PSCI 8357 (Spring 2026)
Statistics for Political Research (STAT) II

WHEN: M/W 4:30 – 5:45PM
WHERE: COMMONS 320

Instructor: Georgiy (Gosha) Syunyaev (g.syunyaev@vanderbilt.edu)

- **Office hours:** Wednesday, 10 AM - 12 PM in Commons 351 (or email me to schedule appointment)
- **Office hourse sign-up link:** https://calendar.app.google/iHhWLRvcGoppGW477

TA: Alexander (Alex) Dean

- **Recitations:** TBD
- **Office hours:** TBD

Course Overview

This course offers an up-to-date exploration of causal inference in quantitative social science research. We will study two main components of causal inference: (1) the analysis of causal identification and (2) statistical inference based on research design. Our focus will be on non-parametric causal identification methods, along with non-parametric and semi-parametric estimation techniques. We will prioritize the principles of research design and robust estimation and inference using frequentist approaches.

Prerequisites

There are two prerequisits for this class. First, students should have a firm grasp of probability theory, statistical inference, and linear models at the level of STAT I or an equivalent course (e.g. do you remember what the law of iterated expectations is? or what does it mean for two random variables to be independent? or the significance of the equation $\beta = (X'X)^{-1}(X'Y)$?). Second, students should have some background in writing scripts to implement statistical analyses in R. The course provides foundational methodological training to Political Science PhD students in their first or second year as part of their required sequence of courses.

Requirements

In-class quizzes

Roughly once a week, we'll start class with a short in-class quiz (no more than 5 questions). The goal is simply to check how the main ideas are landing and to help me identify topics that may need more explanation or practice. These quizzes are *low-stakes* and meant to support your learning—not to add pressure. They are **not graded**, and you should think of them as a quick warm-up and a way for us to calibrate where we are as a class.

Problem sets (7 $\times$ 5 %)

You will receive homework about every two weeks via Brightspace (not GitHub). You will have to submit your completed problem set within a week; exact deadlines will be made clear on the assignment. You can work with others, but to receive credit, your homework must comply with the following guidelines:

- You must turn in a PDF copy of your own homework by the stated deadline to both the instructor and TA.
- The assignment that you turn in must clearly reflect your own thinking. Sets of verbatim copies of homework will have credit reduced by half.
- Include a short disclaimer at the top of the assignment (e.g., after your name/date) indicating whether you used any AI tool(s) beyond spell-checking/light editing and for what purpose(s). If you did, also attach a printed to PDF log of the AI chat.
- Estimates obtained in R must be formatted properly into tables or graphs resembling journal presentation styles. You should use a table formatting function (e.g., `kableExtra`, `apsrtable` or `stargazer` in R). Use a reasonable (2 or at most 3) number of digits after decimal points, report standard errors or confidence intervals along with coefficients, clarify what are the dependent variables in each table or figure, and explain in footnotes to your tables or figures what kinds of estimators or adjustments have been used. Print outs of raw screen output or commented logs will not receive any credit. However, you may include such output as an appendix so that the grader can troubleshoot.
- Mathematical derivations should include all key steps with explanations for important techniques.

- Your assignment should be submitted as a PDF file compiled from LaTeX, R Markdown, or (*ideally*) a Quarto Markdown document.
  - If using raw LaTeX(.tex) for your answers, submit an accompanying R file for any computational tasks, with referenced line numbers corresponding to each specific task.
  - If using R Markdown (.Rmd) or Quarto Markdown (.qmd), include your code as code chunks in the source file. Additionally, submit the source .Rmd or .qmd file along with the compiled PDF to allow us to run your code easily.

Homework will be graded for points as indicated on each assignment and count toward 35% of your grade.

Replications (2 × 20 %)

The primary objective of this course is to provide both a theoretical understanding and hands-on experience implementing the methods we cover. To that end, you must complete **two individual replications** of published studies: one based on an experimental and one based on an observational study. This is an **individual assignment**: you may discuss ideas with classmates, but all code and writing must be your own. **No two students may replicate the same paper.**

Each student must select a paper at least **one month before** the relevant replication deadline and notify both the instructor and the TA. To streamline selection—and to keep the emphasis on replication/extension skills rather than searching for a "perfect" fit—I will provide a curated list of pre-approved papers with publicly available replication materials (e.g., Harvard Dataverse, OSF) in advance of the selection deadline. You should choose from this list unless you have a compelling reason to propose an alternative; any off-list selection must be approved in advance and is still subject to the "no duplicates" rule. If replication materials are not readily accessible online, you may contact the authors or request assistance from the instructor, but deadlines still apply.

