Causal Inference for the Social Sciences

Jake Bowers [*] Thomas (Tom) Leavitt [†] Stephen El-Khatib [‡]
Kathryn Overton [§]

ICPSR Session 2 (July 18, 2022)

Abstract

This course introduces methods and concepts used to infer causal effects from comparisons of intervention and control groups. We'll use the potential outcomes framework of causality to analyze both randomized and observational studies, distinguishing different forms of random assignment and separating observational studies that involve instruments, discontinuities and other devices, highlighting the interplay of study design for statistical analysis. Propensity score matching is treated in depth, with explicit instruction in the use of "optmatch" and related packages in R; other areas of methodological focus include assessment of covariate balance by specification tests and other methods; inference methods that are a robust to small sample sizes, weak instruments, spillover and interaction effects, heterogeneous treatment effects, and/or misspecification of response surfaces; and omitted variable sensitivity analysis.

The course presupposes knowledge of multiple regression at the level of the ICPSR course Regression: II, as well as multiple regression with binary dependent variables (as taught in the ICPSR courses Regression: III or Maximum Likelihood). The part of the course presenting matching requires the use of `R` for computation, but other methods presented in the course are readily implemented either in `R` or in `Stata`.

The course meets 3–5 pm US Eastern Time, Monday through Friday, from July 18 through August 12. There will be no class on July 18, the first day of the course. Tom will be the primary instructor for the first and last weeks of the course. Jake will be the primary instructor for weeks 2 and 3. Stephen and Kathryn will be the TAs for the duration of the course.

You can sign up for office hours with Jake at https://calendly.com/jakebowers/icpsr-office-hours. By default they are on Zoom, but he can meet in-person on request. Tom, Stephen and Kathryn's office hours will be announced closer to the start date of the course.

[*]Political Science and Statistics, University of Illinois @ Urbana-Champaign; jwbowers@illinois.edu

[†]Department of Health Care Policy, Harvard University; leavitt@hcp.med.harvard.edu

[‡]University of California, Riverside, stevekhatib@gmail.com

[§]University of New Mexico, koverton@unm.edu

Overview

We may all warn our freshmen that association is not causation, but inferring causation has always been a central aim both for statisticians and for their collaborators. Until recently, however, inference of causation from statistical evidence depended on murky, scarcely attainable requirements; in practice, the weight of casual arguments was largely determined by the scientific authority of the people making them.

Requirements for causal inference become more clear when they are framed in terms of *potential outcomes*. This was first done by Neyman, who in the 1920s used potential outcomes to model agricultural experiments. Fisher independently proposed a related but distinct, ultimately more influential, analysis of experiments in 1935, and a rich strain of causal analysis developed among his intellectual progeny. It clarified the differing requirements for causal inference with experiments and with observational data, isolating the distinct contributions required of the statistician and of his disciplinary collaborators; generated more satisfying methods with which to address potential confounding due to measured variables; qualitatively and quantitatively advanced our grasp of unmeasured confounding and its potential ramifications; furnished statistical methods with which to eke more out of the strongest study designs, under fewer assumptions; and articulated principles with which to understand study designs as a spectrum, rather than a dichotomy between “good" experiments and “bad" observational studies. Understanding the methods and outlook of the school founded by Fisher’s student W. G. Cochran will be the central task of this course.

The course begins by applying the Fisher and Neyman-Rubin approaches to statistical inference for counterfactual causal efffects to randomized experiments, touching on considerations specific to clustered treatment assignment, “small” sample sizes and treatment effect heterogeneity. The next segment addresses conceptual and methodological challenges of applying the same models to analysis of non-experimental data. This course segment covers ignorability, selection, “common support,” covariate balance, paired comparisons, optimal matching and propensity scores. A short separate section introduces another method aiming to identify experiment-like structures in observational data, namely regression discontinuity, before a return to experiments.

