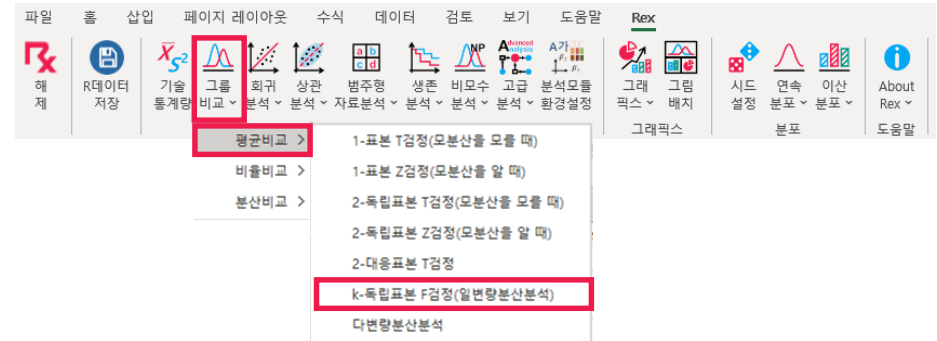


k-독립표본 F검정 (일반량 분산분석)

메뉴 호출하기

- Rex > 그룹비교 > 평균비교 > k-독립표본 F검정(일반량분산분석)



k-독립표본 F검정(일반량분산분석)은 하나 이상의 요인(factor)의 수준에 따라 한 종속변수의 값이 유의한 차이를 보이는지 관찰할 수 있는 통계분석 모듈입니다. 이 분석에서는 각 요인 또는 그 조합의 수준에 따라 종속변수의 값이 얼마나 차이가 있는지 계산하고, 이를 잔차와 비교하여 이 차이가 유의한지 검정함으로써 어느 요인에 따라 종속변수의 값이 큰 차이를 보이는지 확인할 수 있습니다. Rex에서는 분산분석표(ANOVA table)로 결과를 확인할 수 있으며, 다양한 방식의 사후검정을 손쉽게 할 수 있습니다.

• 변수설정 탭

k-독립표본 F검정(일반량분산분석)

변수설정 분석옵션 출력옵션

데이터

전체변수

id
bweight
lowbw
gestwks
preterm
matage
hyp
sex

1 종속변수(1개이상필수)

> <

변수유형

2 요인(1개이상필수)

> <

3 ▼주효과 4 ▼교호작용

5 최종모형(1개이상필수)

삭제

다음말 재설정 확인 취소

메뉴 요소	설명
① 종속변수	평균을 비교하고자 하는 변수를 전체변수로부터 선택합니다. 반드시 1개 이상의 양적 변수가 선택되어야 합니다.
② 요인	요인에 해당하는 변수를 전체변수로부터 1개 이상 선택합니다. 종속변수와 중복하여 선택될 수 없습니다. 선택된 변수들은 질적 변수로 인식되어 분석에 사용됩니다.
③ 주효과	요인으로 지정된 변수를 1개 이상 선택한 상태에서 [주효과] 버튼을 클릭하면, 개별 요인들이 최종모형에 각각 주효과로 포함됩니다.
④ 교호작용	요인으로 지정된 변수를 2개 이상 선택한 상태에서 [교호작용] 버튼을 클릭하면, 선택된 변수들의 교호작용이 최종모형에 포함됩니다.
⑤ 최종모형	요인으로 지정된 변수들 중 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

k-독립표본 F검정(일반량분산분석)

