

In-person session 13

November 14, 2022

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

Control variables and confounding

Ethics + pre-analysis plans

Sharing your stuff

Simulating data

Control variables and confounding

**Do we really not need to
interpret every coefficient?!**

**Is there any harm in
interpreting the coefficients anyway?**

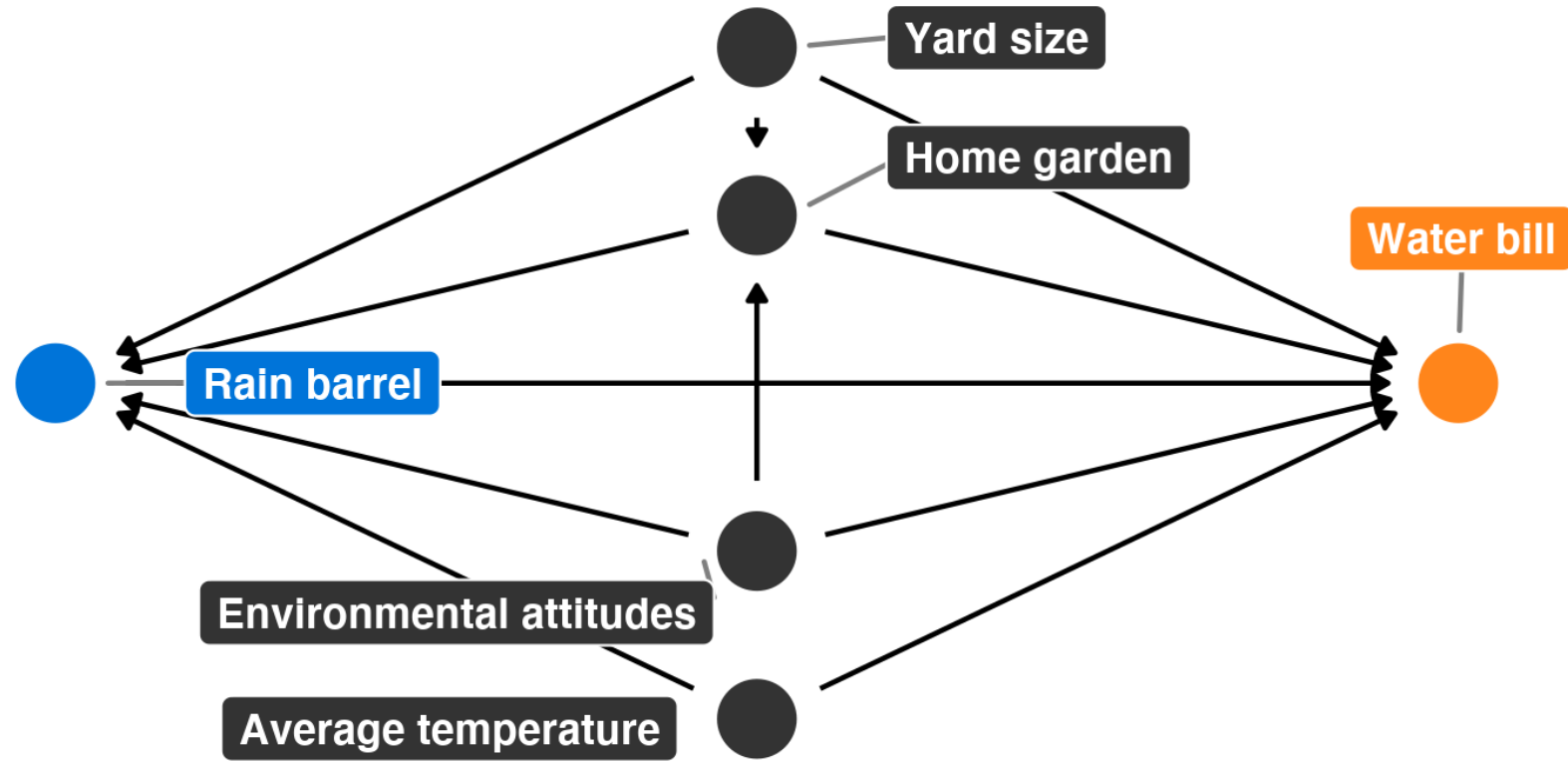
**Should they be significant at least?
Or have the expected \pm sign?**

Table 2 fallacy

Daniel Westreich and Sander Greenland, "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients," *American Journal of Epidemiology* 177, no. 4 (February 2013): 292--98, <https://doi.org/10.1093/aje/kws412>.

