

PP 312: Advanced Statistics for Data Analysis 1, Fall 2024

Instructor: Dan Black

Office: Keller #3047

Drop-in office hours: Tuesday and Thursday, 11:00 to 12:00

Email: danblack@uchicago.edu

Teaching Assistants

Brian Curran, Head TA

Claire Conzelmann

Nidhaan Jain

Gabby Rocco

bcurran@uchicago.edu

cconzelmann@uchicago.edu

nidhaan.jain@chicagobooth.edu

grocco@uchicago.edu

Class time: Tuesday-Thursday: 8:00 to 9:20, Room 0001

Tuesday-Thursday: 9:30 to 10:50, Room 0001

TA sessions: Friday 1:30-2:50, Room 0001

Friday 3:00-4:20, Room 0001

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.

Professional behavior: The Harris School expects faculty, staff, and students to behave professionally. Students engaged in unprofessional behavior will be reported to Academic and Student Affairs for disciplinary action. Please report any inappropriate behavior to your instructors.

In terms of your 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 is not permitted during class. Bring paper and pen (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 use all the modern tools to solve problems correctly.

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. 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 very 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 on the basis of homework assignments given through the term and a midterm.

	<u>Scheduled</u>	<u>Fraction of grade</u>
Assignments 1-5	Various	40%
Assignment 6	December 6 th	20%
Exam	November 15 th (in TA session)	40%
Total		100%

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, but we have been given permission to grade Advanced Stats using a more generous curve. For an advanced course, we will use:

Grade	Fraction
A	0.25
A-	0.25
B+ and below	0.25
B and below	0.25

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.

Tentative Schedule

Date:

Lecture:

October

1 st	Lecture 1: Distributions
3 rd	Lecture 1: Distributions
8 th	Lecture 1: Distributions
10 th	Lecture 1: Distributions & Lecture 2: The science of data
11 th	<i>Homework 1 due</i>
15 th	Lecture 2: The science of data
17 th	Lecture 2: The science of data
18 th	<i>Homework 2 due</i>
22 nd	Lecture 2: The science of data
24 th	Lecture 3: The selection problem
25 th	<i>Homework 3 due</i>
29 th	Lecture 3: The selection problem
31 st	Lecture 4: Estimation and OLS

Tentative Schedule

November

5 th	Lecture 4: Estimation and OLS
7 th	Lecture 4: Estimation and OLS
8 th	Homework 4 due
12 th	Lecture 5: Experiments
14 th	Lecture 6: Matching
15 th	Exam in discussion
19 th	Lecture 6: Matching
21 st	Lecture 6: Matching
22 nd	Homework 5 due
26 th	Thanksgiving break
28 th	Thanksgiving break

December

3 rd	Lecture 7: Distribution of estimates
5 th	Lecture 7: Distribution of estimates
6 th	Homework 6 due

Reading assignments

Lecture 1: Distributions functions

Readings:

Wooldridge, Appendix B

Lecture 2: 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 3: The selection problem

Readings:

P. Krugman, The Accidental Theorist

http://www.slate.com/articles/business/the_dismal_science/1997/01/the_accidental_theorist.html

A. 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)

Lecture 4: Ordinary Least Squares

Readings:

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

Lecture 5: Experiments

Readings:

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

Lecture 6: Matching

Readings:

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.

D. Black, A. Haviland, S. Sanders, and L. Taylor. "Gender Wage Disparities among the Highly Educated" *Journal of Human Resources Summer* 2008 42(3) 630-59.

D. Black and J. Smith. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching" *Journal of Econometrics* August 2004 121(1-2) 99-124

Lecture 7: Distributions of estimates

Readings:

Wooldridge, Appendix C.5 and C.6