

STATISTICS

統計學：原理與應用

本書特色

- 融合數學的理性與文字的感性。說明力求平實流暢、簡明易懂。
- 各章節使用單一範例故事進行串聯，強化概念與具體事例的整合。
- 詳列統計軟體操作步驟與結果解析，不僅見證統計原理於報表中，讓讀者在統計分析上更得心應手。

邱皓政、林碧芳、許碧純、陳育瑜 合著

STATISTICS

統計學：原理與應用

邱皓政、林碧芳、許碧純、陳育瑜 合著

五南圖書出版公司 印行

國家圖書館出版品預行編目資料

統計學：原理與應用 / 邱皓政等著。--
初版。-- 臺北市：五南，2012.11
面；公分
ISBN 978-957-11-6867-8(平裝)

1.統計學

510

101018933



1H78

統計學：原理與應用

作 者—邱皓政、林碧芳、許碧純、陳育瑜

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

排版設計—上驊實業有限公司

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

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

台中市駐區辦公室／台中市中區中山路 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

法律顧問 元貞聯合法律事務所 張澤平律師

出版日期：2012 年 11 月初版一刷

定 價 新臺幣 780 元

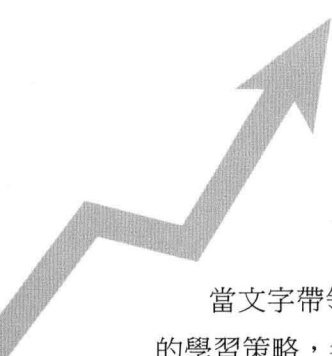


序

如果不同的科學領域之間有需要共同的語言來溝通，那麼統計就是其一；如果真理是愈辯愈明，那麼就更有賴統計發聲。在當代學術的舞臺上，統計學雖不亮眼，但是從不缺席；在真實世界的問題叢林中，統計學雖不是要角，但是絕對是關鍵角色，能夠提供知識與力量。如此誇讚統計雖多所溢美但絕非妄言。統計學作為自然與社會科學絕大多數領域的共同必修課，並不是教授們共同商議的決定，而是眾多學者對於學科基本價值的肯定與專業養成需求的共同默契；在教育應用、社心專業、經濟預測、產業發展、商業經營、管理實務乃至於國家治理，統計程序的應用與分析技術的導入已深入各行各業，都是基於問題解決與預測監控的實際需要。尤其到了科技高度發展的今天，芸芸眾生的日常活動都被數位條碼悄悄紀錄，雲端上的訊息都是數據，益發凸顯統計的存在價值。甚至可以說，誰能掌握數據就像擁有寶藏，但要熟用統計才能點石成金，懂得方法才能登堂入室，在學術界是如此，在實務界亦然。

對於初入大學殿堂的年輕學子而言，要能體會統計的重要性並不容易，若要期望學生能為這個學科的學習廢寢忘食則更是不切實際。只有循循善誘苦口婆心，加上一些日常生活的點滴故事與實際體驗，或許才能在驚喜發現當中找到一絲對於統計的好奇與熱情，相信這是所有統計教師的共同經驗與無奈心情。這些年來，隨著個人電腦的普及與統計套裝軟體的發展，繁瑣的運算終於獲得解套，透過報表即時的反饋與學生實作活動的多樣設計，為統計課堂增添一分樂趣，才使統計老師們稍微寬心。

本書出版的目的，也即是基於前述背景的呼應，希望能用文字的力量來稍解老師們的良言苦口，減少公式推導與繁複運算來降低學生的排斥與恐懼。本書對於統計原理的說明鋪陳大量的文字敘述，雖稱不上是字字珠璣，但力求平實流暢、白話簡明，讓數學的理性融合文字的感性，期使學習者能夠真正透過閱讀來理解統計的道理與邏輯。此其一。



當文字帶領概念前進時，本書有別於傳統統計教材以多重範例反覆演算複習的學習策略，各章節盡可能使用單一範例故事來進行串聯，提升學生課堂學習的專注度，強化概念與具體事例的整合而能避免意識流暢的斷裂，從而得到頓悟的喜悅。因此本書除了各章的課後習作，內文中絕少範例習題，避免捨本而逐末，流於中學補教解題的學習模式。此其二。

