科學圖書大庫

# 統計與機率導論

譯者　劉睦雄　張任業

徐氏基金會出版

021
1005

科學圖書大庫

# 統計與機率導論

譯者　劉睦雄　張任業

30297294

南京航空学院
藏书
图书馆

徐氏基金會出版

357974

# 我們的工作目標

文明的進度，因素很多，而科學居其首。科學知識與技術的傳播，是提高工業生產、改善生活環境的主動力。在整個社會長期發展上，乃對人類未來世代的投資。從事科學研究與科學教育者，自應各就專長，竭智盡力，發揮偉大功能，共使科學飛躍進展，同將人類的生活，帶進更幸福、更完善之境界。

近三十年來，科學急遽發展之收穫，已超越以往多年累積之成果。昔之認爲若幻想者，今多已成爲事實。人類一再親履月球，是各種科學綜合建樹與科學家精誠合作的貢獻，誠令人無限興奮！時代日新又新，如何推動科學教育，有效造就科學人才，促進科學研究與發展，允爲社會、國家的基本使命。培養人才，起自中學階段，此時學生對基礎科學，如物理、數學、生物、化學，已有接觸。及至大專院校專科教育開始後，則有賴於師資與圖書的指導啓發，始能爲蔚爲大器。而從事科學研究與科學教育的學者，志在貢獻研究成果與啓導後學，旨趣崇高，彌足欽佩！

本基金會係由徐銘信氏捐資創辦；旨在協助國家發展科學知識與技術，促進民生樂利，民國四十五年四月成立於美國紐約。初由旅美學人胡適博士、程其保博士等，甄選國內大學理工科優秀畢業生出國深造，前後達四十人，惜學成返國服務者十不得一。另曾贈送國內數所大學儀器設備，輔助教學，尚有微效；然審情度理，仍嫌未能普及，遂再邀請國內外權威學者，設置科學圖書編譯委員會，主持「科學圖書大庫」編譯事宜。以主任委員徐銘信氏爲監修人，編譯委員林碧鏗氏爲編輯人，各編譯委員擔任分組審查及校閱工作。「科學圖書大庫」首期擬定二千種，凡四億言。門分類別，細大不捐；分爲叢書，合則大庫。爲欲達成此一目標，除編譯委員外，本會另聘從事

翻譯之學者五百餘位，於英、德、法、日文出版物中精選最近出版之基本或實用科技名著，譯成中文，供給各級學校在校學生及社會大衆閱讀，內容嚴求深入淺出，圖文並茂。幸賴各學科之專家學者，於公私兩忙中，慨然撥冗贊助，譯著圖書，感人至深。其旅居國外者，亦有感於爲國人譯著，助益青年求知，遠勝於短期返國講學，遂不計稿酬多寡，費時又多，迢迢乎千萬里，書稿郵航交遞，其報國熱忱，思源固本，至足欽仰！

今科學圖書大庫已出版一千餘種，都二億八千餘萬言；尚在排印中者，約數百種，本會自當依照原訂目標，賡續進行，以達成科學報國之宏願。

本會出版之書籍，除質量並重外，並致力於時效之爭取，舉凡國外科學名著，初版發行半年之內，本會即擬參酌國內需要，選擇一部份譯成中文本發行，惟欲實現此目標，端賴各方面之大力贊助，始克有濟。

茲特掬誠呼籲：

自由中國大專院校之教授，研究機構之專家、學者，與從事工業建設之工程師；

旅居海外從事教育與研究之學人、留學生；

大專院校及研究機構退休之教授、專家、學者

主動地精選最新、最佳外文科學名著，或個別參與譯校，或就多年研究成果，分科撰著成書，公之於世。本基金會自當運用基金，並藉優良出版系統，善任傳播科學種子之媒介。尚祈各界專家學人，共襄盛舉是禱！

徐氏基金會　敬啓

中華民國六十四年九月

# 第四版序

本書係用作機率與統計預備課程一學期的教本，亦適用於每週四小時，一共十週的一季課程，或者是每週三小時的兩季課程。本書的第一版，係由曾被加州大學 Davis 校園，以及其他院校之教師們使用數年之久的油印及石版本修訂而得。在這些班級裡，數以百計的學生，幾乎都是來自自然及社會科學之各專門科系。還有若干是來自文學方面的。

