Syllabus

Course Information

Course Number: *Bush 689*

Course Title: *Quantitative Methods for Public Management I*

Section: *600*

Times: *Mondays: 1:10-4pm*

Location: *Allen 1063*

Credit Hours: *3*

Instructor Details

Instructor: *Mike Denly, Ph.D.*

Office: *Allen 1035*

Phone: *979-458-8020*

E-Mail: *mdenly@tamu.edu*

Office Hours: *Mon 10:30-11:30am; Wed 10:30am-12:30pm; Fri 2-3pm; book online at denly.youcanbook.me*

Course Description

We live in an era of data-driven decision-making, and quantitative evidence is fundamental to inform sound governmental policies on both domestic and international issues. This course provides an introduction to quantitative methods for public policy, equipping students with fundamental skills to critically consume and analyze quantitative evidence in international development, public management, and security. Upon successful completion of the course, students will be able to: conduct basic descriptive inference, statistical inference, linear regression, and prediction, using specialized statistical software and, to some extent, MS Excel; and explain the basics of causal inference, using causal diagrams, randomized experiments, and other quasi-experimental methods.

Course Prerequisites

Graduate classification and approval of MPSA or MPIA director.

Special Course Designation

None.

Course Learning Outcomes

- conduct basic descriptive inference, data visualizations, statistical inference, linear regression, and prediction, using the statistical software program R and, to some extent, Microsoft Excel;
- explain the basics of causal inference, using causal diagrams, randomized experiments, and other quasi-experimental methods.

Textbook and Resource Materials

The course does *not* have any required textbooks. The free, online video lectures, which can be found below, cover all of the materials that students need to excel in the course. However, most of the material comes from the

textbooks below, so students may optionally purchase or access the textbooks below if they wish:

- Bueno de Mesquita, Ethan, and Anthony Fowler. 2022. *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press.
- Cumming, Geoff, and Robert Calin-Jageman. 2024. *Introduction to the New Statistics: Estimation, Open Science, and Beyond*. Second Edition. New York: Routledge.
- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2022. *Regression and Other Stories*. Cambridge: Cambridge University Press.
- Gerring, John, and Dino Christenson. 2017. *Applied Social Science Methodology: An Introductory Guide*. Cambridge: Cambridge University Press.
- Huntington-Klein, Nick. 2025. *The Effect: An Introduction to Research Design and Causality*. Second Edition. CRC Press.
- Imai, Kosuke, and Nora Webb Williams. 2022. *Quantitative Social Science: An Introduction in Tidyverse*. Princeton: Princeton University Press.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2023. *An Introduction to Statistical Learning: with Applications in R*. Second Edition. The Netherlands: Springer.
- King, Gary, Robert Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Li, Quan. 2018. *Using R for Data Analysis in Social Sciences: A Research Project-Oriented Approach.* Oxford: Oxford University Press.
- Wickley, Hadley, Mine Cetankanya-Rundell, and Garett Grolemund. 2023. *R for Data Science: Import, Tidy, Transform, and Model Data*. Sebastopol, CA: O'Reilly Media.

Computer and Software Requirements

All students must have access to a computer, not a Chromebook, smartphone, or tablet with limited functionality. Students must also bring a computer to every class. The computer must have R and Microsoft Excel installed. Google Sheets or equivalent programs will not be sufficient to meet course requirements. Prior knowledge of any of these software programs is not required. R is open-source, and students have free access to all Microsoft Office programs, including Excel, through their university email accounts.

Grading Policy

The grading scale below already incorporates very generous grade rounding. Accordingly, there will be no additional rounding of grades under any circumstance.

- >89.5 (A)
- 79.5-89.5 (B)
- 69.5-79.5 (C)
- 59.5-69.5 (D)

- $<$ 59.5 (F)

If students would like to appeal a grade on any assignment, they must make the request to the professor in writing, over email, within 5 days of receiving your grade. In the grade appeal email, students must specify the reason(s) why they think the professor or grader misgraded the assignment. Acceptable reasons include those pertaining to the concepts and material covered during the course. The professor will not consider requests for grade changes that are not germane to the course.

Final grading weights for the individual assignments will be as follows:

- Exams: 30%
- Quizzes: 25%
- Dream Job Assignment: 25%
- In-Class Group Work Other than the Dream Job: 10%
- Dream Job Presentation: 5%
- Attendance: 5%

Exams – Exams will be cumulative and have open-ended questions and answers. Frequently, the professor copies questions from Exam 1 on Exam 2, as well as questions from Exam 2 on Exam 3, to ensure that students retain the information that they learned. There will be no multiple choice or use of cheat sheets or notes on any exam. Per Student Rule 7, students will also only be allowed to make-up exams in the case of university-excused absences, and I will not provide make-up exams for absences that are not university-approved.

Quizzes – As long as students take their own handwritten notes on the online video lectures, they may use those handwritten notes while taking the quiz to start each class. Students may only use a tablet or other computing device for their notes if they have an university-approved accomodation.

Dream Job Assignment – After completing their in-class group work assignments, students must start on their Dream Job assignments. The latter will be individual assignments—i.e., there will no group Dream Jobs. Relatedly, students are not allowed to work on the exact same variable from the exact same dataset for the Dream Job Assignment assignments. If the professor notices that more than one student has the same Dream Job and uses the same variable, the professor will contact the students to let them know about the conflict and ask them to choose different variables.