一旦學習者能夠悠遊於白話的鋪陳與單軸的範例下，完成全章的概念學習而有所體悟之際，本書於各章末節耗費相當篇幅詳細列舉統計軟體操作步驟與結果解析，以實際應用的實作活動來打鐵趁熱，取報表數據來比對觀念而能追根究柢，一方面使學生能夠親眼目睹現代科技的神奇魅力，在光速的運算下，瞬間看到滿天星星，當下決定科學想像的對錯是非，見證統計原理躍現於報表之中，另一方面則能滿足人類原始的好奇慾念，得到真實的學問樂趣。此其三。

多年來，許多莘莘學子帶著忐忑心情走進我的統計課堂，凡人如我也非天賦異稟、無三頭六臂，即使唱趣逗笑、盡心盡力，最後學習成果多是差強人意，無法跳脫統計教師教學評鑑低迷的共同宿命。後來隨著經驗的累積與科技的進步調整教學方式，倡議白話運動，高舉實用主義，終於獲得一些掌聲，在學子身上看到難得的笑容，也決議將這些教學經驗形諸文字，做成教材。

除了已經在多年前轉往管理學院教書的我，本書的其他幾位作者都是長年與我切磋學問分享教學的師生伙伴，雖然各自分散在不同的學門領域，接觸不同的研究課題與學生，但是卻有共同的交集，最具體的部分則展現在這本書的撰述編成中。其中林碧芳博士畢業自國內首屈一指的政治大學教育學院，主修教育心理與計量方法，擔任臺灣統計方法學學會教育訓練處主任多年，推廣教育經驗豐富，可以歸屬於教育與心理學門；許碧純博士則畢業於長春藤名校康乃爾大學，鑽研婦女議題，享有社會學博士頭銜，除了社會理論的深厚根基，資料庫數據的實徵分析經驗豐富，可以歸屬於社會學門；陳育瑜則是輔仁大學心理學系博士候選人，作為學術新秀，追求質量兼備，對於統計方法的運用與結果的詮釋自有不同之見解，可以歸屬於心理學門。雖然四位作者分處管理、教育、社會、心理四大學門，教材術語與範例舉隅自然有其專業脈絡，但本書的適用對象更可跨越這些學門藩

籬，提供統計原理與應用方法的專業教材或自學工具。

常言道，萬丈高樓平地起，本書僅是開啟通往學術殿堂的第一扇窗，但不難體會學問之浩瀚，若要堆起智慧的金字塔，確實要下足功夫，未來才有登高望遠的睥睨。恩師丁興祥教授六十大壽時，歷屆師兄弟妹齊聚內湖碧山巖敘舊慶生，多年不見，好不開心，上山時正當華燈初上，山嵐裊裊，如霧起時。想起自己初作學問時，也是在山下翹首盼望，面對來路茫茫內心忐忑不安，只因為當初聽得長輩一句話，學問在哪？回說，只在此山中，雲深不知處，任聽他人道長短，何不親身探究竟。確實說得好，一步一腳印。那雲深意境，景致如何，悠遊山林採藥人最得明白，千萬別錯過了這一段好風景，那可真是遺憾。

邱皓政

西元 2012 年 10 月
於師大路管理學院



目次

第一章	統計學概論	1
第二章	科學研究與數據	15
第三章	次數分配與統計圖表	37
第四章	描述統計與標準分數	69
第五章	機率與機率分配	113
第六章	抽樣與估計	161
第七章	假設檢定的原理與應用	207
第八章	實驗設計與變異數分析	259
第九章	多因子變異數分析	307
第十章	共變數分析	373
第十一章	相關分析	403
第十二章	迴歸分析	431
第十三章	多元迴歸分析	461
第十四章	調節迴歸與中介迴歸	497
第十五章	無母數檢定：類別資料分析	537
第十六章	無母數檢定：等級資料與關聯分析	577
附錄 A-H		631
索引		647
各章習題參考答案		685
參考文獻		723

Introduction

1

統計學概說

- 1.1 前言
- 1.2 統計學發展的脈絡
- 1.3 統計學的內容是什麼：從五個例子來看
 - 1.3.1 簡單中卻有大道理：描述統計
 - 1.3.2 見微知著：參數估計
 - 1.3.3 是事實還是偶然：假設檢定
 - 1.3.4 無獨有偶、預測未來：相關與迴歸
 - 1.3.5 此消彼長：交叉分析
- 1.4 統計學的分類
- 1.5 結語
 - 1.5.1 真的有必要學統計嗎？
 - 1.5.2 統計真的有這麼難嗎？

