Introduction to causal inference

Session 11

MATH 80667A: Experimental Design and Statistical Methods
HEC Montréal

Outline

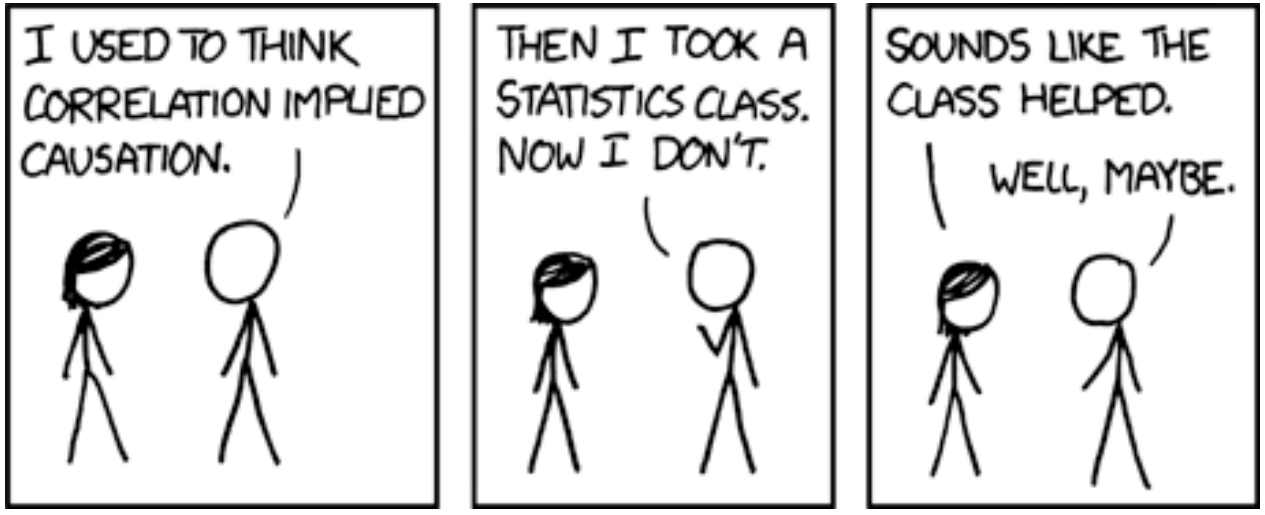
Basics of causal inference

Directed acyclic graphs

Causal mediation

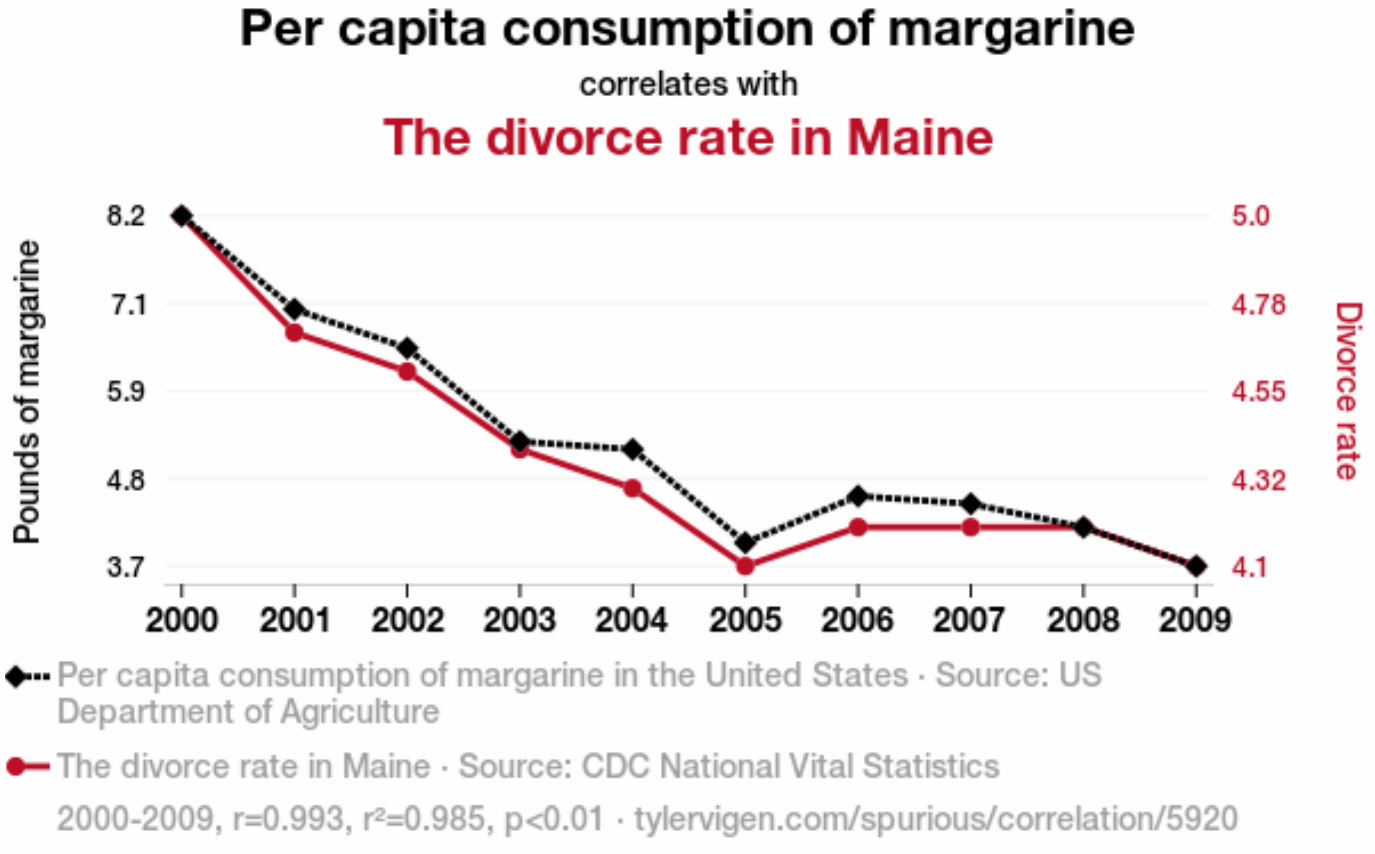
Causal inference

Correlation is not causation



xkcd comic 552 by Randall Munroe, CC BY-NC 2.5 license. Alt text: Correlation doesn't imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'.

