

Introduction to Quantitative Methods

Instructors:

Elias Dinas (elias.dinas@eui.eu), Political Science
Juho Härkönen (juho.harkonen@eui.eu), Sociology

Teaching assistants:

Melanie Sauter (Melanie.Sauter@EUI.eu)
Vicente Valentim (Vicente.Valentim@EUI.eu)

Compulsory Course for all SPS 1st year researchers. No registration required.

Course summary

This course introduces students to the fundamentals of statistical analysis for the social sciences. We will cover the basics, starting from how we can use statistics to summarize information and describe general patterns of interest to how we can implement predictions or arrive to causal claims.

The topics covered are discussed in numerous course books, and below we list the main ones used in this course. For each class, we point to the main chapters that cover each week's topics. There will be general (though not always perfect) overlap between the books. Furthermore, we generally refer to the latest edition, and there will of course be overlap with earlier editions in case the latest one is unavailable.

Some of you are relatively new to statistics (we absolutely recommend taking the “Daily preparatory courses on STATA and data analysis”, given in September). Others have more familiarity through previous courses and own work. While the whole course is compulsory, we offer “additional topics” for those who may be familiar with the core topics of each class.

The lecture (on Mondays) is only one part of the course. At the end of each lecture, we will set up an assignment, which will imply applying what we learned with real data. You will have approximately 36 hours to work on this problem. You can collaborate and you would collaborate even if you could not. By the end of this time you have a lab session, in which you will go through the assignment. You will be asked to provide solutions to the problem set and you will discuss pros and cons of the different approaches.

The lab session will be in Stata, but you can also work with R if you prefer so (see

below). The lab sessions will take place on Thursdays and are given by Melanie Sauter and Vicente Valentim.

Course schedule and contents:

Week 1

Lecture: Monday, 1 Oct, 11:00-13:00, Seminar room 2

LAB (Introduction to R): Wednesday, 3 October, 14:00-17:00, Sala del Capitolo

Asking Questions: Brief overview of the course and discussion of how quantitative methods are useful in answering problems in the social sciences. Types of questions (descriptive, predictive, causal). Types of data (e.g., randomized, observational, survey, macro) and types of variables (and their validity).

Readings: Imai Ch. 2; Agresti Ch. 1; Angrist and Pischke Ch. 1; Gordon Ch. 1.

Week 2

Lecture: Monday, 8 Oct, 11:00-13:00, Seminar room 2

Lab: Wednesday, 10 Oct, 11:00-13:00, Seminar room 3

Measurement 1: The first part of the lecture provides a quick recap of measures of central tendency and measures of dispersion. The second part covers random sampling and sampling error; standard errors and confidence intervals (means and proportions). Finally, we discuss non-random samples and non-random sampling (cluster and stratified sampling), and the use of weights.

Readings: Imai, Ch. 3, 7.1; Agresti, Ch. 2–3, 5.1–5.3; Treiman, Ch. 9.

Additional topics: Calculating weights and assessing sampling bias

Week 3

Lecture: Monday, 15 Oct, 11:00-13:00, Seminar room 2

Lab: Thursday, 18 Oct, 11:00-13:00, Sala del Capitolo

Inference 1: Hypothesis testing: Null and alternative hypotheses; p-values; type I and type II errors. Descriptive analysis of bivariate relations (differences of means, correlations, cross-tabulations).

Readings: Imai, Ch. 7.2; Agresti, Ch. 5.4–7.3; Treiman, Ch. 1.

Additional topics: Power and false positives

Week 4

Lecture: Monday, 22 Oct, 11:00-13:00, Seminar room 2

Lab: Thursday, 25 Oct, 11:00-13:00, Seminar room 2

Regression 1: Bivariate linear regression: short explanation of the idea of

dependent and independent variables. Move from scatterplot and fitting a line “by eye” to discussion of the line equation (e.g. the slope and intercept) to OLS regression (e.g. least squares criterion and the error term). Regression Anatomy: TSS, ESS, RSS and goodness of fit. Interpretation of regression coefficients. Inference for bivariate regression.

Readings: Imai, Ch. 4; Agresti, Ch. 9; Wooldridge Ch. 2 & 3; Field et al., Ch. 7.1–7.5; Dalgaard, Ch. 6; Gordon, Ch. 5; Treiman, Ch. 5.

Additional topics: Calculation of correlation and regression coefficients

Week 5

Lecture: Monday, 29 Oct, 11:00-13:00, Sala del Capitolo

Lab: Wednesday, 31 Oct, 15:00-17:00, Seminar room 4

Regression 2: Multiple regression. Why include more than one independent variable in a regression model. Confounding and mediating variables. How to interpret regression coefficients in the multiple regression model. The *ceteris paribus* criterion. Inference in the multiple regression case.

Readings: Agresti, Ch. 10–11; Wooldridge, Ch. 2 & 3; Field et al., Ch. 7.6–7.8, 7.11–7.12; Dalgaard, Ch. 11; Imai 4.3. Gordon, Ch. 6, 7; Treiman, Ch. 6.

