



SAS、SPSS、STATA

統計軟體檔案格式轉換介紹

陳家玉

一、前言

市面上統計套裝軟體不勝枚舉，目前最常被使用的仍以 SAS、SPSS、STATA 為主流且各有所長。當好不容易申請或取得的資料檔，並非平時熟悉的統計軟體檔案格式時，如何將檔案進行讀取或轉換，是多數人常面臨到的問題。統計軟體檔案格式的轉換，除了利用統計套裝軟體內建的匯入及匯出功能外，Stat/Transfer 是另一個快速且便利的檔案格式轉換工具。本文將統計套裝軟體的匯入/匯出功能分為二部分介紹：第一部分以各軟體資料檔間的轉換為主；第二部分介紹匯入/匯出最通用的純文字檔及普遍用於數據資料整理的 EXCEL 檔，於下以 SAS 9.2 版、SPSS 20 版、STATA 12.1 版逐步分述之。Stat/Transfer 統計數據轉換軟體則針對介面功能操作進行介紹。

二、匯入/匯出統計軟體資料檔功能

在說明如何轉檔前，先瞭解此三種軟體資料檔的差異性（表 1），其中 SPSS、STATA 資料檔可以包含變項說明（variable label）與選項數值說明（value label）兩者，但 SAS 資料檔僅包含變項說明，選項數值說明則是另以格式檔儲存，需套用 format 指令才能取得。

表 1：SAS、SPSS、STATA 資料檔之差異

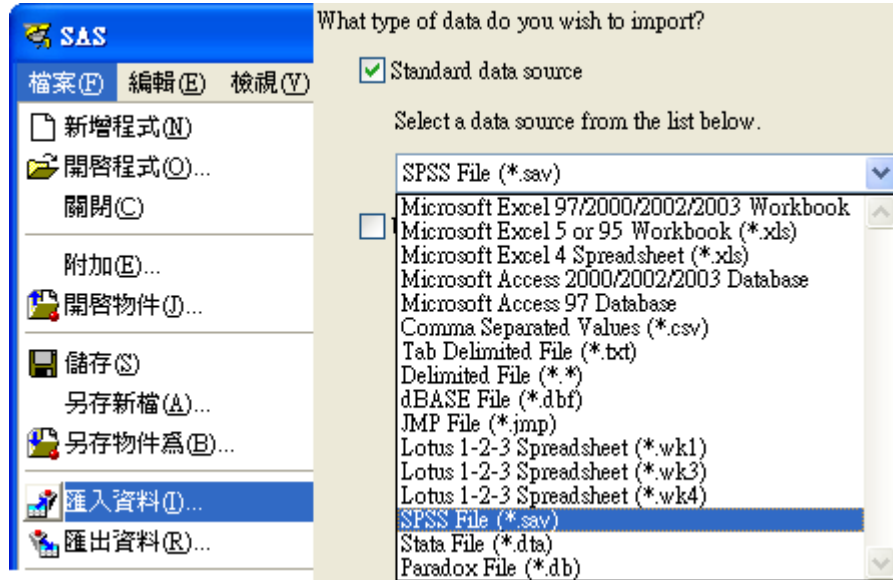
軟體 差異性	SAS	SPSS	STATA
說明	資料與格式分別儲存	資料與格式儲存在同檔案中	資料與格式儲存在同檔案中
資料檔	*.sas7bdat	*.sav	*.dta
格式檔	*.sas7bcac		

數據資料因受限於統計軟體資料檔支援性的問題，較常以純文字檔，搭配語法（syntax）撰寫進行檔案讀取。隨著套裝軟體視窗版的不斷發展、更新且越具易用性，可支援多種檔案類型的匯入與匯出，使得程式語法不再是絕對必要，讓使用者省去自行撰寫程式的麻煩，能快速又方便地讀取或轉換資料。

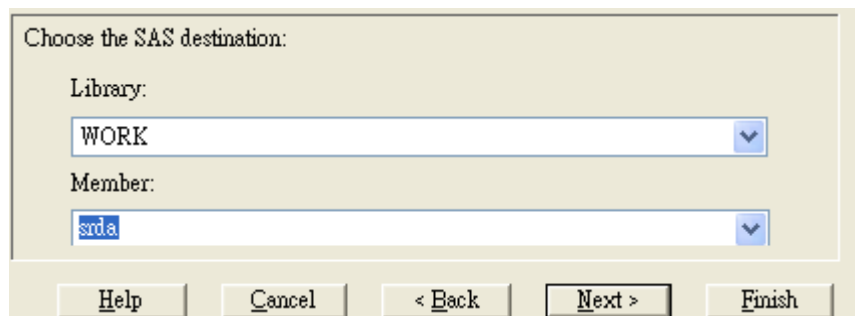
(一) SAS 軟體匯入及匯出 SPSS/STATA 資料檔

1、檔案匯入（圖示以 SPSS 資料檔為例）

由工具列按下檔案 > 匯入資料，直接選取 SPSS 或 STATA 資料檔的格式與檔案。



檔案預設存放於 SAS WORK library 中，並於 Member 自訂檔案名稱，按下 Finish 後即可將檔案匯入。原資料檔中變項說明與選項數值說明，也一併匯入至 SAS Formats 檔。

