

陳正昌 著

SPSS

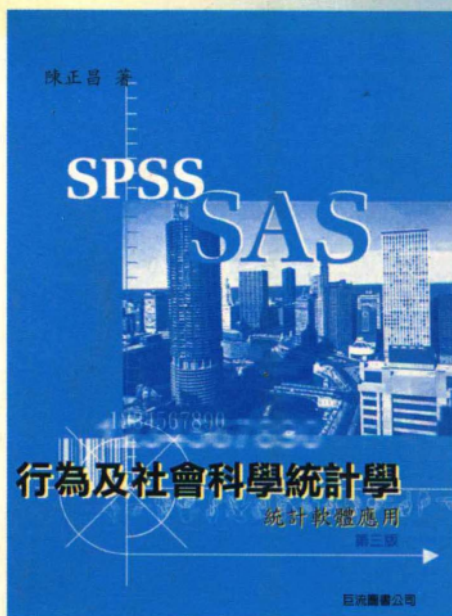
SAS

1234567890

行為及社會科學統計學

統計軟體應用

第三版



在行為及社會科學的研究中，統計學已成為不可或缺的工具。國內近年來各種統計學的論著雖不能說是汗牛充棟，但也頗有可觀的了。為什麼有這麼多優秀的統計學專書之外，筆者還要再寫一本類似的著作呢？首先，目前大多數統計學的論著，在某種程度上仍比較偏向數理統計學，對於數理背景較弱的學生，閱讀上可能有些困難。本書屬於應用統計學，希望能對非數理學系的學生有所助益。其次，多數應用統計學還是比較著重於計算過程的介紹，對於統計方法的適用時機及應注意事項著墨較少，本書希望在這方面多做說明，期能使讀者適當使用統計方法。第三，配合軟體來介紹統計學是目前的趨勢，以往國內的著述較少配合統計軟體來說明，因此讀者還要再花時間學習統計軟體；另一方面，當前許多介紹統計軟體的書籍，又偏向單純說明使用的方法，對於統計學觀念的解說難免有所忽略。本書希望能畢兩者之功於一役，同時引介統計學及統計軟體的使用方法。

ISBN 957-732-206-9



9 789577 322067

行為及社會科學統計學
— 統計軟體應用
(第三版)

陳正昌 著

巨流圖書公司

國家圖書館出版品預行編目資料

行爲及社會科學統計：統計軟體應用／陳正昌
著.--三版.--臺北市：巨流,2004[民93]
面：公分
參考書目：面
含索引
ISBN 957-732-206-9（平裝）

1.統計學 2.統計—電腦程式

511.2

93009878

行爲及社會科學統計：統計軟體應用

出版者：巨流有限圖書公司

創辦人：熊嶺

總編輯：陳巨擘

著者：陳正昌

編輯部：106台北市大安區溫州街48巷5號1F

電話：(02)23695250・23695680

傳真：(02)83691393

E-mail:chuliu@ms13.hinet.net

http://www.liwen.com.tw

總經銷：麗文文化事業機構

地址：802高雄市苓雅區泉州街5號

電話：(07)2261273

傳真：(07)2264697

郵購：郵政劃撥帳號41299514

出版登記證：局版台業字第1045號

ISBN 957-732-206-9（平裝）

2004年7月三版一刷

定價：550元

版權所有・請勿翻印

本書如有破損、缺頁或倒裝，請寄回更換。

作者簡介

陳正昌

學歷：省立臺北師專，1983年

國立政治大學教育學博士，1994年

經歷：臺北縣國小教師六年，1983年 - 1989年

行政院教育改革審議委員會副組長，1994年 - 1995年

現職：國立屏東師範學院初等教育學系副教授，1995年 -

e-mail：chenc@[mail.npttc.edu.tw](mailto:chenc@mail.npttc.edu.tw)

第三版序言

為什麼要寫這本書

在行為及社會科學的研究中，統計學已成為不可或缺的工具。國內近年來各種統計學的論著雖不能說是汗牛充棟，但也頗有可觀的了。為什麼有這麼多優秀的統計學專書之外，筆者還要再寫一本類似的著作呢？

首先，目前大多數統計學的論著，在某種程度上仍比較偏向數理統計學，對於數理背景較弱的學生，閱讀上可能有些困難。本書屬於應用統計學，希望能對非數理學系的學生有所助益。

