

科技統計學

Statistics for Technology

原著者：Christopher Chatfield

譯述者：曾 慧 鶯

複校者：黃 雄 毅

科技圖書股份有限公司

科技統計學

Statistics for Technology

原著者：Christopher Chatfield

譯述者：曾 慧 (驚)

複校者：黃 雄 毅

科技圖書股份有限公司

這是一本介紹統計學基本理論，但不包括數學上的細節，重點置在科技方面實際的應用，並詳細解釋其基本概念。如此可以避免未了解統計公式即從事計算會有誤用的危險。本書列有廣泛的資料分析，與用電算機問題的解例，使讀者有更深一層的了解。

本公司經新聞局核准登記
登記證局版台業字第1123號

書名：科技統計學
原著者：Christopher Chatfield
譯述者：曾慧鶯
複校者：黃雄毅
發行人：趙國華
發行者：科技圖書股份有限公司
台北市復興南路一段360號7樓之三
電話：7073230・7056781
郵政劃撥帳號 15697

七十二年七月初版

特價新台幣100元

有成圖書公司

原 序

寫這本書的目的是使讀者認識統計學在科學與技術上的應用。本書既能作為教科書，以講授統計學的基本課程，亦可作為自修用書。在數學上，儘可能保持簡易，僅需具備微積分的基本知識，與能運算代數公式已足。

在寫這本書時，已經知道有很多優良而具理解性的理工基本統計理論的書。我的目標只是介紹這些基本理論，但不包括數學上的細節，如此可使這本書大部份着重在實際的應用。因為某些結論的敘述不加證明，將不致影響學生的理解能力。為避免如烹調書方式去研究統計學，改用仔細解釋這個題目的基本概念，諸如，機率與樣本的分佈，這是學生必需了解的。當科技工作者試圖分析資料，用不了解的統計公式來代替計算時，會有誤用統計學情事發生。

若不增加數位電算器的使用說明，本書將不能完成。這些機器產生某些科技研究上的巨大影響。首先，若用桌上電算器來計算，試用者能研究比實際更複雜的模式。第二，若艱難的計算問題能先解決將鼓勵許多新方向的研究。如兩種統計的例子，複變數分析與時間數列的分析，電算機亦被廣泛用作例行的資料分析。一種統計程式彙編，正是用在大部份電算機以執行這些複回歸與變異數分析。雖然，讀者使用一個桌上電算器去計算本書指定的問題，將使他對題目有更深一層的了解，同時，可用電算機解更複雜的問題。

一開始，這本書是由 R. B. Abernethy 博士與我合作編寫，Abernethy 博士是任 Pratt and Whitney 公司的航空工程師，不幸由於工作的壓力，Abernethy 博士被迫退出這一計劃。由於他的有價值的幫助，我仍然很感激他的合作。尤其在第三章與第九章的初稿。當然，任何忽略的錯誤，應完全由我負責。

我也感謝 McGill 大學的同事們，閱讀本書，同時給予建設性的批評。對 R. Rigelhof 先生與 H. Schroeder 小姐從事準備手稿的工作亦記下我的感謝。

我感謝 Biometrika 基金會允許轉載 Biometrika 統計表（附錄 B，表 6），Oliver 太太及 Boyd Ltd, Edinburgh。允許刊載生物學、農業與醫學研究統計表（附錄 B，表 7）。我也感謝 Penguin Education 供應在附錄 B 中的表 1、2、3、4 與 5。

最好使 Lewis Carroll 先生知道，從 Charles Luturidge Dodgson 的作品中引用些課題。

這本書分成三篇，第一篇是簡介，第二篇是處理統計理論問題，第三篇是提到某些應用問題。在不同篇中，其內容的分配是有些隨意的。例如迴歸問題，在第二篇中討論的，可能與第三篇中的相同。第三篇的應用問題是着重在經驗設計與分析、品質控制與壽命檢定。讀者若對最後兩個題目之一有特別興趣的，可從第七章直接對第十二章或第十三章作專門性研究。

