PLSC 502: "Statistical Methods for Political Research"

Fall 2024

Professor Christopher Zorn
Department of Political Science
Pennsylvania State University
E-mail: zorn@psu.edu
Mondays, 9:00 a.m. - 12:00 p.m.
Willard Building, Room 069

Course Description

This is the first course in quantitative methods in Penn State's political science Ph.D. program. The course is an introduction to the use of statistics for the social sciences, and political science in particular. There are three main goals of the course: to teach students to read and understand quantitative analyses in published and unpublished work, to provide them with the skills necessary to begin conducting their own quantitative analyses, and to lay the foundation for future courses in quantitative methods. This means learning the fundamentals of data collection, organization, and management; measurement; data visualization and display; univariate, bivariate, and multivariate descriptive statistics; sampling; statistical inference (including essentials of probability and distribution theory); and univariate, bivariate, and multivariate hypothesis testing.

There are no formal prerequisites for this class, mathematical or otherwise, and the only expectation is that students are familiar with high-school level algebra and calculus. The summer "boot camp" should have provided you with more-or-less everything you need for this course. That said, much of the material in this course is somewhat technical. While I have chosen readings that present the material as clearly and with as little jargon as possible, most of it will still require several readings to fully comprehend.

This syllabus is designed to provide an overview to the course. Note that all course materials (including this syllabus, slides, notes, data, code, homework exercises, etc.) will be available on the course Github repo, at https://github.com/PrisonRodeo/PLSC502-2024-git. Clickable links are printed in Penn State blue.

Course Readings

Required Text/Materials

Agresti, Alan. 2017. *Statistical Methods for the Social Sciences*, 5th Ed. Upper Saddle River, NJ: Prentice-Hall.

The 4th Edition of Agresti (with Finlay) is also acceptable, and likely available at a substantially reduced cost. You should purchase this book through Allibris, Amazon, or the like, and do so rela-

tively soon. The first few assigned readings from the text are also available in the Github repository, in case the arrival of your book is delayed. Additional readings, as necessary, be available on the course Github repository, and/or through JSTOR.

Strongly **Recommended**

Verzani, John. 2014. *Using R For Introductory Statistics*, 2nd Ed. Boca Raton, FL: Chapman & Hall. (There's a PDF of the first edition available on the web – e.g., here – which is also pretty good.)

or perhaps:

Thulin, Mans. 2021. *Modern Statistics With R*. Eos Chasma Press. (Very "tidy;" see below.)

A Few Other Useful/Recommended Readings

Math/Statistics/Computing Books

Abelson, Robert P. 1995. *Statistics as Principled Argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Altman, Micah, Jeff Gill and Michael McDonald. 2003. *Statistical Computing for the Social Scientist*. New York: Wiley.

Bohrnstedt, George W., and David Knoke. 2004. *Statistics for Social Data Analysis*, 4th Ed. New York: Wadsworth. (Excerpts below are from the third (1994) edition).

Cleveland, William S. 1985. *The Elements of Graphing Data*. Monterey, CA: Wadsworth.

Cuzzort, R. P., and James S. Vrettos. 1996. *The Elementary Forms of Statistical Reason*. New York: St. Martins.

DeGroot, Morris H., and Mark J. Schervish. 2002. *Probability and Statistics*, 3rd Ed. New York: Addison-Wesley.

Freedman, David, Robert Pisani, and Roger Purves. 2007. *Statistics*, 4th Ed. ew York: W. W. Norton.

Evans, Merran, Nicholas Hastings and Brian Peacock. 2000. *Statistical Distributions*, 3rd Ed. New York: Wiley.

Gill, Jeff. 2006. *Essential Mathematics for Political and Social Research*. Cambridge: Cambridge University Press.

Larsen, Richard J., and Morris L. Marx. 2006. *An Introduction to Mathematical Statistics and its Applications*. 5th ed. Upper Saddle River, NJ: Duxbury.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

Wackerly, Dennis D., William Mendelhall III, and Richard L. Scheaffer. 2008. *Mathematical Statistics with Applications*, 7th Ed. New York: Duxbury.

