國立政治大學「教育心理與研究」
民國81,15期，21-38頁

二次式分配準則
在估計再測信度上的應用

余　民　寧

摘　要

傳統計算再測信度的方法有許多缺失，在理論與方法上也較不嚴謹。本文即在介紹二次式分配準則(quadratic assignment paradigm，簡寫爲QAP)的技術，並將此技術應用在再測信度的計算方法上，同時亦舉例討論說明此技術比傳統技術爲優之處。結論認爲：二次式分配準則的技術，不僅可以提供一個比傳統方法更客觀、嚴謹、和精確的再測信度指標，同時也可以驗證該指標的穩定性是否受施測時機、以及隨機性的影響。本文最後建議在未來的應用上，應該偏向使用此技術，以取代傳統的舊技術。

一、導　論

當今社會及行爲科學的研究，多半以某種工具作爲收集研究資料的依據，尤其以教育與心理學方面的研究，更是以測驗或量表來作爲收集資料的工具。這些工具本身是否客觀、可靠，一直是方法論學者們熱切關心的課題。

一般而言，以測驗或量表測量到的數量或分數的一致性或穩定性，一直是研究者（尤其是心理計量學者）所關心的焦點。這種指相同的個人在不同的時間，以相同的測驗測量（或以複本測驗測量），或在不同的情境下測量，所得結果（數量或分數）的一致性，便叫作「信度」(reliability)(Anastasi, 1988; Cohen, Montague, Nathanson, Swerdlik, 1988; Cronbach, 1990; Feldt & Brennan, 1989)。信度是指測驗分數反映出眞實量數(true measurement)的程度或沒有誤差的程度，亦即是測驗分數的精確性或可靠性。信度高表示測驗分數具有一致性、穩定性、及可預測性；反之，信度低表示測驗分數的誤差大、不可靠、及不能信賴。研究者多半採用信度高的測驗或量表，作爲他們收集資料的工具。由於一個測驗所含有的眞實量數部份，是看不見、觀察不到的一種潛在特質或因素(latent trait or latent factor)，因此，一個測驗或量表的信度高低、大小，無法由測量中獲得，必須由數學公式來估計。最常用的信度估計方法有四種：(1)再測方法(test-retest method)，(2)複本方法(equivalent-forms method)，(3)內部一致性方法(internal-consistency method)，(4)評分者方法(scorer method)（郭生玉，民 79；陳英豪、吳裕益，民 80；黃安邦，民 80；葛樹人，民 78；葉重新，民 76；Carmines & Zeller, 1979)。本文只挑選其中的再測方法作爲探討的對象。

以同一個測驗或量表在不同的時間，針對同一群受試者重覆測量兩次，再根據這兩次的測量結果求得的相關係數，便叫作「再測信度係數」(test-retest reliability coefficient)，或簡稱再測信度。這種經過一段時間差距的重覆測量方法，就是再測方法，而再測信度係數是指測驗結果經過一段期間後的一致性或穩定性。因此，它又叫作「穩定係數」(coefficient of stability)。再測信度高表示前後兩次測驗結果的穩定性高，在假設其他影響信度估計的因素不變的情況下，具有高再測信度的測驗或量表表示測量結果可靠、値得信賴，亦即測驗本身是可靠的，適合用於作爲所欲收集研究資料的工具。然而，在大多數情況下，再測信度受許多因素的影響，以致於再測信度不是一種令人滿意的信度估計方法。這些因素有：(1)兩次施測的間隔時間的長短，間隔愈長，所得之再測信度愈低(Cohen, Montague, Nathanson & Swerdlik, 1988, p.104)；(2)受試者本身情況的改變，例如：練習、

記憶、與成熟等因素，都會使第二次的測量結果顯著地不同於第一次的測量結果；(3)測驗情境的改變，如兩次測驗情境中難保有相同的溫度、聲音、光線、及其他突發事件等情境。因此，選擇一個適當的時間間隔，以減少受試者的練習、記憶、與成熟的影響，再控制好測驗情境，便可獲得一個比較滿意的再測信度，此時，再測信度高就眞的表示測驗結果穩定、可靠。再測信度低就眞的表示測驗結果不可靠、該測驗不適宜用作收集資料的工具。

