Introduction to Quantitative Analysis

DACSS 603 / POLISCI 755 (3 Cred.)

University of Massachusetts Amherst

Fall 2024

Instructor

Dr. Omer Yalcin
Bartlett Hall 259
oyalcin@umass.edu

Course Time and Location

Time: Tuesday & Thursday, 1:00PM - 2:15 PM
Location: Machmer W-13 or Zoom

Please find the Zoom link in Canvas.

Office Hours

You can join office hours either in person in my office or on Zoom. Please book an appointment in advance at https://omerfyalcin.youcanbook.me/. If the time slots listed there do not work for you, then please email me for an appointment. I am also happy to arrange evening Zoom slots for those cannot make it during the day.

Course Description & Objectives

This course serves as a rigorous introduction to quantitative empirical research methods, designed for doctoral students in the social sciences and master's degree students with a data analytics or computational social science focus. Our anchoring framework will be regression as tools to explore and describe associations among variables, to predict outcomes of interest, to generalize beyond a sample, and as part of a broader strategy of causal inference. Simulations and data analysis will be conducted in the R statistical environment. Upon completion, students will be able to understand, evaluate, and criticize the use of common statistical methods for social science research. The course will also serve as a useful basis for more advanced research methods courses.

. I would like to thank Justin Gross for permitting me to adopt portions of his syllabus. The syllabus is subject to change with reasonable advance notice.

Prerequisites

Basic knowledge of algebra, sets, functions, and probability is assumed. More advanced mathematics will be introduced as needed for those who may not have encountered such material previously (and as review for others), but we will focus on conceptual understanding and enhancing our ability to read mathematical notation in papers. It is best to have had some exposure to R prior to this class, preferably at the level of DACSS 601 Introduction to Data Science, but it is possible to succeed without this background if you are willing to put in extra work toward the beginning of the semester.

Textbook

All required course material is either freely available on the public internet, or will be provided for you in our learning management system Canvas, or is available electronically through the UMass Library (details below). You do not need to buy anything unless you want them in a different format, such as in the form of a physical copy. The main textbook is:

- [**RaOS**] Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories.* 1st edition. Cambridge New York, NY Port Melbourne, VIC New Delhi Singapore: Cambridge University Press, July. ISBN: 978-1-107-02398-7. https://users.aalto.fi/~ave/ROS.pdf (**downloadable pdf**)

The textbook also has a homepage with code and data for examples in the book: https://avehtari.github.io/ROS-Examples/index.html

We will also use chapters or excerpts from the following:

- [**LSR**] Navarro, Danielle. 2016. *Learning Statistics with R.* https://learningstatisticswithr.com/lsr-0.6.pdf (**downloadable pdf**)
- [**Rethinking**] McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN.* 2nd edition. Boca Raton: Chapman / Hall/CRC, March. ISBN: 978-0-367-13991-9 (**pdf available through UMass Library**. Either click here or search for the ISBN in the library website. You'll need your UMass credentials to access.)
- [**BayesRules**] Johnson, Alicia A., Miles Q. Ott, and Mine Dogucu. 2022. *Bayes Rules!: An Introduction to Applied Bayesian Modeling.* 1st edition. Boca Raton, Fla London New York NY: Chapman / Hall/CRC, March. ISBN: 978-0-367-25539-8. https://www.bayesrulesbook.com/ (**readable online**)
- [**Effect**] Huntington-Klein, Nick. 2022. *The Effect: An Introduction to Research Design and Causality.* 1st edition. Boca Raton: Chapman / Hall/CRC, January. ISBN: 978-1-03-212578-7. https://theeffectbook.net/ (**readable online**)
- [**Mixtape**] Cunningham, Scott. 2021. *Causal Inference: The Mixtape.* New Haven London: Yale University Press, January. ISBN: 978-0-300-25168-5. https://mixtape.scunning.com/ (**readable online**)

The Course Schedule section lists the required and recommended readings for every week. The Online Resources section lists online resources that we will not directly use, but you may find them useful as quick, easily accessible reference material with applied examples in R. The Further Reference section lists reference material for further or deeper study that you may use for this course, but also (more likely) beyond it.

