# Estimation by Prediction

UCLA SOCIOL 212B
Winter 2025

26 Feb 2025

# Learning goals for today

By the end of class, you will be able to use an **outcome model** in the service of

- describing a population from a non-probability sample
- inferring average causal effects in an observational study

# Predicting outcomes for population inference from non-probability samples

Survey on Amazon Mechanical Turk

1. What is your sex?
2. What is your age?
3. Have you ever had a TikTok account?

Why might the sample mean of $Y$ be a poor estimator of the mean in the full U.S. population?

# A possible (but heroic) assumption

Conditionally exchangeable sampling:

$$\underbrace{S}_{\text{Sampling}} \perp\!\!\!\perp \underbrace{Y}_{\text{Ever had TikTok}} \mid \underbrace{\vec{X}}_{\text{Sex, Age}}$$

Equivalently: $\text{P}(Y \mid S = 1, \vec{X} = \vec{x}) = \text{P}(Y \mid \vec{X} = \vec{x})$

Then sample data on 52-year-old men is informative about the population mean among 52-year-old men

# A possible (but heroic) assumption

Conditionally exchangeable sampling:

$$\underbrace{S}_{\text{Sampling}} \perp\!\!\!\perp \underbrace{Y}_{\text{Ever had TikTok}} \mid \underbrace{\vec{X}}_{\text{Sex, Age}}$$

Equivalently: $\text{P}(Y \mid S = 1, \vec{X} = \vec{x}) = \text{P}(Y \mid \vec{X} = \vec{x})$

Then sample data on 52-year-old men is informative about the population mean among 52-year-old men

**Problems:**

- Many age $\times$ sex subgroups! Small sample.
- How to aggregate the subgroup estimates?

# Step 1: Model the conditional probabilities

$$\text{logit}\left(\hat{\mathsf{P}}(Y = 1 \mid S = 1, \vec{X})\right) = \hat{\beta}_0 + \hat{\beta}_1(\text{Sex} = \text{Female}) + \hat{\beta}_2(\text{Age})$$

# Step 2: Define $\vec{X}$ distribution to aggregate

U.S. Census has estimates of the age $\times$ sex distribution:
$\text{P}(\vec{X} = \vec{x})$ is known

# Overall strategy

$$\hat{\mathrm{P}}(Y) = \sum_{\vec{x}} \hat{\mathrm{P}}(Y \mid \vec{X} = \vec{x})\mathrm{P}(\vec{X} = \vec{x})$$

- measure $\vec{X}$ and $Y$ in a non-probability sample
- measure $\vec{X}$ in a probability sample or census
- assume exchangeable sampling given $\vec{X}$ (heroic!)
- model $\mathrm{E}(Y \mid \vec{X})$ or $\mathrm{P}(Y \mid \vec{X})$ in the non-probability sample
- estimate $\mathrm{P}(\vec{X} = \vec{x})$ in the probability sample or census
- re-aggregate $\hat{\mathrm{E}}(Y \mid \vec{X})$ using the weights $\hat{\mathrm{P}}(\vec{X} = \vec{x})$

# Real example: Xbox survey

Wang et al. 2015

Survey in 2012:
If the election were held today, who would you vote for?
(Obama or Romney)

# Real example: Xbox survey

Wang et al. 2015

Survey in 2012:
If the election were held today, who would you vote for?
(Obama or Romney)

Survey conducted on Xbox. Disproportionately young men.

# Real example: Xbox survey

Wang et al. 2015

Survey in 2012:
If the election were held today, who would you vote for?
(Obama or Romney)

Survey conducted on Xbox. Disproportionately young men.
But over 700,000 responses!

# Real example: Xbox survey

Wang et al. 2015

Measured $\vec{X}$ in hopes of conditional exchangeability:
sex, race, age, education, state, party ID, political ideology, and who they voted for in the 2008 presidential election.

