三、敘述性統計：數量方法

Chapter 3 Descriptive Statistics: Numerical Methods

學習目標

知識(認知)

- 可以描述各種敘述性數值表達方式適用的狀況。
- 分辨各種敘述性統計數值表達方式之間的差異性。
- 評價各種敘述性統計數值表達方式的使用價值。

技能

- 能夠計算各種敘述性統計之數值。
- 能夠計算探索性資料分析之五種數值。
- 能夠依循步驟，製作盒鬚圖。

態度(情意)

- 意識到各種敘述性統計數值表達方式的重要性。
- 可以依循需求判斷，選擇採用適當的敘述性統計數值表達方式。
- 建立自主判斷離群值的觀念。

學習內容

- ■ 中心位置測定值計算
- ■ 分散度測定值計算
- ■ 評量相對位置之計算和偵測離群值
- ■ 探索性資料分析
- ■ 兩個變數相關性的評量計算
- ■ 偏態和鋒度計算

學習建議

【開始上課】下載PDF講義預習

依據各單元順序，觀賞數位教材內容

針對數位教材中練習題目(來賓考試)演練

同步上課學習

相關學習主題討論

學習評量

提示舊經驗：激發學習動機

想要獲得哪些的數值？

■ 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、用餐同伴分布與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。

中心位置測定值

分散度測定值

評量相對位置

偵測離群值

探索性資料分析

兩變數相關程度的評量

偏態與峰度

達到具體客觀清楚呈現的報告目標

三、敘述性統計：數量方法

3.1中心位置測定值
3.2分散度測定值
3.3評量相對位置和偵測離群值
3.4探索性資料分析
3.5兩個變數相關性的評量
3.6偏態與峰度

3.1中心位置測定值

- 3.1.1算術平均值
- 3.1.2中位數
- 3.1.3眾數
- 3.1.4四分位數
- 3.1.5十分位數
- 3.1.6百分位數
- 3.1.7幾何平均值
- 3.1.10縮減式平均值
- 3.1.11溫塞平均值
- 3.1.12加權算術平均值
- 3.1.13全距中點
- 3.1.14中樞紐

中心位置測定值【學習標的】

- ■ 瞭解不同中心位置測定值評量方式
  - ◆意義
  - ◆計算方法
  - ◆使用時機
  - ◆優缺點
  - ◆彼此之間差異

3.1.1算術平均值

■ 平均值(mean)是代表觀測值(屬於區間尺度或比例尺度)分布集中化的位置之數值。

◆觀測值來自母體，平均值μ(讀音mu; muw)符號代表

◆觀測值來自樣本，平均值$\bar{x}$ (讀音x bar)符號代表。

□適用於數值資料分布屬於左右對稱，僅有一個波峰時。

符號	母體	樣本
平均數	μ	$\bar{x}$

未分組資料算術平均值

分組資料算術平均值

未分組資料平均值計算方式

- 母體之特定變數有N個觀測值，其數值分別為$x_1, x_2, x_3, \ldots, x_N$
  - 母體平均值(population mean) $\mu = \sum_{i=1}^{N} \frac{x_i}{N}$
- 樣本之特定變數有n個觀測值，其數值分別為$x_1, x_2, x_3, \ldots, x_n$
  - 樣本平均值(sample mean)$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$
    - $\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$
    - Σ: 希臘大寫字母sigma，意指加法運算

Excel函數：AVERAGE

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

算術平均值之性質

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 $= |2-5| + |3-5| = 5$

◻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 $= 5$。

貢獻不均，極端影響

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)有極端值出現時，算術平均值就非常敏感而受此極端值影響。
  - ◘利用Excel軟體示範說明

加減乘除運算

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，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，其新資料之算術平均值等於原資料算術平均值 ×/÷ 此一常數。

範例加乘(1/2)

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。

範例加乘(2/2)

- 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$5 \times 10 = 50$，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

偏差和等於0

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$\sum_{i=1}^{n}(x_i-\bar{x})=0$。

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

E.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

偏差平方和最小

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。
- 偏差平方和(sum of squares, SS)
  - $\text{SS} = \sum_{i=1}^{n}(x_i - \bar{x})^2$
  - $\text{SS} = \sum_{i=1}^{N}(x_i - \mu)^2$
  - $\sum_{i=1}^{n}(x_i - \bar{x})^2 \le \sum_{i=1}^{n}(x_i - \text{A})^2$，其中A：可以屬於任何數值，此關係式皆成立。

範例：偏差平方和最小

- 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$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{2+3+6+9}{4} = \frac{20}{4} = 5$。
- 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。

3.1.2中位數

- 一般以Me、M_e、Md或M_d符號表示中位數或中量(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，而產生分布傾斜的情況，可以具體的表達出資料的中間特徵。

未分組資料中位數

分組資料中位數

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，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。

未分組資料的中位數計算方式

- 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：
  - n為奇數(odd)
    - 中位數即為第$\frac{n+1}{2}$序位的觀測值。
  - n為偶數(even)
    - 中位數為第$\frac{n}{2}$和$\frac{n}{2}+1$兩序位觀測值的算術平均值。
- 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。

中位數優缺點

優點	缺點
• 不因觀測值中有極大值或極小值而改變中位數的大小。 • 中位數與各數值之差絕對值的總和最小，即$\Sigma\lvert x_i - M_e\rvert$的數值最小。 • 最適用於順序尺度數值型態之變數。	• 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。 • 其他觀測值的數值之高低對中位數沒有任何影響。

中位數(數量奇數)範例

範例3.7 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，求阿文海產店上週來店消費人數的中位數？

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

答案：中位數162人

中位數(數量偶數)範例

範例3.8 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	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$\frac{n}{2} + 1 = 5 + 1 = 6$

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

答案：中位數164人

Excel函數：MEDIAN

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

3.1.3眾數

- 眾數(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，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。

眾數範例一

範例3.10 依據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，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

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

眾數範例二

要搞清楚觀察對象和觀測值喔

範例3.11　依據調查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，求使用主要交通工具的眾數？

名目尺度

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

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。

答案：主要交通工具的眾數為機車

Excel函數：MODE

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

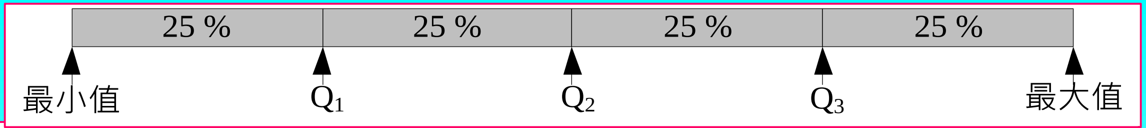
確定

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_1

3.1.4四分位數



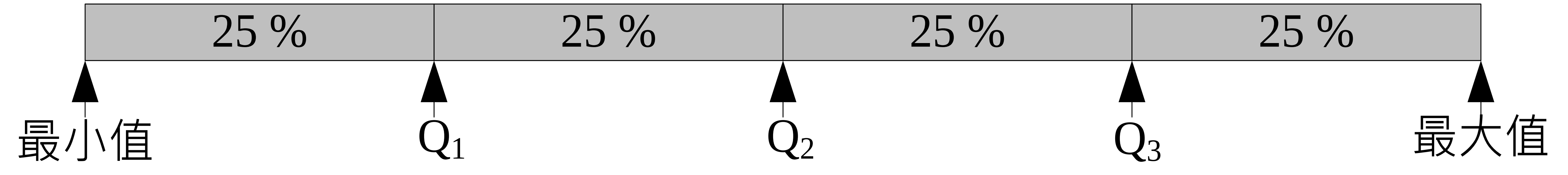
- 將觀測值數值資料由小排至大，再依據觀測值數量n均分成四等分。

第一種為【n - 1模式】，第二種為n + 1模式

- 位於第1個等分位置的數值，即第1(個)四分位數
  - 符號Q_1標示，表示有25 %基本單位數量的觀測值比Q_1小，另有75 %基本單位數量的觀測值比Q_1大
- 位於第2個等分位置的數值，為第2(個)四分位數
  - 符號Q_2標示，表示有50 %基本單位數量的觀測值比Q_2小，另有50 %基本單位數量的觀測值比Q_2大
- 位於第3個等分位置的數值，為第3(個)四分位數
  - 符號Q_3標示，表示有75 %基本單位數量的觀測值比Q_3小，另有25 %基本單位數量的觀測值比Q_3大。

四分位數相對位置

- 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。
  - ◆定義Q_0為最小觀測值之數值；Q_4為最大觀測值之數值。
- 一般四分位數值(Q_1到Q_3)不受極端觀測(數)值的影響，Q_0和Q_4除外。



未分組資料四分位數

分組資料四分位數

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。
- **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。
- 樣本數量n個觀測值，計算Q_i序位：$O(Q_i) = (i \times \frac{n-1}{4}) + 1$
  - ◆其中i = 1, 2, or 3。
- $O(Q_i) = (i \times \frac{n-1}{4}) + 1$序位數值為整數
  - ◆此序位所對應觀測值即為Q_i；
- $O(Q_i) = (i \times \frac{n-1}{4}) + 1$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。

$O(Q_i)$數值為整數範例

範例3.12 阿力餐廳菜單中美食套餐，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、51套，求Q_1、Q_2及Q_3。

題解：

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

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

$O(Q_3) = (i \times \frac{n-1}{4}) + 1 = (3 \times \frac{9-1}{4}) + 1 = 7$，$Q_3$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

$O(Q_i)$數值不為整數範例

範例3.13 阿力餐廳菜單中美食套餐，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。(答案小數取2位)

題解：

$O(Q_1) = (i\times \frac{n-1}{4}) + 1 = (1 \times \frac{10-1}{4}) + 1 = 3.25$，$Q_1$落於第3與4序位之間，$Q_1 = 38 + 0.25 \times (41 - 38) = 38.75$套

$O(Q_2) = (i \times \frac{n-1}{4}) + 1 = (2 \times \frac{10-1}{4}) + 1 = 5.50$，$Q_2$落於第5與6序位之間，$Q_2 = 45 + 0.50 \times (46 - 45) = 45.50$套

$O(Q_3) = (i \times \frac{n-1}{4}) + 1 = (3 \times \frac{10-1}{4}) + 1 = 7.75$，$Q_3$落於第7與8序位之間，$Q_3 = 47 + 0.75 \times (50 - 47) = 49.25$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

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

Excel函數：QUARTILE

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

3.1.5十分位數

第一種為n - 1模式，第二種為n + 1模式

■ 將具有順序性質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。

◆ 定義D_0為最小觀測值之數值；D_{10}為最大觀測值之數值。

◆ 一般十分位數值(D_1到D_9)不受極端觀測(數)值的影響，D_0和D_{10}除外。

未分組資料十分位數

分組資料十分位數

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。
- **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。
- $O(D_i) = (i \times \frac{n-1}{10}) + 1$序位數值為整數
  - 第i十分位數D_i為第$O(D_i)$序位的觀測值。
- $O(D_i) = (i \times \frac{n-1}{10}) + 1$序位數值為非整數
  - 第i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。

未分組十分位數計算範例

範例3.15 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5十分位數D_5和第7十分位數D_7。

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

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

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

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,…, 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。
- ◆ 定義P_0為最小觀測值之數值；P_{100}為最大觀測值之數值。

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。
- **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
  - i：百分位之號碼
  - n：樣本數量
- $O(P_i) = (i \times \frac{n-1}{100}) + 1$序位數值為整數
  - 第i百分位數P_i為第$O(P_i)$序位的觀測值。
- $O(P_i) = (i \times \frac{n-1}{100}) + 1$序位數值為非整數
  - 第i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。

未分組數值的百分位數計算範例

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58百分位數P_{58}和第82百分位數P_{82}。

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58百分位數序位：$(i \times \frac{n-1}{100}) + 1 = (58 \times \frac{15-1}{100}) + 1 = 9.12$序位

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

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

中位數、四、十和百分位數對照表

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}

Excel函數PERCENTILE

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}則輸入0.55)

←

確定

可使用於四分位數與十分位數計算

示範Excel資料剖析方法

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該試卷者，老師即關閉沒有進入者的考卷

上課練習112ch3_2

3.1.7幾何平均值

運算的對象觀測值是【原來數值】而非比例或增加率

■一般資料各取對數後所得之算術平均值為對數單位之資料，由此對數單位的算術平均值再取反對數，所得之值稱為幾何平均值(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

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

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

等號兩邊取對數

■樣本的幾何平均值標示為$\bar{x}_\mathrm{g}$

◆$\bar{x}_\mathrm{g} = \text{antilog}(\frac{1}{n}\times\sum_{i=1}^{n}\log x_i) = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} = (\prod_{i=1}^{n} x_i)^{\frac{1}{n}}$

□$\log \bar{x}_\mathrm{g} = \frac{1}{n}\times\sum_{i=1}^{n}\log x_i$

等號兩邊取對數

幾何平均值的應用

- 幾何平均值使用於指數、滴定濃度、改變率、生長率、投資報酬率、增殖率等資料。
  - 適用於等比數列(在數列$a_1, a_2, a_3, \ldots, a_n$中，$\frac{a_2}{a_1} = \frac{a_3}{a_2} = \frac{a_4}{a_3} = \ldots = \frac{a_n}{a_{n-1}} = r$，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$\bar{x}_{\mathrm{g}}$小於等於算術平均值$\bar{x}$，即$\bar{x}_{\mathrm{g}} \le \bar{x}$。
  - 當各觀測值之數值皆相等時，幾何平均數$\bar{x}_{\mathrm{g}}$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，此狀況下不宜採用幾何平均值。

幾何平均值計算範例

範例3.19 請計算3、4和5三個樣本觀測值的幾何平均值？

題解：觀測值的數量$n = 3$

幾何平均值$\bar{x}_{\mathrm{g}} = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} = \sqrt[3]{3 \times 4 \times 5} = 3.9149$

