Causal Inference (overview)

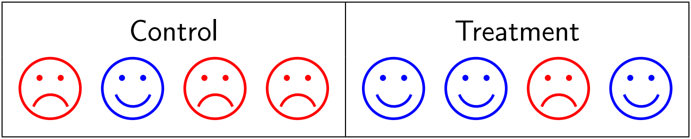
Probability and Statistics for Data Science

Carlos Fernandez-Granda

Key question

Does an observed *statistical* effect imply a causal effect?

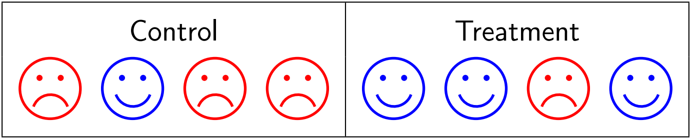
New drug



Control group recovery rate: 25%

Treatment group recovery rate: **75%**

New drug



Control group recovery rate: 25%

Treatment group recovery rate: **75%**

Is a new patient more likely to recover if treated?

3-point shooting

Stephen Curry: 41.7%

Courtney Lee: **43.9%**

3-point shooting

Stephen Curry: 41.7%

Courtney Lee: **43.9%**

Was Lee a better shooter?

Grades in a Portuguese school

With private classes, average grade: **10.94 / 20**

Without private classes, average grade: 9.98 / 20

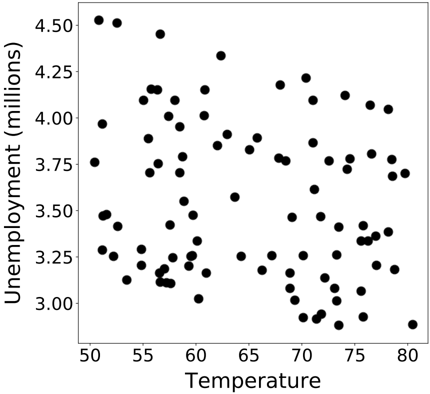
Grades in a Portuguese school

With private classes, average grade: **10.94 / 20**

Without private classes, average grade: 9.98 / 20

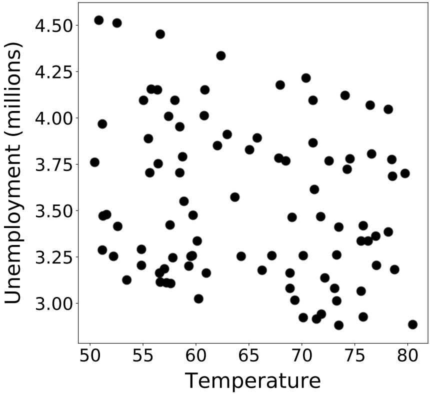
Are the classes useful?

Unemployment and temperature in Spain (2015-2022)



Correlation coefficient: -0.21

Unemployment and temperature in Spain (2015-2022)



Correlation coefficient: -0.21

Would an increase in temperature decrease unemployment?

Plan

Potential outcomes

Confounding factors

Adjusting for confounders

Potential outcomes

Confounding factors

Adjusting for confounders

Probabilistic modeling

Probabilistic modeling

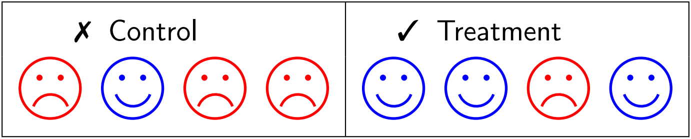
Outcome: $\tilde{y}$

Probabilistic modeling

Outcome: $\tilde{y}$

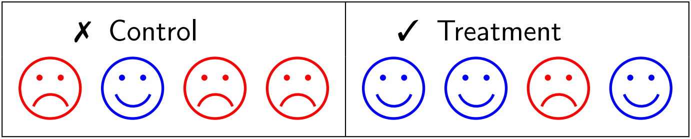
Treatment: $\tilde{t}$

New drug



New drug

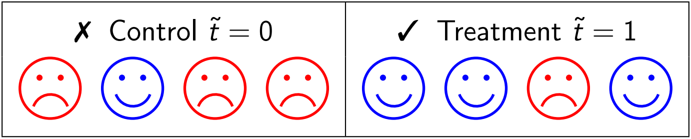
Outcome $\tilde{y}$: If patient recovered $\tilde{y} = 1$, if not $\tilde{y} = 0$



New drug

Outcome $\tilde{y}$: If patient recovered $\tilde{y} = 1$, if not $\tilde{y} = 0$

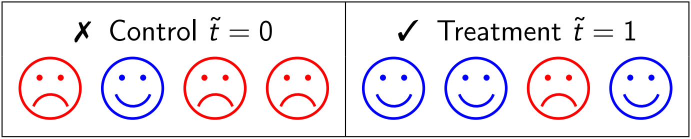
Treatment $\tilde{t}$: If patient received drug $\tilde{t} = 1$, if not $\tilde{t} = 0$



New drug

Outcome $\tilde{y}$: If patient recovered $\tilde{y} = 1$, if not $\tilde{y} = 0$

Treatment $\tilde{t}$: If patient received drug $\tilde{t} = 1$, if not $\tilde{t} = 0$



$$0.25 = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) < p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) = 0.75$$

3-point shooting

3-point shooting

Outcome $\tilde{y}$: If shot goes in $\tilde{y} = 1$, if not $\tilde{y} = 0$

3-point shooting

Outcome $\tilde{y}$: If shot goes in $\tilde{y} = 1$, if not $\tilde{y} = 0$

Treatment $\tilde{t}$: Player who shoots

3-point shooting

Outcome $\tilde{y}$: If shot goes in $\tilde{y} = 1$, if not $\tilde{y} = 0$

Treatment $\tilde{t}$: Player who shoots

$$P\left(\tilde{y} = 1 \,|\, \tilde{t} = \text{Curry}\right) = 0.417$$

$$P\left(\tilde{y} = 1 \,|\, \tilde{t} = \text{Lee}\right) = 0.439$$

Grades in a Portuguese school

Grades in a Portuguese school

Outcome $\tilde{y}$: Grades

Grades in a Portuguese school

Outcome $\tilde{y}$: Grades

Treatment $\tilde{t}$: If student receives private classes $\tilde{t} = 1$, if not $\tilde{t} = 0$

Grades in a Portuguese school

Outcome $\tilde{y}$: Grades

Treatment $\tilde{t}$: If student receives private classes $\tilde{t} = 1$, if not $\tilde{t} = 0$

Outcome is not binary, so we use *conditional mean* to compute average treatment effect (ATE)

Grades in a Portuguese school

Outcome $\tilde{y}$: Grades

Treatment $\tilde{t}$: If student receives private classes $\tilde{t} = 1$, if not $\tilde{t} = 0$

Outcome is not binary, so we use *conditional mean* to compute average treatment effect (ATE)

$$\mu_{\tilde{y} \,|\, \tilde{t}}(1) = 10.94 \qquad \mu_{\tilde{y} \,|\, \tilde{t}}(0) = 9.98$$

Grades in a Portuguese school

Outcome $\tilde{y}$: Grades

Treatment $\tilde{t}$: If student receives private classes $\tilde{t} = 1$, if not $\tilde{t} = 0$

Outcome is not binary, so we use *conditional mean* to compute average treatment effect (ATE)

$$\mu_{\tilde{y} \mid \tilde{t}}(1) = 10.94 \qquad \mu_{\tilde{y} \mid \tilde{t}}(0) = 9.98$$

$$\text{observed ATE} := \mu_{\tilde{y} \mid \tilde{t}}(1) - \mu_{\tilde{y} \mid \tilde{t}}(0) = 0.96$$

