*Last updated August 15, 2023*; latest version here.

# Quantitative Methods for Public Management I

**Instructor Information**
Name: Mike Denly
Email: mdenly@tamu.edu
Office: Allen 1035
Office Hours (OH): Monday 10am-1pm
OH Booking: mikedenly.youcanbook.me
Website: www.mikedenly.com

**Course Information**
Abbreviation: Bush 631
CRN: 53172
Time: 1:30-4:20pm
Room: Allen 1005
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.

## 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.

## 3.3. Required Texbooks

Students must purchase (or borrow from a library) the course's primary textbooks:

- Bueno de Mesquita, Ethan, and Anthony Fowler. 2022. *Thinking Clearly with Data: A Guide to Quantitative Reasoning and Analysis*. Princeton: Princeton University Press.
- Gerring, John, and Dino Christenson. 2017. *Applied Social Science Methodology: An Introductory Guide*. Cambridge: Cambridge University Press.
- Huntington-Klein, Nick. 2022. *The Effect: An Introduction to Research Design and Causality*. 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.
- Li, Quan. 2021. *Using R for Data Analysis in Social Sciences: A Research-Project Oriented Approach*. Oxford: Oxford University Press.

## 3.4. Optional Textbooks

- King, Gary, Robert Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Gelman, Andrew, Jennifer Hill, and AKi Vehtari. 2022. *Regression and Other Stories*. Cambridge: Cambridge University Press.
- West, Carl, and Carl Bergstrom. 2020. *Calling Bullshit: The Art of Skepticism in a Data-Driven World*. New York: Penguin Random House.
- Wickley, Hadley, Mine Cetankanya-Rundell, and Garett Grolemund. 2023. *R for Data Science: Import, Tidy, Transform, and Model Data*. Sebastopol, CA: O'Reilly Media.

For some weeks, I supplement the textbook with other required and optional readings. When these articles can be easily found on the TAMU Library webpage, I will ask students to download the article(s) themselves—to ensure students know how to use the library website. Otherwise, I will post the article(s) on the class website, Canvas. For more information on the specific reading assignments for each week, refer to the Class Schedule (below). Optional readings are not required for each class period, and reading them will not enable students to receive extra credit. However, I may use these readings to supplement the textbook in case it is necessary to facilitate comprehension of important topics.

## 3.5. Attendance, Quizzes, and Participation

All students must come to class prepared, having completed the readings before class. At the beginning of each class, I will give everyone a five-question, multiple-choice quiz.

The quiz 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 both 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 the quiz will only cover the most basic ideas from the required reading, the quiz will serve as a reward: you should receive 100% every time if you read. To give students some cushion for bad days or extenuating circumstances, 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 20% of your final grade.

Third, the quiz will help ensure students are ready to discuss the material and do not rely entirely on my lecture to learn the materials. The material is challenging, and passive learning will generally not suffice for students to perform well in the course. Since 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. As stipulated in the Policies section of this syllabus, I will make every possible effort to ensure that students feel comfortable participating. To ensure that you receive good grade for participation, please try to make at least one meaningful contribution to discussion each class.

# 4. Policies

## 4.1. Grading Rubric

- Attendance: 5%
- Class Participation: 5%
- Exams: 40%
- Homework: 30%
- Quizzes: 20%

## 4.2. Grading Scale

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

- $< 59.5$ (F)

## 4.3. Grade Rounding

The above grading scale already incorporates very generous grade rounding, not to mention the multitude of extra credit opportunities. Accordingly, there will be no additional rounding of grades under any circumstance.

## 4.4. Grade Appeals

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.

## 4.5. Communication, Canvas Discussions, and How to Reach the Professor

If you have questions about homework, class material, or exams, I would kindly ask that you publicly post the question to the respective Discussion on Canvas 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. Additionally, you are more than welcome to ask coursework-related questions during office hours, and you are welcome to reach out to me over email for any other matters. I will endeavor to respond within 24 hours during the work week.

