

計算機使用方法：for CASIO fx-350ES

基本統計功能：Mode 2 1 Shift 1 3 2 (reset data)

x = (input data) when finish AC

Shift 1 3 2 =：平方總和 Sum of Squares

Shift 1 3 2 =：總和 Sum

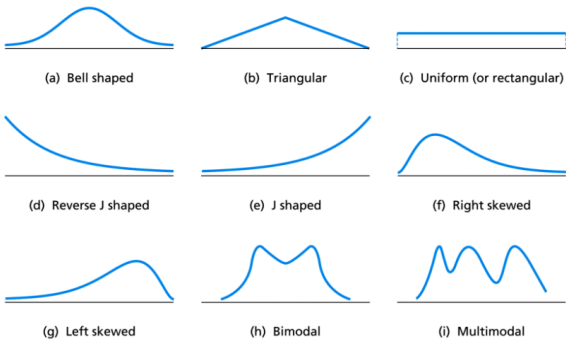
Shift 1 4 2 =：算數平均 Arithmetic Mean

Shift 1 4 3 =：標準差 Standard Deviation

Shift 1 4 4 =：樣本標準差 Sample Standard Deviation

小數分數轉換：Shift Mode 1/2

Common distribution shapes



一、中央趨勢

- 母體平均數 $\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{x_1 + x_2 + \dots + x_n}{N}$
- 樣本平均數 $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$
- 中位數:
 $\tilde{x} = x_{[\frac{n+1}{2}]}$
➢ 若 n 為奇數
➢ 若 n 為偶數 $\tilde{x} = \frac{x_{[\frac{n}{2}]} + x_{[\frac{n+1}{2}]}}{2}$
- 眾數 mode: 出現最多次數的數
- 百分位數: $i = \frac{P}{100} \times n$, 其中 n 為樣本個數,
➢ 若 i 為整數 $\frac{(x_i + x_{i+1})}{2}$ 為第 P 百分位數
➢ 若 i 不為整數, 將 i 無條件進位, x_i 為第 P 百分位數
- 四分位數: 第 1 四分位數為 Q_1 為第 25 百分位數, 第 2 四分位數為 Q_2 為第 50 百分位數, 即中位數, 第 3 四分位數為 Q_3 為第 75 百分位數

二、分散度衡量

- 全距: $R = x_n - x_1$
- 四分位距: $IQR = Q_3 - Q_1$
- 母體變異數: $\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$
- 樣本變異數: $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2}{n-1} - \frac{(\sum_{i=1}^n x_i)^2}{n(n-1)} = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$
- 母體標準差: $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$
- 樣本標準差: $s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n-1} - \frac{(\sum_{i=1}^n x_i)^2}{n(n-1)}} = \sqrt{\frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}}$

名詞解釋：

mean：平均值

InterQuartile Range: $Q_3 - Q_1$ ($Q_1 = n/4$, $Q_2 = n/2$, $Q_3 = n*3/4$)

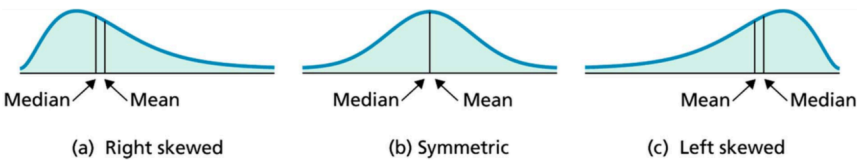
box plot：盒狀圖/盒鬚圖

>> 籬笆 ($Q_1 - 1.5 * IQR$, min, Q_1 , Q_2 , Q_3 , Max, $Q_3 + 1.5 * IQR$)

>> 若有觀測值落在盒鬚圖的籬笆外，則會將之視為離群值。

histogram：柱狀圖

mode：眾數



Conditional Probability	Multiplication Rules
$P(B A) = \frac{P(A \text{ and } B)}{P(A)}$	$P(A \text{ and } B) = P(A) \cdot P(B)$ Independent
	$P(A \text{ and } B) = P(A) \cdot P(B A)$ Dependent

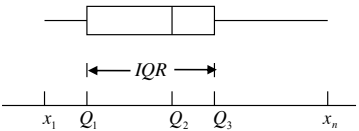
- 母體變異係數: $CV = \frac{\sigma}{\mu} \times 100\%$
- 樣本變異係數: $CV = \frac{S}{\bar{x}} \times 100\%$

三、平均數與標準差之應用

- 母體 Z 分數: $z_i = \frac{x_i - \mu}{\sigma}$
- 樣本 Z 分數: $z_i = \frac{x_i - \bar{x}}{S}$
- 柴比雪夫定理: 當 $k \geq 1$ 時, 任一組資料至少有 $1 - \frac{1}{k^2}$ 比例的觀測值會落在距離平均數 k 個標準之內
- 經驗法則: 若資料呈鐘形分佈, 則
➢ 接近 68% 的資料會落在距離平均數 1 個標準差之內
➢ 接近 95% 的資料會落在距離平均數 2 個標準差之內
➢ 幾乎所有的資料會落在距離平均數 3 個標準差之內

四、離群值之判定:

- Z 分數之判定法: 若資料值 x_i 的 Z 分數為 z_i , 當 $|z_i| > 3$, 則 x_i 可視為離群值
- 盒形圖判定法: 將資料的最小值 x_1 、 Q_1 、 Q_2 、 Q_3 及最大值 x_n 以盒形及直線之方式表示



- 若 Q_1 至 Q_2 的距離比 Q_2 至 Q_3 的距離長, 資料偏右
- 若 Q_1 至 Q_2 的距離比 Q_2 至 Q_3 的距離短, 資料偏左
- $x_i - Q_3 > 1.5 IQR$ 或 $Q_1 - x_i > 1.5 IQR$, x_i 為懷疑之離群值
- $x_i - Q_3 > 3 IQR$ 或 $Q_1 - x_i > 3 IQR$, x_i 為認定之離群值