

***IBM SPSS Regression 26***

**IBM**

**참고**

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 29 페이지의 『주의사항』에 있는 정보를 확인하십시오.

**제품 정보**

이 개정판은 새 개정판에서 별도로 명시하지 않는 한, IBM® SPSS Statistics의 버전 26, 릴리스 0. 수정 0 및 모든 후속 릴리스와 수정에 적용됩니다.

# 목차

회귀분석 . . . . .	1	비선형 회귀분석: 모수 제한조건 . . . . .	18
이분형 로지스틱 회귀 모형에 대한 프로시저 선택	1	비선형 회귀분석: 새 변수로 저장 . . . . .	18
로지스틱 회귀분석 . . . . .	2	비선형 회귀분석: 옵션 . . . . .	19
로지스틱 회귀분석 규칙 설정 . . . . .	3	비선형 회귀분석 결과 해석 . . . . .	19
로지스틱 회귀분석 변수 선택 방법 . . . . .	3	NLR 명령 추가 기능 . . . . .	20
로지스틱 회귀분석 범주형 변수 정의 . . . . .	4	가중 추정 . . . . .	20
로지스틱 회귀분석 새 변수로 저장 . . . . .	5	가중 추정: 옵션 . . . . .	22
로지스틱 회귀분석 옵션 . . . . .	6	WLS 명령 추가 기능 . . . . .	22
LOGISTIC REGRESSION 명령 추가 기능 . . . . .	7	2단계 최소제곱 회귀분석 . . . . .	22
다항 로지스틱 회귀분석 . . . . .	7	2단계 최소제곱 회귀분석 옵션 . . . . .	23
다항 로지스틱 회귀 모형 . . . . .	8	2SLS 명령 추가 기능 . . . . .	23
다항 로지스틱 회귀분석 참조 범주 . . . . .	9	범주형 변수 코딩 체계 . . . . .	24
다항 로지스틱 회귀분석 통계 . . . . .	10	편차 . . . . .	24
다항 로지스틱 회귀분석 기준 . . . . .	11	단순 . . . . .	24
다항 로지스틱 회귀분석 옵션 . . . . .	11	Helmert . . . . .	25
다항 로지스틱 회귀분석 저장 . . . . .	12	차이 . . . . .	25
NOMREG 명령 추가 기능 . . . . .	13	다항 . . . . .	25
프로빗 회귀 . . . . .	13	반복 . . . . .	26
프로빗 회귀 범위지정 . . . . .	14	특수 . . . . .	26
프로빗 회귀 옵션 . . . . .	15	지표 . . . . .	27
PROBIT 명령 추가 기능 . . . . .	15	주의사항 . . . . .	29
비선형 회귀분석 . . . . .	15	상표 . . . . .	31
조건 논리(비선형 회귀분석) . . . . .	16	색인 . . . . .	33
비선형 회귀분석: 모수 . . . . .	17		
비선형 회귀분석 일반 모형 . . . . .	17		
비선형 회귀분석: 손실 함수 . . . . .	18		



---

## 회귀분석

다음 회귀분석 기능은 SPSS® Statistics Standard Edition 또는 회귀분석 옵션에 포함되어 있습니다.

---

### 이분형 로지스틱 회귀 모형에 대한 프로시저 선택

이분형 로지스틱 회귀 모형은 로지스틱 회귀분석 프로시저나 다항 로지스틱 회귀분석 프로시저를 사용하는 것이 적합할 수 있습니다. 각 프로시저에는 다른 프로시저에서는 사용할 수 없는 옵션이 있습니다. 중요한 이론적 특징은 로지스틱 회귀분석 프로시저에서는 공분산 패턴 수가 전체 케이스 수보다 작은지 여부와 데이터 입력 방법과 관계 없이 모든 예측, 잔차, 영향력 통계, 각 개별 케이스 수준의 데이터를 사용하여 적합도 검정을 생성합니다. 반면에 다항 로지스틱 회귀분석 프로시저에서는 예측변수에 대한 동일한 공분산 패턴을 사용하여 부모집단을 형성하도록 내부적으로 케이스를 통합하고 이러한 부모집단에 따른 예측, 잔차, 적합도를 생성합니다. 모든 예측변수가 범주형이거나 연속형 예측변수에서 값의 수가 제한되어 있는 경우에는 각 개별 공분산 패턴에 여러 케이스가 있도록 부모집단 접근법에서 유효한 적합도 검정과 정보를 제공하는 잔차를 생성할 수 있지만 개별 케이스 수준 접근법에서는 생성할 수 없습니다.

로지스틱 회귀분석에서는 다음의 고유한 기능을 제공합니다.

- 모형에 대한 Hosmer-Lemeshow 적합도 검정
- 단계선택 분석
- 모형 모수화를 정의하는 대비
- 분류에 대한 선택적 절단점
- 분류도표
- 하나의 케이스 세트에서 보유된 케이스 세트로 모형 적합
- 예측, 잔차, 영향력 통계 저장

다항 로지스틱 회귀분석에서는 다음의 고유한 기능을 제공합니다.

- 적합도 모형에 대한 Pearson 및 편차 카이제곱 검정
- 적합도 검정의 데이터 그룹에 대한 부모집단 지정 사항
- 부모집단에 의한 개수, 예측 개수, 잔차 목록
- 초과 산포에 대한 분산 추정값 수정
- 모수 추정값의 공분산 행렬
- 모수의 선형결합에 대한 검정
- 중첩 모형에 명시된 지정 사항
- 차이변수를 사용한 1-1 일치 조건 로지스틱 회귀 모형 적합

**참고:** 위와 같은 절차는 이항 분포 및 로짓 연결 함수가 있는 일반화 선형 모형과 같은 이분형 데이터 모형에 적합합니다. 다른 연결 함수가 사용자의 데이터에 더 적합한 경우, 일반화 선형 모형 절차를 사용하십시오.

**참고:** 반복된 이분형 데이터의 측정치가 있거나 상관분석 레코드가 있는 경우 일반화 선형 모형 또는 일반화 추정 방정식 절차를 검토하십시오.

---

## 로지스틱 회귀분석

로지스틱 회귀분석은 예측변수 세트의 값을 기준으로 결정되는 특성이나 결과가 있는지 여부를 예측하려는 상황에서 유용합니다. 이 회귀분석은 선형 회귀 모형과 유사하나 종속변수가 이분형인 모형에 적합합니다. 로지스틱 회귀계수는 모형의 각 독립변수에 대한 승산비를 추정하는데 사용될 수 있습니다. 로지스틱 회귀분석은 판별 분석보다는 좀 더 광범위한 조사에 적절합니다.

**예제.** 라이프스타일이 CHD(관상동맥성심장병)에 대한 위험 요인이 됩니까? 환자 표본의 흡연 상태, 식습관, 운동, 음주, CHD 상태를 측정한다면 네 가지 라이프스타일 변수를 사용하여 모형을 구축한 다음 환자 표본에 CHD가 발생할지 여부를 예측할 수 있습니다. 모형으로 각 요인에 대한 승산비 추정값들(예: 비흡연자들보다 흡연자들에게 CHD가 발병할 가능성이 얼마나 더 많은가)을 구할 수 있습니다.

**통계.** 각 분석에 대해 전체 케이스, 선택한 케이스, 유효 케이스를 선택할 수 있으며 각 범주형 변수에 대해 모수 코딩을 선택할 수 있습니다. 각 단계에서 입력되었거나 제거된 변수, 반복 히스토리, -2 로그-우도, 적합도, Hosmer-Lemeshow 적합도 통계량, 모형 카이제곱, 향상 카이제곱, 분류표, 변수 간의 상관, 관측그룹과 예측 확률 도표, 잔차 카이제곱을 선택할 수 있습니다. 방정식의 각 변수에 대해 계수( $B$ ),  $B$ 의 표준 오차, Wald 통계, 추정된 승산비( $\exp(B)$ ),  $\exp(B)$ 에 대한 신뢰구간, 항이 모형에서 제거된 경우의 로그-우도 통계를 구할 수 있습니다. 방정식에 없는 각 변수에 대해 스코어 통계량을 구할 수 있습니다. 각 케이스에 대해 관측그룹, 예측확률, 예측그룹, 잔차, 표준화 잔차 통계를 구할 수 있습니다.

**방법.** 변수의 블록 입력이나 단계적 방법으로 전진 조건, 전진 LR, 전진 Wald, 후진 조건, 후진 LR, 후진 Wald 중 하나를 사용하여 모형을 추정할 수 있습니다.

## 로지스틱 회귀 데이터 고려사항

**데이터.** 종속변수는 이분형이어야 합니다. 독립변수는 구간 수준이나 범주형이 될 수 있습니다. 독립변수가 범주형인 경우 코딩된 더미나 지표여야 합니다. 프로시저에 범주형 변수를 자동으로 코딩 변경하는 옵션이 있습니다.

**가정.** 로지스틱 회귀분석은 판별 분석에서 사용하는 동일한 개념의 분포 가정을 따르지 않습니다. 그러나 예측변수가 다변량 정규 분포를 따른다면 현재 해법이 좀더 안정적으로 적용될 수 있습니다. 또한 다른 형식의 회귀분석에서처럼 예측변수 간의 다중공선성은 편향된 추정값과 팽창한 표준 오차를 유도할 수 있습니다. 이 프로시저는 소속그룹이 실제로 범주형 변수일 때 가장 효과적입니다. 즉, 소속그룹

이 연속형 변수값에 따라 달라지는 경우(예: "높은 IQ" 대 "낮은 IQ") 연속형 변수를 통해 얻을 수 있는 다양한 정보를 활용하려면 선형 회귀분석을 사용하는 것이 좋습니다.