## 4.6. Office Hours

All students are 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, student may feel free to drop by the professor's office (if in person) or click the Zoom link (if remote office hours). Students will be informed whether office hours will take place each week in person or over Zoom.

## 4.7. Absences

As described in the Course Requirements section of the syllabus (above), it will be very difficult to perform well in the course if you 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, you will need to either provide me with a doctor's note, or you will need to send me an email before class to inform me that you are sick and won't be attending. If you are sick and do not provide me with a doctor's note or email me before class, your absence will not be excused except under very extenuating circumstances.

## 4.8. Late Homework

Unless you receive prior approval from me, I will not accept late assignments without grade penalties, and I will discount most other late assignments as follows:

- 1-15 minutes: 0% (grace period for last-minute issues)
- 15 minutes-24 hours late: -10%
- 24-48 hours late: -25%
- more than 2 days late: -50%
- more than one week: -75%
- more than two weeks: no credit offered

Unless otherwise specified, the only exceptions to the above are the last two assignments, for which I will not offer any credit one week beyond the respective due dates. The reason is that I must heed to university due dates for submission of your final grades.

## 4.9. Homework Policies

Students may consult with other members of the class and/or work in groups for the homework assignments. Regardless of whether students choose to work in groups on the R assignments, students must submit their own copies of their work—i.e., no group submissions. Students are also not allowed to work on the exact same variable with the exact same dataset for the Dream Job homeworks. If the professor notices that more than one student has the same dream job and uses the same variable in the first assignment, the professor will contact the students to let them know about the conflict and ask them to choose different variables.

The professor will drop students' three lowest homework scores when calculating the final portion of the grade dedicated to homework. The professor will not drop additional homework grades. Students who are facing difficulties completing the homework are encouraged to contact the professor and thereby address any standing issues.

If you need help with a particular question, feel free to write the whole class on Canvas (see above) or book an office hours slot. Provided that you attended the class where I covered the material at hand or missed class due to an excused absence (see above), I am very happy to help! I will not provide additional make-up training during office hours if you missed class for a non-excused absence.

## 4.10. 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.

## 4.11. Personal Pronoun and Name Preferences

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. All students are encouraged to place a paper name tag in front of them in class, signalling their preferred name and gender pronoun.

## 4.12. Make-up Policy for Exams

Per Student Rule 7, students will 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. Please read Student Rule 7 in its entirety for relevant rules and regulations.

## 4.13. 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 (http://disability.tamu.edu). 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 indivdiually with that letter, if applicable. I will not provide disability accommodations without a letter from Disability Services under any circumstances.

## 4.14. 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 com-

mitting 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 http://www.tamu.edu/aggiehonor or the latest version of the Texas A&M University Student Rules, under the section "Scholastic Dishonesty." http://rules.tamu.edu. 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 aggiehonor.tamu.edu. 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).

## 4.15. Generative Artificial Intelligence

Artificial Intelligence (AI) text generators and natural language processing tools (colloquially, chatbots - such as ChatGPT), audio, computer code, video, and image generators are explicitly prohibited for quizzes and exams. The professor discourages the use of these tools in the completion of homework assignments. However, as a last-resort measure (i.e. after checking online forums, your classmates, etc.), students may use generative AI tools for coding help on homework. In such instances, these technologies should not be used without appropriate attribution, and students may not use generative AI to perform the write-up of any assignment. 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. The professor may use AI detection tools like GPTZero at random to detect the possibility of academic misconduct in the writing of homework.

## 4.16. 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.

## 4.17. 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.

## 4.18. 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, Readings, and Homework

## Week 1: What Is Quantitative Social Science? (August 22, 2023)

Class:

- Part 1: Introduction
  - Instructor introduction
  - Student introductions
  - Syllabus and class expectations
  - The four characteristics of social scientific research
  - What distinguishes social science from casual conversation
  - Quantitative vs qualitative research
  - Observations vs variables
  - Data types: cross-sectional, time-series and panel data
  - Long vs wide data
  - Variable types: binary, continuous,
- Part 2: Essential skills with Microsoft Excel (using the WGI)
  - Saving and file types (e.g., `.xlsx` vs. `.csv`)
  - Inspecting and filtering data
  - Merging cells, wraping 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

Required Readings:

- Carefully read the course syllabus
- King, Keohane & Verba: Section 1.1 (Pages 3-12).
- Imai & Webb Williams: Sections 1.1-1.2 (Pages 1-8).
- 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

Human Subjects Protection Training in CITI Homework: 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 screen shot of your certificate on Canvas.

