

***IBM SPSS Advanced
Statistics 22***

IBM

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 111 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 Advanced Statistics 소개	1	선형 혼합 모형 통계	32
제 2 장 GLM 다변량 분석	3	선형 혼합 모형 EM 평균.	33
GLM 다변량 모형	4	선형 혼합 모형 저장	33
항 설정	5	MIXED 명령 추가 기능	34
제공합	5	제 6 장 일반화 선형 모형	35
GLM 다변량 대비	6	일반화 선형 모형 반응.	38
대비 유형	6	일반화 선형 모형 참조 범주.	38
GLM 다변량 프로파일 도표	7	일반화 선형 모형 예측변수	39
GLM 다변량 사후분석 비교	7	일반화 선형 모형 옵션.	39
GLM 저장	8	일반화 선형 모형 모형.	39
GLM 다변량 옵션	9	일반화 선형 모형 추정.	40
GLM 명령 추가 기능.	10	일반화 선형 모형 초기값	41
제 3 장 GLM 반복 측도	13	일반화 선형 모형 통계.	42
GLM 반복 측도 요인 정의	15	일반화 선형 모형 EM 평균.	43
GLM 반복 측도 모형.	16	일반화 선형 모형 저장.	44
항 설정.	16	일반화 선형 모형 내보내기	45
제공합	16	GENLIN 명령 추가 기능.	46
GLM 반복 측도 대비.	17	제 7 장 일반화 추정 방정식.	47
대비 유형.	18	일반화된 추정 방정식 모형 유형	49
GLM 반복 측도 프로파일 도표.	18	일반화된 추정 방정식 반응	51
GLM 반복 측도 사후분석 비교.	19	일반화된 추정 방정식 참조 범주	51
GLM 반복 측도 저장.	20	일반화된 추정 방정식 예측변수.	52
GLM 반복 측도: 옵션.	21	일반화된 추정 방정식 옵션	52
GLM 명령 추가 기능.	22	일반화된 추정 방정식 모형	53
제 4 장 분산 성분추정 분석.	23	일반화된 추정 방정식 추정	54
분산 성분추정: 모형	24	일반화된 추정 방정식 초기값	55
항 설정.	24	일반화된 추정 방정식 통계	55
분산 성분추정: 옵션	24	일반화된 추정 방정식 EM 평균	56
제공합(분산 성분추정).	25	일반화된 추정 방정식 저장	57
분산 성분추정: 새 파일로 저장.	26	일반화된 추정 방정식 내보내기.	58
VARCOMP 명령 추가 기능	26	GENLIN 명령 추가 기능.	59
제 5 장 선형 혼합 모형	27	제 8 장 일반화 선형 혼합 모형	61
선형 혼합 모형의 개체/반복변수 선택.	28	일반화 선형 혼합 모형 구하기	62
선형 혼합 모형 고정 효과	29	대상.	62
중첩되지 않은 항 설정.	29	고정 효과.	65
중첩 항 설정.	30	사용자 정의 항 추가	65
제공합	30	랜덤 효과.	66
선형 혼합 모형 랜덤 효과	31	랜덤 효과 블록	66
선형 혼합 모형: 추정	32	가중치 및 오프셋	68
		일반 빌드 옵션	68

추정	69	생명표 옵션	90
평균 추정	69	SURVIVAL 명령 추가 기능	91
저장	70	제 13 장 Kaplan-Meier 생존 분석	93
모형 보기	71	Kaplan-Meier 상태변수에 대한 이벤트 정의	94
모형 요약	71	Kaplan-Meier 요인 수준 비교	94
데이터 구조	71	Kaplan-Meier 새 변수로 저장	94
관찰에 의한 예측	71	Kaplan-Meier 옵션	95
분류	72	KM 명령 추가 기능	95
고정 효과	72	제 14 장 Cox 회귀분석 분석	97
고정 계수	72	Cox 회귀분석 범주형 변수 정의	98
랜덤 효과 공분산	73	Cox 회귀분석 도표	98
공분산 모수	73	Cox 회귀분석 새 변수로 저장	99
평균 추정: 유의한 효과	74	Cox 회귀분석 옵션	99
평균 추정: 사용자 정의 효과	74	Cox 회귀분석 상태변수에 대한 이벤트 정의	100
제 9 장 모형선택 로그선형 분석	77	COXREG 명령 추가 기능	100
로그선형분석 범가지정	78	제 15 장 시간-종속 공변량 계산	101
로그선형분석 모형	78	시간-종속 공변량 계산	101
항 설정	78	시간-종속 공변량이 있는 Cox 회귀분석의 추가 기능	102
모형선택 로그선형 분석 옵션	78	제 16 장 범주형 변수 코딩 체계	103
HILOGLINEAR 명령 추가 기능	79	편차	103
제 10 장 일반 로그선형분석	81	단순	103
일반 로그선형분석 모형	82	Helmert	104
항 설정	82	차이	104
일반 로그선형분석 옵션	83	다항	104
일반 로그선형분석 저장	83	반복	105
GENLOG 명령 추가 기능	83	특수	106
제 11 장 로짓 로그선형 분석	85	지표	106
로짓 로그선형 분석 모형	86	제 17 장 공분산 구조	107
항 설정	87	주의사항	111
로짓 로그선형 분석 옵션	87	상표	113
로짓 로그선형 분석 저장	87	색인	115
GENLOG 명령 추가 기능	88		
제 12 장 생명테이블	89		
생명표 상태변수에 대한 이벤트 정의	90		
생명표 범가지정	90		

제 1 장 Advanced Statistics 소개

Advanced Statistics 옵션은 Statistics Base 옵션보다 고급 모형 옵션을 제공하는 프로시저를 제공합니다.

- GLM 다변량은 GLM 일변량에서 제공하는 일반 선형 모형을 확장하여 여러 종속변수를 사용할 수 있도록 합니다. 추가 확장인 GLM 반복 측정을 사용하면 여러 종속변수를 반복 측정할 수 있습니다.
- 분산 성분추정 분석은 종속변수의 변동을 고정 및 무작위 성분으로 분해하는 특정 도구입니다.
- 선형 혼합 모형은 일반 선형 모형을 확장하여 데이터가 상호 관련되고 일정하지 않은 변동을 나타낼 수 있도록 합니다. 따라서 선형 혼합 모형은 데이터의 평균 뿐만 아니라 데이터의 분산 및 공분산 모형을 만드는 데 유연성을 제공합니다.
- 일반화 선형 모형(GZLM)은 오차항의 정규성 가정을 줄여 종속변수가 변환이나 연결 함수를 통해 예측변수와 선형적으로 관련되도록 해야 합니다. 일반화된 추정 방정식(GEE)은 반복 측정을 허용하도록 GZLM을 확장합니다.
- 일반 로그선형분석을 사용하면 교차 분류 개수 데이터의 모형을 적합시킬 수 있으며 모형선택 로그선형 분석은 모형을 선택할 수 있도록 합니다.
- 로짓 로그선형 분석을 사용하면 범주형 종속변수와 하나 이상의 범주형 예측변수 사이의 관계를 분석하기 위한 로그선형 모형을 적합시킬 수 있습니다.
- 생존 분석은 요인변수의 수준별로 시간-이벤트 변수의 분포를 검토하는 경우 생명표, 요인변수의 수준별로 시간-이벤트 변수의 분포를 검토하거나 층화변수의 수준별로 개별 분석을 생성하는 경우 Kaplan-Meier 생존 분석, 주어진 공변량의 값을 기준으로 하여 지정된 이벤트에 대한 시간을 모델링하는 경우 Cox 회귀분석을 통해 제공됩니다.

제 2 장 GLM 다변량 분석

GLM 다변량 프로시저는 하나 이상의 요인변수나 공변량 별로 다중 종속변수에 대한 회귀분석과 분산 분석을 제공합니다. 요인변수는 모집단을 여러 그룹으로 나눕니다. 이러한 일반선형모형 프로시저를 사용하여 다양한 그룹의 종속변수 결합 분포의 평균에 미치는 요인변수의 효과에 관한 귀무가설을 검정할 수 있습니다. 개별 요인들의 효과뿐만 아니라 요인간 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. 다변량 모형에서 모형내의 효과로 인한 제곱합과 오차 제곱합은 일변량 분석에서 구한 스칼라 식보다는 행렬식으로 나타냅니다. 이들 행렬을 SSCP(제곱합과 교차곱) 행렬이라 합니다. 둘 이상의 종속변수가 지정되면 각 종속변수에 대한 일변량 분산 분석이 제공될 뿐 아니라 근사 F 통계와 함께 Pillai의 트레이스, Wilks의 람다, Hotelling의 트레이스, Roy의 최대근 기준을 사용하는 다변량 분산 분석이 제공됩니다. GLM 다변량은 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 추정값을 통해 모형의 셀에 대한 예측평균값의 추정값을 구할 수 있으며 이들 평균의 프로파일 도표(상호작용 도표)를 사용하여 일부 관계를 보다 쉽게 나타낼 수 있습니다. 사후분석 다중 비교 검정은 각 종속변수에 대해 개별적으로 수행됩니다.

잔차, 예측값, Cook의 거리, 레버리지 값은 가정을 진단하기 위해 데이터 파일에 새로운 변수로 저장할 수 있습니다. 또한 잔차의 교차곱과 제곱합의 제곱 행렬인 잔차 SSCP 행렬, 잔차 SSCP 행렬을 잔차의 자유도로 나눈 잔차 공분산 행렬, 잔차 공분산 행렬의 표준화 양식인 잔차 공분산 행렬 등도 사용할 수 있습니다.

WLS 가중치를 사용하면 서로 다른 측정 정밀도를 보완하기 위해 여러 관측값에 가중치 최소제곱(WLS)에 대한 다른 가중치를 부여하는데 사용되는 변수를 지정할 수 있습니다.

예제. 한 플라스틱 제조업체가 플라스틱 필름의 세 가지 특성인 찢김 저항성, 광택, 불투명도를 측정한다고 두 가지 압출 비율과 두 가지 다른 양의 첨가물을 사용하여 각 압출 비율과 첨가물 양을 조합하고 세 가지 특성을 측정합니다. 이 경우 제조업체는 압출 비율과 첨가물 양이 각기 중요한 결과를 나타내지만 두 가지 요인의 상호작용은 유의하지 않음을 알게 됩니다.

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. 사후분석 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 전

체 셀의 모든 종속변수에 대한 관측평균, 표준 편차 및 개수, 분산 동질성에 대한 Levene 검정, 종속변수의 공분산 행렬 동질성에 대한 Box의 M 검정, Bartlett의 단위행렬 검정을 선택할 수 있습니다.

도표. 평균-산포 도표, 잔차 및 프로파일(상호작용)을 선택할 수 있습니다.

GLM 다변량 데이터 고려 사항

데이터. 종속변수는 양적변수여야 합니다. 요인은 범주형이며 숫자값 또는 문자열 값을 포함할 수 있습니다. 공변량은 종속변수와 관련된 양적변수입니다.

가정. 종속변수의 경우 데이터는 다변량 정규 모집단에서 무작위 표본 추출된 벡터의 확률표본입니다. 즉, 모집단에서는 모든 셀에 대한 분산-공분산 행렬이 동일합니다. 분산 분석은 데이터가 대칭적인 경우에도 정규성에서 이탈해도 영향을 받지 않습니다. 가정을 확인하려면 분산의 동질성 검정(Box의 M 포함)과 평균-산포 도표를 사용할 수 있습니다. 또한 잔차와 잔차 도표를 검토할 수 있습니다.

관련 프로시저. 분산 분석 수행 이전에 데이터 탐색 프로시저를 사용하여 데이터를 검토합니다. 단일 종속변수의 경우 GLM 일변량을 사용합니다. 몇 가지 경우에 각 개체에 대해 동일한 종속변수를 측도하려면 GLM 반복 측도를 사용합니다.

GLM 다변량 분산 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 일반 선형 모형 > 다변량...

2. 두 개 이상의 종속변수를 선택합니다.

선택적으로 모수요인, 공변량, WLS 가중치를 지정할 수 있습니다.

GLM 다변량 모형

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.

제공함. 제공함을 계산하는 방법입니다. 결측 셀이 없는 균형 모형이나 비균형 모형의 경우 제 III 유형 제공함 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 제공합을 계산할 때 이 효과를 포함하지 않는 다른 효과에 맞게 수정되며 이 효과를 포함하는 효과(있는 경우)에 직교하는 제공합 방법을 사용합니다. 제 III 유형 제공합은 일반 추정 형식이 일정하게 유지되는 한 셀 빈도에 관해 변하지 않는 이점이 있습니다. 따라서 이 제공합 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 제공합 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 제공합 방법은 일반적으로 다음 경우에 사용됩니다.

- 제 I 유형과 제 II 유형으로 나열된 모든 모형
- 빈 셀이 있는 모든 균형 모형이나 비균형 모형

GLM 다변량 대비

대비는 효과 수준이 서로 큰 차이가 있는지 여부를 검정하는 데 사용됩니다. 모형의 각 요인에 대한 대비를 지정할 수 있습니다. 대비는 모수의 선형 조합을 나타냅니다.

가설 검정은 귀무가설 $LBM = 0$ 을 기준으로 하며 여기에서 L 은 대비계수(L) 행렬이고 M 은 종속변수의 수와 동일한 차원을 가지는 항등 행렬이며 B 는 모수 벡터입니다. 대비가 지정되면 요인에 해당하는 여러 열이 대비에 일치하는 L 행렬이 작성됩니다. 나머지 열은 L 행렬 추정이 가능하도록 수정됩니다.

F 통계와 모든 종속변수에 걸친 대비 차이에 대한 Student의 t 분포에 따른 Bonferroni 유형 동시 신뢰구간을 사용하는 일변량 검정이 제공될 뿐 아니라 Pillai의 트레이스, Wilks의 람다, Hotelling의 트레이스, Roy의 최대근 기준을 사용하는 다변량 검정이 제공됩니다.

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 다변량 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. 도표에 대한 모든 요인을 사용할 수 있습니다. 각 종속변수에 대한 프로파일 도표가 만들어집니다.

한 요인에 대한 프로파일 도표는 추정된 주변 평균이 여러 수준에서 증가하는지 또는 감소하는지 여부를 보여줍니다. 둘 이상의 요인의 경우 평행선은 두 요인 간 상호작용이 없음을 나타내는데 이것은 한 요인의 수준만 조사할 수 있음을 의미합니다. 평행하지 않는 선들은 상호작용을 나타냅니다.

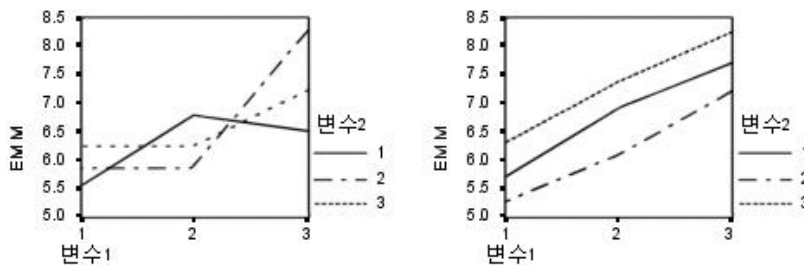


그림 1. 평행하지 않은 도표(왼쪽) 및 평행한 도표(오른쪽)

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

GLM 다변량 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. 사후 검정은 각 종속변수에 대해 개별적으로 수행됩니다.

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. **Tukey**의 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 GT2는 Tukey의 정직유의차 검정과 유사하나 스튜던트화 최대 계수가 사용됩니다. 대개 Tukey의 검정이 더 효과적입니다. **Gabriel의 쌍대 비교 검정** 또한 스튜던트화 최대 계수를 사용하며 셀 크기가 동일하지 않을 경우 일반적으로 Hochberg의 GT2가 더 효과적인 방법입니다. Gabriel의 검정은 셀 크기가 상당히 다양하게 분포할 경우 광범위하게 사용될 수 있습니다.

Dunnett의 대응별 다중 비교 **T** 검정은 단일 통제 평균에 대해 하나의 처리 세트를 비교합니다. 마지막 범주가 기본 제어 범주가 됩니다. 처음 범주를 선택할 수도 있습니다. 또한 양쪽검정이나 한쪽검정을 선택할 수 있습니다. 한 요인 수준(제어 범주 제외)에서의 평균이 제어 범주에서의 평균과 동일하지 않은지 검정하려면 양쪽검정을 사용합니다. 한 요인 수준에서의 평균이 제어 범주에서의 평균보다 작은지를 검정하려면 **< 통제**를 선택합니다. 마찬가지로 한 요인 수준의 평균이 제어 범주에서의 평균보다 큰지를 검정하려면 **> 통제**를 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 **Tamhane**의 **T2**(**t** 검정을 기준으로 하는 보존성 쌍대 비교 검정), **Dunnett**의 **T3**(스튜던트화 최대 계수를 기준으로 하는 쌍대 비교 검정), **Games-Howell** 쌍대 비교 검정(경우에 따라 광범위하게 사용됨) 또는 **Dunnett**의 **C**(스튜던트화 범위를 기준으로 하는 쌍대 비교 검정)를 사용합니다.

Duncan의 다중 범위 검정, Student-Newman-Keuls(**S-N-K**), **Tukey**의 **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**, **Tamhane**의 **T2** 및 **T3**, **Dunnett**의 **C**, **Dunnett**의 **T3**에 제공됩니다. 범위 검정을 위한 동일 서브세트는 **S-N-K**, **Tukey**의 **B**, **Duncan**, **R-E-G-W F**, **R-E-G-W Q** 및 **Waller**에 제공됩니다. **Tukey**의 정직유의차 검정, **Hochberg**의 **GT2**, **Gabriel**의 검정, **Scheffé**의 검정 등은 다중 비교 검정과 범위 검정입니다.

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 다변량 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. 상호작용은 사용자 정의 모형을 지정한 경우에만 사용할 수 있습니다.

- **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. 가설과 오차 **SSCP** 행렬을 출력할 수 있으며 잔차 공분산 행렬에 대한 Bartlett의 단위행렬 검정과 더불어 잔차 **SSCP** 행렬을 출력할 수 있습니다.

동질성 검정은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서의 각 종속변수에 대해 분산의 동질성에 대한 Levene의 검정을 생성합니다. 동질성 검정에는 개체 간 요인 수준의 모든 조합에서의 종속변수들의 공분산 행렬에 대한 Box의 *M* 동질성 검정도 포함됩니다. 평균-산포 도표와 잔차 도표 옵션은 데이터에 대한 가정을 진단하는 데 유용합니다. 이 항목은 요인이 없을 경우에는 사용할 수 없습니다. 각 종속변수에 대한 관측-예측-표준화 잔차 도표를 작성하려면 잔차 도표를 선택합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인하려면 적합결여 검정을 선택합니다. 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.

GLM 명령 추가 기능

이들 기능은 일변량, 다변량, 반복 측도 분석에 적용될 수 있습니다. 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 또는 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 3 장 GLM 반복 측도

GLM 반복 측도 프로시저는 각 개체나 케이스에 대해 동일한 측정을 여러 번 반복할 경우에 분산 분석을 제공합니다. 개체 간 요인이 지정되면 모집단을 여러 그룹으로 분리합니다. 이러한 일반선형모형 프로시저를 사용하면 개체 간 요인과 개체-내 요인 모두의 영향에 대한 귀무가설을 검정할 수 있습니다. 개별 요인들의 효과 뿐만 아니라 요인간 상호작용을 조사할 수 있습니다. 또한 상수 공변량의 영향과 개체 간 요인들과의 공변량 상호작용도 검정할 수 있습니다.

이중 다변량 반복 측도 디자인에서 종속변수는 개체 간 요인이 서로 다른 수준에 있는 둘 이상의 변수 측정값을 나타냅니다. 예를 들어, 각 개체에 대해 3회에 걸쳐 맥박과 호흡 수를 모두 측정할 수 있습니다.

GLM 반복 측도 프로시저는 반복 측도 데이터에 일변량 분석과 다변량 분석을 모두 제공합니다. 균형 모형과 비균형 모형을 모두 검정할 수 있습니다. 모형의 각 셀이 같은 수의 케이스를 포함하는 경우에는 디자인을 수행하는 것이 적절합니다. 다변량 모형에서 모형내의 효과로 인한 제곱합과 오차 제곱합은 일변량 분석에서 구한 스칼라 식보다는 행렬식으로 나타냅니다. 이들 행렬을 SSCP(제곱합과 교차곱) 행렬이라 합니다. GLM 반복 측도는 가설 검정과 더불어 모수 추정값을 생성합니다.

일반적으로 사용되는 사전 대비는 개체 간 요인들에 대한 가설 검정을 수행하는 데 사용할 수 있습니다. 또한 전체 F 검정에서 유의수준이 표시되면 사후 검정을 사용하여 특정 평균들 간의 차이를 평가할 수 있습니다. 주변 평균 추정값을 통해 모형의 셀에 대한 예측평균값의 추정값을 구할 수 있으며 이들 평균의 프로파일 도표(상호작용 도표)를 사용하여 일부 관계를 보다 쉽게 나타낼 수 있습니다.

잔차, 예측값, Cook의 거리, 레버리지 값은 가정을 진단하기 위해 데이터 파일에 새로운 변수로 저장할 수 있습니다. 또한 잔차의 교차곱과 제곱합의 제곱 행렬인 잔차 SSCP 행렬, 잔차 SSCP 행렬을 잔차의 자유도로 나눈 잔차 공분산 행렬, 잔차 공분산 행렬의 표준화 양식인 잔차 공분산 행렬 등도 사용할 수 있습니다.

WLS 가중치를 사용하면 서로 다른 측정 정밀도를 보완하기 위해 여러 관측값에 가중치 최소제곱(WLS)에 대한 다른 가중치를 부여하는데 사용되는 변수를 지정할 수 있습니다.

예제. 12명의 학생을 근심 등급 검정에서 얻은 스코어를 기초로 높은 근심 그룹이나 낮은 근심 그룹으로 할당 하였습니다. 여기서 근심 등급은 개체를 그룹으로 나누는 요인이 되므로 개체 간 요인이라 합니다. 학생들은 학습 과제에 대해 4회의 시험을 실시하며 각 시험에서 틀린 문제의 수를 코딩합니다. 이때 각 시험에서 틀린 문제의 수는 별도의 변수로 코딩되고 개체-내 요인(시험)은 4회의 시험에 대해 네 가지 수준으로 정의됩니다. 시행-근심 상호작용이 유의하지 않은 반면 시행 영향은 유의한 것으로 나타났습니다.

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. 사후분석 범위 검정 및 다중 비교에 대해(오브젝트 간 요인에 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(단

측 및 양측), Tamhane의 T2, Dunnett의 T3, Games-Howell, Dunnett의 C를 선택할 수 있습니다. 기술통계량에 대해 모든 셀의 모든 종속변수에 대한 관측평균, 표준 편차, 개수를 나타낼 수 있습니다 분산 동질성에 대한 Levene 검정을 선택할 수 있으며 Box의 M과 Mauchly의 구형성 검정을 선택할 수 있습니다.

도표. 평균-산포 도표, 잔차 및 프로파일(상호작용)을 선택할 수 있습니다.

GLM 반복 측도 데이터 고려 사항

데이터. 종속변수는 양적변수여야 합니다. 개체 간 요인을 통해 표본은 별도의 하위 그룹(예: 남자와 여자)으로 나누어집니다. 이러한 요인은 범주형이며 숫자값 또는 문자열 값을 포함할 수 있습니다. 개체-내 요인들은 반복 측도 요인 정의 대화 상자에서 정의됩니다. 공변량은 종속변수와 관련된 양적변수입니다. 반복 측도 분석의 경우 공변량은 개체-내 변수의 각 수준에서 일정하게 유지됩니다.

데이터 파일에는 개체에 관한 각 측정 그룹에 해당하는 변수 세트가 포함되어야 합니다. 이 세트는 그룹 내에서 각 측정을 반복할 때마다 각기 하나의 변수를 갖게 됩니다. 반복 횟수와 수준의 수가 같으면 그룹에 대한 개체-내 요인이 정의됩니다. 예를 들어, 몸무게 측정은 각기 다른 날에 실시할 수 있습니다. 동일한 특성을 5일에 걸쳐 측정한 경우 개체-내 요인은 날짜로 지정될 수 있으며 이 날짜는 5개의 수준을 포함합니다.

다중 개체-내 요인들의 경우 각 개체에 대한 측정 횟수는 각 요인 수준의 수의 곱과 같습니다. 예를 들어, 4일 동안 매일 3회 측정을 실시했다면 총 측정 횟수는 각 개체에 대해 12회가 됩니다. 개체-내 요인은 날짜(4)와 횟수(3)로 지정될 수 있습니다.

가정. 반복 측도 분석은 일변량과 다변량의 두 가지 방법으로 접근할 수 있습니다.

일변량 접근법(분할계획법(split-plot) 또는 혼합 모형 접근법(mixed-model approach)이라고도 함)에서 종속변수는 개체-내 요인 수준에 대한 반응입니다. 한 개체에 대한 측정값은 다변량 정규 분포에서 추출된 표본이어야 하며 분산-공분산 행렬은 개체 간 효과들에 의해 형성된 여러 셀에서 동일합니다. 이 경우 종속변수의 분산-공분산 행렬에 관해 확실한 가정이 형성됩니다. 분산-공분산 행렬식이 순환될 경우 일변량 접근법에 사용된 F 통계의 유효성이 보장됩니다(Huynh와 Mandeville, 1979).

이 가정을 검정하기 위해 Mauchly의 구형성 검정을 사용할 수 있는데 이 방법은 직교정규화로 변환된 종속변수의 분산-공분산 행렬에 대해 구형성 검정을 수행합니다. 반복 측도 분석에 대해 Mauchly의 검정이 자동으로 출력됩니다. 표본 결과가 작은 경우에는 이 검정 방법이 그다지 효과적이지 않습니다. 표본 결과가 큰 경우에는 발생한 이탈이 결과에 미치는 영향이 작은 경우에도 검정이 유의할 수 있습니다. 검정의 유의수준이 크면 구형성 가설을 가정할 수 있습니다. 그러나 유의수준이 작고 구형성 가정에 위배되는 것으로 나타나면 일변량 F 통계를 유효화하기 위해 분자와 분모 자유도에 대한 조정을 수행할 수 있습니다. 이 과정에서 엡실론이라는 세 개의 조정 추정값을 GLM 반복 측도 프로시저에 사용할 수 있습니다. 분자와 분모 자유도에 이 엡실론을 곱하고 새로운 자유도를 사용하여 F -비의 유의수준을 계산해야 합니다.

다변량 접근법은 개체에 대한 측정값을 다변량 정규 분포에서 추출한 표본으로 간주하며 분산-공분산 행렬들은 개체 간 효과로 형성된 여러 셀에서 동일한 값을 갖습니다. 여러 셀에서 분산-공분산 행렬이 동일한지 여부를 검정하는 데 Box의 M 검정을 사용할 수 있습니다.

관련 프로시저. 분산 분석 수행 이전에 데이터 탐색 프로시저를 사용하여 데이터를 검토합니다. 각 개체에 관한 반복 측정값이 없으면 GLM 일변량이나 GLM 다변량을 사용합니다. 각 개체에 대해 두 가지 측정값(예: 사전검정과 사후 검정)만 있고 개체 간 요인이 없으면 대응표본 T 검정 프로시저를 사용할 수 있습니다.

GLM 반복 측도 얻기

1. 메뉴에서 다음을 선택합니다.

분석 > 일반 선형 모형 > 반복측도...

2. 개체-내 요인 이름과 수준의 수를 입력합니다.
3. 추가를 클릭합니다.
4. 각 개체-내 요인에 대해 이 단계를 반복합니다.

이중 다변량 반복 측도 디자인에 대한 측도 요인을 정의하려면 다음과 같이 하십시오.

5. 측도 이름을 입력합니다.
6. 추가를 클릭합니다.

요인과 측도를 모두 정의한 후 다음과 같이 하십시오.

7. 정의를 클릭합니다.
8. 목록에서 개체-내 요인들(선택적으로 측도)의 각 조합에 해당하는 종속변수를 하나 선택합니다.

변수 위치를 변경하려면 위/아래 화살표를 사용합니다.

개체-내 요인을 변경하려면 기본 대화 상자를 닫지 않은 채로 반복 측도 요인 정의 대화 상자를 다시 열면 됩니다. 선택적으로 개체 간 요인과 공변량을 지정할 수 있습니다.

GLM 반복 측도 요인 정의

GLM 반복 측도는 동일한 속성 가진 다른 측정값을 나타내는 관련 종속변수 그룹을 분석합니다. 이 대화 상자를 사용하면 GLM 반복 측도에 사용할 하나 이상의 개체-내 요인을 정의할 수 있습니다. 개체-내 요인은 지정하는 순서가 중요합니다. 각 요인은 이전 요인 내에서 한 수준이 됩니다.

반복 측도를 사용하려면 데이터를 정확하게 설정해야 합니다. 이 대화 상자에서 개체-내 요인들을 정의해야 합니다. 이 요인들은 데이터에 존재하지는 않지만 여기에 정의되어 있어야 합니다.

예제. 체중 감량 연구에서 몇 사람의 체중을 5주 동안 매주 측정한다고 가정합니다. 데이터 파일에서 각 사람은 개체나 케이스가 됩니다. 매주 측정된 체중은 변수 체중1, 체중2 등으로 코딩됩니다. 개인의 성별은 다른 변수로 코딩됩니다. 각 개체에 대해 반복적으로 측정한 체중은 개체-내 요인을 정의함으로써 그룹으로 구성될 수 있습니다. 이 요인을 주라고 하고 5가지 수준으로 정의할 수 있습니다. 기본 대화 상자에서 변수 체중1, ..., 체중5를 사용하여 주의 5가지 수준을 할당합니다. 남자와 여자(성별)로 분류한 데이터 파일의 변수를 개체 간 요인으로 지정하여 남자와 여자의 차이를 연구할 수 있습니다.

측도. 개체를 매번 둘 이상의 측도로 측정하려면 측도를 선택하여 측도 방법을 지정합니다. 예를 들어, 각 개체에 대한 맥박과 호흡 비율을 1주일 동안 매일 측정할 수 있습니다. 이들 측도는 데이터 파일에는 변수로 존재하지 않으나 여기에 정의되어 있습니다. 둘 이상의 측도를 포함하는 모형을 이중 다변량 반복 측도 모형이라고 합니다.

GLM 반복 측도 모형

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.

개체 간. 개체 간 요인 및 공변량이 나열됩니다.

모형. 모형은 데이터의 특성에 따라 다릅니다. 사용자 정의를 선택한 다음 분석 시 중요한 개체-내 효과와 상호작용 및 개체 간 효과와 상호작용을 선택할 수 있습니다.

제공합. 개체 간 모형에 대한 제공합 계산 방법입니다. 결측 셀이 없는 균형 모형이나 비균형 모형의 경우 제 III 유형 제공합 방법이 가장 일반적으로 사용됩니다.

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 제공합을 계산할 때 이 효과를 포함하지 않는 다른 효과에 맞게 수정되며 이 효과를 포함하는 효과(있는 경우)에 직교하는 제공합 방법을 사용합니다. 제 III 유형 제공합은 일반 추정 형식이 일정하게 유지되는 한 셀 빈도에 관해 변하지 않는 이점이 있습니다. 따라서 이 제공합 유형은 보통 결측 셀이 없는 비균형 모형에 유용한 것으로 간주됩니다. 이 방법은 결측 셀이 없는 요인 디자인에서 Yates의 가중 평균 제공 기법과 동등합니다. 제 III 유형 제공합 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 제공합 방법은 일반적으로 다음 경우에 사용됩니다.

- 제 I 유형과 제 II 유형으로 나열된 모든 모형
- 빈 셀이 있는 모든 균형 모형이나 비균형 모형

GLM 반복 측도 대비

대비는 개체 간 요인의 수준 차이를 검정하는 데 사용됩니다. 모형의 각 개체 간 요인에 대한 대비를 지정할 수 있습니다. 대비는 모수의 선형 조합을 나타냅니다.

가설 검정은 귀무가설 $LBM=0$ 을 기준으로 하며 여기에서 L 은 대비계수(L) 행렬이고 B 는 모수 벡터이며 M 은 종속변수의 평균 변환에 해당하는 평균 행렬을 나타냅니다. 반복 측도 옵션 대화 상자에서 변환 행렬을 선택하면 이 변환 행렬을 출력할 수 있습니다. 예를 들어, 네 개의 종속변수와 네 가지 수준의 개체-내 요인이 있고 다항 대비(기본값)가 개체-내 요인에 대해 사용되는 경우 M 행렬은 (0.5 0.5 0.5 0.5)이 됩니다. 대비가 지정되면 개체 간 요인에 해당하는 여러 열이 대비에 일치하는 L 행렬이 작성됩니다. 나머지 열은 L 행렬 추정이 가능하도록 수정됩니다.

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.

개체 내 요인에 대해 **없음**이 아닌 대비를 선택해야 합니다.

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 반복 측도 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. 도표에 대한 모든 요인을 사용할 수 있습니다. 각 종속변수에 대한 프로파일 도표가 만들어집니다. 개체 간 요인과 개체 내 요인 모두 프로파일 도표에서 사용할 수 있습니다.

한 요인에 대한 프로파일 도표는 추정된 주변 평균이 여러 수준에서 증가하는지 또는 감소하는지 여부를 보여줍니다. 둘 이상의 요인의 경우 평행선은 두 요인 간 상호작용이 없음을 나타내는데 이것은 한 요인의 수준만 조사할 수 있음을 의미합니다. 평행하지 않는 선들은 상호작용을 나타냅니다.

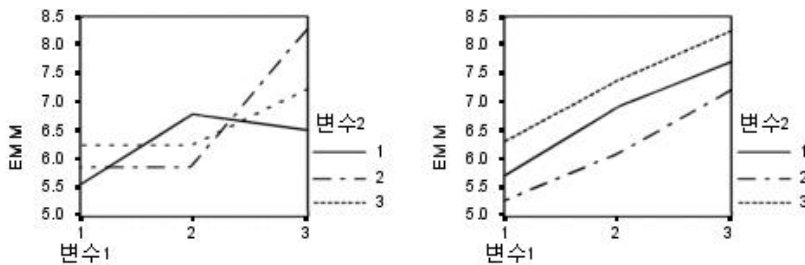


그림 2. 평행하지 않은 도표(왼쪽) 및 평행한 도표(오른쪽)

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

GLM 반복 측도 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. 개체 간 요인이 없으면 이 검정을 사용할 수 없으며 사후분석 다중 비교 검정은 개체-내 요인 수준 간의 평균에 대해 수행됩니다.

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. **Tukey**의 정직유의차 검정은 스튜던트화 범위 통계를 사용하여 모든 그룹 간 쌍대 비교를 수행하며 실험별 오차 비율을 모든 쌍대 비교 콜렉션의 오차 비율로 설정합니다. 다수의 대응평균을 검정할 경우 Tukey의 정직유의차 검정이 Bonferroni 검정보다 효과적인 결과를 나타냅니다. 소수의 대응평균에는 Bonferroni가 보다 효과적입니다.

