

Hypothesis Test

```
import warnings
warnings.filterwarnings('ignore')
```

I. t-Test

- 일반적으로 모집단의 모평균이나 모분산은 알 수 없음
- 표본표준편차로 표본집단을 표준화를 하면, 표준정규분포가 아닌 t-분포를 따름
- 표본표준편차로 변환한 t-분포를 사용하기 때문에 t-Test라고 함

1) 단일집단 t-Test

(1) Hypothesis

- 귀무가설 : 스낵의 평균무게는 50이다.
- 대립가설 : 스낵의 평균무게는 50이 아니다.

(2) 단일표본

```
import pandas as pd

url = 'https://raw.githubusercontent.com/rusita-ai/pyData/master/foodWeight.csv'
DF = pd.read_csv(url)

DF.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 1 columns):
 #   Column  Non-Null Count  Dtype  
---  -
 0   weight  20 non-null      float64
dtypes: float64(1)
memory usage: 288.0 bytes
```

```
DF.head()
```

```

weight
0  58.529820
1  52.353039
2  74.446169
3  53.003369

```

(3) t-Test

- 단이표본 t-Test
- p값이 0.05보다 작기 때문에, '스낵의 평균무게는 50이다'는 귀무가설을 '기각'
 - 집단의 모평균의 차이는 통계적으로 유의미함

```
import scipy.stats
```

```
scipy.stats.ttest_1samp(DF, 50)
```

```
Ttest_1sampResult(statistic=array([2.75033968]), pvalue=array([0.01272559]))
```

2) 두 집단 독립표본 t-Test

- 두 표본집단을 구성(수집)할 때 독립인 경우
 - 독립표본 : 무작위로 남자 100명과 여자 100명을 뽑아 두 집단을 비교
 - 대응표본 : 부부 100쌍을 뽑아 남편 100명과 아내 100명의 두 집단을 비교
- 논리
 - 모평균과 표본평균은 다를 수 있지만 차이가 크지 않을 것
 - 두 표본집단의 평균의 차이는 0이 아니지만, 큰 차이는 보이지 않음
 - 만약, 두 표본평균의 차이가 크다면, 귀무가설이 맞지 않을 수 있음
 - 따라서, 두 표본평균의 차이나는 정도에 대한 가능성(확률)을 계산하여 판단

(1) Hypothesis

- 귀무가설 : 두 집단의 모평균이 동일하다.
- 대립가설 : 두 집단의 모평균은 동일하지 않다.

(2) 독립표본

- Male과 Female 표본집단이 무작위로 추출

```
Male = [117, 108, 105, 89, 101,
        93, 96, 108, 108, 94]
```

```
93, 90, 100, 100, 94,
93, 112, 92, 91, 100,
96, 120, 86, 96, 95]
```

```
Female = [121, 101, 102, 114, 103,
105, 101, 131, 96, 109,
109, 113, 115, 94, 108,
96, 110, 112, 120, 100]
```

(3) Welch's t_Test

- 두 표본분산의 등분산성 검사 없이 진행
- 독립표본 t-Test에 적용(Independent)
 - 'equal_var = False'
- p값이 0.05보다 작기 때문에, '두 집단의 모평균이 동일하다'는 귀무가설을 '기각'
 - 두 집단의 모평균의 차이는 통계적으로 유의미함

```
import scipy.stats
```

```
scipy.stats.ttest_ind(Male, Female, equal_var = False)
```

```
Ttest_indResult(statistic=-2.670573872669349, pvalue=0.01108318824471652)
```

(4) Effect Size

- 두 집단의 평균 차이를 일정한 기준으로 표현
- 절대적이지 않으며, 0.2 정도면 작은 편, 0.5 정도면 중간, 0.8 이면 큰 편

```
DT = scipy.stats.ttest_ind(Male, Female, equal_var = False)
```

- Cohen's d

```
import numpy as np
```

```
t = DT.statistic
```

```
df = len(Male) + len(Female) - 2
```

```
abs(t) / np.sqrt(df)
```

```
0.4332242888591059
```

