

SOCY 7111 Data III: Advanced Data Analysis (Causal inference)

Fall 2019

Mondays 3pm-5:30pm, KTCH 1B24

Professor Amanda Stevenson

Office: KTCH 268

Office Hours: Wednesdays 10am-12pm

Overview

This seminar will survey the major quantitative empirical methods for causal inference in sociology. It is appropriate for students with a strong foundation in regression. Course prerequisites are SOCY Data I and SOCY Data II or equivalent. We will cover experiments, matching and stratification, panel data and fixed effects (including regression discontinuity), difference-in-differences, synthetic control, and instrumental variables in their historical context (including two-stage, latent variable selection, and control function models).

Goals

At the end of this course, students should be able to:

- Describe when using each of the methods studied in the class may be applied to yield inference about causality and when it should not be.
- Make an analysis plan which would establish the grounds for using each of the methods and execute the needed analyses using Stata.
- Execute each method using Stata.
- Provide and respond to constructive criticism.

Materials

- We will use Stata software in this course. You – or at minimum a member of your group – will need to bring a computer with Stata installed to each class meeting.
- Throughout course materials will be drawn from a variety of sources, our primary textbook will be:

Morgan and Winship. *Counterfactuals and causal inference*. Cambridge University Press, 2015. 2nded.

Supplementary readings (rigorous methodologically but written in a more colloquial style) and coverage of additional subjects will be drawn from Scott Cunningham's draft of his forthcoming book on causal inference:

Cunningham. *Causal inference: the mixtape*. Tufte-LaTeX, 2018. Version 1.7, Available from http://scunning.com/cunningham_mixtape.pdf

Structure of the course

Doing quantitative social science is a practice. In this course, you will practice actively reading about methods, talking with the group, analyzing data by following instructions, analyzing data by designing and conducting your own analyses, helping others improve their analytic work, justifying your own analyses, and responding to reviews.

Group work: This class aims to provide you experience working in the way that most quantitative social science actually happens – collaboratively confronting the messiness of real data with a computer and some colleagues – none of whom have much time – to try to figure out what is going on.

Therefore, students will be placed (by the second class meeting) in groups of 3-4. In the interest of inclusion, I will form the groups on the basis of our conversation in the first class meeting, aiming to build groups which share enough substantive overlap that they can productively collaborate on an original analytic project. Students may email me their preferences regarding group members. Graded group work will be composed of four types of assignments:

In-class labs: Most sessions will include an in-class lab, where students will follow guidelines from me to undertake a replication of a published study using the method we are studying. These will often take longer than the class period to complete, so groups should schedule time each week to complete the analysis outlined and to write up their results in a lab report for submission. I recommend setting up a standing weekly group meeting.

Group project development assignments: Groups will also follow structured guidelines to develop an empirical project using one of the methods we will learn to answer a causal question of interest to the group. Most weeks the group will share (publicly within the class on Canvas) an output responding to a prompt. Sharing will be public within the class so that students may all observe their colleagues' idea development, an important component of the research process.

Group project extended abstract and presentation: Each group will eventually generate a polished extended abstract and a 15 minute presentation.

Individual participation in class: Each class will involve discussion and participation in lab activities.

Individual marked-up readings: In general, each week each student will read a few papers and/or textbook sections very carefully with a special focus on annotating equations. As homework, students will turn in their marked-up assigned readings on Canvas. We will discuss marking up and why it matters for the kind of reading we are doing in this class. Students may either scan documents with handwritten mark-up or read on a device and mark-up the file using software. Either way, turn in a PDF.

Reviews of other groups' work: At specific intervals, each individual will provide constructive, helpful peer reviews to each group other than their own.

Grades

This is a graduate seminar. I assume that you are taking it because you want to learn the material. I have selected the required activities specifically because they will help you learn, so please undertake them in that spirit. Doing them just because you are concerned about your grade is probably not going to help you learn what you need to. If at any time you are concerned about your learning or your grade in the course, feel free to talk to me about it.

Final grades will be calculated on the basis of the following distribution:

Individual	Participation - in class	20%
Individual	Mark-ups	20%
Individual	Reviews	10%
Group	Lab reports	20%
Group	Project assignments	15%
Group	Final extended abstract	10%
Group	Final presentation	5%

Assignment of letter grades based on percentages of total points will be as follows, employing standard rounding when percentages fall between whole numbers: 94%+ = A; 90%-93%=A-; 87%-89%=B+; 83%-86%=B; 80%-82%=B-; 77%-79%=C+; 73%-76%=C; 70%-72%=C-; 67%-69%=D+; 63%-66%=D; 60%-62%=D-; 59% & below=F.

Course Policies

This is a graduate course and I assume that you are invested in learning the material.

