SPS Seminar 1st term 2019-2020

Introduction to Data Analysis and STATA for Beginners

Instructor: Mauricio Bucca, Max Weber Fellow

Register online

Contact: Monika.Rzemieniecka@eui.eu

Description

If this is the first time for you studying quantitative methods in social sciences, or need to refresh what you have learn in the past, you might find this course very useful. It is a catch-up preparatory course for the compulsory seminar "Introduction to Quantitative Methods" starting in October. It will provide a basic introduction into statistics and specially Stata, the statistical software package you will be using in the compulsory course. If you already know about statistics but have never used Stata, this course will also be useful for you, since the focus will be on learning Stata.

Newcomers to quantitative methods are strongly encouraged to join the course. Attending this introductory course will allow them to face the compulsory seminar more comfortably. Note that *no credits* will be awarded for attending this seminar.

Please, bring your own laptop to the class, with Stata installed on it, because the course will consist mainly of examples using Stata! To install *STATA* on your laptop, please, visit one of the ICT User Support Offices around the EUI.

Schedule:

The seminar takes place from the 16-20 September 2019 from 13:00-16:00, Seminar Room 2, Badia Fiesolana.

Content:

* Indicates content that will be covered if time permits.

Class 1: A glimpse on data analysis and the Stata environment

- Statistics and how it relates to data: the idea of estimation
- Data structures: standard datasets and how to use them
- Working with commands
- Code: Using do-files, automatization and replicability

- Workflow, good coding practices: Setting up a good do-file template
- A very simple first exercise

Class 2: Data management

- Types of data
- Basic commands: gen, egen, tab, sum, keep, drop, etc.
- Logic operators: if, or, and
- Creating new variables from existing variables: generate, replace, recode, etc.
- Labels: var labels, value labels, etc.
- Dealing with missing values

Class 3: Describing data - Descriptive statistics

- Computing basic statistics: mean, sd, mode, quantiles, etc. (continuous variables)
- Tables for marginal distributions (discrete variables)
* Displaying and using floating outputs
- Basic descriptive plots: bars, pie-charts, histograms, density plots.

Class 4: Describing data - Bivariate association

- Bivariate association between continuous variables: scatter-plots and correlation
- Bivariate association between discrete variables: plots, contingency tables and chi-2 test
* Bivariate association with ordered variables
- Bivariate association between continuous and discrete variables: differences in mean

Class 5: More on data management and statistics

- A last take on workflow and replicability
- A first introduction to inference: standard errors, p-values and statistical testing.
* Merge and append datasets
* Some useful tricks: e.g. “set more off”, “quietly”, “capture”, etc.
* Setting new observations and creating random variables
* Loops: just a demonstration of what can be done.

Recommended reading:

- Statistics without Tears: An Introduction for Non-Mathematicians (Penguin Science), Penguin UK; New Ed edition (May 15, 2000)

Consulting material:

- Stata tutorial: https://tutorials.iq.harvard.edu/Stata/StataIntro/StataIntro.html