

校務研究與數據分析

■ 文／林靜慧·國立中山大學師資培育中心約聘助理教授
吳國鳳·國立臺北護理健康大學護理學院護理助產及婦女健康系助理教授
林怡慧·國立中興大學校務發展中心博士後研究員
林松柏·國立暨南國際大學教育政策與行政學系副教授

臺灣校務研究專業協會於每年一月中旬舉辦會員大會暨研討會，邀請國內外校務研究專業人員及專家學者共襄盛舉，協助各校發揮校務研究的實質效益，形塑以證據為本的校務治理文化，並建立校務研究專業發展的社群網絡。當天論壇場次與會的講者就各自基於所擅長的專業領域結合校務研究議題，並因應資料結構的特性，應用了不同的分析方法，包含了存活分析（Survival Analysis）、資料包絡分析法（Data Envelopment Analysis, DEA）、地理資訊系統（Geographic Information System, GIS）及文字探勘（Text Mining）。此篇文章則重點摘錄當天論壇交流的精華內容，提供不同且多元的角度思考議題分析的可能性，俾利學術研究人員及實務工作者能有效掌握資料分析方法。

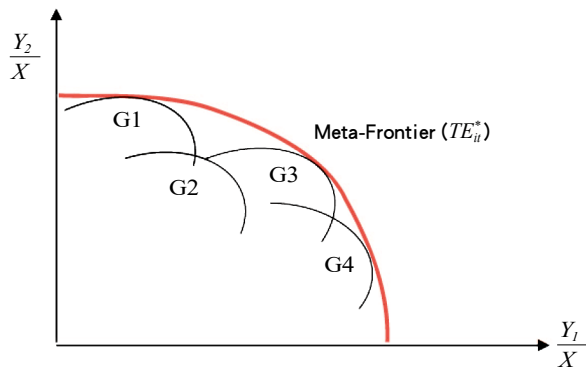
事件歷史分析法／存活分析

校務研究部門工作大致可以區分成兩大面向，亦即日常性的資料匯報及分析（包含回應各單位的資料需求）及學術議題研究。由於校務研究問題的多元與複雜性，往往需運用跨領域知能與多元資料進行處理，同時應用較為複雜與細緻的統

計分析，方能避免犯了相關等於因果的謬誤，進而得到有意義的資訊，提供校務規劃與改進策略的依據。而生物醫學領域最常被使用到的存活分析則是最典型被應用在美國高等教育領域及校務研究議題。存活分析又稱為事件歷史分析法（Event History Analysis, EHA）或簡稱事件史分析，主要在找出造成事件成因的方法。事件的發生是研究者所關注的焦點，因此對於事件發生的比率評估則是風險率，是這個方法的核心，人口統計中的生命表就是事件史的一種表示型態。這個方法最大的特色就是要解決傳統迴歸上的設限資料問題，也就是說針對在研究所觀察的個體在觀察的時期內並沒有發生變化，但因其他原因而退出觀察，或直到觀察結束時仍未發生變化，但因研究停止而被中斷的個體狀態（Singer & Willett, 1993, 2003）。因此，使用此方法需要有明確的研究**起始時間**及**終結時間**，每一個個體的起始及終結時間可能很不相同也不必相同。

學生學習經驗與政策評估

存活分析最常應用在學生學習經驗及政策評估議題，例如大學教授從事教學研究生涯議題、學



圖一 資料包絡分析雙產出單投入共同邊界圖

生休退學、助學金政策等等，研究需考量是否有隨時間化變的解釋變項，且需以縱貫性資料為主，因為若僅用一個獨立時間的資料，則無法分析時間所帶來的影響。由於過往研究多半仍是停留在描述性、或是半結構式訪談為主，無法以數據實徵資料做全面性的嚴謹分析，因此，存活分析將時間變數考量在延畢分析模式中，以學期為單位，對學生從入學開始的時間原點直至事件（延畢）發生的時間分布做研究，將存活時間列入分析，了解在某個時間點之下會發生事件的風險比，同時此種方法的精準程度能彌補在一般線性迴歸及羅吉斯迴歸（Logistic Regression）分析的不足。

