17.800: Quantitative Research Methods I

Fall 2012

Instructor: Teppei Yamamoto
TA: Mark Bell

Department of Political Science
MIT

Contact Information

	Teppei	Mark
Office:	E53–401	E40–444C
Email:	teppei@mit.edu	markbell@mit.edu
Website:	http://web.mit.edu/teppei/www	
Office Hours:	by appointment	T 12.30-2 in E53-485

Logistics

- Lectures: M & W 3:30–5 in 66–160

- Recitations: F 10–11 in E53–438

Note that we will have no class on October 8 (Columbus Day) and November 11 (Veteran's Day).

Overview and Goals

This is the first course in a four-course sequence on quantitative political methodology. Political methodology is a growing subfield of political science which deals with the development and application of statistical methods to problems in political science and public policy. The subsequent courses in the sequence are 17.802, 17.804, and 17.806. By the end of the sequence, students will be capable of understanding and confidently applying a variety of statistical methods and research designs that are essential for political science and public policy research.

This first course provides a graduate-level introduction to regression models, along with the basic principles of probability and statistics which are essential for understanding how regression works. Regression models are routinely used in political science, policy research, and other disciplines in social science. The principles learned in this course also provide a foundation for the general understanding of quantitative political methodology. If you ever want to collect quantitative data, analyze data, critically read an article which presents a data analysis, or think about the relationship between theory and the real world, then this course will be helpful for you.

You can only learn statistics by doing statistics. In recognition of this fact, the homework for this course will be *extensive*. In addition to the lectures and weekly homework assignments,

there will be required and optional readings to enhance your understanding of the materials. You will find it helpful to read these not only once, but 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

Willingness to work hard on unfamiliar materials. Understanding of the basic linear algebra and calculus equivalent to the contents covered in the department's math pre-fresher course. (If you did not complete the math pre-fresher, contact the instructor to see if you have enough background.) In addition, you will benefit more from the class if you have taken one (or more) undergraduate classes in quantitative methodology (e.g., MIT's 17.871, Harvard's Gov 1019 and 1020).

Course 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 (i.e. a typed and well organized write-up with all problems attempted) will be graded on the scale of $(+, \checkmark, -)$. You may re-write one assignment over the semester and have it regraded. If you choose to submit a re-write, it 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 up and submit your own solutions. We also require that you make a solo effort at all the problems before consulting others in your group, and that you write the names of your co-workers on your assignments.

The final assignment of the term will be a special problem set, which will be longer than a regular problem set and weighted more heavily toward the calculation of the final grade. You will not be allowed to collaborate with anybody on the final problem set. This is to test if you have developed sufficient experience to work through problems on your own. No rewrite is permitted on the final assignment.

The in-class, closed-book midterm will take place on October 15 during the regular class time. Plan accordingly.

Finally, please note that *no incompletes will be given in this course.*

Notes on Academic Integrity

Please respect and follow the rules written in MIT's handbook on academic integrity, which is available at:

http://web.mit.edu/academicintegrity/

In particular, the following is a (partial) list of the acts we will consider academically dishonest:

- Obtaining or consulting course materials from previous years

- Sharing course materials with people outside of the class, such as problem sets and solutions

- Copying and pasting someone else's answers to problem sets electronically, even if you collaborated with the person in a legitimate way (as specified above)

Recitation Sessions

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

Course Website

The course website is located at the following URL:

http://stellar.mit.edu/S/course/17/fa12/17.800/

This site will provide homework assignments, datasets, and links to reading materials.

Questions about Course Materials

In this course, we will utilize an online discussion board called *Piazza*. Below is an official blurb from the Piazza team:

> Piazza is a question-and-answer platform specifically designed to get you answers fast. They support LaTeX, code formatting, embedding of images, and attaching of files. The quicker you begin asking questions on Piazza (rather than via individual emails to a classmate or one of us), the quicker you'll benefit from the collective knowledge of your classmates and instructors. We encourage you to ask questions when you're struggling to understand a concept ... See this New York Times article to learn more about their founder's story:
> http://www.nytimes.com/2011/07/04/technology/04piazza.html

In addition to recitation sessions and office hours, please use the Piazza Q & A board when asking questions about lectures, problem sets, and other course materials. You can access the Piazza course page either directly from the below address or the link posted on the Stellar course website:

https://piazza.com/mit/fall2012/17800

Using Piazza will allow students to see other students' questions and learn from them. Both the TA and the instructor 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 or TAs* (unless they are of personal nature) — we will not answer them!

