三、敘述性統計：數量方法

Chapter 3 Descriptive Statistics: Numerical Methods

目錄

圖目錄

表目錄

學習目標

知識(認知)

1.可以描述各種敘述性數值表達方式適用的狀況。

2.分辨各種敘述性統計數值表達方式之間的差異性。

3.評價各種敘述性統計數值表達方式的使用價值。

技能

1.能夠計算各種敘述性統計之數值。

2.能夠計算探索性資料分析之五種數值。

3.能夠依循步驟，製作盒鬚圖。

態度(情意)

1.意識到各種敘述性統計數值表達方式的重要性。

2.可以依循需求判斷和選擇採用適當的敘述性統計數值表達方式。

3.建立自主判斷離群值的觀念。

敘述性統計學的目標是期望將眾多繁雜且毫無章法的觀測值(資料)，利用一個、兩個、三個等少數數值進行概述(summarization)，表達母體或樣本特定變數中大量觀測值(資料)的特徵和屬性。

整理與陳述原始觀測值(資料)，如統計圖、統計表。並不由已知資料推論到未知的領域。利用單一評量變數的原始觀察值(資料)計算特徵量(單變量)，如集中程度的測量、中央趨勢量、集中趨勢量數(measured of central tendency; measures of location)、離散程度的測量、分散數量、離勢量數(measured of dispersion; measures of dispersion)和形狀(shape)[偏態與峰態]。期望解釋和判讀代表特定變數大量原始觀測值之概述的象徵性數值。

當經營一家餐廳、旅行社或旅館時，如何具體地善用數值呈現消費端的年齡層分布、消費金額分布、消費時間分布、用餐同伴分布、在產品面的銷售分布等資料之中心位置、分散程度、偏態與峰度，讓主管、老闆或股東一目了然其數值化的程度高低。就需要使用敘述性統計中的數量呈現方法。

章節結構圖

3.1 中心位置測定值

瞭解原始資料分布中不同**中心位置測定值**、**中心位置量數**(Measures of Central Location)、集中趨勢量數(Measures of Central Tendency)或集中程度測量(Measures of Location)評量方式的意義、計算方法、使用時機、優缺點和彼此之間差異。

3.1.1 平均值

平均值、算術平均值(Arithmetic Mean)、均數、均值或平均數(mean)是代表觀測值(屬於區間尺度或比例尺度)分布集中化(center or middle)位置之數值(集中程度)。若觀測值是來自**母體**，平均值為 μ(讀音 mu; muw)符號代表；若觀測值是來自**樣本**，平均值為$\bar{x}$ (讀音 x bar)符號代表。適用於數值資料分布屬於左右對稱的情況，並且僅有一個波峰時。

符號	母體	樣本
平均值	μ	$\bar{x}$

3.1.1.1 未分組資料之平均值

母體之特定變數有 N 個觀測值，其數值分別標示為 $x_1, x_2, x_3,\ldots, x_N$，母體平均值(population mean) $\mu = \sum_{i=1}^{N}\frac{x_i}{N} = \frac{\sum_{i=1}^{N}x_i}{N}$

樣本之特定變數有 n 個觀測值，其數值分別標示為 $x_1, x_2, x_3,\ldots, x_n$，樣本平均值(sample mean) $\bar{x} = \sum_{i=1}^{n}\frac{x_i}{n} = \frac{\sum_{i=1}^{n}x_i}{n}$

其中：$\sum_{i=1}^{n}x_i = x_1 + x_2 + x_3 +\ldots+ x_n$

Σ：希臘大寫字母 sigma，意指加法運算。

Excel 操作函數

點選欲放置**算術平均值**的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):**AVERAGE**→確定→選擇欲計算算術平均值的資料區間→確定。

3.1.1.2 分組資料的平均值【選擇教材】

在分組資料中，無法得知所有觀測值 x_i 之實際數值，僅利用各組的**組中點(組中值)m_i** 代表各組的**觀測值** x_i。故，分組資料的算術平均值較未分組資料的算術平均值之**準確性較差**。

母體資料平均值 $\mu = \frac{\sum_{i=1}^{k}f_i\times m_i}{\sum_{i=1}^{k}f_i} = \frac{\sum_{i=1}^{k}f_i\times m_i}{N}$

樣本資料平均值$\bar{x} = \frac{\sum_{i=1}^{k}f_i\times m_i}{\sum_{i=1}^{k}f_i} = \frac{\sum_{i=1}^{k}f_i\times m_i}{n}$

其中，f_i 為第 i 組別中觀測值出現的次數，m_i 為第 i 組別的組中點(組中值)，k 為組數，$\sum_{i=1}^{k}f_i = N$ (母體中基本單位數量)，$\sum_{i=1}^{k}f_i = n$ (樣本中基本單位數量)。

若有開放組距時，建議取一般未開放組距之兩倍組距為開放組距之組距，再估算組中點(組中值) m_i。以**約略**(概略)計算算術平均值。

範例 3.1 調查抽樣樣本的年齡層分布如下表，試計算其平均值？(答案有效位數四捨五入後取到小數點後第 1 位)

年齡層	人數
20~29	12
30~39	26
40~49	33
50~59	26
60~69	21

題解：利用表格的製作過程，將計算公式運算過程納入，有系統的呈現計算過程，以提升讀者的學習認知。

年齡層	人數(f_i)	組中點(m_i)	$f_i \times m_i$
20~29	12	24.5	294.0

年齡層	人數(f_i)	組中點(m_i)	$f_i \times m_i$
30~39	26	34.5	897.0
40~49	33	44.5	1468.5
50~59	26	54.5	1417.0
60~69	21	64.5	1354.5
合計	118		5431.0

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i \times m_i}{n} = \frac{5431}{118} = 46.0$$

答案：年齡的算術平均值 46.0 歲

選項為 20~29 歲年齡層，有效的位數為十位數，取樣 118 位，有效位數往左移兩位，所以最後答案算術平均值的有效位數達小數點後一位。

有效數字(significant digits; significant figure)與準確度(accuracy)

在評量的過程中，使用評量工具上可以準確讀取的數字，稱為**精確數值**(certain digits)，例如：一般長度尺規上可以精確的讀到 cm 下一位，如 2.4 cm，因此精確數值即達小數點後第一位，若標示特定物體的長度為 2.45 cm 時，小數點後第二位 5 為**估計值**(uncertain digit)，另亦有一些評量過程不會產生估計值。一般情況下精確數值加上往後一位的估計值，即為有效數字，例如：2.45 cm 即有 3 位有效數字；若無估計值時，精確數值即等於有效數字。

精確度(precision)：任何測量值的精確度代表其測量時，所使用的最小單位，即最後一位有效數字的單位。365200 m 的精確度為 100 m；0.25 sec 的精確度為 0.01 sec；58.5350 kg 的精確度為 0.0001 kg。

Round-off rule

一般情況下，最後答案的所標示的小數點位數，會比原始(觀測值)數值資料多一位小數點位數。若需要更精細的數值時，最後答案所標示的小數點位數可以比原始數值資料多兩位小數點位數。

平均值有效位數：

分組資料算術平均值的另一種算法

$$\bar{x} = AM + \frac{\sum_{i=1}^{k} f_i \times d_i}{n} \times h$$

其中 AM：假設值，一般選擇分布次數最多那一組的組中點

n：總次數(母體或樣本基本單位的數量)　　h：組距

k：組數　　m_i：第 i 組別的組中點

f_i：第 i 組別觀測值出現的次數　　$d_i = \frac{m_i - AM}{h}$

範例 3.2　調查抽樣樣本的年齡層分布如下表，試計算其算術平均值？(答案有效位數四捨五入後取到小數點後第 1 位)

年齡層	人數
20~29	12
30~39	26
40~49	33
50~59	26
60~69	21

題解：

年齡層	人數(f_i)	組中點(m_i)	d_i	$f_i \times d_i$
20~29	12	24.5	-2	-24

年齡層	人數(f_i)	組中點(m_i)	d_i	$f_i \times d_i$
30~39	26	34.5	-1	-26
40~49	33	44.5(AM)	0	0
50~59	26	54.5	1	26
60~69	21	64.5	2	42
合計	118			18

組距 h = 10，組數 k = 5。

$$\bar{x} = \text{AM} + \frac{\sum_{i=1}^{k} f_i \times d_i}{n} \times h = 44.5 + \frac{18}{118} \times 10 = 44.5 + 1.5 = 46.0$$

答案：平均年齡為 46.0 歲

3.1.1.3 平均值之性質

a.平均值是所有觀測值分布的集中點(平衡點和等量點)，代表大於算術平均值的所有觀測值與算術平均值之差的總和，**等於**小於算術平均值的所有觀測值與算術平均值之差的總和。

$$\Sigma|x_{大} - \mu| = \Sigma|x_{小} - \mu|$$

範例 3.3　現有 4 個樣本觀測值數值分別為 2、3、6 和 9，計算其算術平均值$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{2+3+6+9}{4} = \frac{20}{4}$ = 5。

比樣本平均值$\bar{x}$ = 5 高的觀測值有 6 和 9，各觀測值與算術平均值之差的總和 = |6 – 5| + |9 – 5| = 5

比樣本平均值$\bar{x}$ = 5 低的觀測值有 2 和 3，各觀測值與算術平均值之差的總和 = |2 – 5| + |3 – 5| = 5

故比樣本平均值$\bar{x}$ = 5 高的觀測值有 6 和 9，各觀測值與算術平均值之差的總和 = 5 等於比樣本平均值$\bar{x}$ = 5 低的觀測值有 2 和 3，各觀測值與算術平均值之差的總和 = 5。

b.計算平均值時，每一個觀測點的數值皆有納入計算。每一個觀測值對算術平均值皆有貢獻，但是每一個觀測值對算術平均值的貢獻量不均等。當資料(觀測值)有極端值(extreme value)出現時，算術平均值就非常敏感而受此極端值影響。

c.平均值可以進行代數加減乘除運算。原資料之每一樣本點(觀測值) ± 一常數(constant)，其新資料之算術平均值等於原資料算術平均值 ± 此一常數。原資料之每一樣本點(觀測值) ×/÷ 一常數(constant)，其新資料之算術平均值等於原資料算術平均值 ×/÷ 此一常數。

(一)設原觀測值 x_i 與新觀測值 y_i 分別標示為 $x_1, x_2, x_3, ..., x_N$ 及 $y_1, y_2, y_3,...., y_N$。其中 $y_i = x_i + C$，前述 C：常數。

兩母體平均值的關係為 μ_y(調整後平均值) = μ_x(原始平均值) + C

$$\mu_y = \sum_{i=1}^{N} \frac{y_i}{N} = \sum_{i=1}^{N} \frac{x_i + C}{N} = \frac{\sum_{i=1}^{N} x_i + \sum_{i=1}^{N} C}{N} = \frac{\sum_{i=1}^{N} x_i + N \times C}{N} = \frac{\sum_{i=1}^{N} x_i + N \times C}{N} = \frac{\sum_{i=1}^{N} x_i}{N} + \frac{N \times C}{N} = \frac{\sum_{i=1}^{N} x_i}{N} + C = \mu_x + C$$

(二)設原觀測值 x_i 與新觀測值 y_i 分別標示為 $x_1, x_2, x_3, ..., x_N$ 及 $y_1, y_2, y_3,...., y_N$。其中 $y_i = C \times x_i$，前述 C：常數。

兩母體平均值的關係為 μ_y(調整後平均值) = $C \times \mu_x$(原始平均值)

$$\mu_y = \sum_{i=1}^{N} \frac{y_i}{N} = \sum_{i=1}^{N} \frac{C \times x_i}{N} = C \times \frac{\sum_{i=1}^{N} x_i}{N} = C \times \mu_x$$

範例 3.4　現有 4 個樣本觀測值數值分別為 2、3、6 和 9，計算其算術平均值$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{2+3+6+9}{4} = \frac{20}{4}$ = 5。驗證加一常數和乘一常數，其算術平均值也是加一常數和乘一常數。

題解：

原始觀測值	加一常數	新觀測值

	2	+10	12
	3	+10	13
	6	+10	16
	9	+10	19
平均值	5	+10	15

加上一常數後新觀測值的算術平均值$\bar{x}=\sum_{i=1}^{n}\frac{x_i}{n}=\frac{12+13+16+19}{4}=\frac{60}{4}=15$，驗證原始觀測值的算術平均值 5 + 10 = 15，可以獲得新觀測值的算術平均值。

	原始觀測值	乘一常數	新觀測值
	2	×10	20
	3	×10	30
	6	×10	60
	9	×10	90
平均值	5	×10	50

乘上一常數後新觀測值的算術平均值$\bar{x}=\sum_{i=1}^{n}\frac{x_i}{n}=\frac{20+30+60+90}{4}=\frac{200}{4}=50$，驗證原始觀測值的算術平均值 5 × 10 = 50，可以獲得新觀測值的算術平均值。

d.各觀測值(樣本點)與算術平均值之偏差**和**等於 **0**。不論是母體平均值 μ 或樣本平均值$\bar{x}$皆具有此特性。

$\sum_{i=1}^{N}(x_i-\mu)=0$ 和$\sum_{i=1}^{n}(x_i-\bar{x})=0$。

$$\sum_{i=1}^{N}(x_i-\mu)=\sum_{i=1}^{N}x_i-N\times\mu=\sum_{i=1}^{N}x_i-N\times\sum_{i=1}^{N}\frac{x_i}{N}=\sum_{i=1}^{N}x_i-\frac{N}{N}\times\sum_{i=1}^{N}x_i=\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}x_i=0$$

$$\sum_{i=1}^{n}(x_i-\bar{x})=\sum_{i=1}^{n}x_i-n\times\bar{x}=\sum_{i=1}^{n}x_i-n\times\sum_{i=1}^{n}\frac{x_i}{n}=\sum_{i=1}^{n}x_i-\frac{n}{n}\times\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}x_i=0$$

其中：N 為母體觀測值個數，n 樣本觀測值(樣本點)個數。

範例 3.5 現有 4 個樣本觀測值數值分別為 2、3、6 和 9，計算其算術平均值$\bar{x}=\sum_{i=1}^{n}\frac{x_i}{n}=\frac{2+3+6+9}{4}=\frac{20}{4}=5$。驗證以算術平均值為中心運算偏差和的數值為 0。

題解：偏差和$\sum_{i=1}^{n}(x_i-\bar{x})=(2-5)+(3-5)+(6-5)+(9-5)=0$

答案：驗證以算術平均值為中心運算偏差和的數值皆為 0

e.各觀測值(樣本點)與算術平均值之偏差平方和(sum of squares, SS)為**最小**數值。任何非算術平均值與各觀測值的偏差平方和皆會大於前述平方和數值。

平方和(sum of squares, SS) SS = $\sum_{i=1}^{n}(x_i-\bar{x})^2$ SS = $\sum_{i=1}^{N}(x_i-\mu)^2$

$\sum_{i=1}^{n}(x_i-\bar{x})^2 \leq \sum_{i=1}^{n}(x_i-\mathrm{A})^2$，其中 A：可以屬於任何數值，此關係式皆成立。

範例 3.6 現有 4 個樣本觀測值數值分別為 2、3、6 和 9，計算其算術平均值$\bar{x}=\sum_{i=1}^{n}\frac{x_i}{n}=\frac{2+3+6+9}{4}=\frac{20}{4}=5$。驗證以算術平均值為中心運算偏差平方和的數值為最小數值。

題解：偏差平方和(sum of squares, SS) SS = $\sum_{i=1}^{n}(x_i-\bar{x})^2=(2-5)^2+(3-5)^2+(6-5)^2+(9-5)^2=9+4+1+16=30$

$$\sum_{i=1}^{n}(x_i-3)^2=(2-3)^2+(3-3)^2+(6-3)^2+(9-3)^2=1+0+9+36=46$$
$$\sum_{i=1}^{n}(x_i-4)^2=(2-4)^2+(3-4)^2+(6-4)^2+(9-4)^2=4+1+4+25=34$$
$$\sum_{i=1}^{n}(x_i-6)^2=(2-6)^2+(3-6)^2+(6-6)^2+(9-6)^2=16+9+0+9=34$$
$$\sum_{i=1}^{n}(x_i-7)^2=(2-7)^2+(3-7)^2+(6-7)^2+(9-7)^2=25+16+1+4=46$$

答案：在原來平均值的中心點位置分別帶入數值 3、4、6 和 7 運算偏差平方和發現，非平均值的數值當中心點時，運算出的來偏差平方和數值，皆比以平均值運算出來的偏差平方和 30 數值高。故，以平均值為中心點運算出來的偏差平方和數值最小。

f.兩個隨機變數相加時，相加後的隨機變數平均值為原始兩個隨機變數的平均值相加。

X 是一個隨機變數，其平均值為 μ_x，標準(偏)差為 σ_x；Y 為另一個隨機變數，其平均數為 μ_y，標準(偏)差為 σ_y。兩個隨機變數相加獲得的新變數 $Z = X + Y$ 的平均值為 $\mu_x + \mu_y$。

範例 3.7　若 X 隨機變數有三個基本單位其觀測值分別為 1、2 和 3，平均值 2，變異數為 0.6667；Y 隨機變數有三個基本單位其觀測值分別為 2、4 和 6，平均值 4，變異數為 2.6667，請計算隨機變數 X 加上隨機變數 Y 形成的新隨機變數 Z 的平均值。

題解：隨機變數 X 有三個基本單位觀測值加上隨機變數 Y 有三個基本單位觀測值，形成新的隨機變數 Z 一共有 9 個基本單位觀測值，分別為 3、4、5、5、6、7、7、8 和 9，平均值 6，變異數 3.3333。

依據隨機變數之平均值加法定理

$$\mu_{z=x+y} = \mu_x + \mu_y = 2 + 4 = 6$$

即代表隨機找 X 隨機變數的一個觀測值加上隨機找 Y 隨機變數的一個觀測值，合計後，新形成的分布(Z 隨機變數)，其平均值為 6。由此，可以驗證隨機變數之平均值加法定理。

答案：相加後新隨機變數 Z 的平均值 6

3.1.2 中位數

一般以 Me、M_e、Md 或 M_d 符號表示數值資料的中位數或中量(Median)。中位數即是一般白話中的中間值。適用於有極端值(離群值)出現時，而產生數值資料分布傾斜的情況，可以具體的表達出數值資料的中間特徵。

3.1.2.1 未分組數值的中位數

將原始的觀測值依據大到小(或小到大)排序(ranks)，列於最中間序位的觀測值之數值。中位數 = 第 2 個四分位數 = 第 5 個十分位數 = 第 50 個百分位數。中位數適用於順序尺度、區間尺度和比例尺度。代表有 50 %數量基本單位的觀測值比中位數小；亦另有 50 %數量基本單位的觀測值比中位數大。

設一樣本有 n 個樣本數值(觀測值)，依其大小由小排到大(或由大排到小)，中位數定義：

(1)若樣本數量 n 為奇數(odd)，中位數即為第$\frac{n+1}{2}$序位的觀測值。

(2)若樣本數量 n 為偶數(even)，中位數為第$\frac{n}{2}$和$\frac{n}{2}+1$兩序位的觀測值之算術平均值。

若資料量 n 很大，中位數之決定就無法如上述方式求得。而採用百分比的方式決定中位數。例如：公教人員每月收入少於四萬元者佔全體公教人員的 50 %，而大於四萬元者亦佔有 50 %，中位數即為四萬元。

優點：不因觀測值中有極大值或極小值而改變中位數的大小。中位數與各觀測值數值之差絕對值的總和最小，即$\sum_{i=1}^{n}|x_i - M_e|$的數值最小。最適用於順序尺度數值型態之變數。

缺點：中位數僅能決定樣本資料的中間值，而無法反映其他樣本點(觀測值)的真正數值。即其他觀測值的數值之高低對中位數沒有任何影響。

範例 3.8　阿文海產店上週一至日，來店消費人數依序排列為：127、150、162、158、166、210、220 人，求阿文海產店上週來店消費人數分布的中位數？

題解：來店消費人數大小排序：

序位	1	2	3	4	5	6	7
消費人數	220	210	166	162	158	150	127

樣本數量 $n = 7$，屬奇數，$\frac{n+1}{2} = \frac{7+1}{2} = 4$，故中位數為第 4 序位之觀測值 162 人

答案：中位數 162 人

中位數與各觀測值數值之差絕對值的總和最小，即$\sum_{i=1}^{n}|x_i - M_e|$的數值最小。驗證。

序位	1	2	3	4	5	6	7	合計
消費人數 x_i	220	210	166	162	158	150	127	
x_i - M_e	220-162	210-162	166-162	162-162	158-162	150-162	127-162	
x_i - M_e	58	48	4	0	-4	-12	-35	
$\|x_i$ - $M_e\|$	58	48	4	0	4	12	35	161
x_i - 161	220-161	210-161	166-161	162-161	158-161	150-161	127-161	
x_i – 161	59	49	5	1	-3	-11	-34	
$\|x_i$ - 161$\|$	59	49	5	1	3	11	34	162
x_i – 163	220-163	210-163	166-163	162-163	158-163	150-163	127-163	
x_i – 163	57	47	3	-1	-5	-13	-36	
$\|x_i$ - 163$\|$	57	47	3	1	5	13	36	162

前述表格第三列數值是各個觀測值-中位數(162)，第四列數值是運算後數值，第五列數值是第四列數值取絕對值後的數值，合計$\sum_{i=1}^{n}|x_i - M_e|$就是第五列最後一欄數值 161。故$\sum_{i=1}^{n}|x_i - M_e| = 161$。

前述表格第六列數值是各個觀測值-非中位數的隨意值 161，第七列數值是運算後數值，第八列數值是第七列數值取絕對值後的數值，合計$\sum_{i=1}^{n}|x_i - 161|$就是第八列最後一欄數值 162。故$\sum_{i=1}^{n}|x_i - 161| = 162$。

前述表格第九列數值是各個觀測值-非中位數的隨意值 163，第十列數值是運算後數值，第十一列數值是第十列數值取絕對值後的數值，合計$\sum_{i=1}^{n}|x_i - 163|$就是第十一列最後一欄數值 162。故$\sum_{i=1}^{n}|x_i - 163| = 162$。

故，$\sum_{i=1}^{n}|x_i - M_e| = 161$ 與$\sum_{i=1}^{n}|x_i - 161| = 162$ 和$\sum_{i=1}^{n}|x_i - 163| = 162$ 相比是最小的數值，因此，驗證【中位數與各觀測值數值之差絕對值的總和最小，即$\sum_{i=1}^{n}|x_i - M_e|$的數值最小。】。

範例 3.9　阿文海產店過去十天，來店消費人數依序排列為：127、150、162、158、166、210、220、250、159、215 人，求阿文海產店過去十天來店消費人數分布的中位數？(答案有效位數四捨五入取到個位數)

題解：來店消費人數大小排序：

序位	1	2	3	4	5	6	7	8	9	10
消費人數	250	220	215	210	166	162	159	158	150	127

樣本數量 $n = 10$，屬偶數，序位$\frac{n}{2} = \frac{10}{2} = 5$，序位$\frac{n}{2} + 1 = 5 + 1 = 6$。

故中位數為第 5 和第 6 序位觀測值的算術平均值$\frac{166+162}{2} =$ 之 164 人

答案：中位數 164 人

練習 *3.1*　請計算右側觀測值分布的中位數(median) 50, 50, 57, 79, 88, 89, 91 與 95.

Excel 操作函數

點選欲放置**中位數**的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):**MEDIAN**→確定→選擇欲計算中位數的資料區間→確定。

SPSS 操作指令

Analyze(分析)→選擇 Descriptive Statistics(描述性統計)→選擇 Frequencies…(次數分配表)→在 Frequencies 視窗中，將欲統計 Median 數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Frequencies 視窗下方之 Statistics…(統計量)按鈕→在 Frequencies: Statistics 次視窗右上角 Central

Tendency(集中趨勢)內勾選 Median(中位數)，按右邊 Continue(繼續)按鈕→回到 Frequencies 視窗，按 OK(確定)按鈕即會跑出統計結果

Frequencies
Statistics

VAR00001

N	Valid	7
	Missing	0
Median		162.0000

3.1.2.2 分組數值的中位數【選擇教材】

a.計算以下累積次數(頻率)

b.求中位數 M_e 在次數分布表中的組別**位置**，即 $O(M_e) = \frac{n}{2}$

c.由累積次數分布表，計算 M_e 位於哪一組別

d.由 M_e 所在組別中，利用下列公式，計算中位數之數值

$$M_e = L_i + (\frac{n}{2} - F) \times \frac{h_i}{f_i}$$

其中 L_i：中位數所在組別的真正下限值(組下界)

n：總次數 = 樣本數或母體數

F：**小於**中位數所在組別的各組次數和(頻率和、累加次數)

h_i：中位數所在組別的組距

f_i：中位數所在組別觀測值出現的次數(頻率)

範例 3.10　嘿嘿地區 142 家餐廳在上個月份，販售牛排的銷售量分布表，求所有餐廳販售牛排分布的中位數。(答案有效位數四捨五入取到小數點後第 1 位)

牛排銷售量	餐廳家數	以下累積次數(F_i)	
11~15	7	7	
16~20	9	16	
21~25	11	27	
26~30	15	42	
31~35	20	62	F
36~40	22	84	中位數組別
41~45	19	103	
46~50	18	121	
51~55	16	137	
56~60	5	142	

題解：由總累積次數 142，估算中位數所在組別$\frac{n}{2} = \frac{142}{2} = 71$

$$M_e = L_i + (\frac{n}{2} - F) \times \frac{h_i}{f_i} = 35.5 + (\frac{142}{2} - 62) \times \frac{5}{22} = 37.5 \text{ 套(個)}$$

答案：37.5 套(個)

練習 3.2　若某學術文獻指出特定綠色旅館 2015 年的消費者住宿期間每日之中位支出為 NT$ 4010 元。請使用白話方式解釋何謂「中位支出」。以單獨 Microsoft word 電子檔案繳交至數位學習平台，作業名稱：中位支出，Word 檔案主檔名：姓名學號。繳交截止日期時間：以平台設定為準。

3.1.3 眾數

眾數或眾量(Mode)為一組資料中發生觀測值出現次數最多的觀測值之數值或項目，標示為 Mo 或 M_o。眾數計算簡便，且不受極端觀測(數)值的影響，但容易受抽樣變動影響與組距變動影響。觀測值分布中可能有數個眾數，亦有可能完全沒有眾數。眾數受觀測值的多寡與其數值的高低影響不敏銳。

眾數是類別變數(屬於名目尺度)較典型的統計方式。眾數一般數學性質較差，很少被採用。適用於每一個可能觀測值皆會出現很多次數的情況。眾數適用於類別變數(屬於名目尺度)，有兩個或三個以上類別時，僅能夠使用眾數表達該變數資料的分布特性。

範例 3.11 依據某次調查資料顯示前往參加高雄愛河的 20 位日本遊客，其年齡資料分布如下表，求日本遊客年齡的眾數？

18	25	26	25	31	56	62	38	25	26
18	19	20	51	33	45	43	25	29	30

題解：

18	25	26	25	31	56	62	38	25	26
18	19	20	51	33	45	43	25	29	30

由上表資料顯示，其中日本遊客年齡(變數)為 25 歲(觀測值)者有 4 位(觀測值出現次數)，出現次數最多，因此日本遊客年齡的眾數為 25 歲，不是 4；4 為眾數之觀測值出現的次數。

答案：日本遊客年齡的眾數為 25 歲

範例 3.12 依據某次調查資料顯示前往參加屏東燈會的遊客，主要使用的交通工具分布如下表，求使用主要交通工具的眾數？

題解：

交通工具	飛機	火車	遊覽車	計程車	汽車	機車	公車	步行	其他
人數	150	220	215	210	166	262	158	150	12

由上表資料顯示，其中使用機車(觀測值)為交通工具(變數)的人數 262(觀測值出現次數)最多，因此眾數為機車，不是 262；262 為眾數之觀測值出現的次數(人數)。

答案：主要交通工具的眾數為【機車】

分組資料計算眾數

Excel 操作函數

點選欲放置**眾數**的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):**MODE**→確定→選擇欲計算眾數的資料區間→確定。

SPSS 操作指令

Analyze(分析)→選擇 Descriptive Statistics(描述性統計)→選擇 Frequencies…(次數分配表)→在 Frequencies 視窗中，將欲統計 Quartiles 數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Frequencies 視窗下方之 Statistics…(統計量)按鈕→在 Frequencies: Statistics 次視窗右上方 Central Tendency(集中趨勢)內勾選 Mode(眾數)，按右邊 Continue(繼續)按鈕→回到 Frequencies 視窗，按 OK(確定)按鈕即會跑出統計結果

Sample

Frequencies

Statistics

VAR00001

N	Valid	19
	Missing	0
Mode		8.00

中心位置測定值應該採用哪一種的概述方法

	平均值	中位數	眾數
名目尺度			●
順序尺度		●	●
區間尺度 比例尺度	●	●	●

練習 *3.3* 請計算下列樣本資料分布的平均值、中位數和眾數：

(a) 16, 18, 78, 35, 54, 35, 64, 56, 51, 33, 52, 66

(b) 35, 56, 26, 56, 55, 75, 54, 57, 55, 55, 35, 61, 75, 65, 39

練習 *3.4* 台灣 2008 年每月失業人數(單位：千人)如下表所示：

1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
411	424	417	412	416	428	442	452	464	476	507	549

(a)失業人數分布的算術平均值、中位數和眾數？

(b)畢業季節後五個月(6, 7, 8, 9, 10 月)失業人數分布的算術平均值、中位數和眾數？

3.1.4 四分位數

將觀測值數值資料由小排至大，再依據觀測值數量 n 均分成四等分。位於第一個等分位置的數值，即第 1(個)四分位數(first quartile)，利用符號 Q_1 標示之，又稱下四分位數，表示有 25 %基本單位數量的觀測值(資料)比 Q_1 小，另有 75 %基本單位數量的觀測值(資料)比 Q_1 大；位於第二個等分位置的數值，為第 2(個)四分位數(second quartile)，利用符號 Q_2 標示之，即等於中位數 Me，表示有 50 %基本單位數量的觀測值(資料)比 Q_2 小，另有 50 %基本單位數量的觀測值(資料)比 Q_2 大；位於第三個等分位置的數值，稱為第 3(個)四分位數(third quartile)，利用符號 Q_3 標示之，又稱上四分位數，表示有 75 %基本單位數量的觀測值(資料)比 Q_3 小，另有 25 %基本單位數量的觀測值(資料)比 Q_3 大。

中位數 M_e 亦類似四分位數，將次數分布分割為兩等份；四分位數是將次數分布分割為四等份。一般四分位數值(Q_1 到 Q_3)不受極端觀測(數)值的影響。

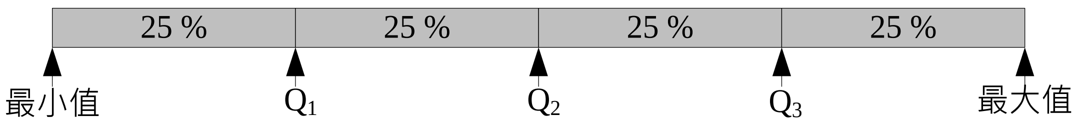


圖 3-1 四分位數位置分布

3.1.4.1 未分組資料的四分位數

四分位數區分為兩種計算方式，第一種為 n - 1 模式，第二種為 n + 1 模式。兩端點定義有些差異，必須區分清楚。

四分位數 *n* - 1 模式：

在 n - 1 模式(即為 Excel 中 QUARTILE.INC/QUARTILE 函數計算方式)中明確定義 Q_0 為最小觀測值之數值；Q_4 為最大觀測值之數值。即代表所有觀測值是以蒐集到的最大觀測值和最小觀測值為上下限，依據觀測值小到大將基本單位平均分配為四等份。

a.在樣本數量 n 個觀測值中，計算 Q_i 序位(Order number)：$O(Q_i) = (i \times \frac{n-1}{4}) + 1$，其中 $i = 1, 2$ 或 3。

b.若 $O(Q_i) = (i \times \frac{n-1}{4}) + 1$ 序位數值為整數，此序位所對應觀測值即為 Q_i；若 $O(Q_i) = (i \times \frac{n-1}{4}) + 1$ 序位數值不為整數，在 $O(Q_i)$整數與 $O(Q_i)$整數+1 兩個序位所對應的兩個觀測值，可利用線性內插法計算前述兩個觀測值的比例平均值，即為 Q_i 數值。

範例 3.13 阿力餐廳菜單中美味套餐，過去 9 個月之月銷售數量經小至大排序後為 32、37、38、41、45、46、47、50 和 51 套，求 Q_1、Q_2 及 Q_3。(答案有效位數四捨五入取到個位數)

題解：

$O(Q_{i=1}) = (i \times \frac{n-1}{4}) + 1 = (1 \times \frac{9-1}{4}) + 1 = 3$ 序位，Q_1 落於第 3 序位的觀測值，$Q_1 = 38$ 套

$O(Q_{i=2}) = (i \times \frac{n-1}{4}) + 1 = (2 \times \frac{9-1}{4}) + 1 = 5$ 序位，Q_2 落於第 5 序位的觀測值，$Q_2 = 45$ 套

$O(Q_{i=3}) = (i \times \frac{n-1}{4}) + 1 = (3 \times \frac{9-1}{4}) + 1 = 7$ 序位，Q_3 落於第 7 序位的觀測值，$Q_3 = 47$ 套

答案：$Q_1 = 38$ 套；$Q_2 = 45$ 套；$Q_3 = 47$ 套

內插法

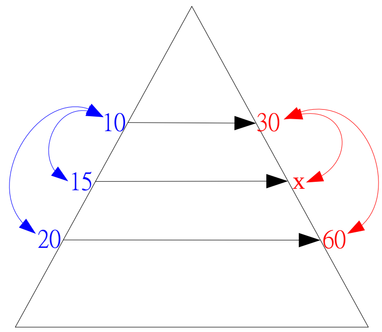


圖 3-2 內插法說明圖

透過上方的三角形，可以協助內插法數值的計算。利用已知的 10→30 和 20→60 的關係，推估 15→x 的未知數值。

$\frac{20-10}{15-10} = \frac{60-30}{x-30}$、$\frac{20-10}{20-15} = \frac{60-30}{60-x}$ 或 $\frac{15-10}{20-15} = \frac{x-30}{60-x}$皆可以推算出內插數值 x。

Excel 操作函數

點選欲放置四分位數的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):QUARTILE→確定→選擇欲計算四分位數的資料區間和第幾四分位數→確定。

範例 3.14 阿力餐廳菜單中美味套餐，過去 10 個月之月銷售數量經小至大排序後為 32、37、38、41、45、46、47、50、51、52 套，求 Q_1、Q_2 及 Q_3。(答案有效位數四捨五入取到小數點後第 2 位)

題解：

$O(Q_{i=1}) = (i\times \frac{n-1}{4}) + 1 = (1 \times \frac{10-1}{4}) + 1 = 3.25$ 序位，Q_1 落於第 3 與 4 序位觀測值之間取其兩者的比例平均值，$Q_1 = 38 + 0.25 \times (41 - 38) = 38.75$ 套

$O(Q_{i=2}) = (i \times \frac{n-1}{4}) + 1 = (2 \times \frac{10-1}{4}) + 1 = 5.50$ 序位，Q_2 落於第 5 與 6 序位觀測值之間取其兩者的比例平均值，$Q_2 = 45 + 0.50 \times (46 - 45) = 45.50$ 套

$O(Q_{i=3}) = (i \times \frac{n-1}{4}) + 1 = (3 \times \frac{10-1}{4}) + 1 = 7.75$ 序位，Q_3 落於第 7 與 8 序位觀測值之間取其兩者的比例平均值，$Q_3 = 47 + 0.75 \times (50 - 47) = 49.25$ 套

答案：$Q_1 = 38.75$ 套；$Q_2 = 45.50$ 套；$Q_3 = 49.25$ 套

3	3.25	4
38	Q_1	41

$$\frac{4-3}{41-38} = \frac{3.25-3}{Q_1-38}$$

練習 3.5　小玟咖啡館販售美式咖啡，過去 12 個月之月銷售數量(杯)依序分別如下，求 Q_1、Q_2 及 Q_3。(答案有效位數四捨五入取到小數點後第 2 位)

352　365　526　655　256　556　265　459　653　265　753　764

題解：依據銷售數量數值大小，由小排到大

序號	1	2	3	4	5	6	7	8	9	10	11	12
數值	256	265	265	352	365	459	526	556	653	655	753	764

$O(Q_{i=1}) = (i\times\frac{n-1}{4}) + 1 = (1\times\frac{12-1}{4}) + 1 = 3.75$ 序位，Q_1 落於第 3 與 4 序位觀測值之間取其兩者的比例平均值，$Q_1 = 265 + 0.75 \times (352 - 265) = 330.25$ 套

$O(Q_{i=2}) = (i\times\frac{n-1}{4}) + 1 = (2\times\frac{12-1}{4}) + 1 = 6.50$ 序位，Q_2 落於第 6 與 7 序位觀測值之間取其兩者的比例平均值，$Q_2 = 459 + 0.50 \times (526 - 459) = 492.50$ 套

$O(Q_{i=3}) = (i\times\frac{n-1}{4}) + 1 = (3\times\frac{12-1}{4}) + 1 = 9.25$ 序位，Q_3 落於第 9 與 10 序位觀測值之間取其兩者的比例平均值，$Q_3 = 653 + 0.25 \times (655 - 653) = 653.50$ 套

答案：$Q_1 = 330.25$ 杯；$Q_2 = 492.50$ 杯；$Q_3 = 653.50$ 杯

四分位數 $n + 1$ 模式：

在 $n + 1$ 模式(即為 Excel 中 QUARTILE.EXC 函數和 SPSS 軟體計算方式)中，沒有定義 Q_0 為最小觀測值之數值和 Q_4 為最大觀測值之數值。即代表不是以蒐集到的最大觀測值和最小觀測值為上下限，依據觀測值小到大將基本單位平均分配為四等份。【選擇教材】

a.在樣本數量 n 個觀測值中，計算 Q_i 序位(Order number)：$O(Q_i) = i\times\frac{n+1}{4}$，其中 $i = 1, 2$ 或 3。

b.若 $O(Q_i) = i\times\frac{n+1}{4}$ 序位數值為整數，此序位所對應觀測值即為 Q_i；若 $O(Q_i) = i\times\frac{n+1}{4}$ 序位數值不為整數，在 $O(Q_i)$ 整數與 $O(Q_i)$ 整數+1 兩個序位所對應的兩個觀測值，可利用線性內插法計算前述兩個觀測值的比例平均值，即為 Q_i 數值。

範例 3.15　阿力餐廳菜單中美味套餐，過去 9 個月之月銷售數量經小至大排序後為 32、37、38、41、45、46、47、50 和 51 套，求 Q_1、Q_2 及 Q_3。(答案有效位數四捨五入取到個位數)

題解：

$O(Q_{i=1}) = i\times\frac{n+1}{4} = 1\times\frac{9+1}{4} = 2.5$ 序位，Q_1 落於第 2 和 3 序位觀測值之間取其兩者的比例平均值，$Q_1 = 37 + 0.5 \times (38 - 37) = 37.5$ 套

$O(Q_{i=2}) = i\times\frac{n+1}{4} = 2\times\frac{9+1}{4} = 5$ 序位，Q_2 落於第 5 序位的觀測值，$Q_2 = 45$ 套

$O(Q_{i=3}) = i\times\frac{n+1}{4} = 3\times\frac{9+1}{4} = 7.5$ 序位，Q_3 落於第 7 和 8 序位觀測值之間取其兩者的比例平均值，$Q_3 = 47 + 0.5 \times (50 - 47) = 48.5$ 套

答案：$Q_1 = 37.5$ 套；$Q_2 = 45$ 套；$Q_3 = 48.5$ 套

SPSS 操作指令

Analyze(分析)→選擇 Descriptive Statistics(描述性統計)→選擇 Frequencies…(次數分配表)→在 Frequencies 視窗中，將欲統計 Quartiles 數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Frequencies 視窗下方之 Statistics…(統計量)按鈕→在 Frequencies: Statistics 次視窗左上角 Percentile

Values(百分位數值)內勾選 Quartile(四分位數)，按右邊 Continue(繼續)按鈕→回到 Frequencies 視窗，按 OK(確定)按鈕即會跑出統計結果。

Frequencies

Statistics

VAR00001

N	Valid	11
	Missing	0
Percentiles	25	38.0000
	50	46.0000
	75	51.0000

3.1.4.2 分組資料的四分位數【選擇教材】

a.計算以下累積次數(頻率)

b.計算 Q_i 序位：$O(Q_i) = (i \times \frac{n-1}{4}) + 1$ 序位，其中 $i = 1, 2$ 或 3

c.由累積次數分配，決定 Q_i 位於那個組別

d.Q_i 所在組別中，利用下列公式，計算 Q_i 數值

$$Q_i = L_i + [(i \times \frac{n-1}{4}) + 1 - F] \times \frac{h_i}{f_i}$$

其中 L_i：Q_i 所在組別的真正下限數值或組下界

n：總次數，母體或樣本總數

F：小於 Q_i 所在組別的各組次數和(頻率和)、累計次數

h_i：Q_i 所在組別的組距

f_i：Q_i 所在組別的次數(頻率)

範例 3.16 嘿嘿地區 142 家餐廳在上個月份，販售牛排的銷售量分布表，求所有餐廳販售牛排分布的 Q_1、Q_2 和 Q_3。(答案有效位數四捨五入取到小數點後第 2 位)

牛排銷售量	餐廳家數	以下累積次數
11~15	7	7
16~20	9	16
21~25	11	27
26~30	15	42
31~35	20	62
36~40	22	84
41~45	19	103
46~50	18	121
51~55	16	137
56~60	5	142

題解：

$O(Q_{i=1}) = (i \times \frac{n-1}{4}) + 1 = (1 \times \frac{142-1}{4}) + 1 = 36.25$ 序位，Q_1 位於 26~30 組別

$Q_1 = L_i + [(i \times \frac{n-1}{4}) + 1 - F] \times \frac{h_i}{f_i} = 25.5 + [(1 \times \frac{142-1}{4}) + 1 - 27] \times \frac{5}{15} = 25.50 + 3.08 = 28.58$ 套

$O(Q_{i=2}) = (i \times \frac{n-1}{4}) + 1 = (2 \times \frac{142-1}{4})+1 = 71.5$ 序位，Q_2 位於 36~40 組別

$Q_2 = L_i + [(i \times \frac{n-1}{4})+1 - F] \times \frac{h_i}{f_i} = 35.5 + [(2 \times \frac{142-1}{4}) + 1 - 62] \times \frac{5}{22} = 35.50 + 2.16 = 37.66$ 套

$O(Q_{i=3}) = (i \times \frac{n-1}{4})+1 = (3 \times \frac{142-1}{4}) + 1 = 106.75$ 序位，Q_3 位於 46~50 組別

$Q_3 = L_i + [(i \times \frac{n-1}{4}) + 1 - F] \times \frac{h_i}{f_i} = 45.5 + [(3 \times \frac{142-1}{4}) + 1 - 103] \times \frac{5}{18} = 45.50 + 1.04 = 46.54$ 套

答案：$Q_1 = 28.58$ 套；$Q_2 = 37.66$ 套；$Q_3 = 46.54$ 套

練習 3.6 高雄所有咖啡館在上個星期，販售卡布奇諾(Cappuccino)咖啡數量(杯數)次數分布表，求咖啡館販售卡布奇諾(Cappuccino)咖啡數量分布的 Q_1、Q_2 和 Q_3。(答案有效位數四捨五入取到小數點後第 2 位)

銷售數量(杯數)	餐廳家數
~100	17
101~200	29
201~300	121
301~400	75
401~500	60
501~600	25
601~700	9
701~800	8
801~900	2
901~	5

3.1.5 十分位數

將具有順序性質的資料均分為十等分的數值。第 i(個)十分位數(Deciles)，標記為 D_i(i = 1, 2, 3,…, 9)。意指至少有$\frac{i}{10}$比率的觀測值(數值點)小於等於 D_i，至少有$\frac{10-i}{10}$比率的觀測值(數值點)大於等於 D_i。一般十分位數值(D_1 到 D_9)不受極端觀測(數)值的影響。

3.1.5.1 未分組數值的十分位數

十分位數區分為兩種計算方式，第一種為 n - 1 模式，第二種為 n + 1 模式。兩端點定義有些差異，必須區分清楚。

十分位數 n - 1 模式：

在 n - 1 模式(即為 Excel 中 PERCENTILE.INC/PERCENTILE 函數計算方式)中，明確定義 D_0 為最小觀測值之數值；D_{10} 為最大觀測值之數值。即代表所有觀測值是以蒐集到的最大觀測值和最小觀測值為上下限，依據觀測值小到大將基本單位平均分配為十等份。

D_i 為一個以由**小**到**大**順序排列的資料集合中之第 $O(D_i) = (i \times \frac{n-1}{10}) + 1$ 序位的觀測值，其中 i：十分位之(序位)號碼；n 樣本數量。

a.若 $O(D_i) = (i \times \frac{n-1}{10}) + 1$ 序位數值為整數時，第 i(個)十分位數 D_i 為第 $O(D_i)$序位的觀測值。

b.若 $O(D_i) = (i \times \frac{n-1}{10}) + 1$ 序位數值為非整數時，第 i(個)十分位數 D_i 為【第 $O(D_i)$整數序位】與【第 $O(D_i)$整數+1 序位】對應的兩個觀測值，利用線性內插法計算前述兩個觀測值的比例平均值。

範例 3.17　高雄市 15 家國際觀光旅館，專職員工數量分別為 89、97、138、91、122、152、147、150、99、121、132、111、120、125、124 人，求第 5(個)十分位數 D_5 和第 7(個)十分位數 D_7。(答案有效位數四捨五入取到小數點後第 1 位)

題解：依據員工數由小至大排列

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
89	91	97	99	111	120	121	122	124	125	132	138	147	150	152

第 5(個)十分位數的序位：$O(D_{i=5}) = (i \times \frac{n-1}{10}) + 1 = (5 \times \frac{15-1}{10}) + 1 = 8$ 序位。第 5(個)十分位數 $D_5 = 122.0$ 人

第 7(個)十分位數的序位：$O(D_{i=7}) = (i \times \frac{n-1}{10}) + 1 = (7 \times \frac{15-1}{10}) + 1 = 10.8$ 序位。第 7(個)十分位數 $D_7 = 125 + 0.8 \times (132 - 125) = 130.6$ 人

答案：第 5(個)十分位數 $D_5 = 122.0$ 人；第 7(個)十分位數 $D_7 = 130.6$ 人

在 Excel 軟體中，沒有直接使用於十分位數的函數，必須借用百分位數的函數 PERCENTILE。因為，第 10 百分位數等於第 1 十分位數；第 20 百分位數等於第 2 十分位數。

十分位數 $n + 1$ 模式：

在 $n + 1$ 模式(即為 Excel 中 PERCENTILE.EXC 函數和 SPSS 軟體計算方式)中，沒有定義 D_0 為最小觀測值之數值；D_{10} 為最大觀測值之數值。即代表不是以蒐集到的最大觀測值和最小觀測值為上下限，依據觀測值小到大將基本單位平均分配為十等份。【選擇教材】

D_i 為一個以由**小**到**大**順序排列的資料集合中之第 $O(D_i) = i \times \frac{n+1}{10}$ 序位的觀測值，其中 i：十分位之(序位)號碼；n 樣本數量。

a.若 $O(D_i) = i \times \frac{n+1}{10}$ 序位數值為整數時，第 i(個)十分位數 D_i 為第 $O(D_i)$ 序位的觀測值。

b.若 $O(D_i) = i \times \frac{n+1}{10}$ 序位數值為非整數時，第 i(個)十分位數 D_i 為【第 $O(D_i)$ 整數序位】與【第 $O(D_i)$ 整數 +1 序位】對應的兩個觀測值，利用線性內插法計算前述兩個觀測值的比例平均值。

範例 3.18　高雄市 15 家國際觀光旅館，專職員工數量分別為 89、97、138、91、122、152、147、150、99、121、132、111、120、125、124 人，求第 5(個)十分位數 D_5 和第 7(個)十分位數 D_7。(答案有效位數四捨五入取到小數點後第 1 位)

題解：依據員工數由小至大排列

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
89	91	97	99	111	120	121	122	124	125	132	138	147	150	152

第 5(個)十分位數的序位：$O(D_{i=5}) = i \times \frac{n+1}{10} = 5 \times \frac{15+1}{10} = 8$ 序位。第 5(個)十分位數 $D_5 = 122.0$ 人

第 7(個)十分位數的序位：$O(D_{i=7}) = i \times \frac{n+1}{10} = 7 \times \frac{15+1}{10} = 11.2$ 序位。第 7(個)十分位數 $D_7 = 132 + 0.2 \times (138 - 132) = 133.2$ 人

答案：第 5(個)十分位數 $D_5 = 122.0$ 人；第 7(個)十分位數 $D_7 = 133.2$ 人

3.1.5.2 分組資料的十分位數【選擇教材】

a.計算以下累積次數(頻率)

b.計算 D_i 序位：$O(D_i) = (i \times \frac{n-1}{10}) + 1$ 序位，其中 $i = 1, 2, \ldots$或 9

c.由累積次數分配，決定 D_i 位於那個組別

d.D_i 所在組別中，利用下列公式，計算 D_i 數值

$$D_i = L_i + [(i \times \frac{n-1}{10}) + 1 - F] \times \frac{h_i}{f_i}$$

其中 L_i：D_i 所在組別的真正下限數值或組下界

n：總次數，母體或樣本總數

F：小於 D_i 所在組別的各組次數和(頻率和)、累計次數

h_i：D_i 所在組別的組距

f_i：D_i 所在組別的次數(頻率)

範例 3.19　嘿嘿地區 142 家餐廳在上個月份，販售牛排的銷售量分布表，求所有餐廳販售牛排分布的 D_3 和 D_6。(答案有效位數四捨五入取到小數點後第 2 位)

牛排銷售量	餐廳家數	以下累積次數
11~15	7	7
16~20	9	16
21~25	11	27
26~30	15	42
31~35	20	62
36~40	22	84
41~45	19	103
46~50	18	121
51~55	16	137
56~60	5	142

題解：

第 3(個)十分位數的序位：$O(D_{i=3}) = (i \times \frac{n-1}{10}) + 1 = (3 \times \frac{142-1}{10}) + 1 = 43.3$ 序位，D_3 位於 31~35 組別

$D_3 = L_i + [(i \times \frac{n-1}{10}) + 1 - F] \times \frac{h_i}{f_i} = 30.5 + [(3 \times \frac{142-1}{10}) + 1 - 42] \times \frac{5}{20} = 30.5 + 0.325 = 30.83$ 套

第 6(個)十分位數的序位：$O(D_{i=6}) = (i \times \frac{n-1}{10}) + 1 = (6 \times \frac{142-1}{10}) + 1 = 85.6$ 序位，D_6 位於 41~45 組別

$D_6 = L_i + [(i \times \frac{n-1}{10}) + 1 - F] \times \frac{h_i}{f_i} = 40.5 + [(6 \times \frac{142-1}{10}) + 1 - 84] \times \frac{5}{19} = 40.5 + 0.42 = 40.92$ 套

答案：$D_3 = 30.83$ 套；$D_6 = 40.92$ 套

3.1.6 百分位數

將具有順序性質的資料均分為一百等分的數值。第 i(個)百分位數(Percentile)，標記為 $P_i(i = 1, 2, 3,\ldots, 99)$。意指至少有 $\frac{i}{100}$ 比率的數值點小於等於 P_i，至少有 $\frac{100-i}{100}$ 比率的數值點大於等於 P_i。一般百分位數值(P_1 到 P_{99})不受極端觀測(數)值的影響。

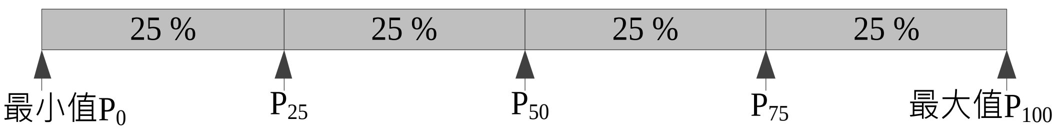


圖 3-3 百分位數分布

3.1.6.1 未分組數值的百分位數

百分位數區分為兩種計算方式，第一種為 n - 1 模式，第二種為 n + 1 模式。兩端點定義有些差異，必須區分清楚。

百分位數 n - 1 模式：

在 n - 1 模式(即為 Excel 中 PERCENTILE.INC/PERCENTILE 函數計算方式)中，明確定義 P_0 為最小觀測值之數值；P_{100} 為最大觀測值之數值。即代表所有觀測值是以蒐集到的最大觀測值和最小觀測值為上下限，依據觀測值小到大將基本單位平均分配為百等份。

P_i 為一個以由小到大順序排列的資料集合中之第 $O(P_i) = (i \times \frac{n-1}{100}) + 1$ 序位的觀測值，其中 i：百分位之號碼；n：樣本數量。

a.若 $O(P_i) = (i \times \frac{n-1}{100}) + 1$ 序位數值為整數時，第 i(個)百分位數 P_i 為第 $O(P_i)$序位的觀測值。

b.若 $O(P_i) = (i \times \frac{n-1}{100}) + 1$ 序位數值為非整數時，第 i(個)百分位數 P_i 為【第 $O(P_i)$整數序位】與【第 $O(P_i)$整數+1 序位】對應的兩個觀測值，利用線性內插法計算前述兩個觀測值的比例平均值。

範例 3.20　高雄市 15 家國際觀光旅館，專職員工數量分別為 89、97、138、91、122、152、147、150、99、121、132、111、120、125、124 人，求第 58(個)百分位數 P_{58} 和第 82(個)百分位數 P_{82}。(答案有效位數四捨五入取到小數點後第 2 位)

題解：依據員工數由小至大排列

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
89	91	97	99	111	120	121	122	124	125	132	138	147	150	152

第 58(個)百分位數的序位：$O(P_{i=58}) = (i \times \frac{n-1}{100}) + 1 = (58 \times \frac{15-1}{100}) + 1 = 9.12$ 序位

$P_{58} = 124 + 0.12 \times (125 - 124) = 124.12$ 人(內插法計算)

第 82(個)百分位數的序位：$O(P_{i=82}) = (i \times \frac{n-1}{100}) + 1 = (82 \times \frac{15-1}{100}) + 1 = 12.48$ 序位

$P_{82} = 138 + 0.48 \times (147 - 138) = 142.32$ 人(內插法計算)

答案：$P_{58} = 124.12$ 人；$P_{82} = 142.32$ 人

Excel 操作函數

點選欲放置百分位數的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):PERCENTILE→確定→選擇欲計算百分位數的資料區間，並輸入欲測量的百分位(將數值縮為 0 和 1 區間的數值，如 P_{55} 即輸入 0.55)→確定。

百分位數 $n + 1$ 模式：

在 $n + 1$ 模式(即為 Excel 中 PERCENTILE.EXC 函數和 SPSS 軟體計算方式)中，沒有定義 P_0 為最小觀測值之數值；P_{100} 為最大觀測值之數值。即代表不是以蒐集到的最大觀測值和最小觀測值為上下限，依據觀測值小到大將基本單位平均分配為百等份。【選擇教材】

P_i 為一個以由小到大順序排列的資料集合中之第 $O(P_i) = i \times \frac{n+1}{100}$序位的觀測值，其中 i：百分位之號碼；n：樣本數量。

a.若 $O(P_i) = i \times \frac{n+1}{100}$序位數值為整數時，第 i(個)百分位數 P_i 為第 $O(P_i)$序位的觀測值。

b.若 $O(P_i) = i \times \frac{n+1}{100}$序位數值為非整數時，第 i(個)百分位數 P_i 為【第 $O(P_i)$整數序位】與【第 $O(P_i)$整數+1 序位】對應的兩個觀測值，利用線性內插法計算前述兩個觀測值的比例平均值。

範例 3.21　高雄市 15 家國際觀光旅館，專職員工數量分別為 89、97、138、91、122、152、147、150、99、121、132、111、120、125、124 人，求第 58(個)百分位數 P_{58} 和第 82(個)百分位數 P_{82}。(答案有效位數四捨五入取到小數點後第 2 位)

題解：依據員工數由小至大排列

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
89	91	97	99	111	120	121	122	124	125	132	138	147	150	152

第 58(個)百分位數的序位：$O(P_{i=58}) = i \times \frac{n+1}{100} = 58 \times \frac{15+1}{100} = 9.28$ 序位

$P_{58} = 124 + 0.28 \times (125 - 124) = 124.28$ 人(內插法計算)

第 82(個)百分位數的序位：$O(P_{i=82}) = i \times \frac{n+1}{100} = 82 \times \frac{15+1}{100} = 13.12$ 序位

P_{82} = 147 + 0.12 × (150 – 147) = 147.36 人(內插法計算)

答案：P_{58} = 124.28 人；P_{82} = 147.36 人

SPSS 操作指令

Analyze(分析)→選擇 Descriptive Statistics(描述性統計)→選擇 Frequencies…(次數分配表)→在 Frequencies 視窗中，將欲統計 Percentile(百分位數)數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Frequencies 視窗下方之 Statistics…(統計量)按鈕→在 Frequencies: Statistics 次視窗左上角 Percentile Values(百分位數值)內勾選 Percentile(s):(百分位數)，於 Percentile(s)後面框框輸入欲統計百分位數的數值，按 Add(新增)按鈕將輸入數值加入右邊的小視窗，按右邊 Continue(繼續)按鈕→回到 Frequencies 視窗，按 OK(確定)按鈕即會跑出統計結果

Frequencies

Statistics

VAR00001

N	Valid	15
	Missing	0
Percentiles	58	124.2800

3.1.6.2 分組數值的百分位數【選擇教材】

a.計算以下累積次數(頻率)

b.計算 P_i 序位：O(P_i) = $(i \times \frac{n-1}{100}) + 1$ 序位，其中 i = 1, 2, …或 99

c.由累積次數分配，決定 P_i 位於那個組別

d.P_i 所在組別中，利用下列公式，計算 P_i 數值

$$P_i = L_i + [(i \times \frac{n-1}{100}) + 1 - F] \times \frac{h_i}{f_i}$$

其中 L_i：P_i 所在組別的真正下限數值或組下界

n：總次數，母體或樣本總數

F：小於 P_i 所在組別的各組次數和(頻率和)、累計次數

h_i：P_i 所在組別的組距

f_i：P_i 所在組別的次數(頻率)

範例 3.22　嘿嘿地區 142 家餐廳在上個月份，販售牛排的銷售量分布表，求所有餐廳販售牛排的 P_{40} 與 P_{56}。(答案有效位數四捨五入取到小數點後第 2 位)

牛排銷售量	餐廳家數	以下累積次數
11~15	7	7
16~20	9	16
21~25	11	27
26~30	15	42
31~35	20	62
36~40	22	84
41~45	19	103
46~50	18	121
51~55	16	137
56~60	5	142

題解：

第 40(個)百分位數的序位：O($P_{i=40}$) = $(i \times \frac{n-1}{100}) + 1 = (40 \times \frac{142-1}{100}) + 1$ = 57.4 序位，P_{40} 位於 31~35 組別

$P_{i=40} = L_i + [(i \times \frac{n-1}{100}) + 1 - F] \times \frac{h_i}{f_i} = 30.5 + [(40 \times \frac{142-1}{100}) + 1 - 42] \times \frac{5}{20} = 30.5 + 3.85 = 34.35$ 套

第 56(個)百分位數的序位：$O(P_{i=56}) = (i \times \frac{n-1}{100}) + 1 = (56 \times \frac{142-1}{100}) + 1 = 78.96$ 序位，P_{56} 位於 36~40 組別

$P_{i=56} = L_i + [(i \times \frac{n-1}{100}) + 1 - F] \times \frac{h_i}{f_i} = 35.5 + [(56 \times \frac{142-1}{100}) + 1 - 62] \times \frac{5}{22} = 35.5 + 3.85 = 39.35$ 套

答案：$P_{40} = 34.35$ 套；$P_{56} = 39.35$ 套

若依例計算百分位數 P_{10}、P_{20}、P_{30}、…、P_{90} 各數值即等於對應的**十分位數** D_1、D_2、D_3、…、D_9。計算方式與百分位數相同。

表 3-1 中位數、四分位數、十分位數和百分位數對照表

中位數	四分位數	十分位數	百分位數
	Q_0	D_0	P_0
		D_1	P_{10}
		D_2	P_{20}
	Q_1		P_{25}
		D_3	P_{30}
		D_4	P_{40}
M_e	Q_2	D_5	P_{50}
		D_6	P_{60}
		D_7	P_{70}
	Q_3		P_{75}
		D_8	P_{80}
		D_9	P_{90}
	Q_4	D_{10}	P_{100}

練習 *3.7*　　高雄所有咖啡館在上個星期，販售卡布奇諾(Cappuccino)咖啡數量(杯數)次數分布表，求咖啡館販售卡布奇諾咖啡數量分布的 Q_1、D_3、D_7、P_{16}、P_{21} 和 P_{83}。(答案有效位數四捨五入取到小數點後第 2 位)

銷售數量(杯數)	餐廳家數	以下累積次數
~100	7	7
101~200	29	36
201~300	41	77
301~400	75	152
401~500	60	212
501~600	25	237
601~700	19	256
701~800	18	274
801~900	2	276
901~	1	277

題解：

$O(Q_{i=1}) = (i \times \frac{n-1}{4}) + 1 = (1 \times \frac{277-1}{4}) + 1 = 70.0$ 序位，Q_1 位於 201~300 組別

$Q_1 = L_i + [(i \times \frac{n-1}{4}) + 1 - F] \times \frac{h_i}{f_i} = 200.5 + [(1 \times \frac{277-1}{4}) + 1 - 36] \times \frac{100}{41} = 200.5 + 82.93 = 283.43$ 杯

$O(D_{i=3}) = (i \times \frac{n-1}{10}) + 1 = (3 \times \frac{277-1}{10}) + 1 = 83.8$ 序位，D_3 位於 301~400 組別

$D_3 = L_i + [(i \times \frac{n-1}{10}) + 1 - F] \times \frac{h_i}{f_i} = 300.5 + [(3 \times \frac{277-1}{10}) + 1 - 77] \times \frac{100}{75} = 300.5 + 9.07 = 309.57$ 杯

$O(D_{i=7}) = (i \times \frac{n-1}{10}) + 1 = (7 \times \frac{277-1}{10}) + 1 = 194.2$ 序位，D_7 位於 401~500 組別

$D_7 = L_i + [(i \times \frac{n-1}{10}) + 1 - F] \times \frac{h_i}{f_i} = 400.5 + [(7 \times \frac{277-1}{10}) + 1 - 152] \times \frac{100}{60} = 400.5 + 70.33 = 470.83$ 杯

$O(P_{i=16}) = (i \times \frac{n-1}{100}) + 1 = (16 \times \frac{277-1}{100}) + 1 = 45.16$ 序位，P_{16} 位於 201~300 組別

$P_{16} = L_i + [(i \times \frac{n-1}{100}) + 1 - F] \times \frac{h_i}{f_i} = 200.5 + [(16 \times \frac{277-1}{100}) + 1 - 36] \times \frac{100}{41} = 200.5 + 22.34 = 222.84$ 杯

$O(P_{i=21}) = (i \times \frac{n-1}{100}) + 1 = (21 \times \frac{277-1}{100}) + 1 = 58.96$ 序位，P_{21} 位於 201~300 組別

$P_{21} = L_i + [(i \times \frac{n-1}{100}) + 1 - F] \times \frac{h_i}{f_i} = 200.5 + [(21 \times \frac{277-1}{100}) + 1 - 36] \times \frac{100}{41} = 200.5 + 56.00 = 256.50$ 杯

$O(P_{i=83}) = (i \times \frac{n-1}{100}) + 1 = (83 \times \frac{277-1}{100}) + 1 = 229.08$ 序位，P_{83} 位於 501~600 組別

$P_{83} = L_i + [(i \times \frac{n-1}{100}) + 1 - F] \times \frac{h_i}{f_i} = 500.5 + [(83 \times \frac{277-1}{100}) + 1 - 212] \times \frac{100}{25} = 500.5 + 68.32 = 568.82$ 杯

答案：$Q_1 = 283.43$ 杯；$D_3 = 309.57$ 杯；$D_7 = 470.83$ 杯；$P_{16} = 222.84$ 杯；$P_{21} = 256.50$ 杯；$P_{83} = 568.82$ 杯

表 3-2 分組資料數值相關中心位置測量值評量公式比較($n - 1$ 法)

	序位	中間數值
中位數	$O(M_e) = \frac{n}{2}$	$M_e = L_i + (\frac{n}{2} - F) \times \frac{h_i}{f_i}$
四分位數	$O(Q_i) = (i \times \frac{n-1}{4}) + 1$	$Q_i = L_i + [(i \times \frac{n-1}{4}) + 1 - F] \times \frac{h_i}{f_i}$
十分位數	$O(D_i) = (i \times \frac{n-1}{10}) + 1$	$D_i = L_i + [(i \times \frac{n-1}{10}) + 1 - F] \times \frac{h_i}{f_i}$
百分位數	$O(P_i) = (i \times \frac{n-1}{100}) + 1$	$P_i = L_i + [(i \times \frac{n-1}{100}) + 1 - F] \times \frac{h_i}{f_i}$

表 3-3 分組資料數值相關中心位置測量值評量公式比較($n + 1$ 法)

	序位	中間數值
中位數	$O(M_e) = \frac{n}{2}$	$M_e = L_i + (\frac{n}{2} - F) \times \frac{h_i}{f_i}$
四分位數	$O(Q_i) = i \times \frac{n+1}{4}$	$Q_i = L_i + [(i \times \frac{n+1}{4}) - F] \times \frac{h_i}{f_i}$
十分位數	$O(D_i) = i \times \frac{n+1}{10}$	$D_i = L_i + [(i \times \frac{n+1}{10}) - F] \times \frac{h_i}{f_i}$
百分位數	$O(P_i) = i \times \frac{n+1}{100}$	$P_i = L_i + [(i \times \frac{n+1}{100}) - F] \times \frac{h_i}{f_i}$

3.1.7 幾何平均值

一般資料各取對數後所得之算術平均值為對數單位(logarithmic scale)之資料，由此對數單位的算術平均值再取反對數(antilog)，所得之平均值稱為幾何平均值或幾何平均數(Geometric Mean, GM)。欲計算幾何平均值的全部觀測值皆必須為正值，不能有零或負值。

母體的幾何平均值標示為 μ_g、G 或 GM。

$$\mu_g = \text{antilog}(\frac{1}{N} \times \sum_{i=1}^{N} \log x_i) = \sqrt[N]{x_1 \times x_2 \times \dots \times x_N} = \sqrt[N]{\prod_{i=1}^{N} x_i} = (\prod_{i=1}^{N} x_i)^{\frac{1}{N}}$$

$$\log \mu_g = \frac{1}{N} \times \sum_{i=1}^{N} \log x_i$$

樣本的幾何平均值標示為$\bar{x}_g$、g 或 gm。

$$\bar{x}_g = \text{antilog}(\frac{1}{n} \times \sum_{i=1}^{n} \log x_i) = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$$

$$\log \bar{x}_g = \frac{1}{n} \times \sum_{i=1}^{n} \log x_i$$

幾何平均值使用於指數、滴定濃度、改變率(rate of change)、生長率(rate of growth)、投資報酬率、增殖率等資料。適用於等比數列(geometric sequence, geometric progression, G. P.，在數列 $a_1, a_2, a_3, \dots, a_n$ 中，$\frac{a_2}{a_1} = \frac{a_3}{a_2} = \frac{a_4}{a_3} = \dots = \frac{a_n}{a_{n-1}} = r$，比值 r 稱為公比。例如：數列 1, 3, 9, 27, 81, 243，$\frac{3}{1} = \frac{9}{3} = \frac{27}{9} = \frac{81}{27} = \frac{243}{81} = 3$)評量資料的中間值，惟不容易進行統計推論。

數值分布資料之幾何平均值$\bar{x}_g$小於等於算術平均值$\bar{x}$，即幾何平均值$\bar{x}_g \le$ 算術平均值$\bar{x}$。當各觀測值之數值皆相等時，幾何平均數$\bar{x}_g$會等於算術平均值$\bar{x}$。

未分組數值資料中，若有數值等於 0 時，或分組數值資料中某一組別的中點值為 0 時，計算獲得的幾何平均值即為 0，不具實質意義，故此狀況下不宜採用幾何平均值。

範例 3.23　請計算 3、4 和 5 三個樣本觀測值分布的幾何平均值？(答案有效位數四捨五入取到小數點後第 4 位)

題解：觀測值的數量 $n = 3$

$$幾何平均值\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n} = \sqrt[3]{3 \times 4 \times 5} = 3.9149$$

答案：幾何平均值 = 3.9149

範例 3.24　某飲料店為反應最近幾年營業額不佳，調降員工薪水，若今年調降 12 %預計明年調降 8 %因應，請計算此飲料店員工薪水之平均調降率？(答案有效位數四捨五入取到小數點後第 4 位)

題解：假設原始薪資數值為 1，x_1是今年調降後的薪資數值，今年調降 12 %，故調降後今年的薪資數值 x_1 = 前一年的數值 × (1 – 平均調降率) = 1 × (1 – 12 %) = 0.88。又假設今年的原始薪資數值為 1，x_2是明年再調降後的薪資數值，明年調降 8 %，調降後明年的薪資數值 x_2 = 前一年的數值 × (1 – 平均調降率) = 1 × (1 – 8 %) = 0.92。觀測值數量 $n = 2$ 年度。

薪資數值的幾何平均值$\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n} = \sqrt[2]{x_1 \times x_2} = \sqrt[2]{0.88 \times 0.92} = 0.8998$

欲計算平均調降率即是原始薪資數值減去薪資數值的幾何平均值，即【平均調降率 = 原始薪資數值 – 薪資數值的幾何平均值】。故，平均調降率 = 1 – 0.8998 = 0.1002 = 10.02 %

新一年的數值 = 前一年的數值 × (1 + 平均增加率)

新一年的數值 = 前一年的數值 × (1 – 平均調降率)