**관련 프로시저.** 산점도 프로시저를 사용하여 다중공선성 관련 데이터를 화면에 나타냅니다. 다변량 정규성과 등분산-공분산 행렬에 대한 가정이 충족되면 판별 분석 프로시저를 사용하여 해법을 빠르게 구할 수 있습니다. 모든 예측변수가 범주형이면 로그선형 프로시저도 사용할 수 있습니다. 종속변수가 연속형이면 선형회귀 프로시저를 사용합니다. 로지스틱 회귀분석 프로시저를 사용하여 저장된 확률을 도표화하는 ROC 곡선 프로시저를 사용할 수 있습니다.

## 로지스틱 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 회귀 > 이분형 로지스틱...

2. 이분형 종속변수를 하나 선택합니다. 이 변수는 숫자 또는 문자일 수 있습니다.
3. 하나 이상의 공분산을 선택합니다. 상호작용 항을 포함하려면 상호작용에 포함된 모든 변수를 선택한 다음 >a\*b>를 선택합니다.

집단(블록)에 변수를 입력하려면 블록에 대한 공변량을 선택한 후 다음을 클릭하여 새 블록을 지정합니다. 블록이 모두 지정될 때까지 이 과정을 반복합니다.

선택적으로 분석할 케이스를 선택할 수 있습니다. 선택변수를 선택한 다음 규칙 기준을 입력합니다.

## 로지스틱 회귀분석 규칙 설정

선택 규칙으로 정의된 케이스는 모형 추정에 포함됩니다. 예를 들어, 변수와 같음을 선택하고 값 5를 지정하면 선택된 변수가 5값을 가지는 케이스만 모형 추정에 포함됩니다.

통계와 분류결과는 선택한 케이스와 선택하지 않은 케이스에 대해 모두 생성됩니다. 생성된 모형에 대한 유효성 검증을 위해 이전 데이터를 기준으로 새로운 케이스를 분류하는 방법이나 데이터를 조정 및 검정 서브세트로 분류하는 방법을 제공합니다.

## 로지스틱 회귀분석 변수 선택 방법

방법을 선택하면 독립변수를 분석에 입력하는 방법을 지정할 수 있습니다. 다른 방법을 사용하여 같은 변수 세트에서 여러 회귀 모형을 작성할 수 있습니다.

### 방법

*Enter* 한 블록에 있는 모든 변수가 단일 단계에 입력되는 변수 선택 프로시저입니다.

전진 선택법(조건)(*Forward Selection (Conditional)*)

스코어 통계의 유의수준을 기반으로 진입 검정을 수행하고 조건 모수 추정값에 따른 우도비 통계의 확률을 기반으로 제거 검정을 수행하는 단계별 선택 방법입니다.

전진 선택법(우도비)(*Forward Selection (Likelihood Ratio)*)

스코어 통계의 유의수준을 기반으로 진입 검정을 수행하고 최대 편우도 추정값에 따른 우도비 통계의 확률을 기반으로 제거 검정을 수행하는 단계별 선택 방법입니다.

전진 선택법(Wald)(*Forward Selection (Wald)*)

스코어 통계의 유의수준을 기반으로 진입 검정을 수행하고 Wald 통계의 확률을 기반으로 제거 검정을 수행하는 단계별 선택 방법입니다.

후진제거법(조건)(*Backward Elimination (Conditional)*)

후진 단계 선택입니다. 조건부 모수 추정값에 기반한 우도비 통계의 확률을 토대로 제거 여부를 검정합니다.

후진제거법(우도비)(*Backward Elimination (Likelihood Ratio)*)

후진 단계 선택입니다. 최대 편우도 추정값에 기반한 우도비 통계의 확률을 토대로 제거 여부를 검정합니다.

후진제거법(Wald)(*Backward Elimination (Wald)*)

후진 단계 선택입니다. Wald 통계의 확률을 기반으로 제거 여부를 검정합니다.

결과의 유의수준은 단일 모형 적합을 기준으로 합니다. 따라서 일반적으로 단계적 방법을 사용하는 경우 유의수준은 유효하지 않습니다.

선택한 모든 독립변수는 단일 회귀 모형에 추가됩니다. 단, 다른 변수 서브세트에 대해 다른 입력 방법을 지정할 수 있습니다. 예를 들어, 단계별 선택을 사용하여 회귀 모형에 변수 블록을 입력하거나 전진 선택법을 사용하여 두 번째 블록을 입력할 수 있습니다. 회귀 모형에 두 번째 변수 블록을 추가하려면 추가(+) 제어를 클릭합니다.

## 로지스틱 회귀분석 범주형 변수 정의

로지스틱 회귀분석 프로시저가 범주형 변수를 처리하기 위한 다음의 세부 사항을 지정할 수 있습니다.

### 공변량

변수 대화 상자에 지정된 모든 공변량의 목록이 레이어에 그 자체로 들어 있거나 한 상호작용의 일부로 들어 있습니다. 이러한 공분산 중 일부가 문자열 변수나 범주형인 경우 범주형 공분산으로만 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 공분산 목록에 존재합니다. 공분산 목록에서 다른 범주형 공분산을 선택한 다음 범주형 공분산 목록으로 이동합니다.

### 대비 변경

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** 예측변수의 각 범주(참조 범주 제외)는 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** 예측변수의 각 범주(마지막 범주 제외)는 후속 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 지표**를 선택하면 참조 범주로서 **처음** 또는 **마지막** 범주가 선택됩니다.

문자 공분산은 범주형 공분산이어야 합니다. 범주형 공변량 목록에서 문자변수를 제거하려면 변수 대화 상자의 공변량 목록에서 해당 변수를 포함하는 모든 항을 제거해야 합니다.

## **로지스틱 회귀분석 새 변수로 저장**

로지스틱 회귀분석 결과를 활성 데이터 세트에 새 변수로 저장할 수 있습니다.

### **예측값**

모형으로 예측한 값을 저장합니다. 사용할 수 있는 옵션은 확률과 소속그룹입니다.

#### **확률(Probabilities)**

각 케이스마다 이벤트의 예측 발생 확률을 저장합니다. 결과의 테이블에 새로운 변수의 이름과 내용이 표시됩니다. "이벤트"는 더 큰 값을 가진 종속 변수의 카테고리입니다. 예를 들면, 종속 변수의 값이 0과 1이면, 카테고리 1의 예측 확률이 저장됩니다.

#### **예측 소속그룹(Predicted Group Membership)**

판별 스코어를 기반으로 가장 큰 사후 확률을 갖는 그룹입니다. 모형에서 케이스가 소속되어 있다고 예측하는 그룹입니다.

### **영향력**

예측값에 대한 케이스의 영향력을 측정하는 통계에서 얻은 값은 구해 저장합니다. 사용할 수 있는 옵션은 Cook의 거리, 레버리지 값, DFBETA입니다.

#### **Cook의 거리(Cook's)**

Cook의 영향력 통계에 대한 로지스틱 회귀분석 아날로그입니다. 특정 케이스를 회귀계수 계산에서 제외할 때 모든 케이스의 잔차가 얼마나 변경될 수 있는지에 대한 척도입니다.

#### **레버리지 값(Leverage Value)**

각 관측값이 모형의 적합도에 미치는 상대적인 영향을 나타냅니다.

## DFBETA

베타 값의 차이는 특정 케이스를 제외하여 나타난 회귀계수의 변화량을 나타냅니다. 값은 모형의 각 항에 대해 계산되며 상수를 포함합니다.

**잔차** 잔차를 저장합니다. 사용할 수 있는 옵션은 비표준화, 로짓, 스튜던트화, 표준화, 편차 등입니다.

### 표준화하지 않은 잔차(*Unstandardized Residuals*)

관측값과 모형에 의해 예측된 값 사이 차이입니다.

### 로짓 잔차(*Logit Residual*)

로짓 척도에서 예측된 경우 케이스의 잔차입니다. 로짓 잔차는 1에서 예측 확률을 뺀 값을 예측 확률에 곱한 값으로 나눈 잔차입니다.

### 스튜던트화 잔차(*Studentized Residual*)

케이스를 제외하는 경우 모형 편차에서의 변화량입니다.

### 표준화 잔차(*Standardized Residuals*)

표준 편차의 추정값으로 나눈 잔차입니다. Pearson 잔차라고도 하는 표준화 잔차는 평균이 0이고 표준 편차가 1이 됩니다.

### 편차(*Deviance*)

모형 편차를 기준으로 하는 잔차입니다.

## XML 파일로 모형정보 내보내기

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

## 로지스틱 회귀분석 옵션

로지스틱 회귀분석을 위한 다음 옵션을 지정할 수 있습니다.

### 통계 및 도표

통계와 도표를 요청할 수 있습니다. 사용할 수 있는 옵션은 분류도표, Hosmer-Lemeshow 적합도, 케이스별 잔차목록, 추정값들의 상관, 반복 히스토리,  $\exp(B)$ 에 대한 신뢰구간입니다. 표시 그룹의 옵션에서 하나를 선택하여 각 단계마다 통계와 도표를 표시하거나 마지막 단계에서 최종 모형에 대해서만 통계와 도표를 표시합니다.

### Hosmer-Lemeshow 적합도 통계량(*Hosmer-Lemeshow goodness-of-fit statistic*)

이 적합도 통계량은 로지스틱 회귀분석, 특히 연속형 공변량을 포함하는 모형과 표본 결과가 작은 연구에 전통적인 적합도 통계량보다 더 효과적입니다. 이 방법은 케이스를 위험도의 십분위수로 그룹화하고 각 십분위수 내의 관측 확률을 기대 확률에 비교합니다.

### 단계선택에 대한 확률

방정식에 변수를 입력하고 제거하는 기준을 제어할 수 있습니다. 변수 입력이나 제거에 대한 기준을 지정할 수 있습니다.

### 단계선택에 대한 확률(*Probability for Stepwise*)

스코어 통계량의 확률이 진입값보다 작을 때는 변수가 모형에 진입되며 제거값보다 크면 제거됩니다. 기본값 설정을 바꾸려면 진입 및 제거에 양수값을 입력합니다. 진입값은 제거값보다 작아야 합니다.

