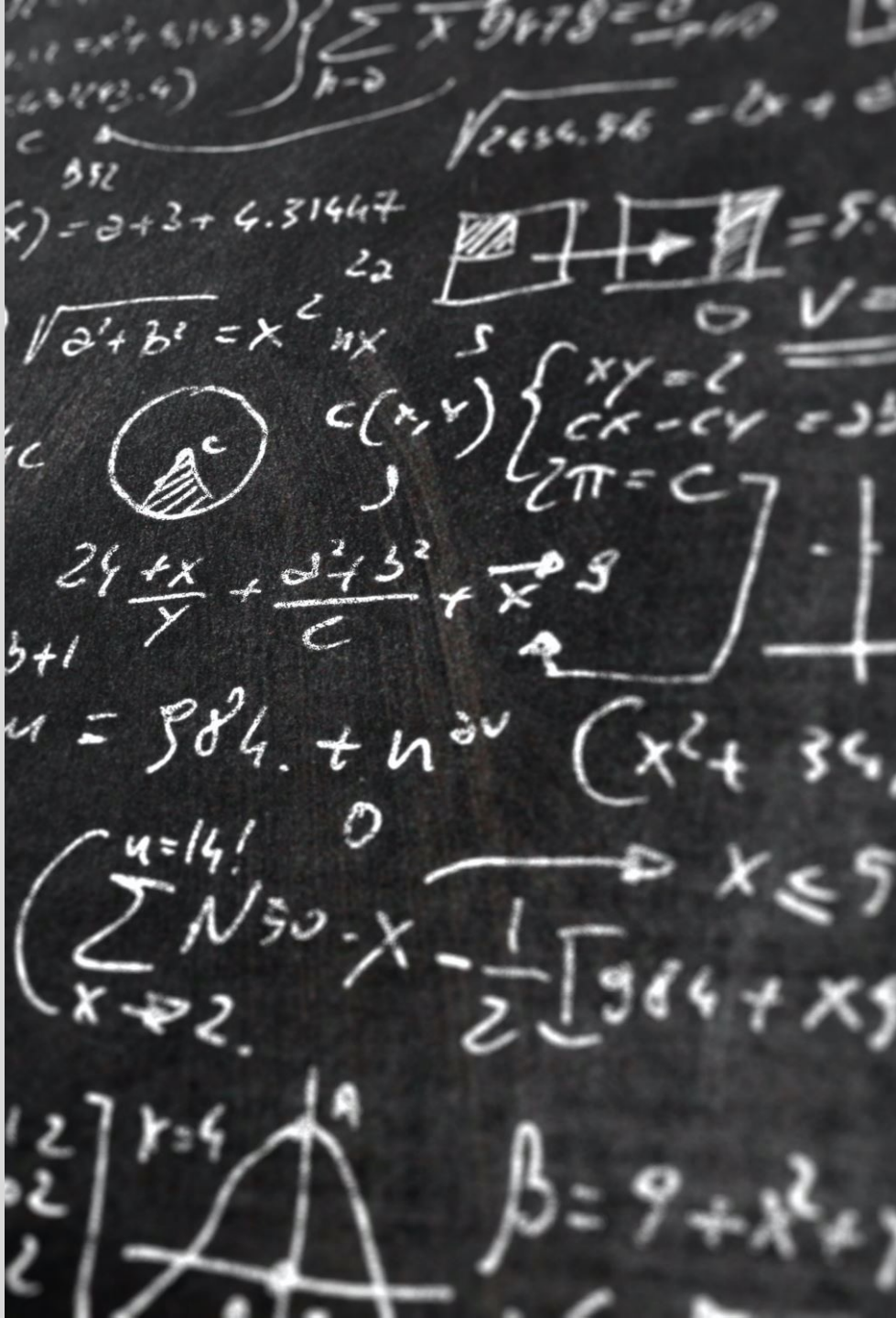
