

PAF 9272: Causal Analysis and Inference

Thomas Leavitt

[Semester Redacted]

E-mail: [Email Redacted]

Class Hours: Thursday 6:05 – 9:00 PM

Office Hours: By Appointment via [Zoom Link Redacted]

Class Room: [Classroom Redacted]

Official Course Description

Causal Analysis and Inference (PAF 9272) is the second course in the research methods sequence. Upon completion of Data Collection and Description (PAF 9270), students can choose between this course and Data Analysis for Public Service (PAF 9271). PAF 9272 is meant for those interested in becoming analysts, researchers, or making quantitative data analysis an important element in their careers. It teaches students how to critically evaluate existing causal analyses of both qualitative and quantitative data and how to conduct statistical analyses to answer causal questions for domestic and international policy and practice. Compared to PAF 9271, PAF 9272 places greater emphasis on observational and experimental data from representative surveys and requires students to write programs (coding) to carry out statistical analyses using advanced statistical software such as Stata or R. The course provides a hands-on introduction to understanding causal evidence, covering logic models and mechanisms, case-oriented vs. variable-oriented approaches, correlation vs. causation, observational vs. (quasi) experimental data, treatment effect, confounding and omitted variable bias, complex survey sampling, generalizability, standard error, confidence interval estimation, hypothesis testing, statistical and practical significance, power analysis, multiple regression, and difference-in-differences estimation. Course sections will use applications tailored towards students' interests and concentrations (e.g., sections more populated with MIA students will have a greater international focus). (Students who took PAF 9170 or PAF 9172 cannot get credit for this course. They, and all other students looking for an advanced causal methods and statistics course, should consider taking PAF 9177.) Open to Austin W. Marxe School of Public and International Affairs MPA and MIA students; others with Marxe School permission.

Course Modality

This course is **Online-Synchronous**. The Meeting ID is [Zoom ID Redacted] and the Zoom link is [Zoom Link Redacted]. There is no passcode.

Course Organization

This course begins by covering many of the topics introduced in PAF 9270 (Data Collection and Description), but does so with greater depth, rigor and accompanying statistical software.

The course then covers concepts and methods to draw conclusions about population quantities from randomly sampled data. After introducing different forms of random sampling, the course shows how to estimate a population average from sample data and test hypotheses about a population average or a difference in population averages.

The subsequent portion of the course covers concepts and methods to draw conclusions about policy impacts from comparisons of intervention and control groups, starting first with randomized experiments, also known as an A/B tests or RCTs (randomized controlled trials). The course then introduces observational studies, which depend on stronger assumptions than randomized experiments.

For all portions of the course, we will engage with real-world examples in a variety of policy domains. We will analyze and interpret existing data, with an emphasis on learning how to conduct analyses in R.

Course Learning Goals

1. Develop and describe causal theories, including mechanisms and alternative explanations of correlations, using qualitative and quantitative research and non-research sources.
2. Demonstrate understanding of causal research concepts, including causation, counterfactual, causal process tracing, generalizability, self-selection, control variables, omitted variable bias, experiments, random assignment, and training and validation datasets.
3. Analyze data with statistical software (such as Stata or R), interpret and write effectively about descriptive and inferential statistics, including hypothesis testing and power analysis
4. Critically evaluate existing causal analyses of both qualitative and quantitative data, to draw valid and useful conclusions, including assessing generalizability, practical significance, and the quality of the causal evidence.
5. Programming (coding) with statistical software to accomplish data analysis tasks including: cleaning data, merging complex datasets, creating new and transforming variables, visualizing data, and preparing tables and charts for professional reports and presentations.
6. Design and conduct research and analyses to estimate causal effects using experimental and observational data, with appropriate methods, such as multiple regression and difference-in-differences.

Course Assignments

Each student should complete 6 problem sets, each of which is due by 5:00 PM on the date listed in the course schedule below. Because I often review problem sets in class, I will not accept late assignments for credit.

Students should also complete a take-home final exam due by 5:00 PM on [Due Date Redacted] and all R assignments on the course's DataCamp site, which can be accessed [DataCamp Link Redacted]. Each DataCamp assignment automatically expires by the beginning of class (6:05 PM) on the day the assignment is due. You will not be able to complete each assignment afterward.

Grades

I will use the standard Baruch College grading scale. I reserve the right to curve the scale depending on the distribution of class scores at the end of the semester. No curve will lower any individual grades relative to what they would have been without the curve. Final course grades will weight the course assessments using the following percentages:

- 50% of your grade will be determined by 6 problem sets (8.333% each).
- 25% of your grade will be determined by the take-home final exam.
- 10% of your grade will be determined by the DataCamp exercises in R.
- 15% of your grade will be determined by course attendance and participation.