Unemployment in Spain

Unemployment in Spain

Outcome $\tilde{y}$: Unemployment

Unemployment in Spain

Outcome $\tilde{y}$: Unemployment

Treatment $\tilde{t}$: Temperature

Unemployment in Spain

Outcome $\tilde{y}$: Unemployment

Treatment $\tilde{t}$: Temperature

Neither is binary, so we use focus on linear effect and evaluate *correlation*: $\rho_{\tilde{y},\tilde{t}} = -0.21$

Unemployment in Spain

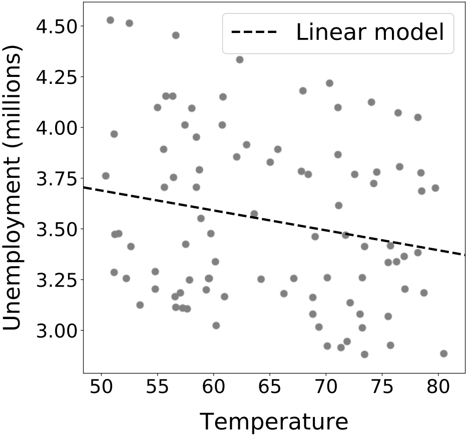
Outcome $\tilde{y}$: Unemployment

Treatment $\tilde{t}$: Temperature

Neither is binary, so we use focus on linear effect and evaluate *correlation*: $\rho_{\tilde{y},\tilde{t}} = -0.21$

Equivalently, *linear model* of outcome given treatment

Linear coefficient: -0.01



Potential outcomes

Alternative scenarios tied to different values of the treatment

Potential outcomes

Alternative scenarios tied to different values of the treatment

They are always defined regardless of the value of the treatment

New drug

$\widetilde{\mathsf{po}}_0$: Outcome if patient is untreated

$\widetilde{\mathsf{po}}_0 = 1$

$\widetilde{\mathsf{po}}_0 = 0$

New drug

$\widetilde{\mathrm{po}}_0$: Outcome if patient is untreated

$\widetilde{\mathrm{po}}_0 = 1$

$\widetilde{\mathrm{po}}_0 = 0$

$\widetilde{\mathrm{po}}_1$: Outcome if patient is treated

$\widetilde{\mathrm{po}}_1 = 1$

$\widetilde{\mathrm{po}}_1 = 0$

New drug

$\widetilde{\text{po}}_0$: Outcome if patient is untreated

$\widetilde{\text{po}}_0 = 1$

$\widetilde{\text{po}}_0 = 0$

$\widetilde{\text{po}}_1$: Outcome if patient is treated

$\widetilde{\text{po}}_1 = 1$

$\widetilde{\text{po}}_1 = 0$

Causal effect: $\mathrm{P}\left(\widetilde{\text{po}}_0 = 1\right)$ compared to $\mathrm{P}\left(\widetilde{\text{po}}_1 = 1\right)$

3-point shooting

$\widetilde{\text{po}}_{\text{Curry}}$: Outcome if Curry shoots

$\widetilde{\text{po}}_{\text{Curry}} = 1$ shot made, $\widetilde{\text{po}}_{\text{Curry}} = 0$ shot missed

3-point shooting

$\widetilde{\text{po}}_{\text{Curry}}$: Outcome if Curry shoots

$\widetilde{\text{po}}_{\text{Curry}} = 1$ shot made, $\widetilde{\text{po}}_{\text{Curry}} = 0$ shot missed

$\widetilde{\text{po}}_{\text{Lee}}$: Outcome if Lee shoots

$\widetilde{\text{po}}_{\text{Lee}} = 1$ shot made, $\widetilde{\text{po}}_{\text{Lee}} = 0$ shot missed

3-point shooting

$\widetilde{\text{po}}_{\text{Curry}}$: Outcome if Curry shoots

$\widetilde{\text{po}}_{\text{Curry}} = 1$ shot made, $\widetilde{\text{po}}_{\text{Curry}} = 0$ shot missed

$\widetilde{\text{po}}_{\text{Lee}}$: Outcome if Lee shoots

$\widetilde{\text{po}}_{\text{Lee}} = 1$ shot made, $\widetilde{\text{po}}_{\text{Lee}} = 0$ shot missed

Causal effect: $\mathrm{P}\left(\widetilde{\text{po}}_{\text{Curry}} = 1\right)$ compared to $\mathrm{P}\left(\widetilde{\text{po}}_{\text{Lee}} = 1\right)$

Grades in a Portuguese school

$\widetilde{\text{po}}_0$: Grade without private classes

Grades in a Portuguese school

$\widetilde{\text{po}}_0$: Grade without private classes

$\widetilde{\text{po}}_1$: Grade with private classes

Grades in a Portuguese school

$\widetilde{\text{po}}_0$: Grade without private classes

$\widetilde{\text{po}}_1$: Grade with private classes

Causal ATE $= \mathrm{E}\left[\widetilde{\text{po}}_1\right] - \mathrm{E}\left[\widetilde{\text{po}}_0\right]$

Unemployment in Spain

$\widetilde{\text{po}}_t$: Unemployment when temperature is t

Unemployment in Spain

$\widetilde{\mathrm{po}}_t$: Unemployment when temperature is t

Defined regardless of temperature value

Unemployment in Spain

$\widetilde{\text{po}}_t$: Unemployment when temperature is t

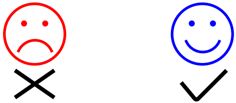
Defined regardless of temperature value

Causal linear effect: $\mathrm{E}\left[\widetilde{\text{po}}_t\right] = \beta_{\text{treat}} t$

Fundamental problem of causal inference

Fundamental problem of causal inference

All potential outcomes are defined regardless of treatment value



Fundamental problem of causal inference

All potential outcomes are defined regardless of treatment value

But we only see the one tied to the observed treatment!

Fundamental problem of causal inference

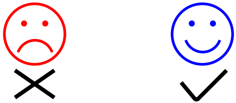
All potential outcomes are defined regardless of treatment value

But we only see the one tied to the observed treatment!

If $\tilde{t} = 1$

Fundamental problem of causal inference

All potential outcomes are defined regardless of treatment value



But we only see the one tied to the observed treatment!

If $\tilde{t} = 1$

All other potential outcomes are unobserved counterfactuals

Key question

When do observed statistics reflect causal effects?

Key question

When do observed statistics reflect causal effects?

- Binary outcomes / treatment

$$\mathrm{P}\left(\widetilde{\mathrm{po}}_0 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) \quad ?$$

$$\mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) \quad ?$$

Key question

When do observed statistics reflect causal effects?

- Binary outcomes / treatment

$$\mathrm{P}\left(\widetilde{\mathrm{po}}_0 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) \quad ?$$

$$\mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) \quad ?$$

- Nonbinary outcomes and binary treatment

$$\mathrm{E}\left[\widetilde{\mathrm{po}}_0\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(0) \quad ?$$

$$\mathrm{E}\left[\widetilde{\mathrm{po}}_1\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(1) \quad ?$$

Key question

When do observed statistics reflect causal effects?

- Binary outcomes / treatment

$$\mathrm{P}\left(\widetilde{\mathsf{po}}_0 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0) \quad ?$$
$$\mathrm{P}\left(\widetilde{\mathsf{po}}_1 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1) \quad ?$$

- Nonbinary outcomes and binary treatment

$$\mathrm{E}\left[\widetilde{\mathsf{po}}_0\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(0) \quad ?$$
$$\mathrm{E}\left[\widetilde{\mathsf{po}}_1\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(1) \quad ?$$

- Nonbinary outcomes / treatment, such that $\mathrm{E}\left[\widetilde{\mathsf{po}}_t\right] = \beta_{\mathsf{treat}}\, t$

$$\beta_{\mathsf{treat}} = \mathrm{Cov}\left[\tilde{y}, \tilde{t}\,\right] = \beta_{\mathsf{MMSE}} \quad ?$$

Key question

When do observed statistics reflect causal effects?

