Guided Practice 6.3

Example 1

Each student in Mr. Lamb's class measured the length of his or her pencil. The following data set shows the pencil lengths in centimeters (cm). What is the expected length of any given pencil? Describe the shape, center, and spread of the data.

Student	Pencil length in cm	Student	Pencil length in cm
1	17.8	11	14.0
2	19.0	12	14.3
3	16.7	13	15.1
4	16.5	14	17.3
5	16.1	15	15.6
6	15.6	16	16.1
7	10.2	17	16.2
8	15.7	18	16.2
9	17.9	19	18.6
10	15.7	20	16.0

1. Order the data from least to greatest.

Student	Pencil length in cm	Student	Pencil length in cm
7	10.2	16	16.1
11	14.0	17	16.2
12	14.3	18	16.2
13	15.1	4	16.5
6	15.6	3	16.7
15	15.6	14	17.3
8	15.7	1	17.8
10	15.7	9	17.9
20	16.0	19	18.6
5	16.1	2	19.0

2. Calculate the interquartile range (IQR).

   To find the IQR, first find the median, first quartile (Q_1), and third quartile (Q_3).

   There are 20 data points in the set. First, find the median.

   The median is the average of the 10th and 11th data points.

   $$\frac{16.1+16.1}{2}=16.1$$

   Next find Q_1, the median of the lower half of the data set.

   Q_1 is the average of the fifth and sixth data values.

   $$\frac{15.6+15.6}{2}=15.6$$

   Next find Q_3, the median of the upper half of the data.

   Q_3 is the average of the 15th and 16th data values.

   $$\frac{16.7+17.3}{2}=17$$

   Now find the IQR, which is the difference between Q_3 and Q_1 ($Q_3 - Q_1$).

   $17 - 15.6 = 1.4$

   IQR = 1.4

3. Multiply 1.5 by the IQR.

   $1.5(\text{IQR}) = 1.5(1.4) = 2.1$

4. Determine if there are any outliers at the lower end of the data.

   Subtract 1.5(IQR) from Q_1.

   $Q_1 - 1.5(\text{IQR}) = 15.6 - 2.1 = 13.5$

   Any data values less than 13.5 are outliers. Examine the data. There is one value in the data set that is less than 13.5. Therefore, there is one outlier: 10.2.

5. Determine if there are any outliers at the upper end of the data.

   Add 1.5(IQR) to Q_3.

   $Q_3 + 1.5(\text{IQR}) = 17 + 2.1 = 19.1$

   Any data values greater than 19.1 are outliers. Examine the data. There are no values in the data set that are greater than 19.1. Therefore, there are no more outliers.

6. List all the outliers.

   All of the outliers are those values that are less than $Q_1 - 1.5(\text{IQR})$ and those values that are greater than $Q_3 + 1.5(\text{IQR})$.

   The only outlier for this data set is 10.2.

7. Calculate a measure of center.

   A measure of center will give us an approximate expected length of any pencil, based on Mr. Lamb's data.

   Because there is an outlier, the mean would be influenced by this value. Instead of using the mean, use the median as a measure of center, because it is not influenced by outliers.

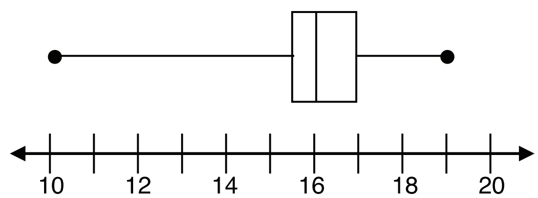
   We already determined that the median is 16.1.

   The expected length of any given pencil is 16.1 cm.

8. Describe the shape, center, and spread in the context of the problem.

   To describe the shape, plot the data using a box plot.

   Create the box plot using the five values found earlier: minimum = 10.2; Q_1 = 15.6; median = 16.1; Q_3 = 17; maximum = 19.0.



   The data is skewed left because the box plot has a long tail on the left side. This is due to the low value of the outlier, 10.2. The spread of the data is large. Notice the width of the box compared to the width of the minimum and maximum values. The center is better described by the median, which is not influenced by outliers. Therefore, the approximate length of a pencil based on Mr. Lamb's data is 16.1 cm, with a large variation in the data.

Example 2

Kayla is trying to estimate the cost of a house painter. She receives the following estimates, in dollars.

1288 1640 1547 1842 1553 1604 2858 1150 1844 1045 1347

She takes the mean of the data and states that the estimated cost of a house painter is $1,610.73. Is her estimate accurate?

1. Order the data from least to greatest.

   1045 1150 1288 1347 1547 1553 1604 1640 1842 1844 2858

2. Calculate the interquartile range (IQR).

   To find the IQR, first find the median, first quartile (Q_1), and third quartile (Q_3).

