

***IBM SPSS Statistics Base 22***

**IBM**

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에 201 페이지의 『주의사항』의 정보를 읽으십시오.

제품 정보

이 개정판은 새 개정판에서 별도로 명시하지 않는 한, IBM SPSS Statistics의 버전 22, 릴리스 0. 수정사항 0 및 모든 후속 릴리스와 수정에 적용됩니다.

# 목차

제 1 장 코드북 . . . . .	1	독립 표본 T 검정 그룹 정의 . . . . .	34		
코드북 결과 탭 . . . . .	2	독립 표본 T 검정: 옵션 . . . . .	34		
코드북 통계 탭 . . . . .	4	대응표본 T검정 . . . . .	35		
제 2 장 빈도분석 . . . . .	5	대응표본 T 검정: 옵션 . . . . .	35		
빈도분석 통계 . . . . .	6	T-TEST 명령 추가 기능 . . . . .	36		
빈도분석 차트 . . . . .	7	일표본 T 검정 . . . . .	36		
빈도분석 형식 . . . . .	7	일표본 T 검정: 옵션 . . . . .	36		
제 3 장 기술통계 . . . . .	9	T-TEST 명령 추가 기능 . . . . .	37		
기술통계 옵션 . . . . .	9	T-TEST 명령 추가 기능 . . . . .	37		
DESCRIPTIVES 명령 추가 기능 . . . . .	10	제 10 장 일원 분산 분석 . . . . .	39		
제 4 장 데이터 탐색 . . . . .	11	일원 분산 분석 대비 . . . . .	40		
데이터 탐색 통계 . . . . .	12	일원 분산 분석 사후 검정 . . . . .	40		
데이터 탐색 도표 . . . . .	12	일원 분산 분석: 옵션 . . . . .	42		
데이터 탐색 거듭제곱값 변환 . . . . .	13	ONEWAY 명령 추가 기능 . . . . .	42		
데이터 탐색: 옵션 . . . . .	13	제 11 장 GLM 일변량 분석 . . . . .	43		
EXAMINE 명령 추가 기능 . . . . .	13	GLM 모형 . . . . .	44		
제 5 장 교차 분석표 . . . . .	15	항 설정 . . . . .	45		
교차 분석표 레이아웃 . . . . .	16	제공함 . . . . .	45		
교차 분석표 묶은 가로 막대형 차트 . . . . .	16	GLM 대비 . . . . .	46		
테이블 레이아웃에 레이아웃 변수를 표시하는 교차 분석 표 . . . . .	16	대비 유형 . . . . .	46		
교차 분석표 통계 . . . . .	17	GLM 프로파일 도표 . . . . .	47		
교차 분석표 셀 표시 . . . . .	18	GLM 옵션 . . . . .	47		
교차 분석표 형식 . . . . .	20	UNIANOVA 명령 추가 기능 . . . . .	48		
제 6 장 요약 . . . . .	21	GLM 사후분석 비교 . . . . .	49		
요약분석 옵션 . . . . .	22	GLM 옵션 . . . . .	50		
사례 요약 통계 . . . . .	22	UNIANOVA 명령 추가 기능 . . . . .	51		
제 7 장 평균 . . . . .	25	GLM 저장 . . . . .	51		
그룹별 평균분석: 옵션 . . . . .	26	GLM 옵션 . . . . .	52		
제 8 장 OLAP 큐브 . . . . .	29	UNIANOVA 명령 추가 기능 . . . . .	53	제 12 장 이변량 상관계수 . . . . .	55
OLAP 큐브 통계 . . . . .	30	이변량 상관: 옵션 . . . . .	56	CORRELATIONS 및 NONPAR CORR 명령 추가 기능 . . . . .	56
OLAP 큐브 차이 . . . . .	31	제 13 장 편상관계수 . . . . .	57	편상관: 옵션 . . . . .	58
OLAP 큐브 제목 . . . . .	32	PARTIAL CORR 명령 추가 기능 . . . . .	58	제 14 장 거리 . . . . .	59
제 9 장 T 검정 . . . . .	33	거리측도: 상이성 측도 . . . . .	59		
T 검정 . . . . .	33				
독립 표본 T 검정 . . . . .	33				

거리측도: 유사성 측도 . . . . .	60
PROXIMITIES 명령 추가 기능 . . . . .	60
<b>제 15 장 선형모형 . . . . .</b>	<b>61</b>
선형 모형 구하기 . . . . .	61
목표 . . . . .	61
기본 사항 . . . . .	62
모형 선택 . . . . .	63
양상블 . . . . .	64
고급 . . . . .	64
모형 옵션 . . . . .	64
모형 요약 . . . . .	64
자동 데이터 준비 . . . . .	65
예측변수 중요도 . . . . .	65
관찰에 의한 예측 . . . . .	65
잔차 . . . . .	65
이상치 . . . . .	66
효과 . . . . .	66
계수 . . . . .	67
평균 추정 . . . . .	67
모형 작성 요약 . . . . .	67
<b>제 16 장 선형 회귀분석 . . . . .</b>	<b>69</b>
선형 회귀분석 변수 선택 방법 . . . . .	70
선형 회귀분석 규칙 설정 . . . . .	70
선형 회귀분석 도표 . . . . .	71
선형 회귀: 새로운 변수로 저장 . . . . .	71
선형 회귀분석 통계 . . . . .	73
선형 회귀분석 옵션 . . . . .	73
REGRESSION 명령 추가 기능 . . . . .	74
<b>제 17 장 순서 회귀분석 . . . . .</b>	<b>75</b>
순서 회귀분석: 옵션 . . . . .	76
순서 회귀분석 결과 . . . . .	76
순서 회귀분석 위치 모형 . . . . .	77
항 설정 . . . . .	77
순서 회귀분석 척도 모형 . . . . .	78
항 설정 . . . . .	78
PLUM 명령 추가 기능 . . . . .	78
<b>제 18 장 곡선 추정 . . . . .</b>	<b>79</b>
곡선 추정 모형 . . . . .	80
곡선 추정: 저장 . . . . .	80
<b>제 19 장 일부 최소제곱 회귀분석 . . . . .</b>	<b>83</b>
모형 . . . . .	85
옵션 . . . . .	85

<b>제 20 장 최근접 이웃 분석 . . . . .</b>	<b>87</b>
이웃 . . . . .	89
기능 . . . . .	90
파티션 . . . . .	90
저장 . . . . .	91
결과 . . . . .	92
옵션 . . . . .	92
모형 보기 . . . . .	92
기능 공간 . . . . .	93
변수 중요도 . . . . .	94
피어 . . . . .	94
가장 가까운 이웃 거리 . . . . .	95
4분위 맵 . . . . .	95
필드선택 오차 로그 . . . . .	95
k 선택 오차 로그 . . . . .	95
k 및 필드선택 오차 로그 . . . . .	95
분류표 . . . . .	95
오차 요약 . . . . .	96
<b>제 21 장 판별 분석 . . . . .</b>	<b>97</b>
판별 분석 범위지정 . . . . .	98
판별 분석 케이스 선택 . . . . .	98
판별 분석 통계 . . . . .	98
판별 분석 단계적 방법 . . . . .	99
판별 분석 분류 . . . . .	100
판별 분석 저장 . . . . .	100
DISCRIMINANT 명령 추가 기능 . . . . .	101
<b>제 22 장 요인 분석 . . . . .</b>	<b>103</b>
요인 분석 선택변수 . . . . .	104
요인 분석: 기술통계 . . . . .	104
요인 분석: 요인추출 . . . . .	104
요인 분석: 회전 . . . . .	105
요인 분석: 요인 스코어 . . . . .	106
요인 분석: 옵션 . . . . .	106
FACTOR 명령 추가 기능 . . . . .	107
<b>제 23 장 군집화 프로시저 선택 . . . . .</b>	<b>109</b>
<b>제 24 장 이단계 군집 분석 . . . . .</b>	<b>111</b>
이단계 군집 분석 옵션 . . . . .	112
이단계 군집 분석 결과 . . . . .	114
군집 뷰어 . . . . .	114
군집 뷰어 . . . . .	114
군집 뷰어 탐색 . . . . .	118
레코드 필터링 . . . . .	119

제 25 장 계층적 군집 분석 . . . . .	121
계층적 군집 분석 . . . . .	122
계층적 군집 분석: 통계 . . . . .	122
계층적 군집 분석: 도표 . . . . .	122
계층적 군집 분석: 새로운 변수로 저장 . . . . .	123
CLUSTER 명령 구문 추가 기능 . . . . .	123
제 26 장 K-Means 군집 분석 . . . . .	125
K-Means 군집 분석 효율성 . . . . .	126
K-Means 군집 분석: 반복 . . . . .	126
K-Means 군집 분석: 저장 . . . . .	126
K-Means 군집 분석: 옵션 . . . . .	127
QUICK CLUSTER 명령 추가 기능 . . . . .	127
제 27 장 비모수 검정 . . . . .	129
일표본 비모수 검정 . . . . .	129
일표본 비모수 검정을 수행하려면 . . . . .	129
필드 탭 . . . . .	129
설정 탭 . . . . .	130
NPTESTS 명령 추가 기능 . . . . .	132
독립 표본 비모수 검정 . . . . .	133
독립 표본 비모수 검정을 수행하려면 . . . . .	133
필드 탭 . . . . .	133
설정 탭 . . . . .	134
NPTESTS 명령 추가 기능 . . . . .	135
대응표본 비모수 검정 . . . . .	135
대응표본 비모수 검정을 수행하려면 . . . . .	136
필드 탭 . . . . .	136
설정 탭 . . . . .	136
NPTESTS 명령 추가 기능 . . . . .	138
모형 보기 . . . . .	139
모형 보기 . . . . .	139
NPTESTS 명령 추가 기능 . . . . .	144
레저시 대화 상자 . . . . .	144
카이제곱 검정 . . . . .	145
이항검정 . . . . .	146
런 검정 . . . . .	147
일표본 Kolmogorov-Smirnov 검정 . . . . .	148
독립 2-표본 비모수 검정 . . . . .	149
대응 2-표본 비모수 검정 . . . . .	151
독립 K-표본 비모수 검정 . . . . .	152
대응 K-표본 비모수 검정 . . . . .	154
제 28 장 다중 응답 분석 . . . . .	157
다중 응답 분석 . . . . .	157
다중 응답 세트 정의 . . . . .	157
다중반응 빈도분석 . . . . .	158

다중반응 교차 분석표 . . . . .	159
다중반응 교차 분석표 범위지정 . . . . .	160
다중반응 교차 분석표: 옵션 . . . . .	160
MULT RESPONSE 명령 추가 기능 . . . . .	161
제 29 장 결과 보고 . . . . .	163
결과 보고 . . . . .	163
행별 요약 보고서 . . . . .	163
요약 보고서 구하기: 행별 요약 . . . . .	164
보고서 데이터 열/구분 형식 . . . . .	164
보고서 요약 행/최종 요약 행 . . . . .	164
보고서 구분: 옵션 . . . . .	164
보고서: 옵션 . . . . .	165
보고서: 윤곽 . . . . .	165
보고서: 제목 . . . . .	165
열별 요약 보고서 . . . . .	166
요약 보고서 구하기: 열별 요약 . . . . .	166
데이터 열 요약 함수 . . . . .	166
총계 열에 대한 데이터 열 요약 . . . . .	167
보고서 열 형식 . . . . .	167
열별 요약 보고서 구분 옵션 . . . . .	167
열별 요약 보고서 옵션 . . . . .	167
열별 요약을 위한 보고서 윤곽 . . . . .	168
REPORT 명령 추가 기능 . . . . .	168
제 30 장 신뢰성 분석 . . . . .	169
신뢰성 분석: 통계 . . . . .	170
RELIABILITY 명령 추가 기능 . . . . .	171
제 31 장 다차원척도법 . . . . .	173
다차원척도법 데이터의 형태 . . . . .	174
다차원척도법 측도 계산 . . . . .	174
다차원척도법: 모형 . . . . .	175
다차원척도법: 옵션 . . . . .	175
ALSCAL 명령 추가 기능 . . . . .	175
제 32 장 비율 통계 . . . . .	177
비율 통계 . . . . .	177
제 33 장 ROC 곡선 . . . . .	179
ROC 곡선: 옵션 . . . . .	179
제 34 장 시뮬레이션 . . . . .	181
모형 파일을 기반으로 시뮬레이션을 디자인하는 방 법 . . . . .	182
사용자 정의 방정식을 기반으로 시뮬레이션을 디자 인하는 방법 . . . . .	182
예측 모형 없이 시뮬레이션 디자인 . . . . .	183

시뮬레이션 계획에서 시뮬레이션을 실행하는 방법	184	시뮬레이션의 차트 결과로 작업	198
시뮬레이션 작성기	185	차트 옵션	199
모형 탭	185	주의사항	201
시뮬레이션 탭	187	상표	203
시뮬레이션 실행 대화 상자	195	색인	205
시뮬레이션 탭	196		
결과 탭	197		

---

## 제 1 장 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 다중 대치를 위해 만든 분할 파일 그룹이 포함됩니다(결측값 추가 기능 옵션에서 사용 가능).

### 코드북 가져오기

1. 메뉴에서 다음을 선택합니다.

분석 > 보고서 > 코드북

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 4 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 DATE11과 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 이분형 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® Statistics 데이터 파일의 이름입니다. 데이터 세트가 IBM SPSS Statistics 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 Statistics 데이터 파일의 디렉토리(폴더) 위치입니다. 데이터 세트가 IBM SPSS Statistics 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 명령에 의해 정의된 파일 설명입니다(해당하는 경우).

**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.

개수 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(Mean). 중심 경향에 대한 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

표준 편차(Standard Deviation). 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.

사분위수(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.

---

## 제 2 장 빈도분석

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 막대형 차트에서는 빈도에 따라 내림차순이나 오름차순으로 개별값을 조정하고 범주 순서를 지정합니다. 변수에 개별값이 많을 때는 빈도분석 보고서가 출력되지 않을 수도 있습니다. 차트에 빈도분석(기본값) 또는 백분율을 레이블로 붙여넣을 수 있습니다.

예제. 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.

통계 및 도표. 빈도 개수, 백분율, 누적 백분율, 평균, 중앙값, 최빈값, 합계, 표준 편차, 분산, 범위, 최소값 및 최대값, 평균의 표준 오차, 왜도와 첨도(표준 오차 포함), 분위수, 사용자 정의 백분위수, 막대형 차트, 원형 차트, 히스토그램 등을 선택할 수 있습니다.

빈도분석 데이터 고려 사항

데이터. 숫자 코드나 문자를 사용하여 범주형 변수(명목 수준이나 순위 수준 측정)를 코딩합니다.

가정. 표와 백분율을 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. 중앙값, 분위수, 백분위수 등의 강력한 통계는 정규성 가정에 상관 없이 양적변수에 적합합니다.

빈도표 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 기술통계량 > 빈도...

2. 범주형 변수와 양적변수를 하나 이상 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 양적변수에 대한 기술통계를 보려면 통계를 클릭합니다.
- 막대형 차트, 원형 차트, 히스토그램 등을 보려면 차트를 클릭합니다.
- 결과가 출력되는 순서를 보려면 형식을 클릭합니다.

## 빈도분석 통계

**백분위수 값.** 특정 백분율은 위에 있고 다른 백분율은 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 수를 선택합니다. 개별 백분위수를 지정할 수도 있습니다. 예를 들어, 제 95 백분위수는 관측값의 95%이하인 값을 나타냅니다.

**중심 경향.** 분포 위치를 설명하는 통계에는 평균, 중앙값, 최빈값, 모든 값 합계 등이 포함됩니다.

- **평균(Mean).** 중심 경향에 대한 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.
- **중앙값(Median).** 전체 케이스의 절반이 위 아래에 해당되는 값으로 제50 백분위수입니다. 케이스 수가 짝수인 경우 중앙값은 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중앙값은 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.
- **합계(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 측도로, 평균으로부터의 제곱합 편차를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.
- **범위(Range).** 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.
- **최소값(Minimum).** 숫자변수의 가장 작은 값입니다.
- **최대값(Maximum).** 숫자변수의 가장 큰 값입니다.
- **평균의 표준 백분위수.** 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.

**분포.** 첨도와 왜도는 분포의 형태와 대칭을 설명하는 통계입니다. 표준 오차와 함께 표시됩니다.

- **왜도(Skewness).** 분포의 비대칭성에 대한 측도입니다. 정규 분포는 대칭이므로 왜도 값이 0입니다. 양의 왜도가 많은 분포는 오른쪽이 깎입니다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.
- **첨도(Kurtosis).** 관측값이 중심에 군집하는 정도에 대한 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 정규 분포에 비해 관측값이 분포 중심에 더 많이 군집되어 있고 분포 극단값까지의 꼬리

가 더 얇다는 의미입니다. 즉, 정규 분포에 비해 급침 분포의 꼬리가 더 두껍습니다. 음의 첨도는 정규 분포에 비해 관측값이 분포 중심에 덜 균집되어 있고 분포 극단값까지의 꼬리가 더 두껍다는 의미입니다. 즉, 정규 분포에 비해 평침 분포의 꼬리가 더 얇습니다.

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 중앙값과 백분위수를 추정할 수 있습니다.

---

## 빈도분석 차트

차트 유형. 원형 차트는 전체에서 부분이 차지하는 정도를 표시합니다. 원형 차트의 각 조각은 단일 그룹화 변수로 정의한 그룹에 해당합니다. 막대형 차트에서 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.

차트화 값. 막대형 차트에서 빈도 개수나 백분율을 척도 축에 레이블로 붙여넣을 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.



---

## 제 3 장 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 차트, 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.

**예제.** 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 척도 및 왜도입니다.

기술통계 데이터 고려 사항

**데이터.** 오차, 이상치, 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.

기술통계 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 기술통계량 > 기술통계...

2. 하나 이상의 변수를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- $z$  스코어를 새로운 변수로 저장하려면 **표준화 값을 변수로 저장**을 선택합니다.
- 선택적인 통계와 출력 순서를 선택하려면 **옵션**을 클릭합니다.

---

## 기술통계 옵션

**평균 및 합계.** 평균이나 산술 평균이 기본적으로 표시됩니다.



산포. 데이터의 산포도나 변동을 측정하는 통계에는 표준 편차, 분산, 범위, 최소값, 최대값, 평균의 표준 오차가 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 측도로, 평균으로부터의 제곱합 편차를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.
- **범위(Range).** 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.
- **최소값(Minimum).** 숫자변수의 가장 작은 값입니다.
- **최대값(Maximum).** 숫자변수의 가장 큰 값입니다.
- **평균의 표준 백분위수.** 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.

분포. 첨도와 왜도는 분포의 형태와 대칭의 특성을 나타내는 통계로 표준 오차와 함께 표시됩니다.

- **첨도(Kurtosis).** 관측값이 중심에 군집하는 정도에 대한 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 정규 분포에 비해 관측값이 분포 중심에 더 많이 군집되어 있고 분포 극단값까지의 꼬리가 더 얇다는 의미입니다. 즉, 정규 분포에 비해 급침 분포의 꼬리가 더 두껍습니다. 음의 첨도는 정규 분포에 비해 관측값이 분포 중심에 덜 군집되어 있고 분포 극단값까지의 꼬리가 더 두껍다는 의미입니다. 즉, 정규 분포에 비해 평침 분포의 꼬리가 더 얇습니다.
- **왜도(Skewness).** 분포의 비대칭성에 대한 측도입니다. 정규 분포는 대칭이므로 왜도 값이 0입니다. 양의 왜도가 많은 분포는 오른쪽이 깎입니다. 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.



## 제 4 장 데이터 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, 이상치 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.

**예제.** 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, 중앙값, 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 이중가중 추정량, 최대 다섯 개 값 및 최소 다섯 개 값, 정규성 검정에서 Lilliefors 유의 수준의 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 이상치 레이블에 사용되며 단문자와 긴 문자(처음 15바이트), 숫자 등이 될 수 있습니다.

**가정.** 데이터 분포가 대칭이거나 정규적일 필요는 없습니다.

데이터 탐색

1. 메뉴에서 다음을 선택합니다.

분석 > 기술통계량 > 탐색...

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, 이상치, 백분위수, 빈도표 등을 확인하려면 통계를 클릭합니다.

- 히스토그램, 정규 확률 도표와 검정, Levene 통계와 평균-산포 도표를 보려면 도표를 클릭합니다.
- 결측값을 처리하려면 옵션을 클릭합니다.

## 데이터 탐색 통계

**기술통계.** 중심 경향과 산포의 측도를 기본값으로 표시합니다. 중심 경향 측도로 분포 위치를 알 수 있으며 여기에는 평균, 중앙값, 5% 잘린 평균 등이 포함됩니다. 산포 측도를 통해 값의 상이성을 알 수 있는데 여기에는 표준 오차, 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 중앙값을 대체합니다. 계산된 추정량은 케이스에 적용한 가중치에서 차이가 납니다. Huber의 M-추정량, Andrew의 웨이브 추정량, Hampel의 redescending M 추정량, Tukey 이중가중 추정량 등이 표시됩니다.

**이상치.** 최대 다섯 개 값과 최소 다섯 개 값을 케이스 레이블과 함께 표시합니다.

**백분위수.** 제 5, 10, 25, 50, 75, 90, 95 백분위수 값을 표시합니다.

## 데이터 탐색 도표

**상자도표.** 이러한 옵션을 사용하여 여러 종속변수가 있을 때 상자도표의 출력을 제어할 수 있습니다. 요인 수준들과 함께 각 종속변수에 따라 출력 형식을 다르게 만들 수 있습니다. 출력 형식에서 상자도표는 요인변수로 정의한 그룹마다 나타납니다. 종속변수와 함께 요인변수를 사용하여 정의한 각 그룹의 출력 형식을 다르게 만들 수 있습니다. 출력 형식에서 상자도표는 각 종속변수에 대해 단계별로 나타냅니다. 이러한 출력 형식은 서로 다른 변수가 다른 시간에 측정된 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 유의 수준을 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 추정으로 셀에서 등분산에 대한 거듭제곱값 변환을 추정할 수 있을 뿐 아니라 모든 셀에서 중앙값의 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. 변환을 사용하여 거듭제곱값 중 하나를 선택하고 거듭제곱값 추정의 설명을 따라 수행하여 변환 데이터 도표를 만들 수 있습니다. 변환 데이터의 사분위수 범위와 중앙값이 도표에 나타납니다. 변환하지 않으므로 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 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. 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 포함되거나 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.



---

## 제 5 장 교차 분석표

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.

교차 분석표 통계와 연관 측도는 이원배치표에 대해서만 계산됩니다. 행, 열, 레이어 요인(제어변수) 등을 지정 하면 각 레이어 요인 값(또는 둘 이상의 제어변수에 대한 값 조합)에 대해 결합된 통계와 측도의 패널을 작성 하게 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.

**예제.** 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 랏다, 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 숫자값이나 문자열 변수(8바이트 이내)값을 사용합니다. 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 > 기술통계량 > 교차 분석표...

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, 백분율, 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 교차 분석표가 작성됩니다. 예를 들어, 행 변수와 열 변수가 하나씩 있고 레이어 변수에 범주가 둘 있을 경우 레이어 변수의 범주마다 이원배치표가 작성됩니다. 제어변수의 다른 레이어를 작성하려면 다음을 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 묶은 가로 막대형 차트

묶은 가로 막대형 차트 출력. 묶은 가로 막대형 차트로 케이스 그룹의 데이터를 요약할 수 있습니다. 행에 지정한 각 변수 값에 대해 하나의 묶은 가로 막대형 차트가 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 묶은 가로 막대형 차트가 작성됩니다.

---

## 테이블 레이어에 레이어 변수를 표시하는 교차 분석표

테이블 레이어에 레이어 변수를 표시합니다. 레이어 변수(제어변수)를 교차 분석표의 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 표시를 선택합니다.
3. 셀 표시 하위 대화 상자에서 열을 선택합니다.
4. 교차 분석표 프로시저를 실행하고 교차 분석표를 두 번 클릭한 다음 교육수준 드롭다운 목록에서 대학 졸업을 선택합니다.

교차 분석표의 선택한 보기는 대학 졸업 수준의 반응자에 대한 통계를 표시합니다.



## 교차 분석표 통계

**카이제곱.** 행 두 개와 열 두 개가 있는 테이블의 경우 카이제곱을 선택하여 Pearson 카이제곱, 우도비 카이제곱, Fisher의 정확 검정 및 Yates의 수정 카이제곱(연속수정)을 계산할 수 있습니다.  $2 \times 2$  테이블의 경우 Fisher의 정확 검정은 더 큰 테이블의 결측 행이나 열로 이루어지지 않은 테이블에 5보다 작은 기대빈도를 가지는 셀이 있을 때 계산됩니다. Yates의 수정 카이제곱은 다른 모든  $2 \times 2$  테이블에 대해 계산됩니다. 임의의 수의 행과 열이 있는 테이블에서는 카이제곱을 선택하여 Pearson 카이제곱과 우도비 카이제곱을 계산합니다. 두 테이블 모두 변수가 양적인 경우 카이제곱으로 선형 대 선형 연관 검정을 수행할 수 있습니다.

**상관.** 행과 열에 정렬된 값이 있는 테이블의 경우 상관을 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 상관을 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.

- **분할계수.** 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 Cramer의 V(Phi and Cramer's V).** 파이는 카이제곱 통계량을 표본 결과로 나누고 그 결과의 제곱근을 구하는 연관 측도에 기반한 카이제곱입니다. Cramer의 V는 카이제곱을 기반으로 한 연관 측도입니다.
- **람다(Lambda).** 독립변수의 값이 종속변수의 값을 예측하는 데 사용될 때 오차 내 비례 축소를 반영하는 연관 측도입니다. 값 1은 독립변수가 종속변수를 완전하게 예측한다는 의미이고, 값 0은 독립변수가 종속변수를 예측하는 데 도움이 되지 않는다는 의미입니다.
- **불확실성 계수.** 하나의 변수값이 다른 변수값을 예측하는 데 사용될 때 오차 내 비례 축소를 나타내는 연관 측도입니다. 예를 들어, 값 0.83은 다른 변수 값을 예측하는 데 하나의 변수 정보가 오차를 83%까지 축소할 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 d를 선택합니다.

- **감마(Gamma).** 범위가 -1에서 1 사이인 두 순서변수 간 연관의 대칭적 측도입니다. 값이 절대값 1에 가까우면 두 변수 간의 관계가 강함을 나타내며 0에 가까울수록 관계가 적거나 없음을 의미합니다. 2원배치표의 경우 0차 감마가, 3원배치 이상의 테이블에서는 조건부 감마가 표시됩니다.
- **Somers의 d(Somers' d).** 범위가 -1부터 1까지인 두 개의 순서변수 사이의 연관 측도입니다. 절대값 1에 가까울수록 두 변수 사이의 연관 정도가 높음을 의미하고, 0에 가까울수록 변수 사이의 관계가 적거나 없음을 의미합니다. Somers의 d는 독립변수에 대해 동물이 아닌 대응 수가 포함된 경우에만 달라지는 감마의 비대칭 확장입니다. 이 통계의 대칭 버전도 계산됩니다.

- *Kendall의 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.
- *Kendall의 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 에타를 선택합니다. 범주형 변수는 숫자로 코딩되어야 합니다.

- *에타*. 0과 1 사이에 존재하는 연관 측도로서, 0은 행 변수와 열 변수 간에 연관이 없음을 나타내고 1에 가까울수록 연관 정도가 높음을 의미합니다. 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 같은 오브젝트에 대해 평가를 내리는 두 등급 평가자의 평가 동의 정도를 측정합니다. 값 1은 완전 동의를 나타내고 값 0은 동의의 여지가 없음을 나타냅니다. Kappa는 행 및 열 값이 동일한 척도를 나타내는 제곱표를 기반으로 합니다. 하나의 변수에 대한 관측값은 있지만 다른 변수에 대한 관측값은 없는 셀에는 개수 0이 지정됩니다. 두 개 변수에 대한 데이터 저장 유형(문자열 또는 숫자)이 동일하지 않을 경우 Kappa가 계산되지 않습니다. 문자열 변수의 경우 두 개 변수 모두 정의된 길이가 동일해야 합니다.

*위험*. 2x2 테이블의 경우, 요인의 존재와 이벤트 발생 간의 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" 디자인에서 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)*. Cochran 및 Mantel-Haenszel 통계는 하나 이상의 레이어(제어) 변수에 의해 정의되는 공변량 패턴의 조건에 따라 요인변수와 이분형 반응변수 간의 독립성 검정에 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, 백분율, 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 <N으로 표시되며 여기서 N은 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 백분율을 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.

**백분율.** 백분율은 행 가로방향으로나 열 세로방향으로 추가될 수 있습니다. 표(하나의 레이어)에 나타난 총 케이스 수의 백분율도 사용할 수 있습니다. 참고: 작은 수 숨기기가 개수 그룹에서 선택된 경우, 숨겨진 수와 연관된 백분율도 숨겨집니다.

**잔차.** 원래의 비표준화 잔차로 인해 관측값과 기대값 간의 차이가 나타납니다. 표준화 잔차와 수정된 표준화 잔차도 사용할 수 있습니다.

- **비표준화.** 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.
- **표준화.** 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.

**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 없음.** 케이스 가중치를 있는 그대로 사용하고 부분 셀 개수를 사용합니다. 그러나 정확한 검정 옵션에서만 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.

---

## 제 6 장 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.

**예제.** 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, 중앙값, 그룹 중앙값, 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  $N$  백분율, 합계 %,  $N$  백분율, 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: 중앙값과 범위)은 정규성 가정을 충족시키거나 충족시키지 않을 수 있는 양적변수에 적절합니다.

사례 요약분석 결과 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 보고서 > 사례 요약...

2. 하나 이상의 변수를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 하나 이상의 그룹화 변수를 선택하여 데이터를 하위 그룹으로 나눕니다.
- 옵션을 클릭하여 결과 제목을 변경하거나 결과 아래에 캡션을 추가하거나 결측값을 가지는 케이스를 제외시킵니다.
- 통계를 클릭하여 선택적 통계를 구합니다.
- 케이스 출력을 선택하여 각 하위 그룹에 케이스를 나열합니다. 기본적으로 시스템은 파일의 처음 100개의 케이스만 나열합니다. 첫 케이스부터의 출력수에 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 줄 바꿈을 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, 중앙값, 그룹 중앙값, 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  $N$  백분율, 합계 %,  $N$  백분율, 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  $n$ 제곱근입니다. 여기서  $n$ 은 케이스 수를 나타냅니다.

**그룹 중앙값(Grouped Median).** 그룹으로 코딩되는 데이터에 대해 계산된 중앙값입니다. 예를 들어, 나이 데이터에서 30대의 각 값은 35로 코딩되고 40대의 각 값은 45로 코딩되는 경우 그룹 중앙값은 코딩된 데이터에 대해 계산된 중앙값입니다.

