POLI210: Political Science Research Methods

Lecture 4.1: Causality I – The potential outcomes framework

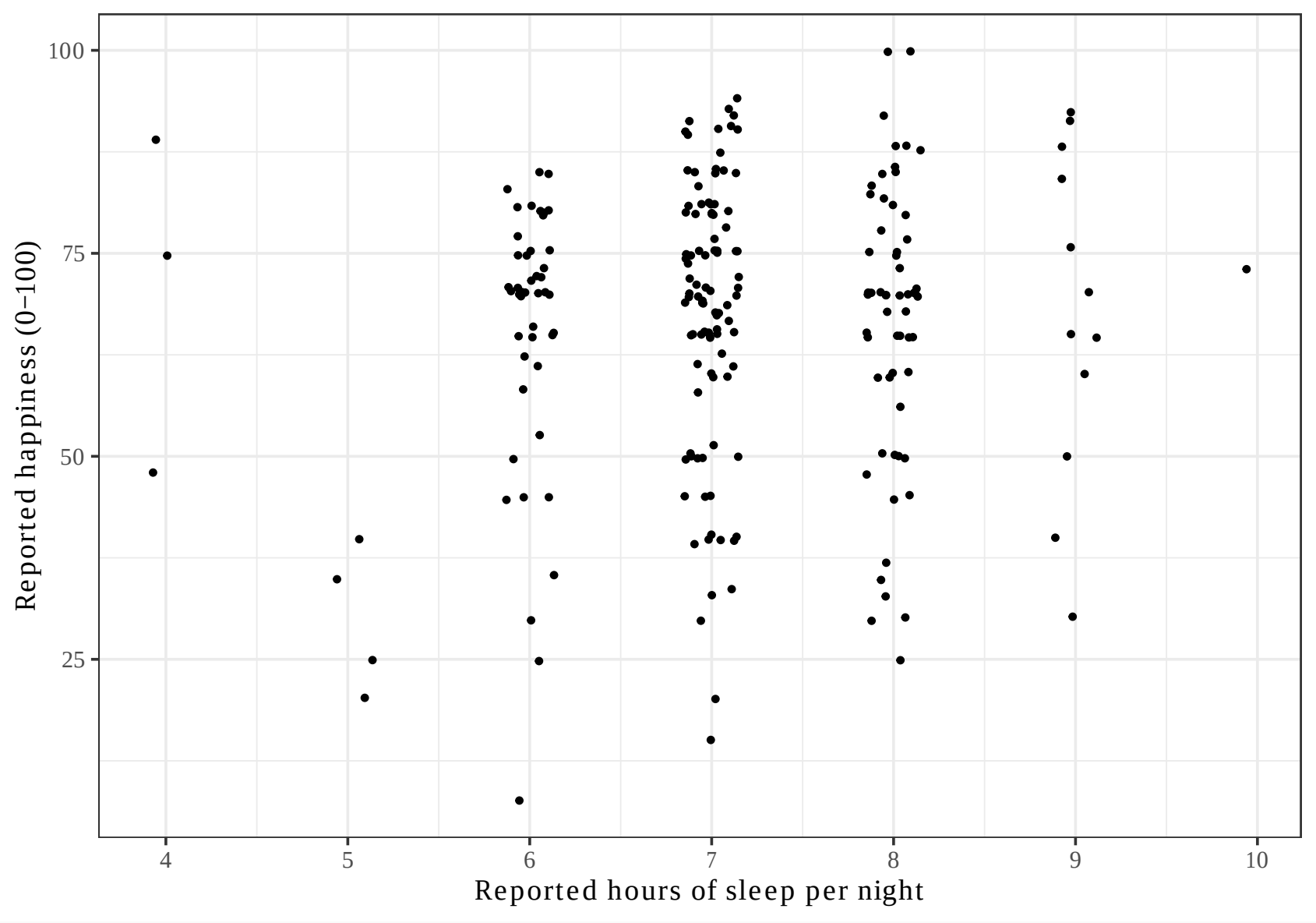
Olivier Bergeron-Boutin

September 21st, 2021

Boring admin stuff

- Thanks for taking the survey! Grades will appear soon
- Assignment 2 will be available shortly
  - Don't stress too much! Very gentle intro

Example from the class survey



Example from the class survey

Reported hours of sleep	Mean happiness
4	70.7
5	30.0
6	65.5
7	67.5
8	66.6
9	67.6
10	73.0

Example from the class survey

- Interpret the data I presented: what is the relationship?
- What's your prior: do you think sleep causally affects happiness?
- Can the data be interpreted causally? Why or why not?
- Come up with a theory:
  - why there would be a *positive* causal effect
  - why there would be *no* causal effect
  - why there would be a *negative* causal effect

Do hospitals hurt people?

