大專用書

統計學

（下冊）

統計學編輯委員會編著

主編：莊　晉　葉金田
周杰之　何丁舜

科教圖書出版社印行

總經銷　科技圖書股份有限公司

大專用書

統計學

（下冊）

統計學編輯委員會編著

主編：莊　晉　葉金田
周杰之　何丁舜

科教圖書出版社印行

總經銷　科技圖書股份有限公司

統計學

（下冊）

行政院新聞局局版台業字第1921號

編　著　者：編輯委員會　主編：葉金田　莊　晉
　　　　　　　　　　　　　　　何丁舜　周杰之

出　版　者：科教圖書出版社

社址：台北市南陽街十五之二號二樓

電話：314-8982　381-0575

發　行　者：周　杰　之

印　刷　者：龍泉印刷有限公司

地址：台北市寶興街八十巷二十九弄八號

電話：301-7461

特價：200元

中華民國七十二年二月修訂版

編者序言

在這科學昌明與進步飛躍的時代中，人類對於瞭解現在，預測未來，以解決實際問題的方法渴望益甚，因而統計學形成爲一種重要的工具科學，應用範圍日漸廣泛，無論在理論科學、應用科學、社會科學、企業管理、工業生產、生物醫學的研究等統計方法成爲必不可缺少的工具。在人們無時不接觸“統計”之結果，與運用統計方法和受統計決策影響的今日，使統計知識已有逐漸成爲現代國民人人必備的基本知識。

爲了提供統計知識的完整性，本書着重於介紹資料之蒐集（Collecting）、組識（Organizing）、分析（Analyzing）、顯示（Presenting）、解釋（Interpreting），及應用（Applicat on）等各項方法和技術，以便讀者能在不確定之情況下，從事推論和預測未來可能的演變，而作出最完善，最合理的決策。

本書爲了在大專教學的需求特按照教育部頒訂最新之大專統計學編輯原則與教學大綱編制章節，且參考中外有關書籍及歷屆高、普、特考試題，大學研究所，插班，轉學試題與各大專商學管理科系期中、期末考精選試題資料充實內容，取材及敘述均力求淺顯易懂，並在每一章節中都附有若干例題與習題俾使讀者實際應用各種方法。因此針對升學、就業、升等等各類考試，亦作了妥善、完備的安排。

編著者才疏學淺，又於倉促中脫稿付梓，雖作愼審校閱，但疏漏謬誤之處在所難免，尙祈碩學先進不吝賜教指正爲盼。

編 著 者　謹 識

民國七十二年二月

統計學

（下冊）

目　錄

第十一章　抽樣方法

第十二章　抽樣分配

附錄

第十一章　抽樣方法

11-1　抽樣的意義

前幾章我們介紹了基本的機率理論和隨機變數的機率分配，現在我們要根據這些原理，從事統計推論的工作，在第一章，我們已經提到統計學上的母體與樣本，爲了進一步認識母體與樣本，今舉例如下：

所有可能的觀測值所成的集合稱爲母體（Population)例如：

(1)某計程車行老板要分析他的每一部車子的耗油情形，此時母體指的是這家車行所有的車子，母體的元素爲這家車行的每一部車子。

(2)生產一批電子計算機組件，欲瞭解其品質情況，此時母體指的是所有這批組件，母體的元素爲這批組件的每一個。

母體的部份集合稱爲樣本。經由樣本，可取得有關母全體的情報。政府機構利用樣本，以取得國民健康、勞工失業等資料。工商企業利用樣本，以取得進料的品質、完成品的品質、新產品受消費者偏好等資料。一般人日常生活也常根據樣本資料作決策，譬如在水菓攤上試吃一個橘子，以決定買或不買；翻閱一本書的幾頁，以決定要不要把整本書讀完。

了解了母體與樣本後，我們可以知道，統計學最主要的是要推論母體的特徵值。例如：

(1)汽車組件檢驗員，自一批進廠的輪胎，抽取幾個檢驗，以決定這批貨是否允收。

(2)從去年度所有藥物傷害請求賠償案中，取一組樣本詳加分析，

以供防止藥物傷害的立法的參考。

(3)從所有請求緊急救濟的家庭，取一組樣本，以瞭解失業與小康計劃的成效。

(4)根據 1,000 家庭的儲蓄情形，以擬定國家的財政政策。

(5)有一電視公司，業務經理，極欲獲知在黃金時間所提供的節目的收視率，以利節目的策劃。他必須從那些擁有電視的人家，抽取若干戶，加以詢問調查。假若經調查某一地區的結果，對其節目反應頗爲良好，而經理便自認得意，那麼這份高興，顯然嫌太早些。因爲，他所調查的地區，可能是全國的某一地區而已。由於地區的環境及教育程度不同，收視情況，自然有異，也就是說，這個節目或許在其他地區不受歡迎。

以上所述的這些個案，均須有效地應用樣本資料來推論母體及解決問題，至於如何得這些樣本，則有賴於抽樣，因此抽樣乃是從母群體中，抽取一部分資料，加以整理，以便推測母群體的特徵，至於**普查**（Census）係指一在特定時間內，對某一事象所要調查群體內每一個子逐一加以查記，藉以觀察全部群體事實之眞象，若再作更深入的研究，則抽樣調查有如下列五個優點：

