

天下文化
知識的世界

01

25%

4%

35%

80%»»

35%

25%

4%

Population*

02

50%

75%

50%

75%

2012年最新修訂版

Graphs*

03

80
75
62
55
50

統計學的世界 II

整合數據

STATISTICS

CONCEPTS AND CONTROVERSIES SEVENTH EDITION

Pie Charts*

04

30%

BY DAVID S. MOORE,
WILLIAM I. NOTZ

淡江大學數學系教授 鄭惟厚 譯

25%

50%

75%

15%*

Icons & Elements*

05



79
折優惠
天下文化

This is a statistics book for readers interested in ideas rather than techniques. It presents, in an accessible and readable way, the important statistical concepts as they are applied to public policy, the human sciences, and everyday life. It is designed to give non-mathematical students critical insight into the uses and misuses of statistics and

統計學的世界 II

整合數據

STATISTICS

CONCEPTS AND CONTROVERSIES
SEVENTH EDITION



by David S. Moore, William I. Notz
淡江大學數學系教授 鄭惟厚 譯

統計學的世界 II

原 著 / 墨 爾、諾 茨
 譯 者 / 鄭惟厚
 顧 問 群 / 林 和、牟中原、李國偉、周成功
 科學館總監 / 林榮崧
 責任編輯 / 林文珠 (第一版)、畢馨云 (第二版)
 封面設計暨美術編輯 / 江儀玲

出版者 / 天下遠見出版股份有限公司
 創辦人 / 高希均、王力行
 遠見·天下文化·事業群 董事長 / 高希均
 事業群發行人 / CEO / 王力行
 出版事業部總編輯 / 王力行
 版權部經理 / 張紫蘭
 法律顧問 / 理律法律事務所陳長文律師 著作權顧問 / 魏啟翔律師
 社址 / 台北市 104 松江路 93 巷 1 號 2 樓
 讀者服務專線 / (02) 2662-0012 傳真 / (02) 2662-0007 ; 2662-0009
 電子信箱 / cwpc@cwgv.com.tw
 直接郵撥帳號 / 1326703-6 號 天下遠見出版股份有限公司

排 版 廠 / 極翔企業有限公司
 製 版 廠 / 東豪印刷事業有限公司
 印 刷 廠 / 立龍藝術印刷股份有限公司
 裝 訂 廠 / 晨捷印製股份有限公司
 登 記 證 / 局版台業字第 2517 號
 總 經 銷 / 大和書報圖書股份有限公司 電話 / (02) 8990-2588
 出版日期 / 2002 年 10 月 30 日第一版
 2012 年 6 月 22 日第二版第 1 次印行
 定 價 / 400 元

國家圖書館出版品預行編目 (CIP) 資料

統計學的世界 / 墨爾 (David S. Moore), 諾茨 (William I. Notz) 作; 鄭惟厚譯. -- 第二版. -- 臺北市: 天下遠見, 2012.06

冊; 公分. -- (知識的世界; 1104-1106)

譯自: Statistics: concepts and controversies, 7th ed.

ISBN 978-986-216-959-9 (第 1 冊: 平裝). --
 ISBN 978-986-216-960-5 (第 2 冊: 平裝). --
 ISBN 978-986-216-961-2 (第 3 冊: 平裝)

1. 統計學

510

101009979

Statistics: Concepts and Controversies, Seventh Edition

by David S. Moore and William I. Notz

First published in the United States by W.H. FREEMAN AND COMPANY, New York

Copyright © 2009, 2006, 2001, 1997 by W. H. Freeman and Company

Complex Chinese Edition Copyright © 2002, 2012 by Commonwealth Publishing Co., Ltd.,

a member of Commonwealth Publishing Group

Published by arrangement with W. H. Freeman and Company through

Bardon-Chinese Media Agency.

All rights reserved.