Angrist and Pischke (2009) report data from the 2005 National Health Interview Survey

- Self-reported health on 1-5 scale (5 = healthiest)

Do hospitals hurt people?

Angrist and Pischke (2009) report data from the 2005 National Health Interview Survey

- Self-reported health on 1-5 scale (5 = healthiest)
- Visit to hospital or not? ("dummy" variable)

Do hospitals hurt people?

Angrist and Pischke (2009) report data from the 2005 National Health Interview Survey

- Self-reported health on 1-5 scale (5 = healthiest)
- Visit to hospital or not? ("dummy" variable)
- Visited hospital: mean health of 3.21

Do hospitals hurt people?

Angrist and Pischke (2009) report data from the 2005 National Health Interview Survey

- Self-reported health on 1-5 scale (5 = healthiest)
- Visit to hospital or not? ("dummy" variable)
- Visited hospital: mean health of 3.21
- Did not visit hospital: mean health of 3.93

Do hospitals hurt people?

Angrist and Pischke (2009) report data from the 2005 National Health Interview Survey

- Self-reported health on 1-5 scale (5 = healthiest)
- Visit to hospital or not? ("dummy" variable)
- Visited hospital: mean health of 3.21
- Did not visit hospital: mean health of 3.93

Did going to the hospital **cause** people to become less healthy?

- Well, maybe...We can probably fit a theory to the data!
  - Hospitals are full of infectious people!

Do hospitals hurt people?

Angrist and Pischke (2009) report data from the 2005 National Health Interview Survey

- Self-reported health on 1-5 scale (5 = healthiest)
- Visit to hospital or not? ("dummy" variable)
- Visited hospital: mean health of 3.21
- Did not visit hospital: mean health of 3.93

Did going to the hospital **cause** people to become less healthy?

- Well, maybe...We can probably fit a theory to the data!
  - Hospitals are full of infectious people!
- But probably not, right?
  - So what's the problem?

The counterfactual: what if?

The relevant question:

- What would be the health status of some person who went to the hospital, *had they not gone to the hospital*?

The counterfactual: what if?

The relevant question:

- What would be the health status of some person who went to the hospital, *had they not gone to the hospital*?
- What would be the health status of some person who did not go to the hospital, *had they gone to the hospital*?

The counterfactual: what if?

The relevant question:

- What would be the health status of some person who went to the hospital, *had they not gone to the hospital*?
- What would be the health status of some person who did not go to the hospital, *had they gone to the hospital*?
- This is the **counterfactual**: what *would have happened*

The counterfactual: what if?

The relevant question:

- What would be the health status of some person who went to the hospital, *had they not gone to the hospital*?
- What would be the health status of some person who did not go to the hospital, *had they gone to the hospital*?
- This is the **counterfactual**: what *would have happened*

The **fundamental problem of causal inference**: we'll never get to observe the counterfactual!

The counterfactual: what if?

The relevant question:

- What would be the health status of some person who went to the hospital, *had they not gone to the hospital*?
- What would be the health status of some person who did not go to the hospital, *had they gone to the hospital*?
- This is the **counterfactual**: what *would have happened*

The **fundamental problem of causal inference**: we'll never get to observe the counterfactual!

- If a person goes to the hospital, I'll never know what would have happened to them if they didn't go

The counterfactual: what if?

The relevant question:

- What would be the health status of some person who went to the hospital, *had they not gone to the hospital*?
- What would be the health status of some person who did not go to the hospital, *had they gone to the hospital*?
- This is the **counterfactual**: what *would have happened*

The **fundamental problem of causal inference**: we'll never get to observe the counterfactual!

- If a person goes to the hospital, I'll never know what would have happened to them if they didn't go
- If a person does not go to the hospital, I'll never know what would have happened to them if they had gone

The counterfactual: what if?

The relevant question:

- What would be the health status of some person who went to the hospital, *had they not gone to the hospital*?
- What would be the health status of some person who did not go to the hospital, *had they gone to the hospital*?
- This is the **counterfactual**: what *would have happened*

The **fundamental problem of causal inference**: we'll never get to observe the counterfactual!

