Numerical Summaries of Data Summary

KEY TERMS

- measure of center
- mode
- median
- balance point
- mean
- measures of variation
- range
- quartile
- interquartile range (IQR)
- box-and-whisker plot
- deviation
- absolute deviation
- mean absolute deviation

LESSON 1 In the Middle

When you analyze a set of data, you often want to describe it numerically. One way to numerically describe a data set is to use a measure of center. A **measure of center** tells you how the data values are clustered, or where the center of a graph of the data is located. There are three measures that describe how a data set is centered: the mean, the median, and the mode.

The **mode** is the data value or values that occur most frequently in a data set. A data set can have more than one mode or no mode. For example, the mode of the data set 12, 6, 12, 26, 4, and 12 is 12.

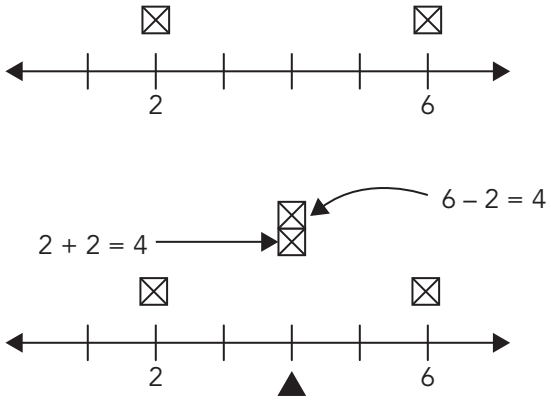
The **median** is the middle number in a data set when the values are placed in order from least to greatest or greatest to least. When a data set has an odd number of data values, you can determine which number is exactly in the middle of the data set. If there is an even number of data values, then the median is calculated by adding the two middle numbers and dividing by 2.

For example, the median of the data set 15, 12, 13, 10, 8, and 14 is 12.5.

The third measure of center is based on leveling off or creating fair shares. For example, if you had a stack of two cubes and a stack of six cubes, you can rearrange the stacks to create two equal stacks of four cubes each.

You can also represent quantities on a number line and create a balance point. When you have all the points at the same value, the number line is balanced. The value where the number line is balanced is called the **balance point**.

For example, consider the data set 2, 6.



The value 2 was moved to the right from 2 to 4. To maintain balance, 6 was moved to the left from 6 to 4. The balance point is 4. The balance point can also be called the mean. The **mean** is the arithmetic average of the numbers in a data set.

For example, determine the mean of the data set: 12, 12, 6, 26, 4, 12.

The mean is calculated by adding all the values in the data set and dividing the sum by the number of values.

Step 1: $12 + 12 + 6 + 26 + 4 + 12 = 72$

Step 2: $\frac{72}{6} = 12$

You can verify that the mean is 12 because the balance point of the data set is 12.

LESSON 2

Box It Up

A **measure of variation** describes the spread of data values. One measure of variation is the range. The **range** is the difference between the maximum and minimum values of a data set.

For example, the range of the data 200, 150, 260, 180, 300, 240, and 280 is $300 - 150 = 150$.

Another set of values that helps to describe variation in a data set is a **quartile**. When data in a set are arranged in order, quartiles are the numbers that split data into quarters (or fourths). Quartiles are often denoted by the letter Q followed by a number that indicates which fourth it represents. Since the median is the second quartile, it could be denoted Q2. The other quartiles are Q1 and Q3. The **interquartile range**, abbreviated IQR, is the difference between the third quartile, Q3, and the first quartile, Q1. The IQR indicates the range of the middle 50 percent of the data.

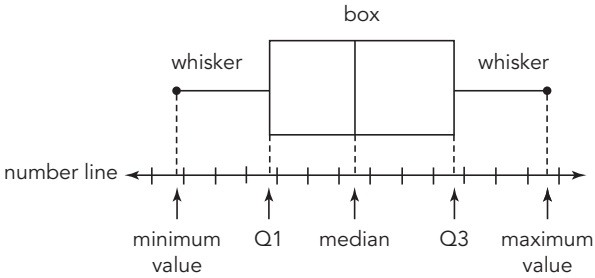
To summarize and describe the spread of the data values, you can use the five-number summary. The five-number summary includes these 5 values from a data set:

- Minimum: the least value in the data set
- Q1: the first quartile
- Median: the median of the data set
- Q3: the third quartile
- Maximum: the greatest value in a data set.

