

S T A T I S T I C S

Theory and Application Vol. 1

統計學

理論與應用 (上)

程大器 編著

智勝
BEST-WISE

統計學 Statistics
理論與應用 (上)
Theory and Application Vol. 1

程大器 編著

智勝文化

統計學理論與應用（上）

Statistics Theory and Application Vol.1

國家圖書館出版品預行編目資料

統計學. 上, 理論與應用 / 程大器編著. -- 初

版. -- 臺北市：智勝文化, 2001[民 90]

面；公分

參考書目：面

含索引

ISBN 957-729-176-7（平裝）

1. 統計學

510

90000095

作 者/程大器

發行人/紀秋鳳

出版/智勝文化事業有限公司

地址/台北市 100 館前路 26 號 6 樓

電話/(02)2388-6368

傳真/(02)2388-0877

郵撥/16957009 智勝文化事業有限公司

登記證/局版臺業字第 5177 號

總經銷/知識達圖書發行有限公司

傳真/(02)2312-2288

出版日期/2003 年 9 月初版三刷

定價/500 元

ISBN 957-729-176-7

Statistics: Theory and Application Vol.1

by Cherng Dah Chih

Copyright 2001 by Cherng Dah Chih

Published by BestWise Co., Ltd.

智勝網址:<http://www.bestwise.com.tw>



本書之文字、圖形、設計均係著作權所有，若有抄襲、模仿、冒用情事，依法追究。
如有缺頁、破損、裝訂錯誤，請寄回本公司調換。

自序

統計學是一門理論與實務應用並重之學科，其理論基礎包含了代數、微積分及或然率等基礎數學，而其應用更是廣泛，可說上至天文、下至地理，一切人文社會與自然生物等科學皆可應用。古人常言：「工欲善其事，必先利其器」，是故進行任何研究工作，必須先研究各種研究方法，如此不僅事半功倍，更對科學領域的開拓有長足進步之影響。

一般人皆認為統計學是一門深奧難懂的學科，因此學習統計的態度總是以背誦公式為主要學習方法，而對於其理論基礎卻不予理會，因此在實務應用中，常會有不知要使用哪種類型的統計方法之挫折感。事實上，學習統計應該理論與應用並重，才能相得益彰，但對於大多數的學習者而言，由於基礎數學的程度較差，因而對於學習統計理論部分時，常會感到窒礙難行，因此常會放棄統計理論部分，而針對統計應用方法加以學習，如此惡性循環下去，更造成了統計方法之濫用，進一步導致錯誤之決策。有鑑於此，筆者為了讓學習者對統計方法之學習有系統性之學習架構，因此才有編寫本書之動機。

本書共分為上、下兩冊，上冊內容主要部分為敘述統計與機率論。敘述統計即大量觀察統計資料之統計分析，包括資料之蒐集、整理與分析（見第一及第二章），而本書上冊主要精華在於機率論部分，由於一般商用統計學書籍皆有共同缺失，即機率論之部分描述太少，使讀者在學習推論統計學時常會造成困擾，不知其推論方法之理論基礎，因此為了使讀者能對上述之缺失能予以改進，故本書對於機率理論部

分有詳細的描述（見第三及第八章），且編排之方式由淺入深，而對於涉及基礎數學部分更是詳細交代其來源及觀念，而對於每個理論之證明皆盡量以較淺顯之方式交代，使讀者能消除學習統計學理論部分之恐懼。

在一般大學統計教學中，由於教學時間有限，因此常常以些許之時間來討論機率部分，而大部分時間以探討推論及應用統計為主。但學生如以此方式學習統計，比較不容易深入瞭解統計方法之奧妙，況且對於一般就業及研究所考試時也無法應付大量之機率理論部分之考題。因此學習統計學若想要融會貫通，則對於機率部分之研習一定要詳細，如此在探討推論及應用統計時，才能瞭解統計方法之架構與應用。在此必須強調一點，即機率與微積分等基礎數學對統計均是工具，而學習統計主要的是分析資料及實務應用，因此所研究或討論的內容均以統計為目標，切勿將機率與統計混淆，而捨本逐末。