Course Readings

This course is ZTC ("Zero Textbook Cost"). All course readings will be available on the course's Blackboard site. The course's readings will draw primarily from the following books:

1. Freedman, D., R. Pisani, and R. Purves (2007). *Statistics* (4th ed.). New York, NY: W. W. Norton & Company.
2. Green, Donald P. (2022). *Social Science Experiments: A Hands-on Introduction*. New York, NY: Cambridge University Press.
3. Rosenbaum, Paul R. (2023). *Causal Inference*. Cambridge, MA: The MIT Press.

Course Standards

Electronics

Please feel free to use your laptops and tablets for class-related purposes. However, please do not use your cell phones or other electronic devices in class. If there are any students for whom cell phones or other electronics are important for course participation, please let me know. I will adjust the electronics policy for the entire class accordingly without singling out anyone.

Attendance, Participation and Collaboration

It goes without saying that attendance and active participation are extremely important to learning from this course. Except for the take-home exam, collaboration with other students is strongly encouraged. However, all take-home assignments should be written individually in students' own words.

If you need to miss class, please do **not** contact me ahead of time. The implications of missing class for the attendance and participation components of your grade are the same regardless of whether you contact me beforehand or not. Please only contact me if you have to miss a class in which you are scheduled to present material.

Incomplete and Late Assignments

If you think that you will be unable to complete an assignment by its due date because of extenuating circumstances, please **do** let me know ahead of time. I may change the assignment's deadline for the entire class. For the final exam, if students turn it in late without prior consent from me, the exam's grade will be penalized by one-third of a letter grade per day (e.g., an A-exam that a student turns in one day late becomes a B+). For problem sets, because I typically review them in class, I will not accept late assignments for credit.

Use of Artificial Intelligence (AI) Tools

Learning the material in this course requires thoughtful completion of all exercises in the course through the completion of your own work. You may use AI tools thoughtfully and with attribution when it supports your learning. If you do use AI tools for an assignment, include at the end of the assignment a brief explanation of which AI tool you used (e.g., Grammarly, ChatGPT, etc.) to accomplish which particular tasks (e.g., coding, language and grammar assistance, etc.)

Suitable uses of AI tools include brainstorming initial ideas, language and grammar assistance, formatting documents, and coding (among others). For other uses, I caution that the content generated by AI tools is frequently (if not usually) incorrect. AI tools will likely be helpful only insofar as they are a complement to (not a replacement of) thoughtful completion of the course material.

In addition, note that content generated via AI tools may not properly attribute or cite sources. You must nevertheless make sure to adhere to the academic integrity standards above. Failure to do so will be regarded as a violation of academic integrity like any other.

Academic Integrity

I will not tolerate violations of academic integrity under any circumstances. Please familiarize yourself with [CUNY's Academic Integrity Policy](#). Additional resources to consult are Graff and Birkenstein (2021), Lipson (2008) and Neville (2016). If you are having difficulty with a written assignment, do not attempt to present another author's work as your own; please come talk to me instead.

Instructions on how to properly cite generative AI are [here](#).

Academic Student Support

Baruch's Marxe School offers academic student support services for both writing and quantitative skills. I highly recommend looking into the Quantitative Student Support department's tutoring services and workshops. More information about these tutoring services and workshops is available [here](#).

Software and computing

This course will use the R statistical programming language. To download R, please go to <https://cloud.r-project.org> and then follow the relevant instructions for installation.

Once you have successfully installed R, please make sure to download and install RStudio, which is an integrated development environment for R. You can download and install RStudio at <https://posit.co/download/rstudio-desktop/>.

You should install both R and RStudio on a personal laptop that you bring to every class. If you do not have access to a personal laptop or other technology, Baruch College's Technology Loan Program may be able to help. You can find more information about this Technology Loan Program [here](#). You may also want to consider use of [Posit Cloud](#), which enables you to use R and RStudio directly in an internet browser without downloading them to a personal device.

A valuable resource for learning R is Wickham et al. (2023), which is available for free at <https://r4ds.hadley.nz/>.

In addition, to facilitate students' learning R, all students should complete all assignments for the course's DataCamp site [DataCamp Link Redacted].

AI tools can be especially helpful for coding with statistical software, especially R. For help with such coding, I recommend ChatGPT, which you can access [here](#). Use of ChatGPT requires signing up for an account, but a free plan is available.

Course Schedule and Readings

Baruch College's academic calendar is available via the Office of the Registrar [here](#).