(1)抽查可節省人力和財力：如台灣每五年舉辦農業普查一次，若將現有農家八十七萬戶逐一調查，人力耗費不貲，經採用 5％或10％之抽查後，效果甚佳。

(2)抽查可縮短調查與整理的時間：抽查可將全盤作業化繁爲簡，處理時間必可大爲縮短，如現行勞動力調查、米穀生產調查，按季抽取小樣本調查，均能以極短時間，迅速產生調查結果，對於研究分析參考，貢獻至大。

(3)抽查可對被調查的份子從事更詳盡的調查，以便提供吾人更多的信息，俾作成更詳盡的調查。如我國之戶口及工商普查，均採事後抽樣複查，以適量之樣本單位，實施精確調查，如此既可補充大規模普查之不逮，又可減輕普查之負擔。

(4)抽查可選取素質較高的人員，加以嚴格訓練，然後從事調查工作，當可提高統計的正確性。

(5)計劃能更周詳，同時資料之搜集與處理較不易發生錯誤。

(6)抽查可配合特定性與機動性之行政措施：政府機關在配合某種特定用途下，如能隨機舉辦抽查，當可發揮更大的統計功效，如奧地利之老人住屋與社會問題調查，荷蘭之電視對兒童影響調查，美國之科學家及工程師就業調查，以及我國舉辦的漁業、林業及交通運輸業等生產成本調查等等，均為抽樣技術適切應用於特殊目的之措施，調查結果深入，成績顯著。

除以上的優點外，下面兩個場合，就非抽查不能盡其功：

(1)為經費、時間或人力所限制，不能從事普遍調查者。

(2)為全體中之各個體，不容許吾人普遍加以觀察者。亦即，各個體被觀察時，將損壞而喪失或減低其原有價值之場合，如電燈泡耐用時間之檢查。

因此，由於抽樣調查之重要，乃造成近幾十年抽樣理論與實務的發展，進而促進統計學之領域及應用更為擴張。

11－2 簡單隨機抽樣

使母體中每一單位已被選出之或然率均相等稱爲隨機抽樣，其樣本稱爲**隨機樣本**。

換言之在抽樣過程中，對母體不加入任何人爲影響，而完全用純粹且偶然的方法抽取個體，使母體中每一個體均有同樣被抽取之機會，即在抽樣時不將母體按某種分類標準加以分類，而對母體中之每一個體（如一人，一事，一物）實施機會均等的抽樣。此法之主要目的在消除人爲偏見而獲得客觀不偏之樣本資料，以使樣本中各種特性之分配近似於母體現象中各種特性之分配。

爲求達到完全隨機之地步，一般可用亂數表法，隨機號碼表或稱亂數表。它係將0～9等十個數字重複按隨機方式抽出，並按其出現先後順序排列所作成之表。由於表中數字係按隨機方式抽出，毫無規律性，故可根據隨機號碼表進行隨機抽樣。最早之隨機號碼表爲1927年英國Tippett所編（共有41,600字）。實地使用隨機號碼表時，係先將其各行各列編號，並製成卡片，然後抽出一張以定縱行，投返後再抽出一張定橫行，行列交叉處之數字即爲開始採用之數字，由此數字起向左、右、上、下任何一方連續取用所需之數字，即可得到所欲之適當樣本。不過，習慣上通常以由左向右，或由上而下抽取者爲多。

【例1】 試自含有500個單位的母群中，抽出一組含10個單位的樣本。

【解】 假設抽出的行之號碼爲12，而行之號碼爲10，並向右連續取號，則自附錄表八的亂數表中，可得連續之數字有：

5841607449983114632242014858845109372887123424064742 9…………

因爲母群體總數爲500，係三位數，故每三位爲一組，則

分別之編號爲

584，160，744，998，311，463，224，201；485，884，510，937，288，712，342，406，474 。

再除去大於500或重複出現的編號，可得被取出的10個樣本之編號應爲

160，311，463，224，201，485，288，342，406，474 。

設母體有N個單位，今從此母體中用隨機方式抽取n個單位構成樣本，如此每一組樣本出現的機率相等，即在放回方式時爲 $\frac{1}{N^n}$，在不放回方式時爲 $\frac{1}{\binom{N}{n}}$ 。由於完全隨機性不易做到，實際應用不廣，但因此法簡單且爲其他抽樣方法的基礎，故在理論上甚爲重要。

11－3 分層隨機抽樣法（Stratified Random Sampling）

將一母體按某種標準分為若干次母體（Sub-population），這些次母體即稱為層（Strata），各層中所包含的單位互不重疊，且其中所包含單位的總和應等於母體中之總單位數；然後自各層中以一定的抽出率抽取隨機樣本，即得分層隨機樣本。統計學上稱此種抽樣法為**分層隨機抽樣法**。

【例1】 ⑴調查國民所得時，可分成少於10000元，10000元至30000元，30000元至60000元，60000元至100000元及高於十萬元的高薪所得者。

