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

Fall 2018 | UC Berkeley

Instructor: Professor Linus Huang
Office Hours (drop-in): Wednesdays, 12:15-2:15 PM, 487 Barrows
E-mail: lbhuang@berkeley.edu
Final Exam: Monday, December 10th, 8-11 AM

Course Overview[1]

Numbers abound at all levels of our everyday lives. Some numbers tell us how the economy is doing overall, who is likely to win an upcoming election, or how many people attended a large gathering like an inauguration ceremony. Other numbers tell us who qualifies for a low-interest loan to go to college or to buy a home. Numbers tell us whether it's likely to rain tomorrow, or how long we might expect to sit in traffic if we go out. Certain numbers tell college applicants which schools are better than others. Other numbers tell schools which applicants are better than others. And once a college-bound person enrolls in a school, yet other numbers will represent the quality of that student's academic performance.

Despite their ubiquity, however, numbers are not always understood. Some seem so transparent that we don't question them. Others are so hopelessly complex that we don't even try to understand them. Many of the numbers that are widely accepted as common knowledge are not even right. Yet, since numbers have the appearance of precision, they continue to influence the way we understand the world. This is part of their double-edged power. As citizens, professionals, social activists, and civic leaders, we need to develop the numerical literacy to recognize bad numbers and either demand, or produce ourselves, better numbers.

This course will introduce you to the basic concepts central to quantitative sociology and equip you to become more savvy consumers of numbers in society today. It seeks to give you an intuitive overview of quantitative tools employed by social scientists and hands-on opportunities to use these tools to examine the world. It is intended specifically for social science majors, and focuses on social science questions. There will be math! However, you do not necessarily need a strong mathematical, statistical, or computing background to succeed in this course. What you *do* need is curiousity about how numbers "work" (what they do and do not say, why some numbers are misleading, etc.) and a desire to try something new. My aim is to show you that quantitative science is useful, and is something that you can do.

To put our lessons into practice, we will use data from the General Social Survey (GSS), *the* premier dataset for assessing Americans' behavior and attitudes on social issues. The GSS is administered by the National Opinion Research Center at the University of Chicago, first in 1972 and almost annually through 1994; it has been administered biennially since then. An enormous amount of data has been collected for the GSS, but fortunately a program developed right here at Berkeley, the Survey Documentation and Analysis (SDA) program (http://sda.berkeley.edu/),

[1] This course overview, and much of the general plan of the course overall, draws heavily from previous iterations of the course offered by Sara Lopus, Michael Schultz, and Mao-Mei Liu.

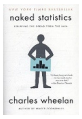
offers a Web-based interface gives us convenient access to the cumulative dataset. SDA gives us a way not only to assess Americans' behavior and attitudes today, but how they have changed over the last nearly half-century. We will use SDA's interface to the GSS throughout the course.

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

Required Readings and Resources

There is one required text for this course, Charles Wheelan's *Naked Statistics: Stripping the Dread from the Data* (2013). It is available at the ASUC Store, or of course wherever savvy consumers today can go to buy books. I will also place it on 1-day reserve at Moffitt Library.



All other readings for this course are in PDF format online in the `FILES` section of bCourses.

Important: **you must have a laptop with internet access and Microsoft Excel or equivalent to take this course**, as you will need it to participate in class. If you want to take this course and do not have a laptop, see me (Professor Huang) immediatedly!

Note that all UC Berkeley students are entitled to free use of Microsoft Office (of which Excel is one component of the "productivity suite") on their personal machines. If you do not already have Office/Excel, go to https://software.berkeley.edu/productivity-software#Microsoft.

Assignments / Grading

Your course grade will consist of three different components: 1) ten individual homework assignments; 2) an in-class midterm exam; and 3) an in-class final exam.

1) Individual homework

There will be ten individual homework assignments designed to accompany current course topics. They will for the most part be due weekly, on Wednesdays by class time (10:00 AM). (There will be no assignments at the very beginning of the course, during the week of the midterm exam, and during the last week of the course.)

Check on bCourses for the exact due dates for each of the ten homework assignments.

Each of the ten homework assignments will be worth 6% of the course grade, or 60% for all ten of them.

2) In-class midterm exam

The midterm exam is scheduled for Friday, October 5th. We will use the Wednesday prior to it as a review session.

The midterm exam will be worth 20% of your course grade.

3) In-class final exam

The final exam will be held during the University's official timeslot for our course, which is on Monday, December 10th, from 8-11 AM. It will be cumulative.

The final exam will be worth 20% of your course grade.

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 incorporates the weightings above and will accurately keep you apprised of your course progress.

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.

There may be multiple opportunities for (primarily in-class) collaborative work in this course. While working in groups is a pedagogical tool and helps us prepare for work beyond this class,

knowing what is acceptable collaboration and what is taking unfair advantage of others can be difficult. If at any point you have any questions about how the honor code applies, or how best to fulfill your obligations as a member of the UC Berkeley community, please ask me.

Reading Schedule

Readings for this course will be either from the required Wheelan text, or in PDF format in the **Readings** folder of the **Files** section on bCourses.

When reading assignments are indicated for a particular date, they should be completed prior to the class on that date.

Part I: Introduction		**Reading**
Aug 22 & 24	Introduction / Reading graphs	(Aug. 24th) Wheelan, ch 1
Aug 27 & 29	Getting to know SDA and the GSS	*No readings.*
Aug 31 & Sep 5[2]	Basic spreadsheet operations	*No readings.*
Sep 7	Types of data	Agresti 2.1 (PDF)
Part II: Analyzing Data		
Sep 10 & 12	Measures of centrality	• Wheelan ch 2 (pp. 15-23) • Agresti 3.2 (PDF)
Sep 14 & 17	Measures of dispersion	• Wheelan ch 2 (pp. 23-35) • Agresti 3.3-3.4 (PDF)
Sep 19, 21, 24	Association	• Wheelan ch 4 • Agresti 3.5
Sep 26	Assessing change over time in data: period, cohort, and age	*No readings.*
Sep 28 & Oct 1	Exploring data visually	**Readings TBD**
Oct 3	Midterm review	
Oct 5	**Midterm exam, in-class**	
Part III: Statistical Inference		
Oct 8 & 10	Introduction to probability	• Wheelan ch 5, 5 ½ • Nate Silver, *The Signal and the Noise* ch 2 (PDF)
Oct 12 & 15	Distributions and the Central Limit Theorem	• Wheelan ch 8 • Agresti 4.3 (PDF)
Oct 17 & 19	Samples and populations	Wheelan ch 6, 7
Oct 22 & 24	Confidence intervals	• Wheelan ch 10 • Agresti 5.1-5.4 (PDF)
Oct 26 & 29	Hypothesis testing	• Wheelan ch 9 • Agresti 6.1-6.6 (PDF)
Oct 31 & Nov 2	Association between categorical variables	Agresti 8.1-8.4 (PDF)
Nov 5 & 7	Comparing two groups	Agresti 7.1-7.3 (PDF)

Commented [LH1]: Assign ch 4.5 in addition to 4.3 next time.

[2] No class on Labor Day, Monday, September 3rd.

Nov 9 & 14[3]	Multivariate analysis	Agresti 10.1-10.3 (PDF)
Nov 16 & 19	Regression	• Wheelan ch 11 • Agresti 9.1-9.5 (PDF) • Agresti 12.1 (PDF) • Agresti 15.1-15.5 (PDF)
Part IV: Social Issues with Numbers		
Nov 26	Lying with data graphics	*No readings.*
Nov 28 & 30	"Big Data" and privacy issues	Stephens-Davidowitz, *Everybody Lies* excerpts TBD
Dec 3, 5, 7	**Reading, Recitation & Review—no class**	
Dec 10	**Final Exam, in-class, 8-11 AM**	

[3] No class on Veterans' Day, Monday, November 12th.