

資訊技術在統計調查之應用與發展

世界各國多將資訊與通訊技術應用在辦理統計調查之範疇，本文著重國內外統計調查應用資訊技術之情形、效益評估與衍生問題及其解決方向，有助於加速運用資訊科技，以提升辦理統計調查之效能。

◎劉天賜（行政院主計處第四局專門委員）

壹、前言

統計調查工作需運用大量經費與人力，多年來各國均戮力研發運用電腦科技，以節省成本，提升效益。由於我國近年來面臨調查環境變遷，民眾配合意願降低，更需加速運用資訊科技，以提升辦理統計調查之效能。資訊與通訊技術（Information and Communication Technique，ICT），係電腦結合通訊之科技，世界各國有很多已將其應用在辦理統計調查之範疇，本文探討重點包括：國內外統計調查應用資訊技術之情形、效益評估與衍生問題及其解決方向。

貳、國外統計調查應用資訊技術之情形

環視世界各國，在辦理統計調查過程中，舉凡資料蒐集、問卷設計、行政管理、普查地圖建構、資料輸入、註號、插補、製表、供應服務、保密等，應用資訊技術甚為廣泛，有些已經行之多年，雖然研發初期面臨諸多問題，惟多經不斷的測試改進，最後才有不錯之成效。由於其研發成果頗為豐碩及多元，本文僅摘述其在人工智慧系統、光學字元辨識系統、地理資訊系統、電腦輔助調查（包含面訪、電話、網路）、統計調查個別資料庫建置等應用及其發展如下：

一、人工智慧系統（Artificial Intelligence，AI）

人工智慧包含兩個學派，第一個學派係指符號系統（Symbolic System），此系統是將專家之智慧以符號方式存於知識庫內，將功能運算存於推理機（Inference Engine），並將之符號化；另一學派為類神經網路（Artificial Neural Networks），即將電腦視為一種模擬大腦神經組織之媒介，藉模擬腦神經細胞組織活動以得到人工智慧，具有平行分散式處理、學習、回憶、歸納、推廣能力，但無法模仿人腦之想像、創造力、自主性與外界溝通能力、情感因素等。目前國外將人工智慧應用在統計調查方面，除預測模型建立、字元辨識、插補技術、資料檢誤外，尚包含行職業、教育程度、及其它分類等自動註號，惟註號功能困難在於比對判斷準則與技術，所以一般多只是停留在電腦輔助註號（Computer Assisted Coding，CAD）上，除少數國家發展運用自動註號（Automatic Code，AC），如英國研發 CLAMOUR（Classification Modeling And Utilities Research And Development）即是一例，美國、加拿大、澳洲、紐西蘭亦有相關之測試實例可供參考。

註解：

二、光學字元辨識系統（Optical Character Recognition，OCR）

此系統基本應用類神經網路之特性，對於不同輸入字元予以辨識，目前有超過二十餘國開發 OCR 系統，均有相當效果之辨識率，因具開放式架構，可在多種作業平台環境下發展，其作業包括掃描、辨識、校登及影像管理，此系統之開發具縮短資料處理時程，減少人力僱用成本，問卷易於管理且降低儲存空間等優點，惟目前國內對於提高中文字元辨識率方面仍有待努力。

三、地理資訊系統（Geographic Information System，GIS）

雖然地理資訊系統之發展受限於地理圖檔之完備情形，但在聯合國推展下已有五十國以上進行不等程度之運用，如美國 MAF/TIGER（Master Address File/Topologically Integrated Geographic Encoding and Referencing）系統、英國 GAPS 系統、日本 CMS（Census Mapping System）系統。美國普查局擬於 2010 年人口普查運用 MAF/TIGER 結合衛星定位系統（Global Positioning System，GPS），使調查員可更有效追蹤到未回答樣本之重複地址，以節省調查作業成本。

四、電腦輔助面訪調查（Computer Assisted Personal Interviewing，CAPI）

CAPI 使用設備，包含筆記型電腦（Laptop）及掌上型電腦（Palmtop PDA），一般都使用在問項相對較多之調查。本系統具有自動導引問項順序、線上檢誤及傳輸、匿名、特性、減少鍵入錯誤等特性，若再加上軟體設計得宜，應可減少整體調查時間。美國除研究於雙端密碼資料加密及電話傳輸指定功能（僅允許以指定之後端電話線執行傳輸，並由後端付費），俾免資料外洩及傳輸資源之濫用情形發生外，並評估以無線傳輸訪查資料之可行性，決定其是否應用於 2010 年人口普查。

