Sociology 7 The Power of Numbers: Quantitative Data in the Social Sciences

Fall 2024 UC Berkeley

Instructor: Professor Linus Huang, Continuing Lecturer
Office Hours: Mondays 1:15-3:30 PM, 487 Social Sciences Building (drop-in), or by appt
Final Exam: in-class, Monday, December 16 2024, 8-11 AM

What will this course be about?

We live in a world of numbers. Numbers say things about out financial status (income, net worth, credit score), the quality of the schools we're considering (*U.S. News & World Report* college ratings), who will win the election in November (polling results), which products we should buy (Amazon, Yelp ratings), how we should prepare for the day (temperature forecasts, chance of rain), and/or our academic status (GPA). Because the world is so full of numbers, not only do social scientists need to understand them to investigate the social world, but we all need to understand numbers to be informed citizens and to just make everyday life choices.

This course provides a starting point for this. We will look at the issues surrounding how we use numbers to describe the world (descriptive statistics) and how we use numbers taken from a sample to generalize about large populations (inferential statistics). But we will also acquaint ourselves with the nuts and bolts of retrieving data and preparing it for analysis—things that will aid us in not just being consumers of data, but also producers of data.

By the end of the course, you will be able to:

- critique and produce basic data visualizations;
- manipulate and analyze data in spreadsheets;
- calculate and explain basic statistical measures of central tendency, variation, and association;
- apply and explain basic concepts of sampling and selection; and
- think critically about reported statistics and quantitative social science.

Instructional Mode

Soc 7 will be in-person only. There will be no recordings of lectures available.

Class time will be used for lecture and (mostly) demonstration. Much of the demonstration will be on the chalkboard/whiteboard. But a lot of it will be using a spreadsheet (Excel), exploring datasets hosted on websites, and/or using various stats web apps. It is **highly** recommended that you follow along during these demonstrations—the problem sets (see below) for the course will require you to do all of these yourself!

Readings and Resources

Our 40 class meetings will be divided into about 14 topics. Most of these topics will have associated required readings, which should be completed prior to the first class meeting scheduled to cover them. All are available in PDF format on the course website; there are no texts or paper readers to purchase for this course.

Especially in the first part of the course, we will be making heavy use of spreadsheet software. I will use Microsoft Excel for all demonstrations. As Berkeley students, you should have access to Excel (and the rest of the Microsoft Office suite) without additional charge. I recommend you have Excel installed on your devices if you do not already, as that will make following the demonstrations and doing the problem set exercises smoother.

However, all spreadsheet apps basically work the same way, and while Excel certainly has proprietary features, we will stay away from them as much as possible. Therefore, if you prefer alternative spreadsheets apps like Google Sheet or Apple Numbers, those *should* work. I do not promise there will be no glitches, however.

Where alternative apps will most vary from Excel is in the creation of data visualizations. Data visualization will be a small (but important) part of the course. Google Sheet and Apple Numbers should be able to produce basic data graphs like column/bar charts, histograms, and scatterplots. The user interface for doing so however will likely be quite different than in Excel, and if you do go with Sheet or Numbers, it will principally be left up to you to figure out the interface.

I recommend that you bring your laptop or tablet devices with you to our class meetings, and follow my demonstrations. Where possible/appropriate, I will make all spreadsheet files I operate on available at class time, if not sooner.

Course grading

Your course grade will consist of three components:

- **8 problem sets** (8% each, 64% total), due roughly weekly (precise schedule below). Problem sets will partly cover material related to the specific topic we are considering that week. But some of the problems will refer back to older topics, to reinforce learning of previous material.

  The problem sets will all be submitted online.

- **an in-class midterm exam** (16% of the course grade), a little before the halfway point of the semester.

  The midterm exam, unlike the problem sets, will be submitted on paper.

- **a final exam** (20% of the course grade). This will be administed in-class, during the University's official timeslot for our course: Monday, December 16 2024, 8-11 AM. The morning timeslot was not my idea.

  Like the midterm, the final exam will be submitted on paper.

For both the in-class midterm and in-class final exam, you will be permitted to use a dedicated **calculator device** to aid in arithmetic computations. For fairness purposes, you may NOT use a calculator app on a general-purpose computing device like a smartphone, tablet, laptop, Google Glasses, etc.

The course grading scale is as follows:

A+	97+	A	93-96	A-	90-92
B+	87-89	B	83-86	B-	80-82
C+	77-79	C	73-76	C-	70-72
D+	67-69	D	63-66	D-	60-62
		F	0-59		

When it comes time to compute overall course grades, I will round to the nearest whole number using standard rounding conventions—89.49 rounds down to 89 and a B+, 89.50 rounds up to 90 and an A-, etc. It doesn't really matter what the letter grade on the individual assignments are.

There are no other discretionary considerations, nor opportunities to earn extra credit on an individual basis, that will factor into your grade.

There are no surprises in how I calculate course grades. The above weightings will be programmed into bCourses so that what you see online will accurately inform you of your course standing. During the semester, with a little arithmetic, you can figure out how well you need to do on subsequent assignments in order to earn a particular desired grade.

Late work policy: All graded work submitted late will be marked down **20%** per 24-hour period it is overdue.

The graded assignments this semester will all be submitted over bCourses. If the wrong file is submitted, or the submitted file cannot be read, you will be charged a late penalty whether or not the mistake was accidental.

