



統計學（一）

第二章 敘述統計 (Descriptive Statistics)

授課教師：唐麗英教授

國立交通大學
工業工程與管理學系

聯絡電話：(03)5731896

e-mail：litong@cc.nctu.edu.tw

2013

☆ 本講義未經同意請勿自行翻印 ☆

本課程內容參考書目

● 教科書

Mendenhall, W., & Sincich, T. (2007). *Statistics for engineering and the sciences*, 5th Edition. Prentice Hall.

● 本課程內容參考書目

1. Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2009). *Basic business statistics: Concepts and applications*, 11th Edition. Upper Saddle River, N.J: Pearson Prentice Hall.
2. Larson, H. J. (1982). *Introduction to probability theory and statistical inference*, 3rd Edition. New York: Wiley.
3. Miller, I., Freund, J. E., & Johnson, R. A. (2000). *Miller and Freund's Probability and statistics for engineers*, 6th Edition. Upper Saddle River, NJ: Prentice Hall.
4. Montgomery, D. C., & Runger, G. C. (2011). *Applied statistics and probability for engineers*, 5th Edition. Hoboken, NJ: Wiley.
5. Watson, C. J. (1997). *Statistics for management and economics*, 5th Edition. Englewood Cliffs, N.J: Prentice Hall.
6. 林惠玲、陳正倉（2009），「統計學：方法與應用」，第四版，雙葉書廊有限公司。
7. 唐麗英、王春和（2013），「從範例學 MINITAB 統計分析與應用」，博碩文化公司。
8. 唐麗英、王春和（2008），「SPSS 統計分析 14.0 中文版」，儒林圖書公司。
9. 唐麗英、王春和（2007），「Excel 2007 統計分析」，第二版，儒林圖書公司。
10. 唐麗英、王春和（2005），「STATISTICA6.0 與基礎統計分析」，儒林圖書公司。
11. 陳順宇（2004），「統計學」，第四版，華泰書局。
12. 彭昭英、唐麗英（2010），「SAS123」，第七版，儒林圖書公司。

第二章 敘述統計 (Descriptive Statistics)

● 第一單元：如何以圖形來展示資料 (Graphs)

一、**定性**資料如何以圖來表示？

— 利用**條圖** (Bar Graph)，**柏拉圖** (Pareto Diagram)，**單圓圖** (Pie Chart)。

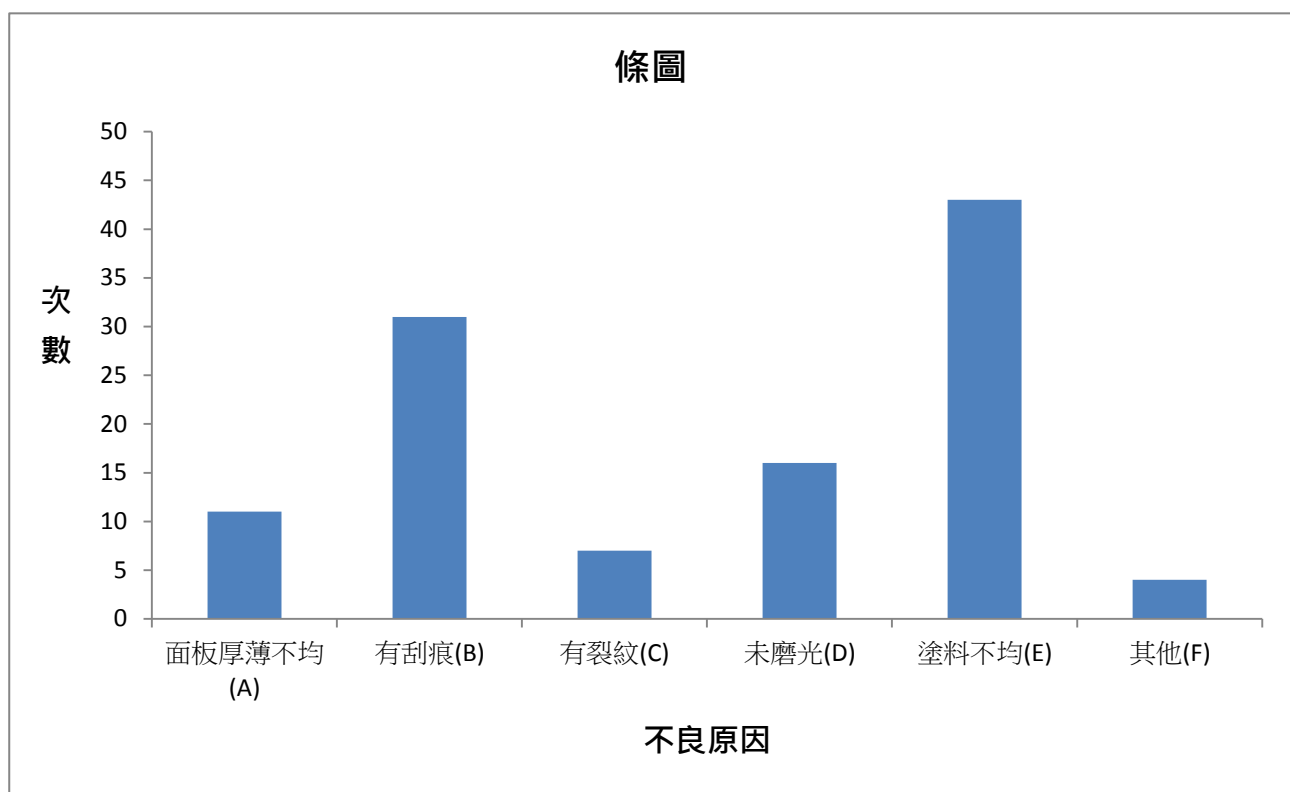
1.**條圖**：條圖是用來比較及對照不同時期或類別間的差異。

作法：

- 1) 水平軸---種類，對每一種類繪入直立條棒。
- 2) 垂直軸---觀察值的次數，相對次數或百分比。
- 3) 每一條棒有相同寬度。
- 4) 條棒彼此之間不相連接。

