12 Maggio 2025

INTRODUCTION TO

CAUSAL MODELLING AND REASONING

Martina Cinquini & Isacco Beretta

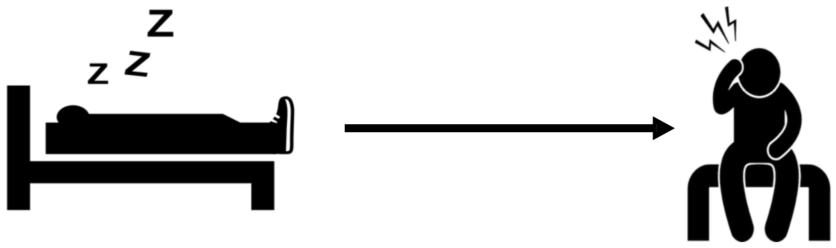
REICHENBACH COMMON CAUSE PRINCIPLE

Let X and Y be two variables such that X and Y are **statistically dependent** then it holds:

- X is indirectly causing Y
- Y is indirectly causing X
- There is a possibly unobserved common cause Z that indirectly causes both X and Y

CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with shoes on is strongly correlated with waking up with a headache

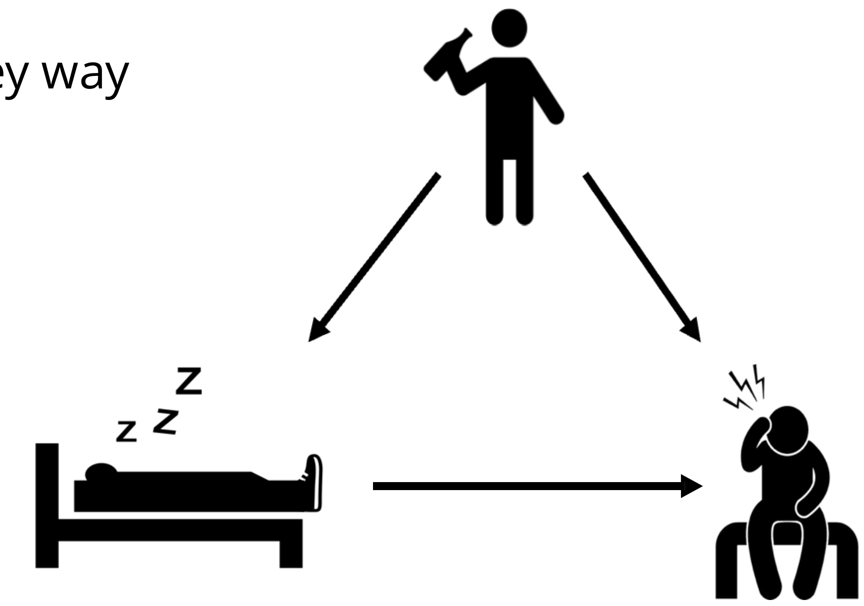


CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way

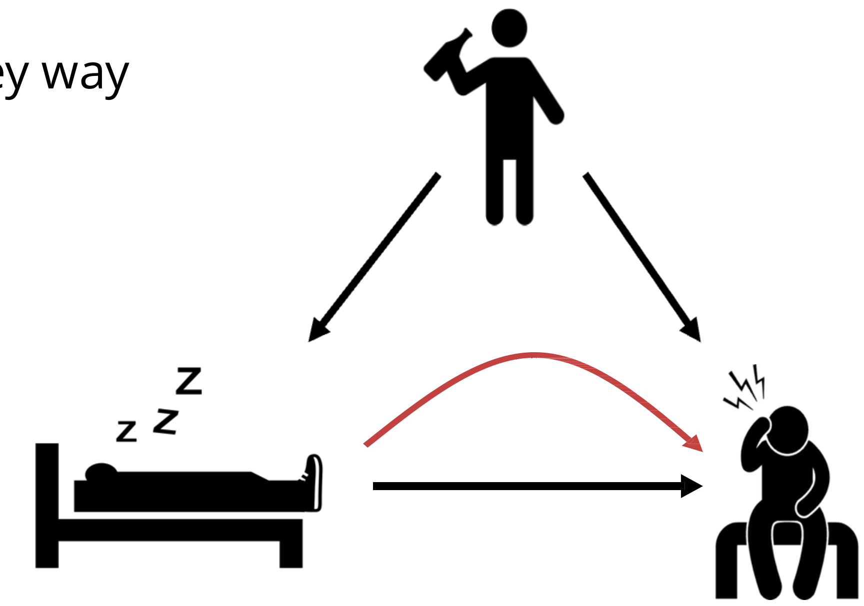


CORRELATION DOES NOT IMPLY CAUSATION

Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding



CORRELATION DOES NOT IMPLY CAUSATION

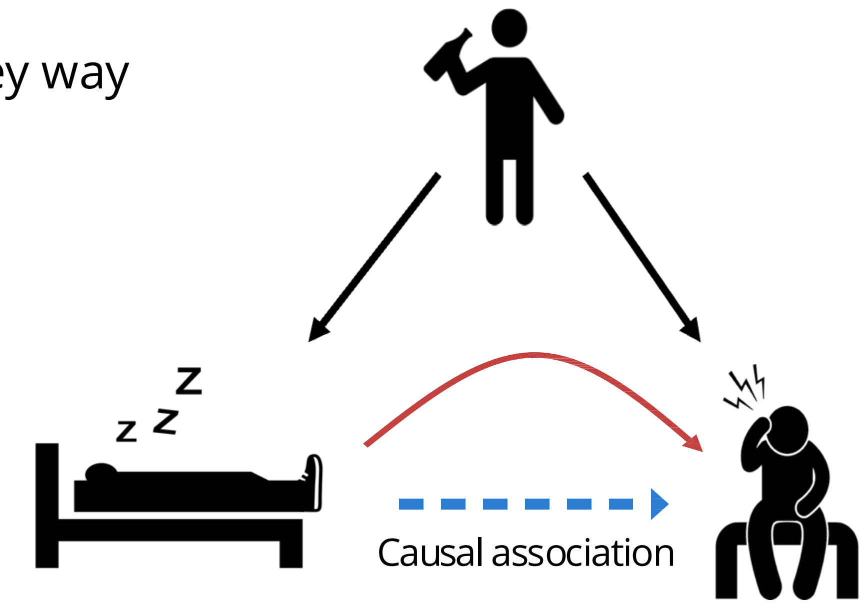
Sleeping with shoes on is strongly correlated with waking up with a headache

Common cause: drinking the night before

1. Shoe-sleepers differ from non-shoe-sleepers in a key way
2. Confounding

Total association (e.g., correlation):

Mixture of causal and confounding association



INGREDIENTS OF A STATISTICAL THEORY OF CAUSALITY

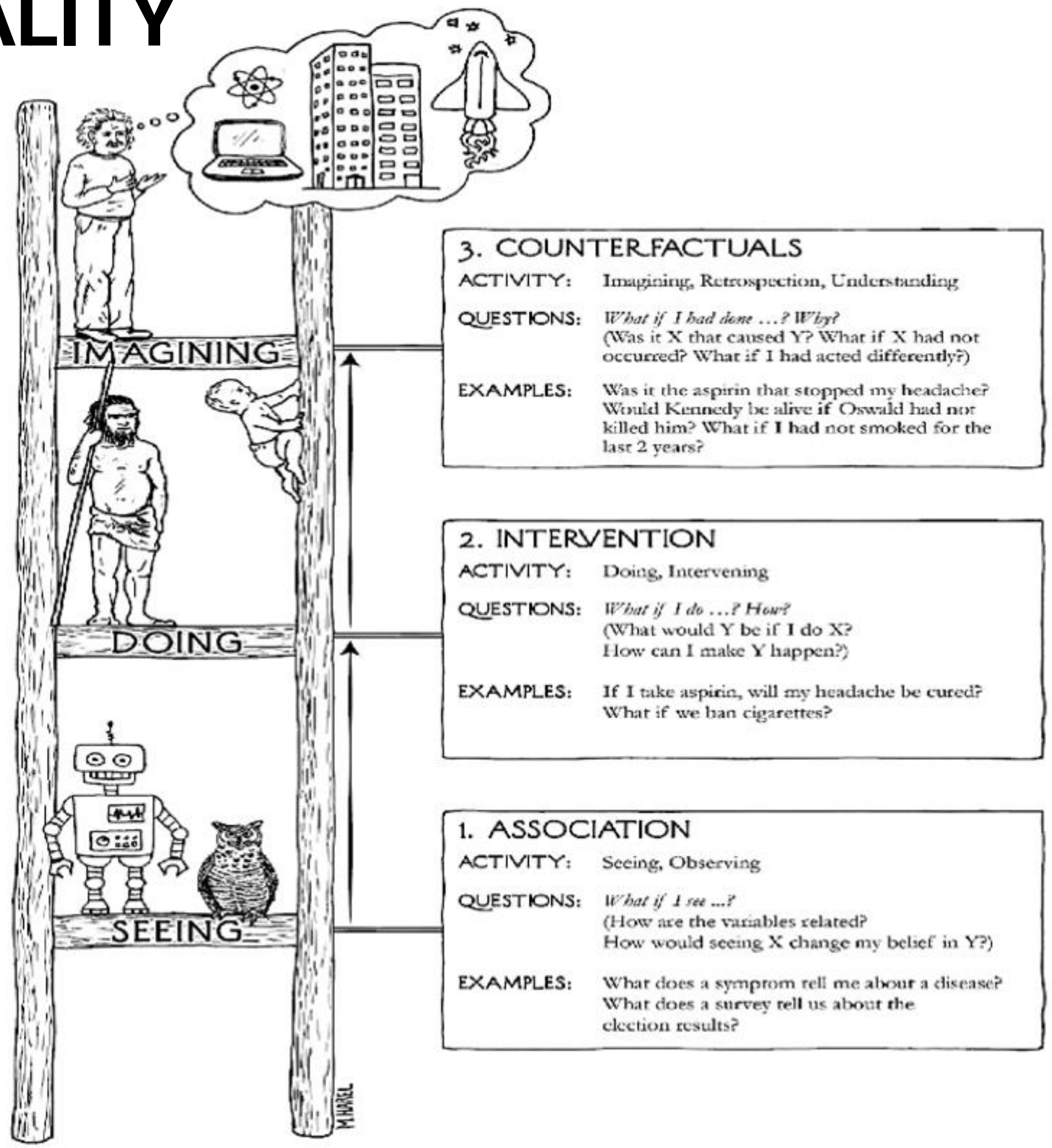
- Working definition of causation
- Method for creating causal models
- Method for linking causal models with features of data
- Method for reasoning over model and data

THE LATTER OF CAUSALITY

“Actual” Causality

“Causality-in-mean”

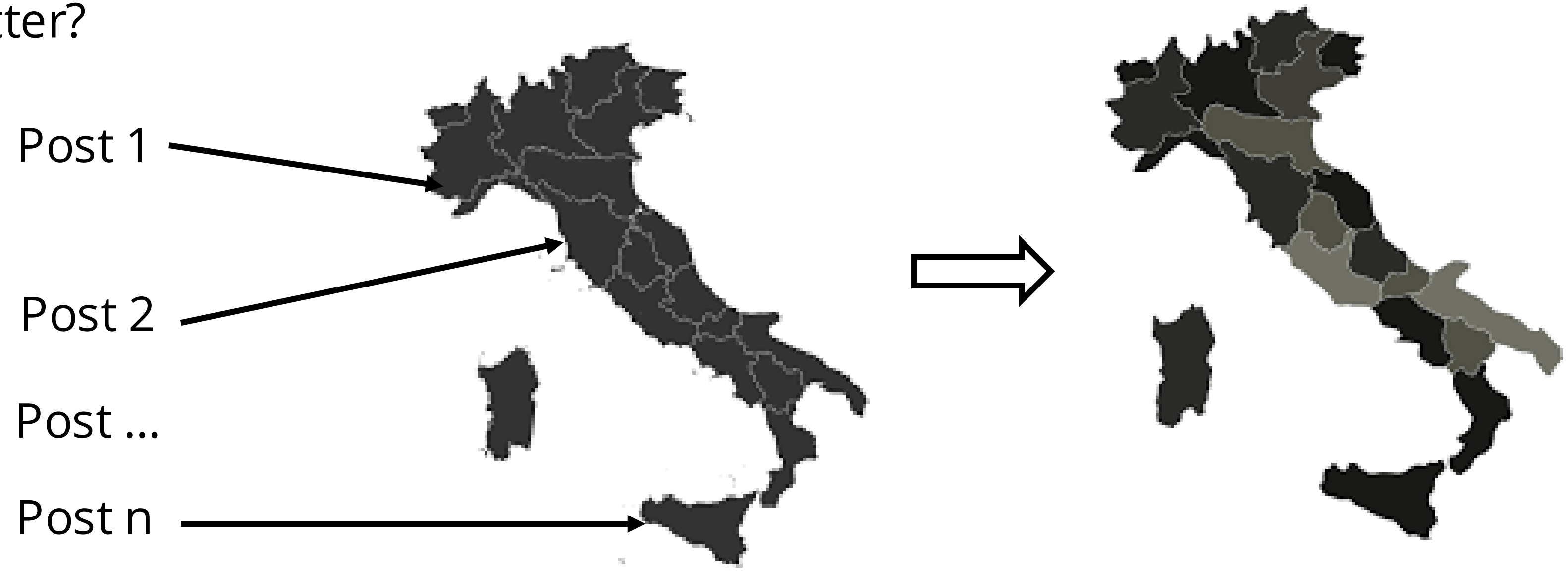
Statistics



RANDOMIZED EXPERIMENTS

Which kind of post works better?

