Lecture 23: 22 April, 2026

Madhavan Mukund

https://www.cmi.ac.in/~madhavan

Data Mining and Machine Learning
January–April 2026

Correlation vs causality

- Apply flame to an old cloth

Correlation vs causality

- Apply flame to an old cloth
- Use heavy gloves when we do this

Correlation vs causality

- Apply flame to an old cloth
- Use heavy gloves when we do this
- Cloth burning is correlated to both applying a flame and using heavy gloves

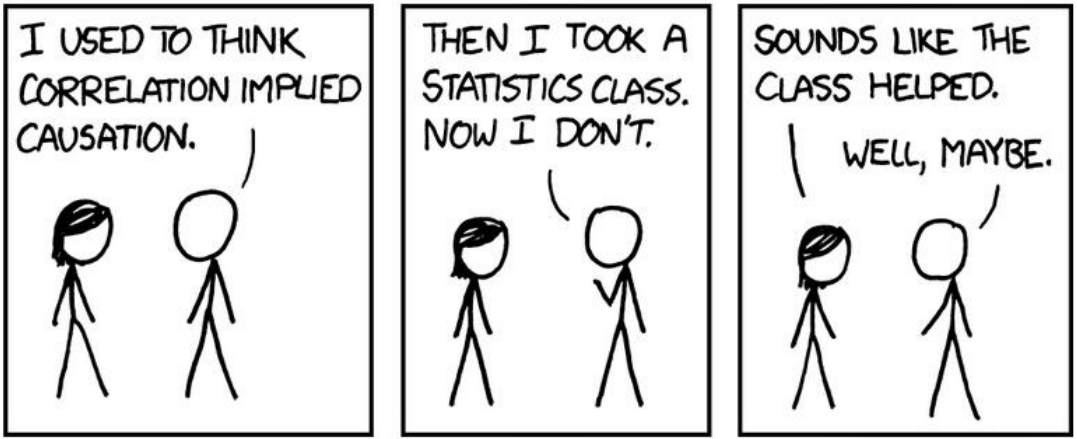
Correlation vs causality

- Apply flame to an old cloth
- Use heavy gloves when we do this
- Cloth burning is correlated to both applying a flame and using heavy gloves
- Which of these correlations implies causation?

Correlation vs causality

- Apply flame to an old cloth
- Use heavy gloves when we do this
- Cloth burning is correlated to both applying a flame and using heavy gloves
- Which of these correlations implies causation?
- Explain using counterfactuals

Correlation vs causality



https://xkcd.com/552

Simpson's paradox

Should we prefer Treatment A or Treatment B?

Mortality rate

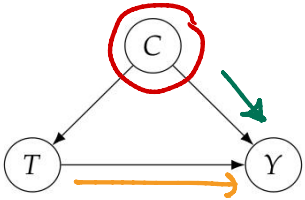
2 treatments for a disease

Availability is 75%, 25%

Treatment	Condition: Mild	Condition: Severe	Condition: Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Simpson's paradox

Condition determines the treatment



- B is preferred for severe cases
- Choose B

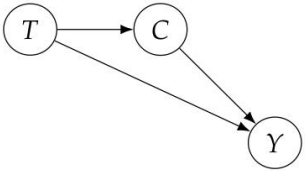
Mortality rate

Treatment	Condition: Mild	Condition: Severe	Condition: Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

⟶ clearly "causal"

Simpson's paradox

Treatment determines the condition



- B is in short supply, delays increase severity
- Choose A

Mortality rate

Treatment	Condition: Mild	Condition: Severe	Condition: Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Confounding

- Observed correlation:

  Go to bed with shoes on, wake up with a headache

Confounding

- Observed correlation:

  Go to bed with shoes on, wake up with a headache

- Does wearing shoes in bed cause headaches?

Confounding

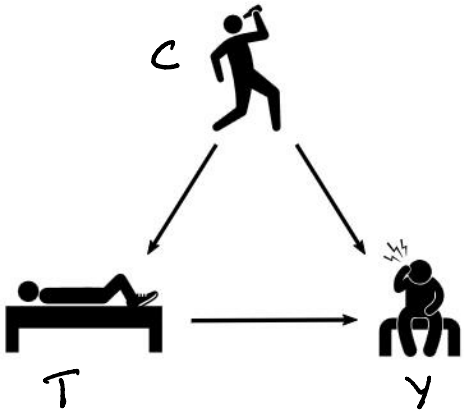
- Observed correlation:

  Go to bed with shoes on, wake up with a headache

- Does wearing shoes in bed cause headaches?

