

# 由一個例子學習六大檢定



PYTHON

# 由一個例子學習六大檢定 (NYSE vs TWSE)



- 使用 Gujarati 的 Appendix D P.488 中所提供的 NYSE 28 檔股票的本益比 (P/E ratio) 資料作為學習範例，並加上 50 檔台灣 TWSE 對應股票本益比資料，來學習 6 大檢定。
- 共六道檢定問題
  - 範例1: 單母體平均值檢定
  - 範例2: 單母體變異數檢定
  - 範例3: 單母體比例檢定
  - 範例4: 雙母體平均數檢定
  - 範例5: 雙母體變異數檢定
  - 範例6: 雙母體比例檢定

# 範例1:單母體平均值檢定(雙尾檢定)



- 假說:  $\begin{cases} H_o: \mu_1 = 30 \\ H_1: \mu_1 \neq 30 \end{cases}$  虛無假說:  $H_o: \mu_1 = 30$
- 檢定統計量:  $t = \frac{\bar{X} - 30}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$  , 遵循一個自由度為  $n - 1$  的  $t$  分配
- 設定顯著水準  $\alpha = 0.05$  , 檢驗NYSE整體的本益比平均值是否等於30?

# Step1.利用pandas套件，讀入NYSE本益比資料



```
#讀入NYSE 本益比資料
import pandas as pd
NYSE= pd.read_csv("NYSE.csv")
NYSE.head()
```

P/E ratio	
0	27.96
1	22.90
2	8.30
3	49.78
4	24.68

## Step2.檢定NYSE的本益比是否平均數為30



```
#單母體平均值檢定  
#檢定NYSE的本益比是否平均數為30  
from scipy import stats  
stats.ttest_1samp(NYSE, 30, alternative="two-sided")
```

```
Ttest_1sampResult(statistic=array([-3.76327226]), pvalue=array([0.00082506]))
```

- 結論:
- (1) statistic=-3.76327226 拒絕域為  $|t| > 2.052$
- $\text{abs}(-3.76327226) = 3.76327226 > 2.052$ ，檢定統計量的值落於拒絕域中
- (2) p-value=0.00082506 < 0.05，在0.05顯著水準下，拒絕虛無假設
- 在0.05的顯著水準下，NYSE本益比平均值不為30。

## 範例2:單母體變異數檢定(雙尾檢定)



- 假說:  $\begin{cases} H_o: \sigma_1^2 = 100 \\ H_1: \sigma_1^2 \neq 100 \end{cases}$  虛無假說:  $H_o: \sigma_1^2 = 100$
- 檢定統計量:  $\chi^2 = \frac{(n-1)}{100} S^2 \sim: \chi^2_{(n-1)}$  , 遵循一個自由度為  $n - 1$  的:  $\chi^2$  分配
- 設定顯著水準  $\alpha = 0.05$  , 檢驗NYSE整體的本益比變異數是否等於30?

# Step1.利用pandas套件，讀入NYSE本益比資料



```
#讀入NYSE本益比資料
import pandas as pd
NYSE= pd.read_csv("NYSE.csv")
NYSE.head()
```

## P/E ratio

0	27.96
1	22.90
2	8.30
3	49.78
4	24.68

## Step2.檢定NYSE的本益比是否變異數為100



```
import numpy as np
#單母體變異數檢定
#檢定NYSE的本益比是否變異數為100

#計算樣本變異數
var=np.var(NYSE,ddof = 1)

#計算統計量
chi_square=((27)*var)/100
chi_square
```

$$\chi^2 = \frac{(n-1)}{100} S^2 \sim: \chi^2_{(n-1)}$$

```
P/E ratio      24.334802
dtype: float64
```

statistic=24.334802

拒絕域為 卡方統計量 > 43.1945 或 < 14.5734  
檢定統計量的值並未落於拒絕域中



## Step2.檢定NYSE的本益比是否變異數為100



```
#計算p-value，自由度為27  
from scipy.stats import chi2  
p_value=chi2.cdf(x=chi_square,df=27)  
p_value  
  
array([0.38831321])
```

