



# 教育資料的多層次模式分析

溫福星\*

## 壹、社會科學研究的兩個統計方法學典範

結構方程模式（Structural Equation Model, SEM）與階層線性模式（Hierarchical Linear Model, HLM）或是多層次模式（Multilevel Model, MLM）可以說是當代社會科學研究兩個重要的方法學典範，結構方程模式主要在探討觀察變項與潛在變項間的關係—測量模式測量誤的問題，以及潛在變項間的影響關係—結構模式，其中測量模式主要是社會科學研究以問卷題項為資料蒐集工具，用來檢驗研究題項構念效信度最主要也最重要的方法。而階層線性模式／多層次模式則是來自於兩階段（或多階段以上）集群隨機抽樣或分層隨機抽樣下，先隨機抽取總體層次的學校／班級或公司／團隊，再隨機抽取個體層次的單位如學校／班級內的學生或是公司／團隊內的員工，然後針對這些學生或是員工進行研究變項的蒐集，因學生與學校／班級或員工與公司／團隊的巢套（nested）關係，所蒐集的變項資料因此形成階層結構，導致許多傳統統計方法針對資料獨立性的假設被違反，因而所發展處理相關（correlated）或集群（clustered）資料的統計方法（Raudenbush & Bryk, 2002）。在過去20年結構方程模式與階層線性模式各自獨立發展的時空背景下，這兩個研究典範陸續發展出處理不同研究問題的應用方法，例如結構方程模式中的中介效果與調節效果的估計與檢定，

---

\* 溫福星 Fur-Hsing Wen，東吳大學國際經營與貿易學系副教授  
電子郵件：wenft@scu.edu.tw

本文評介：O'Connell, A. A., & McCoach, D. B. (Eds.). (2008). *Multilevel modeling for educational data*. New York: Routledge.

階層線性模式的多層次中介效果與跨層次交互作用的估計與檢定，圖 1 即說明了當代這兩個研究典範的發展主軸與應用模型。在圖 1 的三圍新論，X 軸是處理測量誤的結構方程模式主軸、Y 軸是處理階層相關的階層線性模式、Z 軸是處理時間的向度，圖中的各種統計模型即在各自發展下的應用方法。三圍的意義除了是三維以外，亦是指我們現今所用的統計方法是被這三維所圍繞著。

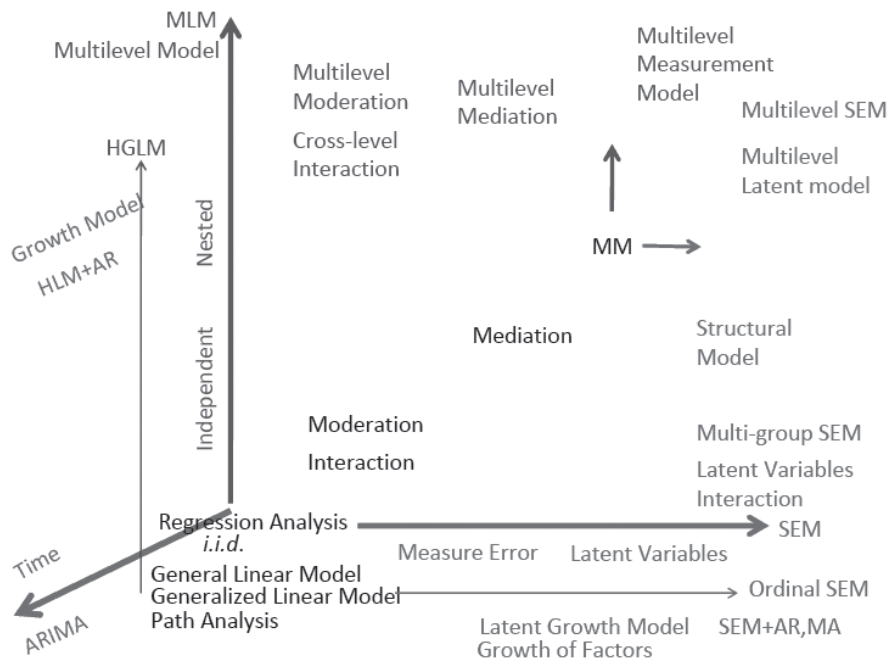


圖 1 三圍新論

三圍新論中的三個座標軸，分別是 X 軸的測量誤 (measurement error) 或是潛在變項 (latent variable)、Y 軸的非獨立性 (nonindependent) 或是嵌套 (nested) 關係、與 Z 軸的時間 (time) 三個向度 (dimensions)。首先說明 X 與 Y 所構成的平面，這是個橫斷

