

統計資料的來源

鄭惟厚教授／淡江大學數學系

我們為什麼需要知道統計資料的來源？

媒體上常有各式各樣和健康有關的報導，我們是不是都應該相信呢？

先來看兩個例子：

有一項最新研究報告指出，常喝咖啡有助於降低罹患成人型糖尿病的機率。如果我的家族有糖尿病史，好像我應該開始努力喝咖啡。但是幾個月之後的另一項報導卻說，咖啡因不利第二型糖尿病控制。

這兩項研究結果是不是有點互相矛盾呢？難道是在說：在還沒得糖尿病之前，我們應該大量喝咖啡以避免得病，但是一旦得病之後，就應該戒掉咖啡以利病情的控制嗎？如果兩項研究結果都可信，似乎只能做出這樣奇怪的結論。那我們似乎應該先確認一下，到底研究結果是否可信。首先讓我們仔細看一下報導內容。

第一項研究的報導內容還說：芬蘭赫爾辛基國家公共衛生研究所的研究人員對一萬六千名芬蘭人進行長期追蹤後發現，每天喝幾杯咖啡，有助於降低罹患成人型糖尿病的機率，如果每天喝十杯以上咖啡，罹患糖尿病的機率更會大幅降低。而第二項研究報導說：美國北卡羅來納州杜克大學醫學院研究人員發現，如果第二型糖尿病患者攝取大量咖啡因，體內葡萄糖和胰島素濃度就會相對提高。而糖尿病臨床治療的目標，就是在降低血液葡萄糖濃度。美國糖尿病學會表示，經診斷患有糖尿病的美國人大約一千七百萬，其中至少百分之九十是第二型糖尿病患者。也就是說對於絕大部分的糖尿病患者來說，似乎應該避免咖啡因的攝取。

兩項研究都是由有公信力的學術單位做的，而且都有數據支持，我們似乎沒有理由不相信；但是看來互相矛盾的結果，又不太可能同時正確，究竟應該如何判斷一項研究的可信程度呢？首先必須了解：研究到底是怎麼做的，也就是說：作結論所根據的數據是怎麼得到的。

蒐集資料的不同方式

蒐集資料的方式之一是**觀測研究**，這是一種被動的數據蒐集方式。我們只觀察、記錄、或度量，但是不對蒐集對象作任何干擾。抽樣調查就是很重要的一種觀測研究，因為我們只被動了解有關調查對象的資訊（比如對某議題的意見），並不做任何干涉。另一種方式叫做**實驗**，它是以主動方式產生數據；做實驗的人會主動介入，他會把某項處理（treatment）加諸於受試對象，來觀察受試對象有何反應。

所有的實驗以及許多觀測研究，都是想要知道一個變數對另一個變數有何影響。我們先介紹幾個必須用到的統計名詞：

實驗用語（見《統計學的世界》124 頁）

反應變數 (response variable)：用來度量研究結果的變數。

解釋變數 (explanatory variable)：我們認為可以解釋或造成反應變數變化的變數。

實驗中所研究的個體，通常稱為**受試對象 (subject)**。

處理 (treatment) 是任何加諸於受試對象的特定實驗條件。

例 1：上網學習的效果。（見《統計學的世界》122 頁）

有一項關於線上學習的樂觀報導，報告了在佛羅里達州羅德岱堡的諾瓦東南大學 (Nova Southeastern University, Lauderdale, Florida) 執行的一項研究。撰寫研究結果的人聲稱，學生在線上 (online) 修習大學部的課，和在教室裡學習的學生「學得一樣好」。如果把教室裡的課用網站取代，可以替大學省很多錢，所以照這項研究結果看來，我們應該全部上線？

