

生物統計學新論

楊志良 著

巨流圖書公司印行

生物統計學新論

羅志良 著

中國農林出版社

北京

生物統計學新論

楊志良 著

◎ 巨流圖書公司印行

生物統計學新論

民國 72 年 2 月一版一印

民國 72 年 9 月一版二印

版權所有 不准翻印

出版登記證：局版臺業字第 1045 號

著	者：楊	志	良
發	行	人：熊	嶺

印行者：巨流圖書公司

臺北市博愛路 25 號(泰華大廈) 613 室 100

電話：(02)3148830•3711031•9246285

郵購：郵政劃撥帳戶 100232 號

定價：臺幣 200 元

如有裝釘錯誤
即請寄回調換

作者簡介

楊志良

1946 年生，台北市人。1968 年畢業於師大衛生教育系，1972 年獲台大公共衛生學碩士，1979 年獲美國密西根大學公共衛生學院哲學博士。

1972 年起執教於台灣大學，講授生物統計，人口與家庭計劃及公共衛生導論等課程。

Q-332

7

目 錄

自 序	1
第一章 統計學基本邏輯及變項種類	5
第二章 母全體、樣本及抽樣	15
第三章 等距變項的統計：等距變項的敘述	29
第四章 等距變項的統計：常態分佈及其應用	45
第五章 等距變項的統計：單一變項的推論	63
(若母全體標準差為已知)	
第六章 等距變項的統計：以二組樣本推論二個母全體	89
(若母全體標準差為已知)	
第七章 等距變項的統計：以一組或二組樣本推論一個或	
二個母全體 (若母全體標準差為未知)	105
第八章 二個類別變項的統計—卡方檢定	129
第九章 二項類別變項 (比率變項) 的統計	151
第十章 變異數的估計與檢定	165

第十一章 數組樣本間等距變項之檢定：一方分類之變異

數分析 177

第十二章 二個等距變項間之統計：簡單直線迴歸及相關 191

第十三章 兩個變項間關係的各種統計方法之比較 215

第十四章 多變數統計分析：複迴歸與複相關 235

第十五章 無母數的統計檢定方法 245

第十六章 生物醫學資料之電腦處理 259

重要統計分佈表

隨機號碼表 26

常態分佈表 (Z-表) 55

t-分佈表 108

卡方分佈表 135

F分佈表 172

重要參考書目 269

附錄：習題解答 270

自序

生物統計學是從事生物健康科學者的工具科學。但至目前為止，不少生物與醫學工作者及研究者仍對它感到陌生，更談不上在運用上得心應手。檢討起來，原因相當複雜：一方面，由於擔任教學者未能認真負起責任，受教者也多有輕忽的態度；另一方面，目前生物統計學的課程在醫學院內多排在一、二年級，因年級低，多未能體會生物統計對其專門科系在工作上及研究上的重要性，等到高年級，甚或畢業後從事研究時，才感到種種的不便與困難，就不免有「書到用時方恨少」的感覺了。

當然統計學的本身也是造成困難的主因之一。因它包涵了不少的邏輯與機率概念，這些概念常不易表達及被全然瞭解，使得講授者不易討好，研習者亦難引起興趣。這些統計學上的邏輯與概念也很難讓研究者以「速成」的方法加以充分掌握與運用。因此要以深入淺出的方式來介紹生物統計學並非易事。拙作希望曾修讀過生物統計者能溫故知新及有助於整理思緒之效。對於這些讀者，希望他們在讀完第一章以後，跳讀第十三章，然後再選擇研讀有興趣的章節。對於初學者，則請注意下列數點：(一)本書特別著重於介紹變項的種類及測量，並強調在統計及處理不同的變項資料時，應選用適當的統計方法。(二)統計學的基本目的，除在於敘述資料以外，係以樣本的統計值來推論母全體的母數，故對統計檢定的假設，樣本統計值及母全體母數間的關係特別著重。(三)「常態分佈」有許多層次，如原來每個個體的測量值的分佈，「樣本平均值」的分佈，「二個樣本平均值差別」的分佈等

。這是初學者最易混淆而難於瞭解之處。此書在這方面特別多加介紹。 (四)就初學者而言，對某項資料應選用何種統計方法常感困難。故書中對基本統計方法間的相通性及差異性嘗試加以比較，以有助於作適當的應用。

如衆所知，國內的環境並不適合教科書的寫作，特別是較冷門的教科書。理由很多，如課多事繁，版權的保障不彰，寫作的收入微薄等等，不一而足。個人固有同感，然竟知其不可爲而爲之，實係諸多先進好友的鞭策，未容稍懈。

