

第一篇

原 理

P
A
R
T

1

- 階層資料結構：一個普遍現象
- 階層資料分析長久以來的困境
- 階層模型統計理論的發展簡史
- 階層線性模式的早期應用
- 本書第 1 版問世以來的新發展
- 本書架構



階層資料結構：一個普遍現象

很多社會研究都涉及階層資料結構。在組織研究中，研究者多會探討工作場中的特徵，例如決策的集中化程度（**centralization**）如何影響員工的生產力（**productivity**）。其中，員工與公司都是分析的單位，變數是在兩個層次同時進行測量。這樣的資料便具有階層結構，即員工是內屬於（**nested**）公司。

在跨國研究中，人口學家探討國家經濟發展的差異是如何與成年人教育程度產生交互作用來影響該國的生育率（參見 Mason, Wong, and Entwistle, 1983）。這種研究包括了國家層次的經濟指標，以及以家戶為單位的教育與生育資訊。家戶與國家都是研究單位，其中家戶內屬於國家，因此這種資料結構也是階層性。

同樣的資料形式也存在於發展研究，其資料的蒐集往往是對同一組個體在不同時點做多次的觀察，這種重複測量包含了每一個體成長的軌跡資訊。心理學家特別感興趣的是一個人的個性及其所處的不同環境如何影響他的成長，例如 Huttenlocher、Haight、Bryk 和 Seltzer（1991）檢視了不同的家庭語言環境對兒童在成長過程中運用字彙數量的影響。當樣本中每一個體被測量的次數相同，通常

會被視為一種個體在不同狀況下的比較分析。但是如果不同個體被觀察的時點數量與時間間隔有所不同時，我們可以把不同的測量時點或狀況視為嵌套（nested）於同一個體中的階層資料。

把許多不同的研究發現利用量化技術來加以整合的研究，則是另一種階層資料分析問題。研究者可能想探討不同研究中所使用的實驗介入策略、研究方法、受試者樣本特徵、實驗效果的脈絡等差異。在這種情況下，研究的樣本是嵌套於不同研究之中。階層模型也為這類研究活動提供了一個十分有效的統計架構（Berkey et al., 1995; Morris & Normand, 1992; Raudenbush and Bryk, 1985）。

教育研究尤其容易遇到這類的挑戰，因為學生學習成長的研究經常會涉及多次重複觀察內屬於個體內的特殊階層結構，這些學生又內屬於某一組織結構。例如，有關教學成效的研究中，研究者所關心的是學生與老師在某一種教學方法下的互動。這種互動通常發生於班級環境中，並且以某一學期為限。那麼研究問題便存在三個焦點：學生在一學年（或其中的某一段）課程學習中的個別成長、教學對學生個人性格和學習收穫上的效果、以及這些關係如何受到班級組織和教師特定行為及個人特徵的影響。由此看來，這種研究資料具有三層結構：第一層的單位是各時點的重複觀察，它內屬於第二層單位中的每一學生，而每一學生又內屬於第三層單位中的班組或學校。

橫斷研究也經常包含三個層次的資料結構。例如，Nye、Hedges 和 Konstantopoulos（2000）重新分析了田納西班級規模實驗的資料，這個實驗涉及不同學校所包含的班級中的學生。Raudenbush、Cheong 和 Fotiu（1999）分析了美國全國教育發展評鑑的資料，其中涉及各州學校的學生。



階層資料分析長久以來的困境

雖然階層資料結構在行為科學和社會研究中普遍存在，但是以往的研究常常不能將其充分地體現於資料分析之中。在相當程度上，這種忽略反映的是傳統統

計技術在分析嵌套結構上的局限性。在社會研究中，這種局限性產生了一連串的問題，如彙總偏差（aggregation bias）、估計精確度誤差以及「分析單位」問題（參見第 5 章）。這些問題還導致了概念上的混亂，阻礙了分析模型中多層次結構的明確化，因而不能檢驗在每個不同層次上的效果，也不能檢驗各種層次之間的關係。相同的，人類發展研究也受到「變化測量」（measurement of change）問題的阻礙（參見第 6 章），此一問題難以處理，甚至導致一些分析得出反對直接建立成長模型的結論。

