



## ***Elementary Data Analysis with STATA***

F. Bernardi & M. Franklin

Monday 15:00 - 17:00 Room 2, Badia Fiesolana

Please register with: [Monika.Rzemieniecka@eui.eu](mailto:Monika.Rzemieniecka@eui.eu)

The aim of this seminar is to give researchers in sociology and political sciences practical knowledge of the application of basic quantitative techniques, which are commonly used in sociology and political sciences, while using one of the major statistical software packages STATA. This introductory course is open to all researchers from all departments, whatever their level of experience with quantitative techniques and STATA.

This seminar has an introductory nature because these techniques should have been taught at a pre-graduate level. Given the diverse methodological backgrounds of the EUI researchers and the different academic cultures within the European universities, such an introductory seminar is necessary for those who want to follow the advanced quantitative seminar *The application of advanced quantitative techniques with STATA*, given in the second term of each academic year.

The seminar centers on the linear regression model. A correct understanding of its proper functioning and limitations is an essential precondition for venturing into more advanced techniques. In this seminar, therefore, you can expect to learn how to perform a linear regression analysis from A to Z. You will also learn some basic notions of statistical inference, some very basic notions of estimation procedures (ordinary least squares and maximum likelihood) and some hints on useful topics in applied research such as how to deal with missing cases and how to account for sample design. The seminar is much focused on real research examples and formalization is kept to a minimum.

As reference book we use '*Quantitative Data Analysis. Doing Research to Test Ideas*', by D. Treiman, Jossey Bass, 2009. The EUI library has a number of copies of this book on the seminar shelf. There is a useful website from UCLA where manuals and annotated output are available for many different statistical models. The explanations are very clear and they take you through the output that STATA provides. <http://www.ats.ucla.edu/stat/stata/dae/>



Those who have a laptop should take it to the seminar because part of the teaching will consist of interactive examples using STATA in the class. At the moment we are trying to sort out how to provide access to STATA for each participant in the seminar. Those who register for the seminar will be informed on this issue.

The most important part of the seminar is not attending the two-hour sessions, but doing the exercises. Doing all exercises is the condition for certification of sufficient attendance. These exercises have to be sent in 24 hours before each session (thus before Sunday 15.00 PM) to the mail-address of prof. Bernardi ([fabrizio.bernardi@eui.eu](mailto:fabrizio.bernardi@eui.eu)) and prof. Franklin ([mark.franklin@eui.eu](mailto:mark.franklin@eui.eu)). Exercises which arrived in time are given back at the start of each meeting with answers, remarks, etc. A successful completion of the course will be judged on the basis of quality of the sent-in exercises. The last assignment is due for the 20<sup>th</sup> of December. It consists of a small paper (4 pages plus tables) where you should carry out a multivariate analysis and interpret the results, using the skills you have learned through-out the seminar. For this final exercise participants are encouraged to use their own data (not necessarily the data they will use in their theses). Researchers without applicable data for this final exercise can use data from the *European Social Survey* (see for more information: <http://www.europeansocialsurvey.org/>).

A special *dropbox* folder will be created where you will find the weekly exercises and the materials used in class. The teacher assistant for this seminar is Pedro Riera, a second year SPS researcher. He will provide assistance with STATA and with the exercises.

More details on seminar practicalities and the final assignment will be given during the first introductory class. Registration before October 4 is obligatory for this seminar. The maximum number of participants is 20 and admission is based on the order of registration.

## Schedule

WEEK 1 (Oct 4<sup>th</sup>) Introduction and overview of seminar. Organizational and practical issues. Descriptive statistics.

**Readings:** Treiman: Chapter 4 *On the manipulation of data by computer*

WEEK 2 (Oct 11<sup>th</sup>) Cross-tabulation. How to make and comment properly a table. Control variable.

**Readings:** Treiman: Chapters 1, 2 and 3



WEEK 3 (Oct 18<sup>th</sup>) Bivariate regression and correlation. The logic of OLS estimation.

**Readings:** Treiman: Chapter 5

WEEK 4 (Oct 25<sup>th</sup>) Basic of statistical inference. Null hypothesis, type I and type II errors. Confidence intervals.

**Readings:** TBA Material will be placed in the seminar boxdrop folder

WEEK 5 (Nov 8<sup>th</sup>) Multiple regression I. The logic of intervening variable (mediating and suppressing variable). Categorical independent variables

**Readings:** Treiman: Chapter 6

WEEK 6 (Nov 15<sup>th</sup>) Multiple regression II. Non linear relationships. Log transformation. Basic introduction to maximum likelihood estimation.

**Readings:** Treiman: Chapter 7

WEEK 7 (Nov 22<sup>nd</sup>) Regression diagnostics.

**Readings:** Treiman: Chapter 10

WEEK 8 (Nov 29<sup>th</sup>) Interactions and limits of the regression model

**Readings:** Brambor, Clark and Golder (2006) *Political Analysis* on interaction effects

[http://homepages.nyu.edu/~mrg217/pa\\_final.pdf](http://homepages.nyu.edu/~mrg217/pa_final.pdf)

WEEK 9 (Dec 6<sup>th</sup>) Students's presentations on their proposal for the analysis for the final assignment.

WEEK 10 (Dec 13<sup>th</sup>) Miscellaneous. Missing data. Sample design and survey estimation. To weight or not to weight?

**Readings:** Treiman: Chapters 8 and 9