Getting to class late, sleeping, talking out of turn, reading, using your cell phone or other device, or otherwise being distracted and distracting, are not acceptable classroom activities.

Please let me know immediately if you have a health problem or disability that necessitates leaving the room during class time, or if you have a conflict that regularly prevents you from getting to class on time.

If you miss a class, please get notes from a classmate rather than asking me for notes. You are responsible for all information communicated in class, whether or not you are in attendance. In addition, information on scheduling changes, assigned work, and grades will be emailed and/or posted on Canvas. Make sure to check your CU email regularly.

All assignments must be completed on time and handed in on Canvas by the start of class on the stated due date. Emailed documents will not be accepted except in extraordinary circumstances. Written work must be spell-checked, grammar-checked, and proofread. The quality of your writing will impact your grade.

I will not give incompletes in this course except under extremely unusual, well-documented circumstances.

Please come to office hours if you have any questions about the material covered in this course.

I want you to do well in this class. If you become concerned about your progress, please see us immediately. Do not wait until the end of the semester when you have no more work to complete that can raise your grade.

Communication

If you email me about the course, be sure to use the course name in the subject line. I will check course email once per day on weekdays. I will strive to respond to your course emails within 24 hours.

Important information will be communicated over email via Canvas, so make sure that you regularly check the email address listed for you in CU's administrative system. The course web page is on Canvas, accessed at canvas.colorado.edu. The syllabus, schedule, handouts, assignments, grades, and other information will be posted on the website, and you will turn

in assignments there. The most recent course information will be available here, as information that is on this syllabus may be changed over the course of the semester. You must regularly log into Canvas so that this important information can reach you.

Because of privacy, we cannot discuss grades over email, so come to office hours if that is your subject of concern. Please be aware that email is not a secure method of communication.

Accommodation for Disabilities

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the [Disability Services website](#). Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition or injury, see [Temporary Medical Conditions](#) under the Students tab on the Disability Services website.

Classroom Behavior

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. For more information, see the policies on [classroom behavior](#) and the [Student Code of Conduct](#).

Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code

(honor@colorado.edu); 303-492-5550). Students who are found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found at the [Honor Code Office website](#).

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

The University of Colorado Boulder (CU Boulder) is committed to fostering a positive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct intimate partner abuse (including dating or domestic violence), stalking, protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, [anonymous reporting](#), and the campus resources can be found on the [OIEC website](#).

Please know that faculty and instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, talk to me in office hours **at least two weeks before the assignment for which an accommodation for a religious obligation is needed**.

See the [campus policy regarding religious observances](#) for full details.

Schedule of topics

8/26/19	Introduction
9/2/19	No class - Labor Day
9/9/19	Theory and math

9/16/19	Experiments
9/23/19	Matching and stratification I: Theory and simulation
9/30/19	Matching and stratification II: Applications
10/7/19	Panel data I: Regression discontinuity
10/14/19	Panel data II: Applications
10/21/19	Difference in differences I
10/28/19	Difference in differences II
11/4/19	Synthetic control
11/11/19	Instrumental variables I
11/18/19	Instrumental variables II
11/25/19	No class - Thanksgiving week
12/2/19	Collaborative peer review
12/9/19	Presentations

Introduction (8/26)

In-class interpretation: Gangl, Markus. "Causal inference in sociological research." *Annual review of sociology* 36 (2010).

Alternative in-class interpretation: Angrist, Joshua D., and Jörn-Steffen Pischke. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *Journal of economic perspectives* 24.2 (2010): 3-30.

Theory and math background (9/9)

Read and mark-up:

- Morgan and Winship Chapter 2
- Morgan and Winship Chapter 3

Optional reading:

- Morgan and Winship Chapter 1
- Cunningham p. 23-66

Experiments and inferential statistics for causal inference (9/16)

Read and mark-up:

- Angus Deaton and Nancy Cartwright. Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, 2018. Also see responses from Pearl and Imbens (as well as from others in the same issue) below.
- Taubman, Sarah L., et al. "Medicaid increases emergency-department use: evidence from Oregon's Health Insurance Experiment." *Science* 6168 (2014): 263-268. General information about the OHIE is available here: <https://www.nber.org/oregon/3.results.html>
- Brand, Jeffrey E. and Juli Simon Thomas. "Causal Effect Heterogeneity." (2013)

Optional readings:

- Pearl, Judea. "Challenging the hegemony of randomized controlled trials: A commentary on Deaton and Cartwright." *Social Science & Medicine*(2018).
- Imbens, Guido. "Understanding and misunderstanding randomized controlled trials: A commentary on Deaton and Cartwright." *Social science & medicine* (1982)210 (2018): 50-52.
- King, Gary, and Langche Zeng. "Estimating risk and rate levels, ratios and differences in case-control studies." *Statistics in medicine*10 (2002): 1409-1427.

