Introduction to Causal Inference

PSY517 Quantitative Analysis III

Derek Powell

Module 4

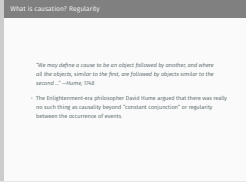
2021-10-05

Introduction to Causal Inference
PSY517 Quantitative Analysis III
Derek Powell
Module 4

What is causation? Regularity

"We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second ..." —Hume, 1748

- The Enlightenment-era philosopher David Hume argued that there was really no such thing as causality beyond "constant conjunction" or regularity between the occurrence of events.



What is causation? Counterfactuals

"We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. ***Or, in other words, where, if the first object had not been, the second never had existed****" —Hume, 1748*

- But he also let slip another view, later developed much further by the philosopher David Lewis (1973), that causality could be defined in terms of ***counterfactuals***.

What is causation? Counterfactuals

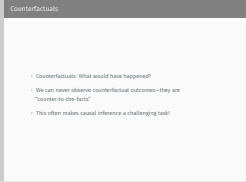
"We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. ***Or, in other words, where, if the first object had not been, the second never had existed****" —Hume, 1748*

- But he also let slip another view, later developed much further by the philosopher David Lewis (1973), that causality could be defined in terms of ***counterfactuals***.

There is still plenty of philosophical debate, but this idea of causality is much closer to our modern conception of it within science.

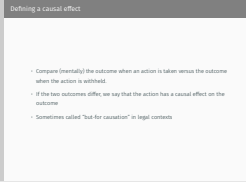
Counterfactuals

- Counterfactuals: What would have happened?
- We can never observe counterfactual outcomes—they are "counter-to-the-facts"
- This often makes causal inference a challenging task!



Defining a causal effect

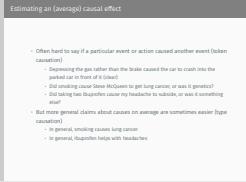
- Compare (mentally) the outcome when an action is taken versus the outcome when the action is withheld.
- If the two outcomes differ, we say that the action has a causal effect on the outcome
- Sometimes called “but-for causation” in legal contexts



- This isn’t perfect, doesn’t identify “proximal” causes
- Could just be an “enabling condition” –e.g. oxygen present when you strike a match
- Can be infinite (useless) causes–by this definition, an asteroid not impacting the earth “caused” Steve McQueen’s cancer b/c it wouldn’t have happened otherwise
- So this is kind of a minimum requirement for a definition

Estimating an (average) causal effect

- Often hard to say if a particular event or action caused another event (token causation)
  - Depressing the gas rather than the brake caused the car to crash into the parked car in front of it (clear)
  - Did smoking *cause* Steve McQueen to get lung cancer, or was it genetics?
  - Did taking two Ibuprofen *cause* my headache to subside, or was it something else?
- But more general claims about causes on average are sometimes easier (type causation)
  - In general, smoking causes lung cancer
  - In general, Ibuprofen helps with headaches



Estimating an (average) causal effect

- There is a causal effet for Zeus if $Y^{a=0}_{Zeus} \neq Y^{a=1}_{Zeus}$
- There is no causal effect for Hera if $Y^{a=0}_{Hera} = Y^{a=1}_{Hera}$
- When estimating the average effects for a group of people or other units of treatment (type causation), we think of causes probabilistically, so there is an effect if:

$$P(Y^{a=0}) \neq P(Y^{a=1})$$

Estimating an (average) causal effect

- There is a causal effet for Zeus if $Y^{a=0}_{Zeus} \neq Y^{a=1}_{Zeus}$
- There is no causal effect for Hera if $Y^{a=0}_{Hera} = Y^{a=1}_{Hera}$
- When estimating the average effects for a group of people or other units of treatment (type causation), we think of causes probabilistically, so there is an effect if:

$$P(Y^{a=0}) \neq P(Y^{a=1})$$

Randomized controlled trials

- Participants (or other study units) are randomly assigned to receive different versions of a treatment (intervention)—often, to receive a treatment or not.
- Allows for direct estimation of the causal effect of a treatment
- *Random assignment* is key: ensures that the groups in each condition will be similar or equivalent *before* the treatment, so that any differences *after* can be attributed to the treatment.

