Bull. Nat'l Taiwan Nor. Univ.
Vol. 37. pp. 133-163 (1992)

師大學報
第37期　第133-163頁(民國81年)

共變數分析功能、假設及使用之限制

范德鑫

教育心理與輔導系

摘　　要

本文共分爲六部分，第一部分敘述共變數分析的理論模式；第二部分說明共變數分析的兩個主要功能 — 消除系統的偏差和增加實驗的精確性；第三部分陳述共變數分析所依據之假設和違反假設所造成的影響，同時也提出部分因應之策略。共變數分析所依據之主要假設包括：(1)隨機分派(2)共變量不受處理影響(3)共變量爲固定且測量無誤差(4)共變量與依變數呈直線關係(5)迴歸斜率相等(6)有關實驗誤差之假設（包括獨立性，變異同質性及常態性）；第四部分討論使用原樣團體(intact group)時應該注意的事項；第五部分指出使用共變數分析應留意的其他問題，最末部分指出研究者使用共變數分析應持之態度。共變數分析自創用至今已超過半世紀之久，由於它具有統計控制和增進研究之精確性的功能，因而爲研究者所樂用。由於現有統計書籍大都強調如何計算而忽略討論此方法使用的限制，以致造成研究者經常誤用、濫用之情事。

本文主要目的在於提醒研究者，共變數分析是一個相當「脆弱」的工具，使用時應小心爲要。也只有對此法之功能與限制透徹了解，研究結果才有意義。

1932年Fisher創用共變數分析方法至今，整整近一甲子的歷史，雖然此法目前已廣為國內外研究者所使用，但是目前它仍是一種易為研究者誤解和誤用的統計方法(Porter & Raudenbush, 1987)，難怪Elashoff (1967)說它是一種脆弱的工具(delicate instrument)。目前國內統計書雖然都有談及這種方法。可惜的是，作者大都把此方法視為變異數分析之延伸，强調如何計算，但對其功用與假設不是未交代淸楚，就是根本未提。研究者在一知半解的情況下使用此種方法進行資料分析，研究的結果令人懷疑，更遑論有多少價值了。本文主要目的在討論共變數分析之理論依據、功能、所需之假設及其他一些使用此統計方法時應注意之事項。希望它有助於研究者對此方法的正確了解與應用。至於計算方法，讀者可以參閱林淸山（民81）心理與教育統計學；朱經明（民78）教育統計學，或謝廣全（民73）最新實用心理與教育統計學。

一、共變數分析的理論模式

在檢定有關母數不偏估計數之假設和計算母數不偏估計數時，為了增加設計的精確性，通常需將因實驗誤差造成的變異減到最小。這些減少無法控制的變異來源的方法一般說來有兩種：(1)實驗的或直接的誤差控制(2)統計的或間接的誤差控制。在變異數分析中將受試者隨機分派至各實驗組；對所有受試者將某一變數維持恆常；將某一變異來源列為研究中之一因子；將受試者區組或分層使成為同質群體等都是使用直接控制的例子。這些方式的設計能減少額外的誤差變異和產生較精確的假設檢定。

在有些研究中，除了依變數外，我們也對與依變數有關的變數加以測量，利用這種方法提供資料，對依變數作統計上的調整，這個增加的變數通常稱為相伴變數(concomitant variable)或共變量(covariate)。而這種設計分析就稱為共變數分析(analysis of covariance, ANCOVA)。

Fisher (1934)指出，共變數分析兼含迴歸與變異數分析兩個廣泛為研究者所用的統計方法的優點，也需要此兩種方法的假設。Scheffe氏(1959)曾對此三種統計方法作簡單的區分，他認為：一個實驗的因子可以是質的（如教學方法），也以是量的（如閱讀時間）。在量的因子中，若不在乎其實際的值，我們可將它分成不同的水準，這種情況下，量的因子亦可視為質的因子。他並認為在變異數分析中，所有的因子都是以質來處理，而在迴歸分析中，所有的因子是屬於量的，且以量來處理。既然共變數分析組合了上述兩個模式，有些因子就需以量來處理，而有些則以質來處理。即，共變量之測量值是量的，且以量來處理。作為自變數的因子則以質來處理。若能檢視一般線性模式之設計矩陣，就更容易了解共變數分析是由變異數分析與迴歸分析組合而成的了。

統合迴歸分析和變異數分析，可使研究者在既定的機率下，確定是否依變數對共變量的迴歸可以解釋所有處理組依變數間之變異。換言之，共變數分析方法可用以檢定當共變量可預測的變異部分扣除以後依變數之平均數。它是利用統計控制來調整依變數之平均數，以達成減少無法控制的變異，同時增加設計的精確性和估計值的有效性的方法。然而使用這種方法並非漫無限制的，爲了檢定統計數能呈必要的分配，分析能有意義，結果能夠解釋，某些假設必須符合，某些條件必須具備。讀者如果把共變數分析看成是依變數調整後的平均數的變異數分析，也許更能了解共變數分析的性質了。對此統計方法雖然不應誤用，但該用時也不能不用。如果對下面的模式及假設有正確的了解，是否適用於特定的研究問題能作批判性評鑑，則共變數分析就不是一個危險且易引起誤解的統計方法了 (Hope, 1968)。

雖然共變數分析可用於兩個或兩個以上的共變量，然而，爲了減少複雜符號的困擾，在此使用簡單一因子固定的分析來說明。事實上，使用一個共變量或多個共變量，基本概念是完全相同的。如果共變量用 X 來表示，依變數以 Y 來表示，則共變數分析模式可寫成：

$$Y_{ij} = u + \alpha_j + \beta\ (X_{ij} - \overline{X}..) + e_{ij}\quad (i = 1,2,\cdots\cdots n_j;\ j = 1,2,\cdots\cdots J)$$

在此，Y_{ij}：第 j 組第 i 個人的依變數值。

u：處理母群總平均數，是一未知常數。

α_j：第 j 組處理效果，第 j 處理組平均數與總平均數之差 $(u_j - u)$。

β： Y 對 X 之迴歸係數。

X_{ij}：第 j 組第 i 個人之共變量值。

$\overline{X}..$：所有受試共變量之平均數。

e_{ij}：第 j 組第 i 個人之實驗誤差或無法控制之變異。

Elashoff (1969) 曾指出，變異數分析模式各個的變異 e_{ij} 在共變數分析模式中分成下面兩部份：⑴ $\beta\ (X_{ij} - \overline{X}..)$，由於 Y 在 X 上直線迴歸的變異⑵ e_{ij} 無法解釋的變異。因此，如果 X 和 Y 呈直線關係時，則在共變數分析中無法解釋的變異顯然地小於變異數分析中之無法解釋的變異。這一點正可說明了共變數分析往往較變異數分析有檢定力的原因。

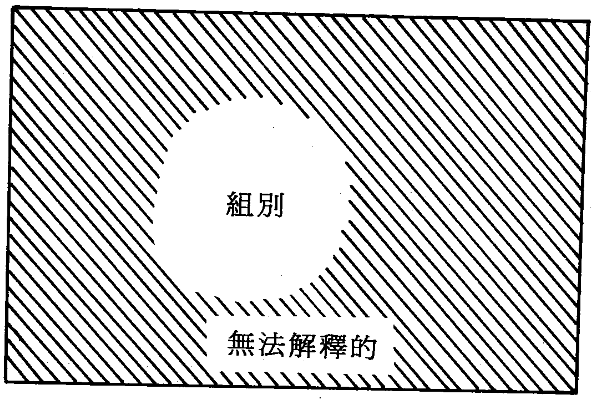
二、共變數分析的功能

按 Snedecor 和 Cochran (1989) 的說法，共變數分析具有下列四種功能：⑴增加隨機化實驗的精確性。⑵調整誤差的來源⑶使隨機實驗中處理效果的性質更加清楚⑷研究多重分數的迴歸。但其中最重要的兩個功能爲：⑴消除系統的偏差⑵減少組內的誤差變異 (Elashoff, 1969; Stevens, 1990)。所謂系統誤差是指與依變數表現有關之最主

要的變數，各組呈現系統差異。如果這種現象存在，各組接受不同處理之後，依變數有顯著差異，除了歸因於不同實驗處理外，實驗開始時，各組在這主要變數不相等可能也是原因之一。使用共變數分析可針對這些主要變項而調整依變數，以校正開始時各組之間的差異。我們現在用一因子設計來加以說明此兩種功能，在此雖然用一因子爲例，但這種概念卻是極易類推到較複雜的情境。

假設某研究者對兩種補强經驗對閱讀成就之影響感到興趣，於是使用包含三種處理之設計，處理一的受試者接受傳統的教學；處理二的受試者除了傳統教學外，每天還接受20分鐘的額外教學；處理三的受試者除了一般教學外，還接受電腦補助教學。研究者使用亂數表從界定的母群中隨機選取30人，再將此30人隨機分派至3組，每組10人，最後隨機將3組分派至各實驗處理。然後按計劃實施，最後對所有30位學生施測閱讀測驗。如果研究者使用變異數分析，所獲得的依變數之變異可分成下面兩部分：(1)可以由自變數（處理組）加以解釋的。(2)無法解釋（誤差）的部分。在這種情況下，組內之差異被視爲無法解釋變異。也許我們可以用圖一來說明：整個長方形表示閱讀分數的變異，斜線的部分表示無法解釋（誤差）的變異。