**조화 평균(Harmonic Mean).** 그룹의 표본 결과가 다를 때 평균 그룹 크기를 추정하는 데 사용됩니다. 조화 평균은 전체 표본 수를 표본 결과의 역수의 합으로 나눈 값입니다.

**첨도(Kurtosis).** 관측값이 중심에 군집하는 정도에 대한 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 정규 분포에 비해 관측값이 분포 중심에 더 많이 군집되어 있고 분포 극단값까지의 꼬리가 더 얇다는 의미입니다. 즉, 정규 분포에 비해 급침 분포의 꼬리가 더 두껍습니다. 음의 첨도는 정규 분포에 비해 관측값이 분포 중심에 덜 군집되어 있고 분포 극단값까지의 꼬리가 더 두껍다는 의미입니다. 즉, 정규 분포에 비해 평침 분포의 꼬리가 더 얇습니다.

**마지막>Last.** 데이터 파일에서 마지막으로 나타나는 데이터 값을 표시합니다.

**최대값(Maximum).** 숫자변수의 가장 큰 값입니다.

**평균(Mean).** 중심 경향에 대한 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

**중앙값(Median).** 전체 케이스의 절반이 위 아래에 해당되는 값으로 제50 백분위수입니다. 케이스 수가 짝수인 경우 중앙값은 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중앙값은 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 %.** 각 범주에서 총 합계의 백분율입니다.

**범위(Range).** 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.

**왜도(Skewness).** 분포의 비대칭성에 대한 측도입니다. 정규 분포는 대칭이므로 왜도 값이 0입니다. 양의 왜도가 많은 분포는 오른쪽이 길다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도 값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.

**표준 편차(Standard Deviation).** 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 정규성 검정에 사용할 수 있습니다. 즉, 비율이 -2보다 작거나 +2보다 큰 경우 정규성을 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 왜도의 비율을 정규성 검정에 사용할 수 있습니다. 즉, 비율이 -2보다 작거나 +2보다 큰 경우 정규성을 거부할 수 있습니다. 왜도가 큰 양의 값인 경우 오른쪽이 길어지고 큰 음의 값인 경우 왼쪽이 길어집니다.

**합계(Sum).** 비결측값을 갖는 전체 케이스 값의 총계입니다.

**분산(Variance).** 평균에 대한 산포 측도로, 평균으로부터의 제곱합 편차를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.



---

## 제 7 장 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.

**예제.** 세 가지의 서로 다른 식용유에 녹아 있는 지방의 평균량을 계산하고 일원 분산 분석을 수행하여 평균의 차이를 확인할 수 있습니다.

**통계.** 합계, 케이스 수, 평균, 중앙값, 그룹 중앙값, 평균의 표준 오차, 최소값, 최대값, 범위, 그룹화 변수의 첫 번째 범주에 대한 변수값, 그룹화 변수의 마지막 범주에 대한 변수값, 표준 편차, 분산, 침도, 침도의 표준 오차, 왜도, 왜도의 표준 오차, 총 합계 %, 전체  $N$  백분율, 합계 %,  $N$  백분율, 기하평균, 조화 평균을 구할 수 있습니다. 옵션에는 분산 분석, 에타, 에타제곱, 선형성 검정  $R$ 과  $R^2$ 이 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. 중앙값 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 검증하려면 일원 분산 분석 프로시저에서 Levene 분산의 동질성 검정을 사용합니다.

하위 그룹 평균 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 평균 비교 > 평균...

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 옵션을 클릭하여 각 통계, 분산 분석표 및 에타, 에타제곱,  $R$ ,  $R^2$ 을 선택합니다.



---

## 그룹별 평균분석: 옵션

각 그룹화 변수의 각 범주 내에 있는 변수에 대해 다음 하위 그룹 통계량인 합계, 케이스 수, 평균, 중앙값, 그룹 중앙값, 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  $N$  백분율, 합계 %,  $N$  백분율, 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  $n$ 제곱근입니다. 여기서  $n$ 은 케이스 수를 나타냅니다.

**그룹 중앙값(Grouped Median).** 그룹으로 코딩되는 데이터에 대해 계산된 중앙값입니다. 예를 들어, 나이 데이터에서 30대의 각 값은 35로 코딩되고 40대의 각 값은 45로 코딩되는 경우 그룹 중앙값은 코딩된 데이터에 대해 계산된 중앙값입니다.

**조화 평균(Harmonic Mean).** 그룹의 표본 결과가 다를 때 평균 그룹 크기를 추정하는 데 사용됩니다. 조화 평균은 전체 표본 수를 표본 결과의 역수의 합으로 나눈 값입니다.

**첨도(Kurtosis).** 관측값이 중심에 군집하는 정도에 대한 척도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 정규 분포에 비해 관측값이 분포 중심에 더 많이 군집되어 있고 분포 극단값까지의 꼬리가 더 얇다는 의미입니다. 즉, 정규 분포에 비해 급침 분포의 꼬리가 더 두껍습니다. 음의 첨도는 정규 분포에 비해 관측값이 분포 중심에 덜 군집되어 있고 분포 극단값까지의 꼬리가 더 두껍다는 의미입니다. 즉, 정규 분포에 비해 평침 분포의 꼬리가 더 얇습니다.

**마지막>Last).** 데이터 파일에서 마지막으로 나타나는 데이터 값을 표시합니다.

**최대값(Maximum).** 숫자변수의 가장 큰 값입니다.

**평균(Mean).** 중심 경향에 대한 척도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

**중앙값(Median).** 전체 케이스의 절반이 위 아래에 해당되는 값으로 제50 백분위수입니다. 케이스 수가 짝수인 경우 중앙값은 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중앙값은 평균과 달리 중심을 벗어난 값에는 영향을 받지 않는 중심 경향 척도이며, 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.

**범위(Range).** 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.



왜도(Skewness). 분포의 비대칭성에 대한 측도입니다. 정규 분포는 대칭이므로 왜도 값이 0입니다. 양의 왜도가 많은 분포는 오른쪽이 길다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도 값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.

표준 편차(Standard Deviation). 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 정규성 검정에 사용할 수 있습니다. 즉, 비율이 -2보다 작거나 +2보다 큰 경우 정규성을 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 왜도의 비율을 정규성 검정에 사용할 수 있습니다. 즉, 비율이 -2보다 작거나 +2보다 큰 경우 정규성을 거부할 수 있습니다. 왜도가 큰 양의 값인 경우 오른쪽이 길어지고 큰 음의 값인 경우 왼쪽이 길어집니다.

합계(Sum). 비결측값을 갖는 전체 케이스 값의 총계입니다.

분산(Variance). 평균에 대한 산포 측도로, 평균으로부터의 제곱합 편차를 케이스 수에서 1을 뺀 값으로 나눈 값과 같습니다. 분산은 변수 자체의 제곱 단위로 측정됩니다.

첫번째 레이어에 대한 통계

분산 분석표 및 에타. 첫 번째 레이어의 각 독립변수에 대해 일원 분산 분석표를 표시하고 에타 및 에타제곱(연관 측도)을 계산합니다.

선형성 검정. 선형 및 비선형 성분과 관련된 제곱합, 자유도 및 평균 제곱과 더불어 F비, R 및 R 제곱을 계산합니다. 독립변수가 단문자열인 경우 선형성은 계산되지 않습니다.



---

## 제 8 장 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.

**예제.** 다양한 영역 및 영역 내의 제품군에 대한 합계 및 평균 판매액을 구합니다.

**통계.** 합계, 케이스 수, 평균, 중앙값, 그룹 중앙값, 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 %, 총 합계 %, 그룹화 변수내의 전체 케이스 %, 그룹화 변수 내의 총 합계 %, 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. 중앙값, 범위 등 강력한 통계는 정규성 가정을 충족시키거나 충족시키지 않을 수 있는 양적 변수에 적절합니다.

OLAP 큐브 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 보고서 > **OLAP 큐브...**

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 <N으로 표시되며 여기서 N은 지정된 정수입니다. 지정된 정수는 2보다 크거나 같아야 합니다.

다중 레이어 차원이 경량 표에서 렌더링되는 방법에 대한 정보는 내용을 참조하십시오.

---

## OLAP 큐브 통계

각 그룹화 변수의 각 범주 내에 있는 요약 변수에 대해 다음 하위 그룹 통계량인 합계, 케이스 수, 평균, 중앙값, 그룹 중앙값, 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  $n$ 제곱근입니다. 여기서  $n$ 은 케이스 수를 나타냅니다.

**그룹 중앙값(Grouped Median).** 그룹으로 코딩되는 데이터에 대해 계산된 중앙값입니다. 예를 들어, 나이 데이터에서 30대의 각 값은 35로 코딩되고 40대의 각 값은 45로 코딩되는 경우 그룹 중앙값은 코딩된 데이터에 대해 계산된 중앙값입니다.

**조화 평균(Harmonic Mean).** 그룹의 표본 결과가 다를 때 평균 그룹 크기를 추정하는 데 사용됩니다. 조화 평균은 전체 표본 수를 표본 결과의 역수의 합으로 나눈 값입니다.

**첨도(Kurtosis).** 관측값이 중심에 군집하는 정도에 대한 측도입니다. 정규 분포의 경우 첨도 통계 값은 0입니다. 양의 첨도는 정규 분포에 비해 관측값이 분포 중심에 더 많이 군집되어 있고 분포 극단값까지의 꼬리가 더 얇다는 의미입니다. 즉, 정규 분포에 비해 급침 분포의 꼬리가 더 두껍습니다. 음의 첨도는 정규 분포에 비해 관측값이 분포 중심에 덜 군집되어 있고 분포 극단값까지의 꼬리가 더 두껍다는 의미입니다. 즉, 정규 분포에 비해 평침 분포의 꼬리가 더 얇습니다.

**마지막>Last).** 데이터 파일에서 마지막으로 나타나는 데이터 값을 표시합니다.

**최대값(Maximum).** 숫자변수의 가장 큰 값입니다.

**평균(Mean).** 중심 경향에 대한 측도입니다. 합계를 케이스 수로 나눈 산술 평균 값입니다.

**중앙값(Median).** 전체 케이스의 절반이 위 아래에 해당되는 값으로 제50 백분위수입니다. 케이스 수가 짝수인 경우 중앙값은 케이스를 오름차순이나 내림차순으로 정렬했을 때 중간에 있는 두 개의 케이스의 평균입니다. 중앙값은 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 백분율입니다. 그룹화 변수가 하나만 있는 경우 이 값은 전체 케이스 수 백분율과 동일합니다.

합계 %. 다른 그룹화 변수의 범주 내에서 지정한 그룹화 변수의 합계 백분율입니다. 그룹화 변수가 하나만 있는 경우 이 값은 총 합계 백분율과 동일합니다.

전체 케이스 수 %. 각 범주에서 전체 케이스 수의 백분율입니다.

총 합계 %. 각 범주에서 총 합계의 백분율입니다.

범위(Range). 숫자변수의 가장 큰 값과 가장 작은 값의 차이로 최대값에서 최소값을 뺀 값을 의미합니다.

왜도(Skewness). 분포의 비대칭성에 대한 측도입니다. 정규 분포는 대칭이므로 왜도 값이 0입니다. 양의 왜도가 많은 분포는 오른쪽이 길다. 유의한 음의 왜도를 가지는 분포에는 왼쪽으로 긴 꼬리가 나타납니다. 왜도 값이 표준 오차의 두 배를 넘는 것은 대칭에서 벗어난 정도를 나타냅니다.

표준 편차(Standard Deviation). 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 정규성 검정에 사용할 수 있습니다. 즉, 비율이 -2보다 작거나 +2보다 큰 경우 정규성을 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 왜도의 비율을 정규성 검정에 사용할 수 있습니다. 즉, 비율이 -2보다 작거나 +2보다 큰 경우 정규성을 거부할 수 있습니다. 왜도가 큰 양의 값인 경우 오른쪽이 길어지고 큰 음의 값인 경우 왼쪽이 길어집니다.

합계(Sum). 비결측값을 갖는 전체 케이스 값의 총계입니다.

분산(Variance). 평균에 대한 산포 측도로, 평균으로부터의 제곱합 편차를 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 백분율을 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. 백분율 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. 백분을 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.

---

## 제 9 장 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 기술 통계는 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 디자인용입니다. 결과에는 검정 변수에 대한 기술 통계, 변수 사이의 상관, 쌍 차이대응에 대한 기술 통계,  $T$  검정 및 95% 신뢰구간이 포함됩니다.

**일표본 T 검정.** 한 변수의 평균을 알려진 값이나 가설 값과 비교합니다. 검정 변수의 기술 통계가  $T$  검정과 함께 표시됩니다. 검정 변수의 평균과 가설 검정 값 사이의 차이에 대한 95% 신뢰구간이 기본값 결과의 쌍입니다.

---

### 독립 표본 T 검정

독립 표본 T 검정 프로시저는 두 개의 케이스 그룹에 대한 평균을 비교합니다. 이 검정의 경우 반응상의 차이는 다른 요인이 아닌 처리(혹은 무처리)로 인한 것이므로 개체를 두 그룹에 무작위로 할당하는 것이 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.

**예제.** 고혈압 환자를 무작위로 거짓 처방 그룹(위안을 주기 위해 효과 없는 약을 제공)과 처리 그룹으로 분류 하였습니다. 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 표본 결과, 평균, 표준 편차, 평균의 표준 오차를 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.

독립 표본 T 검정 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 단



문자(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 이상치가 없음을 볼 수 있습니다.

독립 표본 T 검정 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 평균 비교 > 독립 표본 T 검정...

2. 양적 검정변수를 하나 이상 선택합니다. 각 변수에 대해 별도의  $t$  검정이 계산됩니다.
3. 단일 그룹화 변수를 선택한 후 그룹 정의를 클릭하여 비교하려는 그룹의 두 코드를 지정합니다.
4. 선택적으로 옵션을 클릭하여 결측 데이터의 처리와 신뢰구간의 수준을 제어할 수 있습니다.

## 독립 표본 T 검정 그룹 정의

숫자 그룹화 변수의 경우 두 개의 값이나 하나의 절단점을 지정하여  $t$  검정에 대한 두 그룹을 정의합니다.

- **지정값 사용.** 그룹 1과 그룹 2에 서로 다른 값을 입력합니다. 그 밖의 값을 포함하는 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.

문자 그룹화 변수의 경우 *yes*와 *no*와 같이 그룹 1과 그룹 2에 서로 다른 문자열을 입력합니다. 그 밖의 값을 포함하는 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 T검정

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.

**예제.** 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 디자인으로는 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 평균 차이,  $t$  검정, 평균차에 대한 신뢰구간(신뢰수준 지정 가능)을 선택할 수 있으며 평균차의 표준 편차와 표준 오차를 선택할 수 있습니다.

대응표본 T 검정 데이터 고려 사항

**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...

2. 하나 이상의 대응변수를 선택합니다.

3. 선택적으로 옵션을 클릭하여 결측 데이터의 처리와 신뢰구간의 수준을 제어할 수 있습니다.

### 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.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

---

## 일표본 T 검정

일표본 T 검정 프로시저는 단일-변수의 평균과 지정한 상수와의 차이를 검정합니다.

**예제.** 한 연구원이 학생 그룹의 평균 IQ 스코어가 100에서 얼마나 차이가 나는지를 검정하려는 경우 또는 한 시리얼 제조업체가 생산 라인에서 표본으로 상자들을 선택하여 95% 신뢰수준에서 표본의 평균 무게가 1.3 파운드에서 얼마나 차이가 나는지를 검정하려는 경우에 사용할 수 있습니다.

**통계.** 검정변수에 대해 평균, 표본 결과, 표준 편차, 평균의 표준 오차를 선택할 수 있습니다. 각 데이터 값과 가정된 검정값 사이의 평균 차이, 이 차이가 0인지를 검정하는  $t$  검정, 이 차이에 대한 신뢰구간(사용자가 신뢰 수준 지정 가능)을 선택할 수 있습니다.

일표본 T 검정 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...

2. 동일한 가정값과 비교하여 검정변수를 하나 이상 선택합니다.
3. 각 표본 평균과 비교할 숫자 검정값을 입력합니다.
4. 선택적으로 옵션을 클릭하여 결측 데이터의 처리와 신뢰구간의 수준을 제어할 수 있습니다.

## 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 사용됩니다. 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.

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.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



## 제 10 장 일원 분산 분석

일원 분산 분석 프로시저에서는 단일 요인(독립) 변수로 양적 종속변수에 대한 일원 분산 분석을 작성합니다. 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.

**예제.** 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 그룹에 대해 케이스 수, 평균, 표준 편차, 평균의 표준 오차, 최소값, 최대값, 평균에 대한 95% 신뢰 구간 통계를 구합니다. Levene의 분산 동질성 검정, 분산 분석 테이블 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 유의차입니다.

일원 분산 분석 데이터 고려 사항

**데이터.** 요인변수 값은 정수이어야 하고 종속변수 값은 양적변수이어야 합니다(측도의 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.

일원 분산 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 평균 비교 > 일원 분산 분석...

2. 하나 이상의 종속변수를 선택합니다.
3. 독립 요인변수를 하나만 선택합니다.

---

## 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 추가를 클릭합니다. 각각의 새로운 값은 계수 목록의 하단에 추가됩니다. 추가 대비 변수 세트를 지정하려면 다음을 클릭합니다. 다음과 이전을 사용하여 대비 변수 세트끼리 이동할 수 있습니다.

계수의 순서는 요인변수의 범주 값에 대한 오름차순에 해당하므로 중요합니다. 목록에서 첫 번째 계수는 요인 변수의 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 계수를 합하면 0이 되고 합이 0이 되지 않는 변수 세트도 사용할 수는 있으나 이 경우 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 모든 그룹 평균 간 쌍대 비교를 수행하며 다중 비교에 대한 오차 비율을 수정하지 않습니다.
- **Bonferroni.**  $T$  검정을 사용하여 그룹 평균 간 쌍대 비교를 수행하지만 각 검정에 대한 오차 비율을 전체 검정 수로 나눈 실험별 오차 비율로 설정하여 전체 오차 비율을 제어합니다. 따라서 관측 유의 수준은 다중 비교 작업에 따라 조정됩니다.
- **Sidak.**  $t$  통계를 기준으로 하는 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 수행할 수 있습니다. 표본 결과가 같은 단계별 프로시저를 사용하여 동일서브세트 내의 대응 평균을 비교할 수도 있습니다. 평균은 최고에서 최하까지 정렬되며 극단값치가 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 검정이 아닌 전체 검정에 대한 오차 비율에 대해 보호 수준을 설정합니다. 스튜던트화 범위 통계량을 사용합니다.
- *Hochberg*의 *GT2*. 스튜던트화 최대 계수를 사용하는 다중 비교 및 범위 검정입니다. Tukey의 정직유의차 검정과 유사합니다.
- *Gabriel*. 스튜던트화 최대 계수를 사용하며 셀 크기가 동일하지 않을 때 일반적으로 Hochberg의 *GT2*보다 좀더 효과적인 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.
- *Dunnnett*. 단일 통제 평균에 대해 처리 세트를 비교하는 대응별 다중 비교 t 검정을 수행합니다. 마지막 범주가 기본 제어 범주가 됩니다. 처음 범주를 선택할 수도 있습니다. 한 요인 수준(제어 범주 제외)에서의 평균이 제어 범주에서의 평균과 동일하지 않은지 검정하려면 양쪽검정을 사용합니다. 한 요인 수준에서의 평균이 제어 범주에서의 평균보다 작을지를 검정하려면 <통제를 선택합니다. 한 요인 수준에서의 평균이 제어 범주에서의 평균보다 큰지를 검정하려면 >통제를 선택합니다.

등분산을 가정하지 않음

Tamhane의 T2, Dunnnett의 T3, Games-Howell 및 Dunnnett의 C는 등분산을 가정하지 않는 다중 비교 검정입니다.

- *Tamhane*의 *T2*. t 검정을 기준으로 한 보수적 쌍대 비교를 수행합니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.
- *Dunnnett*의 *T3*. 스튜던트화 최대 계수를 기준으로 하는 쌍대 비교 검정을 수행합니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.
- *Games-Howell*. 경우에 따라 자유롭게 수행되는 쌍대 비교 검정입니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.



- *Dunnnett*의 *C*. 스튜던트화 범위를 기준으로 하는 쌍대 비교 검정을 수행합니다. 이 검정은 분산들이 동일하지 않을 때 적합합니다.

참고: 활성화된 피벗 테이블에서 형식 메뉴의 테이블 특성을 선택하여 나타난 테이블 특성 대화 상자에서 비어 있는 행과 열 감추기를 선택 해제하면 사후 검정 결과를 쉽게 해석할 수 있습니다.

## 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 랜덤 효과.** 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 통계를 계산하여 그룹 평균의 등식을 검정합니다. 등분산을 가정하지 않는 경우 *F* 통계보다 이 통계를 사용하는 것이 좋습니다.
- **Welch.** Welch 통계를 계산하여 그룹 평균의 등식을 검정합니다. 등분산을 가정하지 않는 경우 *F* 통계보다 이 통계를 사용하는 것이 좋습니다.

평균 도표. 하위 그룹 평균을 도표화하는 차트를 표시합니다(각 그룹의 평균은 요인변수 값으로 정의합니다).

결측값. 결측값의 처리를 제어합니다.

- **분석별 결측값 제외.** 분석에서 종속변수나 요인변수에 대한 결측값이 있는 케이스는 해당 분석에서 사용하지 않습니다. 요인변수에 지정한 범위 밖의 케이스도 사용하지 않습니다.
- **목록별 결측값 제외.** 기본 대화 상자의 종속변수 목록에 포함된 종속변수나 요인변수에 대한 결측값을 가지는 케이스는 모든 분석에서 제외됩니다. 다중 종속변수를 지정하지 않은 경우에는 아무런 영향이 없습니다.

## 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, 폴링분산 및 폴링분산의 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.

---

## 제 11 장 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 효과 뿐 아니라 무작위로 선출된 요인간의 상호작용을 조사할 수 있습니다. 더불어 요인에 미치는 공변량의 효과와 공변량 상호작용을 함께 검토할 수 있습니다. 회귀분석의 경우 독립(예측변수)변수가 공변량으로 지정됩니다.

균형 모형과 비균형 모형을 모두 검정할 수 있습니다. 모형의 각 셀이 같은 수의 케이스를 포함하는 경우에는 디자인을 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 예측 평균값의 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.

**예제.** 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 검정을 선택할 수 있습니다.

**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 일반량 테이블 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 일반 선형 모형 > 일반량...

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. 일반량 모형 대화 상자

모형 지정. 완전 요인 모형에는 모든 요인 주효과, 모든 공분산 주효과, 모든 요인-대-요인 상호작용이 포함됩니다. 공변량 상호작용은 포함되지 않습니다. 사용자 정의를 선택하여 하나의 상호작용 서브세트만 지정하거나 요인-대-공변량 상호작용을 지정합니다. 모형에 포함할 모든 항을 지정해야 합니다.

**요인 및 공변량.** 요인 및 공변량이 나열됩니다.

**모형.** 모형은 데이터의 특성에 따라 다릅니다. 사용자 정의를 선택한 다음 분석시 중요한 주효과와 상호작용을 선택할 수 있습니다.

**제공합.** 제공합을 계산하는 방법입니다. 결측 셀이 없는 균형 모형이나 비균형 모형의 경우 제 III 유형 제공합 방법이 가장 일반적으로 사용됩니다.

**모형에 절편 포함.** 이 모형에는 대개 절편이 포함됩니다. 데이터가 원점을 전달한다고 가정할 경우에는 절편을 제외시킬 수 있습니다.

## 항 설정

선택된 요인과 공변량에 대해 다음과 같습니다.

**상호작용.** 선택한 모든 변수의 최고-수준 상호작용 항을 생성합니다. 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.

## 제공합

모형의 경우 한 가지 유형의 제공합을 선택할 수 있습니다. 제 III 유형이 가장 일반적으로 사용되며 기본값입니다.

**제 I 유형.** 이 방법은 또한 제공합의 계층적 분해 방법으로 알려져 있습니다. 각 항은 모형에서 해당 항의 앞에 오는 항에 따라서만 수정됩니다. 제 I 유형 제공합은 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 제공합을 계산합니다. 적합한 효과는 검토 중인 효과를 포함하지 않는 모든 효과입니다. 제 II 유형 제공합 방법은 일반적으로 다음 경우에 사용됩니다.

- 균형 분산 분석 모형
- 주요인 효과만 있는 모형

- 모든 회귀 모형
- 순수 중첩 디자인. (이 중첩 형식은 구문을 통해 지정될 수 있습니다.)

**제 III 유형.** 기본값입니다. 이 방법은 디자인에 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 디자인에서 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 디자인됩니다. 디자인내의 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  $LB=0$ 을 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 선형 도표이며 도표 내의 각 점은 한 요인 수준에서 측정된 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 다변량과 GLM 반복 측도는 Advanced Statistics 옵션을 설치한 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.

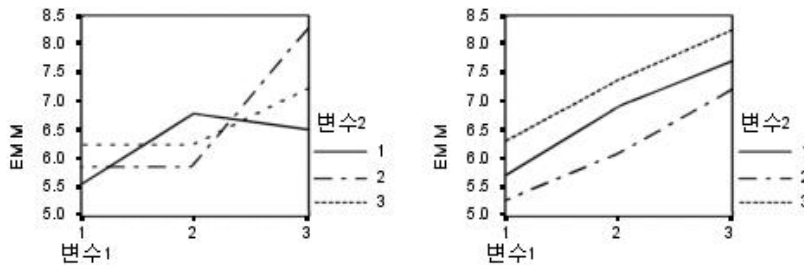


그림 2. 평행하지 않은 도표(왼쪽) 및 평행한 도표(오른쪽)

수평축에 대한 요인을 선택하고 필요에 따라 개별 선과 개별 도표에 대한 요인을 선택하여 도표를 지정하면 도표가 도표 목록에 추가되어야 합니다.

## GLM 옵션

이 대화 상자에서는 선택적 통계를 사용할 수 있습니다. 통계는 모수효과 모형을 사용하여 계산됩니다.

**주변 평균 추정.** 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.
- **신뢰구간 조정.** 신뢰구간과 유의수준에 최소유의차(LSD), Bonferroni, Sidak 조정을 선택합니다. 이 항목은 주효과 비교가 선택되어 있을 경우에만 사용 가능합니다.

**출력.** 전체 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타내려면 기술통계를 선택합니다. 효과 크기 추정값은 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 검정력을 선택합니다. 모수 추정값을 선택하여 각 검정에 대한 모수 추정값, 표준 오차,  $t$  검정, 신뢰구간, 관측 검정력을 생성합니다. 대비계수(L) 행렬을 선택하여 L 행렬을 구합니다.

**동질성 검정**은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서의 각 종속변수에 대해 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 도표를 선택하여 각 종속변수에 대한 관측-예측 표준화 잔차 도표를 작성합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. 적합결여 검정을 선택하여 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인합니다. 일반 추정가능 함수를 사용하여 일반 추정가능 함수를 기준으로 사용자 정의 가설 검정을 작성할 수 있습니다. 대비계수(L) 행렬의 행은 일반 추정가능 함수의 선형결합입니다.

**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. 유의 수준을 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 디자인에 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 변수이름을 지정합니다.



- OUTFILE 하위 명령을 사용하여 상관 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 개체 간 분산 분석 테이블로부터 통계를 포함하는 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 디자인 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 모수요인에 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 다변량과 GLM 반복 측도는 Advanced Statistics 옵션을 설치한 경우에만 사용할 수 있습니다.

Bonferroni와 Tukey의 정직유의차 검정은 보통 다중 비교 검정에 사용됩니다. 스튜던트의  $t$  통계를 기준으로 하는 **Bonferroni** 검정은 다중 비교가 행해지는 사건에 대한 관측유의 수준을 수정합니다. **Sidak의 T** 검정을 사용하는 경우에도 유의 수준이 수정되며 Bonferroni 검정보다 더 강력한 한계를 제공합니다. **Tukey의 정직유의차** 검정은 스튜던트화 범위 통계를 사용하여 모든 그룹 간 쌍대 비교를 수행하며 실험별 오차 비율을 모든 쌍대 비교 컬렉션의 오차 비율로 설정합니다. 다수의 대응평균을 검정할 경우 Tukey의 정직유의차 검정이 Bonferroni 검정보다 효과적인 결과를 나타냅니다. 소수의 대응평균에는 Bonferroni가 보다 효과적입니다.

**Hochberg의 GT2**는 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. 한 요인 수준에서의 평균이 제어 범주에서의 평균보다 작을지를 검정하려면  $<$  통제를 선택합니다. 마찬가지로 한 요인 수준의 평균이 제어 범주에서의 평균보다 큰지를 검정하려면  $>$  통제를 선택합니다.

Ryan, Einot, Gabriel 및 Welsch(R-E-G-W)는 두 가지 다중 하향 범위 검정을 개발하였습니다. 다중 하향 프로시저는 우선 모든 평균이 동일하지 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.

**주변 평균 추정.** 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.
- **신뢰구간 조정.** 신뢰구간과 유의수준에 최소유의차(LSD), Bonferroni, Sidak 조정을 선택합니다. 이 항목은 주효과 비교가 선택되어 있을 경우에만 사용 가능합니다.

**출력.** 전체 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타내려면 기술통계를 선택합니다. 효과 크기 추정값은 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 검정력을 선택합니다. 모수 추정값을 선택하여 각 검정에 대한 모수 추정값, 표준 오차,  $t$  검정, 신뢰구간, 관측 검정력을 생성합니다. 대비계수(**L**) 행렬을 선택하여 **L** 행렬을 구합니다.

**동질성 검정**은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서의 각 종속변수에 대해 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 도표를 선택하여 각 종속변수에

대한 관측-예측-표준화 잔차 도표를 작성합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. 적합결여 검정을 선택하여 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인합니다. 일반 추정가능 함수를 사용하여 일반 추정가능 함수를 기준으로 사용자 정의 가설 검정을 작성할 수 있습니다. 대비계수(L) 행렬의 행은 일반 추정가능 함수의 선형결합입니다.

**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. 유의 수준을 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 디자인에 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 매트릭 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 변수이름을 지정합니다.
- OUTFILE 하위 명령을 사용하여 상관 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 개체 간 분산 분석 테이블로부터 통계를 포함하는 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 디자인 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 Statistics 세션에서 사용할 수 있도록 값을 저장하려면 현재 데이터 파일을 저장해야 합니다.

**예측값.** 각 케이스에 대한 모형 예측값입니다.

- **비표준화.** 모형이 예측하는 종속변수의 값입니다.

- **가중치 부여.** 표준화하지 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.

- **Cook의 거리.** 특정 케이스를 회귀계수 계산에서 제외할 때 모든 케이스의 잔차가 얼마나 변경될 수 있는지에 대한 측도입니다. Cook의 거리가 크면 회귀분석 통계 계산에서 케이스를 제외할 때 계수가 크게 변경됩니다.
- **레버리지 값.** 중심화되지 않은 레버리지 값입니다. 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.