For the data set 24, 32, 16, 18, 30, and 20, the minimum is 16, Q1 is 18, the median is 22, Q3 is 30, and the maximum is 32.

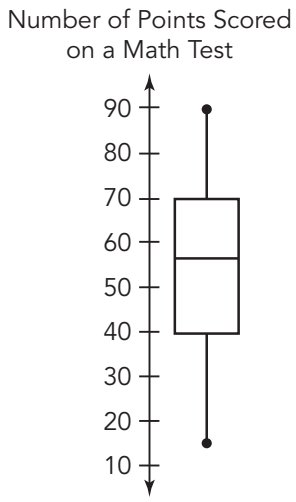
A **box-and-whisker plot**, or just box plot, is a graph that displays the five-number summary of a data set.

Parts of a Box-and-Whisker Plot



Box-and-whisker plots can be represented vertically as well as horizontally.

For example, in this box-and-whisker plot, the minimum of the data set is 15, Q1 is 40, the median of the data set is 56, Q3 is 70, and the maximum of the data set is 90.



LESSON 3

March MADness

Another measure of variation that describes the spread of data values is **deviation**. The deviation of a data value indicates how far that data value is from the mean. To calculate the deviation, subtract the mean from the data value:

$$\text{Deviation} = \text{data value} - \text{mean}$$

For example, the mean of the data set 15, 12, 13, 10, 9, and 13 is 12.

The table describes each data point's deviation from the mean.

Data Point	15	12	13	10	9	13
Deviation from the Mean	3	0	1	−2	−3	1

In order to get an idea of the spread of the data values, you can take the absolute value of each deviation and then determine the mean of those absolute values. The absolute value of each deviation is called the **absolute deviation**. The **mean absolute deviation** (MAD) is the mean of the absolute deviations.

For example, the mean absolute deviation of the data shown in the table is

$$\frac{|3| + |0| + |1| + |-2| + |-3| + |1|}{6} = \frac{10}{6}.$$

So, the MAD is about 1.67.

LESSON 4

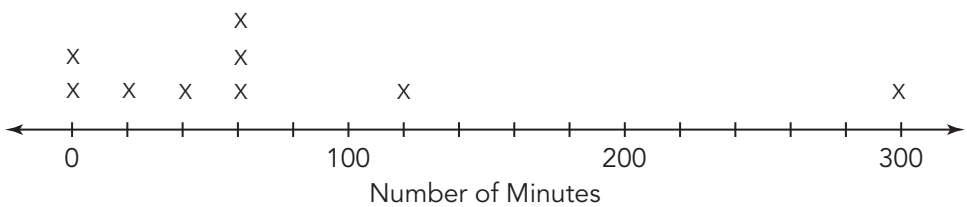
You Chose... Wisely

The distribution of data can affect the measures of center.

The median is not affected by very large or very small data values, but the mean is affected by these large and small values. Therefore, the median is the best measure of center when the data is skewed left or right.

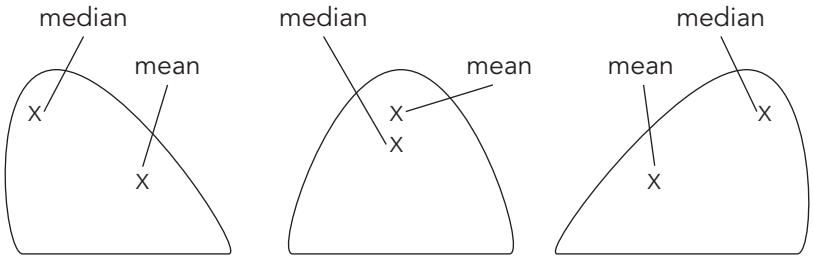
For example, the dot plot shows the amount of time Ben's friends spend exercising on weekdays.

Time Spent Exercising Each Weekday



The data is skewed right, so the mean is greater than the median. The median for the data set is 60 minutes and the mean is 73.33 minutes. The median is a better measure to describe a typical value in the data.

The measure of central tendency and measure of variation used to best describe a data set depends on the values in the data set and the spread of those values. If you use the median to describe the measure of center, you should use the IQR to describe the measure of variation, and if you use the mean to describe the measure of center, you should use the mean absolute deviation to describe the measure of variation.



skewed right

The mean of a data set is greater than the median when the data is skewed to the right.

symmetric

The mean and median are equal when the data is symmetric.

skewed left

The mean of a data set is less than the median when the data is skewed to the left.