

單一母體平均值檢定

Notes-Tests-no-means-and-variances

2 單一母體平均值檢定

- 理論基礎簡述

▷ 虛無假設為 $\rightarrow H_0: \mu = 10$ (對立假設按單雙尾有不同寫法)

▷ 檢定統計量為 $\rightarrow t = \frac{\bar{X} - 10}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$ ，它遵循一個自由度為 $n-1$ 的 t 分配

EX_1: 銀行客戶的等候時間

- 例題 1: 銀行客戶的等候時間

銀行經理懷疑客戶平均等候時間大於 10 分鐘，因而設立虛無假設和對立假設為：

$$H_0: \mu = 10 \text{ (or } \mu \leq 10)$$

$$H_1: \mu > 10$$

銀行經理設定顯著水準 $\alpha = 0.05$ 。(傳統法才需先決定 α ， p -value 法則不用)

EX_1: 銀行客戶的等候時間

根據 $\alpha = 0.05$ ，此次假設檢定的拒絕域為 $t > 1.860$ 。(此為單尾檢定，此處 1.860 就是臨界值，可由查表得知 → 聯想：法官選的 5 億為臨界值。)

銀行經理隨機抽取 9 位客戶，並記錄客戶的等候時間如下：

9.8 7.4 8.3 12.5 10.2 13.2 15.0 12.0 12.3

(注意：現在才完成抽樣，決定是那 9 個客戶被抽到，這時 t 統計量就針對這組樣本的是一個數值了 → 聯想： Y 統計量是針對某甲的數值為 7 億)

經計算 $\bar{X} = 11.1889$ ， $S = 2.4441$ ，可得 (針對這組樣本的) t 統計量的值為

$$t = \frac{\bar{X} - 10}{\frac{S}{\sqrt{9}}} \sim t_{(8)}$$

$$t = \frac{11.1889 - 10}{\frac{2.4441}{\sqrt{9}}} = 1.4593$$

因 $1.4593 < 1.860$ ，檢定統計量的值並未落在拒絕域內，所以，在 0.05 的顯著水準下，不拒絕虛無假設，也就是說，在 0.05 的顯著水準下，根據此 9 位客戶的等候時間，並沒有充分證據顯示銀行客戶 (指母體，所有銀行客戶) 平均等候時間大於 10 分鐘。

EX_1:銀行客戶的等候時間_Python

例題1.銀行客戶的等候時間

```
In [34]: #呼叫統計套件
#scipy.stats版本 1.6.0
from scipy import stats

#輸入資料
dat = [ 9.8, 7.4, 8.3, 12.5, 10.2, 13.2, 15.0, 12.0, 12.3]

#檢定是否平均數為10
scipy.stats.ttest_1samp(dat, 10, alternative="greater")
```

```
Out[34]: Ttest_1sampResult(statistic=1.459298210492694, pvalue=0.09130152865589224)
```

結論:

(1)statistic=1.459298210492694 <1.860，檢定統計量的值並未落於拒絕域內

(2)pvalue=0.09130152865589224 >0.05，在0.05顯著水準下，不拒絕虛無假設

在0.05顯著水準下，根據此9位客戶的等候時間，並沒有充分證據顯示銀行客戶(指母體，所有銀行客戶)平均等候時間大於10分鐘

EX_1:銀行客戶的等候時間_Excel

$$t = \frac{\bar{X} - 10}{\frac{S}{\sqrt{9}}} \sim t_{(8)}$$

	A	B	C	D	E
1	9.8		Mean	11.18889	=AVERAGE(A1:A9)
2	7.4		Standard deviation	2.444097	=STDEV(A1:A9)
3	8.3		Count(n)	9	=COUNT(A1:A9)
4	12.5		Standard error of mean (SEM)	0.814699	=D2/SQRT(D3)
5	10.2		Degree of freedom (df)	8	=D3-1
6	13.2		Hypothesized mean (μ)	10	
7	15				
8	12		t-statistic	1.459298	=(D1-D6)/D4
9	12.3		p-value	0.091302	=TDIST(D8,D5,1)
10					

