Syllabus
Social Science Methodology: Intermediate Statistics

Instructor: Oliver Westerwinter
Spring semester 2018

Time & room
Monday 12:15-14h, 19.2, 26.2, 5.3, 12.3, 19.3, 26.3, 16.4, 23.4, 7.5, 14.5 in 01-U179, 30.4 in 36-108

Office
Oliver Westerwinter
Room: 52-5012, Müller-Friedbergstr. 8, 5th floor
Email: `oliver.westerwinter@unisg.ch`
Office hours: by appointment

This syllabus may be subject to adjustments.

Overview and goals

This course focuses on intermediate statistical methods in the social sciences. The course is divided into three main parts. The first part deals with regression models and ordinary least squares estimation (OLS). Part two covers binary outcome models, nominal and ordered outcome models, and count outcome models. The final part is concerned with panel data models. Throughout the class, we will focus on both the theoretical foundations as well as the practical application of the statistical methods discussed.

The primary goal of the course is to help students to become increasingly skilled users and critical consumers of research employing statistical methods. A special emphasis will lie on the substantive interpretation of the results of statistical models. At the end of the course, students will have acquired solid training in a broad range of statistical models and how they can be used in a practical social science research context. This includes knowledge of the elementary mathematical and probability theory related foundations of statistical methods. Furthermore, students will also be familiar with techniques of preparing, managing, visualizing, and analyzing statistical data using the computer software R.

Prerequisites

A basic understanding of statistical methods (e.g. descriptive statistics, hypotheses tests, confidence intervals) will be helpful to follow the class. Familiarity with calculus, linear algebra, and probability is helpful, but not required.

Class requirements

Final grades will be based on:

- Term paper (55% of final grade)
- Three homework assignments (45% of final grade)
- Participation in lectures

There will be three homework assignments, each of which will contribute 15% to your final grade. The homework assignments will consist of analytical problems and data analysis. The solutions of the assignments have to be submitted either in handwritten or printed form at the beginning of each class at which they are due. You may also submit your solutions via email to `oliver.westerwinter@unisg.ch`. The assignments will be available one week prior to the class at which they are due. No late homework submissions will be accepted. All submitted homework will be graded. Solutions will be available electronically in due course after the submission of the assignments.

The homework assignments are solo exams. Thus, you always need to write and submit your own solutions. Please always make sure that you write your name on every page of your submission. Please note that the homework assignments are designed to support and deepen your understanding of the material discussed in class. They will help you to self-assess your level of accomplishment and plan your term paper project.

The term paper will contribute 55% to your final grade. It has to be submitted either as hard copy or electronically (PDF file per email to: `oliver.westerwinter@unisg.ch`) by June 1. Late submissions will not be accepted. The term paper is expected to present a short research project that uses statistical methods. The maximum length of the paper is 5,000 words (including footnotes, references, etc.). The term paper is a solo exam. You are therefore expected to work on and submit your term paper project individually.

Course website

We use StudyNet as communication platform for the class. The course site at StudyNet can be accessed here: `https://loginpages.unisg.ch/studynet`.

The course website at StudyNet will provide lecture slides, readings, homework assignments, R scripts, and other supplementary materials.

Computation using R

The course will be taught in R. R is an open-source computing language that is widely used in statistics. You can download it for free from `www.r-project.org` and it is recommended that you install R on your private computer prior to the start of class. For those of you who are not yet familiar with R, the course will include an introduction into the basics of R in the first sessions. Furthermore, I will provide detailed example code as well as other resources to learn programming in R including an introductory manual on the StudyNet website of the course. Additional tutorials and other resources to learn about the basics of R are available here: `http://wiki.math.yorku.ca/index.php/R:Gettingstarted`.

The books *Introductory Statistics with R*, *Political Analysis Using R*, and *Quantitative Social Science. An Introduction* (for full references, see below) provide an accessible introduction to R.

Writing using LaTeX

LaTeX is a document processing environment that is also free and open-source. Working in LaTeX is a lot like working with html code for webpages. You write LaTeX documents using a text editor. You add formating features to the document using various codes that are

available in the base LaTeX system or in various packages that you load as part of the document. You then ask the text editor to use the functions available in LaTeX to compile your document. The result is a PDF (or other) file that looks clean and professionally produced. The value of LaTeX over word processors like MS Word is that you have complete control over how your final document looks. This is often trival for memos or letters, but can be quite helpful when doing scientific writing. LaTeX makes it much easier to write mathematical formulas, include figures, construct tables, and build your references. There are a number of features that make it particularly compatible with using R to conduct your statistical analysis.

LaTeX is available from several sources online. TeX Live works for all common platforms. A version called MacTeX is a version of TeX Live specially designed for installation on a Mac. MiKTeX and its newer version, proTeXt, are distributions for Windows machines. You can read more about this online here: `http://latex-project.org/ftp.html`, `http://www.ctan.org/`, and `http://www.tug.org`.

I encourage you to consider using LaTeX for your work in this class. I will provide support and direction with LaTeX, but if you are interested in using LaTeX for this class, you need to take the responsibility yourself to learn the tools you need to do your work. In addition to the sources provided in this course, you can consult the broad range of online sources. Working in groups with other students to work through the challenges of getting started with LaTeX may also be helpful.

Textbooks

The required readings in combination with the slides will be the primary teaching materials. The main textbook for the course is:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western.

Students who are interested in consulting additional textbooks can choose from a broad variety of statistics and econometrics texts with different strengths and weaknesses including:

Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics. An Empiricist's Companion.* Princeton: Princeton University Press.