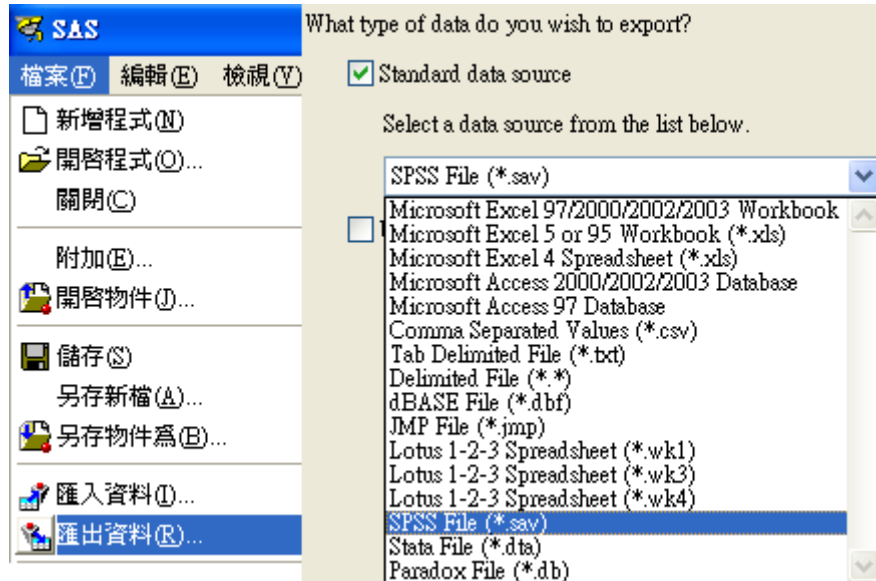


除透過視窗點選外，亦可利用語法讀取檔案，語法請參考如下：

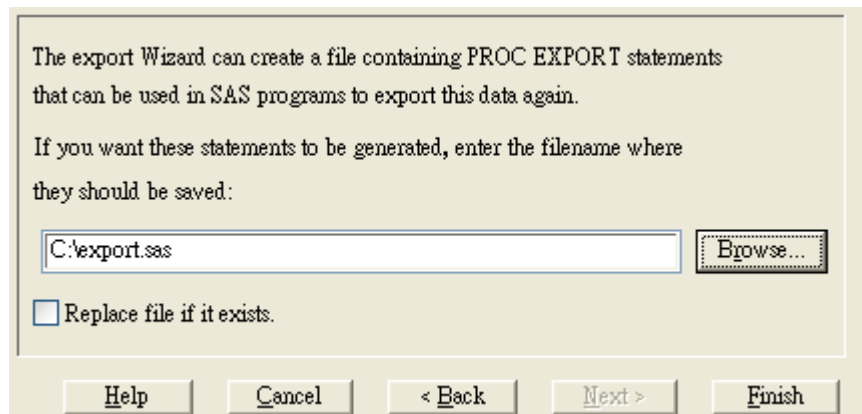
```
PROC IMPORT OUT= WORK.srda  
            DATAFILE= "c:\srda.sav"  
            DBMS=SAV REPLACE;  
RUN;
```

2、檔案匯出（圖示以 SPSS 資料檔為例）

由工具列按下檔案 > 匯出資料，直接選取 SPSS 或 STATA 資料檔格式及儲存路徑；若有 SAS Formats 檔，其變項說明與選項數值說明，也會一併匯出。



整個匯出步驟可透過 Create SAS statements 存取為程式語法，以便往後程式執行時使用。



語法請參考如下：

```
PROC EXPORT DATA= WORK.srda  
              OUTFILE= "c:\srda.sav"  
              DBMS=SAV REPLACE;  
RUN;
```

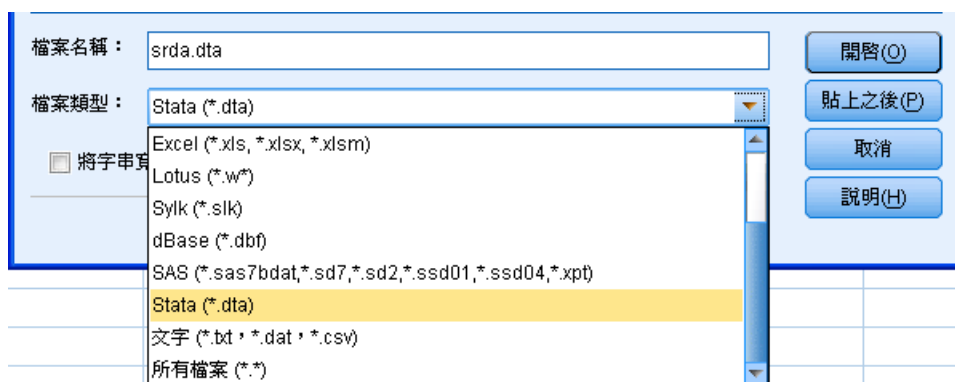
(二) SPSS 軟體匯入及匯出 SAS/STATA 資料檔

1、檔案匯入（圖示以 STATA 資料檔為例）

由工具列按下檔案 > 開啟 > 資料。



直接選取 SAS 或 STATA 資料檔的格式與檔案。



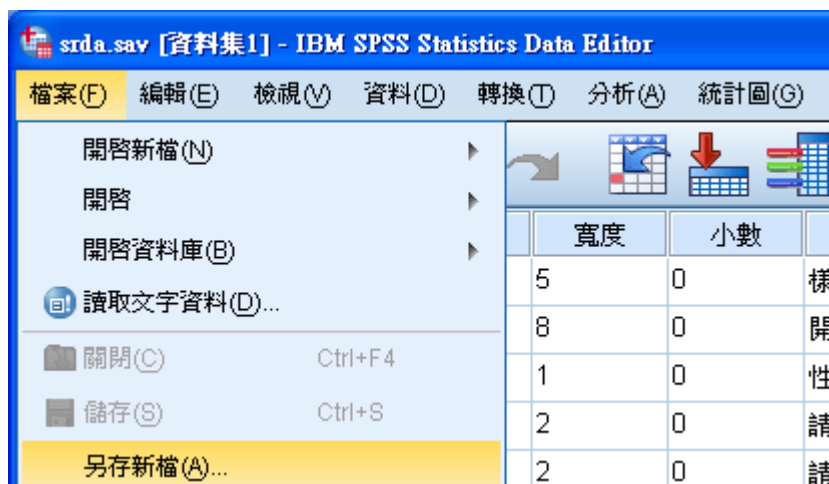
需特別留意因 SAS 資料檔 (.sas7bdat) 與格式檔 (.sas7bcat) 是分別儲存，若以視窗點選介面操作，則無法將對應的選項數值說明一併匯入，須另撰寫程式語法，以 formats 副指令 (subcommand) 宣告格式檔的檔案路徑才可匯入。