- Binary outcomes / treatment

$$\mathrm{P}\left(\widetilde{\mathsf{po}}_0 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0)$$
$$\mathrm{P}\left(\widetilde{\mathsf{po}}_1 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1)$$

- Nonbinary outcomes and binary treatment

$$\mathrm{E}\left[\widetilde{\mathsf{po}}_0\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(0)$$
$$\mathrm{E}\left[\widetilde{\mathsf{po}}_1\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(1)$$

- Nonbinary outcomes / treatment, such that $\mathrm{E}\left[\widetilde{\mathsf{po}}_t\right] = \beta_{\mathsf{treat}}\, t$

$$\beta_{\mathsf{treat}} = \mathrm{Cov}\left[\tilde{y}, \tilde{t}\,\right] = \beta_{\mathsf{MMSE}}$$

When treatment and potential outcomes are independent

Key question

When do observed statistics reflect causal effects?

- Binary outcomes / treatment

$$\mathrm{P}\left(\widetilde{\mathsf{po}}_0 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,0)$$
$$\mathrm{P}\left(\widetilde{\mathsf{po}}_1 = 1\right) = p_{\tilde{y}\,|\,\tilde{t}}(1\,|\,1)$$

- Nonbinary outcomes and binary treatment

$$\mathrm{E}\left[\widetilde{\mathsf{po}}_0\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(0)$$
$$\mathrm{E}\left[\widetilde{\mathsf{po}}_1\right] = \mu_{\tilde{y}\,|\,\tilde{t}}(1)$$

- Nonbinary outcomes / treatment, such that $\mathrm{E}\left[\widetilde{\mathsf{po}}_t\right] = \beta_{\mathsf{treat}}\, t$

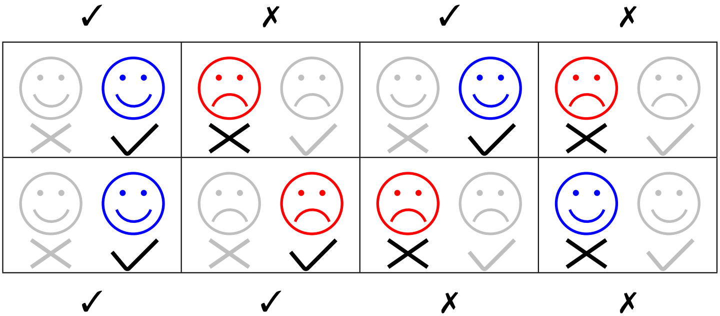
$$\beta_{\mathsf{treat}} = \mathrm{Cov}\left[\tilde{y}, \tilde{t}\,\right] = \beta_{\mathsf{MMSE}}$$

When treatment and potential outcomes are independent

Can be ensured by randomizing the treatment

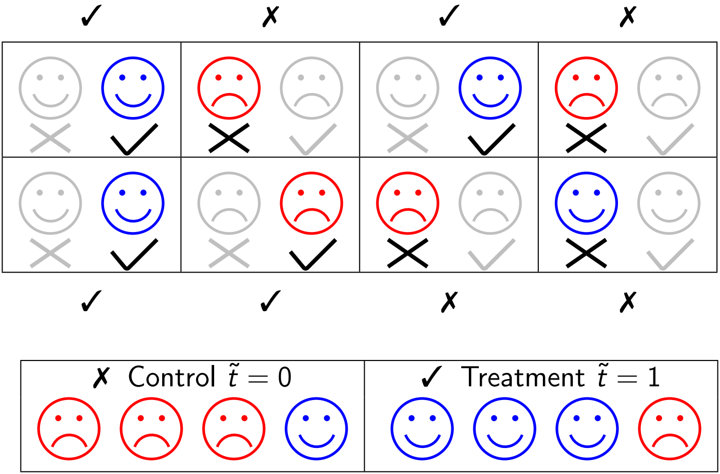
Independence is essential

Avoids systematic differences between control and treatment groups

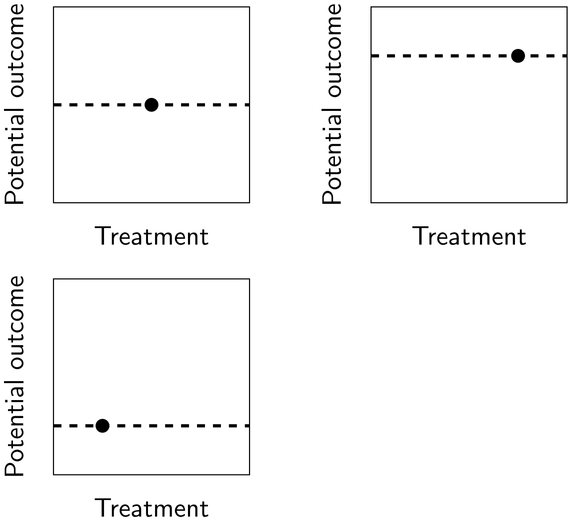


Independence is essential

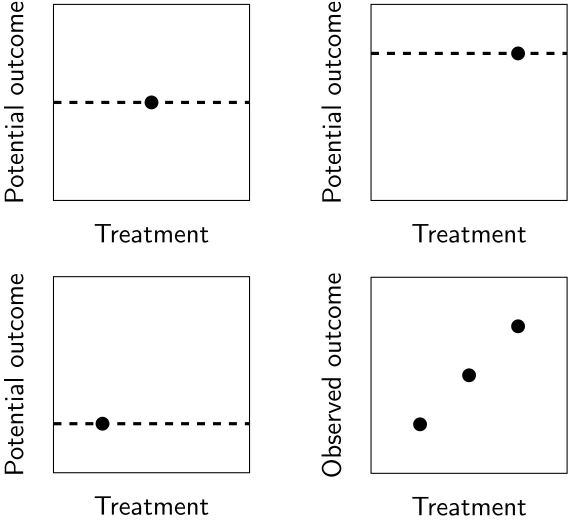
Avoids systematic differences between control and treatment groups



Independence is essential



Independence is essential



Potential outcomes

Confounding factors

Adjusting for confounders

Confounding factor (a.k.a. confounder)

Confounding factor (a.k.a. confounder)

Induces dependence between treatment and potential outcomes

Confounding factor (a.k.a. confounder)

Induces dependence between treatment and potential outcomes

Assumption: Conditional independence of treatment and potential outcomes given confounder

Confounding factor (a.k.a. confounder)

Induces dependence between treatment and potential outcomes

Assumption: Conditional independence of treatment and potential outcomes given confounder

Confounder completely governs dependence

Binary outcome / treatment

Confounder $\tilde{c}$ with C values

$$\mathrm{P}\left(\tilde{y} = 1 \,|\, \tilde{t} = 1\right)$$

Binary outcome / treatment

Confounder $\tilde{c}$ with C values

$$\begin{aligned}
&\mathrm{P}\left(\tilde{y} = 1 \,|\, \tilde{t} = 1\right) \\
&= \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c \,|\, \tilde{t} = 1) \mathrm{P}(\widetilde{\mathsf{po}}_1 = 1 \,|\, \tilde{c} = c)
\end{aligned}$$

