

MSc in Economics for Development
Induction Week: Stata

Binta Zahra Diop
bintazahra.diop@economics.ox.ac.uk

October 11, 2019

Today:

Go over basic commands and then go through and hands on tutorial.

Introduction

Basic Commands

.do File

More commands

Macros and Graphs

IPA training

Stata

- ▶ A statistical software similar to R, Python, SPSS, etc.
- ▶ Stata is the software that economists use the most. It's used to:
 1. Clean datasets
 2. Analyze data
 3. Plot data
- ▶ Arguably: programming and statistical analysis is the most “transferable” skill you'll take away from here.
- ▶ Most/all of you will work with data in your extended essay.

Opening Stata

- ▶ Start → All Programs → Stata 15 → Stata SE 15

OR

- ▶ Open a Stata data file (.dta or .do) by double-clicking.

Main Window

Stata/SE 15.1

Open Save Print Log Viewer Graph Do-file Editor Data Editor Data Browser

Q Search Search Help

Review Command _rc

Results

(R)

Statistics/Data Analysis 15.1

Special Edition

Copyright 1985-2017 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user Stata license expires 8 Oct 2019:
Serial number: 401509015060
Licensed to: Binta Zahra DIOP
Oxford

Notes:

1. Unicode is supported; see [help unicode_advice](#).
2. Maximum number of variables is set to 5000; see [help set_maxvar](#).

Command

Properties

Variables

Name	Label
------	-------

Data

File name	Label
Variables	0
Observations	0
Size	0
Memory	64M
Sorted by	

/Users/binta/Documents/_DPH/Research/Projects/GHS/z_temporary

Main Window

The screenshot displays the Stata/SE 15.1 main window. The interface is divided into several panels:

- Review Panel (Left):** Contains a 'Command' window with the text `_rc`.
- Results Panel (Center):** Displays the Stata startup screen, including the Stata logo, version 15.1, and copyright information. It also shows the single-user license expiration date (8 Oct 2019) and a list of notes.
- Variables Panel (Right):** A table with columns 'Name' and 'Label' for listing variables.
- Properties Panel (Bottom Right):** A panel showing properties for the current dataset, including variables, observations, size, and memory.
- Command Panel (Bottom):** A window for entering Stata commands.

The top of the window features a menu bar with options like Open, Save, Print, Log, Viewer, Graph, Do-file Editor, Data Editor, and Data Browser. A search bar is also present in the top right corner.

RESULTS

Variables

Commands

Basic commands

```
display "Hello"
```


Basic commands

```
display "Hello"
```

```
display 2 * 3
```

Basic commands

```
display "Hello"
```

```
display 2 * 3
```

```
display 2 * 3 + 5
```

Basic commands

```
display "Hello"
```

```
display 2 * 3
```

```
display 2 * 3 + 5
```

```
help
```

Basic commands

```
cd "H:\StataTraining"
```

Basic commands

```
cd "H:\StataTraining"
```

```
use sawages.dta, clear
```

Basic commands

```
cd "H:\StataTraining"
```

```
use sawages.dta, clear
```

```
describe
```

Basic commands

`browse` allows you to search the dataset (in an excel like format).

`browse`

Basic commands

`browse` allows you to search the dataset (in an excel like format).

`browse`

`browse exp prof cler`

Basic commands

`summarize` is a command that gets you summary statistics for variables you are interested.

Note that you can shorten the command. You can also add options.

`summarize`

Basic commands

`summarize` is a command that gets you summary statistics for variables you are interested.

Note that you can shorten the command. You can also add options.

```
summarize
```

```
sum exp
```

Basic commands

`summarize` is a command that gets you summary statistics for variables you are interested.

Note that you can shorten the command. You can also add options.

```
summarize
```

```
sum exp
```

```
sum exp, detail
```

```
sum exp, d
```

Basic commands

tabulate (tab) - lists all the values the variable takes in increasing (or alphabetical) order as well as their frequency. You can also add options.

```
tab exp
```

Basic commands

tabulate (tab) - lists all the values the variable takes in increasing (or alphabetical) order as well as their frequency. You can also add options.

```
tab exp
```

```
tab exp, m
```

```
tab prof male, m
```

Note that missing observations (.) are treated as infinity when sorting values.

Basic commands

`generate (gen)` allows you to create variables.