Class 1: Introduction to course and the R programming language

- ★ Problem set 1 distributed

Class 2: Descriptive statistics — summarizing empirical distributions

- ★ Problem set 2 distributed
- ★ [Problem set 1 due](#)
- ★ [DataCamp Assignment due](#)

- Introduction to R: Intro to basics

Required readings

Theory: Freedman et al. (2007, pp. 57–96)

Class 3: Descriptive statistics — summarizing relationships between variables★ [DataCamp Assignment due](#)

- Introduction to R: Vectors

Required readings

Theory: Freedman et al. (2007, pp. 119–157)

Class 4: Random sampling

★ Problem set 3 distributed

★ [Problem set 2 due](#)★ [DataCamp Assignment due](#)

- Introduction to R: Matrices

Required readings

Theory: Freedman et al. (2007, pp. 333–354)

Application(s): Menachemi et al. (2020, 2021) and Kishore et al. (2018)

Class 5: Estimation and inference of population quantities I★ [DataCamp Assignment due](#)

- Introduction to R: Factors

Required readings

Theory: Freedman et al. (2007, pp. 355–376 and pp. 409–437)

Application(s): Menachemi et al. (2020, 2021) and Kishore et al. (2018)

Class 6: Estimation and inference of population quantities II★ [DataCamp Assignment due](#)

- Introduction to R: Data frames

Required readings

Theory: Freedman et al. (2007, pp. 475–500 and pp. 545–576)

Application(s): Menachemi et al. (2020, 2021) and Kishore et al. (2018)

Class 7: Causality and research design

- ★ Problem set 4 distributed
- ★ **Problem set 3 due**
- ★ **DataCamp Assignment due**
 - Introduction to R: Lists

Required readings

Theory: Green (2022, pp. 1–22)

Class 8: Introduction to randomized experiments

- ★ **DataCamp Assignment due**
 - Intermediate R: Conditionals and Control Flow

Required readings

Theory: Rosenbaum (2023, pp. 1–45)

Application(s): Sherman et al. (1995)

Recommended readings

Green (2022, pp. 42–66)

Class 9: Estimation and inference of causal effects in randomized experiments I

- ★ Problem set 5 distributed
- ★ **Problem set 4 due**

Required readings

Theory: Gerber and Green (2012, pp. 21–50)

Application(s): Sherman et al. (1995)

Class 10: Estimation and inference of causal effects in randomized experiments II

- ★ **DataCamp Assignment due**
 - Introduction to the Tidyverse: Data wrangling

Required readings

Theory: Gerber and Green (2012, pp. 51–66)

Application(s): Sherman et al. (1995)

Class 11: Introduction to observational designs

- ★ Problem set 6 distributed
- ★ **Problem set 5 due**
- ★ **DataCamp Assignment due**
 - Introduction to the Tidyverse: Data visualization

Required readings

Theory: Rosenbaum (2023, pp. 47–65)

Application(s): Silber et al. (2022)

Class 12: Stratification & covariate adjustment in observational designs

- ★ **DataCamp Assignment due**
 - Introduction to the Tidyverse: Grouping and summarizing

Required readings

Theory: Rosenbaum (2023, pp. 67–84)

Application(s): Silber et al. (2022)

Class 13: Pre-post and controlled pre-post designs designs

- ★ **Problem set 6 due**
- ★ **DataCamp Assignment due**
 - Introduction to the Tidyverse: Types of visualizations

Required readings

Theory: Shadish et al. (2002, pp. 103-147)

Application(s): MTA's Fare-Free Bus Pilot Evaluation (MTA, 2024; BronxNet, 2024)

Class 14: Conclusion and pre-final exam review

- ★ **Final exam due on [Due Date Redacted]**

Baruch Policies and Resources

Equal Opportunity and Non-Discrimination

You can submit a report of discrimination and/or retaliation via CUNY's centralized reporting platform [here](#). For more information about CUNY's Policy on Equal Opportunity and Non-Discrimination ("EO Policy"), please access it [here](#). You can also access CUNY's revised Policy on Sexual Misconduct [here](#).

Student Disability Services

Baruch College's Student Disability Services provides accommodations to students with disabilities to promote equal access to Baruch's programs and services. To request accommodations, please do so as soon as possible by directly contacting Student Disability Services [here](#).

Campus Intervention Team (CIT)

The CIT provides a support system for students in crisis. Any member of Baruch's community (faculty, students, staff) can report a concern about a student to CIT. **All reports are confidential.** You can learn more about CIT and the services it offers [here](#).

To report a concern related to concerning behavior, student conduct, academic integrity or student grievance, please use Baruch College's Portal for Incident Reporting available [here](#).

Writing Support

The Schwartz Communication Institute

Writing support services for graduate students in the Marxe school are available through a partnership with the [Bernard L. Schwartz Communication Institute](#) at the Center for Teaching and Learning. Please reach out to the Assistant Director for Writing in Public and International Affairs, [Melina Moore](#), for more information on one-to-one writing consultations. You can also visit the website for writing resources and workshops designed to support graduate students at Marxe. You can learn more [here](#).