Spurious correlation



Spurious correlation by Tyler Vigen, licensed under CC BY 4.0

Correlation vs causation

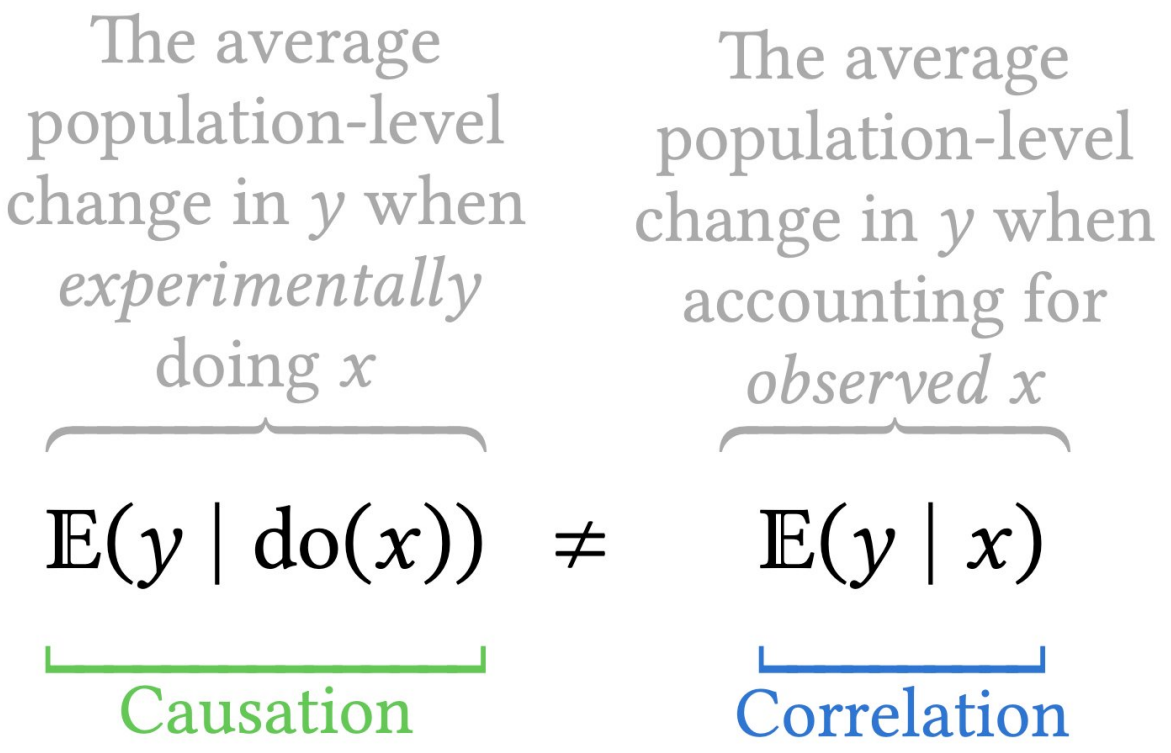


Illustration by Andrew Heiss, licensed under CC BY 4.0

Potential outcomes

For individual i, we postulate the existence of a potential outcomes

- $Y_i(1)$ (response for treatment $X = 1$) and
- $Y_i(0)$ (response for control $X = 0$).

Both are possible, but only one will be realized.

Observe outcome for a single treatment

- Result $Y(X)$ of your test given that you either party $(X = 1)$ or study $(X = 0)$ the night before your exam.

Fundamental problem of causal inference

With binary treatment X_i, I observe either $Y_i \mid \text{do}(X_i = 1)$ or $Y_i \mid \text{do}(X_i = 0)$.

i	X_i	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
1	1	?	4	?
2	0	3	?	?
3	1	?	6	?
4	0	1	?	?
5	0	5	?	?
6	1	?	7	?

Causal assumptions?

Since we can't estimate individual treatment, we consider the **average** treatment effect (average over population) $\mathsf{E}\{Y(1) - Y(0)\}$.

The latter can be estimated as

$$\mathsf{ATE} = \underbrace{\mathsf{E}(Y \mid X = 1)}_{\substack{\text{expected response among} \\ \text{treatment group}}} - \underbrace{\mathsf{E}(Y \mid X = 0)}_{\substack{\text{expected response among} \\ \text{control group}}}$$

When is this a valid causal effect?

(Untestable) assumptions

For the ATE to be equivalent to $\mathsf{E}\{Y(1) - Y(0)\}$, we need:

1. conditional *ignorability*, which states that potential outcomes are independent (denoted with the $\perp\!\!\!\perp$ symbol) of assignment to treatment given a set of explanatories $\boldsymbol{Z}$. In notation $\{Y(0), Y(1)\} \perp\!\!\!\perp X \mid \boldsymbol{Z}$
2. lack of interference: the outcome of any participant is unaffected by the treatment assignment of other participants.
3. consistency: given a treatment X taking level j, the observed value for the response $Y \mid X = j$ is equal to the corresponding potential outcome $Y(j)$.

