# SOC 271B: Methods of Sociological Research - Graduate Statistics I

## UC Berkeley Department of Sociology, Fall 2022

**Lecture: Thursdays, 12PM-1:59PM, 402 SSB**

**Lab: Thursdays, 4-6PM, 402 SSB**

**Instructor: Danya Lagos, dlagos@berkeley.edu**

- Office Hours: Mondays, 10AM-12PM. Book at https://calendly.com/dlagos

**GSI: Paul Salamanca, salamanca@berkeley.edu**

- Office Hours: Wednesdays, 2PM-4PM. Book at https://calendly.com/salamancap/271b

## Course Description

This is the first of two courses on quantitative data analysis designed for sociology Ph.D. students. This course serves as an introduction to statistics for sociological research primarily for sociology graduate students who do not have extensive experience with quantitative methods. Before beginning this course, students are expected to have some familiarity with basic statistical concepts, including populations and sampling, probability, measures of central tendency and variation, the normal distribution, and simple univariate tests of means and proportions. Principal activities include: 1. Exploring the statistical concepts and methods that social scientists most commonly use to gather and analyze quantitative evidence. 2. Using RStudio (an open-source program for data analysis) to put those skills into practice. 3. Applying the skills to sociological data to gain facility and confidence in the use of these methods.

## Course Goals

After successfully completing this course, you will be able to:
1. Use R and tidyverse to manipulate data and estimate statistical models.
2. Apply causal inference concepts to analyzing randomized experiments and observational data.
3. Account for biases in survey samples and identify unobserved groups.
4. Estimate statistical models and use them to evaluate competing social theories.
5. Use probability to estimate and evaluate levels of uncertainty in empirical findings.

This course will emphasize practical application and intuition rather than mathematical details or statistical theory. (Students desiring more mathematically-oriented training are welcome to consult the instructor for other opportunities on campus.) We will work closely with real data throughout the semester in order to learn by doing. Lab sessions will typically include one hour of learning skills necessary to conduct statistical analyses, followed, by an hour of review and individualized support.

## Course Materials

### Textbooks

- Imai, Kosuke and Nora Webb Williams. 2022. *Quantitative Social Science: An Introduction in tidyverse.* Princeton, NJ: Princeton University Press.

- Healy, Kieran. 2017. *Data Visualization: A Practical Introduction.* Princeton, NJ: Princeton University Press. Available for free online at socviz.co

**RStudio Cloud**

Most of the assignments in this course will be completed through RStudio Cloud - a cloud-based interface for running RStudio. It is free for you to use, and does not require you to download anything onto your own computer. I will send all students a link to the course's RStudio Cloud site via email.

**Technology**

Please bring a WiFi-enabled laptop (or tablet with a keyboard) to each course and lab meeting in order to be able to participate in the coding exercises. RStudio does not require installing anything on your computer, so the laptop does not need to be very fancy. If you do not have a laptop already, please let me know as soon as possible so I can work with you to find a solution.

## Schedule of Topics and Readings

| Day | Topic | Readings |
| --- | --- | --- |
| August 25 | Introduction to Course, R, and tidyverse | QSS Chapter 1 |
| September 1 | Causality: Randomized Experiments | QSS Chapter 2.1-2.4 |
| September 8 | Causality: Observational Studies | QSS Chapter 2.5-2.7 |
| September 15 | Measurement and Visualization | QSS Chapter 3 |
| September 22 | Measurement and Visualization | Healy Chapters 4,5 |
| September 29 | Prediction: Prediction Error | QSS Chapter 4.1 |
| October 6 | Prediction: Two-Variable Regression | QSS Chapter 4.2 |
| October 13 | Prediction: Multi-Variable Regression | QSS Chapter 4.3-4.4 |
| October 20 | Probability: Probability Concepts | QSS Chapter 6.1 |
| October 27 | Probability: Conditional Probability | QSS Chapter 6.2 |
| November 3 | Probability: Probability Distributions | QSS Chapter 6.3-6.5 |
| November 10 | Uncertainty: Statistical Inference | QSS Chapter 7.1 |
| November 17 | Uncertainty: Hypothesis Testing | QSS Chapter 7.2 |
| November 24 | No Lecture or Lab (Thanksgiving) | |
| December 1 | Uncertainty: General Linear Model | QSS Chapter 7.3-7.4 |

## Assignments and Deadlines

**Problem Sets (60% of final grade, 12% Each)**

There will be five problem sets assigned throughout the semester designed to give you hands-on experience working through quantitative analysis using RStudio, and to serve as templates for use in your final project and future work. They will be made available via RStudioCloud two weeks in advance of the due date, and are to be submitted through RStudioCloud.

