In-person session 6

February 20, 2025

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

FAQs

Potential outcomes and do()

do-calculus and adjustment

p-values and confidence intervals

FAQs

DAGs vs. Logic models

DAGs are a *statistical* tool

Describe a data-generating process and isolate/identify relationships

Logic models are a *managerial* tool

Oversee the inner workings of a program and its theory

What exactly is a data generating process?

Can we make another DAG together?

The opera!

Randomness

How do we use random.org for things in R?

Are the results from p-hacking actually a threat to validity?

Is a little exploratory p-hacking okay?

Do people actually post their preregistrations?

Yes!

OSF

See this and this for examples

As Predicted

See this

Do you have any tips for identifying the threats to validity in articles since they're often not super clear?

Especially things like spillovers, Hawthorne effects, and John Henry effects?

Using a control group of some kind seems to be the common fix for all of these issues.

What happens if you can't do that? Is the study just a lost cause?

Potential outcomes vs. do() notation

Expectations

$$\mathrm{E}(\cdot), \mathbf{E}(\cdot), \mathbb{E}(\cdot) \quad \text{vs.} \quad \mathrm{P}(\cdot)$$

Basically a fancy way of saying "average"

Potential outcomes and CATEs example

Why can't we just subtract the averages between treated and untreated groups?

When you're making groups for CATE, how do you decide what groups to put people in?

Slides from lecture

Unconfoundedness assumption

How can we assume/pretend that treatment was randomly assigned within each age?

It seems unlikely. Wouldn't there be other factors within the older/younger group that make a person more/less likely to engage in treatment (e.g., health status)?

Slides from lecture

Causal effects with potential outcomes

Potential outcomes notation:

$$\delta \;= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} Y_i(1) - Y_i(0)$$

or alternatively with $\mathbf{E}$

$$\delta \;= \mathbf{E}[Y_i(1) - Y_i(0)]$$

Causal effects with do()

$$\textbf{Pearl notation:}$$

$$\delta \ = \mathbf{E}[Y_i \mid \text{do}(X = 1) - Y_i \mid \text{do}(X = 0)]$$

$$\text{or more simply}$$

$$\delta \ = \mathbf{E}[Y_i \mid \text{do}(X)]$$

$$\mathbf{E}[Y_i \mid \mathrm{do}(X)]$$

$$=$$

$$\mathbf{E}[Y_i(1) - Y_i(0)]$$

We can't see this

$$\mathbf{E}[Y_i \mid \text{do}(X)] \quad \text{or} \quad \mathbf{E}[Y_i(1) - Y_i(0)]$$

So we find the average causal effect (ACE)

$$\hat{\delta} = \mathbf{E}[Y_i \mid X = 1] - \mathbf{E}[Y_i \mid X = 0]$$

$$\overbrace{\mathbf{E}(y \mid \mathrm{do}(x))}^{\text{The average population-level change in } y \text{ when } \textit{directly intervening} \text{ (or doing) } x} \neq \overbrace{\mathbf{E}(y \mid x)}^{\text{The average population-level change in } y \text{ when accounting for } \textit{observed } x}$$

Causation

Correlation

do-calculus and adjustment

DAGs and identification

DAGs are a statistical tool, but they don't tell you what statistical method to use

DAGs help you with the identification strategy

Thomas Massie
@RepThomasMassie

Over 70% of Americans who died with COVID, died on Medicare, and some people want #MedicareForAll ?

11:00 AM · Feb 9, 2022 · Twitter for iPhone

Easist identification

Identification through research design

RCTs

When treatment is randomized, delete all arrows going into it

No need for any do-calculus!

Most other identification

Identification through do-calculus

Rules for graph surgery

Backdoor adjustment and frontdoor adjustment are special common patterns of do-calculus

Where can we learn more about *do*-calculus?

Here!