변수설정 분석옵션 출력옵션

① -제공합 유형
 Type I Type II Type III

② 사후분석수행

③ 사후분석변수

④ 등분산 가정함
 HSD S-N-K
 LSD Duncan
 Scheffe Dunnett

⑤ 등분산 가정하지 않음
 Tamhane's T2 Dunnett T3
 Games-Howell Dunnett C

⑥ 신뢰수준

도움말 재설정 확인 취소

메뉴 요소	설명
① 제공합 유형	<p>각 변수별 F통계량을 계산하기 위한 제공합 계산방식으로 3가지 옵션 중 1개를 선택합니다.</p> <ul style="list-style-type: none"> Type I : 최종모형에 나열된 요인을 입력된 순서대로 검정합니다. Type II : 최종모형에서 교호작용은 고려하지 않고, 서로 다른 주효과만 고려한 상태에서 개별 주효과를 검정합니다. Type III : 최종모형에서 모형 내 다른 주효과 및 교호작용을 모두 고려한 상태에서 개별 요인을 검정합니다.
② 사후분석수행	사후분석을 하고자 하는 경우 선택합니다.
③ 사후분석변수	[사후분석수행]이 선택된 경우 활성화됩니다. 콤보박스에 나열된 최종모형에서 고려된 요인들 중 1개를 선택합니다. 선택된 요인에 대하여 사후분석이 수행됩니다.
④ 등분산 가정함	<p>[사후분석수행]이 선택된 경우 활성화됩니다. [사후분석변수]에서 선택된 요인의 수준별 등분산을 가정한 상태에서 적용할 수 있는 사후분석 방법을 선택할 수 있습니다.</p> <ul style="list-style-type: none"> HSD : Tukey의 HSD에 해당합니다. 검정력이 낮다고 알려져 있습니다. 집단별 대상자수가 동일해야 합니다. LSD : Fisher's LSD에 해당합니다. 가장 엄격하지 않은 사후검정방법으로 최근 연구에서는 선호되지 않습니다. Scheffe : 가장 보수적이고 엄격한 방법으로 알려져 있습니다. S-N-K : Student-Newman-Keuls 방법으로 Tukey 방법의 변형방식에 해당합니다. 집단별 대상자수가 동일해야 합니다. Duncan : Duncan's method에 해당합니다. Dunnett : Dunnett's procedure에 해당합니다. Dunnett이 선택될 경우 [사후분석변수]에서 선택된 요인의 수준이 콤보박스에 나열되며, 선택된 수준이 기저범주(reference)로 간주되어 이와 다른 집단들과의 차이에 대해 분석합니다. 모든 조합에 대한 검정을 하지 않습니다.

• 분석옵션 탭

k-독립표본 F검정(일반량분산분석)

변수설정 분석옵션 출력옵션

① -제공합 유형
 Type I Type II Type III

② 사후분석수행

③ 사후분석변수

④ 등분산 가정함
 HSD S-N-K
 LSD Duncan
 Scheffe Dunnett

⑤ 등분산 가정하지 않음
 Tamhane's T2 Dunnett T3
 Games-Howell Dunnett C

⑥ 신뢰수준

도움말 재설정 확인 취소

메뉴 요소	설명
⑤ 등분산 가정하지 않음	<p>[사후분석수행]이 선택된 경우 활성화됩니다. [사후분석변수]에서 선택된 요인의 수준별 등분산을 가정하지 않는 경우 적용할 수 있는 사후분석 방법을 선택할 수 있습니다. 4가지 옵션 모두 집단별 대상자의 수가 동일하지 않아도 적용이 가능합니다.</p> <ul style="list-style-type: none"> Tamhane's T2 : 집단별 대상자의 수가 동일하지 않아도 적용이 가능합니다. Games-Howell보다 엄격하다고 알려져 있으며, 대상자수가 많아질수록 1종 오류가 높아진다는 단점이 있습니다. Games-Howell : 집단별 대상자의 수가 동일하지 않아도 적용이 가능합니다. 표본수가 6개 미만인 경우 1종 오류가 높아진다는 단점이 있습니다. Dunnett T3 : 집단별 대상자의 수가 동일한 경우에 적용 가능합니다. 집단별 대상자수가 50개 미만인 경우 Games-Howell보다 검정력이 높지만, 50개 이상인 경우는 Games-Howell보다 1종 오류가 높아진다는 단점이 있습니다. Dunnett C : 집단별 대상자의 수가 거의 동일한 경우에 사용하는 것을 권장합니다. 집단별 대상자수가 50개 이상일 때 Games-Howell보다 신뢰구간을 더 좁게 제시하며 검정력이 좋다고 알려져 있습니다.
⑥ 신뢰수준	<p>[사후분석수행]이 선택된 경우 활성화됩니다. 사후분석을 수행할 때 적용되는 신뢰수준으로 0과 1 사이 값을 입력할 수 있습니다. 신뢰수준의 Default는 0.95입니다.</p>

