

五南出版

巷子口統計學

第2版

懂統計 更安心！



許玟斌 著

- 因為算命的告知八字奇佳，命格富貴，就能遊手好閒嗎？
- 名人生日真的都集中在某幾個星座嗎？
- 媒體報導，市售填裝飲料所含菌量超標的比率高達60%，媽媽買得安心嗎？

知道許多人對於明牌的追求，是因為認為中獎數字組合確實能夠被預測。

新貴對於自己的薪資，落於哪個百分比會感到滿意？

實例出發，以插圖輔助說明，自我培養解讀數字的能力，就是這麼簡單。



知識經濟時代的來臨，無論是學術研究、職場應用或日常生活，具備基本的統計素養和邏輯思考，是人人都需要的。但是就一般大眾而言，對於新聞或報章雜誌上，統計訊息的解讀，多數人都是似懂非懂或照單全收。即使大專院校有統計這門學科，因統計資料分析過程，具有高度的連貫性，學習過程中若稍為疏忽，就可能知其然而不知其所以然。此書的目的，就是為了突破傳統教學的窠臼，從有趣而生活化的案例，讓讀者發覺統計應用的廣泛與實用性。透過合邏輯的資料分析，培養問題解決的能力。更進一步說，統計是一門讓數字說話的科學與藝術。



五南文化事業

ISBN 978-957-11-7427-3 (510)



9 789571 174273

RM20

財經・統計

巷子口統計學

第2版



許玟斌 著

五南圖書出版公司 印行

巷子口統計學

作 者 許玟斌
發行人 楊榮川
總編輯 王翠華
主 編 張毓芬
責任編輯 侯家嵐
文字校對 陳俐君
封面設計 盧盈良
封面插畫 蕭育幸
出版者 五南圖書出版股份有限公司
地 址 106台北市大安區和平東路二段339號4樓
電 話 (02)2705-5066
傳 真 (02)2706-6100
劃撥帳號 01068953
戶 名 五南圖書出版股份有限公司
網 址 <http://www.wunan.com.tw>
電子郵件 wunan@wunan.com.tw
法律顧問 林勝安律師事務所 林勝安律師
出版日期 2011年7月初版一刷
2014年1月二版一刷
定 價 新臺幣350元

※本書初版書名為《統計原來這麼生活》

國家圖書館出版品預行編目資料

巷子口統計學／許玟斌著．— 二版．— 臺北

市：五南，2014.01

面；公分

ISBN 978-957-11-7427-3（平裝）

1. 統計學 2. 通俗作品

510

102023682

推薦序

知識經濟時代來臨 不懂統計就落伍了

前東海大學管理學院院長 沈均生博士

二十一世紀是跨入知識經濟時代的關鍵時刻，統計則是把資料轉換成資訊(information)和知識(knowledge)最有效的一種方式。當電腦網路技術爆炸性的發展，資料可以大量即時且無遠弗屆的收集與儲存以及進行快速的計算與繪圖的同時，實用的統計方法就成為最普遍又便利的工具。知識之定義很廣，在資訊管理領域中，界定「知識」以實務的角度來說：舉凡具有相關性和行動性，能解決某特定問題的資訊集合，也都界定為知識，而統整於商業智慧(business intelligence)之範疇。如果管理者具備運用知識的能力，也就是智慧(intelligence)，當然可以形成解決問題的最佳策略。所以資訊和知識將變成一個企業組織的重要資產，大有取代資本之勢。統計之重要性也會隨著全球化的步履，面臨激烈競爭之際更加彰顯。在此新紀元之始推出這本書，甚具意義，個人額手稱慶樂觀其成。

作者在首章透過統計的基本觀念和名詞定義中，闡述了資料、資訊和決策品質之相關性。資料必須經過嚴格的分析流程，轉換成有用的資訊和知識，才能化解我們面臨資料超載的困境。此書展現了作者二十年來教學研究，橫跨資訊電腦和統計兩

