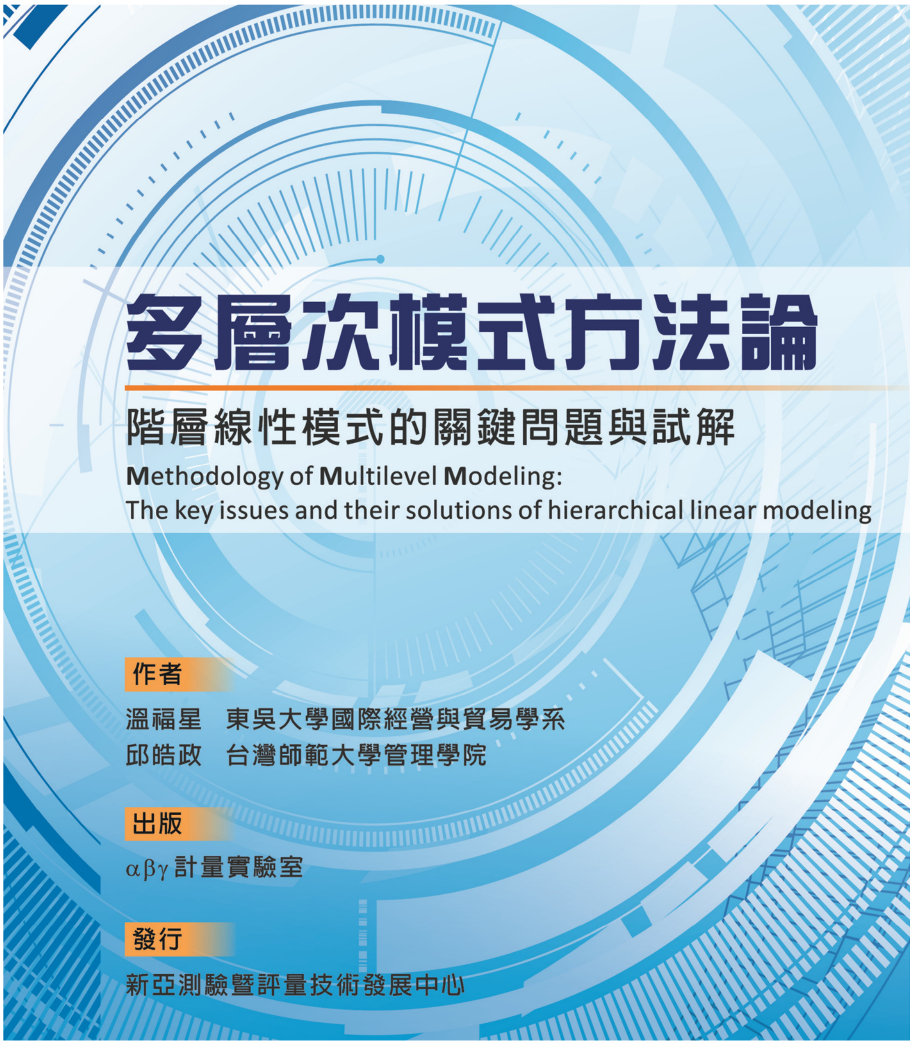

多層次模式方法論
階層線性模式的關鍵問題與試解
Methodology of Multilevel Modeling:
The key issues and their solutions of hierarchical linear modeling
作者
溫福星　東吳大學國際經營與貿易學系
邱皓政　台灣師範大學管理學院
出版
αβγ 計量實驗室
發行
新亞測驗暨評量技術發展中心

Chapter 1

導論：從 GLM 到 HLM

1.1 概說

多層次研究（multilevel research）可以說是當代組織與管理領域最熱門的新興研究典範（Bliese, Chan, & Ployhart, 2007; Kozlowski & Klein, 2000），而階層線性模式（hierarchical linear modeling，HLM；Bryk & Raudenbush,1992）則是一種將迴歸擴展到階層資料結構（hierarchical data structure）的統計分析技術。雖然目前學界已有多種不同的方法與軟體來分析多層次資料，並快速擴展到縱貫資料、非線性模型、潛在變數模型，但若非 Goldstein、Snijders、Bryk 與 Raudenbush 等學者在階層線性模式的原理進行深入探究並持續開發便捷的軟體，多層次研究或許未有今日的成熟發展。

1.1.1 多層次模式概念的發展

事實上，社會科學研究者並非無視於多層次分析的重要性，例如早在一甲子之前，Robinson（1950）就已經指出不同層次的資料若無適當的分析與推論將會犯下嚴重的錯誤。但階層線性模式遲至九十年代才獲得發展，主要的原因之一是受限於一般線性模式的傳統分析框架。在過去，傳統一般線性模式（general linear models，GLM）的迴歸分析或是變異數分析（註[1]），僅能處理單一層次變項間的關係，而無法同時處理包含個體與總體等不同層次的跨層級資料（註[2]）。如果要將不同層次的

註[1] 指固定效果模式的變異數分析。
註[2] 指個體與總體等不同層次的誤差項。

分析單位整合在一個線性模式中，必須將傳統的 GLM 擴充到 HLM。在此同時，模型的提出必須要有充分的理論依據或邏輯推論，再加上資料分析方法及結果解釋均要與研究設計有關，因此階層線性模式不僅是統計分析技術，更是一種統計方法學（statistical methodology）（Courgeau, 2003）。