Interventional data

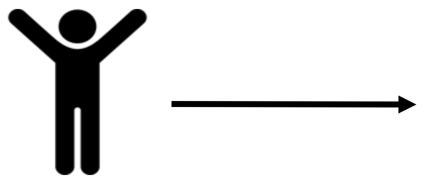


Limitations

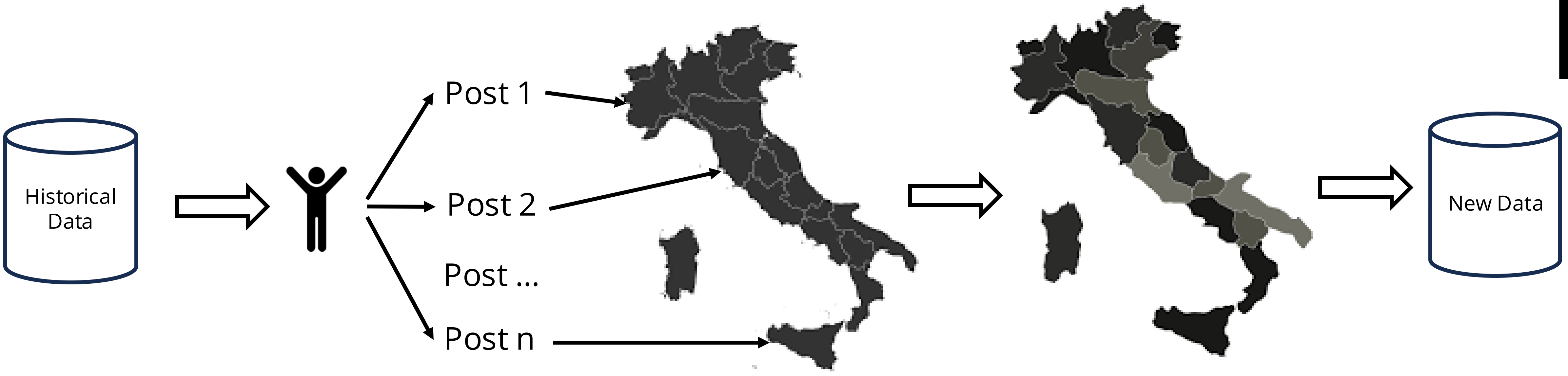
- Can not use **historical** data
- It **cannot** be applied to **certain situations** (e.g., long-term effect, selected demographics, content virality)

BEYOND RANDOMIZED EXPERIMENTS

Associational data



Individual who decides where the post are sent



CAUSAL MODEL FRAMEWORKS

	Potential Outcomes (PO)	Structural Causal Model (SCM)
Antecedents in the earlier econometric literature	Demand and Supply Models (Haavelmo, 1944)	Path analysis (Wright, 1934)

These frameworks are complementary, with different strengths that make them appropriate to address different problems in specific situations.

CAUSAL MODEL FRAMEWORKS

	Potential Outcomes (PO)	Structural Causal Model (SCM)
Antecedents in the earlier econometric literature	Demand and Supply Models (Haavelmo, 1944)	Path analysis (Wright, 1934)
Specifically, to deal with:	Estimating individual-level causal effects	Complex models with a large number of variables

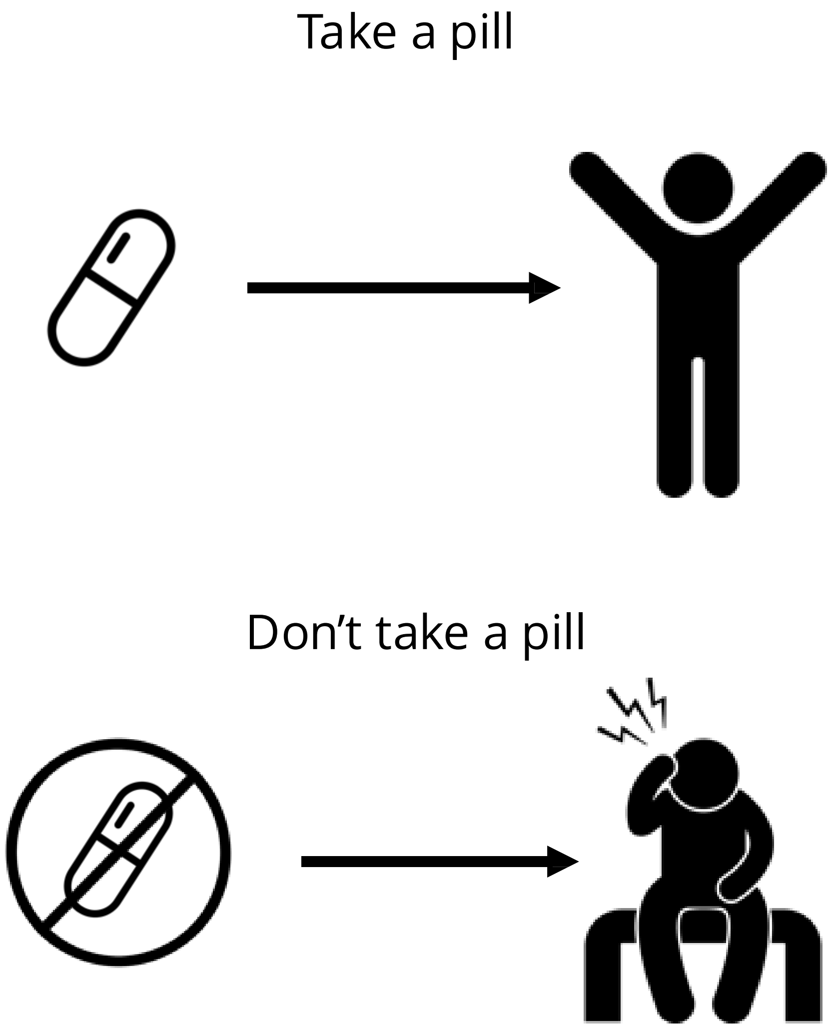
POTENTIAL OUTCOME: INTUITION

Inferring the effect of treatment on some outcome

POTENTIAL OUTCOME: INTUITION

Inferring the effect of treatment on some outcome

Causal Effect?

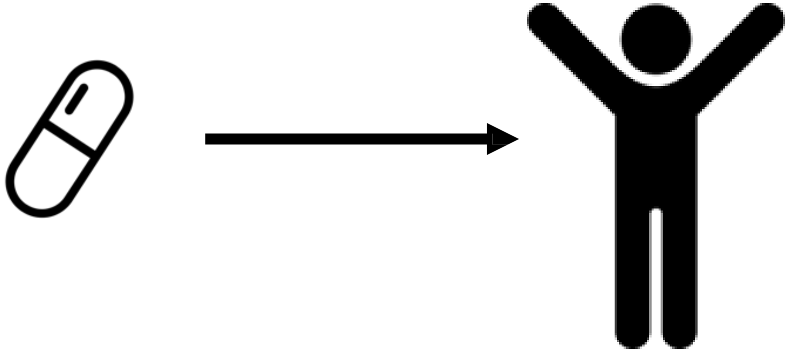


POTENTIAL OUTCOME: INTUITION

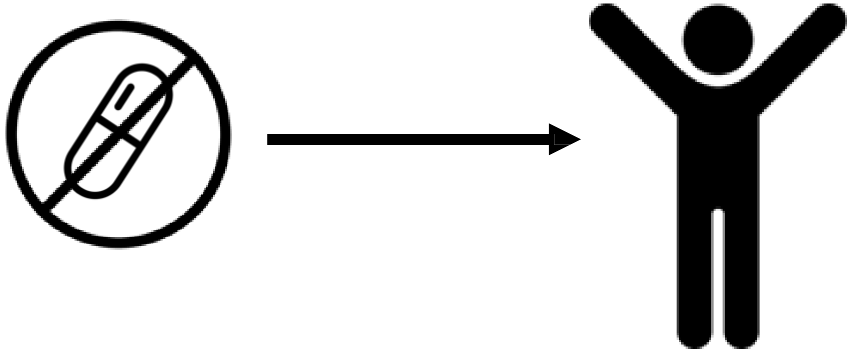
Inferring the effect of treatment on some outcome

No Causal Effect

Take a pill

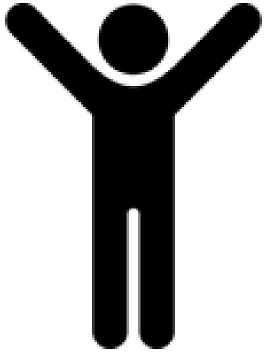


Don’t take a pill



POTENTIAL OUTCOME: NOTATION

$do\,(T = 1)$ $\quad Y_i | do(T = 1) = Y_i(1)$



$do\,(T = 0)$ $\quad Y_i | do(T = 0) = Y_i(0)$

T: Observed Treatment

Y: Observed Outcome

i: used in subscript to denote a specific individual

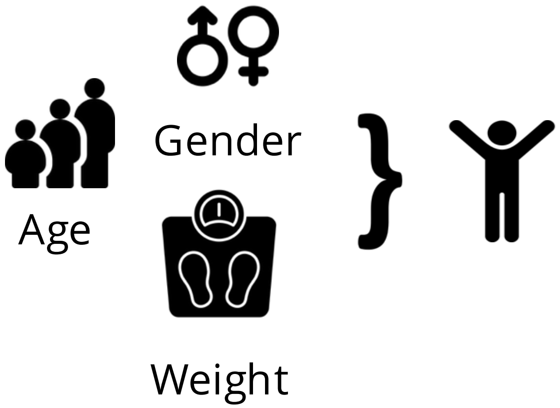
$Y_i(1)$: PO under treatment

$Y_i(0)$: PO under no treatment

OTHER DEFINITIONS

= unit (individual)

= population



= covariates of the individual (Z)

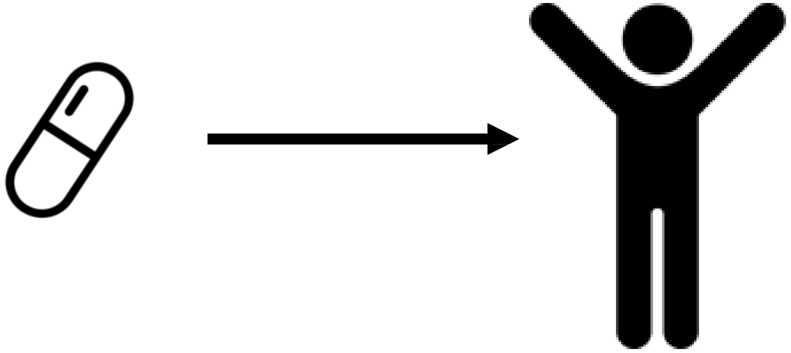
INDIVIDUAL TREATMENT EFFECT (ITE)

The ITE for the i^{th} unit is defined as follows:

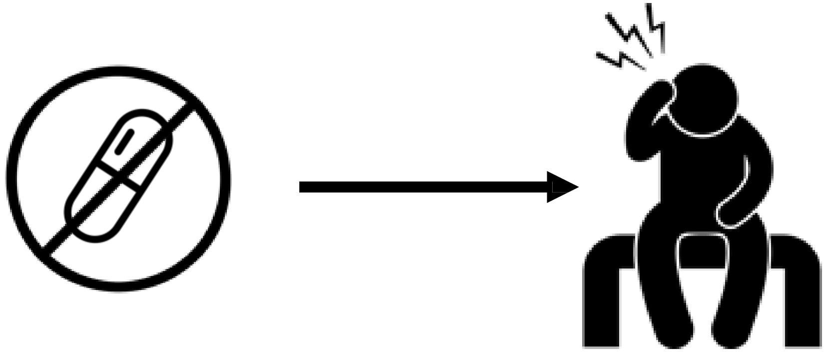
$$Y_i(1) - Y_i(0)$$

POTENTIAL OUTCOME: NOTATION

$do\ (T = 1)$ $Y_i(1) = 1$



$do\ (T = 0)$ $Y_i(0) = 0$



T: Observed Treatment

Y: Observed Outcome

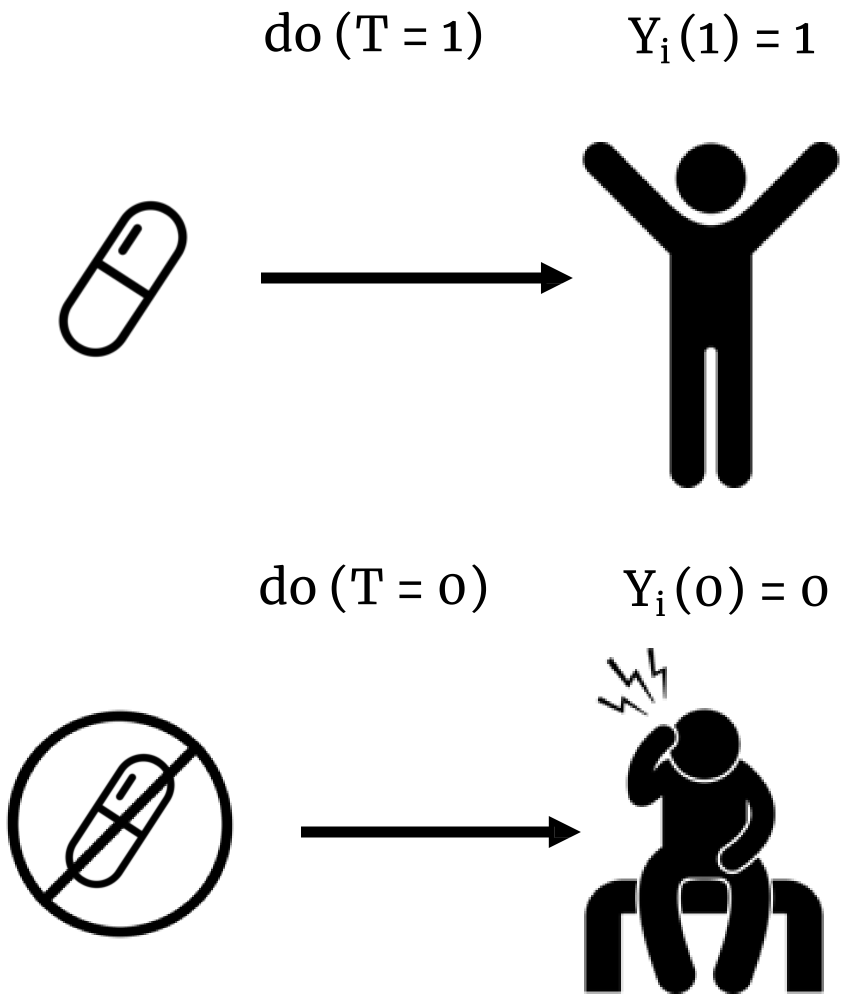
i: used in subscript to denote a specific individual

$Y_i(1)$: PO under treatment

$Y_i(0)$: PO under no treatment

Causal Effect: $Y_i(1) - Y_i(0) = 1$

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE



Fundamental Problem.

We cannot observe both $Y_i(1)$ and $Y_i(0)$, therefore we cannot observe the

Causal Effect: $Y_i(1) - Y_i(0)$

The PO that you do not (and cannot) observe are known as **COUNTERFACTUALS** because they are counter to fact (reality).

Due to the fundamental problem, we know that we can’t access to ITE

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

AVERAGE TREATMENT EFFECT (ATE)

The ATE is obtained by taking an average over the ITEs:

$$E[Y_i(1) - Y_i(0)] = E[Y(1) - Y(0)]$$

where we recall that the average is over the individuals i if $Y_i(x)$ is deterministic.

How would we actually compute the ATE?

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

The question mark means that we do not observe the value

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

$E[Y_i(1) - Y_i(0)] = ?$

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)]$$

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] = E[Y \mid T = 1] - E[Y \mid T = 0]$$

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] = \underline{E[Y \mid T = 1]} - E[Y|T = 0]$$

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0		0	?
2	1	1	1		?
3	1	0	0		?
4	0	0		0	?
5	0	1		1	?
6	1	1	1		?
			2/3		

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] = \underline{E[Y \mid T = 1]} - \underline{E[Y \mid T = 0]}$$

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0		0	?
2	1	1	1		?
3	1	0	0		?
4	0	0		0	?
5	0	1		1	?
6	1	1	1		?
			2/3	1/3	

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] = \underline{E[Y \mid T = 1]} - \underline{E[Y|T = 0]}$$

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0		0	?
2	1	1	1		?
3	1	0	0		?
4	0	0		0	?
5	0	1		1	?
6	1	1	1		?

2/3 - 1/3 = 1/3

The fundamental problem of CI can be seen as a **MISSING DATA PROBLEM**

THE FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

$$E[Y_i(1) - Y_i(0)] = E[Y(1)] - E[Y(0)] \neq E[Y \mid T = 1] - E[Y \mid T = 0]$$

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0		0	?
2	1	1	1		?
3	1	0	0		?
4	0	0		0	?
5	0	1		1	?
6	1	1	1		?

2/3 - 1/3 = 1/3

What does it mean?

causation is not simply association

In general, they are not equal due to **CONFOUNDING**

What **ASSUMPTIONS** would make the ATE equal to the associational difference?

IGNORABILITY - $(Y(1), Y(0)) \perp T$

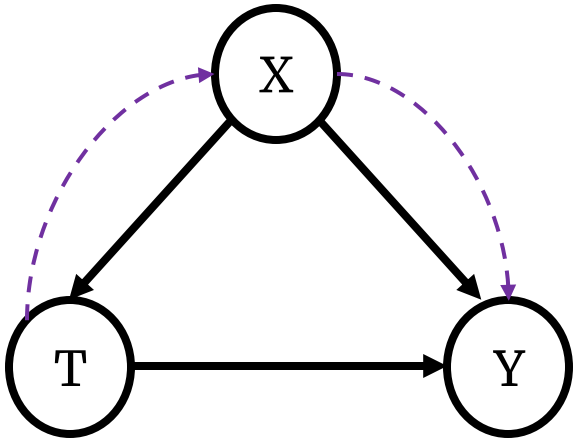
$$E[Y_i(1)] - E[Y_i(0)] = E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0]$$

$$= E[Y \mid T = 1] - E[Y \mid T = 0]$$

- We can ignore how individual ended up in the treatment/control group, and treat their PO as exchangeable. However, it is **unrealistic** in observational data.

- **Unconfoundeness**

$$(Y(1), Y(0)) \perp T \mid X$$



IGNORABILITY - $(Y(1), Y(0)) \perp T$

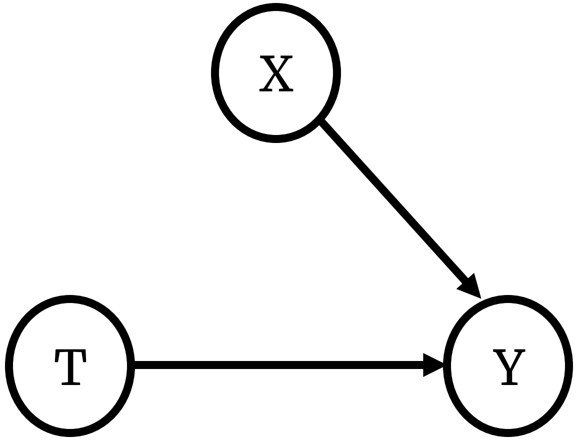
$$E[Y_i(1)] - E[Y_i(0)] = E[Y(1) \mid T = 1] - E[Y(0) \mid T = 0]$$

$$= E[Y \mid T = 1] - E[Y \mid T = 0]$$

- We can ignore how individual ended up in the treatment/control group, and treat their PO as exchangeable. However, it is **unrealistic** in observational data.

- **Unconfoundeness**

$$(Y(1), Y(0)) \perp T \mid \mathbf{X}$$

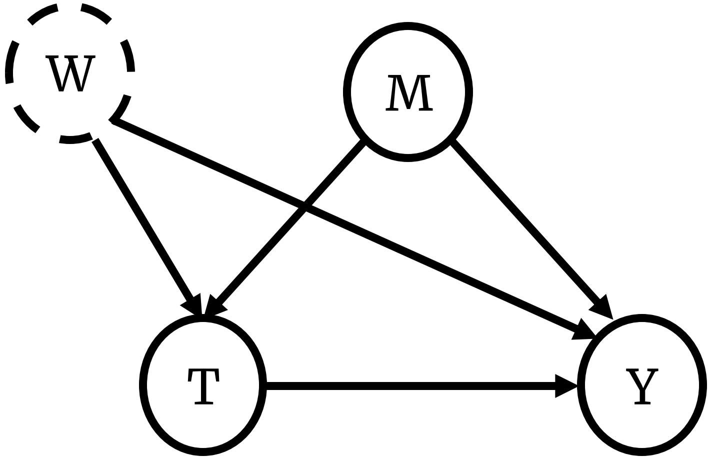


When conditioning on **X**, **non-causal** association between T and Y **no longer exists**.

UNCONFOUNDENESS

- While is not a problem in randomized experiments, it is an **untestable assumption** in observational data

- There may be some **unobserved confounders** that are not part of $X = \{M\}$, meaning unconfoundedness is violated.

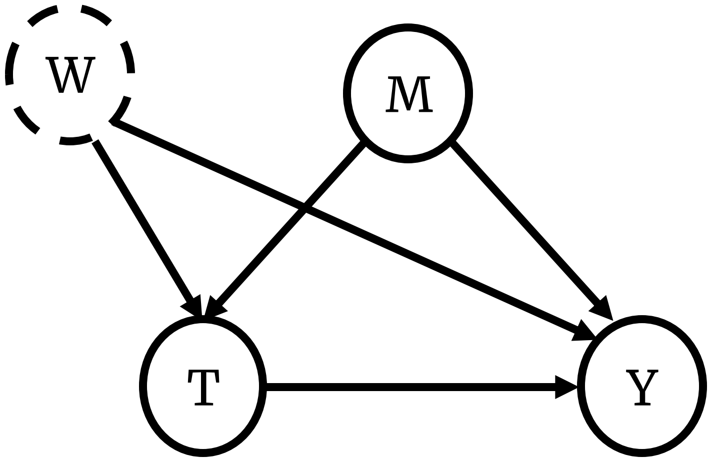


Ignorability

$$(Y(1), Y(0)) \perp T \mid X$$

UNCONFOUNDENESS

- While is not a problem in randomized experiments, it is an **untestable assumption** in observational data
- There may be some **unobserved confounders** that are not part of $X = \{M\}$, meaning unconfoundedness is violated.