答案：幾何平均值 = 3.9149

傳統Excel運算操作方式示範

Excel函數GEOMEAN

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

幾何平均值應用範例(1/2)

範例3.20 黑心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？

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，x_2是明年再調降後的薪資數值

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

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

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 %

幾何平均值應用範例(2/2)

範例3.20 黑心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？

若原來薪水是新臺幣20000元

今年調降12 %：　$20000 \times (1-0.12) = 17600$元

明年調降8 %：　$17600 \times (1-0.08) = 16192$元

利用平均調降率計算

今年調降10.02 %：$20000 \times (1-0.100222) = 17995.6$元

明年調降10.02 %：$17995.6 \times (1-0.100222) = 16192$元 → 兩個金額一樣，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，$\frac{今天薪水}{原始薪水} = \frac{17995.5550}{20000} = \frac{明年薪水}{今年薪水} = \frac{16192}{17995.5550} = 0.899778$，公比數值一樣。

練習

練習3.8 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(年平均成長率)。

題解：欲計算幾何平均值時，若題目提供的是成長率，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；成長率-15%就必須換算為1 + (-15 %) = 0.85。

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

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 %

期間平均增加率

- 幾何平均值亦可使用於計算一段期間內，週期次數為n，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
  - 期間平均增加率$G = \sqrt[n]{\frac{\text{期末值}}{\text{期始值}}} - 1$

期間平均增加率範例

範例3.21 小玟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。

題解：週期次數$n = 10$

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

答案：年平均成長率 = 0.3553 = 35.53 %

範例驗證(1/2)

2000年推估的店數 = 1999年店數 × (1 + 年平均成長率) = 12 × (1 + 0.3553) = 16.3

2001年推估的店數 = 2000年店數 × (1 + 年平均成長率) = 16.3 × (1 + 0.3553) = 22.0

2002年推估的店數 = 2001年店數 × (1 + 年平均成長率) = 22.0 × (1 + 0.3553) = 29.9

2003年推估的店數 = 2002年店數 × (1 + 年平均成長率) = 29.9 × (1 + 0.3553) = 40.5

2004年推估的店數 = 2003年店數 × (1 + 年平均成長率) = 40.5 × (1 + 0.3553) = 54.9

2005年推估的店數 = 2004年店數 × (1 + 年平均成長率) = 54.9 × (1 + 0.3553) = 74.4

2006年推估的店數 = 2005年店數 × (1 + 年平均成長率) = 74.4 × (1 + 0.3553) = 100.8

範例驗證(2/2)

2007年推估的店數 = 2006年店數 × (1 + 年平均成長率) = 100.8 × (1 + 0.3553) = 136.6

2008年推估的店數 = 2007年店數 × (1 + 年平均成長率) = 136.6 × (1 + 0.3553) = 185.2

2009年推估的店數 = 2008年店數 × (1 + 年平均成長率) = 185.2 × (1 + 0.3553) = 251 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。

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_3

3.1.10 縮減式平均值

- 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 (Trimmed Mean)，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。

縮減式平均值計算範例

範例3.24　請計算40、58、60、70、72和99六個樣本觀測值的縮減式平均值 (trimmed mean) $\bar{x}_t$？

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

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

答案：縮減式平均值$\bar{x}_t = 65.0$

3.1.12 加權算術平均值

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

加權算術平均值範例

範例3.26 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。

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

3.1.13全距中點

- 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。
- $全距中點 = \dfrac{最高觀測值 + 最低觀測值}{2}$

3.1.14中樞紐

- 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 $= \dfrac{第_1四分位數+第_3四分位數}{2} = \dfrac{Q_1+Q_3}{2}$

3.1.15 中心位置測定值彼此之差異

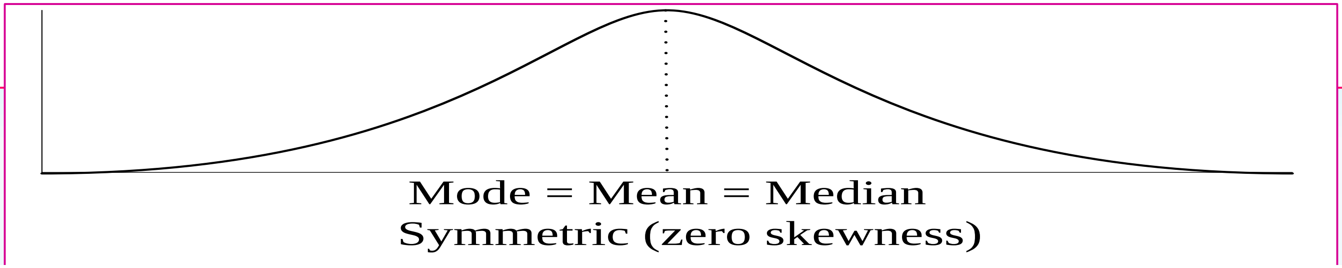
平均數	採用頻率	存在性	觀測值採計	極端值影響
算術平均值	最常使用	總是存在	全部	會
中位數	普通使用	總是存在	一個	無
眾數	很少使用	可以不存在，亦可有多個	數個	無
幾何平均值	有時使用	總是存在	全部	會
中樞紐	極少使用	總是存在	兩個	無

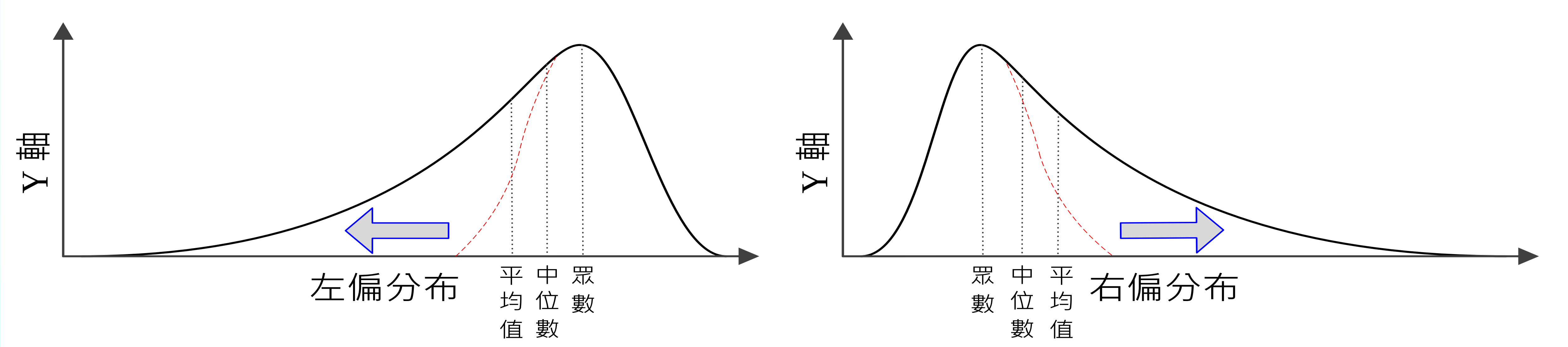
對稱與偏斜

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

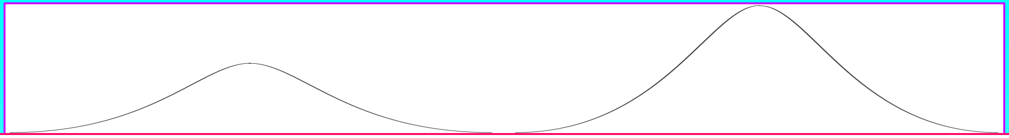
左偏或負偏的次數分布中，算術平均值(mean) μ < 中位數(median) M_e < 眾數(mode) M_o。

右偏或正偏的次數分布中，算術平均值(mean) μ > 中位數(median) M_e > 眾數(mode) M_o。





3.2分散度測定值



3.2.1全距(Range)

3.2.2四分位距(Inter-quartile range, IQR)

3.2.4變異數與標準差

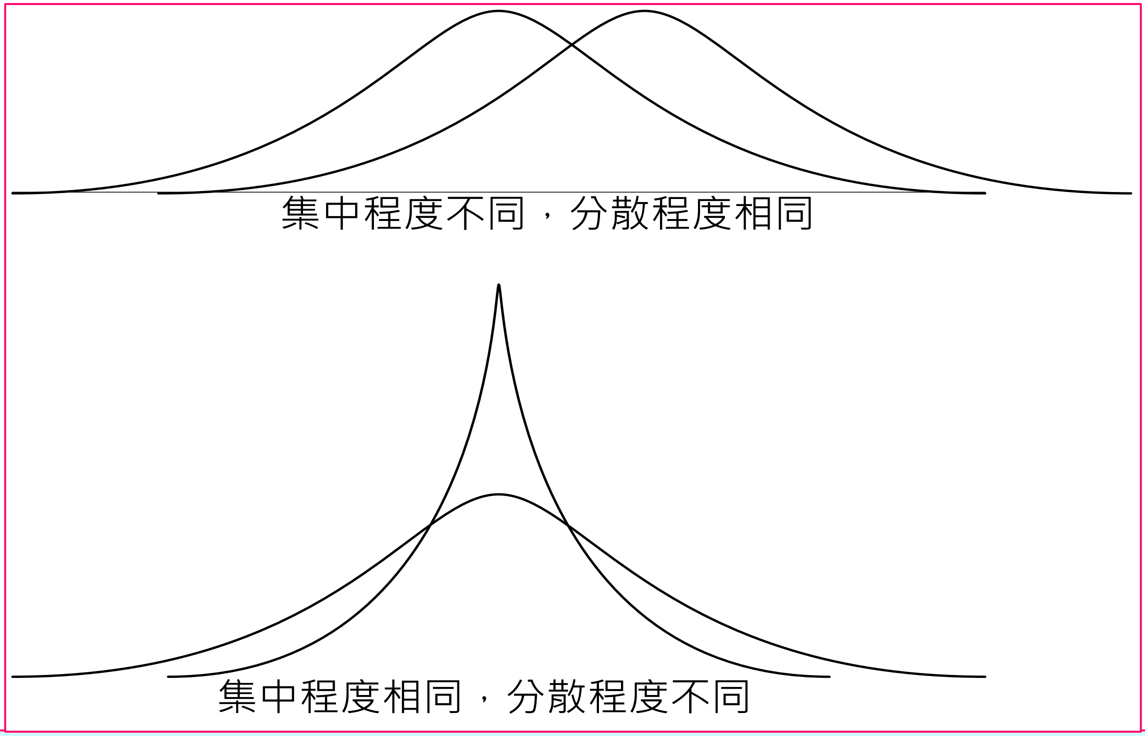
3.2.5樣本變異數(Sample Variance)

3.2.6變異數與標準差之性質

3.2.7相對分散程度

差異量數

- 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。



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
  - 計算容易與方便。
- 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函數

= Large(資料區間,1) – Small(資料區間,1)

3.2.2四分位距

- 第3(個)四分位數與第1(個)四分位數的距離即四分位距(Inter-quartile range, IQR)，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，其他的數值並未被納入計算。

四分位距計算範例

範例3.29 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IQR。

題解：

$$IQR = Q_3 - Q_1 = 89.0 - 59.0 = 30.0$$

答案：$Q_1 = 59.0$元；$Q_3 = 89.0$元；四分位距$IQR = 30.0$元

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_4

3.2.4變異數與標準差

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

標準差

- 標準差(standard deviation, SD)：以σ(讀音sigma)符號表示。
  - 標準差單位與觀測值單位相同。
  - 標準差數值一定都是屬於正值，最小標準差數值為0，代表每一個基本單位的觀測值數值皆相等。
- 母體標準差 SD = σ = $\sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$
  - $\sum_{i=1}^{N}(x_i-\mu)^2 = \sum_{i=1}^{N}x_i^2 - 2\times\mu\times\sum_{i=1}^{N}x_i + \sum_{i=1}^{N}\mu^2 = \sum_{i=1}^{N}x_i^2 - 2\times\frac{\sum_{i=1}^{N}x_i}{N}\times\sum_{i=1}^{N}x_i + N\times\mu^2 = \sum_{i=1}^{N}x_i^2 - 2\times\frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N} + N\times\left(\frac{\sum_{i=1}^{N}x_i}{N}\right)^2 = \sum_{i=1}^{N}x_i^2 - \frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}$
  - $\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N}x_i^2 - \frac{\left(\sum_{i=1}^{N}x_i\right)^2}{N}}{N}} = \frac{\sqrt{N\times\sum_{i=1}^{N}x_i^2-\left(\sum_{i=1}^{N}x_i\right)^2}}{N}$

變異數與標準差範例

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, 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差。

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
$\sum_{i=1}^{N} x_i = 510$		$\sum_{i=1}^{N}(x_i-\mu)^2 = 1880$
$\mu = 102$		

母體變異數$\sigma^2 = \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{N} = \frac{1880}{5} = 376$人^2

母體標準差$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{1880}{5}} = \sqrt{376} = 19.39$ (人)

答案：變異數 = 376人^2；標準差 = 19.39人

Excel函數運算母體

- 母體變異數
  - VAR(.)P → Population
- 樣本變異數
  - VAR(.S) → Sample
- 母體標準差
  - STDEV(.)P → Population
- 樣本標準差
  - STDEV(.S) → Sample

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。
- 將樣本觀測值的偏差平方和$[\sum_{i=1}^{n}(x_i - \bar{x})^2]$除以$n - 1$，獲得樣本變異數，此樣本變異數$S^2$是母體變異數$\sigma^2$的不偏估計值。
- $S^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

樣本變異數說明範例

範例3.34 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$\mu = 5$，母體變異數$\sigma^2 = 6$，從此母體中抽取$n = 2$個觀測值為樣本，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	

所有樣本變異數的平均值即等於母體的變異數。

樣本變異數即為母體變異數的不偏估值。

樣本變異數與標準差

- 樣本變異數(Sample variance)或均方(Mean Square)，S^2符號代表，變異數單位 = [觀測值單位]^2：
  - 樣本變異數$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$
- 樣本標準差(Sample standard deviation)，S符號代表，標準差單位 = 觀測值單位：
  - 樣本標準差$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\left(\sum_{i=1}^{n} x_i^2\right) - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$

