# 淺談統計

魏慶榮

統計是門有系統地研究數據的學科, 大致可以分成三個部分。一是數據的收集, 二是數據的整理, 三是數據的分析。以下, 我們用一些例子, 來說明統計的這三個課題。

## 一. 數據的收集

首先我們談數據的收集, 表面上看來這是很簡單的事情, 只要到圖書館翻翻文獻就能了結的事。這種已經擺在那邊, 我們無法插手的數據, 統計裏叫它做「觀察」得來的資料。這種資料由於不知道它的來源, 有時不太可靠。另外有兩種搜集資料的方法, 那就是「抽樣」和「實驗」。

像民意調查就是抽樣調查的一種。本來抽樣的目的, 是想設計一套方法, 在一群很大的母體 (如選民), 抽出一小部分具有代表性的樣本來, 然後就他們的意見來做分析。爲了具代表性, 不偏那個部分 (如特定族群), 有時還靠丟銅幣的方式 (隨機) 來抽樣。可是報紙上報導的民調, 常常是各個黨派用來造勢而不在反映眞實狀況, 因此選前三黨都號稱民調對他們有利, 造成選民認爲統計是騙人的戲法。還有一些人誤以爲到街頭隨便問幾個人, 就能代表民意, 這跟認爲 Call in 就能代表一般人的意見一樣是不對的, 因爲這種樣本是取自於一特殊的人群 (上街的, 或想 Call in 的), 常常無法代表整個母體。有些廣告也和這個相似, 譬如說,「三個醫生當中有兩個推薦 ×× 藥」。這句話暗示我們, 所有醫生當中, 三分之二的醫生推薦該藥。事實上, 我們應該問一問藥廠, 是不是只問過三個醫生的意見? 而且那三個醫生和藥廠的關係又如何? 譬如說, 其中推薦的那兩位是不是藥廠老闆的親友? 如果是, 這種樣本根本就是數目太少又有偏, 教人如何能相信?

其次, 我們來談實驗。實驗設計是教我們有系統地找「配方」的方法。先舉一個例子。我們常常聽人家說, 有人生病了, 到某個廟裏抽藥籤, 結果吃好了, 因此便說這神明很靈。可是仔細想一想, 我們會發現吃不好的人, 都怪自已沒福氣 (誰敢怪神明?), 因此在報喜不報憂的情況下, 聽到的就只有好的一面了 (有偏樣本)。那麼我們如何才能確定神明是否靈驗呢? 有一個方法便是安排個實驗, 同時找兩組病人, 病症相似, 一組給神明的藥

吃, 另一組給白開水喝, 再比較結果, 便能明白神明究竟靈不靈了。近代藥品的臨床實驗, 就採取類似的設計來檢定新藥有效與否。在工業製造裏, 爲求產量高、品質好, 對進料的成分, 溫度的高低等等, 則常利用一些比較複雜的實驗設計來安排。

## 二. 數據的整理

現在, 我們來講數據的整理。通常數據有時非常雜亂, 有時非常龐大, 我們不可能從這些原始資料直接看出所要的訊息。譬如說, 大專聯考所有考生的成績, 由於多到只能放在磁碟片裏, 如果我們想問, 今年的數學成績是不是比去年考得好, 即使有辦法看遍原始成績, 也很難回答這個問題。這時我們先得把資料重組、分類, 用幾個簡單的量 (叫描述統計量), 或畫個圖形, 清晰而有效地把訊息傳達出來。常常看見的百分比圖或直方圖都屬於資料整理的部份。至於做摘要的統計量, 一般用來描述數據的中心值或最具代表性值, 常見的有平均值、中值和衆數 (最常出現的那個值)。其次, 用來描述這組數據離中心值發散的程度, 如級距 (最大值和最小值的差)、標準差、百分位數 $\cdots$ 等。這些量有什麼用呢? 大家可能都用過中文電腦吧, 其中字音輸入法很大的困擾是同音字太多, 電腦會把所有的同音字列出來教我們選。可是那個字要列在最前面呢? 當然是最常用的字!可是那個字又是最常用的呢? 我們當然不可能把所有的書、雜誌、信件拿出來檢查一下, 只有選擇一些樣本來統計一下各個字的使用次數 (頻數)。譬如說, 有人拿小學六年的課本, 有人拿一年的報紙來分析。分析出來的衆數, 就是排在最前面的字。而輸入法的好壞, 就決定於這些樣本是不是能代表平常使用的習慣。最近有些字音輸入法, 不需要一次選一個字, 電腦會自動選字, 打完一句還會自動修改, 這些功能都是利用比較複雜的統計方法做出來的。