ISBN: 978-986-216-960-5 (英文版 ISBN: 978-1-4292-2991-3)

書號: BW1105

BOOKZONE 天下文化書坊 <http://www.bookzone.com.tw>

簡明目錄 **Brief Contents**



第一部	產生數據	19
第 1 章	數據從何而來？	21
第 2 章	好樣本和壞樣本	43
第 3 章	樣本告訴我們什麼？	63
第 4 章	真實世界中的抽樣調查	89
第 5 章	實驗面面觀	121
第 6 章	真實世界中的實驗	147
第 7 章	資訊倫理	175
第 8 章	度量	201
第 9 章	數字合不合理？	229
第一部	復習	251

第二部 整合數據 265

第10章	好的圖及壞的圖	267
第11章	用圖形呈現分布	297
第12章	用數字描述分布	323
第13章	常態分布	355
第14章	描述關聯的方法：散布圖和相關係數	381
第15章	描述關聯：迴歸、預測及因果關係	409
第16章	消費者物價指數和政府統計	441
第二部	復習	467

第三部 機遇 511

第17章	考慮可能性	513
第18章	機率模型	537
第19章	模擬	555
第20章	賭場的優勢：期望值	577
第三部	復習	597

第四部 推論 607

第21章	什麼是信賴區間？	609
第22章	什麼是顯著性檢定？	641
第23章	統計推論的使用與濫用	669
第24章	雙向表及卡方檢定	689
第四部	復習	717



統計學的世界 II

目錄 Contents

Icons & Elements* 05



第二部 整合數據 265

第 10 章 好的圖及壞的圖 267

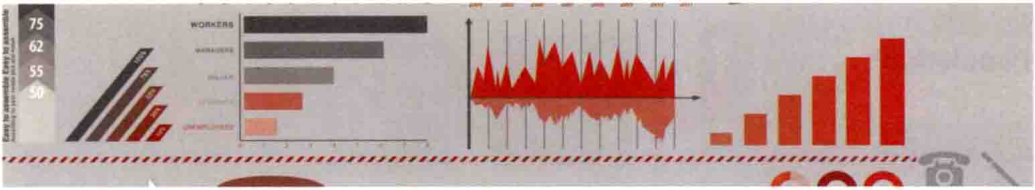
個案研究 267 數據表 268 圓瓣圖及長條圖 270
留意象形圖 274 隨著時間變動的線圖 275 注意刻度 278
怎樣把圖畫好 282 本章重點摘要 284 個案研究隨堂考 285
第10章習題 285 網路尋奇 295 注釋與資料出處 296

第 11 章 用圖形呈現分布 297

個案研究 297 直方圖 298 解釋直方圖 302 莖葉圖 308
本章重點摘要 311 個案研究隨堂考 311
第11章習題 312 網路尋奇 321 注釋與資料出處 322

第 12 章 用數字描述分布 323

個案研究 323 中位數和四分位數 324 五數綜合及盒圖 329
統計學上的爭議：貧富差距 333 平均數和標準差 335
選擇數值描述 339 本章重點摘要 342 個案研究隨堂考 343
第12章習題 343 網路尋奇 352 注釋與資料出處 353



第 13 章 常態分布 355

- 個案研究 355 密度曲線 358 密度曲線的中心和離度 360
 常態分布 361 68-95-99.7 規則 364 標準計分 366
 常態分布的百分位數 369 本章重點摘要 370 個案研究隨堂考 371
 第13章習題 372 網路尋奇 379 注釋與資料出處 380

第 14 章 描述關聯的方法： 散布圖和相關係數 381

- 個案研究 381 散布圖 383 詮釋散布圖 386 相關係數 390
 了解相關係數的意義 392 本章重點摘要 396 個案研究隨堂考 396
 第14章習題 397 網路尋奇 406 注釋與資料出處 407

第 15 章 描述關聯： 迴歸、預測及因果關係 409

- 個案研究 409 迴歸直線 410 迴歸方程式 413
 了解預測的意義 415 相關係數及迴歸 417 因果問題 420
 統計學上的爭議：槍械管制和犯罪 425 因果證據 426
 本章重點摘要 427 個案研究隨堂考 427
 第15章習題 428 網路尋奇 439 注釋與資料出處 439



第16章 消費者物價指數和政府統計 441

- 個案研究 441 指數 442 固定市場總覽物價指數 444
- 如何使用CPI 446 了解CPI的意義 450
- 統計學上的爭議：CPI把通貨膨脹誇大了嗎？ 453
- 政府統計的處境 454 社會統計的問題 456
- 本章重點摘要 457 個案研究隨堂考 458
- 第16章習題 458 網路尋奇 465 注釋與資料出處 466

