Sociology 106: Quantitative Sociological Methods

Spring 2026

Instructor: Kasey Zapatka
Email: kaseyzapatka@berkeley.edu
Lecture: Tuesdays 2:00-3:59 PM, 246 Dwinelle
Office Hours: Tuesdays 11:30 AM-1:00 PM, and 4:00-4:30 PM, Social Science Building Room 444 (signup at https://www.wejoinin.com/kaseyzapatka@berkeley.edu)

Course description and objectives

Sociology 106[1] is an intermediate undergraduate sociology research methods course. It will emphasize the motivation, computation, and interpretation of statistical tests for one or two continuous or categorical variables and give students familiarity with basic regression analysis. The course will also train students in the use of the R statistical programming language for data management and analysis.

Sociology 106 is most appropriate for social science undergraduates who have some familiarity with sociological research methods and wish to learn how to carry out a quantitative sociological research project.

For most weeks, we will have assigned readings and sometimes videos. You should read the assigned readings or watch the assigned videos ***BEFORE*** you come to class. Classroom time will usually be split between a lecture component and a 'lab' component. During the first hour of class, I will introduce the relevant statistical techniques and research concepts for the week in a lecture. Then, in the second hour of class, we will learn how to implement these statistical techniques in the programming language R and apply these techniques to actual data.

By the end of the semester, students should be able to:

- Understand the logic of statistical inference
- Identify appropriate statistical tests for different types of data
- Visualize data and produce descriptive statistics and simple statistical tests using R
- Interpret and communicate statistical results and discuss their relevance in the context of a particular research question

Prerequisites

Previous training in statistics is neither required nor expected. Successful completion of Sociology 5 is a requirement for this course, but other courses that provide an introduction to social science research methods may also suffice. If you have not taken Sociology 5, contact me to obtain permission to enroll.

Required course materials

All of the assigned reading materials will be made available on bCourse. Here's a look at what we'll read:

- **David Lane's Online Statistics** (http://onlinestatbook.com/Online_Statistics_Education.pdf).

[1] I thank Joe LaBriola for their work designing this syllabus in previous semesters, as well as David Harding for his work designing the Sociology 271B class.

- **Hadley Wickham's R for Data Science 2nd Edition** (https://r4ds.hadley.nz), which is also free online.
- We'll read two journal articles to see how social scientists use regression analysis in practice.
  - **Thompson, M.S., & Keith, V. M.** (2001). The Blacker the Berry: Gender, Skin Tone, Self-Esteem, and Self-Efficacy. Gender & Society, 15(3), 336-357. (https://doi.org/10.1177/089124301015003002)
  - **Freeman, L., & Braconi, F.** (2004). Gentrification and Displacement New York City in the 1990s. Journal of the American Planning Association, 70(1), 39–52. (https://doi.org/10.1080/01944360408976337)

You are **required to have a laptop** that can access the internet to take this course.[2] A major component of this course is learning to use the statistical programming language R and getting practice using R to analyze data. So, you'll need:

- **R**[3] (http://cran.rstudio.com), an open-source programming language
- **Positron**[4] (https://positron.posit.co/download.html), a free program that makes working in R much easier and is at the cutting edge of data science right now.

Course requirements

Attendance and participation (10%)

Attendance and active participation in lecture is essential to learn the concepts of the course, as well as how to implement the R commands, data management skills, and workflow strategies you will need to complete the weekly assignments and research paper. Attendance is especially important because this class meets only once per week.

Your grade in this category will be determined by your engagement during both lecture and the data analysis components of class. Phones may not be used during class, and you may use your computer only for course-related purposes. This policy is to support your learning and the learning of other students in the class, and violations of this policy will negatively affect your participation grade.

In addition, your overall grade may be negatively affected by unexcused absences. If you have more than one unexcused absence from lecture, your highest possible grade will be an A-; if you have more than two unexcused absences from lecture, your highest possible grade will be a B+; etc. If you have more than four unexcused absences from lecture, you will receive a failing grade for the course. If an unforeseen emergency arises that causes you to miss lecture, you should let me know as soon as possible the nature of the emergency and prepare to provide some sort of documentation so that the absence can be excused. If documentation is difficult to obtain, please consult with me.