語法請參考如下：

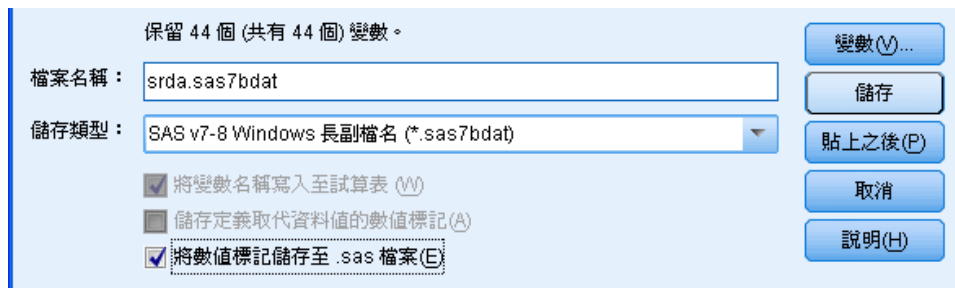
```
GET SAS DATA='c:\srda.sas7bdat'  
/FORMATS='c:\formats.sas7bcat'.
```

2、檔案匯出（圖示以 SAS 資料檔為例）

由工具列按下檔案 > 另存新檔。



直接選取 SAS 或 STATA 資料檔版本後，進行儲存。匯出 SAS 資料檔若勾選「將數值標記儲存至.sas 檔案」，即可一併產生 proc format 語法檔，便於資料之使用。



語法請參考如下：

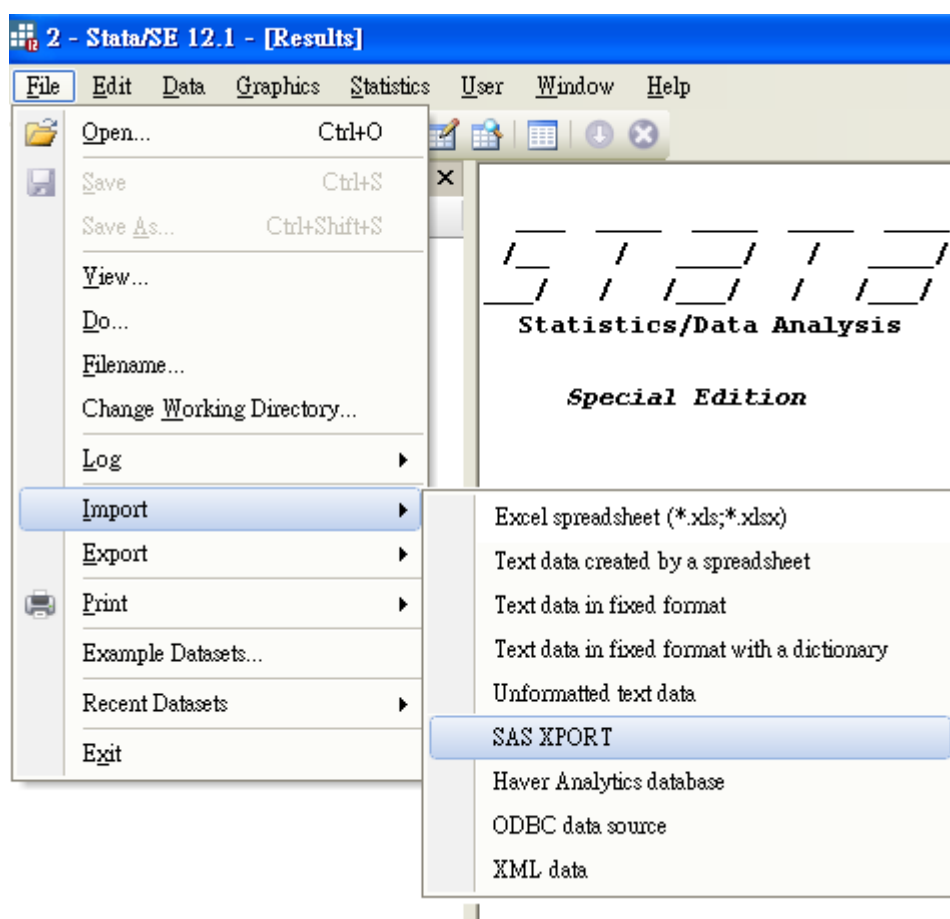
```
SAVE TRANSLATE OUTFILE='c:\srda.sas7bdat'  
/TYPE=SAS  
/VERSION=7  
/PLATFORM=WINDOWS  
/MAP  
/REPLACE  
/VALFILE='c:\srda.sas'
```

(三) STATA 軟體匯入及匯出 SAS Transport 檔、SPSS 資料檔

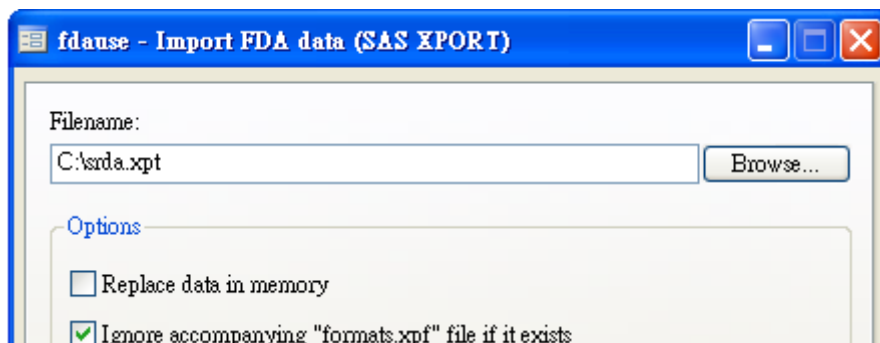
STATA 僅提供 SAS Transport 檔 (*.xpt) 匯入與匯出，並不實際支援 SAS 資料檔 (如 *.sas7bdat) 的讀取，**usesas** 指令也只限於有安裝 SAS 軟體的環境下才可使用；對於 SPSS 檔案支援部分，使用 **usesps** 指令則不需要透過任何軟體，也不需額外花費檔案轉換工具的費用，即可由 STATA 匯入 SPSS 資料檔，但無法由視窗介面操作，須撰寫程式指令處理，目前亦無支援匯出功能。