五、電腦輔助電話調查（Computer Assisted Telephone Interviewing，CATI）

目前各國使用 CATI 系統調查相當普遍，惟使用在問項較簡易之調查，新加坡 2000 年人口普查更首創網際網路、電腦輔助電話訪問、面訪三階段之普查方式，並連結整合公務登記資料，有效減少訪查費用，提升資料品質。各國並已研發成功於 CATI 系統增加電話按鍵式輸入（touch-tone data entry）及語音辨識（voice recognition）等功能。

六、網路調查（Web survey 或 e-Survey）

網路調查包含下列兩種方式：（一）、以電子郵件方式將問卷傳至受訪者，或將問項置於網路中供受訪者自行填答後傳回，優點在於快速，惟因其完全由網路建立之底冊抽樣、填答問卷或採志願性樣本，此種方式不符合機率原理；（二）、底冊來源經由其他方式取得，抽樣方式完全依照機率原理，再透過網路填報，新加坡有 25 項調查採取調查回答引導專家系統（Survey Answering Guide Expert，SAGE）即為一例，新加坡統計局於 2000 年 3 月 1 日由商業預測調查起，陸續針對 12 項調查建置於同一 e-Survey 網頁中，透過網路方式填報，提供快速及簡便之回卷管道，可以一天 24 小時回傳調查資料，透過 SSL（Secure Socket Layer）資料安全傳輸

協定，資料傳輸經過加密保護，禁止未經授權之存取，受訪者利用 e-Survey 回答之比率在 5 至 20%之間。e-Survey 僅為 SAGE 蒐集資料方法之一，受查者亦可以選擇其他方式(使用 e-mail、信件、傳真)回覆調查問卷。SAGE 包括使用者系統操作、多種調查回覆模組、自動化資料轉換、初步資料線上檢誤、整合調查追縱機制。

七、統計調查個別資料庫 (micro-database)

美國、新加坡、荷蘭、芬蘭、瑞典及北歐各國，鼎力發展電腦與統計調查結合之技術，並建置包含企業、住戶、住宅等多項統計調查個別資料庫，運用資料庫存取與倉儲觀念及技術，連結公務登記及統計調查資料，透過網際網路、電子郵件及電話調查系統等途徑，定期蒐集更新資料庫，並經由插補技術，建立合成資料庫 (synthetic database)，除作為抽樣母體應用外，並可產生相關統計結果，有效降低受訪者負擔，顯著提升統計效益。

參、我國資訊技術在統計調查之應用情形

對照世界各國將資訊技術應用在統計調查之蓬勃情形，近年我國年在這一方面亦有了多項研發及精進，茲將我國實施情形摘述如下：

一、OCR 之應用

我國應用 OCR 技術始於 1980 年代工研院所研發系統，目前對於英數字之辨識率超過 95% 以上，其核心辨識係結合統計式與結構式文字特徵，再由高效率特徵轉換模組降低特徵維次，縮小儲存空間，提高辨識效率，最後以類神經網路與字元型態資料庫進行文字辨識。其比對參用之字元型態資料庫，係經長期蒐集累積國人手寫文字型態，並經辨識模組適當學習而建立，未能辨識之字元轉由人工校登，技術上採行較佳之字元、欄位、整表等三階段校登方式，並先經資料庫技術排序處理後顯示，以避免畫面紊亂造成人員誤判，最後仍須進行欄位或整表校登之情形已非常稀少。我國並於九十二年開發完成通用性表單系統，將推廣至其他調查，預期可節省資料處理時間與人工成本。

二、擴展普查 GIS 之建置

普查是一種空間涵蓋面完整之調查，其特點在於可以提供小地區與細分類之統計，惟調查之個體資料 (micro-data) 依法應予保密，因此適當之分類機制益顯重要。歷次普查以鄰為普查區劃分基礎，係僅為分配工作量，而以村里作為普查最小統計地理單元，由於里鄰易隨社經環境變動調整，且人口數量與特徵值之差異大，因而減損普查資料整合及時間數列運用。若能藉由處理空間資料佳之 GIS 技術結合數化地理圖檔，以明顯地物及自然特徵劃分普查區，將可建立範圍完整、適當、穩定且可長遠應用之普查地理單元，供為普查責任區及普查資料整合之用。近年來隨內政部長期推動「國土資訊系統」之成果，已取得台北及高雄兩直轄市地形圖及地址位置等數值圖檔，經適當之 GIS 及資料庫技術處理後，建立普查區、指導區圖層及其相關屬性統計研究，供實地訪查及資料整合連結應用，現正推動至其他縣市地理資訊系統之建

