

EC 220 - Class 1

Malka Guillot

Thursday 6th September, 2016

- 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)
- Reproducibility of your work:
 - Wide academic diffusion
 - Minimal direct hand-editing of data
 - Comments

- 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
- Don't wait to pay that fixed cost!

- 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

The screenshot displays the Stata/MP 13.1 software interface. The top menu bar includes options like Open, Save, Print, Log, Viewer, Graph, Do-file Editor, Data Editor, and Data Browser. The main window is divided into three panels: Review, Results, and Variables.

Review Panel: Shows the command `_rc` and the Stata logo.

Results Panel: Displays the Stata startup screen, including the Stata logo, version 13.1, and copyright information. It also shows the license type (MP - Parallel Edition) and the user's license details (3-user 8-core Stata network perpetual license).

Variables Panel: Contains a search bar and a table with columns for Name and Label. Below the table is a Properties section with tabs for Variables and Data. The Variables tab is currently selected, showing fields for Name, Label, Type, Format, Value Label, and Notes. The Data tab shows fields for Filename, Label, Notes, and Variables (set to 0).

Command Panel: Located at the bottom, it shows the command `_rc` and the user's name `malkaguillot`.

Stata interface

The screenshot displays the Stata/MP 13.1 software interface. The main window is divided into three panes: Command, Results, and Variables.

Command window: This pane at the bottom is where users enter instructions. It contains the text: **Command window: type here the instructions**.

Results window: This pane in the center displays the output of commands. It shows the Stata logo, version 13.1, copyright information (1985-2013 StataCorp LP), and license details (3-user 8-core Stata network perpetual license). It also includes a note about the maximum number of variables (5000). The text **Results are displayed here** is overlaid on this pane.

Variables window: This pane on the right lists the variables in the current dataset. It has a search bar and a table with columns for Name and Label. The text **Variable Window: variable list (name + label) appears here** is overlaid on this pane.

Properties window: This pane at the bottom right shows the properties of the selected variable. It includes a table with columns for Name, Label, Type, Format, Value Label, and Notes. The text **Properties of the dataset** is overlaid on this pane.

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)

- 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** → **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

Data Editor (Edit) - titanic.dta

Filter Variables Properties Snapshots

ID[1] 1

ID	name	age	gender	pclass
1	Allen, Miss Elisabeth Walton	29	Female	1
2	Allison, Miss Helen Loraine	2	Female	1
3	Allison, Mr Hudson Joshua Creighton	30	Male	1
4	Allison, Mrs Hudson J.C. (Bessie Waldo Daniels)	25	Female	1
5	Allison, Master Hudson Trevor	1	Male	1
6	Anderson, Mr Harry	47	Male	1
7	Andrews, Miss Kornelia Theodosia	63	Female	1
8	Andrews, Mr Thomas, jr	39	Male	1
9	Appleton, Mrs Edward Dale (Charlotte Lamson)	58	Female	1
10	Artagaveytia, Mr Ramon	71	Male	1
11	Astor, Colonel John Jacob	47	Male	1
12	Astor, Mrs John Jacob (Madeleine Talmadge Force)	19	Female	1
13	Aubert, Mrs Leontine Pauline	-	Female	1
14	Barkworth, Mr Algernon H.	-	Male	1
15	Baumann, Mr John D.	-	Male	1
16	Baxter, Mrs James (Helene DeLauniere Chaput)	54	Female	1
17	Baxter, Mr Quigg Edmond	24	Male	1
18	Beattie, Mr Thomson	36	Male	1
19	Beckwith, Mr Richard Leonard	37	Male	1
20	Beckwith, Mrs Richard Leonard (Sallie Monypeny)	47	Female	1
21	Behr, Mr Karl Howell	26	Male	1
22	Birnbaum, Mr Jakob	25	Male	1
23	Bishop, Mr Dickinson H.	25	Male	1
24	Bishop, Mrs Dickinson H. (Helen Walton)	19	Female	1
25	Bjornström-Steffansson, Mr Mauritz Hakan	28	Male	1
26	Blackwell, Mr Stephen Weart	45	Male	1
27	Blank, Mr Henry	39	Male	1

Vars: 11 Obs: 1,313 Filter: Off

Variables

Q Enter filter text here

Name	Label
<input checked="" type="checkbox"/> ID	Passenger ID
<input checked="" type="checkbox"/> name	Name
<input checked="" type="checkbox"/> age	Age
<input checked="" type="checkbox"/> gender	Gender
<input checked="" type="checkbox"/> pclass	Passengers cl...
<input checked="" type="checkbox"/> survived	1 if survived,...
<input checked="" type="checkbox"/> ticket	Ticket number
<input checked="" type="checkbox"/> room	Room
<input checked="" type="checkbox"/> boat	Boat
<input checked="" type="checkbox"/> embarked	Port of embar...

Properties

Variables

Name	ID
Label	Passenger ID
Type	Int
Format	%8.0g
Value Label	
Notes	

Data

Filename titanic.dta

Label

Notes

Variables 11

Observations 1,313



Size 229.52K

Memory 84M

Sorted by

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]`

- Creating new variables:

```
generate newvar = EXPRESSION
```

```
egenenerate 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