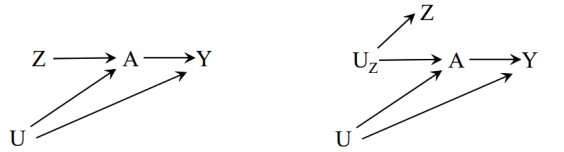
Reader Report 10

H R ch.16 Claudia Shi (js5334)
April 15, 2019

Instrumental variable is defined under three conditions: i) Z is associated with A, ii) Z does not affect Y except through its potential effect on A iii) Z and Y do not share causes.
Apparently there's a distinction between causal instrument and "proxy instrument". Later in 16.4, it explained that if the instrument Z is a proxy instrument, then the standard IV estimand is not a particular sub-population of compilers but a weighted average of the effect all individual in the population. (What is this weighted average?)



(a) Causal Instrument (b) Proxy Instrument

Figure 1: 2 Figures side by side

The smoking example is quite interesting. Initially, I shared Matt's doubts that local culture or socioeconomical status could be a shared cause. For instance, we could argue that people with lower income is likely to be more influenced by the price of cigarettes and more likely to be obese. [1]However, it is no causal relationship between the price and socioeconomical status per se, but rather it is the relationship of smoker's reaction towards the cigarettes price, which can be captured as "smoking cessation".

Technical poiny 16.2 explained how the IVs provide bounds for estimates under partial identification. I didn't really understand how the sharp bound is derived or how it is related to the g-formulation. It was also kinda of surprising to me that IVs only provide a bound. I suppose most social scientists and economists just make a lot of additional assumptions.

16.3 added a fourth condition: homogeneity. There's strong homogeneity, that is a constant effect, and a weak homogeneity, that is a s equality of the average causal effect within levels of Z in both the treated and in the untreated. Is this in a sense co-variate balancing? Intuitively Z just serve as a balanced random assignment? An alternative condition is monotonicity. The basic idea is instrument either has A remains the same or increase A. This is assuming a binary treatment of Z and A. I wonder what would it look at in a categorical case? Would it just be monotone in the sense of a linear "monotone" relationship?

I really the last section. it clearly listed various drawbacks of using IVs, in comparison to previous methods. It still require modeling assumptions even if infinite data were avaliable. Does this make IVs less attractive to computer scientists?

[1]https://www.prb.org/obesity-socioeconomic-status/