Sociology 180B: Introduction to Data Analysis

Spring 2017
Tuesdays and Thursdays, 9-10:20am
Building 160, Room 332

Instructor: Emily Carian
ecarian@stanford.edu
Building 120, Room 053
Office hours: Thursdays 1-2:30pm, and by appointment

Teaching Assistant: Sheridan Stewart
sastew@stanford.edu
Office hours: Mondays 12:30-2:00pm, and by appointment

Course Description

The aim of this course is to provide students with an introduction to the statistical techniques used in sociological research. Students will learn the principles, benefits, and limitations of basic techniques, including linear and logistic regression, in order to both evaluate others' research and produce their own. Students will also practice implementing these techniques on real datasets using *Stata*, a statistical software package used by many sociologists. By the end of the course, students will be able to:

- Apply the appropriate statistical techniques to research questions of interest
- Understand and utilize the basics of *Stata*
- Critically evaluate the statistical methods used in existing sociological research

Required Materials

The following text is required for this course and is available on Amazon. Additional required readings are posted in the Files section on Canvas.

> Wheelan, Charles. 2014. *Naked Statistics: Stripping the Dread from the Data*. New York: W. W. Norton & Company.

We will be using *Stata 14* in this course. This software package is available in several versions (*Small Stata, Stata/IC,* and *Stata/SE*). Both *Stata/IC* and *Stata/SE* can be used for this course. *Stata/SE* is more expensive, but can open larger datasets. If you plan on using statistical techniques for a future project (like a thesis), you may consider buying *Stata/SE*, but this is not necessary for this class. There are several options for accessing *Stata*:

- Purchase a six month ($75), one year ($125), or perpetual ($198) license for use on your computer or laptop (see https://www.stata.com/order/new/edu/gradplans/student-pricing/).

- Use the computers in the library clusters, which are all outfitted with *Stata/IC*. (Because this option will not allow you to follow along in *Stata* in class, I highly recommend you take advantage of option 3 below.)
- Access *Stata* for free through Farmshare on your computer or laptop (instructions here for Mac; here for Windows). If you choose this option, make sure you are able to navigate to *Stata* **before** coming to class. Contact SSDS for help or drop in on their open hours, Monday-Thursday, 2-5pm in the Velma Denning Room in Green Library (note: they are not open for drop in hours week 1 of the quarter).

In addition to *Stata*, we will also make extensive use of the Canvas site. Please make sure to check the site regularly for announcements and to use the Assignment section to turn in all assignments.

Grading and Assignments

Students will be graded according to the following scheme:

Homework:	55%
Attendance and participation:	35%
Final project:	15%
Final exam:	20%

Homework (55%):

The four homework assignments will provide you with opportunities to demonstrate your understanding, interpret statistical results, and apply statistical techniques to real data. They will be essential to your final exam preparation. Each homework assignment will have three parts. First, you will answer big picture questions and explain statistical concepts in your own words. Second, you will interpret the results from statistical output. Third, you will craft your own statistical question and use real data to answer it. Homework due dates are indicated on the course schedule below. **On these dates, turn in your Word document and *Stata* .do file to the Assignment section of Canvas by class time.**

Attendance and participation (10%):

Most learning will take place in class and section. It is crucial that you show up! You will have many opportunities to earn your participation grade by taking part in activities, asking and answering questions, and otherwise engaging with the material. I recognize that some students are uncomfortable speaking in class and section; I encourage these students to speak with me regularly during office hours about the material to demonstrate their understanding.

Final Project (15%):

There are two options for the final project:

- Original research project: Using real data, you will construct a research question and design an analysis plan using one of the statistical techniques discussed in class.
- Analysis of a published research article project: You will identify a piece of sociological research that utilizes one of the techniques used in class, then summarize and evaluate it.

Both require you to present a brief summary of your findings on one of the last two days of class and turn in your final assignment by Wednesday, June 7 at 11:59pm. See Canvas for more information on each of these assignments and other relevant deadlines.

Final exam (20%):
The final exam will be a comprehensive test of the principles, benefits, limitations, and interpretation of the statistical techniques discussed in the course. The final exam will take place Monday, June 12, 8:30-11:30am as scheduled by the Registrar.

Course Policies

Laptops:
Laptops are allowed in class so students can use *Stata* during in-class activities or section. At times when *Stata* is not being used, **students are expected to refrain from using their laptops** (except students with documented disabilities). Not only does research show that laptops are distractions to student learning, taking notes on mathematical concepts is simpler by hand.

