구
사례 통제 연구 (계속)
대응표본 T 검정 100
사분위수
빈도분석 75
사용자 정의 모형
GLM 108
사전
코드북 72
사후분석 다중 비교 104
삭제된 잔차
선형 회귀분석 144
GLM 114
산점도
선형 회귀분석 143
시뮬레이션에서 252
산포 측도
기술통계 77
데이터 탐색 80
비율 통계 233
빈도분석 75
상관
0차 133
교차 분석표 83
시뮬레이션에서 250
이변량 상관 118
편상관 132
상관 행렬
순서 회귀분석 147
요인 분석 181, 182
판별 분석 179
상대적 위험도
교차 분석표 83
상자도표
데이터 탐색 80
변수 비교 80
시뮬레이션에서 252
요인 수준 비교 80
상호작용 항 108, 148, 149
선택 기준
선형 회귀분석 145
선택변수
선형 회귀분석 143
선형 135
선형 라소 회귀분석
변수 152
옵션 153
파티션 152
선형 리지 회귀분석
변수 155
옵션 155
파티션 155
선형 탄력적 순 회귀
변수 149
옵션 150
파티션 149
선형 회귀
가중치 142
결측값 145
도표 143
명령 추가 기능 146
모형 내보내기 정보 144
변수 선택 방법 143, 145
블록 142
새 변수 저장 144

선형 회귀 (계속)
선택변수 143
잔차 144
통계 145
선형모형
결과 복제 139
계수 141
곡선 추정 169
관찰에 의한 예측 140
모형 선택 139
모형 옵션 140
모형 요약 140
모형 작성 요약 142
목표 138
분산분석표(A) 141
신뢰수준 138
앙상블 139
예측변수 중요도 140
이상치 141
자동 데이터 준비 138, 140
잔차 140
정보 기준 140
조합 규칙 139
평균 추정 142
R-제곱 통계 140
선형별 연관
교차 분석표 83
선형성 검정
평균 87
성장모형
곡선 추정 169
소계
열별 요약 보고서 219
수렴
요인 분석 182, 183
K-Means 군집 분석 194
수정된 R 2
선형 회귀분석 145
수정된 R-제곱
선형모형에서 139
순서 회귀분석
링크 147
명령 추가 기능 149
옵션 147
위치 모형 148
척도 모형 148
통계 146
순위 상관계수
이변량 상관 118
스튜던트 t 검정 99
스튜던트화 잔차
선형 회귀분석 144
스트레스
다차원척도법 225
시각화
모형 군집화 188
시계열 필터
박스터 킹 필터 167, 168
크리스티아노 피츠제럴드 필터 167, 168
호드릭 프레스콧 필터 167, 168
시계열도표 분석
예측 170
예측할 케이스 170
시뮬레이션
시뮬레이션 (계속)
결과 251, 252
꼬리 표본추출 250
누적 분포 함수 251
대상 및 입력 형식 표시 252
대상 분포의 백분위수 252
대화형 도표 255
도표 옵션 256
모형 사양 245
민감도 분석 250
방정식 편집기 246
분포 맞춤 247
분포 맞춤 결과 249
분포 맞춤 사용자 정의 249
분포를 새 데이터에 다시 맞춤 253
산점도 252
상자도표 252
새 입력 만들기 247
시뮬레이션 계획 만들기 243, 244
시뮬레이션 계획 실행 244, 253
시뮬레이션 계획 저장 253
시뮬레이션 작성기 245
시뮬레이션한 데이터 저장 253
입력 사이의 상관 250
중지 기준 250
지원하는 모형 245
토네이도 도표 252
확률 밀도 함수 251
what-if 분석 250
시뮬레이션 작성기 245
신뢰구간
대응표본 T 검정 101
데이터 탐색 80
독립 표본 T 검정 100
선형 회귀분석 145
선형 회귀분석에 저장 144
이변량 상관 118, 120
일원 분산 분석 105
일표본 T 검정 102
GLM 109
ROC 곡선 242
ROC 분석에서 240
신뢰성 분석
급내 상관계수 221
기술통계 221
명령 추가 기능 223
분산분석표(A) 221
예 220
통계 220, 221
항목 간 상관 및 공분산 221
Hotelling의 T 2 221
Kuder-Richardson 20 221
Tukey의 가법성 검정 221

