Statistical Analysis in Political Science

GOV 350K
Spring, 2023
University of Texas at Austin

Instructor: Xin Nong — E-mail: xnong@utexas.edu
Classroom Location: PAR 105 — Class Hours: T & TH 5-6:30 p.m.
Office: MEZ 3.228 — Office Hours: T 3-4 p.m. & W 3-5 p.m.

Course Overview

Today, data are available on almost any topic. In a now-classic quote, E.O. Wilson said "We are drowning in information, while starving for wisdom. The world henceforth will be run by synthesizers, people able to put together the right information at the right time, think critically about it, and make important choices wisely."

This course will teach students the skills needed to unlock valuable and important knowledge from data. Students will learn how to obtain data, how to manipulate it into a useful format, how to analyze it to produce insights, and how to present these findings effectively. These skills will be useful in future coursework as well as in a wide range of professional contexts. In other words, the skills you will learn in this class are highly valuable!

This course will teach students how to examine data to uncover patterns, how to estimate simple statistical models to understand relationships between variables, how to build predictive models, and how to learn about cause and effect. There are no mathematical prerequisites for the course and we will not use much beyond basic arithmetic. The main requirement is a willingness to think carefully.

For data analysis, we will use the R statistical programming environment. R is open source and freely available. It is commonly used in academia, business, and other contexts, and includes tools for simple analyses as well as more advanced methods. Many employers view basic statistical knowledge and the ability to use R as great assets when making hiring decisions.

Although this is not a math class per se, it will be somewhat mathematical. There are no prerequisites beyond basic middle and high school math, but students will have to use mathematical reasoning to understand the course's concepts. Students should not worry that they are not a "math person." If you know +, -, ×, and ÷, you just need to be willing to learn the rest.

Required Materials

The main (required) text for the class is

- OpenIntro Statistics (4th Edition), by David Diez, Mine Cetinkaya-Rundel and Christopher Barr.

This book is available as a free pdf here: https://www.openintro.org/book/os/ Students sometimes find it helpful to have multiple sources for material. Below, I list several optional texts. Students should feel free to consult these for slightly different perspectives on the ideas we cover. I will also assign other readings, which will be posted on Canvas. Any assigned readings will be announced to students in advance.

Optional texts for students who want multiple references:

- Data Science: A Gentle Introduction, by James Scott.
  - Nicely written and thoughtful overview of basic probability and statistics, written by a UT professor. Available as free pdf here:
    https://jgscott.github.io/STA371H_Spring2018/files/DataScience.pdf
- Using R for Introductory Statistics, by John Verzani.
  - Available as free pdf here:
    https://cran.rproject.org/doc/contrib/Verzani-SimpleR.pdf
    A good combination of basic intro R and intro statistics. Note: this book uses "=" as the assignment operator instead of "<-" that we will use in the class so you'll have to adjust when looking at the code (not that it makes a difference which one you use).
- An Introduction to R.
  - Basic but fairly comprehensive introduction to R. Avaiable as free pdf here:
    https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
    It's pretty dry (reads more like an instruction manual than a textbook), but that makes it good for reference.
- A Modern Dive into R and the tidyverse, by Albert Kim and Kelly McConville.
  - (Referred to simply as "Modern Dive"). Good introduction to R and RStudio, but relies heavily on some fancier R packages that we won't cover directly (the so-called "tidyverse") so much of the R code will look different from what we use.
- R for Data Science, by Hadley Wickham and Garrett Grolemund.
  - Free online (see https://r4ds.had.co.nz) book that provides a great introduction to modern R for statistics. We may assign some parts of this book on general R and best practices. But it relies heavily on some fancier R packages that we won't cover directly (the so-called "tidyverse") so much of the R code will look different from what we use.

Prerequisites/Corequisites

There are no prerequisites for this course.

Attendance and participation

Attendance and participation are worth 10 percent of your grade. I expect all students to attend class regularly. On some days, we will have in-class pop quizzes. And each time you miss an in-class pop quiz, your final score will be deducted by one point (up to 10 points). Each student has two freebies for pop quizzes (no questions asked).

Problem Sets

Problem sets are worth 20 percent of your grade. They will be posted and turned in through Canvas. Students can work together in small groups on problem sets, but I trust that you all know the difference between helping each other something figure out and copying. Sharing answers in online forums, in large groups, or in shared documents is not allowed.

Lab Assignments

We will be using the R statistical package, which can be downloaded for free at http://www.r-project.org. It is recommended that you run R through a program called RStudio, which is also free to download. We will provide instructions for installing and using both of these programs.

Lab assignments are worth 20 percent of your grade. They will be posted and turned in through Canvas. Students should write their responses in R Markdown (through RStudio) and will turn in "knitted" HTML files with their lab writeups. We will provide instructions for students on using R Markdown, but you can also find a simple introduction here: https://rmarkdown.rstudio.com/lesson-1.html (we will only use the most basic functionality of R Markdown in this course) Students can work together in small groups on lab assignments, but I trust that you all know the difference between helping each other something figure out and copying. All keystrokes must be your own, not copied and pasted from another student's assignment. In other words, you can talk with a (small) group of other students about how to do these assignments, but you must actually do them on your own — no pasting in someone else's code. Sharing answers online or in large groups or shared documents is not allowed.

Midterm Exam

The in-class midterm exam accounts for 25 percent of your grade. It will look like the problem sets but will be more comprehensive.

Final Project

The final project accounts for 25 percent of your grade. Each student will work on a semester long project in which they pose their own research question, then gather and analyze data in order to learn about the answer. The paper will have several intermediate assignments that are designed to help students.

