EDP 384/25: Causal Inference
Fall 2014, Tues/Thur, 2:00-3:20 pm
SZB 432

Instructor: James E. Pustejovsky
Email: pusto@austin.utexas.edu
Phone: 512-471-0683
Office: SZB 538D
Office hours: Mondays, 1-3 pm or by appointment

Course Description

This course introduces the contemporary statistical approach to addressing questions about the causal effects of programs or policies, focusing in particular on applied data-analysis strategies and interpretation. The course begins with an introduction to the potential outcomes framework for expressing causal quantities, followed by an examination of (idealized) simple and block randomized experiments as prototypes for learning about causal effects. The remainder of the course covers theory and data-analysis strategies for drawing causal inferences from observational studies, in which treatment conditions are not randomly assigned. Analysis techniques such as matching methods, propensity-score methods, prognostic score methods, and instrumental variables are covered both in theory and in application. Further, advanced topics will be covered at the discretion of the instructor, based on time and student interest.

Learning Goals

After completing this course, students should be able to:

- Translate research questions into the framework of the potential outcomes model and specify causal quantities to be estimated.
- Articulate and assess the assumptions behind different strategies for estimating treatment effects and drawing causal inferences.
- Understand the conditions under which different causal inference strategies work well.
- Conduct, interpret, and defend a causal analysis of observational data.
- Critically review published research that addresses causal questions.

Prerequisites

- A previous course in research methods, such as EDP 384, Topic 24 – Research Design and Methods for Psychology and Education
- A previous course in regression analysis, such as EDP 382K – Correlation and Regression
- Experience with at least one software package for data management and analysis (e.g., R, SPSS, SAS, Stata). Note that the data-analysis exercises and course project can be completed using your choice of software. ***In-class software demonstrations will use the R environment for statistical computing.*** For a very quick introduction to R, see http://tryr.codeschool.com

Readings

- Required textbook: Morgan, Stephen & Christopher Winship (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press.
- Additional readings posted on Canvas.

Evaluation

- Article presentation (10%). See below.
- Data analysis exercises (20%). Students will complete several short data-analysis exercises involving analysis of real or simulated data. The exercises will involve implementing different analytic methods and interpreting the results.
- Course project (70%). See below.

Article Presentations

Over the course of the semester, each student will make a 10-12 minute presentation on an empirical research article. The presentation should cover: 1) the motivation for the research, 2) the main research question(s), including a description of the target causal effect 3) an explanation of the analytic methods employed, 4) a succinct summary the findings, and 5) a critique of the methods employed.

Course project

Students will work individually or with a partner to analyze an observational study using the causal inference methods developed in class. At the end of the semester, each student/pair will submit a paper (no more than 25 pages, double-spaced) presenting the results of the study. This paper should be of publication quality (as might appear in *Educational Evaluation and Policy Analysis* or *Evaluation Review*) and should focus on the research question, data, empirical strategy, results, and conclusions; the data-analysis code should be submitted as an appendix. Students are free to choose any topic they want, as long as they have a clear research question that concerns the causal effect of some intervention, treatment, policy, or event on some outcome, result, or performance. The breakdown for the grade is as follows:

- **Proposal (15%)**: Proposal for project using Structured Abstract
- **Draft paper (15%)**: Preliminary results from the analysis, including description of the empirical strategy and results, plus relevant tables and figures.
- **Final paper (40%)**

Finding data for the course project

Students will need to identify a dataset (or multiple datasets) for their course project. A good dataset needs to meet several requirements.

1. It has to contain variables that are of theoretical interest, including variables that measure a cause/treatment/intervention and variables that measure outcomes of interest.
2. It has to be large enough to be amenable to the techniques we will be covering in class. A small survey on a convenience sample that you conducted for a previous class will not be very useful. A nationally representative probability sample survey would be an ideal data source. Large administrative data-bases are also often useful.

3. It has to contain a rich set of information about the individuals—the more background variables, the better.
4. If the data are proprietary, you must have any necessary permissions to work with them.
5. Some suggested resources for locating interesting and workable data:

    - The Inter-University Consortium for Political and Social Research: www.icpsr.umich.edu
    - The National Center for Education Statistics: http://ies.ed.gov/funding/datasets.asp
    - The Federal data clearinghouse: https://www.data.gov/
    - Many cities/municipalities now have data clearinghouses. For example:
        - https://data.austintexas.gov/
        - https://data.cityofchicago.org/
        - https://data.cityofnewyork.us/
    - Ask your professors about what data sources they use!

Attendance

Students are responsible for all of the material presented during class meetings. If a student must miss a class, it is their responsibility to obtain from classmates and thoroughly review notes or summaries of the material that they missed.

Academic integrity and plagiarism

Following the University's honor code, students are expected to maintain absolute integrity and a high standard of individual honor in scholastic work. Assignments and projects must be completed with the utmost honesty, which includes acknowledging the contributions of other sources to your scholastic efforts; avoiding plagiarism; and completing assignments independently unless expressly authorized otherwise. ***Homework assignments or projects containing any plagiarized material will not be accepted.***

ADA accommodations

The University of Texas at Austin provides upon request appropriate accommodations for qualified students with disabilities. For more information, please contact the Office of the Dean of Students at 471-6259, 471-4671 TTY.

Religious Holidays

By UT Austin policy, students must notify the instructor of a pending absence due to religious observance at least fourteen days in advance. If the student must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, the student will be given an opportunity to complete the missed work within a reasonable time after the absence, with no penalty.

Course Outline

This outline is tentative and will almost certainly evolve over the course of the semester.

8/28 – Course introduction
9/2 – Probability crash course
9/4 – Potential outcomes models for causal effects
9/9 – Simple randomized experiments
9/11 – Block-randomized experiments
9/16 – Observational studies and confounding
9/18 – Directed acyclic graphs
9/23 – Regression analysis
9/25 – Stratification/matching on a single-covariate
9/30 – Multivariate matching using propensity scores
10/2 – Propensity score analysis: Stratification
10/7 - Propensity score analysis: Matching strategies
10/9 – Balance assessment
10/14 - Some empirical examples
10/16 – Propensity score analysis: Weighting strategies
10/21 – Prognostic scores
10/23 – Other approaches to matching
10/28 – Choosing covariates
10/30 – Instrumental variables
11/4 – Instrumental variables analysis
11/6 – Regression discontinuities
11/11 – Regression discontinuity analysis
11/13 – Difference in differences
11/18 – Fixed effects models
11/20 – TBD
11/25 – TBD
12/2 – TBD
12/4 – TBD

Potential additional topics

1. Sensitivity analysis for unobserved confounders.
2. Causal mediation analysis.
3. Partial identification analysis.
4. Within-study comparison of experimental and non-experimental treatment effect estimates.
5. Generalization from experiments using propensity-score methods.
6. Double-robustness.
7. Multi-level matching.
8. Principal stratification