

## Sociology 7

# The Power of Numbers: Introduction to Quantitative Social Science

Instructor: Michael Schultz  
Email: michael.schultz@berkeley.edu  
Time & Location: TuTh 9:30-11:00am [475 Barrows]  
Office Hours: Th 11:00am-12:00pm [499 Barrows] and by appointment

## Overview

Social statistics have become more widespread and more visible in recent years. From Nate Silver's quantitatively-driven prediction of the 2008 and 2012 elections, to the book *Moneyball*, to the sometimes euphoric claims made for "big data," numbers are now everywhere in the public sphere. But despite their ubiquity, these numbers are not always well understood. Some statistics seem so transparent that we do not think much about them. Others seem so opaque that we give up. Many of the numbers that circulate as common knowledge are not even right. Yet, they have the appearance of precision, and a certain social power, and so they stay on. Wrong numbers can have important social consequences. As citizens, professionals, social activists, and civic leaders, we need to recognize bad statistics and produce better ones.

There are many facets of society that can be effectively understood quantitatively. There are even most facets that can *only* be understood quantitatively. This course will provide students with a set of skills to understand, evaluate, use, and produce quantitative data about the social 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 a basic curiosity about how society is organized and a willingness to try something new. Our aim is to show you that quantitative social science is useful, can be fun, and is something that you can do.

By the end of this course, students will be able to:

- Understand, evaluate, and produce basic graphs
- Find good-quality and relevant
- Manipulate data in a spreadsheet, including producing pivot tables
- Understand and calculate basic statistical measures of central tendency, variation, and correlation
- Understand and apply basic concepts of sampling and selection
- Think critically about reported statistics
- Based on all of these more concrete skills: **Begin to think about social science questions in a quantitative way**

## Books and resources

***You must have a laptop with access to the internet to take this course.*** If you want to take this course and do not have a laptop, see the instructor immediately. There are two textbooks for this course:

Silver, Nate. (2012) The Signal and the Noise: Why so many predictions fail—but some don't. Penguin Books.

Wheelan, Charles. (2013) Naked Statistics: Stripping the dread from the data. Norton.

We will focus on working through real applications, and we will rely heavily on web-based resources and tools. Links to these tools will be available on bCourses.

## Expectations and evaluation

Classroom time will be a combination of lecture, discussion, and practical work. Sometimes you will be required to read in advance of class; sometimes you will be required to watch a video lecture or try out a web-based application in advance of class. All of this prep work must be done as listed in the syllabus. If you are not prepared, the class will not be useful for you, and this will be reflected in your participation grade.

### **Group Project**

Over the course of the semester, you will work in groups on a comprehensive, collaborative project using a real dataset. As you learn new topics in class, you will use your dataset to apply the course material. Your group project will build as the sum of in-class activities and some home assignment. At the end of the semester, each group will do a class presentation in which you identify some interesting relationships you have discovered, presented in graphical forms. This project will be an opportunity to familiarize yourself with data, move it around, and work with it in useful ways.

### **Individual homework**

There will be 10 individual homework assignments designed to introduce new topics or reinforce material covered in recent lectures. Grading of these assignments will focus more upon effort than correct answers, since the purpose is often to prepare you for activities performed in the next class. These assignments must be submitted on bCourses by 9:00am the day they are due. Late homework will not be accepted.

### **Exams**

There will be two exams: a midterm and a final exam. These exams will be much more conceptual than computational, and they will focus on your understanding of the core concepts of the course.

## Grading

<b>Group Project</b>	<b>35%</b>
Written group work (2 assignments)	20%
Group presentation	15%
<b>Individual homework</b>	<b>25%</b>
10 assignments	2.5% each
<b>Exams</b>	<b>30%</b>
Midterm Exam	15%
Final Exam	15%
<b>Participation</b>	<b>10%</b>

## Honor Code

The student community at UC Berkeley has adopted the following Honor Code:

*"As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others."*

There will be a good bit of collaborative work in this course. While this kind of collaboration is an important tool in learning and realistic preparation for the contemporary world of work, it also can make it hard to know what is acceptable collaboration and what is taking unfair advantage of others. Throughout the course, we will indicate whether a given assignment is to be completed alone or in cooperation with others. You will avow on each assignment that you complied with those instructions. 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 the instructor. Almost certainly, someone else has the same question.

## Class Schedule

Aug 27 - Introduction

Sep 1 - Spreadsheets

Stark Chapter 5 (Multivariate Data, Scatterplots, Exercises 5-1 to 5-4)

### Unit 1 - Basic Properties of Data

Sep 3 - Units of Analysis

Wheelan Chapter 1 (What's the Point?)