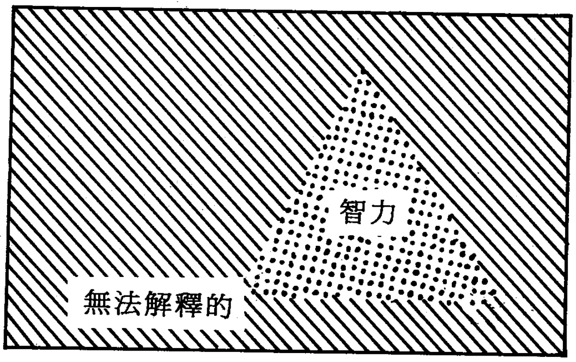
閱讀分數之變異



圖一　閱讀分數變異中可由組別（處理）加以解釋的和剩餘無法解釋的變異

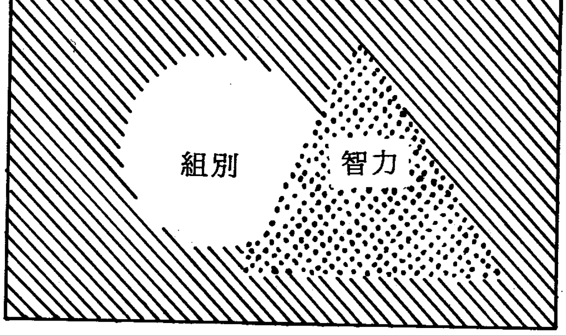
假設研究者在這些學生資料卡中能找到他們的智力分數，同時知道智力測驗的分數和閱讀分數有中度的相關。我們又可以再用圖二來描述。整個長方形仍表示閱讀分數的變異，三角形表示可以用智力測驗分數之變異加以解釋的變異部分。既然知道智力測驗的分數，那麼使用共變數分析應較變異數分析爲佳。

閱讀分數之變異



圖二 閱讀分數變異中可由智力(IQ)
加以解釋的和剩餘無法解釋的變異

共變數分析合併了圖一與圖二所敘述的資料，也就是利用兩個自變數（組別與智力），閱讀成就的變異將分成組別、智力分數和剩餘的誤差三部分，如圖三所示。前述共變數分析之兩種功能皆可利用圖三加以說明。首先，注意到組別與智力分數相重疊（以點表示）部分，它表示組間之變異中可由智力來解釋的部分。因而與組別有關之閱讀分數之變異量減少了。這個特性正好說明了第一個功能，即，爲了某相伴變數（在此爲智力）而調整依變數的分數以達到校正開始時組間的差異。第二個功能則可透過檢視剩下的組別可解釋的變異和無法解釋的變異之間的關係來說明。與組別有關之絕對變異量雖然減少了，但誤差（殘差）變異也因相伴變數之介入而大大減少。因爲誤差較小，故檢定力隨之增加，因而所有假設滿足時使用變異數分析未達顯著之資料，若用共變數分析就可能達到顯著。雖然在某些情況下，也有相反情形，但這情形少之又少 (Huitema, 1980)。



圖三 閱讀分數之變異分割成 (1) 智力可解釋的
變異 (2) 可由組別（處理）解釋附加的
變異，以及 (3) 剩餘的無法解釋的變異

過去有些學者（如Horst, Tallmadge & Wood, 1974; Backer & Engelmann, 1976) 誤認為：如果兩組前測分數相同，就不需要使用計算繁複的共變數分析。國內犯此同樣錯誤觀念的人也不少，國內許多統計書只談及當研究的受試是屬原樣團體時，可用此方法來達到統計控制偏差的效果。這種說法，似乎告訴讀者使用共變數分析只能校正實驗初始時各組不等之共變量。事實上，減少誤差變異，增加研究之精確性是共變數分析兩個重要的功能。 Huck (1972) 指出：各組共變量沒有差異，使用共變數分析仍是極其有用。試看下面兩組之處理資料就可說明共變數分析方法的確有增加研究精確性的功能。兩組之原始資料如下：

	實驗組		控制組	
	前測	後測	前測	後測
	63	70	56	59
	64	76	75	73
	33	47	67	66
	67	73	63	69
	81	85	70	69
	69	77	59	66
	72	77	34	44
	53	68	61	62
	74	82	78	79
	77	80	52	53
	57	66	82	78
	62	67	86	80
	58	71	62	65
	60	68	66	69
	85	86	73	74
平均數	65.00	72.87	65.6	67.07

經檢定發現：此資料各組之前測平均數並無顯著差異 ($F = 0.016$)，若按 Horst, Tallmadge 和 Wood (1974) 的說法，使用變異數分析方法處理即可。用變異數分析結果發現依變數並未達顯著的差異 ($F = 2.66$， $P > .05$)，研究者的結論為：實驗處理沒有效果。倘若使用共變數分析，其結果可能就不一樣了。

對上面的資料在進行共變數分析之前，通常要先進行迴歸係數同質假設之考驗，因 $F = 0.001$，所以共同的組內迴歸係數 (β_w) 存在假設是符合的。在抽出前測分數可以解釋的變異之後，再檢定後測各組平均的假設，結果與前面使用變異數分析之結果完全不同 ($F = 50.56$， $P < .05$)。用共變數分析，所得的結論是：實驗組與控制組有差異。使用不

同的統計方法，結果卻迴異，原因在共變量減少了組內無法解釋的變異所致。這個例子說明了共變量分析確實具有增加精確性的功能。

其實，處理系統偏差最好方法是隨機分派受試者到各處理組。除了誤差因素外，我們才可以說在任何一個變數上各組沒有系統差異。然而，許多的研究我們無法進行隨機分派，在這種情況下，我們可以使各組在某些重要變數上相等，針對主要變數加以配對。當然，這種情況下，我們只能說在這些配對變數上各組是相等的。共變數分析是控制主要變數的統計方法，它和配對方法一樣，只能減少偏差，而無法完全消除偏差 (Stevens, 1990)。至於減少誤差變異，也有許多方法，例如選取較爲同質的受試者，這樣可使依變數之變異較小；有時也可以用多因子設計的方式達成，使用共變數分析只是其中的一種方法。在變異數分析中，檢定統計數 $F = MS_b/MS_w$。式中 MS_w 就是誤差估計數。如果我們能使 MS_w 減小，則 F 比率將增大，就可獲得較敏感或檢定力較強的檢定了。如前所述，在共變數分析中，共變量通常具有使組內變異減小的功能，在共變量不只一個的設計，則要考慮使用相互之間相關低，但各個與依變數相關高的共變量，這樣各個共變量可以移去依變數誤差變異不同部分而達成使誤差變異減得更多的功能。

三、共變數分析所根據之假設和違反假設之影響

前面已經提到過，共變數分析是變異數分析與迴歸分析之合併。因此，此種統計方法除了須具備變異數分析和迴歸分析之假設外，尙應具備一些此統計方法所特有的假設。從一因子共變數分析模式 $Y_{ij} = u + \alpha_j + \beta(X_{ij} - \overline{X}..) + e_{ij}$ 可知，依變數分數是模式中各項之總和。即觀測值 (Y_{ij}) 可視爲下列四個相互獨立成分之總和(1)所有依變數觀測值之總平均 (u) (2)第 j 組處理之效果 (α_j) (3)共變量效果 [$\beta(X_{ij} - \overline{X}..)$] 和 (4)誤差 ($e_{ij}$)，這種現象就稱爲可加性 (additivity)，目前大部分統計書並未把可加性列爲變異數與共變數分析之假設，因此有關違反此假設之影響報導少之又少 (Baker, 1972)。下面所列舉之假設是幾位學者共同的觀點（例如，Elashoff, 1969; Huitman, 1980; Marascuilo & Serlin, 1988)。研究者如果想知道資料是否符合共變數分析理論模式，應先了解其所需之假設及違反假設之結果有什麼影響。若符合假設則使用固定的共變數分析就能適當地檢定或解釋無處理效果之虛無假設。

(1)隨機分派受試者至各處理組。

(2)各組共變量不受處理效果之影響。

(3)共變量測量無誤差。

(4)固定的共變量。

(5)共變量與依變數呈直線關係。

(6)每一組依變數在共變量之迴歸相同。

(7)有關實驗誤差的假設（包括獨立性，變異同質性，及常態性）

就上述各項假設分別討論如下：

(一) 隨機分派受試者至各處理組

隨機化 (randomization) 在實驗設計中是一個非常重要的名詞，特別是在眞正的實驗中，它包含了兩個程序，一是從界定的母群體中隨機抽取受試者，另一是隨機地、獨立地分派受試者至各處理組。如果實驗者不想把研究結果推廣到母群體，則隨機選取這一程序不是很重要，然而，隨機分派才能使共變數分析之F檢定與信賴區間作正確的解釋。按照 Huitema (1980) 的說法，隨機分派之重要性有二：第一，此法可使各組在所有特徵上具有相同的期望值，即是一種形成同等團體的一種方法。第二，隨機分派與共變數分析模式中誤差 (e_{ij}) 之分配有關，e_{ij}必須假設相互間是獨立的。雖然隨機分派無法保證誤差相互間獨立，但使用隨機分派比非隨機分派較有滿足此假設之可能。

未使用隨機方式分派受試者到各處理組，共變數分析F檢定和調整的平均數可能會產生偏差。相反地，若是使用隨機分派，即使共變量之變異數分析結果達到顯著，共變數分析F檢定與調整的平均數也不致於造成偏差的現象。如果實施隨機分派，但因某些原因，各受試者觀測值彼此未相互獨立，則實際犯第一類型錯誤之機率仍將大於設定的α值。如果分數之間相依關係是正向的，誤差變異可能會有低估的現象，自由度也將小於共變數分析計算過程中所指定的。關於這一點，在此舉個例子說明之。假設使用兩種教學方法教統計學，採用兩組隨機的共變數分析設計，每一組 20 位學生。使用前測當作共變量，後測當依變數。如果在第一組中 20 位學生中有 12 位是抄其他 8 人中 1 個人的答案，這表示並非所有 40 位受試的反應彼此間是相互獨立的，一位學生完全控制了其他 12 位同學的作答。這種設計，誤差項自由度爲 40 － 2 － 1（即 N － J － 1），此計算自由度的公式是把依變數中 40 個資料視爲相互獨立計算而得。然而實際上只有 40 － 12 ＝ 28 個獨立的反應，也是第一組只有 8 個獨立分數，因而誤差 (e_{ij}) 是非獨立的。未使用隨機分派或使用隨機分派，但因某些因素未達到觀測值相互獨立，這兩種相依結果都有可能造成α值的增加。