⑵家庭計劃調查時，可分成無孩子人家，一或兩個孩子人家，多於2個孩子人家。

⑶年銷售量之調查時，可分成少於50000單位，50000單位至150000單位及高於150000單位者。

至於各層中樣本單位數的分配，可依下列方式處理之。

⑴**比例配置**（Proportional Allocation）：自各層所抽取的樣本單位數應與各該層的大小成比例，由於樣本必須能代表母體，故大層（內含單位多）應多抽，小層應少抽。如將母體總單位數N分成K層，每層單位數分別為N_1，N_2，……N_k，每層樣本單位數分別為n_1，n_2，…n_k，n為樣本單位數，則由分層抽樣的意義，需合乎

$$\frac{n_i}{n}=\frac{N_i}{N}\left(i=1,2,\cdots\cdots,k \right)$$

故 $$n_i=n\times\frac{N_i}{N}\left(i=1,2,\cdots,k \right)$$

此法簡單易行，且在各層差異間相差不大時仍相當有效。

【例2】 設有一母體分為三層，第一層為200，第二層為400，第三層為600，現欲抽120個為樣本，試問依比例抽樣各層

該抽出多少樣本？

【解】 依比例抽樣公式 $n_i = n \times \frac{N_i}{N}$

由題意知 $N = 200 + 400 + 600 = 1200$，$n = 120$

故 $n_1 = 120 \times \frac{200}{1200} = 20$

$$n_2 = 120 \times \frac{400}{1200} = 40$$

$$n_3 = 120 \times \frac{600}{1200} = 60$$

(2)**非比例配置**（Disprortional Allocation）如果母體單位表徵一致，則抽一個單位爲樣本，就可完全代表母體，但如母體單位表徵差異太大時，則惟有多抽單位構成樣本，對母體才有較大的代表性，故上法僅依單位多寡比例配置樣本單位數並非適宜，實應再考慮各層單位變異的大小，若令σ_i（$i = 1, 2, \cdots\cdots, k$）代表各層單位的標準差，則依Neyman的想法，各層樣本單位數應如下配置：

$$\frac{n_1}{N_1\sigma_1} = \frac{n_2}{N_2\sigma_2} = \cdots\cdots = \frac{n_k}{N_k\sigma_k} = \frac{n}{\sum_{i=1}^{k} N_i\sigma_i}$$

即

$$n_i = n \times \frac{N_i\sigma_i}{\sum_{i=1}^{k} N_i\sigma_i} \quad i = 1, 2, \cdots k$$

即n_i與N_i和σ_i成正比，抽樣時，除考慮結果之代表性大小外，尚須考慮所需付出成本的高低，高成本之群體單位應少抽，低成本者多抽，故Deming進一步提出了考慮因素更多的樣本配置公式，即

$$\frac{n_1}{N_1\sigma_1/\sqrt{c_1}} = \frac{n_2}{N_2\sigma_2/\sqrt{c_2}} = \cdots\cdots = \frac{n_k}{N_k\sigma_k/\sqrt{c_k}} = \frac{n}{\sum_{i=1}^{k} \frac{N_i\sigma_i}{\sqrt{c_i}}}$$

故 $$n_i = n \times \frac{N_i\sigma_i / \sqrt{C_i}}{\sum N_i\sigma_i / \sqrt{C_i}} ,\quad i = 1 , 2 \cdots\cdots\cdots\cdots k$$

式中C_i表各層單位之調查成本（它通常是個平均數），n_i與N_i和σ_i成正比，而與$\sqrt{C_i}$成反比。

【例3】 設將含有1200單位之母體分爲三層，第一層含200單位，第二層含400單位，第三層含600單位。已知各層變量的標準差分別爲$\sigma_1 = 3$，$\sigma_2 = 4$，$\sigma_3 = 3$，今擬於全體中抽出100單位組成樣本，試求各層應抽出之單位？

【解】 固定總抽出數n以使變異數$V(X)$爲最小，可採用下列之牛曼分層抽樣公式，以定各層之抽出數。

$$n_i = n \cdot \frac{N_i\sigma_i}{\sum_{i=1}^{k} N_i\sigma_i} \quad (n_i \text{ 表第 } i \text{ 層抽出數})$$

今已知$N_i = 200$，$N_2 = 400$，$N_3 = 600$，$n = 100$

$$\sigma_i = 3 ,\ \sigma_2 = 4 ,\ \sigma_3 = 3$$

$$\therefore \quad n_1 = n \cdot \frac{N_1\sigma_1}{\sum_{i=1}^{3} N_i\sigma_i}$$

$$= 100 \cdot \frac{200 \times 3}{200 \times 3 + 400 \times 4 + 600 \times 3} = 15$$

$$n_2 = n \cdot \frac{N_2\sigma_2}{\sum_{i=1}^{3} N_i\sigma_i}$$

$$= 100 \cdot \frac{400 \times 4}{200 \times 3 + 400 \times 4 + 600 \times 3} = 40$$

$$n_3 = n \cdot \frac{N_3\sigma_3}{\sum_{i=1}^{3} N_i\sigma_i}$$