Hope for: $Y \perp\!\!\!\perp S \mid \vec{X}$

# Real example: Xbox survey

Wang et al. 2015

Step 1: Estimate conditional means

$$\hat{\text{P}}(Y = 1 \mid S = 1, \vec{X} = \vec{x}) = \text{logit}^{-1}(\text{complicated function of } \vec{x})$$

Step 2: Estimate $\text{P}(\vec{X} = \vec{x})$ using 2008 exit polls

Step 3: Aggregate predictions from (1) weighted by (2):

$$\hat{\text{P}}(Y = 1) = \sum_{\vec{x}} \underbrace{\hat{\text{P}}(\vec{X} = \vec{x})}_{\substack{\text{Stratum size,} \\ \text{estimated from} \\ \text{2008 exit polls}}} \underbrace{\hat{\text{P}}(Y = 1 \mid S = 1, \vec{X} = \vec{x})}_{\substack{\text{Prediction within the stratum,} \\ \text{estimated from Xbox survey}}}$$

# Real example: Xbox survey

Wang et al. 2015

Step 1: Estimate conditional means

$$\hat{\text{P}}(Y = 1 \mid S = 1, \vec{X} = \vec{x}) = \text{logit}^{-1}(\text{complicated function of } \vec{x})$$

Step 2: Estimate $\text{P}(\vec{X} = \vec{x})$ using 2008 exit polls

Step 3: Aggregate predictions from (1) weighted by (2):

$$\hat{\text{P}}(Y = 1) = \sum_{\vec{x}} \underbrace{\hat{\text{P}}(\vec{X} = \vec{x})}_{\substack{\text{Stratum size,} \\ \text{estimated from} \\ \text{2008 exit polls}}} \underbrace{\hat{\text{P}}(Y = 1 \mid S = 1, \vec{X} = \vec{x})}_{\substack{\text{Prediction within the stratum,} \\ \text{estimated from Xbox survey}}}$$

**They did surprisingly well!**

# Takeaways: Prediction to describe

predictive outcome models $\hat{f}:\vec{X}\rightarrow\hat{Y}$
learned in non-probability samples

$+$

known distribution of $\vec{X}$
(e.g., from Census)

$+$

conditional exchangeability
$Y \perp\!\!\!\perp S \mid \vec{X}$

$=$ powerful tool for population inference

# Predicting outcomes for causal claims: An example to work by hand

# An example by hand

Suppose $\{Y^0, Y^1\} \perp\!\!\!\perp A \mid X$.
A researcher estimates a model:

$$\hat{\mathrm{E}}(Y \mid \vec{X}, A) = \hat{\beta}_{\text{Intercept}} + \hat{\beta}_X X + \hat{\beta}_A A + \hat{\beta}_{XA} XA$$

with estimates $\hat{\beta}_{\text{Intercept}} = 0$, $\hat{\beta}_X = 1$, $\hat{\beta}_A = 2$, $\hat{\beta}_{XA} = 1$.

**Task.** Fill in the table. Estimate the average causal effect.

| ID | $X$ | $\hat{Y}^1$ | $\hat{Y}^0$ | $\hat{Y}^1 - \hat{Y}^0$ |
|---|---|---|---|---|
| 1 | 0 | ? | ? | ? |
| 2 | 1 | ? | ? | ? |
| 3 | 1 | ? | ? | ? |
| 4 | 1 | ? | ? | ? |

# Prediction for Causal Inference: Visual Example

# Outcome model: A visual example

I feel confident that I can answer quantitative questions with tools from data science.

- $1 =$ Agree
- $0 =$ Disagree

# Outcome model: A visual example

I feel confident that I can answer quantitative questions with tools from data science.

- 1 = Agree
- 0 = Disagree

What is the average causal effect of taking this class on confidence in data science skills?

# Outcome model: A visual example

Each Row is a Student in This Class

| | |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $Y_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $Y_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $Y_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $Y_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $Y_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $Y_6^{\text{No 212b}}$ |
| Outcome under 212b | Outcome under no 212b |

$Y =$ I feel confident that I can answer quantitative questions with tools from data science

# Outcome model: A visual example

$Y$ = I feel confident that I can answer quantitative questions with tools from data science

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | ? |
| $Y_2^{\text{Takes 212b}}$ | ? |
| $Y_3^{\text{Takes 212b}}$ | ? |
| $Y_4^{\text{Takes 212b}}$ | ? |
| $Y_5^{\text{Takes 212b}}$ | ? |
| $Y_6^{\text{Takes 212b}}$ | ? |

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | ? |
| $Y_2^{\text{Takes 212b}}$ | ? |
| $Y_3^{\text{Takes 212b}}$ | ? |
| $Y_4^{\text{Takes 212b}}$ | ? |
| $Y_5^{\text{Takes 212b}}$ | ? |
| $Y_6^{\text{Takes 212b}}$ | ? |

$Y =$ I feel confident that I can answer quantitative questions with tools from data science

How could we learn about the (?)