Wonnacott, Thomas H., and Ronald J. Wonnacott. 1990. *Introductory Statistics*, 5th ed. New York: John Wiley and Sons.

Books on R (a sample of a very long list; many of these are available electronically via the Penn State Libraries)

Crawley, Michael J. 2014. *Statistics: An Introduction Using R*, 2nd Ed. New York: Wiley.

Daalgard, Peter. 2008. *Introductory Statistics With R*, 2nd Ed. New York: Springer.

Everitt, Brian S., Torsten Hothorn. 2014. *A Handbook of Statistical Analyses Using R*, 3rd Ed. Boca Raton, FL: Chapman & Hall.

Mailund, Thomas. 2022. *Beginning Data Science in R: Data Analysis, Visualization, and Modelling for the Data Scientist*. New York: APress.

Maindonald, John, and John Braun. 2013. *Data Analysis and Graphics Using R: An Example-Based Approach*, 3rd Ed. New York: Cambridge University Press.

Murrell, Paul. 2019. *R Graphics*, 3rd Ed. Boca Raton, FL: Chapman & Hall. (Website is here).

Navarro, Danielle. 2021. *Learning Statistics With R* (and the `lsr` R package on CRAN).

A Few Other R Resources (mostly online)

R Reference Card 2.0 (also in Chinese).

The R Language: A Short Companion.

Robert Kabakoff's Quick-R (really excellent).

Owen, W. J. 2010. The R Guide.

Phillips, Nathaniel D. 2018. YaRrr! The Pirate's Guide to R.

Ricci, Vito. 2005. Fitting Distributions With R.

Ricci, Vito. 2005. R Functions For Regression Analysis.

Santana, Julio Sergio, and Efrain Mateos Farfan. 2014. *El Arte de Programar en R: Un Lenguaje Para la Estadistica*.

Shupinov, Alexay. 2019. *Visual Statistics. UseR!*.

Some Other Useful Resources

The **Political Methodology Section** of the American Political Science Association was created to provide APSA members with an interest in political methodology with a forum in which to meet and discuss ideas. The section publishes a quarterly newsletter (*The Political Methodologist*), a quarterly journal on political methodology (*Political Analysis*), conducts a discussion list on topics relating to political methodology, and maintains an extensive electronic archive of papers, accessible via their homepage.

Also, the **Inter-University Consortium for Political and Social Research** (ICPSR), at the University of Michigan, maintains an extensive archive of data in the social and behavioral sciences. Much of it is accessible via their homepage. They also offer a summer program in quantitative (and other) methods, which we'll discuss a bit later in the year. A similar set of programs (in which I teach) – the **Global School in Empirical Research Methods (GSERM)** – is offered by the University of St. Gallen.

The **Dataverse Project** "increases scholarly recognition and distributed control for authors, journals, archives, teachers, and others who produce or organize data; facilitates data access and analysis for researchers and students; and ensures long-term preservation whether or not the data are in the public domain." It's a repository of all sorts of potentially useful stuff for quantitative researchers.

Grading

Grading will be based on **ten** *numbered*, more-or-less weekly homework exercises (50 points each) and a final paper/poster presentation (500 points). In most instances, exercises will be due

seven days from being assigned (that is, they will be assigned on Monday and typically due the following Monday). Numbered homework exercises will generally involve analysis and discussion/interpretation of actual data, using statistical computer software (see below). In addition, in weeks where there are no numbered homework assignments, you will be given *lettered* assignments ("A," "B," etc.). Those assignments are also required, but are ungraded. Lettered homework assignments will be focused on improving your skills using the R statistical language. Feel free to work on all of the assignments in groups of two or three, but *you must write up all assignments individually*. More details for the homework assignments and the final project will be announced in class.

Also, note that all homework exercises and the final paper should be submitted electronically, as PDF files. Note that **only PDF files will be accepted**, without exception. If you do not yet know how to create a PDF file from a document you've created, please go learn; it is not difficult.