### 분류 분리점

케이스를 분류하기 위한 절단점을 결정할 수 있습니다. 분류 절단점보다 작은 예측값이 있는 케이스는 음수로 분류되고 분류 절사를 넘어서는 예측값이 있는 케이스는 양수로 분류됩니다. 기본값을 변경하려면 0.01과 0.99 사이의 값을 입력합니다.

### 최대반복계산

모형 분석이 종결되기 전에 반복되는 최대 횟수를 변경할 수 있습니다.

### 모형에 상수 포함

모형에 상수항을 포함할지 여부를 지정할 수 있습니다. 이 옵션이 설정되지 않은 경우에는 상수항은 0이 됩니다.

## LOGISTIC REGRESSION 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 변수 값 또는 변수 레이블에 따라 케이스별 결과를 구분합니다.
- 반복 보고서의 간격을 제어합니다. 매 반복 후마다 모수 추정값을 인쇄하지 않고  $n$ 번째 반복 후마다 모수 추정값을 요청할 수 있습니다.
- 반복을 종료하고 중복을 점검하기 위한 기준을 변경합니다.
- 케이스별 목록에 대한 변수 목록을 지정합니다.
- 처리 중에 외부 스크래치 파일에 각 분할 파일 그룹에 대한 데이터를 보관하여 메모리를 보존합니다.

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

---

## 다항 로지스틱 회귀분석

다항 로지스틱 회귀분석은 예측자 변수 세트의 값에 따라 개체를 분류할 때 유용합니다. 다항 로지스틱 회귀분석은 로지스틱 회귀분석과 유사하지만 종속변수가 두 개의 범주로 제한되지 않으므로 로지스틱 회귀분석보다 일반적으로 사용할 수 있습니다.

**예제.** 영화를 흥행시킬 수 있도록 영화사에서는 관객이 선호하는 영화의 종류를 예측하려고 합니다. 이 경우 영화사에서는 다항 로지스틱 회귀분석을 실행하여 개인의 연령, 성별, 데이트 성향이 영화의 선호도에 미치는 영향력 정도를 측정할 수 있습니다. 이 결과에 따라 영화사에서는 특정 영화에 대해 그 영화를 볼 가능성이 높다고 예상되는 관객층을 겨냥하여 집중적으로 광고할 수 있습니다.

**통계.** 반복 히스토리, 모수 상관계수, 근사 공분산 및 상관행렬, 모형 및 부분 효과에 대한 우도비 검정,  $-2$  로그-우도를 선택할 수 있습니다. Pearson 및 편차 카이제곱 적합도, Cox 및 Snell, Nagelkerke,

McFadden  $R^2$ 를 선택할 수 있습니다. 분류: 반응 범주에 의한 관측빈도 대 예측빈도를 선택할 수 있습니다. 교차 분석표: 공분산 패턴 및 반응 범주에 의한 관측 빈도, 예측 빈도(잔차 포함) 및 비율을 선택할 수 있습니다.

**방법.** 다항 로짓 모형은 완전 요인 모형이나 사용자 정의 모형에 모두 적합합니다. 모수 추정에는 반복적 최대 우도 알고리즘을 사용하여 구합니다.

## 다항 로지스틱 회귀 데이터 고려사항

**데이터.** 종속변수는 범주형이어야 합니다. 독립변수는 요인 또는 공분산 변수여야 합니다. 일반적으로 요인은 범주형 변수, 공분산은 연속변수여야 합니다.

**가정.** 임의의 두 범주의 승산비는 다른 모든 반응 범주에 대해 독립적이라고 가정합니다. 예를 들어, 신제품을 출시하는 경우 다른 모든 제품의 시장 점유율에 같은 비율로 영향을 미치며, 공분산 패턴이 주어질 경우에 반응변수는 독립 다항 변수라고 가정합니다.

## 다항 로지스틱 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 회귀 > 다항 로지스틱 회귀...

2. 종속변수를 하나 선택합니다.

3. 요인변수는 선택 사항이며 숫자 또는 범주형 변수 모두 사용할 수 있습니다.

4. 공분산은 선택 사항이지만 지정하는 경우에는 숫자형 변수여야 합니다.

## 다항 로지스틱 회귀 모형

기본적으로 다항 로지스틱 프로시저는 요인 주효과와 공변량 주효과를 사용하여 모형을 작성하지만 이 대화 상자에서 사용자 정의 모형을 지정하거나 단계별 모형 선택을 요청할 수 있습니다.

### 모형설정

주효과 모형에 공분산 주효과와 요인 주효과는 포함되지만 상호작용 효과는 포함되지 않습니다. 완전 요인 모형에는 모든 주효과와 요인 대 요인 상호작용이 포함됩니다. 공분산 상호작용은 포함되지 않습니다. 사용자 정의 모형을 작성하여 요인 상호작용 서브세트 또는 공분산 상호작용 서브세트를 지정하거나 모형 항의 단계 선택을 요청할 수 있습니다.

### 요인 및 공변량

요인 및 공변량이 나열됩니다.

### 강제 입력항

강제 입력 목록에 추가된 항은 모형에 항상 포함됩니다.

### 단계선택항

단계선택 목록에 추가된 항은 다음과 같은 사용자가 선택한 단계선택법 중 하나에 따라 모형에 포함됩니다.

### 전진 입력

이 방법은 모형에서 단계선택항을 제거하는 작업부터 시작합니다. 각 단계에서 모형에 포함되지 않은 단계선택항이 통계적으로 중요한 기여도를 가지지 않을 때까지 가장 중요한 항을 모형에 추가합니다.

### 후진제거법

이 방법은 단계선택 목록에 지정된 모든 항을 모형에 입력하여 시작합니다. 각 단계에서 모형에 남아있는 단계선택항이 모형에 대해 통계적으로 중요한 기여도를 가질 때까지 가장 중요하지 않은 단계선택항을 모형에서 제거합니다.

### 단계별 전진

이 방법은 전진 입력법에서 선택한 모형으로 시작합니다. 여기서 알고리즘은 모형의 단계선택항에 대해 후진제거법을, 모형에 포함되지 않은 대해 전진입력법을 번갈아 진행합니다. 입력 또는 제거 기준에 충족되는 항이 없을 때까지 이 작업을 계속 반복합니다.

### 단계별 후진

이 방법은 후진제거법에서 선택할 모형부터 시작합니다. 여기서 알고리즘은 모형에 포함되지 않은 대해 전진입력법을, 모형의 단계선택항에 대해 후진제거법을 번갈아 진행합니다. 입력 또는 제거 기준에 충족되는 항이 없을 때까지 이 작업을 계속 반복합니다.

### 모형에 절편 포함

모형에 절편을 포함하거나 제외할 수 있습니다.

### 항 설정

선택된 요인과 공분산에 대해 다음과 같습니다.

**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.

### 다항 로지스틱 회귀분석 참조 범주

기본적으로 다항 로지스틱 회귀분석 프로시저는 마지막 범주를 참조 범주로 지정합니다. 변수 대화 상자에서 참조범주 및 범주 정렬 방법을 제어합니다.

### 참조범주

처음, 마지막 또는 사용자 정의 범주를 지정합니다.

## 범주 순서

오름차순으로 가장 작은 값부터 첫 범주를 정의하고 가장 큰 값은 마지막 범주를 정의합니다. 내림차순으로 가장 큰 값부터 처음 범주를 정의하고 가장 작은 값은 마지막 범주를 정의합니다.

## 다항 로지스틱 회귀분석 통계

다항 로지스틱 회귀분석에 대해 다음과 같은 통계를 지정할 수 있습니다.

### 케이스 처리 요약

이 표에는 지정한 범주형 변수에 대한 정보가 포함되어 있습니다.

**모델** 전체 모형에 대한 통계량입니다.

### 유사 R-제곱

Cox 및 Snell, Nagelkerke, McFadden  $R^2$  통계량을 인쇄합니다.

### 단계 요약

이 표에는 단계선택법의 각 단계에서 입력되거나 제거된 효과가 요약되어 있습니다. 모형 대화 상자에 단계 선택 모형을 지정하지 않으면 작성되지 않습니다.

### 모형 적합 정보

이 표에서는 적합 모형과 절편만 있는 모형 또는 널 모형을 비교합니다.

### 정보 기준

이 표에서는 Akaike 정보 기준(AIC)과 Schwartz 베이저안 정보 기준(BIC)을 인쇄합니다.

### 셀 확률

공분산 패턴과 반응 범주에 의한 관측 빈도, 기대 빈도(잔차 포함) 및 비율 표를 출력합니다.

### 분류표

관측반응 대 예측반응 값 표를 출력합니다.

### 적합도

Pearson 및 우도비 카이제곱 통계를 출력합니다. 모든 요인과 공분산 또는 사용자가 정의한 요인 및 공분산의 서브세트에서 지정한 공분산 패턴에 대한 통계를 계산합니다.

### 단조성 측도

일치되는 대응, 비일치되는 대응, 그리고 동률한 대응의 수에 대한 정보를 나타내는 표를 표시합니다. Somer의 D, Goodman과 Kruskal의 감마, Kendall의 타우-a, 그리고 일치 지수 C 또한 표에 표시되어 집니다.

**모수** 모형 모수와 관련된 통계량입니다.

### 추정값

사용자가 지정한 신뢰수준으로 모형 모수의 추정값을 출력합니다.

### 우도비 검정

모형의 부분 효과에 대한 우도비 검정을 출력합니다. 전체 모형 검정은 자동으로 출력됩니다.

### 근사 상관

모수 추정값 상관행렬을 출력합니다.

### 근사 공분산

모수 추정값 공분산 행렬을 출력합니다.

## 부-모집단 정의

요인 및 공분산의 서브세트를 선택하여 셀 확률 및 적합도 검정에서 사용한 공분산 패턴을 정의할 수 있습니다.

## 다항 로지스틱 회귀분석 기준

다항 로지스틱 회귀분석에 대해 다음과 같은 기준을 지정할 수 있습니다.

