1. Do Exercise 0.6 on page 15 of the textbook. (Shown below for those who haven’t been able to purchase a book yet.)

0.6 Wine model. In his book Super Crunchers: Why Thinking by Numbers Is the New Way to Be Smart, Ian Ayres writes about Orley Ashenfelter, who has gained fame and generated considerable controversy by using statistical models to predict the quality of wine. Ashenfelter developed a model based on decades of data from France’s Bordeaux region, which Ayres reports as

\[ \text{WineQuality} = 12.145 + 0.00117\text{WinterRain} + 0.0614\text{AverageTemp} - 0.00386\text{HarvestRain} + \varepsilon \]

where WineQuality is a function of the price, rainfall is measured in millimeters, and temperature is measured in degrees Celsius.

a. Identify the response variable in this model. Is it quantitative or categorical?

b. Identify the explanatory variables in this model. Are they quantitative or categorical?

c. According to this model, is higher wine quality associated with more or with less winter rainfall?

d. According to this model, is higher wine quality associated with more or with less harvest rainfall?

e. According to this model, is higher wine quality associated with more or with less average growing season temperature?

f. Are the data that Ashenfelter analyzed observational or experimental? Explain.

2. Explain why cause and effect conclusions generally cannot be made based on observational studies, but can be made based on randomized experiments.

3. Suppose researchers want to know whether drinking at least two cups of coffee a day increases the risk of kidney cancer.

a. Could they conduct a randomized experiment to test this? Explain.

b. If they conducted an observational study and found that those who drink at least two cups of coffee a day had a higher rate of kidney cancer than those who did not drink at least two cups of coffee a day, could they conclude that drinking coffee increases the risk of kidney cancer? Explain why or why not.

c. Identify a possible confounding variable in this situation, and explain how it meets the two criteria for confounding variables (given in Lecture 1).

For Exercises 4 to 7: Refer to the data set MedGPA accompanying the book and used as an example in class. You can download the data sets from the website accompanying the book, linked to the course webpage. The full list of 11 variables provided in the MedGPA dataset is as follows:

- AcceptStatus: A = accepted to medical school or D = denied admission
- Acceptance Indicator for Accept: 1 = accepted or 0 = denied
- Sex: F = female or M = male
- BCPM: Bio/Chem/Physics/Math grade point average
- GPA: College grade point average
4. Specify whether each of the 11 variables is Categorical or Quantitative. If the variable is Categorical, state whether either of the terms “binary” or “ordinal” are appropriate for that variable.

5. Suppose the following research questions are of interest. In each case, specify which of the variables in this data set would be the response variable, and which would be the explanatory variable(s). Note that each question has only one response variable, but could have multiple explanatory variables.
   a. Do verbal reasoning scores differ on average for males and females?
   b. Are equal proportions of males and females accepted into medical school?
   c. Is GPA a good predictor of MCAT score?
   d. Which of the subscores on the MCAT is the best predictor of whether or not someone is admitted to medical school?
   e. Which of the variables BCPM GPA, full College GPA or MCAT score is the best predictor of whether someone gets into medical school?

6. Read the data set into R.
   a. Compute the mean and a five-number summary for GPA, separately for those who were admitted and those who were denied admission to medical school.
   b. Write a few sentences comparing the GPAs for the two groups (admitted and denied).

7. Continue using the MedGPA data set in R.
   a. Find the regression equation for predicting MCAT score based on GPA. Write the equation.
   b. Interpret the intercept and slope values in the context of this example.
   c. Use the equation to predict the MCAT score for someone who has a GPA = 3.0 and for someone who has a GPA of 4.0.