第三章 敘述統計量

描述統計資料之特性的統計量數有二項：

1.集中趨勢量數：眾數(Mode)、中位數(Median)、平均數(Mean)

2.離散趨勢量數：全距(Range)、標準差(Standard Deviation)、 變異數(Variance) 、變異係數(Coefficient of Variation)、四分位(Quartile)、四分位距(Inter-quartile Range) 、十分位(decile)、百分位(Percentile)

3-1 集中趨勢量數

1.平均數(mean)

(1)母體平均數(population mean)：以 μ 表之，若母體有 N 個，分別為 $x_1, x_2,\ldots,x_N$，則其數學式如下：

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

(2)樣本平均數(sample mean)：以 $\bar{x}$ 表之，若樣本資料有 n 個，分別為 $x_1, x_2,\ldots,x_n$，

則其數學式如下

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

(例 1)：抽樣出來的數據如下：

3,4,1,4,6 求平均數。

解:

$$\bar{x} = \frac{\sum x}{n} = \frac{3+4+1+4+6}{5} = \frac{18}{5} = 3.6$$

2.中位數(median)

將資料由小到大排序，位置居中者，就稱為該組資料之中位數，一般以 Me 表示。資料若有極端值，使用中位數來代表集中趨勢量數最適宜。

(例 2)：抽樣出來的數據如下： 3,4,1,4,6　　求中位數。

解:

(1)將資料排序

1,3,4,4,6

(2)位置

$$i=5*\frac{1}{2}=2.5\cong 3$$

(3) Me=X_3=4

<速解>

(1)由小到大排序

(2)找位置及求值

i=個數*代表值

$$\begin{Bmatrix} \text{小數} \rightarrow \text{無條件進位，所對應值} \\ \text{整數} \rightarrow (\text{對應值}+\text{下一個對應值})/2 \end{Bmatrix}$$

3.眾數(Mode)

一組資料中出現次數最多的數值，一般以 *Mo* 表示，若次數皆相同則無眾數。

(例 3)：抽樣出來的數據如下： 3,4,1,4,6　　求眾數。

解:

因 4 出現次數最多次，所以 Mo=4

3-2 離散趨勢量數

1.四分位數(quartile)：

將一組資料分割成 4 等分，此 3 個數值稱為四分位數，通常以 Qi 表示

$$Q_1=P_{25}，Q_2=P_{50}=Me，Q_3=P_{75}$$

(例 4)： 抽樣出來的數據如下： 3,4,1,4,6　　　求 Q1 及 Q3。

解:

將資料排序

1,3,4,4,6

(1)Q_1 位置

$$i=5*\frac{1}{4}=1.25\cong 2$$

所以　$Q_1=X_2=3$

(2) Q_3 位置

$$i=5*\frac{3}{4}=3.75\cong 4$$

所以　$Q_3=X_4=4$

2.百分位數(percentile)：

將資料依大小順序排列，取 99 個等分點，每一等分點皆稱為百分位數

(例 5)：某班級共 50 人，某次統計學成績如下所示:

22	25	26	29	29	40	42	43	46	56
57	59	60	60	60	61	62	64	66	68
70	71	71	72	75	76	76	77	78	79
82	82	85	85	86	86	86	88	88	86
89	90	91	91	92	92	96	94	96	97

試求出 P_{25},P_{30}。

(1)P_{25} 位置

$$i = 50 * \frac{25}{100} = 12.5 \cong 13$$

所以　P25=X13=60

(2) P_{30} 位置

$$i = 50 * \frac{30}{100} = 15$$

所以　P30=$\frac{X_{15}+X_{16}}{2} = \frac{60+61}{2} = 60.5$

3.全距(range,R)

R=最大值-最小值

(例 6)：設有一組樣本資料如下：

3,4,1,4,6　　試求全距。

解:

$R=X_{MAX}-X_{MIN}=6-1=5$

4.四分位距(Inter-quartile Range，IQR)

$IQR=Q_3-Q_1$

(例 7)：設有一組樣本資料如下：

3,4,1,4,6　　試求四分位距

解:

先排序 1,3,4,4,6

$Q_1=3$

$Q_3=4$

IQR=4-3=1

5.變異數(variance)與標準差(standard deviation)

