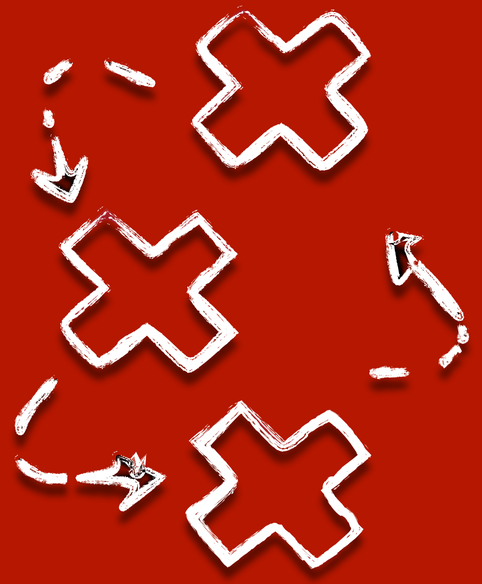


Causal Data Science

Lecture 7:2 Estimating causal effects

Lecturer: Sara Magliacane

UvA - Spring 2022

Estimands for binary treatments

- We generally cannot estimate **unit-level causal effect:** $Y_i(t=1) - Y_i(t=0)$
- We can estimate the average causal effect/**average treatment effect**

$$\text{ATE} = \mathbb{E}[Y(t=1) - Y(t=0)] = \mathbb{E}[Y|\text{do}(T=1)] - \mathbb{E}[Y|\text{do}(T=0)]$$

Estimands for binary treatments

- We generally cannot estimate **unit-level causal effect:** $Y_i(t=1) - Y_i(t=0)$
- We can estimate the average causal effect/**average treatment effect**

$$\text{ATE} = \mathbb{E}[Y(t=1) - Y(t=0)] = \mathbb{E}[Y \mid \text{do}(T=1)] - \mathbb{E}[Y \mid \text{do}(T=0)]$$

- We can estimate the **average causal effect of treatment on the treated**:

$$\text{ATT} = \mathbb{E}[Y(t=1) - Y(t=0) \mid T=1]$$

Estimands for binary treatments

- We generally cannot estimate **unit-level causal effect:** $Y_i(t=1) - Y_i(t=0)$
- We can estimate the average causal effect/**average treatment effect**

$$\text{ATE} = \mathbb{E}[Y(t=1) - Y(t=0)] = \mathbb{E}[Y \,|\, \text{do}(T=1)] - \mathbb{E}[Y \,|\, \text{do}(T=0)]$$

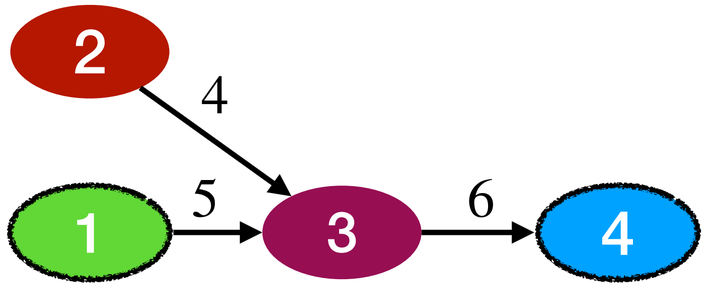
- We can estimate the **average causal effect of treatment on the treated**:

$$\text{ATT} = \mathbb{E}[Y(t=1) - Y(t=0) \,|\, T=1]$$

- We assume that our covariates $\mathbf{X}$ form a valid adjustment set (e.g. we can check them/filter them with backdoor criterion)

Average causal effect/average treatment effect (ATE)

- $\text{ATE} = \mathbb{E}[Y(t=1) - Y(t=0)] = \mathbb{E}[Y|\text{do}(T=1)] - \mathbb{E}[Y|\text{do}(T=0)]$