Ignorability

$(Y(1), Y(0)) \not\perp T \mid X$

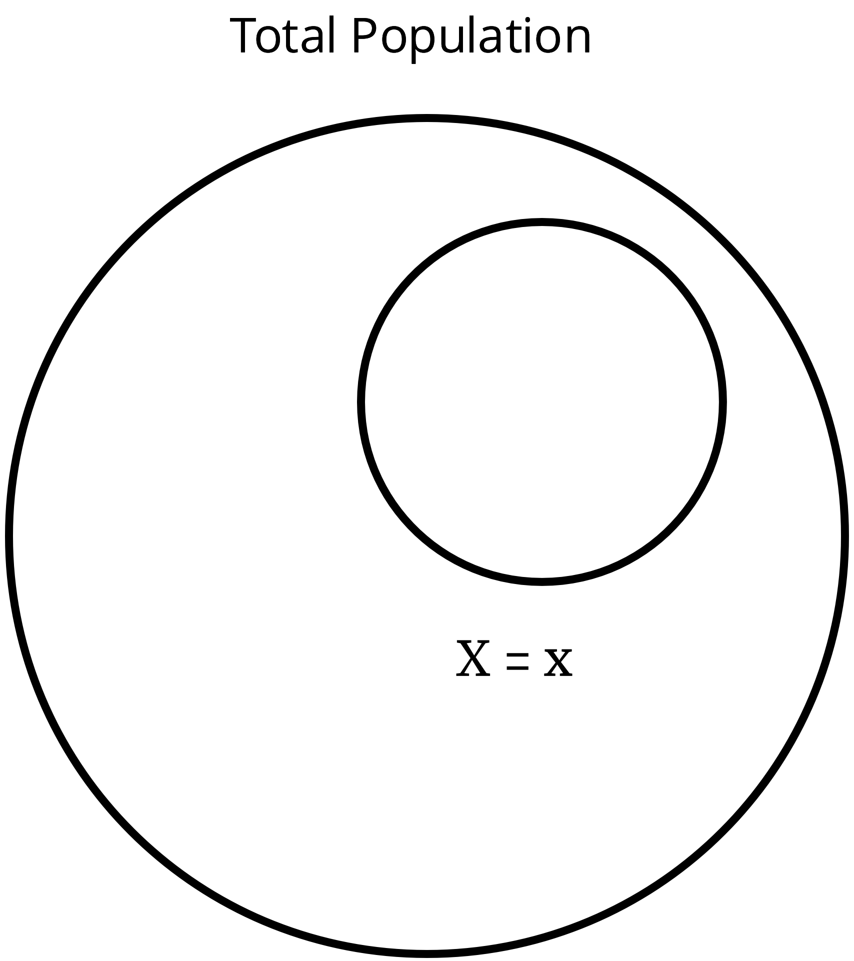
POSITIVITY

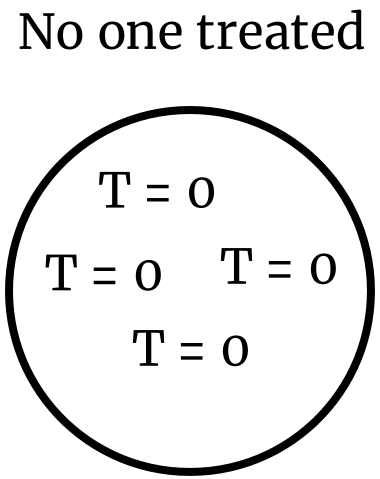
For all values x of covariates $\mathbf{x}$ present in the population of interest (i.e., z such that $P(X = x > 0)$)

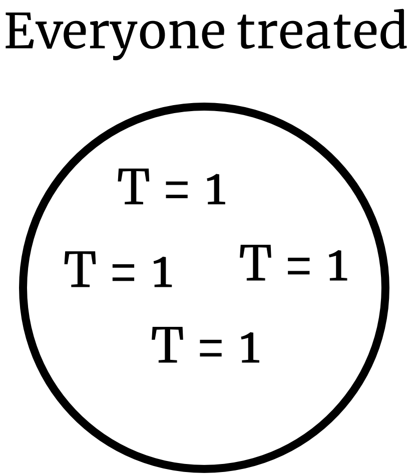
$$0 < P(T = 1 \,|X = x) < 1$$

- Positivity is the condition that **all subgroups of the data** with different value x for covariates $\mathbf{X}$ have some probability of receiving any value of treatment T

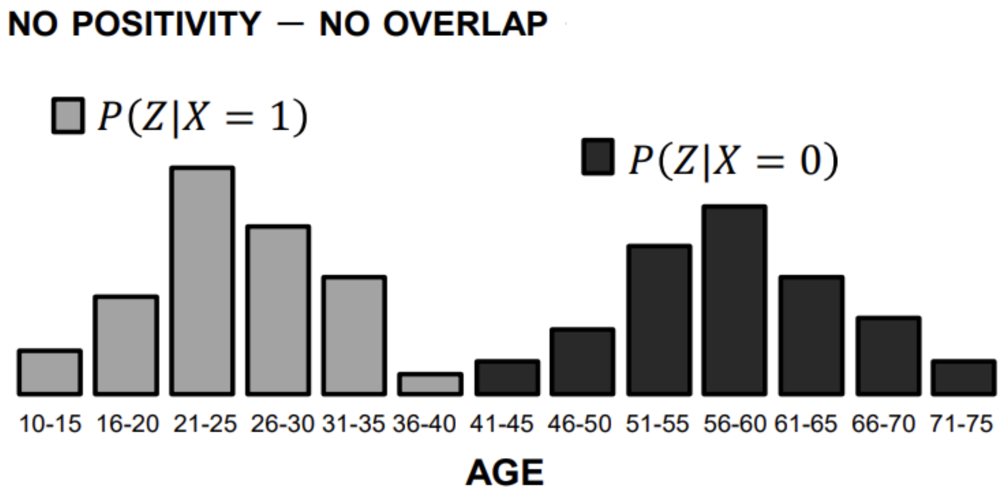
POSITIVITY: INTUITION



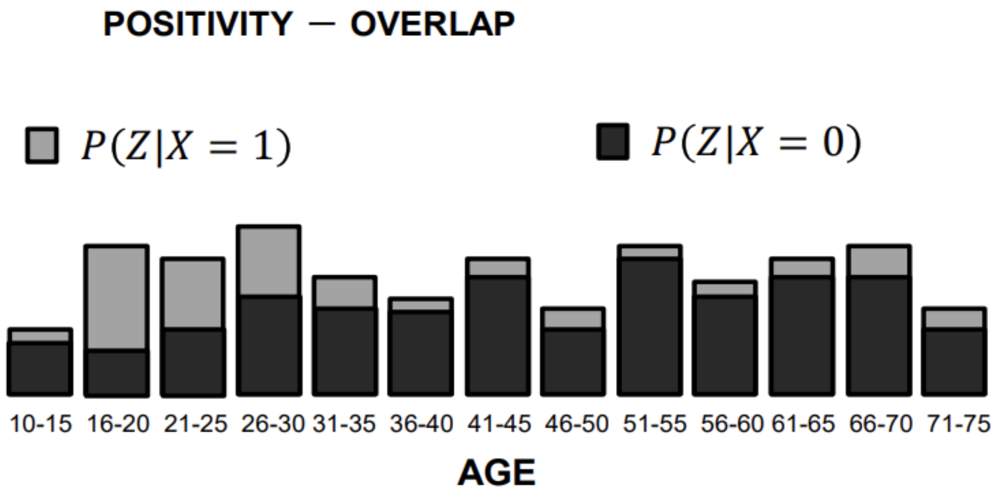




POSITIVITY: OVERLAP



No overlap means severe positivity violation



Complete overlap means no positivity violation adjusting (conditioning) on more covariates $\mathbf{Z}$

could lead to

higher chance of satisfying unconfoundedness

could lead to

higher chance of violating positivity

⇩

demanding too much from models and getting very bad behavior in return

⇩

fit a model to $\mathbb{E}[Y|X, \mathbf{Z}]$ using the available data $(x, y, \mathbf{z})$

increase the "*dimension*" of the covariates $\mathbf{Z}$

⇩

makes the subgroups for any level $\mathbf{z}$ of the covariates $\mathbf{Z}$ smaller

⇩

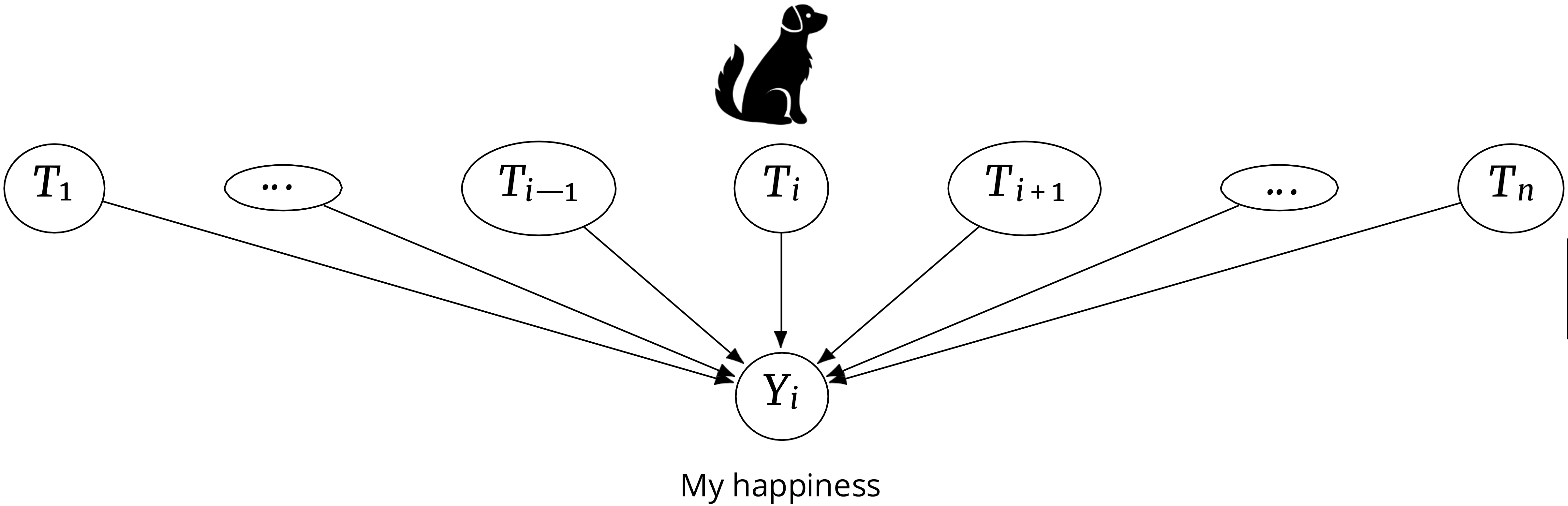
CURSE OF DIMENSIONALITY

NO INTERFERENCE

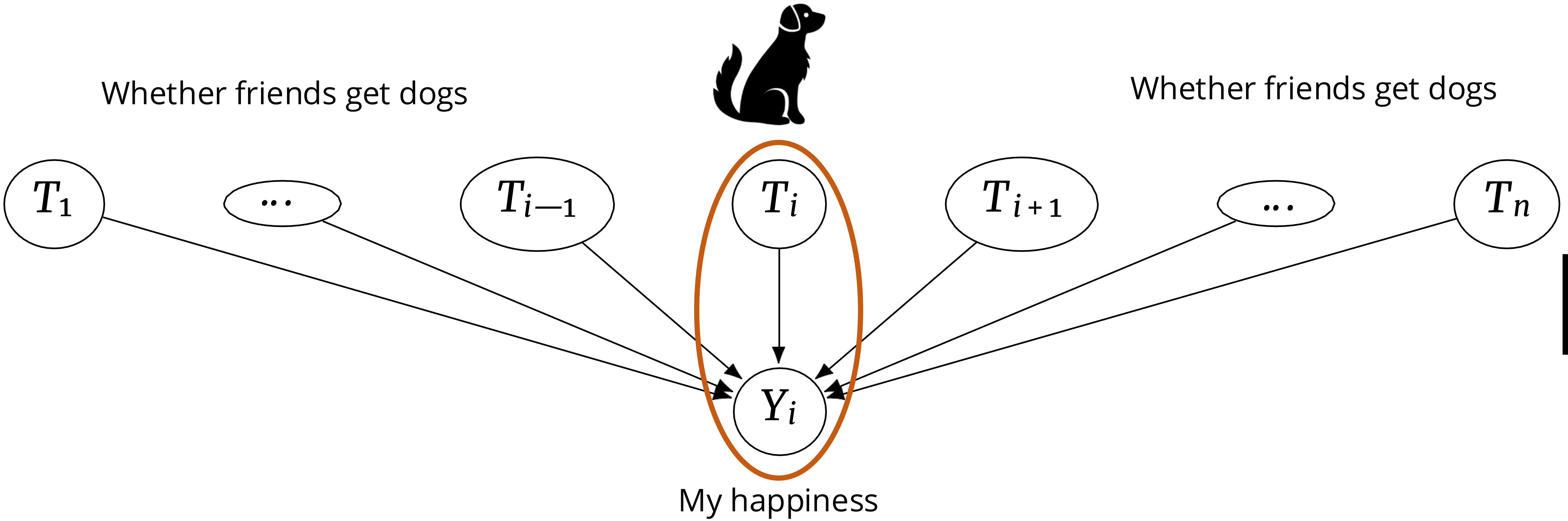
The outcome Y_i of each unit $\mathbf{i}$ is unaffected by anyone else's treatment T_j $j \neq i$

$$Y_i(t_1, t_2, \ldots., t_{i-1}, t_{i+1}, \ldots, t_{n-1}, t_n) = Y_i(t_i)$$

NO INTERFERENCE



NO INTERFERENCE

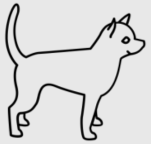


CONSISTENCY

If the treatment is T, then the observed outcome Y is the potential outcome under treatment X.

