

次序性評定資料的 信度分析和因素分析

余 民 寧

(作者為政大教育系副教授兼實驗小學校長)

摘 要

本文的目的旨在應用一種稱作同質性信度係數的量化指標，作為分析次序性評定資料的統計學方法。同質性信度係數的值域介於0與1之間，其值愈大，代表測驗或量表具有較高的內部一致性，或其內容特質被判定為同質性的程度較高，反之，則否。本文亦舉例說明和建議使用同質性信度係數，作為評估次序性評定資料的內部一致性之參考指標，並用作資料分析（如：信度分析和因素分析）的內涵，以作為判定測驗或量表工具是否適當之依據。

壹、導 論

目前，在社會科學的研究領域中，尤其是在心理與教育的研究上，研究者常常使用測驗、量表或問卷作為收集資料的工具。然而這些收集資料的工具，多半是以 Likert 式的五點評定量尺 (Likert-type rating scale) 作為其計分的主要方式。為了確保所收集的資料具有可靠性和正確性，研究者常常會採用兩種統計方法來進行篩選、精簡或考驗其資料：一為信度分析 (reliability analysis)，另一為因素分析 (factor analysis)。因此，這兩種分析方法一直是傳統上研究者所奉為圭臬的標準作法。

在實際作法上， n 位受試者依據 C 個 (如以五點量尺而言，則 $C=5$) 類別的評定等級，在 m 個試題 (或變項) 上進行評定結果，所得的資料是一個 $n \times m$ 階的次序性評定資料 (ordinal rating data) 矩陣，矩陣中的每個元素值是這 C 個連續性整數之一，正值或負值皆可使用 (例如：在 Likert 式五點評定量尺上，這 C 個連續性整數可以是 1,2,3,4,5 或 0,1,2,3,4 或 -2,-1,0,1,2 等皆可)。接著，研究者以此資料作為統計分析的輸入資料，輸入電腦，並且以通行的統計套裝軟體程式 SPSSX(1990)、SAS(1990) 或 BMDP(1990) 中信度分析和因素分析程式，來進行資料分析，之後，並進行分析結果的解釋。

在進行信度分析的過程中，研究者最常使用的信度指標，便是 Cronbach(1951) 的 α 係數，它是一種內部一致性 (internal consistency) 信度，其值介於 0 與 1 之間。 α 係數愈接近於 1，則表示該測驗的內部一致性信度愈高；反之，若 α 係數愈接近於 0，則表示該測驗的內部一致性信度愈低。當然，研究者多半是希望獲得一個較高的 α 係數值，以便證實他所使用的工具是可靠的。

在計算 α 係數時，研究者通常必須根據其所收集到的 $n \times m$ 階的矩陣原始資料，先計算出每個試題的變異數 (即 S_i^2) 和測驗總分的變異數 (即 S_x^2) 之後，再代入下列公式求得 α 係數：

$$\alpha = \frac{m}{m-1} \left[1 - \frac{\sum S_i^2}{S_x^2} \right] \quad (\text{公式 1})$$

α 係數通常被看成是各種信度估計值的下限 (lower bound) (Hopkins, Stanley & Hopkins, 1990; Lord & Novick, 1968)。由於研究者是使用原始資料作為計算 α 係數的單元，並沒有考慮每位受試者在每個試題上作答的誤差大小，僅以單一個 α 係數作為表示整份測驗信度的指標，所以 α 係數僅適用於整份測驗，而不適

用於個別的試題。研究者若想得知每個試題的信度大小，只有求諸他法，這是使用 α 係數的一個限制。此外， α 係數一如其他信度係數一樣，容易受測驗長度、受試者的變異、測驗難度、測驗客觀性和估計方法的影響（郭生玉，民 79；陳英豪、吳裕益，民 80；葉重新，民 81；黃安邦，民 80），這是它的第二個限制。第三個限制是，研究者似乎不太明瞭 α 係數的統計特性，例如： α 成什麼樣的曲線分配？有何統計方法可以考驗它的顯著性？以致於國內學者或測驗方面的教科書都未提及。這些缺失，促使本文作者企欲找尋 α 係數的代替品，和新的分析技術，以彌補使用 α 係數時所遭遇的限制。

