SOC 5811: Social Statistics, or Learning from Social Data

Lecture time: Mon/Weds 1:00-2:15pm
Lecture location: Blegen Hall 210

Lab time: Tues 4:00-5:30pm
Lab location: Hubert H. Humphrey School 85

Instructor: Nick Graetz (he/him/his)
Email: ngraetz@umn.edu
Office hours: Weds 10:00-12:00pm, and by appointment
Office location: 1172 Social Sciences Building

Lab Instructor: Hyunjae Kwon (he/him/his)
Email: kwon0181@umn.edu
Office hours: Weds 12:00-1:00pm, and by appointment
Office location: 980 Social Sciences Building

Course Description

In this course, students will learn core statistical and computational principles that will allow them to perform quantitative analyses using social data. The course is designed for social science students at the beginning of their graduate school careers. However, advanced undergraduates can also take the course.

We will review basic probability, and then move on to univariate inference, the linear regression model, an introduction to causal inference, and panel data. In doing so, students will explore statistical concepts and methods that provide the foundation sociologists use to most commonly collect and analyze numerical evidence.

Sociology 5811 will also provide the foundation for data management and statistical inference using R, a statistical computing environment that is popular in the social sciences.

Prerequisites

This course does not have any formal prerequisites. However, students would benefit from having some familiarity with basic statistical concepts that are taught in introductory statistics courses at the undergraduate level. These include populations and sampling, probability, measures of central tendency and variation, the normal distribution, and simple univariate tests of means and proportions. I assume that you have taken an undergraduate course in statistics. If you have not, please contact me for some introductory materials that you should read at the beginning of the course.

No formal training or background with R or other statistical computing environments is expected.

Course Goals

- Understand the basic logic of quantitative methods, including statistical modeling.
- Learn how to ask questions about quantitative methods and coding.
- Estimate and interpret linear regression models in R.
- Write clean, reproducible, legible code in R.
- Become familiar with presenting quantitative results.

This course focuses on the practical application and substantive understanding of the linear regression model, rather than a full expounding of the mathematical details and statistical theory underlying these models. We will work closely with real data throughout the semester, which will also introduce students to the process of data management.

Course Site

All course materials, including lecture slides and assignments, will be available on my **Github** website; we will discuss how to navigate Github in the first week of class. I will usually push lecture slides and assignments to our course website roughly one week ahead of time. This course will not involve using Git/Github directly[1]; students can simply download the latest version of lecture slides and assignments. Students will then turn in completed assignments on **Canvas**.

Statistical Software

We will use R extensively throughout the course. You can download both R and RStudio here: https://posit.co/download/rstudio-desktop/.

Books

We will read from several books, which are all available for free at the links below:

- *Statistical Inference via Data Science: A ModernDive into R and the Tidyverse (Second Edition)*
- *ggplot2: Elegant Graphics for Data Analysis*
- *Causal Inference: The Mixtape*
- *Regression and Other Stories*
- *An Introduction to Statistical Learning with Applications in R*

Labs

The goal of the lab time will be for you to develop familiarity with programming in R. Labs will be oriented around the weekly assignments, which are organized around (1) data management, (2) the

---

[1] If you'd like to learn more about using Git yourself, this is a useful tutorial. Github is a useful code development platform that uses Git for version control of all your files. This means that your files in Github are tracked and accessible across any computer, and you can always revert to a previous version of any file. This turns out to be particularly helpful when coding, because sometimes you change something and your code doesn't work anymore; with Git, you can simply revert to a previous version of the code that you know worked.

production of good, legible code (3) implementation of statistical concepts, and (4) analysis of results. Lab attendance is an expected part of the course. If you miss multiple lab sessions your final grade may be reduced.

Evaluation

Weekly assignments. Weekly assignments will be relatively short and primarily involve R data analysis. In addition to material relating directly to the weekly content, part of your weekly assignment will involve the selection, cleaning, and analysis of real data. This sequence will provide you with the opportunity to become familiar with the practical necessities involved with quantitative research projects. Each item on the assignments will receive an overall grade of B+, A, or A+. Because our class is relatively small, we have the capacity to individualize any issues or concerns with assignment performance. If you are worried about your performance on assignments, it is your responsibility to come to my office hours, where we can develop an individualized plan to address your concerns. All assignments are due by midnight on the Sunday after they are assigned.

Final paper. Each student will write a final paper that uses statistical methods from the course to examine an interesting research question. The paper will be about 10 double-spaced pages long. In this paper, students will briefly discuss their research questions and data and then spend most of their time discussing analyses. Papers will require a minimum of one table and one figure. More details will be provided during the first weeks of class.

