17.800: Quantitative Research Methods I

Fall 2024

Instructor: F. Daniel Hidalgo

TAs: Serene Ho & Joseph Loffredo

Department of Political Science

MIT

Contact Information

	Danny	**Serene**	**Joe**
Office:	E53–402	E53–436	E53-365
Email:	dhidalgo@mit.edu	sereneho@mit.edu	loffredo@mit.edu
Office Hours:	MW 11:00–12:00 pm	M 11:00–12:00 pm	Tu 12:45-2:10 pm

Logistics

- **Lectures:** M & W 9:30–11:00 am in E51–057
- **Recitations:** F 9:00–10:00 am (location to be determined). The first recitation on Friday, Sept. 13.
- **TA Office Hours:** M 11:00am - 12:00 pm, Tu 12:45-2:10 pm pm in E53 3rd Floor Lounge

Note that the first class meets on Wednesday, September 4, and the last class meets on Wednesday, December 11. There will be no class/recitation/office hours on the following days:

- Friday, September 20 (Student Holiday)
- Monday, October 14 (Indigenous Peoples Day)
- Monday, November 11 (Veterans Day)
- Wednesday, November 27 (Day before Thanksgiving)
- Friday, November 29 (Day after Thanksgiving)

Enrollment is capped at 30 students due to capacity constraints, with priority given to political science graduate students. In case of excess demand, students will be notified whether they are permitted to remain enrolled in the class by the end of the second week.

Overview and Goals

This course is the first in a four-course sequence on quantitative political methodology. Political methodology is a growing subfield of political science that develops and applies statistical methods to problems in political science and public policy. Subsequent courses in the sequence are 17.802, 17.804, and 17.806. By the end of the sequence, students will understand and confidently apply a variety of statistical methods and research designs essential for political science and public policy research.

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

You can only learn statistics by doing statistics. Therefore, the homework for this course will be extensive. In addition to 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 may take the course with the instructor's permission.

Prerequisites

Willingness to work hard on unfamiliar materials. Understanding of 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 sufficient background. Additionally, prior undergraduate courses in quantitative methodology (e.g., 17.803) are beneficial.

Course Requirements

Grades will be based on:

- Weekly homework assignments (45%)
- Final homework assignment (20%)
- Three in-class quizzes (30%)
  - Monday, October 7
  - Monday, November 4
  - Monday, December 2
- Participation (5%)

The weekly homework assignments will consist of analytical problems, computer simulations, and data analysis. They will usually be assigned on Wednesday and due the following Wednesday before the lecture. All assignments must be submitted electronically through the class Canvas site.

No late homework will be accepted unless you request 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 ($\checkmark^+$, $\checkmark$, $\checkmark^-$). Generally speaking, you will receive a $\checkmark^+$ for a few minor mistakes, a $\checkmark$ for one or two major mistakes, and a $\checkmark^-$ for more than two major mistakes. Some problem sets will have extra credit problems, which can improve your grade.

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.

Collaboration on assignments is encouraged, but you must write up and submit your own solutions. For the writeup, we recommend using LaTeX, R Markdown, or Quarto over Word, which is not well suited to displaying code and math. You are also required to make a solo effort at all problems before consulting others in your group, and you must write the names of your collaborators on your assignments.

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

There will be three closed-book **in-class quizzes** conducted on the dates listed above. The quizzes will occur during the first 30 minutes of the regular class time on those dates.

Although this course is primarily lecture-based, **participation** is important. We expect you to attend lectures and recitation sessions regularly. During sessions, we encourage you to participate actively by asking questions and responding to our questions. Participation in the Piazza discussion board (see below) will also count towards the participation grade.

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

Regrading Requests

If you think there is an error in grading your problem set or quiz, start by reading the solution carefully again to ensure you understand it. Then, read your own answer *as if you were reading it for the first time*, and consider whether 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, that should be considered a reasonable basis for the possible grading error. Your answer needs to be self-contained and cannot be supplemented by any information you provide outside of your answer sheet.

If you still believe 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 contact you to 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 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 entire problem set or quiz, not just the disputed problem, and the new grade will replace the original, whether it is higher or lower. 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.

Notes on Academic Integrity

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

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

In particular, the following is a (partial) list of acts considered 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, 9:00–10:00 am in E51–057. The first recitation on Friday, Sept. 6 will be in the Millikan Room E53–482. Recitations will review theoretical material and provide help with computing issues. The teaching assistant will run the sessions and can provide more details. Attendance is strongly encouraged.

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 a valuable resource for getting help on homework and understanding the material. While you are encouraged to talk with the TAs during office hours or via Piazza if you encounter 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 are not expected to answer queries during the weekend. Additionally, 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:

https://canvas.mit.edu/courses/27073

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

Video Recording of Lectures

We will participate in IST's Lightweight Lecture Capture program, which will automatically video record the lectures and post them to the class website. The videos will be available to students in the course but should not be distributed to anyone outside the class. If you have any concerns about the video recording, please contact the instructor.

Even though the course lectures will be recorded, we strongly encourage you to attend the lectures in person. The video will not be available until after the lecture, so you will miss out on the opportunity to ask questions during class. If lecture attendance declines to a low level for a sustained period, we may suspend video recordings of lectures.

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 Canvas course website:

https://piazza.com/mit/fall2024/17800

Using Piazza allows 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 towards his/her class participation grade. *Do not email your questions directly to the instructors or TAs* (unless they are of a personal nature) — we will not answer them!

Notes on Computing

We teach this course in `R`, an open-source statistical computing environment widely used in statistics and political science. You can download it for free from 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 starting point. A quick and nice way to start 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 contains functions for most commonly used statistical procedures. Teaching materials for `R` are available at the course website of the department's math pre-fresher:

https://stellar.mit.edu/S/project/math-camp/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.

Use of Large Language Models

Large language models (LLMs) are quickly transforming how social science research is conducted. Given that LLMs and related technologies will likely be part of your future research, I do not ban their use in this course. Particularly with respect to computation, however, the over-reliance on LLMs can be detrimental to your understanding of how your code works. In many cases, LLMs will give you working code, but you may not understand why it works or when it is actually not doing what you want. Therefore, I strongly encourage you to attempt the homework problems without the use of LLMs, and only use them to check your work or to help you when you are stuck.

With respect to written work and analytical derivations, I expect your words and proofs to be your own. You may use LLMs to check grammar or catch typos, but you should not use them to generate content. If you are unsure about what constitutes appropriate use of LLMs, please ask me.

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.

Optional Books

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

Other good textbooks:

- Freedman, David, Robert Pisani, and Roger Purves. *Statistics*. 4th Edition. New York: Norton. (An excellent introductory book)
- Andrew, Gelman, Jennifer Hill, and Aki Vehtari. *Regression and Other Stories*. Cambridge University Press. (Regression modeling)
- Fox, John and Sanford Weisberg. *An R Companion to Applied Regression*. 3rd ed. (Focus on regression modeling in `R`)
- Aronow, Peter and Benjamin Miller. *Foundations of Agnostic Statistics*. Cambridge University Press. (A recent text that covers the theoretical materials in some depth)

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.

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

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