例 1：某製造 LCD 面板的公司欲改善面板的品質，於是由該公司的品管人員收集了某天產量中不良品的資料，並列出造成不良品之原因的次數資料，如下表所示，請依數據繪製條圖。

不良原因	改善前
面板厚薄不均(A)	11
有刮痕(B)	31
有裂紋(C)	7
未磨光(D)	16
塗料不均(E)	43
其他(F)	4
總和	112



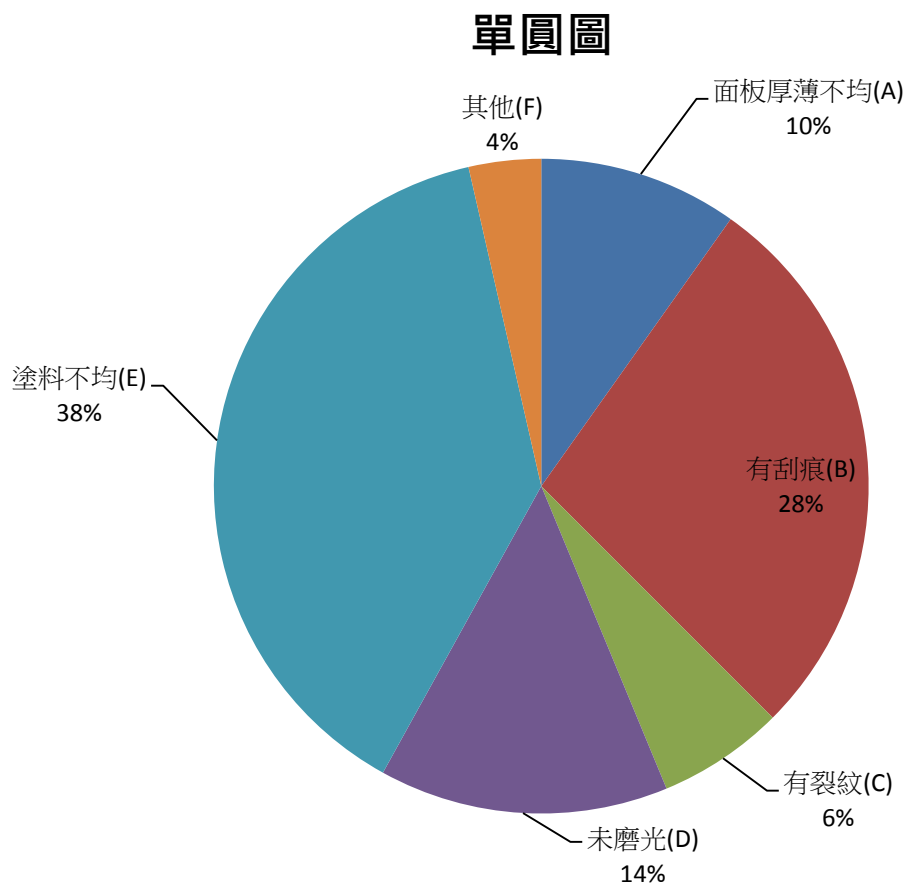
2.單圓圖：單圓圖是用來顯示一個單一總合量如何攤分於各種類別中。

作法：

1)計算出各類別所佔百分比。

2)各類別在單圓圖中所占之角度為該類別在總量中所占百分比 $\times 360^\circ$ 。

例 2：承例 1，畫出單圓圖。

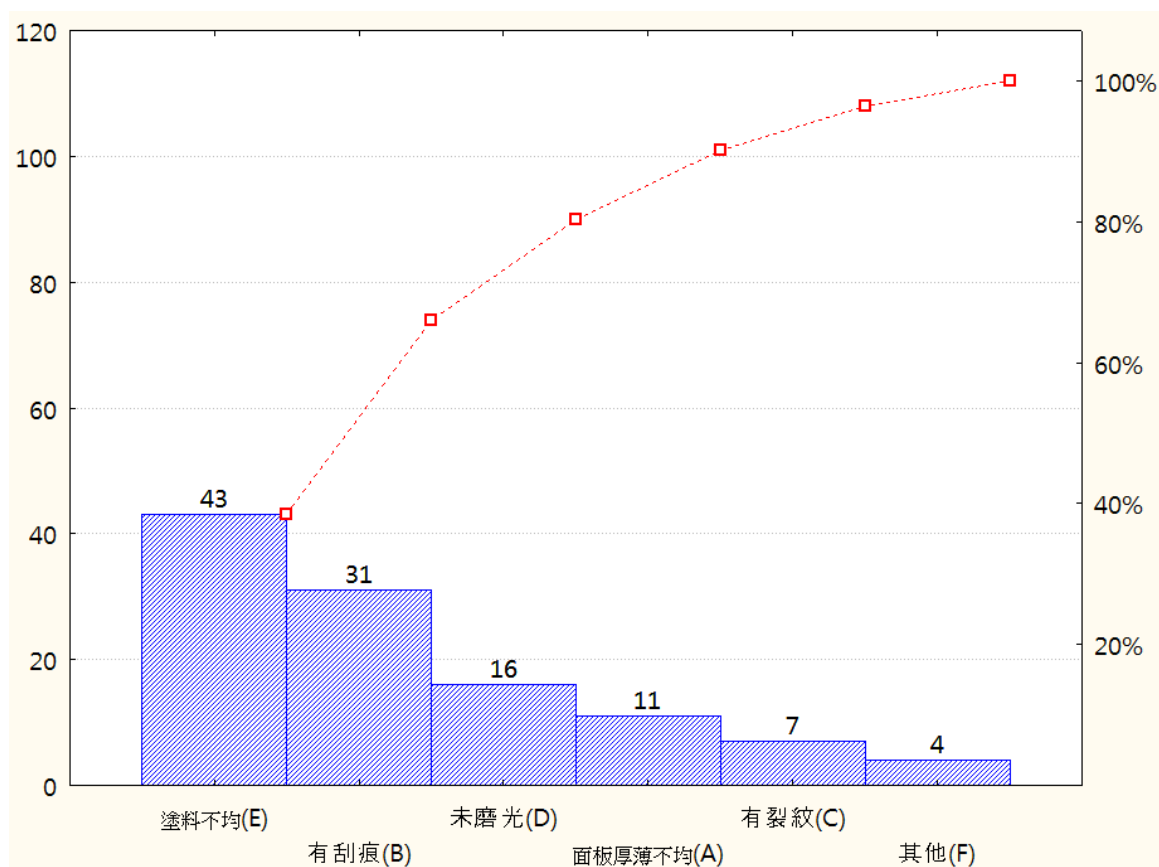


3.柏拉圖：義大利經濟學家柏拉圖(Vilfredo Pareto)認為社會上大部份的財富是操縱在少數人的手中。同理，在改善製程品質時，通常造成品質不良或缺失的因素也符合柏拉圖原理。因此柏拉圖分析是依據「重要少數，瑣細多數」法則，找出造成問題最關鍵之幾個少數因素，以有效改善問題。