Hochberg의 GT2는 Tukey의 정직유의차 검정과 유사하나 스튜던트화 최대 계수가 사용됩니다. 대개 Tukey의 검정이 더 효과적입니다. **Gabriel의 쌍대 비교 검정** 또한 스튜던트화 최대 계수를 사용하며 셀 크기가 동일하지 않을 경우 일반적으로 Hochberg의 GT2가 더 효과적인 방법입니다. Gabriel의 검정은 셀 크기가 상당히 다양하게 분포할 경우 광범위하게 사용될 수 있습니다.

Dunnett의 대응별 다중 비교 T 검정은 단일 통제 평균에 대해 하나의 처리 세트를 비교합니다. 마지막 범주가 기본 제어 범주가 됩니다. 처음 범주를 선택할 수도 있습니다. 또한 양쪽검정이나 한쪽검정을 선택할 수 있습니다. 한 요인 수준(제어 범주 제외)에서의 평균이 제어 범주에서의 평균과 동일하지 않은지 검정하려면 양쪽검정을 사용합니다. 한 요인 수준에서의 평균이 제어 범주에서의 평균보다 작을지를 검정하려면 $<$ 통제를 선택합니다. 마찬가지로 한 요인 수준의 평균이 제어 범주에서의 평균보다 클지를 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.

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 사용되는 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 반복 측도 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.
- **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 거리와 중심화하지 않은 레버리지 값을 사용할 수 있습니다.

- **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.

- **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.

계수 통계. 모수 추정값에 대한 분산-공분산 행렬을 데이터 세트에 저장합니다. 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 데이터는 행렬 파일을 읽는 다른 프로시저에도 사용할 수 있습니다. 데이터 세트는 같은 세션에서 다시 사용할 수 있으나 세션을 마치기 전에 명시적으로 저장하지 않는 한 파일로 저장되지 않습니다. 데이터 세트 이름은 변수이름 규칙을 준수해야 합니다.

GLM 반복 측도: 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. 개체-내 요인과 개체 간 요인을 모두 선택할 수 있습니다.

- **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. 가설과 오차 **SSCP** 행렬을 출력할 수 있으며 잔차 공분산 행렬에 대한 Bartlett의 단위행렬 검정과 더불어 잔차 **SSCP** 행렬을 출력할 수 있습니다.

동질성 검정은 개체 간 요인에 대해서만 개체 간 요인 수준의 모든 조합에서의 각 종속변수에 대해 분산의 동질성에 대한 Levene의 검정을 생성합니다. 동질성 검정에는 개체 간 요인 수준의 모든 조합에서의 종속변수들의 공분산 행렬에 대한 Box의 M 동질성 검정도 포함됩니다. 평균-산포 도표와 잔차 도표 옵션은 데이터에 대한 가정을 진단하는 데 유용합니다. 이 항목은 요인이 없을 경우에는 사용할 수 없습니다. 각 종속변수에 대한 관측-예측-표준화 잔차 도표를 작성하려면 잔차 도표를 선택합니다. 이들 도표는 동일한 분산에 대한 가정을 조사하는 데 유용합니다. 종속변수와 독립변수 간 관계를 해당 모형으로 충분히 설명할 수 있는지 확인하려면 적합결여 검정을 선택합니다. 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.

GLM 명령 추가 기능

이들 기능은 일변량, 다변량, 반복 측도 분석에 적용될 수 있습니다. 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 4 장 분산 성분추정 분석

혼합 효과 모형의 경우, 분산 성분 프로시저가 종속변수의 분산에 미치는 각 랜덤 효과의 영향력을 분석합니다. 이 프로시저는 특히 분할계획법, 일변량 반복 측도, 난괴법(random block design)과 같은 혼합 모형의 분석과 관련되어 있습니다. 분산 성분을 계산하면 분산을 줄이기 위해 주안점을 두어야 할 부분을 결정할 수 있습니다.

분산 성분 추정에 사용할 수 있는 네 가지 방법은 최소 정규 2차 비편향 추정량(MINQUE), 분산 분석(ANOVA), 최대-우도(ML), 제한된 우도비(REML)입니다. 각기 다른 방법에 대해 다양한 지정을 사용할 수 있습니다.

모든 방법에 대한 기본값 결과에는 분산 성분 추정값이 포함됩니다. ML 방법이나 REML 방법을 사용하면 근사 공분산 행렬표가 함께 출력됩니다. 사용할 수 있는 다른 결과로는 분산 분석표와 분산 분석 방법에 대한 기대 평균 제곱이 포함되며 ML 방법과 REML 방법에 대한 반복 히스토리가 포함됩니다. 분산 성분 프로시저는 GLM 일변량 분석 프로시저와 완전히 호환됩니다.

WLS 가중치를 사용하면 측정 정밀도 변동을 보완하기 위해 관측값에 가중 분석에 대해 서로 다른 가중치를 부여하는 변수를 지정할 수 있습니다.

예제. 한 농업 학교에서 여섯 가지의 다른 깔짚을 사용한 우리에서 기른 돼지의 체중 증가량을 1개월 후에 측정하였습니다. 깔짚 변수는 여섯 가지 수준으로 된 변량요인입니다. (연구를 실시한 여섯 가지 깔짚은 대규모의 돼지 깔짚 모집단에서 무작위 표본을 나타냅니다.) 조사자들은 체중 증가에 대한 분산이 생긴 이유는 깔짚의 차이가 깔짚에서 키운 돼지의 차이보다 훨씬 크기 때문이라는 사실을 발견하였습니다.

분산 성분추정 데이터 고려 사항

데이터. 종속변수는 양적변수입니다. 요인은 범주형입니다. 최대 8바이트까지 숫자나 문자열 값을 가질 수 있습니다. 적어도 요인 중의 하나는 변량요인이어야 합니다. 즉, 요인의 수준은 가능한 수준에 대해 무작위 표본해야 합니다. 공변량은 종속변수와 관련된 양적변수입니다.

가정. 모든 방법은 랜덤 효과의 모형 모수가 0 평균과 유한 일정 분산값을 가지며 상호 관련되지 않는 것으로 가정합니다. 다른 랜덤 효과에서 생긴 모형 모수도 상호관련이 없습니다.

잔차 항 역시 0 평균과 유한 일정 분산값을 갖습니다. 또한 어떠한 랜덤 효과의 모형 모수와도 상호관련이 있지 않습니다. 서로 다른 관측으로 생긴 잔차 항도 상호관련되지 않은 것으로 가정합니다.

이러한 가정을 기준으로 하면, 동일한 수준의 변량요인에서 얻은 관측값은 상호관련이 있습니다. 이를 통해 분산 성분 모형을 일반선형 모형과 구분할 수 있습니다.

분산 분석과 MINQUE는 정규성 가정이 필요하지 않습니다. 두 가지 모두 정규성 가정에서 다소 이탈해도 영향을 받지 않습니다.

ML과 REML에서 모형 모수와 잔차 항은 정규 분포이어야 합니다.

관련 프로시저. 분산 성분 분석을 수행하기 전에 데이터를 검사하려면 데이터 탐색 프로시저를 사용합니다. 가설 검정의 경우 GLM 일반요인, GLM 다변량, GLM 반복 측도를 사용합니다.

분산 성분추정 분석표 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 일반 선형 모형 > 분산 성분...

2. 종속변수를 한 개 선택합니다.

3. 현재 데이터에 맞게 모수요인, 변량요인, 공변량에 대한 변수를 선택합니다. 가중 변수를 지정하려면 WLS 가중치를 사용합니다.

분산 성분추정: 모형

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. 모형에는 변량요인이 포함되어야 합니다.

모형에 절편 포함. 모형에는 보통 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.

분산 성분추정: 옵션

방법. 분산 성분추정에 사용되는 다음 네 가지 방법 중 하나를 선택할 수 있습니다.

- **MINQUE**(최소 정규 2차 비편향 추정량)은 모수효과에 대해 일정한 추정값을 생성합니다. 데이터가 정규 분포를 이루고 추정값이 정확하면 이 방법은 모든 비편향 추정량 사이에서 최저 분산을 생성합니다. 랜덤 효과 사전 확률 가중치에 대한 방법을 선택할 수 있습니다.
- **ANOVA**(분산 분석)은 각 효과에 대한 제 I 유형 제곱합이나 제 III 유형 제곱합 중 하나를 사용하여 비편향 추정량을 계산합니다. 분산 분석 방법은 음의 분산 추정값을 생성하는 경우가 있는데 이것은 모형이 잘못되었거나 추정 방법이 부적절하거나 더 많은 데이터가 필요한 경우에 나타납니다.
- **최대-우도(ML)**는 반복을 사용하여 실제로 관측한 데이터와 가장 많이 일치하는 추정값을 생성합니다. 이들 추정값은 편향될 수도 있습니다. 이 방법은 비동시적으로 정규성을 나타냅니다. ML과 REML 추정값은 변환 중에 변하지 않습니다. 이 방법은 고정 효과를 추정하는 데 사용되는 자유도는 고려하지 않습니다.
- **제한된 우도비(REML)** 추정값을 사용하면 많은(전체는 아님) 균형 데이터 케이스에 대한 분산 분석 추정값이 감소됩니다. 이 방법은 고정 효과에 따라 달라지므로 ML 방법을 사용하는 경우보다 표준 오차가 더 작습니다. 이 방법은 고정 효과를 추정하는 데 사용되는 자유도를 고려합니다.

랜덤 효과 사전 확률. 균일은 모든 랜덤 효과와 잔차 항이 관측값에 대해 동일한 영향을 미친다는 것을 의미합니다. **0값** 표시는 0인 랜덤 효과 분산값을 가정하는 것과 동일합니다. 이 옵션은 MINQUE 방법에서만 사용할 수 있습니다.

제곱합. 제 I 유형 제곱합은 계층형 모형에 사용되는 방법으로, 분산 성분추정 분야에 사용됩니다. GLM의 기본값인 제 III 유형을 선택하면 분산 추정값은 제 III 유형 제곱합을 사용하는 가설 검정을 위해 GLM 일반화 분석에서 사용됩니다. 이 옵션은 분산 분석 방법에만 사용할 수 있습니다.

기준. 수렴 기준과 최대반복수를 지정할 수 있습니다. ML 방법이나 REML 방법에서만 사용할 수 있습니다.

출력. 분산 분석 방법에서 제곱합과 기대 평균 제곱을 출력할 수 있습니다. 최대-우도 방법이나 제한된 우도비 방법을 선택한 경우에는 반복 히스토리를 출력할 수 있습니다.

제곱합(분산 성분추정)

모형에 대해 한 가지 유형의 제곱합을 선택할 수 있습니다. 제 III 유형이 가장 일반적으로 사용되며 기본값입니다.

제 I 유형. 이 방법은 또한 제곱합의 계층적 분해 방법으로 알려져 있습니다. 각 항은 모형에서 해당 항의 앞에 오는 항에 따라서만 수정됩니다. 제 I 유형 제곱합 방법은 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 III 유형. 기본값입니다. 이 방법을 사용하면 디자인내의 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.

다. 따라서 이 유형은 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 유형에 나열된 모든 모형
- 빈 셀이 없는 모든 균형 모형이나 비균형 모형

분산 성분추정: 새 파일로 저장

이 프로시저의 몇 가지 결과를 새 IBM SPSS Statistics 데이터 파일에 저장할 수 있습니다.

분산 성분 추정값. 분산 성분 추정값과 추정값 레이블을 데이터 파일 또는 데이터 세트에 저장합니다. 이 옵션은 GLM 프로시저에서 더 많은 통계를 계산하거나 추가 분석을 수행하는 데 사용할 수 있습니다. 예를 들어, 신뢰구간을 계산하거나 가설을 검정하는 데 사용할 수 있습니다.

성분 공변동. 공분산 행렬이나 상관 행렬을 데이터 파일 또는 데이터 세트에 저장합니다. 이 옵션은 최대-우도나 제한된 우도비를 지정한 경우에만 사용할 수 있습니다.

작성된 값의 대상. 분산 성분 추정값 및 행렬을 포함하는 파일의 외부 이름이나 데이터 세트 이름을 지정할 수 있습니다. 데이터 세트는 같은 세션에서 다시 사용할 수 있으나 세션을 마치기 전에 명시적으로 저장하지 않는 한 파일로 저장되지 않습니다. 데이터 세트 이름은 변수이름 규칙을 준수해야 합니다.

MATRIX 명령을 사용하여 필요한 데이터를 데이터 파일에서 추출한 후 신뢰구간을 계산하거나 검정을 수행할 수 있습니다.

VARCOMP 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- DESIGN 하위 명령을 사용하여 디자인에 중첩 효과를 지정합니다.
- MISSING 하위 명령을 사용하여 사용자 결측값을 포함시킵니다.
- CRITERIA 하위 명령을 사용하여 EPS 기준을 지정합니다.

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

제 5 장 선형 혼합 모형

선형 혼합 모형 프로시저는 일반 선형 모형을 확장하여 데이터가 상호 관련되고 일정하지 않은 변동을 나타낼 수 있도록 합니다. 따라서 선형 혼합 모형은 데이터의 평균 뿐만 아니라 데이터의 분산 및 공분산 모형을 만드는 데 유연성을 제공합니다.

또한 선형 혼합 모형 프로시저는 다른 모형을 선형 혼합 모형으로 구성할 수 있도록 지원하는 유연성 있는 도구를 제공합니다. 이러한 모형에는 다중 수준 모형, 계층적 선형 모형, 변량 계수 모형 등이 포함됩니다.

예제. 한 식료품 체인점에서 다양한 쿠폰이 고객의 소비에 미치는 영향을 알아보려고 합니다. 고정 고객 중에서 무작위로 표본 추출하여 10주 동안 각 고객의 소비 행태를 살펴봅니다. 매주마다 다른 쿠폰이 고객에게 배달됩니다. 선형 혼합 모형을 사용하여 10주 동안 각 개체에 대해 반복적으로 발생한 관측값에 따라 상관을 조정하고 여러 가지 쿠폰이 소비에 끼친 영향을 추정할 수 있습니다.

방법. 최대우도(ML) 및 제한 최대우도(REML)를 추정합니다.

통계. 기술통계량에 대해 표본 결과, 평균, 종속변수의 표준 편차, 요인 수준 조합별 공변량을 구합니다. 요인 수준 정보에 대해 각 요인 수준에 대한 정렬값 및 정렬값의 빈도를 구합니다. 또한 모수효과에 대한 모수 추정값 및 신뢰구간, 공분산 행렬의 모수에 대한 Wald 검정 및 신뢰구간을 구합니다. 제곱합의 제 I 유형 및 제 III 유형을 사용하여 다른 가설을 평가할 수 있습니다. 제 III 유형이 기본값입니다.

선형 혼합 모형 데이터 고려 사항

데이터. 종속변수는 양적변수여야 합니다. 요인은 범주형이어야 하며 숫자값 또는 문자열 값을 포함할 수 있습니다. 공변량 변수 및 가중 변수는 양적변수여야 합니다. 개체변수 및 반복변수는 어느 유형이든 관계없습니다.

가정. 종속변수는 모수요인, 변량요인 및 공변량에 선형적으로 관련된 것으로 가정합니다. 모수효과는 종속변수의 평균 모형을 작성하고 랜덤 효과는 종속변수의 공분산 구조 모형을 작성합니다. 다중 랜덤 효과는 상호 독립적인 것으로 간주되며 각각에 대해 개별 공분산 행렬이 계산됩니다. 그러나 동일한 랜덤 효과에 대해 지정된 모형 항을 상호관련시킬 수는 있습니다. 반복 측도는 잔차의 공분산 구조 모형을 작성합니다. 또한 종속변수는 정규 분포에서 비롯된 것으로 가정합니다.

관련 프로시저. 분석 수행 이전에 데이터 탐색 프로시저를 사용하여 데이터를 검토합니다. 상호 관련되거나 일정하지 않은 변동량이 없는 경우에는 GLM 일변량 또는 GLM 반복 측도 프로시저를 사용할 수 있습니다. 랜덤 효과에 분산 성분 공분산 구조가 있고 반복 측도가 없는 경우에는 분산 성분 분석 프로시저를 사용할 수 있습니다.

선형 혼합 모형 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 혼합 모형 > 선형...

2. 선택적으로 개체변수 하나 이상을 선택합니다.
3. 선택적으로 반복변수 하나 이상을 선택합니다.
4. 선택적으로 잔차 공분산 구조를 선택합니다.
5. 계속을 클릭합니다.
6. 종속변수를 한 개 선택합니다.
7. 요인 또는 공변량 하나 이상을 선택합니다.
8. 고정 또는 임의를 클릭하고 고정 효과 또는 랜덤 효과 모형을 지정합니다.

선택적으로 가중 변수를 선택합니다.

선행 혼합 모형의 개체/반복변수 선택

이 대화 상자에서는 개체 및 반복 관측값을 정의하는 변수를 선택하고 잔차에 대한 공분산 구조를 선택할 수 있습니다.

개체. 개체는 다른 개체에 대해 독립적인 것으로 간주할 수 있는 관측 단위입니다. 예를 들어, 의학 연구에서 한 환자의 혈압 기록은 다른 환자의 기록에 대해 독립적인 것으로 간주할 수 있습니다. 개체마다 반복 측정값이 있고 이러한 관측값 간의 상관을 모델링하려는 경우 개체를 정의하는 것이 매우 중요합니다. 예를 들어, 담당 의사에게 지속적으로 진찰을 받는 한 환자의 혈압 기록은 상호 관련된 것으로 예상할 수 있습니다.

또한 개체를 다중변수의 요인 수준 조합별로 정의할 수 있습니다. 예를 들어, 성별 및 연령범주를 개체변수로 지정하여 65세 이상의 남자는 서로 비슷하지만 65세 이하의 남자 및 여자와는 별개라는 사실을 모델링할 수 있습니다.

개체 목록에 지정된 모든 변수는 잔차 공분산 구조에 대한 개체를 정의하는 데 사용됩니다. 이 변수 중 일부 또는 전부 사용하여 랜덤 효과 공분산 구조에 대한 개체를 정의할 수 있습니다.

반복. 이 목록에 지정된 변수는 반복 관측값을 식별하는 데 사용됩니다. 예를 들어, 단일 변수 주는 의학 연구에서의 10주 동안의 관측을 식별하는 데 사용하거나 월 및 일은 1년 동안의 일별 관측을 식별하는 데 함께 사용할 수 있습니다.

반복 공분산 유형. 잔차에 대한 공분산 구조를 지정합니다. 사용 가능한 구조는 다음과 같습니다.

- 선행 종속성: 1차
- AR(1)
- AR(1): 이질적
- ARMA(1,1)
- 복합 대칭
- 복합 대칭: 상관 매트릭
- 복합 대칭: 이질적
- 대각선

- 요인 분석: 1차
- 요인 분석: 1차, 이질적
- Huynh-Feldt
- 척도법 항등
- Toeplitz
- Toeplitz: 이질적
- 비구조적
- 비구조적: 상관

자세한 정보는 107 페이지의 제 17 장 『공분산 구조』 주제를 참조하십시오.

선형 혼합 모형 고정 효과

고정 효과 기본값 모형이 없으므로 고정 효과를 명시적으로 지정해야 합니다. 또는 중첩되거나 중첩되지 않은 항을 설정할 수 있습니다.

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

제공함. 제공함을 계산하는 방법입니다. 결측 셀이 없는 모형의 경우 제 III 유형 방법이 가장 일반적으로 사용됩니다.

중첩되지 않은 항 설정

선택된 요인과 공변량에 대해 다음과 같습니다.

요인. 선택한 변수에 대해 가능한 상호작용 및 주효과를 모두 생성합니다. 기본값입니다.

상호작용. 선택한 모든 변수의 최고-수준 상호작용 항을 생성합니다.

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.

중첩 항 설정

이 프로시저에서는 모형에 대해 중첩 항을 설정할 수 있습니다. 중첩 항은 요인 또는 공변량의 효과를 모델링하는 데 유용합니다. 이들 값은 다른 요인 수준과 상호작용하지 않습니다. 예를 들어, 식료품 체인점은 여러 지점에서의 고객 소비를 살펴볼 수 있습니다. 각 고객은 체인점 중의 한 곳만 자주 가기 때문에 고객 효과는 점포 위치 효과 내에 중첩되었다고 할 수 있습니다.

또한 상호작용 효과를 포함시키거나 여러 수준의 중첩을 중첩 항에 추가할 수 있습니다.

제한. 중첩 항에는 다음과 같은 제한이 있습니다.

- 상호작용 내의 모든 요인은 고유해야 합니다. 따라서 A 가 요인이면 $A*A$ 지정은 유효하지 않습니다.
- 중첩 효과 내의 모든 요인은 고유해야 합니다. 따라서 A 가 요인이면 $A(A)$ 지정은 유효하지 않습니다.
- 공변량 내에 효과를 중첩할 수 없습니다. 따라서 A 가 요인이고 X 가 공변량이면 A 지정은 유효하지 않습니다.

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 항을 선택하는 항에 대해서만 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 III 유형. 기본값입니다. 이 방법을 사용하면 디자인내의 한 효과의 제공합을 계산할 때 이 효과를 포함하지 않는 다른 효과에 맞게 수정되며 이 효과를 포함하는 효과(있는 경우)에 직교하는 제공합 방법을 사용합니다. 제 III 유형 제공합은 일반 추정 형식이 일정하게 유지되는 한 셀 빈도에 관해 변하지 않는 이점이 있습니다. 따라서 이 제공합 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 제공합 방법은 일반적으로 다음 경우에 사용됩니다.

- 제 I 유형에 나열된 모든 모형
- 빈 셀이 없는 모든 균형 모형이나 비균형 모형

선형 혼합 모형 랜덤 효과

공분산 유형. 랜덤 효과 모형에 대한 공분산 구조를 지정할 수 있습니다. 개별 공분산 행렬이 각 랜덤 효과에 대해 추정됩니다. 사용 가능한 구조는 다음과 같습니다.

- 선행 종속성: 1차
- AR(1)
- AR(1): 이질적
- ARMA(1,1)
- 복합 대칭
- 복합 대칭: 상관 매트릭
- 복합 대칭: 이질적
- 대각선
- 요인 분석: 1차
- 요인 분석: 1차, 이질적
- Huynh-Feldt
- 척도법 항등
- Toeplitz
- Toeplitz: 이질적
- 비구조적
- 비구조적: 상관 매트릭
- 분산 성분

자세한 정보는 107 페이지의 제 17 장 『공분산 구조』 주제를 참조하십시오.

랜덤 효과 기본 모형이 없으므로 랜덤 효과를 명시적으로 지정해야 합니다. 또는 중첩되거나 중첩되지 않은 항을 설정할 수 있습니다. 랜덤 효과 모형에 절편 항을 포함하도록 선택할 수도 있습니다.

여러 랜덤 효과 모형을 지정할 수 있습니다. 첫 번째 모형을 설정한 후 다음을 클릭하여 다음 모형을 설정합니다. 이전을 클릭하여 기존 모형을 스크롤할 수 있습니다. 각 랜덤 효과 모형은 다른 모든 랜덤 효과 모형에 대해 독립적인 것으로 가정합니다. 즉, 개별 공분산 행렬이 각각에 대해 계산됩니다. 동일한 랜덤 효과 모형에 지정된 항을 상호관련시킬 수 있습니다.

개체 그룹. 나열된 변수는 개체/반복변수 선택 대화 상자에서 개체변수로 선택한 변수입니다. 이 변수 중 일부 또는 전부 선택하여 랜덤 효과 모형에 대해 개체를 정의합니다.

선형 혼합 모형: 추정

방법. 최대우도 또는 제한 최대우도 추정을 선택합니다.

반복: 다음 옵션을 사용할 수 있습니다.

- 최대반복수. 음수가 아닌 정수를 지정합니다.
- 최대 단계 반복. 각 반복에서 단계 크기는 로그 우도 증가 또는 최대 단계 반복에 도달할 때까지 요인이 0.5씩 감소됩니다. 양의 정수를 지정합니다.
- 다음 단계마다 반복 히스토리 인쇄: **n** 단계. 0^{번째} 반복(초기 추정값)부터 시작하여 각 *n* 반복마다 로그 우도 함수 값 및 모수 추정값이 포함된 테이블이 표시됩니다. 반복 히스토리 인쇄를 선택하면 *n* 값에 관계 없이 항상 마지막 반복이 인쇄됩니다.

로그 우도 수렴. 로그-우도 함수의 절대 변화량 또는 상대 변화량이 지정된 값(음수가 아니어야 함)미만인 경우 수렴을 가정합니다. 지정된 값이 0이면 기준을 사용하지 않습니다.

모수 수렴. 모수 추정값의 최대 절대 변화량 또는 최대 상대 변화량이 지정된 값(음수가 아니어야 함) 미만인 경우 수렴을 가정합니다. 지정된 값이 0이면 기준을 사용하지 않습니다.

Hessian 수렴. 절대값 지정의 경우 Hessian을 기준으로 하는 통계가 지정된 값 미만인 경우 수렴을 가정합니다. 상대값 지정의 경우 수렴은 통계가 로그 우도의 절대값과 지정된 값의 곱보다 작다고 가정합니다. 지정된 값이 0이면 기준을 사용하지 않습니다.

최대 스코어링 단계. 반복 횟수 *n*까지 Fisher 스코어링 알고리즘을 사용하도록 요청합니다. 양수를 지정하십시오.

비정칙성 허용 오차. 이 값은 비정칙성 확인 시 허용 오차로 사용합니다. 양수를 지정하십시오.

선형 혼합 모형 통계

요약 통계. 다음에 대한 테이블을 작성합니다.

- 기술통계. 표본 결과, 평균, 종속변수의 표준 편차, 공변량(지정한 경우)이 표시됩니다. 이러한 통계는 요인 수준 조합별로 표시됩니다.
- 케이스 처리 요약. 정렬된 요인 값, 반복 측도 변수, 반복 측도 개체, 랜덤 효과 개체와 빈도가 표시됩니다.

모형 통계. 다음에 대한 테이블을 작성합니다.

- 모수 추정값. 모수효과 및 랜덤 효과 모수 추정값, 이에 대한 대략적인 표준 오차가 표시됩니다.
- 공분산 모수 검정. 공분산 모수에 대한 근사 표준 오차 및 Wald 검정이 표시됩니다.
- 모수 추정값의 상관. 모수효과 모수 추정값의 근사 상관 행렬이 표시됩니다.
- 모수 추정값의 공분산. 모수효과 모수 추정값의 근사 공분산 행렬이 표시됩니다.
- 랜덤 효과 공분산. 추정된 랜덤 효과의 공분산 행렬이 표시됩니다. 이 옵션은 랜덤 효과를 하나 이상 지정한 경우에만 사용할 수 있습니다. 랜덤 효과에 대해 개체변수가 지정된 경우에는 공통 블록이 표시됩니다.

- 잔차 공분산. 추정된 잔차 공분산 행렬이 표시됩니다. 이 옵션은 반복변수가 지정된 경우에만 사용할 수 있습니다. 개체변수가 지정된 경우에는 공동 블록이 표시됩니다.
- 대비계수(L) 행렬. 이 옵션은 고정 효과 및 사용자 정의 가설을 검증하는 데 사용되는 추정가능 함수를 표시합니다.

신뢰구간. 이 값은 신뢰구간이 구성될 때마다 사용됩니다. 0보다 크거나 같고 100보다 작은 값을 지정합니다. 기본값은 95입니다.

선형 혼합 모형 EM 평균

적합 모형의 주변 평균 추정. 이 옵션을 사용하면 셀에 있는 종속변수의 모형 예측 주변 평균 추정 및 지정된 요인에 대한 주변 평균 추정의 표준 오차를 요청할 수 있습니다. 또한 비교할 주효과의 요인 수준을 요청할 수 있습니다.

- 요인 및 요인 상호작용. 이 목록에는 전체 항을 비롯하여 고정 대화 상자에서 지정한 요인 및 요인 상호작용이 포함됩니다. 공변량에서 설정한 모형 항은 이 목록에서 제외됩니다.
- 평균 출력 기준. 이 프로시저는 이 목록에서 선택한 요인 및 요인 상호작용에 대해 주변 평균 추정을 계산합니다. 전체를 선택하면 모든 요인이 합쳐지고 종속변수의 주변 평균 추정이 표시됩니다. 선택한 모든 요인 또는 요인 상호작용은 연결된 변수가 기본 대화 상자의 요인 목록에서 제거되지 않으면 선택된 상태로 유지됩니다.
- 주효과 비교. 이 옵션을 사용하면 선택한 주효과 수준의 쌍대 비교를 요청할 수 있습니다. 신뢰구간 조정을 사용하면 조정사항을 신뢰구간 및 유의수준에 적용하여 다중 비교를 설명할 수 있습니다. 사용 가능한 메소드는 LSD (조정 없음), Bonferroni 및 Sidak입니다. 마지막으로 각 요인에 대해 비교할 참조 범주를 선택할 수 있습니다. 선택된 참조 범주가 없으면 모든 쌍대 비교가 구성됩니다. 참조 범주에 대한 옵션은 처음, 마지막 또는 사용자 정의(참조 범주 값을 입력한 경우)입니다.

선형 혼합 모형 저장

이 대화 상자에서 작업 파일에 대한 여러 모형 결과를 저장할 수 있습니다.

고정 예측값. 효과가 없는 회귀분석 평균과 관련된 변수를 저장합니다.

- 예측값. 랜덤 효과가 없는 회귀분석 평균입니다.
- 표준 오차. 추정값의 표준 오차입니다.
- 자유도. 추정값과 관련된 자유도입니다.

예측값 및 잔차. 모형 적합값과 관련된 변수를 저장합니다.

- 예측값. 모형 적합값입니다.
- 표준 오차. 추정값의 표준 오차입니다.
- 자유도. 추정값과 관련된 자유도입니다.
- 잔차. 데이터값에서 예측값을 뺀 값입니다.

MIXED 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- TEST 하위 명령을 사용하여 효과에 대한 효과-대 선형 조합에 대한 검정이나 값을 지정합니다.
- MISSING 하위 명령을 사용하여 사용자 결측값을 포함시킵니다.
- EMMEANS 하위 명령의 WITH 키워드를 사용하여 지정한 공변량 값에 대한 주변 평균 추정을 계산합니다.
- EMMEANS 하위 명령을 사용하여 상호작용의 주효과를 비교합니다.

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

제 6 장 일반화 선형 모형

일반화 선형 모형은 종속변수가 지정된 연결 함수를 통해 요인 및 공변량과 선형적으로 관련되도록 일반 선형 모형을 확장합니다. 더욱이 모형을 사용하면 종속변수가 비정규 분포를 갖는 것을 허용합니다. 또한 정상적으로 분포된 반응, 이분형 데이터의 로지스틱 모형, 개수 데이터의 로그선형 모형, 구간 중도절단 생존 데이터에 대한 음의 로그-로그 모형은 물론 매우 일반적인 모형 공식을 통해 다른 많은 통계 모형 같이 널리 사용되는 통계 모형을 포함합니다.

예제. 한 선박 회사는 다른 시기에 건조된 여러 종류의 선박에 대한 손상 개수를 파악하기 위해 포아송 회귀분석에 맞게 일반화 선형 모형을 사용할 수 있으며 결과 모형은 어느 종류의 선박이 가장 쉽게 손상되는지 결정할 수 있습니다.

자동차 보험 회사는 자동차 손해 배상을 파악하기 위해 감마 회귀분석에 맞게 일반화 선형 모형을 사용할 수 있으며 결과 모형은 손해 배상 규모에 가장 많이 기여하는 요인을 결정할 수 있습니다.

의료 연구자들은 의료 조건에 대해 되풀이 되는 시간을 예측하기 위해 구간 중도절단 생존 데이터에 대한 보완 로그-로그 회귀분석에 맞게 일반화 선형 모형을 사용할 수 있습니다.

일반화 선형 모형 데이터 고려 사항

데이터. 반응은 척도, 개수, 이분형 또는 시행한 이벤트가 될 수 있습니다. 요인은 범주형으로 가정됩니다. 공변량, 척도 가중치 및 오프셋은 척도로 가정됩니다.

가정. 케이스는 독립된 관측으로 가정됩니다.

일반화 선형 모형 얻기

메뉴에서 다음을 선택합니다.

분석 > 일반화 선형 모형 > 일반화 선형 모형...

1. 분포 및 연결 함수를 지정합니다(다양한 옵션에 대한 자세한 내용은 아래를 참조하십시오).
2. 반응 탭에서 종속변수를 선택합니다.
3. 예측변수 탭에서 종속변수를 예측하는 데 사용할 요인과 공변량을 선택합니다.
4. 모형 탭에서 선택한 요인과 공변량을 사용하여 모형 효과를 지정합니다.

모형 유형을 사용하면 모형에 대한 분포 및 연결 함수를 지정하여 반응 유형별로 범주화된 여러 공통 모형에 대한 바로 가기를 제공할 수 있습니다.

모형 유형

척도 반응. 다음 옵션을 사용할 수 있습니다.

- 선형. 정규를 분포로, 항등을 연결 함수로 지정합니다.
- 로그 링크가 있는 감마. 감마를 분포로, 로그를 연결 함수로 지정합니다.

순서 반응. 다음 옵션을 사용할 수 있습니다.

- 순서 로지스틱. 다항(순서)을 분포로, 누적 로짓을 연결 함수로 지정합니다.
- 순서 프로빗. 다항(순서)을 분포로, 누적 프로빗을 연결 함수로 지정합니다.

개수. 다음 옵션을 사용할 수 있습니다.

- 포아송 로그선형. 포아송을 분포로, 로그를 연결 함수로 지정합니다.
- 로그 링크가 있는 음수 이항. 음수 이항(보조 모수에 대해 값 1을 가짐)을 분포로, 로그를 연결 함수로 지정합니다. 프로시저 추정값이 보조 모수의 값을 갖도록 하려면 음수 이항 분포를 가진 사용자 정의 모형을 지정하고 모수 그룹에서 추정값을 선택합니다.

이분형 반응 또는 이벤트/시행 데이터. 다음 옵션을 사용할 수 있습니다.

- 이분형 로지스틱. 이항을 분포로, 로짓을 연결 함수로 지정합니다.
- 이분형 프로빗. 이항을 분포로, 프로빗을 연결 함수로 지정합니다.
- 구간 중도절단 생존. 이항을 분포로, 보완 로그-로그를 연결 함수로 지정합니다.

혼합. 다음 옵션을 사용할 수 있습니다.

- 로그 링크가 있는 Tweedie. Tweedie를 분포로, 로그를 연결 함수로 지정합니다.
- 항등 링크가 있는 Tweedie. Tweedie를 분포로, 항등을 연결 함수로 지정합니다.

사용자 정의. 분포와 연결 함수의 자체 조합을 지정합니다.

분포

이 선택은 종속변수의 분포를 지정합니다. 비정규 분포와 항등하지 않은 연결 함수를 지정하는 기능은 일반 선형 모형에서 일반화 선형 모형의 중요한 개선 사항입니다. 많은 분포-연결 함수가 있으며 주어진 데이터 세트에 적합한 함수가 여러 개일 수 있으므로 사전 이론적 고려 사항을 바탕으로 선택하거나 어느 조합이 가장 적합할지를 고려하여 선택하면 됩니다.