Directed acyclic graphs

Types of data

Experimental

You have control over which units get treatment

Observational

You don't have control over which units get treatment

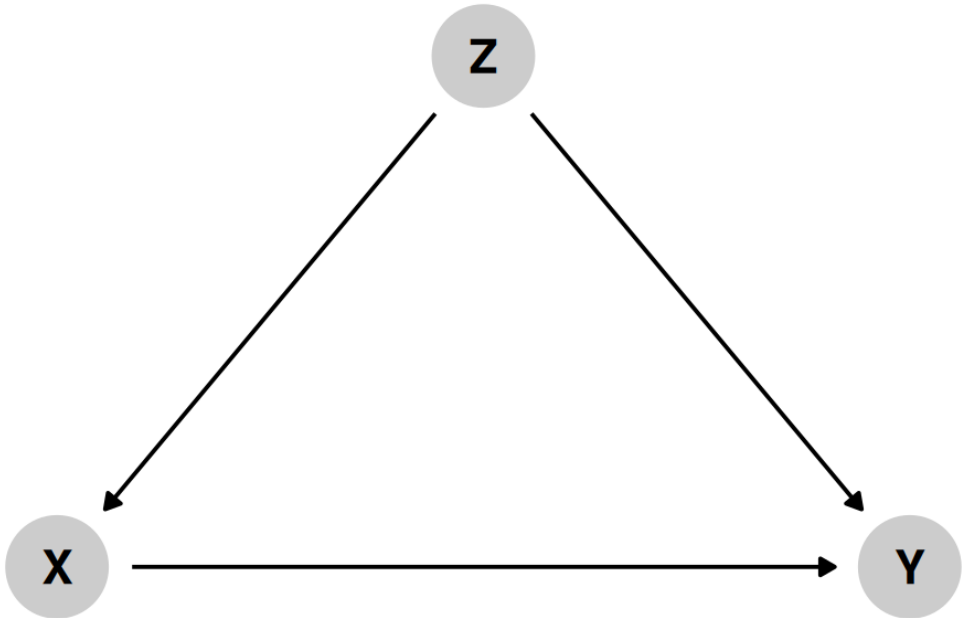
Causal diagrams

Directed acyclic graphs (DAGs)

Directed: Each node has an arrow that points to another node

Acyclic: You can't cycle back to a node (and arrows only have one direction)

Graph: A set of nodes (variables) and vertices (arrows indicating interdependence)

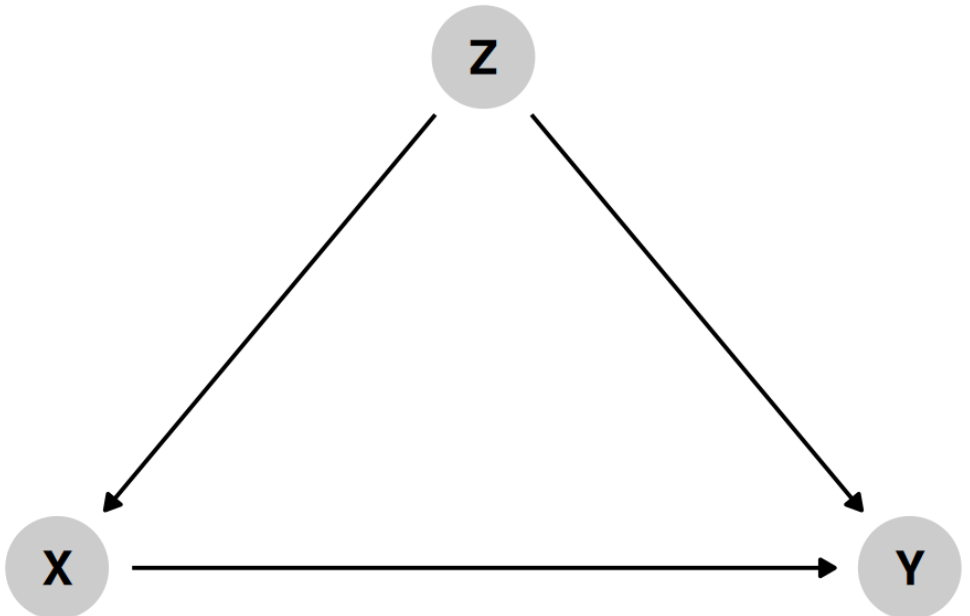


Causal diagrams

Directed acyclic graphs (DAGs)

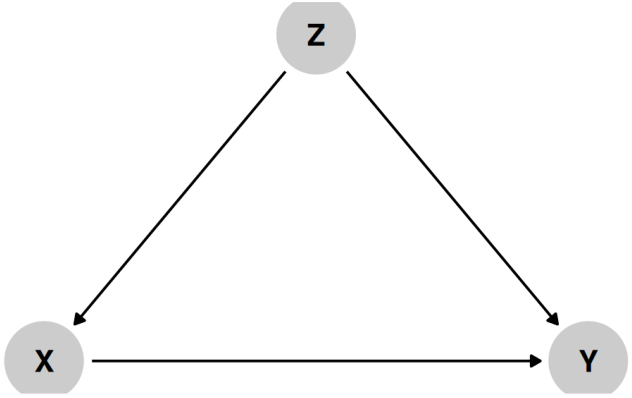
Graphical model of the process that generates the data

Maps your logical model



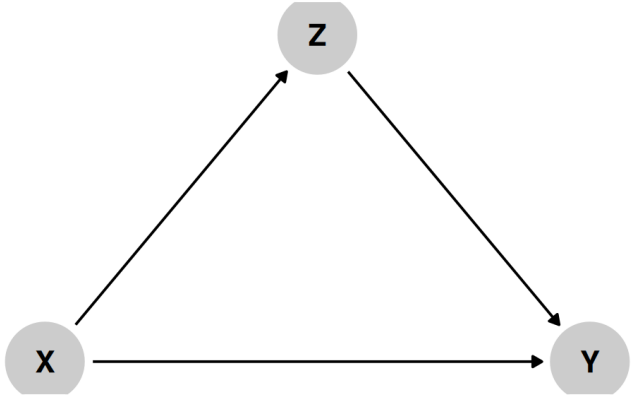
Three types of associations

Confounding



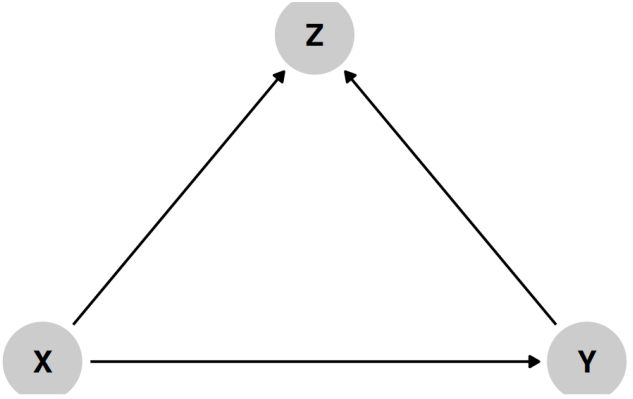
Common cause

Causation



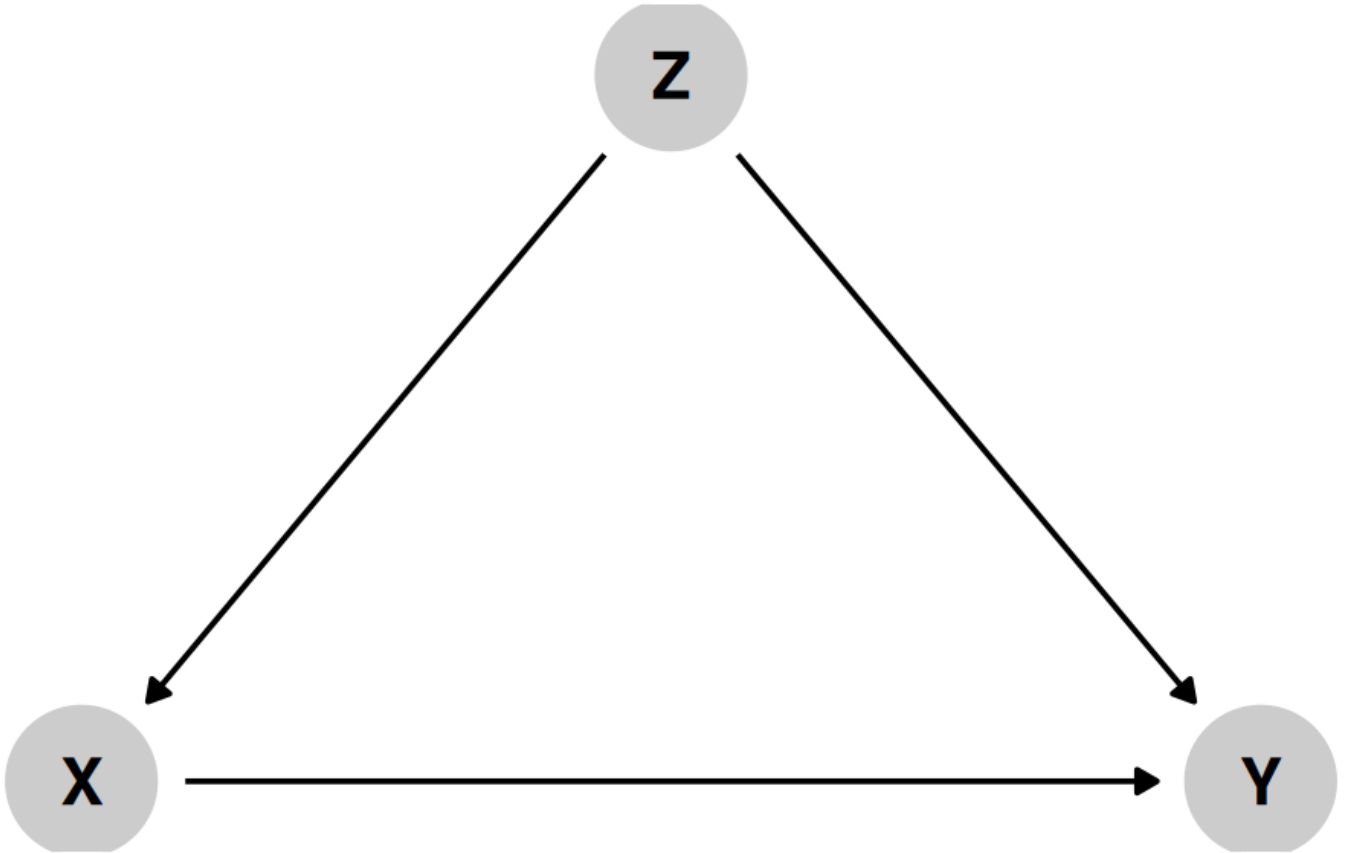
Mediation

Collision



Selection / endogeneity

Confounding



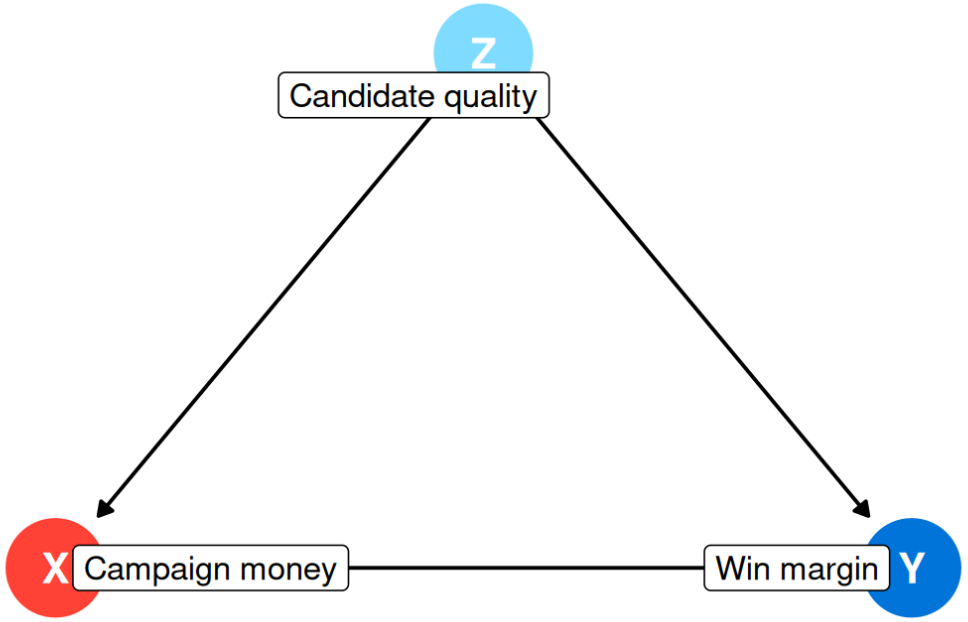
X causes Y

But Z causes both X and Y

Z *confounds* the X → Y association

Confounder: effect of money on elections

What are the paths between money and win margin?



Money → Margin

Money ← Quality → Margin

Quality is a *confounder*

Experimental data

Since we randomize assignment to treatment X, all arrows **incoming** in X are removed.

