

Name _____

Justify all answers by showing your work or by providing a coherent explanation. Please circle your answers

1. In a certain southwestern city the air pollution index averages 62.5 during the year with a standard deviation of 18.0. Assuming that the data follows a bell-shaped distribution, the index falls within what interval 95% of the time?
 - a. (8.5, 116.5)
 - b. (26.5, 98.5)
 - c. (44.5, 80.5)
 - d. (45.4, 79.6)
 - e. There is insufficient information to answer this question.

2. A 1995 poll by the Program for International Policy asked respondents what percentage of the U. S. budget they thought went to foreign aid. The mean response was 18%, and the median was 15%. (The actual amount is less than 1%.) What do these responses indicate about the shape of the distribution of all responses?
 - a. The distribution is skewed to the left.
 - b. The distribution is skewed to the right.
 - c. The distribution is symmetric around 16.5%.
 - d. The distribution is bell-shaped with a standard deviation of 3%.
 - e. The distribution is uniform between 15% and 18%.

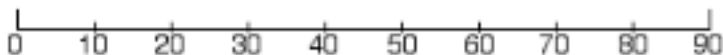
3. When a set of data has suspect outliers, which of the following are preferred measures of center and variability?
 - a. mean and standard deviation
 - b. mean and variance
 - c. mean and range
 - d. median and range
 - e. median and interquartile-range

4. The stem plot below is of a data set of size 127. What is the median of this set?

```

1 | 8
2 | 0 0 1 3 5 6 7 7 8 9
3 | 0 1 1 2 3 3 3 4 4 5 6 6 6 6 7 8 8 9 9
4 | 0 0 0 1 1 1 2 2 2 3 3 3 4 4 5 6 6 7 7 8 8
5 | 0 0 0 0 2 2 2 3 3 3 3 4 4 4 5 5 5 6 6 6 7 7 7 7 7 7 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9
6 | 0 0 0 1 1 1 2 2 2 3 4 4 5 5 6 6 7 7 8 8 9 9
7 | 0 1 2 5 6 6 7 9
8 | 1 2
  
```

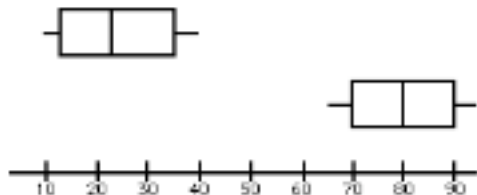
5. Sketch a box plot for the data set above.



6. Given the sample data set 1, 2, 2, 5, 6, 6, 8, 10, find the mean and standard deviation (unbiased).
7. Why is the standard deviation preferred over the variance as a measure of variability?
8. How does the mean absolute deviation (MAD) compare to the standard deviation above. Note that the calculation of MAD involves summing the absolute values of the deviations and then dividing by the number of data items. It is a biased estimator, so we must compare it to the biased standard deviation.
9. Where does the "bias" come from when using these estimators?
10. I can detect a slight skew in the data set from question #6, by comparing the median and mean. Is the skew positive or negative?
11. What is the purpose of z -scores and what would be the z -score corresponding to 2 in the data set described in question #6?

12. Which of the following statements are true?

- | | |
|---|---|
| a. The ranges are the same. | c. Because of symmetry, the medians are the same. |
| b. The interquartile ranges are the same. | d. all of the above |
| | e. a) and b) above |



13. In a group of 10 scores, the largest score is increased by 40 points. What will happen to the mean?
- | | |
|----------------------------------|---|
| a. It will remain the same. | c. It will increase by 10 points. |
| b. It will increase by 4 points. | d. It will increase by 40 points. |
| | e. There is not sufficient information to answer this question. |
14. In a group of 10 scores, the largest score is increased by 40 points. What will happen to the standard deviation?
- | | |
|----------------------------------|---|
| a. It will remain the same. | c. It will increase by 10 points. |
| b. It will increase by 4 points. | d. It will increase by 40 points. |
| | e. There is not sufficient information to answer this question. |
15. A data set has 5 elements. The mean of the data set is 48. If we know four of the data elements to be 20, 30, 50, and 70, what must be the fifth data element?