$p\text{-value}=0.38831321 > 0.05$ ，在0.05顯著水準下，不拒絕虛無假設  
在0.05的顯著水準下，NYSE本益比變異數為100

## 範例3:單母體比例檢定(雙尾檢定)



- 假說:  $\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$  虛無假說:  $H_0: p = p_0$
- 檢定統計量:  $Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$
- $\sigma_{\hat{p}} = \sqrt{p_0 q_0 / n}$
- 設定顯著水準  $\alpha = 0.05$ ，調查本益比大於20的比例約0.5，隨機抽取28位，得知本益比大於25的比例為0.32，查調查結果是否可靠？

# Step1.利用pandas套件，讀入NYSE本益比資料



```
#讀入NYSE本益比資料
import pandas as pd
NYSE= pd.read_csv("NYSE.csv")
NYSE.head()
```

## P/E ratio

0	27.96
1	22.90
2	8.30
3	49.78
4	24.68

Step2.檢定得知本益比大於25的比例為0.32，查調查結果是否可靠？



- 調查本益比大於20的比例約0.5，隨機抽取28位，得知本益比大於25的比例為0.32，查調查結果是否可靠？

```
p=0.5  
q=1-p  
sigma=((p*q)/28)**0.5  
z=(0.32-0.5)/sigma  
z
```

$$\sigma_{\hat{p}} = \sqrt{p_0 q_0 / n}$$

$$Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$$

#z=-1.9049409439665053 ， 臨界值-1.96，不落入拒絕域

-1.9049409439665053

Step2.檢定得知本益比大於25的比例為0.32，查調查結果是否可靠？



```
#計算p-value
```

```
from scipy.stats import norm
```

```
p_value=norm.cdf(2.04)
```

```
p_value
```

```
#0.9793248371339299，不落入拒絕域
```

```
0.9793248371339299
```

## 範例4:雙母體平均值檢定(雙尾檢定)



- 假說:  $\begin{cases} H_o: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$  虛無假說:  $H_o: \mu_1 = 30$

- 檢定統計量

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{(n_1+n_2-2)}$$

- 遵循一個自由度為  $n_1 + n_2 - 2$  的  $t$  分配
- 設定顯著水準  $\alpha = 0.05$ ，檢驗NYSE與TWSE的本益比平均值是否有差異？

# Step1.利用pandas套件，讀入本益比資料



```
#讀入NYSE與TWSE本益比資料
import pandas as pd
PE_ratio= pd.read_csv("PE_RATIO.csv")
PE_ratio.head()
```

	TWSE	NYSE
0	15.13	27.96
1	13.48	22.90
2	23.74	8.30
3	22.40	49.78
4	15.45	24.68

## Step2.檢驗NYSE與TWSE的本益比平均值是否有差異？



```
# 個別指派變數
```

```
TWSE_PE_ratio=[PE_ratio.TWSE]
```

```
NYSE_PE_ratio=[PE_ratio.NYSE]
```

```
from scipy import stats
```

```
stats.ttest_ind(NYSE, TWSE, equal_var = True, alternative='two-sided')
```

```
Ttest_indResult(statistic=array([1.76677961]), pvalue=array([0.0812801]))
```

結論：

(1) statistic=1.76677961，檢定統計量的值並未落於拒絕域中

(2) p-value=0.0812801 > 0.05，在0.05顯著水準下，不拒絕虛無假設  
在0.05的顯著水準下，NYSE與TWSE本益比平均數沒有不同



## 範例5:雙母體變異數檢定(雙尾檢定)



- 假說:  $\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$  虛無假說:  $\sigma_1^2 = \sigma_2^2$

- 檢定統計量:

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1-1, n_2-1)}$$

- 遵循一個自由度為  $n_1 - 1$  和  $n_2 - 1$  的  $F$  分配
- 設定顯著水準  $\alpha = 0.05$ ，檢驗NYSE與TWSE本益比變異數是否相等？

# Step1.利用pandas套件，讀入本益比資料



```
#讀入NYSE與TWSE本益比資料
import pandas as pd
PE_ratio= pd.read_csv("PE_RATIO.csv")
PE_ratio.head()
```