To

Statistics



1.1 前言

科學的目的在探討未知、追求真相。透過科學的研究，人們得以建立系統性知識，發展科學的定律與理論，去除不確定性，做成決策，因此學術研究者的一項基礎訓練，就是學習科學的研究方法，熟習分析工具與技術，其中，統計方法的應用扮演著一個重要的角色。

科學家為了去了解並解釋人類與動物的行為，必須藉由資料（data）的蒐集來達到此一目的。當科學家手中擁有了資料之後，就必須利用一套有效的程序來進行資料的整理、呈現、描述與解釋的工作，並進一步的從中找到決策的根據。從學術上的定義來說，統計是一套處理與分析量化資料的技術，而探究統計方法的原理與應用的學科，稱為統計學（statistics）。統計方法的應用是科學研究的重要環節，統計決策則是科學發現的主要依據。

1.2 統計學發展的脈絡

統計的起源可追溯至十八世紀，德國人將「國家應該注意的事實學問」，包括國家的組織、人口、軍隊與資源的記述工作，以德文的 *statistika* 一詞，正式命名為統計學。換句話說，統計最早是因為統治者對於治理國家的需要而發展出來的一套技術，因此被稱為「政治數學」。由統計（statistic）與國家（state）語出同源來看，就可以明白統計與國家治理的關係。

在古埃及時代，王室為了修建金字塔，就曾對全國人民的財產進行調查。在中國，四千多年前的夏朝，《尚書/禹貢》一文記載了當時的中國九州的各地物產、交通、植物特徵等統計資料，以及田地及貢賦分為九等的作法。這種統計技術與十七世紀德國的國情調查非常相似，但中國甚至早了後者近兩千年。到了周朝，在統計方面更為完善，不僅制定了鄉的定期報表制度，在統計方法上還應用了專門調查方法，使用統計圖示及帳冊，當時的中國人就知道統計分組、平均數、相對量數等近代統計方法。

雖然起跑點甚早，但是中國始終沒有把歷史悠久的統計工作發展成為一門系統的現代科學，西方的統計學雖然到了十九世紀末葉才真正開始，到了四十年代才逐漸成熟，但到了 Cattell 時代，與當時的自然科學、哲學、數學發展相結合，

統計學家們大量引進了概率與數理統計方法，並運用電子化技術來協助進行運算，在統計學領域逐漸獲得領導的地位，直到今日。

目前英文的 *statistics*，是由英國數學家 William Petty (1623~1687) 從德文 *statistika* 翻譯而來，說明統計在「專門研究各種數量」，但如果我們把描述統計正式發展視為當代統計學的起點，那麼十九世紀末的 Galton (Sir Francis Galton, 1822 ~ 1911) 和 Pearson (Karl Pearson, 1857 ~ 1936) 可以說是當代統計學的發展起點。

Galton 是著名的演化論者達爾文的表親，曾為達爾文做過統計分析的工作。1899 年 Galton 以《遺傳學原理》(*Nature Inheritance*) 一書，開啟了統計學的大門，該書除了本身的價值外，最重要的影響是引發 Pearson 對於統計學的興趣。在此之前，Pearson 只是在倫敦大學的一個數學老師。1890 年 Pearson 轉赴格里辛學院 (Gresham College) 教書，在他所開授的《現代科學的範圍與概念》(*the Scope and Concepts of Modern Science*) 中，他開始注意統計的原理及對科學研究的影響，並致力於統計理論的研究。他最大的貢獻與其說是開發了統計這一個處女領域，提出相關係數、標準差這幾個重要的統計名詞，更重要的是引發世人對於統計的重視，改變了人們對統計的態度，並說服科學界承認統計是一門學科。

Galton、Pearson 和他們的學生們除了發展迴歸和卡方檢定等重要的統計概念後，另一個重要的學者 Fisher (Ronald Aylmer Fisher, 1890 ~ 1962) 也出現在統計學名人榜當中，著名的 *F* 檢定的字首 *F*，就是他的姓氏的縮寫。1915 年，他發表了統計量的精確分配的論文，將統計史帶入另一個時期，有人甚至把今天所廣為採用的各種統計理論的絕大多數歸功為 Fisher 的成就。