[2] If you wish to take this course and do not have access to a computer, please let me know immediately. Also, if installing R on a Chromebook may be more difficult; please let me know immediately if you are using a Chromebook for this course, and I will try to connect you with someone who can help with the installation.

[3] If you have a Mac, you need to check whether your laptop has an Apple Silicon (M1/M2) or Intel processor and download the appropriate version of R for Mac. You can check the process by:
1. Click the Apple menu ( ) in the top-left corner of your screen.
2. Select **About This Mac**.
3. Look at the **Processor** or **Chip** line:
- If it says **Apple M1**, **M2**, or similar → you have Apple Silicon.
- If it says **Intel Core i5**, **i7**, etc. → you have an Intel Mac.

[4] You'll want to do the same here or just download and use the "Universal" option, which is a slightly larger file and might not be quite as optimized.

Weekly assignments (30%)

Weekly assignments may include a short problem set designed to test your comprehension of the concepts from lecture, or require that you address a research question of your choice by using R to analyze data of your choice using one of the techniques we discussed in the previous lecture. Most weeks there will be a bit of both. You will be provided a .qmd file on bCourses to use as a template to complete the assignment. It is highly recommended that you spend some time during the first week of the semester to choose a single dataset that interests you that you will use throughout the semester for your assignments and to use for your final paper (see below for where to find data).

I encourage you to find data for and work on these assignments with other students. If you do so, you must report who you worked with on the assignment and still must write up your own work in your own words and submit your own R code and output. All students need practice writing up their analysis of data. Copying is plagiarism and will be treated as such (see UC Berkeley's Academic Code of Conduct).

Weekly assignments will be posted on bCourses. Assignments will be graded on the following ordinal scale:

- 0 = not turned in
- 1 = below expectations
- 2 = meets expectations

Assignments are due on bCourses at 11:59 PM on Thursdays; this timing is so that I can review the assignments and provide feedback to you before class the following Tuesday. There will be 11 weekly assignments over the course of the semester. I will drop your lowest weekly assignment score when calculating your grades on weekly assignments. Each assignment is thus worth 3% of your total grade (3%*10 = 30%).

Research paper (40%)

Each student will develop and present a research question of their choice, address it by using the descriptive and inferential techniques presented in the course to analyze data in R, and write a paper summarizing your findings. You may use your weekly data analysis assignments to work on your research question. You may also (if permitted by the other instructor) integrate this assignment directly into a paper you are writing in another class; please provide me with documentation of permission from your other instructor if you wish to do so. If you do not have an idea about a possible research question to pursue, please sign up for my office hours in the first few weeks and I can try to help you find a question you are interested in.

There will be several milestones throughout the semester to help you smoothly progress towards a final paper. I will give you feedback at each step of the way. Your overall grade on the research paper will be the sum of your grades on the following assignments:

- **Paper proposal (5%)**: A (no more than) two-page double-spaced proposal for the final paper will be due on bCourses by Thursday, February 26 at 11:59 PM. The proposal should succinctly state the research question, why we might care about the answer to this question, hypotheses about what you think the answer is to the research question (and why), the data source you will use to answer this question, and the key independent and dependent variables in the data. At this point, you do not need to discuss what statistical techniques you will use to answer your research question.
- **Annotated bibliography (5%)**: Next, you will identify ten scholarly sources related to your

research question. Articles in academic journals, books, and book chapters from edited volumes are all considered to be scholarly sources. Blogs, internet and newspaper articles, Wikipedia pages, etc., are NOT considered to be scholarly sources. After identifying these ten scholarly sources, for each source, you will briefly (one to two paragraphs) describe how this source relates to your research question. This includes, but is not limited to, the overall argument in the article or book, the types of evidence the author(s) use to support their article, and any possible weaknesses or strengths in the paper. This annotated bibliography will be due on bCourses by Thursday, March 19 at 11:59 PM.