Software, Statistical and Otherwise

You are welcome to make use of whatever statistical software you choose to complete the homework exercises, so long as the manner by which your results are generated and conclusions reached are transparent. However, due to the limits of instructor and TA time and patience, we will support only two (really, 1.5) software packages. Both are available on the machines in the political science computing labs.

R

R is a statistical environment and high-level programming language for data analysis and display. It is effectively the GNU version of the S language; as such, it is free (both as in speech and as in beer) and open source; the current stable version is 4.3.1. R is an *object-oriented* language; it operates mostly by assigning values to objects in the workspace. In the notes, handouts, etc., R commands will be in `monospaced font`, and will be preceded by a caret (">"):

```
> xtab<-table(Y,X)
```

The Comprehensive R Archive Network (CRAN) is the go-to spot for all things R. I cannot begin to list all the R-related resources available on the web; for a sampling, see the Useful R Resources document on the course Github repo.

All in-class examples, code, graphics, and so forth will use R.

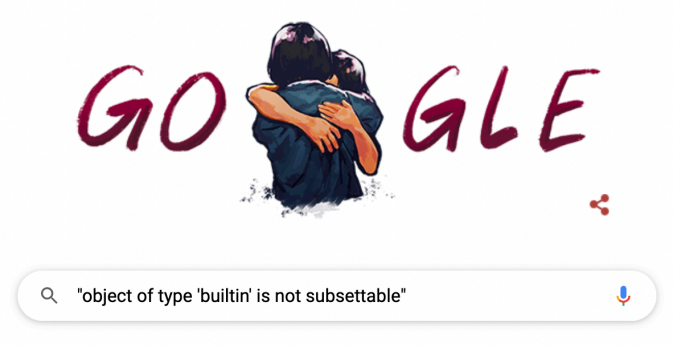
Three other R-related things:

- Don't use R itself; use RStudio / Posit, in either its desktop or cloud versions. It is an IDE for R that makes using R (roughly) 17,311 times easier and more intuitive. It also integrates

with a number of other valuable R-connected things like `rmarkdown`, Shiny, `knitr`, etc.

- For those new to R: When – not if – you encounter a problem using R, avoid the impulse to try to solve it yourself. This is probably not strictly true, but I always begin my R debugging on the premise that there have been many, many people who have created/encountered exactly the same problem as I am having, that they have likely solved it, and that there is quite possibly some documentation on the web of how they accomplished that. Accordingly, a good R debugging routine almost always looks like this:

    1. Encounter a problem with R,
    2. Try for $\leq$ 5-10 minutes to solve it yourself,
    3. Google the details of the (inevitably cryptic) error message R gives you, in quotes; e.g.:



    4. Find the solution, likely posted sometime in 2014 on StackOverflow.[1]
    5. Alternatively, consider familiarizing yourself with one of the good large-language-model-based code helpers (e.g., CoPilot, SuperMaven, etc.) or just ask a general-use LLM (like ChatGPT) to help you out. If/when you do so, bear a few things in mind:
        - Many of these tools are better at more orthodox languages from computer / information science (e.g., Python, Java, etc.) than they are at writing R code.
        - Relatedly: Such tools are usually good either for very basic things, or for getting a "starter" block of code that you can then fix, modify, etc. In my experience, it's pretty rare for (say) ChatGPT to give you the ideal code for your task at hand the very first time you ask for it.

- There is a suite of packages for data management and graphics known collectively in the R world as the "tidyverse." Created by Hadley Wickham, the tidyverse is... hard to characterize. Taken as nothing more than a set of R packages, they're mostly quite useful, particularly if you find yourself working with nonstandard-for-statisticians data formats and structures (JSON, SQL, etc.).

[1] As we noted previously in "boot camp," StackOverflow is your friend.