Additional topics: Elaboration, standardization, and decomposition

Week 6

Lecture: Monday, 5 Nov, 11:00-13:00, Seminar room 2

Lab: Thursday, 8 Nov, 11:00-13:00, Seminar room 3

Inference 2: This lecture considers uncertainty in the linear regression model. It covers: bias; the Gauss-Markov assumptions (What are the OLS assumptions and how can we make sense of them?); re-assessment of OLS inference; robust standard errors.

Readings: Imai, Ch. 7.3; Agresti, Ch. 13.2, 14.1–14.3; Wooldridge, Ch. 4 & 5; Field et al. Ch. 7.7, 7.9; Dalgaard Ch. 11; Gordon, Ch. 11; Angrist and Pischke Ch. 8.

Additional topics: Calculating bias

Week 7

Lecture: Monday, 12 Nov, 11:00-13:00, Seminar room 2

Lab: Thursday, 15 Nov, 11:00-13:00, Seminar room 2

More regression topics: Variable transformations (log transformations and polynomials) and categorical variables, and their interpretation. Reliability of measurement, scales, and measurement error.

Readings: Agresti, Ch. 12.1, 13.1–13.2, 14.4–14.6; Gordon, Ch. 7, 9; Treiman, Ch. 7, 11.

Additional topics: More on measurement and measurement error

Week 8

Lecture: Monday, 19 Nov, 11:00-13:00, Seminar room 2

Lab: Thursday, 22 Nov, 11:00-13:00, Seminar room 3

Regression & moderating effects: How can regression unveil moderating effects. Distinguishing between moderating and mediating effects. Interactions using both binary and continuous moderators. Interpretation of the main effects and the interaction term. Interpretation.

Additional topics: Higher-order interactions and mediation

Readings: Agresti, Ch. 11.4; Gordon, Ch. 8, 10.

Week 9

Lecture: Monday, 26 Nov, 11:00-13:00, Seminar room 2

Lab: Thursday, 29 Nov, 11:00-13:00, Seminar room 3

Extensions 1: Non-Parametric and Quintile Regression : Local linear regression, kernel regression in practice. When use it and what is it good for? Predicting quintiles of the distribution. The logic and practice of quintile regression. When should we use it?

Readings: Keele, Ch. 1, 2. Angrist and Pischke, Ch. 7.

Week 10

Lecture: Monday, 3 Dec, 11:00-13:00, Seminar room 2

Lab: Thursday, 6 Dec, 11:00-13:00, Seminar room 2

Extensions 2: Categorical Choice Models: Binary choice models and categorical choice models in more general. Logistic regression.

Readings: Agresti, Ch. 15; Field et al., Ch. 8; Treiman, Ch. 13–14

Additional topics: Comparing and interpreting estimates in logistic and probit models.

Week 11

Recap: What is Next: Questions and clarifications. What is next and why you should care about it.

Lecture: Monday, 10 Dec, 11:00-13:00, Seminar room 2

More on Stata and R

The clinic session in September is intended to provide you with info and first hands-on experience on Stata. Below you can find some useful information about R:

- To download R, go to: <https://cran.rstudio.com/>
- Many people like R-Studio as a way of managing your work in R. Like R, the

basic version of R-Studio is free.

- You can download it here.
- We suggest that you take a look at these websites, where you will find a number of tutorials:
 - Try R : <http://tryr.codeschool.com/levels/1/challenges/1>
 - swirl : <http://swirlstats.com>
 - Jared Knowles R bootcamp : <https://www.jaredknowles.com/r-bootcamp/>

Online and in-person interaction

Lecture and lab materials (code, worksheets and data) will be distributed via Piazza (<https://piazza.com>) as well as via the instructors' website (exact address and password will be provided in the lab sessions).

Lab instructors will hold regular office hours. Both Elias and Juho will be available during office hours and upon request via email.

Assessment

One assignment per week; you will work on this assignment between lecture and lab and you will discuss solutions in the lab. You will have send your collective or individual work to the lab instructors at least four hour before the lab.

Assessment for this course is based on the evaluation of your work.

Term paper

SPS participants who want to write a term paper for the seminar have to send an outline of about 500 words before 7 December 2018. The full paper (5000 words) has to be submitted by 18 January 2019 to Monika.Rzemieniecka@eui.eu.

Main textbooks

1. Gordon, Rachel (2015). Regression Analysis for the Social Sciences, 2nd Edition. Routledge.
2. Agresti, Alan (2018). Statistical Methods for the Social Sciences, Fifth edition. Pearson.
3. Imai, Kosuke (2017). Quantitative Social Science: An Introduction. Princeton.
4. Wooldridge, Jeffrey M (2015). Introductory econometrics: A modern approach, 4-6th Edition. Nelson Education.
5. Andy Field, Z. F., and Jeremy Miles. (2012). Discovering statistics using R. Sage.
6. Dalgaard, Peter. (2008). Introductory statistics with R. Springer Science & Business Media.
7. Angrist, Joshua and Jörn-Steffen Pischke. (2009). "Mostly Harmless Econometrics". Princeton
8. Keele, Luke. (2008). Semiparametric Regression for the Social Sciences. Wiley.
9. Treiman, Donald J. (2009). Quantitative Data Analysis: Doing Social Research to Test Ideas. Jossey-Bass.