對於多層次資料結構的分析技術，經過半世紀的發展，在不同領域各有不同的專業用語，如表 1.1 所示。本書主要採用 Lindley 與 Smith（1972）以及 Raudenbush 與 Bryk（1992, 2002）的階層線性模式（HLM）來說明多層次資料的分析技術，主要是配合 HLM 軟體的運用，更重要的是因為 HLM 背後的統計語言最容易被理解，同時可以應用在多種不同的研究情境中，例如成長研究、組織效能、多變項分析等，因為這些資料均存在著相同的結構特徵：階層結構。

表 1.1 不同領域對於多層次模式的用語

社會學	多層次線性模型	multilevel linear models
生物統計	混合效果模型	mixed-effects models
	隨機效果模型	random-effects models
計量經濟	隨機係數迴歸模型	random-coefficient regression models
統計學	共變數成分模型	covariance components models
	階層線性模式	hierarchical linear models

階層線性模式最早是由 Lindley 與 Smith（1972）利用線性模型的貝氏估計技術來處理複雜的誤差結構所發展出來的概念，後來被 Raudenbush 與 Bryk（1992, 2002）應用在 HLM 軟體中而獲得普及。事實上，Lindley 與 Smith（1972）最初提出 HLM 模型時並沒有獲得重視，

因為他們的模型需要對非平衡資料進行共變數成分估計，複雜的估計程序使得這套概念僅能應用在一些簡單的問題，直到在期望最大化演算法的進步（EM）（Dempster、Laird & Rubin, 1977），使得共變數成分估計的運用得以具體實現於一般研究議題中，進而被導入多層次資料結構的分析（Dempsteret al.,1981）。例如 Strenio、Weisberg 與 Bryk（1983）即應用此一技術解決橫斷資料中多層次結構的計算問題。後來，又出現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（scoring）演算法（Longford, 1987），使得階層線性模式的數學模型與分析技術得以完備，並導入統計軟體，例如 HLM（Raudenbush et al., 2000）、MIXOR（Hedeker & Gibbons, 1996）、MLwiIN（Rasbash et al., 2000）、SAS 副程式 Proc Mixed（Little et al., 1996）、VARCL（Longford, 1988），以及完全貝氏方法則可應用 BUGS（Spiegelhalter et al., 1994）來進行分析。

1.1.2 多層次模式的當代發展

過去三十年來，隨著研究者對於多層次數據分析的重視以及技術的成熟，關於多層次資料的應用也不斷推陳出新，例如結果變數已經擴展到不同型態的廣義模型（例如二分變數、多分類別變數等離散型變數、順序變數、計數資料等等）。為了能夠處理非連續性的資料，Stiratelli、Laird 與 Ware（1984）以及 Wong 與 Mason（1985）最早導入最大概似估計（ML）進行一階近似處理。Goldstein（1991）和 Longford（1993）沿用他們的概念進一步發展分析軟體來處理不同類型的離散型結果變數的二層和三層的多層次模型，但是不幸的是他們的技術仍有估計不準確的嚴重問題。後來由 Hedeker 與 Gibbons（1993）以及 Pinheiro 與 Bates（1995）提出了高斯-赫米特積分對最大概似值的準確近似計算法，後來

被導入 Mixor 軟體和 SAS Proc Mixed 軟體。HLM 軟體則使用高階拉普拉斯轉換這一種也十分準確，而且計算方便的近似方法（Raudenbush, Yang, & Yosef, 2000），這些技術的發展使多層次模型的原理可以處理多種不同類型的結果變數。

其次，多層次模式不僅可以應用在單純的嵌套資料結構，而且還擴展到交互嵌套（cross classified）的資料結構。所謂單純的嵌套資料結構是指每一個下層單位僅嵌套在單一個上層單位中，例如某班學生嵌套於某一班級，該班級又嵌套於某一學校時，但是在一般研究情形下，嵌套結構可能會更為複雜。例如在管理研究中，常會遇到矩陣組織的狀況，某一個員工屬於某一個功能性部門劃分的結果（例如行銷單位）但是也屬於另一個區域組織的劃分結果（例如大中華事業群），此時即是一個交互嵌套的現象。Raudenbush（1993）對於交互嵌套的資料結構提出了解決方案，並導入 HLM 軟體中。

多層次模式的第三個新趨勢是結果變數從單一變數擴展到多個變數的多變量模式（hierarchical multivariate linear modeling/multivariate multilevel modeling）。此項擴展最重要的應用在於縱貫資料與重複測量資料的分析。縱貫資料基本上是一種平衡設計（在沒有遺漏值的情況下），亦即每個個體的測量次數都相同，並且測量時距間隔也相同。此時即可採用多變量重複測量的共變數模型進行分析，同時處理自我相關誤差和隨機變化的斜率問題，或是擴展到非限定模型，對各時點的變異數進行獨立分析。在 SAS 軟體（1996）的副程式 Proc Mixed、Hedeker 與 Gibbons（1996）的軟體 Mixor、以及 Raudenbush 等人（2000）的 HLM 軟體，都利用這種策略來進行多變量結構的估計。

第四種新的應用取向，是潛在變數模型（latent variable model）的應用，使得多層次模式也能處理潛在變數的估計。其原理也是一種多變量迴歸技術的延伸，利用完整資料（complete data）的推估來預測估計