At the same time, there are at least three places in the tidyverse where things get weird. The first is the idea of maintaining and working only with "tidy" data, something that in the tidyverse (as the name suggests) is tantamount to a Prime Directive. While it's not a bad principle, it can run afoul of the reality of the kind of data that social scientists analyze, which includes things like temporal variation, network structure, and the like. Second is the tidyverse's relatively recent venture into fundamentally changing the syntactical and object structure of R (via "pipes," "tibbles," and the like). While R has recently integrated some of these sorts of functions, they are not (and should not be confused with) "base" R. Tidy expressions also have a habit of getting changed frequently, which can in turn break existing code, something that is never, ever "fun" in any sense of the word. Last is (to be blunt) the vaguely cult-like manner of some of the tidyverse's inhabitants; I'll just leave that here for now.

In terms of advice, I'd say this: It is, in general, a good idea to rely as little as possible on non-base R packages, "tidy" or otherwise. This means that if you can easily (say) draw a nice looking scatterplot using `plot()`, that's usually a better choice from a replicability, transparency, and longevity point of view than using `ggplot2::ggplot()+geom_point()`. That doesn't mean "never use packages" (which would be insane), only that minimizing reliance on them in situations where base R works well is good practice. In this class, for the most part we will not use `ggplot2` (the tidyverse's main graphics/visualization package) for graphics, but will occasionally use some of its data management functions (e.g., `readr`, `dplyr`, and `stringr`). In and beyond PLSC 502, though, you should feel free to explore the "tidyverse," and to learn and use it as you deem useful for your own purposes.

Stata

Once upon a time, not that long ago, Stata was the most widely-used statistical package in the social sciences. (For reasons mysterious yet completely on-brand, it remains the most widely used statistical software in the discipline of economics.) It is a powerful tool for data management, analysis, and display, and boasts some of the best manuals and on-line help of any existing software package. Stata is commercial software; it is not open source, and is expensive to purchase (though Penn State has site licenses you can use while you're affiliated with the university). The current version of Stata is 18.0, but previous versions (back to v. 14, at least) can also be used for the class. In the exceedingly rare instance when they appear in the class notes, handouts, etc., Stata commands will appear in a fixed-width font and will be preceded by a period (".").

```
. regress Y X
```

Stata newbies may want to check out:

Acock, Alan. 2023. *A Gentle Introduction to Stata*, Revised Sixth Edition. College Station, TX: Stata Press.

and/or Stata's dedicated "new users" page:

`https://www.stata.com/links/resources-for-learning-stata/`.

Beyond this, the Stata homepage is a valuable resource for questions about the Stata statistical software. There are a number of useful Stata references on the web, including Scott Long's page at IU and an excellent Stata "help page" sponsored by UCLA.

Other Software Considerations

In no particular order:

- For many years, your instructor did not have a formally-stated preference for either Stata or R. That time has passed: If you want to do professional-level data analysis in political science, and/or be competitive for jobs in industry, government, and the non-profit sector, learn R, not Stata . Be aware that Stata has a far flatter learning curve than R, which means people tend to gravitate toward it given a choice. But R is far more flexible and powerful, and will likely be more valuable and useful to you in the long run.

- While we won't be doing so in this course, you *may* learn that learning another programming language is valuable to you. As of this writing, the most common one people choose to learn is Python, which is widely used in data science applications in the government, non-profit, and private sectors. But there is also potential value to knowing other modern languages (C/C++, Java/Javascript, Julia, etc.), so keep an open mind.

- I also strongly suggest that you learn to use LaTeX, now, while you have the time (and possibly a markdown language, too). You will be glad you did. The department (and other units on campus) occasionally offers workshops in LaTeX. In addition, if you're struggling with LaTeX and would like some example documents, let me know.

- If you insist on using Microsoft Word (or any other WYSIWYG editing program) for writing assignments, papers, etc., **do not under any circumstances cut and paste graphs from Stata and R into those programs**. Save whatever figures you want to use as `.pdf`, `.png`, `.tif`, or `.jpg` files, and import them into the software.

- Similarly: Document creation and editing software (LaTeX, Word, etc.) are not suitable for distribution of work. Unless the person / people reading them will also be editing them, convert all documents to PDFs before sending / submitting / etc. This includes homework assignments and class papers in PLSC 502.