1、SAS Transport 檔 (*.xpt) 匯入/匯出

由工具列按下 File > Import/Export > SAS XPORT。



匯入時選取檔案，若無 SAS 格式檔可於 Options 勾選忽略 formats.xpf 檔，以避免錯誤訊息產生；匯出時可選擇將選項數值說明儲存於 formats.xpf 檔及 SAS 程式檔中。



語法請參考如下：

```
fdause "c:\srda.xpt", novallabels  
fdasave "c:\srda.xpt", rename vallabfile (both)
```

2、SPSS 檔 (*.sav) 匯入

利用 ssc 指令安裝 usespss 模組 (module)，首先於 Command 視窗鍵入 ssc install usespss，安裝完成後，續於指令列撰寫 usespss 語法，即可成功匯入 SPSS 檔。

語法請參考如下：

```
ssc install usespss  
usespss using "c:\srda.sav"
```

三、匯入/匯出「純文字檔」及「Excel 檔」功能

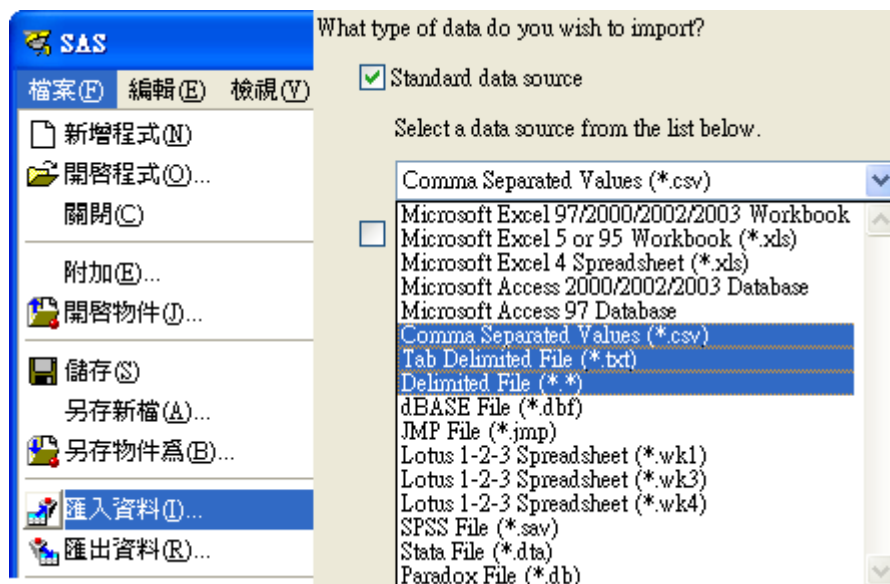
不論電腦檔案格式如何演變，可跨平台使用、不受軟體版本限制、相容性高且最為通用的檔案格式仍是純文字檔；而數據資料的輸入建檔，最普遍且最常使用的則是 Excel 檔，因此上述三種統計套裝軟體大都支援純文字檔及 Excel 檔的匯入及匯出，以下逐一分述之。

(一) SAS 匯入/匯出「純文字檔」及「Excel 檔」

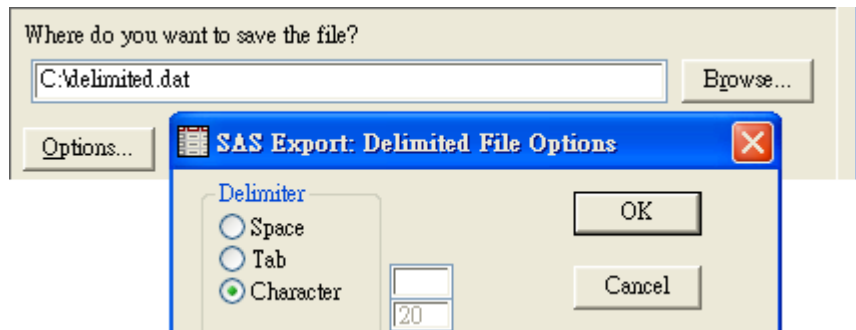
1、純文字檔匯入/匯出由工具列按下檔案 > 匯入資料/匯出資料，依檔案格式可選擇分隔符號文字檔：

- 逗點分隔檔 (Comma Separated Values (*.csv))
- Tab 分隔檔 (Tab Delimited File (*.txt))
- 其他字元分隔檔 (Delimited File (*.dat))

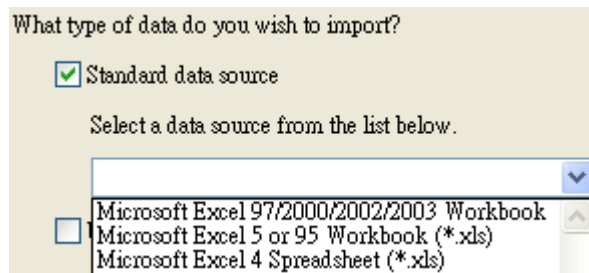
該類檔案格式無法儲存變項說明與選項數值說明的資訊，使用者需再另行撰寫語法宣告之 (LABEL、PROC FORMAT)。若為固定寬度文字檔 (Fixed-Width Text File)，則需採 INFILE 語法讀取外部資料檔並以 INPUT 定義變項欄位。



在輸出設定 (Options) 中，分隔字元 (Delimiter) 可選擇空白分隔 (Space) 或於 Character 輸入自訂字元。

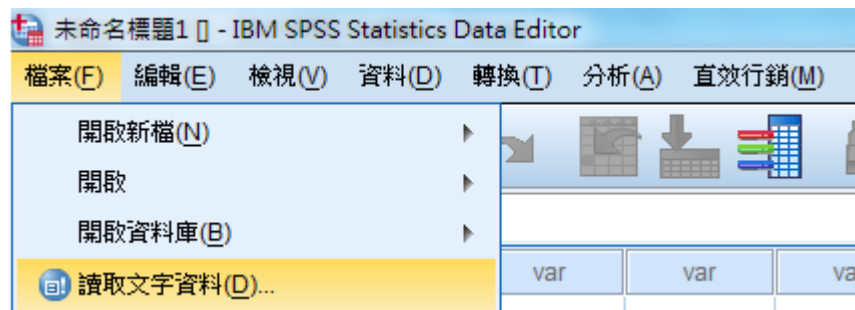


- 2、SAS 9.2 版尚未支援 Microsoft Excel 2007 和 2010 預設檔案格式 (*.xlsx)，匯入/匯出時請先將檔案儲存為較舊的相容版本 (Microsoft Excel 2003 以下)。

