UNIVERSITY OF MANNHEIM
School of Social Sciences

Quantitative Methods in Political Science

(a.k.a. Multivariate Analyses)

Fall 2020

Course Details:
Lecture: Wed, 8h30-10h00 (most weeks: pre-recorded + interactive Q&A Wed, 9h00-9h30)
Lab 1: Wed, 15h30-17h00
Lab 2: Thu, 10h15-11h45

Instructors:

Denis Cohen	Marcel Neunhoeffer (Lab 2)	Oliver Rittmann (Lab 1)
A5, 6 Room A 218	A5, 6 Room B 226	A5, 6 Room B 226
181-2876	181-2542	181-2542
denis.cohen@mzes.uni-mannheim.de	mneunhoe@mail.uni-mannheim.de	orittman@mail.uni-mannheim.de
Wed, 9h45-11h15	Mon, 15h30-17h00	Mon, 15h30-17h00

Course Description:

This course introduces graduate students to quantitative methods in political science. During the first half of the course, we will focus on linear regression models. The topics covered include discussions of the mathematical bases for such models, their estimation and interpretation, model assumptions and techniques for addressing violations of those assumptions, and topics related to model specification and functional forms. During the second half of the course, students will be introduced to the likelihood principle as a theory of inference, including models for binary and count data.

The main goals of this course are to develop sound critical judgment about quantitative studies of political problems, to interpret quantitative analyses in published work, to understand the logic of statistical inference and to recognize and understand basic regression models. It provides the skills necessary to conduct your own quantitative analyses and teaches how to do so using R. This class lays the foundation for "Advanced Quantitative Methods" which will be taught in Spring 2021.

The course language is English. This includes lectures, discussions and assignments.

Recommended for:

Political science graduate students, including M.A. and PhD students (CDSS) and MMDS students.

Course Registration:

The course is divided into a lecture (Multivariate Analyses) and a computer lab session (Tutorial Multivariate Analyses). During the lab sessions, students will apply the statistical models introduced in the lecture. Students who wish to take the course should register for "Multivariate Analyses" at the student portal.

Note that this course is highly demanding and entails a substantial work load for students in the form of weekly homework assignments, a mid-term exam and a data essay. Students who wish to audit this class should notify the instructors in advance. Please note that only registered students will receive feedback on their written work.

Teaching Organization:

Due to the ongoing COVID-19 pandemic, all teaching activities and office hours will take place online. We will use a mix of platforms and teaching methods including:

1. Weekly pre-recorded lecture videos, to be uploaded one week in advance on Thursdays, 12:30, on YouTube.
2. Weekly interactive lecture Q&A sessions, to be held on Wednesdays, 9:00-9:30, on Zoom.
3. Weekly interactive lab sessions on Zoom in two groups:
    - Group 1: Wednesdays, 15:30-17:00
    - Group 2: Thursdays, 10:15-11:45
4. Weekly office hours for the lecture and lab contents on Zoom:
    - Lecture (Denis Cohen): Wednesdays, 9:45-11:15
    - Labs (Marcel Neunhoeffer, Oliver Rittmann): Mondays, 15:30-17:00
5. Weekly exercises, to be distributed on and submitted via ILIAS.
6. A discussion forum on ILIAS for any questions of general interest about lecture and lab contents or the exercises.

Figure 1 below illustrates the weekly schedule for Weeks 2-8 and Weeks 10-11. Lab office hours on Monday and the lab exercise deadline on Tuesday will focus on the materials covered in the preceding week $t-1$ (green).

The lecture Q&A session and lecture office hours on Wednesday as well as the lab sessions on Wednesday and Thursday will focus on the current week t's topic (orange). A corresponding pre-recorded lecture video for week t will be made available in the preceding week $t-1$ on Thursday and should be watched by everyone prior to the corresponding Q&A session in week t on Wednesday, 9:00-9:30. Questions about the recording should