- **Revised paper proposal with outline (5%):** Building on the feedback you received on your initial proposal and the work you've completed in your annotated bibliography, you will submit a revised proposal with a detailed outline for your final paper. This revised proposal will be due on bCourses by Thursday, April 16 at 11:59 PM. The revised proposal should be no more than three to four pages double-spaced and should include: (1) a refined research question that incorporates feedback from your initial proposal, (2) a very brief literature review that synthesizes insights from your annotated bibliography and explains how your research contributes to existing scholarship, (3) updated hypotheses and a clear explanation of the theoretical reasoning behind them, (4) a short description of your data source, key variables, and the statistical techniques you plan to use to test your hypotheses, and (5) a detailed outline of your final paper showing the main sections and key points you will cover in each section. This assignment will help ensure that you are on track to complete a strong final paper and will give you an opportunity to receive additional feedback before your in-class presentation.
- **In-class presentation (5%)**: Then, during our class on April 28, you will give a brief presentation (7-10 minutes) to the class on your research paper. By this point, you should have conducted the statistical analyses you have proposed to answer your research question. Thus, in your presentation, you should discuss the research question, motivation for the research question, hypotheses, data, variables, analyses, and the results of your analyses. This is a great opportunity to get feedback on your (mostly-completed) research project from other members of the class as well as from me.
- **Final paper (20%)**: Your final research paper will be due on bCourses by Thursday, May 7 at 11:59 PM. This paper should be 10-12 pages double-spaced, not counting a cover page, tables, or references. The paper will roughly follow the same format as your in-class presentation.

Final exam (20%)

The final exam for our class will be on Monday, May 11 from 11:30 AM to 2:30 PM. The final will be somewhat similar to the weekly homework assignments and be a mixture of 10 multiple choice and 5-10 short answer/regression analysis/interpretation questions. The test will be closed-book and closed-computer, but I'll give you a clear sense of the types of questions during the review week so you can focus your preparation and you will be provided with any formulas or tables that you need during the exam.

Data for assignments and research paper

Over the course of the semester, you will work with one (or more) datasets of your choosing in your assignments and your final paper. You may use any data you like, and are encouraged to consult with me as you choose and begin working with your data. Below are some examples of online repositories where data are available:

- The **Inter-university Consortium for Political and Social Research (ICPSR)** has thousands of surveys. You may search for surveys at https://www.icpsr.umich.edu/sites/icpsr/find-data
- The **Association of Religion Data Archives (ARDA)** also has a voluminous collection of surveys on religion, accessible at https://thearda.com/Archive/browse.asp.
- The **Integrated Public Use Microdata Series (IPUMS)** is the most accessible way to access lots of U.S. and international census and survey data: https://www.ipums.org
- The **Survey Documentation and Analysis** page at https://sda.berkeley.edu/archive.htm This site has opinion data from the General Social Surveys, American National Election Studies, and various studies on racial attitudes and prejudice, among other surveys on household finance and health.
- The **Roper Center at Cornell University** has a massive database of public opinion polls: https://ropercenter.cornell.edu
- The **Google Dataset Search** is a newer resource that serves as comprehensive search tool and could be another good source of data : https://datasetsearch.research.google.com/

Keys to success

- In some ways, learning statistics is like learning a language, and it is important not to be intimidated by new terms and concepts. It is often helpful to write in plain language the meaning of the quantities or concepts represented by a letter or symbol.
- Please ask questions during lecture or during lab if you do not understand. If something is unclear to you, it is probably unclear to other students as well.
- Lectures and labs are planned to allow time for questions and answers.
- Lecture slides will be made available, but are not a substitute for careful note taking.
- For most students, learning statistics requires thinking through how to solve problems. Statistics cannot be learned simply by reading a book or listening to a lecture. You should not expect to fully understand the material until after you have completed the relevant assignment.
- Because most of the material is cumulative, it is essential that you keep up with the course material. If you find yourself falling behind, seek help immediately from me during office hours.
- The most effective way to study for the exams is to do practice problems. As such, the assignments are a critical part of the course. You are strongly encouraged to do homework assignments and study for exams in groups.