The do-calculus Let G be a CGM, $G_{\overline{T}}$ represent G post-intervention (i.e with all links into T removed) and $G_{\underline{T}}$ represent G with all links *out of* T removed. Let $do(t)$ represent intervening to set a single variable T to t,

Rule 1: $\mathbb{P}(y|do(t), z, w) = \mathbb{P}(y|do(t), z)$ if $Y \perp\!\!\!\perp W|(Z,T)$ in $G_{\overline{T}}$

Rule 2: $\mathbb{P}(y|do(t), z) = \mathbb{P}(y|t, z)$ if $Y \perp\!\!\!\perp T|Z$ in $G_{\underline{T}}$

Rule 3: [illegible] if $Y \perp\!\!\!\perp T|Z$ in $G_{\overline{T}}$, and Z i[illegible]

Rule 1 (Insertion/deletion of observations):

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}} \qquad (3)$$

Rule 2 (Action/observation exchange):

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}} \qquad (4)$$

Rule 3 (Insertion/deletion of act[illegible]

where [illegible] tors o[illegible]

Supplement 2. The *do*-calculus

The *do*-calculus is an axiomatic system for replacing probability formu[illegible] operator with ordinary conditional probabilities. It consists of three axiom schemas th[illegible] graphical criteria for when certain substitutions may be made.

Where $\mathbf{G}$ is the ADMG on variable set $\mathbf{V}$, and P satisfies (MC – *d*-separation), the rules [illegible]

Rule 1 (Insertion/deletion of observations)

$P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{Z}, \mathbf{W}) = P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{W})$ if $\mathbf{Y}$ and $\mathbf{Z}$ are *d*-separated by $\mathbf{X} \cup \mathbf{W}$ in $\mathbf{G}^*$, [illegible] obtained from $\mathbf{G}$ by removing all arrows pointing into variables in $\mathbf{X}$.

Rule 2 (Action/observation exchange)

$P(\mathbf{Y} \mid do(\mathbf{X}), do(\mathbf{Z}), \mathbf{W}) = P(\mathbf{Y} \mid do(\mathbf{X}), \mathbf{Z}, \mathbf{W})$ if $\mathbf{Y}$ and $\mathbf{Z}$ are *d*-separated by $\mathbf{X} \cup \mathbf{W}$ in [illegible] graph obtained from $\mathbf{G}$ by removing all arrows pointing into variables in $\mathbf{X}$ and all arrows pointing out of variables in $\mathbf{Z}$.

Rule 3 (Insertion/deletion of actions)

Theorem 6.2 (Rules of do-calculus) *Given a causal graph G, an associated distribution P, and disjoint sets of variables Y, T, Z, and W, the following rules hold.*

Rule 1:

$$P(y \mid do(t), z, w) = P(y \mid do(t), w) \quad \text{if } Y \perp\!\!\!\perp_{G_{\overline{T}}} Z \mid T, W \qquad (6.18)$$

Rule 2:

$$P(y \mid do(t), do(z), w) = P(y \mid do(t), z, w) \quad \text{if } Y \perp\!\!\!\perp_{G_{\overline{T},\underline{Z}}} Z \mid T, W \qquad (6.19)$$

Rule 3:

$$P(y \mid do(t), do(z), w) = P(y \mid do(t), w) \quad \text{if } Y \perp\!\!\!\perp_{G_{\overline{T},\overline{Z(W)}}} Z \mid T, W \qquad (6.20)$$

[illegible] the set of nodes of Z that aren't ancestors of any node [illegible]

Rule 1: Decide if we can ignore an observation

$$P(y \mid z, \mathrm{do}(x), w) = P(y \mid \mathrm{do}(x), w) \qquad \text{if } (Y \perp Z \mid W, X)_{G_{\overline{X}}}$$

Rule 2: Decide if we can treat an intervention as an observation

$$P(y \mid \mathrm{do}(z), \mathrm{do}(x), w) = P(y \mid z, \mathrm{do}(x), w) \qquad \text{if } (Y \perp Z \mid W, X)_{G_{\overline{X}, \underline{Z}}}$$

Rule 3: Decide if we can ignore an intervention

$$P(y \mid \mathrm{do}(z), \mathrm{do}(x), w) = P(y \mid \mathrm{do}(x), w) \qquad \text{if } (Y \perp Z \mid W, X)_{G_{\overline{X}, \overline{Z(W)}}}$$

$$
\begin{aligned}
&\qquad [\text{Marginalization across } z + \text{chain rule for conditional probabilities}] \\
P(y \mid \mathrm{do}(x)) &= \sum_z P(y \mid \mathrm{do}(x), z) \times P(z \mid \mathrm{do}(x)) \\
&\qquad [\text{Use Rule 2 to treat } \mathrm{do}(x) \text{ as } x] \\
&= \sum_z P(y \mid x, z) \times P(z \mid \mathrm{do}(x)) \\
&\qquad [\text{Use Rule 3 to nuke } \mathrm{do}(x)] \\
&= \sum_z P(y \mid x, z) \times P(z \mid \text{nothing!}) \\
&\qquad [\text{Final backdoor adjustment formula!}] \\
&= \sum_z P(y \mid x, z) \times P(z)
\end{aligned}
$$