# Outcome model: A visual example

# Outcome model: A visual example

For each of you, we could compare

1. your opinion after 212b
2. the average opinion of non-212b students who look like you

# Outcome model: A visual example

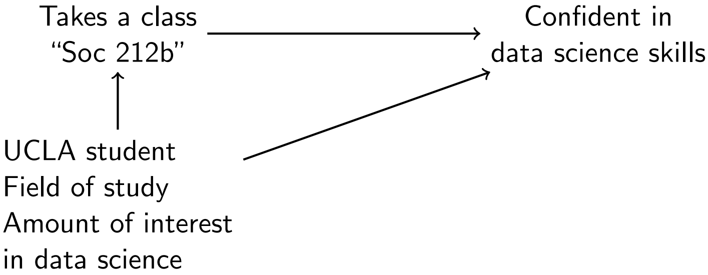
For each of you, we could compare

1. your opinion after 212b
2. the average opinion of non-212b students who look like you

Looks like you in what ways? What else belongs in this DAG?

Takes a class "Soc 212b" ⟶ Confident in data science skills

# Outcome model: A visual example




Suppose these are a sufficient adjustment set.

# Outcome model: A visual example

**Nonparametric estimation:**

For each student in the class, find someone else who

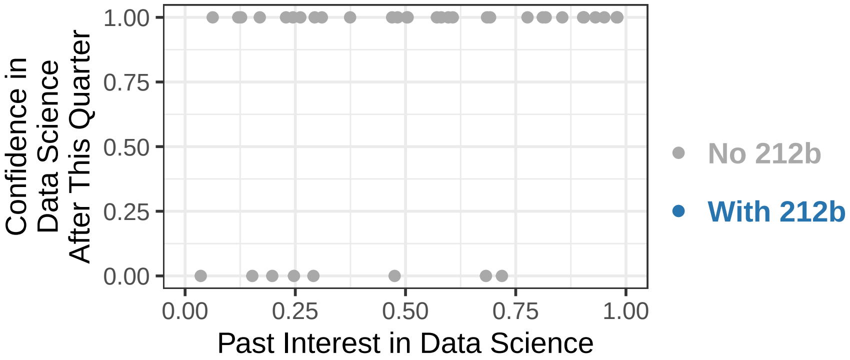
- is a student at UCLA
- shares your field of study
- is exactly as interested in data science as you are
- but did not take this class

Use your **match** to infer your $Y_i^{\text{No 212b}}$ for people like you:

$$E(Y^0 \mid \vec{X} = \vec{x}_i) = \underbrace{E(Y \mid A = 0, \vec{X} = \vec{x}_i)}_{\text{estimated from your match}}$$

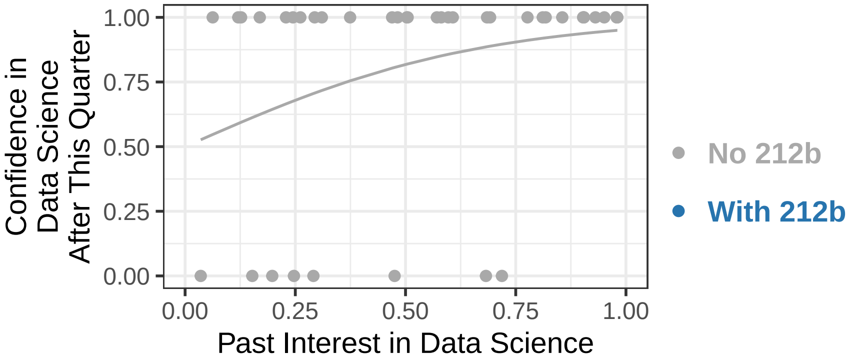
since we have assumed conditional exchangeability given $\vec{X}$.