Randomized controlled trials

- Participants (or other study units) are randomly assigned to receive different versions of a treatment (intervention)—often, to receive a treatment or not.
- Allows for direct estimation of the causal effect of a treatment
- *Random assignment* is key: ensures that the groups in each condition will be similar or equivalent *before* the treatment, so that any differences *after* can be attributed to the treatment.

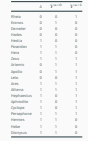
Estimating with a magic wand

Imagine we had a time machine and could observe each person's outcomes under both treatments. Then we could trivially calculate the average causal effect.

	A	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	1
Kronos	0	1	0
Demeter	0	0	0
Hades	0	0	0
Hestia	1	0	0
Poseidon	1	1	0
Hera	1	1	0
Zeus	1	1	1
Artemis	0	1	1
Apollo	0	1	1
Leto	0	0	1
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	0	1
Aphrodite	1	0	1
Cyclope	1	0	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

Estimating with a magic wand

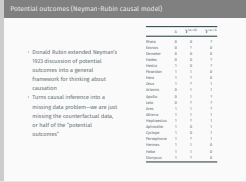
Imagine we had a time machine and could observe each person's outcomes under both treatments. Then we could trivially calculate the average causal effect.



Potential outcomes (Neyman-Rubin causal model)

- Donald Rubin extended Neyman's 1923 discussion of potential outcomes into a general framework for thinking about causation
- Turns causal inference into a missing data problem—we are just missing the counterfactual data, or half of the "potential outcomes"

	A	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	?
Kronos	0	?	0
Demeter	0	0	0
Hades	0	0	?
Hestia	1	0	?
Poseidon	1	1	0
Hera	1	?	0
Zeus	1	?	1
Artemis	0	1	1
Apollo	0	1	?
Leto	0	?	?
Ares	1	1	?
Athena	1	1	1
Hephaestus	1	?	1
Aphrodite	1	0	1
Cyclope	1	0	1
Persephone	1	?	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	?	0

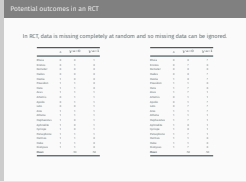


Potential outcomes in an RCT

In RCT, data is missing completely at random and so missing data can be ignored.

	A	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	1
Kronos	0	1	0
Demeter	0	0	0
Hades	0	0	0
Hestia	1	0	0
Poseidon	1	1	0
Hera	1	1	0
Zeus	1	1	1
Artemis	0	1	1
Apollo	0	1	1
Leto	0	0	1
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	0	1
Aphrodite	1	0	1
Cyclope	1	0	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0
Mean		.50	.50

	A	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	0	?
Kronos	0	?	0
Demeter	0	0	0
Hades	0	0	?
Hestia	1	0	?
Poseidon	1	1	0
Hera	1	?	0
Zeus	1	?	1
Artemis	0	1	1
Apollo	0	1	?
Leto	0	?	?
Ares	1	1	?
Athena	1	1	1
Hephaestus	1	?	1
Aphrodite	1	0	1
Cyclope	1	0	1
Persephone	1	?	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	?	0
Mean		.50	.50



- Outside of RCT, we can no longer think of potential outcomes as missing completely at random.
- So, we can't just ignore the missingness and directly estimate causal effects, we need to adjust for it
- Frankly, that gets confusing

Structural Causal Models

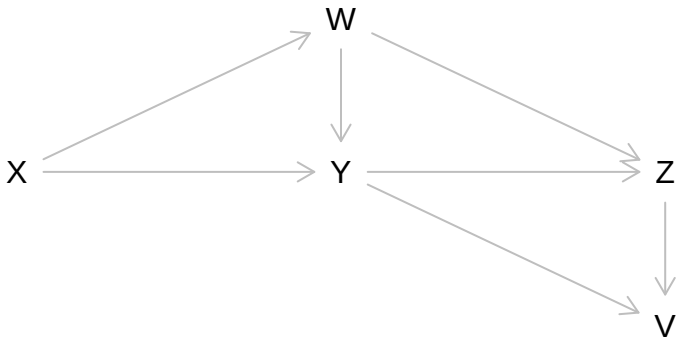
- Introduced by Judea Pearl (2000s and ongoing)
- Represents causal relationships with graphs (diagrams)
- Graphs can be used along with a set of rules called the “do-calculus” to predict outcomes of interventions and determine how to identify causal effects from data

