

## 統計顯著性

我們常常在科學文獻中看到「具有統計顯著性」這樣的論述，可是我們真的了解統計顯著性( statistical significant )的涵意嗎？

首先，我們要注意到，在談統計顯著性時，一定會跟著一個 $\alpha$ ，正確的全稱是：在顯著水準  $\alpha$  下具有統計顯著性。一般選擇  $\alpha = 0.05$ ，意即若以資料來驗證  $H_0$  假設為正確時，這種機率不會超過 5 % (即 20 次中約發生 1 次以下)。如果我們選擇  $\alpha = 0.01$  時，就是指要否定  $H_0$  需要更強的證據，也就是說，當  $H_0$  為真時，會出現該次實驗得到的數據只有 1 %，即 100 次才有 1 次，也就是說：可能性很低。

所謂的「顯著」( significant ) 其意義並不是指「重要」，而是說「以機遇性來說是不容易發生的」。因為  $\alpha$  值是我們選擇的，就是指我們對所下結論願意冒的風險，所以也可以下這樣的結論：「在  $\alpha = 0.05$  水準下具有顯著性，但在  $\alpha = 0.01$  水準下就沒有顯著性。」

我們在應用顯著性檢定時，把原始資料輸進電腦，通過電腦程式很快的就可以計算出相應的檢定統計量( test statistic ) 及 P 值，當然電腦不會告訴我們統計顯著性是什麼涵意，因此一定要真正的弄清楚以上及以下的說明，正確的了解和解釋才是更有價值的。

利用電腦去執行「統計顯著性檢定」往往很容易，可是有智慧的運用檢定並不容易。學術研究報告中常廣泛的使用顯著性檢定，如果你能夠質疑某些不經思考就使用顯著性檢定的報告，並發現問題所在，這才是真的叫做「有深度」！

統計顯著性當然是很有價值的，因為它可以指出光靠機遇是很難會發生的效應。可是我們也要知道：「統計顯著性和實際上的顯著性是不一樣的。」因為：

當我們在常用的水準(  $\alpha = 0.05$  或  $\alpha = 0.01$  )下棄卻  $H_0$  (亦即假設是沒有效應或沒有差別)，它代表的是我們實驗得到的數據指向有效應存在。但是這個效應可能很小，因為我們只要取得夠大的樣本，那麼即使很小的偏離，也可以得到統計的顯著性。還有，有統計顯著性並不一定有價值，以實際的應用目的來說，或許在成本上或時效上是不符合經濟效益的、是沒有實用價值的！

還有，當我們收集到資料後，不妨先畫畫圖仔細檢視一下，看看有沒有離群值或有沒有明顯偏離整體型態的狀況？因為少數的離群觀測值就可能高度影響檢定的結果有沒有顯著，把可能本來很令人信服資料的顯著性給毀了。把圖畫出來，自己看看你認為的效應在圖上看不看得出來？如果看不出來，就想想這樣的效應到底有沒有實際的重要性或實用價值？計算一下有興趣參數的信賴區間，因為那可以估計效應的大小，不要只是在在意有沒有顯著！

傳統上對  $\alpha = 0.05$  的強調也帶來了一些問題。因為對於學術期刊的研究報告，除非達到  $\alpha = 0.05$  的顯著水準，否則就很難被接受和發表。於是，有些另有目的的文獻發表，是在重複做了很多次以後，如果沒有「統計顯著性」的就按下不表，做多了總是會做出「具有統計顯著性」的就拿出來發表(20次就可能有一次吧)，這可是真真實實的實驗數據，完全沒有做假不怕檢驗，可是這樣經不起重複的顯著除了誤導又有何價值？

一項研究如果沒有顯著性就沒有價值嗎？其實也不是！例如醫學上的研究如果能夠驗證，某個新療法(或新藥)「無法」提供顯著的幫助，也就是說效果不顯著，這項研究也是有價值的。因為可以節省資源，避免批准或使用缺乏效果的療法或藥物。如果設計良好的研究結果卻找不到效應，不是也應該值得公開嗎？因為除了可以避免錯誤的引入以外，還可以讓後人不要再重複的做類似的研究！

最後要提醒的是，統計推論不是對所有的資料都適用的，一個設計不良的調查或實驗，當然會得到站不住腳的結果。以民意調查來說，結果的不可靠或變化大，不能全部推到什麼「民意如流水，初一十五不一樣」這種論調！因為，不管是顯著性檢定還是信賴區間的統計應用基本都是基於機率法則，所以抽樣或實驗的隨機化才能保證它的適用性。如果在隨機化上有瑕疵，卻仍用搜集的資料去做統計推論，這才是真正的問題所在，由此而作的推論當然也是會出問題的！

以上的討論相當重要，請同學務必認真研讀，也可以相互討論，真正弄懂了以後自會受用無窮！