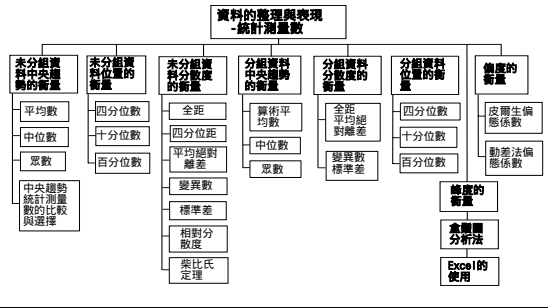


## 4 資料的整理與表現-統計測量數

### 學習目的

1. 了解資料中央趨勢的各種衡量指標如算術平均數、中位數、眾數、加權平均數與幾何平均數等的衡量方法。
2. 熟習各個中央趨勢衡量方法的特性、使用時機與優缺點。
3. 了解資料分散程度的各種衡量指標如全距、四分位距、變異數、標準差、變異係數及柴比氏定理的衡量方法。
4. 熟習各個分散程度衡量方法的特性、使用時機與優缺點。
5. 認識資料相對位置的各種衡量方法如四分位數、十分位數百分位數等的計算。認識與計算資料的偏度、峰度。
6. 了解兩變數間的關係與計算共變變數。
7. 熟習使用EXCEL計算中央趨勢與分散度指標及其他位置之指標。

### 本章結構



### 未分組資料中央趨勢的衡量

#### 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

$$\mu = \frac{\sum_{i=1}^N x_i}{N}$$

##### 樣本平均數

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

林惠玲 陳正倉等 經濟學導論 2000

### 未分組資料中央趨勢的衡量

#### 平均數

##### 加權算術平均數

$$\text{母體: } \mu_w = \frac{\sum_{i=1}^N W_i x_i}{\sum_{i=1}^N W_i} \quad \text{樣本: } \bar{X}_w = \frac{\sum_{i=1}^n W_i x_i}{\sum_{i=1}^n W_i}$$

##### 母體的幾何平均數

$$G = \sqrt[n]{x_1 \cdot x_2 \cdots x_N} = \left( \prod_{i=1}^N x_i \right)^{\frac{1}{N}}$$

##### 樣本的幾何平均數

$$g = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

林惠玲 陳正倉等 經濟學導論 2000

### 算術平均數的特質

- 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 各觀察值與平均數之差的平方和最小
- ④ 優點為考慮到每一個觀察值，缺點為易受極端值的影響。
- ⑤ 可進行代數演算
- ⑥ 可對觀察值予以加權

林惠玲 陳正倉等 經濟學導論 2000

表4.1 電子類股票的價格 (單位: 元)

	A	B	C	D	E	F	G	H	I
1	44.5	167	111	9.75	161	103	122	31.5	25.1
2	36.8	42.7	67	12.8	59	84.5	14.75	17.1	105.5
3	90.5	36.7	14.35	24.9	64.5	53	17.9	17	23.4
4	8	40.6	31	44.5	30.5	78.5	51	185	186
5	11.45	254	32	44.5	170	350	33.2	184	32.6
6	74	17.8	48.2	27.4	14.95	5.1	51	80.5	289
7	145.5	24.2	154	44.1	216	35.3	28	131.5	89
8	68.5	61.5	118	36.9	82	90	58.5	162	71.5
9	504	22	68.5	26.7	30.7	18.9	40.3	116	78
10	64	142	31	37.4	196.5	24.4	7.85	259	122.5
11	30.9	90							

圖4.1 電資料的平衡點

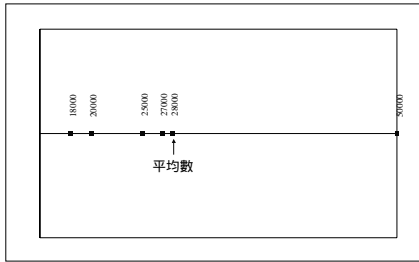


表4.2 電子類股票的價格

	A	B	C
1	電子類股票的平均價格		
2	平均數	75.8173913	
3	標準差	7.07748846	平均數的標準差
4	中間值	51.25	中位數
5	眾數	44.5	

表4.3 學生成績報告單 (加權平均數)

	A	B	C	D	E
1	科目	學分數	學期成績	加權成績	
2	國文	3	81	243	
3	英文	3	82	246	
4	哲學概論	3	83	249	
5	中國歷史	2	90	180	
6	西洋哲學史	2	89	178	
7	中國憲法與政治	2	86	172	
8	經濟學原理	3	78	234	
9	體育	1	75	75	
10	法學概論	2	82	164	
11			1756	83.61905	