由上述可知，再測信度是以兩次測量結果間的相關係數來表示，因此，在理論上，再測信度的值介於負一到正一（即 $-1 \leq r_{xx'} \leq 1$ ）之間。然而，從古典測驗理論對信度的基本定義來看，信度是指眞實分數（true score）（即眞實量數部份）的變異數佔實得分數（observed score）（即觀察值或實際測量值）的變異數之百分比，亦即 $r_{xx'} = S_t^2 / S_x^2$。因此，在實際上，信度的值永遠介於零與一之間（即 $0 \leq r_{xx'} \leq 1$ ），它是屬於A型相關係數的測量（Ott, Larson & Mendenhall, 1987, p.377）。所以，在實際的應用上，我們通常只採用正值的再測信度；負值的再測信度出現時，則表示該測驗的可靠性堪疑，它可能不適宜作爲測量某種潛在特質的工具。

以兩次測量結果間的相關係數作爲再測信度的指標，從統計學觀點來看，它有幾項缺失：(1)相關係數的大小，容易受某個單獨受試者的極端值影響，亦即某個受試者在前後兩次的測量結果中，若有極端改變的情形發生，則誓必會影響到再測信度的大小。(2)兩次測驗結果的計算方式，是以總和分數（summation score）方式來計算相關係數，這樣計算出來的相關係數，完全忽略試題與試題間之相關係數（或共變數）在兩次測量結果中的變化情形。所以，它提供有意義的訊息便很有限。(3)兩次測量結果間的相關係數，容易受樣本分配及樣本大小的影響。(4)以各次測驗的總和分數來計算相關係數，完全漠視受試者的反應組型（response pattern）在兩次測量中所具有的意義，例如兩次測驗的總和分數雖相同，但其反應組型並不會完全一樣，不同的反應組型具有不同的測量意義，因此計算出來的相關係數的意義，應該是不相同的。(5)相關係數的考驗，亦容易受樣本大小的影響，樣本愈大，相關係數愈容易達顯著水準，表示再測信度顯著地不同於零。

上述這些缺失指出，如果我們把試題與試題間的關聯情況也考慮在再測信度的估算過程中，會比單單使用測驗的總和分數算出來的相關係數，還具有意義與方法學上的嚴謹性。此外，關於相關係數的顯著性考驗，顯著的意義僅代表著該相關係數不是零而已，並沒有辦法顯示出受試者在兩次測驗試題上反應組型的變化情形，因此也就無法判斷出該相關係數的值在整個可能出現的相關係數所構成之分配的相

對位置（或機率），所以對該相關係數的顯著性是否有意義，也就無法知曉。想要瞭解再測信度（亦即前後兩次測量結果間的相關係數）的顯著性的意義，我們必須把試題與試題間的關聯情形，考慮在再測信度的估算與檢定的過程裡，這正是本文所欲回答的課題。

本文的目的，即在提出一種估算再測信度的新概念和新的驗證技術，以解決傳統的相關係數所面臨的缺失，和探索新的估算再測信度方法的可能性。

二、二次式分配準則

為了改進再測信度以相關係數作為指標的缺失，我們可以從資料結構的觀點來探索試題與試題間的關聯在前後兩次測量中的變化情形，並且從中提出一種檢定的方法，驗證兩個相關矩陣（亦即，分別為兩次測量中試題與試題間所構成的相關矩陣）之關係是否不受隨機性(randomness)的影響，亦即兩個矩陣間之關係真正達到顯著水準；如能證明該關係真的不受隨機性的影響，則該關係才是真正客觀的再測信度。

針對上述的說明，本節擬提出的「二次式分配準則」(quadratic assignment paradigm，簡寫為QAP)技術，即是用來比較兩個方形近似量數矩陣(proximity matrix)間關聯性的統計技術。透過這個技術，我們可以提出一個比較合理的再測信度指標，並且考驗它的真正顯著性。

理論基礎

假設我們有n個事項(objects)（這個事項指的可能是個人、刺激、社會、測驗、或變項等），記作$O_1, O_2, \ldots, O_n$；又假設我們有兩個$n \times n$矩陣，這兩個矩陣的資料是從上述n個事項上收集到的近似量數的測量(proximity measures)資料(Borg & Lingoes, 1987;Coombs, 1964; Sneath & Sokal, 1973; Young, 1987)。近似量數的測量內涵可能是每一對事項間的相似性(similarity)、偏好性(preference)、距離(distance)、或網路結構(network structure)等測量結果；並且假設這兩個近似量數矩陣間的關係，可用一項指標來予以表示。