## Week 2: Variable Description (August 29, 2023)

Class:

- Univariate description
- Dispersion measures: standard deviation and variance.
- Variable transformations: logs, per capita, deflation of time series.
- Discrete variables: binary, categorical, and ordered variables.
- Continuous variables
- Applications in R
  - Setting the working directory
  - Objects, vectors, entering in data manually, and creating data frames
  - Classes (mumeric, character, factors)

  - Dealing with missing values
  - Installing packages and loading libraries
  - Basic data visualization in `ggplot2`
  - Descriptive statistics (mean, median, mode, quantiles)
  - Tables with `modelsummary`

Required Readings:

- Gerring & Christenson: Chapter 18
- Huntington-Klein: Sections 3.3-3.4
- Imai & Webb Williams: Sections 3.1-3.3

Optional Readings:

- Gerring, John. 2012. "Mere Description." *British Journal of Political Science* (42)4: 721-746.

Excel/Google Sheets Homework Assignment:

- Complete the free "Cells and Formulas" Chapter from Data Camp's Intro to Spreadsheets course. Once you are done, post a screenshot on Canvas to prove that you completed the chapter.

R and R Studio/Posit Homework Setup Assignment:

- You must install R and R Studio/Posit prior to class, and bring your computer with these programs installed to class. See here for relevant installation links. I will not accept late assignments for this assignment.

Dream Job Homework (Part 1):

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 kind of problems that you would need to tackle often as part of your job? What kind of information, data, or analyses would you need in order to tackle those problems?
3. Download a panel dataset from the internet that corresponds to that dream job, state the name of the dataset, state only one variable of focus, and the submit the raw data as an Excel or CSV file with your assignment. Label the file "raw_data".
4. What is the unit of analysis for that 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, median, and mode? (Hint: if there are missing values, you may need to filter them out.)

7. Make a pivot table to summarize your panel dataset into a cross-section of the mean of all periods per unit. Report the mean value for a specific cross-sectional unit in your dataset. Show a screenshot of the pivot table.
8. Produce a bar graph that shows those values across units *sorted* by the average value. Show the figure.
9. 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.
10. Produce a line graph that shows those values *sorted* by period. Show the figure.

## Week 3: Probability Distributions (September 5, 2023)

Class:

- Probability distributions for different variable types
- Central Limit Theorem and the Law of Large Numbers
- $z$-scores
- R programming:
  - Lists
  - Vectors
  - Simulation
  - Loops
  - Functions
  - Apply family commands: `lapply`, `sapply`, etc.
  - Conditionals and control flow (e.g., `ifelse`)
  - Introduction to Quarto
- Probability density functions (pdfs) and cumulative density functions (cdfs)
  - Time-permitting

Required Reading:

- Imai & Webb Williams: Sections 6.3-6.4
- Gerring & Christenson: Chapter 19

Optional Readings:

- Gelman, Hill & Vehtari: Section 3.5
- Wickham, Cetinkaya-Rundell & Grolemund: Chapters 17-21

Dream Job Homework in R (Part 2):