- If a person goes to the hospital, I'll never know what would have happened to them if they didn't go
- If a person does not go to the hospital, I'll never know what would have happened to them if they had gone
- Causal inference as a problem of missing data

Some notation

Y: the outcome/dependent variable

Some notation

Y: the outcome/dependent variable

Y_1: the value of the outcome for the 1st unit

Some notation

Y: the outcome/dependent variable

Y_1: the value of the outcome for the 1st unit

Y_i: the value of the outcome for the i^{th} unit

Some notation

Y: the outcome/dependent variable

Y_1: the value of the outcome for the 1st unit

Y_i: the value of the outcome for the i^{th} unit

D: The treatment status

Some notation

Y: the outcome/dependent variable

Y_1: the value of the outcome for the 1st unit

Y_i: the value of the outcome for the i^{th} unit

D: The treatment status

- Generally, 1 means "treatment" and 0 means "control"

Some notation

Y: the outcome/dependent variable

Y_1: the value of the outcome for the 1st unit

Y_i: the value of the outcome for the i^{th} unit

D: The treatment status

- Generally, 1 means "treatment" and 0 means "control"

D_i: the treatment status for the i^{th} unit

The potential outcomes

A potential outcome: the outcome that *would be observed* under a certain state of the world

The potential outcomes

A potential outcome: the outcome that *would be observed* under a certain state of the world

- $Y_i(1)$: the potential outcome under treatment for the i^{th} unit

The potential outcomes

A potential outcome: the outcome that *would be observed* under a certain state of the world

- $Y_i(1)$: the potential outcome under treatment for the i^{th} unit
- $Y_i(0)$: the potential outcome under control for the i^{th} unit

The potential outcomes

A potential outcome: the outcome that *would be observed* under a certain state of the world

- $Y_i(1)$: the potential outcome under treatment for the i^{th} unit
- $Y_i(0)$: the potential outcome under control for the i^{th} unit

If my treatment is going to the hospital...

- $Y_i(1)$ is value that Y would take if unit i went to the hospital

The potential outcomes

A potential outcome: the outcome that *would be observed* under a certain state of the world

- $Y_i(1)$: the potential outcome under treatment for the i^{th} unit
- $Y_i(0)$: the potential outcome under control for the i^{th} unit

If my treatment is going to the hospital...

- $Y_i(1)$ is value that Y would take if unit i went to the hospital
- $Y_i(0)$ is value that Y would take if unit i did not go to the hospital

The potential outcomes

A potential outcome: the outcome that *would be observed* under a certain state of the world

- $Y_i(1)$: the potential outcome under treatment for the i^{th} unit
- $Y_i(0)$: the potential outcome under control for the i^{th} unit

If my treatment is going to the hospital...

- $Y_i(1)$ is value that Y would take if unit i went to the hospital
- $Y_i(0)$ is value that Y would take if unit i did not go to the hospital

FPCI, restated: you observe only one of $Y_i(1)$ and $Y_i(0)$

The potential outcomes

A potential outcome: the outcome that *would be observed* under a certain state of the world

- $Y_i(1)$: the potential outcome under treatment for the i^{th} unit
- $Y_i(0)$: the potential outcome under control for the i^{th} unit

If my treatment is going to the hospital...

- $Y_i(1)$ is value that Y would take if unit i went to the hospital
- $Y_i(0)$ is value that Y would take if unit i did not go to the hospital

FPCI, restated: you observe only one of $Y_i(1)$ and $Y_i(0)$

- What if we could observe both? We could compute the individual-level treatment effect: $\tau_i = Y_i(1) - Y_i(0)$

Ideal data

Without the fundamental problem of causal inference, what would our data look like?

Ideal data

Without the fundamental problem of causal inference, what would our data look like?

- Let's keep our example on hospitals

Ideal data

Without the fundamental problem of causal inference, what would our data look like?

- Let's keep our example on hospitals
- Y_i: reported health status on 1 to 5 scale (5 = healthiest)

Ideal data

Without the fundamental problem of causal inference, what would our data look like?

- Let's keep our example on hospitals
- Y_i: reported health status on 1 to 5 scale (5 = healthiest)
- D_i: hospital stay in last 12 months ("dummy" variable: 1s and 0s)

Ideal data

Without the fundamental problem of causal inference, what would our data look like?

- Let's keep our example on hospitals
- Y_i: reported health status on 1 to 5 scale (5 = healthiest)
- D_i: hospital stay in last 12 months ("dummy" variable: 1s and 0s)

D_i	$Y_i(1)$	$Y_i(0)$	Y_i
1	2	1	?
1	3	3	?
0	5	4	?
1	3	1	?
0	2	4	?

