Measures of Center and Shape

Inquire: Isn’t Average Enough?

Overview

Every math class teaches measures of central tendency: mean, median, and mode. The most common of these is mean (arithmetic average). However, there are important uses for the other two also. When working with data that has outliers (values an abnormal distance from the other values) the average isn’t always the best choice. If one value is much higher or lower, it will skew the average. The median value might be best in this case. The median is the middle of the data sample. The mode is the only one that will tell you which value occurs the most frequently. When working with data, calculate all three to see which one best represents the truth of the data. There is a reason for each one. Many times the average is enough, but not always.

Big Question: When working with data, why is it important to look at more than one measure of central tendency?

Watch: Central Tendencies in Sports

No matter what sport is your favorite, as you follow the game, you will see statistics at every turn. Whether you are watching professional bull riders or a pre-school track meet, someone has taken the time to figure the averages, percentages, probabilities, and odds of every imaginable outcome. The most common statistics you will see are averages. While the average may not always be the best indicator for a data distribution, it will always be the most common.

The average, also known as the arithmetic mean, is found by adding the values in a data sample and dividing by the number of values. If you have a golfer’s scores from his last five rounds, you can add them together and divide by five to find his average score for those specific rounds. Perhaps you follow football. There are a multitude of Internet sites that will tell you each player’s average yards, sacks, completions, or fumbles per game for this year or any year.

The median isn’t as popular in sports reporting, but it is still there. The median is the middle value in a data sample. If there is an extremely high or low value in a collection of values — say, one person with a high point total in a basketball team full of poor shots— the median is often a more accurate way to measure the central tendency of the group of numbers. It will not give too much attention to that one high point value, but will look at the team overall.

The mode is more of an indicator than an accurate measurement. However, it can be helpful in determining a pattern. The mode is the value with the highest frequency in a data sample. If we want to
look at how many times players had possession of the ball, the mode would quickly tell us which player or
players had the most touches. If that tendency repeats itself in future games, the mode can be a tool for
prediction.

There is always a need for understanding measures of central tendency, because almost everything is
measured and reported, especially in the sports world.

Read: Measures of the Center and Shape of Data

Overview
The center of a data set describes location. The two most widely used measures of the center of data are
the mean (average) and the median. Another measure of the center is the mode. The shape of a data
set can be described as symmetrical or skewed to the right or left. The relationships of the mean, median,
and mode can identify the shape of the data.

Mean
You may be familiar with averaging numbers. You may have averaged grades, hours worked per week, or
money spent on lunch. It’s a simple process of adding and dividing. Add together the amount you spent
on lunch Monday through Friday, then divide by the number of days you bought lunch. This gives you the
average you spent on lunches for the week. In math class this is called the mean. The mean is the most
common measure of the center.

Statisticians and mathematicians find subtle differences in the words “mean,” “average,” and “arithmetic
mean,” but for this lesson we will use them interchangeably. The mean for a data set can be calculated by
adding each value together and dividing the sum by the total number of values.

\[
\text{Mean} = \frac{\text{Sum of all data values}}{\text{number of values}}
\]

To calculate the mean weight of 50 people, add the 50 weights together and divide by 50.

Median
The median is a number that measures the center of the data. You can think of the median as the middle
value, but it does not need to be one of the numbers in the data. It is a number that separates ordered
data into halves. Half the values are smaller than the median, and half the values are larger. The key is
ordered data. This means the data has been arranged in order, usually from the least to the greatest.

To locate the median in a data set, let the letter \( n \) represent the total number of data values in the sample.
If \( n \) is an odd number, the median is the middle value of the ordered data (ordered least to greatest). For
example, a data set with five values has a middle value with two values above it and two values below it.
If the data set is \{3, 5, 7, 9, 11\}, the number 7 is the median.

If \( n \) is an even number, the median is equal to the two middle values added together and divided by two.
Consider the data set \{3, 5, 7, 9, 11, 13\}. There is an even number of values and no one value in the
middle. Take the two middle values (7 and 9), add them together (16), and divide by two (8). The median
is 8. The median does not have to be one of the numbers in the data sample.
Another way to find the location of the median is the formula: \((n+1) \div 2\). Notice that this formula does not tell you the value of the median, but the location of the median. In the first data set above, there were five numbers: \(n=5\). \((5+1)\div 2 = 3\). The third term in the ordered data is the median.

In the second example there were 6 terms in the data set: \(n=6\). \((6+1)\div 2 = 3.5\). The location of the median is at value 3.5 — between the third and fourth terms. They must be averaged by adding them together and dividing by two.

To find the median weight of the 50 people, first order the data. Find the number that splits the data into two equal parts, meaning an equal number of observations on each side. Since 50 is an even number, find the two weights in the center and average. The weights of 25 people are below this weight, and 25 people are heavier than this weight.

The median is generally a better measure of the center when there are extreme values or outliers because it is not affected by the precise numerical values of the outliers.

Mode

Another measure of the center is the mode. The mode is the most frequent value. There can be more than one mode in a data set as long as those values appear the same number of times. A data set with two modes is called bimodal.

Find the mode in the following data set:

50, 53, 59, 59, 63, 63, 72, 72, 72, 72, 72, 76, 78, 81, 83, 84, 84, 84, 84, 90, 93

The most frequent score is 72, which occurs five times. Mode = 72

Symmetrical Shape of Data

The following histogram displays a symmetrical distribution of data. A distribution is symmetrical if a vertical line can be drawn at some point in the histogram such that the shape to the left and the right of the vertical line are mirror images of each other.

