Was Courtney Lee Better Than Stephen Curry?

Simpson's paradox

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

Data

NBA games from the 2014/2015 season

3-point shot percentage

Stephen Curry: 41.7%

Courtney Lee: **43.9%**

Was Lee the better shooter?

A closer look at the data

	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

Causal inference perspective

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

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

From observed data

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

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

What this means

$$P(\tilde{y} = 1 \mid \tilde{t} = \text{Curry}) = 0.417$$

$$P(\tilde{y} = 1 \mid \tilde{t} = \text{Lee}) = 0.439$$

Curry makes 41.7% of his 3-point shots

Lee makes 43.9% of his 3-point shots

What this does not mean

$$P(\tilde{y} = 1 \mid \tilde{t} = \text{Curry}) = 0.417$$

$$P(\tilde{y} = 1 \mid \tilde{t} = \text{Lee}) = 0.439$$

If Lee had taken Curry's shots, he would have made 43.9% instead of 41.7%

Potential outcomes

$\widetilde{\text{po}}_{\text{Curry}}$: 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

What we actually observe:

$$\tilde{y} := \begin{cases} \widetilde{\text{po}}_{\text{Curry}} & \text{if} \quad \tilde{t} = \text{Curry} \\ \\ \widetilde{\text{po}}_{\text{Lee}} & \text{if} \quad \tilde{t} = \text{Lee} \end{cases}$$

Was Lee the better shooter?

$$\mathrm{P}\left(\widetilde{\mathsf{po}}_{\mathsf{Lee}} = 1\right) > \mathrm{P}\left(\widetilde{\mathsf{po}}_{\mathsf{Curry}} = 1\right)?$$

Challenge: We cannot observe them simultaneously!

Observed data

Treatment $\tilde{t}$	Observed outcome $\tilde{y}$	Outcome if Curry $\widetilde{\text{po}}_{\text{Curry}}$	Outcome if Lee $\widetilde{\text{po}}_{\text{Lee}}$
Curry	☹	☹	?
Curry	☺	☺	?
Lee	☺	?	☺
Lee	☹	?	☹
Lee	☺	?	☺

? are counterfactuals

Can we trust the data?

$$\mathrm{P}\left(\widetilde{\mathsf{po}}_{\mathsf{Curry}} = 1\right) \stackrel{?}{=} \mathrm{P}\left(\tilde{y} = 1 \mid \tilde{t} = \mathsf{Curry}\right) = \mathrm{P}\left(\widetilde{\mathsf{po}}_{\mathsf{Curry}} = 1 \mid \tilde{t} = \mathsf{Curry}\right)$$

$$\mathrm{P}\left(\widetilde{\mathsf{po}}_{\mathsf{Lee}} = 1\right) \stackrel{?}{=} \mathrm{P}\left(\tilde{y} = 1 \mid \tilde{t} = \mathsf{Lee}\right) = \mathrm{P}\left(\widetilde{\mathsf{po}}_{\mathsf{Lee}} = 1 \mid \tilde{t} = \mathsf{Lee}\right)$$

True if $\tilde{t}$ and $\widetilde{\mathsf{po}}_{\mathsf{Curry}}$ / $\widetilde{\mathsf{po}}_{\mathsf{Lee}}$ are independent

Are they independent?

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)	$145/366 = 39.6\%$	$19/55 = 34.5\%$

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

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

$$\mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \text{short}, \tilde{t} = \text{Curry}) = 0.5$$

$$\mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \text{short}, \tilde{t} = \text{Lee}) = 0.483$$

$$\mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \text{long}, \tilde{t} = \text{Curry}) = 0.396$$

$$\mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \text{long}, \tilde{t} = \text{Lee}) = 0.345$$

Confounding factor

$\widetilde{\text{po}}_{\text{Curry}}$ / $\widetilde{\text{po}}_{\text{Lee}}$ depend on distance $\tilde{d}$

Distance $\tilde{d}$ depends on $\tilde{t}$

$\widetilde{\text{po}}_{\text{Curry}}$ / $\widetilde{\text{po}}_{\text{Lee}}$ and $\tilde{t}$ are not independent

$$\mathrm{P}\left(\widetilde{\text{po}}_{\text{Curry}} = 1\right) \neq \mathrm{P}\left(\tilde{y} = 1 \,|\, \tilde{t} = \text{Curry}\right) = \mathrm{P}\left(\widetilde{\text{po}}_{\text{Curry}} = 1 \,|\, \tilde{t} = \text{Curry}\right)$$

$$\mathrm{P}\left(\widetilde{\text{po}}_{\text{Lee}} = 1\right) \neq \mathrm{P}\left(\tilde{y} = 1 \,|\, \tilde{t} = \text{Lee}\right) = \mathrm{P}\left(\widetilde{\text{po}}_{\text{Lee}} = 1 \,|\, \tilde{t} = \text{Lee}\right)$$