**반복** 알고리즘 순환의 최대 횟수, 단계 반분에서 최대 단계 수, 로그-우도 및 모수 값 변경에 대한 수렴허용 오차, 반복적 알고리즘 과정의 출력 빈도 및 반복에서 완전한 데이터 분리 또는 완전한 데 가까운 데이터 분리를 확인해야 하는 프로시저를 지정할 수 있습니다.

### 로그-우도 수렴

수렴은 로그 우도 함수의 절대 변화량이 지정된 값보다 작다고 가정합니다. 값이 0인 경우 이 기준은 적용되지 않습니다. 음수가 아닌 값을 지정합니다.

### 모수 수렴

수렴은 모수 추정값의 절대 변화량이 이 값보다 작다고 가정합니다. 값이 0인 경우 이 기준은 적용되지 않습니다.

**델타** 0 이상 1미만의 값을 지정할 수 있습니다. 이 값은 공분산 패턴에 따라 반응 범주 교차 분석표의 빈 셀에 추가됩니다. 안정적인 알고리즘을 만들고 추정값이 편향되지 않도록 합니다.

### 비정칙성 공차

비정칙성 검사에 사용되는 허용 오차 범위를 지정할 수 있습니다.

## 다항 로지스틱 회귀분석 옵션

다항 로지스틱 회귀분석에 대해 다음과 같은 옵션을 지정할 수 있습니다.

### 산포 척도

모수 공분산행렬의 추정값을 정정할 때 사용하는 산포 척도값을 지정할 수 있습니다. **편차**는 편차 함수(우도비 카이제곱) 통계를 사용하여 척도값을 추정합니다. **Pearson**은 Pearson 카이제곱 통계를 사용하여 척도값을 추정합니다. 또한 사용자는 임의의 척도값을 지정할 수 있습니다. 단, 값은 양수로 지정해야 합니다.

### 단계선택 옵션

단계적 방법을 사용하여 모형을 작성하는 경우 이러한 옵션을 사용하여 통계 기준을 제어합니다. 모형 대화 상자에서 단계선택 모형을 지정하지 않으면 이 옵션은 무시됩니다.

### 입력 확률

변수 입력에 필요한 우도비 통계 확률입니다. 지정한 확률이 클수록 변수를 모형에 입력하기 쉽습니다. 전진 입력, 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시됩니다.

### 항 입력 방법

단계선택법으로 항을 입력하는 방법입니다. 우도비 검정과 스코어 검정 중 하나를 선택합니다. 전진 입력, 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시됩니다.

### 제거 확률

변수 제거에 필요한 우도비 통계 확률입니다. 지정한 확률이 클수록 변수가 모형에 남아 있기 쉽습니다. 후진제거법, 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시됩니다.

### 항 제거 방법

단계선택법으로 항을 제거하는 방법입니다. 우도비 검정과 Wald 검정 중 하나를 선택합니다. 후진제거법, 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시됩니다.

### 모형 내 최소 다단효과(후진법)

후진제거법 또는 단계적 후진을 사용하는 경우 모형에 포함할 최소 항 수를 지정합니다. 절편은 모형 항으로 계산되지 않습니다.

### 모형 내 최대 다단효과(전진법)

전진 입력 또는 단계적 전진을 사용하는 경우 모형에 포함할 최대 항 수를 지정합니다. 절편은 모형 항으로 계산되지 않습니다.

### 계층 결정시 입력 및 제거 항에 대한 제약 적용

이 옵션을 사용하여 모형 항의 포함에 제한을 둘 것인지 여부를 선택할 수 있습니다. 포함할 항에 계층을 사용하려면 먼저 포함할 항의 일부인 모든 저차항이 모형에 있어야 합니다. 예를 들어, 계층 요구 사항이 적용 중이면 결혼상태와 성별 요인이 모두 모형에 있어야 결혼상태\*성별 상호작용을 추가할 수 있습니다. 세 개의 단일 선택 단추 중 선택에 따라 계층 확인 시 공분산의 역할이 결정됩니다.

## 다항 로지스틱 회귀분석 저장

저장 대화 상자를 사용하면 작업 파일에 변수를 저장하고 모형정보를 외부 파일로 내보낼 수 있습니다.

### 저장 변수

저장되는 변수:

#### 추정 반응 확률

요인/공변량 패턴을 반응 범주로 분류하는 추정 확률입니다. 반응변수 범주만큼의 확률 추정이 있으며 25개까지 저장됩니다.



## 예측 범주

요인/공분산 패턴에 대한 기대 확률이 가장 큰 반응 범주입니다.

## 예측 범주 확률

반응확률 추정의 최대값입니다.

## 실제 해당 범주에 속할 예측 확률

요인/공분산 패턴을 관측 범주로 분류하는 확률 추정입니다.

## XML 파일로 모형정보 내보내기

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

## NOMREG 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 종속변수의 참조 범주를 지정합니다.
- 사용자 결측값이 있는 케이스를 포함합니다.
- 귀무가설을 모수의 선형결합으로 지정하여 가설 검정을 사용자 정의합니다.

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

---

## 프로빗 회귀

이 프로시저는 자극의 강도와 자극에 대해 특정 반응을 나타내는 케이스 비율 사이의 관계를 측정합니다. 이 프로시저는 일부 종속변수의 수준에 의해 영향을 받거나 이로 인해 유발되는 이분형 결과가 있는 상황에서 유용하며, 특히 실험용 데이터에 적합합니다. 이 프로시저를 사용하면 특정 반응 비율을 감소시키는 데 필요한 자극의 강도(유효 처방 중위수)를 추정할 수 있습니다.

**예제.** 새 살충제가 개미를 죽이는 데 얼마나 효과적이며 적절한 사용 농도는 얼마입니까? 개미 표본에 다른 농도의 살충제를 사용하고 죽은 개미 수와 약에 노출된 개미 수를 기록하는 실험을 수행할 수 있습니다. 이 데이터에 프로빗 회귀를 적용하면 농도와 죽은 개미의 수 사이의 관계 강도를 판별하고, 살충제 투여 개미의 95%를 확실히 죽이려는 경우에 필요한 살충제 농도를 판별할 수 있습니다.

**통계.** 회귀계수와 표준 오차, 절편과 표준 오차, Pearson 적합도 카이제곱, 관측빈도와 기대빈도, 독립변수의 유효 수준에 대한 신뢰구간을 선택할 수 있습니다. 도표: 변환된 반응 도표.

이 프로시저에서는 Gill, Murray, Saunders 및 Wright에 의해 NPSONPSOL<sup>®</sup>에 제시되고 설치된 알고리즘을 사용하여 모형 모수를 추정합니다.

## 프로빗 회귀 데이터 고려사항

**데이터.** 각 독립변수 값(또는 여러 독립변수 값들의 각 조합)의 경우, 반응변수는 중요한 반응을 표출하는 값들을 포함하는 케이스 수의 개수이어야 하며 전체 관측개수 변수는 해당 독립변수에 대한 값들을 포함하는 전체 케이스 수의 개수이어야 합니다. 요인변수는 범주형이어야 하며 정수로 코딩되어야 합니다.

**가정.** 관측값은 독립이어야 합니다. 관측 연구에 나타난 것처럼 관측값 수에 비해 독립변수 값이 많으면, 카이제곱량과 적합도 통계량은 유효하지 않을 수도 있습니다.

**관련 프로시저.** 프로빗 분석은 로지스틱 회귀분석과 밀접한 관계가 있습니다. 실제로 로짓 변환을 선택하면 프로빗 프로시저에서는 본래 로지스틱 회귀분석을 계산합니다. 보통 관측 연구에는 로지스틱 회귀분석이 좀 더 적당한 반면, 계획된 실험에는 프로빗 분석이 더 적절합니다. 결과의 차이를 보면 이러한 두 분석값의 장점을 알 수 있습니다. 로지스틱 회귀분석 프로시저가 독립변수에 대한 승산비 추정값을 보고하는 반면, 프로빗 분석 프로시저는 다양한 반응률에 대한 유효 추정값을 보고합니다(유효 처방 중위수 포함).

## 프로빗 회귀 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 회귀 > 프로빗...

2. 반응 빈도 변수를 선택합니다. 이 변수는 검정 자극에 대한 반응을 표출하는 케이스의 수를 나타냅니다. 이 변수의 값은 음수가 될 수 없습니다.
3. 전체 관측빈도 변수를 선택합니다. 이 변수는 자극이 적용된 케이스의 수를 나타냅니다. 이 변수의 값은 음수가 될 수 없으며 각 케이스에 대한 반응 빈도 변수값 이상이어야 합니다.

선택적으로 요인변수를 선택할 수 있습니다. 선택한 경우 범위지정을 사용하여 그룹의 범위를 지정합니다.

4. 공분산을 한 개 이상 선택합니다. 이 변수에는 각 관측값에 적용되는 자극의 수준이 포함되어 있습니다. 공변량을 변환하려면 변환 드롭 다운 목록에서 변환을 선택합니다. 변환이 적용되지 않으며 대조군이 있는 경우 이 대조군이 분석에 포함됩니다.
5. 프로빗 또는 로짓 모형을 선택합니다.

프로빗 모형(*Probit Model*)

프로빗 변환(누적 표준 정규 분포함수의 역함수)을 반응률에 적용합니다.

로짓 모형(*Logit Model*)

반응 비율에 로짓(로그 승산) 변환을 적용합니다.

## 프로빗 회귀 범위지정

이 대화 상자를 사용하면 분석할 요인변수의 수준을 지정할 수 있습니다. 요인 수준은 연속적인 정수로 코딩해야 하며 지정한 범위 내의 모든 수준이 분석됩니다.

## 프로빗 회귀 옵션

### PROBIT 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 프로빗과 로짓 모형에 대한 분석을 요청합니다.
- 결측값 처리를 제어합니다.
- 공분산을 상용로그 또는 자연 로그 이외의 방법으로 변환합니다.