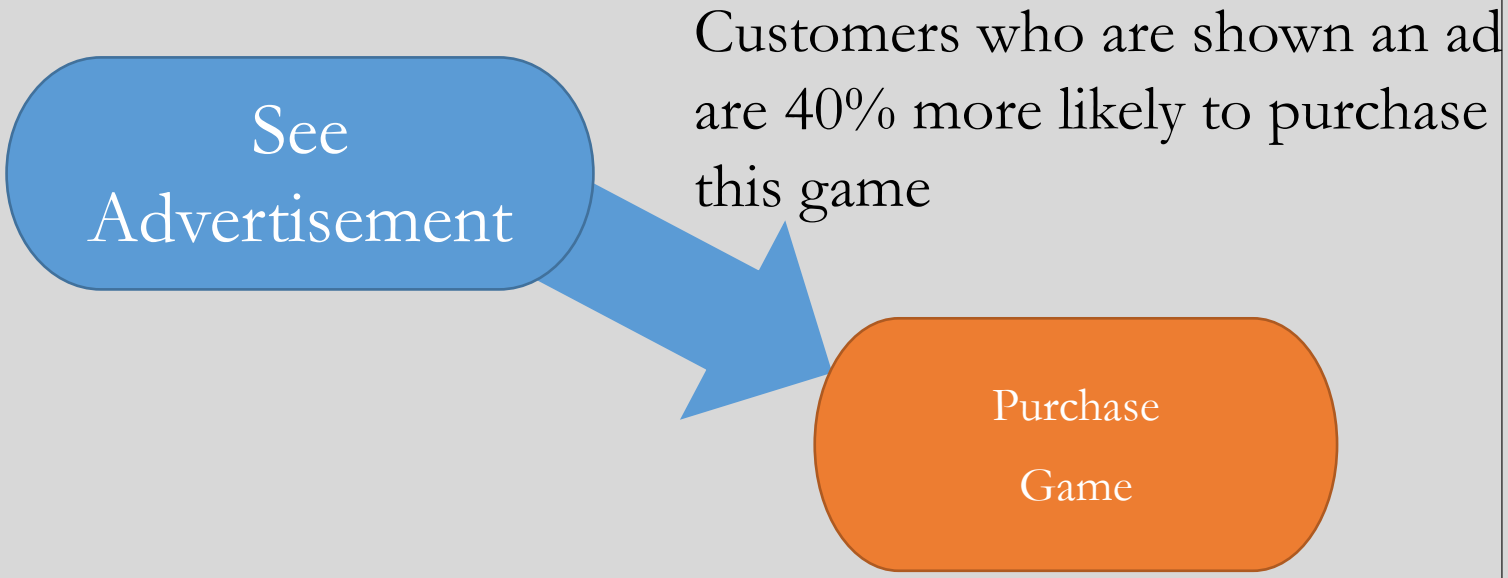
Introduction to Causal Inference