以具體的數學模式來表示，我們有兩個近似量數矩陣，一個叫作**資料矩陣**(data matrix)，其元素值以X_{ij}來表示，另一個叫作**結構矩陣**(structure matrix)，其元素值以Y_{ij}來表示；資料矩陣是實際從研究中收集到的觀察值或實驗值，結構矩陣是根據先前的理論、假設、或經驗所建立起來的資料，也可以是我們想去驗證或推翻

的資料矩陣。這兩個矩陣的元素，都代表每一對事項間的某種近似量數的測量值，其中 i代表列（rows），而 j代表行（ columns），因此，x_{ij}或 Y_{ij}即代表第 i列和第 j行的測量值；當我們以粗寫體字X和Y表示時，即代表它們是由第 i列和第 j行元素所構成之矩陣。

廣義而言，X和Y是定義成一種函數矩陣，亦即每個元素是根據某種測量方式，從 n個事項間收集而得，例如x_{ii}表示對事項 i和 i本身的測量，而x_{ij}表示對事項 i與事項 j之間的測量；換言之，x_{ij}是定義在笛卡爾乘積（Cartesian product）S × S上的一種（非負數）實數值函數，其中S表示是由 $O_1,O_2,...O_n$所構成的集合。為了作業及理論上的方便，x_{ii}和Y_{ii}通常定義為零，這個概念可以解釋成事項本身和本身間的函數（如：距離）為零一般，而x_{ij}和Y_{ij}則為非負數的實數值。

接下來，我們定義一項指標為

$$\Gamma = \Sigma\Sigma\, X_{ij}Y_{ij} \qquad \text{（公式 1）}$$

意即為X和Y矩陣間相對應之元素的乘積和（sum of products）；Γ可以解釋成X和Y矩陣之元素間的一種未常態化的相關係數（unnormalized correlation coefficient），它的值愈大，表示這兩個矩陣間的相對應程度（degrees of correspondence）愈高。Γ指標所測量的，是X和Y矩陣間之高和低元素值的共同組型變化的量，當高元素值在這兩個矩陣中的相同位置作同方向的變化時，Γ指標會變得特別大；因此，Γ指標可以作為衡量X矩陣中元素組型的變化，是否與Y矩陣中元素組型的變化有所相關聯的參考指標。舉例來說，當x_{ij}的元素值大時，其相對應的Y_{ij}也大，或者是x_{ij}的元素值小時，其相對應的Y_{ij}也小，這時所計算出來的Γ值會變大，表示這兩個矩陣間的相對應性很高，或關聯性很高。

到此為止，我們的目的已變得很明顯，那就是找出某種排列方式，使Γ值變得愈大愈好，使這個指標能夠很簡潔有力地表達兩個近似量數矩陣間的關係。尋找這個最佳的排列方式，以使兩個近似量數矩陣間的相關聯指標愈大愈好的問題，便叫作「二次式分配」（quadratic assignment）問題，而它的實作技術便是二次式分配準則。這項二次式分配問題，首由作業研究（operation research）學者 Koopmans & Beckmann(1957)提出，後經 Hubert(1978, 1979), Hubert & Arabie (1989), Hubert & Schultz (1975, 1976), Mantel (1967), Schultz & Hubert (1976)等人的延伸、推導計算公式，目前已是生物計量學（biometrics）和地理學（geography）界廣泛應用的一種技術，值得行為科學與社會科學（甚至是教育研究）界認同、參考的一種新的研究工具。

Γ的排列平均數與變異數

Γ指標雖然可用來計算任何兩個近似量數矩陣間之關聯性程度，到目前爲止，我們都尙未提供評量Γ指標大小的程序作法。關於這種二次式分配問題，我們可以參考非參數統計學中慣用的隨機性模式(randomization model)的作法，如同在列聯表分析中進行Fisher (1934)的精確考驗(exact test)一般，來衡量Γ指標的大小，並考驗該指標的大小是否有達到統計學上顯著水準的意義。以下便是二次式分配準則所擬採行的分析技術。