大領域之學養和功力，融會貫通所使然。

此書另一特色是章節的安排，突破一般論述統計的順序。第二章就先談統計分析流程，以系統導向(system approach)的方式去建構一個研究模式(model)。針對問題明確定義系統範圍、目標與資源限制等，本來就是科學上作各種研究的基石。簡單說，就像在遊戲比賽前先制定一套嚴謹清楚的遊戲規則。然後介紹分析步驟：取樣設計、收集資料、彙整與呈現資料的特徵、參數估計與假設檢定等。而就在第三章到第七章對流程中每一階段，所需的統計理論、概念與方法，逐一條理分明詳細的論述。這種敘述方式在閱讀上可以加強了解和記憶，對統計分析的結果和推論，也能更精確去闡釋。達到知其然，也知其所以然的深度，例如第八章打破彩券有明牌的迷思的推論流程。

從第八章到第十二章之大大小小的統計應用之個案，這些案例都會面對大量原始數據，而是無法直接去解釋和推論的，也是直覺上之局限。面對電腦網路排山倒海的產生資料，人類資訊超載的焦慮日益嚴重，但透過統計的認知，你會驚訝的發現這些案例都能符合統計中最基本又重要的機率分配函數(probability function)：常態(normal)、波亦松(Poisson)、指數(exponential)、柏氏(Bernoulli)、二項(binomial)、均值(uniform)以及離散均等(discrete uniform)。

再根據符合的分配函數理論走過嚴謹的分析流程，看似雜亂無章的原始數據就會把隱藏的有價值的資訊和知識，源源不絕的產生出來了，統計分析真是蠻神奇的。知識經濟時代，我個人認為不但要讓數字說話，更要借助統計電腦網路的運用，讓知識資訊去面對問題，進而採取適當行動和政策。如此能提升人類生活

之品質，也不算太誇張了。

從這些有趣的案例也會發現統計應用之廣，原來統計和人生是如此貼近，無論學術研究上或職場上，即使生活上都息息相關。知識經濟時代之來臨，真如作者所言，人人都需要具備相當程度的統計素養和科學邏輯的思考，也許這是書名的來源。這本書應是一種使命感所驅使的產物，樂於為序，與有榮焉。

自序

統計，是科學，也是藝術

統計是一個通俗的名詞，然而對於新聞與雜誌傳播有關的統計數據資訊，大多數民眾卻只能一知半解的懷疑其準確性，或似懂非懂的照單全收。雖然各級學校也有開設介紹統計基本觀念的課程，但由於統計資料分析的流程具備高度的嚴謹性與連貫性，使得稍為疏忽與幸運之神特別眷顧的學生，可能在望而興嘆之中迷迷糊糊的獲取學分。

統計是一門歷史悠久、應用廣泛的學科，適合於不同程度或背景知識的教科書或論文。但為什麼許多人在學習時卻無法累積足夠的統計基本知識？是否由於大部分介紹基本統計觀念的教學與教材，使用比較教條式的敘述方法：依循名詞定義、定理介紹，舉例說明與習題演算的流程所致？雖然求學吸收知識本來就是一件嚴肅與持續努力的過程，但是多少年輕學子能夠了解與力行。另外有些職場人士，也許從未接觸統計，也許沒有充裕的時間仔細閱讀，也許找不到合適的入門書，以充實自己應用統計方法的能力。

撰寫本書的目的就是為了滿足上述不同的研習需求。一般介紹統計方法、進行資料分析的書籍，比較著重於統計方法的敘述與應用。本書則比較著重於統計方法中的「為什麼」層面的探討。例如，為什麼要在乎資料收集的方式、為什麼要估計系統中

不確定因子的未知參數，或者為什麼需要估計某事件發生的機率？類比於武俠小說的術語，前者比較屬於固定劍招的傳遞，本書則是劍訣導向的養成方式。

根據此一宗旨，本書內容分成三個部分，依序介紹統計基本原理與應用。第一部分先以實際例證說明統計方法的重要性與限制，並定義統計分析的主要步驟。第二部分的重點圍繞在統計資料分析流程中，每個步驟的內涵與意義的介紹與說明。第三部分著重於如何運用第二部分的統計資料分析觀念，並實際探討一些看似普通卻有趣的問題。