By the same token, students are encouraged to consult their classmates as well as the professor for any questions on their Dream Job assignment. Indeed, the purpose of starting the Dream Job assignment during class is to ensure that students have a support system to adequately complete the assignments. Often, students will have adequate class time to complete the Dream Job assignments during class. If students need more time to complete the Dream Job assignments, they will have until approximately 1 hour before the next class to submit the assignment. If students submit the assignment within three days (72 hours) of the end their last class, they will receive two extra points on the respective assignment. When students turn in their assignments within the 72-hour, please add a note to the assignment on Canvas when submitting. This way, the grader will not forget to assign the points.

To ensure that students truly learn the Dream Job Assignment skills, which are largely based on the skills needed to obtain an entry-level position at the State Department or World Bank, the Dream Job assignments will be cumulative. That means a few things:

1. Question numbers will increase from week to week. For example, if Dream Job Assignment 1 ends at question 4, Dream Job Assignment 2 will start at question 5.
2. Students will be responsible for making the corrections to the previously graded Dream Job assignments. When students fail to make the corrections for even one question, they will receive a 15-point penalty on the new Dream Job assignment.
3. Students must submit their entire Dream Job Assignment every week *in one file*. The entire Dream Job Assignment entails: (i) all of the questions from previous weeks; (ii) the (corrected) answers to those questions from previous weeks; (iii) the questions for the respective week; and (iv) the answers to those questions.

Please also ensure that every answer has a complete sentence. In the working world, it would never be acceptable turn in anything formal without complete sentences, so it will be no different here.

The Dream Job Assignment, which will be graded weekly, will account for 25% of students' final grades in the course. The professor will drop students' two lowest weekly Dream Job assignment scores when calculating the respective portion of the final grade. The professor will not drop additional Dream Job assignment grades for any reason. Students who are facing difficulties completing the Dream Job are encouraged to speak to their classmates, the professor, and book office hours, as appropriate. We are all in this together.

Dream Job Presentation – Students will be randomly assigned to present on their Dream Job Assignment progress at one point during the semester. When assessing presentations, the professor will consider the following criteria:

1. *Respect for the 10-minute time limit*. Indeed, all presentations will be timed, and I will sharply end all presentations at 10 minutes. The 10-minute time limit will be in place because: (i) students need to learn how to present their work concisely; (ii) most presentations in the real world have time limits; and (iii) there are many students in this class, so we don't have more time than 10 minutes for each student presentation.
2. *Coverage of the respective material*. That includes all material up to that respective class. Of course, students who are randomly assigned to later classes will have to cover more material and make decisions about which things to mention in passing and other things to cover in more detail.
3. *Aesthetics of the slides*. Pictures and examples often help make points clearly. Slides should only have a few points on them. Too many words is also distracting to the reader.
4. *Ability to present without having to resort to reading the slides*. Reading your slides shows to the audience that you have not spent enough time preparing. Respect your audience and prepare for the presentation.

Students must submit their presentations on Canvas prior to the class when they are presenting. This way, we do not waste time getting the presentations set up during class. Of course, students who present toward the beginning of the semester will necessarily have shorter presentations than those present later in the semester.

In-Class Group Work Other than the Dream Job – Students will work in groups of two for in-class group work. Students will have a different partner for each class, and that partner will be randomly assigned. This way, students gain experience working with a variety of different people with varying skill levels and working styles. There are two additional ideas behind the random assignment: (1) to avoid selection problems in which better students leave the others behind; and (2) to better prepare students for the working world, where most junior employees do not get to pick their colleagues.

Both students in the pair must submit their assignments on Canvas prior to the end of class. Prior to students learning Quarto (i.e., until Class 2), students must submit an R script and a Word document with screenshots. After learning Quarto (i.e., Class 3 going forward), a only PDF file with the questions repeated, answers in text, and code snippets will be sufficient. Students who do not submit their in-class group assignments prior to the end of class will receive a zero on the respective assignment. Both students will receive the same grade on the assignment, so please both submit the same file.

Graded Attendance – Attending class is crucial for learning the material, attendance will be worth 5% of the overall grade. There are 14 classes, so missing each class will result in 7% loss toward students attendance grades. For example, if a student misses one class all semester, they will be receive an attendance grade of 93%.

Late Work Policy

Unless students receive prior approval, the professor will not accept late assignments. That is, students who submit a deliverable after the established deadline will receive a zero. The only exceptions will be make-up for a university-approved absence that is consistent with Student Rule 7.

Course Overview

Class	Topic	Monday Class	Due
1	Intro to data, stats, and social science	1/12/2026	Dream Job 0
	MLK Day (holiday)	1/19/2026	
	Snow (class cancelled)	1/26/2026	
2	Descriptive stats, visualization, and intro to R	2/2/2026	Dream Job 1
3	Probability distributions & Quarto	2/9/2026	Dream Job 2
4	Statistical inference	2/16/2026	Dream Job 3
5	Exam 1	2/23/2026	Dream Job 4
6	Bivariate relationships	3/2/2026	None
	Spring break	3/9/2026	
7	Directed Acyclic Graphs (DAGs)	3/16/2026	Dream Job 5
8	Linear regression 1	3/23/2026	Dream Job 6
9	Linear regression 2	3/30/2026	Dream Job 7
10	Exam 2	4/6/2026	Dream Job 8
11	Prediction and classification	4/13/2026	None
12	Experiments and ethics	4/20/2026	CITI Course & Dream Job 9
13	Natural and quasi-experiments	4/27/2026	Dream Job 10
14	Final Exam	TBD	None

