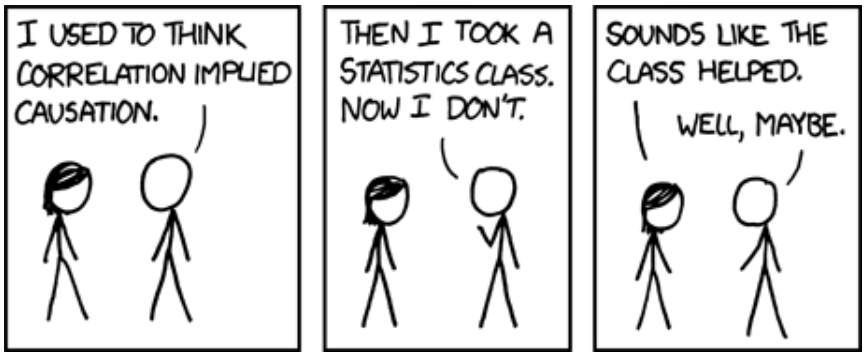
SOCI 210 01: Social Statistics Fall 2024

Professor Erick Axxe
Email: axxe@hendrix.edu
Student hours: MWF 10-11am (or by appt.)
Book an appointment: axxe.youcanbook.me
Office: Mills 220
Class time: MWF 8:10-9am
Class location: Snoddy Computer Lab



https://xkcd.com

Course Description: In this course, we will explore the fundamental principles of statistical inference and develop the programming skills necessary to analyze a wide range of social science questions using data. For example, we will examine questions such as: Who is most likely to win the upcoming presidential election? Do countries become less democratic when their leaders are assassinated? Is there racial discrimination in the labor market? These are just a few of the questions we will address during the course.

Students are not expected to have any prior programming knowledge or experience. The course will focus on bite-sized assignments designed to build coding and statistical skills from the ground up. By the end of the course, students will be equipped to work in any setting that requires a social scientific approach to data analysis, whether in policy non-profits, government, Silicon Valley, Wall Street, or beyond.

Course Website: Link to Teams channel

Required Textbook: Elena Llaudet and Kosuke Imai. *Data Analysis for Social Science: A Friendly and Practical Introduction*. United States, Princeton University Press, 2022.

Required Statistical Softwares: R (www.r-project.org) and RStudio (www.rstudio.com). Both are free!

Tentative Course Schedule: The tentative course schedule is detailed in the table below. (Disclaimer: The schedule, policies, procedures, and assignments in this course are subject to change in the event of extenuating circumstances, by mutual agreement, and/or to ensure better student learning.)

Day	Topic	Materials	Key Concepts and *R* Code	PSet
W 08/28	Course Introduction			
F 08/30	Introduction to R and RStudio	1-1.6	*R*: +, -, *, /, <-, '', (), sqrt(), #	#0
M 09/02	NO CLASS (Labor Day)			
W 09/04	Observations and Variables	1.7	dataframes, observations, variables, unit of observation, *i*, character vs. numeric variables, binary vs. non-binary variables, *n*; *R*: setwd(), read.csv(), View(), head(), dim()	
F 09/06	In-class exercise #1	1.8-1.10	mean or average, $\sum$, unit of measurement; *R*: $, mean()	
M 09/09	Computing and Interpreting Means			#1
W 09/11	Estimating Causal Effects with Randomized Experiments	2-2.4	causal relationships, treatment (X) vs. outcome variables (Y), potential outcomes, factual vs. counterfactual outcomes, fundamental problem of causal inference, individual vs. average causal effects, randomized experiments, random treatment assignment, treatment and control groups, pre-treatment characteristics, the difference-in-means estimator	
F 09/13	In-class exercise #2	2.5-2.7	*R*: ==, ifelse(), []	
M 09/16	Review			#2
W 09/18	Survey Research and Exploring One Variable at a Time	3-3.4	sample, representative sample, random sampling, table of frequencies, table of proportions, histogram, descriptive statistics (mean, median, standard deviation, and variance); *R*: table(), prop.table(), hist(), median(), sd(), var(), ^	
F 09/20	In-class exercise #3			
M 09/23	Review			#3
W 09/25	Exploring the Relationship Between Two Variables	3.5-3.7	scatter plot, correlation; *R*: plot(), cor()	
F 09/27	Predicting Non-Binary Outcomes Using Linear Regression	4-4.4.1	prediction and correlation, predicted ($\widehat{Y}$) vs. actual outcome (Y), prediction errors ($\widehat{\epsilon}$), the least squares method, the linear regression model, $\widehat{Y} = \widehat{\alpha} + \widehat{\beta}X$, interpretation of coefficients, intercept ($\widehat{\alpha}$) and slope ($\widehat{\beta}$), $\triangle\widehat{Y} = \widehat{\beta}\triangle X$; *R*: lm($Y \sim X$), abline()	
M 09/30	Review			#4