Classroom policies

- **Classroom environment:** I intend to help foster a classroom environment where students can seek understanding of the course material free of harassment of any kind. We all come to this class from different backgrounds and with different levels of academic preparation; as such, topics that may seem for some to be 'easy' to understand are likely not easy for everyone. I ask that you respect other students in the class in all ways and remain cognizant of how your words and actions may come off to others. I will not tolerate racist, sexist, homophobic, or other derogatory comments whatsoever in class.

- **Office hours:** I encourage you to come to office hours. Office hours are a good time to introduce yourself, talk about ideas you are thinking about, and ask substantive questions about the course material. Office hours are a great opportunity to start to work on your weekly assignments while being able to ask me for help. I have office hours on Tuesdays before class and right after class (see above). You may sign up for office hours at https://www.wejoinin.com/kaseyzapatka@berkeley.edu or just drop in. I also encourage you to sign up in groups.
- **E-mail:** E-mail is best for asking logistical questions about the course and should not be used for substantive questions about the course materials or about coding. Please re-read the syllabus before sending a logistical question about the course to ensure that your question is not answered here. If you have a substantive question about the course material (i.e., you are not sure you understand a concept from lecture), please make an appointment to meet me in office hours instead. If you cannot meet during my scheduled office hours, e-mail me and we'll figure something out. I will endeavor to respond to e-mails received during the week (8:00 AM Monday-5:00 PM Friday) within 24 hours, and to e-mails received outside this time by 5:00 PM Monday.
- **Accommodation:** I will provide the necessary accommodations to any student who provides me with a letter from a DSP Specialist. Please arrange for me to receive this letter as early in the semester as possible, as I cannot retroactively provide accommodations after I receive this letter. Please notify me over e-mail by the second week of the term about any known or potential extracurricular conflicts (such as religious observances, graduate or medical school interviews, or team activities). I will try my best to help you with making accommodations, but cannot promise them in all cases. In the event there is no mutually-workable solution, you may be dropped from the class.
- **Late work:** The precise due dates of both the weekly assignments and of assignments related to the research paper are listed above. For the weekly assignments, one point (out of two) will be automatically deducted for assignments that are turned in up to one day after the due date; assignments will not be accepted after this. Further, I may not be able to provide timely feedback on late assignments. For assignments related to the research paper, assignments turned in late will be penalized one letter grade for every day late (e.g., from an A to a B). If you have a real emergency, contact me at least 24 hours before the assignment deadline.
  - If you know now that you will have a conflict with the final exam time (**Monday, May 11 from 11:30 AM-2:30 PM**), either do not take the course or speak with me as soon as possible so that we can work out an accommodation. If you have a truly unforeseen emergency that prevents you from attending the exam, contact me as soon as possible.
- **Grading policy:** If you wish to contest a grade, please e-mail me within one week of receiving the grade and outline in writing (1) the assignment you are contesting, (2) the grade you received on the assignment, and (3) why you believe your grade is unfair. I will consider your appeal and may decide to re-grade your assignment. Please note that a re-grade involves a closer scrutiny of your work and may result in you receiving a lower grade.
- **Academic honesty:** UC Berkeley's honor code states that "as a member of the UC Berkeley community, I act with honesty, integrity, and respect for others." I expect that you will follow this code. Copying the work of others (either your classmates or the work of others you find in books or on the internet) and presenting it as your own is plagiarism. This includes output from Generative AI (e.g., ChatGPT). If you are unsure as to whether something constitutes plagiarism, please come talk with me before submitting the assignment! If I find evidence that you have plagiarized work or cheated on an exam, it is grounds to fail you in the course. I take instances of plagiarism or

cheating very seriously and trust that it will not be a problem in our course.