Structural Causal Models

- Introduced by Judea Pearl (2000s and ongoing)
- Represents causal relationships with graphs (diagrams)
- Graphs can be used along with a set of rules called the “do-calculus” to predict outcomes of interventions and determine how to identify causal effects from data

- Have an even better approach: graphs
- lack of arrow is the strong claim (jazz is in the notes that aren’t played)

Directed Acyclic Graphs (DAGs)



- Directed (arrows)
- Acyclic (no cycles or feedback loops)
- Graphs (a diagram!)

Directed Acyclic Graphs (DAGs)

W
X
Y
Z
V

- Directed (arrows)
- Acyclic (no cycles or feedback loops)
- Graphs (a diagram!)

Terminology

- ***Parents***: The nodes with edges flowing in to a given node
- ***Children***: For a given node, the nodes its edges point out toward
- ***Ancestors***: A node's parents, its parents' parents, etc.
- ***Descendants***: A node's children, its children's children, etc.

Terminology

- ***Parents***: The nodes with edges flowing in to a given node
- ***Children***: For a given node, the nodes its edges point out toward
- ***Ancestors***: A node's parents, its parents' parents, etc.
- ***Descendants***: A node's children, its children's children, etc.

Dependence and Independence

- We write $X \perp\!\!\!\perp Y$ to indicate that X and Y are independent.
- $X \perp\!\!\!\perp Y$ if and only if $P(X) = P(X|Y)$ and $P(Y) = P(Y|X)$
- X and Y can also be *conditionally independent* given some other variable(s) Z, which we write $X \perp\!\!\!\perp Y|Z$.

Dependence and Independence
- We write $X \perp\!\!\!\perp Y$ to indicate that X and Y are independent.
- $X \perp\!\!\!\perp Y$ if and only if $P(X) = P(X|Y)$ and $P(Y) = P(Y|X)$
- X and Y can also be conditionally independent given some other variable(s) Z, which we write $X \perp\!\!\!\perp Y|Z$.

d-separation

- Arrows indicate immediate dependencies between nodes
- d-separation (directed separation) is a criteria for determining whether two nodes are independent
- If nodes are independent of each other, we say they are d-separated
  - If X and Y are d-separated, then $X \perp\!\!\!\perp Y$
- If nodes are dependent on one another, we say they are d-connected
  - If X and Y are d-connected, then $X \not\perp\!\!\!\perp Y$
- Nodes can also be conditionally d-separated or conditionally d-connected

d-separation

- Arrows indicate immediate dependencies between nodes
- d-separation (directed separation) is a criteria for determining whether two nodes are independent
- If nodes are independent of each other, we say they are d-separated
  - If X and Y are d-separated, then $X \perp\!\!\!\perp Y$
- If nodes are dependent on one another, we say they are d-connected
  - If X and Y are d-connected, then $X \not\perp\!\!\!\perp Y$
- Nodes can also be conditionally d-separated or conditionally d-connected

Junctions

There are three kinds of elementary structures, or junctions, that might make up larger graphs.

Chain

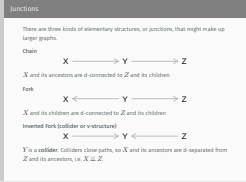
X and its ancestors are d-connected to Z and its children

Fork

X and its children are d-connected to Z and its children

Inverted Fork (collider or v-structure)

Y is a ***collider***. Colliders close paths, so X and its ancestors are d-separated from Z and its ancestors, i.e. $X \perp\!\!\!\perp Z$.