Late work:
Students are expected to turn in all assignments by the deadline noted below in the course schedule. In the case of an emergency that will hinder your ability to turn in an assignment on time, notify me as soon as possible to make arrangements. Exceptions to the deadline will be made on a case-by-case basis. Otherwise, late work will be penalized with a 1/3 of a grade reduction for being late (e.g., A- to B+), and an additional 1/3 of a grade reduction for each day it is not turned in.

Email:
Students are welcome to contact us via email with questions or clarifications. Before emailing, students should consider whether their question is better answered in office hours, in class, or over email. While I make an effort to respond to student emails on the same day, I cannot promise I will respond to emails sent after 5pm until the next day.

The Honor Code:
The Honor Code is Stanford's statement on academic integrity (read it in full [here]). With regard to this class, students are encouraged to work together on the homework assignments. However, students must complete their own work. For instance, all writing on the homework assignments should be your own and students must write their own *Stata* .do files; in the original data analysis plan of the

homework assignments, students should design their own questions. Students doing the Analysis of a Published Research Article Project should take care not to plagiarize. Penalties for violation of the Honor Code can be serious; please make sure you are completing your own work. (Go here for resources on how to avoid plagiarism.)

Resources

Disabilities:

Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is being made. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. The OAE is located at 563 Salvatierra Walk (phone: 650-723-1066).

Office Hours:

Office hours are regularly scheduled (see the top of page 1) and by appointment. When you have questions that were not answered in class, need help with homework or the final project, or want to demonstrate your understanding, please come to office hours.

Counseling and Psychological Services (CAPS):

CAPS is available to help students who experience a wide variety of personal, academic, and relationship concerns. Contact 650-723-3785 (24 hours/day) or go to their website.

Social Science Data and Software (SSDS):

Can't get the *Stata* code quite right? SSDS can help. SSDS is a group within the library that helps faculty, staff, and students with their statistical (and qualitative) software. SSDS has walk-in hours from Monday through Thursday, 2-5pm in the Velma Denning Room in Green Library or you can contact them via their website to make an appointment. Note that SSDS is for software help, not statistical help, so if you're having trouble interpreting your results or choosing a statistical technique, come to office hours.

Course Schedule

Note: Assignments and readings are due *on the date listed*.

Topic 1: Introduction to data analysis and *Stata*

April 4   Introduction to the class, statistics, and data
              Readings due: Wheelan chapter 1
              Assignments due: None

April 6	The scientific method and types of variables Readings due: Frankfort-Nachmias & Leon-Guerrero chapter 1 (Canvas) Assignments due: None
April 11	Descriptive statistics Readings due: Wheelan chapter 2 Assignments due: None (HW 1 distributed)
April 13	Descriptive statistics Readings due: None Assignments due: None

Topic 2: Principles of statistical testing

April 18	Distributions Readings due: Wheelan chapter 5, 8 Assignments due: HW 1 due (HW 2 distributed)
April 20	Standard error Readings due: Frankfort-Nachmias & Leon-Guerrero chapter 7 (Canvas) Assignments due: None
April 25	Hypothesis testing Readings due: Frankfort-Nachmias & Leon-Guerrero chapter 9, 11 (Canvas) Assignments due: None
April 27	Hypothesis testing Readings due: Explore: how do sociologists use hypothesis testing? (Canvas) Assignments due: None
May 2	Correlation Readings due: Wheelan chapter 4 Assignments due: Decision to complete option A or B for final project

Topic 3: Linear regression

May 4	Linear regression Readings due: Wheelan chapter 11 Assignments due: None
May 9	Multiple linear regression Readings due: Wheelan chapter 12 Assignments due: HW 2 due (HW 3 distributed)
May 11	Multiple linear regression Readings due: None Assignments due: None

May 16   Multiple linear regression
            Readings due: Explore: how do sociologists use linear regression? (Canvas)
            Assignments due: Identification of research question/article for final project

Topic 4: Logistic regression

May 18   Multiple linear regression versus logistic regression
            Readings due: Rajaretnam chapter 12 (Canvas)
            Assignments due: None

May 23   Logistic regression
            Readings due: None
            Assignments due: HW 3 due (HW 4 distributed)

May 25   Logistic regression
            Readings due: Explore: how do sociologists use logistic regression? (Canvas)
            Assignments due: None

Topic 4: Error, bias, and bad data

May 30   Error, bias, and bad data
            Readings due: Wheelan chapter 6, 7
            Assignments due: HW 4 due

June 1   Presentations and review
            Readings due: None
            Assignments due: None

June 6   Presentations and review
            Readings due: None
            Assignments due: None

Wednesday, June 7: Final project due, 11:59pm

Monday, June 12: Final exam, 8:30-11:30am