

「AI」驅動之統計調查業務

行政院主計總處國勢普查處

壹、前言

數位時代下，技術創新及運算力加速，驅使資料數量及類型呈指數成長，為統計應用開展了前所未有的可能性。行政院主計總處（以下簡稱本總處）國勢普查處（以下簡稱本處）負責辦理人口及住宅、農林漁牧業、工業及服務業三大普查，以及就業失業、薪資與國富等重要統計及調查工作，亦順應此數位科技趨勢，近年來積極精進統計技術，整合多項調查與公務大數據資料，以持續改進（All-time Improvement）、加速創新（Accelerating Innovation）的精神，並導入人工智慧

（Artificial Intelligence），驅動統計調查業務持續優化，期在快速變動的數據環境中保持實用性（relevance），提供更高品質、更高效率、更細緻分類及更具決策價值之統計資訊。

貳、過去持續改進績效

一、編布睽違 30 年之家庭財富分配統計

隨著經濟快速發展，攸關國民經濟福祉之家庭財富分配備受各界關注，本處曾於民國 80 年辦理調查蒐集相關資料，惟內容涉及高度隱私，執行至為困難；雖可按年依家庭所得分配觀察收入不均程度，然「所

得」與「財富」意涵不盡相同，難以精確評估家庭財富存量分配情形。

考量全球主要國家均定期產製是項統計，且立法院及媒體亦迭有相關資訊需求，本處爰於 113 年排除萬難，透過公、私部門協力，克服資料定義適用性及隱私性等問題，整合本總處抽樣調查、財政部及財團法人金融聯合徵信中心等相關檔案資料，以最佳化統計方法，重現睽違 30 年之重要統計結果（下頁圖 1），做法與主要國家均以統計調查為資料蒐集渠道之方式迥異，創家庭財富分配大數據統計之先例，成本效益大幅提升。

二、銳意革新的薪資分布統計

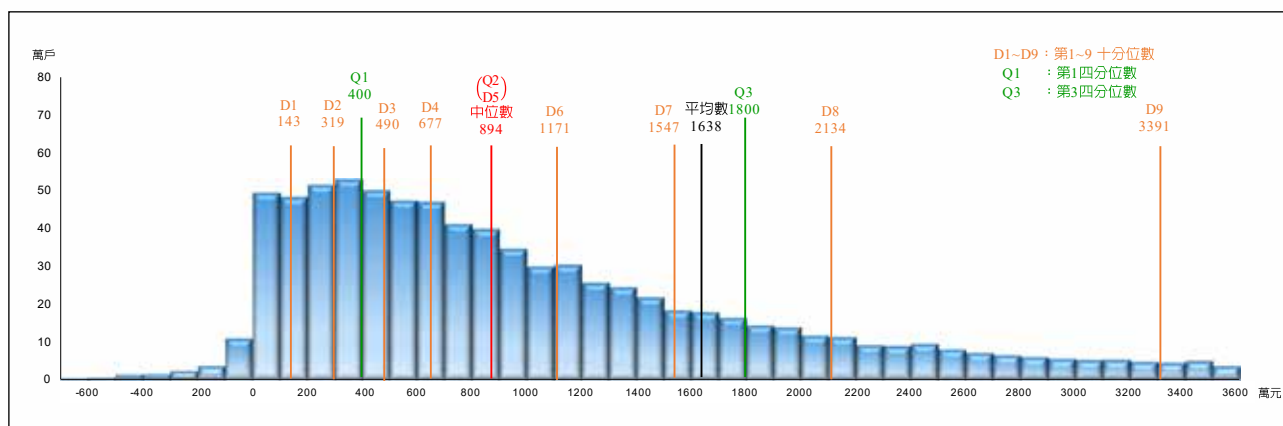
「受僱員工薪資調查」按月蒐集廠商僱用員工之薪資彙總資料，定期發布薪資平均數。惟隨個別員工薪資差異益受關注，本處自 105 年起首次整合家戶面及企業面抽樣調查編製總薪資分布統計；其後順應大數據發展趨勢，整合財稅、勞/健保¹、人口及住宅普查常住人口、工業及服務業普查分支據點等各項資料，精進編製作業，創編員工特性別及縣市別總薪資統計，多方延伸薪資資訊廣度。

113 年考量總薪資內含變動較大之非經常性薪資，且我國係以月薪為主要計薪方式，為貼近民衆感受，再次突破「勞/健保以月薪級距資料為扣（提）繳基準，且具最高、最低級距」之限制，導入線性插補（linear interpolation）法將各級距內資料差異化，並選擇以右偏（right-skewed）之 log normal 分布配適（distribution fitting），設算薪資分布左、右尾資料，建構貼近實際月薪之完整分布，業自 113 年 7 月²起按月發布經常性薪資分布統計。

