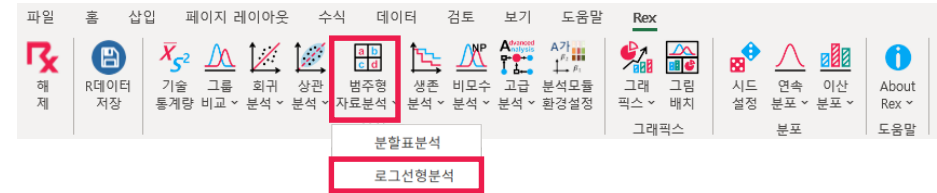


# 로그선형분석

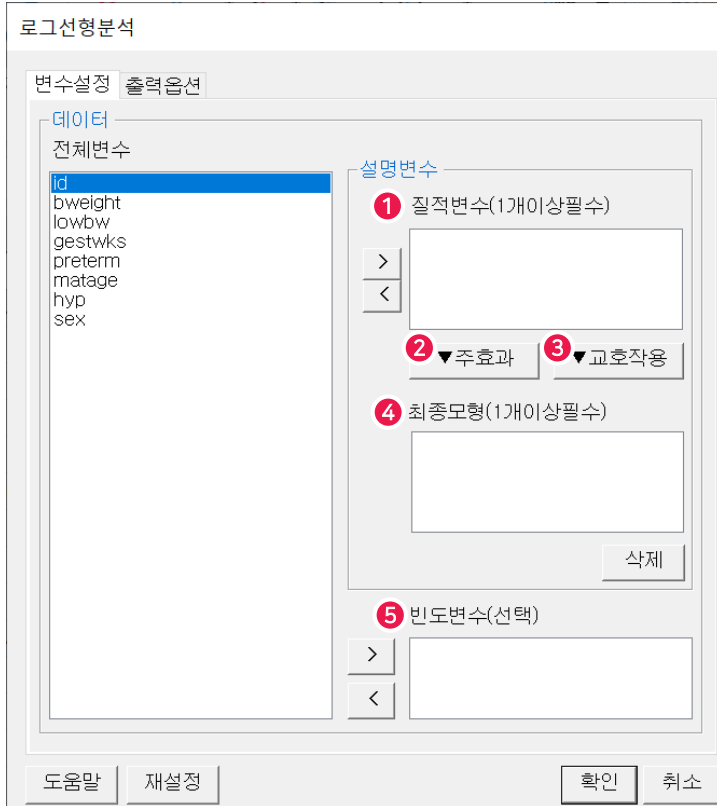
## 메뉴 호출하기

- Rex > 범주형자료분석 > 로그선형분석



로그선형분석은 분할표의 셀 빈도를 범주형 변수들의 관계로 표현하며, 이를 통해 변수들 사이의 연관성(association)과 교호작용(interaction)을 분석할 수 있는 통계분석 모듈입니다. 분할표 분석에서는 변수들 사이의 연관성(association)을 설명할 수 있지만 어떤 범주가 중요한 역할을 하는지는 알기 어렵습니다. 로그선형분석에서는 셀의 기대 빈도의 로그값이 설명변수의 선형함수(linear function)로 표현되어, 변수들 사이의 관계를 설명합니다. 본 모듈에서는 사용자가 세운 선형모형(linear model)에 대한 적합도 검정(goodness-of-fit test), 설명변수에 대한 모수 추정 및 가설 검정을 실시할 수 있습니다.

• 변수설정 탭



메뉴 요소	설명
① 질적변수	설명변수에 포함된 질적변수를 지정해줍니다. 지정된 변수는 문자로 인식되어 분석에 사용됩니다. 엑셀 시트 상에 가장 먼저 등장하는 수준이 기준 범주(reference)로 인식됩니다.
② 주효과	[질적변수]에서 지정된 변수를 1개 이상 선택한 상태에서 [주효과] 버튼을 클릭하면, 해당 변수들이 최종모형에 각각 주효과로 포함됩니다.
③ 교호작용	[질적변수]에서 지정된 변수를 2개 이상 선택한 상태에서 [교호작용] 버튼을 클릭하면, 해당 변수들의 교호작용이 최종모형에 포함됩니다.
④ 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.
⑤ 빈도변수	개별 범주의 조합에 대한 빈도값이 정리된 변수가 있는 경우 빈도변수로 지정합니다. 양의 정수로 입력된 1개의 양적변수를 선택할 수 있습니다. 빈도 변수가 선택되지 않더라도 분석을 수행할 수 있습니다.