在標準的非參數隨機性假設(hypothesis of randomization)下，每個近似量數矩陣中的行（和列）出現在某一特定行（或列）的位置是隨機的。因此，n個事項便可隨機構成n！個可能矩陣，每個矩陣都具有同數的元素値，但這些元素値在每個矩陣中的位置可能都不相同。基於這項基本假設，在實際應用的技術上，我們常常固定某個近似量數矩陣（X或Y），而排列另一個矩陣（Y或X）的行（和列），因此我們有n！種排列方式，可以獲得n！個可能的矩陣，同時也計算出每次排列後的Γ值，共得n！個Γ指標。這n！個Γ指標可以構成一種參考的抽樣分配，只要n值變大，這個抽樣分配會趨近於常態分配(Hubert & Golledge, 1981; Hubert, Golledge & Costanzo, 1981; Hubert, Golledge, Costanzo & Gale, 1985; Hubert & Schultz, 1976; Mantel, 1967; Sen, 1976; Shapiro & Hubert, 1979; Siemiatycki, 1978)。因此，原本計算而得的Γ指標便可放入這個參考分配中，再以檢定常態機率分配的作法，去檢定它在此分配中出現之機率是否爲極端的情形。假使Γ值夠大，它的出現機率小於某種臨界點（通常爲設定之顯著水準），此時便可推翻這n！個矩陣（或者說是n！個Γ值）出現的機率都相等的虛無假設，而認爲該（原本的）Γ值的出現絕非偶然（或隨機）造成的，它必定有某種隱含的意義存在，那就是這兩個近似量數矩陣中的元素值組型作相同方向的變化；亦即這兩個近似量數矩陣的相對應性（或說關聯性）很高，達到某種顯著水準。

爲了理論和作業的方便，我們通常假設X和Y矩陣的對角線元素值爲零，在進行排列(permutation)時，我們固定X（或Y）矩陣，而排列Y（或X）的列和行，亦即每次更換第i列和第j列的同時也更換第i行和第j 行，以形成另一種新排列後的近似量數矩陣。在這種排列情況下，對角線元素經排列後Ｊ還是對角線元素，但在不同的位置上，而非對角線元素則有n(n-1)種非對角線元素位置可移動，每次出現所在位置的機率都相等，對任何一個事項i而言，新排列後的近似量數矩陣的第i行和第j列元素，仍然對應到相同的i事項。

由上述可見，當n增大時，排列後所計算出的Γ值會趨近於常態分配的分佈；換句話說，我們可以使用考驗常態機率分配所慣用的Z分數，來作爲考驗某個Γ值出現的機率是否爲不尋常(unusual)的情形。Z考驗的公式可以表示成：

$$Z=[\Gamma-E(\Gamma)]/\sqrt{Var(\Gamma)} \quad \text{（公式2）}$$

其中的E(Γ)表示對Γ值求期望值，亦即是：

$$\begin{aligned} E(\Gamma) &= E(\Sigma\Sigma X_{ij}Y_{ij}) \\ &= \Sigma\Sigma X_{ij}E(Y_{ij}) \\ &= \Sigma\Sigma X_{ij}\Sigma\Sigma Y_{ij}/n(n-1) \end{aligned} \quad \text{（公式3）}$$

Var(Γ)表示對Γ值求變異數，它的算法比較複雜些，我們可以先計算下列代號：

$$\begin{aligned} A_1 &= (\Sigma\Sigma X_{ij})^2 \\ A_2 &= \Sigma(\sum_j^n X_{ij})^2 \\ A_3 &= \Sigma\Sigma(X_{ij})^2 \\ B_1 &= (\Sigma\Sigma Y_{ij})^2 \\ B_2 &= \Sigma(\sum_j^n Y_{ij})^2 \\ B_3 &= \Sigma\Sigma(Y_{ij})^2 \end{aligned}$$

然後再代入公式(Mantel, 1967, pp.215-217)求得：

$$Var(\Gamma)=-\frac{A_1B_1}{n^2(n-1)^2}+\frac{2A_3B_3}{n(n-1)}+\frac{4(A_2-A_3)(B_2-B_3)}{n(n-1)(n-2)}+\frac{(A_1-4A_2+2A_3)(B_1-4B_2+2B_3)}{n(n-1)(n-2)(n-3)} \quad \text{（公式4）}$$

其實，公式3也可以同樣代號表示成：

$$E(\Gamma)=\sqrt{A_1B_1/n(n-1)}$$

有了E(Γ)和Var(Γ)，我們便可以用來計算Z值，Z值可以被看成是常態化的標準分數，它可以被用來考驗原本的Γ值是否達到顯著水準。

我們也可以接著計算出X和Y矩陣的關聯程度。這種關聯程度值可以Pearson的積差相關係數表示如下：

$$r_{xy} = S_{xy} / S_x S_y$$

$$= \frac{\Sigma XY - \dfrac{(\Sigma X)(\Sigma Y)}{n}}{\sqrt{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}\sqrt{\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{n}}} \quad \text{（公式 5）}$$