樣本標準差範例

範例3.35 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，請計算高雄旅館專職員工數分布之標準差(standard deviation)。

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
$\sum_{i=1}^{n} x_i = 510$		$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 1880$
$\bar{x} = 102$		

樣本標準差$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{1880}{5-1}} = \sqrt{470} = 21.7$ (人)

答案：樣本標準差 = 21.7人

Excel函數運算樣本

- 樣本變異數
  - VAR(.S) → Sample
- 母體變異數
  - VAR(.)P → Population
- 樣本標準差
  - STDEV(.S) → Sample
- 母體標準差
  - STDEV(.)P → Population

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_5

3.2.6變異數與標準差之性質

- 變異數和標準差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差的數值。

全距概略法則

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

沒有極端值出現時，才成立

■ 標準差 $\fallingdotseq \dfrac{全距(range)}{4} = \dfrac{最高數值-最低數值}{4}$

■ 已經知道平均值和標準差的數值，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：

◆最大觀測值 = 平均值 + 2 × 標準差

◆最小觀測值 = 平均值 – 2 × 標準差

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一個常數時，其變異數和標準差數值皆不會改變。

◆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方別為

◆$x_1, x_2, x_3, \ldots., x_N$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$\mu_y = \mu_x + C$

■ 兩母體之變異數分別為σ_x^2及σ_y^2，$\sigma_x^2 = \sigma_y^2$；$\sigma_x = \sigma_y$

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

加減對分散程度影響

範例3.37　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差。

題解：

x_i	$x_i-\mu$	$(x_i-\mu)^2$		x_i	$x_i-\mu$	$(x_i-\mu)^2$
160	-6	36		165	-6	36
165	-1	1		170	-1	1
170	4	16		175	4	16
172	6	36		177	6	36
163	-3	9		168	-3	9
$\sum_{i=1}^{N} x_i = 830$		$\sum_{i=1}^{N}(x_i-\mu)^2 = 98$		$\sum_{i=1}^{N} x_i = 855$		$\sum_{i=1}^{N}(x_i-\mu)^2 = 98$
$\mu = 166$				$\mu = 171$		

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)

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一個常數後，對新形成的母體分布的變異數影響。
- 設原觀測值與新觀測值方別為
  - $x_1, x_2, x_3, \ldots., x_N$及$y_1, y_2, y_3,\ldots., y_N$
  - 其中$y_i = C \times x_i$ (即各樣本乘一常數$C \neq 0$)
- 兩母體平均值的關係為$\mu_y = C \times \mu_x$
- 兩母體變異數分別為σ_x^2及σ_y^2，$\sigma_y^2 = C^2 \times \sigma_x^2$，$\sigma_y = |C| \times \sigma_x$
  - $\sigma_y^2 = \frac{\sum_{i=1}^{N}(y_i-\mu_y)^2}{N} = \frac{\sum_{i=1}^{N}(C\times x_i-C\times\mu_x)^2}{N} = \frac{C^2\times\sum_{i=1}^{N}(x_i-\mu_x)^2}{N} = C^2 \times \sigma_x^2$

乘除對分散程度影響

範例3.38 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差。

題解：

x_i	$x_i-\mu$	$(x_i-\mu)^2$		x_i	$x_i-\mu$	$(x_i-\mu)^2$
160	-6	36		1.60	-0.06	0.0036
165	-1	1		1.65	-0.01	0.0001
170	4	16		1.70	-0.04	0.0016
172	6	36		1.72	-0.06	0.0036
163	-3	9		1.63	-0.03	0.0009
$\sum_{i=1}^{N} x_i =$ 830		$\sum_{i=1}^{N}(x_i-\mu)^2$ = 98		$\sum_{i=1}^{N} x_i =$ 8.30		$\sum_{i=1}^{N}(x_i-\mu)^2$ = 0.0098
$\mu = 166$				$\mu = 1.66$		

乘除對分散程度的變化

- 原始身高(公分)分布母體標準(偏)差σ = $\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{98}{5}} = \sqrt{19.6}$ = 4.4272 (公分)
- 以公尺評量身高分布母體標準(偏)差σ = $\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\frac{0.0098}{5}} = \sqrt{0.00196}$ = 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(公分)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

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，對新形成的母體分布的變異數影響。
- 設原觀測值與新觀測值方別為
  - $x_1, x_2, x_3, \ldots, x_N$及$y_1, y_2, y_3, \ldots, y_N$
  - 其中$y_i = C_1 \times x_i + C_2$，其中$C_1, C_2 \in R$屬於實數
- 兩母體平均值的關係為$\mu_y = C_1 \times \mu_x + C_2$
- 兩母體變異數分別為σ_x^2及σ_y^2，$\sigma_y^2 = C_1^2 \times \sigma_x^2$，$\sigma_y = |C_1| \times \sigma_x$

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_6

3.2.6.4標準差數值最小化

- 利用算術平均值$\bar{x}$為依據所計算獲得的標準差，比其他非算術平均值所計算獲得的標準差為小。
- $S=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} < \sqrt{\frac{\sum_{i=1}^{n}(x_i-A)^2}{n-1}}$，當$\bar{x}\neq A$

範例說明

說明範例(1/2)

範例3.40 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，請計算該五位女性員工身高分布之標準差。

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
$\sum_{i=1}^{N} x_i = 830$		$\sum_{i=1}^{N}(x_i - \mu)^2 = 98$		$\sum_{i=1}^{N}(x_i - 165)^2 = 103$		$\sum_{i=1}^{N}(x_i - 167)^2 = 103$
$\mu = 166$						

說明範例(2/2)

範例3.40 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，請計算該五位女性員工身高分布之標準差。

題解：

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

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，運算出來的標準差 = 4.4272公分為最小數值

3.2.6.5標準差大於等於零

- 一般資料分布中標準差數值皆大於零，$S > 0$；若標準差$S = 0$表示資料中每一個觀測值(數值)皆相等。
  - 標準差$S > 0$，在$x_i \neq x_j$，當$i \neq j$
  - 標準差$S = 0$，在$x_i = x_j$，當$i \neq j$

範例

範例3.41 燕巢一家觀光旅館，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差。

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)，其標準差即為0.0000。

答案：身高分布變異數 = 0.00公分^2；標準差 = 0.0000公分

3.2.6.6兩母體合併後平均值與標準差

資料	數量	平均值	標準差	合併	數量	平均值	標準差
母體A	N_A	μ_A	σ_A	$\rightarrow$	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

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

合併：標準差算法

第1種算法

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

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

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

第2種算法

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

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

合併計算母體平均值與標準差(1/2)

範例3.42 A和B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A、B與A和B兩班合併計算時，該次期中考成績的平均分數和標準差。

A班全班學生成績

80	70	80	90	65	80	75	85	50	60

B班全班學生成績

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

題解：A班母體平均值$\mu_A = 73.50$，A班母體標準差$\sigma_A = 11.63$，B班母體平均值$\mu_B = 78.67$，B班母體標準差$\sigma_B = 11.86$，合併後母體平均值$\mu = 76.32$，合併後母體標準差$\sigma = 12.03$

$$\mu = \frac{N_A \times \mu_A + N_B \times \mu_B}{N} = \frac{N_A \times \mu_A + N_B \times \mu_B}{N_A + N_B} = \frac{10 \times 73.50 + 12 \times 78.67}{10+12} = \frac{735+944}{22} = 76.32$$

合併計算母體平均值與標準差(2/2)

範例3.42 A和B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A、B與A和B兩班合併計算時，該次期中考成績的平均分數和標準差。

標準差第一種算法 $\sigma = \sqrt{\frac{N_A\times(\sigma_A^2+\mu_A^2)+N_B\times(\sigma_B^2+\mu_B^2)}{N_A+N_B} - \mu^2}$

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

標準差第二種算法 $\sigma = \sqrt{\frac{N_A\times\sigma_A^2+N_A\times(\mu_A-\mu)^2}{N} + \frac{N_B\times\sigma_B^2+N_B\times(\mu_B-\mu)^2}{N}} =$

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

$$\sqrt{\frac{1352.5+79.42}{22} + \frac{1686.67+66.18}{22}} = \sqrt{\frac{1431.92}{22} + \frac{1752.85}{22}} = 12.03$$

答案：A班平均值 = 73.50分，標準差 = 11.63分；B班平均值 = 78.67分，標準差= 11.86分；合併AB班平均值 = 76.32分，標準差= 12.03分

合併計算拆解運算範例

範例3.43 A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，標準差5分，其中男生5位學生，平均分數為75分，標準差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差。

題解：母體平均值$\mu = 80 = \frac{N_A \times \mu_A + N_B \times \mu_B}{N} = \frac{5\times75+10\times\mu_F}{15} \rightarrow 80 \times 15 = 5 \times 75 + 10 \times \mu_F \rightarrow$ 女生平均值$\mu_F = 82.5$分

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

$25 \times 15 = 375 = 45 + 125 + 10 \times \sigma_F^2 + 62.5 \rightarrow$ 女生變異數$\sigma_F^2 = 14.25$分^2 $\rightarrow$ 女生標準差$\sigma_F = 3.7749$分

答案：女生平均值 = 82.50分，標準差 = 3.77分

3.2.6.7兩樣本合併後平均值與標準差

資料	數量	平均值	標準差	合併	數量	平均值	標準差
樣本A	n_A	$\bar{x}_A$	S_A	$\rightarrow$	n	$\bar{x}$	S
樣本B	n_B	$\bar{x}_B$	S_B				

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

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

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

合併計算樣本平均值與標準差範例(1/2)

範例3.44 A和B兩班同時修統計學課程，某次期中考成績：A班隨機抽取10位學生；B班隨機抽取12位學生，其成績分別如下所示。試分別估算A、B與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差。

A班學生成績

80	70	80	90	65	80	75	85	50	60

B班學生成績

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

合併計算樣本平均值與標準差範例(2/2)

範例3.44 A和B兩班同時修統計學課程，某次期中考成績：A班隨機抽取10位學生；B班隨機抽取12位學生，其成績分別如下所示。試分別估算A、B與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差。

題解：A班平均分數$\bar{x}_A = 73.50$，A班分數標準差$S_A = 12.2588$，B班平均分數$\bar{x}_B = 78.67$，B班分數標準差$S_B = 12.3828$，合併計算後樣本平均值$\bar{x} = 76.32$，合併計算後標準差$S = 12.3149$

$$\bar{x} = \frac{n_A \times \bar{x}_A + n_B \times \bar{x}_B}{n} = \frac{n_A \times \bar{x}_A + n_B \times \bar{x}_B}{n_A + n_B} = \frac{10 \times 73.50 + 12 \times 78.67}{10 + 12} = \frac{735 + 944}{22} = 76.32$$

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

$$\sqrt{\frac{(10-1) \times 12.2588^2 + 10 \times (73.50 - 76.32)^2}{10 + 12 - 1} + \frac{(12-1) \times 12.3828^2 + 12 \times (78.67 - 76.32)^2}{10 + 12 - 1}} = 12.3149$$

答案：A班平均值 = 73.50分，標準差 = 12.26分；B班平均值 = 78.67分，標準差 = 12.38分；合併AB班平均值 = 76.32分，標準差 = 12.31分

3.2.6.8 k組資料合併平均值與標準差

- 若k組資料母體或樣本的個別數量、平均值和標準差已知，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	$\mu=\frac{\sum_{i=1}^{k}N_i\times\mu_i}{\sum_{i=1}^{k}N_i}$	$\sigma^2=\frac{\sum_{i=1}^{k}\left[N_i\times\sigma_i^2+N_i\times(\mu_i-\mu)^2\right]}{\sum_{i=1}^{k}N_i}$
樣本	$\bar{x}=\frac{\sum_{i=1}^{k}n_i\times\bar{x}_i}{\sum_{i=1}^{k}n_i}$	$S^2=\frac{\sum_{i=1}^{k}\left[(n_i-1)\times S_i^2+n_i\times(\bar{x}_i-\bar{x})^2\right]}{\left(\sum_{i=1}^{k}n_i\right)-1}$

3.2.6.9獨立變數之變異數加法定律

■ 若X和Y是相互獨立的隨機變數($\rho_{xy} = 0$)，獨立變數之變異數加法定律

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

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

變異數加法定律範例

範例3.45　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，計算隨機變數X加上隨機變數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加上隨機變數Y有三個基本單位觀測值，形成新隨機變數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

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

代表隨機找X隨機變數一個觀測值加上隨機找Y隨機變數一個觀測值，合計後，新形成的分布，其變異數為4.8889。

答案：相加後新隨機變數Z分布的變異數4.8889

變異數減法範例

範例3.46 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，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

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

代表隨機找X隨機變數一個觀測值減去隨機找Y隨機變數一個觀測值，合計後，新形成的分布，其變異數為4.8889。

答案：相減後新隨機變數Z分布的變異數4.8889

獨立變數之變異數加法定律範例

範例3.47 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$\sigma_x^2 = 12$ kg²，女性體重分布的變異數$\sigma_y^2 = 10$ kg²，請計算男生體重加女生體重後分布的變異數。

題解：依據獨立變數之變異數加法定理

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

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 kg²。

答案：男生體重加女生體重後分布的變異數22 kg²

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

變異數一般加法定律範例

範例3.48 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$\rho_{xy} = 0.2000$，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。

題解：依據隨機變數之變異數一般加法定律

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

$$= 1300 - 138.56 = 1161.44 \text{ 元}^2$$

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：月平均收入扣除餐飲支出後分布的變異數1161 元^2

有使用到下一個單元CV值概念

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_7

3.2.7相對分散程度

■ 採用全距、四分位距、平均偏差、變異數、標準差等評量觀測值的分散程度時，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。

