

第一章

緒 論

1-1 統計學的意義與分類

統計學 (statistics) 為蒐集、整理、分析、及推論數字資料 (numerical data) 的科學方法。我們蒐集得到的原始資料通常雜亂無章，必須加以整理使其系統化，然後加以分析以了解其特性。並可由已知的樣本資料推論到未知的全體，例如由本市一百位國小六年級學童樣本近視百分比，推論全市國小六年級全體學童近視的百分比，並計算其誤差。統計學根據其是否作統計推論，可區分為敘述統計學 (descriptive statistics) 及推論統計學 (inferential statistics) 兩部分：

(一) 敘述統計學

敘述統計學僅在整理及分析資料本身，並不由已知的資料推論到未知的部分。我們蒐集得到資料通常毫無組織，敘述統計學告訴我們如何加以整理、計算、排列及敘述以使原始資料變得簡明扼要並有意義，使人容易了解。例如某校為了解該校一年級新生的智力水準，可以實施智力測驗後，根據測驗分數予以整理歸納，以表或圖加以表達；並可求出智力的平均數以了解一般的智力水準，或者算出智力變異的情形，以了解學生個別差異的大小。由於僅是就某校一年級新生的資料加以統計分析，並不推論到其他學校的新生，所以是敘述統計學。

(二) 推論統計學

一般科學研究，由於受到時間、金錢、人力、物力的限制，通常無法將研究對象全部加以測量或調查，只能由其中抽取部分樣本加以研究。例如我們要研究全市六年級學生的數學能力，全市國小六年級學生就是我們研究的對象，在統計學上稱為母群體 (population)。通常研究者採用抽樣方式 (sampling method)，由母群體中抽取樣本 (sample)，例如抽出三

百位國小六年級學生，實施數學能力測驗後，計算其平均數學成績等統計量，並據以推論全市國小六年級學生的數學能力，計算其可能的推論誤差。在推論統計裡，研究者真正的目的是推論母群體的性質，而不是敘述樣本的性質。

除了敘述與推論統計之外，有些學者將實驗設計及分析（experimental design and analysis）特別提出。在教育及心理的研究上，我們常需做各種實驗，而實驗必須經過仔細的設計，並選用適當的統計分析方法。統計分析與實驗設計關係密切，不但實驗的結果必須經過統計分析；而且透過統計分析，我們可以事先選用最經濟有效的實驗方式。如此統計學就可區分為(1)敘述統計學、(2)推論統計學、及(3)實驗設計與分析。

統計學所處理的資料是數字資料，因此與數學關係密切，是數學的一種主要分支。同時統計學是一種科學方法，因此被普遍應用到各種社會科學與自然科學中。所以統計學又可區分為數理統計學（mathematical statistics）及應用統計學（applied statistics）；前者在探討各種統計方法的數學原理以及各種統計公式的來源，後者則著重如何將各種統計方法應用到社會及自然科學上。教育及心理統計學為應用統計學的一種，其目的在探討如何應用有關的統計方法以研究各種教育及心理問題與現象。

1-2 變數及其類別

變數（variable）又稱為變項，是指可以不同數值或狀態出現的屬性（property）。例如性別是一種變數，因為有人是男性，有人是女性。身高是一種變數，有人是 170 公分，有人是 165 公分，有人是 150 公分等等。其他如血型、體重、智力、年齡、教育程度、職業、收入、價值、態度、性格、學業成績等都會因人而異，都是變數。燈光的亮度、品種、顏色及教學方法等可以有不同的變化或種類，也是變數。與變數相反的是常數（constant），常數是恆為定值，可視為變數的一種特殊型態。變數通常以

X ， Y ， Z 表示；常數則以 a ， b ， c 表示。

變數的分類，常見之方法如下：

(一) 自變數與依變數

兩個變數 X 與 Y ，當 Y 之值隨 X 之值而變時，在數學上便稱 Y 為 X 的函數，以 $Y=f(X)$ 表示。此時， X 稱為自變數 (independent variable)， Y 稱為依變數 (dependent variable) 或應變數。在教育與心理研究中，我們所研究的變數有時可分為自變數與依變數。例如研究不同教學方法對學生學業成績的影響，此時教學方法為自變數而學業成績為依變數。又如研究燈光亮度對讀書效率的影響，則燈光亮度為自變數而讀書效率為依變數。

(二) 連續變數與間斷變數

連續變數 (continuous variable) 可以有無數個不同之值，任何兩個值之間都可以加以無限制的細分。時間、年齡、身高、體重、智商等都是連續變數。我們測量連續變數所得到的數值會受到測量工具精密度的影響，越精密的測量工具可測得越細微的數值。用精密到 0.1 秒的馬錶測得某人百米賽跑成績為 10.4 秒；若改用精密到 0.01 秒的馬錶，可能發現更精確的數值是 10.42 秒。只要我們有更精密的測量工具，理論上我們可繼續細分下去。事實上，我們測量連續變數所得到的數值應視為一段距離而非一點。例如當我們測得某人為 160 公分，精確至公分為止，則其可能身高應是在 159.5 至 160.5 公分之間。也就是測量值的誤差等於其最末一位數值的半個單位，160 公分只是一個近似值而已。圖 1-1 表明測量連續變數所得的數值應視為一段距離：

