***Answer each question below on separate sheets of paper. Make sure to clearly label each of your answers (e.g., #1 or #4a), put your name on each extra sheet used, and staple these questions to the top of your answer sheets when completed to hand in.***

1. **[4 pts]** What are the two major goals of statistics?
2. **[6 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. Make sure to clearly articulate the situation (e.g., “Consider the situation where …”) and then specifically define the two types of variability within the context of this 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 on the Module 1 Answer Key.
4. **[8 pts]** Identify the specific type of variable for each situation below:
5. Weight (to the nearest gram) of a newly hatched Penguin chick.
6. A respondents answer to “How often in the past month did you have to cut the size of our meal because there was not enough food (never, a few times per year, a few times per month, a few times per week, nearly every day)?”.
7. County of residence (Ashland, Bayfield, Douglas, or Sawyer County).
8. Grade point average (on a four-point scale).
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 sample 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. 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.
    2. Administrators at Northeastern College developed two programs that they hoped would alleviate homesickness among first-year students. In the first program, a student would interact (e.g., eat a meal, go to a movie) weekly with a family in the local community. In the second program, a student 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 of which one would participate in the first program, another in the second program, and the third in neither program. Level of homesickness was recorded at the end of the first semester for all students.
11. **[2 pts]** Define “statistical inference.”
12. **[4 pts]** Determine the type of observational study for each situation below.
    1. Researchers at Northland’s Center for Rural Communities distributed a paper survey to 1000 randomly selected addresses (i.e., households). Of the surveys delivered, 387 surveys were completed and returned.
    2. A Northland Forestry student used a geographic information systems (GIS) map to identify a grid of 10 m2 plots in the ravine on Northland’s campus. She then randomly selected 15 of those plots and recorded the number of invasive Buckthorn plants in each plot.
13. **[4 pts]** Which variable is the response variable in each situation below.
    1. In another part of the study on “homesickness” among students at Southwestern College (see question 6a), the researchers asked the same students how far they were away from home. Ultimately, they wanted to determine if the level of homesickness could be explained by distance from home.
    2. Researchers recorded the distance (meters) that a hitting machine could hit a ball and the elevation (meters above sea level) of the stadium where the ball was hit to determine if elevation affected the distance the ball was hit.
14. **[9 pts]** Students designed a simple experiment to determine if the duration of a D cell battery differed significantly among four different types of batteries (Duracell, Rayovac, Energizer, and Eveready) used at three different temperatures (5, 15, 25oC). Each battery was placed into a high-drain (i.e., high energy usage) machine held in a chamber where the temperature could be strictly controlled. The time (in seconds) until the battery was “dead” was recorded. Twelve batteries of each type were available to the students. Use this information to answer the following questions.
15. What is the response variable?
16. What is/are the factor(s)?
17. What is/are the number of levels?
18. What is the number of treatments?
19. What is the number of replicates per treatment?
20. Physically, what is a replicate?
21. **[6 pts]** Describe the three major principles of experimental design and why each is important.
22. **[10 pts]** A Northland student examined the basal area (cm) of Hemlock at a site in Iron County. A histogram and descriptive statistics for his sample is presented in Figure 1 and Table 1, respectively. Use these results to perform an appropriate exploratory data analysis.

|  |  |
| --- | --- |
| **Figure 1.** Histogram of Hemlock basal area. | **Table 1.** Descriptive statistics of Hemlock basal area.  mean 1.98  sd 1.67  min 0.08  Q1 0.63  median 1.57  Q3 2.71  max 7.94 |

***Answer questions 13-16 using the following data sets.***

**Data Set #1** 🡪 31, 39, 4, 32, 57, 41

**Data Set #2** 🡪 31, 39, 4, 32, 57, 41, 15, 48, 69, 4, 26, 48, 47

***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. **[4 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. **[8 pts]** What is the IQR of Data Set **#2**?