Name: _____ **Date:** _____

PSTAT 5A: Homework 01

Summer Session A 2023, with *Ethan P. Marzban*

As a reminder, homework is neither collected nor graded. We encourage you to stop by Office Hours to ask any questions you may have about your work, or the problems themselves!

1. Consider the list of numbers $X = \{-3,\ -1,\ 0,\ 0.4,\ 0.7,\ 3.9,\ 6\}$.

(a) Compute $\overline{x}$, the mean of X.

Final Answer(s): 1

(b) Compute median(X), the median of X.

Final Answer(s): 0.4

(c) Compute the standard deviation of X.

Final Answer(s): $\frac{\sqrt{8229}}{30} \approx 3.023795$

(d) Compute the range of X.

Final Answer(s): 9

2. In the parts below, you will be provided with the description of a particular dataset. Identify the type of visualization (e.g. histogram, scatterplot, etc.) that you believe would best achieve the stated goal, and provide a brief justification for your answer. Keep in mind that there are potentially multiple "correct" answers- as such, your explanation/justification will be very important!

(a) An environmental scientist would like to see how (if at all) PM2.5 concentration (which is a measure of air quality) varies with temperature (as measured in Centigrade).

Final Answer(s): Scatterplot

(b) A clinical researcher has administered 4 different dosages of a particular medicine to a large set of volunteers, and would like to visualize how (if at all) the insulin levels of subjects varies across dosages.

Name:_____ **Date:**_____

Final Answer(s): Side-by-Side Boxplot, or possibly Scatterplot.

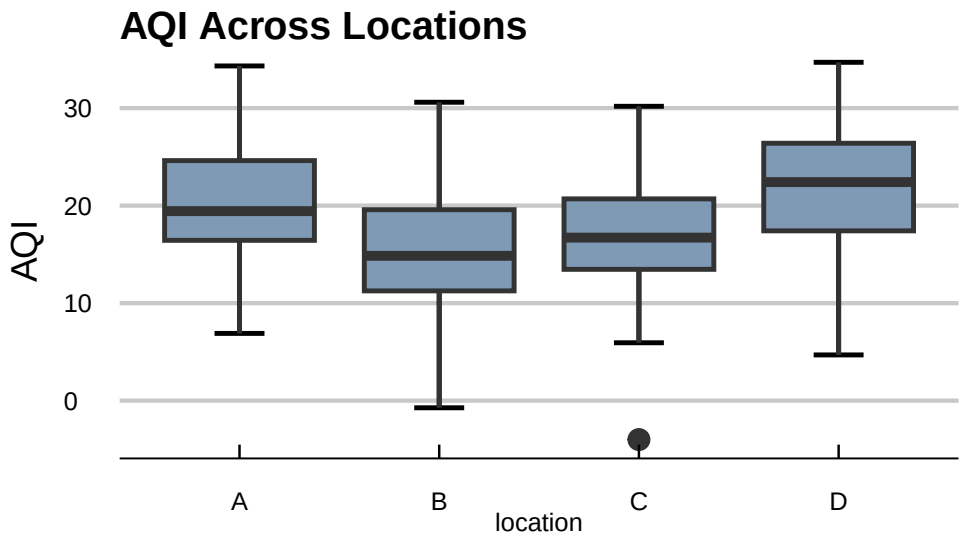
(c) An avid watcher of *Eurovision* has tallied up the number of times each country has won the competition, and would like to visualize their data.

Final Answer(s): Histogram

(d) Alex has collected information on how long it takes a sample of 60 PSTAT 5A students to complete the Final Exam, and would like to visualize the distribution of completion times.

Final Answer(s): Histogram

3. A researcher has collected measurements on the AQI (Air Quality Index) at several locations in Santa Barbara. She has labeled her locations A through D, and collected 100 AQI measurements from each location. The results of her study are displayed graphically below:



(a) What is the (approximate) median AQI at location A?

Final Answer(s): 20

(b) Approximately what percent of AQI readings at Location D were less than 16?

Final Answer(s): 25%

(c) Are there any outliers at any of the locations?

Final Answer(s): Yes, at Location C

(d) From the plot, does it appear that the different locations have different AQI readings? Explain briefly.

Final Answer(s): Yes

4. Consider a list of numbers $X = \{x_i\}_{i=1}^n$, and another list of numbers $Y = \{x_i + b\}_{i=1}^n$ where b is a fixed constant. In other words, the elements of Y are found by taking the elements of X and adding b.

(a) Show that $\overline{y} = \overline{x} + b$.

Final Answer(s): No numerical answer.

(b) What is s_Y^2 in terms of s_X^2?

Final Answer(s): No numerical answer.

5. A recent survey revealed that 40% of UCSB students own an *X-Box*, 70% own a *Playstation*, and 30% own both an *X-Box* and a *Playstation*. A UCSB student is selected at random.

(a) What is the probability that this student owns an *X-Box* or a *Playstation* (or both)?

Final Answer(s): $0.8 = 80\%$

(b) What is the probability that this student owns a *Playstation* but not an *X-Box*?

Final Answer(s): $0.4 = 40\%$

(c) What is the probability that this student owns neither an *X-Box* nor a *Playstation*?

Final Answer(s): $0.2 = 20\%$

6. Two numbers are to be selected at random and with replacement from the set $\{1, 2, 3, 4, 5\}$.

(a) Express the outcome space for this experiment using a table.

Final Answer(s): No numerical answer.

(b) Express the outcome space for this experiment using a tree.

Final Answer(s): No numerical answer.

(c) How many elements are in the outcome space?

Final Answer(s): 25

(d) For each of the following events, write down the mathematical formulation of the event described (i.e. list out the outcomes comprised in each event):

i. A = the first number selected was 3

Final Answer(s): No numerical answer.

ii. B = the second number selected was even

Final Answer(s): No numerical answer.

iii. C = the first number selected was 3 and the second number selected was even.

Final Answer(s): No numerical answer.

iv. D = the first number selected was odd, and the first number selected was even.

Final Answer(s): No numerical answer.

v. E = the first number was strictly greater than the second.

Final Answer(s): No numerical answer.

(e) Are we justified in using the Classical Approach to Probability in this problem? Why or why not?

Final Answer(s): Yes

(f) Use the Classical Approach to Probability to compute the probabilities of the events listed out in part (d) above.

Name: _____ **Date:** _____

Final Answer(s):

(i) $\mathbb{P}(A) = \frac{1}{5} = 0.2 = 20\%$

(ii) $\mathbb{P}(B) = \frac{2}{5} = 0.4 = 40\%$

(iii) $\mathbb{P}(C) = \frac{2}{25} = 0.08 = 8\%$

(iv) $\mathbb{P}(D) = 0$

(v) $\mathbb{P}(E) = \frac{2}{5} = 0.4 = 40\%$

7. On a particular website, passwords must be exactly 7 characters long and consist of 3 letters (A through Z), followed by 2 digits (0 through 9), followed by another letter (A through Z), followed by a special character (!, @, #, $, %).

(a) How many passwords can be created using this scheme, assuming repeated letters, digits, and characters *are* allowed?

Final Answer(s): 228,488,000

(b) How many passwords can be created using this scheme, assuming repeated letters, digits, and characters *are not* allowed?

Final Answer(s): 161,460,000

(c) Suppose now that the letters must still appear together, the digits must still appear together, and the special characters must still appear together, but the order in which these three categories of characters appear is now free to vary. For example, %A122AAB is now a valid password. (Again assume that repeated letters/digits/characters *are* allowed.) How many passwords can be created using this new scheme?

Final Answer(s): 5,483,712,000