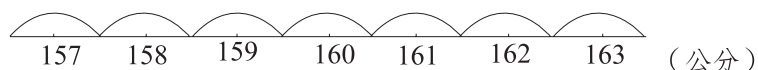


圖 1-1 連續變數各數值所涵蓋之距離

間斷變數 (discrete variable) 一般是由點記 (counting) 所得，例如競選班長，張三得 23 票，李四得 12 票，票數即為間斷變數。票數不是連續的，我們不能加以無限制的細分，只能得到特殊的數值；我們不能說張三得到 23.3 票或李四得到 12.4 票。間斷變數的每一個數值，係代表一個點，而不是一段距離。班級人數、某縣市學校數、家庭人數、擲硬幣得到正面的次數、擲骰子得到的點數等均屬間斷變數。

(三) 名義變數、次序變數、等距變數和比率變數

變數根據其測量尺度 (scale of measurement) 精密的程度，由低而高，區分為名義、次序、等距及比率變數，說明如下：

1. 名義變數 (nominal variable)

又稱為類別變數 (categorical variable)，其目的僅在於分類，性別是一種名義變數，可分為男性與女性，通常我們用 0 代表女性，用 1 代表男性。由於性別僅能分為兩種，故為一種兩分類別變數。其他如血型、國籍、顏色、郵遞區號、電話號碼、球衣背號、學生座號、身分證字號等都是名義變數。名義變數所使用的數字只是作標誌 (label) 用，不適合於加、減、乘、除等算術運算的規則；例如我們不能說 4 號加 2 號等於 6 號，也不能說 20 號比 15 號大。嚴格地說，名義變數並未含有測量尺度的觀念，因為名義變數所使用的數字只是標誌而已，並非測量的結果。

2. 次序變數 (ordinal variable)

如其名稱所示，次序變數除了分類之外，尚能表明各類別的次序、大小或優劣。例如比賽中的冠、亞、季軍不僅能區別比賽者的成績，且可顯示其優劣。其他如名次、等第 (A、B、C 或甲、乙、丙)、軍階等均是次序變數。次序變數能用「>」（大於）或「<」（小於）表示，雖能分別優劣、大小，但不能說各等級之間的差異是相等的。例如第 1 名與第 2 名之間的差異不一定等於第 2 名與第 3 名之間的差異；也不能說第 3 名減第

2 名等於第 2 名減第 1 名。

3. 等距變數 (interval variable)

等距變數的測量尺度較次序變數精密，不但能比較大小，且能計算差別的分量，因為它具有單位相等 (equality of units) 的特性。攝氏及華氏溫度是最常見的等距變數，溫度計上每度之間的距離都是相等的。12°與 24°之間的差別等於 24°和 36°之間的差別。但是攝氏及華氏溫度的零點乃是人定的，並非絕對的零點，因此我們不能說 24°是 12°的兩倍熱，36°是 12°的三倍熱。不過我們可以說 36°和 12°之間的差別是 24°和 12°之間差別的兩倍。智商常被視為等距變數，然而智商 120 與 110 之間的差別是否等於智商 90 與 80 之間的差別並不一定，因此智商也許應是次序變數。不過為了計算方便，我們常將智商視為等距變數，以便作加減乘除的運算。

4. 比率變數 (ratio variable)

比率變數除具有等距變數所有的特點外，更具有絕對零點 (absolute zero)，可以直接形成比率。例如長度就是比率變數，2 公尺是 1 公尺的兩倍。時間也是比率變數，某甲 10 分鐘走完 1 公里，某乙 20 分鐘走完 1 公里，我們可以說某乙用的時間是某甲的兩倍。其他如身高、體重、年齡等均屬於比率變數。

表 1-1 把四種變數特性作一說明，可看出後面較精密的變數均具有前面變數的特性。在教育與心理統計學上，等距與比率變數的區別並不重要，兩者適用的統計方法並無不同，因此兩者可合而為一。如此則有三種變數：(1)名義變數，(2)次序變數，及(3)等距及比率變數。三種變數各有其適用之統計方法，而以有關等距及比率變數之統計方法最多也最重要。有些學者把等距及比率變數合稱為數量變數 (quantitative variable)，因其具有相等單位與一般數字之性質相同，而把名義及次序變數合稱為質的變數 (qualitative variable)。

表 1-1 四種變數特性表

變 數	特 性	舉 例
名義	分類	性別 身分證字號
次序	分類 表示大小或優劣	名次 等第
等距	分類 表示大小或優劣 單位相等	攝氏溫度 華氏溫度
比率	分類 表示大小或優劣 單位相等 具有絕對零點	長度 時間 身高 體重

