創新變革精進獎勵項目

以數據科學探析行政區域複雜社經特性，開創模組化學習型組織

善用數據科學方法，並與時俱進研究大數據應用，可拓展統計調查實務之視野。本文分享家庭收支調查如何有效精進抽樣設計並建立模組化學習型組織之作為。

行政院主計總處（莊科長文寬、陳專員錫慧）

壹、前言

行政院主計總處按年辦理之家庭收支調查（以下簡稱本調查）係爲了解全體家庭收入與支出狀況，供政府施政及各界研究相關議題參考。

本調查係採抽樣調查，如何精進抽樣設計以降低抽樣誤差，是統計調查業務永無止境的課題。隨著科技的發展，統計問題的解方非僅限於統計方法，尚可擴展運用大數據及資訊科學等技術，本文將介紹精進本調查抽樣設計之方法及衍生效益。

貳、以數據科學綜整複雜社經資料分類機制

本調查的抽樣設計係採分層兩段隨機抽樣，先對全部村里分層，以村里爲第一段抽樣單位，戶爲第二段抽樣單位，亦即先抽出一些樣本村里，再從中抽出樣本戶。如果抽出的樣本與母體長得像，此樣本就較具代表性；村里分層作業就是要先釐清母體的各種特徵，再據以抽樣，樣本就較能與母體長得像。在精進村里分層作業的過程中，遭遇的問題包括：

1. 如何客觀有效地將全國 7 千多個村里分層？
2. 如何迅速有效地檢驗分層結果？

一、如何客觀有效地將全國 7 千多個村里分層？

（一）方法評估

影響家庭收支水準的因素相當複雜，經考量與家庭收支調查內容較有關之社經特性及資料可取得性，選取年齡、教育程度及就業人口之產業結構等公務資料作爲分層變數。經由統計專業分析，要同時考量多個分類變數，將全國 7 千多個村里分層，須採用多變量集群分析法，較爲客觀有效，並能建立綜整行政區域複雜社經特性分類機制。

（二）分類機制－多變量集群分析法

分類機制主要目的係將性質相近的村里歸類爲一群，分類準則爲群內差異小、群間差異大。以集群分析之分類機制，須先決定層數，再依相似性進行分群。

1. 決定層數

要如何做到「群內差異小、群間差異大」，係利用變異數分析法將全體資料總變異（SSTO）拆解成組間變異（SSB）與組內變異（SSW），在分層過程中，依相關變異指標大小來判斷，其所使用綜合判定指標如下：

(1) RSQ（R-squared）：組間變異除以全體資料總變異，故該值越大越好。

(2)RMSSTD（Root-Mean-Square Standard Deviation）：將組內變異除以自由度並開更號，故該值越小越好。

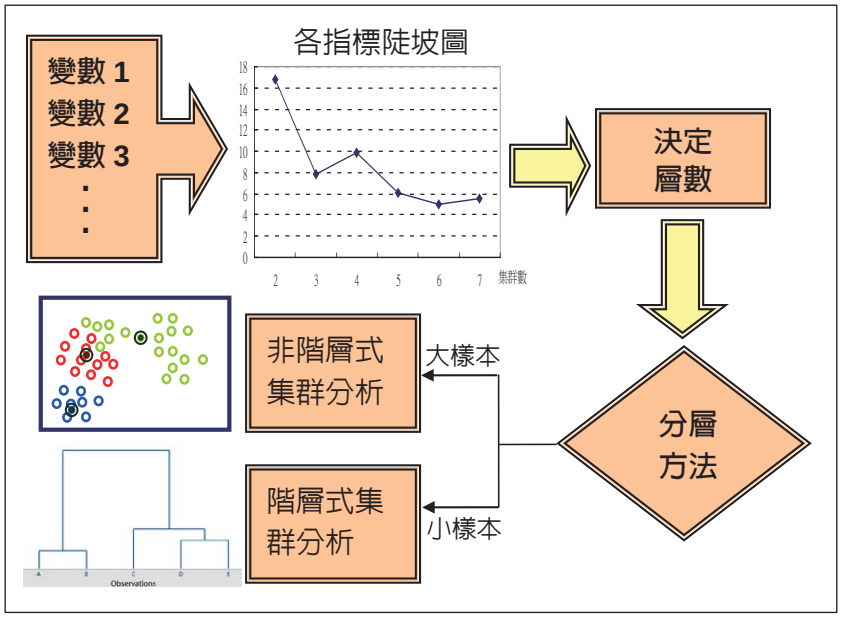
(3)PSF（Pseudo-F statistic）：組間變異與組內變異的比例，值越大越好。

