

第一章 概說

前言

潛在類別模式（latent class modeling, LCM）是探討潛在變數的模型化分析技術。LCM不僅是一種統計方法，更影響了社會科學研究者對於研究的操作與解釋，因此也是一種統計方法學（statistical methodology）。他與傳統因素分析最大的不同在於變數的形式：因素分析處理的是連續變數；潛在類別分析處理的是類別變數。類別變數雖然不像連續變數具有豐富的變異與可以計數的單位，得以進行多樣化的統計分析，但是類別變數卻擁有容易取得、簡單容易操作的特點，例如基本的人口變數、社會變數，都是類別變數。因此，潛在類別模式的分析技術，不但補足了因素分析僅能處理連續潛在變數的缺口，也提供社會科學研究者在面對紛雜的類別資料時，有一更強而有力的分析工具，最重要的是，潛在類別分析把類別資料與潛在變數的觀念加以結合，提高了類別變數的分析價值，也使得社會科學研究者可以一窺潛在類別背後的實證意義，在方法學上具有重要的價值。

第一節 導論

如果說近二十年來，統計原理與技術對社會科學研究最大的衝擊是什麼，最重要的應該是潛在變數模式（latent variable modeling, LCM）的蓬勃發展。主要的理由之一是社會科學研究所關心的研究議題多與抽象而且無法直接觀察的特質有關，為了進行測量與分析，我們往往使用一些間接測量的方法，獲得片面的資料，然後利用潛在變數的概念來整合這些間接測量資料，進而掌握抽象特質的狀態。

以社會學研究當中的重要概念社經地位（socioeconomic status, SES）為例，我們或許可以利用收入、教育程度，或是職業聲望來瞭解人們在社會上的地位或影響力的高低，但是這三者都只涉及社經地位的部分概念，並不能代表社經地位這個概念的全貌，更重要的是社經地位是一個由學者基於理論基礎所提出的抽象概念，客觀上無法直接測量其高低，必須藉由可測量的外顯變數（manifest variables）或指標（indicators）來定義社經地位的狀態，這就有賴潛在變數模式的應用。

潛在變數模式受到重視的另一個重要理由是社會科學所探究的課題是複雜的人類行為經驗與社會現象，研究者所蒐集到的資料往往是片段、交錯複雜的數據，這些原始數據必須進一步加以整理，化簡成為清楚明確的研究變數，以進行後續的統計分析，進而能夠對於現象進行解釋。1970年代以來盛行的主成分分析與因素分析，乃至於當代流行的結構方程模式（structural equation modeling, SEM），最重要的功能就是可以協助研究者將一堆測量數據整併、萃取出少數的幾個主成分（component）或因素（factors）來進行精簡、有效率的分析。潛在變數模式不僅可以協助研究者進行抽象概念的研究，它的另一種功能亦能協助進行資料化簡與整併等這種不需要理論基礎，純粹是一種探索性的資料處理作業。

簡言之，LCM不僅可以從事資料化簡的探索性（exploratory）作業，也可以配合研究者的學理內涵與理論需求，進行驗證性

(confirmatory) 研究，加上電腦普及與功能提升的優勢條件，以模型為基礎 (model-based) 的統計技術與軟體應運而生，促成潛在變數模式近年來的蓬勃發展，未來仍有相當開闊的發展空間。

然而俗語說的好，有得必有失，在高科技掛帥、統計軟體大行其道的今日，如果學術研究者不能透析技術背後的原理脈絡，掌握研究議題下的方法學基礎，一味地只是為分析而分析，那麼充其量只是學「匠」，而不能稱為學「者」。科技始終來自於人性，潛在變數模式雖然富含方法學的哲理與彈性，但是運用者本身的自律與自覺，才是讓學術研究技術得以發揮所長的根本。

類別變數與連續變數在潛在變數模式當中如何正確處理與分析，就是一個很好的例子。基本上，潛在變數模式由潛在變數與外顯變數所構成，當這兩類變數是類別或連續不同形式時，所使用的統計原理即有所不同，再加上研究的議題不同時（例如實驗研究的重複量數資料、發展研究的追蹤資料分析）與資料樣本的結構特性（單層次或多層次），使得潛在變數模式產生了多種不同的型態¹。

本書的目的在介紹以類別變數為主的潛在變數模式分析技術，對象沒有接觸過LCM的社會科學研究人員或是學生，為了適合初學者自行研讀，本書僅對LCM進行原理的介紹與基本模型的範例操作說明（所使用的軟體為MPLUS與LatentGOLD），對進階應用與更複雜或特殊的LCM有興趣的讀者，可以自行參閱專門性著作。然而，雖然本書是緒論性質入門書，但是潛在變數模式本身屬於高等的多變量分析技術，讀者必須具有相當程度的基本統計知識，尤其對於類別變數的分析有相當的瞭解，如果有實際的類別變數的分析與研究經驗，對於本書所介紹的概念與範例會有更好的吸收與瞭解。

