EC 220 - Class 1

Malka Guillot

Thursday 6th September, 2016

About me

- PhD student at the Paris School of Economics
- GTA at LSE
- Research interest: public economics
- Office Hours: Tuesday 11:30 12:30 (32L.1.30)
- E-mail: malkaguillot@gmail.com

Important Information

- Class attendance policy.
- Problem sets:
 - Give it back before Monday, 4pm.
 - I will randomly mark half of them every week.
 - Include your log files
 - By email (preferably) or at my pigeonhole (room 32L.1.01)
- Mock exam at the end of the term (week 11).
- Lent Term Examination in LT Week 0 (duration: 2 hours)

Why a statistical software

- Computation power
- Flexibility and user-friendliness (basic programming)
- Reproductibility of your work:
 - Wide academic diffusion
 - Minimal direct hand-editing of data
 - Comments

About Stata

- Software for statistical analysis
- Not free, but runs on Windows and Unix environment (Mac and Linux)
- Very general human capital
 - One of the most used softwares for Economic Research...
 - ... as well as in many sectors
 - A first experience with statistical software
 - \rightarrow Don't wait to pay that fixed cost!

Help in Stata

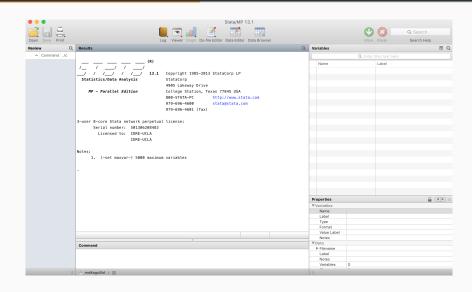
• At LSE:

Table 1: Stata support sessions weeks 3 & 4 - L32.G.18

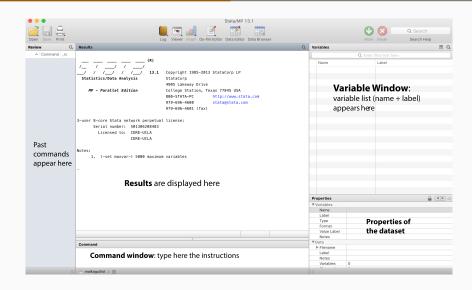
	Monday	Tuesday	Wednesday	Thursday	Friday
Week 3	16-17	14-15	15-16	10-11	12-13
Week 4	16-17	16-17	12-13	10-11	12-13

- Very complete help page accessible by internet or by the command window
 - Just type help mycomand

Stata interface



Stata interface



How do I primarily use Stata?

- Load a dataset in the Stata memory
- Run commands/dofiles to Stata
- Interpret, export, present results

Communicate with Stata:

Toolbar



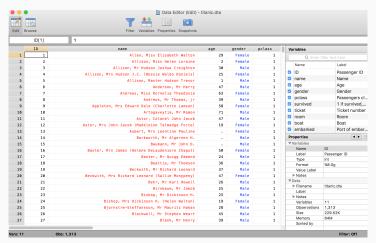
- Command window
- Dofile (next week)

First database

- Opening dataset
 - Go to Moodle Course page and download the dataset titanic, save it in a convenient directory
 - Open Stata: 🖽
 - Open the dataset in Stata format (.dta):
 - in the menu, find $File \rightarrow Open$
 - from the command bar: use "C:/mypath/EC220/data/titanic.dta", clear
- Variables appear in the variable window!

The titanic database

- Information on the passengers of the Titanic
- Each row = 1 individual
- Columns = variables



Have a look at your data

What information do we have?

- Browse = look at the data without being able to modify it → command: browse [var] = [△]
- Edit = allow you to change the data → command: edit [var] = Beware! danger!
- "describing" your data gives you (sometimes) useful info on your dataset (for instance, the types and labels of your variables)
 describe [vars]
- "codebook" allows you to get some more info on variables codebook [vars]

Key statistics

Some questions:

- How many passengers/observations?
- What is the gender composition of the boat?
- What is the average age of the passengers?

Some statistical commands:

- To get the distribution of values of a variable: tabulate myvariable
- To get multiple variable distributions: tabulate myvariable1 myvariable2
- Count the number of observations: count [if]
- Summary statistics: summarize [var], [detail]

Modifying data

- Creating new variables:
 generate newvar = EXPRESSION
 egenerate newvar = FUNCTION(other_variable)
- Modifying the contents of existing variables: replace var = other_var*2
- Dropping and keeping
 - variables: drop var1 var2 or keep var1 var2
 - observations: drop if sexe == "Female"

What to remember from today

- How to download and open datasets.
- Some general commands for describing the data.
- Some general commands for managing the data.
- Use the help files.
- Practice!

Next time

- Dofiles
- Logfiles
- Conditioning commands