其次，多數應用統計學的著述，還是比較著重於計算過程的介紹，對於統計方法的適用時機及應注意事項著墨較少，本書希望在這方面多做說明，期能使讀者適當使用統計方法。

第三，配合電腦軟體來介紹統計學是目前的趨勢。以往國內的著作較少配合統計軟體來說明，因此讀者還要再花時間學習統計軟體；另一方面，當前許多介紹統計軟體的書籍，又偏向單純說明使用的方法，對於統計學觀念的解說難免有所忽略。本書希望能畢兩者之功於一役，同時引介統計學及統計軟體的使用方法。

怎麼使用這本書

筆者在寫作時將讀者假想為對應用統計學有部分認識，但是又不一定有清楚的概念，因此對公式的證明及其來龍去脈並未詳加說明，而比較強調統計概念的介紹，及軟體的使用方法和報表的解說。如果讀者已經上過一個學期的統計學課程，想要利用軟體來進行練習或分析論文資料，本書應是不錯的選擇。

書中首先說明統計學的基本概念(第一章)。其次介紹統計軟體的使用方法，其中並含有資料之除錯及轉換(第二、三章)。如果對於統計學及統計套裝軟體不熟，最好能夠循序閱讀這三章。第三部分(第四章至第十五章)則按變數之數目及

性質，分別介紹適合的統計方法。讀者只要計算研究變數的數目並判斷其性質，就可以找到可用的分析方法。

第四章以後每節均依研究問題、統計思考、程式解說、報表解說，及研究結論的順序呈現。研究問題大多以筆者(陳正昌, 1994) 博士論文的調查資料為例，因此絕大多數都是真實的數據。統計思考部分則針對研究問題加以思考可用的統計分析方法，並大要介紹其公式、適用時機，及統計假設。

除了第十一章因為 SPSS 沒有專門分析母群體比例的方法，SAS 也沒有專門進行第十五章無母數分析的方法外(如果要分析須經過較多的步驟，因此不加以介紹)，每一章的程式解說均分為 SPSS 及 SAS 兩部分。其中 SPSS 同時說明利用選單(menu) 點選及撰寫程式，讀者可以依序點選選單(括號內的中文是中文版的選單)，最後再按「Paste (貼上語法)(syntax)」即可產生程式(或稱為語法)。不過有部分程式指令是系統內定的，不一定要寫出來，所以用選單點選後所貼上的語法會和筆者直接撰寫的程式稍有不同，但是兩者的執行結果是相同的，讀者可以不必疑惑。程式中大寫部分是固定的關鍵字，讀者可以全部照抄；小寫部分通常是變數名稱(或資料集名稱)，應視使用者所要分析的變數加以更改。

目前許多統計科系研究者仍傾向使用 SAS，但是由於 SPSS 比較不須撰寫程式，且大多數的分析結果都已經表格化，所以報表解說以英文版 SPSS 為主(同時也加入中文版之原始註解)，只有在 SPSS 缺乏該種統計方法時才改用 SAS。由於 SAS 與 SPSS 在單變量統計的報表差異不大，所以 SAS 的使用者仍然可以在本書中找到相對應的報表解說。

研究結論是經由問題、思考，及分析後所得的結果，藉此也將整節的內容再加以統整摘要。因為使用的資料多數是筆者親自調查所得，因此研究結論也都是信而有徵的。

本版的改變

本次改版，首先修訂前兩版中已知的錯誤，並對部分敘述不夠詳盡之處再加以補充。

其次，雖然變異數的假設考驗(hypothesis test for variance) 在行為及社會科學

領域較少使用，但是在工業方面卻是相當的課題。另一方面，在進行變異數分析 (analysis of variance, ANOVA) 時，需要針對變異數同質性加以檢定，因此本版又加入變異數的區間估計及假設考驗。

而在經濟學領域中，時間數列是經常用到的統計方法，不過在行為及社會科學領域則較少見。為了增加讀者對單變量時間數列中 ARIMA (autoregressive integrated moving average) 模式的認識，筆者採用九十二學年度國小教師數加以分析，放入附錄中。

