第4章 分析資料–以統計測量數來呈現

學習目的

1.瞭解資料中心位置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. 瞭解資料分散程度的各種衡量指標如全距、四分位距、變異數、標準差、變異係數的衡量方法。
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. 認識資料等分位置的的各種衡量方法如四分位數、十分位數百分位數等的計算
6.熟習使用EXCEL計算中心位置與分散度指標及等分位置之指標。

含課本重點整理，惟仍應研讀課本之詳細內容

第4章 分析資料–以統計測量數來呈現

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

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

樣本平均數

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

第4章　分析資料–以統計測量數來呈現

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

第4章　分析資料–以統計測量數來呈現

中心位置的衡量

❍ 平均數

加權算術平均數

母體：$\mu_w = \dfrac{\sum_{i=1}^{N} W_i x_i}{\sum_{i=1}^{N} W_i}$　　　樣本：$\overline{X}_w = \dfrac{\sum_{i=1}^{n} W_i x_i}{\sum_{i=1}^{n} W_i}$

中心位置的衡量

- 樣本的幾何平均數

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

第4章　分析資料−以統計測量數來呈現

中心位置的衡量

- 幾何平均數的性質

① $\sqrt[n]{\prod_{i=1}^{n}(x_i / y_i)} = \sqrt[n]{\prod_{i=1}^{n} x_i} \Big/ \sqrt[n]{\prod_{i=1}^{n} y_i}$

例如 X 為國民所得，Y 為人口數，則 X/Y 為平均每人所得，要求算其平均成長率可以計算 X/Y 的幾何平均數，或分別計算 X 與 Y 的幾何平均數，再將兩個幾何平均數相除。

②適合衡量等比數列的中央位置，但不易進行統計推論。

現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

中心位置的衡量

- 幾何平均數的投資報酬率

$$\overline{G} = \left[(1+R_1)(1+R_2)\cdots(1+R_n)\right]^{\frac{1}{n}} - 1$$

第4章　分析資料–以統計測量數來呈現

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。

現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

第4章　分析資料–以統計測量數來呈現

中心位置的衡量

❍ 中位數的特質

① $\sum|x_i - m_e|$ 為 $\sum|x_i - A|$ 中之最小，亦即 $\sum|x_i - m_e| \le \sum|x_i - A|$（A 為任意數）。此乃意指一組觀察值中，若欲尋找一個代表值使觀察值與代表值的距離和為最小，則該代表值即為中位數。

②不受極端值的影響。中位數只是觀察值數列中的一個數值，因此當然不受極端值的影響，故對觀察值的變化不敏感。

③不易進行代數演算，亦不易進行統計推論。

第4章　分析資料–以統計測量數來呈現

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或類別。

第4章　分析資料–以統計測量數來呈現

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眾數因可能有多個或一個也沒有，因此眾數比中位數及平均數較少使用。

第4章　分析資料–以統計測量數來呈現

中心位置的衡量

表 4.4 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.資料的中心。資料無極端值或偏態時，具代表性。	1.若有極端值存在時，則不具代表性。
	2.適合代數演算。	2.資料如為偏態，則代表性較差。
	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。	
幾何平均數	1.適合等比資料。	1.不適合一般資料。
	2.敏感度高。	
中位數	1.適用於有極端值的資料。	1.不適合代數演算。
	2.適用於偏態資料。	2.對觀察值敏感性低。
	3.觀察值與中位數絕對差和最小。	
眾數	1.適用於有極端值的資料。	1.可能不只一個或一個也沒有。
	2.適用於偏態資料。	2.敏感性低。
	3.適用於類別資料。	

第4章　分析資料−以統計測量數來呈現

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。四分為數有第1、第2、第3三個寺分位數。

- **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分數值的分割數。

第4章　分析資料−以統計測量數來呈現

分散度的衡量

- **全距**

  $R=$ 最大值 $-$ 最小值

- **四分位距**

  $IQR=$ 第3四分位數 $-$ 第1四分位數 $=Q_3-Q_1$

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

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

式中：$\overline{X}$ ：樣本平均數，n ：樣本數。

第4章　分析資料−以統計測量數來呈現

分散度的衡量

- **變異數的性質**

①變異數的值大於等於 ，若變異數為 時，其意義是所有觀察值均相同，沒有變異（分散）。
②若同一組資料單位不同，其變異數亦不相同。
③單位相同可作比較
④考慮每一個觀察數值
⑤適合代數演算
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。