三、接軌國際的就業、失業統計指標

伴隨少子化及高齡化衍生之工作年齡人口遞減情形，勞動力未被充分利用的程度日益受到關注。為充分反映勞動市場實況並接軌國際，本處依國際勞工組織（International Labour Organization, ILO）倡議，自 113 年 1 月首次編布「4 週失業率³」，未來將俟累積足量資料建立季節調整模型後，正式以其作為我國官方失業率，以利國際比較；另為呈現我國勞動力低度運用（labor underutilization,

圖 1 110 年底我國家庭財富分配



說明：左、右尾持續延伸，惟受限於版面無法顯示。

資料來源：行政院主計總處國勢普查處。

專題

LU) 情形，除前揭官方失業率（即 LU1）外，業修正「人力資源調查」問項，蒐集「工時不足之低度就業者」⁴（time-related underemployment）及「潛在勞動力」⁵（potential labor force）相關資料，並編製 LU2～LU4 指標（圖 2），自

114 年 1 月正式發布，提供各界運用。

四、智慧及數位轉型的農林漁牧業普查構想

研擬規劃普查構想係普查的重要前置工作，114 年農林漁牧業普查以「接軌國際重要

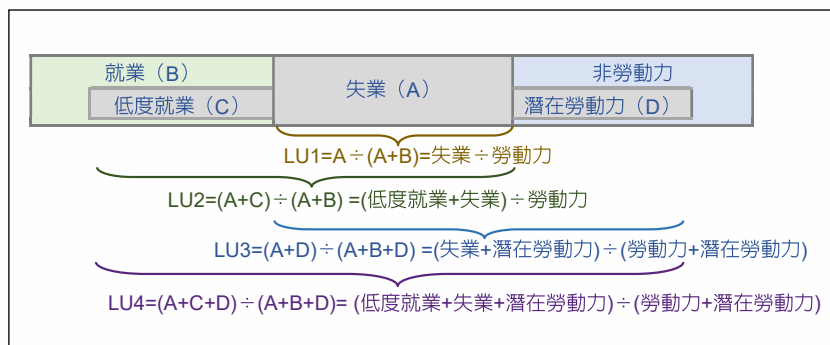
資訊」、「創建整合式母體資料庫」、「智慧及數位轉型」及「階段性訪查流程再造」四大主軸（圖 3）為核心規劃普查構想，業於 113 年擬具普查方案並經行政院核定，作為普查作業之準則。

參、未來加速創新方向

一、精益求精、多元延伸的薪資統計

「薪資」與民衆生計高度相關，係當前政府「包容成長」（inclusive growth）施政目標的重要面向，本處爰不斷推陳出新相關統計指標，以支援政策運用，114 年將進一步連結公教人員保險等相關檔案，行業範圍擴及正規教育及公共行政業；因月薪中位數及分布統計，已運用保險相關大數據產製，取代原「人力運用調查」之「就業者主要工作經常性收入」，爰予停止查填編布，惟為滿足農業部及行政院性別平等會等相關需求，仍持續運用「人力資源調查」全年權數（weights）等統計技術，試編農業從業者、自營作業者、雇

圖 2 勞動力低度運用指標



資料來源：作者自行繪製。

圖 3 114 年農林漁牧業普查規劃構想四大主軸



資料來源：作者自行繪製。

主及各性別 / 職類等大數據統計範圍並未包括之經常性薪資統計，以豐富薪資統計特性別資訊。

另為了解低薪族群背景，除定期產製特性別（年齡、性別、行業別）等橫斷面（Cross-sectional）統計結果，亦將串接歷年個別受僱者追蹤資料（Pannel Data），進行縱貫面（Longitudinal）研析，依循薪資流動（Wage Mobility）理論，探討薪資成長速度、歷時及成因，供相關改善策略規劃之參考。

二、運用常住概念，翻轉就業失業統計

就業及失業統計向為勞動市場重要指標，其資料來源之「人力資源調查」原係按戶籍人數進行抽樣及推估；惟隨跨境移動人口日增，調查推估結果與國內勞動情勢逐漸出現落差，本處爰結合各項公務登記資料，建立常住人口模型，掌握常住人口分布，業初步編算以其為基礎之就業失業統計。本年將進一步參酌美國現住人口調查（Current Population