With these foundations in place, the course’s second half adds conceptual depth and methodological flexibility. Central topics include instrumental variables and local average treatment effects, stratified designs with clustering, interference, omitted variable sensitivity analysis and adapting workhorse techniques such as multiple regression to the demands of causal inference. Over the course of the four weeks the course becomes progressively less conceptual and more applied with increasing emphasis on computing strategies in R.

Administrative

Textbooks

The main texts for the course are

Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017

Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010

Paul R Rosenbaum. *Observational Studies*. Springer, New York, NY, second edition, 2002

These three textbooks are presented in varying difficulty and we will draw from all three. Although we won't follow these books closely, their goals and methods align with the course's, and they will be useful as references and supplements.

Other texts that we draw on include

Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York, NY, 2012

Guido W Imbens and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, NY, 2015

Other readings will be assigned and distributed electronically.

If you're new to `R`, we suggest getting a hold of:

John Fox. *Applied Regression Analysis and Generalized Linear Models*. SAGE Publications, Los Angeles, CA, 3rd edition, 2016

Hadley Wickham and Garrett Grolemund. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Sebastopol, CA, First edition, 2017

`R` software will be required for several specific segments of the course. With some independent effort, students not familiar with `R` in advance should be able to learn enough `R` during the course to complete these assignments. We also recommend some work with `R` — for example, via working through some online `R` courses — before the course for students who have never used it before.

Assignments

Assignments are due each Tuesday, at the beginning of class. Parts of the assignment will be given at the beginning of the week, but other parts will be given during class, over the course of the week. Many of these daily assignments will be given with the expectation that they'll be completed by the next course meeting, although they'll only be collected at the end of the week. Late homework will not be accepted without cause (or prior arrangement with the teaching assistant).

You're welcome to submit a paper at the end of the course, whether or not you're taking the course for credit. In that case we'll return it with comments within a month or so of the course's completion. (If you're taking the course for a grade, the paper won't contribute to the grade unless you're on the borderline between two grades.)

Participation is expected. It can take various forms:

1. Doing in-class exercises and discussing them with your peers;
2. From time to time, making a clarification or raising a clarifying questions;

3. Contributing to in-class discussions.
4. Drop by one of the professor's office hours to share a point that you *and at least one classmate* would like to have clarified or amplified, or to point out a connection to your field;

If you are taking the course for a grade, make a point of doing at least one of 3 and 4.

Course Schedule

The course schedule is below. In the Course content section, we provide more extensive readings on each topic in the schedule, as well as additional “special topics.” The last two days of the course are reserved for special topics chosen by students.

Date	Instructor	Topic	Required readings	Application
Tues, July 19	Tom	Introduction: Randomized experiments and potential outcomes	Holland (1986) Kinder and Palfrey (1993, Section 1.2) Fisher (1935, Introduction) Rosenbaum (2017, Chapter 2)	N/A
July 20	Tom	Randomized experiments: Fisherian Exact tests	Rosenbaum (2017, Chapter 3)	Arceneaux (2005)
July 21	Tom	Randomized experiments: Neymanian Estimation and inference	Gerber and Green (2012, Chapter 2) Aronow and Middleton (2013) Middleton and Aronow (2015) Gerber and Green (2012, pp. 51 – 61)	
July 22				
Mon, July 25	Jake	Randomized Experiments: Covariance adjustment & regression	Gerber and Green (2012, Chapter 4) Rosenbaum (2002) Lin (2013)	
July 26 (HW 1 due)	Jake	Randomized experiments: Noncompliance and attrition	Gerber and Green (2012, Chapters 5 – 6) Rosenbaum (1996) Rosenbaum (2010, Section 5.3)	Albertson and Lawrence (2009)
July 27	Jake			
July 28	Jake	Observational studies: Introduction	Bind and Rubin (2019) Gelman and Hill (2006, Sections 9.0 – 9.2) Berk (2010)	Cerdá et al (2012)
July 29	Jake	Introduction to matching	Rosenbaum (2017, pp. 65 – 90)	
Mon, August 1	Jake	Nuts & bolts of matching	Rosenbaum (2010, Chapters 7 – 8) Rosenbaum (2017, Chapter 11)	
August 2 (HW 2 due)	Jake	Propensity score methods	Rosenbaum (2017, pp. 90 – 96) Hansen (2011)	
August 3	Jake	Assessments of covariate balance	Hansen and Bowers (2008) Gerber and Green (2021, pp. 71 – 79)	
August 4	Jake	Outcome analysis after matching		
August 5	Jake	Nonbipartite matching	Rosenbaum (2010, Chapter 11)	Wong et al (2012, 2020)
Mon, August 8	Tom	Regression discontinuity designs and “natural experiments”	Cattaneo, Titiunik, and Vazquez-Bare (2020) Keele, Titiunik and Zubizarreta (2015) Sales and Hansen (2020)	Caughey and Sekon (2011)
August 9 (HW 3 due)	Tom	Fisherian sensitivity analysis	Rosenbaum (2017, Chapter 9)	Cerdá et al (2012)
August 10	Tom	Neymanian sensitivity analysis	Fogarty (2020)	
August 11	Tom	Special topics	TBD	TBD
August 12	Tom	Special topics	TBD	TBD