- Pearson's r

```
t2 = t ** 2
```

```
np.sqrt(t2 / (t2 + df))
```

```
0.39752319599996255
```

(5) 검증결과 해석

- 두 집단 Male과 Female에 대하여 독립표본 t 검증을 실시한 결과,
 - 집단 Male의 평균값(100)은 집단 Female의 평균값(108)보다 통계적으로 유의미하게 낮았으며($p < 0.05$),
 - 효과 크기는 중간 수준이었다(Cohen's $d = 0.43$, Pearson's $r = 0.39$).

3) 두 집단 대응표본 t-Test

- 두 집단의 자료를 쌍으로 묶을 수 있는 경우
 - 예) 남편과 아내, Before vs. After
- 쌍을 이루고 있는 두 값의 차이를 구함

(1) Hypothesis

- 귀무가설 : 두 집단의 모평균이 동일하다.
- 대립가설 : 두 집단의 모평균은 동일하지 않다.

(2) 대응표본

- 두 표본집합 Before와 After가 순서대로 짝지어 있음

```
Before = [117, 108, 105, 89, 101,
          93, 96, 108, 108, 94,
          93, 112, 92, 91, 100,
          96, 120, 86, 96, 95]
```

```
After = [121, 101, 102, 114, 103,
         105, 101, 131, 96, 109,
         109, 113, 115, 94, 108,
         96, 110, 112, 120, 100]
```

(3) t-Test

- 대응표본 t-Test에 적용(Related)
- p값이 0.05보다 작기 때문에, '두 집단의 모평균이 동일하다'는 귀무가설을 '기각'
 - 두 집단의 모평균의 차이는 통계적으로 유의미함

```
import scipy.stats

scipy.stats.ttest_rel(Before, After)

Ttest_relResult(statistic=-2.9868874599588247, pvalue=0.007578486289181322)
```

(4) Effect Size

- 두 집단의 평균 차이를 일정한 기준으로 표현
- 절대적이지 않으며, 0.2 정도면 작은 편, 0.5 정도면 중간, 0.8 이면 큰 편

```
DT = scipy.stats.ttest_rel(Before, After)
```

- Cohen's d

```
t = DT.statistic

df = len(Before) + len(After) - 2

abs(t) / np.sqrt(df)

0.4845371285121742
```

- Pearson's r

```
t2 = t ** 2

np.sqrt(t2 / (t2 + df))

0.43604662677848566
```

(5) 검증결과 해석

- 두 집단 Before와 After에 대하여 대응표본 t 검증을 실시한 결과,
 - 집단 Before의 평균값(100)은 집단 After의 평균값(108)보다 통계적으로 유의미하게 낮았으며($p < 0.05$),
 - 효과 크기는 중간 수준이었다(Cohen's $d = 0.48$, Pearson's $r = 0.43$).

II. ANOVA Test

- ANalysis Of VAriance(ANOVA)

- 세 개 이상의 집단에 대한 평균을 비교할 때 사용
 - 기존 t-Test 사용 시 과잉검증(Overtesting) 문제 발생
 - t-Test로 두 집단씩 짝지어 분석할 경우 분석횟수가 기하급수적으로 증가
 - 통계적 검증절차를 남용하여, 확률적 의사결정에서 발생 가능한 오류의 확률이 필요 이상으로 증가하는 문제
- 오류의 확률을 통제된 상태에서 전체적인 결과를 확인 가능
- 집단을 구성하는 변수(요인/Factor)가 두 개 이상인 경우, 상호작용 파악에 용이
- 용어
 - 요인(Factor) : 집단을 구별하는 독립변수
 - 수준(Level) : 요인의 수준(예: '성별'요인의 수준은 '남자'와 '여자')
 - 상호작용 : 한 요인의 수준에 따른 종속변수의 차이가 또 다른 요인의 수준에 따라 달라질 때
 - n-way ANOVA
 - One-way ANOVA : 요인이 1개인 분산분석
 - Two-way ANOVA : 요인이 2개인 분산분석

