

# 主成份分析和共同因素分析相關議題之 探究

傅粹馨

國立高雄大學教育系

主成份分析與共同因素分析廣為研究者所使用抽取因素的方法。主成份分析與共同因素分析是不同的。共同因素分析之目的在抽取因素來解釋變項間的關係，而主成份分析之目的在作變項的減縮。

本文探討八個主題：(1)緒論；(2)主成份分析之意義；(3)共同因素分析之意義；(4)主成份分析與共同因素分析之異同；(5)主成份分析與共同因素分析結果近似；(6)共同因素分析較主成份分析為宜；(7)目前普遍使用之因素分析方法；(8)實例分析。

執行因素分析時宜瞭解研究之目的與決定抽取多少個因素是相當重要的，這些會影響因素的抽取與轉軸後的結果，研究者不能太依賴電腦提供之內設選項。

關鍵詞：共同因素分析、因素抽取、主成份分析、轉軸、陡坡圖。

## 一、緒論

目前電腦之軟體硬體發展急速，以較複雜之統計方法分析資料已是輕而易舉之事，執行少數幾個指令，分析工作即大功告成，不論國內外，社會行為科學研究中使用多變量統計分析方法之頻率漸增。本文重點置於使用頻繁之主成份分析(principal component analysis; PCA)和共同因素分析(common factor analysis; CFA)上，此二分析之意義分別為何；二分析之異同為何；認為二分析結果近似之相關研究；認為共同因素分析較主成份分析為宜之相關研究；與目前普遍使用之因素分析方法，均是本文所欲探討之主題。

## 二、主成份分析之意義

成份分析模式(component analysis model)包含了常用的主成份分析(Principal component analysis; PCA)和映象成份分析(image component analysis)二種，因而有主成份分數(principal component score)和映象成份分數(image component score)。主成份分析是由 Pearson 所創用而由 Hotelling 再加以發展的一種統計方法(林清山, 1991)，Pearson 將 PCA 用在正交之迴歸分析(orthogonal regression)為其原意與目的(Jackson, 1991)。於主成份分析中，可將  $m$  個變項加以轉換，使所得線性組合而得  $P$  個 ( $P < m$ ) 成份的變異數變為最大，且成份間彼此無關，這特性也讓研究者將其用在多元迴歸分析中，解決預測變項間的多元共線性問題和在多變項變異數分析中，太多的依變項間具高相關情況下，利用 PCA 使變項變為無關的數個成份分數，以利後續的統計分析。

主成份分析是假設所分析之變項不含誤差，樣本之相關係數矩陣即代表母群之相關係數矩陣。 $N$  個變項經主成份分析會產生  $N$  個成份，一般而言，研究者會從  $N$  個成份中選取前面數個變異量較大之重要成份，而忽略變異量小，不重要之成份，此種方式稱之為 truncated component solution (Gorsuch, 1988)。

主成份分析具以下之特性（林清山，1991；Nunnally & Bernstein, 1994）：

- (一)、主成份分析將成份分數在新軸上的變異數變為最大。
- (二)、每一成份之特徵值(eigenvalue)為該成份之變異數。
- (三)、特徵值為 0 或正值，不為負值。
- (四)、特徵值不為零之個數即代表該相關矩陣之向度(dimensionality)
- (五)、所有特徵值之和為相關係數矩陣之跡(trace)，亦為變項之個數。
- (六)、主成份分析不僅具有幾何學上正交的特性 (geometrically orthogonal)，亦即所有變項每對成份之組型(pattern)係數之乘積和為零，而且具有統計上正交(statistically orthogonal)的特性，亦即成份間彼此無關。
- (七)、所有特徵值之乘積為該相關係數矩陣的行列式(determinant)。

### 三、共同因素分析

共同因素分析是 Spearman 所創用，Thurstone 等加以發揚的一種多變項統計方法（林清山，1991）。於抽取因素時，共同因素分析之過程與主成份分析是相同的，不同處在於相關係數矩陣對角線上的數值，於 PCA 時，對角線之數值為 1；而 CFA 時，對角線上為小於 1 之數值（即共同性之估計值），此矩陣稱之為「縮減成相關係數矩陣」(reduced correlation matrix)（林清山，1991）或「調整的相關係數矩陣」(adjusted correlation matrix)。採此法之研究者其興趣在探討觀察變項間是否能以數個潛在變項(latent variable)來代表觀察變項之間的關係。

共同因素分析中，較常用的方法有主軸法(principal axis factoring)，最大概似法(maximum likelihood factoring)與  $\alpha$  因素分析(alpha 或  $\alpha$  factoring)。

### 四、主成份分析與共同因素分析之異同

在社會科學研究中，研究變項的減縮(reduction)與量表的編製，常以主成份

分析與共同因素分析二種方法抽取成份或因素，這二種方法之異同(林清山，1991；Gorsuch,1997；Joreskog & Sorbom, 1979；Hatcher, 1998；Sharma, 1996；Widaman, 1993)，歸納如下：

### (一)、PCA 與 CFA 相同之處

- 1、二者皆可用來將原有之  $P$  個變項減為  $m$  個成份分數或因素分數( $m < p$ )，以符合精簡(parsimony)的原則。
- 2、當二者之組型矩陣(pattern matrix)經過適當之轉軸後，該矩陣能描述原觀察變項(observed variable)與新成份或新因素之間的關係。
- 3、當觀察變項數目多且共同性高之情況下，資料採此二種方法分析，所得結果近似。

### (二)、PCA 與 CFA 相異之處

欲區別主成份分析與共同因素分析，可分別由理論部分與實證部分加以說明，先就理論部分加以敘述。

Gorsuch(1983,1988)指出二者在理論上的差異在於共同因素分析模式屬於科學或統計的模式(scientific or statistical paradigm)，認為變項之測量含有誤差(error)部分，而成份分析模式屬於數學模式(mathematical paradigm)，將  $p$  個變項分數變為  $m$  個（通常是  $m < p$ ）成份分數，故成份分析模式為共同因素分析模式的一個特例，亦即，成份分析將唯一性(uniqueness)設定為零(Widaman,1990)；使用主成份分析時，相關矩陣之主對角線為 1，假定總變異數是完全由各成份所造成，不含誤差，然而，於因素分析時，主對角線之數值小於 1，表示總變異數中含有誤差(林清山，民 80)。Cattell (1988)將成份分析模式視為一種封閉模式(closed model)，共同因素分析模式視為一種開放模式(open model)。

