DEPARTMENT OF SOCIAL SCIENCES
Center for Doctoral Studies in the Social and Behavioral Sciences
Prof. Thomas Gschwend, Ph.D.

UNIVERSITY OF MANNHEIM

Visitors
D 7, 27 · Room 301/302
68131 Mannheim · Germany
Telephone +49-621-181-2087
Fax +49-621-181-3699
E-Mail: gschwend@uni-mannheim.de
www.sowi.uni-mannheim.de/gschwend

Assistant (Christel Selzer)
Telephone +49-621-181-2413
E-Mail: methods@uni-mannheim.de

empirical and quantitative methods
gess graduate school of economic & social sciences

Advanced Quantitative Methods

Spring 2011

Course Details:

Time:	Wednesday, 8:30–10:00
Place:	A 5, 6 Room C-107 B
Office Hours:	Wednesday, 10:00–11:00

Course Description:

Building on the analytical and theoretical background of the previous course in our MA methods sequence ("Multivariate Analyses"), this course on "Advanced Quantitative Methods" introduces interested graduate students to strategies and tools of how to develop statistical models that are tailored to answer their particular research questions.

You might have noticed by now, the linear regression model is often an inappropriate tool for answering substantive questions in political science. This course serves as an introduction to a multitude of probability models that are appropriate when the linear model is inadequate. After introducing the fundamentals from which statistical models are developed, this course will focus on one specific theory of inference, namely on the statistical theory of maximum likelihood. We will also devote considerable time to statistical programming, simulating and conveying quantities of material interest of such models (using R and Stata) in order to encourage students to switch from a consumer-mode into a producer-mode of social science research.

The goal of the course is three-fold: (1) to prepare students to conduct research using appropriate statistical models and to communicate their results to a nontechnical audience; (2) provide a foundation in the theory of maximum likelihood so students can investigate and implement a wide range of advanced statistical models; and (3) provide students with

the tools necessary to fine-tune existing or to develop new statistical models of political phenomena.

As an elective class, the current curriculum does not offer an accompanying tutorial to spend valuable time to do additional exercises, unfortunately. I like everyone to understand that this puts an even heavier burden on you as a participant. You need to spend more time outside class to comprehend the material than in a methods course with an accompanying tutorial. I expect everyone to come to class fully prepared. Work through the assigned readings. Expect that this will take considerably longer than in a substantive seminar. Do not skip equations! Instead, take notes, prepare questions and team-up with other to answer them, or as last resort, bring them up in class. After every class I expect you to go over the lecture notes and your notes once again. There is no point in getting lost — particularly not in an elective class.

In the weekly class meeting the course will be conducted as mix of lecture and computer lab sessions. During the lecture sessions I will highlight the central concepts and ideas of the readings (and at that time they should not be new to you!). In the computer lab sessions, students will learn to program the statistical models introduced in the lecture. We will also work through examples collaboratively. Our pace will be unpredictable. Nevertheless, understand that the bulk of learning in this course will take place outside of the classroom, by reading, practicing using statistical software, and doing problem sets.

Recommended for:

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

Prerequisites:

M.A. students of political science should have successfully passed the previous course in the MA methods sequence about "Multivariate Analyses" as well as the accompanying "Tutorial Multivariate Analyses". PhD students should have passed equivalent courses. If you can tell me what $(X'X)^{-1}X'y$ is, you have the necessary background to take this class.

Course Registration:

Students who wish to take the course should register for "Advanced Quantitative Methods" at the student portal.

Note that this course is highly demanding and entails a substantial work load for students! Students who wish to audit this class should notify the instructor in advance (participation is subject to free room capacity). Please note that only registered students will receive feedback on their written work.

Readings:

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

Eliason, Scott R. 1993. *Maximum Likelihood Estimation: Logic and Practice.* Newbury Park: Sage.

King, Gary. 1989. *Unifying Political Methodology.* Ann Arbor: University of Michigan Press.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Newbury Park.: Sage.

