

SOCIOL 1156

Statistics for Social Sciences

Fall 2024

Lectures: Mondays and Wednesdays, 10:30-11:45am, WJH 105

Sections:

Thursdays, 3:00-4:00pm, WJH B6

Fridays 10:30-11:30am, WJH 950

Course website: <https://canvas.harvard.edu/courses/136717>

Instructor

Joscha Legewie

Email: jlegewie@fas.harvard.edu

Office hours: Wednesdays 12-1:30pm, WJH 630 (Sign up here: <https://calendly.com/jlegewie/officehours>)

Teaching Fellow

Charlotte O'Herron

Email: charlotteoherron@fas.harvard.edu

Office hours: Thursdays 4:00-4:30pm, WJH B6 (Sign up here: <https://calendly.com/charlotteoherron/charlotte-o-herron-office-hours>)

Catalog Description

This course introduces students to quantitative data analysis in the social sciences. It covers the basics of research design and the use of empirical evidence. Students will learn about descriptive and inferential statistics, including regression analysis. The course aims to equip students with the skills needed to analyze data and effectively communicate their findings in research reports.

Course Goals

This course focuses on the art of drawing meaningful insights from quantitative data generated by surveys, experiments, governments or from countless other sources. The course introduces methods for describing and interpreting these data. It will provide you with useful tools to effectively investigate your substantive sociological interests.

The course is structured to provide a comprehensive understanding of data analysis, starting with foundational concepts and progressing to more advanced techniques. We begin with an introduction to working with data in R and RStudio, spending several sessions to ensure a solid grasp of these essential tools. Next, we delve into describing data, covering various methods to summarize and visualize information. A case study on racial disparities in police stops will allow us to apply these skills to real-world data. Following this, we cover statistical inference and uncertainty by introducing sampling distributions, standard error, confidence intervals and hypothesis testing. After a midterm review and exam, we'll move on to regression analysis, dedicating several sessions to this crucial topic. The course then introduces the concept of causality and its importance in data interpretation. Throughout the semester, we'll intersperse case studies to provide practical application of the concepts learned. This structure allows us to build a strong foundation in data analysis techniques while also providing opportunities to apply these skills to meaningful, real-world scenarios.

This course is aimed at Sociology concentrators, but students from other concentrations are welcome. We will discuss statistical methods used by social scientists but pay relatively little attention to the statistical

theory underlying these methods. Students who have an interest and background in mathematics are encouraged to consider taking more advanced courses.

Grades and Assignments

Any student needing academic adjustments or accommodations should present their letter from the Accessible Education Office (AEO) and speak with me by the end of the second week of the term. All discussions will remain confidential, although AEO may be consulted to discuss appropriate implementation.

If your physical or mental health or other circumstances interfere with your ability to succeed in the class or to complete an assignment on time, please write to me as soon as possible so we can discuss appropriate extensions or accommodations. If you prefer, you can first speak with your resident dean and ask them to be in touch with me.

The course assignments include:

1. *Six Homework Assignments.* You are required to complete six homework assignments throughout the semester. Collaboration on homework assignments is encouraged, for example, during the sociology study tables (33.3% or grade)
2. *Midterm final exams.* You will demonstrate your learning of course concepts through two exams. The first midterm exam will be in-class. The second midterm exam is a take home exam and like an extended homework assignment. Collaboration is not allowed for exams (33.3% or grade)
3. *Final project:* The final project is a data analysis project about a sociological research question of your choice. You will find a dataset, pose a research question, and answer it using data. We recommend that you use the General Social Survey but you can use any dataset of your choice. You will write a short essay (5-7 pages) that introduces your question, describes your data and methods and discusses your findings (33.3% or grade).

As part of the final project, you are required to complete the following assignments:

1. Describe the topic and data for your data essay (300 words) (10%)
2. *Analysis plan and descriptive statistics.* (30%)
3. *Final Paper* (60%)

Submission deadline: All written assignments are due at 10:30am and should be submitted through the Assignments tab on Canvas. Assignments submitted after the deadline will incur a 10% grade deduction immediately. For each additional 24-hour period (or part thereof) that the assignment is late, an extra 5% will be deducted from the grade.

Sometimes unexpected situations arise that prevent students from completing their assignments on time. I aim to provide appropriate accommodations for students'™ circumstances and to do so in a way that is equitable and transparent. You may each take one automatic, no-questions-asked, 48-hour extension (i.e., until 10:30am on Friday) on any of the six homework assignments, or the abstract or analysis plan for the final paper (not the exams or final paper). There is no need to request permission for this extension in advance or to alert the instructors that you are taking it. Your 48-hour extension will automatically be applied to your first late assignment and cannot be reallocated to a different assignment or shared between assignments. For longer extensions, please be in touch *before* the deadline the assignment is due. I will typically work with you and your resident dean to determine an appropriate accommodation.