Binary outcome / treatment

Confounder $\tilde{c}$ with C values

$$
\begin{aligned}
&\mathrm{P}\left(\tilde{y} = 1 \,|\, \tilde{t} = 1\right) \\
&= \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c \,|\, \tilde{t} = 1)\mathrm{P}(\widetilde{\mathrm{po}}_1 = 1 \,|\, \tilde{c} = c) \\
&\neq \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c)\mathrm{P}(\widetilde{\mathrm{po}}_1 = 1 \,|\, \tilde{c} = c)
\end{aligned}
$$

Binary outcome / treatment

Confounder $\tilde{c}$ with C values

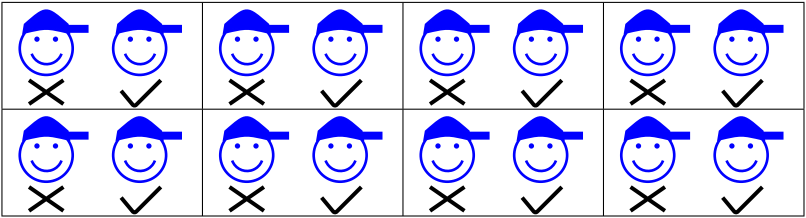
$$
\begin{aligned}
&\mathrm{P}\left(\tilde{y}=1 \mid \tilde{t}=1\right) \\
&= \sum_{c=1}^{C} \mathrm{P}(\tilde{c}=c \mid \tilde{t}=1)\mathrm{P}(\widetilde{\mathrm{po}}_1=1 \mid \tilde{c}=c) \\
&\neq \sum_{c=1}^{C} \mathrm{P}(\tilde{c}=c)\mathrm{P}(\widetilde{\mathrm{po}}_1=1 \mid \tilde{c}=c) \\
&= \mathrm{P}\left(\widetilde{\mathrm{po}}_1=1\right)
\end{aligned}
$$

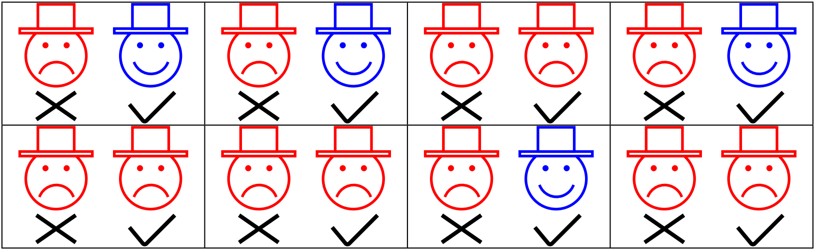
Binary outcome / treatment

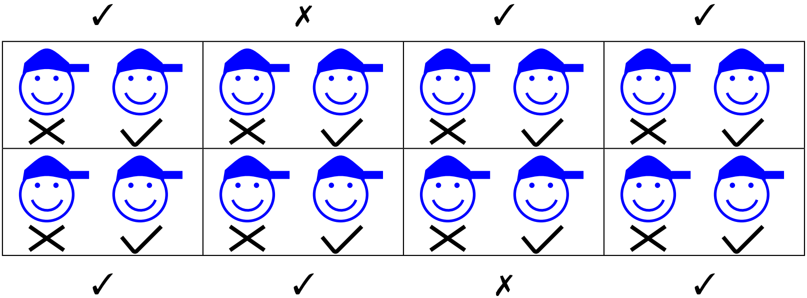
Confounder $\tilde{c}$ with C values

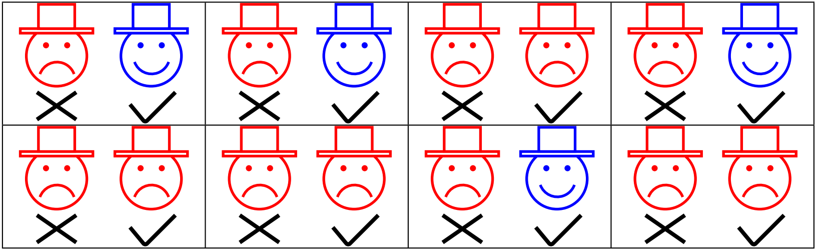
$$
\begin{aligned}
& \mathrm{P}\left(\tilde{y}=1 \,|\, \tilde{t}=1\right) \\
& = \sum_{c=1}^{C} \mathrm{P}(\tilde{c}=c \,|\, \tilde{t}=1) \mathrm{P}(\widetilde{\mathrm{po}}_1 = 1 \,|\, \tilde{c}=c) \\
& \neq \sum_{c=1}^{C} \mathrm{P}(\tilde{c}=c) \mathrm{P}(\widetilde{\mathrm{po}}_1 = 1 \,|\, \tilde{c}=c) \\
& = \mathrm{P}\left(\widetilde{\mathrm{po}}_1 = 1\right)
\end{aligned}
$$

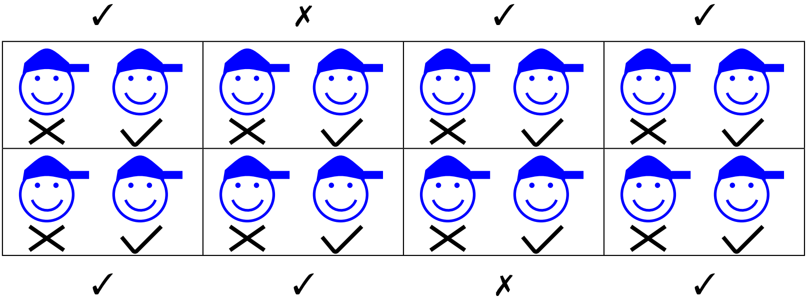
Example: Age as a confounder in new drug evaluation

