——統計專題四——

談統計應用

鄧進財

唸純數的常挖苦唸統計的說：「你們唸統計的常以算術平均數來代表全體（population），那麼你們一手泡在沸水中，另一手浸在冰水中，一定會感到很舒服，因爲你們的平均感受只有 50°C 而已。」而唸統計的也不甘示弱的說：「你們唸純數的說 $a=b$，$b=c$，則 $a=c$，那麼你們一定會熱愛女朋友的另一個愛人，因爲你們愛你們的女朋友，而你們的女朋友愛另一位男朋友，所以你們也會愛你們的情敵。」當然，這都是「自古文人相輕」「外行批評內行」的寫照。事實上，唸過統計的人都知道全體十分集中時，以算術平均數代表全體始有意義；而學過數學的人也都明瞭「等號」是必須具備有傳遞性的（此即，若 $a=b$，$b=c$ 則 $a=c$），可是，上述所說的「愛」並不滿足傳遞性。

在面對不確定情況下，統計是一種能夠幫助我們做出聰明決策的科學方法。下面，讓我們來談談統計應用的一些實例。

例題1. 一計程車問題

國外某地的計程車較少，一位統計學者在該地的某街角等候計程車，眼看來了幾部計程車都載客而過，這位統計學者開始懷疑這個城市到底有幾部計程車，以致於不夠應用。於是他開始記下載客而過的計程車車號，依次如下：

405, 280, 73, 440, 179

接著來了一部空車，載走了統計學者。

假如該城市計程車的編號是從 1 號開始連續編排下來，而且空的計程車是在城裏做隨機性的環繞，那麼，你若是這位統計學者，你將如何從上述記錄的資料來推測該城市共有幾部計程車？

這個問題的估計方法很多，在此我們將簡介兩種簡單的估計方法，並加以粗略的檢定。

一、平均差距法

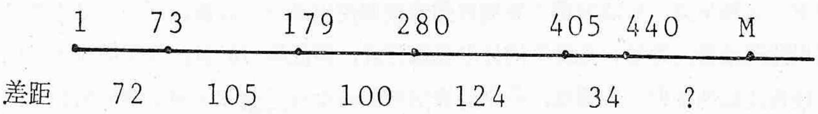

圖 1 平均差距法

假設該城市計程車的編號是連續的，而第一部的編號爲 1，最後一部設爲 M。現在，如果我們能夠猜測第 440 部計程車與第 M 部之間的差距，那麼我們就可以正確的推測 M 的數值。從直覺上，讀者可由圖 M 得知：我們可引用前五部車號的平均差距，來代表第 440 部與最後一部之間的差距，因此我們可以推測

$$M=440+\frac{1}{5}(72+105+100+124+34)$$

$$=440+\frac{1}{5}(440-5)$$
$$=440+87=527$$

統計學者上車後，詢問司機這個城內究竟有幾部計程車，結果司機回答說：城裏共有 550 部計程車。則根據上述的平均差距法所做的估計，其相小誤差僅為 $(550-527)/550=0.04$，因此，上述的估計方法是十分的理想。

二、中位數法

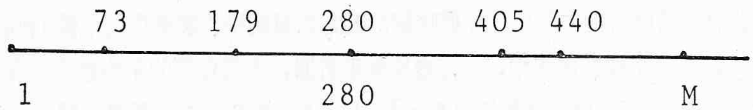

圖 2 中位數法

圖 2 顯示those五部計程車車號的中位數為 280，根據抽樣的特性，亦即樣本足以代表整個全體的特性，我們可以合理的推測 280 也可能是城裏所有計程車車號的中位數，則

$$(M+1)/2=280$$

亦即 $M=559$，此與實際數值 550 部的相對誤差為 $(559\text{-}550)/500=0.02$，因此採用中位數的估計方法，也是十分的美好。

事實上，我們上面所介紹的方法是**數列分析** (serial number analysis) 中最簡單而理想化的解法，因為實際上計程車的編號並不是從 1 號連續下來的，而且計程車所環繞的路徑，通常並不是隨機性的亂兜圈子。數列分析在第二次世界大戰時，盟軍曾用以分析、估計戰場上德軍戰車的數量，戰後發現利用這種統計方法估計德軍戰車的數量相當精確。

例題2. 估計野生動物的存量

在報章雜誌上，我們經常看到某某動物已瀕臨絕種邊緣的報導，你們一定會奇怪這些科學家如何知道那些野生動物的數量呢？例如；鯨魚漫遊在佔全球表面積約 2/3 強的海洋內，而且大部分的時間又都潛伏在水底，我們如何來估計各類鯨魚的存量呢？下面讓我們來介紹二種不同的估計方法，分別說明如下；