Class 1: Introduction to Data, Statistics, and Social Science

Dream Job Assignment (Part 0 [Pre-Course Assignment])

1. Imagine that you have received your dream job after finishing your degree at Texas A&M. What's that dream job?
2. What are the types of problems that you would need to tackle as part of your job? What kinds of information, data, or analyses would you need in order to tackle those problems?
3. After watching the required videos below that explain different data types, download a panel dataset from

the Internet that corresponds to your dream job and has at least 500 observations. State the name of the dataset, declare the source of the dataset, and select only one variable of focus—i.e., delete the rest of the variables if what you download has multiple variables. Submit the raw data as an Excel or CSV file with your assignment and save the file as "raw_data.xlsx" (for Excel files) or "raw_data.csv" (for CSV files). Then, book a meeting with me during my office hours or additional office hours prior to the first class. During the meeting, I can help you reshape/restructure the data or a find a new dataset, if necessary. If you do not meet with me prior to the first class about your dataset, I will be unable to provide you with credit for the homework. If you need help finding a panel dataset online, here are some places—i.e., among MILLIONS of others that you can Google:

- The World Bank's World Development Indicators
- United Nations Conference on Trade and Development Data Center
- United Nations Statistics
- WomanStats.org (started and maintained by Prof. Valerie Hudson)

Note: Please submit a Word document or `.pdf` file with your answers to accompany the Excel or `.csv` file with your data. Please do not write the answers to the above questions on the Excel or `.csv` file, and please do not use a Google Doc for this assignment or any other throughout the course.

Required Videos

- Part 1: Why Am I Taking This Course?
  - Course overview
  - Why statistics?
  - Interview with Major Christopher Swain, US Air Force Intelligence Services
  - Interview with Pascale Schnitzer, Senior Economist at the World Bank
- Part 2: Introduction to Data and Social Science
  - What distinguishes social science from casual conversation?
  - The four characteristics of social scientific research
  - Quantitative vs qualitative research
  - Variable types: binary, continuous, categorical, bounded, etc.
  - Units of analysis and other essential vocabulary
  - Data types: cross-sectional vs time-series vs panel data AND long vs wide

Class

- Part 1: Course Overview
  - Professor introduction
  - Student introductions
  - Syllabus and class expectations

- Part 2: Microsoft Excel Training
  - Saving and file types (e.g., `.xlsx` vs. `.csv`)
  - Inspecting and filtering data
  - Merging cells, wrapping text, and freezing panes
  - Sorting data
  - Pivot tables
  - Missing data
  - Making graphs and troubleshooting
  - Paste special, transposing, formatting, and selecting cells
  - Preparing files for analysis
  - Identifying and creating unique identifiers
  - Relative and absolute cell referencing
  - Basic formulas (IF, SUM, AVERAGE)
  - VLOOKUP

Dream Job Assignment (Part 1)

4. What is the unit of analysis for your Dream Job dataset? What variable(s) identify that unit of analysis?
5. Is the dataset in long or wide format? How do you know?
6. What are the summary statistics for your variable, including the mean? (Hint: if there are missing values, you may need to filter them out.)
7. Is your variable continous, bounded, binary, or categorical ordered? Justify that answer based on the summary statistics.
8. Make a pivot table to summarize your panel dataset into a cross-section of the mean of all periods per unit. For example, if your dataset has different values for France and Canada across different years, you would want to produce the means for France and Canada for all years. Show a screenshot of the pivot table.
9. Produce a bar graph that shows those values across units *sorted* by the average value. If your dataset has too many cross-sectional units to fit into one bar graph, produce the bar graph for just a few of the cross-sectional units. Show the figure.
10. Make a pivot table to summarize your panel dataset into a time series of the sum of all units per period. Report the sum value for a specific period in your dataset. Show a screenshot of the time series.
11. Produce a line graph that shows those values *sorted* by period. Show the figure.

Note: Please submit: (1) a Word or `.pdf` file with the answers to the above questions as well as your graph screenshots; and (2) the Excel file that you used for your calculations, showing the relevant graphs. Otherwise, I will be unable to provide a grade.

Optional Reading

- Bueno de Mesquita & Fowler: Chapter 1 (Pages 1-9)
- Gerring & Christenson: Chapter 1 (Pages 3-7); Chapter 4 (Pages 47-50)
- Huntington-Klein: Sections 3.1-3.2
- Imai & Webb Williams: Sections 1.1-1.2 (Pages 1-8).
- King, Keohane & Verba: Section 1.1 (Pages 3-12)
- Li: Introduction (Pages xv-xvii)

Class 2: Descriptive Statistics, Visualization, and Intro to R

Required Videos

- Why use R, not Excel? [from Andrew Heiss; note Quarto, which we'll learn next, is R Markdown's successor]
- Installing R and R Studio
- Panes in R Studio
- Setting the working directory (Windows/Mac)
- Creating a new project
- Vectors, sequences, and data classes/types
- Data frames, importing data, and installing packages
- Descriptive statistics
- Better understanding the mean, standard deviation, and variance
- Eliminating missing data correctly and piping
- Making nice tables with `modelsummary`
- Making nice graphs in `ggplot2`
- Reshaping/pivoting data wide and long, and dropping variables
- Summarizing/collapsing data
- (Conditionally) creating variables

Class

- Group Work: Civil Wars and the Worldwide Governance Indicators Redux
- Dream Job presentations

Dream Job Assignment (Part 2)

12. Does your variable have missing values? If so, how many? Show screenshots of both your code and output in R.

13. If your variable does have any missing values, show how you would go about properly removing those missing values just for your variable—i.e., without potentially deleting missing values from other variables.
14. What are the summary statistics for your variable, including the mean, standard deviation, and variance? Show screenshots of both your code and output in R. (Hint: if there are missing values, you may need to remove them.)
15. Calculate the mean for the first and second half of your respective time periods. Show screenshots of both your code and output in R.
16. Use `ggplot2` to produce a labeled line graph that shows the mean values for your original, larger data frame by time period. Show screenshots of both your code and output in R.
17. Remove the extra data frames that you created from the last two questions, only keeping your original data frame.

