

統計學， 最強的 商業武器

從買樂透到大數據，全都離不開統計學，
如果不懂統計學，你就等著被騙吧！！

本書最大宗旨在於
如何解讀統計數據資料！

不使用艱深的數學式，就能讓你
看穿數字背後的真相。

你一定聽過「數字會說話」，
但，數字經過加工，也可以騙人。
你也許只看結果不看過程，
但，任何統計都有你意想不到的疏漏。
數據能夠協助決策者做出判斷，
但，也會讓我們誤入歧途。

作者簡介

西內 啓

Nishiuchi Hiromu

畢業於東京大學醫學部，主修生物統計學。曾任東京大學大醫院醫學系研究科醫療傳播學領域的助理講師、大學醫院醫療資訊網路工程研究中心副主任、哈佛癌症研究中心客座研究員等職務。

目前則為各種以數據資料為基礎的社會創新專案，提供研究調查、分析、系統開發以及策略規劃等諮詢服務。

著作包括《科特勒教會我的事》（暫譯）、《上班族的多數煩惱都已有學術性的「解答」》（暫譯）、《世界第一簡單易懂的醫學統計》（暫譯）等。

譯者簡介

陳亦苓

政治大學廣播電視系畢，輔修日文，曾留學並於日本工作近 4 年。目前為自由譯者，擅長資訊類英翻中、日翻中。

個人作品列表：www.anobii.com/bready/books

統計學， 最強的 商業武器

著——西內啓

譯——陳亦苓

目次

PART 01 統計學，憑什麼是最強的武器？ 9

01 不懂統計學的人容易被騙 10

威爾斯的預言／鬼腳圖的必勝策略／掌握了統計學就等於掌握了全世界

02 唯有統計學能最快導出最佳解 17

統計學成為最強武器的理由／「流行病學之父」約翰·斯諾的行動／流行病學延長了人類壽命

03 所有學問都以統計學為基礎 26

「實證」改變了醫療觀念／實證原則亦可應用在教育上／統計學的影響力擴及棒球賽與經濟學

04 大數據時代來臨，讓統計學備受矚目 35

統計學為什麼到現在才嶄露頭角／為何每兩年才能調查一次／讓你對統計學失去興趣的「紙上作業」課程／「大數據」一詞興起的原因／未來十年內最熱門的職業

PART 02 能大幅降低資訊成本的抽樣調查 47

05 統計學家眼中的大數據狂想曲 48

被炒熱的專業術語／活用數據不必花大錢

06 準確度足以嚇死人的抽樣調查 54

二五%的失業率！／支撐新政的統計學家／普查 vs. 抽樣調查／有些人的想法跟八十年前一樣老舊

07 該為了一%的精確度而花費數千萬日圓嗎？ 61

對於抽樣調查的「常見反駁」／誤差的計算方法／即使增加一萬個樣本，標準誤差也只有〇・一%的變化／先從要做出正確判斷所需之最小資料量開始

PART 03 誤差與因果關係是統計學的重點 71

08 拋棄傳統的統計方法吧 72

結論是「喔……」的分析圖表／為了將數據應用於業務的「三個問題」／十九世紀以前的數據分析，都只是在做「彙總」罷了

09 這世界到處充斥著，沒有考慮因果關係的統計分析 79

漏洞百出的圖表／無確實根據的指標運用／哪種食物會導致死亡、犯罪與暴動？

10 改變投遞 DM 的方式，就能多賺六十億 85

買與不買的人，差別在哪裡？／「經驗」很不可靠／改變廣告郵件的散播方式，就能多賺六十億日圓

11 營業額真的提高了？還是只是誤差？ 92

不考慮「誤差」的試算都不切實際／所謂的「A/B測試」是一種親和度的比較／雖然出現了「〇・一％」的差距／輪到「卡方檢定」與「p值」上場了

12 到底該分析什麼樣的數據？ 100

找出密技的第三個關鍵／商業上的明確目標／它與獲利有任何直接相關嗎？／透過CPU的溫度分析來有效降低成本

13 因果關係是個大問題 106

因果關係是有方向性的／電動玩具與青少年犯罪，真有實質的因果關係嗎？／因為是「不公平」的比較，所以沒用／兩種解決方案

PART 04 統計學的終極必殺技——隨機對照實驗 113

14 奶茶是先倒牛奶還是先倒紅茶 114

隨機對照實驗擴大了「科學的目標對象」／為什麼非得隨機不可？／如何泡出一杯完美的紅茶／看穿超自然現象與騙局的方法／也可應用於培訓與廣告郵件的效果測定

15 隨機對照實驗是商場競爭的神器 125

科學來自「觀察」與「實驗」／有「誤差」的科學／對付「誤差」的三種方法／誕生於農場的「實驗設計法」

16 「買兩台就打九折」的銷售策略，真能提高營業額？ 134

以「進攻」為目標的統計學／有夠笨的「錯誤」判定／為公司賺進一・五億美元的客訴處理／隨機其實不容易

17 隨機的三項限制——現實、情感、道德 142

「現實」造成的限制／「道德」造成的限制／「情感」造成的限制

PART 05
好吧，無法做隨機對照實驗，又該怎麼辦？ 151

18 流行病學的進步，證明了抽菸的風險 152

請看看香菸盒吧／「病例對照研究法」的出現／來自天才費雪的反對意見／依據世界各地數據的再抗辯／對於「並不完全一致的條件」，底線在哪裡？

19 分析「回歸平凡」的回歸分析 164

何謂回歸分析／矮個子野村同學的戀愛故事／達爾文的表兄弟與優生學／發現「回歸平凡」的道理／奧運選手的最佳狀態

20 統計學之父的另一項成就 176

高爾頓的回歸分析也有限制／回歸係數本身就具有變動性／連統計學家也無法理解的「真值」概念／回歸分析專用的基本術語

21 讓統計學變得更易理解的一張表 188

統計學教科書中的廣義線性模型可分為兩大類／這張表的用法／以任何方式皆可獲得相同p值的原因／容易造成誤導的用詞「一般線性模型」

22 多元回歸分析與邏輯回歸 199

學者也很常用的統計方法／無法公平比較的辛普森悖論／雖然採用分層分析即可避免自相矛盾，但：／不必分層的多元回歸分析／採用比值比的邏輯回歸／只要懂回歸分析，就能篩去「信口胡謔」的言論

