

제3유형_모집단 2개

✓ 검정방법

1. 대응표본(쌍체) : 동일한 객체의 전 vs 후 평균비교

- (정규성O) 대응표본(쌍체) t검정(paired t-test) : 동일한 객체의 전 vs 후 평균비교
- (정규성X) 윌콕슨 부호순위 검정(wilcoxon) ### 2. 독립표본 : A집단의 평균 vs B집단의 평균
- (정규성O) 독립표본 t검정(2sample t-test)
- (정규성X) 윌콕슨의 순위합 검정(ranksums)

✓ 가설검정 순서(중요!!)

1. 대응표본(쌍체) t검정(paired t-test)

1. 가설설정
2. 유의수준 확인
3. 정규성 검정 (주의) 차이값에 대한 정규성
4. 검정 실시(통계량, p-value 확인)
5. 귀무가설 기각여부 결정(채택/기각)

2. 독립표본 t검정(2sample t-test)

1. 가설설정
2. 유의수준 확인
3. 정규성 검정 (주의) 두 집단 모두 정규성을 따를 경우!
4. 등분산 검정
5. 검정 실시(통계량, p-value 확인) (주의) 등분산여부 확인
6. 귀무가설 기각여부 결정(채택/기각)

✓ 예제문제

Case 1) 대응표본(쌍체) t검정(paired t-test)

문제 1-1

다음은 혈압약을 먹은 전,후의 혈압 데이터이다.

혈압약을 먹기 전, 후의 차이가 있는지 쌍체 t 검정을 실시하시오

(유의수준 5%)

- before : 혈압약을 먹기 전 혈압, after : 혈압약을 먹은 후의 혈압
- H0(귀무가설) : after - before = 0
- H1(대립가설) : after - before ≠ 0

```
In [1]: import pandas as pd
import numpy as np
import scipy.stats as stats
from scipy.stats import shapiro
```

```
In [2]: # 데이터 만들기
df = pd.DataFrame( {
    'before': [120, 135, 122, 124, 135, 122, 145, 160, 155, 142, 144, 135, 167],
    'after' : [110, 132, 123, 119, 123, 115, 140, 162, 142, 138, 135, 142, 160] })
print(df.head(3))

   before  after
0     120    110
1     135    132
2     122    123
```

```
In [3]: # 1. 가설설정
# H0 : 약을 먹기전과 먹은 후의 혈압 평균은 같다(효과가 없다)
# H1 : 약을 먹기전과 먹은 후의 혈압 평균은 같지 않다(효과가 있다)
```

In [4]: # 2. 유의수준 확인 : 유의수준 5%로 확인

In [5]: # 3. 정규성 검정 (차이값에 대해 정규성 확인)
statistic, pvalue = stats.shapiro(df['after']-df['before'])
print(round(statistic,4), round(pvalue,4))

0.9589 0.7363

- p-value 값이 유의수준(0.05) 보다 크다.
귀무가설(H0) 채택(정규성검정의 H0 : 정규분포를 따른다)

In [6]: # 4.1 (정규성0) 대응표본(쌍체) t검정(paired t-test)
statistic, pvalue = stats.ttest_rel(df['after'], df['before'], alternative='two-sided') # alternative='two-side
print(round(statistic,4), round(pvalue,4))

-3.1382 0.0086

In [7]: # 4.2 (정규성X) wilcoxon 부호순위 검정
statistic, pvalue = stats.wilcoxon(df['after']-df['before'], alternative='two-sided')
print(round(statistic,4), round(pvalue,4))
alternative (대립가설 H1) 옵션 : 'two-sided', 'greater', 'less'

11.0 0.0134

In [8]: # 5. 귀무가설 기각여부 결정(채택/기각)
p-value 값이 0.05보다 작기 때문에 귀무가설을 기각한다.
즉, 약을 먹기전과 먹은 후의 혈압 평균은 같지 않다(효과가 있다)

답 : 기각

문제 1-2

다음은 혈압약을 먹은 전,후의 혈압 데이터이다.