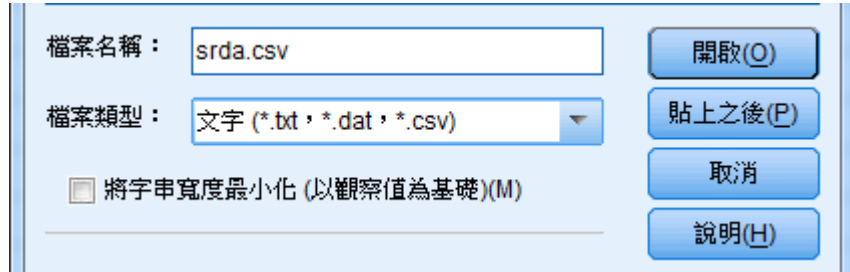


(二) SPSS 匯入/匯出「純文字檔」及「Excel 檔」

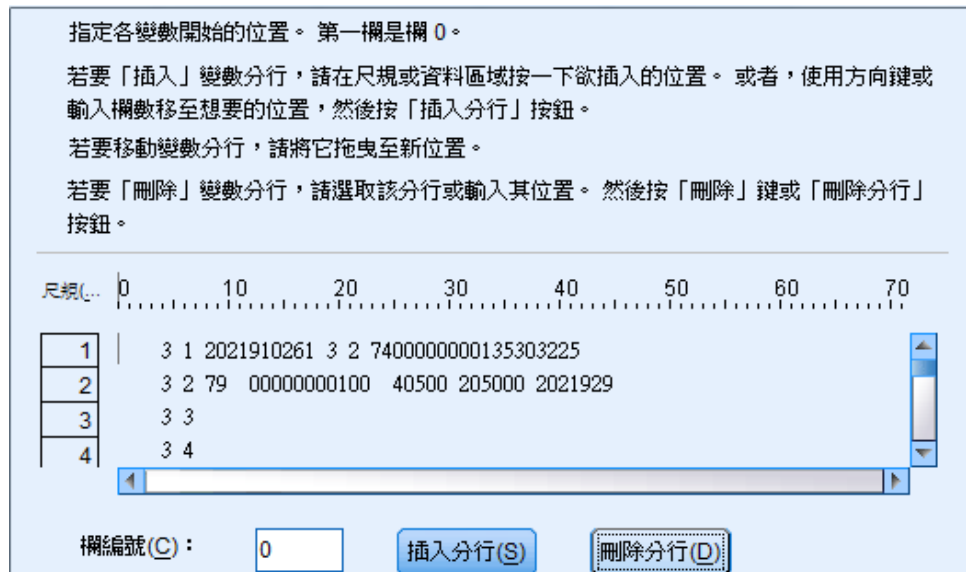
- 1、純文字檔匯入由工具列按下檔案 > 讀取文字資料。



檔案類型可分為Tab分隔檔 (*.txt)、固定寬度文字檔 (*.dat)、逗點分隔檔 (*.csv)，選取檔案按下開啟後，依據文字匯入精靈逐步設定，即可完成資料讀取。¹

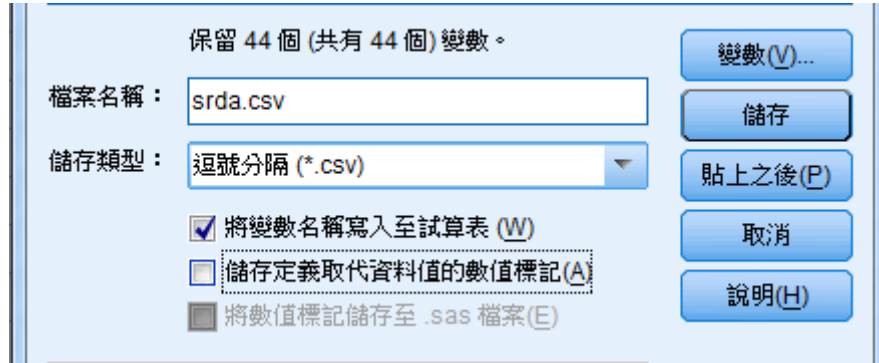


其中固定寬度文字檔，可設定分行讀取（多少行代表一個觀察值），但需以拖曳方式或逐一輸入欄數後按插入分行，指定各變項開始的位置，步驟較為繁雜且容易有誤，因此在讀取該類檔案時，建議採用 DATA LIST FILE 語法讀取資料檔，以利定義變項欄位寬度。