# Outcome model: A visual example



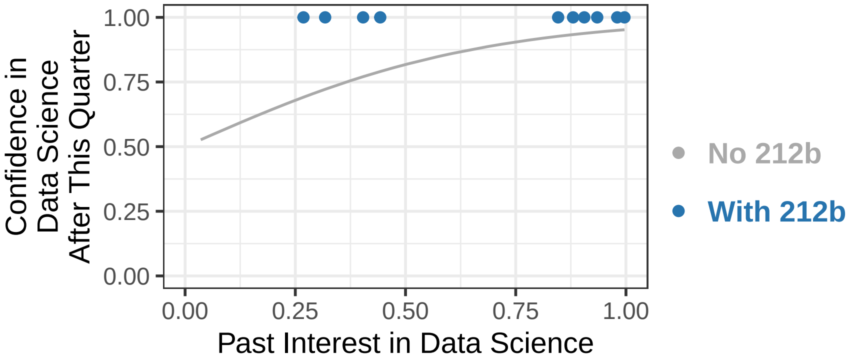

1) Find control units who didn't take this class

# Outcome model: A visual example



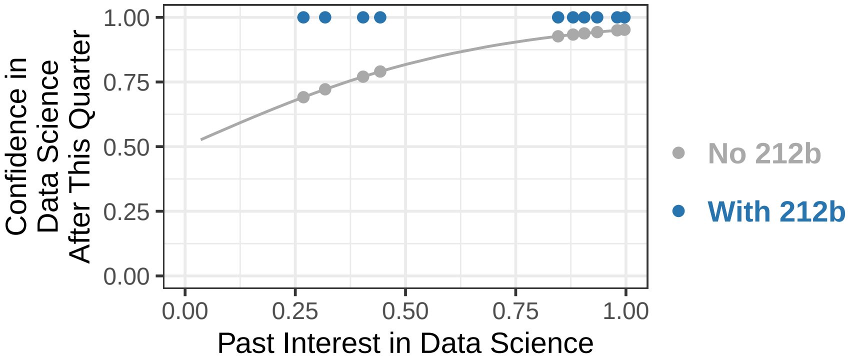

2) Model their outcomes given pre-treatment variables

# Outcome model: A visual example



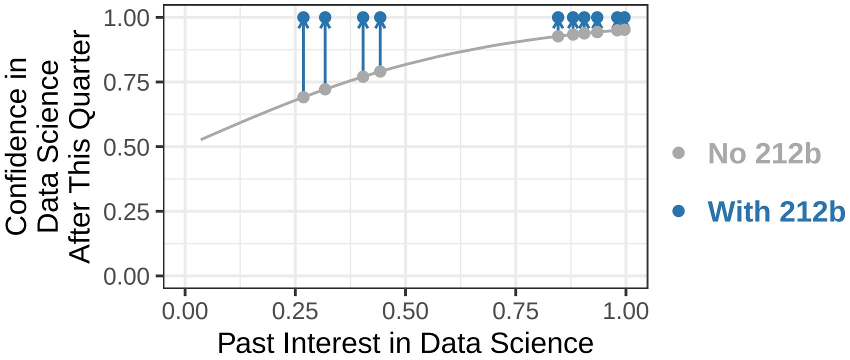

3) Find the treated units of interest

# Outcome model: A visual example




4) Predict their counterfactual outcomes

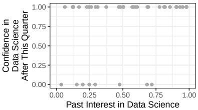
# Outcome model: A visual example



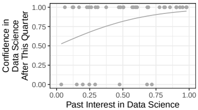

5) Infer causal effect for each person. Average over people

# Outcome model: A visual example

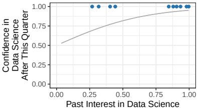
1) Find control units who didn’t take this class



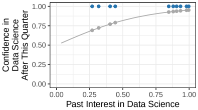

2) Model their outcomes given pre-treatment variables



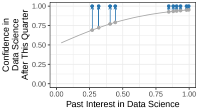

3) Find the treated units of interest




4) Predict their counterfactual outcomes




5) Infer causal effect for each person. Average over people



# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | ? |
| $Y_2^{\text{Takes 212b}}$ | ? |
| $Y_3^{\text{Takes 212b}}$ | ? |
| $Y_4^{\text{Takes 212b}}$ | ? |
| $Y_5^{\text{Takes 212b}}$ | ? |
| $Y_6^{\text{Takes 212b}}$ | ? |

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

**General approach**

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

**General approach**

1) Define potential outcomes

# Outcome model: A visual example

Each Row is a Student in This Class

| | |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

Outcome under 212b

Outcome under no 212b

**General approach**

1) Define potential outcomes
2) Define target population

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

**General approach**

1) Define potential outcomes
2) Define target population
3) Make causal assumptions