Christopher Chatfield
克利司朵夫·賈德菲爾特

伯斯大學（Bath University）

目 錄

原 序

第一篇 簡 介

第一章 統計學概要

第二章 資料分析的簡易法

2.1 簡 介 5

2.2 圖解法 6

2.3 算術法 11

第二篇 理 論

第三章 機率概念

3.1 機率與統計 17

3.2 若干定義 18

3.3 事件的類型 20

3.4 排列與組合 28

3.5 隨機變數 29

第四章 不連續分佈

4.1 不連續機率分佈 34

4.2 二項分佈 35

4.3 二項模式 35

4.4 分佈的類型 40

4.5 二項分佈的平均數與變異數 41

2 科技統計學

4.6 卜桑分佈 44

4.7 二項變量不連續分配 50

第五章 連續分佈

5.1 定 義 55

5.2 連續分佈的平均數與變異數 59

5.3 常態分佈 60

5.4 常態分佈的應用 64

5.5 常態機率紙 66

5.6 指數分佈 68

5.7 雙變量連續分佈 73

第六章 推 估

6.1 點與區間推估 76

6.2 期望值的特性 77

6.3 $\bar{x}$ 的抽樣分佈 80

6.4 s^2 的抽樣分佈 84

6.5 推估值的若干特性 86

6.6 點推估的一般方法 89

6.7 區間推估 93

第七章 重要性檢定

7.1 簡 介 100

7.2 樣本平均數的檢定 105

7.3 兩個樣本平均數的比較 108

7.4 t -檢定應用於成對比較 111

7.5 χ^2 配合適度檢定 112

7.6 F -檢定 117

7.7 免分佈或無參數檢定 119

7.8 勢力與其他考慮 120

第八章 回歸與相關

- 8.1 散佈圖 127
- 8.2 曲線配合 127
- 8.3 回 歸 131
- 8.4 在線性回歸中的信賴區間與顯著性檢定 134
- 8.5 決斷係數 136
- 8.6 複回歸與曲線回歸 137
- 8.7 直交多項式 139
- 8.8 回歸試驗之設計 143
- 8.9 相關係數 144
- 8.10 回歸線估計 148
- 8.11 二變量常態分佈 151
- 8.12 相關係數的解釋 153

第三篇 應 用

第九章 試驗的規畫

- 9.1 主 題 155
- 9.2 測定值 156
- 9.3 誤差的增值 157
- 9.4 用串列及並列法以改進精密度 165
- 9.5 用最小平方法求組合不同的推估值 166

第十章 試驗的設計與分析：I 比較試驗

- 10.1 試驗設計的若干基本考慮 172
- 10.2 簡易比較試驗的數學模式 173
- 10.3 反複次數 174
- 10.4 隨機法 176
- 10.5 隨機性比較試驗的分析 178

4 科技統計學

- 10.6 全距檢定 181
- 10.7 一元變異數分析 182
- 10.8 處理平均數的進一步研究 186
- 10.9 證明模式 188
- 10.10 隨機化區段試驗 188
- 10.11 二元變異數分析 191
- 10.12 拉丁平方 195
- 10.13 均衡不完全區段設計 196

第十一章 試驗的設計與分析：Ⅱ因子配置試驗

- 11.1 簡介 200
- 11.2 完全因子配置試驗的優點 201
- 11.3 完全因子配置試驗的設計 203
- 11.4 完全因子配置試驗的分析 205
- 11.5 緊接步驟 210
- 11.6 2^n 因子配置設計 211
- 11.7 固定效果與隨機效果 215
- 11.8 其他論題 216
- 11.9 殘差的檢查 218
- 11.10 最適條件的決定 219
- 11.11 摘要 223

第十二章 品質控制

- 12.1 合格抽樣 226
- 12.2 操作特徵曲線 227
- 12.3 抽樣計畫類型 230
- 12.4 矯正計畫 231
- 12.5 軍用標準計畫 233
- 12.6 變數抽樣 234
- 12.7 實際問題 234

- 12.8 程序控制 234
- 12.9 樣本的控制圖 236
- 12.10 累積和圖 240
- 12.11 反饋控制 245

第十三章 壽命檢定

- 13.1 可靠性的測定問題 251
- 13.2 損壞時間的數學分佈 255
- 13.3 指數分佈 254
- 13.4 推估條件損壞率 255
- 13.5 韋伯分佈 257