Attendance and participation. You are required to attend both lectures and labs. Your presence in lecture is critical for your full understanding of materials. And your presence in the lab is not only important for your own work, but for the help and assistance you might be able to provide to others. I do not have a formal component of your final grade attributable to attendance because I expect you to treat your studies and this course with the appropriate level of professionalism and dedication that early career PhD and Master's students should hold. If you run into unusual circumstances, it is your responsibility to check in with Nick to get potential issues squared away. With that said, if your attendance becomes unreliable (think missing two lectures/labs by week 8 without an in-person conversation with Nick) your grade risks being lowered due to poor attendance.

Grading

The course grades are based entirely on applied assignments and the final paper but can be lowered below a B+ with seriously delinquent attendance.

Notes

- **As you embark on your PhD journey, I highly recommend you invest in two platforms that will save you tons of time and headache over the years ahead.**
    - First, choose a free platform to use for backing up your files. I use Dropbox and Git.
    - Second, choose a free platform to use for tracking literature and formatting citations. I use Zotero.
- **For those who prefer another programming language (e.g., Stata, Python), it is not strictly necessary for assignments to be completed in R.**
    - For example, if students are more comfortable using Stata, students can access Stata on their lab computers in Anderson Hall or in the computer labs in the Sociology department. Students can also access Stata virtually via AppsToGo. Instructions for accessing AppsToGo can be found here. You navigate to the website https://apps.umn.edu in a non-Chrome browser and access Stata through the "Apps" section. You can contact UMN help here if you encounter difficulties accessing AppsToGo. As this course is designed for early career PhD students, you should think about your choice of statistical software as a long-term investment rather than simply a means for completing the assignments in this course. My personable recommendation is to invest in R for the following reasons:
        - R is free and open source. Stata is proprietary and you must purchase a license.
        - R is an object-oriented programming language. As such, it is much more like other programming languages you might encounter throughout your career than Stata, which is very specifically designed to analyze tabular datasets.
        - R is widely used across academic, industry, and government jobs.

      These features make R an incredibly flexible programming language for a wide variety of tasks related to statistical learning. Ultimately, it is up to students which language they would like to invest in for their PhDs and beyond.
- **It is natural to feel overwhelmed by this course.**
    - I try not to make the workload for this course overwhelming (e.g., assignments), but one thing is inevitable: you will feel like there is a huge set of "rules" you don't know related to quantitative methods, and it will seem impossible to ever learn everything. Part of this course, as an introduction to quantitative methods, is accepting that this feeling will never go away – I still have it! There are a few things I want to note about this feeling as we begin the course:
        - First, when you read published quantitative articles or listen to research presentations, it will seem like the author knows everything there is to know about the methods they are using. We are often trained and socialized to present quantitative research designs and results with total confidence. In turn, consuming these polished papers and talks can make students feel terrified of conducting their own independent quantitative projects: "I can't possibly know everything, I will make mistakes, what happens when I am found out?" The reality is that most great quantitative researchers feel this

way all the time, it has just been scrubbed out of the clean, polished academic products that you see. Remember to be patient with yourself.

- Second, graduate school is an opportunity to figure out how to accept and navigate this feeling of never being able to know everything. How do you learn? How do you read? How far do you want to go on learning certain quantitative methods? Sometimes there is this idea that topics and theory classes teach you to think, while methods course teach you "skills." We will certainly learn skills in this class (e.g., how to implement different models in R), but methods are also theory. This is a class about thinking, about asking research questions, and about what you can and can't learn through quantitative methods. It will be important for us to take a step back frequently to consider this bigger picture.
- Third, this class requires asking questions when you're confused. Especially with coding, there is a high risk of panicking, just trying to make your code run, etc. But I want you to think of coding more like learning a new language. We want to focus on learning to think about data, think about what exactly you're trying to do in your code and how to ask precise questions about that. If you master this skill, then it doesn't really matter which programming language you use, how much you "know" R, etc. Learning to ask questions will take a lot of practice, but eventually you won't get stuck on anything longer than a few minutes because you'll know exactly how to find the answer to your question.

- **Most of the content in this course is cumulative, meaning it is essential for you to keep up with course materials.**
  - If you find yourself falling behind, ask for help! Come to my office and talk with me.