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i
1	2	1	2
1	3	3	3
0	5	4	4
1	3	1	3
0	2	4	4

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	?
1	3	3	3	?
0	5	4	4	?
1	3	1	3	?
0	2	4	4	?

Remember that τ_i is the treatment effect: the difference in potential outcomes for any given unit

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

I can compute τ_i since for each unit i, I have access to both potential outcomes

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

I can compute τ_i since for each unit i, I have access to both potential outcomes

- What do we conclude about the *causal* effect of a hospital stay on health?

Real data

With the fundamental problem of causal inference, we only see one realized outcome Y_i for each unit i:

Real data

With the fundamental problem of causal inference, we only see one realized outcome Y_i for each unit i:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	?	2	?
1	3	?	3	?
0	?	4	4	?
1	3	?	3	?
0	?	4	4	?

Mean health for those who were treated? $\frac{2+3+3}{3} = 2.66$

Real data

With the fundamental problem of causal inference, we only see one realized outcome Y_i for each unit i:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	?	2	?
1	3	?	3	?
0	?	4	4	?
1	3	?	3	?
0	?	4	4	?

Mean health for those who were treated? $\frac{2+3+3}{3} = 2.66$

Mean health for those who were not treated? $\frac{4+4}{2} = 4$

What's the problem?

How would you describe this problem?

What's the problem?

How would you describe this problem?

Selection bias

- If the people who went to the hospital had not been...

What's the problem?

How would you describe this problem?

Selection bias

- If the people who went to the hospital had not been...
- they would be doing much worse...

What's the problem?

How would you describe this problem?

Selection bias

- If the people who went to the hospital had not been...
- they would be doing much worse...
- than the people who, in fact, did not go to the hospital

What's the problem?

How would you describe this problem?

Selection bias

- If the people who went to the hospital had not been...
- they would be doing much worse...
- than the people who, in fact, did not go to the hospital

Stated in more formal terms...

- The potential outcome under control for those who self-selected into the treatment is different, on average, than the potential outcome under control for those who self-selected into the control
- $\mathbb{E}[Y_i(0)|D_i = 1] \neq \mathbb{E}[Y_i(0)|D_i = 0]$

People do stuff for a reason!

What's the problem?

How would you describe this problem?

Selection bias

- If the people who went to the hospital had not been...
- they would be doing much worse...
- than the people who, in fact, did not go to the hospital

Stated in more formal terms...

- The potential outcome under control for those who self-selected into the treatment is different, on average, than the potential outcome under control for those who self-selected into the control
- $\mathbb{E}[Y_i(0)|D_i = 1] \neq \mathbb{E}[Y_i(0)|D_i = 0]$

People do stuff for a reason!

Selection bias in our ideal data

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

- Among those who **self-selected** into the control state, what is the average $Y_i(0)$?

Selection bias in our ideal data

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

- Among those who **self-selected** into the control state, what is the average $Y_i(0)$?
  - $\rightarrow$4

Selection bias in our ideal data

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

- Among those who **self-selected** into the control state, what is the average $Y_i(0)$?
  - $\rightarrow$4
  - Is this something I can observe?

Selection bias in our ideal data

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

- Among those who **self-selected** into the control state, what is the average $Y_i(0)$?
  - $\rightarrow$4
  - Is this something I can observe?
- Among those who **self-selected** into the treatment state, what is the average $Y_i(0)$?

Selection bias in our ideal data

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

- Among those who **self-selected** into the control state, what is the average $Y_i(0)$?
  - $\rightarrow$4
  - Is this something I can observe?
- Among those who **self-selected** into the treatment state, what is the average $Y_i(0)$?
  - $\rightarrow$1.66

Selection bias in our ideal data

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

- Among those who **self-selected** into the control state, what is the average $Y_i(0)$?
  - $\rightarrow$4
  - Is this something I can observe?
- Among those who **self-selected** into the treatment state, what is the average $Y_i(0)$?
  - $\rightarrow$1.66
  - Is this something we can observe?

What's the solution?

Given the FPCI, what should we do?

- We'd like to compare the POs and compute τ_i

What's the solution?

Given the FPCI, what should we do?

