

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

사후 분석 (post hoc)

#01. 사후분석 개요

ANOVA 검증 결과 유의미하다는 결론을 얻었을 때, 구체적으로 어떤 수준(들)에서 평균 차이가 나는지를 검증하는 방법

연구자의 사전 가설(아이디어)없이 ANOVA를 시행한 경우, 탐색적으로 평균 차이가 나는 수준(집단)을 살펴보기 위해 시행하는 방법

조합 가능한 모든 쌍에 대해 비교를 하므로 과잉검증으로 인한 **FWER** 증가

FRWR (가족오류율, Familywise Error Rate)

가설 검정에서 여러 개의 가설을 동시에 비교하는 경우, 각 가설을 독립적으로 검정할 때 발생하는 오류 확률이 존재한다.

FWER는 이러한 가설 중 어느 하나라도 잘못 기각되는 전체적인 오류율을 의미

즉, FWER은 적어도 하나의 거짓 양성(잘못된 기각)이 발생할 확률

가설검정을 많이 할 수록 FWER은 증가

대표적인 사후분석 방법

유의수준을 보정하여 FWER을 0.05로 고정시킴

- 봉페로니 교정 (일반적으로 널리 사용됨)
- 투키의 HSD (일반적으로 널리 사용됨)

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

- 피셔의 LSD : 실제로 보정을 하지 않는 방법이므로 사용하지 않음.
- 셰페의 방법 : 지나치게 보수적이어서 사용하지 않음

#02. 작업 준비

패키지 가져오기

```
from pandas import read_excel

# 분산분석을 위한 라이브러리
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm

# 사후분석을 위한 라이브러리
from statsmodels.sandbox.stats.multicomp import MultiComparison
from statsmodels.stats.multicomp import pairwise_tukeyhsd
from scipy.stats import ttest_ind

# helper 참조
import sys
import os
sys.path.append(os.path.dirname(os.path.dirname(os.getcwd())))
from helper import normality_test, equal_variance_test, independence_test
```

#03. 예제 (1)

품종별 나무 무게 조사 데이터

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E02/tree_weight.xlsx")
df
```

	weight	group
0	4.17	A
1	5.58	A
2	5.18	A
3	6.11	A
4	4.50	A
5	4.61	A
6	5.17	A
7	4.53	A
8	5.33	A
9	5.14	A
10	4.81	B
11	4.17	B
12	4.41	B
13	3.59	B
14	5.87	B

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

	weight	group
15	3.83	B
16	6.03	B
17	4.89	B
18	4.32	B
19	4.69	B
20	6.31	C
21	5.12	C
22	5.54	C
23	5.50	C
24	5.37	C
25	5.29	C
26	4.92	C
27	6.15	C
28	5.80	C
29	5.26	C

데이터 전처리

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

```
df2 = df.copy()
df2['group'] = df2['group'].map({'A': 0, 'B': 1, 'C': 2})
df2
```

	weight	group
0	4.17	0
1	5.58	0
2	5.18	0
3	6.11	0
4	4.50	0
5	4.61	0
6	5.17	0
7	4.53	0
8	5.33	0
9	5.14	0
10	4.81	1
11	4.17	1
12	4.41	1
13	3.59	1
14	5.87	1

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

	weight	group
15	3.83	1
16	6.03	1
17	4.89	1
18	4.32	1
19	4.69	1
20	6.31	2
21	5.12	2
22	5.54	2
23	5.50	2
24	5.37	2
25	5.29	2
26	4.92	2
27	6.15	2
28	5.80	2
29	5.26	2

일원분산분석

정규성, 등분산성, 독립성 검사

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

보통페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

```
all_test(df2['group'], df2['weight'])
```

			statistic	p-value	result
condition	test	field			
정규성	shapiro	group	0.795503	5.434527e-05	False
		weight	0.982683	8.915055e-01	True
	normaltest	group	17.211505	1.830498e-04	False
		weight	0.568335	7.526404e-01	True
	ks_2samp	group vs weight	1.000000	1.691123e-17	False
		weight vs group	1.000000	1.691123e-17	False
등분산성	Bartlett	group vs weight	0.812217	3.674655e-01	True
	Fligner	group vs weight	1.419281	2.335219e-01	True
	Levene	group vs weight	0.849158	3.606080e-01	True
독립성	Chi2	group vs weight	15.905074	9.765907e-01	True

분산분석 수행

