

STATA 統計軟體探討

胡克威

中央研究院調查研究工作室 助研究員

一、前言

STATA 是由美國加州 St. Monica 電腦開發公司於 1985 年推出的統計分析軟體，除了統計分析之外，還包括繪圖與資料管理等程式。雖然根據國科會所做的統計，國內在研究上使用 STATA 的計畫較 SAS 或 SPSS 少了許多，但是在國外，在眾多的統計套裝軟體中，STATA 可以說是後起之秀，使用者正在急遽增加之中。此外，國際上知名的大型調查資料網站，在釋出資料時，除了 SAS 或 SPSS 格式檔之外，現在也開始將 STATA 規格的資料檔列入。

STATA 是一全面性的統計套裝軟體。最初僅是為生物醫學統計而開發的，但在過去的 15 年之中，版本已升級至第七版，用者由原先的生物醫學研究社群擴展到經濟計量分析與社會統計分析的領域中。雖然 STATA 的發展歷史並非久遠，但與 SAS 或 SPSS 等已經相當通用的統計軟體比較，STATA 存在著相當多的特色與優點，在研究分析上的重要性已受到越來越多研究者的重視。由本期起，本刊將對 STATA 的使用方法作全面性的探討。在本期中，本文討論的重心將放在 STATA 特色與

優點上，目的則是讓讀者對 STATA 有更全面的瞭解。在以後各期中，我們將由淺入深，循序漸進地討論有關 STATA 的各種統計運算與資料處理能力。

二、統計軟體概說

統計軟體一般可以分為 (1) 基本統計套裝軟體與 (2) 具備特殊功能的高等統計軟體。在基本統計套裝軟體就有輕量級與重量級之分。Minitab 可以說是輕量級的統計軟體，基本統計功能齊全，易學易用，但在高等統計分析上則有相當的限制。SPSS 或 SAS 則是基本統計套裝軟體中的重量級代表。除了有基本模組外，也開發高等統計的模組，與其它相關的統計處理產品。例如，SAS 就有 SAS/GRAPH, SAS/OR (用於 operational Research), SAS/ETS (用於經濟計量分析、與時間序列分析等), SAS/IML (為 Interactive Matrix Language 的縮寫), SAS/AF (全稱為 Application Facility, 用於建構作業系統之用), 與 SAS/QC (全稱為 Quality Control) 等。

以特殊功能為主的高級統計軟體 (以 Gauss, Limdep, Eview, Rats,

Genstat & GLIM 等為代表) 並不適用於初學者, 其使用對象完全是以統計為專業的社群為主, 在用途上也非常專門。例如 Eview 為預測分析計量軟體, Gauss 為數值分析與統計計量軟體, RATS 為時間序列統計計量軟體。

在使用者數目的多寡上, 當然是以第一類通用型軟體為多, 尤以 SPSS 或 SAS 為其中的佼佼者。SPSS 由於使用上的親和性, 往往成為統計教學上主要的統計軟體, 尤以在社會科學中為然。SAS 則由於其功能多樣化, 不僅為學界所重用, 也是商用的主要統計軟體, 在知識經濟的時代中, 更由於其卓越功能, 成為資料倉儲 (Data warehouse) 與資料採礦 (Data mining) 上不可或缺的工具軟體。

三、 STATA 的特色與優點

與上述統計軟體比較, STATA 算是通用型的統計軟體, 但如 SAS 一樣, STATA 足以處理相當複雜的統計運算, 也具備了高等統計軟體的功能。一般而言, STATA 至少具有以下九種特色與優點。

1. 價格

STATA 在價格上比 SAS 與 SPSS 便宜。以 SPSS 而言, 其基本模組 (SPSS Base) 的教育版在台灣就高達 5 萬元左右, 若要加入 Regression 或

高等統計模組, 則整套價格就至少是 10 萬元了。SAS 在價格上亦與 SPSS 相去不遠。與這兩件統計軟體相比較, STATA 在價格上就便宜許多了。教育版的 STATA 約為 3 萬元左右, 且不需要額外添購其它模組。

2. 佔用硬碟空間的大小

SAS 與 SPSS 除了價格昂貴之外, 其安裝後所佔的硬碟空間也是不容忽視的問題。一般而言, 安裝基本模組後的 SPSS v.10 版會佔用在 85MB 左右的硬碟空間, 對於硬碟空間不夠的使用者而言, 的確會構成困擾。在最新版的 STATA 7.0 方面, 全部安裝約為 10MB 左右, 較 SPSS 更為節省硬碟空間。

3. 運算時間

STATA 的運算速度相當快。這是 Stata 在分析時是將資料全部讀入記憶體 (RAM 或 virtual memory), 在計算全部完成後才和磁片或硬碟交換資料, 因此 STATA 的運算時間與速度要比 SAS 與 SPSS 來的快了許多。

4. 統計功能模組的建構與彈性

STATA 的計算功能模組可以分為兩類。一是 STATA 內建的指令用來進行資料管理、統計分析、圖形展示、矩陣運算等功能。一是以 ASCII 存檔的特殊程式檔 (即 ado 檔, 儲存