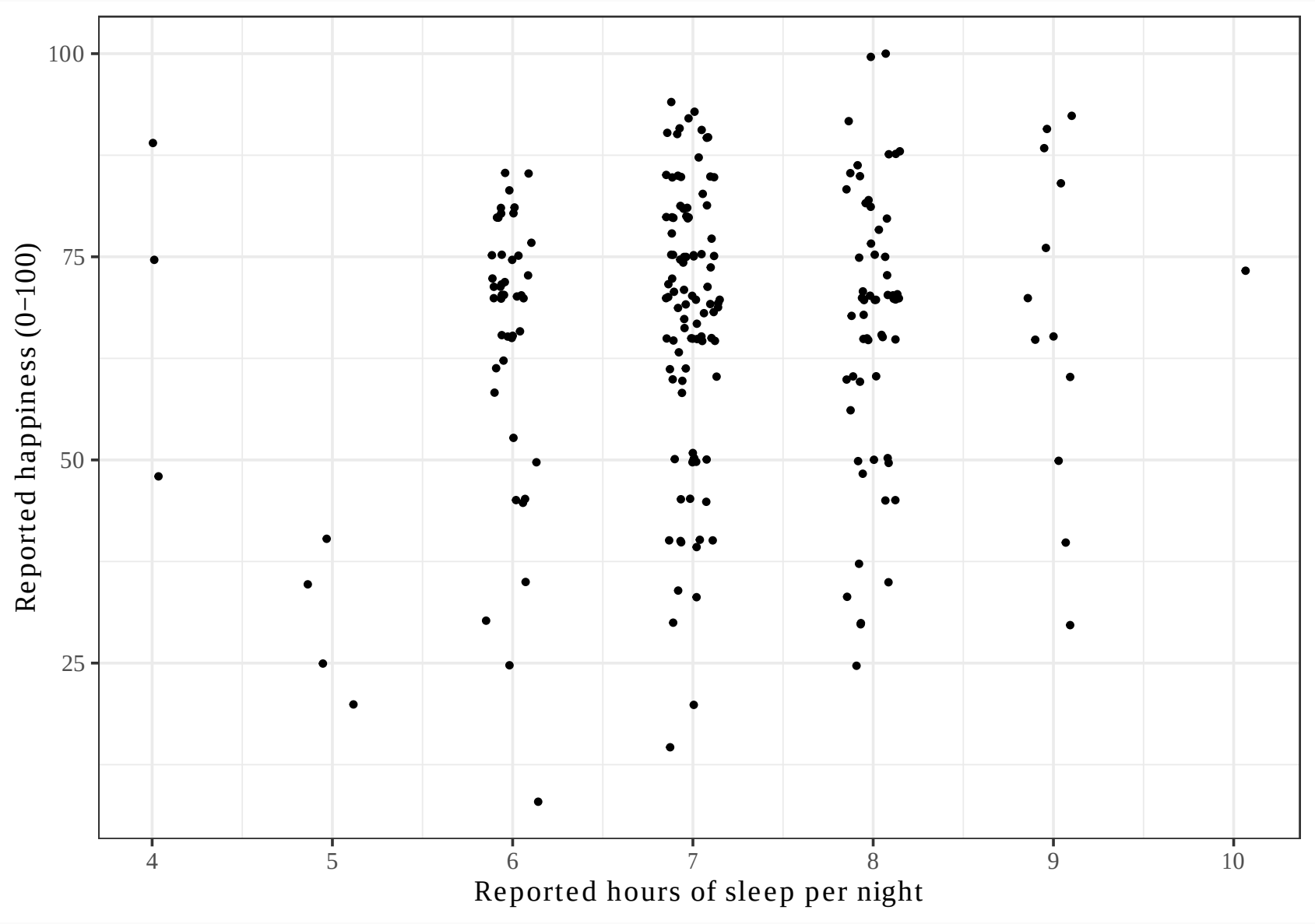
- We'd like to compare the POs and compute τ_i
- But we can't!
- Does this mean we can never draw causal inferences?

What's the solution?

Given the FPCI, what should we do?

- We'd like to compare the POs and compute τ_i
- But we can't!
- Does this mean we can never draw causal inferences?
- Next lecture: the power of randomization

Example from the class survey



Example from the class survey

Reported hours of sleep	Mean happiness
4	70.7
5	30.0
6	65.5
7	67.5
8	66.6
9	67.6
10	73.0

Example from the class survey

- Interpret the data I presented: what is the relationship?
- What's your prior: do you think sleep causally affects happiness?
- Can the data be interpreted causally? Why or why not?
- Come up with a theory:
  - why there would be a *positive* causal effect
  - why there would be *no* causal effect
  - why there would be a *negative* causal effect

References i

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press. https://press.princeton.edu/books/paperback/9780691120355/mostly-harmless-econometrics.