

統計檢定力分析之探討

林佩芬

勤益工商專校企管科

摘要：

在統計教學或一般研究中，研究者通常著重於型 I 錯誤的機率 α ，而採相當嚴苛的標準，在此種作法下，卻忽略了型 II 錯誤的增加，即可能造成研究的失敗，本文從哲學觀點，統計學的觀點，來說明重視型 II 錯誤的重要性，並舉例計算 α 、 β 、 r 、Sample Size 之間的關係，希望提醒研究者進行檢定時，不會再忽視 β 的存在推算。

檢定力分析(Power Analysis)在統計學中並非新鮮的事，只要隨意翻開一本教學課本，都可以找到型 I 錯誤(Type I Error)及型 II 錯誤(Type II Error)的定義，然後會敘述型 II 錯誤和檢定力之關係，爾後，可能再也找不到有系統敘述檢定力分析的文章。

基於以下二個原因，引發了撰寫本文的動機，第一：一般教學課本似乎都太偏重 α 值的設定（即型 I 錯誤的機率），講求“被研究關係”確認之信賴度，至於 β （型 II 錯誤的機率）似乎太過忽視；第二：在講求研究效率的情形下，不得不去重視一個錯誤研究的成本。

本研究主要目的是希望透過有系統之介紹，讓讀者了解到檢定力分析的意義，並適度的導入統計學教材中，使得以往過份重視 α 值分析的情形得以改善。

哲學觀點

社會科學與自然科學一項很大的差異是：自然科學的法則(Law)大都是普遍性法則(Universal Laws)，它是指“每次X事件發生後，Y事件就會發生”；而社會科學直至目前為止尚未找到如此的法則，在社會科學中所使用的法則，僅能被稱之為統計法則(Statistical Law)。(註一)由此可了解到統計對於社會科學的重要性。

自波普(Poper)提出否證論後，統計學的發展亦受其影響。否證論意味著，不論有多少次的成功，只要有一次失敗就可以推翻整套理論。當然，如前面所敘述的，在社會科學中並沒有普遍性法則的存在，任何理論在被測試時，或是任何“被研究關係”在被實証時，都會面臨著機率的問題，為了強調確認程度(the Degree of Confirmation)的提升，所以在欲否定的虛無假設上，設定如此嚴苛的標準（通常 $\alpha = 0.05$ 或 0.01 ）。

在此對型 I 錯誤與型 II 錯誤的定義不再贅述，而直接討論它們對研究之影響。在一般的研究中，研究者是在找尋一些變數間的關係，所以在作統計檢定時，會有以下的形式：

H_0 ：A與B無關

H_1 ：A與B有關

在一般情形下中討論 α ，也就是在“A與B事實上無關”的情形下卻認為“A與B有關”。嚴格的 α 值意味著否定論的說法——不要隨意認定關係。而從另外一方面來考量，在“A與B事實上有關”的情形下，而被認為“A與B無關”，這種錯誤是比較輕微的。對此應該提出二項批判：第一：在研究效果上來看， H_1 被否定的機會太大，而認定研究失敗，是過於草率，其原因可能只是 β 值太大，而造成的失敗；第二：管理學研究的範圍太大，往往一項關係被否定後再被研究之機率不大，而容易造成進步的障礙。因此，在每次作統計檢定時，附上 β 值的計算是有必要的。

問題及其程度

以 t 檢定為例， t 之計算可由兩部份組成

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{df}$$

在式中， r 表示兩個變數間之相關程度，是受兩個變數間關係的強弱而定， $r / \sqrt{1-r^2}$ 則可以解釋成作用力 (Effect Size)， df 為 t 統計值之自由度，其值等於總數減 2 ($N-2$)，受到研究樣本大小的影響， df 則可解釋成樣本大小對統計值 t 之貢獻，又可之為研究力 (Study Size)，將此概念概化可得下式之關係：

$$\text{顯著性檢定} = \text{作用力} \times \text{研究力}$$

作用力是指變數間關係之強弱對顯著性檢定之影響，研究力是指樣本大小對顯著性檢定之影響。

在兩個變數間關係強度不弱時，假定為 0.3，其作用力已達中度之水準（註二）。如果採 $\alpha = 0.05$ ，雙尾的檢定，在總樣本數只有 100 個的情形下，則關係被證實出來的機會不到六成，如果將 α 設定為 0.01 時，則機會立刻下降至不到四成。如果研究的樣本不夠大，就算是有關係存在的事實，也無法被證實。

