Last updated September 19, 2024

Quantitative Methods for Public Management I

Professor Information

Name: Mike Denly

Email: mdenly@tamu.edu

Office: Allen 1035

Office Hours: Monday 10am-1pm

Office Hours Booking: Link

Website: www.mikedenly.com

Course Information

Abbreviation: Bush 631

CRN: 50326 (Monday); 51372 (Tuesday)

Time: 2:25-5:15pm (Monday); 4:40-7:30pm (Tuesday)

Room: Allen 1005 (Monday); Allen 3072 (Tuesday)

Credit Hours: 3

Website: canvas.tamu.edu

1.   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 and security.

This course uses a problem-based learning (PBL) model. The homework each week entails watching a series of online video lectures. 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. After taking the quiz, the professor will release comprehensive lecture notes that students may use to complete group work that will take up the majority of the class. To ensure that everyone gains experience working with others who have different skill levels and working styles, working pairs will be randomly assigned at the start of each class. To conclude each class, students will present on progress toward their Dream Job Assignment, for which students will complete a presentation as well.

2.   Course Learning Outcomes

Upon successful completion of the course, students will be able to:

- conduct basic descriptive inference, statistical inference, linear regression, and prediction, using the statistical software program R and, to some extent, MS Excel.

- explain the basics of causal inference, using causal diagrams, randomized experiments, and other quasi-experimental methods.

3.   Course Requirements

3.1.   Prerequisite Coursework

There are no formal prerequisites for taking this course, other than being in the International Affairs Masters' Program at Texas A&M's Bush School of Government and Public Service.

3.2.   Required Software

This course makes use of `R` and `Excel`. Prior knowledge of any of these software programs is not required. `R` is free, and students should have free access to all Microsoft Office programs, including Excel, through their university email accounts.

3.3.   Textbooks

The course does not have any required textbooks. The free, online video lectures 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 the textbooks below if they wish. Again, though, none of these textbooks are not required.

3.3.1.   Optional Textbooks Not Freely Available Online

- 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.

- Gerring, John, and Dino Christenson. 2017. *Applied Social Science Methodology: An Introductory Guide.* Cambridge: Cambridge University Press.

- King, Gary, Robert Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton: Princeton University Press.

- Imai, Kosuke, and Nora Webb Williams. 2022. *Quantitative Social Science: An Introduction in Tidyverse.* Princeton: Princeton University Press.

- Li, Quan. 2018. *Using `R` for Data Analysis in Social Sciences: A Research Project-Oriented Approach.* Oxford: Oxford University Press.

3.3.2.   Optional Textbooks Freely Available Online

- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2022. *Regression and Other Stories.* Cambridge: Cambridge 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.

- Huntington-Klein, Nick. 2022. *The Effect: An Introduction to Research Design and Causality*. CRC 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.

3.4. Attendance, Quizzes, and Participation

All students must come to class prepared, having watching the online video lectures <u>before</u> class. At the beginning of each class, I will administer a quiz, which serves three purposes.

First, the quiz will help keep track of attendance and serve as a commitment device for students to attend class and on-time. Even if students miss all of the questions on the quiz but are present for class, they will receive full credit toward attendance for the respective class. Overall, attendance will account for 5% of students' final grades.

Second, because students have the option of using their *handwritten* notes from the online videos for the quizzes, the latter should serve as a reward: you should receive 100% every time if you read and take handwritten notes. To give students some cushion for bad days, extenuating circumstances, or university-excused absences, I will drop your lowest 2 scores. I will make no other accommodations should you miss class for any reason or arrive late and miss the quiz. In total, students' average quiz score will comprise 25% of your final grade.

Third, the quiz will help ensure students are ready to participate in class discussion as well as the graded group work during class. The material is challenging, and passive learning will generally not suffice for students to perform well in the course.

Because participation comprises 5% of the final grade, I will post a 1-5 cumulative score for the semester on Canvas for each student after the fourth class and gradually update it during the semester, as appropriate. This way, the final participation grade will not come as a surprise to students at the end of the semester, and students may address me early if they have any concerns. When assigning your participation grades, I will consider your contributions to group work, Canvas Discussions, and in-class discussions, as appropriate.