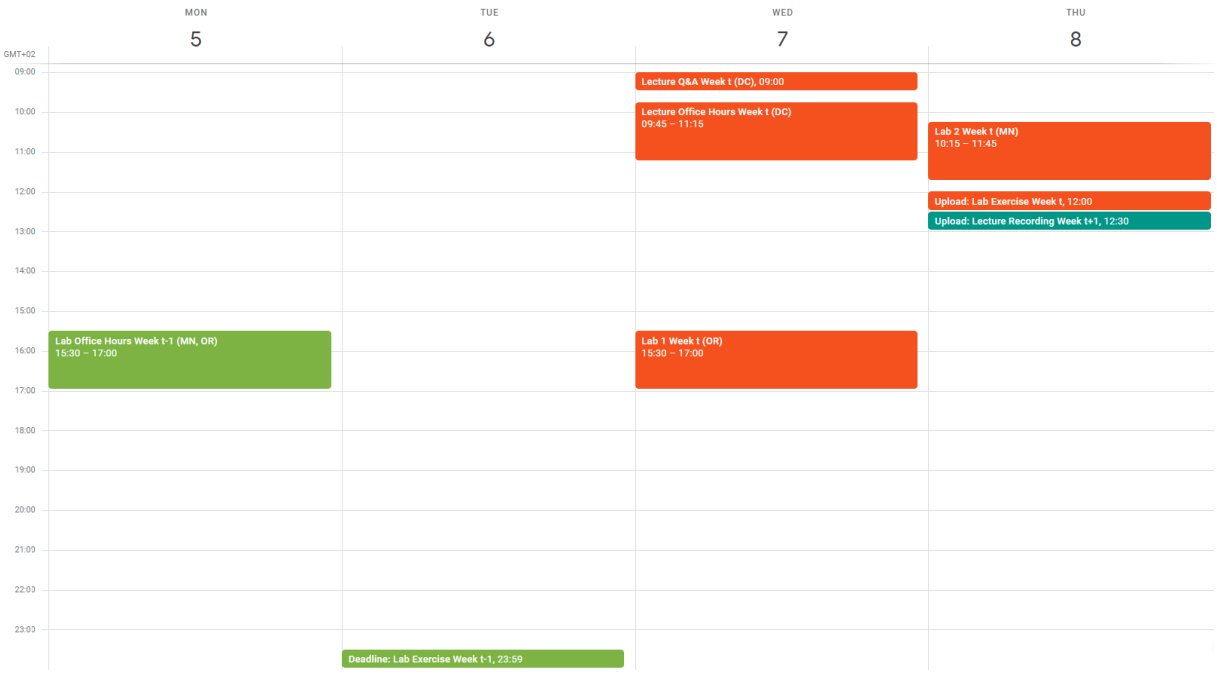

Figure 1: The weekly schedule for Weeks 2-8 and Weeks 11.

be posted on ILIAS at least 24 hours before the Q&A.

After the lab session for week t, we will post the corresponding lab exercise (to be submitted on Tuesday, 23:59, of the following week $t+1$) as well as the lecture recording for week $t+1$ (turqoise).

Please note that unlike Weeks 2-8 and Weeks 10-11, **Weeks 1 & 12** will be taught as fully interactive 90 minute Zoom sessions on **Wednesday, 08:30-10:00**. In **Week 9**, you will have to complete the open-book take-home mid-term exam on **Wednesday, 08:30-10:00**.

Readings:

We will not use a single textbook for this course. Selected readings are available on the course ILIAS site. The following books will be used in the course:

Angrist, Joshua D, and Jörn-Steffen Pischke. 2015. *Mastering Metrics. The Path from Cause to Effect.* Princeton, NJ: Princeton University Press.

Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models.* 2nd edition. Sage.

King, Gary. 1998. *Unifying Political Methodology.* Ann Arbor: University of Michigan Press.

Wooldridge, Jeffrey. 2009. *Introductory Econometrics: A Modern Approach.* 4th edition. South-Western College Pub.

Software:

For all calculations, we will support and use the open-source statistical programming language R. It is particularly suited for carrying out state-of-the-art computer-based simulations and data exercises. It can also be used to generate really nice publication-quality visualizations and 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. A readable introduction is given by Fox, John. 2002. *An R and S-Plus companion to Applied Regression.* Sage. Students with a Stata background can also look at *R for Stata Users* by Muenchen, Robert A. and Hilbe, Joseph M., 2010.

A very good graphical user interface for R (which we will also use during the lab sessions) is RStudio. In recent years a growing number of features have been added to this graphical user interface, which makes it the preferred choice for learning R – especially among beginners. It is cross-platform and open-source. RStudio can be downloaded for free at http://www.rstudio.com/. A style guide to make your code easier to read, share, and verify can be found at http://adv-r.had.co.nz/Style.html.

To facilitate an efficient workflow, we will integrate Github into the course. git is a version control system that makes it easy to track changes and work on code collaboratively. GitHub is a hosting service for git. You can think of it like a public Dropbox for code. We will use it to distribute code and assignments to you. And you will use it to keep track of your code and collaborate in teams. You can find the course on GitHub here .

To get started with R, RStudio and git, please follow the instructions here.

The lab sessions will be devoted to learning the various commands in R and applying the statistical models from the lecture to selected political science data sets. The data sets that we will use cover the major fields in political science.

Prerequisites:

There are formally no prerequisites for this course except an open mind and a good command of high school algebra.

Course Requirements:

You will receive a grade for the seminar "Multivariate Analyses" and a pass/fail for the lab session "Tutorial Multivariate Analyses". Grading will be based on the following components:

- **Homework Assignments (pass/fail)**

  The homework assignments will take the form of problem sets, replications, simulations, or extensions of the analyses in class and the lab. The assignments will be handed out after Lab 2 (Thursday, 12:00) and you are expected to hand in the solution online on GitHub by the next Tuesday 23h59. Late submissions will not be accepted. We provide you with instructions on how to upload your assignemnts on GitHub in the first week.

