Introduction to Quantitative Analysis*
MA Course, Winter 2017

Thursdays 10:00 − 13:00, from October 12 to November 30.
Room 3 of the RRZK.

Instructor:
Bruno Castanho Silva
Cologne Center for Comparative Politics (CCCP)
Email: paula-castanho_bruno@phd.ceu.edu
Office hours: TBD
Office: TBD

Course Description

This course is open to MA and PhD students and provides an introduction to quantitative analysis in political science using R. Topics include the essential elements of sampling, statistical inference, the estimation and interpretation of the linear regression model and data visualization. In the last weeks students will have brief introductions to more advanced topics, including time-series and panel data analysis, multilevel models, and predictive modeling (aka machine learning). Empirical applications will focus primarily on political science. Throughout the course, students will become familiar with the statistical programming language R. The goal of this course is to prepare students to critically interpret and conduct quantitative studies in political science.

---

*This is a preliminary syllabus, subject to minor changes before the course starts.

Prerequisites

Students should have a basic understanding of descriptive statistics in social sciences. No prior knowledge of R is required.

Software

The open-source statistical programming language R is particularly suited for carrying out state-of-the-art computer-based simulations and data exercises. It also generates really nice publication-quality graphics. The software runs under a wide array of operating systems. R can be downloaded for free at http://www.r-project.org/http://www.r-project.org/. Learning R might seem a bit challenging at first, but you will realize that it is incredibly powerful. The lab sessions will be devoted to learning data management and analysis techniques using R.

Students are also strongly encouraged to install RStudio. This is a more user-friendly interface for R with integration to other packages we will use throughout the course. RStudio is available for free at https://www.rstudio.com/.

NB! Installing R can be problematic sometimes. Students must have installed R (and RStudio) before the first class. If you are running into errors when installing, contact the instructor before the first session.

Course Materials

The course is structured as an applied introduction to quantitative analysis with R. Students can choose to do their mandatory readings from one of two sources provided by the instructor:

Gravetter, Frederick J., and Larry B. Wallnau (2008). *Essentials of Statistics for the Behavioral Sciences.* 6th edition. Belmont: Wadsworth.

Diez, David, Christopher Barr, and Mine Cetinkaya-Rundel (2015). *OpenIntro Statistics*, 3rd edition. Available at https://www.openintro.org.

The following materials provide alternative coverage for topics in this course:

Fox, John (2016). *Applied Regression Analysis and Generalized Linear Models*, 3rd edition. Los Angeles: Sage.

Fox, John and Harvey S. Weisberg (2011). *An R Companion to Applied Regression*, 2nd edition. Thousand Oaks: Sage.

Last updated: July 24, 2017

Course Requirements

This course will have a portfolio evaluation. It will be comprised of six data analysis projects during the semester. The first five papers should be around 3 pages, while the last around 5 pages. For the first five, students will have a week to complete the assignment, and two weeks for the last. You can discuss it with other students, but each one must write and submit their own individual answers. Two (nearly) identical assignments will receive a 0. If you work with someone, please indicate their names in your homework. Assignments must be handed in by email with the following deadlines and weights towards the final grade:

Homework	Posted	Deadline	Weight
Assignment 1	October 17	October 24, 23:55 CET	15pts
Assignment 2	October 31	November 06, 23:55 CET	15pts
Assignment 3	November 06	November 13, 23:55 CET	15pts
Assignment 4	November 13	November 20, 23:55 CET	15pts
Assignment 5	November 20	November 27, 23:55 CET	15pts
Assignment 6	November 27	December 10, 23:55 CET	25pts

Points are converted to final grades as follows:

Points	Grade
100–95	1,0
94–90	1,3
89–85	1,7
84–80	2,0
79–75	2,3
74–70	2,7
69–65	3,0
64–60	3,3
59–55	3,7
54–50	4,0
49–0	5,0

Last, the projects submitted must be written with R Markdown (see below).

Last updated: July 24, 2017

A Note on Professional Presentation

RStudio comes with a powerful authoring format called R Markdown. R Markdown documents look like a mix of a text document and R code. They enable easy creation of data analysis reports directly from R. Rather than copying and pasting into Word, your report is created automatically. R Markdown combines the core syntax of markdown (an easy-to-write plain text format) with embedded R code chunks that are run so their output can be included in the final document. R Markdown documents are fully reproducible (they can be automatically regenerated whenever underlying R code or data changes). Markdown is simple to use as it enables the use of a syntax like plain-text.

1. You need to install LaTeX on your machine. This is a free typesetting software which R Markdown uses. Windows and Mac users must install TeXLive 2016 or MikTeX (freely available at https://www.tug.org/texlive/ and https://miktex.org/).

2. To use R Markdown, simply create a new R Markdown document in RStudio. This will load a sample document. Select "Knit PDF" to produce a PDF output file with the write up and the code output.

3. More information on the R Markdown syntax is available here: http://rmarkdown.rstudio.com/.

Schedule

Week 1 (Oct 12): Review of Fundamentals and Intro to R

Reading:
   OpenIntro Statistics, chapter 1, or Gravetter and Wallnau, chapters 1–4.

Week 2 (Oct 19): Probability Theory

Reading:
   OpenIntro Statistics, chapter 2 or Gravetter and Wallnau chapter 6.

Week 3 (Oct 26): Sampling and Hypothesis Testing

Reading:
   OpenIntro Statistics, chapters 4 and 5, or Gravetter and Wallnau, chapters 7,8, 9, 11.

Last updated: July 24, 2017

Week 4 (Nov 2): Correlation and Linear Regression

Reading:

OpenIntro Statistics chapters 7 or Gravetter and Wallnau, chapters 15 and 16.

Week 5 (Nov 9): Analysis of Categorical Data and Dichotomous Dependent Variables

Readings:

Gravetter and Wallnau chapter 16 AND OpenIntro Statistics, chapters 6 and 8 (from p. 386).

Week 6 (Nov 16): Other Considerations: Missing Data, Statistical Power, and Post-Stratification

Readings:

Schafer, J. L., and Graham, J. W. (2002). "Missing Data: Our View of the State of the Art" *Psychological Methods*, 7(2):147177;

TBD

Week 7 (Nov 23): What Else Is out there I: Multilevel Models and Cross-Sectional Time-Series Data Analysis

Readings:

Luke, D. A. (2004). *Multilevel Modeling*. Thousand Oaks, CA: Sage. Chapter 1 (pp. 1–7).

Cowpertwait, P. S. P., and Metcalfe, A. V. (2009). *Introductory Time Series with R*. New York: Springer. Chapter 1 (pp. 1–25).

Finkel, S. E. (1995). *Causal Analysis with Panel Data*. Thousand Oaks, CA: Sage. Chapter 1 (Modeling Change with Panel Data).

Week 8 (Nov 30): What Else is out there II: Predictive Modeling

Reading:

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer. Chapters 1–2.

Last updated: July 24, 2017