本書下冊內容就是以統計方法為基礎，讓讀者能有效的將統計方法應用在日常生活中，而其內容主要區分為推論及應用兩大部分。在推論統計中，主要是將上冊所學之知識用來發展一些推論方法，內容以估計與檢定為主，此內容也是統計學主要之精華，若讀者能瞭解其架構與觀念，則對於學習應用編中之變異數分析及迴歸分析，將會有莫大之助益。變異數分析及迴歸分析是屬於統計應用部分，由於其應用廣泛，而且成效卓著，因此在實務上是最常用之分析工具。而下冊內容中最後將探討無母數統計方法，此部分是統計在近幾年來發展最快速之領域，也是一般升學考試較容易疏忽之部分，且在實務應用中也是常見之分析方法，讀者對於各種方法要釐清觀念，以方便往後應用。

本書因為考慮統計學之整體性與應用性，故其內容較多且廣泛，

因此對於初學者在學習本書內容所花費之時間將會較久，但若讀者對於一些機率論的基礎理論證明、二維隨機變數之機率分配以及變數變換單元無興趣，可直接略過，如此對於數學基礎不好之同學在閱讀本書時，也能快速且有效率地達到學習的效果。

筆者以多年教學經驗編寫本書，希望能對莘莘學子學習統計有所幫助，因此內容盡量翔實，但筆者學識與經歷均屬有限，故對於遺漏或謬誤之處盼請見諒，尚請各方先進及讀者不吝賜教。

最後，感謝周光凱老師給予書籍內容之指正及編排順序上之建議，以及感謝智勝文化事業所有同仁之支持與協助，使本書不論在編排及版面設計上，皆能引人入勝。

程大器

目 *Contents* 錄

第一章 敘述統計 I —— 資料蒐集與整理

第一節	緒論	2
第二節	統計資料之蒐集	9
第三節	統計資料之整理	16

第二章 敘述統計 II —— 統計量數

第一節	集中趨勢量數	44
第二節	離差量數	82
第三節	偏態與峰態量數	109

第三章 機率概論

第一節	集合	129
第二節	樣本空間	141
第三節	機率概念	154
第四節	條件機率與獨立	177
第五節	貝式定理	191

第四章 隨機變數及機率分配

第一節	隨機變數之意義	204
第二節	機率分配	208
第三節	分配函數	227
第四節	聯合機率分配	237
第五節	邊際機率分配	245
第六節	條件機率分配	251
第七節	獨立隨機變數	258
第八節	隨機變數函數之機率分配	266

第五章 期望值

第一節	期望值及其性質	299
第二節	特殊期望值 —— 動差、變異數、偏態係數、 峰態係數	315
第三節	動差母函數	327
第四節	常見不等式	334

第五節	二維隨機變數期望值	339
第六節	兩隨機變數間關係—共變異數與相關係數	345

第六章 特殊間斷機率模式

第一節	均勻分配	382
第二節	伯努利分配	387
第三節	二項分配	388
第四節	多項分配	402
第五節	超幾何分配	410
第六節	多變數超幾何分配	420
第七節	波瓦松分配	423
第八節	負二項分配	435
第九節	幾何分配	442

第七章 特殊連續機率模式

第一節	均勻分配	456
第二節	常態分配	463
第三節	指數分配	485
第四節	伽碼分配	492
第五節	卡方分配	498
第六節	其他連續型分配	504

第八章 抽樣與抽樣分配

第一節	緒論	518
第二節	抽樣分配	529
第三節	樣本平均數的抽樣分配	538
第四節	樣本比例的抽樣分配	548
第五節	兩樣本平均數差的抽樣分配	554
第六節	常態分配有關的三種抽樣分配	560
第七節	順序統計量的抽樣分配	578

附錄

附錄一	二項機率分配機率表	588
附錄二	波瓦松機率分配機率表	595
附錄三	標準常態分配機率值	600
附錄四	卡方分配右尾機率值之臨界點	602
附錄五	T 分配臨界值	603
附錄六	F 分配臨界值	604
附錄七	隨機亂數表	612

