STATA - Read Data

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

Read Data

- Please see **C2_read_data.do**

Set up the Folders

STATA command: global

Give a name to the paths

- You should use a tree-style directory with different folders to organize your work
  - Easier to find things at a later date

STATA command: global

Give a name to the paths

```
1 global do = "D:\st1\do"
2 global rawdata = "D:\st1\rawdata"
3 global workdata = "D:\st1\workdata"
4 global pic = "D:\st1\pic"
5 global log = "D:\st1\log"
```

- Use **global** to give a name to the paths

STATA command: global

Give a name to the paths

```
1  if "`c(username)'" == "ttyang" {
2  global do = "D:\st1\do"
3  global rawdata = "D:\st1\rawdata"
4  global workdata = "D:\st1\workdata"
5  }
6  if "`c(username)'" == "nest" {
7  global do = "D:\st1\do"
8  global rawdata = "D:\st1\rawdata"
9  global workdata = "D:\st1\workdata"
10 }
```

- You might work on the project using multiple devices or work with other people
- Use **if** to give different path based on different username

STATA command: cd

Sets the Default Directory

```
1 cd $rawdata
2 cd "D:\st1\rawdata"
```

- **cd**: Sets the default directory where Stata will look for any files you try to open and save any files you try to save

use: Read data in Stata format (.dta)

STATA command: use

Read Data in Stata Format (.dta)

```
1 use "D:\st1\rawdata\acs_2015.dta", clear
2 use "$rawdata\acs_2015.dta", clear
```

- **use**: Read data in Stata format (.dta)
  - Use **$rawdata** to represent the path
- The **clear** option will clear the revised dataset currently in memory before opening the other one

STATA command: use

Read Data in Stata Format (.dta)

```
1 cd $rawdata
2 use "acs_2015.dta", clear
```

- Use **cd** to change directory to rawdata folder
- Then, load data directly

STATA command: use

Read Data in Stata Format (.dta)

```
1 use pernum sex age year using "$rawdata\acs_2015.dta"
      , clear
```

- If you do not need all the variables from a data set, you can also load only some of the variables (here: pernum sex age year) from a file

import delimited: Import csv or txt Files to STATA

STATA command: import delimited

Import csv or txt Files to STATA

- Syntax:

```
1 import delimited [using] filename [, options]
```

```
1 import delimited using acs_2015.txt,clear
2 import delimited using acs_2015.csv,clear
      encoding("utf-8")
```

- **import delimited**: Load a delimited text file
- Import a .txt (.csv) file from outside
  - **.txt** extension: a kind of computer file that is structured as a sequence of lines of electronic text
  - **.csv** extension: a delimited text file that uses a comma to separate values.
- **encoding("utf-8")**: Make sure you can read Chinese characters correctly

STATA command: import delimited

Import csv or txt Files to STATA

```
1 import delimited ID using acs_2015.csv,clear
      varnames(1)
2 import delimited using acs_2015.csv, colrange
      (1:3) rowrange(3:8) clear
```

- **Line 1:** You can rename specified variable name of imported columns: ID
  - **varnames(1)**: use first row as variable name
- **Line 2:** option **colrange([start][:end])** specifies column range of data to load
- **Line 2:** option **rowrange([start][:end])** specifies row range of data to load

save: Save to STATA format

STATA command: save

Save to STATA format

```
1 save "$rawdata\acs_2015.dta", replace
```

- **save**: save to dta file (STATA format)
- **replace** option: overwrites any previous version of the file in the directory you try saving to
- If you want to keep an old version as back-up, you should save under a different name

export delimited: Export csv or txt Files to STATA

STATA command: export delimited

Export STATA data to csv or txt Files

```
1 export delimited using acs_2015new.csv,replace
```

- export delimited using acs_2015.csv, which can be read in excel
- replace option: replace previous version of acs_2015.csv

import/export excel: Import/Export excel Files to STATA

STATA command: import/export excel

Import/Export excel Files to STATA

```
1 import excel datanum year serial using
      acs_2015new.xlsx, sheet("Data") clear
2 export excel using acs_2015new.xlsx, sheet("Data"
      ) replace
```

- **Line 1:** Import data from a Excel file (*.xls;*.xlsx)
  - option **sheet**: specify variable names and name of spreadsheet that you want to import
- **Line 2:** Export data to Excel file (*.xls;*.xlsx)
  - option **sheet**: specify name of spreadsheet that you want to export

infix: Import ASCII format (dat) Files to STATA

STATA command: infix

Import ASCII format (dat) Files to STATA

```
clear
  infix                  ///
  int      year          1-4    ///
  byte     datanum       5-6    ///
  double   serial        7-14   ///
  using `"$rawdata\acs_2015.dat"'
```

- **infix**: read fixed ASCII format data
  - **ASCII:** American Standard Code for Information Interchange
- This type of data cannot be read directly
- To read this type of data into Stata, we need to use the **infix** command and provide Stata with the information from the codebook
- We need a codebook that explains how the data is stored

STATA command: infix

Import ASCII format (dat) Files to STATA

```
clear
quietly infix                  ///
int      year          1-5     ///
byte     datanum       6-7     ///
double   serial        8-14    ///
using `"$rawdata\acs_2015.dat"'
```

- Make sure you follow the codebook
- Otherwise, you might create a wrong file!!

Other useful commands: compress, preserve, restore

STATA command: compress

```
1 compress
```

- **compress**: to preserve space, only store a variable with the minimum string length necessary

STATA command: preserve/restore

Preserve and Restore Snapshot of the Current Dataset

```
1 preserve
2
3 restore
```

- If you are going to make some revisions but are unsure of whether or not you will keep them, then you have two options
- **preserve**: this command will take a snapshot of the current dataset
- If you want to revert back to that copy later on, just type **restore**

help: Self-Learned Information

STATA command: help

Self-Learned Information

- **help** command displays useful information about the how to use specified command or specific topic
- How to use **help**
  - For example, if you want to know command **use**, you can type **help use**