

Biostatistics

生物統計學

郭寶錚、陳玉敏◎著



生物統計學

郭寶錚・陳玉敏 著

五南圖書出版公司 印行



五南

凝煉知識・品味閱讀

國家圖書館出版品預行編目資料

生物統計學／郭寶鏗，陳玉敏著．-- 二版，--

臺北市：五南，2004 [民93]

面；公分

參考書目：面

ISBN 957-11-3624-7 (精裝)

1. 生物醫學統計學

360.13

93008882

生物統計學

ISBN 957-11-3624-7

作者 郭寶鏗・陳玉敏
編輯 李敏華

發行人 楊榮川
總編輯 王秀珍
主編 翁千雅
出版者 五南圖書出版股份有限公司
地址 台北市大安區(106)和平東路二段339號4樓
電話：(02)2705-5066 傳真：(02)2706-6100
台中駐區辦公室 台中市中國區中山路6號
電話：(04)2223-0891 傳真：(04)2223-3549
高雄市駐區辦公室 高雄市新興區中山一路290號
電話：(07)2358-702 傳真：(07)2350-236

網址 <http://www.wunan.com.tw>

電子郵件 wunan@wunan.com.tw

劃撥帳號 01068953 戶名：五南圖書出版股份有限公司

法律顧問 財團法人資訊工業策進會科技法律中心

出版日期 2000年9月初版一刷
2004年3月初版六刷
2004年9月二版一刷
2006年3月二版四刷

定價 新臺幣 400 元

※版權所有，欲利用本書全部或部分內容，必須徵求本公司同意※

再版序

本書一版獲得市場廣大的迴響，使我們感到欣慰，但這是一種責任。此次再版，我們修正一版書中部分不妥或錯誤的內容，以求正確及完善，另外也增添了 Excel 的應用，希望能藉 Excel 所提供的統計功能，幫助讀者更加熟悉所介紹的統計觀念，本版書中也加入了索引，將有助於讀者查考相關內容。

作者一直希望能提供一本深入淺出，內容清晰詳實且實用的生物統計學教材，供有心學好生物統計學的相關人士使用，但作者才疏學淺，相信書中仍有不周全之處，企盼先進能隨時加以指正，以供未來修正之參考。

作者 謹識

2004 年 7 月

推薦序

郭老師自中興大學畢業後，負笈美國深造，取得愛荷華州立大學統計學碩士及維吉尼亞聯邦大學--維吉尼亞醫學院生物統計學博士後，回國到母校農學院擔任教職工作，在任教期間也發揮所長，曾擔任中華民國超音波學會統計諮詢及維吉尼亞臨床研究組織(VCRO)生物統計諮詢等服務工作。

環顧國內有關生物統計的書籍，能以深入淺出的介紹，並且有系統的整理者，實不多見。郭老師從事生物統計的教學、研究與服務多年，在生物統計領域中是幾位年青有為的佼佼者之一，郭老師將所學應用在公共衛生、醫學、護理及生物學上，並將從事多年的教學經驗寫成「生物統計學」這本書，誠屬難能可貴。郭老師的夫人陳玉敏老師，目前服務於中國醫藥大學護理學系，從事老人長期照護相關研究，共同參與本書的撰寫，而使本書更臻完整。

這本書的特點在於他有極高的可讀性，郭老師以清晰易讀的字句，論述由淺入深、劃分清晰，並將重點有系統地加以詳細描述，寫成一本內容廣泛且富含專業知識的書，這本書對於初入門的同學是一本極優良且難得的課程教材，也是提供從事此方面相關人員的參考書。

賴俊雄

中國醫藥大學教授兼副校長暨公共衛生學院院長

自序

曾經聽過學生詮釋所謂的「統計學」為「通通忘記之學」。的確，當需要面對這麼多的定理公式時，要想不「通通忘記」也難。自己曾思考過為什麼大多數的學生會視統計學為畏途呢？我想其中原因固然很多，但缺乏適當的中文參考教材為主要原因之一。如以生物統計學教科書為例，坊間已有數本不錯的教科書可供選擇，撰寫此書的目的，主要由於作者曾教授過不同背景的學生，包括大學生、碩士班研究生，甚至博士班的在職醫生，另外也常為醫生及研究人員做統計諮詢的工作，深感大多數的學生、醫生及研究人員，都有想把生物統計學「搞懂弄熟」的迫切需求，因此想藉由過去教學及研究經驗，整理出一本以「觀念為導向」的生物統計學教科書，以供大家參考，期貢獻棉薄之力於生物統計觀念的推廣與普及。

本書除適合技術學院、大學及研究所之醫、護、公衛及農學等相關科系學生使用外，另對有心自修者亦很合適。由於各領域的修習學分不同，因此授課老師可自行就教學所需加以取捨，以利教學。本書例題多數具有關聯性，希望藉此建立學習者整體性的觀念。