測量誤差。換言之，研究者蒐集到的資料是一種觀測資料（非完整資料），藉由觀測資料來推估無法被觀測的潛在變數（Little & Rubin, 1987; Little & Schenker, 1995）。在具體作法上，模式的第一層代表了帶有測量誤差的觀測資料，反映著潛在的「真實」資料之間的關聯，在層二即可進行潛在變數的估計。此種模型可進一步的延伸到存在測量誤差的預測變數的直接效果和間接效果的估計與分析。此外，也可以套用於項目反應理論（item response theory），以一種兩層次模型進行測驗的反應概率的項目特徵分析（ICC），找出與個體的「能力」或「潛在特質」的函數。

多層次模式的第五種後期發展是導入貝氏估計法，使得多層次模型的統計模式可進行貝氏推論（Bayesian inference），使得多層次估計更加精準且有彈性。傳統以來，統計推論都基於最大概似法（ML）來進行，此法的優點是參數估計具有一致性，並且具有漸近不偏性和有效性，其前提必須是大樣本下才能符合常態分配的要件。但是在多層次模型中，高層或上層單位未必符合大樣本條件，此外，各組內部的觀察單位數也多呈現不平衡狀況（例如各校抽取的學生數不同），此時不平衡程度也會影響最大概似法的估計正確性。在這些情況下，貝氏方法恰好能夠補足其缺點。從統計原理來看，貝氏估計利用蒙特卡羅技術對特定後驗分配的近似演算，例如資料增廣（data augmentation）（Tanner & Wong, 1987）和吉布斯抽樣（Gibbs sampling）（Gelfand et al., 1990; Gelfand & Smith, 1990），使得複雜的估計得以完成，進而得到更接近實際情況的標準誤，從應用面來看，貝氏估計透過參數的後驗分配資訊的導入，可以對研究議題進行更有彈性的推論與分析。現在這些方法已經廣為被軟體採用（Spiegelhalter et al., 1996; Rasbash et al., 2000），大大提高運算與實際研究應用的便利性。

1.2 迴歸分析的基本概念

文獻上關於 HLM 的原理已經有相當多的介紹（温福星，2006; Goldstein, 2003; Hofmann,1997; Raudenbush & Bryk, 2002; Snijders & Bosker,1999），本書不再贅述這些基本原理與技術，而是著眼於使用 HLM 背後的思考脈絡與所潛藏的各種問題。在使用這技術進行研究之前，研究者除了瞭解分析程序與解釋方法之外，更必須熟悉多層次研究與傳統研究的差異，並進一步瞭解新方法的限制與可能帶來的新問題。例如，HLM 與 GLM 的最大差異在於 HLM 模式引進了不同層次的誤差項，而 GLM 模式只有一個誤差項，而且是屬於個體層次。

過去，研究者若要探討個體與總體等不同層次的跨層級資料，共有兩種類型的迴歸分析可以處理，一個是將總體層次的資料解構（disaggregated）或是進行虛擬化處理（dummy），使成為個體層次資料來進行個體層次的迴歸分析；其次，是將個體層次的資料聚合（aggregated）成總體層次的變項，進行總體層次的迴歸分析。這樣的分析方法是取決於研究目的，如果我們想研究的是個體與總體等不同層次的跨層級資料對個體層次結果變項的影響時，解構方式的個體層次迴歸分析就會遇到資料獨立性與同質性假設被違反的可能，得到所估計的迴歸係數標準誤被低估的情形發生，導致容易拒絕虛無假說的型 I 錯誤膨脹結果。這是因為忽略了相同總體層次下的個體資料間存在高度的相關，亦即個體與總體層次的資料彼此間具有「內屬、巢套、叢集、鑲嵌（nested、clustered、embedded）」的結構特性，例如員工內屬於公司、團隊成員內屬於團隊等，由於多這一層關係，相同公司下的員工、相同團隊下的成員，因組織文化、組織氣氛等因素的潛移默化，使得員工或成員在同一組織（公司或團隊）內較為相似，在不同的組織間較相異。

1.2.1 基本迴歸原理

一般迴歸分析在探討一群來自母體抽樣下的樣本，研究一些自變數對我們所關心依變數的探討，是否這些自變項能夠顯著或是有效地解釋依變項的變異，或是這些自變項能夠提供對依變項的瞭解，進而對其預測。如果僅有單一解釋變數（X）來解釋結果單一結果變數（Y）時，迴歸方程式如(1-1)所示。

$$Y_i = \beta_0 + \beta_1 X_i + e_i \tag{1-1}$$

在 GLM 取向的迴歸分析中，隨機抽樣非常重要，因為隨機抽樣得以使(1-1)式的誤差項（e_i）服從以 0 為平均數、σ_e^2 為變異數的常態分配。如果是對所有母群採用簡單隨機抽樣時，這些樣本是反映母體得部分集合。例如我們關心父母的社經地位（X）是否對孩童的學業成績（Y）產生影響，則我們必須對測量單位（孩童）進行隨機抽樣，透過對母體（全台灣的學童）採用簡單隨機抽樣進行調查，所抽取到的學童來調查其父母親教育程度、職業階層，計算出父母親的社經地位分數，並蒐集孩童的學業成績，最後得以取學業成績（Y）對父母社經地位（X）作迴歸，檢驗學童父母親的社經地位對其學業成績的影響，其影響力由斜率（β_1）參數反映，β_0 為方程式的截距。

