POLI450A: Political Methodology I
(Fall 2024)

Professor: Jens Hainmueller
TAs: Abhinav Ramaswamy and Naiyu Jiang

Class: Mon/Wed 10:30-11:50am; Encina West 219
Section: Friday, time 9:30-11:20; Encina West 219

Contact Information

	Jens Hainmueller	Abhinav Ramaswamy	Naiyu Jiang
Email:	jhain@stanford.edu	acr122@stanford.edu	nyjiang@stanford.edu
Office:	Encina West 409	Encina West 216	Encina West 216
Office Hours:	Book Appointment (or slack if time does not work)	Thursdays 1-3pm	Wednesdays 12:30-2:30pm
Appointment:	calendly.com/jhainmueller	calendly.com/acr122	nyjiang.youcanbook.me

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 450B, 450C, and 450D. 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).

Non-Political Science PhD students must pass an entrance exam to enroll in this class. The entrance exam will be held on Friday 9/20, 9am-12pm. Please contact Sarah Nelson at snelson4@stanford.edu for exam logistics and study materials.

Prerequisites

Willingness to work hard on unfamiliar materials. Understanding of basic linear algebra, calculus, probability, and statistical computing equivalent to the contents covered in the department's math camp course. Students typically have taken one (or more) undergraduate classes in quantitative methodology.

Course Requirements

Grades will be based on

- course participation (5% of final grade)
- homework assignments (30% of final grade)
- midterm exam (30% of final grade)
- final exam (35% of final grade)

The weekly homework assignments will consist of analytical problems, computer simulations, and data analysis. They will usually be assigned on Monday night and due the following Monday, prior to lecture. Homework should be submitted on Canvas. 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 highly 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 write the names of your co-workers on your assignments.

The midterm exam will take place on Monday, October 14. The final exam will be scheduled by the registrar during exam week. Both exams are in-person and closed-book, and students must attend them at the scheduled times. Please plan accordingly.

You will not be allowed to collaborate with anybody on the midterm and final exams. This is to test if you have developed sufficient skills to work through problems on your own. No re-write is permitted on the exams.

Finally, please note that *no incompletes will be given in this course* and that anyone considering auditing the course must complete and submit all homework assignments.

Notes on Academic Integrity

Please respect and follow the rules described in Stanford's Honor code, which is available at:

https://communitystandards.stanford.edu/student-conduct-process/honor-code-and-fundamental-standard

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)

Policy on Using ChatGPT

In an effort to keep up with modern teaching methods and to facilitate an effective learning experience, our class has integrated the use of ChatGPT into its curriculum. The following policy outlines the extent to which students may use ChatGPT in the classroom, for homework, and during exams.

- **In-Class Usage**
  - *Permitted Activities:* Students may consult ChatGPT for clarifications on coding problems, syntax, or other related coding queries. Students can seek guidance on statistics-related issues from ChatGPT.

  - *Prohibited Activities:* Directly copying code or statistical methodologies without attempting the problem first is discouraged. Reliance solely on ChatGPT for answering in-class questions without engagement in class discussions is not allowed.

- **Homework Assignments**
  - *Permitted Activities:* Students are allowed to consult ChatGPT to assist with their homework assignments. If guidance from ChatGPT is utilized, it should serve as a supplementary tool and not as the primary source of information.
  - *Mandatory Reporting*: If a student uses ChatGPT for assistance on a homework assignment, the degree of usage should be detailed explicitly. This includes specifying which portions of the assignment were completed with the help of ChatGPT. Such details should be stated at the beginning of the assignment in a separate section titled **ChatGPT Assistance**. Using ChatGPT appropriately for assignments, as outlined, will not incur any grading penalties.
  - *Prohibited Activities:* Submitting work solely generated by ChatGPT as one's own without any personal understanding or contribution is a breach of this policy.

- **Examinations**
  - ChatGPT is strictly prohibited during the midterm and final exam. **Any** use of ChatGPT in these exams will be considered as violating the Honor Code.

Sections

Weekly sections will be held by the TAs. The sessions will cover a review of the theoretical material and also provide help with computing issues. The teaching assistants will run the sessions and can give more details. Attendance is required.