作法：

- 1)水平軸---種類，對每一種類繪入直立條棒。
- 2)垂直軸---觀察值的次數，相對次數或百分比。
- 3)每一條棒須有相同寬度。

例 3：承例 1，繪製柏拉圖。



二、定量資料如何以圖來表示？

—利用點圖（Dot Diagram）或直方圖（Histogram）。

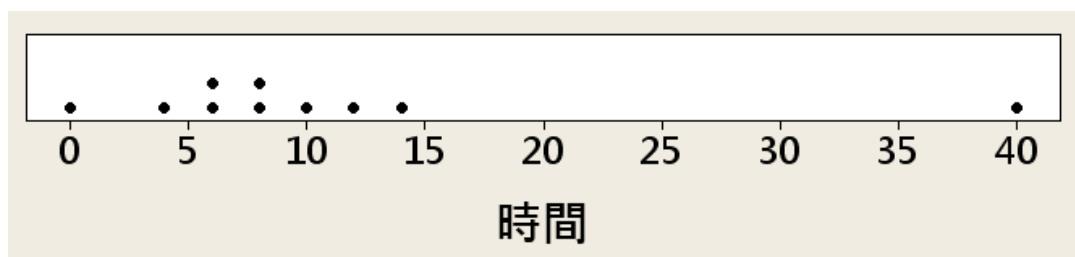
1.點圖：點圖可用以顯現資料之分佈型態。

作法：1)水平軸---數線（標有數據的直線）。

2)在數線上繪入點。

例 4：一位教育人員研究某班級上統計課時能否專心學習的問題，他任選 10 位學生，並分別記錄 50 分鐘內學生開始發生無法專心學習的時間點，得結果如下：

13, 10, 5, 7, 5, 3, 12, 0, 7, 40 請以點圖表之。



2.直方圖：直方圖是次數分佈的圖形表示，是由直立的條狀或矩形所構建。

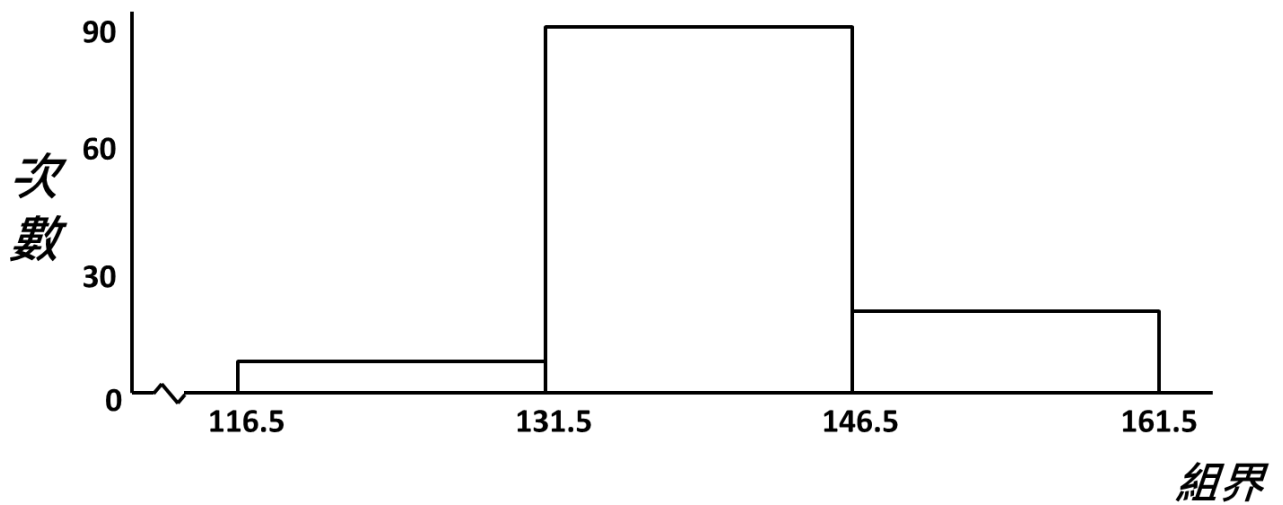
作法：

- 1)水平軸---代表各組之所有組界。
- 2)垂直軸---代表觀察值的次數，相對次數或百分比。
- 3)在水平軸上畫矩形，這些矩形須有相同寬度並且須相連在一起。

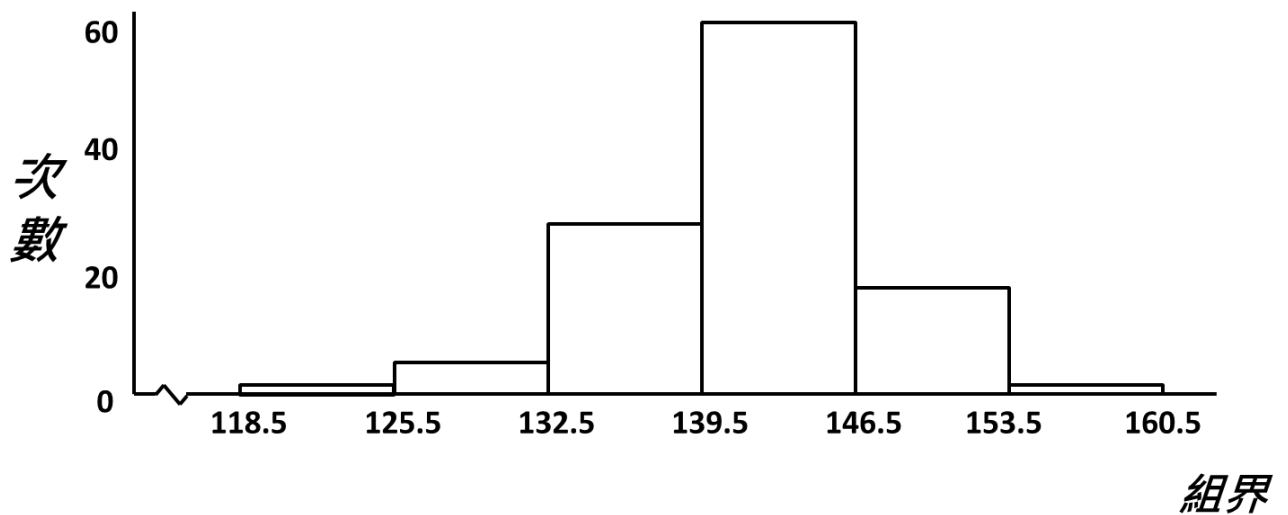
例 5：請依下例資料繪製直方圖。

組限	組界	組中點	次數	相對次數	相對次數 X100
119-125	118.5-125.5	122	1	.009	0.9
126-132	125.5-132.5	129	4	.038	3.8
133-139	132.5-139.5	136	26	.245	24.5
140-146	139.5-146.5	143	59	.557	55.7
147-153	146.5-153.5	150	15	.142	14.2
154-160	153.5-160.5	157	1	.009	0.9
		Total	106	1.000	100.0