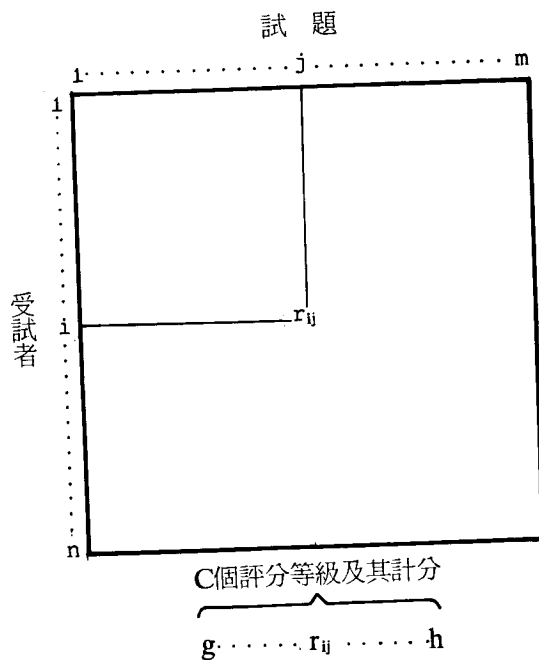
其次，在進行因素分析的過程中，研究者最常使用的步驟是：(1)以多元相關平方 (squared multiple correlation) 法來估計縮減式相關係數矩陣 (reduced correlation matrix) 的共同性 (communality)；(2)接者以主成份分析 (principal component analysis) 法來抽取因素；(3)再以直交轉軸的極變法 (varimax rotation) 進行轉軸，以求出最後的因素負荷量 (factor loading) 矩陣，完成因素分析的工作 (Afifi & Clark, 1990; Dunteman, 1984; Hair, Anderson, Tatham & Black, 1992)。然而，本文作者認為這種因素分析法所使用的仍然是 $n \times m$ 階的原始資料，其元素值僅是來自單一的作答資料，難免受受試者變異及測量誤差的影響很大，而失卻測量的正確性和可靠性。其次，當樣本數或試題數太少時（例如： $n < 5m$ ），所抽出的因素似乎不太穩定，違反因素分析的基本假設和限制，導致因素分析的結果變成廢物。因此，本文作者認為如能改採具有理論意義與統計特性的資料指標，來取代原始資料，則所用來進行因素分析的矩陣資料元素值，會更具有方法論上的意義和特性，所獲得之因素分析結果也會較傳統原始資料所獲得的結果，更具有實用性與理論基礎為背景的意義性。

基於對上述方法的反省與檢討，本文作者認為如果研究者能夠把類似上述所用的 $n \times m$ 階次序性評定資料的內部關聯性考慮在內，並把所使用的信度指標賦予統計機率分配的理論基礎，便能精確找出資料分析所建立的指標在統計學上的真正涵義，以解決或彌補傳統上使用這兩種分析方法的缺憾或不足。本文的目的，便是要依據常用的兩種機率分配理論：(1)適用於小樣本的多項式機率分配 (multinomial probability distribution) 和二項式機率分配 (binomial probability distribution)，和(2)適用於大樣本的常態曲線機率理論 (normal curve probability theory)，提出另一種內部一致性信度係數指標——稱作同質性信度係數 (homogeneity reliability coefficient)，和一種新的資料分析程序，以作為改進或另一種新的分析次序性評定資料的方法，來達到以簡馭繁的統計學目的，並增加量化方法論上的相關文獻。

貳、同質性信度係數之計算

本節的目的，旨在提出一種與內部一致性信度概念相通的信度係數，稱作同質性信度係數，作為判斷測驗試題是否值得信賴的參考指標。

本節所談論的同質性信度係數，其適用對象僅限於次序性評定資料，例如：Likert 式評定量尺、語意分析量尺（semantic differential rating scale）、或二分類評定量尺（dichotomous rating scale）等反應資料。所擬使用的基本假設符號及資料定義為： n 位受試者依據某個具有 c 個評定等級（或 c 個次序性類別）的量尺，在 m 個試題的測驗中作評定工作，這 c 個評定等級必須是連續性的整數，正值或負值皆可使用（例如：在 Likert 式五點評定量尺上，這 c 個連續性整數可以是 1,2,3,4,5 或 0,1,2,3,4 或 -2,-1,0,1,2 等）。其中，受試者 i 在試題 j 的評定結果，記作 r_{ij} ，而在 C 個等級的評分中， r_{ij} 的最小（或最左）值記作 g ，最大（或最右）值記作 h ， g 和 h 值的大小則由研究者根據所採行的量尺類別而自行設定（例如：若採行 1,2,3,4,5 的五點量尺的計分方式，則 g 值為 1， h 值為 5；若採行 -2,-1,0,1,2 的五點量尺的計分方式，則 g 值為 -2， h 值為 2）。茲將這種類型的次序性評定資料，圖示如下，以輔助說明：



接著，我們分成兩種情況來分別計算受試者的同質性信度係數（以 H_i 表示）和試題的同質性信度係數（以 H_j 表示）如下：

(一) n 位受試者分別對個別試題 j 之評分結果

首先，計算受試者兩兩間之評分差距的絕對值，再把總數為 $n(n-1)/2$ 個受試者間之評分差距的絕對值加起來，其值可用下列公式來表示：

$$S_j = \sum_{i'=1}^{n-1} \sum_{i=i'+1}^n |r_{ij} - r_{i'j}| \quad j=1, \dots, m \quad (\text{公式 2})$$

其次，計算試題 j 的同質性信度係數 H_j 如下：

$$H_j = 1 - \frac{4S_j}{(C-1)(n^2 - k)} \quad j=1, \dots, m \quad (\text{公式 3})$$

(二) 單獨一位受試者 i 分別對 m 個試題之評分結果

首先，計算試題兩兩間被評定之評分差距的絕對值，再把總數為 $m(m-1)/2$ 個試題間之評分差距的絕對值加起來，其值可用下列公式來表示：

$$S_i = \sum_{j'=1}^{m-1} \sum_{j=j'+1}^m |r_{ij} - r_{ij'}| \quad i=1, \dots, n \quad (\text{公式 4})$$

其次，計算受試者 i 的同質性信度係數 H_i 如下：

$$H_i = 1 - \frac{4S_i}{(c-1)(m^2 - k)} \quad i=1, \dots, n \quad (\text{公式 5})$$

其中，公式 3 和公式 5 中的 k 值是一種虛擬變數 (dummy variable)，當 n (或 m) 為偶數時， $k=0$ ；當 m (或 n) 為奇數時， $k=1$ 。

上述 H_j 和 H_i 的值皆介於 0 與 1 之間。 H_j 值愈大，表示試題 j 被一群受試者評定 (或判斷) 結果為一致的程度愈高； H_i 值愈大，則表示某一受試者 i 評定 (或判斷) 一群試題結果為相似的程度愈高 (余民寧，民 82; Aiken, 1980, 1985)。簡單的說， H_j 和 H_i 分別代表次序性評定資料中，試題被受試者評定為一致的傾向和受試者將試題評定為相似的傾向，而這個傾向可以 0 到 1 之間的一個指標值來加以表示。

根據機率的定義，如果每個評定等級被單獨挑選到的機會都相等的話，其機率值均為 $1/c$ ，則下列多項式機率分配的公式，可被用來計算 n 次評分中某個隨機組合 (即 n_1, n_2, \dots, n_c) 的機率：

$$P(n_1, n_2, \dots, n_c) = \frac{n!/C^n}{n_1!n_2!\dots n_c!} \quad (\text{公式 6})$$

其中， $n = n_1 + n_2 + \dots + n_c$ ， $n_k (k=1, \dots, c)$ 表示評選第 k 個評定等級（或次序性類別）的人數，全部的總合為 n 。當評定等級只有兩個時（即 $C=2$ ），公式 6 可以簡化成二項式機率分配的公式：

$$P(n_1, n_2) = \frac{n!/c^n}{n_1!n_2!} \quad (\text{公式 7})$$

因此， n 位受試者在具有 c 個評定等級的試題 j 上的評分組合，均可根據公式 2 和公式 3 算出其相對應的 H_j 值和其機率的大小。由於許多種不同的評分組合都可能產生等值的 H_j ，我們可將這些等值的 H_j 所對應的機率值加總起來，結果就可以獲得一個大約或等於所有可能的 H_j 值佔整個機率分配中的右側機率（right-tail probability）大小的參考表（參見附錄一），這個參考表是理論上的機率分配表，我們可以拿它和實際計算而得的 H_j 值作對照，以檢定我們所實際求得的 H_j 值是否有達理論上的顯著水準。

為了說明附錄一的製作過程，茲舉 $n=4$ 和 $c=3$ 的一個試題為例，該試題之計分依序為 1,2,3，如果受試者依序在第一個、第二個、第三個評定等級上作選擇的話。由於四個人 ($n=4$) 在三個等級 ($c=3$) 上的評分組型總共有十五種，我們分別列出這十五種組型的人數分佈型態，並算出每一種人數分佈型態之可能機率值及其評分組型下的 H_j 值，以及列出各種 H_j 值下的累積機率值，並列於表 1 作說明。

由表 1 可知，在 $n=4$ 和 $c=3$ 的情況下，出現 $H=1$ 的機率值，最多僅達 .037，正如附錄一中 $n=4$ 和 $c=3$ 下的 P 值所示者。附錄一所陳列者，僅為 2 至 7 個評定等級 (c) 和 2 至 25 個試題 (m) 或受試者 (n) 範圍內， H 係數的分類機率值中接近（但不大於）.05 和 .01 的右側機率值大小。附錄一僅提供小樣本時，檢定同質性信度係數是否達顯著水準的參照表（余民寧，民 82；Aiken, 1985），例如：假設有 10 位受試者在一個具有 5 個評定等級的試題上（或一位受試者使用一份具有 5 個評定等級的量尺來評定 10 個試題）之評分結果，該試題的同質性信度係數值 (H) 必須大於 .60 才算達到 $\alpha = .01$ 顯著水準，或大於 .51 才算達到 $\alpha = .05$ 的顯著水準。這種經過檢定，符合統計學上顯著水準意義的同質性信度係數值，才能用來認定該試題是個良好試題，可以被保留在測驗裡，作為有效測量工具的單元。

表1 四位受試者在三個評分等級量尺上之同質性信度係數及其累積機率值

人類分佈型態 ($n: 4$) (n_1, n_2, n_3)	評分組型 (r_1, r_2, r_3, r_4)	同質性信度係數	
		S_j	H_j
$P(0,0,4) = \frac{4!/3^4}{0!0!4!} = .0123$	(3,3,3,3)	0	$H=1 - \frac{(4)(0)}{(2)(16)} = 1$
$P(0,4,0) = .0123$	(2,2,2,2)	0	$H=1$
$P(4,0,0) = .0123$	(1,1,1,1)	0	$H=1$
$P(0,1,3) = .0493$	(3,3,3,2)	3	$H = .625$
$P(1,0,3) = .0493$	(3,3,2,2)	4	$H = .5$
$P(0,3,1) = .0493$	(3,2,2,2)	3	$H = .625$
$P(1,3,0) = .0493$	(3,3,3,1)	6	$H = .25$
$P(3,1,0) = .0493$	(3,3,1,1)	8	$H = 0$
$P(3,0,1) = .0493$	(3,1,1,1)	6	$H = .25$
$P(0,2,2) = .0741$	(2,2,2,1)	3	$H = .625$
$P(2,0,2) = .0741$	(2,2,1,1)	4	$H = .5$
$P(2,2,0) = .0741$	(2,1,1,1)	3	$H = .625$
$P(1,1,2) = .1481$	(3,3,2,1)	7	$H = .125$
$P(1,2,1) = .1481$	(3,2,2,1)	6	$H = .25$
$P(2,1,1) = .1481$	(3,1,2,1)	7	$H = .125$

註:以得分組型(3,1,2,1)為例, $S_j = |3-1| + |3-2| + |3-1| + |1-2| + |1-1| + |2-1| = 7$

H 的分類機率值 ($n=4, c=3$)

當 $H=0$ 時 $P = .0741$

$H = .125$ $P = .1481 + .1481 = .2962$

$H = .25$ $P = .0493 + .0493 + .1481 = .2467$

$H = .5$ $P = .0741 + .0741 = .1482$

$H = .625$ $P = .0493 + .0493 + .0493 + .0493 = .1972$

$H = 1$ $P = .0123 + .0123 + .0123 = .0369$

合計 $P = .9993$

當使用大樣本（如： n 或 $m > 25$ ）時，我們可以使用中央極限定理（central limit theorem）來協助決定 H_j 值的平均數的統計顯著性。這種作法很簡單，一般的數理統計學教科書（如：Bickel & Doksum, 1977; Hogg & Craig, 1978）都有說明這種考驗作法；亦即，我們要考驗這 m 個 H_j 值所構成的抽樣分配的平均數，是否達到某種顯著水準。

由於每位受試者都獨立評定每個試題，或每個試題均獨立地被單一受試者評分，因此可知這 m 個 H_j 值的母群體的期望值（ μ ）和標準差（ σ ）可以表示如下（Aiken, 1985）：

$$\mu = \frac{2(c+1) + (n+k)(c-2)}{3c(n+k)} \quad (\text{公式 8})$$

$$\sigma = \left(\frac{2}{3c}\right) \sqrt{\frac{.2(c+1)(n+k-1)[c^2(n+3) - 2(2n-9)]}{(c-1)(n+k)(n^2-k)}} \quad (\text{公式 9})$$

再以下列公式檢定 H_j 的平均數 \bar{H} 是否達顯著水準：

$$Z = \frac{(\bar{H} - \mu)}{\sigma_{\bar{H}}} = \frac{\sqrt{m}(\bar{H} - \mu)}{\sigma} \quad (\text{公式 10})$$

假若 Z 值大於所選定之臨界值（如： $\alpha = .05$ 時之 Z 值為 1.645，或 $\alpha = .01$ 時之 Z 值為 2.33），則可結論認為這整個測驗、量表、或問卷，具有顯著的同質性信度係數 \bar{H} ，亦即是這份工具的內涵是同質的，或值得信賴的。

若公式 10 的 Z 值是根據 n 位受試者所計算出的 H_j 時，其中的 m 必須換成是 n ，而公式 8 和公式 9 中的 n 換成是 m ；其餘符號則與前述定義者相同。

至此，我們總共可以計算出 m 個 H_j 係數值和 n 個 H_i 係數值，每個 H 係數均表示試題或受試者的內部一致性信度係數大小，其值愈大（即愈接近於 1），表示試題或受試者被評分或評定的結果愈趨一致或相似，反之，其值愈小（即愈接近於 0），則表示其被評分或評定的結果愈不一致或相異。

根據 H_j 所計算出的同質性信度係數 \bar{H} ，並非完全是試題間（或整份測驗）的內部一致性信度，而只是代表 n 位受試者在 m 個試題上之評分結果間的一致性程度而已，因此， Z 值（即由公式 10 計算而得者）的解釋方式應該與 Kendall(1970) 所提出的和諧係數（coefficient of concordance）的解釋方式相似（Aiken, 1985, 1989）

參、信度分析與因素分析的新作法

本節即根據上述所求得的 m 個試題的同質性信度係數 H_j ，來重提信度分析和因素分析的新作法。

一、信度分析的作法

依照上述程序，先計算出每個試題的 H_j 值，再求其平均數 \bar{H} ，再代入公式 10 加以檢定，以確定整份測驗的內部一致性信度（即同質性信度）是否達顯著水準。

在計算 \bar{H} 之前，我們亦可分別檢定每個試題的 H_j 值是否也達顯著水準，檢定的方法可以是查附錄一的表，或是代入公式 10 來考驗都行。若有試題的 H_j 值未達顯著者，則建議將其刪除，僅保留 H_j 值達顯著的試題。接著，再計算被保留下的試題 H_j 值的平均數，再代入公式 10 加以考驗。

由於這種新的程序作法，已將同質性信度低的試題予以刪除，所以，被保留下來的試題都是同質性信度高者。因此，所計算出足以代表整份測驗的內部一致性信度的同質性信度係數 \bar{H} ，應該會比以原始資料所計算出的 α 係數還大。我們也可以進一步考驗其差距（即 $\bar{H} - \alpha$ ）是否也達顯著差異，而判斷何種指標比較好。考驗的方式可以用 t 考驗來進行如下：

$$t = \frac{(\bar{H} - \alpha)}{\sigma_{\bar{H}}} = \frac{\sqrt{m} (\bar{H} - \alpha)}{\sigma} \quad (\text{公式 11})$$