答案：平均調降率 = 0.1002 = 10.02 %

若原來薪水是新臺幣 20000 元

今年調降 12 %：20000 × (1 – 0.12) = 17600 元

明年調降 8 %：17600 × (1 – 0.08) = 16192 元

利用平均調降率計算

今年調降 10.02 %：20000 × (1 – 0.100222) = 17995.5550 元

明年調降 10.02 %：17995.6 × (1 – 0.100222) = 16192 元 → 兩個金額一樣，融會貫通原來如此簡單。

去年、今年和明年薪資數值屬於等比數列，依序分別為 20000、17995.6 和 16192 元，$\frac{今年薪水}{原始薪水} = \frac{17995.5550}{20000} = \frac{明年薪水}{今年薪水} = \frac{16192}{17995.5550}$ = 0.899778 就是前面**運算出來的幾何平均值**，【公比】數值一樣。

練習 *3.8*　純純連鎖咖啡館在台灣近 5 年來，販售義式咖啡杯數成長率分別為 10、20、30、-15、122 %。請計算販售義式咖啡杯數成長率的幾何平均值和年平均成長率。(答案有效位數四捨五入取到小數點後第 4 位)

題解：欲計算幾何平均值時，若題目提供的是成長率(相對性數值)，就必須轉換為原來的數值，例如：假設原始數值是 1，成長率 10 %就必須換算為 1 + (10 %) = 1.1(絕對數值)；成長率-15 %就必須換算為 1 + (-15 %) = 0.85(絕對數值)。

幾何平均值$\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \ldots \times x_n} = \sqrt[5]{1.1 \times 1.2 \times 1.3 \times 0.85 \times 2.22} = 1.2649$

獲得幾何平均值後，欲計算其年平均成長率，就必須將原來數值計算獲得的幾何平均值減 1 變為其增加的比率。

年平均成長率 = 幾何平均值 – 1 = 1.2649 – 1 = 0.2649 = 26.49 %

答案：幾何平均值 1.2649；年平均成長率 = 0.2649 = 26.49 %

幾何平均值亦可使用於計算一段期間內，週期次數為 n，僅需要期始值和期末值兩個觀測值，平均百分比的增長率(增加率)。

期間平均增加率(Average percent increase over time)通常以 G 符號代表

$$\text{期間平均增加率 } G = \sqrt[n]{\frac{\text{期末值}}{\text{期始值}}} - 1$$

範例 3.25　震名連鎖咖啡館在台灣近十年來，展店速度驚人，成長迅速，由 1999 年的 12 家店，到 2009 年已經拓展到 251 家店，十年內增加 239 家店。請計算在此十年內年平均成長率。(答案有效位數四捨五入取到小數點後第 4 位，不建議使用百分比標示)

題解：週期次數 $n = 10$

$$\text{期間平均增加率 } G = \sqrt[n]{\frac{\text{期末值}}{\text{期始值}}} - 1 = \sqrt[10]{\frac{251}{12}} - 1 = 1.3553 - 1 = 0.3553 = 35.53\ \%$$

答案：年平均成長率 = 0.3553 = 35.53 %

2000 年推估的店數 = 1999 年店數 × (1 + 年平均成長率) = 12 × (1 + 0.3553) = 16.2641

2001 年推估的店數 = 2000 年店數 × (1 + 年平均成長率) = 16.2641 × (1 + 0.3553) = 22.0435

2002 年推估的店數 = 2001 年店數 × (1 + 年平均成長率) = 22.0435 × (1 + 0.3553) = 29.8764

2003 年推估的店數 = 2002 年店數 × (1 + 年平均成長率) = 29.8764 × (1 + 0.3553) = 40.4928

2004 年推估的店數 = 2003 年店數 × (1 + 年平均成長率) = 40.4928 × (1 + 0.3553) = 54.8817

2005 年推估的店數 = 2004 年店數 × (1 + 年平均成長率) = 54.8817 × (1 + 0.3553) = 74.3835

2006 年推估的店數 = 2005 年店數 × (1 + 年平均成長率) = 74.3835 × (1 + 0.3553) = 100.8152

2007 年推估的店數 = 2006 年店數 × (1 + 年平均成長率) = 100.8152 × (1 + 0.3553) = 136.6392

2008 年推估的店數 = 2007 年店數 × (1 + 年平均成長率) = 136.6392 × (1 + 0.3553) = 185.1930

2009 年推估的店數 = 2008 年店數 × (1 + 年平均成長率) = 185.1930 × (1 + 0.3553) = 251.0000 → 與期始值和期末值相同，驗證年平均成長率運算內容。融會貫通原來如此簡單。

1999 年到 2009 年店數數值屬於等比數列，依序分別為 12.0000、16.2641、22.0435、29.8764、40.4928、54.8817、74.3835、100.8152、136.6392、185.1930 和 251.0000 家，$\frac{2000\text{ 年店數}}{1999\text{ 年店數}} = \frac{16.2641}{12.0000} = \frac{22.0435}{16.2641} = \frac{29.8764}{22.0435} = \frac{40.4928}{29.8764} = \frac{54.8817}{40.4928} = \frac{74.3835}{54.8817} = \frac{100.8152}{74.3835} = \frac{136.6392}{100.8152} = \frac{185.1930}{136.6392} = \frac{251.0000}{185.1930} = 1.3553$，公比數值一樣。

練習 *3.9*　震名連鎖咖啡館在台灣近 7 年來，販售義式咖啡杯數成長迅速，由 2002 年的 10255 杯，到 2009 年販售 265005 杯。請計算在此 7 年內年平均成長率。(答案有效位數四捨五入取到小數點後第 4 位)

Excel 操作函數

點選欲放置**幾何平均值**的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N): **GEOMEAN**→確定→選擇欲計算幾何平均值的資料區間→確定。

SPSS 操作指令

Analyze(分析)→選擇 Reports(報表)→選擇 Case Summaries…(觀察值摘要)→在 Summarize Cases 視窗中，將欲統計 Geometric Mean 數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Summarize Cases 視窗下方之 Statistics…(統計量)按鈕→在 Summary Report: Statistics 次視窗左小視窗 Statistics(統計量)點選 Geometric Mean(幾何平均數)至右小視窗 Cell Statistics(格統計)內，按下方 Continue(繼續)按鈕→回到 Summarize Cases 視窗，按 OK(確定)按鈕即會跑出統計結果

Sample

Summarize

Case Processing Summary[a]

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
VAR00001	6	100.0%	0	.0%	6	100.0%

[a] Limited to first 100 cases.

Case Summaries[a]

		VAR00001
1		9.00
2		1.00
3		4.00
4		5.00
5		75.00
6		7.00
Total	Geometric Mean	6.7490

[a] Limited to first 100 cases.

3.1.8 調和平均值【選擇教材】

各數值倒數的平均值的倒數，即為調和平均值、調和平均數(Harmonic Mean)或倒數平均值(Inverse Mean)，標記為 H。一般數值資料之調和平均值**小於等於**幾何平均值，即調和平均值$\bar{x}_H \leq$ 幾何平均值$\bar{x}_g$或 H ≤ G。當各觀測值之數值皆相等時，調和平均值會$\bar{x}_H$等於幾何平均值$\bar{x}_g$與算術平均值$\bar{x}$。

$$\bar{x}_H = \frac{1}{\frac{1}{n}\times\left(\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}\right)} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

範例 3.26 請計算 3、4 和 5 三個樣本分布的調和平均值？(答案有效位數四捨五入取到小數點後第 4 位)

題解：

$$調和平均值\bar{x}_H = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}} = \frac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} = 3.8298$$

答案：調和平均值 = 3.8298

Excel 操作函數

點選欲放置**調和平均值**的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):**HARMEAN**→確定→選擇欲計算調和平均值的資料區間→確定。

SPSS 操作指令

Analyze(分析)→選擇 Reports(報表)→選擇 Case Summaries…(觀察值摘要)→在 Summarize Cases 視窗中，將欲統計 Geometric Mean 數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Summarize Cases 視窗下方之 Statistics…(統計量)按鈕→在 Summary Report: Statistics 次視窗左小視窗 Statistics(統計量)點選 Harmonic Mean(調和平均值)至右小視窗 Cell Statistics(格統計)內，按下方 Continue(繼續)按鈕→回到 Summarize Cases 視窗，按 OK(確定)按鈕即會跑出統計結果

Sample

Summarize

Case Processing Summary[a]

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
VAR00001	4	100.0%	0	.0%	4	100.0%

a Limited to first 100 cases.

Case Summaries[a]

		VAR00001
1		5.00
2		6.00
3		7.00
4		8.00
Total	Harmonic Mean	6.3039

a Limited to first 100 cases.

3.1.9 平方根平均值【選擇教材】

將各樣本的數值取平方值之和，除以樣本數 n，再將其結果開根號，即可獲得**平方根平均值**(Quadratic Mean or Root Mean Square, RMS)，通常使用$\bar{x}_q$符號代表。

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

範例 3.27 請計算 3、4 和 5 三個樣本分布的平方根平均值？(答案有效位數四捨五入取到小數點後第 4 位)

題解：平方根平均值$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} = \sqrt{\frac{3^2+4^2+5^2}{3}} = 4.0825$

答案：平方根平均值 = 4.0825

3.1.10 縮減式平均值【選擇教材】

為減免極端值出現時，對平均值的影響。一律將最高值和最低值去除後，以剩下的數值來計算其算術平均值，獲得的平均值即為縮減式平均值、裁減式平均值、修正平均值或截尾平均值(Trimmed Mean)，一般使用$\bar{x}_t$或 m_T符號代表。**適用於小組評分時，免除個人關係分數**。

將原始觀測值的數值資料，由大排到小或由小排到大後，分別對前述順序資料中最小和最大的 k %觀測值刪除，計算剩餘觀測值的平均值，稱為 k %截尾平均值。

範例 3.28 請計算 40、58、60、70、72 和 99 六個樣本觀測值分布的縮減式平均值(trimmed mean) $\bar{x}_t$？(答案有效位數四捨五入取到小數點後第 1 位)

題解：去除最小數值 40 與去除最大數值 99，利用剩餘的觀測值計算縮減式平均值

$$縮減式平均值\bar{x}_t = \frac{58+60+70+72}{4} = 65.0$$

答案：縮減式平均值$\bar{x}_t$ = 65.0

Excel 操作函數

點選欲放置**縮減式平均值**的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):**TRIMMEAN**→確定→選擇欲計算縮減式平均值的資料區間和欲裁減的百分比數→確定。

練習 *3.10*　小魚牛肉麵店過去 1 小時，來店消費者結帳金額(單位：新臺幣元)分別為：

95　89　109　89　75　79　49　59　39

請計算：消費者結帳金額的平方根平均值$\bar{x}_q$、調和平均值$\bar{x}_H$和縮減式平均值$\bar{x}_t$。(答案有效位數四捨五入取到小數點後第 1 位)

3.1.11 溫塞平均值【選擇教材】

第一種方法：為減免極端值出現時，對平均值的影響。一律將最高值和最低值去除後，以次高值和次低值取代最高值和最低值(考慮相對權重)，計算取代後的算術平均值，即為溫賽平均值(Winsorized mean)，一般使用$\bar{x}_W$、m_w或 m_W符號代表。

第二種方法：利用第 1 個四分位數 Q_1 取代 Q_1 和其以下的觀測值；以第 3 個四分位數 Q_3 取代 Q_3 和其以上的觀測值，計算其取代後的算術平均值，即為溫賽平均值。

第三種方法：將原始觀測值的數值資料，由大排到小或由小排到大後，分別對前述順序資料中最小的 k %觀測值利用第 k 百分位數 P_k 取代；最大的 k %觀測值利用第 100 – k 百分位數 P_{100-k} 取代，計算取代後所有觀測值的平均值，稱為 k %溫塞平均值。

範例 3.29　請計算 40、58、60、70、72 和 99 六個樣本觀測值分布的溫塞平均值$\bar{x}_W$？(答案有效位數四捨五入取到小數點後第 1 位)

題解：六的樣本觀測值中，最小數值 40，以次小數值 58 取代；最大數值 99，以次大數值 72 取代。取用數值：58, 58, 60, 70, 72, 72 以計算溫賽平均值

$$溫賽平均值\bar{x}_W = \frac{58+58+60+70+72+72}{6} = 65.0$$

答案：溫塞平均值$\bar{x}_W$ = 65.0

3.1.12 加權平均值

加權平均值(Weighted Mean)或**加權算術平均值**(Weighted Arithmetic Mean)在計算一組資料的平均值時，每一個觀測值的貢獻程度並不相等，會依據每一個觀測值的重要性給予不同的權數，計算獲得其平均值。

計算算術平均值時，視各觀測值的重要性皆相同。然而，在特定領域中各觀測值的重要性已被假設成其重要性有差異存在，就需要使用加權平均值呈現其觀測值的中間趨勢。

第 i 個觀測值 x_i 的加權值利用 w_i 符號代表，**母體加權算術平均值** μ_w 為

$$\mu_w = \frac{\sum_{i=1}^{N} w_i \times x_i}{\sum_{i=1}^{N} w_i}$$

第 i 個觀測值 x_i 的加權值利用 w_i 符號代表，**樣本加權算術平均值**$\bar{x}_w$為

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i \times x_i}{\sum_{i=1}^{n} w_i}$$

範例 3.30 某大學餐旅管理學系三年級三位同學上學期修課分數，試計算三位同學的加權算術平均值。(答案有效位數四捨五入取到小數點後第 1 位)

課程	學分數	張☺號		黃貴☺		劉☺實	
		原始成績	加權成績	原始成績	加權成績	原始成績	加權成績
統計學	3	80	240			30	90
資料蒐集與應用	2	60	120	75	150		
生態觀光	2	80	160				
中餐製備	6			85	510		
西餐製備	6	95	570			94	564
研究方法	2			70	140	60	120
營養學	2	75	150	72	144	90	180
餐旅服務	2			91	182	91	182
體育	0	90	0	60	0	80	0
餐旅英文	3	67	201	80	240	75	225
合計	28	547	1441	533	1366	520	1361
算術平均值		78.1		76.1		74.3	
加權算術平均值			80.1		80.4		75.6

題解：因為每位學生修課的內容有相互的差異性，為了達到計算成績的公平性原則，故將學分數視為加權值，以便於計算加權算術平均值。

張小號的加權算術平均值

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i \times x_i}{\sum_{i=1}^{n} w_i} = \frac{3\times80+2\times60+2\times80+6\times95+2\times75+0\times90+3\times67}{3+2+2+6+2+0+3} = \frac{1441}{18} = 80.1$$

答案：加權算術平均值 80.1 分

加權算術平均值

$$\bar{x}_w = \frac{\sum_{i=1}^{k} f_i \times x_i}{n}$$

範例 3.31 研究抽樣樣本其家中小孩數量分布如下表，試計算家中小孩數量分布的平均值？(答案有效位數四捨五入取到小數點後第 1 位)

小孩數(x)	人數(f)	$f \times x$
0	12	0
1	26	26
2	33	66
3	26	78
4	21	84
5	15	75
6	10	60
7	9	63
合計	152	452

題解：$\bar{x}_w = \frac{\sum_{i=1}^{k} f_i \times x_i}{n} = \frac{452}{152} = 2.97$，故平均家中小孩數量 3.0 個

答案：小孩數量平均值 3.0 個

3.1.13 全距中點【選擇教材】

全距中點(Midrange)是將原來觀測值中最高數值和最低數值的中間數值。觀測值中出現極端值的影響最靈敏。

$$全距中點 = \frac{最高觀測值 + 最低觀測值}{2} = \frac{\text{highest value} + \text{lowest value}}{2}$$

3.1.14 中樞紐【選擇教材】

中樞紐(Midhinge)是將原來觀測值中第 1 四分位數 Q_1 和最 3 四分位數 Q_3 的平均值。觀測值中出現極端值的影響可去除。

$$中樞紐 = \frac{第\ 1\ 四分位數 + 第\ 3\ 四分位數}{2} = \frac{Q_1 + Q_3}{2}$$

練習 *3.11* 阿飄牛肉麵店過去 1 小時，來店消費者結帳金額(單位：新臺幣元)分別為：

95 89 109 89 75 79 49 59 39

請計算：消費者結帳金額分布的全距中點和中樞紐。(答案有效位數四捨五入取到小數點後第 1 位)

題解：

$$全距中點\ \text{Midrange} = \frac{最高觀測值 + 最低觀測值}{2} = \frac{109 + 39}{2} = 74.0$$

$$O(Q_1) = (i \times \frac{n-1}{4}) + 1 = (1 \times \frac{9-1}{4}) + 1 = 3\ 序位，Q_1\ 落於第\ 3\ 序位，Q_1 = 59\ 套$$

$$O(Q_3) = (i \times \frac{n-1}{4}) + 1 = (3 \times \frac{9-1}{4}) + 1 = 7\ 序位，Q_3\ 落於第\ 8\ 序位，Q_3 = 89\ 套$$

$$中樞紐\ \text{Midhinge} = \frac{第\ 1\ 四分位數 + 第\ 3\ 四分位數}{2} = \frac{Q_1 + Q_3}{2} = \frac{59 + 89}{2} = 74.0$$

答案：全距中點 Midrange = 74.0 元和中樞紐 Midhinge = 74.0 元

3.1.15 中心位置測定值比較

表 3-4 中心位置測定值差異比較

中心位置	採用頻率	存在性	觀測值採計	極端值影響
算術平均值	最常使用	總是存在	全部	會
中位數	普通使用	總是存在	一個	無
眾數	很少使用	可以不存在，亦可有多個	數個	無
幾何平均值	有時使用	總是存在	全部	會
全距中點	極少使用	總是存在	兩個	會

在左右對稱(symmetric, zero skewness)的次數分布中，算術平均值(mean) μ = 中位數(median) M_e = 眾數(mode) M_o。

在左偏(skewed to the left)或負偏(negatively skewed)次數分布中，算術平均值(mean) μ < 中位數(median) M_e < 眾數(mode) M_o。

在右偏(skewed to the right)或正偏(positively skewed)次數分布中，算術平均值(mean) μ > 中位數(median) M_e > 眾數(mode) M_o。

幾何平均值$\bar{x}_g \leq$ 算術平均值$\bar{x}$。

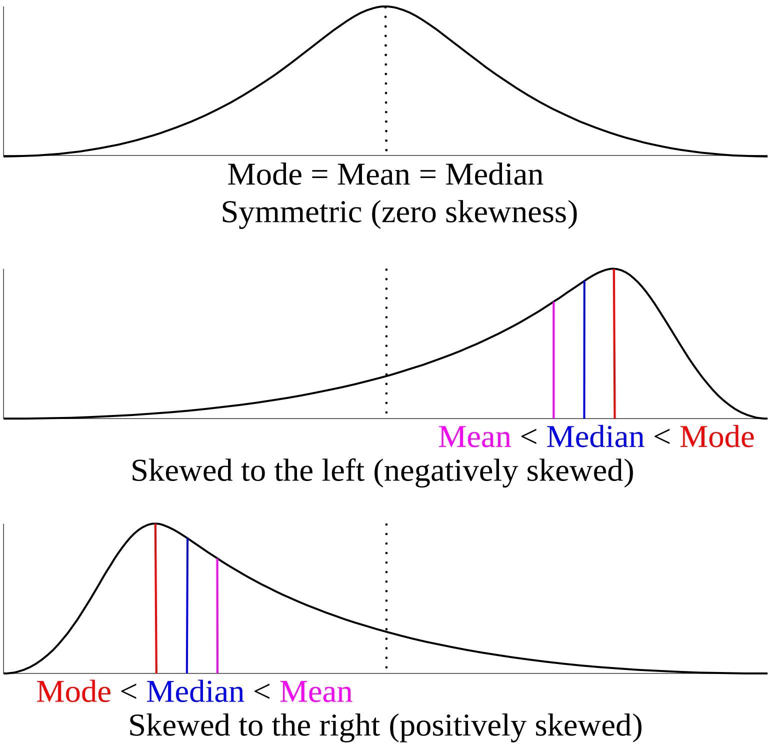


圖 3-4 對稱、左偏和右偏分布

練習 *3.12*　大紅牛肉麵店過去 10 小時內，來店消費者結帳金額(單位：新臺幣元)分別為：

95	89	109	89	75	79	49	59	39
95	95	95	85	75	49	49	59	95
89	109	95	85	89	49	59	39	

請計算：消費者結帳金額分布的**算術平均值**$\bar{x}$、**中位數** M_e 和**眾數** M_o 數值，並比較大小。判斷屬於左偏或右偏的次數分布。(答案有效位數四捨五入取到小數點後第 1 位)

答案：樣本平均值$\bar{x}$ = 76.7 元、中位數 M_e = 85.0 元、眾數 M_o = 95.0 元和左偏

練習 *3.13*　阿潔牛肉麵店過去 10 小時內，來店消費者結帳金額(單位：新臺幣元)分別為：

95	89	109	89	75	79	49	59	39
95	95	95	85	75	49	49	59	95
89	109	95	85	89	49	59	39	

請計算：消費者結帳金額分布的**算術平均值**$\bar{x}$、**中位數** M_e、**眾數** M_o、**幾何平均值**$\bar{x}_g$、**平方根平均值**$\bar{x}_q$、**調和平均值**$\bar{x}_H$、Q_1、Q_3、D_4、P_{38} 數值，並比較大小。判斷屬於左偏或右偏的次數分布。(答案有效位數四捨五入取到小數點後第 2 位)

$\bar{x}$ = 76.69 元, M_e = 85 元, M_o = 95 元, $\bar{x}_g$ = 73.35 元

練習 *3.14*　阿潔牛肉麵店過去 10 小時內，來店消費者結帳金額(單位：新臺幣元)分別為：

49	49	39	59	75	79	49	59	39
49	95	59	85	75	49	49	59	95
89	109	59	85	89	49	59	39	

請計算：消費者結帳金額分布的**算術平均值**$\bar{x}$、**中位數** M_e、**眾數** M_o、**幾何平均值**$\bar{x}_g$、**平方根平均值**$\bar{x}_q$、**調和平均值**$\bar{x}_H$數值，並比較大小。判斷屬於左偏或右偏的次數分布。(答案有效位數四捨五入取到小數點後第 2 位)

答案：$\bar{x}$ = 65 元；M_e = 59 元；M_o = 49 元；$\bar{x}_g$ = 62.14 元

練習 *3.15* 阿花餐廳過去 3 天，來店消費者每人消費金額(單位：新台幣元)分布如下表所示，請分別計算 Q_1、Q_3、D_3、P_{45}、M_e 數值。(答案有效位數四捨五入取到小數點後第 2 位)

消費金額(新台幣：元)	人次
51~60	7
61~70	9
71~80	31
81~90	45
91~100	50
101~110	62
111~120	59
121~130	28
131~140	16
141~150	5

練習 *3.16* 設有一組資料 2、3、7、8、9、9 和 11，則其算術平均值、中位數與眾數之關係為：(A)算術平均值<中位數<眾數；(B)算術平均值>中位數>眾數；(C)中位數<算術平均值<眾數；(D)中位數>算術平均值>眾數。(99 年初等考試統計學大意)

3.2 分散度測定值

評量特定研究變數之各觀測值分布的差異程度，即是觀測值(數據資料)的離中趨勢、變異性(variability)、離勢(dispersion)或離差(derivation)，稱為差異量數、變異量數(measure of variability)、離趨量數、分散量數、離勢量數(measure of dispersion)或變異係數。

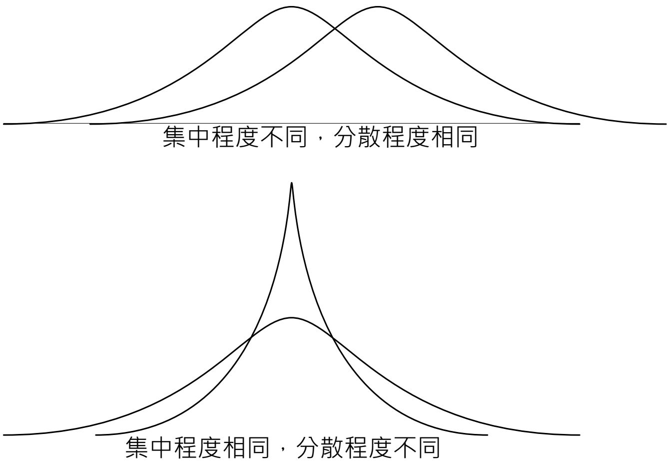


圖 3-5 集中程度和分散程度相同的分布

絕對離散趨勢(因次單位；dimension units)：全距、四分位差、標準(偏)差、變異數

相對離散趨勢(無因次單位；dimensionless units)：變異係數

3.2.1 全距

全距(Range)即樣本中最大與最小觀測值之差，此種表示法僅考慮資料中各觀測值分散的幅度，而不考慮其他觀測值變動的情形，標記為 R。全距單位與觀測值單位相同。

全距 R = 最大觀測值 – 最小觀測值

優點：計算容易。

缺點：遇有極端值時，其分散度(變異性)就顯得特別大。僅考慮觀測值中的最高與最低兩個數值，其他的數值並未被納入計算。

Excel 操作函數

= Large(資料區間,1) – Small(資料區間,1)

SPSS 操作指令

Analyze(分析)→選擇 Reports(報表)→選擇 Case Summaries…(觀察值摘要)→在 Summarize Cases 視窗中，將欲統計 Range(全距)數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Summarize Cases 視窗下方之 Statistics…(統計量)按鈕→在 Summary Report: Statistics 次視窗左小視窗 Statistics(統計量)點選 Range(全距)至右小視窗 Cell Statistics(格統計)內，按下方 Continue(繼續)按鈕→回到 Summarize Cases 視窗，按 OK(確定)按鈕即會跑出統計結果

Sample

Summarize

Case Processing Summary[a]

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
VAR00001	6	100.0%	0	.0%	6	100.0%

a Limited to first 100 cases.

Case Summaries[a]

		VAR00001
1		5.00
2		6.00
3		7.00
4		8.00
5		9.00
6		10.00
Total	Range	5.00

a Limited to first 100 cases.

3.2.1.1 分組資料的全距【選擇教材】

兩種評量法(既然是分組資料，其評量的精準性就會比不分組資料差，因此，計算獲得的數值僅是一個估計值。)

R = 最後一組組中點(組中值) – 第一組組中點(組中值)

R = 最後一組組上限值 – 第一組組下限值

範例 3.32 研究抽樣樣本的年齡層分布如下表，試計算其全距(range)？(答案有效位數四捨五入取到個位數)

年齡層	人數
20~29	12
30~39	26
40~49	33
50~59	26
60~69	21

題解：

年齡層	人數(f_i)	組中點(m_i)
20~29	12	24.5
30~39	26	34.5
40~49	33	44.5
50~59	26	54.5
60~69	21	64.5
合計	118	

R = 最後 1 組組中點(組中值) – 第 1 組組中點(組中值) = 64.5 – 24.5 = 40

R = 最後 1 組組上限值 – 第 1 組組下限值 = 69 – 20 = 49

答案：全距 40 或 49 歲

3.2.2 四分位距

第 3(個)四分位數與第 1(個)四分位數的距離即為**四分位距**、**中四分位數距**(Interquartile range; Inter-quartile range, IQR)或**四分位差**(Quartile deviation, QD)，標記為 *IQR*。衡量原始觀測值中間 50 %的數值，依據此中間 50 %數值中最高值與最低值的差距。四分位距單位與觀測值單位相同。

四分位距 IQR = 第 3(個)四分位數 – 第 1(個)四分位數 = $Q_3 – Q_1$

優點：可以免除極端值的影響。

缺點：僅考慮觀測值中的兩個數值，其他數值並未被納入計算。

範例 3.33 小熊牛肉麵店過去 1 小時，來店消費者結帳金額(單位：新臺幣元)分別為：

95 89 109 89 75 79 49 59 39

請計算消費者結帳金額分布的四分位距(interquartile range, IQR)。(答案有效位數四捨五入取到小數點後第 1 位)

題解：

四分位距 IQR = 第 3(個)四分位數 Q_3 – 第 1(個)四分位數 Q_1 = 89.0 – 59.0 = 30.0

答案：第 1(個)四分位數 Q_1 = 59.0 元；第 3(個)四分位數 Q_3 = 89.0 元；四分位距 IQR = 30.0 元

3.2.3 平均偏差【選擇教材】

在考慮每個觀測值之變化情形時，若以所有觀測值(資料)之算術平均值為中心，各個觀測值與算術平均值之差稱為偏差(deviation)、離差、離均差(deviation from mean)或剩餘。樣本以 $x_i – \bar{x}$表示。母體以 $x_i – \mu$ 表示。

若將所有樣本或母體的觀測值離差之合計，數值會等於 0，$\sum_{i=1}^{n}(x_i-\bar{x}) = 0$ 或$\sum_{i=1}^{N}(x_i-\mu) = 0$，觀測值離差之合計為 0 無法表達出觀測值分布的離散程度。

將偏差取絕對值後，全部將偏差轉成正值，此正值的算術平均值即為**平均偏差**、**平均差**、**平均離差**或**平均絕對離差**(Mean deviation, MD), (Average deviation, AD), (Mean absolute deviation, MAD)。平均偏差愈大者代表觀測值的分布愈分散；平均偏差愈小者代表觀測值的分布愈集中。平均偏差單位與觀測值單位相同。

$$母體資料：MD = \frac{1}{N} \times \sum_{i=1}^{N}|x_i-\mu|$$

$$樣本資料：\mathrm{md} = \frac{1}{n} \times \sum_{i=1}^{n}|x_i-\bar{x}|$$

若以中位數為中心測量觀測值分布程度，稱為**離中差(deviation from median)**。

$$\mathrm{MD} = \frac{1}{n} \times \sum_{i=1}^{n}|x_i-M_e|$$

優點：平均偏差是表示觀測值變異現象良好的方法。

缺點：可是實際上很少採用，因為平均偏差公式在數理上無法做運算或分解。

範例 3.34 高雄市其中 5 家國際觀光旅館，專職員工數量分別為 80、90、130、90 和 120 人，分別利用算術平均值和中位數計算其平均偏差 *MD*。(答案有效位數四捨五入取到小數點後第 1 位)

題解：

x_i	$\lvert x_i - \bar{x}\rvert$, $\bar{x}$=102	$\lvert x_i - M_e\rvert$, M_e=90
80	22	10
90	12	0
130	28	40
90	12	0
120	18	30
$n = 5$	$\sum_{i=1}^{N}\lvert x_i - \bar{x}\rvert = 92$	$\sum_{i=1}^{N}\lvert x_i - M_e\rvert = 80$

平均離均差 MD = $\frac{1}{N} \times \sum_{i=1}^{N}|x_i - \bar{x}| = \frac{92}{5} = 18.4$ 人

平均離中差 MD = $\frac{1}{N} \times \sum_{i=1}^{N}|x_i - M_e| = \frac{80}{5} = 16.0$ 人

答案：平均離均差 = 18.4 人；平均離中差 = 16.0 人

Excel 操作函數

點選欲放置**平均偏差值**的儲存格→插入(I)→函數(F)…→在插入函數的對話方塊中選擇統計類別，選取函數(N):**AVEDEV**→確定→選擇欲計算平均偏差值的資料區間→確定。

3.2.3.1 分組資料的平均偏差【選擇教材】

母體資料：平均偏差 $MD = \frac{1}{N} \times \sum_{i=1}^{k}|m_i - \mu| \times f_i$

樣本資料：平均偏差 $md = \frac{1}{n} \times \sum_{i=1}^{k}|m_i - \bar{x}| \times f_i$

其中 N：母體總數　k：組數　m_i：i 組之組中點、組中值

　　 n：樣本總數　f_i：i 組之次數、頻率

平均偏差數值愈大，其資料之分散程度愈大。平均偏差之計算有納入每一個層級(分組)的資料。

範例 3.35 研究抽樣樣本的年齡層分布如下表，試計算其平均偏差？(答案有效位數四捨五入取到小數點後第 2 位)

年齡層	人數
20~29	12
30~39	26
40~49	33
50~59	26
60~69	21

題解：

年齡層	人數(f_i)	組中點(m_i)	$f_i \times m_i$	$\lvert m_i - \bar{x}\rvert$	$\lvert m_i - \bar{x}\rvert \times f_i$
20~29	12	24.5	294.0	21.5	258.0
30~39	26	34.5	897.0	11.5	299.0
40~49	33	44.5	1468.5	1.5	49.5
50~59	26	54.5	1417.0	8.5	221.0
60~69	21	64.5	1354.5	18.5	388.5
合計	118		5431.0		1216.0

樣本平均值 $\bar{x} = \frac{\sum_{i=1}^{k} f_i \times m_i}{n} = \frac{5431}{118} = 46.0$，故年齡的算術平均值為 46.0 歲

平均偏差 $md = \frac{1}{n} \times \sum_{i=1}^{k}|m_i - \bar{x}| \times f_i = \frac{1216.0}{118} = 10.31$(歲)

答案：平均偏差 = 10.31 歲

3.2.4 變異數與標準(偏)差

設欲調查母體資料有 N 個觀測值(x_i, i = 1, 2, ..., N)，其平均值為 μ，各觀測值偏差平方後之總和[sum of square, SS = $\sum_{i=1}^{N}(x_i-\mu)^2$]除以 N 即得一變異數，特稱此變異數(variance)為變方，統計學上皆以 σ^2(讀音 sigma square)、Var(X)或 V(X)符號表示。變異數單位為觀測值單位的平方，即(觀測值單位)^2。

變異數數值一定都是屬於正值，不會有負值出現，最小變異數數值為 0，其代表每一個基本單位的觀測值數值皆相等。$E(X)$是隨機變數 X 的期望值(重複多次測量隨機變數 X，用所獲得的數值計算之平均值)【隨機變數 X 的期望值 $E(X)$ = 母體平均值 μ】。

$$母體變異數\ \sigma^2 = E[(x-\mu)^2] = \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N} = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} = \frac{1}{N}\times\sum_{i=1}^{N}(x_i-\mu)^2 = \frac{1}{N}\times\sum_{i=1}^{N}\left(x_i^2-2\times x_i\times\mu+\mu^2\right)$$
$$= \frac{1}{N}\times\left[\sum_{i=1}^{N}x_i^2-2\times\sum_{i=1}^{N}x_i\times\mu+\sum_{i=1}^{N}\mu^2\right] = \frac{1}{N}\times\left[\sum_{i=1}^{N}x_i^2-2\times N\times\mu\times\mu+N\times\mu^2\right]$$
$$= \frac{1}{N}\times\left[\sum_{i=1}^{N}x_i^2-N\times\mu^2\right] = \frac{\sum_{i=1}^{N}x_i^2}{N}-\mu^2 = \frac{1}{N}\times\left[\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}\right]$$

