多元迴歸分析中之結構係數與逐步迴歸

傅粹馨

國立高雄師範大學教育系副教授

多元迴歸分析在行爲科學研究領域中，爲使用相當普遍的統計方法，它可用於社會學、心理學、經濟學、政治學和教育上的研究；它亦可使用於實驗的和相關的研究；它可處理連續的和類別的變項，且多元迴歸亦可達成變異數分析（ANOVA）的功能，它也具有平方和（SS）、均方（MS）和F值，甚至可提供更多的訊息。本文擬針對結構係數（structure coefficient）和逐步迴歸分析二項加以探討。

一、結構係數

多元迴歸分析在探討數個預測變項與一個效標變項間的關係，當預測變項間彼此獨立，沒有相關，則迴歸分析的解釋很單純、明瞭，但於實際研究情境中，預測變項間卻有某種程度的關連，稱之爲共線性（collinearity），亦稱之爲多元共線性（multicollinearity），它所造成的影響有三：(一)共線性問題嚴重時，會影響最小平方值（least-squares）計算的精確性（accuracy）；(二)影響統計數據的精確性；(三)影響結果解釋的精確性（Thompson & Borrello, 1985），由此可知多元共線性的影響力。Pedhazur（1982）指出，預測變項間的高度相關，不僅導致迴歸係數之預估值嚴重的扭曲，而且，造成係數之正負符號相反的情形，故多元共線性對標準化的迴歸係數，β_1（β加權／β weights）的解釋亦會造成影響。

由上可知，β頗受多元共線性的影響。早於1964年，Meredith即強調：在典型相關分析時，應解釋結構係數而不是典型函數係數（canonical function coefficient），而多元迴歸是典型相關的一個特例。Colley和Lohnes（1971）、Kerlinger和Pedhazur（1973）、Levine（1977）均認爲在迴歸分析時，應對結構係數加以解釋。

結構係數乃是預測變項與合成或潛在變項（$\hat{Y}$, synthetic/latent variable）的相關（Thompson, 1992），計算方式爲每個預測變項（P）與效標變項（C）的相關係數（r_{pc}）除以複相關係數（R）。茲以Thompson（1990）文中之表加以說明。於表一中，C1表示效標變項，P1至P4是預測變項。

r_{P1C1}＝.107，P1的結構係數爲.107/.527=.203，其餘各預測變項之結構係數依此類推。由表一可知，四個結構係數均比原來各自與C1的相關係數提高了一些，如.203>.107（P1而言）；.395>.208（P2而言）；.858>.452（P3而言）；.826>.435（P3而言）。

結構係數不受共線性的影響，但有時會產生一個現象，即β與結構係數的正負符號

不同（Thompson & Borrello, 1985），導致研究者不知該選擇那一個才是。Thorndike（1978, pp.171-172）提出：當研究者想了解每個預測變項所貢獻的變異量時，則採 β ；當研究者之興趣在了解預測變項與合成變項（$\hat{Y}$）間的關係，則採結構係數。可知，不同的係數回答不同的問題。

表一　相關係數矩陣

變　項	P1	P2	P3	P4	
C1	.107	.208	.452	.432	.527[a]
P1		.449	.262	.311	.203[b]
P2			.332	.282	.395[b]
P3				.460	.858[b]
P4					.826[b]

（a：複相關係數R=.527；b：各預測變項的結構係數。）
（引自Thompson, 1990）

事實上，每個變項的積差相關（預測變項與效標變項的相關）、β 與結構係數，在解釋結果上均扮演著重要的角色。而Cooley和Lohnes（1971,p.55）建議：當樣本不很大時，β 之數據易受嚴重的波動，呈現不穩定，亦即在某個小樣本之 β 值於另一個小樣本時，β 值會有大的變動，在此情形下，建議少用 β 係數來解釋結果。Thompson和Borrelo（1985）和Thompson（1992）強調，當一個預測變項之 β 爲0或接近0，但它可能是個具有強大預測力的變項。

