Causal Inference Working Group

A Practical Introduction to Bayesian Causal Inference

Adam Peterson

University of Michigan

Table of contents

Conceptual Introduction

Motivation

The study of Causal Inference is usually motivated by a question starting with *What if* [2]

Motivation

The study of Causal Inference is usually motivated by a question starting with *What if* [2]:

- *What if* you had been given a vaccine, would you still have gotten sick?
- *What if* you had majored in English, would you still have found Biostatistics?

Samples and Populations

While we can never hope to get a concrete answer to these questions for individuals we *can* get close to meaningful answers for populations

Samples and Populations

While we can never hope to get a concrete answer to these questions for individuals we *can* get close to meaningful answers for populations:

Average Treatment Effect

$$ATE = E[Y^1] - E[Y^0]$$

Samples and Populations

While we can never hope to get a concrete answer to these questions for individuals we *can* get close to meaningful answers for populations:

Average Treatment Effect

$$ATE = E[Y^1] - E[Y^0]$$

... *under certain conditions*

Identifiability Conditions

1. (Conditional) Exchangeability
    - Also called ignorability
2. Consistency
3. No Interference
    - Also called the stable treatment unit value assumption (STUVA)
4. Positivity

Identifiability Conditions

1. (Conditional) Exchangeability
    - Also called ignorability
2. Consistency
3. No Interference
    - Also called the stable treatment unit value assumption (STUVA)
4. Positivity

If assumptions hold then:

Average Treatment Effect

$$ATE = E[Y^1] - E[Y^0] = E[Y|Z = 1] - E[Y|Z = 0]$$

Identifiability Conditions

Q: How do we know if these conditions are satisfied?

Identifiability Conditions

Q: How do we know if these conditions are satisfied?

A: You don't *know* in the sense that you're 100% sure.

Identifiability Conditions

Q: How do we know if these conditions are satisfied?

A: You don't *know* in the sense that you're 100% sure.

⋮

Certain types of experiments or "experimental settings" will be more conducive towards having these conditions satisfied than others

Types of Experiments

- Randomized Control Trials (Treatment assignment is controlled)
  - Conditions *should* be satisfied by trial design
- Observational Data (Treatment assignment is not controlled)
  - Exchangeability: Have you measured all the right confounders?
  - Positivity: Is there sufficient variability in the exposure of interest for you to be able to detect an effect?
  - Consistency: Is the exposure well defined across all observations?
  - Interference: Does treatment assigned to one unit affect another's potential outcome?

Just the start

- Causal Inference is a **Huge** field of study and there is so much more to learn here than what I've gone over.
- We'll now go over a case-study to illustrate some basic concepts using regression to estimate ATE's.
- There are references at the end of this slide deck for those who want to learn further. [1, 3]

Case Study : The Electric Company

Backstory

- In 1970, a set of 192 elementary school classes were randomly assigned to either watch a new educational TV show or not, to see whether the show improved kids' reading ability.
- At the beginning and end of the school year students in all the classes were given a reading test, and the average test score *within each class* was recorded.

Backstory

- In 1970, a set of 192 elementary school classes were randomly assigned to either watch a new educational TV show or not, to see whether the show improved kids' reading ability.
- At the beginning and end of the school year students in all the classes were given a reading test, and the average test score *within each class* was recorded.

What is the question we can answer with these data?

Backstory

- In 1970, a set of 192 elementary school classes were randomly assigned to either watch a new educational TV show or not, to see whether the show improved kids' reading ability.
- At the beginning and end of the school year students in all the classes were given a reading test, and the average test score *within each class* was recorded.

What is the question we can answer with these data?

This is an ecological analysis. Caution is warranted.

Tutorial Time

repository link

References i


A. Gelman, J. Hill, and A. Vehtari.
Regression and other stories.
Cambridge University Press, 2020.


M. A. Hernán and J. M. Robins.
Causal inference, 2010.


A. Oganisian and J. A. Roy.
A practical introduction to bayesian estimation of causal effects: Parametric and nonparametric approaches.
arXiv preprint arXiv:2004.07375, 2020.