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

---

### 비선형 회귀분석

비선형 회귀분석은 종속변수와 독립변수 세트 간의 비선형 관계 모형을 탐색하는 방법입니다. 기존의 선형 회귀분석(선형 모형 추정에만 제한됨)과 달리 비선형 회귀분석에서는 종속변수와 독립변수 간의 임의적 관계를 통해 모형을 추정할 수 있습니다. 이 방법에서는 반복 추정 알고리즘을 사용합니다. 이 프로시저는  $Y = A + BX^{**2}$  형식의 단순 다항에서는 필요하지 않습니다.  $W = X^{**2}$ 를 정의하여 단순 선형 모형인  $Y = A + BW$ 를 구할 수 있으며 이 모형은 기존의 방법(예: 선형 회귀분석 프로시저)을 사용하여 추정할 수 있다는 것은 유의하십시오.

**예제.** 시간을 기준으로 인구를 추정할 수 있습니까? 산점도를 통해서 보면 인구와 시간 간에는 강한 관계가 있는 것처럼 보이지만 이러한 관계는 비선형이기 때문에 비선형 회귀분석 프로시저의 특별한 추정 방법이 요구됩니다. 로지스틱 인구 증가 모형과 같은 적절한 방정식을 세워 이 모형에 적합한 추정값을 구함으로써 실제 측정되지 않은 시간에 따른 인구를 추정할 수 있습니다.

**통계.** 각 반복에 대해 모수 추정값과 잔차 제곱합 통계를 구합니다. 각 모형에 대해 회귀 제곱합, 잔차, 비수정 총계와 수정 합계, 모수 추정값, 근사 표준 오차, 모수 추정값의 근사 상관행렬 통계를 구합니다.

**참고:** 제한 비선형 회귀는 Gill, Murray, Saunders, 그리고 Wright에 의해 NPSOL<sup>®</sup>에 제시되고 구현된 알고리즘을 사용하여 모형 모수를 추정합니다.

### 비선형 회귀 데이터 고려사항

**데이터.** 종속변수와 독립변수는 양적변수이어야 합니다. 범주형 변수(예: 종교, 전공, 거주 지역)는 이분형(더미) 변수나 다른 유형의 대비변수로 코딩 변경되어야 합니다.

**가정.** 종속변수와 독립변수 간의 관계를 정확하게 설명하는 함수를 지정한 경우에만 결과가 유효합니다. 또한 적합한 시작값을 선택하는 것도 매우 중요합니다. 모형의 올바른 함수 형식을 선택한 경우라도 시작값을 잘못 선택하면 모형이 수렴되지 않거나 전체 최적 해법을 얻지 못하고 부분 최적 해법을 얻게 됩니다.

**관련 프로시저.** 비선형 모형에 맨 처음 나타나는 모든 모형을 선형 모형으로 전환할 수 있으며 선형 회귀분석 프로시저를 사용하여 분석할 수 있습니다. 적절한 모형을 잘 모르는 경우에는 곡선 추정 프

로시저를 사용하면 데이터에 유용한 함수 관계식을 확인할 수 있습니다.

## 비선형 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 회귀 > 비선형...

2. 활성 데이터 세트의 변수 목록에서 숫자 종속변수를 하나 선택합니다.
3. 모형 표현식을 작성하려면 모형 표현식 필드에 표현식을 입력하거나 구성요소(변수, 모수, 함수)를 붙여넣습니다.
4. 모수를 클릭하여 모형의 모수를 확인합니다.

세그먼트 모형(다른 도메인 부분에서 서로 다른 형식을 취하는 모형)은 단일 모형문 내에서 조건 논리를 사용하여 지정해야 합니다.

## 조건 논리(비선형 회귀분석)

조건 논리를 사용하여 세그먼트 모형을 지정할 수 있습니다. 모형 표현식 또는 손실 함수 내에 조건 논리를 사용하려면 각 조건마다 하나씩 계열 항의 합계를 작성합니다. 각 항은 논리 표현식이 참인 경우에 결과가 생성되는 표현식을 곱한 논리 표현식으로 구성됩니다.

예를 들어,  $X \leq 0$ 인 경우 0,  $0 < X < 1$ 인 경우  $X$ ,  $X \geq 1$ 인 경우 1인 세그먼트 모형의 표현식은 다음과 같습니다.

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$$

괄호 안의 논리 표현식은 모두 1(true) 또는 0(false)으로 추정됩니다. 그러므로

$X \leq 0$ 인 경우 위 식은  $1 * 0 + 0 * X + 0 * 1 = 0$ 으로 변환됩니다.

$0 < X < 1$ 인 경우 위 식은  $0 * 0 + 1 * X + 0 * 1 = X$ 로 변환됩니다.

$X \geq 1$ 인 경우 위 식은  $0 * 0 + 0 * X + 1 * 1 = 1$ 로 변환됩니다.

다른 논리 표현식과 결과 표현식을 대체하여 훨씬 복잡한 예제를 쉽게 작성할 수 있습니다. 부등식이 두 개인 경우(예:  $0 < X < 1$ )는 복합 표현식(예:  $X > 0 \ \& \ X < 1$ )으로 작성해야 합니다.

문자열 변수는 논리 표현식 내에서 사용할 수 있습니다.

$$(city = 'New York') * costliv + (city = 'Des Moines') * 0.59 * costliv$$

이렇게 하면 뉴욕 거주자에 대한 하나의 표현식(*costliv* 변수값)과 Des Moines 거주자에 대한 다른 표현식(뉴욕에 대한 값의 59%)이 산출됩니다. 문자 상수는 다음과 같이 따옴표나 어포스트로피로 묶어야 합니다.

## 비선형 회귀분석: 모수

모수는 비선형 회귀분석 프로시저에서 추정되는 모형의 일부이며 가법 상수, 승법 계수, 지수 또는 함수 평가에 사용되는 값을 말합니다. 정의한 모든 모수가 초기값과 함께 변수 대화 상자의 모수 목록에 표시됩니다.

**이름** 각 모수에 대한 이름을 지정해야 합니다. 이 이름은 유효한 변수 이름이어야 하며 변수 대화 상자의 모형 표현식에서 사용되는 이름이어야 합니다.

### 시작값

모수에 대한 시작값을 지정할 수 있으며 시작값은 가능한 예측 최종 해법에 가깝게 지정하는 것이 좋습니다. 시작값을 잘못 지정하면 식이 수렴되지 않거나 부분 해법 또는 실제로 적용 불가능한 해법을 수렴하게 됩니다.

### 이전 분석결과를 시작값으로 사용

이 대화 상자에서 비선형 회귀를 이미 실행한 경우에는 이 옵션을 선택하여 이전에 실행한 분석결과를 모수의 초기값으로 사용할 수 있습니다. 이 방법을 사용하면 알고리즘이 서서히 수렴되는 동안 검색을 계속할 수 있습니다. (초기 시작값이 변수 대화 상자의 모수 목록에 계속 표시됩니다.)

**참고:** 이 옵션은 나머지 세션 동안 대화 상자에 그대로 유지됩니다. 모형을 변경하는 경우에는 이 옵션을 선택 해제해야 합니다.

## 비선형 회귀분석 일반 모형

아래의 표는 여러 일반적인 비선형 회귀 모형에 대한 예제 모형 구문입니다. 무작위로 선택한 모형은 데이터에 적합하지 않을 수 있습니다. 모수에는 적합한 시작값이 있어야 하고 일부 모형에는 수렴에 필요한 제한조건이 있어야 합니다.

표 1. 모형 구문 예제

이름	모형 표현식
근사 회귀분석	$b1 + b2 * \exp(b3 * x)$
근사 회귀분석	$b1 - (b2 * (b3 ** x))$
밀도	$(b1 + b2 * x) ** (-1 / b3)$
Gauss	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$
로그 수정	$(b1 + b3 * x) ** b2$
로그 로지스틱	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Metcherlich 수확 체감의 법칙	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Peal-Reed	$b1 / (1 + b2 * \exp(-(b3 * x + b4 * x ** 2 + b5 * x ** 3)))$
3차모형	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
2차 모형	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$

표 1. 모형 구문 예제 (계속)

이름	모형 표현식
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
와이블	$b1 - b2 * \exp(-b3 * x ** b4)$
밀도 산출	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

## 비선형 회귀분석: 손실 함수

비선형 회귀분석의 손실 함수는 알고리즘에 의해 최소화된 함수입니다. 잔차 제곱합을 최소화하려면 잔차 제곱합을 선택하고 다른 함수를 최소화하려면 사용자 정의 손실 함수를 선택합니다.

사용자 정의 손실 함수를 선택하는 경우 모수값을 선택하여 합계가 최소화되도록 손실 함수를 정의해야 합니다.

- 대부분의 손실 함수에는 잔차를 나타내는 특수변수 *RESID\_*가 포함되어 있습니다. 기본값 잔차 제곱합 손실 함수는 *RESID\_\*\*2*로 정확하게 입력되어 있어야 합니다. 손실 함수에 예측값을 사용해야 할 경우 종속변수에서 잔차를 뺀 값과 같습니다.
- 조건 논리를 사용하여 조건 손실 함수를 지정할 수 있습니다.

손실 표현식 필드에 표현식을 입력하거나 표현식의 구성요소를 필드에 붙여넣을 수 있습니다. 문자 상수는 따옴표나 어포스트로피로 묶어야 하며 숫자 상수는 점을 소수점 구분자로 사용하는 미국식으로 입력해야 합니다.

## 비선형 회귀분석: 모수 제한조건

제한조건은 해법을 반복 검색하는 동안 모수에 허용되는 값에 대한 제한 사항입니다. 선형 표현식에서는 하나의 단계가 수행되기 전에 계산되기 때문에 오버플로우되는 것을 방지하도록 선형 제한조건을 사용할 수 있습니다. 그러나 비선형 표현식은 하나의 단계가 수행된 후에 계산이 수행됩니다.

각 방정식이나 부등식을 사용하려면 다음 요소가 있어야 합니다.

- 모형에서 하나 이상의 모수를 포함하는 표현식. 표현식을 입력하거나 숫자, 연산자 또는 괄호를 표현식에 붙일 수 있는 키패드를 사용합니다. 표현식의 나머지 부분과 더불어 필요한 모수를 입력하거나 왼쪽의 모수 목록에서 붙여넣을 수 있습니다. 일반 변수를 상수로 사용할 수 없습니다.
- 세 가지 논리 연산자  $<=$ ,  $=$ ,  $>=$  중 하나
- 논리 연산자를 사용하여 표현식과 비교한 숫자 상수. 상수를 입력합니다. 숫자 상수는 미국식으로 입력하되 구분자로 점(.)을 사용합니다.

## 비선형 회귀분석: 새 변수로 저장

새 변수를 활성 데이터 파일에 저장할 수 있습니다. 사용할 수 있는 옵션에는 잔차, 예측값, 도함수, 손실 함수값이 있습니다. 이러한 변수는 연속적인 분석에 사용하여 모형 적합도를 검정하거나 오류 케이스를 판별할 수 있습니다.

### 예측값(Predicted Values)

변수 이름 pred\_로 예측값을 저장합니다.

### 잔차(Residuals)

변수 이름 resid로 잔차를 저장합니다.

### 도함수(Derivatives)

각 모형 모수에 대해 하나의 도함수가 저장됩니다. 도함수 이름으로는 모수 이름의 처음 여섯 문자 앞에 'd.'가 붙어 사용됩니다.

### 손실 함수 값(Loss Function Values)

이 옵션은 고유한 손실 함수를 지정한 경우 사용할 수 있습니다. 손실 함수 값에는 변수 이름 loss\_가 할당됩니다.

## 비선형 회귀분석: 옵션

옵션을 사용하여 다음과 같은 비선형 회귀분석의 다양한 측면을 제어할 수 있습니다.

### 표준오차의 부트스트랩 추정값

원래 데이터 세트에서 반복되는 표본을 사용하여 통계의 표준 오차를 추정하는 방법입니다. 이 방법은 원래 데이터 세트와 크기가 같은 표본을 많이 추출하기 위해 표본을 복원 추출함으로써 수행됩니다. 이들 표본 각각에 대해 비선형 방정식이 추정됩니다. 각 모수 추정값의 표준 오차는 부트스트랩 추정값의 표준 편차로 계산됩니다. 원래 데이터 변수 세트의 모수 값을 각 부트스트랩 표본에서 시작값으로 사용할 수 있습니다. 이 방법을 사용하려면 축차 2차계획법 알고리즘이 있어야 합니다.

### 추정 방법

가능한 경우 추정 방법을 선택할 수 있습니다. 이 대화 상자나 다른 대화 상자의 일부 선택 사항에는 순차 2차계획법 알고리즘이 있어야 합니다. 순차 2차계획법과 Levenberg-Marquardt를 대신 사용할 수 있습니다.

### 순차 2차계획법(Sequential Quadratic Programming)

이 방법은 제한조건 있는 모형과 제한조건 없는 모형에 대해 사용할 수 있습니다. 제한조건이 있는 모형 또는 사용자 정의 손실 함수 또는 부트스트랩을 지정하면 축차 2차계획법이 자동으로 사용됩니다. 최대 반복수 및 단계 한계에 새로운 값을 입력하고 최적 허용 오차, 함수의 정도 및 무한 단계 크기에 대한 드롭다운 목록에서 선택한 사항을 변경할 수 있습니다.

### Levenberg-Marquardt

제한조건이 없는 모형에 대한 기본값 알고리즘입니다. 제한조건이 있는 모형, 사용자 정의 손실 함수 또는 부트스트랩을 지정하는 경우에는 Levenberg-Marquardt 방법을 사용할 수 없습니다. 최대 반복수에 새로운 값을 입력하고 드롭다운 목록에서 제공합 수렴 및 모수 수렴에 대해 선택한 사항을 변경할 수 있습니다.

## 비선형 회귀분석 결과 해석

비선형 회귀분석에서는 다음과 같은 계산상의 오류가 흔히 발생합니다.

- 모수의 초기값 선택 사항은 수렴에 영향을 미칩니다. 적절한 초기값을 선택합니다. 가능하면 예측 최종 해법에 가깝게 선택하는 것이 좋습니다.
- 경우에 따라 특정 문제 상황에서 보다 잘 수행되는 알고리즘이 있습니다. 옵션 대화 상자에서 사용 가능한 다른 알고리즘을 선택합니다. 손실 함수나 특정 유형의 제한조건을 지정하면 Levenberg-Marquardt 알고리즘을 사용할 수 없습니다.
- 최대 반복 수가 발생하여 반복이 중지된 경우에만 "최종" 모형이 적합한 해법이 아닙니다. 모수 대화 상자에서 이전 분석결과를 시작값으로 사용을 선택하여 반복을 계속하거나 다른 초기값을 선택합니다.
- 누승이나 큰 데이터 값이 필요한 모형에서는 오버플로우나 언더플로우(컴퓨터에 표시하기에 수가 너무 크거나 작음)를 일으킬 수 있습니다. 이러한 문제는 모수에 제한조건을 지정하거나 적합한 초기값을 선택하여 오버플로우나 언더플로우를 방지할 수 있습니다.

## NLR 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 모수 추정값에 대한 초기값을 읽을 파일에 이름을 지정합니다.
- 두 개 이상의 모형문과 손실 함수를 지정합니다. 이러한 기능을 사용하면 세그먼트 모형을 쉽게 지정할 수 있습니다.
- 프로그램에 의해 계산되는 도함수를 사용하지 않고 사용자의 도함수를 제공합니다.
- 생성할 부트스트랩 예제 수를 지정합니다.
- 잔차와 도함수 간의 상관에 대한 수렴을 확인하고 정의하는 도함수에 대한 기준 값 설정을 포함하는 추가 반복수 기준을 지정합니다.

CNLR(제한 비선형 회귀분석) 명령에 대한 추가 기준을 사용하여 다음을 수행할 수 있습니다.

- 각각의 주 반복 내에 허용된 최대 보조 반복수를 지정합니다.
- 도함수 확인에 대한 기준 값을 설정합니다.
- 단계 한계를 설정합니다.
- 초기값이 지정된 경계 내에 있는지 여부를 결정하려면 충돌 허용 오차를 지정합니다.

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

---

## 가중 추정

표준 선형회귀 모형은 연구 중인 모집단에서 분산이 일정하다고 가정합니다. 분산이 일정하지 않으면 (예: 일부 속성이 높게 나타난 케이스가 낮게 나타난 케이스보다 변동이 많은 경우) 보통최소제곱(OLS)을 사용하는 선형회귀에서는 최적의 모형 추정값을 더 이상 제공하지 않습니다. 변동에서의 차이를 다른 변수로부터 예측할 수 있으면 가중 추정 프로시저에서는 가중최소제곱(WLS)을 사용하여 선형회귀 모형의 계수를 계산할 수 있으므로, 회귀계수를 결정할 때 관측값이 정확할수록(즉, 변동이 적을수록) 더 큰 가중치가 주어집니다. 가중 추정 프로시저에서는 가중 변환의 범위를 검정하여 데이터에 가장 적합한 범위를 제공합니다.



**예제.** 인플레이션과 실업 상태가 주식 가격에 미치는 영향은 어떠한 것입니까? 배당액이 높은 주식이 낮은 배당값을 가지는 주식보다 더 큰 변동을 보이는 경우가 많으므로 일반 최소제곱 방법으로는 최적의 추정값을 구할 수 없습니다. 가중 추정 프로시저를 통해 선형 모형 계산시 가격 변동에 대한 주식 배당 가격의 영향력을 설명할 수 있습니다.

**통계.** 검정된 가중치 소스 변수의 각 검정력에 대한 로그-우도값, 다중 R, R-제곱, 수정된 R-제곱, WLS 모형에 대한 분산 분석표, 표준화되었거나 표준화되지 않은 모수 추정값, WLS 모형에 대한 로그-우도를 선택할 수 있습니다.

## 가중추정 데이터 고려사항

**데이터.** 종속변수와 독립변수는 양적변수이어야 합니다. 범주형 변수(예: 종교, 전공, 거주 지역)는 이분형(더미) 변수나 다른 유형의 대비변수로 코딩 변경되어야 합니다. 가중 변수는 양적변수이어야 하며 종속변수의 변동과 관련이 있어야 합니다.

**가정.** 각 독립변수 값에서 종속변수의 분포는 정규 분포이어야 합니다. 종속변수와 각 독립변수 간 관계는 선형적이어야 하며 모든 관측값은 독립적이어야 합니다. 종속변수의 분산은 독립변수의 수준에 따라 달라질 수 있으나 그 차이는 가중 변수를 기준으로 예측할 수 있어야 합니다.

**관련 프로시저.** 데이터 탐색 프로시저는 데이터를 화면 표시하는 데 사용될 수 있습니다. 데이터 탐색 프로시저는 그래픽 표시 뿐만 아니라 분산의 정규성과 동질성에 대해 검정합니다. 종속변수가 독립변수 수준에서 동일한 분산을 가지는 것으로 추정될 경우 선형회귀 프로시저를 사용할 수 있습니다. 데이터가 가정(예: 정규성)을 위반했다면 데이터를 변환해 봅시다. 데이터가 선형으로 관련되지 않고 변환을 수행해도 도움이 되지 않으면 곡선 추정 프로시저에서 대체 모형을 사용합니다. 종속변수가 이분형인 경우(예: 특정 상품이 다 팔렸거나 그렇지 않으면 품목이 부족한 경우), 로지스틱 회귀분석 프로시저를 사용합니다. 종속변수가 중도절단되는 경우(예: 수술 후 회복 시간) 사용자 정의 표 및 고급 통계에서 사용할 수 있는 생명표, Kaplan-Meier 모형, Cox 회귀모형을 사용합니다. 데이터가 독립변수가 아니면(예: 여러 다른 조건에서 동일한 사람을 관측할 경우) 사용자 정의 표 및 고급 통계의 반복측도 프로시저를 사용합니다.