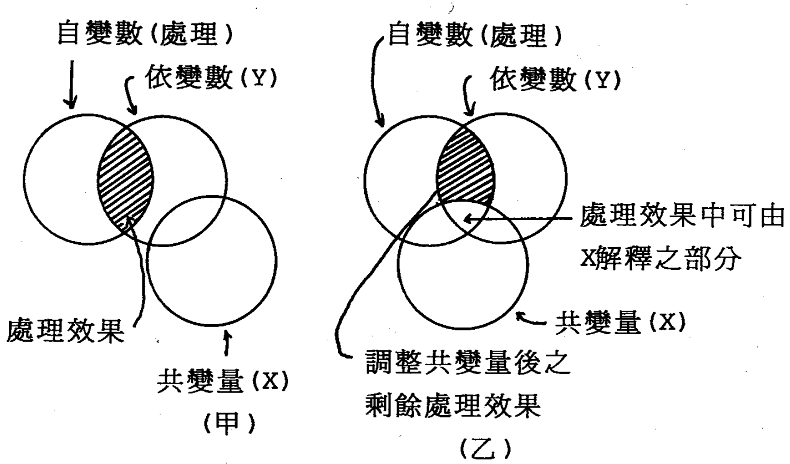
既然獨立性是一個重要的考慮因素，我們就必須知道它是否存在。不幸的是，大部分實驗設計是沒有明確的可用之獨立性檢定。在隨機組的設計例子中， Huitema (1980) 建議實驗者利用所有有關受試行爲，實驗程序，實驗環境等方面之知識來判定獨立性是否存在。若有相依性問題存在，則應重新考慮樣本和抽樣單位 (sampling unit) 之界定。就以前面兩組的例子言， 40 個受試者中只有 28 個觀測值是相互獨立的。這種情形下，分析時應刪除此 12 個特殊的觀測值，因爲它們非但無法提供新的訊息，而且會使F檢定產生偏差。遇到

這種情形，實驗者所要做的是找出可以自由變化的抽樣單位，依據這些單位來進行分析。

若非隨機組設計時，獨立性的問題就更複雜了。重覆量數的設計就有相依性的問題存在，包括了(1)幾種反應來自於同樣的受試(2)觀測值來自於不同的受試。來自於相同受試者相依問題較易鑑定。複合對稱 (compound symmetry) 假設檢定就與獨立性假設直接有關。此部分可參考 Winer (1971)。另一種設計稱爲中斷性的時間系列設計 (interrupted time-series design)，它可用來研究觀測值獨立性的問題。若讀者有興趣可參考 Winer (1971)，Box and Tiao (1975) 和 Glass 等人 (1975) 所著書籍。

(二) 各組共變量不受處理的影響

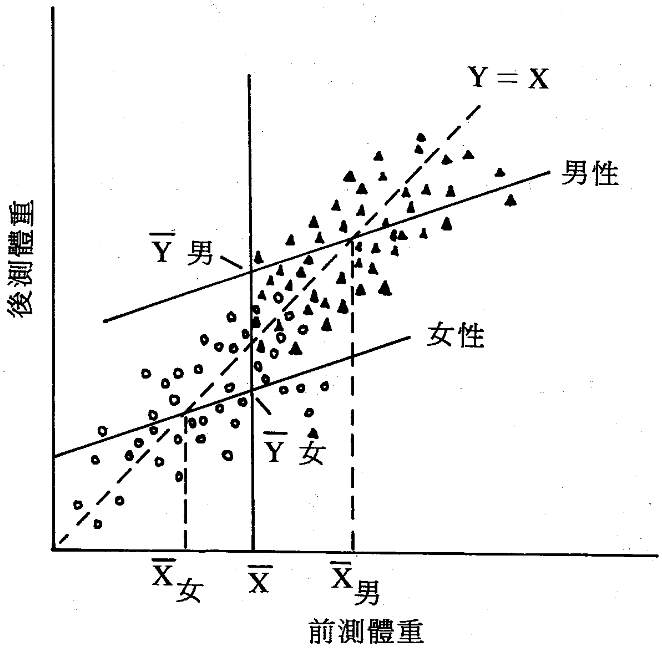
共變量和處理統計上獨立不是共變數分析模式所根據之數學假設，只是邏輯的問題而已。但是如果調整的處理平均數之解釋要有意義，則共變量不受處理影響是十分重要的。共變量若是在處理前獲得，又是隨機分派的話，是項假設則自然會符合。反之，共變量在使用處理之後才測量時，可能就有問題了，因爲處理有可能影響共變量和依變數之值。如果共變量受到處理的影響，處理平均數的差異就無法清楚明確地和處理扯上關係。換言之，處理影響共變數值，計算調整平均數時，將會移去部分處理效果。我們可以用圖四來加以說明：



圖四 處理影響共變量之結果：（甲）處理效果與共變量無關
（乙）處理效果與共變量有關（轉載自 Huitema, 1980）

Evans 和 Anastasio (1968) 亦指出：當處理與共變量相混淆時，將會違反迴歸同質性假設，處理均方之期望值將多包含了表示處理和共變量相關的一項。這種情形下，調整平均數可能有調整過度或不及之情事發生。Evan 和 Anastasio 指出相關之大小將決定違反這個假設之結果。又認爲在大部分處理影響共變量的例子中，相關可能足夠大到無法使用共變數分析。因此，在使用共變數分析方法前，先要了解共變量是否會受到處理之影響。共變量如受到處理的影響，則調整母群平均數相等之假設就毫無意

義了，因為處理與共變量之元素相互重疊之故。例如，有兩個具有不同吵雜聲的處理設計，使用共變數分析等於在比較兩個現實上不存在的處理了。 Marascuilo和Serlin (1988)也指出：共變量不能依賴自變數，假如研究者想比較男女體重減輕情形，男女兩組都予以相同之飲食，受試者在研究開始與結束各量一次體重。研究者想了解調整前測差異後，後測的體重男、女平均數之差異是否一樣。因為男人一般說來較女人重，很容易想像到男人的前測體重大於女人（即$\overline{X}_m > \overline{X}_F$），有些人可能會想到用共變數分析方法來進行分析，倘若這麼不加思索就使用那就錯了，因為共變量（前測體重）與自變數（性別）有高度之相關。在這種情形下，幾種結果會產生，至於產生何種結果，則取決於男、女各組內前測體重與性別之關係如何和後測體重與前測體重之關係如何。這種設計，若使用共變數分析，即使沒效果也會呈現效果的。假設實驗結果如圖五所示：



圖五　共變數分析不正確的使用（轉載自 Marascuilo & Serlin. 1988）

各坐標點若沿 Y = X線散佈，則表示體重沒有減輕，此種飲食對男、女兩組一樣沒有減輕之效，然而，調整平均數卻顯示不同的結果。使用共變數分析者可能會錯說：在調整依變數之後，女性組比男性組為輕，故女性組較男性組體重減輕較多。遇到這種情形， Marascuilo和Serlin (1988)建議應該使用性別×飲食二因子的變異數分析來分析資料，性別包括男、女，飲食包括規定的飲食、非規定飲食各兩個水準，同時使用減輕的體重當依變數。

(三) 共變量測量無誤差

在共變數分析模式中，e_{ij}代表某一個體所有那些影響依變數測量之無法控制的變異來源。實驗誤差是隨機變數，雖然它是所有無法控制變異的組合效果，但卻只與依變數有關。因此，共變數分析所依據之理論及導出是依據共變量測量無誤這個假設的。在共變數分析理論模式中，找不到一項是與共變量有關之隨機誤差。這似乎暗示了測量共變量的工具應該具有完全的信度。如此，所觀測的共變量實得的分數 (X) 與眞正之分數 (T) 才會一樣。在社會科學中，測量工具皆具有某種程度之不可靠性。既然具完全信度之工具不可得，就不能不注意這個假設違反的影響了。由於這個問題牽涉了測量與統計問題，在此稍費一些篇幅討論之。

根據傳統的心理測驗分數理論，一個人實得的分數爲眞正分數（無法測得）和測量之隨機誤差兩相互獨立之部份所組成。我們可以用下面之模式表示：

$$X_i = T_i + E_i$$

式中 $T_i = E(X_i)$，誤差被假設成平均數爲 0 之常態分配，由於 $Cov(T_i, E_i) = 0$，且在 E_i呈常態的假設下，T 和 E 兩個都是相互獨立的。 X 變數之母群變異數於是變成：

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

若雙邊各除以 σ_X^2，則上式變成

$$1 = \frac{\sigma_T^2}{\sigma_X^2} + \frac{\sigma_E^2}{\sigma_X^2}$$

傳統測驗理論信度定義爲眞正分數變異數與實得分數變異數之比率（即 σ_T^2 / σ_X^2）。如果測量工具百分之百可靠，則 σ_X^2應與 σ_T^2相等。測量工具之信度愈低，則誤差的變異部分增大而造成實得分數之變異增大。在這種傳統心理測驗分數理論模式下，試看信度低的工具在共變數分析中，對共變量產生何種影響。

當共變量之測量有誤差時，所產生的問題十分複雜。因爲問題涉及依變數在共變量之迴歸有關。假設依變數和共變量之測量都有誤差，依變數 Y_i之實得值可以用下式表示：

$$Y_i = y_i + d_i$$

式中 Y_i表第 i 個人之眞正分數， d_i表示測量誤差。同樣的共變量值可寫成：

$$X_i = x_i + e_i$$

所需之假設是兩種誤差間無相關即 $r_{de} = 0$，誤差與眞正分數間也無相關。

當變數有誤差存在時，有兩種可能的迴歸需要考慮。一是 Y_i在 X_i上之迴歸。另一是 y_i在 x_i上之迴歸。 Berkson (1950) 稱後者爲兩變數之結構性的關係 (structural relationship)。在應用科學的研究中， X 和 Y 大都是屬隨機變數，因此下面我們針對 X 爲隨機變數來加