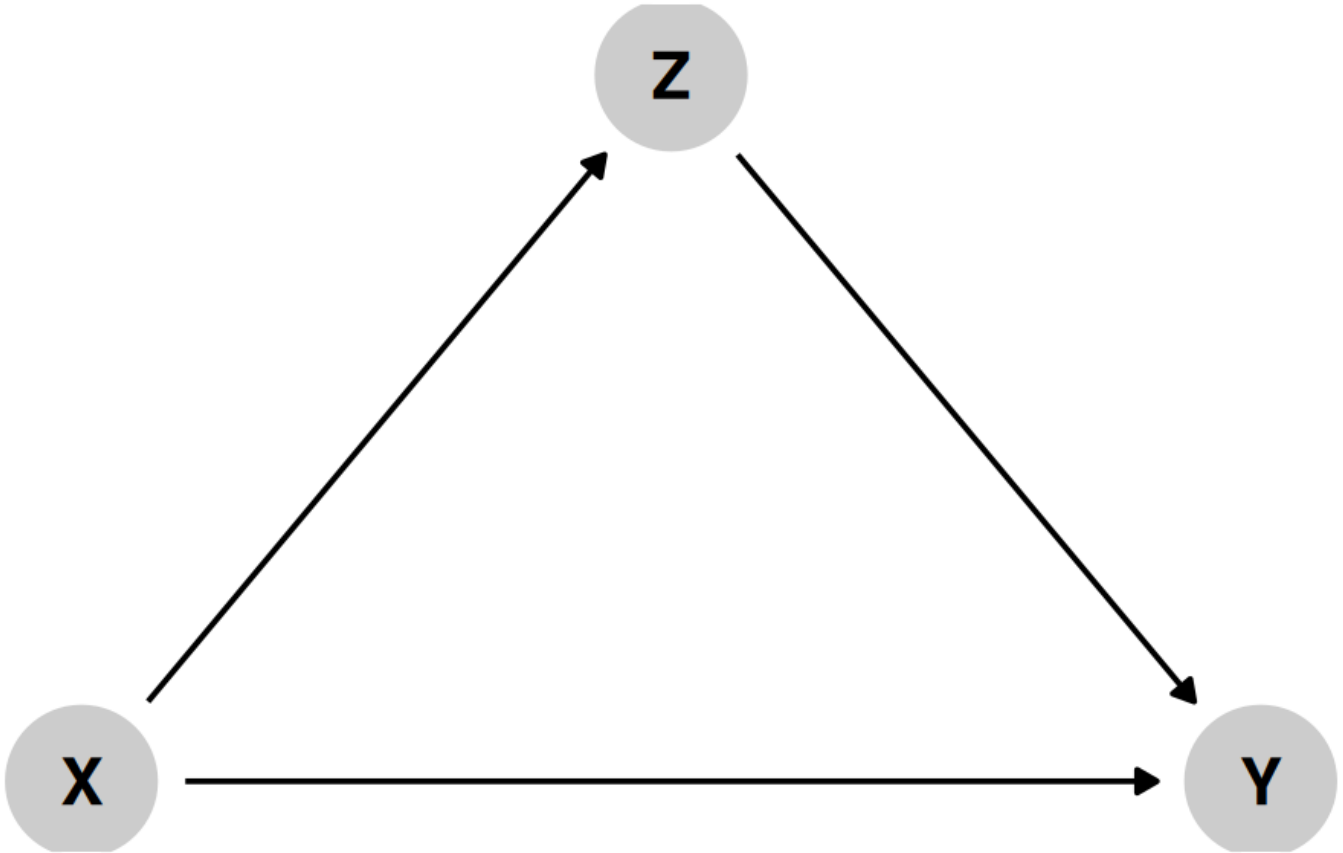
With observational data, we need to explicitly model the relationship and strip out the effect of X on Y.

