Average Treatment Effect

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Goal

Estimate causal effect of a *treatment* from data

All caps titles

Goal: Determine whether all caps titles cause YouTube videos to get more views

Treatment $\tilde{t}$: if title is all caps $\tilde{t} = 1$, if not $\tilde{t} = 0$

Observations: Number of views $\tilde{y}$

Potential outcomes

$\widetilde{\text{po}}_0$: Views if title is proper case

$\widetilde{\text{po}}_1$: Views if title is all caps

Observed data:

$$\tilde{y} := \begin{cases} \widetilde{\text{po}}_0 & \text{if} \quad \tilde{t} = 0 \\ \\ \widetilde{\text{po}}_1 & \text{if} \quad \tilde{t} = 1 \end{cases}$$

Average treatment effect

The average treatment effect is

$$\mathsf{ATE} := \mathrm{E}\left[\widetilde{\mathsf{po}}_1\right] - \mathrm{E}\left[\widetilde{\mathsf{po}}_0\right].$$

Challenge: We do not observe $\widetilde{\mathsf{po}}_0$ and $\widetilde{\mathsf{po}}_1$ directly

Observed data

Treatment $\tilde{t}$	Observed outcome $\tilde{y}$	Outcome if proper case $\widetilde{\text{po}}_0$	Outcome if all caps $\widetilde{\text{po}}_1$
✗	102	102	?
✗	45	45	?
✓	330	?	330
✓	121	?	121
✓	23	?	23

? are counterfactuals

Is $\mu_{\tilde{y}\,|\,\tilde{t}}(1) - \mu_{\tilde{y}\,|\,\tilde{t}}(0)$ a reasonable estimate for the ATE?

Detour: Private classes

Dataset of student grades from school in Portugal

Some students take private classes, but are they useful?

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

Data: Grades $\tilde{y}$

Does the treatment cause the average grade to increase?

Difference in conditional mean

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

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

Case closed?

Possible confounder

We also know whether students previously failed the course ($\tilde{c} = 1$) or not ($\tilde{c} = 0$)

$$\mu_{\tilde{y} \mid \tilde{c},\tilde{t}}(1,1) = 8.95 \qquad \mu_{\tilde{y} \mid \tilde{c},\tilde{t}}(1,0) = 6.66$$

$$\mu_{\tilde{y} \mid \tilde{c},\tilde{t}}(0,1) = 11.20 \qquad \mu_{\tilde{y} \mid \tilde{c},\tilde{t}}(0,0) = 11.31$$

$$p_{\tilde{c} \mid \tilde{t}}(1 \mid 1) = 0.12 \qquad p_{\tilde{c} \mid \tilde{t}}(1 \mid 0) = 0.29$$

Effect of confounder on observed ATE

$$\mu_{\tilde{y} \mid \tilde{t}}(t) = \sum_{c=0}^{1} \int_{y=-\infty}^{\infty} p_{\tilde{c} \mid \tilde{t}}(c \mid t) f_{\tilde{y} \mid \tilde{c}, \tilde{t}}(y \mid c, t) y \, \mathrm{d}y$$

$$= \sum_{c=0}^{1} p_{\tilde{c} \mid \tilde{t}}(c \mid t) \mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}(c, t)$$

Effect of confounder on observed ATE

$$\begin{aligned}
\mu_{\tilde{y}\,|\,\tilde{t}}(1) &= p_{\tilde{c}\,|\,\tilde{t}}(0\,|\,1)\,\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(0,1) + p_{\tilde{c}\,|\,\tilde{t}}(1\,|\,1)\,\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(1,1) \\
&= 0.88 \cdot 11.20 + 0.12 \cdot 8.95 \\
&= \underset{\tilde{c}=0}{\underset{\uparrow}{9.85}} + \underset{\tilde{c}=1}{\underset{\uparrow}{1.09}} = 10.94
\end{aligned}$$

$$\begin{aligned}
\mu_{\tilde{y}\,|\,\tilde{t}}(0) &= p_{\tilde{c}\,|\,\tilde{t}}(0\,|\,0)\,\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(0,0) + p_{\tilde{c}\,|\,\tilde{t}}(1\,|\,0)\,\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(1,0) \\
&= 0.71 \cdot 11.31 + 0.29 \cdot 6.66 \\
&= \underset{\tilde{c}=0}{\underset{\uparrow}{8.08}} + \underset{\tilde{c}=1}{\underset{\uparrow}{1.90}} = 9.98
\end{aligned}$$

How can we avoid effect of confounders?

Randomizing treatment!

