Show **all** of your work on this assignment and answer each question fully in the given context. *Please* staple your assignment!

1. **Chapter 1, Exercise 1 (page 23)**

2. **Chapter 1, Exercise 9 (page 24)**

3. **Lurking variables.**

   There is an common saying in the sciences about establishing cause and effect: "correlation does not imply causation". It summarizes the idea that two events can have a relationship without one causing the other. Examples are easy to think of - suppose I did an observational study in which I looked at the percent of people using an umbrella and the percent of people using windshield wipers, recording the two percentages every hour of the day for two weeks. After collecting my data, I will discover that every time the percent of people using umbrellas rose the percent of people using windshield wipers also rose. The reason for this is not that people using umbrellas cause people to use windshield wipers. Instead both umbrella use and windshield wiper use are responses to a third variable, specifically whether or not it is raining. The danger in a scenario like this is that by just recording the percentage of people using umbrellas and the percentage of people using windshield wipers, I may falsely determine that one causes the other. In statistics we call variables that are unobserved but potentially the true causative factor **lurking variables** - variables that are not directly seen in the data set, but whose influence is.

   Consider the following cases where two variables are gathered and a pattern in the data is noticed. A cause and effect relationship is suggested between the two variables. Suggest a possible lurking variable in each case that might explain why the cause and effect relationship suggested could be wrong.

   note: this question is fairly open ended - as long as you attempt to explain how the lurking variable you suggest could be driving both of the other variables you will get full credit. For example, in the umbrella/windshield wiper example, a satisfactory response would be "The presence of rain may be a lurking variable. When it rains, people use umbrellas to stay dry and windshield wipers in order to drive safely. When it is not raining, neither umbrellas or windshield wipers are needed. Because of this, both percentages would be higher when it is raining and lower when it is not raining."

   (a) Researchers collecting data on outdoor temperature and the number of cases of common illnesses notice that when the temperature is lower more people get sick and when the temperature is higher fewer people get sick. They suggest that cold temperatures are causing these illnesses.

   (b) Researchers collecting data on work history notice that people are more likely to pass away in the five years after retirement than they are to pass away during any five year window of their careers. The conclude that retirement causes the person to pass away.

   (c) Over five years, the Coast Guard gathered monthly data for the number of swimsuit sales and the number of shark attacks. They noticed that when the number of swimsuits sold in a month was higher, the number of shark attacks in that month was also higher. They concluded that swimsuit sales were causing shark attacks.

4. **Hockey game attendance.**

Caroline performs the following study to see if outside temperature has an effect on attendance at her college's hockey games. For each hockey game at her college, Caroline records the outside temperature and the attendance. Here are her results:

Date		*Temperature, deg. F*	*Attendance*
Friday	12/14	35	840
Wednesday	12/19	20	560
Tuesday	1/ 8	−5	340
Friday	1/11	23	775
Wednesday	1/23	14	680
Saturday	2/ 2	30	950
Friday	2/ 8	28	950

(a) Is this an experiment or observational study?

(b) What type of variable is attendance?

Caroline analyzes her results and finds that outside temperature and attendance have a strong positive correlation (i.e., as one increases, the other also increases). She concludes that higher game day temperatures causes higher attendance at their college's hockey games.

(c) Did she come to a proper conclusion for this study? Why or why not?

(d) Look at the day of the week of the hockey games. What type of variable is this?

(e) Rewrite the data table, adding a new column "School Night" (using the values "no" if the game is on a Friday or Saturday, and "yes" if the game is on any other day). How does Attendance relate to School Night?

(f) For what type of studies do you have to worry about possible lurking variables affecting the results?

5. **Washer stretching.**

George works for a company that manufactures rubber washers. He randomly selects 1000 washers off the assembly line throughout two weeks for a study on the durability of these washers under stretching. To make sure that the washers are fit to be used in the real world, George must test the washers. Holding heat constant, George subjects each washer to one of various methods of stretching. The washers are randomly assigned to be stretched under one of five different forces (low, medium-low, medium, medium-high, and high). After each test, George classifies a washer as either defective or non-defective.

(a) Is this an experiment or observational study?

(b) What type of variable is heat?

(c) What type of variable is the amount of stretching?

(d) What type of variable is response to the stretching method?

(e) The 100 selected washers constitutes the sample. What is the population?

(f) George analyzes the results and finds that the defect rate increases with the amount of stretching. Can George conclude that the amount of stretching causes a change in the defect rate of the washers? Why or why not?