GOV 350K: Statistical Analysis in Political Science

Fall 2023, Unique # 38285
TTH 11:00 a.m.-12:30 p.m., PHR 2.116
The University of Texas at Austin

Instructor	Yul Min Park (Department of Government)	Office Hours:	TBA
	E-mail: yulminpark@utexas.edu	Office:	TBA

This course carries the Quantitative Reasoning flag. Quantitative Reasoning courses are designed to equip you with skills that are necessary for understanding the types of quantitative arguments you will regularly encounter in your adult and professional life. You should therefore expect a substantial portion of your grade to come from your use of quantitative skills to analyze real-world problems.

Course Description

This course introduces the necessary skills to gain important knowledge from data within the field of political science, and more broadly, social sciences. The course explores fundamental techniques for describing and analyzing data using statistical methods—more specifically, for obtaining data, manipulating them into a useful format, analyzing them to produce broader insights, and presenting (or visualizing) these findings effectively. These skills will be valuable not just in future coursework but also in a wide range of professional contexts.

For data analysis, we will mainly use the R statistical programming environment. R is used widely in academia, business, and various other contexts, and includes tools for both simple and more sophisticated methods. This course provides in-class lab sessions to help you become better equipped with the ability to use R in a range of statistical analyses. The course has no prerequisites beyond basic middle and high school math (it is not a math class *per se*) but will be somewhat mathematical; it will require using mathematical reasoning to understand the main concepts of the course. You should not worry that you are not a "math person." If you know +, -, ×, and ÷, the main requirement is a willingness to learn the rest.

Readings

The main (required) text for the course is:

- ***OpenIntro Statistics (4th Edition)***, by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr (available as a free pdf here: https://www.openintro.org/book/os/)

All additional materials will be posted online to Canvas. Any assigned reading will be announced to students in advance.

Course Evaluation

Your grade will be determined as follows:

1. 5 Problem Sets (take-home; 8% each)	40%
2. Midterm Exam	20%
3. Final Project (5% Research Question, 5% Data Collection, 10% Data Visualization, 20% Final Paper)	40%

All items listed above are detailed below on the next page. Grades will not be rounded up at the end of the semester. I will grade using the following cutoffs:

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-).

1. Problem Sets

Problem Sets will be handed out, in class, approximately one week prior to their respective due dates, and are to be submitted in class ***before*** class begins. They will include assigned problems as well as coding exercises. Students can work together in small groups, but I trust that you all know the difference between helping each other figure something out and copying. Sharing answers in large groups or shared documents is **not** allowed.

2. Midterm Exam

The Midterm Exam will be held during class time on ***Thursday, October 19***. It will look like the Problem Sets but more comprehensive. The exam will be closed book.

3. Final Project

Each student will work on a semester-long project in which they pose their own research question, gather and analyze data, and provide discussions about their final output. The project will have three intermediate assignments that are designed to help students successfully write a final paper, which will include the statistical concepts and computing tools learned in the class to answer a question of their own.

Submission dates for project assignments (all to be submitted through Canvas) are as follows:

- (1) Research Question (5%) (due by Thursday, October 5, at 11 p.m.)
- (2) Data Collection (5%) (due by Thursday, October 26, at 11 p.m.)
- (3) Data Visualization (10%) (due by Thursday, November 16, at 11 p.m.)
- (4) Final Paper (20%) (due by Thursday, December 7, at 11 p.m.)

Software Requirements

This course will use R statistical software and RStudio, both of which can be downloaded for free at:

- R: http://www.r-project.org
- RStudio: https://www.rstudio.com/products/rstudio/download/#download

Nota Bene: Learning to code often can be a frustrating process, especially when you are new to the process. I will provide instructions for installing and using both programs in class, and I encourage you to reach out if you have any questions. I also encourage you to budget a significant amount of time for the Problem Sets (that include coding exercises).

You will often find answers by searching online. Below are a few resources for R coding:

- Stack Overflow: http://stackoverflow.com
- https://www.r-bloggers.com/
- RStudio Cheat Sheets (uploaded to Canvas under “Files”)

Additionally, the open secret among anyone who codes: typing an error message into Google often will lead you to a solution. If you are running into a problem, the odds are that you are not the first, and someone else has sought for help online. Stack Overflow (linked above) is an especially useful website. And obviously, I am here to help as well.

