

# Lecture 4. The Potential Outcome Framework

---

Vinish Shrestha

Towson University

Motivation

A simple example

Potential Outcome Framework

Average Treatment Effect (ATE)

A completely randomized experiment

Average Treatment Effect on the Treated (ATT)

Unconfoundedness assumption

References

# Motivation

---

# Motivation

- By now, we know are tired of hearing that **correlation is not causality**
  - WE KNOW!!!
- A researcher can perform controlled experiments to determine whether A causes B by controlling for confounders
- The complexities and interrelations of human behavior create a setting starkly different from the controlled environment of a lab
- This complicates things

- Causal Inference: A process to determine whether A causes B in both lab settings and out-of-lab scenarios.
- We'll be focusing on out-of-lab situations in both experimental and non-experimental settings.

## A simple example

---

## A simple example

- Say, we are interested in evaluating the effects of a tutoring program on exam scores for an introductory course.
- For now, we'll assume that the treatment is completely random.
- The class is randomly divided into two groups
  1. treatment
  2. control



## Complete randomization

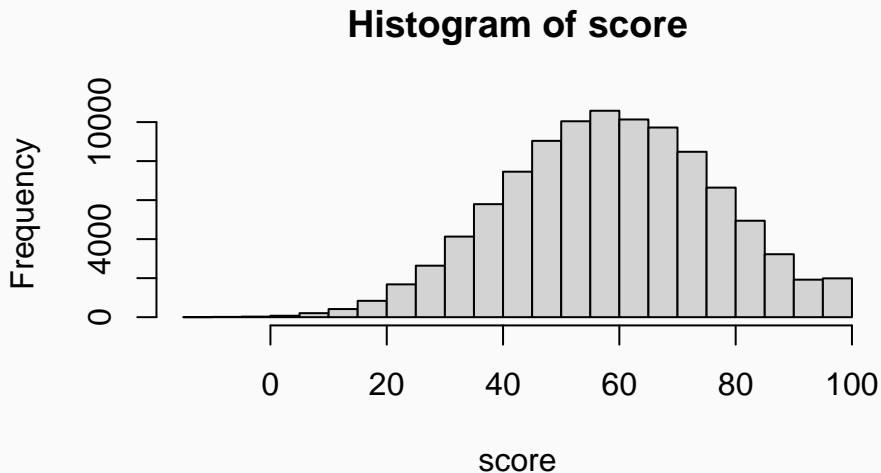
- Each individual has an equal probability of receiving the treatment
- Likely to achieve balance in baseline characteristics (with an arbitrarily high probability) as sample size increases
- Balance: An instance when all pre-treatment covariates between the treatment and control groups are similar.
- If this is attained then it increases confidence that the treatment and the control units are comparable.

## Example

```
set.seed(12577)
n <- 100000 # observations
gender <- rbinom(n, 1, 0.5) # female 1
ability <- rnorm(n, 0, 1)
# income follows the log normal distributing
income <- rlnorm(n, meanlog = 1, sdlog = 0.5)
# multiplying the log normal dist with 20000
income <- income * 20000
lowincome <- ifelse(income < median(income), 1, 0)
highwork <- rbinom(n, 1, 0.3 + 0.2 * lowincome + 0.1 * gender)
# score
score <- 65 - (highwork * 20) - (lowincome * 5) + gender * 10 + ability*10
score <- ifelse(score > 100, 100, score)
```

# Histogram of score

```
hist(score)
```



# Data

```
data <- data.frame(score, lowincome, gender, highwork, ability)
head(data)
```

| ## |   | score    | lowincome | gender | highwork | ability     |
|----|---|----------|-----------|--------|----------|-------------|
| ## | 1 | 30.94561 | 0         | 0      | 1        | -0.33448622 |
| ## | 2 | 52.94731 | 1         | 1      | 1        | -0.11780092 |
| ## | 3 | 61.57899 | 1         | 0      | 0        | 1.48851940  |
| ## | 4 | 57.44281 | 1         | 0      | 1        | 0.81045128  |
| ## | 5 | 41.09817 | 0         | 0      | 0        | -0.95822789 |
| ## | 6 | 68.09644 | 1         | 1      | 0        | -0.09275551 |

