Carnegie Mellon University
Tepper School of Business

45-911 (6 units) **Statistical Models for Management and Control: Causal Inference** Spring 2024

Professor: Fallaw Sowell
Office: TEP 4213
Email: fs0v@andrew.cmu.edu
Phone: (412)-268-3769

Class Meetings:

Section A4 Monday-Wednesday 2:00 pm - 3:50 pm rm: TEP 2118

DELIVERABLES

Students will understand the challenges of and modern approaches to estimating average treatment effects. Students will be able to analyze a business problem and determine if it is a forecasting problem or a control problem. Students will be able to use graphical models to represent relationships in a system. The graphical model will then lead to appropriate models to estimate causal effects. Students will know how to estimate causal effects with outcome regressions, inverse probability weights, and doubly robust estimators. Students will be able to investigate an observational data set and determine if it is appropriate for propensity score matching to estimate average treatment effects. Conditional average treatment effects will be estimated with outcome regression approaches. Extensions to machine learning and big data techniques will be noted.

OBJECTIVE AND OVERVIEW

The course focuses on estimating models to guide management decisions to control a system or an organization. Statistical and machine learning methods are used to estimate regression and classification models. The distinction between correlation and causality is emphasized, and causality will be paramount. Specific topics will include randomized control trials, directed acyclic graphs (DAGs), inverse probability weighting, double robust estimation, propensity score modeling for observational studies, personalized marketing, and personalized medicine.

GRADING AND ASSIGNMENTS

Your course grade is determined by three homework assignments and a final paper. The course grade will be split as follows:

25%	Homework One
25%	Homework Two
25%	Homework Three
25%	Group Final Paper

Homework

On each assignment, I provide a data set. You will then answer a set of questions. I want you to use the R markdown as I do in class. I am okay with you talking, but each person is responsible for creating their answer to the homework. You should submit your R markdown code and a knitted file.

Group Assignment: The team will select the data. Teams of one or two students

- Please send me an email listing the group members by April 1.
- Submit a proposal for the final project by Saturday, April 15, 2024, noon EDT.
- The final paper will be due by 10:00 PM EST on Wednesday, May 3. The final paper should include a statement of the causal effect to be estimated. It should include an explanation of the system that generates the data. It should include a graphical model representation of the system's main features.

  Use appropriate techniques from this course to estimate the causal effect.

  Include an appropriate estimate of a forecasting model for the data set.

  Please email me a single zip file that contains your paper in a pdf file, the data in an excel file, and the R markdown that you used to prepare your paper. The R markdown program should run directly from the directory without path adjustments.

SCHEDULE FOR SPRING 2024

WEEK	DATES	TOPICS
1	March 11 & 13	Introduction to Causal Inference
2	March 18 & 20	Potential Outcome Framework and Randomize Control Trials
	MARCH 25	Homework One due by noon EDT
3	March 25 & 27	Estimating ATE with Outcome Regressions
	APRIL 1	Email Group names to Fallaw
4	April 1 & 3	Inverse Probability weights and double robust estimation
	APRIL 8	Homework Two due by noon EDT
5	April 8 & 10	Propensity Score Matching for observational data
	APRIL 15	Proposal for Final Project
6	April 15 & 17	Conditional Average Treatment Effects
	APRIL 22	Homework Three due by noon EDT
7	April 22	Instrumental variables, regression discontinuity, and differences in differences
	May 1	GROUP ASSIGNMENT: FINAL PAPER. Due by 10:00 PM EDT

COURSE SUMMARY

Week 1:

This is about motivation. The differences between prediction and control are highlighted. The main themes of the course will be noted with examples. We then demonstrate the empirical work that students perform in Homework One. We will analyze two data sets to determine the best model and the causal relationship of X and Y. We then introduce the Graphical Models representation of a system. Nodes and edges will be introduced to represent random variables and the relationships between them. The three basic structures between random variables will be introduced: the chain, the fork, and the collider.

Week 2:

We introduce the potential outcomes framework. Define individual treatment effects, the average treatment effect (ATE), and the conditional average treatment effect (CATE). The fundamental problem of causal inference is presented.

The wonders of Random Control Trials (RCTs) are introduced. The focus is on the covariate balance. We demonstrate this with random assignments and subject selection of the assignments.

Alternatives to simple random sampling are presented. Matched samples, stratified sampling, cluster sampling, systematic sampling. The advantages of these different approaches are presented.