- Hidden cause:

  Both are a consequence of heavy partying the night before!



Confounding

- Observed correlation:

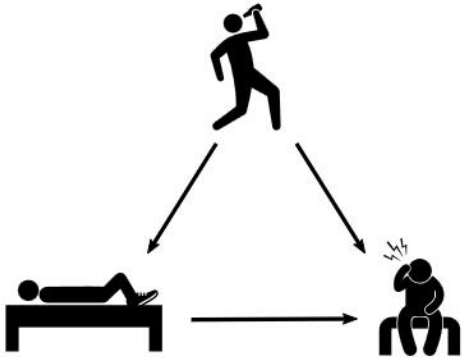
  Go to bed with shoes on, wake up with a headache

- Does wearing shoes in bed cause headaches?

- Hidden cause:

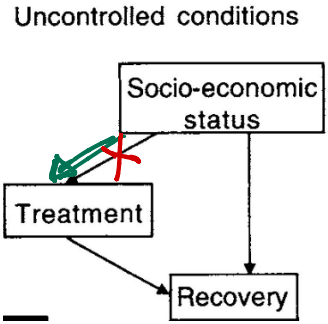
  Both are a consequence of heavy partying the night before!

- Party confounds the association between sleeping with shoes and waking up with a headache



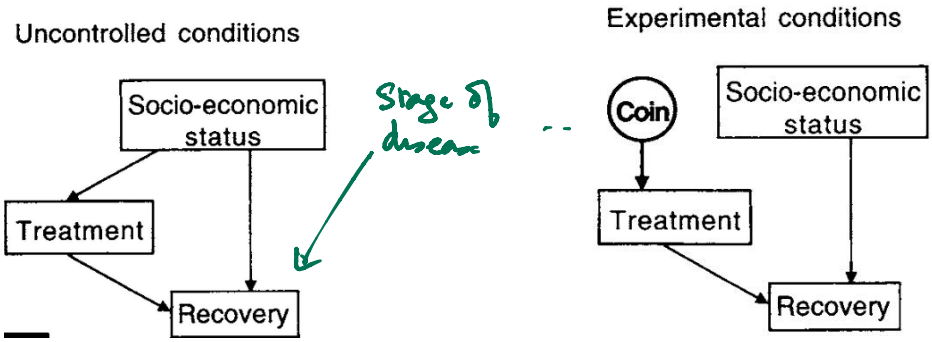
Impact of treatment

Confounding



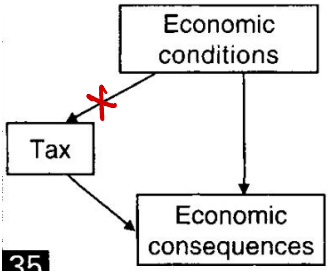
Impact of treatment

Confounding



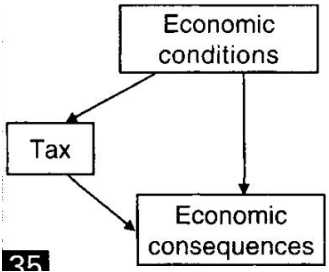
Tax policy

Model underlying data

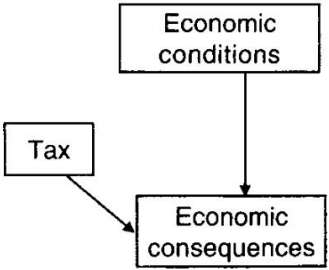


Tax policy

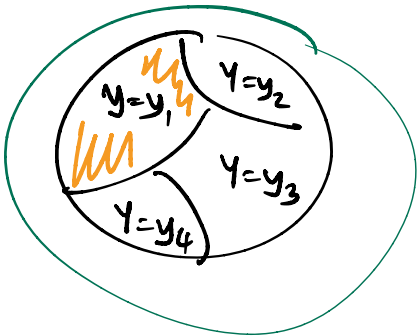
Model underlying data



Model for policy evaluation



$$P(X \mid Y = y_1)$$



