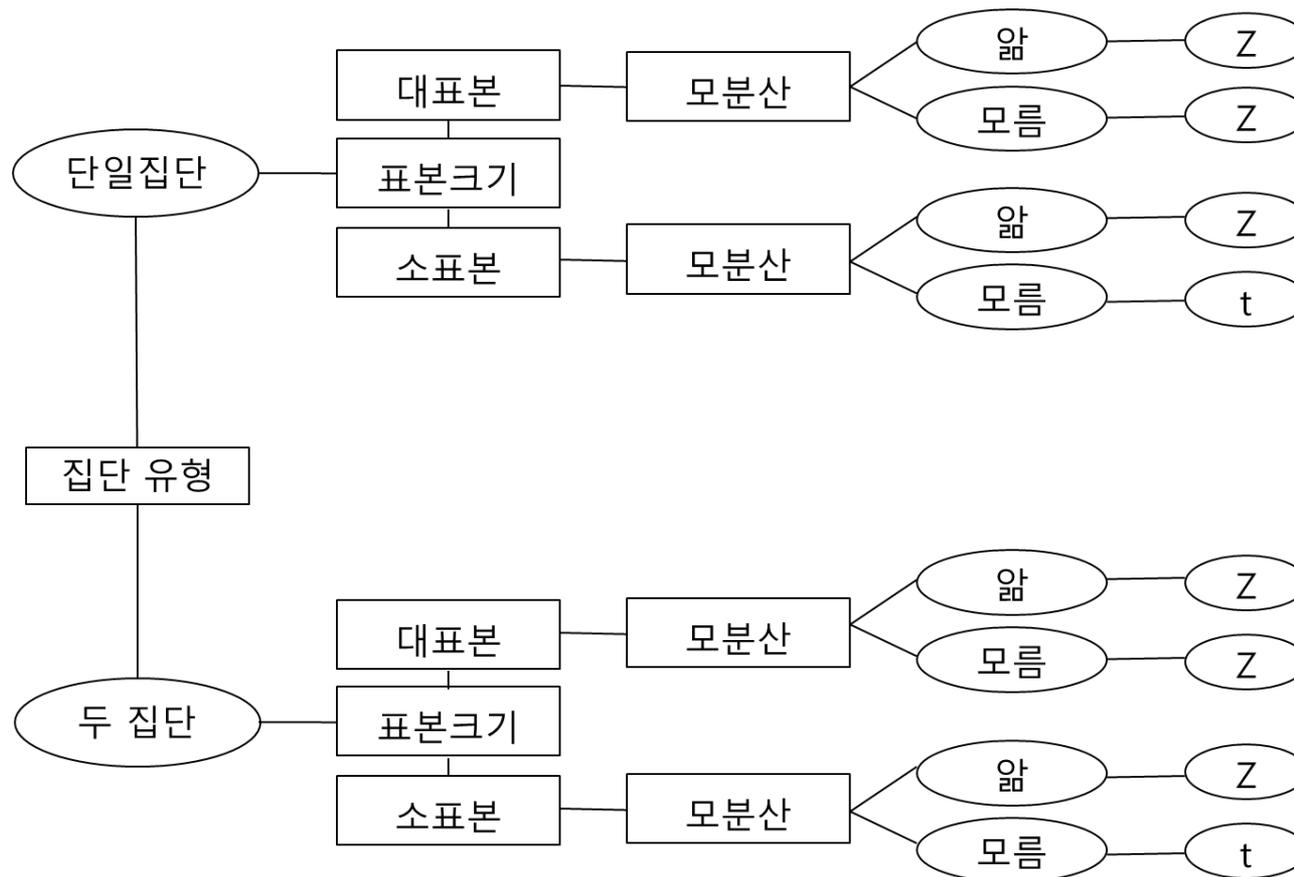


I. 단일집단 가설검정

II. 두 집단 가설검정

### 1. 의사결정 트리

- 모평균에 대한 가설검정은 단일집단이든 두 집단이든, 모분산을 알든 모르든 관계없이
  - 표본의 크기가 30 이상의 대표본이면 Z-통계량을 이용
  - 표본의 크기가 30 미만의 소표본이면 t-통계량을 이용



## 2. 모평균에 대한 가설검정

- (연습 1) 어떤 아이스크림 회사의 영업부 사원은 체인점의 여름 판매량보다 겨울 판매량이 평균 34.5% 감소한다고 한다. 전국 체인점 중 15개를 표본추출하여 판매 감소량을 조사해 보니 다음과 같이 나타났다. 모집단이 정규분포한다고 가정하고 판매량의 감소가 평균 34.5%라는 귀무가설을 5% 유의수준에서 양측검정하라.