面分析的介紹，位於原點位置所在的分析方法是迴歸分析，它的基本假設之一是「*i.i.d.*」就是獨立且同質的分配（independently and identically distributed）。迴歸分析與變異數分析都是屬於一般線性模式（general linear model）的分析方法，它的依變項是屬於連續量尺，而屬於非連續量尺的依變項，例如二分變項、多類別變項或是次序變項的分析方法就是廣義線性模式（generalized linear model）。其中利用迴歸分析於研究架構中，在結構方程模式普及之前，就是屬於多個迴歸分析所串連組合而成的徑路分析（path analysis）最為實用。

在 X 與 Y 平面的對角線是迴歸分析在方法學上的應用，例如是具有交乘項或是交互作用項的調節效果分析（moderation effect）與中介效果分析（mediation effect），這是社會科學研究兩個非常重要的研究方法（Baron & Kenny, 1986; Jose, 2013）。除此之外，調節變項與中介效果的結合，或是中介變項與調節效果的結合所形成的調節式中介效果（Moderated Mediation Effect, MoMe）與中介式調節效果（Mediated Moderation Effect, MeMo）的檢驗，都是可以利用迴歸分析來完成（Muller, Judd, & Yzerbyt, 2005）。但是迴歸分析的使用在方法學上有兩個重要的條件，第一是沒有測量誤問題、第二是資料必須符合獨立性的假設。

由於這兩個條件的限制，使得在社會研究有關構念的問題，涉及到多重指標的測量，因此驗證性因素分析（Confirmatory Factor Analysis, CFA）與 SEM 因應而生，在圖 1 的 X 軸為測量誤的向度，Y 軸是在多層次下具有階層資料結構，使得迴歸分析的測量誤與獨立性假設被凸顯出來。SEM 向度在處理調節效果的問題，可以採用多群體結構方程模式（multi-group SEM）或是潛在變項交互作用結構方程（latent variables interaction SEM）來解決（Marsh, Wen, & Hau, 2006），在中介效果的檢測，更是 SEM 的強項。而在另外一軸有關獨立性問題，所

發展的多層次迴歸分析，基本上是迴歸分析的延伸，凡是迴歸分析的缺點，多層次迴歸分析也都存在，因此在這一軸的研究方法大部分無法處理測量誤的問題。在迴歸分析所應用的調節效果與中介效果的檢驗，多層次迴歸分析都是沿用迴歸分析的步驟或檢驗程序。唯一比較特別的是在圖 1 的右上角，兩個取向在研究多層次測量模式是有點不同，在 MLM 稱為多層次測量模式 (multilevel measurement model; Kamata, Bauer, & Miyazaki, 2008)，在 SEM 稱為多層次潛在模式 (multilevel latent model; Muthen, 1994)，主要的就是多層次驗證性因素分析 (multilevel CFA)，兩者的差異在測量理論不同外，則是在較高層次或是組織層次的測量誤假設。此外，SEM 在多層次潛在模式尚可以進行兩個層次的結構方程模式，稱為多層次結構方程模式 (multilevel SEM)。

在 MLM 與 SEM 處理連續依變項以外，兩者都可以處理類別依變項，在 SEM 稱為類別結構方程模式 (categorical SEM 或 ordinal SEM)，而 MLM 則是將廣義線性模式擴充到多層次廣義線性模式或是階層廣義線性模式 (Hierarchical Generalized Linear Model, HGLM; Raudenbush & Bryk, 2002)。