1) One-way ANOVA

(1) Hypothesis

- 귀무가설 : 모든 집단의 평균은 동일하다
- 대립가설 : 적어도 한 집단의 평균은 다른 집단과 다르다

(2) Load Data

```
import pandas as pd

url = 'https://raw.githubusercontent.com/rusita-ai/pyData/master/PlantGrowth.csv'
DF = pd.read_csv(url)
```

```
DF.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  -
 0  weight  30 non-null      float64
 1  group   30 non-null      object  
dtypes: float64(1), object(1)
memory usage: 608.0+ bytes
```

```
DF.head()
```

	weight	group
0	4.17	ctrl
1	5.58	ctrl
2	5.18	ctrl
3	6.11	ctrl
4	4.50	ctrl

- 3개의 집단 확인
 - 'ctrl', 'trt1', 'trt2'

```
DF['group'].unique()
```

```
array(['ctrl', 'trt1', 'trt2'], dtype=object)
```

(3) 일원분산분석

- $Pr(>F)$ 가 p-value
- p-value 값이 0.05보다 작으면, 통계적으로 유의미한 차이가 존재
- 검증결과 0.0159로 0.05보다 작아서, 귀무가설 '기각'
- 구체적으로 어떤 수준(집단)이 차이가 있는지 확인하려면 사후분석(post hoc test) 필요
- 유의미한 차이가 없는 경우, 사후분석 필요 없음

```
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm

model = ols('weight ~ C(group)', DF).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

(4) 일원분산분석 가정

- 독립성 : 자료의 추출은 독립적으로 이루어졌음
- 정규성 : 자료의 모집단 분포는 정규분포를 따름
- 등분산성 : 모든 집단의 모분산은 동일함

4-1. 독립성

- 자료 수집이 Random Sampling 되었다면 만족하는 것으로 봄

4-2 정규성

- 세 수준 모두 p-value값이 0.05보다 큼
- 모집단의 분포가 정규분포를 따른다는 귀무가설을 '채택'
- 'ctrl' 수준의 정규성

```
import scipy.stats

scipy.stats.shapiro(DF.weight[DF.group == 'ctrl'])

(0.9566815495491028, 0.7474744915962219)
```

- 'trt1' 수준의 정규성

```
scipy.stats.shapiro(DF.weight[DF.group == 'trt1'])

(0.9304108619689941, 0.451945960521698)
```

- 'trt2' 수준의 정규성

```
scipy.stats.shapiro(DF.weight[DF.group == 'trt2'])

(0.9410051107406616, 0.5642509460449219)
```

4-3 등분산성

- 두 가지 테스트 모두 p-value값이 0.05보다 큼
- 모집단의 모분산은 동일하다는 귀무가설을 '채택'
- Levene(레빈 검증)

```
scipy.stats.levene(
    DF.weight[DF.group == 'ctrl'],
    DF.weight[DF.group == 'trt1'],
    DF.weight[DF.group == 'trt2'] )

LeveneResult(statistic=1.1191856948703909, pvalue=0.3412266241254737)
```

- Bartlett(바틀렛 검증)


```

scipy.stats.bartlett(
    DF.weight[DF.group == 'ctrl'],
    DF.weight[DF.group == 'trt1'],
    DF.weight[DF.group == 'trt2'] )

```

BartlettResult(statistic=2.8785737872360935, pvalue=0.23709677363455822)

(5) 일원분산분석 사후분석

- post hoc test
 - 유의미한 검증결과 도출 시, 어떤 수준(들)에서 평균 차이가 나는지를 검증
 - 연구자의 사전 가설(아이디어)없이 ANOVA를 시행한 경우, 탐색적으로 평균 차이가 나는 수준(집단)을 살펴보기 위해 시행
 - 조합 가능한 모든 쌍에 대해 비교하여, 과잉검증으로 인한 FWER 증가
- FWER
 - Family Wise Error Rate
 - 여러 개의 가설 검정을 할 때 적어도 하나의 가설에서 1종 오류가 발생할 가능성
 - 가설검정을 많이 할 수록 FWER은 증가
 - 유의수준을 보정하여 FWER을 0.05로 고정 후 검증

5-1. 사후분석 준비