Course Website

In this course, we will utilize a Slack workspace as the course website. We will use this site to ask questions and also to provide homework assignments, datasets, and links to reading materials.
If you are new to Slack, you can find instructions for how to set it up and join the workspace here.
Below is the official description of the Stanford Slack workspaces:

> Your Stanford Slack account allows you to instantly share messages, links, and files with individuals or groups through selected channels in a workspace, as well as contact people by voice and video call. You can tag individuals or entire channels to get people's attention.

Using Slack will allow students to see other students' questions and learn from them. Both the TAs and the instructor will regularly check the channels 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 a 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 are already familiar with `R` from math camp.

Books

Required Books

Required readings for each section of the course are listed below. Students are expected to complete these readings before the relevant materials are covered in the lectures. The following textbook is required and will be used throughout the course:

- Wooldridge, Jeffrey. *Introductory Econometrics*. New York: South-Western. 6th edition. (5th edition will also work for the course)

We also scanned selected chapters from two alternative textbooks that are helpful. These chapters are posted on the course website. Students who liked these chapters might want to read the entire books in parallel to the Wooldridge book. They cover similar material.

- Peter M. Aronow and Benjamin T. Miller. *Foundations of Agnostic Statistics*. Cambridge University Press.
- R. Carter Hill, William E. Griffiths, and Guay C. Lim. *Principles of Econometrics*. 4th edition.
- John Fox. *Applied Regression Analysis and Generalized Linear Models*. 2nd edition.

To learn `R` you are also required to work through the following book:

- Paul Teetor. 2011. *R Cookbook*. O'Reilly Media.

In addition you can also work through other free R books such as:

- 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:

- 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

First day of class: Sept 23
No classes on Nov 25-29 (Thanksgiving recess)

1 Univariate Statistical Inference

1.1 Point Estimation

- Properties of Estimators
- Sampling Distribution
- Elementary Asymptotic Theory

Required Readings:

- Wooldridge, Appendix C1-C4
- Unless you are well familiar with this material you need to review: Wooldridge, Appendix A & B

1.2 Interval Estimation

- Confidence Intervals

Required Readings:

- Wooldridge, Appendix C5

1.3 Hypothesis Testing

- Logic of Statistical Testing
- p-Values

Required Readings:

- Wooldridge, Appendix C6-C7

2 What is Regression?

- Nonparametric Regression
- Linear Regression
- Bias-Variance Tradeoff

Required Readings:

- Wooldridge, Chapter 1

3 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
- Alternative: Hill, Griffiths, and Lim, Chapters 2 & 3 (course website)

4 Linear Regression with Two Regressors

4.1 Mechanics of Regression with Two Regressors

- Motivation for Multiple Regression
- Mechanics for OLS with Two Regressors

Required Readings:

- Wooldridge, Chapter 3
- Inference for OLS with Two Regressors

Required Readings:

- Wooldridge, Chapter 4 & 5

4.2 Omitted Variables and Multicollinearity

- Omitted Variable Bias
- Multicollinearity

Required Readings:

- Wooldridge, Chapters 6
- Alternative: Hill, Griffiths, and Lim, Chapter 6 (course website)

4.3 Dummy Variables, Interactions and Polynomials

- Dummy Variables
- Interaction Terms
- Polynomials and Logarithms

Required Readings:

- Wooldridge, Chapter 7
- Alternative: Hill, Griffiths, and Lim, Chapter 4 & 7 (course website)

5 Multiple Linear Regression

5.1 Mechanics of Multiple Regression

- Review of Matrix Algebra and Vector Calculus
- Mechanics of Multiple Linear Regression

5.2 Statistical Inference with Multiple Regression

- Statistical Inference for Multiple Linear Regression
- Testing Multiple Hypotheses

Required Readings:

- Wooldridge, Appendix D & E

6 Diagnosing and Fixing Problems in Linear Regression

6.1 Outliers and Influential Observations

- Plotting Residuals
- Standardized and Studentized Residuals
- Added Variable and Component Residual Plots
- Leverage and Influence

Required Readings:

- Fox, Chapter 11 (course website)

6.2 Heteroskedasticity, Serial Correlation and Clustering

- Weighted Least Squares
- Heteroskedasticity-robust Standard Errors
- Cluster-robust Standard Errors
- Autocorrelation

Required Readings:

- Wooldridge, Chapters 8–9
- Fox, Chapter 12 (course website)
- Alternative: Hill, Griffiths, and Lim, Chapter 8 (course website)