在最小平方法（ordinary least squares；OLS）或是最大概似法（maximal likelihood；ML）求解下，迴歸係數 β_1 的估計值可由(1-2)式求得，其意義為共變數佔解釋變數變異數的比例，其數值大小會受到 X 與 Y 單位的影響，而為未標準化係數，從幾何意義來看，未標準化的斜率是指當 X 每變化一個單位時，在 Y 所變化的數量。

$$\beta_1 = \frac{\mathrm{cov}(x,y)}{{s_x}^2} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{SP_{xy}}{SS_x} \quad (1\text{-}2)$$

如果將未標準化係數乘以 X 變項的標準差再除以 Y 變項的標準差，即可去除單位的影響，得到一個**標準化迴歸係數**（standardized regression coefficient），如(1-3)式中的 β'，即為一般俗稱的 Beta 係數（未標準化係數一般習慣上會以英文小寫 b 來標示）。標準化迴歸係數也可視為將 X 與 Y 變項所有數值轉換成 Z 分數後，所得到的 OLS 迴歸斜率。

$$\beta_1' = \beta_1 \frac{s_x}{s_y} \quad (1\text{-}3)$$

由於標準化的結果，(1-1)式中的截距項 β_0 隨即消失，而 β_1 係數的數值類似相關係數，其可能數值範圍應會介於±1 之間（註[3]），其絕對值越大者，表示解釋力越強，正負方向則代表 X 與 Y 變項的關係方向。

如果 X 變數（社經地位）的數值為類別（質性）變數，例如高中低三組而非表示強度的連續變數，X 對 Y 的解釋即為變異數分析模型，以一般線性模式來表現如下：

$$Y_{ij} = \mu + \alpha_j + e_{ij} \quad (1\text{-}4)$$

公式(1-4)中，Y_{ij} 代表的是社經地位為類別 j 的受試者 i 的學業成績，每個受試者的社經地位 (X) 屬於 J 個類別中的一個，社經地位 j 對學業

註[3] 在特殊情況下（例如帶有交互作用項或多項式迴歸中），標準化迴歸係數的數值可能會超過±1 的區間。

成績的影響以 α_j 來反映，e_{ij} 為誤差。相同的，e_{ij} 仍服從以 0 為平均數、σ_e^2 為變異數的常態分配。

解釋變數 X 不論是連續或類別變數，在一般線性模式中都有類似的模式組成要素：常數、解釋變數效果與誤差。常數是指沒有解釋變項時的依變項期望值 μ（例如變異數分析(1-4)公式），或是解釋變項為 0 時的依變項狀態（亦即方程式(1-1)的截距項 β_0），誤差是依變項無法由常數與解釋變項解釋的部份。解釋變數效果為研究者所關心的自變項效果，其影響力大小由斜率表示，斜率若未經過標準化處理，則反映了解釋變數每單位變動量在結果變數上的增減量，但如果經過標準化處理，則類似於相關係數的一種標準化估計數（註[4]），但是此時截距項則會消失。當自變項數目不只一個時，解釋變數效果就可能包含多種不同的解釋效果項（例如主要效果與交互作用效果），此時對於斜率參數的解釋就變得相對複雜。

在一般研究實務中，解釋變數的影響力為研究者所關心的焦點所在，因此研究者所關注的焦點在於斜率而非截距。當自變項數目只有一個時所進行的迴歸分析稱為簡單迴歸（simple regression），斜率數值經過標準化後恰等於相關係數，換言之，當解釋變數只有一個時，可以直接取相關係數作為 $X \rightarrow Y$ 的影響力。當解釋變數超過一個以上時，解釋變數效果就可能包含多種不同的解釋效果項，每一個解釋效果項的影響力由個別的斜率參數所反映。例如(1-5)式中以 X_1 與 X_2 兩個解釋變數所進行的多元迴歸中，β_1 與 β_2 反映了簡單效果（simple effect），β_3 反映了交互作用（interaction effect）。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + e_i \qquad (1\text{-}5)$$

註[4] 在一般的多層次迴歸中，例如 HLM 軟體，為使截距項反映總體層次測量單位的估計數，皆使用未標準化迴歸係數。

當反映交互作用的β_3數值為 0 或其效果可以被忽視時，可以將交互作用項自(1-5)式中移除，此時成為一個帶有兩個解釋變數的多元迴歸（multiple regression），但是當β_3數值不為 0 或其效果不可被忽視時，(1-5)式中的交互作用項就具有重要的討論價值，此種迴歸稱為交互作用迴歸（multiple regression with interaction）（Aiken & West, 1991），可以進行調節效果（moderation effects）的分析討論，因此又稱為調節迴歸（moderation multiple regression; MMR）。

1.2.2 斜率參數的意義

1.2.2.1 控制變數的干擾效果排除

不論是一般的多元迴歸或帶有交互作用項的迴歸分析中，對於斜率參數的解釋，必須注意其統計上的特性。如果模型中只有一個解釋變數，在一般情況下（註[5]），斜率的計算不受截距與誤差項的影響，此時斜率是一種零階估計數（zero-order estimators），對其解釋不必考慮其他因素。但是如果模型中有一個以上的解釋項，β_1是指當其他解釋項對依變數的影響力被估計完畢之後，X_1對Y的邊際解釋力（marginal effect）。以(1-5)式為例，每一個斜率參數是指當其他解釋項（X_1、X_2與X_1X_2交互作用項）被估計後的額外解釋力（incremental effects），而非該解釋項原始的零階影響力。

