Measures of central tendency: The mean, the median, and the mode

Recently, you learned that **measures of central tendency** are values that represent the center of a dataset. When you're working with a new dataset, identifying the central location of your data helps you quickly understand its basic structure.

In this reading, you'll learn more about three measures of central tendency: the mean, the median, and the mode. We'll go over how to calculate each measure, and discuss which measure is best to use based on your specific data.

Measures of central tendency

The mean, median, and mode all describe the center of a dataset in different ways:

- The **mean** is the average value in a dataset.

- The **median** is the middle value in a dataset.

- The **mode** is the most frequently occurring value in a dataset.

Let's explore how to calculate each measure of central tendency.

Calculate the mean, the median, and the mode

Mean

The **mean** is the average value in a dataset. To calculate the mean, you add up all the values in your dataset and divide by the total number of values.

For example, say you have the following set of values: 10, 5, 3, 50, 12. To find the mean, you add all the values for a total of 80. Then, you divide by 5, the total number of values.

$(10 + 5 + 3 + 50 + 12) \div 5 = 80 \div 5 = 16$

The mean, or average value, is 16.

Median

The **median** is the middle value in a dataset. This means half the values in the dataset are larger than the median, and half the values are smaller than the median.

You can find the median by arranging all the values in a dataset from smallest to largest. If you arrange your five values in this way you get: 3, 5, 10, 12, 50. The median, or middle value, is 10.

If there are an even number of values in your dataset, the median is the average of the two middle values. Let's say you add another value, 8, to your set: 3, 5, 8, 10, 12, 50. Now, the two middle values are 8 and 10. To get the median, take their average.

$(8 + 10) \div 2 = 18 \div 2 = 9$

The median is 9.

Mode

The **mode** is the most frequently-occurring value in a dataset. A dataset can have no mode, one mode, or more than one mode.

For example, the set of numbers 1, 12, 33, 54, 75 has no mode because no value repeats. In the set 2, 7, 7, 11, 20 the mode is 7, because 7 is the only value that occurs more than once. The set 3, 12, 12, 40, 40 has two modes: 12 and 40.

When to use the mean, the median, and the mode

Whether you use the mean, median, or mode to describe the center of your dataset depends on the specific data you're working with and what insights you want to gain from your data. Let's discuss some general guidelines for using each measure of central tendency.

Mean versus median

Both the mean and the median describe the central location of a dataset. However, as measures of central tendency, the mean and the median work better for different kinds of data.

The mean has one main disadvantage: it is very sensitive to outliers in your dataset. Recall that an outlier is a value that differs greatly from the rest of the data.

If there are outliers in your dataset, the median is usually a better measure of the center. If there are no outliers, the mean usually works well.

For example, imagine you want to compute the average annual salary for an employee at a small startup company. You have the following salary data:

Employee	#1	#2	#3	#4	#5	#6	#7
Salary	$40,000	$45,000	$45,000	$45,000	$45,000	$50,000	$500,000

You can calculate the mean annual salary by adding up all the values in your dataset and dividing by the total number of values. There are seven salaries in total, and their sum is $770,000.

$770,000 \div 7 = 110,000$

The mean salary for these seven employees is $110,000. However, the data suggests that this mean value does not accurately reflect the typical salary of an employee at this company. Most employees have a salary between $40,000–$50,000. In fact, only one employee has a salary greater than $50,000. The salary of $500,000 is an outlier that pulls up the average, or skews the mean.

In this case, due to the presence of this outlier value, the median is a better measure of central tendency than the mean. The median, or middle value, in this dataset is $45,000. The median gives you a better idea of the typical salary for an employee at this company.

Mode

The mode is useful when working with categorical data because it clearly shows you which category occurs most frequently. Say a company conducts an employee satisfaction survey. The main item on the survey states, "I am satisfied with the opportunity I have to grow within the company." Employees choose among four categories for their response: strongly agree, agree, disagree, strongly disagree. A bar chart summarizes the results.

Employee Survey Results
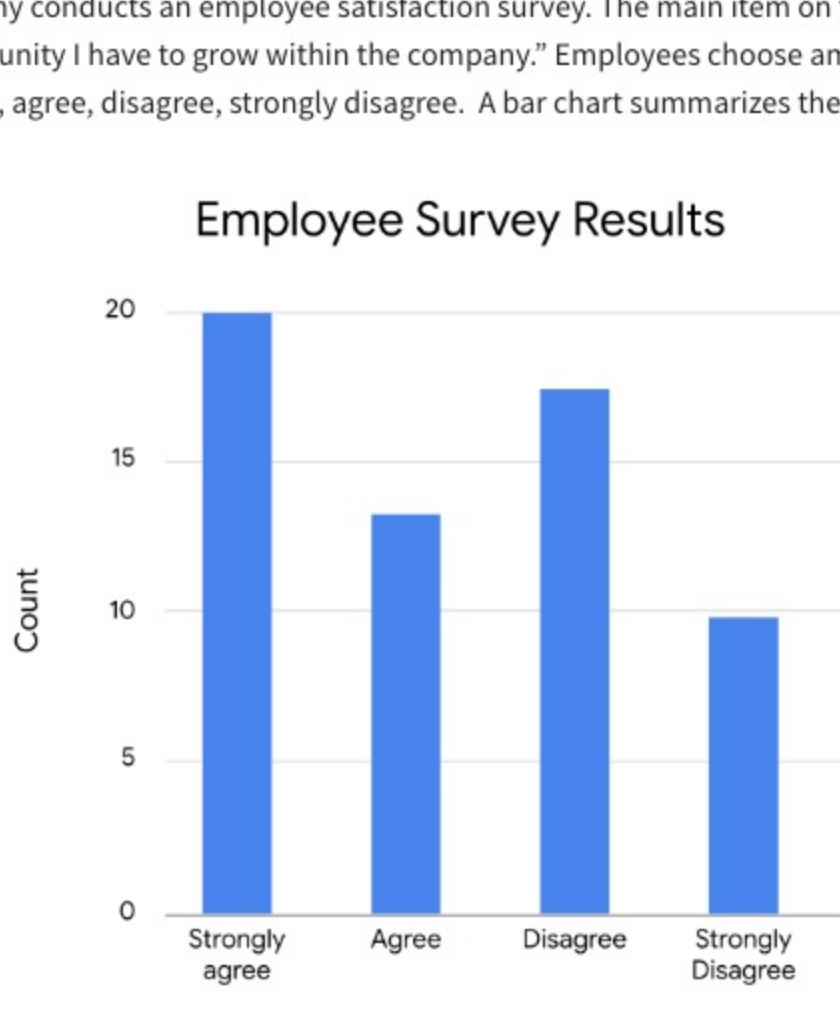
The mode represents the highest bar in the bar chart, which refers to the rating "strongly agree." This is the most frequently-occurring rating in the dataset. The mode gives the company clear feedback on employee satisfaction; in this case, positive feedback.

Key takeaways

Measures of central tendency such as the mean, median, and mode let you describe the center of your dataset using a single value. As a data professional, knowing the center of your dataset helps you quickly understand its basic structure and determine the next steps in your analysis.

Resources for more information

To learn more about measures of central tendency like the mean, median, and mode, explore the following resource:

- This article from the Australian Bureau of Statistics 🗗 offers a useful overview of the mean, median, and mode, and discusses how outliers influence measures of central tendency.

Mark as completed

👍 Like　　👎 Dislike　　⚑ Report an issue