

## 제3유형\_모집단 3개 이상

### ✓ 검정방법

#### 1. 분산분석(ANOVA) : A집단 vs B집단 vs C집단 vs ....

- (정규성O) ANOVA분석
- (정규성X) 크루스칼-왈리스 검정(kruskal-wallis test)

### ✓ 가설검정 순서(중요!!)

1. 가설설정
2. 유의수준 확인
3. 정규성 검정 (주의) 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

#### 문제 1-1

다음은 A, B, C 그룹 인원 성적 데이터이다.

세 그룹의 성적 평균이 같다고 할 수 있는지 ANOVA 분석을 실시하시오.

(유의수준 5%)

- A, B, C : 각 그룹 인원의 성적
- $H_0$ (귀무가설) :  $A(\text{평균}) = B(\text{평균}) = C(\text{평균})$
- $H_1$ (대립가설) : Not  $H_0$  (적어도 하나는 같지 않다)

```
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( {
    '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],
    'C': [130, 120, 115, 122, 133, 144, 122, 120, 110, 134, 125, 122, 122]} )
print(df.head(3))
```

	A	B	C
0	120	110	130
1	135	132	120
2	122	123	115

```
In [3]: # 1. 가설설정
# H0 : 세 그룹 성적의 평균값이 같다. ( A(평균) = B(평균) = C(평균) )
# H1 : 세 그룹의 성적 평균값이 적어도 하나는 같지 않다. (not H0)
```

```
In [4]: # 2. 유의수준 확인 : 유의수준 5%로 확인
```

```
In [5]: # 3. 정규성 검정
print(stats.shapiro(df['A']) )
print(stats.shapiro(df['B']) )
print(stats.shapiro(df['C']) )

# statistic, pvalue = stats.shapiro(df['A'])
# print(round(statistic,4), round(pvalue,4))
```

```
ShapiroResult(statistic=0.9314376711845398, pvalue=0.35585272312164307)
ShapiroResult(statistic=0.9498201012611389, pvalue=0.5955665707588196)
ShapiroResult(statistic=0.9396706223487854, pvalue=0.45265132188796997)
```

- 세 집단 모두 p-value 값이 유의수준(0.05) 보다 크다.  
귀무가설( $H_0$ ) 채택 => 정규분포를 따른다고 할 수 있다.
- 만약 하나라도 정규분포를 따르지 않는다면 비모수 검정방법을 써야 함

```
In [6]: # 4. 등분산성 검정
# H0(귀무가설) : 등분산 한다.
# H1(대립가설) : 등분산 하지 않는다.
print(stats.bartlett(df['A'], df['B'], df['C']))
```

BartlettResult(statistic=4.222248448848066, pvalue=0.12110174433684852)

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

```
In [7]: # 5.1 (정규성0, 등분산성 0) 분산분석(F_oneway)
import scipy.stats as stats
statistic, pvalue = stats.f_oneway(df['A'], df['B'], df['C'])
# 주의 : 데이터가 각각 들어가야 함(밑에 예제와 비교해볼 것)

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

3.6971 0.0346

```
In [8]: # 5.2 (정규성0, 등분산성 X) Welch-ANOVA 분석
# import pingouin as pg # pingouin 패키지 미지원
# pg.welch_anova(dv = "그룹변수명", between = "성적데이터", data = 데이터)
# pg.welch_anova(df['A'], df['B'], df['C']) 형태로 분석불가
```

```
In [9]: # 5.3 (정규성X) 크루스칼 왈리스 검정
import scipy.stats as stats
statistic, pvalue = stats.kruskal(df['A'], df['B'], df['C'])
print(round(statistic,4), round(pvalue,4))
```

6.897 0.0318

```
In [10]: # 6. 귀무가설 기각여부 결정(채택/ 기각)
# p-value 값이 0.05보다 작기 때문에 귀무가설을 기각한다. (대립가설채택)
# 즉, A,B,C 그룹의 성적 평균이 같다고 할 수 없다.

# 답 : 기각
```

## 문제 1-2 데이터 형태가 다를 경우

```
In [11]: # 데이터 만들기
df2 = pd.DataFrame( {
    '항목': ['A','A','A','A','A','A','A','A','A','A','A','A','A','A','A',
            'B','B','B','B','B','B','B','B','B','B','B','B','B','B','B',
            'C','C','C','C','C','C','C','C','C','C','C','C','C','C','C'],
    'value': [120, 135, 122, 124, 135, 122, 145, 160, 155, 142, 144, 135, 167,
              110, 132, 123, 119, 123, 115, 140, 162, 142, 138, 135, 142, 160,
              130, 120, 115, 122, 133, 144, 122, 120, 110, 134, 125, 122, 122]
})
print(df2.head(3))
```

```
항목  value
0   A    120
1   A    135
2   A    122
```

```
In [12]: # 각각 필터링해서 변수명에 저장하고 분석 진행
a = df2[ df2['항목']=='A' ]['value']
b = df2[ df2['항목']=='B' ]['value']
c = df2[ df2['항목']=='C' ]['value']
```

```
In [13]: # 분산분석(F_oneway)
import scipy.stats as stats
statistic, pvalue = stats.f_oneway(a, b, c)
print(round(statistic,4), round(pvalue,4))
```

3.6971 0.0346