Thompson於1990指出，許多使用迴歸分析者，在解釋結果時忽略了結構係數，這是值得注意的。Thompson（1992）認爲思考縝密的研究者在解釋迴歸分析結果時應採以下㈠或㈡的方式：㈠採用 β 和結構係數；㈡採用 β 和預測變項與效標變項間的相關（ r_{pc} ）。

二、逐步迴歸分析

在社會科學的研究中，逐步迴歸（stepwise regression）分析使用得相當普遍（Marascuilo & Serlin, 1988），博士論文中亦有不少人使用該方法作資料分析（Thomspon, 1988）。不幸地，逐步迴歸法會導致分析結果產生嚴重的誤解，許多學者認爲必須審慎使用該方法(Cliff, 1987; Huberty, 1989; Pedhazur, 1982;Snyder,1991; Thompson, 1988, 1989, 1994, 1995)。

首先將逐步迴歸分析的過程作簡扼的敘述，再將使用時會產生的三個問題加以說明，以期使用者對該分析方法有所了解。

㈠逐步迴歸分析

於多元迴歸分析中，有順向解法（forward solution），其特性是一次只能允許一個預測變項進入迴歸公式，第一個被選入迴歸公式者，是預測變項與效標變項（Y）間相關最高的（如X1），第二個進入迴歸公式者乃是其餘預測變項，各剔除了第一個預測變項（X1）的影響力之後，與效標變項的部分相關（part correlation）最高者，此種方式使得每次 R_2 的增加量爲最大，如此循環，直到 R_2 的增加量不再達統計上的顯著水準

爲止，則預測變項不再進入迴歸公式，當預變項進入公式後則留在該公式中，此即爲順向解法的迴歸分析。

逐步迴歸分析可視爲順向選擇的一個改良，它也是每次只容許一個預測變項進入迴歸公式，不同的是，順向法的預測變項只有進入迴歸公式內，而不會從公式中又被移出（remove ），然而，在逐步迴歸法中，於某步驟中（step）進入迴歸公式的預測變項（如X2），可能於下一步驟因其他變項（如X4）進入公式中，而X2不再具有顯著的預測力時，則X2會被排出（Thompson, 1989），亦即每一步驟中，被選入迴歸公式的預測變項，均再重新評估其重要性（林清山，民69）。

逐步迴歸法多半用於選擇變項（variable-selection），從許多的預測變項中，選出少數幾個具有預測力的變項。Huberty（1989）認爲逐步迴歸法的功能有三：㈠選擇或刪除變項；㈡評估變項的重要性；㈢選擇變項且評估其重要性。他指出，人們用此法太頻仍，而不是應該用的時候才用（Huberty, 1989, p.45）。

㈡逐步迴歸所產生的問題

Thompson指出逐步迴歸法所產生的三個問題，茲分述如下：

1.使用不正確的自由度

許多研究者均使用統計套裝軟體作資料分析，然而，在作逐步迴歸時，評估變項在變異量的改變上，軟體並未使用正確的自由度（correct degees of freedom）。茲舉Thompson （1995）的例子來加以說明。該資料有100名受試與50個預測變項，於逐步迴歸分析時的第五個步驟，得到如表二的結果：

表二　逐步迴歸分析結果

分析	變異來源	SS	df	MS	F(計算值)	F（臨界值）	R^2
1	Explained (regression)	20	5	4.0000	4.75	4.41	20.00%
	Unexplained (error)	80	95	0.8421			
	Total	100	100				
2	Explained (regression)	20	50	0.4000	0.25	***	20.00%
	Unexplained (error)	80	50	1.6000			
	Total	100	100				

（引自Thompson, 1995）

就表二上半部，總變異來源（Total）的自由度爲100，是正確的；而預測變項能解釋（Explained）的變異來源的自由度爲5，亦即有五個預測變項已進入迴歸公式；預測變項不能解釋（Unexplained）的變異來源的自由度爲95，亦即101-1-5=95，整個結果的F值達.05統計上的顯著（statistically significant）。

不少研究者（Cliff, 1987; Snyder, 1991; Thompson, 1989）均認爲其中的explained （在SAS中用regression）和unexplained（在SAS中用error）的自由度是錯誤的。假若，這五個預測變項是隨機抽取的，則explained的自由度爲5，尙不致產生問