在典型的美國學院及大學中，許多不同的科系都開有統計預備課程，許多專門學問都注意這門課程，而且對於這門課，也廣大地要求不同的數學預備知識水準。這種現象可能會導致一般人誤認爲基本統計在本質上依其應用範圍之不同而各異。關於這一點，統計學的數學政策似極重要。幾乎在每一門其他之科學中，均有一包含有相當完善之題材的預備課程。任何對於這門科學需要若干基本知識的學生，都可以修習這一預備課程。因此，幾乎每一學院或大學均開有普通化學這門課，凡需要化學基本知識的學生都參加。這樣看來，除了行政或傳統上的因素之外，似乎沒有什麼理由使我們不能對統計學亦採取同樣數學政策。誠然，將基本統計的數學集中爲一個系中的一個課程，可以得到很多好處，目前大多數院校均將這門課置於數學系，或統計系內。若僅開如此一課程，則可以避免許多不必要的重複，以及可以維持單一的數學水準。本書就是爲了適應此一目的而寫成的。

本書所需的數學基礎，只要相當於高中兩年代數的數學知識即可。在第十二章中，有一處用到三角學。但未曾學過三角學的讀者可省略此段落。而亦不致喪失其連貫性。由於本書所需的數學知識有限。故須敘述若干不加證明的定理。但幾乎所有僅需高中兩年代數知識即可證明之定理，我們都予以證明。當讀者發覺由於所需的數學知識有限所致使不加證明的定理相當少時，可能感到驚奇。但當他發現對於所省略的證明之理解，雖經修習初等微積分一兩門課仍不覺有所增進時，就不覺得驚奇了。因讀者通常在未修微積分之前，對於求極限之方法，可能不會熟悉。爲避免此數學上之困難，儘可能將討論限制於有限之情形。特別是，所有總體均假設爲有限。

在 Davis 修習統計課程的學生們，都是來自農業科學，商業管理，經濟學、家庭經濟學、心理學、社會學、地質學、以及醫學等各科系，爲適應此等不同的專業範圍，我們的例題及習題都是從這些科系的範疇內取材。

習題爲本書的一個主要部分。根據我們的經驗，統計學的預備課程之教授方法，以依照數學預備課程的教授方法最爲有效，即在上完每一堂課後，指定許多問題作爲家庭作業。我們特別注意在習題中儘量減少計算，而且避免機械計算（除第十五章的少數問題外）。因此常使我們不得不採用虛構的資料，不過此類資料在許多習題中均代表實際的實驗結果。

雖然在 Davis 校園內大多數修習統計初等課程的學生都是二年級的學生，但仍有一部分爲一年級的學生。我們特別忠告讀者，最好是儘早修習這門課。似乎沒有眞正的理由爲什麼在中學不能開這門課，因爲這門課的難易程度僅與三角學差不多。根據許多學生的反應，甚至於還容易些。提早開統計學的課程，可使學生將其知識運用於許多課程中，在這些課程裡某些形式的統計立可適用。統計學的提前講授，除了具有許多顯著的優點以外，還有一點，統計學實爲修習需要澈底瞭解代數知識的其他數學課程之前，一種複習中學代數便利且有效的工具。因此之故，我們的習題包括幾乎兩年中學代數所常見的題材——學者似覺得以解實際有趣的問題來複習代數，遠比傳統的方法輕鬆得多。

第一章至第十三章構成使用統計學之所有範圍所需資料之基本核心。第十四及十五章主要爲供給學商業管理和經濟的學生。所以特建議一般教一個學期統計課程的教師們，尤其是研究方向爲生物學的，應略去第十四與十五兩章，同時選擇下述的任一途徑作爲代替：其一，完全講授第一至第十三章。其二，講授第一至第十、十二、十三、十六及十七等章。教授一學期課程，研究方向爲社會科學的教師——尤其是研究經濟學的——會覺得以省略十六與十七兩章，而包括十一、十四、及十五等三章爲佳。如欲全部講授，則本書宜用於每週四小時的一個學期的課程，或者是每週三小時十週的兩季課程。

