CHAPTER 1. DESCRIPTIVE STATISTICS

1.1. Statistics

- What is statistics?
- Definitions: Variable, data, population and sample
- Descriptive and Inferential statistics

1.2. Variables and Types of Data

Qualitative, quantitative, categorial, ordinal, discrete, continuous

1.3. Data organization, histograms and bar charts

1.4. Measures of:

- Central Tendency (Location):
    - Mean
    - Median
    - Mode
    - Midrange
- Variation (Dispersion)
    - Range
    - Variance
    - Standard Deviation
    - Coefficient of Variation
- Position
    - Z-score
    - Percentile
    - Decile and Quartile
    - Outlier

1.4. Data representation

- Box-plots
- Frequency distributions
- Graphs: Time series, Pie graphs, Scatter plots

1.5. Shapes of frequency distributions

- Shapes of distributions
- Skewness
- Kurtosis

Chapter 1: Assignments

1. Find the range, variance and standard deviation for the data set for the samples of Brand A and Brand B paint.

Brand A	Brand B
10	35
60	45
50	30
30	35
40	40
20	25

2. Generally, a small standard deviation implies that the measurements are clustered close to the mean. Is this statement true?

3. If the variance of a distribution is 9, the standard deviation is:
   a. 3
   b. 6
   c. 9
   d. 81
   e. impossible to determine without knowing n.

4. Is the variance of a set of negative numbers positive or negative?

5. If a constant were to be added to a set of measurements, the standard deviation would:
   a. remain the same.
   b. increase by the square root of that constant.
   c. increase by the square of that constant.
   d. increase by the magnitude of that constant.
   e. none of the above.

6. The mean of the measured temperature of a set of samples is 87 degree C, and the standard deviation is 5 degree C. The mean of the concentration of one chemical measured in those samples is 5225 mg/L, and the standard deviation is 773 mg/L.
   Compare the variations of the two.

7. Can the IQR be negative?

8. A student scored 65 on a calculus test, that had a mean (for the entire class) of 50 and a standard deviation of 10. She scored 30 on a history test with a mean (for the entire class) of 25 and a standard deviation of 5. Compare her relative positions within the class, on the two tests.

9. A teacher gives a 20-point test to 10 students. Find the percentile rank of a score of 12.
   18, 15, 12, 6, 8, 2, 3, 5, 20, 10

10. A teacher gives a 20-point test to 10 students. Find the value corresponding to the 25^{th} percentile.
18, 15, 12, 6, 8, 2, 3, 5, 20, 10

11. Find Q_1, Q_2, and Q_3 for the data set.
15, 13, 6, 5, 12, 50, 22, 18

12. Can the standard deviation ever be negative?

13. In this data set:
15, 13, 6, 5, 12, 50, 22, 18
Is there any outlier?

14. Let us define a new statistic as the distance between 70th sample percentile and 30th sample percentile. This new statistic would give us information concerning:
    a. central tendency.
    b. variability.
    c. relative position.
    d. skewness.
    e. symmetry.

15. Each year, during a period of seven years, Mrs. Smith gave birth to a child. The standard deviation of the ages (in whole years) of the 7 children of the family Smith is equal to
    a. 2
    b. 4
    c. 7
    d. Cannot be calculated if the present age of the children is not known

16. Is this statement true?: The standard deviation is a point in a distribution.

17. Is this statement true?: The standard deviation of a dataset is 0 when all the values are the same.

18. The mean of the population of ten scores, 78, 91, 91, 94, 74, 23, 63, 22, 78, 89 is 70.3, and the modes are 78 and 91. The skewness of the population is:
    1. negative
    2. zero
    3. positive
    4. not determined
    5. positive or negative depending on the score.

19. The values that have the greatest effect on the variance are those:
a. above the mean.
b. below the mean.
c. nearest the mean.
d. farthest from the mean.

20. A percentile score of 40 indicates that a person:
    a. answered 40% of the questions correctly on the test;
    b. knows 40% of the material covered by the examination;
    c. has earned a score equal to or better than 40 persons in his class;
    d. has earned a score equal to or better than 40% of the persons in his class.

21. Suppose that the 60th percentile of a sample was 1468.3. Explain in simple words what this fact tells you about this sample. You can for example talk to one of your colleagues sitting next to you.

22. The following data are the number of hours worked per week by seven College students:
    3, 7, 4, 6, 2, 8, 19
    Half (50%) of the values in a distribution are:
    a. included in the range
    b. between Q(1) and Q(3)
    c. between the mean and mode
    d. the mode and the highest value

23. The standard deviation of a dataset is 10. If 5 were subtracted from each measurement, the standard deviation of the new dataset would be:
    1. 2
    2. 10/25
    3. 5
    4. none of these.

24. If the mean, median and mode of a distribution are 5, 6, 7 respectively, then the distribution is:
    a. skewed negatively
    b. not skewed
    c. skewed positively
    d. symmetrical
    e. bimodal.

25. Construct two boxplots ("Type 1") for the data.
    33, 38, 43, 30, 29, 40, 51, 27, 42, 23, 31