METHODS OF SOCIOLOGICAL RESEARCH – GRADUATE STATISTICS 1
Sociology 271B, Fall 2017
University of California, Berkeley

Instructor: David J. Harding, dharding@berkeley.edu
Office Hours: Mondays 11am-noon & Fridays 12:30-2:30 pm or by appointment, 462 Barrows
Schedule an appointment online at dharding.youcanbook.me

GSIs:
Robert Pickett, robertempickett@gmail.com, Office Hours: TBA
Aniket Kesari, akesari@berkeley.edu, Office Hours: TBA

Lecture: Wednesdays 9:30-noon, 402 Barrows
Labs: Thursdays 10-noon, 402 Barrows or CSLS ??? (first meeting August 24)

Course Website: https://bcourses.berkeley.edu/
Textbook: Alan Agresti and Barbara Finlay, *Statistical Methods for the Social Sciences*, 4th ed.

Course Description

Sociology 271B is the first of two courses on statistical analysis of numerical data designed for sociology Ph.D. students. This course serves as an introduction to statistics for sociological research primarily for sociology graduate students who do not have extensive experience with quantitative methods. Before beginning this course, students are expected to have some familiarity with basic statistical concepts, including populations and sampling, probability, measures of central tendency and variation, the normal distribution, and simple univariate tests of means and proportions. Principal activities include: 1. Explore the statistical concepts and methods that sociologists most commonly use to gather and analyze quantitative evidence. 2. Use Stata (a popular computer program) to put those skills into practice. 3. Apply the skills to sociological data to gain facility and confidence in the use of these methods.

Course Goals

After successfully completing this course, you will be able to:

(1) Understand the basic logic of statistical modeling
(2) Identify the appropriate type of model given a specific type of data
(3) Construct an appropriate model to shed light on a focused research question
(4) Estimate simple statistical models in Stata
(5) Interpret the results of statistical models and discuss their relevance for evaluating competing theories of a particular phenomenon

This course will emphasize practical application and intuition rather than mathematical details or statistical theory. (Students desiring more mathematically-oriented training are welcome to consult the instructor for other opportunities on campus.) We will work closely with real data throughout the semester in order to learn by doing.

Core Elements of the Course

Lecture and Lab: Attendance and active participation in lecture and lab are essential. If you miss more than one lecture or lab session, your highest possible grade will be A-; if you miss more than two lectures or labs, your highest possible grade will be a B+; etc. If you miss more than four lectures and labs combined, you will receive a failing grade for the course. If an emergency necessitates that you miss one lab or lecture session, please contact the GSI and the instructor (in advance, if possible) so we can arrange for you to make it up. You will need paper, something to write with, and a calculator capable of exponents and logs. If you don't understand something, please ask!

Weekly Assignments: Assignments will involve both hand calculations and, during most weeks, Stata data analysis. In your Stata data analysis you will address a research question of your choice by analyzing data of your choice using one of the techniques we discussed during the previous lecture. The write-up must follow the template provided on the bCourses website and be no longer than 1-2 single-spaced pages (not including the appendix of Stata output). *It is highly recommended that you spend some time during the first week of the semester to choose a single dataset that interests you that you will use throughout the semester for your assignments and possibly for your final paper (see below for where to find data).* The assignments are due in your GSI's mailbox in the Sociology Department on **Mondays by 4 pm**. You must turn in hard copies; electronic submissions will not be accepted. One grade level will be automatically deducted from assignments that are 1-2 days late, and assignments that are more than 2 days late will not be accepted. You may collaborate on these assignments with other students but each individual student must write up his/her own work in his/her own words and submit his/her own Stata output; copying is plagiarism and will be treated as such. Assignments will be graded on the following ordinal scale: 0 = not turned in; 1 = below expectations; 2 = meets expectations; 3 = exceeds expectations (with 2 being the modal grade and a 3 granted only for truly exceptional work). The first assignment will be due on Monday, August 28 and the last assignment will be due Monday, November 13 (there will be no assignment for Lecture 13 – use that week to work on your final paper). *Labor Day is Monday, September 4, so the second assignment will be due Tuesday, September 5.*

Final Paper: Each student will write a final paper that uses statistical methods from the course to examine one or more theoretically motivated hypotheses. The paper should be 12-15 pages double spaced (not counting cover page, tables, or references). The final paper will be due at **5 pm on Friday, December 8**. It should be uploaded on the bCourses website as a Word document. A one-two page single-spaced *proposal* for the final paper will be due at **5 pm on Wednesday, October 18**, also uploaded to the bCourses website. The proposal should state the research question, hypotheses, motivation, data, variables, and analysis methods. Each student will provide peer feedback to two other students on their proposals the following week. More information on the final paper, proposal, and peer feedback will be provided as the deadlines approach.

Take-Home Final Exam: You will have two days to complete the exam, although I expect the typical student to complete it in 4-5 hours. The exam will be comprehensive and will involve both hand calculations and short answer questions. You will <u>not</u> be tested on Stata on the exam. The exam is open-book and open-note, but you may not consult with other individuals (unlike the weekly assignments, the exam is not a group activity). The exam is tentatively scheduled for Tuesday, December 12 and Wednesday, December 13, but if this conflicts with end-of -semester assignments for other first-year courses, we can reschedule it. The exam will be available at 4 pm on December 11, and hard copy is due in your GSI's mailbox in the Sociology Department at 4 pm on December 13.

