Lecture 10: Causality

Intro Stats with Nathan Favero

American University (Washington, DC)

August 2, 2024

Causality

- Five possibilities if X correlates with Y
  1. Correlation between X and Y is a coincidence
  2. Z causes X and Y
  3. Y causes X
  4. Research design problems
  5. X causes Y
- Assessing causality is a bit like detective work; use process of elimination
- If we rule out options 1-4, we can conclude 5 must be true

Causality

Possibility 1: Correlation between X and Y is a coincidence

- Tools to use: confidence intervals and hypothesis tests
- We know there's randomness in our data (correlation will almost never be exactly 0, even if two variables are unrelated to one another)
- Hypothesis test: if the correlation is *statistically significant*, we can rule out possibility 1 (the correlation is strong enough that we think it's unlikely to be the result of dumb luck)
- Confidence interval: after taking randomness into account, we're confident the true correlation falls within some range (maybe we can rule out a correlation of 0)

Causality

Possibility 2: Z causes X and Y

- Tool to use: multiple regression
- If Z has an effect on both X and Y, X will typically be correlated with Y even if X has no effect on Y
- Using multiple regression, we can "control" for the effect of Z (include Z as an independent variable; in Stata, run `reg Y X Z` instead of just `reg Y X`)
- If X has a statistically significant coefficient while controlling for Z, we can rule out possibility 2
- Note: in order to use multiple regression to rule out possibility 2, we have to have good measures of all potential third variables (Z) that might be affecting both X and Y (very difficult in practice)

Causality

Possibility 3: Y causes X

- Tools to use: beyond the scope of this course
- Occasionally, we can logically rule out this possibility (e.g., when studying the relationship between voting intention and economic conditions, hard to imagine that economy changed much because of how people planned to vote)
- Sometimes, collecting data over time will help us assess this possibility (do changes in X precede corresponding changes in Y?)

Causality

Possibility 4: Research design problems

- Here we can consider problems like self-selection, attrition, measurement error, sampling problems, etc.
- Example (self-selection): one hospital has better health outcomes because healthier people go to that hospital for care
- You'll learn more about these types of potential problems throughout the conduct course sequence

Causality

1. Correlation between X and Y is a coincidence
2. Z causes X and Y
3. Y causes X
4. Research design problems
5. X causes Y

- Often, experiments or quasi-experiments are used to help rule out possibilities 2-4
- There are lots of fancy techniques (beyond the scope of this course) that economists and policy analysts use to try to assess causality; they often have complicated assumptions, and thus their reliability is difficult to assess

A few important concepts

- Internal validity: confidence that a *causal conclusion* can be drawn about a set of variables (among the units observed in the study)
- External validity: confidence that the findings of a study can be generalized to a broader set of units (beyond those observed in the study)
- Experimental studies: the researcher manipulates one or more variables of interest
- Observational studies: the researcher observes variation in variables caused by something other than the researcher's own intervention