變異數數值愈大者代表觀測值的分布愈分散；變異數數值愈小者代表觀測值的分布愈集中。

標準(偏)差(standard deviation, SD)以 σ(讀音 sigma)符號表示。標準(偏)差單位與觀測值單位相同。標準(偏)差數值一定都是屬於正值，不會有負值出現，最小標準(偏)差數值為 0，代表每一個基本單位的觀測值數值皆相等。

$$母體標準(偏)差\ SD = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$$

$$\sum_{i=1}^{N}(x_i-\mu)^2 = \sum_{i=1}^{N}x_i^2-2\times\mu\times\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}\mu^2 = \sum_{i=1}^{N}x_i^2-2\times\frac{\sum_{i=1}^{N}x_i}{N}\times\sum_{i=1}^{N}x_i+N\times\mu^2$$
$$= \sum_{i=1}^{N}x_i^2-2\times\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}+N\times\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2 = \sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}}{N}} = \frac{\sqrt{N\times\sum_{i=1}^{N}x_i^2-\left(\sum_{i=1}^{N}x_i\right)^2}}{N}$$

範例 3.36　高雄市全部 5 家國際觀光旅館，專職員工數量分別為 80、90、130、90 和 120 人，請計算高雄市國際觀光旅館專職員工數分布之變異數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：

x_i	$x_i-\mu$	$(x_i-\mu)^2$
80	-22	484
90	-12	144
130	28	784
90	-12	144
120	18	324
$\sum_{i=1}^{N}x_i=510$		$\sum_{i=1}^{N}(x_i-\mu)^2=1880$
μ = 102		

母體變異數 $\sigma^2=\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N}=\frac{1880}{5}$ = 376人^2(亦可使用 Excel 軟體 VAR.P 函數查詢獲得)

母體標準(偏)差 $\sigma=\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}=\sqrt{\frac{1880}{5}}=\sqrt{376}$ = 19.39 人(亦可使用 Excel 軟體 STDEV.P 函數查詢獲得)

答案：母體變異數 = 376 人^2；母體標準(偏)差 = 19.39 人

練習 *3.17*　高雄市全部 10 家國際觀光旅館，專職員工數量(單位：人)分別如下所示，請計算高雄市國際觀光旅館專職員工數分布之標準(偏)差。(答案有效位數四捨五入取到小數點後第 1 位)

125 425 325 425 125 265 325 215 198 265

題解：

x_i	$x_i-\mu$	$(x_i-\mu)^2$
125	-144.3	20822.5
425	155.7	24242.5
325	55.7	3102.5
425	155.7	24242.5
125	-144.3	20822.5
265	-4.3	18.5
325	55.7	3102.5
215	-54.3	2948.5
198	-71.3	5083.7
265	-4.3	18.5
$\sum_{i=1}^{N} x_i$ = 2693		104404.1
μ = 269.3		

母體標準(偏)差 $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{104404.1}{10}} = \sqrt{10440.41} = 102.2$ (人)(亦可使用 Excel 軟體 STDEV.P 函數查詢獲得)

答案：母體標準(偏)差 = 102.2 人

Excel 操作函數

利用 Excel 軟體中插入(I)→函數(F)…→選擇函數類別：統計中選取 **STDEV.P**(number1,number2,…)函數，再選取欲進行統計分析的欄位區間，即可計算所選取欄位區間內數值之母體的標準差。

若提供母體基本單位的觀測值 x_i 和其出現的次數 f_i，一共有 k 種不同的觀測值數值，計算母體平均值、母體變異數和母體標準(偏)差的方式為：

母體平均值 $\mu = \frac{\sum_{i=1}^{k} x_i \times f_i}{N} = \frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_{i}^{k} f_i}$

母體變異數 $\sigma^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N} = \frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_{i}^{k} f_i}$

母體標準(偏)差 $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_{i}^{k} f_i}}$

範例 3.37 務實班級全班同學穿著鞋子的號碼(US)分別有 7.5、8、8.5、9、9.5 和 10 號，其人數依序分別為 2、4、8、16、15 和 12 人，請計算全班同學穿著鞋子號碼分布之變異數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

x_i	f_i	$x_i \times f_i$	$x_i-\mu$	$(x_i-\mu)^2$	$f_i \times (x_i-\mu)^2$
7.5	2	15	-1.6491	2.7196	5.4392
8.0	4	32	-1.1491	1.3205	5.2819
8.5	8	68	-0.6491	0.4214	3.3709
9.0	16	144	-0.1491	0.0222	0.3558
9.5	15	142.5	0.3509	0.1231	1.8467
10.0	12	120	0.8509	0.7240	8.6879
合計	57	521.5			24.9825

母體平均值 $\mu = \frac{\sum_{i=1}^{k} x_i \times f_i}{N} = \frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_{i}^{k} f_i} = \frac{521.5}{57} = 9.1491$

母體變異數 $\sigma^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N} = \frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_{i}^{k} f_i} = \frac{24.9825}{57} = 0.4383$

母體標準(偏)差 $\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_{i}^{k} f_i}} = \sqrt{\frac{24.9825}{57}} = \sqrt{0.4383} = 0.6620$

答案：母體變異數 = 0.4383；母體標準(偏)差 = 0.6620

3.2.4.1 分組資料的變異數和標準(偏)差【選擇教材】

母體資料

母體變異數 $\sigma^2 = \frac{1}{N} \times \sum_{i=1}^{k}[(m_i - \mu)^2 \times f_i]$

母體標準(偏)差 $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{k}[(m_i - \mu)^2 \times f_i]}$

其中 N：母體總數　k：組數　m_i：i 組之組中點、組中值　f_i：i 組之次數、頻率

範例 3.38　進修學院某班級全體學生年齡層分布如下表，試計算其變異數和標準(偏)差？(答案有效位數四捨五入取到小數點後第 2 位)

年齡層(歲)	人數
20~29	12
30~39	26
40~49	33
50~59	26
60~69	21

題解：

年齡層	人數(f_i)	組中點(m_i)	$f_i \times m_i$	$m_i - \mu$	$(m_i - \mu)^2$	$(m_i - \mu)^2 \times f_i$
20~29	12	24.5	294.0	-21.5	462.25	5547.00
30~39	26	34.5	897.0	-11.5	132.25	3438.50
40~49	33	44.5	1468.5	-1.5	2.25	74.25
50~59	26	54.5	1417.0	8.5	72.25	1878.50
60~69	21	64.5	1354.5	18.5	342.25	7187.25
合計	118		5431.0			18125.50

母體平均值 $\mu = \frac{\sum_{i=1}^{k} f_i \times m_i}{N} = \frac{5431}{118} = 46.0$，故年齡的算術平均值為 46.0 歲

母體變異數 $\sigma^2 = \frac{1}{N} \times \sum_{i=1}^{k}[(m_i - \mu)^2 \times f_i] = \frac{18125.50}{118} = 153.61$(歲 ^{2})

母體標準(偏)差 $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{k}[(m_i - \mu)^2 \times f_i]} = \sqrt{153.61} = 12.39$(歲)

答案：母體變異數 = 153.61 歲 ^{2}；母體標準(偏)差 = 12.39 歲

練習 3.18　進修學院某班級全體學生年齡層分布如下表，試計算其平均值、變異數和標準(偏)差？(答案有效位數四捨五入取到小數點後第 1 位)

年齡層(歲)	人數
20~29	11
30~39	16
40~49	15
50~59	6
60~69	2

題解：

年齡層	人數(f_i)	組中點(m_i)	$f_i \times m_i$	$m_i - \mu$	$(m_i - \mu)^2$	$(m_i - \mu)^2 \times f_i$
20~29	11	24.5	269.5	-14.4	207.36	2280.96
30~39	16	34.5	552.0	-4.4	19.36	309.76
40~49	15	44.5	667.5	5.6	31.36	470.40
50~59	6	54.5	327.0	15.6	243.36	1460.16
60~69	2	64.5	129.0	25.6	655.36	1310.72
合計	50		1945.0			5832.00

母體平均值 $\mu = \frac{\sum_{i=1}^{k} f_i \times m_i}{N} = \frac{1945}{50} = 38.9$，故年齡的算術平均值為 38.9 歲

母體變異數 $\sigma^2 = \frac{1}{N} \times \sum_{i=1}^{k}[(m_i - \mu)^2 \times f_i] = \frac{5832.00}{50} = 116.6$(歲 ^{2})

母體標準(偏)差 σ = $\sqrt{\sigma^2}$ = $\sqrt{\frac{1}{N} \times \sum_{i=1}^{k}[(m_i-\mu)^2 \times f_i]}$ = $\sqrt{116.6}$ = 10.8(歲)

答案：母體平均值 = 38.9 歲；母體變異數 = 116.6 歲 ^{2}；母體標準(偏)差 = 10.8 歲

3.2.5 樣本變異數

母體資料數量 N 往往很大或無限大，其全部資料不易取得，且母體平均值 μ 實際上亦未知。透過抽樣程序，皆可以容易獲得樣本的觀測值，利用樣本變異數 S^2 估算母體變異數 σ^2。

若利用母體變異數的計算公式，依據邏輯推論直接將母體資料數量 N 改成樣本資料數量 n 時，所獲得的變異數會有低估(underestimate)的現象發生，因此，直接由母體變異數的計算公式，修改 N 為 n 時會產生偏差的估計值(biased estimate)。

若將樣本觀測值的偏差平方和[sum of square, SS = $\sum_{i=1}^{n}(x_i-\bar{x})^2$]除以 n – 1，即可獲得**樣本變異數**(Sample Variance)、樣本變方或均方(Mean Square)，此樣本變異數 S^2 是母體變異數 σ^2 的不偏估計值(unbiased estimate)。

範例 3.39 設母體觀測值分別為 2、5 和 8，其基本單位總數 N = 3，母體平均值 μ = 5，母體變異數 σ^2 = 6，從此母體中抽取 n = 2 個觀測值為樣本，採用抽出後立即放回，以供下一次抽取之用，其所有可能樣本組合以及其平均值如下：

樣本	平均值 $\bar{x}$	$(\bar{x}-\mu)^2$	$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$	$S_b^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$
2 , 2	2.0	9.00	0.0	0.00
2 , 5	3.5	2.25	4.5	2.25
2 , 8	5.0	0.00	18.0	9.00
5 , 2	3.5	2.25	4.5	2.25
5 , 5	5.0	0.00	0.0	0.00
5 , 8	6.5	2.25	4.5	2.25
8 , 2	5.0	0.00	18.0	9.00
8 , 5	6.5	2.25	4.5	2.25
8 , 8	8.0	9.00	0.0	0.00
和	45.0	27.00	54.0	27.00
平均值	5	3.00	6.0	3.00
	μ	$\frac{\sigma^2}{n}$	σ^2	

將前述三個觀測值的母體中，隨機抽出 2 個觀測值為樣本，一共會有 9 種組合方式，若透過前述樣本變異數 $S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$，一一運算出每一種組合的樣本變異數，9 種組合的樣本變異數之和為 54，平均值為 6，剛剛好等於母體變異數 σ^2 = 6。假設使用 $S_b^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$ 當成樣本變異數時，一一運算出每一種組合的樣本變異數，9 種組合的樣本變異數之和為 27，平均值為 3，不會等於母體變異數 σ^2 = 6，產生偏差。使用樣本變異數 $S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$ 估算母體變異數數值時，不會產生偏差，故，樣本變異數 $S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$ 是母體變異數 σ^2 的不偏估計值(unbiased estimate)。。

樣本變異數(Sample variance)一般利用 S^2 符號代表，變異數單位 = [觀測值單位]^2：

$$\text{樣本變異數 } S^2 = S_x^2 = S_{xx} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1}$$

樣本標準(偏)差(Sample standard deviation)，一般利用 S 符號代表，標準(偏)差單位 = [觀測值單位]：

樣本標準(偏)差 $S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{(\sum_{i=1}^{n}x_i^2)-[2\times(\sum_{i=1}^{n}x_i)\times\bar{x}]+\sum_{i=1}^{n}\bar{x}^2}{n-1}} = \sqrt{\frac{(\sum_{i=1}^{n}x_i^2)-(2\times n\times\bar{x}\times\bar{x})+n\times\bar{x}^2}{n-1}} =$

$\sqrt{\frac{(\sum_{i=1}^{n}x_i^2)-2\times n\times\bar{x}^2+n\times\bar{x}^2}{n-1}} = \sqrt{\frac{(\sum_{i=1}^{n}x_i^2)-n\times\bar{x}^2}{n-1}} = \sqrt{\frac{(\sum_{i=1}^{n}x_i^2)-n\times\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2}{n-1}} = \sqrt{\frac{(\sum_{i=1}^{n}x_i^2)-n\times\frac{(\sum_{i=1}^{n}x_i)^2}{n^2}}{n-1}} =$

$\sqrt{\frac{(\sum_{i=1}^{n}x_i^2)-\frac{(\sum_{i=1}^{n}x_i)^2}{n}}{n-1}}$

樣本變異數(sample variance)另外簡單算式：

$$S^2 = \frac{(\sum_{i=1}^{n}x_i^2)-n\times\bar{x}^2}{n-1} = \frac{(\sum_{i=1}^{n}x_i^2)-\frac{(\sum_{i=1}^{n}x_i)^2}{n}}{n-1} = \frac{1}{n-1}\times\left[(\sum_{i=1}^{n}x_i^2)-\frac{(\sum_{i=1}^{n}x_i)^2}{n}\right]$$

範例 3.40　抽樣調查高雄市旅館專職員工人數，選取其中 5 家旅館當樣本，其專職員工數量分別為 80、90、130、90 和 120 人，請計算高雄市旅館專職員工數分布之標準(偏)差(standard deviation)。(答案有效位數四捨五入取到小數點後第 1 位)

題解：

x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
80	-22	484
90	-12	144
130	28	784
90	-12	144
120	18	324
$\sum_{i=1}^{n}x_i$ = 510		$\sum_{i=1}^{n}(x_i-\bar{x})^2$ = 1880

樣本平均值 $\bar{x} = \sum_{i=1}^{n}\frac{x_i}{n} = \frac{\sum_{i=1}^{n}x_i}{n} = \frac{510}{5} = 102$

樣本標準(偏)差 $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{1880}{5-1}} = \sqrt{470} = 21.7$ (人)(亦可使用 Excel 軟體 STDEV.S 函數查詢獲得)

答案：樣本標準(偏)差 = 21.7 人

練習 3.19　抽樣調查高雄市選取其中 10 家國際觀光旅館當樣本，專職員工數量(單位：人)分別如下所示，請計算高雄市國際觀光旅館專職員工數分布之標準(偏)差(standard deviation)。(答案有效位數四捨五入取到小數點後第 1 位)

225　425　325　425　185　265　325　215　198　265

題解：

	x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
	225	-60.3	3636.09
	425	139.7	19516.09
	325	39.7	1576.09
	425	139.7	19516.09
	185	-100.3	10060.09
	265	-20.3	412.09
	325	39.7	1576.09
	215	-70.3	4942.09
	198	-87.3	7621.29
	265	-20.3	412.09
和	2853		69268.10
平均值	285.3		

樣本標準(偏)差 $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{69268.10}{10-1}} = \sqrt{7696.456} = 87.7$ (人)(亦可使用 Excel 軟體 STDEV.S 函數查詢獲得)

答案：樣本標準(偏)差(sample standard deviation) = 87.7 人

Excel 操作函數

利用 Excel 軟體中插入(I)→函數(F)…→選擇函數類別：統計中選取 STDEV.S(number1,number2,…)函數，再選取欲進行統計分析的欄位區間，即可計算所選取欄位區間內數值之樣本的標準差。

SPSS 操作指令

Analyze(分析)→選擇 Descriptive Statistics(描述性統計)→選擇 Frequencies…(次數分配表)→在 Frequencies 視窗中，將欲統計 Standard Deviation 數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Frequencies 視窗下方之 Statistics…(統計量)按鈕→在 Frequencies: Statistics 次視窗左下角 Dispersion(分散情形)內勾選 Std.deviation(標準差)和 Variance(變異數)，按右邊 Continue(繼續)按鈕→回到 Frequencies 視窗，按 OK(確定)按鈕即會跑出統計結果

Frequencies

Statistics

VAR00001

N	Valid	5
	Missing	0
Std. Deviation		21.67948
Variance		470.000

若提供樣本基本單位的觀測值 x_i 和其出現的次數 f_i，一共有 k 種不同的觀測值數值，計算樣本平均值、樣本變異數和樣本標準(偏)差的方式為：

樣本平均值 $\bar{x} = \frac{\sum_{i=1}^{k} x_i \times f_i}{n} = \frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i}$

樣本變異數 $S^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{\sum_i^k f_i - 1}$

樣本標準(偏)差 $S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{\sum_i^k f_i - 1}}$

範例 3.41 抽樣天使學系數位同學穿著鞋子的號碼(US)分別有 7.5、8、8.5、9、9.5 和 10 號，其人數依序分別為 2、4、8、16、15 和 12 人，請計算樣本同學穿著鞋子號碼分布之變異數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

x_i	f_i	$x_i \times f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i \times (x_i - \bar{x})^2$
7.5	2	15	-1.6491	2.7196	5.4392
8.0	4	32	-1.1491	1.3205	5.2819
8.5	8	68	-0.6491	0.4214	3.3709
9.0	16	144	-0.1491	0.0222	0.3558
9.5	15	142.5	0.3509	0.1231	1.8467
10.0	12	120	0.8509	0.7240	8.6879
合計	57	521.5			24.9825

樣本平均值 $\bar{x} = \frac{\sum_{i=1}^{k} x_i \times f_i}{n} = \frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i} = \frac{521.5}{57} = 9.1491$

樣本變異數 $S^2 = \frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{\sum_i^k f_i - 1} = \frac{24.9825}{57-1} = 0.4461$

樣本標準(偏)差 $S = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i - \bar{x})^2}{\sum_i^k f_i - 1}} = \sqrt{\frac{24.9825}{57-1}} = \sqrt{0.4461} = 0.6679$

答案：樣本變異數 = 0.4461；樣本標準(偏)差 = 0.6679

3.2.5.1 分組資料的變異數和標準(偏)差【選擇教材】

樣本資料

樣本變異數 $S^2 = \frac{1}{n-1} \times \sum_{i=1}^{k}[(m_i - \bar{x})^2 \times f_i]$

樣本標準(偏)差 $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^{k}[(m_i - \bar{x})^2 \times f_i]}$

其中 n：樣本總數 m_i：i 組之組中點、組中值

k：組數 f_i：i 組之次數、頻率

範例 3.42 研究抽樣樣本的年齡層分布如下表，試計算其樣本變異數(sample variance)和標準(偏)差(standard deviation)？(答案有效位數四捨五入取到小數點後第 1 位)

年齡層(歲)	人數
20~29	12
30~39	26
40~49	33
50~59	26
60~69	21

題解：

年齡層	人數(f_i)	組中點(m_i)	$f_i \times m_i$	$m_i - \bar{x}$	$(m_i - \bar{x})^2$	$(m_i - \bar{x})^2 \times f_i$
20~29	12	24.5	294.0	-21.5	462.25	5547.00
30~39	26	34.5	897.0	-11.5	132.25	3438.50
40~49	33	44.5	1468.5	-1.5	2.25	74.25
50~59	26	54.5	1417.0	8.5	72.25	1878.50
60~69	21	64.5	1354.5	18.5	342.25	7187.25
合計	118		5431.0			18125.50

樣本平均值 $\bar{x} = \frac{\sum_{i=1}^{k} f_i \times m_i}{n} = \frac{5431.0}{118} = 46.0$，故年齡的算術平均值為 46.0 歲

樣本變異數 $S^2 = \frac{1}{n-1} \times \sum_{i=1}^{k}[(m_i - \bar{x})^2 \times f_i] = \frac{18125.50}{118-1} = 154.92$(歲 ^{2})

樣本標準(偏)差 $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^{k}[(m_i - \bar{x})^2 \times f_i]} = \sqrt{154.92} = 12.45$(歲)

答案：樣本變異數 = 154.9 歲 ^{2}；樣本標準(偏)差 = 12.5 歲

練習 3.20 研究抽樣樣本的年齡層分布如下表，試計算其平均值(mean)、變異數(variance)和標準(偏)差(standard deviation)？(答案有效位數四捨五入取到小數點後第 1 位)

年齡層(歲)	人數
20~29	11
30~39	16
40~49	15
50~59	5
60~69	3

3.2.6 變異數與標準(偏)差之性質

變異數(variance)和標準(偏)差(standard deviation)數值愈大代表原始資料(數值)愈分散，數值愈小代表原始資料(數值)愈集中。變異數與標準(偏)差數值皆會受到極端觀測值的影響。每一個觀測值的數值皆會影響變異數和標準(偏)差的數值。

全距概略法則(Range rule of thumb)：一般分布情況下，可以利用已知全距(range)數值估算標準(偏)差的大約值：

$$\text{標準(偏)差(Standard deviation)} \fallingdotseq \frac{\text{全距(range)}}{4} = \frac{\text{最高數值}-\text{最低數值}}{4}$$

若已經知道平均值和標準(偏)差的數值，可以利用上述關係反推估一般正常情況(分布)下的最大值和最小值觀測值：

$$\text{最大觀測值} \fallingdotseq \text{平均值} + 2 \times \text{標準(偏)差}$$

$$\text{最小觀測值} \fallingdotseq \text{平均值} - 2 \times \text{標準(偏)差}$$

3.2.6.1 加或減常數對變異數的影響

當每一個觀測值的數值皆加或減上相同一個常數時，其變異數和標準(偏)差數值皆不會改變。即，當每一個觀測值皆加或減一個常數時，其數值的分散程度不會改變，僅是整個分布曲線向右或向左水平移動。

設原觀測值 x_i 與新觀測值 y_i 分別標示為 $x_1, x_2, x_3, \ldots., x_N$ 及 $y_1, y_2, y_3,\ldots., y_N$

其中 $y_i = x_i + C$，前述 C：常數

兩母體平均值的關係為 μ_y(調整後平均值) = μ_x(原始平均值) + C。

若兩母體變異數分別為σ_x^2(原始變異數)及σ_y^2(調整後變異數)，兩母體變異數和標準差的關係為：$\sigma_x^2 = \sigma_y^2$；σ_x(原始標準偏差) = σ_y(調整後標準偏差)。

$$\sigma_y^2 = \frac{\sum_{i=1}^{N}(y_i-\mu_y)^2}{N} = \frac{\sum_{i=1}^{N}(x_i+C-\mu_x-C)^2}{N} = \frac{\sum_{i=1}^{N}(x_i\cancel{+C}-\mu_x\cancel{-C})^2}{N} = \frac{\sum_{i=1}^{N}(x_i-\mu_x)^2}{N} = \sigma_x^2 \quad \rightarrow \quad \sigma_y = \sqrt{\sigma_y^2} = \sqrt{\sigma_x^2} = \sigma_x$$

範例 3.43 燕巢區一家國際觀光旅館，櫃台五位女性員工身高分別為 160、165、170、172 和 163 公分，請計算該旅館櫃檯五位女性員工身高分布之變異數和標準(偏)差。若每位員工都站在 5 公分高的書本上量身高，其身高依序為 165、170、175、177 和 168 公分，請計算五位女性員工身高分布之變異數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：

x_i	$x_i-\mu$	$(x_i-\mu)^2$
160	-6	36
165	-1	1
170	4	16
172	6	36
163	-3	9
$\sum_{i=1}^{N} x_i = 830$		$\sum_{i=1}^{N}(x_i-\mu)^2 = 98$
$\mu = 166$		

x_i	$x_i-\mu$	$(x_i-\mu)^2$
165	-6	36
170	-1	1
175	4	16
177	6	36
168	-3	9
$\sum_{i=1}^{N} x_i = 855$		$\sum_{i=1}^{N}(x_i-\mu)^2 = 98$
$\mu = 171$		

$$\text{母體標準(偏)差 } \sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{98}{5}} = \sqrt{19.6} = 4.4272 \text{ (公分)}$$

答案：原始身高分布和墊了 5 公分厚的書籍後的身高分布，兩者皆為變異數 = 19.6 公分 ^{2}；標準(偏)差 = 4.4272 公分。原始身高分布和墊高後身高分布，前後變異數和標準(偏)差數值都沒有改變。

3.2.6.2 乘或除常數對變異數的影響

當每一個觀測值的數值皆乘或除以相同一個常數後，對新形成的母體分布之平均值、變異數和標準差影響。

設原觀測值 x_i 與新觀測值 y_i 分別標示為 $x_1, x_2, x_3, ..., x_N$ 及 $y_1, y_2, y_3,...., y_N$

其中 $y_i = C \times x_i$ (即各原觀測值乘一常數 $C \neq 0$)

兩母體平均值的關係為 μ_y(調整後平均值) = $C \times \mu_x$(原始平均值)。

若兩母體變異數分別為σ_x^2(原始變異數)及σ_y^2(調整後變異數)，兩母體變異數和標準差的關係為：$\sigma_y^2 = C^2 \times \sigma_x^2$，$\sigma_y$(調整後標準偏差) = $|C| \times \sigma_x$(原始標準偏差)。

$$\sigma_y^2 = \frac{\sum_{i=1}^{N}(y_i-\mu_y)^2}{N} = \frac{\sum_{i=1}^{N}(C\times x_i-C\times\mu_x)^2}{N} = \frac{C^2\times\sum_{i=1}^{N}(x_i-\mu_x)^2}{N} = C^2 \times \sigma_x^2$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{C^2 \times \sigma_x^2} = \sqrt{C^2} \times \sqrt{\sigma_x^2} = |C| \times \sigma_x\text{(原始標準偏差)}$$

範例 3.44 燕巢區一家國際觀光旅館，櫃台五位女性員工身高分別為 160、165、170、172 和 163 公分，請計算該旅館櫃檯五位女性員工身高分布之變異數和標準(偏)差。若每位員工身高都是以公尺評量，其身高依序為 1.60、1.65、1.70、1.72 和 1.63 公尺，請計算五位女性員工身高分布之變異數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：

x_i	$x_i-\mu$	$(x_i-\mu)^2$
160	-6	36
165	-1	1
170	4	16
172	6	36
163	-3	9
$\sum_{i=1}^{N} x_i = 830$		$\sum_{i=1}^{N}(x_i-\mu)^2 = 98$
$\mu = 166$		

x_i	$x_i-\mu$	$(x_i-\mu)^2$
1.60	-0.06	0.0036
1.65	-0.01	0.0001
1.70	-0.04	0.0016
1.72	-0.06	0.0036
1.63	-0.03	0.0009
$\sum_{i=1}^{N} x_i = 8.30$		$\sum_{i=1}^{N}(x_i-\mu)^2 = 0.0098$
$\mu = 1.66$		

原始身高分布母體標準(偏)差 $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{98}{5}} = \sqrt{19.6} = 4.4272$ (公分)

以公尺評量身高分布母體標準(偏)差 $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{0.0098}{5}} = \sqrt{0.00196} = 0.044272$ (公尺)

原始以公分評量身高分布 × 0.01 = 以公尺評量的身高分布 → 乘以係數 C = 0.01

原始身高分布母體標準(偏)差 σ = 4.4272 (公分) × 0.01 = 以公尺評量身高分布母體標準(偏)差 σ = 0.044272 (公尺) → 乘以係數 C = 0.01

原始身高分布母體變異數 σ^2 = 19.6 (公分 ^{2}) × 0.01^2 = 以公尺評量身高分布母體變異數 σ^2 = 0.00196 (公尺 ^{2}) → 乘以係數 $C^2 = 0.01^2 = 0.0001$

答案	原始身高分布(公分)	身高分布(公尺)	乘以係數
變異數	19.6	0.00196	0.0001
標準偏差	4.4272	0.044272	0.01

範例 3.45 已知觀光系甲班修統計學課程，某次期中考全班同學成績之變異數(variance)為 10 分^2，若將全班每位同學成績皆乘以 2，請估算新成績資料分布的變異數數值。

題解：原來成績分布的變異數$\sigma_x^2 = 10$，新成績分布的變異數σ_y^2，常數 $C = 2$

新成績分布的變異數$\sigma_y^2 = C^2 \times \sigma_x^2 = 2^2 \times 10 = 40$ (分 ^{2})

答案：變異數 = 40 分^2

3.2.6.3 乘常數再加常數對變異數的影響

當每一個原始觀測值的數值皆乘以一個常數再加上另一個常數後，對新形成母體分布之平均值、變異數和標準差的影響。

設原觀測值與新觀測值分別標示為 $x_1, x_2, x_3, ..., x_N$ 及 $y_1, y_2, y_3,...., y_N$

其中 $y_i = C_1 \times x_i + C_2$，其中 $C_1, C_2 \in R$ 屬於實數[1]

兩母體平均值的關係為 μ_y(調整後平均值) = $C_1 \times \mu_x$(原始平均值) + C_2

若兩母體變異數分別為σ_x^2(原始變異數)及σ_y^2(調整後變異數)，兩母體變異數和標準差的關係為：$\sigma_y^2 = C^2 \times \sigma_x^2$，$\sigma_y$(調整後標準偏差) = $|C| \times \sigma_x$(原始標準偏差)。

3.2.6.4 標準(偏)差數值最小化

利用算術平均值$\bar{x}$為依據所計算獲得的樣本標準(偏)差 S，比其他非算術平均值$\bar{x}$所計算獲得的標準(偏)差為小。

$$\text{樣本標準(偏)差 } S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} < \sqrt{\frac{\sum_{i=1}^{n}(x_i - A)^2}{n-1}}\text{，當 } \bar{x} \neq A$$

範例 3.46 燕巢區一家國際觀光旅館，櫃台五位女性員工身高分別為 160、165、170、172 和 163 公分，請計算該旅館櫃檯五位女性員工身高分布之標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：

x_i	$x_i - \mu$	$(x_i - \mu)^2$	$x_i - 165$	$(x_i - 165)^2$	$x_i - 167$	$(x_i - 167)^2$
160	-6	36	-5	25	-7	49
165	-1	1	0	0	-2	4
170	4	16	5	25	3	9
172	6	36	7	49	5	25
163	-3	9	-2	4	-4	16
$\sum_{i=1}^{N} x_i = 830$		$\sum_{i=1}^{N}(x_i - \mu)^2 = 98$		$\sum_{i=1}^{N}(x_i - 165)^2 = 103$		$\sum_{i=1}^{N}(x_i - 167)^2 = 103$
$\mu = 166$						

母體標準(偏)差 $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{98}{5}} = \sqrt{19.6} = 4.4272$ (公分)

以 165 公分為中心點運算偏差 $\sqrt{\frac{\sum_{i=1}^{N}(x_i-165)^2}{N}} = \sqrt{\frac{103}{5}} = \sqrt{20.6} = 4.5387$ (公分)

以 167 公分為中心點運算偏差 $\sqrt{\frac{\sum_{i=1}^{N}(x_i-167)^2}{N}} = \sqrt{\frac{103}{5}} = \sqrt{20.6} = 4.5387$ (公分)

以 164 公分為中心點運算偏差 $\sqrt{\frac{\sum_{i=1}^{N}(x_i-164)^2}{N}} = \sqrt{\frac{118}{5}} = \sqrt{23.6} = 4.8580$ (公分)

以 168 公分為中心點運算偏差 $\sqrt{\frac{\sum_{i=1}^{N}(x_i-168)^2}{N}} = \sqrt{\frac{118}{5}} = \sqrt{23.6} = 4.8580$ (公分)

以 165.5 公分為中心點運算偏差 $\sqrt{\frac{\sum_{i=1}^{N}(x_i-165.5)^2}{N}} = \sqrt{\frac{99.25}{5}} = \sqrt{19.85} = 4.4553$(公分)

答案：與其他數值為中心點相比較後，以算術平均值 166 公分為中心點，運算出來的標準(偏)差 = 4.4272 公分為最小數值

[1] 實數(real number)包括所有的有理數和無理數，如+10、0、-1.2、…、π 等。有理數(rational number)是一個整數 a 和另一個非零整數 b 之比率(ratio)，通常表示成 a/b，故又稱作分數。無理數(irrational number)即非有理數之實數，不能表示為兩整數之比率。若將其表示成小數形式，小數點之後的數字有無限多個，且不會循環現象。常見無理數有大部分的開平方根、圓周率(π)、指數(exponent)等。

3.2.6.5 標準(偏)差大於等於零

一般資料分布情形中變異數和標準(偏)差數值皆大於零，即 $S^2 > 0$ 和 $S > 0$；若 $S = 0$ 表示資料中每一個觀測值(數值)皆相等，代表所有觀測值(數值)都一樣，資料完全沒有分散，僅聚焦在單獨一個點上面。

變異數 $S^2 > 0$；標準(偏)差 $S > 0$，在 $x_i \neq x_j$，當 $i \neq j$

變異數 $S^2 = 0$；標準(偏)差 $S = 0$，在 $x_i = x_j$，當 $i \neq j$

範例 3.47 燕巢區一家國際觀光旅館，櫃台五位女性員工身高分別為 165、165、165、165 和 165 公分，請計算該旅館櫃檯五位女性員工身高分布之變異數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：

x_i	$x_i - \mu$	$(x_i - \mu)^2$
165	0	0
165	0	0
165	0	0
165	0	0
165	0	0
$\sum_{i=1}^{N} x_i = 825$		$\sum_{i=1}^{N}(x_i - \mu)^2 = 0$
$\mu = 165$		

母體標準(偏)差 $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{0}{5}} = \sqrt{0} = 0.0000$ (公分)。觀測值數值都相同時(皆為 165 公分)，其標準偏差即為 0.0000。

答案：身高分布變異數 = 0.00 公分 ^{2}；標準(偏)差 = 0.0000 公分

3.2.6.6 兩組母體資料合併後平均值與標準(偏)差的計算方式

若 A 組資料母體的數量 N_A，平均值 μ_A 和標準(偏)差 σ_A；B 組資料母體的數量 N_B，平均值 μ_B 和標準(偏)差 σ_B。合併後母體數量 $N = N_A + N_B$。

資料	數量	平均值	標準(偏)差	合併	數量	平均值	標準(偏)差
母體A	N_A	μ_A	σ_A	→	N	μ	σ
母體B	N_B	μ_B	σ_B				

AB 兩組資料合併之後的平均值 μ 與標準(偏)差 σ 分別為：

$$\text{平均值 } \mu = \frac{N_A \times \mu_A + N_B \times \mu_B}{N} = \frac{N_A \times \mu_A + N_B \times \mu_B}{N_A + N_B}$$

標準(偏)差 σ 第一種算法

$$\sigma_A^2 = \sum_{i=1}^{N_A}\frac{\left(x_{A_i}-\mu_A\right)^2}{N_A} = \sum_{i=1}^{N_A}\frac{x_{A_i}^2-2\times x_{A_i}\times\mu_A+\mu_A^2}{N_A} = \frac{\sum_{i=1}^{N_A}x_{A_i}^2-2\times\sum_{i=1}^{N_A}\left(x_{A_i}\times\mu_A\right)+\sum_{i=1}^{N}\mu_A^2}{N_A} = \frac{\sum_{i=1}^{N_A}\left(x_{A_i}^2\right)-2\times N_A\times\mu_A^2+N_A\times\mu_A^2}{N_A} =$$
$$\frac{\sum_{i=1}^{N_A}\left(x_{A_i}^2\right)-N_A\times\mu_A^2}{N_A} = \frac{(x_{A_1}^2+x_{A_2}^2+\cdots+x_{AN_A}^2)-N_A\times\mu_A^2}{N_A}$$

等號兩側都乘以 N_A → $\sigma_A^2 \times N_A = \sum_{i=1}^{N_A} x_{A_i}^2 - N_A \times \mu_A^2$

