

8.4: Measures of Central Tendency

Learning Objectives

- Find the mean of a data set
- Find the median of a data set
- Find the mode of a data set

It is often desirable to use a few numbers to summarize a distribution. One important aspect of a distribution is where its center is located. Numbers that describe a distribution's center are called measures of central tendency.

Let's begin by trying to find the most "typical" value of a data set.

Mean

Note that we just used the word "typical" although in many cases you might think of using the word "average." We need to be careful with the word "average" as it means different things to different people in different contexts. One of the most common uses of the word "average" is what mathematicians and statisticians call the **arithmetic mean**, or just plain old **mean** for short. "Arithmetic mean" sounds rather fancy, but you have likely calculated a mean many times without realizing it; the mean is what most people think of when they use the word "average".

Mean

The **mean** of a set of data is the sum of the data values divided by the number of values.

We will use n to represent the number of values in a data set.

✓ Example 8.4.1

Marci's exam scores for her last math class were: 79, 86, 82, 94. