

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

Spring 2018 | UC Berkeley

**Instructor:** Linus Huang

**Office Hours (drop-in):** Wednesdays, 3:15 PM - 4:45 PM, 487 Barrows

**E-mail:** [lbhuang@berkeley.edu](mailto:lbhuang@berkeley.edu)

**Final exam:** Exam Group 7, Tuesday, May 8, 2018 3-6 PM

### Course Overview<sup>1</sup>

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 buy a home. Numbers tell us whether it's likely to rain tomorrow, or how long we might be expecting 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 determine 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 and critical consumers of social science research. 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. *You do NOT need a strong mathematical, statistical, or computing background to succeed in this course.* What you do need is curiosity about how is organized, a desire to try something new, and a collaborative and constructive attitude. Our aim is to show you that quantitative science is useful, can be fun, and is something that you can do.

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
- accessing relevant, high-quality data and relevant sociological research

---

<sup>1</sup> 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.

- manipulating and analyzing data in spreadsheets
- calculating and explaining basic statistical measures of central tendency, variation, and correlation
- 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 have also placed 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.

Also, **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) immediately!

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) six individual homework assignments; 2) two exams; and 3) projects.

### 1) Individual homework

There will be 6 individual homework assignments designed to accompany the readings. They are generally due weekly, on Wednesdays by class time (10:00 AM). Ideally, there would be assignments every single week, but there will be no homework the week of the midterm and no homework during weeks where a project (see below) is due.

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

### 2) Exams

There will be two exams: an in-class midterm exam, and an in-class final exam.

### 3) Projects<sup>2</sup>

This semester, in addition to the homework and in the exams, we will apply the tools we'll have developed to conduct a research project using your choice of one of three major datasets:

- **Gapminder** ([www.gapminder.org](http://www.gapminder.org)), a country-level dataset designed to promote an understanding of “basic global facts”.
- **CalEnviroScreen 3.0** (<https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30>), a regional-level (at the level of a census tract) dataset designed to promote a broader understanding of the public health effects of environmental pollution.
- The **General Social Survey**, the most comprehensive survey available of Americans' attitudes on social issues. Collected since 1972, access to its findings has been made immensely more convenient thanks to the Survey Documentation and Analysis (<http://sda.berkeley.edu/>) tool developed right here at UC Berkeley.

Each of these three datasets will be introduced in class early in the semester. Sometime soon following this introduction, students will decide which of the three datasets they're most interested in conducting research with. Different students may select different datasets. Based on these individual preferences, we will form groups of ideally 3-4 students, interested in the same datasets, which will remain together for the duration of the course.

Either individually or in these groups, students will undertake four projects. First, as a group, you will explore and describe the dataset, generate descriptive statistics for key variables, and identify some outcomes of interest. Second, individually, students will read a sociological study of the phenomenon and write a short report. Third, as a group, you will revise the first project, use prior research to develop testable hypotheses, test these hypotheses, and then report the results. Fourth, as a group, you will produce an effective, polished presentation to share their project with the entire class.

Group work can be very productive and rewarding, but—as you're probably aware—can also be tricky to manage. I will do my best to facilitate this process, in part by establishing clear goals and giving significant time in class for groups to meet. Additionally, groups can help themselves by identifying a clear division of labor. Each group member should keep notes on their progress, contributions, and difficulties—and include these where appropriate in their project submissions. Students in well-functioning groups are likely to receive the same (high) grade. But it is possible for students in a poorly-functioning group to receive different grades even if there is only one, group, submission. Please notify me as early as possible as problems arise that you and your group are unable to handle. I will step in when necessary.

Further details about the course projects are at the end of this syllabus document.

---

<sup>2</sup> The design and implementation of the research projects is nearly entirely drawn from Mao-Mei Liu's plan for the Fall 2017 iteration of Soc 7.