## 가중 추정 분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 회귀 > 가중추정...

2. 종속변수를 하나 선택합니다.

3. 하나 이상의 독립변수를 선택합니다.

4. heteroscedasticity 소스에 있는 변수를 가중 변수로 선택합니다.

가중 변수(*Weight Variable*)

거듭제곱된 이 변수의 역수에 의해 데이터가 가중됩니다. 회귀 방정식은 검정력 값에서 지정한 범위 각각에 대해 계산되고 로그-우도 함수를 최대화하는 검정력을 나타냅니다.

### 검정력 범위(Power Range)

가중 변수와 함께 사용되어 가중치를 계산합니다. 여러 회귀 방정식이 사용되며 검정력 범위의 각 값에 대해 하나씩 사용됩니다. 검정력 범위 입력란에 입력한 값은 -6.5부터 7.5까지 이어야 합니다. 검정력 값은 지정한 값에 따라 결정되는 증분만큼 하한값부터 상한값까지 나타납니다. 검정력 범위의 전체 값 수는 150개까지로 제한됩니다.

## 가중 추정: 옵션

가중 추정 분석에 대해 다음 옵션을 지정할 수 있습니다.

### 최적의 가중값을 새 변수로 저장

현재 파일에 가중 변수를 추가합니다. 이 변수는  $WGT\_n$ 이며, 여기에서  $n$ 은 변수에 고유 이름을 부여하기 위해 선택한 숫자입니다.

### 분산분석표 및 추정값 출력 기준

결과에 통계가 표시되는 방식을 제어할 수 있게 합니다. 사용할 수 있는 옵션은 최적 거듭제곱 결과만 및 각 거듭제곱 별로입니다.

## WLS 명령 추가 기능

명령 구문을 사용하여 수행할 수 있는 추가 기능은 다음과 같습니다.

- 검정력에 단일 값을 지정합니다.
- 검정력 값의 목록을 지정하거나 값의 범위와 검정력 값의 목록을 합합니다.

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

---

## 2단계 최소제곱 회귀분석

표준 선형 회귀 모형은 종속변수의 오차가 독립변수와 상관되지 않는 것으로 가정합니다. 이런 경우가 아니면 (예를 들어, 변수 간 관계가 양방향인 경우) 보통최소제곱(OLS)을 사용하는 선형 회귀분석은 더 이상 최적의 모형 추정값을 제공하지 않습니다. 2단계 최소제곱 회귀분석은 오차항과 상관되지 않은 도구변수를 사용하여 추정된 불확실한 예측변수의 값을 계산한 다음(첫 번째 단계) 계산한 값으로 종속변수의 선형 회귀 모형을 추정합니다(두 번째 단계). 계산된 값은 해당 오류와 상관되지 않는 변수를 기준으로 달라지므로 2단계 모형의 결과는 최적 상태를 나타냅니다.

**예제.** 상품 수요가 가격과 소비자의 수입에 관련이 있습니까? 이 모형에서 고려할 점은 가격과 수요가 서로에 대해 역효과를 가진다는 것입니다. 즉, 가격도 수요에 영향을 미칠 수 있으며 수요도 가격에 영향을 미칠 수 있습니다. 2단계 최소제곱 회귀 모형은 소비자의 수입과 떨어진 가격을 사용하여 수요에 대한 측정 오차에 상관하지 않는 가격의 대리값을 계산합니다. 초기에 지정된 모형에서 원래 가격 대신 이 대리값을 사용하여 결과를 계산합니다.

**통계.** 각 모형에 대해 표준화 및 비표준화 회귀계수, 다중  $R$ ,  $R^2$ , 수정된  $R^2$ , 추정값의 표준 오차, 분산 분석표, 예측값, 잔차가 있으며 또한 각 회귀계수에 대한 95% 신뢰구간과 모수 추정값의 상관행렬과 공분산 행렬 통계를 구합니다.

## 2-단계 최소제곱 회귀 데이터 고려사항

**데이터.** 종속변수와 독립변수는 양적변수이어야 합니다. 범주형 변수(예: 종교, 전공, 거주 지역)는 이분형(더미) 변수나 다른 유형의 대비변수로 코딩 변경되어야 합니다. 내생 설명변수는 범주형이 아닌 양적변수이어야 합니다.

**가정.** 각 독립변수 값에서 종속변수의 분포는 정규 분포이어야 합니다. 종속변수 분포의 분산은 모든 독립변수 값에 대해 항상 일정해야 합니다. 종속변수와 각 독립변수 간 관계는 선형이어야 합니다.

**관련 프로시저.** 종속변수의 오차와 상관되는 예측변수가 없다고 생각되면 선형 회귀분석 프로시저를 사용할 수 있습니다. 데이터로 인해 지정한 가정 중 하나(예: 정규성이나 일정한 분산)를 위반하게 되는 경우 데이터를 변환합니다. 데이터가 선형으로 관련되지 않고 변환을 수행해도 도움이 되지 않으면 곡선 추정 프로시저에서 대체 모형을 사용합니다. 종속변수가 이분형이면 (예: 특정 상품의 완전 판매여부) 로지스틱 회귀분석 프로시저를 사용합니다. 데이터가 독립변수가 아니면(예: 여러 다른 조건에서 동일한 사람을 관측할 경우) 반복측도 프로시저를 사용합니다.

## 2단계 최소제곱 회귀분석 구하기

1. 메뉴에서 다음을 선택합니다.

분석 > 회귀 > 2-단계 최소제곱...

2. 종속변수를 하나 선택합니다.

3. 하나 이상의 설명변수(예측변수)를 선택합니다.

4. 하나 이상의 도구변수를 선택합니다.

- 도구변수(*Instrumental*). 2단계 최소제곱 분석의 첫 번째 단계에서 예측된 내생변수값을 계산하는 데 사용되는 변수입니다. 동일한 변수가 설명 변수 및 도구변수 목록 상자 모두에 나타날 수 있습니다. 도구변수 수는 설명 변수 수보다 크거나 같아야 합니다. 나열된 설명 변수와 도구변수가 모두 동일하면 결과는 선형 회귀분석 프로시저의 결과와 같아집니다.

도구변수로 지정되지 않은 설명변수는 내생으로 간주됩니다. 일반적으로 설명 목록의 모든 외생 변수도 도구변수로 지정됩니다.

## 2단계 최소제곱 회귀분석 옵션

분석시 다음 옵션을 선택할 수 있습니다.

### 저장할 새 변수 지정

활성 파일에 새 변수를 추가할 수 있습니다. 사용할 수 있는 옵션은 예측값과 잔차입니다.

### 모수의 공분산 표시

모수 추정값의 공분산 행렬을 인쇄할 수 있습니다.

## 2SLS 명령 추가 기능

명령 구문을 사용하면 여러 방정식을 동시에 추정할 수도 있습니다. 명령에 대한 자세한 내용은 *Command Syntax Reference*를 참조하십시오.

## 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 부명령문에서 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 부명령문은 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 부명령문은 두 번째 범주를 생략하는 대비 행렬을 구합니다.

/CONTRAST(FACTOR) = SIMPLE(2)

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

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$

## Helmert

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

$$\begin{matrix} \text{mean} & (1/k & 1/k & \dots & 1/k & 1/k & 1/k) \\ \text{df}(1) & ( 1 & -1/(k-1) & \dots & -1/(k-1) & -1/(k-1) & -1/(k-1)) \\ \text{df}(2) & ( 0 & 1 & \dots & -1/(k-2) & -1/(k-2) & -1/(k-2)) \\ & \vdots & & & & & \\ & \vdots & & & & & \\ \text{df}(k-2) & ( 0 & 0 & \dots & 1 & -1/2 & -1/2) \\ \text{df}(k-1) & ( 0 & 0 & \dots & 0 & 1 & -1) \end{matrix}$$

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

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & -1/3 & -1/3 & -1/3 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

## 차이

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

$$\begin{matrix} \text{mean} & ( 1/k & 1/k & 1/k & \dots & 1/k) \\ \text{df}(1) & ( -1 & 1 & 0 & \dots & 0) \\ \text{df}(2) & ( -1/2 & -1/2 & 1 & \dots & 0) \\ & \vdots & & & & \\ & \vdots & & & & \\ \text{df}(k-1) & (-1/(k-1) & -1/(k-1) & -1/(k-1) & \dots & 1) \end{matrix}$$

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

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ -1 & 1 & 0 & 0 \\ -1/2 & -1/2 & 1 & 0 \\ -1/3 & -1/3 & -1/3 & 1 \end{pmatrix}$$

## 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 생략하면 1에서  $k$ 까지 연속적인 정수로 지정할 수 있습니다. 여기서  $k$ 는 범주 수입니다. 변수 *drug*에 세 개의 범주가 있으면 부명령문은 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 제공 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 부명령문을 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 아니어야 합니다. 그런 경우 프로시저는 선형 종속성을 보고하고 처리를 중지합니다. Helmert, 차이 및 다항 대비는 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.





---

## 주의사항

이 정보는 미국에서 제공되는 제품 및 서비스용으로 작성된 것입니다. 본 자료는 다른 언어로도 제공될 수 있습니다. 그러나 자료에 접근하기 위해서는 해당 언어로 된 제품 또는 제품 버전의 사본이 필요할 수 있습니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. 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.

07326

서울특별시 영등포구

국제금융로 10, 31FC

한국 아이.비.엠 주식회사

대표전화서비스: 02-3781-7114

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

*Intellectual Property Licensing*

*Legal and Intellectual Property Law*

*IBM Japan Ltd.*

*19-21, Nihonbashi-Hakozakicho, Chuo-ku*

*Tokyo 103-8510, Japan*

IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 묵시적이든 명시적이든 어떠한 종류의 보증 없이 이 책을 "현상태대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및/또는 프로그램을 사전 통지 없이 언제든지 개선 및/또는 변경할 수 있습니다.