Learning Management System

Classroom material, including slides, resources, assignments, announcements, deadlines, and more will be posted in our learning management system, Canvas. Canvas can be accessed at https://www.umass.edu/it/canvas. We will also use Canvas for asynchronous discussion. More on this in the participation section.

Lectures

Most weeks, there will be pre-recorded short lecture videos posted on Canvas. Students should watch those lectures **before** attending class alongside doing all the required readings. We will use synchronous class time to discuss any remaining issues and do classroom activities. It is essential that students come prepared having done the readings and watched any lecture videos posted on Canvas; ready not just to listen, but to interact. All class sessions will be recorded via Zoom and posted on Canvas afterwards. Students who cannot attend class synchronously should watch them later.

Grading

Grades are calculated as follows:

- Participation (10%)
- Homeworks (50%)
- Final Project (40%)

Participation: It is imperative that students actively and regularly participate in class discussion. Canvas's discussion section will be the primary way of participating in class discussion asynchronously. Students are expected to regularly ask and/or answer questions about R, statistics, math, readings, assignments and so on. You can also participate by sharing insights, material, and other resources on Canvas. Participation does not need to reflect expertise. Synchronous participation during class is another additional opportunity to participate, but online students who may be unable to join synchronously will not be disadvantaged in grading.

Homework Assignments: There will be a homework assignment most weeks, a total of ten. The assignments will include exercises to help you better understand concepts and methods covered during class. Collaboration is acceptable, but please write up your own answers; do **not** hand in identical written responses. Two of the ten assignments will check your progress on the final project and provide you with timely feedback. So please get started on the final project early and start making progress. See which weeks we have assignments and which of them are related to the final in Course Schedule

Deadlines and Late Work: The assignments will be due on a Friday at 11:59 pm. Each assignment—but not the final project—will have two full days of a grace period following the deadline. That means you can return the assignment penalty-free until two days after the deadline, i.e. Sunday at 11:59 pm. However, if you do make use of the grace period, then you may also receive feedback later than those that return on time. If, without a valid excuse, an assignment is returned even later than the Sunday following the Friday it was due on, then it will be penalized by a 10% deduction of the full grade for every day that it is late (returning on Monday counts as one day late, Tuesday is two days, etc.). If you think you will not be able to meet a deadline for a good reason and you contact me by email at least 24 hours in advance of the Friday deadline, we can work out a new deadline.

Final Project: The final project will be in the form of a data analysis paper that investigates a social scientific question using methods covered in the course, primarily linear regression. It should identify a research question, posit a hypothesis, and use regression methods to assess the evidence for or against that hypothesis. One difference between the final project and a typical published paper is that your final project paper should include not just text, tables, figures, etc. but also the R code (see authoring assignments). Details of how the final project paper should look like can be found in Canvas.

It is important that students make progress on the final project throughout the semester. To ensure that this is the case, two assignments out of the ten will check in on your final project progress.

At the end of every semester, DACSS holds a poster session on campus. It is a great opportunity to get feedback and connect with other students and faculty on campus. You are highly encouraged to (but not required) present your project in the form of a poster at the poster session.

Authoring: All assignments and the final project should be completed using R as a Quarto document. Quarto is a scientific and technical publishing system that allows for easily integrating narrative text, equations, figures, code, and output from code into a single document. Using the RStudio IDE makes authoring and rendering Quarto documents easy. Please provide both the source code (.qmd file) and a Quarto-rendered pdf document (.pdf) for every assignment. For more information on Quarto, visit https://quarto.org/.

