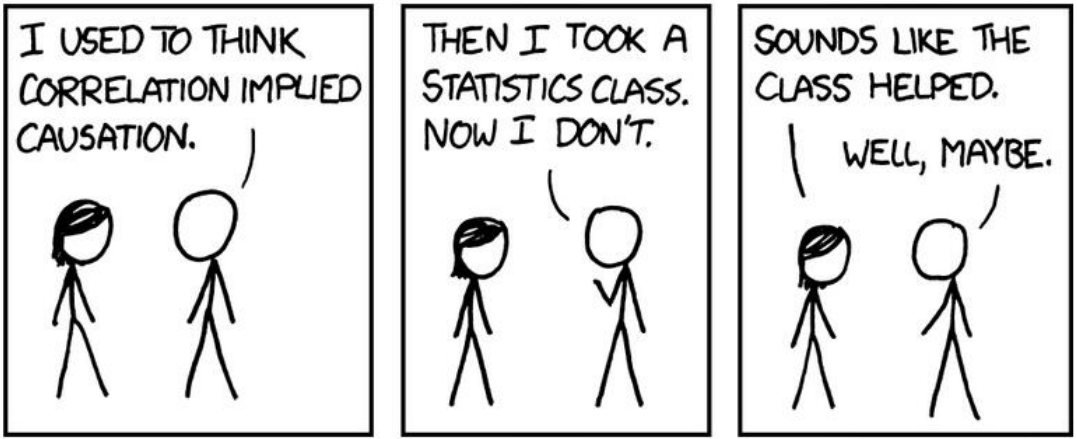
Lecture 25: 24 April, 2025

Madhavan Mukund

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

Data Mining and Machine Learning
January–April 2025

Correlation vs causality



https://xkcd.com/552

Simpson's paradox

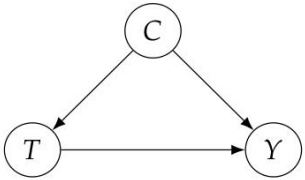
Should we prefer Treatment A or Treatment B?

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)

Simpson's paradox

Condition determines the treatment



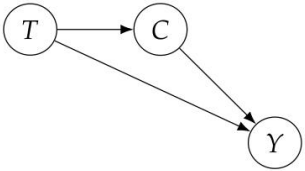
- B is preferred for severe cases
- Choose B

Mortality rate

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


Simpson's paradox

Treatment determines the condition



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

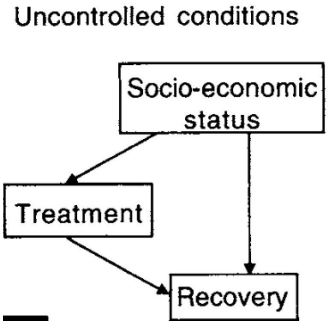
Mortality rate

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)

