PP 312: Advanced Statistics for Data Analysis 1, Fall 2023

Instructor: Dan Black
Office: Keller #3047
Drop-in office hours: Tuesday, 9:30 to 10:50, Keller 3047
Email: danblack@uchicago.edu
Also, by appointment: please contact my administrator Lydia Veliko at lydiav@uchicago.edu to schedule an appointment.

Teaching Assistants

Danielle Nemschoff, Head TA	dinemschoff@uchicago.edu
Azim Shivji	ashivji@uchicago.edu
Marina Smalling	msmallin@uchicago.edu
Anita Zaldivar	azaldivar@uchicago.edu

Class time: Tuesday-Thursday: 8:00 to 9:20, Room 1022
TA session: Thursday, 2:00 – 3:20, Room 1022

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

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.

- Please avoid using the internet in class, whether on your laptop or your phone. This is detrimental to one's learning and very distracting to other students.

- Please bring your name tags to every class. It helps your instructors and your classmates learn who you are.

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.

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.

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) that the program creates.** We are interested in your ability to use all the modern tools to arrive at the *correct* solutions to problems.

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, Mendenhal, 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, less than $15 for the print. More mathematical than Wooldridge and more detailed.

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.

	Scheduled	Fraction of grade
Homeworks	Various	60%
Midterm	November 11	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 who is 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	1/3
A-	1/3
B+ and below	1/3

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

Tentative Schedule

Date:	Lecture:
September	
26th	L1: Models and causality
28th	L1: Models and causality and L2: Random variables and distributions
October	
3rd	L2: Random variables and distributions
5th	L2: Random variables and distributions
10th	L2: Random variables and distributions
12th	L2: Random variables and distributions and L3: Estimation, part 1
17th	L3: Estimation, part 1
19th	L3: Estimation, part 1
24th	L3: Estimation, part 1 and L4: Application of estimation: Experiments
26th	L4: Application of estimation: Experiments and L5: Estimation, part 2
31st	L5: Estimation, part 2
November	
2nd	L5: Estimation, part 2
7th	L5: Estimation, part 2 and L6: Distributions of estimates
9th	L6: Distributions of estimates
9th	**Midterm exam in discussion sections**
14th	L6: Distributions of estimates and L7: The science of data
16th	L7: The science of data
21st	**Thanksgiving break**
23rd	**Thanksgiving break**
28th	L7: The science of data and L8: The bootstrap
30th	L8: The bootstrap
December	
2nd	**Final Homework Project Due**

Reading assignments

Lecture 1: Models and causality
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 2: Random variables and distributions
Readings:
Wooldridge, Appendix B

Lecture 3: Estimation, part 1
Readings:
Wooldridge, Chapters 2, 3, and Appendix C.1 to C.4

Lecture 4: Application of estimation: Experiments
Readings:
James Heckman and Jeffrey Smith. "Assessing the Case for Social Experiments" *Journal of Economic Perspectives* 9(2) Spring 1995 85-110.

Lecture 5: Estimation, part 2
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 6: Distributions of estimates
Readings:
Wooldridge, Appendix C.5 and C.6

Lecture 7: 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 8: The bootstrap

Readings:

David Brownstone and Robert Valletta, 2001. "The Bootstrap and Multiple Imputation: Harnessing Increased Computing Power for Improved Statistical Inference" *Journal of Economic Perspectives* 15(4): 129-41