The mean, median, and mode are each 7 for this data. In a perfectly symmetrical distribution, the mean and the median are the same. This example has one mode, and the mode is the same as the mean and median.
Skewness

Skewness is asymmetry (not symmetric) in a statistical distribution. A graph appears distorted or skewed either to the left or to the right when it is not symmetrical. The following histogram is skewed to the right.

![Histogram](image)

The mean is 6.3, the median is 6.5, and the mode is 7. Notice that the mean is less than the median, and they are both less than the mode. The mean and the median both reflect the skewing, but the mean reflects it more so.

Reflect Poll: Learned Anything New?

After studying this lesson, do you feel like you learned anything new about measures of central tendency?

- Nah, I knew it already.
- It was mostly review, but it helped.
- Yes, I never really understood it before.

Expand: Measures of Central Tendency

Overview

The most common measures of central tendency are the arithmetic mean, the median, and the mode. How do the various measures of central tendency compare with each other? For symmetric distributions, the mean, median, and mode are equal, except when the distribution is bimodal. Differences among the measures occur with skewed distributions. It is important to not only know the differences in these tendencies, but also how to calculate them.

Consider the following data. Person A is concerned about how much money he is spending on coffee on his way to work. He buys either a tall, grande, or venti five mornings a week. For the last three weeks, he has spent the following:

1.85, 1.85, 2.45, 2.10, 1.85, 2.10, 2.10, 1.85, 1.85, 2.10, 2.10, 1.85, 1.85, 1.85, 2.45

Before any calculations can be done, the list should be ordered from least to greatest.

1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 1.85, 2.10, 2.10, 2.10, 2.10, 2.10, 2.45, 2.45
The mean is the total of all values divided by the number of values: $30.20/15 = $2.01

The median is the middle number in the list of ordered values: $1.85

The mode is the number found most frequently in the data: $1.85

A graph of this data would be slightly skewed. The median and the mode are equal, but the mean is slightly larger. As was seen in earlier examples, the mean will show skewness more than the median and the mode.

The following example will help to explain the need for looking at all measures of central tendency, not just one.

Suppose that in a neighborhood of 50 people, one person earns $5,000,000 per year and the other 49 each earn $30,000. Which is the better measure of the center — the mean or the median?

There are 50 terms in our data. $30,000 x 49 and $5,000,000 x 1. The median is found by averaging the two middle terms, $30,000 and $30,000. The median income is $30,000.

To find the mean, add all the values and divide by 50. $30,000 x 49 = $1,470,000.

$$\frac{1,470,000 + 5,000,000}{50} = \frac{6,470,000}{50} = $1,570,000$$

The mean income is $1,570,000

The median is a better measure of the center than the mean because 49 of the values are $30,000 and one is $5,000,000. The $5,000,000 is an outlier. The $30,000 gives us a better sense of the middle of the data.

Lesson Toolbox

Additional Resources and Readings

A video covering this lesson’s topics in an interesting and fast-paced way

- Link to resource: https://youtu.be/kn83BA7cRNM

A video giving information about how and why central tendencies are useful

- Link to resource: https://youtu.be/AdH5vfobH5E

A video offering a very good description, including graphs, of skewness

- Link to resource: https://youtu.be/XSSRvVMOqIQ

Lesson Glossary

bimodal: when there are two values in a data set that appear the same number of times and share the highest frequency

ordered data: when data has been arranged in a particular order, usually from the least to the greatest

mean: the same as the arithmetic average found by adding all of the values and dividing by the number of values
**median**: a value in the center of a data set; there are as many data points below the median as above it

**mode**: the value that occurs most frequently in a data set

**skewness**: asymmetry in a statistical distribution, in which the graph appears distorted or skewed either to the left or to the right

**symmetric distributions**: types of distribution where the left side of the distribution mirrors the right side

---

### Check Your Knowledge

1. Which measure of central tendency is the value that occurs most frequently?
   a. mean
   b. mode
   c. median
   d. measurement

2. The mean, median, and mode will be the same when the distribution is...
   a. skewed to the right.
   b. skewed to the left.
   c. asymmetrical.
   d. symmetrical.

3. What is the mode in the following data sample? \{4, 6, 6, 5, 4, 7, 8, 5, 6, 9, 8, 6\}
   a. 5
   b. 6
   c. 8
   d. 7

---

**Answer Key:**

---

### Citations

**Lesson Content:**
Authored and curated by Kathryn R. Price, M.Ed., NBCT for The TEL Library. CC BY NC SA 4.0

**Adapted Content:**
License CC BY 4.0. Link to resource: https://cnx.org/contents/tWu56V64@35.1:Obtk_9tJ@21/2-6-Skewness-and-the-Mean-Median-and-Mode

**Attributions**
“Figure 2.11” By OpenStax is licensed under CC BY 4.0. Link to resource: https://cnx.org/contents/tWu56V64@35.1:Obtk_9tJ@21/2-6-Skewness-and-the-Mean-Median-and-Mode

“Figure 2.12” By OpenStax is licensed under CC BY 4.0. Link to resource: https://cnx.org/contents/tWu56V64@35.1:Obtk_9tJ@21/2-6-Skewness-and-the-Mean-Median-and-Mode