Impact of treatment

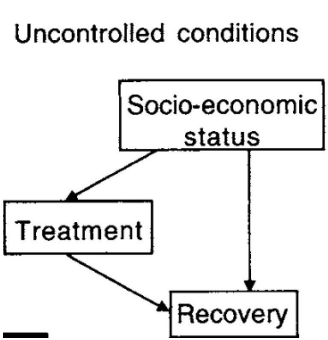
Confounding

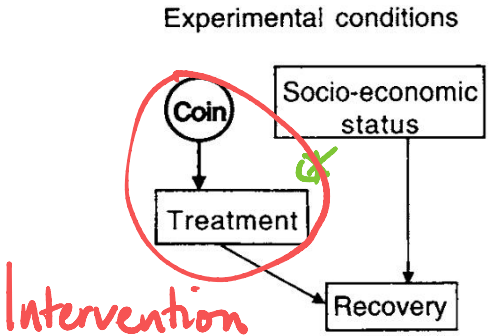
Judea Pearl



Impact of treatment

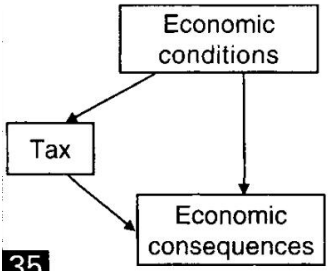
Confounding





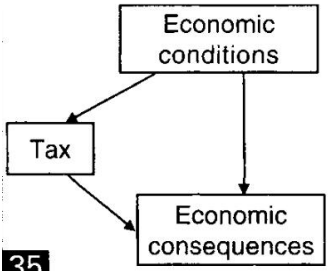
Tax policy

Model underlying data

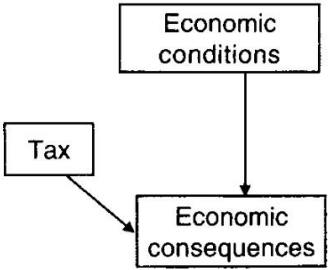


Tax policy

Model underlying data



Model for policy evaluation



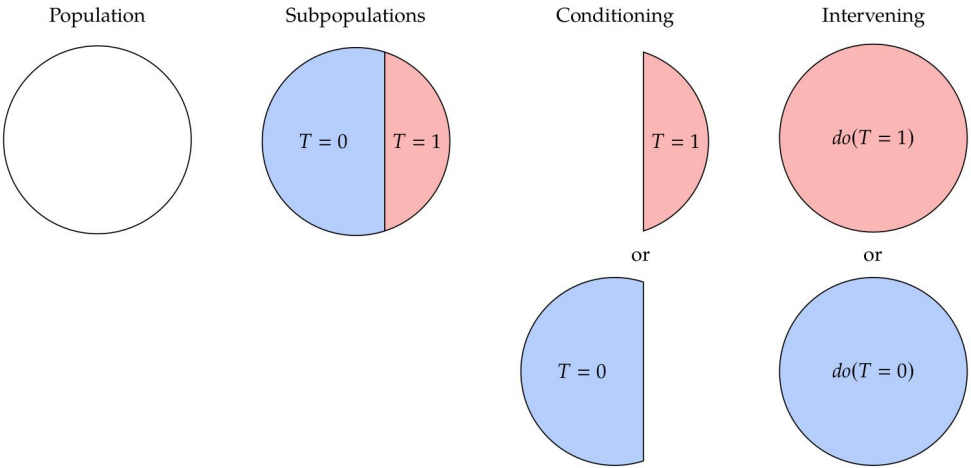
Conditioning vs intevention

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

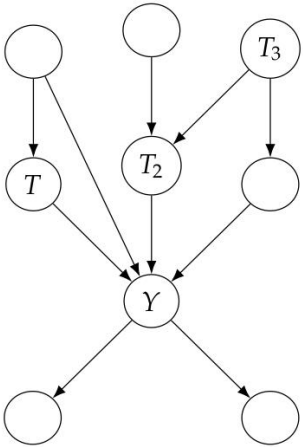
↑
Pearl's notation

Conditioning vs intevention

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

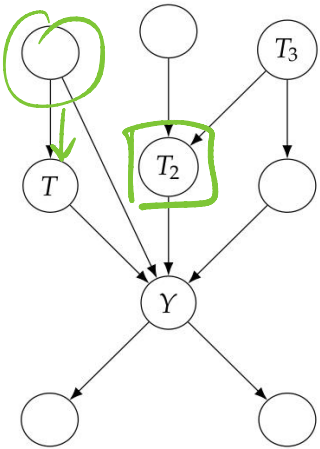


Conditioning vs intevention

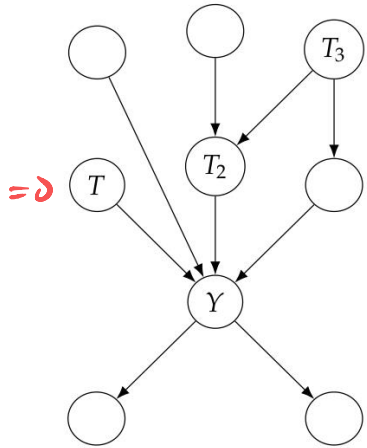


Conditioning vs intevention

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

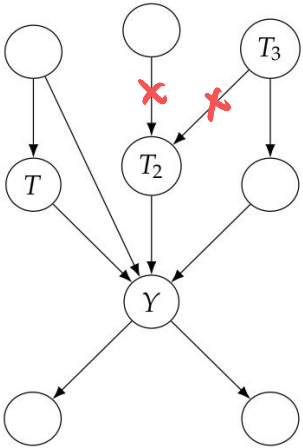


Intervene on T

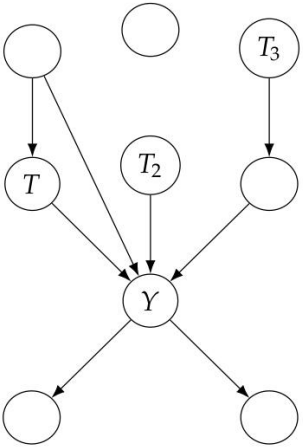


Conditioning vs intevention

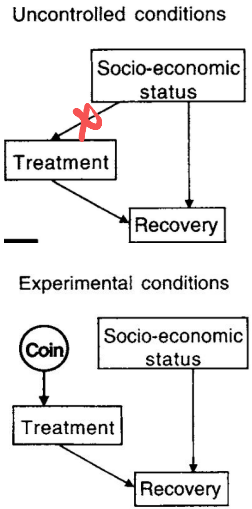
$P(Y \mid \text{do}(T_2 = 1))$



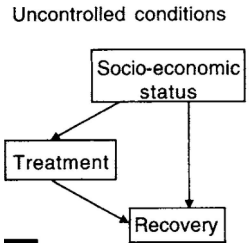
Intervene on T_2

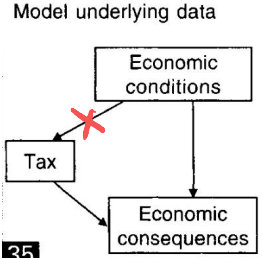


Conditioning vs intevention



Conditioning vs intevention





$(V_1, V_2, \ldots, V_n)$

$P(x_1, x_2, \ldots, x_n)$

$$\prod_i P(V_i = x_i \mid pa_i)$$

Set some subset $S \subseteq \{V_1, \ldots, V_n\}$

If $x_1 \ldots x_n$ is compatible with S

$$\prod_{i \notin S} P(V_i = x_i \mid pa_i)$$