```
model = ols('weight ~ C(group)', data=df2).fit()
anova_lm(model)
```

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

	df	sum_sq	mean_sq	F	PR(>F)
C(group)	2.0	3.76634	1.883170	4.846088	0.01591
Residual	27.0	10.49209	0.388596	NaN	NaN

사후분석

봉페로니 교정

- Bonferroni correction
- 모든 집단을 짝지어 t-test
- FWER이 중간 정도

```
comp = MultiComparison(df['weight'], df['group'])
result = comp.allpairtest(ttest_ind, method='bonf')
result[0]
```

Test Multiple Comparison ttest_ind FWER=0.05 method=bonf
alphacSidak=0.02, alphacBonf=0.017

group1	group2	stat	pval	pval_corr	reject
A	B	1.1913	0.249	0.7471	False
A	C	-2.134	0.0469	0.1406	False
B	C	-3.0101	0.0075	0.0226	True

투키의 HSD

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

Tuckey's Honestly Significant Difference = "진정으로 유의미한 차이"

FWER이 중간 정도

```
hsd = pairwise_tukeyhsd(df['weight'], df['group'], alpha=0.05)
hsd.summary()
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	B	-0.371	0.3909	-1.0622	0.3202	False
A	C	0.494	0.198	-0.1972	1.1852	False
B	C	0.865	0.012	0.1738	1.5562	True

결과보고

1. 전체 모형에 대한 분석결과 보고
2. 사용한 사후분석 방법에 대한 보고
3. 유의미한 사후분석 결과들에 대한 보고

group에 따른 weight의 평균 차이는 유의미하였다($F(2, 27) = 4.846, p < 0.05$). Tukey의 HSD를 이용하여 사후분석을 실시한 결과, B 조건과 C 조건에서 유의미한 평균 차이가 있었다($p < 0.05$).

예제 (2)

독극물 실험 데이터

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

필드	설명
Time	동물의 생존시간
poison	사용된 독극물 종류
treat	사용되는 치료 유형

데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E02/poisons.xlsx")
df
```

	time	poison	treat
0	0.31	1	A
1	0.45	1	A
2	0.46	1	A
3	0.43	1	A
4	0.36	2	A
5	0.29	2	A
6	0.40	2	A
7	0.23	2	A
8	0.22	3	A
9	0.21	3	A

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

	time	poison	treat
10	0.18	3	A
11	0.23	3	A
12	0.82	1	B
13	1.10	1	B
14	0.88	1	B
15	0.72	1	B
16	0.92	2	B
17	0.61	2	B
18	0.49	2	B
19	1.24	2	B
20	0.30	3	B
21	0.37	3	B
22	0.38	3	B
23	0.29	3	B
24	0.43	1	C
25	0.45	1	C
26	0.63	1	C
27	0.76	1	C

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

	time	poison	treat
28	0.44	2	C
29	0.35	2	C
30	0.31	2	C
31	0.40	2	C
32	0.23	3	C
33	0.25	3	C
34	0.24	3	C
35	0.22	3	C
36	0.45	1	D
37	0.71	1	D
38	0.66	1	D
39	0.62	1	D
40	0.56	2	D
41	1.02	2	D
42	0.71	2	D
43	0.38	2	D
44	0.30	3	D
45	0.36	3	D

사후 분석 (post hoc)

#01. 사후분석 개요

FRWR (가족오류율, Familywise Error Rate)

대표적인 사후분석 방법

#02. 작업 준비

패키지 가져오기

#03. 예제 (1)

데이터 가져오기

데이터 전처리

일원분산분석

정규성, 등분산성, 독립성 검사

분산분석 수행

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

	time	poison	treat
46	0.31	3	D
47	0.33	3	D