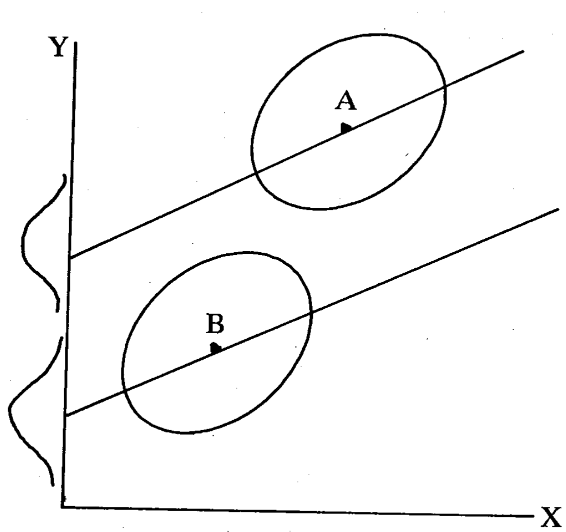
以討論。

在簡單迴歸中，大家知道，當要找出迴歸線之斜率和截距時，通常是用最小平方法來算得，即令依變數 (Y) 之殘差 (e) 平方和爲最小求得。如果兩個變數都有誤差，結構性的或眞正的關係的最適合線將會使得兩變數之誤差達到最小。 因此最小平方迴歸斜率估計數 $b_{Y.X}$是有偏差的。 Berkon 研究指出：最小平方估計數並非眞正迴歸斜率之不偏估計數，其平均値小於眞正之斜率。按照他的看法兩個斜率之間的關係應爲：

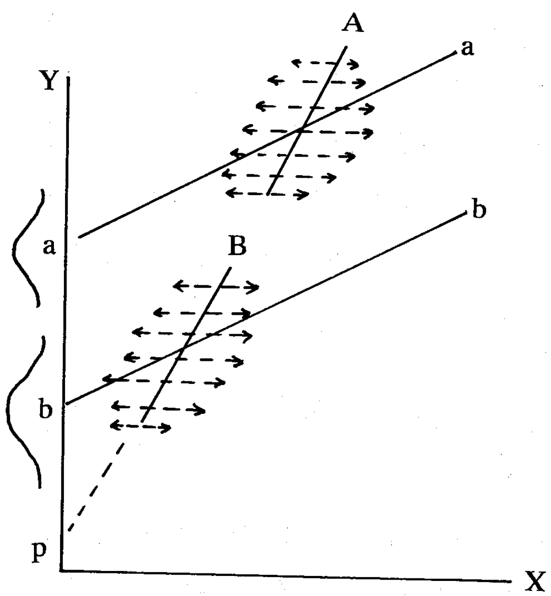
$$\beta_{Y \cdot X} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} \beta_{Y \cdot X}$$

式中 σ_T^2 爲共變量眞正分數之變異數，σ_E^2 爲共變量誤差部分之變異數。根據前面所述，已知道心理測驗分數理論之基本假設爲 $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$，信度定義成：眞正的變異與實得變異之比率。從公式得之，偏差的程度決定於共變量之信度。 Berkson (1950) 和 Scheffe (1959) 曾指出，X 是固定的有誤差的變數，則斜率之最小平方估計數就是結構性關係的不偏估計數。然而，在社會科學中的研究 X 很少是固定的。

由於最小平方斜率估計數可使縱坐標（依變數）的誤差減至最小，所以上述的偏差與依變數之誤差無關。 Mandansky (1959) 指出，斜率的偏差問題所以會發生，是因爲也有橫坐標（共變量）的誤差存在，這在估計 $\beta_{Y \cdot X}$ 時必須考慮的。同時他也指出，當誤差假設爲常態分配時，兩個有誤差存在的變數資料結構性關係是無法確定的。 Lord (1960) 也表示，當共變量有誤差存在時，一般所用之共變數分析方法是無法正確地調整研究開始時兩組之差異。圖六和圖七也許能幫助說明共變數誤差對調整平均數之檢定的可能影響。

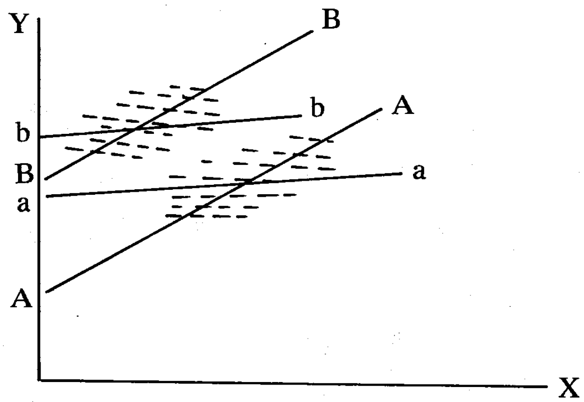


圖六　兩處理組共變數分析模式



圖七　因共變量有誤差造成截距的差異

圖六說明一般的變異數分析情境，A，B是反映兩個處理觀測值之分散圖，迴歸斜率是以最小平方法求得。Lord假設X和Y是遵照組內雙變量常態分配，如此，次數形成了兩個常態分配。因爲兩組並未有重疊，因此說調整平均數之檢定達到顯著。圖七中，結構性關係的迴歸線以A與B表示。由於在此兩變數是完全的相關，因而兩迴歸線是一致的，兩組的截距也相同，皆爲P。正確的結論是：一旦y在x上的迴歸移除之後，兩組間之差異就未達顯著。點狀線表示共變量之測量介入誤差之效果。共變量值此時呈水平方向之分佈，且形成的分散圖與圖六相似，y在誤差的共變量(X)之迴歸形成a，b兩線（和圖六一樣），截距是有顯著差異的。顯然地，某些資料使用共變數分析發現調整平均數達到顯著的差異，然而檢視結構性的關係，則差異就不存在了。同樣的，共變量測量如有誤差時，共變數分析有時也無法偵測到統計上地顯著差異。如圖八所示：眞正的迴歸線A，B之截距有差異，一旦共變量有了測量誤差則此差異就減少了。即[(B－A)＞(b－a)]。



圖八　因共變量有誤差造成截距之減小

上面討論的兩個例子，是共變數分析用於非隨機實驗之結果。Cochran (1968)指出，如果有誤差的共變量(X)各組平均數相同時，則斜率的最小平方估計數就不致於影響調整平均數之檢定。在隨機的實驗中，共變量的平均數理論上應該相等，或近於相等。在這種情況下，使用有誤差的共變量之唯一影響是：依變數的誤差變異數無法減少很多。Cochran (1975)和Gourlay (1953)指出，如果σ_e^2表示變異數分析中之實驗誤差之變異數，共變數介入之後，變異值($\sigma_{e'}$)之公式爲：

$$\sigma_{e'}^2=\sigma_e^2\ (1-\rho^2)(1+\frac{1}{f_e-2})$$

此式中之f_e表示誤差之自由度數，Cochran (1968)又導出了另一種公式：

$\sigma_{e'}^2=\sigma_e^2\ (1-\rho^2 r_{XX}\ r_{YY})$，式中$r_{XX}$爲共變量之信度估計值，$r_{YY}$爲依變數之信度

估計值。雖然 Cochran 在此未說明自由度之損失情形，但主要目的在於說明：在隨機實驗中，共變量有測量誤差存在只會使得檢定不那麼敏感。

然而，如前面所述，若使用共變數分析於原樣團體將可能造成低估或高估調整平均數之情形，這是因為使用原樣團體，共變量之平均數很容易有所不同。因此，如 Cochran (1968) 指出：有誤差的變數調整平均數之期望值和無誤差情況下的調整平均數之差異並不相等。 Kahneman (1965) 研究發現，當各組共變量之平均數眞正有差異時，共變量之誤差對合併的組內變異數影響大於組間的變異數；使得組間的迴歸線比合併組內的迴歸線要陡峭。因此，他認為實際犯第一類型錯誤率要大於設定的水準，且隨樣本數之增加和信度之減小而增加。因此共變數分析用於準實驗設計，當共變量具有測量誤差時，應遵循何種方法是值得研究的。

Lord (1960) 是第一位提出可用於有誤差之共變量和原樣團體設計方法者，不過他的方法限使用於兩個處理組和大樣本（用在大樣本才能導出 $u_{Y1} - u_{Y2}$ 之不偏樣本估計值）。為了計算共變量之信度估計值，需要對共變量測量兩次。他發展的檢定統計數被用來和單值常態曲線 (unit normal curve) 值相比較，以確定顯著性。可惜的是，他並未告訴使用這種方法要多大的樣本才適當，而且只能用於兩組情境也是一大限制。 Porter (1967) 曾檢視樣本數，共變量的信度， X 和 Y 的相關對 Lord 所提出的統計數分配的影響。結果發現此統計數輻合常態分配的速度取決於這三種變數。共變量與依變數的相關對輻合之速度有反向的影響。當共變量信度只有 .50 時， Lord 的統計數未能十分近於常態分配，因此，在使用 Lord 的方法前，必須考慮這些變數。

Porter (1967) 曾發展一種方法，以估計的眞正共變量分數替代實得共變量分數。結果顯示：調整組內平方和與實得分數分析而得的相同，但全部的和調整的組間平方和卻不相同，原因在於斜率之最小平方估計數改變了。 Porter 發現，即使信度估計數是來自於樣本資料， F 分配仍可以使用。如果有了共變量之信度估計值，則眞正分數可以用下面公式來估計：

$$T = \overline{X} + r_{XX}\ (X - \overline{X}\ )\ 或\ T = r_{XX}\ X + \overline{X}\ \ (1 - r_{XX}\ )$$

從此公式可知，眞正分數是實得分數的直線轉換，估計的眞正分數之平均數會等於實得分數之平均數，估計的眞正分數之變異數為信度的平方乘以實得分數。而且直線轉換不致於影響到 X 和 Y 的相關。因此若使用估計眞正分數，σ_e^2 減少量與前面提到的因共變量之調整而使得 σ_e^2減小之公式（見前頁）求得者是一樣的。

Porter 曾進行模擬研究，研究中，他使用各種的樣本數、共變量的信度、依變數與共變量的相關，結果發現：每一處理組樣本數至少要 20 才能使估計的眞正分數用於 F 分配。又當共變量信度只有 .5 時，使用四個處理組之實徵分配比 .7 和 .9 者顯現更大的差距。而且，使用四個處理組，隨著依變數與共變量相關之增加，實徵的與理論

的分配之一致性隨之減少。除了兩組的情況外，這種影響是非常有系統的。使用估計的眞正分數能校正估計結構性關係斜率之偏差。因此，當共變量不可靠時，Porter的方法頗適合使用。

因此，當共變量未具完全信度時，能否使用共變數分析呢？讀者看完上述之後，對此問題一定感到非常納悶，這是在所難免。本來這個領域就比較複雜，雖然了解使用這種統計方法，有許多問題存在，卻沒有人能淸楚說出使用有誤差的共變量將有何種後果。也沒有人能提出其他替代的方法。我們認爲當使用此法於原樣團體之設計時，回憶一下，使用最小平方之迴歸斜率估計可能造成的錯誤結論恐怕是一項明智之舉。