題，但是這五個被選中的預測變項，在這五個步驟中，每個步驟均有50個預測變項參與其中，故於表二的下半部，explained（regression）的自由度改爲50，unexplained（error）的自由度改爲50，則F值由4.75度爲.25，由統計上的顯著變爲不顯著，可知，軟體之統計分析趨於高估結果，造成易於犯第一類型錯誤。當遇上以下之(一)或(二)的情況，則犯第一類型錯誤之情形將更爲嚴重：(一)預測變項數目多而步驟數目卻少；或(二)樣本人數少。

Huberty （1989）亦指出，於逐步迴歸分析時，我們對於電腦報表所提供之F統計與顯著機率（probability），不宜太認眞（should not be taken too serously），電腦對於進入或移出公式的變項，所提供的機率不是往常看見的傳統的顯著水準（Afifi & Clark， 1990，p.215），一則因爲逐步迴歸涉及許多變項的統計考驗，易趨於第一類型錯誤增加的機率（wilkison，1979），再則，因爲F統計數是個相當複雜的問題。

Thompson（1995）指出，多數研究者在使用逐步迴歸時，均採用了統計軟體所提供的錯誤的自由度，這不意謂著我們必須遺棄這種方法，只要我們用正確的自由度取而代之，再計算其餘的相關部分即可。

逐步迴歸在以下的三種情形，或許問題的嚴重性會降低：(一)樣本人數相當多；(二)預測變項的數目不多；(三)explained的離均差平方和（SS），於每個步驟中均接近0（Thompson，1995）。

2.無法辨認q個最佳組合的預測變項

許多研究者誤認爲使用逐步迴歸法可以辨認出那些變項是最佳q個變項的組合（the "best" subset of size q）（Thompson，1989，1995），從Thompson （1995）的研究顯示，假若有四個預測變項爲X1~X4與效標變項爲Y，當採用逐步迴歸法時，X1與X2被選出，其複相關（R^2）爲.45；但當相同的資料用所有可能組合法（all-possible-subsets）分析時，q=2的最佳組合卻爲X3與X4，R^2=.475。由此例可知，當q=2的最佳組合的預測變項（X3，X4）卻未包含使用逐步迴歸法所選出的那兩個預測變項（X1與X2）。這例子引申的含意爲：以逐步迴歸法所選出的q個預測變項未必即是所謂的最佳的q變項的組合（Huberty，1989）。例如：有30個預測變項，採逐步迴歸分析，只作三個步驟（step），則選出的三個預測變項有可能與最佳組合只含三個預測變項的結果全然不同，亦即最佳組合所含的三個變項，卻不包括逐步迴歸所選的三個變項。

由上可知，逐步迴歸法是每一步驟選出「次最好」的預測變項（next-best-predictor）進入公式，但需考慮已在公式內的預測變項。故於q的步驟中，每一步驟選出一個次最佳（next-best）的預測變項，這情形與選擇最好的q個組合的變項是不同的。換言之，逐步迴歸分析從不同時考慮所有預測變項的組合（all the combinations of the predictor variables）。

Thompson（1988）指出學生誤將變項進入公式的次序與最佳q個變項組合的選擇混爲一談。該學生利用逐步迴歸分析，作如下的結論：學電腦經驗（r=-.288）與數學焦慮(r=-.141)對被預測變項（電腦焦慮）的變異量有顯著的貢獻，故只提及這兩個變項的重要性，然而第三個預測變項（以往上數學課的次數）與電腦焦慮的相關(r=-.175)，比數學焦慮與電腦焦慮的相關要高些，學生卻未考慮，只因爲在逐步迴歸分析中，只有選中兩個預測變項（電腦經驗與數學焦慮）。