1. Imagine that you have received your dream job after finishing your degree at Texas A&M. What's that dream job? [Yes, this is the same question from last week; kindly repeat the answer.]
2. What are the kind of problems that you would need to tackle often as part of your job? What kind of information, data, or analyses would you need in order to tackle those problems? [Yes, this is the same question from last week; kindly repeat the answer.]
3. Download a panel dataset from the internet that corresponds to that dream job, state the name of the dataset, state only one variable of focus, and the submit the raw data as an Excel or CSV file with your assignment. Label the file "raw_data". [Yes, this is the same question from last week; kindly repeat the answer.]
4. What is the unit of analysis for that dataset? What variable(s) identify that unit of analysis? [Yes, we know this is the same question from last week.]
5. Is the dataset in long or wide format? How dod you know? [Yes, we know this is the same question from last week.]
6. What are the summary statistics for your variable, including the mean, median, mode, and standard deviation? Show screenshots of both your code and output in R. (Hint: if there are missing values, you may need to remove them.)
7. Collapse or summarize your panel dataset into a cross-section of the mean of all periods per unit. Report the mean value for a specific cross-sectional unit in your dataset. Show screenshots of both your code and output in R.
8. Use `ggplot2` to produce a bar graph that shows those values across units *sorted* by the average value. Show screenshots of both your code and output in R.
9. Collapse or summarize your panel dataset into a time series of the sum of all units per time period. Report the sum value for a specific period in your dataset. Show screenshots of both your code and output in R.
10. Use `ggplot2` to produce a line graph that shows those values *sorted* by time period. Show screenshots of both your code and output in R.

## Week 4: Statistical Inference (September 12, 2023)

Class:

- Sampling: random sampling, stratified samples, etc.
- Sampling distributions
- Standard errors
- Confidence intervals
- $p$-values

- Hypothesis testing
- Error types: Type I and Type II errors
- Margin of error
- Statistical power

Required Readings:

- Imai and Webb Williams: Sections 7.2
- Gerring & Christenson: Chapter 4 (pages 53-60); Chapter 20.
- Bueno de Mesquita & Fowler: Chapter 6 (pages 94-105).

Functions, Loops, and Central Limit Theorem Homework:

1. Create a vector of 10 draws from a normal distribution that is centered at 5 and has a standard deviation of 2. What is the mean of that vector? What is its variance? What is its standard deviation? How do the standard deviation and variance relate to each other? Is the mean of the vector equal to 5? If not, why is this the case?
2. Repeat this same process once again. Is the mean of this second vector equal to 5? Is it equal to the mean of the first vector? Why is this the case? Show screenshots of both your code and output in R.
3. Write a function that takes the (a) mean and (b) standard deviation of a normal distribution and (c) number of draws as inputs, and produces the mean and standard deviation of the resulting vector as output.
4. Write a loop that performs that function 1,000 times and collects the mean of each iteration of the function into a separate vector. Repeat this process three times, first setting the number of draws in the function to 10, then setting the number of draws to 100, and finally setting the number of draws to 1,000.
5. Report your findings on all on the same `ggplot2` figure as three overlaying histograms, each with a different color that is duly identifying the number of draws in the figure's legend.
6. What do you see in the figure? Do the three histograms center around the same value? What value is this? Do the histograms differ from each other? If so, why? How does this relate to the topics we discussed in our last class?
7. 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. Note: you may need to install the `tinytex` package in `R` in order to produce the relevant `.pdf` file.

## Week 5: Exam 1 (September 19, 2023)

Class:

- Exam

Confidence Intervals, Margins of Error, and $p$-values Homework:

1. For each of the three vectors/histograms from last week's homework, determine the 2.5% and 97.5% percentiles. How do these ranges (confidence intervals) compare to each other? What are the margins of error in each?
2. Considering that the three distributions are centered at 5, what is the p-value associated with a value of 6.5 in each of them? Could we reject a null hypothesis that the true mean of the distribution is 5 if we observed a value of 6.5 in all of them?
3. Consider the higher end of the three confidence intervals above. Replicate the distributions as in last week's homework, but center them at 6.5. How likely is it to observe values smaller that those higher critical values if the distributions are centered at 6.5? What does that tell us about sample size and statistical power?

## Week 6: Decision Theory and Bayes (September 26, 2023)

Class:

- Probability
- Sample space and complements
- Law of total probability
- Law of addition
- Independence
- Conditional independence
- Bayes' theorem
- Bayesian updating
- Applications in R
  - Merging datasets (see here)
  - Bayesian updating in R to account for false positives

Required Reading:

- Imai & Webb Williams: Sections 6.1.1-6.1.2
- Bueno de Mesquita & Fowler: Pages 314-331
- Denly, Michael. 2021. "Making Merges Go Through Using `tidylog` and `anti_join`." Blog Post.