Note: Please submit (1) a Word or `.pdf` file with the answers to the above questions; (2) your Excel or CSV file with the data; and (3) your R script. Otherwise, I will be unable to provide a grade.

Optional Reading

- Gerring & Christenson: Chapter 18
- Huntington-Klein: Sections 3.3-3.4
- Imai & Webb Williams: Sections 1.3.1-1.3.5 (Pages 8-17)
- Gerring, John. 2012. "Mere Description." *British Journal of Political Science* (42)4: 721-746.

Class 3: Probability Distributions and Learning Quarto

Required Videos

- Part 1: Learning and Installing Quarto
  - Quarto tutorial
- Part 2: Probability Distributions
  - Random variables, frequency distributions, and probability distributions
  - Frequency meets probability: the Law of Large Numbers
  - The Central Limit Theorem and normal distribution
  - The standard normal distribution
  - Understanding z-scores
  - The Bernoulli distribution
  - Assessing data "normality" (boxplots, histograms, Q-Q plots, density & counts)
  - Deflating (adjusting the data for inflation)
  - Variable transformations: natural logs

Class

- Group Work: Quarto and the normality of the Worldwide Governance Indicators
- Dream Job presentations

Dream Job Assignment (Part 3)

18. Invent hypothetical but logical values for μ and σ, then calculate the z-score for $\bar{X}$. Show your work.
19. Interpret the z-score-score. What is it telling you?
20. If your variable is continous or bounded, make a Q-Q plot for it in R using `ggplot2`, and interpret the Q-Q plot. If your variable is binary or categorical, make a bar graph for it in R using `ggplot2`, and provide a brief overview of the results.
21. If your variable is continuous or bounded, make a density plot for it in R using `ggplot2`, and interpret whether the data appear to be skewed and have kurtosis. If your variable is binary or categorical, make a bar graph of expected vs observed counts in R using `ggplot2`, and interpret the results.
22. Based on your responses to the previous two questions, state outright whether it is necessary to deflate your and/or take the log. If so, re-run the plots according to the deflated and/or logged values and asssess whether the data are now more normally distributed.

Note: Please submit your homework as a Quarto `.qmd` file *and* its accompanying `.pdf` file, showing all code, tables, and figures. I will not accept homework submitted in a regular R script and MS Word documents. Given that you will need to submit all previous (and corrected) Dream Job Assignments along with this one, you will need to enter the results from the previous assignments into Quarto as well. Recall that if you do not previously have LaTeX on your computer, you will need to install the `tinytex` package in R in order to produce the relevant `.pdf` file.

Optional Reading

- Gerring & Christenson: Chapter 19
- Imai & Webb Williams: 1.3.6-1.3.12 (Pages 18-33); 6.3-6.4
- Gelman, Hill & Vehtari: Section 3.5

Class 4: Statistical Inference

Required Videos

- What is statistical inference?
- The sampling distribution and theory of sampling
- Different types of samples
- Reliability, validity, and components of estimators
- Precision: standard errors
- Precision: margin of error

- Precision: confidence intervals
- Precision: p-values
- Null Hypothesis Significance Testing (NHST)
- Statistical power (Part 1)
- Example: test scores
- Example: polling

Class

- Group work: snowstorms in Narnia and eras of Texas A&M Aggie Football
- Dream Job presentations
- Exam review

Dream Job Assignment (Part 4)

23. Based on the values for μ and σ that you invented for last week's assignment, test the hypothesis that the sample and population means are the same. Assume that the population size is twice your sample size.

Note: Please submit your homework as a Quarto `.qmd` file and its accompanying `.pdf` file, showing all code, tables, and figures. I will not accept homework submitted in a regular R script and MS Word documents. **Given the exam and the need for an answer key prior to the exam, I will not accept late assignments—not even by one minute.**

Optional Reading

- Gerring & Christenson: Chapter 20 (pages 302-309).
- Bueno de Mesquita & Fowler: Chapter 6 (pages 94-105).
- Li: Chapter 3 (pages 94-101)
- Imai and Webb Williams: Section 7.2
- Gelman, Andrew. 2023. "What Is a Standard Error?" *Journal of Econometrics* 237(105516): 1-2.