- **비표준화.** 관측값과 모형에 의해 예측된 값 사이 차이입니다.
- **가중치 부여.** 표준화되지 않은 가중치 잔차입니다. WLS 변수가 먼저 선택되어야만 이 항목을 사용할 수 있습니다.
- **표준화.** 표준 편차의 추정값으로 나눈 잔차입니다. Pearson 잔차라고도 하는 표준화 잔차는 평균이 0이고 표준 편차가 1이 됩니다.
- **스튜던트화.** 케이스에 따라 다양한 표준 편차의 추정값으로 나눈 잔차로서 독립변수 평균으로부터의 독립변수의 케이스 값 거리에 따라 달라집니다.
- **삭제됨.** 회귀계수 계산에서 케이스를 제외할 때의 케이스에 대한 잔차로, 종속변수 값과 조정된 예측값 간 차이입니다.

**계수 통계.** 모형에서 모수 추정값의 분산-공분산 행렬을 현재 세션의 새 데이터 세트나 외부 IBM SPSS Statistics 데이터 파일에 씁니다. 또한 각 종속변수에 대해 모수 추정값이 포함된 한 행과 모수 추정값에 해당하는  $t$  통계의 유의수준이 포함된 한 행, 잔차 자유도로 구성된 한 행이 있습니다. 다변량 모형의 경우 각 종속변수에 대해 이와 유사한 행들이 존재합니다. 이 행렬 파일은 행렬 파일을 읽는 다른 프로시저에도 사용할 수 있습니다.

---

## GLM 옵션

이 대화 상자에서는 선택적 통계를 사용할 수 있습니다. 통계는 모수효과 모형을 사용하여 계산됩니다.

**주변 평균 추정.** 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.

- 신뢰구간 조정. 신뢰구간과 유의수준에 최소유의차(LSD), Bonferroni, Sidak 조정을 선택합니다. 이 항목은 주효과 비교가 선택되어 있을 경우에만 사용 가능합니다.

출력. 전체 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타내려면 기술통계를 선택합니다. 효과 크기 추정값은 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 검정력을 선택합니다. 모수 추정값을 선택하여 각 검정에 대한 모수 추정값, 표준 오차,  $t$  검정, 신뢰구간, 관측 검정력을 생성합니다. 대비계수(L) 행렬을 선택하여 L 행렬을 구합니다.

동질성 검정은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서의 각 종속변수에 대해 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 도표를 선택하여 각 종속변수에 대한 관측-예측-표준화 잔차 도표를 작성합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. 적합결여 검정을 선택하여 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인합니다. 일반 추정가능 함수를 사용하여 일반 추정가능 함수를 기준으로 사용자 정의 가설 검정을 작성할 수 있습니다. 대비계수(L) 행렬의 행은 일반 추정가능 함수의 선형결합입니다.

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. 유의 수준을 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 디자인에 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 매트릭 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 변수이름을 지정합니다.
- OUTFILE 하위 명령을 사용하여 상관 행렬 데이터 파일을 구성합니다.

- OUTFILE 하위 명령을 사용하여 개체 간 분산 분석 테이블로부터 통계를 포함하는 행렬 데이터 파일을 구성합니다.
- OUTFILE 하위 명령을 사용하여 디자인 행렬을 새 데이터 파일에 저장합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



---

## 제 12 장 이변량 상관계수

이변량 상관계수 프로시저는 Pearson 상관계수, Spearman의 Rho 및 Kendall의 타우- $b$ 를 계산하고 각 유의 수준과 함께 출력합니다. 상관은 변수나 순위 순서의 관련성을 측정합니다. 상관계수를 계산하기 전에 잘못된 결과를 나타낼 수 있는 이상치의 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.

**예제.** 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.

이변량 상관 데이터 고려 사항

**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 상관 구하기

메뉴에서 다음을 선택합니다.

분석 > 상관분석 > 이변량...

1. 둘 이상의 숫자변수를 선택합니다.

다음 옵션을 사용할 수도 있습니다.

- **상관계수.** 정규 분포된 양적변수에 대해서는 **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. 상관 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.
- **유의수준 검정.** 한쪽 확률이나 양쪽 확률을 선택할 수 있습니다. 연관 방향을 이미 알고 있는 경우에는 한쪽을 선택합니다. 모를 때는 양쪽을 선택합니다.
- **유의한 상관 플래그.** 상관의 유의수준이 0.05이면 별표 하나로 표시되고 0.01이면 별표 두 개로 표시됩니다.



---

## 이변량 상관: 옵션

통계. Pearson 상관계수의 경우 다음 중 하나 또는 모두를 선택할 수 있습니다.

- 평균과 표준 편차. 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  $N-1$ 로 나눈 값과 같습니다.

결측값. 다음 중 하나를 선택할 수 있습니다.

- 대응별 결측값 제외. 한 상관의 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.

---

## 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 상관 행렬을 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.

---

## 제 13 장 편상관계수

편상관계수 프로시저는 한 개 이상의 추가 변수의 효과를 제어하는 동시에 두 변수간 선형 관계를 설명하는 편상관을 계산합니다. 상관은 선형 연관의 척도입니다. 두 변수가 완벽하게 연관될 수는 있지만 두 변수의 관계가 선형이 아니면 상관계수가 연관 측정에 적합한 통계가 될 수 없습니다.

**예제.** 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 유의 수준이 있는 편상관 행렬과 0차 상관 행렬을 선택할 수 있습니다.

편상관 데이터 고려 사항

**데이터.** 대칭이고 양적변수를 사용합니다.

**가정.** 편상관 프로시저는 각 대응변수가 정규 이변량인 것으로 가정합니다.

편상관 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 상관분석 > 편상관...

2. 편상관을 계산할 숫자변수를 둘 이상 선택합니다.

3. 숫자 제어변수를 한 개 이상 선택합니다.

다음 옵션을 사용할 수도 있습니다.

- **유의수준 검정.** 한쪽 확률이나 양쪽 확률을 선택할 수 있습니다. 연관 방향을 이미 알고 있는 경우에는 한 쪽을 선택합니다. 모를 때는 양쪽을 선택합니다.
- **관측 유의 수준 출력.** 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 편상관 행렬과 0차 상관 행렬 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 상관 행렬을 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 상관과는 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 상관 행렬을 읽거나 편상관 행렬을 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 상관 행렬을 표시합니다(STATISTICS 하위 명령 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

---

## 제 14 장 거리

이 프로시저는 대응 케이스나 대응변수 간에 유사성이나 상이성(거리)을 측정하는 여러 통계를 계산합니다. 이러한 유사성이나 거리측도는 요인 분석, 군집 분석, 다차원척도법 등과 마찬가지로 복잡한 데이터 세트를 분석 하는데 사용할 수 있습니다.

**예제.** 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 코사인인 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, 램다, 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...

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.

---

### 거리측도: 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, 블럭, 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 Williams 를 선택할 수 있습니다. 유와 무에 값을 입력하여 의미 있는 두 개의 값을 지정하면 거리측도는 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 거리측도를 통해 생성한 값을 변환할 수 있습니다. 이 값은 거리측도를 계산한 다음 적용됩니다. 옵션으로는 절대값, 부호 바꾸기, 0-1 범위로 척도 조정 등을 사용할 수 있습니다.

---

## 거리측도: 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, 램다, 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 거리측도는 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 거리측도를 통해 생성한 값을 변환할 수 있습니다. 이 값은 거리측도를 계산한 다음 적용됩니다. 옵션으로는 절대값, 부호 바꾸기, 0-1 범위로 척도 조정 등을 사용할 수 있습니다.

---

## PROXIMITIES 명령 추가 기능

거리측도 절차에서는 PROXIMITIES 명령 구문을 사용합니다. 명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- Minkowski 거리측도에 대한 거듭제곱으로 정수를 지정합니다.
- 사용자 정의된 거리측도에 대한 거듭제곱 및 근으로 정수를 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

---

## 제 15 장 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 트리)과 비교하여 잘 이해되며 일반적으로 아주 빨리 작성될 수 있습니다.

**예제.** 주택 소유자의 보험 청구에 대해 조사하기 위한 리소스가 제한된 보험 회사가 보험 청구액을 추정하기 위한 모형을 만들고자 합니다. 이 모형을 서비스 센터에 배포하면 영업 담당자는 고객과 통화하면서 청구 정보를 입력하여 지난 데이터를 토대로 '예상' 청구 비용을 즉시 산출할 수 있습니다. 자세한 정보는 주제를 참조하십시오.

**필드 요구사항.** 대상과 하나 이상의 입력이 있어야 합니다. 기본적으로 모두 또는 없음의 사전 정의된 역할이 있는 필드는 사용되지 않습니다. 대상이 연속형(척도)이어야 합니다. 예측변수(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 옵션을 클릭하여 선택적 작성 및 모형 설정을 지정합니다.
3. 모형 옵션을 클릭하여 활성 데이터 세트에 스코어를 저장하고 외부 파일로 모형을 내보냅니다.
4. 실행을 클릭하여 프로시저를 실행하고 모형 오브젝트를 만듭니다.

---

### 목표

원하는 주요 목표 해당하는 목표를 선택합니다.

- **표준 모형 만들기.** 이 방법은 예측변수를 사용하여 대상을 예측하는 단일 모형을 설정합니다. 일반적으로 표준 모형은 부스팅되었거나 배깁되었거나 큰 데이터 세트 앙상블보다 해석하기 쉽고 스코어링이 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 스코어링합니다. 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 스코어링합니다. 사용 가능한 규칙은 대상의 측정 수준에 따라 다릅니다.

- **아주 큰 데이터 세트의 모형 만들기((IBM SPSS Statistics Server 필요).** 이 방법은 데이터 세트를 별도의 데이터 블록으로 분할하여 앙상블 모형을 작성합니다. 데이터 세트가 너무 커서 위의 모형을 만들 수 없거나 증분 모형 작성의 경우 이 옵션을 선택하십시오. 이 옵션은 표준 모형보다 작성하는 데 시간이 덜 걸릴 수 있지만 스코어링하는 데는 더 오래 걸릴 수 있습니다. 이 옵션에는 IBM SPSS Statistics Server 연결이 필요합니다.

부스팅, 배깅 및 아주 큰 데이터 세트와 관련된 설정은 64 페이지의 『앙상블』을 참조하십시오.

## 기본 사항

**자동으로 데이터 준비.** 이 옵션은 모형의 예측력을 최대화하기 위해 프로시저에서 대상 및 예측변수를 내부적으로 변환할 수 있습니다. 모든 변형은 모형과 함께 저장되며 스코어링을 위해 새 데이터에 적용됩니다. 변환된 필드의 원래 버전은 모형에서 제외됩니다. 기본적으로 다음 자동 데이터 준비가 수행됩니다.

- **날짜 및 시간 처리.** 각 날짜 예측변수가 참조 날짜(1970-01-01) 이후의 경과 시간이 포함된 새 연속형 예측변수로 변환됩니다. 각 시간 예측변수가 참조 시간(00:00:00) 이후의 경과 시간이 포함된 새 연속형 예측변수로 변환됩니다.
- **측정 수준 조정.** 고유 값이 5개 미만인 연속형 예측변수가 순서 예측변수로 다시 캐스팅됩니다. 고유 값이 10개보다 많은 순서 예측변수가 연속형 예측변수로 다시 캐스팅됩니다.
- **이상치 처리.** 절사 값을 넘는 연속형 예측변수 값(평균에서 3배 표준 편차)이 절사 값으로 설정됩니다.
- **결측값 처리.** 명목 예측변수의 결측값이 훈련 파티션의 최빈값으로 대체됩니다. 순서 예측변수의 결측값이 훈련 파티션의 중앙값으로 대체됩니다. 연속형 예측변수의 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 예측변수로서 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 향으로 사용 가능한 예측변수를 모두 입력하는 모든 예측변수 포함을 선택하십시오. 기본적으로 단계적 전진이 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 통계는 모형 오차에서 향상도의 통계 검정을 기준으로 합니다. 수정된 R-제곱은 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 우도를 기준으로 하며 너무 복잡한 모형에 페널티를 부여하도록 조정됩니다. 수정된 R-제곱은 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 작성하는데 상당히 많은 시간이 걸릴 수 있습니다.

---

## 양상블

이 설정은 부스팅, 배깅 또는 아주 큰 데이터 세트가 목표에서 요청될 때 발생하는 양상블 동작을 결정합니다. 선택한 목표에 해당하지 않는 옵션은 무시됩니다.

배깅 및 아주 큰 데이터 세트. 양상블을 스코어링할 때 양상블 스코어값을 계산하기 위해 기본 모형에서 예측값을 조합하는 데 사용되는 규칙입니다.

- 연속형 대상의 기본 조합 규칙. 연속형 대상에 대한 양상블 예측값은 기본 모형의 예측값 평균 또는 중앙값을 사용하여 조합될 수 있습니다.

모형 정확도를 향상시키는 것이 목표인 경우 조합 규칙 선택이 무시됩니다. 부스팅은 항상 가중 다수 투표를 사용하여 범주형 대상을 스코어링하고 가중 중앙값을 사용하여 연속형 대상을 스코어링합니다.

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 생성을 클릭하십시오. 기본값은 54752075입니다.

---

## 모형 옵션

데이터 세트에 예측값 저장. 기본 변수 이름은 *PredictedValue*입니다.

모형 내보내기. 외부 .zip 파일에 모형을 씁니다. 스코어링을 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.

차트. 차트는 최종 모형의 정확도를 표시하며 더 크게 표시된 것이 더 나은 형식입니다. 최종 모형의 경우 값은  $100 \times$  수정된  $R^2$ 입니다.

---

## 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.
- **이상치 자름** 절사 값을 넘는 연속형 예측변수 값(평균에서 3배 표준 편차)을 절사 값으로 설정합니다.
- **결측값 바꾸기** 명목 필드의 결측값을 최빈값으로, 순서 필드를 중앙값으로, 연속형 필드를 평균으로 바꿉니다.
- **범주를 합쳐** 대상과의 연관 최대화 입력과 대상 간의 관계를 기반으로 '유사한' 예측변수 범주를 식별합니다. 유의적으로 다르지 않은 범주(즉 0.05보다 큰  $p$ -값을 가지는 범주)가 합쳐집니다.
- **상수 예측변수 제외 / 이상치 처리 후 / 범주 합치기 후** 다른 ADP 작업을 마친 후 단일 값을 가진 예측변수를 제거합니다.

---

## 예측변수 중요도

일반적으로 가장 중요한 예측변수 필드에 모델링 노력을 집중하고 가장 쓸모 없는 예측변수를 삭제하거나 무시하기를 원합니다. 예측변수 중요도 차트를 사용하면 모형 추정 시 각 예측변수의 상대적인 중요도를 표시하여 원하는 작업을 수행할 수 있습니다. 값이 상대적이므로 표시된 모든 예측변수에 대한 값의 합은 1.0이 됩니다. 예측변수 중요도는 모형 정확도와는 관련이 없습니다. 단지 예측 시 각 예측변수의 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 차트를 표시합니다.

**차트 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.

## 이상치

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 이상치를 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.** 예측변수 중요도를 줄여 효과가 위에서 아래로 정렬되는 차트입니다. 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. 기본값입니다.
- **테이블.** 전체 모형과 개별 모형 효과에 대한 ANOVA 테이블입니다. 예측변수 중요도를 줄여 개별 효과가 위에서 아래로 정렬됩니다. 기본적으로는 테이블이 축소되어 전체 모형의 결과만 보여줍니다. 개별 모형 효과의 결과를 보려면 테이블에서 수정된 모형 셀을 클릭합니다.

**예측변수 중요도.** 보기에 표시되는 예측변수를 제어하는 예측변수 중요도 슬라이더가 있습니다. 모형을 바꾸지는 않지만 가장 중요한 예측변수에 집중할 수 있습니다. 기본적으로 상위 10개 효과가 표시됩니다.

**유의수준.** 예측변수 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 예측변수)이 모형 내에서 코딩된 지표이므로 요인을 포함하는 효과에는 일반적으로 여러 관련 계수가 있으며, 중복(참조) 모수에 해당하는 범주를 제외하고 각 범주에 하나씩 있습니다.

**유형.** 여러 가지 표시 유형이 있으며 유형 드롭다운 목록에서 액세스할 수 있습니다.

- **다이아그램.** 먼저 절편을 표시한 다음 예측변수 중요도를 줄여 위에서 아래로 효과를 정렬하는 차트입니다. 요인을 포함하는 효과 내에서 계수가 데이터 값의 오름차순으로 정렬됩니다. 다이아그램의 연결선은 계수의 부호에 따라 색상이 지정되고(다이아그램 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.
- **테이블.** 개별 모형 계수의 값, 유의수준 검정 및 신뢰구간을 표시합니다. 절편 이후, 예측변수 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 예측변수의 경우) 모형 모수와 연결된 값 레이블을 보여주는 도구 팁이 표시됩니다. 이것은 자동 데이터 준비에서 유사한 범주의 범주형 예측변수를 합칠 때 만들어지는 새 범주를 보려는 경우에 특히 유용합니다.

**예측변수 중요도.** 보기에 표시되는 예측변수를 제어하는 예측변수 중요도 슬라이더가 있습니다. 모형을 바꾸지는 않지만 가장 중요한 예측변수에 집중할 수 있습니다. 기본적으로 상위 10개 효과가 표시됩니다.

**유의수준.** 예측변수 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 예측변수에 대해 표시되는 차트입니다. 차트는 다른 모든 예측변수를 상수로 유지하면서 수직축에 대상의 모형 추정값을 표시하고 수평축에 예측변수의 각 값을 표시합니다. 대상에서 각 예측변수 계수의 효과를 시각화하는 데 유용합니다.

**참고:** 유의한 예측변수가 없는 경우 평균 추정이 생성되지 않습니다.

---

## 모형 작성 요약

모형 선택 알고리즘으로 모형 선택 설정에서 없음 외에 다른 항목을 선택한 경우 모형 작성 프로세스에 대한 세부 정보를 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.



## 제 16 장 선형 회귀분석

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.

**예제.** 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, 상관 행렬, 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-공분산 행렬, 분산팽창요인, 허용 오차, Durbin-Watson 검정, 거리 측도(Mahalanobis, Cook, 레버리지 값), DFBETA, DFFIT, 예측구간, 케이스별 진단 정보 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...

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.
- 단계적(*Stepwise*). 각 단계에서 최소 확률 F를 가지면서 방정식에 없는 독립변수가 입력됩니다. 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*). 방정식에 모든 변수를 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 모형에 순차적으로 입력되는 단계별 변수 선택 프로시저입니다. 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 다음을 클릭합니다.

## 선형 회귀분석 규칙 설정

선택 규칙으로 정의한 케이스는 분석에 포함됩니다. 예를 들어, 변수를 선택하고 같음을 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 이상치, 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(\*ZPRED와 \*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.

- 비표준화. 모형이 예측하는 종속변수의 값입니다.
- 표준화. 각 예측값을 표준화 형식으로 변환합니다. 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.
- 평균예측 표준 오차. 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 전체 케이스 평균과 얼마나 달라지는지에 대한 측도입니다. *Mahalanobis* 거리가 크면 케이스가 독립변수 하나 이상에 대해 극단값을 갖는 것으로 식별합니다.

- *Cook의 거리*. 특정 케이스를 회귀계수 계산에서 제외할 때 모든 케이스의 잔차가 얼마나 변경될 수 있는지에 대한 척도입니다. *Cook의 거리*가 크면 회귀분석 통계 계산에서 케이스를 제외할 때 계수가 크게 변경됩니다.
- *레버리지 값*. 회귀분석의 적합도에 점이 미치는 영향을 측정합니다. 중심화된 레버리지 값의 범위는 0(적합도에 아무 영향도 미치지 않음)부터  $(N-1)/N$ 까지입니다.

**예측 구간**. 평균과 개별 예측 구간에 대한 상한 및 하한 경계를 말합니다.

- *평균(Mean)*. 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*. 1과 99.99 사이의 값을 입력하여 두 예측 구간에 대한 신뢰수준을 지정합니다. 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.

- *비표준화*. 관측값과 모형에 의해 예측된 값 사이 차이입니다.
- *표준화*. 표준 편차의 추정값으로 나눈 잔차입니다. Pearson 잔차라고도 하는 표준화 잔차는 평균이 0이고 표준 편차가 1이 됩니다.
- *스튜던트화*. 케이스에 따라 다양한 표준 편차의 추정값으로 나눈 잔차로서 독립변수 평균으로부터의 독립변수의 케이스 값 거리에 따라 달라집니다.
- *삭제됨*. 회귀계수 계산에서 케이스를 제외할 때의 케이스에 대한 잔차로, 종속변수 값과 조정된 예측값 간 차이입니다.
- *삭제된 스튜던트화*. 표준 오차로 나뉘어진 케이스에 대한 삭제된 잔차입니다. 삭제된 스튜던트화 잔차와 관련된 스튜던트화 잔차 간 차이는 케이스가 예측을 통해 그 차이를 얼마나 해소하는지를 나타냅니다.

**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.

- *DFBETA*. 베타 값의 차이는 특정 케이스를 제외하여 나타난 회귀계수의 변화량을 나타냅니다. 값은 모형의 각 항에 대해 계산되며 상수를 포함합니다.
- *표준화 DFBETA(Standardized DfBeta)*. 베타 값의 표준화된 차이입니다. 특정 케이스를 제외하여 나타난 회귀계수의 변화량을 나타냅니다. 절대값이 2보다 큰 케이스를  $N$  제곱근( $N$ 은 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 표준화된 차이입니다. 특정 케이스를 제외하여 나타난 예측값의 변화량을 나타냅니다. 절대값이 2보다 큰 표준화 값을  $p/N$  제곱근( $p$ 는 모형에서 모수의 수이고  $N$ 은 케이스 수)으로 나누어 결과 값을 확인합니다.
- *공분산 비율*. 회귀계수 계산에서 모든 케이스가 포함된 공분산 행렬 행렬식에 대한 특정 케이스가 제외된 공분산 행렬 행렬식의 비율입니다. 이 비율이 1에 가까우면 케이스로 인해 공분산 행렬이 크게 달라지지 않습니다.

계수 통계. 데이터 세트 또는 데이터 파일에 회귀계수를 저장합니다. 데이터 세트는 같은 세션에서 다시 사용할 수 있으나 세션을 마치고 전에 명시적으로 저장하지 않는 한 파일로 저장되지 않습니다. 데이터 세트 이름은 변수이름 규칙을 준수해야 합니다.

**XML 파일에 모형정보 내보내기.** 모수 추정값과 필요에 따라 해당 공분산을 XML(PMML) 형식의 지정된 파일로 내보냅니다. 스코어링을 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.

---

## 선형 회귀분석 통계

다음의 통계를 사용할 수 있습니다.

**회귀계수.** 추정값에서는 회귀계수  $B$ , 표준 오차  $B$ , 표준화 계수 베타,  $B$ 에 대한  $t$  값 및 양방향 유의 수준  $t$ 를 표시합니다. 신뢰구간에서는 각 회귀계수 또는 공분산 행렬에 대해 지정된 신뢰 수준을 가진 신뢰구간을 표시합니다. 공분산 행렬은 대각선이 있으면 분산을 사용하고 대각선이 없으면 공분산을 사용하여 회귀계수의 분산-공분산 행렬을 표시합니다. 상관 행렬도 표시됩니다.

**모형 적합.** 모형에 진입되었거나 모형에서 제거한 변수를 나열하며 다음과 같은 다중  $R$ ,  $R^2$  및 수정된  $R^2$ , 추정값의 표준 오차 및 분산 분석표의 적합도 통계량이 표시됩니다.

**R 제곱 변화량.** 독립변수를 추가하거나 삭제하여 생성되는  $R^2$  통계의 변화량입니다. 하나의 변수와 연결된  $R^2$  변화량이 큰 경우 해당 변수는 종속변수의 예측변수로 적합합니다.

**기술통계.** 각 변수에 대해 유효한 케이스 수, 평균, 표준 편차 등을 분석에서 사용할 수 있습니다. 한쪽 방향 유의 수준이 있는 상관 행렬과 각 상관에 대한 케이스 수도 표시됩니다.

**편상관(Partial Correlation).** 다른 변수의 상호 연관으로 인해 발생한 상관을 제거한 후에도 두 변수 간에 남아 있는 상관입니다. 또한 종속변수와 독립변수 모두에서 모형에 있는 다른 독립변수의 선형 효과를 제거할 때 발생하는 종속변수와 독립변수 간 상관을 나타냅니다.

**부분상관(Part Correlation).** 독립변수에서 모형에 있는 다른 독립변수의 선형 효과를 제거할 때 발생하는 종속변수와 독립변수 간 상관을 나타냅니다. 변수가 방정식에 추가될 때의 R 제곱 변화량과 관련이 있습니다. 준편상관이라고도 합니다.

**공선성 진단.** 공선성 또는 다중공선성은 하나의 독립변수가 다른 독립변수의 선형 함수가 되는 부적절한 상황입니다. 척도와 비중심화 교차곱 행렬, 조건 지수, 분산-분해 비율 등의 고유 값이 분산팽창요인(VIF)과 개별 변수에 대한 허용 오차와 함께 표시됩니다.

**잔차.** 잔차의 연속 상관에 대해 Durbin-Watson 검정을 표시하고 선택 기준( $n$  표준 편차 위의 이상치)을 만족하는 케이스에 대해 대응별 진단 정보를 표시합니다.

---

## 선형 회귀분석 옵션

다음 옵션을 사용할 수 있습니다.

선택법 기준. 이러한 옵션은 전진, 후진, 단계별 변수 선택법을 지정할 때 적용됩니다. 변수는  $F$  값의 유의수준 (확률)이나  $F$  값 자체에 따라 모형에서 진입하거나 삭제할 수 있습니다.

- $F$  확률 사용(Use Probability of  $F$ ).  $F$  값의 유의 수준이 진입값보다 작으면 모형에 변수가 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).  $F$  값이 진입값보다 크면 모형에 변수가 입력되고  $F$  값이 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.

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,  $R^2$ 은 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 상관 행렬을 쓰거나 행렬을 읽어 회귀분석을 연습니다(MATRIX 하위 명령 사용).
- CRITERIA 하위 명령을 사용하여 허용 오차 수준을 지정합니다.
- 같거나 다른 종속 변수에 대한 여러 모형을 연습니다(METHOD 및 DEPENDENT 하위 명령 사용).
- 추가 통계를 연습니다(DESCRIPTIVES 및 STATISTICS 하위 명령 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



## 제 17 장 순서 회귀분석

순서 회귀분석을 사용하여 예측변수 세트(요인 또는 공변량)에 대한 다중 순서 반응의 종속성을 모델링할 수 있습니다. 순서 회귀분석 디자인은 McCullagh(1980, 1998) 방법을 기초로 하며 이 순서 회귀분석 프로시저는 구문에서 PLUM이라고도 합니다.

표준 선형 회귀분석은 반응(종속) 변수와 예측(독립) 변수 기중 결합간의 차이 제곱합 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.

**예제.** 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 공변량 행렬, 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...

2. 종속변수를 하나 선택합니다.

3. 확인을 클릭합니다.

---

## 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 예측변수 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.

- **로지.**  $f(x)=\log(x/(1-x))$ . 일반적으로 고르게 분포된 범주에 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 분포인 경우 사용됨.
- **Cauchit(역 Cauchy).**  $f(x)=\tan(\pi(x-0.5))$ . 일반적으로 잠복 변수에 극단값이 많은 경우 사용됨.

---

## 순서 회귀분석 결과

결과 대화 상자에서 뷰어에 표시할 테이블을 작성하고 변수를 작업 파일에 저장할 수 있습니다.

**출력.** 다음에 대한 테이블을 작성합니다.

- **반복 히스토리 인쇄.** 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 상관 행렬입니다.
- **모수 추정값의 근사 공분산.** 모수 추정값의 공분산 행렬입니다.



- **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.

저장된 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 인쇄.** 로그-우도의 출력을 제어합니다. 다항 상수항 포함을 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 공변량 주효과와 요인 주효과는 포함되지만 상호작용 효과는 포함되지 않습니다. 사용자 정의 모형을 작성하여 요인 상호작용이나 공변량 상호작용 서브세트를 지정할 수 있습니다.

**요인/공변량.** 요인 및 공변량이 나열됩니다.

**위치 모형.** 모형은 선택한 주효과와 상호작용 효과에 따라 다릅니다.

### 항 설정

선택된 요인과 공변량에 대해 다음과 같습니다.

**상호작용.** 선택한 모든 변수의 최고-수준 상호작용 항을 생성합니다. 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.

---

## 순서 회귀분석 척도 모형

척도 대화 상자에서 분석에 대한 척도 모형을 지정할 수 있습니다.

**요인/공변량.** 요인 및 공변량이 나열됩니다.

**척도 모형.** 모형은 선택한 주효과와 상호작용 효과에 따라 다릅니다.

### 항 설정

선택된 요인과 공변량에 대해 다음과 같습니다.

**상호작용.** 선택한 모든 변수의 최고-수준 상호작용 항을 생성합니다. 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.

---

## 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.

---

## 제 18 장 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.

**예제.** 인터넷 서비스 공급자가 해당 네트워크에서 바이러스에 감염된 채로 전달되는 전자 메일의 트래픽 백분율을 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 시간을 선택하면 곡선 추정 프로시저를 통해 케이스 간 시간 길이가 균일한 시간변수가 작성됩니다. 시간을 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...

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 차트가 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  $Y = b_0 + (b_1 * t)$ 입니다. 계열 값은 시간의 선형 함수로 모델링됩니다.

**로그 변환.** 모형식은  $Y = b_0 + (b_1 * \ln(t))$ 입니다.

**역.** 모형식은  $Y = b_0 + (b_1 / t)$ 입니다.

**2차.** 모형식은  $Y = b_0 + (b_1 * t) + (b_2 * t^{**2})$ 입니다. 이 2차 모형은 "줄어드는" 계열 또는 감소하는 계열을 모델링하는 데 사용할 수 있습니다.

**3차.** 모형식은  $Y = b_0 + (b_1 * t) + (b_2 * t^{**2}) + (b_3 * t^{**3})$ 입니다.

**거듭제곱.** 모형식은  $Y = b_0 * (t^{**b_1})$  또는  $\ln(Y) = \ln(b_0) + (b_1 * \ln(t))$ 입니다.

**복합.** 모형식은  $Y = b_0 * (b_1^{**t})$  또는  $\ln(Y) = \ln(b_0) + (\ln(b_1) * t)$ 입니다.

**S-곡선.** 모형식은  $Y = e^{**}(b_0 + (b_1/t))$  또는  $\ln(Y) = b_0 + (b_1/t)$ 입니다.

**로지스틱.** 모형식은  $Y = 1 / (1/u + (b_0 * (b_1^{**t})))$  또는  $\ln(1/y-1/u) = \ln(b_0) + (\ln(b_1) * t)$ 로,  $u$ 는 상한값을 나타냅니다. 로지스틱을 선택한 후 상한값을 지정하여 회귀 방정식에서 사용합니다. 이 값은 양수여야 하며 최대 종속변수 값보다 커야 합니다.