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
| --- | --- |
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

**General approach**

1) Define potential outcomes
2) Define target population
3) Make causal assumptions
4) Model unobserved outcomes

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

**General approach**

1) Define potential outcomes
2) Define target population
3) Make causal assumptions
4) Model unobserved outcomes
5) Predict them

# Outcome model: A visual example

Each Row is a Student in This Class

| Outcome under 212b | Outcome under no 212b |
|---|---|
| $Y_1^{\text{Takes 212b}}$ | $\hat{Y}_1^{\text{No 212b}}$ |
| $Y_2^{\text{Takes 212b}}$ | $\hat{Y}_2^{\text{No 212b}}$ |
| $Y_3^{\text{Takes 212b}}$ | $\hat{Y}_3^{\text{No 212b}}$ |
| $Y_4^{\text{Takes 212b}}$ | $\hat{Y}_4^{\text{No 212b}}$ |
| $Y_5^{\text{Takes 212b}}$ | $\hat{Y}_5^{\text{No 212b}}$ |
| $Y_6^{\text{Takes 212b}}$ | $\hat{Y}_6^{\text{No 212b}}$ |

**General approach**

1) Define potential outcomes
2) Define target population
3) Make causal assumptions
4) Model unobserved outcomes
5) Predict them
6) Report an average

# Outcome model: A coding example

Since this part is code, it will follow the website instead of slides.

# Predicting outcomes for causal inference with machine learning

# Outcome model: You are now an expert

1. Assume a DAG

$$
\begin{tikzcd}
\vec{L} \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}
$$

2. By consistency, exchangeability, and positivity,

$$\underbrace{\mathrm{E}(Y^a \mid \vec{L} = \vec{\ell})}_{\text{Causal}} = \underbrace{\mathrm{E}(Y \mid A = a, \vec{L} = \vec{\ell})}_{\text{Statistical}}$$

3. Using regression, estimate $\hat{\mathrm{E}}(Y \mid A, \vec{L})$
4. Predict unknown potential outcomes and average

$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathrm{E}}\left(Y \mid A = a, \vec{L} = \vec{\ell}_i\right)$$

**Big idea:** Why constrain ourselves to regression for $\hat{\mathrm{E}}(Y \mid A, \vec{L})$?

Hill, Jennifer L. 2011.
"Bayesian nonparametric modeling for causal inference."
Journal of Computational and Graphical Statistics 20.1:217-240.

- Binary treatment (simulated)
- Continuous confounder $X$ (simulated)

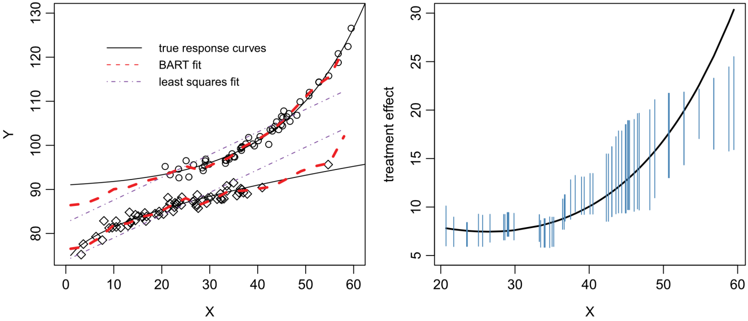



Figure 1. Left panel: simulated data with linear regression and BART fits. Right panel: BART inference for treatment effect on the treated. A color version of this figure is available in the electronic version of this article.

# How did she do that?[1]

1) Learn an automated partitioning of the data (aka a “tree”)

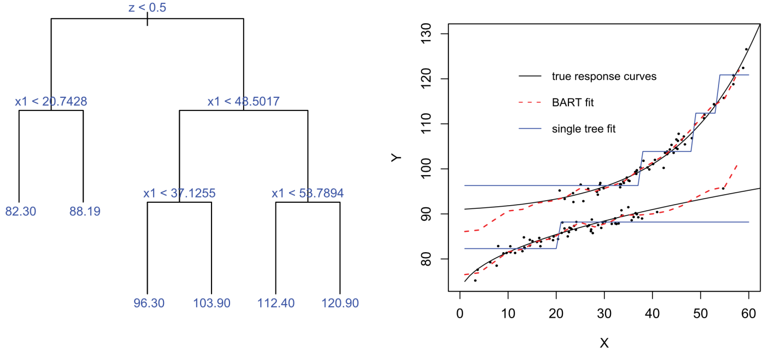



Figure 2. Left panel: the binary tree fit to the data from Figure 1. Right panel: single-tree fits (solid lines) and BART fits (dashed lines). A color version of this figure is available in the electronic version of this article.