- 이항검정. 이 분포는 이분형 반응이나 이벤트 수를 나타내는 변수의 경우에만 적합합니다.
- 감마. 이 분포는 더 큰 양의 값 쪽으로 비대칭되는 양의 척도 값을 가진 변수에 적합합니다. 데이터 값이 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- 역 가우스. 이 분포는 더 큰 양의 값 쪽으로 비대칭되는 양의 척도 값을 가진 변수에 적합합니다. 데이터 값이 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- 음수 이항. 이 분포는 k 성공을 관측하는 데 필요한 시행 횟수로 생각할 수 있으며 양의 정수 값을 갖는 변수에 적합합니다. 데이터 값이 양수이거나, 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다. 음수 이항 분포의 보조 모수 값은 0보다 크거나 같은 숫자가 될 수 있습니다. 이 값을 고정값으로 설정하거나 프로시저에서 추정하도록 할 수 있습니다. 보조 모수가 0으로 설정되면 분포를 사용하는 것은 포아송 분포를 사용하는 것과 동일합니다.

- 정상. 이것은 중앙(평균) 값에 대해 값이 대칭되는 종 형태의 분포를 띠는 척도변수에 적합합니다. 종속변수는 숫자여야 합니다.
- 포아송. 이 분포는 고정 기간 동안 중요 이벤트의 발생 수로 생각할 수 있으며 양의 정수 값을 갖는 변수에 적합합니다. 데이터 값이 양수이거나, 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- Tweedie. 이 분포는 감마 분포의 포아송 혼합으로 표현할 수 있는 변수에 적합합니다. 분포는 연속 특성(음이 아닌 실수 값 사용)과 이산형 분포(단일 값 0에서 양의 확률 매스)의 관점에서 "혼합"된 것입니다. 종속변수는 데이터 값이 0보다 크거나 같은 숫자가 되어야 합니다. 데이터 값이 0보다 작거나 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다. Tweedie 분포에서 모수의 고정 값은 1보다 크고 2보다 작은 숫자가 될 수 있습니다.
- 다항. 이 분포는 순서 반응을 나타내는 변수에 적합합니다. 종속변수는 숫자나 문자열이 될 수 있으며 최소 두 개의 유효한 개별 데이터 값을 가져야 합니다.

연결 함수

연결 함수는 종속변수의 변환으로 모형을 추정할 수 있습니다. 다음 함수를 사용할 수 있습니다.

- 항등. $f(x)=x$. 종속변수가 변환되지 않습니다. 이 링크는 분포에 사용할 수 있습니다.
- 보 로그-로그. $f(x)=-\log(1-x)$. 이항 분포에만 적합합니다.
- 누적 Cauchit. $f(x) = \tan(\pi (x - 0.5))$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.
- 누적 보 로그-로그. $f(x)=-\ln(1-x)$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.
- 누적 로짓. $f(x)=\ln(x / (1-x))$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.
- 누적 음수 로그-로그. $f(x)=-\ln(-\ln(x))$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.
- 누적 프로빗. $f(x)=\Phi^{-1}(x)$, 반응에서 각 범주의 누적 확률에 적용됩니다. 여기서 Φ^{-1} 은 표준 정규 누적 분포의 역함수입니다. 다항 분포에만 적합합니다.
- 로그. $f(x)=\log(x)$. 이 링크는 분포에 사용할 수 있습니다.
- 로그 보. $f(x)=\log(1-x)$. 이항 분포에만 적합합니다.
- 로짓. $f(x)=\log(x / (1-x))$. 이항 분포에만 적합합니다.
- 음수 이항. $f(x)=\log(x / (x+k^{-1}))$, 여기서 k 는 음수 이항 분포의 보조 모수입니다. 음수 이항 분포에만 적합합니다.
- 음수 로그-로그. $f(x)=-\log(-\log(x))$. 이항 분포에만 적합합니다.
- 승산비. $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$, $\alpha \neq 0$ 인 경우. $f(x)=\log(x)$, $\alpha=0$ 인 경우. α 는 필요한 숫자 지정이며 실수가 되어야 합니다. 이항 분포에만 적합합니다.
- 프로빗. $f(x)=\Phi^{-1}(x)$, 여기서 Φ^{-1} 은 표준 정규 누적 분포의 역함수입니다. 이항 분포에만 적합합니다.

- 거듭제곱. $f(x)=x^{\alpha}$, $\alpha \neq 0$ 인 경우. $f(x)=\log(x)$, $\alpha=0$ 인 경우. α 는 필요한 숫자 지정이며 실수가 되어야 합니다. 이 링크는 분포에 사용할 수 있습니다.

일반화 선형 모형 반응

많은 경우에 종속변수를 간단히 지정할 수 있지만 두 값만 사용하고 시행에 이벤트를 기록하는 변수에 주의를 기울여야 합니다.

- 이분형 반응. 종속변수가 두 값만 사용할 때는 모수 추정에 대해 참조 범주를 지정할 수 있습니다. 이분형 반응변수는 문자 또는 숫자가 될 수 있습니다.
- 시행 세트에서 발생하는 이벤트의 수. 반응이 시행 세트에서 발생하는 많은 이벤트면 종속변수에는 이벤트 수가 포함되며 시행 수가 포함되어 있는 추가 변수를 선택할 수 있습니다. 또한 시행 수가 모든 개체에서 동일한 경우 고정 값을 사용하여 시행을 지정할 수 있습니다. 각 케이스에 대해 시행 수는 이벤트 수보다 크거나 같아야 합니다. 이벤트는 양의 정수가 되어야 하며 시행 수는 양의 정수가 되어야 합니다.

순서 다항 모형에 대해 응답의 범주 순서를 오름차순, 내림차순 또는 데이터로 지정할 수 있습니다(데이터 순서란 데이터에 있는 첫 번째 값이 첫 번째 범주를 정의하고 마지막 값은 마지막 범주를 정의합니다)

척도 가중치. 척도 모수는 반응의 변수와 관련한 추정된 모형 모수입니다. 척도 가중치는 "알려진" 값으로, 관측할 때마다 달라질 수 있습니다. 척도 가중 변수를 지정한 경우 반응의 변수와 관련한 척도 모수는 각 관측에 대해 나눈 것입니다. 0보다 작거나 같고 또는 값이 없는 척도 가중치를 가진 케이스는 분석에 사용되지 않습니다.

일반화 선형 모형 참조 범주

이분형 반응의 경우 종속변수에 대한 참조 범주를 선택할 수 있습니다. 이것은 모수 추정값과 저장된 값 같은 특정 결과에 영향을 미칠 수 있지만 모형 적합을 변경해서는 안 됩니다. 예를 들어, 이분형 반응이 0과 1의 값을 사용하는 경우:

- 기본적으로 프로시저는 마지막(가장 높은 값) 범주 또는 참조 범주인 1을 만듭니다. 이 상황에서 모형 저장된 확률은 주어진 케이스가 값 0을 사용하는 변화를 추정하고 모수 추정값은 범주 0의 우도와 관련하여 해석해야 합니다.
- 첫 번째(가장 낮은 값) 범주 또는 참조 범주로 0을 지정하는 경우 모형 저장된 확률은 주어진 케이스가 값 1을 사용하는 변화를 추정합니다.
- 사용자 정의 범주를 지정하고 변수에 정의된 레이블이 있는 경우 목록에서 값을 선택하여 참조 범주를 설정할 수 있습니다. 이는 모형을 지정하는 동안 정확히 특정 변수를 어떻게 코딩했는지 기억이 나지 않을 때 편리할 수 있습니다.

일반화 선형 모형 예측변수

예측변수 탭을 사용하면 모형 효과를 작성하고 옵션 오프셋을 지정하는 데 사용되는 요인과 공변량을 지정할 수 있습니다.

요인. 요인은 범주형 예측변수로 숫자 또는 문자가 될 수 있습니다.

공변량. 공변량은 척도 예측변수로 숫자가 되어야 합니다.

참고: 반응이 이분형 형식을 갖는 이항이면 프로시저는 선택된 요인과 공변량을 관측한 값의 교차 분류를 기준으로 하는 부모집단에 의해 편차와 카이제곱 접합도 통계를 계산합니다. 일관성 있는 수의 부모집단을 보장하기 위해 프로시저의 여러 실행에서 동일한 세트의 예측변수를 유지해야 합니다.

오프셋. 오프셋 항은 "구조" 예측변수입니다. 계수는 모형으로 추정되지 않지만 값 1을 갖는 것으로 가정합니다. 따라서 오프셋 값은 단순히 대상의 선형 예측변수에 추가됩니다. 이는 각 케이스가 중요 이벤트마다 다른 노출 수준을 가질 수 있는 포아송 회귀 모형에 특히 유용합니다.

예를 들어, 개별 운전자의 사고 비율을 모델링할 때 3년 경력 중 한 번의 사고를 낸 운전자와 25년 경력 중에 한 번 사고를 낸 운전자 사이에는 중요한 차이가 있습니다! 사고 수는 운전 경력의 자연 로그가 오프셋 항으로 포함되는 경우 로그 링크와 함께 포아송 또는 음수 이항 반응으로 모델링할 수 있습니다.

다른 분포와 링크 유형의 조합은 오프셋 변수의 다른 변환이 필요할 수 있습니다.

일반화 선형 모형 옵션

이러한 옵션은 예측변수 탭에 지정한 모든 요인에 적용됩니다.

사용자 결측값. 케이스가 분석에 포함되려면 요인은 유효한 값을 가져야 합니다. 이 제어를 사용하면 요인변수 사이에서 사용자 결측값을 유효한 데이터로 취급할지 여부를 결정할 수 있습니다.

범주 순서. 이것은 요인의 마지막 수준을 결정하는 것과 관련이 있으며 추정 알고리즘에서 중복 모수와 관련이 있을 수 있습니다. 범주 순서를 변경하면 이 모수 추정값이 "마지막" 수준과 관련하여 계산되기 때문에 요인 수준 효과 값을 변경할 수 있습니다. 요인은 가장 낮은 값에서 가장 높은 값으로 오름차순으로, 가장 높은 값에서 가장 낮은 값으로 또는 "데이터 순서"로 저장될 수 있습니다. 즉, 데이터에서 첫 번째 값은 첫 번째 범주를 정의하고 마지막 고유 값은 마지막 범주를 정의합니다.

일반화 선형 모형 모형

모형 효과 지정. 기본 모형은 절편 전용이므로 다른 모형 효과를 명시적으로 지정해야 합니다. 또는 중첩되거나 중첩되지 않은 항을 설정할 수 있습니다.

중첩되지 않은 항

선택된 요인과 공변량에 대해 다음과 같습니다.

주효과. 선택한 각 변수에 대한 주효과 항을 생성합니다.

상호작용. 선택한 모든 변수의 최고 수준 상호작용 항을 생성합니다.

요인. 선택한 변수에 대해 가능한 상호작용 및 주효과를 모두 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.

중첩 항

이 프로시저에서는 모형에 대해 중첩 항을 설정할 수 있습니다. 중첩 항은 요인 또는 공변량의 효과를 모델링하는 데 유용합니다. 이들 값은 다른 요인 수준과 상호작용하지 않습니다. 예를 들어, 식료품 체인점은 여러 점포에서의 고객의 소비 성향을 살펴볼 수 있습니다. 각 고객은 체인점 중의 한 곳만 자주 가기 때문에 고객 효과는 점포 효과 내에 중첩되었다고 할 수 있습니다.

또한 같은 공변량을 포함하고 있는 다항 항 같이 상호작용 효과를 포함시키거나 여러 수준의 중첩을 중첩 항에 추가할 수 있습니다.

제한. 중첩 항에는 다음과 같은 제한이 있습니다.

- 상호작용 내의 모든 요인은 고유해야 합니다. 따라서 A가 요인이면 A*A 지정은 유효하지 않습니다.
- 중첩 효과 내의 모든 요인은 고유해야 합니다. 따라서 A가 요인이면 A(A) 지정은 유효하지 않습니다.
- 공변량 내에 효과를 중첩할 수 없습니다. 따라서 A가 요인이고 X가 공변량이면 A 지정은 유효하지 않습니다.

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

다항 순서 분포가 있는 모형에는 단일 절편 항이 없습니다. 대신 인접 범주 사이의 전환점을 정의하는 임계값 모수가 있습니다. 임계값은 모형에 항상 포함됩니다.

일반화 선형 모형 추정

모수 추정. 이 그룹의 제어를 사용하면 추정 방법을 지정하고 모수 추정값에 대한 초기값을 제공할 수 있습니다.

- **방법.** 모수 추정 방법을 선택할 수 있습니다. Newton-Raphson, Fisher 스코어링 또는 Fisher 스코어링 반복이 Newton-Raphson 방법으로 전환하기 전에 수행되는 하이브리드 방법 중에서 선택합니다. 하이브리드 방법의 Fisher 스코어링 단계 동안 Fisher 반복의 최대 수에 도달하기 전에 수렴이 얻어진 경우 알고리즘은 Newton-Raphson 방법으로 계속됩니다.

- **척도 모수 방법.** 척도 모수 추정 방법을 선택할 수 있습니다. 최대 우도는 모형 효과와 함께 척도 모수를 추정합니다. 이 옵션은 반응에 음수 이항, 포아송, 이항 또는 다항 분포가 있는 경우 유효하지 않습니다. 편차와 Pearson 카이제곱 옵션은 이러한 통계값에서 척도 모수를 추정합니다. 또한 척도 모수에 대한 고정 값을 지정할 수 있습니다.
- **초기값.** 프로시저는 모수에 대한 초기값을 자동으로 계산합니다. 또한 모수 추정값에 대한 초기값을 지정할 수 있습니다.
- **공분산 행렬.** 모형 기반 추정량은 Hessian 행렬의 일반화 역의 음수입니다. 동질성(Huber/White/sandwich 라고도 함) 추정량은 변수와 연결 함수의 지정이 잘못된 경우에도 공분산의 일관성 있는 추정을 제공하는 "수정된" 모형 기반 추정량입니다.

반복. 다음 옵션을 사용할 수 있습니다.

- **최대반복수.** 알고리즘에서 실행할 반복의 최대 횟수입니다. 음수가 아닌 정수를 지정합니다.
- **최대 단계-반분.** 각 반복에서 단계 크기는 로그 우도 증가 또는 최대 단계 반분에 도달할 때까지 요인이 0.5씩 감소됩니다. 양의 정수를 지정합니다.
- **데이터 점의 분리 확인.** 이 옵션을 선택하면 모수 추정값이 중복되지 않았는지 확인하기 위한 검정이 수행됩니다. 모든 케이스를 올바르게 분류하는 모형을 프로시저에서 생성할 수 있는 경우에 분리가 발생합니다. 이 옵션은 다항 반응의 이항 반응 및 이분형 형식의 이항 반응에 사용할 수 있습니다.

수렴 기준 다음 옵션을 사용할 수 있습니다.

- **모수 수렴.** 이 옵션을 선택하면 모수 추정값의 절대 변화량 또는 상대 변화량이 지정된 값보다 크거나 같아 지는 시점에 반복을 멈춥니다. 지정된 값은 0 또는 양의 정수이어야 합니다.
- **로그-우도 수렴.** 이 옵션을 선택하면 로그-우도 함수의 절대 변화량 또는 상대 변화량이 지정된 값보다 작아 지는 시점의 계산 후에 반복을 멈춥니다. 지정된 값은 0 또는 양의 정수이어야 합니다.
- **Hessian 수렴.** 절대값 지정의 경우 수렴은 Hessian 수렴을 기준으로 하는 통계가 지정된 양의 값보다 작다고 가정합니다. 상대값 지정의 경우 수렴은 통계가 로그 우도의 절대값과 지정된 양의 값의 곱보다 작다고 가정합니다.

비정칙성 허용 오차. 비정칙(또는 비가역) 행렬에는 추정 알고리즘에 심각한 문제를 일으킬 수 있는 선형 종속 열이 있습니다. 거의 비정칙인 행렬은 잘못된 결과를 초래할 수 있으므로 프로시저는 행렬식이 허용 오차보다 작은 행렬은 비정칙으로 취급합니다. 양수값을 지정합니다.

일반화 선형 모형 초기값

초기값을 지정하는 경우 모형에 있는 모든 모수(중복 모수 포함)에 대해 제공해야 합니다. 데이터 세트에서 왼쪽에서 오른쪽으로 변수를 순서 지정할 때 다음과 같이 되어야 합니다. *RowType_*, *VarName_*, *P1*, *P2*, ..., where *RowType_* 및 *VarName_*은 문자열 변수이고 *P1*, *P2*, ...는 모수의 순서 목록에 해당하는 숫자 변수입니다.

- 초기값은 변수 *RowType_*의 경우 값 *EST*를 갖는 레코드에 제공되고 실제 초기값은 *P1*, *P2*, ... 아래 제공됩니다. 프로시저는 *RowType_*이 *EST* 이외의 값을 갖고 첫 번째 *RowType_*을 벗어나는 모든 레코드가 *EST*와 같은 모든 레코드를 무시합니다.

- 모형에 포함된 경우 절편 또는 반응에 다항 분포가 있는 경우 임계값 모수는 초기값을 먼저 나열해야 합니다.
- 척도 모수 및 반응에 음수 이항 분포가 있는 경우 음수 이항 모수는 초기값을 마지막에 지정해야 합니다.
- 분할 파일이 효과에 있는 경우 변수는 분할 파일을 만들 때 지정한 순서로 분할 파일 변수 또는 변수, 이어서 위와 같이 *RowType_*, *VarName_*, *P1*, *P2*, ... 등으로 시작해야 합니다. 분할은 원래 데이터 세트에서와 같은 순서로 지정한 데이터 세트에서 발생해야 합니다.

참고: 변수 이름 *P1*, *P2*, ... 는 필요하지 않습니다. 프로시저는 모수에 대한 변수 매핑이 변수 이름이 아닌 변수 위치를 기준으로 하기 때문에 모수의 유효한 변수 이름을 받아들입니다. 마지막 모수를 초과하는 변수는 무시됩니다.

초기값의 파일 구조는 모형을 데이터로 내보낼 때 사용한 구조와 같습니다. 따라서 프로시저에서 한 실행의 최종 값을 후속 실행의 입력으로 사용할 수 있습니다.

일반화 선형 모형 통계

모형 효과. 다음 옵션을 사용할 수 있습니다.

- 분석 유형. 생성할 분석 유형을 지정합니다. 제 I 유형 분석은 일반적으로 모형에서 순서 예측변수에 대한 사전 이유가 있을 때 적합한 반면 제 III 유형은 보다 일반적으로 적용됩니다. Wald 또는 우도비 통계는 카이제곱 통계 그룹에서 선택한 것을 기준으로 계산됩니다.
- 신뢰구간. 50보다 크고 100보다 작은 신뢰수준을 지정합니다. Wald 구간은 모수가 점근 정규 분포를 갖고 있다는 가정을 전제로 합니다. 프로파일 우도 구간은 더 정확하지만 계산 비용이 많이 소요될 수 있습니다. 프로파일 우도 구간의 허용 오차 수준은 구간을 계산하는 데 사용되는 반복 알고리즘을 중지하는 데 사용되는 기준입니다.
- 로그-우도 함수. 이 함수는 로그-우도 함수의 출력 형식을 제어합니다. 전체 함수에는 모수 추정에 관해 일관성 있는 추가 항이 포함되어 있습니다. 모수 추정에는 효과가 없으며 일부 소프트웨어 제품에서는 출력이 되지 않습니다.

인쇄. 다음과 같은 결과를 사용할 수 있습니다.

- 케이스 처리 요약. 분석에 포함되었거나 제외된 케이스 수와 백분율 및 상관 데이터 요약 테이블을 표시합니다.
- 기술통계. 종속변수, 공변량 및 요인에 대한 기술 통계와 요약 정보를 표시합니다.
- 모형 정보. 데이터 세트 이름, 종속변수 또는 이벤트 및 시행 변수, 오프셋 변수, 척도가중 변수, 확률 분포 및 연결 함수를 표시합니다.
- 적합도 통계량. 편차와 척도화된 편차, Pearson 카이제곱 및 척도화된 Pearson 카이제곱, 로그 우도, Akaike 정보기준(AIC), 유한 표본 수정된 AIC(AICC), 베이저안 정보 기준(BIC) 및 일관된 AIC(CAIC)를 표시합니다.
- 모형 요약 통계. 각 효과에 대한 제 I 유형 또는 제 III 유형 대비의 우도비 통계 및 모형 적합 총괄 검정을 포함한 모형 적합 검정을 표시합니다.

- **모수 추정값.** 모수 추정값과 해당 검정 통계량 및 신뢰구간을 표시합니다. 원래 모수 추정값 외에 선택적으로 누승 모수 추정을 표시할 수 있습니다.
- **모수 추정값의 공분산 행렬.** 추정된 모수 공분산 행렬이 표시됩니다.
- **모수 추정값의 상관 행렬.** 추정된 모수 상관 행렬이 표시됩니다.
- **대비계수(L) 행렬.** EM 평균 탭에서 요청하는 경우 기본 효과 및 주변 평균 추정에 대한 대비계수를 표시합니다.
- **일반 추정가능 함수.** 대비계수(L) 행렬을 생성하는 지표를 표시합니다.
- **반복 히스토리.** 모수 추정값과 로그 우도에 대한 반복 히스토리를 표시하고 기울기 벡터와 Hessian 행렬의 마지막 평가를 인쇄합니다. 반복 히스토리 테이블은 0^{번째} 반복(초기 추정값)부터 시작하여 매 n 번째 반복에 대한 모수 추정값을 표시합니다. 여기서 n 은 인쇄 구간 값입니다. 반복 히스토리를 요청하는 경우 n 에 관계 없이 마지막 반복은 항상 표시됩니다.
- **LM 검정.** 편차 또는 Pearson 카이제곱이나 정규, 감마, 역 가우스 및 Tweedie 분포에 대한 고정된 수의 세트를 사용하여 계산되는 척도 모수의 타당성을 평가하기 위한 LM 검정 통계량을 표시합니다. 음수 이항 분포의 경우 고정된 보조 모수를 검정합니다.

일반화 선형 모형 EM 평균

이 탭을 사용하면 요인 대 요인 상호작용의 수준에 대해 주변 평균 추정을 표시할 수 있습니다. 전체 추정 평균을 표시하도록 요청할 수도 있습니다. 주변 평균 추정은 순서 다항 모형에는 사용할 수 없습니다.

요인 및 상호 작용. 이 목록에는 예측변수 탭에 지정된 요인과 모형 탭에 지정된 요인 상호작용이 포함되어 있습니다. 공변량은 이 목록에서 제외됩니다. 이 목록에서 항을 직접 선택하거나 **By** * 단추를 사용하여 상호작용 항에 결합할 수 있습니다.

평균 출력 기준. 추정된 평균은 선택한 요인 및 요인 상호작용에 대해 계산됩니다. 대비는 추정된 평균을 비교하기 위해 가설 검정을 어떻게 설정할지 결정합니다. 단순대비에는 참조 범주 또는 다른 항을 비교하는 데 기준으로 삼을 요인 수준이 필요합니다.

- **대응별.** 쌍대 비교는 지정했거나 내재된 요인의 모든 수준 조합에 대해 계산됩니다. 이는 요인 상호작용의 대비에만 사용할 수 있습니다.
- **단순.** 각 수준의 평균을 지정한 수준의 평균과 비교합니다. 이 대비 유형은 대조군이 있는 경우 유용합니다.
- **편차.** 요인의 각 수준은 총 평균에 비교됩니다. 편차 대비는 직교 대비가 아닙니다.
- **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.

척도, 주변 평균 추정은 종속변수의 원래 척도를 기준으로 하는 반응 또는 연결 함수에 의해 변환되는 종속변수를 기준으로 하는 선형 예측변수에 대해 계산할 수 있습니다.

다중 비교에 대한 조정. 다중 대비를 통해 가설 검정을 수행하는 경우 포함된 대비에 대한 유의 수준에서 전반적인 유의 수준을 조정할 수 있습니다. 이 그룹을 사용하면 조정 방법을 선택할 수 있습니다.

- **최소유의차.** 이 방법은 특정 선형 대비가 귀무가설 값과 다르다는 가설을 거부할 전체 확률을 제어하지 않습니다.
- **Bonferroni.** 이 방법은 관측 유의 수준을 다중 대비가 검정되는 요인에 맞춥니다.
- **순차 Bonferroni(Sequential Bonferroni).** 순차 단계별로 낮아지는 거부 Bonferroni 프로시저로, 개별 가설은 거부하지만 동일한 전체 유의 수준을 유지한다는 점에서 훨씬 덜 보수적인 방법입니다.
- **Sidak.** 이 방법은 Bonferroni 접근법보다 더 엄격한 한계를 제공합니다.
- **순차 Sidak(Sequential Sidak).** 순차 단계별로 낮아지는 거부 Sidak 프로시저로, 개별 가설은 거부하지만 동일한 전체 유의 수준을 유지한다는 점에서 훨씬 덜 보수적인 방법입니다.

일반화 선형 모형 저장

확인된 항목은 지정된 이름으로 저장됩니다. 같은 이름을 가진 새 변수와 같은 이름을 가진 기존 변수를 덮어쓰도록 선택하여 새 변수 이름을 고유하게 만들 수 있습니다.

- **반응 평균의 예측값.** 원래 반응 매트릭에서 각 케이스에 대한 모형 예측값을 저장합니다. 반응 분포가 이항이고 종속변수가 이분형인 경우 프로시저는 예측 확률을 저장합니다. 반응 분포가 다항이면 항목 레이블은 누적 예측 확률이 되고 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 누적 예측 확률을 저장합니다.
- **반응 평균에 대한 신뢰구간의 하한.** 반응 평균에 대한 신뢰구간의 하한을 저장합니다. 반응 분포가 다항이면 항목 레이블은 누적 예측 확률에 대한 신뢰구간의 하한이 되고 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 하한을 저장합니다.
- **반응 평균에 대한 신뢰구간의 상한.** 반응 평균에 대한 신뢰구간의 상한을 저장합니다. 반응 분포가 다항이면 항목 레이블은 누적 예측 확률에 대한 신뢰구간의 상한이 되고 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 상한을 저장합니다.
- **예측 범주.** 이항 분포와 이분형 종속변수가 있는 모형 또는 다항 분포의 경우 각 케이스에 대한 예측 반응 범주를 저장합니다. 이 옵션은 다른 반응 분포에는 사용할 수 없습니다.
- **선형 예측변수의 예측값.** 선형 예측변수의 매트릭에서 각 케이스에 대한 모형 예측값을 저장합니다(지정한 연결 함수를 통해 변환된 반응). 반응 분포가 다항이면 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 예측값을 저장합니다.
- **선형 예측변수의 예측된 값의 예측된 표준 오차.** 반응 분포가 다항이면 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 예측 표준 오차를 저장합니다.

다음 항목은 반응 분포가 다항일 때 사용할 수 있습니다.

- *Cook*의 거리. 특정 케이스를 회귀계수 계산에서 제외할 때 모든 케이스의 잔차가 얼마나 변경될 수 있는지에 대한 측도입니다. *Cook*의 거리가 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.
- 원시 잔차. 관측값과 모형에 의해 예측된 값 사이 차이입니다.
- **Pearson** 잔차. 원시 잔차의 부호가 있는 **Pearson** 카이제곱 통계에 대한 케이스의 기여도 제공근.
- 표준화된 **Pearson** 잔차. 케이스에 대한 척도 모수 1-레버리지 곱의 역의 제공근을 곱한 **Pearson** 잔차.
- 편차 잔차. 원시 잔차의 부호가 있는 편차 통계에 대한 케이스의 기여도 제공근.
- 표준화된 편차 잔차. 케이스에 대한 척도 모수와 1-레버리지 곱의 역의 제공근을 곱한 편차 잔차입니다.
- 우도 잔차. 원시 잔차의 부호가 있는 표준화된 **Pearson** 및 표준화된 편차 잔차의 제공의 가중 평균의 제공근(케이스 평균을 기준).

일반화 선형 모형 내보내기

데이터로 모형 내보내기. 모수 추정값, 표준 오차, 유의수준 및 자유도와 함께 모수 상관 행렬이나 공분산 행렬이 포함되어 있는 IBM SPSS Statistics 형식의 데이터 세트를 씁니다. 행렬 파일에서 변수의 순서는 다음과 같습니다.

- 분할변수. 사용된 경우 분할을 정의하는 모든 변수.
- **RowType_**. 값(및 값 레이블) *COV* (공분산), *CORR* (상관), *EST* (모수 추정값), *SE* (표준 오차), *SIG* (유의 수준) 및 *DF* (표본추출 디자인 자유도)를 사용합니다. 각 모형 모수에 대해 행 유형 *COV* (또는 *CORR*)을 갖는 별도의 케이스는 물론 각기 다른 행 유형에 대해 별도의 케이스가 있습니다.
- **VarName_**. 모수 추정값 테이블에 표시된 모수 문자에 해당하는 값 레이블과 함께 행 유형 *COV* 또는 *CORR*에 대해 모든 추정 모형 모수(척도 또는 음수 이항 모수)의 순서 목록에 해당하는 값 *P1*, *P2*, ...,을 사용합니다. 다른 행 유형의 셀은 비어 있습니다.
- **P1, P2, ...** 이러한 변수는 모수 추정값 테이블에 표시된 모수 문자에 해당하는 변수 레이블과 함께 모든 모형 모수(해당되는 경우 척도와 음수 이항 모수를 포함)의 순서 목록에 해당하며 행 유형에 따라 값을 사용합니다.

중복 모수의 경우 모든 공분산은 0으로 설정되고 상관은 시스템 결측값으로 설정됩니다. 모든 모수 추정값은 0으로 설정되고 모든 표준 오차, 유의 수준 및 잔차의 자유도는 시스템 결측값으로 설정됩니다.

척도 모수의 경우 공분산, 상관, 유의 수준 및 자유도는 시스템 결측값으로 설정됩니다. 척도 모수를 최대 우도를 통해 추정할 경우 표준 오차가 제공되며, 그렇지 않으면 시스템 결측값으로 설정됩니다.

음수 이항의 경우 공분산, 상관, 유의 수준 및 자유도는 시스템 결측값으로 설정됩니다. 음수 이항 모수를 최대 우도를 통해 추정할 경우 표준 오차가 제공되며, 그렇지 않으면 시스템 결측값으로 설정됩니다.

분할이 있는 경우에는 모수 목록을 모든 분할에 대해 누적해야 합니다. 해당 분할에서 일부 모수는 관련이 없을 수 있으며 이는 중복과 같지 않습니다. 관련이 없는 모수의 경우 모든 공분산이나 상관, 모수 추정값, 표준 오차, 유의 수준 및 자유도는 시스템 결측값으로 설정됩니다.

이 행렬 파일을 추가 모형 추정을 위한 초기값으로 사용할 수 있습니다. 이러한 프로시저가 여기에 내보낸 모든 행 유형을 받아 들이지 않는 경우 이 파일은 다른 프로시저에서 추가 분석에 즉시 사용할 수 없습니다. 그러면 이 행렬 파일에 있는 모든 모수가 파일을 읽는 프로시저에 같은 의미를 가지므로 주의해야 합니다.

XML로 모형 내보내기. 모수 추정값과 모수 공분산 행렬(선택한 경우)을 XML(PMML) 형식으로 저장합니다. 스코어링을 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.

GENLIN 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 모수 추정값의 초기값을 숫자 목록으로 지정합니다(CRITERIA 하위 명령 사용).
- EMMEANS 하위 명령을 사용하여 추정된 주변 평균을 계산할 때 해당 평균 이외의 다른 값으로 공변량을 고정할 수 있습니다.
- 주변 평균 추정의 사용자 정의 다항 대비를 지정합니다(EMMEANS 하위 명령 사용).
- 지정된 대비 유형을 사용하는 것과 비교하여 주변 평균 추정이 표시될 요인의 서브세트를 지정합니다(EMMEANS 하위 명령의 TABLES 및 COMPARE 키워드 사용).

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

제 7 장 일반화 추정 방정식

일반화된 추정 방정식 프로시저는 일반화 선형 모형을 확장하여 반복 측정이나 군집된 데이터 같은 기타 상호 관련이 있는 관측을 분석할 수 있습니다.

예제. 공공 의료 담당자는 일반화된 추정 방정식을 사용하여 공기 오염이 아이들에게 미치는 영향을 연구하기 위해 반복 측도 로지스틱 회귀분석을 맞출 수 있습니다.

일반화된 추정 방정식 데이터 고려 사항

데이터. 반응은 척도, 개수, 이분형 또는 시행한 이벤트가 될 수 있습니다. 요인은 범주형으로 가정됩니다. 공변량, 척도 가중치 및 오프셋은 척도로 가정됩니다. 개체 또는 개체-내 반복 측정을 정의하는 데 사용되는 변수는 반응을 정의하는 데 사용할 수 없지만 모형에서 다른 역할을 수행할 수 있습니다.

가정. 케이스는 개체 내에 종속되고 개체 간에 독립되는 것으로 간주됩니다. 상관 행렬은 개체-내 종속성이 모형의 일부로 추정되는 것을 나타냅니다.

일반화된 추정 방정식 얻기

메뉴에서 다음을 선택합니다.

분석 > 일반화 선형 모형 > 일반화 추정 방정식...

1. 하나 이상이 개체변수를 선택합니다(자세한 옵션은 아래 참조).

지정된 변수의 값의 조합은 데이터 세트 내에서 개체를 고유하게 정의해야 합니다. 예를 들어, 단일 환자 ID 변수는 단일 병원에서 개체를 정의하기에 충분해야 하지만 병원 ID와 환자 ID는 환자 식별 번호가 병원마다 고유하지 않은 경우 필요할 수 있습니다. 반복 측도 설정에서 각 개체마다 여러 관측을 기록하므로 각 개체는 데이터 세트에서 여러 케이스를 차지할 수 있습니다.

2. 모형 유형 탭에서 분포와 연결 함수를 지정합니다.

3. 반응 탭에서 종속변수를 선택합니다.

4. 예측변수 탭에서 종속변수를 예측하는 데 사용할 요인과 공변량을 선택합니다.

5. 모형 탭에서 선택한 요인과 공변량을 사용하여 모형 효과를 지정합니다.

선택적으로 반복 탭에서 다음을 지정할 수 있습니다.

개체-내 변수. 개체-내 변수의 값의 조합은 개체 내의 측정 순서를 정의하므로 개체-내 및 개체변수의 조합은 각 측정을 고유하게 정의합니다. 예를 들어, 기간, 병원 ID 및 환자 ID의 조합은 각 케이스마다 특정 병원 내의 특정 환자에 대한 특정 사무실 방문을 정의합니다.

데이터 세트가 이미 정렬되어 각 개체의 반복 측정이 케이스의 연속 블록에서 발생하는 경우 반드시 개체-내 변수를 지정할 필요는 없으며 개체 및 개체내 변수별로 케이스 정렬을 선택 해제하고 (임시) 정렬을 수행하는 데 필요한 처리 시간을 절약할 수 있습니다. 일반적으로 적절한 측정 순서를 보장하려면 개체-내 변수를 사용하는 것이 좋습니다.

개체 및 개체-내 변수는 반응을 정의하는 데 사용할 수 없지만 모형에서 다른 기능은 수행할 수 있습니다. 예를 들어, 병원 ID는 모형에서 요인으로 사용할 수 있습니다.