Adjusting for backdoor confounding

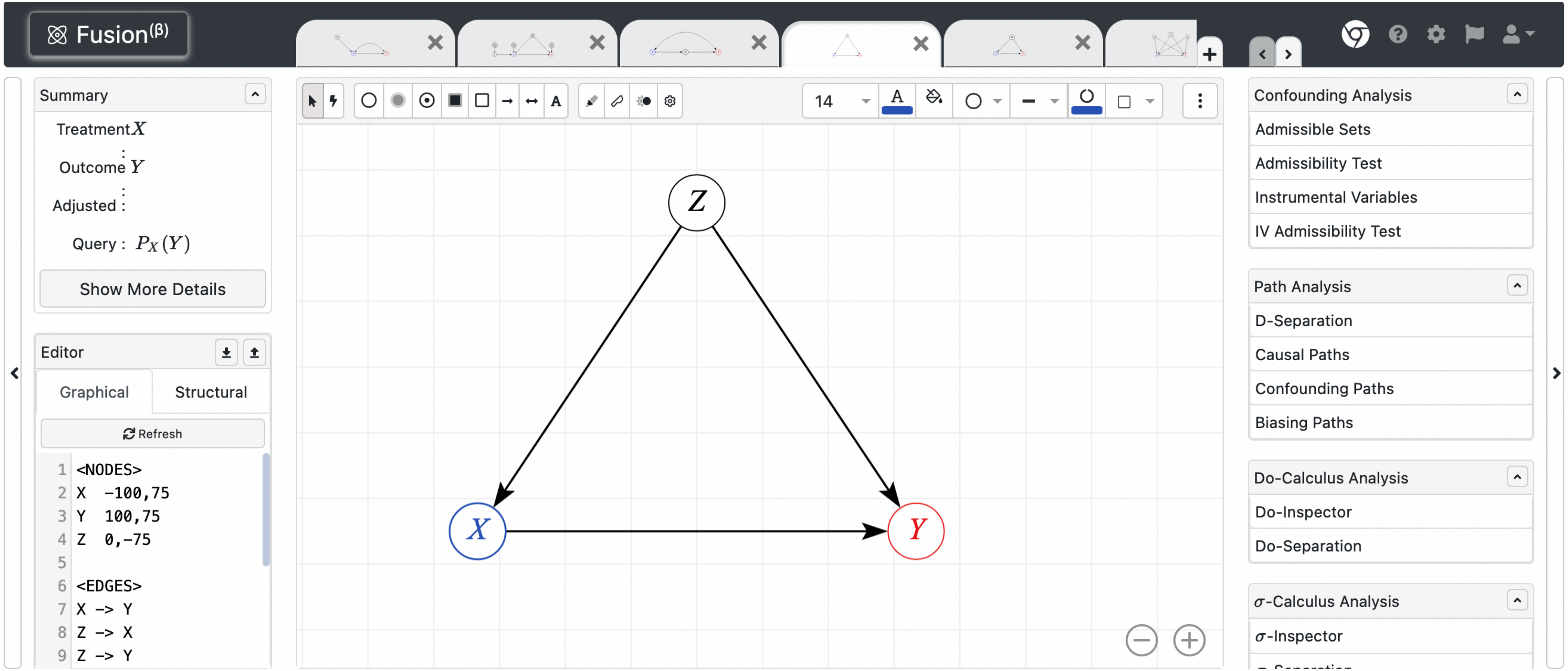
$$\overbrace{\mathbf{E}(y \mid \operatorname{do}(x))}^{\text{Causal effect of } x \text{ on } y} = \underbrace{\sum_{z}}_{\text{Sum across all values of } z} \overbrace{\mathbf{E}(y \mid x, z)}^{\text{Conditional mean of } y, \text{ given } x \text{ and } z\ldots} \times \overbrace{\mathbf{P}(z)}^{\ldots \text{ weighted by } z}$$

Adjusting for frontdoor confounding

More complex DAGs without obvious backdoor or frontdoor solutions

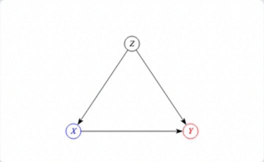
Chug through the rules of do-calculus to see if the relationship is identifiable

Causal Fusion


Fusion(β)
Summary
Treatment : X
Outcome : Y
Adjusted :
Query : $P_X(Y)$
Show More Details
Editor
Graphical
Structural
Refresh
1 <NODES>
2 X -100,75
3 Y 100,75
4 Z 0,-75
5
6 <EDGES>
7 X -> Y
8 Z -> X
9 Z -> Y
14
A
Z
X
Y
Confounding Analysis
Admissible Sets
Admissibility Test
Instrumental Variables
IV Admissibility Test
Path Analysis
D-Separation
Causal Paths
Confounding Paths
Biasing Paths
Do-Calculus Analysis
Do-Inspector
Do-Separation
σ-Calculus Analysis
σ-Inspector

Compute The causal effect of X on Y conditional on with do : (Query: $P_X(Y)$ Non-Parametric Clear

1 $$P_X(Y) = \sum_Z P(Y|X,Z)\,P(Z)$$



Load
Estimation
Derivation
Remove

Simplified Obtained by Back-Door adjustment with an admissible set $\{Z\}$

Do-Calculus

$P_X(Y)$ (1)

$\sum_Z P_X(Y|Z)\,P_X(Z)$ Summing over: Z (2)

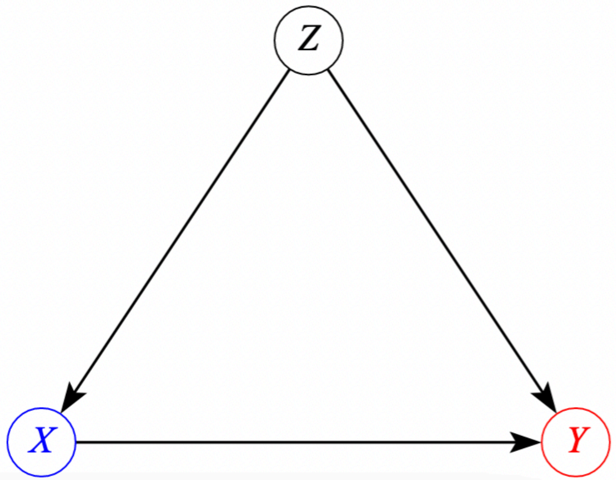
$\sum_Z P(Y|X,Z)\,P_X(Z)$ Rule 2: $(X \perp Y|Z)_{G_{\underline{X}}}$ (3)