$N_A \times \mu_A^2$ 從等號右側移到左側 → $\sigma_A^2 \times N_A + N_A \times \mu_A^2 = \sum_{i=1}^{N_A} x_{A_i}^2$

等號左側共同有的 N_A 提出來 → $N_A \times (\sigma_A^2 + \mu_A^2) = \sum_{i=1}^{N_A} x_{A_i}^2$

等號左右兩側對調 → $\sum_{i=1}^{N_A} x_{A_i}^2 = N_A \times (\sigma_A^2 + \mu_A^2)$

$\sigma_B^2 = \frac{\left(x_{B1}^2+x_{B2}^2+x_{B3}^2+\cdots+x_{BN_B}^2\right)-N_B\times\mu_B^2}{N_B}$ 同理，類似 A 的變異數推理方式，B 變異數亦可轉換為下列的表達方式

$$\sum_{i=1}^{N_B} x_{B_i}^2 = N_B \times (\sigma_B^2 + \mu_B^2)$$

$$\text{變異數 } \sigma^2 = \frac{\left(\sum_{i=1}^{N_A}x_{A_i}^2+\sum_{i=1}^{N_B}x_{B_i}^2\right)-(N_A+N_B)\times\mu^2}{N_A+N_B} = \frac{N_A\times(\sigma_A^2+\mu_A^2)+N_B\times(\sigma_B^2+\mu_B^2)-(N_A+N_B)\times\mu^2}{N_A+N_B} = \frac{N_A\times(\sigma_A^2+\mu_A^2)+N_B\times(\sigma_B^2+\mu_B^2)}{N_A+N_B} - \mu^2$$

標準(偏)差 $\sigma = \sqrt{\frac{N_A\times(\sigma_A^2+\mu_A^2)+N_B\times(\sigma_B^2+\mu_B^2)}{N_A+N_B}-\mu^2}$

標準(偏)差 σ 第二種算法

變異數 $\sigma^2 = \frac{N_A\times\sigma_A^2+N_A\times(\mu_A-\mu)^2}{N}+\frac{N_B\times\sigma_B^2+N_B\times(\mu_B-\mu)^2}{N} = \frac{N_A\times\sigma_A^2+N_A\times(\mu_A-\mu)^2}{N_A+N_B}+\frac{N_B\times\sigma_B^2+N_B\times(\mu_B-\mu)^2}{N_A+N_B}$

標準(偏)差 $\sigma = \sqrt{\frac{N_A\times\sigma_A^2+N_A\times(\mu_A-\mu)^2}{N}+\frac{N_B\times\sigma_B^2+N_B\times(\mu_B-\mu)^2}{N}} = \sqrt{\frac{N_A\times\sigma_A^2+N_A\times(\mu_A-\mu)^2}{N_A+N_B}+\frac{N_B\times\sigma_B^2+N_B\times(\mu_B-\mu)^2}{N_A+N_B}}$

範例 3.48 甲和乙兩班同時修統計學課程，某次期中考全班成績分別如下所示。試分別估算甲、乙與甲和乙兩班合併計算時，該次期中考成績分布的平均分數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

甲班全班學生成績

80 70 80 90 65 80 75 85 50 60

乙班全班學生成績

75 80 80 80 82 85 85 82 90 95 60 50

題解：甲班母體平均值 $\mu_{甲}$ = 73.50 分，甲班母體標準(偏)差 $\sigma_{甲}$ = 11.63 分，乙班母體平均值 $\mu_{乙}$ = 78.67 分，乙班母體標準(偏)差 $\sigma_{乙}$ = 11.86 分，合併後母體平均值 μ = 76.32 分，合併後母體標準(偏)差 σ = 12.03 分。

合併甲乙班母體平均值 $\mu = \frac{N_{甲}\times\mu_{甲}+N_{乙}\times\mu_{乙}}{N} = \frac{N_{甲}\times\mu_{甲}+N_{乙}\times\mu_{乙}}{N_{甲}+N_{乙}} = \frac{10\times73.50+12\times78.67}{10+12} = \frac{735+944}{22}$ = 76.32 分

合併甲乙班母體標準(偏)差第一種算法 $\sigma = \sqrt{\frac{N_{甲}\times\left(\sigma_{甲}^2+\mu_{甲}^2\right)+N_{乙}\times\left(\sigma_{乙}^2+\mu_{乙}^2\right)}{N_A+N_B}-\mu^2}$

$= \sqrt{\frac{10\times(11.63^2+73.50^2)+12\times(11.86^2+78.67^2)}{10+12}-76.32^2}$ = 12.03 分

合併甲乙班母體標準(偏)差第二種算法 $\sigma = \sqrt{\frac{N_{甲}\times\sigma_{甲}^2+N_{甲}\times\left(\mu_{甲}-\mu\right)^2}{N}+\frac{N_{乙}\times\sigma_{乙}^2+N_{乙}\times\left(\mu_{乙}-\mu\right)^2}{N}} =$

$\sqrt{\frac{10\times11.63^2+10\times(73.50-76.32)^2}{22}+\frac{12\times11.86^2+12\times(78.67-76.32)^2}{22}} = \sqrt{\frac{1352.5+79.42}{22}+\frac{1686.67+66.18}{22}} = \sqrt{\frac{1431.92}{22}+\frac{1752.85}{22}}$

= 12.03 分

答案：甲班母體平均值 = 73.50 分，母體標準(偏)差 = 11.63 分；乙班母體平均值 = 78.67 分，母體標準(偏)差= 11.86 分；合併甲乙班母體平均值 = 76.32 分，母體標準(偏)差 = 12.03 分

範例 3.49 某系甲班學生修統計學課程，某次期中考成績全部 15 位學生平均分數為 80 分，標準(偏)差 5 分，其中男生 5 位學生，平均分數為 75 分，標準(偏)差 3 分，試估算 10 位女生，該次期中考成績分布的平均分數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：母體平均值 $\mu = 80 = \frac{N_A\times\mu_A+N_B\times\mu_B}{N} = \frac{5\times75+10\times\mu_F}{15}$ → $80\times15 = 5\times75+10\times\mu_F$ → 女生平均值 μ_F = 82.5 分

變異數 $\sigma^2 = 5^2 = 25 = \frac{N_A\times\sigma_A^2+N_A\times(\mu_A-\mu)^2}{N}+\frac{N_B\times\sigma_B^2+N_B\times(\mu_B-\mu)^2}{N} = \frac{5\times3^2+5\times(75-80)^2}{15}+\frac{10\times\sigma_F^2+10\times(82.5-80)^2}{15}$ →

$25\times15 = 375 = 45+125+10\times\sigma_F^2+62.5$ → 女生變異數 σ_F^2 = 14.25 分 ² → 女生標準(偏)差 σ_F = 3.7749 分

答案：女生平均值 = 82.50 分，標準(偏)差 = 3.77 分

3.2.6.7 兩組樣本資料合併後平均值與標準(偏)差的計算方式

若 A 組資料樣本的數量 n_A，平均值$\bar{x}_A$和標準(偏)差 S_A；B 組資料樣本的數量 n_B，平均值$\bar{x}_B$和標準(偏)差 S_B。合併後樣本數量 $n = n_A + n_B$。

資料	數量	平均值	標準(偏)差	合併	數量	平均值	標準(偏)差
樣本A	n_A	$\bar{x}_A$	S_A	→	n	$\bar{x}$	S
樣本B	n_B	$\bar{x}_B$	S_B				

AB 兩組資料合併之後的樣本平均值$\bar{x}$與樣本標準(偏)差 s 分別為

$$\text{平均值}\bar{x} = \frac{n_A\times\bar{x}_A+n_B\times\bar{x}_B}{n} = \frac{n_A\times\bar{x}_A+n_B\times\bar{x}_B}{n_A+n_B}$$

$$\text{樣本變異數 } S^2 = \frac{(n_A-1)\times S_A^2+n_A\times(\bar{x}_A-\bar{x})^2}{n-1} + \frac{(n_B-1)\times S_B^2+n_B\times(\bar{x}_B-\bar{x})^2}{n-1} = \frac{(n_A-1)\times S_A^2+n_A\times(\bar{x}_A-\bar{x})^2}{n_A+n_B-1} + \frac{(n_B-1)\times S_B^2+n_B\times(\bar{x}_B-\bar{x})^2}{n_A+n_B-1}$$

$$\text{標準(偏)差 } S = \sqrt{\frac{(n_A-1)\times S_A^2+n_A\times(\bar{x}_A-\bar{x})^2}{n-1} + \frac{(n_B-1)\times S_B^2+n_B\times(\bar{x}_B-\bar{x})^2}{n-1}} = \sqrt{\frac{(n_A-1)\times S_A^2+n_A\times(\bar{x}_A-\bar{x})^2}{n_A+n_B-1} + \frac{(n_B-1)\times S_B^2+n_B\times(\bar{x}_B-\bar{x})^2}{n_A+n_B-1}}$$

範例 3.50 甲和乙兩班同時修統計學課程，某次期中考成績：甲班隨機抽取 10 位學生；乙班隨機抽取 12 位學生，其成績分別如下所示。試分別估算甲、乙與甲和乙兩班合併計算時，該次期中考成績分布的平均分數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

甲班學生成績

80	70	80	90	65	80	75	85	50	60

乙班學生成績

75	80	80	80	82	85	85	82	90	95	60	50

題解：甲班樣本平均分數$\bar{x}_{甲}$ = 73.50 分，甲班分數樣本標準(偏)差 $S_{甲}$ = 12.2588 分，乙班樣本平均分數$\bar{x}_{乙}$ = 78.67 分，乙班分數樣本標準(偏)差 $S_{乙}$ = 12.3828 分，合併計算後樣本平均值$\bar{x}$ = 76.32 分，合併計算後標準(偏)差 S = 12.3149 分。

$$\text{合併甲乙班樣本平均值}\bar{x} = \frac{n_{甲}\times\bar{x}_{甲}+n_{乙}\times\bar{x}_{乙}}{n} = \frac{n_{甲}\times\bar{x}_{甲}+n_{乙}\times\bar{x}_{乙}}{n_{甲}+n_{乙}} = \frac{10\times73.50+12\times78.67}{10+12} = \frac{735+944}{22} = 76.32$$

$$\text{合併甲乙班樣本標準(偏)差 } S = \sqrt{\frac{\left(n_{甲}-1\right)\times S_{甲}^2+n_{甲}\times\left(\bar{x}_{甲}-\bar{x}\right)^2}{n_{甲}+n_{乙}-1} + \frac{\left(n_{乙}-1\right)\times S_{乙}^2+n_{乙}\times\left(\bar{x}_{乙}-\bar{x}\right)^2}{n_{甲}+n_{乙}-1}} =$$

$$\sqrt{\frac{(10-1)\times12.2588^2+10\times(73.50-76.32)^2}{10+12-1} + \frac{(12-1)\times12.3828^2+12\times(78.67-76.32)^2}{10+12-1}} = 12.3149$$

答案：甲班樣本平均值 = 73.50 分，樣本標準(偏)差 = 12.26 分；乙班樣本平均值 = 78.67 分，樣本標準(偏)差 = 12.38 分；合併甲乙班樣本平均值 = 76.32 分，樣本標準(偏)差 = 12.31 分

3.2.6.8 *k* 組資料合併後平均值與標準(偏)差的計算方式

若 k 組資料母體或樣本的個別數量、平均值和標準(偏)差已知，可以估算 k 組資料合併後整體資料的平均值和標準(偏)差。

	平均值	變異數
母體	$\mu = \frac{\sum_{i=1}^{k} N_i\times\mu_i}{\sum_{i=1}^{k} N_i}$	$\sigma^2 = \frac{\sum_{i=1}^{k}[N_i\times\sigma_i^2+N_i\times(\mu_i-\mu)^2]}{\sum_{i=1}^{k} N_i}$
樣本	$\bar{x} = \frac{\sum_{i=1}^{k} n_i\times\bar{x}_i}{\sum_{i=1}^{k} n_i}$	$S^2 = \frac{\sum_{i=1}^{k}[(n_i-1)\times S_i^2+n_i\times(\bar{x}_i-\bar{x})^2]}{(\sum_{i=1}^{k} n_i)-1}$

3.2.6.9 獨立變數之變異數加法定律

若 X 和 Y 是相互獨立的兩個隨機變數，代表兩個隨機變數的相關係數為 0($\rho_{xy} = 0$)，獨立變數之變異數加法定律(addition rule for variances of independent random variables)：

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$$
$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2$$

範例 3.51 若 X 隨機變數有三個基本單位其觀測值分別為 1、2 和 3，變異數為 0.6667；Y 隨機變數有三個基本單位其觀測值分別為 1、4 和 6，變異數為 4.2222，請計算隨機變數 X 加上隨機變數 Y 形成的新隨機變數 Z 的變異數。

題解：隨機變數 X 有三個基本單位觀測值加上隨機變數 Y 有三個基本單位觀測值，形成新的隨機變數 Z 一共有 9 個基本單位觀測值，分別為 2、5、7、3、6、8、4、7 和 9，變異數 4.8889。

依據獨立變數之變異數加法定理

$$\sigma_{z=x+y}^2 = \sigma_x^2 + \sigma_y^2 = 0.6667 + 4.2222 = 4.8889$$

即代表隨機找 X 隨機變數的一個觀測值加上隨機找 Y 隨機變數的一個觀測值，合計後，新形成的分布，其變異數為 4.8889。由此，可以驗證獨立變數之變異數加法定理。

答案：相加後新隨機變數 Z 的變異數 4.8889

範例 3.52 若 X 隨機變數有三個基本單位其觀測值分別為 1、2 和 3，變異數為 0.6667；Y 隨機變數有三個基本單位其觀測值分別為 1、4 和 6，變異數為 4.2222，請計算隨機變數 X 減去隨機變數 Y 形成新隨機變數 Z 分布的變異數。

題解：隨機變數 X 有三個基本單位觀測值減去隨機變數 Y 有三個基本單位觀測值，形成新的隨機變數 Z 一共有 9 個基本單位觀測值，分別為 0、-3、-5、1、-2、-4、2、-1 和-3，變異數 4.8889。

依據獨立變數之變異數加法定理

$$\sigma_{z=x-y}^2 = \sigma_x^2 + \sigma_y^2 = 0.6667 + 4.2222 = 4.8889$$

即代表隨機找 X 隨機變數的一個觀測值減去隨機找 Y 隨機變數的一個觀測值，合計後，新形成隨機變數 $Z = X - Y$ 的分布，其變異數為 4.8889。由此，可以驗證獨立變數之變異數加法定理。

答案：相減後新隨機變數 Z 分布的變異數 4.8889

範例 3.53 若 X 代表男生體重的隨機變數，Y 代表女生體重的隨機變數，假設男生體重與女生體重兩者之間是相互獨立，男生體重分布的變異數 $\sigma_x^2 = 12$ kg^2，女性體重分布的變異數 $\sigma_y^2 = 10$ kg^2，請計算男生體重加女生體重後分布的變異數(答案有效位數四捨五入取到個位數)。

題解：依據獨立變數之變異數加法定理

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 = 12 + 10 = 22 \text{ kg}^2$$

即代表隨機找一位男生與隨機找一位女生，兩位體重合計後，新形成的體重分布，其變異數為 22 kg^2。

答案：男生體重加女生體重後分布的變異數 22 kg^2

3.2.6.10 隨機變數之變異數一般加法定律

若 X 和 Y 是相互不獨立的隨機變數，兩變數之間具有相關係數 $\rho_{xy} \neq 0$，隨機變數之變異數一般加法定律(general addition rule for variances of random variables)：

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2 \times \rho_{xy} \times \sigma_x \times \sigma_y$$
$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2 \times \rho_{xy} \times \sigma_x \times \sigma_y$$

範例 3.54 若 X 代表每位台灣人月平均收入的隨機變數，Y 代表每位台灣人月平均餐飲支出的隨機變數，假設台灣人月平均收入與月平均餐飲支出兩者之間是相互不獨立，台灣人月平均收入的變異數 $\sigma_x^2 = 1200$ 元^2，台灣人月平均餐飲支出的變異數 $\sigma_y^2 = 100$ 元^2，台灣人月平均收

入與月平均餐飲支出的相關係數 ρ_{xy} = 0.2000，請計算台灣人月平均收入扣除月平均餐飲支出後分布的變異數(答案有效位數四捨五入取到個位數)。

題解：依據隨機變數之變異數一般加法定律

$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2 \times \rho \times \sigma_x \times \sigma_y = 1200 + 100 - (2 \times 0.2000 \times \sqrt{1200} \times \sqrt{100}) = 1300 - 138.56 = 1161.44$ 元 ^{2}

即代表隨機找一位台灣人調查其月平均收入與隨機找另一位台灣人調查其月平均餐飲支出，兩位月平均收入扣除月平均餐飲支出後，新形成的分布，其變異數為 1161.44 元 ^{2}。

答案：月平均收入扣除月平均餐飲支出後分布的變異數 1161 元 ^{2}

練習 *3.21* A 和 B 兩班同時修統計學課程，某次期中考成績：A 班全部 45 位學生平均分數為 80 分，標準(偏)差 5 分；B 班全部 50 位學生平均分數為 75 分，標準(偏)差 10 分，試估算 A 和 B 兩班合併計算時，該次期中考成績分布的平均分數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

練習 *3.22* A 和 B 兩班同時修統計學課程，某次期中考成績：A 班全部 15 位學生平均分數為 80 分，標準(偏)差 5 分；B 班全部 12 位學生分數分別為 75、80、80、80、82、85、85、82、90、95、60、50，試估算 A 和 B 兩班合併計算時，該次期中考成績分布的平均分數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

練習 *3.23* 火辣咖啡車上週販售各種咖啡飲料的售價，平均值為新台幣 65.8 元，標準(偏)差新台幣 15.5 元，全球咖啡豆原物料價格飛漲，欲全面性提高每一種咖啡飲品售價 5 %。請估算調高後各種咖啡飲料售價分布之平均值和標準(偏)差。

練習 *3.24* 全球咖啡豆原物料價格飛漲，火辣咖啡車欲調整每一種咖啡飲品的售價，請估算下列兩種方式對於原始售價之平均值、中位數、眾數和標準(偏)差的影響。

(a)每一種咖啡飲品皆調漲新台幣 10 元。

(b)每一種咖啡飲品皆調漲其原始售價的 15 %。

範例 3.55 奇遇餐廳今天每位消費者消費金額(新台幣：元)的次數分布如下表所示，試計算今天每位消費金額分布的平均值、中位數、第 1 四分位數 Q_1、變異數和標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

消費金額	次數
51~60	5
61~70	24
71~80	26
81~90	31
91~100	21
101~110	8

題解：

消費金額	次數 f_i	組中點 m_i	$f_i \times m_i$	以下累積次數	$m_i - \mu$	$(m_i - \mu)^2$	$(m_i - \mu)^2 \times f_i$
51~60	5	55.5	277.5	5	–25.48	649.14	3245.71
61~70	24	65.5	1572.0	29	–15.48	239.58	5749.84
71~80	26	75.5	1963.0	55	-5.48	30.01	780.29
81~90	31	85.5	2650.5	86	4.52	20.45	633.83
91~100	21	95.5	2005.5	107	14.52	210.88	4428.50
101~110	8	105.5	844.0	115	24.52	601.32	4810.53
合計	115		9312.5				19648.70

母體平均值 $\mu = \frac{\sum_{i=1}^{k} f_i \times m_i}{N} = \frac{\sum_{i=1}^{k} f_i \times m_i}{\sum_{i=1}^{k} f_i} = \frac{9312.5}{115} = 80.98$ 元

$O(M_e) = \frac{n}{2} = \frac{115}{2} = 57.5$，因此中位數位於 81~90 組別中

$M_e = L_i + (\frac{n}{2} - F) \times \frac{h_i}{f_i} = 80.5 + (\frac{115}{2} - 55) \times \frac{10}{31} = 80.5 + 2.5 \times \frac{10}{31} = 81.31$ 元

$O(Q_{i=1}) = (i \times \frac{n-1}{4}) + 1 = (1 \times \frac{115-1}{4}) + 1 = 29.5$，$Q_1$ 位於 71~80 組別

$Q_1 = L_i + [(i \times \frac{n-1}{4}) + 1 - F] \times \frac{h_i}{f_i} = 70.5 + [(1 \times \frac{115-1}{4}) + 1 - 29] \times \frac{10}{26} = 70.50 + 0.19 = 70.70$ 元

變異數 $\sigma^2 = \frac{1}{N} \times \sum_{i=1}^{k} (m_i - \mu)^2 \times f_i = \frac{19648.70}{115} = 170.86$ 元 ^{2}

標準(偏)差 $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{k} (m_i - \mu)^2 \times f_i} = \sqrt{170.86} = 13.07$ 元

答案：平均值 = 80.98 元，中位數 = 81.31 元，第 1 四分位數 Q_1 = 70.70 元，變異數 = 170.86 元 ^{2}，標準(偏)差 = 13.07 元

範例 3.56　依據 20 個樣本觀測值，計算獲得平均值為 50，標準(偏)差為 10 後，發現其中有兩個觀測值(數值)35 及 62 必須移除。試估算移除兩個前述觀測值後，(A)平均值與(B)標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：設 n_x = 20 個樣本觀測值的隨機變數為 X，移除兩個樣本後的隨機變數為 Y。

先透過已知算術平均值$\bar{x} = \frac{\sum_{i=1}^{n_x} x_i}{n_x} = \frac{\sum_{i=1}^{20} x_i}{20} = 50$，獲得原始樣本觀測值的和$\sum_{i=1}^{20} x_i = 1000$

移除 2 個樣本觀測值後新組成資料分布的樣本觀測值之和$\sum_{i=1}^{18} y_i = \sum_{i=1}^{20} x_i - x_{19} - x_{20} = 1000 - 35 - 62 = 903$

移除 2 個樣本觀測值後新組成資料分布的樣本觀測值之算術平均值$\bar{y} = \frac{\sum_{i=1}^{n_y} y_i}{n_y} = \frac{\sum_{i=1}^{n_x-2} y_i}{n_x-2} = \frac{\sum_{i=1}^{20-2} y_i}{20-2} = \frac{\sum_{i=1}^{18} y_i}{18} = \frac{903}{18} = 50.16667$

由已知原始觀測值的樣本變異數$S_x^2 = \frac{1}{n-1} \times \left[\sum_{i=1}^{n_x} x_i^2 - \frac{(\sum_{i=1}^{n_x} x_i)^2}{n_x}\right] = 10^2 = 100 = \frac{1}{20-1} \times \left[\sum_{i=1}^{20} x_i^2 - \frac{(1000)^2}{20}\right]$

獲得原始觀測值數值平方和$\sum_{i=1}^{20} x_i^2 = 100 \times 19 + \frac{(1000)^2}{20} = 1900 + 50000 = 51900$

移除 2 個樣本觀測值後新組成資料分布的樣本觀測值之數值平方和$\sum_{i=1}^{18} y_i^2 = \sum_{i=1}^{20} x_i^2 - 35^2 - 62^2 = 51900 - 35^2 - 62^2 = 46831$

移除 2 個樣本觀測值後新組成資料分布的樣本觀測值之變異數$S_y^2 = \frac{\sum_{i=1}^{18} y_i^2 - n_y \times \bar{y}^2}{n_y - 1} = \frac{46831 - 18 \times (50.16667)^2}{18-1} = \frac{46831 - 45300.5}{17} = \frac{1530.5}{17} = 90.0294$

移除 2 個樣本觀測值後新組成資料分布的樣本觀測值之標準(偏)差 $S_y = \sqrt{S_y^2} = \sqrt{90.0294} = 9.4884$

答案：(A)樣本平均值 = 50.17；(B)樣本標準(偏)差 = 9.49

練習 3.25　依據 20 個樣本觀測值，計算獲得平均值為 50，標準(偏)差為 10 後，又發現另有兩個觀測值(數值)35 及 62 必須納入計算。試估算納入兩個前述觀測值後，(A)平均值與(B)標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：設原本 n_x = 20 個樣本觀測值的隨機變數為 X，納入 2 個樣本(觀測值)後的隨機變數為 Y

先透過已知算術平均值$\bar{x} = \frac{\sum_{i=1}^{n_x} x_i}{n_x} = \frac{\sum_{i=1}^{20} x_i}{20} = 50$，獲得原始樣本觀測值的和$\sum_{i=1}^{20} x_i = 1000$

加入 2 個樣本觀測值後新組成資料分布的樣本觀測值之和$\sum_{i=1}^{22} y_i = \sum_{i=1}^{20} x_i + x_{21} + x_{22} = 1000 + 35 + 62 = 1097$

加入 2 個樣本觀測值後新組成資料分布的樣本觀測值之平均值$\bar{y} = \frac{\sum_{i=1}^{n_y} y_i}{n_y} = \frac{\sum_{i=1}^{n_x+2} y_i}{n_x+2} = \frac{\sum_{i=1}^{20+2} y_i}{20+2} = \frac{\sum_{i=1}^{22} y_i}{22} = \frac{1097}{22} = 49.8636$

由已知原始觀測值的樣本變異數$S_x^2 = \frac{1}{n-1} \times \left[\sum_{i=1}^{n_x} x_i^2 - \frac{\left(\sum_{i=1}^{n_x} x_i\right)^2}{n_x}\right] = 10^2 = 100 = \frac{1}{20-1} \times \left[\sum_{i=1}^{20} x_i^2 - \frac{(1000)^2}{20}\right]$

獲得原始觀測值數值平方和$\sum_{i=1}^{20} x_i^2 = 100 \times 19 + \frac{(1000)^2}{20} = 1900 + 50000 = 51900$

加入 2 個樣本觀測值後新組成資料分布的樣本觀測值數值平方和$\sum_{i=1}^{22} y_i^2 = \sum_{i=1}^{20} x_i^2 + 35^2 + 62^2 = 51900 + 35^2 + 62^2 = 56969$

加入 2 個樣本觀測值後新組成資料分布的樣本觀測值變異數$S_y^2 = \frac{\sum_{i=1}^{22} y_i^2 - n_y \times \bar{y}^2}{n_y - 1} = \frac{56969 - 22 \times (49.8636)^2}{22-1} = \frac{56969 - 54700.41}{21} = \frac{2268.59}{21} = 108.0281$

加入 2 個樣本觀測值後新組成資料分布的樣本觀測值標準(偏)差 $S_y = \sqrt{S_y^2} = \sqrt{108.0281} = 10.3937$

答案：(A)樣本平均值 = 49.86；(B)樣本標準(偏)差 = 10.39

練習 3.26 依據 40 個樣本觀測值，計算獲得平均值為 49.21，標準(偏)差為 11.61 後，又發現另有 3 個觀測值(數值)15、82 及 75 必須納入計算。試估算納入三個前述觀測值後，(A)平均值與(B)標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：設原本 $n_x = 40$ 個樣本觀測值的隨機變數為 X，納入 3 個樣本(觀測值)後的隨機變數為 Y。

先透過已知算術平均值$\bar{x} = \frac{\sum_{i=1}^{n_x} x_i}{n_x} = \frac{\sum_{i=1}^{40} x_i}{40} = 49.21$，獲得原始樣本觀測值的和$\sum_{i=1}^{40} x_i = 1968.4$

加入 3 個樣本觀測值後新組成資料分布的樣本觀測值之和$\sum_{i=1}^{43} y_i = \sum_{i=1}^{40} x_i + x_{21} + x_{22} + x_{23} = 1968.4 + 15 + 82 + 75 = 2140.4$

加入 3 個樣本觀測值後新組成資料分布的樣本觀測值之平均值$\bar{y} = \frac{\sum_{i=1}^{n_y} y_i}{n_y} = \frac{\sum_{i=1}^{n_x+3} y_i}{n_x+3} = \frac{\sum_{i=1}^{40+3} y_i}{40+3} = \frac{\sum_{i=1}^{43} y_i}{43} = \frac{2140.4}{43} = 49.7767$

由已知原始觀測值的樣本變異數$S_x^2 = \frac{1}{n-1} \times \left[\sum_{i=1}^{n_x} x_i^2 - \frac{\left(\sum_{i=1}^{n_x} x_i\right)^2}{n_x}\right] = 11.61^2 = 134.7921 = \frac{1}{40-1} \times \left[\sum_{i=1}^{40} x_i^2 - \frac{(1968.4)^2}{40}\right]$

獲得原始觀測值數值平方和$\sum_{i=1}^{40} x_i^2 = 134.7921 \times 39 + \frac{(1968.4)^2}{40} = 5256.892 + 96864.964 = 102121.856$

加入 3 個樣本觀測值後新組成資料分布的樣本觀測值數值平方和$\sum_{i=1}^{43} y_i^2 = \sum_{i=1}^{40} x_i^2 + 15^2 + 82^2 + 75^2 = 102121.856 + 15^2 + 82^2 + 75^2 = 114695.856$

加入 3 個樣本觀測值後新組成資料分布的樣本觀測值變異數$S_y^2 = \frac{\sum_{i=1}^{43} y_i^2 - n_y \times \bar{y}^2}{n_y - 1} = \frac{114695.856 - 43 \times (49.7767)^2}{43-1} = \frac{114695.856 - 106542.1}{42} = \frac{8153.713}{42} = 194.136$

加入 2 個樣本觀測值後新組成資料分布的樣本觀測值標準(偏)差 $S_y = \sqrt{S_y^2} = \sqrt{194.136} = 13.9333$

答案：(A)樣本平均值 = 49.78；(B)樣本標準(偏)差 = 13.93

練習 3.27 依據 20 個母體觀測值，計算獲得平均值為 50，標準(偏)差為 10 後，發現其中有兩個觀測值(數值)35 及 62 必須移除。試估算移除兩個前述觀測值後，(A)平均值與(B)標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：設 20 個母體觀測值的隨機變數為 X，移除兩個基本單位後的隨機變數為 Y。

先透過已知算術平均值 $\mu_x = \frac{\sum_{i=1}^{N_x} x_i}{N_x} = \frac{\sum_{i=1}^{20} x_i}{20} = 50$，獲得原始樣本觀測值的和$\sum_{i=1}^{20} x_i = 1000$

移除 2 個觀測值後新組成母體資料分布之和$\sum_{i=1}^{18} y_i = 1000 - 35 - 62 = 903$

移除 2 個觀測值後新組成母體資料分布之算術平均值 $\mu_y = \frac{\sum_{i=1}^{18} y_i}{18} = \frac{903}{18} = 50.16667$

由已知原始觀測值的母體變異數$\sigma_x^2 = \frac{1}{N_x} \times \left[\sum_{i=1}^{N_x} x_i^2 - \frac{\left(\sum_{i=1}^{N_x} x_i\right)^2}{N_x}\right] = 10^2 = 100 = \frac{1}{20} \times \left[\sum_{i=1}^{20} x_i^2 - \frac{(1000)^2}{20}\right]$

獲得原始觀測值數值平方和$\sum_{i=1}^{20} x_i^2 = 100 \times 20 + \frac{(1000)^2}{20} = 2000 + 50000 = 52000$

移除 2 個觀測值後新組成母體資料分布之數值平方和$\sum_{i=1}^{18} y_i^2 = \sum_{i=1}^{20} x_i^2 - 35^2 - 62^2 = 52000 - 35^2 - 62^2 = 46931$

移除 2 個觀測值後新組成母體資料分布之變異數$\sigma_y^2 = \frac{\sum_{i=1}^{18} y_i^2 - n \times \mu_y^2}{N_y} = \frac{46931 - 18 \times (50.16667)^2}{18} = \frac{46931 - 45300.5}{18} = \frac{1630.5}{18} = 90.58333$

移除 2 個觀測值後新組成母體資料分布之標準(偏)差 $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{90.58333} = 9.5175$

答案：(A)母體平均值 = 50.17；(B)母體標準(偏)差 = 9.52

練習 *3.28* 依據 20 個母體觀測值，計算獲得平均值為 50，標準(偏)差為 10 後，又發現另有兩個觀測值(數值)35 及 62 必須納入計算。試估算納入兩個前述觀測值後，(A)平均值與(B)標準(偏)差。(答案有效位數四捨五入取到小數點後第 2 位)

題解：設 20 個母體觀測值的隨機變數為 X，納入兩個基本單位(觀測值)後的隨機變數為 Y。

原來母體平均值 $\mu_x = \frac{\sum_{i=1}^{N_x} x_i}{N_x} = \frac{\sum_{i=1}^{20} x_i}{20} = 50 \rightarrow \sum_{i=1}^{20} x_i = 50 \times 20 = 1000$

$\sum_{i=1}^{22} y_i = \sum_{i=1}^{20} x_i + y_{21} + y_{22} = 1000 + 35 + 62 = 1097$

納入兩個觀測值後新組成母體資料分布之母體平均值 $\mu_y = \frac{\sum_{i=1}^{N_y} y_i}{N_y} = \frac{\sum_{i=1}^{22} y_i}{22} = \frac{1097}{22} = 49.8636$

原來母體變異數$\sigma_x^2 = \frac{1}{N_x} \times \left[\sum_{i=1}^{N_x} x_i^2 - \frac{\left(\sum_{i=1}^{N_x} x_i\right)^2}{N_x}\right] = 10^2 = 100 = \frac{1}{20} \times \left[\sum_{i=1}^{20} x_i^2 - \frac{(1000)^2}{20}\right]$

$\sum_{i=1}^{20} x_i^2 = 100 \times 20 + \frac{(1000)^2}{20} = 2000 + 50000 = 52000$

$\sum_{i=1}^{22} y_i^2 = \sum_{i=1}^{20} x_i^2 + 35^2 + 62^2 = 52000 + 35^2 + 62^2 = 57069$

納入兩個觀測值後新組成母體資料分布之母體變異數$\sigma_y^2 = \frac{\sum_{i=1}^{22} y_i^2 - N_y \times \mu_y^2}{N_y} = \frac{57069 - 22 \times (49.8636)^2}{22} = \frac{57069 - 54700.41}{22} = \frac{2268.591}{22} = 107.6632$

納入兩個觀測值後新組成母體資料分布之母體標準(偏)差 $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{107.6632} = 10.3761$

答案：(A)母體平均值 = 49.86；(B)母體標準(偏)差 = 10.38

3.2.7 相對分散程度

採用全距、四分位距、平均偏差、變異數、標準(偏)差等變數(參數)評量觀測值(原始資料)的分散程度時，皆會受到平均值和觀測值的**測量單位**的影響，致評量分散程度的數值會有差異，無法客觀比較分析。若欲比較兩組觀測資料的變異性，為達客觀比較之基準，採用分子標準(偏)差與分母平均值單位可相互抵消的比率為分散程度的衡量標準—變異係數(Coefficient of variation: *CV*)。變異係數愈大者，代表其資料的分散程度愈大，變異係數的數值完全不受原始評量單位的影響。變異係數屬於無因次單位。

變異係數 $CV = \frac{\text{標準(偏)差}}{\text{平均值}}$ 母體資料：變異係數 = $\frac{\text{母體標準(偏)差}}{\text{母體平均值}} = \frac{\sigma}{\mu}$ 樣本資料：變異係數 = $\frac{\text{樣本標準(偏)差}}{\text{樣本平均值}} = \frac{S}{\bar{x}}$

範例 3.57 抽樣調查某班級選取其中 5 位學生當樣本，其身高分別為 160, 159, 178, 166, 170 cm；其體重分別為 53, 49, 62, 53, 56 kg，請計算身高和體重的變異係數，並比較何者分散程度較高。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