再者，某些學者（行為統計學家）視因素分析為相關（或共同變數）取向(correlation or covariance oriented)，視主成份分析為變異數取向(variance oriented)

(林清山, 1991; Joreskog & Sorbom, 1979) 因素分析之目的在於再製(reproduce)變項的相關係數矩陣，而主成份分析之目的在再製變項的總變異量(Joreskog & Sorbom, 1979)，換言之，主成份分析的重點在解釋資料的變異量，而因素分析之重點在解釋變項間的相關。在主成份分析中，全部的成份都要用到，才能再製原來的相關矩陣；在因素分析時，只要少數幾個因素即可再製原來的相關矩陣(林清山, 1991)。

此外，於主成份分析中，成份是觀察變項的線性組合(linear combination)；而在因素分析中，觀察變項為各因素之線性組合加上誤差，二者在概念上是不相同的。

主成份分析與共同因素分析在實徵研究部分之異同，可藉由這二種方法所得之成份分數或因素分數間之相似性或因素組型(pattern)間之相似性來說明；有些研究採真實資料分析，有些是採電腦模擬的資料分析(computer-generated or simulated data or artificial data)。下面擬就實徵部分之結果，歸納為二大類加以說明：(一)主成份分析與共同因素分析之結果近似；(二)共同因素分析較主成份分析為宜。

在以下許多之成份分析與共同因素分析研究中，主成份分析(PCA)佔成份分析的大部分，而在共同因素分析中，主軸因素分析(principal axis factoring)使用較多。主成份分析是將相關係數矩陣(主對角線之值為 1)加以分析，從理論的觀點，認為每個變項沒有測量誤差；而共同因素分析是將共同性(communalities)置於相關係數矩陣之主對角線上，而共同性為每個變項為共同因素所解釋的變異量，它表示了變項是包含了誤差部分。最常被用來當共同性之估計值者是複相關係數平方(squared multiple correlation; SMC)，它是指每個變項與其他變項間之複相關係數平方，從此可知，變項之測量誤差亦包含在共同因素分析中，而至於初始與最終之共同性估計值(initial and final communality estimates)間之差異與抽取之因素之關係有關，換言之，假若一個變項與抽取之因素有相關，則共同性會增加，若變項與抽取之因素無關，則共同性維持不變。

此外，亦有不少研究者使用最大概似法執行因素分析(maximum likelihood

factor analysis)，將結果與其他方法作比較。當各種分析方法之結果作比較時，彼此採用相同的方法決定因素或成份的數目（如特徵值大於 1 的方法）和同樣的轉軸方式（如直交之最大變異法，varimax），如此以利結果之相互比較。

## 五、主成份分析與共同因素分析之結果近似

從變項數目的觀點而言，當變項數目較多（多於 40 個）且共同性均高（高於 .40）時，主成份分析與共同因素分析之結果十分近似(Acito & Anderson, 1980; Fava & Velicer, 1992a, 1992b; Gorsuch, 1983; Velicer & Jackson, 1990a, 1990b; Velicer, Peacock & Jackson, 1982; Wilkinson, 1989)

此外，Nunnally (1978; 引自 Benson & Nasser, 1998)指出，當變項數目多於 20 個時，則相關係數矩陣主對角線上之數值是否為 1，主成份分析與共同因素之結果十分近似。Gorsuch(1983)指出當變項數目增加時，相關係數矩陣變大，相對地，主對角線上之數值對整個的分析影響減少；例如研究含十個變項，則主對角線上數值個數之比重占整個相關矩陣中一百個相關係數的十分之一，當變項數目變為 40 個時，則主對角線上之個數所佔之比重成為 0.025（亦即 40/1600），故對二者分析之結果影響不大，雖是如此，但 Benson 和 Nasser(1998)仍認為應顧及理論部分，不能太著重實徵部分之結果。

Velicer 的 1990 年與 1992 年研究顯示，認為沒有任何理由去偏愛主成份分析或共同因素分析，因為從實際觀點而言，方法的選擇對實際研究結果的影響不大，反而是決定保留成份或因素之個數的抉擇是更重要的，作如此的結論是基於研究中，成份分析與共同因素分析二者在組型的比較(comparison of patterns)(Velicer et al., 1982)和成份或因素分數的比較(comparison of derived scores)(Fava & Velicer, 1992b)上均有高度的相似性，因而在資料之結構完備且能正確地決定因素數目下，則更有理由相信這二種方法所得的結果更近似了。

## 六、共同因素分析較主成份分析為宜

雖然 Velicer 等人於 1990 年與 1992 年之研究結果顯示，當變項數目多且共同性高時，成份分析與共同因素分析之結果十分近似(Fava & Velicer, 1992a, 1992b; Velicer & Jackson, 1990a, 1990b; Velicer, Peacock & Jackson, 1982)，但仍有不少研究結果並非如此(Bentler & Kano, 1990; Borgatta, 1989; Borgatta, Kercher & Stull, 1986; Daniel, 1990; Gorsuch, 1990; Hubbard & Allen, 1987; Lee & Comrey, 1979; Loehlin, 1990; McArdle, 1990; Mulaik, 1990; Snook & Gorsuch, 1989; Widaman, 1990, 1993)。

於上述之研究中，Bentler 和 Kano(1990)與 Widaman(1990)，以代數描述方式界定二者差異的範圍，其餘的研究者以真實或模擬資料作實證研究，歸納得知，當以下二種情形同時存在時，主成份分析和共同因素分析間產生之結果差異會最大：(1)變項與因素之比例太小（或另一種說法為變項的個數）且(2)變項在因素上之負荷量太低（或另一種說法為共同性太低），這二種情境使成份負荷量(component loading)比用來比照之因素負荷量(亦即真實負荷量 true loading)為大。事實上，二者主要之差異來自於相關矩陣之主對角線上的數值不同，主成份分析時主對角線上之數值是 1，而共同因素分析時是在對角線上放置共同性(Gorsuch, 1983)。