1:單尾檢定
2:雙尾檢定

不拒絕虛無假設

EX_2:我國境內基金現金股息的檢定

- 例題 2: 我國境內基金現金股息的檢定

國外財務規劃專業人員認為，我國境內基金因全球金融海嘯影響到其績效表現，而懷疑整體基金（指母體）過去 5 年累計現金股息發放金額的平均數是低於 2.2 元，因此設立虛無假設與對立假設為：

$$H_0: \mu = 2.2 \quad (\text{or } \mu \geq 2.2)$$

$$H_1: \mu < 2.2$$

該專員設定顯著水準 $\alpha = 0.1$ (10%)。

該專員決定抽取 6 家基金（小樣本），故在母體為常態分配的假設下，檢定統計量（test statistic）選擇為：

$$t = \frac{\bar{X} - 2.2}{\frac{S}{\sqrt{6}}} \sim t_{(5)}$$

此次假設檢定的拒絕域為 $t < -1.476$ （查表練習）。

EX_2:我國境內基金現金股息的檢定

該專員隨機抽取 6 家基金，並分別查詢其過去 5 年現金股息累積金額如下：

配息期間	日盛 基金	兆豐 國際國民	建弘 雙福	富邦 基金	寶來 台灣金融	元大 多元
2007~2011 年	1.48	1.12	2.35	0.96	0.7	2.66

經計算 $\bar{X} = 1.5450$ ， $S = 0.7915$ ， t 統計量的值為

$$t = \frac{\bar{X} - 2.2}{\frac{S}{\sqrt{6}}} \sim t_{(5)}$$

$$t = \frac{1.5450 - 2.2}{\frac{0.7915}{\sqrt{6}}} = -2.0271$$

因 $-2.0271 < -1.476$ ，檢定統計量的值落在拒絕域內，所以，在 0.1 的顯著水準下，拒絕虛無假設。也就是說，在 0.1 的顯著水準下，由此 6 家基金過去 5 年發放現金股息的情形，有充分證據顯示我國境內基金（指母體）過去 5 年現金股息發放總金額的平均數是低於 2.2 元。

EX_2:我國境內基金現金股息的檢定_Python

例題2:我國境內基金現金股息的檢定

```
In [35]: #呼叫統計套件
#scipy.stats版本 1.6.0
from scipy import stats

#輸入資料
dat = [1.48, 1.12, 2.35, 0.96, 0.7, 2.66]

#檢定是否平均數為10
stats.ttest_1samp(dat, 2.2, alternative="less")
```

```
Out[35]: Ttest_1sampResult(statistic=-2.0270608498954967, pvalue=0.04923975317006294)
```

結論:

(1)statistic=-2.0270608498954967 < -1.476，檢定統計量的值落於拒絕域中

(2)pvalue=0.04923975317006294 < 0.1，在0.1顯著水準下，拒絕虛無假設

在0.1的顯著水準下，由此6家基金過去5年發放現金股利的情形，有充分證據顯示我國境內基金(指母體)過去5年現金股息發放總金額的平均數是低於2.2元

EX_2:我國境內基金現金股息的檢定_Excel

	A	B	C	D	E	F
1	1.48		Mean	1.545	=AVERAGE(A1:A6)	
2	1.12		Standard deviation	0.791499	=STDEV(A1:A6)	
3	2.35		Count(n)	6	=COUNT(A1:A6)	
4	0.96		Standard error of mean (SEM)	0.323128	=D2/SQRT(D3)	
5	0.7		Degree of freedom (df)	5	=D3-1	
6	2.66		Hypothesized mean (μ)	2.2		
7						
8			t-statistic	-2.02706	=(D1-D6)/D4	
			p-value	0.04924	=TDIST(ABS(D8),D5,1)	

$$t = \frac{\bar{X} - 2.2}{\frac{S}{\sqrt{6}}} \sim t_{(5)}$$

拒絕虛無假設

1:單尾檢定
2:雙尾檢定

EX_3: 白氏雞精標示焦糖比重的檢定

- 例題 3: 白氏雞精標示焦糖比重的檢定

白氏雞精標示焦糖比重是 0.32%，公司品保人員定期抽樣檢驗白氏雞精的焦糖比重，以確保出廠的雞精有最佳的口感，在抽樣檢驗時，品保人員設立虛無假設和對立假設為：(太大太小都不好 → 雙尾檢定)

$$H_0: \mu = 0.32$$

$$H_1: \mu \neq 0.32$$

品保人員設定顯著水準 $\alpha = 0.01$ 。

由於檢驗一瓶就耗損一瓶，為節省成本，品保人員決定抽取 10 瓶 (小樣本)，故在母體為常態分配的條件下，檢定統計量選擇為：

$$t = \frac{\bar{X} - 0.32}{\frac{S}{\sqrt{10}}} \sim t_{(9)}$$

此次假設檢定的拒絕域為 $|t| > 3.250$ 。(注意：雙尾檢定的拒絕域寫法不同)