而在二次式分配問題下，我們可以仿照 Pearson 的公式，把公式 5 擴充爲適用於二次式分配問題中兩個近似量數矩陣間之關聯程度的係數，該係數可以表示如下：

$$r_r = \frac{\Sigma\Sigma X_{ij}Y_{ij} - \dfrac{(\Sigma\Sigma X_{ij})(\Sigma\Sigma Y_{ij})}{n(n-1)}}{\sqrt{\Sigma\Sigma X_{ij}^2 - \dfrac{(\Sigma\Sigma X_{ij})^2}{n(n-1)}}\sqrt{\Sigma\Sigma Y_{ij}^2 - \dfrac{(\Sigma\Sigma Y_{ij})^2}{n(n-1)}}}$$

$$= \frac{\Gamma - E(\Gamma)}{\sqrt{A_3 - \dfrac{A_1}{n(n-1)}}\sqrt{B_3 - \dfrac{B_1}{n(n-1)}}} \quad \text{（公式 6）}$$

這個係數 r_r 可以用來當作再測信度的指標，這個指標會比根據總分計算出的相關係數來當作再測信度的作法，更具有理論上的合理性，因爲它同時考慮到試題與試題間的變化情形，這點是傳統上以相關係數作爲再測信度指標所難以媲美的。

三、應用實例

爲了使讀者明瞭二次式分配準則如何應用到再測信度的檢定上，下列例子僅提供作技術分析的討論，不作任何研究結果的參考或推論。

爲了便利於舉例示範應用的過程，本節用來作爲討論工具的是簡式考試焦慮量表（余民寧，民 76）（參見附錄一）。該量表共有十個題目（每個題目可被視作一個事項，共計十個事項），適用於高中生等有升學壓力的受試者。原本修訂時的再

測信度爲 .83（樣本數爲 211 人）。

假設某位研究者已知考試焦慮的特性，而他想再修訂該量表以適合某種特殊受試對象，同時亦採用再測信度作爲指標。假設他現有八十九名再職進修的中等教師，進行兩次施測，前測在期末考試前兩週，後測在期末考試前一週，兩次測得結果均以總和分數表示其考試焦慮的程度，分數愈高，代表焦慮程度愈強。同樣這批受試者在兩次測量結果間的相關係數，即是本次修訂量表的再測信度，其值爲 .3394，且達 .001 的顯著水準，顯示該修訂量表的再測信度偏低，但其值的大小卻顯著的異於零，表示不是偶然造成的。該量表還是可用，只是測量分數的可靠性較低，它不是一種穩定的測量考試焦慮的工具。

上述的作法是傳統上修訂一份量表常用的檢定過程：計算兩次測量結果間的相關係數，即作爲其再測信度的指標，並且下結論是否採用該修訂的量表。其實，這種考驗方法和結論，都不太正確，理由是該相關係數的計算是以總和分數來進行的，而總和分數無法反映出變項與變項間在兩次測量中的變化情形，也就是說它沒有考慮到同一位受試者在兩次測量中的反應組型所代表的意義。例如，總和分數同爲 30 分，其前測的反應組型假設爲 (2323232323)，後測的反應組型假設爲 (4141414141)，二者之間的變化表示在奇數變項上是隨著考試的迫近而增加，而偶數變項上的反應則隨著考試的迫近而降低。如果說奇數變項所測量到的潛在特質是「情緒化反應」，而偶數變項所測量到的潛在特質是「憂慮」，這是符合考試焦慮特性的現象。但是這兩種反應組型所代表的意義並不相同，也無法單由總和分數中加以看出，這就是使用總和分數的缺失。

其實，傳統上估計再測信度所遭遇的困難，可以經由使用方法較嚴謹的技術來予以克服。把二次式分配準則的技術應用到再測信度的估計上，正是一個可行的例子。

首先，先將本例中的前測和後測的變項間相關係數矩陣（參見表一）求出；然後，再以 1 矩陣（即元素值皆爲 1 的同階矩陣）減去前測矩陣和後測矩陣；再把減後所得的前測矩陣看成是 X 矩陣，後測矩陣看成是 Y 矩陣，分別計算出 A_1 到 A_3，以及 B_1 到 B_3 的代號值；再將該代號值代入公式 3 求出 Γ 指標的排列平均數，代入公式 4 求出 Γ 指標的排列變異數，再代入公式 2 求出 Γ 指標的常態化標準分數 Z，並且代入公式 6 求出 Γ 指標的關聯程度係數，以作爲再測信度的指標。這整個計算過程，可以參見附錄二所列的 SAS/IML 程式。