This renders treatment $\tilde{t}$ independent to potential outcomes $\widetilde{\text{po}}_0$ and $\widetilde{\text{po}}_1$

If treatment is randomized

$$\begin{aligned}
\mu_{\tilde{y} \mid \tilde{t}}(1) &= \mu_{\widetilde{\mathsf{po}}_1 \mid \tilde{t}}(1) \\
&= \int_x x f_{\widetilde{\mathsf{po}}_1 \mid \tilde{t}}(x \mid 1) \, \mathrm{d}x \\
&= \int_x x f_{\widetilde{\mathsf{po}}_1}(x) \, \mathrm{d}x \\
&= \mathrm{E}\left[\widetilde{\mathsf{po}}_1\right]
\end{aligned}$$

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

$$\mathsf{ATE} = \mu_{\tilde{y} \mid \tilde{t}}(1) - \mu_{\tilde{y} \mid \tilde{t}}(0)$$

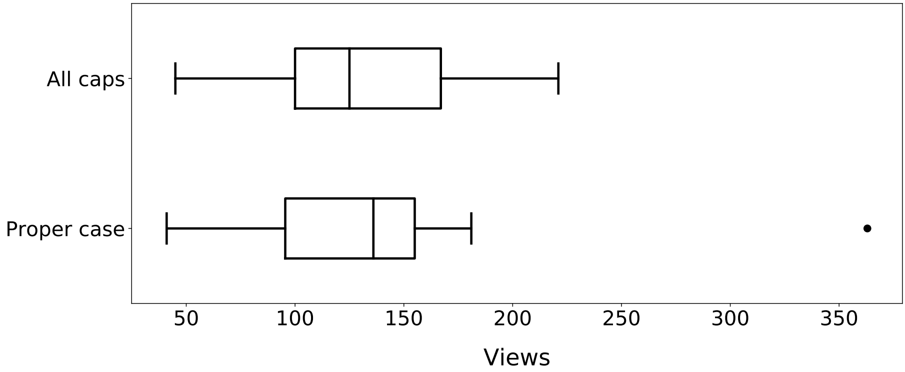
YouTube videos

All caps: 19

No all caps: 26

$$\begin{aligned} \text{ATE} &= \mu_{\tilde{y}\,|\,\tilde{t}}(1) - \mu_{\tilde{y}\,|\,\tilde{t}}(0) \\ &= 133 - 132 \approx 0 \end{aligned}$$

YouTube videos



What if we cannot randomize?

$$\begin{aligned} \mathrm{E}\left[\widetilde{\mathsf{po}}_1\right] &= \mathrm{E}\left[\mu_{\widetilde{\mathsf{po}}_1 \,|\, \tilde{c}}(\tilde{c})\right] \\ &= \sum_{c \in C} p_{\tilde{c}}(c) \mu_{\widetilde{\mathsf{po}}_1 \,|\, \tilde{c}}(c) \end{aligned}$$

Do we know $p_{\tilde{c}}$?

Do we know $\mu_{\widetilde{\mathsf{po}}_0 \,|\, \tilde{c}}$ and $\mu_{\widetilde{\mathsf{po}}_1 \,|\, \tilde{c}}$?

$\mu_{\widetilde{\text{po}}_0 \mid \tilde{c}}$ and $\mu_{\widetilde{\text{po}}_1 \mid \tilde{c}}$

Assumption: $\widetilde{\text{po}}_0$ and $\tilde{t}$ are conditionally independent given $\tilde{c}$

$$\begin{aligned}
\mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}(c, 0) &= \mu_{\widetilde{\text{po}}_0 \mid \tilde{c}, \tilde{t}}(c, 0) \\
&= \int_x x f_{\widetilde{\text{po}}_0 \mid \tilde{c}, \tilde{t}}(x \mid c, 0) \, \mathrm{d}x \\
&= \int_x x f_{\widetilde{\text{po}}_0 \mid \tilde{c}}(x \mid c) \, \mathrm{d}x \\
&= \mu_{\widetilde{\text{po}}_0 \mid \tilde{c}}(c)
\end{aligned}$$

$$\mu_{\tilde{y} \mid \tilde{c}, \tilde{t}}(c, 1) = \mu_{\widetilde{\text{po}}_1 \mid \tilde{c}}(c)$$

Adjusting for a confounding factor

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

$$= \sum_{c \in C} p_{\tilde{c}}(c) \mu_{\tilde{y} \,|\, \tilde{c}, \tilde{t}}(c, 1)$$

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

$$\text{ATE} = \sum_{c \in C} p_{\tilde{c}}(c) \mu_{\tilde{y} \,|\, \tilde{c}, \tilde{t}}(c, 1) - \sum_{c \in C} p_{\tilde{c}}(c) \mu_{\tilde{y} \,|\, \tilde{c}, \tilde{t}}(c, 0)$$

Private classes

$$
\begin{aligned}
\text{adjusted ATE} &= \sum_{c=0}^{1} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,1) - \sum_{c=0}^{1} p_{\tilde{c}}(c)\mu_{\tilde{y}\,|\,\tilde{c},\tilde{t}}(c,0) \\
&= (0.79 \cdot 11.20 + 0.21 \cdot 8.95) - (0.79 \cdot 11.31 + 0.21 \cdot 6.66) \\
&= 0.39 < 0.93
\end{aligned}
$$

Are our assumptions correct? No

Is this a better measure of the effect of the private classes? Yes

What have we learned

Confounding factors can completely distort the average treatment effect

Randomization neutralizes confounders

We can adjust for known confounders under conditional independence assumptions