置。

三、推廣 e-Survey

鑒於各事業單位應用網際網路情形日趨普遍，且「受雇員工薪資調查」受查廠商，迭有反映郵寄填表之不便，或傳真表件常遇線路不暢、資料不清晰或缺漏情形。遂於八十八年九月建置完成網路線上填報系統，以人性化操作介面，提供另一填報管道，該系統具資料保密、線上檢誤、縣市直接鍵入資料之功能，可確保資料安全及調查品質，提升審核工作之效率，目前已推廣至 1600 家廠商。另九十年「工商及服務業普查」、經濟部之「外銷訂單調查」亦提供受查廠商採網路填報方式辦理，確能提升資料品質及時效。

四、開發 AI 系統

行政院主計處九十二年運用光學文字辨識、語辭斷句、索引分類等技術，初步開發完成行職業代碼自動註號之人工智慧辨識系統，以替代部分人工判讀作業，未來經測試改進後，逐步推廣至各項統計調查應用，達成資源共享目的。

五、推動 CAPI 系統

行政院主計處刻正推動電腦輔助面訪調查，自九十二年七月起至九十六年十二月止，採循序漸進方式，分短、中、長程辦理，先期將「人力資源調查」納入試辦，並檢討其辦理成效後，再推廣至其他家戶面調查。由於電腦輔助面訪調查具有線上檢誤、自動導引問項順序、自動選擇正確受訪者及匿名訪問等特質，可有效降低人工作業成本，提升調查資料品質。另物價調查並自九十三年三月起全面應用 PDA 進行查價工作。

肆、應用資訊技術之效益評估

一、普查作業

我國八十九年戶口及住宅普查採用 OCR 作業，資料處理費時約五個月，每個月運用人力 110 人，分別較 1990 年普查之八個月，每個月運用人力 180 人，無論在時效或表件調閱人力等方面均大幅縮減，免除列印錯誤報表、節省存放空間，並具線上檢核、中文地址辨識光碟影像管理、標準化等特性及功能。八十九年農林漁牧業普查、九十年工商及服務業普查、九十二年攤販調查續應用所研發之 OCR 系統，資源充分利用，並擷取辨識普查表影像中文地址，據以首次建立普查中文母體名冊，供為抽樣應用。另開發行政電子作業系統，使相關作業程序標準化，便於管理及減少錯誤，減輕人力負擔。

註解：

二、抽樣調查作業

受雇員工薪資調查推動網路調查後，在提升效率及品質、降低受訪者及人力負荷方面，確

具成效。另 CAPI 效益方面，因我國目前在試辦初期並無量化數據可資比較，惟國外有頗多論述探討 CAPI、PAPI 在調查品質、時效、成本之差異等方面之差異，一般 CAPI 較具優勢，惟常因調查辦理方式、規模、週期、軟體設計產生不同。

伍、衍生問題及解決方向

雖然應用資訊技術後，統計調查在提升資料品質、時效及降低成本等方面頗具成效，但也相對衍生很多問題：如資料保護、各界認知及接受度、一致性與標準化、調查制度與流程再造、人員訓練與知識管理、樣本代表性等，以下列述如下：

一、資料保密

在保護個人隱私及提供可造福人類知識之間，如何求取平衡點，為資訊道德之主要困境。由於資料需求者要求統計機構儘量提供資訊，以供制定政策及學術研究應用，另在民眾心目中，私人、行政或統計資料庫間沒有差別，再加上網路易存取、無線網路盛行之特性及資訊安全機制相對不足，民眾疑慮從未消除。高度運用資訊技術，雖可提升整體統計調查效率，但也相對增加個人隱私暴露之風險，為消除民眾之疑慮，提高其信任，應加強資料保護安全措施，如隱藏 (mask)、加總 (micro-aggregation)、干擾 (perturbation)、風險評估及防火牆等技術之研發，以防止個別資料不當洩漏，維護民眾隱私，確保受訪者權益。

