

## Class 11 Causal Inference & Potential Outcome Framework

Dr. Wei Miao

UCL School of Management

November 6, 2024

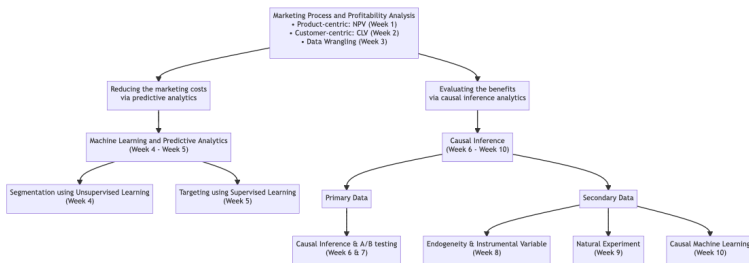
## Section 1

# Causal Inference

# Our Journey So Far

- Any business activity brings benefits and costs. We're given the benefit information in all the case studies so far
  - **Apple (Week 1):** influencer marketing increases sales by 2.5%
  - **1st assignment:** loyalty program increases retention rate to 95%
- In reality, this benefit information is often not readily available, and we need to measure them using **causal inference tools**.

# Causal Inference Road Map

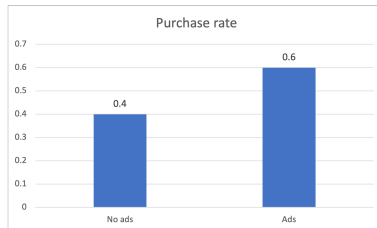


# Learning Objectives

- Understand key concepts of **causal inference**
  - Rubin's potential outcome framework
  - Fundamental problem of causal inference
  - Average treatment effect (ATE) and randomization
- Learn the steps to conduct **randomized controlled trials** (RCTs)

# Why Causal Inference Matters? Example 1

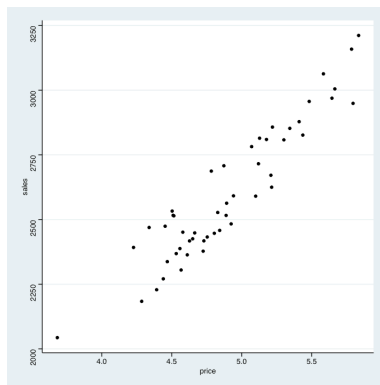
Tom purchases paid ads on Instagram to advertise his new bubble tea shop. Instagram ads are targeted to individuals who are predicted to have a higher likelihood of being bubble tea lovers. In the end, some Instagram users saw no ads and some saw the ads. The **purchase rates** for each group are shown below.



**Question:** Can Tom be confident to conclude the Instagram ads are effective in converting new customers?

## Why Causal Inference Matters? Example 2

Tom bought a marketing survey data from a consulting agency. The survey collected prices and store visits (sales) for different bubble tea shops in Canary Wharf. Tom finds that there seems to be a positive relationship between prices and store visits.

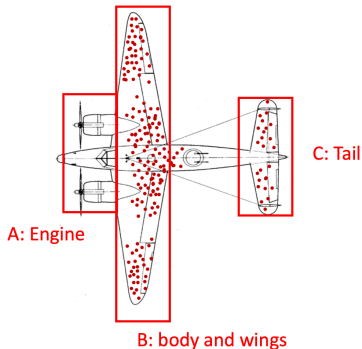


**Question:** Can Tom conclude that he should also increase the prices for his bubble tea shop to increase the store visits?

## Why Causal Inference Matters? Example 3

This is a fighter plane that just returned from the battlefield. Red dots are bullet holes.

Which part of A, B, and C would you reinforce to increase the pilot's survival rate?





## Why Causal Inference Matters? Example 4

I have held a secret from you for a long time, it's time for me to confess ....

# Nobel Prize in Economics (2021)



Ill. Niklas Elmehed © Nobel Prize Outreach.

**David Card**

Prize share: 1/2



Ill. Niklas Elmehed © Nobel Prize Outreach.