33.46	33.38	32.73	32.15	33.99	34.10	33.97	34.34
22.95	33.85	34.23	34.05	34.13	34.45	34.19	

```

b1-ch7-2.py

import pandas as pd
import numpy as np
import scipy.stats as stats
from scipy.stats.distributions import t as tdist
data1 = pd.read_stata("http://kanggc.ptime.org/book/data/chap11-2-1.dta")
data1.head()
var1 = data1['var1']
var1.describe()
print("Data Frame of data1 is : ",f"\n{data1}\n")
print("Summary Statistics of data1 is : ", f"\n{data1.describe()}\n")
xbar = np.mean(var1)
xbar
xstd = np.std(var1)
xstd
t_cal = (xbar-34.5)/(xstd/np.sqrt(14))
t_cal
t14_1 = tdist.ppf(0.025, df=14)
t14_1
t14_2 = tdist.ppf(0.975, df=14)
t14_2
print("Mean of var1 is : ",f"\n{xbar}\n")
print("Standard Deviation of data1 is : ", f"\n{xstd}\n")
print("Calculated t-statistic is : ",f"\n{t_cal}\n")
res = stats.ttest_1samp(data1, popmean=34.5, alternative='two-sided')
print(res)
ci = res.confidence_interval(confidence_level=0.95)
print(ci)
ci.low
ci.high
print("The Lower Value of 95% Confidence Interval is :", ci.low)
print("The Upper Value of 95% Confidence Interval is :", ci.high)
    
```

```

Mean of var1 is :
33.79800033569336

Standard Deviation of data1 is :
0.6089244484901428

Calculated t-statistic is :
-4.31357656270602
    
```

```

TtestResult(statistic=array([-4.3135763]), pvalue=array([0.00071453]), df=array([14]))
ConfidenceInterval(low=array([33.44895364]), high=array([34.14704703]))
The Lower Value of 95% Confidence Interval is : [33.44895364]
The Upper Value of 95% Confidence Interval is : [34.14704703]
    
```

- 계산된 검정통계량의 값은 -4.3136이고 5% 유의수준 하에서 자유도가 14인 t-분포의 임계치는 -2.145와 2.145이므로 5% 유의수준 하에서 모평균이 34.5라는 귀무가설을 기각

### 3. 모분산에 대한 가설검정

- (연습 2) 한 기술연구소에서 휴대전화 배터리 무게의 분산이 62g이라는 주장에 대한 양측검정을 하려고 한다. 휴대전화 7개를 무작위 선정하여 조사한 무게가 다음과 같으며, 휴대전화 배터리 무게는 정규분포한다고 하자. 5% 유의수준에서 휴대전화 배터리 무게의 분산이 62g이라는 귀무가설을 양측검정하라.

b1-ch7-3.py	36	37	38	39	39	44	47
-------------	----	----	----	----	----	----	----

```
import numpy as np
import scipy.stats as stats
weight = [36, 37, 38, 39, 39, 44, 47]
print("Data of time variable is :", weight)
xbar = np.mean(weight)
xbar
xvar = np.var(weight, ddof=1)
xvar
print("Mean of data is :", xbar)
print("Variance of data is :", xvar)
df=6
q=len(weight)-1
chi_cal = (df*xvar)/62
chi_cal
1-stats.chi2.cdf(chi_cal, 6)
print("Calculated Chi-statistic is :", chi_cal)
print("Probability of between 0 and Calculated Chi-statistic is :", 1-stats.chi2.cdf(chi_cal, 6))
chi_u = stats.chi2.ppf(0.975, df)
chi_l = stats.chi2.ppf(0.025, df)
print("Upper Critical Value of Chi2 w/ df=15 is :", chi_u)
print("Lower Critical Value of Chi2 w/ df=15 is :", chi_l)
```