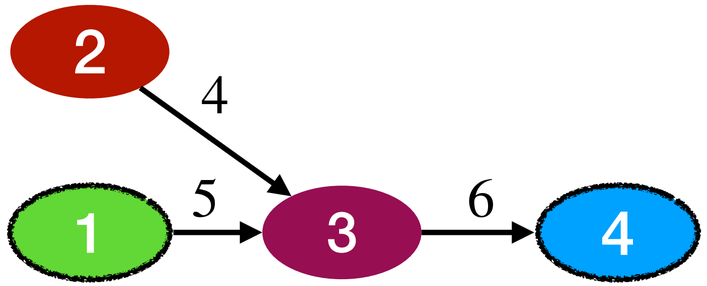

```
x2_1 = randn(n_samples)
x1_1 =  1
x3_1 = 5 * x1_1 + 4 * x2_1 + randn(n_samples)
x4_1 = 6 * x3_1 + randn(n_samples)

x2_0 = randn(n_samples)
x1_0 =  0
x3_0 = 5 * x1_0 + 4 * x2_0 + randn(n_samples)
x4_0 = 6 * x3_0 + randn(n_samples)
diff = np.mean(x4_1) - np.mean(x4_0)
print(diff)
```

```
30.514748479180785
```

Average causal effect/average treatment effect (ATE)

- $\text{ATE} = \mathbb{E}[Y(t=1) - Y(t=0)] = \mathbb{E}[Y \mid \text{do}(T=1)] - \mathbb{E}[Y \mid \text{do}(T=0)]$




```
x2_1 = randn(n_samples)
x1_1 =  1
x3_1 = 5 * x1_1 + 4 * x2_1 + randn(n_samples)
x4_1 = 6 * x3_1 + randn(n_samples)

x2_0 = randn(n_samples)
x1_0 =  0
x3_0 = 5 * x1_0 + 4 * x2_0 + randn(n_samples)
x4_0 = 6 * x3_0 + randn(n_samples)
diff = np.mean(x4_1) - np.mean(x4_0)
print(diff)
```

```
30.514748479180785
```

- How well does the treatment work on the patients who choose it?

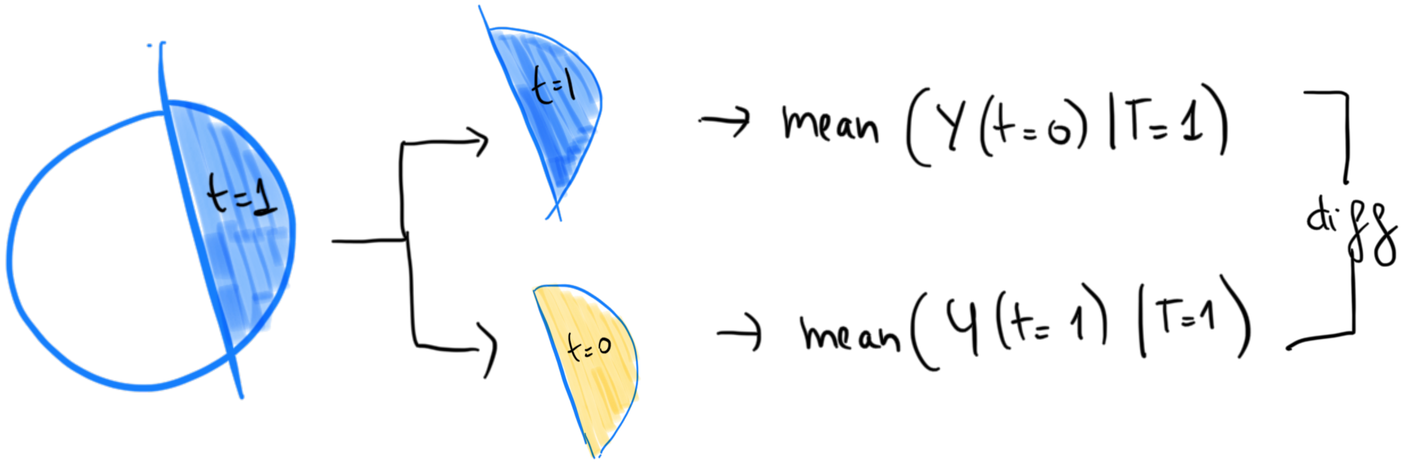
$$\text{ATT} = \mathbb{E}[Y(t=1) - Y(t=0) \mid T=1]$$

cannot write in do() notation

Average causal effect of treatment on the treated (ATT)

- How well does the treatment work on the patients who choose it?

