**Answer questions 1 and 2 with your final answer clearly identified (e.g., circled) and ALL of your work shown (i.e., just providing the final answer will not receive full, if any, credit).**

**Data Set #1** 🡪 78, 89, 54, 62, 90, 71 **Data Set #2** 🡪 78, 89, 54, 62, 90, 71, 45, 78, 99, 34, 56, 78, and 77

1. **[4 pts]** What is the mean of Data Set **#1**?

2. **[8 pts]** What is the standard deviation of Data Set **#1**?

3. **[8 pts]** What is the IQR for Data Set **#2**?

4. **[4 pts]** What is the median for Data Set **#2**?

**Answer questions 5-9 in the space provided. Please be as specific as possible.**

5. **[12 pts]** Two Northland students were interested in determining the average amount of playing time (mins) per game for NHL rookies that were drafted straight from professional (minors or European) leagues. To answer this question they generated a random sample of 25 NHL rookies that were drafted from a professional league. Use this information to identify the **I**ndividual, **V**ariable, **Po**pulation, **Pa**rameter, **Sa**mple, and **St**atistic.

6. **[9 pts]** Maret and Collins (1996) designed an experiment to test the effects of food ration (low, medium, and high densities of brine shrimp per day) and density of frog tadpoles (0, 1, or 2 tadpoles present) on the growth of larval salamanders. A total of 45 identical aquaria that could be maintained in a controlled laboratory environment were available for the experiment. Use this information to answer the following questions.

* What is the response variable? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* What is/are the factor variable(s)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* What is/are the number of levels? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* What is/are the number of treatments? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* What is the number of replicates per treatment? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

7. **[4 pts]** A recent Suffolk poll asked a sample of 1000 voters this question: “Intelligence agencies concluded that Russia tried to meddle in the U.S. presidential election last year. How serious of an issue do you think this is?” The percentage responses are shown in Table 1. Perform a thorough univariate EDA with these data.

**Table 1.** Percentage of respondents by answer to the seriousness of Russian meddling question.

Very Somewhat Not Very Not At All

Serious Serious Serious Serious

46.3% 16.3% 11.5% 19.8%

8. **[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.** Descriptive 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 |

9. **[6 pts]** Researchers examined the “broad” types of animals (“Mammal”, “Bird”, and “Amph(ibian)/Rep(tile)”) among four zoos. The number of animal types in each zoo is shown in Table 3. Use these results to answer the questions below Table 3. Round all answers *to one decimal place* and *show your work*.

Table 3. Frequency of animals by broad type and zoo location.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Broad Type of Animal** | | |
| **Zoo** | **Amph/Rep** | **Bird** | **Mammal** |
| **Chicago** | 27 | 66 | 70 |
| **Minnesota** | 4 | 13 | 52 |
| **San Antonio** | 168 | 218 | 69 |
| **San Diego** | 27 | 40 | 109 |

* What percentage of birds are in the Minnesota zoo?
* What percentage of animals in the Chicago zoo are birds?
* What percentage of animals in the San Diego zoo are mammals?

**library(NCStats)**

**distrib(val,mean=meanval,sd=sdval,lower.tail=FALSE,type=”q”)**

where **val** is a value of the quantitative variable or the area (i.e., percentage as a proportion).

**meanval** is the population mean.

**sdval** is the population standard deviation.

**lower.tail=FALSE** is included for “right-of” calculations.

**type=”q”** is included for reverse calculations.

10. **[12 pts]** The average number of acres burned per year by forest and range fires in Socorro County, New Mexico is approximately normally distributed with a mean of 4300 acres and a standard deviation of 750 acres. Use this information and RStudio to answer the questions below *to one decimal place*. *Include your R code*.

* What percentage of years have more than 5000 acres burned?
* What is the top 5% of amount of acres burned per year?
* What is the IQR for number of acres burned per year? [*Hint: Think carefully about what the IQR is.*]

**Multiple Choice [30 pts] -- choose the ONE BEST answer for each question by writing the corresponding letter in the blank to the left of the question.**

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

\_\_\_\_\_\_\_

1. New York Yankee Aaron Judge had three, zero, one, and one hits in four games. This difference between individuals (i.e., games) is an example of what type of variability?

**A.** Natural **B.** Extreme **C.** Process **D.** Variable **E.** Sampling

1. Three samples of houses in Puerto Rico found that 91%, 85%, and 93%, respectively, were severely damaged by Hurricane Maria. This difference among statistics is an example of what type of variability?

**A.** Natural **B.** Extreme **C.** Process **D.** Variable **E.** Sampling

1. What is the summary of all individuals, whether actually examined or not?

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

1. What is the individuals that were actually examined?

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

1. What type of variable is the category of hurricane (i.e., 1, 2, 3, 4, and 5)?

**A.** Nominal **B.** Ordinal **C.** Response **D.** Continuous **E.** Discrete

1. What type of variable is wind speed (measured in miles per hour)?

**A.** Nominal **B.** Ordinal **C.** Response **D.** Continuous **E.** Discrete

1. Which graph would best be used to examine the distribution of people’s favorite breakfast meat (e.g., none, bacon, sausage, ham, spam)?

**A.** Bar Chart **B.** Dot Plot **C.** Histogram **D.** Scatterplot **E.** Stemplot

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

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

1. The mean is \_\_\_\_\_\_\_\_ the median for an extremely left-skewed distribution.

**A.** less than **B.** equal to **C.** greater than **D.** five times **E.** a sibling of

1. On any normal distribution, what proportion of the individuals are within +1 of ?

**A.** 0.680 **B.** 0.900 **C.** 0.950 **D.** 0.997 **E.** 1.000

1. What type of normal distribution question is “What is the square footage of a house such that 40% of houses are smaller?”

**A.** forward, left-of **B.** forward, right-of **C.** forward, between **D.** reverse, left-of **E.** reverse, right-of

1. What type of normal distribution question is “What percentage of college softball teams have more than 20 players on their roster?”

**A.** forward, left-of **B.** forward, right-of **C.** forward, between **D.** reverse, left-of **E.** reverse, right-of

1. What is the name of the variable that we are interested in predicting or explaining?

**A.** Continuous **B.** Discrete **C.** Explanatory **D.** Response **E.** Sampling

1. What type of study is it if individuals choose to be (or not to be) part of the sample?

**A.** Convenience **B.** Inference **C.** Simple Random **D.** Regression **E.** Voluntary Response

1. What type of study is it if the researcher makes it so that every individual has the same chance of being selected for inclusion in the sample?

**A.** Convenience **B.** Inference **C.** Simple Random **D.** Regression **E.** Voluntary Response

**Short (Paragraph) Answers [15 pts] -- Answer the following questions with complete sentences.**

26. Define natural and sampling variability. Provide a thoughtful (real or realistic) example that depicts each type of variability.

27. Describe the major principles of experimental design and why each is important.

28. Provide at least three reasons why observational studies may be valuable (even though strict cause-and-effect statements cannot be made).