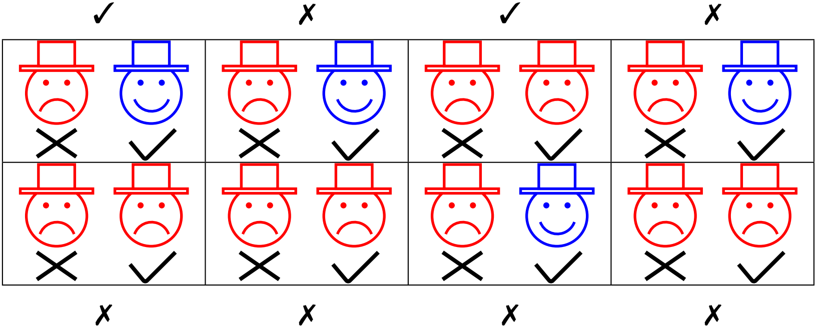


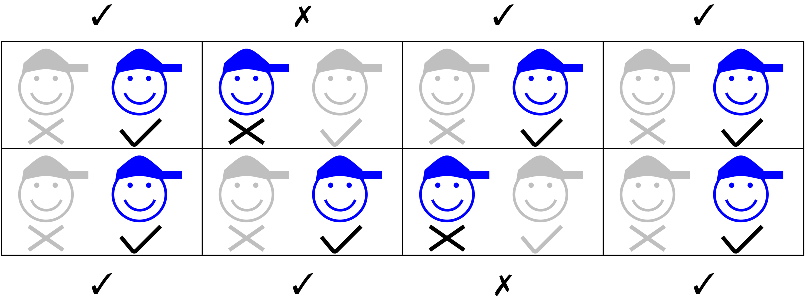


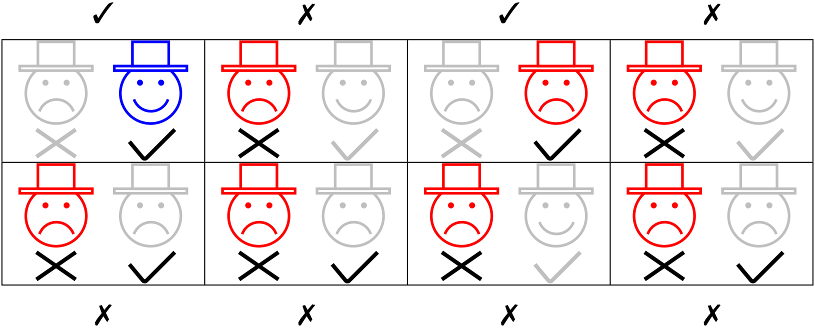


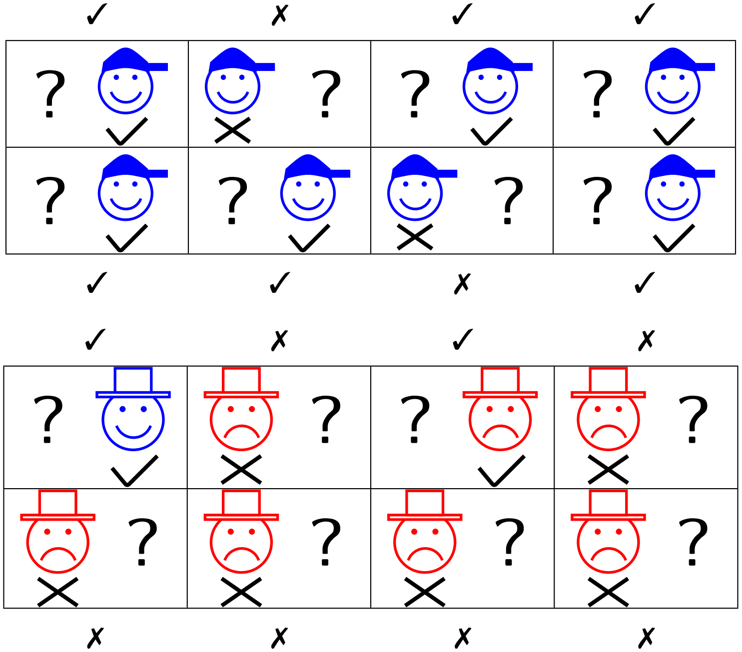








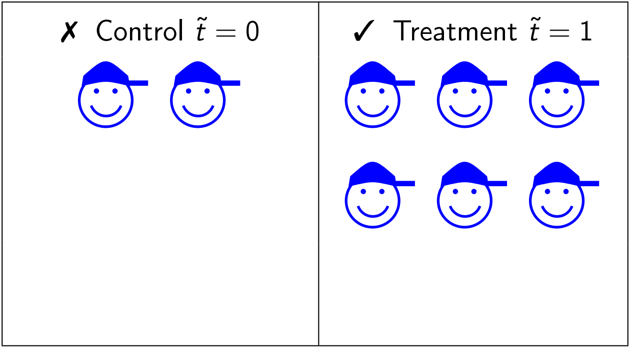




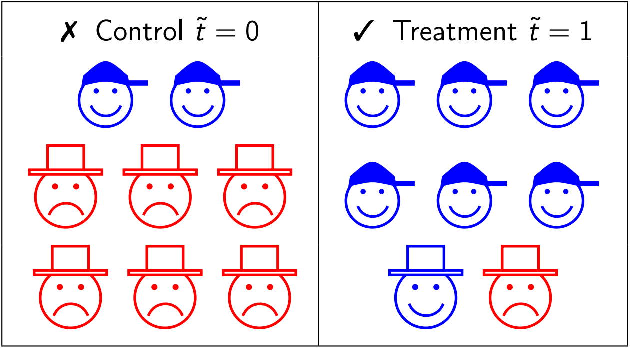
Data

✗ Control $\tilde{t} = 0$	✓ Treatment $\tilde{t} = 1$

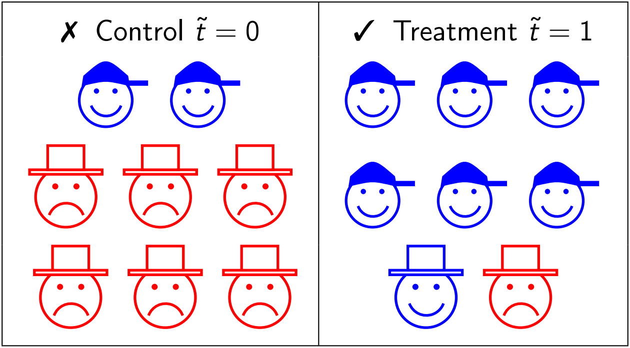
Data



Data



Data



$$0.25 = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0\right) < \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1\right) = 0.875$$

3-point shooting

Confounder?

	Stephen Curry	Courtney Lee
Total	$190/456 = 41.7\%$	$75/171 = \mathbf{43.9\%}$

3-point shooting

Confounder? Shot distance $\tilde{d}$

	Stephen Curry	Courtney Lee
Total	$190/456 = 41.7\%$	$75/171 =$ **43.9%**

3-point shooting

Confounder? Shot distance $\tilde{d}$

	Stephen Curry	Courtney Lee
Short threes ($\leq$ 24 feet)	$45/90 = \mathbf{50.0\%}$	$56/116 = 48.3\%$
Total	$190/456 = 41.7\%$	$75/171 = \mathbf{43.9\%}$

3-point shooting

Confounder? Shot distance $\tilde{d}$

	Stephen Curry	Courtney Lee
Short threes ($\leq$ 24 feet)	$45/90 =$ **50.0%**	$56/116 = 48.3\%$
Long threes ($>$ 24 feet)	$145/366 =$ **39.6%**	$19/55 = 34.5\%$
Total	$190/456 = 41.7\%$	$75/171 =$ **43.9%**

3-point shooting

Confounder? Shot distance $\tilde{d}$

	Stephen Curry	Courtney Lee
Short threes ($\leq$ 24 feet)	$45/90 =$ **50.0%**	$56/116 = 48.3\%$
Long threes ($>$ 24 feet)	$145/366 =$ **39.6%**	$19/55 = 34.5\%$
Total	$190/456 = 41.7\%$	$75/171 =$ **43.9%**

Simpson's paradox

3-point shooting

Confounder? Shot distance $\tilde{d}$

	Stephen Curry	Courtney Lee
Short threes ($\leq$ 24 feet)	45/90 = **50.0%**	56/116 = 48.3%
Long threes ($>$ 24 feet)	145/366 = **39.6%**	19/55 = 34.5%
Total	190/456 = 41.7%	75/171 = **43.9%**

Simpson's paradox

$$\mathrm{P}(\tilde{d} = \text{short} \mid \tilde{t} = \text{Curry}) = 0.197$$
$$\mathrm{P}(\tilde{d} = \text{short} \mid \tilde{t} = \text{Lee}) = 0.678$$

Nonbinary outcome binary treatment

Confounder $\tilde{c}$ with C values

$$\mu_{\tilde{y} \mid \tilde{t}}(1)$$

Nonbinary outcome binary treatment

Confounder $\tilde{c}$ with C values

$$\begin{aligned}
&\mu_{\tilde{y} \mid \tilde{t}}(1) \\
&= \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c \mid \tilde{t} = 1) \mu_{\widetilde{\mathbf{po}_1} \mid \tilde{c}}(c)
\end{aligned}$$

