Introduction to Stata – Problem Set 2

Salvatore Viola
Universitat de Barcelona - MSc in Economics

Academic Year 2025/2026

Submission

Students are allowed to work in groups of up to 3 people. Before the beginning of the following class on **Friday, October 31st, at 10:00h**, each group must submit a single .do file containing the code and answers to the questions (commented out) to sviola@ub.edu in this format:

- `PS#_LASTNAME_LASTNAME_LASTNAME.do`

Submissions will be graded primarily based on the correctness of the responses, but partial credit will also be rewarded where students have made a noticeable effort to solve exercises using the methods covered in class and the available Stata documentation.

Question 1

1. Go to this website: `https://data.oecd.org/`
   - click on indicators - look through datasets - select one
   - scroll down and click access now
   - change some of the filters (if you want)
   - select only one country over multiple years or multiple countries in one year
2. Download the data as **FILTERED DATA IN TABULAR TEXT (CSV)**
3. Load the data in your .do file and try to understand the structure and different variables. Describe how the data looks and what the most comon variable type is.
4. Execute some descriptive commands (i.e., browse, sum, drop, keep, etc.). Can you find a command or method that will give you an example of each of the variables?
5. Trim the data to keep a specific set which removes the unnecessary variables from the original download.
6. Using your trimmed data, save a table of descriptive statistics with the -estout- command for only one year or country in you data. Include the number of observations, the mean, the standard deviation, the minimum and maximum values. The table should include a title, a note at the bottom and the column headers and variable names should be appropriately labeled (i.e., not just the variable names as they appear in Stata).
7. Save the data both as a .dta (Stata file format) AND as an excel table.

Question 2

Many Stata commands store some of their output in the background in variables called "macros" that you can reference. The example covered in class showed how the results of the summarize command can be found in "r()" macros using the "return list" command. Often times we want to reference metrics generated using estimation commands like "regress" in Stata. Unlike the "summarize" command, "regress" only stores some of its information in "r()" macros. Taking this into consideration, respond to the following questions:

1. Where are the results of the regress command stored apart from "r()" macros?
2. Where within this location are the coefficients of the estimation stored and in what sort of format?
3. Is this the only other location where Stata commands can store information relating to their output? What are the other ones and how do they differ from one another?
4. Aside from referencing specific values like in the example from class, in what other ways might these variables be used? List a few different ideas.
   (Hint: think back to the previous problem set)