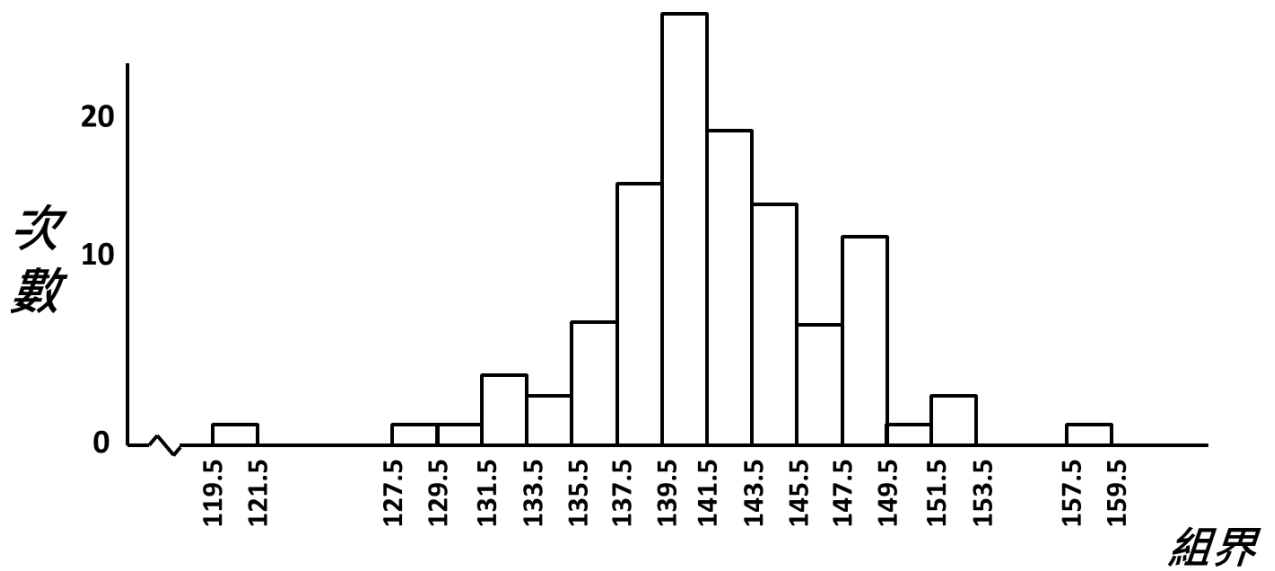
利用例 5 資料繪製直方圖時，組數不宜太多或太少



(a) 組數 = 3, 組距 = 15 → 組數過少



(b) 組數 = 6, 組距 = 7 → 組數恰當



(b) 組數 = 20, 組距 = 2 → 組數過多

● 第二單元 常用統計量或指標

一、原始數據特徵值之計算

原始**連續型**數據分析之特徵主要可分為以下四大類：

1. **集中趨勢**(Central Tendency of Location)
2. **變異**(Dispersion)
3. **偏態**(Skewness)
4. **峰態**(Kurtosis)

1. **集中趨勢**：「集中趨勢指標」是表示一組數據中央點位置所在的一個指標。

● 最常用的集中趨勢指標：平均數、中位數、眾數。

1) **平均數**： 群體平均數： $\mu = \frac{\sum Xi}{N}$

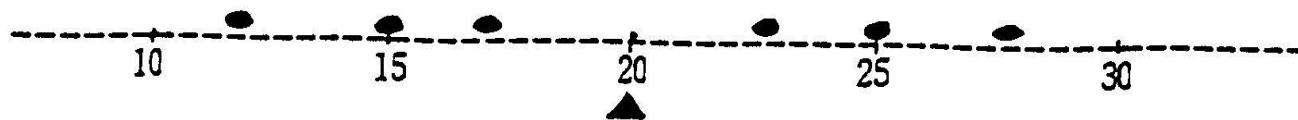
樣本平均數： $\bar{X} = \frac{\sum Xi}{n}$

其中 N 表群體大小，n 表樣本大小。

例 6：請找出下列群體數據之平均數：0, 7, 3, 9, -2, 4, 6

例 7：請找出下列樣本數據之平均數：25, 12, 23, 28, 17 and 15。

例 8：請將例 7 之資料繪入下面之點圖中，並標出平均數。此平均數為數據之「平衡點」。



2) **中位數**：將一組數據由小至大排序後，最中間的那一個數值稱為中位數。

● 找中位數之方法：

(1) 當 n =奇數， \tilde{X} = 排序第 $(n+1)/2$ 位之數值。

(2) 當 n =偶數， \tilde{X} = 排序第 $(n/2)$ 位及第 $(n/2)+1$ 位的兩數值之平均數。

例 9：請找出下列樣本數據之中位數：9, 2, 7, 11, 14

例 10：請找出下列樣本數據之中位數：9, 2, 7, 11, 14, 6

3) **眾數**：在一組數據，出現次數最多的數值稱之。

例 11：3, 3, 2, 1, 4, 2, 3

例 12：3, 1, 4, 2

● 何時用平均數？何時用中位數或眾數？

平均數對**離群值**非常敏感，而中位數或眾數則對離群值較不敏感。因此，當資料中有離群值時，則使用中位數或眾數，否則，使用平均數。

例 13：1, 3, 4, 6, 6, 9, 13

$$\bar{X} = \frac{1+3+\dots+13}{7} = 6, \text{ 中位數} = \underline{\hspace{2cm}}, \text{ 眾數} = \underline{\hspace{2cm}}$$

若此組數據最後一筆資料改成 **70**：1, 3, 4, 6, 6, 9, 70

$$\text{則 } \bar{X} = 14.14, \text{ 中位數} = \underline{\hspace{2cm}}, \text{ 眾數} = \underline{\hspace{2cm}}$$

例 14：設有甲和乙兩個國家，其國民年所得如下表

（以美金計）：

	甲	乙
平均數	\$3,850	\$4,650
中位數	\$4,000	\$1,200

假如兩個國家的其他因素（如：稅收等等）均相似，則哪一國有較高之生活水準？理由為何？

答： 甲 國家有較高之生活水準。因其平均數和中位數非常接近，表示了該國有許多國民年所至少超過該國國民年所得之平均數或中位數。然而， 乙 國家之財富卻集中在少數人的手中。

2. 變異 (Dispersion)

「變異指標」是表示一組數據間差異大小或數值變化的一個量數。

●三個主要量測變異之指標：

全距 (Range)，**變異數**或**標準差** (Variance and Standard Deviation) 及 **變異係數** (Coefficient of Variation)