為什麼進行資料分析時，必須嚴謹遵守一貫的流程？簡單的說，因為我們希望藉由資料，產生具備說服力、完整與正確的資訊。這種正確且完整的知識，正是人們作出適當決策的依據。適當決策的重要性，應該是淺而易懂的道理。但是，為什麼形成適當決策的設計卻是相對的困難重重？主要原因是，我們面對的決策問題大多充滿了不確定性或變異性的因子。例如選舉結果的預測，颱風路徑的變化、影響地區與範圍以及風速與雨量強度的估計，投資組合的選擇，公共政策的效果，農作物的培育，醫療衛生藥品劑量的訂定與政府政策的制定等，無一不具有不確定性或變異性的因素。

不確定性的另一種說法是，事件的出現或發生具有隨機性或不可預測性，表示或儲存這類物件性質的度量結果，就稱為隨機變數。而在一般的共識中，處理自然現象的不確定性與度量物件性質的變異性，最主要的工具就是統計。因為統計方法的功能就是萃取資料中隱藏的資訊，包括敘述資料性質的特徵值，視覺化的圖形與表格；隨機變數的機率行為模式；未知參數的估

計值；以及某問題的陳述與實際資料的隨機行為互相矛盾的機率等。因此，統計方法就是一種將觀察與紀錄的數據轉換成為輔助決策資訊的工具。

統計資料分析雖然不能直接預測某個事件會不會出現。不過在了解問題中的隨機因子，並建立一個能夠描述這個系統的機率模式之後，的確能夠用來估計某事件發生的機率。

面對一個充滿變異的真實環境，無論是哪一個行業的知識工作者，沒有基本統計的素養，實在不容易勝任職場上與生活上的挑戰，更不用奢談能夠在職場上出人頭地。作者相信，了解本書內容者，無論在生活上與工作上都能夠正確的應用統計分析的方法，解決各自面對的問題，或是至少了解為什麼有些問題的解答並不符合科學精神。

作者深深以為，學習統計就是養成以科學邏輯的資料分析流程，思考解決問題的習慣與累積讓數字說話的藝術與科學的知識。

許玟斌

簡要目錄

推薦序：知識經濟時代來臨 不懂統計就落伍了 沈均生
博士 iii

自序：統計，是科學，也是藝術 vii

第一部分 什麼是統計？

1 統計和我有關嗎？——統計、資訊與決策 3

2 怎麼統，如何計——統計分析流程 29

第二部分 悠遊統計大觀園——統計方法

3 他的民調比較高？——資料收集與樣本品質 59

4 神奇的數字會說話——資料彙整與呈現 87

5 地球自轉一周的時間每次不一樣？
——隨機變數與機率函數 117

6 樣本分配與參數估計 171

7 拿出證據，否則無罪！——假設檢定
213



第三部分 在生活中與統計相遇——個案應用

- 8 彩券號碼算得出來嗎？——從均等與均值分配 253
- 9 比較臺南與恆春年雨量——常態分配 275
- 10 我們吃的食物安全嗎？——柏氏與二項分配 325
- 11 某年颱風侵襲的次數可以計算嗎？——波亦松分配 349
- 12 計算有發警報颱風間隔時間的機率——指數分配 371
- 結語——統計推論的假設 391

目錄

推薦序：知識經濟時代來臨 不懂統計就落伍了 沈均生
博士 iii

自序：統計，是科學，也是藝術 vii

第一部分 什麼是統計？

- 1 統計和我有關嗎？——統計、資訊與決策 3
 - 成為社會贏家第一步——掌握資訊、正確決策 4
 - 正確決策的指南針：統計與資訊 6
 - 掌握三大基本術語，輕鬆和統計Say Hello！ 8
 - 資料與資訊 8
 - 系統與模式 10
 - 樣本與母體 12
- 到底，什麼是統計學？——統計學定義 15
- 生活無處不統計——統計應用實例 19
- 2 怎麼統，如何計——統計分析流程 29



一目了然的統計分析流程圖	30
開始統計的第一步——定義問題與建立模式	32
民意調查可信度的關鍵——收集資料	36
取樣設計——決定研究過程的品質與效果	37
試驗設計——因果關係建立的重要過程	39
度量個體性質——決定資料品質與算術運算意義	41
表格、圖形的作用——彙整與呈現資料	43
颱風到底在哪兒登陸？——參數估計有答案	48
無罪推定的檢查機制——假設檢定	51
最終目的——形成結論	53