Conditioning

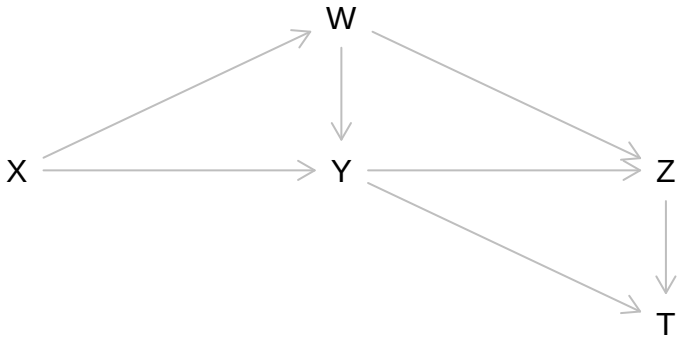
- Conditioning = controlling for, holding constant, or “once I know”.
- Conditioning changes paths, opens or closes them.
- Conditioning on a node at the center of a fork or chain blocks the open path
- Conditioning on a ***collider*** opens the closed path

Conditioning

- Conditioning = controlling for, holding constant, or “once I know”.
- Conditioning changes paths, opens or closes them.
- Conditioning on a node at the center of a fork or chain blocks the open path
- Conditioning on a *collider* opens the closed path

Local Markov property

- Conditional on its parent nodes, a child node is independent of all other ancestors.



Local Markov property

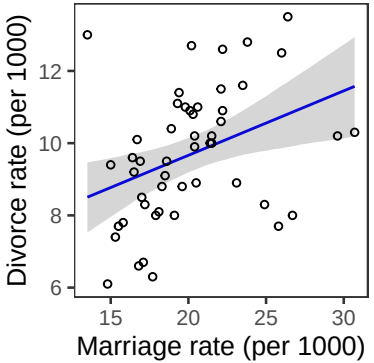
- Conditional on its parent nodes, a child node is independent of all other ancestors.

W
X
Y
Z
T

For T, don't need to know about X and W if you know about Y and Z

Confounding: spurious association

Does marriage cause divorce?

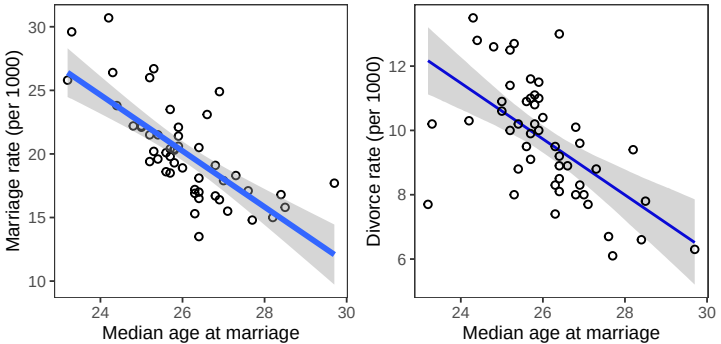


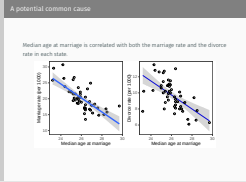
Confounding: spurious association
Does marriage cause divorce?
Divorce rate (per 1000)
Marriage rate (per 1000)

- of course, in some trivial sense it does
- but seems like more marriage could be due to cultural value of marriage, which should mean less divorce right?

A potential common cause

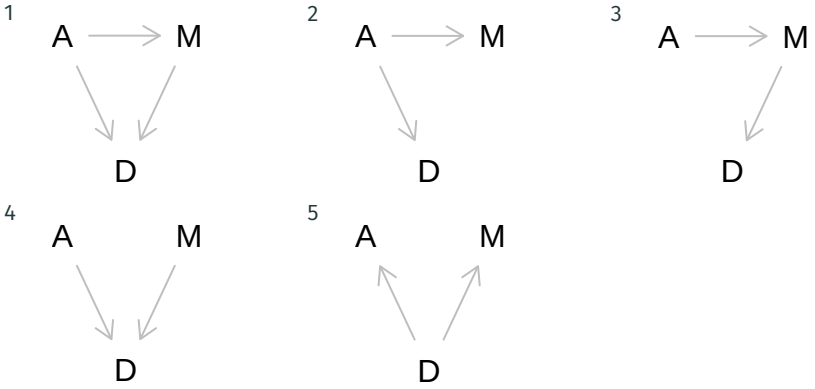
Median age at marriage is correlated with both the marriage rate and the divorce rate in each state.

