SOSC 13200: Social Science Inquiry II

DETAILS
Remote for the Winter 2021 quarter
TuTh 1PM – 2:20PM (section 3) / 2:40 PM – 4 PM (section 4)
Course website on Canvas
wschultz@uchicago.edu

INSTRUCTOR
Dr. William Schultz
Virtual "Town hall" office hours

- WEDNESDAY from 1PM-2PM
- WEDNESDAY from 9PM-10PM

One-on-one Zoom meetings by appointment

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less." — Marie Curie

COURSE DESCRIPTION
It is easy to come up with stories that *sound* like they explain social phenomena around the world, or what the impacts of a public policy choice will be. But how do we determine whether those stories are accurate, or useful? The answer: science.

As the course catalog explains, this sequence provides students with an introduction to the philosophy of social science inquiry, a sense of how that inquiry is conducted, and an understanding of how policy implications can be drawn responsibly from evidence provided by empirical social science. The sequence's objective is to expose students to both classic and contemporary perspectives on the promise of social science, as well as a sense of its uses and abuses.

In the Fall quarter, we covered a philosophy of science that is common in social research and discussed different issues you need to think about when planning a social science research project. Now that you've thought through a plan for your study, it's time to put that plan into action and conduct some statistical tests! This quarter will introduce you to the basics of statistical techniques commonly used in the social sciences. You will also learn statistical programming using R, so that you can apply these techniques yourself to real world data.

COURSE STRUCTURE
I'm providing the course material in both synchronous and asynchronous formats to make sure that it is accessible to students who are not able to easily participate synchronously during the pandemic. I only request that students reach out to me if you need to take in the class material asynchronously.

I'll post presentation slides for each class day on Canvas. I'll also host virtual class sessions in which I'll present those slides and discuss the material with students as in a normal in-person course. I'll record those virtual sessions and post them to Canvas. I'm not grading attendance in the virtual class sessions (as you can see below), but students should keep in mind that there will sometimes be important material in the recorded classes that are not in the slides. Some class sessions this quarter will be more applied (where I walk everyone through some coding problem), while other sessions will be more lecture oriented.

I'll happily answer questions over email, but I encourage students to post questions about the course material on the Canvas discussion board. Often, if something is confusing you, it is confusing to others as well. It's helpful for me to know what material students are struggling with.

My scheduled office hours for this course will be virtual "town hall" Zoom meetings for students to drop in either individually or as a group and ask questions about the material. You can come and go from these meetings whenever you like. I've scheduled time for these later in the evening that may be more reasonable for students who are not in the Central US time zone. All students are, of course, welcome to attend the later timeslot. Private meetings are available by appointment for anyone who wants them; we can always find a time that works for you.

COURSE MATERIALS

Most readings will be found on Canvas. Others can be found by following instructions in the Course Schedule below.

I've based the structure of this class off of a great textbook called *Quantitative Social Science: An Introduction*, by Kosuke Imai. You can find more information about the book here: http://qss.princeton.press/. The website has some supplemental materials you may find useful, like example code for different problems we will work on and a package that runs tutorials in R.

Instead of assigning readings from this book, though, I'll be asking you to read research articles that different chapters of this textbook drawn on as examples. That way, you'll come into class with a solid understanding of the theory we're interested in each day and the dataset we'll use to explore it. The only exception is that we will read some selections from Chapters 6 and 7. However, I'll provide those readings on Canvas, so you don't have to buy anything. That said, the book is a great resource to consider picking up if you want to dive more into this material after the class is over.

It's important to your own education that you do the assigned readings for each day. It's not a good use of your time to take a class without putting in the effort to learn something from it. I do my best to limit the amount of reading I assign without threatening your education by leaving something important out. That's my end of the bargain. Your end of the bargain is keeping up with the readings I do assign.

GRADING PHILOSOPHY AND EXPECTATIONS

Grades in college have several purposes: assessment; signaling a student's overall ability; and signaling a student's effort to master the course material. In my personal view, the last purpose is most important. That is my priority when designing and grading assignments.

EARNING YOUR GRADE

Your overall grade will be out of 300 points.

Four problem sets: 192 points = 64%

- Each problem set is worth 48 points
- These are due throughout the quarter, indicated in the schedule below. Problem sets will ask you to complete coding exercises in R that demonstrate your understanding of the course material.
- Completing a problem set requires submitting two things: your R code, and a write-up interpreting or explaining your results. I will give partial credit relative to how much of your coding was done correctly. Your answers in the writeup *must* match the results produced by your code.
- If you aren't happy with your grade on a problem set, you can resubmit it *once* with corrections and earn back up to half of the points you lost the first time. This resubmission is due a week after your writeup is returned to you with a grade. *This is not an option for the final problem set.*
- Each problem set is associated with specific class days. The problem set will be available on Canvas after the last associated class session is over (e.g., the first problem set will be available after the end of our third class session).

Lab exercises (participation): 48 points = 16%

- You will occasionally have to complete lab exercises in the R studio cloud (I'll explain in our first session, and in an announcement) to show that you're following along as I'm teaching R coding. These will be graded leniently, and should be relatively quick to complete. Think of them as participation exercises. There will be roughly seven of these (precise number TBD). I will announce lab exercises on Canvas. They'll be due at 11:59PM CT the day after our class (so either Wednesday or Friday).

Open notes final (on Canvas): 60 points = 20%

- This final exam will test your understanding of the statistical topics covered over the course of the quarter. It will focus on statistical topics applied in the problem sets. It will involve a mix of multiple choice and open-ended questions.
- You will have two days to take the final. It is due at the end of our final exam time scheduled by the University. It is designed to be completed within our allocated final exam time (two hours), but I'm making it available two days early to provide you some flexibility in your schedule. *Once you start it, you must finish within two hours.*

MISSED/LATE ASSIGNMENTS

I accept missed assignments without penalty for documented medical reasons, family crises, call to active military duty or jury duty, religious holy days, and official University activities. I also give deference to parents with dependent children who are sick.

