***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., #1 or #4a), 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?
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. Person’s home continent (Africa, Asia, Australia, Europe, N America, S America, Antarctica).
6. Amount of rainfall (cm) that Jasper, Alberta receives per year.
7. Household income (<$20000, $20000-50000, $50000-100000, $100000-500000, >$500000).
8. How long (years) someone has owned their “smart phone.”
9. **[12 pts]** A Sustainable Community Development student was interested in the mean amount of money that Wisconsin rural communities spent on so-called “green amenities” (e.g., solar array fields, runoff storage wetlands). To examine this question, the student sampled 34 rural communities in 2012 and determined from their published budgets the amount that each spent on “green amenities.” 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 were interested in whether fat thickness on seals (a measure of energy reserves) differed between wild and domestic seals or between the sex of the seal. To examine this, the researchers sampled both male and female seals from the Seattle and Portland (OR) zoos (i.e., the domestic seals) and from off the Olympic Peninsula (i.e., the wild seals). For each of 37 seals sampled they measured and recorded fat thickness (cm).
    2. Students in the Vertebrate Physiology class examined data to determine if caffeine affected the resting heart rate of humans. For a random group of 28 student subjects, the student researchers recorded resting heart rate both before and after the subjects were given a caffeinated soft drink. The amount of soft drink provided to each subject depended on the subject’s body weight (to adjust for the effect of body weight on uptake of caffeine).
11. **[4 pts]** Determine the type of observational study for each situation below.
    1. Sociologists gathered and analyzed English language tweets from 2.4 million people in 84 countries for over a year. The authors of the tweets were unaware that their tweets were being studied. The researchers used software that analyzed the meaning of words in the tweets and assessed their connections to moods and emotions, among other things. From these results they made conclusions about mood swings of humans throughout the day.
    2. The Gallup Poll organization contacted 3404 random households from June 5-8, 2014. From those households, 1027 adults agreed to answer questions about their confidence in various parts of the government (the president, Supreme Court, etc.). From this, they computed the percentage of respondents in each level of confidence.
12. **[4 pts]** Which variable is the response variable in each situation below.
    1. The Wisconsin DNR is interested in determining if how much money a person makes affects that person’s level of satisfaction (a five-point ranking from “completely satisfied” to “not satisfied at all”) with the new deer harvest regulations.
    2. Researchers want to determine if the winning percentage of a hockey coach depends on the coach’s number of years of coaching experience and whether the coach is from Canada or not.
13. **[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.
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]** The Strategic Research Initiative polled 802 Wisconsin residents in Fall, 2013 and asked them “From what you know about the Affordable Care Act, also known as Obamacare, would you say that you strongly support, somewhat support, somewhat oppose, or strongly oppose this policy?” Table 1 contains the percentages of respondents by their level of support. Perform an EDA from these results.

**Table 1.** Percentage of respondents by level of support for the Affordable Care Act.

Strongly Somewhat Somewhat Strongly Not

Support Support Oppose Oppose Sure

22% 32% 12% 32% 2%

1. **[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 in Figure 1 and Table 2, respectively. Perform an EDA from these results.

|  |  |
| --- | --- |
| **Figure 1.** Histogram of Hemlock basal area. | **Table 2.** Summary 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 18-21 using the following data sets.***

**Data Set #1** 🡪 163, 70, 37, 248, 37

**Data Set #2** 🡪 163, 70, 37, 248, 37, 181, 92, 392, 59, 37, 114

***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**?