Fisher 最重要的成就在於適用於小樣本的統計方法的發展，他與他的同事也發展出許多樣本統計量的機率分配，推論統計的決策模式也多半在他的手中完成。Fisher 可以稱得上是一個天才兒童，在他很小的時候就已精通立體三角之類的數學問題。1912 年畢業於劍橋大學，得到天文學的學士學位，1915 年發表令他一夕成名的文章，也開啟統計領域對於樣本統計分配的研究。他的兩本名著《*Statistical Methods for Research Workers*》和《*Design of Experiments*》分別於 1925 年和 1935 年出版，對於統計有重大的影響。最大概似估計法 (maximum likelihood estimation) 觀念的提出，也是歸功於 Fisher，可見得他在統計領域的重要地位。

統計的現代化也與 Pearson 有關，他的兒子 Egon Pearson 在 1928 年與 Jerzy

Neyman 共同發表幾篇重要的論文，主要在探討統計決策的基本問題，例如第一與第二類型錯誤，檢定力和信賴區間之類的觀念。在這期間，實務界開始大量採用統計技術來處理品管問題，興起了抽樣理論與調查方法的研究。Egon Pearson 整理並修正了 Fisher 早期所留下的問題，完成了今天統計領域所使用的決策理論核心概念。

到了今天，統計成為一門科學方法，其應用範圍遍及自然科學及社會科學的整個領域中的絕大部分，舉凡農業、工業、商業、經濟、教育、醫藥、政治、社會、心理等各領域無不適合採用統計方法進行研究，可見得統計學有其發揮的空間，也是提升整體科學發展的奠基之一。電子計算機與電腦的普及，無疑是統計學高度發展的推手，當 SAS、SPSS、BMDP 等統計套裝軟體出現之後，統計運算更為方便，只要知道應採用何種統計方法就可以快速得到結果。總而言之，統計學是一套讓科學得以實踐並發揮效益的工具，工具必須要有使用的空間及場合，才能見到真正的效益。這個空間，就是學習統計者的舞台。

1.3 統計學的內容是什麼：從五個例子來看

在社會與行為科學領域中，對於統計方法有幾種基本的應用方式，以下我們將介紹五種常見的應用範例，並同時介紹一些專有名詞，關於這五種範例與專有名詞的相關細節，將在後面的章節詳細討論。

1.3.1 簡單中卻有大道理：描述統計

關於各位讀者，有一個值得探討的問題，是你願意花多少時間來學習統計？先不談統計學，現在的學生究竟願意花多少時間讀書就很值得關心。如果在速食店隨機找 10 個學生來問一下，他們可能會認為一個星期花 7 小時在課業上就已經不錯了，有人多一點十幾個小時，也有人平常都不會讀書，除非有考試。但是如果拿一樣的問題去問 10 個大學老師，他們可能會認為一個星期花 14 個小時讀書也不為過。

現在，有一位統計老師希望透過科學的方法來瞭解這個問題，他編製了一份

簡單的問卷，列舉一些關於統計課程學習的問題，然後透過統計學會裡面的朋友協助發放給選修統計課程的學生來填寫，最後回收了將近 500 份問卷，他發現有 15% 的學生會進行課前預習，45% 的學生會做課後複習，他們練習統計習題的時間每週平均只有 0.8 小時，但是花了 2.9 小時在使用統計軟體來做作業，尤其是當統計學是選修課而非必修課時，使用軟體做作業的時間越長，學生課後複習的比率越高。這位老師對於研究結果感到驚訝，因為他自己的統計課並沒有使用統計軟體，他認為統計知識的建立一定要從演算中學習。他開始思考是否改變教學方式，因為研究數據透露著，有使用電腦來輔助學習統計的學生似乎對於課程的滿意度較高，也比較樂意在課前課後自我學習。

前面的例子說明了描述統計的奧妙，它甚至改變了某位老師的教學方式。簡單來說，**描述統計**（descriptive statistics）是指利用某些簡單的數字來描述一群對象的某些特徵，這個「簡單的數據」稱為**統計量**（statistic），而數據通常來自於一群人、動物、物體或事件，統稱為研究對象。以統計學的術語來說，這一群研究對象就是**樣本**（sample），而樣本是從母體裡所抽取得到一個子集合，**母體**（population）則是指帶有某共同特徵的一群研究對象的全部。例如，在大學當中修統計學的學生可以是一個母體，因為他們擁有一個共同的特徵，是在大學四年間都要修一門相同的課。同樣的，在大學當中教統計的老師也可被認為是一個母體，這些老師不論是系上的專任老師或兼課老師，只要是在大學開課教統計學的老師，都擁有共同的特徵，但是助教就不算是這個母體的一部份，因為他們不是老師，而是協助老師教課的人。

