Introduction to Causal Inference

Hans Jarett Ong   Paula Tse Wing

September 8, 2022

Contents

1. Why Statistics is not Enough
2. Causal Inference
3. Learning Objectives
4. Causal Inference in Real Life
5. References

Why Statistics is not Enough

"Correlation is not Causation"

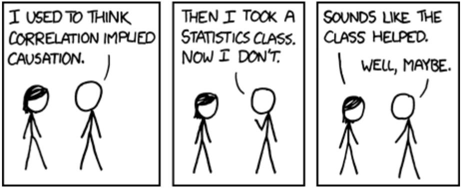

I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Image Source: XKCD

- It is difficult to be confident in our results with mere correlations.
- We lack "skin in the game" – if all goes wrong, we can just say "correlation is not causation".

"Correlation is not Causation"

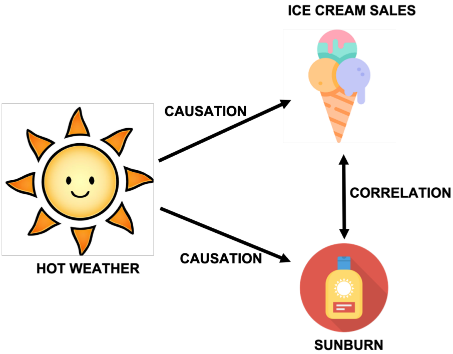


Image Source: Roy Riachi

Data cannot "speak for itself".

	Drug	No drug
Men	81 of 87 recovered (93%)	234 of 270 recovered (87%)
Women	192 of 263 recovered (73%)	55 of 80 recovered (69%)
Combined data	273 of 350 recovered (78%)	289 of 350 recovered (83%)

Example 1: Taken from *Pearl 2016 - Causal Inference in Statistics: A Primer*

- Is this drug helpful or harmful?
- How would you present this finding?
- Which is the correct interpretation?

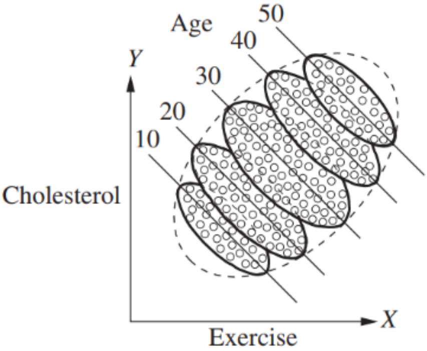
Data cannot "speak for itself".

	No drug	Drug
Low BP	81 of 87 recovered (93%)	234 of 270 recovered (87%)
High BP	192 of 263 recovered (73%)	55 of 80 recovered (69%)
Combined data	273 of 350 recovered (78%)	289 of 350 recovered (83%)

Example 2: Taken from *Pearl 2016 - Causal Inference in Statistics: A Primer*

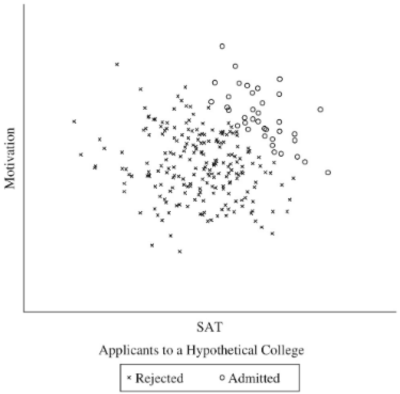
- Is this drug helpful or harmful?
- How would you present this finding?
- Which is the correct interpretation?

Data cannot "speak for itself".



Example 3: Taken from *Pearl 2016 - Causal Inference in Statistics: A Primer*

Data cannot "speak for itself".



Example 4: Taken from *Morgan and Winship - Counterfactuals and Causal Inference*

Note: "Motivation" refers to motivation during the interview.

Causal Inference

Two Kinds of Studies/ Data

Experimental Data:

- The investigator has complete or partial control of the system.
- The treatment or exposure is assigned by the investigator.
- e.g. A/B testing, Randomized Controlled Trials (RCTs)

Observational Data:

- The investigator is only a passive observer and has no control of the system.
- The treatment or exposure is not assigned by the investigator.
- This comprises most of the data we encounter as data scientists (we rarely conduct our own experiments).

Two Kinds of Studies/ Data

While experiments are still the best way to infer causality, some experiments are simply impossible to conduct. Consider these queries:

- Does smoking cause lung cancer?
- What is the effect of education on salaries?
- What is the effect of CO_2 emissions on global climate?

