Soc 7 The Power of Numbers: Quantitative Data in the Social Sciences
Fall 2023 | UC Berkeley

Instructor: Professor Linus Huang, Lecturer
Office Hours:
- *In-person, strictly drop-in:* Mondays, 1:15-3:00 PM, 487 Social Sciences Building
- *Zoom, sign-up needed:* Tuesdays, 1:00-3:00 PM, sign up on Google Calendar

E-mail: lbhuang@berkeley.edu
Final Exam: Monday, December 11th, 8-11 AM

What will this course be about?

Even if you are not going to become a quantitative academic or other researcher, it is important to understand numbers and statistics. We live in a world of numbers: GPA, credit scores, *U.S. News & World Report* college indices, Amazon customer ratings, chance of rain tomorrow. Answers to important questions are offered based on numerical analysis: is a Covid vaccine effective? are humans the cause of global warming? should you go to college or not? should you drive, or take the BART, to get to an event in San Francisco? should you buy or rent your next home?

To be an informed citizen in today's world, whether you are a STEM major or not, and even if you are a college graduate or not, it is important to have (1) a degree of numerical literacy (or "numeracy"), and (2) a basic understanding of the science of statistics. The two are not the exact same thing. Working toward both are the key objectives of the course. This course by itself will not provide all you will ever need to know about numeracy and statistics. But, it will help you start down this path!

One of the problems students encounter learning statistics is that it appears as an arcane language. Statistics comes off not as a way to help us understand the social world, which is in fact what it is, but as an esoteric body of knowledge whose main purpose seems to be to show that you know it, and others don't. Every scientific discipline has its jargon and concepts that seem alien to the unitiated—this goes for qualitative scientific disciplines as well as quantitative ones. I will not claim to offer a magical way of learning statistics that avoids mathematical technicalities. Yes, there will be math! But, compared to other introductory statistics courses, Sociology 7 will take the position that statistics are a set of tools whose value is that they help us understand the world, not something to be learned for the sake of impressing folks back home.

Toward that end, both the presentation of material and the quasi-weekly problem sets will emphasize the application of statistical concepts to practical, "real world" sociological problems, and not just the standard mathematical quiz-type questions (though there will be those, too). We will take much more time understanding the "why" behind conventions in statistical analysis than in other introductory courses.

Sociology 7, then, is about dealing with quantitative data—finding it; handling it; analyzing and interpreting it; and reporting it, including in visual form. By the end of the semester, you will be able to understand, evaluate, use, and produce quantitative data about the social world by:

- critiquing and producing basic graphs
- manipulating and analyzing data in spreadsheets
- calculating and explaining basic statistical measures of central tendency, variation, and association
- applying and explaining basic concepts of sampling and selection
- thinking critically about reported statistics and quantitative social science more broadly.

Readings and Resources

All course readings are available in PDF format on the bCourses site. There are no books or paper readers to purchase.

All indicated readings should be completed *prior* to the first class we begin discussing them.

Important: Optimally, you should have Microsoft Excel for Windows or for the Mac. UC Berkeley students should have access to Excel and the rest of Microsoft Office by virtue of student technology fees you've paid. Throughout the course but especially near the beginning, you should bring your PC to class and follow along while I demonstrate Excel—and later, a handful of assorted web apps—on the screen.

It *may* be possible to get by using Google Sheets, the Google Docs substitute for Excel. The basics of spreadsheet manipulation and calculation should be 100% functionally identical between Sheets and Excel. Where there may be differences that are relevant to us, however, is in producing data visualizations.

Instruction Mode

Sociology 7 will be in-person only for the duration of the semester. There will be no recordings of classes available.

Grading

Your course grade will be determined by the following components:

- **8 problem sets** (6% each, 48% total), due roughly weekly. Problem sets will partly cover material related to the statistical analysis topic we are considering at the moment. But others 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** (26% of the course grade) at about the one-third point in the course.

  The midterm exam, unlike the problem sets, will be submitted on paper. Likely this will be an exam booklet. But precise logistics will be announced as the time approaches.

- **a final exam** (26% of the course grade). This will be administered in-class during the official final exam time slot for the course, which is Monday, December 11th, 2023, 8-11 AM. The morning timeslot was not my idea.

  Like the midterm exam, the final exam will be submitted on paper. Further details will be announced as we approach the final.

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. It doesn't really matter what the letter grade on the individual assignments are.

There are no surprises in how I calculate course grades. The GRADES section on bCourses will incorporate the weightings above and will accurately keep you apprised of your course progress.

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

Academic Honesty

The UC Berkeley Honor Code states that "As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others" (https://teaching.berkeley.edu/berkeley-honor-code). I expect you will follow these principles. You may not copy specific text or ideas from others, whether from fellow students, from authors of our readings or other material you find, without specific attribution. To do otherwise is to plagiarize. You may not cheat on any of the homework assignments or exams by bringing in illicit outside material, copying from fellow students, or engaging in other dishonest practices. Violation of these rules will result in an immediate **-0-** on the *entire* assignment in question, plus a report to the Office of Academic Affairs at my discretion.