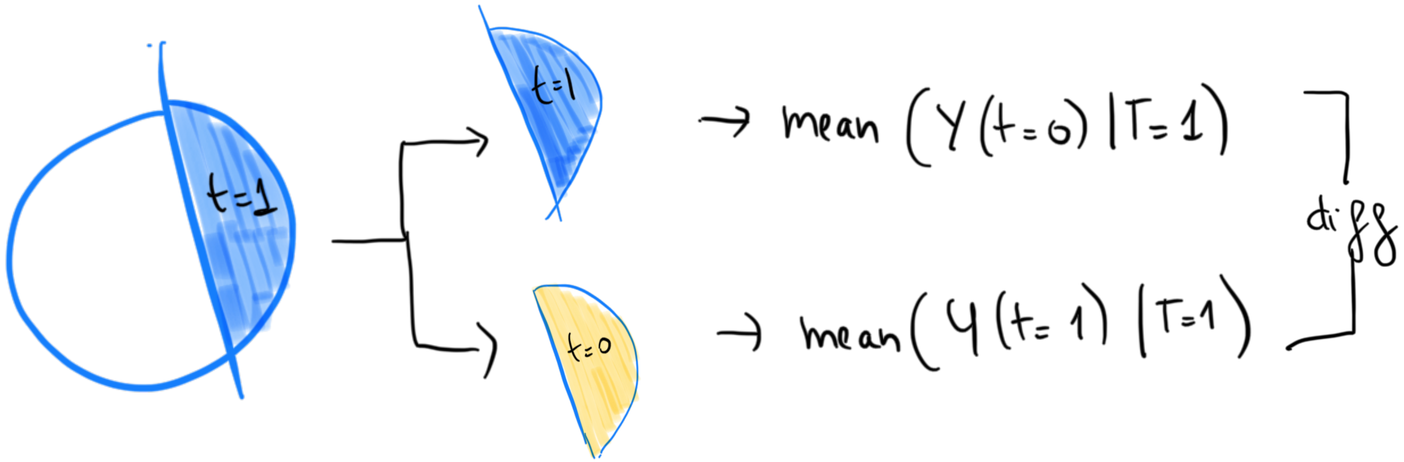
$$\text{ATT} = \mathbb{E}[Y(t=1) - Y(t=0) \,|\, T=1]$$



Average causal effect of treatment on the treated (ATT)

- How well does the treatment work on the patients who choose it?

$$\text{ATT} = \mathbb{E}[Y(t=1) - Y(t=0) \,|\, T=1]$$



- **Not the same as ATE:** For example, people who choose a treatment could be more health-conscious, which means they get anyway better outcomes