EX_3: 白氏雞精標示焦糖比重的檢定

品保人員隨機抽取 10 瓶，經測量焦糖比重紀錄如下：

0.312 0.314 0.309 0.318 0.321 0.310 0.312 0.318 0.321 0.315

經計算： $\bar{X} = 0.3150$ ， $S = 0.0043$ ， t 統計量的值為

$$t = \frac{\bar{X} - 0.32}{\frac{S}{\sqrt{10}}} \sim t_{(9)} \qquad t = \frac{0.3150 - 0.32}{\frac{0.0043}{\sqrt{10}}} = -3.6771$$

因 $|t| = 3.6771 > 3.250$ ，檢定統計量之值落入拒絕域內，所以，在 0.01 的顯著水準下，拒絕虛無假設。也就是說，在 0.01 的顯著水準下，此 10 瓶的焦糖比重有充分證據顯示白氏雞精的平均焦糖比重不是 0.32%。

EX_3:白氏雞精標示焦糖比重的檢定_Python

例題3:白氏雞精標示焦糖比重的檢定

```
In [36]: #呼叫統計套件
#scipy.stats版本 1.6.0
from scipy import stats

#輸入資料
dat = [0.312, 0.314, 0.309, 0.318, 0.321, 0.310, 0.312, 0.318, 0.321, 0.315]

#檢定是否平均數為0.32
stats.ttest_1samp(dat, 0.32, alternative="two-sided")
```

```
Out[36]: Ttest_1sampResult(statistic=-3.6380343755449545, pvalue=0.005417433388261108)
```

結論:

(1)statistic=-3.6380343755449545

$\text{abs}(-3.6380343755449545) = 3.6380343755449545 > 0.3250$ ，檢定統計量的值落於拒絕域中

(2)pvalue=0.005417433388261108<0.01，在0.01顯著水準下，拒絕虛無假設

在0.01的顯著水準下，此10瓶的焦糖比重有充分證據顯示白氏雞精的平均焦糖比重不是0.32%

#文件t統計量算出-3.6771，而程式算出-3.638，取決於文件取S=0.0043，而程式取S=0.004346

EX_3:白氏雞精標示焦糖比重的檢定_Excel

	A	B	C	D	E	F
1	0.312		Mean	0.315	=AVERAGE(A1:A10)	
2	0.314		Standard deviation	0.004346	=STDEV(A1:A10)	
3	0.309		Count(n)	10	=COUNT(A1:A10)	
4	0.318		Standard error of mean (SEM)	0.001374	=D2/SQRT(D3)	
5	0.321		Degree of freedom (df)	9	=D3-1	
6	0.31		Hypothesized mean (μ)	0.32		
7	0.312					
8	0.318		t-statistic	-3.63803	=(D1-D6)/D4	
9	0.321		p-value	0.005417	=TDIST(ABS(D8),D5,2)	
10	0.315					