3.以進入迴歸公式之次序來評定預測變項的重要性

也許常常會看到以下的敘述：「最重要的預測變項爲V3，因爲它是第一個進入迴歸公式，第二個重要的預測度變項是V1，因爲它是第二個進入迴歸公式…。」有些人云：「…有些變項不“重要”的，因爲它們不符合進入迴歸公式所訂的標準（Huberty，1989，p.46）。」，以上的敘述不宜採用，因爲預測變項進入迴歸公式的次序不能用來評估變項的重要性。Huberty （1989）作如下的解釋：於逐步迴歸分析時，當預測變項與效標變項相關最高者，第一個進入迴歸公式……，話說第三個預測變項進入公式時，已受到前面兩個已在迴歸公式內的變項所影響，若其中的一個或兩個變項更換了，則第三個變項也跟著不同了，亦即，若第一個變項選錯了，則其後的步驟全盤皆錯，這種依賴性或條件性（dependence or conditionality）的特性，使得以逐步迴歸分析來評估變項的重要性是有待商榷的。Snyder （1991）即舉了一個例子說明逐步迴歸分析易受取樣誤差（sampling error）的影響，造成分析結果不易複製（replicate），亦即不容易類推到其他樣本，也就是所謂的樣本特定性（sample-specific）（Thompson，1994）。該例子中的樣本爲12人，有四個預測變項（X1-X4）與效標變項（Y），其相關矩陣列於表三。

表三　相關矩陣

	ZY	ZX1	ZX2	ZX3	ZX4
ZY	—	.497	.444	.384	.319
ZX1	24.7%	—	.018	-.074	-.004
ZX2	19.7%	.0%	—	-.054	.099
ZX3	14.7%	.5%	.3%	—	.103
ZX4	10.2%	.0%	1.0%	1.1%	—

（對角線以上之數據爲相關係數；對角線以下爲r^2）
（引自Snyder，1991）

樣本爲12人，該研究想了解抽樣誤差所造成的影響，於是每次以取樣9人的方式，如第一個樣本（N=9）是去掉了1,2,3號的受試，第二個樣本（N=9）是去掉了1,2,4號的受試，如此則有220種N=9的組合，然後採逐步迴歸分析220個樣本，所得結果爲：有87個樣本的X1 爲第一個變項進入公式，在整個N=12的情況下，X1爲第一個進入公式的變項；有68個樣本是以X2爲第一個進入公式的變項；有25個樣本是以X4爲第一個進入公式的變項，結論並不一致。由此例可知，由小樣本所得到的變項進入公式的次序並不能推論（generalize）至母群中變數的重要性。

欲使抽樣誤差的影響減少，可從三方面加以注意：(一)採用較大的樣本；(二)採用較少的預測變項；和(三)具有較大的效應量（effect size），在此情況下使用逐步迴歸法或許罪惡感會稍微減輕些。

三、逐步迴歸法問題的解決途徑

Huberty（1989）提出了他對逐步迴歸法的替代方案，但事前先有其所依據的假設：(1)研究中所使用的變項必須已經過審慎的抉擇；(2)研究變項數目能少於30個；(3)有理論的或充份的理由要刪除某些變項；(4)研究者要透過統計軟體處理資料。首先就變項的選取加以敘述，再就變項的重要性加以說明。

㈠變項的選取

1.所有可能變項的組合

可利用SAS的RSQUARE的指令或利用BMDPQR，電腦可提供2^P-1的迴歸公式，如有3個變項X1-X3 ，則會產生X1,X2,X3,（X1,X2），（X1,X3），（X2,X3）和（X1,X2,X3）七種組合的R^2與C_p，當變項數目增加時，則組合數目會急遽的增加。研究者可在各組合中，挑選複相關平方（R^2）最高者，爲該組合中最佳的組合，例如於前述之（X1,X2），（X1,X3）和（X2,X3）中，以（X1,X2）之R^2最高，則該組即是所有q=2中的最佳組合。若研究者有八個變項，則他可以知道誰是q=7的最佳組合，次佳組合；他也可知道q=6時的最佳、次佳組合；q=5至q=2均依此類推。