Grading Rubric for Final Course Grade:

Final letter grades are assigned using the University's Plus-Minus Grading Scale according to following rubric:

- A (94-100%)
- A- (90-93%)
- B+ (86-89%)
- B (81-85%)
- B- (77-80%)
- C+ (74-76%)

- C (70-73%)
- F (Below 70%)

Software

The course will be taught in the R programming language and the RStudio IDE. R is a free software environment for statistical computing and graphics; it has become the standard in statistics and among increasing numbers of social scientists and policy analysts. Aside from the advantage of being completely free of charge and open-source, R is versatile and powerful, with new packages becoming available all the time for handling tasks central to data analytics and computational social science (e.g. Bayesian inference, social network analysis, text analysis). If you have not done so already, please start by downloading R and RStudio on your own computer. You may also wish to sign up for a free RStudio Cloud account to work in R in your online browser. For more memory-intensive code, this will not be ideal, but it can be a good way to quickly work on small projects with others. You will also need Quarto, a scientific publication system for data science. While Quarto can be used as a standalone software, it comes with RStudio and it is easiest to use it through RStudio.

Online Resources

R

- R For Data Science (https://r4ds.hadley.nz/)
- YARRR! The Pirate's Guide to R (https://bookdown.org/ndphillips/YaRrr/)
- R for the Rest of Us: A Statistics-Free Introduction (https://book.rfortherestofus.com/)

Statistics

- Statistical Inference Via Data Science: R and the Tidyverse (https://moderndive.com/index.html)
- Introduction to Modern Statistics (2e) (https://openintro-ims.netlify.app/)
- Econometrics with R (https://www.econometrics-with-r.org/)
- Handbook of Regression Modeling in People Analytics https://peopleanalytics-regression-book.org/index.html
- A User's Guide to Statistical Inference and Regression (https://mattblackwell.github.io/gov2002-book/)
- Collection of Introductory Statistics Resources (https://stats.libretexts.org/Bookshelves/Introductory_Statistics)
- Applied Statistics with R https://book.stat420.org/
- A Language, not a Letter: Learning Statistics in R (https://ademos.people.uic.edu/index.html)
- University of Virginia StatLab Articles (https://library.virginia.edu/data/articles)
- Data Analysis Examples, OARC Stats, UCLA (https://stats.oarc.ucla.edu/other/dae/)

Causal Inference

- Causal Inference in R (https://www.r-causal.org/)
- Statistical Tools for Causal Inference (https://chabefer.github.io/STCI/)

Visualization & Communication

- Telling Stories with Data (https://tellingstorieswithdata.com/)
- R Graphics Cookbook, 2nd edition (https://r-graphics.org/)
- Fundamentals of Data Visualization (https://clauswilke.com/dataviz/index.html)

Miscelleneous

- Download R from homepage: http://cran.r-project.org/index.html
- RStudio, recommended for using R: https://www.rstudio.com/products/rstudio/download/
- Free Manual: https://cran.r-project.org/doc/manuals/R-intro.pdf
- Resources via UCLA: https://stats.idre.ucla.edu/r/
- Many have found stack overflow invaluable: https://stackoverflow.com/questions/tagged/r

Email Policy

I expect you to have lots of questions throughout the course. Please ask them! When you have a question about course material, it is likely that others in the class have the same question. It is also likely that someone else in the class can answer the question faster or better. Therefore, students are encouraged to ask their questions on Canvas's discussion section, publicly. For questions that are relevant for you but not for other students or for other types of concerns, please feel free to email me. In most cases, you can expect a response within one business day.

