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. **Chapter 2, Section 2, Exercise 1 (page 37)**

   Answer: Following the process described in Lecture 3:

   - $N = 38$,
   - $n = 5$,
   - $m = 2$,

   The instructions tell us to repeat the selection process where the previous process ended. The first two rows of table B.1 are:

   `97842 69095 25982 03484 25173 05982 14624 31653 17170 92785`

   `53047 13486 69712 33567 82313 87631 03197 02438 12374 40329`

   Since $m = 2$ we can rewrite this as the two digit values we will be looking at:

   `97 84 26 90 95 25 98 20 34 84 25 17 30 59 82 14 62 43 16 53 17 17 09 27 85`

   `53 04 71 34 86 69 71 23 35 67 82 31 38 76 31 03 19 70 24 38 12 37 44 03 29`

Working left to right, keeping new entries that are between 1 and 36 gives us the following samples:

- Sample 1: 26, 25, 20, 34, 25
- Sample 2: 17, 30, 14, 16, 09
- Sample 2: 27, 04, 34, 23, 35
- Sample 2: 31, 38, 31, 03, 19

(5 pts.)

4. **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. If I had made note of that during the process, I might have noticed that. In statistics we call variables that are unobserved but potentially the true causative factor **lurking variables** - they variables are not directly seen in the data set, but their influence is.

Consider the following cases where two variables are related and suggest a possible lurking variable in each case that might explain why the relationship is not a cause and effect.

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, an 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."

(a) People are more likely to get sick during the winter. They are less likely to get sick during the summer.
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 who may be sick which causes the change in the number of people falling ill. (5 pts.)

(b) People are more likely to pass away after retiring from work than they are if they are not planning to retire any time soon.
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. So, perhaps **age** is a potential lurking variable. (5 pts.)

(c) When lots of swimsuits are being sold, the number of shark attacks is higher. When few people are buying swimsuits, the number of shark attacks is lower.
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 get

themselves new swimsuits but they also get in the ocean with sharks in those swimsuits. (5 pts.)

5. **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) How does game attendance relate to the day of the week (regarding weekday games vs. weekend games)? **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.)

6. **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 degrees of stretching. The washers are assigned to the different stretching groups randomly. 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 the factor under study (the variable George wants to see if it affects the response), thus it is an experimental variable (which is a type of supervised variable). (5 pts.)

(d) What type of data is recorded, quantitative or qualitative? **Answer:** Qualitative data is recorded 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, specifically all washers produces by this assembly line during these 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:** 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 stretching of the washers does affect the defect rate. (5 pts.)

7. **Measurement characteristics.**

Students Jack and Jill each use the same ruler to measure the thicknesses of various books book to the nearest millimeter. All books have the same thickness of 3.6 cm, a fact that is unknown to them. After doing the measurement five times, the following data are recorded:

Jack: 3.5 cm, 3.7 cm, 4.0 cm, 3.3 cm, 3.5 cm
Jill: 3.9 cm, 3.8 cm, 4.0 cm, 3.9 cm, 3.8 cm

(a) Who has more accurate data? **Answer:** Jack's data is more accurate as it is on average close to the true value of 3.6 cm. (5 pts.)

(b) Who has more precise data? **Answer:** Jill's data is more precise as her measurement-to-measurement variation is the smallest. But her observations are on average not close to the true thickness of the books (i.e., they have bias), so her observations are not as accurate as Jack's measurements. (5 pts.)