

다항자료회귀분석

다항자료회귀분석은 종속변수가 범주형 변수(categorical variable)일 때 일반화선형모형을 이용하여 회귀모형을 적합시키는 통계분석 모듈입니다. 종속변수가 다항 분포를 따른다고 가정하며, 자료가 각 범주에 포함될 확률을 통계적 모형에 사용합니다. 종속변수가 명목형일 경우, 기저범주에 포함될 확률 대비 각 범주에 포함될 확률의 비율이, 종속변수가 순서형일 경우, 각 순서 이하의 범주에 포함될 확률이 통계적 모형에 들어갑니다. 순서형 종속변수의 경우 설명변수가 양적변수일 때, 비례오즈모형을 사용할 수 있으며, 이 모형에서는 설명변수의 변화량에 따른 확률의 변화를 범주마다 일정하다고 가정할 수 있습니다.

메뉴 호출하기

- 회귀분석 > 다항자료회귀분석



• 변수설정 탭

다항자료회귀분석

변수설정 분석옵션 출력옵션 변수선택

데이터

전체변수

X2
X1X2
X1X3
X1

① 종속변수(필수)

> treatment
<

설명변수

② 질적변수(선택-1개이상가능)

> X3
<

③ 양적변수(선택-1개이상가능)

> Y
<

④ ▼주효과 ⑤ ▼교호작용

⑥ 최종모형

X3
Y

삭제

도움말 재설정 확인 취소

메뉴 요소	설명
① 종속변수	모형화하고자 하는 종속변수를 전체변수로부터 선택합니다. 반드시 1개 이상의 질적 변수가 선택되어야 합니다. 이 질적변수는 3개 이상의 수준을 가지는 범주형 자료(categorical data)만 가능합니다.
② 질적변수	설명변수에 포함된 변수들의 유형을 지정해줍니다. 질적변수로 지정된 변수는 문자로 인식되어 분석에 사용됩니다. 엑셀 시트 상에 가장 먼저 등장하는 수준이 기저범주(reference)로 인식됩니다.
③ 양적변수	설명변수에 포함된 변수들의 유형을 지정해줍니다. 문자형 변수는 선택될 수 없으며, 선택된 경우 분석에서 제외됩니다. 질적변수에 지정된 변수와 중복되어 선택될 수 없습니다.
④ 주효과	[질적변수]와 [양적변수]에서 유형이 지정된 변수를 1개 이상 선택한 상태에서 [주효과] 버튼을 클릭하면, 해당 변수들이 최종모형에 각각 주효과로 포함됩니다.
⑤ 교호작용	[질적변수]와 [양적변수]에서 유형이 지정된 변수를 2개 이상 선택한 상태에서 [교호작용] 버튼을 클릭하면, 해당 변수들의 교호작용이 최종모형에 포함됩니다.
⑥ 최종모형	주효과 또는 교호작용으로 정의된 변수들이 설명변수로 간주되어 모형에 포함됩니다. 포함된 주효과 또는 교호작용 중 삭제하고자 하는 항목이 있는 경우, 해당 항목을 선택한 뒤 [삭제] 버튼을 클릭하면 최종 모형에서 제외됩니다.

• 분석옵션 탭

다항자료회귀분석

변수설정 | **분석옵션** | 출력옵션 | 변수선택

① 종속변수타입

② ☐ 종속변수 범주 개수 20개 초과 허용

• 명목형

③ 기저범주 선택 A

④ ☐ 평행회귀모형 (Parallel regression)

• 순서형

⑤ 범주순서(오름차순)

☒ 알파벳순 ☐ 사용자정의

B
C

▲ Low
↓
▼ High

⑥ 연결함수

☒ logit ☐ probit ☐ cloglog
☐ cauchit ☐ logc

⑦ ☐ 평행회귀모형 (Parallel regression)

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
1. 종속변수타입	<p>종속변수에 지정된 다항자료의 유형으로 2가지 중 하나를 선택합니다.</p> <ul style="list-style-type: none"> • 명목형 (Default) : 종속변수가 명목형(nominal)인 경우 선택합니다. [연결함수]로 'logit' 함수만 지원합니다. • 순서형 : 종속변수가 순서형(ordinal)인 경우 선택합니다. [연결함수]로 5가지 함수를 지원합니다.
2. 종속변수 범주 개수 20개 초과 허용	종속변수에 포함된 범주의 개수가 20개를 넘어갈 경우 이 옵션을 선택합니다.
③ 기저범주 선택	<p>[종속변수타입]으로 [명목형]을 선택하는 경우 활성화됩니다.</p> <p>콤보박스에 나열된 종속변수의 범주 중 하나를 기저범주(baseline category)로 지정합니다.</p>
4. 명목형 > 평행회귀모형 (Parallel regression)	<p>[종속변수타입]으로 [명목형]을 선택하는 경우 활성화됩니다.</p> <p>설명변수의 효과가 기저범주와 임의의 한 범주의 비율에 대하여 모두 동일하다고 가정하는 경우 이 옵션을 선택합니다. 모든 범주에 대해 회귀모형의 기울기가 동일하게 출력됩니다.</p>