參考書目

中文書目	614
英文書目	615

索引

中英文索引	618
-------------	-----

1

敘述統計 I — 資料蒐集與整理

緒論

1.1

統計資料之蒐集

1.2

統計資料之整理

1.3



緒 論

1.1.1 統計的認識

統計學 (statistics) 是在面對不確定情況下，使人們能做出聰明決策的科學方法，其過程必須經過資料的蒐集、整理與分析。

一般人皆認為統計只不過是一些常在報章雜誌上看到的數字或類似煙囪的圖形，例如棒球的打擊率、離婚率、進出口數字、股票的價格變化……等，但這種認識與統計的定義——蒐集、整理、陳示、分析數字資料——相差甚遠。統計資料除了是有意義的數字資料外，也是我們分析某種現象的重要依據，因此統計學應為一種用以建立宇宙複雜現象之通理通則的科學方法，亦為解決問題的指南，因此統計不僅代表統計資料，更代表一種方法。

生活在現代的人們，常被一些經由報章、雜誌、書籍、廣播、電視等各種人類活動傳送之大量資料所干擾及困惑，因此若要將此大量資料轉換成有用之資訊以方便使用及分析，唯有透過統計方法加以整理及歸納。事實上，統計方法對科學家、工程師、企業家或其他參與解決問題活動的人而言，則更為重要，他們可藉由蒐集、整理、分析及推論結果等方法將小量資料推及大量資料，而將一些不確定之現象利用機率之原理來判斷，以做出正確的決策行為，達到最高的效用。

統計分析在各行各業的許多問題中常被廣泛地應用，例如一個科學家如何評價一個新理論的正確性？一個學生如何去試驗一種新藥品的效用？最近失業率的改變是否是季節性的變動或是隨機性的差異？為了要正確地判定一批商品的品質應抽樣檢查幾個商品？為何只要訪問幾百個投票者就可預測選票的結果？國民所得如何用稻米的產量、石油的價格、出超或入超的數字來預測？只要想探討不明現象的真相，

幾乎都可以利用統計方法去解決。尤其近代統計學主要的作用是用以發展原理及方法，來幫助我們在面對不確定的情況下做出正確決策。事實上，今天很多經濟的、社會的、工業的、商業的、心理的現象都是不確定的，因此想要在不確定的現象下做出正確決策，統計就是一個很有效率及價值的工具，因此在一百多年前統計學家 H.G Wells 就曾說過：「統計方法的思考，有一天將會和讀、寫一樣，成為一般國民所必須具備的基本能力。」

本書之重點將放在統計方法的運用上，以幫助研究者解決問題。各種統計資料之討論僅為基礎，而資料蒐集的方法、資料整理的原則，以及資料事實之分析則均將逐章討論。由於統計方法日新月異，且為解決不確定問題所必須具備之知識，因此 Wallis 與 Robert 稱統計學是對不確定事件做出最明智決定 (making wise decision) 的一種方法。

1.1.2 統計方法過程

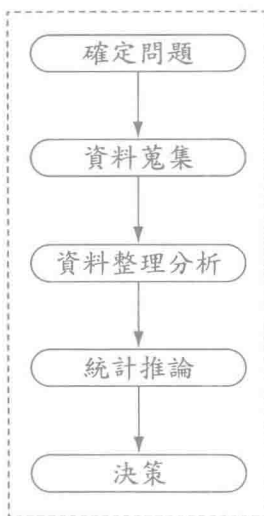
統計工作基本上可大略分成下列幾個重要階段，為了使讀者更加明瞭何謂統計方法及其內容，我們以圖 1.1 來說明其流程，並詳述各階段工作內容於後。

一、確定問題

根據研究目的來確定所要研究的問題，並作為統計調查之對象。此處所謂的對象，在統計學中稱為全體或母體 (population)，而母體中的每一份子就稱為基本單位 (elementary unit)，每個基本單位所對應的數字則稱為觀測值 (observation)。例如，某工廠要決定某一月份生產燈泡的平均壽命，則該月份生產的所有燈泡就是全體，每個燈泡就是基本單位，每個燈泡的壽命就是觀測值，欲得觀測值即須先蒐集資料。

Figure

●圖 1.1 統計方法過程



二、資料蒐集

資料的蒐集常因人為的錯誤而使樣本失去真實性，例如金賽博士 (Dr. Kinsey) 研究性問題，常問及某人一個月行房幾次？每次行房高潮的程度？由於這類問題對大部分人而言都非常敏感，所以人們通常不會據實回答，其所得到的資料便不真實，以此資料所做的推論自然也就不會正確了。一般來說，蒐集資料是整個統計調查過程中最費時、也最花錢的階段，不過，其亦有難易之分。例如要檢查某種布料顏色的好壞，只需看看布料就可決定，但有些問題的處理則既複雜又要有一些技術，如試驗某種藥品含某種化學成分的比率、某種飼料對豬體重增加的效用、某公司調查消費者對其商品之印象如何等。

