***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 sheets of questions to the top of your completed answer sheets 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 in the Module 1 Answer Key.
4. **[8 pts]** Identify the specific type of variable for each situation below:
5. Number of visits to derekogle.com in a month.
6. Cell phone service used by a student (AT&T, Sprint, T-Mobile, US Cellular, Verizon, Other).
7. Number of crayfish (i.e., crawdads) captured in a trap.
8. Amount of time (in seconds) that a user spends on derekogle.com per visit.
9. **[12 pts]** A researcher is interested in determining the mean density of Eastern Tent Caterpillars (*Malacosoma americanum*) in the Chequamegon Forest in Bayfield County in 2018. They recorded the number of Eastern Tent Caterpillars observed in each of twenty 1-hectare square plots in the Chequamegon Forest in Bayfield County in 2018. Their goal is to compute the mean number of caterpillars per hectare (i.e., per plot). 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 are interested in the effect of amount of light (measured in units called lux) on the growth of a plant. To test this, the researcher planted 40 seedlings each in separate pots. Each pot was held in a greenhouse with controlled temperature, watering, etc. Each pot, however, was exposed to one of three different levels of light. After one month, the researcher recorded the amount that each seedling had grown.
    2. A politician was interested in gauging the opinions of her electorate on environmental policy. To assess the electorate’s collective opinion, she sent a written survey to 1200 randomly selected registered voters in her district. A total of 238 surveys were returned.
11. **[2 pts]** Define “statistical inference.”
12. **[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 observations of students getting coffee while the student researcher was eating breakfast in the cafeteria.
    2. Prof. Schanning distributed a survey through the mail to 2000 randomly-selected households in Wisconsin, Minnesota, and Michigan to determine attitudes towards the reestablishment of Wolves in the Midwest. A total of 486 surveys were completed and returned to Prof. Schanning.
13. **[4 pts]** Which variable is the response variable in each situation below.
    1. A travel agency is interested in determining if the amount of time it takes to travel between Duluth, MN and Chicago, IL differs depending on the mode of transportation (personal vehicle, bus, train, or airplane).
    2. Consultants for the insurance industry are interested in determining if the number of firefighters needed to fight a structure (e.g., home) fire can be used to predict the financial damage to the structure caused by the fire.
14. **[9 pts]** Students performed a simple study to determine if gas mileage for a Toyota Prius depended on the type of gasoline (87 octane with 10% ethanol, 89 octane with 10% ethanol, and 90 octane with no ethanol) and the amount of hills driven (<10% hills, 10-25% hills, >25% hills). The students had enough resources (time and money) to drive the same car a total of 36 times for 100 miles at approximately 55 mph. At the end of each 100 mile drive, the students recorded the number of gallons of gasoline used by the Prius. 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. In this particular study, what is a replicate?
21. **[6 pts]** Describe the three major principles of experimental design and why each is important.
22. **[10 pts]** The current ages of all individuals in the United States House of Representatives are summarized in Figure 1 and Table 2. Perform a thorough univariate EDA with these data.

|  |  |
| --- | --- |
| **Figure 1.** Histogram of Representative’s ages. | **Table 2.** Summary statistics of Representative’s ages.  Mean 58.1  St. Dev. 10.1  Min. 33.0  1st Qu. 51.0  Median 59.0  3rd Qu. 66.0  Max. 79.0 |

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

**Data Set #1** 🡪 227, 134, 101, 312, 101

**Data Set #2** 🡪 227, 134, 101, 312, 101, 245, 156, 456, 123, 101, 178

***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**?