2.1 資料的形態

對一組統計資料加以彙總或描述之前，應先區別**資料的形態**(type of data)，因為不同形態的資料其所採用的彙總方法（統計方法）不同。

統計資料一般是由一個或多個變數之值所組成，所謂**變數**(variable)乃是一種具有不同值或結果的特徵之衡量，凡一切可計量的特徵皆稱為變數，而變數的量（或值）稱為**變量**(variate)。例如：年齡為一變數，其變量可能為 1 至 120；另外，性別、所得、長度、物價等等，亦皆為變數。

統計資料依據其變量的特性，而可分為**量的資料**(quantitative data)與**質的資料**(qualitative data)。茲分別說明如下：

質的資料亦稱屬性資料，係依據資料的屬性或類別之尺度來區分的資料，故又稱為**類別資料**(category data)。

例如血型、等級、膚色、教育程度、性別、國別、地區別、良品或不良品等資料皆屬之。由於統計學所處理的資料皆為數字性的，故欲對質的資料進行統計分析之前，必先加以量化。例如我們可用 0、1 表示男女性別，用 0、1、2 表示黃、黑、白種人等等。

量的資料亦稱為屬量資料，係依據數字尺度所衡量得出的資料。

例如年齡、身高、體重、分數、意外事件次數、溫度、速度等等皆屬之，它們皆可用數字表達。

量的資料一般又可區分為**間斷資料**(discrete data)與**連續資料**(continuous data)。間斷資料其特性是**可計數的**(countable)，且具有一最小的計數單位，而相鄰兩個最小計數單位的間斷資料之間存有**間隙**(gap)。例如某班級的人數即為間斷資料，其最小計數單位為一個人，此班級可能有 50 人，但絕不會有 50.5 人。間斷資料的其他例子包括：車禍件數、產品良品個數、籃球得分數等等。至於連續的資料一般係由**量測**(measure)而得，這是不可計數的，且任二個連續資料之間尚存有無限多個值，亦即不存在有間隙。例如人的身高即為連續資料，身高 175 公分與 176 公分之間尚存在許多其他值，如 175.1, 175.25, 175.365 等等。一般而言，凡屬於物理基本單位所得的資料皆屬連續資料，例如公克、公分、秒（CGS 制）分別代表重

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING

量、長度與時間的資料；此外，由其所衍生的衡量單位之資料亦屬於連續資料，例如面積、密度、濃度、速度等等。

例 2.1

請說明下列資料屬於質的資料或量的資料？若為量的資料，請指出其為連續或間斷的資料。

(a) 最多大學生追劇的前五名。
(b) 每月接到的疑似詐騙電話通數。
(c) 20～30 歲的人每天玩線上遊戲的時間。
(d) 進入 MLB 季後賽的球隊。
(e) 黃金的價格。

解 質的資料：(a)、(d)。
量的資料：(b)、(c)、(e)；其中 (b) 屬於間斷資料；(c)、(e) 屬於連續資料。

綜合言之，欲區分質的資料或量的資料，只要想想當別人問你某一變數時，你若以數字回答，則為量的資料；否則便為質的資料。例如：「你是哪種『血型』？」回答：「A」、「B」、「AB」或「O」；「你的出生地為何？」回答：「高雄」、「彰化」、「香港」……。因此，血型與出生地皆屬於質的資料。同理，量的資料亦可依相同方式來判斷。最後要提醒讀者注意的是，質的資料雖經常用數字代表，但這些代表的數字並不能做某些數學運算。例如血型 A、B、AB、O 分別以 0、1、2、3 表示，但求這些數字的平均數（平均血型）並無意義。

有關資料的形態，歸納如圖 2.1 所示。

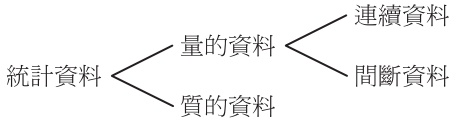


圖 2.1　資料的形態

2.2 次數分配

統計資料經蒐集完成之初，通常都是一群雜亂的數字，必須加以整理或彙總才能顯示資料所具有的特徵。最基本的整理或彙總方法是編製成統計表或統計

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING

圖，而製表或繪圖之前，一般皆先將其歸類分組。此一步驟所進行的工作，即是本節所要介紹的次數分配之主題。

2.2.1 次數分配的意義

所謂**次數分配**(frequency distribution)係將資料依數量大小或類別而分成若干組，並計算各組的資料個數（次數），由此可顯示資料分布的情況。次數分配一般可用圖或表的形式表示，因而有次數分配圖與次數分配表之分。本節先討論次數分配表，下一節再討論次數分配圖。

