

INTRODUCTION TO QUANTITATIVE ANALYSIS USING R*

MA Course, Winter 2020/21

Lecture videos will be uploaded every Friday from 06/11 until 18/12
Online consultations will take place on Thursdays 11:00 - 12:00, from
12/11 until 17/12, and then once more on 14/01.

Instructor:

Bruno Castanho Silva, PhD

Cologne Center for Comparative Politics (CCCP)

Email: bcsilva@wiso.uni-koeln.de

Office hours: By appointment.

1 How This Course is Structured Online

Usually this course would be structured with weekly 3-hour meetings, in which half the session is a lecture/seminar, and the other half is practice in R. Due to the extraordinary circumstances this term, the weekly schedule will be the following:

- A few short video lectures adding up to 45-60 minutes on the topic of that week, uploaded every Friday;
- A video lecture of the R tutorial, also uploaded every Friday;

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

- A 1h online consultation through Zoom, every Thursday at 11:00 starting from November 12. The consultation will always focus on the topics of the videos uploaded the previous Friday, so you have almost a week to watch them and start with the weekly assignments before our online meeting. If this slot does not work for you, please write me an email and we can talk another time.
- Ilias forums for questions.

2 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. Empirical applications will focus primarily on political science. Throughout the course, students will become familiar with the statistical programming language R. The goal is to prepare students to critically interpret and conduct quantitative studies in political science.

Prerequisites

Students should know high school arithmetics (i.e., can you add, subtract, multiply, divide, and calculate an average?). No prior knowledge of R is required.

Nevertheless, students are strongly advised to take the Online workshop on Intro to R at CCCP offered by Daniel Saldivia Gonzatti on October 26 -28. Taking the workshop will make this course easier. Link for more information:

<https://cccp.uni-koeln.de/de/news/single-view/introduction-to-r-workshop-2>

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. We will have plenty of practice on R during the course.

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.

3 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.

4 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 assignment is a short data analysis paper of 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, while the deadline for the last is a month after the last videos are uploaded. You can discuss your assignments 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:

Last updated: August 28, 2020

Homework	Posted	Deadline	Weight
Assignment 1	November 13	November 20, 23:55 CET	15pts
Assignment 2	November 20	November 27, 23:55 CET	15pts
Assignment 3	November 27	December 4, 23:55 CET	15pts
Assignment 4	December 4	December 11, 23:55 CET	15pts
Assignment 5	December 11	December 18, 23:55 CET	15pts
Assignment 6	December 18	January 22, 23:55 CET	25pts

Points are converted to final grades as follows:

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

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

Bonus Points

It is possible to increase your grade by up to 10 points in this course, with a bonus points assignment, which will be posted at the end of the semester. The bonus points exercise **cannot** be used to replace one of the mandatory assignments. It means that you can only receive bonus points if you submit all six mandatory assignments. Moreover, you can only receive bonus points if you reach the minimum passing grade with the mandatory assignments. If you fail to reach a passing grade (51) with the six mandatory assignments, the bonus points will not be added to them in order to reach the passing grade.

4.1 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 \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.

5 Schedule

This is a preliminary reading list which will be updated before the semester starts!

Week 1 (Nov 6): Review of Fundamentals, General introduction and Intro to R

Reading:

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

Week 2 (Nov 13): Distributions and Sampling

Reading:

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

Week 3 (Nov 20) Models, Inference and Hypothesis Testing

Reading:

OpenIntro Statistics, p. 127–137; AND OpenIntro Statistics chapters 4 and 5, OR Gravetter and Wallnau chapters 8, 9, and 10.

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 (Nov 27): Correlation and Intro to Linear Regression

Reading:

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

Week 5 (Dec 04): Linear Regression II

Gravetter and Wallnau chapter 17 (from 17.3 onwards) AND James et al. 2013, chapter 3.

Week 6 (Dec 11): Linear Regression III

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

Week 7 (Dec 18): Logistic Regression and Wrap Up

Readings:

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