(四) 共變量爲固定的變數

此假設與共變量測量無誤之假設都是一般教科書比較不談的。Scheffe (1959)指出，爲使能與變異數分析一般理論相一致，在共變數分析模式中，隨機變數爲 $\{Y_{ij}\}$ 和 $\{e_{ij}\}$，即依變數和誤差變數。因此共變量的觀測值假設是實驗者選定者。例如，在某一固定的藥量、動物的活動即是。實驗者把藥納入作爲共變量，其值是由實驗者加以選定的，且以數量來表示。雖然共變量之固定性質可以用於物理或生理科學，但在社會科學中，共變量觀測值幾乎都是隨機變數值。僅就使用的設計和變數之性質言，共變量爲固定變數之假設經常是無法滿足的。爲了使用隨機變數作爲共變量，伴隨變數之觀測值上之條件分配誤差呈常態，變異數相等兩假設必須符合（關於這兩個假設將在下面討論）。雖然假設檢定之分配理論也是條件的，但如果所有共變量值假設都能符合，則顯著水準對非條件或固定的變數情況是相同的。

共變量經常是隨機而非固定的變數，因此，有必要去了解違反此假設對分析將有何種影響。Atquillah (1964)曾提出一種情境分析檢視方法。當共變量之觀測值爲隨機變數，且能假設抽自常態的母群體時，則F分配的期望值、變異數和共變量是固定的情境是完全一樣。

(五) 共變量與依變數呈直線關係（迴歸直線性）

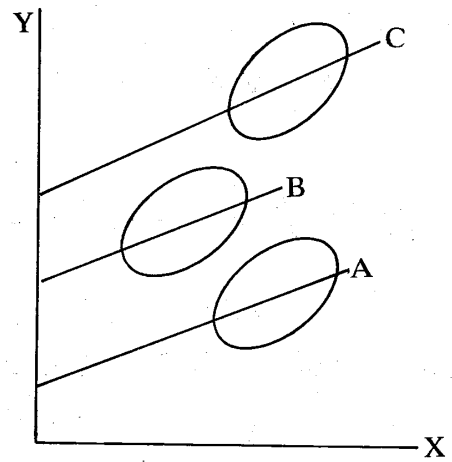
在確定適合資料的模式時，研究者必須確定共變量和依變數間關係的正確形態。共變數分析模式需假設所有處理母群之共變量與依變數有直線的關係。Elashoff (1969)曾指出：此假設是否滿足，可以根據理論或先前實驗之結果，或根據X,Y分散圖來分析。她認爲也可以對 $\beta = 0$ 進行統計檢定，如果拒絕了，則表示共變量和依變數有顯著之直線關係。

線性假設最常違反是發生在二次關係較適合代表資料時。Cochran (1957)指出，當受試者與處理是隨機分派時，顯著性檢定並未受到嚴重之影響。由於使用共變數分

析之設計大都屬非隨機的設計，因此處理之共變量經常不相同，這種情形，使用不正確之迴歸形式，可能導致變異數之異質，也會導致調整平均數解釋上的困難。如果錯誤使用直線模式，且X值各組間有很大的差別，則調整部分 $\beta\ (X_{ij}-\overline{X}..)$ 將導致各組不等之變異數，也會減損假設檢定之敏感度。 Atquillah (1964)研究直線假設違反，調整平均數所產生的問題。他檢視了當X和Y的眞正關係爲二次時，在直線模式下，兩處理組平均數和多組平均數之情形。他發現在兩樣本之情況， $E(\hat{\alpha}_1-\hat{\alpha}_2)$ 或估計平均數差異之期望值是有偏差的，除非X值能假設是來自於相同之常態母群。既然一般研究經常未使用隨機分派，處理之平均數差異就會產生解釋上的問題。至於兩處理組以上之分析就更複雜了。在組數趨於無限大下，均方之期望值將產生嚴重偏差現象，造成偏差之總F檢定和處理平均數差異解釋上的困難。即使X觀測值是抽自相同的常態母群，也有上述現象發生。如果使用迴歸的正確形式，偏差的程度決定於二次項的係數。因此， Cochran的說法在某些狀況下不一定正確，研究者使用直線迴歸模式時經常未考慮有關共變量與依變數之眞正關係。上述假設違反的結果正告訴我們在使用共變數分析之前應多考慮所使用迴歸之形式是否正確。

(六) 共同的迴歸斜率

在一般共變數分析模式 $(Y_{ij}=u+\alpha_j+\beta\ (X_{ij}-\overline{X}..)+e_{ij})$ 中，除了迴歸直線假設以外，$\beta\ (X_{ij}-\overline{X}..)$ 這一項除了抽樣誤差以外也需要有各組迴歸線之斜率都是相同的假設。也就是說，沒有處理與斜率之交互作用，從下圖九就可以了解一般。圖九中迴歸線是相互平行的（即斜率不變），每一X水準都是處理C最好，都超過固定的數量。



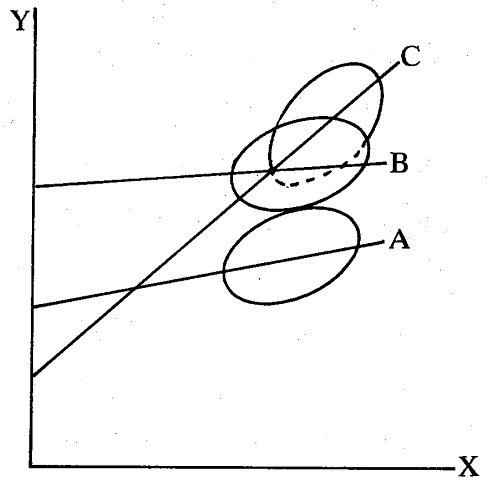
圖九　迴歸線相互平行

反之，圖十之斜率不相等，形成非平行的迴歸線。在這種情況下，雖然平均而言，處理C最爲有效，但顯然地，並非所有X之水準都最好，這種現象就是一般所謂的

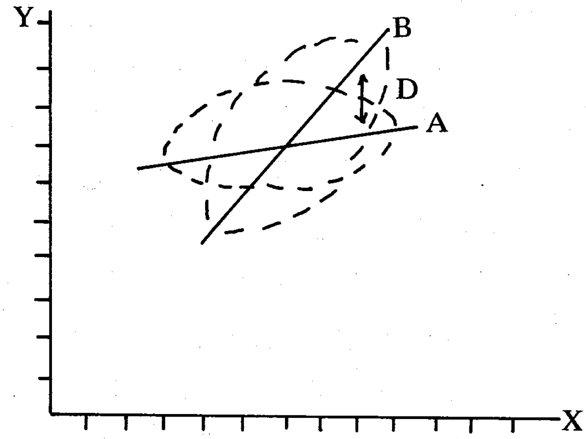
處理與共變量間之交互作用。Feldt (1958)研究發現如果把共變量作爲變異數中之區組變數，則各組迴歸線之不同等於處理和區組變數水準之顯著交互作用。表示圖十之模式，應是：

$$Y_{ij}=u+\alpha_j+\beta_j(X_{ij}-\overline{X}..)+e_{ij}\quad (i=1,\ \ldots\ n,\ j=1,2,3)$$

式中迴歸係數有足標，目的在表示各組斜率不相同。如果這個模式適合資料，則不宜使用具有同斜率之迴歸線。Feldt (1958)主張，當各組斜率有顯著不同，所求得之誤差變異數可能有高估眞正誤差變異數的現象，因而造成減損假設之檢定力。Glass 等人(1972)也指出，如果有相當的理由懷疑迴歸斜率不相等，則宜使用共變數當作區組變數之多因子變異數分析設計。倘若只有兩處理組時，可用 Johnson 和 Neyman (1936)創用的方法，此法並不需要迴歸斜率相等之假設，使用各處理組斜率就可以適合迴歸線。可以用圖十一加以說明：



圖十　迴歸線非平行



圖十一　Johnson-Neyman 檢定

此 Johnson-Neyman 法除了不需有迴歸斜率同質外，其他假設如各組 Y 與 X 上的迴歸要呈直線，對每一 X 值，Y 呈常態分配，且變異數相等也都需要滿足。此法之功能在於找到 Y_A-Y_B 值之顯著區域。換言之，X 要在什麼值以上或以下，差異才會達顯著。同樣地，Li (1964)也提出兩樣本以上迴歸同質無法滿足時，調整平均數之檢定方法。一般使用的調整平均數之檢定，只是這個方法的特例。所檢定的假設是：在共變量分配的某一點，處理母群平均數相等。他指出，所選擇的點通常是所有的共變量平均值，這與一般使用的檢定相同。可惜的是很少有書把這個可爲一般使用的檢定列入。

到底迴歸線平行假設不符合，有什麼影響呢？若資料使用不等斜率模式較恰當時，進行一般共同斜率模式之共變數分析之影響如何？Atquillah (1964)曾以兩處理組、多個處理組兩種情況加以研究。用於兩組情境的方法是：針對兩種模式，分別導出調整的處理平均數差異之期望值與變異數。他發現期望值顯示有明顯的差異，除非共變數相等或兩組共

變數之變異相等。雖然兩模式之平均數差異之變異數公式相同，但是 Atquillah 無法說明兩模式 σ_e^2 之估計數將有何不同，變異數公式如下：

$$V(\hat{\alpha}_1' - \hat{\alpha}_2') = \sigma_e^2\left[\frac{2}{n} + \frac{(\overline{X}_1 - \overline{X}_2)^2}{\Sigma\Sigma(X_{ij} - \overline{X}_{.j})^2}\right]$$

然而 Feldt (1958) 和 Elashoff (1969) 發現，如果斜率不平行，誤差變異數將會增加，增加多少決定於斜率差異之程度。這種情形將導致保守的 F 檢定，這也許是 Atquillah 認爲兩斜率有差異，則要分開畫出迴歸線以 X 爲函數來估計其效果之理由。