Class 5: Exam 1

Class

- Exam

Class 6: Bivariate Relationships

Required Videos

- One-sample t-test
- Cross-tabs, merging, and country codes

- Covariance and correlation
- Two-sample difference-in-means/t-tests
- Statistical power (part 2)

Class

- Group work: democracies, autocracies, and transparency
- Dream Job presentations

Dream Job Assignment (Part 5)

24. Subset the data to keep only two years or time periods.
25. Test the hypothesis that the distribution of the smaller, subsetted data is the same as the original data at the 95% confidence level. Make sure to explain your results.
26. Pick a new variable from a dataset on the Internet that (i) shares the same unit of analysis (i.e., panel structure) as your original variable; (ii) you think helps explain your original variable; and (iii) is not a subset or related to the construction of your original variable. Then, (iv) bring that variable into R. Also, (v) state the source of the data; and (vi) explain in 2-4 sentences why you think variable is a good independent variable to explain your original variable.
27. Merge the new dataset with your larger one, ensuring that everything merges in. (Hint: see my blog post.)
28. Test if your original variable is correlated with the new variable that you imported. Explain your results.
29. Run a t-test of the two variables and interpret the results.
30. Calculate the statistical power of your last test, using your sample data to calculate Cohen's d. Interpret your result.

Optional Reading

- Gerring & Christenson: Chapter 20 (pages 309-312) and Chapter 21
- Bueno de Mesquita & Fowler: Chapter 2
- Li: Chapter 3 (Pages 116-127)
- Imai & Webb Williams: Section 3.6

Class 7: Directed Acyclic Graphs

Required Videos

- Bivariate one-way and two-way relationships
- Forks: close the back door!
- Chains/mediators: to close or not to close the back door?

- Colliders: keep the back door open!
- Irrelevant variables: there is no back door!
- (Ir)relevant modifiers: no back door but a side door?
- Unobservable/unmeasured variables: beware of hidden doors!
- R: Coding basic DAGs
- R: Capturing the complexity of the social world with more variables

Class

- Relationships between variables
- Group Work: DAG in Texas A&M Football
- Dream Job presentations

Dream Job Assignment (Part 6)

31. Find a third variable from a dataset on the Internet that (i) shares the same unit of analysis (i.e., panel structure) as your original variable; (ii) you think helps explain your original variable; (iii) is not a subset or related to the construction of your original variable or second variable; and (iv) is not a mediator or collider. State the source of the data, explain in 4-7 sentences why you think variable is a good independent variable to explain your original variable. When doing so, ensure to explain why this third variable does not enter as a mediator or collider. Then, (a) draw a DAG using ggdag to show that the third variable is not a mediator or collider; (b) bring that variable into R; and (c) merge that variable into your larger dataset.

Optional Reading

- Denly, Michael, and Graham Goff. 2026. "Aligning Estimands, Conclusions, and Results." Working Paper.
- Huntington-Klein, Nick. 2025. *The Effect: An Introduction to Research Design and Causality*. Second Edition. CRC Press.
  - See Chapters 6-8
- Heiss, Andrew. 2020. "Causal Inference." In *R for Political Data Science: A Practical Guide*, ed. Urdinez, Francisco, and Andrés Cruz. Chapman Hall/CRC chapter 10, pp 235-273.

Class 8: Linear Regression 1

Required Videos

- Overview of linear regression (example: education and income)
- Learning linear regression through James Bond and Rotten Tomatoes
- Creating variables to test the ageism and nostalgia hypotheses
- Adjusting the data for inflation

- Variable transformations: natural logs
- Linear regression with a continuous independent variable
- Linear regression with a binary independent variable
- Presenting results in tables and coefficient plots

Class

- Group Work
- Dream Job presentations

Dream Job Assignment (Part 7)

32. Create a scatterplot with your original variable on the *y*-axis and your second variable on the *x*-axis.
33. Add the line of best fit to the scatter plot. Interpret the results and explain what's special about this line. (Hint: Be sure to mention residuals.)
34. With your original variable as the dependent variable and second variable as the independent variable, run a linear regression.
35. "Tidy" your output from the linear regression that you just ran using the `broom` package.
36. Output the results of your linear regression in a table and a coefficient plot using `modelsummary`.
37. Interpret the linear regression, considering its practical/substantive significance, statistical significance, and R^2 value.

Optional Reading

- Gerring & Christenson: Chapter 22 (pp. 331-343)
- Bueno de Mesquita & Fowler: Chapter 5 (pp. 74-79); Chapter 5 (pp. 105-109)
- Li: Chapter 5
- Imai & Webb Williams: Sections 4.2.1-4.2.3; Section 7.3
- James, Witten, Hastie & Tibshirani: Section 3.1

Class 9: Linear Regression 2

Required Videos

- Multivariate regression, F-statistic, and Adjusted R^2
- Interpreting coefficients in multivariate regression
- How coefficients and standard errors change as you add regression controls
- Gauss-Markov assumption 1: Linearity and additivity (and interactions)
- Gauss-Markov assumption 2: No high/perfect collinearity

- Gauss-Markov assumption 3: Constant error variance/homoskedasticity
- Outliers

Class

- Group Work: Civil Wars and the Worldwide Governance Indicators Redux
- Dream Job presentations
- Exam review

Dream Job Assignment (Part 8)

38. Using your same bivariate linear regression from last week, produce a scatterplot that takes the regression residuals on the *y*-axis and the independent variable of that regression on the *x*-axis. Does the figure suggest that the regression errors are homoscedastic or heteroskedastic?
39. Replicate the regression table from above but considering heteroskedasticity-robust standard errors. Did the regression coefficients change? Did the standard errors change? If so, did they become larger or smaller? Why or why not?
40. Run a multivariate linear regression, using your original variable as the dependent variable, and your two newer variables as your independent variables. Then, interpret the output, including the coefficients, statistical significance, R^2, Adjusted R^2, and F-statistic.
41. Run a new linear regression that controls for both of your independent variables and their interaction term. Compare this new model with the interaction to the model that you ran above without the interaction term. Then, statistically test whether we need the interaction term. Thereafter, interpret only the interaction term in the new regression and produce a coefficient plot using the `interplot` and/or `interflex` package(s). What do we learn?
42. Examine the correlation between your two independent variables. Is there a risk of potentially high collinearity between them? Why would this be a problem? Could you run a regression in the presence of perfect collinearity between some of the independent variables?

