

# Introduction to Causal Inference & Potential Outcomes

David M. Vock

PubH 7485/8485

# For discussion

- What is causal inference?

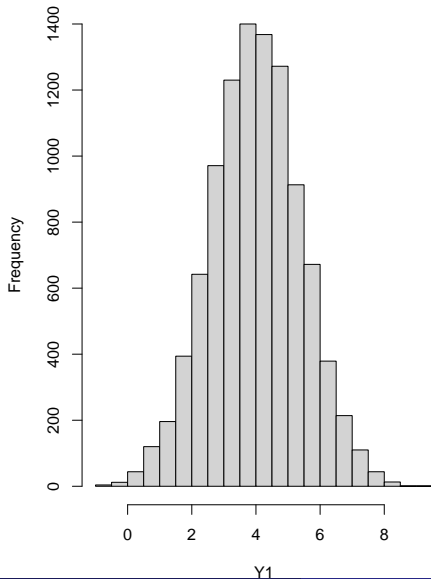
# Potential outcome

- The outcome that would have been observed had possibly contrary to fact an individual a specific intervention/treatment/action
- Denote potential outcomes using a superscript notation; i.e.,  $Y^a$  is the value of the variable  $Y$  for a randomly selected individual if s/he received intervention  $a$
- Can think of potential outcomes as inherent characteristics of an individual
- This notation implicitly assumes that the treatment is well defined and that outcomes do not depend on the treatment others received (more on this shortly)

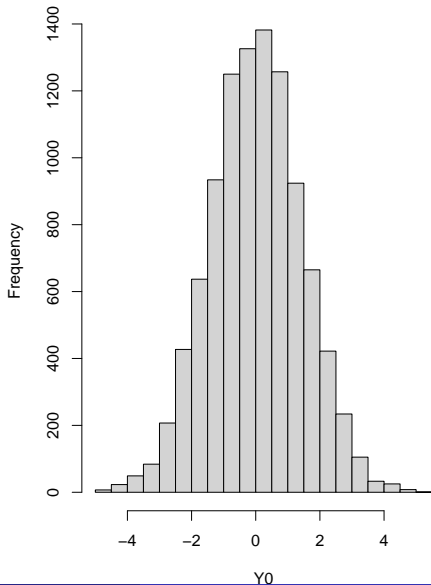
- Causal inference concerns learning about (e.g., estimating, making inferences about) the distribution of potential outcomes and summary measures of those distributions using observed data.
- Next slide has the distribution if everyone in the population were to receive treatment 1 (i.e., distribution of  $Y^1$ ) and distribution if everyone in the population were to receive treatment 0 (i.e., distribution of  $Y^0$ )

# Example of Distribution of Potential Outcomes

Histogram of Y1



Histogram of Y0



# Summary Measures

- As is the case with any analysis, we need some way of summarizing these distributions
- Common summary measures from “associational” analyses also apply here

# Key Population Causal Estimands for Binary Treatment

- (population) average treatment effect (ATE):

$$E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$$

- (population) average treatment effect among treated (ATT):

$$E(Y^1 - Y^0|A = 1) = E(Y^1|A = 1) - E(Y^0|A = 1), \text{ where } A \text{ is treatment received (1 = trt, 0 = control)}$$

- (population) conditional average treatment effect (CATE):

$$E(Y^1 - Y^0|X) = E(Y^1|X) - E(Y^0|X)$$

# Other Population Causal Estimands for Binary Treatment

- For binary outcome data: relative risk or odds ratio

$$P(Y^1 = 1)/P(Y^0 = 1) \text{ or } \frac{P(Y^1=1)}{P(Y^1=0)} / \frac{P(Y^0=1)}{P(Y^0=0)}$$

- For survival data: survival difference at  $t$

$$P(Y^1 > t) - P(Y^0 > t)$$



# Observed Data Notation

- Do not observe potential outcomes on individuals

For a randomly selected individual

- $Y$  - observed response/outcome
- $A$  - treatment received; for now we will assume that  $A$  has only two levels: 1 (treated group) and 0 (control group)
- $X$  - (vector of) covariates

