

Doctoral Program in Sociology
CUNY-Graduate Center

Sociological Statistics I

[Subject to Modification]

[Acknowledgement: Based originally on Professor Jeremy Porter's 2016 Sociological Statistics I course]

Version: November 15, 2022

Semester: Fall 2022

Course Code: SOC 71500

Classroom: Graduate Center 6418

Class time: Tuesdays, 4:15 – 6:15 pm

INSTRUCTORS

Professor: Leslie McCall

Office: Room 6112.09

Office Hours: Mondays, 4:00 – 5:00 pm, via Zoom, or by appointment

Phone: 212-817-8776 (office)

E-mail: lmccall@gc.cuny.edu

Teaching Assistant: Yue (Juliette) Zhu

Office: TBA

Office Hours: TBA

E-mail: Yue Zhu <yzhu3@gradcenter.cuny.edu>

COURSE DESCRIPTION

This course provides an overview of introductory statistics as applied to sociological and other social scientific research. Topics covered include single-variable data description (measures of central tendency, measures of variability, and graphing), fundamentals of theory-based inferential statistics (probability and the normal curve, samples and populations, hypothesis testing), and associations between two variables (ANOVA, Chi-square, correlation, and bivariate linear regression analysis). The course will also introduce students to the software package **R** for the analysis of social science data. No prior knowledge of statistics or R is necessary.

COURSE LEARNING GOAL

The goal of this course is to provide students with an up-to-date understanding of the principles underlying commonly used introductory statistical approaches within the social sciences. Specifically, students will acquire an understanding of some fundamental concepts in data description and analysis, become proficient in the performance of corresponding statistical procedures, be able to apply these statistical concepts and skills in new and real-world contexts, and have the knowledge to read and evaluate statistical research.

BLACKBOARD

All students must have access to Blackboard. Class materials, including slides, assignments, data, labs, readings, supplement materials and announcements, will be posted on Blackboard.

Make sure the email address and contact information in Blackboard is current so I can inform you of any changes in the course schedule or announcements. Make sure to check Blackboard often to keep updated.

COURSE MATERIALS AND RESOURCES

Textbooks:

Required: Lane, David M., et al. 2015. *Introduction to Statistics*. This book is free [online](#) (click the link for the pdf version on left side of screen). Primary textbook (unreferenced page numbers below refer to this text).

Required: Frankfort-Nachmias, Chava, and Anna Leon-Guerroero. 2017. *Social Statistics for a Diverse Society, Eighth Edition* (Pine Forge Press). An earlier edition of this book is available online through [Internet Archives](#) and pdfs are available upon request. [Shorthand below: FN-LG.]

Required: Ismay, Chester, and Albert Kim. *Statistical Inference via Data Science: A Modern Dive into R and the Tidyverse* (CRC Press). This book is free [online](#). [Shorthand below: MD.]

Other reading/review materials and links are provided on Blackboard (e.g., references to help with R) and will be added as needed.

Computer programs: R is open source and freely available. Links to program are provided on Blackboard.

Calculator: Examples of the use of computer programs to describe data and perform statistical analyses will be given during class time but students will also be expected to calculate some statistics by hand or calculator and thus to master use and interpretation of the relevant formulas. Simple calculating may be required, but can usually be performed on included phone calculators. Also, the scientific view of the calculator that comes included in the MacOS and Windows programs will work.

COURSE ASSESSMENT

Throughout the semester there will be lab assignments, a take-home mid-term exam, and a final paper. Late work will be accepted for full earned credit if and only if special arrangements are made with me prior to the due date; otherwise 5% of the total points possible will be deducted for each day the assignment is late. There will be no incomplete grades given in this course unless the situation is deemed unusual by the instructor. All assignments must be submitted to pass the course.

Assignments (Labs): Assignments will be discussed during each week's lab session, which will take place during the second half of each class. These weekly assignments will allow the student to learn how to produce the relevant statistics covered that week in class, usually by using R to do so. The objective is to obtain and interpret the appropriate results from R, and to write up the findings in a short lab report in the style of a memo or

essay with nicely and clearly formatted tables and output as appropriate. Each assignment is due on Blackboard the following week before class.

Proportion of final grade: *Assignments* are worth a total of ~40%.

[3 points for first 6 assignments; 7 points for last 3 assignments]

Exam: The mid-term exam will be similar to the assignments but larger and more encompassing in scope, covering several topics at once.

Proportion of final grade: 20%.

Final Project: The final paper provides an opportunity for students to compose their own quantitative research paper from scratch. Students will develop research questions and obtain the appropriate data to investigate those questions. The only requirement is that the data are in a machine-readable format so that they can be entered into a software program (i.e., R) for analysis; the data also must be appropriate for quantitative data analysis (meaning that there must be a sufficient sample size and numeric representation of all response categories). Students will be responsible for obtaining data from a real study. We will cover potential data resources in class. The final project report should resemble an early version of a research article. All students will be required to present their project in a “conference style” presentation at the end of the semester, which will be described in class (and will be 10-12 minutes in length).

Proportion of final grade: *presentation* is worth 15%, and *final paper* is worth 25% for a total of ~40%.