Formally, $T = t \Rightarrow Y = Y(t)$

$(T = 1) \implies Y = 1$ (I'm happy)



$(T = 1) \implies Y = 0$ (I'm not happy)

Consistency assumption violated

SUTVA

A combination of consistency and no interference. Specifically, the PO of a unit **do not depend** on the treatments assigned to others.

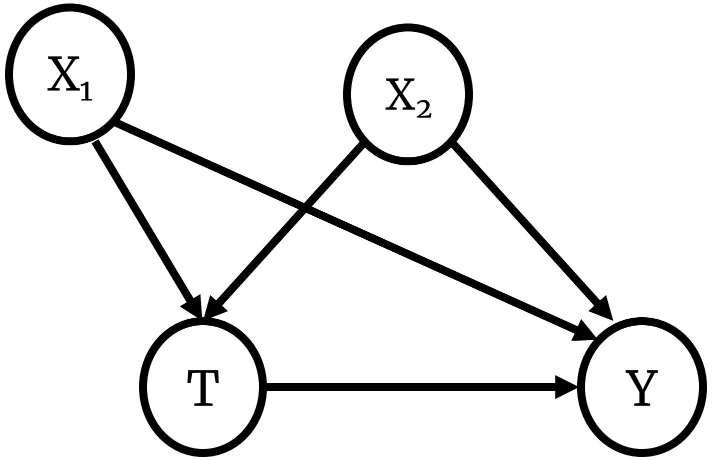
But in real world …

 $(T = 1) \implies Y = 1$ (I'm happy)

HOW TO USE THE PO: AN EXAMPLE

PROPENSITY SCORE MATCHING (PSM)

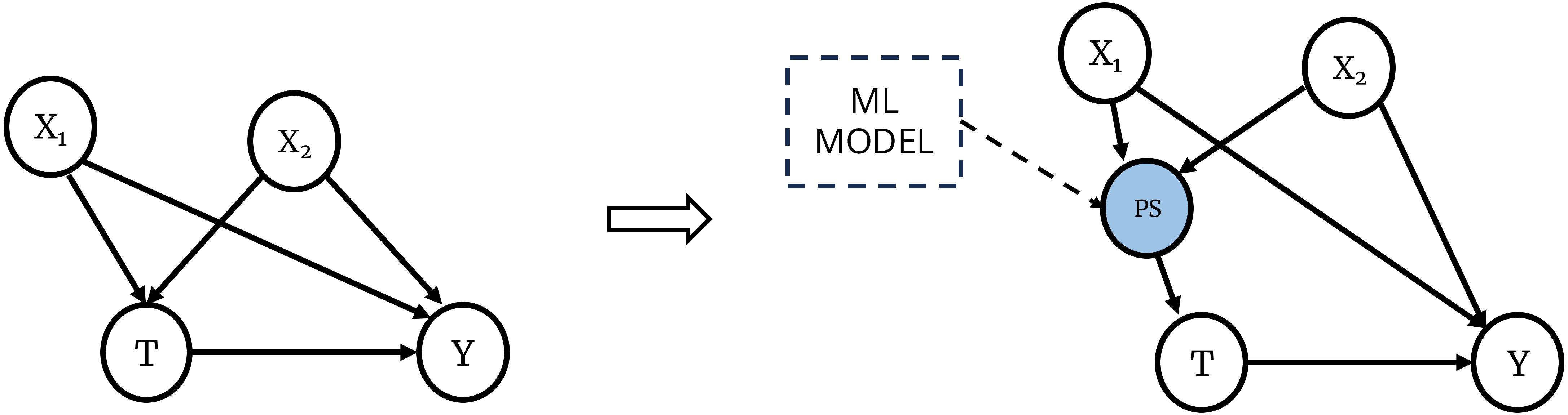
It match $T=0$ and $T=1$ observations on the estimated probability of being treated.



HOW TO USE THE PO: AN EXAMPLE

PROPENSITY SCORE MATCHING (PSM)

It match T=0 and T=1 observations on the estimated probability of being treated.



PO RECAP

- Mainly used for estimating average effects of binary treatments
- Convincing empirical applications

LIMITATIONS:

- An **expert of the field** should **verify** whether **all** the previous **assumptions** are **valid**.

  It is **challenging** and you need **some people working on it**.
- **No** use of causal diagrams

CAUSAL MODEL FRAMEWORKS

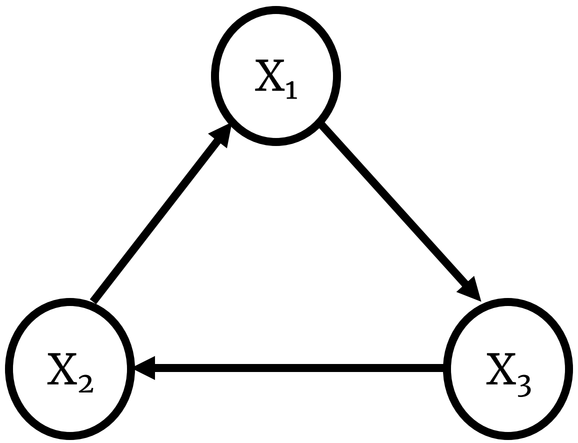
	Potential Outcomes (PO)	Structural Causal Model (SCM)
Antecedents in the earlier econometric literature	Demand and Supply Models (Haavelmo, 1944)	Path analysis (Wright, 1934)
Specifically, to deal with:	Estimating individual-level causal effects	Complex models with a large number of variables

STRUCTURAL CAUSAL MODEL

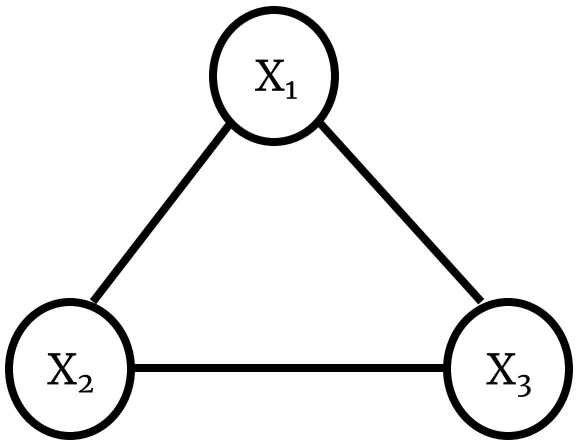
Mathematically, a Structural Causal Model (SCM) consists of **a set of Endogenous (V)** and a set of **Exogenous (U)** variables connected by **a set of functions (F)** that determine the values of the the variables in V based on the values of the variables in U.

Each SCM is associated with a **graphical model** where **each node** is a **variable in V** and each edge is a **function f**.