Incomplete Grade Policy

For your reference, a copy of the essential sections of the UMass policy regarding Incomplete grades has been provided below. (More information can be found on page 28 of the following document: https://www.umass.edu/registrar/sites/default/files/academicregs.pdf):

> "Students who are unable to complete course requirements within the allotted time because of severe medical or personal problems may request a grade of Incomplete from the instructor of the course. Normally, incomplete grades are warranted only if a student is passing the course at the time of the request and if the course requirements can be completed by the end of the following semester. Instructors who turn in a grade of "INC" are required to leave a written record of the following information with the departmental office of the academic department under which the course is offered: (1) the percentage of work completed, (2) the grade earned by the student on the completed work, (3) a description of the work that remains to be completed, (4) a description of the method by which the student is to complete the unfinished work, and (5) the date by which the

work is to be completed. In the case of an independent study where the entire grade is determined by one paper or project, the instructor should leave with the department information pertaining to the paper or project, which will complete the course. To avoid subsequent misunderstanding it is recommended that the student also be provided with a copy of this information.

Grades of Incomplete will be counted as F's until resolved. If not resolved by the end of the following semester, they will automatically be converted to an F if taken before Fall 2004, to an IF if taken thereafter. Faculty wishing to extend this deadline must write to the Registrar's Office stipulating a specific date by which the incomplete will be completed."

AI Policy

AI-powered tools for generating and manipulating text, code, image and alike are fast becoming ubiquitous. ChatGPT is the most recent well-known example. GitHub Copilot is another tool that is especially relevant for coding. There are many more. Learning how to use these tools effectively is likely to become an important skill soon. You can use these tools in your work, but please remember that AI can make mistakes, plagiarize, or veer off the task you gave it. AI is best used as an assistant, rather than as a replacement for yourself. This usually means giving detailed, specific prompts and having AI help with bottlenecks in a project rather than trying to have it do the whole job by itself for you. If I do mention AI use as a problem in feedback on your work, it will not be because of AI use per se (as mentioned, that is fine and allowed) but because of these reasons. With these caveats, feel free to use it.

Academic Honesty

Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. The procedures outlined below are intended to provide an efficient and orderly process by which action may be taken if it appears that academic dishonesty has occurred and by which students may appeal such actions. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent. For more information about what constitutes academic dishonesty, please see the Dean of Students' website: http://umass.edu/dean_students/codeofconduct/acadhonesty/

Statement on Disabilities

The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier free campus. If you are in need of accommodation for a documented disability, register with Disability Services to have an accommodation letter sent to your faculty. It is your responsibility to initiate these services and to communicate with faculty ahead of time to manage accommodations in a timely manner. For more information, consult the Disability Services website.

Title IX Statement

In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: https://www.umass.edu/titleix/resources. You do not need to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day / 7 days a week / 365 days a year at the SASA Hotline 413-545-0800.

Course Schedule

Week 1 (Sept 3 and 5): Introduction & Descriptive Statistics

- **Summary:** We will start by thinking about the kinds of research questions we can answer using data and statistics. This is important not least because you will need to find and answer one in the final project. We will also cover descriptive statistics such as mean, median, mode, etc. which we use when describing a sample.
- **Required**:
  - [**Effect**] Ch. 2 Research Questions
  - [**RaOS**] Ch. 1 Overview, Ch. 2 Data and Measurement
  - [**LSR**] Ch. 2.2. Scales of measurement, Ch. 5 Descriptive Statistics (only 5.1-5.3 and 5.6-5.7)
- **Recommended:**
  - Murphy, Kevin R. 2021. "In praise of Table 1: The importance of making better use of descriptive statistics." *Industrial and Organizational Psychology* 14 (4): 461–477

Week 2 (Sept 10 and 12): Probability

- **Summary:** A fundamental goal of statistics is inference: being able to make statements about a population using a sample from that population. This week we cover probability theory, which is used by both frequentist and Bayesian statistical traditions—in different ways—to make statistical inference.
- **Required:**
  - [**LSR**] Ch. 9 Introduction to Probability
  - [**RaOS**] Ch. 3 Some basic methods in mathematics and probability
  - [**BayesRules**] Ch. 2 Bayes' Rule
- **Recommended:**
  - Stark, Philip B, et al. 2003. "What is the chance of an earthquake." *NATO Science Series IV: Earth and Environmental Sciences* 32:201–213
  - Univariate Distribution Relationships (interactive webpage at `https://www.math.wm.edu/ leemis/chart/UDR/UDR.html`)
  - Watch: Binomial distributions — Probabilities of probabilities, part 1 (`https://www.youtube.com/watch?v=8idr1WZ1A7Q`)
  - Watch: Why "probability of 0" does not mean "impossible" — Probabilities of probabilities, part 2 (`https://www.youtube.com/watch?v=ZA4JkHKZM50`)

★ **Assignment 1** due on Friday, Sept 13.