Optional Readings

Optional texts for students who want multiple references are as follows:

- *The Cartoon Guide to Statistics*, by Larry Gonick and Woollcott Smith
    - 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. Any version should be fine.
- *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
    - 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.) Available as free pdf here: https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf
- *An Introduction to R*, by William N. Venables, David M. Smith, and the R Core Team
    - Basic but fairly comprehensive introduction to R. It's pretty dry (reads more like an instruction manual than a textbook), but that makes it good for reference. Available as free pdf here: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
- *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"). 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. I may assign some parts of this book on general R and best practices.

Course Policies

Course Syllabus

This syllabus is subject to change or adjustment by the instructor. Consistent class attendance is the best way to ensure that no important announcements are missed.

Technology

Students may not use cell phones in class, except for when in need of calculators. Computers can only be used in class for academic purposes. They should not be used for e-mail, social networking, etc.

Assignment Submission

Problem Sets must be submitted ***in person, as hard copies, before class begins*** on the specified due dates. Final Project assignments must be submitted ***through Canvas, as Word documents, by 11 p.m.*** on the due dates. It is your responsibility to ensure that your file has been correctly submitted by the deadlines.

Late Assignments or Make-up Exams

In the absence of extreme circumstances, all assignments (Problem Sets + Final Project assignments) must be submitted by the due date. Assignments turned in late will be penalized by *10 points per day* over the deadline. For example, a Final Project assignment would get 10 points off if it is 1 minute late, 20 points off if it is 24 hours and one minute late, etc. For Problem Sets, an assignment would get 10 points off if it is 10 minutes late (after class has begun), 20 points off if it is 24 hours and one minute late, etc.

Make-up exams and late assignments will be allowed only if you provide proper documentation (such as a letter from your doctor or the health services center in case of illness, or a letter from the appropriate university authority for official university events or activities). Please check with the instructor to see what constitutes an approved absence. **There will be no deadline extensions without <u>prior</u> written consent from the instructor.** 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. For those with approved absence for the Midterm Exam, I will schedule a *single time* at the end of the semester for a make-up exam. I will not schedule make-up during the semester.

Grade Appeals

You have the right to appeal one of the instructor's grading decisions. Any appeals about grades must be initiated in writing via e-mail, containing explanation of why the decisions behind the assignment of your grade should be revisited. Appeals are due within 10 days of receiving the grade in question. After that period, all grades will be considered final. I will regrade the assignment or exam without referencing the original grade. Your grade may go up or down.

Attendance

This class will not have an attendance policy. However, you will be responsible for keeping up with assignments and contents of the course. Keeping up is especially important in a statistics course, whose material be cumulative over the semester. If you miss class, you should obtain notes from a classmate.

Sharing of Course Material is Prohibited

No material used in this class, including, but not limited to, lecture slides, exams, papers, Problem Sets, in-class materials, and videos, may be shared online or with anyone outside of the class unless you have my explicit, written permission. Unauthorized sharing of materials promotes cheating. It is a violation of the University's Student Honor Code and an act of academic dishonesty. I am well aware of the sites used for sharing materials, and any materials found online that are associated with you, or any suspected unauthorized sharing of materials, will be reported to Student Conduct and Academic Integrity in the Office of the Dean of Students. These reports can result in sanctions, including failure in the course.

Class Recording

Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

Academic Integrity and Scholastic Honesty

"Scholastic dishonesty" "includes, but is not limited to, cheating, plagiarism, collusion, falsifying academic records, misrepresenting facts, and any act designed to give unfair academic advantage to the student (such as, but not limited to, submission of essentially the same written assignment for two courses without the prior permission of the instructor), or the attempt to commit such an act." Section 11-802 (b), Institutional Rules on Student Services and Activities.

I take such academic dishonesty very seriously. If you use words or ideas that are not your own (or that you have used in a previous class), you must cite your sources. Otherwise, you will be guilty of plagiarism and subject to academic disciplinary action, including, but not limited to, failure of the assignment and course. You are responsible for understanding the University Academic Honesty and the University Honor Code. If you have any questions about what constitutes scholastic dishonesty, you should consult with me and the following website: http://deanofstudents.utexas.edu/conduct/.

Special Needs

UT respects and welcomes students of all backgrounds, identities, and abilities. If there are circumstances that make your learning environment and activities difficult, if you have medical information that you need to share with me, or if you need specific arrangements in case the building needs to be evacuated, please notify the instructor before the need arises. I am committed to creating an effective learning environment for all students, but I can only do so if you discuss your needs with me as early as possible. You may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 512-471-6259, http://diversity.utexas.edu/disability/.