Note: Please submit your homework as a Quarto `.qmd` file and its accompanying `.pdf` file, showing all code, tables, and figures. I will not accept homework submitted in a regular R script and MS Word documents. Given the exam and the need for an answer key, I will not accept late homework.

Optional Reading

- Gerring & Christenson: Chapter 22 (pp. 343-352)
- Li: Chapter 6
- James, Witten, Hastie & Tibshirani: Section 3.2

Class 10: Exam 2

Class

- Exam

Class 11: Prediction and Classification

Required Videos

- Within-sample prediction and extrapolation
- (Root) Mean-Squared Error (R/MSE)
- Out-of-sample predictions and overfitting
- Classification
- Linear probability model (LPM)
- Logistic regression: basics
- Logistic regression: prediction
- Logistic regression: prediction assessment (confusion matrix)
- Logistic regression: prediction assessment (precision and recall)
- Logistic regression: coefficients (log odds)
- Logistic regression: coefficients (odds ratios)
- Logistic regression: coefficients (average marginal effects)
- R exercise (part 1): log odds, odds ratios, and average marginal effects
- R exercise (part 2): confusion matrix, precision, and recall

Class

- Review of exam
- Group work
- Dream Job presentations

Dream Job Assignment (Part 9)

45. If your original dependent variable is not binary, create a binary version of it, dichotomizing values by the median. (Hint: think of what we did during the group work assignment on democracy, autocracy, and transparency.) If your original dependent variable is binary, please just say so here and proceed.
46. Create separate random training and test datasets, reserving 25% of your data to the test sample.
47. On the training dataset, take the binary version of your dependent variable and estimate a linear probability model using the same multivariate specification as last week. Tell us what you find in terms of practical/substantive significance, statistical significance, R^2, and Adjusted R^2.
48. Obtain the predictions for the model and ascertain whether all of the predictions make sense.
49. Run a logistic regression model using the exact same specification as above. Can you interpret these coefficients? If so, how?

50. Obtain the odds ratios for the coefficients that you estimated in the previous step, and interpret these odds ratios.
51. Obtain the average marginal effects for the coefficients and interpret them.
52. Make a confusion matrix for your results based on whether the predicted probability is above or below the median predicted probability in the training dataset. Make probability predictions on the test dataset, and build a confusion matrix for your results on the test dataset—just as you did on the training dataset.
53. Build a table that has the precision and the recall scores for your model on the training and the test dataset. What do you observe? Which ones are higher? What value is most important whenever considering the classification accuracy of different models? Why?

Note: Please submit your homework as a Quarto `.qmd` file and its accompanying `.pdf` file, showing all code, tables, and figures. I will not accept homework submitted in a regular R script and MS Word documents.

Optional Reading

- Bueno de Mesquita & Fowler: Pages 79-89.
- James, Witten, Hastie & Tibshirani: Pages 29-31.
- Li: Pages 313-322.
- Imai & Webb Williams: Section 4.1

Class 12: Experiments and Ethics

Required Human Subjects Protection Assignment

: In order to be able to perform any kind of research at the university, you need to take a training course on Human Subjects Data Protection. To do so:

- Click here. Select "register" under Create an account.
- Search for "Texas A&M University" and click on "Continue to step 2".
- Enter your contact information and create your username, password, and security question.
- On question 1, select "Social and Behavioral Research Investigators and Key Personnel". For all other questions, select "Not at this time".
- Subsequently, you will see a button to start the IRB Social Basic Course. Finish the course. Then, provide a PDF of your certificate on Canvas.

Required Videos

- Comparing observational and experimental data
- Experiments' secret sauce: random assignment
- Overcoming the fundamental problem of causal inference
- Potential outcomes framework: basics

- Potential outcomes framework: key assumptions
- Potential outcomes framework: estimands and calculations
- Field experiments
- Lab experiments and ethics
- Survey experiments
- Validity threats: attrition, noncompliance, and spillover (Cohen & Dupas application)
- Internal vs external validity
- R application (Bertrand & Mullainathan)

Class

- Group work
- Dream Job presentations

Dream Job Assignment (Part 10)

54. What is a causal question that you will need to answer as part of your dream job?
55. What would be the ideal field experiment that you would run to be able to answer that question? Why is a field experiment generally the best method to be able to discern a causal effect for your particular question—barring no problems that you will discuss below? Note: your answer can be be unrealistic, especially if you are working on a sensitive topic like crime, corruption, or war.
56. What would be the constraints to performing such an experiment? Hint: you can talk about ethics, resources, external validity, or other things.
57. While the ideal field experiment may not be possible to run, a survey or a lab experiment is likely feasible. Provide a description of either a feasible lab or survey experiment.
58. What are some challenges to inference in that lab or survey experiment? Hint: you can talk about attrition, non-compliance, spillover/interference, Hawthorne effects, demand effects, or other things.
59. Do the above challenges affect your estimand of interest? Explain why or why not.