Course content

1 Potential outcomes and random assignment

Required

Paul W Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986, Sections 1 – 4. (The article that brought the "Rubin Causal Model" to statisticians' attention.)

Section 1.2, "Experimentation defined," of Donald R Kinder and Thomas R Palfrey. On behalf of an experimental political science. In Donald R Kinder and Thomas R Palfrey, editors, *Experimental Foundations of Political Science*, Michigan Studies in Political Analysis, chapter 1, pages 1–39. University of Michigan Press, Ann Arbor, MI, 1993. (Particularly pp. 5 – 10.)

Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017, Chapter 2.

Chapter 1, "Introduction," of Ronald Aylmer Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, SCT, 1935 and pages 131 – 135 of Joan Fisher Box. *R. A. Fisher, the Life of a Scientist*. Wiley, New York, NY, 1978 for historical context.

Recommended

Jake Bowers and Thomas Leavitt. Causality and design-based inference. In Luigi Curini and Robert Franzese, editors, *The SAGE Handbook of Research Methods in Political Science and International Relations*, volume 2, chapter 41, pages 769–804. SAGE Publications, Thousand Oaks, CA, 2020

2 Random assignment as a basis for inference

Application

Kevin Arceneaux. Using cluster randomized field experiments to study voting behavior. *The Annals of the American Academy of Political and Social Science*, 601(1):169–179, 2005

2.1 Inference for causal effects: the Neyman tradition

2.1.1 Estimation of average causal effects

Required

Chapter 2 and pp. 51 – 61 of Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York, NY, 2012

Peter M Aronow and Joel A Middleton. A class of unbiased estimators of the average treatment effect in randomized experiments. *Journal of Causal Inference*, 1(1):135–154, 2013

Joel A Middleton and Peter M Aronow. Unbiased estimation of the average treatment effect in cluster-randomized experiments. *Statistics, Politics and Policy*, 6(1-2):39–75, 2015

2.1.2 Variance estimation and hypothesis testing

Required

Chapter 3 of Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York, NY, 2012.

Endnote spanning pages A-32 and 33, David A Freedman, Robert Pisani, and Roger Purves. *Statistics*. W. W. Norton & Company, New York, NY, 3rd edition, 1998. (This can be read as a précis of: Jerzy Splawa-Neyman, D M Dabrowska, T P Speed, et al. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472, 1990.)