公式 11 所計算出的 t 值，將與自由度為 $m-1$ 的查表 t 值作比較，以判斷它是否達到 .05 或 .01 的顯著差異，進而判斷何種指標（即 \bar{H} 或 α ）為優。

二、因素分析的作法

因為 H_j 是代表試題的同質性信度係數，屬於是內部一致性信度的一種，因此，它可以被用來當作因素分析中縮減式的相關係數矩陣的共同性，作為因素分析的起點資料矩陣。因為，共同性的概念和本質，相當於試題本身所具有的信度特性一般，信度被看成是共同性的上限（upper bound）（Tatsuoka, 1988, p.182），因此，以同質性信度係數來取代共同性，以作為進行因素分析的起點資料作法，不論在理論上或實用上而言，都算是合理的。

接著，依照傳統的作法，進行主成份分析抽取因素，再進行極變法的直交轉軸

，以獲得符合簡單結構原則（simple structure principle）的因素負荷量矩陣，再加以解釋結果。

由於這種新程序的作法，其共同性的使用，具有統計特性的理論背景意義，因此，所獲得的因素結構，將更富有解釋上的意義。

此外，由於具有不顯著的 H_j 值的試題已被刪除，因此，使用 H_j 值來作為共同性，可以事先作到初步精簡資料的目標，使得後續因素分析的工作得以節省運算的時間和成本。

肆、計算實例及其比較

為了能夠詳細說明同質性信度係數之計算過程，及其在信度分析與因素分析過程中之應用，並舉 10 個試題的簡式考試焦慮量表（余民寧，民 76）為例。本例子僅供作說明使用，不作為其他推論之用。

從國立政治大學教育研究所中等學校教師在職進修班的暑期班中，隨機抽取 25 名至少有五年以上輔導經驗的老師，就這 10 個試題在 5 點評定量尺上，分別評定每個試題能否有效測量到學生的考試焦慮，評定的計分方式為：評「無法有效測量」者，給 1 分；依序對次一等級評定者加 1 分；直到評「可以有效測量」者，則給 5 分。因此，這些計分值，分別是五個連續的正整數值之一，從 1 到 5 不等。這種資料型態，為 25×10 階的次序性評定資料矩陣，如表 2 的原始資料所示。茲援用此原始資料說明如下。

一、信度分析

首先，試算每個試題的 S_j 值。茲以第一題為例說明：

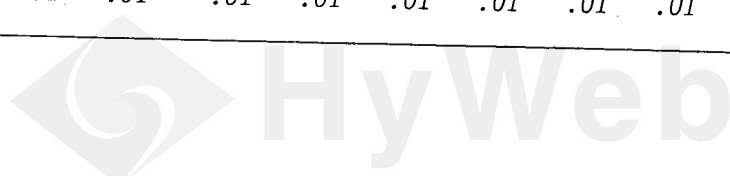
$$S_{.1} = \sum_{j'=1}^{24} \sum_{i=i'+1}^{25} |r_{ij'} - r_{ij}|$$

$$= |4-4| + |4-5| + \dots + |5-4| + |5-4| = 126$$

（共有 $25(25-1)/2 = 300$ 個差值的絕對值）

表2 25名教師在10個試題上之5點評定資料及其H值

試 題 評 分 者										
	1	2	3	4	5	6	7	8	9	10
1	4	3	4	5	5	4	4	4	5	4
2	4	4	4	4	3	4	4	5	5	4
3	5	4	4	4	5	4	4	5	4	4
4	4	3	3	4	4	3	4	4	4	4
5	4	4	4	4	4	4	4	4	4	4
6	4	4	4	4	4	4	4	5	4	4
7	4	3	4	5	4	4	4	4	4	4
8	5	3	4	3	4	3	5	4	4	4
9	4	3	4	4	3	4	3	4	3	4
10	4	4	4	5	5	4	4	5	4	4
11	4	3	4	5	4	4	5	4	3	3
12	5	4	5	5	5	4	5	5	5	4
13	4	4	3	4	4	4	4	5	4	4
14	5	5	5	4	4	3	4	4	5	4
15	4	4	3	4	5	5	4	5	4	5
16	4	4	4	5	5	4	4	4	4	4
17	4	4	4	5	5	4	4	4	4	4
18	5	4	3	4	4	4	3	4	4	4
19	4	5	5	4	4	4	5	4	3	4
20	4	4	4	4	5	4	5	5	4	4
21	4	5	4	4	5	5	5	4	4	4
22	4	4	4	5	4	4	4	4	4	5
23	5	5	4	3	4	4	3	4	5	4
24	5	4	5	4	3	4	4	4	4	4
25	4	4	3	4	5	4	4	5	4	4
H_j	.80	.68	.71	.71	.65	.82	.72	.77	.71	.89
$p <$.01	.01	.01	.01	.01	.01	.01	.01	.01	.01