附錄A 常態、 χ^2 、 t 、 F 分佈間的關係

附錄B 統計表

- 表 1 常態曲線下的面積表 265
- 表 2 司徒頓 t 分佈的百分比點 266
- 表 3 χ^2 分佈的百分比點 267
- 表 4 F 分佈的上百分比點 268
- 表 5 e^{-x} 值 271
- 表 6 司徒頓全距分佈的百分比點 272
- 表 7 隨機數 274

進一步閱讀書籍之介紹 276

習題解答 279

第一篇 簡 介

第一章 統計學概要

統計學 (statistics) 是門日漸重要的課程，有利於各種科學的研究。已成為用最佳的可能方法以收集、分析並解釋資料的科學。統計對於試驗不確定性 (experimental uncertainty) 情況下特別有用，且常被設定為“對不確定事實作決定的科學”。我們先舉出幾個試驗的不確定性的科學例子。

◀ 例 題 1 ▶

在同樣操縱情況下，於十分鐘時距內所測得的火箭推進，記下下列三十個觀察值。

999.1	1003.2	1002.1	999.2	989.7	1006.7
1012.3	996.4	1000.2	995.3	1008.7	993.4
998.1	997.9	1003.1	1002.6	1001.8	996.5
992.8	1006.5	1004.5	1000.3	1014.5	998.6
989.4	1002.9	999.3	994.7	1007.6	1000.9

這些觀察值，在 989.4 與 1014.5 之間變動，其平均值約為 1000。對這些小於每一觀察值的絕對值變動，無明顯的理由。且這種變動亦無系統化，任何無形態或規則的變動，稱為隨機變異 (random variation)。

2 科技統計學

在本例題中，若運轉情況保持一致，則可預測下一個觀察值，將是一千加上一個正或負的小隨機量（small random quantity）。

◀ 例題 2 ▶

在一分鐘內，大量宇宙微粒撞擊儀器的四十次連續週期記錄如下：

3	0	0	1	0	2	1	0	1	1
0	3	4	1	2	0	2	0	3	1
1	0	1	2	0	2	1	0	1	2
3	1	0	0	2	1	0	3	1	2

這些觀察值，自 0 至 4 之間變動，所得的觀察值，0 與 1 要比 2，3 及 4 的次數更多。且尚有試驗的不定性，因為無法正確地預測下一個觀察值為何。但可預期其將在 0 至 4 之間，且更接近 0 或 1。在第四章中，將了解到這個資料，雖不能預測個別觀察值，但確有其型態（pattern）存在。

◀ 例題 3 ▶

二十台冷凍機馬達，被工作條件所迫，陷於毀滅或失敗（每小時）的次數記載如下：

104.3	158.7	193.7	201.3	206.2
227.8	249.1	307.8	311.5	329.6
358.5	364.3	370.4	380.5	394.6
426.2	434.1	552.6	594.0	691.5

我們無法正確地預測每一個別馬達能維持多久。但如可能，願意預測一組馬達的行為型態。例如也許想知道全部持續超過一星期（168 小時）的馬達比率。這個問題將在第十三章中討論。

當一位習科技的學生，已完成他的學業而第一次進入工業界，他必需準備經常面對着包括試驗不確定性的情況。本書的目的，是提供技術人員處理這些不確定性的方法。這些方法對於工業與研究非常有用。

工業上試驗，常有下面某些或是全部的特性：

- (1) 控制試驗的物理法則沒有全盤了解。
- (2) 在試驗從前未曾做過，或至少未成功地做過的情況下，手術與技巧沒有完全發展出來。
- (3) 儘早有力的鼓勵推行最少量的簡易試驗。
- (4) 試驗人員可能並不客觀。例如當發明人試驗他自己的發明品，或當一個公司試驗有競爭性的產品時，均有此種傾向。