	TWSE	NYSE
0	15.13	27.96
1	13.48	22.90
2	23.74	8.30
3	22.40	49.78
4	15.45	24.68

## Step2.檢驗NYSE與TWSE的本益比變異數是否有差異?



```
# 個別指派變數
```

```
TWSE_PE_ratio=[PE_ratio.TWSE]
```

```
NYSE_PE_ratio=[PE_ratio.NYSE]
```

```
#左尾檢定
```

```
from scipy.stats import f
```

```
n1=50; n2=28
```

```
alpha=0.05
```

```
var_NYSE=np.var(NYSE,ddof = 1)
```

```
var_TWSE=np.var(TWSE,ddof = 1)
```

```
fc=f.ppf(alpha,n1-1,n2-1)
```

```
fc #臨界值
```

```
0.583314124057284
```

```
f0=var_TWSE/var_NYSE
```

```
f0 #檢定統計量
```

```
#統計量>臨界值，不落入拒絕域
```

```
P/E ratio    3.377378
```

```
dtype: float64
```

$$F = \frac{S_1^2}{S_2^2} \sim F_{(n_1-1, n_2-1)}$$

## Step2.檢驗NYSE與TWSE的本益比變異數是否有差異?



```
#p-value  
pvalue = f.cdf(f0,n1-1,n2-1)  
pvalue  
#pvalue>0.05, 不落入拒絕域  
  
array([0.99944467])
```

結論:NYSE與TWSE整體本益比變異數相等

## 範例6:雙母體比例檢定(右尾檢定)

- 假說:  $\begin{cases} H_o: p_1 - p_2 \leq 0 \\ H_1: p_1 - p_2 > 0 \end{cases}$  虛無假說:  $H_o: p_1 - p_2 \leq 0$
- 檢定統計量:  $Z = \frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{S_{\widehat{P}_1 - \widehat{P}_2}}$
- $S_{\widehat{P}_1 - \widehat{P}_2} = \sqrt{\frac{P_1}{n_1} + \frac{P_2}{n_2}}$
- 設定顯著水準  $\alpha = 0.05$ ，調查建議投資NYSE還是TWSE，隨機抽出NYSE 28家，本益比大於20有19家，再隨機抽出TWSE 50家，本益比大於20有10家，檢定NYSE中本益比大於20的比例是否大於TWSE？

# Step1.利用pandas套件，讀入本益比資料



- 可讀或不讀入

#讀入NYSE本益比資料

```
import pandas as pd
NYSE= pd.read_csv("NYSE.csv")
NYSE.head()
```

P/E ratio	
0	27.96
1	22.90
2	8.30
3	49.78
4	24.68

#讀入TWSE本益比資料

```
import pandas as pd
TWSE= pd.read_csv("TWSE.csv")
TWSE.head()
```

P/E ratio	
0	15.13
1	13.48
2	23.74
3	22.40
4	15.45

## Step2.檢驗NYSE中本益比大於20的比例是否大於TWSE?

```
#右尾檢定
#調查建議投資NYSE還是TWSE，隨機抽出NYSE 28家，本益比大於20，有19家
#再隨機抽出TWSE 50家，本益比大於20，有10家
#檢定NYSE中本益比大於20的比例是否大於TWSE
p0=0
n1=50; n2=28
pbar1=19/28
pbar2=10/50
pbar=(n1*pbar1+n2*pbar2)/(n1+n2)
se=np.sqrt(pbar*(1-pbar)*(1/n1+1/n2))
pbar0=pbar1-pbar2
alpha=0.05
zc=norm.ppf(1-alpha,0,1)
zc #臨界值
```

1.6448536269514722

```
#pbarc=p0+zc*se
z0=(pbar0-p0)/se
z0
```

4.05539694760702

```
#p-value
1-norm.cdf(pbar0,p0,se)
1-norm.cdf(z0,0,1)
#落入拒絕域，代表NYSE的本益比大於20的比例高於TWSE
```

2.502461739117834e-05