**Joshua D. Angrist**

Prize share: 1/4



Ill. Niklas Elmehed © Nobel Prize Outreach.

**Guido W. Imbens**

Prize share: 1/4

*[...] the other half jointly to Joshua D. Angrist and Guido W. Imbens “for their methodological contributions to the analysis of causal relationships.”*

# Causal Inference

- Causal inference is the process of estimating the unbiased causal effects of a particular policy/business intervention on the outcome variables of interest.
- Correlation  $\neq$  Causation: **Machine learning** models are good at finding correlations, but not causations. For example, on rainy days, we observe more umbrellas on the street
  - correct correlation statement: number of umbrellas is **positively correlated** with rainfall
  - correct causal statement: heavier rain **leads to** more umbrellas
  - incorrect causal statement: more umbrellas **lead to** heavier rain
- Causality becomes more complex in the business world. Managers can easily make mistakes without causal inference training.
  - Imagine if the actual causal effect of Instagram ads on incremental profit per customer is £1 for Tom, and Tom pays £1.5 for each click

## Section 2

# Potential Outcome Framework

# Rubin Causal Model and the Potential Outcome Framework

- The **Rubin causal model (RCM)** or the **Potential Outcome Framework** is the well accepted framework for thinking about causal effects.
- For each customer  $i$ , we can define the **potential outcomes** in order to evaluate the causal effect of a treatment on their outcomes:
  - $Y_i^1$ : the outcome if the customer is exposed to the treatment, ceteris paribus
  - $Y_i^0$ : the outcome if the customer is not exposed to the treatment, ceteris paribus

## Definition: Individual Treatment Effects

- The causal effect of the treatment on the individual  $i$  is the difference between the two potential outcomes. We define the **individual treatment effect**  $\delta_i$  as follows

$$\delta_i = Y_i^1 - Y_i^0$$

### ! Important

These two potential outcomes do not depend on whether or not the individual is exposed to the treatment in reality.

## Example

**Tom would like to measure the causal effect of seeing a displayed ad on customer purchase probability.**

## Examples of Individual Treatment Effects

Let's say we have retrieved all infinity stones from Thanos, and have created 2 parallel universes

- In Universe 1, with a displayed ad
  - Dr Strange has a rate of 70%
- In Universe 2, without a displayed ad
  - Dr Strange has a rate of 60%

| Customer   | Y1  | Y0  | TE  |
|------------|-----|-----|-----|
| Dr Strange | 0.7 | 0.6 | 0.1 |



# A Motivating Example: A Group of Customers

- Conceptually, we can collect a sample of customers, and estimate individual treatment effect for each of them

| Customer   | Y1   | Y0   | TE    |
|------------|------|------|-------|
| Dr Strange | 0.70 | 0.60 | 0.10  |
| Iron Man   | 0.55 | 0.50 | 0.05  |
| Thor       | 0.80 | 0.72 | 0.08  |
| Hulk       | 0.60 | 0.62 | -0.02 |

# Fundamental Problem of Causal Inference

- To measure the individual treatment effect of the ads on a customer's purchase rate, we need to observe **all potential outcomes** of the **same individual** in **parallel universes** (i.e., with and without being exposed to the treatment).
- We use  $D_i = 1$  to denote the treated/treatment group and  $D_i = 0$  to denote the untreated/control group in reality.
- We only observe one potential outcome, the **realized outcome**, in our reality
  - For treated customers, we only observe their realized outcomes of being treated:  $Y_i^1 | D_i = 1$
  - For untreated customers, we only observe their realized outcomes of being untreated  $Y_i^0 | D_i = 0$
- The remaining potential outcomes, i.e., **counterfactual outcomes**, are never observed in our reality
  - For treated customers, we never observe their outcomes when being untreated:  $Y_i^0 | D_i = 1$
  - For untreated customers, we never observe their outcomes when treated:  $Y_i^1 | D_i = 0$