공분산 행렬. 모형 기반 추정량은 Hessian 행렬의 일반화 역의 음수입니다. 동질성 추정량(Huber/White/sandwich 추정량이라고도 함)은 작업 중인 상관 행렬을 잘못 지정한 경우에도 공분산의 일관성 있는 추정을 제공하는 "수정된" 모형 기반 추정량입니다. 이 지정은 일반화된 추정 방정식의 선형 모형 일부에 있는 모수에 적용되지만 추정 탭에 있는 지정은 초기의 일반화 선형 모형에만 적용됩니다.

작업 중인 상관 행렬. 이 상관 행렬은 개체-내 종속성을 나타냅니다. 그 크기는 측정 수, 즉 개체-내 변수 값의 조합에 따라 결정됩니다. 다음 구조 중 하나를 지정할 수 있습니다.

- **독립적.** 반복 측정은 상관성이 없습니다.
- **AR(1).** 반복 측정은 1차 자기회귀 관계를 갖습니다. 두 요소 사이의 상관은 인접 요소의 경우 r 이며, 세 번째 요소로 분리된 두 요소의 경우 r^2 등의 방식으로 적용됩니다. $-1 << 1$ 로 제한됩니다.
- **교환 가능.** 이 구조는 요소 간의 동질적상관입니다. 또한 복합 대칭 구조라고도 합니다.
- **M-종속.** 연속 측정은 공통 상관계수를 가지며 제3자로 분리된 측정 쌍은 공통 상관계수를 가지며 이런 형식으로 계속 적용되어 $m-1$ 기타 측정으로 분리된 측정 쌍을 가집니다. 예를 들어, 3학년부터 7학년까지 학생들을 대상으로 매년 표준화 검사를 수행합니다. 이 구조에서는 3학년과 4학년, 4학년과 5학년, 5학년과 6학년 및 6학년과 7학년 점수의 상관이 동일하며, 3학년과 5학년, 4학년과 6학년 및 5학년과 7학년의 상관이 동일하고, 3학년과 6학년, 4학년과 7학년의 상관이 동일하다고 가정합니다. m 보다 큰 분리를 사용한 측정은 상관되지 않는다고 가정합니다. 이 구조를 선택할 때는 작업 중인 상관 행렬의 순서보다 m 작은 값을 지정합니다.
- **비구조적.** 일반 상관 행렬입니다.

기본적으로 프로시저는 중복되지 않은 모수의 수로 상관 추정값을 조정합니다. 데이터에서 개체 수준 복제 변경에 따라 추정량이 변하지 않도록 하려면 이 조정을 제거하는 것이 바람직할 수 있습니다.

- **최대반복수.** 일반화된 추정 방정식 알고리즘에서 실행할 반복의 최대 횟수입니다. 음수가 아닌 정수를 지정합니다. 이 지정은 일반화된 추정 방정식의 선형 모형 일부에 있는 모수에 적용되지만 추정 탭에 있는 지정은 초기의 일반화 선형 모형에만 적용됩니다.
- **행렬 업데이트.** 작업 중인 상관 행렬의 요소는 알고리즘을 반복할 때마다 업데이트되는 모수 추정값을 기준으로 추정됩니다. 작업 중인 상관 행렬이 전혀 업데이트되지 않을 경우 초기 작업 중인 상관 행렬은 추정 프로세스 전반에 걸쳐 사용됩니다. 행렬이 업데이트되는 경우 작업 중인 상관 행렬 요소를 업데이트할 반복 구간을 지정할 수 있습니다. 1보다 큰 값을 지정하면 처리 시간을 줄일 수 있습니다.

수렴 기준. 이 지정은 일반화된 추정 방정식의 선형 모형 일부에 있는 모수에 적용되지만 추정 탭에 있는 지정은 초기의 일반화 선형 모형에만 적용됩니다.

- **모수 수렴.** 이 옵션을 선택하면 모수 추정값의 절대 변화량 또는 상대 변화량이 지정된 값보다 크거나 같아 지는 시점에 반복을 멈춥니다. 지정된 값은 0 또는 양의 정수이어야 합니다.
- **Hessian 수렴.** 수렴은 Hessian을 기준으로 하는 통계가 지정된 값보다 작은 경우 양수여야 한다고 가정합니다.

일반화된 추정 방정식 모형 유형

모형 유형을 사용하면 모형에 대한 분포 및 연결 함수를 지정하여 반응 유형별로 범주화된 여러 공통 모형에 대한 바로 가기를 제공할 수 있습니다.

모형 유형

척도 반응. 다음 옵션을 사용할 수 있습니다.

- **선형.** 정규를 분포로, 항등을 연결 함수로 지정합니다.
- **로그 링크가 있는 감마.** 감마를 분포로, 로그를 연결 함수로 지정합니다.

순서 반응. 다음 옵션을 사용할 수 있습니다.

- **순서 로지스틱.** 다항(순서)을 분포로, 누적 로짓을 연결 함수로 지정합니다.
- **순서 프로빗.** 다항(순서)을 분포로, 누적 프로빗을 연결 함수로 지정합니다.

개수. 다음 옵션을 사용할 수 있습니다.

- **포아송 로그선형.** 포아송을 분포로, 로그를 연결 함수로 지정합니다.
- **로그 링크가 있는 음수 이항.** 음수 이항(보조 모수에 대해 값 1을 가짐)을 분포로, 로그를 연결 함수로 지정합니다. 프로시저 추정값이 보조 모수의 값을 갖도록 하려면 음수 이항 분포를 가진 사용자 정의 모형을 지정하고 모수 그룹에서 추정값을 선택합니다.

이분형 반응 또는 이벤트/시행 데이터. 다음 옵션을 사용할 수 있습니다.

- **이분형 로지스틱.** 이항을 분포로, 로짓을 연결 함수로 지정합니다.
- **이분형 프로빗.** 이항을 분포로, 프로빗을 연결 함수로 지정합니다.
- **구간 중도절단 생존.** 이항을 분포로, 보완 로그-로그를 연결 함수로 지정합니다.

혼합. 다음 옵션을 사용할 수 있습니다.

- **로그 링크가 있는 Tweedie.** Tweedie를 분포로, 로그를 연결 함수로 지정합니다.
- **항등 링크가 있는 Tweedie.** Tweedie를 분포로, 항등을 연결 함수로 지정합니다.

사용자 정의. 분포와 연결 함수의 자체 조합을 지정합니다.

분포

이 선택은 종속변수의 분포를 지정합니다. 비정규 분포와 항등하지 않은 연결 함수를 지정하는 기능은 일반 선형 모형에서 일반화 선형 모형의 중요한 개선 사항입니다. 많은 분포-연결 함수가 있으며 주어진 데이터 세트에 적합한 함수가 여러 개일 수 있으므로 사전 이론적 고려 사항을 바탕으로 선택하거나 어느 조합이 가장 적합할지를 고려하여 선택하면 됩니다.

- **이항검정.** 이 분포는 이분형 반응이나 이벤트 수를 나타내는 변수의 경우에만 적합합니다.
- **감마.** 이 분포는 더 큰 양의 값 쪽으로 비대칭되는 양의 척도 값을 가진 변수에 적합합니다. 데이터 값이 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- **역 가우스.** 이 분포는 더 큰 양의 값 쪽으로 비대칭되는 양의 척도 값을 가진 변수에 적합합니다. 데이터 값이 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- **음수 이항.** 이 분포는 k 성공을 관측하는 데 필요한 시행 횟수로 생각할 수 있으며 양의 정수 값을 갖는 변수에 적합합니다. 데이터 값이 양수이거나, 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다. 음수 이항 분포의 보조 모수 값은 0보다 크거나 같은 숫자가 될 수 있습니다. 이 값을 고정 값으로 설정하거나 프로시저에서 추정하도록 할 수 있습니다. 보조 모수가 0으로 설정되면 분포를 사용하는 것은 포아송 분포를 사용하는 것과 동일합니다.
- **정상.** 이것은 중앙(평균) 값에 대해 값이 대칭되는 종 형태의 분포를 띠는 척도변수에 적합합니다. 종속변수는 숫자여야 합니다.
- **포아송.** 이 분포는 고정 기간 동안 중요 이벤트의 발생 수로 생각할 수 있으며 양의 정수 값을 갖는 변수에 적합합니다. 데이터 값이 양수이거나, 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- **Tweedie.** 이 분포는 감마 분포의 포아송 혼합으로 표현할 수 있는 변수에 적합합니다. 분포는 연속 특성(음이 아닌 실수 값 사용)과 이산형 분포(단일 값 0에서 양의 확률 매스)의 관점에서 "혼합"된 것입니다. 종속변수는 데이터 값이 0보다 크거나 같은 숫자가 되어야 합니다. 데이터 값이 0보다 작거나 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다. Tweedie 분포에서 모수의 고정 값은 1보다 크고 2보다 작은 숫자가 될 수 있습니다.
- **다항.** 이 분포는 순서 반응을 나타내는 변수에 적합합니다. 종속변수는 숫자나 문자열이 될 수 있으며 최소 두 개의 유효한 개별 데이터 값을 가져야 합니다.

연결 함수

연결 함수는 종속변수의 변환으로 모형을 추정할 수 있습니다. 다음 함수를 사용할 수 있습니다.

- **항등.** $f(x)=x$. 종속변수가 변환되지 않습니다. 이 링크는 분포에 사용할 수 있습니다.
- **보 로그-로그.** $f(x)=-\log(1-x)$. 이항 분포에만 적합합니다.
- **누적 Cauchit.** $f(x) = \tan(\pi (x - 0.5))$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.
- **누적 보 로그-로그.** $f(x)=-\ln(1-x)$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.
- **누적 로짓.** $f(x)=\ln(x / (1-x))$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.

- 누적 음수 로그-로그. $f(x)=-\ln(-\ln(x))$, 반응에서 각 범주의 누적 확률에 적용됩니다. 다항 분포에만 적합합니다.
- 누적 프로빗. $f(x)=\Phi^{-1}(x)$, 반응에서 각 범주의 누적 확률에 적용됩니다. 여기서 Φ^{-1} 은 표준 정규 누적 분포의 역함수입니다. 다항 분포에만 적합합니다.
- 로그. $f(x)=\log(x)$. 이 링크는 분포에 사용할 수 있습니다.
- 로그 보. $f(x)=\log(1-x)$. 이항 분포에만 적합합니다.
- 로짓. $f(x)=\log(x / (1-x))$. 이항 분포에만 적합합니다.
- 음수 이항. $f(x)=\log(x / (x+k^{-1}))$, 여기서 k 는 음수 이항 분포의 보조 모수입니다. 음수 이항 분포에만 적합합니다.
- 음수 로그-로그. $f(x)=-\log(-\log(x))$. 이항 분포에만 적합합니다.
- 승산비. $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$, $\alpha \neq 0$ 인 경우. $f(x)=\log(x)$, $\alpha=0$ 인 경우. α 는 필요한 숫자 지정이며 실수가 되어야 합니다. 이항 분포에만 적합합니다.
- 프로빗. $f(x)=\Phi^{-1}(x)$, 여기서 Φ^{-1} 은 표준 정규 누적 분포의 역함수입니다. 이항 분포에만 적합합니다.
- 거듭제곱. $f(x)=x^\alpha$, $\alpha \neq 0$ 인 경우. $f(x)=\log(x)$, $\alpha=0$ 인 경우. α 는 필요한 숫자 지정이며 실수가 되어야 합니다. 이 링크는 분포에 사용할 수 있습니다.

일반화된 추정 방정식 반응

많은 경우에 종속변수를 간단히 지정할 수 있지만 두 값만 사용하고 시행에 이벤트를 기록하는 변수에 주의를 기울여야 합니다.

- 이분형 반응. 종속변수가 두 값만 사용할 때는 모수 추정에 대해 참조 범주를 지정할 수 있습니다. 이분형 반응변수는 문자 또는 숫자가 될 수 있습니다.
- 시행 세트에서 발생하는 이벤트의 수. 반응이 시행 세트에서 발생하는 많은 이벤트면 종속변수에는 이벤트 수가 포함되며 시행 수가 포함되어 있는 추가 변수를 선택할 수 있습니다. 또한 시행 수가 모든 개체에서 동일한 경우 고정 값을 사용하여 시행을 지정할 수 있습니다. 각 케이스에 대해 시행 수는 이벤트 수보다 크거나 같아야 합니다. 이벤트는 양의 정수가 되어야 하며 시행 수는 양의 정수가 되어야 합니다.

순서 다항 모형에 대해 응답의 범주 순서를 오름차순, 내림차순 또는 데이터로 지정할 수 있습니다(데이터 순서란 데이터에 있는 첫 번째 값이 첫 번째 범주를 정의하고 마지막 값은 마지막 범주를 정의합니다)

척도 가중치. 척도 모수는 반응의 변수와 관련한 추정된 모형 모수입니다. 척도 가중치는 "알려진" 값으로, 관측할 때마다 달라질 수 있습니다. 척도 가중 변수를 지정한 경우 반응의 변수와 관련한 척도 모수는 각 관측에 대해 나눈 것입니다. 0보다 작거나 같고 또는 값이 없는 척도 가중치를 가진 케이스는 분석에 사용되지 않습니다.

일반화된 추정 방정식 참조 범주

이분형 반응의 경우 종속변수에 대한 참조 범주를 선택할 수 있습니다. 이것은 모수 추정값과 저장된 값 같은 특정 결과에 영향을 미칠 수 있지만 모형 적합을 변경해서는 안 됩니다. 예를 들어, 이분형 반응이 0과 1의 값을 사용하는 경우:

- 기본적으로 프로시저는 마지막(가장 높은 값) 범주 또는 참조 범주인 1을 만듭니다. 이 상황에서 모형 저장된 확률은 주어진 케이스가 값 0을 사용하는 변화를 추정하고 모수 추정값은 범주 0의 우도와 관련하여 해석해야 합니다.
- 첫 번째(가장 낮은 값) 범주 또는 참조 범주로 0을 지정하는 경우 모형 저장된 확률은 주어진 케이스가 값 1을 사용하는 변화를 추정합니다.
- 사용자 정의 범주를 지정하고 변수에 정의된 레이블이 있는 경우 목록에서 값을 선택하여 참조 범주를 설정할 수 있습니다. 이는 모형을 지정하는 동안 정확히 특정 변수를 어떻게 코딩했는지 기억이 나지 않을 때 편리할 수 있습니다.

일반화된 추정 방정식 예측변수

예측변수 탭을 사용하면 모형 효과를 작성하고 옵션 오프셋을 지정하는 데 사용되는 요인과 공변량을 지정할 수 있습니다.

요인. 요인은 범주형 예측변수로 숫자 또는 문자가 될 수 있습니다.

공변량. 공변량은 척도 예측변수로 숫자가 되어야 합니다.

참고: 반응이 이분형 형식을 갖는 이항이면 프로시저는 선택된 요인과 공변량을 관측한 값의 교차 분류를 기준으로 하는 부모집단에 의해 편차와 카이제곱 접합도 통계를 계산합니다. 일관성 있는 수의 부모집단을 보장하기 위해 프로시저의 여러 실행에서 동일한 세트의 예측변수를 유지해야 합니다.

오프셋. 오프셋 항은 "구조" 예측변수입니다. 계수는 모형으로 추정되지 않지만 값 1을 갖는 것으로 가정합니다. 따라서 오프셋 값은 단순히 대상의 선형 예측변수에 추가됩니다. 이는 각 케이스가 중요 이벤트마다 다른 노출 수준을 가질 수 있는 포아송 회귀 모형에 특히 유용합니다.

예를 들어, 개별 운전자의 사고 비율을 모델링할 때 3년 경력 중 한 번의 사고를 낸 운전자와 25년 경력 중에 한 번 사고를 낸 운전자 사이에는 중요한 차이가 있습니다! 사고 수는 운전 경력의 자연 로그가 오프셋 항으로 포함되는 경우 로그 링크와 함께 포아송 또는 음수 이항 반응으로 모델링할 수 있습니다.

다른 분포와 링크 유형의 조합은 오프셋 변수의 다른 변환이 필요할 수 있습니다.

일반화된 추정 방정식 옵션

이러한 옵션은 예측변수 탭에 지정한 모든 요인에 적용됩니다.

사용자 결측값. 케이스가 분석에 포함되려면 요인은 유효한 값을 가져야 합니다. 이 제어를 사용하면 요인변수 사이에서 사용자 결측값을 유효한 데이터로 취급할지 여부를 결정할 수 있습니다.

범주 순서. 이것은 요인의 마지막 수준을 결정하는 것과 관련이 있으며 추정 알고리즘에서 중복 모수와 관련이 있을 수 있습니다. 범주 순서를 변경하면 이 모수 추정값이 "마지막" 수준과 관련하여 계산되기 때문에 요인 수준 효과 값을 변경할 수 있습니다. 요인은 가장 낮은 값에서 가장 높은 값으로 오름차순으로, 가장 높은 값에서 가장 낮은 값으로 또는 "데이터 순서"로 저장될 수 있습니다. 즉, 데이터에서 첫 번째 값은 첫 번째 범주를 정의하고 마지막 고유 값은 마지막 범주를 정의합니다.

일반화된 추정 방정식 모형

모형 효과 지정. 기본 모형은 절편 전용이므로 다른 모형 효과를 명시적으로 지정해야 합니다. 또는 중첩되거나 중첩되지 않은 항을 설정할 수 있습니다.

중첩되지 않은 항

선택된 요인과 공변량에 대해 다음과 같습니다.

주효과. 선택한 각 변수에 대한 주효과 항을 생성합니다.

상호작용. 선택한 모든 변수의 최고 수준 상호작용 항을 생성합니다.

요인. 선택한 변수에 대해 가능한 상호작용 및 주효과를 모두 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.

중첩 항

이 프로시저에서는 모형에 대해 중첩 항을 설정할 수 있습니다. 중첩 항은 요인 또는 공변량의 효과를 모델링하는 데 유용합니다. 이들 값은 다른 요인 수준과 상호작용하지 않습니다. 예를 들어, 식료품 체인점은 여러 점포에서의 고객의 소비 성향을 살펴볼 수 있습니다. 각 고객은 체인점 중의 한 곳만 자주 가기 때문에 고객 효과는 점포 효과 내에 중첩되었다고 할 수 있습니다.

또한 상호작용 효과를 포함시키거나 여러 수준의 중첩을 중첩 항에 추가할 수 있습니다.

제한. 중첩 항에는 다음과 같은 제한이 있습니다.

- 상호작용 내의 모든 요인은 고유해야 합니다. 따라서 A 가 요인이면 $A*A$ 지정은 유효하지 않습니다.
- 중첩 효과 내의 모든 요인은 고유해야 합니다. 따라서 A 가 요인이면 $A(A)$ 지정은 유효하지 않습니다.
- 공변량 내에 효과를 중첩할 수 없습니다. 따라서 A 가 요인이고 X 가 공변량이면 A 지정은 유효하지 않습니다.

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

다항 순서 분포가 있는 모형에는 단일 절편 항이 없습니다. 대신 인접 범주 사이의 전환점을 정의하는 임계값 모수가 있습니다. 임계값은 모형에 항상 포함됩니다.

일반화된 추정 방정식 추정

모수 추정. 이 그룹의 제어를 사용하면 추정 방법을 지정하고 모수 추정값에 대한 초기값을 제공할 수 있습니다.

- **방법.** 모수 추정 방법을 선택할 수 있습니다. Newton-Raphson, Fisher 스코어링 또는 Fisher 스코어링 반복이 Newton-Raphson 방법으로 전환하기 전에 수행되는 하이브리드 방법 중에서 선택합니다. 하이브리드 방법의 Fisher 스코어링 단계 동안 Fisher 반복의 최대 수에 도달하기 전에 수렴이 얻어진 경우 알고리즘은 Newton-Raphson 방법으로 계속됩니다.
- **척도 모수 방법.** 척도 모수 추정 방법을 선택할 수 있습니다.

최대 우도는 모형 효과와 함께 척도 모수를 추정합니다. 이 옵션은 반응에 음수 이항, 포아송 또는 이항 분포가 있는 경우 유효하지 않습니다. 우도의 개념은 일반화된 추정 방정식으로 입력되지 않기 때문에 이 지정은 초기 일반화 선형 모형에만 적용됩니다. 이 척도 모수 추정값은 자유도로 나눈 Pearson 카이제곱에 의해 척도 모수를 업데이트하는 일반화된 추정 방정식으로 전달됩니다.

편차와 TPearson 카이제곱 옵션은 초기 일반화 선형 모형의 통계 값에서 척도 모수를 추정합니다. 이 척도 모수 추정값은 고정 값으로 취급되는 일반화된 추정 방정식으로 전달됩니다.

또한 척도 모수에 대한 고정 값을 지정합니다. 초기 일반화 선형 모형과 일반화된 추정 방정식을 추정할 때 고정 값으로 간주됩니다.

- **초기값.** 프로시저는 모수에 대한 초기값을 자동으로 계산합니다. 또한 모수 추정값에 대한 초기값을 지정할 수 있습니다.

이 탭에 지정된 반복과 수렴 기준은 초기 일반화 선형 모형에만 적용됩니다. 일반화된 추정 방정식을 적합시키는 데 사용되는 추정 기준은 반복 탭을 참조하십시오.

반복. 다음 옵션을 사용할 수 있습니다.

- **최대반복수.** 알고리즘에서 실행할 반복의 최대 횟수입니다. 음수가 아닌 정수를 지정합니다.
- **최대 단계-반분.** 각 반복에서 단계 크기는 로그 우도 증가 또는 최대 단계 반분에 도달할 때까지 요인이 0.5씩 감소됩니다. 양의 정수를 지정합니다.
- **데이터 점의 분리 확인.** 이 옵션을 선택하면 모수 추정값이 중복되지 않았는지 확인하기 위한 검정이 수행됩니다. 모든 케이스를 올바르게 분류하는 모형을 프로시저에서 생성할 수 있는 경우에 분리가 발생합니다. 이 옵션은 다항 반응의 이항 반응 및 이분형 형식의 이항 반응에 사용할 수 있습니다.

수렴 기준 다음 옵션을 사용할 수 있습니다.

- **모수 수렴.** 이 옵션을 선택하면 모수 추정값의 절대 변화량 또는 상대 변화량이 지정된 값보다 크거나 같아지는 시점에 반복을 멈춥니다. 지정된 값은 0 또는 양의 정수이어야 합니다.
- **로그-우도 수렴.** 이 옵션을 선택하면 로그-우도 함수의 절대 변화량 또는 상대 변화량이 지정된 값보다 작아지는 시점의 계산 후에 반복을 멈춥니다. 지정된 값은 0 또는 양의 정수이어야 합니다.

- **Hessian 수렴.** 절대값 지정의 경우 수렴은 Hessian 수렴을 기준으로 하는 통계가 지정된 양의 값보다 작다고 가정합니다. 상대값 지정의 경우 수렴은 통계가 로그 우도의 절대값과 지정된 양의 값의 곱보다 작다고 가정합니다.

비정칙성 허용 오차. 비정칙(또는 비가역) 행렬에는 추정 알고리즘에 심각한 문제를 일으킬 수 있는 선형 종속 열이 있습니다. 거의 비정칙인 행렬은 잘못된 결과를 초래할 수 있으므로 프로시저는 행렬식이 허용 오차보다 작은 행렬은 비정칙으로 취급합니다. 양수값을 지정합니다.

일반화된 추정 방정식 초기값

프로시저는 초기 일반화 선형 모형을 추정하며 이 모형의 추정값은 일반화된 추정 방정식의 선형 모형 부분에서 모수 추정값에 대한 초기값으로 사용됩니다. 행렬 요소는 모수 추정값을 기준으로 하기 때문에 작업 중인 상관 행렬에는 초기값이 필요하지 않습니다. 이 대화 상자에서 지정한 초기값은 추정 탭의 최대 반복을 0으로 설정하지 않은 경우 일반화된 추정 방정식이 아닌 초기 일반화 선형 모형에 대한 출발점으로 사용됩니다.

초기값을 지정하는 경우 모형에 있는 모든 모수(중복 모수 포함)에 대해 제공해야 합니다. 데이터 세트에서 왼쪽에서 오른쪽으로 변수를 순서 지정할 때 다음과 같이 되어야 합니다. *RowType_*, *VarName_*, *P1*, *P2*, ..., where *RowType_* 및 *VarName_*은 문자열 변수이고 *P1*, *P2*, ...는 모수의 순서 목록에 해당하는 숫자 변수입니다.

- 초기값은 변수 *RowType_*의 경우 값 *EST*를 갖는 레코드에 제공되고 실제 초기값은 *P1*, *P2*, ... 아래 제공됩니다. 프로시저는 *RowType_*이 *EST* 이외의 값을 갖고 첫 번째 *RowType_*을 벗어나는 모든 레코드가 *EST*와 같은 모든 레코드를 무시합니다.
- 모형에 포함된 경우 절편 또는 반응에 다항 분포가 있는 경우 임계값 모수는 초기값을 먼저 나열해야 합니다.
- 척도 모수 및 반응에 음수 이항 분포가 있는 경우 음수 이항 모수는 초기값을 마지막에 지정해야 합니다.
- 분할 파일이 효과에 있는 경우 변수는 분할 파일을 만들 때 지정한 순서로 분할 파일 변수 또는 변수, 이어서 위와 같이 *RowType_*, *VarName_*, *P1*, *P2*, ... 등으로 시작해야 합니다. 분할은 원래 데이터 세트에서와 같은 순서로 지정한 데이터 세트에서 발생해야 합니다.

참고: 변수 이름 *P1*, *P2*, ... 는 필요하지 않습니다. 프로시저는 모수에 대한 변수 맵핑이 변수 이름이 아닌 변수 위치를 기준으로 하기 때문에 모수의 유효한 변수 이름을 받아들입니다. 마지막 모수를 초과하는 변수는 무시됩니다.

초기값의 파일 구조는 모형을 데이터로 내보낼 때 사용한 구조와 같습니다. 따라서 프로시저에서 한 실행의 최종 값을 후속 실행의 입력으로 사용할 수 있습니다.

일반화된 추정 방정식 통계

모형 효과. 다음 옵션을 사용할 수 있습니다.

- **분석 유형.** 모형 효과를 검정하기 위해 생성할 분석 유형을 지정합니다. 제 I 유형 분석은 일반적으로 모형에서 순서 예측변수에 대한 사전 이유가 있을 때 적합한 반면 제 III 유형은 보다 일반적으로 적용됩니다. Wald 또는 일반화 스코어 통계는 카이제곱 통계 그룹에서 선택한 것을 기준으로 계산됩니다.

- 신뢰구간. 50보다 크고 100보다 작은 신뢰수준을 지정합니다. Wald 구간은 선택한 카이제곱 통계 유형에 관계 없이 항상 생성되며 모수가 접근 정규 분포를 갖고 있다는 가정을 전제로 합니다.
- 로그 준-우도 함수. 이 함수는 로그 준-우도 함수의 출력 형식을 제어합니다. 전체 함수에는 모수 추정에 관계 없이 일관성 있는 추가 항이 포함되어 있습니다. 모수 추정에는 효과가 없으며 일부 소프트웨어 제품에서는 출력이 되지 않습니다.

인쇄. 다음과 같은 결과를 사용할 수 있습니다.

- 케이스 처리 요약. 분석에 포함되었거나 제외된 케이스 수와 백분율 및 상관 데이터 요약 테이블을 표시합니다.
- 기술통계. 종속변수, 공변량 및 요인에 대한 기술 통계와 요약 정보를 표시합니다.
- 모형 정보. 데이터 세트 이름, 종속변수 또는 이벤트 및 시행 변수, 오프셋 변수, 척도기중 변수, 확률 분포 및 연결 함수를 표시합니다.
- 적합도 통계량. 모형 선택에 대한 Akaike의 정보 기준의 두 가지 확장을 표시합니다. 즉, 최고의 상관 구조를 선택하는 경우 준-우도 독립성 모형 기준(QIC)을, 최고의 예측변수 서브세트를 선택하는 경우 또다른 QIC 측도 등을 표시합니다.
- 모형 요약 통계. 각 효과에 대한 제 I 유형 또는 제 III 유형 대비의 우도비 통계 및 모형 적합 총괄 검정을 포함한 모형 적합 검정을 표시합니다.
- 모수 추정값. 모수 추정값과 해당 검정 통계량 및 신뢰구간을 표시합니다. 원래 모수 추정값 외에 선택적으로 누승 모수 추정을 표시할 수 있습니다.
- 모수 추정값의 공분산 행렬. 추정된 모수 공분산 행렬이 표시됩니다.
- 모수 추정값의 상관 행렬. 추정된 모수 상관 행렬이 표시됩니다.
- 대비계수(L) 행렬. EM 평균 탭에서 요청하는 경우 기본 효과 및 주변 평균 추정에 대한 대비계수를 표시합니다.
- 일반 추정가능 함수. 대비계수(L) 행렬을 생성하는 지표를 표시합니다.
- 반복 히스토리. 모수 추정값과 로그 우도에 대한 반복 히스토리를 표시하고 기울기 벡터와 Hessian 행렬의 마지막 평가를 인쇄합니다. 반복 히스토리 테이블은 0^{번째} 반복(초기 추정값)부터 시작하여 매 n 번째 반복에 대한 모수 추정값을 표시합니다. 여기서 n 은 인쇄 구간 값입니다. 반복 히스토리를 요청하는 경우 n 에 관계 없이 마지막 반복은 항상 표시됩니다.
- 작업 중인 상관 행렬. 개체-내 종속성을 나타내는 행렬의 값을 표시합니다. 그 구조는 반복 탭의 지정에 따라 다릅니다.

일반화된 추정 방정식 EM 평균

이 탭을 사용하면 요인 대 요인 상호작용의 수준에 대해 주변 평균 추정을 표시할 수 있습니다. 전체 추정 평균을 표시하도록 요청할 수도 있습니다. 주변 평균 추정은 순서 다항 모형에는 사용할 수 없습니다.

요인 및 상호 작용. 이 목록에는 예측변수 탭에 지정된 요인과 모형 탭에 지정된 요인 상호작용이 포함되어 있습니다. 공변량은 이 목록에서 제외됩니다. 이 목록에서 항을 직접 선택하거나 **By** * 단추를 사용하여 상호작용 항에 결합할 수 있습니다.

평균 출력 기준. 추정된 평균은 선택한 요인 및 요인 상호작용에 대해 계산됩니다. 대비는 추정된 평균을 비교하기 위해 가설 검정을 어떻게 설정할지 결정합니다. 단순대비에는 참조 범주 또는 다른 항을 비교하는 데 기준으로 삼을 요인 수준이 필요합니다.

- **대응별.** 쌍대 비교는 지정했거나 내재된 요인의 모든 수준 조합에 대해 계산됩니다. 이는 요인 상호작용의 대비에만 사용할 수 있습니다.
- **단순.** 각 수준의 평균을 지정한 수준의 평균과 비교합니다. 이 대비 유형은 대조군이 있는 경우 유용합니다.
- **편차.** 요인의 각 수준은 총 평균에 비교됩니다. 편차 대비는 직교 대비가 아닙니다.
- **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.

척도. 주변 평균 추정은 종속변수의 원래 척도를 기준으로 하는 반응 또는 연결 함수에 의해 변환되는 종속변수를 기준으로 하는 선형 예측변수에 대해 계산할 수 있습니다.

다중 비교에 대한 조정. 다중 대비를 통해 가설 검정을 수행하는 경우 포함된 대비에 대한 유의 수준에서 전반적인 유의 수준을 조정할 수 있습니다. 이 그룹을 사용하면 조정 방법을 선택할 수 있습니다.

- **최소유의차.** 이 방법은 특정 선형 대비가 귀무가설 값과 다르다는 가설을 거부할 전체 확률을 제어하지 않습니다.
- **Bonferroni.** 이 방법은 관측 유의 수준을 다중 대비가 검정되는 요인에 맞춥니다.
- **순차 Bonferroni(Sequential Bonferroni).** 순차 단계별로 낮아지는 거부 Bonferroni 프로시저로, 개별 가설은 거부하지만 동일한 전체 유의 수준을 유지한다는 점에서 훨씬 덜 보수적인 방법입니다.
- **Sidak.** 이 방법은 Bonferroni 접근법보다 더 엄격한 한계를 제공합니다.
- **순차 Sidak(Sequential Sidak).** 순차 단계별로 낮아지는 거부 Sidak 프로시저로, 개별 가설은 거부하지만 동일한 전체 유의 수준을 유지한다는 점에서 훨씬 덜 보수적인 방법입니다.

일반화된 추정 방정식 저장

확인된 항목은 지정된 이름으로 저장됩니다. 같은 이름을 가진 새 변수와 같은 이름을 가진 기존 변수를 덮어쓰도록 선택하여 새 변수 이름을 고유하게 만들 수 있습니다.

- **반응 평균의 예측값.** 원래 반응 매트릭에서 각 케이스에 대한 모형 예측값을 저장합니다. 반응 분포가 이항이고 종속변수가 이분형인 경우 프로시저는 예측 확률을 저장합니다. 반응 분포가 다항이면 항목 레이블은 누적 예측 확률이 되고 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 누적 예측 확률을 저장합니다.
- **반응 평균에 대한 신뢰구간의 하한.** 반응 평균에 대한 신뢰구간의 하한을 저장합니다. 반응 분포가 다항이면 항목 레이블은 누적 예측 확률에 대한 신뢰구간의 하한이 되고 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 하한을 저장합니다.
- **반응 평균에 대한 신뢰구간의 상한.** 반응 평균에 대한 신뢰구간의 상한을 저장합니다. 반응 분포가 다항이면 항목 레이블은 누적 예측 확률에 대한 신뢰구간의 상한이 되고 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 상한을 저장합니다.
- **예측 범주.** 이항 분포와 이분형 종속변수가 있는 모형 또는 다항 분포의 경우 각 케이스에 대한 예측 반응 범주를 저장합니다. 이 옵션은 다른 반응 분포에는 사용할 수 없습니다.
- **선형 예측변수의 예측값.** 선형 예측변수의 매트릭에서 각 케이스에 대한 모형 예측값을 저장합니다(지정한 연결 함수를 통해 변환된 반응). 반응 분포가 다항이면 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 예측값을 저장합니다.
- **선형 예측변수의 예측된 값의 예측된 표준 오차.** 반응 분포가 다항이면 프로시저는 마지막을 제외하고 저장할 지정된 범주 수까지 반응의 각 범주에 대한 예측 표준 오차를 저장합니다.

다음 항목은 반응 분포가 다항일 때 사용할 수 있습니다.

- **원시 잔차.** 관측값과 모형에 의해 예측된 값 사이 차이입니다.
- **Pearson 잔차.** 원시 잔차의 부호가 있는 Pearson 카이제곱 통계에 대한 케이스의 기여도 제공근.

일반화된 추정 방정식 내보내기

데이터로 모형 내보내기. 모수 추정값, 표준 오차, 유의수준 및 자유도와 함께 모수 상관 행렬이나 공분산 행렬이 포함되어 있는 IBM SPSS Statistics 형식의 데이터 세트를 씁니다. 행렬 파일에서 변수의 순서는 다음과 같습니다.