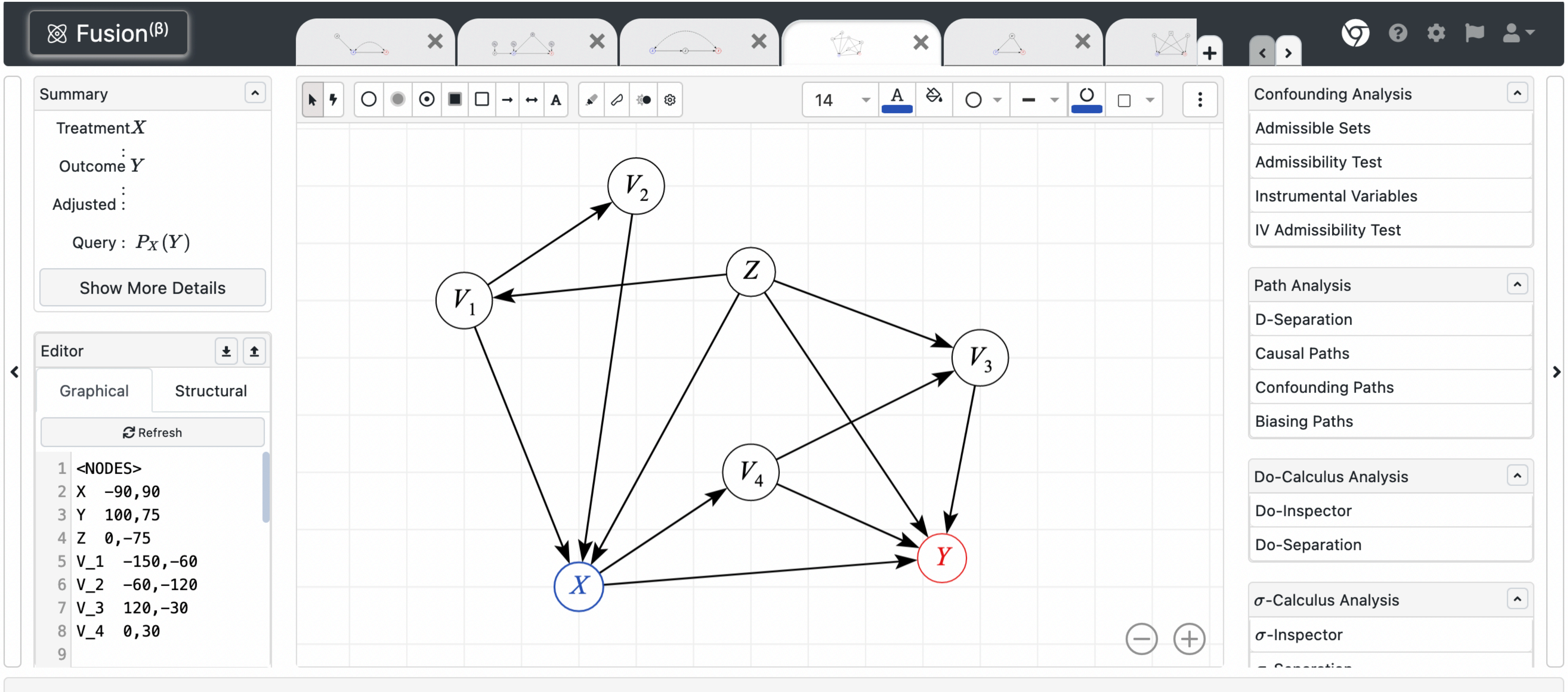
$\sum_Z P(Y|X,Z)\,P(Z)$ Rule 3: $(X \perp Z)_{G_{\overline{X}}}$ (4)

Finally we get: $\sum_Z P(Y|X,Z)\,P(Z)$

Subgraph:

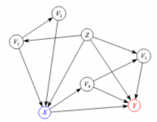
Show non-active nodes/edges




Fusion(β)
Summary
Treatment X
Outcome Y
Adjusted :
Query : $P_X(Y)$
Show More Details
Editor
Graphical
Structural
Refresh
1 <NODES>
2 X -90,90
3 Y 100,75
4 Z 0,-75
5 V_1 -150,-60
6 V_2 -60,-120
7 V_3 120,-30
8 V_4 0,30
9
14
V_2
V_1
Z
V_3
V_4
X
Y
Confounding Analysis
Admissible Sets
Admissibility Test
Instrumental Variables
IV Admissibility Test
Path Analysis
D-Separation
Causal Paths
Confounding Paths
Biasing Paths
Do-Calculus Analysis
Do-Inspector
Do-Separation
σ-Calculus Analysis
σ-Inspector

Compute The causal effect of X on Y conditional on with do : (Query: $P_X(Y)$ Non-Parametric Clear

1 $P_X(Y)$ is not identifiable from $P(V_1, V_2, V_3, V_4, X, Y, Z)$ and $P_X(V_1, V_2, V_3,$



Load

Remove

**When things are identified, there are still arrows leading into Y.
What do we do with those?
How do you explain those relationships?**

**Outcomes have multiple causes.
How do you justify that your proposed cause is the most causal factor?**

Does every research question need an identification strategy?

No!

Correlation alone is okay!
Can lead to more focused causal questions later!

Forbes

BREAKING | Jan 14, 2022, 12:34pm EST | 145,393 views

Moderna Starts Human Trials Of mRNA Vaccine For Virus That Likely Causes Multiple Sclerosis

Robert Hart Forbes Staff
Business
I cover breaking news.