**성장.** 모형식은  $Y = e^{**}(b_0 + (b_1 * t))$  또는  $\ln(Y) = b_0 + (b_1 * t)$ 입니다.

**지수.** 모형식은  $Y = b_0 * (e^{**}(b_1 * t))$  또는  $\ln(Y) = \ln(b_0) + (b_1 * t)$ 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 시간을 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.



---

## 제 19 장 일부 최소제곱 회귀분석

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 예측변수가 서로 밀접하게 상호 관련되어 있거나 예측변수 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.

PLS(Partial Least Squares) 회귀분석 데이터 고려 사항

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 저장됩니다, 첫 번째 범주는 (1,0,...,0), 다음 범주는 (0,1,0,...,0), ..., 그리고 마지막 범주는 (0,0,...,0,1)로 나타냅니다. 범주형 종속변수는 더미 코딩을 사용하여 나타냅니다. 즉, 참조 범주에 해당하는 지표를 생략합니다.

빈도 가중치. 가중치는 사용하기 전에 가장 가까운 정수로 반올림됩니다. 결측 가중치나 0.5 미만의 가중치를 갖는 케이스는 분석에 사용되지 않습니다.

결측값. 사용자 및 시스템-결측값을 유효하지 않은 값으로 처리함

척도 조정. 모든 모형 변수는 범주형 변수를 나타내는 지시변수를 포함하여 가운데 위치하고 표준화됩니다.

PLS(Partial Least Squares) 회귀분석 얻기

메뉴에서 다음을 선택합니다.

분석 > 회귀분석 > PLS(Partial Least Squares)...

1. 종속변수를 한 개 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.

## 전제조건

부분 최소제공 회귀분석 프로시저는 Python 확장 명령이며 기본적으로 IBM SPSS Statistics 제품과 함께 설치되는 IBM SPSS Statistics - Essentials for Python가 필요합니다. 무료로 사용 가능한 NumPy 및 SciPy Python 라이브러리도 필요합니다.

**참고:** 분산 분석 모드로 작업 중인 사용자(IBM SPSS Statistics Server 필수)의 경우, 서버에 NumPy 및 SciPy가 설치되어 있어야 합니다. 시스템 관리자에게 지원을 요청하십시오.

### Windows 및 Mac 사용자

Windows 및 Mac의 경우, IBM SPSS Statistics와 함께 설치된 버전과 별개인 Python 2.7 버전에 NumPy 및 SciPy를 설치해야 합니다. 별도의 Python 2.7 버전이 없는 경우 <http://www.python.org>에서 다운로드할 수 있습니다. 그런 다음, Python 버전 2.7의 NumPy 및 SciPy를 설치하십시오. 설치 프로그램은 <http://www.scipy.org/Download>에 있습니다.

NumPy 및 SciPy를 사용하도록 설정하려면 Python 위치를 NumPy 및 SciPy가 설치된 Python 2.7 버전으로 설정해야 합니다. Python 위치는 옵션 대화 상자(편집 > 옵션)의 파일 위치 탭에서 설정합니다.

### Linux 사용자

소스를 다운로드하여 직접 NumPy 및 SciPy를 빌드하도록 권장합니다. 소스는 <http://www.scipy.org/Download>에 있습니다. IBM SPSS Statistics와 함께 설치된 Python 2.7 버전에 NumPy 및 SciPy를 설치할 수 있습니다. IBM SPSS Statistics가 설치된 위치의 Python 디렉토리에 있습니다.

IBM SPSS Statistics와 함께 설치된 버전 이외의 Python 2.7 버전에 NumPy 및 SciPy를 설치하도록 선택하는 경우 해당 버전을 가리키도록 Python 위치를 설정해야 합니다. Python 위치는 옵션 대화 상자(편집 > 옵션)의 파일 위치 탭에서 설정합니다.

### Windows 및 Unix 서버

IBM SPSS Statistics와 함께 설치된 버전과 별개인 Python 2.7 버전에 NumPy 및 SciPy를 서버에서 설치해야 합니다. 서버에 별개의 Python 2.7 버전이 없는 경우에는 <http://www.python.org>에서 다운로드할 수 있습니다. Python 2.7의 NumPy 및 SciPy는 <http://www.scipy.org/Download>에서 제공됩니다. NumPy 및 SciPy를 사용하려면 서버의 Python 위치를 NumPy 및 SciPy가 설치된 Python 2.7 버전으로 설정해야 합니다. Python 위치는 IBM SPSS Statistics Administration Console에서 설정됩니다.

---

## 모형

**모형 효과 지정.** 주효과 모형은 모든 요인과 공변량 주효과를 포함합니다. 사용자 정의를 선택하여 상호작용을 지정합니다. 모형에 포함할 모든 항을 지정해야 합니다.

**요인 및 공변량.** 요인 및 공변량이 나열됩니다.

**모형.** 모형은 데이터의 특성에 따라 다릅니다. 사용자 정의를 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. 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 예측변수에 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 요인 스코어도 도표화합니다.
- **잠재 요인에 대한 추정값을 저장합니다.** 잠재 요인 로딩과 잠재 요인 기중치를 저장합니다. 잠재 요인 기중치도 도표화합니다.
- **독립변수에 대한 추정값을 저장합니다.** 회귀 모수 추정값과 투영에서 변수 중요도(VIP)를 저장합니다. 잠재 요인별로 VIP도 도표화합니다.



## 제 20 장 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 (가장 가까운 이웃)이 기록되고 가장 많은 수의 가장 가까운 이웃이 포함되어 있는 범주에 새 케이스가 놓입니다.

검사할 최근접 이웃의 수를 지정할 수 있습니다. 이 값은  $k$ 라고 합니다.

또한 최근접 이웃 분석은 연속적인 대상 값을 계산하는 데 사용할 수 있습니다. 이 경우, 가장 가까운 이웃의 평균 또는 중앙값 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 범주를 나타내는 경우 해당 변수는 명목으로 취급될 수 있습니다. 예를 들어, 직원이 근무하는 회사의 부서가 있습니다. 종교, 우편번호 또는 종교 단체 등이 명목 변수에 해당합니다.
- **순서(Ordinal)**. 변수의 값이 고유한 순위가 있는 범주를 나타내는 경우 해당 변수는 순서로 취급될 수 있습니다. 예를 들어, 매우 불만족에서 매우 만족에 이르는 서비스 만족도 수준이 있습니다. 순서변수의 예로는 만족도나 신뢰도를 나타내는 태도 스코어 및 선호도 등급 스코어가 있습니다.
- **척도(Scale)**. 해당 값이 의미 있는 메트릭으로 순서가 지정된 범주를 나타내므로 값 간 거리 비교가 적합한 경우 해당 변수는 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 변수 유형을 식별합니다.

표 1. 측정 수준 아이콘

	숫자	문자열	날짜	시간
척도(연속형)		해당 없음		
순서				

표 1. 측정 수준 아이콘 (계속)

	숫자	문자열	날짜	시간
명목				

범주형 변수 코딩. 프로시저는 임시로 프로시저 기간 동안  $c$  코딩 중 하나를 사용하여 범주형 예측변수를 다시 코딩합니다. 변수의  $c$  범주가 있는 경우 변수는  $c$  벡터로 저장됩니다. 첫 번째 범주는 (1,0,...,0), 다음 범주는 (0,1,0,...,0), ..., 그리고 마지막 범주는 (0,0,...,0,1)로 나타냅니다.

이 코딩 체계는 기능 공간의 차원을 증가시킵니다. 특히 총 차원 수는 모든 범주형 예측변수에 대한 척도 예측 변수와 수 범주의 수를 합한 값입니다. 그 결과 이 코딩 체계로 인해 훈련이 늦어지게 됩니다. 가장 가까운 이웃 훈련이 매우 느리게 진행될 경우 프로시저를 진행하기 전에 유사한 범주를 조합하거나 거의 사용하지 않는 범주를 가진 케이스를 제외하여 범주형 예측변수에서 범주의 수를 줄일 수 있습니다.

검증용 표본이 정의되어 있다 해도 모든  $c$  코딩 중 하나는 훈련 데이터를 기준으로 합니다(90 페이지의 『파티션』 참조). 따라서 검증용 표본에는 훈련 데이터에 나타나지 않는 예측변수 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(90 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 복제. 이 프로시저에서는 파티션의 임의 할당 및 교차 검증 중첩 과정에서 난수 생성을 사용합니다. 결과를 정확하게 복제하고 동일한 프로시저 설정을 사용하려면 Mersenne Twister에 대한 시드(90 페이지의 『파티션』 참조)를 설정하거나 변수를 사용하여 파티션 및 교차 검증 중첩을 정의하십시오.

최근접 이웃 분석을 얻으려면

메뉴에서 다음을 선택합니다.

분석 > 분류 > 가장 가까운 이웃 항목...

1. 대상이 있을 경우 독립변수 또는 예측변수로 간주할 수 있는 하나 이상의 기능을 지정하십시오.

대상(선택적). 대상(종속변수 또는 반응변수)이 지정되지 않은 경우 프로시저는 가장 가까운 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,  $[2*(x_{\text{최소}})/(x_{\text{최대}})]-1$ 이 사용됩니다. 조정된 정규화 값은 -1과 1 사이의 값에 해당합니다.

**초점 케이스 식별자(선택적).** 초점 케이스 식별자를 사용하면 원하는 케이스를 표시할 수 있습니다. 예를 들어, 조사자는 한 학교 구역의 시험 스코어인 초점 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 스코어를 가장 가까운 이웃의 학교 구역과 비교합니다.

초점 케이스는 또한 임상 연구에도 사용되어 임상 케이스와 유사한 제어 케이스를 선택할 수 있습니다. 초점 케이스는 가장 가까운 이웃 ( $k$ ) 및 거리 표, 기능 공간 차트, 피어 차트 및 4분위 맵에 표시됩니다. 초점 케이스에 대한 정보는 결과 탭에 지정된 파일에 저장됩니다.

지정된 변수에 대한 양의 값을 가진 케이스는 초점 케이스로 처리됩니다. 이는 양의 정수가 없는 변수를 지정하는 데 유효한 값입니다.

**케이스 레이블(선택적).** 케이스는 기능 공간 차트, 피어 차트 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

**가장 가까운 이웃( $k$ ) 수.** 가장 가까운 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. 가장 가까운 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 적도일 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. 예측변수에 대한 기능 중요도는 오차 비율 또는 모형에서 오차 비율을 없앤 예측변수를 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, 가장 가까운 이웃의 중앙값을 기준으로 계산하는지 여부를 지정합니다.

## 기능

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 95 페이지의 『필드선택 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 표본은 가장 가까운 이웃 모형을 훈련하는 데 사용되는 데이터 레코드로 구성되어 있으며 데이터 세트의 일부 케이스 백분율은 모형을 얻기 위해 훈련 표본에 할당해야 합니다. 검증용 표본은 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 백분율을 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 케이스를 분석에서 제외합니다. 분할변수에 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 V까지 번호가 매겨집니다.
- 변수를 사용하여 케이스를 할당합니다. 활성 데이터 세트에 있는 각 케이스를 중첩에 할당하는 숫자 값을 지정합니다. 변수는 숫자이어야 하며 1부터 V까지의 값을 가집니다. 이 범위 내의 모든 값이 누락되어 있고 모든 분할 값에서 분할 파일이 유효한 경우 오차가 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 시드를 설정하면 난수 생성기의 현재 상태가 유지되고 분석이 완료되면 해당 상태로 복원된다는 차이점이 있습니다.

---

## 저장

저장된 변수의 이름. 자동 이름 생성은 모든 작업을 유지할 수 있도록 해줍니다. 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 팀의 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 결과

- 케이스 처리 요약. 케이스 진행 요약 테이블을 표시합니다. 케이스 진행 요약 테이블에는 분석에서 포함된 케이스 수와 제외된 케이스 수가 합계 및 훈련 및 검증용 표본별로 요약되어 표시됩니다.
- 차트 및 테이블. 테이블과 차트를 포함한 모형 관련 결과를 표시합니다. 모형 보기에 있는 테이블에는 초점 케이스, 범주형 반응변수 및 오차 요약에 대한 가장 가까운 이웃( $k$ )과 거리가 포함됩니다. 모형 보기의 그래픽 결과에는 선택 오차 로그, 기능 중요도 차트, 기능 공간 차트, 피어 차트 및 4분위 맵이 포함됩니다. 자세한 정보는 『모형 보기』 주제를 참조하십시오.

### 파일

- XML에 모형 내보내기. 스코어링을 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다. 분할 파일이 정의된 경우에는 이 옵션을 사용할 수 없습니다.
- 초점 케이스와 가장 가까운 이웃( $k$ ) 간의 거리 내보내기. 각 초점 케이스의 경우 각 초점 케이스의 가장 가까운 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 차트 및 테이블을 선택하면 프로시저는 뷰어에서 가장 가까운 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 차트를 보여줍니다. 변수 중요도 차트를 사용할 수 없는 경우, 즉 기능 탭에서 중요도별 가중치 기능이 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 차트는 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 차트 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. 초점 케이스는 연결된 가장 가까운 이웃  $k$ 에 나타납니다.

제어 및 상호작용성 차트에 있는 제어 수를 통해 기능 공간을 탐색할 수 있습니다.

- 차트에 나타나는 서브셋 기능군에서 특정 기능을 선택할 수 있으며 차원에 표시된 기능을 변경할 수 있습니다.
- “초점 케이스”는 기능 공간 차트에서 선택한 간단한 점이 됩니다. 초점 케이스 변수를 지정한 경우 초점 케이스를 나타내는 점이 처음으로 선택됩니다. 그러나 어떠한 점도 사용자가 선택하면 임시로 초점 케이스가 될 수 있습니다. 점 선택 시 “일반적인” 제어가 적용됩니다. 점을 클릭하면 해당 점을 선택하며 모든 다른 모든 점을 선택 취소합니다. 점을 컨트롤-클릭하면 해당 점을 선택한 점 세트에 추가합니다. 피어 차트와 같은 연결된 보기는 기능 공간에서 선택한 케이스를 기준으로 자동으로 업데이트합니다.
- 가장 가까운 이웃 수( $k$ )를 변경하여 초점 케이스에 표시할 수 있습니다.
- 차트에서 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 표시를 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 표시를 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 템플릿에 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 차트를 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.

## 피어

이 차트에서는 초점 케이스와 각 기능 및 대상에 대한 가장 가까운 이웃( $k$ )이 표시됩니다. 기능 공간에서 초점 케이스를 선택한 경우 이 도표를 사용할 수 있습니다.

링크 동작. 피어 차트는 두 가지 방법으로 기능 공간에 연결됩니다.

- 기능 공간에서 선택한 케이스(초점)은 가장 가까운 이웃( $k$ )과 함께 피어 차트에 표시됩니다.
- 기능 공간에서 선택한  $k$  값은 피어 차트에 사용됩니다.

## 가장 가까운 이웃 거리

이 테이블에는 가장 가까운 이웃( $k$ )과 초점 케이스의 거리만 표시됩니다. 변수 탭에 초점 케이스 식별자가 지정된 경우 이 테이블을 사용할 수 있으며 이 변수를 기준으로 식별한 초점 케이스만 표시합니다.

각 행:

- 초점 케이스 열에는 초점 케이스에 대한 케이스 레이블 변수의 값이 포함됩니다. 케이스 레이블이 정의되지 않은 경우 이 열에는 초점 케이스의 케이스 번호가 포함됩니다.
- 가장 가까운 이웃 그룹 아래에 있는  $i$ 번째 열에는 초점 케이스의  $i$ 번째로 가장 가까운 이웃에 대한 케이스 레이블 변수의 값이 포함됩니다. 케이스 레이블이 정의되지 않은 경우 이 열에는 초점 케이스의  $i$ 번째로 가장 가까운 이웃의 케이스 번호가 포함됩니다.
- 가장 가까운 이웃 그룹 아래에 있는  $i$ 번째 열에는 초점 케이스에 대해  $i$ 번째로 가장 가까운 이웃이 포함됩니다.

## 4분위 맵

이 차트에는  $y$ 축에 대상 및  $x$ 축에 척도 기능이 있는 산점도(또는 대상의 측정 수준에 따라 다른 점 도표)에서 초점 케이스와 가장 가까운 이웃( $k$ )이 표시됩니다. 기능 공간에서 대상을 선택하고 기능 공간에서 초점 케이스를 선택할 경우 이 도표를 사용할 수 있습니다.

- 훈련 파티션의 변수 평균에서 연속적인 변수에 대해 참조선이 그려집니다.

## 필드선택 오차 로그

차트에 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)를 표시합니다. 대상과 필드선택이 유효하면 이 차트를 사용할 수 있습니다.

## $k$ 선택 오차 로그

차트에 있는 점은  $x$ 축에 가장 가까운 이웃 수( $k$ 축)를 가진 모형의  $y$ 축에 오차(대상의 측정 수준에 따라 오차 비율 또는 오차 제곱합)를 표시합니다. 대상과  $k$  선택이 유효하면 이 차트를 사용할 수 있습니다.

## $k$ 및 필드선택 오차 로그

$k$ 에 의해 패널된 필드선택 차트(『필드선택 오차 로그』 참조)입니다. 대상이 있고  $k$  및 필드선택이 모두 유효하면 이 차트를 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.

## 오차 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.



## 제 21 장 판별 분석

판별 분석은 그룹소속에 대한 예측 모형을 작성합니다. 모형은 프로시저로 그룹 간에 가장 큰 차이를 나타내는 예측변수의 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 예측 변수 측정을 통해 새로운 케이스에 적용될 수는 있습니다.

참고: 그룹화 변수에는 셋 이상의 값이 있습니다. 그룹화 변수에 대한 코드는 정수이어야 하며 최소값과 최대 값을 지정해야 합니다. 이러한 한계를 넘는 값이 있는 케이스는 분석에서 제외됩니다.

예제. 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.

$$D = a * \text{기후} + b * \text{도시} + c * \text{인구} + d * \text{일인당 국내총생산(GDP)}$$

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 상관 행렬, 그룹-내 공분산 행렬, 개별-그룹 공분산 행렬, 전체 공분산 행렬 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 더미 변수나 대비 변수로 코딩이 변경되어야 합니다.

가정. 케이스는 독립적이어야 합니다. 예측변수에는 다변량 정규 분포가 있어야 하고 그룹-내 분산 공분산 행렬은 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...



2. 정수값으로 된 그룹화 변수를 선택하고 범위지정을 클릭하여 특정 범주를 지정합니다.
3. 독립변수나 예측변수를 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. 값을 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.
- Box의  $M$ . 그룹 공분산 행렬의 등식에 대한 검정을 수행합니다. 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.

- Fisher의 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 상관 행렬, 그룹-내 공분산 행렬, 개별-그룹 공분산 행렬, 전체 공분산 행렬입니다.

- 그룹-내 상관. 상관을 계산하기 전에 모든 그룹에 대한 개별 공분산 행렬의 평균을 구하여 그룹 내 풀링 상관 행렬을 표시합니다.
- 그룹-내 공분산. 그룹 내 풀링 공분산 행렬을 표시하는데 이는 전체 공분산 행렬과 다를 수 있습니다. 이 행렬은 모든 그룹에 대해 개별 공분산 행렬을 평균하여 구합니다.
- 개별-그룹 공분산. 각 그룹에 대해 개별 공분산 행렬을 표시합니다.
- 전체 공분산. 단일 표본으로 작성한 것처럼 모든 케이스로부터 공분산 행렬을 표시합니다.

---

## 판별 분석 단계적 방법

방법. 새로운 변수 입력 및 제거에 사용할 통계를 선택합니다. 사용할 수 있는 옵션은 Wilk의 람다, 설명되지 않는 분산, Mahalanobis의 거리, 최소  $F$ -비, Rao의  $V$  등입니다. Rao의  $V$ 를 사용하면 입력할 변수에 대한  $V$ 에 최소값 증가를 지정할 수 있습니다.

- Wilk의 람다. 단계별 판별 분석의 변수 선택 방법으로, Wilk의 람다를 낮추는 정도에 따라 방정식에 입력할 변수를 선택합니다. 각 단계에서 전체 Wilk의 람다를 최소화할 변수를 입력합니다.
- 설명되지 않는 분산. 각 단계에서 그룹 간 설명되지 않은 변동 합계를 최소화하는 변수를 입력합니다.
- Mahalanobis 거리. 독립변수의 케이스 값이 전체 케이스 평균과 얼마나 달라지는지에 대한 척도입니다. Mahalanobis 거리가 크면 케이스가 독립변수 하나 이상에 대해 극단값을 갖는 것으로 식별합니다.
- 최소  $F$ -비. 그룹 간 Mahalanobis 거리로부터 계산한  $F$ -비를 최대화하는 단계별 분석의 변수 선택 방법입니다.
- Rao의  $V$ . 그룹 평균 간 차이에 대한 척도입니다. Lawley-Hotelling 트레이스라고도 하며 각 단계에서 Rao의  $V$ 의 증가를 최대화하는 변수가 입력됩니다. 이 옵션을 선택한 다음 변수가 가져야 하는 최소값을 입력하여 분석에 사용합니다.

기준. 사용할 수 있는 옵션은 **F-값 사용** 및 **F-확률** 사용입니다. 변수 입력 및 제거를 위한 값을 입력합니다.

- $F$ -값 사용.  $F$  값이 진입값보다 크면 모형에 변수가 입력되고  $F$  값이 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.  $F$  값의 유의 수준이 진입값보다 작으면 모형에 변수가 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.

출력. 단계 요약에서는 각 단계 다음에 모든 변수에 대한 통계를 표시하며 대응별 거리에 대한  $F$ 에서는 각 대응 그룹에 대한 대응별  $F$ 비 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.

출력. 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.
- 요약표. 판별 분석을 기준으로 각 그룹에 정확하게 할당되거나 잘못 할당된 케이스의 수로, "혼동행렬"이라고도 합니다.
- 순차제거복원 분류. 분석의 각 케이스가 해당 케이스가 아닌 다른 모든 케이스에서 파생된 함수에 따라 분류됩니다. 이 방법을 "U-방법"이라고도 합니다.

결측값을 평균으로 바꾸기. 이 옵션을 선택하면 분류 단계에서만 결측값에 대한 독립변수의 평균을 바꿀 수 있습니다.

공분산 행렬 사용. 이 옵션을 선택하여 그룹-내 공분산 행렬이나 개별-그룹 공분산 행렬을 사용하여 케이스를 분류할 수 있습니다.

- 그룹 내. 그룹 내 풀링 공분산 행렬이 케이스 분류에 사용됩니다.
- 개별-그룹. 개별-그룹 공분산 행렬이 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 값에 대해 전체 그룹화 산점도를 작성합니다. 함수가 하나만 있는 경우에는 산점도 대신 히스토그램이 표시됩니다.
- 개별-그룹. 처음 두 판별 함수 값의 개별 그룹 산점도를 작성합니다. 함수가 하나만 있는 경우에는 산점도 대신 히스토그램이 표시됩니다.
- 영역 맵. 함수 값에 따라 케이스를 그룹으로 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. 스코어링을 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 상관 행렬을 읽고 분석합니다.
- MATRIX 하위 명령을 사용하여 나중에 분석하기 위해 상관 행렬을 기록합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



## 제 22 장 요인 분석

요인 분석을 사용하여 기본 변수 또는 요인을 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 유용성을 제공합니다.

- 일곱 가지 요인 추출 방법을 사용할 수 있습니다.
- 비직교 회전에 대한 직접 오블리민과 프로맥스 방법을 포함하여 다섯 가지 회전 방법을 사용할 수 있습니다.
- 요인 스코어를 계산하는 데 세 가지 방법이 있으며 스코어는 이후의 분석에 대한 변수로 저장됩니다.

**예제.** 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 상관 행렬(유의 수준, 판별계수, 역 상관 행렬 포함), 재연된 상관 행렬(역-이미지 행렬 포함), 초기해법(공통성, 고유 값, 설명한 분산 백분율), 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 공분산 행렬을 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.

요인 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 차원 축소 > 요인...

2. 요인 분석에 대한 변수를 선택합니다.

---

## 요인 분석 선택변수

분석할 케이스를 선택하려면

1. 선택변수를 선택합니다.
2. 값을 클릭하고 선택 값으로 정수를 입력합니다.

요인 분석에서는 선택변수 값이 있는 케이스만을 사용합니다.

---

## 요인 분석: 기술통계

통계. 일변량 기술통계에는 각 변수에 대한 평균, 표준 편차, 유효 케이스 수 등이 포함됩니다. 초기해법에서는 공통성, 고유 값, 설명된 분산 백분율 등을 표시합니다.

상관 행렬. 사용할 수 있는 옵션은 계수, 유의 수준, 행렬식, 역 모형, 재연된 상관 행렬, 역-이미지, KMO와 Bartlett의 단위행렬 검정입니다.

- *KMO*와 *Bartlett*의 단위행렬 검정. 표본 적합도에 대한 Kaiser-Meyer-Olkin 측도는 변수 간 편상관이 작 은지 여부를 검정합니다. *Bartlett*의 단위행렬 검정은 상관 행렬이 요인 모형이 부적합함을 나타내는 항등 행렬인지 여부를 검정합니다.
- *재연됨*. 요인 해법에서 추정된 상관 행렬입니다. 잔차(추정된 상관과 기대된 상관 간 차이)도 표시됩니다.
- *역-이미지*. 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 측도 는 역 이미지 상관 행렬의 대각선에 표시됩니다.

---

## 요인 분석: 요인추출

방법. 요인추출 방법을 지정할 수 있습니다. 사용할 수 있는 방법은 주성분, 가중치를 부여하지 않은 최소제곱 법, 일반화 최소제곱법, 최대 우도, 주축 요인 추출, 알파 요인 추출, 이미지 요인 추출 등입니다.

- *주성분 분석(Principal Components Analysis)*. 관측된 변수의 상관되지 않은 선형 조합을 형성하는 데 사 용되는 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 상관 행렬이 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 간의 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 것 같은 모수 추정값을 생성하는 요인 추출 방법입니다. 상관은 변수의 역고유성에 의해 가중되고 반복 알고리즘이 사용됩니다.
- 주축 요인 추출(*Principal Axis Factoring*). 공통성의 초기 추정값으로서 대각선으로 배치된 제곱 다중 상관계수를 사용하여 원래의 상관 행렬로부터 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(*Alpha*). 분석 변수를 잠재적 변수의 모집단에서 추출된 표본으로 간주하는 요인 추출 방법입니다. 이 방법은 요인의 알파 신뢰도를 최대화합니다.
- 이미지 요인 추출(*Image Factoring*). Guttman이 개발한 요인 추출 방법으로, 이미지 이론을 기반으로 합니다. 편 이미지라고 하는 변수의 공통 부분은 가설 요인의 합수보다는 남아 있는 변수에서 선형 회귀분석으로 정의됩니다.

분석. 상관 행렬이나 공분산 행렬을 지정할 수 있습니다.

- 상관 행렬. 분석의 변수가 다양한 척도에서 측정된 경우 유용합니다.
- 공분산 행렬. 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.

출력. 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.
- 스크리 도표. 각 요인과 관련된 분산 도표로, 그대로 유지할 요인의 수를 결정하는 데 사용됩니다. 일반적으로 도표는 큰 요인들의 가파른 기울기와 나머지 요인들의 점진적 꼬리 부분 (스크리) 간의 뚜렷한 구분을 보여 줍니다.

수렴에 대한 최대반복수. 해법을 추정할 때 알고리즘에서 사용할 최대 단계수를 지정할 수 있습니다.

## 요인 분석: 회전

방법. 요인 회전 방법을 선택할 수 있습니다. 사용할 수 있는 방법은 베리맥스, 이퀴맥스, 쿼티맥스, 직접 오블리민, 프로맥스입니다.

- 베리맥스 방법(*Varimax Method*). 각 요인의 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.

- **쿼티맥스 방법(Quartimax Method)**. 각 변수를 설명하는 데 필요한 요인 수를 최소화하는 회전 방법입니다. 이 방법을 사용하면 관측된 변수의 해석을 단순화할 수 있습니다.
- **이쿼맥스 방법(Equamax Method)**. 요인을 단순화하는 베리맥스 방법과 변수를 단순화하는 쿼티맥스 방법을 조합한 회전 방법입니다. 요인에 읽어들인 변수의 수와 변수 설명에 사용할 요인 수는 최소화됩니다.
- **프로맥스 회전(Promax Rotation)**. 요인이 상관되도록 하는 오블리크 회전입니다. 이 회전은 직접 오블리민 회전보다 좀더 빨리 계산될 수 있으므로 큰 데이터 세트에 유용합니다.

출력. 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 상관 행렬이 표시됩니다.
- **요인로딩 도표(Factor Loading Plot)**. 처음 세 요인에 대한 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.

방법. 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 스코어**. 요인 스코어 계수 추정 방법입니다. 생성된 스코어의 평균은 0이고 변수 범위에서 고유 요인의 제곱합이 최소화됩니다.
- **Anderson-Rubin 방법(Anderson-Rubin Method)**. 요인 스코어 계수를 추정하는 방법으로서, 추정 요인의 직교성을 보장하는 Bartlett 방법을 수정한 방법입니다. 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 핸들 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.
- 상관 행렬 또는 요인로딩 행렬을 나중에 분석하기 위해 디스크에 저장합니다.
- 상관 행렬이나 요인로딩 행렬을 읽고 분석합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



---

## 제 23 장 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 축소 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 Statistics 파일로 저장하는 기능.

K-Means 군집 분석 프로시저에서는 대용량 데이터 파일도 분석할 수 있습니다.



---

## 제 24 장 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.

**예제.** 소매 및 소비 제품 판매 회사는 정기적으로 고객의 구매 습관, 성별, 연령, 수입 등을 설명하는 데이터를 군집하는 기법을 사용합니다. 회사는 각 고객 그룹에 맞는 마케팅 및 제품 개발 전략을 세워 판매를 촉진하는 물론 상품 가치를 높일 수 있습니다.

**거리 척도.** 이 기능을 선택하면 두 군집 간 유사성이 계산되는 방식을 결정할 수 있습니다.

- **로그-우도.** 로그-우도 척도를 선택하면 변수에 대한 확률 분포가 계산됩니다. 연속형 변수는 정규 분포로, 범주형 변수는 다항분포로 계산된다고 가정하며 모든 변수를 독립변수로 가정합니다.
- **유클리디안.** 유클리디안 척도를 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 요약 을 제공합니다. 자세한 정보는 112 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 베이저안 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 트리와 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(Gaussian)분포를 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...

2. 범주형 변수와 연속형 변수를 하나 이상 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 군집 구성 기준을 조정합니다.
- 잠음 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

이상치 처리. 이 옵션을 사용하면 CF(군집 기능) 트리가 채워진 경우 군집화를 수행하는 동안 이상치를 처리할 수 있습니다. CF 트리 리프 노드에 케이스를 더 이상 추가할 수 없거나 리프 노드를 더 이상 분할할 수 없으면 CF 트리가 채워진 것입니다.