第二部分 悠遊統計大觀園——統計方法

3 他的民調比我高？——資料收集與樣本品質	59
資料來源	60
應用有限的歷史紀錄	61
大肚溪沿海，冬季候鳥數量——直接觀察	65
加入特定條件的試驗樣本	65
無法普查下的解決之道——取樣設計	66
紙牌的分派牌型——簡單隨機取樣設計	67
某商店某日每位顧客平均消費額——系統取樣設計	69
某地區國小高年級學童月平均零用金——分層抽樣設計	70
鄰、路、街、段——分團取樣設計	71
多階段取樣設計	72
某大學生族群的讀書風氣性向——非機率取樣設計	73

颱風路徑、雨量、風速轉變——隨機樣本 74

下午茶的意外收穫——試驗設計 75

一群人喝酒，脈搏與血壓改變各不同——完全隨機設計
77

性別、社區、產品——隨機化區塊設計 80

生活中最頻繁的活動——度量個體性質 81

預報溫度24度，卻感覺好冷——度量有效性與準確性
82

男性、學歷、身高、視力——不同度量尺規的應用 85

4 神奇的數字會說話——資料彙整與呈現 87

為什麼賺不到平均國民所得——數值表示方式 89

178公分的兒子身高不如187公分的爸爸——趨中性 89

來自世界各國的觀光客——分散程度 98

昨天美國的股票大漲，今天要大量賣出——兩個變數之
間的關聯 100

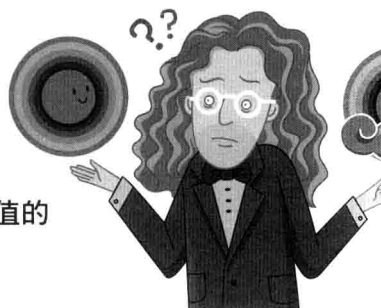
哪一種表格比較好？——表格表示方式 102

傳遞更多的視覺資訊——圖形表示方式 109

線圖：以直線連接變數在相鄰的時空的反應值 109

圓形圖：以圓形代表變數所有數
值的整體，扇形大小
代表某屬性的比例
110

條狀圖：以線條的長度表示觀察值的
變化 111



	直方圖：以連續長方形代表變數落入不同數值區間的分配	112
	莖葉圖：一種將觀察值分成兩部分，左部分稱為莖，而右部分稱為葉，然後將所有觀察值根據莖部在葉部依次排列的呈現方式	114
	盒子圖：五數彙整的圖形表示方式	115
	母體資料彙整	116
5	地球自轉一周的時間每次不一樣？ ——隨機變數與機率函數	117
	彩券號碼期期都不同——隨機現象與隨機試驗	118
	明天一定會下雨，今天必須將橘子採收完——計算機率	121
	投擲兩枚硬幣——理論法	122
	臺灣每年遭受中級颱風侵襲的機率——相對次數法	124
	感覺快樂或幸福——主觀法	125
	兩枚硬幣的例子——隨機變數	126
	隨機變數的種類	128
	隨機變數的分配函數	129
	隨機變數的參數	133
	隨機變數的期望值	135
	明星國中本區與越區就讀人數——聯合機率分配	142
	新生嬰兒身高與體重的比較研究——相關係數	143
	獨立隨機變數	146
	條件隨機變數	149

簡單實用的機率函數	151
擲骰子機率——均等分配	153
只有兩種結果——二項分配	155
平均每年颱風侵襲臺灣的次數——波亦松分配	159
某站牌公車抵達時間——均值分配	161
計算活多久的機率——指數分配	164
罐裝汽水的裝填量——常態分配	166
6 樣本分配與參數估計	171
樣本統計量	172
參數估計式	175
動差法	177
最大概似法	180
優良估計式的性質	183
點估計值	187
樣本分配函數	188
Z-分配	189
χ^2 -分配	190
t-分布	192
參數區間估計	194
建立信賴區間的理論基礎	195
變異數已知的常態母體平均數	197
變異數未知常態母體平均數	199
常態母體變異數 σ^2 的區間估計	202