Optional Reading

- Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4): 991-1013.
- Cohen, Jessica, and Pascaline Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." *Quarterly Journal of Economics* 125(1): 1–45.
- Gomila, Robin. 2021. "Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis." *Journal of Experimental Psychology: General* 150(4): 700-709.

- Bueno de Mesquita & Fowler: Chapter 3
- Gerring & Christenson: Chapter 7 and Chapter 23 (Pages 353-357)
- Imai & Webb Williams: Sections 2.3-2.4
- Huntington-Klein: Chapter 7

Class 13: Natural Experiments and Quasi-Experiments

Required Videos

- Overview of natural experiments and quasi-experiments
- Matching
- Difference-in-differences (dif-in-dif)
- Instrumental variables (IV)
- Regression discontinuity designs (RD/RDD)

Class

- Group work
- Dream Job presentations

Dream Job Assignment (Part 11)

:

60. Go back to the causal question you identified last week. Assume that you cannot run an experiment to address it directly, so you need to find observational data on the cause and the consequence of interest and assess how they correlate with each other. Can you interpret that correlation causally? What potential concerns would you have?
61. What would a data generating process that yields those concerns look like? Characterize that process in the form of a causal diagram.
62. Are measures of potential confounders observable? If so, how would you use regression analysis or matching methods to approximate the causal effect of interest? Can these methods help you tackle your causal question?
63. Can you think of sources of exogenous variation in your treatment of interest? Hint: It could be natural events, the timing of policy choices, discontinuities in assignment of the treatment, etc.
64. Based on that source of exogenous variation, what specific quasi-experimental method could you leverage to tackle your causal question of interest? Hint: think of instrumental variables, regression discontinuity, difference-in-differences, or matching. Explain your design in detail.

Optional Reading

- Gerring & Christenson: Chapter 8 and Chapter 23 (Pages 357-369)
- Lipsky, Ari, and Sander Greenland. 2021. "Causal Directed Acyclic Graphs." *Journal of the American Medical Association* 327(11): 1083-1084.
- Imai & Webb Williams: Section 2.5
- Huntington-Klein: Chapters 5, 16, and 19
- Rohrer, Julia. 2018. "Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data." *Advances in Methods and Practices in Psychological Science* 1(1): 27-42.
- Angrist, Joshua. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80(3): 313-336.
- Galiani, Sebastian, and Ernesto Schargrodsky. 2010. "Property Rights for the Poor: Effects of Land Titling." *Journal of Public Economics* 94(9–10): 700-729.
- Dell, Melissa. 2015. "Trafficking Networks and the Mexican Drug War." *American Economic Review* 105(6): 1738-79.

Final Exam Date: TBD at 10:30am

University Policies

Attendance Policy

The university views class attendance and participation as an individual student responsibility. Students are expected to attend class and to complete all assignments.

Please refer to Student Rule 7 in its entirety for information about excused absences, including definitions, and related documentation and timelines.

Makeup Work Policy

Students will be excused from attending class on the day of a graded activity or when attendance contributes to a student's grade, for the reasons stated in Student Rule 7, or other reason deemed appropriate by the instructor.

Please refer to Student Rule 7 in its entirety for information about makeup work, including definitions, and related documentation and timelines.

Absences related to Title IX of the Education Amendments of 1972 may necessitate a period of more than 30 days for make-up work, and the timeframe for make-up work should be agreed upon by the student and instructor" (Student Rule 7, Section 7.4.1).

"The instructor is under no obligation to provide an opportunity for the student to make up work missed because of an unexcused absence" (Student Rule 7, Section 7.4.2).

Students who request an excused absence are expected to uphold the Aggie Honor Code and Student Conduct Code (See Student Rule 24).

Academic Integrity Statement and Policy

"An Aggie does not lie, cheat or steal, or tolerate those who do."

"Texas A&M University students are responsible for authenticating all work submitted to an instructor. If asked, students must be able to produce proof that the item submitted is indeed the work of that student. Students must keep appropriate records at all times. The inability to authenticate one's work, should the instructor request it, may be sufficient grounds to initiate an academic misconduct case" (Section 20.1.2.3, Student Rule 20).

You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at aggiehonor.tamu.edu.

Notice of Nondiscrimination

Texas A&M University is committed to providing safe and non-discriminatory learning, living, and work environments for all members of the University community. The University provides equal opportunity to all employees, students, applicants for employment or admission, and the public regardless of race, color, sex (including pregnancy and related conditions), religion, national origin, age, disability, genetic information, or veteran status. Texas A&M University will promptly, thoroughly, and fairly investigate and resolve all complaints of discrimination, harassment (including sexual harassment), complicity and related retaliation based on a protected class in accordance with System Regulation 08.01.01, University Rule 08.01.01.M1, Standard Administrative Procedure (SAP) 08.01.01.M1.01, and applicable federal and state laws. In accordance with Title IX and its implementing regulations, Texas A&M does not discriminate on the basis of sex in any educational program or activity, including admissions and employment. The following person has been designated to handle inquiries and complaints regarding the non-discrimination policies: Jennifer M. Smith, TAMU Associate VP & Title IX Coordinator at YMCA Ste 108, College Station, TX 77843, 979-458-8407, or email civilrights@tamu.edu. For other reporting options, visit https://ocrcas.ed.gov/contact-ocr to locate the address and phone number of the office that serves your area, or call 1-800-421-3481.

