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)**

   Answer: (5 pts.) During calibration, our goal is to get the equipment to reflect the truth. This involves the use of standards that have some known quantity that should be measuring. Our main goal in this process is to be *accurate*, meaning that, as the book states on page 17, our instrument is to bring our measurements in line with the standard. (5 pts.)

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

   Answer: In this case, we can take the use the scare-quotes around the word "the" as a possible hint - while there may be some **ideal version** of the dowel in question, we have real, incosistent, physical dowels. The exact nature of each dowel, though identical in the large picture, may be less than identical in the details - slight differences in the texture or moisture content of the wood for instance may not have a great influence on the dowels strength, but may be just enough to make the strengths of any two dowels non-identical. Measuring the strengths of several of these dowels may help us get a better picture of what "the" strength of the ideal dowel we have in mind actually is.

   Of course, even if each dowel were exactly the same, we have the issue of measurement error. Since we hope that the machine is well calibrated, we can imagine that there is no systematic flaw in the way the observed measurements differ from the true quantities we are measuring. We can write this idea in an equation:

   $$\text{Observed Measurement} = \text{Acutal Quantity} + \text{Random Measurement Error}$$

   If the error is truly random, making multiple measurements allows us to get a better idea what part of the measurements comes from the truth and what comes from the error.

   In both cases, testing several dowels gives us a chance to determine what the true value is by account for the existence of errors. (5 pts.)

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.

Answer: During the winter, people also tend to spend more time inside. Perhaps it is the **amount of time spent inside**, in closer contact with other people and their germs which may actually be driving up the number of people falling ill. (5 pts.)

(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. **Answer:** People who retire are also generally older. If a person has no plans to retire any time soon, they may be earlier in their career. Perhaps **the individual's age** is a lurking variable. (5 pts.)

(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. **Answer:** In this case the a potential lurking variable is the **the number of people swimming in the ocean** or perhaps, the season. In the summer people not only buy swimsuits but they also get in the ocean with sharks while wearing those swimsuits. (5 pts.)

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? **Answer:** Observational study. Caroline's role is passive; there is no assignment of temperatures to the games. In an experiment, the temperatures would be manipulated to make a greater temperature difference between the games, making the results more telling of the underlying relationship between attendance and outside temperature. (5 pts.)

(b) What type of variable is attendance? **Answer:** The response (this is the focus of the study). (5 pts.)

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? **Answer:** No, you cannot come up with a causation conclusion for observational studies—you can do so only for experiments. There may be (and is) a lurking variable responsible for the effect on attendance that is also correlated with outside temperature, thus making it *seem* that the temperature is causing the change in the response. (5 pts.)

(d) Look at the day of the week of the hockey games. What type of variable is this? **Answer:** This is a lurking variable as it could affect the response (attendance). (5 pts.)

(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? **Answer:** Attendance is noticeably higher for weekend games, during which the weather is also warmer. It may be that games have higher attendance when they are played on the weekend or when the weather is warmer. We do not know from this given study. (5 pts.)

(f) For what type of studies do you have to worry about possible lurking variables affecting the results? **Answer:** Observational studies. These results demonstrate the problem of drawing conclusions from observational studies. (5 pts.)

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? **Answer:** Experiment. George manipulates the washers by assigning them to certain treatment groups. In an observational study, one does not assign units to treatment groups. (5 pts.)

(b) What type of variable is heat? **Answer:** This is a controlled variable as heat is kept constant for all units. Controlled variables are a type of supervised variables. (5 pts.)

(c) What type of variable is the amount of stretching? **Answer:** The amount of stretching is one of the responses we record. In this case, it is a numeric variable. (5 pts.)

(d) What type of variable is response to the stretching method? **Answer:** By this I had in mind whether the the washer was rendered useless or not. In that case, it is qualitative data as the possible results (defective or non-defective) are categories, not numbers. (5 pts.)

(e) The 100 selected washers constitutes the sample. What is the population? **Answer:** The population is all the rubber washers produced by George's company on the assembly line during the two weeks. The company doesn't really care how these specific 100 washers perform–they want to learn about how their washers perform in general to stretching. (5 pts.)

(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? **Answer:** This problem kind of broke - here I meant the amount of force applied during stretching increased the defect rate of the washers. George can make a causation statement here. As long as he ensures that the extraneous variables are properly handled, either through controlling or adequate randomization of the units and testing order, then he has a high degree of confidence that increasing the force does affect the defect rate. (5 pts.)