Mulaik(1990)指出 Velicer 和 Jackson(1990a)並未提及兩個研究，其一為 Borgatta, Kercher 和 Stull(1986)之模擬資料分析，其二為 Hubbard 和 Allen(1987)之七組眾所週知之實際資料分析，研究者均將資料採主成份分析與共同因素分析且保留相同數目的因素作比較，二者差異明顯，Hubbard 和 Allen 指出，不僅在組型係數(pattern coefficient)之大小和正負符號(sign)上有差異，更在因素解釋上有所不同。Velicer 和 Jackson(1990a)認為會產生如此的差異，因為在因素數目作了不當的決定，採特徵值（艾根值）大於 1 的準則或稱 Kaiser 的準則(eigenvalue greater than one; Kaiser's rule)，保留了太多的因素之故，亦即造成此差異是因抽取太多的因素(overextraction)。然而，Mulaik(1990)認為不宜模糊(blurring)主成份分析和共同因素分析二者之不同。Bentler 和 Kano(1990)亦認為

如此，不應期望在任何情況下，二者的結果都近似，當變項數目少，欲瞭解變項間之結構時，以共同因素分析為宜；雖是如此，但當研究之主要考量為資料的整理，特別是向度的縮減(dimension reduction)時，則以成份分析為宜。Gorsuch(1990)的觀點則認為宜將共同因素分析視為一種標準的分析形式，因為此分析考慮到變項內是應含有誤差(error)，且此法提供了不偏(unbiased)的因素負荷量。Snook 與 Gorsuch(1989)之 Monte Carlo 研究使用 36 個變項顯示，PCA 之組型負荷量(pattern loading)是偏誤的(biased)，而 CFA 則否。

Widaman(1993)指出當變項之共同性低且變項在因素或成份之負荷量不高時，則主成份分析與共同因素分析之結果呈現明顯地差異，而且重作 Snook 與 Gorsuch(1989)之研究，指出 PCA 與 CFA 二者之差異與每個因素下有多少個變項有關，而非與變項之總數有關。最後建議當研究者欲瞭解變項之潛在建構(latent constructs)或因素時，採 CFA 而不是 PCA。

McArdle(1990)指出 CFA 優於 PCA 基於以下理由：(一)精確的理由(mathematical grounds):主成份分析之負荷量與真實的負荷量(true loading)比較之下，有高估(overestimate)的現象，當真實負荷量愈來愈小時，其主成份分析之負荷量之精確度就愈來愈差，而共同因素分析不會有這種現象；(二)統計的理由(statistical grounds)：CFA 在複雜情境下（如偏態或峰度），與 PCA 結果之比較，CFA 較能提供穩定的數值。而 PCA 優於 CFA 係基於實用的觀點(practical grounds)，PCA 計算速度快些。從 Mulaik(1990)的觀點，指出 Velicer 和 Jackson(1990a)過度強化了主成份分析與共同因素分析結果之相似處，卻將不同處淡化了，Mulaik 認為在設計良好的心理實驗情境下，共同因素分析是較精確與可信的方法。

綜上可知，經由提出理論與實徵結果，支持共同因素分析論點者多於主成份分析，然而於期刊上的研究結果，使用主成份分析者佔大多數，其原因容後再述。



## 七、目前普遍使用之因素分析方法

Benson 和 Nasser(1998)指出執行因素分析時，須考量下列四項：(一)概念的或理論的考量(*conceptual theoretical consideration*)；(二)設計的考量(*design consideration*)；(三)統計的考量(*statistical consideration*)；和(四)報告的考量(*reporting consideration*)，本文重點置於(一)理論與(三)統計的考量上。理論考量部分即研究者須決定採用成份模式或因素分析模式(二者之異同前面已敘述)，統計的考量則包括了抽取因素的方法、決定因素的數目和轉軸三項，其中以抽取因素數目是最重要的一項(Comrey & Lee,1992;Gorsuch, 1983)，茲簡扼分別敘述如下：

### (一)、抽取因素的方法

於成份分析模式中，研究者最常使用的是主成份分析(PCA)，而於共同因素分析中，則是主軸法(*principal axis factoring;PAF*)，主成份分析廣受使用的原因如下：

Gorsuch(1997)指出，主成份分析雖在理論上有其不適切性(*theoretical inappropriateness*)，但仍有許多的使用者，其原因不外下列五項：1.共同因素分析有技術上的問題(*technical problem*)：亦即因素分數的不確定性(*indeterminacy of factor score*)，因素分數必須以估計的方式取得，且不是只有唯一的一個因素分數；成份分數則沒有這種問題；但少數研究者使用成份或因素分數；2.易於計算，於今日而言，這對 CFA 未必是一大問題；3.有些研究以最大概似因素分析(*maximum likelihood factor analysis, MLFA*)和主成份分析作比較(主成份分析不需要作疊代的工作)，因 MLFA 需要多次之疊代解法之程序(*iterative procedure*)以求得最後之共同性，則聚斂(*convergence*)有時是個問題，然而在主軸法(*principal axis*)之共同因素分析只須重複二或三次，即可得共同性，且 Snook 和 Gorsuch(1989)指出共同因素分析所得之因素，負荷量與母群之數值(真實負荷量)接近，然而成份分析所得之數值有偏高的傾向(*inflated loadings*)，基於此項理由，對成份分析之偏好有所矯正；4.當只有變項間具高相關或界定每個因素

之變項數目多之情況下，則主成份分析與共同因素分析所得結果近似；5. 普遍使用之統計軟體中，執行因素分析時，主成份分析為其內設(default)的指令，易讓使用者由電腦軟體來決定抽取因素之方法，此項似乎是最主要的原因。

## (二)、決定因素數目

決定因素數目後加以轉軸和解釋是因素分析過程中較為複雜的步驟，畢竟不當之因素數目經過轉軸後，易使研究結果扭曲(Fava & Velicer, 1996)，決定因素數目的方法有諸多的研究（林清山，1991; Bentler & Kano, 1990; Comrey & Lee, 1992; Gorsuch, 1983; Zwick & Velicer, 1982, 1986），決定何者是最適當的因素模式的方法之一是以何者最能精確地再製(reproduce)原來的相關係數矩陣。可知，抽取因素的數目與其再製原來之相關係數矩陣有關，事實上，研究者期盼以最少的因素數目卻能包含最多的訊息，緣此精簡原則，因素數目之問題似乎沒有唯一的解。