Dates for project assignments (all to be submitted through Canvas)

- Choose a Research Question (due by Th 2/16 at 11pm)
- Gather Data (due by Th 3/28 at 11pm)
- Final Paper (due by Th 4/27 at 11pm)

Grading Policy

- Attendance and participation: 10%
- Problem Sets: 20%
- Lab assignments: 20%
- Midterm exam: 25%
- Final project: 25%

The following grade cutoffs will be used to calculate final course grades:

93-100	A	90-92	A-	87-89	B+
83-86	B	80-82	B-	77-79	C+
73-76	C	70-72	C-	60-69	D
0-59	F				

Course Policies

Make-up Exams and Late Assignment

Late problem sets, lab assignments, and the final project will be penalized by 10 points for each day late (so 10 points off if it's 1 minute late, 20 points off if it's 24 hours and one minute late, etc.). Problem sets and lab assignments will not be accepted once the answer key has been posted.

There will be no deadline extensions without prior written consent from the professor. These special accommodations will be granted only for valid medical or other serious reasons, defined at the professor's discretion. Students who miss the midterm exam and provide an excuse afterwards will not be granted a makeup and may be given zero percent on the exam.

Academic Integrity and Honesty

UT students should seek to be totally honest in their dealings with others. They should complete their own work and be evaluated based upon that work. They should avoid academic dishonesty and misconduct in all its forms, including plagiarism, fabrication or falsification, cheating, and other academic misconduct. Students are expected not only to be honest but also to assist other students in fulfilling their commitment to be honest.

While students should make a general commitment to proper academic conduct, there are still specific skills most students need to master over time in order to correctly cite sources, especially in the age of the internet, as well as deal with the stress and strain of college life without resorting to cheating. Please know that as your professor I will notice instances of cheating on exams or plagiarizing on papers.

Please also see the University Honor Code site for more information at:
http://registrar.utexas.edu/catalogs/gi09-10/ch01/index.html.

Accommodations for Disabilities

The University of Texas at Austin is committed to providing a working and learning atmosphere that reasonably accommodates qualified persons with disabilities. If you have any disability which may impair your ability to complete this course successfully, please contact the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471- 6259, http://www.utexas.edu/diversity/ddce/ssd/.

All students needing special accommodations must contact both the students with disabilities and the instructor at least two weeks prior to the examination in order to assure special accommodations can be made. Students not taking action by this time may be required to take the exam in the standard time and place.

Religious Holidays

By UT Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

Diversity and Inclusion

UT-Austin is committed to creating an environment of inclusion among the diverse identities represented at the University of Texas at Austin. And I would like to create a learning environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ability, etc.)

If you have a name and/or set of pronouns that differ from those that appear in your official records, please let me know! I will gladly honor your request to address you by a name that is different from what appears on the official roster, and by the gender pronouns you use. You can add a "preferred name" with the Gender and Sexuality Center:
http://diversity.utexas.edu/genderandsexuality/publications-and-resources/.

For instructions on how to add your pronouns to Canvas, visit:
https://utexas.instructure.com/courses/633028/pages/profile-pronouns

Emergencies

Please also see the following recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767, http://www.utexas.edu/safety/. 1) Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside. 2) Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building. 3) Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class. 4) In the event of an evacuation, follow the instruction of faculty or class instructors. 5) Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office. 6) Behavior Concerns Advice Line (BCAL): 512-232-5050. 7) Link to information regarding emergency evacuation routes and emergency procedures can be found at: www.utexas.edu/emergency.

Course Outline

Jan 10: Syllabus and Introduction

Jan 12: Probability and Statistics

Jan 17: Samples and Sampling Methods

Jan 19 In-Class Lab Session
Problem Set 1 Due at 11 pm on Jan 19

Jan 24 Probability and Probability Distributions

Jan 26 Important Distributions Lab Assignment 1 Due at 11 pm on Jan 26

Jan 31 Important Distributions

Feb 2 Sampling Distributions

Feb 7 Estimating Proportions and Confidence Intervals

Feb 9 Estimating Means and Confidence Intervals

Feb 14 In-Class Lab Session
Problem Set 2 Due at 11pm on Feb 14

Feb 16 Hypothesis Testing: Single Group Proportion and Mean

Feb 21 Hypothesis Testing: Two groups Proportions and Means
Lab Assignment 2 Due at 11 pm on Feb 21

Feb 23 Hypothesis Testing: Categorical Variables

Feb 28 In-Class Lab Session
Problem Set 3 Due at 11 pm on Feb 28

March 2 Review Session

Mar 7 Midterm
Lab Assignment 3 Due at 11 pm on Mar 8

Mar 9 Associations and Linear Regressions: Introduction

Mar 14, 16 Spring Break

Mar 21 Linear Regressions: Interpretations

Mar 23 In-Class Lab Session
Problem Set 4 Due at 11 pm on Mar 23

Mar 28 Multivariate Regressions: Introduction

Mar 30 Multivariate Regressions: Extension
Lab Assignment 4 Due at 11 pm on Mar 30

Apr 4 In-Class Lab Session

Apr 6 Logistic Regression
Problem Set 5 Due at 11 pm on Apr 6

Apr 11 In-Class Lab Session

Apr 13 Causal Inference and Counterfactual Framework

Apr 18 Difference-in-Differences
Lab Assignment 5 Due at 11 pm on Apr 18

Apr 20 Instrumental Variables and Regression Discontinuity

Apr 27 Final Project Due