2.決定那一種組合是最好的

當得知各種組合的R^2之後，要決定到底是q=4或q=5或q=6，該如何決定，可採用下述的方法：

・利用調整後複相關係數的平方（adjusted R^2）。茲舉Afifi & Clark （1990）的資料說明該方法。於表四中爲效標變項（Y）與五個預測變項（X1-X5）的相關矩陣，表五爲所有可能組合的R^2，調整後的R^2（$\hat{R}^2$）與Cp（Mallows, 1973）、Up（Bendel & Afifi, 1977 ）。對$\hat{R}^2$，Cp，Up稍作解釋，其計算方式分別列於公式(1)、(2)、(3)，以表五之q=2，（X5, X6）的那一組爲例說明R^2與Up之計算，因資料不足，故Cp之計算從略。（N=30）

$$\hat{R}^2=R^2-\frac{P(1-R^2)}{N-P-1} \qquad \text{公式(1)，} \hat{R}^2=.404-\frac{2(1-0.44)}{30-2-1}=.360$$

$$U_p=\frac{1-R^2}{(N-P-1)(N-P-2)} \qquad \text{公式(2)，} U_p=\frac{1-.404}{(30-2-1)(30-2-2)}=.000849$$

$$.000849\times1000=8.5=U_p$$

$$C_p=(SSE\ /MSE)-(N-2P)+1 \quad \text{公式(3)，} C_p=\text{（從略）}$$

表四　相關係數矩陣

	Y	X_1	X_2	X_3	X_4	X_5	X_6
Y	1						
X_1	0.32	1					
X_2	-0.47	-0.46	1				
X_3	0.13	0.36	-0.02	1			
X_4	-0.20	0.56	0.09	0.39	1		
X_5	0.35	0.62	-0.22	0.25	0.15	1	
X_6	0.33	-0.30	-0.16	-0.45	-0.57	-0.43	1

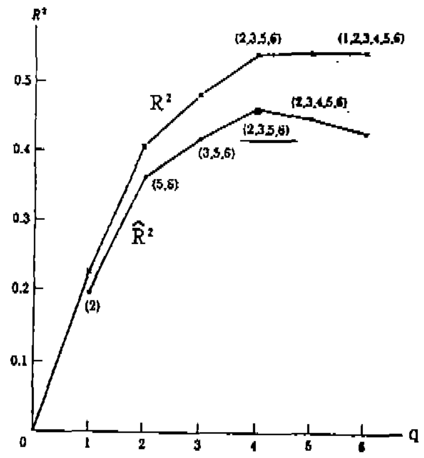
（引自Afifi & Clark, 1990）

表五　所有可能組合之R^2，$\hat{R}^2$，C_p與$U_p \times 1000$

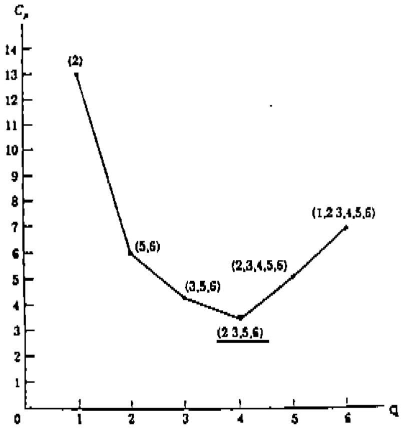
變項數目	變項	R^2	$\hat{R}^2$	C_p	$U_p \times 1000$
1	X2*	0.244	0.197	13.0	1.03
1	X5	0.123	0.092	18.1	
1	X6	0.106	0.074	18.9	
2	X5,X6	0.404	0.360	6.0	0.85
2	X2,X6*	0.288	0.235	11.8	
2	X2,X5	0.287	0.234	11.8	
3	X5,X6,X3	0.477	0.147	4.3	0.80
3	X2,X6,X5*	0.472	0.410	4.6	
3	X5,X6,X1	0.428	0.362	6.8	
4	X2,X5,X6,X3	0.534	0.459	3.4	0.78
4	X5,X6,X3,X4*	0.494	0.413	5.4	
4	X5,X6,X3,X1	0.484	0.402	5.9	
5	X2,X5,X6,X3,X4*	0.541	0.445	5.1	0.83
6	X1-X6	0.542	0.423	7.0	0.90