於此，研究者探究大學生延畢現象（四年為正常畢業年限），資料對象為南部某研究型大學，時間範圍自2006-2007學年度入學的大學生直至2012-2013學年度，資料面向包含了學生的學籍基本資料、學業準備度、大學學習經驗，包含學習投入、社會投入及學校相關措施與政策，如出國交換及就學貸款等等。分析發現學習動機強且目標明確的大學生（如雙主修），順利畢業的情形相較於只是修輔系的學生有顯著差異。值得注意的是，政府的助學金補助對於大學生能否順利畢業並無顯著差異，然而就學貸

款對學生有其顯著差異。此外，延畢的學生人數因所就讀的科系不同而有所差異，細節討論與分析可進一步參考相關文章（Lin, Chen, & Borden, 2020）。

資料包絡分析法

資料包絡分析最初的目的是為了進行非營利組織的效率評估（Charnes, Cooper, and Rhodes, 1978），後來被廣泛應用到各生產事業及公私部門組織上，例如醫院經營效率分析、政府公部門營運效率分析等。資料包絡分析為非參數估計，無樣本數上之限制，其原理係透過數學規劃的方式，透過投入導向與產出導向，將所有受評單位（Decision Making Unit, DMU；DEA常以DMU 代替各受評單位）估算出一條共同的效率邊界，透過各DMU至效率邊界的距離，探討績效達成是否有努力的空間。

1. 考量成本效用概念進行評估分析

此方法對於校務研究中的各項議題分析納入成本概念進行評估，進而說明學校經營效率是相當適切可行的，校務研究人員首先可找出同質性高之大專校院，接著透過資料包絡分析法進行各項效率值計算，形成一共同邊界（如圖一所示），最後找出同屬性下之學校相對效率值，進一步說明性質相近之學校經營效率。

舉例來說，研究者可找出屬性相近的學校4間，分別為，G1、G2、G3、G4四間學校，以學生人數X為投入項、總休學人數Y1、總退學人數Y2為產出項，進行資料包絡分析，從下圖中可知，紅色部分為透過投入、產出項所估算出來之共同邊界，G1、G3、G4均與邊界相切點，其效率值=1，表示G1、G3、G4三間學校透過資料包絡分析發現，其學生人數與休退學比率經計算是符合效率的。而G2部分，其效率值小於1，在同屬

性學校中屬於效率值較差之學校。若學校屬於G1、G3、G4則可持續維持學生人數、休、退學人數之比率。G2學校則須要調降休退學人數，才可提高學校經營效率值。

2. 資料包絡分析與分項效率值之應用

資料包絡分析法除單年期資料分析外，還可以進一步採用多年期資料進行動態資料包絡分析，例如納入近幾年來學校招生人數、休、退學人數及一穩定變項作為連結變數進行多年期學校經營效率成效分析，透過分析結果可了解多年來在同領域學校間，貴校的經營狀況。一般資料包絡分析結果，所呈現的結果為總體效率值，無法分辨出造成總體效率值差的原因，若想進一步了解其他原因，可以採用Hu與Chang（2016）提出分項效率值之概念（計算方式如下列公式），藉由目標值與實際值的比例計算出分項投入及產出的效

$$0 \leq \frac{\text{目標非意欲產出}}{\text{實際非意欲產出}} \leq 1 \dots \dots \dots \text{Equation (1)}$$