Lab sessions will provide a space for you to ask questions and seek practical assistance from your GSI. You are also encouraged to work on these problem sets in pairs, but writing related to the analyses and interpretation of findings should be your own individual work.

**Problem Set #1 (Due September 15)**   Will be posted on RStudioCloud September 1

**Problem Set #2 (Due September 29)**   Will be posted on RStudioCloud September 15

**Problem Set #3 (Due October 20)**   Will be posted on RStudioCloud October 6

**Problem Set #4 (Due November 10)**   Will be posted on RStudioCloud October 27

**Problem Set #5 (Due December 1)**   Will be posted on RStudio Cloud November 17

**Final Paper (40% of final grade, due Wednesday, December 14)**

Over the course of the semester, you will be encouraged to start looking at one (or more) datasets of your choosing in order to complete a final paper at least 8-10 pages and no longer than 15 pages (excluding citations). This paper should take the form of a journal article, with an introduction, literature review, methods section, results section, discussion section / conclusion, references, tables, and figures. You may use any data and work on any topic that you'd like to explore, but you must meet with both the instructor and GSI by **October 13**.

**Data Sources**   Below are some online repositories where data are available for final projects:

| Data Source | Description |
| --- | --- |
| General Social Survey (GSS) | Omnibus survey collected regularly since 1972 by NORC at the University of Chicago, focused on attitudes, beliefs, practices, and experiences of people living in the United States |
| American National Election Studies (ANES) | National surveys of voters conducted before and after every United States presidential election since 1948 |
| Cooperative Election Study (CES) | Formerly the Cooperative Congressional Election Study, this is a national online survey conducted before and after United States presidential and midterm elections since 2006 |
| ICPSR | Based out of the University of Michigan, the ICPSR hosts a massive database of publicly available and restricted-use data collected by social scientists working on a number of topics |
| Harvard Dataverse | Similar to ICPSR, database of publicly available and restricted-use data collected by social scientists working on a number of topics |
| The ARDA Data Archive | International, US-Based, and State-based data on religion |
| United States Census Bureau | Includes data from U.S. Census Bureau, including decennial census, American Community Survey, Current Population Study, and others |

## Course Policies

### Attendance

Your in-person attendance and participation in all lectures and lab sessions is expected (and vital to the success of the course). You may miss one day without penalty, but any further absence will result in a reduction of your final grade by 5% for each session (lab or lecture) missed. Class and labs will not be recorded, and you will be responsible for catching up with the course material before the next class session.

### Deadlines

You are responsible for turning in all assignments on or before the posted deadlines. Any late work will have its final grade reduced by 10% for each day that it is turned late.

## Supplementary Reading

### Essential Reference Resources

- Choosing the Correct Statistical Test in SAS, STATA, SPSS, and R. *UCLA: Statistical Consulting Group.* https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/

- Damico, Anthony Joseph. *Analyze Survey Data for Free: Step by Step Instructions to Explore Public Microdata from an Easy to Type Website.* http://asdfree.com

**The Politics of Quantitative Methods**

- Johfre, Sasha Shen and Jeremy Freese. 2021. "Reconsidering the Reference Category." *Sociological Methodology* 51(2), 253-269. https://doi.org/10.1177/0081175020982632

- Sprague, Joey and Mary K. Zimmerman. 1989. "Quality and Quantity: Reconstructing Feminist Methodology." *The American Sociologist* 20, 71-86. https://doi.org/10.1007/BF02697788

**Modeling Strategies**

- Martin, John Levi. 2018. *Thinking Through Statistics.* Chicago: University of Chicago Press. ISBN-13: 978-0226567461

- Mood, Carina. 2009. "Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It." *European Sociological Review* 26(1), 67-82. https://doi.org/10.1093/esr/jcp006

**Publishing Guidelines**

- Mustillo, Sarah A., Omar A. Lizardo, and Rory M. McVeigh. 2018. "Editors' Comment: A Few Guidelines for Quantitative Submissions." *American Sociological Review* 83(6), 1281-1283. https://doi.org/10.1177/0003122418806282

**Programming**

- Wickham, Hadley et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4(43), 1686. https://doi.org/10.21105/joss.01686

## Attributions

Portions of this syllabus are adapted from:
David Harding's 271B Syllabus
Robert Apel's Introductory Statistics Course at Rutgers University.
Gabriel Rossman's 210B and 212B Syllabi at UCLA.