

# PP 312: Advanced Statistics for Data Analysis 1, Fall 2025

**Instructor:** Dan Black

**Office:** Keller #3047

**Drop-in office hours:** Wednesday, 9:00 to 10:20 and Thursday, 1:30 to 2:20

**Email:** danblack@uchicago.edu

## Teaching Assistants

**Matt Stadnicki, Head TA**

**Andrew Doherty Munro**

**Ganon Evans**

mstadnicki@uchicago.edu

adohertymunro@uchicago.edu

ganon@uchicago.edu

**Class time:** Tuesday-Thursday: 9:30 to 11:20, Room 0023

Tuesday-Thursday: 11:30 to 12:20, Room 0023

**TA sessions:** Friday 10:30-11:50, Room 0010

Friday 1:30-2:50, Room 0010

## Course Objectives

- To introduce students to statistics that are useful in the analysis of public policy data.
- To provide students with basic training in the necessary computation skills to analyze data.
- To provide students with basic training in the necessary statistical skills to analyze data.
- To introduce students to the evaluation of their epistemology

**Professional Behavior:** The Harris School expects faculty, staff, and students to conduct themselves in a professional manner. Students engaged in unprofessional behavior will be reported to Academic and Student Affairs for disciplinary action. Please report any inappropriate behavior to your instructors.

## **Classroom Behavior:**

- Please be on time for class. While I understand that despite their best efforts, people occasionally run late, tardiness disrupts your colleagues and is unprofessional.
- Because there is substantial evidence that using electronics in the classroom negatively affects learning outcomes, students' use of electronic devices during class is prohibited: **no computers, tablets, or phones**. Bring a pen (or pencil) to class to take notes. Lecture slides are available on the Canvas site.
- Please bring your name tags to every class. It helps your instructors and your classmates learn who you are.

**Use of AI:** The use of AI programs (e.g., ChatGPT) is strongly encouraged. You should, however, save the output from the program and cite the fact that you are using it. **You are responsible for any errors and other issues (e.g, plagiarism) the program creates.** We are interested in your ability to utilize all modern tools to solve problems effectively.

## Harris Integrity Policy for Problem Sets Involving Programming Code

Academic dishonesty will not be tolerated. If you commit plagiarism, you may receive an F and be referred to the Area Disciplinary Committee. All work must be your own. **Do not:**

- Show other students your code
- Ask for another student's code
- Use online solutions to textbook questions
- Copy large portions of code from online repositories (e.g., replication code)

Every submission begins with "This submission is my work alone and complies with the 31202-integrity policy. Add your initials to indicate your agreement: \*\* \_\_ \*\*"

How should you collaborate? You can clarify ambiguities in problem set questions, discuss conceptual aspects of problem sets, show output on screen (e.g., a graph or table), and show helpful documentation files.

**Title IX Reporting Responsibilities:** Your instructor and TAs for this class are designated as "responsible employees" under the US law known as Title IX. We have to report incidents of sexual harassment, including sexual violence, domestic violence, dating violence, stalking, or other misconduct, to appropriate school officials.

## Text

Jeffrey W. Wooldridge, *Introductory Econometrics: A Modern Approach*. **Any edition will do.** This book is used next term in *Advanced Statistics for Data Analysis*, 2.

**Optional Text:** If you want to make statisticians rich by buying standard textbooks, previous students have found these handy:

- Wackerly, Mendenhall, and Scheafer, *Mathematical Statistics with Applications*. Any edition will do.
- John Rice, *Mathematical Statistics and Data Analysis* is used in Stat 244 and is fine (but expensive).
- M.G. Bulmer, *Principles of Statistics*, New York: Dover Classics, 1979. Less than \$10 for the Kindle edition and less than \$15 for the print version. It is more mathematical and detailed than Wooldridge.

These books are interesting takes on Statistics.

- Charles Wheelan's *Naked Statistics: Stripping the Dread from Data*. Try this one if you want something with virtually no math, but good intuitive explanations. Charlie used to teach at the Harris School.
- Carl T. Bergstrom and Jevin D. West. *Calling Bullshit* New York: Random House, 2020. An excellent approach to understanding when you are being misled and how to combat misinformation.

- Tim Harford. *The Data Detective: Ten Easy Rules to Make Sense of Statistics* New York: Riverside Books, 2021.
- David Salsburg. *The Lady Tasting Tea*. This book provides an overview of many of the ideas of statistics.
- Nate Silver. *Signal and the Noise: Why So Many Predictions Fail – but Some Don't* New York: Penguin Press, 2012.

