8.3

You Are Too Far Away!

Calculating IQR and Identifying Outliers

LEARNING GOALS

In this lesson, you will:

- Calculate and interpret the interquartile range (IQR) of a data set.
- Determine if a data set contains outliers.

KEY TERMS

- interquartile range (IQR)
- outlier
- lower fence
- upper fence

Everywhere in our world there are boundaries that show where something begins and ends. The walls to your classroom are boundaries. The lanes on the road are boundaries. There are boundaries on sports fields and boundaries for each state and country. But what about the universe? Is there a boundary to show where the universe begins and ends?

That is a question that astronomers and physicists have debated for quite some time. For example, in the early 1900s, astronomer Harlow Shapely claimed that the entire universe was located within the Milky Way galaxy (the same galaxy where the Earth is located). He determined the galaxy was 300,000 light-years in diameter and in his opinion, could be thought of as the boundary of the universe. It was not until 1925, when Edwin Hubble showed that there are stars located much farther than 300,000 light years away. At this point, most scientists agreed that the universe must be larger than the Milky Way galaxy.

So we know the universe is larger than the Milky Way galaxy, but will we ever know just how large it is? Scientists have studied the idea that the universe is actually expanding for quite some time. What does this mean in terms of boundaries? Do you think we will ever know the size of the universe?

8

PROBLEM 1 Touchdown!

Coach Petersen's Middletown 9th grade football team is having a tough season. The team is struggling to win games. He is trying to determine why his team has only won a few times this year. The table shows the points scored in games in 2011 and 2012.

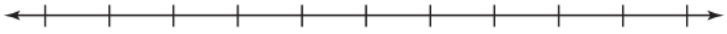
Points Scored (2011)	10	13	17	20	22	24	24	27	28	29	35
Points Scored (2012)	0	7	17	17	18	24	24	24	25	27	45

1. Analyze the data sets in the table.

   a. In which year do you think the football team performed better? Explain your reasoning.

b. Calculate the five number summary for each year.

c. Construct box-and-whisker plots of each year's scores using the same number line for each.



8

d. Evita states that because the medians are the same, both teams performed equally well. Is she correct? Explain your reasoning.

e. What conclusions can you draw about the points scored each year?

Another measure of data distribution Coach Petersen can use to compare the teams is the *interquartile range* or *IQR*. The **interquartile range, IQR,** measures how far the data is spread out from the median. The IQR gives a realistic representation of the data without being affected by very high or very low data values. The IQR often helps show consistency within a data set.The IQR is the range of the middle 50 percent of the data. It is calculated by subtracting Q3 − Q1.

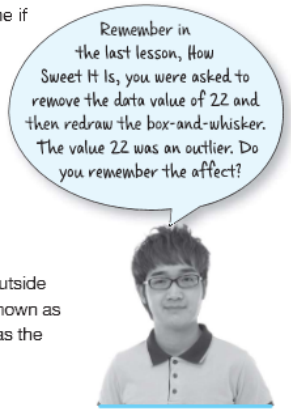
2. Calculate the IQR for the points scored each year. Then interpret the IQR for each year.

8

PROBLEM 2 Get Outta Here!

Another useful statistic when analyzing data is to determine if there are any *outliers*. An **outlier** is a data value that is significantly greater or lesser than other data values in a data set. It is important to identify outliers because outliers can often affect the other statistics of the data set such as the mean.

An outlier is typically calculated by multiplying the IQR by 1.5 and then determining if any data values are greater or lesser than that calculated distance away from Q1 or Q3. By calculating Q1 − (IQR · 1.5) and Q3 + (IQR · 1.5), you are determining a lower and upper limit for the data. Any value outside of these limits is an outlier. The value of Q1 − (IQR · 1.5) is known as the **lower fence** and the value of Q3 + (IQR · 1.5) is known as the **upper fence**.



Let's analyze the data set given to see how outliers can be represented on a box-and-whisker plot.

2, 5, 6, 6, 7, 9, 10, 11, 12, 12, 14, 28, 30

Minimum = 2, Q1 = 6, Median = 10, Q3 = 13, Maximum = 30

IQR = 7

Using the five number summary and IQR, calculate the upper and lower fence to determine if there are any outliers in the data set.