아

알파 계수
신뢰성 분석 220, 221
알파 요인 추출 182
앙상블
선형모형에서 139
에타
교차 분석표 83
평균 87

에타제곱
  평균 87
역 모형
  곡선 추정 169
연구 조사 117, 118
연속형 결과
  메타분석 32
  사전 계산된 효과 크기 데이터 32
  원시 효과 크기 데이터 32
열 백분율
  교차 분석표 84
열 비율 통계
  교차 분석표 84
열별 요약 보고서
  결측값 220
  명령 추가 기능 220
  소계 219
  열 형식 217
  쪽 레이아웃 218
  쪽 매김 220
  쪽 제어 219
  총계 220
  총계 열 219
예측
  곡선 추정 170
예측 구간
  곡선 추정으로 저장 170
  선형 회귀분석에 저장 144
예측값
  곡선 추정으로 저장 170
  선형 회귀분석에 저장 144
예측변수 중요도
  선형모형 140
오차 요약
  최근접 이웃 분석에서 178
완전 요인 모형
  GLM 108
왜도
  기술통계 77
  데이터 탐색 80
  빈도분석 75
  열별 요약 보고서 219
  요약 86
  평균 87
  행별 요약 보고서 217
  OLAP 큐브 89
왜도의 표준 오차
  요약 86
  평균 87
  OLAP 큐브 89
요약
  옵션 85
  통계 86
요인 분석
  개요 181
  결측값 184
  계수표시 형식 184
  기술통계 182
  로딩 도표 183
  명령 추가 기능 184
  수렴 182, 183
  예 181
  요인 스코어 184
  추출 방법 182
요인 분석 (계속)
  케이스 선택 182
  통계 181, 182
  회전 방법 183
요인 스코어 184
우도비 구간
  일표본 비모수 검정 196
우도비 카이제곱
  교차 분석표 83
  순서 회귀분석 147
원형 도표
  빈도분석 76
위치 모형
  순서 회귀분석 148
위험도
  교차 분석표 83
유사성 측도
  거리 측도 134
  계층적 군집 분석 192
유클리드 거리
  거리 측도 134
  최근접 이웃 분석에서 173
유클리드 거리 제곱값
  거리 측도 134
이단계 군집 분석
  옵션 186
  외부 파일에 저장 187
  작업 파일에 저장 187
  통계 187
이미지 요인 추출 182
이변량 상관계수
  결측값 120
  명령 추가 기능 120
  상관계수 118
  신뢰구간 118, 120
  옵션 120
  유의수준 118
  통계 120
이분산성 검정의 디자인
  GLM 115
이분형 결과
  사전 계산된 효과 크기 데이터 32
  원시 효과 크기 데이터 32
이상치
  데이터 탐색 80
  선형 회귀분석 143
  이단계 군집 분석 186
이쿼멕스 회전
  요인 분석 183
이항검정
  결측값 206
  명령 추가 기능 206
  옵션 206
  이분형 205
  일표본 비모수 검정 196
  통계 206
인쇄
  in 블랜드 알트만 117
인쇄 대화 상자
  메타분석 연속형 38
  메타분석 연속형 효과 크기 45
  메타분석 이분형 53
  메타분석 이분형 효과 크기 61
  메타분석 회귀분석 65

일반화 최소제곱법
요인 분석 182
일원 분산 분석
결측값 105
다중 비교 104
다항 대비 103
대비 103
명령 추가 기능 106
사후 검정 104
옵션 105
요인변수 103
통계 105
일표본 비모수 검정
런 검정 197
이항검정 196
카이제곱 검정 197
콜모고로프-스미르노프 검정 197
필드 196
일표본 비율 91, 92, 94
일표본 Kolmogorov-Smirnov 검정
검정 분포 207
결측값 208
명령 추가 기능 209
옵션 208
통계 208
Lilliefors 검정 207, 208
일표본 T 검정
결측값 102
명령 추가 기능 101, 102
신뢰구간 102
옵션 102
일표본 Wilcoxon 검정 199