特徵值大於 1 和陡坡圖是最常用來決定因素數目的方法，尤其是前者，其為電腦統計軟體中的內設指令。茲將上述二種方法之適用時機敘述如下：

### 1、kaiser 的特徵值大於 1 的方法

特徵值(eigenvalue)大於 1 之方法原是為分析母群相關矩陣(population correlation matrix)主對角線為 1 而設計，且保留特徵值大於或等於 1 之成份。然而，此特徵值大於 1 之方法卻被誤用在分析樣本(sample)相關矩陣以決定保留共同因素數目，因而，在共同因素模式和樣本資料(sample data)情境下，通常造成高估(overestimate)或偶而低估(underestimate)因素數目之情形 (Cliff, 1988; Hakstian et al., 1982; Lee & Comrey, 1978)，雖然此法易於執行，且為統計電腦軟體中的內設選項，但使用時宜謹慎，因在某些情境下，此法會將差異微小的特徵值加以區分，例如，兩個連續的特徵值分別為 1.02 與 .98，第一個成份得以保留，而第二個卻被淘汰，然而這兩個成份的差異是微乎其微，這是值得注意

的。

Kaiser 之取特徵值大於 1 的論點是若特徵值小於 1，則成份分數之信度係數是負的(Kaiser,1960;引自 Cliff,1988)，而 Cliff(1988)指出 Kaiser 的論點是誤將 KR20 公式(Kuder and Richardson)用在求組合分數的信度(reliability of a composite)，認為有多少個特徵值大於 1 即表示有多少個可信賴的因素，Cliff(1988)表示這是沒有任何邏輯的依據。即使 Kaiser 的論點遭到 Cliff(1988)之反駁，且亦廣為研究者所批評(Gorsuch,1983; Lee & Comrey,1979; Tzeng,1992; Zwick & Velicer,1982, 1986)，但此法仍是最常被用為決定成份或因素數目的方法(Ford, MacCallum & Tait, 1986)，原因不外是特徵值大於 1 之方法為多數統計電腦軟體之內設選項。

## 2、陡坡圖之方法

陡坡圖(scree plot)之形成乃是將未轉軸前之特徵值畫在 Y 軸，而因素數目依序畫在 X 軸，再以研究者主觀的判斷決定該圖之決斷點，此點以上的因素表示共同因素，以下的因素則是唯一(unique)的因素，不予採用。

目視陡坡圖的方法有其缺點，一是，此法是用來決定母群相關係數矩陣之精確(exact)因素數目；二是，當採用樣本資料作分析時，難以決定決斷點；三是，由於涉及主觀的判斷與研究者經驗之多寡，因而造成因素數目頗不一致的情形。研究(Zwick & Velicer, 1982, 1986)採五種方法決定成份數目（卡方檢定、MAP、PA 此三項下段有簡述、特徵值大於 1 與陡坡圖等五法），結果顯示，利用陡坡圖所得的結果比用特徵值大於 1 或 Bartlett 卡方考驗方法來得正確，換言之，當資料的結構適當（如大樣本、變項之共同性高）與具經驗的研究者作判斷，則陡坡圖不愧是一種更簡單且較其他方法更能提供準確因素數目的方法(Gorsuch, 1983; Hakstian et al., 1982; Tzeng, 1992)

除了上述二種方法外，尚有 Bartlett 的卡方檢定(Chi-square test)、Velicer 的 MAP 法 (minimum-average partial procedure)、Horn 的 PA(parallel analysis)(Velicer & Jackson, 1990)、迴歸為主的方式(regression-based procedures)

(Nasser, 1997,引自 Benson & Nasser, 1998)，後三種方法無法自現成的統計軟體中取得，故使用上較不普遍。

事實上，抽取多少個因素，而後加以轉軸，頗有爭議(Fava & Velicer, 1992a; Wood, Tataryn & Gorsuch, 1996)，抽取太少的因素，可能遺漏了重要的因素，而抽取太多的因素，轉軸後可能有些因素只有一個或二個變項加以界定，有失妥當。

### (三)、轉軸

轉軸之主要目的為協助因素更具心理意義(psychological meaningfulness)的解釋，亦即達成簡單結構(simple structure)的原則，最常使用的方法為正交轉軸(orthogonal rotation)，部分原因為它是多數統計軟體中的內設選項，部分原因為正交轉軸之結果簡單，易於解釋，認為因素間是沒有相關的，斜交轉軸(oblique rotation)之結果會產生三種矩陣：因素結構矩陣(factor structure matrix)、因素組型矩陣(factor pattern matrix)和因素相關矩陣(factor correlation matrix)，結果解釋上不若正交轉軸之簡易。

然而某些研究者主張斜交轉軸，如 Reise, Waller 和 Comrey(2000)舉出五項考慮使用斜交轉軸之原因：1.若執行斜交轉軸，則可再進行高階(higher-order)之分析；2.斜交轉軸之結果較正交更能符合簡單結構的準則；3.某些研究指出斜交轉軸產生之因素，其複製性較優，亦即另以一個類似之樣本作分析，易於得到相同之因素結構；4.認定任何心理變項間沒有相關是不合理的，因而斜交轉軸似乎較能反映真實的心理現象；5.斜交轉軸下，因素間的相關可以估計而不若正交轉軸之相關被設定為零，因素間的相關可以提供有價值的訊息。總之，選擇何種轉軸方式，研究者必須以理論為規準。

### (四)、普遍使用之因素分析之步驟

主成份分析 → 特徵值大於 1 → 正交轉軸（最大變異法）(varimax)。國內

外研究似乎普遍存在此種現象。

Ford 等人(1986)分析三種期刊(1975~1984)之 152 個使用因素分析研究之結果，發現不少研究者以下述之步驟作分析：

- 1、使用主成份分析。
- 2、抽取特徵值大於 1 之成份。
- 3、正交轉軸。
- 4、通常將負荷量大於.30 或.40 的變項加以解釋。

執行因素分析若以上述之四個「自動式的決定」(automatic decision)，易引起錯誤的結果(Comrey 1978; Gorsuch, 1997; Lee & Comrey, 1979; MacCallum, 1983),Comrey(1978)不鼓勵以 PCA 方式，取特徵值大於 1 之成份作轉軸，理由如下：高估了共同性、抽取了太多的因素與扭曲了轉軸後的結果。Armstrong(1967,引自 MacCallum,1983)將模擬之資料以上述之四個步驟分析後，所得結果與已知的結構不合，而 MacCallum(1983)將其資料重新分析一次，但改用共同因素分析模式、檢視陡坡圖和利用斜交轉軸，發現結果與已知之結構吻合。