¹ 對潛在變數模型的全貌有興趣的讀者，可以參見Skronal與Rabe-Hesketh (2004) 在 *Generalized Latent Variable Modeling* 專書中的詳細介紹。而Eye與Clogg (1994)、Heinen (1996) 的書中則有關於類別變數的潛在變數模型的討論。

第二節 潛在變數模式

一、潛在變數的特性

潛在變數 (latent variables) 是指無法直接測量的變數，必須以統計方法來估計出變數的狀態。一般我們所蒐集的研究資料，都是可以直接測量觀察的變數資料，因此稱為外顯變數 (manifest variables)、觀察變數 (observed variables) 或測量變數 (measured variables)。基本上，潛在變數是一個透過統計程式來加以定義的數據型態，至於潛在變數的內容是什麼，到了社會科學研究者的眼中，就成為不同的內涵。例如心理學家就會把潛在變數看做是抽象的心理特質，透過操作型定義，將潛在變數與心理學現象加以連結，利用潛在變數來反映構念的強度 (Nunnally, 1978)。儘管如此，潛在變項做為一個統計術語，必須符合特定的要件 (Bollen, 2002)，才可做為實際研究的工具²。

最早對潛在變數感到興趣並發展出一套運算與分析程序者，是一批從事遺傳與智力研究的生物與心理學家，例如Gotton、Spearman、Thurston等人。為了瞭解智力的內涵，並能夠測量智力的高低，這些學者發展出因素分析技術，來估計「智力」這一個潛在構念。

在因素分析的架構下，潛在變數是一種具有強弱高低的計量變數 (metrical variable)，例如智力測驗的題目 (觀察變數) 得分會有相關，就是因為背後存在一個共同影響源 (common cause)，亦即「智力」所造成。潛在變數的分數高低代表受測者的個別差異，個別差異的程度可以從潛在變數的變異數來加以估計，而不同的潛在變數

² 有關潛在變數的討論，除了Bollen (2002) 的文章之外，還可以參考Denny Borsboom, Gideon J. Mellenbergh, and Jaap van Heerden (2003) 與Westen, D. & Rosenthal, R. (2003) 等人的討論。

之間所具有的顯著共變，則反映了構念之間具有相關。

潛在變數除了可以以連續變數來定義，亦可以以類別變數來定義，亦即影響觀察變數的分數變動，是因為背後所存在的潛在類別（latent class）。此時，受測者在潛在變數上的個別差異不再是數量上的程度差異，而是質性的類型差異，例如來自「城市」與「鄉村」地區學生在各科學業成績高低的相關差異。由此可知，不僅我們所測量的變數有類別與連續之別，潛在變數也可以被定義成「連續」與「類別」兩種不同形式，因此潛在變數模式可以區分成四種類型，如表1.1所示。

由表1.1可知，四種潛在變數模式中，最為眾人所熟知，也是發展最早的，是外顯變數為連續變數且潛在變數也是連續變數的「因素分析」。其他三類模型均涉及類別變數，其中發展年代較早的是將潛在變數定義成連續變數的潛在特質分析（latent trait analysis），由於多應用在能力測驗的試題分析（題目答對或答錯，為二分觀察變數），因此也被稱為項目反應理論（item response theory）。

表1.1 不同典型的潛在變數模式與開創人物

潛在變數	外顯變數	
	類別	連續
類別	潛在類別分析 Latent Class Analysis	潛在剖面分析 Latent Profile Analysis
	Lazarsfeld & Henry(1968) Goodman(1974) Clogg(1981)	Gibson(1959) Lazarsfeld & Henry(1968) Rost(1985)
連續	潛在特質分析 Latent Trait Analysis (also Item Response Theory)	因素分析 Factor Analysis (include EFA and CFA)
	Richardson(1936) Lawley(1943) Rasch(1960) Haberman(1974)	Spearman(1904) Thurstone(1935, 1947) Jöreskog(1967)

從發展年代可以看出，潛在變數模式的發展脈絡，係延續最早由Spearman基於對「智力」的好奇而發展的因素分析傳統，將潛在變數定義成連續變數，其後才將觀察變數擴展到類別變數，發展出潛在特質分析。但遲至1960年代，才陸續有學者將潛在變數以類別變數來定義，發展出純粹為類別變數模式的潛在類別分析（LCA; Lazarsfeld & Henry, 1968; Goodman, 1974）與觀察變數為連續變數的潛在剖面分析（Latent Profile Analysis）（LPA; Gibson, 1959; Lazarsfeld & Henry, 1968），其中最關鍵的文獻是Lazarsfeld與Henry（1968）所完成的著作《潛在結構分析》（*latent structure analysis*），同時涵蓋了LCA與LPA兩種技術，這就是為什麼後來許多學者將這兩種模型統稱為「潛在結構分析」的起源。