4. Policies

4.1. Grading

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 you would like to appeal your grade on any assignment, you must make the request to me in writing, over email, within 5 days of receiving your grade. In your grade appeal, you must specify the reason(s) why you think I misgraded the assignment. Acceptable reasons include those pertaining to the concepts and material covered during the course. I 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: 20%

- Dream Job Assignment: 25%

- In-Class Group Work Other than the Dream Job: 10%

- Dream Job Presentation: 5%

- Attendance: 5%

- Class Participation: 5%

4.2. Canvas Discussions, Office Hours, and Getting Help

If you have questions about homework, class material, or exams, please do not email me. Instead, please publicly post the question to the respective Canvas Discussion page so that everyone can see it. This way, all of the students will be able to benefit from my responses as well as those of the other students. Providing that one of your classmates has not already responded to your question, I will endeavor to respond within 12 hours during the work week. Most days, I will respond within a few hours if I do not have other immediate commitments.

All students are also welcome and encouraged to visit the professor during office hours. Given that many students may want to attend, kindly book the office hours in advance using the online booking tool. Of course, if no one has booked the time slot in advance, students may feel free to drop by the professor's office *during the alloted time spot.* Office hours will take place in person, unless communicated otherwise. If you are unable to attend my scheduled office hours time due to a class or work conflict, please email me to book an appointment. I would be delighted to meet with you at other times, providing that you make an appointment with me. I will not provide office hours help with assignments that are late.

4.3. Absences

It will be very difficult to perform well in the course if students do not attend regularly. The only absences that I will consider legitimate include those pertaining to religious

holidays, illness, extenuating circumstances due to an emergency, and university-excused absences. For illnesses, students will need to either provide me with a doctor's note or send me an email before class to inform me that they are sick and won't be attending. If students are sick and do not provide me with a doctor's note or email me before class, I will not excuse the absence except under very extenuating circumstances. An unexcused absence will result in students receiving a 0 on the quiz and the in-class group work assignment.

4.4.   Late Homework

Unless you receive prior approval from me, I will not accept late assignments. In other words, all late homeworks will receive a zero.

4.5.   In-Class Group Work Policies

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 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.

4.6.   Dream Job Assignment Policies

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 me 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.

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 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, you would never turn in anything formal without a complete sentence, 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, post on the respective Canvas Discussion page, and book office hours, as appropriate. We are all in this together.

4.7.   Dream Job Presentation Policies

Students will be randomly assigned to present on their Dream Job Assignment progress at one point during the semester. When assessing presentations, I 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 a points. Slides should only have a few points on them. Too many words is 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.

4.8.   Students' Rights and Responsibilities

- You have a right to a learning environment that supports mental and physical wellness.

- You have a right to respect.

- You have a right to be assessed and graded fairly.

- You have a right to freedom of opinion and expression.

- You have a right to privacy and confidentiality.

- You have a right to meaningful and equal participation, to self-organize groups to improve your learning environment.

- You have a right to learn in an environment that is welcoming to all people. No student shall be isolated, excluded or diminished in any way.

With these rights come these responsibilities:

- You are responsible for taking care of yourself, managing your time, and communicating with the instructor if things start to feel out of control or overwhelming.

- You are responsible for acting in a way that is worthy of respect and always respectful of others.

4.9.   Exams and Make-up Policy for Exams

Exams will be cumulative and involve open-ended answers. There will be no multiple choice or use of cheat sheets. 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.

4.10.   Disability Policy

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning

environment that provides reasonable accommodation for their disabilities. If you believe you have a disability requiring an accommodation, please contact Disability Services. Provided that I receive an accommodation letter from Disability Services, I will be more than happy to accommodate any disability, and I would encourage students to contact me individually with that letter, if applicable. I will not provide disability accommodations without a letter from Disability Services under any circumstances.

4.11.   Academic Dishonesty/Plagiarism Statement

As commonly defined, plagiarism consists of passing off as one's own the ideas, words, writings, etc., which belong to another. In accordance with the definition, you are committing plagiarism if you copy the work of another person and turn it in as your own, even if you should have the permission of the person. Plagiarism is one of the worst academic sins, for the plagiarist destroys the trust among colleagues without which research cannot be safely communicated. If you have any questions regarding plagiarism or any other form of academic misconduct, please consult the Aggie Honor System Office website) or the latest version of the Texas A&M University Student Rules, under the section "Scholastic Dishonesty." Always remember: "An Aggie does not lie, cheat or steal, or tolerate those who do."

You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at student-rules.tamu.edu/aggiecode. Importantly: "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).

4.12.   Generative Artificial Intelligence

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, but students may only use generative AI tools for *coding help* as a last-resort measure—i.e., after checking with your classmates 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 incredibly easy for me to spot if you use ChatGPT or another generative AI tool, so particularly in this first course, try do everything yourself. It will pay off in the long run.