==You are __NOT__ allowed to use ChatGPT for Sociology 7==

Many—but not all—of the questions on the problem sets this semester will be of the typical "textbook" variety; they will require application of statistical concepts and will be a test of your understanding of them. It is ESSENTIAL that you not simply feed them to ChatGPT or other large-language model/generative AI bot and mindlessly paste in the bot's results. First, this is academic **dis**honesty for the purposes of the course. Second, there would be no point to taking the course if you do so.

Working through these problem sets is not a hazing ritual. It is to develop numerical literacy or "numeracy", one of the objectives of the course. If I am doing my job in lecture, these problems will not seem so insurmountable. But encountering difficulty with the material is also okay! You need not resort to cheating to try to get by. Although you will be required to rely on your own resources for the midterm and final exams, I am **happy** ☺ to assist you on the problem sets. That is in fact one of the reasons this course's enrollment limit is being kept low—to allow more interaction between you and me. … Don't look for shortcuts. Come see me instead!

Bottom line: You are not simply discouraged from using ChatGPT on this semester's assignments. You are prohibited from doing so.

Further note: The problem sets this semester will not solely be an exercise in restraining yourself from using AI. Although the "textbook"-variety questions are designed to (mostly) have a right or wrong answer, many of the questions will deliberately embody the practical uncertainties of "real-world" data analysis. This does not mean that any answer will be as good as any other; instead, how you communicate your reasoning and judgment will figure into your score on such questions. It would probably be over-reaching to say that generative AI like ChatGPT is useless for these problems. But it is likely to be of less help, particularly without accurate prompting.

Reading, Assignment, and Exam Schedule

All readings for this course are available on bCourses in the `READINGS` folder of the `FILES` section. Readings associated with a date are to be completed prior to the class meeting on that day.

Introduction			
Introduction: why study numbers?	Aug 23	W	*No readings.*
	Aug 25	F	
	Aug 28	M	
Basic spreadsheet operations	Aug 30	W	*No readings.*
	Sep 1	F	
	Sep 4	M	Labor Day: No class.
	Sep 6	W	

Descriptive Statistics			
Types of data	Sep 8	F	*Read:* Agresti §2.1
Measures of centrality	Sep 11	M	*Read:* • Wheelan ch 2 pp. 15-23 • Agresti §3.2
	Sep 13	W	Problem Set #1 due on bCourses

Measures of dispersion	Sep 15	F	*Read:*
	Sep 18	M	• Wheelan ch 2 pp. 23-35 • Agresti §3.3-§3.4
	Sep 20	W	Problem Set #2 due on bCourses
Association	Sep 22	F	*Read:*
	Sep 25	M	• Wheelan ch 4 • Agresti §3.5, §10.1-§10.3
	Sep 27	W	Problem Set #3 due on bCourses
	Sep 29	F	
Lying with data visualizations	Oct 2	M	*No readings.*
Midterm exam, in-class	Oct 4	W	

Introduction to Probability

Introduction to probability, and why we should care about it	Oct 6	F	*Read:* Wheelan ch 5
	Oct 9	M	
	Oct 11	W	
Samples & populations	Oct 13	F	*No readings.*
Distributions & the Central Limit Theorem	Oct 16	M	*Read:* • Wheelan ch 8 • Agresti §4.3-§4.5 Problem Set #4 due on bCourses
	Oct 18	W	
	Oct 20	F	
	Oct 23	M	Problem Set #5 due on bCourses

Inferential Statistics

Hypothesis testing	Oct 25	W	*Read:*
	Oct 27	F	• Wheelan ch 9
	Oct 30	M	• Agresti §6.1-§6.6
	Nov 1	W	Problem Set #6 due on bCourses
Confidence intervals	Nov 3	F	*Read:*
	Nov 6	M	• Wheelan ch 10 • Agresti §5.1-§5.4
	Nov 8	W	Problem Set #7 due on bCourses
	Nov 10	F	Veterans Day: No class
	Nov 13	M	
Association between categorical variables	Nov 15	W	*Read:* Agresti §8.1-§8.4
	Nov 17	F	
	Nov 20	M	Problem Set #8 due on bCourses

	Nov 22	W	Thanksgiving: No class
	Nov 24	F	Thanksgiving: No class
Regression	Nov 27	M	*Read:*
	Nov 29	W	• Wheelan ch 11
	Dec 1	F	• Agresti §9.1-§9.5

Reading, Review & Recitation week	Dec 4	M	No class during RRR week. There will be an optional review session on the Wednesday of this week at our normal time and place.
	Dec 6	W	
	Dec 8	F	

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