

統計學之意義

黃登源

國立臺灣師範大學數學研究所

數據與由數據得結論在我們的日常生活裡是常遇到的事。爲了評估數據的健全性我們必須發展一套具體而嚴格方法以得到數據，並以嚴謹的態度使用它們。若要由數據得到適切而有意義的結論，數據必須由小心計畫及小心執行調查得到。

爲了使數據對我們有意義，我們需有規則地獲得某一班學生中每個學生的身高、體重，及年齡。此種調查，或數據之蒐集，表現組成該班個體群的特徵。若該班所有個體之集合表現在調查中所有有興趣個體，則該班稱爲母群體。不過，若母群體由更廣之個體群所組成，如該大學中所有學生，該班學生僅爲一部份，或稱該母群體之樣本。在母群體中某特徵之完全調查稱爲普查，但若僅調查母群體之一部份稱爲抽樣調查。

母群體可由人類特徵，農場面積，城市居民等組成。母群體由數據分類及描述，例如我國農業可由每年產量描述。數據用來描述母群體之特徵而由母群體中所有個體所表達出來的數值稱爲參數。參數表示母群體之事實。由樣本導出之數據可用來估計或評估參數值者稱爲統計量。在某學期中學生的平均分數爲估計該學生在大學四年中的平均分數；在此例中，該學期平均分數爲一統計量，而且它估計全部八學期平均分數。八學期平均分數爲參數；母群體爲全部在八學期中該學生所修課程分數。

針對這些平均數、百分比、比例等有一種科

學方法稱爲統計學專門研究統計方法的應用與發展。此種科目主要關心下列項目：

1. 設計或計畫實驗及抽樣調查，
2. 由實驗及抽樣調查蒐集之數據獲得資訊，
3. 利用由實驗或樣本之事實對母群體作推論。

上述可做爲統計學之定義。我們可綜合敘述如下：統計學之意爲蒐集，分類及評估事實做爲推論之基礎。詳細一點可敘述爲“由不完全資訊所需精確智識之技術主體，對蒐集、組織、分析、解釋及陳述可以數據形式表示之資訊”。

因爲所有人類以各種不同方式牽涉到統計學，所以必須有興趣於對他們所從事的課題有用的統計技術或工具。大部份的人有興趣於瞭解或使用統計學的工具，而不是發展新的統計方法。

當應用統計方法於實際問題時有些假設可能未被驗證。統計方法的應用總是牽涉到某種程度的主觀性，爲了適當使用所考慮之統計方法，對此主觀性應該常常發出疑問及評估。

一、資料蒐集

假如我們收集某大學學生的電話號碼，可得一堆數據，若僅為獲得一大堆數據，這種收集法是沒用的，若這些電話號碼是五年前的紀錄，則更是沒有用。

假設五年前電話號碼有身高與體重之紀錄，而無目的與用途，這堆數據仍為一堆沒有意義的數據。另一方面，若這些身高與體重用以表示某種事實，則這些數據便可稱為資料。資料可定義為可得結論之事實。資料與不能得結論之數據的區別是很重要的。假如我們僅需一堆數據的話，高速計算機可幫我們打出一大堆。資料收集應考慮下列幾個觀點。

為什麼收集資料？

對於已知項目必須有收集資料之理由，而不是由於容易收集或是某人、某地、某時可用於實驗研究或調查之目的。

何種資料應收集？

在開始研究之前，我們必須決定所欲收集資料之性質與特徵。

資料如何收集？

決定收集資料方法必須考慮所做研究之條件。記錄資料方法應小心，不小心記錄將降低資料之價值，甚至使之無用。

何處及何時收集資料？

在某些調查中僅有一個可能地點作調查，此時沒有困難去選擇地點，不過仍需決定舉行調查時間。在其他調查中通常有很多地點舉行調查，必須選擇適當地點。

何人收集資料？

若許多人牽涉資料之收集時必須精確決定何

人記錄何種特殊觀測使重要項目不至省略。若可能時為標準化結果最好全由特定個人或測度器械測量或觀測。

已收集資料已完全描述否？

必須完全描述可用資料於可用形式。例如五年前電話數據所附身高及體重測度，若無適當標以符號，則這些僅是一堆數據而已。

資料如何處理與貯藏？

資料收集以後必須計劃貯藏直到下一步驟開始。而且在資料利用及完全研究後必須決定對該資料如何處理。假設所有資訊在該研究中已被收集，資料必須毀掉。資料若無進一步使用價值不應佔據有價值空間。