■ 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$CV = \dfrac{\text{標準差}}{\text{平均值}}$

- ▫ 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。

相對分散程度的計算範例

範例3.51 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。

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
$\Sigma x_i = 273$		$\Sigma(x_i-\bar{x})^2 = 93.2$
$\bar{x} = 54.6$		

$$\text{身高}S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{243.2}{5-1}} = \sqrt{60.8} = 7.80 \text{ cm}$$

$$\text{身高}CV = \frac{S}{\bar{x}} = \frac{7.80}{166.6} = 0.0468$$

$$\text{體重}S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{93.2}{5-1}} = \sqrt{23.3} = 4.83 \text{ kg}$$

$$\text{體重}CV = \frac{S}{\bar{x}} = \frac{4.83}{54.6} = 0.0884$$

身高$CV = 0.0468 <$ 體重$CV = 0.0884$

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

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_8

3.3 評量相對位置和偵測離群值

- 3.3.1 z-分布
- 3.3.2柴比氏定理
- 3.3.3經驗法則
- 3.3.4偵測離群值
- 3.3.5標準誤(差)

3.3.1 z-分布

- 透過平均值和標準差可以使用於標定特定觀測值在一個母體(樣本)分布中的相對位置。
  - 利用平均值(中心點)為0，變異數為1和標準差為1為標準化的條件。
- z-值：可視為觀測值x_i數值與樣本平均值$\bar{x}$之間有多少個標準差的差距。
  - $z_5 = 1$，代表觀測值x_5高於樣本平均值$\bar{x}$達到1個單位標準差的距離。
  - $z_6 = -2$，代表觀測值x_6低於樣本平均值$\bar{x}$有2個單位標準差的距離。
    - $z_i > 0$，代表觀測值$x_i > \bar{x}$；$z_i < 0$，代表觀測值$x_i < \bar{x}$。
  - z-值屬於無因次單位(dimensionless unit)。

z值運算公式

■ $z_i = \frac{x_i - \bar{x}}{s}$

◆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。

z值運算範例

範例3.52 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，請分別計算其z值。

題解：

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

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

$z_1 = \frac{x_1-\bar{x}}{S} = \frac{160-166.6}{7.80} = –0.8462$

答案：z值依序為-0.8464、-0.9747、1.4620、-0.0769和0.4360

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

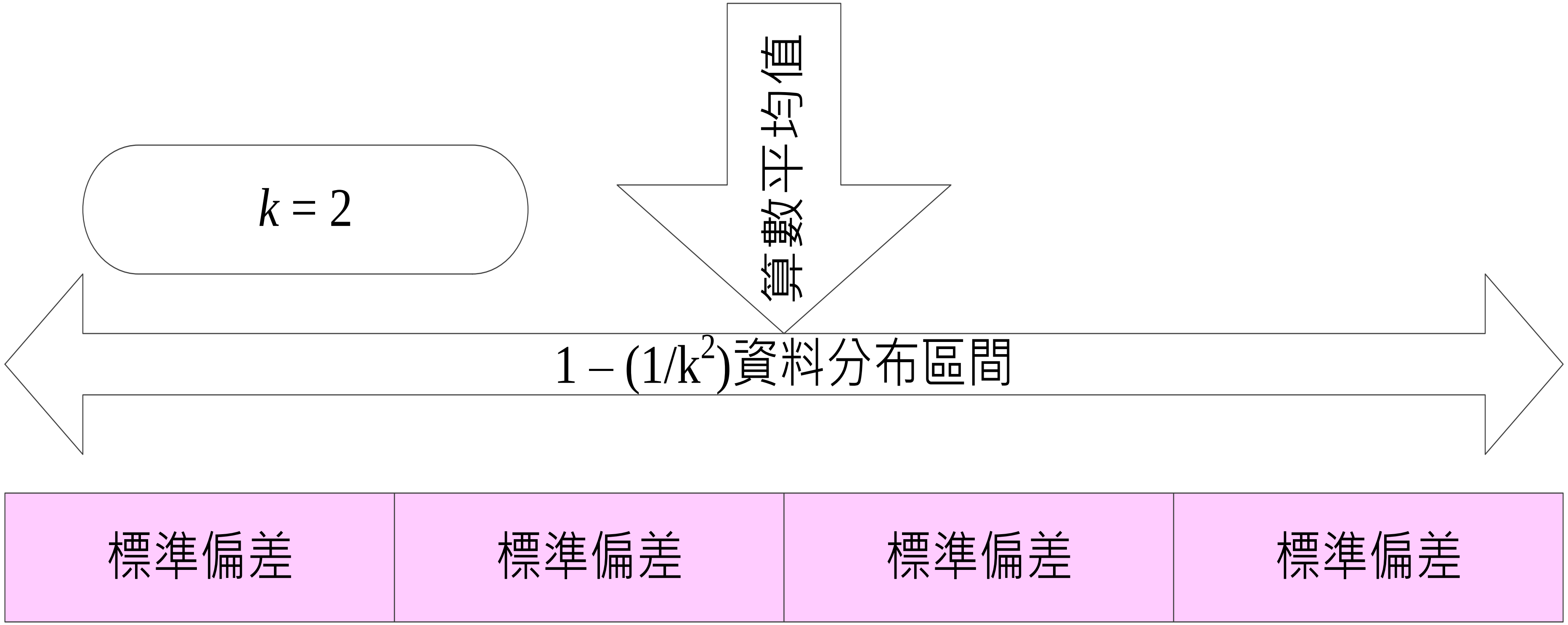
上課練習112ch3_9

3.3.2柴比氏定理

- 在任何觀測值分布資料中，至少有$(1-\frac{1}{k^2})$比率的資料，分布在算術平均值μ為中心，$\pm k$個標準差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。
  - ◆柴比氏定理推估的觀測值比率比較保守(少)。

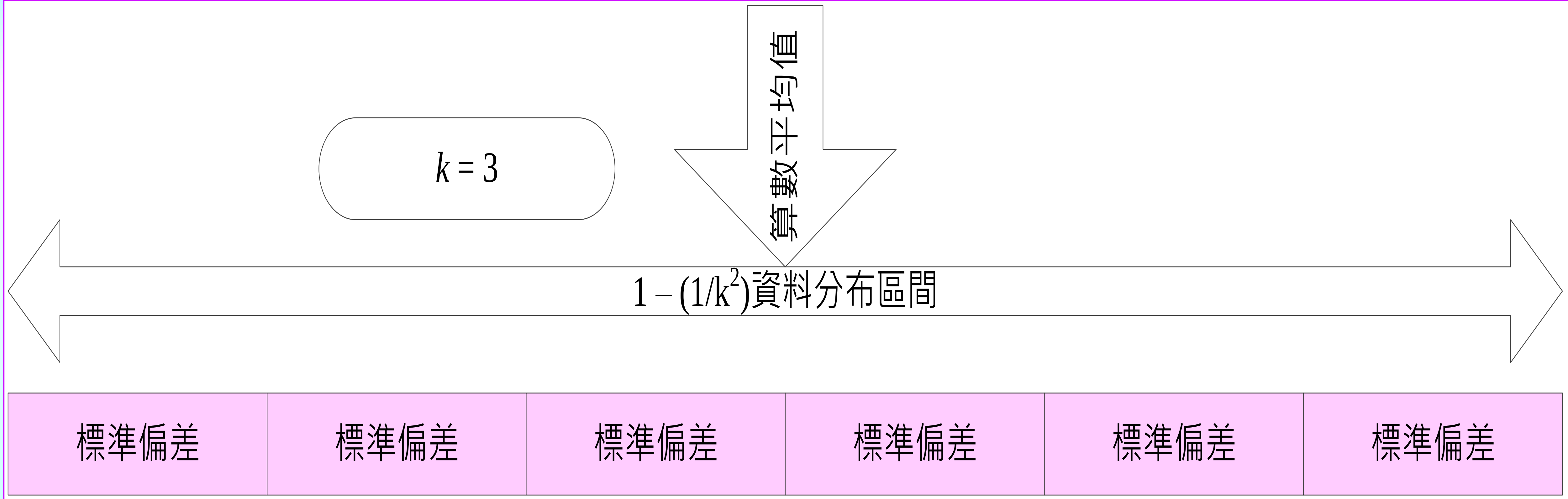
$k=2$

- $k = 2$，至少有$(1-\frac{1}{k^2}) = (1-\frac{1}{2^2}) = 0.75$比率的觀測值，介於與算術平均值為中心，±2個標準差S的範圍內。



$k = 3$

■ $k = 3$，至少有$(1 - \frac{1}{k^2}) = (1 - \frac{1}{3^2}) = 0.89$比率的觀測值，介於與算術平均值為中心，±3個標準差S的範圍內。



k值與觀測值比率

- $k = 4$，至少有$(1 - \frac{1}{k^2}) = (1 - \frac{1}{4^2}) = 0.94$比率的觀測值，介於以算術平均值為中心，$\pm 4$個標準差$S$的範圍內。
- $k = 5$，至少有$(1 - \frac{1}{k^2}) = (1 - \frac{1}{5^2}) = 0.96$比率的觀測值，介於以算術平均值為中心，$\pm 5$個標準差$S$的範圍內。
- $k = 6$，至少有$(1 - \frac{1}{k^2}) = (1 - \frac{1}{6^2}) = 0.97$比率的觀測值，介於以算術平均值為中心，$\pm 6$個標準差$S$的範圍內。

柴比氏定理推論母體分布比例

k = 2	柴比氏定理	0.75比例的觀測值
k = 3		0.89比例的觀測值
k = 4		0.94比例的觀測值
k = 5		0.96比例的觀測值
k = 6		0.97比例的觀測值

柴比氏定理計算範例

範例3.53 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？

題解：

分數介於61與69分之間，以平均分數65分為中心，前後各2個($k=\frac{平均值-分數下限}{標準差}=\frac{分數上限-平均值}{標準差}=\frac{69-65}{2}=\frac{4}{2}=2$)標準差的間距，透過柴比氏定理，推測至少有$(1-\frac{1}{k^2})=(1-\frac{1}{2^2})=0.75$比率的學生分數介於61與69分之間。

分數介於60與70分之間，以平均分數65分為中心，前後各2.5個($k=\frac{平均值-分數下限}{標準差}=\frac{分數上限-平均值}{標準差}=\frac{70-65}{2}=\frac{5}{2}=2.5$)標準差的間距，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

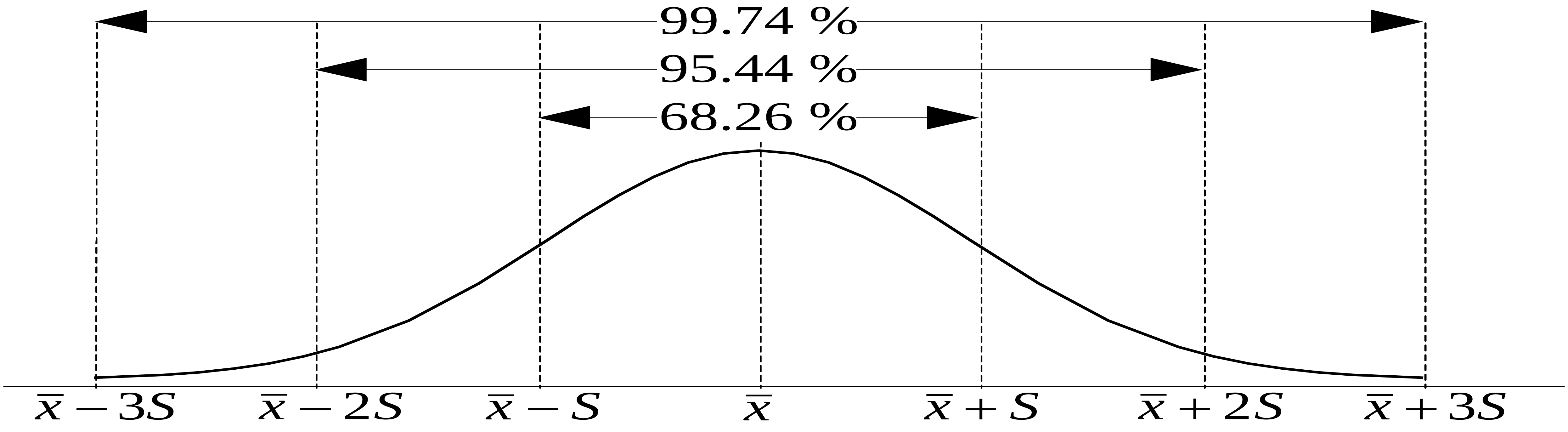
輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_10

3.3.3經驗法則

■觀測值資料分布具有鐘型的分布型態，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：



3.3.4偵測離群值

- 觀測值資料中，若有少數的觀測值是屬於極端值或異常值，此類數值稱為異常值或離群值。
  - 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)

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σ² = 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	$\frac{\sigma^2}{n}$	σ²	

樣本平均值分散程度

標準誤(差)

- 標準誤(差)(standard error, SE) = $\sigma_{\bar{x}} = \sqrt{\dfrac{\sigma^2}{n}}$
  - 標準誤(差)SE實際上是樣本平均值$\bar{x}$抽樣分布的標準差(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

- 探索性資料分析可以利用圖形瞭解資料分布情況。
  - 製作此圖形時，必須利用五數綜合(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
- 當觀測值之分布圖型產生分布偏斜傾向時，五數綜合會比算術平均值及標準差的表達效果更佳。
  - 當觀測值分布趨近於對稱狀況時，使用算術平均值及標準差表達效果為佳。

五數綜合運算範例

範例3.54 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。

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$Q_0 = 122$

第1四分位數Q_1

$\mathrm{O}(Q_1) = (i \times \frac{n-1}{4}) + 1 = (1 \times \frac{12-1}{4}) + 1 = 3.75$，$Q_1$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

$\mathrm{O}(Q_2) = (i \times \frac{n-1}{4}) + 1 = (2 \times \frac{12-1}{4}) + 1 = 6.5$，$Q_2$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