由於不同的潛在變數模式均起源於因素分析，如果讀者能夠瞭解因素分析的原理與特性，將可以很容易地跨入其他三種涉及類別變數的潛在變數模式，因此，以下將略述因素分析的概念，以利讀者銜接這幾種模型的概念。從另一個角度來看，多數研究者對於因素分析並不陌生，甚至於已經從傳統的探索性因素分析延伸到驗證性因素分析（confirmatory factor analysis, CFA）（Anderson & Rubin, 1956; Joreskog, 1967），但是對於如何處理類別變數的潛在變數模式卻甚少接觸，以致於在面對各種形式的變數的潛在模型分析上，欠缺一個完整的瞭解與應用技術的熟稔，本書的介紹或許有助於潛在變數模式的瞭解，但更重要的意義在於透過潛在類別分析的介紹，使各位能夠瞭解各種不同模型之間的異同，適當的使用時機，以提高資料分析的正確操作，而有助於社會科學研究的發展。

二、因素分析模式

1904年Spearman首度提出因素分析的概念以來，至今已有百年歷史。基本上，因素分析的目的除了在尋找觀察變項背後共同潛在

構念，也廣泛被用來簡化變項、探討變項間的群組關係，但是只有在被應用於定義共同因素的因素分析方法，才是潛在變數模式的一種模式。芝加哥大學的Thurstone教授指出，因素分析之所以能夠用來估計潛在變數，最主要的關鍵在於簡化結構（simple structure）原則，他認為外顯變項間的複雜關係，主要是因為背後的潛在構念具有多面向的結構，因此因素分析的主要任務是在釐清變數間的關係，找出最簡化、最清楚、最有解釋力的因素結構，來反映潛在構念的結構，此一最簡化原則也就成為因素分析最根本的原則（Mulaik, 1972）。

（一）因素分析的基本模型

以因素分析去估計構念（constructs）時，研究者的假設為抽象的心理構念是影響外顯行為的決定因素，由於構念隱含在可觀察的事物背後，因此我們所測量到的同一個構念的任兩個觀察變數的相關高低，事實上是由於該潛在構念影響的結果（Borsboom, Mellenbergh, & Heerden, 2004）。以統計的術語來說，某一個構念對於其測量分數的影響力，由因素負荷（factor loading）表示。如果因素變異數為1.0，此時用來測量同一個構念的任兩個觀察變數的相關係數（ ρ_{ij} ）可被兩個標準化因素負荷量（ λ_i 與 λ_j ）的乘積所取代：

$$\rho_{ij} = \lambda_i \lambda_j$$

此時外顯變數之間的關係被其背後存在的共同潛在構念所解釋。透過此一潛在變數的統計界定程式，我們得以解釋外顯變數之間的關係。如果外顯變數之間的相關能夠被潛在變數充分解釋，使得外顯變數之間不再具有關聯，亦即具有局部獨立性，為理想的因素模型。

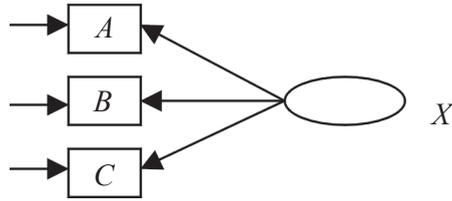


圖1.1：基本的潛在變數模式圖示

圖1.1呈現了帶有三個外顯變數與一個潛在變數的因素分析模型。模型中， X 做為 A 到 C 三個外顯變數的共同影響源，可以完全解釋外顯變數之間的關係，使得三個變數之間沒有關聯，亦即局部獨立性現象的存在。外顯變數無法被共同影響源解釋的獨特變異（uniqueness）彼此相互獨立且服從常態分配，又可稱為測量殘差（measurement residuals），由圖1.1當中 A 到 C 三個外顯變數背後的三個獨立箭頭表示之。由於外顯變數的變異是由背後共同影響源（潛在變數 X ）所反映（reflective）或投射得出，因此這種描述觀察變數與潛在變數的模型又稱為反映性模型（reflective model）³。

（二）探索性與驗證性分析

傳統上，研究者在進行因素分析之前，並未對資料的因素結構有任何預期與立場，而是藉由觀察資料來界定因素的組成與結構，此種因素分析策略帶有濃厚的發現、試探、嘗試的意味，因此稱為探索性因素分析（exploratory factor analysis, EFA）。然而，有時研究者

