GOV 350K – Statistical Analysis in Political Science
Unique # 37375
Spring 2024

Professor: Stephen Jessee
Email: sjessee@utexas.edu
Office hours: Tu 9-12

Teaching Assistants: TBD

All office hours will be on Zoom. See links in "Calendar" tab in Canvas.

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. The class will be taught online, but extensive in-person help will be available to students through online office hours and optional help sessions which will be conducted via video conference.
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.

Textbook

The main 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/
If you'd like a hard copy, you can order one through Amazon for around $20 (see the same link above). To download the book for free, you will have to "Add Ebook to Cart" and "Check Out" but you can choose how much you pay (i.e. you can pay nothing if you want). There is a standard pdf version and a "Tablet-Friendly PDF", which is basically a pdf with very small margins that looks better on a tablet (and maybe on a computer too depending on your preferences).
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:

- *The Cartoon Guide to Statistics,* by Larry Gonick and Woollcott Smith
    - This is a somewhat silly perspective on the basic concepts in probability and statistics. As the name suggests, it's full of cartoons. But the explanations of many of the ideas are actually very clear and useful. I have heard that pdfs of this book can be found online. Any version should be fine – don't need to worry if it's the latest one.

- *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.r-project.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.

- *An Introduction to R.*

    - Basic but fairly comprehensive introduction to R. Available 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.

Troubleshooting with R

Other than these books, you will often find answers by searching online (Google an error message you received or search for "histograms in R"). Or you can search for your question on http://stackoverflow.com where it's likely someone has already asked something similar. Obviously, the professors and TAs are here to help as well. I would suggest that you try to figure out things on your own at first. Searching for solutions to error messages isn't the most exciting thing to do, but it's a skill that will serve you well later on, so while you should always feel free to ask us any questions you have, there's a benefit to attempting to figure these things on your own for a bit. When you really are stuck, though, definitely let us know. Just make sure you start working on assignments early enough so that if you do hit a snag, there's time to talk to us.

Class Structure

This course will consist of seven "modules" of material, each lasting two weeks. A module will include recorded video lecture segments, software walkthroughs and reading assignments. A new module will be posted every other Friday and quizzes (to be taken after doing reading and watching lecture segments) for a given module must be taken by the second Monday after posting (i.e. 10 days after a module is posted) at 11pm. Each module will also have a problem set, which will be due the second Wednesday (12 days after the module is posted) at 11pm and a computing assignment (a "lab") which will be due Friday (14 days after the module is posted) by 11pm.

Quizzes
Students will complete online quizzes to assess their understanding of the material after doing readings and viewing pre-recorded lecture videos.
After their due date passes, the quizzes will be closed and students will receive a grade of zero for any quizzes they did not complete.
Students may use notes or other materials during the quizzes but should work alone. We will drop the lowest two quiz grades for each student.
******Students must work entirely by themselves on the quizzes.***
******Collaboration is NOT allowed on quizzes.***

Problem Sets
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 large groups or shared documents is **not** allowed.

Computing Assignments (AKA "Labs")
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.

Each module will include a computing assignment, which we will call "labs". 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 figure something out and copying. All keystrokes must be your own, not copied and pasted from another student's assignment. Sharing answers online or in large groups or shared documents is **not** allowed.

Semester Project
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 project will have several intermediate assignments that are designed to help students end up with a final paper that uses the statistical concepts and computing tools learned in the class to answer a question chosen by each student.

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

Choose a Research Question	(due by Th 2/29 at 11pm)
Gather Data	(due by Th 4/4 at 11pm)
Final Paper	(due by Th 5/2 at 11pm)

The semester project grade will be a combination of the three above project assignments, weighted 10%, 20% and 70%, respectively.

Late Assignments

Late computing assignments and problem sets will be penalized by 10 points for each day late. For example, an assignment would get 10 points off if it's 1 minute late, 20 points off if it's 24 hours and one minute late, etc.
(Note: I will generally make the official Canvas due times at 5 minutes after, e.g. 11:05pm when an assignment is due at 11pm, to account for random glitches and clocks that are slightly off, but Canvas will mark as late anything after that).

Grading

Course grades will be based on the following percentages:

Quizzes:	20%
Computing Activities:	25%
Problem Sets:	25%
Project:	30%

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

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

I will round up half points but not less (for example, 92.5 = A, but 92.49 = A-).

Academic Honesty and Integrity
Issues of academic dishonesty will be taken very seriously and any finding of cheating or other unethical behavior will be pursued to the full extent possible under university policy. Students should consult the University of Texas Honor Code.

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 exams and provide an excuse afterwards will not be granted a makeup and may be given zero percent on the exam in question.

*** If you have a potential issue, let me or a TA know as soon as you can. We will work with you. But it's much more likely we'll be able to find an option that's satisfactory for everyone if we know earlier, rather than if we hear after you've already missed a deadline or fallen way behind in the class. In other words: ***we're here to help, but you'll have to let us know if you need help***.

Special Needs
Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 471-6259, https://diversity.utexas.edu/disability/
All students needing special accommodations should email the professor as soon as possible at the start of the semester so that we can work out accommodations to help you succeed.

Below is a set of technical support guidelines from UT IT:

Technical Requirements

All students are required to have access to a **laptop** or **desktop computer**, running either Windows or macOS operating systems. Tablets, smartphones, and Microsoft Surface devices are not supported. Your computer should meet the following requirements:

- Modern and up-to-date operating system (macOS or Windows)
- Browser: Chrome (highly recommended), Safari or Firefox. If using Safari or Firefox, be prepared to download Chrome and use it.
- Internet connection speed: 5 Mbps download speed. Check your speed here.
- Functional webcam and microphone
- Zoom installed and configured

Confirm that your computer is able to stream video by visiting https://www.laits.utexas.edu/tower/tech.php

For the best experience:

- Close all unnecessary browser windows and tabs and programs
  - Streaming music (Pandora, Spotify, etc.)
  - Social media sites
  - YouTube or other video sites
  - Online/Offline Gaming
- Check your computer is free of viruses, malware, and spyware (UT recommendations)
- Clear the browser's cache before class (here's how.)

Technical Support

If you experience a technical problem, click on the "***Online Course Tech Support***" item in the left-side navigation bar. The Online Course Tech Support chatbot, called "LAITS Bot", will assist with technical problems and can escalate your question to a human if it cannot readily answer your question.