George Mason University
Schar School of Policy and Government

Draft Syllabus (subject to change)
Introduction to Data Analysis for PhD Students (POGO 511)

Fall 2026, Mondays 4:30 pm

Course Instructor: John S. Earle (earle@gmu.edu)
Stata Instructor: Sungbin Park (spark224@gmu.edu)
Office: Founders Hall 539
Office hours: Wednesdays, 2-4 p.m., or by appointment

This course introduces PhD students to statistical and econometric methods used in policy evaluation and research. Pre-requisites include basic mathematics (algebra and elementary calculus), probability, and statistics (high school or introductory college level). No economics background is assumed, and examples will be drawn not only from economics but also political science and other social sciences as well. Relevant statistical theory will be discussed, but the course will emphasize intuition, applications, and research practice rather than rigorous proofs.

Most class meetings will consist of lectures and discussions of statistical methods. Four meetings will be reserved for learning Stata, the leading statistical program used by economists, political scientists, and sociologists. These classes will be led by Sungbin Park.

The assignments for the course are designed not only to further understanding of topics in data analysis but also to develop students' abilities as policy researchers. There are eight sets of "homework" and four computer assignments, all through MindTap, the Cengage Publisher's program to accompany the textbook. These assignments can be worked on by groups of students together, but each student should submit separate answers, and it is in each student's interest that they fully understand the answers submitted. Some of the problems from the textbook are slightly confusing, but many of them permit multiple attempts. These assignments are due by 11:59 pm on the Sunday after the class for which they are listed below. Late answers may be submitted, but they are subject to a 30 percent penalty in order to disincentivize falling behind. In determining the course grade, I will drop the lowest two homework scores and lowest computer assignment scores. The purpose of all these provisions is to encourage students to use the homework as learning tools and to discourage requests for extensions, which will generally be denied.

In addition to an in-class midterm and cumulative final exam, the course requires each student to choose a research paper and to make a short presentation focused on the methods of the paper; the presentations will be on the last day of class, and more details about requirements for the presentation will be provided soon. The final course grade will be determined by the homework (5% each, 30% total), the computer assignments (5% each, 15% total), the midterm exam (15%), the presentation (10%), and the final exam (30%).

By the end of the course, students should be able to understand, interpret, and criticize simple statistical analyses in professional social science journals and to carry out similar analyses themselves. The most important tool for policy analysis and empirical research in the social science more generally is regression, and thus the goal of the course is for students to understand the uses and limitations of this tool and to be able to apply it in simple contexts. By the end of the course, students should also have sufficient preparation for more advanced courses in specialized fields of statistics and econometrics.

The main textbook for the class is Wooldridge, *Introductory Econometrics*, 8th Edition, 2025. This book is already a classic in social science, having revolutionized the teaching and practice of program evaluation and estimation of "causal effects" more generally, and it is also the very best book that a PhD student (and even many scholars) can own as a reference both when reading journal articles and when designing research projects. Links for purchasing the e-book (or hard copy if you prefer) and MindTap are on the course Canvas site.

Statistical software: Stata 14 or newer, IC or SE (SE is better for larger data sets, but IC is adequate for this class; see http://www.stata.com/products/which-stata-is-right-for-me/). You can purchase a license or get free access two ways: computer labs in Arlington or "Virtual Computing Lab" (https://its.gmu.edu/knowledge-base/what-is-patriot-virtual-computing-labs/).

Other useful resources:

Angrist, Joshua D., and Jörn-Steffen Pischke. 2017. "Undergraduate Econometrics Instruction: Through Our Classes, Darkly." *Journal of Economic Perspectives*, 31 (2): 125-44.

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives*, 31 (2): 3-32.

Bailey, Michael A. 2016. *Real Stats: Using Econometrics for Political Science and Public Policy.* Oxford University Press. eISBN-13: 9780190262198.

Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. Yale University Press.

Healey, Joseph, *Statistics: A Tool for Social Research*, 10th edition, Cengage.

Stata, *Getting Started with Stata*, Stata Press, Release 14. Free download: http://www.stata.com/bookstore/getting-started-windows/

UCLA statistics website: http://www.ats.ucla.edu/stat/stata/

Outline of Topics and Schedule by week (preliminary and subject to change)

0. (before semester): Course preparation

    Wooldridge, Mathematics Appendix A
    Test your mathematics with MindTap Exercise: Basic Mathematical Tools (ungraded)

1. (August 24): Course Overview and Review of Probability and Statistics

    Wooldridge, Appendices B and C
    Wooldridge, Chapter 1
    MindTap Homework 1 with Chapter 1 (graded, but no deadline)
    MindTap Exercises on Probability and Statistics (ungraded)

2. (August 31): Simple Linear Regression

    Wooldridge, Chapter 2
    MindTap Homework 2 due (graded)

3. (Sept 7): no class (Labor Day holiday)

4. (Sept 14): Introduction to Stata

    MindTap Video. Stata Tutorial: The Basics. How to Use Stata.

MindTap Video. Stata Tutorial: The Basics. Basic Estimation Commands.
MindTap Videos. Stata Tutorial: The Basics. Commands and Variables.
Stata Assignment 1 (on MindTap)

5. (Sept 21): Multiple Regression I (Estimation)

Wooldridge, Chapter 3
MindTap Homework 3 (graded)

6. (Sept 28): Descriptive statistics and basic regression with Stata

Stata Assignment 2

7. (Oct 5): Multiple Regression II (Inference)

Wooldridge, Chapter 4
Amrhein, Valentin, Sander Greenland, Blake McShane, et. al., "Retire statistical significance." *Nature*, 567, 305, 21 April 2019.
MindTap Homework 4 (graded)

8. (Oct 12): no class (Fall "Break")

8. (Oct 19): Midterm Quiz (in class)

9. (Oct 26): Multiple Regression III (Applications)

Wooldridge, Chapter 6
MindTap Homework 5 (graded)

10. (Nov 2): Multiple regression with Stata

Stata Assignment 3

11. (Nov 9): Qualitative Variables

Wooldridge, Chapter 7
MindTap Homework 6 (graded)

12. (Nov 16): Time Series and Panel Data

Wooldridge, Chapters 10 and 13
MindTap Homework 7 (graded)
MindTap Homework 8 (graded)

13. (Nov 23): no class (Thanksgiving holiday)

14. (Nov 30): Stata Applications

Stata Assignment 4

15. (Dec 7): Presentations

Final exam: Dec 14 (set by Registrar)

Academic Accommodation for a Disability
If you are a student with a disability and you need academic accommodations, please see me and contact the Disability Resource Center (DRC) at (703) 993-2474. All academic accommodations must be arranged through the DRC as early as possible in the term.

Academic standards policy: see https://catalog.gmu.edu/policies/academic-standards/.

AI policy: The use of generative AI tools (e.g., ChatGPT, Claude, Gemini) is not permitted for any coursework in this class unless explicitly authorized. Unauthorized use will be treated as a violation of academic standards.

GMU common course policies: see https://stearnscenter.gmu.edu/home/gmu-common-course-policies/.