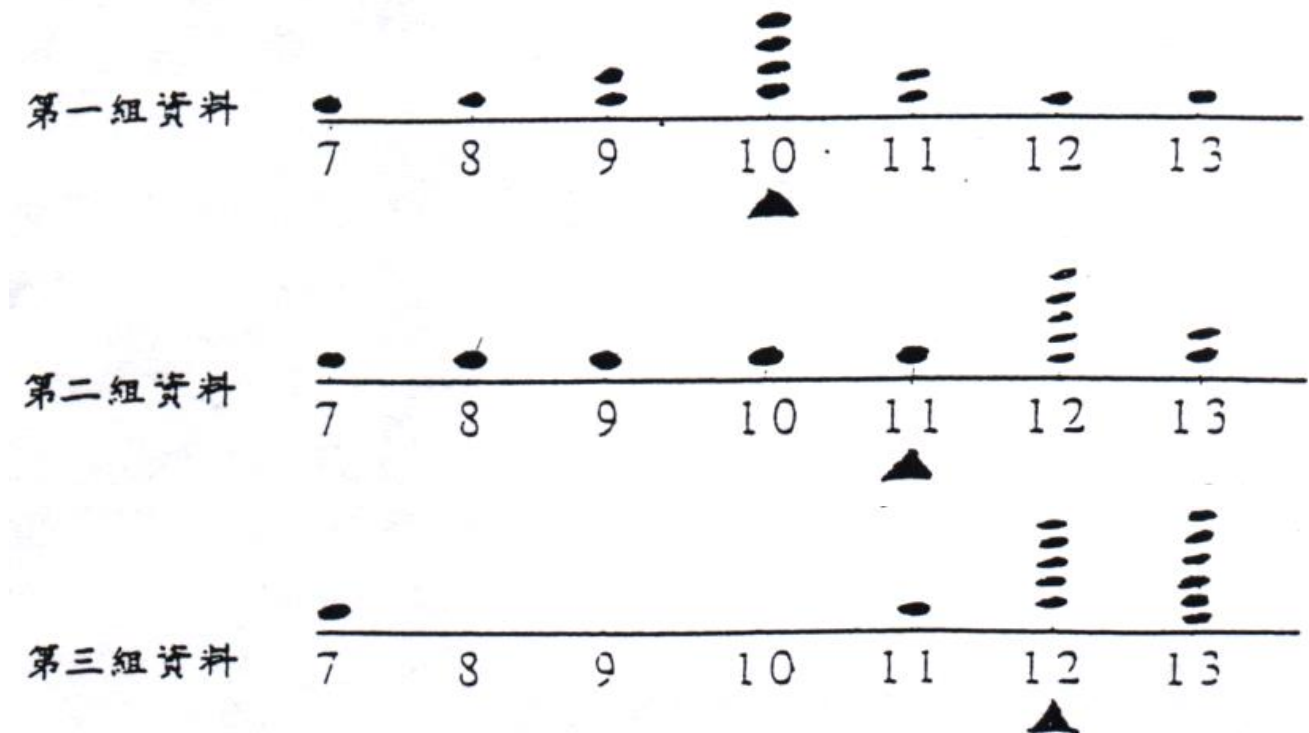
1) **全距** (R)：全距是用來衡量一組數據差異大小或數值變化最簡單的方法

公式： $R = \boxed{\text{最大值} - \text{最小值}}$

●用全距之缺點：

當一組數據中有 離群值 出現或資料量太大($n > 10$)時，全距並非一個很好的衡量數據變異的量數，因其無法解釋最小與最大值之間，數據分佈的情形。

例 16：以下三組數據有相同之全距，不同之分佈。



2) 離中趨勢指標：變異數和標準差

$$\text{群體變異數：} \sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1}$$

樣本變異數：

$$= \frac{\text{平方和} - (\text{和的平方} / \text{數據總和})}{\text{數據總和} - 1}$$

$$\text{群體標準差：} \sigma = \sqrt{\sigma^2}$$

$$\text{樣本標準差：} S = \sqrt{S^2}$$

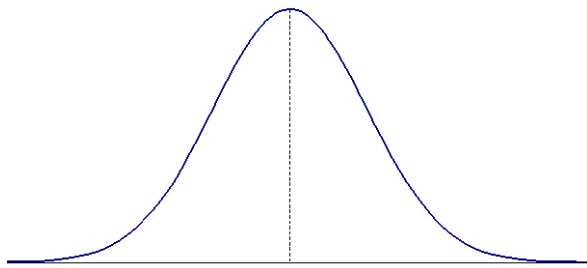
例 17：請找出下列樣本數據之平均數、變異數及標準差：5, 8, 1, 2, 4

3. 偏態 (Skewness)

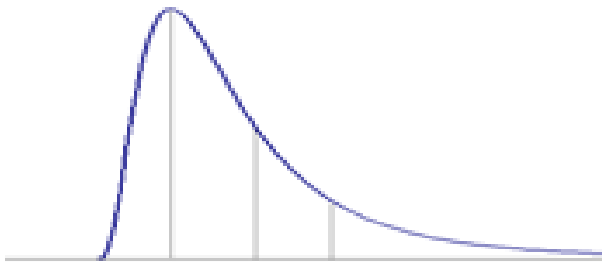
「偏態」是用來說明一組數據分佈的形態。

● 單峰分佈有三種形態之偏態：

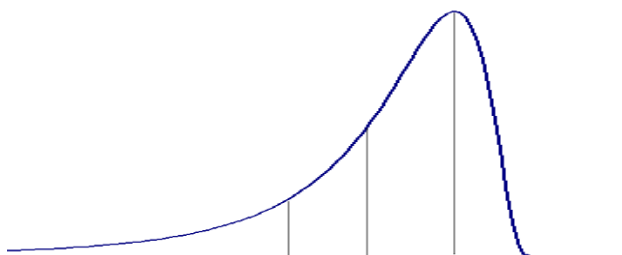
1) 對稱：平均數_____中位數



2) 右偏，正偏：平均數_____中位數



3) 左偏，負偏：平均數_____中位數



●偏態係數

樣本偏態係數之公式如下：

$$g_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3 / (n-1)}{S^3}$$

1. 偏態係數 = _____ 表示樣本分佈是對稱的。
2. 偏態係數 = _____ 表示樣本分佈是偏右的。
3. 偏態係數 = _____ 表示樣本分佈是偏左的。

