

ECON B253: INTRODUCTION TO ECONOMETRICS

Sebastian Anti, Assistant Professor of Economics
Bryn Mawr College

Lecture 1 b:

Stata

Orientation

Stata Orientation

2

- **What is Stata?**
- **How can I access Stata?**
- **What are the basics of its use?**
 - ▣ Interface
 - ▣ File structure
 - ▣ Coding resources
 - ▣ Troubleshooting
- **Stata vs. R**

Before we begin, open up the course's Moodle page.

What is Stata?

3

- **Stata is data analysis software**
- **It is (still) the most commonly used data analysis software used by economists**
 - ▣ R is catching up though
- **It is extremely good at handling regression analysis, the core of econometrics**

How Can I Access Stata?

4

- **Stata costs money**
 - ▣ But Bryn Mawr has purchased access for you!
- **You can access Stata through Bryn Mawr's computer lab**
 - ▣ Go to <https://techdocs.blogs.brynmawr.edu/9708> and follow the instructions
 - ▣ For Haverford Students, try:
<https://iitskb.sites.haverford.edu/knowledge-base/software-downloads/>
 - ▣ Swarthmore: <https://kb.swarthmore.edu/display/SW/Stata>

Using Stata

5

- **Navigate to Stata and open it!**

Button to view data in spreadsheet form

Button to stop code before it finishes running

The screenshot shows the Stata/MP 15.1 interface. The top toolbar includes icons for Open, Save, Print, Log, Viewer, Graph, Do-file Editor, Data Editor, and Data Browser. The Data Browser icon is highlighted with an arrow pointing to the text 'Button to view data in spreadsheet form'. On the right, the 'More' button (a green circle with a white arrow) is highlighted with an arrow pointing to the text 'Button to stop code before it finishes running'. The main window is divided into three panes. The left pane, titled 'Results', shows the output of several Stata commands: `. hist var4` (with bin=9, start=-9.7014265, width=2.1766433), `. sum var4` (displaying a table of statistics for var4), `. *(1 variable, 2241 observations pasted into data editor)`, `. hist var5` (with bin=33, start=-9.9814386, width=.60542283), `. sum var5` (displaying a table of statistics for var5), and `. label variable var1 "Description of var1"`. The middle pane, titled 'Command', shows the command `reg var1 var2`. The right pane, titled 'Variables', shows a table of variables in the dataset. The table has two columns: 'Name' and 'Label'. The variables listed are var1, var2, var4, and var5. The label for var1 is 'Description of var1'. An arrow points from the text 'Description of variables' to the label for var1. Another arrow points from the text 'List of variables in your dataset' to the list of variables. A third arrow points from the text 'Field to search for variables in your dataset' to the search bar at the bottom of the Variables pane. A fourth arrow points from the text 'Display area for results of code you run' to the Results pane. A fifth arrow points from the text 'Field where you enter code' to the Command pane.

Display area for results of code you run

Field where you enter code

Description of variables

List of variables in your dataset

Field to search for variables in your dataset

Variable	Obs	Mean	Std. Dev.	Min	Max
var4	81	.3097248	5.688299	-9.701427	9.888363

Variable	Obs	Mean	Std. Dev.	Min	Max
var5	2,241	-.0798875	5.784554	-9.981439	9.997515

Name	Label
var1	Description of var1
var2	
var4	
var5	

Command
`reg var1 var2`

/Users/sebastiananti/Dropbox/New Cambodia Folder/Data for Stata

Stata File Structure

7

- **You'll be working with two types of Stata files: (1) .do files, and (2) .dta files**
- **A code “script” is a file that contains your code**
- **The Stata file where you save your code scripts is called a “Do” file. Its file extension is “.do”**
 - ▣ You will need to submit your .do files of your problem sets for your problem set submissions to be considered complete
- **Files of data that are readable to Stata have the extension “.dta”**

Using Stata

8

- **Navigate to Stata and open it!**
- **Open up a .do file (File >> New >> Do-File)**