Obligatory Statement on Academic Integrity

Academic integrity is the pursuit of scholarly activity in an open, honest and responsible manner. Academic integrity is a basic guiding principle for all academic activity at The Pennsylvania State University, and all members of the University community are expected to act in accordance with this principle. Consistent with this expectation, the University's Code of Conduct states that all students should act with personal integrity, respect other students' dignity, rights and property, and help create and maintain an environment in which all can succeed through the fruits of their efforts.

Academic integrity includes a commitment by all members of the University community not to engage in or tolerate acts of falsification, misrepresentation or deception. Such acts of dishonesty violate the fundamental ethical principles of the University community and compromise the worth of work completed by others.

In cases of any violation of academic integrity it is the policy of the Department of Political Science to follow procedures established by the College of the Liberal Arts. More information on academic integrity and procedures followed for violation can be found here.

Obligatory Statement on Accommodations for Disabilities

Penn State welcomes students with disabilities into the University's educational programs. Every Penn State campus has an office for students with disabilities. Student Disability Resources (SDR) website provides contact information for every Penn State campus, at:

http://equity.psu.edu/sdr/disability-coordinator.

For further information, please visit the Student Disability Resources website, at:

http://equity.psu.edu/sdr/.

In order to receive consideration for reasonable accommodations, you must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: See documentation guidelines at:

http://equity.psu.edu/sdr/guidelines.

.

If the documentation supports your request for reasonable accommodations, your campus disability services office will provide you with an accommodation letter. Please share this letter with your instructors and discuss the accommodations with them as early as possible. You must follow this process for every semester that you request accommodations.

Obligatory Statement on Counseling and Psychological Services

Many students at Penn State face personal challenges or have psychological needs that may interfere with their academic progress, social development, or emotional wellbeing. The university offers a variety of confidential services to help you through difficult times, including individual and group counseling, crisis intervention, consultations, online chats, and mental health screenings. These services are provided by staff who welcome all students and embrace a philosophy respectful of clients' cultural and religious backgrounds, and sensitive to differences in race, ability, gender identity and sexual orientation.

Counseling and Psychological Services at University Park (CAPS)
(http://studentaffairs.psu.edu/counseling/): 814-863-0395

Counseling and Psychological Services at Commonwealth Campuses
(http://senate.psu.edu/faculty/counseling-services-at-commonwealth-campuses/)

Penn State Crisis Line (24 hours / 7 days/week): 877-229-6400. Crisis Text Line (24 hours / 7 days/week): Text LIONS to 741741.

Obligatory Statement on Educational Equity and Reporting Bias

Penn State takes great pride to foster a diverse and inclusive environment for students, faculty, and staff. Consistent with University Policy AD29, students who believe they have experienced or observed a hate crime, an act of intolerance, discrimination, or harassment that occurs at Penn State are urged to report these incidents as outlined on the University's Report Bias webpage (http://equity.psu.edu/reportbias/).

Obligatory Statement on Religious Observances

The Religious and Spiritual Observances Calendar is compiled by the Center for Spiritual and Ethical Development in consultation with campus and community religious leaders. It specifies those holy days of the major world religions for which observance may require students to depart from their normal routine at the University. Please note that only those holy days which occur when Penn State classes are in session are listed. This is not, therefore, an exhaustive list of all major holy days in each religious tradition.

Non-Obligatory Statement on Generative AI, Large Language Models, etc.

You're undoubtedly well aware of the existence of large language models (LLMs) – e.g., ChatGPT – and other artificial intelligence (AI) tools for language / image creation. Having been described as everything from making everyone their own version of Tony Stark to a Lovecraftian shoggoth (and most things in between), LLMs are currently creating a sometimes-depressing, sometimes-hilarious panic among faculty in legacy academic disciplines and programs. Most of that panic

revolves around the use of LLMs to "cheat," in the traditional sense: to create work that deceptively gives the impression that the student knows something they do not. Beyond its intrinsically duplicitous nature, such use in a conventional classroom setting gives rise to concerns about equity and (potentially) devalues the experience / credential for other class members.