Luke Keele, Randolph T. Stevenson, and Felix Elwert, "The Causal Interpretation of Estimated Associations in Regression Models," *Political Science Research and Methods* 8, no. 1 (January 2020): 1--13, <https://doi.org/10.1017/psrm.2019.31>.





Why do we teach stats with the idea of throwing in a bunch of control variables if that's really not the best way to evaluate impact?

What is the point of using control variables if they aren't going to be interpreted?

How do we know which coefficients we need to ultimately worry about and interpret?

**Why did we control for things
in the RCT section of problem set 8?**

How can you NOT have unobserved confounders unless you're doing a laboratory study where you can control everything?

Design-based inference seems easier since there's no unobserved confounding?

Why don't we just do that all the time?

Model-based inference seems easier since there's no need for a special situation?

Why don't we just do that all the time?

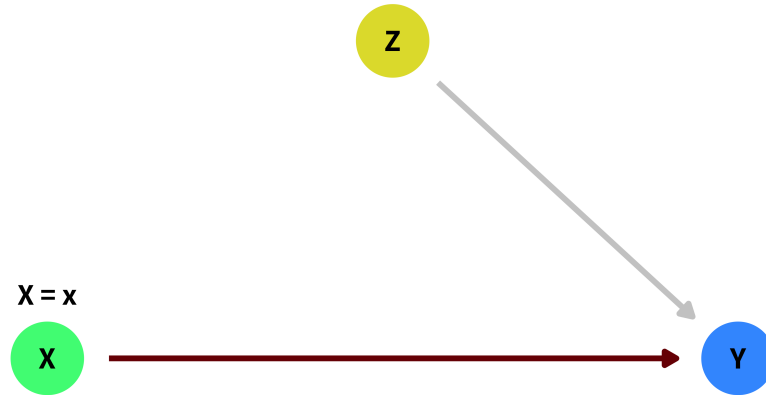
**You said that choosing between
model-based inference versus design-based
inference is based on the situation that we are in.**