Day	Topic	Materials	Key Concepts and *R* Code	PSet
W 10/02	Predicting Binary Outcomes Using Linear Regression	4.6-4.9 (skip 4.8)	R^2, relationship between R^2 and correlation	
F 10/04	Review			#5
M 10/07	Estimating Causal Effects with Observational Data and the Problem of Confounders	5-5.3.1	observational studies vs. randomized experiments, confounders (Z), interpretation of $\widehat{\alpha}$ and $\widehat{\beta}$ when X is binary and identifies treatment assignment	
W 10/09	Controlling for Confounders Using Multiple Linear Regression	5.3.2-5.4.2	multiple vs. simple linear regression models, new interpretation of coefficients	
F 10/11	NO CLASS (Dr. Axxe is traveling)			
M 10/14	In-class exercise #4			
W 10/16	Review			#6
M 10/21	Internal vs. External Validity	5.5-5.7	internal validity, external validity	
W 10/23	Review			#7
F 10/25	MIDTERM EXAM			
M 10/28	Midterm Review			
W 10/30	Probability	6-6.8	probability, random variables, probability distributions, Bernoulli vs. normal distribution, the standard normal distribution, population parameters vs. sample statistics, the law of large numbers, the central limit theorem	
F 11/01	Hypothesis Testing with Coefficients	7-7.6	hypothesis testing, confidence intervals, test statistic, standard error of $\widehat{\beta}$, *R*: summary()$coef	
M 11/04	Review			#8
W 11/06	In-class exercise #5	7.7 (PDF)		
F 11/08	In-class exercise #6			
M 11/11	Hypothesis testing continued T-test and ANOVA	T-tests & ANOVA	*R*: t.test(), aov()	

Day	Topic	Materials	Key Concepts and *R* Code	PSet
W 11/13	Chi-square (χ^2) test	Chi-square tests	*R*: chisq.test()	
F 11/15	Review			#9
M 11/18	NO CLASS (Dr. Axxe is traveling)			
W 11/20	Data visualizations Tidyverse & GGplot2	Data visualizations, pt. 1		
F 11/22	Data visualizations GGplot2 & Stargazer	Data visualizations, pt. 2	*R*: ggplot(), stargazer()	
M 11/25	Review			#10
M 12/02	Final exam review			
W 12/04	Final exam			
F 12/06	Exam review & final presentations preparation			
F 12/13	Final presentations (2-5pm)			

Note: In the event of a class cancellation such as for severe weather, students are expected to continue with readings as originally scheduled. Any assignments scheduled remain due unless other instructions are posted on the course website or communicated on the discussion board.

Evaluation: The final grade will be based on the following:

1. *Class Participation* (*10%*). Based on (1) performance (not participation) in daily polls and (2) participation in in-class discussions.
2. *Problem Sets* (*50%*). Take-home problem sets will be given throughout the course. There are 10 required problem sets in total.
    - *Collaboration policy*: While working on problem sets, students are *encouraged* to work with a partner.
    - *Due date*: Problem sets are due *before* the class in which they are discussed. (See tentative schedule above.)
    - *Second submission policy*: On days that assignments are due, students will have the opportunity to revise and resubmit their original submission in small groups during class. For each corrected answer, students will receive 50% credit on their homework grade.
    - *Late policy*: Late submissions will not be accepted unless special permission is granted in advance. (We will review the homework on the days that it is due.) If something comes up, please send me an email petitioning an extension. IMPORTANT NOTE: No more than two extensions will be granted per student.
    - *Grading and feedback*: I will grade and return assignments within seven days of the original due date.
3. *Midterm* (*10%*) and *Final Exam* (*15%*). No collaboration is permitted during the exams. However, students are allowed to use a single letter-sized sheet of *hand-written* notes (two-sided).
4. *Final Presentation* (*15%*). On December 13th from 2-5 pm, students will present a statistical analysis to the class.

Course Policies: This this course adheres to policies and procedures that apply to all Hendrix courses with regard to academic accommodations, academic integrity, diversity, health accommodations, Title IX, class attendance and Hendrix Career Competencies. (Follow links to see the policies.)

Resources: Please make use of campus resources available to you. These include Bailey Library and the Writing Center.

Any student who has difficulty affording groceries or accessing sufficient food to eat every day, or who lacks a safe and stable place to live, and believes this may affect their performance in the course, is urged to contact the Dean of Students for support. Furthermore, please notify me if you are comfortable in doing so. This will enable me to provide any resources that I may possess. Here is a list of local food pantries. For note-taking, I recommend OneNote. Grammarly is free software which checks and suggests fixes to grammatical errors. Zotero is a free citation management software.

Communication preferences: General questions about the course or assignments should be posted in the Teams channel titled "Q-and-A". Otherwise, please send me an email (axxe@hendrix.edu). Please *do not* use Teams' chat function to communicate with me.

Attendance policy: (1) class attendance is mandatory, (2) for an absence to be excused, students must send an email to Dr. Axxe *before* the missed classed period, (3) students whose unexcused absences reach or exceed 50% over a three-week period will be withdrawn from the course, (4) students are

expected to come to class on time and ready to participate (having done the readings and completed the problem sets listed for that day), and (5) make-up exams will only be offered in justified, special circumstances.

Course Goals and Learning Objectives:

Goals	Objectives
1. Students will know how to recognize and interpret quantitative information	(a) Students will be able to read and understand quantitative data in various formats (b) Students will be able to communicate the meaning of quantitative data and the results of data analysis
2. Students will understand the theoretical basis of quantitative reasoning	(a) Students will be able to explain the basic concepts of quantitative reasoning, such as variables, constants, and estimates (b) Students will be able to understand how inferences are drawn from quantitative analysis (c) Students will be able to recognize the limitations of quantitative methods
3. Students will understand the practical application of quantitative data analysis	(a) Students will be able to determine and use appropriate quantitative methods to solve problems (b) Students will be able to accurately interpret the results of data analyses (c) Students will be able to assess results for reasonableness