當兩個以上處理組之設計時，Atquillah 研究發現，即使共變量和依變數爲常態分配，組間 F 檢定之漸進平均數 (asymtotic mean) 和變異數會有偏差現象。這種結果之推衍需要接近無限大之處理組才能獲得，所以即使在這種不實際的情況下，F 檢定仍然存有偏差現象。

和此分析性研究相對的是 Peckham (1968) 的模擬研究，他的組數和每一組的受試者作九種組合對迴歸斜率進行系統改變研究。許多程序限制了此研究結果之類推性。爲產生實徵的抽樣分配，將共變量固定，即使斜率不一，也要將共變量之變異數維持恆常。固定共變量是我們擔心的，因爲如前述的在許多研究中，共變量將產生不等之誤差變異數，由於介入了另一假設之違反而使分析愈趨於複雜。然而 Peckham 發現實徵的 F 統計數之抽樣分配和理論分配在 .20，.10，.05，.01 等水準上並沒有太大之差別，例如，在五個處理組的設計中，斜率一直到 .20 和 .80 之差時，和常態理論之差距才需要擔心。各處理組之共變量平均數相等或不等，這種結果都是一樣的。許多有關 Peckham 之後續之研究可使得結果更具類推性。雖然迴歸線相等之假設應主動的列入共變數分析章節中，然而對於違反迴斜線同質對第一類型錯誤之影響一直都未能獲得一個明確之結論，解釋也仍是一個問題。

（七）有關實驗誤差的假設

在使檢定的統計數能符合某已知分配，除了模式中所敘述之一些假設外，還需要一些其他之假設。在變異數分析和共變數分析中，爲了均方比例能形成 F 分配，需對有關無法控制或誤差變異數作某種分配上的假設。有些書對觀測值 Y_{ij} 之分配作假設而不是對誤差作假設。其實這兩種方法都是一樣的，因爲在任何一處理母群中，觀測值之唯一變異來源就是誤差部分。有關誤差的假設如下：

(1)誤差呈獨立性分配（獨立性）

(2)在每一處理母群內，對每一共變量值，誤差呈常態分配（常態性）

(3)在每一處理母群內，對每一共變量值，誤差變異數相等（變異同質性）

這三個假設通常用下面的符號來表示：

$e_{ij} \sim NID(0, \sigma^2)$

此符號只在說明誤差呈平均數爲零，變異數爲 σ^2 之常態與獨立的分配。

爲什麼這三個變異數分析和共變數分析所依據之假設是那麼重要呢？因爲在此兩種分析中，是根據這些假設來建立數學模式的，而且所有的模式都接近眞實，由於違反假設是無可避免的，統計學者因而把研究問題轉向「假設違反達到多嚴重才會對第一與第二類型錯誤率有嚴重的影響？」也許設定的 α 爲 .05，只允許有 5% 時候犯了錯誤的拒絕，然而在某一假設違反下，可能有 40% 的時候犯了錯誤的拒絕，遇到這種情形，勢必要能偵測到這種違反情形並採取某些補救辦法。可是所有假設違反都不嚴重的情況下，要知道那一個假設需特別注意，在什麼情境之下要特別注意，恐怕是十分重要的。

在討論此三種假設之前，應先對幾個常用之名詞需先了解，設定的顯著水準 (α) 是指因研究者指定犯第一類型錯誤之水準，即當虛無假設爲眞且所有的假設都符合時，錯誤拒絕時候之百分比。實際的 α 是指如果一個或多個假設違反，研究者錯誤拒絕之百分比。如果實際 α 非常接近設定的 α，就說此檢定統計數是强韌的 (robust)。

有許許多多的研究在探討違反變異變分析或共變數分析假設的影響，其中以 Glass 等人 (1972) 討論最爲深入。他們指出非常態對第一類型錯誤率上有些許之影響，即使是偏態或呈高狹或低闊峰的分配影響也不大。例如，對某些非常態母群，實際之 α 只有 .055 或 .06，此值與設定的水準 .05 十分接近。像這種情形，對於常態假設，F 統計數是强韌的。也許有些讀者對於這一點感到迷惑。基本理由是中央極限定理 (central limit theorem)。此定理在說明任何分配之獨立觀測值之總和（或平均）會隨著觀測值數之增加而接近常態分配。Bock (1975) 研究指出，即使明顯偏離常態之分配，50 個或更多的觀測值之總和就接近常態了。就是中度之非常態，分配少至 10 到 20 個觀測值就很接近常態了。Glass 和 Hopkins (1984) 也發現母群體即使呈矩形或偏態，樣本數少到 5 或 10，其抽樣分配也慢慢接近常態分配了。既然獨立觀測值之總和隨著觀測值數之增加很快就接近常態，平均數也同樣有這種情形。又 F 抽樣分配是依據平均數，因此在非常態之情況下，F 之抽樣分配只是稍稍偏離常態而已。所以常態或非常態分配所得之抽樣分配，其臨界值差是非常微小的。還有，偏態所造成的非常態對檢定力之影響也是微乎其微。可是低闊峰會影響檢定力，特別是樣本數 (n) 小的時候更爲明顯 (Baker, 1972)。

至於母群變異同質的假設，若各組人數相等或相近時（最大與最小之比小於 1.5），變異數雖然不相等，F 檢定統計數卻呈强韌的。也就是，實際的顯著水準與設定的顯著水準十分相近。唯在各組人數有明顯不相同（最大與最小之比大於 1.5）而且統計檢定顯示母群變異數不相等時，才要特別注意。在這種情況，研究發現，如果人數

少的組共變異數較大，則F統計數是激進的 (liberal)，所謂激進的統計數意指：錯誤拒絕太多，即實際 α 大於設定的 α 。也就是說，研究者原認爲有5%的時候錯誤拒絕虛無假設（設定的 α ），但事實上，他眞正拒絕率，可能達到11%（實際 α ）。反之，當人數多的一組，變異數較大，則F統計數是保守的 (conservative)，意指實際 α 小於設定的 α ，有許多研究者誤認爲這沒什麼不好。然而，我們應注意到， α 值小會造成檢定力減小的情形。

目前有許多變異數同質之檢定方法，例如，Bartlett法，Cochran法，Hartley法，只可惜上述三種方法對非常態性相當敏感。也就是說，研究者使用這些檢定，可能拒絕了虛無假設而說母群變異數不相等；事實上，拒絕虛無假設可能是由於母群非常態所造成。一般言，Levene提出的檢定方法比較不會發生此種問題 (Kennedy & Bush, 1985)。

表一 一因子變異數分析中觀測值有相關造成實際第一類型之錯誤率*（設定的 α ＝.05）

		ρ								
m	n	.00	.01	.10	.30	.50	.70	.90	.95	.99
2	3	.0500	.0522	.0740	.1402	.2374	.3819	.6275	.7339	.8800
	10	.0500	.0606	.1654	.3729	.5344	.6752	.8282	.8809	.9475
	30	.0500	.0848	.3402	.5928	.7205	.8131	.9036	.9335	.9708
	100	.0500	.1658	.5716	.7662	.8446	.8976	.9477	.9640	.9842
3	3	.0500	.0529	.0837	.1866	.3430	.5585	.8367	.9163	.9829
	10	.0500	.0641	.2227	.5379	.7397	.8718	.9639	.9826	.9966
	30	.0500	.0985	.4917	.7999	.9049	.9573	.9886	.9946	.9990
	100	.0500	.2236	.7791	.9333	.9705	.9872	.9966	.9984	.9997
5	3	.0500	.0540	.0997	.2684	.5149	.7808	.9704	.9923	.9997
	10	.0500	.0692	.3151	.7446	.9175	.9798	.9984	.9996	1.0000
	30	.0500	.1192	.6908	.9506	.9888	.9977	.9998	1.0000	1.0000
	100	.0500	.3147	.9397	.9945	.9989	.9998	1.0000	1.0000	1.0000
10	3	.0500	.0560	.1323	.4396	.7837	.9664	.9997	1.0000	1.0000
	10	.0500	.0783	.4945	.9439	.9957	.9998	1.0000	1.0000	1.0000
	30	.0500	.1594	.9119	.9986	1.0000	1.0000	1.0000	1.0000	1.0000
	100	.0500	.4892	.9978	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ρ 表示級內相關　　* 轉載自Scariano & Davenport, 1987,第124頁
m 表示組數
n 表示每組觀測值數

獨立性假設可說是三個假設中最爲重要者，即使違反情形不嚴重，對於F檢定顯著水準和檢定力卻有相當大的影響。誤差間少量的相依，造成的實際 α 要大於設定的 α 好幾倍之多，這種相依程度可利用級內相關 (intraclass correlation) 來測量之：

$$R = (MS_b - MS_w) / (MS_b + (n-1)MS_w)$$

式中，MS_b 和 MS_w 分別是F統計數之分子與分母，n則爲每一組受試者人數。表一顯示，相依性對第一類型錯誤之影響十分嚴重。例如三個處理組，每組10人，且中度相依 (R = .30) 爲例，實際 α 值卻高達 .5379。三個處理組，每組30人，相依性極小 (R = .10)，實際的 α 也達 .4917，此值幾乎是設定的 α 值 .05 之十倍。從表中也可看出：在同一區的級內相關，實際之 α 隨著人數之增加而增加。

現在讓我們來思考社會科學研究中，一些觀測值（或誤差）相依性可能出現之情境。教學方法的研究就是一個典型的相依性可能出現的情境。例如，一個班級中，其中幾個特殊人物可能對其他同學的成績有決定性的影響。因此，一個班級同學後測之成績至少部分取決於不良的班級氣氛。反之，即使在良好的班級氣氛中，相依性仍然可能發生，因爲許多同學的成績會受到積極性學習情境而增強。因此，不管是正性或負性的班級氣氛，同學之成績彼此間並非相互獨立的。 Glass 和 Hopkins (1984) 認爲無論何時，如果對受試者個別實施處理，所得的觀測值相互間可視爲獨立的；但是如果處理牽涉到人與人之間之交互影響，例如討論法，或團體諮商，觀測值可能相互間會有影響。因此隨機分派雖然較非隨機分派較有可能使得分數之間相互獨立，但不一定保證獨立性假設一定符合，因此研究者在整個研究的過程中須隨時注意可能發生相依的情況。

