新編

生物統計學

包興林　著

新竹黎明書店經銷

新編
生物統計學

新編 生物統計學

包興林 著

新竹黎明書店經銷

新編生物統計學

中華民國七十三年四月初版

編著者：包興林
發行人：楊國藩
發行所：國興出版號
新竹市西門街262號
電話：(035)223129號
行政院新聞局局版台業字第0465號
總經銷：黎明書店
新竹市中正路72號
電話：(035)229418號
郵政劃撥七〇二七三號

實價：200元

目　錄

第一章
統計學係搜集分析與整理統計資料並推論其結果的科學

一、一般統計資料的處理應保留多少位數合宜須加斟酌

幾乎世界上所有的資料，均須符合下述四種特點，方可叫做統計資料：

①數量性，所謂的數量性，即是可以數字來表現它特點的資料，如果是品質資料，則必須事先轉變爲數量資料之後再加以整理。

②眞實性，資料的統計，必須根據測量或登記、調查、實驗獲之，來自眞確而客觀之事實，並不是個人主觀意志來決定。

③變異性，統計上所探討的課題，絕大多數屬於變異的問題。如果一份資料中各方面的特性量的特徵或性質的特徵完全相符時，則此份資料極爲單純，即不必利用統計方法探究之。

④群體性，統計最重要的，是在研討群體的數量特徵，而不是零散的個別情事。譬如觀察中華民國的婦女，在生育期所生育的子女總數，而不是以特定的某一婦女或某一城市的婦女而言，而是必須囊括台澎金馬所有地區爲範圍的。

一般所稱之統計資料，是根據個體特性以點計（數數）或度量（測驗）自然景象、生物景象或社會景象各方獲得之群體的資料。這種統計資料，必須含有時間、空間與質量三個要素，群體內個體要素的多寡，和群體的恰成相反，群體中個體的要素多，群體的範圍也就愈小，反之

則大。

以生物現象的「人」來說，倘若僅根據「人」生理上必備的質的要素爲主，（例如靈活運作的雙手、直立行走的特徵、以及胎生等等），而沒有參雜時間、空間以及量等要素爲範疇，如此，地球上自古至今所有的人類，便都屬於人的群體。如果參與時間、空間以及量三大要素及其他更多層面的條件，則其群體的局限就將會狹小許多了。

統計學可分爲應用統計學和純理統計學兩種，應用統計學是探討統計的方法於各種現實問題的處理上的有效運用，純理統計學則探討統計方法的原理，在本書中討論的生命、人口、醫學以及生物統計學，都是應用統計學的一部分。

我們所說的統計學，就是採集、整理和統計資料的分析，並且推論與詮釋最後結果的一種工具科學。依據我國所謂的統計學這個名詞，乃是從英文「Statistics」直接翻譯過來，這個字的德文拼音是「Statistik」、法文是「Statistique」，它們的讀音都頗爲近似。直接從意大利文「Statisti」此字轉變而成，原本的意思是政治家「Statesman」；間接的起源是依據拉丁文「Status」而得之。「Status」原來的意思，是說明一國的政治情況（A political state）。然而，政治情況主要的也就是人口和財富，尤其是以人口爲最重要的現象，人口是一國情勢表徵，國家爲了需要徵兵，理所當然要詳加調查及登記這國家的武力根源——人口，來作爲計劃兵源的根本依據。

最早的統計是起源於政治的需要而對人口大概的統計而來的。

人類的知識，難以千萬計，粗略的區分，大致可分爲三種。一是經驗的知識，必須要長時間的親身體驗、經歷方可獲得，當時的人們往往只知其然而不知其所以然；第二種是科學的知識，不僅知其然，並也知其所以然，是經過專門的研究探討和反覆的實際應証或客觀的查証而獲之。就如同當今進步的文明社會中衆多機械工程、電、力、聲光與衞生

醫學各層各面的原理探究及發明，都是歸屬科學的知識。

第三種卽是逾越經驗知識與科學知識的統計知識，譬如生命現象中的「巨數恒性」的知識、生物在遺傳上顯性與隱性比例關係的知識、醫學治療上各個年齡以及不同性別的各項疾病發生率的知識，都是屬於這方面的知識。如果人類的知識中，缺乏了統計知識，則相信必有許許多多難以理解或茫然無知的情形發生。

天地之間的各種人、事、物，都是參差不齊、變化多端、難以捉摸，甚至時而消失、忽又出現。無論它如何的變異，只要仔細觀察，統計、研究，仍是有跡可尋的，這就是統計現象在一個龐大群體時，多是處於常態分配的情況，就拿人類的體質來說，人們的高矮、胖瘦、身強、體弱，普通是以中庸者爲多數，而愈向兩方極端則愈少。統計人員就依照此種事實，可用一部分來代替一個龐大群體，來攫取所需的代表數值，來應用於同一種類的事物。醫學與生物的研究實驗上，大部分應用統計現象的常態分配來整理資料。

至於科學的分類，各個科學家都有他們不同的獨到分法，例如孔德（Comte）將科學分爲化學、生物學、社會學、天文學、物理學、數學等六類，斯賓塞則將之分爲抽象與具體科學。大致來說，應分爲下述三類較適當，第一類是研討物質的無機現象的物質科學，也有人將之稱爲自然科學，包括物理學、化學、天文學、冶金學、工程學、建築學、地質學等均屬之；第二類是研究有機的生物現象的生物科學，此外，也有些學者將物質科學與生物科學併稱爲自然科學，有人口學、生物學、動物學、植物學、心理學、生理學、醫學、育種學、優生學等。第三類是研討超機的社會現象的社會科學，如社會學、政治學、經濟學、法理學、倫理學、商學、教育學等。

任何一類的科學都是需要用到統計方法，才能獲得越發精密且有價值的結果；社會科學的日精月益，絕大多是應用統計方法而得，社會現

象是雜亂無章的，必須藉著統計科學來整理，含混不清的事態也要靠著統計來分析，錯綜複雜的因果關係的問題也須利用統計來解答，奧秘難測的趨向亦是憑藉著統計來推斷，這種科學的利用價值眞可謂是博大精深了。

有句話說「數字是不會撒謊的」（Figures won't lie），可堪稱眞理。

以簡單的數字表示極其繁複的狀況，是統計法的優點。

統計的資料，必定是數量的資料。統計的結果，也必定是以數字來表示。統計上所有的數量，是由點計或者測度而得，不容易達到絕對的準確，僅是準確到足以取信之程度罷了。實際上，凡是經由測度而來的數量，姑且不管它是自然現象、生物現象或者是社會現象，要達到絕對的準確根本是不可能的，例如人體的脈搏每分鐘跳動七十二次、體溫多少、高低血壓各是多少等都是。點計所得到的數字或許眞能夠達到百分之百的絕對準確，但自點計數字演化而來的其他數字，如男女性別的比例、出生及死亡的比率等，都只是近似值，不可能成爲絕對正確值。如此，便產生了誤差。

通常統計資料的整理，保留多少位數方爲合宜，必須詳加研究。統計數值的計算，如果保留數字過少，會影響準確性，過多，又會加增計算上的繁複與混淆不清。一般均考慮到二方面來決定準確的單位，一：素材本身的特性，準確性的容許程度爲何？二：我們所需要的準確度爲何？例如平常算計長度，道路就以公里或市里計算，布料以尺或寸計算，身高論公分。至於重量、體重以公斤計算，黃金以兩或錢或分計算……等。基於事實的需要，不同的情況便要保留不同的數值，所以經常會刪除或者簡化尾數。通常是用四捨五入法，但是這種簡化法，入的機率爲$\frac{5}{9}$，捨的機會爲$\frac{4}{9}$，似乎缺乏精準，因此，便有人提議，比五大的入，比五小的捨，剛好爲五的就看前一位數爲奇數時加一，前一位數爲偶

數時就捨去不要了，使其入捨機會均等。

誤差

①可能誤差與相對誤差

統計數值最末一位數的半個單位即爲可能誤差，譬如說一個人的身高是178公分，則可能誤差就是0.5公分，體重50.5公斤，那麼，可能的誤差就是0.05公斤。相對誤差爲可能誤差與近似值（統計數值）之比，以公式表示：

$$\text{相對誤差}=\frac{\text{可能誤差}}{\text{統計數值}}\times\frac{100}{100}$$

例如上述之身高的相對誤差$=\frac{0.5}{178}\times\frac{100}{100}=0.28\%$

$$\text{體重的相對誤差}=\frac{0.05}{50.5}\times\frac{100}{100}=0.1\%$$

誤差可能有單位，這個單位與原來統計數值的單位一樣，各統計數值誤差的大小，較不適宜用可能誤差來比較，以相對誤差比較較適當，原因是，相對誤差沒有單位，誤差的程度可明白的顯示出來。誤差的另一面則爲確度，所以相對誤差小，其準確度較高，相對誤差大，準確度便要低了，由此可得一結論，即爲，一數值的相對誤差和準確度之間的關係是互爲相反的。

②絕對誤差與實際誤差

所謂的絕對誤差就是，近似值與眞值之差，而實際誤差就是絕對誤差和眞值的比了。因爲眞值不容易求得，所以絕對誤差和實際誤差也就不好算出，通常即以可能誤差和相對誤差來替代。

所謂的有效數字，是指用來表示一數值的準確程度的數字。

通常計算有效數字的方法，是由數字的左邊數到右邊，而不是由零的第一位數字向左算到最後一位準確數字爲止，通常單純整數最末的零

不能算爲有效數字，而小數末尾的零就必須算是有效數字了。

相對誤差的大小與有效數字位數的多少之間的關係是成相反的關係，有效數字位數多者相對誤差小，反過來說，有效數字位數少的話，相對誤差便大；有效數字位數若是相同，那麼數值中數字小的相對誤差大，數值中數字大的相對誤差就小。

此外，對於有效數字位數的計算方法，統計數值若是沒有整數，則小數點後面的零就不算作是所謂的有效數字了。

近似值的計算法

①近似值差數的兩個可能誤差，等於兩近似值的可能誤差之和。因爲減法是加法的還原，所以減法準確度的計算和加法的計算並無軒輊，也是以有效數字位數最小的一個準確度而決定差數的準確度。近似值的和與差的算法也就是近似值的和數和差數的準確度的決定法。胡元璋在他所著的「統計計算的確度」一書中引述了金氏（W. I. King）的一段話：「一鍊的強度決定於最後的一環，一個總和的確度亦由其確度最低之一項決定之。」實在是精簡明瞭。

②幾個近似值和數的可能誤差，等於各近似值的可能誤差之和。

設有近似值x_1，x_2，x_3……x_n其可能誤差分別爲$\pm e_1$，$\pm e_2$，$\pm e_3$，……$\pm e_n$，則x_1之眞值應爲$x_1 \pm e_1$，……x_n的眞值應爲$x_n \pm e_n$。設x_1，x_2，……x_n之總合爲x，x的可能誤差爲E，則可以得到下列二式：

$$x = x_1 + x_2 + \cdots\cdots + x_n \qquad ⓐ$$

$$x \pm E = (x_1 \pm e_1) + (x_2 \pm e_2) \cdots\cdots + (x_n \pm e_n) \qquad ⓑ$$

ⓑ式－ⓐ式 $\pm E = \pm e_1 \pm e_2 \cdots\cdots \pm e_n$ 此明顯的表示總和的可能誤差爲原來各數值的可能誤差之和。

計算一準確度不同之近似值的和數時，應先把各數值化簡到相同的

單位時再相加，其和數的準確度以原來各數值中的有效數字位數最少的近似值的單位爲準。

商的準確度

乘法的還原是除法，所以除法計算的確度和乘法的是一樣的，除法的相對誤差，大抵等於被除數與除數之相對誤差的和，而商數的有效數字的位數，最多等於（一般情況是少於）兩近似值中有效數字的位數最少者。倘若相除的二個數，一爲精確值，一爲近似值，這個商數的有效數字位數，一般都是以近似值的有效數字位數爲準，如剛好可除盡，商數的有效數字位數少於其中近似值有效數字位數也可。

積的準確度

①積數的相對誤差，大抵等於相乘的二個數之相對誤差之和。

設 x_1 ，x_2 二數相乘，其相對誤差 e_1' 及 e_2'

則 x_1 之眞值爲 $x_1(1\pm e_1')$

x_2 之眞值爲 $x_2(1\pm e_2')$

二者相乘後之最大積數爲 $x_1(1+e_1')x_2(1+e_2')$，最小積數爲 $x_2(1-e_1')x_2(1-e_2')$。

其積數的相對誤差爲：

$$\frac{x_1(1+e_1')x_2(1+e_2')-x_1x_2}{x_1x_2}$$

$$=\frac{x_1x_2[(1+e_1')(1+e_2')-1]}{x_1x_2}$$

$$=(1+e_1')(1+e_2')-1$$

$$=1+e_1'+e_2'+e_1'e_2'-1$$

$$=e_1'+e_2'+e_1'e_2'$$

或爲

$$\frac{x_1x_2 - x_1(1-e_1')x_2(1-e_2')}{x_1\ x_2}$$

$$= \frac{x_1x_2\,[\,1-(1-e_1')(1-e_2')\,]}{x_1\ x_2}$$

$$= 1-(1-e_1')(1-e_2')$$
$$= 1-1+e_1'+e_2'-e_1'\ e_2'$$
$$= e_1'+e_2'-e_1'\ e_2'$$

以上左右二式中之 $\pm e_1'\ e_2'$ 極爲微小，爲了方便計算，故以省略，如此左右二式均可改爲：$e_1'+e_2'$，即積數之相對誤差爲二數之相對誤差之和。故乘積的眞值爲：$x_1x_2\,[\,1\pm(e_1'+e_2')\,]$。

影響乘法最大的是準確度最低的一項，因此，積數的有效數字位數最多等於（或少於）兩近似值中有效數字之位數最少者。如果相乘的二個數，一爲精確數，一爲近似值，它的積數之有效數字位數，一般是以近似值的有效數字位數爲準。

②積數的可能誤差，大抵等於乘數及被乘數和其可能誤差二者相乘之積的和。

設 x_1，x_2 二數相乘，其可能誤差爲 e_1'，e_2'，則積數的最大可能爲：

$$(x_1+e_1)(x_2+e_2)$$
$$= x_1x_2+e_1e_2+(x_1e_2+x_2e_1)$$

積數的最小可能爲：

$$(x_1-e_1)(x_2-e_2)$$
$$= x_1x_2+e_1e_2-(x_1e_2+x_2e_1)$$

故 $(x_1\pm e_1)(x_2\pm e_2)$ 之積必爲：

$$x_1x_2+e_1e_2\pm(x_1e_2+x_2e_1)$$

因 e_1e_2 之值甚小，爲了方便計算，大多省略不計，故 $(x_1\pm e_1)$

$(x_2 \pm e_2)$之積可改寫爲：

$$x_1 x_2 \pm (x_1 e_2 + x_2 e_1)$$

計算積數可能誤差的簡單方法可以下式代入：

$$(x_1 \pm e_1)(x_2 \pm e_2)$$
$$= x_1 x_2 \pm (x_1 e_2 + x_2 e_1)$$

偏誤與簡化尾數

上述之近似值的誤差，有正與負二個方向，設有近似值若干個，不一定誤差會相同，往往會有偏高或偏低者，計算後，可能正負號會互相抵銷，並非一定能達到誤差的極限，這叫作「非偏誤」（Unbiased error）或者是「補償誤差」（Compensative error）。倘若各數值都爲同號的誤差，則叫做「偏誤」（Biased error），或「累積誤差」（Cummulative error）。有一點須加以注意，原來的數值爲非偏誤時，和數與積數的誤差或許會減少，而商數與差數的誤差就不一定了。如果被除數及被減數偏高，而除數及減數偏低的話，商數和差數的誤差增大偏高，反之則增大偏低。再者，原來數質爲相同方向偏誤時，和數和積數增大偏誤，商數與差數便有相互抵銷而減輕偏觀的作用。這在四捨五入尾數的簡化也是一樣的，通常四捨五入後的數值，如果有偏誤，則會若干項偏高，若干項偏低，相乘或相加的誤差可降到最少。不過，在除法與減法中，若四捨五入後，一邊偏高一邊偏低的話，商數和差數的誤差就會增大。倘四捨五入後有相同方向偏誤，則商數和差數不變，但和數及積數便會增大差誤。所以，在四捨五入簡化尾數時，應該謹愼才好。

二、統計研究問題成級資料最爲簡便經濟

爲了解所有實際狀況，必須依據調查，如果將所要研討的全部實際

狀況詳加調查，以所有的資料作個分析與探討，則必爲最實在且可靠的資料。這種以某一特定情形進行總體個別調查的方法稱爲普查（ Census ），例如人口普查便是。可是，由於天地之間，無論何種現象，都是十分繁複，而且散佈又廣，往往會受人力、時間、經濟等各因素的限制，不可能作全面的調查，因此，除了特別重要的資料，很少應用到普查。即使是一國最重要的人口資料，也是要隔許多年方能舉行一次普查，並不是經常性的。

一般蒐集靜態原始資料，大部份是採用抽樣調查（ Sampling Survey ），意思就是，在全部的事物中抽取一部分，將所抽查的樣本（ Sample ）作爲全部事物的依據，以樣本特性來闡明總體的特性，如果抽樣得宜，則其正確度可達很高，如此，誤差便會降至最低程度。

抽樣的方法大致可區分如下：

①部落抽樣（ Cluster random sampling ）

此法與分類比例近似，都是先把待研究的事物依某項規則分爲若干類，每類就叫一部落，其次，自所有的部落中，隨便選幾個部落來進行調查，這種方法就叫作部落抽樣法。不過這要是各個部落間相差不懸殊，並且部落內部的分子差異大時比較準確。假使各個部落間差異懸殊，而部落之內部大致相仿，則這個部落抽樣就不適合使用了。

②隨意抽樣（ Random Sampling ）

另一名稱是「機會抽樣」，一個群體中的每一分子都有相同被列爲抽樣對象的機會，且不參雜研究人員的任何主觀意見。這種抽樣方式，是自所有全部的資料中挑選，而不是只在一部分資料中抽選，而且，所抽取的樣本範圍非常廣，通常是代表優質的。一般使用抽樣工具決定調查樣本。

③兩段抽樣（ Two stage sampling ）

首先，把待研究的事物分爲幾個組，用隨機法抽取其中幾組，其