PP 312: Advanced Statistics for Data Analysis 1, Fall 2022

Instructor: Dan Black
Office: Keller #3047
Drop-in office hours:
Tuesday-Thursday, 10:00 to 10:50

Also, by appointment: please contact my administrator Lydia Veliko at lydiav@uchicago.edu to schedule an appointment.

Email: danblack@uchicago.edu

Teaching Assistants

Angela Wyse, Head TA	awyse@uchicago.edu
Jieyi Chen	jieyi@uchicago.edu
Carolyn Liu	crliu@uchicago.edu
Sam Zuckert	szuckert@uchicago.edu

Class Time: Tuesday-Thursday: 2:00 to 3:20, Room 1022

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.

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

The specific skills are as follows:

1. Learn the basic properties of the basic functions of probabilities: cumulative distribution functions (cdf's), probability mass functions (pmf's), and probability density functions (pdf's).
2. Learn the basic descriptive statistics: means, standard deviation (variance), skewness, covariances, correlation coefficients, and quantiles. Learn when these parameters are informative.
3. Learn that "parameters" are fixed values and that "estimates of parameters" are random variables.
4. Learn how to calculate estimates of these basic statistics in both Stata and R.
5. Learn the basics of hypotheses testing. Learn how to construct both null and alternative hypotheses.
6. Learn to draw the distinctions between exact and asymptotic tests. Learn when asymptotic tests will perform well and when they will not.
7. Learn how to use simulations to help understand complex statistical problems. Learn how to program estimates using simulations.
8. Learn how and when to use the bootstrap to improve on asymptotic tests. Learn how to program bootstrap estimations.
9. Understand the basics of sample design including both stratification and clustering.
10. Understand the problems associated with both unit and item nonresponse and the assumptions behind the "correction" of these problems.
11. Understand the consequences of measurement error.
12. Understand how experiments allow you to draw causal inference.

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

- Dekking, F.M., C. Kraaikamp, H.P. Lopuhaä, and L.E. Meester, *A Modern Introduction to Probability and Statistics: Understanding Why and How*.
- 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.
- Charles Wheelan's *Naked Statistics: Stripping the Dread from Data*. If you want something with virtually no math, but good intuitive explanations, try this one. Charlie used to teach at the Harris School.

Other Resources

- 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.

- David Spiegelhalter. *The Art of Statistics: How to Learn from Data* New York: Basic Books, 2019.

There are hundreds of statistics books. They seldom make the bestseller lists, but they are often excellent.

There are lots of 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://data.princeton.edu/stata/
http://tutorials.iq.harvard.edu/Stata/StataIntro/StataIntro.html
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 xx	40%
Total		100%

If you believe that your grade on an assignment is incorrect or unfair, please submit your concerns in writing to the head TA within a week of its being returned. 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 advance 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 homeworks, I will assign homework groups where you are free to discuss issues. Groups will change over the quarter.

Professional behavior: The Harris School expects faculty, staff, and students to behave always 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.

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 a duty to report incidents of sexual harassment, including sexual violence, domestic violence, dating violence, and stalking, or other misconduct to appropriate school officials**.**

Tentative Schedule

Date:	Lecture:
September	
	L1 – Models
27th	
29th	L1 – Models
October	
4th	L2 – Random variables
6th	L2 – Random variables
11th	L3 – Multivariate distributions
13th	L3 – Multivariate distributions
18th	L4 – Summary statistics
20th	L4 – Summary statistics
25th	L5 – Central limit theorem & simulations
27th	L5 – Central limit theorem & simulations
November	
	L6 – Hypothesis Testing
1st	
3rd	L6 – Hypothesis Testing
8th	L7 – The bootstrap
10th	L7 – The bootstrap
15th	L8 – The science of data
17th	L8 – The science of data
22nd	**Thanksgiving break**
24th	**Thanksgiving break**
29th	L9 – Experiments
December	
	L9 – Experiments
1st	
xx	**Final Homework Project Due**

Reading assignments

Lecture 1:

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)

Video:

Newton's model

Lecture 2:

Readings:

Wooldridge, Appendix B

Videos:

Bored Work: Discrete distributions
Binomial distribution
Making your own distributions

Lecture 3:

Readings:

Wooldridge, Appendix B

Lecture 4:

Readings:

Wooldridge, Appendix B

Video:

Regression

Lecture 5:

Readings:
Wooldridge, Appendix C

Lecture 6:

Readings:

Wooldridge, Appendix C

Lecture 7:

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

Lecture 8:

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 9:

Readings:

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

Video:

Internal and external validity