



Universität zu Köln

Quantitative Analysis of European Politics

Summer Semester 2017

Thursday, 12:00 - 13:30 (Seminarraum 1.31, Universitätsstrasse 91)

(Note: First session takes place in Hörsaal 2 in Gottfried-Keller-Str. 6)

Prof. Sven-Oliver Proksch

Cologne Center for Comparative Politics (CCCP)

E-mail: so.proksch@uni-koeln.de

Office Hours: Thursday 13:30-14:30 (by appointment)

Office: Universitätsstrasse 91

1. Course Description

This course is open to MA or PhD students and provides an introduction to quantitative analysis in political science using R. Topics include the estimation and interpretation of the linear regression model, models for binary data, and quantitative analysis of political text. Empirical applications will focus primarily on European politics. 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.

Prerequisites

Students should have a basic understanding (undergraduate level) of statistical methods in political science. No prior experience with 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/>. 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.

2. Course Materials

The course is structured as an applied introduction to quantitative analysis with R. It is recommended that students obtain a copy of a statistics textbook as a resource for the course and the term paper. The following textbooks are recommended:

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

While this text is no-doubt expensive, it is worth having. This particular statistics textbook is highly regarded and widely used for similar courses. The following companion book includes a range of useful practical examples of the various statistical techniques using R:

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

The following books provide alternative coverage of the topics covered in this course (and more). The Diez et al. is an open source textbook for introductory statistics (with specific R examples) and can be downloaded for free. Gailmard's book focuses more heavily on the connection between statistical procedures and social science theory. Wooldridge and Kennedy are popular introductory econometrics textbooks.

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

Gailmard, Sean. 2014. *Statistical Modeling and Inference for Social Science*. Cambridge University Press.

Wooldridge, Jeffrey. 2016. *Introductory Econometrics: A Modern Approach*. 6th edition. South-Western College Pub.

Kennedy, Peter. 2008. *A Guide to Econometrics*. 6th edition. Blackwell Publishing.

3. Course Requirements

Grading for the course will be based on the following components:

- Research Paper (70 points)

You will write a 15 page research paper, to be handed in on **August 10**. A list of references and appendices do not count towards the page limit. Your paper should be an original quantitative analysis and take the form of a journal submission. It should focus on a research question in European politics and you should plan to use the statistical techniques introduced in the class (while it can include a quantitative content analysis component, this can only be in addition to a regression analysis). The largest component of the paper should be on data analysis and interpretation (meaning it should include a shorter literature review and theory section). The paper needs to be written in \LaTeX (see below). Students are encouraged to use the research

project as an opportunity to start work on their thesis, or to work on a potentially publishable paper. You can pursue a paper that is a replication and extension of an existing published journal article. Replication means going back to a primary data source to reconstruct a dataset (rather than using a replication file from the author) and to replicate the original result. Extension means an original additional contribution to an existing study, involving for instance a different measurement approach, consideration of additional variables, changes to model specifications, inclusion of additional data, etc. Extensions should be derived from theoretical propositions or a methodological critique.

You are expected to hand in a 3-page proposal for the paper by **May 18** (email). It should describe the puzzle/research question, working hypotheses, and the data and methods you are planning to use. Moreover, you need to include the dataset that you will use. If you do a replication, attach a copy of the paper to be replicated. It is your responsibility to start the process of choosing your dataset early enough.

- Research Project Presentation (30 points)

You will present your research results to the class during the final weeks of the course. You need not have the paper finished at this point, but the empirical analysis should be concluded. This means: start your research project early. Your presentation will be modelled after a panel presentation at a conference. Other students will provide feedback to the presentation. This feedback should be respectful and offer constructive methodological criticism, including ideas for how the research can be improved. Speaking about empirical methods can be difficult. The presentation is included to give students a chance to practice this skill.

Points are converted to the final grade 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

4. Bonus Work

There will be five bonus assignments in the form of replications, simulations, or extensions of the analysis in class and the lab distributed throughout the semester. The assignments will be handed out at the end of a class and you have one week to complete them. Assignments must be written in R Markdown (see below) and handed in prior to the next class meeting (unless noted otherwise).

