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? a
   1. The diagram should include a variable for “number of celebrities in the cast”
   2. The diagram should contain one variable for “show airs in the evening” and another for “show doesn’t air in the evening”
   3. The diagram should not contain a variable for “show budget” because budgets are often secret and the researcher can’t measure them
   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.
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
   2. 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
   3. 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

Doctors’FatigueLevel

NewPolicieSReducingShiftTimes

ShiftHours

QualityOfCare

Doctors’Experience

Hospital

FundingLevel funding level funding level

OtherPoliciesImprovingQualityOfCare

AmountOfVisitors

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? The treatment is the funding level of public schools. The outcome of interest is national student achievement.
   2. Write down a list of relevant variables.

Faculty performance in public schools (PSs), public school’s administration, student support services in public schools, education resources, education quality of private schools, family income, macro (economic) environment, and education policy.

* 1. Which of the variables in your list in part b are causes of both treatment and outcome?

EducationPolicy and MacroEnvironment are the causes of both treatment and outcome.

* 1. Why might we want to pay extra attention to the variables listed in part c?

Thet may offer alternative explanations that EducationPolicy or MacroEnvironment causes both FundingLevelOfPublicSchools and NationalStudentAchievement, and so FundingLevelOfPublicSchools and NationalStudentAchievement might be related just because a positive EducationPolicy or MacroEnvironment causes both to rise.

* 1. Draw a causal diagram of the variables listed in part b.

EducationPolicy

PSs’StudentSupportServices

PSs’Administration

MacroEnvironment

NationalStudentAchievement

FundingLevelOfPublicSchools

PSs’FacultyPerformance

FamilyIncome

EducationResources

EducationQualityOfPrivateSchools

* 1. Simplify the diagram from part e.

FundingLevelOfPublicSchools

EducationPolicy

NationalStudentAchievement

MacroEnvironment

1. Describe the kinds of situations that each of the following could be applied to in order to simplify a causal diagram.
   1. Unimportance: If the arrows coming in and out of a variable are likely to be tiny and unimportant effects, we can probably remove the variable.
   2. Redundancy: If there are any variables on the diagram that occupy the same space - they have the arrows coming in and going out of them from/to the same variables - we can probably combine them and describe them together (this works even if there are arrows between some of the variables being grouped together).
   3. Mediators: If one variable is only on the graph as a way for one variable to affect another (i.e. B in A → B → C where nothing else connects to B), then we can probably remove it and just have A → C directly.
   4. Irrelevance: Some variables are an important part of the data generating process but irrelevant to the research question at hand. If a variable isn’t on any path between the treatment and outcome variables, we can probably remove the variable.
2. How can a causal diagram be modified so as to avoid cyclic relationships?

As in the true data generating process and common statistical applications time is always a factor, whenever we have a cycle in our diagram, we can get out of it by thinking about adding a time dimension. It works because the cycles pop up when the arrows loop back on themselves, but time’s arrow only moves in one direction. Another way to break a cycle in a causal diagram is to find a source of random variation, like a randomized experiment, for one of the variables in the cycle, then if we just focus on the part of the variable driven by randomness, the effect can’t loop back on itself.

1. Think of a research question in your field of interest.
   1. What is the cause variable and what is the outcome variable?

The research question I think of is “does dieting really help an adult reduce weight?” The cause variable is dieting. The outcome variable is weight.

* 1. Write down a list of between 5 and 10 relevant variables in the data generating process.

Calorie intake, food structure, metabolic rate, how much exercise one does, eating disorders, genetics, and hormone level.

* 1. Draw a causal diagram incorporating all the variables from part b.

EatingDisorder

MentalHealth

CalorieIntake

HormoneLevel

Weight

FoodStructure

Diet

MetabolicRate

Genetics

Exercise

* 1. 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.

One’s mental health will affect one’s hormone level, but it is also possible for hormone level to impact one’s mental health which would create a cyclical relationship between hormone level and mental health if no time dimension is added. So for clarification, I had better annotate “MentalHealtht” and “HormoneLevelt+1.”

1. 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.  
   

Letʼs refer to different time periods as t , t + 1, t + 2, and so on, where t is “some particular time,” t + 1 is “the time right after that,” and so on.

Motivationt+1

Motivationt

StudentAchievementt+1

StudentAchievementt