¹ 可將該次設定，儲存成文字精靈預先定義格式 (*.tpf)，供日後再次匯入使用。

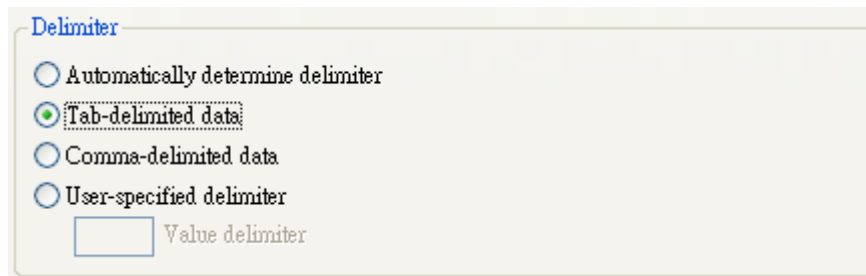
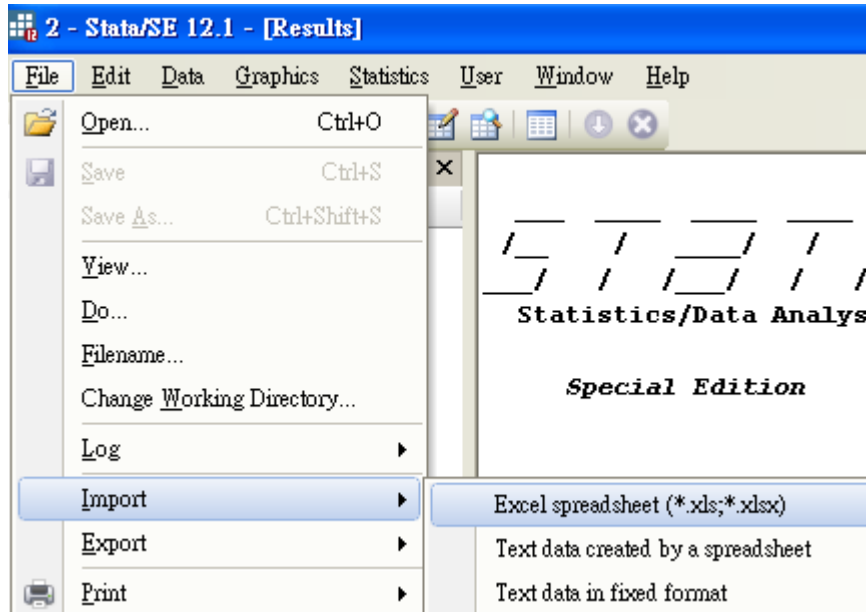
檔案匯出則由工具列按下檔案 > 另存新檔，檔案類型分為 Tab 分隔檔 (*.dat)，逗點分隔檔 (*.csv)，固定 ASCII (*.dat)，可勾選是否將變數名稱寫入至檔案中，即完成資料匯出。



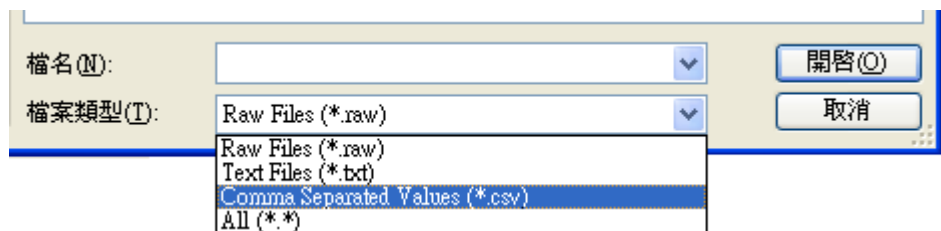
- 2、SPSS 16.0 版以上始支援 Microsoft Excel 2007、2010 (*.xls, *.xlsx, *.xlsm) 檔匯入和匯出。

(三) STATA 匯入/匯出「純文字檔」及「Excel 檔」

由工具列按下 File > Import，檔案類型依序為 Excel (*.xls, *.xlsx)、分隔符號文字檔（自動判定分隔符號、Tab、逗點、自訂字元分隔）、固定寬度文字檔，選取格式與檔案後即可匯入。STATA 12.0 版以上開始支援 Excel 檔的匯入與匯出。



接續選取檔案來源，檔案類型預設為 Raw Files (*.raw)，需在選取檔案時下拉重新選取，方可開啟檔案。



固定寬度文字檔在匯入時需先建立 STATA dictionary 欄位定義檔 (*.dct) 或於 Specifications 逐一填入各變項欄位定義等資訊。

Use dictionary file:

Specifications:

[Examples](#)

rate 1-4 speed 6-7 acc 9-11
2 lines 1: id 1-6 str name 7-36 2: age 1-2 sex 4

STATA 匯出部分提供三項功能可供勾選，使用者可自行依據需求設定。其中，若套有選項數值說明的變項，在匯出時記得勾選已套用選項數值說明的變項以數值輸出（Output numeric values (not labels) of labeled variables），反之則以數值說明內容輸出（請見下圖所示）。

- Do not write variable names on the first line
- Output numeric values (not labels) of labeled variables
- Do not enclose strings in double quotes

	A	B	C	D	E
1	id	stime	a1	a2	a3
2	3	2021910	2	61	3
3	19	2021925	9	21	10
4	42	2021946	2	61	3
5	91	2022000	2	59	3

以數值輸出

	A	B	C	D	E
1	id	stime	a1	a2	a3
2	3	2021910	女性	61	小學
3	19	2021925	9	21	二三專/軍
4	42	2021946	女性	61	小學
5	91	2022000	女性	59	小學