- **분할변수.** 사용된 경우 분할을 정의하는 모든 변수.
- **RowType_.** 값(및 값 레이블) *COV* (공분산), *CORR* (상관), *EST* (모수 추정값), *SE* (표준 오차), *SIG* (유의 수준) 및 *DF* (표본추출 디자인 자유도)를 사용합니다. 각 모형 모수에 대해 행 유형 *COV* (또는 *CORR*)을 갖는 별도의 케이스는 물론 각기 다른 행 유형에 대해 별도의 케이스가 있습니다.
- **VarName_.** 모수 추정값 테이블에 표시된 모수 문자에 해당하는 값 레이블과 함께 행 유형 *COV* 또는 *CORR*에 대해 모든 추정 모형 모수(척도 또는 음수 이항 모수)의 순서 목록에 해당하는 값 *P1*, *P2*, ...,을 사용합니다. 다른 행 유형의 셀은 비어 있습니다.
- **P1, P2, ...** 이러한 변수는 모수 추정값 테이블에 표시된 모수 문자에 해당하는 변수 레이블과 함께 모든 모형 모수(해당되는 경우 척도와 음수 이항 모수를 포함)의 순서 목록에 해당하며 행 유형에 따라 값을 사용합니다.

중복 모수의 경우 모든 공분산은 0으로 설정되고 상관은 시스템 결측값으로 설정됩니다. 모든 모수 추정값은 0으로 설정되고 모든 표준 오차, 유의 수준 및 잔차의 자유도는 시스템 결측값으로 설정됩니다.

척도 모수의 경우 공분산, 상관, 유의 수준 및 자유도는 시스템 결측값으로 설정됩니다. 척도 모수를 최대 우도를 통해 추정한 경우 표준 오차가 제공되며, 그렇지 않으면 시스템 결측값으로 설정됩니다.

음수 이항의 경우 공분산, 상관, 유의 수준 및 자유도는 시스템 결측값으로 설정됩니다. 음수 이항 모수를 최대 우도를 통해 추정한 경우 표준 오차가 제공되며, 그렇지 않으면 시스템 결측값으로 설정됩니다.

분할이 있는 경우에는 모수 목록을 모든 분할에 대해 누적해야 합니다. 해당 분할에서 일부 모수는 관련이 없을 수 있으며 이는 중복과 같지 않습니다. 관련이 없는 모수의 경우 모든 공분산이나 상관, 모수 추정값, 표준 오차, 유의 수준 및 자유도는 시스템 결측값으로 설정됩니다.

이 행렬 파일을 추가 모형 추정을 위한 초기값으로 사용할 수 있습니다. 이러한 프로시저가 여기에 내보낸 모든 행 유형을 받아 들이지 않는 경우 이 파일은 다른 프로시저에서 추가 분석에 즉시 사용할 수 없습니다. 그러면 이 행렬 파일에 있는 모든 모수가 파일을 읽는 프로시저에 같은 의미를 가지므로 주의해야 합니다.

XML로 모형 내보내기. 모수 추정값과 모수 공분산 행렬(선택한 경우)을 XML(PMML) 형식으로 저장합니다. 스코어링을 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.

GENLIN 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 모수 추정값의 초기값을 숫자 목록으로 지정합니다(CRITERIA 하위 명령 사용).
- 고정된 작업 중인 상관 행렬을 지정합니다(REPEATED 하위 명령 사용).
- EMMEANS 하위 명령을 사용하여 추정된 주변 평균을 계산할 때 해당 평균 이외의 다른 값으로 공변량을 고정할 수 있습니다.
- 주변 평균 추정의 사용자 정의 다항 대비를 지정합니다(EMMEANS 하위 명령 사용).
- 지정된 대비 유형을 사용하는 것과 비교하여 주변 평균 추정이 표시될 요인의 서브세트를 지정합니다(EMMEANS 하위 명령의 TABLES 및 COMPARE 키워드 사용).

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

제 8 장 일반화 선형 혼합 모형

일반화 선형 혼합 모형은 선형 모형을 확장하여

- 대상이 지정된 연결 함수를 통해 요인 및 공변량과 선형적으로 관련되도록 합니다.
- 대상은 비정규 분포를 가질 수 있습니다.
- 관측값은 상호 관련될 수 있습니다.

일반화 선형 혼합 모형은 단순 선형 회귀 모형에서 비정규 경시적 자료에 대한 복합 다중 수준 모형까지 다양한 모형을 포함합니다.

예제. 교육청은 일반화 선형 혼합 모형을 사용하여 실험적인 교수법이 수학 스코어 향상에 효과적인지 아닌지를 알 수 있습니다. 같은 클래스의 학생들은 같은 교사에게 배웠으므로 상호 관련되어야 하며, 같은 학교에 있는 클래스도 상호 관련될 수 있으므로 학교와 클래스 수준에서 랜덤 효과를 포함하여 변동의 다른 소스를 설명할 수 있습니다. 자세한 정보는 을 참조하십시오.

의료 연구자들은 항경련제가 환자의 발작률을 줄일 수 있는지의 여부를 알아보기 위해 일반화 선형 혼합 모형을 사용할 수 있습니다. 같은 환자에게서 얻은 반복 측정은 대체로 양의 상관을 가지므로 일부 랜덤 효과가 있는 혼합 모형이 적합합니다. 발작 횟수인 대상 필드는 양의 정수 값을 가지므로 포아송 분포 및 로그 링크를 가진 일반화 선형 혼합 모형이 적합할 수 있습니다. 자세한 정보는 주제를 참조하십시오.

TV, 전화 및 인터넷 서비스의 케이블 제공업체 대표는 일반화 선형 혼합 모형을 사용하여 잠재 고객에 대해 더 잘 알 수 있습니다. 가능한 반응은 명목 측정 수준을 갖고 있으므로 회사 분석가는 임의 절편이 있는 일반화된 로짓 혼합 모형을 사용하여 주어진 설문조사 반응자의 반응 내에서 서비스 유형(TV, 전화, 인터넷)에 대한 서비스 사용 질문에 대한 반응 간의 상관을 알 수 있습니다. 자세한 정보는 주제를 참조하십시오.

데이터 구조 탭에서는 관측값들을 연결할 때 데이터 세트를 구성하는 레코드 간의 구조적 관계를 지정할 수 있습니다. 데이터 세트의 레코드가 독립된 관측값인 경우 이 탭에서 아무 것도 지정할 필요가 없습니다.

개체. 지정된 범주형 필드 값의 조합은 데이터 세트 내의 개체를 고유하게 정의해야 합니다. 예를 들어, 단일 환자 ID 필드는 단일 병원의 개체를 정의하기에 충분해야 하지만, 환자 식별 번호가 병원마다 고유하지 않은 경우 병원 ID와 환자 ID의 조합이 필요할 수 있습니다. 반복 측도 설정에서 각 개체마다 여러 관측을 기록하므로 각 개체는 데이터 세트에서 여러 레코드를 차지할 수 있습니다.

개체는 다른 개체에 대해 독립적인 것으로 간주할 수 있는 관측 단위입니다. 예를 들어, 의학 연구에서 한 환자의 혈압 기록은 다른 환자의 기록에 대해 독립적인 것으로 간주할 수 있습니다. 개체마다 반복 측정값이 있고 이러한 관측값 간의 상관을 모델링하려는 경우 개체를 정의하는 것이 매우 중요합니다. 예를 들어, 담당 의사에게 지속적으로 진찰을 받는 한 환자의 혈압 기록은 상호 관련된 것으로 예상할 수 있습니다.

데이터 구조 탭에서 개체로 지정된 모든 필드는 잔차 공분산 구조에 대한 개체를 정의하는 데 사용되며 랜덤 효과 블록에서 랜덤 효과 공분산 구조에 대한 개체를 정의하는 데 가능한 필드 목록을 제공합니다.

반복 측도. 여기에 지정된 필드는 반복 관측값을 식별하는 데 사용됩니다. 예를 들어, 단일 변수 주는 의학 연구에서의 10주 동안의 관측을 식별하는 데 사용하거나 월 및 일은 1년 동안의 일별 관측을 식별하는 데 함께 사용할 수 있습니다.

공분산 그룹 정의 기준. 여기에 지정된 범주형 필드는 독립된 반복 효과 공분산 모수 세트를 정의합니다. 각 범주의 모수는 그룹 필드의 교차 분류에 의해 정의됩니다. 모든 개체의 공분산 유형은 동일하며, 동일한 공분산 그룹 내의 개체는 모수에 대해 같은 값을 가집니다.

반복 공분산 유형. 잔차에 대한 공분산 구조를 지정합니다. 사용 가능한 구조는 다음과 같습니다.

- 1차 자기회귀(AR1)
- 자기회귀 이동 평균(1,1) (ARMA11)
- 복합 대칭
- 대각선
- 척도법 항등
- Toeplitz
- 비구조적
- 분산 성분

일반화 선형 혼합 모형 구하기

이 기능을 사용하려면 Advanced Statistics 옵션이 필요합니다.

메뉴에서 다음을 선택합니다.

분석 > 혼합 모형 > 일반화된 선형...

1. 데이터 구조 탭에서 데이터 세트의 개체 구조를 정의합니다.
2. 필드 및 효과 탭에는 측정 수준 또는 이벤트/시행 지정을 가질 수 있는 단일 대상이 있어야 합니다. 어떤 경우든, 이벤트와 시행 지정은 연속형이어야 합니다. 선택적으로 분포 및 연결 함수, 고정 효과, 랜덤 효과 블록, 오프셋 또는 분석 가중치를 지정합니다.
3. 작성 옵션을 클릭하여 선택적 작성 설정을 지정합니다.
4. 모형 옵션을 클릭하여 활성 데이터 세트에 스코어를 저장하고 외부 파일로 모형을 내보냅니다.
5. 실행을 클릭하여 프로시저를 실행하고 모형 오브젝트를 만듭니다.

대상

이 설정은 대상, 분포 및 연결 함수를 통한 예측변수에 대한 관계를 정의합니다.

대상. 대상은 필수입니다. 어떤 측정 수준도 가질 수 있으며, 대상의 측정 수준은 적합한 분포 및 연결 함수를 제한합니다.

- 시행 수를 분모로 사용합니다. 대상 반응이 시행 세트에서 발생하는 많은 이벤트면 대상 필드에는 이벤트 수가 포함되며 시행 수가 포함되어 있는 추가 필드를 선택할 수 있습니다. 예를 들어, 새 살충제를 실험할 때 개미 표본에 다른 농도의 살충제를 사용하고 죽은 개미 수와 각 표본의 개미 수를 기록할 수 있습니다. 이 경우 죽은 개미 수를 기록한 필드를 대상(이벤트) 필드로 지정하고, 각 표본의 개미 수를 기록한 필드를 시행 필드로 지정해야 합니다. 각 표본의 개미 수가 동일한 경우 시행 수를 고정 값으로 지정할 수 있습니다.

각 레코드에 대해 시행 수는 이벤트 수보다 크거나 같아야 합니다. 이벤트는 양의 정수가 되어야 하며 시행 수는 양의 정수가 되어야 합니다.

- 참조 범주 사용자 정의. 범주형 대상의 경우, 참조 범주를 선택할 수 있습니다. 이것은 모수 추정값과 같은 특정 결과에 영향을 미칠 수 있지만 모형 적합을 변경해서는 안 됩니다. 예를 들어 대상이 0, 1, 2 값을 가지면, 기본적으로 프로시저는 마지막(가장 높은 값) 범주 또는 참조 범주인 2를 만듭니다. 이 상황에서 모수 추정값은 범주 2의 가능도에 비례하여 범주 0 또는 1의 가능도와 관련된 것으로 해석해야 합니다. 사용자 정의 범주를 지정하고 대상에 정의된 레이블이 있는 경우 목록에서 값을 선택하여 참조 범주를 설정할 수 있습니다. 이는 모형을 지정하는 동안 정확히 특정 필드를 어떻게 코딩했는지 기억이 나지 않을 때 편리할 수 있습니다.

대상 분포 및 선형 모형과의 관계(링크). 예측변수의 값을 고려할 때, 모형은 지정된 형태를 따르기 위해 대상 값의 분포를 예상하고, 대상 값의 경우 지정된 연결 함수를 통해 예측변수와 선형적으로 관련됩니다. 여러 공통 모형에 대한 바로 가기가 제공됩니다. 바로 가기 목록에는 없지만 맞추고자 하는 특정한 분포 및 연결 함수 조합이 있는 경우 사용자 정의를 선택합니다.

- 선형모형. 정규 분포를 항등 링크와 함께 지정합니다. 이는 대상이 선형 회귀분석 또는 ANOVA 모형을 사용하여 예측될 수 있을 때 유용합니다.
- 감마 회귀분석. 감마 분포를 로그 링크와 함께 지정합니다. 이는 대상에 모든 양수값이 포함되고 더 큰 값 쪽으로 비대칭될 때 사용되어야 합니다.
- 로그선형분석. 포아송 분포를 로그 링크와 함께 지정합니다. 이는 대상이 고정 기간 동안 발생 개수를 나타낼 때 사용되어야 합니다.
- 음수 이항 회귀분석. 음수 이항 분포를 로그 링크와 함께 지정합니다. 이는 대상과 분모가 k 성공을 관측하는 데 필요한 시행 수를 나타낼 때 사용되어야 합니다.
- 다항 로지스틱 회귀분석. 다항 분포를 지정합니다. 이는 대상이 다범주 반응일 때 사용되어야 합니다. 누적 로짓 링크(순서 결과)나 일반화된 로짓 링크(다범주 명목 반응)를 사용합니다.
- 이분형 로지스틱 회귀분석. 이항 분포를 로짓 링크와 함께 지정합니다. 이는 대상이 로지스틱 회귀 모형에 의해 예측된 이분형 반응일 때 사용되어야 합니다.
- 이분형 프로빗. 이항 분포를 프로빗 링크와 함께 지정합니다. 이는 대상이 기본 정규 분포가 있는 이분형 반응일 때 사용되어야 합니다.
- 구간 중도절단 생존. 이항 분포를 보 로그-로그 링크와 함께 지정합니다. 이는 몇몇 관측값에 종료 이벤트가 없을 때 생존 분석에서 유용합니다.

분포

이 선택은 대상의 분포를 지정합니다. 비정규 분포와 항등하지 않은 연결 함수를 지정하는 기능은 선형 혼합 모형에서 일반화 선형 혼합 모형의 중요한 개선 사항입니다. 많은 분포-연결 함수가 있으며 주어진 데이터 세트에 적합한 함수가 여러 개일 수 있으므로 사전 이론적 고려 사항을 바탕으로 선택하거나 어느 조합이 가장 적합할지를 고려하여 선택하면 됩니다.

- **이항검정.** 이 분포는 이분형 반응이나 이벤트 수를 나타내는 대상의 경우에만 적합합니다.
- **감마.** 이 분포는 더 큰 양의 값 쪽으로 비대칭되는 양의 척도 값을 가진 대상에 적합합니다. 데이터 값이 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- **역 가우스.** 이 분포는 더 큰 양의 값 쪽으로 비대칭되는 양의 척도 값을 가진 대상에 적합합니다. 데이터 값이 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.
- **다항.** 이 분포는 다범주 반응을 나타내는 대상에 적합합니다. 모형 형태는 대상의 측정 수준에 따라 다릅니다.

명목 대상은 대상의 각 범주(참조 범주 제외)에 대해 별도의 모형 모수 세트가 추정되는 명목 다항 모형을 생성합니다. 주어진 예측변수에 대한 모수 추정값은 대상의 각 범주의 우도와 해당 예측변수 사이의 참조 범주와 관련한 관계를 보여줍니다.

순서 대상은 일반적인 절편 항이 대상 범주의 누적 확률과 관련된 임계값 모수의 세트로 대체되는 순서 다항 모형을 생성합니다.

- **음수 이항.** 음수 이항 회귀분석은 대상이 높은 분산의 발생 개수를 나타낼 때 사용되는 음수 이항 분포를 로그 링크와 함께 사용합니다.
- **정상.** 이것은 중앙(평균) 값에 대해 값이 대칭되는 종 형태의 분포를 띠는 연속형 대상에 적합합니다.
- **포아송.** 이 분포는 고정 기간 동안 중요 이벤트의 발생 수로 생각할 수 있으며 양의 정수 값을 갖는 변수에 적합합니다. 데이터 값이 양수이거나, 0보다 작거나 같고 또는 없는 경우 해당 케이스는 분석에 사용되지 않습니다.

연결 함수

연결 함수는 모형을 추정할 수 있는 대상의 변환입니다. 다음 함수를 사용할 수 있습니다.

- **항등.** $f(x)=x$. 대상은 변환되지 않습니다. 이 링크는 다항을 제외한 모든 분포에 사용할 수 있습니다.
- **보 로그-로그.** $f(x)=-\log(-\log(1-x))$. 이항 또는 다항 분포에만 적합합니다.
- **Cauchit.** $f(x) = \tan(\pi (x - 0.5))$. 이항 또는 다항 분포에만 적합합니다.
- **로그.** $f(x)=\log(x)$. 이 링크는 다항을 제외한 모든 분포에 사용할 수 있습니다.
- **로그 보.** $f(x)=\log(1-x)$. 이항 분포에만 적합합니다.
- **로짓.** $f(x)=\log(x / (1-x))$. 이항 또는 다항 분포에만 적합합니다.
- **음수 로그-로그.** $f(x)=-\log(-\log(x))$. 이항 또는 다항 분포에만 적합합니다.
- **프로빗.** $f(x)=\Phi^{-1}(x)$, 여기서 Φ^{-1} 은 표준 정규 누적 분포의 역함수입니다. 이항 또는 다항 분포에만 적합합니다.

- 거듭제곱. $f(x)=x^{\alpha}$, $\alpha \neq 0$ 인 경우. $f(x)=\log(x)$, $\alpha=0$ 인 경우. α 는 필요한 숫자 지정이며 실수가 되어야 합니다. 이 링크는 다항을 제외한 모든 분포에 사용할 수 있습니다.

고정 효과





고정 효과 요인은 일반적으로 해당 값이 모두 데이터 세트에 나타나는 필드로 볼 수 있으며 스코어링에 사용할 수 있습니다. 기본적으로 대화 상자에 지정되지 않은 사전 정의된 입력 역할이 있는 필드가 모형의 고정 효과 부분에 입력됩니다. 범주형(명목 및 순서) 필드는 모형에서 요인으로 사용되고 연속형 필드는 공변량으로 사용됩니다.

소스 목록에서 하나 이상의 필드를 선택하고 효과 목록으로 끌어 모형에 효과를 입력합니다. 생성되는 효과의 유형은 선택을 끄는 핫스팟에 따라 다릅니다.

- 기본. 끌어 놓은 필드가 효과 목록 하단에 별도의 주효과로 나타납니다.
- 2원배치. 끌어 놓은 필드의 모든 가능한 쌍이 효과 목록 하단에 이원 상호작용으로 나타납니다.
- 3원배치. 끌어 놓은 필드의 모든 가능한 세 개로 구성된 세트가 효과 목록 하단에 3원배치 상호작용으로 나타납니다.
- *. 끌어 놓은 모든 필드의 조합이 효과 목록 하단에 단일 상호작용으로 나타납니다.

효과 작성기의 오른쪽에 있는 단추를 사용하면 여러 조치를 수행할 수 있습니다.

표 1. 효과 작성기 단추 설명.

아이콘	설명
	삭제할 항목을 선택한 후 삭제 단추를 클릭하여 고정 효과 모형에서 항목을 삭제할 수 있습니다.
 	순서를 변경하려는 항목을 선택한 후 위 또는 아래 화살표를 클릭하여 고정 효과 모형에서 항목의 순서를 바꿀 수 있습니다.
	『사용자 정의 항목 추가』 대화 상자에서 사용자 정의 항목 추가 단추를 클릭하여 중첩 항목을 모형에 추가할 수 있습니다.

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

사용자 정의 항목 추가

이 프로시저에서는 모형에 대해 중첩 항목을 설정할 수 있습니다. 중첩 항목은 요인 또는 공변량의 효과를 모델링하는 데 유용합니다. 이들 값은 다른 요인 수준과 상호작용하지 않습니다. 예를 들어, 식료품 체인점은 여러 점포에서의 고객의 소비 성향을 살펴볼 수 있습니다. 각 고객은 체인점 중의 한 곳만 자주 가기 때문에 고객 효과는 점포 효과 내에 중첩되었다고 할 수 있습니다.

또한 같은 공변량을 포함하고 있는 다항 항 같이 상호작용 효과를 포함시키거나 여러 수준의 중첩을 중첩 항에 추가할 수 있습니다.

제한. 중첩 항에는 다음과 같은 제한이 있습니다.

- 상호작용 내의 모든 요인은 고유해야 합니다. 따라서 A 가 요인이면 $A*A$ 지정은 유효하지 않습니다.
- 중첩 효과 내의 모든 요인은 고유해야 합니다. 따라서 A 가 요인이면 $A(A)$ 지정은 유효하지 않습니다.
- 공변량 내에 효과를 중첩할 수 없습니다. 따라서 A 가 요인이고 X 가 공변량이면 A 지정은 유효하지 않습니다.

중첩 항 작성

1. 다른 요인 내에 중첩된 요인 또는 공변량을 선택한 다음 화살표 단추를 클릭합니다.
2. (포함)을 클릭합니다.
3. 이전 요인이나 공변량이 중첩된 요인을 선택한 다음 화살표 단추를 클릭합니다.
4. 항 추가를 클릭합니다.

선택적으로 상호작용 효과를 포함시키거나 여러 수준의 중첩을 중첩 항에 추가할 수 있습니다.

랜덤 효과

랜덤 효과 요인은 데이터 파일의 값을 더 큰 모집단 값의 무작위 표본으로 간주할 수 있는 필드입니다. 대상의 과도한 변동을 설명할 때 유용합니다. 기본적으로 데이터 구조 탭에서 개체를 둘 이상 선택한 경우, 가장 인쪽에 있는 개체 너머에 각 개체에 대해 랜덤 효과 블록이 생성됩니다. 예를 들어 데이터 구조 탭에서 학교, 클래스 및 학생을 개체로 선택한 경우, 다음 랜덤 효과 블록이 자동으로 생성됩니다.

- 랜덤 효과 1: 개체가 학교임(효과 없음, 절편만)
- 랜덤 효과 2: 개체가 학교 * 클래스임(효과 없음, 절편만)

다음과 같은 방법으로 랜덤 효과 블록을 사용할 수 있습니다.

1. 새 블록을 추가하려면 블록 추가...를 클릭합니다. 『랜덤 효과 블록』 대화 상자가 열립니다.
2. 기존 블록을 편집하려면 편집하려는 블록을 선택하고 블록 편집... 을 클릭합니다. 『랜덤 효과 블록』 대화 상자가 열립니다.
3. 블록을 하나 이상 삭제하려면 삭제하려는 블록을 선택하고 삭제 단추를 클릭합니다.

랜덤 효과 블록





소스 목록에서 하나 이상의 필드를 선택하고 효과 목록으로 끌어 모형에 효과를 입력합니다. 생성되는 효과의 유형은 선택을 끄는 핫스팟에 따라 다릅니다. 범주형(명목 및 순서) 필드는 모형에서 요인으로 사용되고 연속형 필드는 공변량으로 사용됩니다.

- 기본. 끌어 놓은 필드가 효과 목록 하단에 별도의 주효과로 나타납니다.
- 2원배치. 끌어 놓은 필드의 모든 가능한 쌍이 효과 목록 하단에 이원 상호작용으로 나타납니다.

- 3원배치. 끌어 놓은 필드의 모든 가능한 세 개로 구성된 세트가 효과 목록 하단에 3원배치 상호작용으로 나타납니다.
- *. 끌어 놓은 모든 필드의 조합이 효과 목록 하단에 단일 상호작용으로 나타납니다.

효과 작성기의 오른쪽에 있는 단추를 사용하면 여러 조치를 수행할 수 있습니다.

표 2. 효과 작성기 단추 설명.

아이콘	설명
	삭제할 항목을 선택한 후 삭제 단추를 클릭하여 모형에서 항목을 삭제할 수 있습니다.
 	순서를 변경하려는 항목을 선택한 후 위 또는 아래 화살표를 클릭하여 모형에서 항목의 순서를 바꿀 수 있습니다.
	65 페이지의 『사용자 정의 항목 추가』 대화 상자에서 사용자 정의 항목 추가 단추를 클릭하여 중첩 항목을 모형에 추가할 수 있습니다.

절편 포함. 절편은 기본적으로 랜덤 효과 모형에 포함되지 않습니다. 데이터가 원점을 전달한다고 가정할 경우에는 절편을 제외시킬 수 있습니다.

공분산 그룹 정의 기준. 여기에 지정된 범주형 필드는 독립된 랜덤 효과 공분산 모수 세트를 정의합니다. 각 범주의 모수는 그룹 필드의 교차 분류에 의해 정의됩니다. 각 랜덤 효과 블록에 다른 그룹 필드 세트를 지정할 수 있습니다. 모든 개체의 공분산 유형은 동일하며, 동일한 공분산 그룹 내의 개체는 모수에 대해 같은 값을 가집니다.

개체 조합. 데이터 구조 탭의 사전 설정된 개체 조합에서 랜덤 효과 개체를 지정할 수 있습니다. 예를 들어 데이터 구조 탭에 학교, 클래스 및 학생이 순서대로 개체로 정의된 경우, 개체 조합 드롭다운 목록에는 없음, 학교, 학교 * 클래스 및 학교 * 클래스 * 학생이 옵션으로 표시됩니다.

랜덤 효과 공분산 유형. 잔차에 대한 공분산 구조를 지정합니다. 사용 가능한 구조는 다음과 같습니다.

- 1차 자기회귀(AR1)
- 자기회귀 이동 평균(1,1) (ARMA11)
- 복합 대칭
- 대각선
- 척도법 항등
- Toeplitz
- 비구조적
- 분산 성분

가중치 및 오프셋

분석 가중치. 척도 모수는 반응의 변수와 관련한 추정된 모형 모수입니다. 분석 가중치는 "알려진" 값으로, 관측할 때마다 달라질 수 있습니다. 분석 가중 필드를 지정한 경우 반응의 변수와 관련한 척도 모수는 각 관측에 대해 분석 가중치로 나눈 것입니다. 0보다 작거나 같고 또는 값이 없는 분석 가중치를 가진 레코드는 분석에 사용되지 않습니다.

오프셋. 오프셋 항은 "구조" 예측변수입니다. 계수는 모형으로 추정되지 않지만 값 1을 갖는 것으로 가정합니다. 따라서 오프셋 값은 단순히 대상의 선형 예측변수에 추가됩니다. 이는 각 케이스가 중요 이벤트마다 다른 노출 수준을 가질 수 있는 포아송 회귀 모형에 특히 유용합니다.

예를 들어, 개별 운전자의 사고 비율을 모델링할 때 3년 경력 중 한 번의 사고를 낸 운전자와 25년 경력 중에 한 번 사고를 낸 운전자 사이에는 중요한 차이가 있습니다! 사고 수는 운전 경력의 자연 로그가 오프셋 항으로 포함되는 경우 로그 링크와 함께 포아송 또는 음수 이항 반응으로 모델링할 수 있습니다.

다른 분포와 링크 유형의 조합은 오프셋 변수의 다른 변환이 필요할 수 있습니다.

일반 빌드 옵션

이 선택은 모형을 작성하는 데 사용되는 몇몇 고급 기준을 지정합니다.

정렬 순서. 이 제어는 "마지막" 범주를 결정하기 위해 대상 및 요인(범주형 입력)의 범주 순서를 결정합니다. 대상이 범주형이 아니거나 사용자 정의 참조 범주가 62 페이지의 『대상』 설정에 지정된 경우, 대상 정렬 순서 설정이 무시됩니다.

중지 규칙. 알고리즘에서 실행할 최대 반복 횟수를 지정할 수 있습니다. 알고리즘은 내부 루프 및 외부 루프로 구성된 이중 반복 프로세스를 사용합니다. 최대 반복 수에 지정된 값이 두 루프 모두에 적용됩니다. 비음수 정수를 지정하십시오. 기본값은 100입니다.

사후 추정 설정. 이 설정은 몇몇 모형 결과가 보기에 대해 어떻게 계산되는지를 결정합니다.

- **신뢰수준.** 모형 계수의 구간 추정값을 계산하는 데 사용되는 신뢰수준입니다. 0보다 크고 100보다 작은 값을 지정하십시오. 기본값은 95입니다.
- **자유도.** 이 옵션은 유의수준 검정에 대해 자유도가 어떻게 계산되는지를 지정합니다. 표본 결과가 충분히 크거나, 데이터가 균형을 이루거나, 척도법 항등 또는 대각선처럼 모형이 간단한 공분산 유형을 사용하는 경우 모든 검정에 대해 고정(잔차 방법)을 선택합니다. 기본값입니다. 표본 결과가 작거나, 데이터가 비균형적이거나, 비구조적처럼 모형이 복잡한 공분산 유형을 사용하는 경우 검정마다 다름(Satterthwaite approximation)을 선택합니다.
- **고정 효과 및 계수 검정.** 모수 추정값 공분산 행렬을 계산하는 방법입니다. 모형 가정을 위반할 염려가 있는 경우 강력한 추정을 선택하십시오.

추정

모형 작성 알고리즘은 내부 루프와 외부 루프로 구성된 이중 반복 프로세스를 사용합니다. 다음 설정은 내부 루프에 적용됩니다.

모수 수렴.

모수 추정값의 최대 절대 변화량 또는 최대 상대 변화량이 지정된 값(음수가 아니어야 함) 미만인 경우 수렴을 가정합니다. 지정된 값이 0이면 기준을 사용하지 않습니다.

로그 우도 수렴.

로그-우도 함수의 절대 변화량 또는 상대 변화량이 지정된 값(음수가 아니어야 함)미만인 경우 수렴을 가정합니다. 지정된 값이 0이면 기준을 사용하지 않습니다.

Hessian 수렴.

절대값 지정의 경우 Hessian을 기준으로 하는 통계가 지정된 값 미만인 경우 수렴을 가정합니다. 상대 값 지정의 경우 수렴은 통계가 로그 우도의 절대값과 지정된 값의 곱보다 작다고 가정합니다. 지정된 값이 0이면 기준을 사용하지 않습니다.

최대 Fisher 스코어링 단계.

반복 횟수 n 까지 Fisher 스코어링 알고리즘을 사용하도록 요청합니다. 양수를 지정하십시오.

비정칙성 허용 오차.

이 값은 비정칙성 확인 시 허용 오차로 사용합니다. 양수를 지정하십시오.

참고: 기본적으로 모수 수렴이 사용되며 허용 오차 1E-6의 최대 절대 변경을 확인합니다. 이 설정은 버전 22 이전의 버전에서 얻은 결과와는 다른 결과를 얻을 수 있습니다. 22 이전 버전에서 결과를 재현하려면 모수 수렴 기준으로 상대값을 사용하고 기본 허용 오차를 1E-6으로 유지하십시오.

평균 추정

이 탭을 사용하면 요인 대 요인 상호작용의 수준에 대해 주변 평균 추정을 표시할 수 있습니다. 주변 평균 추정은 다항 모형에는 사용할 수 없습니다.

항. 범주형 필드로만 구성된 고정 효과의 모형 항은 다음과 같습니다. 모형이 주변 평균 추정을 생성할 각 항을 확인합니다.

- **대비 유형.** 대비 필드의 수준에 대해 사용할 대비 유형을 지정합니다. **없음**이 선택되면 대비가 생성되지 않습니다. **대응별**은 지정된 요인의 모든 수준 조합에 대해 쌍대 비교를 생성합니다. 이는 요인 상호작용의 대비에만 사용할 수 있습니다. **편차 대비**는 요인의 각 수준을 총 평균과 비교합니다. **단순 대비**는 마지막을 제외한 요인의 각 수준을 마지막 수준과 비교합니다. "마지막" 수준은 작성 옵션에 지정된 요인의 정렬 순서에 의해 결정됩니다. 이러한 대비 유형은 모두 직교하지 않습니다.
- **대비 필드.** 선택된 대비 유형을 사용하여 비교되는 수준인 요인은 지정합니다. **없음**을 대비 유형으로 선택한 경우 대비 필드를 선택할 수 없거나 선택할 필요가 없습니다.

연속형 필드. 열거된 연속형 필드는 연속형 필드를 사용하는 고정 효과의 항에서 추출됩니다. 주변 평균 추정을 계산할 때 공변량이 지정된 값으로 고정됩니다. 평균을 선택하거나 사용자 정의 값을 지정하십시오.

평균 추정을 표시합니다. 주변 평균 추정을 대상의 원래 척도를 기준으로 계산할지 연결 함수 변환을 기준으로 계산할지 지정합니다. 원래 대상 척도는 대상의 주변 평균 추정을 계산합니다. 대상이 이벤트/시행 옵션을 사용하여 지정되었을 때 이벤트 수보다는 이벤트/시행 비율에 대한 주변 평균 추정을 제공합니다. 연결 함수 변환은 선형 예측변수의 주변 평균 추정을 계산합니다.

다중 비교 사용에 대한 조정. 다중 대비를 통해 가설 검정을 수행하는 경우 포함된 대비에 대한 유의 수준에서 전반적인 유의 수준을 조정할 수 있습니다. 조정 방법을 선택할 수 있습니다.

- **최소유의차.** 이 방법은 특정 선형 대비가 귀무가설 값과 다르다는 가설을 거부할 전체 확률을 제어하지 않습니다.
- **순차 Bonferroni(Sequential Bonferroni).** 순차 단계별로 낮아지는 거부 Bonferroni 프로시저로, 개별 가설은 거부하지만 동일한 전체 유의 수준을 유지한다는 점에서 훨씬 덜 보수적인 방법입니다.
- **순차 Sidak(Sequential Sidak).** 순차 단계별로 낮아지는 거부 Sidak 프로시저로, 개별 가설은 거부하지만 동일한 전체 유의 수준을 유지한다는 점에서 훨씬 덜 보수적인 방법입니다.

최소유의차 방법은 순차 Sidak 방법보다 덜 보수적이므로 순차 Bonferroni 방법보다도 덜 보수적입니다. 다시 말해, 최소유의차는 적어도 순차 Sidak만큼 개별 가설을 거부하므로 순차 Bonferroni만큼 개별 가설을 거부하게 됩니다.

저장

선택한 항목이 지정한 이름으로 저장됩니다. 기존 필드 이름과 충돌하는 이름은 허용되지 않습니다.

예측값. 대상의 예측값을 저장합니다. 기본 필드 이름은 *PredictedValue*입니다.

범주형 대상의 예측 확률. 대상이 범주형인 경우, 이 키워드가 저장할 최대 범주로 지정된 값까지 첫 번째 n 범주의 예측 확률을 저장합니다. 계산된 값은 순서 대상의 누적 확률입니다. 기본 루트 이름은 *PredictedProbability*입니다. 예측 범주의 예측 확률을 저장하려면 신뢰도를 저장합니다(아래 참조).

신뢰구간. 예측값 또는 예측 확률에 대한 신뢰구간의 상한 및 하한을 저장합니다. 다항을 제외한 모든 분포에 대해 두 변수를 생성하고 기본 루트 이름은 *CI*이며 *_Lower* 및 *_Upper*가 접미사로 붙습니다.