4.峰度（Kurtosis）

●樣本峰度係數之公式如下：

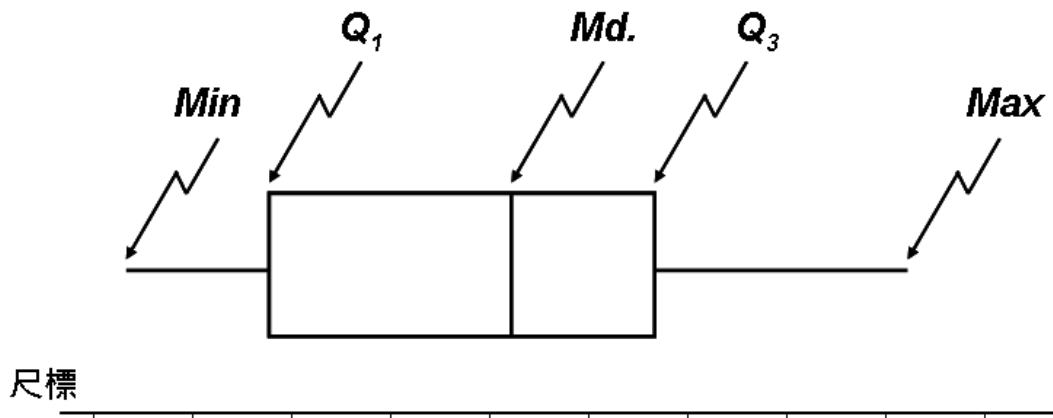
$$g_2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4 / (n-1)}{S^4} - 3$$

1. 峰度係數 = 0 表示資料呈常態峰。
2. 峰度係數 > 0 表示資料呈高峽峰。
3. 峰度係數 < 0 表示資料呈低闊峰。

● 盒鬚圖

● 何謂盒鬚圖（Box-Whisker Plot，簡稱 Box Plot）？

盒鬚圖是資料的一種圖形展示法。此圖可同時標出資料之集中趨勢、變異、偏態、最小值、最大值等。此圖又稱「五指標摘要圖」(five-number summary plot)