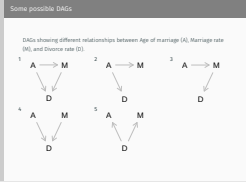




Some possible DAGs

DAGs showing different relationships between Age of marriage (A), Marriage rate (M), and Divorce rate (D).





take 20 seconds and think about it, get an answer in your head

- #1 = two related causes (consistent)
- #2 = common cause / confounded (consistent)
- #3 = full mediation (consistent)
- #4 = independent causes (ruled out because implies $A \perp\!\!\!\perp M$)
- #5 = common effects (seems inconsistent with domain knowledge)

Testing implications with models

Model 1: `D ~ M`

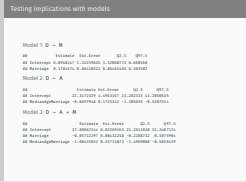
```
##             Estimate  Est.Error       Q2.5    Q97.5
## Intercept 6.0948247 1.32259626 3.52060772 8.668568
## Marriage  0.1784574 0.06418522 0.05463465 0.303502
```

Model 2: `D ~ A`

```
##                       Estimate Est.Error      Q2.5      Q97.5
## Intercept          32.3472329 4.4963167 23.282333 41.2050635
## MedianAgeMarriage -0.8697948 0.1725342 -1.205826 -0.5207514
```

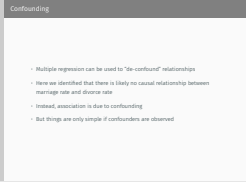
Model 3: `D ~ A + M`

```
##                       Estimate  Est.Error       Q2.5      Q97.5
## Intercept          37.00667244 8.02269265 21.2611038 52.3407134
## Marriage           -0.05712297 0.08432258 -0.2268732  0.1075904
## MedianAgeMarriage -1.00431032 0.25711073 -1.4969008 -0.5028419
```

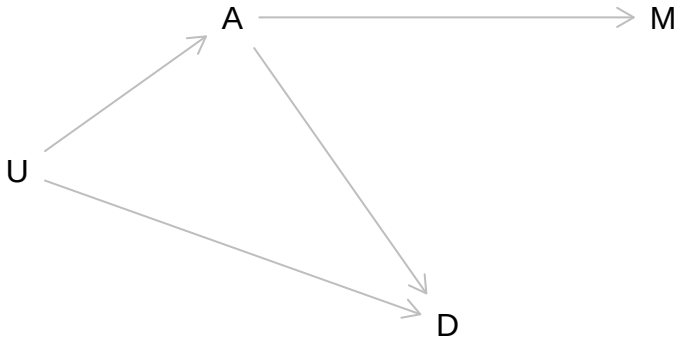
Confounding

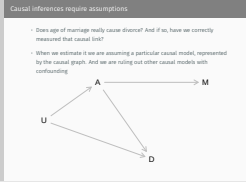
- Multiple regression can be used to “de-confound” relationships
- Here we identified that there is likely no causal relationship between marriage rate and divorce rate
- Instead, association is due to confounding
- But things are only simple if confounders are observed



Causal inferences require assumptions

- Does age of marriage really cause divorce? And if so, have we correctly measured that causal link?
- When we estimate it we are assuming a particular causal model, represented by the causal graph. And we are ruling out other causal models with confounding





The do-operator

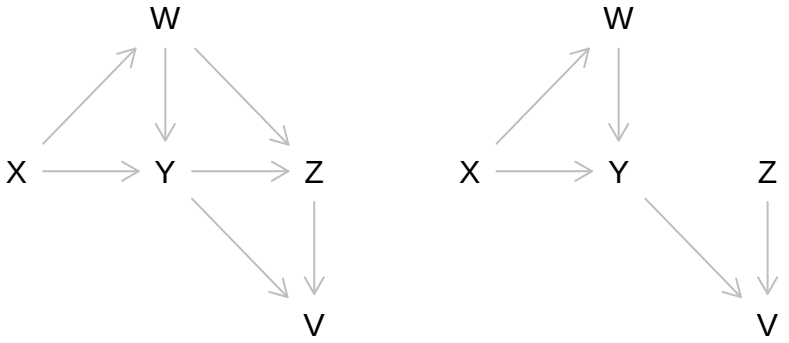
- Just because $X \not\perp\!\!\!\perp Y$, or $P(Y) \neq P(Y|X)$, this does not imply that X causes Y or vice versa.
- We say X has a causal effect on Y if $P(Y) \neq P(Y|do(X))$