若資料保留為將來使用，我們應確實存在安全、乾燥、防火處。

二、統計學之基本概念

統計學的功能可大致分為(1)降低大量資訊到可管理的大小；(2)提供基於資料對未知量之估計；(3)決定支持假設之程度及(4)提供接受假設做為決策或採取行動風險之估計。

統計學一般分為敘述與推論統計學。敘述統計學為任何數據資料之處理，不牽涉由樣本對母群體之綜合。即，當我們有興趣於描述一群元素（人類，考試分數等），我們使用敘述統計學。當我們做綜合、預測、估計或其他達到面臨不確定性之決策時，我們使用歸納統計學，其過程稱為統計推論。

例1：送小孩上幼稚園的家庭平均收入為一萬元。（敘述統計）

例2：在視力敏銳度實驗中有24位正常視力及

6位色盲。有正常視力者機率很高。(歸納統計)

在統計學上有四個重要名詞如下：

- (1)母群體：在某一特定範圍內至少有一共同特徵之所有事物。例如在某種考試中學生之所有考試分數，在某診所病人的所有心臟舒張血壓讀數及在某一社會團體的人的所有宗教聯盟。
- (2)樣本：母群體之一部份。通常經由隨機選擇樣本作爲代表母群體。
- (3)資料：數據或度量收集作爲樣本觀測之結果。例如在某種課程中學生的考試分數。
- (4)統計量：依照特殊程序由樣本中所收集原始資料之使用所得之數據。例如，在新陳代謝診斷中血液膽固醇高於年齡可容忍的極限病人的百分比爲一統計量。

我們知道統計學之目的之一爲降低大量資料爲更可管理及可解釋的形式。次數分配及圖形表示的使用爲這種過程的一個步驟。

對於任何一堆資料，第一步驟在組織次數分配以決定分數或度量值的全距。全距爲最大與最小分數之差。全距可幫助決定在次數表中使用區間單位之大小。在次數表中之區間可將原始資料分爲更少種類。區間大小總是與資料有關。例如，在血壓中，個體度量單位爲毫米水銀柱高，若以 50 毫米爲區間，則由於太大以至無意義。血壓以 5 毫米的區間爲適當。另一方面，在某社區的收入分配可能最好一萬元或以上表示。我們建

立次數分配必須決定適當區間數。

在表一中取 15 個區間可能是合理的數目，因爲它可給我們區間大小有意義解釋沮喪分數。

我們可得下表：

表二 精神病樣本標準沮喪刻度次數分配

區 間	分 割	次 數
81-83	1	1
78-80	1	1
75-77		0
72-74	1	1
69-71		0
66-68	11	2
63-65	1	1
60-62	1111	4
57-59	11	2
54-56	1111 111	8
51-53	111	3
48-50	1111	4
45-47	111	3
42-44	1111	5
39-41	11	2
36-38	11	2
33-35		0
30-32		0
27-29	1	1
總 和		40

注意在我們所造的次數分配圖，有 19 個分離區間，雖然最初選 15 個區間，由於事實全距 54 被 15 個區間除得區間大小爲 3 到 4。因此，在次數分配圖上稍爲多於 15 個區間。在次數分配圖中檢查所組織的資料，注意 40 個分數中主要落在分配的中下部份；即，大部份精神病人有沮喪分數在 42 及 62 之間。在此方法次數分配將

表一 以標準刻度衡量精神病人沮喪的原始分數

56	78	62	37	54	39	62	60
28	82	38	72	62	44	54	42
42	55	57	65	58	47	42	56
56	56	55	66	42	52	48	48
47	41	50	52	47	48	53	68

40 個數安排在更有意義及更容易分析的方法。

做次數分配表的步驟如下：

1. 決定全距
2. 習慣上將全距以 15 或小於 15 除之。任意大於 15 的數將使所得的表難讀或解釋。最重要的是區間大小的選擇與資料有關。
3. 列出區間由一行之底端開始。令最低區間開始的數為區間大小的倍數。
4. 記錄次數。
5. 將次數總數記在底下。

以前例說明區間大小取法。由 54 被 15 除介於 3 及 4 之間，不過區間大小一般為整數，所以使用 3 或 4。更方便的是使用奇數做區間大小因每個區間中點為整數。因此如 3 及 5 為常用區間大小。在此情形我們將選擇 3。現在區間大小決定，我們能夠開始做區間刻度。一個決定開始的方法為允許最低區間以區間大小的倍數為起點；即最低區間的最低數應為區間大小乘以其他數。此數必須與原始數據中最低值一樣或更小。因為在我們所使用的例中，最低分數為 28，最低區間應為 27 至 29 最適當。在找到最低區間之後，區間刻度在一行中簡單列出區間直到最高區間。最

高區間為區間大小倍數且含該堆值中之最大者。

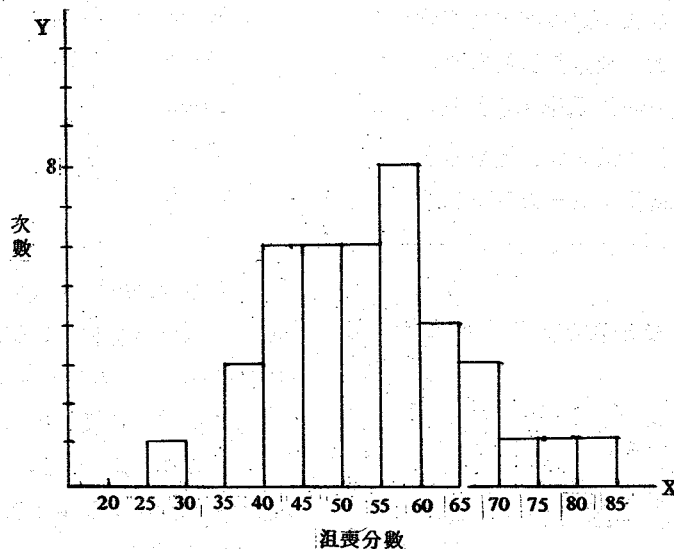
在決定區間及每個區間的次數之後，另一行可以記錄次數，這樣可以算出累積次數或在該區間個案數加上在該區間之下（或上）的個案數。找到這一行的每個區間的值，注意該區間上限的分數，例如最底下區間（27-29），在上限下發生次數為 1，由底下往上算第四個區間（36-38），累積次數為 3，因 3 個分數低於 38。

我們將前述資料畫成長條圖如下圖。

在劃長條圖時將會發現如何處理 65，60 及 55 的問題；即要將分數 65 劃在區間 60-65 或在區間 65-70？為此問題我們必須建立真正極限。

真正極限開始低於低極限 0.5 單位及高於高極限 0.5 單位；例如區間 65 及 69，開始在 64.5 止於 69.5，因此我們可以毫不混淆的將 65 放在區間 65-69。區間 40-44 開始在 39.5 止於 44.5。這些數 34.5，39.5，44.5 等稱真正極限。長條圖以簡便方法表示 35-39，40-44 等以代替真正極限 34.5-39.5，39.5-44.5 等。

上述方法為分析資料之一種工具，還有許多方法也可用來獲取有用資訊。本文僅介紹常用方法以為參考。



沮喪分數長條圖