Chapter 6, pp. 87 – 98 of Guido W Imbens and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, NY, 2015

Recommended

Chapter 6 of Thad Dunning. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press, New York, NY, 2012

Xinran Li and Peng Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520):1759–1769, 2017

Peng Ding. A paradox from randomization-based causal inference. *Statistical Science*, 32(3):331–345, 2017

Peter M Aronow, Donald P Green, Donald KK Lee, et al. Sharp bounds on the variance in randomized experiments. *The Annals of Statistics*, 42(3):850–871, 2014

2.2 Inference for causal effects: the Fisherian tradition

Required

Chapter 3 of Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017.

Recommended

Pages 27 – 49 of Paul R Rosenbaum. *Observational Studies*. Springer, New York, NY, second edition, 2002

Chapter 2 of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010.

2.3 Covariance adjustment in randomized experiments

Required

Chapter 4 of Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York, NY, 2012.

Paul R Rosenbaum. Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17(3):286–327, 2002.

Winston Lin. Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, 7(1):295–318, 2013

Recommended

Luke W Miratrix, Jasjeet S Sekhon, and Bin Yu. Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(2):369–396, 2013

David A Freedman. On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2):180–193, 2008

David A Freedman. Randomization does not justify logistic regression. *Statistical Science*, 23(2):237–249, 2008

David A Freedman. On regression adjustments in experiments with several treatments. *The Annals of Applied Statistics*, 2(1):176–196, 2008

Cyrus Samii and Peter M Aronow. On equivalencies between design-based and regression-based variance estimators for randomized experiments. *Statistics & Probability Letters*, 82(2):365–370, 2012

Peter M Aronow and Cyrus Samii. Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60(1):250–267, 2016

Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1):265–296, 2020

3 Noncompliance and Attrition

Applications

Bethany Albertson and Adria Lawrence. After the credits roll: The long-term effects of educational television on public knowledge and attitudes. *American Politics Research*, 37(2):275–300, 2009

3.1 Noncompliance and instrumental variables

Required

Chapters 5 and 6 of Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation.* W.W. Norton, New York, NY, 2012.

Section 5.3, "Instruments," of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010

Paul R Rosenbaum. Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association*, 91(434):465–468, 1996

Recommended

Section 2.3 of Paul R Rosenbaum. Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17(3):286–327, 2002

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996

Guido W Imbens and Paul R Rosenbaum. Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):109–126, 2005

Hyunseung Kang, Laura Peck, and Luke Keele. Inference for instrumental variables: A randomization inference approach. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 181(4):1231–1254, 2018

Ben B Hansen and Jake Bowers. Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23(2):219–236, 2008

3.2 Attrition, or missing outcomes

Recommended

Chapter 7 of Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York, NY, 2012

David S Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102, 2009

Peter M Aronow, Jonathon Baron, and Lauren Pinson. A note on dropping experimental subjects who fail a manipulation check. *Political Analysis*, 27(4):572–589, 2019

Joel L Horowitz and Charles F Manski. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, 95(449):77–84, 2000

Alexander Coppock, Alan S Gerber, Donald P Green, and Holger L Kern. Combining double sampling and bounds to address nonignorable missing outcomes in randomized experiments. *Political Analysis*, 25(2):188–206, 2017

4 Observational Studies

4.1 Introduction: "Controlling for" in observational studies

Required

Marie-Abele C Bind and Donald B Rubin. Bridging observational studies and randomized experiments by embedding the former in the latter. *Statistical Methods in Medical Research*, 28(7):1958–1978, 2019

Sections 9.0 – 9.2 (especially discussion of interpolation and extrapolation) of Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY, 2006

Richard Berk. What you can and can't properly do with regression. *Journal of Quantitative Criminology*, 26(4):481–487, 2010.