Estimation method: Matching

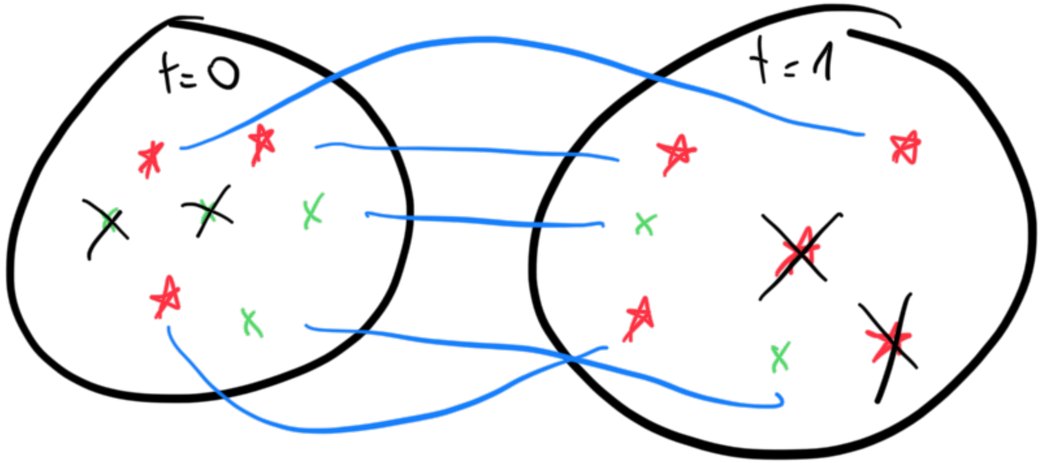
- **Usually for ATT,** sometimes for ATE
- **Intution:** find the most similar couple of patients in terms of covariates $\mathbf{X}$, such that one is in the treatment and the other in the control group

Estimation method: Matching

- **Usually for ATT,** sometimes for ATE
- **Intution:** find the most similar couple of patients in terms of covariates $\mathbf{X}$, such that one is in the treatment and the other in the control group
  - If successful, it's like an RCT
  - For example: I want to compare the outcomes of other people of my age

Estimation method: Matching

- **Usually for ATT,** sometimes for ATE
- **Intution:** find the most similar couple of patients in terms of covariates $\mathbf{X}$, such that one is in the treatment and the other in the control group
  - If successful, it's like an RCT
  - For example: I want to compare the outcomes of other people of my age



$$P(A|T=0) = P(A|T=1)$$

Matching - continuous covariates, greedy/optimal

- If exact matching on the value is not possible, e.g. because we have continuous covariates, we can use any **distance**, e.g Mahalanobis distance

Matching - continuous covariates, greedy/optimal

- If exact matching on the value is not possible, e.g. because we have continuous covariates, we can use any **distance**, e.g Mahalanobis distance

- Many variants exist, in general two types of algorithms:
  - **Greedy matching:** greedily and incrementally match treated with control based on distance
  - **Optimal matching:** optimize for the smallest total distance, can be slow

Matching - continuous covariates, greedy/optimal

- If exact matching on the value is not possible, e.g. because we have continuous covariates, we can use any **distance**, e.g Mahalanobis distance

- Many variants exist, in general two types of algorithms:
  - **Greedy matching:** greedily and incrementally match treated with control based on distance
  - **Optimal matching:** optimize for the smallest total distance, can be slow

- Need to check **covariate balancing** after matching (e.g. std mean difference)

Estimation method: Propensity score matching

- **Assumptions**: binary treatment T, $\mathbf{X}$ is valid adjustment set
- **Propensity score:** the probability of getting assigned the treatment

$$\pi := P(T = 1 \mid \mathbf{X} = x)$$

- We then do **matching on propensity scores**

Estimation method: Propensity score matching

- **Assumptions**: binary treatment T, $\mathbf{X}$ is valid adjustment set
- **Propensity score:** the probability of getting assigned the treatment

$$\pi := P(T = 1 \mid \mathbf{X} = x)$$

- We then do **matching on propensity scores**

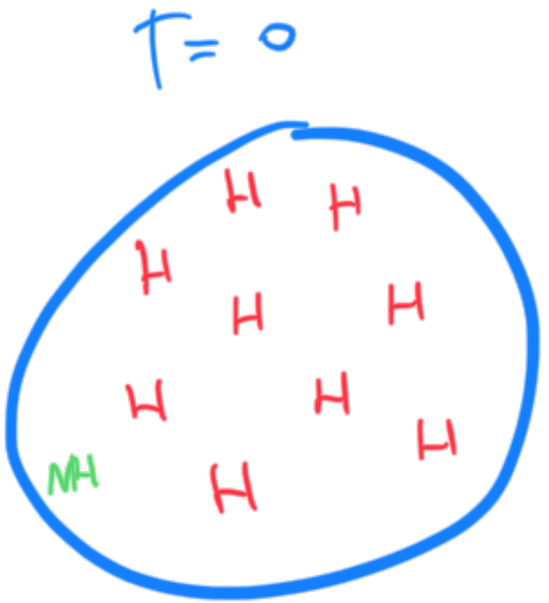
- π encodes all information of $\mathbf{X}$ that is useful for T, i.e. $T \perp\!\!\!\perp \mathbf{X} \mid \pi$
  - If $\mathbf{X}$ has a lot of covariates, it might easier to match for since it's a number
  - π is estimated from data, e.g. with **logistic regression**