描述統計可以說是研究者能夠從手邊資料所得到的最基本的資訊，雖然這些統計量都很簡單（例如平均數或標準差），但是卻有大功用，它對我們所蒐集的資訊提供我們客觀的摘述，關於描述統計的相關細節我們會在第三章與第四章進行討論。然而，描述統計也是有所限制，例如它們無法獲知數據之間的相互關係，也無法瞭解不同狀況下的數據差異的意義。例如當統計課是必修或選修課時，學生學習的狀況是否真的有所差別？使用統計軟體來協助教學是否真的會提高學生的學習興趣？如果僅看粗略的描述統計量，我們並無法得知自主學習與課程的教學方式是否有關，而學生們願意多花時間讀書的確實原因為何也不得而知。所以我們需要推論統計來協助我們。

1.3.2 見微知著：參數估計

前面描述統計的例子中，我們曾經舉例：在速食店隨機找 10 個學生調查得知平均每週讀書時間是 7 小時，那麼這句話到底能不能反映大學生的現況呢？也就是說，這 10 個學生所蒐集得到的統計量是否能夠推論到具有相同特徵的所有大學生的這個母體呢？此外，那位認真的統計老師調查 500 位學生所得到的「每週平均花費 2.9 小時使用統計軟體來做作業」，是否就是現在學生學習統計的普遍狀況？描述統計量只能反映我們從一個樣本裡所蒐集的測量數據是甚麼，但是無法得知我們沒有掌握的資料會如何。顯然一個樣本所得到的資料並不能代表全體，但是為了得到更全面的結果，研究者必須進行統計推論，從手中所獲得的樣本資訊（掌握已知）來對其所出自的母體得出結論（推論未知），而參數估計（parameter estimation）就是利用樣本統計量來推知母體參數的過程。

為了確保樣本能夠代表母體，樣本的獲得必須遵循一定的規則，最常見抽樣方式是隨機抽樣，也就是從母體中以某種隨機方法挑選成員，母體中的每一個成員都有相同（或特定）的機率被選擇作為樣本，而且每個被挑選的成員彼此相互是獨立不互相影響，此時所建立的樣本可稱為隨機樣本（random sample）。利用隨機樣本所獲得的資料所計算出描述統計量之後（例如 10 位學生平均每週讀書 7 小時），得以用來推論母體的狀況，此一被推論的母體特徵被稱為參數（parameter），例如大學生每週平均讀書幾小時。有很多常見的以描述統計量來推論母體參數的例子：

- * 初入社會的社會新鮮人平均起薪為 25000 元。
- * 一般上班族每天花 23 分鐘才能到達工作地點。
- * 臺灣地區平均每戶成年人數為 2.53 人。
- * 臺灣地區平均每人每月消費支出為 18007 元。

以上每個數值都是從樣本所獲得的描述統計所估計而來的參數。我們不可能去一一詢問每一個人的起薪是多少，或是去調查每一個上班族每天花多少時間通勤，因此每個數值都是母體的估計值。利用描述統計量是否能夠準確估計母體的參數，取決於抽樣過程與抽取的觀察值數目。關於抽樣方法，以及相關機率問題

與估計過程，將在第五至六章討論。

1.3.3 是事實還是偶然：假設檢定

科學家除了想對母體特徵進行瞭解之外，更對事情的因果關係感興趣。為了掌握因果（causality），最好的方式就是進行實驗（experiment）。一般來說，實驗的進行必須先去自變數（independent variable; IV）與依變數（dependent variable; DV）。自變數是指研究者能夠操弄的變數，藉由操弄某個變數的不同狀況（實驗水準），觀察一群實驗受試者（subject）或參與者（participant）的行為表現，此時被研究者預期會被自變數所影響的行為結果稱為依變數。