在 2001 年第五版的 APA 出版手冊 (Publication Manual of the American Psychological Association) 中提及，研究報告中除了顯示研究結果是否達到統計上的顯著之外，仍要告訴讀者研究的重要如何，因此應該在論文中說明效果量 (effect size) 及關聯的強度，而提供效果量也有利於後續的整合分析 (meta analysis)。本書亦針對此部分加以介紹。

此外，統計軟體的版本不斷推陳出新，因此本書主要配合 SPSS 12.0.1 及 SAS 9.0 版加以說明。雖然版本有所差別，但是常用的分析方法並無太大差異，因此本書仍可適用於使用 SPSS 6.1 版至 11.5 版，及 SAS 6.12 版至 8.2 版的讀者。

雖然筆者一直認為針對統計軟體的語法加以了解，才能掌握其精髓。不過為了因應視窗軟體直接選取選單的方便性，本版也擷取分析過程的畫面，希望能讓讀者更快熟悉統計軟體，以便將比較多的心力用在分析及決策中。

感謝與期許

本書能夠出版，巨流圖書公司楊老闆伉儷及陳巨擘總編輯的支持及督促居功厥偉；因為他們的用心，使得本書能夠採用雙色印刷，不僅使品質提升，更有利於讀者閱讀；而李麗娟小姐負責全書的出版業務，也是功不可沒。

這段期間，非常感謝大家採用本書。其中，部分讀者來函索取資料檔，有幾位讀者與我討論書中的疑問，這都使我感受大家的支持與督促，因而此次改版的工作也就更加審慎。

本書的大部分內容都在個人的授課過程中使用過，此期間部分學生指出本書難以理解之處，更有許多學生發現疑問與我討論，使我得以更正錯誤。所以，本

版如果能夠比較沒有疏漏之處，這些學生功不可沒。在教學中，看到學生對統計學從陌生而熟悉，從畏懼而喜愛（多數人至少是不害怕），從茫然不解到獨立完成論文資料分析，就是筆者最大的成就。而在與年輕學子的互動中，我發現多數學生都相當活潑並具創意，這是值得個人學習之處。

而最應感佩的是屏東師院陳新豐教授，在本版付梓前詳細閱讀初稿，除了指出錯誤之處，更提出許多寶貴意見，使得內容更臻完善。在此衷心感謝您的協助。

本書從撰寫——初版——改版，經過了將近六年的時間。書中的每一個字、每一條公式，都是由我輸入，連排版工作也是獨力完成，所以期間投注了不少心力。但是筆者一直相信最好的作品都尚未完成。如果您在閱讀過程中，發現有疑問或是錯誤之處（SPSS 分析步驟如果有錯誤，可以參考程式設計部分），請直接與我聯繫。期望在各位的支持及個人不斷地努力下，以後能將比較接近理想的成品呈現給大家。如果需要書中的資料檔進行練習，煩請上筆者的網頁，在「個人著作」之「專書」部分下載 (<http://cclearn.npttc.edu.tw/tuition/ccchen-web/>)。

陳正昌 於屏東師院
2004年6月

目次

第三版序言	I
第一章 統計學簡介	1
第一節 統計學的分類	1
第二節 學習統計學的理由	3
第三節 學習統計軟體的理由	6
第四節 測量的尺度	8
第五節 變數的分類	10
第六節 變數性質及統計方法	12
第二章 統計軟體使用簡介	15
第一節 較知名之統計軟體	15
第二節 安裝與執行程式	17
第三節 資料之蒐集及登錄	32
第四節 統計方法及程式選擇	42
第三章 資料之存取、預檢與轉換	45
第一節 資料之讀取與儲存	45
第二節 資料之預檢	54
第三節 資料之轉換	59
第四章 次數分配表及圖示法	73
第一節 質的變數之次數分配表	73
第二節 量的變數之次數分配表	77
第三節 圖示法	84
第五章 一個變數之描述統計	97
第一節 名義變數之描述統計	97
第二節 次序變數之描述統計	101
第三節 等距及比率變數之描述統計	106