Nonbinary outcome binary treatment

Confounder $\tilde{c}$ with C values

$$
\begin{aligned}
&\mu_{\tilde{y} \mid \tilde{t}}(1) \\
&= \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c \mid \tilde{t} = 1) \mu_{\widetilde{\mathbf{po}_1} \mid \tilde{c}}(c) \\
&\neq \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c) \mu_{\widetilde{\mathbf{po}_1} \mid \tilde{c}}(c)
\end{aligned}
$$

Nonbinary outcome binary treatment

Confounder $\tilde{c}$ with C values

$$\begin{aligned}
&\mu_{\tilde{y}\,|\,\tilde{t}}(1) \\
&= \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c \,|\, \tilde{t} = 1)\, \mu_{\widetilde{\mathrm{po}}_1\,|\,\tilde{c}}(c) \\
&\neq \sum_{c=1}^{C} \mathrm{P}(\tilde{c} = c)\, \mu_{\widetilde{\mathrm{po}}_1\,|\,\tilde{c}}(c) \\
&= \mathrm{E}\left[\widetilde{\mathrm{po}}_1\right]
\end{aligned}$$

Grades in a Portuguese school

Confounder: Whether students previously failed the course ($\tilde{c} = 1$) or not ($\tilde{c} = 0$)

Grades in a Portuguese school

Confounder: Whether students previously failed the course ($\tilde{c} = 1$) or not ($\tilde{c} = 0$)

Conditional mean function $\mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}$

Previously failed	Private classes	No private classes
Yes		
No		

Grades in a Portuguese school

Confounder: Whether students previously failed the course ($\tilde{c} = 1$) or not ($\tilde{c} = 0$)

Conditional mean function $\mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}$

Previously failed	Private classes	No private classes
Yes	8.95	6.66
No		

Grades in a Portuguese school

Confounder: Whether students previously failed the course ($\tilde{c} = 1$) or not ($\tilde{c} = 0$)

Conditional mean function $\mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}$

Previously failed	Private classes	No private classes
Yes	8.95	6.66
No	11.20	11.31

Grades in a Portuguese school

Confounder: Whether students previously failed the course ($\tilde{c} = 1$) or not ($\tilde{c} = 0$)

Conditional mean function $\mu_{\tilde{y} \,|\, \tilde{c}, \tilde{t}}$

Previously failed	Private classes	No private classes
Yes	8.95	6.66
No	11.20	11.31

$$\mathrm{P}\left(\text{Previously failed} \,|\, \text{Private classes}\right) = 0.122$$

$$\mathrm{P}\left(\text{Previously failed} \,|\, \text{No private classes}\right) = 0.285$$

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\widetilde{\mathrm{po}}_t := \beta_{\mathrm{treat}} t + \beta_{\mathrm{conf}} \tilde{c} + \tilde{z}$$

$\tilde{z}$ independent from $\tilde{t}$ and $\tilde{c}$

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\widetilde{\text{po}}_t := \beta_{\text{treat}} t + \beta_{\text{conf}} \tilde{c} + \tilde{z}$$

$$\tilde{y} = \beta_{\text{treat}} \tilde{t} + \beta_{\text{conf}} \tilde{c} + \tilde{z}$$

$\tilde{z}$ independent from $\tilde{t}$ and $\tilde{c}$

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\widetilde{\text{po}}_t := \beta_{\text{treat}} t + \beta_{\text{conf}} \tilde{c} + \tilde{z}$$

$$\tilde{y} = \beta_{\text{treat}} \tilde{t} + \beta_{\text{conf}} \tilde{c} + \tilde{z}$$

$\tilde{z}$ independent from $\tilde{t}$ and $\tilde{c}$

"Short" linear regression model of $\tilde{y}$ given $\tilde{t}$

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\widetilde{\mathsf{po}}_t := \beta_{\mathsf{treat}} t + \beta_{\mathsf{conf}} \tilde{c} + \tilde{z}$$

$$\tilde{y} = \beta_{\mathsf{treat}} \tilde{t} + \beta_{\mathsf{conf}} \tilde{c} + \tilde{z}$$

$\tilde{z}$ independent from $\tilde{t}$ and $\tilde{c}$

"Short" linear regression model of $\tilde{y}$ given $\tilde{t}$

$$\beta_{\mathsf{MMSE}} = \mathrm{Cov}\left[\tilde{y}, \tilde{t}\,\right]$$

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\widetilde{\mathsf{po}}_t := \beta_{\mathsf{treat}} t + \beta_{\mathsf{conf}} \tilde{c} + \tilde{z}$$

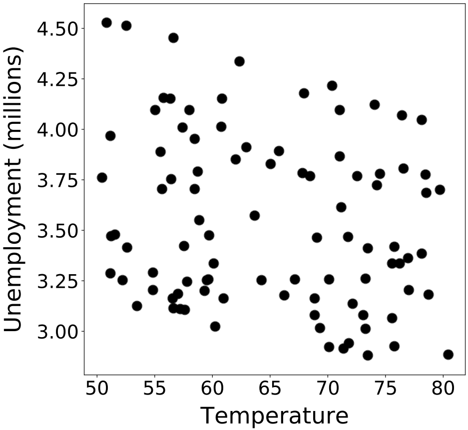
$$\tilde{y} = \beta_{\mathsf{treat}} \tilde{t} + \beta_{\mathsf{conf}} \tilde{c} + \tilde{z}$$

$\tilde{z}$ independent from $\tilde{t}$ and $\tilde{c}$

"Short" linear regression model of $\tilde{y}$ given $\tilde{t}$

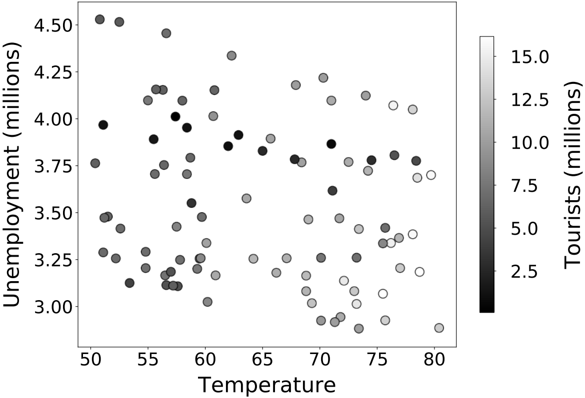
$$\begin{aligned}\beta_{\mathsf{MMSE}} &= \mathrm{Cov}\left[\tilde{y}, \tilde{t}\,\right] \\ &= \beta_{\mathsf{treat}} + \beta_{\mathsf{conf}} \sigma_{\tilde{t},\tilde{c}}\end{aligned}$$

Unemployment and temperature in Spain



Confounder?

Tourists!



4.50
4.25
4.00
3.75
3.50
3.25
3.00
Unemployment (millions)
50
55
60
65
70
75
80
Temperature
15.0
12.5
10.0
7.5
5.0
2.5
Tourists (millions)

Potential outcomes

Confounding factors

Adjusting for confounders

Motivation

Motivation

Randomization neutralizes all confounders, even if they are unknown!

Motivation

Randomization neutralizes all confounders, even if they are unknown!

Problem: Randomization is very costly, and often *not possible*

Motivation

Randomization neutralizes all confounders, even if they are unknown!

Problem: Randomization is very costly, and often *not possible*

Solution: Correct for known confounders

Motivation

Randomization neutralizes all confounders, even if they are unknown!