Greg Lewis (Microsoft Research)



Correlation vs Causation

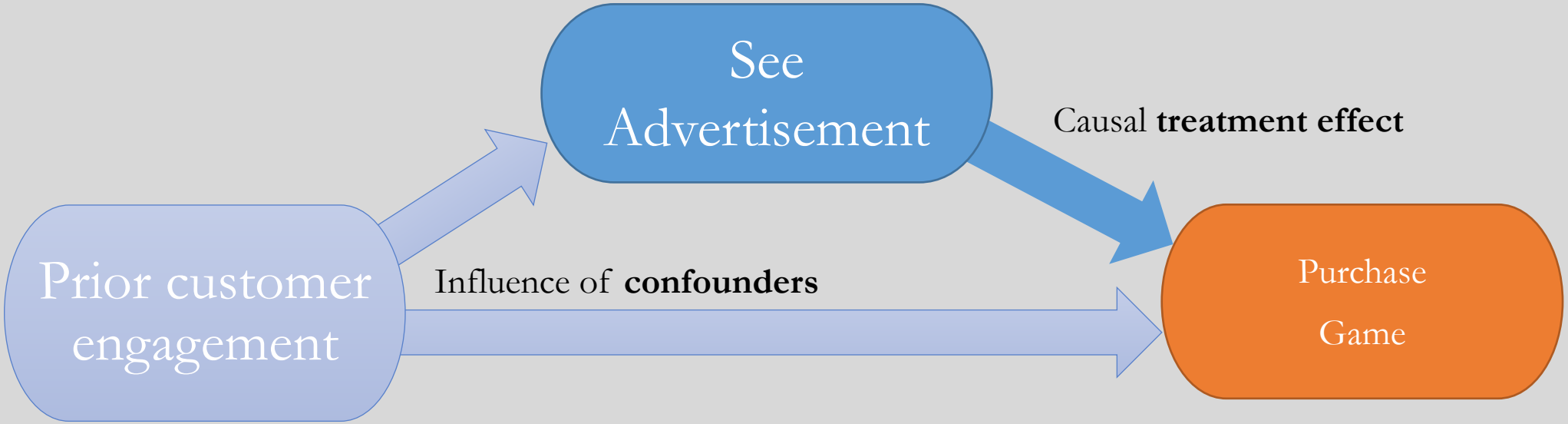
A decision maker budgeting for future ad spending may be encouraged by a past correlation: customers who have seen advertisements in the past are far more likely to later purchase the advertised game.



Uncovering Causal Pathways

This correlation pattern combines multiple causal pathways

1. The **treatment** may have a direct causal effect on the **outcome.**
2. **Confounding features** may influence both the probability of treatment and the outcome, creating additional non-causal correlation.



Forecasting Tools vs Causal Tools

- **Forecasting:** if you want to predict sales (i.e. for inventory) advertisements may be the single best feature—they capture both causal effects and other customer traits

- **Causal:** if you want to estimate the treatment effect of the ad (i.e. to budget for future campaigns), observed correlations can be misleading

How do you identify causal relationships?

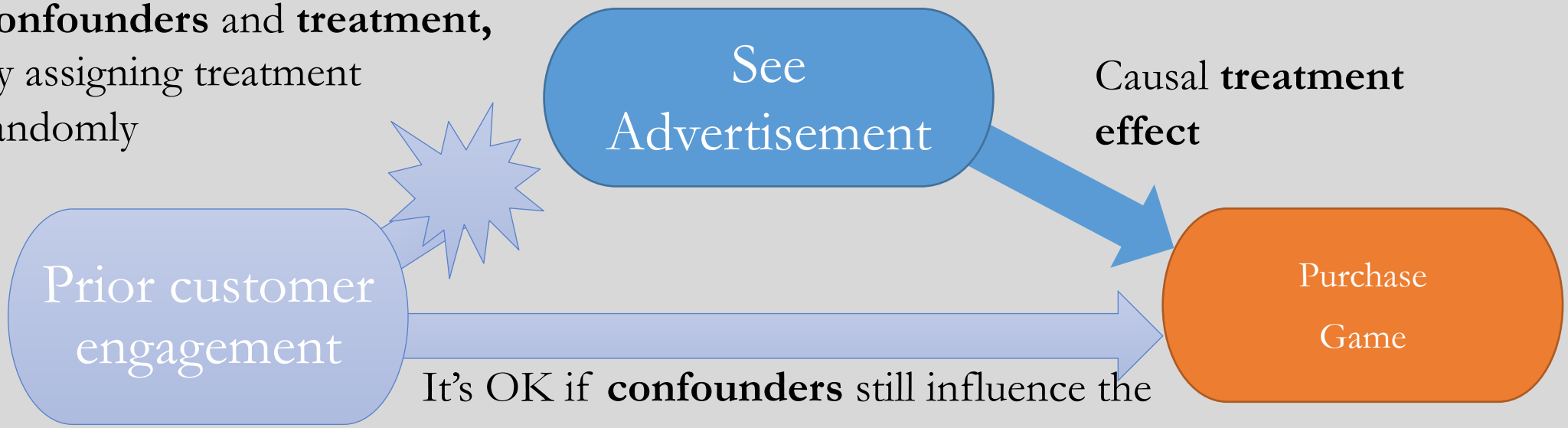
To identify a causal relationship between some treatment T and an outcome Y, you need a quasi-experiment.