第二部復習 467

- 第二部重點摘要 469 第二部復習習題 472
- 第二部報告作業 480 注釋與資料出處 482

「輪到你練習了」習題解答 483

習題解答（單數題） 494

表B 常態分布的百分位數 509

第二部

整合數據



文字本身不構成故事。得由作者把文字組成句子，再把句子組成故事。如果文字組合得不好，故事可能讓人看不大懂。數據也一樣，如果要讓人看清楚數字隱含的訊息，一樣需要經過整合。字用得太囁嚅，會讓主題變模糊而不是變清楚。一大堆的數據更是叫人難以消化，因此我們常常需要一個精簡的摘要，來凸顯出重要內容。應該如何整理、綜合及呈現數據，就是本書第二部的主題。

整理及彙整大量的事實時，最容易把事實扭曲，其中有些是無心的，有些卻是故意的。不管呈現事實用的是數字還是文字，上述情況發生的機會都差不多。我們將指出在呈現數據時，會讓不小心的人上當的一些陷阱。把統計看成說謊藝術的人，看統計時注意力都放在數據的彙整及呈現上。我們卻主張，誤導的綜合資料及選擇性的呈現資料，早在亞當、夏娃偷食禁果之後和上帝的對話中，就看得到了。不要怪統計。要記得那句老話：「數字不會說謊，但騙子會算計。」所以要小心。



10 好的圖及壞的圖



個案研究

日益升高的醫療保健支出成為大家關注的焦點已二十年了。醫療保健服務網試圖降低花費，而那些能成功做到這點、卻沒有犧牲醫療品質的服務網路，有理由覺得驕傲。

數年之前，圖 10.1 出現在美國某家地方報紙《哥倫布日報》上面。登載這則廣告的目的，是要讓讀者相信，翡翠醫療保健集團在控制會員保費支出的成長方面表現良好。保費在圖中呈現平緩成長、令人看了放心。事實真是這樣嗎？

統計處理的就是數據，而我們常用圖表來呈現數據。圖表可以幫助我們看出數據當中所含的訊息。不過，並不是所有的圖表都能夠正確呈現、或者清楚呈現數據當中所含的訊息。在本章當中你會學到呈現數據的基本方法，以及如何評估你在媒體上看到的圖表品質好壞。讀完這章你就能夠判斷，圖 10.1 是好的圖還是壞的圖了。



圖 10.1 翡翠醫療保健集團1993年6月8日刊登在《哥倫布日報》的廣告（此圖為重製）。



數據表

建議你看看《美國統計精粹》，它每年出一本，裡面有各式各樣的數值資訊。私立小學和中學的數目是不是有成長？這些學校的學生中，少數族裔占多少？過去幾年當中，每年共有多少人得到學士學位？這些學位若根據學習領域來分，或者根據畢業生的年齡、種族、性別來分的話，各有多少？所有這些以及更多其他的資訊都可以在《美國統計精粹》的教育一節裡找到。這些數據表把資料做了摘要。我們並不想看每一個大學學位的資訊，只想知道我們感興趣的一些類別當中的計數。

例1 怎麼樣的表才清楚？

30歲左右的年輕人，教育程度如何？表10.1呈現了25歲以上的人

的資料。這個表是數據表的一個好的示範。表的標示很清楚，所以資料的主題一目了然。主標題描述了資料的總主題，並且列出年份，因為這種資料會逐年改變。表裡面的標題闡明了變數，並且說明了度量變數所用的單位，例如，你可以注意一下，計數以千為單位。資料來源出現在表的底部。這份普查局發表的結果，內容資料事實上是從「當前人口調查」取得的。

表 10.1 25 歲以上人士的教育程度，2006 年

教育程度	人數 (千)	百分比
高中以下	27,896	14.5
高中畢業	60,898	31.7
大專肄業	32,611	17.0
專科畢業	16,760	8.7
學士學位	35,153	18.3
更高學位	18,567	9.7
總 數	191,884	100.0

資料來源：美國普查局《2006年美國教育實況》

表10.1先列出不同教育程度的25歲以上社會人士的計數。比率（或百分比、比例）通常要比計數清楚——聽到有14.5%的社會人士沒有讀完高中，比起聽到有27,896,000人沒有讀完高中，前者的資訊要清楚得多。表10.1也列出了百分比。中間和右邊的這兩行數字，用兩種不同方式呈現了「教育程度」這個變數的分布；每一行都在告訴我們這個變數可能有些什麼值（即6種教育程度），以及該變數等於其中每一個值的比率。