• 출력옵션 탭



메뉴 요소	설명
① 잔차진단 그래프	최종모형에 설정된 선형모형을 적합시킨 후 계산되는 잔차(residual), 쿡의 거리(Cook's distance), 지렛값(leverage value)를 이용한 잔차진단그래프(diagnostic graph)를 생성합니다.
② 박스그림	요인의 수준별로 종속변수의 박스그림을 생성합니다.
③ 적합값	모형에 의해 해당 요인에 대해 예측된 종속변수의 값을 엑셀 시트에 저장합니다. ANOVA_Fitted라는 변수명으로 저장됩니다.
④ 잔차	모형에 의해 해당 요인에 대해 예측된 적합값과 실제 관측값의 차이인 잔차를 엑셀 시트에 저장합니다. ANOVA_ResidOriginal 이라는 변수명으로 저장됩니다.
⑤ 표준화잔차	잔차를 표준편차로 나누어 표준화한 값을 엑셀 시트에 저장합니다. ANOVA_ResidStandardized라는 변수명으로 저장됩니다.
⑥ 스튜던트화잔차	해당 개체를 제외한 상태에서 계산된 표준편차로 잔차를 나눈 스튜던트화 잔차를 엑셀 시트에 저장합니다. ANOVA_ResidStudentized라는 변수명으로 저장됩니다.
⑦ 쿡의 거리	개별 개체들이 모형에 미치는 영향력을 평가하기 위해, 잔차와 지렛값을 동시에 고려한 척도인 쿡의 거리(Cook's distance)를 엑셀 시트에 저장합니다. ANOVA_CookDist라는 변수명으로 저장됩니다.

예제

소의 연령에 따른 육류의 셀레늄(selenium) 농도를 비교하였다. 연령은 총 4가지 그룹으로 나누었다. 4 가지 연령별로 육류에 존재하는 셀레늄 농도에 차이가 있는지 가설검정을 수행하고, 필요한 경우 Tukey의 HSD검정을 통해 사후검정하라. [데이터 : manual_data.xlsx > 셀레늄]

1. [변수설정] > [종속변수] '셀레늄' 지정
2. [변수설정] > [요인] '연령' 입력 > [주효과] 클릭 > [최종모형] '연령' 추가 확인
3. [분석옵션] > [사후분석수행] 체크 > [사후분석변수] '연령' 확인 > [HSD] 체크
4. [출력옵션] > [박스그림] 체크
5. [확인]

k-독립표본 F검정(일반량분산분석)

변수설정 분석옵션 출력옵션

데이터

전체변수

종속변수(1개이상필수)

셀레늄

변수유형

요인(1개이상필수)

연령

주효과 교호작용

최종모형(1개이상필수)

연령

삭제

도움말 재설정 확인 취소

k-독립표본 F검정(일반량분산분석)

변수설정 분석옵션 출력옵션

제공할 유형

Type I Type II Type III

사후분석수행

사후분석변수 연령

등분산 가정함

HSD S-N-K

LSD Duncan

Scheffe Dunnett A

등분산 가정하지 않음

Tamhane's T2 Dunnett T3

Games-Howell Dunnett C

신뢰수준 0.95

도움말 재설정 확인 취소

k-독립표본 F검정(일반량분산분석)

변수설정 분석옵션 출력옵션

출력

잔차진단 그래프

박스그림

저장

적합값

잔차

표준화잔차

스튜던트화잔차

록의 거리

도움말 재설정 확인 취소

예제 - 결과창

Analysis of Variance (ANOVA)

Data Structure

No. of total observations	113
No. of used observations	113
No. of used variable(s)	2

Variable List

Category	Variable	N	N.valid	(% .valid)	N.miss	(% .miss)
Dependent variable	셀레늄	113	113	(100.00%)	0	(0.00%)
Factor	연령	113	113	(100.00%)	0	(0.00%)

Analysis Description

Model	셀레늄 = 연령 + error
SS (sum of squares)	Type II
Posthoc analysis	Performed with 연령
Posthoc analysis method	Tukey's HSD
Confidence level for posthoc analysis	95%

Descriptive Statistics

Group	N.valid	Mean	SD
연령=A	22	1448.3636	629.126
연령=B	14	992.6429	627.263
연령=C	29	873.8276	304.7435
연령=D	48	851.7292	385.9067