# Consistency and individual causal effect

- Individual causal effect for individual  $i$ :  $Y_i^1 - Y_i^0$
- Consistency assumption -  $Y = Y^A$  or another way of writing this is if treatment is binary  $Y = Y^1A + Y^0(1 - A)$ . In other words we observe the potential outcome for the treatment actually received
- FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE: only observe one potential outcome on each subject and without ridiculously strong assumptions the individual causal effect is not identified
- Causal inference is inherently a missing data problem

# Interference and SUTVA

- Interference: The distribution of the potential outcome under treatment  $a$  depends on the treatment others receive. Examples?
- In the presence of interference, the counterfactual outcome  $Y_i^a$  for an individual  $i$  is not well defined because an individual's outcome depends also on other's treatment assignment.
- Another implicit assumption in our definition of a individual's counterfactual outcome under treatment  $a$  is that there is only one version of treatment  $A = a$
- Stable-unit-treatment-value assumption contains both of these assumptions

- Want to estimate a “causal effect” of  $A$
- Defined the causal treatment effect for individual  $i$  as  $Y_i^1 - Y_i^0$
- As we will establish, there is no possibility of measuring the subject-specific causal treatment effect for any individual  $i$ .
- However, it may be possible, under certain assumptions to estimate a population-level causal treatment effect

# Remark on Causality

- The definition of causal treatment effect may not get at the heart of the mechanism of causality, especially if, the intervention  $A = a$  may trigger a series of events prior to the response.
- However (for the time being), we will not be considering factors that may occur between intervention and response since we are considering point exposure studies. We are only interested in the end result however that may occur.
- Still, there is some appeal to the notion of “what would be the difference in response for the same individual subject to two different interventions”
- Later in the course we will study time-dependent interventions where we will consider the effect that a sequence of interventions given over time will have on the response of interest (including causal mediation)

# Example Dataset

##		Y0	Y1	A	X
## 1		-0.7	2.4	0	-0.2
## 2		-0.5	3.5	1	0.2
## 3		-0.9	4.0	0	-0.3
## 4		-1.8	3.3	1	-0.5
## 5		-1.4	2.7	0	-0.2
## 6		0.6	5.0	1	0.3
## 7		0.8	2.2	0	-0.2
## 8		-3.6	0.9	1	-2.7
## 9		-0.1	2.9	0	0.3
## 10		0.5	4.8	1	0.6

- NOTE: In the real world we would not observe  $\{Y^0, Y^1\}$
- Inherent missing data problem

# Example Dataset: What we really get to see

##		Y0	Y1	Y	A	X
## 1		-0.7	NA	-0.7	0	-0.2
## 2		NA	3.5	3.5	1	0.2
## 3		-0.9	NA	-0.9	0	-0.3
## 4		NA	3.3	3.3	1	-0.5
## 5		-1.4	NA	-1.4	0	-0.2
## 6		NA	5.0	5.0	1	0.3
## 7		0.8	NA	0.8	0	-0.2
## 8		NA	0.9	0.9	1	-2.7
## 9		-0.1	NA	-0.1	0	0.3
## 10		NA	4.8	4.8	1	0.6



# Fundamental Problem of Causal Inference

- Causal models are statistical models for potential outcomes
- E.g., we might assume that  $Y^0 \sim N(\mu, \sigma^2)$  and  $Y^1 \sim N(\mu + \delta, \sigma^2)$  so that the average treatment effect is  $\delta$
- We do not observe  $\{Y^0, Y^1\}$  but instead observe  $(X, A, Y)$
- Key question: can we use the observed data to learn about causal models/parameters?
- Key answer: yes but not without some assumptions

- Typically assume that we have sampled data randomly from some population of interest (maybe a somewhat contrived population - “all individuals with abnormal heart rate who would have enrolled in a Medtronic study”)
- LOTS of statistical tools to estimate parameters and corresponding uncertainty
- For example, if I take a sample of size 10 from population of interest, we could estimate population mean, construct a confidence interval, etc.

```
## [1] "Sample of Y1"
```

```
## [1] 2.4 3.5 4.0 3.3 2.7 5.0 2.2 0.9 2.9 4.8
```