Recommended

Chapter 5 of Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017

William G Cochran. The planning of observational studies of human populations. *Journal of the Royal Statistical Society. Series A (General)*, 128(2):234–266, 1965

On the problem of kitchen sink regressions, Christopher H Achen. Toward a new political methodology: Microfoundations and ART. *Annual Review of Political Science*, 5:423–450, 2002

Chapters 11 and 19 (on overly influential points) of John Fox. *Applied Regression Analysis and Generalized Linear Models*. SAGE Publications, Los Angeles, CA, 3rd edition, 2016

4.2 Matching: An introduction

Application

Magdalena Cerdá, Jeffrey D Morenoff, Ben B Hansen, Kimberly J Tessari Hicks, Luis F Duque, Alexandra Restrepo, and Ana V Diez-Roux. Reducing violence by transforming neighborhoods: A natural experiment in Medellín, Colombia. *American Journal of Epidemiology*, 175(10):1045–1053, 2012

Required

Pages 65 – 90 of Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017

Paul R Rosenbaum. Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application*, 7(1):143–176, 2020

Recommended

Chapter 3 of Paul R Rosenbaum. *Observational Studies*. Springer, New York, NY, second edition, 2002, specifically Sections 3.1 – 3.2 and 3.4 – 3.5.

Paul R Rosenbaum. Observational studies: Overview. In Neil J Smelser and Paul B Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 10808–10815. Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], 2001

Robert Bifulco. Can nonexperimental estimates replicate estimates based on random assignment in evaluations of school choice? A within-study comparison. *Journal of Policy Analysis and Management*, 31(3):729–751, 2012.

Kevin Arceneaux. A cautionary note on the use of matching to estimate causal effects: An empirical example comparing matching estimates to an experimental benchmark. *Sociological Methods & Research*, 39(2):256–282, 2010.

4.3 Nuts & bolts of matching

Required

Chapters 7 – 8 of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010,

Chapter 11 Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017,

4.4 Propensity scores methods

Required

Pages 90 – 96 of Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017

Ben B Hansen. Propensity score matching to extract latent experiments from nonexperimental data: A case study. In Neil J. Dorans and Sandip Sinharay, editors, *Looking Back: Proceedings of a Conference in Honor of Paul W. Holland*, volume 202 of *Lecture Notes in Statistics*, chapter 9, pages 149–181. Springer, New York, NY, 2011

Recommended

Donald B Rubin. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74(366a):318–328, 1979

James M Robins, Miguel Ángel Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, 2000

Daniel E Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236, 2007.

Chapters 13 of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010

Paul R Rosenbaum and Donald B Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38, 1985

4.5 Covariate balance and outcome analysis after matching

Required

Ben B Hansen and Jake Bowers. Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23(2):219–236, 2008

Pages 71 – 79 of Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York, NY, 2012

Recommended

Colin B Fogarty. On mitigating the analytical limitations of finely stratified experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):1035–1056, 2018

Nicole E Pashley and Luke W Miratrix. Insights on variance estimation for blocked and matched pairs designs. *Journal of Educational and Behavioral Statistics*, 2020

Kosuke Imai. Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in Medicine*, 27(24):4857–4873, 2008

Kosuke Imai, Gary King, and Elizabeth A Stuart. Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2):481–502, 2008

4.6 Nonbipartite matching

Applications

Cara J Wong, Jake Bowers, Tarah Williams, and Katherine Drake Simmons. Bringing the person back in: Boundaries, perceptions, and the measurement of racial context. *The Journal of Politics*, 74(4):1153–1170, 2012

Cara J Wong, Jake Bowers, Daniel Rubenson, Mark M Fredrickson, and Ashlea Rundlett. Maps in people's heads: Assessing a new measure of context. *Political Science Research and Methods*, 8(1):160–168, 2020

Required

Chapter 11 of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010

Recommended

Bo Lu, Elaine Zanutto, Robert Hornik, and Paul R Rosenbaum. Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association*, 96(456):1245–1253, 2001

Chapter 11 of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010

Bo Lu, Robert Greevy, Xinyi Xu, and Cole Beck. Optimal nonbipartite matching and its statistical applications. *The American Statistician*, 65(1):21–30, 2011

Bo Lu. Propensity score matching with time-dependent covariates. *Biometrics*, 61(3):721–728, 2005