The other side of the LLM equation is that they are powerful tools for augmenting learning and creating new knowledge. Experience suggests that, in line with other technological advances (the printing press, personal computers, search engines, etc.), it is wiser to adapt to LLMs than to attempt to limit or ban their use. This is especially true in a course like this one, where (a) LLMs are particularly useful tools for learning technical skills (e.g., the R programming language) and (b) the long-term, repeated nature of graduate school creates disincentives for "cheating" in a conventional sense. Accordingly, enrollees in PLSC 502 are welcome to use generative AI tools, such as ChatGPT, to assist them with their work in the course. In doing so, it is important to remember that such AI tools are capable of making errors, and that it is each student's responsibility to verify the information they receive from the such a tool. In addition, any information obtained from a generative AI source must be noted/cited in the student's work, just as they would cite any other source.

Course Schedule

Readings should be completed prior to coming to class on the assigned day. Page numbers are generally for Agresti and Finlay (4th Ed., 2008); the 5th Edition page numbers are similar. Note that we will not, in general, hew closely (or at all) to the readings themselves, other than topically. Links are generally to DOIs or to stable PDFs at JSTOR. I won't assign readings from either Verzani (2014) or Thulin (2021), but students should consult the relevant parts of those texts for software guidance.

August 26: Introduction / Overview + How To Read Tables and Figures

- **Readings**
  - *Required*:
    - Gomez, Brad T., Thomas G. Hansford, and George A. Krause. 2007. "The Republicans Should Pray for Rain: Weather, Turnout, and Voting in U.S. Presidential Elections." *Journal of Politics* 69(3):649-663.
    - Rosh, Robert M. 1987. "Ethnic Cleavage as a Component of Global Military Expenditures." *Journal of Peace Research* 24(1):21-30.
  - *Recommended (as examples)*:
    - Apodaca, Clair, and Michael Stohl. 1999. "United States Human Rights Policy and Foreign Assistance." *International Studies Quarterly* 43(1):185-198
    - Holbrook, Thomas M., and Emily Van Dunk. 1993. "Electoral Competition in the American States." *American Political Science Review* 87(4):955-962

        - Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83(2):557-565.

- No exercise assigned.

September 2: NO CLASS – Labor Day

September 9: Data: Structure, Measurement, and Management

- **Readings**
    - *Required*:
        - Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95:529-546.
        - Agresti and Finlay, pp. 1-7, 11-14.
        - Nagler, Jonathan. 1995. "Coding Style and Good Computing Practices." *The Political Methodologist* 6(2):2-8 (and the follow-up blog post in 2015).
        - Stevens, S. S. 1946. "On the Theory of Scales of Measurement." *Science* 103:677-680.
        - Workman, Samuel. 2020. "Four Principles of Data Collection." *Towards Data Science* blog, May 12, 2020.
    - *Recommended*:
        - Cox, Eli P. 1980. "The Optimal Number of Response Alternatives for a Scale: A Review." *Journal of Marketing Research* 17:407-422.
        - de Jonge, Edwin, and Mark van der Loo. 2013. "An Introduction To Data Cleaning With R." Discussion paper, Statistics Netherlands.
        - Heise, D. R. 2001. "Scaling and Classification in Social Measurement." In *Internattional Encyclopedia of Social and Behavioral Sciences*. Amsterdam: Elsevier.
        - Mitchell, Joel. 1986. "Measurement Scales and Statistics: A Clash of Paradigms." *Psychological Bulletin* 100:398-407.
        - Rossi, Giovanni Battista. 2014. *Measurement and Probability: A Probabilistic Theory of Measurement with Applications*. Springer.
        - Schedler, Andreas. 2012. "Judgment and Measurement in Political Science." *Perspectives on Politics* 10:21-36.