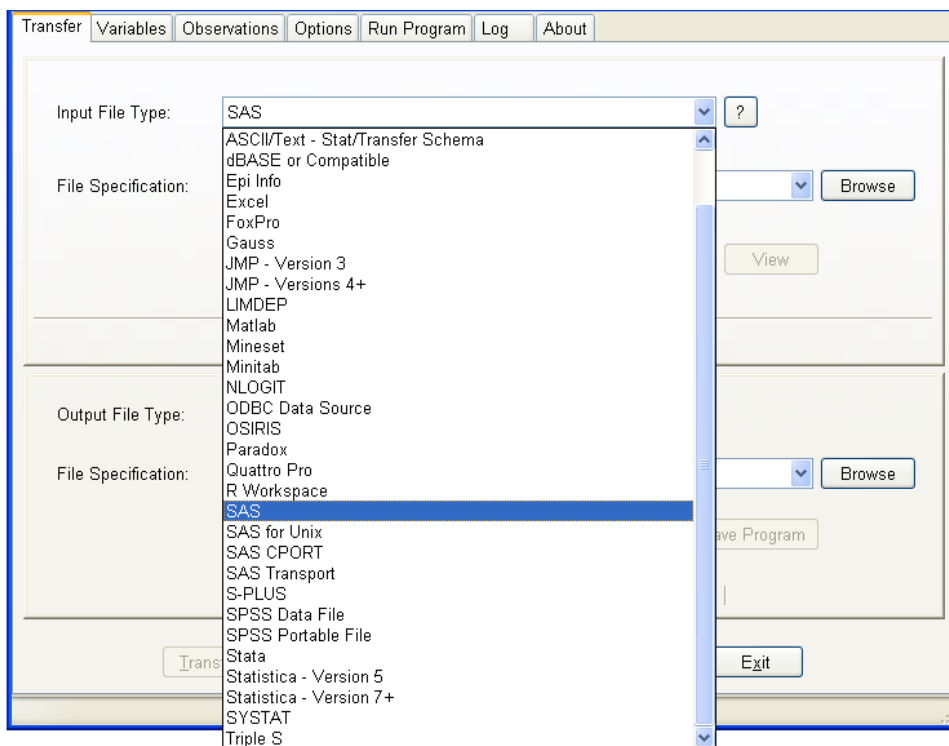
以數值說明
內容輸出

四、使用 Stat/Transfer 統計數據轉換軟體

Stat/Transfer 軟體所能支援的數據資料檔案格式眾多，只要透過標準的圖形介面設定，操作上不但簡便且快速，所需檔案類型即可直接取得。介面上主要提供了四大選單（Transfer、Variables、Observations 及 Options），其相關功能與操作設定以 Stat/Transfer 10 版分述如下：

（一）轉換（Transfer）設定

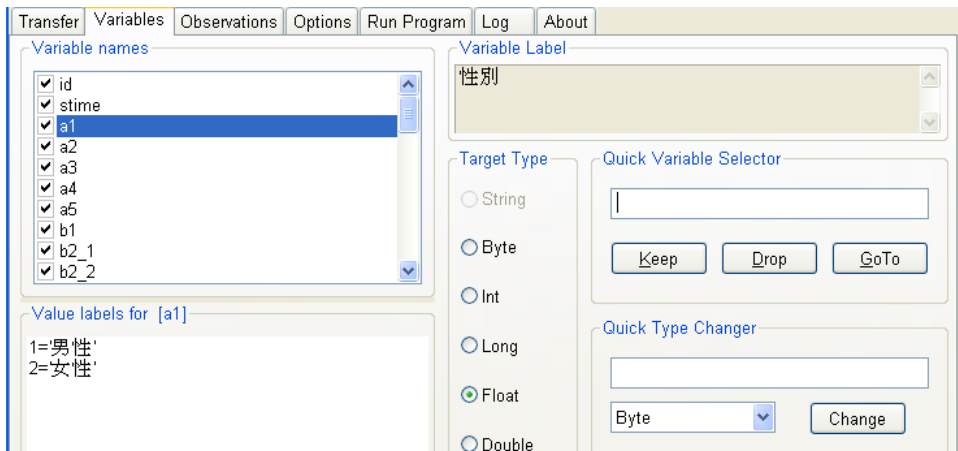
開啟 Stat/Transfer 軟體後，於輸入檔案類型（Input File Type）中選擇包括常用的 Excel、SAS/SPSS/STATA 資料檔與可攜式檔案，並從檔案位置（File Specification）中選取來源檔案的路徑及名稱。



接續於輸出檔案類型（Output File Type）中點選，格式包括Excel 97-2003,2007、SAS/SPSS/STATA資料檔及程式檔（含ASCII）等，再由檔案位置（File Specification）指定檔案的路徑及名稱，相關輸出設定皆完成後，按下「**T**ransfer」即進行檔案之轉換。

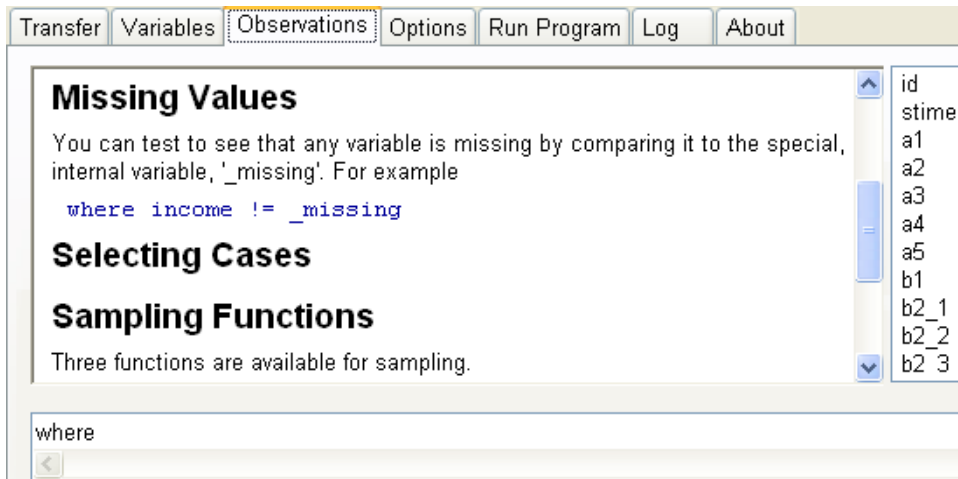
(二) 變項 (Variables) 設定：

檔案輸出若只想轉換部分變項，可於「Variables」介面的左側變項清單中勾選，或在 Quick Variable Selector 中輸入變項名稱(以逗點區隔)，一併按選 Keep 或 Drop。而 Target Type 通常為預設，另可手動自行更改變項的欄寬以及屬性，亦可於 Quick Type Changer 中輸入變項名稱(以逗點區隔)選擇屬性後，可快速更改。



