

# 17.800: Quantitative Research Methods I

Fall 2019

Instructor: F. Daniel Hidalgo

TAs: Matias Giannoni & Rori Lekalake

Department of Political Science

MIT

## Contact Information

	Danny	Matias	Rori
Office:	E53-401	E53-434	E53-418
Email:	<a href="mailto:dhidalgo@mit.edu">dhidalgo@mit.edu</a>	<a href="mailto:giannoni@mit.edu">giannoni@mit.edu</a>	<a href="mailto:lekalake@mit.edu">lekalake@mit.edu</a>
Office Hours:	Th 3-5 pm	T 4-6 pm	M 2-4 pm

## Logistics

- Lectures: M & W 9:30–11:00 in [E51-372](#)
- Recitations: F 9:00–10:00 in [E51-361](#)
- TA Office Hours: M 2:00–4:00, T 4:00–6:00 in [E53 3rd Floor Lounge](#)

Note that the first class meets on September 4, and the last class meets on December 11. There will be no class/recitation on the following days: Friday, September 20 (Student Holiday); Monday, October 14 (Columbus Day); Monday, November 11 (Veterans Day); Wednesday, November 27 (Day before Thanksgiving); Friday, November 29 (Day after Thanksgiving)

Please also note that enrollment is capped at 30 students due to capacity constraints and priority is given to political science graduate students. Therefore we cannot guarantee a spot for students from other departments. The available spots will be assigned by a lottery in the first week of class in case there is excess demand.

## 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 graduate students from other departments. Qualified undergraduates can also take the course subject to permission of the instructors.

## 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., 17.803).

## Course Requirements

Grades will be based on:

- Weekly homework assignments (45% of final grade)
- Final homework assignment (20% of final grade)
- Three in-class quizzes (30% of final grade)
  - Monday, September 30
  - Monday, October 28
  - Monday, December 2
- Participation (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. All assignments must be submitted electronically through the class Stellar site. No late homework will be accepted unless you ask for special permission from the instructor in advance of the deadline. All sufficiently attempted homework (i.e. a typed and well organized write-up with all problems attempted) will be graded on a three-point scale ( $\checkmark^+$ ,  $\checkmark$ ,  $\checkmark^-$ ). 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. For the writeup, we recommend that you use L<sup>A</sup>T<sub>E</sub>X or [R Markdown](#)

over Word, which is not well suited to displaying code and math. 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.

There will be three closed-book in-class quizzes. The first quiz will take place on September 30th during the regular class time, the second on October 28th, and the third on December 2nd.

If you think that there is an error in grading your problem set or quiz answer, you should always start by reading the solution carefully one more time, making sure you understand it well. Then, read your own answer *as if you were reading it for the first time*, and ask yourself whether any of the differences might reasonably have caused the point reduction. Note that it is what you actually wrote in your answer, not what you *intended* to communicate by writing your answer, that should be considered a reasonable basis for the possible grading error. That is, your answer needs to be self-contained and cannot be supplemented by any information you provide outside of your answer sheet.

If you still think there was truly an error in your grade, you can request a re-grading by emailing one of us about your intent. The TA who graded your answer will get in touch with you and discuss your concerns. Unless the request is about a simple factual error the TA can resolve without any uncertainty (e.g. scores were not added up correctly), the TA will forward your entire problem set or quiz for re-grading by either the other TA or the instructor. The new grader will re-grade the whole problem set or quiz anew, not only the particular problem you had a dispute about, without regard to the original scores. The new grade will then replace your original grade for the problem set or quiz and become final, regardless of whether it is higher or lower.

Finally, please note that *no incompletes will be given in this course* except in highly unusual circumstances.

## 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 and TA Help

Weekly recitation sessions will be held on Friday mornings, 9–10 in **E51-361**. 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.

In addition to recitation, each TA will hold office hours once per week in the 3rd floor lounge of E-53. Questions will be addressed on a first-come, first-served basis.

The TAs for the course are very valuable resource for getting help on homework and understanding the material. While you are encouraged to talk with the TAs in office hours or via piazza if you run into problems with any aspect of the class, please respect their time. In general, TAs are only expected to answer questions submitted via piazza within 24 hours and they are not expected to answer queries during the weekend. In addition, in person queries should be restricted to the TA's posted office hours or recitation.

## Course Website

The course website is located at the following URL:

<http://stellar.mit.edu/S/course/17/fa19/17.800/>

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

## Questions about Course Materials

In this course, we will utilize an online discussion board called *Piazza*. This is a question-and-answer platform that is easy to use and designed to get you answers to questions quickly. We encourage you to use the Piazza Q & A board when asking questions about lectures, problem sets, and other course materials outside of recitation sessions and office hours. 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/fall2019/17800>

Using Piazza will allow students to see other students' questions and learn from them. Both the TAs 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](http://www.r-project.org). We recommend that you use the RStudio IDE (integrated development environment) to work with R, which can be downloaded for free from <http://www.rstudio.com>.

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 is the R tutorial created by *Data Camp*: [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 statistical procedures. Teaching materials of R are available at the course website of the department's math pre-fresher:

<http://stellar.mit.edu/S/project/mathprefresher/materials.html>

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 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. 5th edition.

To learn R you are expected to work through the following tutorial, along with the materials used in the math preresher: *Introduction to R* by Data Camp.

### 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.