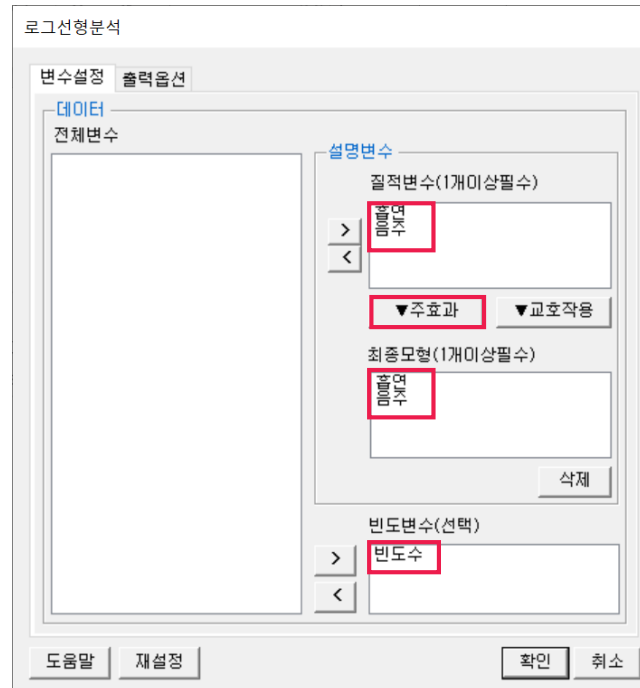
출력옵션 탭

메뉴 요소	설명
① 출력 > 모수 (회귀계수)	<ul style="list-style-type: none"> <li>신뢰구간 : 회귀계수 추정값의 신뢰구간을 출력합니다.</li> <li>신뢰수준 : [신뢰구간] 옵션이 선택되는 경우 활성화됩니다. 신뢰수준은 0과 1 사이 값을 입력할 수 있습니다. 신뢰수준의 Default는 0.95입니다.</li> <li>지수 모수 추정값 : 회귀계수의 지수변환값을 출력합니다. 연결함수로 logit을 선택한 경우, 지수 모수 추정값은 오즈비(Odds ratio)에 해당합니다.</li> </ul>
② 출력 > 분산분석표	<p>모형식에 대한 분산분석표(ANOVA table)이 출력됩니다. 회귀모형의 설명력을 판단하기 위한 제공할 계산 방식으로 다음의 3가지 옵션 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> <li>Type I : 설명변수를 순차적으로 하나씩 추가하면서 제공합의 증가량을 계산하는 방식입니다.</li> <li>Type II (Default) : 전체 회귀모형에서 주효과를 하나씩 제거하면서 제공합의 감소량을 계산하는 방식입니다. 최종모형이 주효과만으로 구성된 경우에 적합합니다.</li> <li>Type III : 전체 회귀모형에서 주효과와 교호작용을 포함한 모든 효과를 하나씩 제거하면서 제공합의 감소량을 계산하는 방식입니다. 최종모형에 교호작용도 포함된 경우 적합합니다.</li> </ul>
③ 그래프 > 막대그래프	두 종속변수에 대한 막대그래프가 출력됩니다. 변수의 수준별 빈도를 확인할 때 유용합니다.
④ 그래프 > 풍선그래프	두 종속변수에 대한 풍선그래프가 출력됩니다. 변수의 수준별 상대빈도를 확인할 때 유용합니다.
⑤ 그래프 > 모자이크그래프	두 종속변수에 대한 모자이크그래프가 출력됩니다. 변수의 수준별 상대빈도를 확인할 때 유용
⑥ 출력 > 분할표	설명변수로 지정된 질적변수들의 분할표가 출력됩니다. 질적변수의 수가 3개 이하 일 때에만 가능합니다.