第六章	兩個質的變數之關聯分析	115
	第一節 兩個名義變數之關聯分析	115
	第二節 兩個次序變數之關聯分析	128
	第三節 名義變數與次序變數之關聯分析	139
第七章	兩個量的變數之關聯分析	149
	第一節 兩個量的變數之關聯分析	149
	第二節 兩個量的變數之迴歸分析	162
第八章	質的變數與量的變數之關聯分析	173
	第一節 名義變數與量的變數之關聯分析	173
	第二節 次序變數與量的變數之關聯分析	177
第九章	兩個質的變數之差異分析	185
	第一節 卡方之適合度考驗	185
	第二節 卡方之百分比同質性考驗	198
	第三節 卡方之獨立性考驗	209
	第四節 卡方之改變顯著性考驗	215
第十章	平均數的區間估計及假設考驗	225
	第一節 平均數的區間估計	225
	第二節 一個平均數之假設考驗	236
	第三節 兩個獨立樣本平均數之假設考驗	248
	第四節 兩個相依樣本平均數之假設考驗	259
第十一章	百分比的區間估計及假設考驗	269
	第一節 百分比的區間估計	269
	第二節 一個百分比之假設考驗	274
	第三節 兩個獨立樣本百分比之假設考驗	283
第十二章	變異數的區間估計及假設考驗	291
	第一節 變異數的區間估計	291
	第二節 一個變異數之假設考驗	297

第三節	兩個獨立樣本變異數之假設考驗	304
第四節	三個以上獨立樣本變異數之假設考驗	309
第十三章	變異數及共變數分析	317
第一節	獨立樣本單因子變異數分析	317
第二節	相依樣本單因子變異數分析	345
第三節	獨立樣本二因子變異數分析	356
第四節	獨立樣本單因子共變數分析	380
第十四章	三個以上量的變數之關聯分析	395
第一節	淨相關及部分相關	395
第二節	多元迴歸分析	403
第三節	徑路分析	422
第十五章	質的變數之差異考驗—無母數統計分析	431
第一節	一個樣本的差異性考驗	431
第二節	兩個獨立樣本的差異考驗	440
第三節	兩個相依樣本的差異考驗	447
第四節	三個以上獨立樣本的差異考驗	451
第五節	三個以上相依樣本的差異考驗	454
附 錄	時間數列分析	461
第一節	統計概念簡介	461
第二節	SPSS 程式	464
第三節	SAS 程式	478
第四節	分析摘要表	484
參考書目		487
索引		491
中文部分		491
英文部分		502

第一章

統計學簡介

第一節 統計學的分類

統計學 (statistics) 是蒐集、整理、分析，及推論數字資料 (numerical data) 的科學方法 (朱經明, 2001, p.1)。它可以使研究者更精確地描述資料，以適宜且有意義的方式彙整結果，進而得到更概括性的結論。其分類方式大略有以下兩種：

壹、數理統計學與應用統計學

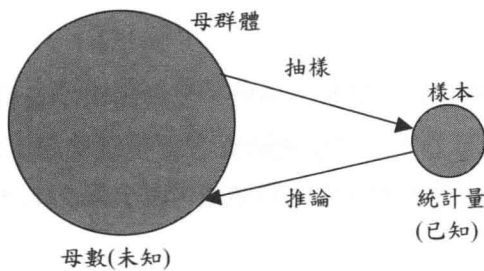
數理統計學 (mathematical statistics) 比較著重統計方法的數學原理及統計公式的推導，是一門抽象、理論，及有系統的學術領域。在這個領域，機率論、微積分，及線性代數等數學的基礎非常重要。目前國內有許多大學校院都開設有統計學系，其中有研究所課程者也不少。

應用統計學 (applied statistics) 著重於將統計學的知識應用在其他領域 (如：教育學、心理學、社會學、管理學……)，主要將統計學當成一個工具，因此對於統計之數學方法及原理的探究相對較不著重。不過這也不意味對於各種統計假設及應注意要點可以忽略，因為對統計學原理不了解，就會造成誤用及濫用，形成「垃圾進，垃圾出」(garbage in, garbage out; GIGO) 的情形。

貳、描述統計學與推論統計學

描述統計學 (descriptive statistics) 著重於資料整理及描述，此外，也包含製作圖表。使用描述統計，可以將一堆龐雜而無頭緒的數據 (例如：50 個學生的 3 科成績，或 1,000 人的身高、體重)，分析整理成有意義及可以理解的資料。描述統計的發展與政府調查有很大的關聯。