▶ 變數的分布

一個變數的分布（**distribution**），告訴我們變數有些什麼可能值，以及每一個值出現的比率。

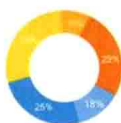
例2 捨入誤差

你有沒有檢查一下表 10.1 中的數字是否相合？人數的總數應該是

$$\begin{aligned} &27,896 + 60,898 + 32,611 + 16,760 + 35,153 + 18,567 \\ &= 191,885 \text{ (千人)} \end{aligned}$$

可是表裡面的總數是 191,884。怎麼回事？裡面每一單項的數字，在換成以千人為單位時經過四捨五入。因為是每一項個別做四捨五入，加起來和總數不合是正常的。在這章開始，這種**捨入誤差 (roundoff error)**在我們做算術的時候，都會一直跟著我們。

表裡面出現捨入誤差不算稀奇。比如說當表裡面的數字是百分比或比例的時候，總數加起來可能會和 100% 或 1 有微小差距。表 10.1 裡面的百分比加起來就等於 99.9%，而不是 100%。



圓瓣圖及長條圖

表 10.1 當中的分布很簡單，因為「教育程度」只有 6 種可能值。要把這個分布用圖來表示的話，可以用**圓瓣圖 (pie chart)**。圖 10.2 就是 25 歲以上人士教育程度的圓瓣圖。圓瓣圖可以顯示一個整體怎樣分為幾個部分。要畫圓瓣圖，先得畫個圓，圓代表全體，在這個例子中，就是所有 25 歲以上人士。圓裡面的楔形就代表各部分，各楔形的圓心角和各部分的大小成比例。比如說，有 18.3% 的人有學士學位但沒有更高的學位。因為一個圓的內角為 360 度，所以代表「學士學位」的楔形，圓心角就有：

$$0.183 \times 360 = 66 \text{ 度}$$

圓瓣圖的好處是讓我們看到，所有的部分合起來，的確是全體。但是角度比長度難比較，所以圓瓣圖並不是比較各部分大小的方法。

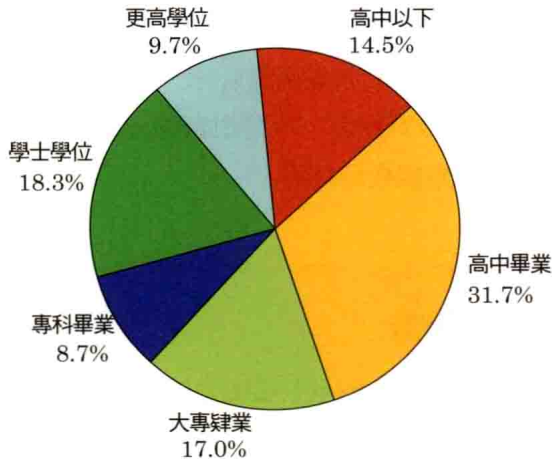


圖 10.2 2006年25歲以上人士教育程度分布的圓瓣圖。

圖 10.3 是根據同樣數據所畫出的長條圖 (bar graph)。每個長條的高度顯示出，25 歲以上的人當中，合於該長條底部標示的教育程度的人，占多少百分比。從這個長條圖可以清楚看出，高中畢業生比大專肄業的人多，因為「高中畢業」的長條比較高。這種差異在圓瓣圖的楔形之間不容易看出來，所以我們得在每個楔形旁都標示百分比。除非是用電腦繪圖，否則長條圖一般來說比圓瓣圖好畫。