Survey, CPS）辦理方式，以有人居住之有效地址為抽樣母體清冊，精進以常住人口為基礎之就業失業統計。

本處長期蒐集戶籍地址、村里門牌地址等資訊，地址涵蓋範圍完善，近年受惠於大數據及人工智慧技術快速發展，本處將整合房屋稅籍檔、用水（電）等大數據，導入機器學習等人工智慧工具及統計方法，判定並註記可供居住之住宅、有人居住之住宅等有效地址，建置地址母體，提供「人力資源調查」抽樣運用，從基礎翻轉就業失業統計辦理方式，期能分別從「國人」及「國內」的角度，全面完整呈現我國勞動市場實況。

三、結合財稅大數據及統計技術，擘劃工業及服務業普查新圖像

工業及服務業普查為減輕受訪廠商填報負擔，自 85 年起抽選代表性樣本填報「各業專屬調查表」（含細緻的財務問項），其餘填報「普查表」（財務問項僅須填報總金額，細項由推計產生）；惟財務問項較

為敏感，且訪查與報稅期間重疊，致廠商及訪員對普查相關規劃設計迭有抱怨，影響配合意願及資料品質，訪員招募亦面臨挑戰。

考量財政部提供資料之損益科目分類不敷普查及國民所得統計編算產值所需，115 年普查仍保留代表性樣本填報「各業專屬調查表」，「普查表」的查填則參考美國、南韓做法，大規模、特定廠商（如多場所企業）續填財務問項（總金額），其餘規模小、同質性高且業務單純者，採「抽樣」及「插補」方式取代全查（下頁圖 4）；而訪員人力來源亦規劃予以多元化，刻正評估納入記帳業者之可行性。如此一來，預計可大幅緩解查填、估算衍生之問題及普查工作負荷，有效提升普查執行效率。

四、創意無邊界、應用零極限的三大普查母體整合式資料庫

最新母體產製係普查重要功能之一，除為普查底冊基礎，亦為各機關抽樣調查之母體來源。現行三大普查母體整編來

專題

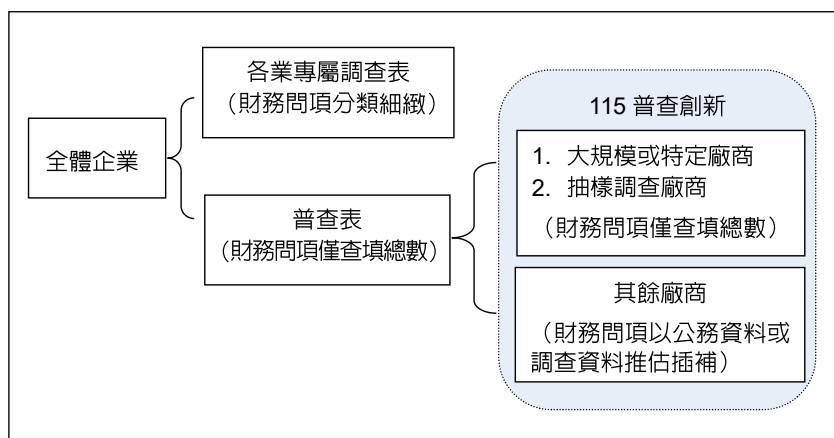
源檔案近百個，對於資料檔案傳輸、清洗、比對及插補的處理常有重複作業或格式不一等情形，且以檔案形式儲存，不易進行跨檔案之關聯查詢。

為提升整編及運用效能，

本處建置三大普查母體整合式之關聯性資料庫（圖 5），從強化資料傳輸安全、建立標準化作業、齊一資料格式、快速關聯查詢，到機敏資料去識別化及資料取用權限化，充分發

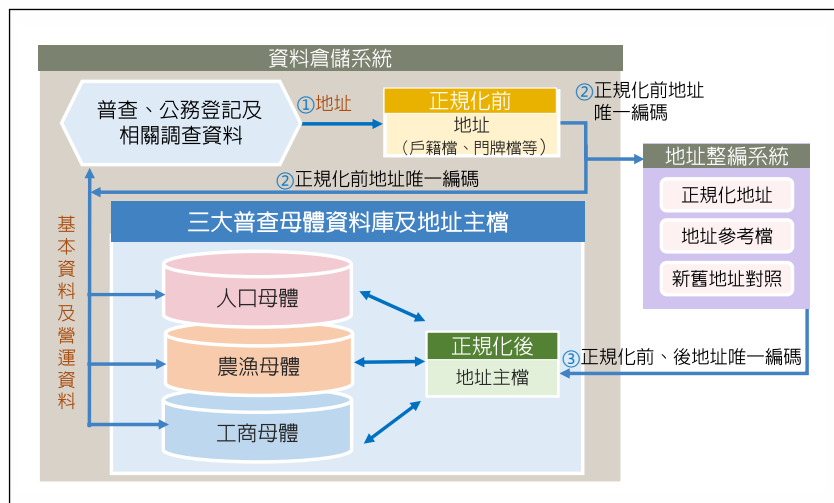
揮資料庫之優點，再造母體整編流程，有效減省作業時間及維護成本；而定期更新機制則有助於非普查年間掌握最新農、工、服務業及人口等基礎國勢重要資訊，並提供統計調查抽樣及樣本輪換應用，有效減輕受訪者填報負荷。