*表示由逐步迴歸所選出的組合。
（引自Afifi & Clark, 1990）

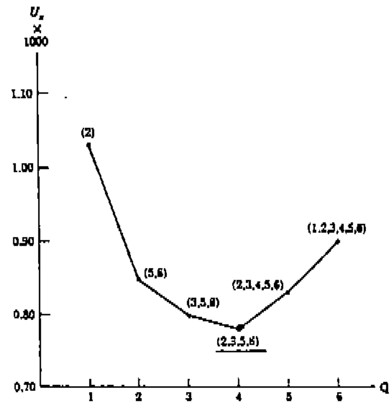
根據表五的數據，可畫出圖一至圖三的曲線。圖一表示不同的最佳組合（q=1到q=6）與相對應的R^2，$\hat{R}^2$的曲線，由圖可以看出$\hat{R}^2$在（2,3,5,6）處爲最高點，當q=5或q=6時，曲線有下斜下之傾向，表示（2,3,5,6）這四個變項是最佳的組合。圖二爲Cp與不同的q值的曲線，Cp愈小愈好，表示誤差平方的總和愈小，故此曲線之最低點正好是（2,3,5,6）處。圖三亦和圖二有相似的結果（Up值愈小愈好）。故由此三個圖所得的結論是：q=4是最佳的組合，變項爲X2,X3,X5,X6，而不是其他的變項；q=5或q=3的組合都未必比q=4的組合爲佳。



圖一　R^2與$\hat{R}^2$分別與不同之q的曲線圖
（引自Afifi & Clark, 1990）



圖二　Cp與不同之q的曲線圖
（引自Afifi & Clark, 1990）



圖三　Up與不同之q的曲線圖
（引自Afifi & Clark, 1990）

㈡變項的重要性

當找出變項的最佳組合後，再著手於各變項相對的重要性，其方法如下：

1.$R^2_p - R^2_{(i)}$，i=1,2,……,P-1，茲舉例說明之。假設有X-X4四個預測變項與效標變項Y。$R^2_{y.1234}$代表R^2_p，而$R^2_{y.234}$則代表$R^2_{(1)}$，表示X1不在其中，而只有X2-X4的三個預測變項，則$R^2_{y.1234}-R^2_{y.234}$就代表X1所造成的差異部分，$R^2_{y.1234}-R^2_{y.134}$代表X2所造的差異，X3與X4依此類推，再比較此四個的差異值，最大者可謂相對性的重要性較高，次大者其重要性次之。

2.(1)從Huberty（1989）的觀點認為各個變項的相對貢獻，R^2（即第一項所敘述的，$R^2_p - R^2_{(i)}$）大的即較為重要。(2)以往的認為第一個進入迴歸公式的變項即為最重要的（important），第二個進入即為次重要的，現今已不如是說了。(3)以標準化的迴歸係數的大小來表示其重要性，亦有可議之處，因為預測變項間的多元共線性對其產生影響。(4)Huberty （1989）認為Thompson和Borrello（1985）、Cooley和lohnes（1971）所提的結構係數，可用來表示預測變項的重性。用此係數產生的問題是此法基本上忽略了所有預測變項的系統（"system" of predictors），從公式

可得知，即每個r_{xy}（或r_{xc}）除以一個常數（R），原來之r_{xy}（或r_{xc}）最大者，其結構係數亦是最大。

3.Achen（1982, pp.68-77）提出了三種解釋預測變項重要性的方法，基本上，這些方法仍未離開使用標準化（β）與未標準化（b）的迴歸係數。

四、結語：

全文對逐步迴歸分析的缺失作了簡扼的分析，並提供了Huberty（1989）所建議的方法，先選變項，再決定其相對的重要性。然而，使用迴歸分析時，仍須注意如下的事項（Huberty, 1989）：

(一)模式的特殊性

原本爲p個變項的模式，經過迴歸分析後，選出q個變項，必須注意的是，當這p個變項增加或減少一些變項後，這原來的q個最佳組合便未必是最佳組合。再者，若X1在三個預測變項的情形下，它是最重要的（不論是用任何的係數來決定），但當變項變爲四個時，雖仍擁有同樣的係數，但有可能相對地X1已變爲不重要了。由此可知，變項最佳組合與其各個變項相對的重要性與該研究所包含的全部變項有密切的關係。