倘若觀測值（或誤差）有關連，研究者應如何處理？從表一數據知道，即使觀測值間之相關非常低，其實際所犯之 α 值大於設定的 α 值很多，因此， Stevens (1990) 認爲如果研究者有相當的理由認爲觀測值間相互依賴（或關連），則可使用較嚴謹之顯著水準，如此做，即使實際所犯之 α 大於設定的，仍是個能容忍的數值。如3或5組，每組10個觀測值，級內相關爲 .10 的例子，則實際 α 大於設定的顯著水準 .05 達六倍之多。像這種情況，使用 $\alpha = .01$ 來進行檢定恐怕好些，因爲實際 α 值可能爲 .05 或大一點點而已。

如果在每一處理中，又分幾個小組進行實驗（如諮商），這種情況，雖然各處理之間之觀測值相互間可能獨立，但是各處理組內受試之觀測值間極可能會有相關存在。遇到這種情形，研究者宜用各小群體的平均數作爲分析單位。這麼做，當然會大大減少有效的樣本數，然而檢定力的減小並不如一些人所想像的。理由是使用平均數較使用各個的觀測值來得穩定，因此，組內之變異將會減小。現在我們利用表二來說明：

表二　在兩處理設計中，檢定力大於.80每組所需之人數

		級內相關					
			.10			.20	
	效果大小	.20	.50	.80	.20	.50	.80a
α水準	每組人數						
.05	10	73	13	6	107	18	8
	15	62	11	5	97	17	8
	20	56	10	5	92	16	7
	25	53	10	5	89	16	7
	30	51	9	5	87	15	7
	35	49	9	5	86	15	7
	40	48	9	5	85	15	7
.10	10	57	10	5	83	14	7
	15	48	9	4	76	13	6
	20	44	8	4	72	13	6
	25	41	8	4	69	12	6
	30	39	7	4	68	12	6
	35	38	7	4	67	12	5
	40	37	7	4	66	12	5

(a) .20表示低度效果大小
.50表示中度效果大小　　轉載自Barcikowski, 1981
.80表示高度效果大小

根據表中資料，若中度或高度之效果大小，檢定力大於.80，則每一處理所需之小組數不需要太大。例如，$\alpha = .10$，級內相關爲.10，中度之效果大小，每一處理只要10小組，每一小組10人就夠了，如果研究者認爲檢定力只要大於.75，顯著水準.15可以忍受，那麼每一處理只要用5或6小組，每一小組10人即可。由於係使用雙重的外插法來計算，所以這些數字只是一個概略估計數而已。

討論了有關共變分析的假設之後，讀者一定覺得相當困擾，正確使用共變數分析方法需要有那麼多的假設。我們不能不承認一組資料要符合上述所有的假設，實在太困難。然而這並非表示我們今後就完全不能使用這種方法。重要的是我們使用這種方法時，要隨時就上述假設去思考，如果違反假設其結果之影響可接受時，仍可使用之；若違反很大，而仍然使用，其研究結果可能毫無意義。Elashoff (1969)認爲上述所有

假設中，以隨機分派，共變量與處理效果無關，迴歸同質三項假定，對於共變數分析的結果之解釋最爲重要。而 Stevens (1990) 則認爲在使用統計套裝程式之前需先檢定下面兩個假設：一是依變數與共變量是否有直線之趨勢，第二是否迴歸斜率是否同質。然而這並非說明其他假設不重要，研究者應盡可能使這些假設滿足，即使辦不到，也應採取一些前面所述之補救方法使研究結果得以合理解釋。

四、使用原樣團體進行共變數分析應注意之要點

有一些學者（如 Anderson, 1963; Lord 1969）强烈反對研究者使用共變數分析方法來分析原樣團體的資料。但是，事實上，目前許多有關心理或教育方面之研究是使用原樣團體。完全反對使用未免太苛，濫用無度也非正確態度。研究者應先了解使用共變數分析於原樣團體將有那些限制或危險。詳加考慮這些因素，再使用此種統計方法，才能正確發揮此統計方法之功能。Stevens (1990) 提出使用原樣團體要注意下面幾個問題：

(1)使用原樣團體進行研究，即使用了幾個共變量，恐怕仍無法使各組原樣群體在各方面完全相等，在一些未知的變數上仍可能有差異，而且就某一個變數使各組相等，說不定還會加大了其他變項的差異。

(2)使用共變數分析目的之一爲調整後測平均數，彷彿所有各組在實驗開始各組共變量是一樣的情境。然而我們要考慮在這共變量相等的團體在現實世界上是否眞正存在。例如研究者想要知道 A，B 兩種教學法那一種好，使用 A 教學法的班級都是高能力的學生，而使用 B 教學法的是低能力學生。雖然共變數分析方法，可以使這兩班的能力一樣，再比較調整之後的後測成績，但是「想用平均能力的學生來比較 A，B 兩種方法那一個好」，是毫無意義的，因爲 A，B 兩種教學法是分別爲某一能力水準而設計的。將來，這兩種方法沒有一種是要用於平均能力的學生的。

(3)由於直線與迴歸斜率同質兩假設必須滿足，使用共變數分析法才適當。可是原樣團體是否在此兩假設較不易符合，這是值得注意的。

(4)研究結果容易受到原樣團體在依變數上不同之成長而受到混淆，因而不易解釋處理之效果。如果某一處理組自然成長大於控制組，兩組共變量又有顯著差異，調整前測差異之後，後測之差異到底是由於處理效果？不同成長？或此兩者的效果，就很難判定了。

(5)測量誤差的問題雖然也會發生在隨機設計的研究，但是在這種研究中之影響只有減少檢定力而已，在非隨機研究中，測量誤差會嚴重造成處理之偏差。Reichardt

(1979)指出，前測（共變量）的測量誤差，就是沒有處理效果，也會產生假性的處理效果。Pedhadzar (1982)也曾討論使用原樣團體測量誤差之影響，他一再提醒研究者正視測量誤差的問題而能達成下列兩點共識：①應設法使共變量之量數具有高度之信度②忽視這個問題（正是目前大部分使用共變數分析情況）也無法使它消失。

了解這五個限制之後，讀者也許會懷疑是否使用原樣團體，是否應放棄使用共變數分析方法。其實，其他用於分析這一類資料的統計方法如樣本配對，淨得分數之變異數分析也會遭遇到同樣的問題，例如，嚴重造成處理效果之偏差。根據原樣團體來推論因果，不管何種統計方法同樣辦不到。因此，我們要做的是盡我們最大努力，注意所有可能產生的問題，如同Pedhadzar (1982)所言，進行這種研究（事實上所有科學研究都一樣）需要研究者作合理的思考、持續的警惕，對使用的方法之功能與限制完全的了解。

五、使用共變數分析模式其他應注意的事項

研究者使用共變數分析時，除了對上述之各假設及使用原樣團體要注意的要點以外，尚有一些概念在使用共變數分析時也需考慮。Marascuilo和Serlin (1988)認爲如果迴歸線不是直線，迴歸線彼此間未平行，等分散性（變異同質性）未能滿足，殘差不呈常態，研究者使用共變數分析模式就不恰當了，因爲這些假設都是用來獲取在虛無假設爲眞的情況下F抽樣分配。等分散性和常態雖也是使用隨機區組之假設，然而非直線迴歸與迴歸線未相互平行，則非隨機區組之必要假設。Marascuilo和Serlin (1988)曾介紹另一種統計控制的方法稱爲調整平均數法(adjusted average)，此法適用於不等之樣本數，除此，對細格變異數不相等也有效。在某些情況中，研究者想要使用共變數分析，但等分散性假設不符合時，可使用調整平均數法。研究者究竟選擇共變數分析還是區組設計呢？他們認爲：若r_{xy}大於0.6，則使用共變數分析較佳，因爲能顯現較大的檢定力；若r_{xy}低於.40則以隨機區組設計較好；若r_{xy}介於.40與.60之間，使用那一種並無太大的差別。

另外，如何選擇共變量也是值得研究者注意的。一般言，任何理論上與依變數有相關的變數，都應視爲共變量。但是，理想上，要選擇與依變數有顯著相關，而彼此之間相關又低的變數。若兩共變量相關高（如.80），則它們從依變數移除很多相同的誤差變異。這樣，加入了第二個共變量，事實上並沒有太多助益。反之，若兩共變量之相關很低，則它們從依變數中移除很多不同的誤差變異，也就是說，誤差變異可以減小很多。

雖然共變數分析可使用於兩個或兩個以上之共變量情形，但也不能漫無限制。Huitema (1980)曾使用下列公式，以限制使用的共變量數。

$$\frac{[C+(J-1)]}{N} \leqq .10$$

式中C爲共變量數目，J爲組數，N爲總樣本人數。例如我們有四個處理組，總共80位受試者，則$(C+3)/80 \leqq .10$，即$C \leqq 5$，換言之我們最多只能用五個以下的共變量。他認爲：若$[C+(J-1)]/N > .10$，則共變數分析F檢定雖正確，但調整平均數之估計值可能十分不穩定。

在許多研究中，研究者常會要比較兩組或多組前後測的資料。這種研究，我們可以把前測資料當作共變量，後測當依變量使用共變數分析方法來處理以外，可能研究者會用下列兩者之一：

1. 使用前後測之差異分數或淨得分數進行單因子變異數分析。
2. 使用二因子重覆量數變異數分析；一個因子爲組別變數，一個因子爲處理變數。

Huck和McLean (1975)和Jennings (1988)曾以上述兩種分析模式與共變數分析模式相比較。他們都認爲共變數分析方法最優，Huck和Mclean認爲使用重覆量數法，交互作用F值能指出處理定否有不同效果，而非處理的主要效果。試看下面兩個情境平均數組型就可以了解。

情　境　一

	前測	後測
處理組	70	80
控制組	60	70

情　境　二

	前測	後測
處理組	65	80
控制組	60	68

在第一個情境中，處理的主要效果可能達到顯著，因爲橫列之平均數相差10點，然而後測兩組10點之差只是初測的差異（10點）之延轉而已。實驗組與控制組兩組之後測差並非由於處理不同而造成；反之，情境二中，即使處理組在前測分數雖較控制組爲高，但在後測增加了15點，而控制組只增加8點，此表示兩組的表現有不同的改變。因爲在二因子變異數分析中，交互作用的效果可視爲差異之差異。所以情境二正可視爲是顯著的交互作用效果。又Huck和McLean (1975)表示：重覆量數分析之交互作用F值和對淨得分數進行變異數分析所得之F比例是完全一樣。又迴歸係數不等於1時（大多數都是這種情形），共變數分析之誤差項將小於淨得分數之變異數分析之誤差項，因而共變數分析是一種較爲敏感或較具檢定力之分析。