推論統計學 (inferential statistics) 包括估計 (又含點估計及區間估計) 及考驗 (test, 或譯為檢定。常見的考驗有 z 考驗、 t 考驗、 F 考驗, 及 χ^2 考驗) 兩部分。後者的目的是希望將樣本 (sample) 分析所獲得的結論, 推論至母群體 (population); 前者則是希望經由樣本的性質 (稱為統計量或統計數 (statistic), 一般使用羅馬字母來代表) 來推論母群體的性質 (稱為母數或參數 (parameter), 一般使用希臘字母來代表)。其意義可以用下圖表示:



平常煮湯時「嚐湯頭」的動作, 其實就隱含了推論的意義。我們常會從一鍋湯裡舀一小匙來試一下鹹度 (或甜度) 是否足夠。為什麼我們不需要把全部的湯喝完就知道它不夠鹹 (或太甜)? 那是因為我們知道這一小匙的樣本, 與整鍋母群體的性質是一樣的, 所以可以「見微知著」, 從樣本推論母群體的性質。

有些學者 (如林清山, 1992) 特別將實驗設計 (design of experiments) 單獨列出。它是推論統計學的一支, 主要著重於變數間因果關係 (cause-and-effect relation) 的探討, 統計方法最常使用變異數分析 (analysis of variance)。實驗設計的發展來自於農業的研究, 英國統計學家 Fisher 對此領域有極大的貢獻。

至於為什麼要抽樣而不選取全部的母群體呢? 理由有四 (林惠玲、陳正倉 1999, p.240; 詹志禹、賴世培, 1996, p.33): 一是母群體太大, 或是無限大, 不可能全部選取。例如要調查收視率或是候選人的支持度, 如果全臺灣的家庭或選民都訪問, 則可能一年後才能知道結果, 最後調查就失去意義了。又如要了解河川受污染的情形, 一定也是隨機選取一些河水裝入試管, 再加以檢驗, 不可能把整條河的水都拿來分析。二是選取後可能會破壞觀察體, 例如工廠要測試燈泡的壽命或是車子的耐撞程度, 不可能把全廠生產的產品都拿來測試, 一定只會「隨機」選取一些樣本而已。又如煮湯, 我們也不可能把整鍋湯都喝完了, 才說湯太鹹。平時市場水果攤會提供一些試吃的樣本 (是否隨機則不敢保證), 也是同樣的

道理。三是有時對樣本加以測量反而比較精確。例如研究者想在一週內了解臺灣地區每個年齡層人口的平均身高，如果要對二千三百萬人逐一測量，不僅不可能，而且也有可能隨便交差了事，此時倒不如隨機找些有代表性的樣本，仔細測量，如此反而比較準確。四是資源有限，在時間、人力及金錢等資源有限之下，只好採用抽樣方式。

第二節 學習統計學的理由

統計學在目前的許多學術領域幾乎都是必修學科，行為及社會科學方面也是如此。依據林清山(1992)及余民寧(1995)的見解，學習統計學的理由有五點(其中一點筆者與兩位學者持不同見解)，以下將其中四點加以闡釋，並另外加上一點補充。

壹、實際工作的需要

不管是教育工作者、社會工作者，或是心理工作者，在專業工作上常會運用統計學的知識。

以教師為例，學生學習後的教學評量，常須「統計」學生的得分、最高分、最低分、平均分數，或是將分數加以分組，其實都在應用統計學的知識。又如輔導人員(含社會工作人員)經常會經由各種測驗或量表來了解受輔導者的心理特質，此時也需要有統計學的知識。

貳、研究工作的需要

目前許多研究都須使用量的取向(quantitative approach)，此時，統計方法的適當運用就非常重要。即使是質的取向(qualitative approach)的研究，有時也會採用一些簡單的統計方法(如：計算次數、百分比……)。