再舉一個例子, 假設你是成衣廠的老板, 如何決定成衣大小型號的件數呢? 某種型號, 做太少搶不到市場的先機, 做太多又怕賣不出去, 怎麼辨呢? 這時要是有個成衣型號調查的頻率圖那就好, 我們就會明白特大號、大號 $\cdots$、小號等等佔人口的比率有多少, 就能幫我們做決定。

在選取描述統計量時, 有時要很小心。底下是個有關大學入學是否歧視女性的例子。實際的例子發生在美國柏克萊大學, 他們沒有大專聯考, 憑介紹信申請的。我們把數據簡化, 並假設全校只有兩個系:

| | 男 | | 女 | |
|---|---|---|---|---|
| | 接受 | 拒絕 | 接受 | 拒絕 |
| 電機系 | 30 | 30 | 10 | 10 |
| 英文系 | 5 | 15 | 10 | 30 |
| 全 校 | 35 | 45 | 20 | 40 |

女權運動的人說學校有性歧視, 因爲男生入學率是 $35/80=0.4375$, 大於女生的入學率 $20/60=0.3333\cdots$。學校緊張了, 就去問各個系。每個系都說, 我們絕對男女平等。查一下的結果, 電機系男、女生入學率都是 $1/2$, 英文系男、女生入學率都是 $1/4$。這怎麼可能呢? 那裡出了錯誤呢? 我們看看電

機系總共收了40個學生, 遠比英文系的15來得大, 而大部分的的女生,60個中的40個, 卻又申請英文系, (相對的, 男生80個中, 只有20個申請英文系), 雖然拒絕率相同, 總個數卻相當可觀, 因此造成各個系沒歧視而整體看來貌似歧視的現象。換句話說, 在系的層次, 以男女入學率做爲描述統計量是合適的, 在學校的層次就會產生問題了。一個比較合理的全校性統計量要考慮到各個系佔的比重, 所以用

男 (女) 生入甲系的比率×申請甲系的學生比率 +男 (女) 生入乙系的比率×申請乙系的學生比率

比較能反映出全校的男女生個別入學率。上面的例子, 依照這個方法算, 男女生入學率都是 $1/2$, 相當合理。

## 三. 機率論

最後, 我們來談談數據的分析。這部份的統計叫做推論統計, 是要從計算出來的統計量做出結論。像剛剛提到成衣廠的例子, 有了頻率圖後, 就要決定各個型號的製造個數。由於統計量是由部份樣本算出來的, 和由母體算出來的量總難免有誤差, 誤差的大小也會隨樣本的不同而變化。機率論就是來描述這些誤差大小的工具。

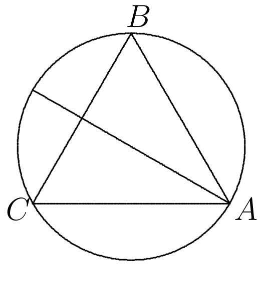
機率論是個很有趣的題材, 常有出人意料之外的結果。假設有一天我們走在街上, 突然被邀去檢查, 看看有沒有感染到某種危險疾病 (如愛滋病), 而結果是陽性反應。碰到這種情況, 我們不免會怨天尤人, 說這種疾病非常少, 總人口的一小部份 (假設是$10^{-6}$) 才有, 爲何偏偏就落在我的身上。接著可能會有人來安慰, 說這個檢查不是百分之百準的, 沒有病而測出陽性反應是有的, 而且大到十分之一。可是我們還是憂心忡忡, 直覺地認爲有百分之九十的機會還是會得到。眞的是 $9/10$ 嗎? 其實, 檢查還有一種誤差, 也就是有病卻得到陰性反應。如果這個機率是 $1/100$ 的話, 那麼我們眞 正得病的機率應該是

$$
\begin{aligned}
& P(\text{得病} \mid \text{陽性反應}) \\
=\ & 10^{-6} \cdot (1-0.01)/[10^{-6} \\
& \times 0.99 + (1-10^{-6}) \times 0.1] \\
<\ & 10^{-6}/0.1 \\
=\ & 10^{-5}。
\end{aligned}
$$