Week 3 (Sept 17 and 19): Statistical Inference

- **Summary:** Armed with probability theory, we will look into how both frequentists and Bayesians go about doing statistical inference. Frequentism focuses on how samples from a population would behave for *given* values of parameters. Bayesians express their beliefs about the value of a parameter using probability theory (prior), then use data to update that belief (posterior). This difference in whether beliefs about parameters can be probabilistic is the basis of the difference between the two traditions.
- **Required:**
  - [**LSR**] Ch. 10 Estimating unknown quantities from a sample
  - [**RaOS**] Ch. 4 Statistical Inference (only 4.1-4.3)
  - [**Rethinking**] Ch 2. Small Worlds and Large Worlds
- **Recommended:**
  - [**BayesRules**] Ch. 3 The Beta-Binomial Bayesian Model
  - Bradley, Valerie C, et al. 2021. "Unrepresentative big surveys significantly overestimated US vaccine uptake." *Nature* 600 (7890): 695–700
  - Nagele, Peter. 2003. "Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals." *British Journal of Anaesthesia* 90 (4): 514–516
  - Watch: But what is the Central Limit Theorem? (`https://www.youtube.com/watch?v=zeJD6dqJ5lo`)

  - Watch: Bayes theorem, the geometry of changing beliefs (`https://www.youtube.com/watch?v=HZGCoVF3YvM`)

★ **Assignment 2** due on Friday, Sept 20.

Week 4 (Sept 24 and 26): Hypothesis Testing

- **Summary:** Hypotheses are statements about a population. We now know how to make inferences about (or "form beliefs about," if you will) a population using samples. Hence, we can now check or "test" whether our hypotheses are supported by our data ("sample"). Frequentist approach is based on whether we can reject a "null hypothesis" in favor of alternative; the Bayesian approach allow us to directly compare the believability of two hypotheses with Bayes factors.
- **Required:**
  - [**LSR**] Ch. 11 Hypothesis Testing, Ch. 13 Comparing two means (only 13.1-13.7), Ch. 17 Bayesian statistics (only 17.2 and 17.5),
  - [**RaOS**] Ch. 4 Statistical Inference (remainder from last week, i.e 4.4-4.7)
- **Recommended:**
  - Simonsohn, Uri. 2014. *Confidence Intervals Don't Change How We Think about Data,* October. http://datacolada.org/28
  - Gill, Jeff. 1999. "The insignificance of null hypothesis significance testing." *Political research quarterly* 52 (3): 647–674
  - Lempert, Richard. 2009. "The Significance of Statistical Significance:: Two Authors Restate An Incontrovertible Caution. Why A Book?" *Law & Social Inquiry* 34 (1): 225–249
  - Watch: The medical test paradox, and redesigning Bayes' rule (`https://www.youtube.com/watch?v=lG4VkPoG3ko`)

★ **Assignment 3** due on Friday, Sept 27.