Each assignment will be graded according to the following scale: 0 points, 0.5 points, 1 point. Late, incomplete or assignments with major errors will receive 0 points. Assignments with some minor errors will receive 0.5 points. Assignments with no errors will receive 1 point. It means that a student can receive up to 5 bonus points in total (the total number of points cannot exceed 100 points). Of course, it is possible to obtain 100 points in the course without any bonus points. You may organize in groups of up to 3 students, but you must write up each assignment on your own and perform any analyses yourself. If you have worked with another student, please indicate with whom on your homework.

4. A Note on Professional Presentation

R Studio 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 email. Your bonus data analysis assignments need to be written in R Markdown.

1. You need to install \LaTeX on your machine. This is a free typesetting software which R Markdown uses.
On Windows: <https://www.tug.org/protext/>
On Mac: <https://tug.org/mactex/>
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/>.

The research paper needs to be written in \LaTeX . It is a free typesetting software package and enables you to typeset and print your work at the highest typographical quality, using a predefined, professional layout. The main advantages of \LaTeX over normal word processors include professionally crafted layouts, support for typesetting of mathematical formulae in a convenient way (which makes formal theory papers easier to write), a few easy-to-understand commands that specify the logical structure of a document, more complex structures such as footnotes, references, table of contents, bibliographies that can all be generated easily, and free add-on packages for specific tasks (e.g. make a reference list adhere to the exact standards of a scientific journal). A short introduction can be found [here](#). Importantly, there are a number of features that make it particularly compatible with R. It is essential that you complete your paper with a professional presentation of your results including – but not limited to – a professional layout of text, meaningful and professionally looking tables and publication-quality graphs. Final papers should be written in font size 12pt, double-spaced, and with at least 2.5cm margins on all sides.

5. Course Outline

Week 1 (20 April): Introduction and Overview of R

(Note: First session takes place in Hörsaal 2 in Gottfried-Keller-Str. 6)

- Gailmard 2014, “Observable Data and Data-Generating Processes”
- Fox 2016, chapters 1-2 *or* OpenIntro Stats, chapter 1

Week 2 (27 April): Fundamentals of Probability Theory

- Fox 2016, appendix D *or* OpenIntro Stats, chapters 2-3

Week 3 (4 May): Sampling and Statistical Inference

- Fox 2016, chapter 6 *or* OpenIntro Stats, chapters 4-6

Week 4 (11 May): Linear Regression

- Fox 2016, chapter 5 and 7 *or* OpenIntro Stats, chapter 7

Week 5 (18 May): Linear Regression: Interaction Terms, Inference, and Hypothesis Testing

- Fox 2016, chapter 6 *or* OpenIntro Stats, chapter 8
- Thomas Brambor, William Roberts Clark, and Matt Golder. 2006. “Understanding Interaction Models: Improving Empirical Analyses,” *Political Analysis* 14: 63-82

Week 6 (25 May): No class (Ascension Day)

Week 7 (1 June): Presenting Regression Results

- Gary King, Michael Tomz, and Jason Wittenberg. 2000. “Making the Most of Statistical Analyses: Improving Interpretation and Presentation”, *American Journal of Political Science* 44(2): 341-35

Week 8 (8 June): No class (Study Break)

Week 9 (15 June): No class (Corpus Christi)

Week 10 (22 June): Generalized Linear Models: Binary Dependent Variables

- Fox 2016, chapter 14

Week 11 (29 June): Text as Data: Quantitative Content Analysis

- Grimmer, Justin and Brandon Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts”, *Political Analysis*
- Jonathan B. Slapin and Sven-Oliver Proksch. 2008. “A Scaling Model For Estimating Time-Series Policy Positions from Texts”, *American Journal of Political Science* 52(3), 705-722.

Week 12 (6 July) - Week 15 (27 July): Project Presentations