4.13.   Title IX and Statement on Limits to Confidentiality

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 gender-based discrimination and sexual harassment, including sexual assault, sexual exploitation, domestic violence, dating violence, and stalking. With the exception of some medical and mental health providers, all university employees (including full and part-time faculty, staff, paid graduate assistants, student workers, etc.) are Mandatory Reporters and must report to the Title IX Office if the employee experiences, observes, or becomes aware of an incident that meets the following conditions (see University Rule 08.01.01.M1):

- The incident is reasonably believed to be discrimination or harassment.

- The incident is alleged to have been committed by or against a person who, at the time of the incident, was (1) a student enrolled at the University or (2) an employee of the University.

Mandatory Reporters must file a report regardless of how the information comes to their attention—including but not limited to face-to-face conversations, a written class assignment or paper, class discussion, email, text, or social media post. Although Mandatory Reporters must file a report, in most instances, a person who is subjected to the alleged conduct will be able to control how the report is handled, including whether or not to pursue a formal investigation. The University's goal is to make sure you are aware of the range of options available to you and to ensure access to the resources you need. Students wishing to discuss concerns in a confidential setting are encouraged to make an appointment with Counseling and Psychological Services (CAPS). Students can learn more about filing a report, accessing supportive resources, and navigating the Title IX investigation and resolution process on the University's Title IX webpage.

4.14.   Statement on Mental Health and Wellness

Texas A&M University recognizes that mental health and wellness are critical factors that influence a student's academic success and overall wellbeing. Students are encouraged to engage in healthy self-care by utilizing available resources and services on your campus Students who need someone to talk to can contact Counseling & Psychological Services (CAPS) or call the TAMU Helpline (979-845-2700) from 4:00 p.m. to 8:00 a.m. weekdays and 24 hours on weekends. 24-hour emergency help is also available through the National Suicide Prevention Hotline (800-273-8255) or at suicidepreventionlifeline.org.

Graduate school is demanding; you will face many unexpected challenges. Your health and wellbeing, however, are of paramount importance. If you are feeling overwhelmed, stressed, or facing any other obstacle which seems to be getting in the way of your wellbeing and/or academic achievement, resources and help are available both on-line and in-person free of charge for university students. For more information, see caps.tamu.edu.

In the event that you need an in-person physician or dial-a-nurse medical care (including women's health and pharmacy services), please take advantage of the TAMU Student Health

Services. Regardless of your health insurance status, services are available to you as an enrolled student for a very small fee. For more information, visit shs.tamu.edu.

4.15.   TAMU Writing Center

The University Writing Center (UWC) is here to help you develop and refine the communication skills important to your success in college and beyond. The UWC provides this help in a welcoming atmosphere that respects all Aggies backgrounds and abilities. Our trained peer consultants are available to work with you on any kind of writing or speaking project, including research papers, lab reports, application essays, or creative writing, and at any stage of your process, whether you're deciding on a topic or reviewing your final draft. You can also get help with public speaking, presentations, and group projects. We can work with you in person at our Evans or BLCC locations or via Zoom or email. To schedule an appointment or to view our handouts, videos, or interactive learning modules, visit writingcenter.tamu.edu. If you have questions, need help making an appointment, or encounter difficulty accessing our services, email uwc@tamu.edu.

5.   Class Schedule, Homework, and Videos

Week 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

- US Census Data

- 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)

Week 2: Descriptive Statistics and Intro to R

Required Videos:

- 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.

Week 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)

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. 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.

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

Week 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):

22. 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.

Week 5: Exam 1

Class:

- Exam

Week 6: Probability and Bayes' Rule

Required Videos:

- Probability basics

- Mutually-exclusive and not mutually-exclusive events

- Independent and non-independent events

- Multiplying probabilities of independent and mutually-exclusive events

- Adding probabilities of independent and mutually-exclusive events

- Sample space (Ω) and complements (^{C})

- Conditional probability of independent and non-independent events

- Bayes' theorem/rule

- Bayesian updating

- Example: finding terrorists in an airport

- Example: was Elvis Presley an identical twin?

Class:

- Review of exam

- Group Work: doping at the Olympics

- Dream Job presentations

Dream Job Assignment (Part 5):

23. Invent a question involving what you have learned in Bayesian statistics that applies to your Dream Job. Then, answer that question using Bayes Rule.

Optional Reading:

- Gerring & Christenson: Pages 292-293 (Review)

- Bueno de Mesquita & Fowler: Pages 1-3 and 314-331

- Imai & Webb Williams: 6.1.1-6.1.2 (Pages 279-284); 6.2.1-6.2.3 (Pages 291-309)

- McGrayne, Sharon Bertsch. 2012. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy.* New Haven, CT: Yale University Press.