例如，酒駕是危害民眾安全的行為，因為喝酒會影響駕駛人的反應能力，但是要喝下多少酒精才會影響駕駛呢？如果喝下帶有米酒調味的湯頭是否會影響駕駛呢？政府對於酒駕的定義必須有一個合理的標準才能作為取締的依據，此時就需要進行實驗。如果我們以喝下酒精的多寡為自變數，測量駕駛對於信號的反應速度作為依變數。研究者認為酒精會影響對反應速度的這個想法稱為研究假設（research hypothesis），一般來說，研究假設是研究者對於兩個或兩個以上變數關係的預期或預測，在實驗中，研究假設是指自變數與依變數的預測關係。

為了檢驗研究假設是否成立，科學家挑選一群實驗參與者，將他們隨機分配到幾個不同的組別，然後操弄不同的狀況來進行實驗處理（treatment），例如讓參與者飲用不同比例的含酒精飲料，最後觀察他們的行為反應靈敏度。之所以要進行隨機分派（random assignment）的目的，是為了確保參與者有相同的機會被分配到各種實驗狀況下，而且參與者被分派到實驗組與其他參與者的分派是獨立的，因此每一組參與者的特徵都十分相似，使得每一組之下的參與者在依變數上的表現，可以被視為是自變數影響的結果。

對於實驗操弄是否影響依變數，需要進行一系列的統計分析與考驗。首先，研究者必須計算每個實驗狀況下受試者在依變數得分的描述統計量，此一步驟跟一般描述統計量的計算過程完全相同。第二個步驟則是決定不同實驗組的依變數描述統計量的差異，是否效果大到能夠歸因於自變數的影響，而非偶然性的差異，此一步驟需要使用統計假設檢定（statistical hypothesis testing）。

基本上，任何兩個群組的描述統計量都會存在一些無法預期的差異，稱之為

隨機差異（chance difference），即使自變數對於依變數完全沒有影響，隨機差異也會發生於實驗研究中的不同群組當中。假設檢定的原理，就是將實驗中的各群組實際所觀察到的差異，來與群組間預期的隨機差異進行比較。如果獲得隨機差異的機會很小，而實際所觀察到的差異夠大，那研究者就可以得到受試者們的差異並非是隨機差異的結論，亦即可將觀察值的差異歸因於自變數的效果。為了要決定不同實驗群組間在依變數上的平均數差異是否不是機會所造成，最常使用的統計方法為 *t* 檢定與**變異數分析**（analysis of variance），前者適用於當只有兩個組別的平均數相互比較時，後者則可應用在超過兩個組別的平均數差異的比較。

統計假設檢定在學術研究上被廣泛使用，要了解它所涉及的概念，需要有描述統計、機率與抽樣理論的統計知識，這些主題將在第四章到第六章進行討論。然後我們將在第七章正式介紹統計假設檢定，在第八章開始進行一連串關於實驗設計與變異數分析的說明，一直到第十章。

1.3.4 無獨有偶、預測未來：相關與迴歸

許多社會科學研究無法以實驗方式來進行研究，例如許多變數無法進行操作，例如性別、年齡、左撇子或右撇子、體重、身高、收入狀況、教育程度、藥物使用狀況等**受試者變數**（subject variable）。但是科學家還是對於這些受試者變數與其他相關變數之間是否具有共變關係感到興趣。如果一個變數的改變與另一個變數的一致性改變有關，則兩個變數稱為具有共變關係。例如，如果一個人的學業成績與個人的努力有關係，那麼成績與努力就具有共變關係，這兩個變數的變化具有關聯性，此時可以計算介於 -1.00 到 1.00 之間的**相關係數**（correlation coefficient），作為描述兩組分數的相關程度和方向的統計量。

相關係數雖然是一個統計量，但也需要使用統計假設檢定來決定這個係數是否是隨機現象。如果兩組分數的相關並非是隨機發生的，那麼我們就可以嘗試從一個變數去預測另一個變數。假設你知道一個人的努力程度，就可以預測他的學業表現，但是以努力程度去預測學業成績到底有多麼準確呢？也就是如何有效的以某個變數的分數去預測另一個變數的分數，必須使用**迴歸分析**（regression analysis）。相關與迴歸分析技術可以說是社會科學的領域中最被廣泛使用的一種統計技術，例如下列各種狀況的研究與應用：

- * 孩童年齡增長與智力發展的關係。
- * 廣告投資金額越高是否有效提高產品銷售情形。
- * 以歷年來的出生與死亡人口來預測未來的人口結構。
- * 要獲得多少產品需要投入多少生產成本與人力。