由於採取季制的大學數字日見增加，目前更多更多的院校，都希望將基礎的機率與統計，作每週四小時，一共十週的兩季課程授完。由經驗知，對於此一課程，本書之第一至第十三章可完全講授，但須略去下述各章節：5～4節、第十一章，12～6、12～7、12～8、12～9 等各節，以及13～5節。同時最好是將第十一章指定爲授完第十章後之參考資料。

現下之第四版與第三版不同之點，不僅爲許多小的改進，而且還有下述四點主要的改變。

1. 因由經驗顯示，許多同仁都希望將機率部分寫得比前版詳盡些。故我們已將第五章予以擴充。我們曾鄭重考慮，在寫此章時要以集合的術語來定義機率。但因多數學生尚未熟悉集合之基本概念，如此一來，我們必須另闢一章來討論集合。於是就要增大本書的篇幅，使我們無法在一學期或一季課程中，將機率與統計同時授完。此舉與本書的一貫目的相違背。所以經與許多同仁商討後，決定取消這個主張。

不過我們已將討論機率的一章稍加擴充。擴充的方法係在條件機率與 Bayes 定理加上一節，還加上許多說明的範例。尤其是有些範例是迎合學生的興趣而選的，例如點帽及生日問題便是。此外又加上許多困難程度不同的習慣。前 62 題是依照該章各節所包含之題材應順序而排列的，雜題則是困難程度遞增之順序而排列，從第 63 題開始。討論條件機率之 5 – 4 節，以及此節之習題（習題 28 至 36 題）可略去而不致影響其一貫性。目前由利用某種非參量性檢定所導致之優良結果的工作顯示，曾經長期考慮這些方法的懷疑主義，是不容許存在的了。似乎無庸置疑，最近所證明若干這些檢定的高度效率的驚人結果，將會引起在初等統計課程中對非參量檢定的更大重視。此外，大多數這些檢定方法，對於數學知識的要求不多，當然不需要微積分的知識。所以這些檢定方法的內容，似乎特別適合於僅需兩年中學代數基礎的課程。

因此之故，第十一章在第三版中，除符號檢定外，已擴充至對小與大樣品之 Wilcoxon 兩樣品檢定，現在此版中，又更進一步擴充至包括在對於不成對與成對兩比較法之級（rank）中有相持（ties）之情形。

3. 第六章已擴充至好像更從容地引入了二項分配。

4. 大約六十五題難易程度不同的習題加入第五至第十三章，少數習題已被刪去，一部分重新修正，又有少部分已加以改寫使其意義更爲明顯。

我們要感謝 Cambridge 的 Sir Ronald A. Fisher 教授，Rothamsted 的 Frank Yates 博士，以及感謝 Edinburgh 的 Messrs. Oliver and Boyd 有限公司允許我們從他們的 Statistical Tables for Biological, Agricultural and Medical research 一書摘要重印表 Ⅱ、Ⅲ、及 Ⅳ。

我們衷心感謝 Hubert A. Arnold, George A. Baker 及 Curtis M. Fulton，當他們在他所教授的班裡使用這本書早期的草稿時，曾給我許多有價值的建議與批評。我們亦要感謝 Jerry Foytik 教授對於我們原稿的忠告與建議，尤其是在第十五章中有關經濟學方面之資料的批評。特別

感謝 Vi Raggi 太太對於手稿最仔細與最正確的打字。

Henry L. Alder
Edward B. Roessler
1967年12月

# 目　錄


第一章　引言……1

第二章　資料之處理……4

2-1. 引言……4
2-2. 表現資料之表及圖解法……4
2-3. 度數分配及其圖表以及圖解表示……13

第三章　連加記號……20

第四章　資料分析……27

4-1. 引言……27
4-2. 集中趨勢之測度……28
4-3. 散佈之測度……38

第五章　初等機率、排列、及組合……52

5-1. 機率的定義……52
5-2. 期望值……57
5-3. 三個初等機率定律……58
5-4. 條件機率……65
5-5. 組合與排列……69
5-6. 重複試驗……73

第六章　二項分配……86

第七章　常態分配……93

7-1. 視同二項分配之近似之常態分配……93
7-2. 視同一連續變數之度數分配的一極限的常態分配……98
7-3. 在常態曲線下之特殊面積……100
7-4. 曲線上之級配……101

第八章 隨機抽樣與大樣品理論……106

8-1. 引言……106
8-2. 樣品平均數的分配……108
8-3. 樣品標準偏差之分配……113
8-4. 兩樣品平均差的分配……113
8-5. 標準誤差……117

第九章 檢定假設、有意水準、信賴界限、大樣品法……120

9-1. 檢定假設……120
9-2. 信賴界限……127

第十章 Student之t-分配、小樣品法……133

10-1. 全域標準偏差之最佳推定……133
10-2. Student 之 t- 分配……134
10-3. 平均數之分配……135
10-4. 平均數差之分配……139
10-5. 配對變量之情形……142

第十一章 非參數統計……152

11-1. 引言……152
11-2. 非配對情形之Wilcoxon 兩樣品檢定（試驗）……152
11-3. 配對情形的符號檢定（試驗）……161
11-4. 對於配對情形的Wilcoxon 檢定（試驗）……163

第十二章 迴歸與相關……171

12-1. 引言……171
12-2. 線性迴歸方程式……172

12-3. 估計之標準誤差……180
12-4. 相關係數……182
12-5. 相關係數之顯著性……186
12-6. $Y$在$X$上與$X$在$Y$上迴歸線之比較……192
12-7. 過原點之迴歸線……193
12-8. 多項式迴歸……194
12-9. 多重迴歸……195

第十三章 卡方分配……201

13-1. 定義……201
13-2. 卡方之分配……202
13-3. 在發生學上之應用……203
13-4. 對於列聯表之應用……204
13-5. 對於常態試驗之應用……207
13-6. 調整之卡方……208

第十四章 指數……217

14-1. 引言……217
14-2. 基期之選擇……217
14-3. 簡單指數……218
14-4. 加權指數……220

第十五章 時間序列……224

15-1. 引言……224
15-2. 長期傾向……226
15-3. 季節性變化……232
15-4. 循環消長……240

第十六章 F分配……247

16-1. 定義……247
16-2. 試驗兩分散之齊次性……250

第十七章 分散分析，一個分類準則……255

**17-1.** 引言……255
**17-2.** 平方和之分割，大小相等樣品……256
**17-3.** 平方和之分割，大小不等樣品……260
**17-4.** 分散分析表中的分散比較，大小相等樣品……262
**17-5.** 最小顯著差……266
**17-6.** 分散分析表中分散之比較，大小不等樣品……268
**17-7.** 兩個樣品的特例……270
**17-8.** 具有一個分類準則的分散分析中的假設……272
**17-9.** 變換……274

**附錄**……283

參考書籍介紹……285
表Ⅰ 常態機率曲線下之面積……288
表Ⅰa 常態機率曲線之縱坐標……291
表Ⅱ Student的 $t$ 分配……294
表Ⅲ 由 $r$ 至Z之變換……295
表Ⅳ 卡方分配……296
表Ⅴ F分配……298
表Ⅵ Wilcoxon 分配（無配對）……304
表Ⅶ Wilcoxon 分配（有配對）……305
表Ⅷ 由百分率至 $\sin^{-1}\sqrt{\text{百分率}}$ 之變換……307
表Ⅸ 平方與平方根……310

**奇數習題答案**……320
**索引**……327

# 第一章　引　言

統計學爲一門處理數值資料之搜集，組織，分析及解釋之科學。

“資料之搜集”爲獲得數量之步驟。當吾人適當的搜集或由代表性之資料，才能獲致有效之結論。雖然此爲統計之程序中極重要之部分，由於簡潔起見，吾人將考慮所處理之資料業已有用，而不再深入討論。

“資料之組織”爲將所搜集之數或量以一適合於導出合乎邏輯之結論之形式表現出之工作。組織之代表性方法以及藉著表及圖所表現之資料，將於第二章中論及。

“資料之分析”爲由已得之數或量抽取相關之情報之程序，由此可以形成一摘要及可資理解之數值描述。爲此目的常用之最重要數量爲平均數，中位數、變程、標準偏差及其他，此等數量將於第四章中論及。

