Syllabus
17.800 Quantitative Research Methods I: Regression

Professors: Jens Hainmueller & Teppei Yamamoto
TF: Krista Loose
Fall Semester 2011

Time & Room
Class: M & W 3:30-5 in 66-160
Recitation: F 10-11 in E53-438

Office
Jens Hainmueller
Room: E53-459
Email: jhainm@mit.edu
Office Hours: by appointment

Teppei Yamamoto
Room: E53-463
Email: teppei@mit.edu
Office Hours: by appointment

Krista Loose
Room: E53-408
Email: loosek@mit.edu
Office Hours: M 2:00-3:30.

Overview and Class Goals
This is the first course in a three-course sequence on quantitative political methodology, by which we mean the application of statistical methods to problems in political science and public policy (the subsequent classes are 17.802 and 17.804). The goal of the three-course sequence is to teach you to understand and to confidently apply a variety of statistical methods and research designs that are essential for political science and public policy research.

In this first course, we give a graduate level introduction to regression models (primarily linear regression) which are routinely used in political science, policy research, and all other social science disciplines. The principles learned in this course also provide a foundation for a general understanding of quantitative political methodology. If you ever want to collect data, analyze data, critically read an article which presents a data analysis, or think about the relationship between theory and the real world, then you will (hopefully) find this class useful.

You can only learn statistics by doing statistics. In recognition of this fact, the homework for this course will be extensive. The homework and lectures are accompanied by required readings. You may find it helpful to read these multiple times (before, during, and after the corresponding homework).

The class is open to interested students from other departments and qualified undergraduates, subject to permission of the instructor.

Prerequisites
A willingness to work hard on possibly unfamiliar material. The math pre-fresher course (or equivalent) must be completed. A basic understanding of statistics is very helpful meaning that you will benefit more from the class if you have taken at least one undergraduate class in quantitative methodology (such as for example MIT's 17.871 or Harvard's Gov 1019 and 1020).

Class Requirements
Grades will be based on

- homework assignments (70% of final grade)

- a midterm exam (25% of final grade)
- participation and presentation (5 % of final grade).

The weekly homework assignments will consist of analytical problems, computer simulations, and data analysis. They will usually be assigned on Wednesday night and due the following Wednesday, prior to lecture. No late homework will be accepted. All sufficiently attempted homework (ie. a typed and well organized write-up with all problems attempted) will be graded on a (+,✓,-) scale, and may be re-written and re-graded once. The re-write is due before the Wednesday lecture one week after the assignment is returned. We encourage students to work together on the assignments, but you always need to write your own solutions, and we require that you make a solo effort at all the problems before consulting others in your group. We also ask that you write the names of your co-workers on your assignments. Notice that for the last problem set in the term, you will not be allowed to collaborate with anybody. This is to test if you have developed sufficient experience to work through problems on your own. ***No incompletes will be given in this course.***

The Midterm Exam (in-class) will take place on October 12. Please plan accordingly.

Recitation Sections
Weekly recitation sections will be held on Fridays 10-11 in E53-438. The section will cover a review of the theoretical material and also provide help with computing issues. The teaching assistant will run the sections and can give more detail. Attendance is strongly encouraged.

Course Website
The course website is located at the following URL:

http://stellar.mit.edu/S/course/17/fa11/17.800/

This site will provide homework assignments, datasets, and supplementary materials.

Course Forum
The course website has a discussion board in the "forum" section. This discussion board provides an opportunity to post questions regarding the course material and/or computing. In addition to precepts and office hours, please use this Forum on the Stellar course website when asking questions about lectures, problem sets, and other course materials. This will allow students to see other students' questions and learn from them. Both the TA and the instructors will regularly check the Board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. A student's respectful and constructive participation on the forum will count toward his/her class participation grade. Do not email your questions directly to the instructors (unless they are of personal nature) — we will not answer them!

Schedule
Please notice the following scheduling issues: No class on 9/21 (Student Holiday) and no class on 10/10 (Columbus Day).

Computation
We teach the course in `R`, which is an open-source computing language that is very widely used in statistics. You can download it for free from www.r-project.org. The web provides many great tutorials and resources to learn R. A list of these is provided at http://wiki.math.yorku.ca/index.php/R:_Getting_started. A nice way to start you off are the two video tutorials provided by Dan Goldstein here and also here. `R` runs on a wide variety of UNIX platforms, Windows and MacOS - you can download and use it even if your computer is 10 years old. `R` makes programming very easy, has strong graphical capabilities, and also contains canned functions for most commonly used estimators.

If you are already well versed in another statistical software you can also use this for the course at your own risk.

Required Reading

The assigned required readings are listed on the syllabus for each topic. You should read this material closely!

Required Books

The following required textbooks are available at the COOP and will be used for the class.

Bertsekas, Dimitri and Tsitsiklis, John. *Introduction to Probability.* 2nd edition.

Wooldridge, Jeffrey. *Introductory Econometrics.* New York: South-Western. 4th edition.

