

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

결과보고

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

결과보고

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

데이터 가져오기

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

| | weight | group |
|---|--------|-------|
| 0 | 4.17 | A |
| 1 | 5.58 | A |
| 2 | 5.18 | A |
| 3 | 6.11 | A |
| 4 | 4.50 | A |

데이터 전처리

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

| | weight | group |
|---|--------|-------|
| 0 | 4.17 | 0 |
| 1 | 5.58 | 0 |
| 2 | 5.18 | 0 |

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

| | weight | group |
|---|--------|-------|
| 3 | 6.11 | 0 |
| 4 | 4.50 | 0 |

일원분산분석

정규성, 등분산성, 독립성 검사

```
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 |

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

| | | | statistic | p-value | result |
|-----------|------|-----------------|-----------|--------------|--------|
| condition | test | field | | | |
| 독립성 | Chi2 | group vs weight | 15.905074 | 9.765907e-01 | True |

분산분석 수행

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

| | 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 |

결과보고

p-value가 유의수준 0.05보다 작으므로 품종에 따른 나무에 무게는 통계적으로 유의미한 차이가 있다고 볼 수 있다.

사후분석

봉페로니 교정

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

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

pval_corr 값이 0.05 보다 작은 항목만 통계적으로 유의함

```
# MultiComparison(결과값, 조건값)
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

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 |

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

| | | | | | | |
|---|---|-------|-------|---------|--------|-------|
| 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

| 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")
```

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

```
df.head()
```

| | 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 |

데이터 전처리

```
df2 = df.copy()
df2['treat'] = df2['treat'].map({'A': 1, 'B': 2, 'C': 3, 'D': 4})
df2.head()
```

| | time | poison | treat |
|---|------|--------|-------|
| 0 | 0.31 | 1 | 1 |
| 1 | 0.45 | 1 | 1 |
| 2 | 0.46 | 1 | 1 |
| 3 | 0.43 | 1 | 1 |
| 4 | 0.36 | 2 | 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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

df2.dtypes

```
time      float64
poison     int64
treat      int64
dtype: object
```

이원분산분석

정규성, 등분산성, 독립성 검정은 생략함

```
#model = ols('time ~ C(poison) + C(treat) + C(poison):C(treat)', data=df)
model = ols('time ~ C(poison) * C(treat)', data=df).fit()
anova_lm(model)
```

| | df | sum_sq | mean_sq | F | PR(>F) |
|--------------------|------|----------|----------|-----------|--------------|
| C(poison) | 2.0 | 1.033012 | 0.516506 | 23.221737 | 3.331440e-07 |
| C(treat) | 3.0 | 0.921206 | 0.307069 | 13.805582 | 3.777331e-06 |
| C(poison):C(treat) | 6.0 | 0.250138 | 0.041690 | 1.874333 | 1.122506e-01 |
| Residual | 36.0 | 0.800725 | 0.022242 | NaN | NaN |

결과해석

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

| 요인 | 분석모델 | 결과 |
|--------------|-------------------------------|-----------------------------------|
| poison | $F(2, 36) = 23.222, p < 0.05$ | poison의 수준에 따라 평균에 차이가 난다고 볼 수 있음 |
| treat | $F(3, 36) = 13.806, p < 0.05$ | treat의 수준에 따라 평균에 차이가 난다고 볼 수 있음 |
| poison:treat | $F(6, 36) = 1.874, p > 0.05$ | 유의미하지 않음. 상호작용 효과는 발견하지 못함 |

사후분석

poison에 대한 봉페로니

```
comp = MultiComparison(data=df2['time'], groups=df2['poison'])
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 |
|--------|--------|--------|--------|-----------|--------|
| 1 | 2 | 0.8189 | 0.4193 | 1.0 | False |
| 1 | 3 | 6.2474 | 0.0 | 0.0 | True |
| 2 | 3 | 3.6234 | 0.0011 | 0.0032 | True |

treat에 대한 봉페로니

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

```
comp = MultiComparison(data=df2['time'], groups=df2['treat'])
result = comp.allpairtest(ttest_ind, method='bonf')
result[0]
```

Test Multiple Comparison ttest_ind FWER=0.05 method=bonf
alphacSidak=0.01, alphacBonf=0.008

| group1 | group2 | stat | pval | pval_corr | reject |
|--------|--------|---------|--------|-----------|--------|
| 1 | 2 | -3.7291 | 0.0012 | 0.007 | True |
| 1 | 3 | -1.3855 | 0.1798 | 1.0 | False |
| 1 | 4 | -3.1478 | 0.0047 | 0.028 | True |
| 2 | 3 | 2.7215 | 0.0125 | 0.0748 | False |
| 2 | 4 | 1.27 | 0.2174 | 1.0 | False |
| 3 | 4 | -1.7796 | 0.089 | 0.5338 | False |

```
hsd = pairwise_tukeyhsd(df2['time'], df2['poison'], alpha=0.05)
hsd.summary()
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|--------|--------|----------|--------|---------|---------|--------|
| 1 | 2 | -0.0731 | 0.5882 | -0.2525 | 0.1063 | False |
| 1 | 3 | -0.3412 | 0.0001 | -0.5206 | -0.1619 | True |
| 2 | 3 | -0.2681 | 0.0021 | -0.4475 | -0.0887 | True |

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

결과보고

사후분석

봉페로니 교정

투키의 HSD

결과보고

예제 (2)

데이터 가져오기

```
hsd = pairwise_tukeyhsd(df2['time'], df2['treat'], alpha=0.05)
hsd.summary()
```

Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|--------|--------|----------|--------|---------|---------|--------|
| 1 | 2 | 0.3625 | 0.001 | 0.1253 | 0.5997 | True |
| 1 | 3 | 0.0783 | 0.8143 | -0.1589 | 0.3156 | False |
| 1 | 4 | 0.22 | 0.0778 | -0.0172 | 0.4572 | False |
| 2 | 3 | -0.2842 | 0.0132 | -0.5214 | -0.0469 | True |
| 2 | 4 | -0.1425 | 0.387 | -0.3797 | 0.0947 | False |
| 3 | 4 | 0.1417 | 0.3922 | -0.0956 | 0.3789 | False |