Disabled Student Program (DSP)

DSP students should have their arrangements made as soon as possible. The DSP office will automatically send me a digital copy of your letter, with an explanation of your

accommodations. It will not be necessary to also provide me with a physical copy of the letter.

If your accommodations are specifically for in-class assignments, they will not be automatically applied to take-home assignments (like the problem sets) as well.

For the in-class exams (the midterm and the final), I will proactively contact those with DSP accommodations ahead of time, to make arrangements.

Recommendation Letters

Writing recommendation letters is part of an instructor's job and one which I embrace readily. However, it is difficult for me, and unhelpful to all parties involved, to write letters for students I do not know very well. I can and have exercised the discretion to decline to write letters when I don't know students beyond what grades they've earned in the course. If you anticipate asking me to write a letter, it is in your interest to participate in class discussion, visit me during my office hours, etc., and let me know what your plans and interests are above and beyond course material. (And this goes for any professor you're contemplating asking for a letter!)

Academic Honesty & Generative AI

While I encourage collaborative learning efforts organized amongst yourselves (e.g., study groups), these should be suspended while the take-home midterm exams are on. Similarly, while you are free and even encouraged to discuss the readings, the reading response papers themselves should be your own work.

Plaigiarism—passing off the work of someone else, or of AI engines like ChatGPT, as your own—will result in an immediate **zero** on the assignment. Depending on the nature of the infraction, I may also be compelled to report the plaigiarism to the appropriate authorities at the Division of Student Affairs. If you are uncertain about what does and does not count as plaigiarism, you should consult with me.

Use of generative AI (e.g., ChatGPT) is RESTRICTED.*

*Guidelines and language adapted from Professor Heather Haveman in the Department of Sociology, UC Berkeley.

Generative AI has earned a lot of attention since ChatGPT was released. You may *not* use any generative AI app or tool to complete any of the graded assignments for this course. If I suspect the use of generative AI or other such assistance on an assignment—answers without any shown work on problem sets are a particular red flag—I may simply assign a **zero** to the assignment in addition to taking other measures.

That said, there are a wide range of uses of AI tools by students, such as:

- correcting grammar
    - Though, note that you will not be graded for spelling or grammar *per se* on any of the course assignments, unless they are so incoherent that I cannot understand your argument.
- explaining unfamiliar terms found in readings
- generating practice quiz questions

I encourage you to explore these tools, at least the free versions of LLMs. They are becoming more prominent in the "real world" outside academia, so you will undoubtedly encounter them wherever you work after graduation.

Reading, Assignment, and Exam Schedule

All readings are available in PDF format on the course website. Each should be completed prior to the first meeting on schedule associated with it.

Setting the stage			
Introduction	Aug 28	W	*No readings*.
	Aug 30	F	
	Sep 2	M	Labor Day: no class.
Basic spreadsheet operations	Sep 4	W	*No readings*.
	Sep 6	F	
	Sep 9	M	

Descriptive Statistics			
Types of data	Sep 11	W	*Read:* Agresti §2.1
Measures of central tendency	Sep 13	F	*Read:* • Wheelan ch 2 pp. 15-23 • Agresti §3.2
	Sep 16	M	
	Sep 18	W	Problem Set #1 due on bCourses
Measures of dispersion	Sep 20	F	*Read:* • Wheelan ch 2 pp. 23-35 • Agresti §3.3-§3.4
	Sep 23	M	
	Sep 25	W	Problem Set #2 due on bCourses
Association	Sep 27	F	*Read:* • Wheelan ch 4 • Agresti §3.5, §10.1-§10.3
	Sep 30	M	
	Oct 2	W	
	Oct 4	F	Problem Set #3 due on bCourses
Midterm exam, in-class	Oct 7	M	

Introduction to Probability			
Introduction to probability, and why we should care about it	Oct 9	W	*Read:* Wheelan ch 5
	Oct 11	F	
	Oct 14	M	
Samples and populations	Oct 16	W	*No readings*. Problem Set #4 due on bCourses

Distributions and the Central Limit Theorem	Oct 18	F	*Read:* • Wheelan ch 8 • Agresti §4.3-§4.5
	Oct 21	M	
	Oct 23	W	
	Oct 25	F	Problem Set #5 due on bCourses

Inferential Statistics			
Hypothesis testing	Oct 28	M	*Read:* • Wheelan ch 9 • Agresti §6.1-§6.6
	Oct 30	W	
	Nov 1	F	
	Nov 4	M	Problem Set #6 due on bCourses
Confidence intervals	Nov 6	W	*Read:* • Wheelan ch 10 • Agresti §5.1-§5.4
	Nov 8	F	
	Nov 11	M	Veterans Day: no class
	Nov 13	W	
	Nov 15	F	Problem Set #7 due on bCourses
Comparing two groups	Nov 18	M	*Read:* Agresti §7.1-§7.3
Association between categorical variables	Nov 20	W	*Read:* Agresti §8.1-§8.4
	Nov 22	F	
	Nov 25	M	Problem Set #8 due on bCourses
	Nov 27	W	Thanksgiving: no class
	Nov 29	F	Thanksgiving: no class
Regression	Dec 2	M	*Read:* • Wheelan ch 11 • Agresti §9.1-§9.5
	Dec 4	W	
	Dec 6	F	

Reading, Review & Recitation week	Dec 9	M	No class during RRR week.
	Dec 11	W	
	Dec 13	F	

Final exam, in-class	Dec 16	M	We are in the 8-11 AM timeslot.