• 분석옵션 탭

다항자료회귀분석

변수설정 | **분석옵션** | 출력옵션 | 변수선택

① 종속변수타입

② ☐ 종속변수 범주 개수 20개 초과 허용

• 명목형

③ 기저범주 선택 A

④ ☐ 평행회귀모형 (Parallel regression)

• 순서형

⑤ 범주순서(오름차순)

☒ 알파벳순 ☐ 사용자정의

B
C

▲ Low
↓
▼ High

⑥ 연결함수

☒ logit ☐ probit ☐ cloglog

☐ cauchit ☐ logc

⑦ ☐ 평행회귀모형 (Parallel regression)

도움말 | 재설정 | **확인** | 취소

메뉴 요소	설명
5. 범주순서(오름차순)	<p>[종속변수타입]으로 [순서형]을 선택하는 경우 활성화됩니다.</p> <p>종속변수의 범주의 순서로, 다음 2가지 방식 중 1개를 선택합니다.</p> <ul style="list-style-type: none"> 알파벳 (Default) : 범주 순서를 알파벳 순으로 자동 설정합니다. 사용자정의 : 사용자가 범주의 순서를 지정할 수 있습니다. 이 옵션이 선택되면 우측 박스가 활성화되어 종속변수의 범주가 나열됩니다. 원하는 범주를 선택하고 우측의 이동버튼을 이용하여 순서를 변경합니다. 맨 위(Low)에 위치한 범주가 기저범주(baseline category)로 지정됩니다.
6. 연결함수	<p>[종속변수타입]으로 [순서형]을 선택하는 경우 활성화됩니다. 다항자료에 대한 일반화선형모형을 적합하기 위해 필요한 연결함수로, 다음 5가지 옵션 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> logit (Default) : $f(x) = \log\left(\frac{x}{1-x}\right)$ probit : $f(x) = \Phi^{-1}(x)$ (Φ^{-1}은 표준정규분포의 역함수) cloglog (Complementary log-log function) : $f(x) = \log(-\log(1-x))$ cauchit (Cauchy quantile function) : $f(x) = \tan(\pi(x-0.5))$ logc : $f(x) = \log(1-x)$
7. 순서형 > 평행회귀모형 (Parallel regression)	<p>[종속변수타입]으로 [순서형]을 선택하는 경우 활성화됩니다.</p> <p>설명변수의 효과가 순서대로 범주 간의 비율에 대하여 모두 동일하다고 가정하는 경우 이 옵션을 선택합니다. 모든 범주에 대해 회귀모형의 기울기가 동일하게 출력됩니다.</p>

• 출력옵션 탭

다항자료회귀분석

변수설정 분석옵션 **출력옵션** 변수선택

출력

① 회귀계수

☒ 신뢰구간 ☐ 지수 모수 추정값
신뢰수준

② ☒ 분산분석표

☒ Type II ☐ Type III

③ ☒ 분류표 (관측 vs 예측)

요약표

④ ☐ 단순다항회귀 추가

⑤ 저장

☐ 예측 확률
☐ 잔차
☐ 표준화 잔차
☐ 선형예측값
☐ 헤트 행렬의 대각원소

도움말 재설정 **확인** 취소

메뉴 요소	설명
① 회귀계수	<ul style="list-style-type: none"> 신뢰구간 : 회귀계수 추정값의 신뢰구간을 출력합니다. 신뢰수준 : [신뢰구간] 옵션이 선택되는 경우 활성화됩니다. 0에서 1 사이 값을 입력할 수 있으며, Default는 0.95입니다. 지수 모수 추정값 : 회귀계수의 지수변환값을 출력합니다. 연결함수로 logit을 선택한 경우, 지수 모수 추정값은 오즈비(Odds ratio)에 해당합니다.
2. 분산분석표	<p>회귀모형식에 대한 분산분석표(ANOVA table)가 출력됩니다. 종속변수의 변동량에 대한 회귀모형의 설명력을 판단하기 위한 제공합 계산 방식 2가지 중 하나를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> Type II (Default) : 전체 회귀모형에서 주효과를 하나씩 제거하면서 제공합의 감소량을 계산하는 방식입니다. 최종모형이 주효과만으로 구성된 경우에 적합합니다. Type III : 전체 회귀모형에서 주효과와 교호작용을 포함한 모든 효과를 하나씩 제거하면서 제공합의 감소량을 계산하는 방식입니다. 최종모형에 교호작용도 포함된 경우 적합합니다.
③ 분류표 (관측 vs 예측)	<p>종속변수의 관측값과 예측값의 분류표(classification table)을 출력합니다. 예측값은 해당 개체의 각 범주 별 예측확률 중 가장 높은 범주로 계산됩니다.</p>