(二)樣本的特殊性

在某個樣本中，某些變項是屬最佳的組合且各自有其相對的重要性，但換成另一個樣本，情形未必是如此。亦即適合甲樣本迴歸公式的，未必適合乙樣本，這就牽涉到推論的問題。當樣本人數與變數數目之比例愈大（如60：1比5：1要大），則應用上的問題較少，但並不保證沒有問題，有一點要注意的是，須檢視另一個樣本中，各變項間相關的型態（pattern）是否與原來樣本的變項間相關型態是否近似，愈近似則推論愈有效。研究者指出可用統計的方法如再抽樣（resampling）的策略，如jackknifing或bootstrapping，使所得的結果能應用到其他的樣本上。

(三)資料的探究

研究者必須仔細地檢視資料，在迴歸分析中，界外值（outlier）對分析結果頗具影響，應審慎處理。再者，在選取變項與排序其重要性時，亦得依據常識與價值標準來作判斷（Tatsuoka, 1982）。Gordon（1968, p.614）曾說過一件關於利用常識的事情。設若以乳牛的特性、人力與農場面積三個變項來預估乳酪廠的收入，即使分析結果，顯示了乳牛不是重要的變項，應該沒有任何人會認爲乳牛與乳酪廠的收入沒有關聯，任何有常識的人，應該都會了解這一點。

參考書目

林清山（民69）。多變項分析統計法，台北：東華。

Achen, C. H. (1982). Interpreting and using regression. Beverly hills:Sage.

Afifi, A. A. & Clark, V. (1990). Computer-aided multivariate analysis. 2nd ed., New York: Chapman & Hall.

Bendel, R. B. & Afifi, A. A. (1977). Comparison of stopping rules in forward stepwise regression. Journal of the American Statistical Association, 72, 46-53.

Cliff, N. (1987). Analyzing multivariate data. New York: Harcourt Brace Jovanovich.

Cooley, W. W., & Lohnes, P. R. (1971). Multivariate data analysis. New York : Wiley.

Gordon, R. A. (1968). Issues in multiple regression. The American Journal of Sociology, 73, 592-616.

Huberty, C. J. (1989). Problems with stepwise methods-Better alter-natives. In B. Thompson (Ed.), Advances in Social Science Methodology, (Vol. 1, pp.43-70). Greenwich: JAI Press Inc.

Kerlinger, F. W., & Pedhazur, E.J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Levine, M. S. (1977). Canonical analysis and factor comparison. Beverly Hills: Sage.

Mallows, C. L. (1973). Some comments on Cp. Technometrics, 15, 661-676.

Marascuilo, L. A., & Serlin, R. C. (1988). Statistical methods for the social and behavioral sciences. New York: W. H. Freeman.

Meredith, W. (1964). Canonical correlation with fallible data. Psychometrika, 29, 55-65.

Pedhazur, E. J. (1982). Multiple regression in behavioral research:Explanation and prediction. 2nd ed., New York: Holt, Rinehart and Winston.

Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researchers. In B. Thompson(Ed.), Advances in Educatonal Research: Substantive Findings, Methodological Developments (Vol. 1, pp. 99-105). Greenwich: JAI Press Inc.

Tatsuoka, M. M. (1982). Statistical methods. In H. E. Mitzel(Ed.), Encyclopedia of educational research (pp. 1780-1808). New York: Free Press.

Thompson, B. (1988, November). Common methodology mistakes in dissertations : Improving dissertation quality. Paper presented at the annual meeting of the Mid-South Educational Research Association, Louisville. (ERIC Document Reproduction Service No. ED 301 595).

Thompson, B (1989). Why won't stepwise methods die? Measurement and Evaluation in Counseling and Development, 21, 146-148.

Thompson, B. (1990). Don't forget the structure coefficients. Measurement and Evaluation in Counseling and Development, 22, 178-180.

Thompson, B. (1992, April). Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897).

Thompson, B. (1994, April). Common methodology mistakes in dissertation, revisited. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 368 771).

Thompson, B. (1995, April). Stepwise regression and stepwise discriminant analysis need not to apply. Paper Presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 382 635).

Thompson, B., & Borrello, G. M. (1995). The importance of structure coefficients in regression research. Educational and psychological Measurement, 45, 203-209.

Thorndike, R. M. (1978). Correlational procedures for research. New York: Gardner.