$\mathrm{O}(Q_3) = (i \times \frac{n-1}{4}) + 1 = (3 \times \frac{12-1}{4}) + 1 = 9.25$，$Q_3$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$Q_4 = 188$

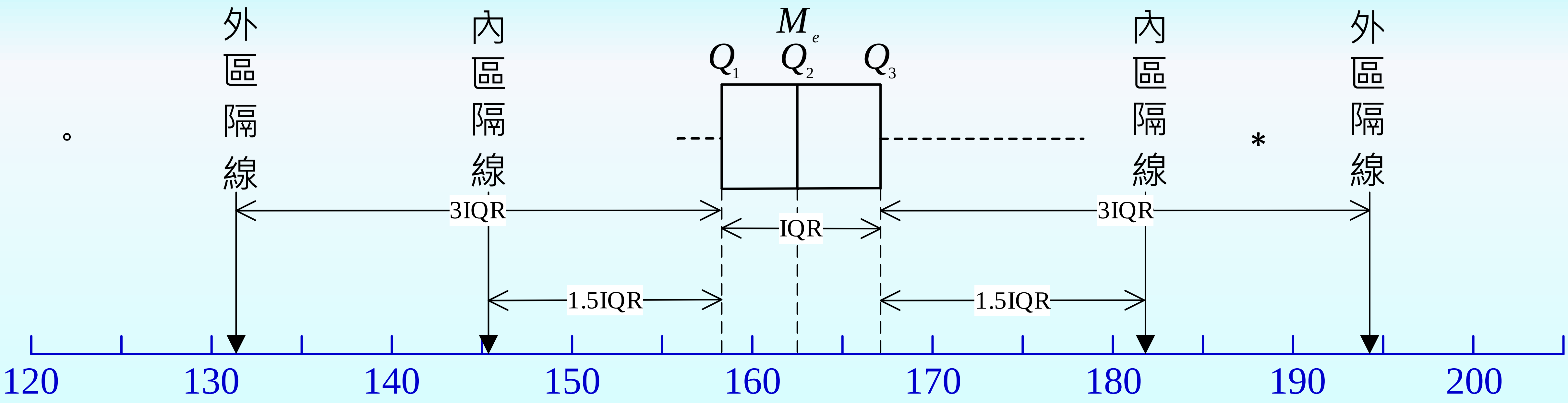
盒鬚圖

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。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。

使用前一個範例數值繪製盒形圖

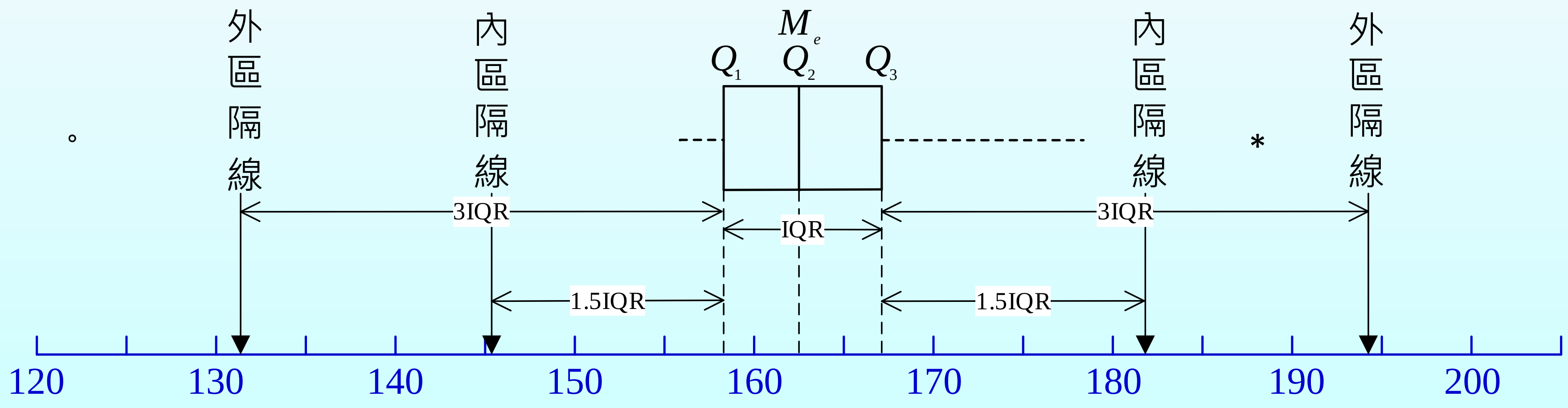


盒鬚圖：內、外區隔和鬚

c.利用四分位距IQR作為設定區隔的基本單位。

- 內區隔為Q_1以下1.5單位IQR與Q_3以上1.5單位IQR。
  - $Q_1 - 1.5IQR = 158.75 - 1.5 \times 9 = 145.25$
  - $Q_3 + 1.5IQR = 167.75 + 1.5 \times 9 = 181.25$
- 外區隔為Q_1以下3單位IQR與Q_3以上3單位IQR。
  - $Q_1 - 3IQR = 158.75 - 3 \times 9 = 131.75$
  - $Q_3 + 3IQR = 167.75 + 3 \times 9 = 194.75$

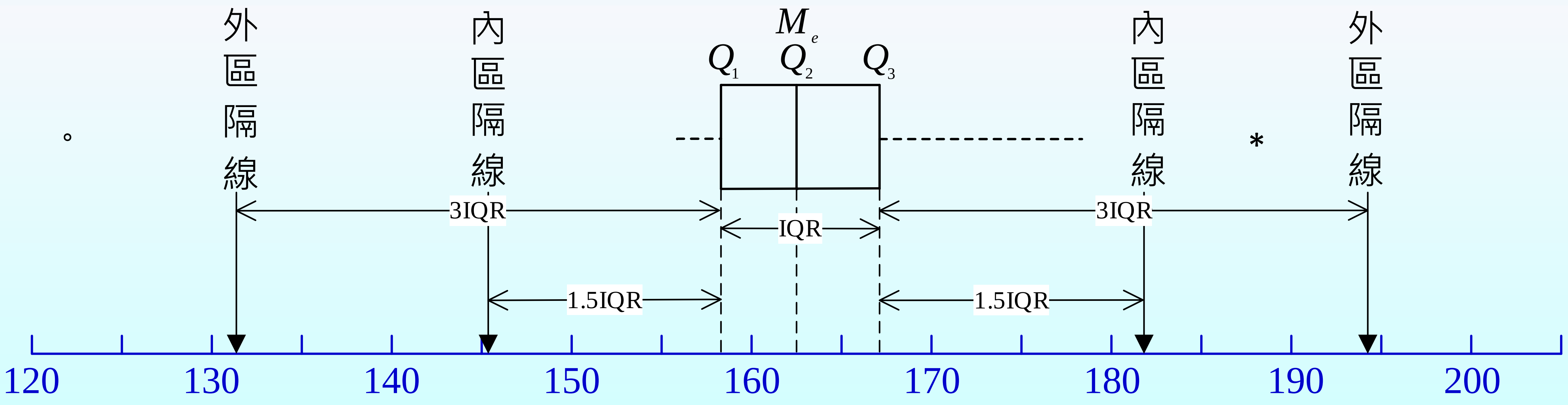
d.盒鬚圖中水平虛線為鬚whiskers，即為在內區隔區間內最大的觀測值到最小的觀測值。此例內區隔、內籬區間內最大的觀測值為178；最小觀測值為156。



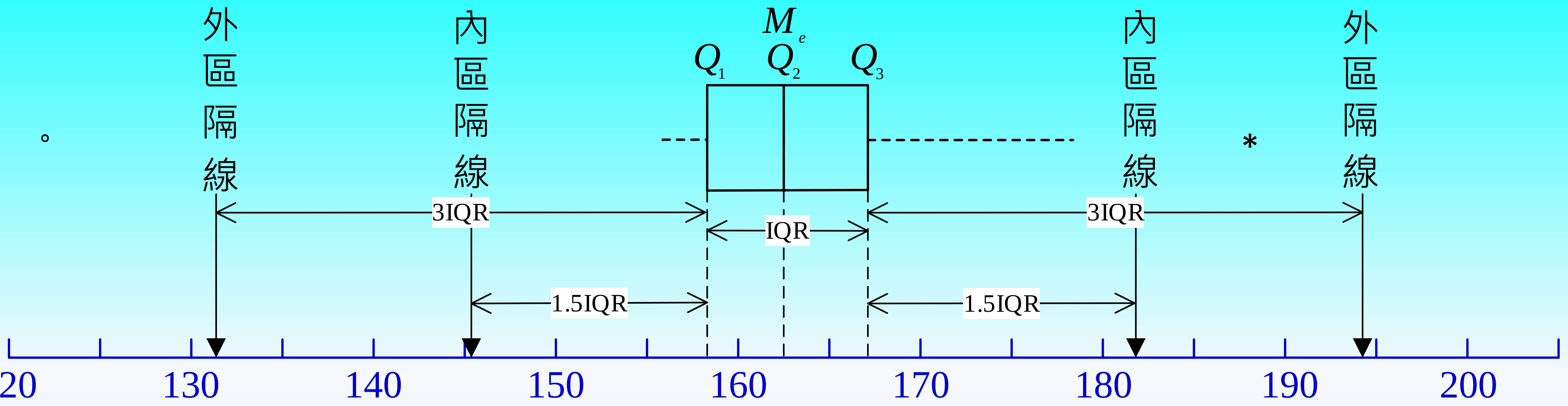
盒鬚圖：輕微和極度離群值

e. 利用內區隔和外區隔協助判定是否具有離群值(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；若有觀測值在外區隔線以外者，判定為極度離群值。
- 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：判定為輕微離群值有觀測值188；判定為極度離群值有觀測值122，已分別在盒鬚圖中標示「*」和「。」。



盒鬚圖：分布趨勢判讀



透過盒鬚圖可以簡單的瞭解資料分布的情況，以上圖為例：資料的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。

盒鬚圖不需計算母體(樣本)分布的算術平均值和標準差數值，即可瞭解資料的分布情況。

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

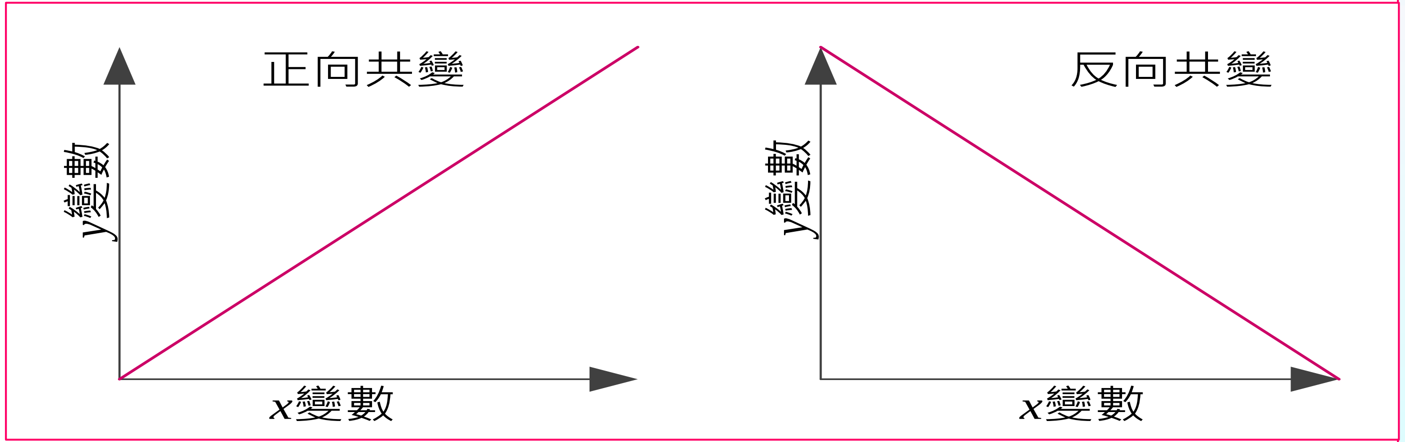
上課練習112ch3_11

3.5兩個變數相關性的評量

- ■ 一個基本單位有兩個配對隨機變數(雙變量資料)
- ■ 3.5.1共變異數(Covariance)
  - ◆受評量單位影響
- ■ 3.5.2皮爾森積差相關係數(Coefficient of correlation)
  - ◆不受評量單位影響

兩個變數之間的關係

- 兩個變數之間的關係，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。
  - 評量兩個變數相關程度，可以採用**共變(異)數**(Covariance)或**相關係數**(Coefficient of correlation)進行。



- 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、拋物線關係等。

3.5.1共變(異)數

- 母體資料中變數X和變數Y的共變(異)數以$Cov(X, Y)$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」：
  - $Cov(X, Y) = \sigma_{xy} = \frac{\sum_{i=1}^{N}[(x_i-\mu_x)\times(y_i-\mu_y)]}{N} = \frac{\sum_{i=1}^{N}[x_i\times y_i-x_i\times\mu_y-\mu_x\times y_i+\mu_x\times\mu_y]}{N} = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \frac{\sum_{i=1}^{N}x_i\times\mu_y}{N} - \frac{\sum_{i=1}^{N}\mu_x\times y_i}{N} + \frac{\sum_{i=1}^{N}\mu_x\times\mu_y}{N} = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \mu_x\times\mu_y - \mu_x\times\mu_y + \frac{N\times\mu_x\times\mu_y}{N} = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \mu_x\times\mu_y - \mu_x\times\mu_y + \mu_x\times\mu_y = \frac{\sum_{i=1}^{N}x_i\times y_i}{N} - \mu_x\times\mu_y = E(X\times Y) - E(X)\times E(Y)$
- 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」：
  - $cov(X, Y) = S_{xy} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1}$

共變(異)數可能出現的三種情境

- 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。
- 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，亦有可能存在抛物線、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。
- 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)數特性