Try:

```
generate one = 1
```

```
browse
```

```
br
```

Basic commands

if allows you to add conditions.

```
gen edcat = 0 if edyrs == 0  
tab edcat
```

Basic commands

if allows you to add conditions.

```
gen edcat = 0 if edyrs == 0  
tab edcat
```


Basic commands

if allows you to add conditions.

```
gen edcat = 0 if edyrs == 0
tab edcat
```

replace allows you to change the values of a variable

Replacing variables

```
replace edcat = 1 if edyrs < 5
replace edcat = 2 if edyrs >= 5 & edyrs < 10
replace edcat = 3 if edyrs >= 10
drop if edyrs < 0 | edyrs == .
tab edcat
tab edcat, m
```

Basic commands

Now that you have created variables, you can label them.

```
label var edcat "Category, years in education"
```

Basic commands

Now that you have created variables, you can label them.

```
label var edcat "Category, years in education"
```

```
label define education_cat 0 "No education" 1 "Less than 5y" 2 "5 to  
9 years" 3 "10y or more", replace
```

```
label values edcat education_cat  
tab edcat, m
```

Basic commands

```
gen index = _n  
gen count = _N  
bys clustnum: gen indcount = _N
```

Basic commands

rename to rename variables

One variable at a time:

```
rename edcat education_category
```

Several variables at a time:

```
rename (serv education_category) (service edcat)
```

Note that spaces are not allowed

Basic commands

`rename` to rename variables

One variable at a time:

```
rename edcat education_category
```

Several variables at a time:

```
rename (serv education_category) (service edcat)
```

Note that spaces are not allowed

Best practice: don't use upper case.

Try: `rename *, lower` (to make sure there are no upper cases in your variables). This will help also spot variables that have the same information but slightly different names

Basic commands

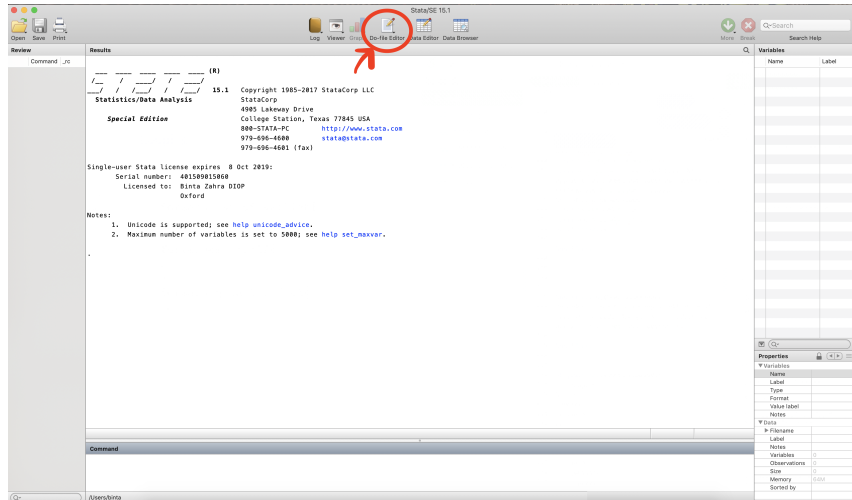
Now that you have your dataset. You can save it.

```
save "sawages_191011.dta", replace
```

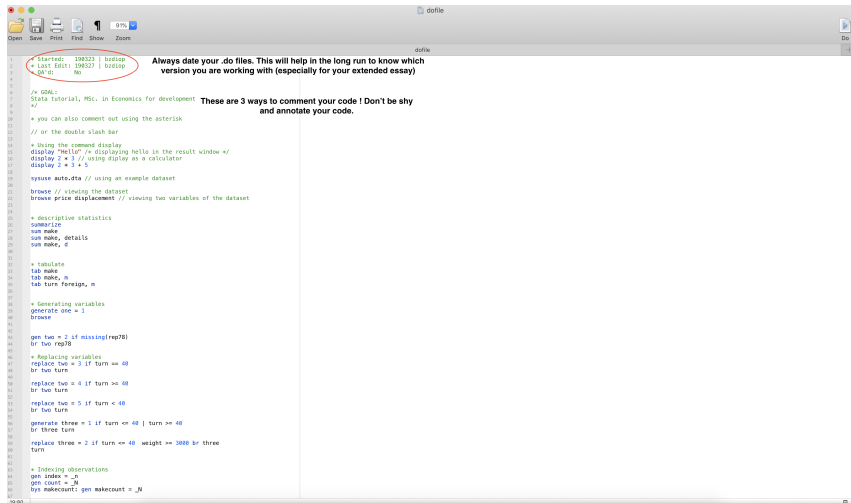
```
export excel using "sawages_191011.xlsx", firstrow(varlabels) replace
```

Best Practice

In order to not have to repeat all the work that you do each time you start working on Stata: use the do file editor.