Mean of data is : 40.0  
Variance of data is : 16.0

Calculated Chi-statistic is : 1.5483870967741935  
Probability of between 0 and Calculated Chi-statistic is : 0.9562158210771494

Upper Critical Value of Chi2 w/ df=15 is : 14.44937533544792  
Lower Critical Value of Chi2 w/ df=15 is : 1.2373442457912027

- 계산된 검정통계량의 값은 1.54838이고 5% 유의수준 하에서 자유도가 6인  $\chi^2$ -분포의 임계치는 1.237과 14.449이므로 5% 유의수준 하에서 모분산이 62g이라는 귀무가설을 허용

# 1. 모평균에 대한 가설검정

## (1) 두 집단의 분산을 모르지만 같다고 가정할 경우

- (연습 3) 대기업과 중소기업에 근무하는 근로자의 직장 만족도가 다음과 같이 조사되었다. 두 모집단의 분산은 동일하다고 가정하고, 중소기업 근로자들의 평균 직장 만족도가 대기업 근로자와 같다는 귀무가설 ( $\mu_1 = \mu_2$ ) 및 대립가설( $\mu_1 \neq \mu_2$ )을 설정하고 5% 유의수준에서 양측검정하라.

b1-ch7-4.py

```
import numpy as np
import scipy.stats as stats
from scipy.stats.distributions import t as tdist
small = [41, 45, 42, 62, 68, 54, 52, 55, 44, 60]
large = [74,74,70,52,76,91,71,78,76,78,83,50,52,66,65,53,72]
print("Data Frame of small is :",f'\n{small}\n')
print("Data Frame of large is :",f'\n{large}\n')
s_mean = np.mean(small)
l_mean = np.mean(large)
print("Mean of small is :",s_mean)
print("Mean of large is :",l_mean)
s_var = np.var(small, ddof=1)
l_var = np.var(large, ddof=1)
print("Variance of small is :",s_var)
print("Variance of large is :",l_var)
s_p = np.sqrt((9*s_var+16*l_var)/25)
print("Standard deviation of pooled variance is :",s_p)
t_cal = (s_mean-l_mean)/(s_p*(np.sqrt((1/10)+(1/17))))
t_cal
print("Calculated Value of t Statistic is :",t_cal)
res = stats.ttest_ind(a=small, b=large, equal_var=True, alternative='two-sided')
print(res)
t25 = tdist.ppf(0.025, df=25)
t25
print("Critical value of t distribution w/ p=0.05 and df=25 is :",t25)
```

표본 1(중소기업 근로자,  $n_1 = 10$ )

41	45	42	62	68	54	52	55	44	60
----	----	----	----	----	----	----	----	----	----

표본 1(대기업 근로자,  $n_2 = 17$ )

74	74	70	52	76	91	71	78	76	78
83	50	52	66	65	53	72			

Mean of small is : 52.3  
 Mean of large is : 69.47058823529412  
 Variance of small is : 85.12222222222223  
 Variance of large is : 138.76470588235296

Standard deviation of pooled variance is : 10.929474450526241  
 Calculated Value of t Statistic is : -3.9421075895786686

TtestResult(statistic=-3.9421075895786677, pvalue=0.0005747836765740096, df=25.0)  
 Critical value of t distribution w/ p=0.05 and df=25 is : -2.0595385527532946

- 계산된 검정통계량의 값은 -3.9421이고 5% 유의수준 하에서 자유도가 25인 t-분포의 임계치는 -2.06이므로 5% 유의수준 하에서 귀무가설을 기각

### (3) 독립이 아닌 짝진 표본

- (연습 5) 통계학을 수강하는 경영학부 학생들을 대상으로 보충수업이 학생들에게 도움이 되는지 알아보기 위해 6명을 임의로 선정하였다. 보충수업 전에 시험을 보게 하고 보충수업을 수강한 후 다시 시험을 보게 하였는데 보충수업이 학생들의 성적 향상에 도움이 되는지 5% 유의수준에서 단측검정하라.

