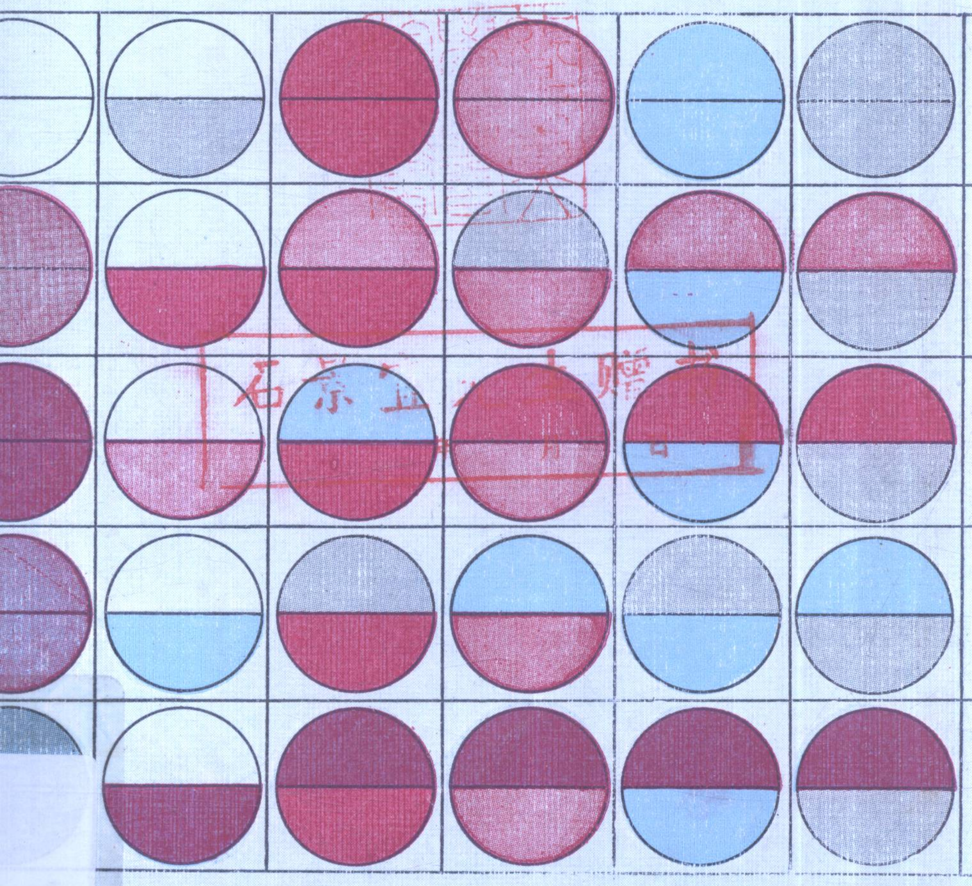
統計學

上冊

編著者

洪春雄 林昇平



東華書局印行

統計學

上　册

洪春雄
林昇平　編著

東華書局印行

中華民國六十 年五月 初 版
中華民國七十四年二月 五 版

大學用書 統計學（全二册）

上册

定價 新台幣九十元整

（外埠酌加運費滙費）

編著者 洪春雄 林昇平
發行人 卓鑫淼
出版者 臺灣東華書局股份有限公司
臺北市博愛路一〇五號
電話：3819470 郵撥：6481
印刷者 中臺印刷廠
臺中市公園路三十七號

行政院新聞局登記證 局版臺業字第零柒貳伍號

（59012）

統　計　學

上册目次

章節　頁碼

第一章

緒　　論

1.1 統計學的性質 (Nature of Statistics)

統計學家主要是在處理科學觀察上可能出現的事件。他可能對某一交叉路口發生意外事件的次數、擲一個骰子所出現的點數、或在某一次化學試驗中留下的沉澱量等感到興趣。因此，他經常處理的是**可點計或數字上的量** (counts or numerical measurements)。凡是以原始搜集形式記錄下來未經過整理的資料，謂之**原始資料**或**初級資料** (raw data)；反之，現成的已整理過的資料，謂之**次級資料** (secondary data)。統計學家可能喜歡用多種不同方法，來描述大量原始資料的現象，以提供所需的情報 (information)；或喜歡以一部分資料的片面知識，對整個資料作判斷及下結論。

