Stata - Computer Tools Course

Structure

The course will be taught for two weeks with 2 hours a day in the afternoons. There will be 2 parallel groups in order to enable each of you to follow the course material on a computer and do some exercises in the classroom on your own.

Aim

The aim of this course is to make you familiar with using Stata. We will teach in a very applied way using real-world datasets. In the end, you should be able to start doing analyses with Stata in the following "Statistics and Econometrics" course from day 1. We will NOT teach any econometrics in these two weeks, since this will be the subject of your regular courses in the following trimesters.

Exam

At the end of the course, we will give you a short in-class exam and discuss the solutions together.

Contents

1. Introduction to STATA

- Working with Stata: menu vs. command line vs. do files
- Help files, online PDF documentation since Stata 11
- Creating empty datasets and copy/pasting data
- Data import: different ways of importing data
- Describing the data
  - o Describe
  - o Sum
  - o Tabulate

2. Data sources

- Import data from main public data sources: World Bank (WDI), Penn Tables, Eurostat, ECB, …
- Missing values: "." vs. "99"

3. Data manipulation
- Generating new variables. "Generate" vs. "Egen".
- Dropping variables
- Sorting
- Recode, group
- Labelling variables and values
- Logical expressions

4. Basis statistical routines
- Mean, standard deviation, correlation
- Percentiles
- (t-)Test on mean difference. Compare groups within one variable, compare two variables.
- Cross-tabulation of two binary variables and corresponding tests (Pearson)
- Cross-tabulation of two discrete variables and corresponding tests (Pearson)
- OLS with one explanatory variable
- Internal variables: _coef, _se
- More stored information: "Ereturn list", "matrix list e(vce)"
- Postestimation commands.

5. Programming in do files
- If condition
- Loops
- Commenting

6. Graphing (here menu can be useful)
- Line plot. Legend, labels, shapes, colors, …
- Scatter plot
- Combining graphs: "twoway", e.g. scatter with regression line
- Histogram
- Kernel density, intuitive discussion of bandwidth
- Step function for cdf

7. Panel data
- Data structure: Wide vs. long
- Reshape
- Xtset
- Xtdes

8. Time Series data
- Tsset
- Lag and forward operator
- First difference and dlog

9. Presenting results