依照上述步驟，將這 1 矩陣減去前測矩陣（即表一中對角線以右的對稱矩陣）和後測矩陣（即表一中對角線以左的對稱矩陣），便可得到我們所需要的兩個近似

表1　前後兩次考試焦慮量表測量結果的相關係數矩陣

1.0000	.3496	.4502	.3455	.5171	.5520	.5822	.3559	.5254	.4044
.5521	1.0000	.3604	.4092	.4281	.3517	.4020	.5170	.5667	.4490
.5617	.5683	1.0000	.2662	.4386	.3188	.4009	.2949	.3530	.2466
.4178	.6671	.4493	1.0000	.6362	.4503	.5523	.5352	.5711	.4926
.6467	.6529	.6357	.6588	1.0000	.5648	.6438	.5264	.6979	.5424
.5733	.6881	.5031	.6709	.7390	1.0000	.6679	.5433	.6674	.6132
.6331	.5742	.5655	.6076	.6885	.6813	1.0000	.6311	.7289	.5771
.4887	.7288	.5488	.6466	.6821	.7145	.6226	1.0000	.7480	.6868
.5795	.7084	.5837	.6208	.7960	.7520	.6924	.7253	1.0000	.7261
.5626	.6844	.5325	.6787	.7051	.7152	.6524	.7736	.7649	1.0000

（註：爲節省空間起見，前後測結果的相關係數矩陣均列印在同一個表中，其中，對角線以右爲前測結果，對角線以左爲後測結果。）

量數矩陣。此時，這兩個矩陣的對角線元素都是零，且爲對稱性矩陣。接下來，便可代入公式，計算出我們所需要的指標來。附錄二的程式可以提供整個計算過程，研究者只要輸入原始的相關係數矩陣（X和Y）的元素值便可。

在本例中，所得到的各種指標，分別爲：

$$\Gamma = 16.99$$
$$E(\Gamma) = 16.21$$
$$Var(\Gamma) = .06$$
$$r_\Gamma = .77$$
$$Z = 3.18(p<.001)$$

由於Z分數達到 .001的顯著水準，可見前測的相關矩陣和後測的相關矩陣間具有明顯的關聯程度，其值爲 .77。表示出在考慮變項與變項間的變化情形，以及前後兩次測量的反應組型變化的意義後，眞正的再測信度指標比傳統的再測信度指標較大。由此可見，測量的時機縱然會影響到受試者的反應變化（這是考試焦慮的特性之一），但是根據傳統計算再測信度的方法，卻會得到「該量表不是一種良好的測量考試焦慮的工具」等相反的結論。而依據二次式分配準則的技術所求出的再測信度

，卻可以把眞正適當的再測信度值找出來，不受不相關因素（如：測量的時機等）的影響，而得到一個比較正確的估計值和結論。這也就是二次式分配準則的優點。它能彌補傳統技術在理論與方法上的不足，進而取代再測信度的傳統估計法，並將計算出的關聯係數作爲再測信度的新指標。

四、討論與結論

根據公式 1的定義，Γ 指標爲兩個近似量數矩陣間相對應之元素乘積和，亦可看成是兩個近似量數矩陣間的一種未常態化相關係數；該值愈大，表示這兩個近似量數矩陣的相對應程度愈高。根據二次式分配問題的目的，二次式分配準則的技術旨在找出某種或某些矩陣元素的排列方式，使得Γ 值愈大愈好。由於在上述例子中，我們有十個變項，所以我們將有 10！=3628800 種排列方式，可得 3628800 個Γ 值，這些Γ 值自成一個參考分配。我們可將原本的Γ 值（在本例中爲 16.99）放入這個參考分配中。根據中央極限定理及大數定理可知，這個參考分配將成常態分配。因此，我們可用常態化標準分數來考驗這個原本的Γ 值出現的機率是否小於某個設定的顯著水準。在本例中，Z 值即等於公式 2所表示者，其值爲 $(16.99-16.21)/\sqrt{.06}=3.18$，達 .001的顯著水準，它足以推翻每個Γ 值出現之機率均相等的虛無假設，而支持這兩個近似量數矩陣（即前後測的相關係數矩陣）間有某種顯著的關聯存在的說法，其值可由公式 6求出，爲 .77。這種由二次式分配準則的技術所計算出的關聯係數，才可以看成是比較恰當的再測信度，而顯著性考驗也支持它的存在。