上述之 1 和 2 項已於前文敘述過，而第 3 項之正交轉軸為多數研究者所採用，有些研究者即使認知了因素間有相關，但仍執行正交轉軸，以致影響研究的結果(Ford et al., 1986)，第 4 項之顯著負荷量數值之界定，若以負荷量大於.30 之變項來界定因素，而負荷量為.29 者，卻不為所用，這似乎有點武斷，這和其他數值為決斷點.40 者情況亦同。

## 八、實例分析

本文以 PCA 與 CFA (均執行正交之 varimax 轉軸)分析實際資料，將結果作說明與比較 (PFA 為 CFA 常用的方法，本文採 PFA 方法)。資料為高雄市國小四年級與六年級 584 名學生在「兒童情緒智力量表」(31 題，五個分量表)的作答情形 (劉清芬，民 88)。資料以 SAS 軟體分析，分為二部分加以說明：(一)採 PCA 與 CFA，由統計軟體之內設(default)指令分析，結果作比較；(二)

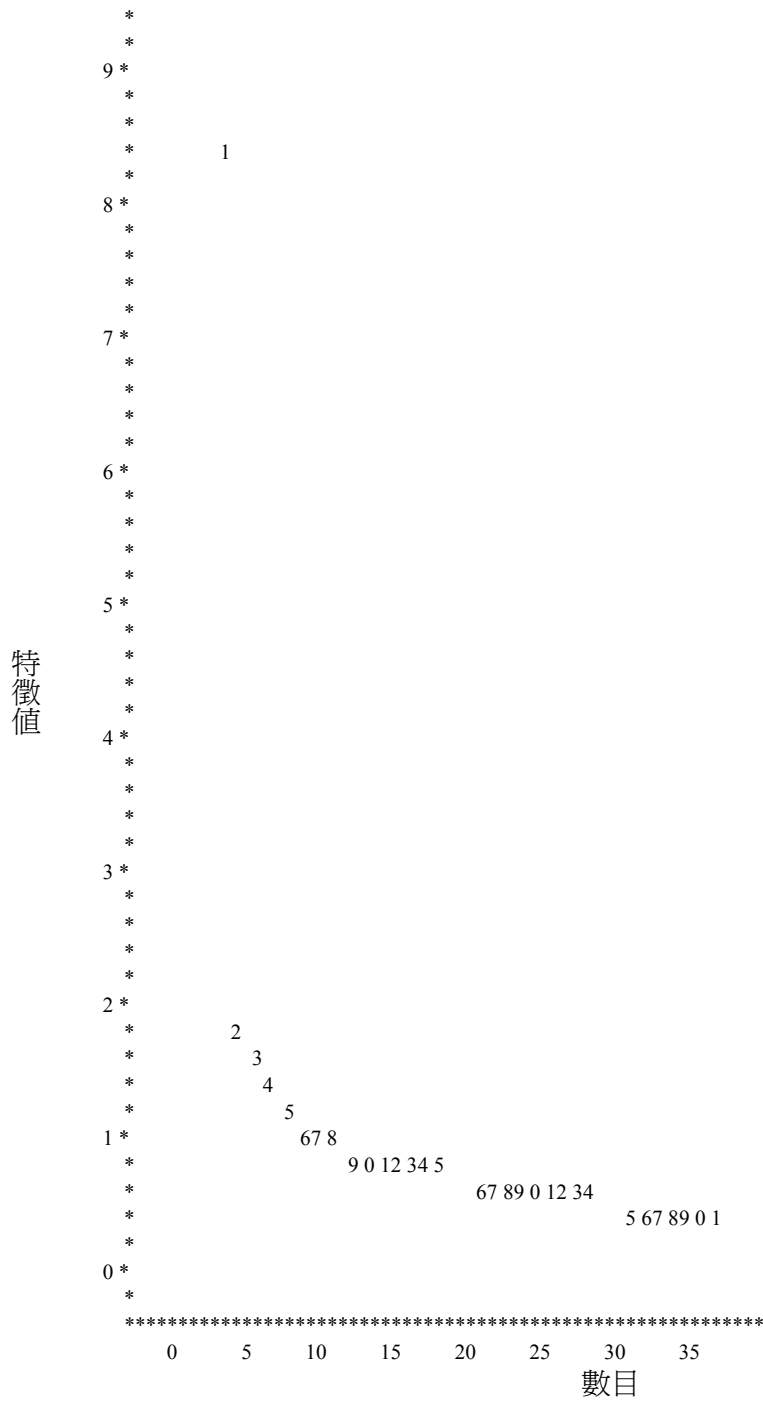
採 PCA 與 CFA，指定五個因素，再以正交轉軸最大變異法(varimax)分析，結果作比較。