다항 분포 및 명목 대상의 경우, 각 종속변수 범주에 대해 필드 하나가 생성됩니다. 이는 저장할 최대 범주로 지정된 값까지 첫 번째 n 범주의 예측 확률에 대한 상한 및 하한을 저장합니다. 기본 루트 이름은 *CI*이며 기본 필드 이름은 대상 범주의 순서에 따라 *CI_Lower_1*, *CI_Upper_1*, *CI_Lower_2*, *CI_Upper_2*...입니다.

다항 분포 및 순서 대상의 경우, 마지막을 제외한 각 종속변수 범주에 대해 필드 하나가 생성됩니다(자세한 정보는 68 페이지의 『일반 빌드 옵션』 주제를 참조하십시오). 저장할 최대 범주로 지정된 값까지(마지막 포함하지 않음) 첫 번째 n 범주의 누적 예측 확률의 상한과 하한을 저장합니다. 기본 루트 이름은 *CI*이며 기본 필드 이름은 대상 범주의 순서에 따라 *CI_Lower_1*, *CI_Upper_1*, *CI_Lower_2*, *CI_Upper_2*...입니다.

Pearson 잔차. 각 레코드에 대해 Pearson 잔차를 저장하며, 모형 적합의 사후 추정 진단에 사용될 수 있습니다. 기본 필드 이름은 *PearsonResidual*입니다.

신뢰도, 범주형 대상에 대한 예측값의 신뢰도를 저장합니다. 계산되는 신뢰도는 예측값의 확률(가장 높은 예측 확률) 또는 가장 높은 예측 확률과 두 번째로 가장 높은 예측 확률의 차이를 기준으로 할 수 있습니다. 기본 필드 이름은 *Confidence*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 보기

이 프로시저에서는 뷰어에 모형 오브젝트를 만듭니다. 이 오브젝트를 활성화(두 번 클릭)하면 모형에 대한 대화형 보기를 얻을 수 있습니다.

기본적으로 모형 요약 보기가 표시됩니다. 다른 모형 보기를 보려면 보기 축소판 그림에서 선택하십시오.

모형 오브젝트에 대안으로서 옵션 대화 상자(편집 > 옵션)의 결과 탭에 있는 결과 표시 그룹에서 피벗 테이블 및 차트를 선택하여 피벗 테이블 및 차트를 생성할 수 있습니다. 다음 항목은 모형 오브젝트를 설명합니다.

모형 요약

이 보기는 모형과 그 적합성을 한 눈에 파악할 수 있도록 요약한 스냅샷입니다.

테이블. 표는 대상 설정에 지정된 대상, 확률 분포 및 연결 함수를 식별합니다. 대상이 이벤트와 시행에 의해 정의되는 경우, 셀이 이벤트 필드 및 시행 필드 또는 고정 시행 수를 표시하기 위해 분할됩니다. 또한 유한 표본 수정된 Akaike 정보 기준(AICC) 및 베이지안 정보 기준(BIC)이 표시됩니다.

- *Akaike* 수정. -2(제한) 로그 우도에 기반한 혼합 모형을 선택 및 비교하기 위한 척도입니다. 값이 작을수록 모형이 우수함을 나타냅니다. AICC는 작은 표본 결과의 AIC를 "수정합니다". 표본 결과가 증가함에 따라 AICC는 AIC로 수렴됩니다.
- *베이지안*. -2 로그 우도에 기반한 모형을 선택 및 비교하기 위한 척도입니다. 값이 작을수록 모형이 우수함을 나타냅니다. 또한 BIC는 초과 모수화된 모형에 페널티를 부여하며 AIC보다 훨씬 엄격합니다.

차트. 대상이 범주형인 경우 차트는 최종 모형의 정확도를 표시하며 이는 정확한 분류의 백분율입니다.

데이터 구조

이 보기는 지정된 데이터 구조의 요약을 제공하며 개체와 반복 측도가 올바르게 지정되었는지 확인할 수 있도록 합니다. 첫 번째 개체에 대한 관측 정보가 각 재체 필드와 반복 측도 필드 및 대상에 대해 표시됩니다. 또한 각 개체 필드와 반복 측도 필드에 대한 수준 수가 표시됩니다.

관찰에 의한 예측

이벤트/시행으로 지정된 대상을 포함하여 연속형 대상의 경우, 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.

분류

범주형 대상의 경우, 정확한 전체 퍼센트와 함께 관측값 대 예측값의 교차 분류를 히트 맵에 표시합니다.

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

- 행 퍼센트. 셀의 행 백분율(전체 행의 백분율로 표시되는 셀 개수)을 표시합니다. 기본값입니다.
- 셀 개수. 셀의 셀 개수를 표시합니다. 히트 맵의 음영이 행 백분율의 기준입니다.
- 히트 맵. 셀의 값은 표시하지 않고 음영만 표시합니다.
- 압축. 셀의 행 또는 열 머리말, 값을 표시하지 않습니다. 대상에 범주가 많은 경우에 유용할 수 있습니다.

결측. 레코드에 대상의 결측값이 있으면 모든 유효한 행 아래의 (결측) 행에 표시됩니다. 결측값이 있는 레코드는 정확한 전체 퍼센트에 기여하지 않습니다.

여러 대상. 여러 범주형 대상이 있는 경우, 각 대상이 별도의 테이블에 표시되고 표시되는 대상을 제어하는 대상 드롭다운 목록이 있습니다.

큰 테이블. 표시된 대상에 범주가 100개 이상 있으면 테이블이 표시되지 않습니다.

고정 효과

이 보기는 모형에서 각 고정 효과의 크기를 표시합니다.

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

- 다이어그램. 고정 효과 설정에 지정된 순서로 효과가 위에서 아래로 정렬되는 차트입니다. 다이어그램의 연결선은 효과 유의수준을 기준으로 가중되며 선이 굵을수록 더 유의한 효과(더 작은 p -값)입니다. 기본값입니다.
- 테이블. 전체 모형과 개별 모형 효과에 대한 ANOVA 테이블입니다. 고정 효과 설정에 지정된 순서로 각 효과가 위에서 아래로 정렬됩니다.

유의수준. 보기에 표시되는 효과를 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.

고정 계수

이 보기는 모형에서 각 고정 계수의 값을 표시합니다. 요인(범주형 예측변수)이 모형 내에서 코딩된 지표이므로 요인을 포함하는 효과에는 일반적으로 여러 관련 계수가 있으며, 중복 계수에 해당하는 범주를 제외하고 각 범주에 하나씩 있습니다.

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

- 다이어그램. 절편을 먼저 표시한 다음 고정 효과 설정에 지정된 순서로 효과가 위에서 아래로 정렬되는 차트입니다. 요인을 포함하는 효과 내에서 계수가 데이터 값의 오름차순으로 정렬됩니다. 다이어그램의 연결선은 계수 유의수준을 기준으로 색상이 지정되고 가중되며 선이 굵을수록 더 유의한 계수(더 작은 p -값)입니다. 기본값 유형입니다.

- **테이블.** 개별 모형 계수의 값, 유의수준 검정 및 신뢰구간을 표시합니다. 절편 다음으로, 고정 효과 설정에 지정된 순서로 효과가 위에서 아래로 정렬됩니다. 요인을 포함하는 효과 내에서 계수가 데이터 값의 오름차순으로 정렬됩니다.

다항. 효과에 다항 분포가 있는 경우, 다항 드롭다운 목록이 표시할 대상 범주를 제어합니다. 목록에서 값의 정렬 순서는 작성 옵션 설정의 지정 사항에 의해 결정됩니다.

지수. 이분형 로지스틱 회귀분석(이항 분포 및 로짓 링크), 명목 로지스틱 회귀분석(다항 분포 및 로짓 링크), 음수 이항 회귀분석(음수 이항 분포 및 로그 링크) 및 로그 선형 모형(포아송 분포 및 로그 링크) 등 특정 모형 유형에 대한 지수화한 계수 추정값 및 신뢰구간을 표시합니다.

유의수준. 보기에 표시되는 계수를 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.

랜덤 효과 공분산

이 보기는 랜덤 효과 공분산 행렬(**G**)을 표시합니다.

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

- **공분산 값.** 고정 효과 설정에 지정된 순서로 효과가 위에서 아래로 정렬되는 공분산 행렬의 히트 맵입니다. 셀 값에 해당하는 상관도표의 색은 키에 표시된 것과 같습니다. 기본값입니다.
- **상관도표.** 공분산 행렬의 히트 맵입니다.
- **압축.** 행 및 열 머리글이 없는 공분산 행렬의 히트 맵입니다.

블록. 여러 랜덤 효과 블록이 있는 경우, 표시할 블록을 선택할 수 있는 블록 드롭다운 목록이 있습니다.

그룹. 랜덤 효과 블록에 그룹 지정이 있는 경우, 표시할 그룹 수준을 선택할 수 있는 그룹 드롭다운 목록이 있습니다.

다항. 효과에 다항 분포가 있는 경우, 다항 드롭다운 목록이 표시할 대상 범주를 제어합니다. 목록에서 값의 정렬 순서는 작성 옵션 설정의 지정 사항에 의해 결정됩니다.

공분산 모수

이 보기는 잔차 및 랜덤 효과에 대한 공분산 모수 추정값과 관련 통계를 표시합니다. 이것은 공분산 구조가 적합한지에 대한 정보를 제공하는 고급이지만 기본적인 결과입니다.

요약표. 잔차(**R**) 및 랜덤 효과(**G**) 공분산 행렬에서 모수의 수, 고정 효과(**X**) 및 랜덤 효과(**Z**) 디자인 행렬의 순위(열 수), 데이터 구조를 정의하는 개체 필드에 의해 정의되는 개체 수에 대한 빠른 참조입니다.

공분산 모수 표. 선택된 효과에 대해, 각 공분산 모수의 추정값, 표준 오차 및 신뢰구간이 표시됩니다. 표시되는 모수의 수는 효과 및 랜덤 효과 블록에 대한 공분산 구조와 블록의 효과 수에 따라 다릅니다. 비대각선 모수가 유의하지 않은 경우, 더 간단한 공분산 구조를 사용할 수 있습니다.

효과. 랜덤 효과 블록이 있는 경우, 표시할 잔차 또는 랜덤 효과 블록을 선택할 수 있는 효과 드롭다운 목록이 있습니다. 잔차 효과는 항상 사용할 수 있습니다.

그룹. 잔차 또는 랜덤 효과 블록에 그룹 지정이 있는 경우, 표시할 그룹 수준을 선택할 수 있는 그룹 드롭다운 목록이 있습니다.

다항. 효과에 다항 분포가 있는 경우, 다항 드롭다운 목록이 표시할 대상 범주를 제어합니다. 목록에서 값의 정렬 순서는 작성 옵션 설정의 지정 사항에 의해 결정됩니다.

평균 추정: 유의한 효과

3원배치 상호작용으로 시작하여 이원 상호작용, 마지막으로 주효과로 끝나는 10개의 "가장 유의한" 고정 모든 요인 효과를 표시하는 차트입니다. 이 차트는 수직축에 대상의 모형 추정값을 표시하고 수평축에 주효과(또는 상호작용에서 첫 번째 나열된 효과)의 각 값을 표시합니다. 상호작용에서 두 번째 나열된 효과의 각 값에 대해 별도의 선이 만들어집니다. 3원배치 상호작용에서 세 번째 나열된 효과의 각 값에 대한 별도의 차트가 만들어집니다. 기타 모든 예측변수는 상수로 유지됩니다. 대상에서 각 예측변수 계수의 효과를 시각화하는 데 유용합니다. 유의한 예측변수가 없는 경우 평균 추정이 생성되지 않습니다.

신뢰도. 작성 옵션의 일부로 지정된 신뢰수준을 사용하여 주변 평균의 신뢰 한계 상한 및 하한을 표시합니다.

평균 추정: 사용자 정의 효과

사용자가 요청한 고정 모든 요인 효과에 대한 테이블과 차트입니다.

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

- **다이아그램.** 이 유형은 수직축에 대상의 모형 추정값 선형 차트를 표시하고 수평축에 주효과(또는 상호작용에서 첫 번째 나열된 효과)의 각 값을 표시합니다. 상호작용에서 두 번째 나열된 효과의 각 값에 대해 별도의 선이 만들어집니다. 3원배치 상호작용에서 세 번째 나열된 효과의 각 값에 대한 별도의 차트가 만들어집니다. 기타 모든 예측변수는 상수로 유지됩니다.

대비가 요청된 경우, 대비 필드의 수준을 비교하기 위해 또 다른 차트가 표시됩니다. 상호작용의 경우, 대비 필드 외에 효과의 각 수준 조합에 대해 차트가 표시됩니다. **대응별** 대비의 경우, 거리 네트워크 차트입니다. 즉 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.

편차 대비의 경우, 수직축에 대상의 모형 추정값을 표시하고 수평축에 대비 필드의 값을 표시하는 막대형 차트가 표시됩니다. 상호작용의 경우, 대비 필드 외에 효과의 각 수준 조합에 대해 차트가 표시됩니다. 막대는 대비 필드의 각 수준과 전체 평균 간의 차이를 표시하며, 검은색 가로선으로 나타냅니다.

단순 대비의 경우, 수직축에 대상의 모형 추정값을 표시하고 수평축에 대비 필드의 값을 표시하는 막대형 차트가 표시됩니다. 상호작용의 경우, 대비 필드 외에 효과의 각 수준 조합에 대해 차트가 표시됩니다. 막대는 대비 필드의 각 수준(마지막 제외)과 마지막 수준 간의 차이를 표시하며, 검은색 가로선으로 나타냅니다.

- **테이블.** 이 유형은 대상의 모형 추정값, 표준 오차 및 효과에서 필드의 각 수준 조합에 대한 신뢰구간의 테이블을 표시합니다. 기타 모든 예측변수는 상수로 유지됩니다.

대비가 요청된 경우, 추정값, 표준 오차, 유의수준 검정 및 각 대비에 대한 신뢰구간이 있는 또 다른 테이블이 표시됩니다. 상호작용의 경우, 대비 필드 외에 효과의 각 수준 조합에 대한 별도의 행 세트가 있습니다. 또한 전체 검정 결과가 있는 테이블이 표시됩니다. 상호작용의 경우, 대비 필드 외에 효과의 각 수준 조합에 대한 별도의 전체 검정이 있습니다.

신뢰도. 작성 옵션의 일부로 지정된 신뢰수준을 사용하여 주변 평균의 신뢰 한계 상한 및 하한 표시를 전환합니다.

윤곽. 대응별 대비 다이어그램의 윤곽을 전환합니다. 원 윤곽은 망 윤곽보다 대비를 덜 나타내지만 선이 겹쳐지지 않습니다.

제 9 장 모형선택 로그선형 분석

모형선택 로그선형 분석 프로시저에서는 다원배치 교차 분석표(분할표)를 분석합니다. 또한 반복 비율-적합 알고리즘을 사용하여 계층적 로그선형 모형을 다차원 교차 분석표에 적합시킵니다. 이 프로시저를 사용하면 결합된 범주형 변수를 알 수 있습니다. 모형 설정에 강제 입력과 후진제거법을 사용할 수 있습니다. 항등 모형의 경우 부분 연관에 대한 모수 추정값과 검정을 요청할 수 있습니다. 항등 모형에서는 모든 셀에 0.5를 더합니다.

예제. 두 가지 세제에 대한 사용자 선호도 연구에서 조사자는 각 그룹에 속한 사용자의 수를 세고 물의 연성(연함, 보통, 강함), 두 세제의 상표 중 이전에 사용한 상표, 세척수 온도(냉수나 온수) 등의 다양한 변수 범주를 조합합니다. 이를 통해 온도가 물의 연성과 상표 선호도와 어떤 관계를 있는지 알 수 있습니다.

통계. 빈도, 잔차, 모수 추정값, 표준 오차, 신뢰구간, 부분 연관에 대한 검정을 의미합니다. 사용자 정의 모형의 경우 잔차 도표와 정규 확률 도표를 의미합니다.

모형선택 로그선형 분석 데이터 고려 사항

데이터. 요인변수는 범주형입니다. 분석할 모든 변수는 숫자여야 합니다. 모형선택 분석을 시작하기 전에 범주형 문자열 변수를 숫자변수로 코딩변경할 수 있습니다.

수준이 많은 변수를 많이 지정하지 않도록 하십시오. 이러한 양식을 사용하면 다수의 셀이 소수의 관측값을 가지는 상황이 발생할 수 있으며 카이제곱 값을 사용할 수 없게 됩니다.

관련 프로시저. 모형선택 프로시저를 사용하면 모형에서 필요한 항을 쉽게 구분할 수 있습니다. 그런 다음 일반 로그선형분석이나 로짓 로그선형 분석을 사용하여 모형의 값을 계속 계산할 수 있습니다. 또한 자동 코딩변경을 사용하여 문자열 변수의 코딩을 변경할 수 있습니다. 숫자변수에 빈 범주가 있을 경우 코딩변경을 사용하여 연속적인 정수 값을 작성합니다.

모형선택 로그선형 분석 구하기

메뉴에서 다음을 선택합니다.

분석 > 로그선형 > 모형 선택...

1. 둘 이상의 숫자 범주형 요인을 선택합니다.
2. 요인 목록에서 하나 이상의 요인변수를 선택한 다음 범위지정을 클릭합니다.
3. 각 요인변수값에 대한 범위를 정의합니다.
4. 모형 작성에서 옵션을 선택합니다.

선택적으로 셀 가중 변수를 선택하여 구조적 0을 지정할 수 있습니다.

로그선형분석 범위의 지정

각 요인변수에 대한 범주의 범위를 표시해야 합니다. 최소값과 최대값은 요인변수의 최저 범주와 최대 범주에 해당합니다. 두 값은 모두 정수이어야 하며 최소값은 최대값보다 작아야 합니다. 경계를 벗어난 값을 포함하는 케이스는 제외됩니다. 예를 들어, 최소값을 1로, 최대값을 3으로 지정하면 1, 2, 3 값만 사용됩니다. 각 요인변수에 대해 이 과정을 반복합니다.

로그선형분석 모형

모형 지정. 항등 모형에는 모든 요인 주효과와 모든 요인 대 요인 상호작용이 있습니다. 사용자 정의를 선택하여 불항등 모형에 대한 생성 클래스를 지정합니다.

생성 클래스. 생성 클래스는 요인이 나타나는 최고수준 항의 목록을 나타냅니다. 계층적 모형은 생성 클래스와 관련 저차 사항을 모두 정의하는 항을 포함하고 있습니다. 요인 목록에서 변수 A , B , C 를 선택한 다음 항 설정 드롭다운 목록에서 상호작용을 선택한다고 가정합니다. 그 결과 생성된 모형에는 지정된 3원 배치 상호작용 $A*B*C$, 2원 배치 상호작용 $A*B$, $A*C$, 그리고 $B*C$, 그리고 A , B , C 에 대한 주 효과가 포함됩니다. 생성 클래스에 관련 저차 사항을 지정하지 마십시오.

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.

모형선택 로그선형 분석 옵션

출력. 빈도, 잔차 또는 둘 다 선택할 수 있습니다. 항등 모형의 경우 관측빈도와 예측빈도가 동일하며 잔차는 0과 같습니다.

도표. 사용자 정의 모형의 경우 잔차 및 정규확률의 두 가지 도표 유형 중 하나 또는 모두를 선택할 수 있습니다. 모형이 데이터에 얼마나 적합한지 판별하는데 도움이 됩니다.

항등 모형에 대한 출력. 항등 모형의 경우 모수 추정값을 선택할 수 있습니다. 모수 추정값을 사용하면 모형에서 삭제할 수 있는 항을 결정할 수 있습니다. 부분 연관에 대한 검정 방법을 나열하는 연관표도 사용할 수 있습니다. 이 옵션을 다수의 요인이 있는 테이블에 사용하면 계산하기 어렵습니다.

모형 설정 기준. 반복 비율-적합 알고리즘을 사용하여 모수 추정값을 구합니다. 최대반복수, 수렴 또는 델타(항 등 모형에 대한 모든 셀 빈도에 추가되는 값)를 지정하면 하나 이상의 추정 기준을 바꿀 수 있습니다.

HILOGLINEAR 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- CWEIGHT 하위 명령을 사용하여 행렬식에 셀 가중치를 지정할 수 있습니다.
- DESIGN 하위 명령을 사용하여 단일 명령으로 몇 가지 모형의 분석을 생성할 수 있습니다.

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

제 10 장 일반 로그선형분석

일반 로그선형분석 프로시저에서는 교차 분석표나 분할표에서 각 교차 분류 범주로 구분되는 관측값의 빈도 개수를 분석합니다. 테이블의 각 교차 분류를 사용하여 셀을 구성하고 각 범주형 변수를 요인이라고 합니다. 종속변수는 교차 분석표 셀의 케이스 수(빈도)를 나타내고 설명변수는 요인 및 공변량을 나타냅니다. 이 프로시저에서는 Newton-Raphson 방법을 사용하여 계층적 및 비계층적 로그선형 모형의 최대-우도 모수를 추정합니다. 포아송 분포나 다항분포를 분석할 수 있습니다.

최대 10개까지 요인을 선택하여 테이블의 셀을 정의할 수 있습니다. 셀 구조변수를 사용하면 불완전한 테이블에 대해 구조적 0을 정의하거나 모형에 오프셋 항을 포함시키거나 로그-비율 모형을 적용하거나 또는 주변 테이블들을 조정하는 방법을 이행할 수 있습니다. 대비 변수를 사용하면 일반화 로그-승산비(GLOR)를 계산할 수 있습니다.

모형 정보와 적합도 통계량은 자동으로 표시됩니다. 또한 다양한 통계와 도표를 출력하거나 활성 데이터 세트에 잔차와 예측값을 저장할 수 있습니다.

예제. 플로리다의 교통 사고에 관한 보고서 데이터는 안전 벨트 착용 여부와 부상 정도와의 관계를 결정하는데 사용됩니다. 승산비는 이 관계가 얼마나 중요한지를 나타냅니다.

통계. 관측빈도 및 기대빈도, 원시 잔차, 수정된 잔차, 편차 잔차, 디자인 행렬, 모수 추정값, 승산비, 로그-승산비, GLOR, Wald 통계, 신뢰구간을 선택할 수 있습니다. 도표에 대해 수정된 잔차, 편차 잔차, 정규확률을 선택할 수 있습니다.

일반 로그선형분석 데이터 고려 사항

데이터. 요인은 범주형이며 셀 공변량은 연속형입니다. 모형에 공변량이 있으면 셀의 케이스에 대한 평균 공변량 값이 해당 셀에 적용됩니다. 대비 변수는 연속형입니다. 이 대비 변수는 일반화 로그-승산비를 계산하는데 사용됩니다. 대비 변수의 값은 셀 기대개수의 로그선형결합에 대한 계수에 해당됩니다.

셀 구조변수에서는 가중치를 지정합니다. 예를 들어, 일부 셀이 구조적 0이면 셀 구조변수는 0이나 1의 값을 가집니다. 통합 데이터 가중치를 부여하는 데는 셀 구조변수를 사용하지 마십시오. 대신 데이터 메뉴에서 가중치 케이스를 선택합니다.

가정. 일반 로그선형분석에서는 포아송 분포와 다항분포 두 가지 분포를 사용할 수 있습니다.

포아송 분포의 가정은 다음과 같습니다.

- 연구 이전에 전체 표본 결과가 정해지지 않았거나 전체 표본 결과에 대한 분석 조건이 없습니다.
- 한 셀의 관측 이벤트는 다른 셀의 셀 개수에 대해 통계적으로 독립적입니다.

다항 분포 가정은 다음과 같습니다.

- 전체 표본 결과가 정해지거나 전체 표본 결과에 대한 분석 조건이 있습니다.

- 셀 개수가 통계적으로 독립적이지 않습니다.

관련 프로시저. 교차 분석표 프로시저를 사용하여 교차 분석표를 검토합니다. 하나 이상의 범주형 변수를 반응 변수로 간주하고 다른 변수를 설명변수로 간주하는 것이 타당하면 로짓 로그선형 프로시저를 사용합니다.

일반 로그선형분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 로그선형 > 일반...

2. 일반 로그선형분석 대화 상자에서 최대 10개까지 요인변수를 선택합니다.

선택적으로 다음을 수행할 수 있습니다.

- 셀 공변량을 선택합니다.
- 셀 구조변수를 선택하여 구조적 0을 정의하거나 오프셋 항을 포함시킵니다.
- 대비 변수를 선택합니다.

일반 로그선형분석 모형

모형 지정. 항등 모형에는 요인변수와 관련된 모든 주효과와 상호작용이 있습니다. 공변량 항은 포함되지 않습니다. 사용자 정의를 선택하여 하나의 상호작용 서브세트만 지정하거나 요인-대-공변량 상호작용을 지정합니다.

요인 및 공변량 요인 및 공변량이 나열됩니다.

모형에 포함된 항. 모형은 데이터의 특성에 따라 다릅니다. 사용자 정의를 선택한 다음 분석시 중요한 주효과와 상호작용을 선택할 수 있습니다. 모형에 포함할 모든 항을 지정해야 합니다.

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.

일반 로그선형분석 옵션

일반 로그선형분석 프로시저에서는 모형 정보와 적합도 통계량을 출력합니다. 또한 다음 중 하나 이상을 선택할 수 있습니다.

출력. 셀 관측빈도와 기대빈도, 원시 잔차, 수정된 잔차, 편차 잔차, 모형의 디자인 행렬, 모형에 대한 모수 추정값과 같은 여러 통계를 출력할 수 있습니다.

도표. 도표는 사용자 정의 모형에서만 사용할 수 있으며 두 개의 산점도 행렬(셀 관측개수와 기대개수에 대한 수정된 잔차나 편차 잔차)을 포함합니다. 또한 수정된 잔차나 편차 잔차의 정규 확률 도표와 디트렌딩된 정규 도표를 출력할 수 있습니다.

신뢰구간. 모수 추정값에 대한 신뢰구간을 조정할 수 있습니다.

기준. Newton-Raphson 방법은 최대 우도 모수 추정값을 구하는데 사용됩니다. 최대반복수, 수렴 기준, 델타 (초기 근사값에 대한 모든 셀에 추가되는 상수)에 대해 새로운 값을 입력할 수 있습니다. 델타는 항등 모형에 대한 셀에 남아 있습니다.

일반 로그선형분석 저장

활성 데이터 세트에 새로운 변수로 저장하려는 값을 선택합니다. 새 변수이름의 접미문자 n 은 차례로 증분하여 저장된 각 변수에 대한 고유 이름을 만듭니다.

데이터가 데이터 편집기에서 개별 관측값으로 코딩되는 경우에도 저장된 값은 통합 데이터(분할표의 셀)를 참조합니다. 비통합 데이터용으로 잔차나 예측값을 저장할 경우 분할표에 저장된 셀 값은 해당 셀의 각 케이스에 대해 데이터 편집기에 입력됩니다. 저장된 값의 의미를 이해하려면 데이터를 통합하여 셀 개수를 구해야 합니다.

잔차의 네 가지 유형인 원시 잔차, 표준화 잔차, 수정된 잔차 및 편차 잔차를 저장할 수 있습니다. 예측값도 저장할 수 있습니다.

- **잔차.** 단일 잔차나 원시 잔차라고도 하며 셀 관측개수와 셀 기대개수 간 차이를 나타냅니다.
- **표준화 잔차.** 표준 오차의 추정값으로 나눈 잔차입니다. 표준화 잔차는 Pearson 잔차라고도 합니다.
- **수정 잔차.** 표준화 잔차를 표준 오차 추정값으로 나눈 값입니다. 선택한 모형이 정확할 때 수정된 잔차는 표준 정규 분포에 가깝기 때문에 정규성을 확인하는 데 표준화 잔차보다 수정된 잔차가 사용됩니다.
- **편차 잔차.** 우도비 카이제곱 통계량(G 제곱)에 대한 부호가 있는 개별 기여도 제곱근입니다. 여기서 부호는 관측개수에서 기대개수를 뺀 잔차의 부호입니다. 편차 잔차는 점근 표준 정규 분포를 이룹니다.

GENLOG 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- GERESID 하위 명령을 사용하여 셀 관측빈도와 기대빈도의 선형 조합을 계산하고 해당 조합의 잔차, 표준화 잔차, 수정된 잔차를 인쇄합니다.
- CRITERIA 하위 명령을 사용하여 중복 확인에 사용하는 기본값 임계값을 변경합니다.
- PRINT 하위 명령을 사용하여 표준화 잔차를 출력합니다.

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

제 11 장 로짓 로그선형 분석

로짓 로그선형 분석 프로시저에서는 종속(또는 반응)변수와 독립(또는 설명)변수 간의 관계를 분석합니다. 독립 변수가 범주형(요인)이 될 수도 있는 반면 종속변수는 항상 범주형입니다. 기타 독립 변수(셀 공변량)는 연속형일 수 있지만 케이스마다 적용되지는 않습니다. 셀에 대한 가중 공변량 평균이 이러한 셀에 적용됩니다. 종속 변수의 승산 로그는 모수의 선형결합으로 표현됩니다. 이 경우 자동적으로 다항 분포가 가정되는데 이러한 모형을 다항 로짓 모형이라고도 합니다. 이 프로시저에서는 Newton-Raphson 알고리즘을 사용하여 로짓 로그선형 모형의 모수를 추정합니다.

조합된 종속변수와 요인변수를 1에서 10까지 선택할 수 있습니다. 셀 구조변수를 사용하면 불완전한 테이블에 대해 구조적 0을 정의하거나 모형에 오프셋 항을 포함시키거나 로그-비율 모형을 적용하거나 또는 주변 테이블들을 조정하는 방법을 이행할 수 있습니다. 대비 변수를 사용하면 일반화 로그-승산비(GLOR)를 계산할 수 있습니다. 대비 변수의 값은 셀 기대개수의 로그선형결합에 대한 계수에 해당됩니다.

모형 정보와 적합도 통계량은 자동으로 표시됩니다. 또한 다양한 통계와 도표를 출력하거나 활성 데이터 세트에 잔차와 예측값을 저장할 수 있습니다.

예제. 플로리다에서 219마리의 악어에 대한 연구가 수행되었습니다. 악어의 먹이 유형이 악어의 크기와 서식하는 네 곳의 호수에 따라 어떻게 달라질 수 있는지에 대한 연구였으며 연구 결과로는 어류보다 양서류를 더 좋아하는 작은 악어에 대한 승산이 큰 악어의 경우보다 0.70배 더 낮은 것으로 나타났고 어류대신 양서류를 주로 선택하는 승산은 호수 3에서 가장 높았습니다.

통계. 관측빈도 및 기대빈도, 원시 잔차, 수정된 잔차, 편차 잔차, 디자인 행렬, 모수 추정값, 일반화 로그 승산비, Wald 통계, 신뢰구간. 도표에 대해서는 수정된 잔차, 편차 잔차, 정규 확률 도표를 선택할 수 있습니다.

로짓 로그선형 분석 데이터 고려 사항

데이터. 종속변수는 범주형입니다. 요인은 범주형입니다. 셀 공변량은 연속형일 수 있으나 모형에 공변량이 있을 경우 한 셀의 케이스에 대한 평균 공변량 값을 해당 셀에 적용합니다. 대비 변수는 연속형입니다. 이러한 변수는 일반화 로그 승산비(GLOR)의 계산에 사용됩니다. 대비 변수의 값은 셀 기대개수의 로그선형결합에 대한 계수에 해당됩니다.

셀 구조변수에서는 가중치를 지정합니다. 예를 들어, 몇몇 셀이 구조적 0이면 셀 구조변수는 0이나 1의 값을 가집니다. 데이터 통합 가중치를 부여하는 데는 셀 구조변수를 사용하지 않습니다. 대신 데이터 메뉴에서 가중치 케이스를 사용합니다.

가정. 설명변수의 각 범주 조합내의 개수는 다항 분포를 나타내는 것으로 가정합니다. 다항 분포 가정은 다음과 같습니다.

- 전체 표본 결과가 정해지거나 전체 표본 결과에 대한 분석 조건이 있습니다.
- 셀 개수가 통계적으로 독립적이지 않습니다.

관련 프로시저. 교차 분석표 프로시저를 사용하여 분할표를 출력합니다. 하나의 관측개수와 설명변수 세트 간의 관계를 분석하려는 경우 일반 로그선형분석 프로시저를 사용합니다.

로짓 로그선형 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 로그선형 > 로짓...

2. 로짓 로그선형 분석 대화 상자에서 하나 이상의 종속변수를 선택합니다.

3. 하나 이상의 요인변수를 선택합니다.

종속 변수와 요인변수의 총 수는 10보다 작거나 같아야 합니다.

선택적으로 다음을 수행할 수 있습니다.

- 셀 공변량을 선택합니다.
- 셀 구조변수를 선택하여 구조적 0을 정의하거나 오프셋 항을 포함시킵니다.
- 하나 이상의 대비 변수를 선택합니다.

로짓 로그선형 분석 모형

모형 지정. 항등 모형에는 요인변수와 관련된 모든 주효과와 상호작용이 있습니다. 공변량 항은 포함되지 않습니다. 사용자 정의를 선택하여 하나의 상호작용 서브세트만 지정하거나 요인-대-공변량 상호작용을 지정합니다.

요인 및 공변량 요인 및 공변량이 나열됩니다.

모형에 포함된 항. 모형은 데이터의 특성에 따라 다릅니다. 사용자 정의를 선택한 다음 분석시 중요한 주효과와 상호작용을 선택할 수 있습니다. 모형에 포함할 모든 항을 지정해야 합니다.

종속 항의 가능한 모든 조합을 구하고 각 조합을 모형 목록의 각 항에 대응함으로써 항이 디자인에 추가됩니다. 종속변수에 대해 상수항 포함 옵션을 선택하면 단위 항 (1)도 모형 목록에 함께 추가됩니다.

예를 들어, 변수 $D1$ 과 $D2$ 를 종속변수로 가정합니다. 로짓 로그선형 분석 프로시저($D1$, $D2$, $D1*D2$)에 의해 종속 항 목록이 작성됩니다. 모형에 포함된 항 목록에 $M1$, $M2$, 상수가 포함되면 모형 목록에는 1, $M1$, $M2$ 가 포함됩니다. 결과적으로 작성된 디자인에는 각 모형 항과 각 종속 항의 조합이 포함됩니다.

$D1$, $D2$, $D1*D2$

$M1*D1$, $M1*D2$, $M1*D1*D2$

$M2*D1$, $M2*D2$, $M2*D1*D2$

종속변수에 대해 상수항 포함. 사용자 정의 모형에서 종속변수에 대한 상수항을 포함시킵니다.

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.

로짓 로그선형 분석 옵션

로짓 로그선형 분석 프로시저에서는 모형 정보와 적합도 통계량을 출력합니다. 또한 다음 옵션 중에서 하나 이상을 선택할 수 있습니다.

출력. 셀 관측빈도와 기대빈도, 원시 잔차, 수정된 잔차, 편차 잔차, 모형의 디자인 행렬, 모형에 대한 모수 추정값과 같은 여러 통계를 출력할 수 있습니다.

도표. 도표는 사용자 정의 모형에서만 사용할 수 있으며 두 개의 산점도 행렬(셀 관측개수와 기대개수에 대한 수정된 잔차나 편차 잔차)을 포함합니다. 또한 수정된 잔차나 편차 잔차의 정규 확률 도표와 디트렌딩된 정규 도표를 출력할 수 있습니다.