一、記號法

石缸內有一大堆彈珠，假設我們不能夠將彈珠整個倒出來一一計數，而又想知道整缸內彈珠的個數，我們可以利用隨機抽樣的理論，先從缸內任取幾個彈珠，假設取 10 個，作記號後再放回缸內，加以完全的攪亂，然後再從缸內抽取一些彈珠，分別計算這些彈珠中有記號和沒有記號的個數比，假設計算結果是沒有記號的彈珠為有記號彈珠的四倍，那麼應用隨機抽樣：樣本足以代表全體的特性，我們可以推測整個有記號與沒有記號的彈珠比為 1：4 因此可推知共有 40 個沒有記號的彈珠，也就是說，總共有 50 個彈珠在石缸內。

利用相同的方法，我們可先捕捉一些鯨魚，在厚脂的鯨背上烙入 0.3 公尺的圓柱形金屬，再放回海中任其游動。其後我們可由捕獲的鯨魚中，計算有記號與沒有記號的比率，即可推測出鯨魚的總數。

這種簡單的設計，雖然在執行上有很多困難，但仍然被用來估計很多野生動物的總數。可是，利用這種方法來推測鯨魚存量，在執行上尚有很多因素必須加以化慮，譬如；將金屬烙入鯨背，是否就不會因跳

躍而脫落？工人在寒冷的暴風雨中開剖像房子那麼大的鯨魚，是否會注意到那個小小的柱形金屬？而且鯨魚不像彈珠，牠是有一定的壽命（若不被捕殺，一般鯨魚能活到四十多歲）。爲了解決這些複雜的問題，我們需要連續好幾年的實驗和比較過程，把前述的方法加以更精密的推展。例如：我們連續兩年捕捉相同數目的鯨魚，作完記號後放回，則兩年後我們從被捉的有記號鯨魚中，計算第1年與第2年被作記號的比率，即可推算這段期間鯨魚被捕或自然死亡的比率。

二、捕捉法

海底鯨魚的存量多，則鯨魚被發現的次數就多，於是鯨魚被捕量亦增加，反之則否。因此我們可假設鯨魚被捕的機率是固定的；那麼我們即可由每年捕到鯨魚的數目，求得鯨魚存量的估計值。例如：這一期我們捉到 25,000 條鯨魚，比前一期少了10%，則我們可由此推知這一期共有 25,000/0.10=250,000 條鯨魚，其原因說明如下：

設前一期有x條鯨魚，且鯨魚被捕的機率爲p，則這一期應有（$x-xp$）條鯨魚（不考慮自然死亡與新生的鯨魚），再由已知條件，得下列聯立方程式：

$$(x-xp)p=25000 \qquad \text{（公式 1）}$$

$$xp\left(1-\frac{1}{10}\right)=25000 \qquad \text{（公式 2）}$$

由公式 1 和 2 得知，

$$1-p=1-\frac{1}{10}\Longrightarrow p=\frac{1}{10}$$

因此這一期原有的鯨魚數爲

$$(x-xp)=25000\times 10=250000 \text{ 條}$$

然而，事實上並非如此單純。例如：捕鯨船在海上各種不同的氣候下作業，捕獲的鯨量必然大有出入；又捕鯨公司不斷地採用新技術，改善效率；加以鯨魚不斷的自然死亡和新生幼鯨，我們豈可假定鯨魚被捕的機率永遠不變？爲了克服這些困難，每隔一段期間之後，我們必須將p根據實際情形，加以適當的調整。

例題3. 如何調查敏感性的問題

假如我們所調查的問題涉及個人名譽道德、私人秘密及其他利害相關的事項，被調查的人多半不會忠實回答，若勉強直接調查，也難獲得可靠的資料。例如：我們想瞭解商人的逃稅率，如果直接詢問他們所得到的答案，必然是否定的。諸如此類的問題，我們要如何設計調查方式，以獲取被調查者的坦誠合作，提高調查效果？在此我們舉個例子來談談如何利用一種**「隨檢問答」**（randomized response）的技巧，以消除被調查者的疑慮；爭取精誠合作。

如果我們要調查商人的逃稅率，我們可設計兩個問題：**問題 1**，爲敏感性的問題──「你是否曾經逃稅？」，**問題 2**，爲無關的問題──「你的身分證號碼是否爲奇數？」

對每一個被調查的商人，我們可請他自行在放有 7 個紅球，3 個白球的袋中任意選取一球，然後按照下列規則回答問題：

若選到白球，則回答問題 2，

若選到紅球，則回答問題 1。

注意：被調查者只須回答「是」或「否」，而不須告訴我們選到何種球、所回答的是何種問題，如此我們即無法確知對方所回答的是何種問題，因此被調查者可以毫無顧忌的回答問題。