Mike Baiocchi, Dylan S Small, Scott Lorch, and Paul R Rosenbaum. Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association*, 105(492):1285–1296, 2010

José R Zubizarreta, Dylan S Small, Neera K Goyal, Scott Lorch, and Paul R Rosenbaum. Stronger instruments via integer programming in an observational study of late preterm birth outcomes. *The Annals of Applied Statistics*, 7(1):25–50, 2013

4.7 Regression discontinuity designs and "natural experiments"

Application

Devin Caughey and Jasjeet S Sekhon. Elections and the regression discontinuity design: Lessons from close us house races, 1942–2008. *Political Analysis*, 19(4):385–408, 2011

Required

Matias D Cattaneo, Rocío Titiunik, and Gonzalo Vazquez-Bare. The regression discontinuity design. In Luigi Curini and Robert J Franzese Jr., editors, *Sage Handbook of Research Methods in Political Science & International Relations*. Sage Publications, Washington, D.C., 2020

Adam Sales and Ben B Hansen. Limitless regression discontinuity. *Journal of Educational and Behavioral Statistics*, 45(2):143–174, 2020

Luke Keele, Rocío Titiunik, and José R Zubizarreta. Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1):223–239, 2015

Recommended

Jasjeet Sekhon and Rocío Titiunik. On interpreting the regression discontinuity design as a local experiment. In Matias D. Cattaneo and Juan Carlos Escanciano, editors, *Regression Discontinuity Designs: Theory and Applications*, volume 38 of *Advances in Econometrics*, chapter 1. Emerald Group Publishing, Bingley, UK, 2017

Jasjeet S Sekhon and Rocío Titiunik. Understanding regression discontinuity designs as observational studies. *Observational Studies*, 2:174–182, 2016

Ben B Hansen and Adam Sales. Comment on cochran's "observational studies". *Observational Studies*, pages 184–193, 2015

Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001

Guido W Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008

David S Lee. Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2):675–697, 2008

Andrew Gelman and Guido W Imbens. Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456, 2019

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):698–714, 2008

Chapter 6 of Joshua D Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ, 2008

5 Sensitivity analysis

Application

Magdalena Cerdá, Jeffrey D Morenoff, Ben B Hansen, Kimberly J Tessari Hicks, Luis F Duque, Alexandra Restrepo, and Ana V Diez-Roux. Reducing violence by transforming neighborhoods: A natural experiment in Medellín, Colombia. *American Journal of Epidemiology*, 175(10):1045–1053, 2012

5.1 Sensitivity analysis for sharp nulls

Required

Chapter 11 Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017,

Recommended

Chapter 4 of Paul R Rosenbaum. *Observational Studies*. Springer, New York, NY, second edition, 2002

Paul R Rosenbaum. Sensitivity analysis for stratified comparisons in an observational study of the effect of smoking on homocysteine levels. *Annals of Applied Statistics*, 12(4):2312–2334, 2018

Paul R Rosenbaum and Abba M Krieger. Sensitivity of two-sample permutation inferences in observational studies. *Journal of the American Statistical Association*, 85(410):493–498, 1990

Ben B Hansen, Paul R Rosenbaum, and Dylan S Small. Clustered treatment assignments and sensitivity to unmeasured biases in observational studies. *Journal of the American Statistical Association*, 109(505):133–144, 2014

Jesse Y Hsu and Dylan S Small. Calibrating sensitivity analyses to observed covariates in observational studies. *Biometrics*, 69(4):803–811, 2013

5.2 Sensitivity analysis for weak nulls

Required

Colin B Fogarty. Testing weak nulls in matched observational studies. Working Paper, 2020

Recommended

Colin B Fogarty. Studentized sensitivity analysis for the sample average treatment effect in paired observational studies. *Journal of the American Statistical Association*, 115(531):1518–1530, 2020

Colin B Fogarty, Pixu Shi, Mark E Mikkelsen, and Dylan S Small. Randomization inference and sensitivity analysis for composite null hypotheses with binary outcomes in matched observational studies. *Journal of the American Statistical Association*, 112(517):321–331, 2017

6 Additional topics

6.1 Interference

Paul R Rosenbaum. Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007.

