

第6章 分析資料-以統計測量數呈現 基礎統計學二版

6 分析資料-以統計測量數呈現

○ 學習目的

1. 瞭解資料中心位置之各種衡量指標如算術平均數、中位數、眾數、等的衡量方法。
2. 熟習各個中心位置衡量方法的特性、使用時機與優缺點。
3. 瞭解資料分散程度之各種衡量指標如四分位距、變異數、標準差、變異係數的衡量方法。
4. 熟習各個等分位數衡量方法的特性、使用時機與優缺點。
5. 熟習使用 EXCEL 計算中心位置與分散度指標及分位數之指標。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

本章結構

分析資料-以統計測量數呈現

- 中心位置的
衡量
 - 平均數
 - 中位數
 - 眾數
 - 中心位置
各統計測
量數之比
較與選擇
- 等分位置
的衡量
 - 四分位數
 - 十分位數
 - 百分位數
- 盒鬮圖
分析法
- 分散度
的衡量
 - 平均絕對
離差
 - 變異數
 - 標準差
 - 相對分散度
- Excel 的
使用
 - 柴比氏定理

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

○ 平均數

算術平均數的意義

所有觀察值的總和除以觀察值的個數即為算術平均數。算術平均數在數線上代表資料的平衡點。

母體平均數

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

樣本平均數

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

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

表6.1 銀行業與證券業的月薪

	A	B	C	D	E	F	G	H	I
1 銀行業	25	26	28	28	29	31	32	33	
2 證券業	20	21	22	22	24	27	28	76	

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

○ 算數平均數的特質

- ① 資料的平衡點
- ② 各觀察值與平均數間的差的總和最小
- ③ 優點為考慮到每一個觀察值，缺點為易受極端值的影響。
- ④ 可進行代數演算
- ⑤ 可對觀察值予以加權

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

圖6.1 銀行業的平均月薪

圖6.2 證券業的平均月薪

林惠玲 陳正倉著 雙葉書報發行 2008

中心位置的衡量

○ 平均數

加權算術平均數

$$\bar{X}_w = \frac{\sum_{i=1}^n W_i x_i}{\sum_{i=1}^n W_i}$$

中心位置的衡量

表6.2 朱文慶大一下的成績單

	A	B	C	D	E
1 科目		學分數	學期成績	加權成績	
2 國文		3	86	258	
3 英文		3	87	261	
4 歷史		2	95	190	
5 體育		1	87	87	
6 微積分		3	82	246	
7 經濟學原理		4	86	344	
8 會計學		3	89	267	
9 社會心理學		3	90	270	
10 工程數學與社會適應		2	91	182	
11		24	2105	87.71	

中心位置的衡量

○ 中位數

中位數是位於依數值大小順序排列的觀察值中央的那一個數值。

中心位置的衡量

○ 中位數的特質

中位數只是觀察值數列中的一個數值，因此不受極端值的影響，故對觀察值的變化不敏感。

中心位置的衡量

○ 眾數

眾數是指觀察值中其出現次數最多的那一個數值。

中心位置的衡量

○ 眾數的性質

- ① 不受極端值的影響
- ② 可能有多個或一個也沒有
- ③ 對觀察值的個數或數值變化的感應不靈敏
- ④ 眾數大都應用於類別資料

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

表6.3 中心位置統計測量數的比較

	A	B	C	D	E
1	銀行業		證券業		
2					
3	平均數	29	平均數	30	
4	中間值	28.5	中間值	23	中位數
5	眾數	28	眾數	22	

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

表6.4 中心位置統計測量數之比較

統計測量數	優點	缺點
算術平均數	1.資料的重心。資料無極端值或偏態時，具代表性。 2.適合代數演算 3.考慮所有觀察值，敏感度高。 4.觀察值與平均數平方和最小	1.若有極端值存在時，則不具代表性。 2.資料如為偏態，則代表性較差。
中位數	1.適用於有極端值的資料 2.適用於偏態資料 3.觀察值與中位數絕對差和最小	1.不適合代數演算 2.對觀察值敏感性低
眾數	1.適用於有極端值的資料 2.適用於偏態資料 3.適用於質的資料	1.可能不只一個或不存在 2.敏感性低

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

表6.5 士林與桃園地院訴訟案件審理日數

士林地院				桃園地院			
65	110	58	72	66	67	68	65
73	78	64	77	63	73		
78	90	82	86	105	61	81	59

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

中心位置的衡量

表6.6 地院訴訟案件審理日數的平均數中位數與眾數

	A	B	C	D	E
1	士林地院		桃園地院		
2					
3	平均數	75.58	平均數	73.4	
4	中間值	70	中間值	73	中位數
5	眾數	65	眾數	73	

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

等分位置的衡量

- **四分位數**
四分位數是將順序資料分成四等分數值的分位數。四分位數有第1、第2、第3三個四分位數。
- **十分位數**
十分位數是將資料均分為十等份數值的分位數。
- **百分位數**
百分位數是將順序資料均分為一百等分數值的分位數。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

