 

**Course title: Causal Inference**

**Name of the instructors: Xi Chen**

**Prerequisites (knowledge of topic):**

* The basic knowledge of probability and statistics such as conditional probability, hypothesis testing and regression.
* Experiences of working with data (with any software package).
* No prior knowledge on causality and causal inference is required.

**Hardware:**

A laptop for the practical sessions.

**Software:**

R and R Studio (both are freely-available).

**Learning objectives.**

Upon the completion of the course, your will be able to:

* **Recognize** research problems that require causal inference.
* **Describe** different perspectives and models in causal inference.
* **Implement** a selection of tools in causal inference.
* **Evaluate** critically in-lab and field experiments and other research designs.
* **Develop** research designs to address causal inference problems.

**Course content.**

Causal questions in the form of how X influences Y are pervasive in real life. It is therefore imperative for us to know how to address these questions, especially given the “big data revolution” in the last decade. Moreover, without the understanding of causal inference, we can easily fall victim to misinformation. For example, in response to Apple’s new privacy policy on the mobile system, Facebook launched a series of full-page newspaper ads, claiming that Apple’s new privacy policy would hurt small business advertisers. Facebook concluded that for small business advertisers, the new policy would lead to “a cut of 60% in their sales for every dollar they spend.” However, is the claim credible? How do we judge its credibility? To answer these questions, in this course, you are introduced to the exciting area of causal inference.

This course provides you with conceptual understandings, as well as tools to learn causality from data. These understandings and tools come from the rapidly developing science of causal inference. On the conceptual level, the course covers basic concepts such as causation vs. correlation, causal inference, causal identification and counterfactual. It also presents perspectives and tools to help you formalize and conceptualize causal relationships. These perspectives and tools are synergized from multiple disciplines, including statistics (e.g., Robin Causal Model or Potential Outcome Framework), computer science (e.g., Pearlian Causal Model or Causal Graph), and econometrics (e.g., identification strategies and local average treatment effect).

In this course, we will also discuss a selection of tools in causal inference. We will start with the completely randomized experiment and discuss the assignment mechanism, Fisher’s exact p-value, and Neyman’s repeated sampling approach. From then on, we will gradually relax the assumptions of complete randomization and discuss situations where the complete randomization does not hold. Specifically, we will discuss the following: First, block randomization and conditional random assignment, with a focus on matching and weighting estimators; Second, non-compliance where the random assignment fails and the local average treatment effects; Third, attrition where some outcomes are missing and the bounding approach to attrition; Fourth, research designs when the assignment mechanism is unknown to or uncontrolled by us, including the difference-in-difference approach and regression discontinuity design. Moreover, in the last day of the course, we will discuss the ethics of causal inference and the assessment of the unconfoundedness assumption. As the “finale,” we will go through the most recent developments in the intersection between causal inference and machine learning, where machine learning techniques are used to address causal inference problems.

One major distinction of the course is its emphasis on practical relevance. Throughout the course, you are given cases and real data to apply what you learn to real causal inference problems. The course is split between lectures and practical sessions. Cases and data will be provided by the instructor before class.

**Structure.**

Day 1:

1. Introduction to causality and causal inference.
2. Overview of the course.
3. Causal graph: a new language of causality.
4. Causal identification.
5. The formulation of an empirical strategy.
6. Practical session: Using Daggity to draw and analyze causal graphs.

Day 2:

1. The potential outcome framework.
2. The assignment mechanism.
3. Fisher vs. Neyman’s treatment of completely randomized experiments.
4. Counterfactual: Robin vs. Pearl.
5. Block randomization.
6. Selection on observables: subclassification, matching and weighting.
7. Practical session: Matching and weighting estimation of the effect of virtual-fitting on sales.

Day 3:

1. Non-compliance problem in the treatment assignment.
2. The intention-to-treat and the local average treatment effect (LATE).
3. The link of LATE to the instrumental variables strategy.
4. The attrition problem after randomization.
5. Classic treatments of the attrition problem.
6. The bounding approach and the Lee bound.
7. Practical session: Estimating LATE in a field experiment of coupons.

Day 4:

1. Canonical difference-in-difference (DID) model.
2. Two-way fixed estimation of the multiple-period DID.
3. DID with staggered adoptions of the treatment.
4. The basic idea of the regression continuity design (RDD).
5. How to estimate the LATE with RDD.
6. Regression kink design and the bunching method.
7. Practical session: DID analysis of the Low Carbon London Electricity Trial.

Day 5:

1. Ethics of causal inference: being transparent about your assumptions.
2. Assessing the assumption of unconfoundedness.
3. Sensitivity analysis in regression adjustment.
4. Placebo and falsification test in DID and RDD.
5. Recent applications of machine learning techniques in causal inference.

**Literature.**

**Lecture notes will be shared. You are recommended to read them before class.**

**Books (all supplementary)**

* Gerber, Alan S., and Donald P. Green. Field experiments: Design, analysis, and interpretation. WW Norton, 2012.
* Guido W. Imbens and Donald B. Rubin. Causal inference for statistics, social, and biomedical sciences: An Introduction. Cambridge University Press, 2015.
* Lee, Myoung-jae. Matching, regression discontinuity, difference in differences, and beyond. Oxford University Press, 2016.
* Morgan, Stephen L., and Christopher Winship. Counterfactuals and causal inference. Cambridge University Press, 2015.
* Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. Causal inference in statistics: A primer. John Wiley & Sons, 2016.

**Articles**

**(starred ones are recommended readings and others are optional)**

**(arranged in the same order as the course contents)**

Day 1:

* \* Pearl, J. (1995). Causal diagrams for empirical research. Biometrika, 82(4), 669-688. [Section 1-3].
* Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. Journal of Economic Literature, 57(4), 835-903.
* \* Angrist, J. (2022). Empirical strategies in economics: Illuminating the path from cause to effect (No. w29726). National Bureau of Economic Research.

Day 2:

* \* Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469), 322-331.
* Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. Journal of Economic Literature, 58(4), 1129-79.
* \* Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. Annual Review of Statistics and Its Application, 7, 143-176.
* Imbens, G. W. (2015). Matching methods in practice: Three examples. Journal of Human Resources, 50(2), 373-419.

Day 3:

* Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. Econometrica, 62(2), 467-475.
* \* Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91(434), 444-455.
* \* Manski, C. F. (1990). Non-parametric bounds on treatment effects. American

Economic Review, 80(2), 319-323.

* \* Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Review of Economic Studies, 76(3), 1071-1102.
* Martel García, F. (2013). When and Why is Attrition a Problem in Randomized Controlled Experiments and How to Diagnose it. Available at SSRN 2267120.

Day 4:

* \* Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. Foundations and Trends® in Econometrics, 4(3), 165-224.
* Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277.
* \* Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. Annual Review of Economics, 14, 821-851.
* Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. Journal of Economic Literature, 48, 281-355.
* Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of Econometrics, 142, 615-635.

Day 5:

* \* Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. Biometrika, 74(1), 13-26.
* Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. Journal of Political Economy, 113(1), 151.
* \* Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187-204.
* \* Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242.
* Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. The Econometrics Journal, 21(1). C1-C68.

**Examination part.**

The performance of participants will be assessed by a take-home assignment at the end of the course (100%). The assignment will be due in 3 weeks.

**Supplementary aids.**

The examination aids and documents are listed below:

1. Lecture notes for all the lectures.
2. Starred articles in the literature list.
3. R notebooks shared after the practical sessions.

**Examination content.**

The list of topics that are covered in the assignment are:

* The basics of the causal graph.
* Formulating an identification strategy with the causal graph.
* Evaluating empirical strategies.
* The basics potential outcome framework.
* Calculating Fisher’s exact p-value and Neyman’s repeated sampling statistics.
* Matching and weighting estimators.
* Propensity scores.
* The non-compliance problem.
* Calculating local average treatment effects.
* The attrition problem in experiments.
* The bounding approach.
* Basics of the difference-in-difference approach.
* Basics of the regression discontinuity design.
* Sensitivity analysis.
* Falsification and placebo tests.

**Examination relevant literature.**

Only the starred article in the literature list are necessary for the examination. Books and articles are only supplementary.