Week	Day	Topic	Reading, etc.
1	9/3	Introduction to the course	
2	9/8	**No class!**	
	9/10	Introduction to R and ggplot2	*ModernDive*, Getting Started with Data in R 1.1-1.3 *R for Data Science*, Data Visualization 1.1-1.2
3	9/15	**No class!**	
	9/17	Quantitative research methods	*Causal Inference: The Mixtape*, Chapter 2: Probability and Regression Review 2.1-2.12
Part 1: Probability and regression			
4	9/22	Linear regression I	*Causal Inference: The Mixtape*, Chapter 2: Probability and Regression Review: 2.13-2.25
	9/24	Linear regression II	
5	9/29	Sampling distributions and statistical inference I	*Regression and Other Stories*, Chapter 4: Statistical Inference
	10/1	Sampling distributions and statistical inference II	
6	10/6	Simulation I	*Regression and Other Stories*, Chapter 5: Simulation
	10/8	Simulation II	
7	10/13	Regression with multiple variables I	*Regression and Other Stories*, Chapter 10: Linear Regression with Multiple Predictors, 10.0-10.3
	10/15	Regression with multiple variables II	
Part 2: Introduction to causal inference			
8	10/20	DAGs	*Causal Inference: The Mixtape*, Chapter 3: Directed Acyclic Graphs
	10/22	Moderation and mediation	
9	10/27	Potential outcomes I	*Causal Inference: The Mixtape*, Chapter 4: Potential Outcomes Causal Model 4.0-4.1
	10/29	Potential outcomes II	
Part 3: Introduction to panel data			

10	**11/3**	Panel data I	*Causal Inference: The Mixtape,* Chapter 8: Panel Data
	11/5	Panel data II	
11	**11/10**	Panel data III	
	11/12	Panel data IV	
Part 4: Regression diagnostics
12	**11/17**	Assumptions and diagnostics	*Regression and Other Stories,* Chapter 11: Assumptions, Diagnostics, and Model Evaluation
	11/19		
13	**11/24**		
	11/26	**Thanksgiving week – no class!**	
14	**12/1**	Transformations and regression	*Regression and Other Stories,* Chapter 12: Transformations and Regression
	12/3		
Part 5: Review
15	**12/8**	What have we learned?	
	12/10	What good is a regression?	Spirling, A. & Stewart, B. (2024). "What good is a regression? Inference to the best explanation and the practice of political science research." *Working Paper.*

COLLEGE OF LIBERAL ARTS POLICIES

FINAL EXAMINATIONS: (see schedule on the Calendar web site at onestop.umn.edu/calendar/final-exam-times): You are required to take final examinations at the scheduled times. Under certain circumstances, however, you may request final examination schedule adjustment in your college office. Instructors are obligated to schedule make-up examinations within the final examination period for students who have three final examinations within one calendar day. Instructors also are encouraged to reschedule examinations for students with religious objections to taking an examination on a given day. You must submit your request for an adjustment in your schedule at least two weeks before the examination period begins. For assistance in resolving conflicts, call the CLA Student Information Office at 625-2020. If you miss a final, an F or N is recorded. You must obtain the instructor's permission to make up the examination. Final examinations may be rescheduled by the instructor only through the official procedure for that purpose (as noted on the above web page). Final examinations may not be scheduled for the last day of class or earlier or for a Study Day. If an examination is rescheduled at the instructor's request, and you have an examination conflict because of it, you are entitled to be given the final examination at an alternative time within the regularly scheduled examination period for that semester.

CLASS ATTENDANCE: As a student, you are responsible for attending class and for ascertaining the particular attendance requirements for each class you are enrolled in. You should also learn each instructor's policies concerning make-up of work for absences. Instructors and students may consult the CLA Classroom, Grading, and Examination Procedures Handbook for more information on these policies (https://policy.umn.edu/education/makeupwork).

CLASSROOM BEHAVIOR: You are entitled to a good learning environment in the classroom. Students whose behavior is disruptive either to the instructor or to other students will be asked to leave (the policies regarding student conduct are outlined on-line at https://communitystandards.umn.edu/policies-and-procedures/what-kind-consequences-could-i-face

ELECTRONIC DEVICES: University instructors may restrict or prohibit the use of personal electronic devices in his or her classroom, lab, or any other instructional setting. For the complete policy, visit: http://policy.umn.edu/education/studentresp

SCHOLASTIC CONDUCT: The University Student Conduct Code defines scholastic dishonesty as follows:

> Scholastic Dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis. Students cannot evade (intentionally or unintentionally) a grade sanction by withdrawing from a course before or after the misconduct charge is reported. This also applies to late withdrawals, including discretionary late cancellation (also known as the "one-time-only drop").

The use of Artificial Intelligence (AI) may constitute Academic Misconduct, please see your course instructor for their specific policy on the use of such technologies in their course. For the complete policy, visit: http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf

STUDENT MENTAL HEALTH AND STRESS MANAGEMENT: As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. University of Minnesota services are available to assist you with addressing these and other concerns you may be experiencing. You can learn more about the broad range of confidential mental health services available on campus via http://www.mentalhealth.umn.edu/

ACADEMIC ADVISING: Students need to see their College adviser (i.e. 'Advisor of Record' or 'Degree Advisor') at least one a year.   Most CLA students are advised out of Johnston Hall and all CLA students can schedule an appointment or do a Drop-In same day appointment here: https://cla.umn.edu/undergraduate-students/advising/cla-advising-johnston-hall.

CAREER ADVISING: CLA Students are encouraged to connect with CLA Career Services no later than Spring of their Sophomore year. Various online resources, courses, and scheduling help for both appointments and Drop-Ins can be found here: https://cla.umn.edu/undergraduate-students/cla-career-services

SOCIOLOGY DEPARTMENT POLICIES

GRADE INFORMATION: Grades are due in the Office the Registrar within 3 business days after the final examination. The Department of Sociology requires A-F registration in courses required for the major/minor. University regulations prescribe the grades that will be reported on your transcript. No information regarding grades will be released by the department office staff to anyone except designated personnel in Records and college offices. Students may access their own grades through the MyU Portal. Your Canvas grade may differ from your final grade, please see the instructor for details. Instructors establish ground rules for their courses in conformity with their department policies and are expected to explain them at the first course meeting. This includes announcement of office hours and location, the kind of help to be expected from the instructor and teaching assistants, and tutorial services, if available

INCOMPLETES: It is the instructor's responsibility to specify conditions under which an Incomplete (I) grade is assigned. Students should refer to the course syllabus and talk with the instructor as early as possible if they anticipate not completing the course work. Coursework submitted after the final examination will generally be evaluated down <u>unless</u> prior arrangements are made <u>in writing</u> by the instructor. University policy states that if completion of the work requires the student to attend class in substantial part a second time, assigning an "I" grade is NOT appropriate. Incompletes are appropriate only if the student can make up the coursework independently with the same professor. Students need to have completed a substantial portion of the course in order to be even considered for an Incomplete. The college does not permit you to submit extra work to raise your grade unless all students in the class are afforded the same opportunity.

MAKE-UP EXAMINATIONS: Arrangements for special examinations must be made directly with the instructor who taught the course and who is responsible for approving and supervising the examination or making individual arrangements. Circumstances for missing an exam include, but are not necessarily limited to: verified illness, participation in athletic events or other group activities sponsored by the University, serious family emergencies, subpoenas, jury duty, military service, and religious observances. It is the responsibility of the student to notify faculty members of such circumstances as far in advance as possible.

GRADE CHANGES: Grades properly arrived at are not subject to renegotiation unless all students in the class have similar opportunities. Students have the right to check for possible clerical errors in the assignment of grades by checking with the instructor and/or teaching assistant.

Students with justifiable complaints about grades or classroom procedures have recourse through well-established grievance procedures. You are expected to confer first with the course instructor. If no satisfactory solution is reached, the complaint should be presented in writing to the department Director of Undergraduate Studies or the Coordinator of Undergraduate Advising (909 Soc Sci). If these informal processes fail to reach a satisfactory resolution, other formal procedures for hearing and appeal can be invoked.

DISABILITY SERVICES: Students with disabilities that affect their ability to participate fully in class or to meet all course requirements are encouraged to bring this to the attention of the instructor so that appropriate accommodations can be arranged. For more info contact Disability Resource Center in 180 McNamara. https://disability.umn.edu

SEXUAL HARASSMENT: "Sexual harassment" means unwelcome sexual advances, requests for sexual favors, and/or other verbal or physical conduct of a sexual nature. Faculty, Graduate Students, and full time Staff are Mandated Reporters for prohibited conduct. Such behavior is not acceptable in the University setting. If you have experienced sexual misconduct, discrimination, harassment, or related retaliation; or if you have questions about any EOAA-related issue, please contact Equal Opportunity and Affirmative Action (EOAA) at (612) 624-9547 or eoaa@umn.edu For the complete policy, visit policy.umn.edu/hr/sexharassassault

SOCIOLOGY PROGRAMS INFORMATION: The Sociology Department offers two options for the Bachelor of Arts degree and a Bachelor of Science degree. Students interested in majoring in Sociology should view the online-information session about the major. Further information can be obtained from the following persons via email or online at http://cla.umn.edu/sociology

      General information, Sociology Department – soc@umn.edu
      Coordinator of Undergraduate Advising, Bobby Bryant – socadvis@umn.edu
      Director of Undergraduate Studies, Prof. Jane VanHeuvelen – jvanheuv@umn.edu
      Soc Honors Faculty Representative, Prof. Jane VanHeuvelen – jvanheuv@umn.edu
      Director of Graduate Studies, Professor Michelle Phelps – phelps@umn.edu
      Graduate Program Associate, Evva Parsons – eparsons@umn.edu

Undergraduate jobs, internships, volunteer and research opportunities, scholarships, and much more can be found in the Undergraduate Resources site - https://sociologyundergrad.wordpress.com/