## Score by baseline characteristics

```
reg <- lm(score ~ lowincome + gender + highwork + ability, data)
# Print the summary output cleanly
print(summary(reg))
```

```
##
## Call:
## lm(formula = score ~ lowincome + gender + highwork + ability,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -43.662  -6.706   0.061   6.781  43.430
##
## Coefficients:
```

## Random assignment of treatment (flipping of a coin)

```
treat  <- rbinom(n, 1, 0.5) # random assignment of treatment
data   <- cbind(treat, data)

# get summary by treatment
data_sum <- data %>%
  group_by(treat) %>%
  summarize(lowincome = mean(lowincome),
            gender = mean(gender),
            highwork = mean(highwork),
            ability = mean(ability),
            score = mean(score))
```

## Summary of baseline characteristics by treatment

```
data_sum %>%
```

```
kable()
```

| treat | lowincome | gender    | highwork  | ability    | score    |
|-------|-----------|-----------|-----------|------------|----------|
| 0     | 0.4989493 | 0.4988292 | 0.4500951 | 0.0020368  | 58.46954 |
| 1     | 0.5010493 | 0.5022884 | 0.4462077 | -0.0068307 | 58.49362 |

# Potential Outcome Framework

---



# Setup

- $W$  denotes the treatment status
  - $W_i \in \{0, 1\}$
- $Y_i$  is the exam score following the treatment Assignment
- $X_i$  are the covariates (e.g., gender, race).
- The subscript  $i$  indicates an individual or unit of observation

# Potential Outcome Framework: Neyman-Rubin Causal Model

- Use the potential outcome framework to describe treatment effects
- Follow the Neyman-Rubin causal model (Splawa-Neyman, Dabrowska, and Speed 1923 [1990]; Rubin 1974)

## More Notations

- $Y_i(0)$ : the potential outcomes for an individual  $i$  in the case of no treatment
- $Y_i(1)$  as the potential outcomes for an individual  $i$  in the case of treatment
- Potential outcomes are not realized yet

## What is observed and what isn't?

- a)  $[Y_i(0)|W_i = 0]$ . The potential outcome of an unit  $i$  in the no-treatment state conditional upon  $i$  actually not receiving the treatment. *observed outcome*
- b)  $[Y_i(0)|W_i = 1]$ . The potential outcome of an unit  $i$  who received the treatment ( $W_i = 1$ ) would be in absence of the treatment. This is not observed.

Similarly,

- c)  $[Y_i(1)|W_i = 1]$ . The potential outcome of an unit  $i$  in the treatment state conditional upon  $i$  receiving treatment.
- d)  $[Y_i(1)|W_i = 0]$ . The potential outcome of an unit  $i$  in the treatment state conditional upon  $i$  not receiving treatment. This is not observed.





# Illustration of the Potential Outcome Framework

## Potential Outcome



The potential outcomes in filled circles are observed.

The respective counterfactual outcomes are in empty circles.

potential outcome    no treated cf    no treated observed    treated cf    treated observed

## Counterfactual outcomes

- What would've been the outcome for an observed unit  $i$  in a different treatment state.