採用凝聚式集群法計算出上述幾個指標並畫出陡坡圖，集群分析的目標在於使群內同質性大、群間異質性大，因此，若觀察值的合併使得上述各判定指標值突然增加或減少，表示應停止合併，據此決定較佳層數。

2. 分層方法

由於家庭收支調查非直轄市部分係以縣市爲副母體，直轄市則以行政區爲副母體，然非直轄市之村里數屬大樣本，而直轄市行政區村里數屬小樣本，因此非直轄市採非階層式集群分析法 K-mean

圖 1　分類機制



資料來源：作者自行整理繪製。

創新變革精進獎勵項目

（Non-hierachical cluster-K-mean）進行分群，直轄市則分別按行政區以凝聚式階層式集群分析法－華德法（Hierarchical cluster-Wald's method）進行分群，相關程序詳上頁圖 1。

二、如何迅速有效地檢驗分層結果？

經上述多變量集群分析完成 7 千多筆村里之分層作業，傳統上係逐筆檢視分層結果，惟此法費時且難以綜觀村里層別之地理位置分布情形。經相關專業領域同仁研討，結合國土資訊系統村里經緯度數據，並運用相關統計軟體，將 7 千多筆數據轉化爲視覺化地理圖形，可具體展現地理行政區域資訊。

（一）國土資訊系統

國土資訊系統是結合全國各種具有空間分布特性之地理資料，將國土空間的組成物件表現在圖形上，使用者可依需要將相關的圖資加以套疊，進行資料存取、處理及分析。藉由國土資訊系統開放資料之經緯度數據，可讓使用者將其與具地理空間資料，運用軟體繪製成地圖。

（二）使用軟體

目前採用 SAS 及 R 軟體進行編撰程式語言以繪製各縣市地圖，SAS（全名爲 Statistical Analysis System）係以 C 語言爲基礎的軟體，須租用，所需成本較多，其係以內建套件（語法爲 PROC GMAP）或 GUI 介面產製地圖；而 R 是具資料處理及統計分析之軟體，爲免費的公開軟體，可於官方網站（http://www.r-project.org/）搜尋繪製地圖之相關套件（表 1）。

（三）繪製地圖

將數量龐大的 7 千多筆村里分層結果資料，結合國土資訊系統村里經緯度數據，以統計軟體 R 或 SAS 自製程式繪出各縣市村里分層地圖，由於可以不同顏色標示各層（下頁圖 2），分層結果適切性一目瞭然，相較於用表格檢視較省時且有效。

表 1　SAS 與 R 軟體比較

	成本	繪製地圖套件
SAS	付費	內建
R	免費	須於網站搜尋下載

資料來源：作者自行整理。

參、建立模組化之學習型組織

在前述精進村里分層作業中，須先擬訂策略，然後就問題部分提出解決方法，其中需要各種領域專業，包括程式設計、統計理論及社經分析等。經由本項精進作業實作，各專業領域同仁透過研討及互相學習，不僅完成任務，同時也建立了模組化之學習型組織（下頁圖 3）。

個體需要學習方能進步，組織亦需具學習的機制始不致僵化枯朽，還能輔助個體的學習成長，形成個體與組織間相輔相成的良好循環。在實務上，單位中的成員各有專長，將每一種專長視爲一個模組，模組中持續鑽研該領域，模組的串聯