Matching and stratification I (9/23)

Read and mark up:

- Morgan and Winship Chapter 4
- Colson, K. Ellicott, et al. "Optimizing matching and analysis combinations for estimating causal effects." *Scientific reports*6 (2016): 23222.

Optional Reading:

- King, Gary, et al. "Comparative effectiveness of matching methods for causal inference." Unpublished manuscript 15 (2011): 41.

Matching and stratification II (9/30)

Read and mark-up:

- Morgan and Winship Chapter 5

- Gough, Margaret. 2017. "Birth Spacing, Human Capital, and the Motherhood Penalty at Midlife in the United States." *Demographic Research* 37:363–416.

Optional reading:

- Iacus, Stefano M., Gary King, and Giuseppe Porro. "A theory of statistical inference for matching methods in applied causal research." URL: <http://gking.harvard.edu/publications/how-Coarsening-Simplifies-Matching-Based-Causal-Inference-Theory> (2015).

Panel Data I: Regression Discontinuity (10/7)

Read and mark up:

- Meyersson (2014): Islamic Rule and the Empowerment of the Poor and Pious, *Econometrica* 82(1): 229-269.
- Imbens, Guido W. & Lemieux, Thomas, 2008. "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, vol. 142(2), pages 615-635.

Optional reading:

- *(Highly recommended)* Cunningham p. 245-262
- *(Highly recommended)* Calonico, Sebastian, et al. "Regression discontinuity designs using covariates." *Review of Economics and Statistics* 3 (2019): 442-451.
- Morgan and Winship remainder of Chapter 11
- Cattaneo, Matias D., Nicolás Idrobo, and Rocio Titiunik. "A Practical Introduction to Regression Discontinuity Designs: Volume II." (2018). (Pages TBA)
- (Forthcoming final version of (1)) Cattaneo, M., Idrobo, N., & Titiunik, R. (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations* (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge: Cambridge University Press.
- *Alternative for replication:* Amarante, Verónica, et al. "Do cash transfers improve birth outcomes? Evidence from matched vital statistics, program, and social security data." *American Economic Journal: Economic Policy* 2 (2016): 1-43.

Panel Data II: Issues in RD (10/14)

Read and mark up:

- Athey, Susan, et al. "Ensemble Methods for Causal Effects in Panel Data Settings." *AEA Papers and Proceedings*. Vol. 109. 2019.

- Cattaneo, Titiunik and Vazquez-Bare. Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality, *Journal of Policy Analysis and Management* 36(3): 643-681, Summer 2017.

Optional reading:

- Guido Imbens, Karthik Kalyanaraman, 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator, *The Review of Economic Studies*, vol. 79(3), pages 933–959.
- The remainder of Cattaneo et al. (2019) above (*To be selected by AJS on the basis of concerns among groups who are considering RD for their projects*)

Difference-in-differences I (10/21)

Read and mark up:

- Card, David, and Alan B. Krueger. "Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania" *American Economic Review* (1994). http://davidcard.berkeley.edu/data_sets.html
- Cunningham p. 263-286

Optional reading:

- (*HIGHLY recommended if you are new to DID*) Dimick, Justin B., and Andrew M. Ryan. "Methods for evaluating changes in health care policy: the difference-in-differences approach." *Jama* 22 (2014): 2401-2402.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How much should we trust differences-in-differences estimates?." *The Quarterly journal of economics* 1 (2004): 249-275.
- Imbens, Guido W., and Michal Kolesar. "Robust standard errors in small samples: Some practical advice." *Review of Economics and Statistics* 4 (2016): 701-712.

Difference-in-differences II (10/28)

Substantive focus TBD, depending on progress thus far and course interests.

Options for additional topics in DID:

- Triple difference (DDD)
- Semi-parametric and nonparametric vs. parametric approaches
- Problems associated with the number of treated units
- Heterogeneity of effects

Examples

- Cornwell, Christopher, and Scott Cunningham. "The long-run effect of abortion on sexually transmitted infections." *American law and economics review*1 (2013): 381-407.
- Donohue III, John J., and Steven D. Levitt. "The impact of legalized abortion on crime." *The Quarterly Journal of Economics*2 (2001): 379-420.
- Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez. "Designing difference in difference studies: best practices for public health policy research." *Annual review of public health*39 (2018).
- Ferman, Bruno, and Cristine Pinto. "Inference in differences-in-differences with few treated groups and heteroskedasticity." *Review of Economics and Statistics*3 (2019): 452-467.

Synthetic control (11/4)

Read and mark up:

- Abadie, Alberto, 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*490 (2010): 493-505.
- Cunningham p. 287-313

Optional reading:

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Comparative politics and the synthetic control method." *American Journal of Political Science*2 (2015): 495-510.
- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synth: An r package for synthetic control methods in comparative case studies." *Journal of Statistical Software*13 (2011).