이 정보에서 언급되는 비IBM 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

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

07326

서울특별시 영등포구

국제금융로 10, 31FC

한국 아이.비.엠 주식회사

대표전화서비스: 02-3781-7114

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

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

인용된 성능 데이터와 고객 예제는 예시 용도로만 제공됩니다. 실제 성능 결과는 특정 구성과 운영 조건에 따라 다를 수 있습니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 제품들을 테스트하지 않았으므로, 비IBM 제품과 관련된 성능의 정확성, 호환성 또는 기타 청구에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다.

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

저작권 라이선스:

이 정보에는 여러 운영 플랫폼에서의 프로그래밍 기법을 보여주는 원어로 된 샘플 응용프로그램이 들어 있습니다. 귀하는 이러한 샘플 프로그램의 작성 기준이 된 운영 플랫폼의 애플리케이션 프로그래밍 인터페이스(API)에 부합하는 애플리케이션을 개발, 사용, 판매 또는 배포할 목적으로 IBM에 추가 비용을 지불하지 않고 이들 샘플 프로그램을 어떠한 형태로든 복사, 수정 및 배포할 수 있습니다. 이러한

샘플 프로그램은 모든 조건하에서 완전히 테스트된 것은 아닙니다. 따라서 IBM은 이러한 프로그램의 신뢰성, 서비스 가능성 또는 기능을 보증하거나 진술하지 않습니다. 본 샘플 프로그램은 일체의 보증 없이 "현상태대로" 제공됩니다. IBM은 귀하의 샘플 프로그램 사용과 관련되는 손해에 대해 책임을 지지 않습니다.

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

© IBM 2019. 이 코드의 일부는 IBM Corp.의 샘플 프로그램에서 파생됩니다.

© Copyright IBM Corp. 1989 - 2019. All rights reserved.

---

## 상표

IBM, IBM 로고 및 [ibm.com](http://ibm.com)은 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 "저작권 및 상표 정보" 웹 페이지([www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml))에 있습니다.

Adobe, Adobe 로고, PostScript 및 PostScript 로고는 미국 및/또는 기타 국가에서 사용되는 Adobe Systems Incorporated의 등록상표 또는 상표입니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Intel Centrino, Intel Centrino 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation 또는 그 계열사의 상표 또는 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.



## 색인

### [가]

- 가중 추정 20
  - 로그-우도 20
  - 명령 추가 기능 22
  - 반복 히스토리 22
  - 분산 분석표 및 추정값 출력 22
  - 예제 20
  - 최적의 가중치를 새 변수로 저장 22
  - 통계 20
- 공분산
  - 로지스틱 회귀분석 4
- 공분산 행렬
  - 다항 로지스틱 회귀분석 10
- 관측값이 0인 셀
  - 다항 로지스틱 회귀분석 11
- 근사 회귀분석
  - 비선형 회귀분석 17

### [다]

- 다변량 선형 회귀 18, 19
- 다항 로지스틱 회귀분석 7, 10
  - 기준 11
  - 명령 추가 기능 13
  - 모형 8
  - 모형 내보내기 정보 12
  - 저장 12
  - 참조 범주 9
  - 통계 10
- 단계 반분
  - 다항 로지스틱 회귀분석 11
- 단계 선택
  - 다항 로지스틱 회귀분석 8
  - 로지스틱 회귀분석 3
- 대비
  - 로지스틱 회귀분석 4
- 델타
  - 관측값이 0인 셀 보정 11

### [라]

- 레버리지 값
  - 로지스틱 회귀분석 5

- 로그 수정 모형
  - 비선형 회귀분석 17
- 로그-우도
  - 가중 추정 20
  - 다항 로지스틱 회귀분석 10
- 로지스틱 회귀 분석 2
- 로지스틱 회귀분석 2
  - 계수 2
  - 규칙 설정 3
  - 단계선택 확률기준 6
  - 대비 4
  - 명령 추가 기능 7
  - 문자 공분산 4
  - 반복 6
  - 범주형 공분산 4
  - 변수 선택 방법 3
  - 분류 질사 6
  - 상수항 6
  - 새 변수 저장 5
  - 선택 규칙 정의 3
  - 영향력 측도 5
  - 예제 2
  - 예측값 5
  - 이분형 1
  - 잔차 5
  - 출력 옵션 6
  - 통계 2
  - 통계 및 도표 6
  - Hosmer-Lemeshow 적합도 통계량 6

### [마]

- 모수 제한조건
  - 비선형 회귀분석 18
- 모수 추정값
  - 다항 로지스틱 회귀분석 10
- 문자 공분산
  - 로지스틱 회귀분석 4
- 믿을만한 신뢰구간
  - 프로빗 회귀 15
- 밀도 모형
  - 비선형 회귀분석 17
- 밀도 산출 모형
  - 비선형 회귀분석 17

### [바]

- 반복
  - 다항 로지스틱 회귀분석 11
  - 로지스틱 회귀분석 6
  - 프로빗 회귀 15
- 반복 히스토리
  - 다항 로지스틱 회귀분석 11
- 범주형 공분산 4
- 분류
  - 다항 로지스틱 회귀분석 7
- 분류표
  - 다항 로지스틱 회귀분석 10
- 분리
  - 다항 로지스틱 회귀분석 11
- 비선형 모형
  - 비선형 회귀분석 17
- 비선형 회귀 18, 19
- 비선형 회귀분석 15
  - 결과 해석 19
  - 도함수 18
  - 명령 추가 기능 20
  - 모수 17
  - 모수 제한조건 18
  - 부트스트랩 추정값 19
  - 새 변수로 저장 18
  - 세그먼트 모형 16
  - 손실 함수 18
  - 순차 2차계획법 19
  - 시작값 17
  - 예제 15
  - 예측값 18
  - 일반 회귀 모형 17
  - 잔차 18
  - 조건 논리 16
  - 추정 방법 19
  - 통계 15
  - Levenberg-Marquardt 알고리즘 19
- 비정칙성
  - 다항 로지스틱 회귀분석 11

## [사]

- 사용자 정의 모형
  - 다항 로지스틱 회귀분석 8
- 산포 척도 값
  - 다항 로지스틱 회귀분석 11
- 상관행렬
  - 다항 로지스틱 회귀분석 10
- 상대적인 중위수 설명력
  - 프로빗 회귀 15
- 상수항
  - 선형 회귀분석 6
- 선형 회귀분석 18, 19
  - 가중 추정 20
  - 2단계 최소제곱 회귀분석 22
- 셀 확률표
  - 다항 로지스틱 회귀분석 10
- 수렴 기준
  - 다항 로지스틱 회귀분석 11
- 신뢰구간
  - 다항 로지스틱 회귀분석 10

## [아]

- 완전 요인 모형
  - 다항 로지스틱 회귀분석 8
- 우도비
  - 산포 척도 값 추정 11
  - 적합도 10
- 이분형 로지스틱 회귀분석 1, 2

## [자]

- 적합도
  - 다항 로지스틱 회귀분석 10
- 전진 선택법
  - 로지스틱 회귀분석 3
- 절편
  - 포함 또는 제외 8
- 제약된 회귀분석
  - 비선형 회귀분석 18
- 주효과 모형
  - 다항 로지스틱 회귀분석 8

## [차]

- 참조 범주
  - 다항 로지스틱 회귀분석 9

## [타]

- 통계 회귀 분석 18, 19

## [파]

- 편차 함수
  - 산포 척도 값 추정 11
- 평행성 검정
  - 프로빗 회귀 15
- 프로빗 분석
  - 명령 추가 기능 15
- 프로빗 회귀
  - 기준 15
  - 믿을만한 신뢰구간 15
  - 반복 15
  - 범위지정 14
  - 상대적인 중위수 설명력 15
  - 예제 13
  - 자연 반응을 15
  - 통계 13, 15
  - 평행성 검정 15

## [하]

- 회귀 분석 18, 19
  - 회귀 분석 소프트웨어 18, 19
- 후진제거법
  - 로지스틱 회귀분석 3

## [숫자]

- 2단계 최소제곱 회귀분석 22
  - 도구변수 22
  - 명령 추가 기능 23
  - 모수의 공분산 23
  - 새 변수 저장 23
  - 예제 22
  - 통계 22
- 2차 모형의 비율
  - 비선형 회귀분석 17
- 3차모형의 비율
  - 비선형 회귀분석 17

## C

- Cook의 거리
  - 로지스틱 회귀분석 5

- Cox 및 Snell R-제곱
  - 다항 로지스틱 회귀분석 10

## D

- DfBeta
  - 로지스틱 회귀분석 5

## G

- Gauss 모형
  - 비선형 회귀분석 17
- Gompertz 모형
  - 비선형 회귀분석 17

## H

- Hosmer-Lemeshow 적합도 통계량
  - 로지스틱 회귀분석 6

## J

- Johnson-Schumacher 모형
  - 비선형 회귀분석 17

## M

- McFadden R-제곱
  - 다항 로지스틱 회귀분석 10
- Metcherlich 수확 체감의 법칙
  - 비선형 회귀분석 17
- Michaelis Menten 모형
  - 비선형 회귀분석 17
- Morgan-Mercer-Florin 모형
  - 비선형 회귀분석 17

## N

- Nagelkerke R-제곱
  - 다항 로지스틱 회귀분석 10

## P

- Peal-Reed 모형
  - 비선형 회귀분석 17
- Pearson 카이제곱
  - 산포 척도 값 추정 11
  - 적합도 10

## R

Richards 모형

비선형 회귀분석 17

## S

SPSS 로지스틱 회귀 2

## V

Verhulst 모형

비선형 회귀분석 17

Von Bertalanffy 모형

비선형 회귀분석 17

## W

Weibull 모형

비선형 회귀분석 17