How to adjust with observational data

Include covariate in regression

Matching | **Stratifying** | **Inverse probability weighting**

Causation

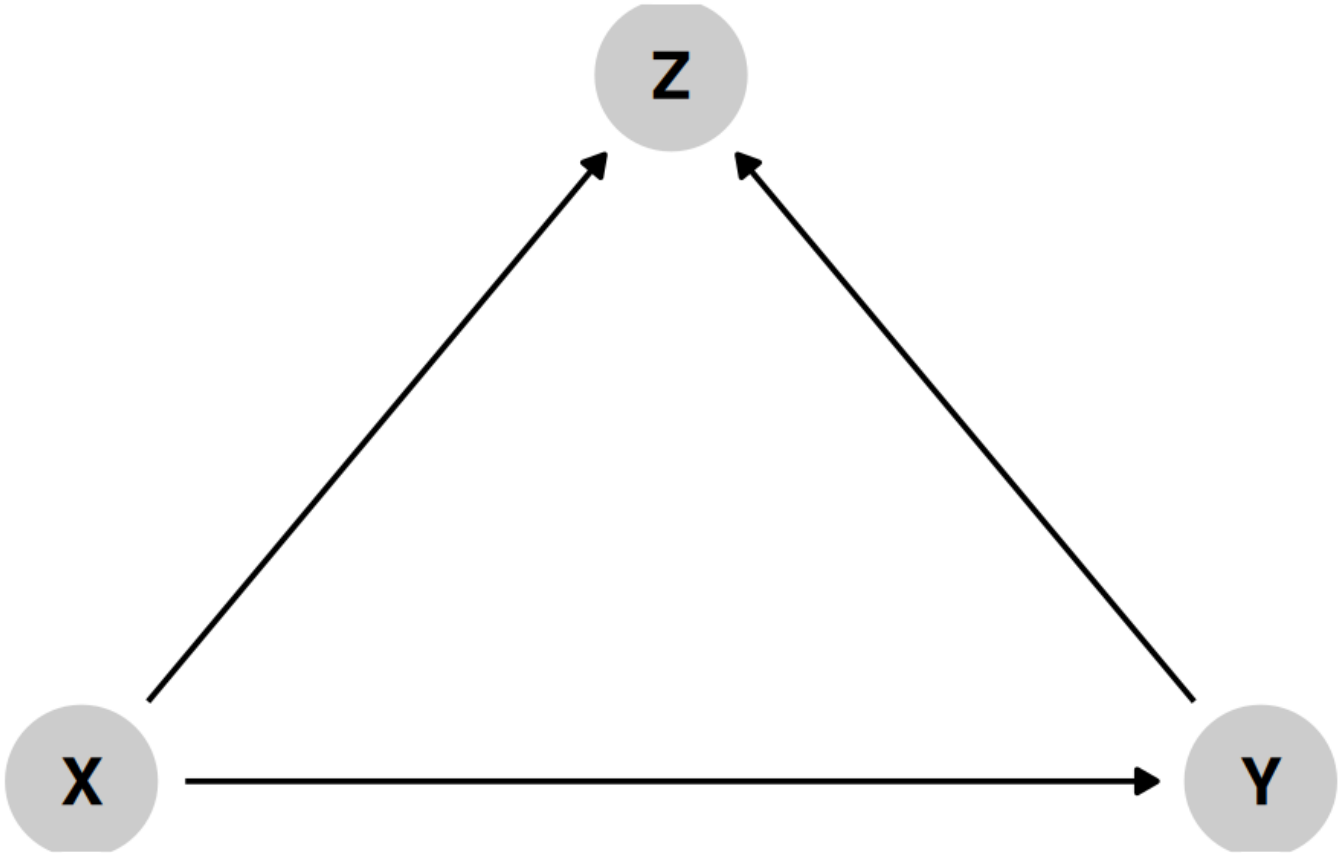


X causes Y

X causes Z which causes Y

Z is a mediator

Colliders



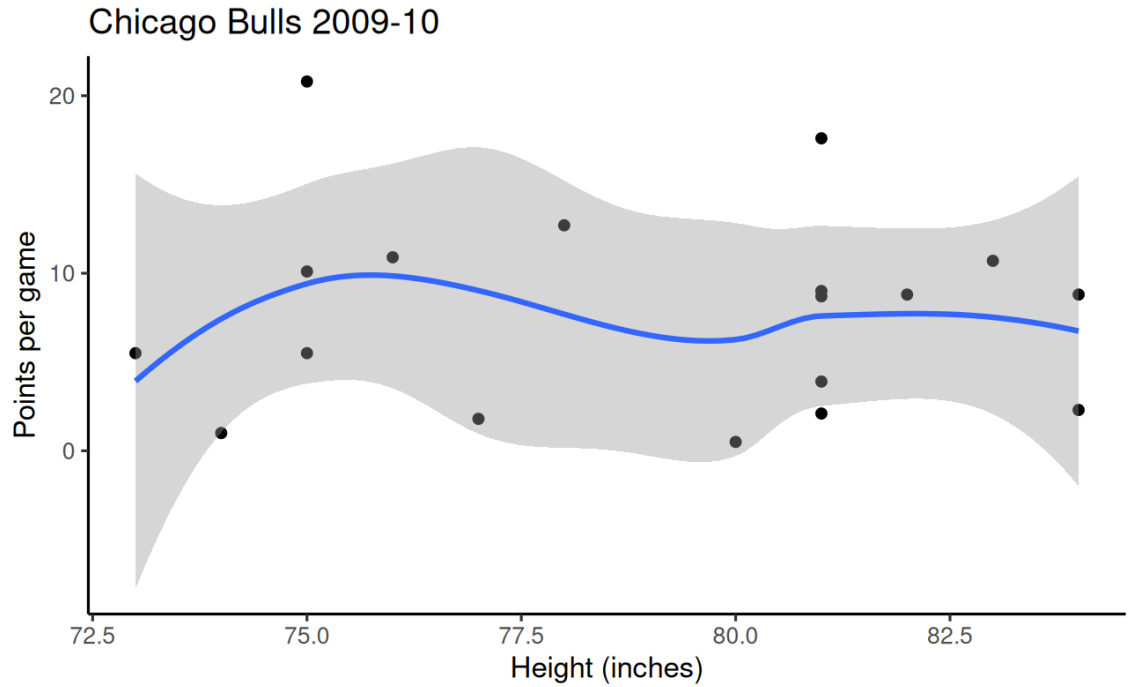
X causes Z

Y causes Z

Should you control for Z?

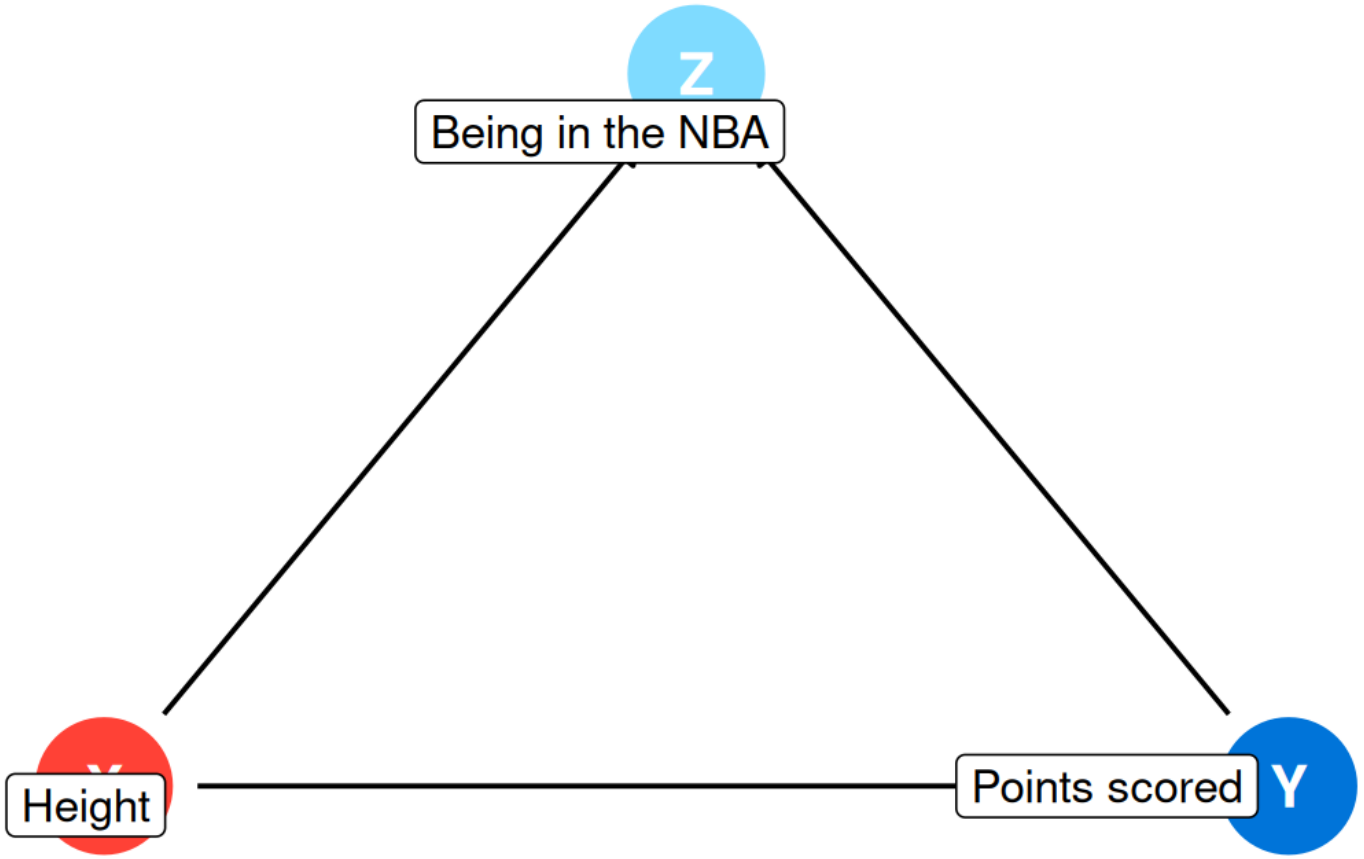
Colliders can create fake causal effects