統計學 (statistics) (註) 是一門處理用於搜集、表達、分析與解釋資料的方法的科學。有關統計方法應用的所有問題，均可分別屬於**敍述統計學** (descriptive statistics) 或**推論統計學** (inductive statistics) 的範圍。凡對資料之處理，而導致有關一大群資料之預測或推論時，謂之**推論統計學**或**統計推論**。反之，若僅限於手中資料，而對一大群資料不作進一步之分析與歸納時，謂之**敍述統計學**。

(註) Statistics 有二種意義：①數字資料的**收集**及②處理資料的**方法**。本書採用第二種意義。見 Richard P. Runyon and Audrey Haber: Fundamentals of Behavioral Statistics, © 1967, Addison-Wesley Publishing Company.

爲了能淸楚區分敍述統計學和推論統計學起見，玆就最近三十年來每一年七月份在某地區總雨量之一些數値爲例來說明。任何表示資料的數値，如最近三十年間七月份的平均雨量、或最近三十年內最乾燥之七月份的平均雨量等，都是敍述統計學範圍內的數値。由這三十年所獲得的資料，我們並不想對這三十年以外的任何一年的雨量，作任何說明。若在最近三十年間這一地區七月份的平均雨量是 1.3 吋，而且我們欲對下一個七月有所陳述，如預期下一個七月的雨量在 1.2 吋與 1.4 吋之間，我們這樣推論，就是推論統計學的範圍了。

推論統計學的歸納是相當不確定的，乃因我們所處理的僅是研究對象之部分片面資料而已。爲應付不確定的現象，就必須瞭解**機率理論**。本書將作最低限度的討論。機率理論最好是運用**集合** (set) 符號。因此，第二章將先介紹集合及集合的特性。如要精確的證明，則統計課程是需要數學的。然而，由於省掉某些需要微積分的證明，所以只要具備高中代數的基本觀念，就可學習很多統計方法。

本書爲使科學、哲學、商學、農學或醫學等各方面的學者都能獲益而寫。無論在那一方面應用，其搜集與分析的基本技術是相同的。譬如，化學家作三個變數的試驗，以及測量所欲產出物的數量。最後用統計方法分析其結果。這和試施三種肥料，測量作物的生產量，或和使用三部相同機器所產生劣品數量等結果的分析是相同的。很多統計方法，最初均由農業上的應用所導出，而後證明在其他方面應用也有同樣的價值。

統計學可能被用來敍述一群資料的集中趨勢，以及提供與此集中數値 (central value, 或中値) 相對變動或差異的量數。兩個變數的關係，可用**相關係數** (correlation coefficient) 的數値來衡量 (註)。

(註) 均屬於敍述統計學的範圍。

統計學不僅可使我們作預測和檢定假設，而且可決定其判斷的正確性(註)。今日每一個進步的產業，均聘請統計學家來指導他們的品質管制方法，以及協助建立良好的廣告和產品的銷售計劃。在商業上，統計學家負責時間數列的分析與指數的編製。

的確！如果運用適當，統計學將是一個非常有效的工具。反之，若誤用了統計方法，就會導致錯誤的結果。故在某些特定情況下，必須謹慎應用最正確和最有效的方法，使能自可資利用的資料獲得最大情報。

用來分析一群資料的方法，與用來搜集情報的方法是有高度關聯的。因此，任何一種研究，自"計劃"開始設計迄"結果"的分析與解釋，均需與統計學家磋商。

1.2 統計學史 (History of Statistics)

敍述統計學方面的歷史應追遡自有人類開始。在早期聖經時代，統計學乃用以提供有關賦稅、戰爭、農作物收穫、甚至競技等方面的情報。自從十六世紀以後，因有機率理論之產生，推論統計學始得以突飛猛進。事實上，目前的統計學大部份是最近四百年來多數人在各方面努力研究的成果。