Problem: Randomization is very costly, and often *not possible*

Solution: Correct for known confounders

Assumption: Treatment is conditionally independent of the potential outcomes given the confounder

Adjusting for a confounder

If treatment is conditionally independent of the potential outcomes given the confounder

Adjusting for a confounder

If treatment is conditionally independent of the potential outcomes given the confounder

(1) Conditional statistics given confounder are OK

Adjusting for a confounder

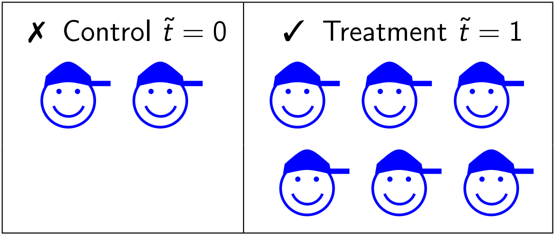
If treatment is conditionally independent of the potential outcomes given the confounder

(1) Conditional statistics given confounder are OK

(2) They can be combined to estimate true causal effect

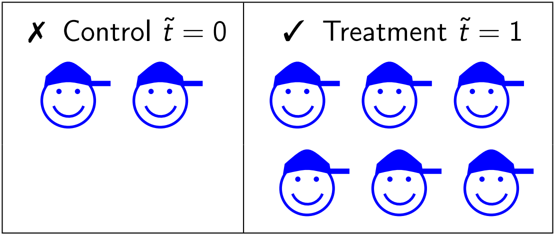
New drug: Young subjects

Confounder: Age $\tilde{a}$



New drug: Young subjects

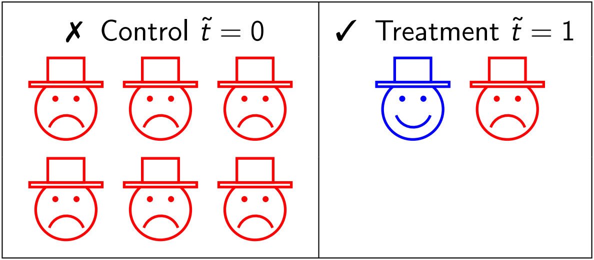
Confounder: Age $\tilde{a}$



$$1 = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = \text{young}\right) = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = \text{young}\right) = 1$$

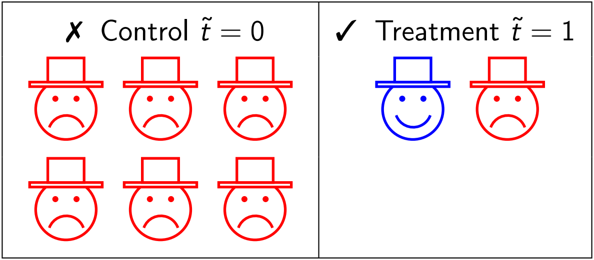
New drug: Old subjects

Confounder: Age $\tilde{a}$



New drug: Old subjects

Confounder: Age $\tilde{a}$



$$0 = \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = \text{old}\right) < \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = \text{old}\right) = 0.5$$

Adjusted causal effect

$$p_{\widetilde{\text{po}_0}}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = a\right)$$

$$p_{\widetilde{\text{po}_1}}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = a\right)$$

Adjusted causal effect

$$p_{\widetilde{\text{po}}_0}(1) = \sum_{a \in \{\text{young}, \text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = a\right)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0 = \mathbf{0.5}$$

$$p_{\widetilde{\text{po}}_1}(1) = \sum_{a \in \{\text{young}, \text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = a\right)$$

Adjusted causal effect

$$p_{\widetilde{\text{po}}_0}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = a\right)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0 = \mathbf{0.5}$$

$$p_{\widetilde{\text{po}}_1}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = a\right)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0.5 = \mathbf{0.75}$$

Adjusted causal effect

$$p_{\widetilde{\mathrm{po}}_0}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = a\right)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0 = \mathbf{0.5}$$

Observed: 0.25

$$p_{\widetilde{\mathrm{po}}_1}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = a\right)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0.5 = \mathbf{0.75}$$

Adjusted causal effect

$$p_{\widetilde{\mathrm{po}}_0}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 0, \tilde{a} = a\right)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0 = \mathbf{0.5}$$

Observed: 0.25

$$p_{\widetilde{\mathrm{po}}_1}(1) = \sum_{a \in \{\text{young},\text{old}\}} p_{\tilde{a}}(a) \mathrm{P}\left(\tilde{y} \mid \tilde{t} = 1, \tilde{a} = a\right)$$

$$= 0.5 \cdot 1 + 0.5 \cdot 0.5 = \mathbf{0.75}$$

Observed: 0.875

3-point shooting

Confounder: Shot distance $\tilde{d}$

$$p_{\widetilde{\text{po}}_{\text{Curry}}}(1) = \sum_{d \in \{\text{short}, \text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d}, \tilde{t}}(1 \mid d, \text{Curry})$$

$$p_{\widetilde{\text{po}}_{\text{Lee}}}(1) = \sum_{d \in \{\text{short}, \text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d}, \tilde{t}}(1 \mid d, \text{Lee})$$

3-point shooting

Confounder: Shot distance $\tilde{d}$

$$p_{\widetilde{\text{po}}_{\text{Curry}}}(1) = \sum_{d \in \{\text{short},\text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d},\tilde{t}}(1 \mid d, \text{Curry})$$
$$= 0.329 \cdot 0.5 + 0.671 \cdot 0.396 = \mathbf{0.430}$$

$$p_{\widetilde{\text{po}}_{\text{Lee}}}(1) = \sum_{d \in \{\text{short},\text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d},\tilde{t}}(1 \mid d, \text{Lee})$$

3-point shooting

Confounder: Shot distance $\tilde{d}$

$$p_{\widetilde{\mathrm{po}}_{\mathrm{Curry}}}(1) = \sum_{d \in \{\text{short},\text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d},\tilde{t}}(1 \mid d, \text{Curry})$$

$$= 0.329 \cdot 0.5 + 0.671 \cdot 0.396 = \mathbf{0.430}$$

$$p_{\widetilde{\mathrm{po}}_{\mathrm{Lee}}}(1) = \sum_{d \in \{\text{short},\text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d},\tilde{t}}(1 \mid d, \text{Lee})$$

$$= 0.329 \cdot 0.483 + 0.671 \cdot 0.345 = \mathbf{0.390}$$

3-point shooting

Confounder: Shot distance $\tilde{d}$

$$p_{\widetilde{\text{po}}_{\text{Curry}}}(1) = \sum_{d \in \{\text{short},\text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y}\,|\,\tilde{d},\tilde{t}}(1\,|\,d, \text{Curry})$$
$$= 0.329 \cdot 0.5 + 0.671 \cdot 0.396 = \mathbf{0.430}$$

Observed: 0.417

$$p_{\widetilde{\text{po}}_{\text{Lee}}}(1) = \sum_{d \in \{\text{short},\text{long}\}} p_{\tilde{d}}(d) p_{\tilde{y}\,|\,\tilde{d},\tilde{t}}(1\,|\,d, \text{Lee})$$
$$= 0.329 \cdot 0.483 + 0.671 \cdot 0.345 = \mathbf{0.390}$$

3-point shooting

Confounder: Shot distance $\tilde{d}$

$$p_{\widetilde{\mathrm{po}}_{\mathrm{Curry}}}(1) = \sum_{d \in \{\mathrm{short}, \mathrm{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d}, \tilde{t}}(1 \mid d, \mathrm{Curry})$$

$$= 0.329 \cdot 0.5 + 0.671 \cdot 0.396 = \mathbf{0.430}$$

Observed: 0.417

$$p_{\widetilde{\mathrm{po}}_{\mathrm{Lee}}}(1) = \sum_{d \in \{\mathrm{short}, \mathrm{long}\}} p_{\tilde{d}}(d) p_{\tilde{y} \mid \tilde{d}, \tilde{t}}(1 \mid d, \mathrm{Lee})$$

$$= 0.329 \cdot 0.483 + 0.671 \cdot 0.345 = \mathbf{0.390}$$

Observed: 0.439

Grades in a Portuguese school

Confounder: Whether student previously failed

$$\mathrm{E}\left[\widetilde{\mathsf{po}}_1\right] = \sum_{c \in \{0,1\}} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,1)$$

$$\mathrm{E}\left[\widetilde{\mathsf{po}}_0\right] = \sum_{c \in C} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,0)$$