在這個例子當中：大學生是諾瓦東南大學研究中的受試對象，而解釋變數是學習環境（是教室還是在線上）；反應變數是修完課後，學生的考試成績。

研究報告說：上網修課的學生，和在傳統教室修同樣課的學生學得一樣好，這個結果是否可信呢？我們首先來看看研究是怎樣做的。

諾瓦東南大學的研究不是實驗而是觀測研究，因為它沒有對學生受試對象加諸任何處理，而是任由學生自行選擇要在教室上課還是上網。研究中只度量了他們的學習成果。然而會選擇上網的學生，是否本來就可能和選擇教室的學生不太一樣呢？比如說程度較好以及較為自動自發的學生，是否會偏向於選擇上網組呢？因為可以自己控制上課進度。如此一來，這兩組的學生一開始就在本質上有差異，而這些差異極可能影響學習效果；也就是說，將這樣兩組學生拿來做比較，在基礎上就已不公平了，如此做出的結果，當然也就明顯不可靠了。換句話來說就是：上網學習以及教室學習的效果，已經和一些潛藏在背景裡的因素，無可救藥的混雜在一起了。這叫做交絡。

潛在變數 (lurking variable) 是對研究中其他變數間的關係有重要影響，卻並未被列為解釋變數的變數。當兩個變數對反應變數的影響混在一起、無法區分時，我們稱這兩個變數是**交絡的**。交絡的變數可以是解釋變數，也可以是潛在變數。

在上面的例子當中，潛在變數包括學生原本的程度，學生的學習習慣等等，任何（除了學習環境之外）可能影響修課後考試成績的因素。讓學生自己決定加入上網組還是教室組，則同一組的同學之間必定會有一些共同特質，而這些特質是會影響最後的評估結果的；換句話說，用這樣的分組方式，則交絡現象無從避免。

既然上述研究結果不可靠，那要如何做，才能比較教室上課和線上上課的效果呢？要知

道答案最好的方法是：

指定一些學生到教室上課，其他學生上網。這樣就是一項實驗，因為我們有主動介入，規定學生要用何種學習環境（指定學生上網、或者到教室上課，就是我們加諸學生的「處理」）。而為了盡量「公平」、避免人為因素的影響，分組應該用隨機方式進行。像這樣隨機分組、並且同時比較兩組或者更多組的實驗，叫做**隨機化比較實驗**。

例 2：胃冷凍（見《統計學的世界》126 頁）

實際在病人身上研究醫療效果的實驗，叫做「臨床試驗」（clinical trials）。胃冷凍曾經是胃潰瘍的標準療法，而使得胃冷凍變成胃潰瘍一般療法的臨床試驗，其架構是一種「單軌」設計：

給病人治療 → 度量病人的回應

胃冷凍 → 減低痛楚？

接受治療後病人的確表示比較不痛了，但是我們不能據此宣稱，胃冷凍導致痛楚的減輕。有可能只是「安慰劑效應」（placebo effect）。

許多病人對任何治療都有正面反應，即使只是安慰劑。這種對假治療的反應，就稱為安慰劑效應。實驗使用了單軌設計，代表安慰劑效應會和胃冷凍可能會有的任何效應交絡在一起。

數年之後做了另一項臨床試驗，把潰瘍病人先分成兩組。

一組就像前次試驗一樣，接受胃冷凍治療；另一組接受的是安慰劑治療，也就是打入汽球的溶液溫度和體溫一樣，而不是經過冷凍的。

結果：治療組的 82 位病人中，有 34% 病情改善，但是安慰劑組的 78 位病人中，也有 38% 有改善。這表示胃冷凍的效應，不過是和安慰劑差不多罷了，從此醫師就不再使用這種方法。

觀測研究和單軌實驗都常常因為和潛在變數的交絡問題，而產生沒有用的數據。如果只能觀察，交絡現象很難避免。做實驗，情況就好得多。想要釐清因果關係，在可行的前提下，都應該做隨機化比較實驗。

設計實驗的第一個目標：

確定實驗可以顯示解釋變數對於反應變數的影響。

單軌實驗常常因為交絡而達不到這個目標。

補救方法：

同時比較二個或多個處理，比如胃冷凍例子中的處理組（真正接受胃冷凍治療）及對照組（即安慰劑組）。

安慰劑組也叫做**控制組**（control group），因為藉著比較處理組和控制組，使我們能夠

控制潛在變數(例如安慰劑效應)的影響。控制組不一定是接受像安慰劑那樣的假治療，臨床試驗常常會把一種新的治療方法和已經在使用的方法比較，而不是和安慰劑比。隨機指派到現有療法的病人，就構成控制組。

