Segment 1: Fundamentals of Causal Inference

Section 01: What is Causal Inference?

What is causal inference and how do we mathematically formalize "cause" and "effect"?

Data Science

- ▶ (Most) common goal in statistics and data science: **Prediction**
  - ▶ E.g., regression:

Data Science

- ▶ (Most) common goal in statistics and data science: **Prediction**
  - ▶ E.g., regression:

- ▶ (Another) common goal: **Decision Making**
  - ▶ If I take an action, what will be the result?
  - ▶ Questions of **causal inference**.

Data Science

- ▶ (Most) common goal in statistics and data science: **Prediction**
  - ▶ E.g., regression:

- ▶ (Another) common goal: **Decision Making**
  - ▶ If I take an action, what will be the result?
  - ▶ Questions of **causal inference**.
- ▶ Can prediction models/algorithms tell us about impacts of decisions?
- ▶ That is, can they tell us about underlying causal relationships in the data?

Data Science

- ▶ (Most) common goal in statistics and data science: **Prediction**
    - ▶ E.g., regression:

- ▶ (Another) common goal: **Decision Making**
    - ▶ If I take an action, what will be the result?
    - ▶ Questions of **causal inference**.
- ▶ Can prediction models/algorithms tell us about impacts of decisions?
- ▶ That is, can they tell us about underlying causal relationships in the data?
    - ▶ Maybe...but there's more work to do

Correlation is Not Causation!

So what is causation?

Correlation is Not Causation!

So what is causation?

- ▶ Sounds very intuitive...
- ▶ Technical definition can be elusive

Correlation is Not Causation!

So what is causation?

- ▶ Sounds very intuitive...
- ▶ Technical definition can be elusive

So what is "causal inference?"

- ▶ The **causal inference methodology** in this class will give:
  - ▶ One way of formalizing what is meant by "causal effect" in data science
  - ▶ A set of perspectives/tools/methods for data-based decision making that:
    1. Frame questions about cause and effect to be answered with data
    2. Analyze data in a way specifically designed to estimate the consequences of actions or decisions
    3. Inform the design of prospective studies to help uncover causal effects

Different Meanings of "Causation"

Optional Reading: Holland (1986) Statistics and Causal Inference, *Journal of the American Statistical Association* 81(396) pp. 945-960

1. Ultimate meaningfulness of causation
   - Aristotle!

Different Meanings of "Causation"

1. Ultimate meaningfulness of causation
   - ▶ Aristotle!
2. Deducing the causes of a given effect
   - ▶ Which lifestyle factors caused her to have this disease?

Different Meanings of "Causation"

1. Ultimate meaningfulness of causation
   - Aristotle!
2. Deducing the causes of a given effect
   - Which lifestyle factors caused her to have this disease?
3. Understanding details of causal mechanisms
   - What is the biological pathway through which inhaled air pollution causes cardiovascular disease?

Different Meanings of "Causation"

1. Ultimate meaningfulness of causation
   - ▶ Aristotle!
2. Deducing the causes of a given effect
   - ▶ Which lifestyle factors caused her to have this disease?
3. Understanding details of causal mechanisms
   - ▶ What is the biological pathway through which inhaled air pollution causes cardiovascular disease?
4. Measuring the effects of causes
   - ▶ What is the effect of providing a customer with targeted advertising?
   - ▶ Statistics/data science has the most to say about this one

Examples

- The aspirin I took caused my headache to go away.
- I did not get the flu this year because I was vaccinated.
- Her colon cancer got diagnosed too late because she is black.
- His diabetes caused him to have a heart attack.
- Turning the car's steering wheel caused the car to turn.
- Does he have lung cancer because he smokes?
- Will studying for the test cause a high score?
- Is there a causal link between air pollution and cardiovascular disease?

Example

Three statements, all using different meanings of "because" to explain the same "effect:"

(A) She did well on the exam because she was a woman.
  - ▶ "Cause" as an attribute she possesses

(B) She did well on the test because she studied.
  - ▶ "Cause" is a voluntary activity that was performed

(C) She did well on the test because she was coached.
  - ▶ "Cause" is an activity that was imposed

Key Point

Causal inference methods in this course will:

- ▶ Define explicit actions/states to determine "the cause" of interest
- ▶ Define causal effects as comparisons between outcomes under competing states
- ▶ Discuss ways to use **data** to understand what would have happened under some competing state using:
    1. Thoughtful study design (either prospective or retrospective)
    2. Statistical methods and/or models

    (The better we are at 1, the less 2 matters)

Correlation is Not Causation!

Then what is "causal inference"?

Correlation is Not Causation!

Then what is "causal inference"?

A general analytic perspective to:

- ▶ Frame questions about cause and effect to be answered with data
- ▶ Analyze data in a way specifically designed to estimate the consequences of actions or decisions
- ▶ Clarify common threats to the validity of analyses that (implicitly or explicitly) aim to characterize causal relationships
- ▶ Increase transparency about the assumptions required to estimate causal effects with data

Some Data Science Frameworks and Perspectives

1. **Potential Outcomes Framework**, AKA, Rubin Causal Model, counterfactual framework, Neyman-Rubin Causal Model
2. Causal diagram framework
3. Structural equation modeling
4. Econometrics, "Quasi Experiments"

These are related frameworks. One **goal** of this course is to give the fundamental tools to navigate, understand, and apply tools from this literature.