也許是由於人類對賭博的貪慾，乃導致早期機率理論的發展。在增加贏錢的努力下，賭徒訪請數學家就各種機遇的賭博，提供最有利的對策。當時提供這些答案的一些數學家，有巴斯卡 (Pascal)、萊布尼茲 (Leibnitz)、法莫特 (Fermat) 及詹姆斯・百努里 (James Bernoulli) 諸人。

(註) 屬於統計推論的範圍。

一七三三年廸莫弗爾 (De Moivre) 提出了很多推論統計學所依據的常態分配 (normal distribution) 方程式。又爲紀念高斯 (Gauss. 1777-1855), 這與鐘形分配 (bell-shaped distribution) 相同的分配往往就被稱爲**高斯分配** (Gaussian distribution), 因爲高斯就相同數量重覆測量的誤差研究中, 導出了常態分配的方程式。統計學在天文學上的應用, 拉普萊斯 (Laplace) 的著作, 在當時亦是極享盛名的。

十九世紀, 比利時的統計學家阿道夫・奎特菲特 (Adolph Quttelet, 1796-1874), 致力將統計方法 用於教育學與社會學的領域中。奎氏是指出由某一方面的研究導出之統計技術, 亦可適用於其他方面的第一位統計學家。

將統計學應用到社會科學, 貢獻最大的或許要算法蘭斯・高登爵士 (Sir Francis Galton, 1822-1911)。他最有名的貢獻, 乃將統計學應用於遺傳學和優生學的領域。同時, 高登並提出百分位數 (percentiles) 致被讚譽。數學家皮爾遜 (Karl Pearson, 1857-1936) 與高登共同發展廻歸和相關的理論。皮爾遜對統計學貢獻良多, 目前的抽樣理論大多是由他所提出的。

廿世紀初葉, 高謝 (William S. Gosset) 提出了多種方法以便根據少量資料作決策。由於他所服務的啤酒廠不允許其出版有助於其競爭者的文章, 他乃以"學生" ("Student") 的筆名發表著作 (註)。之後, 在小樣本理論 (small sample theory) 方面的貢獻更大。對小樣本之貢獻與試驗設計 (experimental designs) 方法之建立者, 就是費雪爵士 (Sir Ronald Fisher, 1890-1962), 他至今仍被視爲是本世紀最傑出的統計學家。

(註) 原文登在Biometrika 6 ,1-25(1908),標題爲"The probable error of a mean。"

廿世紀出了幾位著名的統計學家，他們之中大部份均仍繼續致力於統計學之新理論與應用的發展。自 1950 年以後，在多數的美國大專院校，統計學業已形成爲一門單獨研究的課程。且自 1955 年起，大部分的研究所也都增開較高深的統計學課程。而電子計算機 (electronic computers) 的應用，則是促進現代統計學發展的主要因素。

以上所述，乃統計科學的過去與目前狀況，至於未來我們將可看到很多新理論的發展。在數理統計學、機率論、遊戲理論、線型規劃、隨機過程 (stochastic processes) 以及實驗設計等多方面的研究是可預見的。現代的研究人員，無不認爲統計學乃是他們最有用的工具之一。

第二章

集合及機率

2.1 集合及子集合 (Sets and Subsets)

研究機率和統計學所需要的數學之基本觀念就是**集合** (set)。集合的定義是一群明確的事物（或客體)。例如，佛吉尼亞州的河流、英文字母的字、參議院的議員、以及銷售皮鞋每月的收入，這些均可構成集合。甚至，一條線段亦可視爲無限點的集合。在集合內的每一事物（或客體)，叫做集合的**元素** (an element of the set or a member of the set)。

通常，集合是用 A、B、X 或 Y 等的大寫英文字母表示，而 a、b、x 或 y 之類的小寫英文字母，則用來表示集合中的元素。爲要有一個完善定義的集合，不管集合的元素是不是既定的事物，我們對它需要有一些表示的方法。