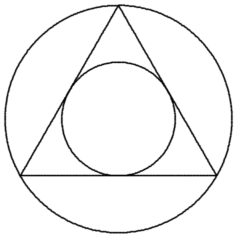
上式表明的是: 有了陽性反應後, 得病的機率小於十萬分之一。這個數字相當小, 因此除非有其它症狀, 否則不必那麼緊張。

和機率相關的一個常用語是「隨機」。如果我們從 1 到 10 隨機抽出一個數字來, 通常我們認爲任一特定數字被抽中的機會是 $1/10$。我們事實上“私下”假設了機率的分布是均匀的。這種把「隨機」和等機率等同起來的做法 , 在可能値, 如 1 到 10, 是有限個時還好。可是在可能値是無限、連續時, 這種直覺就會產生麻煩了。想想看, 給定一個單位圓,「隨機」選取一條弦, 那麼這條弦的長度大於內接正三角形的邊長的機率有多大?

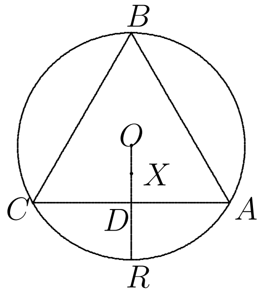
第一個答案: 由於弦的一個端點落在那裡都一樣, 我們把它固定在 $A$, 由 $A$ 點做一個內接正三角形 $ABC$。由於另一端點只有落在弧 $BC$ 上, 弦長才會大於邊長, 因此機率是 $BC$ 弧長/圓周長 $= 1/3$。




第二個答案: 由於每一條弦都由其中點唯一決定, 所以隨機取一條弦等於隨機取圓內一個點。若點落在半徑爲$1/2$ 的同心圓內, 則弦長大於內接正三角形的邊長, 因此機率是同心小圓面積/圓面積 $= 1/4$。



第三個答案: 跟上面一樣只考慮弦的中點 $X$, 由圓心經過 $X$ 畫一條半徑 $OR$, 其中點爲 $D$。經 $D$ 點畫一垂直於 $OR$ 的弦 $AC$, 由 $AC$ 可造一內接三角形 $ABC$。不難看出: $X$ 只有落在 $OD$ 上, 弦長才會大於 $AC$ 長, 因此機率是 $OD$ 長/ $OR$ 長 $= 1/2$。




同一個問題, 居然有三個答案, 什麼地方出了問題? 在第一個答案中, 隨機取弦變成隨機在圓周上取一點, 而在第二個答案變成隨機在圓內取一點, 在第三個答案卻又變成隨機在半徑上取一點。每個說法都滿足直覺, 可是答案卻都不同。「隨機」就是等機率, 這個直覺行不通了。因此我們談機率時, 一定要確定樣本空間 (是圓周? 是圓內? 是弦? ) 和機率 (是否等機率? )。

有了樣本空間和機率函數, 現在我們能描述數據發生的機率了。數據 $x$ 的可能値如果是有限或可數個 (離散型), 我們用機率密度函數 $f(x) = P[X = x]$ 來描述。如果是連續型的, 我們用密度函數$f(x)$在 (a,b) 的積分,

$$P[a < x < b] = \int_a^b f(x)dx$$

來表達 $x$ 落在 $a, b$ 之間的機率。最有名的密度函數就是鐘形的正態密度函數

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

其中$\mu$ 代表中心點 (平均値),$\sigma$ 是標準差, 可以用來量數據離中心散開的程度。正態密度在許多領域經常出現。物理學家說數學家證明了在正常狀況下, 密度是正態的, 而數學家卻說, 物理學家發現了自然界是正態的。當然這都只是部分事實, 一般人認爲在穩定的狀況下, 數據 (如產品的規格) 可以當做是正態的。心理界與教育界認爲上智與下愚人很少, 大部份人在中間, 因此構成正態分布 (常態分班的由來)。底下, 我們舉一個實例來看看正態分布在統計推論上的應用。