Required Homework Assignments:

- None.

## Week 7: Bivariate and Multivariate Relationships (October 3, 2023)

Class:

- Cross-tabulations
- Difference in proportions
- Covariance
- Correlation
- Difference in means
- $t$-tests
- Applications in R
  - Overlapping histograms
  - Scatter plots
  - Adding features to `ggplot2` figures (e.g., line types, colors, shapes)
  - Adding multiple plots to the same figure (e.g., `ggarrange`, `facets`)

Required Reading:

- Imai & Webb Williams: Section 3.6
- Bueno de Mesquita & Fowler: Chapter 2.
- Gerring & Christenson: Chapter 21.

Bayesian Homework:

1. Assume that you have been called by an international anti-doping organization to testify on the probability that athletes are using performance-enhancing drugs (PEDs). The first athlete was randomly selected through the organization's random testing program. It is thought that overall 5 percent of the athletes in this particular sport are using PEDs. In this instance, the athlete ends up testing positive for PEDs, and the test is known to give positive results 95% of the time that the athlete in question is actually using PEDs. However, the test also gives a positive result 3% of the time when the athlete in question is not actually using PEDs. What is the probability that this first athlete was using PEDs?
2. After you testify, you find out that the sample size for the initial estimate of athletes from this particular sport using PEDs is 20. Meanwhile, 5 more drug test results become available, and 60% of those tests come back positive. What is the new prior probability of athletes from this particular sport using PEDs?
3. The international anti-doping agency was happy with your work during your first testimony, so they call you to testify again after a gold-medal winning athlete tests positive for PEDs. Using the same test as before, does the probability that the athlete

was actually using PEDs change? Do you think the athlete's gold medal should be taken away?

4. The first athlete protested the result of the initial test and demanded that a new test be performed. The second test also comes up positive. Assuming independence between the two tests, what is the posterior probability of the athlete using PEDs given that both tests came up positive?

## Week 8: Fall Break - No Class (October 10, 2023)

## Week 9: Linear Regression 1 (October 17, 2023)

Class:

- Dependent and independent variables
- Line of best fit in a scatter plot
- Measures of goodness of fit
- $F$-test
- Interpreting a regression coefficient
- Binary and categorical variables as independent variables
- R regression exercises

Required Reading:

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

Optional Reading:

- Li: Chapter 5
- Gareth, Witten, Hastie & Tibshirani: Section 3.1

Dream Job Homework (Part 3):

1. From your dream dataset, pick one continuous variable of interest and dichotomize it into two new numeric variables. (Note: "dichotomize" means separate into binary—i.e., 0 or 1 according to whether the value is below or above the median of the variable).
2. Create two new character/string variables on the basis dummy/indicator variable from the last question, and adequately name those variables according to your context. Hint: one variable should capture when that dummy variable == 1; and the other variable should capture that dummy variable == 0. For example, if you are working with a corruption variable, one variable should capture more corrupt countries, and the other variable should capture less corrupt countries.

3. Pick another variable from a different dataset that shares the same unit of analysis (i.e., panel structure), and bring in that variable to R.
4. Merge the two datasets together, making sure to that everything merges in. (Hint: see Mike's blog post).
5. Subset the data to only keep only one year of the data.
6. Test if the the original variable from the first two Dream Job Homeworks is correlated with the new variable that you imported, by showing (a) a pairwise correlation table; and (b) a labeled scatter plot. Explain in words what your table and scatterplot suggest.
7. Create a crosstab with the binary versions of both variables that you created above. Explain in words what your crosstab suggests.
8. Consider your original variable in continuous form and the new variable in binary form. Use a $t$-test to assess whether the mean of the original variable is the same when the new variable $== 0$ vs. when the new variable $== 1$.
9. Produce an overlapping histogram of the distribution of your original variable in the sample when the new variable $== 0$ vs. when the new variable $== 1$. Put two vertical lines at the average value for both samples. Your histograms should have different colors and must be duly identified in the figure's legend.
10. 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. Note: you may need to install the `tinytex` package in `R` in order to produce the relevant `.pdf` file.