實驗設計的邏輯

隨機化比較實驗是統計學裡面最重要的概念之一。它的設計是要讓我們能夠得到釐清因果關係的結論。因果關係是指：一件事情會導致某項結果的發生，例如若按時服用少量阿斯匹靈能減少心臟病的發作，則這兩件事之間就有因果關係；按時服用阿斯匹靈是「因」，心臟病發作次數減少則是「果」。

隨機化比較實驗的邏輯如下：

- 用隨機化的方法將受試對象分組，所分出的各組在實施處理之前，應該在各方面都類似。
- 用「比較」的設計以確保：除了實驗上的處理（如胃冷凍例子中的胃冷凍治療及安慰劑治療）外，其他所有因素都會同樣作用在所有的組上。
- 因此，反應變數的差異必是處理的效用所致。

用隨機方法選組，可避免人為指派時可能發生的系統性偏差。

法律已有規定，任何新藥必須用隨機化比較實驗來證明其安全性及有效性。但是對於其他醫療方法，比如手術，就沒有這項規定。

隨機指派受試對象時，也有可能運氣不好，把比如抽菸的人幾乎全放在同一組。解決方法是：用很多受試對象。此時利用隨機指派方式分的組，就有可能和母體有類似的組成。這個意思大致是這樣子的：比如有一組人、其中女生佔三成，然後我們要把這些人隨機平均分成兩組，其中一組會被分派麻煩的任務、另一組不會。如果女生都被分進任務組，是否會讓人懷疑不公平呢？但隨機分組是可能發生這種情況的，然而機率大小卻隨著人數而有不同；在總共 10 人、女生 3 人的情況下，女生全部被分到 5 人任務組的機率不算很小，但是若總共有 100 人、其中女生 30 人，則全部 30 個女生都被選進 50 人的任務組之機率就會非常接近 0 了。

有了基礎知識，我們再回到最初的媒體報導例子，判斷一下研究結果是否可信。兩項和糖尿病有關的研究，是觀測研究還是實驗呢？第一項研究很清楚的說明了是對一萬六千名芬蘭人進行長期追蹤後的結果，所以它是觀測研究。第二項研究雖沒說得那麼明白，但是用常理判斷一下，也就知道不可能是實驗；因為如果要進行實驗，必須把參與研究的糖尿病患隨機分成兩組，然後指定其中一組人要大量攝取咖啡因，另一組人則避免攝取咖啡因，一段時間之後再做比較。大量攝取咖啡因對身體有害，所以不可能做這種實驗。既然是觀測研究，代表兩項研究中大量喝咖啡的人，都是自己選擇喝咖啡的，而喝很多咖啡的人和少喝的人之間，會不會有其他的差異，而這些差異會影響研究結果呢？非常有可能。比如說，工作忙常常需要熬夜的人可能會喝較多咖啡，但這些人生活作息也較不正常、飲食也可能不規律，壓力也比較大，那我們如何知道對糖尿病的影響完全

是由於咖啡因而不是其他這些因素造成的呢？現在我們已經了解，觀測研究不能當作因果關係的依據，那麼有人根據觀測研究做出互相矛盾的結論，也就不足為奇了。

如果留意一下，會發現媒體常常報導有關健康的研究結果。其中有些結果未必可信，所以我們不應將這類資訊照單全收；但是若一律不信，又會錯過許多有用的訊息。現在我們知道應該怎樣「對付」這類研究結果了，就是先要了解資料的來源。如果資料是經由觀測研究取得，對於因果關係的結論，通常需要存疑；如果是經過嚴謹的隨機化比較實驗才得到因果關係的結論，可信度就大致沒問題了。

參考資料：統計學的世界 天下文化出版