- 잠음 처리를 선택하고 CF 트리가 채워진 경우 산재되어 있는 리프의 케이스를 "잠음" 리프에 추가하면 트리가 다시 작성됩니다. 리프는 최대 리프 크기 케이스의 지정된 백분율보다 더 적은 백분율의 케이스를 포함하고 있을 때 산재하고 있는 것으로 간주됩니다. CF 트리가 다시 작성되면 이 트리에 이상치가 적용됩니다. 그렇지 않으면 이상치는 무시됩니다.
- 잠음 처리를 선택하지 않고 CF 트리가 채워진 경우 거리 변경 임계값을 늘려 트리를 다시 작성할 수 있습니다. 최종 군집화 후에 군집에 할당될 수 없는 값은 이상치로 설명됩니다. -1이라는 식별 번호가 지정된 이상치 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 트리 튜닝 기준.** 다음 군집화 알고리즘 설정은 CF(군집 기능) 트리에 적용되기 때문에 변경 시 주의해야 합니다.

- 초기 거리 변경 임계값. 이 옵션은 CF 트리를 작성할 때 사용되는 초기 임계값입니다. 특정 케이스를 CF 트리 리프에 삽입하면 임계값보다 트리가 좁게 작성되므로 리프가 분리되지 않습니다. 임계값보다 트리가 넓게 작성되면 리프가 분리됩니다.
- 최대 분기(리프 노드당). 리프 노드가 가질 수 있는 최대 하위 노드 수입니다.
- 최대 트리 깊이. CF 트리가 가질 수 있는 최대 리프 수입니다.
- 가능한 최대 노드 수.  $b$ 가 최대 분기이고  $d$ 가 최대 트리 깊이인 함수  $(b^{d+1} - 1) / (b - 1)$ 를 기반으로 프로시저에서 생성될 수 있는 최대 CF 트리 노드 수를 나타냅니다. CF 트리 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 트리가 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 114 페이지의 『이단계 군집 분석 결과』 주제를 참조하십시오.

군집 모형 업데이트를 지정하면 원래 모형에 대해 지정된 CF 트리의 생성과 관련된 옵션이 사용됩니다. 즉, 저장된 모형에 대한 거리 측도, 잡음 처리, 메모리 할당 또는 CF 트리 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 차트 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 트리는 XML 형식으로 내보낼 수 있는 두 가지 유형의 결과 파일입니다.

- 최종 모형 내보내기. 최종 군집 모형을 XML(PMML) 형식의 지정된 파일로 내보냅니다. 스코어링을 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.
- CF 트리 내보내기. 이 옵션을 사용하면 군집 트리의 현재 상태를 저장한 다음 나중에 새로운 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 고수의 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 소속 군집을 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 『모형 요약 보기』 주제를 참조하십시오.
- 군집. 자세한 정보는 『군집 보기』 주제를 참조하십시오.

4개의 링크/보조 보기가 있습니다.

- 예측변수 중요도. 자세한 정보는 117 페이지의 『군집 예측변수 중요도 보기』 주제를 참조하십시오.
- 군집 크기(기본값). 자세한 정보는 117 페이지의 『군집 크기 보기』 주제를 참조하십시오.
- 셀 분포. 자세한 정보는 117 페이지의 『셀 분포 보기』 주제를 참조하십시오.
- 군집 비교. 자세한 정보는 117 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 모델링 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  $(B-A) / \text{최대}(A,B)$ 에 따라 전체적인 레코드의 평균을 계산합니다. 여기서 A는 레코드와 군집 중심 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 예측변수라고도 합니다.
- 군집 해법에 있는 여러 개의 군집입니다.

## 군집 보기

군집 보기에는 각 군집의 군집 이름, 크기 및 프로파일이 포함된 기능별 군집 눈금이 들어 있습니다.

눈금의 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 백분율로 표시됨). 눈금에 있는 각 크기 셀에는 해당 군집 내에 크기 백분율을 표시하는 세로 막대, 숫자 형식의 크기 백분율, 군집 케이스 개수가 표시됩니다.
- **지형.** 개별 입력 또는 예측변수로서 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(전형) 범주의 이름과 백분율이 표시됩니다.

군집 보기 내에서 다양한 방법을 선택하여 군집 정보를 표시할 수 있습니다.

- **군집 및 기능 전치.** 자세한 정보는 『군집 및 기능 전치』 주제를 참조하십시오.
- **기능 정렬.** 자세한 정보는 『기능 정렬』 주제를 참조하십시오.
- **군집 정렬.** 자세한 정보는 『군집 정렬』 주제를 참조하십시오.
- **셀 내용 선택.** 자세한 정보는 『셀 내용』 주제를 참조하십시오.

**군집 및 기능 전치:** 기본적으로 군집은 열로 표시되고 기능은 행으로 표시됩니다. 이 출력결과를 역으로 바꾸려면 기능 정렬 기준 단추 왼쪽에 있는 **군집 및 기능 전치** 단추를 누르십시오. 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 문숫자 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 백분율과 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 막대형 차트가 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

예측변수 중요도 보기는 모형 추정에서의 각 필드의 상대적 중요도를 보여줍니다.

## 군집 크기 보기

군집 크기 보기에는 각 군집이 들어 있는 원형 차트가 표시됩니다. 원형 차트의 조각마다 각 군집의 백분율 크기가 표시됩니다. 각 조각에 마우스를 올려 놓으면 해당 조각 내의 개수가 표시됩니다.

차트 아래에 다음과 같은 크기 정보가 나열된 테이블이 있습니다.

- 가장 작은 군집의 크기(전체 개수 및 백분율)
- 가장 큰 군집의 크기(전체 개수 및 백분율)
- 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 눈금 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 중요도를 선택하면 필드가 항상 전체 중요도별로 정렬됩니다.

배경 도표에 각 기능의 전체 분포가 표시됩니다.

- 범주형 기능은 점 도표로 표시됩니다. 점 크기는 각 군집에 대해 빈도가 가장 높은/전형 범주를 나타냅니다 (기능별).
- 연속형 기능은 상자도표로 표시됩니다. 이 도표에는 전체 평균 및 사분위수 범위가 표시됩니다.

이러한 배경 보기에 겹쳐그려진 항목은 선택된 군집의 상자도표입니다.

- 연속형 기능의 경우, 사격형 마커와 가로 선은 각 군집의 중앙값 및 사분위수 범위를 나타냅니다.
- 각 군집엔 보기 맨 위에 다른 색상으로 표시됩니다.

## 군집 뷰어 탐색

군집 뷰어는 대화식 출력결과입니다. 다음을 수행할 수 있습니다.





- 필드 또는 군집을 선택하여 보다 자세하게 볼 수 있습니다.
- 원하는 항목을 선택하여 군집을 비교할 수 있습니다.
- 출력결과를 변경할 수 있습니다.
- 축을 전치할 수 있습니다.

### 도구 모음 사용

도구 모음 옵션을 사용하여 왼쪽 및 오른쪽 패널에 표시되는 정보를 제어할 수 있습니다. 도구 모음 제어를 사용하여 출력결과와 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 115 페이지의 『군집 보기』 주제를 참조하십시오.

표2. 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.

**지형.** 기본적으로 선택됩니다. 모든 입력 기능을 숨기려면 확인란을 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: 사용 가능한 평가 필드가 없으면 이 확인란을 사용할 수 없습니다.

**군집 설명.** 기본적으로 선택됩니다. 모든 군집 설명을 숨기려면 확인란을 선택 해제하십시오.

**군집 크기.** 기본적으로 선택됩니다. 모든 군집 크기 셀을 숨기려면 확인란을 선택 해제하십시오.

**최대 범주 수.** 범주형 기능의 차트에 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. 확인을 클릭합니다.



---

## 제 25 장 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 거리측도나 유사성 측도를 작성합니다. 각 단계마다 통계가 표시되므로 가장 적절한 선택을 할 수 있습니다.

**예제.** 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 검토하는데 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.** 거리측도나 유사성 측도를 적절하게 사용하여 데이터를 분석합니다. 거리측도 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...

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, 중앙값 군집화, Ward의 방법을 사용할 수 있습니다.

**측도.** 거리측도나 유사성 측도를 지정하여 군집화에 사용할 수 있습니다. 다음 데이터 유형과 적합한 거리측도 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, 램다, 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.** 거리측도에 의해 생성된 값을 변환할 수 있습니다. 이 값은 거리측도를 계산한 다음 적용됩니다. 절댓값, 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.

**군집 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 척도에서 거듭제곱 및 루트 값을 지정합니다.
- 저장된 변수에 대한 이름을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



---

## 제 26 장 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 군집 중심을 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.

**예제.** 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 옵션**을 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 옵션**을 사용하지 않으면 케이스 순서와 관련된 문제를 방지할 수 있습니다. 그러나 케이스에서 군집중심까지의 거리가 동등인 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...

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. 최종 값 저장을 선택합니다. 그런 다음 전체 데이터 파일을 복구하고 분류만 하기 방법을 선택한 후 초기값 읽기를 선택하여 표본에서 추정된 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 분류하기를 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 최대반복수를 1로 설정합니다.

수렴 기준. 반복을 중단할 시기를 결정합니다. 이 시기는 군집중심초기값 사이의 최소 거리 비율을 나타내므로, 0보다 크고 1보다 작거나 같아야 합니다. 예를 들어, 기준이 0.02이면, 반복은 군집중심초기값 사이의 최소 거리 퍼센트가 2 이상인 거리만큼 군집중심을 옮기지 않을 때 중지됩니다.

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.

- **군집중심초기값.** 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 케이스의 군집정보.** 케이스마다 최종 군집 할당을 표시하고 케이스 분류에 사용된 군집 중심과 케이스 간의 유클리드 거리를 표시합니다. 최종 군집 중심 간 유클리드 거리도 표시합니다.

**결측값.** 사용 가능한 옵션은 목록별 결측값 제외 또는 대응별 결측값 제외입니다.

- **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 군집 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.



---

## 제 27 장 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.

원하는 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 검정을 수행하려면

메뉴에서 다음을 선택합니다.

분석 > 비모수 검정 > 일표본...

1. 실행을 클릭합니다.

선택적으로 다음을 수행할 수 있습니다.

- 목표 탭에서 목표를 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. 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 이항 검정을 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 『이항 검정 탭』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 131 페이지의 『카이제곱 검정 탭』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 131 페이지의 『콜모고로프-스미르노프 옵션』의 내용을 참조하십시오.
- 평균과 가설 값 비교(윌콕슨 부호순위 검정). Wilcoxon 부호 순위 검정은 연속형 및 순서 필드에 적용됩니다. 이 옵션은 필드의 중앙값에 대한 일표본 검정을 생성합니다. 숫자로 가설 중앙값을 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 131 페이지의 『런 검정 옵션』의 내용을 참조하십시오.

**이항검정 탭:** 이항검정은 플래그 필드(범주가 2개만 있는 범주형 필드)에 적용되지만 "성공" 정의 규칙을 사용하면 모든 필드에 적용할 수 있습니다.

**가설 비율.** 이 옵션은 "성공"으로 정의된 레코드의 기대 비율 또는  $p$ 를 지정합니다. 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 베이저언 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 사용은 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. 기본값입니다.
- 성공값 지정은 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 중심점은 최소값 및 최대값의 평균에서 절단점을 설정합니다.
- 사용자 정의 절단점을 사용하여 절단점의 값을 지정할 수 있습니다.

**카이제곱 검정 탭:** 모든 범주 확률 동일. 이 옵션은 표본에 있는 모든 범주에 같은 빈도를 생성합니다. 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.

**클모고로프-스미르노프 옵션:** 이 대화 상자에서는 검정할 분포와 가설 분포의 모수를 지정합니다.

**정규 분포.** 표본 데이터 사용에서는 관측된 평균과 표준 편차를 사용하며, 사용자 정의를 통해 값을 지정할 수 있습니다.

**균일.** 표본 데이터 사용에서는 관측된 최소값과 최대값을 사용하며, 사용자 정의를 통해 값을 지정할 수 있습니다.

**지수.** 표본 평균에서는 관측된 평균을 사용하며, 사용자 정의를 통해 값을 지정할 수 있습니다.

**포아송.** 표본 평균에서는 관측된 평균을 사용하며, 사용자 정의를 통해 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 있음은 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 코딩변경은 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 중앙값은 표본 중앙값에서 절단점을 설정합니다.
- 표본 평균은 표본 평균에서 절단점을 설정합니다.
- 사용자 정의를 사용하여 절단점의 값을 지정할 수 있습니다.

### 검정 옵션

**유의 수준.** 이 옵션은 모든 검정의 유의 수준(알파)을 지정합니다. 0과 1 사이의 숫자 값을 지정합니다. 기본값은 0.05입니다.

**신뢰구간(%).** 이 옵션은 생성된 모든 신뢰구간의 신뢰수준을 지정합니다. 0과 100 사이의 숫자 값을 지정합니다. 기본값은 95입니다.

**제외된 케이스.** 이 옵션은 검정의 케이스 기준을 결정하는 방법을 지정합니다.

- 목록별 결측값 제외는 필드 탭에 이름 지정된 모든 필드에 대해 결측값이 있는 레코드가 모든 분석에서 제외됨을 의미합니다.
- 검정별 결측값 제외는 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.



---

## 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.

원하는 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 중앙값 비교. 이 목표는 중앙값 검정을 사용하여 그룹 전체에서 관측된 중앙값을 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. 실행을 클릭합니다.

선택적으로 다음을 수행할 수 있습니다.

- 목표 탭에서 목표를 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 중앙값, 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. 모든 대응별 다중 비교 또는 단계별 하향 비교 중 하나로  $k$  표본의 다중 비교를 요청할 수 있습니다(선택 사항).

**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. 대체 가설 순서를 지정합니다. 가장 작은 값에서 가장 큰 값으로는 첫 번째 그룹의 위치 모수가 두 번째와 같지 않고, 두 번째와 세 번째가 같지 않은 방식의 대체 가설을 규정합니다. 가장 큰 값에서 가장 작은 값으로는 마지막 그룹의 위치 모수가 끝에서 두 번째와 같지 않고, 끝에서 두 번째와 끝에서 세 번째가 같지 않은 방식의 대체 가설을 규정합니다. 모든 대응별 다중 비교 또는 단계별 하향 비교 중 하나로  $k$  표본의 다중 비교를 요청할 수 있습니다(선택 사항).

- **그룹 전체에서 범위 비교.** 이 옵션은 표본의 범위가 동일한지 확인하는 독립 표본 검정을 생성합니다. 모제의 **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 중앙값 비교. 이 옵션은 표본의 중앙값이 동일한지 확인하는 독립 표본 검정을 생성합니다. 중앙값 검정(k 표본)에서는 풀링된 표본 중앙값(데이터 세트의 모든 레코드에서 계산됨) 또는 사용자 정의 값을 가설 중앙값으로 사용할 수 있습니다. 모든 대응별 다중 비교 또는 단계별 하향 비교 중 하나로 k 표본의 다중 비교를 요청할 수 있습니다(선택 사항).
- 그룹 전체에서 신뢰구간 추정. 핫지스-레만 추정(2개 표본)은 두 그룹의 중앙값 차이에 대해 독립 표본 추정과 신뢰구간을 생성합니다. 그룹 필드에 값이 3개 이상 있으면 이 검정이 생성되지 않습니다.

## 검정 옵션

**유의 수준.** 이 옵션은 모든 검정의 유의 수준(알파)을 지정합니다. 0과 1 사이의 숫자 값을 지정합니다. 기본값은 0.05입니다.

**신뢰구간(%).** 이 옵션은 생성된 모든 신뢰구간의 신뢰수준을 지정합니다. 0과 100 사이의 숫자 값을 지정합니다. 기본값은 95입니다.

**제외된 케이스.** 이 옵션은 검정의 케이스 기준을 결정하는 방법을 지정합니다. 목록별 결측값 제외는 하위 명령에 이름 지정된 모든 필드에 대해 결측값이 있는 레코드가 모든 분석에서 제외됨을 의미합니다. 검정별 결측값 제외는 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.

원하는 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 다른 필드가 지정된 경우, 먼저 측정 수준별로 분리된 다음 각 그룹에 적절한 검정이 적용됩니다. 예를 들어, 관측 데이터와 가설 데이터 자동으로 비교를 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. 실행을 클릭합니다.

선택적으로 다음을 수행할 수 있습니다.

- 목표 탭에서 목표를 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 목표와 함께 사용할 수 없는 사항으로 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 필드가 2개 지정되면 범주형 데이터에 맥니머의 검정을 적용하고, 필드가 3개 이상 지정되면 범주형 데이터에 Cochran의 Q를 적용합니다. 또한 필드가 2개 지정되면 연속형 데이터에 월콕슨 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 데이터에서 변경 검정.** 맥니머의 검정(2개의 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 『McNemar의 검증: 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. 모든 대응별 다중 비교 또는 단계별 하향 비교 중 하나로  $k$  표본의 다중 비교를 요청할 수 있습니다(선택 사항). 검정 설정에 대한 자세한 내용은 138 페이지의 『Cochran의 Q: 성공 정의』의 내용을 참조하십시오.
- **다항 데이터에서 변경 검정.** 주변 동질성 검정(2개의 표본)은 쌍을 이루는 2개의 순서 필드 사이 값 조합이 동일할 가능성이 있는지 확인하는 대응표본 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.
- **중앙값 차이와 가설 값 비교.** 이러한 검정은 각각 두 필드 사이의 중앙값 차이가 0이 아닌지를 확인하는 대응표본 검정을 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 중앙값 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. 모든 대응별 다중 비교 또는 단계별 하향 비교 중 하나로  $k$  표본의 다중 비교를 요청할 수 있습니다(선택 사항).
- **분포 비교.** 프리드만의 순위별 2원 ANOVA( $k$  표본)는  $k$  대응표본이 동일한 모집단에서 비롯되었는지를 확인하는 대응표본 검정을 생성합니다. 모든 대응별 다중 비교 또는 단계별 하향 비교 중 하나로  $k$  표본의 다중 비교를 요청할 수 있습니다(선택 사항).

**McNemar의 검증: 성공 정의:** 맥니머의 검정은 플래그 필드(범주가 2개만 있는 범주형 필드)에 적용되지만 "성공" 정의의 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 사용은 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. 기본값입니다.



- 성공값 지정은 지정된 값 목록으로 검정을 수행하여 "성공"을 정의합니다. 문자 또는 숫자 값 목록을 지정합니다. 목록에 있는 값이 표본에 없어도 됩니다.

**Cochran의 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 사용은 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. 기본값입니다.
- 성공값 지정은 지정된 값 목록으로 검정을 수행하여 "성공"을 정의합니다. 문자 또는 숫자 값 목록을 지정합니다. 목록에 있는 값이 표본에 없어도 됩니다.

## 검정 옵션

유의 수준. 이 옵션은 모든 검정의 유의 수준(알파)을 지정합니다. 0과 1 사이의 숫자 값을 지정합니다. 기본값은 0.05입니다.

신뢰구간(%). 이 옵션은 생성된 모든 신뢰구간의 신뢰수준을 지정합니다. 0과 100 사이의 숫자 값을 지정합니다. 기본값은 95입니다.

제외된 케이스. 이 옵션은 검정의 케이스 기준을 결정하는 방법을 지정합니다.

- 목록별 결측값 제외는 하위 명령에 이름 지정된 모든 필드에 대해 결측값이 있는 레코드가 모든 분석에서 제외됨을 의미합니다.
- 검정별 결측값 제외는 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.

---

## 모형 보기

### 모형 보기

이 프로시저에서는 뷰어에 모형 뷰어 오브젝트를 만듭니다. 이 오브젝트를 활성화(두 번 클릭)하면 모형에 대한 대화형 보기를 얻을 수 있습니다. 모형 보기에는 2개 패널로 된 창이 있으며 왼쪽에는 기본 보기, 오른쪽에는 링크 또는 보조 보기가 있습니다.

2개의 기본 보기가 있습니다.

- 가설 요약. 기본 보기입니다. 자세한 정보는 『가설 요약』 주제를 참조하십시오.
- 신뢰구간 요약. 자세한 정보는 『신뢰구간 요약』 주제를 참조하십시오.

7개의 링크/보조 보기가 있습니다.

- 일표본 검정. 일표본 검정이 요청된 경우 기본값 보기입니다. 자세한 정보는 140 페이지의 『일표본 검정』 주제를 참조하십시오.
- 대응 표본 검정. 대응 표본 검정이 요청되고 일표본 검정이 요청되지 않은 경우 기본값 보기입니다. 자세한 정보는 141 페이지의 『대응 표본 검정』 주제를 참조하십시오.
- 독립 표본 검정. 대응 표본 검정이 요청되지 않고 일표본 검정이 요청된 경우 기본값 보기입니다. 자세한 정보는 142 페이지의 『독립 표본 검정』 주제를 참조하십시오.
- 범주형 필드 정보. 자세한 정보는 143 페이지의 『범주형 필드 정보』 주제를 참조하십시오.
- 연속형 필드 정보. 자세한 정보는 143 페이지의 『연속형 필드 정보』 주제를 참조하십시오.
- 쌍대 비교. 자세한 정보는 143 페이지의 『쌍대 비교』 주제를 참조하십시오.
- 동일서브세트. 자세한 정보는 144 페이지의 『동일서브세트』 주제를 참조하십시오.

### 가설 요약

모형 요약 보기는 비모수 검정을 한 눈에 파악할 수 있도록 요약한 스냅샷입니다. 이 보기는 귀무가설 및 결정을 강조하여 중요한  $p$ -값에 주목합니다.

- 각 행은 별도의 검정에 해당합니다. 행을 클릭하면 링크 보기의 검정에 대한 추가 정보가 표시됩니다.
- 열 헤더를 클릭하면 해당 열의 값별로 행이 정렬됩니다.
- 재설정 단추를 사용하면 모형 뷰어를 원래 상태로 되돌릴 수 있습니다.
- 필드 필터 드롭다운 목록을 사용하여 선택된 필드를 포함하는 검정만 표시할 수 있습니다.

### 신뢰구간 요약

신뢰구간 요약은 비모수 검정에 의해 생성된 모든 신뢰구간을 보여줍니다.

- 각 행은 별도의 신뢰구간에 해당합니다.
- 열 헤더를 클릭하면 해당 열의 값별로 행이 정렬됩니다.



## 일표본 검정

일표본 검정 보기는 요청된 일표본 비모수 검정과 관련된 세부사항을 보여줍니다. 이 정보는 선택된 검정에 따라 표시여부가 결정됩니다.

- 검정 드롭다운을 사용하여 지정된 유형의 일표본 검정을 선택할 수 있습니다.
- 필드 드롭다운을 사용하면 검정 드롭다운에서 선택한 검정을 사용하여 검정한 필드를 선택할 수 있습니다.

### 이항검정

이항검정은 누적 막대형 차트와 검정 테이블을 보여줍니다.

- 누적 막대형 차트에는 검정 필드의 "성공" 및 "실패" 범주에 대한 관측 빈도 및 가설 빈도가 표시됩니다. "실패" 범주는 "성공" 범주 맨 위에 누적됩니다. 막대 위에 마우스를 올려 놓으면 도구 팁에 범주 백분율이 표시됩니다. 막대에 표시되는 차이는 검정 필드에 가설 이항 분포가 없을 수 있음을 나타냅니다.
- 테이블에 검정 세부사항이 표시됩니다.

### 카이제곱 검정

카이제곱 검정 보기는 묶은 가로 막대형 차트와 검정 테이블을 보여줍니다.

- 묶은 가로 막대형 차트에는 검정 필드의 각 범주에 대한 관측 빈도 및 가설 빈도가 표시됩니다. 막대에 마우스를 올려 놓으면 도구 팁에 관측 빈도 및 가설 빈도와 이 빈도 간의 차이(잔차)가 표시됩니다. 관측 막대와 가설 막대에 표시되는 차이는 검정 필드에 가설 분포가 없을 수 있음을 나타냅니다.
- 테이블에 검정 세부사항이 표시됩니다.

### 월콕슨 부호순위

월콕슨 부호순위 검정 보기는 히스토그램과 검정 테이블을 보여줍니다.

- 히스토그램에는 관측 중앙값 및 가설 중앙값을 보여주는 세로선이 있습니다.
- 테이블에 검정 세부사항이 표시됩니다.

### 런 검정

런 검정 보기는 차트와 검정 테이블을 보여줍니다.

- 차트에는 관측된 런 검정 수와 함께 정규 분포가 세로선으로 표시됩니다. 정확한 검정이 수행되는 경우 이 검정은 정규 분포를 기반으로 하지 않습니다.
- 테이블에 검정 세부사항이 표시됩니다.

### 콜모고로프-스미르노프 검정

콜모고로프-스미르노프 검정 보기는 히스토그램과 검정 테이블을 보여줍니다.

- 히스토그램에는 가설된 균일, 정규, 포아송 또는 지수 분포의 확률 밀도 함수의 오버레이가 포함됩니다. 이 검정은 누적 분포를 기반으로 하므로 보고서에 표시된 최대 차이는 누적 분포와 관련하여 해석해야 합니다.
- 테이블에 검정 세부사항이 표시됩니다.

## 대응 표본 검정

일표본 검정 보기는 요청된 일표본 비모수 검정과 관련된 세부사항을 보여줍니다. 이 정보는 선택된 검정에 따라 표시여부가 결정됩니다.

- 검정 드롭다운을 사용하여 지정된 유형의 일표본 검정을 선택할 수 있습니다.
- 필드 드롭다운을 사용하면 검정 드롭다운에서 선택한 검정을 사용하여 검정한 필드를 선택할 수 있습니다.

### McNemar 검정

McNemar 검정 보기는 묶은 가로 막대형 차트와 검정 테이블을 보여줍니다.

- 묶은 가로 막대형 차트에는 검정 필드에 의해 정의된 2x2 테이블의 비대각선 셀에 대한 관측 빈도 및 가설 빈도가 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

### 부호 검정

부호 검정 보기는 누적 히스토그램과 검정 테이블을 보여줍니다.

- 누적 히스토그램은 누적 필드와의 차이를 나타내는 부호를 사용하여 필드 간의 차이를 표시합니다.
- 테이블에 검정 세부사항이 표시됩니다.

### 윌콕슨 부호순위 검정

윌콕슨 부호순위 검정 보기는 누적 히스토그램과 검정 테이블을 보여줍니다.

- 누적 히스토그램은 누적 필드와의 차이를 나타내는 부호를 사용하여 필드 간의 차이를 표시합니다.
- 테이블에 검정 세부사항이 표시됩니다.

### 주변 동질성 검정

주변 동질성 검정 보기는 묶은 가로 막대형 차트와 검정 테이블을 보여줍니다.

- 묶은 가로 막대형 차트에는 검정 필드에 의해 정의된 테이블의 비대각선 셀에 대한 관측 빈도가 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

### Cochran의 Q 검정

Cochran의 Q 검정 보기는 누적 막대형 차트와 검정 테이블을 보여줍니다.

- 누적 막대형 차트에는 검정 필드의 "성공" 및 "실패" 범주에 대한 관측 빈도가 표시됩니다. "실패" 범주는 "성공" 범주 맨 위에 누적됩니다. 막대 위에 마우스를 올려 놓으면 도구 팁에 범주 백분율이 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

### Friedman의 이원 분산 분석(순위별)

Friedman의 이원 분산 분석(순위별) 보기는 패널된 히스토그램과 검정 테이블을 보여줍니다.

- 히스토그램에는 검정 필드에 의해 패널된 순위의 관측 분포가 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### 켄달의 일치계수(W)

켄달의 일치계수(W) 보기는 패널된 히스토그램과 검정 테이블을 보여줍니다.

- 히스토그램에는 검정 필드에 의해 패널된 순위의 관측 분포가 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### 독립 표본 검정

독립 표본 검정 보기는 요청된 독립 표본 비모수 검정과 관련된 세부사항을 보여줍니다. 이 정보는 선택된 검정에 따라 표시여부가 결정됩니다.

- 검정 드롭다운을 사용하여 지정된 유형의 독립 표본 검정을 선택할 수 있습니다.
- 필드 드롭다운을 사용하면 검정 드롭다운에서 선택한 검정을 사용하여 검정한 그룹 필드 및 검정 조합을 선택할 수 있습니다.

#### Mann-Whitney 검정

Mann-Whitney 검정 보기는 인구 피라미드 차트와 검정 테이블을 보여줍니다.

- 인구 피라미드 차트에는 그룹 필드의 범주를 기준으로 하여 각 그룹의 레코드 수와 그룹의 평균 순위를 표시하는 연속형 히스토그램이 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### 콜모고로프-스미르노프 검정

콜모고로프-스미르노프 검정 보기는 인구 피라미드 차트와 검정 테이블을 보여줍니다.

- 인구 피라미드 차트에는 그룹 필드의 범주를 기준으로 하여 각 그룹의 레코드 수를 표시하는 연속형 히스토그램이 표시됩니다. 누적 단추를 클릭하여 관측 누적 분포 행을 표시하거나 숨길 수 있습니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### Wald-Wolfowitz run 검정

Wald-Wolfowitz run 검정 보기는 누적 막대형 차트와 검정 테이블을 보여줍니다.

- 인구 피라미드 차트에는 그룹 필드의 범주를 기준으로 하여 각 그룹의 레코드 수를 표시하는 연속형 히스토그램이 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### 크루스칼-왈리스 검정

크루스칼-왈리스 검정 보기에는 상자도표와 검정 테이블을 보여줍니다.

- 그룹 필드의 각 범주에 대해 별도의 상자도표가 표시됩니다. 상자 위에 마우스를 올려 놓으면 도구 팁에 평균 순위가 표시됩니다.

- 테이블에 검정 세부사항이 표시됩니다.

#### Jonckheere-Terpstra 검정

Jonckheere-Terpstra 검정 보기는 상자도표와 검정 테이블을 보여줍니다.

- 그룹 필드의 각 범주에 대해 별도의 상자도표가 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### Moses의 극단값반동 검정

Moses의 극단값반동 검정 보기는 상자도표와 검정 테이블을 보여줍니다.

- 그룹 필드의 각 범주에 대해 별도의 상자도표가 표시됩니다. 레코드 ID 단추를 클릭하여 점 레이블을 표시하거나 숨길 수 있습니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### 중앙값 검정

중앙값 검정 보기는 상자도표와 검정 테이블을 보여줍니다.

- 그룹 필드의 각 범주에 대해 별도의 상자도표가 표시됩니다.
- 테이블에 검정 세부사항이 표시됩니다.

#### 범주형 필드 정보

범주형 필드 정보 보기에는 필드 드롭다운에서 선택한 범주형 필드의 막대형 차트가 표시됩니다. 사용 가능한 필드 목록은 가설 요약 보기에서 현재 선택된 검정에 사용되는 범주형 필드로 제한됩니다.