[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
- SS (sum of squares) : 분산분석표 계산 시 사용될 제곱합 유형 (Type II)
- Posthoc analysis : 사후검정 분석변수 (연령)
- Posthoc analysis method : 사후검정 분석방법 (Tukey's HSD)
- Confidence level for posthoc analysis : 사후검정 신뢰수준 (0.95)

[Descriptive Statistics]

- 연령그룹 A의 유효개체수는 22개이며, 셀레늄 평균은 1448.3636, 표준편차는 629.126입니다.
- 연령그룹 B의 유효개체수는 14개이며, 셀레늄 평균은 992.6429, 표준편차는 627.263입니다.
- 연령그룹 C의 유효개체수는 29개이며, 셀레늄 평균은 873.8276, 표준편차는 304.7435입니다.
- 연령그룹 D의 유효개체수는 48개이며, 셀레늄 평균은 851.7292, 표준편차는 385.9067입니다.

• 예제 - 결과창

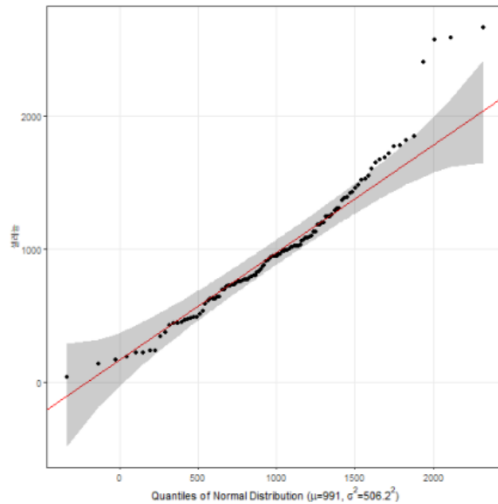
Assessing Test Assumptions

Univariate Normality

Shapiro-Wilk Test

	W	P-value
셀레늄	0.9456	0.0002

Univariate QQ plot



복사 저장

Levene's Test for Homogeneity of Variances between Groups in Factors

	F	D.F1	D.F2	P-value
연령	4.2728	3	109	0.0068

[Shapiro-Wilk test]

- 종속변수에 대한 정규성 검정 결과입니다.
- $p=0.0002$ 로 유의수준 0.05 하에서 종속변수가 정규성을 따른다고 할 수 없습니다.

[Univariate QQ plot]

- 종속변수의 양 끝 점이 직선에서 떨어져 있어 정규성을 따른다고 할 수 없습니다.

[Levene's Test for Homogeneity of Variance between Groups]

- 집단별 분산이 동일한지 검정한 결과, $p=0.0068$ 으로 유의수준 0.05 하에서 종속변수에 대한 집단별 분산이 동일하다고 할 수 없습니다.

• 예제 - 결과창

Results of Analysis of Variance (ANOVA)

Model Fitness Measurements

	Value	D.F
Deviance	23026499.9223	109
log-likelihood	-851.0395	5
AIC	1712.0789	
AICc	1712.6397	
BIC	1725.7159	

ANOVA Table

	D.F	SS	MS	F	P-value
연령	3	5931208.0423	1977069.3474	9.3588	1.476x10 ⁻⁰⁵
Residuals	109	23026499.9223	211252.2929		

Welch's ANOVA

F	D.F1	D.F2	P-value
5.752	3.000	38.960	0.0023

[Results of Analysis of Variance (ANOVA)]

[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]

- 최종모형의 통계적인 유의성을 보여줍니다.
- 등분산 가정이 만족되지 않으므로 본 결과는 해석하지 않습니다.

[Welch's ANOVA]

- 등분산 가정이 만족되지 않아, 로버스트한 Welch's ANOVA 결과를 확인합니다.
- p=0.0023으로 유의수준 0.05 하에서 최종모형이 셀레늄의 농도를 설명하기에 통계적으로 유의하다고 판단합니다.