• 예제

흡연 경험(Yes=1, No=0)과 음주 경험(Yes=1, No=0) 사이에 연관성이 있는지 알아보기 위하여 조사한 데이터이다. 포화 로그선형모형과 독립성 로그선형모형 중 더 적절한 모형을 선택하고, 해석 가능한 효과들을 해석하라. [데이터 : manual\_data.xlsx > 흡연]

1. [변수설정] > [질적변수] > '흡연', '음주' 지정 > [주효과] 클릭
2. [변수설정] > [최종모형] > '흡연', '음주' 추가 확인
3. [변수설정] > [빈도변수] > '빈도수' 지정
4. [출력옵션] > [출력] > [모수(회귀계수)] > [지수모수추정값] 체크
5. [출력옵션] > [출력] > [분할표] 체크
6. [확인]



예제 - 결과창

## Log Linear Analysis

### Data Structure

No. of total observations	4
No. of used observations	4
No. of used variable(s)	3

### Variable List

Category	Variable	N	N.valid	(%.valid)	N.miss	(%.miss)
Categorical variable	흡연	4	4	(100.00%)	0	(0.00%)
Categorical variable	음주	4	4	(100.00%)	0	(0.00%)
Frequency variable	빈도수	4	4	(100.00%)	0	(0.00%)

- Warning : The type of variable '흡연, 음주' was numeric, but was coerced into character. (Log linear analysis is supported only for categorical variables.)

### Analysis Description

Model	빈도수 ~ 흡연 + 음주
Saturated model	FALSE

- All hierarchical terms of interactions are included in the model. **Bold type** indicates added terms. (\*) indicates the interaction term.)

### Contingency Table

	음주(0)	음주(1)	Total
흡연(0)	36	14	50
흡연(1)	52	131	183
Total	88	145	233

[Data Structure]

- No. of total observations : 전체 개체 수
- No. of used observations : 분석에 사용된 개체 수
- No. of used variable(s) : 분석에 사용된 변수의 수

[Variable List]

- 분석에 사용된 개별 변수들의 유효개수(N.valid), 유효퍼센트(%.valid), 결측수(N.miss), 결측퍼센트(%.miss)가 출력됩니다.

[Analysis Description]

- Rex 분석 메뉴에서 지정된 내용이 출력됩니다.
- Model : [변수설정] > [최종모형]에 지정된 모형 (빈도수 ~ 흡연 + 음주)
- Saturated model : 교호작용까지 포함된 경우 Saturated model로 판단됩니다.

[Contingency Table]

- 흡연과 음주의 빈도표입니다.
- 비흡연의 경우 음주하는 경우는 50명 중 14명이고, 흡연의 경우 음주하는 경우는 183명 중 131명입니다.

예제 - 결과창

Results of Log Linear Analysis

Coefficient Estimates

	Estimate	exp (Estimate)	SE	Z-value	P-value	Lower bound of 95% CI (Wald)	Upper bound of 95% CI (Wald)	Lower bound of 95% CI (Profile likelihood)	Upper bound of 95% CI (Profile likelihood)
(Intercept)	2.9383	18.8841	0.1645	17.8583	2.490x10 <sup>-71</sup>	2.6158	3.2608	2.6033	3.2491
흡연1	1.2975	3.6600	0.1596	8.1307	4.268x10 <sup>-16</sup>	0.9847	1.6102	0.9933	1.6203
음주1	0.4994	1.6477	0.1351	3.6957	0.0002	0.2345	0.7642	0.2370	0.7675

Model Fitness Measurements

	Value	DF
Deviance	31.1836	1
Pearson's chi-square	31.7391	1
-2*log-likelihood	53.6052	3
AIC	59.6052	
BIC	57.7641	

• A model with a smaller value is better.

ANOVA Table

	SS	DF	F	P-value
흡연	80.6969	1	2.5425	0.3566
음주	14.0867	1	0.4438	0.6259
Residuals	31.7391	1		

Used R Packages

- Main results : 'glm' of R package 'stats'
- Confidence interval : 'confint' of R package 'stats'
- Model fitness measurements : 'deviance', 'residuals', 'logLik', 'AIC', 'BIC' of R package 'stats'
- ANOVA table : 'anova' of R package 'stats', 'Anova' of R package 'car'
- All results other than those mentioned above were written with basic functions of R.

[Coefficient Estimates]

- 표에 제시된 신뢰구간은 estimate 값에 대한 신뢰구간입니다. 오즈비의 신뢰구간을 구하기 위해서는 지수변환이 필요합니다.
- 음주에 대한 흡연의 오즈비는 3.6600 (95% 신뢰구간: exp(0.9847)=2.6770, exp(1.6102)=5.0038)배이며, 이는 통계적으로 유의합니다 (p<0.0001).
- 흡연에 대한 음주의 오즈비는 1.6477 (95% 신뢰구간: exp(0.2345)=1.2643, exp(0.7642)=2.1473)배이며, 이는 통계적으로 유의합니다 (p=0.0002).