### (一)、由統計軟體之內設指令分析結果作比較

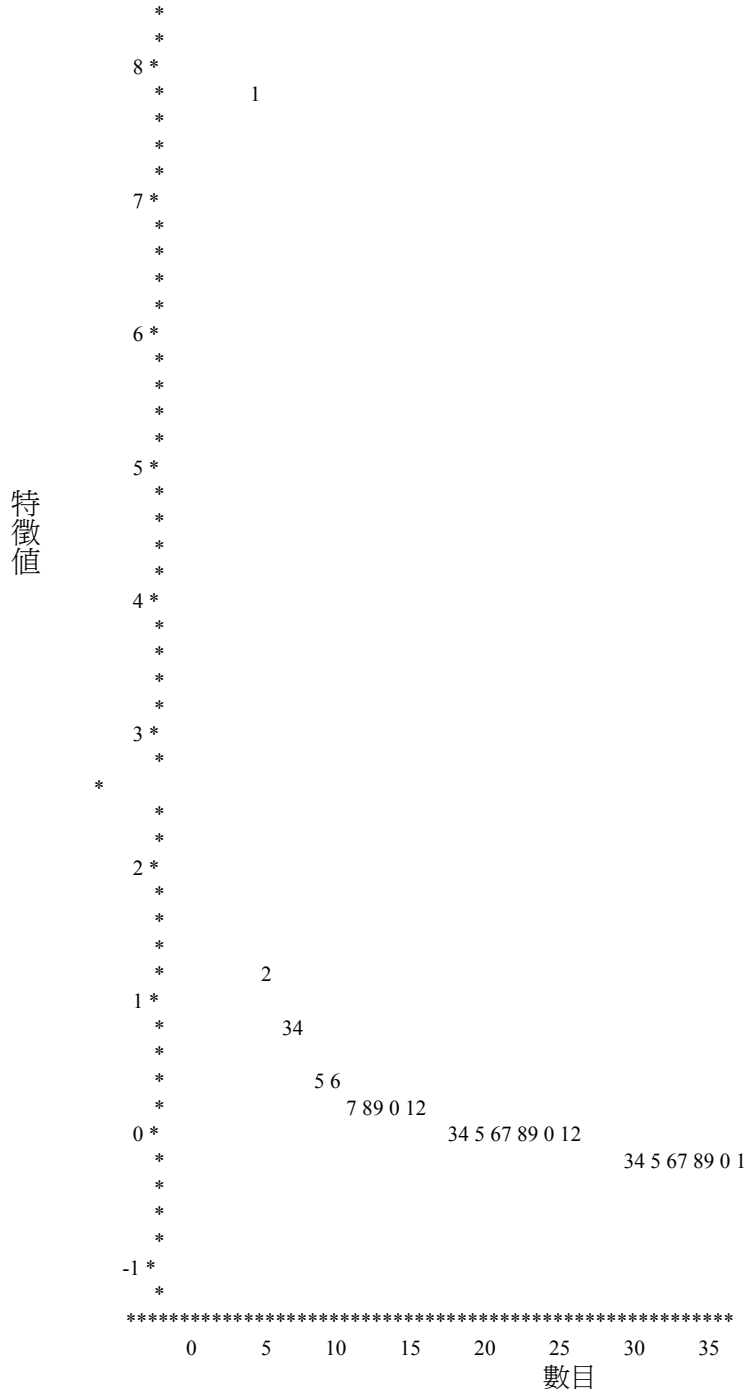
當資料採 PCA 與 CFA 分析時，研究者由電腦內部的運作自動來決定因素數目，正交轉軸分析結果如下：

- 1、在 PCA 法下，得到六個大於 1 的特徵值，由大至小依序如下：8.4146；1.8315；1.5140；1.3481；1.1906；1.0258；第七、八個特徵值分別為 0.9622 與 0.9203，與第六個特徵值差異微小，但當電腦軟體以特徵值大於 1 的標準自動取決時，第七個因素就不為採用，本文亦曾敘述此問題。
- 2、在 CFA 法下，電腦自動以比例(proportion)來決定因素數目，非以特徵值大於 1 的方式，分析結果取得四個因素，其特徵值分別如下：7.7873；1.2145；0.8215 與 0.7178，因其累積比例已達 1.0153，故只選四個因素，第五個特徵為.4621，其餘之特徵值愈來愈小，亦出現負的特徵值，直至第 31 個特徵值時，累積比例回復為 1，CFA 分析結果會產生此種現象，而 PCA 之分析，所有特徵值均為正值（因篇幅所限，未能列出所有數值）。
- 3、PCA 與 CFA 法下所得之陡坡圖可知，圖一(PCA 法)中，可採五為決斷點，取五個因素，而圖二(CFA 法)中，五或六為決斷點，取五或六個因素，這與從上述之數值分析結果不一致(PCA 法時取六個因素，而 CFA 法時取四個因素)，此時研究者可藉理論或驗證性因素分析(confirmatory factor analysis)來加以定奪，而非盲目由電腦分析結果來決定因素數目。

圖一、PCA 特徵值之陡坡圖



圖二、CFA 特徵值之陡坡圖





## (二)、採 PCA 與 CFA 法之結果比較

採 PCA 與 CFA 法取五個因素，再以正交之最大變異法分析，結果之比較如下：

- 1、就各變項之負荷量與共同性而言，表一與表二分別為 PCA 與 CFA 正交轉軸後之組型矩陣(rotated factor pattern)，表一(PCA)中各因素下之變項，其因素負荷量均高於表二(CFA)中相對應之變項（如表一中 EQ4 之負荷量為.69367，表二中 EQ4 之負荷量為.58339），且各變項之共同性亦具有此特性（如表一之 EQ4 共同性為.52312，表二中之 EQ4 共同性為.38807）。
- 2、就轉軸前之特徵值與轉軸後之負荷量平方和(sum of square loading; SSL)在 PCA 與 CFA 上之差異（表三）。

表一、PCA 正交轉軸後之組型矩陣

	因素一	因素二	因素三	因素四	因素五	共同性
EQ8	0.69367	0.11922	0.09113	0.09449	0.05189	0.51532
EQ5	0.66564	0.10955	0.14118	0.19026	-0.03039	0.51214
EQ4	0.65965	0.20821	-0.06621	0.01081	-0.20031	0.52312
EQ7	0.65083	0.14062	0.17749	0.11886	0.10292	0.49958
EQ14	0.64916	0.15061	0.07113	0.13629	0.12068	0.48229
EQ20	0.49227	0.20759	0.21674	0.14384	0.29531	0.44030
EQ27	0.46313	0.21857	0.26460	0.15580	-0.30164	0.44754
EQ22	0.45579	0.42005	0.24505	0.12242	0.04324	0.46109
EQ17	0.44019	0.22409	0.14757	0.09860	0.20874	0.31905
EQ9	0.43808	0.18855	-0.02900	0.25030	0.25202	0.35448
EQ11	0.43094	0.09754	0.14425	0.36371	0.28472	0.42938
EQ24	0.36465	0.18666	0.32354	0.22799	0.27460	0.39987
EQ25	0.14457	0.75552	0.10822	0.10475	0.03729	0.61578
EQ28	0.24911	0.63390	0.18864	0.04878	-0.12003	0.51625
EQ31	0.21574	0.61126	0.23098	0.12538	0.08377	0.49628
EQ13	0.13214	0.59323	0.13080	-0.00181	0.25999	0.45409
EQ16	0.01018	0.57910	0.17901	0.16413	0.28400	0.47510
EQ30	0.21521	0.56692	0.26582	0.13278	-0.04198	0.45777
EQ10	0.24592	0.52448	-0.27714	0.20774	-0.12153	0.47029
EQ23	0.32909	0.49159	0.30594	0.08560	0.08586	0.45826
EQ29	0.24085	0.26783	0.77031	0.06813	-0.03959	0.72933
EQ21	0.16358	0.22422	0.70820	0.17739	-0.01578	0.61029
EQ15	0.13050	0.24748	0.63639	0.12943	0.04597	0.50213
EQ18	0.12611	0.24843	-0.00370	0.62409	0.05959	0.47068
EQ6	0.24933	0.23157	0.00569	0.61228	0.00956	0.49080
EQ1	-0.00609	0.05511	0.11673	0.58465	-0.03697	0.35988
EQ3	0.19723	-0.02821	0.01493	0.55441	0.19533	0.38544
EQ2	0.13062	-0.00156	0.24896	0.55427	0.01576	0.38650
EQ26	0.18026	0.27486	0.17693	0.35616	-0.24860	0.32800
EQ12	0.17960	0.05580	-0.19919	-0.04363	0.55246	0.38216
EQ19	0.01969	0.10900	0.24785	0.17177	0.47164	0.32565
SSL	4.236466	3.774136	2.493384	2.450559	1.344293	14.2988

