Social Statistics

Wrapping Up Descriptives

October 11, 2021

Transition to Multivariate Descriptives

Bivariate description examines association between two variables

Multivariate description begins to explain associations by including other variables

Common variables in sociology:

- Race, sex, age, education, income, class, marital status, etc.

Often more interesting as *control variables*

- How does the relationship between education and age at first birth *differ by race*?

Language of Multivariate Explanations

Controlling for race, the association between years of education and age at first birth is...

Independent of race, the association between years of education and age at first birth is...

Net of race, the association between years of education and age at first birth is...

Holding race constant, the association between years of education and age at first birth is...

Language of Multivariate Explanations

Will rarely *eliminate* the X1-Y relationship but may: weaken, reduce, reverse

Does not have to change X1-Y, but should not say controlling "proves" anything

Note: The focal relationship is still the X1-Y relationship. The new variables are X2, X3, X4, etc.

Finding Lurking Variables

A&F on GPA and income (p. 307)

- "An alternative explanation is that high GPA and high income could both be caused by a lurking variable such as individuals' tendency to work hard."

How would you measure working hard?

- But hours worked/studied, etc. could also be associated with income!

Finding Lurking Variables

Preferable to think of multivariate models as attempts to offer "alternative explanations"

Remember: theory is lurking, but magic variables rarely exist

Remember also: do not try to control for too many things. Inference requires variation!

Why Multivariate Explanations Matter

"So often, when we show the deep logic of what once seemed odd or incomprehensible behavior, bystanders often nod knowingly and say, 'It's obvious.' ...On the other hand, if they don't like your finding...they will announce that your results are either (a) 'spurious' or (b) biased. In either case, what they mean is that *the way you chose the sample influenced the outcome.* This is the thing that we must avoid at all costs."

-- Kristin Luker, *Salsa Dancing into the Social Sciences*, p. 110

Types of Multivariate Explanations

Spurious Associations

X2 related to X1 and Y; controlling for X2 reduces association between X1 and Y

- As ice cream sales increase, drowning injuries also tend to increase

- Fire damage tends to increase when there are more fire trucks responding

Types of Multivariate Explanations

Chain Relationships

X1 influences Y *through* X2 ("Intervening" or "Mediating" relationship)

- Neighborhood poverty -> lack of transportation -> unemployment

- Common in health research

  - Smoking -> Tar Buildup -> Lung Disease

Types of Multivariate Explanations

Suppressor Variables

No (or weak) association between X1 and Y until controlling for X2

- Controlling for age reveals association between education and income

Types of Multivariate Explanations

Some neat examples in the Chetty et al (2014) mobility data

Let's return to R to consider the correlation between racial segregation (`racial_seg`) and median household income (`hh_income`). Would you expect this correlation to be positive or negative? Would you expect this correlation to be strong or weak?