於 c:\stata\ado 之中)。此外，熟悉 STATA 程式語言的使用者，亦可以自行撰寫合乎自己需要的 ado 程式檔（置於 c:\ado\personal 內）。對一般使用者而言，STATA 已提供相當齊全的統計分析模組，在運算能力上並不亞於 SAS，更較 SPSS 為優。單以回歸統計分析而言，STATA 除了提供基本回歸分析的功能外，也包含相當多的回歸診斷程式（regression diagnostics），包括 Standardized and Studentized residuals，DFBETAs，COVARATIO，Durbin-Watson 等檢定。

對於從事學術研究而言，STATA 最為使用者所稱道的程式與指令上的特點有二。第一、STATA 提供相當具有彈性的統計程式運算，容許使用者自行開發 STATA 巨集語言程式，建立自己的 ADO 或 DO 檔。因此，STATA 在運算上可以因應使用者之需求作修改、添加或自行編寫程式。筆者本人，就曾以此開發與回歸分析有關的報表轉換程式。此類程式的開發將會在以後介紹。第二、STATA 是以鍵入指令的互動式計算（interactive computing）為探勘資料的基礎，使用者可以用 log 的指令將運算過程與結果存檔，作為撰寫批次檔（即 do 檔）的基礎。此類批次的功能有助於研究者對所探討的變項進行更為細緻的

探討。

5. 快速的圖形展示能力

STATA 也提供相當不錯的統計圖形展示能力，而且由於是將資料全部讀入記憶體之故，可以很快速地將統計的結果以圖形展現。雖然 STATA 不能如 EXCEL 或 S-Plus 一般，建構 3D 的圖形，但是卻具備充分處理 2D 圖形的能力，並且能夠將 PostScript 或視窗所使用的 wmf 規格轉換成為 Word 的圖形檔。此外，STATA 亦開發出 STATA/GRAPHIC 軟體，可以修飾由 STATA 內建程式中所產生的 STATA 圖形（即 gph 檔）。

6. 資料檔之「可攜帶性」(portability)

STATA 在開發之初即致力於運算指令與資料檔在不同電腦平台之間的相容性（compatibility）。執此之故，STATA 的資料格式檔，具有「可攜帶性」，例如，windows 下所儲存的資料格式就不需要做轉換，就可以在 DOS、Unix、Linux 或 Macintosh 等平台中使用 STATA。因此，STATA 並沒有 SAS 或 SPSS 等軟體在跨越其它運算平台時所遇到的資料格式轉換的困擾。

7. 程式通用性

STATA 在不同電腦平台之間的相容性亦展現在其指令使用上。

STATA 的內建指令一般都可以在各電腦平台中使用，而使用者自行撰寫的程式（如 DO 檔）只需要改變路徑（path）後，就可以在其它平台中運作。

8. 網上功能

由 STATA 第 6 版起，STATA 增加了 use, type, copy 等指令的網頁功能。例如，若使用者鍵入 use <http://www.sinica.edu.tw/~ahu/data/example>，則可以在網路上使用在此網頁中的 example 資料檔。由於 STATA 是將資料檔讀入記憶體中，使用者若鍵入 copy <http://www.sinica.edu.tw/~ahu/data/example.dta>，則可以將此檔案存入自己工作的電腦中。若要查閱網上的文件，我們鍵入 type 指令、網址、該文件路徑與檔案名，就可以查閱到所需要的文件。

由於 STATA 在各平台上的操作指令都是一樣的，這項 STATA 新開發出的網上功能對於研究與教學都具有相當的意義。在研究上，研究者及其和作者可以將資料置於網路中，以網路連結的方式使用資料。在教學上，學生也可以自行到網路上使用有關該科目之資料。

9. 網上更新程式功能、使用者討論區、與 STATA 工具書

STATA 在安裝後有多種方法用來添加其統計處理上的功能。首先，最簡單的方式是在 STATA 中，鍵入 update 的指令，以網路連結的方式，到 Stata 網站尋找並下載最新的升級程式檔（即 ado 檔）。其次，我們也可以參與 Statalist，由電子郵件中參與討論區，或者到 Boston college 的 STATA 網頁中（網址為 <http://ideas.repec.org>）查閱或下載最新編寫程式。此外，我們也可以使用網上搜尋的方法（例如，在 www.google.com 鍵入 Stata & ado），找到其他使用者開發的 STATA 程式檔。

目前市面上也有幾本不錯的 STATA 工具書。除了 STATA 本身所提供的工具書外，筆者個人認為一本由 Lawrence C. Hamilton 所撰寫的 Statistics with STATA（Duxbury press 發行）是一本非常有用的工具書。對統計的初階者而言，不但可以由此書中學會 STATA 的基本指令，也會由其中領略到一些高等統計計算與解釋的問題。對我而言，這本書讓我認識了 STATA，也讓我認識了統計。在今期本文的最後，我鼓勵那些有興趣學習 STATA 卻又是統計新手的讀者去圖書館找這本書，我相信這一趟一定會有收穫。