23 統計學家的極致因果推論 214

運用回歸模型時，需注意交互作用／陷入僵局的變數選擇工作／無限趨近於隨機的「傾向分數」

PART 06 統計學之間的殘酷戰爭 225

24 社會調查法 vs. 流行病學及生物統計學 226

統計學的六大領域／追求正確性的社會調查專家／追求「合理判斷」的流行病學及生物統計學家／永無止盡的爭論

25 創造了「IQ」一詞的心理統計學 238

「一般智力」的發明／將智力分成七種的多元智能理論／心理統計學家的想法和作法／「問卷調查」是心理統計學家的命脈／對IQ的結論

26 做為商業用途的資料採礦 254

資料採礦的歷史很短／「尿布與啤酒」的購物籃分析／與其採取購物籃分析，不如運用卡方檢定／誕生於人工智慧研究的進階技術／資料採礦專家為什麼會認為回歸模型「已過時」？／有助於「預測」的資料採礦

27 用於分析語言的文字探勘 272

被文獻計量學所否定的「莎士比亞II培根」理論／「形態分析法」與支撐著Google的「N-Gram」／文字探勘的商業應用／充分活用文字探勘的關鍵就在於其他的統計知識

28 著重演繹的計量經濟學與著重歸納的統計學 283

統計學與計量經濟學在「表面上」的差異／統計學與計量經濟學在「本質上」的差異／追求更理想模型的計量經濟學家／影響力漸增的計量經濟學

29 貝葉斯派與頻率論派在機率觀念上的對立 295

頻率論派的想法較單純／貝葉斯派會考慮「先驗機率」與「後驗機率」／與計量經濟學志同道合的貝葉斯統計／在垃圾郵件的判別上大顯神威的貝葉斯統計

PART 07 統計學，讓你得以站在巨人的肩膀上 307

30 用統計學找出最佳解 308

實證的層級／最高層次的實證為「系統性回顧」與「整合分析」／
「最佳解答」已經公布

31 讓我們來尋找實證 318

英文文獻資料的搜尋方法／顯而易見的課題

結語 328

PART _____

01

統計學，憑什麼是 最強的武器？

01 不懂統計學的人容易被騙

威爾斯的預言

威爾斯於一九〇三年預言，統計思維和讀寫能力一樣，終將成為良好公民所需具備的必要條件。

這段文字就寫在哈佛大學醫學院所使用的統計學教科書的開頭。

作家H·G·威爾斯(Herbert George Wells)人稱「科幻小說之父」，也是一位思想家。時光機器及透明人等概念都是因其著作而聞名，據說，他的科學知識之廣博、眼光之高遠，早就預言了核子武器和國際聯盟的出現，甚至還預言了類似現在流行的維基百科般的百科全書會問世。

在現代統計學仍處於黎明期的一九〇三年當時，我們無從確認威爾斯為何會如此預言。但經過了一百年之後，正如同讀寫能力的重要，我們也確實需要統計學的思維。讀寫能力視為教養或素養（literacy），因此，缺乏統計學方面的識讀能力，亦即缺乏「統計學素養」，尤其對於現今社會發展與時代潮流所趨，是極為不智的。

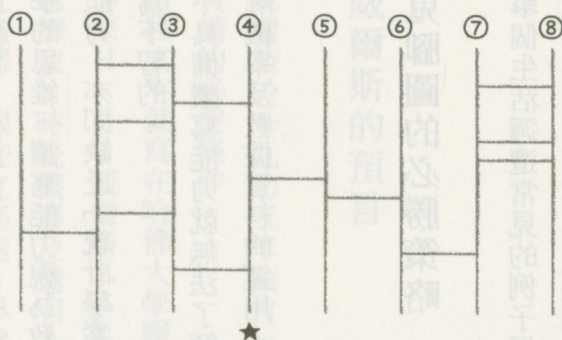
不具備讀寫能力就無法了解合約及法律等內容，而不具備統計學素養則無法了解機率及數據資料的運用。無論是以上何者，都可能陷入弱勢族群的處境。

鬼腳圖的必勝策略

舉個生活週遭常見的例子好了，過去我在讀研究所的時候，經常為了決定誰該去便利商店替大家買東西，和同一研究室的同学玩鬼腳圖。由於其助人替

雖說鬼腳圖有好多種形式，不過，我們採取的方法是依據參加人數四位

圖表 1 鬼腳圖示例



(包含我在內)來繪製兩倍數量，也就是八條的垂直線，然後由我在其他人看不見的情況下，於左邊數來第四條垂直線的下方標記星號，再由另外三位友人在我看不見的情況下，每人分別畫四條橫線。接下來，所有參加者以猜拳的方式決定順序，從最贏的人開始，每人依序在①～⑧號的垂直線中各選一條線，待每個人都選完一輪後，再以顛倒的順序選一輪。而所選直線若連到了★標記(也就是抽中)的人，就要替大家跑腿，到便利商店買東西(如圖表1)。

若是只靠直覺來玩鬼腳圖，那麼，你輸的機率很高。