Colliders can hide real causal effects



Height is unrelated to basketball skill... among NBA players

Colliders and selection bias

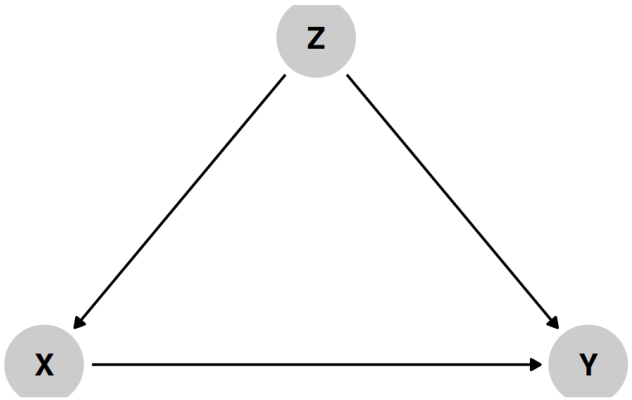


Conditioning on colliders

- Omnipresent in the literature
- Example: When and how does the number of children affect marital satisfaction? An international survey
- Example: The Predictive Validity of the GRE Across Graduate Outcomes

Three types of associations

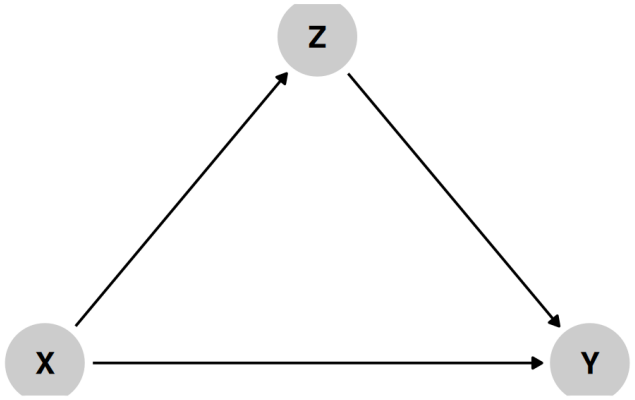
Confounding



Common cause

Causal forks $X \leftarrow Z \rightarrow Y$

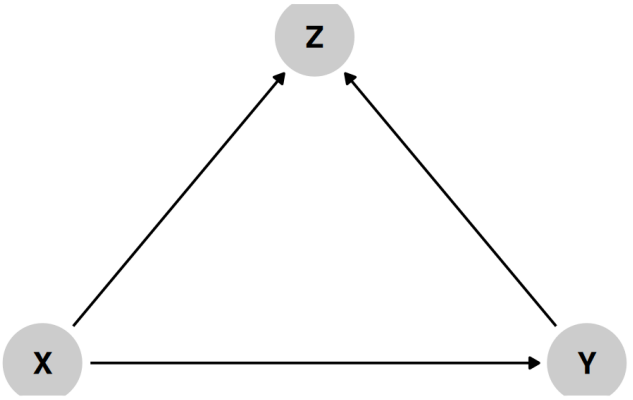
Causation



Mediation

Causal chain $X \rightarrow Z \rightarrow Y$

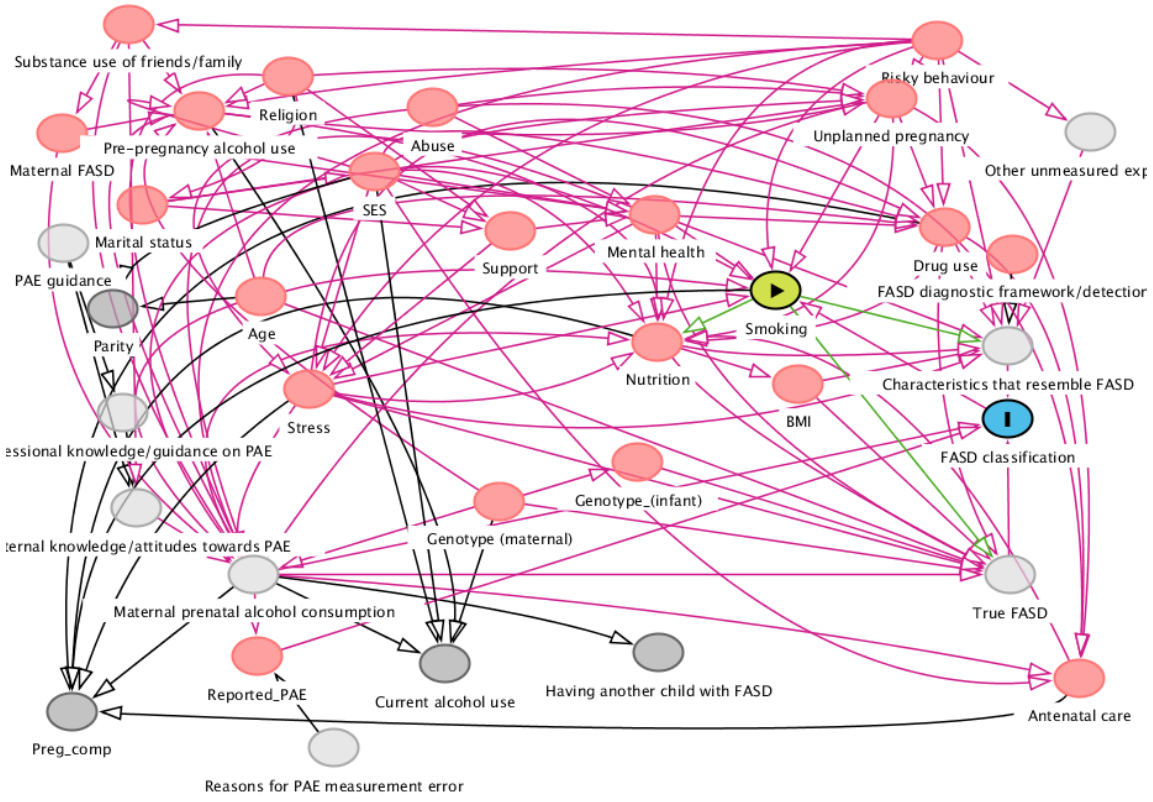
Collision



Selection / endogeneity

inverted fork $X \rightarrow Z \leftarrow Y$

Life is inherently complex



Postulated DAG for the effect of smoking on fetal alcohol spectrum disorders (FASD)

Causal mediation

Key references

- Imai, Keele and Tingley (2010), A General Approach to Causal Mediation Analysis, *Psychological Methods*.
- Pearl (2014), Interpretation and Identification of Causal Mediation, *Psychological Methods*.
- Baron and Kenny (1986), The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations, *Journal of Personality and Social Psychology*

Limitations:

- Bullock, Green, and Ha (2010), Yes, but what's the mechanism? (don't expect an easy answer)
- Uri Simonsohn (2022) Mediation Analysis is Counterintuitively Invalid

Sequential ignorability assumption

- potential mediation given treatment x as $M_i(x)$ and
- potential outcome for treatment x and mediator m as $Y_i(x, m)$.
- Given pre-treatment covariates $\boldsymbol{Z}$, potential outcomes for mediation and treatment are conditionally independent of treatment assignment.

$$Y_i(x', m), M_i(x) \perp\!\!\!\perp X_i \mid \boldsymbol{Z}_i = \boldsymbol{z}$$

- Given pre-treatment covariates and observed treatment, potential outcomes are independent of mediation.

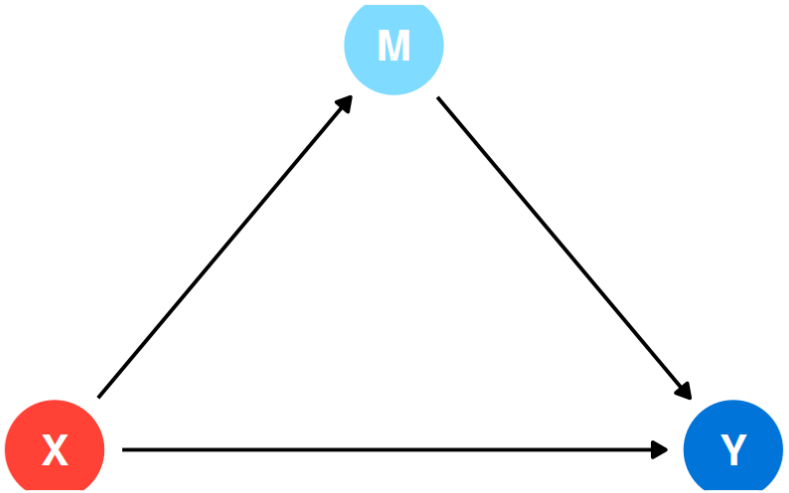
$$Y_i(x', m) \perp\!\!\!\perp M_i(x) \mid X_i = x, \boldsymbol{Z}_i = \boldsymbol{z}$$

