POLSCI 630: Introduction to Regression Methods

Lectures: Mo & We at 3:20 pm to 4:10 pm
Instructor: Jiawei Fu (jiawei.fu@duke.edu)
TA:

Location:
Office Hour: [link], Gross Hall 294A or online
Office Hour:

Course Description

This is a required graduate-level methods course that provides a rigorous and detailed introduction to the empirical techniques used in contemporary research. Regression methods serve as the cornerstone for using data to answer research questions. Even amid the recent revolution in causal inference, regression remains a foundational tool. Therefore, this course lays the groundwork for mastering more advanced methodological approaches in the future.

This course emphasizes both the theoretical foundations and the practical implementation of empirical methods using real data. As a graduate-level class, it covers not only core theoretical concepts but also recent advances in the literature. The accompanying lab sessions provide hands-on experience, guiding students from data preparation to the application of various regression techniques.

By the end of this course, students will be able to conduct rigorous empirical research. Beyond simply running regressions in statistical software, students are expected to understand the logic underlying each method and to develop a clear grasp of its assumptions, scope, and limitations. You will learn to interpret and critically evaluate published research, assess methodological choices, and engage with methodological papers at a technical level. For those interested in specializing in research methods, this course also provides the essential foundation for more advanced methodological training.

This course assumes the basic background knowledge of calculus, probability theory, linear algebra, and statistics that is covered in our Fall semester course POLSCI 609: Fundamentals of Research.

Requirements and Grading

Scribing (10%): This is a technical course, and reading the mathematical material on your own can be challenging. Students are therefore expected to take detailed notes during lectures and attend all lab sessions, which are required components of the course. As is traditional in the Statistics Department, each student is expected to serve as the "scribe" for one or two lectures, taking notes and producing LaTeX-formatted lecture notes that will be shared publicly with everyone.

Problem Sets (50%): The most effective way to encourage learning and deepen understanding of the material is through hands-on assignments. Consequently, there will be problem sets designed to reinforce key concepts and provide practical experience. Each problem set will typically include two types of questions: 1) Simple algebra: These questions only require the direct application of formulas discussed in class; 2) Programming: These tasks involve using R to write code and implement methods.

Exams (20% each): There will be two, in-person exams. Exams will focus on theoretical understanding, interpretation, and proficiency in R. One class session will be dedicated to the midterm exam, and the final exam will take place during the officially scheduled exam period.

AI and Collaboration Policy

AI and your peers are valuable resources for study and research. However, they are most effective when you have built a strong foundation of knowledge yourself. For example, it is well known that generative AI can produce incorrect or misleading information (hallucinations). As a user, you can only identify these mistakes if you have a solid understanding of the subject. Taking classes, studying, and practicing are essential steps to develop the foundational skills needed to effectively engage with AI in the future.

Therefore, please do not upload your homework and rely on AI to solve problems for you. Simply reading AI-generated answers will not help you learn. Real learning happens when you think deeply, and struggle through challenges.

You are encouraged to discuss with your peers, but the final work must be written by you. For any submitted work, please indicate: which parts, if any, were generated with AI assistance, and who you discussed the assignment with. Honest documentation of your learning process ensures that you develop real understanding and academic integrity.

Textbooks

We will strike a balance between two textbooks. The first is intended to be read during the winter break, and relevant chapters should be re-read before each lecture. The second textbook contains more rigorous and advanced material, which will deepen your understanding of the topics covered in class.

1. Basic: Wooldridge, Jeffrey M. Introductory econometrics a modern approach. South-Western cengage learning, 2016.

2. Advanced: Hansen, Bruce. Econometrics. Princeton University Press, 2022.

Resources

These textbooks are also excellent general references that may be helpful for this class.

- Wasserman, L. (2013). All of statistics: a concise course in statistical inference. Springer Science & Business Media.
- Hansen, Bruce. Probability and statistics for economists. Princeton University Press, 2022.
- Wooldridge, Jeffrey M. Econometric analysis of cross section and panel data. MIT press, 2010.
- Hsiao, Cheng. Analysis of panel data. No. 64. Cambridge university press, 2022.
- Baltagi, Badi Hani, and Badi H. Baltagi. Econometric analysis of panel data. Vol. 4. Chichester: John Wiley & Sons, 2008.

- Gelman, Andrew, and Jennifer Hill. Data analysis using regression and multilevel/hierarchical models. Cambridge university press, 2007.
- Imbens, Guido W. "Causal inference in the social sciences." Annual Review of Statistics and Its Application 11 (2024).
- Ding, Peng. A first course in causal inference. Chapman and Hall/CRC, 2024.

(Temporary) Topics

Least Squares Regression: Estimation

- Conditional Expectation and Projection
- Algebra of Least Squares
- Property of OLS Estimator
- Various Variance Estimation
- Asymptotic Theory of OLS
- Causal Interpretation

Least Squares Regression: Inference

- Hypothesis Testing
- Bootstrap

Instrumental Variables

- Traditional Wald and 2SLS estimators
- Property and Inference
- Various Specification Tests
- Weak and Many IVs
- Local Average Treatment Effect

Panel Data

- Random Effects Model
- Fixed Effects Model
- Dynamic Panel Models
- Difference in Difference
- Staggered Design

Other Topics

- Time series
- Nonparametric regression
- Regression Discontinuity