以統計術語來說，多元迴歸方程式中的斜率參數是指當「其他解釋項」的效果維持恆定時的淨解釋力，此時「其他解釋項」的混淆或干擾效果（confounding effects）係透過統計方法排除（partial-out）或被統計加以控制，使得X_1對Y的解釋不受其他解釋項的干擾，因此又被稱為非

註[5] 在特殊情況下，例如時間序列數據，會發生誤差項的自我相關現象而影響斜率的估計，在此服從迴歸分析基本假設。

零階的淨效果（partial effects）。在(1-5)式時，因存在 X_1X_2 交互作用項，此時 X_1 對 Y 的解釋不稱為主要效果（main effect），而是稱為簡單效果。如果不存在交互作用項，此時「其他解釋項」又可被稱為控制變數（control variables），在社會科學研究中，不論是實驗研究或非實驗研究，都經常會將一些會對研究結果可能會產生干擾作用的變數加以測量（例如年齡、智力、年資），然後放入方程式加以控制，藉以觀測研究者所關心的解釋變數如何對依變數產生「真正的」、「乾淨」解釋力。

但是特別值得注意的是，並非那些被研究者宣稱作為「控制變數」者會對特定解釋變數的解釋力產生控制作用，只要在放在迴歸方程式的所有變數，包括控制變數與其他解釋變數，都會對某一個解釋變數的影響力產生控制效果，例如研究者納入了「動機」與「努力」去預測學業成就，同時以「年齡」與「智力」作為控制變數，此時對於「動機」變數而言，控制變數除了「年齡」與「智力」以外，只要「努力」變數也同時存在於方程式，那麼控制變數就有三個。

1.2.2.2 交互作用的調節效果分析

前面一節所提及的控制效果，是指在一般多元迴歸中，多個解釋項之間的相互干擾效果進行統計排除下，對於斜率進行解釋的一種特殊狀況。另一種更為特殊的狀況，是當兩個解釋變數對於依變數具有交互作用時，對於斜率的解釋，又比前一節所介紹的控制效果更為複雜，在方法學上稱為調節效果（moderation effect）分析。

以(1-5)式為例，X_1X_2 是一個由兩個解釋變數相乘得到的交互作用項，這一項不僅是多元迴歸的第三個解釋項，更是一個與前面兩個解釋項所衍生出來的解釋項，因此其性質迥異於一般的解釋項。

在(1-5)式中，如果沒有交互作用項，β_1 都是指當 X_2 為固定時 X_1 的斜率；β_2 也都是指當 X_1 為固定時 X_2 的斜率，β_1 與 β_2 是控制彼此後

對 Y 的淨解釋力，我們可以把此一模型稱為控制模型。

但是，如果(1-5)式的交互作用項存在時，β_1 與 β_2 則是控制彼此並控制高階交互作用項（X_1X_2）後的解釋力，由於 X_1X_2 與 X_1 及 X_1X_2 與 X_2 之間具有相依性，因此帶有交互作用項的 β_1 與 β_2 並非控制模型中的「淨」效果，若要以一般的迴歸係數的解釋策略：控制其他變數後的解釋變數影響力通常無法合理解釋 β_1 與 β_2 的意義。而要改採調節效果的概念，把數學中的條件化概念加以導入，才能正確解釋迴歸係數的意義，因此交互作用迴歸又被稱為調節模型。

在調節模型中，各解釋變項的解釋力是一種條件化解釋力，亦即其中一個解釋變數對依變數的影響，是在另一個解釋變數（做為調節變數）的不同水準下會有不同的解釋力，稱為「調節解釋力」，(1-5)式中的迴歸係數 β_3 表示了調節效果的強弱。由(1-5)式整理得到(1-6)即可得知其意義（為了便於說明，將各變數足標 i 予以省略）。

$$Y = \beta_0 + (\beta_1 + \beta_3 X_2)X_1 + \beta_2 X_2 + e \tag{1-6}$$

由(1-6)式可知，X_1 對 Y 的影響不只由 β_1 所反映，而是由 $\beta_1 + \beta_3 X_2$ 決定。當 β_3 估計值達顯著水準時，表示交互作用存在，兩個解釋變數會彼此調節其對 Y 的影響力。X_1 對 Y 的影響力隨 X_2 的不同狀況而變（被調節）；如果交互作用不顯著（β_3=0），X_2 的調節影響即可忽略，X_1 對 Y 的影響力不隨 X_2 的不同狀況而變（不被調節），交互作用可以移除，退化成控制模型。此時模型中雖然沒有交互作用項，但是調節變數 X_2 的並沒有完全消失，方程式中仍保有 $\beta_2 X_2$ ，此時 X_2 不再作為調節變數而成為 X_1 的控制變數。