Once you have selected your paper, you are expected to complete two tasks:

1. **Replication:** Replicate the primary analyses reported in the main body of the paper **using your own code** (you do not need to replicate appendices unless explicitly required). You should not rely on the authors' original code as your main implementation.
2. **Extensions (at least two):** Conduct **two distinct extensions** of the replicated analysis. Each extension must be **clearly delineated** (labelled as "Extension 1" and "Extension 2"), **well-motivated**, and explained as a specific contribution beyond the main replication. Extensions may take one of two broad forms:
   - **Robustness checks:** e.g., alternative estimators/specifications, alternative measurement choices, additional tests of identifying assumptions.
   - **Theory-driven analyses:** e.g., additional tests implied by the theory/hypotheses (such as heterogeneity analyses). You may do two robustness checks, two theory-driven analyses, or one of each, as long as each is distinct, motivated, and clearly separated.

For both replication and extensions, you must post a brief replication plan on the Brightspace forum at least **one week** before submitting the replication report. I am happy to discuss any issues and extension ideas during the office hours.

Your final submission must include (1) the data files you used, (2) the original authors' replication code (if available), (3) a replication write-up compiled to PDF, and (4) the raw, fully reproducible R Markdown or Quarto Markdown source file that would allow the instructor and TA to re-run your analyses. The compiled PDF write-up should be **10–20 double-spaced pages** and should include:

- A brief summary of the paper's theory and hypotheses.
- A concise description of the data, model, and main results (screenshots from the paper are allowed for reference).
- A report on your replication of the main results (including code, output, and discussion of any deviations from the original paper).
- A report on your two extensions (including code, output, and interpretation).
- *(Optional)* An appendix with ancillary details, tables, or figures that does not count toward the page limit and may be referenced in the main text.

Because part of the goal is to practice professional research communication, a small portion of your grade will be based on **formatting and presentation quality**. For example: do not dump large data frames; do not present a separate table for every minor model variation; format tables and figures in a paper-ready style rather than showing raw console output.

Final Exam (25%)

A final exam will be scheduled during the final examination week (the week of April 20th). The exam aims to assess individual progress, thus allowing me to provide personalized recommendations for improving your methodological foundations. If you cannot take the exam during the scheduled period, you must provide notice at least one week in advance so that we can arrange an alternative time. The final will constitute 20% of your grade.

Grading and Deadlines

- **35% Problem sets** (every two weeks, see schedule below)
- **40% Replications** (due March 2nd and April 13th)
- **25% Take-home final exam** (Week of April 20th)

Late work will not be accepted without documented proof of a family or medical emergency.

Problem sets and exam will have opportunities to earn extra credit, theoretically meaning you could score above 100%. All grades are curved without taking the extra credit results into consideration. The class follow the standard Vanderbilt grading system: A 94+ | A- 90-93 | B+ 87-89 | B 84-86 | B- 80-83 | C+ 77-79 | C 74-76 | C- 70-73 | D+ 67-69 | D 64-66 | D- 60-63 | F <60

Resources

Software

You will have to work in R in this class. I encourage using Quarto for your assignments. This is a great investment that will pay off in the long run in terms of productivity as well as reproducibility. Quarto Markdown runs easily through RStudio or VS Code (or even many of its wrappers like Cursor or soon Positron). You can also check materials in the repository I prepared for the Scientific Workflow workshop at Vanderbilt here: github.com/gerasy1987/workflow_workshop.

Textbooks

We will draw on textbooks and papers for the course. Here are the **required** textbooks:

- Angrist, Joshua D., and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton university press, 2009.
- Gerber, Alan S and Donald P Green. *Field Experiments: Design, analysis, and interpretation.* W.W. Norton, 2012.
- Morgan, Stephen L., and Christopher Winship. *Counterfactuals and Causal Inference.* 2nd Ed. Cambridge University Press, 2015.

And here are the **recommended** ones:

- Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. *A Practical Introduction to Regression Discontinuity Designs: Foundations.* Cambridge University Press, 2019.
- de Chaisemartin, Clément, and Xavier D'Haultfoeuille. *Difference-in-Differences for Simple and Complex Natural Experiments.* Forthcoming from Princeton University Press, 2023.
- Ding, Peng. *A First Course in Causal Inference.* CRC Press, 2024.
- Huber, Martin. *Causal Analysis.* MIT Press, 2023.
- Humphreys, Macartan, and Alan M. Jacobs. *Integrated Inferences: Causal Models for Qualitative and Mixed-Method Research.* Cambridge University Press, 2023.
- Imbens, Guido W., and Donald B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences.* Cambridge University Press, 2015.
- Lohr, Sharon L. *Sampling: Design and Analysis.* CRC press, 2021.