次數分配表亦簡稱為次數表，包括簡單次數表與分組次數表。所謂簡單次數表係將變量相同者（或屬於同一類別者）逐一歸併彙總而得者，適用於質的資料或規模不大的間斷資料。所謂分組次數表則為將全部變量依其大小次序，劃分為若干段落（分組），以每一段落為一**組**(class)，由此所編製的表，它適用於連續資料或規模較大的間斷資料。

以分組次數表的形式所表示的統計資料，稱為**分組資料**(grouped data)。雖然資料一經分組後會喪失許多原有的細節，但卻可讓我們了解原始資料所蘊含的規律性及其分布的情況，因此具有彙總與簡化原始資料的功用。

2.2.2 次數分配表的編製

簡單次數表的編製較為簡單，而分組次數表的編製通常很複雜。以下我們先談分組次數表的編製步驟，然後再以實例來說明此二種表的編製。

1. 分組次數表之編製步驟

分組次數表的編製雖較為複雜，但若能依循下列的步驟，將可使得編製的工作更為順利。

◢ 步驟 1：排序

將原始資料中的各觀測值依大小順序排成一序列，此一步驟的完成有助於往後編表的工作。

◢ 步驟 2：求全距

全距(range)係指資料中最大數值與最小數值之差，通常以 R 表示，亦即：

$$R = \text{最大值} - \text{最小值}$$

著作權所有 侵害必究

華泰文化 HWA TAI PUBLISHING

步驟 3：決定組數及決定組距

此步驟即在決定將全距劃分為多少組（段）及決定**組距**(class interval)之大小（即決定每一段的長度）。組數之多寡應視資料量的多少而定，一般的法則是：當觀測值個數 n 在 50～99 時，以 5～10 組為宜；當 n 在 100～249 時，以 7～12 組為宜；當 $n \geq 250$ 時，以 10～20 組為宜。組數之多寡乃與組距成反比，亦即在同一全距內，組距愈小，則所劃分組數愈多。以組距除全距即可得大約的組數（組數必須為整數），或者以組數除全距則可得組距（此為本書所採用者）。令 k 代表所決定組數，d 代表組距，則 d, R, k 之間的關係如下：

$$d = \frac{R}{k} \left(\text{組距} = \frac{\text{全距}}{\text{組數}} \right)$$

根據上式所算出的 d 可能為小數，但請記住一原則，組距的有效位數應與資料之數值的有效位數相同，故可做捨入，但不能比未被捨入的值還小。此外，為了方便起見，亦可取較易處理的組距（參閱後面實例之說明）。

步驟 4：定組限

組數與組距皆已決定之後，便應確定每一組之上、下界限，此即**組限**(class limit)。每一組皆有兩個組限，數值小的稱為**組下限**(lower class limit)，數值大的稱為**組上限**(upper class limit)。確定組限時應記住一個原則，即組限的小數位數應與觀測值之小數位數相同。又組下限 (L)、組上限 (U) 與組距 (d) 三者之間的關係如下所示：

$$d = U - L + (\text{最小計算單位})$$

所謂最小計算單位是指觀測值之有效數字的最小單位，例如觀測值為 $2.03, 3.12, 2.47, \cdots$，則最小計算單位為 0.01。此時若取 $L = 2.10$，$d = 0.50$，則 U 應為 $0.50 = U - 2.10 + 0.01$，即 $U = 2.59$。又，其下一組應為 2.60～3.09，餘此類推。

由此可知，只要最小一組組下限及組上限決定後，隨即可決定出其他各組的組限，直至最後一組編列完成。

著作權所有 侵害必究

華泰文化 HWA TAI PUBLISHING

步驟 5：歸類與劃記

當組限確定後，即可將每一觀測值審慎分別歸入其所屬之組內，並用記號「正」或「卌」劃記。切記不可有重複，亦不可有所遺漏。

步驟 6：計算次數

歸類劃記完畢之後，即計算各組次數，稱為**組次數** (class frequency)，並登載於整理表之最後一行；同時將各組次數加總，求其總次數，惟總次數應與原始資料之觀測值全部個數相同，否則即為錯誤。

2. 次數分配表編製實例

茲以兩個實例來說明如何編製次數分配表，其中第一個例子為間斷資料之簡單次數分配表，可不必依前述的編製步驟進行；至於第二個例子為連續資料之分組次數表，其編表工作較為複雜，我們將依循前述之編製步驟來說明。

著作權所有 侵害必究

例 2.2

開心公司 30 位員工 5 月份遲到的天數紀錄如下，請編製次數分配表。

1, 3, 1, 1, 0, 1, 0, 1, 1, 0, 2, 2, 0, 0, 5
1, 2, 1, 2, 0, 0, 1, 6, 4, 3, 3, 1, 2, 4, 0