## 四. 數據的分析 — 推論統計

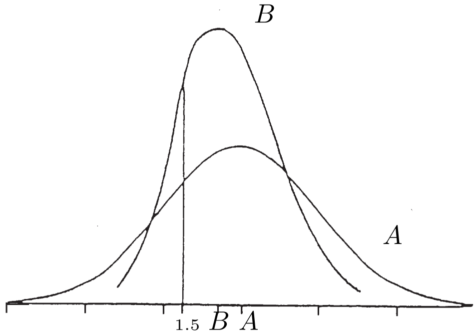
這個實例與衛星接收器相關。台灣做的衛星接收器曾在中東戰爭出了名。爲了便於說明, 我們把故事稍微加油添醋。有個美國廠商, 想要買台灣工廠製造的衛星接收器。他要求兩家工廠各做三個樣本, 從三個樣本他要判斷那個工廠的產品不良率合乎標準, 比如說, 那家工廠產品不良率小於 20%, 就選那個。現在假設對某個規格的要求是大於 1.5 (工業界叫做望大需求),$A$、$B$ 兩廠做出來的結果如下:

| 廠商 / 數據 | $A$ | $B$ |
|---|---|---|
| 1 | 1.51 | 1.51 |
| 2 | 1.53 | 1.53 |
| 3 | 1.7 | 1.48 |

很顯然,$A$ 廠三個都合乎規格, 而 B 廠卻有一個接收器, 其規格小於要求。猜猜看, 美商選了那個工廠? $B$ 廠! A 廠很不服氣, 要求說明。美商就傳眞一份計算如下:

| 廠商 / 統計 | $A$ | $B$ |
|---|---|---|
| 平均值 | 1.58 | 1.506 |
| 標準差 | 0.0109 | 0.000048 |
| 不良率 | 0.2524 | 0.1928 |

傳眞中並說明了, 不良率是根據正態分布計算的。A 廠還是一頭霧水。事實上, 我們用白話來說是這樣的: 美商假設工廠的製程是穩定的, 產品規格的大小依正態密度來分佈。$A$ 廠商雖然都合乎規格, 但是產品值的大小發散程度 (標準差) 太大, 如果正態分布是對的話, 將來出現不合規格的產品機會就比 B 廠大。$B$ 廠雖然有個不合規格, 但是其大小變化很小, 依正態分布計算, 不良率小於 20%, 所以可取。用圖來說明




$A$ 密度較肥胖,B 密度較陡瘦, 因此 1.5 左邊的面積 B 比 A 小。

這個例子似乎違反常理, 當然還有很多可討論的地方。比如說, 用三個樣本來估計平均值和標準差是不是太少了? A 廠是不是逐步在改善他的製程, 因此愈做愈好, 正態分布是不是還適合呢? 這些都是很有趣的問題, 特別是要選多少樣本才合乎道理, 正是推論統計裡一個重要的課題。

## 五. 結論

在這篇文章裡, 我們簡單地介紹了統計的內容, 並且用幾何的例子來說明統計的數學工具 – 機率論, 也用了民意調查、藥物臨床實驗、電腦輸入法、工廠製造、性別歧視和病症檢定等做例子, 來說明統計應用的廣泛。事

實上, 做爲方法論的統計, 在人文社會科學、生命科學、自然科學都扮演著重要的角色, 歡迎大家走進這個領域共同努力。

註: 本文將同時發表於通識教育季刊。

## 參考資料


(1) 比較通識性的入門書, 可以參考
Statistics, Concepts and Controversies 2nd ed., D. S. Moore, 1985, Freeman and Company, New York.

(2) 有關抽樣調查的入門書, 可以參考
抽樣調查, 林進田,1993, 華泰書局。
抽樣方法之應用, 韋端,1990, 中國統計學社。

(3) 有關實驗設計的入門書, 可以參考
Statistics for Experimenters. G.E.P. Box, W.G. Hunter and J.S. Hunter 1978, Wiley, New York.

(4) 有關統計的誤用, 可以參考
跨越數字陷阱, 董時叡,1991, 遠流。

(5) 第三節裡舉的隨機選取弦的例子, 叫做 Bertrand's paradox, 比較詳細的說明及其它相關例子可以參考
Paradoxes in Probability Theory and Mathematical Statistics, C.J. Szekely, 1986 Akademiai Kiado, Budapest.

(6) 有關常態分班與自學方案的討論, 可以參看
論國中自學案班級常模五分制之公平性, 林妙香, 國科會研究彙刊, 人文及社會科學, 民83,7月,4卷,2期,246-263頁。

(7) 統計在生物、政治、社會和物理世界的許多實例, 可以參看
Statistics: A Guide to the Unknown. 3rd ed. J. M. Tanur et al., 1989, Wadsworth, California.

(8) 統計在品管方面的應用以及 Deming 之貢獻可參考
Dr. Deming, The American Who Taught the Japanese About Quality, R. Aguayo, 1990, Carol.
台灣有翻譯本: 品管大師戴明博士, 汪益譯,1991, 聯經。


—本文作者任職於中央研究院統計科學研究所—