Grades in a Portuguese school

Confounder: Whether student previously failed

$$\mathrm{E}\left[\widetilde{\mathsf{po}}_1\right] = \sum_{c \in \{0,1\}} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,1)$$

$$= 0.79 \cdot 11.20 + 0.21 \cdot 8.95 = 10.73$$

$$\mathrm{E}\left[\widetilde{\mathsf{po}}_0\right] = \sum_{c \in C} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,0)$$

Grades in a Portuguese school

Confounder: Whether student previously failed

$$\mathrm{E}\left[\widetilde{\mathrm{po}}_1\right] = \sum_{c \in \{0,1\}} p_{\tilde{c}}(c) \mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}(c, 1)$$

$$= 0.79 \cdot 11.20 + 0.21 \cdot 8.95 = 10.73$$

$$\mathrm{E}\left[\widetilde{\mathrm{po}}_0\right] = \sum_{c \in C} p_{\tilde{c}}(c) \mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}(c, 0)$$

$$= 0.79 \cdot 11.31 + 0.21 \cdot 6.66 = 10.33$$

Grades in a Portuguese school

Confounder: Whether student previously failed

$$\mathrm{E}\left[\widetilde{\mathrm{po}}_1\right] = \sum_{c \in \{0,1\}} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,1)$$

$$= 0.79 \cdot 11.20 + 0.21 \cdot 8.95 = 10.73$$

$$\mathrm{E}\left[\widetilde{\mathrm{po}}_0\right] = \sum_{c \in C} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,0)$$

$$= 0.79 \cdot 11.31 + 0.21 \cdot 6.66 = 10.33$$

$$\text{adjusted ATE} = 10.73 - 10.33 = \mathbf{0.4}$$

Grades in a Portuguese school

Confounder: Whether student previously failed

$$\mathrm{E}\left[\widetilde{\text{po}}_1\right] = \sum_{c \in \{0,1\}} p_{\tilde{c}}(c) \mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}(c, 1)$$

$$= 0.79 \cdot 11.20 + 0.21 \cdot 8.95 = 10.73$$

$$\mathrm{E}\left[\widetilde{\text{po}}_0\right] = \sum_{c \in C} p_{\tilde{c}}(c) \mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}(c, 0)$$

$$= 0.79 \cdot 11.31 + 0.21 \cdot 6.66 = 10.33$$

adjusted ATE $= 10.73 - 10.33 = \mathbf{0.4}$ Observed: 0.93

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\tilde{y} = \beta_{\text{treat}}\tilde{t} + \beta_{\text{conf}}\tilde{c} + \tilde{z}$$

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\tilde{y} = \beta_{\text{treat}} \tilde{t} + \beta_{\text{conf}} \tilde{c} + \tilde{z}$$

We adjust for the confounder including it in the model

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

$$\tilde{y} = \beta_{\text{treat}}\tilde{t} + \beta_{\text{conf}}\tilde{c} + \tilde{z}$$

We adjust for the confounder including it in the model

- Response $\tilde{y}$
- Feature vector $\tilde{x} := \begin{bmatrix} \tilde{t} \\ \tilde{c} \end{bmatrix}$

Nonbinary outcome nonbinary treatment

Assuming linear model given confounder $\tilde{c}$ and treatment $\tilde{t}$ (both standardized)

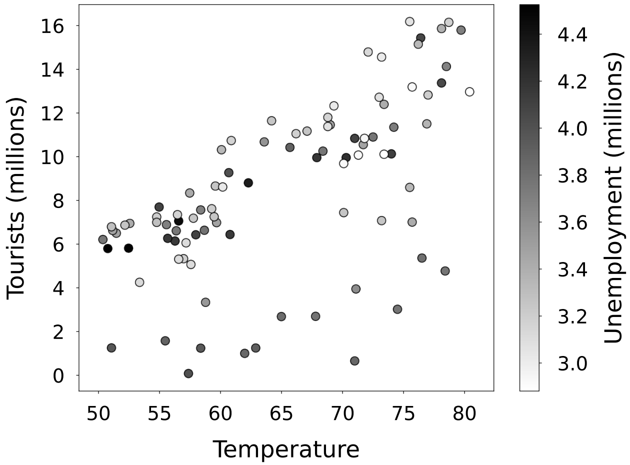
$$\tilde{y} = \beta_{\text{treat}} \tilde{t} + \beta_{\text{conf}} \tilde{c} + \tilde{z}$$

We adjust for the confounder including it in the model

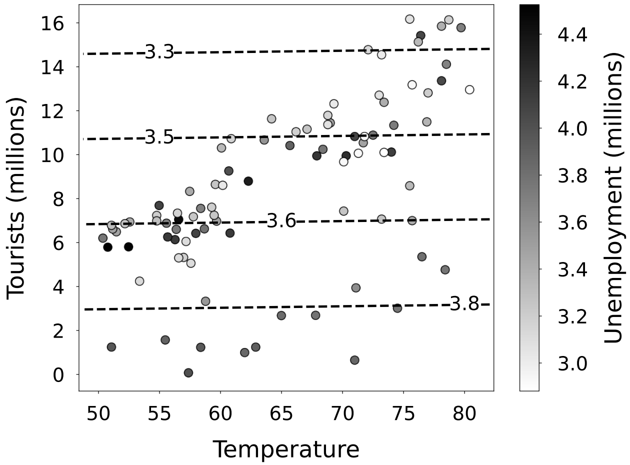
- Response $\tilde{y}$
- Feature vector $\tilde{x} := \begin{bmatrix} \tilde{t} \\ \tilde{c} \end{bmatrix}$

$$\beta_{\text{MMSE}} = \begin{bmatrix} \beta_{\text{treat}} \\ \beta_{\text{conf}} \end{bmatrix}$$

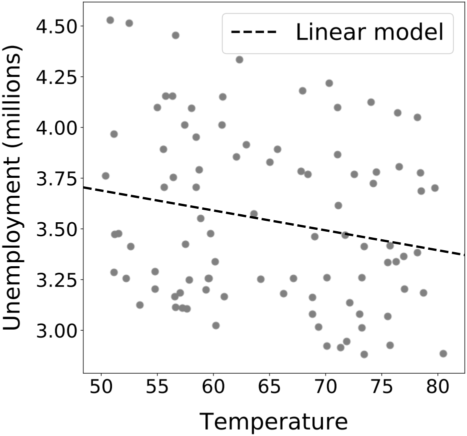
Incorporating the confounder



With confounder, temperature coefficient: 0.0003



Without confounder, temperature coefficient: -0.01



What have we learned?

What have we learned?

Observed statistics do not necessarily reflect causal effects

What have we learned?

Observed statistics do not necessarily reflect causal effects

They do if treatment and potential outcomes are independent, usually due to randomization

What have we learned?

Observed statistics do not necessarily reflect causal effects

They do if treatment and potential outcomes are independent, usually due to randomization

Otherwise confounders can completely distort observed effects

What have we learned?

Observed statistics do not necessarily reflect causal effects

They do if treatment and potential outcomes are independent, usually due to randomization

Otherwise confounders can completely distort observed effects

Under conditional independence assumptions, we can adjust for *known* confounders