Possible reasons for using causal inference:

- Risk - when the conducting the experiment is risky or unethical
- Costs - running experiments may be too costly and take too long to conduct
- Data Collection - some kinds of data are hard to collect, e.g. long-term studies with lots of follow-up
- Past - you can't conduct experiments on historical data or on events that already happened
- Inform next experiment - identify biases and which experiments to prioritize

Causal Inference as the "Science" part of Data Science

- Causal Inference is not a single method or some black-box technique that magically gives us causality.
- It requires *domain knowledge*. As seen earlier, data cannot be divorced from its context (i.e. the process that generated the data).
- It provides tools and a language for us to formally and clearly state our assumptions or hypotheses (theories).
- It goes beyond pattern recognition and aims to uncover actual *scientific knowledge*.

Doing Science is not Easy

"In so far as a scientific statement speaks about reality, it must be falsifiable: and in so far as it is not falsifiable, it does not speak about reality." – Karl Popper

- Popper needed a criterion to separate science from non-science (pseudoscience) in response to the 2 famous theories of his time: Einstein's Theory of Relativity and Freud's Psychoanalysis.
- In short, you can never be right! You can only be wrong.
- Scientific theories need to have "skin in the game".

Going beyond Patterns

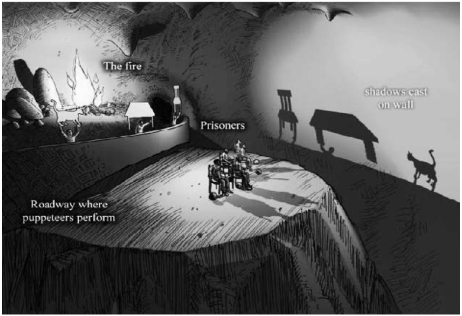


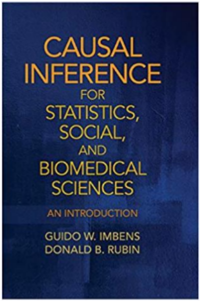
Image Source: Markus Maurer

Aims of Causal Inference:

- Shadows → Reality
- Patterns and Correlations → Causal/Scientific Knowledge
- Data → Data Generating Process

Learning Objectives

Different Schools of Thought

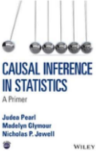


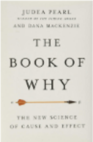
Potential Outcome Framework

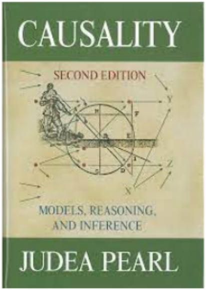
ANALYTICAL METHODS FOR SOCIAL RESEARCH
Counterfactuals and Causal Inference
Methods and Principles for Social Research
SECOND EDITION
STEPHEN L. MORGAN
CHRISTOPHER WINSHIP

Causal Inference for The Brave and True

[Causal Inference for The Brave and True (matheusfacure.github.io)](matheusfacure.github.io)







Causal Graphs Framework

- Although they have different starting points, these 2 are already proven to be equivalent.
- Many textbooks (in the middle) introduce causal inference using a mix of the 2 frameworks, taking advantage of the strengths of each one.

Methods used in Causal Inference

Simple Methods

- Matching
- Stratification
- Propensity Matching
- Inverse Probability Weighting
- Wald Estimator
- Instrumental Variables
- Doubly Robust Estimators

Advanced Methods

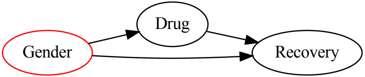
- Causal Discovery
- Double ML (Chernozhukov et al. 2016)
- Orthoforests
- T-learner
- X-learner (Kunzel et al. 2017)
- Intent to Treat DRIV
- ... and so much more!

- Causal Inference is still a growing field with a lot of ongoing research.
- In this series, we'll focus on the basics (the simple methods).

Example of Causal Reasoning

	Drug	No drug
Men	81 of 87 recovered (93%)	234 of 270 recovered (87%)
Women	192 of 263 recovered (73%)	55 of 80 recovered (69%)
Combined data	273 of 350 recovered (78%)	289 of 350 recovered (83%)

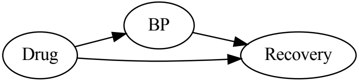
Example 1: Taken from *Pearl 2016 - Causal Inference in Statistics: A Primer*



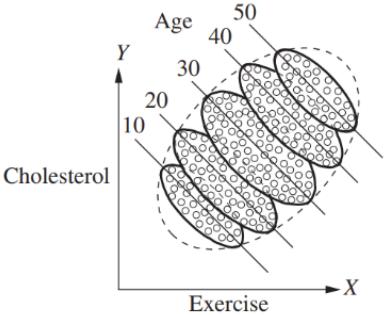
Example of Causal Reasoning

	No drug	Drug
Low BP	81 of 87 recovered (93%)	234 of 270 recovered (87%)
High BP	192 of 263 recovered (73%)	55 of 80 recovered (69%)
Combined data	273 of 350 recovered (78%)	289 of 350 recovered (83%)