Q1：第一四分位數或第 25 百分位數。

Q2：第二四分位數或中位數(Md.)。

Q3：第三四分位數或第 75 百分位數。

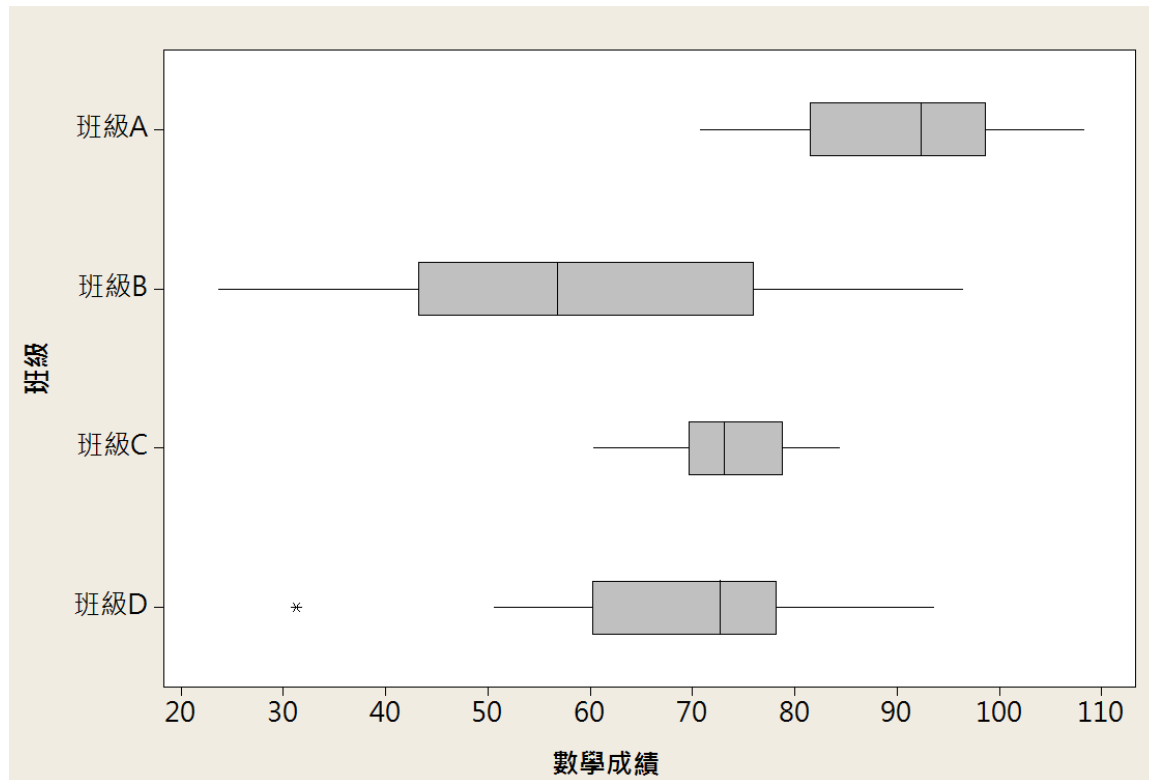
● 盒鬚圖之主要功用：

從視覺上即可有效的找出資料之主要的表徵值。

●盒鬚圖之其它功用：

1.可同時比較數組資料

例 19：比較四個班級的數學成績



2.可辨認出離群值

●何謂離群值（Outliers）？

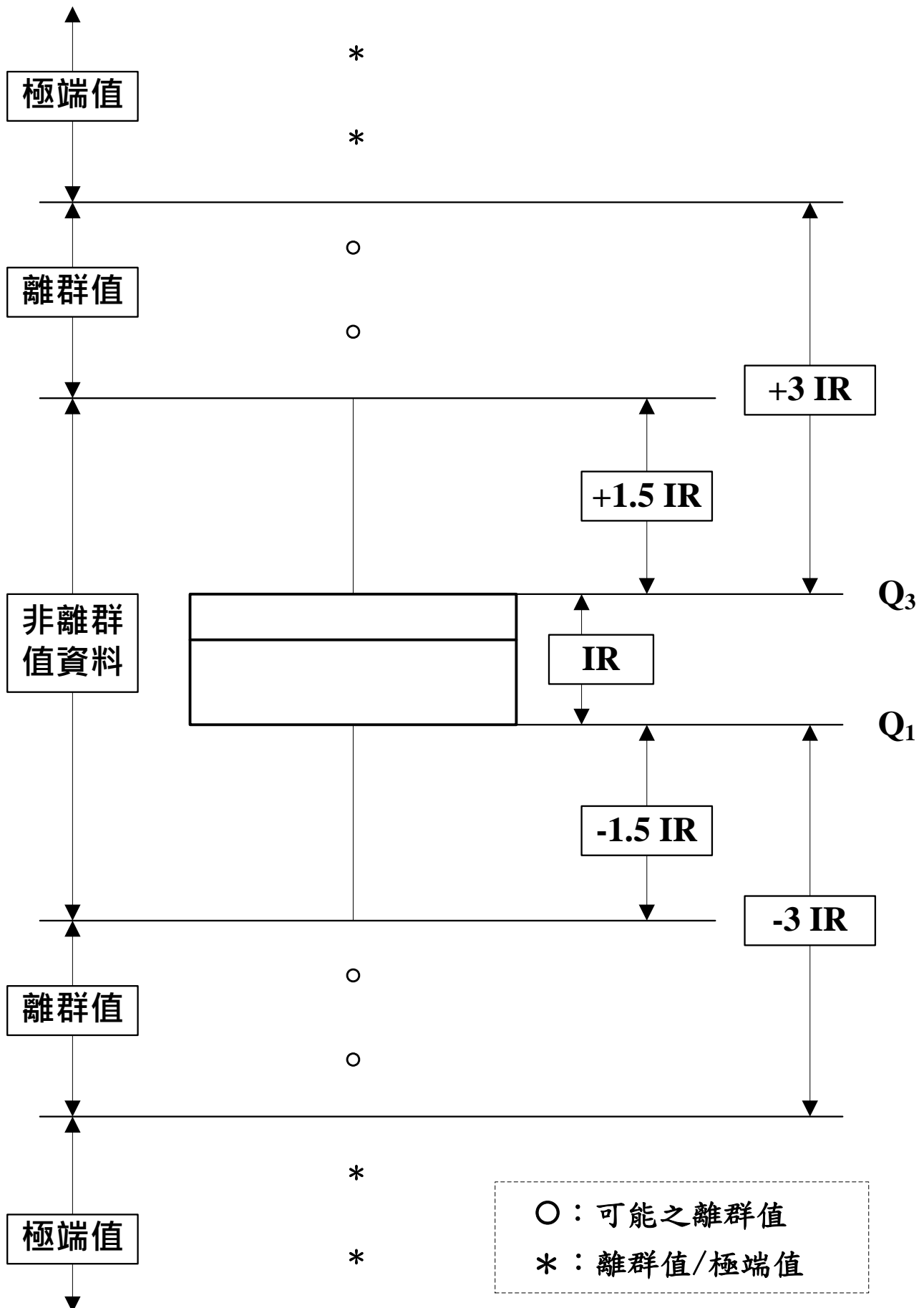
離群值是遠大於或遠小於同一筆數據中之其它值之數據。

●如何利用盒鬚圖辨認出離群值？

1)超過盒鬚圖之盒 $1.5 (Q_3 - Q_1)$ 至 $3 (Q_3 - Q_1)$ 距離內之值可當作可能之離群值或極端值。

2)超過盒鬚圖之盒 $3 (Q_3 - Q_1)$ 距離外之值可當作非常可能之離群值。

註： $Q_3 - Q_1$ = 第 75 百分位數 - 第 25 百分位數 = 中四分位距 (Interquartile Range, IR)



例 20：下列資料為三條生產線的良率，請依下列資料繪製盒鬚圖。

生產線 1	生產線 2	生產線 3
0.996	0.992	0.998
0.997	0.991	0.993
0.995	0.990	0.997
0.998	0.991	0.998
0.999	0.992	0.996
0.998	0.996	0.988
0.986	0.995	0.991
0.998	0.995	0.997
0.997	0.997	0.998
0.988	0.990	0.996