從前述的討論中，我們可以得到兩個結論，第一，控制模型巢套在交互作用模型中，因為控制模型比交互作用少了交互作用參數，或是說，控制模型是將交互作用項參數設為 0 的一個特殊模型。換言之，控制模型是交互作用中的一個特例，這兩種模型可以套用巢套比較策略來檢定模式適配度的變化。第二，交互作用模型中的斜率意義的解釋方式與控制模型不同。如果交互作用顯著，必須對斜率參數進行條件化解釋。一般在研究實務中，對於調節效果的解釋必須取其中一個解釋變數作為調節變數，另一個解釋變數則作為主要解釋變數，進行單純效果（simple effects）的討論，分析在不同的調節變數水準下的斜率與截距，稱為簡單斜率（simple slope）與簡單截距（simple intercept）分析（註[6]）。

1.2.3 截距參數的意義

在一般的迴歸方程式中（例如 (1-1)式），截距是指當解釋變數為 0 的依變數數值。例如當 Y_i 是孩童的學業成績，X_i 是父母親的社經地位，β_0 為截距項，β_1 為斜率，當學生父母的社經地位等於 0 時，學生學業成績的預測值才會等於 β_0。

為了解釋的緣故，進行迴歸分析時需進行一些修正，不直接用父母親的社經地位原始數值作為自變項，而是以父母親社經地位的標準分數做為自變項，也將學童學業成績標準化，此種設計就是標準化的迴歸分析，截距項的解釋就不必進行特殊處理，即能反映平均情形。但是將自變數加以標準化不僅改變截距，同時也會將變數的變異數固定為 1，當解釋變數數目不只一個時，同時模式複雜時（例如包括交互作用項時），將會改變解釋變數的相對位置，影響斜率的估計。

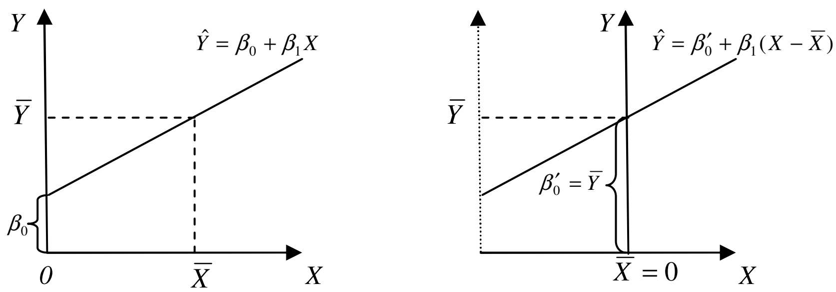
註[6] 對於調節迴歸的分析策略可參考邱皓政(2010)的第十二章的說明。

另一種廣為被採用的替代作法，是以解釋變數中心化來移動截距的位置。例如將父母社經地位（X）距離所有樣本中父母社經地位平均值（$\overline{X}$）的差距，亦即離均差分數（X - $\overline{X}$）作為迴歸分析的自變項，此一離均差分數的獲得歷程稱為中心化（centering），在本書稱為平減（因為是減去平均數之故），此時(1-1)式可轉換(1-7)式。

$$Y = \left(\beta_0 + \beta_1 \overline{X}\right) + \beta_1\left(X - \overline{X}\right) + e \tag{1-7}$$

在簡單迴歸中，截距（β_0）的一般最小平方法估計值是$\overline{Y} - \beta_1 \overline{X}$，代入(1-7)式，可得到(1-8) 式。

$$Y = \overline{Y} + \beta_1\left(X - \overline{X}\right) + e = \beta_0' + \beta_1\left(X - \overline{X}\right) + e \tag{1-8}$$



(a)X 變數未平減　　(b)X 變數減去平均數

圖 1.1 解釋變數中心化（平減）後的截距意義變化圖示

在未進行解釋變數中心化之前，截距 β_0 是指當解釋變數為 0 時 Y 的起始值，如圖 1.1(a)所示。如果以學童學業成績對學童父母親社經地位的離均值進行迴歸分析，如(1-8)式，但仍然以符號 β_0' 作為截距項，β_1 為斜率，則 $\beta_0' = \bar{Y}$ 。此時截距即為學童平均學業成績，斜率仍然為學童父母社經地位對孩童學業成績的影響，不受平減的影響（如圖 1.1(b)）。

由前述的平減方程式(1-8)可知，對解釋變數進行平減的主要影響，是將截距轉換成 $\bar{Y}$ 且其他參數（斜率與誤差）維持不變。換言之，如果 X 未經平減，則截距所反映的是當 X=0 時 Y 的數值。如果 X=0 代表絕對零點或 X 數值範圍包括 0，此時截距可解釋成 Y 的起始值，但如果 X 的數值範圍不包括 0，X 未平減時的截距則無解釋上的意義。

傳統上，研究者所關注的是迴歸方程式中的斜率而非截距，因為斜率 β_1 代表 X 對 Y 的影響力，經顯著性考驗若證明 β_1 顯著不等於 0，研究者可宣稱 X 對 Y 的影響具有統計意義。相對之下，研究者對 β_0 截距的估計與顯著性考驗並不關心。但是在多層次模式中，截距具有重要的意義，隨機截距用來反映各組平均的差異，因此截距的意義與檢定不可忽略，連帶使得會改變截距意義的平減操作成為重要議題。

另外，解釋變數平減對於交互作用迴歸也有重要的價值，如果交互作用項為經過平減的解釋變數相乘，將可使交互作用項與主要效果項的相關大幅降低，有利於迴歸係數的解釋與呈現。同時也可以使截距回到平均數的位置，如新方程式如(1-9)式，在多層次分析中非常關鍵。

$$Y = \beta_0' + \beta_1' X_1' + \beta_2' X_2 + \beta_3' X_1' X_2' + e \tag{1-9}$$