b1-ch7-6.py

```
import pandas as pd
import numpy as np
import scipy.stats as stats
from scipy.stats.distributions import t as tdist
data = pd.read_excel("http://kanggc.iptime.org/book/data/chap11-3-2.xlsx")
data
var1 = data['var1']
var2 = data['var2']
d = var1 - var2
print("Data Frame of data is : ",f'\n{data}\n')
print("Data Frame of d is : ",f'\n{d}\n')
d_mean = np.mean(d)
d_mean
d_var = np.var(d, ddof=1)
d_var
t_cal = d_mean/(np.sqrt(d_var/6))
t_cal
t5_1 = tdist.ppf(0.05, df=5)
t5_1
t5_2 = tdist.ppf(0.95, df=5)
t5_2
print("Mean of d is : ",d_mean)
print("Variance of d is : ", d_var)
print("Calculated t-statistic is : ",t_cal)
res = stats.ttest_1samp(d, popmean=0, alternative='less')
print(res)
ci = res.confidence_interval(confidence_level=0.95)
print(ci)
ci.low
ci.high
print("The Lower Value of 95% Confidence Interval is : ", ci.low)
print("The Upper Value of 95% Confidence Interval is : ", ci.high)
print("Critical value of t distribution w/ p=0.05 and df=5 is : ",t5_1)
```

학생	보충수업 전 ( $X_1$ )	보충수업 후 ( $X_2$ )	점수 차이 ( $d = X_1 - X_2$ )
1	75	82	-7
2	71	73	-2
3	52	59	-7
4	46	48	-2
5	70	69	1
6	83	93	-10

Mean of d is : -4.5  
Variance of d is : 17.1

Calculated t-statistic is : -2.6655699499159153  
TtestResult(statistic=-2.665569949915916, pvalue=0.022292411754899415, df=5)

Critical value of t distribution w/ p=0.05 and df=5 is : -2.0150483726691575

- 계산된 검정통계량의 값은 -2.6655이고 5% 유의수준 하에서 자유도가 5인 t-분포의 임계치는 -2.015이므로 5% 유의수준 하에서 귀무가설을 기각

## 2. 모분산에 대한 가설검정

- (연습 6) 경제학과 및 경영학과 학생들의 계량경제 기말고사 성적이 다음과 같다. 두 모집단 모분산의 동일성 여부를 5% 유의수준에서 단측검정하라.

b1-ch7-7.py

```
import scipy.stats as st
import numpy as np
from scipy.stats.distributions import f as fdist
econ = [41,45,42,62,68,54,52,55,44,60]
mgt = [74,74,70,52,76,91,71,78,76,78,83,50,52,66,65,53,72]
print("Data Frame of econ is :",f"\n{econ}\n")
print("Data Frame of mgt is :", f"\n{mgt}\n")
v_econ = np.var(econ, ddof=1)
v_mgt = np.var(mgt, ddof=1)
print("Variance of econ is :",v_econ)
print("Variance of mgt is :", v_mgt)
F = v_mgt / v_econ
print("Calculated F-statistic is :",F)
df1 = len(mgt) - 1
df2 = len(econ) - 1
print("df of numerator is :", df1)
print("df of denominator is :", df2)
F_1 = fdist.ppf(0.95, df1, df2)
F_1
print("Value of F Distribution w/ df1=16 & df2=9 under 5% significance level is :",F_1)
```

경제					n <sub>1</sub> = 10)				
41	45	42	62	68	54	52	55	44	60
경영									
					n <sub>2</sub> = 17)				
74	74	70	52	76	91	71	78	76	78
83	50	52	66	65	53	72			

Variance of econ is : 85.12222222222223  
 Variance of mgt is : 138.76470588235296

Calculated F-statistic is : 1.6301818991530823

df of numerator is : 16  
 df of denominator is : 9

- 계산된 검정통계량의 값은 1.6302이고 5% 유의수준 하에서 분자의 자유도가 16이고, 분모의 자유도가 9인 F-분포의 임계치는 2.9889이므로 5% 유의수준 하에서 귀무가설을 허용

Value of F Distribution w/ df1=16 & df2=9 under 5% significance level is : 2.988965557308775