這種根據二次式分配準則所計算出的再測信度，它之所以在理論與方法上皆較傳統以總和分數計算者爲嚴謹，乃因爲它在計算的過程中已考慮到變項與變項間的變化情形（即相關係數的大小）和反應組型的不同意義（即矩陣經過 n！次的排列，每次排列的結果均代表不同意義的矩陣，亦即是代表不同意義的反應組型），所以能夠得到比較客觀和嚴謹的再測信度指標值。

若從測量標準誤（standard error of measurement，簡寫成 SE）的觀點來看，其計算公式如下所示：

$$SE = S_x\sqrt{1-r_{xx'}} \qquad \text{（公式 7）}$$

其中 S_x代表測驗的標準差，而 $r_{xx'}$代表信度係數。把本例中的前後測資料代入公式 7，可以求出前後測的測量標準誤分別如下：

$$\text{前測的 } SE = 7.12\sqrt{1-.3394} = 5.79$$

$$後測的 SE = 8.95\sqrt{1-.3394} = 7.27$$

其中 7.12 爲前測的標準差值，8.95 爲後測的標準差值，而 .3394 爲該量表的傳統再測信度係數。我們也可以根據二次式分配準則所提供的統計數，計算出它的測量標準誤如下：

$$二次式分配準則的 SE = .06\sqrt{1-.77} = .03$$

可見後者所得的測量標準誤比前者所得的測量標準誤還小，顯示後者的測量精確性比前者爲大；亦即根據二次式分配準則的技術所得之再測信度指標，比根據傳統方法所得之再測信度指標，還要精確、可靠。在運用再測信度作爲一份測量工具的可靠程度的指標時，我們應該偏向使用較精確的二次式分配準則下所計算出的估計值。

傳統的再測信度的誤差來源，主要來自時間抽樣上的誤差（郭生玉，民 79，頁 67）。在本文所舉的例子中，施測的時機更是影響考試焦慮分數穩定性的一大因素。若研究者此時不明瞭考試焦慮量表的特性，並且使用傳統的再測信度作爲修訂量表的指標，很可能就獲得一個與預期相反的結論，而使修訂工具時所花費的金錢和心力白白浪費掉。此時若改用理論與方法較嚴謹的分析技術，就可以彌補傳統方法的缺失，而驗證先前已修訂過工具的穩定性。

由上述的分析和討論，我們可以獲致下列的結論，那就是：二次式分配準則的技術，不僅可以提供一個比傳統方法更客觀、嚴謹、和精確的再測信度指標，同時也可以驗證該指標的穩定性是否受施測時機和隨機性的影響。在未來的修訂量表過程中，建議應該偏向使用理論與方法皆較嚴謹的技術，來取代缺失的傳統技術。

附錄一

考試焦慮量表

作答說明：請依據你在考試時的感受，以「ｖ」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	□	□	□	□	□
2.我會感覺到後悔	□	□	□	□	□
3.我會有緊張、不舒服的感覺	□	□	□	□	□
4.我會覺得應該好好準備才對	□	□	□	□	□
5.我會有不安、心煩的感覺	□	□	□	□	□
6.我覺得別人會對我的考試成績感到失望	□	□	□	□	□
7.我會感覺到焦慮、手心流汗	□	□	□	□	□
8.我會覺得沒有盡力把試題答好	□	□	□	□	□
9.我會感覺到恐慌，不知如何是好	□	□	□	□	□
10.我會覺得對這次考試沒有信心	□	□	□	□	□

附錄二

二次式分配準則的 SAS/IML 程式

```
PROC IML;
X={?};
Y={?};
E=(X|Y);
X1=E-X;
Y1=E-Y;
N=NROW(X);
GAMMA=SUM(X1#Y1);
A1=SUM(X1)**2;
A2=X1[*,+][##,1];
A3=SSQ(X1);
B1=SUM(Y1)**2;
B2=Y1[*,+][##,1];
B3=SSQ(Y1);
MEAN=SQRT(A1*B1)/(N*(N-1));
VAR=-(A1*B1)/(N**2*(N-1)**2)+(2*A3*B3)/(N*(N-1))+(4*(A2-A3)*(B2-B3))/
    (N*(N-1)*(N-2))+(A1-4*A2+2*A3)*(B1-4*B2+2*B3)/(N*(N-1)*(N-2)*(N-3));
CORR=(GAMMA-MEAN)/(SQRT(A3-A1/(N*(N-1)))*SQRT(B3-B1/(N*(N-1))));
Z=(GAMMA-MEAN)/SQRT(VAR);
PRINT  GAMMA;
PRINT  MEAN;
PRINT  VAR;
PRINT  CORR;
PRINT  Z;
```

?表示由研究者自行輸入之矩陣資料。

參考文獻

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。政大教育研究所碩士論文（未發表）。

郭生玉（民79）。心理與教育測驗（5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76）。心理測驗（7版）。台北：大洋。

葛樹人（民78）。心理測驗學（再版）。台北：桂冠。

Anastasi, A. (1988). Psychological testing(6th ed.). New York: Macmillan.