有一點必須注意的是，(1-9)式中經過平減的 X_1' 與 X_2' 兩個新解釋變數本身的平均數為 0，理論上會使得方程式(1-9)的截距回到 $\bar{Y}$ 的位置，但是平減後相乘的交乘積項 $X_1'X_2'$ 並非平均數為零，而是期望值為 X_1 與 X_2 的共變數這個新變數（Bohrnstedt & Gikdberger, 1969; Hays, 1988）。

$$E((X_1 - \bar{X}_1)(X_2 - \bar{X}_2) = Cov(X_1, X_2) \qquad (1\text{-}10)$$

如果(1-9)式的截距要回到 $\bar{Y}$ 的位置，$X_1'X_2'$ 還必須再進行一次平減：共變數中心化（減去 X_1' 與 X_2' 的共變數），否則僅針對 X_1 與 X_2 進行平減忽略乘積項的共變數，則中心化無法還原截距成為結果變數的平均數。

X_1' 、X_2' 與 $X_1'X_2'$ 三者經過平減之後所進行的調節迴歸分析，交互作用的參數估計與顯著性考驗不會改變，但低階的迴歸係數一來因為變數間的相關降低而得以避免共線性問題，二來各參數變得更容易解釋，β_1' 是指當 X_2' 為零時 X_1' 的斜率、β_2' 是指當 X_1' 為零時 X_2' 的斜率的條件效果。更重要的是，各項經過原點歸 0，截距才可反映結果變數的平均數 $\bar{Y}$ 。

1.3 多層次模型分析的方法學概念

1.3.1 隨機取樣與誤差獨立性

在一般迴歸方程式中，例如(1-1)式，可以發現(Y_i, X_i)代表一個個體，也就是我們所分析的單位，一個家庭的孩童(i)學業成績(Y)與其父母親的社經地位(X)。此時的研究設計是以簡單隨機抽樣為抽樣方法，抽出的分析單位為學童，因此迴歸分析的資料單位是學童，就是(1-1）式的下標i。但是實際上的研究設計可能不是如此，特別是在教育或組織研究，我們常常採取的抽樣方式是隨機集群抽樣，通常是多階段的隨機集群抽樣（multi-stage random cluster sampling ），此時雖然統計分析的單位是學童(i)，但是要抽取學童之前，會先抽取班級，而且將這個被抽中的班級中每一個學童都進行全面的資料蒐集。

換言之，要抽個體(i)之前先抽取班級，此時班級就是隨機抽樣的單位，所有合乎調查研究的班級都會被彙整一起成為抽樣架構。有時彙整班級可能更為複雜，在抽取班級之前，先抽取學校，甚至在抽取學校之前，先抽取鄉鎮或縣市，此時即是多階段集群抽樣，最後所抽取的班級，所有的學生都一一接受施測或調查。組織管理或是國際企業的研究也類似，如果研究者想分析的單位是員工或是子公司，探討這些員工的特性與其績效的關係，或是組織特性與員工績效甚至滿意度的關係時，往往員工或子公司資料的蒐集是透過公司組織或是母公司的方式取得，由於企業組織員工的研究，在資料的取得相當困難，當公司或企業數不多時，我們會不排斥來自同一家公司的許多員工資料。換言之，在這個研究裡，因為所蒐集的「樣本」資料中，有許多員工資料來自同一公司或企業組織，儼然就是多階段的架構，好像員工樣本是先抽取公司別，

再從各個公司別抽取不同的員工，進行調查的樣子，這就是巢套或是內嵌、鑲嵌的樣本設計。

在這樣多階段隨機集群抽樣架構下所抽取的個別樣本，都隱含帶有受不同階段抽樣架構的影響，例如不同鄉鎮縣市下，可能隱含著城鄉差距的問題，有些縣市的教育資源豐沛，有些則欠缺，最明顯的例子就是偏遠地區的英文與資訊教師，相對省、直轄市的師資設備是相當的匱乏缺乏，這根本就直接影響到學童的學業成績。同樣地，不同的班級學校，可能又隱含學校老師、甚至校長的教學理念與教學方法，同時學區的因素亦隱含父母社經地位的高低，以及學童受教育資源分配的影響等等。因此，造成孩童學業成績的變異，不單單只有簡單迴歸分析一個解釋變數：父母的社經地位所影響，其變異可能可以被城鄉差距或教育方法不同所影響。

在商管研究中也有類似的現象，例如我們常提到組織文化、組織階層等，以員工為個體的研究，都會參雜組織層面的課題。因此在許多社會科學的研究裡，有如經濟學有所謂的個體經濟（Micro Economy）與總體經濟（Macro Economy）之別，亦即在研究課題的分析上，分析單位或是資料蒐集單位有是個人與組織的區別。

1.3.2 多層次分析的方法學概念

根據 Courgeau（2004）的分類，所研究的模式有兩類的區別，一個是總體的層次（Macro 或 Aggregate Level），一個是個體的層次（Micro 或 Individual Level），以圖 1-3 說明之。