率值，可以進一步了解各分項投入或產出對於決策單位之重要性，了解造成效率值差的原因來自於哪個變項。

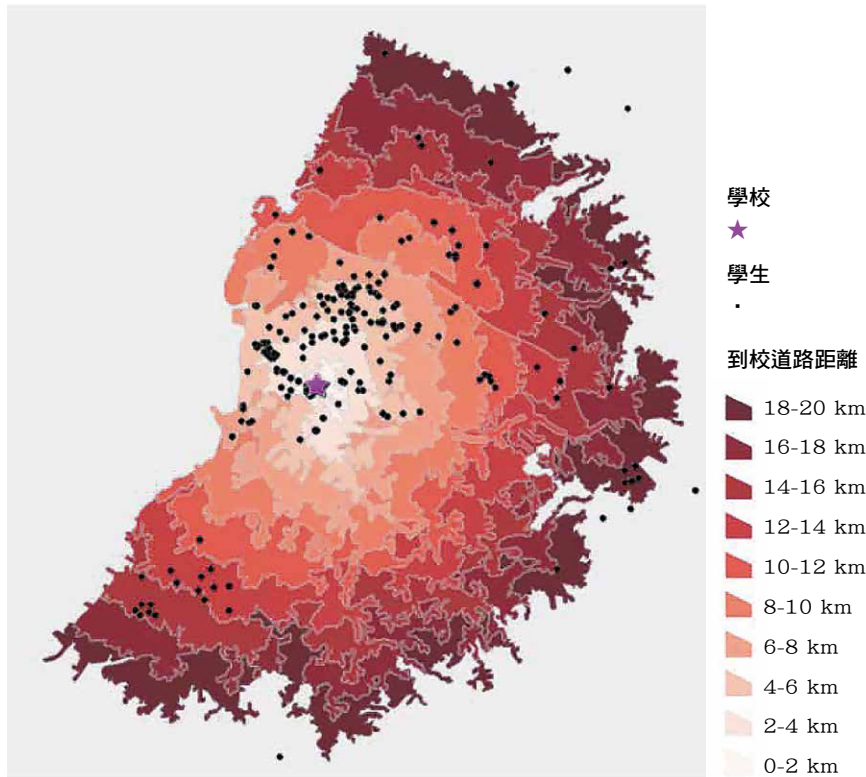
透過分項效率值的呈現，可了解G1、G2、G3、G4學校各項變數實際表現狀況與調整方向。例如要增加招生人數進而改善招生策略，或是透過校內各項輔導機制降低休退學人數。到底要招生到幾位學生才達到有效率？降低休、退學人數多少人數才達到有效率？透過資料包絡分析之結果可精確地告訴校務研究人員學校經營要達到有效率，各項變數需調整至多少之精確值。未來校務研究在後高教深耕時期，亦可應用至高教深耕計畫共同績效指標上計算出同屬性學校，透過教育部補助經費下，各項指標成效達成效率值。

地理資訊系統與空間分析

校務研究分析普遍採用統計方法與圖表來呈現資料及分析結果，然而對於具有空間特質的校務研究資料，需透過空間分析技術與方法以更有效的呈現其空間分布與關係（Jardine & Teodorescu, 2003）。空間分析包含「空間資料」（例如：學生居住地、學生實習與就業地點等）的收集與使用，以及透過地圖的呈現探究資料的「空間分布」（如：群聚或隨機分散）與「空間關係」（如：距離關係）。

早期透過在圖紙上繪製空間資料並進行分析的方式有其限制，隨著電腦科技的發展，地理資訊系統的開發強化了空間資料的儲存、處理、分析與呈現，尤其是在個人電腦的發展與普及下，GIS相關軟體與功能更加著重在能讓更多使用者容易學習與操作。此外，透過全球定位系統與遙感探測，再加上逐步釋出的各種公開資訊（如：人口與教育統計）與相關空間資料（如：行政區界、各級學校位置），空間資訊的收集更加多元。科技的進步與資料便利性的提升，使得空間資訊的運用與分析得以更廣泛的運用在不同的領域議題上。

空間分析在校務研究上可用來協助了解學生來源的分布情形（如：Granados, 2003）。依據不同的分析目的，資料可依不同層級的空間單元進行分析，當分析國際學生來源分布時，國家為適當的空間單元，來自各個國家的學生經統計後於地圖上呈現。就區域型的學校而言，較小的空間單元如縣市、鄉鎮市區、來源高中等則更能夠顯示區域內學生來源的細微分布差異，或將學生地址透過地理編碼繪製成GIS圖層進行更細緻的呈現與分析，包含使用熱點分析呈現高密度群聚的區域。透過呈現學生來源的空間分布，顯示需加強招生的區域做為策略規



圖二 空間分析圖

劃的參考。

此外，分析人員可藉由圖層套疊探討資料間的空間關係，例如，套疊學生分布情形與學校位置，並進一步分析其間的距離關係，例如透過區域分析（Buffer analysis）找出離學校特定距離內的學生，以便呈現並探討在地學生分布情形與佔比（如：美國加州大學績效問責報告）。相似的分析也可以應用於探討學生通勤距離與學習歷程的關聯性，例如透過網絡分析（Network analysis）以道路距離（或行駛時間）計算到校距離，或繪製道路距離的範圍，如圖二，將學生依所在範圍分類，藉以探討通勤距離與修課表現及校內活動參與等的關聯性。

文字探勘

校務研究有助大學治理採取數據導向的證據型

決策思維，數據資料庫中的結構化資料適合資料採礦方法，而針對非結構化文字資料進行自然語言處理的方法即為文字探勘。現代的網路世代學生習慣透過網路社群平台分享意見與表達情緒，由此所形成的輿情傳播相當廣泛、快速與多變。校務研究若能妥善運用分析網路社群的文字資料，將有助於了解學生對校園問題所反應的即時真實性意見，為學校領導者提供更「接地氣」的決策智慧。