身高 x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
160	-6.6	43.56
159	-7.6	57.76
178	11.4	129.96
166	-0.6	0.36
170	3.4	11.56
$\Sigma x_i = 833$		$\Sigma(x_i-\bar{x})^2 = 243.2$
$\bar{x} = 166.6$		

體重 x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$
53	-1.6	2.56
49	-5.6	31.36
62	7.4	54.76
53	-1.6	2.56
56	1.4	1.96
$\Sigma x_i = 273$		$\Sigma(x_i-\bar{x})^2 = 93.2$
$\bar{x} = 54.6$		

身高$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{243.2}{5-1}} = \sqrt{60.8} = 7.80$ cm　　身高變異係數$CV = \frac{S}{\bar{x}} = \frac{7.80}{166.6} = 0.0468$

體重$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{93.2}{5-1}} = \sqrt{23.3} = 4.83$ kg　　體重變異係數$CV = \frac{S}{\bar{x}} = \frac{4.83}{54.6} = 0.0884$

身高變異係數 $CV = 0.0468$ < 體重變異係數 $CV = 0.0884$

答案：身高 $CV = 0.0468$；體重 $CV = 0.0884$；體重的分散程度較高

練習 3.29　抽樣調查高雄市選取其中 10 家國際觀光旅館當樣本，專職員工數量分別如下所示，請計算高雄市國際觀光旅館專職員工數之變異係數(Coefficient of variation)。(答案有效位數四捨五入取到小數點後第 4 位)

225　317　325　425　185　265　325　215　198　265

練習 3.30　100 位學生身高分布中，男生平均身高 170 cm，標準(偏)差 5 cm；女生平均身高 165 cm，標準(偏)差 5 cm。試問男生或女生何者相對分散程度較高？

練習 3.31　觀光系一年級有四班，同時修「統計學」課程，某次期中考成績分別為：

班級	人數 N_i	平均成績 μ_i	標準(偏)差 σ_i
4T1A	51	76.50	12.53
4T1B	60	68.23	14.25
4T1C	56	62.28	15.23
4T1D	55	60.28	10.25

請問哪一個班級「統計學」課程學習成效較一致？

練習 3.32　觀光系一年級天使班全班 65 位修統計學學生，期中考試統計學成績最低 5 分，最高 95 分，算術平均值 72.5 分，標準(偏)差為 12.5 分，請計算下列條件的平均值(mean)、標準(偏)差(standard deviation)和變異係數(coefficient of variation)數值。

(a) 全班學習表現良好，每位學生統計學成績皆加 10 分。

(b) 全班學習表現良好，每位學生統計學成績皆乘以 1.2。

(c) 全班學習表現太差，每位學生統計學成績皆扣 15 分。

(d) 全班學習表現太差，每位學生統計學成績皆除以 1.5。

題解：

(a) 平均值 = 72.5 + 10 = 82.5 分；標準(偏)差= 12.5 分；變異係數 CV = $\frac{\sigma}{\mu} = \frac{12.5}{82.5} = 0.1515$

(b) 平均值 = 72.5 × 1.2 = 87.0 分；標準(偏)差= 12.5 × 1.2 = 15.0 分；變異係數 CV = $\frac{\sigma}{\mu} = \frac{15.0}{87.0} = 0.1724$

(c) 平均值 = 72.5 - 15 = 57.5 分；標準(偏)差= 12.5 分；變異係數 CV = $\frac{\sigma}{\mu} = \frac{12.5}{57.5} = 0.2174$

(d) 平均值 = $\frac{72.5}{1.5} = 48.3333$ 分；標準(偏)差= $\frac{12.5}{1.5} = 8.3333$ 分；變異係數 C

練習 3.33 設有一組資料 2、3、7、8、9、9 和 11，其四分位差(inter-quartile range, IQR)為：(A)5；(B)6；(C)7；(D)8。(99 年初等考試統計學大意) B

練習 3.34 在某寒冷的冬天，連續 10 天的溫度都低於 0 度。則此 10 天溫度的標準差：(A)因為每天的溫度都是負的，所以標準差是負的；(B)標準差大於大於或等於 0；(C)因為每天的溫度都是負的，所以標準差不能算；(D)標準差是可正可負的。(99 年初等考試統計學大意) B

動差

所有資料內所有觀測值與算術平均值差異 r 次方的平均值，即稱為 r 級動差

$$M_r = \frac{\sum_{i=1}^{N}(x_i-\mu)^r}{N}$$

3.3 評量相對位置和偵測離群值

3.3.1 z-分布

純純一個月賺新台幣 10 萬元，此數值是高或低，需要一個族群比對其相對數值，才能夠客觀的論定。透過平均值和標準(偏)差可以使用於標定(決定)特定觀測值在一個母體(樣本)分布中的相對位置(Measures of relative location)。評量觀測值(資料)相對位置的表達方法，利用平均值(中心點)為 0，變異數為 1 和標準(偏)差為 1 是標準化的條件。

z-值、z-分布(z-distribution)、z-分數、z-score 或標準化值(standardized value)：可視為觀測值 x_i 數值與樣本平均值$\bar{x}$之間有多少個標準(偏)差的差距。若 $z_5 = 1$，代表觀測值 x_5 高於樣本平均值$\bar{x}$達到 1 個單位標準(偏)差的距離。若 $z_6 = -2$，代表觀測值 x_6 低於樣本平均值$\bar{x}$有 2 個單位標準(偏)差的距離。因此，若 $z_i > 0$，代表觀測值大於平均值 $x_i > \bar{x}$；若 $z_i < 0$，代表觀測值小於平均值 $x_i < \bar{x}$。z-值屬於無因次單位(dimensionless unit)。

$$z_i = \frac{x_i-\bar{x}}{S}$$

其中 x_i：第 i 個觀測值 $\bar{x}$：樣本平均值 S：樣本標準(偏)差

通常一般正常分布的觀測值之標準化 z 值會介於–2 到+2 之間；觀測值之標準化 z 值若高於+2 或低於–2 皆屬於不正常的情況。判斷離群值的一種指標。

範例 3.58 抽樣調查某班級選取其中 5 位學生當樣本，其身高分別為 160、159、178、166 和 170 公分，依據 5 位學生資料當比較基礎，請分別計算個別學生身高對應的標準化 z 值。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

順序	身高 x_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	z_i	$z_i-\bar{z}$	$(z_i-\bar{z})^2$
1	160	–6.6	43.56	–0.8464	-0.8464	0.7164
2	159	–7.6	57.76	–0.9747	-0.9747	0.9500
3	178	11.4	129.96	1.4620	1.4620	2.1375
4	166	–0.6	0.36	–0.0769	-0.0769	0.0059
5	170	3.4	11.56	0.4360	0.4360	0.1901
合計	833		243.2	0.0000		4.0000

樣本平均值$\bar{x} = \frac{\sum_{i=0}^{n}x_i}{n} = \frac{833}{5}$ =166.6 cm；樣本標準(偏)差 $S = \sqrt{\frac{\sum_{i=0}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{243.2}{5-1}} = \sqrt{60.8} = 7.80$ cm

特定一個觀測值 x_1 對應的標準化值 $z_1 = \frac{x_1-\bar{x}}{S} = \frac{160-166.6}{7.80} = -0.8462$

所有樣本觀測值對應的標準化值 z 分布的樣本標準(偏)差 $S_z = \sqrt{\frac{\sum_{i=0}^{n}(z_i-\bar{z})^2}{n-1}} = \sqrt{\frac{4.0000}{5-1}} = \sqrt{1} = 1.0000$ cm

答案：原始樣本觀測值對應的標準化 z 值依序為-0.8464、-0.9747、1.4620、-0.0769 和 0.4360

練習 3.35　抽樣調查高雄市選取其中 10 家國際觀光旅館當樣本，專職員工數量分別如下所示，以此 10 家國際觀光旅館資料當比較基準，請依序計算高雄市各國際觀光旅館專職員工數之標準化 z 值。(答案有效位數四捨五入取到小數點後第 4 位)

K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
254	317	325	425	185	265	325	215	198	260

練習 3.36　假設有 6 個樣本觀測值分別為 1、2、3、4、5 和 30。(A)請計算前述觀測值的 Z 分數(Z-score)；(B)請說明觀測值 30 是否可以視為一個離群值(outlier)？(數值答案有效位數四捨五入取到小數點後第 3 位)

題解：

順序	觀測值 x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	z_i
1	1	-6.5	42.25	-0.5849
2	2	-5.5	30.25	-0.4949
3	3	-4.5	20.25	-0.4049
4	4	-3.5	12.25	-0.3149
5	5	-2.5	6.25	-0.2250
6	30	22.5	506.25	2.0246
合計	45		617.50	0.0000

樣本平均值 $\bar{x} = \frac{\sum_{i=0}^{n} x_i}{n} = \frac{45}{6} = 7.5$；樣本標準差 $S = \sqrt{\frac{\sum_{i=0}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{617.50}{6-1}} = \sqrt{123.5} = 11.1131$

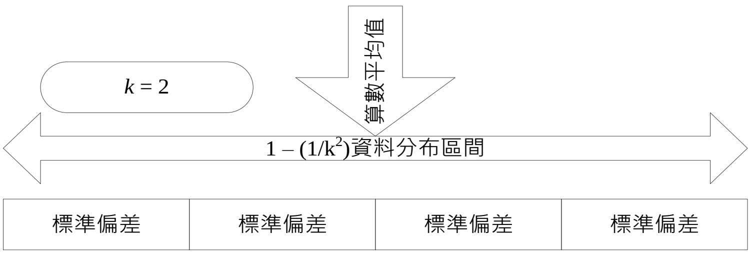
特定一個觀測值 x_1 對應的標準化值 $z_1 = \frac{x_1-\bar{x}}{S} = \frac{1-7.5}{11.1131} = -0.5849$

答案：(A)z 值依序為-0.5849, -0.4949, -0.4049, -0.3149, -0.2250, 2.0246；(B)觀測值 30 因為其 z > 2，故可以判斷為離群值

3.3.2 柴比氏定理

柴比氏定理、柴比雪夫定理或謝比雪夫定理(Chebyshev's theorem, Bienayme-Chebyshev rule)陳述，在任何觀測值的分布資料中，至少有$(1-\frac{1}{k^2})$比率或$(1-\frac{1}{k^2}) \times 100$ %的資料，分布在算術平均值 μ 為中心，$\pm k$ 個標準(偏)差 σ 的範圍內。

樣本資料分布在$\bar{x} \pm k \times S$ 區間內，母體資料分布在 $\mu \pm k \times \sigma$ 區間內，至少有$(1-\frac{1}{k^2})$比率的資料(觀測值)。必須 $k > 1$，k 可以是非整數。



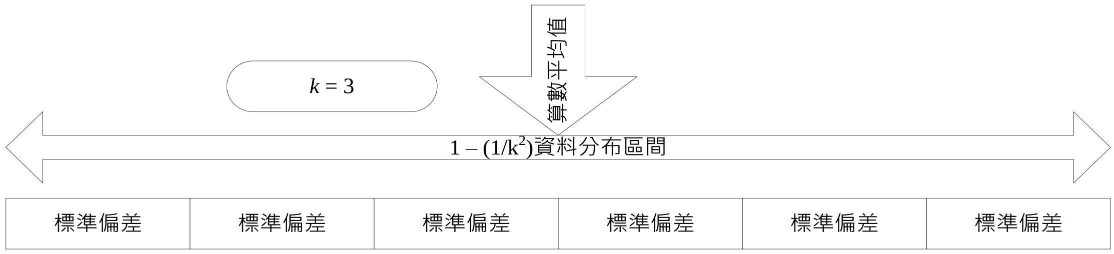


圖 3-6 柴比氏定理說明

柴比氏定理推估的觀測值比率比較保守(少)。

$k = 2$，至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{2^2}) = 0.75$ 比率或 75 %的觀測值，介於以算術平均值為中心，±2 個標準(偏)差 S 的範圍內。

$k = 3$，至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{3^2}) = 0.89$ 比率或 89 %的觀測值，介於以算術平均值為中心，±3 個標準(偏)差 S 的範圍內。

$k = 4$，至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{4^2}) = 0.94$ 比率或 94 %的觀測值，介於以算術平均值為中心，±4 個標準(偏)差 S 的範圍內。

$k = 5$，至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{5^2}) = 0.96$ 比率或 96 %的觀測值，介於以算術平均值為中心，±5 個標準(偏)差 S 的範圍內。

$k = 6$，至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{6^2}) = 0.97$ 比率或 97 %的觀測值，介於以算術平均值為中心，±6 個標準(偏)差 S 的範圍內。

範例 3.59　餐旅系修讀統計學 50 位學生，平常考算術平均分數 65 分，標準(偏)差 2 分。有多少比率的學生分數介於 61 與 69 分之間？有多少比率的學生介於 60 與 70 分之間？(答案有效位數四捨五入取到小數點後第 2 位)

題解：

分數介於 61 與 69 分之間，剛好是以平均分數 65 分為中心(基準)，上下各有 2 個($k = \frac{平均值-分數下限}{標準差} = \frac{分數上限-平均值}{標準差} = \frac{69-65}{2} = \frac{4}{2} = 2$)標準(偏)差(2 分)的間距，透過柴比氏定理，推測至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{2^2}) = 0.75$ 比率或 75 %的學生分數介於 61 與 69 分之間。

分數介於 60 與 70 分之間，剛好是以平均分數 65 分為中心(基準)，上下各有 2.5 個($k = \frac{平均值-分數下限}{標準差} = \frac{分數上限-平均值}{標準差} = \frac{70-65}{2} = \frac{5}{2} = 2.5$)標準(偏)差(2 分)的間距，透過柴比氏定理，推測至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{2.5^2}) =$ 0.84 比率或 84 %的學生分數介於 60 與 70 分之間。

答案：0.75 比率的學生分數介於 61 與 69 分之間；0.84 比率的學生分數介於 60 與 70 分之間

練習 3.37　某班全部 100 位學生參加統計學能力測驗，其平均分數為 65 分，標準(偏)差為 5 分。甲：試估算 45 到 85 分之間的人數？乙：以平均分數為中心點，$\frac{8}{9}$的學生人數之成績範圍？(答案有效位數四捨五入取到小數點後第 2 位)

題解：

分數介於 45 與 85 分之間，剛好是以平均分數 $\mu = 65$ 分為中心點基準，上下各有 4 個($k = \frac{平均值-分數下限}{標準差} = \frac{分數上限-平均值}{標準差} = \frac{85-65}{5} = \frac{20}{5} = 4$)標準(偏)差($\sigma = 5$ 分)的分數間距【下限分數：$\mu - k \times \sigma = 65 - 4 \times 5 = 45$ 分；

上限分數：$\mu + k \times \sigma = 65 + 4 \times 5 = 85$ 分】，透過柴比氏定理，推測至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{4^2}) = 0.94$ 比率或 94 %的學生分數介於 45 與 85 分之間，全部學生有 100 位，94% × 100 = 94 位，因此會有 94 位學生。

透過柴比氏定理，推測至少有$(1-\frac{1}{k^2}) = \frac{8}{9} \rightarrow 1-\frac{8}{9} = \frac{1}{k^2} \rightarrow \frac{9}{9}-\frac{8}{9} = \frac{1}{9} = \frac{1}{k^2} \rightarrow k^2 = 9 \rightarrow k = \sqrt{k^2} = \sqrt{9} = 3$，在標準(偏)差 5 分時，學生分數介於分數下限：平均值 - (k×標準差) = 65 – (3 × 5) = 50 與分數上限：平均值 + (k×標準差) = 65 + (3 × 5) = 80 分之間。

若以平均分數為中心點，90%的學生人數之成績範圍，依據柴比氏定理，推測至少有$(1-\frac{1}{k^2}) = 90\ \%$。$1 - 90\% = 0.1 = \frac{1}{k^2} \rightarrow k^2 = \frac{1}{0.1} = 10 \rightarrow k = \sqrt{k^2} = \sqrt{10} = 3.1623$。在標準(偏)差 5 分時，學生分數介於分數下限：平均值 - (k×標準差) = 65 – (3.1623 × 5) = 49.19 與分數上限：平均值 + (k×標準差) = 65 + (3.1623 × 5) = 80.81 分之間。

答案：45 與 85 分之間有 94 位學生；$\frac{8}{9}$學生之成績範圍 50~80 分之間

3.3.3 經驗法則

若觀測值資料的分布具有鐘型的分布型態，可以依據經驗法則(Empirical rule)，將資料分布的比率與以平均值$\bar{x}$為中心距離標準(偏)差 S 數之間的相關性，以類似在常態分布中，以算術平均值$\bar{x}$為中心，建立以下分布情況：

$\bar{x} \pm S$ 區間內，包含觀測值資料總次數(頻率)的 68.26 %。

$\bar{x} \pm 1.2S$ 區間內，包含觀測值資料總次數(頻率)的 76.99 %。

$\bar{x} \pm 1.5S$ 區間內，包含觀測值資料總次數(頻率)的 86.64 %。

$\bar{x} \pm 2S$ 區間內，包含觀測值資料總次數(頻率)的 95.45 %。

$\bar{x} \pm 3S$ 區間內，包含觀測值資料總次數(頻率)的 99.74 %。

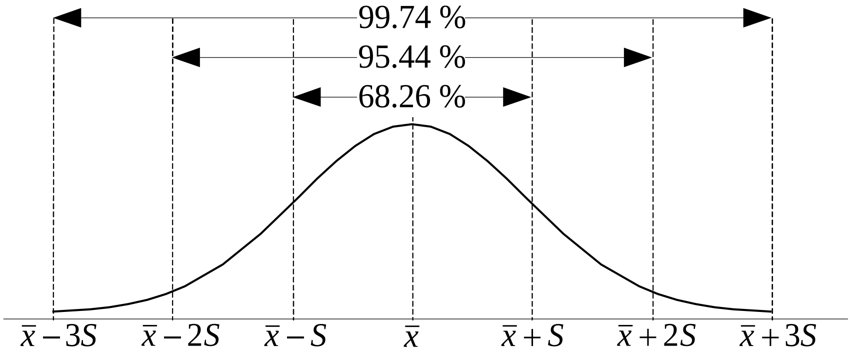


圖 3-7 經驗法則分布

3.3.4 偵測離群值

在觀測值的紀錄資料中，若有少數的觀測值是屬於極端值(極大值、極小值)或異常值(異常大、異常小)，此類的數值稱為異常值或離群值(outlier)。必須判定是否納入後續統計分析，極端值的出現亦有可能是輸入錯誤所致。亦有可能是因為抽樣機會的關係。

3.3.5 標準誤差【選擇教材】

設母體觀測值分別為 2、5 和 8，母體基本單位數 $N = 3$，母體平均值 $\mu = 5$，母體變異數 $\sigma^2 = 6$，從此母體中抽取 $n = 2$ 個觀測值為樣本，其所有可能樣本以及其平均值如下表：

樣本	平均值$\bar{x}$	$(\bar{x}-\mu)^2$	$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$	$S_b^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$

2 , 2	2	9	0	0
2 , 5	3.5	2.25	4.5	2.25
2 , 8	5	0	18	9
5 , 2	3.5	2.25	4.5	2.25
5 , 5	5	0	0	0
5 , 8	6.5	2.25	4.5	2.25
8 , 2	5	0	18	9
8 , 5	6.5	2.25	4.5	2.25
8 , 8	8	9	0	0
和	45	27	54	27
平均值	5	3	6	3
	μ	$\frac{\sigma^2}{n}$	σ^2	

標準誤(差)(standard error, SE)

$$標準誤差\ SE = \sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

標準誤(差)(standard error of mean value, SE)實際上是所有樣本平均值$\bar{x}$抽樣分布的標準(偏)差(standard deviation of sample mean)，表示抽樣時之抽樣誤差(sampling error)。

3.4 探索性資料分析

在探索性資料分析(Exploratory data analysis)時，可以利用圖形瞭解資料分布情況。在製作此圖形時，必須利用五種數值、五數綜合或五量數彙總(five-number summary)進行資料分析：

最小值Q_0	第1四分位數Q_1	第2四分位數Q_2、中位數M_e	第3四分位數Q_3	最大值Q_4

當觀測值(資料)之分布圖型(直方圖)產生分布偏斜傾向時，五數綜合會比算術平均值(mean)及標準(偏)差(standard deviation, SD)的表達和概述(summarization)效果更佳；當觀測值(資料)分布趨近於對稱狀況時，使用算術平均值及標準(偏)差表達或概述效果為佳。

範例 3.60　抽樣調查某班級選取其中 12 位學生當樣本，其身高分別為 162, 159, 178, 166, 170, 160, 122, 158, 188, 167, 156, 163 公分，請計算最小值 Q_0、第 1 四分位數 Q_1、第 2 四分位數 Q_2、第 3 四分位數 Q_3 和最大值 Q_4。(答案有效位數四捨五入取到小數點後第 1 位)

題解：依據身高大小排序為：122, 156, 158, 159, 160, 162, 163, 166, 167, 170, 178, 188

最小值 $Q_1 = 122$

第 1 四分位數 Q_1

$O(Q_1) = (i \times \frac{n-1}{4}) + 1 = (1 \times \frac{12-1}{4}) + 1 = 3.75$，$Q_1$ 落於第 3 和第 4 序位之間。$Q_1 = 158 + 0.75\times(159–158) = 158.75$

第 2 四分位數 Q_2、中位數 M_e

$O(Q_2) = (i \times \frac{n-1}{4}) + 1 = (2 \times \frac{12-1}{4}) + 1 = 6.5$，$Q_2$ 落於第 6 和第 7 序位之間。$Q_2 = 162 + 0.5 \times (163 – 162) = 162.5$

第 3 四分位數 Q_3

$O(Q_3) = (i \times \frac{n-1}{4}) + 1 = (3 \times \frac{12-1}{4}) + 1 = 9.25$，$Q_3$ 落於第 9 和第 10 序位之間。$Q_3 = 167 + 0.25 \times (170 – 167) = 167.75$

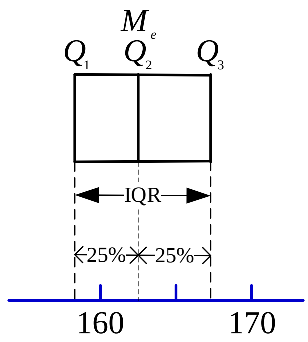
最大值 $Q_4 = 188$

答案：最小值 $Q_0 = 122$；$Q_1 = 158.75$；$Q_2 = 162.5$；$Q_3 = 167.75$；最大值 $Q_4 = 188$

盒鬚圖(Box and whisker plot)

盒鬚圖、盒形圖、盒狀圖、盒圖(Box plot)目的：期望表達出觀測值的分布偏斜程度，同時瞭解離群值的分布狀況。

a.盒鬚圖以第 1 和第 3 四分位數為其左右邊界，左右邊界之間為四分位距 IQR = $Q_3 - Q_1$ = 167.75 – 158.75 = 9，此左右邊界之間應佔觀測值的 50 %資料量。

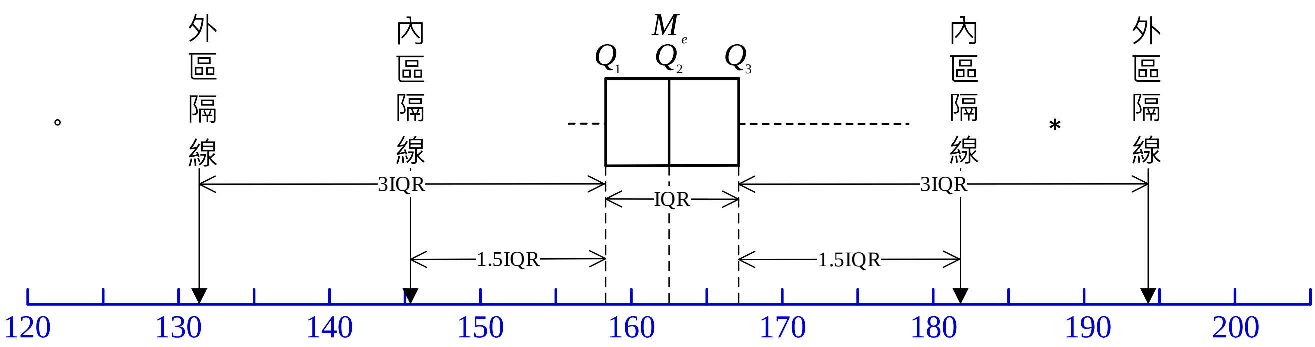


b.盒鬚圖中之垂直線代表第 2 四分位數 Q_2 或中位數 M_e。故垂直線左邊的盒狀圖和右邊的盒狀圖內，所佔觀測值的資料量應相等。

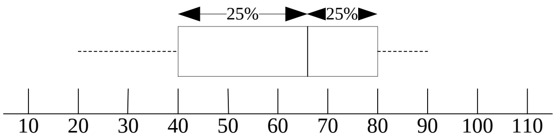
c.利用四分位距 IQR 作為設定區隔(fence)的基本單位。內區隔、內隔離或內籬(Inner fences)為 Q_1 以下 1.5 單位 IQR 與 Q_3 以上 1.5 單位 IQR。即為 $Q_1 - 1.5IQR$ = 158.75 – 1.5 × 9 = 145.25 與 $Q_3 + 1.5IQR$ = 167.75 + 1.5 × 9 = 181.25 區間即為內區隔。外區隔、外隔離或外籬(Outer fences)為 Q_1 以下 3 單位 IQR 與 Q_3 以上 3 單位 IQR。即為 $Q_1 - 3IQR$ = 158.75 – 3 × 9 = 131.75 與 $Q_3 + 3IQR$ = 167.75 + 3 × 9 = 194.75 區間即為外區隔。

d.盒鬚圖中水平虛線為鬚 whiskers，即為在內區隔或內籬區間內最大的觀測值到最小的觀測值。此例內區隔、內籬區間內最大的觀測值為 178；最小觀測值為 156。當觀測值資料數量少時，鬚有可能沒有。

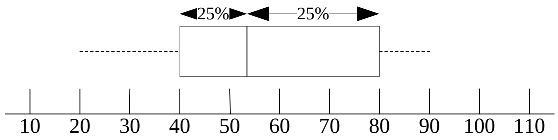
e.利用內區隔和外區隔協助判定是否具有離群值或界外值(Outliers)。若有觀測值介於內區隔線與外區隔線之間，判定為輕微離群值(Mild outliers)；若有觀測值在外區隔線以外者，判定為極度離群值(Extreme outliers)。在觀測值中若有輕微離群值時，在盒鬚圖中標示「*」；若在觀測值中有極度離群值時，在盒鬚圖中標示「。」。此範例：判定為輕微離群值者有觀測值 188；判定為極度離群值者有觀測值 122，已分別在盒鬚圖中標示「*」和「。」。輕微離群值和極度離群值有幾個觀測值，就分別標示幾個符號，亦有可能沒有輕微離群值和極度離群值，就不用標示符號。



透過盒鬚圖可以簡單的瞭解資料分布的情況，以上圖為例：資料分布趨近於左右對稱(常態分布)，但資料分布稍微右偏或正偏。在盒鬚圖中欲辨識資料分布的左偏或右偏分布，可以參閱下列圖形。



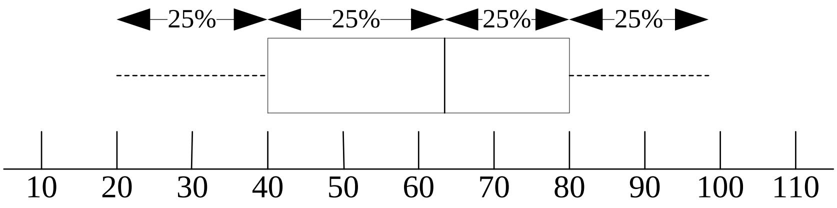
左偏分布



右偏分布

繪製盒鬚圖不需計算母體(樣本)分布的算術平均值和標準(偏)差數值，即可瞭解資料分布的情況。

在盒鬚圖中若沒有離群值時，左右兩個鬚範圍內會各佔有 25%的資料量(如下圖所示)；盒鬚圖中若有離群值時，左右兩個鬚範圍內會少於 25%的資料量，必須由 Q_3 到最高離群值才會有 25%的資料量，由 Q_1 到最低離群值會有 25%的資料量。



練習 3.38　為了瞭解會逛六合夜市的消費者年齡結構，隨機選取 20 位消費者，其年齡(歲)如下所示：15, 26, 35, 41, 18, 19, 18, 17, 16, 32, 55, 49, 11, 24, 26, 28, 35, 45, 34, 41，請繪製：(A)枝葉圖；(B)盒鬚圖。

3.5 兩個變數相關性評量

一個基本單位有兩個配對變數之間(雙變量資料)的關係，可以分為是否有相關，及相關的程度兩部分。線性的共變關係即是兩個變數之間存在一個變數變動會導致另一個變數變動。若一個變數增加導致另一個變數亦增加，屬於正向共變關係；若一個變數增加導致另一個變數的減少，屬於反向共變關係。評量兩個變數相關程度時，可以採用共變異數、共變數(Covariance)或相關係數(Coefficient of correlation)進行測量。非線性的共變關係即是兩個變數之間存在指數關係、對數關係、拋物線、圓形關係等。

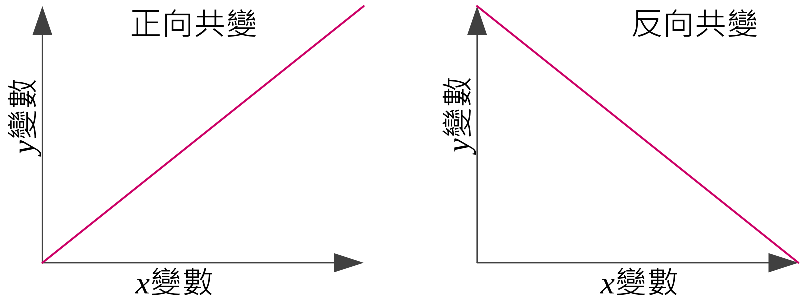


圖 3-8 正向共變和反向共變對照圖

3.5.1 共變異數【選擇教材】

母體資料中變數 X 和變數 Y 的共變(異)數(Population covariance)以 $Cov(X, Y)$或 σ_{xy} 符號表示，共變異數單位為「X 觀測值單位」×「Y 觀測值單位」：

母體共變(異)數 $Cov(X,\ Y) = \sigma_{xy} = \frac{\sum_{i=1}^{N}[(x_i-\mu_x)\times(y_i-\mu_y)]}{N} = \frac{\sum_{i=1}^{N}[x_i\times y_i - x_i\times\mu_y - \mu_x\times y_i + \mu_x\times\mu_y]}{N} = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \frac{\sum_{i=1}^{N}x_i\times\mu_y}{N} - \frac{\sum_{i=1}^{N}\mu_x\times y_i}{N} + \frac{\sum_{i=1}^{N}\mu_x\times\mu_y}{N} = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \mu_x\times\mu_y - \mu_x\times\mu_y + \frac{N\times\mu_x\times\mu_y}{N} = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \mu_x\times\mu_y - \mu_x\times\mu_y + \mu_x\times\mu_y = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \mu_x\times\mu_y = E(X\times Y) - E(X)\times E(Y)$

樣本資料中變數 X 和變數 Y 的共變(異)數(Sample covariance)以 $cov(X, Y)$或 S_{xy} 符號表示，共變異數單位為「X 觀測值單位」×「Y 觀測值單位」：

樣本共變(異)數 $cov(X, Y) = S_{xy} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1}$

共變異數(Covariance)可能出現的三種情境：

1.共變異數大於 0(正值)，表示 x_i 與 y_i 有正向的線性關係，當 x_i 值愈高時，y_i 值亦愈高。

2.共變異數趨近於 0，表示 x_i 和 y_i 之間無線性關係存在，x_i 與 y_i 值沒有線性的共變關係；共變異數趨近於 0 時，亦有可能存在拋物線、鐘形、圓形的關係，故當共變異數趨近於 0 時，僅能推測 x_i 與 y_i 值之間沒有線性關係存在，不能推論 x_i 與 y_i 值之間沒有關係存在，其有可能具有圓形的關係。

3.共變異數小於 0(負值)，表示 x_i 與 y_i 有反向的線性關係，當 x_i 值愈高時，y_i 值愈低。

共變異數(Covariance)特性：

1.共變異數絕對值的大小，代表線性化程度的高低或共同變異程度(關係)的大小。

2.共變異數數值高低會受 X 與 Y 觀測值(數值)評量單位的影響。

3.共變異數數值有受 X 與 Y 觀測值(數值)評量單位影響，其意涵解釋不易。

4.將其中一個變數乘(除)以一個常數(C)，另一個變數不動，其新的共變異數為原始共變異數乘(除)該常數。

新樣本共變(異)數 $cov(X, \mathrm{C}\times Y) = S_{x\mathrm{C}\times y} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(\mathrm{C}\times y_i-\mathrm{C}\times\bar{y})]}{n-1} = \frac{\mathrm{C}\times\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1} = \mathrm{C}\times S_{xy}$

範例 3.61　抽樣調查某班級選取其中 10 位學生當樣本，其身高依序分別為 162, 159, 178, 166, 170, 160, 122, 158, 188, 167 cm，體重依序分別為 60, 59, 65, 63, 64, 50, 35, 55, 70, 65 kg，請計算身高與體重的共變異數(Covariance)。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

學生編號	身高 x_i	體重 y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})\times(y_i-\bar{y})$
1	162	60	–1	1.4	–1.4
2	159	59	–4	0.4	–1.6
3	178	65	15	6.4	96.0
4	166	63	3	4.4	13.2
5	170	64	7	5.4	37.8
6	160	50	–3	–8.6	25.8
7	122	35	–41	–23.6	967.6
8	158	55	–5	–3.6	18.0
9	188	70	25	11.4	285.0
10	167	65	4	6.4	25.6
合計	1630	586	0	0.0	1466.0

身高平均值$\bar{x} = \frac{\sum_{i=1}^{n}x_i}{n} = \frac{1630}{10} = 163$ cm，體重平均值$\bar{y} = \frac{\sum_{i=1}^{n}y_i}{n} = \frac{586}{10} = 58.6$ kg

共變異數 $cov(X, Y) = S_{xy} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1} = \frac{1466}{10-1} = 162.89$ (cm×kg)

答案：共變異數 = 162.89 cm×kg

練習 3.39　延續前一題，變更 X 與 Y 變數的單位，重新計算共變異數(covariance)。身高變成 m(公尺)，體重變成 g(公克)。(答案有效位數四捨五入取到小數點後第 4 位)

練習 3.40　抽樣調查深水城市居民選取其中 12 位居民當樣本，其平均週交通支出和平均週餐飲支出(單位：新台幣元)依序列於下表，請計算平均週交通支出與平均週餐飲支出的共變異數(Covariance)。(答案有效位數四捨五入取到個位數)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
交通	1620	590	2530	2580	440	356	2750	1500	1652	2552	565	421
餐飲	1623	1256	1751	1762	842	895	1743	1542	1610	1655	1045	913