- Sometimes we have missing data
- If the missingness is completely random, then it is as if we just took a smaller sample from the population

```
## [1] "Sample of Y1"
```

```
## [1] NA 3.5 NA 3.3 NA 5.0 NA 0.9 NA 4.8
```

```
## [1] "Observed Data Indicator"
```

```
## [1] 0 1 0 1 0 1 0 1 0 1
```

- Note: observed data indicator is independent of response  $Y^1$

# Randomization

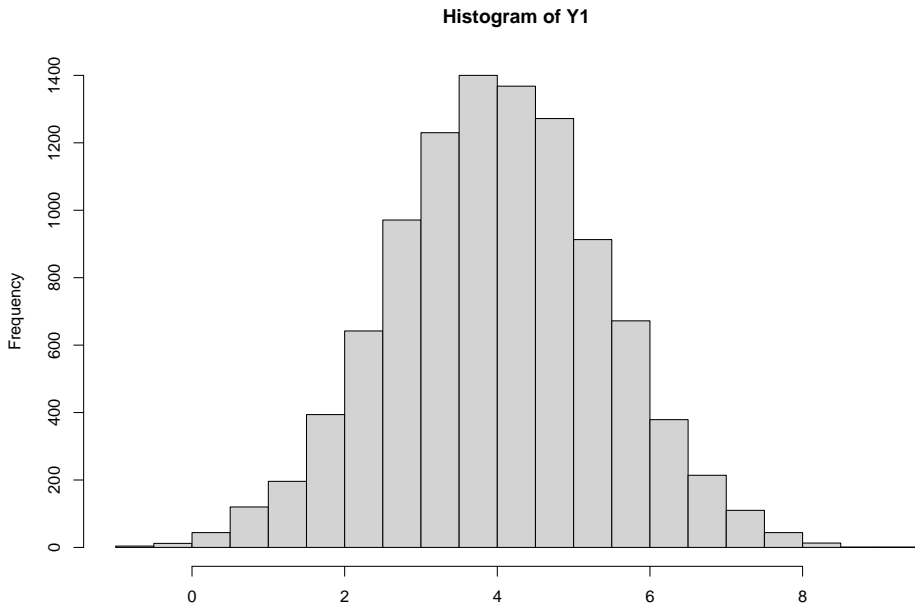
- Intuitively, it has been accepted that the use of a randomized will result in an unbiased estimator of causal treatment effect
- Patients are assigned to treatment interventions according to a random mechanism that is independent of all other factors
- Therefore, individuals randomized to say two treatment groups are similar, on average, with respect to all characteristics except for the treatment intervention they are assigned to
- Consequently, we feel confidence that differences in response between two randomized groups can be reasonably attributed to (assigned) treatment and not other factors prior to treatment assignment
- Can we develop this more formally

# Potential outcomes as baseline covariate

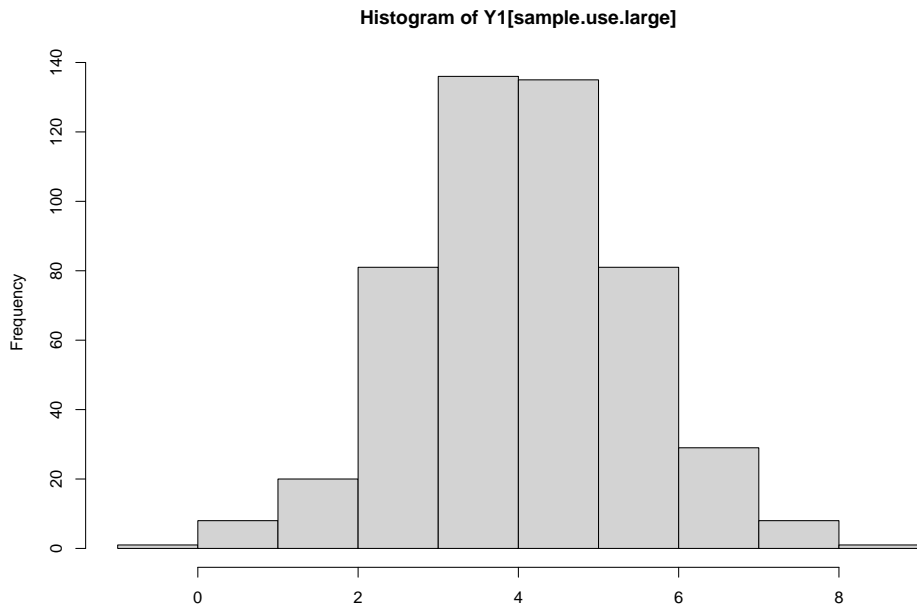
- If  $X$  is a baseline covariate, then randomization guarantees that  $A$  is independent of  $X$
- Remember, it is helpful to think of the potential outcomes  $\{Y_i^1, Y_i^0\}$  as inherent characteristics of patient  $i$  before any treatment is assigned to them
- That is, in some sense, these potential outcomes can be viewed as baseline characteristics of patient  $i$  that are unobserved
- Thus, for a randomized trial, it is reasonable to assume that  $\{Y_i^1, Y_i^0\}$  is independent of  $A_i$