在實務上，研究者不必要也不可能去調查母體之每一個基本單位。例如要調查燈泡的壽命，如果把生產的燈泡全部拿來試驗，最後雖然

試驗出壽命，但燈泡也沒有了；再者，若要預測選舉的結果而對所有的選民進行訪問，這與舉辦一次選舉有何不同？事實上，根據統計抽樣的理論，我們只要調查母體的一小部分對象就可利用優良的統計方法去推論母體的真實狀況。這一小部分在統計中稱為樣本 (sample)，得到樣本的方法稱為抽樣 (sampling)。當然我們希望抽到很好的樣本，也就是說，抽出的樣本要足以代表母體，否則便會發生很嚴重的錯誤。

很明顯地，因金錢、時間……等因素的限制，我們必須採用抽樣的方法，從母體中抽取一些樣本，並以這些樣本為基礎，對整個母體做推論，因此，樣本是否能適切地代表母體是很重要的。基本上，要獲得足以代表母體的樣本，並不困難，只要善加利用統計抽樣理論方法，便可以幫助我們得到一個具有代表性的樣本，這也就是說，蒐集哪種資料？如何去蒐集？要蒐集多少個樣本？關於這些問題皆可以利用所謂的抽樣設計 (sample design) 方法幫忙解決。在計劃及設計抽樣時要非常小心，否則無法得到正確的結論。要好的推論與結論，樣本代表性的好壞就非常重要，當然要樣本代表性高，就必須有好的抽樣方法，因此現代統計學都是建立於抽樣理論上，故讀者宜深切體會抽樣方法及其觀念。

三、資料整理及分析

當資料蒐集好之後，要把它整理成讓人容易瞭解的形式，並分類成某種有系統的形式、表格或圖形，再計算一些統計中所需要的測度值，如比例 (proportion)、平均數 (average)、眾數 (mode)、中位數 (median)、變異數 (variance)、標準差 (standard deviation)、全距 (range)、偏態 (skewness)、峰度 (kuotosis) 等。而利用這些樣本資料所計算出來的測度值就稱為樣本統計量，簡稱為統計量 (statistic)，而經

過普查，從整個母體計算出來的測度值，稱為母數或參數 (parameter)。通常這些母數都是未知的，而統計主要工作就是為了決定這些母數是什麼。既然樣本足以代表整個母體，則從樣本資料所計算出來的統計量就可代替這些未知母數。但是這種計算的階段必須包括資料的組織化 (organization)、表現化 (presentation) 及簡化 (summarization) 等，即所謂的敘述統計 (descriptive statistics)；至於如何用統計量去決定母體的未知母數？用哪種統計量去估計母體的母數？估計準確性如何？則是所謂的推論統計 (inference statistics)。

因此「Statistics」這個字，除可以稱為統計資料或統計方法外，其內容可更進一步區分為敘述統計 (descriptive statistics) 與推論統計 (inference statistics or inductive statistics) 兩大類。

敘述統計是經由蒐集、整理與分析的統計方法，表現統計資料的特性，內容包括集中趨勢、差異分析、週期性之變動、圖形之狀態與其他事項的關係等；而推測統計則是採用機率原理，尋求如何經由最適當的抽樣方法取得或蒐集樣本，並依此樣本資料的整理求得樣本的特徵統計量，再經由假設與檢定，以樣本統計量從事推定與檢定未知母數的方法，即在解決樣本與母體間推理的問題，其內容包括機率、抽樣方法、機率分配、樣本統計量、抽樣分配、統計之假設與檢定等。

四、統計推論

資料整理完成後，研究者就可利用樣本資料所計算出的樣本統計量來估計母體中未知的母數，或是用假設檢定的方法判斷母體的母數是否為某個常數。統計推論方法可分為估計 (estimation) 及假設檢定 (testing hypothesis) 兩大類。所謂估計是指利用樣本統計量去推論母體未知的母數，例如某批產品不良率的調查，在不能檢驗整批產品的情