- 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。
- 共變(異)數數值高低會受*X*與*Y*觀測值評量單位的影響。
  - 共變(異)數數值有受*X*與*Y*觀測值評量單位影響，其意涵解釋不易。

共變(異)數計算範例(1/2)

範例3.55 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)。

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

共變(異)數計算範例(2/2)

COVARIANCE.S

範例3.55 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)。

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
合計	1630	586	0	0.0	1466.0

身高平均值$\bar{x}=\frac{1630}{10}=163$ cm，體重平均值$\bar{y}=\frac{586}{10}=58.6$ 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

3.5.2相關係數

- 為了避免X與Y變數數值之「單位」差異對共變異數數值的影響，採取相關係數評量X與Y變數的相關程度。
  - 相關係數將原本共變異數之單位與標準差之單位透過分子和分母單位相互抵銷的方式，獲得此相關係數為無因次單位，故計算相關係數時皆不會受到原本X與Y變數評量單位的影響。
- 母體相關係數$\rho_{xy} = \dfrac{Cov(X,Y)}{\sigma_x \times \sigma_y} = \dfrac{\sigma_{xy}}{\sigma_x \times \sigma_y}$

母體資料相關係數

- 母體資料中變數X和變數Y的相關係數以ρ_{xy}符號表示(讀音：rho)：
- 母體相關係數 $\rho_{xy} = \dfrac{Cov(X,Y)}{\sigma_x \times \sigma_y} = \dfrac{\sigma_{xy}}{\sigma_x \times \sigma_y} = \dfrac{\frac{\sum_{i=1}^{N}(x_i-\mu_x)\times(y_i-\mu_y)}{N}}{\sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu_x)^2}{N}}\times\sqrt{\frac{\sum_{i=1}^{N}(y_i-\mu_y)^2}{N}}} =$

$$\frac{\sum_{i=1}^{N}(x_i-\mu_x)\times(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i-\mu_x)^2}\times\sqrt{\sum_{i=1}^{N}(y_i-\mu_y)^2}} = \frac{\sum_{i=1}^{N} x_i\times y_i-\frac{\sum_i^N x_i\times\sum_i^N y_i}{N}}{\sqrt{\sum_{i=1}^{N} x_i^2-\frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}\times\sqrt{\sum_{i=1}^{N} y_i^2-\frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N}}} = \frac{E(X\times Y)-E(X)\times E(Y)}{\sqrt{V(X)}\times\sqrt{V(Y)}}$$

  - $Cov(X, Y) = \sigma_{xy}$：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的標準差

樣本資料相關係數

- ■ 樣本資料中變數X和變數Y的相關係數以r_{xy}或γ_{xy}符號表示：
- ■ 樣本相關係數r_{xy} $= \dfrac{cov(X,Y)}{S_x \times S_y} = \dfrac{S_{xy}}{S_x \times S_y} = \dfrac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\times\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}} =$

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

- ◆$cov(X,Y) = S_{xy}$：樣本資料共變異數
- ◆S_x和S_y：分別為變數X和變數Y的樣本資料標準差

相關係數特性

- 相關係數的數值分布介於–1與1之間。
  - 相關係數只解釋兩變數之間，線性關係的強度。
  - 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變數順序前後對調，其相關係數數值不會改變。
  - 相關係數會受到極端值(離群值)的影響。
- 相關係數 = 1，代表x_i與y_i變數之間的呈現正向完全直線之關係，每一對x_i與y_i變數皆落在斜率為正的直線上。x_i與y_i變數之間的變動方向一致，當x_i變數增加時，y_i變數亦同時增加。
- 相關係數 = –1，代表x_i與y_i變數之間的呈現反向完全直線之關係，每一對x_i與y_i變數皆落在斜率為負的直線上。x_i與y_i變數之間的變動方向相反，當x_i變數增加時，y_i變數同時減少。
- 相關係數 = 0，代表x_i與y_i變數之間無直線關係，可能是x_i與y_i變數之間無關係、x_i與y_i變數之間具有拋物線關係、圓形關係、非線性關係。

相關係數運算範例(1/2)

範例3.56 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。

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

相關係數運算範例(2/2)

範例3.56 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。

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
合計	1630	586	268346	35246	96984	0	2656	0.0	906.40	1466.0

身高平均值$\bar{x} = \frac{1630}{10} = 163$　　體重平均值$\bar{y} = \frac{586}{10} = 58.6$

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

$$身高S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{2656}{10-1}} = \sqrt{295.11} = 17.18$$

$$體重S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}} = \sqrt{\frac{906.4}{10-1}} = \sqrt{100.71} = 10.04$$

$$相關係數r_{xy} = \frac{cov(X,Y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{162.89}{17.18\times10.04} = 0.9448$$

Excel操作：資料分析

線上討論：操作問題

點選檔案下面選項

出現Excel選項視窗，點選增益集，選擇【分析工具箱】，按執行

在【資料】點選【分析工具】

在分析工具視窗中點選【相關係數】後按【確定】

在輸入範圍選擇欲計算相關係數的資料區間

確定

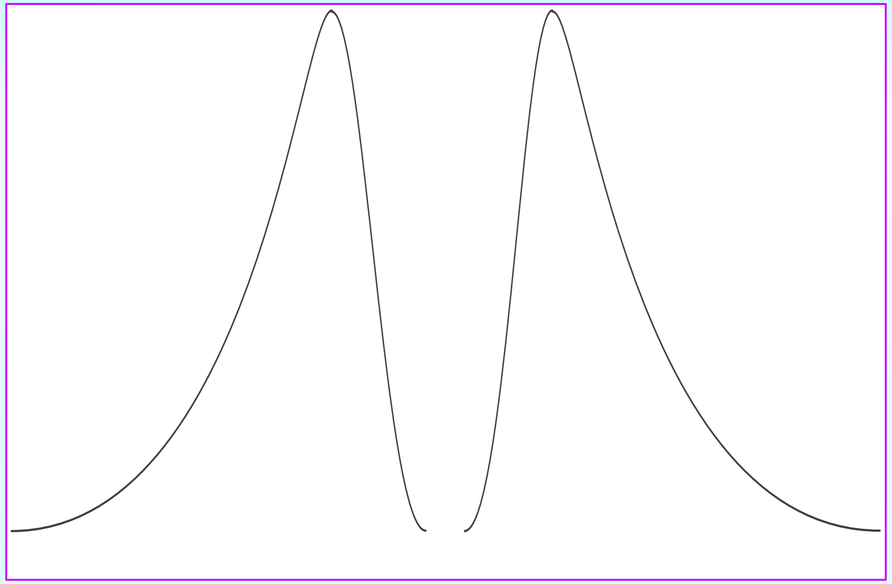
輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

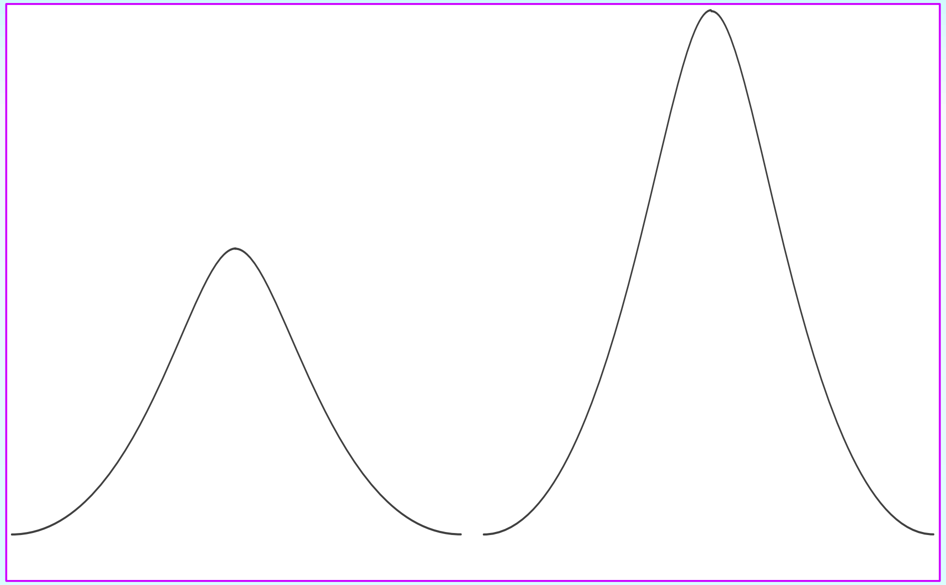
上課練習112ch3_12

3.6偏度與峰度

3.6.1偏度 (Skewness)

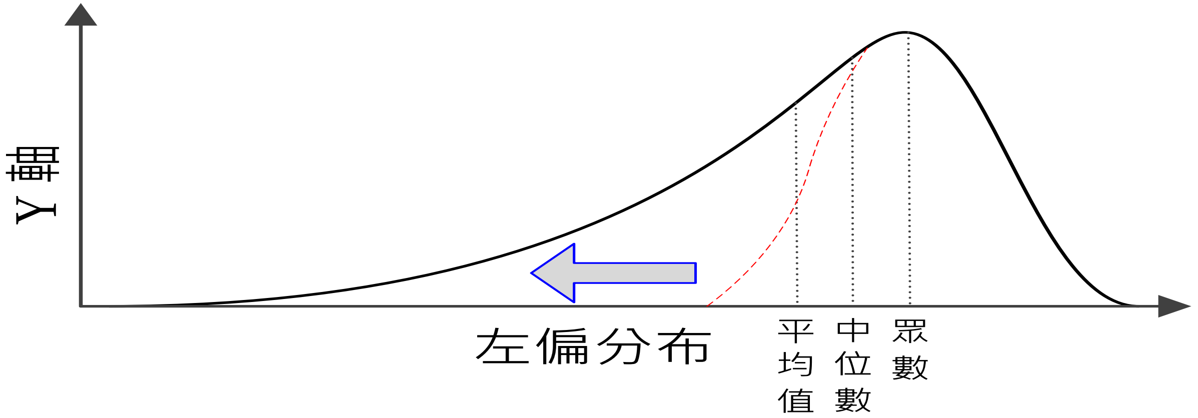


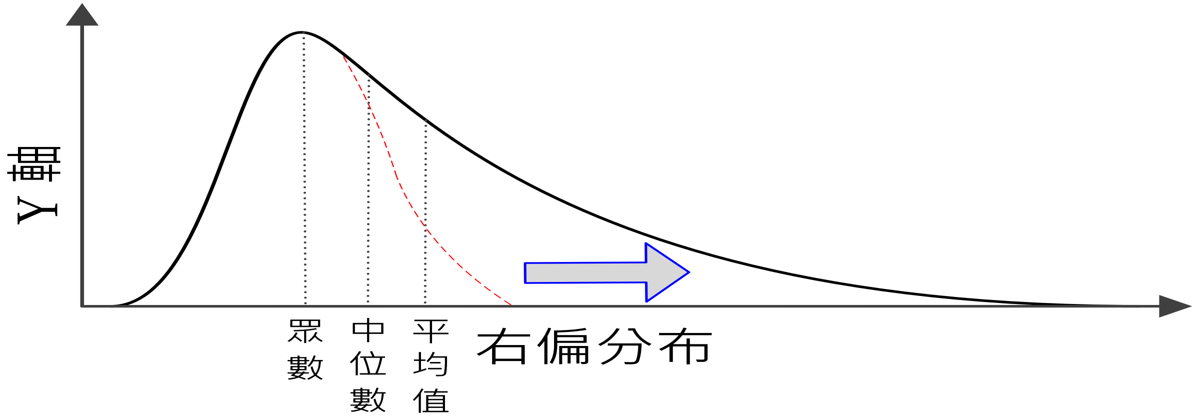
3.6.2峰度 (Kurtosis)



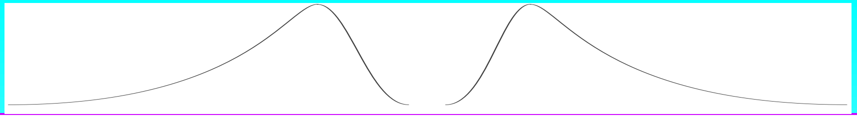
3.6.1偏度

- ■ 評量觀測資料數值分布對稱性時，資料分布可能是右偏或左偏。
  - ◆偏度或偏態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。
- ■ 偏度評量方法
  - ◆Bowley偏態係數SK_B
  - ◆皮爾森(Karl Pearson)偏態係數SK_p
  - ◆動差法計算偏態係數α_3
  - ◆Excel軟體偏態係數的計算方式





Bowley偏態係數SK_B



$$SK_B=\frac{(Q_3-M_e)-(M_e-Q_1)}{(Q_3-M_e)+(M_e-Q_1)}=\frac{(Q_3-M_e)-(M_e-Q_1)}{Q_3-Q_1}=\frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1}=\frac{Q_3-2\times Q_2+Q_1}{Q_3-Q_1}=\frac{Q_1+Q_3-2\times Q_2}{Q_3-Q_1}$$

若$Q_3=M_e=Q_2$，即$SK_B=\frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1}=\frac{(Q_2-Q_2)-(Q_2-Q_1)}{Q_2-Q_1}=\frac{\cancel{(Q_2-Q_2)}-(Q_2-Q_1)}{Q_2-Q_1}=\frac{-(Q_2-Q_1)}{Q_2-Q_1}=-1$，此為最小值；$Q_1=M_e=Q_2$，即$SK_B=\frac{(Q_3-Q_2)-(Q_2-Q_1)}{Q_3-Q_1}=\frac{(Q_3-Q_1)-(Q_1-Q_1)}{Q_3-Q_1}=\frac{(Q_3-Q_1)-\cancel{(Q_1-Q_1)}}{Q_3-Q_1}=\frac{(Q_3-Q_1)}{Q_3-Q_1}=1$，此為最大值。因此偏態係數$SK_B$會介於$-1\sim1$，$-1<SK_B<1$。