談及此書的寫作，頗多值得憶念的，那是在七十年暑假，承公共衛生前輩泰斗王金茂先生之推薦、鼓勵及阮院長仲垠之邀請，五度前往高雄阮綜合醫院演講「生物統計」。同時期，省公共衛生研究所王所長國裕及該所譚主任文海邀往該所對同仁講授全系列之同一科目，使筆者不得不在尋找合適課本失敗之餘，提筆自行寫作講義。至該年秋，受命於公共衛生研究所，開課講授「生物統計推論」，只得每週乙篇，繼續將該系列講義完成。想到當時該班同學們的熱情：或協助抄寫、或提供意見、或指出錯誤，坦誠相與，均足令人感動。還有更值一提的，是「當代醫學社」的蔡敦仁博士虛懷下顧，於蒞臨旁聽課程之餘，力邀筆者將講稿以專欄方式逐章發表，其盛情與鼓勵，更使筆者感奮於心。該社並同意筆者在專欄刊載完畢後，自行印行單行本，亦值稱道。

七十一年上半年，同仁張珏、蘇喜、李蘭、吳淑瓊諸老師，於擔任醫學院內各系之生物統計課程時，以拙作原稿爲教材之一部份，提供不少高見。其中吳淑瓊老師更對拙作之習題解析、誤謬指正及校對抄寫惠與許多協助，鄭玉娟老師則中途拔刀相助，花費大量時間，逐字審閱，指正甚多。公共衛生研究所同仁洪琇清先生代爲繪製書中圖

表，至為辛勞。又巨流圖書公司負責人熊嶺先生及編輯趙天池先生，熱誠支持本書的出版，我們一見如故，至感投契。以上來自各方的鼓勵與協助，均值感懷，於此並致謝忱。

最後，也是最重要的，拙作真若對我國健康科學有少許貢獻，則最應歸功於導引筆者對公共衛生教學與研究產生濃厚興趣的本所已故陳拱北教授及大學時的恩師李叔佩教授，還有生物統計啟蒙師的吳新英教授及以電腦在統計上運用的林家青教授。

拙作膚淺謬誤之處必不能免，祈望各界不吝指正。為自我策勵，且將本書當為「上册」，希望在未來能將多變數統計方法、生命表及其運用、衛生統計及統計在檢驗室之品質管制等運用上，逐一完成合而當為「下册」，以力求系統的完整。

楊志良 謹誌

民國七十一年十二月五日於
台大醫學院公共衛生研究所

第一章

統計學基本邏輯及變項種類

一、概 論

統計學 (Statistics) 是什麼？簡單的說，就是對觀察到的或測量到的資料，如何加以處理，以及處理後的資料如何加以利用，以協助研究者做判斷及推論的一門學問。如果處理或分析的對象是有關生物的事務或現象，那就是生物統計學了 (Biostatistics)。

如果依照 (生物) 統計學的功能來分，統計學可分為二個主要部份：敘述統計學 (Descriptive Statistics) 及推論統計學 (Inferential Statistics)。例如有 5, 10, 50, 90, 95，五個資料 (當然也可以是 50 個，甚至五萬個或任何數量的資料)，我們可以利用幾個指標 (index) 來敘述這些資料。例如其個數為 5，最大值為 95、最小值為 5，平均值為 50 等等。當然在前述的例子裏，因只有 5 個資料，且以 5 或 0 為尾數，一下子就可全部明瞭，因此用這些指標來敘述上述 5 個資料，對於這些資料之瞭解並無多大幫助。但如果是 500 個人的身高或 1,000 人的血壓資料，記錄下來密密麻麻，恐怕從頭到尾看了很多次，還是不能對這些資料有什麼整體的印象。若能用幾個指標加以敘述，例如 500 人的身高最高是 185cm，最矮是 143cm，平均是 166cm 等等，則由繁化簡，使研究者很快就可對很多的資料，有綜合性的瞭解，並掌握其特性。當然，敘述資料的指

標與方法很多，將在本書其他章節中討論。

統計學的另一個功能是協助人們做推論或下結論，舉個例子來說，由筆者做莊家，與諸位讀者用骰子賭博。賭時用四粒骰子，每粒六面，每面的點數最小是 1，最大是 6，賭時莊家先擲，若四粒出現之點數均同，則表示「最大」通贏（通吃），押注者不必再擲而賭注一律由莊家收取。現由筆者為莊家先擲，第一次就擲出四個 6 點，因此通吃。第二、第三，連續至第四次均如此，次次出現四個 6 點通吃。則請問讀者諸君將做何反應？想必判定莊家舞弊或骰子不均勻，因而採取必要之行動（從要求退還賭注到揮拳相向）。然而是如何做此種的（莊家舞弊或骰子不均勻）判斷或推論的呢？此判斷的邏輯是什麼？平常大家只是直覺認定，並沒有詳細的加以考慮判斷的過程。現把它分析為下列諸點：

(1) 每粒骰子出現 6 點之機會為 6 分之 1，擲一次四粒均出現 6 點的機會為 $(1/6)^4$ ，而連擲四次，每次均出現四個 6 點的機會為 $((1/6)^4)^4$ 或 $(1/6)^{16}$ ，亦即 3.5×10^{-10} 。

(2) 先假設莊家沒有舞弊或骰子沒有不均勻。在此假設下，莊家連擲四次，每次均出現四個 6 點的機會極小 $(1/6^{16})$ ，雖然不能說完全不可能發生，但發生之可能性幾乎接近於 0。

(3) 因假設莊家沒有作弊而連續出現四次 4 個骰子都是 6 點的機會非常小，亦即假設成立的機會很小，因而推翻原先之假設，認為莊家舞弊，而採取必要之行動。

在上述的判斷過程中，(1)項為敘述性的統計，敘述觀察到的事件，(2)與(3)項就是推論統計了。

因此要做統計推論之先，必須先做敘述性的統計。但資料加以敘述以後，不一定會適合做統計推論，這必須依情形而定。（將在以後

章節中說明)

二、資料的種類

資料 (data) 的獲得必需靠觀察或經由測量而來。資料中的單一項目稱為變項 (variable)，例如體檢資料中，身高為一個變項，性別、體重、血壓、視力等均各別為一個變項。資料或變項依其觀察或測量的特性可分為下列三大類：類別變項 (nominal variable)，序位變項 (ordinal variable)，等距變項 (interval variable) 及等比變項 (ratio variable)。不同類別的資料，要用不同的方法或指標加以敘述，在做推論時也要用不同的統計方法，否則非但牛頭不對馬嘴，且常會做出極其荒謬的結論。因此，判定變項的類別極為重要。很多醫學相關系、科的學生學了統計以後，在應用時常感到困惑，主要原因也就是以往的生物統計課本，在這方面的討論常不詳實，而致使學習的人雖面對資料，但常不知道從何着手。

(一)類別尺度(nominal scale)

類別尺度的變項，顧名思義，是在測量或觀察時，將對象加以分類所獲得的變項。例如人可以分為男人及女人 (性別)，或中國人、美國人、英國人、日本人等等 (種族別)。再如顏色種類、水果的種類、疾病的種類等等。通常為了點計 (counting 或 enumerating) 的方便，特別是利用電腦處理時，常將類別變項的各類別加以編號，每一類賦予一個數字。例如男為 1，女為 2。又如中國人為 1，美國人為 2，英國人為 3，……等等。在統計時，只要點計有幾個 1，有幾個 2，……等，就可知道有多少男人、女人，或者中國人、美國人

、英國人。在利用電腦處理資料時，類別變項幾乎全部運用此種方式。顯然的，類別變項的各類別雖賦予一個數目，但數目只是代號而不代表任何關係。男為 1，女為 2，並不代表男為女之二分之一，也不代表女大於男（ $2 > 1$ ），因為也可以把男給予 2 的代號，而女給予 1 的代號。數目的給予完全是任意的或武斷的，人種別亦是如此。

(二)序位尺度(ordinal scale)

此種變項在顯示某項測量或觀察對象的等級。例如我們可以請教某人對崔苔菁、鳳飛飛及張琍敏三人的喜好程度，若回答為最喜歡張琍敏，次喜歡鳳飛飛，再次喜歡崔苔菁。則為了方便，我們可將張給 3 的數目，鳳給 2 的數目，崔給 1 的數目。（當然也可反過來，張給 1，鳳給 2，崔給 3），3 大於 2，2 大於 1，3 大於 1，代表喜好的強度序位。亦即喜好張大於喜好鳳，喜好鳳大於喜好崔，而喜好張也大於喜好崔。如果反過來排列，1 代表張（最喜歡）亦無不可，其排列所代表的意義與順序仍然不變。此地的數目，顯然地，較類別變項的數目具較多的意義與訊息。但此種序位尺度的數目只代表序位，而不能說 3（喜歡張的程度）是 1（喜歡崔的程度）的三倍。亦即我們僅知道其程度或強度的不同，而不知其間的差別有多大。某人可能喜好鳳的程度僅只略微低於喜好張一點點，却極厭惡崔；也可能極喜好張，而非常厭惡鳳及崔，而厭惡鳳的程度略少於崔而已。此二種情形無法從序位中分辨出，因二者喜好的序位是一致的。很多關於態度的測量資料多屬於序位變數。例如「您贊成家庭計劃嗎？」(1)非常反對，(2)反對，(3)無意見，(4)贊成，(5)十分贊成。又如考試成績也應屬序位變項，我們不能說考 90 分的人比考 10 分的人懂的（或學習到的）程度大 9 倍。85 分比 80 分多 5 分，65 分比 60 分多 5 分，

但也不能說相差一樣多。我們只能說，如果試題適當批改又公正的話，90 分的人比 85 分的好，85 分比 80 分好，再來是 65 分、60 分，而 10 分最差。（又如硬度亦是序位變數）

(三)等距尺度(interval scale)及比率尺度(ratio scale)

顧名思義，此種變項 1 與 2 的差別，101 與 102 與的差別是相同的。（1 cm 與 2 cm 的差別與 101 cm 與 102 cm 的差別均是 1 cm，一樣長）。也就是說每個單位都是一樣大小。此種資料（變項）所具有的訊息最為詳實。舉凡長度、重量、時間、溫度及它們所導出的種種度量（如面積、比重），都屬等距變項。等距尺度中，具有絕對的 0（亦即包涵了「沒有」者），稱為比例變項（ratio）。例如長度、時間、重量可以為 0。因此，此種比例變項可以計算倍數。例如體重 80 公斤的人是體重 40 公斤者的 2 倍。但如溫度，雖是等距尺度，但不能計算倍數。我們不能說氣溫 40°C 是 20°C 的兩倍，因為 0°C 是武斷給予的，是人們為了方便，以水的冰點為 0°C 。我們當然也可給予某一金屬之融點（如 1000°C ）定為零點。如果以該一金屬融點為準，則通常所稱氣溫 40°C 則為該金屬融點零下 960 度， 20°C 則為零下 980 度，則顯然二者不是 2:1 或 -960 度是 -980 度的二倍了。溫度是不能「沒有」的，（人死了，量不到體溫不能說不能量溫度或沒有任何「溫度」），因為沒有真正的零，因此也就不能算倍數了；雖然不能算倍數（比例），但因為是等距變項，因此可以算平均。例如某地區之年平均氣溫。

至於有幾個人，生幾次病，有幾個蘋果都是等距變數，且是比例變數，因為有真正具有意義的零及可以計算倍數或比例。

三、各類變數間的關係及轉化

類別、序位及等距（或比例）變項間有何關係呢？可不可能把等距變項轉化成序位或類別變項呢？或者反過來把類別變項轉成等距變項呢？現在以十個人的體重做例子，加以說明：

44.8	52.1	52.4	52.5	55.8	56.3	58.6	58.7	62.3	62.5
1	2	3	4	5	6	7	8	9	10
(1)輕	(1)輕	(1)輕	(1)輕	(2)重	(2)重	(2)重	(2)重	(2)重	(2)重

上述十個人的體重，如果以公斤計算，當然是等距（比例）資料，我們可以知道最重的人（編號為 10，體重 62.5 公斤）比次重者（62.3 公斤）多了 0.2 公斤，而次重者比第三名多 3.6 公斤，我們也可以說最重的是最輕的（44.8 公斤）1.395 倍，我們也可以把全部體重加起來除以 10，而得知平均體重為 55.6 公斤。如果這十個人的體重不用體重計來量，改用翹翹板來測量（當然比體重計的測量來得粗略），每次二個人比誰重，最重的號碼大，最輕的號碼小，則可將 10 個人從 1 號排到 10 號，就成了序位變數，第 10 號的比第 9 號重，也比其他任何一位重。而第 5 號比第 1, 2, 3, 4 號要重，而比第 6, 7, 8, 9, 10 號為輕，因此我們知道誰重、誰輕，也就是各個人體重間的次序，但我們不知道第 10 號與第 9 號相差幾公斤，也不知第 4 號與第 5 號差多少，更不知道這二個差距是不是相等。如果第 1 到第 9 號總共只相差 1 公斤，而第 9 號與第 10 號相差 30 公斤，或者每一號間均差 2 公斤，排列起來的順序都是一樣的。因此我們絕不能只根據序位，而說 10 號是 1 號的十倍重（ $10 \div 1 = 10$ ），因