Jake Bowers, Mark Fredrickson, and Costas Panagopoulos. Reasoning about interference between units: A general framework. *Political Analysis*, 21(1):97–124, 2013.

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics*, 11(4):1912–1947, 2017.

Charles F Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, 2013.

Susan Athey, Dean Eckles, and Guido W Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.

6.2 Factorial and complex experiments

Tirthankar Dasgupta, Natesh S Pillai, and Donald B Rubin. Causal inference from 2^k factorial designs by using potential outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4):727–753, 2015

Jens Hainmueller, Daniel J Hopkins, and Teppei Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1):1–30, 2014

Naoki Egami and Kosuke Imai. Causal interaction in factorial experiments: Application to conjoint analysis. *Journal of the American Statistical Association*, 114(526):529–540, 2019

Alan S Gerber, Donald P Green, Edward H Kaplan, and Holger L Kern. Baseline, placebo, and treatment: Efficient estimation for three-group experiments. *Political Analysis*, 18(3):297–315, 2010

6.3 Difference-in-Differences

Pages 162 – 167 of Paul R Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press, Cambridge, MA, 2017

Section 4.1 of Alan S Gerber and Donald P Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton, New York, NY, 2012

Chapter 5 of Joshua D Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ, 2008

Michael Lechner. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, 4(3):165–224, 2011

Charles F Manski and John V Pepper. How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions. *The Review of Economics and Statistics*, 100(2):232–244, 2018

Peng Ding and Fan Li. A bracketing relationship between difference-in-differences and lagged-dependent-variable adjustment. *Political Analysis*, 27(4):605–615, 2019

Kosuke Imai and In Song Kim. On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, 29(3):405–415, 2021

6.4 Special topics in matching: Multilevel, risk set, cardinality, generalized full matching and fine balance

José R Zubizarreta and Luke Keele. Optimal multilevel matching in clustered observational studies: A case study of the effectiveness of private schools under a large-scale voucher system. *Journal of the American Statistical Association*, 112(518):547–560, 2017

Shu Yang, Guido W Imbens, Douglas E Cui, Zhanglin abd Faries, and Zbigniew Kadziola. Propensity score matching and subclassification in observational studies with multi-level treatments. *Biometrics*, 72(4):1055–1065, 2016

Samuel D Pimentel, Lindsay C Page, Matthew Lenard, and Luke Keele. Optimal multilevel matching using network flows: An application to a summer reading intervention. *Annals of Applied Statistics*, 12(3):1479–1505, 2018

Chapter 12 of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010

Yunfei Paul Li, Kathleen J Propert, and Paul R Rosenbaum. Balanced risk set matching. *Journal of the American Statistical Association*, 96(455):870–882, 2001

Chapter 10 of Paul R Rosenbaum. *Design of Observational Studies*. Springer, New York, NY, 2010

Cinar Kilcioglu and José R Zubizarreta. Maximizing the information content of a balanced matched sample in a study of the economic performance of green buildings. *The Annals of Applied Statistics*, 10(4):1997–2020, 2016

Fredrik Sävje, Michael J Higgins, and Jasjeet S Sekhon. Generalized full matching. *Political Analysis*, 2021

6.5 Attributable effects and inference of random causal quantities

Paul R Rosenbaum. Effects attributable to treatment: Inference in experiments and observational studies with a discrete pivot. *Biometrika*, 88(1):219–231, 2001

Paul R Rosenbaum. Exact confidence intervals for nonconstant effects by inverting the signed rank test. *The American Statistician*, 57(2):132–138, 2003

Ben B Hansen and Jake Bowers. Attributing effects to a cluster-randomized get-out-the-vote campaign. *Journal of the American Statistical Association*, 104(487):873–885, 2009