Week 7: Bivariate Relationships

Required Videos:

- One-sample t-test/difference in means

- How t-ratio statistics and z-scores differ

- 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 6):

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 variable from a different dataset on the Internet that (i) shares the same unit of analysis (i.e., panel structure) as your original variable; and (ii) you think helps explain your original variable. Then, (iii) bring that variable into R. Also, (iv) state the source of the data; and (v) explain in 2-3 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. Test the statistical power of your last test. Assume that your original variable is your dependent variable, and calculate Cohen's d based on your sample. Interpret your results.

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

Week 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):

31. Create a scatterplot with your original variable on the y-axis and your new variable that you added in the last assignment in the x-axis.

32. 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.)

33. With your original variable as the dependent variable and new variable as the independent variable, run a linear regression.

34. "Tidy" your output from the linear regression that you just ran using the `broom` package.

35. Output the results of your linear regression in a table and a coefficient plot using `modelsummary`.

36. 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

Week 9: Linear Regression 2

Required Videos:

- Multivariate regression, F-statistics, 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):

37. 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?

38. 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?

39. Use the Internet to find another independent variable that you think explains your original variable, which will continue to act as your dependent variable. Describe the new independent variable, state the source of the data, and explain why you think it explains your dependent variable.

40. Bring that new variable into R and merge it into your larger dataset, ensuring that the merge goes through correctly. If it doesn't make sure to correct all potential errors.

41. 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.

42. 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?

43. 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

Week 10: Exam

Class:

- Exam

Week 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

- Logistic regression basics

- Logistic regression example: getting an A (Part 1–equations)

- Logistic regression example: getting an A (Part 2–odds ratio & marginal effects in R)

- Logistic regression example: getting an A (Part 3–confusion matrix)

- Logistic regression example: getting an A (Part 4–precision & recall)

Class:

- Review of exam

- Group work

- Dream Job presentations

Dream Job Assignment (Part 9):

44. 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.

45. Create separate random training and test datasets, reserving 25% of your data to the test sample.

46. 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. If your dependent variable is already binary, please just refer the reader to question 39 from last week.

47. Obtain the predictions for the model and ascertain whether all of the predictions make sense.

48. Run a logistic regression model using the exact same specification as above. Can you interpret these coefficients? If so, how?

49. Obtain the odds ratios for the coefficients that you estimated in the previous step, and interpret these odds ratios.

50. Obtain the average marginal effects for the coefficients and interpret them.

51. 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.

52. 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

Week 12: Randomized Experiments

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:

- Experiment basics and the fundamental problem of causal inference

- Types of randomization

- Potential outcomes, assumptions, and estimands

- Applications of experiments

- Field experiments

- Lab experiments and ethics

- Survey experiments and demand effects

- Application: Malaria bednet distribution in Africa

- Validity threats: attrition, noncompliance, and spillover

Class:

- Group work

- Dream Job presentations

Dream Job Assignment (Part 10):

53. What is a causal question that you will need to answer as part of your dream job?

54. 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.

55. What would be the constraints to performing such an experiment? Hint: you can talk about ethics, resources, external validity, or other things.

56. 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.

57. 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.

58. 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.

- 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

Week 13: Natural Experiments and Quasi-Experiments

Required Videos:

- Causal diagrams and identification assumptions

- Matching

- Difference-in-differences

- Regression discontinuity designs

- Instrumental variables

Class:

- Group work

- Dream Job presentations

Dream Job Assignment (Part 11):

59. 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?

60. What would a data generating process that yields those concerns look like? Characterize that process in the form of a causal diagram.

61. 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?

62. 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.

63. 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.

Week 14: Critical Consumption of Quantitative Information

Required Videos:

- Case studies

- External validity

- Key takeaways from this course

Class:

- Group work

- Dream Job presentations

- Exam review

Optional Reading:

- Bueno de Mesquita & Fowler: Chapter 16

- Gerring & Christenson: Chapter 9

- Gelman, Hill & Vehtari: Chapters 1.

- Findley, Michael, Kikuta, Kyosuke, and Denly, Michael. 2021. "External Validity." *Annual Review of Political Science* 24: 365-393.

  - Read: pages 365-373; the rest of the article is optional.

- Williams, Martin. 2021. "External Validity and Policy Adaptation: From Impact Evaluation to Policy Design." *World Bank Research Observer* 35(2): 158-191.

Final Exam Date: December 6, 2024 at 10:30am

- Both the Monday and Tuesday sections will meet at this time in Room 3072.