- 막대 위에 마우스를 올려 놓으면 도구 팁에 범주 백분율이 표시됩니다.

#### 연속형 필드 정보

연속형 필드 정보 보기에는 필드 드롭다운에서 선택한 연속형 필드의 히스토그램이 표시됩니다. 사용 가능한 필드 목록은 가설 요약 보기에서 현재 선택된 검정에 사용되는 연속형 필드로 제한됩니다.

#### 쌍대 비교

쌍대 비교 보기는 대응별 다중 비교가 요청되었을 때  $k$ -표본 비모수 검정에서 생성된 비교 테이블 및 거리 네트워크 차트를 보여줍니다.

- 거리 네트워크 차트는 비교 테이블을 그래픽으로 나타낸 것으로서 네트워크 노드 간의 거리는 표본 간의 차이에 해당합니다. 노란색 선은 통계적으로 유의차에 해당하며, 검정색 선은 비유의차에 해당합니다. 네트워크의 선에 마우스를 올려 놓으면 선으로 연결된 노드 간의 조정된 유의수준차가 도구 팁에 표시됩니다.
- 비교 테이블은 모든 쌍대 비교의 결과를 숫자로 보여줍니다. 각 행은 별도의 쌍대 비교에 해당합니다. 열 헤더를 클릭하면 해당 열의 값별로 행이 정렬됩니다.

## 동일서브세트

동일서브세트 보기는 단계별 하향 비교가 요청되었을 때  $k$ -표본 비모수 검정에서 생성된 비교 테이블을 보여줍니다.

- 표본 그룹의 각 행은 별도의 관련 표본(별도의 필드로 데이터에 표시됨)에 해당합니다. 통계적으로 현저히 다르지 않은 표본은 동일한 색상이 지정된 서브세트에 그룹화됩니다. 식별된 각 서브세트에 별도의 열이 사용됩니다. 모든 표본이 통계적으로 현저히 다른 경우 각 표본에 별도의 서브세트가 사용됩니다. 통계적으로 현저히 다른 표본이 없을 경우 단일 서브세트가 사용됩니다.
- 둘 이상의 표본이 있는 각 서브세트에 대해 검정 통계량, 유의수준 및 조정된 유의수준이 계산됩니다.

---

## 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. 이러한 대화 상자에서는 정확한 검정 옵션이 제공하는 기능을 사용할 수 있습니다.

**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, 중앙값 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.

**예제.** 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 범주에 대한 기대빈도는 1 이상이어야 합니다. 범주의 20%에서만 기대빈도가 5보다 작습니다.

카이제곱 검정 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 비모수 검정 > 레거시 대화 상자 > 카이제곱...

2. 하나 이상의 검정변수를 선택합니다. 각 변수로 서로 다른 검정을 만들 수 있습니다.

3. 선택적으로 기술통계, 사분위수, 결측데이터 처리 방법을 보려면 옵션을 클릭합니다.

## 카이제곱 검정 기대범위와 기대값

**기대범위.** 기본적으로 변수의 고유 값은 하나의 범주로 정의됩니다. 특정 범위 내에 범주들을 설정하려면 지정한 범위 사용을 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. 값을 선택하고 검정변수의 각 범주에 0보다 큰 값을 입력한 다음 추가를 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.

## NPART 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.

예제. 동전을 던질 때 앞면이 나올 확률은 1/2입니다. 이러한 가설을 기초로 동전을 40번 던져 결과(앞면 또는 뒷면)를 기록합니다. 이항검정을 사용한 결과 던진 횟수의 3/4이 앞면이고 관측된 유의 수준이 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 변수는 *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 > 레저시 대화 상자 > 이항...

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.

예제. 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 > 레저시 대화 상자 > 실행...

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, 중앙값 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.

## NPART 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 Z는 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.

**예제.** 여러 모수 검정에는 정규적으로 분포된 변수가 있어야 합니다. 일표본 Kolmogorov-Smirnov 검정은 변수(예: 가구소득)가 정규적으로 분포되었음을 검정할 때 사용됩니다.

**통계.** 평균, 표준 편차, 최소값, 최대값, 누락되지 않은 케이스 수, 사분위수를 선택할 수 있습니다.

일표본 Kolmogorov-Smirnov 검정 데이터 고려 사항

**데이터.** 양적변수(측도 구간 또는 비율 수준)를 사용합니다.

가정. Kolmogorov-Smirnov 검정에서는 검정 분포의 모수가 미리 지정되어 있는 것으로 가정합니다. 이 프로 시저로 표본에서 모수를 추정합니다. 표본 평균과 표본 표준 편차는 정규 분포에 대한 모수가 되고 표본 최소 값과 최대값은 균일 분포의 범위를 정의하며 표본 평균은 포아송 분포의 모수가 되며 표본 표준 편차는 지수 분포에 대한 모수가 됩니다. 가설 분포로부터의 이탈을 발견하기 위한 검정력은 크게 감소할 수 있습니다. 추정 모수 값을 사용하여 정규 분포에 대해 검정을 수행하려면 수정된 K-S Lilliefors 검정을 사용하는 것이 좋습니다. 이 검정은 데이터 탐색 프로시저에서 수행할 수 있습니다.

일표본 Kolmogorov-Smirnov 검정 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 비모수 검정 > 레저시 대화 상자 > 1-표본 K-S...

2. 하나 이상의 숫자 검정변수를 선택합니다. 각 변수로 서로 다른 검정을 만들 수 있습니다.
3. 선택적으로 기술통계, 사분위수, 결측데이터 처리 방법을 보려면 옵션을 클릭합니다.

### 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.

예제. 더 편안하고 보기도 좋으며 빠르게 치아를 교정할 수 있도록 새로운 치열 교정기가 개발되었습니다. 새 치열 교정기를 착용할 수 있는 기간을 이전 교정기와 비교하기 위해 이전 교정기를 착용한 어린이 10명과 새 교정기를 착용한 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 > 레저시 대화 상자 > 2 독립 표본...

2. 하나 이상의 숫자변수를 선택합니다.

3. 그룹화 변수를 선택한 후 그룹 정의를 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  $U$  검정**은 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 동렬시 할당되는 평균 순위로 순위가 지정됩니다. 동렬의 수는 전체 관측값 수보다 작아야 합니다. 모집단의 위치가 똑같은 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 이들 두 수 중 작은 쪽에 해당합니다. 윌콕슨 순위 합계  $W$  통계도 표시됩니다.  $W$ 는 평균 순위가 작은 그룹의 순위합이며 평균 순위가 같은 그룹은 제외합니다. 그러한 경우 독립 2-표본 비모수 검정 그룹 정의 대화 상자에서 마지막으로 이름이 지정된 그룹의 순위합입니다.

**Kolmogorov-Smirnov의  $Z$  검정**과 **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 같은 모집단에서 비롯된 경우 두 그룹은 순위 전반에 걸쳐 무작위로 흩어져 있게 됩니다.

**Moses의 극단값반동 검정**은 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. 대조군의 범위는 대조군에서 가장 큰 순위와 가장 작은 순위의 차에 1을 더한 값으로 계산됩니다. 기회 이상치로 범위가 왜곡되기 쉬우므로 통제 케이스의 양끝 5%는 자동으로 절단됩니다.

## 독립 2-표본 비모수 검정: 그룹 정의

파일을 두 그룹이나 표본으로 분할하려면 그룹 1과 그룹 2에 서로 다른 정수 값을 입력합니다. 다른 값을 가진 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.

예제. 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. 정확한 검정 옵션을 설치한 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 > 레저시 대화 상자 > 2 대응 표본...

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 검정은 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 검정은 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 검정을 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 디자인에서 실험 개입에 따라 반응에 다르게 나타나는 변화를 파악하는 데 유용합니다.

데이터가 범주형이면 주변 동질성 검정을 사용합니다. 이 검정은 McNemar 검정을 이분형 반응에서 다항 반응으로 확장한 검정 방법입니다. 카이제곱 분포를 사용하여 반응의 변화를 확인하고 전후 디자인에서 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.

## 독립 K-표본 비모수 검정

독립 K-표본 비모수 검정 프로시저는 한 변수에 대해 둘 이상의 케이스 그룹을 비교합니다.

예제. 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  $H$ , 중앙값 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 > 레저시 대화 상자 > K 독립 표본...

2. 하나 이상의 숫자변수를 선택합니다.

3. 그룹화 변수를 하나 선택하고 범위지정을 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, 중앙값 검정, Jonckheere-Terpstra 검정은 모두 여러 개의 독립 표본이 동일한 모집단에서 비롯된 것인지를 검정합니다.

Mann-Whitney의  $U$  검정이 확장된 **Kruskal-Wallis H** 검정은 일원 분산 분석의 비모수 동일 분석으로 분포 위치에서의 차이를 찾아냅니다. 강력하지는 않지만 더 일반적인 중앙값 검정은 위치와 형태에서 분포 차이를 찾아냅니다. Kruskal-Wallis의  $H$  검정과 중앙값 검정은 해당 표본을 가져온  $k$  모집단의 연역 순위가 없다고 가정합니다.

$k$  모집단에 원래부터 연역 순위(오름차순이나 내림차순)가 있는 경우 **Jonckheere-Terpstra** 검정이 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 최소값과 최대값으로 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.



## NPART TESTS 명령 추가 기능(독립 K-표본 비모수 검정)

또한 명령 구문을 사용하여 중앙값 검정에 대해 관측된 중앙값 이외의 값을 지정할 수 있습니다(MEDIAN 하위 명령 함께 사용).

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

## 대응 K-표본 비모수 검정

대응 K-표본 비모수 검정 프로시저는 둘 이상의 변수 분포를 비교합니다.

**예제.** 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...

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 디자인의 비모수 분석이거나 셀당 하나의 관측값을 갖는 2원배치 분산 분석을 나타냅니다. Friedman은 대응변수  $k$ 가 동일한 모집단에서 비롯되었음을 나타내는 귀무가설을 검정합니다. 각 케이스마다  $k$  변수는 1부터  $k$ 까지 순위가 정해집니다. 검정 통계량은 이러한 순위를 기반으로 합니다.

**Kendall의 W**는 Friedman 통계의 정규성을 나타냅니다. Kendall의 W는 등급 평가자 간의 동의 측도인 일치 계수로 해석할 수 있습니다. 각 케이스는 감정 평가나 등급 평가자이며 각 변수는 평가되는 품목이나 사람을 나타냅니다. 각 변수의 경우 순위합이 계산됩니다. Kendall의 W는 0(동의 안함)과 1(완전 동의) 사이의 순위를 차지합니다.

**Cochran의 Q**는 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.

### NPART TESTS 명령 추가 기능(대응 K-표본 비모수 검정)

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



---

## 제 28 장 다중 응답 분석

---

### 다중 응답 분석

다중 이분형 및 다항 범주형 세트를 분석하는 데 두 프로시저를 사용할 수 있습니다. 다중반응 빈도분석 프로시저는 빈도표를 표시합니다. 다중반응 교차 분석표 프로시저는 2-3차원 교차 분석표를 표시합니다. 프로시저를 사용하기 전에 다중 응답 세트를 정의해야 합니다.

**예제.** 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 맵핑해야 합니다. 이를 수행하는 방법은 두 가지입니다. 하나는 각 선택(예: American, United, TWA, USAir 및 기타)에 해당하는 변수를 정의하는 것입니다. 승객이 United를 선택하는 경우 변수 *united*가 코드 1에 할당되고, 그렇지 않으면 0에 할당됩니다. 이는 맵핑 변수의 다중 이분형 방법입니다. 반응을 맵핑하는 다른 방법은 질문에 가능한 최대 반응 수를 평가하고 항공사를 지정하는 데 사용되는 코드를 사용하여 같은 수의 변수를 설정하는 다중 범주형 방법입니다. 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 맵핑 방법이나 사용할 수 있지만 선택하는 방법은 반응 분포에 따라 다릅니다.

---

### 다중 응답 세트 정의

다중 응답 세트 정의 프로시저로 기본 변수를 다중 이분형 세트나 다항 범주형 세트로 그룹화하여 빈도표와 교차 분석표를 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 제거를 클릭합니다. 변수 세트를 바꾸려면 목록에서 선택한 다음 변수 세트 정의 특성을 수정하고 바꾸기를 클릭합니다.

기본 변수를 이분형이나 범주형으로 코딩할 수 있습니다. 이분형 변수를 사용하려면 이분형을 선택하여 다중 이분형 세트를 작성합니다. 개수화 값에 대해 정수 값을 입력합니다. 개수화 값의 발생이 한 번 이상인 변수는

다중 이분형 세트의 범주가 됩니다. 범주형을 선택하여 성분 변수와 값 범위가 같은 다항 범주형 세트를 작성합니다. 다항 범주형 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...

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 추가를 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 이분형 세트에서는 결과에 나타난 범주 이름이 그룹의 기본 변수로 정의된 변수 레이블에서 비롯됩니다. 변수 레이블이 정의되지 않은 경우에는 변수 이름을 레이블로 사용합니다. 다항 범주형 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 이분형 세트의 표 작성으로부터 임의의 변수에 대한 결측값이 있는 케이스는 제외합니다. 이분형 변수 세트로 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다중 이분형 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.** 다항 범주형 세트의 표 작성에서 임의의 변수에 대한 결측값이 있는 케이스는 제외됩니다. 범주형 변수 세트로 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다항 범주형 세트의 성분 변수에 지정한 범위 내 유효한 값이 없을 때 결측된 것으로 간주됩니다.

**예제.** 설문조사 질의로부터 작성한 각 변수는 기본 변수가 됩니다. 다중반응 항목을 분석하려면 이러한 변수를 다중 이분형 세트나 다항 범주형 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 설문 조사하고 이분형 변수를 사용하여 **다중 이분형 세트**를 정의할 경우 해당 변수 세트의 세 변수는 각각 그룹화 변수의 범주가 됩니다. 세 항공사의 개수와 백분율이 하나의 빈도표에 나타납니다. 세 항공사를 모두 선택한 반응자가 없을 경우 두 변수를 작성하여 각각 세 항공사에 하나씩 코드를 부여할 수 있습니다. **다항 범주형 세트**를 정의할 경우에는 기본 변수에 같은 코드를 추가하여 값의 표로 작성할 수 있습니다. 결과로 나타나는 값의 변수 세트는 각 기본 변수에 대한 값과 같습니다. 예를 들어, United로 반응한 30명의 반응자 중 5명은 동 항공사를 항공사 1로, 25명은 항공사 2로 반응했을 수 있습니다. 세 항공사에 대한 개수와 백분율이 하나의 빈도표에 표시됩니다.

**통계.** 개수가 표시된 빈도표, 반응 백분율, 케이스 백분율, 유효 케이스 수, 결측 케이스 수를 선택할 수 있습니다.

다중반응 빈도분석 데이터 고려 사항

**데이터.** 다중 응답 세트를 사용합니다.

**가정.** 개수와 백분율을 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 교차 분석표

다중반응 교차 분석표 프로시저로 정의한 다중 응답 세트, 기본 변수, 조합 등을 교차 분석표로 만들 수 있습니다. 또한, 케이스나 반응을 기초로 셀 백분율을 구하고 결측값 처리를 수정하거나 대응 교차 분석표를 만들 수 있습니다. 먼저 하나 이상의 다중 응답 세트를 정의해야 합니다. "다중 응답 세트 정의"를 참조하십시오.

다중 이분형 세트에서는 결과에 나타난 범주 이름이 그룹의 기본 변수로 정의된 변수 레이블에서 비롯됩니다. 변수 레이블이 정의되지 않은 경우에는 변수 이름을 레이블로 사용합니다. 다항 범주형 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 8 자씩 세 줄로 표시합니다. 단어가 잘리지 않도록 하려면 행과 열 항목을 뒤바꾸고 레이블을 다시 정의합니다.

**예제.** 다중 이분형 세트와 다항 범주형 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 세트  
트로 조합하면 서비스와 스케줄이 포함된 질문으로 항공사 선택을 교차 분석표화할 수 있습니다.

통계. 셀, 행, 열, 총계와 셀, 행, 열, 전체 백분율을 사용한 교차 분석표를 선택할 수 있습니다. 셀 백분율은  
케이스나 반응에 따라 달라집니다.

다중반응 교차 분석표 데이터 고려 사항

데이터. 다중 응답 세트나 숫자 범주형 변수를 사용합니다.

가정. 개수와 백분율을 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 교차 분석표 차원에 대해 여러 숫자변수나 다중 응답 세트를 선택합니다.

3. 각 기본 변수의 범위를 지정합니다.

선택적으로 제어변수나 다중 응답 세트의 범주마다 2차원 교차 분석표를 사용할 수 있습니다. 레이어 목록에  
대한 항목을 하나 이상 선택합니다.

## 다중반응 교차 분석표 범주지정

값 범위는 교차 분석표의 기본 변수에 대해 정의해야 합니다. 표로 만들 범주의 최소값과 최대값을 정수로 입  
력합니다. 범위 밖의 범주는 분석 시 제외됩니다. 범위 내 값은 정수로 가정되며 정수가 아닌 수는 잘립니다.

## 다중반응 교차 분석표: 옵션

셀 백분율. 셀 개수는 항상 표시됩니다. 행 백분율, 열 백분율, 두 가지 방법의 테이블(전체) 백분율 등을 선택  
하여 표시할 수 있습니다.

백분율 계산 기준. 셀 백분율을 케이스(또는 반응자)에 대한 기준으로 삼을 수 있습니다. 다항 범주형 세트에  
대해 변수 일치를 선택할 때는 이렇게 할 수 없습니다. 셀 백분율을 반응에 대한 기준으로 삼을 수 있습니다.  
다중 이분형 세트에서는 반응 수가 케이스에 대한 개수화 값 수와 같습니다. 다항 범주형 세트에서는 반응 수  
가 지정 범위 내 값의 수와 같습니다.

결측값. 다음 중 하나 또는 모두를 선택할 수 있습니다.

- 이분형 결측데이터의 목록별 제외. 다중 이분형 세트의 표 작성으로부터 임의의 변수에 대한 결측값이 있는  
케이스는 제외합니다. 이분형 변수 세트에 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다  
중 이분형 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. 다항 범주형 세트의 표 작성에서 임의의 변수에 대한 결측값이 있는 케이스는 제외됩니다. 범주형 변수 세트로 정의된 다중 응답 세트에만 적용됩니다. 기본적으로 케이스는 다항 범주형 세트의 성분 변수에 지정한 범위 내 유효한 값이 없을 때 결측된 것으로 간주됩니다.

기본적으로 두 다항 범주형 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 백분율을 기준으로 합니다. 다중 이분형 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.



---

## 제 29 장 결과 보고

---

### 결과 보고

케이스 목록 및 기술 통계는 데이터를 연구하고 설명하는 기본 도구입니다. 데이터 편집기나 요약 프로시저를 사용하여 케이스 목록을, 빈도 프로시저를 사용하여 빈도 개수와 기술 통계를, 평균 프로시저를 사용하여 부모 집단 통계를 얻을 수 있습니다. 각각 정보를 쉽게 파악할 수 있도록 디자인된 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.

**예제.** 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.

**미리보기.** 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 > 행별 요약 보고서...

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 요약을 클릭하여 요약 축도를 지정합니다.
5. 전체 요약 통계가 표시된 보고서에 대해 요약을 클릭하여 요약 축도를 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 맞춤 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: 윤곽

보고서 윤곽에서 각 보고서 쪽의 너비와 길이, 쪽에서 보고서 위치, 공백줄과 레이블 삽입 등을 제어합니다.

쪽 윤곽. 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 10 줄까지 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.

**예제.** 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.

**미리보기.** 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 > 열별 요약 보고서...

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 요약을 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 삽입을 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 윤곽

열별 요약 보고서를 위한 보고서 윤곽은 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.
- 중앙값, 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.

---

## 제 30 장 신뢰성 분석

신뢰성 분석을 사용하여 측정 척도의 특성과 척도를 구성하는 항목을 연구할 수 있습니다. 신뢰성 분석 프로시저에서는 공통적으로 사용된 척도 신뢰도의 척도 수를 계산하고 척도의 개별 항목 간 관계에 대한 정보를 제공합니다. 급내 신뢰도 추정값을 계산하도록 급내 상관계수를 사용할 수 있습니다.

**예제.** 현재 질문지의 척도 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, 항목 간 상관 및 공분산, 신뢰도 추정값, 분산 분석표, 급내 상관계수, Hotelling의  $T^2$  및 Tukey의 가법성 검정을 포함합니다.

**모형.** 다음의 신뢰도 모형을 사용할 수 있습니다.

- **알파(Cronbach).** 이것은 내적일치(internal consistency) 모형으로 평균 항목 간 상관을 기준으로 합니다.
- **반분계수.** 이 모형에서는 해당 척도를 두 부분으로 분리하고 부분간 상관을 검사합니다.
- **Guttman.** 이 모형에서는 실제 신뢰도에 대한 Guttman의 하한값을 계산합니다.
- **동형.** 이 모형에서는 여러 복제의 모든 항목에는 등분산과 동일한 오차 분산이 있는 것으로 가정합니다.
- **절대 동형.** 이 모형에서는 동형 모형을 가정하며 여러 항목에서 평균은 동일한 것으로 가정합니다.

신뢰성 분석 데이터 고려 사항

**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.

**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.

신뢰성 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 척도 > 신뢰성 분석...

2. 둘 이상의 변수를 가법 척도의 잠재적 성분으로 선택합니다.
3. 모형 드롭다운 목록에서 모형을 선택합니다.

## 신뢰성 분석: 통계

척도와 문항을 기술하는 다양한 통계를 선택할 수 있습니다. 기본적으로 보고되는 통계에는 케이스의 수, 문항의 수가 있으며 다음과 같은 신뢰도 추정값도 포함됩니다.

- **Alpha 모형.** 알파계수; 이분형 데이터의 경우 이 모형은 Kuder-Richardson 20(KR20) 계수와 동일합니다.
- **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.
- **동형 및 절대 동형 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 기술통계. 여러 케이스의 척도나 항목에 대한 기술통계를 생성합니다.

- **항목.** 여러 케이스의 항목에 대한 기술통계를 생성합니다.
- **척도.** 척도에 대한 기술통계를 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.

**요약값.** 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.
- **공분산.** 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.
- **상관.** 항목 간 상관관계에 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.

**항목 간.** 항목 간 상관 행렬이나 공분산 행렬을 생성합니다.

**분산 분석표.** 동일한 평균에 대한 검정을 생성합니다.

- **F 검정.** 반복 측도 분산 분석표를 출력합니다.
- **Friedman 카이제곱.** Friedman 카이제곱과 Kendall 일치계수를 표시합니다. 이 옵션은 순위 형식의 데이터에 적합합니다. 카이제곱 검정은 분산 분석표의 일반적인 F 검정을 대체합니다.
- **Cochran 카이제곱.** Cochran의 Q를 표시합니다. 이 옵션은 이분형 데이터에 적합합니다. Q 통계는 분산 분석표에서 일반적인 F 통계를 대체합니다.

**Hotelling의 T 제곱.** 척도의 모든 항목의 평균은 0이라는 귀무가설에 대해 다변량 검정을 생성합니다.

**Tukey의 가법성 검정.** 항목 간에 다변량 반복을 하지 않는 가정에 대한 검정을 생성합니다.

급내 상관계수. 케이스 내에서 측도의 일치 또는 값의 일치를 생성합니다.

- **모형.** 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 변량을 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.

- 상관 행렬을 읽고 분석합니다.
- 나중에 분석하기 위해 상관 행렬을 씁니다.
- 반분 방법으로 동일반분 이외의 분할 방법을 지정합니다.

명령 구문에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.



## 제 31 장 다차원척도법

다차원척도법은 오브젝트간 또는 케이스간 거리측도 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 측정된 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.

**예제.** 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 행렬이나 데이터로부터 거리 행렬 계산하기를 선택합니다.
4. 데이터로부터 거리 행렬 계산하기를 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 거리측도를 지정합니다.

---

## 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 정방대칭형을 선택할 수 없습니다.

---

## 다차원척도법 척도 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 척도를 작성하기 위한 세부사항을 지정할 수 있습니다.

척도. 분석에 필요한 상이성 척도를 지정할 수 있습니다. 해당 데이터 유형에 해당하는 척도에서 하나를 선택한 후 척도 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 척도 또는 파이제곱 척도를 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 측정될 경우) 근접성(이분형 데이터에는 적용할 수 없음)을 계산하기 전에 값을 표준화해야 할 경우가 있습니다. 표준화 드롭다운 목록에서 표준화 방법을 선택합니다. 표준화할 필요가 없으면 없음을 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 허용을 선택할 수 있습니다.

---

## 다차원척도법: 옵션

다차원척도법 분석을 위해 다음 옵션을 지정할 수 있습니다.

**출력.** 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 **결측값**으로 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.



---

## 제 32 장 비율 통계

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.

**예제.** 다섯 지역에서 주택의 견적가와 판매가 간의 비율이 균일합니까? 결과를 통해 비율의 분포가 각 지역마다 매우 다르다는 것을 알 수 있을 것입니다.

**통계.** 중앙값, 평균, 가중 평균, 신뢰구간, COD(산포계수), 중앙값 중심 변동 계수, 평균 중심 변동 계수, PRD(가격 관련 차), 표준 편차, AAD(평균 절대 편차), 범위, 최소값 및 최대값, 중앙값 비율 내 사용자가 지정한 범위 또는 백분율에 대해 계산된 집중 지수를 구할 수 있습니다.

비율 데이터 고려 사항

**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 통계 구하기

1. 메뉴에서 다음을 선택합니다.

    분석 > 기술통계량 > 비율...

2. 분자변수를 선택합니다.

3. 분모변수를 선택합니다.

또는 다음 작업을 수행합니다.

- 그룹화 변수를 선택하고 결과에서 그룹의 순서를 지정합니다.
- 결과를 뷰어에 표시할지 여부를 선택합니다.
- 나중에 다시 사용할 수 있도록 결과를 외부 파일에 저장할지 여부를 선택하고 결과를 저장할 파일의 이름을 지정합니다.

---

## 비율 통계

**중심 경향.** 중심 경향 측도는 비율의 분포를 보여주는 통계입니다.

- **중앙값.** 이 값보다 작은 비율의 수와 이 값보다 큰 비율의 수가 같은 값입니다.
- **평균.** 비율을 합한 값을 비율의 전체 수로 나눈 결과입니다.
- **가중 평균.** 분자의 평균을 분모의 평균으로 나눈 결과입니다. 가중 평균은 분모에 의해 기중치를 부여한 비율의 평균이기도 합니다.

- 신뢰구간. 요청하는 경우 평균, 중앙값 및 기중 평균에 대한 신뢰구간을 보여 줍니다. 신뢰수준으로 0 이상, 100 미만의 값을 지정합니다.

산포. 관측값에 분포되어 있는 변동의 양을 측정하는 통계입니다.

- **AAD.** 평균 절대 편차는 중앙값에 대한 비율의 절대 편차를 합하여 비율의 전체 수로 나눈 결과입니다.
- **COD.** 산포계수는 평균 절대 편차를 중앙값 백분율로 표시한 결과입니다.
- **PRD.** 가격 관련 차는 회귀 지수로도 알려져 있으며 평균을 기중 평균으로 나눈 결과입니다.
- **중앙값 중심 COV.** 중앙값 중심 변동 계수는 중앙값 편차의 평균 제곱근을 중앙값 백분율로 표시한 결과입니다.
- **평균 중심 COV.** 평균 중심 변동 계수는 표준 편차를 평균 백분율로 표시한 결과입니다.
- **표준 편차.** 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.** 최소값은 가장 작은 비율입니다.
- **최대값.** 최대값은 가장 큰 비율입니다.

집중 지수. 집중 계수는 구간 내 해당하는 비율 백분율을 측정합니다. 다음 두 가지 방법으로 계산할 수 있습니다.

- **비율 범위.** 여기에서 구간은 구간의 상한값과 하한값을 지정하여 정확하게 정의합니다. 비율 하한값 및 비율 상한값을 입력하고 추가를 클릭하여 구간을 정합니다.
- **비율 범위.** 여기에서 구간은 중앙값 백분율을 지정하여 함축적으로 정의합니다. 0과 100 사이의 값을 입력하고 추가를 클릭합니다. 구간에서 낮은 값은  $(1 - 0.01 \times \text{값}) \times \text{중앙값의 값}$ 과 같으며 높은 값은  $(1 + 0.01 \times \text{값}) \times \text{중앙값의 값}$ 과 같습니다.

---

## 제 33 장 ROC 곡선

이 프로시저는 대상을 분류하는 두 개의 범주를 갖는 변수가 있는 분류 방법의 성능을 평가하는 유용한 방법입니다.

**예제.** 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

**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.

분석 > ROC 곡선...

2. 하나 이상의 검정 확률 변수를 선택합니다.
3. 하나의 상태변수를 선택합니다.
4. 상태변수에 대한 양의 값을 확인합니다.

---

### 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 표준 오차에 대한 모수.** 곡선 아래 영역의 표준 오차 추정 방법을 지정할 수 있습니다. 사용할 수 있는 방법에는 비모수 방법과 이변량 음지수 방법이 있습니다. 또한 신뢰구간의 수준을 설정할 수 있습니다. 사용할 수 있는 범위는 50.1에서 99.9% 까지 해당됩니다.

**결측값.** 결측값 핸들 방법을 지정할 수 있습니다.

---

## 제 34 장 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 재료를 포함하는 이윤 모형이 있지만 시장 변동성으로 인해 비용에 불확실성이 있습니다. 불확실성이 있는 모형에 시뮬레이션을 사용하여 이윤에 대한 효과를 확인할 수 있습니다.

IBM SPSS Statistics의 시뮬레이션은 Monte Carlo 방법을 사용합니다. 불확실한 입력은 삼각분포와 같은 확률 분포로 모델링되며 그러한 입력에 대한 시뮬레이션 값은 이러한 분포에서 생성됩니다. 값이 알려진 입력은 알려진 값에 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 Statistics에서 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 디자인하고 실행할 사용자를 위한 고급 인터페이스입니다. 시뮬레이션 작성기는 시뮬레이션 디자인, 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 디자인하는 시뮬레이션 작성기의 전체 공정능력은 제공하지 않습니다.



---

## 모형 파일을 기반으로 시뮬레이션을 디자인하는 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 시뮬레이션...

2. **SPSS 모형 파일** 선택과 **계속**을 차례로 클릭합니다.

3. 모형 파일을 엽니다.

모형 파일은 IBM SPSS Statistics 또는 IBM SPSS Modeler에서 작성된 모형 PMML을 포함하는 XML 파일입니다. 자세한 정보는 185 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 맞춤을 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. 실행을 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 디자인하는 방법

1. 메뉴에서 다음을 선택합니다.

분석 > 시뮬레이션...

2. 방정식으로 입력 및 계속을 차례로 클릭합니다.