Follow

Listen to article 3 minutes

TOPLINE Moderna recently launched early stage clinical trials for an mRNA vaccine against the Epstein-Barr virus (EBV), a common pathogen that infects almost everyone at some point in their lives, is the primary cause of mononucleosis and, according to a study published in the journal *Science* Thursday, likely causes multiple sclerosis (MS), offering hope the devastating neurological condition might be prevented.

p-values and confidence intervals

In the absence of *p*-values, I'm confused about how we report... significance?

Imbens and *p*-values

Nobody really cares about *p*-values

Decision makers want to know a number or a range of numbers—some sort of effect and uncertainty

Nobody cares how likely a number would be in an imaginary null world!

Imbens's solution

Report point estimates and some sort of range

> "It would be preferable if reporting standards emphasized confidence intervals or standard errors, and, even better, Bayesian posterior intervals."

Point estimate

The single number you calculate (mean, coefficient, etc.)

Uncertainty

A range of possible values

Greek, Latin, and extra markings

Statistics: use a sample to make inferences about a population

Greek

Letters like β_1 are the ***truth***

Letters with extra markings like $\hat{\beta}_1$ are our ***estimate*** of the truth based on our sample

Latin

Letters like X are ***actual data*** from our sample

Letters with extra markings like $\bar{X}$ are ***calculations*** from our sample

Estimating truth

Data → Calculation → Estimate → Truth

Data	X
Calculation	$\bar{X} = \frac{\sum X}{N}$
Estimate	$\hat{\mu}$
Truth	μ

$$\bar{X} = \hat{\mu}$$

$$X \rightarrow \bar{X} \rightarrow \hat{\mu} \xrightarrow{\text{🤞 hopefully 🤞}} \mu$$

Population parameter

Truth = Greek letter

An single unknown number that is true for the entire population

Proportion of left-handed students at GSU

Median rent of apartments in Atlanta

Proportion of red M&Ms produced in a factory

Treatment effect of your program

Samples and estimates

We take a sample and make a guess

This single value is a *point estimate*

(This is the Greek letter with a hat)

Variability

You have an estimate, but how different might that estimate be if you take another sample?

Left-handedness

You take a random sample of 50 GSU students and 5 are left-handed.

If you take a different random sample of 50 GSU students, how many would you expect to be left-handed?

3 are left-handed. Is that surprising?

40 are left-handed. Is that surprising?

Nets and confidence intervals

How confident are we that the sample picked up the population parameter?

Confidence interval is a net

We can be X% confident that our net is picking up that population parameter

If we took 100 samples, at least 95 of them would have the true population parameter in their 95% confidence intervals

> A city manager wants to know the true average property value of single-owner homes in her city. She takes a random sample of 200 houses and builds a 95% confidence interval. The interval is ($180,000, $300,000).

We're 95% confident that the interval ($180,000, $300,000) captured the true mean value