將圖 1 的 X 與 Y 平面延伸到第三軸時間 Z，在傳統的時間數列資料有時間數列迴歸，或是時間數列分析的自我迴歸整合移動平均模式 (ARIMA)。在經濟與財務金融的時間數列分析，所蒐集的資料點或波次是相當多，不像組織管理與教育研究，要蒐集四波的重複觀測資料是相當艱難的工程，因此 MLM 與 SEM 在處理重複觀測量仍然局限在不長的時間點上。在 MLM 上有成長模式 (growth model)，可以用一般的 HLM 來處理外，亦可以利用階層多變項線性模式 (Hierarchical Multivariate Linear Model, HMLM; Thum, 1997) 來處理帶有相關殘差的自我迴歸模式。而在 SEM 方面，利用 CFA 的方法來處理重複觀

測資料，稱為潛在成長模式（latent growth model; McArdle & Epstein, 1987），或是去估計殘差項自我迴歸或是移動平均的 SEM（Sivo & Willson, 2000）。X 與 Y 平面的方法模型亦可以延伸應用到重複觀測資料上。

## 貳、本書評論

全書共有 521 頁分為四個部分十三個章節，由 37 位學者撰寫編輯而成，關於本書的優缺點陳述如下：

- 一、除了第一章簡介與最後三章結果呈現、軟體、數理模式外，各章節內容淺顯易懂，每章都先說明原理方法，再搭配實徵資料分析與結果說明，文章前後呼應，條理分明，適合初學者自學。
- 二、全書幾乎已包含多層次模式重要的概念與內容，例如為何巢套資料在傳統迴歸分析上的問題，在第一章就開宗明義的介紹組內相關係數（ICC）。各章節都介紹多層次模式分析上的重要觀念與分析模式，例如固定效果與隨機效果、完全訊息最大概似估計法與受限最大概似估計法、ICC（1）與 ICC（2）等，而示範的分析結果，本書對以上重要統計量的意義也做詳實的解釋，可以視為研究者或初學者學習的工具書或手冊，特別是第十一章研究結果的呈現，可以做為多層次研究寫作的檢核。
- 三、部分章節最後作者都整理相關方法比較的大表，可幫助自學者減少摸索與整理時間。此外，全書重要的相關議題與概念各章節都有所連結，例如第二章的樣本數與第八章的檢力相互呼應、第六章的二元資料分析與第十章的項目反應分析互相配合。最後，重要的參考或引用文獻在內文都有所交代，可視為進階閱讀與學習的參考依據，省去蒐尋與摸索的時間。雖然本書是 2008 年出版，距離現在

也有一段時間，可能有更新的文獻與方法問世，但基本與重要的觀念仍不變。

四、美中不足地方限於篇幅，無法提供各章節所介紹方法的軟體資料匯入與模式設定畫面（第十章有提供相關軟體語法），雖然本書有提供連結網站的示範資料庫，但對初學者自學則有部分進入障礙。另外，全書的分析對個體層次解釋變項大都進行總平減，然對組平減方法則著墨不多。

## 參、本書章節簡介

第一章有關於階層資料的特徵，來自於集群抽樣或是實驗操縱的結果，使得來自相同的群體內的兩個受試者因獲得相同的共同事件或背景較不同群體的受試者間來得較相似。ICC 即是用來衡量這樣資料相似或是同質的測度，反應受試者背後環境間的差異，也代表組內受試者間資料的相依性。而這樣的資料結構也違反傳統統計方法所假設的獨立性，這也稱為設計效果，忽視這樣的資料階層結構的特性，使得參數估計值的標準誤被低估，導致增加拒絕虛無假設的型一錯誤。多層次模式即考慮到這樣的資料特性，用來確保統計分析結果的統計結論效度。

第二章探討大型資料庫的設計與分析，大型資料庫最常要克服的問題是：不是一個簡單的抽樣架構與要求要有足夠的不同特徵受試者樣本。因此大型資料庫的設計是一個分層多階段的集群抽樣，這樣複雜抽樣設計的可能問題包含多階段與過度抽樣必須在分析時被考慮，以確保參數估計值標準誤的正確估計。接著介紹了集群與多階段抽樣、分層抽樣、非比例抽樣與加權的方法與優缺點，以及簡介三個美國國家教育統計中心（NCES）的公開教育資料庫（一個橫斷面 SASS、兩個追蹤調查 ECLS 與 NELS）抽樣設計、研究內容、資料結構與以多層次分析發