³ 在潛在變數模型中，定義外顯變數與潛在變數的關係的模型稱為測量模型。除了潛在變數得以影響外顯變數，稱為反映性模型之外，也有學者主張外顯變數得以影響潛在變數，稱為形成性模型（formative models），有興趣的讀者可以參閱2007年《心理學方法》（*Psychological Methods*, 41(2)）當中 Howell, Breivik與Wilcox（2007）合撰的論文，以及兩篇由Bollen、Bagozzi所寫的精彩回應文。

在研究之初即已提出某種特定結構關係的假設，例如某一個概念的測量問卷是由數個不同子量表所組成，此時因素分析可以被用來確認觀察資料的模型是否即為研究者所預期的形式，此種因素分析稱為**驗證性因素分析（CFA）**，具有理論檢驗與確認的功能。

基本上，同一組觀察資料可以進行探索性與驗證性分析，但何時應採用探索性或驗證性分析，主要取決於研究者的目的與文獻的內涵。EFA的主要功能是在簡化龐雜的資料，希望能夠透過變數組合或因素萃取來獲得較少的變數（稱為主成分或因素）來反映某一群具有關聯的資料。在心理與教育測量上，一系列的問卷題目的重組或測驗得分的解釋，也可以透過EFA，找出測驗題目或分數背後的因素結構來達成。EFA的主要特色之一，是因素結構的決定取決於資料本身與統計程式，而非理論或文獻觀點，因此適合理論尚未發展完全的研究課題的資料分析，或是不須採用理論觀點做為指引的調查研究。也因為EFA的目的在簡化資料、尋找因素結構，因此EFA的統計程式將會尋求最少的新變數來反映最大程度的原來的測量變數的內容，對於無法被歸類於某特定因素的測量變數將被忽視或建議刪除，藉以獲得一個最簡化結構。

相對之下，CFA的目的則在檢驗研究者所提出的理論模式是否可以透過研究過程所蒐集得到的資料來獲得證實。當理論模式與觀察資料十分吻合時，我們將可宣稱理論模式與實際資料達到擬合的狀態，相反地，當理論模式與觀察資料不吻合時，研究者所提出的理論模式將被推翻，此時研究者可以藉由**模型修飾（model modification）**或**替代模型（alternative model）**的提出來繼續他的研究。

換言之，在驗證性分析中，理論與文獻決定了統計模型，並先於資料之前而存在，資料的分析是基於研究者所提出的模型架構來進行參數估計，而非從資料的關係來推導出潛在變數的內涵與型態。也因此，在驗證性分析中，因素結構及其組成方式可以由研究者決定，不必受限於資料的本身，因此操作上有較大的彈性，應用範圍也較EFA更為廣泛，但也必須能夠找到合宜的理論或合理的邏輯推理

來支援所提出的因素結構，需要更豐富的文獻與理論論述做為基礎。因此一般將CFA稱為理論推導（theory driven）研究，將EFA稱為資料推導（data driven）研究。從研究的角度來看，EFA與CFA兩者目的不同，使用的時機也不一樣，兩者具有相輔相成的功效（Coste, Bouee, Ecosse, & Pouchot, 2005）。

三、潛在類別分析

如果說因素分析是用來探討連續外顯變數（continuous manifest variables）背後的連續潛在變數（continuous latent variables）的最佳技術，那麼潛在類別分析是用來探討類別外顯變數（categorical manifest variables）背後的類別潛在變數（categorical latent variables）的最佳技術。因素分析與潛在類別分析主要的差異，就是變數的型態是連續或類別的差異。而LCM即是用以探討潛在變數的一種模型化的分析技術。以先前討論的圖1.1為例，如果把圖中的各外顯變數與潛在變數視為類別變數時，即成為一個潛在類別模型（latent class model）。

基本上，因素分析是以潛在變數來解釋外顯變數之間的線性關係（linear relationship），達到局部獨立性；LCM的目的即在於以最少的潛在類別數目來解釋外顯變數之間的關聯，以達到局部獨立性。除此之外，雖然LCM與因素分析背後的數理基礎迥然不同，但是在方法學上，LCM與因素分析在功能與目的上幾乎沒有什麼不同。例如，當代因素分析被區分為探索性與驗證性兩種，LCM也有類似的分類，探索性的LCM主要由資料來決定潛在變數的內容（第三章），驗證性LCM則是由理論來引導潛在變數的內容（第四章）。

傳統以來，潛在變數的估計除了使用因素分析來進行潛在因素的萃取之外，項目反應理論則以非線性模型來進行潛在特質的測定。LCM與因素分析最大的不同，在於LCM所處理的是類別潛在變數，而因素分析所處理的則是連續潛在變數。另一方面，IRT模型雖然可