Notes on Computing

We teach this course in R, an open-source statistical computing environment that is very widely used in statistics and political science. You can download it for free from www.r-project.org. The web provides many great tutorials and resources to learn R: This list is a good list to start. A quick 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-based platforms (including Mac OS X), Windows and Linux – 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 are free to use it, but you will be on your own.

Books

Required Books

There will be required readings for each section of the course. Students are expected to complete them before the relevant materials are covered in the lectures. The following textbooks are required (available at the COOP) and will be used throughout the course.

- 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, along with the materials used in the math prefresher.

- 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. (statistics basics)

- Andrew, Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press. (regression modeling)

- Fox, John and Sanford Weisberg. *An R Companion to Applied Regression.* 2nd ed. (R, with focus on regression modeling)

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.

Course Schedule and Reading Assignments

1 Introduction

- Overview and Course Requirements
- Course Outline

2 Elementary Probability Theory

- Why Do We Need Probability?
- Probability Axioms
- Marginal, Joint and Conditional Probability
- Law of Total Probability
- Bayes' Rule
- Independence

Required Readings:

- Bertsekas and Tsitsiklis, Chapter 1
- Wooldridge, Appendix A

3 Random Variables and Probability Distributions

- Discrete and Continuous Random Variables
- Measures of Location
- Measures of Dispersion
- Probability Distributions

Required Readings:

- Bertsekas and Tsitsiklis, Chapters 2.1–2.4 & 3.1–3.3
- Wooldridge, Appendix B.1 & B.3

4 Multiple Random Variables

- Joint and Conditional Distributions
- Conditional Expectation
- Covariance and Independence

Required Readings:

- Bertsekas and Tsitsiklis, Chapters 2.5–2.8, 3.4–3.7, 4.2 & 4.3
- Wooldridge, Appendix B.2 & B.4–B.5

5 Univariate Statistical Inference

5.1 Point Estimation

- Properties of Estimators

- Sampling Distribution

- Elementary Asymptotic Theory

5.2 Interval Estimation

- Confidence Intervals

5.3 Hypothesis Testing

- Logic of Statistical Testing

- p-Values

Required Readings:

- Wooldridge, Appendix C

- Bertsekas and Tsitsiklis, Chapter 5

6 What is Regression?

- Nonparametric Regression

- Linear Regression

- Bias-Variance Tradeoff

Required Readings:

- Wooldridge, Chapter 1

7 Simple Linear Regression

- 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 Readings:

- Wooldridge, Chapter 2

Optional Readings:

- 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.

8   Linear Regression with Two Regressors

8.1   Mechanics of Regression with Two Regressors

- Motivation for Multiple Regression

- Mechanics and Inference in OLS with Two Regressors

8.2   Omitted Variables and Multicollinearity

- Omitted Variable Bias

- Multicollinearity

8.3   Dummy Variables, Interactions and Polynomials

- Dummy Variables

- Interaction Terms

- Polynomials and Logarithms

Required Readings:

- Wooldridge, Chapters 3–7

9   Multiple Linear Regression

9.1   Mechanics of Multiple Regression

- Review of Matrix Algebra and Vector Calculus

- Mechanics of Multiple Linear Regression

9.2   Statistical Inference with Multiple Regression

- Statistical Inference for Multiple Linear Regression

- Testing Multiple Hypotheses

Required Readings:

- Wooldridge, Appendix D & E

10 Diagnosing and Fixing Problems in Linear Regression

10.1 Outliers and Influential Observations

- Plotting residuals

- Standardized and studentized residuals

- Added variable and component residual plots

- Leverage and influence

10.2 Heteroskedasticity, Serial Correlation and Clustering

- Weighted least squares

- Generalized least squares

- Heteroskedasticity-robust standard errors

- Cluster-robust standard errors

- Autocorrelation

10.3 Measurement Error

- Types of measurement errors

- Measurement error in the dependent variable

- Measurement error in an independent variable

Required Readings:

- Wooldridge, Chapters 8–9

Optional Readings:

- 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.

11 Extensions and Advanced Topics (time permitting)

- Nonlinear Regression Models

  - Logit and Probit Models
  - Generalized Linear Models

- Semiparametric and Nonparametric Regression Models

– Generalized Additive Models

Required Readings:

- Wooldridge, Chapter 17.1

Optional Readings:

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