



## Introduction to Quantitative Methods

Organised by Fabrizio Bernardi

Monday, 11:00 - 13:00

Seminar Room 2, Badia Fiesolana

Please register [online](#)

(Contact: [Monika.Rzemieniecka@eui.eu](mailto:Monika.Rzemieniecka@eui.eu))

### Introduction and Aim

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. The seminar centers on the linear regression model and follows the approach proposed in Berk (2004). A correct understanding of its proper functioning and limitations is an essential precondition for venturing into more advanced techniques. We will discuss the (mis)use of statistical significance testing in current research in sociology and political science and new developments such as the type M and type S errors (Gelman and Carlin 2014). We will also see how to estimate and correctly interpret interactions. We will discuss the logistic regression model and show how and when the linear probability model can work well with dichotomous dependent variables. Finally, we will sketch the insights that might be gained using quantile regression.

Starting from class 6, the class will be divided in two groups. The advanced group (group A) will go through the full program up to the logistic regression and quantile regression. The other group (group B) will spend more classes on additional practical exercises on data management and multiple regression. See the schedules below.

### Basic references

1. Berk, R. (2004). *Regression Analysis: A Constructive Critique*. Thousand Oaks, CA: Sage Publications.
2. Treiman, D. (2009). *Quantitative Data Analysis. Doing Research to Test Ideas*, Jossey Bass.
3. Firebaugh, G. (2008). *Seven Rules for Social Research*, Princeton, NJ: Princeton University Press.

The EUI library has a number of copies of 2. and 3. on the seminar shelf. The Berk's book has been ordered.

### Organisation

You should bring your laptop to the seminar because part of the teaching will consist of interactive examples using STATA in the class. The most important part of the seminar is not attending the two-hour sessions, but making the exercises that are assigned each week. A special *dropbox* folder will be created where you will find the weekly exercises and the materials used in class. The teacher assistants for this seminar are Lea Kröger ([Lea.Kroger@EUI.eu](mailto:Lea.Kroger@EUI.eu)) and Reto Bürgisser ([Reto.Burgisser@EUI.eu](mailto:Reto.Burgisser@EUI.eu)), both second

year SPS researchers. Participants will be divided in two groups and the TAs will provide feedback on your exercises. The TAs will also have office hours to provide assistance with STATA and with the exercises (again time and place tbc).

### **Credits**

Complying with the attendance requirements and making all exercises is the condition for being awarded the seminar credits. These exercises have to be posted in a dedicated folder of the seminar dropbox 24 hours before each session (thus before Sunday 11.00 AM). You can team in groups of **two** for making the exercises, ideally an "advanced" and a "beginner" working together. The last exercise is due for the 14<sup>th</sup> of December. This last exercise consists of a small paper (4 pages plus tables) where **researchers of group B** should carry out a simple multivariate analysis and interpret the results. **Researchers of group A** should carry out a more complex analysis that should include one of the following features: the specification of an interaction, the estimation of a linear probability model or of a quantile regression. For this final exercise participants are encouraged to use their own data (not necessarily the data they will use in their thesis). Researchers without applicable data have to contact me and the TAs in order to find proper data for this exercise. If you want to write a **term paper** for this seminar, you should replicate a published article and carry out an informed benchmarking of the results and/or estimate the type S and type M errors associated to a given coefficient (see class 6 for group A). More details on this and on other seminar practicalities will be given during the first introductory class.

### **Schedule**

CLASS 1 (Oct 5): Introduction and overview of seminar. Process of elaboration: spurious, intervening and conditional relationship.

**Readings:** Treiman: Chapters 1, 2, 3 and 4

CLASS 2 (Oct 12): Bivariate regression and correlation. The logic of OLS estimation.

How to comment your findings.

**Readings:**

Berk: Chapter 2

Treiman: Chapter 5

Firebaugh: The Second Rule. Look for Differences That Make a Difference, and Report Them (Chapter 2)

More technical: Berk: Chapter 3

CLASS 3 (Oct 19): Problems related to linear regression. Outliers, truncation, regression fallacy, ecological fallacy. Basic of statistical inference.