[1] Chipman, Hugh A., Edward I. George, and Robert E. McCulloch. “BART: Bayesian additive regression trees.” The Annals of Applied Statistics 4.1 (2010): 266-298.

# How did she do that?

2) Repeat many times. Take the average.

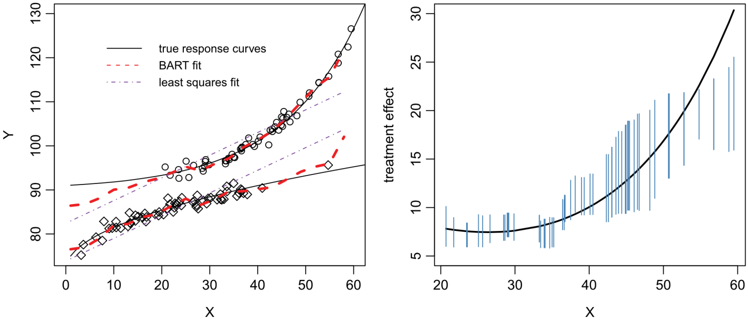



Figure 1. Left panel: simulated data with linear regression and BART fits. Right panel: BART inference for treatment effect on the treated. A color version of this figure is available in the electronic version of this article.

# Core idea: Causal assumptions unlock machine learning[2]

Once you make this assumption

$$\begin{tikzcd} \vec{L} \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

you get to do this

| **Inputs** | $\Rightarrow$ | **Black Box Algorithm** | $\Rightarrow$ | **g-Formula** |
|---|---|---|---|---|
| $\{\vec{L}, A, Y\}$ | | $\boxed{\hat{Y} = \hat{f}(\vec{L}, A)}$ | | $\hat{\text{E}}(Y^a) = \frac{1}{n}\sum_{i=1}^n \hat{f}(a, \vec{\ell}_i)$ |

[2] Caveat: There are ways to do even better. This is just a start.
See Van der Laan, M. J., & Rose, S. (2018). Targeted learning in data science. Springer International Publishing.

There are **so many** algorithms you might use!

*Statistical Science*
2019, Vol. 34, No. 1, 43–68
https://doi.org/10.1214/18-STS667


# Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition[1]

**Vincent Dorie, Jennifer Hill, Uri Shalit, Marc Scott and Dan Cervone**

*Abstract.* Statisticians have made great progress in creating methods that reduce our reliance on parametric assumptions. However, this explosion in research has resulted in a breadth of inferential strategies that both create opportunities for more reliable inference as well as complicate the choices that an applied researcher has to make and defend. Relatedly, researchers advocating for new methods typically compare their method to at best 2 or 3 other causal inference strategies and test using simulations that may or may not be designed to equally tease out flaws in all the competing methods. The causal inference data analysis challenge, "Is Your SATT Where It's At?", launched as part of the 2016 Atlantic Causal Inference Conference, sought to make progress with respect to both of these issues. The researchers creating the data testing grounds were distinct from the researchers submitting methods whose efficacy would be evaluated. Results from 30 competitors across the two versions of the competition (black-box algorithms and do-it-yourself analyses) are presented along with post-hoc analyses that reveal information about the characteristics of causal inference strategies and settings that affect performance. The most consistent conclusion was that methods that flexibly model the response surface perform better overall than methods that fail to do so. Finally new methods are proposed that combine features of several of the top-performing submitted methods.

*Key words and phrases:* Causal inference, competition, machine learning, automated algorithms, evaluation.