儘管分析單位問題和變化測量問題在文獻中一直都各自被拿來討論，但其實它們有著共同的原因，即傳統統計技術無力分析多層次模型。在過去，對於某些具體研究問題總是能夠找到處理方法，至少可以部分地解決問題。隨著近年來階層線性模式統計理論的發展，一套完整的方法已經建立起來，可以廣泛應用，得到有效的估計。

從我們的角度而言，更重要的是過去發展多層次模型的障礙現在已經被破除。我們現在能夠建立關於各層次內關係與各層次間關係的假設，並且還可以估計出各個層次上的變化量。從實務的角度來看，階層線性模式更加接近行為和社會科學研究中的基本現象。它們不僅可以應用於新問題的探索，而且還可以提供過去所不能得到的實證發現。



階層模型統計理論的發展簡史

本書所討論的模型在不同領域的文獻中有不同的稱呼。在社會學研究中，它經常被稱為多層次線性模型（multilevel linear models，參見 Goldstein, 1995; Mason et al., 1983）；在生物統計研究中更經常用的名字是混合效果模型（mixed-effects models）和隨機效果模型（random-effects models，參見 Elston and Grizzle, 1962; Laird and Ware, 1982; Singer, 1998）；計量經濟學文獻稱之為隨機係數迴歸模型（random-coefficient regression models，參見 Rosenberg, 1973; Longford,

1993)；統計學文獻則稱之為共變數成分模型 (covariance components models, 參見 Dempster, Rubin, and Tsutakawa, 1981; Longford, 1987)。

我們採用階層線性模式 (hierarchical linear models) 的稱謂，因為它指出了即使在不同應用中，例如成長研究、組織效能、統合研究，他們的資料都存在相同的重要結構特徵。此一稱呼最早由 Lindley 和 Smith (1972) 提出，是他們對線性模型的貝氏估計方面所做出貢獻的其中一部分，他們為複雜誤差結構的嵌套數據發展出一個通用的研究典範。

遺憾的是，Lindley 和 Smith (1972) 的研究有一段時間停滯不前，因為其模型應用需要對非平衡資料進行共變數成分估計。在 1970 年代早期，他們的模型只能解決一些極簡單的問題，尚不可能提供一個通用的估計方法。後來，Dempster、Laird 和 Rubin (1977) 在 EM 演算法上的進展突破了此一技術障礙，形成了概念上切實可行又可以廣泛應用的共變數成分估計方法。Dempster 等人 (1981) 說明了這種方法對多層次資料結構的適用性。Laird 和 Ware (1982) 以及 Strenio、Weisberg 和 Bryk (1983) 都是應用此一技術方法解決了橫斷資料中多層次結構的計算問題。

後來，又產生了其他透過疊代與加權的一般化最小平方法的共變數成分估計方法 (Goldstein, 1986) 和 Fisher 得分 (score) 演算法 (Longford, 1987)。現在已經有好多統計軟體可用來估計這些模型，包括 HLM (Raudenbush et al., 2000)、MIXOR (Hedeker & Gibbons, 1996)、MLWIN (Rasbash et al., 2000)、SAS 副程式 Proc Mixed (Little et al., 1996)、VARCL (Longford, 1988)。完全貝氏方法也已經由 Gelfand 等人 (1990) 和 Seltzer (1993) 研製出來，配合的軟體如 BUGS (Spiegelhalter et al., 1994) 現在也得到了廣泛應用。



階層線性模式的早期應用

如上所述，行為科學和社會科學資料常常具有嵌套結構。例如，包括了內

屬於個人的重複觀測，這些個人可能又內屬於某種組織單位，例如學校。而組織單位本身又可能內屬於社區，再內屬於州，再內屬於國家。利用階層線性模式，此一結構中的每一層次都用子模型來代表。這些子模型反映了本層變數之間的關係，並且定義了某一層次的變數對另一層次變數之間的關係所產生的影響。儘管不論有多少個層次都可以這樣來代表，但所有基本統計特徵可以透過一個兩層的模型反映出來。