   There are 11 data points in the set. First, find the median.

   The median is the sixth data point: 1,553.

   Next find Q_1, the median of the lower half of the data set.

   Q_1 is the third data value: 1,288.

   Next find Q_3, the median of the upper half of the data.

   Q_3 is the ninth data value: 1,842.

   Now find the IQR, which is the difference between Q_3 and Q_1 ($Q_3 - Q_1$).

   $1842 - 1288 = 554$

3. Multiply 1.5 by the IQR.

   $1.5(\text{IQR}) = 1.5(554) = 831$

4. Determine if there are any outliers at the lower end of the data.

   Subtract 1.5(IQR) from Q_1.

   $Q_1 - 1.5(\text{IQR}) = 1288 - 831 = 457$

   Any data values less than 457 are outliers. Examine the data. There are no values in the data set that are less than 457.

5. Determine if there are any outliers at the upper end of the data.

   Add 1.5(IQR) to Q_3.

   $Q_3 + 1.5(\text{IQR}) = 1842 + 831 = 2673$

   Any data values greater than 2,673 are outliers. Examine the data. There is one value in the data set that is greater than 2,673. Therefore, there is one outlier: 2,858.

6. List all the outliers.

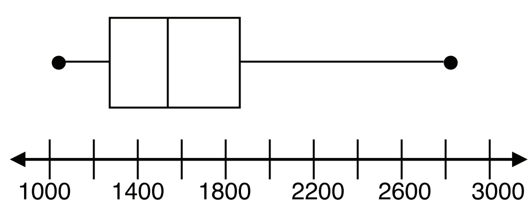
   All of the outliers are those values that are less than $Q_1 - 1.5(IQR)$ and those values that are greater than $Q_3 + 1.5(IQR)$.

   The only outlier for this data set is 2,858.

7. Create a box plot of the data set.

   Use the values we found for minimum, maximum, Q_1, median, and Q_3 to create a box plot.

   The minimum is the lowest value, 1,045; the maximum is the greatest value, 2,858; Q_1 is 1,288; the median is 1,553; and Q_3 is 1,842.



8. Describe how the outlier has influenced the center of the data.

   Compare the calculated mean to the median.

   The outlier is larger than the rest of the data set, and increased the mean. This can also be seen when comparing the mean and median; Kayla's mean is greater than the median.

9. Describe how the outlier has influenced the shape of the data.

   Identify if the data is skewed to the left or right.

   Data represented in the box plot that is skewed to the left has a longer tail on the left side of the box plot.

   Similarly, data that is skewed to the right has a longer tail on the right side of the box plot.

   The data is skewed to the right because of the influence of the large outlier.

10. Describe how the outlier has influenced the spread of the data.

The overall range of the data has been increased by the outlier. Without the outlier, the maximum would be 1,844, making the range (the difference between the maximum and minimum) 799:

$$1844 - 1045 = 799$$

With the outlier, the range is 1,813:

$$2858 - 1045 = 1813$$

11. Determine a reasonable estimate for the cost of a house painter.

Either the median or the mean without the outlier could be used as an estimate. The mean without the outlier is the mean of the remaining 10 data points:

$$\frac{1045+1150+1288+1347+1547+1553+1604+1640+1842+1844}{10}=1486$$

Kayla's estimate is not correct. One of the values is an outlier, which increased her estimate. A more accurate estimate is the mean without the outlier ($1,486) or the median ($1,553).

Example 3

The National Basketball Association has strict regulations about the dimensions of basketballs used during games. The circumference of the basketball must be between 749 mm and 762 mm. A basketball manufacturer is sending a shipment of various brands of basketballs to the NBA. Brandon, who works in the shipping department, notices that the mean circumference of the basketballs is in this range, so he decides to send all the basketballs to the NBA. The NBA gets the shipment, and is not happy. Many of the basketballs in the shipment do not have the right circumference. What could have happened? How could the mean circumference be in this range, but many of the individual basketballs have the wrong circumference?

1. Consider how the mean is calculated. Is there any way the mean could not accurately represent the rest of the data?

   The mean is an average of all the data, but the mean doesn't address the variability. The maximum and minimum sizes of the basketballs must lie within the given range. In addition, if there are very large or very small data values, the mean could be greater than or less than many of the data points.

2. Describe how the mean could misrepresent the data, and what Brandon could do to ensure he doesn't send the wrong size basketballs.

   Brandon should use more measures of center and spread to understand the circumferences of the basketballs being delivered. More specifically, he should focus on variability, and make sure the circumferences lie within the necessary range.