GRAPH TERMINOLOGY

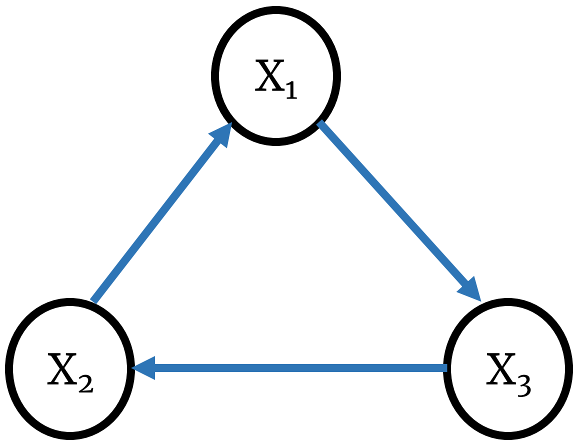


Directed Graph

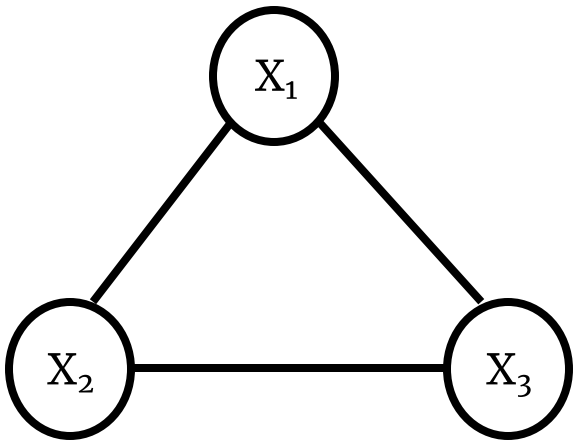


Undirected Graph

GRAPH TERMINOLOGY



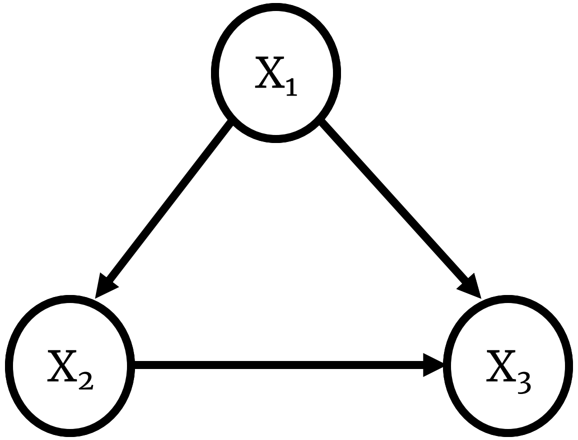
Directed Graph



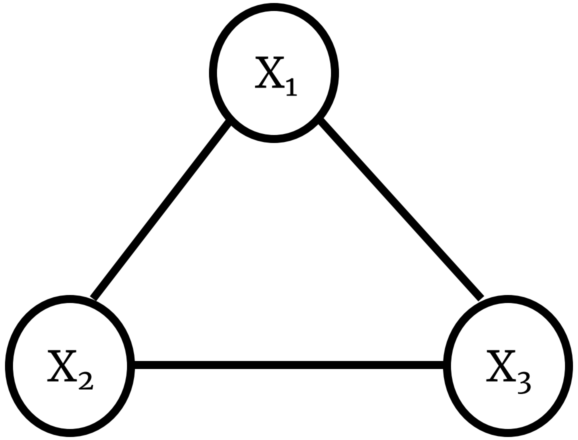
Undirected Graph

This graph contains a cycle

GRAPH TERMINOLOGY

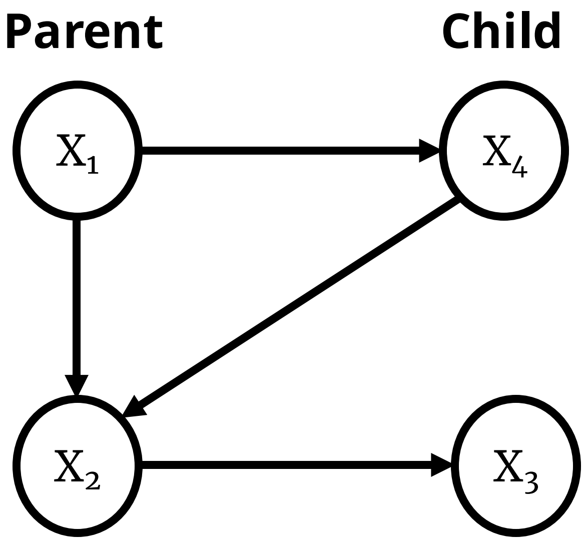


Directed Acyclic Graph

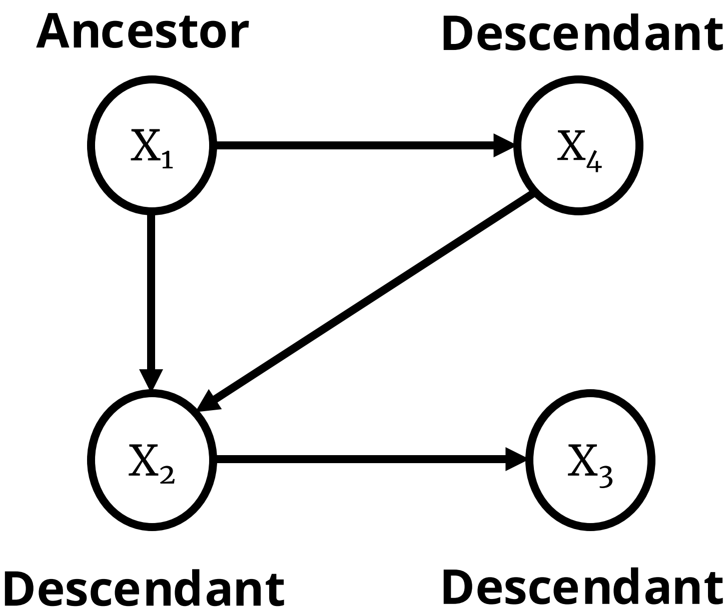


Undirected Graph

GRAPH TERMINOLOGY



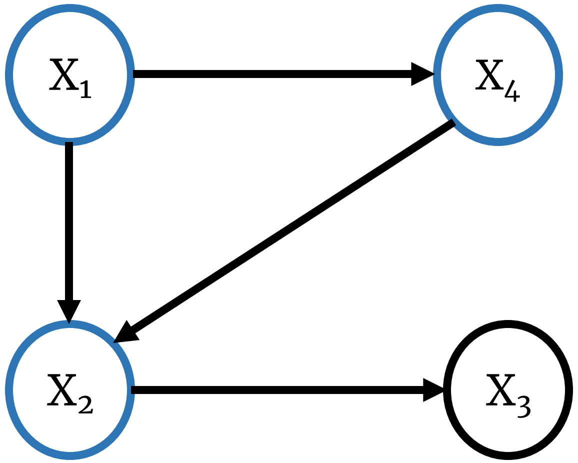
GRAPH TERMINOLOGY



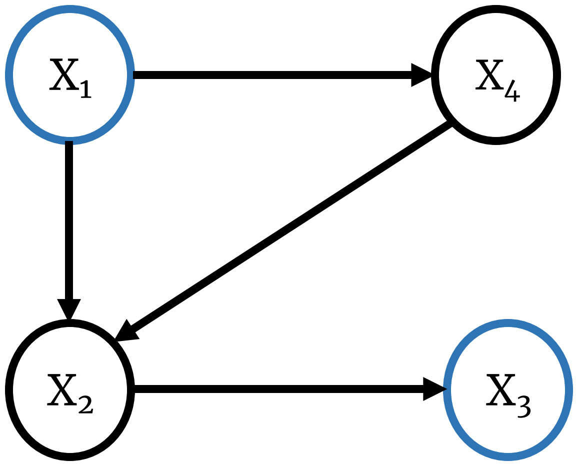
Descendant is a **broader** term than child because it includes **not only the immediate children** but also **their children and so forth**

GRAPH TERMINOLOGY

Adjacent



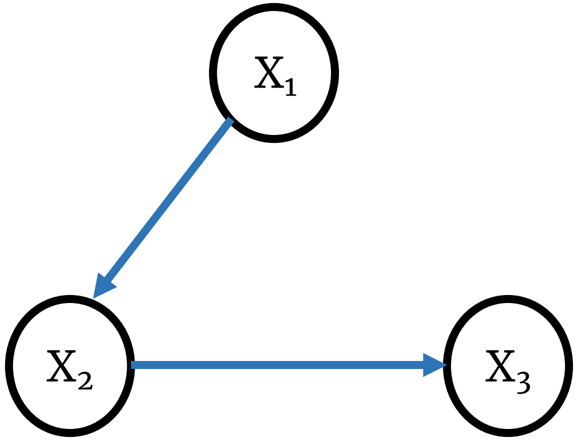
Not Adjacent



Ajdacent is a node that is **directly connected** to another node within a graph

GRAPH TERMINOLOGY

Path



A **path** is a sequence of nodes where each node is connected to the next node by an edge

STRUCTURAL CAUSAL MODEL: EXAMPLE

$\mathbf{X} = \{X_1, X_2, X_3\}$

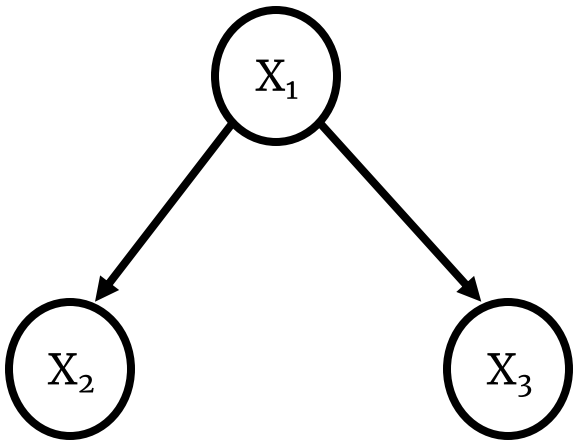
$X_1 := \text{Uniform}(0, 1)$

$X_2 := \sin(X_1) + \text{Normal}(0, 1)$

$X_3 := 2 * X_1 + \text{Normal}(0, 1)$

Structural Equation

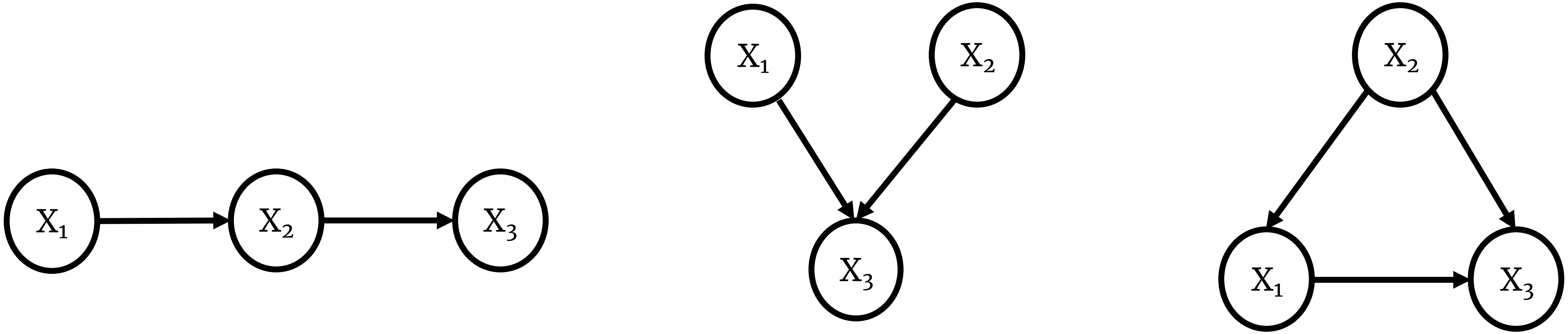
(SE)



Directed Acyclic Graph

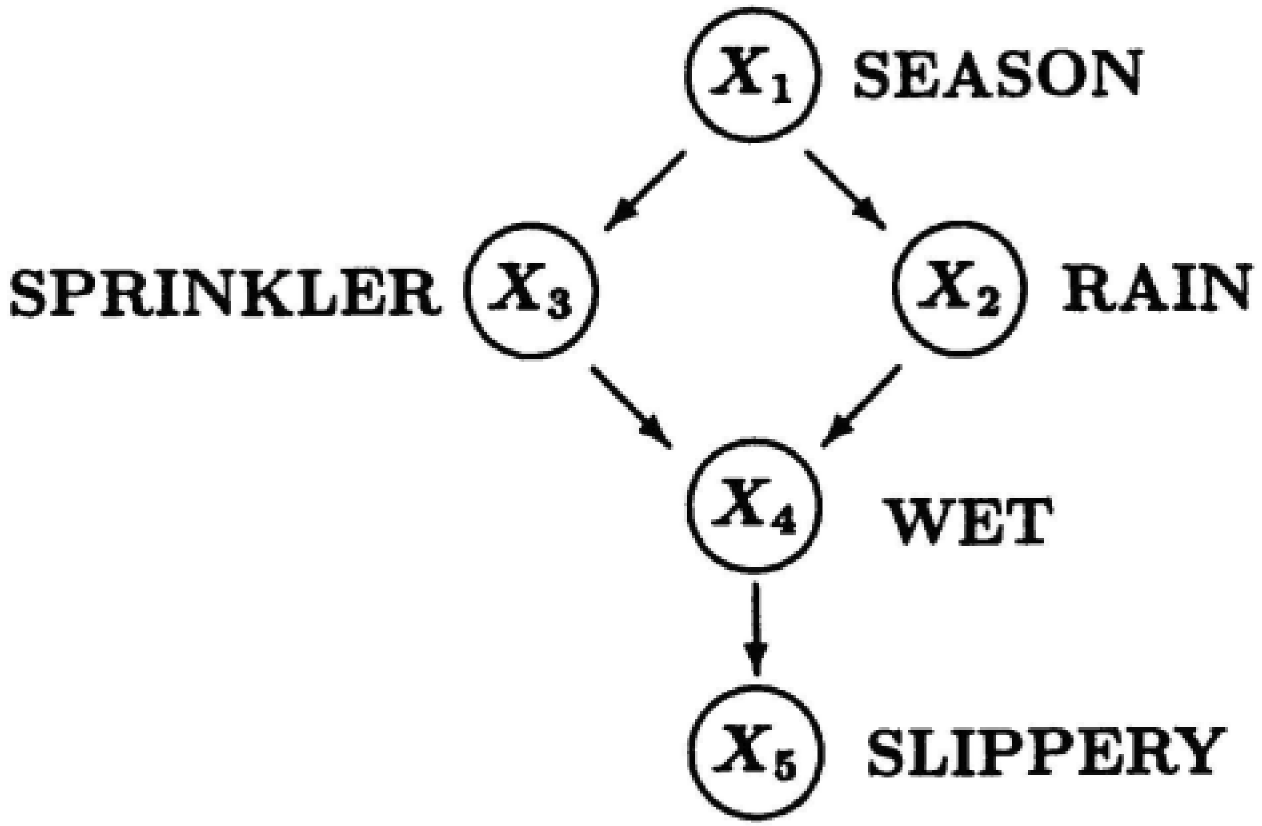
(DAG)