- *Exercise A: Do some things with data...*

September 16: Descriptive Statistics: Graphics

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 31-38.
    - Fox (2008), pp. 28-50.
  - *Recommended*:
    - Wilke, Claus O. 2022. *Fundamentals of Data Visualization*. O'Reilly Media.
    - From Data To Viz (various; this week, see especially here, here, here, and here.)
    - Lane, David. 2022. *Introductory Statistics*. (cc) LibreTexts. Chapter 2.
- *Exercise One: Plotting data.*

September 23: Central Tendency and Variation

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 38-55.
    - McNichol, Daniel. 2018. "On Average, You're Using the Wrong Average: Geometric & Harmonic Means in Data Analysis." *Towards Data Science* blog, January 26, 2018.
  - *Recommended*:
    - Carr, Kareem. 2023. "TEN types of statistical averages." Twitter thread, August 4, 2023.
    - Groeneveld, Richard A., and Glen Meeden. 1977. "The Mode, Median, and Mean Inequality." *The American Statistician* 31(3):120-21.
    - Lane, David. 2022. *Introductory Statistics*. (cc) LibreTexts. Chapter 3.
    - Weisberg, Herbert F. 1992. *Central Tendency and Variation*. Newbury Park, CA: Sage Publications.
- *Exercise Two: Central tendency and variation.*

September 30: Probability and Random Variables

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 73-85.
    - Fox (2008), Appendix D., pp. 65-74.
  - *Recommended*:
    - Grinstead, Charles M., and J. Laurie Snell. 2006. *Introduction to Probability*, 2nd Ed.. American Mathematical Society.
    - Lane, David. 2022. *Introductory Statistics*. (cc) LibreTexts. §5.1-5.6.
    - Rudas, Tamás. 2004. *Probability Theory: A Primer*. Newbury Park, CA: Sage Publications.
    - Stark, B. and A. Freedman. 2016. "What is the Chance of an Earthquake?" Technical Report 611, Department of Statistics, University of California - Berkeley.
- *Exercise B: More data things...*

October 7: Probability Distributions

- **Readings**
  - *Required*:
    - Fox (2008), Appendix D., pp. 75-86.
    - Univariate Distribution Relationships (interactive website at `http://www.math.wm.edu/~leemis/chart/UDR/UDR.html`).
  - *Recommended*:
    - Chen, William, and Joe Blitzstein. 2015. *Probability Cheatsheet 2.0*.
    - Evans, Merran, Nicholas Hastings and Brian Peacock. 2000. *Statistical Distributions*, 3rd Ed. New York: Wiley.
    - Thomopoulos, Nick T. 2017. *Statistical Distributions: Applications and Parameter Estimates*. New York: Springer.
    - The Probability Distributome Project on the web.
- *Exercise Three: Distributions*.

October 14: Randomization, Sampling, and Sampling Distributions

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 15-25, 58-61, 85-99.
    - Meng, Xiao-Li. 2018. "Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 U.S. Presidential Election." *Annals of Applied Statistics* 12:685-726.
  - *Recommended*:
    - Bradley, Valerie C., Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao-Li Meng, and Seth Flaxman. 2021. "Unrepresentative Big Surveys Significantly Overestimated U.S. Vaccine Uptake." *Nature* 600: 695-700.
    - Kaltom, Graham. 1983. *Introduction to Survey Sampling*. Newbury Park, CA: Sage Publications.
    - Thompson, Steven K. 2002. *Sampling*, 2nd Ed. New York: Wiley.
- *Exercise Four: Sampling*.

October 21: Estimation and Estimators

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 107-109.
    - Fox (2008), Appendix D., pp. 89-92.
    - Knoblauch, Kenneth, and Laurence T. Maloney. 2013. *Modeling Psychophysical Data in R*, Appendix B, pp. 335-346.
  - *Recommended*:
    - Fisher, R. A. 1922. "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London, A*, 222:309-368.
- *Exercise C: Even More Data Things*

October 28: Statistical Inference

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 109-123, 147-174.

- *Recommended*:
  - 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." *Law and Social Inquiry* 34:225-249.
  - Pomeranz, Janet Bellcourt. 1982. "Confidence in Confidence Intervals." *Mathematics Magazine* 55(1):12-18.
  - Sekhon, Jasjeet. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals." *Perspectives on Politics* 2:281-293.

- *Exercise Five: Estimation and Inference*.