Best Practices



The screenshot shows a Notepad++ window titled 'dofile'. The menu bar includes 'Open', 'Save', 'Print', 'Find', 'Show', and 'Zoom'. The status bar at the bottom shows 'Ln 92, Col 1'. The code is as follows:

```
1 * Started: 198323 | bzdlop
2 * Last Edit: 198327 | bzdlop
3 * Author: No
4
5
6 /* COME:
7 Stata tutorial, MSc. in Economics for development
8 */
9
10 * you can also comment out using the asterisk
11 // or the double slash bar
12
13 * Using the command display
14 display "Hello" /* displaying hello in the result window */
15 display 2 * 3 // using display as a calculator
16 display 2 * 3 + 5
17
18 sysuse auto.dta // using an example dataset
19
20 browse // viewing the dataset
21 browse price displacement // viewing two variables of the dataset
22
23
24
25 * descriptive statistics
26 summarize
27 sum make
28 sum make, details
29 sum make, d
30
31
32 * tabulate
33 tab make
34 tab make, m
35 tab turn foreign, m
36
37
38 * Generating variables
39 generate one = 1
40 browse
41
42
43 gen two = 2 if missing(rep78)
44 br two rep78
45
46 * Replacing variables
47 replace two = 3 if turn == 40
48 br two turn
49
50 replace two = 4 if turn >= 40
51 br two turn
52
53 replace two = 5 if turn < 40
54 br two turn
55
56 generate three = 1 if turn <= 40 | turn >= 40
57 br three turn
58
59 replace three = 2 if turn <= 40 weight >= 3000 br three
60 turn
61
62
63 * Indexing observations
64 gen index = _n
65 gen count = _N
66 bys makecount: gen makecount = _N
67
```

Annotations in the image include:

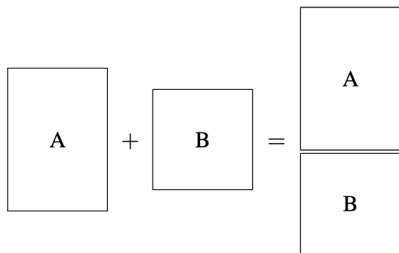
- A red circle around lines 1-3 with the text: "Always date your .do files. This will help in the long run to know which version you are working with (especially for your extended essay)".
- A green box around line 7 with the text: "These are 3 ways to comment your code ! Don't be shy and annotate your code."

Opening an excel file

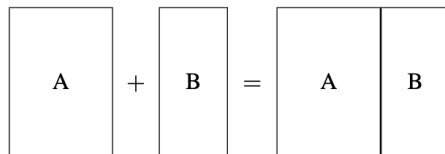
```
import excel "sawages_191011.xlsx", firstrow clear
```

Combining datasets

append



merge



Append: harmonize the names of your variables and their contents.

NOTE that you can type "help (h) [command]" to get more information for all commands on Stata

Try: help merge OR help append

Merging files

When you merge two datasets, make sure that both have at least one variable in common.

The common variable/s must have the same name.

`merge 1:1 index using "auto.dta", update replace`

also, `merge 1:m`, `merge m:1`, `merge m:m` (avoid the latter at all cost !!!)

Other useful commands

- ▶ `recode` for making multiple changes to a variable.
- ▶ `egen` offers a wide variety of useful functions. Example: `egen hhsize = count(pcode), by(hhid)`
- ▶ `encode` and `decode`
- ▶ `destring` to turn variables that appear to be number but coded as strings into strings and `tostring` to reverse it
- ▶ `keep` and `drop`
- ▶ `by` and `bysort`: `<command>`
- ▶ `levelsof`

Macros

```
foreach i of numlist 1/10 {  
  display 'i'  
}
```

Macros

```
foreach i of numlist 1/10 {  
  display 'i'  
}
```

```
forvalues i = 1/10 {  
  display 'i'  
}
```

Macros

```
foreach i of numlist 1/10 {  
  display 'i'  
}
```

```
forvalues i = 1/10 {  
  display 'i'  
}
```

```
foreach i in a b c d e f g {  
  display "'i'"  
}
```


Graphing

Histogram:

```
histogram edyrs
```

Graphing

Histogram:

```
histogram edyrs
```

```
scatter logwphy exp
```

```
scatter logwphy exp, by(edyrs)
```

BREAK

Stata Tutorial

Open folder.

Other helpful links

- ▶ Links to cheat sheets:
 - Data processing
 - Data analysis
 - Data visualization
- ▶ IPA stata training
- ▶ Tutorial (JPAL presentation)
- ▶ Best coding practices