- **Technology policy:** Since much of this course involves data analysis on your computer, you may use a computer during class for class purposes only. Please do not use your cell phone during class. Using your phone or your computer for purposes unrelated to the class will negatively affect your attendance and participation grade.
- **Generative AI**: You are encouraged to use AI tools (e.g., ChatGPT), Google, and online resources (e.g., StackOverflow or StackExchange) to help you understand statistical concepts, debug code, and clarify new terminology. These tools can provide explanations, suggest libraries or packages, and offer sample code to guide your learning. However, do not copy or paraphrase AI-generated content in any assignment, as this counts as plagiarism. The goal is to use these resources to support your learning and problem-solving in statistics and programming, not to replace your own work. If you are going to spend the time taking this course, you should learn as much as you can.

Academic and non-academic resources

Like any other skill, effective writing can be developed with practice and assistance from others. I am happy to discuss your writing during office hours. Other resources that may be helpful include the Student Learning Center Writing Program and Academic Centers located in residence halls. The Student Learning Center also provides tips on effectively managing time and overcoming procrastination. Students struggling with academic or non-academic issues that may be causing elevated levels of stress and/or anxiety may be interested in the services offered by Counseling and Psychological Services at the Tang Center. Counselors at the Tang Center assist students by offering strategies for managing stress and anxiety, and working with counselors can help students develop self-understanding and resolve issues. Counselors at the Tang Center also offer emergency consultations, appointments, workshops, and self-help resources. I note that I (alongside other academic advisers and professors at UC Berkeley) am designated as a 'responsible employee', which means that I must report incidents of sexual violence and/or harassment that I am told about to the Office for the Prevention of Harassment and Discrimination. Students who are concerned about maintaining confidentiality are advised to use confidential resources such as the PATH to Care Center, UHS Social Services, Be Well at Work Employee Assistance, and the Ombuds Office for Students and Post-Doctoral Appointees.

Schedule

Week	Date	Topic	Readings/Videos	Assignments
1	1/20	Introduction		
2	1/27	Statistical computing in R	Positron tutorial; Positron + Quarto tutorial; Wickham 7.1, 7.2, 7.2.1, 7.2.3, 7.5	HW #1 due 2/5
3	2/3	Visualizing distributions	Lane p. 82-85, 92-108, 165-169; Exploratory Data Analysis in R with Positron; Wickham 1.1, 1.3, 1.4, 1.5, 1.6	HW #2 due 2/12

Week	Date	Topic	Readings/Videos	Assignments
4	2/10	Summarizing distributions	Lane p. 123-128, 131-135, 140-153, 170-175	HW #3 due 2/19
5	2/17	Basics of probability theory	Lane p. 189-197, 212-214	HW #4 due 2/26; paper proposal due 2/26
6	2/24	Probability models of distributions	Lane p. 248-265	HW #5 due 3/5
7	3/3	Sampling distributions	Lane p. 300-315, 319-321; seeing theory- central limits theorem	HW #6 due 3/12
8	3/10	Estimates and confidence intervals	Lane p. 328-355, 358-359; seeing theory-confidence intervals	HW #7 due 3/19; annotated bibliography due 3/19
9	3/17	Hypothesis testing	Lane p. 369-391, 398-411, 448-457, 597-607	HW #8 due 4/2
10	3/24	Spring Recess		
11	3/31	Linear regression	Lane p. 461-481; Thompson and Keith (Introduction p. 336-340, Data and Methods p.342-344, Results p. 344-348)	HW #9 due 4/9
12	4/7	Logistic regression	Freeman and Braconi (focus on Methodology, Results)	HW #10 due 4/16; revised proposal with outline due 4/16
13	4/14	Extensions to regression	Thompson and Keith (Results p.349-351, Discussion p. 351-354)	HW #11 due 4/23
14	4/21	Mediation analysis and multicollinearity		
15	4/28	Research presentations 1		
16	5/5	Presentations 2 & RRR Week		Final paper due 5/7
17	5/11	Final	11:30 AM - 2:30 PM	