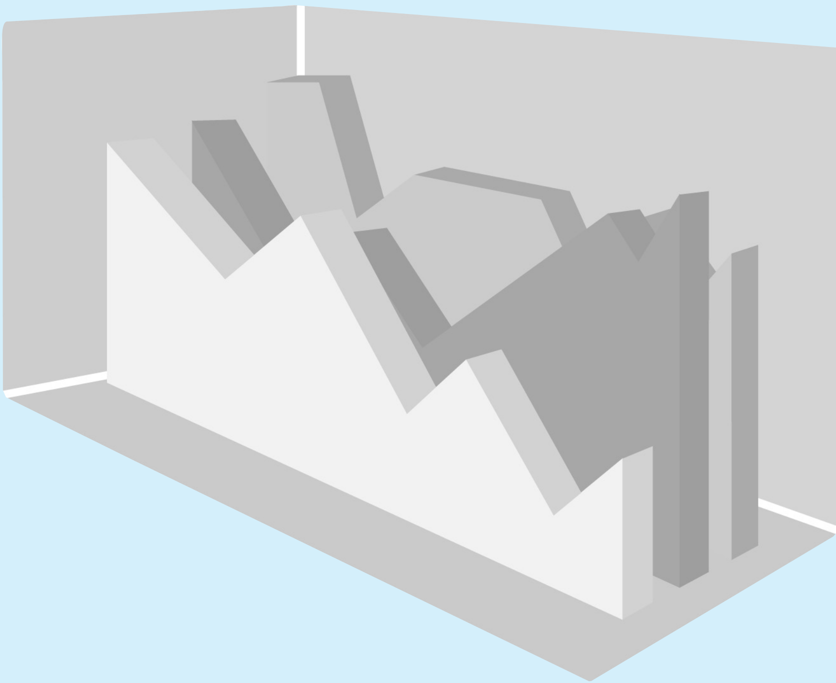
1h271.tpf-1 7/22/2004 10:39:14

第一章

統計基本概念



統計學（Statistics）是一門決策常使用的科學，它提供了蒐集、組織、分析及歸納各種數據資料的科學程序。

統計學的應用範圍非常廣泛，其應用範圍包括各種社會科學與自然科學。例如，政府機構可以利用統計學了解一般民意的反應；選舉時，各候選人用統計方法預測得票率；教育家可以藉此得知不同的教學方法之功效；心理學家可以憑此測知人類心理的特質；醫學研究者可以了解某種病因的影響情形。

統計學的一些基本概念

統計學中常常見到「母體」（Population）和「樣本」（Sample）等字樣，其意義說明於後。

統計學的目的在於研究問題，然後解決此問題。而欲達到研究的目的，必先蒐集相關的資料，若是所蒐集的資料，其所包含的範圍為全部有關的資料時，則此全部範圍稱之為「母體」。但是，通常我們可能限於時間或成本，無法調查全部母體，可以從其中具有代表性的一部分資料，加以調查，則此部分的資料，即稱之為「樣本」。

通常對於用來描述母體特徵的數，稱之為「母數」（Parameter），而用來說明樣本特徵的數，則稱之為「統計量」（Statistic）。常見的母數中的平均數 μ，母數的標準差 σ；其相對應的樣本統計量，平均數為 $\overline{X}$，樣本標準差為 S。

常用的各種統計量

2.1 集中趨勢統計量

能代表一組資料的數，常用的有「平均數」（Mean）以及「加權平均數」（Weight Mean）。

⑴平均數

將所有的樣本資料總合，除以樣本數，即為樣本平均數。例如，有 n 筆資料數據為 x_1，x_2，…，x_n 則樣本平均數 $\overline{X}$ 為：

$$\overline{X} = \frac{x_1 + x_2 + \cdots + x_n}{n} \qquad (1.1)$$

【例 1】

為調查全班統計學的成績狀況，於全班中抽樣 10 位同學之成績如下：

80，75，60，83，72，70，55，80，92，78

試計算此 10 位同學統計學之平均成績。

解：

$$\Sigma x = 80+75+60+83+72+70+55+80+92+78$$

$$= 745$$

則 $\overline{X} = \frac{745}{10} = 74.5$

⑵加權平均數

如果每一個資料x_i之權重並不相同（所謂權重，是指重要程度之比例），則 n 個資料的平均數不應該都以 1/n 作為權重。

設有 n 個資料，其權重比例各為 w_1，w_2，w_3，…，w_n，則其加權平均數為：

$$\widetilde{X} = \sum_{i=1}^{n} w_i x \qquad (1.2)$$

【例 2】

某校教學每週國文課 5 小時、數學 7 小時、歷史 3 小時。

今有甲、乙兩學生，其成績如下：

甲生：國文 65 分、數學 70 分、歷史 30 分；

乙生：國文 60 分、數學 75 分、歷史 30 分；

試求此二位同學學業成績加權平均數。

解：

由於此三個科目每週教學時數並不相同，首先應將此三科授課時數總數相加：

$5+7+3=15$，

而國文科之授課時數加權為 $w_1=\frac{5}{15}$

數學科之授課時數加權為 $w_2=\frac{7}{15}$

歷史科之授課時數加權為 $w_3=\frac{3}{15}$

因此，此二位學生之學業成績加權平均數各為：

$$\tilde{x}_{甲}=65\times\frac{5}{15}+70\times\frac{7}{15}+30\times\frac{3}{15}$$
$$=60.3$$

$$\tilde{x}_{乙}=60\times\frac{5}{15}+75\times\frac{7}{15}+30\times\frac{3}{15}$$
$$=61$$

