



## PSTAT 5A: Discussion Worksheet 01

Spring 2023, with Ethan P. Marzban

---

Welcome to the first PSTAT 5A Discussion Section! We encourage you to solve the following problems in groups. Statistics and Data Science are not meant to be lonely fields- we have quite a bit we can learn from each other! Your TA will go over what you need to do in order to receive credit for Discussion Section, so make sure to attend!

- Consider the list of numbers  $X = \{-1, 0, 1.5, 2, 3, 3, 4, 5, 10\}$ .
  - Compute  $\bar{x}$ , the mean of  $X$ .
  - Compute  $\text{median}(X)$ , the median of  $X$ . (HW01 provides a formula for how to compute the median of a list of numbers.)
  - Compute the standard deviation of  $X$ .
- 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!
  - A statistician is interested in visualizing the relationship between heights and weights of students at UCSB.
  - 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.
  - A soccer fan has tallied up the number of times every country has made it into the World Cup, and would like to visualize their data.
  - Morgan has collected information on how long it takes a sample of 100 athletes to complete an obstacle course, and would like to visualize the distribution of completion times.
- (A Bit of Math)** Consider two sets of data  $X = \{x_i\}_{i=1}^n$  and  $Y = \{y_i\}_{i=1}^m$  (note that  $X$  and  $Y$  have different sizes!) Define the set  $Z = X \cup Y$  (i.e.  $Z = \{x_1, \dots, x_n, y_1, \dots, y_m\}$ ); show that

$$\bar{Z} = \left(\frac{n}{n+m}\right) \cdot \bar{X} + \left(\frac{m}{n+m}\right) \cdot \bar{Y}$$

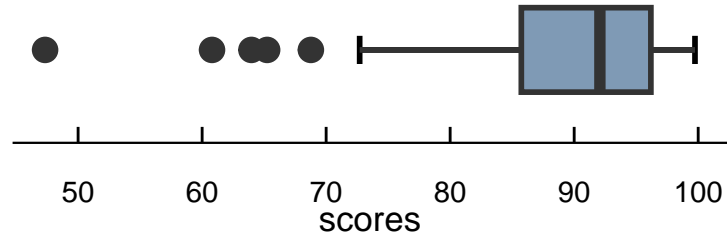
How does the formula for  $\bar{Z}$  simplify if  $n = m$  (i.e. if the two sets of data have the same size)?

Name: \_\_\_\_\_

Date: \_\_\_\_\_

4. The boxplot of scores on a particular exam is given below:

### Boxplot of Scores



- What score was the median score on this exam?
- Kara knows that she performed better than 25% of the class. What was Kara's score?
- How many outliers are there?
- Provide the 5-number summary for the distribution of exam scores.
- Compute the Interquartile Range (IQR) of exam scores.