Set $Y = y_1$

Intervention

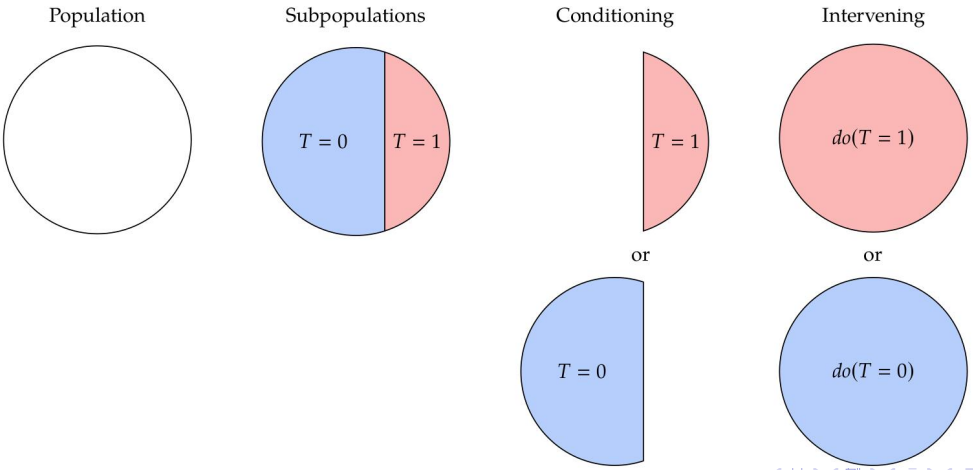
Conditioning vs intevention

$P(Y \mid T = 0)$ vs $P(Y \mid \text{do}(T = 0))$

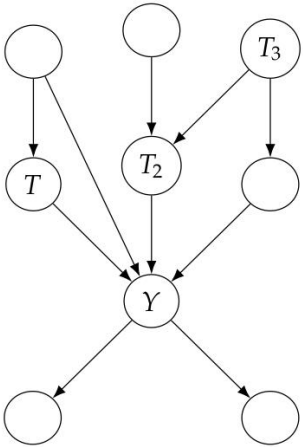
Judea Pearl

Conditioning vs intevention

$P(Y \mid T = 0)$ vs $P(Y \mid \text{do}(T = 0))$

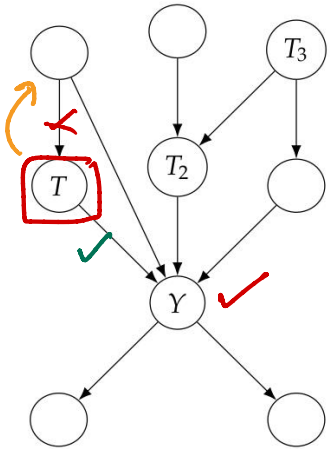


Conditioning vs intevention

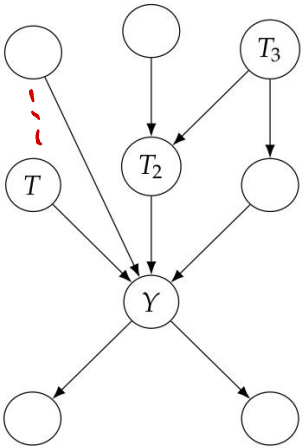


Conditioning vs intevention

$P(Y \mid \text{do}(T = 0))$

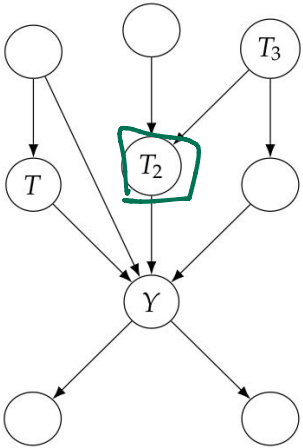


Intervene on T Delete incoming edges

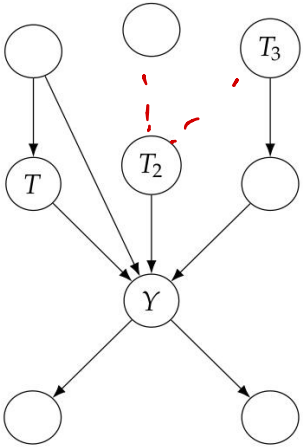


Conditioning vs intevention

$P(Y \mid \text{do}(T_2 = 1))$

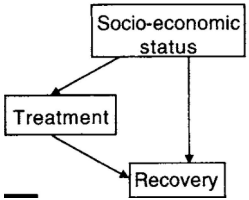


Intervene on T_2

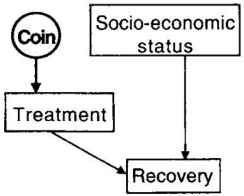


Conditioning vs intevention

Uncontrolled conditions



Experimental conditions



Conditioning vs intevention