WARNING

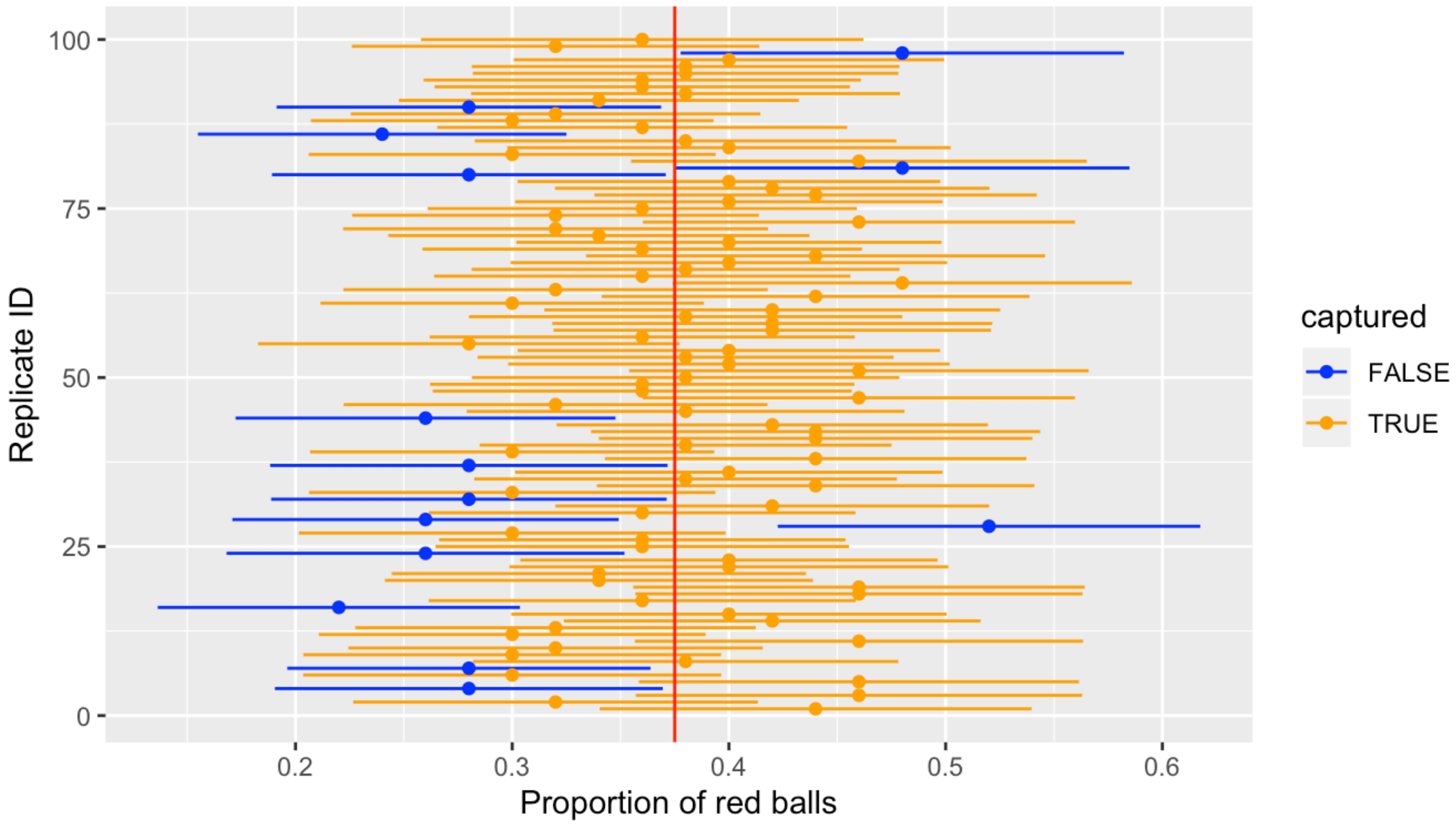
It is way too tempting to say “We’re 95% sure that the population parameter is X”

People do this all the time! People with PhDs!

YOU will do this too

Nets

If you took lots of samples,
95% of their confidence intervals
would have the single true value in them



100
75
50
25
0
Replicate ID
0.2
0.3
0.4
0.5
0.6
Proportion of red balls
captured
FALSE
TRUE

Frequentism

This kind of statistics is called "frequentism"

The population parameter θ is fixed and singular while the data can vary

$$P(\text{Data} \mid \theta)$$

You can do an experiment over and over again; take more and more samples and polls

Frequentist confidence intervals

"We are 95% confident that this net captures the true population parameter"

~~"There's a 95% chance that the true value falls in this range"~~

Bayesian statistics

Rev. Thomas Bayes

$$P(\theta \mid \text{Data})$$

$$\color{orange}{P(\text{H} \mid \text{E})} = \frac{\color{red}{P(\text{H})} \times \color{blue}{P(\text{E} \mid \text{H})}}{P(\text{E})}$$

$$P(\mathrm{H} \mid \mathrm{E}) = \frac{P(\mathrm{H}) \times P(\mathrm{E} \mid \mathrm{H})}{P(\mathrm{E})}$$

$$P(\text{Hypothesis} \mid \text{Evidence}) = \frac{P(\text{Hypothesis}) \times P(\text{Evidence} \mid \text{Hypothesis})}{P(\text{Evidence})}$$

But the math is too hard!

So we simulate!

(Monte Carlo Markov Chains, or MCMC)

Bayesianism and parameters

In the world of frequentism,
there's a fixed population parameter
and the data can hypothetically vary

$$P(\text{Data} \mid \theta)$$

In the world of Bayesianism,
the data is fixed (you collected it just once!)
and the population parameter can vary

$$P(\theta \mid \text{Data})$$

Bayesian credible intervals

(AKA posterior intervals)

"Given the data, there is a 95% probability that the true population parameter falls in the credible interval"

Intervals

Frequentism

There's a 95% probability that the range contains the true value

Probability of the range

Few people naturally think like this

Bayesianism

There's a 95% probability that the true value falls in this range

Probability of the actual value

People *do* naturally think like this!