這種調查方式，雖然我們並不知道被調查者個別回答的問題是何種問題，但是在全體上，我們利用簡單的機率理論，即可輕易地求得所要調查的問題，茲說明如下：

設 P(X) 表示陳述X成立的機率，P(X|Y) 表示陳述Y成立時陳述X成立的條件機率，則

$$P\text{（回答「是」）}=P\text{（選到問題 1）}\times P\text{（回答「是」選到問題 1）}$$

$$+P（選到問題2）\times P（回答「是」選到問題2）$$

或者寫得更簡潔點：

$$\lambda=p\pi+(1-p)\theta \qquad （公式3）$$

式中 λ 表示回答「是」的機率， 表示商人中曾逃稅過的比率，p 表示選到敏感性問題的機率，而 θ 表示商人之身分證號碼爲奇數的機率。假設調查統計的結果，得到回答「是」的比率爲0.44，也就是 $\lambda\approx0.44$ 則將 $p=0.70$，及 $\theta=0.50$ 代入公式3中，得

$$0.44=0.70\times\pi+(1-0.70)\times0.50\Longrightarrow\pi=0.41$$

因此我們可估計商人中曾逃稅的比率爲 41%。

上面我們所用的方法，若碰到一位多疑，而身分證號碼又不是奇數的商人，儘管事實上我們不知道他的身分證號碼，而他依然很可能會害怕：假如他回答「是」，而我們若查出他的身分證號碼，那不就證明他曾逃稅嗎？我們爲了避免因此而產生資料的偏差，可以再應用下面的例子。

假如我們想研究贊成墮胎的比率，那麼可以將均一的紅色、白色及藍色玻璃珠放入盒中，已知各色玻璃珠出現的機率爲 p_r、p_w p_b。我們可設計如下的隨機問答，被調查者先從盒中隨機抽取一玻璃珠，然後按下列規則回答。

取到紅色玻璃珠：回答「你是否贊成墮胎？」，
取到白色玻璃珠：回答「是」，
取到藍色玻璃珠：回答「否」。

同前例，我們可以得到下列的數學模型

$$\lambda=p_r\times\pi+p_w\times1+p_b\times0 \qquad （公式4）$$

式中 λ 表示回答「是」的機率，π 表示贊成墮胎的機率。因爲公式4中的 p_r 與 p_w 爲已知，因此只要計算回答「是」的比率，卽可求得 π 的估計值。這種方式的隨機問題，可以完全消除被調查者的疑慮，也可利用 π $\lambda\geq p_w$ 的關係，檢定調查的資料是否可靠。

上面我們提到一些簡單的統計應用，而實際從事統計工作，那經常是一件十分繁雜的工作，而且稍微不小心，常會鬧笑話。例如：⑴當日光燈剛問世時，一些人深信，暴露在日光燈的幅射下將喪失生殖能力。某條鐵路已裝妥日光燈，爲了破除這個觀念，承辦了一次實驗，將老鼠分成兩組，一組暴露在白熱燈下，另一組生活在日光燈下，經過一段時間，在白熱燈下的老鼠已有了正常數目的子孫，但是另一組卻沒有一個後代，這個實驗反而更加令人相信，暴露在日光燈下將喪失生殖能力。後來經過一些懷疑者重新詳細的檢驗，卻意外地發現第二組的老鼠竟然都是同性的。

讓我們再舉個例子來談談把因果關係顚倒所造成的笑話。⑵新希伯利德的土著們相信虱子有益於身體健康，因爲根據他們數百年的觀察經驗，身體好的人都有虱子，只有生病的人身上沒有。於是他們由此推論：虱子使人健康，每個人都該有一些。這種推論顯然有問題，但是他們卻是確信不移。後來經過一些有經驗的人仔細研究之後，發現在新希伯利德地方差不多每個人都長虱子，可是當有人生病發燒的時候，因爲體溫太高，虱子們便喬遷他去，另覓住宅。這個例子告訴我們，如果把因（生病發燒）與果（身上沒虱子）顚倒，混淆一談，所得的結論將令人啼笑皆非。

這類的例子很多，又如：⑶曾經有人費了很多工夫找出在大學生中，吸煙學生的成績是要比不吸煙的差，於是許多人（尤其是不吸煙的家長）就很高興的推論：如果要成績好的話，大學生非戒煙不可。事實上，這個推論是以「抽煙」爲因，「成績差」爲果所導得的，但是這樣的假設是毫無根據的。我們同樣可以把因果關係顚倒過來，以「成績差」爲因，「抽煙」爲果，也就是說：學生因成績差，只好藉煙解愁。如此一來，就是把煙戒了也無法提高成績。

資 料 來 源

1. 科技發展小組，統計能爲你做些什麼，文理出版社發行
2. 夏沛然，統計魔術，科學月刊發行