There are many interesting online guides to the software used in this class. They include:

**R resources:**

<http://r4ds.had.co.nz/>  
<https://www.statmethods.net/>

**Stata:**

<http://web.mit.edu/14.31/www/stata.html>  
<http://www.stata.com/links/video-tutorials/>

**Grades**

We will assign grades for this course based on homework assignments given throughout the term, attendance, and three quizzes:

**Fraction**

5 assignments:	0.50
3 quizzes:	0.45
Attendance:	0.05

**Assignments:** Homework assignments will be done individually. To have some people to talk with about the assignments, we will assign homework groups where you are free to discuss issues.

If you believe your assignment grade is incorrect or unfair, please submit your concerns in writing to the head TA within a week of its return. Explain fully in writing why you believe what the problems are. The TA responsible for the relevant question will respond in writing. If you still have concerns, you may submit them in writing to me.

Core courses at the Harris School are graded on a curve; however, we have been granted permission to use a more generous curve for Advanced Stats. For an advanced course, we will use:

**Fraction**

**Grade**

A	0.25
A-	0.25
B+	0.25
B and below	0.25

# Tentative Schedule

**Date:**

**Lecture:**

## **September**

30<sup>th</sup> Lecture 1: Review of distributions

## **October**

2<sup>nd</sup> Lecture 1: Distributions

7<sup>th</sup> Lecture 1: Distributions

9<sup>th</sup> Lecture 1: Distributions

14<sup>th</sup> Lecture 2: Review of hypothesis testing

16<sup>th</sup> Lecture 2: Hypothesis testing

21<sup>st</sup> Lecture 2: Hypothesis testing

23<sup>rd</sup> Lecture 3: Estimation and the selection problem

28<sup>th</sup> Lecture 3: Estimation

30<sup>th</sup> Lecture 3: Estimation

## **November**

4<sup>th</sup> Lecture 3: Estimation

6<sup>th</sup> Lecture 4: The science of data

11<sup>th</sup> Lecture 4: Data

13<sup>th</sup> Lecture 4: Data

14<sup>th</sup> Lecture 4: Data

18<sup>th</sup> Lecture 5: Randomized Control Trials (RCTs)

20<sup>th</sup> Lecture 5: RCTs

25<sup>th</sup> **Thanksgiving break**

27<sup>th</sup> **Thanksgiving break**

## **December**

2<sup>nd</sup> Lecture 5: RCTs

4<sup>th</sup> Lecture 5: RCTs

# Reading assignments

## **Lecture 1: Review of distribution functions and their empirical counterparts**

### **Readings:**

Wooldridge, Appendix B

## **Lecture 2: Review of hypothesis testing**

### **Readings:**

Wooldridge, Appendix C.5 and C.6

## **Lecture 3: The selection problem, describing data, and estimation**

### **Readings:**

P. Krugman, The Accidental Theorist

[http://www.slate.com/articles/business/the\\_dismal\\_science/1997/01/the\\_accidental\\_theorist.html](http://www.slate.com/articles/business/the_dismal_science/1997/01/the_accidental_theorist.html)

D. Roy. "Some Thoughts on the Distribution of Earnings" Oxford Economic Papers 3(2) June 1951 135-46.

M. Friedman, "The Methodology of Positive Economics" (1953)

D. Black. "Matching as a Regression Estimator - How to Avoid Making Assumptions about the Functional Form of the Regression Equation" *IZA World of Labor* September 2015 1-10.

Wooldridge, Chapters 2, 3, and Appendix C.1 to C.4

## **Lecture 4: The science of data**

### **Readings:**

D. Black, Y. Hsu, S. Sanders, L. Schofield, and L. Taylor. "Black-White Mortality Differentials at Older Ages: Inferential Difficulties and New Evidence" *Demography*, 54(6) December 2017 2001-24.

D. Black, S. Sanders, and L. Taylor. "Measurement of Higher Education in the Census and CPS" *Journal of the American Statistical Association* September 2003 98(463) 545-54.

## **Lecture 5: Randomized control trials**

### **Readings:**

J. Heckman and J. Smith. "Assessing the Case for Social Experiments" *Journal of Economic Perspectives* 9(2) Spring 1995 85-110.

Fernando P Polack *et al.* "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine" *New England Journal of Medicine* 383(27) December 31<sup>st</sup>, 2020 2603-15.

John P.H. Wilding *et al.* "Once-Weekly Semaglutide in Adults with Overweight or Obesity" *New England Journal of Medicine* 384(11) March 18<sup>th</sup> 2021, 989-1002