解 簡單次數分配表可採列舉式的分組方式，即按數值大小依序排列。由於員工遲到天數從 0 到 6 天，故共分為 7 組，然後再計算各組次數，最後計算總次數，參閱表 2.1。

表 2.1 開心公司員工 5 月份遲到天數次數分配表

遲到天數	劃記	員工人數
0	正下	8
1	正正	10
2	正	5
3	下	3
4	丁	2
5	一	1
6	一	1
總計		30

華泰文化 HWA TAI PUBLISHING

例 2.3

某工會共有 48 位會員，其去年平均月收入如下表（單位：千元）。

81,	52,	76,	62,	79,	62,	72,	31,	71,	32,	60,	73
40,	40,	59,	39,	58,	38,	90,	49,	52,	59,	65,	28
83,	48,	68,	60,	39,	69,	54,	75,	42,	72,	52,	93
58,	81,	58,	53,	56,	58,	77,	57,	72,	45,	88,	61

解 (a) 排列：依數值大小重新排列，如下所示：

28,	31,	32,	38,	39,	39,	40,	40,	42,	45,	48,	49
52,	52,	52,	53,	54,	56,	57,	58,	58,	58,	58,	59
59,	60,	60,	61,	62,	62,	65,	68,	69,	71,	72,	72
72,	73,	75,	76,	77,	79,	81,	81,	83,	88,	90,	93

(b) 求全距：最大值為 93，最小值為 28，故全距：

$$R = 93 - 28 = 65$$

(c) 定組數及組距：由於總觀測值個數 $n = 48$，故取組數 $k = 5$。接著計算組距 $d = \frac{R}{k} = \frac{65}{5} = 13$；為便利分組，組距往往取 5 或 10 的倍數。就本例而言，取 $d = 15$（需大於 13）較為方便。

(d) 定組限：由於觀測值皆為整數，故組限亦取為整數。就本例而言，最小值為 28，最大值為 93，故取第一組之下限為 23，最後一組（第 5 組）之上限為 97。準此，由 23 為起點，每隔 15 為一組，計可分為 23～37，38～52，53～67，68～82，83～97 等 5 組，參閱表 2.2。

(e) 歸類和劃記：將 48 位會員月收入分別歸入所屬的組內，並在各該組內劃記，參閱表 2.2。

(f) 計算次數：將劃記欄中各組的記號個數用數字記載於次數欄中，然後將各組之次數加總，可得總次數 $n = 48$，與原始資料總人數相符合。

表 2.2　48 位會員月收入之次數分配

組限	劃記	次數
23～37	下	3
38～52	正正丅	12
53～67	正正正一	16
68～82	正正下	13
83～97	丌	4
總計		48

著作權所有 侵害必究

華泰文化
HWA TAI PUBLISHING

3. 編製次數分配表之原則

前述的六個編製步驟中，由於步驟 3 與 4 涉及組數、組距與組限之選取，故不同的人所做的選擇也許不一樣。由此可知，次數分配表的編製並無唯一正確的答案，最重要的是要能使所編製出來的次數分配表發揮其提供有用資訊的功用。雖說組數、組距與組限之選取並無唯一的結果，但皆須符合下列幾個原則：

(1) 關於組數

次數分配表旨在顯示一組資料的集中情形與分布狀態，且為便於計算與運用，故組數不宜太多，即其組距不宜過小；但為保持原始資料的確實性，組數亦不宜太少，即其組距又不宜過大。在前述編製步驟 3 中我們曾提及幾個參考原則，但實際選取時又當如何呢？我們或可利用**史塔基法則** (Sturge's rule) 作為客觀的判定；此法則係一經驗法則，其公式如下：

$$k = 1 + 3.3 \log n$$

式中的 k 代表組數，n 表示觀測值個數。

(2) 關於組距

組距的選取除了考慮前述的規定外（參見步驟 3），一般皆選用相等的組距。使各組的組距相等，此舉具有下列的優點：

① 便於比較各組的次數。

② 便於使用。

③ 便於計算各種統計量數（參見第 3 章）。

組距相等雖具有這些優點，但在某些特殊的分配上，基於事實上的需要，可採用不等組距。譬如說，某地區的人口依其所有財富而編製次數分配表，若取相等的組距，則組數必多至數百甚至數千（例如，1 萬元為一組，財富達 1 千萬元時，即有 1 千組）。但若取不等組距，0～50, 50～100, 100～500, ⋯, 100,000～500,000（單位：千元），則不再有上述的困擾。

(3) 關於組限

有關組限的選取，其有效數字位數必須與觀測值之有效位數相同；此外，分組的結果必須能將所有的觀測值皆涵蓋在裡面，亦即第一組的下限必須小於最小的觀測值，而最後一組的上限必須大於最大的觀測值。另外，盡可能地避免採用**敞開組** (open class)，如此才便利某些統計量數的計算。所謂敞開組係指第一組的下限與最後一組的上限並無界定，例如：37 以下、83 以上等等，都是敞開組的例子。

著作權所有 侵害必究

華泰文化 HWA TAI PUBLISHING

綜合言之，編製次數分配表的工作或可說是一種藝術，它並無唯一確定的答案。近年來由於電腦已普遍地運用於各種統計方法上，因此可藉助電腦之迅速處理能力，輸入各種不同的組數、組距與組限之選擇值，然後再從中選出最符合個人研究目的的次數分配表（參見第 2.5 節 Excel 的範例）。

2.2.3 相對次數分配與累積次數分配

1. 相對次數分配

所謂**相對次數** (relative frequency)，係指將各組次數除以總次數。若以 f_i 表示第 i 組的組次數，n 表示總次數（等於 $\sum f_i$），則將各組之相對次數依次數表的方式列出相對次數表，則稱為**相對次數分配表** (relative frequency distribution table)。相對次數通常以小數或百分比表示，其功用在於能顯示各組次數對總次數所占的比重，並能用以表示兩個或更多個次數分配同組間之合理的比較。此外，各組之相對次數的總和應等於 1（若以百分比表示則為 100%），參見例 2.4。

$$\text{第 } i \text{ 組相對次數} = \frac{f_i}{n} = \frac{f_i}{\sum f_i}$$

2. 累積次數分配

所謂**累積次數** (cumulative frequency)，係指將各組次數依次累加；列出累加次數的表即稱為**累積次數分配表** (cumulative frequency distribution table)。累積次數分配之所以重要，乃因為在許多情況下，我們所感興趣的可能不是落在某一組內的觀測值個數，而是落在某一特定值之上或之下的觀測值之個數。

累積次數依累加的方式不同，一般有以下累積與以上累積之分。以下累積乃是由數值較小的組之組次數，依序累加至數值較大的組之次數而得；以上累積則相反，即由數值較大的組之組次數，依序累加至數值較小的組之累積次數；參見例 2.4。

例 2.4

參考例 2.3，試就表 2.2 編製其相對次數分配與累積次數分配。

解 表 2.2 之相對次數與累積次數（包括以下累積與以上累積）分別計算於表 2.3 中：

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING

表 2.3　表 2.2 之相對次數分配與累積次數分配

組限	次數 (f_i)	相對次數		累積次數	
		$\frac{f_i}{n}$	$\frac{f_i}{n} \times 100\%$	以下累積	以上累積
23～37	3	0.0625	6.25	3	48
38～52	12	0.25	25.00	15	45
53～67	16	0.3333	33.33	31	33
68～82	13	0.2708	27.08	44	17
83～97	4	0.0834	8.34	48	4
總計	48	1.00	100		

由表 2.3 可知，48 位會員月收入落在 53～67 千元之區間的比重最大，約占 $\frac{1}{3}$；根據以上累積次數知，月收入 68 千元以上的會員人數有 17 人，而由以下累積次數可知，月收入 67 千元以下的會員人數有 31 人。

最後，本節所述的相對次數與累積次數之觀念，亦同樣地適用於簡單次數分配表的情況。讀者可試著將表 2.1 之資料，依相同的方式求其相對次數分配與累積次數分配。

2.3 次數分配的圖示法

次數分配圖乃是一種統計圖，用以表示次數分配表中次數分配的情況。因此，它只是將次數分配中的統計資料由表格形式改以圖形來陳示，故欲繪製次數分配圖之前，必須先編好次數分配表。

次數分配圖的功用在於：使閱圖者僅花費少許時間便可對統計資料得到一明確的綜合性概念。此外，其亦與次數分配表一樣，具有比較的功用。次數分配圖的種類很多，基本上亦與資料的形態有關。常見的次數分配圖有：直方圖、多邊形圖、累積次數分配圖等三種。這些統計圖皆只適用於連續型的統計資料，至於間斷型的統計資料另有其他構圖方式，諸如：長條圖、面積圖等，本節將分別探討這些統計圖的意義、繪法及特性。

2.3.1 直方圖

直方圖 (histogram) 係以長方形面積大小，表示次數分配中各組次數的一種統計圖。此種圖形的橫座標代表各組的**組界** (class boundary)，而縱座標則表示組次

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING