Causal inference for the social sciences
2011 ICPSR Summer Program, Session 2

Abstract: This course provides an introduction to statistical methods used in causal inference. The content is geared specifically toward students and researchers in the social sciences. Using the potential outcomes framework of causality, topics covered include research designs such as randomized experiments and observational studies. Classes will explore the impact of noncompliance in randomized experiments, as well as nonrandom assignment in observational studies. To analyze these research designs, the methods covered include matching, instrumental variables, difference-in-difference, and regression discontinuity. Examples are drawn from economics, political science, public health and sociology.

Prerequisites for the course are knowledge of multiple regression using linear algebra and some familiarity with limited dependent variables. The course will rely on R for computation.

Instructor: Ben Hansen, Associate Professor of Statistics, University of Michigan; ben.hansen@umich.edu.

Teaching Assistant: Yun-Jia Lo, Ph.D. candidate in Measurement and Quantitative Methods, College of Education, Michigan State University; loyunjia@msu.edu

Course Meetings: 1 Modern Languages Building, 3-5p.m. M-F
Instructor office hours: 11:30 a.m.-12:30 p.m., MTuThF; 445F West Hall.
T.A. office hours: 2-3 p.m., MTuW; 5-6 p.m., M-Th. 220 Newberry Hall.

Description: 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.

The requirements for causal inference become clearer when it is framed in terms of potential outcomes, as was first done early in the 20th century but became standard only by that century’s end. In 1922 Neyman first proposed such a statistical framework, using it for analysis of randomized experiments. More recently, his intellectual disciples have clarified the special demands placed on common statistical methods by attempting to draw causal inferences with them. 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. They clarified the requirements for causal inference with observational data, isolating the distinct contributions required of the statistician and of his disciplinary collaborators; used both Fisher’s and Neyman’s frameworks to describe vividly more attainable targets of causal estimation; offered methods with which to address potential confounding due to measured variables more comprehensively and more satisfyingly than was previously possible; qualitatively and quantitatively advanced our grasp of unmeasured confounding and its potential ramifications; 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 this school, founded by Fisher’s student W.G. Cochran, will be the central task of this course.

The course begins with the strengths of random assignment and the limitations of statistical
modeling as bases for causal analysis, as presented by Neyman and his latter-day champion David Freedman. It then turns to Donald Rubin, a leading figure in the Cochran school, and his “counterfactual” or potential outcome modeling framework. The framework yields methods of analysis for experiments that would previously have been described as “broken” and for observational studies that might previously have been dismissed as hopelessly confounded. We then turn to Paul Rosenbaum’s development of Cochran’s themes and methods. At a technical level, this involves matching, permutation inference, sensitivity analysis and design sensitivity; more broadly, it emphasizes simplicity of statistical results, a modular design of statistical methods (which facilitates diagnostics and the checking of computations), robustness, transparency of assumptions, and planning data collection and analysis so as to minimize potential for bias and maximize credibility.

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. Most of the course will require little in the way of statistical computing, although computing oriented students will have the option of satisfying the course paper requirement with a data analysis—for which it will be permissible to use packages other than R, perhaps in combination with R for specific functions.

**Required Texts:**
If you’re new to R, you might also get hold of:
Other readings will be assigned and distributed electronically through a password-protected “ctools” Web site (http://ctools.umich.edu).

**Assignments**
Computer or pencil and paper exercises will be assigned periodically and collected on Mondays and on Thursdays at the beginning of class. Late homework will not be accepted without cause. Students will be expected to do readings and sometimes exercises in preparation for most class meetings; these assignments will be the basis for periodic unannounced quizzes at the beginning of the lecture period.

Course participants will be expected to give presentations based on assigned readings, in small teams to be designated by the instructor. Each participant should plan to participate in one or two such teams, each consisting of two or three students and responsible for a prepared, 30 minute presentation.

Participants taking the course for grade will also be expected to submit a written assignment. One possibility is to prepare a short paper of 4 to 8 pages, two days in advance of your in-class presentation, laying out an argument you might present. (This argument might apply what’s described in the reading to debates or problems you’re familiar with, critically evaluate what you find in the reading in light of other methodological principles, or critically evaluate other methodological principles or practices in light of what you find in the reading.) Feedback on this paper should inform a revision of the short paper to be submitted on or before the last day of class. Another possibility is to prepare a similar short paper engaging with some paper from the causal inference literature that doesn’t make it onto our reading list (which the student would select in consultation with the instructor). A third possibility is to execute and briefly write up a data analysis performed using methods discussed in the course.