## Exercise

- In the previous table, let's say, Dr Strange and Hulk are treated (by the ads) in reality, while Iron Man and Thor are not treated.
- Based on this information, we can find out the values of the realized outcomes and counterfactual outcomes for each customer.
- For example, for Dr Strange:
  - Realized outcome:  $Y_i^1 | D_i = 1 = 0.7$  (Note that we only observe this value in reality)
  - Counterfactual outcome:  $Y_i^0 | D_i = 1 = 0.6$  (Note that we don't observe this value in reality)

| Customer   | Y1  | Y0   | Treated |
|------------|-----|------|---------|
| Dr Strange | 0.7 | ?    | Yes     |
| Iron Man   | ?   | 0.5  | No      |
| Thor       | ?   | 0.72 | No      |
| Hulk       | 0.6 | ?    | Yes     |

# Fundamental Problem of Causal Inference

- Since it is impossible to see both potential outcomes at once, one of the potential outcomes is always missing, so we can never quantify the individual treatment effects. This dilemma is called the **Fundamental Problem of Causal Inference**.

| subject    | Treated | Y1  | Y0   | Y1-Y0 |
|------------|---------|-----|------|-------|
| Dr Strange | Yes     | 0.7 | ?    | ?     |
| Iron Man   | No      | ?   | 0.5  | ?     |
| Thor       | No      | ?   | 0.72 | ?     |
| Hulk       | Yes     | 0.6 | ?    | ?     |

## Section 3

# Average Treatment Effects

# The Average Treatment Effect

- Since individual treatment effects are unobservable, we often care more about the **average treatment effects** (ATE) on the population level. The ATE is defined as the average of individual treatment effects across the population.

$$ATE = E[Y_i^1 - Y_i^0] = \frac{1}{N} \sum_{i=1}^N (Y_i^1 - Y_i^0)$$

- For display ads, can we obtain the ATE by directly calculating the difference in the average outcomes between the treatment group and control group. Why?

$$\begin{aligned} & E[Y^1 | D_i = 1] - E[Y^0 | D_i = 0] \\ &= E[Y^1 | D_i = 1] - E[Y^0 | D_i = 1] \\ &+ E[Y^0 | D_i = 1] - E[Y^0 | D_i = 0] \end{aligned}$$

# Data Example

- Please load `data_treatmenteffect` in the data folder into your RStudio.

```
data_treatmenteffect <- read.csv("https://www.dropbox.com/scl/fi/ndvvwr
```

- **Exercise:** This data are generated from Instagram's paid ads, treated customers are those who see Instagram's ads.
  - Compute the difference in the average rates between the treated and untreated customers.
  - Compute the ATE based on the individual treatment effects.
  - Compare the two results.

# The Average Treatment Effect

- To quantify the correct ATE, we must **randomize who receives the treatment**, instead of targeting or letting the individuals choose the treatment.
- After randomization, we can then obtain the ATE by comparing the difference in the average outcomes across the treatment group and control group. Because randomization ensures that
  - Selection bias is fully removed<sup>1</sup>
  - The treatment effects on the treatment group individuals and the control group individuals should be equal. The former is called the **average treatment effects on the treated (ATT)**, and the latter is called the **average treatment effects on the untreated (ATU)**.

**Exercise:** Let's go back to the previous data example and compute the ATT and ATU after randomization.

---

<sup>1</sup>Selection bias refers to the pre-existing difference between the treatment group and control group even without the treatment



# Gold Standard of Causal Inference

- Mathematically, the previous slide can be represented by the **Basic Identity of Causal Inference**:

$$\begin{aligned} & E[Y^1 | D_i = 1] - E[Y^0 | D_i = 0] \\ &= E[Y^1 | D_i = 1] - E[Y^0 | D_i = 1] \\ &+ E[Y^0 | D_i = 1] - E[Y^0 | D_i = 0] \end{aligned}$$

Outcome for treated – Outcome for untreated  
= [Outcome for treated – Outcome for treated if not treated]  
+ [Outcome for treated if not treated  
– Outcome for untreated]  
= Impact of treatment on treated + selection bias.

- Randomized experiments makes ATT equal to ATE, and removes selection bias. Thus, randomized experiments are the **gold standard of causal inference**.