本書所討論的應用問題包括三個方面的研究用途：一是如何對個體單位取得較好的效果估計（例如透過借助其他學校也有類似估計的事實來改進某一個學校的迴歸模型）；二是對各層次之間的效果建立模型並進行假設考驗（例如學校規模如何影響學生的社會特徵與學校的教學成績之間的關係）；三是分解各層次間的變異數和共變數成分（例如將學生層次的變數之間的共變分解為校內與校間兩種成分）。以下將會提到每一種模型已發表的例子。

個體效果的估計改善

Braun、Jones、Rubin 和 Thayer（1983）所關注的問題是管理學院研究生入學考試使用標準化考試分數來選取少數民族申請者。許多學校的錄取決策部分取決於一道以考試成績來預測以後學業成績的公式。然而，由於大多數學校所錄取的大多數學生都是白人，所以白人學生的資料主宰了這道預測公式的估計。結果，這些公式可能在選擇少數民族學生時並不能提供合理的排序。

原則上，較公正的作法是每個學校應該建立一個選擇少數民族學生的公式，但是估計此一公式有其困難，因為大多數學校只有幾個少數民族學生，而資料太少不能建立可靠的預測公式。Braun 等人（1983）使用的是 59 個工商管理學校的研究生資料，其中有 14 個學校根本沒有少數民族學生，有 20 個學校的少數民族學生的人數在 1~3 人之間。這些學校便不可能應用標準迴歸方法來建立少數民族學生的預測公式。此外，即使是對於有較多少數民族學生的其他 25 個學校而言，雖然可以建立估計模型，但少數民族學生的樣本量還是太少，因此其估計係數精確度會很差。

如果換用另一種方法，可以不管學生屬於哪個學校，將所有學校的樣本彙集在一起。但這種方法也會產生問題，當某些學校的少數民族學生較多而另一些學

校的少數民族學生很少時，由於公式不能反映這種選擇性現象，因而導致預測係數的估計偏差。

Braun 等人（1983）採用了階層線性模式來解決此一兩難問題。透過彙集所有資料，他們得以有效利用所有資訊為每個學校提供一個能區分白人學生和少數民族學生的模型。每個學校的預測值實際上是此一學校的各項資訊與所有樣本之中的變數關係的加權組合，正如人們所期望的，這些對每一成分的相對權數取決於它們的精確度。此一估計程式將在第 3 章加以描述，在第 4 章給予示範。有關的應用還可以參見 Rubin（1980）的論文，這些統計發展的回顧可以參見 Morris（1983）的論文。

跨層次效果的模式化

階層模型的第二種用途關係到用模型來反映在某一層次測量的變數如何影響其他層次變數之間的關係，並且對相應的假設進行檢驗。由於這種跨層次效果在行為科學和社會研究中十分普通，因此多層次模型的架構比傳統方法有明顯的優越性。

社會研究例證

Mason 等人（1983）檢驗了 15 個國家中，母親的教育程度和城鄉居住類型對其生育率的影響。大家都知道，在許多國家，教育程度高和城市化水準高會導致生育率較低。然而，研究者提出的問題是，這種影響也許取決於各個國家的特徵，包括以國民生產總值（GNP）所測量的國家經濟發展狀況和計畫生育工作的強度。

Mason 等人發現，在所有的國家中，母親教育程度較高的確與較低的生育率有關聯。但是，城鄉生育率的差別在各國之間有所不同，處於差別兩極的情況是國民生產總值很高的國家和幾乎沒有組織化計畫生育工作的國家。有關差別效果（differentiating effects）的識別，例如城鄉差距的識別，以及它們變化的預測，將在第 5 章加以討論。

發展研究例證

語言能力發展研究假設個人詞彙的獲取取決於兩個方面：一是適當的語言環境，二是從環境中學習的先天能力。人們普遍認為，兒童詞彙發展上的差別主要是由先天能力上的差別決定的。然而，實證研究結果並不太支持這種假設。遺傳可能性研究發現，父母的標準化詞彙測驗得分只能解釋子女在該測驗得分差別中的 10%~20%。語言環境研究的結果也好不了多少，大部分個人詞彙獲取上的差別仍然無法得到解釋。