To learn `R` you are expected to work through one of the following free tutorials (they cover similar material, just pick the one you like best):

Owen. *The R Guide.* At: http://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf
Venables and Smith. *An Introduction to R.* At: http://cran.r-project.org/doc/manuals/R-intro.pdf
Verzani. *Simple R.* At: http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf

Optional Books

The following books are optional but may prove useful to students looking for additional coverage of some of the course topics.

Other good textbooks:

Freedman, David; Robert Pisani; and Roger Purves. *Statistics.* 4rd Edition. New York: Norton.

Andrew, Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

For math background:

Gill, Jeff. *Essential Mathematics for Political and Social Research.* 1st Edition. 2nd printing. New York: Cambridge University Press.

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

For visualizing data (conceptual):

Cleveland, William S. *Visualizing Data.* Summit, NJ: Hobart Press.

Tufte, Edward. *The Visual Display of Quantitative Information, 2nd Edition.* Cheshire, CN: Graphics Press.

For visualizing data (implementation in R):

Murrell, Paul. *R Graphics.* Chapman & Hall.

Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis.* Springer.

Sarkar, Deepayan. *Lattice: Multivariate Data Visualization with R.* Springer.

Schedule

1 Introduction

Topics covered

- Overview and Course Requirements
- Course Outline

2 Probability Useful for Statistics

Topics covered

- Elementary Probability Theory
- Random Variables and Functions of Random Variables
- Elementary Asymptotics

Required Reading

Bertsekas and Tsitsiklis, Chapters 1-4

Wooldridge, Appendix A & B

3 Univariate Statistical Inference

Topics covered

- Sampling Distributions
- Point Estimation
- Interval Estimation
- Small and Large Sample Properties of Estimators
- Hypothesis Testing

Required Reading

Wooldridge, Appendix C

Bertsekas and Tsitsiklis, Chapter 5 & 9.1

4 What is Regression Analysis?

Topics covered

- Summarizing and Plotting Bivariate Data
- Marginal, Joint and Conditional Distributions
- Conditional Expectation
- Nonparametric Regression
- Bias-Variance Tradeoff
- Uses for Linear Regression

Required Reading

Wooldridge, Chapter 1

5 Simple Linear Regression

Topics covered

- Mechanics of Ordinary Least Squares
- Linear model assumptions
- Properties of the Least Squares Estimator
- Gauss-Markov Theorem
- Testing and Confidence Intervals
- Large Sample Inference

Required Reading

Wooldridge, Chapter 2

Bertsekas and Tsitsiklis, Chapter 9.2-9.5

Optional Reading

Tatem, Andrew J; Carlos A. Guerra; Peter M. Atkinson; and Simon I. Hay. 2004. "Momentous Sprint at the 2156 Olympics." *Nature* 431 (30 September): 525.

6 Linear Regression with Two Regressors

Topics covered

- Mechanics of OLS with Two Regressors
- Omitted Variable Bias
- Multicollinearity
- Dummy Variables
- Interaction Terms
- Polynomials and Logarithms

Required Reading

Wooldridge, Chapter 3-7

Optional Reading

Brambor, Thomas, William Clark, and Matt Golder. 2005. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis*. 13: 1-20.

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

7 Multiple Linear Regression

Topics covered

- Review of Matrix Algebra and Vector Calculs
- Mechanics of Multiple Linear Regression
- Statistical Inference for Multiple Linear Regression
- Testing Multiple Hypotheses

Required Reading

Wooldridge, Appendix D & E

8 Diagnosing and Fixing Problems I

Topics covered

- Nonconstant Error Variance
- Weighted Least Squares
- Heteroskedasticity-Robust Standard Errors
- Correlated Errors
- Generalized Least Squares

Required Reading

Wooldridge, Chapter 8

9 Diagnosing and Fixing Problems II

Topics covered

- Leverage Points
- Outliers
- Influence Points
- Nonnormality
- Nonlinearity

Required Reading

Wooldridge, Chapter 9

Optional Reading

Jackman, Robert W. 1987."The Politics of Economic Growth in the Industrial Democracies, 1974-80: Leftist Strength or North Sea Oil?" *The Journal of Politics*, Vol. 49, No. 1, pp. 242-256. (available via JSTOR)

Wand, Jonathan; Kenneth Shotts; Jasjeet Sekhon; Walter Mebane; Michael Herron; and Henry Brady. 2001 "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *APSR*. 95: 793-810.

Beck, Nathaniel and Simon Jackman. 1998. "Beyond Linearity by Default: Generalized Additive Models." *AJPS*. 42: 596-627.

10 Advanced Regression Topics

Topics covered

- Robust Regression (in particular MM-Estimation)
- Semi-parametric and Non-parametric Regression (in particular General Additive Models)
- Nonlinear Regression: Logit and Probit Models

Required Reading

Wooldridge, Chapter 17.1

Optional Reading

Rousseeuw, Peter J. and Bert C. van Zomeren. 1990. "Unmasking Multivariate Outliers and Leverage Points (with Discussion)." *JASA*. 85: 633-651.

Hampel, Frank. 2001. "Robust Statistics: A Brief Introduction and Overview." Link.