Precursors of instrumental variables approaches (11/11)

Read and mark up:

- Cunningham p. 205-214
- Winship, Christopher, and Robert D. Mare. "Models for sample selection bias." *Annual review of sociology*1 (1992): 327-350.

Optional reading:

- Morgan and Winship 291-324

Instrumental variables approaches (11/18)

Read and mark up (these readings may change):

- Cunningham p. 215-245
- Card, David. *Using geographic variation in college proximity to estimate the return to schooling*. No. w4483. National Bureau of Economic Research, 1993.

Optional reading:

- Heckman, James J. "Econometric causality." *International statistical review* 1 (2008): 1-27.
- Morgan and Winship Chapter 9 (focus only on specific sections discussed in class)

Collaboration on peer review (12/2)

Presentations (12/9)

Schedule of labs and assignments (tentative)

Class date	Topic	Reading mark-up due*	Reviews*	Group work due – project **	Group work due – lab reports*
8/26/19	Introduction	None	None	None	None
9/2/19	No class (Labor Day)				
9/9/19	Theory and math	Gangl; M+W ch. 2+3	None	Upload two papers you like from your area that make causal claims. Write a couple sentences about how each makes these claims.	None

9/16/19	Experiments	Deaton + Carthwright; Taubman et al.	None	(1) Outline a causal question in your area and describe two experiments you would like conduct to answer it in an ideal world. (2) List and summarize observational data which you have access to and permission to use.	DAGs and summaries assignment
9/23/19	Matching and stratification I	M+W ch. 4; Colson et al.	None	Write a ~1 paragraph summary of the role of endogeneity for each of the two causal questions you outlined last week. Then for each, list the observable and unobservable sources of endogeneity. This will take the form of a list of lists or a matrix.	OHIE report
9/30/19	Matching and stratification II	M+W ch. 5; Gough	Brief	Choose 2-3 combinations of causal questions and datasets and for each combination, outline the advantages and disadvantages of each of the matching procedures in Colson. State which you consider the most promising and explain why.	Colson et al. report
10/7/19	Panel data I: Regression discontinuity	Meyersson; Imbens + Lemieux	None	Describe 1-2 RD designs which could address your question/s and outline the assumptions in each.	Gough report
10/14/19	Panel data II: Concerns	Athey et al.; Cattaneo et al.	Brief	Do a preliminary implementation of any	Meyersson report

				RD design using your data and write <2 page reflecting on its strengths and weaknesses.	
10/21/19	Difference in differences: I	Cunningham p. 263-286; Card + Krueger	None	Even if your data do not allow it, write ~1 page about a natural experiment where the cause whose effect you're interested in changed for some units but not for all.	Athey report
10/28/19	Difference in differences: II	TBD	Brief	Outline DID design for your question - a hypothetical one if none exists or a real one if one does.	None
11/4/19	Synthetic control approaches	Cunningham p.287-313; Abadie 2010	Brief	Write a principled ~1 page supporting either a DID or SC design addressing your question.	Card and Krueger report
11/11/19	Instrumental variables: I	Cunningham p.205-214; Winship + Mare	Brief	Write a proposal for the analysis you think would best answer your causal question. Include at minimum preliminary descriptive statistics.	Abadie report
11/18/19	Instrumental variables: II	Cunningham p.215-245; Card 1995	Brief	Revise your proposals to respond to our conversation in class or otherwise strengthen them. Include at minimum a table or figure or other main effects result. Include your output as an appendix.	Morgan or Card report
11/25/19	No class - Thanksgiving week				

12/2/19	Collaborative peer review	None	Extended	Write a polished extended abstract summarizing your question, briefly motivating it (~1 paragraph), describing your materials and methods, outlining your results, and concluding with the meaning of your finding. Include your output as an appendix.	None
12/9/19	Presentations	None	None	Revise your extended abstract in light of the conversation 12/2 and develop a 15 minute presentation on your work. Give your presentation in class.	None

* Due before class begins.

**Due before class begins or 5pm the night before for work to reviewed by classmates

Some Additional Resources

(If you are serious about using these methods, buying and reading this book is an excellent move. -AJS) Angrist, Joshua D. and Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

(If you are interested in the theoretical debates around how causality is established, Pearl is an essential read and this is his broadest book on the subject. - AJS) Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press. 2nd edition.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments*. W. W. Norton.

Imbens, Guido W. and Jeffrey Wooldridge. 2009. *Recent Developments in the Econometrics of Program Evaluation*. Journal of Economic Literature, 47(1): 5–86.

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

Manski, Charles. 1995. *Identification Problems in the Social Sciences*. Cambridge: Harvard University Press.

Rosenbaum, Paul R. 2009. *Design of Observational Studies*. Springer Series in Statistics.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.