### Using the illustration

- Note that person A is treated
- Outcome for person A is observed in the treatment state
  - $Y_A(1)|W_A = 1$
- The ideal *counterfactual* for A is  $A_0$ : the outcome for person A in the state of no treatment
  - $Y_A(0)|W_A = 1$

# Fundamental Problem in causal inference

- Write the observed variable  $Y_i$ , can be written as a function of the potential outcomes as follows:

$$Y_i = W \times \underbrace{Y_i(1)}_{\text{outcome in treatment state}} + (1 - W) \times \underbrace{Y_i(0)}_{\text{outcome in no-treatment state}} \quad (1)$$

- The problem is that one cannot observe both  $Y_i(0)$  and  $Y_i(1)$  at the same time.
- The causal inference through the lens of Neyman-Rubin causal framework can be seen as the missing data problem

*If one has the data for  $Y_i(1)$  then the  $Y_i(0)$  counterpart is missing and vice-versa. Much of causal inference is finding ways to deal with the missing-data problem.*

## To proceed further: we need assumptions

- **A1. Independence assumption.** States that the treatment is independent of potential outcomes.
- $W_i \perp Y_i(0), Y_i(1)$
- Quite literally it means that the person A has 50% chance for being in the treatment and no-treatment state each.
- Also, states that the treatment assignment is independent of any covariates  $X_i$ .
  - male and female;
  - high income and low income;
  - employed vs. not working

## Other assumptions

**A2. Overlap.** States that observations in both the treatment and control groups fall within the common support.

- For instance, this assumption is violated if the treatment group consist of all females and the control group consist of all males as one would not be able to attain balance in covariates.

**A3. Stable Unit Treatment Value Assumption (SUTVA).** no interference assumption

- Treatment status of one unit should not affect the potential outcome for other units
- This assumption breaks down if there are spillover effects.



## Average Treatment Effect (ATE)

---

## Individual specific treatment effect

- Say, parallel universes exist
- Say, you can observe unit A in treatment and no-treatment states
- You can evaluate the treatment effect specific to person A
- Do the same for B, C, ...

*This is (not yet) possible. However, one can identify ATE under the governing assumptions.*

- Let's start with individual treatment effect

$$ITE_i = Y_i(1) - Y_i(0) \quad (2)$$

- not feasible
- New Target:

$$ATE = E(Y_i(1)) - E(Y_i(0)) \quad (3)$$

- $Y_i(1)$ : outcome for an unit  $i$  in the treatment state,
- $Y_i(0)$ : outcome for the same unit  $i$  in absence of the treatment.

Using the independence assumption:  $W_i \perp Y_i(0), Y_i(1)$

$$ATE = E(Y_i(1)) - E(Y_i(0)) \quad (4)$$

$$= E(Y_i|W_i = 1) - E(Y_i|W_i = 0) \quad (5)$$

- $E(Y_i(1)) = E(Y_i|W_i = 1)$  &  $E(Y_i(0)) = E(Y_i|W_i = 0)$
- We have expressed expected potential outcome as observed outcome using the independence assumption
- ATE is concerned with the whole population

# A completely randomized experiment

---

# Randomized Controlled Trials (RCT)

- Is the cornerstone of causal inference
- Often referred to as a gold standard
- ATE is identified through the randomization of treatment assignment
- $W \perp Y_i(0), Y_i(1)$  ensures that treatment and control groups are comparable in expectation
- RCT a straightforward yet immensely powerful

## A complete random assignment

- The probability of treatment  $e$  is same for all units
- coin toss or a bernoulli trail
- In an RCT setting, the difference-in-means estimator is given as:

$$\hat{\tau} = \frac{1}{N_t} \sum_{W_i=1} Y_i - \frac{1}{N_c} \sum_{W_i=0} Y_i \quad (6)$$

*The difference-in-mean estimator is unbiased and consistent for the average treatment effect.*