表文章的相關訊息。文中在抽樣加權係數的設計說明較多，利用簡單的數字來說明讓人看了淺顯易懂；另外在兩層結構各層樣本數的決定中，作者整理出幾篇重要的模擬研究文獻，說明在固定與隨機效果參數不偏估計下，第二層的組數大小要比組內的樣本數來得重要。

第三章示範多層次分析應用在學校效果 (school effect) 的橫斷面研究，以 PISA 資料庫為例，透過 HLM 軟體進行分析，利用五個多層次模型依序在控制了學生與家庭背景資料後，檢視學校硬體的環境 (context) 與軟體的氣候 (climate)，以及其交互作用對學生數學成就的影響。這五個模型分別為零模型 (沒有任何解釋變項)、學生模型 (個體層次學生與家庭變項)、脈絡模型 (同時含有學生層次與學校層次變項，學校層次變項為由學生變項聚合平均的脈絡變項)、評估模型 (存在學生層次與學校層次脈絡變項與環境變項外，另外引入學校氣候變項) 與互動模型 (含有學校層級環境與氣候變項的交互作用項)。該示範介紹了標準化分數 (連續變項)、總平減 (虛擬與連續變項) 方法、與其迴歸係數的意義、信度 (ICC2) 的解釋、與虛假決定係數的計算與解釋 (隨機截距模型)，以及部份以截距與斜率為結果變項模型的考慮。最後，對平減的不同方法與意義做出說明 (其中個人認為部份總平減的解釋有些錯誤 [p.95])。

第四章是利用多層次與結構方程模式進行追縱資料成長模式的估計，傳統分析重複觀測的變異數分析必須不能有遺失值且要所有觀測必須在相同時間上量測，因此這樣限制使得多層次模式脫穎而出，它允許受試者在不同時間量測、可以不要有相同測量間距、可以有不同資料筆數，可以進行不同成長軌跡估計、探討與共變數之間的關係、以及估計受試者內與受試者間的變異。該章以多層次模式示範了線性、二次曲線與分段的多層次成長模式，以結構方程模式示範線性與非線性潛在成長曲線模型，並對非線性模型的成長參數有很清楚的解釋，以及示範成長

混合模型，最後整理比較了多層次模式、潛在成長曲線與成長混合模型的假設和優缺點。

第五章非單純巢套設計的交叉分類隨機模型，也就是個體層次的受試者同時巢套於兩個上層組織內，但這兩個上層組織是獨立的情況，例如同一個學校的學生未必來自相同的社區，同一個社區的學生未必到相同學校就讀，學校與社區就是兩個上層組織單位；或是學生目前的學業成績同時受到國中與高中兩個求學階段的影響。本章有較多的方程式來解釋這個模型，並利用一些圖表來加強說明。最後作者以 HLM 示範兩層、以 SAS 介紹兩層與三層的交叉分類隨機模型，並摘錄部分 HLM 軟體的畫面與報表結果，以及 SAS 語法。（(5.13) 公式有誤、圖 5.13 與分析的模式不一致）。

第六章是類別資料的多層次模式，主要介紹二元資料與次序資料的多層次二元邏吉斯與多層次次序邏吉斯迴歸分析。一開始作者有非常清楚的單一層次次序邏吉斯迴歸的原理與模式解析，然後延伸到多層次模式的二元邏吉斯與次序邏吉斯，文中並搭配許多方程式說明，並以 HGLM 示範多層次二元邏吉斯與多層次次序邏吉斯迴歸分析，最後並對現有的軟體作一介紹。該章特別地方是類別依變項的 ICC 計算說明，以及對多層次次序邏吉斯的比例勝敗比（proportional odds）假設的檢視。

第七章是多層次模式適配度或適切性與解釋力的評估，在模式的選擇上可以根據精簡程度、競爭模型比較與參數的意義來考慮，可以進行離異數差異的卡方檢定與 AIC、BIC 的比較；在解釋力方面可以考慮變異數的削減程度虛假決定係數（pseudo  $R^2$ ）與預測誤差的變異數減少百分比來衡量。該章對 AIC、BIC 與卡方差異有不少篇幅介紹之間的關係，以及利用虛假決定係數可能造成問題的原因進行討論，因此建議使用預測誤差的變異數減少百分比。