자

자동 데이터 준비
선형모형에서 140
자동 분포 맞춤
시뮬레이션에서 247
자르기 및 채우기 대화 상자
메타분석 연속형 37
메타분석 연속형 효과 크기 44
메타분석 이분형 52
메타분석 이분형 효과 크기 59
잔차
곡선 추정으로 저장 170
교차 분석표 84
선형 회귀분석에 저장 144
잘린 평균
데이터 탐색 80
잡음 처리
이단계 군집 분석 186
저장 대화 상자
메타분석 연속형 39
메타분석 연속형 효과 크기 46
메타분석 이분형 54
메타분석 이분형 효과 크기 61
메타분석 회귀분석 65
적합도
순서 회귀분석 147
전진 선택법
선형 회귀분석 143
최근접 이웃 분석에서 174
전체 퍼센트
전체 퍼센트 (계속)
교차 분석표 84
절대 동형 모형
신뢰성 분석 220, 221
점수 92
점수(연속수정) 92
정규 확률 도표
데이터 탐색 80
선형 회귀분석 143
P-p 도표 234
Q-Q 236
정규성 검정
데이터 탐색 80
정보 기준
선형모형에서 139
정확검정 이항
대응표본 비율 95
제곱합
GLM 108
제목
OLAP 큐브 90
제어변수
교차 분석표 82
제프리스
일표본 비율 92
제프리스 구간
일표본 비모수 검정 196
조합 규칙
선형모형에서 139
조화 평균
요약 86
평균 87
OLAP 큐브 89
주변 동질성 검정
대응 2-표본 비모수 검정 210
대응표본 비모수 검정 202
주성분 분석 181, 182
주축 요인 추출 182
줄기와 잎그림 도표
데이터 탐색 80
중간 확률 수정된 이항
대응표본 비율 95
중심 경향 측도
데이터 탐색 80
비율 통계 233
빈도분석 75
중위수
데이터 탐색 80
비율 통계 233
빈도분석 75
요약 86
평균 87
OLAP 큐브 89
중위수 검정
독립 2-표본 비모수 검정 211
지리 공간 모형화 256–265
지수모형
곡선 추정 169
직접 오블리민 회전
요인 분석 183
집중 지수
비율 통계 233
쪽 매김
열별 요약 보고서 220

쪽 매김 (계속)
행별 요약 보고서 218
쪽 제어
열별 요약 보고서 219
행별 요약 보고서 218

차

차이 대비
GLM 109
참조 범주
GLM 109
척도
가중 카파에서 224
다차원척도법 225
신뢰성 분석 220
척도 모형
순서 회귀분석 148
첨도
기술통계 77
데이터 탐색 80
빈도분석 75
열별 요약 보고서 219
요약 86
평균 87
행별 요약 보고서 217
OLAP 큐브 89
첨도의 표준 오차
요약 86
평균 87
OLAP 큐브 89
첫 번째
요약 86
평균 87
OLAP 큐브 89
초기 임계값
이단계 군집 분석 186
총계
열별 요약 보고서 220
총계 열
보고서 219
최근접 이웃 거리
최근접 이웃 분석에서 177
최근접 이웃 분석
결과 175
모형 보기 176
변수 저장 175
옵션 175
파티션 174
필드선택 174
항목 173
최대 분기
이단계 군집 분석 186
최대값
기술통계 77
데이터 탐색 80
보고서 열 비교 219
비율 통계 233
빈도분석 75
요약 86
평균 87
OLAP 큐브 89
최대우도
요인 분석 182
최소값
기술통계 77
데이터 탐색 80
보고서 열 비교 219
비율 통계 233
빈도분석 75
요약 86
평균 87
OLAP 큐브 89
최소유의차
일원 분산 분석 104
GLM 112
최적 서브세트
선형모형에서 139
추론 대화 상자
메타분석 연속형 35
메타분석 연속형 효과 크기 42
메타분석 이분형 50
메타분석 이분형 효과 크기 57
메타분석 회귀분석 64

카

카이저 정규화(Z)
요인 분석 183
카이제곱
결측값 205
교차 분석표 83
기대값 205
기대범위 205
독립성 83
선형별 연관 83
옵션 205
우도비 83
일표본 검정 204
통계 205
통계적 가설 테스트 117, 118
Fisher의 정확 검정 83
Pearson 83
Yates의 연속성 보정 83
카이제곱 거리
거리 측도 134
카이제곱 검정
일표본 비모수 검정 196, 197
케이스 목록 85
케이스 수
요약 86
평균 87
OLAP 큐브 89
케이스별 진단 정보
선형 회귀분석 145
켄달의 일치계수(W)
대응표본 비모수 검정 202
코드북
결과 73
통계 74
콜모고로프-스미르노프 검정
일표본 비모수 검정 196, 197
Lilliefors 검정 197
쿼티멕스 회전
요인 분석 183
크기 차이 측도
거리 측도 134
클로퍼-피어슨 구간