- (5) 試驗結果並非是預期的，或令人失望的。（工程人員解釋令人失望的結果，如馬費定律 (Murphy's law) 的說明。統計的相關法則可用來表示“若有兩宗事件可能同時發生，而遇到的將是最壞的情況”。）
- (6) 雖然可能出現試驗的不確定性，但許多工業方面情勢必需作決定，而不需另外試驗或理論研究。

爲了說明統計如何幫助不同階段的試驗，我們假定已求出改進太空泵運行的作業，這機器的基本設計說明在圖 1 中：汽油通過泵並以電動馬達轉動。

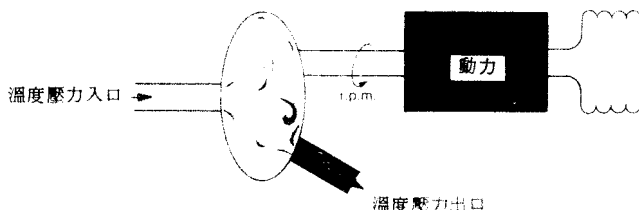


圖 1 太空唧筒圖

第一步驟，是研究目前的工業技術。統計方法，不是一項通曉物理法則 (physical laws) 的代替品。統計是技術人員的工具；與微分方程式或數位電算機有相同意義。

第二步驟，是儘可能精密地說明測驗項目的目標。例如我們也許對改進流動、壓力的提高，與效率的增加，或對降低重量與聲音有興趣，或者有興趣對延長故障與故障間的時距，以改進機器的可靠性 (reliability)。

目標確定後，可列出變數 (variables) 或因數 (factors) 表，而在測驗項目時，這許多變數或因數將發生變動或被變動。在太空泵試驗中，變數包括動力、通路溫度、通路壓力與速率。測驗項目必需設計用來尋求在不同因素下選擇連續值 (successive value) 的最適方法。設計試驗包含的問題將在第十及十一章中討論。

爲求了解目標是否已達成，必需注意其產出 (output) 或反應 (responses)。在太空泵例中，這些產出或反應包括壓力升高、流動、能率與可靠性。必需注意，在這些反應間，可能相互關連。例如，若結果能率降

4 科技統計學

低或失去可靠性（即泵經常故障）則增加流動，可能毫無意義。

在試驗期間，將作因素與反應的測定。被測驗的項目，應作為獲得最精密與正確資料的問題，將在第九章中討論。有結果的資料分析，不僅可決定每一變數對反應的個別影響，且可決定幾個變數間相互連合的影響。推估（estimating）這些結果的過程，稱為統計推論（inference）。這些推估，應分辨出統計的性質，諸如缺少定向誤差，稱為不偏性（unbiasedness）。最後得出的結論或建議，應包含風險（risk），與（或在）推估中的可能的誤差。

為求成功的實現上面的程序，建立一個數學模式（mathematical model），以描述這些物理條件，是相當有利的。這個模式，企圖以數學觀點來制定體系，例如：可能由一系列的有關利息變數方程式所組成。這模式可能包括未知係數（unknown coefficients）或參數（parameter）而該項係數或參數必需從資料中加以推估。在完成這些步驟後，下一步驟是測驗模式的可靠性。最後，若模式被確立，即可得出問題的解決方案。

進一步對統計原理的討論，可參閱下列各書。

參考文獻

- HOOKE. R. (1963). Scientific Inference. Holden-Day.
MORONEY. M. J. (1956). Facts from Figures. 3rd edn.
Penguin Books.
REICHMANN. W. J. (1961). Use and Abuse of Statistics.
Methuen: paperback edn (1964). Penguin Books.