題解：

居民編號	交通 x_i	餐飲 y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})\times(y_i-\bar{y})$
1	1620	1623	157.0	236.6	37143.6
2	590	1256	-873.0	-130.4	113853.8
3	2530	1751	1067.0	364.6	389010.4
4	2580	1762	1117.0	375.6	419526.6
5	440	842	-1023.0	-544.4	556938.3
6	356	895	-1107.0	-491.4	543998.3
7	2750	1743	1287.0	356.6	458922.8
8	1500	1542	37.0	155.6	5756.6
9	1652	1610	189.0	223.6	42257.3
10	2552	1655	1089.0	268.6	292487.3
11	565	1045	-898.0	-341.4	306592.2
12	421	913	-1042.0	-473.4	493300.2
合計	17556	16637	0.0	0.0	3659787.0

交通支出平均值$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{17556}{12}=1463.0$ 元；餐飲支出平均值$\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{16637}{12}=1386.4$ 元

共變異數 $cov(X,Y)=S_{xy}=\frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1}=\frac{3659787.0}{12-1}=332707.9$ 元 ^{2}

答案：共變異數 covariance = 332708 元 ^{2}

Excel 操作函數

在 Excel 2007 軟體內，利用插入→函數→COVAR，計算母體共變異數數值。若要計算的是樣本共變異數(sample covariance)時，將工作表格中的數值乘上$\frac{n}{n-1}$。若以上述範例：樣本共變異數 $= 146.6\times\frac{10}{10-1}=$ 162.89。

在 Excel 2010 軟體內，利用插入→函數→COVARIANCE.P(母體共變數)或 COVARIANCE.S(樣本共變數)，計算共變異數數值。

利用 Excel 軟體中工具(T)→資料分析(D)…→在資料分析對話方塊中選取分析工具(A)共變數→確定。在共變數對話方塊中，輸入範圍(I)拖拉選擇欲統計比較的數值區間，其分組方式選擇逐欄(C)，輸出選項勾選新工作表(P):→確定。即在新增加的工作表中出現選取數值區間內各欄之間的母體共變異數(population covariance)，若要計算的是樣本共變異數(sample covariance)時，將工作表格中的數值乘上$\frac{n}{n-1}$。若以上述範例：樣本共變異數 $= 146.6\times\frac{10}{10-1}=162.89$。

	欄 1	欄 2
欄 1	265.6	
欄 2	146.6	90.64

3.5.2 皮爾森積差相關係數

為了避免 X 與 Y 變數數值之「單位」差異對共變異數(covariance)數值的影響，採取相關係數評量 X 與 Y 變數的相關程度。皮爾森積差相關係數(Pearson product moment correlation coefficient)、皮爾森相關係數或相關係數(Coefficient of correlation)係將原本共變異數(covariance)之單位與標準(偏)差(standard deviation)之單位透過分子和分母單位相互抵銷的方式，獲得此皮爾森積差相關係數為無因次單位(dimensionless unit)，故計算皮爾森積差相關係數時皆不會受到原本 X 與 Y 變數評量單位的影響。

母體資料中變數 X 和變數 Y 的皮爾森積差相關係數(Pearson product moment correlation coefficient; population correlation coefficient)以 ρ_{xy} 符號表示(讀音：rho)：

母體相關係數 $\rho_{xy} = \frac{Cov(X,Y)}{\sigma_x \times \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} = \frac{\frac{\sum_{i=1}^{N}(x_i-\mu_x)\times(y_i-\mu_y)}{N}}{\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu_x)^2}{N}} \times \sqrt{\frac{\sum_{i=1}^{N}(y_i-\mu_y)^2}{N}}} = \frac{\sum_{i=1}^{N}(x_i-\mu_x)\times(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i-\mu_x)^2} \times \sqrt{\sum_{i=1}^{N}(y_i-\mu_y)^2}} = \frac{\sum_{i=1}^{N}x_i \times y_i - \frac{\sum_i^N x_i \times \sum_i^N y_i}{N}}{\sqrt{\sum_{i=1}^{N}x_i^2 - \frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}} \times \sqrt{\sum_{i=1}^{N}y_i^2 - \frac{\left(\sum_{i=1}^{N}y_i\right)^2}{N}}} = \frac{E(X\times Y)-E(X)\times E(Y)}{\sqrt{V(X)}\times\sqrt{V(Y)}}$

其中 $Cov(X, Y) = \sigma_{xy}$：母體資料觀測值 X 與觀測值 Y 的共變異數

σ_x 和 σ_y：母體資料觀測值 X 與觀測值 Y 的母體標準(偏)差(population standard deviation)

樣本資料中變數 X 和變數 Y 的皮爾森積差相關係數(Pearson product moment correlation coefficient; sample correlation coefficient)以 r_{xy} 或 γ_{xy} 符號表示：

樣本相關係數 $r_{xy} = \frac{cov(x,y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} \times \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{n}x_i \times y_i - \frac{\sum_i^n x_i \times \sum_i^n y_i}{n}}{\sqrt{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}} \times \sqrt{\sum_{i=1}^{n}y_i^2 - \frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}}}$

其中 $cov(X,Y) = S_{xy}$：樣本資料共變異數(sample covariance)

S_x 和 S_y：分別為變數 X 和變數 Y 的樣本資料標準(偏)差(sample standard deviation)

相關係數(Coefficient of correlation)特性：

1.相關係數的數值分布介於–1 與 1 之間。
2.相關係數只解釋兩變數之間，線性(直線)關係的強度。
3.相關係數不受自變數與依變數認定的影響。即將 x_i 與 y_i 變量順序前後對調，其相關係數數值不會改變。
4.相關係數會受到極端值(離群值)的影響。
5.相關係數 = 1，代表 x_i 與 y_i 變量之間的呈現正向完全直線之關係，每一對 x_i 與 y_i 變量皆落在斜率為正的直線上。x_i 與 y_i 變量之間的變動方向一致，當 x_i 變量增加時，y_i 變量亦同時增加。
6.相關係數 = –1，代表 x_i 與 y_i 變量之間的呈現反向完全直線之關係，每一對 x_i 與 y_i 變量皆落在斜率為負的直線上。x_i 與 y_i 變量之間的變動方向相反，當 x_i 變量增加時，y_i 變量同時減少。
7.相關係數 = 0，代表 x_i 與 y_i 變量之間無直線關係，可能是 x_i 與 y_i 變量之間無關係、x_i 與 y_i 變量之間具有拋物線關係、圓形關係、非線性關係。
8.其中一個變數觀測值數值統一加(減)一個常數，對相關係數數值沒有影響。一個變數觀測值統一加(減)一個常數，另一個變數觀測值統一加(減)另一個常數，對相關係數數值也沒有影響。
9.其中一個變數觀測值數值統一乘(除)一個常數，對相關係數數值沒有影響。一個變數觀測值統一乘(除)一個常數，另一個變數觀測值統一乘(除)另一個常數，對相關係數數值也沒有影響。

範例 3.62　抽樣調查某班級選取其中 10 位學生當樣本，其身高依序分別為 162, 159, 178, 166, 170, 160, 122, 158, 188, 167 cm，體重依序分別為 60, 59, 65, 63, 64, 50, 35, 55, 70, 65 kg，請計算身高與體重的相關係數(Coefficient of correlation)。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

順序	x_i	y_i	x_i^2	y_i^2	$x_i \times y_i$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\times(y_i-\bar{y})$
1	162	60	26244	3600	9720	-1	1	1.4	1.96	-1.4
2	159	59	25281	3481	9381	-4	16	0.4	0.16	-1.6
3	178	65	31684	4225	11570	15	225	6.4	40.96	96.0
4	166	63	27556	3969	10458	3	9	4.4	19.36	13.2
5	170	64	28900	4096	10880	7	49	5.4	29.16	37.8
6	160	50	25600	2500	8000	-3	9	-8.6	73.96	25.8
7	122	35	14884	1225	4270	-41	1681	-23.6	556.96	967.6
8	158	55	24964	3025	8690	-5	25	-3.6	12.96	18.0
9	188	70	35344	4900	13160	25	625	11.4	129.96	285.0
10	167	65	27889	4225	10855	4	16	6.4	40.96	25.6
合計	1630	586	268346	35246	96984	0	2656	0.0	906.40	1466.0

身高平均值$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1630}{10} = 163$　　體重平均值$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{586}{10} = 58.6$

共變異數 $cov(x,y) = S_{xy} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1} = \frac{1466}{10-1} = 162.89$

身高 $S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{2656}{10-1}} = \sqrt{295.11} = 17.18$　　體重 $S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}} = \sqrt{\frac{906.4}{10-1}} = \sqrt{100.71} = 10.04$

相關係數 $r_{xy} = \frac{cov(x,y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{162.89}{17.18\times10.04} = 0.9444$

另一種公式算法：相關係數 $r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_i^n x_i \times \sum_i^n y_i}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}} \times \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}}} = \frac{96984-\frac{1630\times586}{10}}{\sqrt{268346-\frac{1630^2}{10}}\times\sqrt{35246-\frac{586^2}{10}}} =$
$\frac{96984-95518}{\sqrt{268346-265690}\times\sqrt{35246-34339.6}} = \frac{1466}{51.5364\times30.1065} = 0.9448$

答案：相關係數 = 0.9448

練習 *3.41*　抽樣調查深水城市居民選取其中 12 位居民當樣本，其平均週交通支出和平均週餐飲支出(單位：新台幣元)依序列於下表，請計算平均週交通支出與平均週餐飲支出的相關係數(Coefficient of correlation)。(答案有效位數四捨五入取到小數點後第 4 位)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
交通	1620	590	2530	2580	440	356	2750	1500	1652	2552	565	421
餐飲	1023	1156	1751	1762	842	1195	1743	1542	1610	1655	1045	913

題解：

順序	x_i	y_i	x_i^2	y_i^2	$x_i \times y_i$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})\times(y_i-\bar{y})$
1	1620	1023	2624400	1046529	1657260	157	24649	-330.08	108955.01	-51823.1
2	590	1156	348100	1336336	682040	-873	762129	-197.08	38841.84	172053.8
3	2530	1751	6400900	3066001	4430030	1067	1138489	397.92	158337.67	424577.1
4	2580	1762	6656400	3104644	4545960	1117	1247689	408.92	167212.84	456759.9
5	440	842	193600	708964	370480	-1023	1046529	-511.08	261206.17	522838.3
6	356	1195	126736	1428025	425420	-1107	1225449	-158.08	24990.34	174998.3
7	2750	1743	7562500	3038049	4793250	1287	1656369	389.92	152035.01	501822.8
8	1500	1542	2250000	2377764	2313000	37	1369	188.92	35689.51	6989.9
9	1652	1610	2729104	2592100	2659720	189	35721	256.92	66006.17	48557.3
10	2552	1655	6512704	2739025	4223560	1089	1185921	301.92	91153.67	328787.3
11	565	1045	319225	1092025	590425	-898	806404	-308.08	94915.34	276658.8
12	421	913	177241	833569	384373	-1042	1085764	-440.08	193673.34	458566.8
合計	17556	16237	35900910	23363031	27075518	0	10216482	0.00	1393016.92	3320787.0

交通支出平均值$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{17556}{12} = 1463$　　餐飲支出平均值$\bar{y} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{16237}{12} = 1353.083$

共變異數 $cov(x,y) = S_{xy} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1} = \frac{3320787}{12-1} = 301889.7$

交通支出標準(偏)差 $S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{10216482}{12-1}} = \sqrt{928771.1} = 963.73$

餐飲支出標準(偏)差 $S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}} = \sqrt{\frac{1393016.92}{12-1}} = \sqrt{126637.9} = 355.86$

相關係數 $r_{xy} = \frac{cov(x,y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{301889.7}{963.73 \times 355.86} = 0.8803$

答案：相關係數 = 0.8803

練習 3.42　若 Var(X) = 15.21、Var(Y) = 4.25 和 Cov(X,Y) = -6.25，請計算 (A) X 和 Y 的相關係數；(B) $3X + 5$ 和 $4Y – 7$ 的相關係數。(答案有效位數四捨五入取到小數點後第 3 位)

題解：

(A) $\rho_{xy} = \frac{Cov(x,y)}{\sigma_x \times \sigma_y} = \frac{Cov(x,y)}{\sqrt{Var(x) \times Var(y)}} = \frac{-6.25}{\sqrt{15.21 \times 4.25}} = \frac{-6.25}{8.040056} = -0.7774$

(B) Cov(3X+5,4Y-7) = 3 × 4 × Cov(X,Y) = 12 × -6.25 = -75.00

Var(3X + 5) = 3^2 × Var(X) = 9 × 15.21 = 136.89

Var(4Y – 7) = 4^2 × Var(Y) = 16 × 4.25 = 68.00

$\rho_{xy} = \frac{Cov(3x+5,4y-7)}{\sigma_{3x+5} \times \sigma_{4y-7}} = \frac{Cov(3x+5,4y-7)}{\sqrt{Var(3x+5) \times Var(4y-7)}} = \frac{-75.00}{\sqrt{136.89 \times 68.00}} = \frac{-75.00}{96.48067} = -0.7774$

答案：(A)-0.777；(B)-0.777

範例 3.63　已知 X 與 Y 兩個變數，相關資訊：$\sum_{i=1}^{n} x_i = 120$，$\sum_{i=1}^{n} x_i^2 = 45662$，$\sum_{i=1}^{n} x_i \times y_i = 21000$，$\sum_{i=1}^{n} y_i = 140$，$\sum_{i=1}^{n} y_i^2 = 34562$，樣本數量 n = 100，請計算 X 與 Y 兩個變數的相關係數。(答案有效位數四捨五入取到小數點後第 4 位)

題解：

相關係數 $r_{xy} = \frac{cov(x,y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{\sum_{i=1}^{n} x_i \times y_i - \frac{\sum_i^n x_i \times \sum_i^n y_i}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \times \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}} = \frac{21000 - \frac{120 \times 140}{100}}{\sqrt{45662 - \frac{120^2}{100}} \times \sqrt{34562 - \frac{140^2}{100}}} = \frac{21000-168}{\sqrt{45662-144} \times \sqrt{34562-196}} = \frac{20832}{\sqrt{45518} \times \sqrt{34366}} = \frac{20832}{213.3495 \times 185.3807} = 0.5267$

答案：相關係數 r_{xy} = 0.5267

範例 3.64　小魚連鎖咖啡館蒐集 12 個營業點資料，分析人事費用 x_i(單位：新台幣百元)與銷售金額 y_i(單位：新台幣百元)的相關係數。獲得相關係數為 0.5678。若將原來人事費用資料由單位是新台幣百元，調整為新台幣元；同時，銷售金額單位不變，請計算調整單位後的相關係數。

題解：將原來人事費用資料由單位是新台幣百元，調整為新台幣元。假設人事費用調整單位前後的變量分別標示為 x_i 和 x_{in}，即 x_{in}(調整後變量) = 100 × x_i(調整前變量)。人事費用調整單位前後的平均值分別標示為$\bar{x}$和$\bar{x}_n$，即$\bar{x}_n$(調整後平均值) = 100 × $\bar{x}$(調整前平均值)。

相關係數 $r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x}) \times (y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = 0.5678$

相關係數$r_{x_n y} = \frac{\sum_{i=1}^{n}(x_{in}-\bar{x}_n) \times (y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_{in}-\bar{x}_n)^2} \times \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{n}(100 \times x_i - 100 \times \bar{x}) \times (y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(100 \times x_i - 100 \times \bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{100 \times \left[\sum_{i=1}^{n}(x_i-\bar{x}) \times (y_i-\bar{y})\right]}{100 \times \left[\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right]} = \frac{\sum_{i=1}^{n}(x_i-\bar{x}) \times (y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = r_{xy} = 0.5678$

答案：調整後的相關係數 = 0.5678，對無因次單位(相關係數)而言，調整評量時的單位(人事費用資料由單位是新台幣百元，調整為新台幣元)，對無因次單位(相關係數)的數值沒有影響，原來是相關係數為 0.5678，調整單位後相關係數還是 0.5678。原變數觀測值數值統一乘(除)一個常數，對相關係數數值沒有影響。

範例 3.65　現有 X 與 Y 兩個隨機變數，並分別以 E 與 V 各代表其期望值與變異數。已知 E(X) = 5，E(Y) = 6，E(XY) = 21，V(X) = 9，V(Y) = 10，請計算 x 與 y 兩個隨機變數的相關係數。

題解：期望值即是重複多次測量特定隨機變數，用所獲得的數值計算之平均值。期望值具備線性函數特性具有相加特性。$E[aX + bY] = a\times E[X] + b\times E[Y]$。其中：$X$ 和 Y 代表在同一個機率空間內的兩個隨機變數(相互獨立或非相互獨立)，a 和 b 為實數。

$$E(XY) = \sum_{i=1}^{N}\frac{x_i\times y_i}{N} = \frac{1}{N}\times\sum_{i=1}^{N}x_i\times y_i = 21$$

$$V(X) = \sigma_x^2 = \sum_{i=1}^{N}\frac{(x_i-\mu_x)^2}{N} = \frac{\sum_{i=1}^{N}(x_i-\mu_x)^2}{N} = \frac{1}{N}\times\sum_{i=1}^{N}(x_i-\mu_x)^2 = \frac{1}{N}\times\sum_{i=1}^{N}(x_i^2-2\times x_i\times\mu_x+\mu_x^2) = \frac{1}{N}\times[\sum_{i=1}^{N}x_i^2-2\times\sum_{i=1}^{N}x_i\times\mu_x+\sum_{i=1}^{N}\mu_x^2] = \frac{1}{N}\times[\sum_{i=1}^{N}x_i^2-2\times N\times\mu_x\times\mu_x+N\times\mu_x^2] = \frac{1}{N}\times[\sum_{i=1}^{N}x_i^2-N\times\mu_x^2] = \frac{\sum_{i=1}^{N}x_i^2}{N}-\mu_x^2 = \frac{1}{N}\times\left[\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}\right] = 9$$

$$V(Y) = \sigma_y^2 = \sum_{i=1}^{N}\frac{(y_i-\mu)^2}{N} = \frac{\sum_{i=1}^{N}(y_i-\mu)^2}{N} = \frac{1}{N}\times\sum_{i=1}^{N}(y_i-\mu_y)^2 = \frac{1}{N}\times\sum_{i=1}^{N}(y_i^2-2\times y_i\times\mu_y+\mu_y^2) = \frac{1}{N}\times[\sum_{i=1}^{N}y_i^2-2\times\sum_{i=1}^{N}y_i\times\mu_y+\sum_{i=1}^{N}\mu_y^2] = \frac{1}{N}\times[\sum_{i=1}^{N}y_i^2-2\times N\times\mu_y\times\mu_y+N\times\mu_y^2] = \frac{1}{N}\times[\sum_{i=1}^{N}y_i^2-N\times\mu_y^2] = \frac{\sum_{i=1}^{N}y_i^2}{N}-\mu_y^2 = \frac{1}{N}\times\left[\sum_{i=1}^{N}y_i^2-\frac{\left(\sum_{i=1}^{N}y_i\right)^2}{N}\right] = 10$$

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x\times\sigma_y} = \frac{\sigma_{xy}}{\sigma_x\times\sigma_y} = \frac{\frac{\sum_{i=1}^{N}(x_i-\mu_x)\times(y_i-\mu_y)}{N}}{\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu_x)^2}{N}}\times\sqrt{\frac{\sum_{i=1}^{N}(y_i-\mu_y)^2}{N}}} = \frac{\sum_{i=1}^{N}(x_i-\mu_x)\times(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i-\mu_x)^2}\times\sqrt{\sum_{i=1}^{N}(y_i-\mu_y)^2}} = \frac{\sum_{i=1}^{N}x_i\times y_i-\frac{\sum_i^N x_i\times\sum_i^N y_i}{N}}{\sqrt{\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}}\times\sqrt{\sum_{i=1}^{N}y_i^2-\frac{\left(\sum_{i=1}^{N}y_i\right)^2}{N}}}$$

$$= \frac{\frac{1}{N}\times\left(\sum_{i=1}^{N}x_i\times y_i-\frac{\sum_i^N x_i\times\sum_i^N y_i}{N}\right)}{\frac{1}{N}\times\left(\sqrt{\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}}\times\sqrt{\sum_{i=1}^{N}y_i^2-\frac{\left(\sum_{i=1}^{N}y_i\right)^2}{N}}\right)} = \frac{\frac{\sum_{i=1}^{N}x_i\times y_i}{N}-\frac{\sum_i^N x_i\times\sum_i^N y_i}{N^2}}{\sqrt{\frac{1}{N}\times\left[\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}\right]}\times\sqrt{\frac{1}{N}\times\left[\sum_{i=1}^{N}y_i^2-\frac{\left(\sum_{i=1}^{N}y_i\right)^2}{N}\right]}} =$$

$$\frac{\frac{\sum_{i=1}^{N}x_i\times y_i}{N}-\left(\frac{\sum_i^N x_i}{N}\right)\times\left(\frac{\sum_i^N y_i}{N}\right)}{\sqrt{\frac{1}{N}\times\left[\sum_{i=1}^{N}x_i^2-\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}\right]}\times\sqrt{\frac{1}{N}\times\left[\sum_{i=1}^{N}y_i^2-\frac{\left(\sum_{i=1}^{N}y_i\right)^2}{N}\right]}} = \frac{E(x\times y)-E(x)\times E(y)}{\sqrt{V(x)}\times\sqrt{V(y)}} = \frac{21-5\times 6}{\sqrt{9}\times\sqrt{10}} = \frac{21-30}{9.4868} = -0.9487$$

答案：相關係數 = -0.9487

Excel 操作指令

利用 Excel 軟體中工具(T)→資料分析(D)…→在資料分析對話方塊中選取分析工具(A)**相關係數**→確定。在相關係數對話方塊中，輸入範圍(I)拖拉選擇欲統計比較的數值區間，其分組方式選擇逐欄(C)，輸出選項勾選新工作表(P):→確定。即在新增加的工作表中出現選取數值區間內各欄之間的**相關係數(**coefficient of correlation)。

	欄 1	欄 2
欄 1	1	
欄 2	0.944844	1

3.6 偏度與峰度

3.6.1 偏度

評量觀測資料數值分布的對稱性時，資料分布可能是右偏或左偏。偏度或偏態(Skewness)可以使用偏度係數或偏態係數(coefficient of skewness, SK)來衡量。偏度係數屬於無因次單位。介紹四種不同的偏度運算公式，其運算獲得的偏度係數數值皆不同，皆是表示偏斜程度的一種指標。

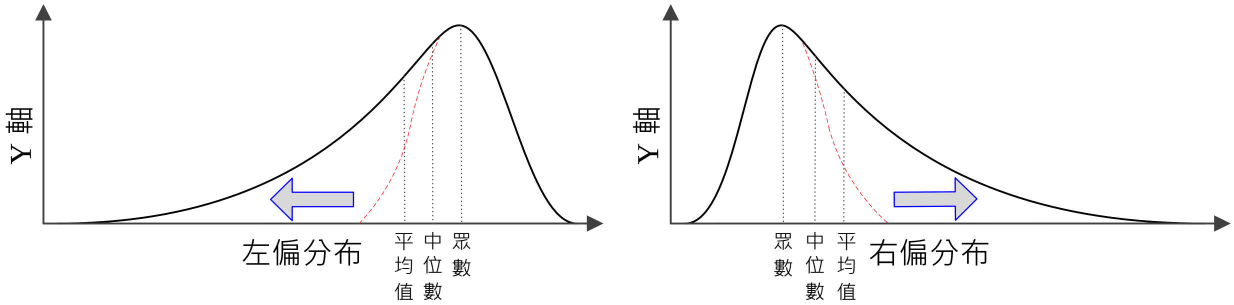


圖 3-9 左偏和右偏分布圖

Bowley 偏態係數 SK_B 或四分位偏態係數(Quartile skewness coefficient)

$$SK_B = \frac{(Q_3-M_e)-(M_e-Q_1)}{(Q_3-M_e)+(M_e-Q_1)} = \frac{(Q_3-M_e)-(M_e-Q_1)}{Q_3-Q_1} = \frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1} = \frac{Q_3-2\times Q_2+Q_1}{Q_3-Q_1} = \frac{Q_1+Q_3-2\times Q_2}{Q_3-Q_1}$$

若 $Q_3 = M_e = Q_2$，即 $SK_B = \frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1} = \frac{(Q_2-Q_2)-(Q_2-Q_1)}{Q_2-Q_1} = \frac{(\cancel{Q_2}-\cancel{Q_2})-(Q_2-Q_1)}{Q_2-Q_1} = \frac{-(Q_2-Q_1)}{Q_2-Q_1} = -1$，此為最小值；$Q_1 = M_e = Q_2$，即 $SK_B = \frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1} = \frac{(Q_3-Q_1)-(Q_1-Q_1)}{Q_3-Q_1} = \frac{(Q_3-Q_1)-(\cancel{Q_1}-\cancel{Q_1})}{Q_3-Q_1} = \frac{(Q_3-Q_1)}{Q_3-Q_1} = 1$，此為最大值。因此偏態係數 SK_B 會介於–1 ~ 1，$-1 < SK_B < 1$。

$|SK_B| \fallingdotseq 0.0$，偏斜程度屬於「沒有」。一般 $|SK_B| \fallingdotseq 0.1$，偏斜程度屬於「輕微」。一般 $|SK_B| > 0.3$，偏斜程度屬於「嚴重」。

若 $Q_3 - M_e > M_e - Q_1$，致 $SK_B > 0$，屬於右偏分布或正偏分布(positively skewed)。

若 $Q_3 - M_e = M_e - Q_1$，致 $SK_B = 0$，屬於對稱分布。

若 $Q_3 - M_e < M_e - Q_1$，致 $SK_B < 0$，屬於左偏分布或負偏分布(negatively skewed)。

皮爾森(Karl Pearson)偏態係數 SK_p (Pearson's skewness coefficients)、皮爾森中位數偏態係數(Pearson median coefficient of skewness)或第二偏態係數(second skewness coefficient)

母體資料：$SK_p = \frac{3\times(\mu-M_e)}{\sigma}$，其中 M_e：母體中位數 = 第 2 四分位數 Q_2

樣本資料：$SK_p = \frac{3\times(\bar{x}-m_e)}{s}$，其中 m_e：樣本中位數 = 第 2 四分位數 Q_2

若算術平均值(mean) $\mu >$ 中位數(median) M_e，致 $SK_p > 0$，屬於右偏分布(skewed right distribution)或正偏分布(positively skewed distribution)。

若算術平均值(mean) $\mu =$ 中位數(median) M_e，致 $SK_p = 0$，屬於對稱分布(normal distribution)。

若算術平均值(mean) $\mu <$ 中位數(median) M_e，致 $SK_p < 0$，屬於左偏分布(skewed left distribution)或負偏分布(negatively skewed distribution)。

動差法(method of moments)計算偏態係數 α_3

母體資料：$\alpha_3 = \frac{\frac{\sum_{i=1}^{N}(x_i-\mu)^3}{N}}{\left[\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}\right]^3} = \frac{\frac{\sum_{i=1}^{N}(x_i-\mu)^3}{N}}{\sigma^3}$　　樣本資料：$\alpha_3 = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{n}}{\left[\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\right]^3} = \frac{m_3}{s^3}$

$\alpha_3 = 0$，資料對稱分布

$\alpha_3 > 0$，資料右偏分布

$\alpha_3 < 0$，資料左偏分布

Excel 軟體中偏態係數的計算方式：即使用 Excel 函數 SKEW 運算的偏態係數

$$SK_p = \frac{n}{(n-1)\times(n-2)} \times \sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^3$$

$SK_p = 0$ 為對稱分布(symmetric distribution)

算術平均值(mean) μ = 中位數(median) M_e = 眾數(mode) M_o

$SK_p > 0$ 為右偏分布(skewed to the right)或正偏分布(positively skewed)

算術平均值(mean) μ > 中位數(median) M_e > 眾數(mode) M_o

$SK_p < 0$ 為左偏分布(skewed to the left)或負偏分布(negatively skewed)

算術平均值(mean) μ < 中位數(median) M_e < 眾數(mode) M_o

$|SK_p|$數值愈大，代表資料分布愈具有偏斜的型態。

範例 3.66 抽樣調查高雄市選取其中 12 家國際觀光旅館當樣本，專職員工數量(人)分別如下所示，請依序計算高雄市國際觀光旅館專職員工數之 Bowley 偏態係數 SK_B 和皮爾森(Karl Pearson)偏態係數 SK_p。(答案有效位數四捨五入取到小數點後第 4 位)

254 317 324 331 325 425 185 265 325 215 198 260

題解：第 1 四分位數 $Q_1 = 244.25$ 人，$M_e = Q_2 = 291$ 人，$Q_3 = 325$ 人，樣本平均值$\bar{x} = 285.3$ 人，標準(偏)差 SD = 68.7 人。

Bowley 偏態係數 $SK_B = \frac{(Q_3-M_e)-(M_e-Q_1)}{Q_3-Q_1} = \frac{(325-291)-(291-244.25)}{325-244.25} = \frac{-12.75}{80.75} = -0.1579$

皮爾森(Karl Pearson)偏態係數 $SK_p = \frac{3\times(\bar{x}-m_e)}{s} = \frac{3\times(285.3-291)}{68.7} = -0.2475$

答案：Bowley 偏態係數 $SK_B = -0.1579$；皮爾森(Karl Pearson)偏態係數 $SK_p = -0.2475$

練習 3.43 抽樣調查高雄市選取其中 12 家國際觀光旅館當樣本，專職員工數量分別如下所示，請依序計算高雄市國際觀光旅館專職員工數之 Bowley 偏態係數 SK_B 和皮爾森(Karl Pearson)偏態係數 SK_p。(答案有效位數四捨五入取到小數點後第 4 位)

154 266 341 331 325 425 185 265 122 215 198 123

題解：$Q_1 = 177.25$，$M_e = Q_2 = 240$，$Q_3 = 326.5$，樣本平均值$\bar{x} = 245.8$，SD = 95.9

Bowley 偏態係數 $SK_B = \frac{(Q_3-M_e)-(M_e-Q_1)}{Q_3-Q_1} = \frac{(326.5-240)-(240-177.25)}{326.5-177.25} = \frac{23.75}{149.25} = 0.1591$

皮爾森(Karl Pearson)偏態係數 $SK_p = \frac{3\times(\bar{x}-m_e)}{s} = \frac{3\times(245.8-240)}{95.9} = 0.1814$

答案：Bowley 偏態係數 $SK_B = 0.1591$；皮爾森(Karl Pearson)偏態係數 $SK_p = 0.1814$

練習 3.44 奇遇餐廳今天消費者消費金額的次數分布如下表所示：

消費金額	次數
51~60	5
61~70	24
71~80	26
81~90	31
91~100	25
101~110	5

試計算消費金額的平均值、中位數、眾數、變異數、標準(偏)差和偏態係數。

在 Excel 軟體內，利用插入→函數→SKEW，計算偏度值(係數)。

SPSS 操作指令一

Analyze(分析)→選擇 Reports(報表)→選擇 Case Summaries…(觀察值摘要)→在 Summarize Cases 視窗中，將欲統計 Skewness(偏態)數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Summarize Cases 視窗下方之 Statistics…(統計量)按鈕→在 Summary Report: Statistics 次視窗左小視窗

Statistics(統計量)點選 Skewness(偏態)至右小視窗 Cell Statistics(格統計)內，按下方 Continue(繼續)按鈕→回到 Summarize Cases 視窗，按 OK(確定)按鈕即會跑出統計結果

Summarize

Case Processing Summary(a)

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
VAR00001	12	100.0%	0	.0%	12	100.0%

a Limited to first 100 cases.

Case Summaries(a)

		VAR00001
1		5.00
2		65.00
3		6.00
4		5.00
5		88.00
6		5.00
7		8.00
8		77.00
9		5.00
10		88.00
11		5.00
12		8.00
Total	Skewness	.893

a Limited to first 100 cases.

