The Effect of Extremes

Let's see how statistics change with the data.

10.1 At Least I Did Well

Several online video games match players with other players at random to compete in a team game.

1. What information could you use to determine the top players in a team game like this? Explain your reasoning.

2. There are two games of a similar type. One person claims that the best players play game A. Another person claims that game B has better players. How could you display data to help inform their discussion? Explain your reasoning.

10.2 Separated by Skew

1. Use technology to create a dot plot that represents the distribution of the data, then describe the shape of the distribution.

   6 7 8 8 9 9 9 10 10 10 10 11 11 11 12 12 13 14

2. Find the mean and median of the data.

3. Find the mean and median of the data with 2 additional values included as described.

   a. Add 2 values to the original data set that are greater than 14.

   b. Add 2 values to the original data set that are less than 6.

   c. Add 1 value that is greater than 14 and 1 value that is less than 6 to the original data set.

   d. Add the two values, 50 and 100, to the original data set.

4. Share your work with your group. What do you notice is happening with the mean and median based on the additional values?

5. Change the values so that the distribution fits the description given to you by your teacher, then find the mean and median.

6. Find another group that created a distribution with a different description. Explain your work and listen to their explanation, then compare your measures of center.

10.3 Plots Matching Measures

Create a possible dot plot with at least 10 values for each of the conditions listed. Each dot plot must have at least 3 values that are different.

1. a distribution that has both mean and median of 10

2. a distribution that has both mean and median of -15

3. a distribution that has a median of 2.5 and a mean greater than the median

4. a distribution that has a median of 5 and a median greater than the mean

Are you ready for more?

The mean and the median are by far the most common measures of center for numerical data. There are other measures of center, though, that are sometimes used. For each measure of center, list some possible advantages and disadvantages. Be sure to consider how it is affected by extremes.

1. *Interquartile mean*: The mean of only those points between the first quartile and the third quartile.

2. *Midhinge*: The mean of the first quartile and the third quartile.

3. *Midrange*: The mean of the minimum and maximum value.

4. *Trimean:* The mean of the first quartile, the median, the median again, and the third quartile. So we are averaging four numbers because the median is counted twice.

Lesson 10 Summary

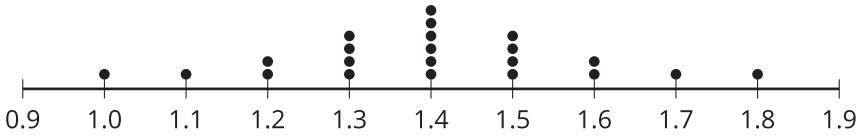
Is it better to use the mean or median to describe the center of a data set?

The mean gives equal importance to each value when finding the center. The mean usually represents the typical values well when the data have a symmetric distribution. On the other hand, the mean can be greatly affected by changes to even a single value.

The median tells you the middle value in the data set, so changes to a single value usually do not affect the median much. So, the median is more appropriate for data that are not very symmetrically distributed.

We can look at the distribution of a data set and draw conclusions about the mean and the median.

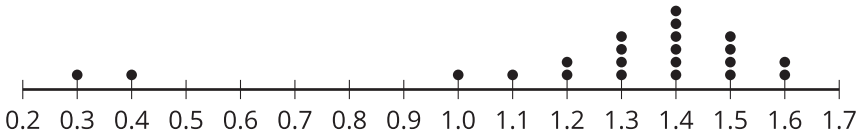
Here is a dot plot showing the amount of time a dart takes to hit a target in seconds. The data produce a symmetric distribution.



time to hit dartboard in seconds

When a distribution is symmetric, the median and mean are both found in the middle of the distribution. Since the median is the middle value (or the mean of the two middle values) of a data set, you can use the symmetry around the center of a symmetric distribution to find it easily. For the mean, you need to know that the sum of the distances away from the mean of the values greater than the mean is equal to the sum of the distances away from the mean of the values less than the mean. Using the symmetry of the symmetric distribution you can see that there are four values 0.1 second above the mean, two values 0.2 seconds above the mean, one value 0.3 seconds above the mean, and one value 0.4 seconds above the mean. Likewise, you can see that there are the same number of values the same distances below the mean.

Here is a dot plot using the same data, but with two of the values changed, resulting in a skewed distribution.



time to hit dartboard in seconds

When you have a skewed distribution, the distribution is not symmetric, so you are not able to use the symmetry to find the median and the mean. The median is still 1.4 seconds since it is still the middle value. The mean, on the other hand, is now about 1.273 seconds. The mean is less than the median because the lower values (0.3 and 0.4) result in a smaller value for the mean.

The median is usually more resistant to extreme values than is the mean. For this reason, the median is the preferred measure of center when a distribution is skewed or if there are extreme values. When using the median, you would also use the IQR as the preferred measure of variability. In a more symmetric distribution, the mean is the preferred measure of center, and the MAD is the preferred measure of variability.