3. 시뮬레이션 작성기의 모형 탭에 있는 새 방정식을 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 맞춤을 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. 실행을 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.
- 시뮬레이션된 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 시뮬레이션된 입력의 분포모수가 다른 효과를 조사합니다.
- 생성할 최대 케이스 수를 설정하거나 꼬리 표본추출을 요청하는 등 고급 옵션을 지정합니다.
- 결과를 사용자 정의합니다.
- 시뮬레이션된 데이터를 데이터 파일에 저장합니다.

---

## 예측 모형 없이 시뮬레이션 디자인

1. 메뉴에서 다음을 선택합니다.

분석 > 시뮬레이션...

2. 시뮬레이션된 데이터 작성을 클릭하고 계속을 클릭하십시오.
3. 시뮬레이션 빌더의 모형 탭에서 시뮬레이션할 필드를 선택하십시오. 활성 데이터 세트에서 필드를 선택하거나 새로 작성을 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 맞춤을 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. 실행을 클릭하여 시뮬레이션을 실행합니다. 기본적으로 시뮬레이션된 데이터는 저장 설정에 지정된 새 데이터 세트에 저장됩니다. 또한 시뮬레이션의 세부사항을 지정하는 시뮬레이션 계획은 저장 설정에 지정된 위치에 저장됩니다.

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 디자인된 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 열기를 클릭합니다.
3. 시뮬레이션 작성기에서 열기 확인란이 선택되지 않았는지 확인하고 계속을 클릭합니다.
4. 시뮬레이션 계획을 엽니다.
5. 시뮬레이션 실행 대화 상자에서 실행을 클릭합니다.

시뮬레이션 작성기에서 시뮬레이션을 실행하려면 다음을 수행합니다.

1. 메뉴에서 다음을 선택합니다.

분석 > 시뮬레이션...

2. 기존 시뮬레이션 계획 열기를 클릭합니다.
3. 시뮬레이션 작성기에서 열기 확인란을 선택하고 계속을 클릭합니다.
4. 시뮬레이션 계획을 엽니다.
5. 시뮬레이션 탭에서 수정하려는 설정을 수정합니다.
6. 실행을 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 맞추습니다.
- 시뮬레이션한 입력의 분포를 변경합니다.
- 결과를 사용자 정의합니다.
- 시뮬레이션한 데이터를 데이터 파일에 저장합니다.

---

## 시뮬레이션 작성기

시뮬레이션 작성기는 시뮬레이션을 디자인하고 실행하기 위한 전체적인 공정능력을 제공합니다. 시뮬레이션 작성기를 사용하여 다음과 같은 일반 작업을 수행할 수 있습니다.

- PMML 모형 파일에서 정의한 IBM SPSS 모형에 대한 시뮬레이션을 디자인하고 실행합니다.
- 지정한 사용자 정의 방정식 세트로 정의한 예측 모형에 대한 시뮬레이션을 디자인하고 실행합니다.
- 예측 모형이 없이 데이터를 생성하는 시뮬레이션을 디자인하고 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 Statistics 또는 IBM SPSS Modeler에서 작성된 모형 PMML을 포함하는 XML 파일입니다. 예측 모형은 IBM SPSS Statistics 내의 선형 회귀분석 및 의사결정 트리와 같은 프로시저로 만들어지며 모형 파일로 내보낼 수 있습니다. 찾아보기를 클릭하고 원하는 파일을 탐색하여 다른 모형 파일을 사용할 수 있습니다.

시뮬레이션이 지원하는 PMML 모형

- 선형 회귀분석
- 일반화 선형 모형
- 일반 선형 모형
- 이분형 로지스틱 회귀분석
- 다항 로지스틱 회귀분석
- 순서 다항 회귀분석
- Cox 회귀분석
- 트리
- 부스팅된 트리(C5)
- 판별 분석
- 두 단계 군집
- K-Means 군집 분석
- 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 세트에 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 방정식을 클릭하여 방정식을 만듭니다. 이렇게 하면 방정식 편집기가 열립니다. 기존 방정식을 수정하고 이를 복사하여 새 방정식의 템플릿으로 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.

$$\text{수입} = \text{가격} * \text{판매량}$$

$$\text{지출} = \text{고정비} + \text{판매량} * (\text{단위\_비용\_재료} + \text{단위\_비용\_임금})$$

$$\text{이윤} = \text{수입} - \text{지출}$$

모형 없이 시뮬레이션된 데이터 작성. 예측 모형 없이 데이터를 시뮬레이션하려면 이 옵션을 선택하십시오. 활성 데이터 세트에서 필드를 선택하거나 새 필드를 정의하도록 새로 작성을 클릭하여 시뮬레이션할 필드를 지정하십시오.

## 방정식 편집기

방정식 편집기를 사용하여 예측 모형에 대한 사용자 정의 방정식을 만들거나 수정할 수 있습니다.

- 방정식의 표현식은 활성 데이터 세트의 필드나 방정식 편집기에서 정의한 새 입력 필드를 포함할 수 있습니다.
  - 측정 수준, 값 레이블 및 결과가 대상에 대해 생성되었는지에 대한 여부와 같은 대상의 특성을 지정할 수 있습니다.
  - 이전에 정의된 방정식의 대상을 현재 방정식에 입력으로 사용하여 결합된 방정식을 만들 수 있습니다.
  - 방정식에 구체적인 설명을 첨부할 수 있습니다. 설명은 모형 탭에서 방정식과 함께 표시됩니다.
1. 대상 이름을 입력합니다. 선택적으로 대상 입력란 아래의 편집을 클릭하여 정의된 입력 대화 상자를 열어 대상의 기본 특성을 변경할 수 있습니다.
  2. 표현식을 작성하려면 숫자 표현식 필드에 성분을 붙여넣거나 숫자 표현식 필드에 직접 입력합니다.
- 활성 데이터 세트의 필드를 사용하여 표현식을 작성하거나 새로 만들기 단추를 클릭하여 새 입력을 정의할 수 있습니다. 이렇게 하면 입력 정의 대화 상자가 열립니다.
  - 함수 그룹 목록에서 그룹을 선택하고 함수 목록의 함수를 두 번 클릭하여 함수를 붙여 넣거나 함수를 선택한 다음 함수 그룹 목록 옆에 있는 화살표를 클릭할 수 있습니다. 물음표로 표시된 모수를 입력합니다. 모

두리는 레이블이 있는 함수 그룹에서는 사용 가능한 모든 함수 목록을 제공합니다. 대화 상자의 지정된 영역에는 현재 선택한 함수에 대한 간략한 설명이 표시됩니다.

- 문자열 상수는 따옴표로 묶어야 합니다.
- 값에 소수점이하자리수가 포함된 경우 소수점 구분자로는 마침표(.)를 사용해야 합니다.

참고: 시뮬레이션은 문자열 대상이 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.

**대상.** 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 맞춤을 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 차트 및 테이블에 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 예측변수가 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 맞춤을 클릭합니다.

가장 근접한 분포 및 연관된 모수가 과거 데이터의 히스토그램(또는 막대형 차트)에 겹친 분포 도표와 함께 분포 열에 표시됩니다. 시뮬레이션한 입력 사이의 상관은 상관 설정에 표시됩니다. 입력에 대한 행을 선택하고 세부 사항 맞춤을 클릭하여 맞춤 결과를 검사하고 특정 입력에 대한 자동 분포 맞춤을 사용자 정의할 수 있습니다. 자세한 정보는 190 페이지의 『세부사항 맞춤』 주제를 참조하십시오.

입력에 대한 행을 선택하고 맞춤을 클릭하여 특정 입력에 대한 자동 분포 맞춤을 실행할 수 있습니다. 활성 데이터 세트의 필드와 일치하는 시뮬레이션한 모든 필드에 대한 상관도 자동으로 계산됩니다.

참고: 연속형 및 순서 입력의 경우 검정된 분포에 대해 사용 가능한 적합을 찾을 수 없으면 경험 분포가 가장 근접한 것으로 나타납니다. 연속형 입력의 경우 경험 분포는 과거 데이터의 누적 분포 함수입니다. 순서 입력의 경우 경험 분포는 과거 데이터의 범주형 분포입니다.

수동으로 분포 지정. 유형 드롭다운 목록에서 분포를 선택하고 모수 눈금에서 분포모수를 입력하여 시뮬레이션한 입력에 대한 확률 분포를 수동으로 지정할 수 있습니다. 분포의 모수를 입력하고 나면 지정한 모수에 근접한 분포의 도표 예가 모수 눈금 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 눈금의 왼쪽 옆을 클릭하고 숫자 값으로 범주를 지정합니다. 오른쪽 옆에 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 구간에 대한 확률 합계가 1과 같은 그런 각 구간에 할당된 확률을 가진 구간 세트에 포함되어 있습니다. 주어진 구간 내의 값은 해당 구간에 정의된 균일분포로부터 작성됩니다. 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 모델링하여 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). 자세한 정보는 주제를 참조하십시오.

**고정 입력 값 지정.** 분포 옆의 유형 드롭다운 목록에서 고정을 선택하고 고정 값을 입력하여 고정 입력 값을 지정합니다. 입력 값이 숫자나 문자열이나에 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 옆에서 확인란을 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 통해 각각 지정한 값에 대해 시뮬레이션한 케이스의 독립 세트(사실상 별도의 세트)를 생성하여 고정 입력 또는 시뮬레이션한 입력 값의 분포모수에 대한 계통 변화 효과를 조사할 수 있습니다. 민감도 분석을 지정하려면 시뮬레이션한 입력을 선택하고 민감도 분석을 클릭합니다. 민감도 분석은 단일 고정 입력 또는 시뮬레이션한 입력 값에 대한 단일 분포모수로 제한됩니다. 자세한 정보는 191 페이지의 『민감도 분석』 주제를 참조하십시오.

## 맞춤 상태 아이콘

다음에 맞춤 열에 있는 아이콘은 각 입력 필드의 맞춤 상태를 나타냅니다.

표 3. 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 반복이 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 막대형 차트)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.

**모수.** 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). 자세한 정보는 주제를 참조하십시오. 범주형 분포에서 모수 이름은 범주이며 모수 값은 연관된 확률입니다.

**사용자 정의된 분포 세트에 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 실행을 클릭하십시오.

**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 반복이라고 합니다.

**반복.** 이 옵션을 선택하면 다양해질 입력에 대해 값의 세트를 지정할 수 있습니다.

- 분포 모수의 값을 변경하는 경우 드롭 다운 목록에서 모수를 선택하십시오. 반복 눈금으로 모수 값에 값의 세트를 입력합니다. 계속을 클릭하면 연관된 입력의 모수 눈금에 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 눈금으로 모수 값에서 확률 세트를 지정하십시오. 다른 범주 또는 구간의 확률이 이에 따라 자동으로 조정됩니다.

**반복하지 않음.** 이 옵션을 사용하여 입력에 대한 반복을 취소합니다. 계속을 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 맞춤 때 시뮬레이션한 입력 사이의 상관이 자동으로 계산되도록 지정합니다.

**맞춤 시 상관을 다시 계산하지 않음.** 수동으로 상관을 지정하고 분포를 활성 데이터 세트에 자동으로 맞춤 때 덮어써지는 것을 방지하려면 이 옵션을 선택합니다. 상관 눈금에 입력한 값은 -1과 1 사이여야 합니다. 값이 0 이면 연관된 입력 쌍 사이에 상관이 없도록 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 세트의 분할표 계산.** 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 분할표 사용.** 기본적으로 분할표가 포함된 시뮬레이션 계획을 로드할 때 계획의 테이블을 사용합니다. 활성 데이터 세트의 분할표 계산을 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 제한을 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.

**경험 분포.** 연속형 입력의 경우 경험 분포는 과거 데이터의 누적 분포 함수입니다. 연속형 입력의 경험 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 생성을 클릭하십시오. 기본값은 629111597입니다.

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 막대형 차트도 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 백분율을 표시하는 막대형 차트를 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 막대형 차트가 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 차트에서 이동할 수 있는 참조선의 초기 위치를 지정할 수 있습니다. 하단 및 상단 선에 지정된 값은 백분위수가 아니라 가로축의 위치를 나타냅니다. -무한을 선택하여 하단 선을 제거하거나 무한을 선택하여 상단 선을 제거할 수 있습니다. 기본적으로 선은 5번째 및 95번째 백분위수에 있습니다. 다중 대상 또는 민감도 분석 반복 결과로 인해 다중 분포 함수가 단일 차트에 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 빼 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 반복 결과로 인해 다중 분포 함수가 단일 차트에 표시되는 경우 참조선은 첫 번째 반복이나 첫 번째 대상의 분포에만 적용됩니다. PDF나 CDF 차트에서 액세스하는 차트 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 차트 한 개 및 누적 분포 함수에 대한 다른 차트로써 단일 차트에 표시합니다. 이 옵션을 선택하지 않을 경우 각 대상의 결과는 별도의 차트에 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 막대형 차트를 생성하는 데 사용됩니다. 예측 범주를 선택하면 막대형 차트가 생성됩니다. 예측 확률을 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 막대형 차트가 모든 반복의 결과를 포함한 묶은 가로 막대형 차트로 표시됩니다. 범주를 함께 그룹화하도록 선택하거나 반복을 함께 그룹화할 수 있습니다.

## 결과

**토네이도 차트.** 토네이도 차트는 다양한 메트릭을 사용하여 대상과 시뮬레이션한 입력 사이의 관계를 표시하는 막대형 차트입니다.

- **입력 값이 있는 대상의 상관.** 이 옵션은 주어진 대상과 시뮬레이션한 각 입력 사이에 대한 상관계수의 토네이도 차트를 만듭니다. 이 유형의 토네이도 차트는 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 차트를 만들어 사용자가 각 입력이 대상의 전체 불확실성에 기여하는 정도를 평가할 수 있게 해줍니다. 이 토네이도 차트 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 빼으로써 각 시뮬레이션한 입력 조절이 대상에 미치는 효과를 표시하는 토네이도 차트를 만듭니다. 이 토네이도 차트 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 차트에 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 히트 맵으로 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 구간을 선택하고 수를 지정합니다. 사용자 정의 백분위수를 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, 중앙값, 최소값 및 최대값, 지정한 수준의 평균의 신뢰구간, 대상 분포의 제 5와 제 95 백분위수가 포함됩니다. 범주형 대상의 경우 테이블에는 대상의 각 범주

에 속하는 케이스의 백분율이 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 상관관계수 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 차트, 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 하한을 선택하고 최소값 열에 값을 입력하십시오. 상한을 설정하려면 한계 열에서 상한을 선택하고 최대값 열에 값을 입력하십시오. 하한 및 상한을 모두 설정하려면 한계 열에서 둘 다를 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, 차트 및 테이블에 대한 결과 설정, 맞춤에 대한 고급 옵션 설정, 경험 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 디자인되었습니다. 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 맞춤니다.
- 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 눈금은 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 맞추는 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 맞추는 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 눈금에서 입력에 대한 항목을 선택하십시오.

### 민감도 분석 지정

민감도 분석을 통해 각각 지정한 값에 대해 시뮬레이션한 케이스의 독립 세트(사실상 별도의 세트)를 생성하여 고정 입력 또는 시뮬레이션한 입력 값의 분포모수에 대한 계통 변화 효과를 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 191 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 맞춤을 선택한 다음 목록에서 필드를 선택하여 일치하는 필요한 필드를 수정합니다.
3. 모두 맞춤을 클릭합니다.

맞춰진 각 입력의 경우 데이터에 가장 근접한 분포가 과거 데이터의 히스토그램(또는 막대형 차트)에 겹친 분포 도표와 함께 표시됩니다. 허용 가능한 맞춤을 찾을 수 없는 경우 경험 분포가 사용됩니다. 경험 분포에 맞춰진 입력의 경우 경험 분포가 해당 히스토그램으로 나타나므로 과거 데이터의 히스토그램만 볼 수 있습니다.

참고: 상태 아이콘의 전체 목록은 188 페이지의 『시뮬레이션한 필드』 주제를 참조하십시오.

## 확률 분포 수정

시뮬레이션한 입력의 확률 분포를 수정하고 필요한 경우 시뮬레이션한 입력을 고정 입력으로 변경할 수 있습니다(그 반대로 가능).

1. 입력을 선택하고 수동으로 분포 설정을 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 188 페이지의 『시뮬레이션한 필드』 주제를 참조하십시오.

맞춤 시 범주형 입력의 사용자 결측값 포함. 활성 데이터 세트의 데이터에 다시 맞춤 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 포함되면 모든 입력 값이 유효해야 합니다.

## 결과 탭

결과 탭을 사용하여 시뮬레이션으로 생성된 결과를 사용자 정의할 수 있습니다.

밀도 함수. 밀도 함수는 시뮬레이션의 결과 세트를 살펴보는 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 백분율을 표시하는 막대형 차트가 생성됩니다.
- **누적 분포 함수.** 누적 분포 함수는 지정한 값보다 적거나 같은 대상 값의 확률을 표시합니다.

토네이도 차트. 토네이도 차트는 다양한 메트릭을 사용하여 대상과 시뮬레이션한 입력 사이의 관계를 표시하는 막대형 차트입니다.

- **입력 값이 있는 대상의 상관.** 이 옵션은 주어진 대상과 시뮬레이션한 각 입력 사이에 대한 상관계수의 토네이도 차트를 만듭니다.
- **분산에 대한 기여도.** 이 옵션은 시뮬레이션한 각 입력으로부터 대상의 분산에 대한 기여도를 표시하는 토네이도 차트를 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 차트를 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 계획 이 분할표에서 범주형 데이터를 생성하도록 지정하면 범주형 분포를 사용하는 입력 사이의 연관 분할표가 표시 됩니다.

별도의 대상의 결과로부터 오버레이. 시뮬레이션하는 예측 모형에 대상이 여러 개 들어 있을 경우 별도의 대상 결과를 단일 차트에 표시할 것인지를 지정할 수 있습니다. 이 설정은 확률 밀도 함수, 누적 분포 함수 및 상자 도표의 차트에 적용됩니다. 예를 들어 이 옵션을 선택한 경우 모든 대상의 확률 밀도 함수는 단일 차트에 표 시됩니다.

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 184 페이지의 『시뮬레이션 계획에서 시뮬레이션을 실행하는 방법』 주제를 참조하십시오.

---

## 시뮬레이션의 차트 결과로 작업

시뮬레이션에서 생성된 여러 차트는 대화형 기능이 있어 표시를 사용자 정의할 수 있습니다. 대화형 기능은 결과 뷰어에서 차트 오브젝트를 활성화(두 번 클릭)하여 사용할 수 있습니다. 모든 시뮬레이션 차트는 그래프보드 시각화입니다.

연속형 대상의 확률 밀도 함수 차트. 이 차트에는 차트를 별도의 영역으로 나누는 두 개의 이동 수직 참조선이 있습니다. 차트 아래의 테이블은 대상이 각 영역에 있을 확률을 표시합니다. 여러 개의 밀도 함수가 같은 차트에 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 차트의 차트 옵션 단추를 클릭하여 사용할 수 있습니다. 특히 슬라이더 위치를 명시적으로 설정하고 고정 참조선을 추가하며 차트 보기를 연속 곡선에서 히스토그램으로(그 반대도 가능) 이동할 수 있습니다. 자세한 정보는 199 페이지의 『차트 옵션』 주제를 참조하십시오.

연속형 대상의 누적 분포 함수 차트. 이 차트에는 동일한 두 개의 이동 가능한 수직 참조선과 위의 확률 밀도 함수 차트에서 설명한 연관된 테이블이 있습니다. 또한 차트 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 『차트 옵션』 주제를 참조하십시오.

민감도 분석 반복이 있는 범주형 대상의 막대형 차트. 민감도 분석 반복이 있는 범주형 대상의 경우 예측 대상 범주의 결과는 모든 반복의 결과를 포함하는 묶은 가로 막대형 차트로 표시됩니다. 차트에는 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 차트에서 모든 대상의 상자도표를 표시하도록 선택하면 군집 상자도표가 생성됩니다. 차트에는 드롭다운 목록이 있어 대상 또는 반복을 군집화할 수 있습니다.

## 차트 옵션

차트 옵션 대화 상자를 통해 시뮬레이션에서 생성된 확률 밀도 함수 및 누적 분포 함수의 활성화된 차트 표시를 사용자 정의할 수 있습니다.

보기. 보기 드롭다운 목록은 확률 밀도 함수 차트에만 적용됩니다. 보기를 사용하여 차트 보기를 연속 곡선에서 히스토그램으로 전환할 수 있습니다. 이 기능은 여러 개의 밀도 함수가 같은 차트에 표시되는 경우 사용할 수 없습니다. 이 경우 밀도 함수는 연속 곡선으로만 보입니다.

순서. 순서 드롭다운 목록은 누적 분포 함수 차트에만 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. -무한을 선택하여 왼쪽 선을 제거할 수 있으며(음의 무한대로 위치 설정) 무한을 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 반복 결과로 인해 여러 함수가 단일 차트에 표시되는 경우 선이 적용될 특정 함수를 지정할 수 있습니다.

- **시그마.** 대상의 평균에서 지정한 수의 표준 편차를 더하고 빼 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.

차트 옵션 대화 상자에서 연관된 옵션을 선택 취소하고 계속을 클릭하면 참조선이 제거됩니다.



---

## 주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다.

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. 이 책에서 IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산을 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 라이선스까지 부여하는 것은 아닙니다. 라이선스에 대한 의문사항은 다음으로 문의하십시오.

135-700

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

전화번호: 080-023-8080

2바이트(DBCS) 정보에 관한 라이선스 문의는 한국 IBM 고객만족센터에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

Intellectual Property Licensing

Legal and Intellectual Property Law

IBM Japan Ltd.

1623-14, Shimotsuruma, Yamato-shi

Kanagawa 242-8502 Japan

다음 단락은 현지법과 상충하는 영국이나 기타 국가에서는 적용되지 않습니다. 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 비IBM의 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.

(1) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함)간의 정보 교환 및 (2) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이선스 사용자는 다음 주소로 문의하십시오.

135-700

서울특별시 강남구 도곡동 467-12, 군인공제회관빌딩

한국 아이.비.엠 주식회사

고객만족센터

이러한 정보는 해당 조건(예를 들면, 사용료 지불 등)하에서 사용될 수 있습니다.

이 정보에 기술된 라이선스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이선스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 국제 프로그램 라이선스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

본 문서에 포함된 모든 성능 데이터는 제한된 환경에서 산출된 것입니다. 따라서 다른 운영 환경에서 얻어진 결과는 상당히 다를 수 있습니다. 일부 성능은 개발 단계의 시스템에서 측정되었을 수 있으므로 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동일하게 나타날 것이라고는 보증할 수 없습니다. 또한 일부 성능은 추정을 통해 추측되었을 수도 있으므로 실제 결과는 다를 수 있습니다. 이 책의 사용자는 해당 데이터를 본인의 특정 환경에서 검증해야 합니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 비IBM 제품을 반드시 테스트하지 않았으므로, 이들 제품과 관련된 성능의 정확성, 호환성 또는 기타 주장에 대해서는 확인할 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 예제는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.

이러한 샘플 프로그램 또는 파생 제품의 각 사본이나 그 일부에는 반드시 다음과 같은 저작권 표시가 포함되어야 합니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.



이러한 샘플 프로그램 또는 파생 제품의 각 사본이나 그 일부에는 반드시 다음과 같은 저작권 표시가 포함되어야 합니다.

© 귀하의 회사명) (연도). 이 코드의 일부는 IBM Corp.의 샘플 프로그램에서 파생됩니다.

© Copyright IBM Corp. \_연도\_. All rights reserved.

---

## 상표

IBM, IBM 로고 및 ibm.com은 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹의 『저작권 및 상표 정보』([www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml))에 있습니다.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 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

## [가]

### 가설 요약

비모수 검정 139

### 가장 가까운 이웃 거리

최근접 이웃 분석에서 95

### 가중 예측값

GLM 51

### 가중 최소제곱법

선형 회귀분석 69

### 가중 평균

비율 통계 177

### 가중치를 부여하지 않은 최소제곱법

요인 분석 104

### 감마

교차 분석표 17

### 거듭제곱 모형

곡선 추정 80

### 거듭제곱값 추정

GLM 일변량 47, 50, 52

### 거리 59

값 변환 59, 60

명령 추가 기능 60

변수 간 거리 계산 59

상이성 측도 59

예제 59

유사성 측도 60

측도변환 59, 60

케이스 간 거리 계산 59

통계 59

### 거리측도

거리측도 59

계층적 군집 분석 122

최근접 이웃 분석에서 89

### 검증용 표본

최근접 이웃 분석에서 90

### 결측값

다중반응 교차 분석표 160

다중반응 빈도분석 158

대응 2-표본 비모수 검정 152

대응표본 T 검정 35

데이터 탐색 13

독립 표본 T 검정 34

독립 2-표본 비모수 검정 151

### 결측값 (계속)

독립 K-표본 비모수 검정 153

런 검정 148

선형 회귀분석 73

열별 요약 보고서 167

요인 분석 106

이변량 상관 56

이항검정 147

일원 분산 분석 42

일표본 Kolmogorov-Smirnov 검정 149

일표본 T 검정 36

최근접 이웃 분석에서 92

카이제곱 검정 146

편상관 58

행렬 요약 보고서 165

ROC 곡선 179

### 계층적 군집 분석 121

값 변환 122

거리 행렬 122

거리측도 122

고드름 도표 122

군집 변수 121

군집 케이스 121

군집방법 122

군집화 스케줄 122

텐드로그램 122

도표 출력방향 122

명령 추가 기능 123

새 변수 저장 123

소속군집 122, 123

예제 121

유사성 측도 122

측도변환 122

통계 121, 122

### 계층적 분해 45

### 고드름 도표

계층적 군집 분석 122

### 고유 값

선형 회귀분석 73

요인 분석 104

### 곡선 추정 79

모형 80

분산 분석 79

상수 포함 79

### 곡선 추정 (계속)

예측 80

예측 구간 저장 80

예측값 저장 80

잔차 저장 80

### 곱셈

보고서 열 곱하기 167

### 공분산 비율

선형 회귀분석 71

### 공분산 행렬

선형 회귀분석 73

순서 회귀분석 76

판별 분석 98, 100

GLM 51

### 공선성 진단 정보

선형 회귀분석 73

### 과적합 방지 기준

선형모형에서 63

### 관측개수

교차 분석표 18

### 관측빈도

순서 회귀분석 76

### 관측평균

GLM 일변량 47, 50, 52

### 교차 분석표 15

교차 분석표 15

교차 분석표 출력없음 15

다중 응답 159

레이어 16

묶은 가로 막대형 차트 16

셀 표시 18

제어변수 16

통계 17

형식 20

### 군집 분석

계층적 군집 분석 121

효율 126

K-Means 군집 분석 125

### 군집 뷰어

개요 114

군집 모형 정보 114

군집 및 기능 이동 116

군집 및 기능 전지 116

군집 보기 115

군집 뷰어 (계속)

- 군집 비교 보기 117
- 군집 예측변수 중요도 보기 117
- 군집 정렬 116
- 군집 출력결과와 정렬 116
- 군집 크기 보기 117
- 군집의 비교 117
- 군집의 크기 117
- 군집중심 보기 115
- 기능 정렬 116
- 기능 출력결과와 정렬 116
- 기본 보기 116
- 레코드 필터링 119
- 모형 요약 115
- 사용 118
- 셀 내용 정렬 116
- 셀 내용 출력결과와 116
- 셀 분포 보기 117
- 셀의 분포 117
- 예측변수 중요도 117
- 요약 보기 115

군집 빈도

- 이단계 군집 분석 114
- 군집화 114
- 군집 보기 114
  - 전체 출력결과와 114
  - 프로시저 선택 109

그룹 간의 차이

- OLAP 큐브 31
- 그룹 비교
- OLAP 큐브 31
- 그룹 중앙값
- 요약 22
  - 평균 26
  - OLAP 큐브 30

그룹 평균 25, 29

극단값

- 데이터 탐색 12

근접성

- 계층적 군집 분석 121

기능 공간 차트

- 최근접 이웃 분석에서 93

기대개수

- 교차 분석표 18

기대빈도

- 순서 회귀분석 76

기술통계 9

- 기술통계 9

기술통계 (계속)

- 데이터 탐색 12
  - 명령 추가 기능 10
  - 비율 통계 177
  - 빈도분석 6
  - 요약 22
  - 이단계 군집 분석 114
  - 출력 순서 9
  - 통계 9
  - GLM 일변량 47, 50, 52
  - z 스코어 저장 9
- 기하평균
- 요약 22
  - 평균 26
  - OLAP 큐브 30

## [나]

나누기

- 보고서 열 나누기 167
- 누적 분포 함수
- 시뮬레이션에서 193
- 누적 빈도
- 순서 회귀분석 76

## [다]

다중 비교

- 일원 분산 분석 40

다중 응답

- 명령 추가 기능 161

다중 응답 분석

- 교차 분석표 159
- 다중반응 교차 분석표 159
- 다중반응 빈도분석 158
- 빈도표 158

다중 응답 세트

- Codebook 1
- 다중 응답 세트 정의 157
- 범주 157
  - 변수 세트 레이블 157
  - 변수 세트 이름 157
  - 이분형 157

다중 회귀분석

- 선형 회귀분석 69

다중 R

- 선형 회귀분석 73
- 다중반응 교차 분석표 159

다중반응 교차 분석표 (계속)

- 값 범위지정 160
- 결측값 160
- 반응 백분율 계산 기준 160
- 반응변수 세트의 일치 변수 160
- 셀 백분율 160
- 응답 백분율 계산 기준 160

다중반응 빈도분석 158

- 결측값 158
- 다차원척도법 173
- 값 변환 174
  - 거리 행렬 만들기 174
  - 거리측도 174
  - 기준 175
  - 데이터 형태 정의 174
  - 명령 추가 기능 175

예제 173

조건부 175

차원 175

척도화 모형 175

출력 옵션 175

측정 수준 175

통계 173

다항 대비

- 일원 분산 분석 40

GLM 46

단계 선택

- 선형 회귀분석 70

단계적 전진

- 선형모형에서 63

단순 대비

- GLM 46

대비

- 일원 분산 분석 40

GLM 46

대응 연구

- 대응표본 T 검정 35

대응 표본 151, 154

대응 2-표본 비모수 검정 151

검정 유형 151

결측값 152

명령 추가 기능 152

옵션 152

통계 152

대응 K-표본 비모수 검정 154

검정 유형 154

명령 추가 기능 155

통계 155

대응표본 비모수 검정 135  
 필드 136  
 Cochran의 Q 검정 138  
 McNemar 검정 137  
 대응표본 T검정 35  
 결측값 35  
 대응변수 선택 35  
 옵션 35  
 데이터 탐색 11  
 거듭제곱값 변환 13  
 결측값 13  
 도표 12  
 명령 추가 기능 13  
 옵션 13  
 통계 12  
 덴드로그램  
 계층적 군집 분석 122  
 도시-블록 거리  
 최근접 이웃 분석에서 89  
 독립 표본 검정  
 비모수 검정 142  
 독립 표본 비모수 검정 133  
 필드 탭 133  
 독립 표본 T 검정 33  
 결측값 34  
 그룹 정의 34  
 그룹화 변수 34  
 문자열 변수 34  
 신뢰구간 34  
 옵션 34  
 독립 2-표본 비모수 검정 149  
 검정 유형 150  
 결측값 151  
 그룹 정의 150  
 그룹화 변수 150  
 명령 추가 기능 151  
 옵션 151  
 통계 151  
 독립 K-표본 비모수 검정 152  
 검정 유형 153  
 결측값 153  
 그룹화 변수 153  
 명령 추가 기능 154  
 범위지정 153  
 옵션 153  
 통계 153  
 독립 T 검정  
 대응표본 T 검정 35