Civil Rights, Free Speech, and Title IX Policies

Texas A&M University is committed to fostering a learning environment that is safe and productive for all. University policies and federal and state laws prohibit discrimination and harassment based on an individual's race, color, sex, (including pregnancy and related conditions), religion, national origin, age, disability, genetic information, veteran status, or any other legally protected characteristic. This includes forms of sex-based violence, such as sexual assault, sexual harassment, sexual exploitation, dating/domestic violence, and stalking.

Students can report discrimination/harassment, access supportive resources, or learn more about their options for resolving complaints on the University's Civil Rights & Title IX webpage.

Students should be aware that all university employees (except medical or mental health providers) are mandatory reporters, which means that if they observe, experience or become aware of an incident that they reasonably believe to be discrimination/harassment alleged to have been committed by or against a person who was a student or employee at the time of the incident, the employee must report the incident to the university.

Americans with Disabilities Act (ADA) Policy

Texas A&M University is committed to providing equitable access to learning opportunities for all students. If you experience barriers to your education due to a disability or think you may have a disability, please contact the Disability Resources office on your campus (resources listed below) Disabilities may include, but are not limited to attentional, learning, mental health, sensory, physical, or chronic health conditions. All students are encouraged to discuss their disability related needs with Disability Resources and their instructors as soon as possible.

To request academic accommodations, contact the designated ADA office based on your location:

- Texas A&M University, College of Nursing, College of Dentistry, Irma Lerma Rangel College of Pharmacy College Station, College of Medicine, School of Public Health, Institute of Biosciences and Technology, EnMed Program, Bush School in Washington DC, Mays Business School – CityCentre, TAMU Engineering Academies, Texas A&M University Higher Education Center at McAllen and Texas A&M University at Galveston should contact Disability Resources at (979) 845-1637 or disability@tamu.edu.

If you are experiencing difficulties with your approved accommodations, contact the office responsible for approving your accommodations or the Texas A&M ADA Coordinator Julie Kuder at ADA.Coordinator@tamu.edu or (979) 458-8407.

Pregnancy Accommodations

Texas A&M provides reasonable accommodations to students due to pregnancy and/or related conditions, such as childbirth, recovery and lactation. Students should contact the University's Pregnancy Coordinator as soon as they become aware of the need for accommodation. Depending on the circumstances, accommodations could include extended time to complete assignments or exams, changes in course sequence, or modifications to the physical classroom environment. Texas A&M will also allow a voluntary leave of absence, ensure the availability of lactation space, and maintain grievance procedures to provide for the prompt and equitable resolution of complaints of sex discrimination. For information regarding pregnancy accommodations, email TIX.Pregnancy@tamu.edu.

Statement on Mental Health and Wellness

Texas A&M University recognizes that mental health and wellness are critical factors influencing a student's academic success and overall wellbeing. Students are encouraged to engage in healthy self-care practices by utilizing the resources and services available through University Health Services on its mental health webpage. The TELUS Health Student Support app provides access to professional counseling in multiple languages anytime, anywhere by phone or chat, and the 988 Suicide & Crisis Lifeline offers 24-hour emergency support at 988 or 988lifeline.org.

Texas A&M College Station

Students needing a listening ear can contact University Health Services (979.458.4584) 24-hour emergency help is also available through the 988 Suicide & Crisis Lifeline (988) or at 988lifeline.org.

Statement on the Family Educational Rights and Privacy Act (FERPA)

FERPA is a federal law designed to protect the privacy of educational records by limiting access to these records, to establish the right of students to inspect and review their educational records and to provide guidelines for the correction of inaccurate and misleading data through informal and formal hearings. Currently enrolled students wishing to withhold any or all directory information items can do so within howdy.tamu.edu using the Directory Information Withholding Form. The complete FERPA Notice To Students and the student records policy is available on the Office of the Registrar webpage. Items that can never be identified as public information are a student's social security number, citizenship, gender, grades, GPR or class schedule. All efforts will be made in this class to protect your privacy and to ensure confidential treatment of information associated with or generated by your participation in the class.

Directory items include name, UIN, local address, permanent address, email address, local telephone number, permanent telephone number, dates of attendance, program of study (college, major, campus), classification, previous

institutions attended, degrees honors and awards received, participation in officially recognized activities and sports, medical residence location and medical residence specialization.

Free Speech and Civil Discourse

Texas A&M recognizes that the pursuit of truth through open and robust discourse is critical to academic inquiry. However, as a community of scholars, the university has an aspirational expectation that such discourse will be conducted in accordance with Aggie Core Values. In this "marketplace of ideas," we encourage civil dialogue creating an environment that allows individuals to express their ideas and to have their ideas challenged in respectful and responsible ways. Students can learn more about Freedom of Expression and Free Speech on the University's website about the First Amendment.

Artificial Intelligence Statement

Generative Artificial Intelligence (AI) text generators and natural language processing tools, including but not limited to ChatGPT and Claude, are explicitly prohibited for quizzes and exams in this course. The professor also highly discourages the use of these tools to complete in-class group work and Dream Job assignments. However, students may use generative AI tools for *coding help* as a last-resort measure—i.e., after checking with your classmates, the professor, and online help forums, etc. In any case, students may not use generative AI tools for the write up of their assignments. Submitting work with a significant percentage of AI-generated content can be considered academic misconduct under Texas A&M University Student Rule 20. Exceptions including pre-existing software additions such as spelling and grammar checkers, which are acceptable. Honestly, it is very easy for professors to spot if you use ChatGPT or another generative AI tool, so particularly in this first course, try to do everything yourself. It will pay off in the long run.