註：SSL：負荷量平方和

表二、CFA 正交轉軸之組型矩陣

	因素一	因素二	因素三	因素四	因素五	共同性
EQ4	0.58339	0.20424	-0.00551	0.06846	-0.03596	0.38807
EQ8	0.58162	0.13698	0.12448	0.14249	0.19295	0.43008
EQ5	0.57747	0.13090	0.15546	0.22280	0.10891	0.43628
EQ7	0.54206	0.16317	0.18997	0.16861	0.20588	0.42736
EQ14	0.53182	0.16407	0.10976	0.17289	0.24802	0.41321
EQ27	0.43460	0.22223	0.23523	0.19282	-0.15813	0.35578
EQ22	0.40555	0.39591	0.24831	0.16739	0.11683	0.42454
EQ20	0.37708	0.21422	0.22728	0.18878	0.33627	0.38845
EQ17	0.35370	0.22626	0.15675	0.15896	0.18585	0.26068
EQ9	0.34070	0.17836	0.04750	0.25126	0.25581	0.27871
EQ25	0.15026	0.67793	0.14110	0.12579	0.08242	0.52470
EQ28	0.26935	0.56601	0.19369	0.09174	-0.06245	0.44275
EQ31	0.20006	0.55127	0.23657	0.15621	0.12442	0.43977
EQ30	0.22703	0.50824	0.25365	0.16071	0.00231	0.40002
EQ13	0.12337	0.49324	0.16341	0.04960	0.22193	0.33692
EQ160	0.01845	0.48399	0.19193	0.17653	0.22415	0.35283
EQ23	0.29523	0.46218	0.29365	0.13208	0.12273	0.41950
EQ10	0.24437	0.38569	-0.10650	0.18144	0.00859	0.25281
EQ29	0.22448	0.26699	0.71543	0.11847	-0.04239	0.64935
EQ21	0.14539	0.22739	0.62103	0.20842	0.00082	0.50197
EQ15	0.12205	0.24384	0.53651	0.16539	0.03754	0.39096
EQ24	0.27494	0.20434	0.28923	0.25103	0.28542	0.34549
EQ6	0.22724	0.21246	0.05090	0.52625	0.06230	0.38019
EQ18	0.12377	0.21199	0.05230	0.49292	0.11408	0.31898
EQ3	0.15921	0.01817	0.03803	0.43351	0.14453	0.23594
EQ2	0.12319	0.05811	0.18224	0.42694	0.04906	0.23645
EQ1	0.03919	0.06855	0.10607	0.40237	-0.00524	0.17941
EQ11	0.31975	0.11940	0.16201	0.35286	0.32542	0.37315
EQ26	0.20322	0.24105	0.16218	0.30229	-0.12371	0.23239
EQ12	0.11108	0.04211	-0.08586	0.02974	0.22882	0.07473
EQ19	0.03617	0.12490	0.17510	0.17505	0.18331	0.11182
SSL	3.102849	3.060661	2.032992	1.927915	0.878854	11.0032

註：SSL：負荷量平方和。

表三、PCA 與 CFA 轉軸前之特徵值與正交轉軸後之負荷量平方和之比較

		因素					和	%
		1	2	3	4	5		
轉軸前	PCA	8.4146	1.8315	1.5140	1.3481	1.1906	14.2988	46.12
特徵值	CFA	7.7873	1.2145	.8215	.7178	.4621	11.0032	35.49
轉軸後	PCA	4.2365	3.7741	2.4934	2.4506	1.3443	14.2988	46.12
	SSL	3.1028	3.0606	2.0329	1.9279	.8788	11.0033	35.49

註：1、%表示：和÷31(14.2988/31=46.12)。

2、SSL：負荷量平方和。

由表三中可知：

- (1) PCA 法中之五個因素不論轉軸前後所能解釋總變異的百分比要比 CFA 法來得高(46.12%與 35.49%)。
- (2) PCA 法下，轉軸前之五個特徵值均大於 CFA 法，轉軸後之情況亦如此。

綜合上述，會產生數值的差異，是因 PCA 分析時認為變項之變異量中不含誤差，而 CFA 中則認為應顧及誤差。

再者，從分析中顯示，PCA 法再正交轉軸，其 root-mean-square error(RMSR)為.0518,大於 CFA 法再正交轉軸所得之 RMSR(.0272)，RMSR 數值較小者，顯示具較佳的適配(Nunnally & Bernstein, 1994)。

綜合言之，本研究之資料分別以 PCA 與 CFA 分析，結果大致相同，但仍仍有相異處，即某些因素所屬的變項數目不同，可能是變項間的相關係數低者居多，變項之共同性不高，再者，變項數目不多，亦是原因之一。

## 九、結語

二十世紀下半期，許多研究藉著因素分析以瞭解變項間相關的組型，達成資料的縮減、資料的轉換或假設的檢定等目的，當研究者之旨趣在於潛在變項間的關係時，理應使用共同因素模式，但多數人卻過度依賴成份模式，亦即將相關係數矩陣主對角線之數值以 1 取代共同性估計值。多數研究指出，當變項

數目少於 30 個（或 40 個，若共同性低時），選擇主成份分析或共同因素分析是個重要的抉擇，但當變項數目更多時，選擇何種方法抽取因素不是如此重要，但當變項數目少時，用共同因素分析較宜，因為此法所產生之負荷量較保守且較精確，不若主成份分析提供高估之數值。