We strive to grade all assignments submitted on time within one week of the submission deadline.

Sections and Teaching Fellows

Charlotte O'Herron (charlotteoherron@fas.harvard.edu), doctoral student in Sociology, will be your teaching fellow (TF) for this course.

The TF will hold section meetings most weeks starting in the week of September 9th. Section meetings will cover elaboration of materials from lecture, instruction in coding, and worked examples illustrating course material, including questions from the homework assignments.

The TF will also hold student hours.

Textbooks

Textbooks are not required for this course. However, if you want to read about the topics in the course or learn about them from a slightly different perspective, we recommend the following two books:

Imai/Williams. 2022. Quantitative Social Science: An Introduction in tidyverse. Princeton University Press.

Wickham/Jetinkaya-Rundel/ Grolemund. 2023. R for Data Science. O'Reilly Media (freely available [here](#))

Collaboration and Academic Integrity

Collaboration. Discussion and the exchange of ideas are essential to academic work. You are encouraged to work with your classmates on homework assignments. You may also consult your classmates on the choice of paper topics and share sources such as possible datasets. However, you should ensure that any written work you submit for evaluation is the result of your own research and writing and that it reflects your own approach to the topic. You must also adhere to standard citation practices in this discipline and properly cite any books, articles, websites, lectures, etc. that have helped you with your work. If you received any help with your writing (feedback on drafts, etc. from peers or course instructional staff), you must also acknowledge this assistance.

Harvard College Honor Code. Members of the Harvard College community commit themselves to producing academic work of integrity – that is, work that adheres to the scholarly and intellectual standards of accurate attribution of sources, appropriate collection and use of data, and transparent acknowledgement of the contribution of others to their ideas, discoveries, interpretations, and conclusions. Cheating on exams or problem sets, plagiarizing or misrepresenting the ideas or language of someone else as one’s own, falsifying data, or any other instance of academic dishonesty violates the standards of our community, as well as the standards of the wider world of learning and affairs.

Plagiarism and other forms of academic dishonesty will be dealt with severely. All students are expected to read and comply with Harvard College’s Academic Integrity and Academic Dishonesty [guidelines](#).

Generative AI. We encourage and teach you how to use Generative AI such as ChatGPT for data analysis. However, we impose certain restrictions on the use of generative AI for course work.

- **Dos:** Use AI/LLMs to explain, comment, debug, and improve R code that you drafted. Also use AI/LLMs to get summaries of text passages and concepts that you do not understand. This will help you to read complex texts and better understand difficult topics. You can also use AI/LLMs to discuss your understanding of the course materials and ask things that confuse you.
- **Don'ts:** Don't use AI/LLMs to generate, draft, and write your R code. Don't use AI/LLMs to write your assignments.

Course Schedule

The schedule of topics listed below is approximate; topics may take more or less time than anticipated. Remember that all written assignments are due at 10:30am and should be submitted through the Assignments tab on Canvas.

Week	Date	Topic	Assignments
	4-Sep	Introduction	
Week 1	9-Sep	Working with Data in R and R Studio	
	11-Sep	Working with Data in R and R Studio	
Week 2	16-Sep	Working with Data in R and R Studio	
	18-Sep	Describing Data	HW1 due
Week 3	23-Sep	Describing Data	

	25-Sep	Describing Data	
Week 4	30-Sep	Case Study	HW2 due
	2-Oct	Statistical Inference and Uncertainty	
Week 5	7-Oct	Statistical Inference and Uncertainty: Confidence Intervals	Final Project Abstract due
	9-Oct	Statistical Inference and Uncertainty: Hypothesis Testing	
Week 6	14-Oct	No class	
	16-Oct	Review	HW3 due
Week 7	21-Oct	Midterm Exam	Midterm exam
	23-Oct	Regression	
Week 8	28-Oct	Regression	
	30-Oct	Regression	HW4 due
Week 9	4-Nov	Regression	
	6-Nov	Case Study	Final Project analysis plan due
Week 10	11-Nov	No class	HW5 due
	13-Nov	Causality	
Week 11	18-Nov	Causality	
	20-Nov	Causality	
Week 12	25-Nov	Case Study	HW6 due
	27-Nov	No class	
Week 12	2-Dec	Final Review	

4-Dec Final Review

Midterm exam
due

TBD

Final paper due