“資料之解釋”爲由資料之分析而取得結論之工作，且時常包含由相同物質中之少量搜集之情報而推測該物質大量搜集之資料。資料之解釋乃成爲此文之主要部分。

統計學乃成爲處理能以數值情報作有限度解答問題之科學，亦即由計算或計量所得之情報。在搜集時，對於生物學之研究中所作之昆蟲數量之搜集，或者對於工廠中工人或工時之調查，兩者顯然有些不同。而統計學者之職責，首先爲選擇情報所需之種類，然後指出適當及有效之搜集方法且繼續此情報之程序，最後將此結果表示。在表示此結果時，特別當資料爲未完全時，統計學家必須應用已經得到有效的結論之法則及技巧。在面臨這未確實性中，統計學家必須有一明智之決定。

統計學一辭有二極大不同之意義。當其作爲一單有名辭時，其意義誠如上　所述。此在科學之過程中，作爲數值資料之研討及評價之用。當其作爲一複數名辭時，其與“數值資料”一辭爲同義字。因此，吾人若言在世界年鑑（World Almanac）或美國之統計摘要（Statistical Abstract of the United states）中有統計數字，此乃意指其中含有數值資料；此爲較古老且更一般之意義。本來，統計學乃做爲搜集資料以便政府首長處理行

政事務之用。欲將如此之情報表成數字資料，吾人必須回溯到Aristotle及其所發表之論文“國家之行政”(matter of state)上，事實上，吾人能印證“統計學”與“國家”二辭乃由相同之語根所導致。在古代，許多文明的國家，爲軍事及財政上之理由，已經編輯大量之“統計量”以便確定一個國家之人力及物力是否雄厚。在聖經中，吾人可研讀到如此之人口調查，且在羅馬帝國之大部分地區，此爲便於抽稅所常採取之措施。

在十六及十七世紀，機遇之遊戲特別盛行於貴族中；且由他們提出許多包含機率之有趣問題，而由當代有名之數學家如 Cardano, Galileo, Pascal, Fermat, Leibniz 以及 James Bernoulli 所解決。

同一時期，統計學之發展亦有不同之進展。在英國，John Graunt 作了關於重要之統計學、保險學以及經濟統計學之半 - 數學研討。其工作爲 William Petty 爵士及 Edmund Halley 所推廣。首先將理論視爲一整體之論文爲 Jacob Bernoulli 所提出，他詳細說明大數法則之原理，而此在Poisson 之論文中亦曾述及。在1773年，正規分配爲 de Moivre 所發現，Quetelet 證明生物學及人類學之測度依循相近於Gaussian之誤差論。在德國，Knapp 發覺大量關於死亡率之統計數字，且 Lexis 從而發展一即今日稱爲一線分散分析之步驟。關於此理論，其他之貢獻者爲 Stirling, Bayes, Daniel Bernoulli, Euler, Nicolas 及 John Bernoulli, Simpson, D'Alembert, Lagrange, Buffon, Montmort, Condorcet, 以及偉大之 Laplace 。後者提供最具發展性之論文於其不朽之著作“機率之解析理論”(Thèorie Analytique des Probabilitès)上。

於十九世紀之後十五年，Francis Galton 爵士，爲英國優生學學校之創辦者，及其得力之助手 Karl Pearson 利用發生學所產生之問題，發展了迴歸及相關係數。爾後，Pearson 及 Spearman 將此理論推廣，且應用至社會科學之研究，Pearson 並且廣泛的研究抽樣誤差之效用理論。

在此世紀之初葉，服務於一愛爾蘭之啤酒廠之 Guinness 之統計學家 William S. Gosset 以“Student”之筆名發表了許多關於統計分析之文章。他特別著重於由抽樣所獲取之資料之表示法。首先，他發展由少量樣本中抽取信賴情報之方法，由 Gosset 所發展之方法，後來在英國，由 R. A. Fisher 及其同事將其理論及應用之領域加以推廣，特別是在農業方面。

吾人每日之生活越來越由奠基於定量之情報之決定所左右。對於活動領

域之問題，以科學之程序，即臆說，經驗以及臆說之測定爲今日熟悉之過程。今日，時新之統計方法奠基於機率之理論，乃爲物理學及生物學、經濟學及社會學、心理學及教育學、醫學及農學，以及政府及工業上不可或缺之助力。天文學家預測天體之未來位置，主要是根據統計之方法；符合於發生學之分離爲一確定之統計量；生命保險之保險費及年金之付給亦由根據統計記錄所得死亡數之統計表所決定；電力公司若無負荷需要量之統計資料將無法有效供應電力；研究工作者由統計之商榷決定農場試驗所得結果之意義；工程師發現抽樣理論對於生產品質之控制相當有效；商業之主管以及政府之分析學者利用統計之程序以便決定事務之處理，雖然，其具有廣泛不同領域之應用，而大部份統計方法之使用則一致。統計分析之某一觀點，可能在某一領域之應用較其他方面著重，但是一般而言，相同之統計程序可適用於整個領域。