You can obtain these as PDFs on Brightspace or on the authors' websites and online preprint archives (SSRN or arXiv). Papers are listed below according to topic.

Course Pages

Syllabus, readings, assignments, and announcements for the course will be posted on the course page on Brightspace (course management system used by Vanderbilt). I will post announcements and changes to the home page of the site, please keep an eye out. In addition, we will have discussion forums for any class related questions, replication plans and class related news/social media posts on Brightspace.

In addition to Brightspace, we will maintain a class GitHub repository at https://github.com/gerasy1987/PSCI8357. All current lecture slides, handouts, code, and other publicly shareable materials will be posted there. Please **clone** this repository to your laptop early in the semester and **pull** regularly so you always have the latest versions. The GitHub repository will also include the **raw Quarto (.qmd)** source files used to produce many of the course materials, which can serve as concrete examples of reproducible scientific writing and analysis workflows.

Getting comfortable with **git/GitHub and version control** is a valuable skill in its own right. Even if you have not used it before, this course is a low-stakes opportunity to practice a workflow that will help you on future projects-keeping track of changes, collaborating cleanly, and making it easy to return to earlier versions of your work when (not if) something breaks.

AI Tools

Students may use AI tools (e.g., ChatGPT, Claude, Copilot) for **all non in-class activities** in this course, including brainstorming, debugging, coding assistance, explaining unfamiliar concepts or coaching for the exam. However, AI should be treated as an **assistant**, not a substitute for your own writing, critical thinking, and coding.

This course is designed to make you an informed and professional user of causal inference tools—not someone who can only run pre-written snippets of code without understanding the underlying identification logic, assumptions, inference and code. Using AI to fully replace substantive reasoning, original writing, or core implementation work undermines that goal and may be treated as a form of academic misconduct.

If you use any AI tool **other than grammar/spell-checking tools** (e.g., *Grammarly*-style editing) for any part of any class assignment, you must do **both** of the following:

1. **Disclosure statement:** Include a short disclaimer at the top of the assignment (e.g., after your name/date) indicating what tool(s) you used and for what purpose(s).
2. **Conversation log:** In addition to other submitted materials, submit a PDF printout of the full conversation log(s) with the AI tool(s) used for that assignment. For example, if you used ChatGPT in a browser, open the relevant conversation and print/save it to PDF, then submit it with your assignment.

If you have questions or concerns about what is allowed, or how to document AI use, please ask during class or contact me individually.

Course Schedule

Note: *The schedule is subject to change. Readings will be distributed through Brightspace ahead of each class.*

Date	Title	Topics	Readings
1/5	Introduction and Syllabus		
	PRIMER ON CAUSAL INFERENCE		
1/7	Probability Review		• Samii (2016)
1/12	What is Identification?	observation versus intervention, potential outcomes (ATE, ATT, probability of necessity), ATE under SUTVA and strong ignorability, causal effects with potential outcomes	• Gerber and Green (2012, Ch. 2) • Angrist and Pischke (2009, Ch. 1-2) • Morgan and Winship (2015, Ch. 1-2) • Holland (1986)
1/14	No class	SPSA, MLK Holiday	
1/19			
1/21	What is Identification (cont.)?	observation versus intervention, potential outcomes (ATE, ATT, probability of necessity), ATE under SUTVA and strong ignorability, causal effects with potential outcomes,	• Gerber and Green (2012, Ch. 2) • Angrist and Pischke (2009, Ch. 1-2) • Morgan and Winship (2015, Ch. 1-2) • Holland (1986)
1/26	Regression and Causality	CEF and its properties, selection on observables, conditional ignorability, **problem set 1 is posted on 1/21**	• Angrist and Pischke (2009, Ch. 3.1-3.2.2) • Morgan and Winship (2015, Ch. 6.1-6.2) • Cinelli, Forney, and Pearl (2024)
1/28			
2/2	The Truth about Regression	regression anatomy, omitted variable bias, positivity assumption	• Angrist and Pischke (2009, Ch. 3.2.3-3.5) • Morgan and Winship (2015, Ch. 6.3) • Clarke (2005) • Aronow and Samii (2016)
	EXPERIMENTAL DESIGNS		
2/4 2/9	Basics of Experimental Design	target quantities (SATE, PATE), inference for the SATE and PATE in an idealized experiment, randomization inference, imbalance, **problem set 2 is posted on 2/2**	• Gerber and Green (2012, Ch. 3)
2/11	Complex Experimental Designs and Power Analysis	cluster and block randomization, factorial designs, MDE and power analysis	• Gerber and Green (2012, Ch. 4)
2/16			
	OBSERVATIONAL DESIGNS FOR CAUSAL INFERENCE		

Date	Title	Topics	Readings
2/18 **2/23** **2/25**	Matching	distance metrics, exact matching, nearest neighbor matching, propensity score matching, balancing property, sensitivity analysis, weighting, doubly robust estimators **problem set 3 is posted on 2/16**	• Angrist and Pischke (2009, Ch. 3) • Morgan and Winship (2015, Ch. 5) • Sekhon (2009) • Caliendo and Kopeinig (2008)
3/2 3/4	Instrumental Variables	Local Average Treatment Effects, IV estimator, one-/two-sided non-compliance, 2SLS estimator and its bias, **problem set 4 is posted on 3/2**, **experimental replication is due on 3/2**	• Angrist and Pischke (2009, Ch. 4) • Morgan and Winship (2015, Ch. 9.1-9.3) • Angrist, Imbens, and Rubin (1996) • Sovey and Green (2011)
3/4	Difference-in-Differences	two-period DID, pre-trends, event study	• Angrist and Pischke (2009, Ch. 5.2) • Bertrand, Duflo, and Mullainathan (2004)
3/9 **3/11**	No Class	Spring Break	
3/16 **3/18**	Difference-in-Differences (cont.)	conditional DID, continuous treatment DID, event-by-event estimation, **problem set 5 is posted on 3/16**	• Angrist and Pischke (2009, Ch. 5.2) • Bertrand, Duflo, and Mullainathan (2004)
3/23 **3/25**	Panel Data	staggered adoption design, fixed effects, TWFE estimator and its bias	• Angrist and Pischke (2009, Ch. 5.1, 5.4) • Morgan and Winship (2015, Ch. 11.3)
3/30	Synthetic Control	identification, estimation and statistical inference, interactive FEs, **problem set 6 is posted on 3/25**	• Abadie, Diamond, and Hainmueller (2010)
4/1 4/6	Regression Discontinuity Designs	parametric, non-parametric local regression, optimized RD and honest inference, threats, fuzzy RDD	• Angrist and Pischke (2009, Ch. 6) • Morgan and Winship (2015, Ch. 11.2) • Imbens and Lemieux (2008) • De la Cuesta and Imai (2016)
4/8 **4/13**	Mediation Effects	natural direct and indirect effects, controlled direct and indirect effects **problem set 7 is posted on 4/6**, **observational replication is due on 4/13**	• Gerber and Green (2012, Ch. 9-10)

Date	Title	Topics	Readings
4/15	Effect Heterogeneity	interaction effects, decomposing effect heterogeneity, optimal treatment regimes	TBD
4/20	Course Wrap-Up		

Course Policies

Cell Phones, Laptops, Tablets, etc.

You are asked to silence your cell phone / tablet / smart watch before class begins.

Academic Honor Code

Students are expected to be familiar with and adhere to the Vanderbilt University Academic Honesty policy, available at www.vanderbilt.edu/student_handbook/the-honor-system/.

While collaboration is a key component of the social sciences, it is imperative that each student's work on assignments reflects their own efforts. Care must be taken to avoid plagiarism. Collaboration is allowed on problem sets, but strictly prohibited on final exams and replications.

Academic misconduct, which includes cheating, fabrication, plagiarism, altering graded examinations for additional credit, having another person take an examination, falsification of results, and facilitating academic dishonesty, as specified further in the university policy, is unacceptable and may result in penalties such as failure of the assignment or course, as well as disciplinary actions at the program or university level.

Accommodations for Learning or Access Disabilities

This course is designed to be inclusive and respectful of students of all backgrounds, identities, and abilities. If there are barriers that affect the learning environment or require specific arrangements (such as those related to building evacuations), students are encouraged to discuss these with the instructor as early as possible. The confidentiality of these discussions will be respected. Students should also contact the Vanderbilt Student Access office (www.vanderbilt.edu/student-access/) to learn about specific accommodations and ensure they are provided in a timely manner. Accommodations requests should be made within the first three weeks of the semester, except under unusual circumstances.

