

# INTRODUCTION TO QUANTITATIVE ANALYSIS USING R\*

## MA Course, Winter 2019/20

---

**Thursdays 10:00 – 13:30, from October 10 to November 28.**  
**Seminarroom S105, IBW Building**

**Instructor:**

Bruno Castanho Silva, PhD

Cologne Center for Comparative Politics (CCCP)

Email: [bcsilva@wiso.uni-koeln.de](mailto:bcsilva@wiso.uni-koeln.de)

Office hours: By appointment.

Office: IBW Gebäude, Herbert-Lewin Str. 2, 3.44c, third floor..

---

## 1 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 (i.e., can you calculate an average?). No prior knowledge of R is required.

**Nevertheless, students who take the 2-day workshop on Intro to R at CCCP offered by Ayjeren Rozyjumayeva will have an easier time in the first sessions.**

## 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/>. 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.**

## 2 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.

### 3 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. **It is mandatory to deliver all assignments to pass the course.** For the first five, students will have a week to complete the assignment, and three 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 through Ilias, with the following deadlines and weights towards the final grade:

Homework	Posted	Deadline	Weight
Assignment 1	October 15	October 22, 23:55 CET	15pts
Assignment 2	October 29	November 05, 23:55 CET	15pts
Assignment 3	November 05	November 12, 23:55 CET	15pts
Assignment 4	November 12	November 19, 23:55 CET	15pts
Assignment 5	November 26	December 3, 23:55 CET	15pts
Assignment 6	December 3	December 24, 23:55 CET	25pts

Points are converted to final grades as follows:

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

#### Bonus Points

It is possible to increase your grade by up to 12 points in this course, with two bonus points assignments. Each one will have a maximum grade of six. They will be posted on **October 22** and **December 3**, and you will have one week to do each. Bonus points exercises **cannot** be used a replacement for one of the mandatory assignments. It means that you can only receive bonus points if you submit all six mandatory assignments.

Points	Grade
100–96	1,0
95–91	1,3
90–86	1,7
85–81	2,0
80–76	2,3
75–71	2,7
70–66	3,0
65–61	3,3
60–56	3,7
55–51	4,0
50–0	5,0

### 3.1 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  $\text{\LaTeX}$  on your machine. This is a free typesetting software which R Markdown uses. Mac users should install TeXLive, freely available at <https://www.tug.org/texlive/>; Windows users should install `tinytex`, which can be done directly within RStudio, with the following three commands:
  - `install.packages('tinytex')`
  - `library(tinytex)`
  - `install_tinytex()`
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/>.

**NB! Installing and running R and RMarkdown can be tricky. All students must make sure they have an operational installation of both before the first class. If not, please contact the instructor in advance.**

## 4 Schedule

### **Week 1 (Oct 10): Review of Fundamentals, Probability Theory, and Intro to R**

Reading:

OpenIntro Statistics, chapter 1 and 2, OR Gravetter and Wallnau, chapters 1–4 and 6.

### **Week 2 (Oct 17): The Normal Distribution and Sampling**

Reading:

OpenIntro Statistics, p. 127–137; AND OpenIntro Statistics chapter 4, OR Gravetter and Wallnau chapters 7 and 8;

### **Week 3 (Oct 24) Hypothesis Testing and the t-test**

Reading:

OpenIntro Statistics, p. 127–137; AND OpenIntro Statistics chapters 4 and 5, OR Gravetter and Wallnau chapters 10 and 11.

Plus, watch these two videos in sequence:

<https://www.youtube.com/watch?v=vYVr50hjFbQ>

and

<https://www.youtube.com/watch?v=xh20btybjp4&t=386s>

### **Week 4 (Oct 31): Correlation and Bivariate Linear Regression**

Reading:

OpenIntro Statistics, chapters 4 and 5, or Gravetter and Wallnau, chapters 16 and 17 (up to 17.3).

**Week 5 (Nov 07): Multivariate Linear Regression**

Gravetter and Wallnau chapter 17 (from 17.3 onwards)

**Week 6 (Nov 14): Regression 3: Interactions, Polynomials, and Power**

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political analysis*, 14(1), 63-82.

**Week 7 (Nov 21): 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 8 (Nov 28): Further Considerations and What Else is Out There**

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).

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.

Schafer, J. L., and Graham, J. W. (2002). “Missing Data: Our View of the State of the Art” *Psychological Methods*, 7(2):147–177;