在圖 1-3 中，圖的下半部是屬於個體層次的分析架構，探討個體例如學童的父母親社經地位(X)對孩童學業成績(Y)的影響，分析的層次

是學童，也就是(1-4)式的 i。而圖 1-3 上半部分析的架構是整體或是總體的層次，在學童父母親社經地位對學童學業成績的影響這個例子，是指學校、班級或是縣市這個層級，簡單來說，就是班級或學校的學童父母平均社經地位(Z)，是否對整個班級或學校學童平均學業成績(Y^*)的影響。在組織研究中，這個總體層次的變項：學童父母平均社經地位(Z)稱為情境變項（Contextual Variable）。圖 1-3 的水平實線就是探討不同層次的解釋變項對結果變項的影響，而垂直直線則是個體層次的研究變項與總體層次研究變項間的關係，總體層次的研究變項可能是不一樣的總體層次特性或特徵，或是個體層次研究變項的加總或平均。

階層線性模式所要研究的是即圖 1.2 虛線的部份，配合集群抽樣或是鑲嵌巢套設計，除了探討個體學童的父母親社經地位(X)對孩童學業成績(Y)的影響外，亦探討研究總體層次的解釋變項(Z)如城鄉差距、教育方法或整體父母親社經地位對個體學童學業成績的影響，同時參雜總體變項與個體變項的解釋變數對個體被解釋變項的影響。

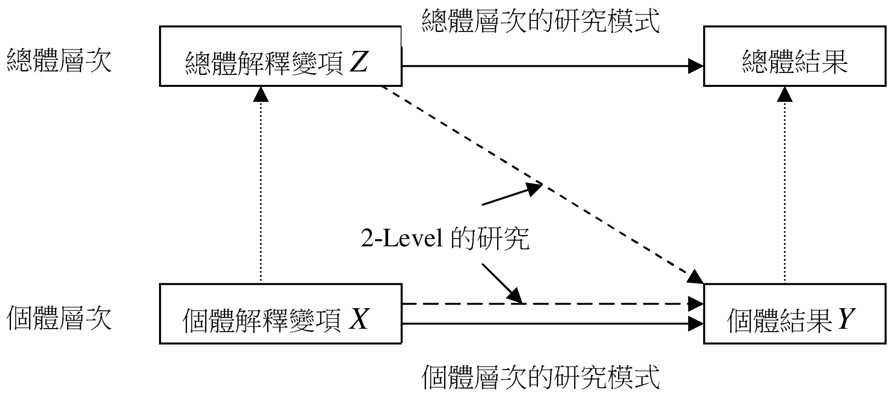
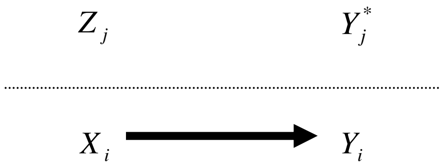
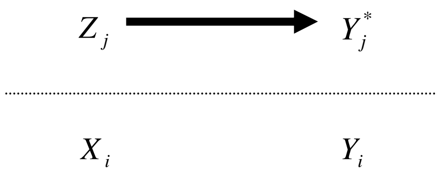


圖 1.2 總體與個體層次的研究方法設計

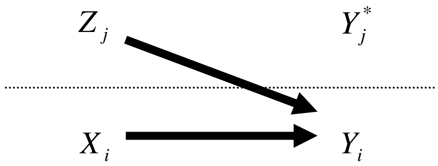
我們若將圖 1.2 簡化，簡單來說，如果我們研究的架構是下圖，則屬於是個體層次的研究，例如我們一般的簡單隨機抽樣，抽取出學童，蒐集其學業成績，與其父母的社經地位，研究父母社經地位對孩童學業成績的影響（注意變數符號的下標示是 i ）。



如果我們研究的架構是下圖，則屬於是總體層次的研究，例如我們一般的隨機抽樣，抽取出公司，蒐集其公司組織文化與其公司的經營績效，研究公司組織文化對公司經營績效的影響。或是，研究班級或學校的平均父母社經地位對該班級或校孩童學業成績的影響（注意變數符號的下標示是 j ）。



如果我們研究的架構是下圖，則屬於是階層線性模式的研究，如圖 1.2 的說明（注意變數符號的下標示有 i 和 j ）。



我們將上一頁圖示的說明改以統計符號來呈現，如(1-11)式所示：

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + \gamma Z_j + e_{ij} \quad (1\text{-}11)$$

(1-11)式中，j代表的是公司或是班級（學校），而i 代表在每個公司內的員工，或是班級學校下的學童，學童與員工是巢套（鑲嵌）在班級和公司裡，j為$1,2,3,\cdots J$，共有J 公司或班級，而i 是附屬於j下，所以這J 公司或班級各有$n_1, n_2, \cdots, n_j, \cdots, n_J$ 的學童或員工。而Y_{ij}代表學童的學業成績或員工的績效，X_{ij} 代表學童父母親的社經地位或員工的滿意度，Z_j 則代表該班級學校的平均父母親的社經地位（contextual variable），或是員工所屬公司的福利誘因（aggregate variable）。這裡的迴歸係數截距項α_j所有樣本中可能相同，也可能在每個班級或公司中都不同，斜率項β_j也一樣，在所有樣本中可能相同，也可能在每個班級或公司中都都不同，代表每個公司員工滿意度對其績效影響的程度不一樣，或是每個班級學童父母親的社經地位對學童學業成績影響的程度不一樣。而γ代表的是各公司的福利誘因對員工績效的影響，或是各班級平均的父母親社經地位對學童學業成績的影響。