表4.4 消費者物價指數(幾何平均數)

年別	消費者物價指數	變動率
80	83.60	
81	87.33	1.0446
82	89.90	1.0294
83	93.58	1.0409
84	97.02	1.0367
85	100.00	1.0307
86	100.90	1.0090

資料來源: "Taiwan Statistical Data Book 1998", 行政院經建會。

消費者物價指數變動率為  $CPI_t / CPI_{t-1}$

## 未分組資料中央趨勢的衡量

## ○ 中位數

## 中位數的意義

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。

## ○ 眾數

## 眾數的意義

眾數是指觀察值中其出現次數最多的那一個數值。

圖4.2 中位數

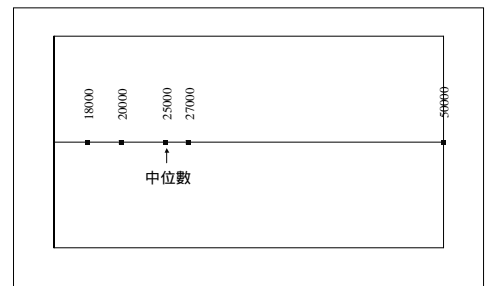


表4.5 中央趨勢統計測量數之比較

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. 觀察值與平均數差平方和最小 5. 適合統計推論的工作	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 4. 可做無母數統計推論	1. 不適合代數演算 2. 對觀察值敏感性低 3. 不易進行母數統計推論
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. 可能不止一個或不存在 2. 敏感性低 3. 不能做統計推論
幾何平均數	1. 適合等比資料 2. 敏感度高	1. 不適合一般資料 2. 不適合作統計推論

## 未分組資料位置的衡量(其他測量數)

## ○ 四分位數

四分位數是將順序資料分成四等份數值的分位數。

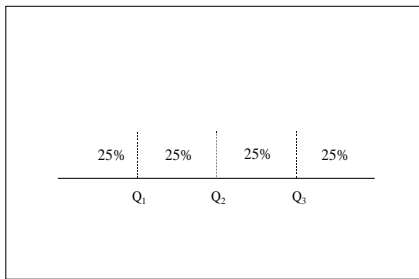
## ○ 十分位數

十分位數是將資料均分為十等份數值的分割數。

## ○ 百分位數

百分位數是將順序資料均分為一百等份數值的分割數。

圖4.6 四分位數的圖解



## 未分組資料分散度的衡量

## ○ 全距

$$R = \text{最大值} - \text{最小值}$$

## ○ 四分位距

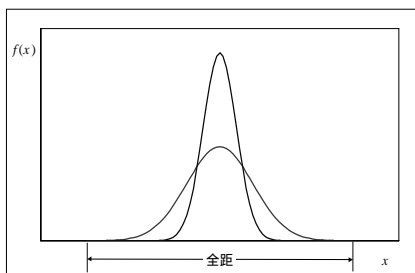
$$IQR = \text{第三四分位數} - \text{第一四分位數} = Q_3 - Q_1$$

## ○ 平均絕對離差

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

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

圖4.9 資料的分散情形



## 未分組資料分散度的衡量

## ○ 變異數

## 母體變異數

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

式中： $\mu$ ：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。

## 未分組資料分散度的衡量

## ○ 標準差

## 母體標準差

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

## 樣本標準差

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

## 未分組資料分散度的衡量

## ○ 相對分散度

## 變異係數

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

$$\text{母體資料: } CV = \frac{\sigma}{\mu}$$

$$\text{樣本資料: } CV = \frac{s}{\bar{X}}$$

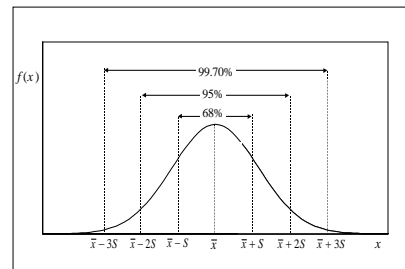
## ○ 柴比氏定理

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  $k > 1$ 。

表4.8 兩種股票的投資報酬率

	A	B	C	D
1	聯電		宏電	
2				
3	平均數	16.65	平均數	12.79
4	標準差	5.40307	標準差	3.320457
5	標準差	17.88681	標準差	10.50021
6	變異數	291.9317	變異數	110.2543

