

SOCI 385 Social Statistics

Introduction

September 12, 2023

- Today's plan: overview of the course outline and expectations, introductions, discussion of Ralston article
- I arrived somewhat late as a quantitative sociologist. Arrived somewhat late as a sociologist at all.
 - Avoided any math or stats classes as an undergrad.
 - By grad school, a few things had changed:
 - * Numbers were everywhere. Sociology is way more quantitative than I knew, and way more quantitative than our department here.
 - * Interests in inequalities in education were largely studied quantitatively
 - * Realized I actually liked thinking quantitatively. Having to write a research question and a hypothesis in a line of code helped me clarify my thoughts. Having to think about which variables to include - and, perhaps more importantly, which to exclude - was hugely influential.
 - As a result, I approach quantitative analysis as a way of thinking and as a way of telling stories.
 - That said, I am comfortable saying that stats have limitations. I don't believe quantitative sociology is always the best way to do sociology. But I do believe that many questions can benefit from approaching them quantitatively. And I know that engaging sociology and actually doing the work of sociology requires some familiarity with quantitative methods.
- About the course
 - Purpose: use statistics as a way to introduce principles of social science research
 - * Different from data science. Our focus is on explanation rather than prediction, and we are question-driven rather than data-driven.
 - Become exposed to some common survey sources and methods in sociology, namely the General Social Survey
 - Develop tools to be able to complete a quantitative senior project, read and understand quantitative research, be prepared for quantitative methods courses you might take in graduate school, and learn how to navigate some of the software that is in high demand on the job market
 - Uses math to explain concepts, but not a math course. No calculus required or expected. No previous statistics course required or expected
- About you
 - Name, major, any quantitative courses, senior work, particular interests to connect to datasets

Course Overview

- Broken into four topics
 - This week, introductions and review of research methods (variables, hypotheses,

- samples, etc)
 - Next three weeks, descriptive statistics. Center, spread, and shape. One variable, two variables, three variables. With three variables, will start to bring in causality which is where most quantitative research in social science is focused.
 - By that point you will have seen a lot of data. Hoping we can take a break for a day and talk about finding data for your own projects.
 - Then a pause for catch up or for probability. Probability is kind of random at this point in the semester, but we need a few basics to be able to transition to the next modules.
 - Next four weeks, inference. How to use the statistics we estimate about a sample to say something about the parameters of a population. Hypothesis testing, confidence intervals. Allows to ask if differences are “statistically significant”
 - Final three weeks, regression. How is a change in an explanatory variable associated with a change in an outcome? Only linear in this course, but hope to show you the very basics of non-linear models (at least logistic models)
- That calendar structures the schedule of assignments
 - Anchored by two big problem sets: one after descriptives, one after inference
 - In between, ungraded weekly assignments. Feels like a lot of work, but the course would be way harder without them!
 - Problem sets and assignments will be handed out on Thursday and due through Canvas by 10:00 AM the following Tuesday.
 - * An assignment could ask you to clean up a data set, run some basic commands, make a graph, etc. Assignments are short practices connecting readings to class discussions to the software to give you practice.
 - * A problem set would ask you to do the same, but to also think sociologically and practice writing quantitatively by offering and incorporating substantive explanations for your findings.
- Problem sets and assignments lead up to a final project. You’ll write a full report using pieces from all the modules, with an emphasis on regression.
- Totally encouraged to collaborate on assignments but submit your own. But I want problem sets done on your own.

Textbook

- We are using the free online “Introduction to Modern Statistics” textbook
 - No book is perfect. Better to give you one that’s free and supplement with other pieces as necessary.
 - Lots of additional practice online.

Software

- We are using R, through R Studio and with Quarto (not R Markdown). I think Quarto is what someone last year called “the new R.” They are all free.
- R is the actual program. R Studio is an easier interface. Quarto is a way to write reproducible and transparent reports that integrate your code and your results in your

writing.

- I want everyone to have R on your own computer if possible. We'll install the programs today.
- R has a steep learning curve. It is writing code, not point and click. It is also very valuable if you go to grad school or if you want any kind of job requiring analysis or research skills.
- We'll start with R on Thursday

Next class

- Read Jung et al paper on hurricanes. Stop before the “materials and methods” section.
- Think about the research design. What is the research question, the theories being tested, the hypotheses, the variables. We'll start there on Thursday.

Ralston (2020)

- New piece getting lots of attention
- Should note that UK sociology is much less quantitative than US sociology. Most of Ralston's comments about feelings toward statistics among students mirroring feelings toward statistics in the discipline overall wouldn't hold here. But it does capture some themes that seem fitting at Middlebury.
- How does statistical anxiety resonate with you? Would you agree or disagree with the statement "the idea of learning statistics makes me feel anxious"?
 - Certainly resonated with me as an undergrad.
- What is epistemological legitimacy: "idea that students do not engage with statistics because they are not seen as either appropriate or legitimate methods for researching social life." Also "learners do not engage with the legitimacy of mathematically based methods for understanding the social world."
 - "Using statistics detaches you from your research topic". That's the one I'm worried about most; we can deal with the anxiety but how do we deal with that?
- Surprised by art or science scale: 1 = closer to an art, 10 = closer to a science, mean of 2.8
 - And 56% agreed that "my degree will help me get a good job"
- Conclusion is that there needs to be "a pedagogical approach that situates strategies for teaching methods within strategies designed to demonstrate the legitimacy of statistical approaches." How do we do that?