There are three main ways to do this.

Randomized Experiments—Why They Work

When it's possible and cost effective to run an experiment, they remain the gold standard for answering causal problems

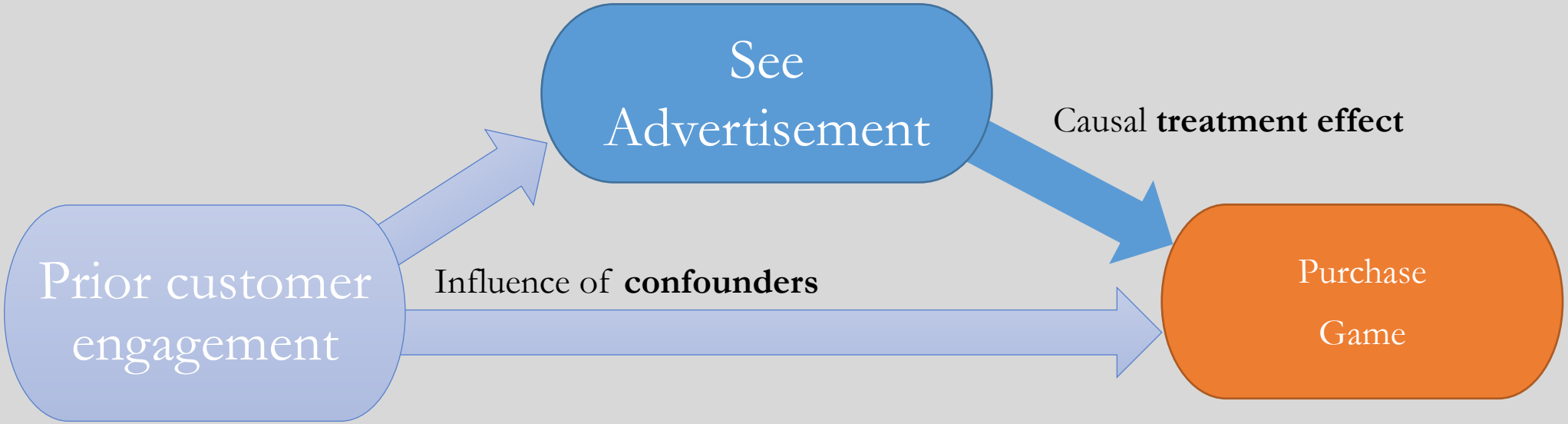
An experiment removes the relationship between **confounders** and **treatment,** by assigning treatment randomly



Measure Confounders—Why It Works

- If you can plausibly measure all confounding influences, carefully designed statistical models can separately estimate each causal path in the graph

