Homework for Chapter 11: Causality with Less Modeling

1. Suppose that you are analyzing the effect of universities and colleges opening during a pandemic on increase in the number of positive cases. Name one strategy that you can use to avoid having to collect data on all types of campus characteristic variables that are constant over time that you may have to control for in your analysis.

We can use the method of fixed effects, where universities and colleges are observed multiple times when they are open and closed during the pandemic. A control for universities is added. This effectively controls for everything constant over time about universities and colleges.

1. Intuitively, why would a method that isolates front doors allow you to ignore back doors related to unmeasured variables?

The key idea here is that we can partition the variation in Treatment. By either selecting a particular sample (tossing out data subject to the rest of the variation) or using certain approaches to statistical adjustment (focusing just on the variation we estimate to be due to an exogenous variable), we can throw out the part that is driven by back doors, and leave ourselves to do analysis with just another part that isn’t (isolate just the variation in the front door paths). Then we can focus on estimating the Front Doors directly and don’t have to worry about back doors.

1. On robustness tests:
   1. What are robustness tests?

A robustness test is a way of either checking whether we can disprove an assumption, or redoing our analysis in a way that doesn’t rely on that assumption and seeing if the result changes.

* 1. What is the purpose of conducting a robustness test?

It is a good avenue for checking whether our assumptions seem false and ferret out bad assumptions, which is much easier to do than checking whether they seem true when it comes to checking assumptions using data rather than theory.

* 1. What are placebo tests?

Placebo tests are tests where we pretend that treatment is being assigned somewhere it isn’t, and we check whether we estimate an effect. If we find an effect of “treatment,” that tells us that there must be a bad assumption somewhere, since we’re finding that the effect of nothing is something.

1. Suppose you want to study the effect of attending tutoring sessions on grade point averages (GPA). List at least five variables that impact both attendance of tutoring sessions and students’ GPA. Is it feasible to measure and control for all of the variables?

Student’s acquisition level of relevant knowledge, student’s willingness to learn, student’s time spent on studying, student’s intelligence quotient, support from student’s family, peers’ performance (especially if GPA is on a curve). It is not feasible to directly measure and control for all the above variables.

1. Describe partial identification in your own words.

Empirical results might be partly driven by a priori assumptions rather than data which rejects point identification. But even though some data do not allow for an exact quantification, we can find out which values of the true parameter of interest are compatible with the observations we made and get non-trivial insights about the objects of interest. Partial identification allows us to make fewer assumptions (instead of keeping making assumptions to guarantee point identification) and use them to generate bounds on the parameters of interest when we work with a model where it is not possible to infer the true value of the parameter of interest even with an infinitely large data set. Instead, we make the assumptions that weʼre pretty certain about. Then, we use a range of possibilities about the remaining things we must assume. Finally, we figure out what our estimate is over that range, giving us a range of possibilities for the estimate itself. Under the partial identification approach, there is a trade-off between the precision of the result(s) and strength of the assumption(s): adding precision to the results by imposing heavier assumptions or reducing the precision of the results as the assumptions become less informative.

1. Pick any causal diagram from the book other than Figure 11.2.
   1. Reproduce that diagram here.

Diagram

Description automatically generated

* 1. Select two variables on the diagram without a direct link between them (i.e. no single arrow straight from one of them to the other).

OthersBadDriving, YourBadDriving

* 1. What variables would you need to control for that will eliminate any relationship between the two variables (you might not need any).

None.

* 1. If you looked at the relationship between your two variables from part b, while controlling for the variables from part c, and you got a nonzero result, what would you conclude?

Because paths between OthersBadDriving and YourBadDriving on this diagram contain a collider which makes us assume that even if we don’t control anything, OthersBadDriving and YourBadDriving should be unrelated. If we got a nonzero result showing they are related, I will conclude that it’s the evidence against the assumptions we made for our research design.

1. What does it mean to say that the effect of financial deregulation on the rate at which firms go bankrupt is “bounded from above” at 2 percentage points? d
   1. The effect is 2 percentage points, and it’s a positive effect
   2. The effect is 2 percentage points, and it’s a negative effect
   3. The effect is at least as large as 2 percentage points
   4. The effect is no larger than 2 percentage points
   5. If we’re willing to make an additional, stronger set of assumptions, the effect would be larger than 2 percentage points, but without those assumptions it’s bounded to be lower.