Stata class notes (1)

About Stata

As Stata opens, you can see the Command line at the bottom of the window - that is where we are going to write (for now).

We are going to use the command line in class at least at first. Feel free to use the graphical interface instead (i.e. click around) or use do-files, but on the exam you will need to demonstrate that you can get the desired output either by writing down the right command, or telling me exactly where to click. The same is true for Excel. None of the students liked the clicking by the end of the course, so you may want to embrace the script.

The result of the command you entered (after hitting 'Enter') is going to show up on the output window above the Command window.

You can see the history of your commands to the left of the Command window (Review), and on the upper-right (Variables) you will see what variables your data set contains if you have something in the memory. In the lower-right little window (Properties) you can see the properties of a variable if you click on one.

A Stata command: basic syntax

A Stata line in today's class starts with a command. This is the main command that tells Stata what procedure you want to do. Then WITHOUT any commas, the arguments of the commands follow. These are the objects you want Stata to do something with. Finally, you put a comma, and that is going to signal Stata that the arguments are all enumerated, and now come the options of the command. Options modify the procedure the command does. All in all, the generic simple syntax for a command line looks like:

```
command argument1 argument2, option1 option2
```

In the GREAT Stata help, sometimes you see that the options or some of the arguments are in square brackets. Then those particular options/arguments are optional, meaning if you do not specify them, the command will still run. (On the other hand, if you do not specify the compulsory arguments/options, you will get a red error message.)

For example below, the arguments of summarize are optional

```
summarize [var_name1 varname2 …]
```

When we are feeding a file name (its path, that is) to a command, we usually need to use the word 'using' to give that type of argument.

Our Stata sessions

We will always start with the following steps, so it would be great if you could do this as fast as possible before/at the beginning of class. First, I would like you to define the working directory of your session. Stata is not going to look for your data set everywhere on your computer, and you would also like to collect your results in an ordered fashion.

1. For this reason, you need to start with downloading your data files right away after you connect the Remote Desktop Connection in the lab, and put it into your favorite directory where you also want to collect your results.
2. Then copy the path of that directory to Stata following the 'cd' command. For example, my directory is \\files\users\ptoth\Desktop\StataClasses. Then I write

```
cd \\files\users\ptoth\Desktop\StataClasses
```

3. The first 'real' thing you always do when starting the session is to start a log

```
log using my_first_log, t
```

   This line will create a text file (you can open it on any platform) called 'my_first_log.txt' in your folder. If you would like to rewrite an already existing log file, you need to add the option 'replace' to the end of the line above.

4. Then you open you data set. For example, I would like to open the catholic.dta data set. This is done by the 'use' command.

```
use catholic.dta
```

   Make sure you download data with the .dta extension, as that is the only extension you can 'Open' in Stata with this command (you need another command to 'Import' other extensions like csv, txt or xlsx).
5. You can browse your data set either by clicking on the 'browse' or the 'edit' icons on the menu bar of Stata, or putting the 'browse' command to the command line.
6. *Here comes the exercise. We will do stuff here……..*
7. If you need to open a new data set, because for example you are done with one exercise, you do that by using the 'replace' option with the 'use' command. Say, now I am done with C10 and would like to do C1 now. For that, I need to open the '401K.dta' file (already downloaded in the working directory).

```
use 401K.dta, replace
```

8. *We do stuff here too……*
9. Finally, never forget to CLOSE YOUR LOG! Otherwise—trouble. You do this by the log close command, so enter to the command line:

```
log close
```

Commands for today

See log using, use, log close above.

Name and syntax (for our purposes)	**Purpose**
`summarize [variable_list]`	To get summary statistics (mean, standard deviation, min and max) about the variables you put in 'variable_list' – without commas. If you do not specify variables, Stata calculates the descriptives for every variable in the data set.
`regress LHS_var RHS_var`	To run OLS (i.e. fit a line) – regress LHS_var on RHS_var