독립성 검정  
 카이제곱 17  
 동일서브세트  
 비모수 검정 144  
 동형 모형  
 신뢰성 분석 169, 170  
 디트렌딩된 정규도표  
 데이터 탐색 12

## [ 라 ]

람다  
 교차 분석표 17  
 런 검정  
 결측값 148  
 명령 추가 기능 148  
 옵션 148  
 일표본 비모수 검정 130, 131  
 절단점 147, 148  
 통계 148  
 레버리지 값  
 선형 회귀분석 71  
 GLM 51  
 레이어  
 교차 분석표 16  
 로그 모형  
 곡선 추정 80  
 로딩 도표  
 요인 분석 105  
 로지스틱모형  
 곡선 추정 80  
 링크  
 순서 회귀분석 76

## [ 마 ]

마지막  
 요약 22  
 평균 26  
 OLAP 큐브 30  
 막대형 차트  
 빈도분석 7  
 메모리 할당  
 이단계 군집 분석 112  
 모드  
 빈도분석 6  
 모수 추정값  
 순서 회귀분석 76

모수 추정값 (계속)  
 GLM 일변량 47, 50, 52  
 모형 보기  
 비모수 검정 139  
 최근접 이웃 분석에서 92  
 민감도 분석  
 시뮬레이션에서 191

## [ 바 ]

반복  
 요인 분석 104, 105  
 K-Means 군집 분석 126  
 반복 대비  
 GLM 46  
 반복 히스토리  
 순서 회귀분석 76  
 반분계수 신뢰도  
 신뢰성 분석 169, 170  
 배경  
 선형모형에서 61  
 백분위수  
 데이터 탐색 12  
 빈도분석 6  
 시뮬레이션에서 194  
 백분율  
 교차 분석표 18  
 범위  
 기술통계 9  
 비율 통계 177  
 빈도분석 6  
 요약 22  
 평균 26  
 OLAP 큐브 30  
 범주형 필드 정보  
 비모수 검정 143  
 베리맥스 회전  
 요인 분석 105  
 베타 계수  
 선형 회귀분석 73  
 변수 간의 차이  
 OLAP 큐브 31  
 변수 비교  
 OLAP 큐브 31  
 변수 중요도  
 최근접 이웃 분석에서 94  
 변환 행렬  
 요인 분석 103

보고서

- 열 값 곱하기 167
- 열 값 나누기 167
- 열 비교 167
- 열별 요약 보고서 166
- 총계 열 167
- 행별 요약 보고서 163
- 혼합 전체 167

복합모형

- 곡선 추정 80

부스팅

- 선형모형에서 61

부호 검정

- 대응 2-표본 비모수 검정 151
- 대응표본 비모수 검정 136

분류

- ROC 곡선 179

분류표

- 최근접 이웃 분석에서 95

분산

- 기술통계 9
- 데이터 탐색 12
- 빈도분석 6
- 열별 요약 보고서 166
- 요약 22
- 평균 26
- 행별 요약 보고서 164
- OLAP 큐브 30

분산 동질성 검정

- 일원 분산 분석 42
- GLM 일변량 47, 50, 52

분산 분석

- 곡선 추정 79
- 모형 44
- 선형 회귀분석 73
- 선형모형에서 66
- 일원 분산 분석 39
- 평균 26
- GLM 일변량 43

분산팽창요인

- 선형 회귀분석 73

분포 맞춤

- 시뮬레이션에서 188

분포 추도

- 기술통계 9
- 빈도분석 6

분할계수

- 교차 분석표 17

분할표 15

- 불확실성 계수
- 교차 분석표 17

블록 거리

- 거리측도 59

비모수 검정

- 대응 2-표본 비모수 검정 151
- 대응 K-표본 비모수 검정 154
- 독립 2-표본 비모수 검정 149
- 독립 K-표본 비모수 검정 152
- 런 검정 147
- 모형 보기 139
- 일표본 Kolmogorov-Smirnov 검정 148

- 카이제곱 145

비율 통계 177

- 통계 177

빈도 5

- 교차 분석표 출력없음 7
- 차트 7
- 출력 순서 7
- 통계 6
- 형식 7

빈도표

- 데이터 탐색 12
- 빈도분석 5

[ 사 ]

사례 통제 연구

- 대응표본 T검정 35

사분위수

- 빈도분석 6

사용자 정의 모형

- GLM 44

사전

- Codebook 1

사후분석 다중 비교 40

삭제된 잔차

- 선형 회귀분석 71
- GLM 51

산점도

- 선형 회귀분석 71
- 시뮬레이션에서 194

산포 추도

- 기술통계 9
- 데이터 탐색 12
- 비율 통계 177
- 빈도분석 6

상관

- 교차 분석표 17
- 시뮬레이션에서 191
- 이변량 상관 55
- 편상관 57
- 0차 58

상관 행렬

- 순서 회귀분석 76
- 요인 분석 103, 104
- 관별 분석 98

상대적 위험도

- 교차 분석표 17

상자도표

- 데이터 탐색 12
- 변수 비교 12
- 시뮬레이션에서 194
- 요인 수준 비교 12

상호작용 항 45, 77, 78

선택변수

- 선형 회귀분석 70

선형 회귀분석 69

- 가중치 69
- 결측값 73
- 도표 71
- 명령 추가 가능 74
- 모형 내보내기 정보 71
- 변수 선택 방법 70, 73
- 블록 69

새 변수 저장 71

선택변수 70

잔차 71

통계 73

선형모형 61

결과 복제 64

계수 67

곡선 추정 80

관찰에 의한 예측 65

모형 선택 63

모형 옵션 64

모형 요약 64

모형 작성 요약 67

목표 61

분산 분석표 66

신뢰수준 62

양상블 64

예측변수 중요도 65

이상치 66

자동 데이터 준비 62, 65

선형모형 (계속)  
 잔차 65  
 정보 기준 64  
 조합 규칙 64  
 평균 추정 67  
 R-제곱 통계 64  
 선형별 연관  
 교차 분석표 17  
 선형성 검정  
 평균 26  
 성장모형  
 곡선 추정 80  
 소개  
 열별 요약 보고서 167  
 수렴  
 요인 분석 104, 105  
 K-Means 군집 분석 126  
 수정된 R 2  
 선형 회귀분석 73  
 수정된 R-제곱  
 선형모형에서 63  
 순서 회귀분석 75  
 링크 76  
 명령 추가 기능 78  
 옵션 76  
 위치 모형 77  
 척도 모형 78  
 통계 75  
 순위 상관  
 이변량 상관 55  
 스튜던트 t 검정 33  
 스튜던트화 잔차  
 선형 회귀분석 71  
 스트레스  
 다차원척도법 173  
 시각화  
 모형 군집화 114  
 시계열도표 분석  
 예측 80  
 예측할 케이스 80  
 시뮬레이션 181  
 결과 193, 194  
 꼬리 표본추출 191  
 누적 분포 함수 193  
 대상 및 입력 형식 표시 194  
 대상 분포의 백분위수 194  
 대화형 차트 198  
 모형 사양 185

시뮬레이션 (계속)  
 민감도 분석 191  
 방정식 편집기 186  
 분포 맞춤 188  
 분포 맞춤 결과 190  
 분포 맞춤 사용자 정의 190  
 분포를 새 데이터에 다시 맞춤 196  
 산점도 194  
 상자도표 194  
 새 입력 만들기 187  
 시뮬레이션 계획 만들기 182, 183  
 시뮬레이션 계획 실행 184, 195  
 시뮬레이션 계획 저장 195  
 시뮬레이션 작성기 185  
 시뮬레이션한 데이터 저장 195  
 입력 사이의 상관 191  
 중지 기준 191  
 지원하는 모형 185  
 차트 옵션 199  
 토네이도 차트 194  
 확률 밀도 함수 193  
 what-if 분석 191  
 시뮬레이션 작성기 185  
 신뢰구간  
 대응표본 T 검정 35  
 데이터 탐색 12  
 독립 표본 T 검정 34  
 선형 회귀분석 73  
 선형 회귀분석에 저장 71  
 일원 분산 분석 42  
 일표본 T 검정 36  
 GLM 46, 47, 50, 52  
 ROC 곡선 179  
 신뢰구간 요약  
 비모수 검정 139, 140, 141  
 신뢰성 분석 169  
 급내 상관계수 170  
 기술통계 170  
 명령 추가 기능 171  
 분산 분석표 170  
 예제 169  
 통계 169, 170  
 항목 간 상관 및 공분산 170  
 Hotelling의 T 2 170  
 Kuder-Richardson 20 170  
 Tukey의 가법성 검정 170  
 쌍대 비교  
 비모수 검정 143

## [ 아 ]

알파 계수  
 신뢰성 분석 169, 170  
 알파 요인 추출 104  
 앙상블  
 선형모형에서 64  
 에타  
 교차 분석표 17  
 평균 26  
 에타제곱  
 평균 26  
 GLM 일변량 47, 50, 52  
 역 모형  
 곡선 추정 80  
 연속형 필드 정보  
 비모수 검정 143  
 열 백분율  
 교차 분석표 18  
 열 비율 통계  
 교차 분석표 18  
 열별 요약 보고서 166  
 결측값 167  
 명령 추가 기능 168  
 소개 167  
 열 형식 164  
 쪽 매김 167  
 쪽 윤곽 165  
 쪽 제어 167  
 총계 167  
 총계 열 167  
 예측  
 곡선 추정 80  
 예측 구간  
 곡선 추정으로 저장 80  
 선형 회귀분석에 저장 71  
 예측값  
 곡선 추정으로 저장 80  
 선형 회귀분석에 저장 71  
 예측변수 중요도  
 선형모형 65  
 오차 요약  
 최근접 이웃 분석에서 96  
 완전 요인 모형  
 GLM 44  
 왜도  
 기술통계 9  
 데이터 탐색 12



- 왜도 (계속)
  - 빈도분석 6
  - 열별 요약 보고서 166
  - 요약 22
  - 평균 26
  - 행별 요약 보고서 164
  - OLAP 큐브 30
- 왜도의 표준 오차
  - 요약 22
  - 평균 26
  - OLAP 큐브 30
- 요약 21
  - 옵션 22
  - 통계 22
- 요인 분석 103
  - 개요 103
  - 결측값 106
  - 계수표시 형식 106
  - 기술통계 104
  - 로딩 도표 105
  - 명령 추가 기능 107
  - 수렴 104, 105
  - 예제 103
  - 요인 스코어 106
  - 추출 방법 104
  - 케이스 선택 104
  - 통계 103, 104
  - 회전 방법 105
- 요인 스코어 106
- 우도비 구간
  - 일표본 비모수 검정 130
- 우도비 카이제곱
  - 교차 분석표 17
  - 순서 회귀분석 76
- 원형 차트
  - 빈도분석 7
- 위치 모형
  - 순서 회귀분석 77
- 위험도
  - 교차 분석표 17
- 유사성 측도
  - 거리측도 60
  - 계층적 군집 분석 122
- 유클리드 거리
  - 거리측도 59
  - 최근접 이웃 분석에서 89
- 유클리드 거리 제곱값
  - 거리측도 59

- 이단계 군집 분석 111
  - 옵션 112
  - 외부 파일에 저장 114
  - 작업 파일에 저장 114
  - 통계 114
- 이미지 요인 추출 104
- 이변량 상관계수
  - 결측값 56
  - 명령 추가 기능 56
  - 상관계수 55
  - 옵션 56
  - 유의 수준 55
  - 통계 56
- 이상치
  - 데이터 탐색 12
  - 선형 회귀분석 71
  - 이단계 군집 분석 112
- 이퀴맥스 회전
  - 요인 분석 105
- 이항검정 146
  - 결측값 147
  - 명령 추가 기능 147
  - 옵션 147
  - 이분형 146
- 일표본 비모수 검정 130
  - 통계 147
- 일반화 최소제곱법
  - 요인 분석 104
- 일부 최소제곱 회귀분석 83
  - 모형 85
  - 변수 내보내기 85
- 일원 분산 분석 39
  - 결측값 42
  - 다중 비교 40
  - 다항 대비 40
  - 대비 40
  - 명령 추가 기능 42
  - 사후 검정 40
  - 옵션 42
  - 요인변수 39
  - 통계 42
- 일표본 비모수 검정 129
  - 런 검정 131
  - 이항검정 130
  - 카이제곱 검정 131
  - 콜모고로프-스미르노프 검정 131
  - 필드 129
- 일표본 Kolmogorov-Smirnov 검정 148

- 일표본 Kolmogorov-Smirnov 검정 (계속)
  - 검정 분포 148
  - 결측값 149
  - 명령 추가 기능 149
  - 옵션 149
  - 통계 149
- 일표본 T 검정 36
  - 결측값 36
  - 명령 추가 기능 36, 37
  - 신뢰구간 36
  - 옵션 36

## [ 자 ]

- 자동 데이터 준비
  - 선형모형에서 65
- 자동 분포 맞춤
  - 시뮬레이션에서 188
- 잔차
  - 곡선 추정으로 저장 80
  - 교차 분석표 18
  - 선형 회귀분석에 저장 71
- 잔차 도표
  - GLM 이변량 47, 50, 52
- 잘린 평균
  - 데이터 탐색 12
- 잡음 처리
  - 이단계 군집 분석 112
- 적합도
  - 순서 회귀분석 76
- 전진 선택법
  - 선형 회귀분석 70
  - 최근접 이웃 분석에서 90
- 전체 백분율
  - 교차 분석표 18
- 절대 동형 모형
  - 신뢰성 분석 169, 170
- 정규 확률 도표
  - 데이터 탐색 12
  - 선형 회귀분석 71
- 정규성 검정
  - 데이터 탐색 12
- 정보 기준
  - 선형모형에서 63
- 제공함 45
  - GLM 44
- 제목
  - OLAP 큐브 32

- 제어변수
  - 교차 분석표 16
- 제프리스 구간
  - 일표본 비모수 검정 130
- 조합 규칙
  - 선형모형에서 64
- 조화 평균
  - 요약 22
  - 평균 26
  - OLAP 큐브 30
- 주변 동질성 검정
  - 대응 2-표본 비모수 검정 151
  - 대응표본 비모수 검정 136
- 주변 평균 추정값
  - GLM 일변량 47, 50, 52
- 주성분 분석 103, 104
- 주축 요인 추출 104
- 줄기와 잎그림 도표
  - 데이터 탐색 12
- 중심 경향 측도
  - 데이터 탐색 12
  - 비율 통계 177
  - 빈도분석 6
- 중앙값
  - 데이터 탐색 12
  - 비율 통계 177
  - 빈도분석 6
  - 요약 22
  - 평균 26
  - OLAP 큐브 30
- 중앙값 검정
  - 독립 2-표본 비모수 검정 152
- 지수모형
  - 곡선 추정 80
- 직접 오블리민 회전
  - 요인 분석 105
- 집중 지수
  - 비율 통계 177
- 쪽 매김
  - 열별 요약 보고서 167
  - 행별 요약 보고서 165
- 쪽 제어
  - 열별 요약 보고서 167
  - 행별 요약 보고서 165

## [ 차 ]

- 차이 대비
  - GLM 46
- 차트
  - 케이스 레이블 79
  - ROC 곡선 179
- 참조 범주
  - GLM 46
- 처음
  - 요약 22
  - 평균 26
  - OLAP 큐브 30
- 척도
  - 다차원척도법 173
  - 신뢰성 분석 169
- 척도 모형
  - 순서 회귀분석 78
- 첨도
  - 기술통계 9
  - 데이터 탐색 12
  - 빈도분석 6
  - 열별 요약 보고서 166
  - 요약 22
  - 평균 26
  - 행별 요약 보고서 164
  - OLAP 큐브 30
- 첨도의 표준 오차
  - 요약 22
  - 평균 26
  - OLAP 큐브 30
- 초기 임계값
  - 이단계 군집 분석 112
- 총계
  - 열별 요약 보고서 167
- 총계 열
  - 보고서 167
- 최근접 이웃 분석 87
  - 결과 92
  - 모형 보기 92
  - 변수 저장 91
  - 옵션 92
  - 파티션 90
  - 필드선택 90
  - 항목 89
- 최대 분기
  - 이단계 군집 분석 112

- 최대값
  - 기술통계 9
  - 데이터 탐색 12
  - 보고서 열 비교 167
  - 비율 통계 177
  - 빈도분석 6
  - 요약 22
  - 평균 26
  - OLAP 큐브 30
- 최대우도
  - 요인 분석 104
- 최소값
  - 기술통계 9
  - 데이터 탐색 12
  - 보고서 열 비교 167
  - 비율 통계 177
  - 빈도분석 6
  - 요약 22
  - 평균 26
  - OLAP 큐브 30
- 최소유의차
  - 일원 분산 분석 40
  - GLM 49
- 최적 서브세트
  - 선형모형에서 63

## [ 카 ]

- 카이제곱 145
  - 결측값 146
  - 교차 분석표 17
  - 기대값 145
  - 기대범위 145
  - 독립성 17
  - 선형별 연관 17
  - 옵션 146
  - 우도비 17
  - 일표본 검정 145
  - 통계 146
  - Fisher의 정확 검정 17
  - Pearson 17
  - Yates의 연속성 보정 17
- 카이제곱 거리
  - 거리측도 59
- 카이제곱 검정
  - 일표본 비모수 검정 130, 131
  - 케이스 목록 21

케이스 수  
 요약 22  
 평균 26  
 OLAP 큐브 30  
 케이스별 진단 정보  
 선형 회귀분석 73  
 켄달의 일치계수(W)  
 대응표본 비모수 검정 136  
 코드북 1  
 콜모고로프-스미르노프 검정  
 일표본 비모수 검정 130, 131  
 콰티맥스 회전  
 요인 분석 105  
 크기 차이 측도  
 거리측도 59  
 클로퍼-피어슨 구간  
 일표본 비모수 검정 130

## [ 타 ]

타우-b  
 교차 분석표 17  
 타우-c  
 교차 분석표 17  
 토네이도 차트  
 시뮬레이션에서 194  
 트리 깊이  
 이단계 군집 분석 112

## [ 파 ]

파이  
 교차 분석표 17  
 파이제곱 거리측도  
 거리측도 59  
 판별 분석 97  
 결측값 100  
 공분산 행렬 100  
 그룹화 변수 97  
 기술통계 98  
 기준 99  
 단계적 방법 97  
 도표 100  
 독립변수 97  
 명령 추가 기능 101  
 모형 내보내기 정보 100  
 범위지정 98  
 분류분석 변수 저장 100

판별 분석 (계속)  
 사전 확률 100  
 예제 97  
 출력 옵션 99, 100  
 케이스 선택 98  
 통계 97, 98  
 판별 방법 99  
 함수의 계수 98  
 행렬 98  
 Mahalanobis의 거리 99  
 Rao의 V 99  
 Wilks의 람다 99  
 패턴 차이 측도  
 거리측도 59  
 패턴 행렬  
 요인 분석 103  
 편상관계수 57  
 결측값 58  
 명령 추가 기능 58  
 선형 회귀분석 73  
 옵션 58  
 통계 58  
 0차 상관 58  
 편차 대비  
 GLM 46  
 편회귀잔차 도표  
 선형 회귀분석 71  
 평균 25  
 기술통계 9  
 다중 보고서 열 167  
 데이터 탐색 12  
 비율 통계 177  
 빈도분석 6  
 열별 요약 보고서 166  
 옵션 26  
 요약 22  
 일원 분산 분석 42  
 통계 26  
 평균 26  
 하위 그룹 25, 29  
 행별 요약 보고서 164  
 OLAP 큐브 30  
 평균의 표준 오차  
 요약 22  
 평균 26  
 OLAP 큐브 30  
 평균-산포 도표  
 데이터 탐색 12

평균-산포 도표 (계속)  
 GLM 일변량 47, 50, 52  
 평행선 검정  
 순서 회귀분석 76  
 표준 오차  
 기술통계 9  
 데이터 탐색 12  
 빈도분석 6  
 GLM 47, 50, 51, 52  
 ROC 곡선 179  
 표준 편차  
 기술통계 9  
 데이터 탐색 12  
 비율 통계 177  
 빈도분석 6  
 열별 요약 보고서 166  
 요약 22  
 평균 26  
 행별 요약 보고서 164  
 GLM 일변량 47, 50, 52  
 OLAP 큐브 30  
 표준화  
 이단계 군집 분석 112  
 표준화 값  
 기술통계 9  
 표준화 잔차  
 선형 회귀분석 71  
 GLM 51  
 표준화하지 않은 잔차  
 GLM 51  
 프로파일 도표  
 GLM 47  
 피어  
 최근접 이웃 분석에서 94  
 필드선택  
 최근접 이웃 분석에서 95

## [ 하 ]

하위 그룹 평균 25, 29  
 합계  
 기술통계 9  
 빈도분석 6  
 요약 22  
 평균 26  
 OLAP 큐브 30  
 핫지스-레만 추정  
 대응표본 비모수 검정 136

항 설정 45, 77, 78  
 행 백분율  
   교차 분석표 18  
 행별 요약 보고서 163  
   결측값 165  
   구분 구간 164  
   구분 열 163  
   꼬리말 165  
   데이터 열 163  
   명령 추가 기능 168  
   열 형식 164  
   정렬 시퀀스 163  
   제목 165  
   제목변수 165  
   쪽 매김 165  
   쪽 윤곽 165  
   쪽 제어 164  
 허용 오차  
   선형 회귀분석 73  
 형식  
   보고서 열 164  
 확률 밀도 함수  
   시뮬레이션에서 193  
 회귀계수  
   선형 회귀분석 73  
 회귀분석  
   다중 회귀분석 69  
   도표 71  
   선형 회귀분석 69  
 효과-크기 추정값  
   GLM 일변량 47, 50, 52  
 후진제거법  
   선형 회귀분석 70  
 혼련 표본  
   최근접 이웃 분석에서 90  
 히스토그램  
   데이터 탐색 12  
   빈도분석 7  
   선형 회귀분석 71

## [ 숫자 ]

0차 상관  
   편상관 58  
 2차 모형  
   곡선 추정 80  
 2-표본 T 검정  
   독립 표본 T 검정 33

3차모형  
   곡선 추정 80  
 4분위 맵  
   최근접 이웃 분석에서 95

## A

AAD(평균 절대 편차)  
   비율 통계 177  
 Akaike 정보 기준  
   선형모형에서 63  
 Anderson-Rubin 요인 스코어 106  
 Andrews의 웨이브 추정량  
   데이터 탐색 12

## B

Bartlett 요인 스코어 106  
 Bartlett의 단위행렬 검정  
   요인 분석 104  
 Bonferroni  
   일원 분산 분석 40  
   GLM 49  
 Box의 M 검정  
   판별 분석 98  
 Brown-Forsythe 통계  
   일원 분산 분석 42

## C

Chebychev 거리  
   거리측도 59  
 Cochran의 통계  
   교차 분석표 17  
 Cochran의 Q  
   대응 K-표본 비모수 검정 154  
 Cochran의 Q 검정  
   대응표본 비모수 검정 136, 138  
 Codebook  
   결과 2  
   통계 4  
 COD(산포계수)  
   비율 통계 177  
 Cohen의 kappa  
   교차 분석표 17  
 Cook의 거리  
   선형 회귀분석 71  
   GLM 51

COV(변동 계수)  
   비율 통계 177  
 Cox 및 Snell R2  
   순서 회귀분석 76  
 Cramér의 V  
   교차 분석표 17  
 Cronbach의 알파  
   신뢰성 분석 169, 170

## D

d  
   교차 분석표 17  
 DfBeta  
   선형 회귀분석 71  
 DFFIT  
   선형 회귀분석 71  
 Duncan의 다중 범위 검정  
   일원 분산 분석 40  
   GLM 49  
 Dunnett의 C  
   일원 분산 분석 40  
   GLM 49  
 Dunnett의 T 검정  
   일원 분산 분석 40  
   GLM 49  
 Dunnett의 T3  
   일원 분산 분석 40  
   GLM 49  
 Durbin-Watson 통계  
   선형 회귀분석 73

## F

F 통계  
   선형모형에서 63  
 Fisher의 정확 검정  
   교차 분석표 17  
 Fisher의 LSD  
   GLM 49  
 Friedman 검정  
   대응 K-표본 비모수 검정 154  
   대응표본 비모수 검정 136

## G

Gabriel의 쌍대 비교 검정  
   일원 분산 분석 40

Gabriel의 쌍대 비교 검정 (계속)  
GLM 49

Games-Howell의 대응 비교 검정  
일원 분산 분석 40  
GLM 49

GLM  
모형 44  
변수 저장 51  
사후 검정 49  
제공함 44  
프로파일 도표 47  
행렬 저장 51

GLM 일변량 43, 48, 51, 53  
대비 46  
옵션 47, 50, 52  
주변 평균 추정값 47, 50, 52  
진단 정보 47, 50, 52  
출력 47, 50, 52

Goodman과 Kruskal의 감마  
교차 분석표 17

Goodman과 Kruskal의 람다  
교차 분석표 17

Goodman과 Kruskal의 타우  
교차 분석표 17

Guttman 모형  
신뢰성 분석 169, 170

## H

Hampel의 redescending M 추정량  
데이터 탐색 12

Helmert 대비  
GLM 46

Hochberg의 GT2  
일원 분산 분석 40  
GLM 49

Hotelling의 T 2  
신뢰성 분석 169, 170

Huber의 M-추정량  
데이터 탐색 12

## I

ICC(급내 상관계수)  
신뢰성 분석 170  
ICC. 급내 상관계수 참조 170

## K

k 및 필드선택  
최근접 이웃 분석에서 95

k 선택  
최근접 이웃 분석에서 95

kappa  
교차 분석표 17

Kendall의 타우-b  
교차 분석표 17  
이변량 상관 55

Kendall의 타우-c 17  
교차 분석표 17

Kendall의 W  
대응 K-표본 비모수 검정 154

Kolmogorov-Smirnov의 Z  
독립 2-표본 비모수 검정 150  
일표본 Kolmogorov-Smirnov 검정 148

KR20  
신뢰성 분석 170

Kruskal의 타우  
교차 분석표 17

Kruskal-Wallis의 H  
독립 2-표본 비모수 검정 152

Kuder-Richardson 20(KR20)  
신뢰성 분석 170

K-Means 군집 분석  
개요 125  
결측값 127

군집 거리 126  
군집 정보 저장 126  
명령 추가 기능 127

반복 126  
방법 125  
소속군집 126  
수렴 기준 126

예제 125  
통계 125, 127  
효율 126

## L

Lance와 Williams 상이성 측도 59  
거리측도 59

Levene 검정  
데이터 탐색 12  
일원 분산 분석 42  
GLM 일변량 47, 50, 52

Lilliefors 검정  
데이터 탐색 12

## M

Mahalanobis의 거리  
선형 회귀분석 71  
판별 분석 99

Manhattan 거리  
최근접 이웃 분석에서 89

Mann-Whitney U  
독립 2-표본 비모수 검정 150

Mantel-Haenszel 통계  
교차 분석표 17

McFadden R2  
순서 회귀분석 76

McNemar 검정  
교차 분석표 17  
대응 2-표본 비모수 검정 151  
대응표본 비모수 검정 136, 137

Minkowski 거리  
거리측도 59

Monte Carlo 시뮬레이션 181

Moses의 극단값반동 검정  
독립 2-표본 비모수 검정 150

M-추정량  
데이터 탐색 12

## N

Nagelkerke R2  
순서 회귀분석 76

Newman-Keuls  
GLM 49

## O

OLAP 큐브 29  
제목 32  
통계 30

## P

Pearson 상관계수  
교차 분석표 17  
이변량 상관 55  
Pearson 잔차  
순서 회귀분석 76

Pearson 카이제곱  
교차 분석표 17  
순서 회귀분석 76  
PLUM  
순서 회귀분석 75  
PRD(가격 관련 차)  
비율 통계 177

## R

R 2  
선형 회귀분석 73  
평균 26  
R 2 변화량 73  
r 상관계수  
교차 분석표 17  
이변량 상관 55  
R 통계  
선형 회귀분석 73  
평균 26

Rao의 V  
판별 분석 99

rho  
교차 분석표 17  
이변량 상관 55

ROC 곡선 179  
통계 및 도표 179

Ryan-Einot-Gabriel-Welsch 다중 범위  
일원 분산 분석 40  
GLM 49

Ryan-Einot-Gabriel-Welsch 다중 F  
일원 분산 분석 40  
GLM 49

R-E-G-W의 F  
일원 분산 분석 40  
GLM 49

R-E-G-W의 Q  
일원 분산 분석 40  
GLM 49

R-제곱  
선형모형에서 64

## S

Scheffé 검정  
일원 분산 분석 40  
GLM 49

Shapiro-Wilk의 검정  
데이터 탐색 12  
Sidak의 T 검정  
일원 분산 분석 40  
GLM 49  
Somers의 d  
교차 분석표 17  
Spearman 상관계수  
교차 분석표 17  
이변량 상관 55  
Spearman-Brown 신뢰도  
신뢰성 분석 170  
Student-Newman-Keuls  
일원 분산 분석 40  
GLM 49  
S-모형  
곡선 추정 80  
S-스트레스  
다차원척도법 173

## T

T 검정  
대응표본 T 검정 35  
독립 표본 T 검정 33  
일표본 T 검정 36  
GLM 일변량 47, 50, 52

Tamhane의 T2  
일원 분산 분석 40  
GLM 49

Tukey의 가법성 검정  
신뢰성 분석 169, 170

Tukey의 이중가중 추정량  
데이터 탐색 12

Tukey의 정직유의차  
일원 분산 분석 40  
GLM 49

Tukey의 B 검정  
일원 분산 분석 40  
GLM 49

## V

V  
교차 분석표 17

## W

Wald-Wolfowitz run 검정  
독립 2-표본 비모수 검정 150

Waller-Duncan T 검정  
일원 분산 분석 40  
GLM 49

Welch 통계  
일원 분산 분석 42

what-if 분석  
시뮬레이션에서 191

Wilcoxon 부호순위 검정  
대응 2-표본 비모수 검정 151  
대응표본 비모수 검정 136  
일표본 비모수 검정 130

Wilks의 람다  
판별 분석 99

## Y

Yates의 연속성 보정  
교차 분석표 17

## Z

z 스코어  
기술통계 9  
변수로 저장 9