這本書的順利完成，除要感謝中國醫藥學院賴主任俊雄的鼓勵與題序，也要感謝多位學生幫忙打字及繪圖，倉促完成，疏忽與錯誤在所難免，還希望先進不吝賜教，以期能盡善盡美。

郭寶錚 謹誌

民國八十九年九月四日於

中興大學農藝學系生物統計研究室

目錄

第1章 緒論 1

- 一、生物統計學的定義及重要性 [3]
- 二、統計資料 [4]
- 三、變數、觀測單位及觀測值 [5]
- 四、量測尺度 [5]
- 五、測量工具的信度與效度 [7]
- 習題 [9]

第2章 抽樣及實驗 11

- 一、抽樣觀念的介紹 [13]
- 二、使用抽樣的原因 [14]
- 三、使用抽樣時所產生的誤差 [15]
- 四、抽樣方法 [15]
- 五、實驗與樣本調查 [20]
- 六、隨機比較實驗 [20]
- 七、區集設計 [22]
- 習題 [23]

第 3 章 資料的整理、摘要與呈現 ... 25

一、未分組資料及分組資料 [27]

二、次數分布表 [28]

三、圖 [33]

習題 [42]

第 4 章 資料集中趨勢及變異性的測度 43

一、集中趨勢的測度 [45]

二、變異性的測度 [51]

習題 [55]

第 5 章 機率理論 57

一、實驗及事件 [59]

二、事件的運算 [60]

三、機率的定義 [61]

四、機率加法規則 [63]

五、條件機率 [65]

六、機率乘法規則 [66]

七、獨立事件 [66]

八、貝氏定理 [67]

九、敏感度、特異性、偽陰性率及偽陽性率

[68]

習題 [70]

第 6 章 隨機變數及分立機率分布 73

- 一、隨機變數的種類 [75]
- 二、機率分布 [76]
- 三、分立隨機變數的期望值與變異數 [76]
- 四、二項分布 [79]
- 五、布瓦松分布 [84]
- 六、兩個隨機變數的聯合機率分布 [86]
- 七、共變異數 [89]
- 八、兩隨機變數間互相獨立 [90]
- 九、兩隨機變數的期望值及變異數性質 [91]
- 習題 [93]

第 7 章 連續機率分布及常態分布 ... 95

- 一、連續機率分布 [97]
- 二、常態分布 [99]
- 三、常態分布近似二項分布 [107]
- 習題 [109]

第 8 章 抽樣分布 111

- 一、樣本平均數的抽樣分布 [113]
- 二、中央極限定理 [114]
- 三、兩樣本平均數差的抽樣分布 [116]
- 四、樣本比例的抽樣分布 [118]
- 五、兩樣本比例差的抽樣分布 [120]

|| 習題 [122]

第 9 章 估 計 125

- 一、點估計 [127]
- 二、單一母群體平均數的信賴區間 [128]
- 三、t 分布 [131]
- 四、兩個母群體平均數差的信賴區間 [135]
- 五、配對樣本下兩個母群體平均數差的信賴區間 [139]
- 六、單一母群體比例的信賴區間 [141]
- 七、兩母群體比例差的信賴區間 [142]
- 八、單一母群體變異數的信賴區間 [143]
- 九、兩個母群體變異數比值的信賴區間 [145]
- 重點提示 [148]
- 習題 [150]

第 10 章 假設檢定 153

- 一、基本概念 [155]
- 二、單一母群體平均數的假設檢定 [161]
- 三、兩個母群體平均數差的假設檢定 [167]
- 四、配對樣本下兩母群體平均數差的假設檢定 [171]
- 五、單一母群體比例的假設檢定 [173]
- 六、兩個母群體比例差的假設檢定 [174]
- 七、母群體變異數的假設檢定 [176]
- 八、兩個母群體變異數比值的假設檢定 [178]
- 重點提示 [180]

習題	182
----	-----

第 11 章 檢力及樣本大小的決定 185

一、錯誤的種類	187
二、檢力	190
三、樣本大小的決定	192
習題	194

第 12 章 卡方檢定 195

一、適合度檢定	197
二、獨立性檢定	199
三、Fisher's 精準檢定	204
四、McNemar's 檢定	207
習題	209

第 13 章 變異數分析 211

一、單向變異數分析	214
二、多重比較	217
習題	221

第 14 章 迴歸與相關 223

一、簡單線性迴歸	226
二、複迴歸	234
三、相關係數與決定係數	235

四、邏輯迴歸及勝算比 238

習題 242

第 15 章 無母數檢定 243

一、符號檢定 245

二、Wilcoxon 符號等級檢定 248

三、Wilcoxon 等級和檢定 250

四、Kruskal-Wallis 檢定 252

五、Spearman 等級相關係數 255

習題 258

附錄 1	Excel 之應用	261
附錄 2	連加符號	319
附錄 3	習題解答	327
附 表		345
參考書目		369
索 引		373

第 1 章

緒 論

統計學(statistics)是一門探討如何蒐集、整理及呈現資料(data)，並藉這些資料做出推論與預測的科學。其中資料的蒐集、整理與呈現屬於敘述統計學(descriptive statistics)的範圍。分析所蒐集到的資料並做出推論與預測，則隸屬推論統計學(inferential statistics)的研究領域。經由機率理論(probability theory)的貫穿其中，而形成了統計學的基本架構。除此之外，統計學就其所應用領域的不同，大致可區分成商業統計學、工業統計學、教育統計學、社會統計學及生物統計學等。