圖4.11 經驗法則



## 分組資料中央趨勢的衡量

## ○ 算術平均數

$$\text{母體均數: } \mu = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i} \quad \text{樣本均數: } \bar{X} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i}$$

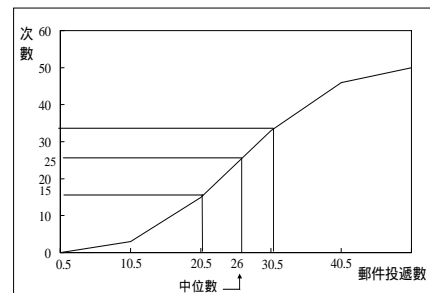
## ○ 中位數

$$m_e = L_{m_e} + W_{m_e} \left( \frac{\frac{n}{2} - F_L}{f_{m_e}} \right)$$

式中： $L_{m_e}$ ： $m_e$  所在組的組下界， $W_{m_e}$ ： $m_e$  所在組的組距，

$f_{m_e}$ ： $m_e$  所在組的組次數， $F_L$ ： $m_e$  前一組的累加次數。

圖4.12 郵件投遞中位數的圖解





## 分組資料中央趨勢的衡量

## ○ 眾數

## 粗略法眾數

$$m_0 = \frac{(\text{組上界} + \text{組下界})}{2}$$

## 金氏法眾數

$$m_0 = L_{m_0} + a = L_{m_0} + \frac{f_{-1}W_{m_0}}{f_{-1} + f_{+1}}$$

式中： $L_{m_0}$ ： $m_0$ 所在組的組下界， $f_{-1}$ ： $m_0$ 前一組的次數，  
 $f_{+1}$ ： $m_0$ 後一組的次數， $W_{m_0}$ ： $m_0$ 所在組的組矩。

## 皮爾生眾數

$$m_0 = \bar{X} - 3(\mu - m_e)$$



## 分組資料分散度的衡量

## ○ 全距

①  $R = \text{最後一組組中點} - \text{第一組組中點}$

②  $R = \text{最後一組組上限} - \text{第一組組下限}$

## ○ 平均絕對離差

$$\text{母體：} MAD = \frac{1}{n} \sum_{i=1}^k |x_i - \mu| f_i$$

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

$x_i$ ：組中點， $k$ 為組數， $f_i$ ：組次數。



## 分組資料分散度的衡量

## ○ 變異數與標準差

## 母體變異數與標準差

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k (x_i - \mu)^2 f_i \quad \sigma = \sqrt{\sigma^2}$$

式中： $x_i$ ：組中點， $f_i$ ：組次數， $N$ ：母體個數， $k$ ：組數。

## 樣本變異數與標準差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{X})^2 f_i \quad S = \sqrt{S^2}$$

式中： $x_i$ ：組中點， $f_i$ ：組次數， $n$ ：母體個數， $k$ ：組數。



## 分組資料位置(其他測量數)的衡量

## ○ 四分位數

## 第1四分位數

$$Q_1 = L_{Q_1} + \frac{\frac{1}{4}n - F_{Q_1}}{f_{Q_1}} W_{Q_1}$$

式中： $L_{Q_1}$ ： $Q_1$ 所在組的組下界， $f_{Q_1}$ ： $Q_1$ 所在組的組次數，  
 $W_{Q_1}$ ： $Q_1$ 所在組的組矩， $F_{Q_1}$ ： $Q_1$ 前一組的累加次數。

## 第3四分位數

$$Q_3 = L_{Q_3} + \frac{\frac{3}{4}n - F_{Q_3}}{f_{Q_3}} W_{Q_3}$$

式中： $L_{Q_3}$ ： $Q_3$ 所在組的組下界， $f_{Q_3}$ ： $Q_3$ 所在組的組次數，  
 $W_{Q_3}$ ： $Q_3$ 所在組的組矩， $F_{Q_3}$ ： $Q_3$ 前一組的累加次數。



## 分組資料位置(其他測量數)的衡量

## ○ 十分位數

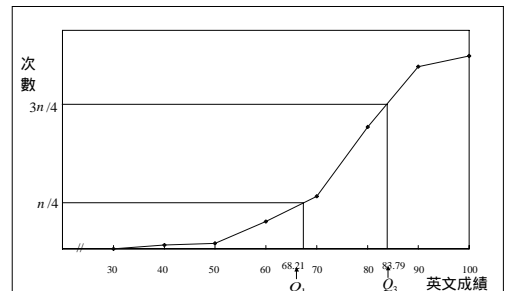
$$D_i = L_{D_i} + \frac{\frac{n \cdot i}{10} - F_{D_i}}{f_{D_i}} W_{D_i}$$