1-3 總和（以 Σ 表示，讀做 Sigma 或 Summation）的計算

變數可以各種數值出現，這些數值稱為變量（variate）。例如身高是一種變數，而各個人的身高就是變量。處理變量的方法，最常見的是求它們的總和。X 變數若有三個變量 X_1 ， X_2 和 X_3 ，則其總和用 $\sum_{i=1}^3 X_i$ 表示：

$$\sum_{i=1}^3 X_i = X_1 + X_2 + X_3$$

Σ 符號下面的 $i=1$ 及上面的 3，分別代表第一個及最後一個變量，即 X_1 與 X_3 ；求總和即依序由 X_1 連加至 X_3 為止。如 X 有 N 個變量 X_1 ， X_2 ， X_3 ，……， X_N ，則其總和為：

$$\sum_{i=1}^N X_i = X_1 + X_2 + X_3 + \cdots + X_N$$

如果計算總和的起迄點已經很清楚，通常可將 Σ 上下的小符號省略，只寫 ΣX_i ；或者連 X_i 的 i 也省略，只寫 ΣX 。

表 1-2 有兩個變數 X 及 Y ，以及其各種總和的計算過程：

表 1-2 總和計算之說明

X	Y	X^2	Y^2	XY	X+Y
10	3	100	9	30	13
15	5	225	25	75	20
12	1	144	1	12	13
8	1	64	1	8	9
10	3	100	9	30	13
總和 55	13	633	45	155	68

$$\Sigma X = 10 + 15 + 12 + 8 + 10 = 55$$

$$\Sigma Y = 3 + 5 + 1 + 1 + 3 = 13$$

$$\begin{aligned}\Sigma X^2 &= 10^2 + 15^2 + 12^2 + 8^2 + 10^2 \\ &= 100 + 225 + 144 + 64 + 100 = 633\end{aligned}$$

$$\begin{aligned}\Sigma Y^2 &= 3^2 + 5^2 + 1^2 + 1^2 + 3^2 \\ &= 9 + 25 + 1 + 1 + 9 = 45\end{aligned}$$

$$(\Sigma X)^2 = (55)^2 = 3025$$

$$(\Sigma Y)^2 = (13)^2 = 169$$

$$\begin{aligned}\Sigma XY &= (10 \times 3) + (15 \times 5) + (12 \times 1) + (8 \times 1) + (10 \times 3) \\ &= 30 + 75 + 12 + 8 + 30 = 155\end{aligned}$$

$$\Sigma X \Sigma Y = 55 \times 13 = 715$$

$$\begin{aligned}\Sigma (X + Y) &= (10 + 3) + (15 + 5) + (12 + 1) + (8 + 1) + (10 + 3) \\ &= 13 + 20 + 13 + 9 + 13 = 68\end{aligned}$$

$$\Sigma X + \Sigma Y = 55 + 13 = 68$$

由表 1-2 中各種總和，可以看出：(1) $\sum X^2 \neq (\sum X)^2$ ， $\sum Y^2 \neq (\sum Y)^2$ ；(2) $\sum XY \neq \sum X \sum Y$ ；但是(3) $\sum(X+Y) = \sum X + \sum Y$ 。有關總和之運算規則，主要有下列三個：

- (1) $\sum_{i=1}^N C = NC$
- (2) $\sum CX = C \sum X$
- (3) $\sum(X+Y) = \sum X + \sum Y$

上面三個等式中，C 為常數，X 與 Y 為變數。將上面三個規則說明如下：

- (1) $\sum_{i=1}^N C = \underbrace{C + C + \cdots + C}_N = NC$
- (2) $\begin{aligned} \sum CX &= CX_1 + CX_2 + \cdots + CX_N \\ &= C(X_1 + X_2 + \cdots + X_N) \\ &= C \sum X \end{aligned}$
- (3) $\begin{aligned} \sum(X+Y) &= (X_1 + Y_1) + (X_2 + Y_2) + \cdots + (X_N + Y_N) \\ &= (X_1 + X_2 + \cdots + X_N) + (Y_1 + Y_2 + \cdots + Y_N) \\ &= \sum X + \sum Y \end{aligned}$

由上面三個規則，我們可以看出：

$$\begin{aligned} \sum(X-C)^2 &= \sum(X^2 - 2CX + C^2) \\ &= \sum X^2 - \sum 2CX + \sum C^2 \\ &= \sum X^2 - 2C \sum X + NC^2 \end{aligned}$$

習題一

1. 試寫出三種連續變數及三種間斷變數。
2. 試分別寫出二種名義變數、次序變數、等距變數和比率變數。
3. 假設有 5 位小朋友體重分別為：
14，13，15，11，16（公斤）
試問：
 - (1) X_3 ， X_5 分別為多少？
 - (2) 計算 ΣX ， $(\Sigma X)^2$ 和 ΣX^2 。
 - (3) 計算 $\Sigma X/N$ ， N 為人數。
 - (4) $\Sigma X/N$ 一般稱為什麼？