신뢰구간. 모수 추정값에 대한 신뢰구간을 조정할 수 있습니다.

기준. Newton-Raphson 방법은 최대 우도 모수 추정값을 구하는데 사용됩니다. 최대반복수, 수렴 기준, 델타(초기 근사값에 대한 모든 셀에 추가되는 상수)에 대해 새로운 값을 입력할 수 있습니다. 델타는 항등 모형에 대한 셀에 남아 있습니다.

로짓 로그선형 분석 저장

활성 데이터 세트에 새로운 변수로 저장하려는 값을 선택합니다. 새 변수이름의 접미문자 n 은 차례로 증분하여 저장된 각 변수에 대한 고유 이름을 만듭니다.

데이터가 데이터 편집기에 개별 관측값으로 코딩되는 경우에도 저장된 값은 통합 데이터(분할표의 셀)를 참조합니다. 비통합 데이터용으로 잔차나 예측값을 저장할 경우 분할표에 저장된 셀 값은 해당 셀의 각 케이스에 대해 데이터 편집기에 입력됩니다. 저장된 값의 의미를 이해하려면 데이터를 통합하여 셀 개수를 구해야 합니다.

잔차의 네 가지 유형인 원시 잔차, 표준화 잔차, 수정된 잔차 및 편차 잔차를 저장할 수 있습니다. 예측값도 저장할 수 있습니다.

- 잔차. 단일 잔차나 원시 잔차라고도 하며 셀 관측개수와 셀 기대개수 간 차이를 나타냅니다.
- 표준화 잔차. 표준 오차의 추정값으로 나눈 잔차입니다. 표준화 잔차는 Pearson 잔차라고도 합니다.
- 수정 잔차. 표준화 잔차를 표준 오차 추정값으로 나눈 값입니다. 선택한 모형이 정확할 때 수정된 잔차는 표준 정규 분포에 가깝기 때문에 정규성을 확인하는 데 표준화 잔차보다 수정된 잔차가 사용됩니다.
- 편차 잔차. 우도비 카이제곱 통계량(G 제곱)에 대한 부호가 있는 개별 기여도 제곱근입니다. 여기서 부호는 관측개수에서 기대개수를 뺀 잔차의 부호입니다. 편차 잔차는 접근 표준 정규 분포를 이룹니다.

GENLOG 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- GERESID 하위 명령을 사용하여 셀 관측빈도와 기대빈도의 선형 조합을 계산하고 해당 조합의 잔차, 표준화 잔차, 수정된 잔차를 인쇄합니다.
- CRITERIA 하위 명령을 사용하여 중복 확인에 사용하는 기본값 임계값을 변경합니다.
- PRINT 하위 명령을 사용하여 표준화 잔차를 출력합니다.

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

제 12 장 생명테이블

고용 기간(고용에서 퇴직까지의 시간)과 같은 두 이벤트 사이의 시간 분포를 검토하려는 여러 상황이 있습니다. 그러나 이러한 종류의 데이터에는 두 번째 이벤트가 코딩되지 않는 몇 가지 케이스가 있습니다(예: 연구 중반 시점에도 여전히 퇴직하지 않고 회사에 근무하는 사람들). 이는 몇 가지 이유로 발생할 수 있습니다. 연구가 끝나기 전에 해당 이벤트가 단순히 일어나지 않은 케이스도 있으며 연구가 끝나기 전에 조사자가 케이스를 계속 조사하지 못하여 그 상태를 모르게 되는 경우도 있으며 연구와는 상관없는 이유(피고용자가 병가를 얻어 결근한 경우)로 인해 연구를 계속할 수 없는 케이스가 있습니다. 이러한 케이스들을 총괄하여 **중도절단 케이스**라 하며 이 경우 t 검정이나 선형 회귀분석과 같은 전통적 기법으로 연구하기에 적합하지 않습니다.

이러한 유형의 데이터에 유용한 통계적 기법을 사후검토 **생명표**라고 합니다. 이 생명표의 기본 개념은 관측 기간을 더 작은 시간 구간으로 나누는 것입니다. 각 구간에서 적어도 그 기간 이상 관측된 모든 사람들은 해당 구간에 발생하는 종료 이벤트의 확률 계산에 적용됩니다. 각 구간에서 추정된 확률은 다른 시점에서 발생하는 이벤트의 전체 확률을 추정하는 데 사용됩니다.

예제. 기존의 니코틴 패치 치료 요법보다 향상된 새로운 니코틴 패치 요법이 흡연자의 금연에 도움이 되었습니까? 한 그룹은 기존의 치료요법을 받는 그룹, 다른 그룹은 새로운 치료요법을 받는 그룹을 통해 실험할 수 있습니다. 이 데이터에 의해 생명표를 작성하면 두 그룹 간의 전체 금연 비율을 비교하여 실험 요법이 전통 요법보다 더 나은지 판별할 수 있습니다. 또한 생존 함수나 위험 함수를 도표로 작성하여 보다 자세한 내용을 시각적으로 비교할 수 있습니다.

통계. 참여자 수, 빠져나간 수, 위험에 노출된 수, 종료 이벤트의 수, 종료 비율, 생존 비율, 누적 생존 비율(및 표준 오차), 확률 밀도(및 표준 오차), 각 그룹의 매 시간 구간별 위험률(및 표준 오차); 각 그룹의 평균 생존 시간; 그룹간의 생존 분포를 비교하기 위한 Wilcoxon(Gehan) 검정을 선택할 수 있습니다. 도표에 대해 생존 함수, 로그생존 함수, 밀도 함수, 위험 함수, 1-생존 함수 도표를 선택할 수 있습니다.

생명표 데이터 고려 사항

데이터. 시간변수는 양적변수여야 합니다. 이벤트를 단일값이나 연속 값의 범위로 코딩한 경우 상태변수는 정수로 코딩한 이분형이나 범주형이어야 합니다. 요인변수는 정수로 코딩한 범주형이어야 합니다.

가정. 중요 이벤트에 대한 확률은 초기 이벤트 이후에만 좌우되므로 절대 시간과 비교하여 안정적인 것으로 가정합니다. 즉, 여러 다른 시점에서 연구를 시작한 케이스들(예 : 다른 시점에서 치료를 받기 시작한 환자들)은 유사한 반응을 나타내야 합니다. 또한 중도절단 케이스와 중도절단되지 않은 케이스 사이에는 규칙적인 차이가 없어야 합니다. 예를 들어, 많은 중도절단 케이스가 보다 심각한 조건의 환자들이면 결과는 한 쪽으로 치우치게 됩니다.

관련 프로시저. 생명표 프로시저에는 이러한 종류의 분석(일반적으로 생존 분석으로 알려져 있음)에 대해 보편 통계 접근법을 사용합니다. Kaplan-Meier 생존 분석 프로시저는 관측 기간을 더 작은 시간 구간으로 분할하는 방법을 사용하지 않는 다소 다른 생명표 계산 방법을 사용합니다. 이 방법은 관측값의 수가 작아 각 생존

시간 구간에 소수의 관측값만 있는 경우에 사용하기 좋습니다. 공변량으로 제어하려는 생존 시간이나 변수와 관련된 것으로 추정되는 변수가 있으면 Cox 회귀분석 프로시저를 사용합니다. 공변량이 다른 시점에서 동일한 케이스에 대해 다른 값을 가질 수 있는 경우에는 Cox 회귀분석 시간-종속 공변량을 사용합니다.

생명표 작성

1. 메뉴에서 다음을 선택합니다.

분석 > 생존 > 생명표...

2. 하나의 숫자 생존변수를 선택합니다.
3. 검토할 시간 구간을 지정합니다.
4. 상태변수를 선택한 다음에 종료 이벤트가 발생한 케이스를 정의합니다.
5. 이벤트 정의를 클릭하여 이벤트가 발생했음을 나타내는 상태변수 값을 정의합니다.

선택적으로 1차 요인변수를 하나 선택할 수 있습니다. 그러면 생존변수에 대한 보험 통계표가 요인변수의 각 범주에 대해 생성됩니다.

요인 변수에 따른 2차 순서를 선택할 수 있습니다. 1차와 2차 요인변수의 모든 조합에 대해 생존변수에 대한 보험 통계표가 작성됩니다.

생명표 상태변수에 대한 이벤트 정의

선택한 값이나 상태변수에 대한 값이 나타나면 해당 케이스에 대한 종료 이벤트가 발생했음을 나타냅니다. 다른 케이스들은 모두 중도절단된 것으로 간주됩니다. 중요 이벤트를 나타내는 단일값이나 값의 범위를 입력합니다.

생명표 범위지정

지정한 범위 내의 요인변수 값이 있는 케이스가 분석에 포함되며 범위의 각 유일한 값에 대해 별도의 테이블 (요청되는 경우 도표)이 생성됩니다.

생명표 옵션

생명표 분석을 다양한 측면으로 제어할 수 있습니다.

생명표. 결과에 생명표를 출력하지 않으려면 생명표를 선택 해제합니다.

도표. 생존 함수 도표를 출력할 수 있습니다. 요인변수를 정의한 경우에는 요인변수로 정의한 각 하위 그룹에 대한 도표가 생성됩니다. 생존 함수, 로그생존 함수, 위험 함수, 밀도 함수, 1 - 생존 함수 도표를 사용할 수 있습니다.

- 생존. 선형 척도로 누적 생존 함수를 표시합니다.
- 로그 생존. 로그 척도에 누적 생존 함수를 표시합니다.
- 위험. 선형 척도에 누적 위험 함수를 표시합니다.

- 밀도, 밀도 함수를 표시합니다.
- 1 - 생존 함수. 선형 척도에 1 - 생존 함수를 도표화합니다.

처음 요인의 수준 비교. 1차 제어변수가 있는 경우에는 이 그룹에서 선택 사항을 하나 선택하여 하위 그룹의 생존확률을 비교하는 Wilcoxon (Gehan) 검정을 수행할 수 있습니다. 그러면 1차 요인에 대한 검정이 실행됩니다. 2차 요인을 정의한 경우에는 2차 변수의 각 수준에 대한 검정이 수행됩니다.

SURVIVAL 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 둘 이상의 종속변수를 지정합니다.
- 구간을 다르게 지정합니다.
- 둘 이상의 상태변수를 지정합니다.
- 일부 요인과 일부 제어변수만 포함하는 비교를 지정합니다.
- 정확한 비교값보다는 근사 비교값을 계산합니다.

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

제 13 장 Kaplan-Meier 생존 분석

고용 기간(고용에서 퇴직까지의 시간)과 같은 두 이벤트 사이의 시간 분포를 검토하려는 여러 상황이 있습니다. 그러나 이러한 종류의 데이터에는 대개 몇 가지 중도절단 케이스가 있습니다. 중도절단 케이스는 두 이벤트 중 두 번째 이벤트는 코딩되지 않는 케이스를 나타냅니다(예: 연구 중반 시점에도 여전히 퇴직하지 않고 회사에 근무하는 사람들). Kaplan-Meier 프로시저는 중도절단 케이스가 존재할 경우 이벤트-시간 모형을 추정하는 방법입니다. Kaplan-Meier 모형은 이벤트가 발생하는 각 시점에서 조건부 확률을 추정하고 그 확률들의 곱한계값을 구하여 각 시점에서 생존 비율을 추정하는 방법을 사용합니다.

예제. 새로운 AIDS 치료법이 생명 연장에 있어서 치료상의 이점이 있습니까? 전통적인 치료법을 사용하는 한 그룹과 실험적 치료법을 사용하는 다른 그룹으로 이루어진 두 그룹의 AIDS 환자를 통해 연구를 수행할 수 있습니다. 데이터로부터 하나의 Kaplan-Meier 모형을 만들어 두 그룹 간 전체 생존 비율을 비교함으로써 실험적 치료법이 전통적 치료법에 비해 개선책이 될 수 있는지 여부를 판별할 수 있습니다. 또한 생존 함수나 위험 함수를 도표로 작성하여 보다 자세한 내용을 시각적으로 비교할 수 있습니다.

통계. 시간, 상태, 누적 생존확률, 표준 오차를 포함하여 생존표, 누적 이벤트 및 잔여 수(표준 오차와 95%의 신뢰구간을 가지는 평균 및 중앙값 생존 시간)를 선택할 수 있습니다. 도표에 대해 생존 함수, 위험 함수, 로그 생존 함수, 1 - 생존 함수를 선택할 수 있습니다.

Kaplan-Meier 데이터 고려 사항

데이터. 시간변수는 연속형이어야 하며 상태변수는 범주형이나 연속형이 될 수 있습니다. 또한 요인변수와 층화변수는 범주형이어야 합니다.

가정. 중요 이벤트에 대한 확률은 초기 이벤트 이후에만 좌우되므로 절대 시간과 비교하여 안정적인 것으로 가정합니다. 즉, 여러 다른 시점에서 연구를 시작한 케이스들(예 : 다른 시점에서 치료를 받기 시작한 환자들)은 유사한 반응을 나타내야 합니다. 또한 중도절단 케이스와 중도절단되지 않은 케이스 사이에는 규칙적인 차이가 없어야 합니다. 예를 들어, 많은 중도절단 케이스가 보다 심각한 조건의 환자들이면 결과는 한 쪽으로 치우치게 됩니다.

관련 프로시저. Kaplan-Meier 프로시저는 각 이벤트 발생 시간에 생존 함수나 위험 함수를 추정하는 생명표의 계산 방법을 사용합니다. 생명표 프로시저는 관측 기간을 더 작은 시간 구간으로 분할하는 방법을 사용하는 생존 분석에 대해 보험 통계 접근법을 사용하며 대형 표본을 다루는데 유용할 수 있습니다. 공변량으로 제어하려는 생존 시간이나 변수와 관련된 것으로 추정되는 변수가 있으면 Cox 회귀분석 프로시저를 사용합니다. 공변량이 다른 시점에서 동일한 케이스에 대해 다른 값을 가질 수 있는 경우에는 Cox 회귀분석 시간-종속 공변량을 사용합니다.

Kaplan-Meier 생존 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 생존 > Kaplan-Meier...

2. 시간변수를 하나 선택하십시오.
3. 상태변수를 하나 선택하여 종료 이벤트가 발생한 케이스들을 구분합니다. 이 변수는 숫자나 단문자일 수 있습니다. 그런 다음 이벤트 정의

선택적으로 요인변수를 하나 선택하여 그룹 차이를 검토할 수 있습니다. 또한 변수의 각 수준(계층)에 대해 별도의 분석을 생성하는 층화변수를 선택할 수 있습니다.

Kaplan-Meier 상태변수에 대한 이벤트 정의

종료 이벤트가 발생했음을 나타내는 값을 입력합니다. 단일값, 값의 범위 또는 값 목록을 입력할 수 있습니다. 값의 범위 옵션은 상태변수가 숫자인 경우에만 사용할 수 있습니다.

Kaplan-Meier 요인 수준 비교

수준이 다른 요인의 생존 분포에 대한 등식을 검정하는 통계를 구할 수 있습니다. 사용 가능한 통계는 Log 순위, Breslow, Tarone-Ware입니다. 비교 방법을 지정하려면 모든 계층들 풀링, 각 계층별, 모든 계층의 대응별 또는 각 계층의 대응별 등 선택 중 하나를 선택합니다.

- 로그 순위. 생존 분포의 등식을 비교하는 검정을 수행합니다. 이 검정에서는 모든 시점이 동일하게 가중됩니다.
- Breslow. 생존 분포의 등식을 비교하는 검정을 수행합니다. 각 시점에서 위험 상태의 케이스 수에 따라 시점이 가중됩니다.
- Tarone-Ware. 생존 분포의 등식을 비교하는 검정을 수행합니다. 시점은 각 시점에서 위험 상태에 있는 케이스 수의 제곱근으로 가중됩니다.
- 모든 계층들 풀링. 단일 검정의 모든 요인 수준을 비교하여 생존 곡선의 등식을 검정합니다.
- 모든 계층의 대응별. 별개의 대응 요인 수준을 각각 비교합니다. 대응별 추세 검정은 사용할 수 없습니다.
- 각 계층별. 각 계층별 모든 요인 수준에 대해 별도의 등식 검정을 수행합니다. 층화변수가 없으면 검정은 수행되지 않습니다.
- 각 계층의 대응별. 각 계층에 대한 별개의 대응 요인 수준을 각각 비교합니다. 대응별 추세 검정은 사용할 수 없습니다. 층화변수가 없으면 검정은 수행되지 않습니다.

요인 수준들의 선형추세. 여러 요인 수준의 선형추세를 검정할 수 있습니다. 이 옵션은 요인 수준의 전체 비교에(쌍대 비교가 아닌) 대해서만 사용할 수 있습니다.

Kaplan-Meier 새 변수로 저장

Kaplan-Meier 테이블에서 얻은 정보를 새 변수로 저장할 수 있으며 후속 분석에 이 변수를 사용하여 가설을 검정하거나 가정을 확인할 수 있습니다. 또한 생존확률, 생존확률의 표준 오차, 위험, 누적 이벤트를 새 변수로 저장할 수 있습니다.

- 생존. 누적 생존 확률 추정값입니다. 기본값 변수 이름에는 시퀀스 번호와 함께 접두부 sur_가 붙습니다. 예를 들어, sur_1이 이미 존재하는 경우 Kaplan-Meier는 변수 이름에 sur_2를 할당합니다.
- 생존 표준 오차. 누적 생존 추정값의 표준 오차입니다. 기본값 변수 이름에는 시퀀스 번호와 함께 접두부 se_가 붙습니다. 예를 들어, se_1이 이미 존재하는 경우 Kaplan-Meier는 변수 이름에 se_2를 할당합니다.
- 위험. 누적 위험 함수 추정값입니다. 기본값 변수 이름에는 시퀀스 번호와 함께 접두부 haz_가 붙습니다. 예를 들어, haz_1이 이미 존재하는 경우 Kaplan-Meier는 변수 이름에 haz_2를 할당합니다.
- 누적 이벤트. 케이스가 생존 시간 및 상태 코드별로 정렬될 경우의 이벤트 누적 빈도입니다. 기본값 변수 이름에는 시퀀스 번호와 함께 접두부 cum_가 붙습니다. 예를 들어, cum_1이 이미 존재하는 경우 Kaplan-Meier는 변수 이름에 cum_2를 할당합니다.

Kaplan-Meier 옵션

Kaplan-Meier 분석에서 다양한 결과 유형을 요청할 수 있습니다.

통계. 계산된 생존 함수에 대해 생존표, 생존시간의 평균과 중앙값, 사분위수를 포함하는 통계가 출력되도록 선택할 수 있습니다. 요인변수를 포함시키면 각 그룹에 대해 별도의 통계가 생성됩니다.

도표. 도표를 사용하여 생존 함수, 1-생존 함수, 위험 함수, 로그생존 함수를 시각적으로 검토할 수 있습니다. 요인변수를 포함시키면 각 그룹에 대한 함수를 도표로 나타낼 수 있습니다.

- 생존. 선형 척도로 누적 생존 함수를 표시합니다.
- 1 - 생존 함수. 선형 척도에 1 - 생존 함수를 도표화합니다.
- 위험. 선형 척도에 누적 위험 함수를 표시합니다.
- 로그 생존. 로그 척도에 누적 생존 함수를 표시합니다.

KM 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 손실된 케이스를 중도절단 케이스에서 비롯된 별도의 한 범주로 간주하는 빈도표를 구합니다.
- 선형추세 검정에 대해 동일하지 않은 간격을 지정합니다.
- 생존 시간변수에 대해 사분위수가 아닌 백분위수를 구합니다.

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

제 14 장 Cox 회귀분석 분석

Cox 회귀분석은 시간-이벤트 데이터의 예측 모형을 작성합니다. 모형은 예측변수의 주어진 값에 대해 특정 시간 t 에 중요 이벤트가 발생했을 확률을 예측하는 생존 함수를 생성합니다. 생존 함수의 형태와 예측변수의 회귀계수는 관측된 개체에서 추정됩니다. 그런 다음 예측변수의 측정값을 갖고 있는 새로운 케이스에 모형을 적용할 수 있습니다. 관측하는 동안 중요 이벤트가 발생하지 않는 중도절단 개체의 정보는 모형 예측에 유용하게 기여합니다.

예제. 흡연에 따른 폐암 발생 위험도가 남자와 여자에서 차이를 나타내니까? 일일 흡연량과 성별을 공변량으로 입력하는 Cox 회귀 모형을 구성하여 폐암 발병 시기에 성별과 흡연량이 미치는 영향에 대한 가설을 검정할 수 있습니다.

통계. 각 모형에 대해 $-2LL$, 우도비 통계, 전체 카이제곱 통계를 구합니다. 모형에 있는 변수에 대해 모수 추정값, 표준 오차, Wald 통계를 구합니다. 모형에 없는 변수에 대해 스코어 통계 및 잔차 카이제곱 통계를 구합니다.

Cox 회귀분석 데이터 고려 사항

데이터. 시간변수는 양적변수여야 하지만 상태변수는 범주형이나 연속형이 될 수 있습니다. 독립변수(공변량)는 연속형이나 범주형이 될 수 있습니다. 독립변수가 범주형인 경우 코딩된 더미나 지표여야 합니다. 프로시저에 범주형 변수를 자동으로 코딩변경하는 옵션이 있습니다. 층화변수는 정수나 단문자로 코드화된 범주형이어야 합니다.

가정. 관측값은 독립적이어야 하며 위험 비율은 시간에 관계 없이 일정해야 합니다. 즉, 케이스마다 위험 비율이 시간에 따라 달라지면 안됩니다. 후자의 가정을 비례 위험 가정이라 합니다.

관련 프로시저. 비례 위험 가정이 적용되지 않는 경우(위 관련 항목) Cox 회귀분석 시간-종속 공변량 프로시저를 사용할 필요가 있습니다. 공변량이 없거나 범주형 공변량이 하나만 있으면 생명표나 Kaplan-Meier 프로시저를 사용하여 표본에 대한 생존확률이나 위험 함수를 검토할 수 있습니다. 표본에 중도절단 데이터가 없으면 (모든 케이스가 종료 이벤트에 도달한 경우) 선형 회귀분석 프로시저를 사용하여 예측변수와 시간-이벤트 간의 관계를 모형으로 만들 수 있습니다.

Cox 회귀분석 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 생존 > Cox 회귀분석...

2. 시간변수를 하나 선택하십시오. 시간값이 음수인 케이스는 분석되지 않습니다.
3. 상태변수를 선택한 다음 이벤트 정의를 클릭합니다.
4. 하나 이상의 공변량을 선택합니다. 상호작용 항을 포함시키려면 상호작용에 포함된 모든 변수를 선택한 다음 **>a*b>**를 클릭합니다.

선택적으로 층화변수를 정의하여 다른 그룹에 대한 개별 모형을 계산할 수 있습니다.

Cox 회귀분석 범주형 변수 정의

Cox 회귀분석 프로시저로 범주형 변수를 처리하는 방법의 세부 사항을 지정할 수 있습니다.

공변량. 기본 대화 상자에 지정된 모든 공변량의 목록이 레이어에 그 자체로 들어 있거나 한 상호작용의 일부로 들어 있습니다. 이러한 공변량 중 일부가 문자열 변수나 범주형인 경우 범주형 공변량으로만 사용할 수 있습니다.

범주형 공변량. 범주형으로 구분된 변수를 나열합니다. 각 변수에는 사용될 대비 코딩을 나타내는 표기법이 괄호에 포함되어 있습니다. 문자열 변수(변수이름 뒤에 기호 <가 붙음)는 이미 범주형 공변량 목록에 존재합니다. 공변량 목록에서 다른 범주형 공변량을 선택한 다음 범주형 공변량 목록으로 이동합니다.

대비 바꾸기. 대비 방법을 바꿀 수 있습니다. 사용할 수 있는 대비 방법은 다음과 같습니다.

- **지표.** 대비는 소속범주가 있는지 여부를 나타냅니다. 참조 범주는 대비 행렬에서 0행으로 표현됩니다.
- **단순.** 참조 범주를 제외한 예측변수의 각 범주는 참조 범주와 비교됩니다.
- **차이.** 처음 범주를 제외한 예측변수의 각 범주는 이전 범주의 평균 효과와 비교됩니다. 이것을 역 Helmert 대비라고도 합니다.
- **Helmert.** 마지막 범주를 제외한 예측변수의 각 범주는 후속 범주들의 평균 효과와 비교됩니다.
- **반복.** 처음 범주를 제외한 예측변수의 각 범주는 선행하는 범주와 비교됩니다.
- **다항.** 직교 다항 대비. 범주는 동일한 간격으로 떨어져 있어야 합니다. 다항 대비는 숫자변수에 대해서만 사용할 수 있습니다.
- **편차.** 참조 범주를 제외한 예측변수의 각 범주는 전체 효과와 비교됩니다.

편차, 단순 또는 지표를 선택하면 참조 범주로서 처음 또는 마지막 범주가 선택됩니다. 이 방법은 바꾸기를 누른 후에야 실제로 변경됩니다.

문자 공변량은 범주형 공변량이어야 합니다. 범주형 공변량 목록에서 문자열 변수를 제거하려면 기본 대화 상자의 공변량 목록에서 해당 변수를 포함하는 모든 항목을 제거해야 합니다.

Cox 회귀분석 도표

도표를 사용하면 추정된 모형을 계산하고 결과를 해석하는데 도움이 됩니다. 생존 함수, 위험 함수, 로그 - 로그, 1 - 생존 함수를 도표로 나타낼 수 있습니다.

- **생존.** 선형 척도로 누적 생존 함수를 표시합니다.
- **위험.** 선형 척도에 누적 위험 함수를 표시합니다.
- **로그 - 로그.** $\ln(-\ln)$ 변환이 추정값에 적용된 후의 누적 생존확률을 추정합니다.
- **1 - 생존 함수.** 선형 척도에 1 - 생존 함수를 도표화합니다.

이러한 함수는 공변량 값에 따라 다르므로 공변량에 대한 상수값을 사용하여 함수-시간을 도표로 나타내야 합니다. 값 변경 옵션에서 기본값으로 설정된 각 공변량 평균을 상수값으로 사용하거나 값 단추를 선택하여 사용자 고유 값을 입력할 수 있습니다.

범주형 공변량을 선구분 그룹화 변수 입력란으로 이동하면 해당 공변량의 각 값에 대한 개별 선을 도표로 나타낼 수 있습니다. 이 옵션은 도표화되는 공변량 값 목록에서 이름 뒤에 **(Cat)**로 표시되는 범주형 공변량에 대해서만 사용할 수 있습니다.

Cox 회귀분석 새 변수로 저장

다양한 분석 결과를 새 변수로 저장할 수 있습니다. 그런 다음 이러한 변수를 후속 분석에 사용하여 가설을 검정하거나 가정을 확인할 수 있습니다.

모형 변수 저장. 회귀 모형에 대한 생존 함수와 표준 오차, 로그-로그, 위험 함수, 편잔차, DFBETA를 저장하고 선형 예측변수 X^* 베타를 새 변수로 저장할 수 있습니다.

- **생존 함수.** 제공된 시간 동안 누적 생존 함수의 값입니다. 해당 기간에 대한 생존 확률에 해당합니다.
- **생존 함수 로그-로그.** $\ln(-\ln)$ 변환이 추정값에 적용된 후의 누적 생존확률을 추정합니다.
- **위험 함수.** 누적 위험 함수 추정값을 저장합니다. Cox-Snell 잔차라고도 합니다.
- **편잔차.** 비례 위험 가정을 검정하기 위해 생존 시간에 대한 편잔차를 도표화할 수 있습니다. 최종 모형에서 각 공변량마다 하나의 변수가 저장됩니다. 편잔차는 공변량이 하나 이상 포함되어 있는 모형에만 사용할 수 있습니다.
- **DFBETA.** 케이스가 제거된 경우 추정되는 계수의 변화량입니다. 최종 모형에서 각 공변량마다 하나의 변수가 저장됩니다. DFBETA는 공변량이 하나 이상 포함되어 있는 모형에만 사용할 수 있습니다.
- **X^* 베타(X^* Beta).** 선형 예측변수 스코어입니다. 평균 중심 공변량 값과 각 케이스에 대한 해당 모수 추정값 곱의 합입니다.

시간중속 공변량으로 Cox 회귀분석을 실행하는 경우 저장할 수 있는 유일한 변수는 DFBETA 및 선형 예측변수 X^* 베타입니다.

XML 파일에 모형정보 내보내기. 모수 추정값은 XML 형식으로 지정된 파일로 내보내집니다. 스코어링을 목적으로 이 모형 파일을 사용하여 모형 정보를 다른 데이터 파일에 적용할 수 있습니다.

Cox 회귀분석 옵션

분석과 결과를 다양한 측면으로 제어할 수 있습니다.

모형 통계. 추정값의 $\exp(B)$ 와 추정값의 상관에 대한 신뢰구간을 포함하여 모형 모수 통계를 구할 수 있습니다. 이러한 통계는 각 단계마다 또는 마지막 단계에서만 구할 수 있습니다.

단계선택에 대한 확률. 단계선택에 대한 확률을 선택한 경우 모형에 진입하는 확률이나 모형에서 제거되는 확률을 지정할 수 있습니다. 진입된 F 의 유의 수준이 진입값보다 작으면 변수가 입력되고 유의 수준이 제거값보다 크면 변수가 제거됩니다. 진입값은 제거값보다 작아야 합니다.

최대반복수. 모형에 대한 최대반복수를 지정할 수 있으므로 프로시저에서 해법을 찾는 데 걸리는 시간을 제어할 수 있습니다.

기준선 위험 함수 표시. 공변량의 평균에서 기준선 위험 함수와 누적 생존 함수를 출력할 수 있습니다. 이러한 표시는 시간-종속 공변량을 지정한 경우에는 사용할 수 없습니다.

Cox 회귀분석 상태변수에 대한 이벤트 정의

종료 이벤트가 발생했음을 나타내는 값을 입력합니다. 단일값, 값의 범위 또는 값 목록을 입력할 수 있습니다. 값의 범위 옵션은 상태변수가 숫자인 경우에만 사용할 수 있습니다.

COXREG 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 손실된 케이스를 중도절단 케이스에서 비롯된 별도의 한 범주로 간주하는 빈도표를 구합니다.
- 편차, 단순 및 지표 대비 방법에 대해 처음이나 마지막 이외의 참조 범주를 선택합니다.
- 다항 대비법 사용시 적용되는 범주들에 동일하지 않은 간격을 지정합니다.
- 추가 반복수 기준을 지정합니다.
- 결측값 처리를 제어합니다.
- 저장된 변수에 대한 이름을 지정합니다.
- 외부 IBM SPSS Statistics 데이터 파일에 결과를 기록합니다.
- 처리 중에 각 분할 파일 그룹에 대한 데이터를 외부 스크래치 파일에 보유합니다. 이렇게 하면 대량의 데이터 세트를 사용하여 분석을 실행할 경우 메모리 자원을 보존할 수 있습니다. 이 방법은 시간-종속 공변량과 함께 사용할 수 없습니다.

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

제 15 장 시간-종속 공변량 계산

Cox 회귀 모형 계산시 비례 위험 가정이 적용되지 않는 특정 상황이 있습니다. 즉, 위험 비율은 시간이 지남에 따라 변경되며 하나 또는 그 이상의 공변량에 대한 여러 값이 서로 다른 시점에서 서로 다르게 나타납니다. 이러한 경우 확장된 Cox 회귀 모형을 사용해야 하는데 이를 통해 시간-종속 공변량을 지정할 수 있습니다.

이러한 모형을 분석하려면 먼저 시간-종속 공변량을 정의해야 합니다. 다중 시간-종속 공변량은 명령 구문을 사용하여 지정할 수 있습니다. 이를 쉽게 하기 위해 시간을 나타내는 시스템 변수를 사용할 수 있습니다. 이 변수를 $T_$ 라고 합니다. 이 변수를 사용하여 다음의 두 가지 일반적 방법으로 시간-종속 공변량을 정의할 수 있습니다.

- 특정 공변량과 관련된 비례 위험 가정을 검정하거나 비례하지 않은 위험을 허용하는 확장 Cox 회귀 모형을 추정하는 경우 문제에서 시간-종속 공변량을 해당 시간변수 $T_$ 와 공변량의 함수로 정의할 수 있습니다. 일반적 예는 시간변수와 공변량의 단순 곱이지만 더 복잡한 함수도 지정할 수 있습니다. 시간-종속 공변량 계수의 유의수준을 검정하면 비례 위험 가정이 적합한지 알 수 있습니다.
- 일부 변수는 다른 시간 주기에서 다른 값을 가질 수 있으나 시간에 대해 체계적인 관계가 없습니다. 이러한 경우 논리 표현식을 사용하여 수행할 수 있는 구간별 시간-종속 공변량을 정의해야 합니다. 논리 표현식은 참인 경우 1의 값을, 거짓인 경우 0의 값을 취합니다. 일련의 논리 표현식을 사용하면 여러 측정에서 시간-종속 공변량을 작성할 수 있습니다. 예를 들어, 4주동안 일주일에 한 번씩 혈압을 측정한 경우($BP1$ 에서 $BP4$ 로 나타냄) 시간-종속 공변량을 $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \geq 2 \ \& \ T_ < 3) * BP3 + (T_ \geq 3 \ \& \ T_ < 4) * BP4$ 로 정의할 수 있습니다. 괄호로 묶인 항 중 정확히 한 개만 지정된 케이스에 대해 1과 같고 나머지는 0이 된다는 점에 유의하십시오. 이 함수는 시간이 1주일 미만이면 $BP1$ 을 사용하고, 1주일이 넘고 2주일 미만이면 $BP2$ 를 사용하고, 나머지도 마찬가지로 사용합니다.

시간-종속 공변량 계산 대화 상자에서 함수-설정 제어를 사용하여 시간-종속 공변량의 표현식을 설정하거나 $T_COV_$ 입력란에 직접 입력할 수 있습니다. 문자 상수는 따옴표나 어포스트로피로 묶어야 하며 숫자 상수는 점을 소수점 구분자로 사용하는 미국식으로 입력해야 합니다. 결과로 생성된 변수를 $T_COV_$ 라고 하며 Cox 회귀 모형에 공변량으로 포함됩니다.

시간-종속 공변량 계산

1. 메뉴에서 다음을 선택합니다.

분석 > 생존 > 시간종속 Cox 회귀분석...

2. 시간-종속 공변량에 대한 표현식을 입력합니다.

3. 모형을 클릭하여 Cox 회귀분석을 계속 수행합니다.

참고: 새로운 변수 $T_COV_$ 를 공변량으로 Cox 회귀 모형에 포함되었는지 확인하십시오.

자세한 정보는 97 페이지의 제 14 장 『Cox 회귀분석 분석』 주제를 참조하십시오.

시간-종속 공변량이 있는 Cox 회귀분석의 추가 기능

명령 구문을 사용하면 다중 시간-종속 공변량도 지정할 수 있습니다. 다른 명령 구문 기능은 시간-종속 공변량의 포함 여부에 관계 없이 Cox 회귀분석에 대해 사용할 수 있습니다.