Wilkison, L.(1979). Tests of significance in stepwise regression. Psychological Bulletin, 86, 168-174.

中國人

大同文教基金會

究竟什麼是「中國人」呢？

根據我們老祖宗的解釋：「中」是「禮義」，「國」是「邦」，中國就是「禮儀之邦」。

可是，外國人對我們不作這樣解釋。西方人稱我們是China man。在亞洲我們的東鄰日本，卻稱我們叫「支那」。其實China或「支那」，都是「秦」的意思—中國是在「秦」時纔完成大一統的—姑不論這China或「支那」是友善的稱呼，或是輕蔑的俚語，都不能改變我們，因為我們血管中流著的是綿延不斷的炎黃血液，我們所秉承的是五千年精深博大的歷史精神。

中國和外國人怎樣區別呢？在西方往往是以膚色、語言和吃食來區分各種民族的，但在中國人就不同了，因為東方的日本人、韓國人、越南人等等，都是黃皮膚、黑眼睛，東南亞和東北亞各個國家的人民，更是以稻米作為日常的主要食物，所以以膚色和食物來分別中國人和外國人是不對的。我們常看到一些不是黃皮膚、黑眼睛的中國人，譬如我國新疆省的前民意代表立法委員阿伯都拉先生，他有著高鼻子、藍眼睛，乍看起來，你一定會以為他是美國人或歐洲人的；但是除了長相像外國人以外，無論在言行、生活和思想上的表現，他都是一個不折不扣、道道地地的中國人。阿伯都拉說：

「我是新疆伊寧人，我父親生在那兒長在那兒，我母親生在那兒長在那兒，我也生在那兒長在那兒，所以我是道道地地的中國人。要知道中國人是不以膚色來區別中國人和外國人的，而是以他的思想和生活方式來區分的。」

阿伯拉先生說得很對。因為中國的土地幅員廣大，吸收包容外來文化的能力又特別強，所以有許多外國人，在中國待久了，就被中國「同化」變成中國人了。例如在唐朝的時候，有波斯人居住在河南、福建；明朝時還有大量的猶太人移民到中國來，但是到現在，根本就看不見波斯人和猶太人了。這時的波斯人、猶太人，並不是被消滅掉了，而是被我們悠久博大的文化「同化」成中國人了。所以中國人並不是以膚色長相來區分中國人和外國人的，而是以思想、言行和生活習慣來區別的。

在這裡，再談一談趙麗蓮博士的一段小故事。趙麗蓮博士在台灣，可說是「桃李滿天下」。從外形上看，她雖然有著外國人的血統成份，但是從她的言行思想，都是十足的中國人。譬如有一天有位姓李的學生去求她幫忙，趙博士和藹地問她的學生：

「我能辦得到嗎？你說說看？」

「小事情，您一定辦得到的，」姓李的學生說：「我這次分發在苗栗一所偏遠的山地國中，搭車時會很不方便；而王××卻分發在後龍鎮，他祇是一個人，我想請老師能夠幫忙說服他，和我對調一下工作崗位。」

「他會同意嗎？」

「我想，祇要老師開口，他一定會同意，因為他最敬佩老師了。」

「這個，我不能答應你，」趙博士委婉地拒絕了：「既然那個地方很偏僻，你都不願意去，他也不會願意去的，將心比心呀！你忘了中國一句古話嗎？『己所不欲，勿施於人』呀！」

趙麗蓮博士的母親是德國人，可是從上面這個小故事裡可顯示出，她的思想言行，十足地表現了中國人謙恭和睦的處世態度。

從古至今，有著外國血統的中國人，實在很多：譬如唐朝最了不起的大詩人李白，和專門寫長詩的元微之，都不是漢人；有「大將軍」之稱的名將郭子儀也不是漢族人；唐朝的[illegible]，[illegible]是土耳其人，但是他們思想和行為的表現，都是忠心耿耿百分之百的中國人。

所以，大體上說來，任何一個人，只要他的思想、行為和他的生活方式、處世方法，都非常的中國化，就可以說是「中國人」了。也就是說，一個人的思想行為合乎中國的傳統思想和生活，才是真正的中國人。