Hardware/Software:

Students are welcome to use their own computers. I will primarily support and use two software packages in the course: R and Stata. For the majority of problems, R will be the software package of choice.

The open-source statistical programming language R is particularly suited for carrying out state-of-the-art computer-based simulations and programming advanced statistical models. 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/`. A readable introduction is given by Fox, John. 2002. *An R and S-Plus companion to Applied Regression.* Sage.

Stata is a commercial software package and probably the most widely-used software package in political science. Its appeal stems from many pre-programmed ("canned") procedures and its ease of use. An introduction is given in Kohler, Ulrich and Kreuter, Frauke. 2005. *Data Analysis using Stata.* Stata Press.

Course Requirements:

Grading will be based on the following components:

- **Homework Assignments (50%)**

  There will be a series of (mostly bi-weekly) homework assignments that will take the form of problem sets, replications, simulations, or extensions of the analysis in class and the lab. The assignments will be handed out at the end of class and you are expected to hand in the solution at the beginning of the next meeting (unless noted otherwise). You need to work through every homework assignment. Late assignment will not be accepted.

  I encourage you to work in small groups on the assignments. Usually 3-4 people per group seems to be best. While collaboration is encouraged, each student must write up his or her problem set individually. If you have worked with another student, please indicate with whom you did so on your homework. Moreover, you are strongly encouraged to seek advice from the instructor during office hours or by email. Note that

instructive discussions about the material are best done during office hours rather than by email.

- **Final Paper (50%)**

  There will be a final draft paper but no final exam. Each student will produce a co-authored manuscript (or a solo-authored manuscript, with permission of the instructor) that applies or develops the appropriate statistical model to an important substantive problem. Students will choose their own topics. What works particularly well is to start replicating an already published article. My advice is to pic an article that interests you, was published within the last few years in a good journal, and uses methods we have or will talk about in class (or uses different methods at about the same level of sophistication).

  The draft paper must include all analysis, tables, figures, and description of the results. A good write-up of the draft paper should read like the third quarter of a journal article. The rest of the draft may be in detailed outline form, although it would be better to have it fully written.

  You also need to provide all necessary information to replicate your analysis. The replication material must include your dataset and computer code to be able to reproduce all tables and figures that make it in the paper. I expect you to comment your computer code heavily to explain what you are doing. Your code must be neatly formatted and run cleanly. To that end, please avoid writing computer-specific lines into your code that will prevent it from running on other machines. I will award partial credit if necessary.

  The final draft paper together with all replication material are due on **June 9^{th}, 2011**. Please submit all files electronically and, additionally, a hard-copy of your draft paper by **10am** that day. Late submissions will not be accepted.

What to do today?

Find a coauthor and start working on the the draft paper very soon.

Other Considerations:

A great website with many R and Stata code examples is the UCLA Stat Consulting Site. Another good site that introduces R to SPSS or Stata users is Quick-R.

Learn to use LaTeXwhile you can. 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 LaTeXover normal word processors include professionally crafted layouts, support for typesetting of mathematical formulae in a convenient way, 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.

Course Outline:

The following is the list of topics I hope to cover in this course. Because the pace of the course will vary throughout the semester, I have chosen not to assign dates for the topics below. After the third topic the ordering is fairly arbitrary. We will approach the material as slowly, or as quickly, as necessary, and will cover as much, or as little, as possible. Each week after class the readings for the following week will be assigned and can be downloaded from the course website. A list of topics we will cover includes the following:

- Probability Theory and the Role of Statistical Simulation
- A Review of Linear Models (in Matrix Form) as well as Probability Theory
- Introduction to Maximum Likelihood Estimation (MLE)
- Heteroscedastic Regression
- Logit and Probit
- Models for Ordinal Dependent Variables
- Choice Models
- Truncation, Censoring and Selection
- Models for Pooled Cross-Sectional Time-Series Data
- Event Count Models
- Baby-Bayes