Week 5 (Oct 1 and 3): Simple Linear Regression

- **Summary:** We get started with regression! In the "simple" case, we have one predictor predicting a numeric/continuous outcome.
- **Required:**
  - [**RaOS**] Ch. 6 Background on regression modeling, Ch. 7 Linear Regression with a single predictor, Ch. 8 Fitting Regression Models
- **Recommended:**
  - [**Rethinking**] Ch. 4. Geocentric Models
  - [**BayesRules**] Ch. 9 Simple Normal Regression
  - Alley, Joshua. 2021. *An Open Collection of Political Science Research with OLS Models and Cross-Sectional Data.* https://polmeth.org/blog/open-collection-political-science-research-ols-models-and-cross-sectional-data

★ **Assignment 4** due on Friday, Oct 4. This assignment will check in on your Final Project progress. By this stage, you should at least have one research question, one hypothesis that you want to test, and ideas about what kind of data you can use to test them. Details in Canvas.

Week 6 (Oct 8 and 10): Multiple Regression I

- **Summary:** After doing descriptive statistics in the beginning, then statistical inference, we are adding another layer to our analysis: causal inference. A primary (but not the only) reason people use linear regression is to isolate the causal effect of a variable on an outcome. Directed Acyclic Graphs (DAGs) provide a principled framework for choosing control variables in a way that can isolate and estimate causal effects. When there are more than one predictor in a regression, it is called "multiple regression."
- **Required:**
  - [**Effect**] Ch 5 - Identification
  - [**Mixtape**] Ch 3 - Directed Acyclic Graphs
- **Recommended:**
  - [**Effect**] Chapters 6-9
  - [**Rethinking**] Chapters 5 & 6
  - Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2022. "A crash course in good and bad controls." *Sociological Methods & Research,* 00491241221099552
  - Wysocki, Anna C, Katherine M Lawson, and Mijke Rhemtulla. 2022. "Statistical control requires causal justification." *Advances in Methods and Practices in Psychological Science* 5 (2): 25152459221095823
  - Westreich, Daniel, and Sander Greenland. 2013. "The table 2 fallacy: presenting and interpreting confounder and modifier coefficients." *American journal of epidemiology* 177 (4): 292–298
  - Keele, Luke, Randolph T Stevenson, and Felix Elwert. 2020. "The causal interpretation of estimated associations in regression models." *Political Science Research and Methods* 8 (1): 1–13
  - Berk, Richard. 2010. "What you can and can't properly do with regression." *Journal of Quantitative Criminology* 26:481–487

★ **Assignment 5** due on Friday, Oct 11.

Week 7 (Oct 17): Multiple Regression II

—No class on Tuesday per academic calendar.—

- **Summary:** Regression models, beside continuous predictors, can also have dichotomous or categorical predictors, and multiplicative interaction terms. This week we cover how to interpret them.
- **Required:**
  - [**RaOS**] Ch 10. Linear regression with multiple predictors
- **Recommended:**
  - Beck, Nathaniel, and Fundación Juan March. 2010. "Making regression and related output more helpful to users." *The Political Methodologist* 18 (1): 4–9
  - Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. "Understanding interaction models: Improving empirical analyses." *Political analysis* 14 (1): 63–82

★ **Assignment 6** due on Friday, Oct 18.

Week 8 (Oct 22 and 24): Transformations

- **Summary:** Linear regression imposes a functional form restriction on the predictors: the outcome is a linear combination of the predictors + error. We can use logarithmic and polynomial transformations of the predictors to more accurately model some relationships while staying in the linear framework.
- **Required:**
  - [**RaOS**] Ch. 12 Transformations and regression
  - Benoit, Kenneth. 2011. *Linear regression models with logarithmic transformations.* https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf
- **Recommended:**
  - Rittmann, Oliver, Marcel Neunhoeffer, and Thomas Gschwend. 2023. "How to improve the substantive interpretation of regression results when the dependent variable is logged." *Political Science Research and Methods,* 1–9
  - Chen, Jiafeng, and Jonathan Roth. 2024. "Logs with zeros? Some problems and solutions." *The Quarterly Journal of Economics* 139 (2): 891–936

★ **Assignment 7** due on Friday, Oct 25.

