

제3유형_모집단 1개(집단의 평균 vs 특정값)

✓ 검정방법

1) (정규성O) 단일표본 t검정(1sample t-test)

2) (정규성X) 윌콕슨 부호순위 검정

✓ 가설검정 순서(중요!!)

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

✓ 데이터 불러오기

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: # 데이터 불러오기 mtcars
df = pd.read_csv("mtcars.csv")
df.head(3)
```

```
Out[2]:
```

	car	mpg	cyl	displacement	horsepower	drat	weight	qsec	vs	am	gear	carb
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1

예제문제

1. mtcars 데이터셋의 mpg 열 데이터의 평균이 20과 같다고 할 수 있는지 검정하십시오. (유의수준 5%)

```
In [3]: import scipy.stats as stats
from scipy.stats import shapiro
```

```
In [4]: # 1. 가설설정
# H0 : mpg 열의 평균이 20과 같다.
# H1 : mpg 열의 평균이 20과 같지 않다.
```

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

```
In [6]: # 3. 정규성 검정
# H0(귀무가설) : 정규분포를 따른다.
# H1(대립가설) : 정규분포를 따르지 않는다.
statistic, pvalue = stats.shapiro(df['mpg'])
print(round(statistic,4), round(pvalue,4))

result = stats.shapiro(df['mpg'])
print(result)

0.9476 0.1229
ShapiroResult(statistic=0.9475648403167725, pvalue=0.1228824257850647)
```

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

```
In [7]: # 4.1 (정규성만족 O) t-검정 실시
statistic, pvalue = stats.ttest_1samp(df['mpg'], popmean=20, alternative='two-sided') # H1 : 왼쪽값이 오른쪽 값과
print(round(statistic,4), round(pvalue,4))
# alternative (대립가설 H1) 옵션 : 'two-sided', 'greater', 'less'

0.0851 0.9328
```

```
In [8]: # 4.2 (정규성만족 X) wilcoxon 부호순위 검정
statistic, pvalue = stats.wilcoxon(df['mpg']-20, alternative='two-sided')
print(round(statistic,4), round(pvalue,4) )

249.0 0.7891
```

```
In [9]: # 5. 귀무가설 기각여부 결정(채택/ 기각)
# p-value 값이 0.05보다 크기 때문에(0.9328) 귀무가설을 채택한다
# 즉, mpg 열의 평균이 20과 같다고 할 수 있다.

# 답 : 채택
```

```
In [10]: # 실제로 평균을 구해보면
df['mpg'].mean()
```

```
Out[10]: 20.090625000000003
```

2. mtcars 데이터셋의 mpg열 데이터의 평균이 17보다 크다고 할 수 있는지 검정하시오.(유의수준 5%)

```
In [11]: # 1. 가설설정
# H0 : mpg 열의 평균이 17보다 작거나 같다(mpg mean <= 17)
# H1 : mpg 열의 평균이 17보다 크다(mpg mean > 17)
```

```
In [12]: # 2. 유의수준 확인 : 유의수준 5%로 확인
```

```
In [13]: # 3. 정규성 검정
statistic, pvalue = stats.shapiro(df['mpg'])
print(round(statistic,4), round(pvalue,4) )

0.9476 0.1229
```

```
In [14]: # 4.1 (정규성만족 0) t-검정 실시
statistic, pvalue = stats.ttest_1samp(df['mpg'], popmean= 17, alternative='greater') # H1 : 왼쪽값이 오른쪽 값보다
print(round(statistic,4), round(pvalue,4) )

2.9008 0.0034
```

```
In [15]: # 4.2 (정규성만족 X) wilcoxon 부호순위 검정
statistic, pvalue = stats.wilcoxon(df['mpg']-17, alternative='greater')
print(round(statistic,4), round(pvalue,4) )

395.5 0.0066
```

```
In [16]: # 5. 귀무가설 기각여부 결정(채택/ 기각)
# p-value 값이 0.05보다 작기 때문에 귀무가설을 기각한다(대립가설 채택)
# 즉, mpg 열의 평균이 17보다 크다고 할 수 있다.

# 답 : 기각
```

3. mtcars 데이터셋의 mpg열 데이터의 평균이 17보다 작다고 할 수 있는지 검정하시오.(유의수준 5%)

```
In [17]: # 1. 가설설정
# H0 : mpg 열의 평균이 17보다 크거나 같다(mpg mean >= 17)
# H1 : mpg 열의 평균이 17보다 작다(mpg mean < 17)
```

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

```
In [19]: # 3. 정규성 검정
statistic, pvalue = stats.shapiro(df['mpg'])
print(round(statistic,4), round(pvalue,4) )

0.9476 0.1229
```

```
In [20]: # 4.1 (정규성만족 0) t-검정 실시
statistic, pvalue = stats.ttest_1samp(df['mpg'], popmean= 17, alternative='less') # H1 : 왼쪽값이 오른쪽 값보다 작
print(round(statistic,4), round(pvalue,4) )

2.9008 0.9966
```

```
In [21]: # 4.2 (정규성만족 X) wilcoxon 부호순위 검정
statistic, pvalue = stats.wilcoxon(df['mpg']-17, alternative='less')
print(round(statistic,4), round(pvalue,4) )

395.5 0.9938
```

```
In [22]: # 5. 귀무가설 기각여부 결정(채택/ 기각)
# p-value 값이 0.05보다 크기 때문에 귀무가설을 채택한다.
# 즉, mpg 열의 평균이 17보다 작다고 할 수 없다

# 답 : 채택
```