CAUSAL STRUCTURES

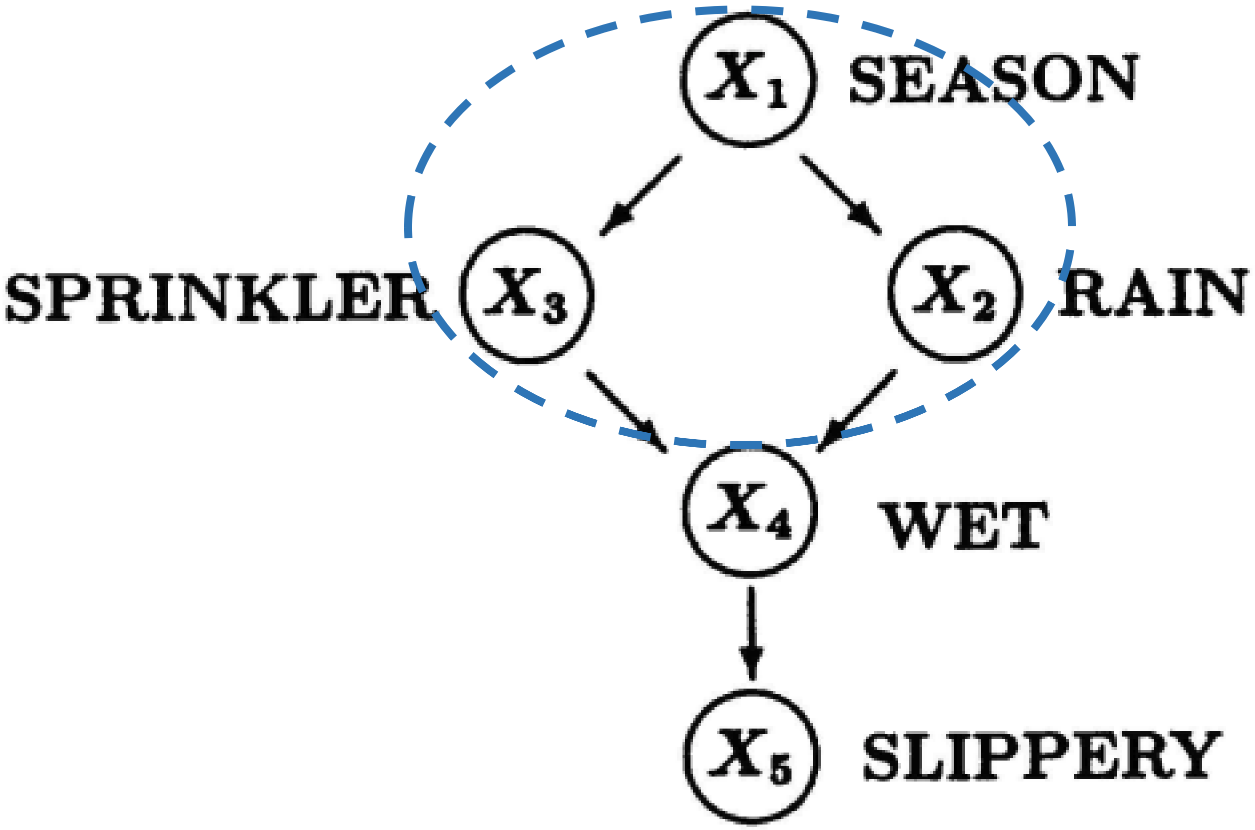


Chain **Collider** **Confounder**

CAUSAL STRUCTURES: EXAMPLE

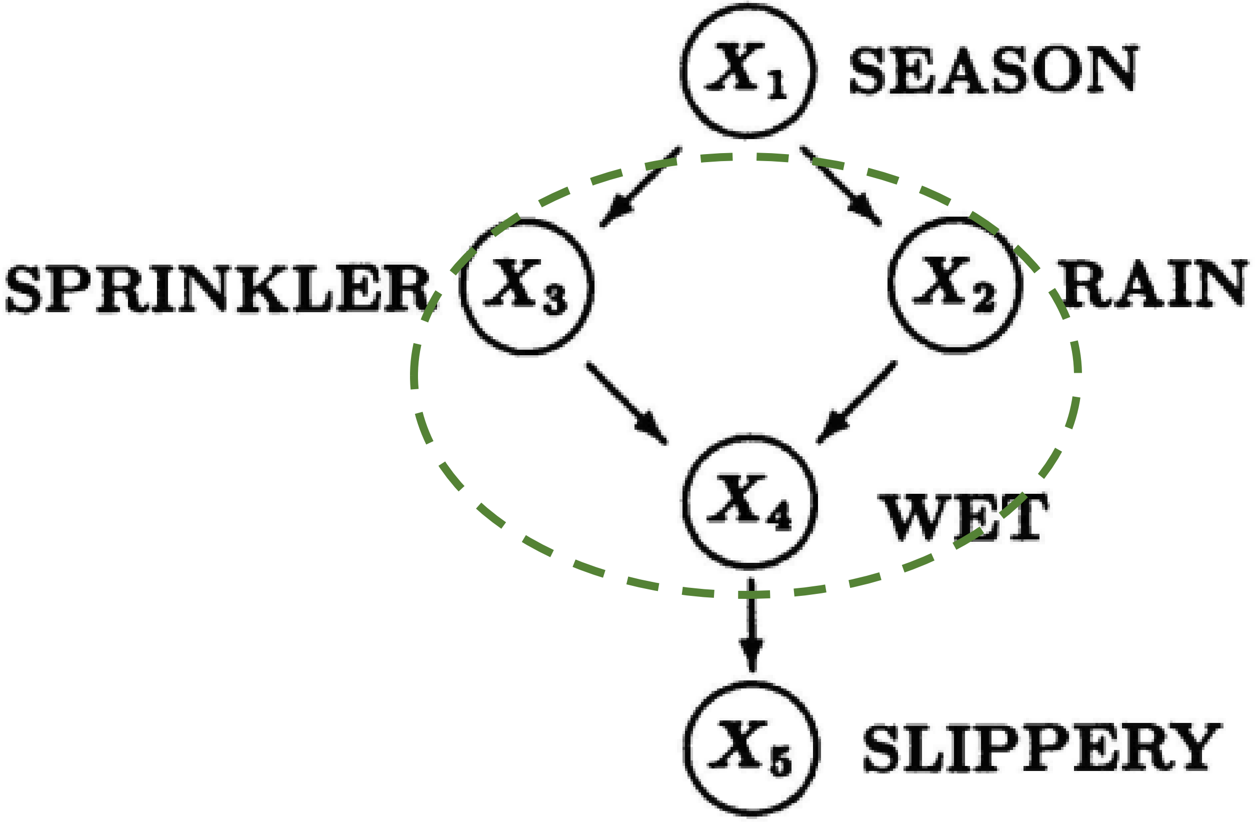


CAUSAL STRUCTURES: EXAMPLE



Confounder

CAUSAL STRUCTURES: EXAMPLE



Collider

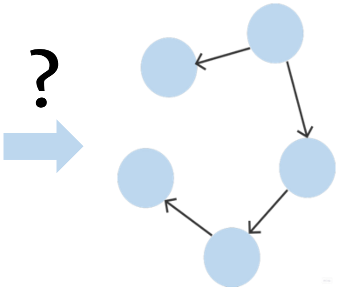
LEVELS OF INVESTIGATION

Causal Discovery (CD)

Given a set of variables,

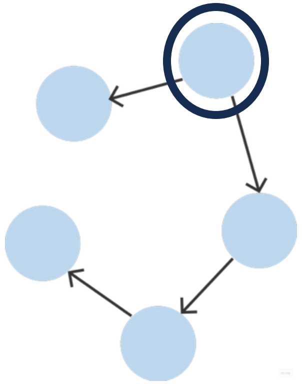
is it possible to **determine the causal relationship** between them?

A	B	C	D	E
3.2	2.2	1.6	7.5	2.4
2.9	3.1	1.3	8.2	5.1
...	...	...	...	...

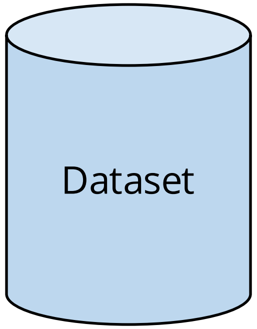


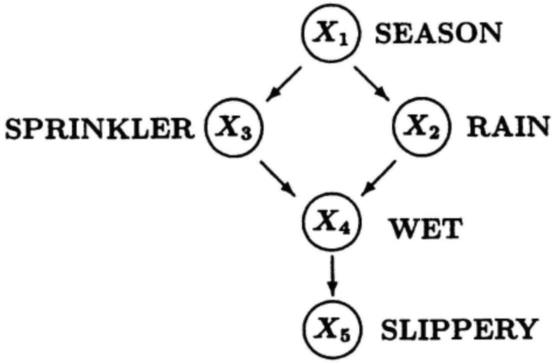
Causal Inference (CI)

If we manipulate the value of one variable, **how much would the others change**?



CAUSAL PIPELINE





Causal Discovery

What are the consequences of turning on the sprinkler?
(The floor gets wet)

Causal Inference

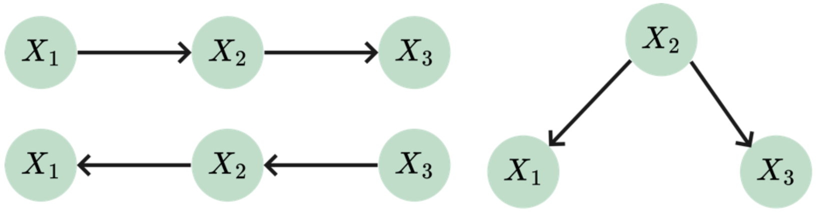
CAUSAL DISCOVERY: METHODS

Constraint-based

Score-based

Markov Equivalence Class

$$X_1 \perp\!\!\!\perp X_3 \mid X_2 \quad and \quad X_1 \not\perp\!\!\!\perp X_3$$



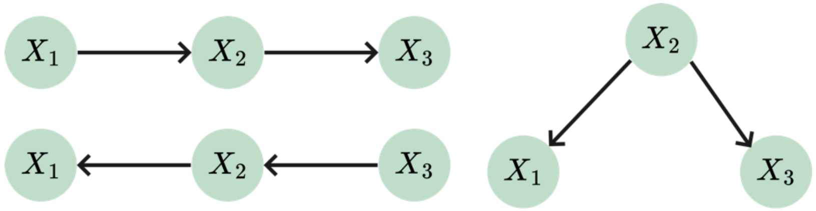
CAUSAL DISCOVERY: METHODS

Constraint-based

Score-based

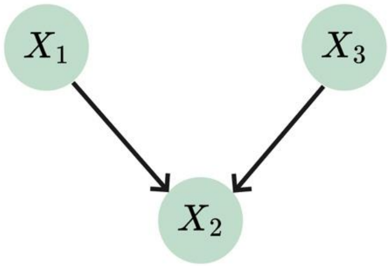
Markov Equivalence Class

$$X_1 \perp\!\!\!\perp X_3 \mid X_2 \quad \textit{and} \quad X_1 \not\!\perp\!\!\!\perp X_3$$



V-structure

$$X_1 \not\!\perp\!\!\!\perp X_3 | X_2, \quad X_1 \perp\!\!\!\perp X_3$$



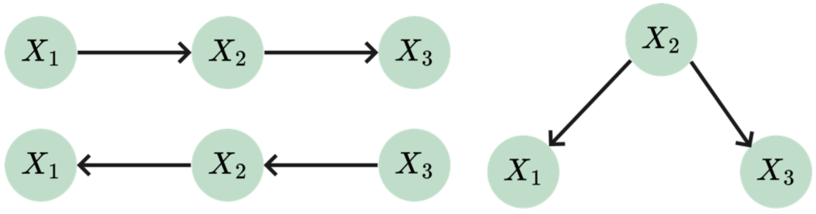
CAUSAL DISCOVERY: METHODS

Constraint-based

Score-based

Markov Equivalence Class

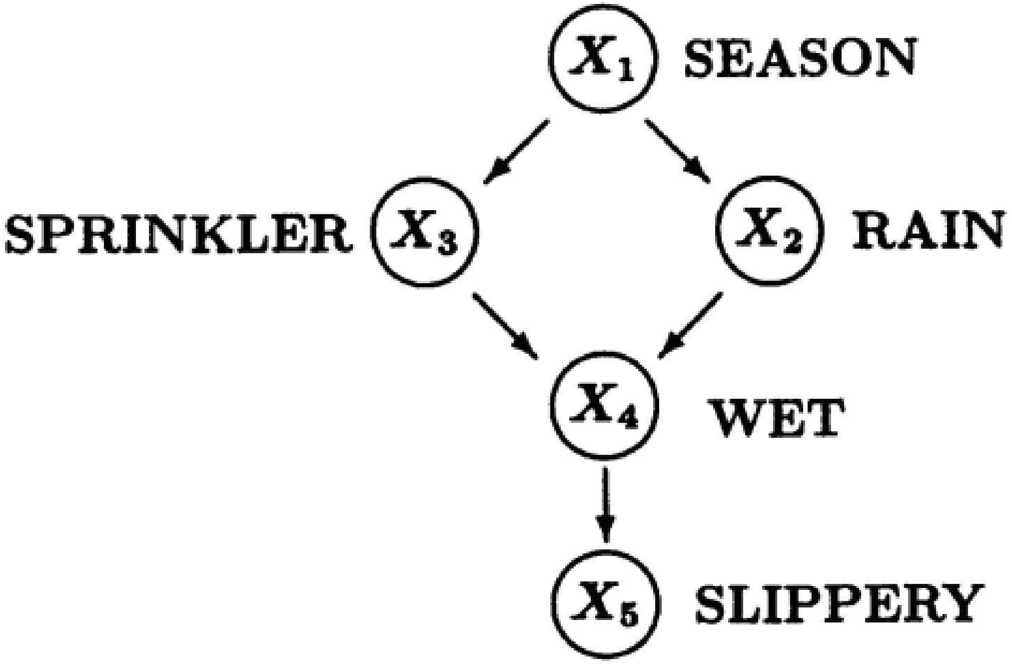
$X_1 \perp\!\!\!\perp X_3 \mid X_2$ *and* $X_1 \not\!\perp\!\!\!\perp X_3$