## Week 10: Linear Regression 2 (October 24, 2023)

Class:

- Multivariate regression
- Interpreting coefficients in multivariate regression
- How coefficients and standard errors change as you add regression controls
- Multicollinearity
- Gauss-Markov Assumptions
- Heteroskedasticity-robust and clustered standard errors
- How the goodness of fit changes as you add regression controls
  - $R^2$
  - Adjusted $R^2$
- Interaction terms and polynomials as independent variables

- Fixed effects in panel data
- R regression exercises

Required Reading:

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

Dream Job Homework (Part 4):

1. Merge two additional variables that you think predict your original variable of interest from from Dream Job Homework 1. Transform them into binary variables around their median values, as you did in the last assignment. You should now have four variables: The original one, the one you added during the last assignment and the two new ones from this week.
2. Create a scatterplot that has the continuous value of your original variable in the $y$-axis and the continuous value of the variable you added in the last assignment in the $y$-axis. Additionally, make each point to have a different color and a different shape according to the binary values of the two new variables you just added in the last step. The colors and shapes should be duly identified in the figure's legend. Be sure to make variable names understandable in your figure.
3. Add the line of best fit to the scatter plot. Make sure your line is colored in black and with a dash shape. What's special about this line? (Hint: Be sure to mention residuals.)
4. With your original variable as the dependent variable, run three linear regressions separately, changing the independent variable with the other three variables.
5. Output the results of your three linear regressions in a table using `modelsummary`. Interpret the two coefficients in each of the three regressions, considering their practical/substantive and statistical significance. Which regression has the higher $R^2$ value?
6. "Tidy" your output from each of the individual regressions that you just ran using the `broom` package.
7. Produce a single coefficient plot to capture all of three regressions. The legend on the coefficient plot should identify each of the individual regression outputs.
8. 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. Note: you may need to install the `tinytex` package in `R` in order to produce the relevant `.pdf` file.

## Week 11: Exam (October 31, 2023)

Class:

- Exam

Dream Job Homework (Part 5):

1. Perform the same regression table as last week, adding a fourth regression column that controls for all three independent variables at the same time. Do the coefficients associated with each variable change substantively? If so, why do you think the coefficients changed that much?
2. Consider one of the first three regressions. 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?
3. 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?
4. Compare the $R^2$ score for the fourth regression. Is it larger or smaller than those of the first three regressions? Why would this be the case? Is the Adjusted $R^2$ necessarily larger for regressions with a higher number of independent variables?
5. Produce a pairwise correlation table of the three 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?
6. Take two of your independent variables. Run a regression that controls for them and for their interaction term. Interpret each of the regression coefficients. Produce a coefficient plot using the `interplot` package.
7. 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. Note: you may need to install the `tinytex` package in `R` in order to produce the relevant `.pdf` file.

## Week 12: Prediction and Classification (November 7, 2023)

Class:

- Prediction in linear regression
- Overfitting
- Model complexity
- In-sample vs. out-of-sample prediction
- Linear probability model
- Logistic regression

- Marginal effects
- Confusion matrix
- Precision
- Recall
- ROC curves
  - Time-permitting

Required Reading:

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

Required Homework Assignment:

- None

## Week 13: Randomized Experiments (November 14)

Class:

- Fundamental problem of causal inference
- Potential outcomes framework
- Correlation vs. causation
- Omitted variable bias
- Random assignment
- Reverse causality
- Randomized experiments: field, survey, and lab
- Internal validity
- Estimands
- Challenges with experiments
  - External validity
  - Ethical considerations
  - Attrition
  - Non-compliance
  - Spillover

  - Hawthorne effects
  - Demand effects

- R Application
  - 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.