過去，研究者們主要採用兩時點設計來研究語言環境效果。例如，先在 14 個月大時檢測嬰兒的詞彙量，同時取得母親的語言能力或整體語言能力。然後間隔一段時間後，例如在嬰兒 26 個月大的時候，再次檢測嬰兒的詞彙量。這些資料採用傳統線性模型進行分析，其研究焦點是，在控制嬰兒的「初始能力」（即 14 個月時掌握的詞彙量）的條件下，看母親的語言能力對其 26 個月大的嬰兒的詞彙量的影響。

Huttenlocher 等人（1991）蒐集了一些關於兒童的縱貫資料，在其年齡為 14 個月至 26 個月期間進行觀測，觀測最多的達到了 7 次。此一資料可以用每個孩子的縱貫重複檢測值對其個人詞彙量的發展軌跡建立模型，每個孩子的成長由一套參數來反映。第二個模型應用了關於這個孩子的其他資訊，包括孩子的性別、母親在家中的說話量，來預測孩子詞彙量的成長參數。有關此一模型的實際模式化與估計將在第 6 章中詳細討論。

此一分析應用了階層線性模式，最後顯示語言環境在嬰幼兒期詞彙量增長中的作用遠比以往研究所報告的水準要大得多。（實際上，如果對比原來僅用兩時點檢測值的傳統方法分析結果，所顯示的差異還要更大。）此一應用研究顯示了傳統方法在成長與其相關因素關聯研究方面的效度很差。

變異數共變數成分的分解

階層線性模式的第三種用途是對非平衡嵌套數據中的變異數共變數成分進行分解估計。例如，教育研究常常分析學生個體在不同的班級和學校組織脈絡的成長過程，對此一現象建立正式模型需要包括三個層次。

Bryk 和 Raudenbush（1988）使用持續效果研究（Sustaining Effect Study;

Carter, 1984) 的縱貫資料的一部分樣本來示範此一方法。此一資料包含 86 個學校中的 618 名學生在一年級至三年級期間所進行的 5 次數學考試成績。首先，他們對各學校的每個孩子建立個體學業成績的成長（或重複測量值）模型。三層次方法使這些個體成長軌跡的差別得以分解為學校內部的差別成分和學校之間的差別成分。此一應用研究的細節將在第 8 章中進行討論。

該研究的結果令人驚訝，成長率中有 83% 的變異數反映的是學校之間的差別。相對的，只有大約 14% 的成長變異數反映的是學校之間在學生進校時水準上的差別，這種結果與關於學校影響橫斷研究的典型結果是一致的。此一分析說明，學校之間的差別非常大，然而傳統模型卻檢測不出來，因為傳統模型沒有包含將學習指標進行校內差別成分與校間差別成分的分解。



本書第 1 版問世以來的新發展

隨著對階層模型興趣的增加，方法上的創新也日益加速，社會科學和醫學方面都產生了許多極富創造力的應用。第一，模型所採用的結果變數已經擴展了範圍；第二，模型現在不僅可以包括純粹的嵌套資料結構，而且還可以包括交互嵌套的資料結構；第三，多變項結果變量影響問題的應用已經成為主流；第四，潛在變數已經被植入階層模型；第五，對階層模型的貝氏推論獲得了更廣泛的普及和應用。這些新發展再與現在大大提高的運算能力結合起來，使得這種模型的擴展引發了整個系列的新應用。

結果變數範圍的擴展

本書的第1版僅將其關注限於連續型分配的層 1 結果。對於這種類型的結果，經常可以有理由假定模型的誤差是常態分配，統計推論可以在常態分配理論範圍內進行較容易的運算。當結果屬於離散型變數時，推論便更為複雜。明顯的例子包括二分變數的結果（如就業與失業）、計數資料（如某鄰里社區一年中所發生