Homework for Chapter 23: Under the Rug

1. You are interested in whether drinking alcohol increases your chances of developing dementia later in life. For simplicity, let’s say you plan to do this by regressing whether someone has dementia at age 75 on the number of drinks per week they took at age 35. In developing your regression model, list three things that you would be uncertain about, like whether certain controls should be included, or what functional forms should be used. These should be things for which you could reasonably argue it should be done either way.
2. Which of the following is least likely to have *construct validity* issues?
   1. You want to know whether strange responses to opinion polls are just people doing the survey quickly and not paying attention. To measure whether people do the survey quickly, you offer an online survey and record the times they started and finished.
   2. You want to know the effect of earning more than $2,000 per month on health issues. So you use a measurement of monthly wage income as your income measure.
   3. You want to know the effect of social redistribution policies on a population’s happiness. As a measure of happiness, you use the results from a survey that asks people how happy they are on a scale of 1-10.
   4. You want to know how someone’s willingness to take risks affects whether they buy or rent their home. To measure risk-taking, you ask people a hypothetical question about whether they’d take a risk by running into a burning building to save someone.
3. An apparent pattern in political polling in the United States in the late 2010s is that polls about who you would vote for to be US President produce results that are more in favor of the Democratic candidate and less in favor of the Republican candidate than the actual vote ends up being. Assuming this isn’t driven by an actual big shift towards Republicans just before election day, this indicates a statistical bias in favor of Democrats in the polls relative to the election. How might the observer effect offer an explanation of this phenomenon?
4. You are looking at a program that installs very visible traffic cameras in an attempt to scare drivers into not speeding (since the drivers know the camera would see and ticket them). You have the camera set up to record the speed of each car as well, so you can tell whether it’s speeding or not. However, due to the way it recognizes cars, the camera will sometimes not realize that smaller cars or motorcycles are there until it’s too late, so it will recognize something was there (and will be able to tell you what kind of car is there) but sometimes not record its speed. So there’s some missing values of speed. Is the missingness in speed:
   1. Not in universe
   2. Missing completely at random
   3. Missing at random
   4. Missing not at random
5. You want to know whether installing a stop sign at an intersection reduces speeding through the intersection. Very conveniently for you, the city council recently wanted to add a bunch of stop signs but couldn’t afford them all and so chose which ones to install at random. You want to identify the effect of the signs by looking at how car speed changed after installation at those intersections, and comparing those changes to how speed changed at (arguably comparable) nearby intersections. How would SUTVA be violated here?
6. Which of the following is most likely to have a power law distribution?
   1. Variation in height-at-birth across babies
   2. Variation in number-of-copies-sold across books
   3. Variation in test scores across students
   4. Variation in debt-to-GDP ratio across countries