SPSS 操作指令二

Analyze(分析)→選擇 Descriptive Statistics(描述性統計)→選擇 Frequencies…(次數分配表)→在 Frequencies 視窗中，將欲統計 Skewness(偏態)數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Frequencies 視窗下方之 Statistics…(統計量)按鈕→在 Frequencies: Statistics 次視窗右下角 Distribution(分布/分配)內勾選 Skewness(偏態)，按右邊 Continue(繼續)按鈕→回到 Frequencies 視窗，按 OK(確定)按鈕即會跑出統計結果

Frequencies

Statistics

VAR00001

N	Valid	12
	Missing	0
Skewness		.893
Std. Error of Skewness		.637

VAR00001

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5.00	5	41.7	41.7	41.7
	6.00	1	8.3	8.3	50.0
	8.00	2	16.7	16.7	66.7
	65.00	1	8.3	8.3	75.0
	77.00	1	8.3	8.3	83.3
	88.00	2	16.7	16.7	100.0
	Total	12	100.0	100.0	

3.6.2 峰度

當次數分布(分配)有集中的現象時，即會出現波峰(peak)，波峰可為單峰、雙峰與多峰型態。在單峰的情況下，依據平坦陡峭的分布程度可分為高峽峰(lepto kurtosis)、低闊峰(platy kurtosis)或常態峰(normal kurtosis)，峰度(Kurtosis)的高低以峰度係數進行評量。介紹四級動差法和 Excel 軟體兩種峰度係數的運算方式，兩種運算出來的峰度係數數值不同，不過皆可以代表分布平坦陡峭程度的一種數值指標。

峰態係數或峰度係數(coefficient of kurtosis)之度量一般以 β_2、K、β_k 或 γ_2 (γ 讀音 gamma)符號表示，利用 SPSS 和 Excel 軟體所計算峰度是以 γ_2 係數呈現。峰態係數屬於無因次單位。

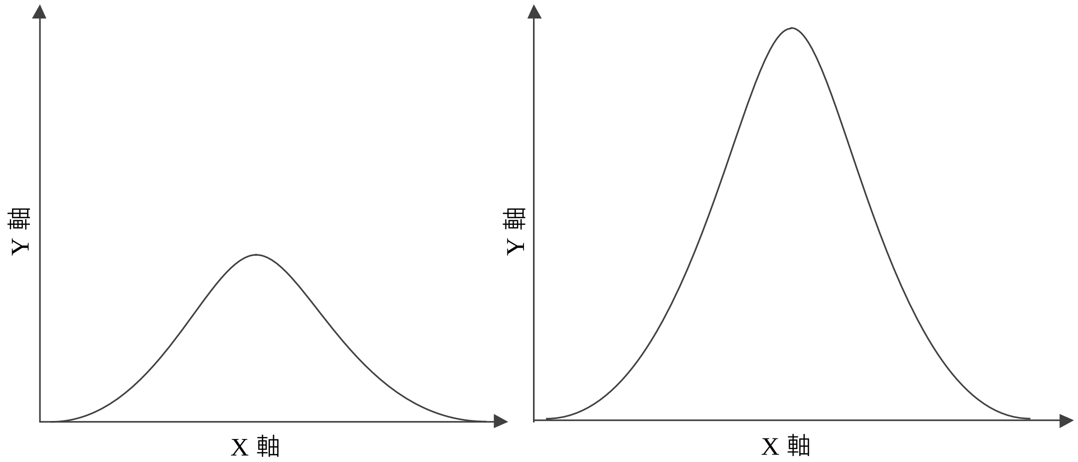


圖 3-10 峰度分布對照圖

四級動差法計算峰度係數

若原始母體資料中有 N 個觀測值，可以依據觀測值數值相等性區分為 k 組，在第 i 組內有 f_i 個觀測值，組內觀測值數值皆相等，觀測值數量 $N=\sum_{i=1}^{k} f_i$。其母體平均值 $\mu=\frac{\sum_{i=1}^{k} x_i \times f_i}{N}=\frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i}$ 和母體標準(偏)差 $\sigma=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N}}=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_i^k f_i}}$。

若原始樣本資料中有 n 個觀測值，可以依據觀測值數值相等性區分為 k 組，在第 i 組內有 f_i 個觀測值，組內觀測值數值皆相等，觀測值數量 $n=\sum_{i=1}^{k} f_i$。其平均值 $\bar{x}=\frac{\sum_{i=1}^{k} x_i \times f_i}{n}=\frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i}$ 和標準(偏)差 $S=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\bar{x})^2}{\sum_i^k f_i-1}}$。

	母體	樣本
觀測值數量	$N=\sum_{i=1}^{k} f_i$	$n=\sum_{i=1}^{k} f_i$
平均值	$\mu=\frac{\sum_{i=1}^{k} x_i \times f_i}{N}=\frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i}$	$\bar{x}=\frac{\sum_{i=1}^{k} x_i \times f_i}{n}=\frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i}$
標準(偏)差	$\sigma=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N}}=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_i^k f_i}}$	$S=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\bar{x})^2}{\sum_i^k f_i-1}}$
峰度β_2	$\beta_2=\frac{\frac{1}{N}\times\sum_{i=1}^{k}[f_i \times (x_i-\mu)^4]}{\sigma^4}$	$\beta_2=\frac{\frac{1}{n-1}\times\sum_{i=1}^{k}[f_i \times (x_i-\bar{x})^4]}{S^4}$
峰度γ_2	$\gamma_2=\beta_2-3$	$\gamma_2=\beta_2-3$

Excel 軟體中峰度係數的計算方式

$$\gamma_2=\left[\frac{n\times(n+1)}{(n-1)\times(n-2)\times(n-3)}\times\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^4\right]-\frac{3\times(n-1)^2}{(n-2)\times(n-3)}$$

常態峰(mesokurtosis)表示資料的分布呈現一般的型態，峰度係數 $\beta_2=3$；$\gamma_2=0$。

高峽峰(leptokurtosis)表示資料分布過度集中於算術平均值或眾數附近，峰度係數 $\beta_2 > 3$；$\gamma_2 > 0$。

低闊峰(platykurtosis)表示資料分布比較平均分散，峰度係數 $\beta_2 < 3$；$\gamma_2 < 0$。

在 Excel 軟體內，利用插入→函數→KURT，計算峰度值(γ_2 係數)。使用 Excel 軟體欲直接使用 KURT 函數計算時，必須輸入每一個基本單位的觀測值，才能直接運算。

範例 3.67　計算下列母體數值資料的四級動差法計算峰度係數(答案有效位數四捨五入取到小數點後第 3 位)

組序	x_i	f_i	$x_i \times f_i$	$x_i - \mu$	$(x_i-\mu)^2$	$f_i \times (x_i-\mu)^2$	$(x_i-\mu)^4$	$f_i \times (x_i-\mu)^4$
1	5	5	25	-25.42	646.01	3230.03	417325.0	2086624.9
2	6	1	6	-24.42	596.17	596.17	355423.0	355423.0
3	8	2	16	-22.42	502.51	1005.01	252513.2	505026.5
4	65	1	65	34.58	1196.01	1196.01	1430432.6	1430432.6
5	77	1	77	46.58	2170.01	2170.01	4708930.1	4708930.1
6	88	2	176	57.58	3315.84	6631.68	10994796.7	21989593.5
	合計	12	365			14828.92		31076030.5

題解：前述資料一共有 $k = 6$ 組不同觀測值數值資料。

母體觀測值數量 $N = \sum_{i=1}^{k} f_i = 5 + 1 + 2 + 1 + 1 + 2 = 12$

母體平均值 $\mu = \frac{\sum_{i=1}^{k} x_i \times f_i}{N} = \frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i} = \frac{365}{12} = 30.42$

母體標準(偏)差 $\sigma = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_i^k f_i}} = \sqrt{\frac{14828.92}{12}} = \sqrt{1235.74} = 35.15$

$\beta_2 = \frac{\frac{1}{N} \times \sum_{i=1}^{k}[f_i \times (x_i-\mu)^4]}{\sigma^4} = \frac{\frac{1}{12} \times 31076030.5}{35.15^4} = \frac{2589669.2}{1527061} = 1.696$

$\gamma_2 = \beta_2 - 3 = 1.696 - 3 = -1.304$

答案：$\beta_2 = 1.696$；$\gamma_2 = -1.304$

練習 3.45　計算下列樣本數值資料的峰度係數(答案有效位數四捨五入取到小數點後第 3 位)

組序	x_i	f_i	$x_i \times f_i$	$x_i - \bar{x}$	$(x_i-\bar{x})^4$	$f_i \times (x_i-\bar{x})^4$	$(x_i-\bar{x})^2$	$f_i \times (x_i-\bar{x})^2$
1	41	1	41	-22.462	254541.1	254541.1	504.521	504.5
2	45	1	45	-18.462	116164.0	116164.0	340.828	340.8
3	57	1	57	-6.462	1743.2	1743.2	41.751	41.8
4	63	2	126	-0.462	0.0	0.1	0.213	0.4
5	64	2	128	0.538	0.1	0.2	0.290	0.6
6	65	2	130	1.538	5.6	11.2	2.367	4.7
7	69	1	69	5.538	940.9	940.9	30.675	30.7
8	72	1	72	8.538	5315.2	5315.2	72.905	72.9
9	77	1	77	13.538	33595.2	33595.2	183.290	183.3
10	80	1	80	16.538	74813.6	74813.6	273.521	273.5
	合計	13	825			487124.7		1453.2

題解：前述資料一共有 $k = 10$ 組不同觀測值數值資料。

樣本觀測值數量 $n = \sum_{i=1}^{k} f_i = 1 + 1 + 1 + 2 + 2 + 2 + 1 + 1 + 1 + 1 = 13$

樣本平均值 $\bar{x} = \frac{\sum_{i=1}^{k} x_i \times f_i}{n} = \frac{825}{13} = 63.462$

樣本標準(偏)差 $S = \sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{1453.2}{13-1}} = \sqrt{121.1} = 11.005$

$\beta_2 = \frac{\frac{1}{n-1} \times \sum_{i=1}^{k}[f_i \times (x_i-\bar{x})^4]}{S^4} = \frac{\frac{1}{13-1} \times 487124.7}{11.005^4} = \frac{40593.7}{14665.83} = 2.7679$

$\gamma_2 = \beta_2 - 3 = 2.7679 - 3 = -0.2321$

答案：樣本平均值 $\bar{x} = 63.5$；樣本標準(偏)差 $S = 11.0$；$\beta_2 = 2.768$；$\gamma_2 = -0.232$

SPSS 操作指令一

Analyze(分析)→選擇 Reports(報表)→選擇 Case Summaries…(觀察值摘要)→在 Summarize Cases 視窗中，將欲統計 Kurtosis(峰度)數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Summarize Cases 視窗下方之 Statistics…(統計量)按鈕→在 Summary Report: Statistics 次視窗左小視窗 Statistics(統計量)點選 Kurtosis(峰度)至右小視窗 Cell Statistics(格統計)內，按下方 Continue(繼續)按鈕→回到 Summarize Cases 視窗，按 OK(確定)按鈕即會跑出統計結果峰度係數 γ_2。

Summarize

Case Processing Summary[a]

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
VAR00001	12	100.0%	0	.0%	12	100.0%

[a] Limited to first 100 cases.

Case Summaries[a]

		VAR00001
1		5.00
2		65.00
3		6.00
4		5.00
5		88.00
6		5.00
7		8.00
8		77.00
9		5.00
10		88.00
11		5.00
12		8.00
Total	Kurtosis	-1.339

[a] Limited to first 100 cases.

SPSS 操作指令二

Analyze(分析)→選擇 Descriptive Statistics(描述性統計)→選擇 Frequencies…(次數分配表)→在 Frequencies 視窗中，將欲統計 Kurtosis(峰度)數值的變數，由左小視窗勾選至右邊 Variable(s)(變數)小視窗→點選 Frequencies 視窗下方之 Statistics…(統計量)按鈕→在 Frequencies: Statistics 次視窗右下角 Distribution(分布/分配)內勾選 Kurtosis(峰度)，按右邊 Continue(繼續)按鈕→回到 Frequencies 視窗，按 OK(確定)按鈕即會跑出統計結果峰度係數 γ_2。

Frequencies

Statistics

VAR00001

N	Valid	12
	Missing	0
Kurtosis		-1.339
Std. Error of Kurtosis		1.232

VAR00001

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	5.00	5	41.7	41.7	41.7
	6.00	1	8.3	8.3	50.0
	8.00	2	16.7	16.7	66.7
	65.00	1	8.3	8.3	75.0

	Frequency	Percent	Valid Percent	Cumulative Percent
77.00	1	8.3	8.3	83.3
88.00	2	16.7	16.7	100.0
Total	12	100.0	100.0	

3.7 敘述性統計 SPSS 操作方法【選擇教材】

次數分布敘述統計

1. Analyze/Statistics(統計分析) → Descriptive Statistics(敘述統計) → Frequencies (次數分配表)，打開 Frequencies(次數分配表)對話視窗
2. 將左邊的變項清單之變數，點選進入右邊的 Variable(s):對話方塊中
3. 點選 Statistics…按鈕，出現 Frequencies: Statistics 對話方塊
4. 在左下角 Dispersion (分散趨勢)的方塊中，勾選 Std. deviation 選項
5. 在右上角 Central Tendency (集中趨勢)的方塊中，勾選 Mean 選項
6. 點選 Continue 按鈕，回到 Frequencies 的對話方塊
7. 點選 OK 按鈕，執行敘述統計分析
8. 獲得下列成果

Frequencies

Statistics

		項目問題 1	項目問題 2	項目問題 3	項目問題 4
N	Valid	32	32	32	32
	Missing	268	268	268	268
Mean		2.47	1.94	2.16	2.63
Std. Deviation		0.915	0.948	0.884	1.008

Frequency Table

項目問題 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	5	1.7	15.6	15.6
	2	10	3.3	31.3	46.9
	3	15	5.0	46.9	93.8
	4	1	0.3	3.1	96.9
	5	1	0.3	3.1	100.0
	Total	32	10.7	100.1	
Missing	System	268	89.3		
Total		300	100.0		

項目問題 2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	14	4.7	43.8	43.8
	2	7	2.3	21.9	65.6
	3	10	3.3	31.3	96.9
	4	1	0.3	3.1	100.0
	Total	32	10.7	100.0	
Missing	System	268	89.3		
Total		300	100.0		

項目問題 3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	8	2.7	25.0	25.0
	2	13	4.3	40.6	65.6
	3	9	3.0	28.1	93.8
	4	2	0.7	6.3	100.0
	Total	32	10.7	100.0	
Missing	System	268	89.3		
Total		300	100.0		

項目問題 4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	4	1.3	12.5	12.5
	2	11	3.7	34.4	46.9
	3	11	3.7	34.4	81.3
	4	5	1.7	15.6	96.9
	5	1	0.3	3.1	100.0
	Total	32	10.7	100.0	
Missing	System	268	89.3		
Total		300	100.0		

描述性敘述統計

1. Analyze/Statistics(統計分析) → Descriptive Statistics(敘述統計) → Descriptives… (描述性統計量，打開 Descriptives (描述性統計量)對話視窗
2. 點選左邊方塊中，欲進行敘述統計的變數，進入右邊 Variable(s):方塊中
3. 點選 Options…按鈕，進入 Descriptives: Options 對話方塊
4. 勾選 Mean 和 Std. deviation 兩個選項
5. 按 Continue 按鈕，回到 Descriptives 對話方塊
6. 按 OK 按鈕，執行敘述性統計
7. 獲得結果如下

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
項目問題 1	32	1	5	2.47	0.915
項目問題 2	32	1	4	1.94	0.948
項目問題 3	32	1	4	2.16	0.884
項目問題 4	32	1	5	2.63	1.008
Valid N (listwise)	32				

問題型態

遊客的遊憩體驗、旅遊滿意度、遊憩動機的現況？

分析方法

以算術平均值與標準(偏)差表示最為適宜，顯示有關變數的分布(配)情況。

敘述性統計 SPSS 操作方法

次數分布敘述性統計

1. Analyze/Statistics(統計分析) → Descriptive Statistics(敘述統計) → Frequencies (次數分配表)，打開 Frequencies(次數分配表)對話視窗
2. 將左邊的變項清單之變數，點選進入右邊的 Variable(s): 對話方塊中
3. 點選 Statistics…按鈕，出現 Frequencies: Statistics 對話方塊
4. 在左下角 Dispersion (分散趨勢)的方塊中，勾選 Std. deviation 選項
5. 在右上角 Central Tendency (集中趨勢)的方塊中，勾選 Mean 選項
6. 點選 Continue 按鈕，回到 Frequencies 的對話方塊
7. 點選 OK 按鈕，執行敘述統計分析
8. 獲得下列成果

Frequencies

Statistics

		行為	動機	態度	認知
N	Valid	40	40	40	40
	Missing	7	7	7	7
Mean		3.025	2.8375	2.925	3.2333
Std. Deviation		1.07387	0.73935	1.09515	0.72087

Frequency Table

行為

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	4	8.5	10	10
	2	6	12.8	15	25
	3	19	40.4	47.5	72.5
	4	7	14.9	17.5	90
	5	4	8.5	10	100
	Total	40	85.1	100	
Missing	System	7	14.9		
Total		47	100		

動機

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	4.3	5	5
	1.5	2	4.3	5	10
	2	2	4.3	5	15
	2.25	3	6.4	7.5	22.5
	2.5	3	6.4	7.5	30
	2.75	4	8.5	10	40
	3	11	23.4	27.5	67.5
	3.25	6	12.8	15	82.5
	3.75	5	10.6	12.5	95
	4	2	4.3	5	100
	Total	40	85.1	100	
Missing	System	7	14.9		
Total		47	100		

態度

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	3	6.4	7.5	7.5
	2	14	29.8	35	42.5
	3	8	17	20	62.5
	4	13	27.7	32.5	95
	5	2	4.3	5	100
	Total	40	85.1	100	
Missing	System	7	14.9		
Total		47	100		

認知

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.67	1	2.1	2.5	2.5
	2	2	4.3	5	7.5
	2.33	3	6.4	7.5	15
	2.67	7	14.9	17.5	32.5
	3	5	10.6	12.5	45
	3.33	7	14.9	17.5	62.5
	3.67	5	10.6	12.5	75
	4	8	17	20	95
	4.67	2	4.3	5	100
	Total	40	85.1	100	
Missing	System	7	14.9		
Total		47	100		

描述性敘述統計

1. Analyze/Statistics(統計分析) → Descriptive Statistics(敘述統計) → Descriptives… (描述性統計量 ，打開 Descriptives (描述性統計量)對話視窗
2. 點選左邊方塊中，欲進行敘述統計的變數，進入右邊 Variable(s):方塊中
3. 點選 Options…按鈕，進入 Descriptives: Options 對話方塊
4. 勾選 Mean 和 Std. deviation 兩個選項
5. 按 Continue 按鈕，回到 Descriptives 對話方塊
6. 按 OK 按鈕，執行敘述性統計
7. 獲得結果如下

Descriptives

Descriptive Statistics

	N	Mean	Std. Deviation
行為	40	3.025	1.07387
動機	40	2.8375	0.73935
態度	40	2.925	1.09515
認知	40	3.2333	0.72087
Valid N (listwise)	40		

練習 3.46 欲瞭解食品近一年價格波動情況，而至市場實地訪查，得下列兩類食品資料：(2007 普考統計學概要)

食品	平均值	中位數	標準差	最小值	最大值	樣本數
雞蛋(x)	29.16	30.00	2.72	24.00	33.00	12
牛肉(y)	109.33	110.00	6.63	98.00	120.00	12

$\sum_{i=1}^{12} x_i \times y_i = 38314$

1.求兩者價格的全距和變異係數(coefficient of variation)

2.何者價格波動大？兩者價格波動的相關程度(相關係數)為何？

練習 3.47 某單位主管欲採購每棵至少高 3 公尺的行道樹 150 棵。甲林場負責人報告目前樹木生長概況：林場有 600 棵樹，平均高度是 2.5 公尺，標準差是 0.5 公尺，樹木高度的分布接近對稱。乙林場負責人報告其林場有 500 棵樹，平均高度是 2.6 公尺，標準差是 0.2 公尺，樹木高度的分布並不對稱。(2007 普考統計學概要)

1.甲林場大約有幾棵樹符合條件？(提示：柴比雪夫不等式或經驗法則)

2.乙林場大約有幾棵樹符合條件？(提示：柴比雪夫不等式或經驗法則)

3.該主管應如何採購？

練習 3.48 一組「樣本」(sample)的觀察值為 3、7、5、6、7 和 2，則「樣本眾數」(sample mode)為：(A)5；(B)5.5；(C)7；(D)2。(2007 初等考試統計學大意)

練習 3.49 風速可利用地面上的風速儀或天上的氣象衛星測定。另 X 與 Y 分別代表由風速儀及氣象衛星測得某地之風速，則通常 X 與 Y 為：(A)正相關；(B)負相關；(C)相關係數為 0；(D)不一定。(2007 初等考試統計學大意)

練習 3.50 設有一組資料 2、3、7、8、9、9 和 11，其四分位差(inter-quartile range)為：(A)4；(B)5；(C)6；(D)7。(99 年公務人員初等考試統計學大意) A

練習 3.51 X, Y, Z 互為獨立的隨機變數，而且 Var(X) = 4, Var(Y) = 9, Var(Z) = 16，則 Var(X – Y + Z) = ？(A)81；(B)29；(C)11；(D)9。(2007 初等考試統計學大意)

練習 3.52 觀測值的單位與下列哪些數值相同？arithmetic mean, quartile, deciles, trimmed mean, interquartile range, variance, covariance, coefficient of variation, standardized value, coefficient of correlation, coefficient of skewness

練習 3.53 是非題(○)在右偏的資料分布中，平均值大於中位數與衆數。

練習 3.54 是非題(○)在一組樣本資料中，將每一個觀測值皆乘以一常數 C，則其平均值與標準(偏)差皆變成原始數值的 C 倍。

練習 3.55　下列何者不是用來評量測量資料的分散程度？(A)全距(range)；(B)變異數(variance)；(C)算術平均(arithmetic mean)；(D)標準差(standard deviation)

範例 3.68　阿登餐廳進貨一批鱸魚，平均每隻魚進貨價 200 元，標準(偏)差 20 元。依據標準菜單每一盤清蒸鱸魚需要一隻鱸魚製作，現欲使用鱸魚進貨價的 160 %為每盤清蒸鱸魚的售價。請計算(A)每盤清蒸鱸魚的平均售價(元)；(B)每盤清蒸鱸魚售價的變異數(variance)(元 ^{2})(答案有效位數四捨五入取到個位數)

題解：

每盤清蒸鱸魚售價 = 160 % × 每隻鱸魚進貨價

每盤清蒸鱸魚售價的平均值= 160 % × 200 元 = 320 元

每盤清蒸鱸魚售價的變異數(variance) = $(160\ \%)^2 \times 20^2 = 1024$ 元 ^{2}

答案：(A)平均售價 = 320 元；(B)售價的變異數 = 1024 元 ^{2}

練習 3.56　請每位組員蒐集自己一個星期三餐支出費用，利用 Microsoft excel 軟體分別計算早餐、中餐和晚餐支出費用的算術平均值、中位數、變異係數、標準(偏)差、偏度和峰度，並計算早餐、中餐和晚餐支出費用的相關係數。先將上述組員個別資料並列於同一個 Microsoft word 檔案中，再將整組組員的上述資料，合併製作成表格，放在同一份 Microsoft word 檔案中，整組組員一起比較分析，並將比較分析結果撰寫在同一份檔案中。

T score(T 分數)

假設母體屬於常態分布時，T score 屬於一種轉換分數，其平均值 = 50，標準差 = 10。 $T = 50 + 10Z = 50 + 10 \times \frac{x_i - \bar{x}}{s}$。由 Z 分數運算出 T 分數。

PR 值(percentile rank)：PR90 代表排前 10%

第 r 級原動差(rth zero moment)　$E(X^r) = \frac{\sum_{i=1}^{N}(x_i)^r}{N}$，其中 $r = 1, 2, 3, 4, \ldots$

一級原動差(1th zero moment)　$\mu_1' = \frac{\sum_{i=1}^{N} x_i}{N}$

二級原動差(2th zero moment)　$\mu_2' = \frac{\sum_{i=1}^{N}(x_i)^2}{N}$

三級原動差(3th zero moment)　$\mu_3' = \frac{\sum_{i=1}^{N}(x_i)^3}{N}$

四級原動差(4th zero moment)　$\mu_4' = \frac{\sum_{i=1}^{N}(x_i)^4}{N}$

一級主動差(1th principle moment)　$\mu_1 = \frac{\sum_{i=1}^{N}(x_i-\mu)}{N}$

二級主動差(2th principle moment)　$\mu_2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$

三級主動差(3th principle moment)　$\mu_3 = \frac{\sum_{i=1}^{N}(x_i-\mu)^3}{N}$

四級主動差(4th principle moment)　$\mu_4 = \frac{\sum_{i=1}^{N}(x_i-\mu)^4}{N}$

討論議題

1.學習者同步討論議題：最需要評量的分散程度

第一回合請於 D+3 早上 1130 以前，從「議題討論」區【張貼】標題：「分散程度」，本文：請論述在何種狀況下，最需要知道欲評量變數的分散程度(變異數和標準偏差)？請具體說明理由(50 個以上中文字詮釋)。

期望可以透過議題討論資訊的交流，相互激勵，提升學習效益。待有 50 篇第一回合【張貼】回應後或第一回合【張貼】時間結束後，檢視其他同學的回應。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「最佳詮釋」，本文：請說明哪一位同學詮釋得最好，其理由(20 個字以上)。透過同學之間的討論分享，可以提升學習效益。加油。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

重點整理

Excel 函數彙整

Excel 函數	統計功能	輸入資料	輸出資料
AVERAGE	平均值	統計欄位	算術平均值
MEDIAN	中位數	統計欄位	中位數
MODE	眾數	統計欄位	眾數
QUARTILE	四分位數$(n-1)$，有第 0 和 4 四分位數	統計欄位，四分位序	四分位數
QUARTILE.EXC	四分位數$(n+1)$，無第 0 和 4 四分位數	統計欄位，四分位序	四分位數
QUARTILE.INC	四分位數$(n-1)$，有第 0 和 4 四分位數	統計欄位，四分位序	四分位數
PERCENTILE	百分位數$(n-1)$，有第 0 和 100 百分位數	統計欄位，百分位序	百分位數
PERCENTILE.EXC	百分位數$(n+1)$，無第 0 和 100 百分位數	統計欄位，百分位序	百分位數
PERCENTILE.INC	百分位數$(n-1)$，有第 0 和 100 百分位數	統計欄位，百分位序	百分位數
GEOMEAN	幾何平均值	統計欄位	幾何平均值
HARMEAN	調和平均值	統計欄位	調和平均值
TRIMMEAN	裁減式平均值	統計欄位，裁減比例	裁減式平均值
LARGE	最大值	統計欄位，序位	序位最大值
MAX	最大值	統計欄位	最大觀測值
SMALL	最小值	統計欄位，序位	序位最小值
MIN	最小值	統計欄位	最小觀測值
STDEV(.)P	母體標準(偏)差	統計欄位	母體標準(偏)差
STDEV(.S)	樣本標準(偏)差	統計欄位	樣本標準(偏)差
VAR(.)P	母體變異數	統計欄位	母體變異數
VAR(.S)	樣本變異數	統計欄位	樣本變異數
SKEW	樣本偏度	統計欄位	樣本偏度
SKEW.P	母體偏度	統計欄位	母體偏度
KURT	峰度	統計欄位	峰度
COVARIANCE.P	母體共變異數	配對分析統計欄位	母體共變異數
COVARINACE.S	樣本共變異數	配對分析統計欄位	樣本共變異數

科學記號 E：在 Microsoft Excel 軟體中經常檢視到

1.23E04 = 1.23×10^4 = 12300

1.23E-04 = 1.23×10^{-4} = 0.000123

未分組資料之算術平均值

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

分組資料的算術平均值

$$\mu = \frac{\sum_{i=1}^{k} f_i \times m_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i \times m_i}{N}$$
$$\bar{x} = \frac{\sum_{i=1}^{k} f_i \times m_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i \times m_i}{n}$$

未分組數值的中位數

(1)若樣本數量 n 為奇數(odd)，中位數即為第$\frac{n+1}{2}$序位的觀測值。

(2)若樣本數量 n 為偶數(even)，中位數為第$\frac{n}{2}$和$\frac{n}{2}+1$兩個序位觀測值的算術平均值。

分組數值的中位數

$$M_e = L_i + \left(\frac{n}{2} - F\right) \times \frac{h_i}{f_i}$$

未分組資料的四分位數

a.在 n 個觀測值(樣本)中，確認 Q_i 序位：$O(Q_i) = (i \times \frac{n-1}{4}) + 1$，其中 $i = 1, 2$ 或 3。

b.若 $O(Q_i)$數值為整數，此序位所對應觀測值之數值即為 Q_i；若 $O(Q_i)$數值不為整數，在 $O(Q_i)$整數與 $O(Q_i)$整數+1 兩個位置所對應觀測值之數值之間，利用線性內插法計算 Q_i 數值。

分組資料的四分位數

$$Q_i = L_i + \left[\left(i \times \frac{n-1}{4}\right) + 1 - F\right] \times \frac{h_i}{f_i}$$

未分組數值的十分位數

若$(i \times \frac{n-1}{10}) + 1$ 數值為整數時，第 i 個十分位數 D_i 為第$(i \times \frac{n-1}{10}) + 1$ 項的數值(觀測值)。

若$(i \times \frac{n-1}{10}) + 1$ 數值為非整數時，第 i 個十分位數 D_i 為第$(i \times \frac{n-1}{10}) + 1$ 整的數項與第$(i \times \frac{n-1}{10}) + 1$ 整數+1 項的數值(觀測值)，利用線性內插法計算之間的內插值。

分組資料的十分位數

$$D_i = L_i + \left[\left(i \times \frac{n-1}{10}\right) + 1 - F\right] \times \frac{h_i}{f_i}$$

未分組數值的百分位數

P_i 為一個以由小到大順序排列的資料集合中之第$(i \times \frac{n-1}{100}) + 1$ 項的數值，i: 百分位之號碼；n: 樣本數量。若$(i \times \frac{n-1}{100}) + 1$ 為整數時，P_i 為第$(i \times \frac{n-1}{100}) + 1$ 項的數值(觀測值)。若$(i \times \frac{n-1}{100}) + 1$ 為非整數時，P_i 為第$(i \times \frac{n-1}{100}) + 1$ 整數項的數值(觀測值)與第$(i \times \frac{n-1}{100}) + 1$ 整數+1 項的數值(觀測值)利用線性內插法計算。

分組數值的百分位數

$$P_i = L_i + \left[\left(i \times \frac{n-1}{100}\right) + 1 - F\right] \times \frac{h_i}{f_i}$$

幾何平均值

$$\mu_g = \text{antilog}\left(\frac{1}{N} \times \sum_{i=1}^{N} \log x_i\right) = \sqrt[N]{x_1 \times x_2 \times \dots \times x_N} = \sqrt[N]{\prod_{i=1}^{N} x_i} = \left(\prod_{i=1}^{N} x_i\right)^{\frac{1}{N}}$$
$$\bar{x}_g = \text{antilog}\left(\frac{1}{n} \times \sum_{i=1}^{n} \log x_i\right) = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$

期間平均增加率

$$G = \sqrt[n]{\frac{\text{期末值}}{\text{期始值}}} - 1$$

調和平均值

$$\bar{x}_H = \frac{1}{\frac{1}{n}\times\left(\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}\right)} = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

平方根平均值

$$\bar{x}_q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

加權算術平均值

$$\mu_w = \frac{\sum_{i=1}^{N} w_i \times x_i}{\sum_{i=1}^{N} w_i}$$

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i \times x_i}{\sum_{i=1}^{n} w_i}$$

全距中點

$$\text{全距中點 Midrange} = \frac{\text{highest value} + \text{lowest value}}{2}$$

中樞紐

$$\text{中樞紐 Midhinge} = \frac{Q_1+Q_3}{2}$$

偏斜分布

左右對稱(symmetric, zero skewness)，算術平均值(mean) μ = 中位數(median) M_e = 眾數(mode) M_o。

左偏(skewed to the left)、負偏(negatively skewed)，算術平均值(mean) μ < 中位數(median) M_e < 眾數(mode) M_o。

右偏(skewed to the right)、正偏(positively skewed)，算術平均值(mean) μ > 中位數(median) M_e > 眾數(mode) M_o。

全距

$$R = \text{最大觀測值} - \text{最小觀測值}$$

分組資料的全距

R = 最後一組組中點(組中值) – 第一組組中點(組中值)

R = 最後一組組上限值 – 第一組組下限值

四分位距

$$\text{IQR} = \text{第 3(個)四分位數} - \text{第 1(個)四分位數} = Q_3 - Q_1$$

平均偏差

$$\text{母體資料：} MD = \frac{1}{N} \times \sum_{i=1}^{N} |x_i - \mu|$$

$$\text{樣本資料：} \text{md} = \frac{1}{n} \times \sum_{i=1}^{n} |x_i - \bar{x}|$$

離中差

$$\text{MD} = \frac{1}{n} \times \sum_{i=1}^{n} |x_i - M_e|$$

分組資料的平均偏差

母體資料：$MD = \frac{1}{N} \times \sum_{i=1}^{k} |m_i - \mu| \times f_i$

樣本資料：$md = \frac{1}{n} \times \sum_{i=1}^{k} |m_i - \bar{x}| \times f_i$

變異數與標準(偏)差

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{N} = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N} = \frac{1}{N} \times \sum_{i=1}^{N}(x_i-\mu)^2 = \frac{1}{N} \times \sum_{i=1}^{N}\left(x_i^2 - 2 \times x_i \times \mu + \mu^2\right)$$

$$= \frac{1}{N} \times \left[\sum_{i=1}^{N} x_i^2 - 2 \times \sum_{i=1}^{N} x_i \times \mu + \sum_{i=1}^{N} \mu^2\right] = \frac{1}{N} \times \left[\sum_{i=1}^{N} x_i^2 - 2 \times N \times \mu \times \mu + N \times \mu^2\right]$$

$$= \frac{1}{N} \times \left[\sum_{i=1}^{N} x_i^2 - N \times \mu^2\right] = \frac{\sum_{i=1}^{N} x_i^2}{N} - \mu^2 = \frac{1}{N} \times \left[\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}\right]$$

$$\text{Population SD} = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}} = \frac{\sqrt{N \times \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}}{N}$$

分組資料的變異數和標準(偏)差

母體變異數 $\sigma^2 = \frac{1}{N} \times \sum_{i=1}^{k}[(m_i - \mu)^2 \times f_i]$

母體標準(偏)差 $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{k}[(m_i - \mu)^2 \times f_i]}$

樣本變異數

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1}$$

樣本標準(偏)差

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\left(\sum_{i=1}^{n}x_i^2\right)-\left[2\times\left(\sum_{i=1}^{n}x_i\right)\times\bar{x}\right]+\sum_{i=1}^{n}\bar{x}^2}{n-1}} = \sqrt{\frac{\left(\sum_{i=1}^{n}x_i^2\right)-\left[2\times n\times\bar{x}\times\bar{x}\right]+n\times\bar{x}^2}{n-1}} = \sqrt{\frac{\left(\sum_{i=1}^{n}x_i^2\right)-2\times n\times\bar{x}^2+n\times\bar{x}^2}{n-1}} =$$

$$\sqrt{\frac{\sum_{i=1}^{n}x_i^2-n\times\bar{x}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2-n\times\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2-n\times\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n^2}}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1}}$$

分組資料的樣本變異數和標準(偏)差

樣本變異數 $S^2 = \frac{1}{n-1} \times \sum_{i=1}^{k}[(m_i - \bar{x})^2 \times f_i]$

樣本標準(偏)差 $S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^{k}[(m_i - \bar{x})^2 \times f_i]}$

變異係數(Coefficient of variation: CV)

母體資料：$CV = \frac{\sigma}{\mu}$

樣本資料：$CV = \frac{S}{\bar{x}}$

z-分布

$$z_i = \frac{x_i-\bar{x}}{S}$$

柴比氏定理

在任何觀測值的分布資料中，至少有$(1-\frac{1}{k^2})$比率或$(1-\frac{1}{k^2}) \times 100$ %的資料，分布在算術平均值 μ 為中心，$\pm k$ 個標準(偏)差 S 的範圍內。

共變異數

$$Cov(x, y) = \sigma_{xy} = \frac{\sum_{i=1}^{N}[(x_i-\mu_x)\times(y_i-\mu_y)]}{N}$$

$$cov(x, y) = S_{xy} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1}$$

皮爾森積差相關係數

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x\times\sigma_y} = \frac{\sigma_{xy}}{\sigma_x\times\sigma_y}$$

$$r_{xy} = \frac{cov(x,y)}{S_x\times S_y} = \frac{S_{xy}}{S_x\times S_y} = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\times\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\times\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{n}x_i\times y_i-\frac{\sum_i^n x_i\times\sum_i^n y_i}{n}}{\sqrt{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}\times\sqrt{\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}}}$$

Bowley 偏態係數 SK_B

$$SK_B = \frac{(Q_3-M_e)-(M_e-Q_1)}{(Q_3-M_e)+(M_e-Q_1)} = \frac{(Q_3-M_e)-(M_e-Q_1)}{Q_3-Q_1}$$

皮爾森(Karl Pearson)偏態係數 SK_p

母體資料：$SK_p = \frac{3\times(\mu-M_e)}{\sigma}$，$M_e$: 母體中位數

樣本資料：$SK_p = \frac{3\times(\bar{x}-m_e)}{S}$，$m_e$: 樣本中位數

四級動差法計算峰度係數

母體峰度：

$\beta_2 = \frac{\frac{1}{N}\times\sum_{i=1}^{k}[f_i\times(x_i-\mu)^4]}{\sigma^4}$

$\gamma_2 = \beta_2 - 3$

樣本峰度：

$$\beta_2 = \frac{\frac{1}{n-1}\times\sum_{\mathrm{i=1}}^{\mathrm{k}}[f_i\times(x_i-\bar{x})^4]}{S^4}$$

$$\gamma_2 = \beta_2 - 3$$

Excel 軟體中峰度係數的計算方式

$$\gamma_2 = \left[\frac{n\times(n+1)}{(n-1)\times(n-2)\times(n-3)}\times\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\mathrm{S}}\right)^4\right]-\frac{3\times(n-1)^2}{(n-2)\times(n-3)}$$