一、生物統計學的定義及重要性

生物統計學(biostatistics)主要是將統計學運用於生物(biological)或生命(life)科學等領域。近來由於人們對其周遭的環境品質越來越關心，也更重視疾病的預防及衛生服務的提供，並且積極研發新的醫藥技術等因素，生物統計學在醫學、藥學、護理學及流行病學等研究領域已被廣泛的應用，所以能快速的发展成為統計學的一支。

我們在日常生活中，常會接觸到如下的訊息：

中央健保局 90 年西醫門診資料分析結果顯示，國人一年平均就醫 14.1 次；其中有 3.46 次是看一般感冒，另外還有 0.15 次是看流行性感冒。

根據內政部警政署的統計，91 年我國每 10 萬人當中，道路交通事故死亡率為 19.2 人，高於新加坡的 5.2 人及日本的 6.6 人。

衛生署管制藥品管理局公布 92 年台灣地區藥物濫用統計，醫療院所通報的案件共 8256 件，較 91 年增加 602 件。藥物濫用最大宗為海洛因，佔 8 成，其次為安非他命、

FM2、搖頭丸、強力膠等。

1111 人力銀行於 93 年 5 月，進行護理人員辛酸問卷調查，共回收 1622 份有效問卷，在 95%信心水準下，誤差值為正負 2.4%。結果顯示，有 78.61%的受訪護理人員想要轉業，原因包括職業倦怠、壓力大以及工作乏味等。

上述這些訊息對大眾、醫護人員、研究人員和行政人員，甚至立法者都有直接、間接的助益。但前面已提過，資料的蒐集、整理、呈現以及推論和預測，都需藉由統計學，但是如何正確解讀及看待這些訊息，也唯有靠著對統計知識的了解，以及統計觀念的推廣才能達成。隨著時代的進步，在各個領域當中，有愈來愈多的人要接觸或處理統計資料，不論是自己處理或尋求統計諮詢，甚至單單只是接受經統計處理後所得到的結果。當面對所得到的結果時，人們往往有二種極端的反應，一種為全然否定，總認為統計只是數字遊戲；另一種則為不論結果的來源及品質良窳，全然接受。因此如能具備正確的統計觀念，並熟悉統計的實施步驟及處理的限制等相關課題，相信必能提昇對統計分析所提供訊息的判斷能力及正確的態度。

二、統計資料

統計學中所處理的資料，可能是由量測(measurement)或計數(count)所構成的數值型資料(numerical data)，例如病患的收縮壓、舒張壓等量測，或醫院中急診人數及空病床數等計數的數值型資料。此外，亦有依照某些規則以區分成不同類別的類別型資料(categorical data)，例如病患的性別及血型等，均可區分成不同的類別，而我們經常以數字來代表各個不同的類別，例如當性別是男性者以 1 代表，性別是女性者則以 2 代表。至於資料的來源，可經由樣本調查、實驗、長

期儲存的記錄及已發表的文獻、報告，或是商業資料庫等。

三、變數、觀測單位及觀測值

在統計學中，我們常把所觀察或測量的特性，稱為變數(variable)。之所以稱為變數的原因，主要是因為所觀察或測量的特性會產生不同的結果。所觀察或測量的對象，我們稱之為觀測單位(observational unit)，而變數的結果則構成觀測值(observation)。例如某校獸醫學系5位大一新生入學時身體檢查的資料如下：

學生	性別	血型	身高	體重	今年內就醫次數
1	男	A	175	67	0
2	男	AB	169	62	1
3	女	O	160	51	0
4	男	B	181	70	0
5	女	O	164	53	2

以上的資料組是由5位新生所構成，每一位新生都是一個觀測單位，我們分別記錄其性別、血型、身高、體重，以及今年就醫次數等5個變數的結果，而每一個變數的結果就是一個觀測值，例如學生1，其有5個觀測值，分別為：(男，A，175，67，0)。

四、量測尺度

變數所使用的量測尺度一般包括類別尺度(nominal scale)、序位尺度(ordinal scale)、等距尺度(interval scale)及等比尺度(ratio scale)等四種。因為變數為所觀察或測量的特性，因此為了使用正確的統計方法，對於變數所使用的量測尺度應該有所了解。