The Baruch Writing Center

The Writing Center offers free, professional writing support for all undergraduate and graduate students at Baruch, through one-to-one consultations, workshops, peer review groups, written feedback, online resources, and a journal of outstanding student writing. The Writing Center support faculty through classroom visits, in-class workshops, referral forms, and workshop lesson plans, and it is always available for conversations about teaching and writing. More information is available [here](#), by calling 646-312-4012 or by emailing the center at writing.center@baruch.cuny.edu.

Academic Advisement

The Marxe school provides a range of academic advisement services to ensure the successful completion of students' degree programs. For undergraduate advisement (Bachelor of Science

in Public Affairs (BSPA), you can write an e-mail to mSPIA.bspa@baruch.cuny.edu. For graduate advisement (Master of Public Administration, Executive MPA, Master of International Affairs, and Master of Science in Education-HEA or any of Marxe's certificate programs), you can write an e-mail to mSPIA.advisement@baruch.cuny.edu.

You can also schedule an appointment with an academic advisor via Baruch College's Navigate platform available [here](#).

Career Services and Resources

For personalized career guidance, please make an appointment by logging in to [Marxe Career-Connect](#). For general inquiries, you can write an e-mail to mSPIA.careerservices@baruch.cuny.edu. More information about career services is available [here](#); a variety of career resources, including important links, documents and opportunities, is available [here](#).

References

- BronxNet (2024, April). Today's Verdict: Push to expand bus pilot program. <https://www.youtube.com/watch?v=oUbbA8GBAw0>.
- Freedman, D., R. Pisani, and R. Purves (2007). *Statistics* (4th ed.). New York, NY: W. W. Norton & Company.
- Gerber, A. S. and D. P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W.W. Norton.
- Graff, G. and C. Birkenstein (2021). *They Say / I Say* (5th ed.). New York, NY: W. W. Norton & Company.
- Green, D. P. (2022). *Social Science Experiments: A Hands-on Introduction*. New York, NY: Cambridge University Press.
- Kishore, N., D. Marqués, A. Mahmud, M. V. Kiang, I. Rodriguez, A. Fuller, P. Ebner, C. Sorensen, F. Racy, J. Lemery, L. Maas, J. Leaning, R. A. Irizarry, S. Balsari, and C. O. Buckee (2018). Mortality in Puerto Rico after Hurricane Maria. *The New England Journal of Medicine* 379(2), 162–170.
- Lipson, C. (2008). *Doing Honest Work in College: How to Prepare Citations, Avoid Plagiarism, and Achieve Real Academic Success* (2nd ed.). Chicago, IL: University of Chicago Press.
- Menachemi, N., B. E. Dixon, K. K. Wools-Kaloustian, C. T. Yiannoutsos, and P. K. Halverson (2021). How many SARS-CoV-2-Infected people require hospitalization? Using random sample testing to better inform preparedness efforts. *Journal of Public Health Management and Practice* 27(3), 246–250.
- Menachemi, N., C. T. Yiannoutsos, B. E. Dixon, T. J. Duszynski, W. F. Fadel, K. K. Wools-Kaloustian, N. U. Needleman, K. Box, V. Caine, C. Norwood, L. Weaver, and P. K. Halverson (2020). Population point prevalence of SARS-CoV-2 infection based on a statewide random sample – Indiana, April 25–29, 2020. *The Morbidity and Mortality Weekly Report (MMWR)* 69, 960–964.
- MTA (2024, July). Fare-free bus pilot evaluation. <https://mta.info/document/147096>.
- Neville, C. (2016). *The Complete Guide to Referencing and Avoiding Plagiarism* (3rd ed.). Buckingham, UK: Open University Press.
- Rosenbaum, P. R. (2023). *Causal Inference*. The MIT Press Essential Knowledge Series. Cambridge, MA: The MIT Press.
- Shadish, W. R., T. D. Cook, and D. T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company.
- Sherman, L. W., D. P. Rogan, T. Edwards, R. Whipple, D. Shreve, D. Witcher, W. Trimble, The Street Narcotics Unit, R. Velke, M. Blumberg, A. Beatty, and C. A. Bridgeforth (1995). Deterrent effects of police raids on crack houses: A randomized, controlled experiment. *Justice Quarterly* 12(4), 755–781.

Silber, J. H., P. R. Rosenbaum, J. G. Reiter, A. S. Hill, S. Jain, D. A. Wolk, D. Small, S. Hashemi, B. A. Niknam, M. D. Neuman, L. A. Fleisher, and R. Eckenhoff (2022). Alzheimer’s dementia after exposure to anesthesia and surgery in the elderly: A matched natural experiment using appendicitis. *Annals of Surgery* 276(5), e377–e385.

Wickham, H., M. Çetinkaya Rundel, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (2nd ed.). Sebastopol, CA: O’Reilly Media.