企業界研究如何減少成本、提高產品的良率，甚至減少工作人員的錯誤或提

高工作士氣，都會應用到統計學的知識。最近相當受重視的「六個標準差」(6 sigma) 品質管制也與統計學有關。根據六個標準差之概念，一般流程之能力約在三個標準差，其不良率約在百萬分之 66807；當進步到四個標準差時，不良率便為百萬分之 6210；五個標準差時，不良率為百萬分之 233；而達到六個標準差時，則只有百萬分之 3.4。六個標準差的目標在於挑戰極限、追求完美，經由將流程操作的可能失誤減少到最低限度，以使組織做到：品質與效率最高、成本最低、流程循環時間最短、利潤貢獻最大，及全方位顧客滿意。商業界對顧客的消費行為進行分析，常常也有助於提高營業額*。

醫學界不管是新藥的發明，或是新的治療方法之採用，在普及化之前都需要嚴謹的實驗，畢竟人命關天不可不慎，此時容許犯錯(第一類型錯誤, type one error)的機率就應訂得更小(理論上應是不容許犯錯，但是事實上幾乎不可能)。

學會適當使用統計方法是進行量的研究的必要條件(necessary condition。換言之，不會使用適當的統計方法，就無法進行量的研究)，但絕對不是充分條件(sufficient condition)。研究者會因為使用了不當的統計方法而使研究成果品質不佳；但是一個設計不當的研究，絕不會因為使用了高深的統計方法而有所改善。所以，妥善規劃並細心執行仍是研究重點，統計只是個工具。除了專注於統計學原理的研究者外，統計學極少成為研究的最終目的。

此外，比較精確的統計方法當然可以提供更多的訊息，讓研究者了解更多的資訊；但是，研究者也應體認：統計方法的運用以適切為最重要，千萬不要只為了炫耀學識而賣弄複雜的統計方法，那就捨本逐末了。

參、學術溝通的需要

實務工作者或學術研究者，不管是在平時的口語溝通或是閱讀各種研究報告，經常都會碰到統計學的名詞或觀念，如果對統計學不了解的話，就很難對這些內容(語言或文字)有所領會。近年來，行為及社會科學的論文經常會使用一些

*一家國外知名百貨公司分析客戶的資料庫後發現，最多與尿布一起購買的商品是啤酒。因此各賣場就將兩者放在一起，結果尿布和啤酒的銷售量都增加了 30%(楊蕙菁, 2004.3.8)。

較複雜的統計方法 (例如：結構方程模式 (structural equation modeling, SEM) 或是階層線性模式 (hierarchical linear model, HLM))，對這些統計方法我們不一定要會計算，但是對它的概念一定要了解，不然常會發生「每一個字都看懂，但就是不知道在說什麼」的情形。所以近十幾年來美國研究所階段，統計學甚至取代傳統要求的外國語言成為必修學科 (Glass & Hopkins, 1996, p.2)。

肆、科學訓練的需要

科學家常會說：「數學是科學之母。」數學重要的功能之一是學習如何解決問題，統計學本來就與數學有密切的關係；而且目前統計學比較著重在問題解決的過程，所以在學習統計學的過程中常會使用推理 (含演繹及歸納) 及思考的方法，這也是一種科學的訓練。

根據報紙報導：「九十學度國民中小學中輟學生有 9,464 人，其中臺北縣 1,376 人，佔了 14.5% 最多。」你會做何聯想？是不是臺北縣國民教育階段學生中途輟學的問題比較嚴重？如果你得知：九十學度臺北縣國中小學學生佔了全國的 16.8%，則你的想法又是什麼呢？

又如 (此為虛設例子)：研究者分析大學一年級新生的家庭社會階層 (分為 5 個階層) 後發現，來自最低兩個階層的學生佔了 50%，因此最後獲得「大學聯考非常公平，因為來自中下階層的學生比例最高」的結論，你是否會相信呢？如果你得知：最低兩個階層的家庭在全國的比例為 60%，則是否會接受這個研究結論呢？

再舉一例子：假設有一家電視臺宣稱他們新一齣八點檔連續劇的「收視率」是前一齣的 3 倍，或許你會認為這齣新的電視劇滿受歡迎的。但是，如果你知道上一部戲的收視率是 0.1%，你會有什麼想法呢？

有一本書名為「真實的謊言」(張美惠譯, 1996)，內容在描述怎麼利用統計學來說謊，如果對統計學有正確的認識，受騙的可能性也會比較小一些。另外有一本名為「數字的陷阱」的書 (盛逢時譯, 2001)，描述福爾摩斯利用統計學知識辦案的過程，也是本可以一讀的科普著作。