Required Readings:

- Imai & Webb Williams: Sections 2.3-2.4
- Bueno de Mesquita & Fowler: Chapter 3
- Gerring & Christenson: Chapters 8 and 23.

Optional Reading:

- Huntington-Klein: Chapter 7

Dream Job Homework (Part 6):

1. Create separate random training and test datasets, reserving 25% of your data to the test sample.
2. On the training dataset, take the binary version of your original variable from Dream Job Homework Part 1. Estimate a linear probability model using the multivariate specification with three independent variable from Dream Job Homework Part 5. Tell us what you find in terms of practical/substantive significance, statistical significance, $R^2$, and Adjusted $R^2$.
3. Obtain the predictions for the model and ascertain whether all of the predictions make sense.
4. Run a logistic regression model using the exact same specification as above. Can you interpret these coefficients? If so, how?
5. Obtain the odds ratios for the coefficients that you estimated in the previous step, and interpret these odds ratios.
6. Obtain the average marginal effects for the coefficients and interpret them.
7. 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.
8. 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?

9. 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. Note: you may need to install the `tinytex` package in `R` in order to produce the relevant `.pdf` file.

## Week 14: Natural Experiments and Quasi-Experiments (November 21, 2023)

Class:

- Causal diagrams and identification assumptions
- Closing back doors
  - Regression
  - Matching
    - Application: Galiani, Sebastian, and Ernesto Schargrodsky. 2010. "Property Rights for the Poor: Effects of Land Titling." *Journal of Public Economics* 94(9–10): 700-729.
- Natural experiments and finding front doors
  - Difference-in-differences
    - Application: Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *Industrial and Labor Relations Review* 43(2): 245-257.
  - Regression discontinuity designs
    - Application: Dell, Melissa. 2015. "Trafficking Networks and the Mexican Drug War." *American Economic Review* 105(6): 1738-79.
    - Application: Morales-Arilla, José Ramon. 2023. "Export Side Effects of Wars on Organized Crime: The Case of Mexico." *Journal of International Economics* September 2023(103775): 2-22.
  - Instrumental variables
    - Application: Angrist, Joshua. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80(3): 313-336.

Required Reading:

- Imai & Webb Williams: Sections 2.5 and 4.3
- Gerring & Christenson: Chapter 8

- 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.
- Lipsky, Ari, and Sander Greenland. 2021. "Causal Directed Acyclic Graphs." *Journal of the American Medical Association* 327(11): 1083-1084.

Optional Reading:

- Huntington-Klein: Chapters 5, 16, and 19

Dream Job Homework (Part 7):

1. What is a causal question that you will need to answer as part of your dream job?
2. 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.
3. What would be the constraints to performing such an experiment? Hint: you can talk about ethics, resources, external validity, or other things.
4. 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.
5. 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.
6. Do the above challenges affect your estimand of interest? If so, what is another estimand that you can examine?

## Week 15: Critical Consumption of Quantitative Information (November 28, 2023)

Class:

- Case studies
- Application: Did the 2017 financial sanctions on Venezuela kill 40,000 people?
- Critical consumption of quantitative analyses
  - Practical significance and measurement
  - External validity
  - Selected samples
- Key takeaways from this course

Required Reading:

- Bueno de Mesquita & Fowler: Chapter 16
- Gerring & Christenson: Chapter 9
- West and Bergstrom: Chapter 10
- Sachs, Jeffrey. 2019. "Economic sanctions as collective punishment: The case of Venezuela." Center for Economic and Policy Research.
- Bahar, Dany, Sebastian Bustos, José Ramon Morales-Arilla, and Miguel Santos. 2021. "Impact of the 2017 sanctions on Venezuela: Revisiting the evidence." *Social Science Research Network* 3809344.
- 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.

Dream Job Homework (Part 8):

1. 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?
2. What would a data generating process that yields those concerns look like? Characterize that process in the form of a causal diagram.
3. 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?
4. 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.
5. 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, or difference-in-differences. Explain your design in detail.

## Final Exam Date: December 8, 2023 at 10:30am