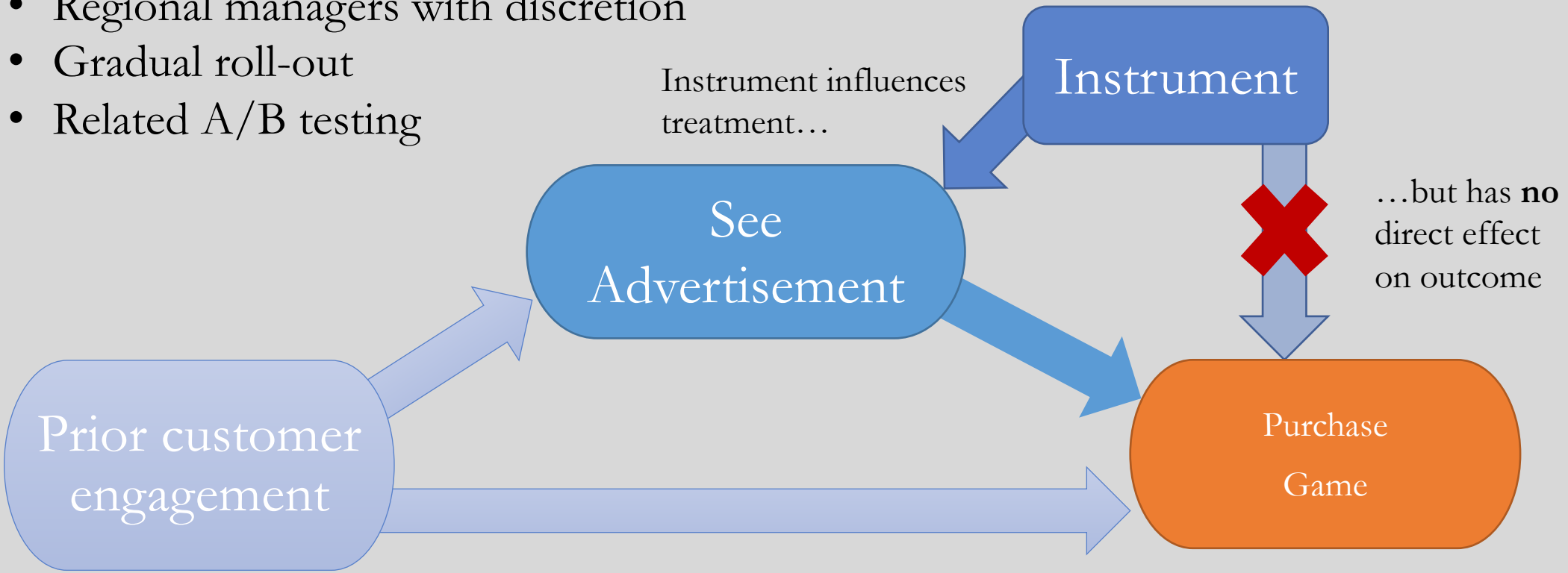
- This case of full observability is known as **unconfoundedness**



Instrumental Variables—Why They Work

Sometimes you can't randomize your treatment, but you can discover randomization that happened naturally or in a related context

- Regional managers with discretion
- Gradual roll-out
- Related A/B testing



How do you identify causal relationships?

To identify a causal relationship between some treatment T and an outcome Y, you need a quasi-experiment

A quasi-experiment is some context in which the distribution of treatments varies with the assignment of some draw Z whose assignment is independent of the potential distribution of outcomes, conditional on some covariates X.

How do you identify causal relationships?

Experiment: the experimental assignment mechanism is Z

Unconfoundedness: the unobserved determinants of treatment assignment are Z ("found experiments")

Instrumental variables: the instrument is Z

Terminology

Start with the following mathematical model:

$$\underbrace{Y_i}_{outcome} = \underbrace{\theta(X_i)}_{ITE} \times \underbrace{T_i}_{treatment} + \underbrace{f(W_i)}_{\cancel{other}\ features} + \underbrace{\varepsilon}_{noise}$$

Individual Treatment Effect (ITE): treatment effect on an individual unit. Not identifiable in a cross-section (nothing to average over).

