Stata Workshop: DAY 1

Taisei NODA

Graduate School of Economics, Osaka University

June 11, 2019

Overview

About this course

- Instructor: Taisei NODA(tge016nt@student.econ.osaka-u.ac.jp)
- Class time: 9:00-10:30, June 11 and June 18
- This course is supplementary course for Econometric Method I
- I expect that students have never used Stata before, but have basic understanding of statistics (e.g. mean, standard deviation).
- Everything we will cover is on printed "do-file". Go through on your pace.
- Don't hesitate to ask for help. We have TAs.

Schedule

DAY 1: June 11

- Introduction to Stata
- Set up
- Basic Syntax/operation
- Explore and summarize the data
- Create new variables
- Select and replace parts of data
- Label Data

Schedule

DAY 2: June 18

- Import data
- Visualization
- Data analysis
- Automation; loop and macro (if we have time)
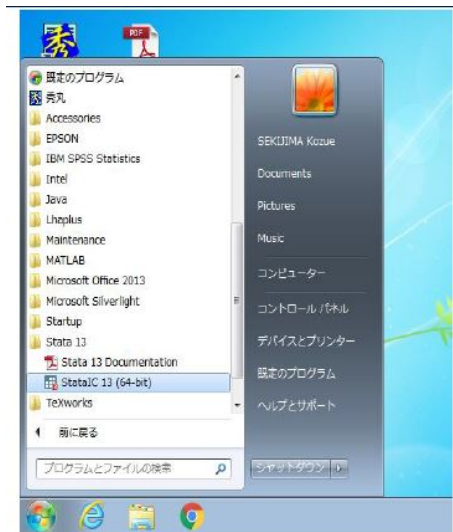
Important Reference

- Stata Cheat Sheet by Tim Essam
  - Everything you need (at least for beginners) is nicely summarized
  - Google "Stata Cheat Sheet"

What is Stata and Why?

- **Stata** is a statistical software package, widely used especially in the field of social science.
- In these days, many of top academic journals publish Stata code for replication of empirical results found in papers.
    - Thus, being familiar with Stata is critical to follow up the latest studies.
- All the OSIPP students have access to this software.

Launch your Stata

- "Start button" $\Rightarrow$ "All the programs" $\Rightarrow$ "Stata 13" $\Rightarrow$ "StataIC 13"

Stata has windows

Three types of files

- **do file**: Script file. You place all the commands you want to perform in the file.
- **dta file**: This is data format Stata uses. Stata stores in an unique format.
- **log file**: This is record where every your work and output displayed in reuslts window is documented.
- If Stata is cooking, do file is a recipe, dta file is ingredient and log file is a dish.

Principles of data analysis

- Keep your work reproducible
    - It is thus important to save what you did as do file and log file.
- Automate repetitive tasks whenever you can
    - You want to avoid making mistakes.
    - It is recommended to Write everything you want to perform on do-file

Launch your do-file editor

- Click "New Do-file"
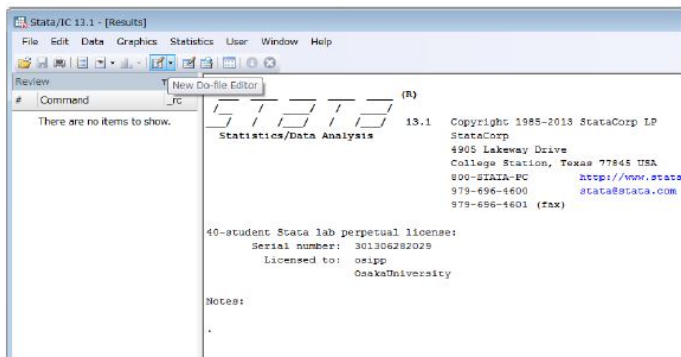
Do-file editor
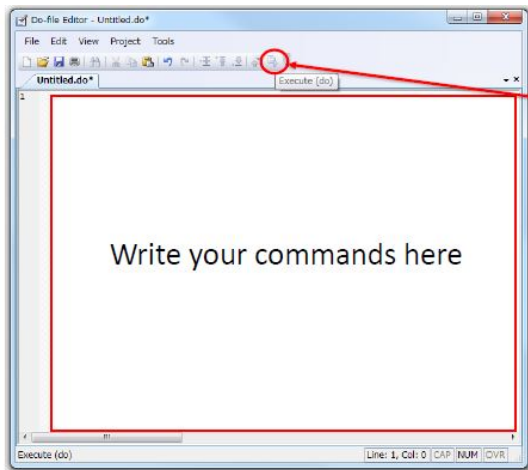
To **execute your do-file**, click the "Execute (do)" button.

Write your commands here

- Or, hit Ctrl+d on your key board

Execute your code


```
  stataworkshop2019_day1.do   Untitled.do
 1
 2        /*************************************************
 3                             Stata Workshop 2019 DAY 1
 4                             2019/6/11 GSE Taisei NODA
 5        *************************************************
 6
 7        /*********************************
 8                     Set up
 9        *********************************/
10      *"Routine work". Ensure that no dataset or log is remain
11      clear
12      capture log close
13      *avoid interruption
14      set more off
15
16      *display your current working directory
17      pwd
18      *change working directory
19      *cd "D:"
20      *display filenames in working directory
21      dir
22
```

"Startup routine"

- Make sure that your stata has nothing restored before starting.
  - **clear**
  - **capture log close**
- Avoid interruption during running your code
  - **set more off**

Make comments on do-file

- You can make comments on dofile by using asterisk (*)
- Stata ignores line which started by asterisk and parts enclosed /* and */ (colored by green)

```
*You can make comments on do-file by putting asterisk in the begining of line
/*
If you want to make comment in multiple lines, enclose your comment by /* and */
(the order does matter), as I am doing.
*/
```

Set up

- Display your current working directory (working directory is the place where Stata search data file and save your output, such as log file and do file).
    - **pwd**
- Change working directory
    - **cd** "D:"
- Display file names in working directory
    - **dir**
- Today, we use a sample file installed in Stata ("Auto" data). This data describes characteristics of cars. Each observation is a car.
    - **sysuse auto,clear**

Basic Syntax

Every program in Stata has following structure.

`[by varlist1:] command [varlist2] [=exp] [if exp] [in range] [weight] [using filename] [,options]`

- Each line consists of nine parts (at most).
  - **by** *varlist* **:**
  - **command**
  - *varlist*
  - **=** *value*
  - **if**
  - **in** *range*
  - *weight*
  - **using** *filename*
  - **,options**
- "varlist" means variable list (e.g. "wage education.....")

Basic Syntax (cont'd)

- You do not always write all of the nine parts (,we often do not).
- We must include "**command**", usually "*varlist*", and often "**if**" and ",options".
    1. Stata is sensitive to order and UPPER CASE/lower case. Be aware of that.
    2. Stata recognizes separated lines as separated codes. Don't split into multiple lines.
    3. Stata considers space separated letters as a independent word. For instance, "**sysuseauto,clear**" or "**sysuse au to, clear** don't work

Basic Syntax - Example

- We now work on "summarize" command as an example
- We will come back to "bysort :" later.

```
*Example: summarize

bysort rep78: summarize price if foreign==0&price<=9000,detail

*summarize scalculates and displays a variety of univariate summary statistics.

summarize
summarize price
*"detail" option displays additional statistics, such as median and quartiles
summarize price,detail
*Stata allows us to use abbreviation
sum price,d
```

- See do-file

Next Week

- June 18-
- We will cover how to import excel and visualize and analysis your data
- Bring today's do-file