## Assignments, Weightings, and Due Dates

Assignment	% of Grade	Due Date
Homeworks (6)	6 x 5% each = 30%	<i>throughout the semester</i>
Mid-Term Exam	15%	March 14 <sup>th</sup> , in-class
Final Exam	15%	May 8 <sup>th</sup> , 3-6 PM, in-class
Project 1: Exploring & Describing the Data	8%	February 26 <sup>th</sup> , 10:00 AM
Project 2: Diving into the Literature	8%	March 23 <sup>rd</sup> , 10:00 AM, on bCourses
Project 3: Hypotheses & Beyond	12%	April 13 <sup>th</sup> , 10:00 AM
Project 4: Group Presentation	12%	April 23 <sup>rd</sup> , 25 <sup>th</sup> & 27 <sup>th</sup> , in-class

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 will be a significant amount of 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

In addition to Wheelan's *Naked Statistics*, readings will come from the following:

- Alan Agresti (2017), *Statistical Methods for the Social Sciences* (5<sup>th</sup> ed.). Pearson.
- Joel Best (2001), *Damned Lies and Statistics: Untangling Numbers from the Media Politicians and Activists*. University of California Press.
- Darrell Huff (1954), *How to Lie with Statistics*. W.W. Norton & Company.
- Nate Silver (2012), *The Signal and the Noise: Why Most Predictions Fail—but Some Don't*. Penguin Group.
- Edward R. Tufte (2001), *The Visual Display of Quantitative Information* (2<sup>nd</sup> ed.). Graphics Press.

Readings from the above five are in PDF format in the FILES section on bCourses.

Overall, there 40 class meetings and 18 topics, depending on how you count them—meaning that on average we spend about two class meetings per topic. Below is the schedule for those 18 topics, and the associated readings. Except where otherwise indicated, whenever a reading is listed, it is due by the first day scheduled for the topic, even if we spend multiple days on the topic overall.

<b>Part I: Introduction</b>		<b>Reading</b>
Jan. 17 & 19	Introduction / Reading Graphs	(Jan. 19 <sup>th</sup> ) Wheelan, ch 1
Jan. 22	Units of Analysis	<i>No readings.</i>
Jan. 24 & 26	Basic Spreadsheet Operations	<i>No readings.</i>
<b>Part II: Properties of Data</b>		
Jan. 29	Types of Data	Agresti 2.1 (PDF)
Jan. 31	What are the scales of things?	Best, “The worst social statistic ever” + “The public as innumerate audience” (PDF)
Feb. 2 & 5	Measures of centrality	<ul style="list-style-type: none"> <li>• Wheelan ch 2 (pp. 15-23)</li> <li>• Agresti 3.2 (PDF)</li> </ul>
Feb. 7 & 9	Measures of dispersion	<ul style="list-style-type: none"> <li>• Wheelan ch 2 (pp. 23-35)</li> <li>• Agresti 3.3-3.4 (PDF)</li> </ul>
Feb. 12 & 14	Association	<ul style="list-style-type: none"> <li>• Wheelan ch 4</li> <li>• Agresti 3.5 (PDF)</li> <li>• Agresti 9.4, pp. 259-262 (PDF)</li> </ul>
Feb. 16, 21, 23	Introduction to probability	<ul style="list-style-type: none"> <li>• Wheelan ch 5, 5 ½</li> <li>• Silver ch 2 (PDF)</li> </ul>
Feb. 26 & 28	Distributions & the Central Limit Theorem	<ul style="list-style-type: none"> <li>• Wheelan ch 8</li> <li>• Agresti 4.3 (PDF)</li> </ul>

<b>Part III: Statistical Inference</b>		
Mar. 2, 5, 7	Samples and populations	Wheelan ch 6, 7
Mar. 9	Selection bias	<i>No readings.</i>
Mar. 12	Midterm review	
Mar. 14	Midterm Exam, in-class	
Mar. 16 & 19	Hypothesis Testing	<ul style="list-style-type: none"> <li>• Wheelan ch 9</li> <li>• Agresti 6.1-6.6 (PDF)</li> </ul>
Mar. 21 & 23	Estimation	<ul style="list-style-type: none"> <li>• Wheelan ch 10</li> <li>• Agresti 5.1-5.4 (PDF)</li> </ul>
Mar. 26, 28, 30	<b>SPRING BREAK</b>	
Apr. 2	Midterm Retrospective	
Apr. 4 & 6	Regression	Wheelan ch 10, 11
<b>Part IV: Interpreting &amp; Representing Data</b>		
Apr. 9 & 11	Representing & visualizing data	Tufte <i>excerpts TBD</i>
Apr. 13 & 16	Data sources	<ul style="list-style-type: none"> <li>• Huff ch 10 (PDF)</li> <li>• Silver ch 12 (PDF)</li> </ul>
Apr. 18 & 20	“Big Data” / Data science	<i>Readings TBD</i>
Apr. 23, 25, 27	Group presentations, in-class	
Apr. 30, May 2 & 4	<b>Reading, Recitation &amp; Review—no class</b>	
May 8	Final Exam, in-class, 3-6 PM	

