**Samuel Snelson**

**SOC 723 Spring 2023**

**17 Feb 2023**

Homework for Chapter 7: Drawing Causal Diagrams

1. You are making a simplified causal diagram to represent the data generating process of viewership for a TV show. Which of the following is true?
   1. The diagram should include a variable for “number of celebrities in the cast”
      1. Yes. The number of celebrities in the case of ant given show is likely to influence people’s decision to watch. Personally, although in the context of film, I often literally search for films by an actor’s name.
   2. The diagram should contain one variable for “show airs in the evening” and another for “show doesn’t air in the evening”
      1. Yes. The evening is often referred to as “prime time” because, as is common knowledge (not strictly empirically justified), people get home from work/school and this is the time when viewership is high.
   3. The diagram should not contain a variable for “show budget” because budgets are often secret and the researcher can’t measure them
      1. Even if a researcher can’t measure a variable, this alone is not a good reason to exclude it from their DAG. Unobserved confounds (of many types) should be included so a researcher can account for their potential influence (e.g., sensitivity analysis). In the case of a given shows budget, this is likely causally related to the viewership of a show. From my own assessment of this causal system, although again from the view of film, big budget films are quite likely to have high viewership (e.g., James Cameron’s Avatar).
   4. The diagram should contain the variable “review score in the Jefferson Weekly,” which is the newspaper published by the students at Jefferson High School, with a readership of about 120 people.
      1. There is a simple and a more thoughtful answer to this . The simple answer is likely not. The more thoughtful answer is that it depends on the specific estimand a researcher is investigating. If one is studying the viewership of a show on a major network broadcasted across the whole US, then this newspaper is essentially negligible. However, if our estimand is what is the viewership of some set of shows in Jefferson city, then this newspaper may account for more some non-trivial amount of variability in viewership (assuming for the sake of the example that it’s a small town where people read the newspaper). Of course, I think Nick is asking this in the context of a large scale show. I make the point about the thoughtful answer because it is important to think carefully about what we are and are not putting in our DAGs and the implications.
2. Draw a causal diagram for the research question “do long shift hours make doctors give lower-quality care?” that incorporates the following features (and only the following features):
   1. Long shift hours affect both how tired doctors are, and how much experience they have, both of which affect the quality of care

Diagram

Description automatically generated

* 1. How long shifts are is often decided by the hospital the doctor works at. There are plenty of other things about a given hospital that also affect the quality of care, like its funding level, how crowded it is, and so on

Diagram

Description automatically generated

* 1. New policies that reduce shift times may be implemented at the same time (with the timing determined by some unobservable change in policy preferences) as other policies that also attempt to improve the quality of care

Diagram

Description automatically generated

1. Consider this research question: Does the funding level of public schools affect student achievement for students in your country?
   1. What is the treatment and what is the outcome of interest?
      1. Treatment: funding level of public schools
      2. Outcome: achievement for students in your country
   2. Write down a list of relevant variables.
      1. Community crime/police presence
      2. School delinquency/School Resource Office (SRO) presence
         1. I have elected to collapse these variables for i. and ii. because there is some debate about which comes first. My goal is not to resolve this but to recognize their mutual influence on educational finance and achievement.
      3. Lag student achievement
      4. Quality of educational materials (e.g., books; hardware)
      5. Quality of instruction (e.g., teacher competency; teacher work burdens, access to AP/IB)
      6. Extracurricular support access from school services (e.g., standardize test prep, tutoring, low student-to-counselor ratio)
      7. Community wealth
      8. Local tax contribution to school finance
      9. Supplementary extracurricular support access external to school services (e.g., standardized test prep, tutoring)
   3. Which of the variables in your list in part b are causes of both treatment and outcome?
      1. In this causal system which I will display below, there are two separate confounding structures. Community crime/police prevalence influences school finance through local/state/federal investment and influences student achievement through school delinquency/SRO prevalence. Secondly, community wealth influences school finance through local tax contributions to school finance and influences student achievement through access to supplementary extracurricular supports (outside school).
   4. Why might we want to pay extra attention to the variables listed in part c?
      1. As recognized above, these are confounding structures and imply that both the treatment and outcome are common causes of the confounders. This means that if we were to disregard the confounders in our estimation of the effect of the treatment on the outcome, we would generate a biased estimate of the total (and direct) effect.
   5. Draw a causal diagram of the variables listed in part b.