Accommodations for Religious Holidays

As per the University's policy, you must notify the instructor 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, an 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.

Drop Policy

If you want to drop a class after the 12th class day, you will need to execute a Q drop before its deadline, which typically occurs near the middle of the semester. See: https://ugs.utexas.edu/vick/academic.

Emergency Evacuation Policy

In the event of a fire or other emergency, it may be necessary to evacuate a building rapidly. Once evacuated, no one may re-enter the building unless given instructions by the following: Austin Fire Department, University of Texas at Austin Police Department, or Fire Prevention Services office.

Students should familiarize themselves with all the exit doors of each room and building they occupy at the university, and should remember that the nearest exit routes may not be the same as the way they typically enter buildings. Students requiring assistance in evacuation shall inform the instructor in writing during the first week of class. Information regarding emergency evacuation routes and emergency procedures can be found at: http://www.utexas.edu/emergency. Behavior Concerns Advice Line (BCAL): 512-232-5050, http://operations.utexas.edu/units/csas/bcal.php.

Course Schedule: *(Note: This is subject to change as we proceed through the semester.)*

Wk	Day	Subject	Readings	Assignments Due
1	8/22	Course Introduction: What is Data?		
	8/24	Introduction to Probability and Statistics	1.1; p 22	
2	8/29	Samples and Sampling Methods	1.3 (pp 23-28)	
	8/31	Samples and Sampling Methods *(continued)*	*(continued)*	
3	9/5	Descriptive Statistics	1.2; 2.1; 2.2 (*skim*)	
	9/7	***Lab Session 1**: Introduction to R*	"Installing R" (Canvas)	**Problem Set 1** *(submit in class)*
4	9/12	Probability and Probability Distributions	3.1; 3.4-3.5	
	9/14	Binomial, Uniform, Normal Distribution	4.1; 4.2 (*skim)*; 4.3	
5	9/19	Normal Distribution *(continued)*, Sampling Distributions	*(continued)*	
	9/21	***Lab Session 2**: How to Load Data into R*		
6	9/26	Estimation: (1) Introduction Estimation: (2) Estimating Proportions	5.1-5.2 6.1.1-6.1.2; 6.1.5	
	9/28	Estimation: (3) Estimating Means ***Lab Session 3**: R for Estimation*	7.1.1-7.1.4	**Problem Set 2** *(submit in class)*
7	10/3	Hypothesis Testing: (1) Introduction	5.3	
	10/5	Hypothesis Testing: (2) Single Group	6.1.3; 7.1.5 6.1.1-6.1.2 *(review)*; 7.1.1-1.7.4 (review)	**Project Assignment #1** *(submit on Canvas)*
8	10/10	Hypothesis Testing: (3) Two Groups	6.2; 7.3.1-7.3.2	
	10/12	Hypothesis Testing: (4) Categorical Variables	2.2-2.3 *(skim)*; 6.3-6.4	**Problem Set 3**
9	10/17	***Lab Session 4**: R for Hypothesis Testing*		
	10/19	**Midterm Exam**		
10	10/24	Linear Regression: (1) Introduction, Estimation	8.1	
	10/26	Linear Regression: (2) R-Squared, Confidence Intervals for Coefficient Estimates	8.2; 8.4	**Project Assignment #2** *(submit on Canvas)*
11	10/31	Linear Regression: (3) Hypothesis Testing	*(continued)*	
	11/2	***Lab Session 5**: R for Regression*		**Problem Set 4** *(submit in class)*
12	11/7	Multivariate Regression: (1) Introduction	9.1	
	11/9	Multivariate Regression: (2) Interpretations	*(continued)*	
13	11/14	Linear Regression Extensions (e.g., dummy variable regressions, interaction effects)		
	11/16	Discuss Final Projects ***Lab Session 6**: R for Multivariate Regression*		**Project Assignment #3** *(submit on Canvas)*
14	11/21	*Fall Break/Thanksgiving (No Class)*		
	11/23	*Fall Break/Thanksgiving (No Class)*		
15	11/28	Logistic Regression ***Lab Session 7**: R for Logistic Regression*	9.5; 1.4	
	11/30	Conceptualization and Measurement Experiments		**Problem Set 5** *(submit in class)*
				Final Paper DUE 12/7 *(submit on Canvas)*