

450A Political Methodology I

Lecture 1: Introduction

Stanford University

- Why are we here? (learning goals)
- What is it about? (big picture)
- How do we do it? (logistical matters)

What is Political Methodology?

- Methods subfield of political science, a relative of econometrics, psychometrics, biostatistics, sociological methodology, etc.
- Historically, political methodologists were trained in many different areas, and so the field is heavily interdisciplinary.
- As the crossroads for other disciplines, it is one of the best places to learn about methods broadly. It reflects the diverse nature of political science.

What is Political Methodology?

- An APSA organized section (Political Methodology, Section 10), founded in 1986
- Extremely valuable for the job market
- PolMeth mailing list and annual conference (<https://polmeth.org/mailling-list>)
- *Political Analysis* ranked 2nd of 332 political science journals by 2025 impact factor (Clarivate JCR)

- Math Camp
- 450A: Linear regression (Foundation)
- 450B: Causal Inference, with a focus on design-based methods
- 450C: Machine Learning and Model-based Inference
- 450D: Advanced Topics
- ...
- ECON 272: Applied Econometrics
- ECON 293: Machine Learning and Causal Inference

This Course

- A graduate-level introduction to regression analysis
- After this course, you will be:
 - Capable of conducting and interpreting regression analysis
 - Able to critically read and understand many quantitative empirical studies published in academic journals
 - Prepared to learn more advanced statistical methods

We're All "Methodologists"

- New methodologies emerge regularly
- Innovations in ML, AI, and large language models will transform social science
- Novel data types, such as text and images, demand new methods
- Techniques and "routines" you develop can benefit others
- Numerous user-friendly tools are at your disposal
- Improving your "outside options" can lead to a happier research life
- Diversifying your researcher portfolio (a little) may also be beneficial

Logistics

Class Time

- Two 100-minute lectures a week, plus a weekly section led by the TA
- Lectures are laptop-free and phone-free; a tablet for note-taking (iPad, reMarkable, etc.) is fine

Who Should Take This Course?

- Take this course if you want to:
 - analyze social science data in your research
 - understand regression *rigorously, but not as pure mathematics*
 - be a political methodologist and make a solid start
 - be more proficient in R
 - * Class is open to PhD students from other departments, but you need to meet the prerequisites
- This course may not be for you if:
 - You just want a cookbook-style introduction to regression analysis
 - You have taken a Ph.D.-level econometrics/statistics sequence

What are the Prerequisites?

- Math (linear algebra and calculus) equivalent to the contents covered in the department's math camp course
- Some undergraduate probability, statistics, 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.
- **By this point, you should already have either completed the math camp or passed the qualifying exam for this course.**

Requirements

- Course participation, 5% of the final grade
- Weekly homework assignments, 35% of final grade
 - Posted weekly and due six days later
 - Working in groups is encouraged, so long as you write up your work on your own and list co-workers
 - Graded on (+,√,-) scale; rewrite policy
- Midterm exam (Week 6), closed-book (one two-sided cheat sheet), 25% of final grade
- Final exam, closed-book (two two-sided cheat sheets), 35% of final grade

- **Homework** comes in two parts
 - Part 1, foundational proofs and data analysis: we strongly encourage you to work without AI; the exams test these skills
 - Part 2, one harder problem, deliberately beyond the lectures: AI is encouraged; end with one paragraph on how it helped and where it fell short
 - Taking fewer than 5 units? Part 2 is not required
- **Ideal use:** a tutor and a collaborator, not an answer machine — ask why, check its work against known answers, make the judgment calls yourself
- **Examinations**
 - GenAI is strictly prohibited during exams. Any use of GenAI in such exams will be considered as violating the Honor Code.

- Yiqing Xu

Research interest: causal inference (panel data); AI-assisted social science research; comparative politics (China)

How will I learn?

- Weekly sections led by the TA (attendance required)
- If you have questions...
 - Ask during lectures and sections
 - Try GenAI (esp. related to coding)
 - Come to office hours
- **Please do not email us questions directly** except if it is of personal nature

- Active learning will be used to reinforce both conceptual and technical aspects of regression.
- Each class includes a worksheet (not graded), with time to work through problems during lecture.
- The goal is to deepen understanding and bridge the gap between coding-heavy homework and concept- and proof-oriented exams.

- Open Source
- Statistical programming language + canned packages
- Relatively steep start up costs, but large long-term gains for learning
- Programming
- Graphics
- Data handling
- Many free tutorials and reference books

Strongly encouraged: LaTeX / RMarkdown for problem sets.

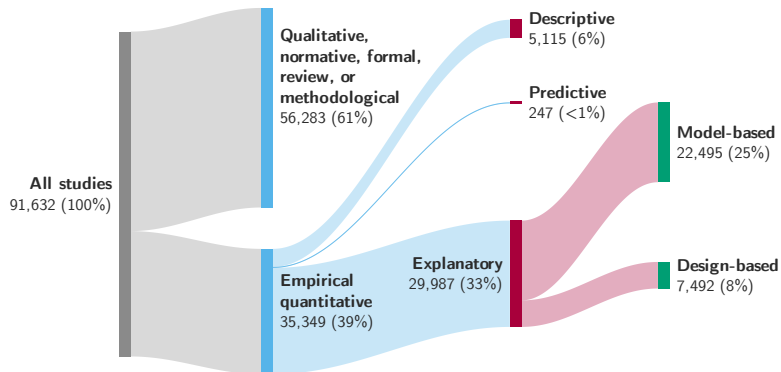
- Required textbooks:
 - Wooldridge, Jeffrey. *Introductory Econometrics: A Modern Approach*.
 - Readings listed on syllabus. Need to follow along each week.
- Selected chapters from:
 - Peng Ding, *Linear Model and Extensions*. arXiv preprint arXiv:2401.00649.
 - P. M. Aronow and Benjamin T. Miller. *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.

This Course

Social Science Objectives

- What do political scientists do in their empirical work?
- Answer three types of questions:
 - Descriptive
 - Predictive
 - Causal (explaining)

Social Science Objectives



Model-based: no clear research design; the causal claim relies on modeling assumptions. **Design-based:** a clear research design; the causal claim relies on how the cause was assigned.

Source: Torreblanca, Dinneen, Grossman & Xu (2026), "The Credibility Revolution in Political Science," arXiv:2601.11542, Figure 2: 91,632 articles in 156 political science journals, 2003–2023, classified with large language models. Percentages are shares of all 91,632 articles.

We ask questions to make a descriptive inference—that is, learn about some unobserved phenomenon on the basis of some set of observed facts

- Are elections for the U.S. House of Representatives less competitive in the last several decades than previously?
- Do civil wars last longer in countries with rough terrain?

de Kadt & Grzymala-Busse (2025), “Good Description” (SocArXiv, doi:10.31235/osf.io/e74a9_v2):

- Uncover facts in need of explanation, build theory, or help to evaluate and revise theory
- Clear, comparable, and complete

We ask questions about observations not in our data

- Who will win the Democratic primary?
- Which party will control the U.S. House after the next midterm elections?

We ask questions about counterfactuals.

- Would citizens have voted at a higher rate if a campaign worker had contacted them in person?
- Would France have developed more quickly if it had developed institutions of limited government in the late 17th and early 18th century?

Questions of the form what would have happened to the behavior or outcomes of people, cities or states if they had been subjected to some stimulus, T .

Course Plan

- Univariate Statistical Inference
- What is Regression?
- Simple Linear Regression
- Multiple Linear Regression
- Diagnosing and Fixing Problems
- Extensions