Thinking Bayesianly

We all think Bayesianly,
even if you've never heard of Bayesian stats

Every time you look at a confidence interval, you inherently think that the parameter is around that value, but that's wrong!

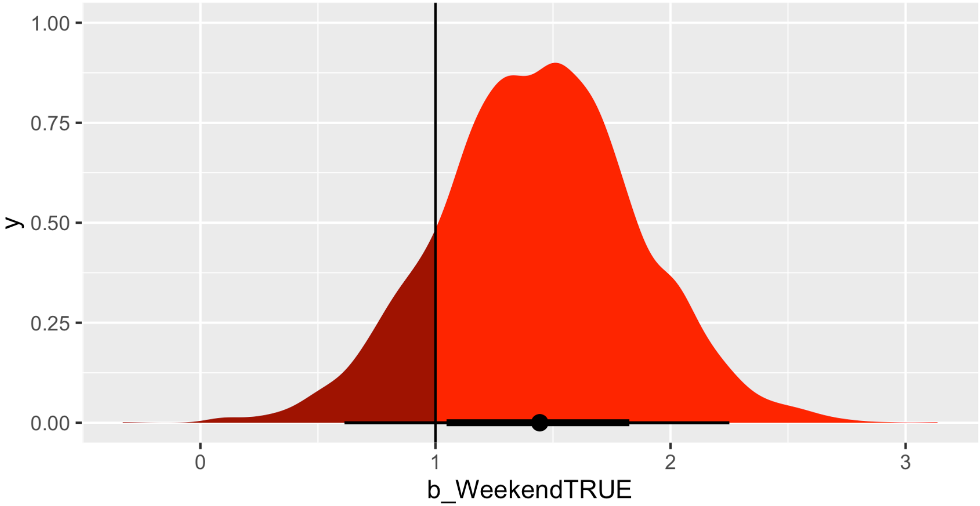
BUT Imbens cites research that that's actually generally okay

Often credible intervals are super similar to confidence intervals

Bayesian inference

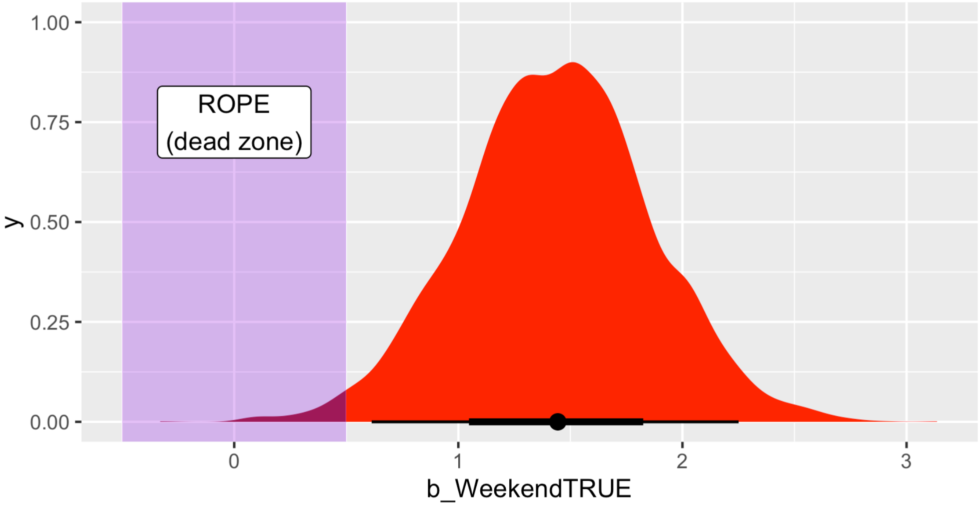
Inference without *p*-values!

Probability of direction



Point shows median value;
thick black bar shows 66% credible interval;
thin black bar shows 95% credible interval

Region of practical equivalence (ROPE)



Point shows median value;
thick black bar shows 66% credible interval;
thin black bar shows 95% credible interval