## 1. INTRODUCTION

In the absence of a controlled randomized or natural experiment,[2] inferring causal effects involves the difficult task of constructing fair comparisons between ob-

*Vincent Dorie is Associate Research Scientist, Data Science Institute, Columbia University, 475 Riverside Drive, Room 320L, New York, New York 10115, USA (e-mail: vdorie@gmail.com). Jennifer Hill is Professor of Applied Statistics and Data Science, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: jennifer.hill@nyu.edu). Uri Shalit is Assistant Professor, Faculty of Industrial Engineering and Management, Technion, Technion—Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: urishalit@technion.ac.il). Marc Scott is Professor of Applied Statistics, Department of Applied Statistics, Social Science, and Humanities, New York University, 246 Greene Street, 3rd Floor, New York, New York 10003, USA (e-mail: marc.scott@nyu.edu). Dan Cervone is Director of Quantitative Research, Los Angeles Dodgers, Dodger Stadium, 1000 Vin Scully Ave., Los Angeles, California 90012, USA (e-mail: dcervone@gmail.com).*

[1]Discussed in 10.1214/18-STS684; 10.1214/18-STS680; 10.1214/18-STS690; 10.1214/18-STS689; 10.1214/18-STS679; 10.1214/18-STS682; 10.1214/18-STS688

[2]We use natural experiment to include (1) studies where the causal variable is randomized not for the purposes of a study (for instance, a school lottery), (2) studies where a variable is randomized but the causal variable of interest is downstream of this (e.g., plays the role of an instrumental variable), and (3) regression discontinuity designs.

# Dorie et al. 2019[3]: Is Your SATT Where It's At?

[3] Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." Statistical Science, 34(1), 43-68. See also https://jenniferhill7.wixsite.com/acic-2016/competition

# Dorie et al. 2019[3]: Is Your SATT Where It's At?

- Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} \left(Y_i^1 - Y_i^0\right)$$

[3] Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." Statistical Science, 34(1), 43-68. See also https://jenniferhill7.wixsite.com/acic-2016/competition

# Dorie et al. 2019[3]: Is Your SATT Where It's At?

- Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} \left(Y_i^1 - Y_i^0\right)$$

- Simulated data. SATT was known to organizers

[3] Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." Statistical Science, 34(1), 43-68. See also https://jenniferhill7.wixsite.com/acic-2016/competition

# Dorie et al. 2019[3]: Is Your SATT Where It's At?

- Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} \left(Y_i^1 - Y_i^0\right)$$

- Simulated data. SATT was known to organizers
- Confounders were defined by the organizers

[3]Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." Statistical Science, 34(1), 43-68. See also https://jenniferhill7.wixsite.com/acic-2016/competition

# Dorie et al. 2019[3]: Is Your SATT Where It's At?

- Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} \left(Y_i^1 - Y_i^0\right)$$

- Simulated data. SATT was known to organizers
- Confounders were defined by the organizers
- Participants could use any algorithm to estimate SATT

[3] Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." Statistical Science, 34(1), 43-68. See also https://jenniferhill7.wixsite.com/acic-2016/competition

# Dorie et al. 2019[3]: Is Your SATT Where It's At?

- Goal: The Sample Average Treatment Effect on the Treated

$$\text{SATT} = \frac{1}{n_{\text{Treated}}} \sum_{i:A_i=1} \left(Y_i^1 - Y_i^0\right)$$

- Simulated data. SATT was known to organizers
- Confounders were defined by the organizers
- Participants could use any algorithm to estimate SATT
- 30 teams attempted the task

[3] Dorie, V., Hill, J., Shalit, U., Scott, M., & Cervone, D. (2019). "Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition." Statistical Science, 34(1), 43-68. See also https://jenniferhill7.wixsite.com/acic-2016/competition

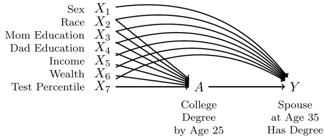
# Words of Warning

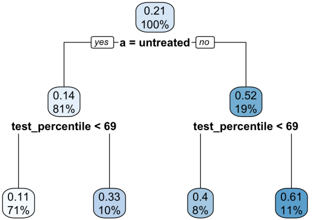
# Good prediction of $Y$ does not guarantee good estimation of $Y^1 - Y^0$

# Good prediction of $Y$ does not guarantee good estimation of $Y^1 - Y^0$

- there may be unmeasured confounding
- the regularization in machine learning models can induce a large bias
- to predict $Y^1$, the model is trained on treated units. But untreated units may have a very different distribution of $\vec{X}$

# Good prediction of $Y$ does not guarantee good estimation of $Y^1 - Y^0$: An example




What is troubling about this tree, if used for causal inference?

# Learning goals for today

By the end of class, you will be able to use an **outcome model** in the service of

- describing a population from a non-probability sample
- inferring average causal effects in an observational study