***Choose the one best answer for each question below by writing the corresponding letter in the blank to the left of the question. Each question is worth 1 point.***

**1.** What is the summary of all possible individuals of interest, whether examined or not, called?

\_\_\_\_\_ **A.** Parameter **B.** Population **C.** Gang **D.** Sample **E.** Statistic

**2.** What is the group of individuals actually examined in a statistical study called?

\_\_\_\_\_ **A.** Parameter **B.** Population **C.** Gang **D.** Sample **E.** Statistic

**3.** What is the symbol used to represent the sample standard deviation?

\_\_\_\_\_ **A.**  **B.**  **C.** Q3 **D.** x **E.** s

**4.** What is the symbol used to represent the sample mean?

\_\_\_\_\_ **A.**  **B.**  **C.** Q3 **D.** x **E.** s

**5.** Which measures should be used if the distribution is strongly right-skewed?

\_\_\_\_\_ **A.** x & s **B.** x & IQR **C.** x & range **D.** Median & s **E.** Median & IQR

***Answer each question below on separate sheets of paper. Make sure to clearly label each of your answers (e.g., #6 or #9a), put your name on each extra sheet used, and staple these questions to the top of your completed answer sheets to hand in.***

1. **[2 pts]** What are the two major goals of statistics? Define any statistical jargon in your answer!
2. **[3 pts]** What are three “realities” that, if they did not exist, would eliminate the need for the field of statistics?
3. **[5 pts]** Describe natural and sampling variability within the context of a “real-life” (or realistic) situation of interest to you. Clearly describe the situation (e.g., “Consider the situation where …”) and then specifically define the two types of variability within the context of that situation (e.g., *“Within this situation, natural variability is when XXX and sampling variability is when XXX”*). Please use a different example than I used in the Module 1 Answer Key.
4. **[8 pts]** Identify the specific type of variable for each situation below:
5. Loudness (db) of bathroom hand dryers.
6. Individual sheets of paper towel used in the 2nd floor CSE bathroom each day.
7. Employment position in a hospital (doctor, nurse, support staff, executive administration).
8. How long (years) a patient survives following removal of a cancerous tumor.
9. **[12 pts]** The Northland College Student Association was interested in the percentage of Northland students enrolled in Winter, 2017 that supported a resolution that the Board of Trustees divest Northland’s investment portfolio of companies with socially or environmentally questionable business practices. To estimate this percentage they asked a random selection of 90 students whether they supported the resolution or not. Use this information to identify the **I**ndividual, **V**ariable, **Po**pulation, **Pa**rameter, **Sa**mple, and **St**atistic. [*Clearly label your answers with I, V, Po, Pa, Sa, and St.*]
10. **[6 pts]** Determine if each situation below represents an experimental or observational study. *Explain your reasoning.*
    1. Administrators at Northeastern College developed two programs that they hoped would alleviate homesickness among first-year students. Students in the first program would interact (e.g., eat a meal, go to a movie) weekly with a family from the local community. Students in the second program would participate in weekly outdoor experiences (e.g., go walking, go skiing). To test the effectiveness of these programs, the administrators randomly allocated incoming first-year students into three groups -- one would participate in the first program, another in the second program, and the third in neither program. Level of homesickness was measured for all students from an end-of-the-semester questionnaire.
    2. Researchers gave students at Southwestern College a questionnaire that ultimately could be summarized into a measure of “homesickness.” The researchers were interested in determining if homesickness differed between “in-state” and “out-of-state” students.
11. **[4 pts]** Determine the type of observational study for each situation below.
    1. A Northland student was interested in the percentage of Northland students that used their own mug for coffee, rather than the paper cups provided. To investigate this, the student researcher recorded visual observations of students getting coffee while the student researcher was eating breakfast in the cafeteria.
    2. Researchers at George Mason University contacted 1860 random individuals in early December 2018 and asked them to participate in a web-based survey about climate change. A total of 1114 people completed the web-based survey. One summary they computed from this survey was the percentage of respondents that agree with the statement that “global warming is happening.”
12. **[4 pts]** Which variable is the response/dependent variable in each situation below.
    1. Researchers at George Mason University wanted to determine if a person’s level of agreement with the statement “global warming is happening now” differed based on the person’s level of education (less than high school, high school, college, advanced degree).
    2. Researchers want to determine if the winning percentage of a hockey coach depends on the coach’s years of coaching experience and whether the coach is from Canada or not.
13. **[9 pts]** A Bayfield High School student was interested in determining the effects of seat height (inches) and tire pressure (PSI; pounds per square inch) on his time to complete a particular route. He tried seat heights of 26, 30 and 32 inches and tire pressures at the level recommended by the manufacturer, 2 PSI below the recommendation, and 5 PSI below the recommendation. The student rode his bike a total of 45 times (he did this by riding his bike three times on each of 15 days with randomly chosen seat heights and tire pressures so as not to be too tired on some rides). He recorded the time to complete the route for each ride. Use this information to answer the following questions.
14. What is the response variable?
15. What is/are the factor(s)?
16. What is/are the number of levels?
17. What is the number of treatments?
18. What is the number of replicates per treatment?
19. In this particular study, what is a replicate?
20. **[6 pts]** Describe the three major principles of experimental design and why each is important.
21. **[4 pts]** Researchers at George Mason University asked 1114 people how much they thought global warming will harm plants and animals (excluding humans). Table 1 contains the percentages of respondents by their response. Perform an EDA from these results.

**Table 1.** Percentage of respondents by response to global warming’s harm of plants and animals.

A Great Moderate Only A Not Don’t

Deal Amount Little At All Know

58% 16% 8% 9% 9%

1. **[10 pts]** A brewery in Minneapolis was interested in how their new cider matched up to other ciders in terms of alcohol by volume (ABV; %). A histogram and descriptive statistics for their sample of 225 ciders is in Figure 1 and Table 2, respectively. Perform an EDA from these results.

|  |  |
| --- | --- |
| **Figure 1.** Histogram of alcohol by volume for ciders. | **Table 2.** Statistics for alcohol by volume for ciders.  mean 4.64  sd 0.83  min 0.50  Q1 4.00  median 4.50  Q3 5.00  max 7.40 |

***Answer questions 18-21 using the following data sets.***

**Data Set #1** 🡪 13, 224, 13, 157, 68.

**Data Set #2** 🡪 13, 224, 13, 157, 68, 368, 35, 13, 90, 139, 46.

***Make sure to clearly identify (e.g., circle) your final answer and show ALL of your work (i.e., just providing the final answer will not receive full, if any, credit).***

1. **[3 pts]** What is the mean of Data Set **#1**?
2. **[8 pts]** What is the standard deviation of Data Set **#1**?
3. **[4 pts]** What is the median of Data Set **#2**?
4. **[7 pts]** What is the IQR of Data Set **#2**?