現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

第4章　分析資料−以統計測量數來呈現

分散度的衡量

- 標準差

母體標準差

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

樣本標準差

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

現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

第4章　分析資料−以統計測量數來呈現

分散度的衡量

- 相對分散度

變異係數

$$變異係數(CV) = \frac{標準差}{平均數}$$

$$母體資料：CV = \frac{\sigma}{\mu}$$

$$樣本資料：CV = \frac{S}{\overline{X}}$$

現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

第4章　分析資料–以統計測量數來呈現

柴比氏定理與經驗法則

❍**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$k>1$ 。

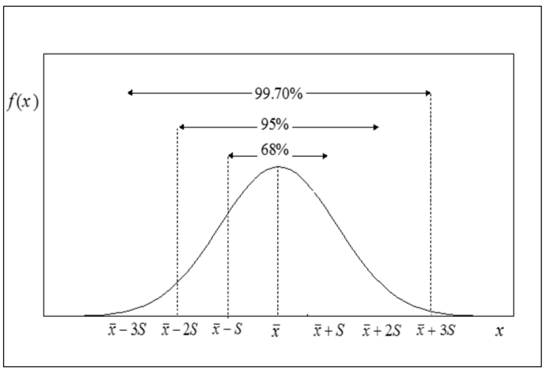
❍**經驗法則**

若資料為鐘形分配，則有68%的觀察值落在$\bar{X}\pm S$內，有95%的觀察值落在$\bar{X}\pm 2S$內，有99.7%的觀察值落在$\bar{X}\pm 3S$內（S為標準差）。

第4章　分析資料–以統計測量數來呈現

柴比氏定理與經驗法則

圖4.10　經驗法則



第4章　分析資料–以統計測量數來呈現

Z 值

- Z 值

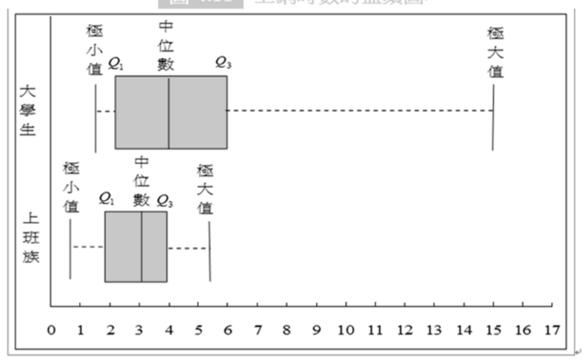
樣本 x 值的 Z 值：$\dfrac{x-\bar{x}}{s}$　，母體 X 值的 Z 值：$\dfrac{X-\mu}{\sigma}$

現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

第4章　分析資料–以統計測量數來呈現

盒鬚圖分析法（5數彙總）

圖 4.11 上網時數的盒鬚圖



現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

盒鬚圖分析法（5數彙總）

- 如何檢查極端值（outliers）

所謂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：計算出第一四分位數Q_1與第三四分位數Q_3

步驟 3：計算四分位距$IQR = Q_3 - Q_1$

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。

第4章　分析資料–以統計測量數來呈現

分組資料中心位置的衡量

- 算術平均數

母體均數：$\mu = \dfrac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$　　樣本均數：$\overline{X} = \dfrac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$

現代統計學 林惠玲 陳正倉 合著 雙葉書廊發行 2015

分組資料中心位置的衡量

- **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。

分組資料中心位置的衡量

- **眾數**

粗略法眾數

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

兩組數量資料相關性的衡量

- 母體相關係數

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- 樣本相關係數

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

4.7【是非題】

①平均數、中位數、眾數三者皆是用來測定一群資料分散度情況的統計測定數，而平均數總是優於中位數和眾數。

②當平均數為零時，標準差亦等於零，反之亦然。

③任何一組統計資料，可能有很多眾數，也可能沒有眾數。

④當所有的資料值均為負數時，平均數、眾數、中位數及變異數亦均是負值。

⑤若分配是單峰對稱分配，則平均數=眾數=中位數。

⑥若分配是單峰對稱分配，則算術平均數 ，幾何平均數 相等。

4.9 某電池生產商為測試其所生產的電池壽命，隨機選取20個電池，測驗結果如下（以小時為單位）：

41.0 40.9 39.8 42.6 41.4 42.5 42.4 40.8 42.5 42.8 39.7 41.8 42.2 42.5 42.6 40.0 41.0 42.4 42.7 43.6

請利用Excel計算下列各數值：

①算術平均數、眾數、中位數。

②第10及第90百分位數。

③全距、MAD（平均絕對離差）及四分位全距。

④變異數及變異係數。

第4章　分析資料–以統計測量數來呈現

4.13 設一組資料含有 $X_1, X_2, \cdots, X_n$ 共有 n 個，且 $\bar{X}$ 與 S_X^2 分別表示其平均數與變異數，則：

①設 $Y_i = X_i + k, i = 1,2,\ldots,n$，試問 $Y_1, Y_2, \ldots, Y_n$ 這 n 個數的平均數與變異數與原來的 $\bar{X}, S_X^2$ 有何關係？

②設 $Y_i = kX_i, i = 1,2,\ldots,n$，試問 $Y_1, Y_2, \ldots, Y_n$ 此 n 個數的平均數與變異數又與原來之 $\bar{X}, S_X^2$ 有何關係？

4.14 設台中市醫生和律師去年全年所得的分配情形如下（單位：萬元）：

職業	人數	平均所得	中位數	眾數	標準差
醫生	240	200	150	120	50
律師	160	150	120	100	50

①計算該市從事這二種職業的人的總平均所得。

②計算二種職業所得的標準差。

③哪一種職業的所得差異較大？

④各種職業所得之分配呈現何種型態？何者較為偏態？