Cohen 曾提出一套測定 α 及 β 相對嚴重性的指標，稱之為 β / α 比，它的計算方式很簡單，即在一定的作用力及固定的樣本大小下，每設定一個 α 值，就相對產生一個 β 值，例如 $\alpha = 0.05$ 及 $\beta = 0.6$ 時，其 β / α 比為 $0.6 / 0.05 = 12$ 。（註三）。藉此觀念，可以整理出如表一的計算值。

表一：作用力與顯著水準之關係

N	Effect sizes(r)and significance levels(.05 and .10)					
	r = .10		r = .30		r = .50	
	.05	.10	.05	.10	.05	.10
10	19	9	17	8	13*	5+
20	19	9	15	6	7*	2+
30	18	8	13	5	3	1
40	18	8	10	4	2	
50	18	8	9	3		
60	18	8	7	2		
70	17	8	6	2		
80	17	8	4	1		
90	17	8	4	1		
100	17	7	3			
120	16	7	2			
140	16	7	1			
160	15	6				
180	15	6				
200	14	6				
300	12	5				
400	10	4				
500	8	3				
600	6	2				
700	5	2				
800	4	1				
900	3					
1000	2					

Note: Entries are to nearest integer; blanks indicate values < 1 .

*For $r = .70$ these ratios would drop to 6 and < 1 , respectively.

+For $r = .70$ these ratios would drop to 2 and < 1 , respectively.

資料來源：同註二

由表一可以了解到問題的嚴重性，當關係強度不是很高時($r = 0.1$)，如果樣本數不是很大(樣本數不超過100)，型II錯誤的機率將高出型I錯誤機率很多，其意味著：很容易忽略或否認兩個變數間既有的關係。

檢定力分析

檢定力可以定義為當 H_0 為偽時，拒絕 H_0 之機率，簡單地說就是 $1 - \beta$ ，或者可以解釋為不犯型II錯誤之機率。

在前面已經討論到，檢定力和作用力大小， α 值之設定，及樣本大小皆有關係。對於一個

研究者而言，了解這些數值之間的關係是十分重要的。以Cohen的研究為例，他的著眼點是先假定作用力之大小及設定 α 值，再找出不同檢定力下所需要之樣本數大小為何。

利用 r 來檢定變數關係時，在顯著水準為0.05的情形，不同作用力及檢定力水準下，所對應之樣本數大小如表二所示。

表二 $\alpha=0.05$ 時 r .power及sample size之關係

Power	Effect sizes(r)						
	.10	.20	.30	.40	.50	.60	.70
.15	85	25	10	10	10	10	10
.20	125	35	15	10	10	10	10
.30	200	55	25	15	10	10	10
.40	300	75	35	20	15	10	10
.50	400	100	40	25	15	10	10
.60	500	125	55	30	20	15	10
.70	600	155	65	40	25	15	10
.80	800	195	85	45	30	20	15
.90	1000	260	115	60	40	25	15

資料來源：同註二 pp92-93

r 經常被使用，因為其他的指標很容易轉換成 r （註四）。而在比較兩組關係間之差異時， $(r_1 - r_2)$ ，必須先將 r 值轉換成Fisher Z值，其公式為： $Z = 1/2 \ln \{ (1+r)/(1-r) \}$ ，再求出兩個Z值之差異 $(q = Z_{r1} - Z_{r2})$ ，此 q 值將比直接差異值 $(r_1 - r_2)$ 為大，故可獲得較為顯著之作用力，由表三中可以找出在不同 q 值及期望檢定力下，所需要之樣本大小。作Fisher Z轉換之目的是要擴張作用力的大小。