雖然實徵研究顯示，在變項數目不少和共同性皆不低的情況下，二者多數是得到近似的結果，但有一點研究者仍須注意的是：因素模式的選擇與研究目的是有關的，不能例行地將主成份分析當成共同因素分析來使用，任由電腦統計軟體之內設選項來決定因素的數目，一味使用正交轉軸，因而理論的依據是不可或缺的。

## 參考文獻

- 林清山(1990)。《多變項分析統計法》。台北：東華書局。
- 劉清芬(1990)。「國小學生批判思考、情緒智力與學業成就之相關研究」。《國立高雄師範大學教育系碩士論文》。
- Acito, F., & Anderson, R. D. (1980). A Monte Carlo comparison of factor analytic methods. *Journal of Marketing Research*, 17, 228-236.
- Benson, J., & Nasser, F. (1998). On the use of factor analysis as a research tool. *Journal of Vocational Education Research*, 23(1), 13-33.
- Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research*, 25(1), 67-74.
- Borgatta, E. F. (1989). A note on using factor analysis and other procedures in research: A reply to Wilkinson. *Sociological Methods and Research*, 17(4), 460-464.
- Borgatta, E. F., Kercher, K., & Stull, D. E. (1986). A cautionary note on the use of principal component analysis. *Sociological Methods and Research*, 15(1-2), 160-168.
- Cattell, R. B. (1988). The meaning and strategic use of factor analysis. In J. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 131-203). New York: Plenum Press.
- Cliff, N. (1988). The eigenvalue-greater-than-one rule and the reliability of components. *Psychological Bulletin*, 103, 276-279.
- Comrey, A. L. (1978). Common methodological problem in factor analytic studies. *Journal of Consulting and Clinical Psychology*, 46(4), 648-659.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). New York: Academic Press.
- Daniel, L. G. (1990, November). *Common factor analysis or component analysis: An update on an old debate*. Paper presented at the Mid-South Educational Research Association, New Orleans. (Document Reproduction Service ERIC No. 325 531).
- Fava, J. L., & Velicer, W. F. (1992a). The effect of overextraction on factor and component analysis. *Multivariate Behavioral Research*, 27(3), 387-415.
- (1992b). An empirical comparison of factor, image, component, and scale scores. *Multivariate Behavioral Research*, 27(3), 301-322.
- (1996). The effects of underextraction in factor and component analyses. *Educational and Psychological Measurement*, 56, 907-929
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*,

39, 291-314.

- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- (1988). Exploratory factor analysis. In J. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 231-258). New York: Plenum Press.
- (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, 25(1), 33-39.
- (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68(3), 532-560.
- Hakstian, A. R., Rogers, T., & Cattell, R. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, 17, 193-219.
- Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.
- Hubbard, R., & Allen, S. J. (1987). A cautionary note on the use of principal components analysis: Supportive empirical evidence. *Sociological Methods and Research*, 16(2), 310-308.
- Jackson, J. E. (1991). *A user's guide to principal components*. New York: John Wiley & Sons.
- Joreskog, K. G., & Sorbom, F. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt.
- Lee, H. B., & Comrey, A. L. (1979). Distortions in a commonly used factor analysis procedure. *Multivariate Behavioral Research*, 14, 301-321.
- Loehlin, J. C. (1990). Component analysis versus common factor analysis: A case of disputed authorship. *Multivariate Behavioral Research*, 25(1), 29-31.
- MacCallum, R. (1983). A comparison of factor analysis program in SPSS, BMDP, and SAS. *Psychometrika*, 48(2), 223-231.
- McArdle, J. J. (1990). Principles versus principals of structural factor analysis. *Multivariate Behavioral Research*, 25(1), 81-87.
- Mulaik, S. A. (1990). Blurring the distinctions between component analysis and common factor analysis. *Multivariate Behavioral Research*, 25(1), 53-59.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12(3), 287-297.
- Sharma, S. (1996). *Applied multivariate techniques*. New York: John Wiley & Sons.
- Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, 106(1), 148-154.
- Tzeng, O. S. (1992). On reliability and number of principal components together with Cliff and Kaiser. *Perceptual and Motor Skill*, 75, 929-930.
- Velicer, W. F., & Jackson, D. N. (1990a). Component analysis versus common factor analysis:

- Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25(1), 1-28.
- (1990b). Component analysis versus common factor analysis: Some further observations. *Multivariate Behavioral Research*, 25(1), 97-114.
- Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor pattern: A Monte Carlo approach. *Multivariate Behavioral Research*, 17, 371-388.
- Widaman, K. F. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioral Research*, 25(1), 89-95.
- (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28(3), 263-311.
- Wilkinson, L. (1989). A cautionary note on the use of factor analysis: A response to Borgatta, Kercher, and Stull, and Hubbard and Allen. *Sociological Methods and Research*, 17(4), 449-459.
- Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis varimax rotation. *Psychometrical Methods*, 1(4), 354-365.
- Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, 17, 253-269.
- Zwick, W. R., & Velicer, W. F. (1986). A comparison of five rules for determining the number of factors to retain. *Psychological Bulletin*, 99, 432-442.



# Issues Related to Principal Component Analysis and Common Factor Analysis

Tsuey-shing Fu

National Kaohsiung Normal University Department of Education

The principal component analysis (PCA) and common factor analysis (CFA) are the most basic and frequently used factor analytic models. In fact, PCA and CFA are two different procedures. Their goals are also divergent. That is, CFA is used to extract as many factors as necessary to explain the correlations among the variables. On the other hand, PCA is meant to create summaries of variables.

This paper consists of eight main parts: (1) to introduce the purpose of this study; (2) to review the basic concepts of PCA; (3) to review the basic concepts of CFA; (4) to compare the characteristics of PCA and CFA; (5) to describe the similarity between PCA and CFA based on empirical researches; (6) to discuss the reason that CFA is more appropriate than PCA; (7) to present the common methods for using factor analysis; and (8) to display an example by using PCA and CFA.

The critical decisions in the selection of factor analysis are to understand the purpose of the study and to determine the number of factors to extract. The results of extraction and rotation are affected by these decisions. Users should not rely on the default options provided by computer programs.

keywords: common factor analysis; factor extraction; principal component analysis; rotation; scree plot