

掌握量化社會研究之鑰： 評謝宇《迴歸分析》

李瑞中

中央研究院歐美研究所與社會學研究所

《迴歸分析》，謝宇著，北京：社會科學文獻出版社，2010年，387頁。

只有先走進西方社會學的深處去掌握其傳統的底蘊，並佐以對本土之種種關照，再回過頭來反照西方做深切的反思，才可能為臺灣社會學尋找出一條有個性、富創意、且能在全世界當中佔一席之地的希望之路來。（葉啟政 2005：213）

量化社會研究乃是西方社會學的重要傳統之一，但國內社會學界（特別是晚近）對這個傳統的批判與否定往往先於瞭解，於是在追求本土社會學的過程，我們對西方社會學的掌握常常侷限於西方社會學理論和質化研究的傳統，因此儘管葉啟政已經為我們指出一條「希望之路」，台灣社會學若要「在全世界當中佔一席之地」，特別在量化社會學這個傳統之下，似乎還有相當一段距離要努力。

對於量化社會學傳統的批判與否定，除了「人」的因素之外（瞿海源 2009），知識上的因素，尚且來自於隨著計量理論的快速發展，運算功能的大幅提升，以及大型社會調查等各類型量化資料益發繁複，不實際從事量化資料分析的學者對於量化社會學的現況就未必能夠全然掌握。葉啟政(1978)曾經替研究方法（在漢語世界中）的經典教科書《社會及行為科學研究法》撰寫〈因徑分析〉的專章，距離 Duncan (1975)出版同一主題的經典 *Introduction to Structural Equation*

Models 不過短短三年左右時間，可謂站在量化社會學的最前沿，但他晚近對統計社會學傳統的詮釋與批判（葉啓政 2001, 2005）卻遭到譚康榮質疑，認為與目前實際量化社會研究有很大的落差（Tam 2005）。兩位學養深厚的頂尖社會學者意見歧異，看在對統計社會學的科學哲學與量化社會研究的實際運作都不甚了了的芸芸衆生眼裡，就更如同「隙中窺月」了吧！

謝宇最近出版《迴歸分析》，開宗明義表示「希望那些對定量研究持負面態度和批評意見的學者也能學習統計知識，因為只有在真正理解了統計學思想之後，一個人才能對定量研究方法進行評價」（第 3 頁）。誠如本書扉頁所稱，迴歸分析是「社會科學領域中最基礎同時也是最經典的定量分析方法」，其他非迴歸分析的量化方法，如「序列分析」（sequence analysis），在社會學裡的應用價值，仍存爭議（Wu 2000）。因此這本書所探討的主題，恰好可以作為深化讀者對量化社會學理解的重要入門。由量化社會學巨擘謝宇引領，從最基礎的統計概念開始（如隨機變數、機率分配），到進階的追蹤資料分析和因果推論，配合豐富的研究實例與統計軟體 Stata 程式範例，讀者將對量化社會研究的方法與邏輯有極為通盤且深入的認識。相信無論是有志親身從事量化社會研究的學者，或是對量化社會學持懷疑與批判態度的學者，都將獲益良多。

除了傳統迴歸分析的教科書大多會涵蓋的課題，謝宇的《迴歸分析》最大的不同，也是與質化社會研究有所呼應之處，在於本書更深入淺出地介紹實際研究工作中，「檢驗社會學理論」常面臨到的課題，讓統計分析不只是描述資料，而得以與理論扣連。譬如教育社會學認為教育不單單只是人力資本的累積，文憑本身傳遞了人們地位的相對高低，才是真正影響就業與薪資的因素。因此有人提出教育年限作為獨變項，除了連續變項之外，還要將一系列虛擬變項（代表高中畢業、大學畢業等）投入迴歸方程式中。基於「統計」的理由，我曾指出此乃期期不可。旁觀者以為這是量化研究者「輕理論、重方法」的本位主義，其實不然。本書第十二章即有清楚推導，說明為何當教

育年限作為一個連續變項，事實上與一系列虛擬變項互為「嵌套模型」(nested models)：亦即前者為後者加上教育對依變項影響為線性的限制條件。因此在連續變項上加入虛擬變項，會由於「完全多重共線性」導致迴歸方程式無解，技術上不可行；同時線性關係的限制條件，恰好為理論爭議之所在，可以透過比較兩個嵌套模型的配適程度，以資料檢驗這個看似純粹理論的爭論。這個簡單卻實用的推導，在一般統計教科書上並不常見，但突顯了量化社會研究不流於理論流派的口舌之爭，而以經驗證據做客觀且直接交鋒的優點。同樣的，第十一章看似討論三種數學函數，但這三類函數卻是以統計模型表達社會科學理論極具彈性並且好用的工具。讀者較為熟知的約莫是多項式迴歸(polynomial regression)和階躍函數迴歸(step function regression)，而樣條函數迴歸(spline function regression)則是晚近實徵研究逐漸興起的新趨勢。

謝宇的《迴歸分析》不但對於檢驗社會學理論的資料分析，提供讀者如前段所舉例子的指引，也同時釐清了許多研究者對基礎統計分析的誤解。譬如所謂的「多重共線性(multicollinearity)問題」，本書以專章處理，是極明智的抉擇。很多研究者只知道多重共線性會產生問題，但並不瞭解問題的本質，因此我曾經看到審稿意見宣稱作者的統計分析有多重共線性的問題，卻未必知道如果作者得到的結果已經達到統計顯著性，妥善處理多元共線性會讓標準誤變小，提高統計檢定力，將會增強支持相同結論經驗證據的強度，而不會導致實質結論的逆轉。因此這個批評不應該作為反對該文發表的理由，而是支持該文的發表。其他類似例子諸如：加入不相關的獨變項或者忽略有關的獨變項所造成的影響（第八章），以及是否可以刪去交互作用項中的低次（主要效果）項（第十三章），都是常見於實徵研究中，由於對基本統計問題的誤解，所造成應用上的不當。