表三 q 值、power及sample size之關係

power	Effect sizes($q = Z_{r1} - Z_{r2}$)*						
	.10	.20	.30	.40	.50	.60	.70
.15	170	45	20	15	10	10	10
.20	250	65	30	20	15	10	10
.30	400	105	50	30	20	15	15
.40	600	150	70	40	30	20	15
.50	800	195	90	50	35	25	20
.60	1000	240	115	65	45	30	25
.70	1250	315	140	80	55	40	30
.80	1600	400	175	100	65	50	35
.90	2100	540	235	135	90	65	45

資料來源：同註二 p125-126

以下舉兩個例子來說明檢定力在研究中的重要性。假設要研究A及B兩個變數間之關係，事

前預估它們之間的相關程度在0.3至0.6之間，如果統計檢定設定 $\alpha = 0.05$ 雙尾，而要求有0.9的檢定力時，由表二得知所需要之樣本大小在25個至115個之間。再假設以前曾有研究者研究過相同之問題，其只採用了20組的樣本，而得到統計值 $t = 1.59$ ，以此研究來分析，當 $t = 1.59$ 時，相對應之 r 值為 $0.35 (r = \sqrt{t^2 / (t^2 + df)})$ ，依表二可查出在 $\alpha = 0.05$ ，雙尾的檢定下，此研究之檢定力只有三成左右，因此犯型II錯誤之機會很高（約七成左右），換句話說，在A及B變數間有關係而誤判為無關之機率將高達七成。因此若要進行這樣的一個研究，可將作用力預估為0.3到0.4之間，而樣本數之選擇在115個至60個之間。

第二個例子是分析一個實際失敗的研究（註五）。此研究在討論高年級幼童及低年級幼童對一種非語言溝通測驗之敏感性差異。本研究共採用了50位高年級幼童及30位低年級幼童，分別得到兩個相關係數0.37及0.6，在經過Fisher Z轉換後得到 $Z_{0.37} = 0.39$ ， $Z_{0.6} = 0.69$ ，故 $q = Z_{0.6} - Z_{0.37} = 0.69 - 0.39 = 0.3$ ，在查表三時，是假定兩組之樣本數大小要一致，否則以調和平均數為其樣本數，故本研究之約當樣本數為：

$$n = \frac{1}{1/2(1/n_1 + 1/n_2)} = \frac{1}{1/2(1/30 + 1/50)} = 37.5$$

在表三中，可查出此研究之檢力只有0.2至0.3之間，無怪乎研究會失敗，因為型II錯誤發生之機率太高，其實作用力 $(r_1 - r_2)$ 在0.3時，已屬中度相關水準，所以本研究要成功必須增加樣本數。

增加作用力的影響力

欲增加作用力的影響力可藉由四個方向著手（註六）。第一：選用適當之統計值，例如在組間或組內有線性迴歸關係存在時，選用 r 比 t 值好，因為 t 值無法取得線性迴歸之趨勢。第二： α 值之設定，在可容許的範圍內適度提高 α 值，而使 β 值下降，或是在一般顯著水準下，檢定結果不具顯著性時，應該將 P 值計算出來。第三：樣本大小的調整，樣本愈大時， β 值及 α 值可以同時兼顧。第四：作用力大小亦可影響檢定力，如果作用力愈大，則 β 值愈小，但是作用力是被研究的對象，通常是一個未知數。

結 論

透過適當的討論來顯現型II錯誤之嚴重性，是有其必要的，雖然在統計教科書中常以一句話，型I錯誤比較嚴重，而不再理會型II錯誤，藉由本文的探討，希望提醒研究者不應該再輕忽 β 的推算。

註 釋

註一：Shelby D. Hunt, *Modern Marketing Theory: Critical Issues in the philosophy of Marketing Science*, Ohio, Cincinnati, U.S.A.: South-western Publishing Co. chap5.

註二：J.Cohen, Statistical power Analysis, for Behavioral Sciences, Academic Press
 , New York, U.S.A.1977

註三：同註二

註四：Robert Rosenthal and Ralph L. Rosnow, Essentials of Behavioral Research:
Methods and Data Analysis, McGRAW-HILL, Inc, New York U.S.A.1991, 2ed editor
P452

註五：B.M. Depaulo and R. Rosenthal, "Age change in nonverbal Decoding skill:Evi-
dence for Increasing Differentiation" Merrill-Palmer Quarterly v.25.1979 PP
145-150

註六：同註四，P440