Diagram

Description automatically generated

* 1. Simplify the diagram from part e.
     1. If I’m being honest, I don’t really want to simplify further because these are such important dimensions which shape the landscape of education finance and achievement. However, if I really needed to simplify, then I would acquiesce by focusing on the central mediating structure with acknowledgement of the influence from the confounding structures. I know that Nick references mediators as something to exclude to simplify a DAG. However, I care a lot about the identification of those mediating pathways, so they’re not going anywhere unless we’re interested in the total effect more so (then we’ll swap the mediators for the confounders).

1. Describe the kinds of situations that each of the following could be applied to in order to simplify a causal diagram.
   1. Unimportance
      1. Simply put, if the causal effect of a variable on another is essentially negligible, then it can be disregarded. An example of this is the decision to include or exclude the Jefferson Weekly newspaper in the model for viewership of a TV show. Even if the newspaper causes the 120 readers to watch a show, this will only trivially influence the probability distribution of the outcome for millions of possible viewers.
   2. Redundancy
      1. The point of this idea is to focus on distinct variables. Variables which are similar to each other (ultimately reflected in accounting for overlapping variance in the outcome) can be combined or one of them selected without much harm to the estimation.

* 1. Mediators
     1. I understand Nick’s point here to be that in the estimation of the total effect of a treatment on an outcome, mediators are not necessary to get an accurate effect estimate. However, as I stated in the previous question, I’ll push back when it concerns the estimation of the effects of those mediating pathways. Like Steve has said many times in class, sometimes the direct effect doesn’t make sense because a treatment may only influence the outcome through a mediator. However, when one is interested precisely in the effect of the mediating pathway (e.g., the effect of wealth on SAT score through test prep), then mediators matter!
  2. Irrelevance
     1. Nick’s point here is that even if a variable is important as a causal influence on the treatment xor the outcome, we are ultimately interested in factors which influence the relationship (that is, the treatment and the outcome together). Thus, when a variable isn’t related to both and is not part of our research question, then it doesn’t need to be included.

1. How can a causal diagram be modified so as to avoid cyclic relationships?

*“We live and we die by time, and we must not commit the sin of turning our back on time.”*

*– Tom Hanks, Castaway*

* 1. DAG cycles (oxymoronic as those two things are together) are generally an illusion which obfuscates the temporal relation between some set of time-dependent variables. X 🡨🡪 Y conceals that generally X1 🡪 Y1 🡪 X2 🡪 Y2 and so forth. Causal diagrams can therefore be modified to resolve cyclic relationships by noting the time at which causal influence is transmitted from X to Y and vice versa.
  2. Also, as Steve has noted in class about cycles, unless there is a ceiling or floor, they’ll blow up or down endlessly. Last time I checked, there weren’t any sociological black holes, so no infinities or infinitesimals.

1. Think of a research question in your field of interest.
   1. What is the cause variable and what is the outcome variable?
      1. Cause: ethnoracial variability in maltreatment rates
      2. Outcome: ethnoracial inequalities in foster care placement proportions
   2. Write down a list of between 5 and 10 relevant variables in the data generating process.
      1. Race
      2. Poverty (inability for parent to meet child’s basic needs)
      3. Parent mental health problems
      4. Child mental health problems
      5. Child delinquency
      6. Ethnoracial disparities in CPS intervention
   3. Draw a causal diagram incorporating all the variables from part b.
   4. Stop working on this problem for fifteen minutes and do something else. Then come back, look at your causal diagram from part c again, and describe one bad assumption you think it made, or something it left out.
      1. This is a model approximately reflecting one side in a theoretical debate over the causes of ethnoracial inequalities in foster care placement. One (ideal type) side of the debate holds that ethnoracial disparities in foster care placement are caused by ethnoracial differences in underlying maltreatment rates. The other (ideal type) side of the debate holds that ethnoracial disparities in foster care placement are caused by racial disparate likelihoods of surveillance and intervention by child welfare officials. This model assumes that intervention by CPS (child protective services) is only a function of variability in maltreatment rates. These rates vary as a function of race, poverty, and parents’ mental health problems. The ethnoracial disparities in surveillance/intervention approach would construct the model such that CPS intervention is a function of (but not limited to) race directly.
2. Consider the diagram below. It depicts a cyclical relationship between student achievement and motivation. If students achieve more (i.e., score well on exams), then their motivation goes up, and if their motivation goes up, they achieve more. Change the diagram so that the relationship is not cyclic anymore.  
   

Diagram

Description automatically generated