Week 3:

The average effects of treatment are estimated with linear regression using the outcome variable. The need to carefully control the covariate is emphasized. The noncausal paths in the system are blocked so that only the causal effect is measured. The assumptions needed for accurate conclusions are noted and examples demonstrate when the assumptions hold and fail.

The estimation of ATE is presented when the data are from an RCT. The ATE is estimated using the backdoor adjustment when all the confounders are observed. ATE is estimated with front door adjustment when an appropriate moderator is observed.

Week 4:

Inverse probability weights are used to estimate average treatment effects. The pseudosamples are created and used.

Double robust estimation is presented. This is a combination of the regression approach and the inverse probability weighting.

Week 5:

When we have an observational study, how close can we get to an RCT? The basic idea is that we might be able to create a sub-sample of our data that mimics a randomized control trial. It is not always possible, but sometime causal effects can be estimated with observation data. The intuition and steps to perform the ATE estimation with propensity scores are presented. The use of machine learning models as an intermediate step is noted.

Week 6:

We consider different approaches to estimating conditional average treatment effects. This is the foundation for personalized medicine and personalized marketing. Almost all of these approaches use machine learning tools. We will focus on the outcome regression, but more modern approaches will be noted.

Week 7:

This week we consider special situations that allow for the estimation of causal effects. These are not as general as the previous situations but do occur enough to be important to understand: instrumental variables, regression discontinuity designs, and differences in differences (longitudinal data sets).

Causal Inference Course Mini 4 2024

This course is designed for the Tepper MBA program. There is no textbook required because no single book presents the material to complement and support your Tepper MBA. There is a very large literature on causal inference from several different disciplines: e.g., economics, computer science, biostatistics, and political science. Here is a list of different books that I think you might find useful.

References

Angrist, J., & Pischke, J. (2014). *Mastering 'metrics: The path from cause to effect*. Princeton University Press. Retrieved from https://books.google.com/books?id=s2eYDwAAQBAJ

Angrist, J., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion* (1st ed.). Princeton University Press. Retrieved from https://EconPapers.repec.org/RePEc:pup:pbooks:8769

Bai, H., & Clark, M. (2018). *Propensity score methods and applications*. SAGE Publications. Retrieved from https://books.google.com/books?id=7BZoDwAAQBAJ

Cunningham, S. (2021). *Causal inference*. Yale University Press. Retrieved from https://books.google.com/books?id=DZ4REAAAQBAJ

Gerber, A., & Green, D. (2012). *Field experiments: Design, analysis, and interpretation*. W. W. Norton. Retrieved from https://books.google.com/books?id=yxEGywAACAAJ

Guo, S., & Fraser, M. (2015). *Propensity score analysis*. SAGE Publications. Retrieved from https://books.google.com/books?id=ANY5DQAAQBAJ

Hernan, M., & Robins, J. (2023). *Causal inference*. CRC Press. Retrieved from https://books.google.com/books?id=_KnHIAAACAAJ

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press. doi: doi:10.1017/CBO9781139025751

Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press. Retrieved from https://books.google.com/books?id=TFjPDwAAQBAJ

Lawson, J. (2014). *Design and analysis of experiments with r*. CRC Press. Retrieved from https://books.google.com/books?id=TOxMBgAAQBAJ

Leite, W. (2016). *Practical propensity score methods using r*. SAGE Publications. Retrieved from https://books.google.com/books?id=4wecDQAAQBAJ

Luca, M., & Bazerman, M. (2021). *The power of experiments: Decision making in a data-driven world*. MIT Press. Retrieved from https://books.google.com/books?id=XDsZEAAAQBAJ

MA, M. H., & Robins, J. (2019). *Causal inference*. Chapman & Hall/CRC.

Morgan, S., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press. Retrieved from https://books.google.com/books?id=Q6YaBQAAQBAJ

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). New York, NY, USA: Cambridge University Press.

Pearl, J., Glymour, M., & Jewell, N. (2016). *Causal inference in statistics: A primer*. Wiley. Retrieved from https://books.google.com/books?id=IqCECwAAQBAJ

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Penguin Books Limited. Retrieved from https://books.google.com/books?id=EmY8DwAAQBAJ

Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference - foundations and learning algorithms*. Cambridge, MA, USA: The MIT Press.

Rosenbaum, P. (2017). *Observation and experiment: An introduction to causal inference*. Harvard University Press. Retrieved from https://books.google.com/books?id=KtcuDwAAQBAJ

VanderWeele, T. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press. Retrieved from https://books.google.com/books?id=K6cgBgAAQBAJ