另外，使用淨得分數進行變異數分析研究者要考慮的一個問題是：一般而言，淨

得分數之信度較差。明確的言，若前測與後測分數之間的相關與測驗的信度係數愈相近，則差異分數之信度係數愈近於0 (Stevens, 1990; Murphy & Davidshofer, 1990)。差異分數之信度估計值之公式如下

$$r_{DD}=\frac{\dfrac{r_{xx}+r_{yy}}{2}-r_{xy}}{1-r_{xy}}$$

式中，r_{xx} 爲前測分數之信度，r_{yy} 爲後測分數之信度。若前後測使用相同測驗則 (r_{xx} + r_{xy})／2正是測驗之信度，r_{xy} 爲前後測分數之相關。心理計量學者（如Cronbach和Furby 1970）亦指出前後改變的測量存在著許許多多的問題，前後改變的量數應予以放棄。

總　結

共變數分析是一個最常用和最不易爲人了解的統計方法，由於此法是迴歸分析與變異數分析之組合，因而這兩種統計方法所需之假設也是共變數分析所必要的。此法和其他統計方法一樣，要完全滿足所需之假設雖十分困難，但研究者必須注意違反假設對結果之影響有多大。誠如Elashoff (1969)所言，研究者必須小心使用與解釋共變數分析及其結果。對於共變數分析方法之使用，Baker (1972)曾提出非常簡潔而又實用之原則。玆簡單敘述以爲讀者參考：如果研究者想透過使用共變數分析以對一變數進行統計上的控制，則應檢討研究中所關心的問題以確定調整共變量值是否合乎邏輯，是否獲得的結論有意義。倘未隨機分派受試者到各處理組，則應檢視所使用的原樣團體之性質，如果共變量數是在實施處理之後獲得的，除非有相當的理由相信處理和共變量未相混淆，否則應放棄使用共變數分析方法。

若上述問題都考慮而沒問題之後，研究者需進一步檢視依變量與共變量之間的關係，以決定是否依變數直線迴歸調整能代表資料。倘若共變量是隨機變數，則應考慮其分配之形狀是否影響到依變數非常態檢定之敏感性。研究者也要檢視用來獲得共變量量數之工具的信度，因爲，信度低，斜率之最小平方估計數將會受到重大的影響，依變數平均數之調整可能有過之或不及的現象。除了上述這些考慮以外，爲了使用共同斜率以調整依變數平均數，檢定迴歸線平行也是必要的。

想要利用共變數分析以對一變數作統計上的控制之前，研究者應該先作所有重要的與困難的決定。而且，也要考慮是否依變數與共變量數間之相關大到使用此法較使用多因子變異數分析方法（以共變數作爲區組變數）爲佳。變異數分析所需要的假設較少，要求也沒有像共變數分析那麼嚴格。因此，使用共變數分析方法所得到的精確性應從假設性質所帶來的問題來評鑑之。

閱讀本文之後，明顯地，共變數分析誠如Elashoff所言是一相當「脆弱」的工具，這並非意味我們不應使用這種統計方法，而是研究者必須徹底了解它，且能辨別該不該用，其結果之解釋是否合理，並非所有教育領域的研究統計，共變數分析都能適用，所以必須仔細考慮其使用之適當性。然而由於此種統計方法仍有可能被繼續濫用，研究者應採批判觀點，檢視共變數分析之研究結果。最後，Baker建議，爲使有助於研究者正確使用共變數分析方法，有關統計之教科書或課程對此方法之討論應更爲周詳和審慎。使用此法之前，研究者也應重新溫習這個統計方法，如此，才能反映研究者是完全了解而非隨隨便便的決定。

參考書目

朱經明（民78）。教育統計學。台北市：五南。

林清山（民81）。心理與教育統計學。台北市：東華。

謝廣全（民73）。實用心理與教育統計學。高雄市：復文。

Anderson, N. H. (1963). Comparison of different populations: Resistance to extinction and transfer. *Psychological Bulletin,* **70**, 162-179.

Atquillah, M. (1964). The robustness of the covariance analysis of a one-way classification. *Biometrika,* **51**, 365-372.

Baker, C. W. (1972). The analysis of covariance: A review and discussion. Working paper No. 26 of the Department of Educational Research.

Barcikowsk, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics,* **6**, 267-285.

Becker, W. C., & Engelmann, S. (1976). Teaching 3: Evaluation of instruction. Chicago: Science Research Associates, Inc.

Berkson, J. (1950). Are there two regressions? *American Statistical Association Journal,* **45**, 164-180.

Bock, R. D. (1975). Multivariate statistical methods in behavioral research. New York: McGraw-Hill.

Box, G. E., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association,* **70**, 70-79.

Cochran, W. G. (1957). Analysis of covariance: its nature and uses. *Biometrics,* **13**, 261-281.

Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics,* **10**, 637-666.

Cronbach, L. J., & Furby, L. (1970). How we should measure change or should we? *Psychological Bulletin,* **74**, 68-80.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal,* **6**, 383-401.

Evans, S. H., & Anastasio, E. H. (1968). Misuse of analysis of covariance when treatment effect and covariate are confounded. *Psychological Bulletin,* **69**, 225-234.

Feldt, L. S. (1958). A comparison of the precision employing a concomitant variable. *Psychometrika,* **23**, 335-353.

Fisher, R. A. (1934). The design of experiments. London: Oliver and Boyd.

Glass, G. V., & Hopkins, K. D. (1984). Statistics in education and psychology (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research,* **42**, 237-288.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). Design and analysis of time series experiments. Colorado Associated University Press, Boulder.

Gourlay, N. (1953). Covariance analysis and its applications in psychological research. *British Journal of Psychological Statistics,* **6**, 25-34.

Hope, K. (1968). Methods of multivariate analysis. New York: Godon & Breach.

Horst, D. P., Tallmadge, C. K., & Wood, C. T. (1974). Measuring achievement gains in educational projects (RMC. Rep. UR-243). Los Altos, Ca: RMC Research Corporation.

Huck, S. W. (1972). The analysis of covariance: Increased power through reduced variability. *Journal of Experimental Education,* **41**, 42-46.

Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin,* **82**, 511-518.

Huitema, B. (1980). The analysis of covariance and alternatives. New York: Wiley.

Jennings, E. (1988). Models for pretest-posttest data: repeated measures ANOVA revisted. *Journal of Educational Statistics,* **13**, 273-280.

Kahneman, D. (1965). Control of spurious association and the reliability of the controlled variable. *Psychological Bulletin,* **64**, 326-329.

Kennedy, J. J., & Bush, A. J. (1985). An introduction to the design and analysis of experiments in behavioral research. New York: University Press of America.

Li, J. C. R. (1964). Statistical inference, I. Ann Arbor, Michigan: Edwards Brothers, Inc.

Lord, F. M. (1960). Large sample covariance analysis when the control variable is falliable. *American Statistical Association Journal,* **55**, 307-321.

Lord, F. M. (1969). Statistical adjustments when comparing pre-existing groups. *Psychological Bulletin,* **70**, 162-179.

Mandansky, A. (1959). The fitting of the straight lines when both variables are subject to error. *American Statistical Association Journal,* **54**, 173-205.

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

Murphy, K. R., & Davidshofer, C. O. (1990). Psychological testing: Principles and applications. Englewood Cliffs, N.J.: Prentice-Hall.

Peckham, P. D. (1968). An investigation of the effects of non-homogeneity of regression slops upon the F-test of analysis of covariance. Laboratory of Educational Research Report No. 16, Boulder, Colorado: University of colorado.

Pedhazur, E. (1982). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Porter, A. C. (1967). The effects of using fallible variables in the analysis of covariance, Ph. D. thesis. Madison, Wisconsin: University of Wisconsin.

Porter, A. C., & Raudenbush, S. W. (1987). Analysis of covariance: Its model and use in psychological research. *Counseling Psychology,* **34**, 383-392.

Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent group designs. in T. Cook & D. Campbell (Eds.), Quasi-experiment: Design and analysis issues for field settings. Rand McNally, Chicago, Ill.

Scariano, S., & Davenport, J. (1987). The effects of violations of independence assumptions in the one-way ANOVA. *The American Statistician,* **41**, 123-129.

Scheffe, H. (1959). The analysis of variance. N.J.: John Wiley & Sons.

Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods. Iowa State University Press/AMES.

Stevens, J. (1990). Intermediate statistics: A modern approach. New Jersey: Lawrence Erlbaum Associates.

Winer, B.J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

THE FUNCTIONS, ASSUMPTIONS, AND LIMITATIONS FOR ANALYSIS OF COVARIANCE

Der-Hsin Fan

Department of Educational Psychology and Counseling

Abstract

The analsyis of covariance (ANCOVA) is one of the statistical techniques frequently used by educational researchers because it has the functions of statistical control and reduction of within group or error variance. However, it is a very delicate and commonly misunderstood procedure. So that users can accurately use ANCOVA, the author discusses it in terms of the following six parts.

The first part describes its rationale and model. The second part depicts ANCOVA's two major functions-elimination of systematic bias and reduction of within group or error variance. The assumptions of the model, effects of voilation, and corresponding strategies to be used are illustrated in the third part. The assumptions examind are: (1) that cases are randomly assigned to treatment conditions, (2) that the covariate is independent of the treatment effect, (3) that the covariate is fixed and measured without error, (4) that the covariate is linearly related to the dependent variable, (5) that regression of the dependent variable on the covariate is the same for each group, and (6) that the assumptions regarding experimental error (including independence, homogeneity of variance, and nomality) have not been violated. The use of ANCOVA with intact groups is discussed in the fourth part. Some miscellaneous considerations are raised in the fifth part. The final part summarizes the pratical steps involved in utilizing ANCOVA.