Format: All written work for this class must be clear and legible in essay/memo format and the statistical/data results must be presented in proper technical format, which will be described in class. Also, the print format should be set to double-spaced, 12-point Times Roman font, with one-inch margins on all sides of the page.

GRADING SYSTEM

Grades are assigned based on the percentage of accumulated points:

Grades are assigned based on percentage of accumulated points:

A : 93-100; *B+*: 87-89; *C+*: 77-79; *D+*: 67-69; *F*: 0-<60

A-: 90-92; *B* : 83-86; *C* : 73-76; *D* : 60-66;

B-: 80-82; *C-*: 70-72;

ATTENDANCE

Students are responsible for all material presented in lectures and labs. It is expected that students will attend lectures/lab; however, attendance will not typically be taken past the first few weeks while I am learning names and faces. If a student misses a class, then the student is responsible for obtaining both the course material that was missed as well as any class announcements from his/her classmates.

Students are expected to arrive on time for the beginning of class. Students anticipating late arrival should do their best to get situated in class without being disruptive. A pattern of tardiness can negatively affect your grade. Lab assignments are due at the beginning of class on the due date specified, and are always due one week later unless otherwise noted. No credit will

be awarded for late work without prior approval from instructor. Late work is strongly discouraged.

Cell Phones must be turned off during class (unless using for class).

PLAGIARISM AND COOPERATION ON ASSIGNMENTS

Plagiarism and cheating are violations of CUNY’s policy on academic integrity. By registering in this course, you are promising to abide by all the requirements stated in this policy. Students in breach of this policy are liable to severe penalty, including disciplinary action.

Students are expected to do their own assignments. Discussion of the course material among students is allowed in the spirit of increasing common understanding. Each student, however, is expected to do their own written assignments. A major section of verbatim text common to submitted assignments is not acceptable. For the final paper, each student should write their own paper and perform their own statistical analysis without assistance from outside persons, except the professor and the teaching assistant (minor help from other students with R syntax and data sources is permitted).

WEEKLY SCHEDULE (NOTE: SUBJECT TO CHANGE)

Week 1: August 30

Topic: Course introduction and organization
Basic concepts and terms

Week 2: September 6

Lab: Introduction to R
Reading: MD, Chp. 1

Week 3: September 13

Topic: Organization of data
Data description
Reading: 10-122 (pages refer to online textbook unless otherwise noted); MD, Chp. 2
Lab A: Data exploration [Assignment A handed out]

Week 4: September 20

Topic: Measures of central tendency
Measures of dispersion
Reading: 123-163; MD, Chp. 2
Lab B: Data description [Assignment B handed out]

Week 5: September 27 (GC holiday; no class)

Week 6: October 4 (GC holiday; no class)

Week 7: October 11

Topic: Correlation
Reading: 164-184; FN-LG pp. 282-291 (in Chp. 10); MD, Chp. 5 (intro through 5.1.1)
Lab C: Correlation; Data cleaning and preparation [Assignment C handed out]

Week 8: October 18

Topic: Bivariate regression, Part I; Z-scores
Reading: 461-471; 482-494; FN-LG pp. 325-342 (in Chp. 12), pp. 130-132 (in Chp. 5); MD, Chp. 5 (all except 5.2)
Lab D: Bivariate regression; Data cleaning and preparation [Assignment D handed out]
Other: Hand out take-home mid-term exam

Week 9: October 25

Topic: Probability and the normal curve
Samples and populations
Reading: Main reading: FN-LG Chps. 5-6; MD, Chp. 7 and Chp. 5 (section 5.2); Additional (e-text): 185-202; 248-271; 299-310.
Lab E: Bivariate regression, cont'd; Intro to dummy variables
Sampling distributions [Assignment E, FN-LG, p. 170, handed out; due 11/8]

Week 10: November 1

Topic: Statistical inference (point estimates and confidence intervals)
Reading: Main reading: FN-LG Chps. 6-7; MD, Chp. 8; Additional (e-text): 328-368.
Lab: Exploring datasets/topics for final paper [Paper proposal Assignment handed out]
Other: Take-home exam due

Week 11: November 8

Topic: Hypothesis testing and significance tests (differences in means)
Reading: Main reading: FN-LG Chp. 8; MD, Chp. 9; Additional/optional (e-text): 311-315; 369-446.
Lab F: Differences in means, two groups [Assignment F handed out]

Week 12: November 15

Topic: ANOVA
Reading: Main reading: FN-LG Chp. 11; Additional (e-text): (222-247); 515-575.
Lab G: Differences in means, multiple groups [Assignments G handed out]

Other: **Final paper proposal due Friday, Nov. 18, 5pm (email to me)**

Week 13: November 22

Topic: Bivariate regression, Part II; Intro to Multiple Regression
Reading: Main reading: FN-LG Chp. 12; MD, Chp. 10; Additional (e-text): 461-514.
Lab H: Regression coefficient significance testing [Assignment H handed out]

Week 14: November 29

Topic: Chi-square tests of association
Reading: Main reading: FN-LG Chp. 10; MD, Chp. 11; Additional (e-text): 597-615.
Lab I: Crosstabs and tests of association [Assignment I handed out]

Week 15: December 6

Other: Student presentations

Week 16: December 13

Other: Student presentations

Week 17: December 20, Final Papers Due @ 5 pm.