$|SK_B|\fallingdotseq 0.0$，偏斜程度屬於「沒有」。

一般$|SK_B|\fallingdotseq 0.1$，偏斜程度屬於「輕微」。

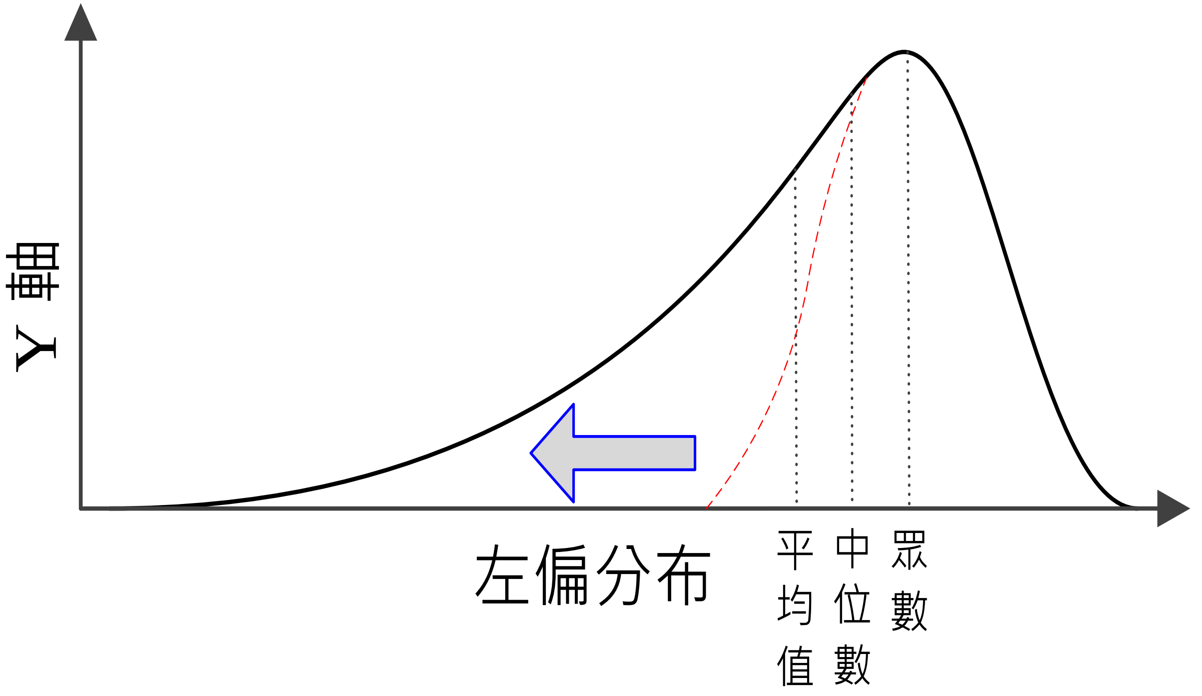
一般$|SK_B|>0.3$，偏斜程度屬於「嚴重」。

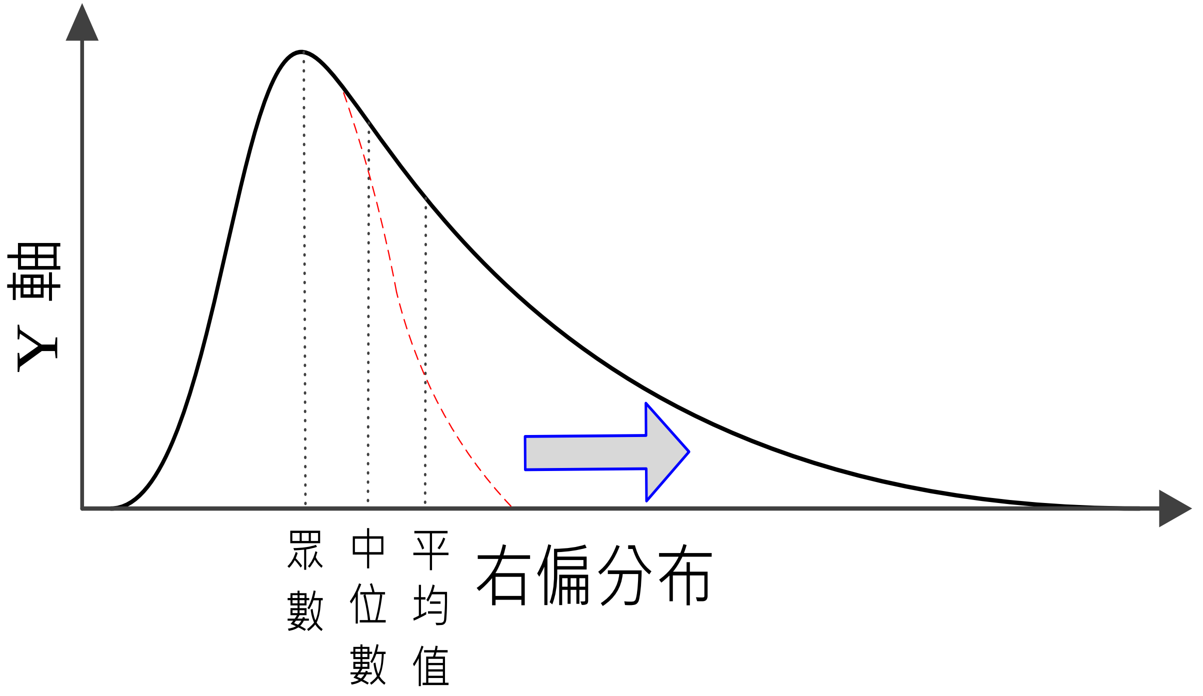
右偏、對稱和左偏

若$Q_3 - M_e > M_e - Q_1$，致$SK_B > 0$，屬於右偏分布、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、負偏分布(negatively skewed)。





皮爾森(Karl Pearson)偏態係數SK_p

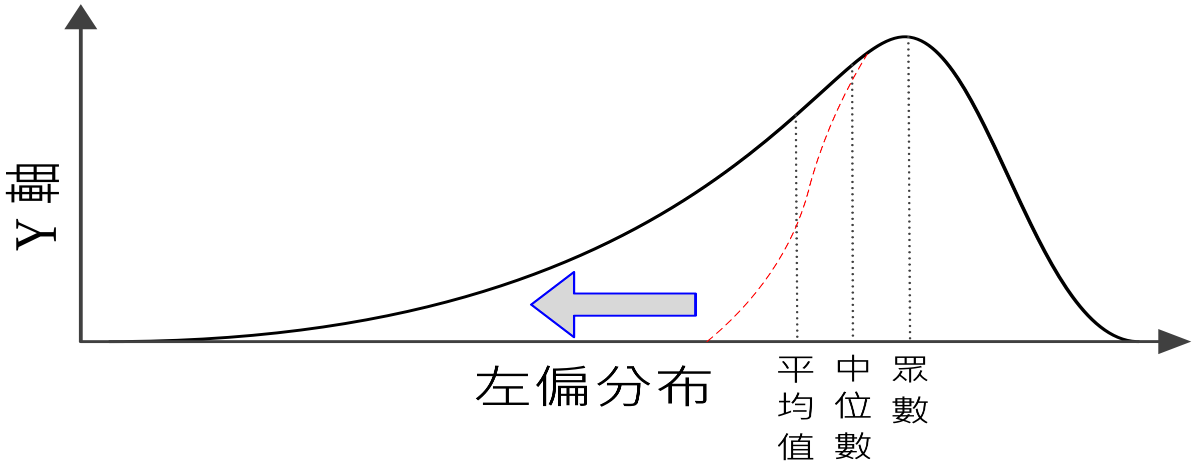
母體資料：$SK_p = \frac{3\times(\mu - M_e)}{\sigma}$，其中$M_e$：母體中位數

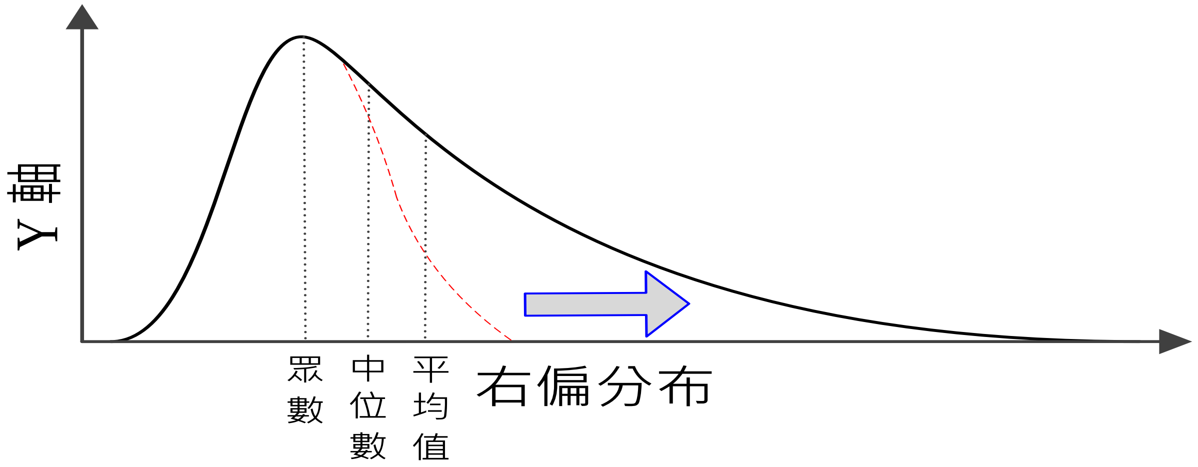
樣本資料：$SK_p = \frac{3\times(\bar{x} - m_e)}{s}$，其中$m_e$：樣本中位數

若算術平均值(mean) μ > 中位數(median) M_e，致$SK_p > 0$，屬於右偏分布(skewed right distribution)、正偏分布。

若算術平均值(mean) μ = 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) μ < 中位數(median) M_e，致$SK_p < 0$，屬於左偏分布(skewed left distribution)、負偏分布。



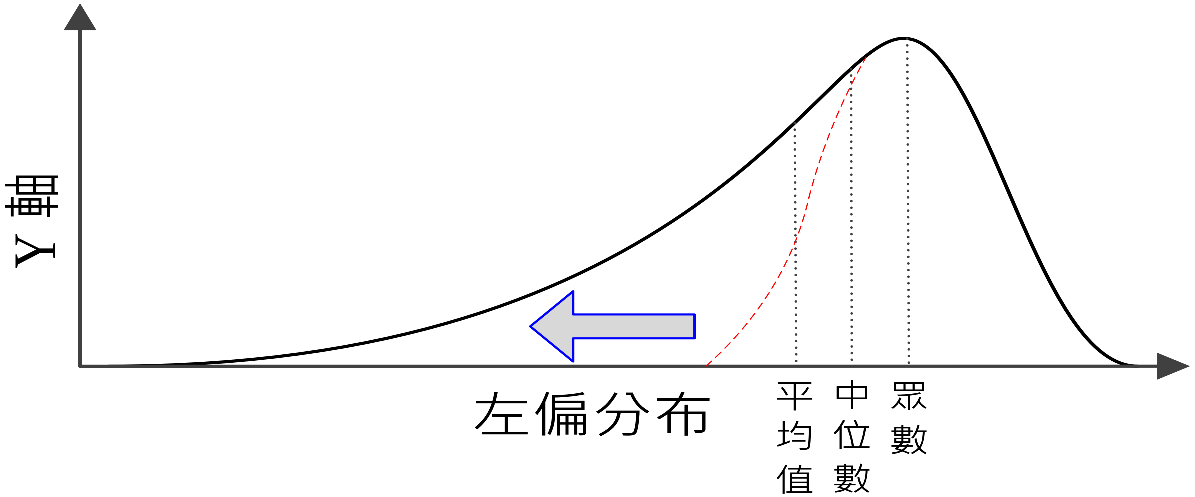


動差法計算偏態係數α_3

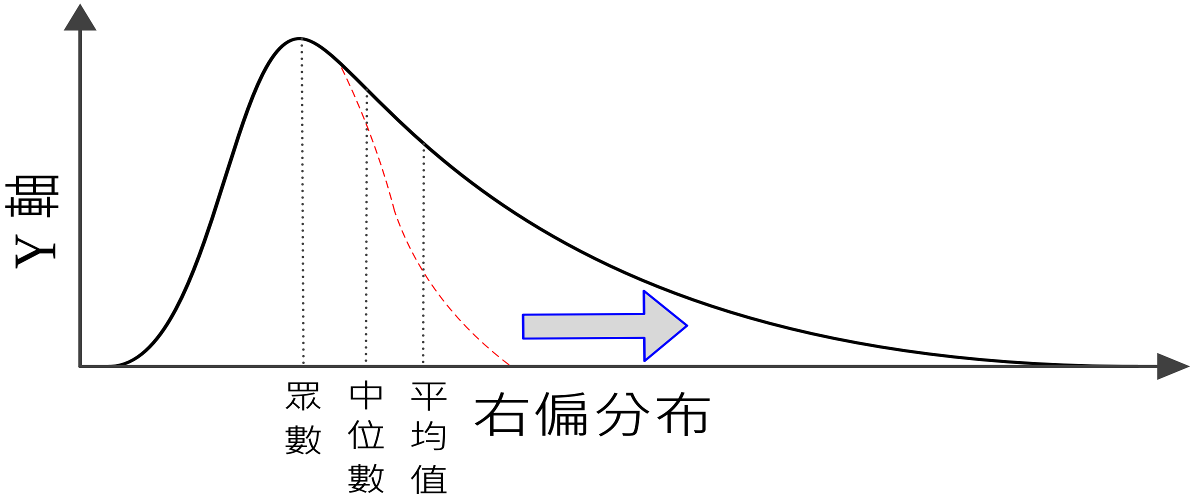
母體資料：$\alpha_3 = \dfrac{\dfrac{\sum_{i=1}^{N}(x_i-\mu)^3}{N}}{\left[\sqrt{\dfrac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}\right]^3}$

