

Instructor: Jacob Habinek
Office Hours: Fridays, 2-4pm

Email: habinek@berkeley.edu
Office: 494 Barrows Hall

Meetings: Fridays, 12-2pm, B5 Hearst Annex
Website: <https://bcourses.berkeley.edu/courses/1363500>

Course description and learning goals

Sociology 106 is an intermediate undergraduate social science research methods course. It will emphasize the motivation, computation, and interpretation of statistical tests for one or two continuous or categorical variables. 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 research project.

Classroom time will include both a lecture and a laboratory component. During the first hour of class, the instructor will introduce the relevant statistical concepts and tools for the week. The second hour of class will consist of a data analysis practicum in which students will learn how to apply statistical techniques to the analysis of real data using R. After successfully completing this course, you will be able to:

1. Understand the basic logic of statistical inference
2. Identify the appropriate statistical test given a specific type of data
3. Compute descriptive statistics, graphs, tables, and simple statistical tests using R
4. Interpret the results of statistical tests and discuss their relevance to a particular research question

Prerequisites

Previous training in statistics in 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 the instructor to obtain permission to enroll.

Required course materials

There are no required readings for the course. The recommended textbooks are:

Sheldon Ross, *Introductory Statistics* (3rd edition), OR
David Lane, *Online Statistics* (<http://onlinestatbook.com/>)
Jane Miller, *The Chicago Guide to Writing about Numbers* (2nd edition)

You are not required to read anything from the textbooks before lectures. We will go over all concepts in lecture first. The Ross and Lane books are interchangeable, and intended to help you consolidate and review the statistical concepts covered in lecture. The Miller book is intended to help you with completing the data analysis assignments and the final paper.

You are required to have a laptop that can access the internet to take this course. In order to complete the assignments, you will be downloading, installing, and using statistical computing software, as well as finding data on the web. If you wish to take this course and do not have access to a computer, please speak to the instructor immediately.

Course requirements

The course involves the following activities:

Active participation during lectures and workshops (10% of your final grade): class time is your chance 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 assignments. Checking email, using Facebook, looking something up on Wikipedia, online gambling, etc. during class time is disrespectful; doing so will affect your grade in a meaningful way.

Weekly homework assignments (20%): each week you will have a homework assignment. The assignment 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 analyzing data of your choice using one of the techniques discussed in the previous lecture. Most weeks there will be a bit of both. The bCourses site has links to many data sources, but you may use any data you like. Homework assignments will be due by 5:00pm on bCourses the night before class; late homework will not be accepted. Homework will be graded on the following ordinal scale: 0 = not turned in; – = below expectations; + = meets expectations; ++ = exceeds expectations.

Midterm exam (15%): the midterm exam is scheduled for 12:00pm on Friday, October 23rd. The midterm questions will be similar to the weekly homework assignments, except that you will not be asked to analyze data using R. The test will be closed book and closed notes, but you will be permitted a calculator. Make-up exams are only given when the circumstance clearly warrant, and then only when the student has given the instructor advanced notice of his or her absence.

Final paper (40%): the final paper will address a research question of your choice using the descriptive and inferential techniques presented in the course. It is essentially a more polished and comprehensive version of the weekly data analysis assignments; students may use their weekly homework assignments to “build up” to their final paper, or conduct an entirely new analysis. You are encouraged (if permitted by the other instructor) to integrate this assignment directly into a paper you are writing for another class. We will discuss the paper further later in the semester. It should be about 10 to 20 pages in length. The final paper will be due on bCourses by 5:00pm on Monday, December 14th.

Final exam (15%): the final exam is scheduled for 11:30am on Friday, December 18th. It will be similar in format to the midterm exam. The final exam will be comprehensive in content, but will focus on evaluating your ability to use and interpret the results of the statistical tests covered in the latter part of the course.

Schedule

| Week | Date | Topic | Ross | Lane |
|--|-------|--|---------------|-----------------------|
| 1 | 8/28 | Introduction | --- | --- |
| Part I: finding and describing data | | | | |
| 2 | 9/4 | Statistical computing and finding data | --- | --- |
| 3 | 9/11 | Visualizing distributions | Ch. 2.1-2.3 | Ch. 2.a, b.2, b.4-7 |
| 4 | 9/18 | Summarizing distributions | Ch. 3.1-3.5 | Ch. 3.a-b |
| Part II: probability and sampling | | | | |
| 5 | 9/25 | Basics of probability theory | Ch. 4.2-4.5 | Ch. 4.a-e |
| 6 | 10/2 | Probability models of distributions | Ch. 5.1-6.7 | Ch. 4.g-h, Ch. 7.a-e |
| 7 | 10/9 | Sampling distributions | Ch. 7.1-7.5 | Ch. 8.a-e |
| 8 | 10/16 | Estimates and confidence intervals | Ch. 8.1-8.7 | Ch. 10.a-e.4 |
| 9 | 10/23 | MIDTERM | --- | --- |
| Part III: statistical inference and tests | | | | |
| 10 | 10/30 | Hypothesis testing | Ch. 9.1-9.5 | Ch. 11.a-i, Ch.12.a-c |
| 11 | 11/6 | Analysis of variance | Ch. 10.1-10.5 | Ch. 15.a-c |
| 12 | 11/13 | Chi-square test | Ch. 10.6 | Ch. 17.a-e |
| 13 | 11/20 | Linear regression | Ch. 12.1-12.5 | Ch. 14.a-e |
| 14 | 11/27 | <i>No Class: Thanksgiving</i> | --- | --- |
| 15 | 12/4 | Logistic regression | --- | --- |
| End of class: review and final exam | | | | |
| 16 | 12/11 | REVIEW SESSION | --- | --- |
| 17 | 12/14 | FINAL PAPER DUE BY 5:00pm | --- | --- |
| | 12/18 | FINAL EXAM, 11:30am-2:30pm | --- | --- |