(1)母體變異數

(公式 1)

$$\sigma^2 = \frac{\sum (x-\mu)^2}{N}$$

(公式 2)

$$\sigma^2 = \frac{\sum x^2 - N\mu^2}{N}$$

(公式 3)

$$\sigma^2 = \frac{N\sum x^2 - (\sum x)^2}{N^2}$$

(2)母體標準差

$$\sigma = \sqrt{\sigma^2}$$

(3)樣本變異數

(公式 1)

$$S^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

(公式 2)

$$S^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1}$$

(公式 3)

$$S^2 = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$$

(4)樣本標準差

$$s = \sqrt{s^2}$$

證明：$\sum(X-\mu)^2 = \sum X^2 - N\mu^2$

證：$\sum(X-\mu)^2 = \sum(X^2 - 2X\mu + \mu^2) = \sum X^2 - 2\mu\sum X + \sum \mu^2$

$=\sum X^2 - 2\mu N\mu + N\mu^2 = \sum X^2 - 2N\mu^2 + N\mu^2 = \sum X^2 - N\mu^2$

註：$\sum X = N\mu$;$\sum \mu^2 = N\mu^2$

(例 8)： 設有一組樣本資料如下：
3,4,1,4,6 試求變異數及標準差

解:

利用公式 1:

$$S^2=\frac{\sum(x-\bar{x})^2}{n-1}$$

X	$x-\bar{x}$	$(x-\bar{x})^2$
3	-0.6	0.36
4	0.4	0.16
1	-2.6	6.76
4	0.4	0.16
6	2.4	5.76
18		13.2

$$\bar{x}=\frac{\sum x}{n}=\frac{18}{5}=3.6$$

$$S^2=\frac{\sum(x-\bar{x})^2}{n-1}=\frac{13.2}{5-1}=\frac{13.2}{4}=3.3$$

$$S=\sqrt{3.3}=1.8166$$

利用公式 3:

$$S^2=\frac{n\sum x^2-(\sum x)^2}{n(n-1)}$$

X	X^2
3	9
4	16
1	1
4	16
6	36
18	78

$$\bar{x}=\frac{\sum x}{n}=\frac{18}{5}=3.6$$

$$S^2=\frac{n\sum x^2-(\sum x)^2}{n(n-1)}=\frac{5*(78)-(18)^2}{5(5-1)}=\frac{390-324}{20}$$

$$=\frac{66}{20}=3.3$$

$$S=\sqrt{3.3}=1.8166$$

6.變異係數(coefficient of variation)

變異係數為標準差除以平均數再乘以 100%

$$\text{變異係數(CV)}=\frac{\text{標準差}}{\text{平均數}}*100\%$$

(例 9) ： 設有一組樣本資料如下： 3,4,1,4,6 試求變異係數

解:

$$cv=\frac{s}{\bar{x}}=\frac{1.8166}{3.6}\times 100\%=0.5046\times 100\%=50.46\%$$

3-3 柴比雪夫與經驗法則

1.柴比雪夫不等式(Chebyshev's Theorem)

◎柴比雪夫不等式主要用在不知母體分配的情況下，估計某變數所涵蓋範圍的機率

◎必須先知道母體平均數與變異數才能利用柴氏不等式求其機率

◎至少有 $1-1/k^2$ 的資料，在距離平均數 k 個標準差的範圍內

2.經驗法則(Empirical Rule)

◎經驗法則主要用在資料呈單峰對稱分配或鐘型分配時，估計某變數所涵蓋範圍的機率

◎約有 68%的資料，在距離平均數 1 個標準差的範圍內

約有 95%的資料，在距離平均數 2 個標準差的範圍內

約有 99.7%的的資料，在距離平均數 3 個標準差的範圍內

3.柴比雪夫與經驗法則之比較表

K(多少個σ)	柴比雪夫	經驗法則
1	無	68%
2	至少 3/4	95%
3	至少 8/9	99.7%
4	至少 15/16	100%

(例 10)：有 100 位學生修課，期中考成績之平均數為 70，標準差為 5。有多少學生的分數介於 60 與 80 之間(1)用柴比雪夫(2)用經驗法則，分別算出?