  Homework assignments for Weeks 1 & 2 must be handed in individually. Homework assignments from Week 3 onward must be handed in in groups of 2-3 students. We

will randomly assign groups in Weeks 3, 6, and 10. All group members will receive the same grade. Throughout the years, we learned that the only reasonable way to manage the workload in the first semester is to work together. Group work saves you time, as not everyone has to type up his or her own answers. Although it is completely up to you how you share the work, you learn best if everyone tries to answer the problem sets individually first. From our experience, not getting strongly involved in each of the weekly homework assignments lowers group work quality and increases the risk of failing both the mid-term exam and the data essay project. For this reason, indicate about how much percentage points each group member contributed towards the final product.

We are aware that the current situation complicates group work. To make communication among group members as easy as possible, we set-up a slack-channel. Click on this link to join the slack-channel. Note that you can only *sign-in with your @mail.uni-mannheim.de adress*.

Moreover, you are strongly encouraged to seek advice from the instructors during office hours or through the ILIAS forum. Homework assignments are graded only as *pass* or *fail*. However, all homework assignments have to be passed!

- **Midterm Exam (1/2)**

  The midterm will be a 90-min open-book take-home exam that covers the first half of the course materials.

- **Data Essay (1/2)**

  Towards the end of the course, you will work on a data analysis project (we will hand out a data set and problem description). The project will involve the creative application of the statistical techniques to a substantive problem in political science. Your paper should have 2000 words ($\pm$ 10%, without bibliography). The essay will be marked down if you go below or above the word count. The data essay should emphasize the substantive, statistical, and causal significance of your analysis and the write-up should read very much like the results section of a published article. No collaboration is permitted on the data analysis project. You are welcome to seek advice from the instructors during office hours. Details on the data analysis project will be provided at a later stage. Your data essay is due on **22 December 2020** by **23h59**. Late submissions will not be accepted.

Other Considerations:

A great website with many R code examples is the UCLA Stat Consulting Site. Another good site that introduces R to SPSS or Stata users is Quick-R. The standard site to get help for R on the net is Stack Overflow. There are also great resources (Primers and Cheat Sheets) within the rstudio.cloud ecosystem.

Within your assignments, you will be required to write mathemtical expressions and formulas. A great typesetting software package to do that is LaTeX. Rstudio supports LaTeX-style math formulas and we strongly encourage you to make use of it. You can find an introuction to mathematical expressions in LaTeX here and a list of mathematical symbols here.

Course Outline:

Week 1 (30 September 2020): Introduction. Visualizing Data.

Fox, John. 2008. *Applied Regression Analysis*. Chapters 1 and 3.

Week 2 (7 October 2020): Fundamentals of Probability.

Moore, Will H., and David A. Siegel. 2009. *A Mathematics Course for Political & Social Research*. Chapters $9 - 11$.

Week 3 (14 October 2020): Sampling & Statistical Inference.

Wooldridge, Jeffrey. 2009. *Introductory Econometrics*. Appendix C.

Week 4 (21 October 2020): Linear Regression: Basics & Hypothesis Testing.

Kennedy, Peter. 2008. *A Guide to Econometrics*. Chapters3.

Fox, John. 2008. *Applied Regression Analysis*. Chapters $5 + 6$.

Week 5 (28 October 2020): Linear Regression: Statistical Control & Causality.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2015. *Mastering Metrics. The Path from Cause to Effect.* Chapter 2.

Cinelli, Carlos, Andres Forney and Judea Pearl. 2020. *A Crash Course in Good and Bad Controls.* Retrieved from https://ucla.in/2ZcRpRq.[1]

Week 6 (4 November 2020): Linear Regression: Dummies & Interactions.

Fox, John. 2008. *Applied Regression Analysis*. Chapter 7.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14: 63-82.

Hainmueller, Jens, Jonathan Mummolo, und Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27(2): 163–192.

Week 7 (11 November 2020): Linear Regression: Interpreting Substantive Effects via the Simulation Method.

Gary King, Michael Tomz, and Jason Wittenberg. 2001. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation" *American Journal of Political Science* 44(2): 347-361

Week 8 (18 November 2020): Linear Regression: Diagnostics.

Fox, John. 2008. *Applied Regression Analysis*. Chapter 11-13.2.1

Week 9 (25 November 2020): Midterm Take-Home Exam

Good Luck!

[1] For a background on causal graphs, first read this accessible *Introduction to Directed Acyclic Graphs* by Malcolm Barrett.

Week 10 (2 December 2020): Non-linear Probability Models - The Likelihood Theory of Statistical Inference.

King, Gary. 1998. *Unifying Political Methodology.* Chapters 1, 2, 4.

Week 11 (9 December 2020): Binary Data

Fox, John. 2008. *Applied Regression Analysis.* Chapter 14.1

Week 12 (16 December 2020): Data Essay Q&A. Semester Wrap-Up.