```

from statsmodels.sandbox.stats.multicomp import MultiComparison
import scipy.stats

```

```
MC = MultiComparison(DF['weight'], DF['group'])
```

5-2. Bonferroni Correction

- 봉페로니 교정
 - 모든 집단을 짝지어 t-Test
- 'trt1'과 'trt2' 수준 간의 평균차이만 유의미함

```

result = MC.allpairtest(scipy.stats.ttest_ind, method = 'bonf')

result[0]

```

Test Multiple Comparison ttest_ind FWER=0.05
method=bonf alphasidak=0.02.

5-3. Tuckey's HSD

- 투키의 HSD(Honestly Significant Difference)
- 'trt1'과 'trt2' 수준 간의 평균차이만 유의미함

```
from statsmodels.stats.multicomp import pairwise_tukeyhsd

THSD = pairwise_tukeyhsd(DF['weight'], DF['group'], alpha = 0.05)

THSD.summary()
```

Multiple Comparison of Means - Tukey HSD,
FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
ctrl	trt1	-0.371	0.3921	-1.0621	0.3201	False
ctrl	trt2	0.494	0.198	-0.1971	1.1851	False
trt1	trt2	0.865	0.012	0.1739	1.5561	True

(6) 일원분산분석 결과 해석

- group에 따른 weight의 평균 차이는 유의미함($p < 0.05$)
 - 사후분석을 실시 결과, 'trt1' 수준과 'trt2' 수준에서 유의미한 평균 차이가 있었음($p < 0.05$)

2) Two-way ANOVA

- Two-way ANOVA : 요인이 2개인 분산분석
- 요인 간 상호작용 파악이 주요 목적
- 다원분산분석 가정
 - 독립성
 - 정규성
 - 등분산성
- 주요 용어
 - 주효과 : 다른 요인(집단구분 변수)과 상관없이, 한 요인의 수준(집단)에 따라 효과가 유의미하게 달라질 때 '주효과가 있다'고 함
 - 상호작용효과 : 한 요인의 수준에 따른 효과차이가 또 다른 요인의 수준에 따라 달라질 때, '요인들 간 상호작용이 존재한다'고 함

(1) Hypothesis

- 귀무가설 : 모든 집단의 평균은 동일하다
- 대립가설 : 적어도 한 집단의 평균은 다른 집단과 다르다

(2) Load Data

```
import pandas as pd
```

```
url = 'https://raw.githubusercontent.com/rusita-ai/pyData/master/poisons.csv'
DF = pd.read_csv(url)
```

```
DF.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48 entries, 0 to 47
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ------  -
0   Unnamed: 0    48 non-null    int64
1   time         48 non-null    float64
2   poison       48 non-null    int64
3   treat       48 non-null    object
dtypes: float64(1), int64(2), object(1)
memory usage: 1.6+ KB
```

```
DF.head()
```

	Unnamed: 0	time	poison	treat
0	1	0.31	1	A
1	2	0.45	1	A
2	3	0.46	1	A
3	4	0.43	1	A
4	5	0.36	2	A

(3) 균형설계

- 각 집단과 조건별로 표본의 수가 동일한 경우
- 비균형설계의 경우 계산 방법이 달라짐

3-1. 독립변수 별 균형설계 확인

- 'poison': 집단별 표본수 16으로 동일

```
DF.groupby('poison').agg(len)
```

	Unnamed: 0	time	treat
poison			
1	16	16.0	16
2	16	16.0	16
3	16	16.0	16

- 'treat': 집단별 표본수 12로 동일