- The assumption that treatment assignment is independent of potential outcomes should not be confused with treatment assignment being independent of the observed response
- That is,  $A_i$  independent of  $\{Y_i^1, Y_i^0\}$  does NOT imply  $A_i$  independent of  $Y_i = A_i Y_i^1 + (1 - A_i) Y_i^0$
- The observed response  $Y_i$  is a consequence of the assigned treatment  $A_i$  whereas the potential outcomes  $\{Y_i^1, Y_i^0\}$  are inherent characteristics of the individual which are not affected by the assigned treatment

# Original Population



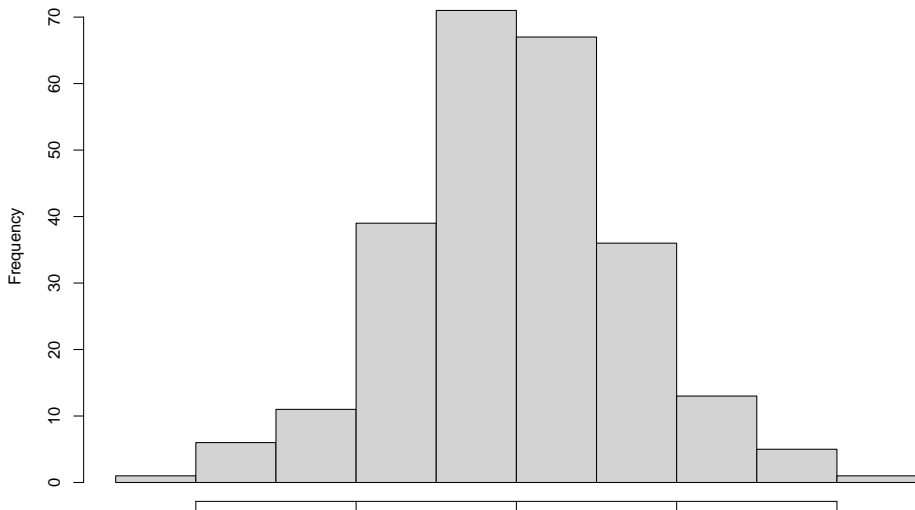
# Identify 500 Participants for Study





# Randomly Assign 250 to Receive Treatment (only observe $Y^1$ on these participants)

Histogram of Y1[sample.use.large.A1]



# Natural estimators

- Assume we have a sample of data from a randomized study  $(Y_i, A_i), i = 1, \dots, n$
- Consider the difference in the treatment-specific sample averages  $\bar{Y}_1 - \bar{Y}_0$  where  $\bar{Y}_1 = \sum_{i=1}^n A_i Y_i / \sum_{i=1}^n A_i = \sum_{i=1}^n A_i Y_i^1 / \sum_{i=1}^n A_i$  and  $\bar{Y}_0 = \sum_{i=1}^n (1 - A_i) Y_i / \sum_{i=1}^n (1 - A_i) = \sum_{i=1}^n (1 - A_i) Y_i^0 / \sum_{i=1}^n (1 - A_i)$
- This is an unbiased, consistent, and asymptotically normal estimator of the ATE

- Challenge occurs when missingness is not random

```
## [1] 2.4 NA NA NA 2.7 NA 2.2 0.9 2.9 NA
```

- Now sample is no longer representative of population of interest. Standard estimation and inference techniques do not hold