Sep 8 – Scales

**HW 1 Due  
(Interpreting Graphs)**

Best Chapter 1, section "The public as innumerate audience"

Sep 10 - Types of Data

Stark Chapter 3 (Variables; Exercise 3-1) (online)

Sep 15 - Centrality

**HW 2 Due  
(Making Graphs)**

Stark Chapter 4 (Measures of Location; Videos of Exercises; Exercises 4-1 through 4-5) (online)

Wheelan Chapter 2 - "Descriptive Statistics" pp15-23

Sep 17 - Dispersion

Stark Chapter 4 (Spread or Variability; The Range, IQR, and SD; Videos of Exercises; Exercises 4-6 to 4-8) (online)

Wheelan Chapter 2 - pp23-35 "Descriptive Statistics"

Sep 22 – Association

**HW 3 Due  
(Centrality)**

Wheelan Chapter 4: "Correlation"

## Sep 24 - Histograms & Probability

Stark Chapter 13 (Theories of Probability, Random Events, Equally Likely Outcomes, Frequency Theory, Exercises 13-1 to 13-5) (online)  
Silver Chapter 2 ("Are You Smarter than a Television Pundit?")  
Silver Chapter 4 pp134-138 ("How to know if your forecasts are all wet")  
Wheelan Chapter 5: Basic Probability  
Wheelan Chapter 5 ½: The Monty Hall Problem

## Sep 29 - Distributions / Central Limit Theorem

**HW 4 Due**  
**(Variance & Correlation)**

Stark Chapter 23 (The Normal Curve) (online)  
Wheelan Chapter 8 ("The central limit theorem")

## Unit 2 - Statistical Inference

### Oct 1 - Samples and Population

Stark Chapter 24 (Sampling, Parameters and Statistics, Why Sample?, Sampling from Hypothetical Populations, Exercises 24-1 and 24-2) (online)  
Wheelan Chapter 6: Problems with Probability  
Radio Lab - "A very lucky wind" from Stochasticity (online)

### Oct 6 - Random Samples / Counterfactuals

**HW 5 Due**  
**(Histograms)**

Stark Chapter 24 (Simple Random Samples, Systematic Random Samples, Exercise 24-4) (online)  
Wheelan Chapter 7 "The Importance of Data: 'Garbage in, garbage out'" pp110-113

### Oct 8 - Modes of Inference (Bayesian / Frequentist)

Wheelan Chapter 9 ("Inference")  
Stark Chapter 27 (Hypothesis Testing: Does Chance Explain the Results?, Examples of Hypothesis Testing Problems, Significance Level and Power, Test Statistics and P-Values, Exercises 27-1 to 27-3) (online)  
The Data skeptic podcast #4 [p-values] (online)

Oct 13 - Midterm Review

**Oct 15 - Midterm**

Oct 20 - Hypothesis Testing on continuous data

**Group Project #1 Due**

Silver Chapter 8 ("Less and less and less wrong")

The Data Skeptic Podcast: #2 [Type I/Type II errors] (start with minute 2:30) (online)

The Data Skeptic Podcast: #24 [The T-Test] (online)

Oct 22 - Hypothesis testing on categorical data

Stark Chapter 31 (The chi-square statistic, The chi-square test for goodness of fit, Exercises 31-2, 31-4) (online)

The Data Skeptic Podcast: #40 [The Chi-Squared Test] (online)

Oct 27 – Regression

**HW 6 Due**

**(Hypothesis Testing)**

Wheelan Chapter 11 ("Regression Analysis")

Wheelan Chapter 12 ("Common regression mistakes")

**Unit 3 - Interpreting and Representing Data**

Oct 29 - Data sources

Huff Chapter 10 ("How to Talk Back to a Statistic")

Silver Chapter 12 ("A Climate of Healthy Skepticism")

The Data Skeptic Podcast: #36 [Data Provenance] (online)

Nov 3 - Good graphs / sources

**HW 7 Due**

**(Regression)**

Wheelan Chapter 10 "Polling: How we know that 64 percent of Americans support the death penalty (with a sampling error +/- 3 percent)" pp180-183

Best Chapter 1, section "Thinking about statistics as social products"

Nov 5 - Pivot Tables

Steele Chapter 2 ("Identity and Intellectual Performance")  
Excel Pivot Table Tutorial

Nov 10 - Lying with graphs

**HW 8 Due  
(Pivot Tables &  
Types of Graphs)**

Huff "How to lie with statistics" pp60-73  
Wheelan Chapter 3: "Deceptive Description"

Nov 12 - Visualization

Tufte "The Visual Display of Quantitative Information"

## Unit 4 - Social Issues with Data

Nov 17 - Selection Biases

**HW 9  
(Deceptive Graphs &  
Proper Visualization)**

Berk "An introduction to sample selection bias in sociological data" (online)  
Stark Chapter 24 (Bias in Surveys, Exercise 24-3) (online)

Nov 19 - Sampling as a social process

Wheelan Chapter 10 "Polling: How we know that 64 percent of Americans support the death penalty (with a sampling error +/- 3 percent)" pp169-180  
The Data Skeptic Podcast: #21 [Selection Bias] (online)

Nov 24 - Big Data / Data Science

**HW 10  
(Big Data )**

Lewis - Moneyball -Chapter 4 ("Field of ignorance")  
Freedman "Statistical Models and Shoe Leather" (online)  
Silver Chapter 5 ("Desperately seeking signal")  
Wheelan Chapter 7 "The Importance of Data: 'Garbage in, garbage out'" pp113-126

Nov 26 – No Class (Thanksgiving)

Dec 1 - TBA

Dec 3 - Presentations

**Group Project #2 Due**

Dec 7-11 – RRR Week – Final Exam Review (TBD)

**Dec 15 - Final Exam (3-6pm, Room TBD)**