其次，計算每個試題的 H_i 值。仍以第一題為例說明：

$$H_1 = 1 - \frac{4(126)}{(5-1)(25^2-1)} = .80$$

第三，查附錄一的表，得知 H_i 值為 .80，已達 $\alpha = .01$ 的顯著水準，該試題應予保留。同理，其餘九題的 H_i 值也可以算出，並且查表得知其顯著水準如表 2 所示。結果，全部試題都應予以保留，表示這些試題都具有頗高的同質性信度係數，皆能作為測量考試焦慮的有效工具。

第四，計算這些被保留試題的 H_i 值的平均數 \bar{H} ，並考驗其顯著性如下：

$$\bar{H} = (.80 + .68 + \dots + .71 + .89) / 10 = .750$$

$$\mu = \frac{2(5+1) + (25+1)(5-2)}{(3)(5)(25+1)} = .231$$

$$\sigma = \left[\frac{2}{3(5)} \right] \sqrt{\frac{.2(5+1)(25+1-1)[25(25+3) - 2(50-9)]}{(5-1)(25+1)(625-1)}} = .071$$

$$Z = \frac{\sqrt{10} (.750 - .231)}{.071} = 23.12 \quad (p < .01)$$

第五，計算表 2 所具有的 α 係數值為 .450。

第六，考驗 \bar{H} 和 α 之間的差異是否達顯。茲考驗如下：

$$t = \frac{\sqrt{10} (.750 - .450)}{.071} = 13.36 \quad (p < .01)$$

由此可見，表 2 中每個試題的同質性信度係數皆達 .01 的顯著水準，並且整份量表的同質性信度係數為 .750，亦達 .01 的顯著水準，可見這份量表具有顯著良好的內部一致性信度。不僅是每個試題，連整份量表都具有顯著的同質性信度指標，它們都可以作為收集考試焦慮反應的可靠工具。

由表 2 的原始資料亦可看出，由於每個試題被評定的分數都介於 3 到 5 之間，由於變異過小，所以計算出來的 α 係數自然偏低。但從相反觀點來看，正因為反應資料的變異小，這表示結果被評定為一致的程度偏高，亦即是試題的內容特質被判定為同質的成份居多數，評分者間之評定結果的一致性自然偏高所致。

根據公式 11 的差異顯著性考驗得知， \bar{H} 顯著的比 α 係數為優。這項結果顯示出，類似表 2 的次序性評定資料適合使用 \bar{H} 來當作內部一致性信度指標，而不適合使用 α 係數來當作指標。

在一般的人文及社會科學（尤以心理與教育學為最）的測量資料裡，資料通常是受試者在試題上作自陳式（self-report）的評定結果，是受試者根據本身的實況或標準作反應；而本研究表 2 所示的資料，乃受試者（亦即是評分者）根據外在的標準（如：專業知識、外在效標、或限定的評定量尺等）作客觀評定、判斷、或給分的評量工作，因此，資料並不是在測量受試者本身（而前者是在測量受試者本身），這一點必須加以分辨及說明。而從本例可知，當遇到的分析資料是屬於後者似的評定資料時，使用 \bar{H} 指標會比使用 α 係數來得恰當。

二、因素分析

由於這種資料不是自陳式的評定資料，因此在進行因素分析後，所抽取的因素並不代表該量表所測量到的受試者的共同特質，而是根據效標評定結果之一致性的集群（clusters）（亦即，將評定結果一致性高的試題集結成數群），這種集群將無法反映出受試者反應資料中的共同特質或屬性。因此，類似表 2 的次序性評定資料，將無法再以傳統的方式來進行因素分析，而需要以不同方式來進行因素分析，才能解決因素抽取後之代表性困境的難題。下列的因素分析程序，將可提供解決此一難題的途徑。

首先，以每個試題的同質性信度係數作為共同性，代入縮減式相關係數矩陣的對角線，以作為因素分析的起點資料矩陣。

其次，再以主成份分析方法抽取共同因素。

第三，再將抽取出的因素負荷量矩陣加以極變法（varimax）的直交轉軸，以期符合簡單結構的原則。

第四，將轉軸後之因素負荷量矩陣加以解釋。

為了說明起見，茲以表 2 的資料進行傳統的因素分析和新法因素分析，並比較說明其差異如下。

表3 傳統式因素分析轉軸後之因素負荷量矩陣及解釋率

試題	因素一	因素二	因素	特徵值	解釋變異數的百分比	累積的百分比
1	.8200	-.1393	1	2.203	22.0	22.0
4	-.5988	.0823	2	1.875	18.8	40.8
9	.5966	.3569				
2	.5256	.4987				
3	.3820	-.2167				
7	-.2365	.0509				
5	-.2714	.6492				
6	-.3457	.6407				
8	-.0695	.6247				
10	.1134	.5727				

由表3的傳統式因素分析結果得知，第1、4、9、2、3和7試題主要是測量第一個因素，而第5、6、8和10試題主要是測量第二個因素。由附錄二的原始量表可知，第一個因素是代表這25位教師評定結果較為一致所組成的一個集群或因素，可以暫定名為「主要的外顯特徵」，而第二個因素是代表另一個評定結果為一致的集群或因素，可以暫定名為「次要的內隱特徵」。所得之因素結構與原始量表所得之因素結構迥異，原始量表所得的第一個因素為「情緒化」反應，代表的測量試題為第1、3、5、7和9試題；而原始量表所得的第二個因素為「憂慮」反應，代表的測量試題為第2、4、6、8和10試題。由此可見，所得之結果，意義相差甚遠。