樣本資料：$\alpha_3 = \dfrac{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{n}}{\left[\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\right]^3}$

$\alpha_3 = 0$，資料對稱分布
$\alpha_3 > 0$，資料右偏分布
$\alpha_3 < 0$，資料左偏分布



左偏分布



右偏分布

Excel軟體偏態係數計算方式：SKEW

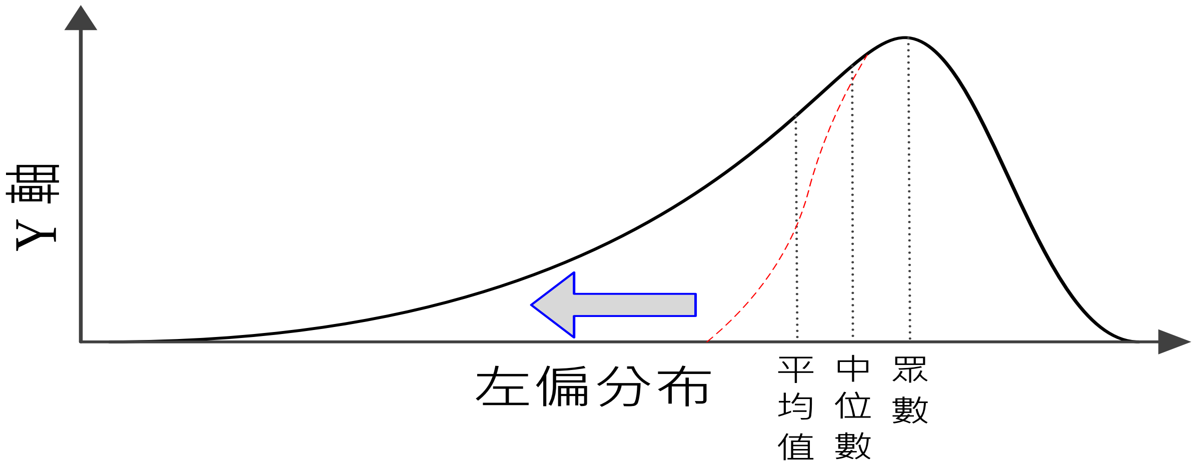
$$SK_p = \frac{n}{(n-1)\times(n-2)} \times \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^3$$

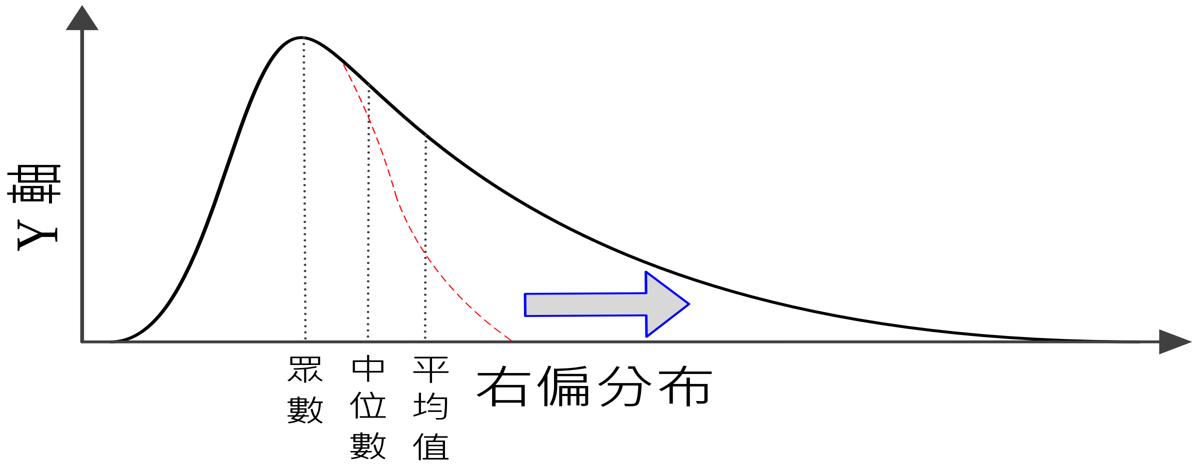
$SK_p = 0$為對稱分布(symmetric distribution)
算術平均值(mean) μ = 中位數(median) M_e = 眾數(mode) M_o

$SK_p > 0$為右偏分布(skewed to the right)、正偏分布
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)、負偏分布
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。





Excel函數：SKEW

線上討論：操作問題

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):SKEW

確定

選擇欲計算偏態的資料區間

確定

偏態係數運算範例

範例3.60 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。

254	317	324	331	325	425	185	265	325	215	198	260

題解：第1四分位數$Q_1 = 244.25$，$M_e = Q_2 = 291$，$Q_3 = 325$，樣本平均值$\bar{x} = 285.3$，標準差SD = 68.7

$$SK_B = \frac{(Q_3-M_e)-(M_e-Q_1)}{Q_3-Q_1} = \frac{(325-291)-(291-244.25)}{325-244.25} = \frac{-12.75}{80.75} = -0.1579$$

$$SK_p = \frac{3\times(\bar{x}-m_e)}{s} = \frac{3\times(285.3-291)}{68.7} = -0.2475$$

答案：$SK_B = -0.1579$；$SK_p = -0.2475$

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_13

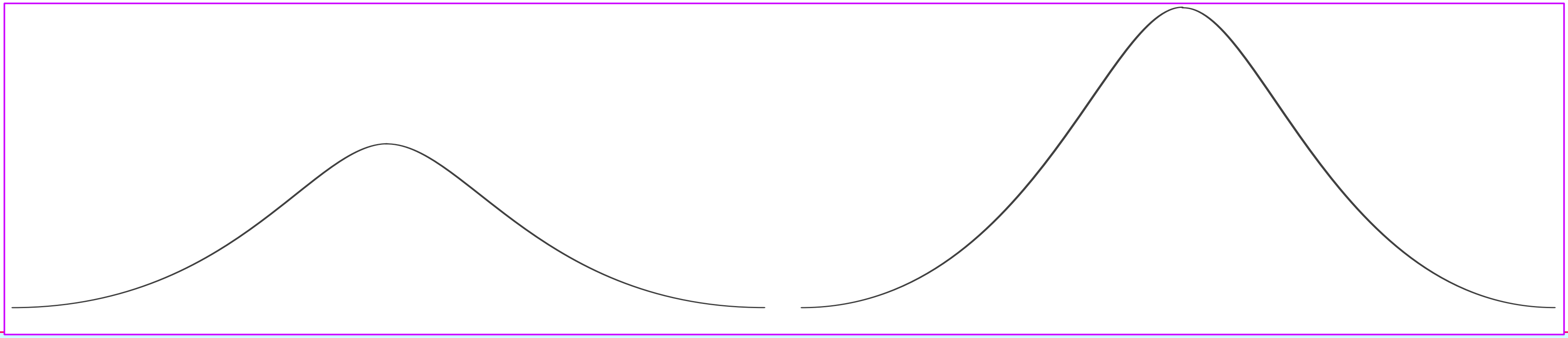
3.6.2峰度

■ 當次數分布有集中的現象時，會出出現波峰(peak)，波峰可為單峰、雙峰與多峰型態。

◆單峰情況，依據平坦陡峭的分布程度可分為高峽峰、低闊峰或常態峰，峰度高低以峰度係數進行評量。

◆峰態係數或峰度係數之度量一般以β_2、K、β_k或γ_2 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。

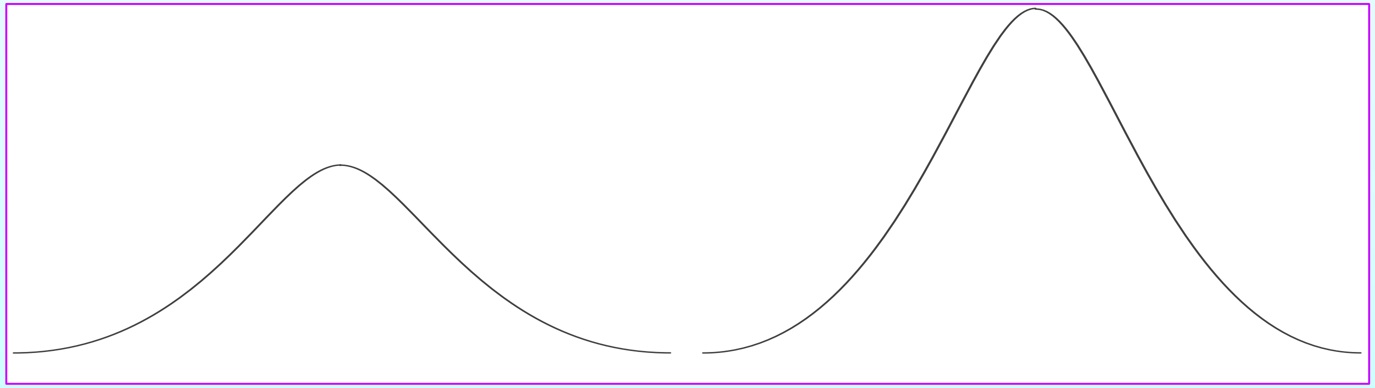


觀測值相同，出現頻率：運算方式

- 母體資料有N個觀測值，依據觀測值數值相等性區分為k組，第i組內有f_i個觀測值，組內觀測值皆相等
  - 母體觀測值數量$N=\sum_{i=1}^{k} f_i$
  - 平均值$\mu=\frac{\sum_{i=1}^{k} x_i \times f_i}{N}=\frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i}$
  - 標準差$\sigma=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{N}}=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\mu)^2}{\sum_i^k f_i}}$
- 樣本資料有n個觀測值，依據觀測值數值相等性區分為k組，第i組內有f_i個觀測值，組內觀測值皆相等
  - 樣本觀測值數量$n=\sum_{i=1}^{k} f_i$
  - 平均值$\bar{x}=\frac{\sum_{i=1}^{k} x_i \times f_i}{n}=\frac{\sum_{i=1}^{k} x_i \times f_i}{\sum_i^k f_i}$
  - 標準差$S=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{\sum_{i=1}^{k} f_i \times (x_i-\bar{x})^2}{\sum_i^k f_i-1}}$。

四級動差法計算峰度係數

	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軟體峰度係數計算方式：KURT

$$\gamma_2 = \left[\frac{n\times(n+1)}{(n-1)\times(n-2)\times(n-3)} \times \sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^4\right] - \frac{3\times(n-1)^2}{(n-2)\times(n-3)}$$

常態峰(mesokurtosis)

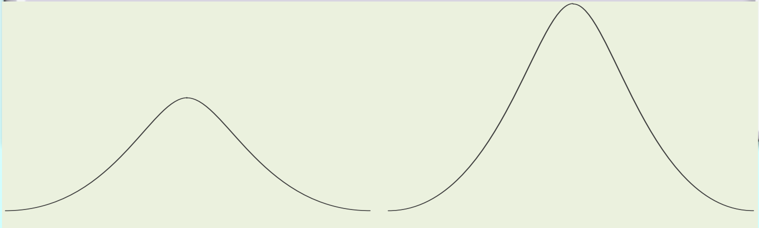
- 表示資料的分布呈現一般的型態
- 峰度係數$\beta_2 = 3$；$\gamma_2 = 0$

高峽峰(leptokurtosis)

- 表示資料分布過度集中於平均數或眾數附近
- 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$\beta_2 < 3$；$\gamma_2 < 0$



Excel函數：KURT

線上討論：操作問題

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):KURT

確定

選擇欲計算峰度的資料區間

確定

峰度係數運算範例

範例3.61 計算下列母體資料四級動差法計算峰度係數

x_i	f_i	$x_i \times f_i$	$x_i - \mu$	$(x_i - \mu)^2$	$f_i \times (x_i - \mu)^2$	$(x_i - \mu)^4$	$f_i \times (x_i - \mu)^4$
5	5	25	-25.42	646.01	3230.03	417325.0	2086624.9
6	1	6	-24.42	596.17	596.17	355423.0	355423.0
8	2	16	-22.42	502.51	1005.01	252513.2	505026.5
65	1	65	34.58	1196.01	1196.01	1430432.6	1430432.6
77	1	77	46.58	2170.01	2170.01	4708930.1	4708930.1
88	2	176	57.58	3315.84	6631.68	10994796.7	21989593.5
合計	12	365			14828.92		31076030.5

題解：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$\mu = \dfrac{\sum_{i=1}^{k} x_i \times f_i}{N} = \dfrac{\sum_{i=1}^{k} x_i \times f_i}{\sum_{i}^{k} f_i} = \dfrac{365}{12} = 30.42$

母體標準差$\sigma = \sqrt{\dfrac{\sum_{i=1}^{k} f_i \times (x_i - \mu)^2}{\sum_{i}^{k} f_i}} = \sqrt{\dfrac{14828.92}{12}} = \sqrt{1235.74} = 35.15$

$$\beta_2 = \frac{\frac{1}{N} \times \sum_{i=1}^{k} [f_i \times (x_i - \mu)^4]}{\sigma^4} = \frac{\frac{1}{12} \times 31076030.5}{35.15^4} = \frac{2589669.2}{1527061} = 1.696$$

$\gamma_2 = \beta_2 - 3 = 1.696 - 3 = -1.304$

答案：$\beta_2 = 1.696$；$\gamma_2 = -1.304$

輕鬆練習一下

此投影片播放後3分鐘內，沒有進入該練習卷者，老師即關閉沒有進入者的練習卷(本練習5 分鐘完成)

上課練習112ch3_14

先使用『來賓練習』熟習課程內容

正式平常考每一題留下運算檔案

匡列足夠時間，一次登入，一口氣完成，一次繳交

關閉時間：D日1200

平常考112ch3

您辛苦了

第三章課程結束

聆聽您的想法與建議