Diez, David M., Christopher D. Barr and Mine Cetinkaya-Rundel. 2014. *OpenIntro Statistics.* Second edition. The book and the complementary materials are available at `https://www.openintro.org/stat/textbook.php`.

Fox, John. 2016. *Applied Regression Analysis and Generalized Linear Models.* Third edition. London: Sage.

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson.

Imai, Kosuke. 2017. *Quantitative Social Science. An Introduction.* Princeton: Princeton University Press.

Monogan III, James E. 2015. *Political Analysis Using R.* New York: Springer.

Ugarte, Maria Dolores, Ana F. Militino and Alan T. Arnholt. 2016. *Probability and Statistics with R.* Second edition. London: Taylor & Francis.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data.* Second edition. Cambridge: MIT Press.

Schedule

The weekly coverage might change as it depends on the progress of the class. The assigned required and optional readings are listed in the class schedule for each session. The required readings should be completed prior to the session for which they are listed and studied carefully. In addition, it is recommended to consult the optional readings. The optional readings may prove useful to students looking for additional coverage of some of the course topics as well as for developing the topic of your term paper project. This schedule is subject to adjustments.

February 19 – Introduction & regression and OLS I

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 2+3

Optional readings:

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 2-4

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables.* London: Sage. Ch. 2

February 26 – Regression and OLS II

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 4+7

Optional readings:

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

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 5

Monogan III, James E. 2015. *Political Analysis Using R.* New York: Springer. Ch. 6

March 05 – Regression and OLS III

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 6

Optional readings:

Achen, Christopher H. 2005. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22: 327-339.

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

Braumoeller, Bear F. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58: 807-820.

March 12 – Regression and OLS IV (problem set 1 available)

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 8+9

Optional readings:

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 9

Monogan III, James E. 2015. *Political Analysis Using R.* New York: Springer. Ch. 6

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 15

March 19 – Binary outcome models I (problem set 1 due)

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 17

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables.* London: Sage. Ch. 3

Optional readings:

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 17

Horrace, William C. and Ronald L. Oaxaca. 2006. "Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model." *Economic Letters* 90: 321-327.

Monogan III, James E. 2015. *Political Analysis Using R.* New York: Springer. Ch. 7

March 26 – Binary outcome models II

Required readings:

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables.* London: Sage. Ch. 3

Optional readings:

Bilder, Christopher R. and Thomas M. Loughlin. 2015. *Analysis of Categorical Data with R.* London: Taylor and Francis. Ch. 2

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 17

April 16 – Nominal and ordered outcome models I (problem set 2 available)

Required readings:

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables.* London: Sage. Ch. 5+6

Optional readings:

Bilder, Christopher R. and Thomas M. Loughlin. 2015. *Analysis of Categorical Data with R.* London: Taylor and Francis. Ch. 3

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 18

Monogan III, James E. 2015. *Political Analysis Using R.* New York: Springer. Ch. 7

April 23 – Nominal and ordered outcome models II (problem set 2 due)

Required readings:

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables.* London: Sage. Ch. 5+6

Optional readings:

Bilder, Christopher R. and Thomas M. Loughlin. 2015. *Analysis of Categorical Data with R.* London: Taylor and Francis. Ch. 3

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 18

Monogan III, James E. 2015. *Political Analysis Using R.* New York: Springer. Ch. 7

April 30 – Count models

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 17

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables.* London: Sage. Ch. 8

Optional readings:

Bilder, Christopher R. and Thomas M. Loughlin. 2015. *Analysis of Categorical Data with R.* London: Taylor and Francis. Ch. 4

Greene, William H. 2012. *Econometric Analysis. International Edition.* Seventh edition. New York: Pearson. Ch. 18

Monogan III, James E. 2015. *Political Analysis Using R.* New York: Springer. Ch. 7

May 07 – Panel data models I (problem set 3 available)

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 13

Optional readings:

Andress, Hans-Jurgen, Katrin Golsch and Alexander W. Schmidt. 2013. *Applied Panel Data Analsis for Economic and Social Surveys.* New York: Springer. Ch. 2+3

May 14 – Panel data models II (problem set 3 due)

Required readings:

Wooldridge, Jeffrey. 2013. *Introductory Econometrics. A Modern Approach.* Fifth edition. New York: South-Western. Ch. 14

Optional readings:

Andress, Hans-Jurgen, Katrin Golsch and Alexander W. Schmidt. 2013. *Applied Panel Data Analsis for Economic and Social Surveys.* New York: Springer. Ch. 4+5

Further optional readings

The following texts may turn out helpful for students who want to deepen some of the materials covered in class and review the basics of calculus, probability theory, research design, and programming in R.

Research design:

King, Gary, Robert O. Keohane and Sidney Verba. 1994 *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

Probability theory:

Bertsekas, Dimitri and Tsitsiklis, John. 2002. *Introduction to Probability.* Second edition.

Blitzstein, Joseph K. and Jessica Hwang. 2015. *Introduction to Probability.* New York: Taylor & Francis.

Mathematics:

Gill, Jeff. 2006. *Essential Mathematics for Political and Social Research.* New York: Cambridge University Press.

Moore, Wil H. and David A. Siegel. 2013. *A Mathematical Course for Political and Social Science.* Princeton: Princeton University Press.

Simon, Carl and Blume, Lawrence. 2010. *Mathematics for Economists.* New York: Norton.

Introduction to R:

Dalgaard, Peter. 2008. *Introductory Statistics with R.* Second edition. New York: Springer.

Field, Andy, Jeremy Miles and Zoë Field. 2012. *Discovering Statistics Using R.* Los Angeles: Sage.

Imai, Kosuke. 2017. *Quantitative Social Science. An Introduction.* Princeton: Princeton University Press.