關於相關與迴歸的原理，我們將在第十一與十二章介紹，然後在第十三與十四章討論比較進階的多元迴歸與中介及調節迴歸分析。

1.3.5 此消彼長：交叉分析

每當選舉季節，電視上最熱鬧的新聞之一就屬候選人的支持度調查。我們經常可以聽到某某民調公司以電話訪問台灣地區 18 歲以上的 1000 名選民，詢問他們如果明天是投票日，他們是否支持某一政黨的候選人，然後得到一個百分比數字，同時交代一下抽樣誤差的比​​例（例如正負三個百分點），然後作為選舉情勢分析的證據。

事實上，民調的分析也是一種統計推論，民調公司所訪問的選民就是樣本，從而獲得支持意願的數據，進而推知全體選民的支持意向，此時為了確保樣本能夠代表母體，抽樣過程是其關鍵，樣本越大，抽樣誤差（sampling error）越小。若從統計的角度來看，民意調查與前述的統計推論有兩個主要的差異。

第一個差異，是民調所使用的變數通常不是反映強度變化的連續變數，而是數人頭類別變數。連續變數的資料可以計算各種描述統計量，但是類別變數所能夠使用的描述統計量相當有限，通常只有人數的次數資料，或是把人數除以總人數得到百分比。例如 1000 個選民當中，有 250 人支持甲政黨，有 400 人支持乙政黨，150 人支持丙政黨，其餘則是未表態者，除以總人數之後就可得到百分比或機率：.25、.40、.15 與 .20。

進一步的，民調通常會把選民的投票意向，再就他所屬的政黨別，進行兩者的交叉分析，看看不同政黨背景的人是否支持特定的政黨，還是說會有跨政黨的投票意向。例如偏向甲政黨的選民，理應支持甲政黨的候選人，偏向乙政黨的選民，則應支持乙政黨的候選人，如果有相當比例的選民投向對立或不同的政黨，那麼就有玄機可以探討了。

在統計上，對於類別的次數分配是否具有特殊的比例（例如一個班級的性別比例是否為 1 比 1），或是兩個類別變數的分佈比例是否具有連帶關係（例如選民自己的政黨傾向與候選人的支持情形），可以利用卡方檢定（chi-square test）來檢驗。由於所牽涉的變數主要是類別變數，因此不需要對於母體分配是否為特殊分配進行假設，因此這一類的統計分析又被稱為**無母數檢定**（non-parametric test）。關於卡方檢定在類別資料與等級資料的分析應用，我們將在本書的最後兩章：第十五與十六章介紹。

1.4 統計學的分類

從統計所涉及的知識內涵來看，統計學可區分為**理論統計**（theoretical statistics）與**應用統計**（applied statistics）兩個層次。前者所關心的問題是統計的基本原理與理論內涵，多以數學模式的推導或模擬方法來探討統計的相關問題；後者則強調統計方法如何解決各種研究問題，多涉及研究場域的實徵資料分析與具體研究課題的解決，兩者之間相輔相成，可以說是理論與實踐的關係。

若從統計的功能來看，統計學則可以區分為**描述統計**（descriptive statistics）與**推論統計**（inferential statistics）兩大範疇。描述統計的目的在整理與描述研究者所獲得的數據，以描繪出數據的全貌與特徵；推論統計的目的則在進行統計的檢驗與決策，尋找數據背後的科學意義。在自然科學領域（包括行為科學），由於側重實驗研究法，因此衍生出一門與統計分析關係密切的學科，稱為**實驗設計**（experimental design）。實驗設計的內涵除了影響了一個實驗進行的程序與方式，更決定了研究的數據如何分析，因此，亦有學者將實驗設計從描述統計與推論統計獨立出來，成為另一系列的統計技術（如 Kirk, 1995）。

如果從統計技術的複雜度來看，統計學可以區分為**單變量統計**（univariate statistics）、**雙變量統計**（bivariate statistics）與**多變量統計**（multivariate statistics）。單變數統計涉及單獨一個變數的處理，雙變數則涉及兩個變數關係的探討，多變量統計則牽涉到多個變數關係的分析。不論是哪一種分類，都說明了統計學的內涵包含甚廣，無法用一個單一的分類系統來涵括各種不同的統計方法與概念。但是一般來說，除了專門以統計原理為主要學習內涵的數學與統計系所之外，其他各學門對於統計學的探究，主要在於如何配合該學門的特性，有效的