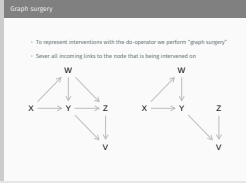
The do-operator

- Just because $X \not\perp\!\!\!\perp Y$ or $P(Y) \neq P(Y|X)$, this does not imply that X causes Y or vice versa.
- We say X has a causal effect on Y if $P(Y) \neq P(Y|do(X))$

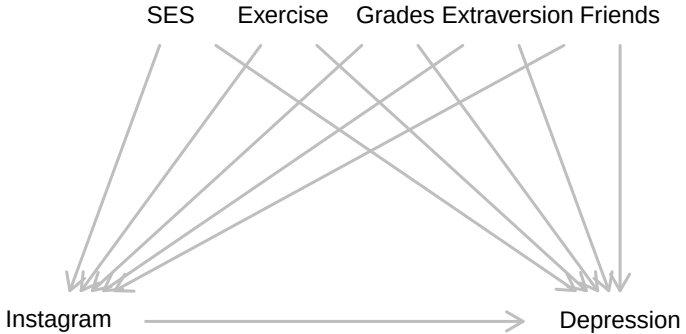
Graph surgery

- To represent interventions with the do-operator we perform “graph surgery”
- Sever all incoming links to the node that is being intervened on

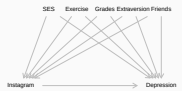




Graph surgery in RCT



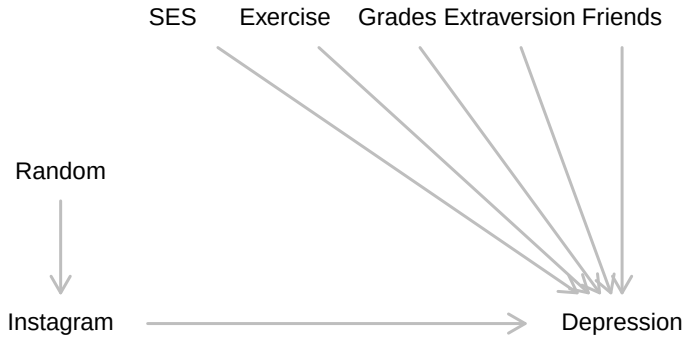
Graph surgery in RCT



Consider effect of instagram use on teenage depression

Graph surgery in RCT

- All parents of manipulated variable are severed, replaced with random assignment (which has no parent causes)



Graph surgery in RCT
All parents of manipulated variable are severed, replaced with random assignment (which has no parent causes)
SES Exercise Grades Extraversion Friends
Random
Instagram
Depression

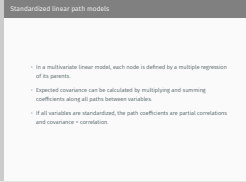
Path analysis

Path analysis

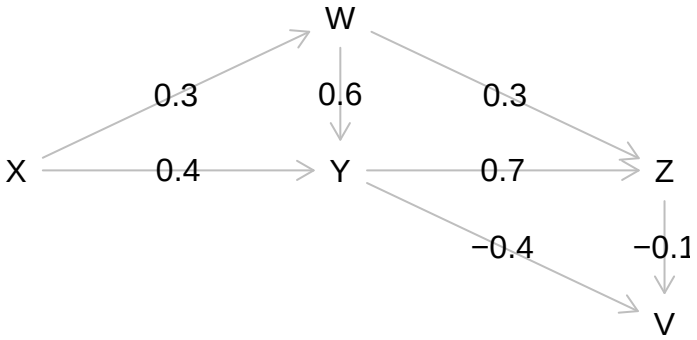
- Introduced by Sewall Wright in a 1920 paper
- A (simpler) form of Structural Equation Modeling, where all variables are directly observed
- A powerful but simple tool for causal inference

Standardized linear path models

- In a multivariate linear model, each node is defined by a multiple regression of its parents.
- Expected covariance can be calculated by multiplying and summing coefficients along all paths between variables.
- If all variables are standardized, the path coefficients are partial correlations and covariance = correlation.