## Projects<sup>3</sup>

### Project 1: Exploring & Describing the Data

Due Feb. 26<sup>th</sup>. 8% of the course grade. **Group** project.

We will work in small groups to conduct a research project using one of the Gapminder, CalEnviroScreen, or General Social Survey datasets. Project 1 involves exploring and describing the dataset and generating descriptive statistics.

Working with other members of your group, combine your notes and write a 3-5 page report (5 pages maximum) explaining the following:

<sup>3</sup> The design and implementation of the projects for Soc 7 is nearly entirely drawn from Mao-Mei Liu’s presentation of the course during the Fall 2017 semester at UC Berkeley. All imperfections in execution are my (Dr. Linus’) own.

- Introduce data and variables. Describe the dataset. How was the data collected? Which population does it represent? Which outcomes (name at least 2) are you most interested in examining and understanding? Which variables best capture these? Choose 2-5 other variables of interest. What do these variables measure? How were they collected?
- Provide descriptive statistics of variables and relationships. Provide appropriate descriptive statistics (centrality and dispersion) for each variable. Which variables do you predict to be related to one another? For at least three pairs of variables, describe their association.
- Discuss. Explain what the data and your analyses suggest. What are the take-aways? What can we learn about this phenomenon using this data?

### **Project 2: Diving into the Literature**

Due Mar. 23<sup>rd</sup>. 8% of the course grade. **Individual** project.

Start by identifying the outcome you want to explore this semester. Find 1 influential sociology article that examines this. Describe the data using language we learned about in class. Then, identify 1 dependent variable and 1 independent variable. Make sure that these variables are predicted to be related to each other. Justify your choices of variables by copying and pasting (or typing) one or more short excerpts. Explain in your own words: (1) why the author expects that the independent and dependent variables are related and (2) how the independent variable is actually related to the dependent variable. This project should be 2 pages maximum.

### **Project 3: Hypotheses & Beyond**

Due Apr 13<sup>th</sup>. 12% of the course grade. **Group** project.

This is your opportunity to demonstrate your knowledge, show off your skills and pursue your interests! Use prior research to develop testable hypotheses, test these hypotheses, report and discuss these results. All decisions should be justified in writing. Consider projects 1 and 2 feedback as you write this project. Choose and develop the best representation for your results, both in table and figures form. Write a 10-15 page report (15 pages maximum not including references, figures/tables and appendix). A suggested outline for Project 3:

- Intro
- Theoretical framework - Summarize and discuss the theoretical and empirical knowledge that previous studies provide. Based on these, develop 2-3 hypotheses. Explain and justify why they are interesting hypotheses. Relate to existing sociology studies (be sure to reference!).
- Data & Variables - Introduce data and variables (dependent, independent, etc.). Provide descriptive statistics of variables (same instructions as in Project 1 + for each point estimate, provide a confidence interval)
- Results & Analysis - Test each hypothesis with data and the appropriate tests (i.e. means, proportions, associations - each paper should include and test one hypothesis about an association). Provide a full write-up, including statistics and confidence intervals (where necessary). Determine whether you should reject the hypothesis or not. Explain the reasoning and conclusion in words. Connect back to your theoretical framework
- Conclusion

- References
- Figures and Tables
- Appendix - full 5-step write-up of hypothesis testing, spreadsheets

**Project 4: Presentations**

Last week of instruction: Apr. 23<sup>rd</sup>, 25<sup>th</sup>, 27<sup>th</sup>. 12% of the course grade. **Group** project.

Using powerpoint or a short movie, present a 5-8 minute interesting, informative and effective audio/visual overview of your group project (Intro to real sociology; data and variables; Descriptive Statistics; Hypotheses & Inference). Include a reflection of the problems and difficulties encountered and how the group solved them.