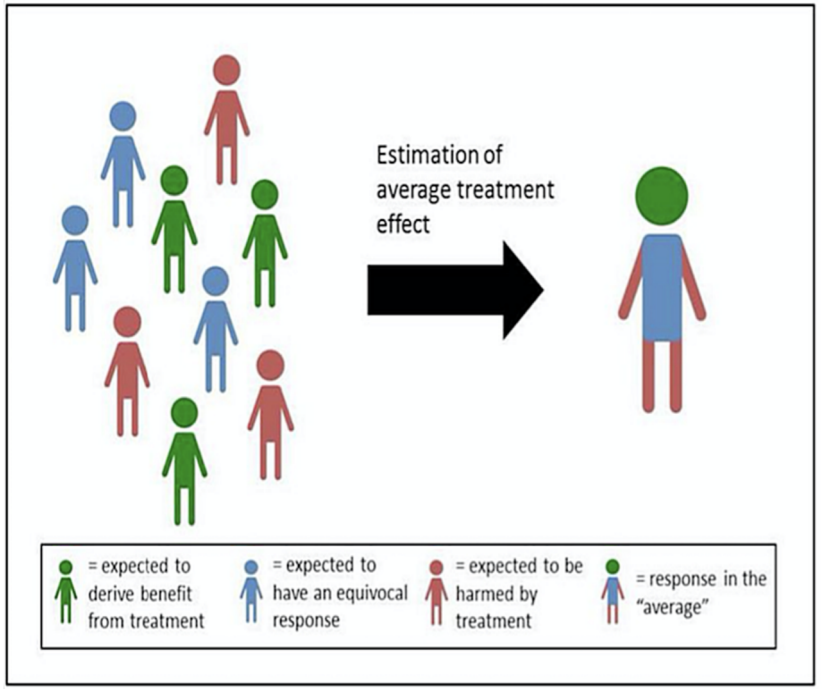
Average Treatment Effect (ATE): average of the ITEs across the population.

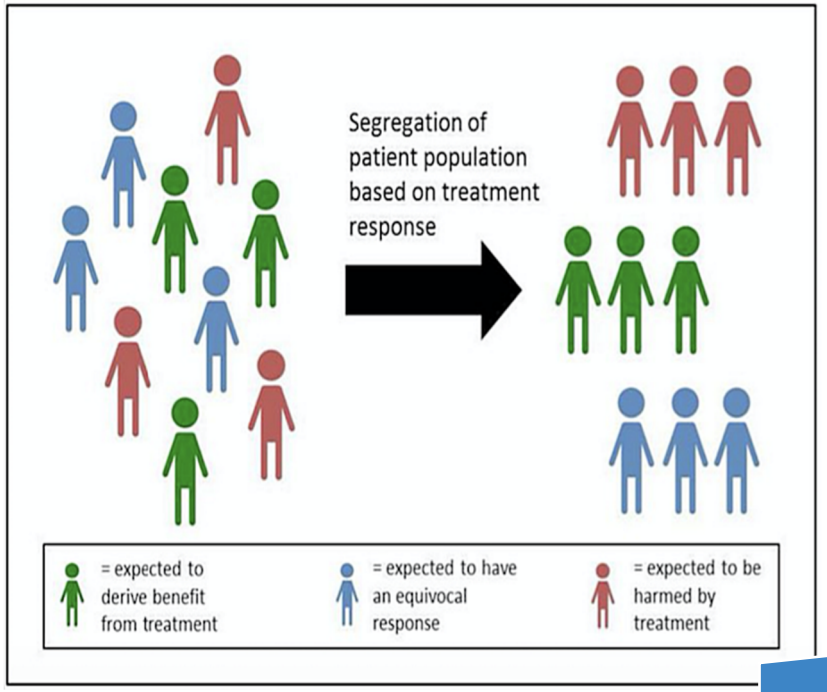
Average Treatment Effect (ATT): average of the ITEs among the population that was treated.

Conditional Average Treatment Effect (CATE): average of the ITEs among some subset of the population (e.g. men).