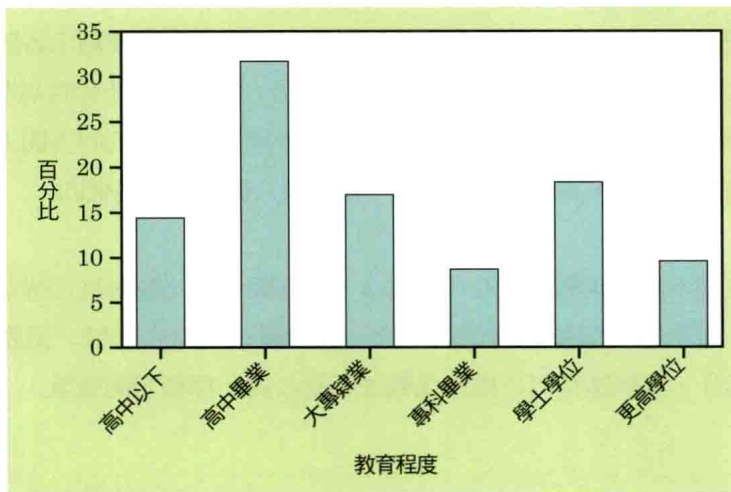


圖 10.3 2006年25歲以上人士教育程度分布的長條圖。

當我們在考慮各種圖的時候，把變數稍加分類會有幫助，有的變數具備有意義的數值刻度（例如身高幾公分、SAT分數等），而有的變數例如性別、職業或者教育程度，只是把個體分到不同類別而已。圓瓣圖和長條圖對於第二類變數最有用。

▶ 類別變數和數量變數

類別變數 (categorical variable) 是把個體歸類到其中一個組或類別。

數量變數 (quantitative variable) 的值是數值的，因此拿來做算術比如加法或平均的時候，是有意義的。

要表示類別變數的分布，可以用圓瓣圖或長條圖。

雖然圓瓣圖和長條圖都可以用來表示像教育程度這樣的類別變數的分布（不論是計數還是百分比），但長條圖的用途還是比較廣些。

例3 稅太高？

圖 10.4 比較了八大民主國家的稅負高低。長條的高度顯示各國的所得稅加上社會安全稅，占工資總額的百分比；這些數字計算的基準，是沒有小孩的單身普通勞工的工資水準。習於抱怨稅率太高的美國人恐怕會很驚訝的發覺，美國的稅率 23%，幾乎是八個國家中最低的。

我們不能用圓瓣圖來取代圖 10.4，因為圖 10.4 比較的是八個單獨的數量，而不是一個整體的各個部分。圓瓣圖只能用來比較一個整體的各個部分，長條圖卻可以用來比較並不屬於同一個整體的數量。

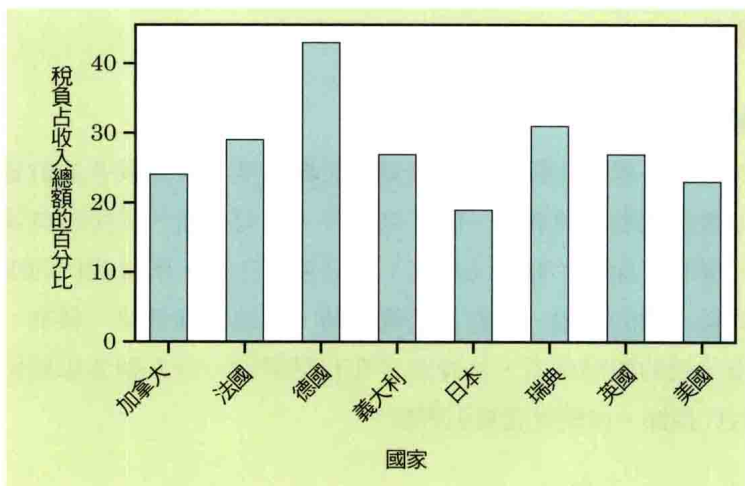


圖 10.4 2006年八個國家國民所得稅以及社會安全稅占收入總額的百分比。百分比的計算對象是沒有小孩的單身中等收入勞工。(數據來源是經濟合作發展組織，網址 www.oecd.org/document/17/0,3343,en_2649_34533_38148433_1_1_1_1,00.html。)

輪到你 練習了

10.1 稅。美國「退休生活資訊中心」整理出2007年50個州各自的預計州稅及地方稅占收入百分比的資料。下面的表列出了中西部某幾州的數據。

州名	稅占收入百分比
伊利諾	10.8%
印第安納	10.7
愛荷華	11.0
密西根	11.2
明尼蘇達	11.5
俄亥俄	12.4
威斯康辛	12.3

「州名」是類別變數還是數量變數？如果要呈現表中數據的分布，你會用圓餅圖還是長條圖？