POLI450A: Political Methodology I
(Fall 2026)

Yiqing Xu, Stanford University

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).

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

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 benefit more from the class if they 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 (35% of final grade)
- midterm exam (25% 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 **Wednesday** night and due the following **Tuesday** by 11:59pm. 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 quarter 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 takes place in **Week 6**. The final exam takes place in the **Exam Week**.

The midterm and final exams will be closed book. You may bring one two-sided cheat sheet. You will not be allowed to collaborate with anybody on the midterm and final exams. This is to test if you have developed sufficient experience 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.

Variable unit distinction. Each homework assignment will consist of two parts. The first part, required of all students, focuses on foundational proofs and data analysis; students complete it without the assistance of AI, writing and running their own code. The second part is one harder problem, set deliberately beyond the lectures so that students learn something the course does not teach; students are encouraged to use AI as a tool in developing their solutions (see the Policy on Using GenAI). Students enrolled in fewer than five units are not required to complete the second part.

Submission. Students upload one PDF with both parts, typed in LaTeX, Quarto or R Markdown, together with their `R` code (a script file, or code shown in the PDF). The second part ends with one paragraph on the student's AI use: how the AI helped, and where it was not enough on its own. Students must be able to explain everything they submit.

Active Learning

In this course, we will use active learning to reinforce both conceptual and technical aspects of regression analysis. Each class will include a short worksheet, which consists of in-class exercises, where we will pause the lecture and allow time for you to work through problems directly. These worksheets are not graded, but they are designed to deepen understanding, provide opportunities for immediate practice, and bridge the gap between coding-oriented homework assignments and exams, which place greater emphasis on conceptual reasoning and proofs.

Laptops and phones. Lectures are laptop-free and phone-free. Please keep both put away during class, so that you can follow the derivations on the whiteboard and work through the worksheets. Tablets used for taking notes, such as an iPad or a reMarkable, are fine. The rule applies to lectures only, not to the sections.

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, including by posting them on a public website or repository (uploading course materials to an AI tool for your own study is allowed; see the Policy on Using GenAI)
- 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 GenAI

Generative AI tools such as ChatGPT, Claude Code, and Google Gemini can be valuable for learning, but they are most useful when students have first developed the ability to reason through fundamental problems on their own. In this course, the goal is therefore not to avoid AI altogether, but to use it selectively and in ways that strengthen rather than replace independent problem-solving skills.

During class, students are expected to follow the lecture closely, with particular attention to the derivations and arguments developed on the whiteboard. The pace and structure of the course are designed to make active engagement with these derivations an important part of the learning process. Students may use AI as an occasional supplementary resource for clarification, but it should not distract from following the lecture or working through the material as it is presented.

Homework assignments are designed to develop both independent problem-solving skills and the ability to use modern computational tools effectively. For foundational problems, including basic proofs, derivations, and data analysis, students are strongly encouraged to work without AI assistance. These problems are intended to build fluency with the core concepts and techniques that students will be expected to reproduce independently on examinations.

The second part of each assignment is one harder problem, set deliberately beyond the lectures so that working on it teaches something the course itself does not: sequential testing, inference with a handful of clusters, measurement error, the numerical linear algebra inside `lm()`. On these problems students are encouraged to use AI as a tool, and the ideal use is a tutor and a collaborator rather than an answer machine: ask it to propose approaches, ask why they work, test its output against known answers, and make the judgment calls it cannot make. Each assignment therefore ends with one paragraph on how the AI helped and where it was not enough on its own. When using AI, students remain responsible for understanding, checking, and being able to explain the work they submit; an AI-generated solution that a student cannot independently evaluate or explain is not a satisfactory solution.

Students may upload the course's lecture slides, handouts, and problem sets to AI tools for their own study, for example to ask about a slide they did not follow. This use does not count as "sharing course materials with people outside of the class" under the Notes on Academic Integrity. It does not cover posting course materials or solutions where others can see them, or uploading the textbook chapters and readings provided in the course.

Midterm and final examinations will be completed without the use of Generative AI. Students will therefore be assessed directly on their own understanding of the fundamental concepts, proofs, derivations, and methods developed in the course. This is one reason that independent practice on the foundational portions of the homework is especially important: students who rely on AI for skills they have not yet mastered are likely to be at a disadvantage on the exams.

Sections

Sections will be held once a week. The sessions 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 required.

Course Website

The slides, worksheets, short videos, R code, and data are on the course site:

https://teaching.yiqingxu.org/regression-lectures/

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 will also scan or provide selected chapters from the following textbooks that are helpful. These chapters are provided to students. Students who liked these chapters might want to read the entire books in parallel to the Wooldridge book. They cover similar material.

- P. Aronow and Benjamin T. Miller. *Foundations of Agnostic Statistics*. Cambridge University Press.
- Peng Ding, *Linear Model and Extensions*. arXiv preprint arXiv:2401.00649.
- Matthew Blackwell. *A User's Guide to Statistical Inference and Regression*. Free online: https://mattblackwell.github.io/gov2002-book/
- 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*. 4th Edition. New York: Norton. (statistics basics)
- Gelman, Andrew 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, CT: 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 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

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, Chapters 4 & 5

4.2 Omitted Variables and Multicollinearity

- Omitted Variable Bias
- Multicollinearity

Required Readings:

- Wooldridge, Chapter 6
- Alternative: Hill, Griffiths, and Lim, Chapter 6

4.3 Dummy Variables, Interactions and Polynomials

- Dummy Variables
- Interaction Terms
- Polynomials and Logarithms

Required Readings:

- Wooldridge, Chapter 7
- Alternative: Hill, Griffiths, and Lim, Chapters 4 & 7

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

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
- Alternative: Hill, Griffiths, and Lim, Chapter 8