Handout 1

PO8006 - Andrea Salvi
Hilary Term 2018/2019

Reference: STATA Companion 1, 11

1 Admin Stuff

What's the purpose of this?

- Furthering what we did in class and explore its applications;

- Deepening the concepts of Quants and statistical modeling;

- Hands-on STATA;

- Real-world applications of theories, concepts and quantitative methods;

- Revision of Homework (Reminder: Homework 1 is due on the 8th of February but it will not require *STATA*);

- Q&A;

- **NB: tutorials do not replace the lectures and vice-versa!.**

2 Hi STATA!

2.1 Get to know the interface:

The interface is built upon 4 main panels:

- Command: Your console, useful for some commands.

- Results: Outputs

- Variables: Names and labels of the variables from the current data-set

- Review: History of previously executed commands.

  - Allows to reiterate commands by simply clicking on them.
  - You can sort commands by clicking on *Command* (decreasing, increasing).

  Other important features:

- Data Editor (browse) for a look at your data.

  - Let's check-it-out: type *use auto.dta* in the Command panel
  - Open the Data Editor (browse): from here we are able to do several things: view filters etc.
  - Now open Data Editor (Edit). What's the main difference with the former?
  - IMPORTANT: you must close the Data Editor and Data Browser before Stata will run any commands.
  - Now, let's get rid of this data. Type: "clear all" in the Command panel (guide on the "clear" command)

2.2 Ways to interact with STATA

You have several options here:

- **Pull-down menus**
- **Command Window**
- **Do files!**

Let's explore them using *educ99gdp.dta.* Type: *sysuse.dta.* That allows us to load built-in data-sets through the "Command Window". Alternatively, we can use the pull-down menus (File - Example Data-sets - Example data-sets installed with Stata)

1. Pull-down menus: You will see in the pull-down menus under Statistics, for example, a series of options for various types of statistical tests (in reality, they show just a fraction of what STATA can do!). Let's try something:

   - Click on "Statistics"
   - Let's calculate mean and variance of the percentage of GDP spent on private education.

2. We can do the same thing working interactively through the Command Panel.

   - As a first step, it is good practice to use the *"describe"* command;
   - Now type *"summarize"*. What do we get?
   - No give a try to *"summarize, detail"*
   - What if we want to summarize just the "private" variable? *"summarize private, detail"*

3. However, if you want to be a PRO, I would suggest to pursue another strategy: *Write a DO FILE!*.

   This is probably the best way to have a reproducible and clear code. Instead of just typing commands into the Command Panel or using the pull-down menus, it is likely that you will want to keep a record of what you did so that you can refer to them (and run them again) later. Let's have a look at the following Do file:

```
clear all
use auto.dta
summarize price , detail
scatter price mpg
```

2.3 More details on Do Files:

So far, we have been working with built-in data. What if you want to bring it to the next level?

Let's introduce the "Working Directory".

- To check your directory type: *pwd*;
- To change it: cd "directory path"
- You can do that using the UI as well (File - Change Working Directory).
- Whatever you save will end up in this folder.
- STATA will try to load stuff (e.g. data) from here.

Now are working environment is up and running. Let's create a new do file!

1. Use the menu bar to create a new do file
2. Save it!

Some basic syntax :

- If you can put a * before a line, STATA will not execute that line. Two main uses:
  - First, you can rerun your do-file while leaving out certain commands.
  - You can annotate your file as shown below

```
* This is a comment
clear all
use auto.dta
summarize price, detail
* summarize, detail
scatter price mpg
*Scatter plot with price and mileage
```

Other ways to write your comments:

- begin the line with *;

- begin the comment with // (useful at the end of a command)

- place the comment between /* and */ delimiters. (works within lines as well!)

From STATA help guide (`https://www.stata.com/manuals13/pcomments.pdf`):
"The comment indicator * may be used only at the beginning of a line, but it does have the advantage that it can be used interactively. * indicates that the line is to be ignored. The // comment indicator may be used at the beginning or at the end of a line. However, if the // indicator is at the end of a line, it must be preceded by one or more blanks. That is, you cannot type the following:

```
tabulate region// there are 4 regions in this dataset
```

// indicates that the rest of the line is to be ignored."

Delimit From the STATa help guide (`https://www.stata.com/manuals13/pdelimit.pdf`): "The delimit command resets the character that marks the end of a command. It can be used only in do-files".

The default would be the so-called "carriage return". However, for long lines of code that is not so practical. The command *#delimit ;* changes the delimiter to a semicolon. To restore the carriage return delimiter inside a file, use *#delimit cr.*

Logs File - Log - View: Opens and allows to save the logs with all the previous outputs (usually not necessary).

Help & Installing packages

- *search*: use it when you know what you want to do, but you need the specific command.

- *help*: use it when you know the command but do not remember how it works.

- scc install "package name": to install package (try textitssc dscribe a)

2.4 More on Data

As shown, there are several ways to interact with STATA and load new data:

1. File - Import

2. "Open" icon if you are using Stata files (.dta)

3. use filename.dta

What if you have an Excel file? Firstly, make sure that the cells type in excel is numeric (change it on excel). You have several options here. Let's explore them: `https://www.stata.com/support/faqs/data-management/converting-excel-files/`

"Copy and paste" is a bit funky with variable names and labels (you can amend them with Tools - Variable Manager.

Commands:

- Excel: import excel filename (or use File–¿import)

- CSV: import delimited using "filename.csv", varn(1) clear

- usespss using "pathandfilename", clear

- For other formats see page 6 of PDF Handbook.

- You can also insert data into the File Editor and then save as .dta

2.5 Describe Datasets:

Let's walk through an example: Open gss1012 It is probably better to simply set your wd and then type -use- and file name "use gss2012".

- describe: describe dataset that you have just opened.

- codebook varname-

- codebook

What does "." indicates?

Keeping or dropping variables If you want to keep or drop variables:

- keep varnames

- drop varnames