由表4大略可以得知，由於是以同質性信度係數作為共同性，所以所得之特徵值及其解釋變異數的百分比會較傳統式以主成份分析法為主之因素分析所得結果為低，但是表4所得之因素結構卻較接近原始量表所得的因素結構，除了第4和第5試題之評定結果較不一致外，表4比表3更接近原始量表的因素結構，也就是說，在類似表2的次序性評定資料下，新法的因素分析會比傳統式因素分析更為適當。

表4 新法因素分析轉軸後之因素負荷量矩陣及解釋率

試題	因素一	因素二	因素	特徵值	解釋變異數的百分比	累積的百分比
1	.8588	-.0801	1	1.630	16.3	16.3
9	.4473	.2961	2	1.206	12.1	28.4
4	-.4279	.0696				
3	.3126	-.1005				
7	-.1306	.0721				
5	-.2186	.5472				
6	-.3349	.5428				
8	-.0851	.4642				
2	.3832	.4559				
10	.0176	.3926				

伍、結 論

同樣是採用 Likert 式五點評定量尺，但所收集到的資料卻可分成兩類：一為自陳式的評定資料，另一為外評式的評定資料。前者適合傳統式的分析方法，如：使用 α 係數作為信度分析，或以主成份分析為主式的因素分析；而後者比較適合以同質性信度係數 \bar{r} 作為信度分析的依據，並以試題的同質性信度係數作為共同性，來進行新式的因素分析。

本文即考量傳統式的資料分析方法，可能無法勝任類似表 2 的次序性評定資料的分析，因此嘗試新的解決辦法。本文即根據所考量的傳統方法的缺失，同時亦兼顧所選定之量化指標背後所應有的統計學特性，介紹一種類似評分者信度的量化指標——同質性信度係數，作為內部一致性信度的另一種參考指標。同質性信度係數的值域介於 0 與 1 之間，其值愈大，代表測驗或量表具有較高的內部一致性，或其內容特質被判定為同質性的程度較高，反之，則否。

由於同質性信度係數具有嚴謹的機率分配等統計學理論背景，它比 α 係數較適用於外評式的次序性評定資料（如表 2 的資料所示）。同時，若以同質性信度係數

代替共同性來進行因素分析，其結果也會比傳統式（即以1來代替共同性）因素分析之結果，更接近原始測驗或量表的因素結構。上述結論已由舉例說明中，獲得支持。

最後，本文建議以同質性信度係數作為評估次序性評定資料的内部一致性之參考指標，並用作資料分析方法（如：信度分析和因素分析）的內涵，以作為判定測驗或量表工具是否適當之依據。



附錄一

同質性信度係數(H)之考驗的右側機率值對照表

試題數 (m) 或 受試者數 (n)	評 定 等 級 數 目 (C)											
	2		3		4		5		6		7	
	H	p	H	p	H	p	H	p	H	p	H	p
3							1.00	.040	1.00	.028	1.00	.020
4									1.00	.005	1.00	.003
4			1.00	.037	1.00	.016	1.00	.008	.85	.035	.83	.038
5					1.00	.004	1.00	.002	.87	.007	.89	.004
5			1.00	.012	.78	.033	.75	.040	.73	.035	.72	.050
6			1.00	.004	.81	.010	.86	.003	.78	.010	.81	.005
6	1.00	.031	.72	.037	.67	.046	.72	.024	.67	.041	.67	.046
7			1.00	.001	.72	.010	.75	.009	.73	.007	.72	.008
7	1.00	.016	.75	.014	.67	.030	.62	.041	.63	.036	.61	.050
8	1.00	.008	.78	.005	.71	.006	.67	.007	.67	.009	.69	.009
8			.56	.033	.56	.030	.56	.046	.57	.045	.57	.050
9	1.00	.004	.65	.009	.67	.007	.65	.009	.66	.007	.65	.009
9	.60	.039	.55	.031	.53	.047	.52	.048	.54	.047	.55	.046
10	1.00	.002	.64	.006	.60	.010	.60	.010	.62	.008	.62	.009
10	.64	.021	.52	.028	.48	.045	.51	.046	.51	.050	.53	.047
11			.60	.006	.58	.009	.58	.009	.59	.009	.59	.010
11	.67	.012	.50	.031	.49	.032	.48	.046	.51	.040	.50	.048
12	.69	.006	.56	.009	.55	.010	.56	.010	.57	.010	.57	.010
12	.44	.039	.44	.034	.44	.049	.47	.048	.48	.045	.49	.046
13	.71	.003	.52	.007	.52	.009	.55	.009	.55	.009	.56	.010
13	.48	.022	.43	.031	.43	.049	.45	.045	.47	.047	.48	.049
14	.73	.002	.51	.009	.51	.009	.52	.010	.53	.010	.54	.010
14	.51	.013	.41	.037	.42	.046	.44	.048	.45	.048	.46	.048
15	.54	.007	.48	.009	.50	.009	.51	.009	.52	.009	.52	.010
15	.36	.035	.39	.030	.40	.048	.43	.046	.44	.047	.45	.049
16	.56	.004	.48	.007	.48	.008	.49	.010	.51	.010	.51	.010
16	.39	.021	.38	.040	.39	.049	.41	.049	.43	.047	.44	.048
17	.58	.002	.47	.006	.46	.009	.48	.010	.49	.010	.50	.010
17	.28	.049	.35	.043	.39	.046	.41	.046	.42	.049	.44	.048
18	.44	.008	.42	.010	.45	.009	.47	.010	.48	.010	.49	.010
18	.31	.031	.33	.046	.37	.049	.40	.050	.41	.049	.43	.050
19	.47	.004	.41	.010	.44	.010	.46	.009	.48	.010	.49	.010
19	.33	.019	.33	.040	.37	.043	.39	.046	.41	.048	.42	.049
20	.49	.003	.40	.010	.43	.010	.45	.010	.47	.010	.48	.010
20	.25	.041	.32	.047	.36	.049	.38	.050	.40	.050	.41	.048
21	.38	.007	.39	.010	.42	.010	.45	.009	.46	.010	.47	.010
21	.27	.027	.33	.042	.35	.048	.38	.048	.40	.048	.41	.048
22	.40	.004	.37	.010	.41	.010	.44	.010	.45	.010	.46	.010
22	.30	.017	.30	.050	.35	.049	.37	.049	.39	.050	.40	.050
23	.42	.003	.37	.010	.41	.010	.43	.010	.45	.010	.46	.010
23	.23	.035	.30	.043	.34	.050	.37	.044	.38	.050	.40	.048
24	.34	.007	.36	.010	.40	.010	.42	.010	.44	.010	.45	.010
24	.25	.023	.30	.048	.34	.049	.36	.050	.38	.048	.39	.049
25	.36	.004	.36	.010	.39	.010	.42	.010	.43	.010	.44	.010
25	.19	.043	.29	.048	.33	.047	.36	.049	.38	.049	.39	.050