• 예제 - 결과창

Post-hoc Analysis for variable '연령'

Post-hoc Analysis

Method	Comparison	Difference	95% LCI	95% UCI	P-value
Tukey	A-B	455.7208	45.7349	865.7066	0.023
	A-C	574.5361	235.4829	913.5892	0.0001
	A-D	596.6345	287.8823	905.3867	1.093x10 ⁻⁰⁵
	B-A	-455.7208	-865.7066	-45.7349	0.023
	B-C	118.8153	-271.4539	509.0844	0.8569
	B-D	140.9137	-223.3402	505.1676	0.7443
	C-A	-574.5361	-913.5892	-235.4829	0.0001
	C-B	-118.8153	-509.0844	271.4539	0.8569
	C-D	22.0984	-259.9468	304.1436	0.997
	D-A	-596.6345	-905.3867	-287.8823	1.093x10 ⁻⁰⁵
	D-B	-140.9137	-505.1676	223.3402	0.7443
	D-C	-22.0984	-304.1436	259.9468	0.997

Grouping

Method	Label	셀레늄	Groups
Tukey	연령=A	1448.3636	a
	연령=B	992.6429	b
	연령=C	873.8276	b
	연령=D	851.7292	b

[Post-hoc Analysis]

- 연령그룹에 대해 쌍별 비교분석하는 사후분석 결과가 출력됩니다.
- Tukey's HSD 방법을 적용한 결과, A-B, A-C, A-D에서 유의한 차이를 보이는 것으로 확인됩니다.

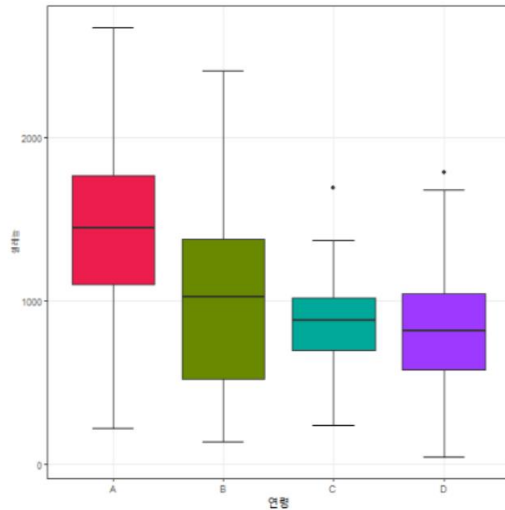
[Grouping]

- 사후분석 결과, A / B,C,D로 그룹화할 수 있음을 알 수 있습니다.

예제 - 결과창

Box Plot

연령



복사 | 저장

Used R Packages

- Main results : `'anova'` of R package `'stats'`, `'Anova'` of R package `'car'`
- Shapiro-Wilk Test : `'shapiro.test'` of R package `'stats'`
- Levene's Test : `'leveneTest'` of R package `'car'`
- Post-hoc Analysis : Tukey's HSD : `'TukeyHSD'` of R package `'stats'`, `'HSD.test'` of R package `'agricolae'`
- Post-hoc Analysis : Scheffe test : `'ScheffeTest'` of R package `'DescTools'`, `'scheffe.test'` of R package `'agricolae'`
- Post-hoc Analysis : Fisher's LSD : `'LSD.test'` of R package `'agricolae'`
- Post-hoc Analysis : Duncan's test : `'Duncan.test'` of R package `'agricolae'`
- Post-hoc Analysis : Student-Newman-Keuls test : `'SNK.test'` of R package `'agricolae'`
- Post-hoc Analysis : Tamhane's T2 : `'tamhaneT2Test'` of R package `'PMCMRplus'`
- Post-hoc Analysis : Dunnett's T3 : `'dunnettT3Test'` of R package `'PMCMRplus'`
- Post-hoc Analysis : Games-Howell test : `'gamesHowellTest'` of R package `'PMCMRplus'`
- Post-hoc Analysis : Dunnett's test : `'DunnettTest'` of R package `'DescTools'`
- Residual : `'residuals'` of R package `'stats'`
- Standard residual : `'rstandard'` of R package `'stats'`
- Studentized residual : `'rstudent'` of R package `'stats'`
- Cook's distance : `'cooks.distance'` of R package `'stats'`
- All results other than those mentioned above were written with basic functions of R.

[Box Plot]

- 연령 그룹에 따른 셀레늄 농도에 대한 박스그림이 출력됩니다. A그룹의 셀레늄 농도가 다른 그룹에 비해 높은 것을 확인할 수 있습니다.

[Used R packages]

- 본 분석에 사용된 R 패키지들이 나열되어 있습니다.