在統計學之研究過程中，有一二言值得吾人注意。最重要的是，吾人必須瞭解統計程序之本身並無法保證其能避免產生誤差、錯誤、不充分之理由或不正確之結論。原有之資料必須正確，所採之方法亦需適當，且其結果須由不但瞭解其方法亦爲應用者本身之領域之學者加以闡釋。在此書所討論之統計方法，僅做爲以適當之方式及應用至所設計之情況，而能產生有用之結果之工具，其本身並無法創造奇蹟。

# 第二章　資料之處理

## 2-1 引言

通常，情報之搜集常需大量之資料，而此項資料若可資瞭解或有效地表示，吾人可以某種形式將其摘錄。明晰及有力之表示將對此資料之瞭解及正確之表示有一重大之助益。一般，吾人使用二種表示定量資料之方法，一種方法包含數量本身之摘要表示，通常以“表”之形式表示，另一種方法爲將定量資料以“圖”之形式表示——圖解，圖表，或其他類似之表現。一群資料以此二種方法表現爲統計分析之一部份，而此可以導致對此項資料之較佳理解。

## 2-2 表現資料之表及圖解法

幾乎所有科學及商業之出版物，政府之施政報告，雜誌、及報紙、五花八門之資料均藉著表、圖或照片將其說明之。

吾人時常將定量之資料視爲本文材料之一部份，但若以表說明時，將有某些助益。根據聯邦準備銀行制度之報告，在美國大陸所有銀行之總存款，於1930年爲58.1 兆美元，於1935年爲55.4兆美元，於1940年爲 76兆美元，於1945年爲165.5兆美元，於1950年爲175.3 兆美元，而於1955年爲220.4兆美元。毫無疑問的，此種表現法將使人感到索然無味，而無法把握此資料之重點。實際上，吾人可將此種表現以一欄表之形式表示，而使人一目瞭然。相同之情報可以下述之方式表示：在美國大陸之所有銀行，從1930年至1955年，每五年爲一區間之總存款（以兆元爲單位）爲：

| 年　度 | 總存款 |
|---|---|
| 1930 | 58.1 |
| 1935 | 55.4 |
| 1940 | 76.0 |

| | |
|---|---|
| 1945 | 165.6 |
| 1950 | 175.3 |
| 1955 | 220.4 |

此種排列使得吾人更加明瞭此資料之重要特性。吾人立即明白在 25 年間，銀行之存款急劇上昇，且在 1940 年及 1945 年之間，幾乎上昇一倍有餘。此種特徵，以此欄表說明時，將更加顯得清晰。然而，欲完全瞭解在右方欄表之資料之意義。吾人必須描述在原文中之資料，由於此，吾人依然必須以原文之表示法描述之。

在表之表現中，資料以行及列排列，此形式尙保留欄表之特性。附加於此表之說明，必需包括所有之情報，以便使此表一目瞭然，因此，表之表示法可以省略原文之說明。關於上述關於銀行存款之資料，可以表 **2-1** 之形式表示之。

表 **2-1**　由 1930 年至 1955 年，每五年爲一區間，在美國大陸所有銀行之總存款（以兆元爲單位）。

| 年　度 | 總 存 款 |
|---|---|
| 1930 | 58.1 |
| 1935 | 55.4 |
| 1940 | 76.0 |
| 1945 | 165.6 |
| 1950 | 173.3 |
| 1955 | 220.4 |

來源：聯邦準備銀行制度之主管局。

關於表之編號一般用以迅速及正確之識別，且任意在原文中所提之資料，可以由特別表之數字加以引證。

**定義 2-1**　一可測之特性，稱爲變數。

**定義 2-2**　一變數之個別可測量，稱爲變量。

例如在表 2-1 之資料，銀行之總存款爲一變數，而每一存款之登錄數爲一變量。