An example of a path model



An example of a path model

Simulating data from a path model

```r
set.seed(13235124)
N <- 1000
d1 <- tibble(
  X = rnorm(N),
  W = .3*X + rnorm(N),
  Y = .4*X + .6*W + rnorm(N),
  Z = .3*W +  .7*Y + rnorm(N),
  V = -.1*Z + -.4*Y + rnorm(N)
)
```


Simulating data from a path model
set.seed(13235124)
N <- 1000
d1 <- tibble(
X = rnorm(N),
W = .3*X + rnorm(N),
Y = .4*X + .6*W + rnorm(N),
Z = .3*W + .7*Y + rnorm(N),
V = -.1*Z + -.4*Y + rnorm(N)
)

Path tracing rules

$\text{cov}(X, Y)$ can be calculated by tracing all paths from X to Y

Rules for standardized variables and coefficients

- Trace backward along an arrow and then forward, or simply forwards from one variable to the other but never forward and then back
- Pass through each variable only once in each chain of paths
- Trace through at most one two-way arrow in each chain of paths

Path tracing rules
$\text{cov}(X, Y)$ can be calculated by tracing all paths from X to Y
Rules for standardized variables and coefficients
- Trace backward along an arrow and then forward, or simply forwards from one variable to the other but never forward and then back
- Pass through each variable only once in each chain of paths
- Trace through at most one two-way arrow in each chain of paths

Calculating some paths

$$\text{cov}(X, Y) = .4 + (.3)(.6) = .58$$

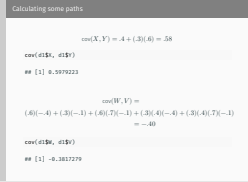
```
cov(d1$X, d1$Y)
```

```
## [1] 0.5979223
```

$$\text{cov}(W, V) =$$
$$(.6)(-.4) + (.3)(-.1) + (.6)(.7)(-.1) + (.3)(.4)(-.4) + (.3)(.4)(.7)(-.1)$$
$$= -.40$$

```
cov(d1$W, d1$V)
```

```
## [1] -0.3817279
```

DRAW THIS OUT ON THE BOARD

Note: I am being a bit lazy as I these are not truly standardized data/coefficients, but they are close enough to illustrate the point without introducing further complications

Paths to causes

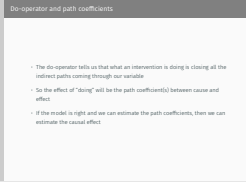
"Ok, so this is all great and everything, but it's all just about correlations! What does it mean for causation?"—You, possibly

- The do-operator tells us that, if the model is right, the path coefficients represent causal effects

Paths to causes
"Ok, so this is all great and everything, but it's all just about correlations! What does it mean for causation?"—You, possibly
The do-operator tells us that, if the model is right, the path coefficients represent causal effects

Do-operator and path coefficients

- The do-operator tells us that what an intervention is doing is closing all the indirect paths coming through our variable
- So the effect of “doing” will be the path coefficient(s) between cause and effect
- If the model is right and we can estimate the path coefficients, then we can estimate the causal effect



Back-door criterion

- To deconfound two variables X and Y, must block every noncausal path between them without affecting any causal paths
- A ***back-door path*** from X to Y is any path from X to Y that starts with an arrow pointing into X.
- One way to deconfound two variables is to find and condition on a set of variables Z that blocks all of the back-door paths between X and Y.
- Need to be careful that Z does not affect any causal paths between X and Y.

Back-door criterion
- To deconfound two variables X and Y, must block every noncausal path between them without affecting any causal paths
- A ***back-door path*** from X to Y is any path from X to Y that starts with an arrow pointing into X.
- One way to deconfound two variables is to find and condition on a set of variables Z that blocks all of the back-door paths between X and Y.
- Need to be careful that Z does not affect any causal paths between X and Y.

- can do an example with the paths from the path tracing
- two backdoor paths, both blocked by conditioning on X.
- what about conditioning on Y instead? No good, blocks a causal path

Extra slides

Conditioning on a collider

- Conditioning on a ***collider*** opens the path
- Sometimes called ***Berkson's Paradox***, but it's better thought-of as "explaining away"