Otherwise, late assignments drop 10% of their full worth for each day they are late, but you can still submit them to me through email.

PETITIONS FOR A GRADE CHANGE

I consider written petitions for a change of grade on assignments. However, I require students to wait two business days after receiving a grade before submitting their petition.

EMAIL POLICY

Please include your first and last name and course information in the subject line of your email. When I receive your email, I will make effort to respond in a timely manner, usually within 48 hours. You may receive a reply sooner than that, but you should not expect an immediate response. Please treat all email correspondences with your instructor as you would treat any other professional exchange. I expect emails to be respectful and polite, to use correct grammar and complete sentences.

Schedule

Getting started

(Class 1) Tuesday, Jan 12th — What is R; levels of measurement
No readings (start on readings for Thursday)

(Class 2) Thursday, Jan 14th — Our first R lab
Read for today:

- Wickham and Grolemund. *R for Data Science.* Available free here: https://r4ds.had.co.nz/index.html
    - Chapter 1, 4, 6, 8.

Analyzing single variables and bivariate relationships

(Class 3) Tuesday, Jan 19th — The potential outcomes framework ("Rubin's causal model")
Read for today:

- Holland (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*.
    - From the beginning through section 4 (pages 945 – 949)
    - It's only a few pages, but it's dense. Make sure you take time to think through and process what you're reading.
- Bertrand and Mullainathan (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*
    - Pages 991 – 997
    - Read enough to make sure you understand the theory they want to test and the dataset they produce to test it

(Class 4) Thursday, Jan 21st — Estimating causal relationships
Read for today:

- Card and Krueger (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*
    - Pages 772 - 778
- Gerber, Greene and Larimer (2008). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *The American Political Science Review*
    - First two sections (Introduction and "Social Norms, the Calculus Of Voting, and Prior Research"), and 36-38

(Class 5) Tuesday, Jan 26th — Measuring and visualizing single variables
Read for today:

- Lyall, Blair, and Imai (2013). Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan. *American Political Science Review*.
    - Beginning to page 688

(Class 6) Thursday, Jan 28th — Measuring and visualizing the relationships between variables
Read for today:

- McCarty, Poole, and Rosenthal (2006). Polarized America: The Dance of Ideology and Unequal Riches. *MIT Press.* Chapter 2 (working paper)
    - Beginning through page 13

PROBLEM SET 1 (CLASS 1-3): DUE JAN 30TH

Linear regression: a Swiss army knife for applied research?

(Class 7) Tuesday, Feb 2nd — The intuition for linear regression
Read for today:

- Kellstedt and Whitten. *The Fundamentals of Political Science Research*. Chapter 8.
    - Pages 172-178

(Class 8) Thursday, Feb 4th — Bivariate regression
Read for today:

- Todorv et al. (2005). Inferences of competence from faces predict election outcomes. *Science*.
    - Entire paper

PROBLEM SET 2 (CLASS 4-6): DUE FEB 13TH

(Class 9) Tuesday, Feb 9th — Model fit and multiple regression
Read for today:

- Chattopadhyay and Duflo (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica.*
    - Beginning through 1414; 1423 - 1427

How confident are we in our findings?

(Class 10) Thursday, Feb 11th — Some important probability topics
Read for today:

- Imai (2017). Quantitative Social Science. *Princeton University Press*. Chapter 6.
    - Pages 242-247
    - Skim 248 and 249, what is the “birthday problem?”
    - Pages 250-252
    - Pages 278-286

(Class 11) Tuesday, Feb 16th — The normal distribution and the central limit theorem

Read for today:

- Imai (2017). Quantitative Social Science. *Princeton University Press*. Chapter 6.
    - Pages 286-296
    - Pages 300-306

(Class 12) Thursday, Feb 18th — Uncertainty estimation (part 1)
Read for today:

- Imai (2017). Quantitative Social Science. *Princeton University Press*. Chapter 7.
    - Pages 314-326

PROBLEM SET 3 (CLASS 7-9): DUE FEB 27TH

(Class 13) Tuesday, Feb 23rd — Uncertainty estimation (part 2)
Read for today:

- Imai (2017). Quantitative Social Science. *Princeton University Press*. Chapter 7.
    - Pages 326-342

(Class 14) Thursday, Feb 25th — Hypothesis testing (part 1)
Read for today:

- Fisher (1935). *The Design of Experiments*. Chapter 2.
    - Pages 11-26
    - What is the null hypothesis, and how do we use it in scientific research?
- Bohannon (2015). I Fooled Millions into Thinking Chocolate Helps Weight Loss. Here's How. *Gizmondo*.
    - Not on Canvas; you can find this news story easily by searching online.
    - What is multiple testing, and what does it have to do with “p values”?

(Class 15) Tuesday, March 2nd — Hypothesis testing (part 2)
Read for today:

- Amrhein et al. (2019). Retire Statistical Significance. *Nature.*
    - Entire paper
- Gerber and Malhotra (2008). Do Statistical Reporting Standards Effect What is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science.*
    - Pages TBD

For a deeper understanding (not required):

- McShane et al. (2019). Abandon Statistical Significance. *The American Statistician.*

(Class 16) Thursday, March 4th — Interpreting linear regression results (part 1)
Readings TBD

(Class 17) Tuesday, March 9th — Interpreting linear regression results (part 2)
Readings TBD

(Class 18) Tuesday, March 11th — Review

PROBLEM SET 4 (CLASS 10-15): DUE MARCH 12TH