클로퍼-피어슨 구간 (계속)
일표본 비모수 검정 196

타

타우-b
교차 분석표 83
타우-c
교차 분석표 83
탐색
거듭제곱값 변환 80
결측값 81
도표 80
명령 추가 기능 81
옵션 81
통계 80
토네이도 도표
시뮬레이션에서 252

파

파이
교차 분석표 83
파이제곱 거리 측도
거리 측도 134
판별 분석
결측값 180
공분산 행렬 180
그룹화 변수 178
기술통계량 179
기준 180
단계적 방법 178
도표 180
독립변수 178
명령 추가 기능 181
모형 내보내기 정보 181
범위지정 179
분류분석 변수 저장 181
사전 확률 180
예 178
출력 옵션 180
케이스 선택 179
통계 178, 179
판별 방법 180
함수의 계수 179
행렬 179
Mahalanobis의 거리 180
Rao의 V 180
Wilks의 람다 180
패턴 차이 측도
거리 측도 134
패턴 행렬
요인 분석 181
퍼센트
교차 분석표 84
편상관계수
0차 상관 133
결측값 133
명령 추가 기능 133
선형 회귀분석 145
옵션 133
통계 133
편차 대비
편차 대비 (계속)
GLM 109
편최소제곱회귀
모델 171
변수 내보내기 171
편향 대화 상자
메타분석 연속형 36
메타분석 연속형 효과 크기 43
메타분석 이분형 51
메타분석 이분형 효과 크기 59
편회귀잔차 도표
선형 회귀분석 143
평가자 간 일치도 221
평균
기술통계 77
다중 보고서 열 219
데이터 탐색 80
비율 통계 233
빈도분석 75
열별 요약 보고서 219
옵션 87
요약 86
일원 분산 분석 105
통계 87
평균 87
하위 그룹 87, 88
행별 요약 보고서 217
OLAP 큐브 89
평균-산포 도표
데이터 탐색 80
평균의 표준 오차
요약 86
평균 87
OLAP 큐브 89
평행선 검정
순서 회귀분석 147
포리스트 도표 66
포리스트 도표 탭
메타분석 연속형 66
메타분석 연속형 효과 크기 66
메타분석 이분형 66
메타분석 이분형 효과 크기 66
표준 오차
기술통계 77
데이터 탐색 80
빈도분석 75
GLM 114
ROC 곡선 242
ROC 분석에서 240
표준편차
기술통계 77
데이터 탐색 80
비율 통계 233
빈도분석 75
열별 요약 보고서 219
요약 86
평균 87
행별 요약 보고서 217
OLAP 큐브 89
표준화
이단계 군집 분석 186
표준화 값
기술통계 77
표준화 잔차

표준화 잔차 (계속)
선형 회귀분석 144
GLM 114
표준화하지 않은 잔차
GLM 114
프로파일 도표
GLM 110
프록시맵 분석 123
플롯
in 블랜드 알트만 117
피어
최근접 이웃 분석에서 177
필드선택
최근접 이웃 분석에서 178

하

하위 그룹 평균 87, 88
합계
기술통계 77
빈도분석 75
요약 86
평균 87
OLAP 큐브 89
핫지스-레만 추정
대응표본 비모수 검정 202
항 설정 108, 148, 149
행 백분율
교차 분석표 84
행별 요약 보고서
결측값 218
구분 구간 217
구분 열 216
꼬리말 218
데이터 열 216
명령 추가 기능 220
열 형식 217
정렬 시퀀스 216
제목 218
제목변수 218
쪽 레이아웃 218
쪽 매김 218
쪽 제어 217
허용 오차
선형 회귀분석 145
형식
보고서 열 217
확률 도표
P-p 도표 234
Q-Q 236
확률 밀도 함수
시뮬레이션에서 251
회귀계수
선형 회귀분석 145
회귀분석
다중 회귀분석 142
도표 143
메타분석 32
선형 회귀 142
효과크기
대응표본 T 검정 100
독립 표본 T 검정 99
후진제거법
선형 회귀분석 143
훈련 표본
최근접 이웃 분석에서 174
히스토그램
데이터 탐색 80
빈도분석 76
선형 회귀분석 143