혈압약을 먹은 후 혈압이 감소했는지 확인하기 위해 쌍체 t 검정을 실시하시오

(유의수준 5%)

- before : 혈압약을 먹기 전 혈압, after : 혈압약을 먹은 후의 혈압
- H0(귀무가설) : after - before >= 0
- H1(대립가설) : after - before < 0

In [9]: # 데이터 만들기
df = pd.DataFrame({
 'before': [120, 135, 122, 124, 135, 122, 145, 160, 155, 142, 144, 135, 167],
 'after' : [110, 132, 123, 119, 123, 115, 140, 162, 142, 138, 135, 142, 160] })
print(df.head(3))

	before	after
0	120	110
1	135	132
2	122	123

In [10]: # 1. 가설설정
H0 : 약을 먹은 후 혈압이 같거나 증가했다. (after - before >= 0)
H1 : 약을 먹은 후 혈압이 감소했다. (after - before < 0)

In [11]: # 2. 유의수준 확인 : 유의수준 5%로 확인

In [12]: # 3. 정규성 검정 (차이값에 대해 정규성 확인)
statistic, pvalue = stats.shapiro(df['after']-df['before'])
print(round(statistic,4), round(pvalue,4))

0.9589 0.7363

In [13]: # 4.1 (정규성0) 대응표본(쌍체) t검정(paired t-test)
statistic, pvalue = stats.ttest_rel(df['after'],df['before'],alternative='less')
print(round(statistic,4), round(pvalue,4))

alternative (대립가설 H1) 옵션 : 'two-sided', 'greater', 'less'

-3.1382 0.0043

In [14]: # 4.2 (정규성X) wilcoxon 부호순위 검정
statistic, pvalue = stats.wilcoxon(df['after']-df['before'], alternative='less')
print(round(statistic,4), round(pvalue,4))

11.0 0.0067

In [15]: # 5. 귀무가설 기각여부 결정(채택/기각)
p-value 값이 0.05보다 작기 때문에 귀무가설을 기각한다.
즉, 약을 먹은 후 혈압이 감소했다고 할 수 있다.

답 : 기각

Case 2) 독립표본 t검정 (Independent t-test)

문제 2-1

다음은 A그룹과 B그룹 인원의 혈압 데이터이다.

두 그룹의 혈압평균이 다르다고 할 수 있는지 독립표본 t검정을 실시하시오.

(유의수준 5%)

- A : A그룹 인원의 혈압, B : B그룹 인원의 혈압
- H0(귀무가설) : $A = B$
- H1(대립가설) : $A \neq B$

```
In [16]: # 데이터 만들기
df = pd.DataFrame( {
    'A': [120, 135, 122, 124, 135, 122, 145, 160, 155, 142, 144, 135, 167],
    'B' : [110, 132, 123, 119, 123, 115, 140, 162, 142, 138, 135, 142, 160] })
print(df.head(3))
```

```
   A    B
0 120  110
1 135  132
2 122  123
```

```
In [17]: # 1. 가설설정
# H0 : A그룹과 B그룹의 혈압 평균은 같다.      (A = B)
# H1 : A그룹과 B그룹의 혈압 평균은 같지 않다. (A ≠ B)
```

```
In [18]: # 2. 유의수준 확인 : 유의수준 5%로 확인
```

```
In [19]: # 3. 정규성 검정
# H0(귀무가설) : 정규분포를 따른다.
# H1(대립가설) : 정규분포를 따르지 않는다.

statisticA, pvalueA = stats.shapiro(df['A'])
statisticB, pvalueB = stats.shapiro(df['B'])
print(round(statisticA,4), round(pvalueA,4))
print(round(statisticB,4), round(pvalueB,4))
```

```
0.9314 0.3559
0.9498 0.5956
```

- p-value 값이 유의수준(0.05) 보다 크다.
귀무가설(H0) 채택
- 만약 하나라도 정규분포를 따르지 않는다면 비모수 검정방법을 써야 함
(윌콕슨의 순위합 검정 ranksums)

```
In [20]: # 4. 등분산성 검정
# H0(귀무가설) : 등분산 한다.
# H1(대립가설) : 등분산 하지 않는다.
statistic, pvalue = stats.bartlett(df['A'], df['B'])
print(round(statistic,4), round(pvalue,4) )
```

```
0.0279 0.8673
```