二、各界認知及接受度

以 CAPI 系統試辦初期為例，由於訪查員認為攜帶電腦不方便、且擔心遺失或損害電腦，受訪戶擔心隱私受侵犯，再加上都市化地區之大樓及公寓不易進入、軟體設計等因素，相對影響實施 CAPI 之預期目標，未來宜加強宣導、針對不同特性的受訪戶採取不同調查方式，如評估採取網路或電話調查之可行性。另在軟體設計方面，宜朝更易操作，以節省整體作業時間，硬體設備之汰換採購朝更易攜帶，以臻預期目標。

三、一致性與標準化

作業過度依賴系統平台，擴充及整合不易，作業環境與技術障礙，造成互聯困擾。各作業所建立之電子檔案，在分類、中文內碼、定義、涵蓋範圍、品質、格式、建置時間等仍有諸多問題尚待克服，致使資訊互聯困難，未來宜持續推動開放標準與建構環境，增進各機關電子檔案之標準化，採共通性開放平台以利作業整合。

四、調查制度與流程再造

由於調查環境隨時在變，未來很難以同一種方式完成調查 (form will find you)，應用資訊技術後，相對影響人力調配、資料處理、行政作業等調查制度及其過程，須以流程再造 (business process redesign, BPR) 理念，整合不同調查方式於同一平台，朝整合、簡化、創新等方向予以重新設計，以達成既定目標。

五、知識管理與人員訓練

由於網路發展，促使資訊接取多元化與更具時效，各國皆將其研發理念及成果，上載於網站，推動網路學習，以利知識流通及觀念分享，增加附加價值。另隨資訊技術多元化發展且規模日趨複雜，人力素質無法跟上資訊技術之腳步，因此應積極加強人員資訊知能之訓練，以延續並創造更多的運用及成果。

六、樣本代表性問題

網路及電話訪問，雖可在短時間及合理成本內完成調查工作，但樣本代表性仍有諸多問題亟待克服。在 CATI 方面，應用隨機撥號雖可以降低樣本之偏誤，但以我國目前手機普及情況，其對於 CATI 調查樣本代表性之影響有待評估。另知識網路可採住戶電話號碼當作抽樣底冊，以處理網路調查涵蓋面誤差問題，惟其底冊來源、母體分配及樣本代表性等理論發展有待克服。

註解：

陸、結語

當世界以數位速度前進時，應採取「轉變策略」，也惟有不斷創新才有競爭力，且不應只將資訊技術視為一種高效率之工具，而需與統計調查之目標相聯結才有意義。為適應外在迅速及多元之環境變遷，未來以單一方式完成統計調查，將愈趨困難，宜研究建構多元化（面訪、通信、CATI、CAPI、網路、傳真等）調查方式體系，整合線上、非線上調查於同一平台，並在調查制度、人力配置及行政作業各方面，研擬一套更具彈性之配套措施，並經長期測試及循序漸進，以謀求克服之道。

參考文獻

1. Marek Fuchs、Mick P. Couper, Sue Ellen Hansen, Technology Effects: Do Capi or Papi Interviews Take longer?, Journal of Official Statistics. Vol. 16. No .3.2000. pp.273-286
2. Chares Louis Kincannon, U.S. Census Bureau Strategic Plan FY2004-2008, September 2003
3. Rosemary Crocker, Rob Edmondson, Computer Assisted Personal Interview Solutions in Australia, ESACP Workshop Paper, 1999
4. Warren J. Mitofsky, Internet and Innovative Data Collection, 53rd Session Proceedings, 2001
5. Stephen E. Fienberg, Discussion of Three Papers on “Disclosure Control and Facilitating Access by Users”, 53rd Session Proceedings, 2001
6. Svein Nordbotten “Neural network Imputation Applied to the Norweigan Population Census data”, Journal Of Official Statistics ,Vol.12, No.4, 1996, pp.385-401
7. Svein Nordbotten “Editing Statistical Records by Neural Networks”, Journal Of Official Statistics, Vol.11, No.4, 1995, pp.391-411
8. 羅國華、劉惠玲、周元暉，資訊技術在戶口及住宅普查上之應用與發展，主計月刊第 576 期，民國九十二年
9. 潘寧馨，我國發展電腦輔助面訪調查之研究，行政院主計處，民國九十二年
10. 劉天賜，神經網路應用於調查資料之研究，八十七年國際統計學術研討會，民國八十七年