SECTION 3.4 – Measures of Position and Outliers

z-Score

The ***z*-score** represents the distance that a data value is from the mean in terms of the number of standard deviations. We find it by subtracting the mean from the data value and dividing this result by the standard deviation. There is both a population *z-score* and a sample *z-score:*

Population z-Score

$$z = \frac{x - \mu}{\sigma}$$

Sample z-Score

$$z = \frac{x - \overline{x}}{s}$$

The z-score is unitless. It has mean 0 and standard deviation 1. This is also sometimes referred to the **standardized version** of x or the **standardized variable** corresponding to the variable x.

***k*th Percentile**

The ***k*th percentile** denoted P_k, of a set of data is a value such that k percent of the observations are less than or equal to the value.

Quartiles

The most common percentiles. **Quartiles** divide data sets into fourths or four equal parts.

First Quartile (or Lower Quartile)

The first quartile, denoted Q_1, divides the bottom 25% of the data from the top 75%. Therefore, the first quartile is equivalent to the 25th percentile.

Second Quartile (or Middle Quartile)

The second quartile, denoted Q_2, divides the bottom 50% of the data from the top 50%. Therefore, the second quartile is equivalent to the 50th percentile or the median.

Third Quartile (or Upper Quartile)

The third quartile, denoted Q_3, divides the bottom 75% of the data from the top 25%. Therefore, the third quartile is equivalent to the 75th percentile.

Finding Quartiles

Step 1 Arrange the data in ascending order.

Step 2 Determine the median, M, or second quartile, Q_2.

Step 3 Divide the data set into halves: the observations below (to the left of) M and the observations above M. The first quartile Q_1, is the median of the bottom half of the data and the third quartile Q_3, is the median of the top half of the data.

Interquartile Range, IQR

The **interquartile range,** or **IQR**, is the range of the middle 50% of the observations in the data set. That is, the IQR is the difference between the third and first quartiles and is found using the formula:

$$\text{IQR} = Q_3 - Q_1$$

Outlier

A value in the data set that is an extreme value. An extreme value is any value that is significantly larger (or smaller) than the other values. We can calculate outliers by using quartiles.

SECTION 3.4 – Measures of Position and Outliers

Checking for Outliers by Using Quartiles

Step 1 Determine the first and third quartiles of the data set.
Step 2 Compute the interquartile range, IQR.
Step 3 Determine the **fences (or limits)**. Fences serve as the cutoff points for determining outliers.

$$\text{Lower fence} = Q_1 - 1.5(\text{IQR})$$
$$\text{Upper fence} = Q_3 + 1.5(\text{IQR})$$

Step 4 If a data value is less than the lower fence or greater than the upper fence, it is considered an outlier.

Note: You could have no outliers, or exactly one, or multiple outliers in any data set.

☺ **Exercises**:

1) Find the z-score for the value 104, when the mean is 92 and the standard deviation is 8.

2) A highly selective boarding school will only admit students who place at least 1.5 z-scores above the mean on a standardized test that has a mean of 110 and a standard deviation of 12. What is the minimum score that an applicant must make on the test to be accepted?

3) **Men versus Women.** The average 20– to 29–year-old man is 69.6 inches tall, with a standard deviation of 3.0 inches, while the average 20– to 29–year-old woman is 64.1 inches tall, with a standard deviation of 3.8 inches. Who is relatively taller, a 67-inch man or a 62-inch woman?

4) In 1995, Jerry Rice of the San Francisco 49ers had the most receiving yards of any wide receiver in the 90's era with 1848 receiving yards. In 2012, Calvin Johnson of the Detroit Lions had the most receiving yards of any wide receiver in the 2010's era with 1964 receiving yards. In the 90's, the mean and standard deviation for receiving yards was 288.4 and 342.8 respectively. In the 2010's, the mean and standard deviation for receiving yards was 269.2 and 329.6 respectively. (Note: The standard deviation is very high here since these are all players calculated throughout each decade that had at least one reception, thus the range is very large and skewed.) Who had a better performance in their decade?

SECTION 3.4 – Measures of Position and Outliers

☺ **Exercises**:

For exercises 5 through 8, find the first, second, and third quartiles of the following data sets.

5) 1, 1, 2, 3, 4, 6

1 1 2 3 4 6

Q_1 = _________ Q_2 = _________ Q_3 = _________

6) 2, 2, 3, 4, 5, 7, 9

2 2 3 4 5 7 9

Q_1 = _________ Q_2 = _________ Q_3 = _________

7) 3, 4, 4, 5, 6, 7, 9, 11

3 4 4 5 6 7 9 11

Q_1 = _________ Q_2 = _________ Q_3 = _________

8) 4, 4, 5, 7, 8, 9, 10, 14, 15

4 4 5 7 8 9 10 14 15

Q_1 = _________ Q_2 = _________ Q_3 = _________

9) The owner of a supermarket recorded the number of customers who came into his store each hour in a day. The results were 12, 8, 10, 6, 15, 3, 6, 7, 12, 8, and 9.

a) Find Q_1, Q_2, and Q_3.

b) Compute the interquartile range, IQR.

c) Determine the lower and upper fences. Are there any outliers?

SECTION 3.4 – Measures of Position and Outliers

☺ **Exercises**:

10) Here are the highest temperatures ever recorded (in °F) in 32 different U.S. states.

93 , 105 , 105 , 105 , 106 , 106 , 107 , 107 , 108 , 110 , 110 , 112 , 112 , 112 , 114 , 114
114 , 115 , 116 , 117 , 118 , 118 , 118 , 118 , 118 , 119 , 121 , 124 , 127 , 129 , 132 , 134

a) Find Q_1, Q_2, and Q_3.

b) Compute the interquartile range, IQR.

c) Determine the lower and upper fences. Are there any outliers?