Mental Health

Students may encounter stressors that impact both their academic performance and personal well-being. These can include academic pressures and challenges related to relationships, mental health, substance use, identities, or finances. If these challenges interfere with academic success, students should reach out to the instructor to explore potential solutions together. Vanderbilt offers the following resources:

- University Counseling Center: Provides individual and group therapy, psychiatric services, and assessments. Urgent Care Counseling is available in person from 9 am - 4 pm, Monday through Friday, or by phone 24/7. For more information, call 615-322-2571, visit vu.edu/scn, or connect with Student Care Coordination (SCC) where most referrals to the UCC begin.
- Center for Student Wellbeing: Aims to support personal and academic success. Contact by calling 615-322-0480 or emailing healthydores@vanderbilt.edu.
- Student Care Coordination: Assists students in accessing campus and community resources for academic and personal support. Call 615-343-9355 or visit vu.edu/scn.
- Crisis Text Line: For free 24/7 support, text VANDY to 741741.
- National Suicide & Crisis Lifeline: Call or text 988 for confidential, round-the-clock support.
- Vanderbilt Psychiatric Hospital: Offers 24/7 crisis assessment and admissions. Immediate help is available by calling 615-327-7000.

Mandatory Reporting

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and support applied to offenses against other protected categories such as race and national origin. If students or someone they know has been harassed or assaulted, they can call the Project Safe 24-hour crisis/support hotline at 615-322-7233. A list of resources can be found at Project Safe. The University's Title IX Coordinator (615-322-4705) is another contact point, where appropriate resources and contacts for confidential support are available: www.vanderbilt.edu/title-ix/.

As faculty members, professors have responsibilities to help create a safe learning environment on campus, regardless of identity or circumstances. Professors also have a mandatory reporting responsibility. It is the intention that students feel able to share information related to their life experiences in classroom discussions, written work, and one-on-one meetings. Faculty will seek to keep information shared as private as possible. However, as representatives of an institution that strives for safety for all people, professors are mandatory reporters. University faculty, many staff members, and some student leaders are required to report incidents of sexual assault, sexual harassment, dating violence, domestic violence, stalking, and child abuse, as well as any suspected discrimination (regarding age, race, color, creed, religion, ancestry, national or ethnic origin, sex/gender, sexual orientation, disability, genetic information, military status, familial status, or other protected categories under local, state, or federal law) to the University's Title IX Coordinator (615-322-4705), as required by University policy and state and federal law. If an experience of interpersonal violence and/or child abuse is disclosed to faculty or classmates with mandatory reporting duties, whether in class discussion, through a course assignment, or in private communication, the disclosure will be kept as private as possible but may not remain confidential.

Diversity Statement

Social science centers around creative thinking aimed at answering challenging questions. Such creativity flourishes through exposure to diverse perspectives that stem from varied experiences. Diversity in all its forms, including age, ability or disability, ethnicity, national origin, race, religion,

sex, gender, sexual orientation, and family and marital status, is highly valued in this class. It is expected that all students will respect these differences and strive to understand how others' perspectives, behaviors, and worldviews may differ from their own.

Religious Holidays

Observing religious holidays and cultural practices is an important part of reflecting diversity. As an instructor, the commitment is to provide equivalent educational opportunities to students of all belief systems. Students should review the course requirements at the beginning of the semester to identify any foreseeable conflicts with assignments, exams, or other required attendance. If possible, students are encouraged to contact the instructor within the first two weeks of the first class meeting to discuss and make fair and reasonable adjustments to the schedule and/or tasks.

Acknowledgments

This course is largely based on the materials prepared for a previous iteration of Statistics for Political Research III taught by Bradley Smith at Vanderbilt. Additionally, some weeks draw on materials from similar classes taught by Cyrus Samii at NYU and Naoki Egami at Columbia.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Wconometrics: An Wmpiricist's Companion*. Princeton University Press.

Aronow, Peter M, and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250–67.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1): 249–75.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22 (1): 31–72.

Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2024. "A Crash Course in Good and Bad Controls." *Sociological Methods & Research* 53 (3): 1071–1104.

Clarke, Kevin A. 2005. "The Phantom Menace: Omitted Variable Bias in Econometric Research." *Conflict Management and Peace Science* 22 (4): 341–52.

De la Cuesta, Brandon, and Kosuke Imai. 2016. "Misunderstandings about the Regression Discontinuity Design in the Study of Close Elections." *Annual Review of Political Science* 19 (1): 375–96.

Gerber, Alan S, and Donald P Green. 2012. "Field Experiments: Design, Analysis, and Interpretation." *(No Title)*.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60.

Imbens, Guido W, and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

Morgan, Stephen L, and Christopher Winship. 2015. *Counterfactuals and Causal Inference*. Cambridge University Press.

Samii, Cyrus. 2016. "Causal Empiricism in Quantitative Research." *The Journal of Politics* 78 (3): 941–55.

Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12 (1): 487–508.

Sovey, Allison J, and Donald P Green. 2011. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science* 55 (1): 188–200.