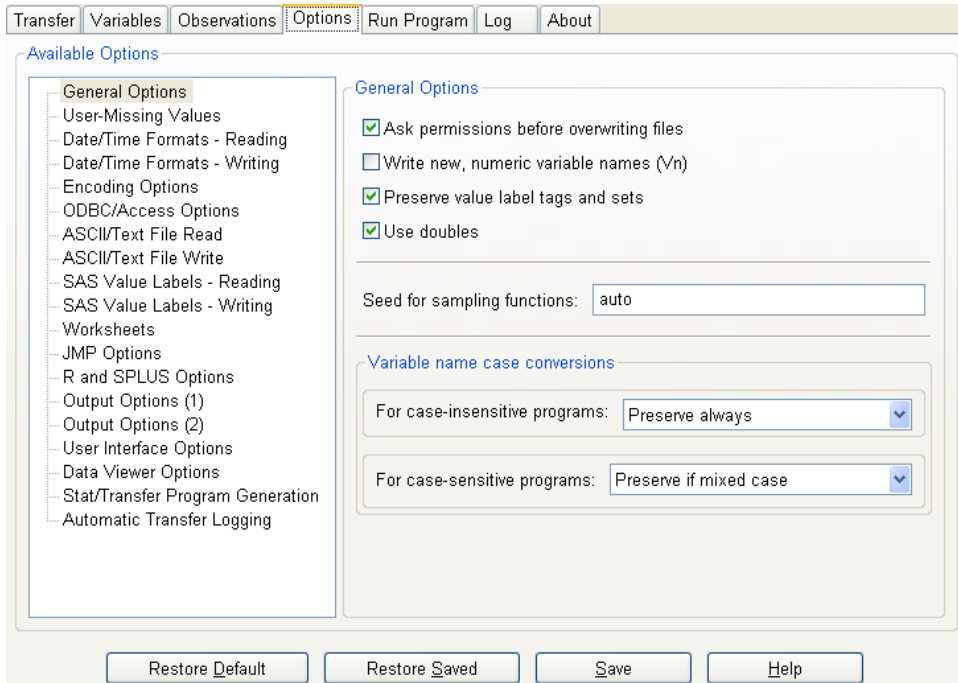
(三) 觀察值 (Observations) 設定：

可於右列看到檔案中所有的變項，在左下方 Where 一欄中，可輸入左上方的指令來篩選，也可進行 Missing Values 處理及簡單隨機抽樣 (simple random sampling)、固定樣本隨機抽樣 (random sampling of fixed size) 和等距抽樣 (systematic sampling) 三種隨機抽樣的功能。

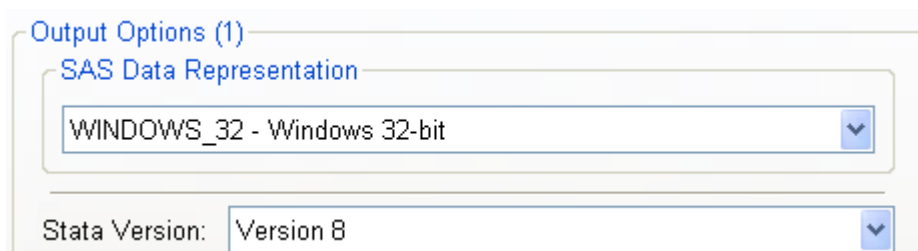


(四) 選項 (Options) 設定：

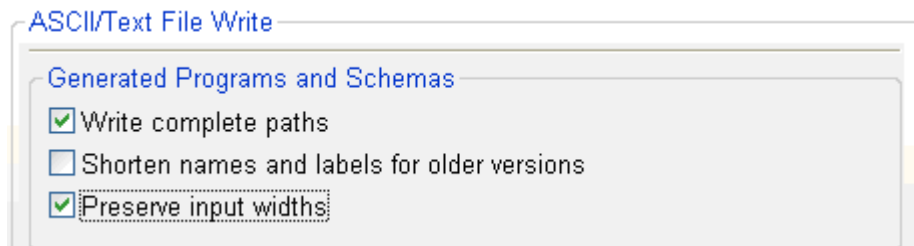
選項設定包含一般、日期/時間格式、編碼、資料庫 (ODBC/ACCESS)、純文字檔 (ASCII/Text File)、SAS、Excel 工作表 (worksheet)、JMP、R 和 SPLUS，以及輸出選項 (Output Options) 等 19 個設定，通常以預設值 (Default) 使用即可。



選項設定主要介紹在轉檔時較常使用的兩項設定功能，分述如次：在使用 STATA 資料檔時，因軟體升級易造成舊版程式無法開啟新版檔案的問題，故在設定轉檔時，可選擇版本設定，由 Options 的 Output Options (1)，於 Stata Version 選定所需版本，亦可按下 Save 將此設定儲存。



另外一提的是關於純文字檔匯出 (ASCII/Text File Write) 的「Preserve input widths」設定，在 Stat/Transfer 10 版以後即使勾選此設定，目前僅能匯出以資料觀察值為基礎其字元寬度最小化的檔案，仍無法維持原資料檔所設定的欄位寬度。



舉例來說，範例檔中「a4 請問您的婚姻狀況」變項，於 SPSS 資料檔中所設定的欄位寬度為 2。

名稱	類型	寬度	小數	標記	值	
a1	數字的	1	0	性別	{1, 男性}...	1 = "未婚"
a2	數字的	2	0	請問您今年__歲?	{98, 拒答}...	2 = "已婚"
a3	數字的	2	0	請問您的教育程度...	{1, 無}...	3 = "離婚"
a4	數字的	2	0	請問您的婚姻狀況...	{1, 未婚}...	4 = "分居"
						5 = "寡居 (喪偶)"
						6 = "同居"
						98 = "拒答"

在該檔案中因數值只有 1~6，雖經 Stat/Transfer 勾選「Preserve input widths」設定，但轉換後欄位寬度仍縮減為「1」，因此使用時需加以留意。

五、結語

統計套裝軟體因應時代的趨勢不斷地改版更新，可支援的檔案類型將更為廣泛及完善，藉由統計套裝軟體內建匯入及匯出功能或 Stat/Transfer 軟體工具的運用，從中找尋利於自己最為便利及有效的方式，成為資料分析最首要的課題。

參考資料

IBM SPSS Statistics 20 Core System User's

Guide http://library.uvm.edu/services/statistics/SPSS20Manuals/IBM_SPSS_Statistics_Core_System_User_Guide.pdf

Sergiy Radyakin 2008 Using SPSS files in

Stata http://www.stata.com/meeting/snasug08/radyakin_usesps.ppt