[Model Fitness Measurements]

- 모형의 적합도를 나타내는 지표가 출력됩니다. 동일 데이터에 대해 여러 개의 모형에서 적합도 지표가 산출된 경우 비교 가능하지만, 현재 모형은 하나 뿐이므로 해석에 의미가 없습니다.

[ANOVA Table]

- 최종모형 내에서 개별 변수들의 유의성을 확인할 수 있습니다.
- 흡연 변수의 경우 p=0.3566로 유의수준 0.05 하에서 통계적으로 유의하지 않다고 판단됩니다.
- 음주 변수의 경우 p=0.6259로 유의수준 0.05 하에서 통계적으로 유의하지 않다고 판단됩니다.

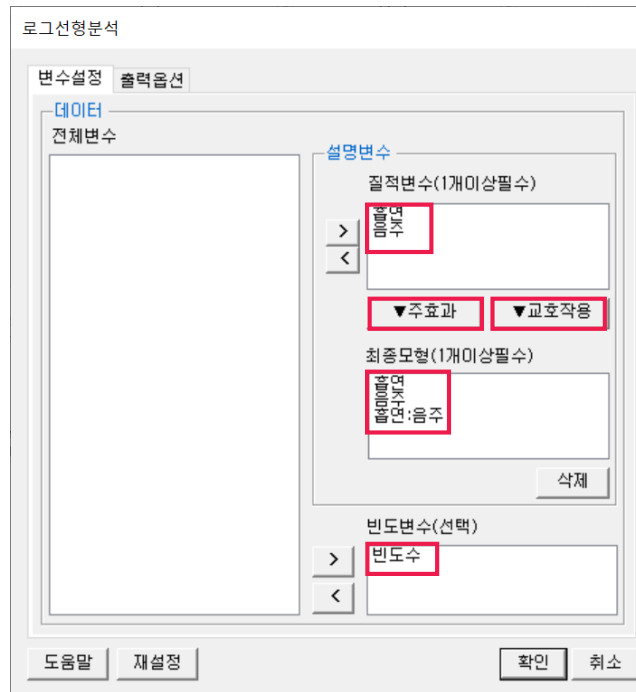
[Used R Packages]

- 본 분석에 사용된 R 패키지들이 나열되어 있습니다.

• 예제

흡연 경험(Yes=1, No=0)과 음주 경험(Yes=1, No=0) 사이에 연관성이 있는지 알아보기 위하여 조사한 데이터이다. 포화 로그선형모형과 독립성 로그선형모형 중 더 적절한 모형을 선택하고, 해석 가능한 효과들을 해석하라. [데이터 : manual\_data.xlsx > 흡연]

1. [변수설정] > [질적변수] > '흡연', '음주' 지정 > [주효과] 클릭
2. [변수설정] > [질적변수] > '흡연', '음주' 지정 > [교호작용] 클릭
3. [변수설정] > [최종모형] > '흡연', '음주', '흡연:음주' 추가 확인
4. [변수설정] > [빈도변수] > '빈도수' 지정
5. [출력옵션] > [출력] > [모수(회귀계수)] > [지수모수추정값] 체크
6. [출력옵션] > [출력] > [분할표] 체크
7. [확인]



예제 - 결과창

## Log Linear Analysis

### Data Structure

No. of total observations	4
No. of used observations	4
No. of used variable(s)	3

### Variable List

Category	Variable	N	N.valid	(%.valid)	N.miss	(%.miss)
Categorical variable	흡연	4	4	(100.00%)	0	(0.00%)
Categorical variable	음주	4	4	(100.00%)	0	(0.00%)
Frequency variable	빈도수	4	4	(100.00%)	0	(0.00%)

Warning : The type of variable '흡연, 음주' was numeric, but was coerced into character. (Log linear analysis is supported only for categorical variables.)

### Analysis Description

Model	빈도수 ~ 흡연 + 음주 + 음주:흡연 + 흡연:음주
Saturated model	TRUE

All hierarchical terms of interactions are included in the model. **Bold type** indicates added terms. (\*) indicates the interaction term.)

### Contingency Table

	음주(0)	음주(1)	Total
흡연(0)	36	14	50
흡연(1)	52	131	183
Total	88	145	233

[Data Structure]

- No. of total observations : 전체 개체 수
- No. of used observations : 분석에 사용된 개체 수
- No. of used variable(s) : 분석에 사용된 변수의 수

[Variable List]

- 분석에 사용된 개별 변수들의 유효개수(N.valid), 유효퍼센트(%.valid), 결측수(N.miss), 결측퍼센트(%.miss)가 출력됩니다.