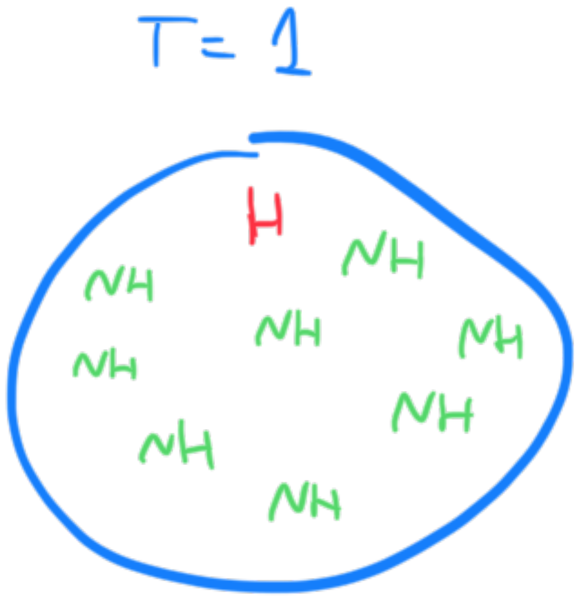
Inverse probability weighting - example

Propensity scores:

$$P(V=1 \mid H=1) = 0.1$$

$$P(V=1 \mid H=0) = 0.9$$

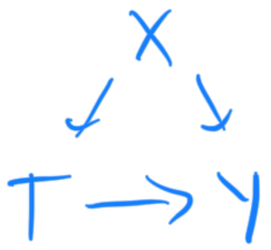




Estimation method: Inverse probability weighting (IPW)

- **Idea:** rather than match, reweight (downweight or upweight) observations
- **Inverse probability (of treatment) weighting:** weight by inverse of probability of treatment **received**:
  - For treated $T = 1$: weight by the inverse of $\pi = P(T = 1 \mid \mathbf{X})$
  - For untreated $T = 0$: weight by the inverse of $1 - \pi = P(T = 0 \mid \mathbf{X})$

IPW Example



$P(T=1 \mid X=1) = 0.1$

$P(T=1 \mid X=0) = 0.8$

$P(T=0 \mid X=1) = 0.9$

$P(T=0 \mid X=0) = 0.2$

Population:

	X=0	X=1
T=0	1	9
T=1	4	1

	X=0	X=1
T=0	1/0.2	9/0.9
T=1	4/0.8	1/0.1

pseudo-population

	X=0	X=1
T=0	5	10
T=1	5	10

Estimation method: Inverse probability weighting (IPW)

- **Inverse probability (of treatment) weighting:** weight by inverse of probability of treatment **received**:
  - For treated $T = 1$: weight by the inverse of $\pi = P(T = 1 \mid \mathbf{X})$
  - For untreated $T = 0$: weight by the inverse of $1 - \pi = P(T = 0 \mid \mathbf{X})$

$$\hat{\mathbb{E}}(Y(t = 1)) = \frac{1}{n} \sum_{i=1}^{n} Y_i \cdot 1\{T = 1\} \cdot \frac{1}{P(T = 1 \mid X_i)}$$

$$\hat{\mathbb{E}}(Y(t = 0)) = \frac{1}{n} \sum_{i=1}^{n} Y_i \cdot 1\{T = 0\} \cdot \frac{1}{P(T = 0 \mid X_i)}$$

Questions?