等分位置的衡量

圖6.3 第四分位數的圖解

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

等分位置的衡量

產業經濟學的學期成績

78	79	80	81	82	83	83	84	84
85	86	87	88	89	90	91	92	95

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

盒鬚圖分析法

如果我們想要得知整個資料的分布情形，我們可以利用 5 個重要的測量數最小值： (min) 、 Q_1 、 M 、 Q_3 、及最大值 (max) 來表示資料的分布，這 5 個數稱為 5 個摘要數 (five numbers summary)。5 個摘要數可以盒鬚圖 (box-and-whisker plot) 來描述，而從盒鬚圖可更清楚地得知資料的分布。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

盒鬚圖分析法

表6.7 鍾樂水與朱碧霞的業績

鍾樂水	30	63	66	78	82	96	106	270
朱碧霞	64	82	88	90	96	108	128	166

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

盒鬚圖分析法

圖6.4 業務員業績的盒鬚圖

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

最常用來描述資料分布的是以平均數來描述資料的中心位置，以標準差 (standard deviation) 來描述資料的分散度。

分散程度或變異性的測量數主要有平均絕對離差 (average absolute deviation)，變異數 (variance)，標準差 (standard deviation) 以及衡量相對分散程度的變異係數 (coefficient of variation) 等。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

圖6.5 A股票價格的分配

圖6.6 B股票價格的分配

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

○ 平均絕對離差

母體： $MAD = \frac{1}{N} \sum_{i=1}^N |x_i - \mu|$

樣本： $mad = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{X}|$

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

表6.8 縱貫路與中山高的開車時間

縱貫路	38	36	40	39	42
中山高	32	19	41	47	56

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

表6.9 縱貫路開車時間的平均絕對離差

開車時間	$X - \bar{X}$	$ X - \bar{X} $	$(X - \bar{X})^2$
38	-1	1	1
36	-3	3	9
40	1	1	1
39	0	0	0
42	3	3	9
合計	0	8	20

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

○ 變異數

母體變異數

$$\sigma^2 = \frac{1}{N} \sum (x_i - \mu)^2$$

式中： μ ：母體平均數， N ：母體個數。

樣本變異數

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

式中： \bar{X} ：樣本平均數， n ：樣本數。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

○ 變異數的性質

- ① 變異數的值大於等於 0，若變異數為 0 時，其意義是所有觀察值均相同，沒有變異（分散）。
- ② 考慮每一個觀察數值
- ③ 適合代數演算
- ④ 若同一組資料單位不同，其變異數亦不相同。
- ⑤ 單位相同可作比較
- ⑥ 適合利用樣本變異數對母體變異數做統計推論
- ⑦ 具有複名數（如元²），不易解釋。如電腦價格的變異數的單位為平方元（元²），不具意義。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

○ 標準差

母體標準差

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

樣本標準差

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

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

表6.10 開車時間的變異數與標準差

	A	B	C	D
1	經貴路開車時間		中山高開車時間	
2				
3	平均數	39	平均數	39
4	標準差	2.23606798	標準差	14.19507
5	變異數	5	變異數	201.5

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

○ 相對分散度

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

母體資料: $CV = \frac{\sigma}{\mu}$

樣本資料: $CV = \frac{S}{\bar{X}}$

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

表 6.11 兩種基金報酬率的平均數與標準差

基金類別	平均數 (%)	標準差 (%)
甲基金	11.32	9.63
乙基金	7.21	4.87

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

柴比氏定理

○ 柴比氏定理

不論資料為何種分配，至少有 $(1-1/k^2)$ 的資料落在距離平均數 k 個標準差的範圍內。 k 為大於 1 的任意數，即 $k > 1$ 。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

分散度的衡量

觀念與思考 如何檢查極端值

極端值是指與其他大部分的數值比較起來為極小或極大的數值，利用下列步驟可檢查是否有極端值。

步驟 1：將觀察值由小而大排列

步驟 2：計算出第 1 四分位數 Q_1 與第 3 四分位數

步驟 3：計算四分位距 $IQR = Q_3 - Q_1$ （四分位距是第 1 四分位數與第 3 四分位數的差距）

步驟 4：計算 $Q_1 - 1.5 \times IQR$ 及 $Q_3 + 1.5 \times IQR$

步驟 5：若觀察值 x 小於 $Q_1 - 1.5 \times IQR$ 或大於 $Q_3 + 1.5 \times IQR$ 則為極端值。

林惠玲 陳正倉著 雙葉書報發行 2008

第6章 分析資料-以統計測量數呈現 基礎統計學二版

EXCEL 的使用

圖6.7 四分位數的對話方塊

林惠玲 陳正倉著 雙葉書報發行 2008

EXCEL 的使用

表6.12 海之濱春季夏季的營業收入

	A	B	C
1	海之濱春季的營業收入		
2			
3	平均數	54	
4	中間值	54	中位數
5	眾數	45	
6	標準差	15.3438	
7	變異數	235.4333	
8	範圍	68	全距
9	最小值	21	
10	最大值	89	
11	個數	61	元素個數

林惠玲 陳正金著 雙葉書報發行 2008