```
DF.groupby('treat').agg(len)
```

	Unnamed: 0	time	poison
treat			
A	12	12.0	12
B	12	12.0	12
C	12	12.0	12
D	12	12.0	12

- 'poison', 'treat': 집단별 표본수 4로 동일

```
DF.groupby(['poison', 'treat']).agg(len)
```

		Unnamed: 0	time
poison	treat		
1	A	4	4.0
	B	4	4.0
	C	4	4.0
	D	4	4.0
2	A	4	4.0
	B	4	4.0
	C	4	4.0
	D	4	4.0
3	A	4	4.0
	B	4	4.0
	C	4	4.0
	D	4	4.0

(4) 이원분산분석

- 'poison': $p < 0.05$ 로 유의미. 즉 poison의 수준에 따라 평균에 차이가 난다고 볼 수 있음
- 'treat': $p < 0.05$ 로 유의미. 즉 treat의 수준에 따라 평균에 차이가 난다고 볼 수 있음
- 'poison:treat': $p > 0.05$ 로 유의미하지 않음. 상호작용 효과는 발견하지 못함

```
from statsmodels.formula.api import ols
from statsmodels.stats.anova import anova_lm

TWA = ols('time ~ C(poison) * C(treat)', DF).fit()

anova_lm(TWA)
```

	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	0.250137	0.041690	1.874333	1.122506e-01
Residual	36.0	0.800725	0.022242	NaN	NaN

(5) 비균형설계

- `anova_lm(TWA, typ = 3)`
 - 'typ = 3' 옵션 추가

(6) 이원분산분석 결과 해석

- 'time'에 대한 'poison'과 'treat'을 요인으로 하는 이원분산분석 실시 결과,
 - 'poison'의 주효과 유의($p < 0.05$),
 - 'treat'의 주효과 유의($p < 0.05$),
 - 'poison'과 'treat'의 유의미한 상호작용효과는 발견할 수 없음($p > 0.05$).
- 만약 상호작용효과가 유의하다면,
 - 'poison' 1, 2, 3 집단별 'treat'의 단순효과분석 수행

III. Chi-Squared Test

- 독립성 검정
- 관찰된 빈도가 기대된 빈도와 의미있게 다른지 여부를 검증
- 명목척도 자료의 분석에 사용

(1) Hypothesis

- 귀무가설 : '가사노동의 종류'(행)와 '수행하는 사람'(열)은 독립이다.
- 대립가설 : '가사노동의 종류'(행)와 '수행하는 사람'(열)은 독립이 아니다.

(2) Load Data

```
url = 'https://raw.githubusercontent.com/rusita-ai/pyData/master/housetasks.csv'
DF = pd.read_csv(url,
                  index_col = 0)
```

```
DF.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 13 entries, Laundry to Holidays
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Wife             13 non-null    int64
1   Alternating      13 non-null    int64
2   Husband          13 non-null    int64
3   Jointly          13 non-null    int64
dtypes: int64(4)
memory usage: 520.0+ bytes
```

- 행 : 가사노동의 종류
- 열 : 수행하는 사람

```
DF.head()
```

	Wife	Alternating	Husband	Jointly
Laundry	156	14	2	4
Main_meal	124	20	5	4
Dinner	77	11	7	13
Breakfast	82	36	15	7
Tidying	53	11	1	57

(3) 카이제곱검증

```
import scipy.stats
```

```
chi2, pvalue, dof, expected = scipy.stats.chi2_contingency(DF)
```

3-1. chi-square 검정통계량

```
chi2
```

```
1944.4561959955277
```

3-2. 자유도

```
dof
```

```
36
```

3-3. p-value

```
pvalue
```

```
0.0
```

(4) 카이제곱검증 결과 해석

- $p < 0.05$ 이므로, '가사노동의 종류'(행)와 '수행하는 사람'(열)은 독립이 아니다.

```
#
```

```
#
```

```
#
```

The End

```
#
```

```
#
```

```
#
```