式中： $D_i$ ：第*i*個十分位數， $L_{D_i}$ ： $D_i$ 所在組的組下界， $f_{D_i}$ ： $D_i$ 所在組的組次數， $W_{D_i}$ ： $D_i$ 所在組的組矩， $F_{D_i}$ ： $D_i$ 前一組的累加次數。

## ○ 百分位數

$$P_i = L_{P_i} + \frac{\frac{n \cdot i}{100} - F_{P_i}}{f_{P_i}} W_{P_i}$$

式中： $P_i$ ：第*i*個百分位數， $L_{P_i}$ ： $P_i$ 所在組的組下界， $f_{P_i}$ ： $P_i$ 所在組的組次數， $W_{P_i}$ ： $P_i$ 所在組的組矩， $F_{P_i}$ ： $P_i$ 前一組的累加次數。

圖4.14 分組資料時  $Q_1$ 、 $Q_3$  的解法



## 偏度

## ○ 皮爾生偏度係數

$$\text{母體: } SK_p = \frac{3(\mu - M_e)}{\sigma}$$

$$\text{樣本: } SK_p = \frac{3(\bar{X} - m_e)}{S}$$



圖4.15 對稱的次數分配

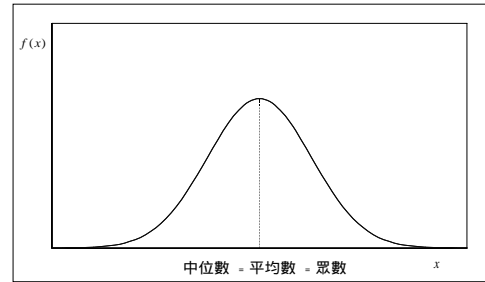


圖4.16 左偏的次數分配

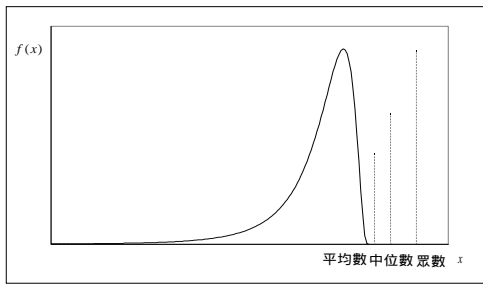
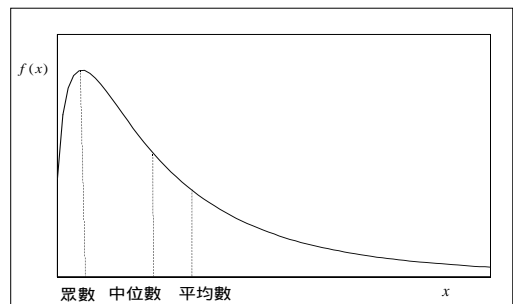


圖4.16 左偏的次數分配



## 兩變數間的關係

## ○ 共變數

## 母體共變數

$$COV(X, Y) = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{N}$$

## 樣本共變數

$$cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{n-1}$$



## 動差法的偏態係數

## ○ 動差法的偏態係數

$$\text{母體: } \alpha_3 = \frac{M_3}{\sigma^3} = \frac{M_3}{(\sqrt{M_2})^3}$$

$$\text{樣本: } \alpha_3 = \frac{m_3}{(\sqrt{m_2})^3}$$



兩變數間的關係

○ 相關係數

母體相關係數

$$\rho_{XY} = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

式中：COV(X,Y) 為母體共變數， $\sigma_X$ ， $\sigma_Y$  為標準差。

樣本相關係數

$$r_{XY} = \frac{cov(X,Y)}{S_X S_Y}$$

式中 cov(X,Y) 為樣本共變數， $S_X$ ， $S_Y$  為樣本標準差。



偏度

○ 皮爾生偏度係數

母體：  $SK_p = \frac{3(\mu - M_2)}{\sigma}$

樣本：  $SK_p = \frac{3(\bar{X} - m_2)}{S}$



動差法的偏態係數

○ 動差法的偏態係數

母體：  $\alpha_3 = \frac{M_3}{\sigma^3} = \frac{M_3}{(\sqrt{M_2})^3}$

樣本：  $\alpha_3 = \frac{m_3}{(\sqrt{m_2})^3}$



圖4.25 三種峰度的圖形

