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.
2. Which of the following is the best definition of the term *identified* as in “this variation has identified the effect we’re interested in”?
   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.s
   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.
3. 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?
4. 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?
   2. What sort of variation would identify the answer to the research question?
   3. Give one suggestion for how the study authors might isolate variation that would identify the answer to the research question
5. 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.”
   2. “Dates are announced for the downtown summer concert series, driving up sales at downtown restaurants.”
   3. “Unsanitary? Hospital visits linked to 20% increased risk of disease.”
   4. “Dress for success! Every CEO follows this office-wear rule.”
6. 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.
7. 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).
   2. What are possible alternative explanations for this claim, aside from the shoe making the person run faster?
   3. 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?
8. Which of the following terms describe how a variable changes from observation to observation?
   1. Data-generating process
   2. Variation
   3. Raw data
   4. Determinants