$$K = \frac{80-70}{5} = 2$$

(1)柴比雪夫

100 x 3/4 = 75，至少 75 人

(2)經驗法則

100 x 0.95 = 95，95 人

第三章之 EXCEL 應用

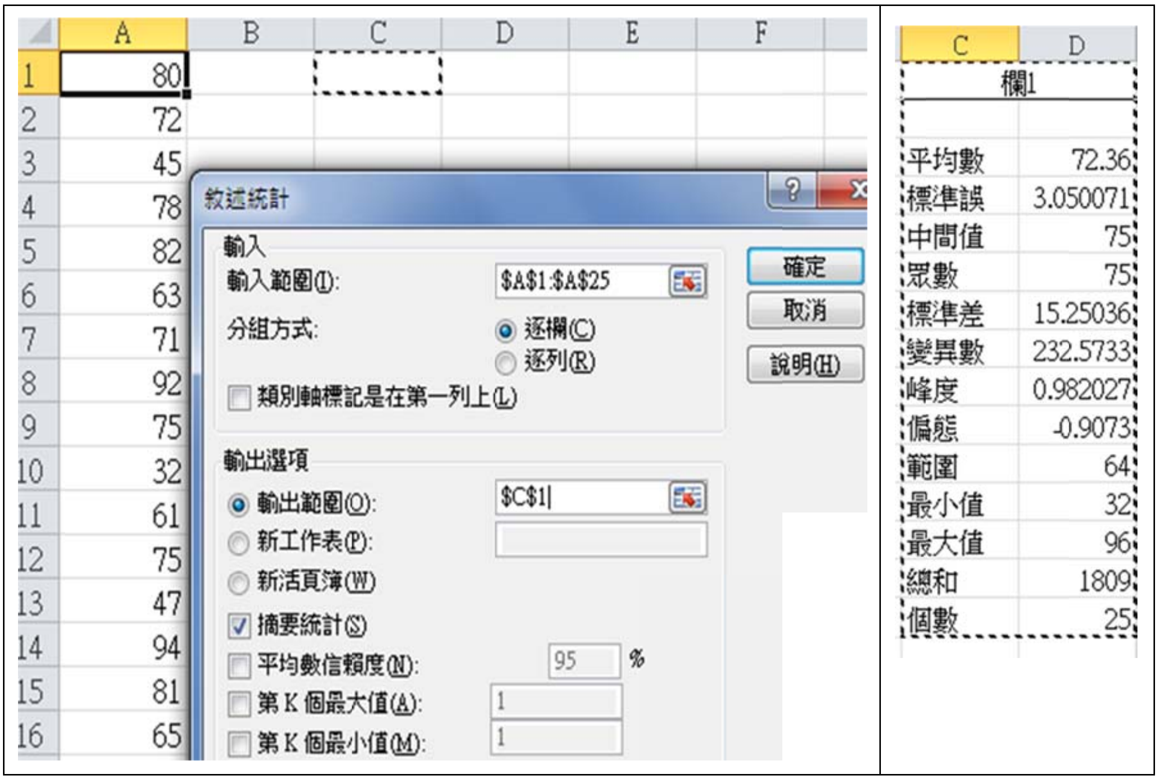
(實例 1):下列資料，利用 Excel 之資料分析，算出摘要統計量

80,72,45,78,82,63,71,92,75,32,61,75,47,94,81,65,68,73,84,75,79,58,78,85,96

解:

1.資料分析→敘述統計→確定

2.輸入範圍☐→⊙輸出範圍☐→☑摘要統計→確定



綜合練習 3

1.有一組樣本資料值為 8, 3, 10, 3, 11, 4 請計算其平均數、眾數、中位數、全距、Q1、Q2、Q3、IQR、變異數、標準差、變異係數。

2.若母體有 10 個，平均數為 16 ，標準差為 5，其中有一數正確應為 23，卻誤為 32，問修正後平均數及標準差應為多少？

3.全班有 55 人，若平均數為 72 分，標準差為 4 分，欲算出 64 分到 80 分有多少人？

(1)若利用柴比雪夫，有多少人？

(2)若利用經驗法則，有多少人？