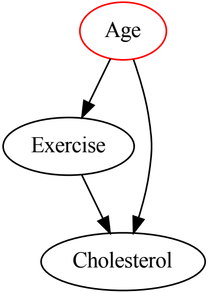
Example 2: Taken from *Pearl 2016 - Causal Inference in Statistics: A Primer*



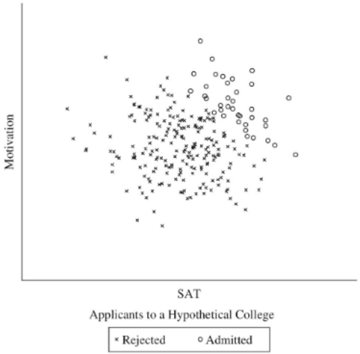
Example of Causal Reasoning



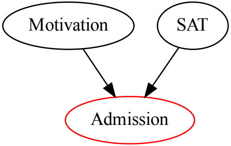
Example 3: Taken from *Pearl 2016 - Causal Inference in Statistics: A Primer*



Example of Causal Reasoning



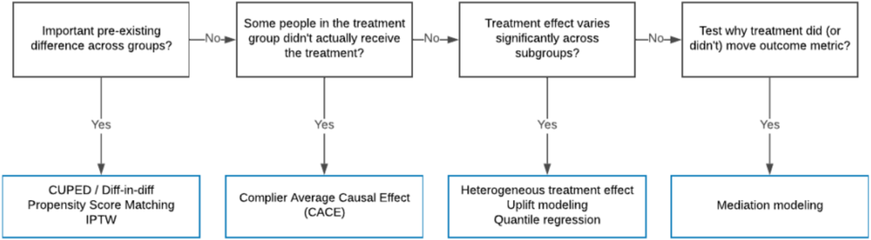
Example 4: Taken from *Morgan and Winship - Counterfactuals and Causal Inference*



Causal Inference in Real Life

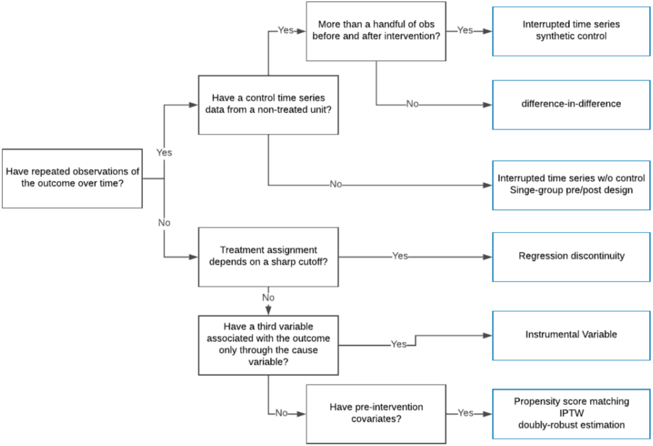
Casual Inference at Uber Labs

Uber uses Causal Inference with both *experimental data* (to guide experiments and to make post-experiment adjustments) and *observational data*.


Important pre-existing difference across groups?
No
Some people in the treatment group didn't actually receive the treatment?
No
Treatment effect varies significantly across subgroups?
No
Test why treatment did (or didn't) move outcome metric?
Yes
Yes
Yes
Yes
CUPED / Diff-in-diff
Propensity Score Matching
IPTW
Complier Average Causal Effect (CACE)
Heterogeneous treatment effect
Uplift modeling
Quantile regression
Mediation modeling

Causal Inference methods for experimental data used at Uber Labs.

Casual Inference at Uber Labs



Have repeated observations of the outcome over time?

Yes

Have a control time series data from a non-treated unit?

Yes

More than a handful of obs before and after intervention?

Yes

Interrupted time series
synthetic control

No

difference-in-difference

No

Interrupted time series w/o control
Singe-group pre/post design

No

Treatment assignment depends on a sharp cutoff?

Yes

Regression discontinuity

No

Have a third variable associated with the outcome only through the cause variable?

Yes

Instrumental Variable

No

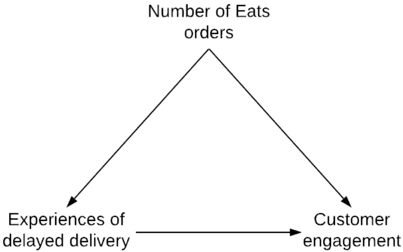
Have pre-intervention covariates?