Week 9 (Oct 29 and 31): Assumptions, Diagnostics, and Model Evaluation

- **Summary:** The validity of inferences from a linear regression model is based on assumptions like normality of errors, independence of observations, equal variance of compared groups (homoskedasticity). We will look into tools that can be used to check whether these assumptions hold. We will also look into evaluating the model fit as a whole, for example using R-squared (about which there is considerable debate, if you check the recommended readings!).
- **Required:**
  - [**RaOS**] Ch. 11 Assumptions, diagnostics, and model evaluation
- **Recommended:**
  - [**BayesRules**] Ch. 10 Evaluating Regression Models
  - Lewis-Beck, Michael S, and Andrew Skalaban. 1990. "When to use R-squared." *The Political Methodologist* 3 (2): 9–11
  - King, Gary. 1990. "When not to use R-squared." *The Political Methodologist* 3 (2): 11–12
  - Ford, Clay. *Is R-squared Useless? — UVA Library.* https://library.virginia.edu/data/articles/is-r-squared-useless

★ **Assignment 8** due on Friday, Nov 1. This assignment is our second check in point for the final project. At this point you should have integrated / addressed feedback you received at the first check in point, created a DAG, and ran a one regression model that tests your hypothesis. Details in Canvas.

Week 10 (Nov 7): Logistic Regression I

—No class on Tuesday (Election Day)—

- **Summary:** Logistic regression is the quintessential model for binary outcomes. Predictors are still linearly combined, but the output is in log-odds and needs to be transformed into a probability using the sigmoid function.
- **Required:**
  - [**RaOS**] Ch. 13. Logistic regression
- **Recommended:**
  - [**BayesRules**] Ch. 13 Logistic Regression (13.1-13.3)
  - Mood, Carina. 2010. "Logistic regression: Why we cannot do what we think we can do, and what we can do about it." *European sociological review* 26 (1): 67–82
  - King, Gary, and Langche Zeng. 2001. "Logistic regression in rare events data." *Political analysis* 9 (2): 137–163
  - Ford, Clay. *Simulating a Logistic Regression Model — UVA Library.* https://library.virginia.edu/data/articles/simulating-a-logistic-regression-model

Week 11 (Nov 12 and 14): Logistic Regression II

- **Summary:** More on logistic regression, especially interpretation and model fit evaluation.
- **Required:**
  - [**RaOS**] Ch. 14. Working with logistic regression
- **Recommended:**
  - [**BayesRules**] Ch. 13 Logistic Regression (13.4-13.6)
  - Ford, Clay. *Getting Started with Multinomial Logit Models — UVA Library.* https://library.virginia.edu/data/articles/getting-started-with-multinomial-logit-models
  - Ford, Clay. *Fitting and Interpreting a Proportional Odds Model — UVA Library.* https://library.virginia.edu/data/articles/fitting-and-interpreting-a-proportional-odds-model

★ **Assignment 9** due on Friday, Nov 15.

Week 12 (Nov 19 and 21): Generalized Linear Models

- **Summary:** Logistic regression was a type of "generalized linear model," one of many generalizations of the linear model that can handle various distributions for the outcome variable. We now consider more GLMs, especially Poisson and Negative Binomial Regression for count data.
- **Required:**
  - [**RaOS**] Ch. 15. Other generalized linear models
- **Recommended:**
  - [**BayesRules**] Ch. 12 Poisson & Negative Binomial Regression

- Zeileis, Achim, Christian Kleiber, and Simon Jackman. 2008. "Regression models for count data in R." *Journal of statistical software* 27 (8): 1–25. https://www.jstatsoft.org/article/view/v027i08
- Ford, Clay. *Getting Started with Hurdle Models — UVA Library.* https://library.virginia.edu/data/articles/getting-started-with-hurdle-models
- Ford, Clay. *Getting Started with Negative Binomial Regression Modeling — UVA Library.* https://library.virginia.edu/data/articles/getting-started-with-negative-binomial-regression-modeling

★ **Assignment 10** due on Friday, Nov 22.