Local Average Treatment Effect (LATE): for IV specifically, average of the treatment effects weighted by compliance with the IV

Why Machine Learning? : Personalization





From ATE to CATE

Enables personalized experience &
Optimize for incremental effects

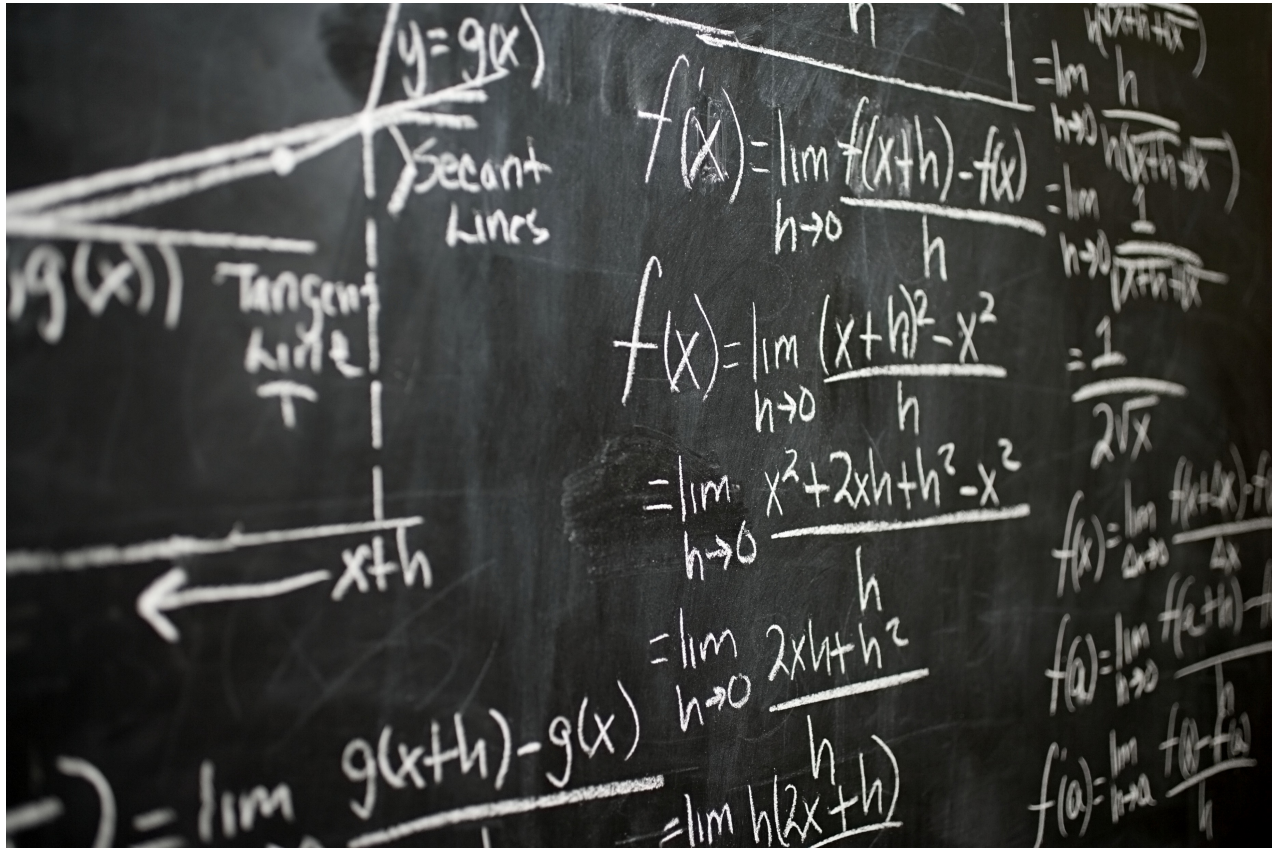
Machine Learning + Causal Inference

Develop optimization applications based on the unlocked insights

Graph credit: Xie, Yuxiang, Nanyu Chen, and Xiaolin Shi. "False discovery rate controlled heterogeneous treatment effect detection for online controlled experiments." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

Why Machine Learning? : Flexibility

We want to control for confounding variables W flexibly.

But we don't know exactly which variables are genuine confounds (i.e. affect both treatment and outcome), or the exact functional form of the relationship.

Learning which variables to include and how is exactly what machine learning is good at!