Total effect

Total effect: overall impact of X (both through M and directly)

$$\mathsf{TE}(x, x^*) = \mathsf{E}[Y \mid \mathrm{do}(X = x)] - \mathsf{E}[Y \mid \mathrm{do}(X = x^*)]$$

X → M → Y
plus
X → Y



Average controlled direct effect

$$\begin{aligned}\mathsf{CDE}(m, x, x^*) &= \mathsf{E}[Y \mid \mathrm{do}(X = x, m = m)] - \mathsf{E}[Y \mid \mathrm{do}(X = x^*, m \\ &= \mathsf{E}\{Y(x, m) - Y(x^*, m)\}\end{aligned}$$

Expected population change in response when the experimental factor changes from x to x^* and the mediator is set to a fixed value m.

Problem: this forces manipulation of the mediator, and only gives outcome for a fixed value m.

Direct and indirect effects

The **natural direct effect** is the expected change in Y under treatment x if M is set to whatever value it would take under control x^*:

$$\mathsf{NDE}(x, x^*) = \mathsf{E}[Y\{x, M(x^*)\} - Y\{x^*, M(x^*)\}].$$

The **natural indirect effect** is the expected change in Y if we set X to its control value and change the mediator value which it would attain under x:

$$\mathsf{NIE}(x, x^*) = \mathsf{E}[Y\{x^*, M(x)\} - Y\{x^*, M(x^*)\}].$$

Total effect

Counterfactual conditioning reflects a physical intervention, not mere (probabilistic) conditioning.

We define the **total effect** as

$$\mathsf{TE}(x, x^*) = \mathsf{NDE}(x, x^*) - \mathsf{NIE}(x^*, x).$$

Necessary and sufficiency of mediation

From Pearl (2014):

> The difference $\mathsf{TE} - \mathsf{NDE}$ quantifies the extent to which the response of Y is owed to mediation, while NIE quantifies the extent to which it is explained by mediation. These two components of mediation, the necessary and the sufficient, coincide into one in models void of interactions (e.g., linear) but differ substantially under moderation

- In linear systems, changing the order of arguments amounts to flipping signs
- This definition works under temporal reversal and gives the correct answer (the regression-slope approach of the linear structural equation model does not).