第八章是檢力、樣本數與設計間的關係，由於教育研究大都探討介入對學生的影響，而這個介入通常是發生在班級或學校上，因此多層次模式要研究的是學校層級實驗組與控制組間的差異效果。本章以實驗設計變異數分析的角度探討四種不同的多層次研究設計，因此本章有許多的公式用來推導與檢力相關的因子，包含了各層級的樣本數與 ICC 等，另外配合 Optimal Design 軟體的圖形進行說明。本章中出現非中化（non-centrality）參數，這屬於數理統計是教育學者較不熟悉的部份，但可從型 I 與型 II 錯誤的關係圖來理解。

第九章是多層次模式應用到統合分析上，作者介紹了單變量與多變量的統合分析資料，利用 HLM 軟體進行兩群平均數效果量的差異檢定與研究效果量變異數的異質性檢定。由於統合分析的資料是已發表的樣本描述統計量平均數、變異數與樣本數，透過公式可以計算出標準化的兩組平均數差異，再經轉換成不偏的效果量估計值與效果量抽樣變異數，最後匯入 HLM 軟體進行統合分析，但此方法與以一般原始資料輸入的多層次模式有點不同。

第十章是應用多層次測量模式到傳統的古典測量理論的 Essentially Tau-Equivalent 模型與項目反應理論的 Rasch 模型上，並延伸到巢套的集群資料。全章分為三個部份，第一部份是應用三層 HLM 與兩層 HMLM 模組用來估計學生層次與學校層次測量題項的內部一致性信度、第二部分利用三層 HGLM 模組分析一組二元題項的 Rasch 模型；前述兩個應用都可以加入學生或學校層次的解釋變項，例如在 HGLM 加入二分解釋變項就可形成對 Rasch 模型 DIF 的檢測。有關於 HLM 與 HGLM 應用到多層次測量的優缺點則有相當多篇幅討論，針對其缺點，作者介紹了第三部份，利用多層次結構方程的多層次驗證性因素分析來解決。

第十一章是多層次分析的結果報導，作者建議研究者要呈現什麼與

如何利用文字、表或圖來呈現這些結果。作者認為在所撰寫的文章中，盡可能在（一）研究問題與（二）文獻探討去強調層多層次研究的目的與適切性，在（三）方法中呈現各層次樣本、研究的抽樣設計、變項蒐集、定義與衡量、所配適的模式（包含設定、變項平減、模式發展過程、模式評估）、估計與推論中報導相關的多層次結果、（四）結果的呈現與（五）討論如何與讀者溝通，最後以 5 個大項 37 個小項的檢核表來確認讀者是否可以從文章回答這些問題，做為一篇好的多層次研究文章的判斷標準。

第十二章是可以進行多層次模式分析的軟體簡介，包含 MLwiN、HLM、SAS、SPLUS、R、SPSS、Mplus、STATA，最後作者用一張大表來比較這些軟體之間的差異。

第十三章作者介紹多層次模式的理論基礎，用來說明參數估計的過程，本章類似 Raudenbush 與 Bryk（2002）階層線性模式專書第十四章估計理論一樣，完全是屬於數理統計的部分。

## 參考文獻

- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical consideration. *Journal of Personality and Social Psychology*, *51*, 1173-1182.
- Jose, P. E. (2013). *Doing statistical mediation and moderation*. New York: The Guilford Press.
- Kamata, A., Bauer, D. J., & Miyazaki, Y. (2008). Multilevel measurement modeling. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 345-388). Charlotte, NC: Information Age Publishing.
- Marsh, H. W., Wen, Z., & Hau, K.-T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller

- (Eds.), *Structural equation modeling: A second course* (pp. 225-265). Greenwich, CT: IAP.
- McArdle, J. J., & Epstein, D. (1987). Latent growth curve within developmental structural equation models. *Child Development*, 58, 110-133.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89, 852-863.
- Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376-398.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Sivo, S. A., & Willson, V. L. (2000). Modeling causal error structures in longitudinal panel data: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, 7, 174-205.
- Thum, Y. M. (1997). Hierarchical linear models for multivariate outcomes. *Journal of Educational and Behavioral Statistics*, 22(1), 77-108.