第二章 資料分析的簡易法

2.1 簡 介

工程人員所面臨的最普遍問題之一，是選擇資料的重要特性，並說明若干經過試驗後的觀察值。只有如此，才能測驗並解釋所得的結果。本章所討論資料分析的簡易法，很普遍的被使用着，且常給予一個共同的名稱為說明統計（descriptive statistics）。在此簡介中，描述要比在本書理論部份的開端，更為方便。

觀察值，常是一個樣本（sample），由全部試驗結果中產生的。〔有時稱為母群（population）〕。這是很重要，應自開始就要認識清楚。因為採用所有可能的測定值太不經濟而且費時，唯有抽取樣本一途。統計，是基於一種觀念，即樣本具有“代表性”（typical），且樣本能對全體母群作預測（predictions）。

資料（data），經常是由一系列測定值所組成。此一系列測定值是根據試驗狀況的某些特性，或根據物體的性質而來。所進行調查的現象，通常稱為變量（variate）。在初步值得強調的是，不連續與連續資料（discrete and continuous data）。有關不連續與連續資料的區別如下：不連續資料，只能採用個別值的順序，且經常由若干可點清的觀察值，依一系列整數所組成。因此，在第一章例題 2 中，一分鐘內放射的宇宙微粒數目，必定是依 0, 1, 2, 3, ……順序，而不可能觀察出 1.5。其他不連續變量的例子是，若干飛彈及不下雨天數等的連續試驗。

在另一方面，連續資料是採用在連續範圍內，任何有價值的觀察值。因此，在第一章例題 1 中，火箭的推進，可能為任何正數。其他連續變量（continuous variates）為長度、重量、力與時間。工程師及科學家較喜用連續資料，因其較有利便與益處。

6 科技統計學

不論是不連續或連續資料，資料的結論與顯示，可由圖解與計算法聯合來完成。

2.2 圖解法

資料儘可能用不同的方法繪出。因而只需注意所產生的圖形，就可獲得許多資料，這是一項很好的構想。

2.2.1 線條圖

用第一章例題 2 的不連續資料為例。第一步是找出個別值發生的次數。例如“1”發生 13 次，而“4”僅發生 1 次。

表 1

宇宙微粒數	次數
0	13
1	13
2	8
3	5
4	1
共計 40	

在右欄中的值，形成一個次數分佈 (frequency distribution)。如將這次數分佈繪出如圖 2，則每線的高度與某值發生的次數成比例。

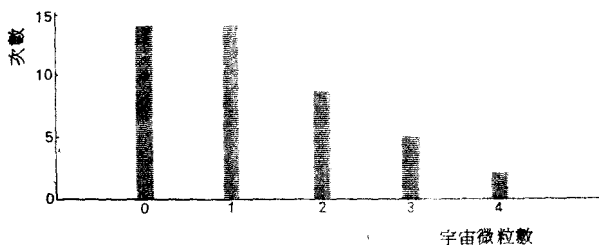


圖 2 表 1 資料線條圖

此圖稱為線條圖 (bar chart) [或線圖 (line graph) 或次數圖 (frequency diagram)]，是分析不連續資料非常有用的視覺助手。

2.2.2 直方圖

大多數人對直方圖非常熟悉，此圖除可用於連續資料外，與線條圖型態相同。茲舉例說明如下。

◀ 例題 1 ▶

100 個學生的身高，測得最接近吋為止，列表如下：

身高 (英寸)	學生人數
60-62	5
63-65	15
66-68	40
69-71	30
72-74	10
共計	100

將這些資料劃分為五組，與不同組發生的次數，形成一項次數分佈。此項資料可繪成直方圖，其中長條形，是與每組的次數成比例而構成的。

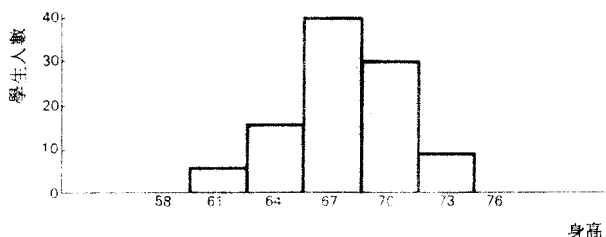


圖 3 表 2 資料的直方圖

如何繪製直方圖：

- (1) 在 5 與 20 組距 (class intervals) 間分佈觀察值。在例題 1 中 (60~62) 吋為組距。
- (2) 組代表值 (class mark) 是組距的中點。間距內的所有值，視為集中組代表值。
- (3) 決定每一間距的觀察值數目。
- (4) 用組代表值為中心，以面積與組次數成比例，繪出若干長條。若所有長條同寬，則高度與組次數成正比例。