附錄二

考試焦慮量表

余民寧修定

中華民國七十六年六月

作答說明：請根據你在參加考試時的感受，以「✓」的方式，回答下列問題。

- | | 1.
從
未
發
生 | 2.
很
少
發
生 | 3.
偶
爾
發
生 | 4.
經
常
發
生 | 5.
一
直
發
生 |
|----------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1. 我會感覺到心跳得很厲害。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. 我會感到很後悔。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. 我會緊張，胃有不舒服的感覺。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 4. 我才覺得應該好好準備才對。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 5. 我會有不安，心煩的感覺。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 6. 我覺得可能別人對我的考試成績感到失望。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 7. 我會感到焦慮，手心流汗。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 8. 我會覺得沒有盡力把試題答好。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 9. 我會感到恐慌，不知如何是好。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 10. 我會覺得對這次考試沒有信心。 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

參考書目

- 余民寧。(民76)。**考試焦慮、成就動機、學習習慣與學業成績之研究**。國立政治大學教育研究所碩士論文。
- 余民寧。(民82)。**次序性資料的內容效度係數和同質性信度係數之計算**。**測驗年刊**，40輯，199-214頁。
- 郭生玉。(民79)。**心理與教育測驗**(五版)。台北：精華。
- 陳英豪、吳裕益。(民80)。**測驗與評量**(修訂一版)。高雄：復文。
- 黃安邦。(民80)。**心理測驗**。台北：五南。
- 葉重新。(民81)。**心理測驗**。台北：三民。
- Affifi, A. A., & Clark, V. (1990). *Computer-aided multivariate analysis* (2nd ed.). New York : Van Nostrand Reinhold.
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955-959.
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142.
- Aiken, L. R. (1989). Coefficients of response homogeneity and congruence. *Educational and Psychological Measurement*, 49, 321-324.
- Bickel, P. J., & Doksum, K. A. (1977). *Mathematical statistics: Basic ideas and selected topics*. Oakland, CA: Holden-day.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Dixon, W. J. (Ed.), (1990). *BMDP statistical software*. Berkeley, CA: University of California Press.
- Dunteman, G. H. (1984). *Introduction to multivariate analysis*. Beverly Hills, CA: SAGE.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). *Multivariate data analysis with readings* (3rd ed.). New York : Macmillan.
- Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.) New York : Macmillan.
- Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psy-*

- chological measurement and evaluation* (6th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Kendall, M. G. (1970). *Rank correlation methods* (4th ed.). London: Griffin.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- SAS Institute (1990). *SAS user's guide: Statistics* (Version 6). Cary, NC: SAS Institute Inc.
- SPSS Inc. (1990). *SPSS-X user's guide* (4th ed.). Chicago: SPSS Inc.
- Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research* (2nd ed.). New York: Macmillan.

Reliability Analysis and Factor Analysis of Ordinal Rating Data

Min-Ning Yu

ABSTRACT

The purpose of this paper is to apply a quantitative index, the coefficient of homogeneity, to analyze ordinal rating data. Values of the coefficient of homogeneity are ranged from 0 to 1. The higher the value, the higher the internal consistency of the test or scale; the lower the value, the lower the internal consistency. Finally, the coefficient of homogeneity is applied to an example in which it is used as an internal consistency index for reliability analysis and as a starting point for factor analysis. This kind of index can be used as a basis of judgement if the test or scale is appropriate.