Untitled

Open Save Print Find Show Zoom 116%

Results File Robustness Ch... Refugee_Draft... Rubustness Ch... Data Prep for... AJPS_Checkp... Data Processin... Data Processing CSES Cambodi... Post_Prelim_A... Untitled

```
1 * An asterisk tells Stata not to read this line of code. It is useful for inserting comments to yourself
2 // Two slashes do the same thing
3
4 /* Text bookended by "/*" and "*/" block out large chunks of code
5 that span multiple lines. It is useful for longer notes to yourself */
6
7 /* When you write a line of code and you want Stata to run it, you highlight it
8 and then click the "Do" button in the upper right-hand corner of the window.
9 Then check Stata's output window to see what the results are. */
10
11 /* If you have code that is really long, you can continue it on the next line by inserting "///".
12 Here is an example: */
13
14 reghdfe migrated treat_0_5 treat_6_10 treat_11_15 treat_16_20 ///
15 camp_0_5 camp_6_10 camp_11_15 camp_16_20 ///
16 no_members sex age hh_head_sex educ_yrs hh_head_age_yrs rural_binary ///
17 if min_dist_camp_1999_2016<=100 & age<=49 & country_sub_group==1, ///
18 absorb(district_code district_code#c.year province_code#i.year) vce(cluster dhsid)
```

To run your code, just highlight what you want to run and press this button

Here are some useful tips for coding in a .do file

83:18

Using Stata

10

- ❑ **Navigate to Stata and open it!**
- ❑ **Open up a .do file (File >> New >> Do-File)**
- ❑ **Download PS1's dataset (the .dta file on Moodle)**
- ❑ **Open it in Stata**

Stata File Structure

11

- **It's best practice to keep all your files for a project in one folder in the location you are using to save your project (wherever that may be)**
 - ▣ This location is called the “working directory”
- **You should to tell Stata where that working directory is whenever you settle in to work with Stata on a project**
- **You can do this by going to File >> Change Working Directory, and then selecting the file where you are storing everything**

Using Stata

12

- ❑ **Navigate to Stata and open it!**
- ❑ **Open up a .do file (File >> New >> Do-File)**
- ❑ **Download PS1's dataset (the .dta file on Moodle)**
- ❑ **Open it in Stata**
- ❑ **Write in your .do file: `scatter wage educ`**
- ❑ **Run this by highlighting it and clicking the button in the top right of the .do file window**

Coding Resources

13

- ❑ **Coding can be really fun. It's kind of like a puzzle. All the pieces have to fit together correctly**
- ❑ **Most people who code a lot aren't "fluent" in the coding language. We are constantly researching the right code to achieve what we want**
- ❑ **Don't be afraid to Google how to do things when you get stuck**
 - ❑ In fact I've designed some questions specifically for you to practice this

Coding Resources

14

- **Stata's User Manual is great:**
<https://www.stata.com/manuals/u.pdf>
 - ▣ Use control F on this. It's too big to leaf through
- **UCLA has some nice tutorials on how to do basic things**
 - ▣ <https://stats.idre.ucla.edu/other/annotatedoutput/>
- **StataList is a great online forum**
 - ▣ Many people have had your questions answered many times over here already
 - ▣ <https://www.statalist.org/forums/>

Troubleshooting

15

- **You will get errors when you code. Don't get discouraged!**
 - ▣ Look at what you've done. Try something different based on your intuition, Google your problem, try again, repeat
 - I have worked coding problems like this for months(!) before I found out how to achieve what I wanted
 - ▣ If you get stuck and can't work it through and you've given it a good shot by consulting resources online come to me in office hours
 - ▣ Don't wait to the last minute to do problem sets!

Stata vs R

16

- **Many of you may be interested in learning R. Sorry to disappoint you**
- **R is great and much more versatile than Stata**
 - ▣ It can do econometrics *and* other kinds of analysis that economists do all the time that Stata does poorly or not at all like,
 - Geospatial analysis and map creation
 - Monte Carlo simulations
 - Computable General Equilibrium (CGE) analysis
 - Many, many, others
 - ▣ I personally use both about equally in my own work

Stata vs R

17

- **For this reason R is becoming more popular every day among economists, especially younger ones**
- **However, when I want to do *econometrics*, I always still work in Stata**
 - ▣ Stata is faster, more efficient, and easier to work with than R when doing econometrics
 - ▣ Stata is like a sports car, R is a like a Jeep
- **Since econometrics is obviously our focus, we'll be using Stata**

An Aside

18

- **Also, my view is once you've gained competency in one coding language it is easy to jump over to another one**
 - ▣ So if you want to eventually learn R, learning Stata is a great way to start!

Some Notes on the Problem Set

19

- ❑ **Be sure to read the “Problem Set Instructions and Tips” document in the Problem Sets Section of the Moodle Page**
- ❑ **Start with plenty of time to get stuck and potentially see me in office hours to resolve difficulty**
- ❑ **Have fun!**