Grading: Your grade for the course will be based on the Weekly Assignments (40%), the Final Paper (20%), the Final Exam (25%), Peer Feedback on the final paper proposal (5%), and participation in lecture/lab (10%).

Data for Weekly Assignments and Final Paper

Over the course of the semester you will work with one (or more) datasets of your choosing in weekly assignments and a final paper. You will then build upon, improve, and revise this final paper in the spring semester in 271C. You may use any data that you like, and are encouraged to consult with the instructor and/or GSI as you choose your data. Below are some online repositories where data are available:

sda.berkeley.edu/archive.htm
icpsr.umich.edu/icpsrweb/ICPSR/access/index.jsp
norc.uchicago.edu/GSS+Website/
thearda.com/Archive/browse.asp
www.census.gov
www.ropercenter.uconn.edu/

Statistical Computing

The primary software for this course will be Stata (http://ww.stata.com). Stata is flexible, relatively user-friendly, and commonly used by social scientists. It has a large and diverse user community with many user-written commands that keep Stata continually up to data with new developments. You will need to use Stata for most weekly assignments and for your final paper. Most Stata instruction will occur in lab, though many examples will be presented in lecture. You are encouraged to seek help with Stata questions at Berkeley's D-lab (http://dlab.berkeley.edu), located on the third floor of Barrows Hall. Another good resource is UCLA's Stata website (http://www.ats.ucla.edu/stat/stata/).

Students in 271B will have access to the Demography Computer Lab, a Linux server with a windows and terminal-based computing environment that can be accessed remotely from anywhere you have internet access. The Demography Computer Lab includes Stata and other statistical and database computing applications. This means you will not need to purchase your own copy of Stata for the course. Basics of using this system will be covered the first week of lab. See also http://lab.demog.berkeley.edu/.

Keys to Success in an Introductory Statistics Course

Since this is a statistics course, it will be very different from the typical sociology course. Because most of the material is cumulative, it is absolutely essential that you keep up with the course material.

- The readings are relatively short, but they are dense and need to be read carefully. You will have a much easier time understanding the lecture material if you have done the reading ahead of time.
- For most students, learning statistics requires thinking through how to solve problems. Statistics cannot be learned simply by reading a book or listening to a lecture. You should not expect to fully understand the material until after you have completed the relevant assignment.
- Learning statistics is in some ways like learning a language, and it is important not to be intimidated by new terms or the use of letters (Greek letters, even) to represent quantities or

concepts. It is often helpful to write in plain language the meaning of the quantities or concepts represented by a letter or symbol.

- The most effective way to study for the exams is to do practice problems. This is why the assignments are a critical part of the course. Also note that the Agresti and Finlay textbook includes answers to many odd-numbered questions; if you feel that you do not adequately understand some part of the material, these may help you to work through it. You are strongly encouraged to do homework assignments and study for exams in groups.
- Lecture slides will be made available in advance on the bCourses website. This is so you do not have to copy formulas and diagrams during lecture, but lecture slides are not a substitute for careful note taking.
- If you find yourself falling behind, seek help *immediately* from your GSI or the professor during office hours.
- Please ask questions during lecture or during lab if you do not understand. If something is unclear to you, it is probably unclear to other students as well. Lectures and labs are planned to allow time for questions (and answers).

Weekly Schedule of Lectures/Topics and Readings

1	August 23	Introduction; Comparing Two Groups	AF Ch. 7
2	August 30	Association between two categorical variables	AF Ch. 8
3	September 6	Bivariate OLS and Correlation	AF Ch. 9
4	September 13	Matrices and Likelihood	Weisburg App. 2A1*, Myung 2003*
5	September 20	Multiple Regression	AF Ch. 10, 11
6	September 27	ANOVA and Regression with Mixed Predictors	AF Sections 12.1-3, 13.1-4
7	October 4	Polynomials and Model Selection	AF Sections 14.1, 14.5-6
8	October 11	Regression Diagnostics	AF Sections 14.2-3
9	October 18	The Generalized Linear Model *(Final Paper Proposal due 5 pm)*	AF Section 14.4
10	October 25	Logistic Regression I *(Peer Feedback due 5 pm)*	AF Sections 15.1-2
11	November 1	Logistic Regression II	AF Sections 15.1-3
12	November 8	Other Models for Discrete Outcomes	AF Sections 15.4-5
13	November 15	Missing Data	Acock, *JMF* 2005*
	November 22	*Thanksgiving Week, No Lecture or Lab*	
14	November 29	Review for final exam	
	December 8 (Friday)	Final Paper Due at 5 pm (upload Word file on course website)	
	December 12-13 (Tuesday-Wednesday)	Take Home Final Exam *(tentative)* Due in GSI mailbox 4pm, Dec. 14	

AF = Agresti and Finlay textbook
* reading available as PDF on course website

See also the supplemental videos for each week on the bcourses website under Files. These provide another avenue of introduction to some of the material we will cover in the course. I recommend watching the videos and doing the reading before the corresponding lecture.