[Analysis Description]

- Rex 분석 메뉴에서 지정된 내용이 출력됩니다.
- Model : [변수설정] > [최종모형]에 지정된 모형 (빈도수 ~ 흡연 + 음주 + 흡연:음주)
- Saturated model : 교호작용까지 포함된 경우 Saturated model로 판단됩니다.

[Contingency Table]

- 흡연과 음주의 빈도표입니다.
- 비흡연의 경우 음주하는 경우는 50명 중 14명이고, 흡연의 경우 음주하는 경우는 183명 중 131명입니다.



예제 - 결과창

Results of Log Linear Analysis

Coefficient Estimates

	Estimate	exp (Estimate)	SE	Z-value	P-value	Lower bound of 95% CI (Wald)	Upper bound of 95% CI (Wald)	Lower bound of 95% CI (Profile likelihood)	Upper bound of 95% CI (Profile likelihood)
(Intercept)	3.5835	36.0000	0.1667	21.5011	1.520x10 <sup>-102</sup>	3.2569	3.9102	3.2381	3.8933
흡연1	0.3677	1.4444	0.2168	1.6960	0.0899	-0.0572	0.7927	-0.0533	0.7998
음주1	-0.9445	0.3889	0.3150	-2.9986	0.0027	-1.5618	-0.3271	-1.5952	-0.3501
흡연1:음주1	1.8684	6.4780	0.3551	5.2622	1.424x10 <sup>-07</sup>	1.1725	2.5643	1.1924	2.5922

Model Fitness Measurements

	Value	DF
Deviance	-3.553x10 <sup>-15</sup>	0
Pearson's chi-square	1.430x10 <sup>-28</sup>	0
-2*log-likelihood	22.4215	4
AIC	30.4215	
BIC	27.9667	

- Note : Since the proposed model is saturated, degrees of freedom of Deviance and Pearsort's chi-square are zero.
- A model with a smaller value is better.

ANOVA Table

- Warning : Anova table is not supported for the saturated model.

Used R Packages

- Main results : 'glm' of R package 'stats'
- Confidence interval : 'confint' of R package 'stats'
- Model fitness measurements : 'deviance', 'residuals', 'logLik', 'AIC', 'BIC' of R package 'stats'
- ANOVA table : 'anova' of R package 'stats', 'Anova' of R package 'car'
- All results other than those mentioned above were written with basic functions of R.

[Coefficient Estimates]

- 표에 제시된 신뢰구간은 estimate 값에 대한 신뢰구간입니다. 오즈비의 신뢰구간을 구하기 위해서는 지수변환이 필요합니다.
- 음주에 대한 흡연의 오즈비는 1.444 (95% 신뢰구간:  $\exp(-0.0572)=0.9444$ ,  $\exp(0.7927)=2.2094$ )배이며, 이는 통계적으로 유의합니다 ( $p=0.0899$ ).
- 흡연에 대한 음주의 오즈비는 0.3889 (95% 신뢰구간:  $\exp(-1.5618)=0.2098$ ,  $\exp(-0.3271)=0.7210$ )배이며, 이는 통계적으로 유의합니다 ( $p=0.0027$ ).
- 교호작용의 경우  $p<0.0001$ 로 행효과인 흡연여부와 열효과인 음주여부 간의 교호작용이 존재한다고 볼 수 있으며, 두 변수는 통계적으로 독립이라고 볼 수 없습니다. 이 경우, 교호작용만 해석이 가능하고 각각의 주효과를 해석하는 것은 신뢰할 수 없습니다. 흡연여부와 음주여부의 2X2 빈도표에 대한 오즈비만 유의하며 이것은  $\exp(1.8684)=6.4780$ 으로 추정할 수 있습니다.

[Model Fitness Measurements]

- 모형의 적합도를 나타내는 지표가 출력됩니다.
- 교호작용이 없던 모형의 경우 Deviance 31.1836, AIC=59.6052, BIC=57.7641로 흡연과 음주의 교호작용이 추가된 모형이 그보다 Deviance, AIC, BIC가 모두 낮아져 데이터에 더 적합한 모형으로 판단됩니다.

[ANOVA Table]

- 포화모형에서는 개별 변수들의 유의성을 확인할 수 없습니다.

[Used R Packages]

- 본 분석에 사용된 R 패키지들이 나열되어 있습니다.