출력옵션 탭

다항자료회귀분석

변수설정 분석옵션 **출력옵션** 변수선택

출력

① 회귀계수

☒ 신뢰구간 ☐ 지수 모수 추정값
신뢰수준

② ☒ 분산분석표

☒ Type II ☐ Type III

③ ☒ 분류표 (관측 vs 예측)

요약표

④ ☐ 단순다항회귀 추가

⑤ 저장

☐ 예측 확률
☐ 잔차
☐ 표준화 잔차
☐ 선형예측값
☐ 해트 행렬의 대각원소

도움말 재설정 **확인** 취소

메뉴 요소	설명
④ 단순다항회귀 추가	[변수설정] 탭에서 2개 이상의 설명변수가 최종모형에 포함된 경우 이 옵션이 활성화됩니다. 이 옵션을 선택할 경우, 최종모형에 포함된 2개 이상의 설명변수에 대해 단순회귀모형을 적합한 회귀계수표가 출력됩니다.
⑤ 저장	모형 적합 후 다음 중 선택되는 통계량을 엑셀시트에 저장합니다. 괄호 안에 표기된 변수명으로 저장됩니다. 변수명의 i는 범주번호로, 범주가 n개인 경우 개별 변수가 n개씩 생성됩니다. 단, 예측 확률을 제외한 나머지 통계량은 기저범주는 제외한 (n-1)개의 변수가 생성됩니다. <ul style="list-style-type: none"> 예측 확률 : 최종모형으로 예측된 범주 별 확률 (MultinomGLM_Fitted_i) 잔차 : 적합값과 실제값의 차이 (MultinomGLM_Resid_log_mu_i). 표준화 잔차 : 잔차를 표준편차로 나눈 값 (MultinomGLM_stdResid_log_mu_i) 선형예측값 : 일반화선형모형에서 도출된 역변환되기 전 선형예측값 (MultinomGLM_LinearPred_log_mu_i) 해트 행렬의 대각원소 : 해당 개체와 나머지 개체의 평균의 차이인 지렛값 (MultinomGLM_HatValue_log_mu_i)

• 변수선택 탭

다항자료회귀분석

변수설정 분석옵션 출력옵션 변수선택

① ☒ 변수선택

② 변수선택방법

☒ 전진선택법 ☐ 후진제거법 ☐ 단계별선택법

설명변수 ③ 고정변수

X3
Y

>
<

도움말 재설정 확인 취소

메뉴 요소	설명
① 변수선택	[변수설정] 탭에서 2개 이상의 설명변수가 최종모형에 포함된 경우 이 옵션이 활성화됩니다. 이 옵션을 선택할 경우, AIC(Akaike Information Criterion)을 기반으로 최적의 모형에 해당하는 변수를 자동으로 선택합니다. 최적의 모형은 Rex output의 Variable Selection에서 확인할 수 있습니다.
② 변수선택방법	<p>[변수선택]을 선택한 경우 활성화됩니다. 다음 3가지 옵션 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> 전진선택법 (Default) : 상수항만 있는 모형에서 시작하여 최종모형에 선택된 설명변수들 중 종속변수에 가장 큰 영향력을 갖는 변수부터 하나씩 차례대로 추가하면서 최적의 모형을 선택하는 방식입니다. 후진제거법 : 최종모형에 선택된 설명변수들을 모두 포함한 모형에서 시작하여 종속변수에 가장 작은 영향력을 갖는 변수부터 하나씩 차례대로 제거하면서 최적의 모형을 선택하는 방식입니다. 단계별선택법 : 상수항만 있는 모형에서 시작하여 최종모형에 선택된 설명변수들 중 하나씩 모형에 추가하면서 매번 모든 설명변수를 평가하는 방식으로 전진선택법과 후진선택법을 반복적으로 수행하는 방식입니다.
③ 고정변수	[변수설정] 탭에서 선택된 2개 이상의 변수가 나열된 설명변수 창에서 변수선택과정과 상관없이 최종 모형에 반드시 포함되어야 하는 변수들을 고정변수로 지정합니다.