Lower Fence:
= Q1 − (IQR · 1.5)
= 6 − (7 · 1.5)
= −4.5

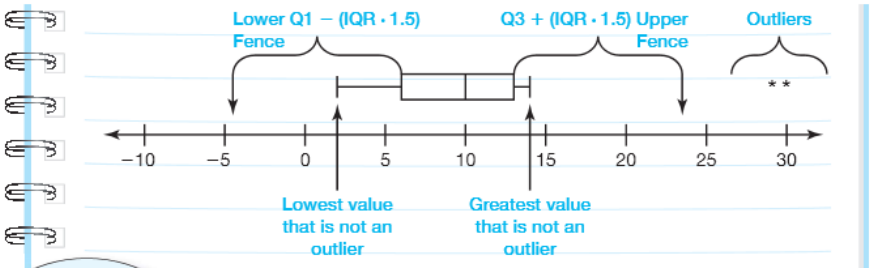
There are no values less than −4.5.

Upper Fence:
= Q3 + (IQR · 1.5)
= 13 + (7 · 1.5)
= 23.5

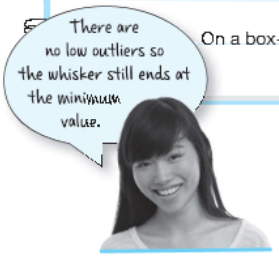
Both 28 and 30 are greater than 23.5.

If there are outliers, the whisker will end at the lowest or highest value that is not an outlier.

Since 28 and 30 are both outliers, 14 is the greatest data value that is not an outlier.



On a box-and-whisker plot, it is common to denote outliers with an asterisk.



8

Recall the data sets from Problem 1, *Touchdown!* The five number summary and IQR for each data set is shown.

2011:	2012:
Minimum = 10	Minimum = 0
Q1 = 17	Q1 = 17
Median = 24	Median = 24
Q3 = 28	Q3 = 25
Maximum = 45	Maximum = 45
IQR = 11	IQR = 8

1. Use the formulas to determine if there are any outliers in either data set.

a. Determine the upper and lower fence for each year's data set.

b. Identify any outliers in either set of data. Explain your reasoning.

8

2. Remove any outliers for the 2012 data set and, if necessary, reconstruct and label the box-and-whisker plot(s). Compare the IQR of the original data to your new calculations. What do you notice?

PROBLEM 3 Hurry Up!

Brenda needs to get the oil changed in her car, but she hates to wait! Quick Change and Speedy Oil are two garages near Brenda's house. She decides to check an online site that allows customers to comment on the service at different local businesses and record their wait times. Brenda chooses 12 customers at random for each garage. The wait times for each garage are shown.

Wait Times (minutes)

Quick Change				Speedy Oil			
10	60	22	15	5	60	45	24
12	24	20	18	40	26	55	30
16	23	22	15	32	85	45	30

1. Create a box-and-whisker plot of each data set.

2. Calculate and interpret the IQR for each data set.

3. Describe each data distribution and explain its meaning in terms of this problem situation.

4. Identify the measure of central tendency that best represents each data set. Explain your reasoning.

5. Identify any outliers in the data sets.

6. Remove any outliers in each data set and, if necessary, reconstruct the box-and-whisker plot. Compare the IQR of the original data to your new calculations. What do you notice?

7. Does your choice for the best measure of center from Question 4 still hold true?

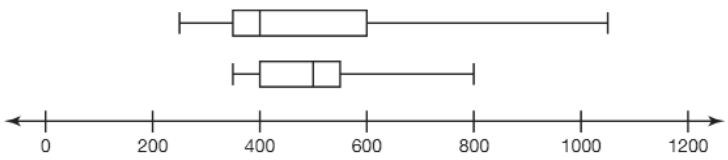
8. Based on the data gathered, which garage should Brenda choose if she is in a hurry?

Talk the Talk

8

1. Why is the IQR not affected by extremely high or low data values in a data set? Explain your reasoning.

2. Use the two box-and-whisker plots shown to answer each question.



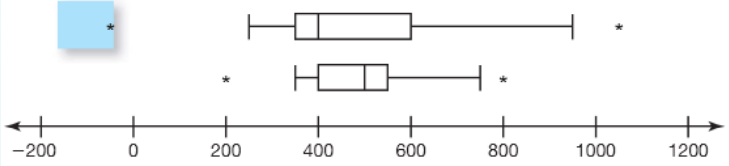
a. Estimate the five number summary for each box plot to the nearest 50.

b. Based on your estimates, calculate the IQR of both box-and-whisker plots.

8

c. Determine if there are any outliers in either data set shown in the box-and-whisker plots.

d. Lydia was told to assume that each data set has one outlier and that there are data values at the upper and lower fences. Lydia recreated the two box plots from Question 2 to represent the outliers. Her box plots are shown.



Are Lydia's box plots correct? Explain why or why not.

Be prepared to share your solutions and methods.