- p-value 값이 유의수준(0.05) 보다 크다.
귀무가설(H0) 채택 => 등분산성을 따른다고 할 수 있다.

```
In [21]: # 5.1 (정규성0, 등분산성 0/X) t검정
statistic, pvalue = stats.ttest_ind(df['A'], df['B'],
                                     equal_var=True,
                                     alternative='two-sided')
# 만약 등분산 하지 않으면 False로 설정

print(round(statistic,4), round(pvalue,4) )
```

```
0.8192 0.4207
```

```
In [22]: # 5.2 (정규성X) 윌콕슨의 순위합 검정
statistic, pvalue = stats.ranksums(df['A'], df['B'], alternative='two-sided')
print(round(statistic,4), round(pvalue,4) )
```

```
0.8462 0.3975
```

```
In [23]: # 6. 귀무가설 기각여부 결정(채택/기각)
# p-value 값이 0.05보다 크기 때문에 귀무가설을 채택한다
# 즉, A그룹과 B그룹의 혈압 평균은 같다고 할 수 있다.

# 답 : 채택
```

```
In [24]: # (참고) 평균데이터 확인
print(round(df['A'].mean(),4))
```

```
print(round(df['B'].mean(),4))
```

```
138.9231
133.9231
```

문제 2-2

다음은 A그룹과 B그룹 인원의 혈압 데이터이다.

A그룹의 혈압 평균이 B그룹보다 크다고 할 수 있는지 독립표본 t검정을 실시하시오.

(유의수준 5%)

- A : A그룹 인원의 혈압, B : B그룹 인원의 혈압
- H_0 (귀무가설) : $A - B \leq 0$ (or $A \leq B$)
- H_1 (대립가설) : $A - B > 0$ (or $A > B$)

```
In [25]: # 데이터 만들기
df = pd.DataFrame( {
    'A': [120, 135, 122, 124, 135, 122, 145, 160, 155, 142, 144, 135, 167],
    'B' : [110, 132, 123, 119, 123, 115, 140, 162, 142, 138, 135, 142, 160] })
print(df.head(3))
```

```
      A      B
0  120   110
1  135   132
2  122   123
```

```
In [26]: # 1. 가설설정
# H0 : A그룹의 혈압 평균이 B그룹보다 작거나 같다. (A <= B)
# H1 : A그룹의 혈압 평균이 B그룹보다 크다.      (A > B)
```

```
In [27]: # 2. 유의수준 확인 : 유의수준 5%로 확인
```

```
In [28]: # 3. 정규성 검정 (차이값에 대해 정규성 확인)
# H0(귀무가설) : 정규분포를 따른다.
# H1(대립가설) : 정규분포를 따르지 않는다.

statisticA, pvalueA = stats.shapiro(df['A'])
statisticB, pvalueB = stats.shapiro(df['B'])
print(round(statisticA,4), round(pvalueA,4))
print(round(statisticB,4), round(pvalueB,4))

0.9314 0.3559
0.9498 0.5956
```

```
In [29]: # 4. 등분산성 검정
# H0(귀무가설) : 등분산 한다.
# H1(대립가설) : 등분산 하지 않는다.
statistic, pvalue = stats.bartlett(df['A'], df['B'])
print(round(statistic,4), round(pvalue,4) )

0.0279 0.8673
```

```
In [30]: # 5.1 (정규성0, 등분산성 0/X) t검정
statistic, pvalue = stats.ttest_ind(df['A'], df['B'],
                                     equal_var=False,
                                     alternative='greater')
# 만약 등분산 하지 않으면 False로 설정

print(round(statistic,4), round(pvalue,4) )

0.8192 0.2104
```

```
In [31]: # 5.2 (정규성X) 윌콕슨의 순위합 검정
statistic, pvalue = stats.ranksums(df['A'], df['B'], alternative='greater')
print(round(statistic,4), round(pvalue,4) )

0.8462 0.1987
```

```
In [32]: # 6. 귀무가설 기각여부 결정(채택/기각)
# p-value 값이 0.05보다 크기 때문에 귀무가설을 채택한다
# 즉, A그룹의 혈압 평균이 B그룹보다 작거나 같다고 할 수 있다.
# (A그룹의 혈압 평균이 B그룹보다 크다고 할 수 없다)

# 답 : 채택
```