有兩種表示集合的方法。第一個方法：假若集合之元素數目有限，則可將這些元素加以**排列** (list)，用逗點（註）分開，再用大括號括起來。因此，若集合 A 含有 2、4、6 和 8 的元素，則可寫成

$$A=\{2, 4, 6, 8\},$$

或集合 B，擲一硬幣之可能結果，可以寫成

$$B=\{H, T\},$$

此處的的 H 和 T 依次代表正面 (heads) 和反面 (tails)。

（註）逗點有"和"的意義。

第二個方法：一個集合可以一個**敍述或法則** (statement or rule) 來描寫集合元素具有的共同性質(註)。例如，令C代表世界上人口超過一百萬人的城市的集合。若x是C的任一元素，則可寫成

$$C=\{x|x\text{ 是人口超過一百萬人的城市}\},$$

上式讀作“C 是全部x 的集合,而x 是人口超過一百萬人的城市。”縱線條讀作“而……是” (such that)。同樣的，X 是英文字母中母音的集合，則可寫成

$$X=\{x|x\text{ 是英文字母中的母音}\}。$$

不管我們用概述方法 (rule method) 或用列舉元素來描述集合，均須依據手中的特定問題而定。如想列舉藍眼睛的人民做爲集合之所有元素，是相當困難的。另一方面，我們也無簡單的法則來概述下列集合

$$Y=\{\text{書本，狗，硬幣，戰爭}\}。$$

在集合符號中的$\in$，其意義是“爲……的一個元素”或“屬於” (“is an element of” or “belong to”)，而$\notin$的本意爲“不是……的一個元素”或“不屬於”(“is not an element of” or “does not belong to”)。若x是集合A的一個元素，而y不是，則可寫成

$$x\in A \quad \text{及} \quad y\notin A。$$

例 2.1 令 $A=\{2,4,6,8\}$ 及 $B=\{x|x\text{ 是可被 3 整除的整數}\}$。則 $4\in A$，$9\in B$，$8\notin B$ 及 $3\notin A$。

定義：假若兩個集合均恰好含有相同的元素時，則此兩集合相等。

(註) 參閱王義濡譯：有限數學之理論與問題 (上)，P. 50。

若集合 A 與集合 B 相等或相一致。例如 $A=\{3,4\}$ 及 $B=\{4,3\}$ 則屬於集合 A 的每一元素亦屬於集合 B，及屬於集合 B 的每一元素也必屬於集合 A，我們用 $A=B$ 表示此種相等的關係。然而，若集合 A 或集合 B 中，至少含有一個元素非兩者所共有時，則稱此兩集合不相等，可寫成 $A\neq B$。

例 2.2 令 $A=\{1,3,5\}$ 及 $B=\{3,1,5\}$，$C=\{1,3,5,7\}$。則 $A=B$，$A\neq C$ 及 $B\neq C$。

注意：元素之順序重新排列，並不改變該集合（註）。

> **定義：空集合** (null set or empty set) **是不含元素的集合。我們用 $\varnothing$ 表示之。**

若令 A 是體重超過五噸之人類的集合，則 A 一定是空集合。又若 $B=\{x|x$ 是 7 的非素因數 (nonprime factor)$\}$，由於 7 的僅有因數是素數 1 和 7，故 B 也必定是一個空集合。

若集合 $A=\{1,2,3,4,5\}$ 和集合 $B=\{2,4\}$。我們知道集合 B 的每一元素也是集合 A 的元素。集合 B 就叫做集合 A 的**子集合**或**部分集合** (a subset of A)。以符號表示，可寫成 $B\subset A$，此處 $\subset$ 意爲"是……的子集合"或"是含於……內"("is a subset of" or "is contained in")。

> **定義：若集合 A 内的每一元素也是集合 B 的元素時，則 A 稱爲 B 的子集合。**

（註）對一集合而言，我們不考慮其元素間的次序。且同一元素不容許在一集合中出現一次以上，若有重複，則將重複者删去。見樊以楨等編之集合函數與關係一書（東華書局出版）p.13-14。