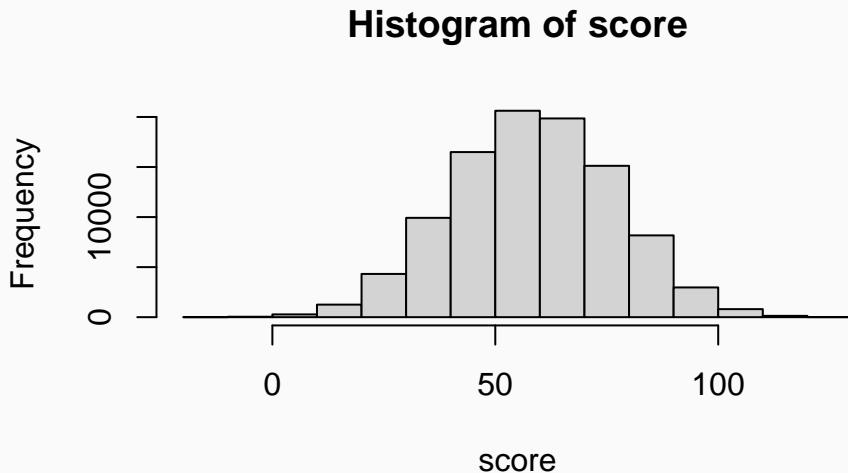
## Example (ATE = 5)

```
set.seed(12577)
n <- 100000 # observations
gender <- rbinom(n, 1, 0.5) # female 1
ability <- rnorm(n, 0, 1)
# income follows the log normal distributing
income <- rlnorm(n, meanlog = 1, sdlog = 0.5)
# multiplying the log normal dist with 20000
income <- income * 20000
lowincome <- ifelse(income < median(income), 1, 0)
highwork <- rbinom(n, 1, 0.3 + 0.2 * lowincome + 0.1 * gender)
# score
score <- 65 - (highwork * 20) - (lowincome * 5) + gender * 10 + ability*10
```



# Histogram of score

```
hist(score)
```



# Data

```
data <- data.frame(score, lowincome, gender, highwork, ability)
head(data)
```

| ## |   | score    | lowincome | gender | highwork | ability     |
|----|---|----------|-----------|--------|----------|-------------|
| ## | 1 | 30.94561 | 0         | 0      | 1        | -0.33448622 |
| ## | 2 | 52.94731 | 1         | 1      | 1        | -0.11780092 |
| ## | 3 | 61.57899 | 1         | 0      | 0        | 1.48851940  |
| ## | 4 | 57.44281 | 1         | 0      | 1        | 0.81045128  |
| ## | 5 | 41.09817 | 0         | 0      | 0        | -0.95822789 |
| ## | 6 | 68.09644 | 1         | 1      | 0        | -0.09275551 |

## Score by baseline characteristics

```
reg <- lm(score ~ lowincome + gender + highwork + ability, data)
# Print the summary output cleanly
print(summary(reg))
```

```
##
## Call:
## lm(formula = score ~ lowincome + gender + highwork + ability,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -43.845  -6.754   0.022   6.779  43.533
##
## Coefficients:
```

## Random assignment of treatment (flipping of a coin)

```
treat  <- rbinom(n, 1, 0.5) # random assignment of treatment
data   <- cbind(treat, data)
score_later <- score + (5 * treat) + rnorm(n, 5, 2.5)
data    <- cbind(data, score_later)

# get summary by treatment
data_sum <- data %>%
  group_by(treat) %>%
  summarize(lowincome = mean(lowincome),
            gender = mean(gender),
            highwork = mean(highwork),
            ability = mean(ability),
            score = mean(score))
```

## Summary of baseline characteristics by treatment

```
data_sum %>%
```

```
kable()
```

| treat | lowincome | gender    | highwork  | ability    | score    |
|-------|-----------|-----------|-----------|------------|----------|
| 0     | 0.4989493 | 0.4988292 | 0.4500951 | 0.0020368  | 58.52175 |
| 1     | 0.5010493 | 0.5022884 | 0.4462077 | -0.0068307 | 58.54164 |

## ATE (when treatment is independent of baseline)

```
tau <- mean(data$score_later[data$treat == 1]) -  
mean(data$score_later[data$treat == 0])  
  
cat("The ATE estimate is:", tau)  
  
## The ATE estimate is: 5.009525
```

## Case when treatment depends on baseline

```
treat2 <- rbinom(n, 1, 0.5 - 0.3 * data$highwork) # dependent on baseline  
data <- cbind(data, treat2)  
data$score_later2 <- data$score + (5 * data$treat2) + rnorm(n, 5, 2.5)
```

## ATE (when treatment is dependent of baseline)

```
tau2 <- mean(data$score_later2[data$treat2 == 1]) -  
mean(data$score_later2[data$treat2 == 0])  
  
cat("The ATE estimate when \n treatment is dependent:", tau2)  
  
## The ATE estimate when  
## treatment is dependent: 11.4312
```



- Simple difference in means across the treatment and no-treatment group will be enough to recover the ATE if independence assumption holds
- The difference in means will generate unbiased estimate of ATE
- However, if the treatment depends on baseline characteristics, then the simple difference in means estimate will likely be biased
- In the upcoming lecture we will see how to alleviate these bias.

## Average Treatment Effect on the Treated (ATT)

---

- ATT pertains to the evaluation of treatment effects for only those units that are treated.

Formally, it is defined as:

$$ATT = E(Y_i(1) - Y_i(0) | W_i = 1) \quad (7)$$

## ATT (using the independence assumption)

- Re-write expected potential outcomes as observed outcomes using the independence assumption:

$$ATT = E(Y_i(1) - Y_i(0)|W_i = 1) \quad (8)$$

$$= E(Y_i(1)|W_i = 1) - \underbrace{E(Y_i(0)|W_i = 1)}_{=E(Y_i(0)|W_i=0)} \quad (9)$$

$$= E(Y_i(1)|W_i = 1) - E(Y_i(0)|W_i = 0) \quad (10)$$

$$= E(Y_i|W_i = 1) - E(Y_i|W_i = 0) \quad (11)$$

# Unconfoundedness assumption

---

## Independence Assumption is too strict

- Treatment assignment being fully random is very rare
- Assignment is driven by some selective covariates
- Referring to the tutoring example, it may be unethical to disallow someone in the control group who wants to attend the tutoring sessions.
- Tutoring sessions may be voluntarily held

## Other times treatment targets specific subgroup

- Say, low income subgroups are more likely to be treated
- Here, the treatment assignment is not completely random by systematically depends on the baseline characteristics
- We require adjustments before being able to estimate treatment effects in such cases
- As long as: i) treatment assignment mechanism is known; and ii) variables governing assignment are observed, we can proceed head with **unconfoundedness assumption**

# Unconfoundedness assumption

- Sometimes called the conditional independence assumption
- Formally, this states that  $Y_i(0), Y_i(1) \perp W_i | X_i$ .
- This means that conditional upon the covariates the treatment assignment is random.
- To estimate ATT
  - estimate ATT within each strata (high and low income groupings)
  - ATT estimate = weighted average of the strata-estimates
  - weights = proportion of high vs. low income in the sample



## ATT estimation with unconfoundedness assumption

- We know that people who work more are less likely to receive treatment

```
## ATT estimate within highwork group
```

```
tau1 <- mean(data$score_later2[data$treat2 == 1 & data$highwork == 1]) -  
  mean(data$score_later2[data$treat2 == 0 & data$highwork == 1])
```

```
## ATT estimate within no-highwork (low work) group
```

```
tau2 <- mean(data$score_later2[data$treat2 == 1 & data$highwork == 0]) -  
  mean(data$score_later2[data$treat2 == 0 & data$highwork == 0])
```

```
# weight
```

```
wgt <- mean(data$highwork)
```

```
# weighted average
```

```
tau2 <- (wgt * tau1) + (1 - wgt) * (tau2)
```

## ATT estimation with unconfoundedness assumption

```
## The ATE estimate under unconfoundedness is: 5.022411
```

## References

---

- Rubin, Donald B. 1974. “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.” *Journal of Educational Psychology* 66 (5): 688.
- Splawa-Neyman, Jerzy, Dorota M Dabrowska, and Terrence P Speed. 1923 [1990]. “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.” *Statistical Science*, 1923 [1990], 465–72.