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

제 16 장 범주형 변수 코딩 체계

대부분의 프로시저에서 범주형 독립변수가 대비 변수 세트에 자동으로 대체되도록 요청하여 블록과 같은 형식으로 방정식에서 변수를 입력하거나 제거할 수 있습니다. 일반적으로 CONTRAST 하위 명령에서 대비 변수 세트 코딩 방법을 지정할 수 있습니다. 이 부록에는 CONTRAST에 요청된 다른 대비 유형이 실제로 어떻게 수행되는지에 대해 설명되어 있습니다.

편차

총 평균의 편차. 행렬 항에서 이러한 대비의 형식은 다음과 같습니다.

```
mean ( 1/k  1/k  ...  1/k  1/k)
df(1) (1-1/k -1/k  ... -1/k -1/k)
df(2) (-1/k  1-1/k  ... -1/k -1/k)
      .
      .
df(k-1) (-1/k -1/k  ... 1-1/k -1/k)
```

여기서 k 는 독립변수에 대한 범주의 수이며 마지막 범주는 기본적으로 생략됩니다. 예를 들어, 세 개의 범주가 있는 독립변수에 대한 편차 대비는 다음과 같습니다.

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3  2/3 -1/3)
```

마지막 범주 이외의 다른 범주를 생략하려면 DEVIATION 키워드 다음의 괄호 안에 몇 번째 범주를 생략할지 숫자를 지정합니다. 예를 들어, 다음 하위 명령은 첫 번째 및 세 번째 범주에 대한 편차를 구하고 두 번째 범주를 생략합니다.

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

요인에 세 개의 범주가 있다고 가정합니다. 결과로 나타나는 대비 행렬은 다음과 같습니다.

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3  2/3)
```

단순

단순 대비. 요인의 각 수준을 마지막 범주와 비교합니다. 다음은 일반 행렬 형식입니다.

```
mean (1/k  1/k  ...  1/k  1/k)
df(1) ( 1  0  ...  0  -1)
df(2) ( 0  1  ...  0  -1)
      .
      .
df(k-1) ( 0  0  ...  1  -1)
```

여기서 k 는 독립변수에 대한 범주 수입니다. 예를 들어, 네 개의 범주가 있는 독립변수에 대한 단순 대비는 다음과 같습니다.

```
(1/4  1/4  1/4  1/4)
( 1  0  0  -1)
( 0  1  0  -1)
( 0  0  1  -1)
```

참조 범주로 마지막 범주 대신 다른 범주를 사용하려면 SIMPLE 키워드 다음의 괄호 안에 참조 범주의 시퀀스 번호를 지정합니다. 여기서 참조 범주의 시퀀스 번호는 반드시 해당 범주와 연관된 값일 필요는 없습니다. 예를 들어, 다음 CONTRAST 하위 명령은 두 번째 범주를 생략하는 대비 행렬을 구합니다.

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

요인에 네 개의 범주가 있다고 가정합니다. 결과로 나타나는 대비 행렬은 다음과 같습니다.

```
(1/4  1/4  1/4  1/4)
(  1  -1   0   0)
(  0  -1   1   0)
(  0  -1   0   1)
```

Helmert

Helmert 대비. 독립변수의 범주와 후속 범주의 평균을 비교합니다. 다음은 일반 행렬 형식입니다.

```
mean (1/k      1/k      ...      1/k      1/k      1/k)
df(1) (  1  -1/(k-1)  ...  -1/(k-1)  -1/(k-1)  -1/(k-1))
df(2) (  0      1      ...  -1/(k-2)  -1/(k-2)  -1/(k-2))
      .
      .
df(k-2) (  0      0      ...      1      -1/2      -1/2)
df(k-1) (  0      0      ...      0      1      -1)
```

여기서 k 는 독립변수에 대한 범주 수입니다. 예를 들어, 네 개의 범주가 있는 독립변수는 다음 형식의 Helmert 대비 행렬을 가지고 있습니다.

```
(1/4  1/4  1/4  1/4)
(  1  -1/3  -1/3  -1/3)
(  0   1  -1/2  -1/2)
(  0   0   1   -1)
```

차이

차이 또는 역 Helmert 대비. 독립변수의 범주와 변수의 이전 범주의 평균을 비교합니다. 다음은 일반 행렬 형식입니다.

```
mean ( 1/k      1/k      1/k      ...  1/k)
df(1) (  -1      1      0      ...  0)
df(2) (  -1/2    -1/2    1      ...  0)
      .
      .
df(k-1) (-1/(k-1)  -1/(k-1)  -1/(k-1)  ...  1)
```

여기서 k 는 독립변수에 대한 범주 수입니다. 예를 들어, 네 개의 범주가 있는 독립변수의 차이 대비는 다음과 같습니다.

```
( 1/4  1/4  1/4  1/4)
( -1   1   0   0)
(-1/2 -1/2  1   0)
(-1/3 -1/3 -1/3  1)
```

다항

직교 다항 대비. 첫 번째 자유도에는 모든 범주에 걸친 선형 효과가 포함되고 두 번째 자유도에는 2차 효과, 세 번째 자유도에는 3차 효과 등 차원이 더 높은 효과가 적용됩니다.

주어진 범주형 변수로 측정된 처리 방법 수준 간 간격을 지정할 수 있습니다. 기본값은 동일한 간격입니다. 메트릭을 생각하면 1에서 k 까지 연속적인 정수로 지정할 수 있습니다. 여기서 k 는 범주 수입니다. 변수 *drug*에 세 개의 범주가 있으면 하위 명령은 다음과 같습니다.

```
/CONTRAST(DRUG)=POLYNOMIAL
```

또는

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

일부 경우에는 동일한 간격이 필요하지 않습니다. 예를 들어, *drug*가 세 그룹에 대해 각각 다른 복용량을 지정한다고 가정합니다. 두 번째 그룹에 대해 투여한 복용량이 첫 번째 그룹에 대해 두 배이고 세 번째 그룹에 대해 투여한 복용량은 첫 번째 그룹에 대해 세 배이면, 처리 범주는 동일한 간격이며 이러한 상황에 적합한 메트릭은 연속적인 정수로 구성됩니다.

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

그러나 두 번째 그룹에 대해 투여한 복용량이 첫 번째 그룹에 대해 네 배이고 세 번째 그룹에 대해 투여한 복용량이 첫 번째 그룹에 대해 일곱 배이면, 적합한 메트릭은 다음과 같습니다.

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

두 가지 경우 모두 대비 지정 사항의 결과로 *drug*에 대한 첫 번째 자유도에는 복용량 수준에 대한 선형 효과가 포함되며 두 번째 자유도에는 2차 효과가 포함됩니다.

다항 대비는 추세 검정 및 표면상의 반응 특성을 조사하는 경우에 특히 유용합니다. 또한 다항 대비를 사용하여 곡선형 회귀분석과 같은 비선형 곡선형 차트 적합 모형을 수행할 수 있습니다.

반복

독립변수의 근접도 비교. 다음은 일반 행렬 형식입니다.

```
      mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
      .
      .
df(k-1) ( 0 0 0 ... 1 -1)
```

여기서 k 는 독립변수에 대한 범주 수입니다. 예를 들어, 네 개의 범주가 있는 독립변수에 대한 반복 대비는 다음과 같습니다.

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

이러한 대비는 프로파일 분석 및 다양한 스코어가 필요한 경우에 특히 유용합니다.

특수

사용자 정의 대비. 주어진 독립 변수가 많을 경우 필요에 따라 여러 행 및 열을 사용하여 제곱 행렬 형식으로 특수 대비를 입력할 수 있습니다. MANOVA 및 LOGLINEAR의 경우 첫 번째 행에는 항상 평균이나 대비, 효과가 입력됩니다. 주어진 변수가 있을 경우 이 행에는 주어진 변수와 비교하여 기타 독립 변수의 평균을 계산하는 방법이 표시됩니다. 일반적으로 이 대비는 변수의 벡터입니다.

행렬의 나머지 행에는 변수 범주 간 비교를 나타내는 특수 대비가 포함됩니다. 일반적으로 직교 대비가 가장 유용합니다. 직교 대비는 통계적으로 독립적이며 중복되지 않습니다. 다음과 같은 경우가 직교 대비입니다.

- 각 행에 대한 대비계수의 합계가 0입니다.
- 모든 대응 행의 해당 상관계수의 곱도 합계가 0입니다.

예를 들어, 네 가지 처리 수준이 있고 이러한 여러 처리 수준을 서로 비교하려는 경우 적합한 특수 대비는 다음과 같습니다.

```
(1  1  1  1)  weights for mean calculation
(3 -1 -1 -1)  compare 1st with 2nd through 4th
(0  2 -1 -1)  compare 2nd with 3rd and 4th
(0  0  1 -1)  compare 3rd with 4th
```

MANOVA, LOGISTIC REGRESSION 및 COXREG에 대해 다음 CONTRAST 하위 명령을 사용하여 지정합니다.

```
/CONTRAST(TREATMNT)=SPECIAL( 1  1  1  1 3 -1 -1 -1 0  2 -1 -1 0  0  1 -1 )
```

LOGLINEAR의 경우 다음을 지정해야 합니다.

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1  1  1  1 3 -1 -1 -1 0  2 -1 -1 0  0  1 -1 )
```

평균 행을 제외한 각 행의 합계는 0입니다. 각 대응 행의 곱도 합계가 0입니다.

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

특수 대비가 반드시 직교 대비가 아닐 수도 있습니다. 그러나 특수 대비 및 직교 대비는 서로 선형 결합이 아
니어야 합니다. 그런 경우 프로시저는 선형 종속성을 보고하고 처리를 중지합니다. Helmert, 차이 및 다항 대
비는 모두 직교 대비입니다.

지표

지시변수 코딩. 더미 코딩이라고도 하며 LOGLINEAR 또는 MANOVA에서 사용할 수 없습니다. 코딩된 새 변수의 수는 $k-1$ 입니다. 참조 범주에 있는 케이스는 모든 $k-1$ 변수에 대해 0으로 코딩됩니다. i 번째 범주에 있는 케이스는 i 번째가 1로 코딩되는 것을 제외하고 모든 지시변수에 대해 0으로 코딩됩니다.

제 17 장 공분산 구조

이 절에서는 공분산 구조에 대한 추가 정보를 제공합니다.

선행 종속성: 1차. 이 공분산 구조는 근접 요소 간의 이질적 분산 및 이질적상관을 포함하고 있습니다. 근접하지 않은 두 요소 간의 상관은 중요한 요소 간에 위치한 요소 간의 상관 곱입니다.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_4\sigma_1\rho_1\rho_2\rho_3$
$(\sigma_2\sigma_1\rho_1$	σ_2^2	$\sigma_3\sigma_2\rho_2$	$\sigma_4\sigma_2\rho_2\rho_3$
$(\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_3\sigma_2\rho_2$	σ_3^2	$\sigma_4\sigma_3\rho_3$
$(\sigma_4\sigma_1\rho_1\rho_2\rho_3$	$\sigma_4\sigma_2\rho_2\rho_3$	$\sigma_4\sigma_3\rho_3$	σ_4^2

AR(1). 동질적분산이 있는 1차 자기회귀 구조입니다. 모든 두 요소 간의 상관은 근접 요소에 대한 rho와 같습니다. 세 번째 분리된 요소에 대해서는 ρ^2 가 되며 이러한 형식으로 계속 적용됩니다. $-1 < \rho < 1$ 로 제한됩니다.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3$
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho^2$
$(\sigma^2\rho^2$	$\sigma^2\rho$	σ^2	$\sigma^2\rho$
$(\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	σ^2

AR(1): 이질적. 이질적 분산이 있는 1차 자기회귀 구조입니다. 두 요소 사이의 상관은 인접 요소의 경우 ρ 이며, 세 번째 요소로 분리된 두 요소의 경우 ρ^2 등의 방식으로 적용됩니다. -1 과 1 사이로 제한됩니다.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho^2$	$\sigma_4\sigma_1\rho^3$
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho^2$
$(\sigma_3\sigma_1\rho^2$	$\sigma_3\sigma_2\rho$	σ_3^2	$\sigma_4\sigma_3\rho$
$(\sigma_4\sigma_1\rho^3$	$\sigma_4\sigma_2\rho^2$	$\sigma_4\sigma_3\rho$	σ_4^2

ARMA(1,1). 1차 자기회귀 이동 평균 구조이며 동형분산을 포함하고 있습니다. 두 요소 간의 상관은 근접 요소에 대한 ρ 와 같습니다. 세 번째 분리된 요소에 대해서는 ρ^2 가 되며 이러한 형식으로 계속 적용됩니다. ϕ 와 θ 는 각각 자기회귀 모수 및 이동 평균 모수이며, 이들 값은 -1 에서 1 사이 값으로 제한됩니다.

$(\sigma^2$	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2$	$\sigma^2\phi\rho^3$
$(\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2$
$(\sigma^2\phi\rho^2$	$\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho$
$(\sigma^2\phi\rho^3$	$\sigma^2\phi\rho^2$	$\sigma^2\phi\rho$	σ^2

복합 대칭. 이 구조는 일정한 분산 및 일정한 공분산을 포함하고 있습니다.

$(\sigma^2 + \sigma_1^2$	σ_1	σ_1	σ_1
$(\sigma_1$	$\sigma^2 + \sigma_1^2$	σ_1	σ_1
$(\sigma_1$	σ_1	$\sigma^2 + \sigma_1^2$	σ_1
$(\sigma_1$	σ_1	σ_1	$\sigma^2 + \sigma_1^2$

복합 대칭: 상관 매트릭. 이 공분산 구조는 요소 간에 동질적분산 및 동질적상관을 포함하고 있습니다.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho$
$(\sigma^2\rho$	$\sigma^2\rho$	σ^2	$\sigma^2\rho$
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$	σ^2)

복합 대칭: 이질적. 이 공분산 구조는 요소 간에 이질적 분산 및 일정한 상관을 포함하고 있습니다.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho$)
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho$)
$(\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	σ_3^2	$\sigma_4\sigma_3\rho$)
$(\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	σ_4^2)

대각선. 이 공분산 구조는 요소 간에 이질적 분산 및 영상관을 포함하고 있습니다.

$(\sigma_1^2$	0	0	0)
(0	σ_2^2	0	0)
(0	0	σ_3^2	0)
(0	0	0	σ_4^2)

요인 분석: 1차. 이 공분산 구조는 요소 전체에서 이질적인 항과 동질적인 항으로 구성된 이질적 분산을 포함하고 있습니다. 모든 두 요소 간의 공분산은 요소의 이질적 분산 항에 대한 곱의 제곱근입니다.

$(\lambda_1^2 + d$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1$)
$(\lambda_2\lambda_1$	$\lambda_2^2 + d$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2$)
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d$	$\lambda_4\lambda_3$)
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d$)

요인 분석: 1차, 이질적. 이 공분산 구조는 요소 전체에서 이질적인 두 개의 항으로 구성된 이질적 분산을 포함하고 있습니다. 모든 두 요소 간의 공분산은 요소의 이질적 분산 항에 대한 첫 번째 곱의 제곱근입니다.

$(\lambda_1^2 + d_1$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1$)
$(\lambda_2\lambda_1$	$\lambda_2^2 + d_2$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2$)
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d_3$	$\lambda_4\lambda_3$)
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d_4$)

Huynh-Feldt. "원형" 행렬로 모든 두 요소 간의 공분산은 요소의 분산 평균에서 상수를 뺀 값과 같습니다. 분산 및 공분산 모두 상수가 아닙니다.

$(\sigma_1^2$	$[\sigma_1^2 + \sigma_2^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_4^2]/2 - \lambda$)
$([\sigma_1^2 + \sigma_2^2]/2 - \lambda$	σ_2^2	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda$)
$([\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	σ_3^2	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda$)
$([\sigma_1^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda$	σ_4^2)

척도법 항등. 이 구조는 일정한 분산을 포함하고 있습니다. 모든 요소 간에 어떠한 상관도 없다고 가정합니다.

$(\sigma^2$	0	0	0)
-------------	---	---	----

$$\begin{pmatrix} 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

Toeplitz. 이 공분산 구조는 요소 간에 동질적 분산 및 동질적 상관을 포함하고 있습니다. 근접 요소 간 상관은 근접 요소 대응과 동일합니다. 또한 제3자로 분리된 요소 간 상관도 동일합니다.

$$\begin{pmatrix} \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 & \sigma^2\rho_3 \\ \sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 \\ \sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1 \\ \sigma^2\rho_3 & \sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2 \end{pmatrix}$$

Toeplitz: 이질적. 이 공분산 구조는 요소 간에 이질적 분산 및 이질적상관을 포함하고 있습니다. 근접 요소 간 상관은 근접 요소 대응과 동일합니다. 또한 제3자로 분리된 요소 간 상관도 동일합니다.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_2 & \sigma_4\sigma_1\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_1 & \sigma_4\sigma_2\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_4\sigma_3\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{pmatrix}$$

비구조적. 일반 공분산 행렬입니다.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{2\ 1} & \sigma_{31} & \sigma_{41} \\ \sigma_{2\ 1} & \sigma_2^2 & \sigma_{32} & \sigma_{4\ 2} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{4\ 3} \\ \sigma_{41} & \sigma_{4\ 2} & \sigma_{4\ 3} & \sigma_4^2 \end{pmatrix}$$

비구조적: 상관 매트릭. 이 공분산 구조는 이질적 분산 및 이질적상관을 포함하고 있습니다.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \sigma_3\sigma_1\rho_{31} & \sigma_4\sigma_1\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_3\sigma_2\rho_{32} & \sigma_4\sigma_2\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_4\sigma_3\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{pmatrix}$$

분산 성분. 이 구조는 지정된 랜덤 효과 각각에 대한 척도법 항등 구조를 할당합니다.

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 예제에는 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 예제에는 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)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]

가중 예측값
GLM 8
GLM 반복 측도 20
감마 분포
일반화 선형 모형에서 35
일반화된 추정 방정식 49
개체변수
선형 혼합 모형 28
거듭제곱값 추정
GLM 다변량 9
GLM 반복 측도 21
경시적 모형
일반화 선형 혼합 모형 61
계층적 로그선형 모형 77
계층적 모형
일반화 선형 혼합 모형 61
계층적 분해 5, 16
분산 성분추정 25
고정 예측값
선형 혼합 모형 33
공변량
Cox 회귀분석 98
공분산 구조 107
선형 혼합 모형 107
공분산 모수 검정
선형 혼합 모형 32
공분산 분석
GLM 다변량 3
공분산 행렬
선형 혼합 모형 32
일반화 선형 모형에서 40, 42
일반화된 추정 방정식 54, 55
GLM 8
관측평균
GLM 다변량 9
GLM 반복 측도 21
교차 분석표
모형선택 로그선형 분석 77
교차곱
가설 및 오차행렬 9
구간별 시간-종속 공변량
Cox 회귀분석 101

기술통계
선형 혼합 모형 32
일반화 선형 모형에서 42
일반화된 추정 방정식 55
GLM 다변량 9
GLM 반복 측도 21

[나]

누적 로짓 연결 함수
일반화 선형 모형에서 35
일반화된 추정 방정식 49
누적 보 로그-로그 연결 함수
일반화 선형 모형에서 35
일반화된 추정 방정식 49
누적 음수 로그-로그 연결 함수
일반화 선형 모형에서 35
일반화된 추정 방정식 49
누적 프로빗 연결 함수
일반화 선형 모형에서 35
일반화된 추정 방정식 49
누적 Cauchit 연결 함수
일반화 선형 모형에서 35
일반화된 추정 방정식 49

[다]

다변량 분산 분석 3
다변량 회귀분석 3
다변량 GLM 3
다중 수준 모형
일반화 선형 혼합 모형 61
다항 로지스틱 회귀분석
일반화 선형 혼합 모형 61
다항 로짓 모형 85
다항 분포
일반화 선형 모형에서 35
일반화된 추정 방정식 49
단계 반분
선형 혼합 모형 32
일반화 선형 모형에서 40
일반화된 추정 방정식 54
대비
로짓 로그선형 분석 85

대비 (계속)
일반 로그선형분석 81
Cox 회귀분석 98
대비계수(L) 행렬
일반화 선형 모형에서 42
일반화된 추정 방정식 55
도표
로짓 로그선형 분석 87
일반 로그선형분석 83

[라]

랜덤 효과
선형 혼합 모형 31
랜덤 효과 사전 확률
분산 성분추정 24
랜덤 효과공분산 행렬
선형 혼합 모형 32
레버리지 값
일반화 선형 모형에서 44
GLM 8
GLM 반복 측도 20
로그 보 연결 함수
일반화 선형 모형에서 35
일반화된 추정 방정식 49
로그 연결 함수
일반화 선형 모형에서 35
일반화된 추정 방정식 49
로그 우도 수렴
선형 혼합 모형 32
일반화 선형 모형에서 40
일반화된 추정 방정식 54
로그선형분석 77
로짓 로그선형 분석 85
일반 로그선형분석 81
일반화 선형 혼합 모형 61
로지스틱 회귀분석
일반화 선형 혼합 모형 61
로짓 로그선형 분석 85
기준 87
대비 85
도표 87
모형 사양 86
변수 저장 87

로지 로그선형 분석 (계속)

- 셀 개수의 분포 85
- 셀 공변량 85
- 셀 구조 85
- 신뢰구간 87
- 예측값 87
- 요인 85
- 잔차 87
- 출력 옵션 87

로지 연결 함수

- 일반화 선형 모형에서 35
- 일반화된 추정 방정식 49

[마]

모수 공분산 행렬

- 선형 혼합 모형 32

모수 수렴

- 선형 혼합 모형 32
- 일반화 선형 모형에서 40
- 일반화된 추정 방정식 54

모수 추정값

- 로지 로그선형 분석 85
- 모형선택 로그선형 분석 78
- 선형 혼합 모형 32
- 일반 로그선형분석 81
- 일반화 선형 모형에서 42
- 일반화된 추정 방정식 55
- GLM 다변량 9
- GLM 반복 측도 21

모수효과

- 선형 혼합 모형 29

모형 보기

- 일반화 선형 혼합 모형 71

모형 정보

- 일반화 선형 모형에서 42
- 일반화된 추정 방정식 55

모형선택 로그선형 분석 77

- 명령 추가 기능 79
- 모형 78
- 옵션 78
- 요인 범위지정 78

문자 공변량

- Cox 회귀분석 98

[바]

반복

- 모형선택 로그선형 분석 78
- 일반화 선형 모형에서 40
- 일반화된 추정 방정식 54

반복 측도 변수

- 선형 혼합 모형 28

반복 히스토리

- 선형 혼합 모형 32
 - 일반화 선형 모형에서 42
 - 일반화된 추정 방정식 55
- 보 로그-로그 연결 함수
- 일반화 선형 모형에서 35
 - 일반화된 추정 방정식 49

분리

- 일반화 선형 모형에서 40
- 일반화된 추정 방정식 54

분산 동질성 검정

- GLM 다변량 9
- GLM 반복 측도 21

분산 분석

- 분산 성분추정 24
- 일반화 선형 혼합 모형 61
- GLM 다변량 3
- GLM 반복 측도 13

분산 성분

- 결과 저장 26
- 명령 추가 기능 26
- 모형 24
- 옵션 24

분할표

- 일반 로그선형분석 81

비례 위험 모형

- Cox 회귀분석 97

비정칙성 허용 오차

- 선형 혼합 모형 32

빈도분석

- 모형선택 로그선형 분석 78

[사]

사용자 정의 모형

- 모형선택 로그선형 분석 78
- 분산 성분추정 24
- GLM 반복 측도 16

삭제된 잔차

- GLM 8

삭제된 잔차 (계속)

- GLM 반복 측도 20

상관 행렬

- 선형 혼합 모형 32
- 일반화 선형 모형에서 42
- 일반화된 추정 방정식 55

상호작용 항 5, 16, 24, 78, 82, 87

- 선형 혼합 모형 29

생명테이블 89

- 도표 90

- 명령 추가 기능 91

- 생존 상태 변수 90

- 생존 함수 89

- 예제 89

- 요인 수준 비교 90

- 요인변수 90

- 위험률 89

- 테이블 출력하지 않음 90

- 통계 89

- Wilcoxon(Gehan) 검정 90

생성 클래스

- 모형선택 로그선형 분석 78

생존 분석

- 생명표 89
- 시간-중속 Cox 회귀분석 101
- Cox 회귀분석 97
- Kaplan-Meier 93

생존 함수

- 생명표 89

선형 혼합 모형 27, 107

- 공분산 구조 107

- 랜덤 효과 31

- 명령 추가 기능 34

- 모수효과 29

- 모형 32

- 상호작용 항 29

- 저장할 변수 33

- 주변 평균 추정값 33

- 추정 기준 32

- 항 설정 29, 30

스코어링

- 선형 혼합 모형 32

승산비

- 일반 로그선형분석 81

승산비 연결 함수

- 일반화 선형 모형에서 35
- 일반화된 추정 방정식 49

승수 연결 함수
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49
 신뢰구간
 로짓 로그선형 분석 87
 선형 혼합 모형 32
 일반 로그선형분석 83
 GLM 다변량 9
 GLM 반복 측도 21

[아]

에타제곱
 GLM 다변량 9
 GLM 반복 측도 21
 역 가우스 분포
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49
 연결 함수
 일반화 선형 혼합 모형 62
 예측값
 로짓 로그선형 분석 87
 선형 혼합 모형 33
 일반 로그선형분석 83
 완전 요인 모형
 분산 성분추정 24
 GLM 반복 측도 16
 요인
 GLM 반복 측도 15
 요인 수준 정보
 선형 혼합 모형 32
 우도 잔차
 일반화 선형 모형에서 44
 위험률
 생명표 89
 음수 로그-로그 연결 함수
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49
 음수 이항 분포
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49
 음수 이항 연결 함수
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49
 이항 분포
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49

일반 로그선형분석
 기준 83
 대비 81
 도표 83
 명령 추가 기능 83
 모형 사양 82
 변수 저장 83
 셀 개수의 분포 81
 셀 공변량 81
 셀 구조 81
 신뢰구간 83
 예측값 저장 83
 요인 81
 잔차 83
 출력 옵션 83
 일반 추정가능 함수
 일반화 선형 모형에서 42
 일반화된 추정 방정식 55
 일반선형모형
 일반화 선형 혼합 모형 61
 일반화 로그-승산비
 일반 로그선형분석 81
 일반화 선형 모형 35
 모형 내보내기 45
 모형 사양 39
 모형 유형 35
 반응 38
 범주형 요인에 대한 옵션 39
 변수를 활성 데이터 세트에 저장 44
 분포 35
 연결 함수 35
 예측변수 39
 이분형 반응의 참조 범주 38
 일반화 선형 혼합 모형 61
 주변 평균 추정값 43
 초기값 41
 추정 기준 40
 통계 42
 일반화 선형 혼합 모형 61
 고정 계수 72
 공분산 모수 73
 관찰에 의한 예측 71
 대상 분포 62
 데이터 구조 71
 랜덤 효과 66
 랜덤 효과 공분산 73
 랜덤 효과 블록 66
 모수효과 65, 72

일반화 선형 혼합 모형 (계속)
 모형 내보내기 70
 모형 보기 71
 모형 요약 71
 분류표 72
 분석 기준치 68
 사용자 정의 항 65
 연결 함수 62
 오프셋 68
 주변 평균 추정값 69
 평균 추정 74
 필드 저장 70
 일반화 추정 방정식 47
 모형 내보내기 58
 모형 사양 53
 모형 유형 49
 반응 51
 범주형 요인에 대한 옵션 52
 변수를 활성 데이터 세트에 저장 57
 예측변수 52
 이분형 반응의 참조 범주 51
 주변 평균 추정값 56
 초기값 55
 추정 기준 54
 통계 55

[자]

잔차
 로짓 로그선형 분석 87
 모형선택 로그선형 분석 78
 선형 혼합 모형 33
 일반 로그선형분석 83
 일반화 선형 모형에서 44
 일반화된 추정 방정식 57
 잔차 공분산 행렬
 선형 혼합 모형 32
 잔차 도표
 GLM 다변량 9
 GLM 반복 측도 21
 잔차 SSCP
 GLM 다변량 9
 GLM 반복 측도 21
 적합도
 일반화 선형 모형에서 42
 일반화된 추정 방정식 55
 정규 분포
 일반화 선형 모형에서 35

정규 분포 (계속)
 일반화된 추정 방정식 49

정규 확률 지표
 모형선택 로그선형 분석 78

제공합 5, 16
 가설 및 오차행렬 9
 분산 성분추정 25
 선형 혼합 모형 30

제한된 우도비 추정
 분산 성분추정 24

주변 평균 추정값
 선형 혼합 모형 33
 일반화 선형 모형에서 43
 일반화된 추정 방정식 56
 GLM 다변량 9
 GLM 반복 측도 21

중도절단 케이스
 생명표 89
 Cox 회귀분석 97
 Kaplan-Meier 93

중첩 항
 선형 혼합 모형 30
 일반화 선형 모형에서 39
 일반화된 추정 방정식 53

[차]

참조 범주
 일반화 선형 모형에서 38
 일반화된 추정 방정식 51, 52

척도 모수
 일반화 선형 모형에서 40
 일반화된 추정 방정식 54

최대-우도 추정
 분산 성분추정 24

최소유의차
 GLM 다변량 7
 GLM 반복 측도 19

[카]

케이스 처리 요약
 일반화 선형 모형에서 42
 일반화된 추정 방정식 55

[파]

편차 잔차
 일반화 선형 모형에서 44

평균-산포 지표
 GLM 다변량 9
 GLM 반복 측도 21

포아송 분포
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49

포아송 회귀분석
 일반 로그선형분석 81
 일반화 선형 혼합 모형 61

표준 오차
 GLM 8
 GLM 다변량 9
 GLM 반복 측도 20, 21

표준 편차
 GLM 다변량 9
 GLM 반복 측도 21

표준화 잔차
 GLM 8
 GLM 반복 측도 20

표준화하지 않은 잔차
 GLM 8
 GLM 반복 측도 20

프로빗 분석
 일반화 선형 혼합 모형 61

프로빗 연결 함수
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49

프로파일 지표
 GLM 다변량 7
 GLM 반복 측도 18

[하]

항 설정 5, 16, 24, 78, 82, 87

항등 모형
 모형선택 로그선형 분석 78

항등 연결 함수
 일반화 선형 모형에서 35
 일반화된 추정 방정식 49

혼합 모형
 일반화 선형 혼합 모형 61

1차 27

효과-크기 추정값
 GLM 다변량 9

효과-크기 추정값 (계속)
 GLM 반복 측도 21

후진제거법
 모형선택 로그선형 분석 77

B

Bartlett의 단위행렬 검정
 GLM 다변량 9

Bonferroni
 GLM 다변량 7
 GLM 반복 측도 19

Box의 M 검정
 GLM 다변량 9

Breslow 검정
 Kaplan-Meier 94

C

Cook의 거리
 일반화 선형 모형에서 44
 GLM 8
 GLM 반복 측도 20

Cox 회귀분석 97
 공변량 97
 기준선 함수 99
 단계선택 진입과 제거 99
 대비 98
 지표 98
 명령 추가 기능 100
 문자 공변량 98
 반복 99
 범주형 공변량 98
 새 변수 저장 99
 생존 상태변수 100
 생존 함수 99
 시간-종속 공변량 101, 102
 예제 97
 위험 함수 99
 이벤트 정의 100
 통계 97, 99
 편잔차 99
 DFBETA 99

D

Duncan의 다중 범위 검정
 GLM 다변량 7

Duncan의 다중 범위 검정 (계속)
 GLM 반복 측도 19

Dunnett의 C
 GLM 다변량 7
 GLM 반복 측도 19

Dunnett의 T 검정
 GLM 다변량 7
 GLM 반복 측도 19

Dunnett의 T3
 GLM 다변량 7
 GLM 반복 측도 19

F

Fisher 스코어링
 선형 혼합 모형 32

Fisher의 LSD
 GLM 다변량 7
 GLM 반복 측도 19

G

Gabriel의 쌍대 비교 검정
 GLM 다변량 7
 GLM 반복 측도 19

Games-Howell의 대응 비교 검정
 GLM 다변량 7
 GLM 반복 측도 19

Gehan 검정
 생명표 90

GLM
 변수 저장 8
 행렬 저장 8

GLM 다변량 3, 10
 공변량 3
 사후 검정 7
 옵션 9
 요인 3
 종속변수 3
 주변 평균 추정값 9
 진단 9
 출력 9
 프로파일 도표 7

GLM 반복 측도 13
 명령 추가 기능 22
 모형 16
 변수 저장 20
 사후 검정 19

GLM 반복 측도 (계속)
 옵션 21
 요인 정의 15
 주변 평균 추정값 21
 진단 21
 출력 21
 프로파일 도표 18

GLOR
 일반 로그선형분석 81

H

Hessian 수렴
 일반화 선형 모형에서 40
 일반화된 추정 방정식 54

Hochberg의 GT2
 GLM 다변량 7
 GLM 반복 측도 19

K

Kaplan-Meier 생존 분석 93
 도표 95
 명령 추가 기능 95
 사분위수 95
 새 변수 저장 94
 생존 상태 변수 94
 생존표 95
 예제 93
 요인 수준 비교 94
 요인 수준들의 선형추세 94
 이벤트 정의 94
 통계 93, 95
 mean and median survival time 95

L

L 행렬
 일반화 선형 모형에서 42
 일반화된 추정 방정식 55

Levene 검정
 GLM 다변량 9
 GLM 반복 측도 21

LM 검정
 일반화 선형 모형에서 42

Log 순위 검정
 Kaplan-Meier 94

M

Mauchly의 구형성 검정
 GLM 반복 측도 21

MINQUE
 분산 성분추정 24

N

Newman-Keuls
 GLM 다변량 7
 GLM 반복 측도 19

Newton-Raphson 방법
 로짓 로그선형 분석 85
 일반 로그선형분석 81

P

Pearson 잔차
 일반화 선형 모형에서 44
 일반화된 추정 방정식 57

R

Ryan-Einot-Gabriel-Welsch 다중 범위
 GLM 다변량 7
 GLM 반복 측도 19

Ryan-Einot-Gabriel-Welsch 다중 F
 GLM 다변량 7
 GLM 반복 측도 19

R-E-G-W의 F
 GLM 다변량 7
 GLM 반복 측도 19

R-E-G-W의 Q
 GLM 다변량 7
 GLM 반복 측도 19

S

Scheffé 검정
 GLM 다변량 7
 GLM 반복 측도 19

Sidak의 T 검정
 GLM 다변량 7
 GLM 반복 측도 19

SSCP
 GLM 다변량 9
 GLM 반복 측도 21

Student-Newman-Keuls

GLM 다변량 7

GLM 반복 측도 19

T

T 검정

GLM 다변량 9

GLM 반복 측도 21

Tamhane의 T2

GLM 다변량 7

GLM 반복 측도 19

Tarone-Ware 검정

Kaplan-Meier 94

Tukey의 정직유의차

GLM 다변량 7

GLM 반복 측도 19

Tukey의 B 검정

GLM 다변량 7

GLM 반복 측도 19

Tweedie 분포

일반화 선형 모형에서 35

일반화된 추정 방정식 49

W

Wald 통계

로지 로그선형 분석 85

일반 로그선형분석 81

Waller-Duncan T 검정

GLM 다변량 7

GLM 반복 측도 19

Wilcoxon 검정

생명표 90