Confounding factor

$$
\begin{aligned}
&\mathrm{P}\left(\tilde{y} = 1 \,|\, \tilde{t} = \text{Lee}\right) \\
&= \mathrm{P}(\tilde{y} = 1, \tilde{d} = \text{short} \mid \tilde{t} = \text{Lee}) + \mathrm{P}(\tilde{y} = 1, \tilde{d} = \text{long} \mid \tilde{t} = \text{Lee}) \\
&= \mathrm{P}(\tilde{d} = \text{short} \mid \tilde{t} = \text{Lee})\mathrm{P}(\tilde{y} = 1 \,|\, \tilde{d} = \text{short}, \tilde{t} = \text{Lee}) \\
&\quad + \mathrm{P}(\tilde{d} = \text{long} \mid \tilde{t} = \text{Lee})\mathrm{P}(\tilde{y} = 1 \,|\, \tilde{d} = \text{long}, \tilde{t} = \text{Lee}) \\
&= 0.678 \cdot 0.483 + 0.322 \cdot 0.345 = 0.439
\end{aligned}
$$

Confounding factor

$$
\begin{aligned}
&\mathrm{P}\left(\tilde{y}=1 \,|\, \tilde{t}=\text{Lee}\right) \\
&= 0.678 \cdot 0.483 + 0.322 \cdot 0.345 = 0.439
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{P}\left(\tilde{y}=1 \,|\, \tilde{t}=\text{Curry}\right) \\
&= \mathrm{P}(\tilde{y}=1, \tilde{d}=\text{short} \mid \tilde{t}=\text{Curry}) \\
&\quad + \mathrm{P}(\tilde{y}=1, \tilde{d}=\text{long} \mid \tilde{t}=\text{Curry}) \\
&= \mathrm{P}(\tilde{d}=\text{short} \mid \tilde{t}=\text{Curry})\mathrm{P}(\tilde{y}=1 \,|\, \tilde{d}=\text{short}, \tilde{t}=\text{Curry}) \\
&\quad + \mathrm{P}(\tilde{d}=\text{long} \mid \tilde{t}=\text{Curry})\mathrm{P}(\tilde{y}=1 \,|\, \tilde{d}=\text{long}, \tilde{t}=\text{Curry}) \\
&= 0.197 \cdot 0.500 + 0.803 \cdot 0.396 = 0.416
\end{aligned}
$$

Adjusting for the confounding factor

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

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

Do we know $p_{\widetilde{\text{po}}_{\text{Curry}} \mid \tilde{d}}$?

$p_{\widetilde{\text{po}}_{\text{Curry}} \mid \tilde{d}}$

Assumption: $\widetilde{\text{po}}_{\text{Curry}}$ and $\tilde{t}$ are conditionally independent given $\tilde{d}$

$$\begin{aligned}
\mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \text{short}, \tilde{t} = \text{Curry}) &= \mathrm{P}(\widetilde{\text{po}}_{\text{Curry}} = 1 \mid \tilde{d} = \text{short}, \tilde{t} = \text{Curry}) \\
&= \mathrm{P}(\widetilde{\text{po}}_{\text{Curry}} = 1 \mid \tilde{d} = \text{short}) = 0.5
\end{aligned}$$

$$\begin{aligned}
\mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \text{long}, \tilde{t} = \text{Curry}) &= \mathrm{P}(\widetilde{\text{po}}_{\text{Curry}} = 1 \mid \tilde{d} = \text{long}, \tilde{t} = \text{Curry}) \\
&= \mathrm{P}(\widetilde{\text{po}}_{\text{Curry}} = 1 \mid \tilde{d} = \text{long}) = 0.396
\end{aligned}$$

$p_{\widetilde{\mathsf{po}}_{\mathsf{Lee}} \mid \tilde{d}}$

Assumption: $\widetilde{\mathsf{po}}_{\mathsf{Lee}}$ and $\tilde{t}$ are conditionally independent given $\tilde{d}$

$$\begin{aligned} \mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \mathsf{short}, \tilde{t} = \mathsf{Lee}) &= \mathrm{P}(\widetilde{\mathsf{po}}_{\mathsf{Lee}} = 1 \mid \tilde{d} = \mathsf{short}, \tilde{t} = \mathsf{Lee}) \\ &= \mathrm{P}(\widetilde{\mathsf{po}}_{\mathsf{Lee}} = 1 \mid \tilde{d} = \mathsf{short}) = 0.483 \end{aligned}$$

$$\begin{aligned} \mathrm{P}(\tilde{y} = 1 \mid \tilde{d} = \mathsf{long}, \tilde{t} = \mathsf{Lee}) &= \mathrm{P}(\widetilde{\mathsf{po}}_{\mathsf{Lee}} = 1 \mid \tilde{d} = \mathsf{long}, \tilde{t} = \mathsf{Lee}) \\ &= \mathrm{P}(\widetilde{\mathsf{po}}_{\mathsf{Lee}} = 1 \mid \tilde{d} = \mathsf{long}) = 0.345 \end{aligned}$$

Adjusting for the confounding factor

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

$$= 0.329 \cdot 0.5 + 0.671 \cdot 0.396 = {\color{blue}0.430} \quad (\neq 0.417)$$

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

$$= 0.329 \cdot 0.483 + 0.671 \cdot 0.345 = {\color{red}0.390} \quad (\neq 0.439)$$

Are our assumptions correct? No

Is this a better measure of shooting ability? Yes

What have we learned?

Causal inference from observational data is very challenging

Confounders can severely distort conditional probabilities (Simpson's paradox)

We can adjust for known confounders