可以及時解決複雜的專業問題，並提供模組間相互學習的管道。此種模組化的學習型組織就能讓組織細胞活化並持續成長。

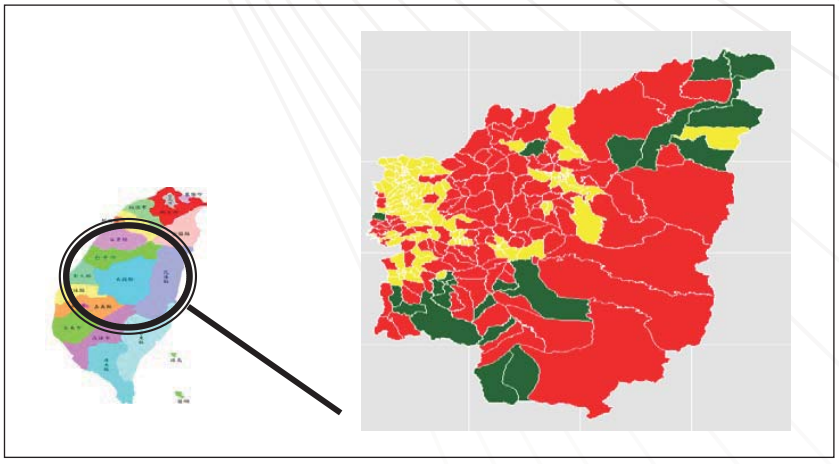
肆、結論

精進抽樣設計是抽樣調查業務須持續努力的作為。透過專業分析及與時俱進的技術學習，不僅完成家庭收支調查之抽樣設計精進作業，同時也達成下列兩項效益：

一、建立數據科學分類機制：針對具複雜社經特性之行政區域資料分類，將資料蒐集、高階統計分析方法、大數據應用及實證等，透過家庭收支調查村里分層作業實作，建立標準作業機制。

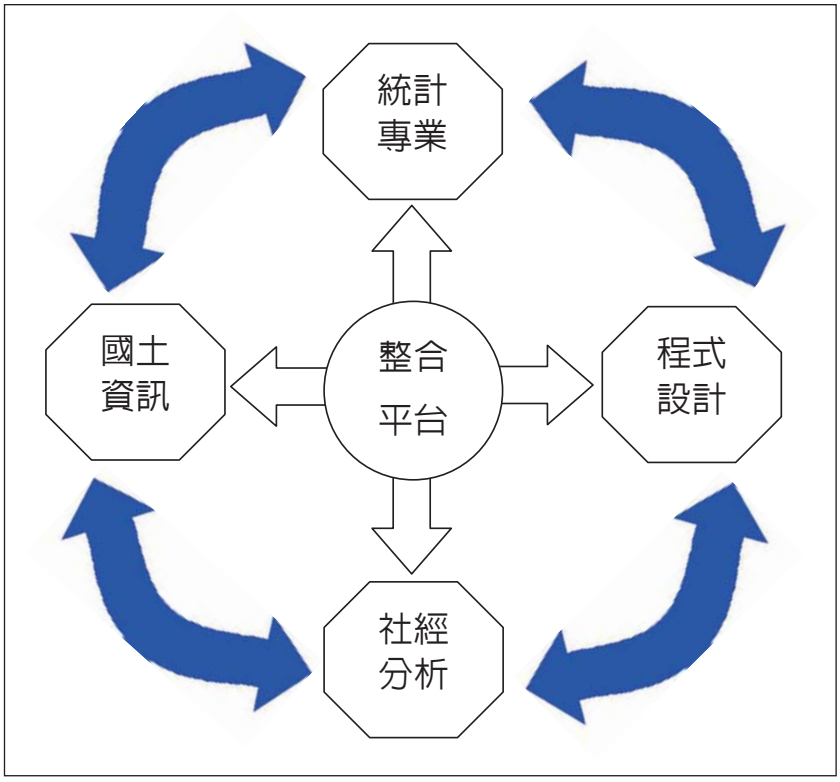
二、建立模組化學習型組織：以零成本，將單位內具社經分析、抽樣設計、多變量統計、國土資訊及程式設計專業之同仁，就程式設計、統計理論及分析等領域，進行本項村里分層作業細部分工，再整合各專業領域成果，建立專業且互相學習的模組化學習型組織。❖

圖 2　地圖呈現分層結果



資料來源：作者自行整理繪製。

圖 3　模組化之學習型組織



資料來源：作者自行繪製。