Borg, I., & Lingoes, J. (1987). Multidimensional similarity structure analysis. New York:Springer-Verlag.

Carmines, E.G., & Zeller, R.A. (1979). Reliability and validity assessment. Beverly Hills, CA:SAGE.

Cohen, R.J., Montague, P., Nathanson, L.S., & Swerdlik, M.E.(1988).Psychological testing:An introduction to tests and measurement. Mountain View, CA:Mayfield.

Coombs, C.H. (1964). A theory of data. New York: John Wiley & Sons.

Cronbach, L.J.(1990). Essentials of psychological testing (5th ed.). New York: Harper & Row.

Feldt, L.S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), Educational measurement (3rd ed.)(pp.105-146). New York:Macmillan.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. Proceedings of the Royal Society, 144A, 285-307.

Hubert, L. J. (1978). Generalized proximity function comparisons. British Journal of Mathematical and Statistical Psychology, 31, 179-192.

Hubert, L. J. (1979). Generalized concordance. Psychometrika, 44,135-142.

Hubert, L. J., & Arabie, P. (1989). Combinatorial data analysis: Confirmatory comparisons between sets of matrices. Applied Stochastic Models and Data Analysis, 5, 273-325.

Hubert, L.J., & Golledge, R.G. (1981). Matrix reorganization and dynammic programming: Applications to paired comparisons and unidimensional seriation.Psychometrika, 46, 429-441.

Hubert, L.J., Golledge, R.G., & Costanzo, C.M.(1981). Generalized procedures for evaluating spatial autocorrelation. Geographical Analysis, 13, 224-233.

Hubert, L.J., Golledge, R.G., Costanzo, C.M, & Gale, N. (1985). Tests of randomness: Unidimensional and multidimensional. Environment and Planning, A17, 373-385.

Hubert, L.J., & Schultz, J.V. (1975). Maximun likelihood paired-comparison ranking and quadratic assignment. Biometrika, 62, 655-659.

Hubert, L.J., & Schultz, J.V. (1976). Quadratic assignment as a general data analysis strategy. British Journal of Mothematioal and Statistical Psychology,29, 190-241.

Koopmans, T.C., & Beckmann, M.(1957). Assignment problems and the location of economic activities. Econometrica, 25, 53-76.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Research, 27, 209-220.

Ott, L., Larson, R.F., & Mendenhall, W. (1987). Statistics: A tool for the social sciences (4th ed.). Boston, MA: Duxbury.

Schultz, J.V., & Hubert, L.J. (1976). A nonparametric test for the correspondence between two proximity matrices. Journal of Educcational Statistics, 1, 59-67.

Sen, A.K. (1976). Large sample-size distribution of statistics used in testing for spatial correlation. Geographical Analysis, 8, 175-184.

Shapiro, C.P., & Hubert, L.J. (1979). Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. Annals of Statistics, 7, 788-794.

Siemiatycki, J. (1978). Mantel's space-time clustering statistics: Computing higher moments and a comparison of various data transforms. Journal of Statistical Computation and Simulation, 7, 13-31.

Sneath, P., & Sokal, R. (1973). Numerical taxonomy. San Francisco, CA:

Freeman.

Young, F.W. (1987). Multidimensional soaling: History, theory, and applications. Hillsdale, NJ: Lawrence Erlbaum Associates.

An Application of Quadratic Assignment Paradigm on Estimating Test-Retest Reliability

Min-Ning Yu

ABSTRACT

The traditional method for computing test-retest reliability suffers from a number of defects. Its underlying methodology is not rigorous either. The purpose of this paper is to introduce the technique of quadratic assignment paradigm(QAP) and to apply this methodology to the computation of test-retest reliability. It also provides an example to show the advantages of the QAP over the traditional method. It is concluded that the QAP can be applied to the computation of test-retest reliability, provide a more objective, rigorous and precise estimate of test-retest reliability than the traditional method does, and be used to test for significance of stability and randomness. For future applications, the QAP technique is recommended as a substitute for the traditional method in computing test-retest reliability.