此外，謝宇的《迴歸分析》也涵蓋了相當篇幅進階的課題：第九章介紹了「因果推斷和路徑分析」，由極其根本而重要的科學方法論開始鋪陳，並由時下流行的「反事實」(counterfactual)觀點切入，接著

則避開這塊文獻晚近常見與「傾向分數」(propensity score)相關的分析方法，而回到更傳統卻也與迴歸分析扣連更緊密的「路徑分析」(path analysis)。有些讀者也許會對這樣的編排感到訝異，我卻認為這樣的編排不但有助於破除大家對一些特殊統計方法的迷思（如認為傾向分數分析方法就能夠認定因果關係），同時討論了一些迴歸分析裡常見的重要概念，如直接效果與間接效果的區別以及所謂「模型識別」(model identification)的問題，更重要的是這段討論指出了量化社會研究一個核心命題：「理論的一個主要目標就在於提供一個模型能使一些外生變量(exogenous variable)變成內生變量(endogenous variable)」(第 182 頁)。而對量化研究抱持負面看法的學者，特別有關量化研究忽略個體間異質性(heterogeneity)的批評，如果詳細推敲這一章，將發現量化研究背後的邏輯與質化研究背後的邏輯其實有相當高的雷同。這一章也呼應了謝宇在第一章所強調「社會現象的異質性是〔無論質化量化的〕社會科學研究中面臨的最大難題」，量化研究比「思辨、內省、個人體驗、觀察和直覺等」更能妥善地「反映社會現象的異質性」(第 2 頁)。但無論讀者是否同意謝宇的宣稱，我認為質化與量化取徑真正大異其趣之處，唯有在某些質化社會研究擺明不關心因果關係（例如古典的「深描」(thick description)取徑）。只要描述社會現象的目的是檢驗帶有因果關係命題的社會理論，則無論資料如何取得（或者以何種方式分析），其認識論的最最根本之道仍萬變不離其宗。第十五章介紹的縱貫數據(longitudinal data)的分析和第十六章介紹的多層線性模型(multilevel models)都是進階、但在頂尖期刊論文中很常見的分析方法。儘管過去沒有接觸這些方法的讀者也許需要輔以其他的工具書才能夠順暢地在實際研究中運用這些方法，但以本書對核心觀念介紹之簡潔、正確、清楚，坊間（包含翻譯的）著作（在內）可說無出其右。特別在國內長期追蹤資料庫蓬勃發展，資料分析與研究成果的發表趕不上資料蒐集的進度而需人孔急之際，讓分析複雜資料的入門者閱讀謝宇的《迴歸分析》不失為絕佳的速成訓練方式。

目前謝宇的《迴歸分析》只有簡體字版，而書中許多專業術語的中文翻譯也沿用中國的法，與台灣習慣用法不同。例如 ANOVA 譯為「方差分析」，而不是「變異數分析」；normal distribution 譯為「正態分佈」，而不是「常態分配」等。但繁簡字體之間的辨識問題，以及專業術語的翻譯，相信不致影響到大多數讀者對於本書內容的理解。不過倘若容我稍稍吹毛求疵，以謝宇的科學史功力之深厚，也許本書的參考文獻可以再補強，給讀者更多第一手或經典的出處，譬如第十章除了兩本對岸出版的統計教科書之外，讀者大概也會感興趣知道 Arthur Goldberger 對多重共線性問題的貢獻。

由於相當程度地以數學（含微積分與線性代數）作為表達的工具，且涵蓋了相當篇幅的進階統計方法，謝宇的《迴歸分析》乍看之下確實會嚇跑不少原本就對數學與統計心存畏懼的讀者，但如果讀者克服這層障礙，靜下心來咀嚼本書的每個章節，就會逐漸體會量化社會研究的內在理路，進而掌握量化社會研究之鑰。即使讀者讀畢全書，真正理解迴歸分析的內涵，卻仍舊沒有被謝宇說服：「儘管帶有自身的缺陷、侷限和不完善，定量方法依然是理解社會及其變遷的最佳途徑」（第 1 至 2 頁），相信持懷疑甚至否定態度的讀者對量化社會研究的批評也會更能夠一劍封喉，而非僅止於隔靴搔癢，徒然劈落一地的稻草。

誌謝：感謝國科會專題研究計畫(NSC99-2410-H-001-081-MY2)以及傑出人才發展基金會的部份經費補助；並感謝張晉芬、謝麗玲、饒雨涵、何思瑩、張怡娟與陳也禎閱讀初稿，提供修改意見。

參考文獻

- 葉啓政(1978)因徑分析。見楊國樞、文崇一、吳聰賢、李亦園編，社會及行爲科學研究法，頁 859-905。台北：東華。
- (2001)均值人與離散人的觀念巴貝塔：統計社會學的兩個概念基石。台灣社會學 1: 1-63。
- (2005)觀念巴貝塔：當代社會學的迷思。台北：群學。
- 瞿海源(2009)最後的一些疑問。人文與社會科學簡訊 10(3): 3。
- Duncan, Otis Dudley (1975) *Introduction to Structural Equation Models*. New York: Academic Press.
- Tam, Tony (2005) Crushing the Berlin Wall of Sociology: A Review Essay on the Babel Tower of Ideas. *Taiwanese Sociology* 9: 215-240.
- Wu, Lawrence L. (2000) Some Comments on "Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect." *Sociological Methods and Research* 29(1): 41-64.