**Readings:** Treiman: Chapter 5

Chapter 3 from Bohrnstedt and Knoke, *Statistics for social data analysis*, in the dropbox.

CLASS 4 (Oct 26): Statistical inference. Null hypothesis, type I and type II errors. Confidence intervals.

**Readings:**

Chapters 16, 26 and 29 from Freedman, Pisani and Purves, *Statistics*, in the dropbox.

CLASS 5 (Nov 2<sup>th</sup>): Multiple regression.

**Readings:** Treiman: Chapter 6

Additional readings

Young, C. 2009. "Model Uncertainty in Sociological Research: An Application to Religion and Economic Growth." *American Sociological Review*. 74:380–97.

<http://asr.sagepub.com/content/74/3/380.full.pdf+html>

Special issue of *Conflict management and peace studies*, V. 22, 4, 2005

## GROUP A (seminar room 2)

CLASS 6 (Nov 9<sup>th</sup>): Substantive versus statistical significance.

**Readings:**

Bernardi, F., Chakhaia, L. and Leopold, L. (2015), "*Sing Me a Song With Social Significance*": *The (mis)use of significance testing in European sociological research*, unpublished manuscript.

Gelman, A. and Carlin, J. (2014). "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science*:1-11.

Gelman, A. and Weakliem, D. (2009), "Of Beauty, Sex and Power. Too little attention has been paid to the statistical challenges in estimating small effects." *American Scientist* 97:310-316.

CLASS 7 (Nov 16<sup>nd</sup>): Interactions

**Readings:**

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

[http://mattgolder.com/files/research/pa\\_final.pdf](http://mattgolder.com/files/research/pa_final.pdf)

Macdonald, K. (2011), "Family Investments in Children: What the Interactions and the Data Do Not Say." *European Sociological Review*.

<http://esr.oxfordjournals.org/content/early/2011/02/26/esr.jcr021>

CLASS 8 (Nov 23<sup>rd</sup>): More examples on interactions. With TAs

CLASS 9 (Nov 27<sup>th</sup>)\* Introduction to logistic regressions. Comparison with the linear probability model. Examples.

**Readings:**

Mood, C. (2010). "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It." *European Sociological Review* 26:67-82.

<http://esr.oxfordjournals.org/content/26/1/67.abstract>

CLASS 10 (Dec 4<sup>th</sup>)\* Introduction to quantile regression. Regression diagnostics.

**Readings:** Material to be distributed

\* Note change in schedule: Friday instead of Monday

### **GROUP B (Seminar Room 3)**

CLASS 6 (Nov 9<sup>th</sup>): Additional examples on data management.

CLASS 7 (Nov 16<sup>nd</sup>): Additional examples on multiple regressions.

CLASS 8 (Nov 23<sup>rd</sup>): Interactions

#### **Readings:**

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

[http://mattgolder.com/files/research/pa\\_final.pdf](http://mattgolder.com/files/research/pa_final.pdf)

Macdonald, Kenneth I. 2011. "Family Investments in Children: What the Interactions and the Data Do Not Say." *European Sociological Review*.

<http://esr.oxfordjournals.org/content/early/2011/02/26/esr.jcr021>

CLASS 9 (Nov 27<sup>th</sup>)\*: More examples on interactions.

CLASS 10 (Dec 4<sup>th</sup>)\* More examples on identified problems. Or Substantive versus statistical significance

#### **Readings:**

Bernardi, F., Chakhaia, L. and Leopold, L. (2015), "*Sing Me a Song With Social Significance*": *The (mis)use of significance testing in European sociological research*, unpublished manuscript.

Gelman, A. and Carlin, J. (2014). "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science*:1-11.

Gelman, A. and Weakliem, D. (2009), "Of Beauty, Sex and Power. Too little attention has been paid to the statistical challenges in estimating small effects." *American Scientist* 97:310-316.

\* Note change in schedule: Friday instead of Monday