**Which one of the approaches do you think
will be used the most in the real world?**

**Do we not need to use a DAG
if we use one of
the design-based methods?**

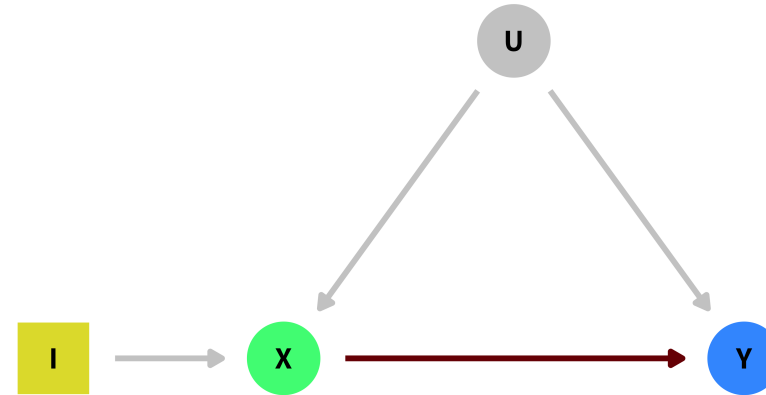
Randomized trial

Randomization deletes all arrows into X



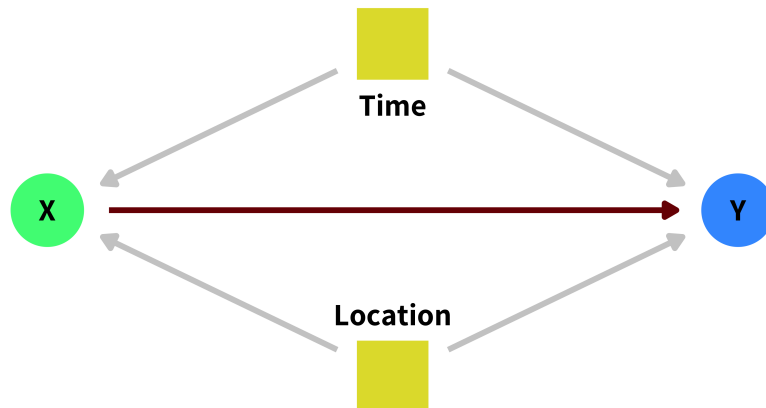
Instrumental variable

Find effect of instrument (I) on X, then find effect of $(X | I)$ on Y



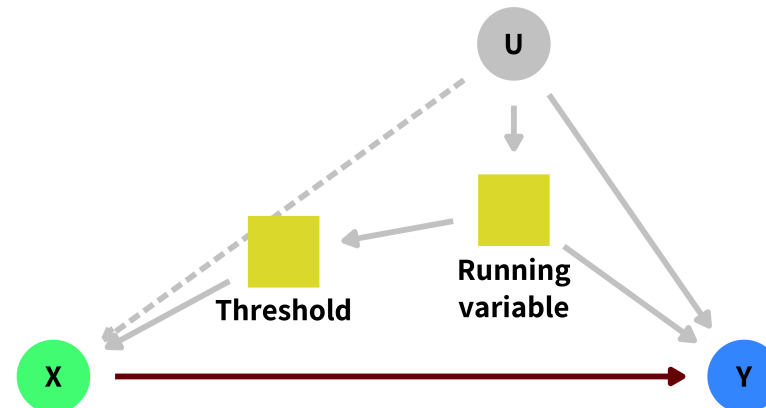
Difference-in-differences

Adjust for both time (e.g., year) and location (e.g., country, state)



Regression discontinuity

Adjust for both the running variable and the threshold



Ethics + pre-analysis plans

**Where can one publish their assumptions
before running the numbers in research?**

OSF.io AsPredicted.org

**Do people really provide all their code and analysis in advance?
What if you realize you made a mistake
or want to explore something later?**

**What if we have additional variables to investigate that we missed
out during our simulation of fake data and pre-registration/pre-
analysis stage? Can we add them once they have been registered?**

**Does pre-registering include interactions
or quadratics you want to test?**

Standard operating procedures

—

Departures from preregistered plan

—

Example

**Should you share the synthetic data
as part of a preanalysis plan?**

Does pre-registration kill creativity and "ah-ha" moments?

Example of confirmatory vs.
exploratory preregistration

Sharing your stuff

**What can we put
on our resumes now?**

**Can I really just post R stuff
on a website or on Twitter?**

**It seems like that's what
the experts do, not me!**

(Public work)

Websites for sharing R stuff

Super easiest (but least flexible): **RPubs**

Easiest (but less flexible): **R Markdown websites**

Example; example

What all the cool kids are starting to use: **Quarto websites**

Example; example; example

Hardest (but most flexible): **Blogdown**

Example; example

Quarto

(R Markdown 2.0)

Quarto

Example PDF and Word templates

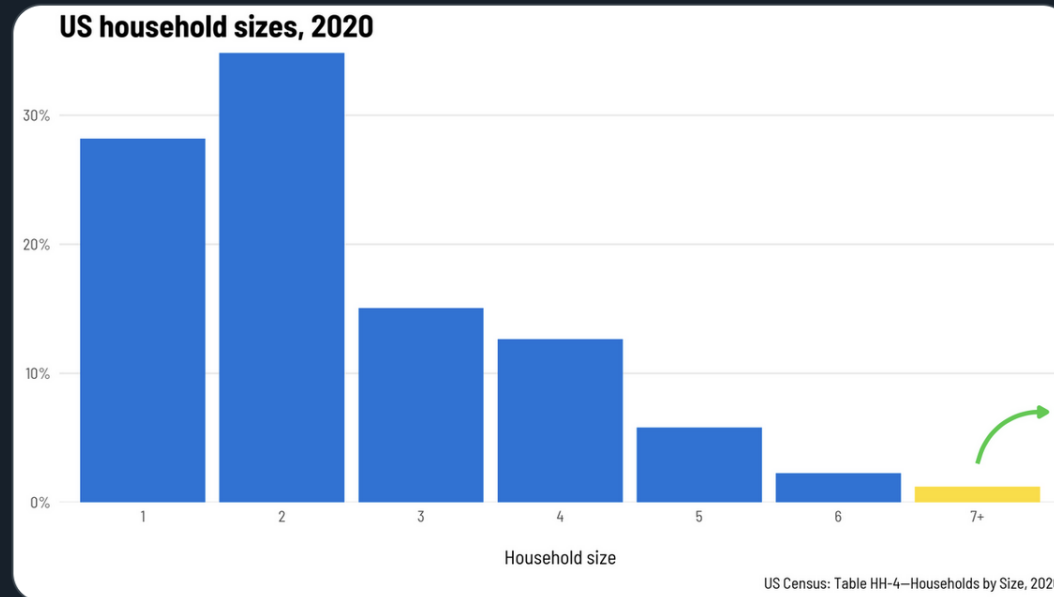
Simulating data

Poisson in real life



Andrew Heiss
@andrewheiss

we just really wanted to yeet ourselves to the far edge of this distribution



10:12 AM · May 25, 2021 · Twitter Web App

Basic process

1: Draw a DAG

2: Create standalone exogenous columns

3: Connect endogenous columns

Baseline + effect

4: Polish columns

Iterate. Iterate so so much.