Week 13 (Nov 26): Research Design Considerations

—No class on Thursday (Thanksgiving)—

- **Summary:** We now take a step back and think about bigger questions: What types of design decisions should we think about before even collecting data? Why are published research findings biased toward finding large effects—i.e. effects presented as larger in scientific publications than they are in reality? We will also touch on a little bit on the "replication crisis" in academic research.
- **Required:**
  - [**RaOS**] Ch. 16. Design and sample size decisions
- **Recommended:**
  - Larsen, Erik Gahner. 2022. *Effect sizes in political science,* December. Accessed August 20, 2024. https://erikgahner.dk/2022/effect-sizes-in-political-science/

Week 14 (December 3 and 5): Review & Catch Up

- **Summary:** This is a week for us to catch up on anything we missed so far. We can also use it as open office hours for questions about the final project or anything else. We have one recommended reading that I hope would be helpful when thinking about framing your final project papers.
- **Recommended:**
  - Lundberg, Ian, Rebecca Johnson, and Brandon M Stewart. 2021. "What is your estimand? Defining the target quantity connects statistical evidence to theory." *American Sociological Review* 86 (3): 532–565

Week 15 (December 10): Review & Catch Up

—Tuesday is the last day of classes.—

- **Summary:** Same as last week.

★ ★ **Final Project** due on December 13, Friday.

Further Reference

Statistics / Regression:

- Agresti, Alan. 2017. *Statistical Methods for the Social Sciences.* 5th edition. Boston: Pearson, January. ISBN: 978-0-13-450710-1
- Fox, John, and Sanford Weisberg. 2018. *An R Companion to Applied Regression.* 3rd edition. Los Angeles London New Delhi Singapore Washington, DC Melbourne: SAGE Publications, Inc, October. ISBN: 978-1-5443-3647-3
- Shalizi, Cosma Rohilla. *The Truth about Linear Regression.* https://www.stat.cmu.edu/~cshalizi/TALR/TALR.pdf
- Weisberg, Sanford. 2013. *Applied Linear Regression.* 4th edition. Hoboken, NJ: Wiley, December. ISBN: 978-1-118-38608-8

Econometrics:

- Greene, William. 2017. *Econometric Analysis.* 8th edition. New York, NY: Pearson, March. ISBN: 978-0-13-446136-6
- Kennedy, Peter. 2008. *A Guide to Econometrics. 6th edition.* 6th edition. Malden, Mass.: Wiley-Blackwell, February. ISBN: 978-1-4051-8257-7
- Stock, James, and Mark Watson. 2018. *Introduction to Econometrics.* 4th edition. New York, NY: Pearson, November. ISBN: 978-0-13-446199-1
- Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach.* 7th edition. Boston, MA: Cengage Learning, January. ISBN: 978-1-337-55886-0

Bayesian:

- Clayton, Aubrey. 2021. *Bernoulli's Fallacy: Statistical Illogic and the Crisis of Modern Science.* Pages: 368 Pages. Columbia University Press, August. ISBN: 978-0-231-55335-3
- Gelman, Andrew, et al. 2013. *Bayesian Data Analysis.* 3rd edition. Boca Raton: Chapman / Hall/CRC, November. ISBN: 978-1-4398-4095-5. http://www.stat.columbia.edu/~gelman/book/BDA3.pdf
- Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* 1st edition. Cambridge: Cambridge University Press, December. ISBN: 978-0-521-68689-1

Causal Inference:

- Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect.* With French flaps edition. Princeton, NJ Oxford: Princeton University Press, December. ISBN: 978-0-691-15284-4
- Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect.* 1st edition. New York: Basic Books, May. ISBN: 978-0-465-09760-9

Probability:

- Ross, Sheldon. 2019. *A First Course in Probability, Global Edition.* 10th edition. Harlow, United Kingdom: Pearson, October. ISBN: 978-1-292-26920-7