圖 4 115 年工業及服務業普查填表規劃



資料來源：作者自行繪製。

圖 5 三大普查母體整合資料庫



資料來源：作者自行繪製。

肆、AI 賦能，調查統計增能

隨人工智慧快速發展，聯合國官方統計大數據及數據科學專家委員會（UN Committee of Experts on Big Data and Data Science for Official Statistics, UN-CEBD）、聯合國歐洲經濟委員會（United Nations Economic Commission for Europe, UNECE）官方統計現代化高級別小組（High-Level Group for the Modernization of Official Statistics, HLG-MOS）均積極推動 AI 在統計調查領域的普及與創新應用，目前生成式 AI 在政府統計的應用包括：

一、程式碼和 IT（Information Technology）開發，例如程式碼翻譯、產生。

二、資料處理的文字產生，例如合成資料產生、行業代碼分類。

三、文本生成，例如產生統計分析報告、新聞稿等。

四、通訊影像或視訊資料生成。

五、資訊搜尋。

六、聊天機器人，例如針對調查受訪者、訪談者提供客服。

七、行政任務，例如內部電子郵件、會議紀錄、準備簡報投影片等。

本處業於統計調查規劃設計階段，透過大型語言模型（Large Language Model, LLM）蒐集、分析國內外趨勢，輔助重點規劃，精進普抽查問項設計，未來可運用機器學習演算法的聚類、降維等功能，優化樣本分層、抽樣，提高代表性及資料品質；訪問調查階段亦規劃運用智慧客服，輔助調查疑難排除，未來可進一步透過自動語言翻譯及交互式訪談培訓，提升跨語言溝通與訪談技巧；資料處理階段使用 LLM 輔助程式語言轉換（Fortran to SAS）及語法錯

誤、邏輯錯誤之修正，農林漁牧業普查亦規劃辦理農作物自動標籤與分類，提高資料處理效率；至統計分析及報告編製，將生成初步分析內容，再輔以專業人員檢視，減輕工作負荷並提升效率；另外亦整合公務大數據資料，運用機器學習（Machine Learning）與統計方法建立模型（如非普查年間常住人口模型），充實統計資訊，以及輔助母體名冊編製（如判定實際從農者、判定有居住者之有效地址），提升調查執行效率，期透過 AI 賦能，全面助益統計調查各階段業務的效能化、精準化與智慧化。

伍、結語

統計調查資訊是政府決策的基石，但其作業環節繁複，傳統模式難以快速應對變遷的環境與不斷湧現的新需求，唯有透過持續的技術變革與加速創新，並積極應用 AI 工具，調查統計才能提升效率並實現永續發展。另外，基本國勢調查是政府制定社會及產業相關政策的重要數據來源，值此農

林漁牧業普查、工業及服務業普查即將陸續展開之際，亟盼中央及地方政府無縫協作，發揮普查數據力量，驅動中央與地方施政決策，共創雙贏。

註釋

1. 包括勞工保險檔、勞退月提繳工資檔、全民健康保險檔等按月扣（提）繳薪資資料。
2. 指資料時間，以下 4 週失業率、LU2 ～ LU4 之發布亦均為資料時間。
3. 指失業者尋職期間為「含資料標準週的過去 4 週」。
4. 指因經濟因素，資料標準週工時低於 35 小時，且希望大於或等於 35 小時之就業業者。
5. 指無工作、想工作、可工作但過去 4 週未找工作，或無工作、過去 4 週有找工作但不能開始工作之非勞動力。

參考文獻

1. 聯合國統計司（United Nations Statistics Division, UNSD），Handbook on Management and Organization of National Statistical Systems
2. 聯合國歐洲經濟委員會網站，<https://unece.github.io/genAI/>❖