Quiz 1

02/01/2023

This quiz is due on Monday, February 6$^{\text{th}}$ *before* class (that is, by 6:05 Eastern Time). All nine problems below come from the exercises at the end of the assigned chapters. In a couple of instances I made very slight changes to their wording or notation, but nothing that should change an answer. You can submit your answer as a Word document or a .pdf, and you can submit it via email or Brightspace. If you submit some weird format, I have to be able to open it by the time it's due. No shenanigans!

Questions from Chapter 2

Question 1

Consider the following three statements. Which ones describe a correlation, and which ones do not? Why?

(a) Most professional data analysts took a statistics course in college.
(b) Among Major League Baseball players, pitchers tend to have lower-than-average batting averages.
(c) Whichever presidential candidate wins Ohio tends to win the Electoral College.

Question 2

Consider the last statement from Question 1 about Ohio and presidential elections. Do you think it's useful for description? Forecasting? Causal inference? Why or why not?

Question 3

The table below shows some data on which countries are major oil producers and which countries experienced a civil war between 1946 and 2004. Are being a major oil producer and experiencing civil war positively correlated, negatively correlated, or uncorrelated? Explain your answer.

	Civil War	No Civil War
Oil Producer	7	12
Non-Oil Producer	55	94

Questions from Chapter 3

Question 4

Sarah says that she is hungry. John hands her a piece of pizza. Sarah eats the pizza and then declares that she is no longer hungry.

   (a) The fundamental problem of causal inference seems to say that you can't know that Sarah eating the pizza had a causal effect on her no longer being hungry. Is that right? Explain.
   (b) Do you think you nonetheless have good reasons to believe that eating the pizza had an effect on Sarah no longer being hungry? Explain why or why not.
   (c) Do you have good reasons for believing that John handing Sarah the pizza had a causal effect on her no longer being hungry? In your assessment, are the reasons to believe John's actions had a causal effect better or worse than the reasons to believe Sarah eating the piece of pizza had a causal effect?

Question 5

A government is considering making alcohol consumption illegal as part of a public health campaign. Let's think of making alcohol illegal as the treatment T. Write $T = 1$ if the government makes alcohol illegal and $T = 0$ if the government leaves alcohol legal. We will think of a binary outcome for each person either they drink alcohol or they do not. If person i drinks at treatment status T, we write her potential outcome as $Y_i^T = 1$, and if she doesn't drink, we write it as $Y_i^T = 0$.

Suppose the society is made up of three groups: the always drinkers, the legal drinkers, and the never drinkers. The always drinkers will drink whether or not alcohol is legal. The legal drinkers will drink if and only if alcohol is legal. The never drinkers won't drink whether or not alcohol is legal.

   (a) Write down, in potential outcomes notation and as a number (0 or 1), each of the two potential outcomes for each of the three groups.
   (b) Write down, in both potential outcomes notation and as a number (0 or 1), the causal effect of making alcohol illegal on drinking for each of the three groups.
   (c) Is there an effect, on average, of banning alcohol in this society?
   (d) Suppose you are out to lunch with some friends and one of them says, "My uncle lives in a place where they banned alcohol and all of his friends kept drinking, so I don't think the ban does anything." Explain, in terms of our example, why this isn't a convincing argument.

Question from Chapter 4

Question 6

"At least twenty billionaires dropped out of college before earning their fortunes, including Bill Gates and Mark Zuckerberg."

   (a) Does this mean that dropping out of college is correlated with becoming a billionaire? Why or why not?
   (b) Draw the two-by-two table that would allow you to assess whether dropping out of college is correlated with becoming a billionaire. Let's assume that exactly twenty people have dropped out of college and become billionaires, so you know what to put in one of the four cells. Make your best guess for the other cells. At the time of this writing, there are about 7.8 billion people in the world, and about two thousand billionaires. Do you think there is a positive or negative correlation between dropping out of college and becoming a billionaire?

(c) Given your guesses from part (b), what proportion of the non-billionaires would need to be college dropouts in order for the correlation to be negative? What proportion of the non-billionaires would need to be college dropouts in order for the correlation to be positive?

(d) If you're currently a college student deciding whether you want to drop out in the hopes of becoming a billionaire, you may want to restrict attention to people who actually started college. Do you think the correlation between dropping out of college and becoming a billionaire is more or less likely to be positive if we restrict attention to just people who start college?

(e) About 7 percent of the world's population has a college degree. And about a third of people who start college complete it. If we assume that everyone who becomes a billionaire started college, you should now have all the information you need to assess the correlation between becoming a billionaire and dropping out of college among those who start college. Is it positive, negative, or zero?

Questions from Chapter 9

Question 7

One of the authors of the readings was once at a meeting where he was briefed on the ways in which data analytics can improve universities' operations. The example the presenter was most excited about was from a data analytics team in a major research university's development (which is jargon for *fundraising*) department. The data analytics team had discovered the following correlation by analyzing years of data: alumni who donate to the university six years in a row are way more likely to be lifelong givers than are alumni who only donate five years in a row.

The presenter was excited because, in their view, this finding from the analytics team suggested a clear strategy to improve fundraising and alumni engagement. In particular, on the basis of this analysis, they had decided to make a major push to encourage alumni who had already given for five years in a row to give a sixth—the idea being that the evidence of a correlation between giving for six years and giving in the future suggested that giving in that sixth year had a big causal effect on future giving, so resources spent encouraging five-year givers to become six-year givers were being put to the best possible use. Provide two arguments, using the clear thinking skills you acquired in this chapter, to explain why this might not be a good plan.

Question 8

At the end of the discussion of violent and non-violent resistance in Chapter 1 of the assigned readings from *Thinking Clearly with Data*, the authors asked the following: *Why might the fact that there are more government crack-downs following violent protests than non-violent protests not mean that switching from violence to non-violence will reduce the risk of crack-downs?* They promised that you would be able to give a compelling answer by the end of Chapter 9. Let's see if they made good on their promise, and try identify at least one reason why the fact that violent protests are more often met with a government crack-down than non-violent protests is not compelling evidence that the use of violent protest tactics causes government crack-downs.