依據此定義，任何集合均是它本身的子集合如 $A\subset A$。而一集合之任何一個子集合不是這集合的本身，則稱爲集合的**眞子集或眞部分集合** (proper subset)。因此，若 $B\subset A$ 而 $B\neq A$，則 B 是 A 的眞子集。

例 2.3 集合 $B=\{2,4\}$ 是集合 $A=\{1,2,3,4,5\}$ 的眞子集。而集合 $C=\{3,2,5,1,4\}$ 是 A 的子集合，但不是眞子集，因 $A=C$ 之故。

所有的集合是某一特定集合的子集合時，則此特定集合稱爲**廣集合或全集合** (universal sets)，通常以 U 表示之。實數 (real numbers) 可視爲廣集合。所有大專學生之智商 (I. Q.) 的集合可視爲廣集合，因某一特定院校學生的 I. Q. 就是它的一個子集合。

例 2.4 廣集合 $U=\{1,2,3\}$ 的全部子集合，即爲 $\{1,2,3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$，$\{1\}$，$\{2\}$，$\{3\}$，及 ϕ。

注意：一含有三個元素的廣集合，其子集合之個數是 $2^3=8$。故一般含 n 個元素之集合，有 2^n 個子集合。

子集合與其對應之廣集合，可用**范氏圖示法** (Venn diagrams) 說明之。在范氏圖上，令廣集合爲一長方形，而子集合是長方形內之圓形。因此，在圖 2.1 上我們可以看出 A,B 及 C 都是廣集合 U 的子集合。同時很清楚的表示下列關係；$B\subset A$；B 和 C 沒有共同元素；A 和 C 至少有一共同元素。

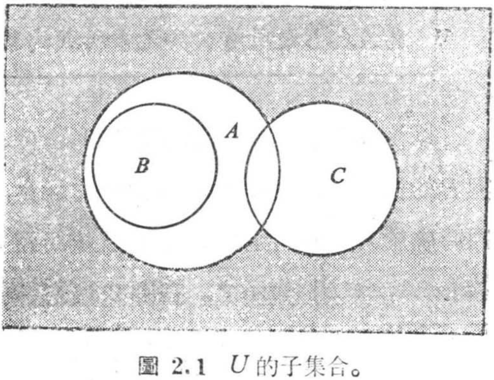


圖 2.1 U 的子集合。

有時爲了方便起見，常就圖形上之各部份劃線如圖 2.2 所示。在此我們以某一大學之學生作爲廣集合。子集合用某方向直線部份代表選讀數學的學生，而以另一方向直線代表選讀歷史的學生之子集合。重疊或交叉部份，代表數學與歷史兩者均選讀的學生，至於圖中未劃線部份表示未選讀數學或歷史的學生。

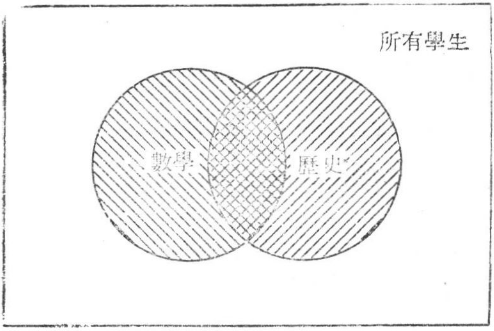


圖 2.2 用劃線表示的子集合。

2.2 集合的運算 (Set Operations) (註)

現在讓我們討論某些集合運算，據以作成新的集合。這些新的集合和已知的集合一樣，也是同一廣集合的子集合。

> **定義：兩集合 A 與 B 之交集 (intersection) 是一個集合，爲 A 與 B 之所有公共元素所成的集合。**

以符號表示，A 和 B 的交集可記作 $A\cap B$ （讀作"A 交 B" 或 "A 與 B 的交集"）。在 $A\cap B$ 集合中之元素，必須是這些（也只有這些）同時屬於 A 和 B 的元素 。這些元素可用**列舉法**或**概括法**——$A\cap B=\{x\,|\,x\epsilon A, x\in B\}$ 表示之。在圖 2.3 范氏圖上，劃線部份就是 $A\cap B$。