1:單尾檢定
2:雙尾檢定

拒絕虛無假設

$$t = \frac{\bar{X} - 0.32}{\frac{S}{\sqrt{10}}} \sim t_{(9)}$$

補充:利用Excel資料分析，跑出敘述統計
以EX_1:銀行客戶的等候時間為例

- Step1:利用資料分析，跑出敘述統計

檔案常用插入版面配置公式資料校閱檢視開發人員增益集說明告訴我您想做什麼

取得資料

查詢與連線

排序與篩選

資料工具

預測

大綱

取得及轉換資料

查詢與連線

排序與篩選

資料工具

預測

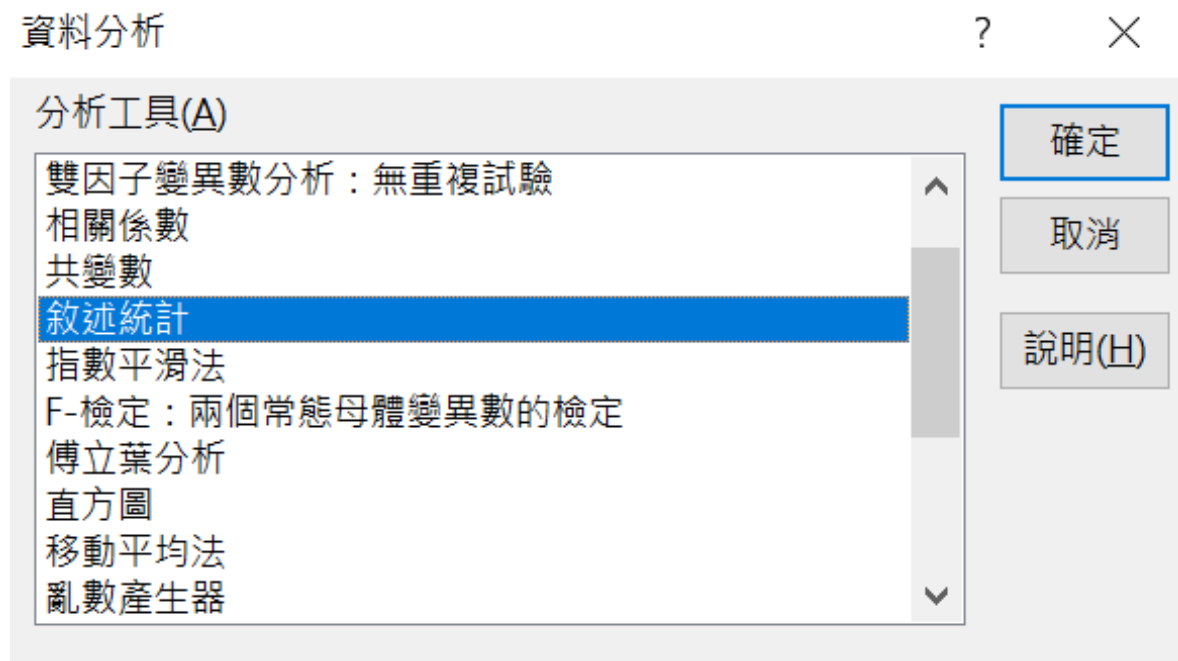
大綱

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	9.8		Mean	11.18889										
2	7.4		Standard deviation	2.444097										
3	8.3		Count(n)	9										
4	12.5		Standard error of mean (SEM)	0.814699										
5	10.2		Degree of freedom (df)	8										
6	13.2		Hypothesized mean (μ)	10										
7	15													
8	12		t-statistic	1.459298										
9	12.3		p-value	0.091302										

資料分析工具
財務和科學資料分析的
工具

補充:利用Excel資料分析，跑出敘述統計 以EX_1:銀行客戶的等候時間為例

- Step2:敘述統計



補充：利用Excel資料分析，跑出敘述統計
以EX_1：銀行客戶的等候時間為例

- Step3:選取資料範圍，並選出輸出摘要統計

A	B	C	D	E
9.8		Mean	11.18889	
7.4		Standard deviation	2.444097	
8.3		Count(n)	9	
12.5		Standard error of mean (SEM)	0.814699	
10.2		Degree of freedom (df)	8	
13.2		Hypothesized mean (μ)	10	
15				
12		t-statistic	1.459298	
12.3		p-value	0.091302	

敘述統計

輸入

輸入範圍(I):
\$A\$1:\$A\$9

分組方式:
☒ 逐欄(C)
☐ 逐列(R)

☐ 類別軸標記是在第一列上(L)

輸出選項

☐ 輸出範圍(O):
☐ 新工作表(P):
☐ 新活頁簿(W)

☒ 摘要統計(S)

☐ 平均數信賴度(N):
☐ 第 K 個最大值(A):
☐ 第 K 個最小值(M):

95

1

1

%

確定

取消

說明(H)

補充:利用Excel資料分析，跑出敘述統計 以EX_1:銀行客戶的等候時間為例

- Step4:輸出結果

	A	B
1	欄1	
2		
3	平均數	11.1888889
4	標準誤	0.81469907
5	中間值	12
6	眾數	#N/A
7	標準差	2.4440972
8	變異數	5.97361111
9	峰度	-0.72331716
10	偏態	-0.15961283
11	範圍	7.6
12	最小值	7.4
13	最大值	15
14	總和	100.7
15	個數	9
16		