Homework for Chapter 5: Identification

1. Think about the last time you sat down in a chair (perhaps right now). When you did that, you probably predicted you would *observe* that you would end up sitting in the chair, rather than passing through it or the chair breaking. List three assumptions you made about the data generating process when you made that prediction.

The effect of gravity: all things with mass or energy are attracted to one another.

The chair and I are both solid physical objects that cannot pass through each other.

The chair can support my weight.

1. Which of the following is the best definition of the term *identified* as in “this variation has identified the effect we’re interested in”? d
   1. We’ve generated the data by conducting a controlled experiment in which treatment is randomly assigned.
   2. In the data generating process, the only reason why we see variation in the outcome variable is because of the treatment variable.
   3. The relationship we are looking at in the data actually tests a hypothesis.
   4. In the variation we use, there’s no reason we’d see any relationship at all except for the effect we’re interested in.
2. Go to your favorite news source and find an article that describes the results of a new empirical study. Describe what you think are some features of the data generating process. What are some ways to explain their result other than the interpretation they had? Did the study (as described in the article) have ways of blocking out the alternate explanation you thought of?

The study funded by the National Institutes of Health of more than 2,000 couples shows that COVID-19 vaccination does not affect the chances of conceiving a child. The data generating process of chances of vaccinated and unvaccinated couples’ conception is complex in terms of potential confounders, diverse from couple to couple with various backgrounds, and dependent on different life stages and time phases the couples are going through. Factors such as time spent at home, personal planning on having children which could be greatly affected by a couple’s relationship, physical and mental health, phased income, and expectations on the macro environment, etc. all could have effects on chances of conception in addition to the COVID-19 vaccination. It could be that COVID-19 vaccination does reduce chances of conception, but at the same time, the time for the couples spent at home increases a lot during the pandemic due to quarantine, lockdowns, and remote working which increases couples’ contact time and frequency of sex. Being in a period of unrest may also trigger couples’ tendency towards growing intimate dependence on each other which could also raise the reproduction willingness and activities while offsetting the negative effect of vaccination. By looking into factors they thought that could potentially influenced the results, such as types, doses, and recency of vaccination as well as the nationality, occupation, and history of infertility of participants, the researchers (as described in the article) did not introduce ways to block out the alternate explanation I thought of.

1. You read about a new study with the headline “eating caviar linked to longer lifespan.” The study’s research question is “does eating caviar make you live longer?” In the study’s data, they find that people who eat caviar have, on average, longer lifespans than people who don’t.
   1. What are some alternate explanations for this relationship?

It could be that people who eat caviar often tend to be wealthier than people who don’t (as we know caviar is usually expensive) and thus can afford to live healthier lifestyles (better nutrition, more time and space to exercise) and have more access to medical resources. This could all make high-income people who happen to eat caviar have, on average, longer lifespans than people with lower income who don’t eat caviar.

* 1. What sort of variation would identify the answer to the research question?

A variation in people’s lifespans that is only driven by whether or not they eat caviar when we control for all the other factors potentially impacting the lifespan.

* 1. Give one suggestion for how the study authors might isolate variation that would identify the answer to the research question

Researchers may randomly sample participants among people with similar income or family background and compare the average life spans of the people who eats caviar and those who don’t to see if there is still considerable variation.

1. For each of the following news headlines, assume that the underlying data actually only shows a correlation between the two variables mentioned. Give an alternate explanation for the correlation other than the causal relationship implied by the headline.
   1. “As stock market drops, presidential approval ratings decline.”

Temporary recessions may lead to both the declines of the stock market and presidential approval ratings.

* 1. “Dates are announced for the downtown summer concert series, driving up sales at downtown restaurants.”

The downtown famous of its lotus garden tends to have more visitors during the summer which drives up sales at downtown restaurants. As more people would visit the downtown to enjoy the lotuses during their summer vacation, some organizers decide to hold summer concert series at downtown.

* 1. “Unsanitary? Hospital visits linked to 20% increased risk of disease.”

A majority of people who visit the hospital have concerns over their health. People concerned about their health conditions are likely to already develop some symptoms or have a history of risky activities before they visit the hospital which make them go to the hospital in the first place. So it makes sense that we observe that people who go to the hospital are more likely to have diseases compared to those who do not.

* 1. “Dress for success! Every CEO follows this office-wear rule.”

It is not necessarily dressing well at work will lead to success, but could be that people with higher family income have more human and educational resources than people with lower family income which makes the former more likely to obtain senior positions like CEOs and thus considered to be successful. Meanwhile people from wealthier family tend to have more money to spend on what to wear and are likely to pay more attention to their outfits than average people do to befit their social status. So dressing well and being successful may both be results of one’s family background.

1. Why is a variable that causes both the “treatment” and “outcome” variables especially concerning for identification? You may want to use the phrase “alternate explanation” in your answer.

A variable that causes both the “treatment” and “outcome” variables may offer an “alternate explanation” for the results other than the variation of interest, but to identify the answer to a research question, we need to isolate only the variation on the part of our variable of interest by closing off these possible alternate explanations, namely getting rid of variation on the part of other variables.

1. Shoe company Crikey claims that people who wear their fancy and expensive professional running-shoe Cool Mistrunner brand run 4 to 5% faster than if they wore an average shoe.
   1. In a few sentences, describe the data-generating process (you will probably leave some things out, that’s okay).

The fancy and expensive professional running-shoe Cool Mistrunner produced by Crikey provides the runners with the mesh upper and sweat-absorbing padding that improve breathability, as well as the wide forefoot and responsive high foam cushioning that provide stability to prevent rolling. Such technological construction makes the Cool Mistrunner lighter and more comfortable to wear even in hot weather so that runners can run smoother and faster than in average shoes without the design.

* 1. What are possible alternative explanations for this claim, aside from the shoe making the person run faster?

Buying and wearing any fancy and expensive running shoes, not necessarily the Cool Mistrunner, makes people feel good about themselves and have more faith in running (effect of autosuggestion). Being in a good mood and with the relatively big pre-investment in the running gears, people tend to be more motivated to pay more effots to running and run regularly, to make the best use of the shoes or probably to show off their fancy running gears, which will improve their running capability. Experienced runners are also more likely to buy more expensive professional shoes than beginners which may let us observe that people wearing expensive professional shoes run faster than people who wore average shoes.

* 1. In running their study, the researchers accounted for some alternative explanations, including: gender, enthusiasm for running, and whether runners have participated in marathons and/or half marathons. Think of an alternative explanation not on this list. What is the implication of not accounting for this alternative explanation?

The venue for running matters in terms of running: people are more likely to run faster in an indoor venue on some treadmills or flat ground without being affected by weather and other external uncertainties than in an open-air field with complex and uneven topography such as trails in the woods. It is possible that people often choose to wear the fancy and expensive shoes indoors in order not to easily ruin the shoes under sudden bad weathers or with the mud and rocks outside. Not accounting for this alternative explanation will probably lead to the “conclusion” that people who wear Crikey’s fancy and expensive professional running-shoe Cool Mistrunner brand run faster than if they wore an average shoe while in fact it is just that people run faster in a relatively more comfortable environment (indoors) than in the else (outdoors).

1. Which of the following terms describe how a variable changes from observation to observation? b
   1. Data-generating process
   2. Variation
   3. Raw data
   4. Determinants