文字探勘立基於資料科學的發展，當前資料科學發展具有巨量資料的儲存、演

算法的運行、機械學習的運用，以及更快的處理器等特徵。資料科學發展歷經三個階段：著重建置資料庫蒐集大量數據資料的「布拉赫」（Brahe），運用資料採礦分析數據再轉換為資訊的「克卜勒」（Kepler），基於前兩個階段產生知識與決策智慧的「牛頓」（Newton）。三個發展階段亦呼應DIKW（Data, Information, Knowledge, Wisdom）模式，從蒐集事實資料、形塑組織資訊、分析整合知識，進而產出決定與行動的情境智慧。當處理巨量的文本資料，較難仰賴人工判斷標記的編碼與判讀，文字探勘則是以演算法辨識文件的特徵表現，並不採人工判斷，所以能有較佳的一致性與可驗證性。

學生輿論議題分析

文字探勘的分析流程包括擷取文字、界定分

詞、定義停用詞、建立文字矩陣、計算詞頻，以及塑模分析等六個步驟。針對臺灣學生使用習慣，擷取輿論文字可以從FB、Dcard或PTT等平台，不同平台所使用的資料擷取方式亦大不相同。識別字詞可以利用中研院輿情分析系統、社群聆聽平台eyeSocial或OpView社群口碑資料庫等，但若是校務研究者要針對學校進行文字探勘，則可以自行建構詞典內容，採取字典法能依不同需求新增或删除字詞內容，也較有彈性因應校務研究者的需求。分析軟體則可採用R語言的相關套件，如Rwordseg或tm等。

文字探勘分析資料內容為文字，字句之間均有關聯性，無法如同資料採礦以隨機方法將內容分為訓練組與測試組進行模式驗證。文字探勘驗證除了邀請相關領域學者專家進行評估外，可隨機抽取部分文字內容，再分別比較人為識別與分析結果的一致性。具體而言，驗證流程可分為三階

段，首先確定以人為識別方法對隨機抽取文本內容分析結果是否有一致性，接著則撰寫電腦文字探勘分析程序對同樣的分析文本進行識別，最後則比較人為與電腦識別在分析文本的結果是否相符。因此，校務研究人員若能善用文字探勘分析學生輿情，將可避免落入決策者的主觀偏誤及研究對象的社會期望偏誤，提供即時正確的決策參考資訊。

未來展望

校務研究係屬應用型實徵研究，除了配合大學院校行政運作管理之資料匯報及量化描述分析外，考量研究議題的多元複雜及資料型態的差異，需要不同學門的學理作為研究理論的基礎，選取適切的資料分析方法，形塑有意義且具學術性的高等教育實徵研究，更有助於提升決策品質及大學治理的有效運作（彭森明，2018；黃榮村、周懷樸、林世昌，2018）。

◎參考文獻

- 彭森明（2018）。**高等教育校務研究的理念與應用**。臺北：智勝出版社。
- 黃榮村、周懷樸、林世昌（2018）。**臺灣校務研究理論與技術**。臺北：高等教育。
- Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of DMUs. *European Journal of Operational Research*, 2, 429-444.
- Granados, M. (2003). Mapping data on enrolled students. *New Directions for Institutional Research*, 120, 23-35. Retrieved from <https://doi.org/10.1002/ir.90>
- Hu, J. L., and Chang, T. P. (2016). Total-Factor Energy Efficiency and Its Extensions: Introduction, Computation, and Application. In J. Zhu (Ed.), *Data Envelopment Analysis: A Handbook of Empirical Studies and Applications* (pp.45-69). New York, NY: Springer.
- Jardine, D. D., & Teodorescu, D. (2003). An Introduction of GIS: Concepts, Tools, Data Sources, and Types of Analysis. *New Directions for Institutional Research*, 120, 5-13.
- Lin, C. H., Chen, J. H., and Borden, V. (2020). Making graduation matter: initial evidence and implications for policy and research. *Studies in Higher Education*. London, England: SRHE.
- Singer, J. D., and Willett, J. B. (1993). It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics*, 18(2), 155-95. doi:10.3102/10769986018002155.
- Singer, J. D., and Willett, J. B. (2003). *Applied longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York, NY: Oxford University Press.