（註）本節的修改部份參閱東華書局出版之"集合函數與關係"一書。

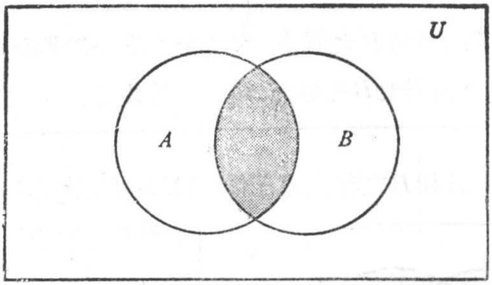


圖 2.3 A和B之交集。

例 2.5 若 $A=\{1,2,3,4,5\}$ 及 $B=\{2,4,6,8\}$；則 $A\cap B=\{2,4\}$。

例 2.6 若R 是全部納稅義務人的集合，S 是年齡超過 65 歲人的集合，則 $R\cap S$ 是所有納稅義務人年齡超過 65 歲的集合。

例 2.7 若 $P=\{a,e,i,o,u\}$ 及 $Q=\{r,s,t\}$；則 $P\cap Q=\phi$。亦即 P 和 Q 沒有共同元素。

> 定義：若 $A\cap B=\phi$，則A和B兩個集合互相分離；亦即A和B沒有共同的元素。

圖 2.4 范氏圖上表示的，是 A 和 B 兩個相離的集合。以灰暗不同的陰影面積分別表示兩集合A和B，我們發見沒有重疊的陰影面積來表示 $A\cap B$ 的集合。因此，$A\cap B$ 是空集合。

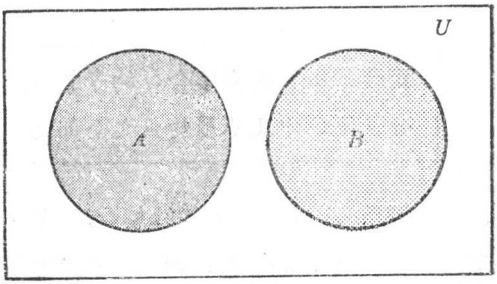


圖 2.4 分離的集合。

定義：兩集合A與B之聯集 (union) 是一個集合，其元素屬於A或屬於B或屬於A與B所共有。

以符號表示，A和B之聯集可記作 $A\cup B$ (讀作"A聯B"或"A與B的聯集")。$A\cup B$中的元素可以列擧或概括為 $A\cup B=\{x|x\in A$ 或 $x\in B\}$。圖 2.5 范氏圖上的陰影面積代表$A\cup B$集合之元素。

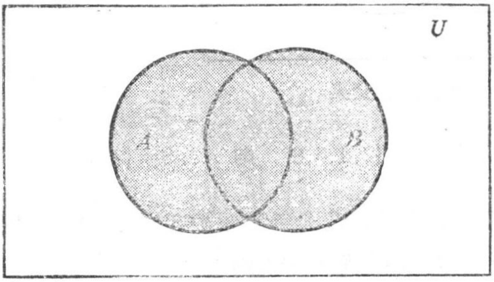


圖 2.5 A和B之聯集。

例 2.8 若 $A=\{2,3,5,8\}$ 及 $B=\{3,6,8\}$；則 $A\cup B=\{2,3,5,6,8\}$。

例 2.9 若 $M=\{x|3<x<9\}$ 及 $N=\{y|5<y<12\}$；則 $M\cup N=\{z|3<z<12\}$。

假如我們把某些工廠之員工視為廣集合U。令全部吸煙者形成一子集合。則所有不吸煙者作成之集合，也是U的一個子集合，後者稱為吸煙者集合之**餘集合** (complement of the set)。

定義：若A是廣集合U的子集合，就U而論，A的餘集合就是U中凡不屬於A之全部元素的集合。

我們用 A' 表示 A 的餘集合。A' 之元素可列擧或概括表示為 $A'=\{x|x\in U,\ x\notin A\}$。圖 2.6 范氏圖上的陰影面積就是集合 A' 之元素。