2.2 離勢統計量

表達資料分散的情形，常常使用的統計量有：變異數（Variance）、標準差（Standard Deviation）和變異係數（Coefficient of Variation）。

⑴變異數

n 個資料 x_1，x_2，…，x_n 的變異（Variation）是指各 x_i 與 $\overline{X}$ 差距之平方和，以 Sxx 表示：

$$Sxx=\sum_{i=1}^{n}(x_i-\overline{X})^2 \tag{1.3}$$

亦可證出

$$Sxx=\sum_{i=1}^{n}x_i^2-n\overline{X}^2 \tag{1.4}$$

而，變異數 S^2 即變異 Sxx 除以自由度（n－1），即

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

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}x_i^2-n\overline{X}^2\right] \qquad (1.5)$$

所謂自由度，是指可以自由活動的程度。例如，若無其他條件要求，對於 n 個數據可以有任意的 n 個決定的自由度。通常，若是既定有 k 個統計量，則 n 個數據之自由度為 $n-k$。因而，若已知 $\overline{X}$，則此 n 個數據，由於受到了 $\overline{X}$ 的限制，此時只能任意指定其中的 $n-1$ 個數據，也就是說，只有 $n-1$ 個自由度。因此，若是已知母體之平均數 μ，沒有統計量之限制，則自由度為 n，所求出的母體變異數為：

$$\sigma^2=\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} \qquad (1.6)$$

【例 2】

試以例 1 之數據，計算樣本變異數 S^2。

解：

$$S^2=\frac{1}{n-1}\left[\Sigma x_i^2-n\overline{X}^2\right]$$

$$=\frac{1}{9}[6400+5625+3600+6889+5184+4900+3025$$

$$+6400+8464+6084-10\times 74.5^2]=118.72$$

⑵標準差

標準差之定義，就是變異數開根號，亦即

$$S=\sqrt{S^2} \tag{1.7}$$

以例 3 而言，

$$S=\sqrt{118.72}=10.896$$

⑶變異係數

對於不同單位的各種資料，所計算出變異數的大、小，不能就此作為變異程度之大、小，也不能用以作為離散程度之比較。

【例 3】

有 5 位同學之身高，分別以公分及公尺之記錄如下：

以公分計	175	180	165	170	172
以公尺計	1.75	1.80	1.65	1.70	1.72

試以兩種不同長度單位，計算此 5 位同學身高之變異數。

解：

設以公分計之身高變數為 X，以公尺計之身高變數為 Y，然後先求各變數之平均數：

$\overline{X} = (175 + 180 + 165 + 170 + 172) \div 5 = 172.4$

$\overline{Y} = 1.724$

$Sxx = (175^2 + 180^2 + 165^2 + 170^2 + 172^2) - 5 \times (172.4)^2 = 125.2$

$Syy = (1.75^2 + 1.80^2 + 1.65^2 + 1.70^2 + 1.72^2) - 5 \times (1.724)^2$

$\quad = 0.01252$

同理，$S_x^2 = 125.2 \div 4 = 31.3$

$S_y^2 = 0.01252 \div 4 = 0.00313$

從這個例子可以看出，同樣的一件事，如果所使用的單位不同，其變異數所得的結果竟然會有 10000 倍之差距，因此，對於不同單位之變異情形，宜以「變異係數」計算之。變異係數之定義為：

$$C.V = \frac{S}{X} \times 100 \qquad (1.8)$$

再以變異係數計算本例：

$$C.V_x = \frac{\sqrt{31.3}}{172.4} \times 100 = 3.245$$

$$C.V_y = \frac{\sqrt{0.00313}}{1.724} \times 100 = 3.245$$

此時可以看出，對於同樣的一件事情，即使不同單位之衡量，以變異係數就可以表現出相同的變異情形。

標準分數

所謂的標準分數（Standard Score），是將原來的數據轉換成標準值的一種方法，最常用的是「Z分數」（Z-Score）。

所謂 Z 分數，是將原始數據減去平均數，再除以標準差，所得出的數據稱之為Z分數。Z分數之功用為：對於不同單位的兩組數據，應先化成Z分數，才可加以比較。原始分數轉化成Z分數之後，可以相互加減運算，並能比較數值之大、小。

Z 分數轉換之方法為：

3.1 母體資料

$$Z=\frac{x-\mu}{\sigma} \tag{1.9}$$

3.2 樣本資料

$$Z=\frac{x-\overline{X}}{S} \tag{1.10}$$

【例 4】

假設某項考試，甲、乙二人參加應試。但是甲生選考統計學，分數 85 分，該場考試之統計學平均分數 80 分，