November 4: Two-Group Comparisons + Statistical Power

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 183-209.
    - Gelman, Andrew, and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9:641-651.
    - Larsen, Erik Gahner. 2022. "Effect Sizes in Political Science." Blog post, December 31, 2022.
  - *Recommended*:
    - Arel-Bundock, Vincent, Ryan C. Briggs, Hristos Doucouliagos, Marco M. Aviña, and T.D. Stanley. 2023. "Quantitative Political Science Research is Greatly Underpowered." Working paper (OSF preprints).
    - Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, Second Edition. Mahwah, NJ: Lawrence Erlbaum.
    - Wood, Graham R., and David J. Saville. 2002 . "A New Angle on the t-Test." *The Statistician* 51(1):99-104.
    - Zaller, John. 2002. "The Statistical Power of Election Studies to Detect Media Exposure Effects in Political Campaigns." *Electoral Studies* 21:297-329.

- *Exercise Six: Conduct two-group comparisons*.

November 11: Measures of Association

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 221-246, 255-259, 269-273.
    - Bollen, Kenneth A., and Kenney H. Barb. 1981. "Pearson's r and Coarsely Categorized Measures." *American Sociological Review* 46:232-239.
    - Khamis, Harry. 2008. "Measures of Association: How To Choose?" *Journal of Diagnostic Medical Sonography* 24:155-162.
  - *Recommended*:
    - Bohrnstedt and Knoke (1994), pp. 155-181.
- *Exercise Seven: Measures of association*.

November 18: Linear Regression, I

- **Readings**
  - *Required*:
    - Agresti and Finlay, pp. 259-279.
  - *Recommended*:
    - Bohrnstedt and Knoke (1994), pp. 191-204.
    - Kahane, Leo H. 2001. *Regression Basics*. Thousand Oaks, CA: Sage Publications.
    - Alley, Joshua. 2021. "An Open Collection of Political Science Research with OLS Models and Cross-Sectional Data." *The Political Methodologist* blog, September 8, 2021.
- *Exercise Eight: Linear regression*.

November 25: NO CLASS - Thanksgiving Break

December 2: Linear Regression, II

- **Readings**
  - *Required*:

        - Agresti and Finlay, pp. 280-289.
        - Westreich, Daniel, and Sander Greenland. 2013. "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients." *American Journal of Epidemiology* 177:292-298.
        - Keele, Luke, Randolph T. Stevenson, and Felix Elwert. 2020. "The Causal Interpretation of Estimated Associations in Regression Models." *Political Science Research and Methods* 8:1-13.
    - *Recommended*:
        - Kahane, Leo H. 2001. *Regression Basics*. Thousand Oaks, CA: Sage Publications.
        - Xiao, Qinyu, et al. 2023. *Effect Sizes and Confidence Intervals Guide*. https://doi.org/10.17605/OSF.IO/D8C4G

- *Exercise Nine: More linear regression.*

December 9: Miscellany: Missing Data, Bayes, Multivariate Statistics...

- **Readings**
    - *Required*:
        - Agresti and Finlay, pp. 301-314.
        - Fox (2008), Appendix D., pp. 101-106.
    - *Recommended*:
        - Daniel, Rhian M., Michael G. Kenward, Simon N. Cousens, and Bianca L. De Stavola. 2012. "Using Causal Diagrams to Guide Analysis in Missing Data Problems," *Statistics in Medicine* 21:243-256.
        - Fisher, Bonnie S., et al. 1998. "How Many Authors Does It Take To Publish An Article? Trends and Patterns In Political Science." *PS: Political Science and Politics* 31: 847-858.
        - Gill, Jeff. 2007. *Bayesian Methods: A Social and Behavioral Sciences Approach*, 2nd Ed. London: Chapman & Hall.
        - Loredo, T. J. 1990. "From LaPlace to Supernova SN 1987A: Bayesian Inference in Astrophysics." In Forgere, P. F. *Maximum Entropy and Bayesian Methods*. Dordrecht: Kluwer.
        - Tufte, Edward R. 1969. "Improving Data Analysis in Political Science." *World Politics* 21(4):641-54.

- *Exercise Ten: Miscellaneous...*

December 16-20: Finals Week