Functional Causal Models

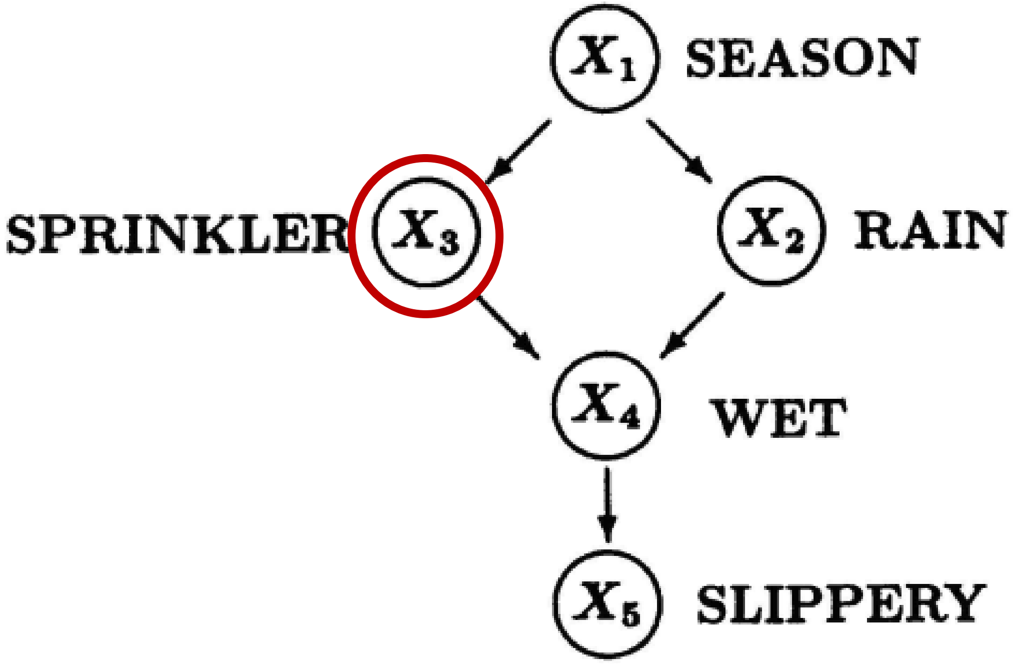
- **Strong** assumptions but they **can uniquely identify the true DAG**
- Linear and non-Gaussian, Additive noise, Post-nonlinear

INTERVENTION



Interpreting edges as cause-effect relationships enable reasoning about the outcome of **interventions** using the **do-operator**

INTERVENTION

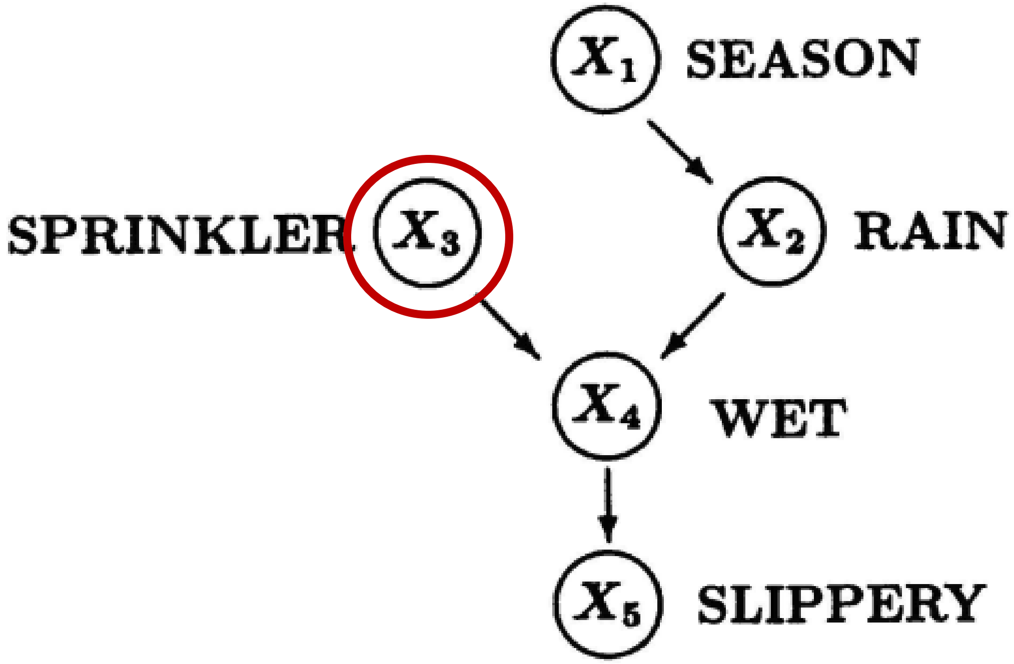


The notation $\mathrm{do}(\mathrm{Sprinkler} := \mathrm{ON})$ denotes an intervention by which variable Sprinkler is set to value ON.

Externally forcing the variable to assume a particular value makes it **independent of its causes** and **breaks their causal influence on it.**

INTERVENTION

Interventional Data



Graphically, the effect of an intervention can be captured by **removing all incoming edges to the intervened variable**.

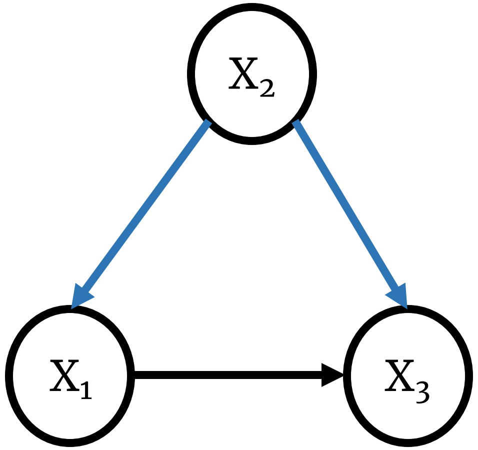
BACK-DOOR CRITERION

The best-known technique to find causal estimands given a graph.

A set of variables $\mathbf{Z}$ satisfies the **back-door criterion** relative to an ordered pair of variables (X_i, X_j) in a DAG G if:

- no node in $\mathbf{Z}$ is a descendant of X_i
- $\mathbf{Z}$ blocks every path between X_i and X_j that contains an arrow into X_i.

BACK-DOOR CRITERION: EXAMPLE



Backdoor path

$X_1 <- X_2 -> X_3$

This path is **not causal**.

It is a process that creates **spurious correlations** between X_1 and X_3 that are driven solely by fluctuations in the X_2 random variable.

If we can **close all of the open backdoor paths**, then we can isolate the causal effect of X_1 and X_3 using an identification strategy.

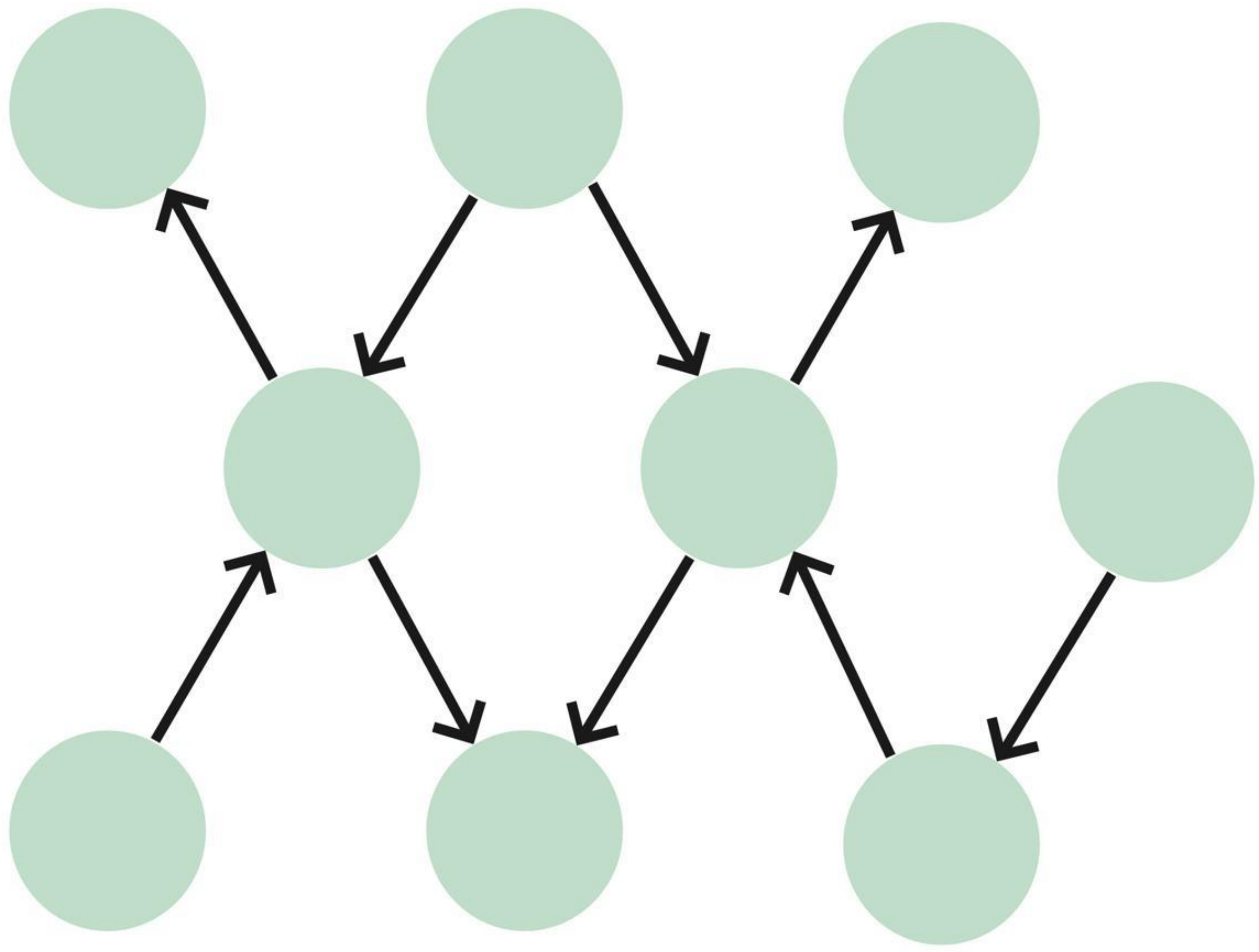
$$P(X_3 \mid do(X_1) = \sum_{X2} P(X_3 \mid X_1, X_2)\, P(X_2)$$

EXERCISE

Find

the discovered

graph



REFERENCES

Pearl, Judea, and Dana Mackenzie. The book of why: the new science of cause and effect. Basic books, 2018.

Imbens, Guido W. "Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics." *Journal of Economic Literature* 58.4 (2020): 1129-1179.

Nogueira, Ana Rita, et al. "Methods and tools for causal discovery and causal inference." *Wiley interdisciplinary reviews: data mining and knowledge discovery* 12.2 (2022): e1449.

Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

https://www.bradyneal.com/causal-inference-course

THANK FOR YOUR ATTENTION