Jasjeet S Sekhon and Yotam Shem-Tov. Inference on a new class of sample average treatment effects. *Journal of the American Statistical Association*, 116(534):798–804, 2020

Luke Keele, Dylan Small, and Richard Grieve. Randomization-based instrumental variables methods for binary outcomes with an application to the 'IMPROVE' trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(2):569–586, 2017

Paul R Rosenbaum. Confidence intervals for uncommon but dramatic responses to treatment. *Biometrics*, 63(4):1164–1171, 2007

6.6 Regression-based sensitivity analysis

Carrie A Hosman, Ben B Hansen, and Paul W Holland. The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder. *The Annals of Applied Statistics*, 4(2):849–870, 2010.

Carlos Cinelli and Chad Hazlett. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1):39–67, 2020.

Guido W Imbens. Sensitivity to exogeneity assumptions in program evaluation. *The American Economic Review*, 93(2):126–132, 2003.

Emily Oster. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204, 2019.

6.7 External validity

Holger L Kern, Elizabeth A Stuart, Jennifer Hill, and Donald P Green. Assessing methods for generalizing experimental impact estimates to target populations. *Journal of Research on Educational Effectiveness*, 9(1):103–127, 2016

Luke W Miratrix, Jasjeet S Sekhon, Alexander G Theodoridis, and Luis F Campos. Worth weighting? How to think about and use weights in survey experiments. *Political Analysis*, 26(3):275–291, 2018

Magdalena Bennett, Juan Pablo Vielma, and José R Zubizarreta. Building representative matched samples with multi-valued treatments in large observational studies. *Journal of Computational and Graphical Statistics*, 29(4):744–757, 2020

Jeffrey H Silber, Paul R Rosenbaum, Richard N Ross, Justin M Ludwig, Wei Wang, Bijan A Niknam, Nabanita Mukherjee, Philip A Saynisch, Orit Even-Shoshan, Rachel R Kelz, and Lee A Fleisher. Template matching for auditing hospital cost and quality. *Health Services Research*, 48(5):1446–1474, 2014

Daniel Westreich, Jessie K Edwards, Catherine R Lesko, Stephen R Cole, and Elizabeth A Stuart. Target validity and the hierarchy of study designs. *American Journal of Epidemiology*, 188(2):438–443, 2019

6.8 Weighting methods

Ambarish Chattopadhyay, Christopher H Hase, and José R Zubizarreta. Balancing versus modeling approaches to weighting in practice. *Statistics in Medicine*, 39(24):3227–3254, 2020

Kosuke Imai and Marc Ratkovic. Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263, 2014

José R Zubizarreta. Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922, 2015

Raymond K W Wong and Kwun Chuen Gary Chan. Kernel-based covariate functional balancing for observational studies. *Biometrika*, 105(1):199–213, 2018

Zhiqiang Tan. Regularized calibrated estimation of propensity scores with model misspecification and high-dimensional data. *Biometrika*, 107(1):137–158, 2020

Yixin Wang and José R Zubizarreta. Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika*, 107(1):93–105, 2020

Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46, 2012

6.9 Synthetic control

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 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, 2010

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2012

Eli Ben-Michael, Avi Feller, and Jesse Rothstein. The augmented synthetic control method. *Journal of the American Statistical Association*, 2021

6.10 Bayesian causal inference

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6(1):34–58, 1978

Chapter 8 of Guido W Imbens and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, New York, NY, 2015

Guido W Imbens and Donald B Rubin. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25(1):305–327, 1997

Luke Keele and Kevin M Quinn. Bayesian sensitivity analysis for causal effects from 2×2 tables in the presence of unmeasured confounding with application to presidential campaign visits. *The Annals of Applied Statistics*, 11(4):1974–1997, 2017

Peng Ding and Luke Miratrix. Model-free causal inference of binary experimental data. *Scandinavian Journal of Statistics*, 46(1):200–214, 2019