Yes

Propensity score matching
IPTW
doubly-robust estimation

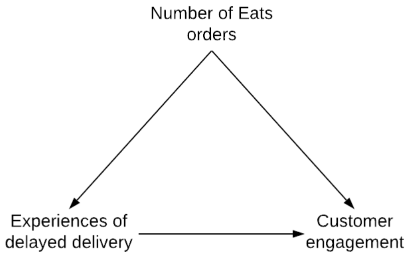
Causal Inference methods for observational data used at Uber Labs.

Causal Inference at Uber Labs



- Uber relies observational data when experiments are implausible. For example, they might want to know the effect of delays in food delivery on customer engagement (they wouldn't want to intentionally delay some of the orders).

Causal Inference at Uber Labs



- Naively comparing *experiences of delayed delivery* and *customer engagement* might lead us to conclude that delays *increase* customer engagement.
- But *number of Eats orders* is a confounder because those who order more are also more likely to experience delays. So, they must control for this in order to get the actual effect.

ALICE project by Microsoft

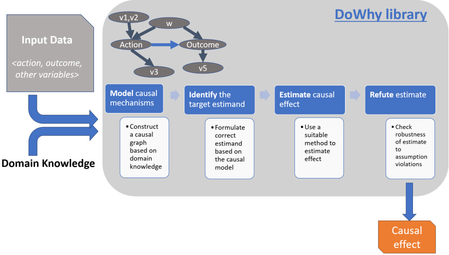
ALICE or Automated Learning and Intelligence for Causation and Economics aims to go from predictions to decision-making.

- For decision-making, we need to find the features that **cause the outcome** and *to estimate how the outcome would change if the features are changed.*
- We cannot infer the effect of interventions from correlations.
- The important features in a prediction model don't necessarily affect the outcome directly. Plus, these features may not be the best features to act on.

ALICE project by Microsoft

ALICE project combines **DoWhy** and **EconML** – packages developed by Microsoft for causal inference.

- DoWhy - a tool for encoding domain knowledge using causal graphs and making inferences and estimations using a combination of said domain knowledge and observational data.


DoWhy library
Input Data
<action, outcome, other variables>
Domain Knowledge
v1,v2
w
Action
Outcome
v3
v5
Model causal mechanisms
• Construct a causal graph based on domain knowledge
Identify the target estimand
• Formulate correct estimand based on the causal model
Estimate causal effect
• Use a suitable method to estimate effect
Refute estimate
• Check robustness of estimate to assumption violations
Causal effect

Schematic of Microsoft's DoWhy Library.

- EconML - a tool that handles high-dimensional data where the causal effect varies over different sub-populations.

ALICE project by Microsoft

Many data science questions are actually causal questions. Here are some sample questions ALICE aims to answer:

- **A/B experiments:** If I change the algorithm, will it lead to a higher success rate?
- **Policy decisions:** If we adopt this treatment/policy, will it lead to a healthier patient/more revenue/etc.?
- **Policy evaluation:** knowing what I know now, did my policy help or hurt?
- **Credit attribution:** are people buying because of the recommendation algorithm? Would they have bought anyway?

Nobel Prize in Economics 2021

The prize was awarded jointly to:

- David Card
    - *"for his empirical contributions to labour economics"*
    - used **natural experiments** to determine the labour market effects of *minimum wages, immigration, and education* and showed that
        - increasing minimum wage doesn't necessarily lead to fewer jobs
        - incomes of locals in a country can benefit from new immigration
        - resources in schools are far more important for student's future labour market success than was previously thought
- Joshua D. Angrist and Guido W. Imbens
    - *"for their methodological contributions to the analysis of causal relationships"*
    - they developed the natural experiments methodology and demonstrated how precise conclusions about cause and effect can be drawn from natural experiments.

References

References

1. Hernan, Miquel A., and James M. Robins. Causal Inference. CRC Press, 2019.
2. Roy, Jason. “A Crash Course in Causality: Inferring Causal Effects from Observational Data — Coursera.” Coursera, https://www.coursera.org/learn/crash-course-in-causality. Accessed 15 Aug. 2022.
3. Matheus, Facure. “Causal Inference for The Brave and True.” Matheus Facure, https://matheusfacure.github.io/python-causality-handbook/landing-page.html. Accessed 15 Aug. 2022.
4. Morgan, Stephen L., and Christopher Winship. Counterfactuals and Causal Inference. Cambridge University Press, 2014.
5. Pearl, Judea, et al. Causal Inference in Statistics. John Wiley & Sons, 2016.
6. Yao, Liuyi, et al. “A Survey on Causal Inference.” ACM Transactions on Knowledge Discovery from Data, no. 5, Association for Computing Machinery (ACM), Oct. 2021, pp. 1–46. Crossref, doi:10.1145/3444944.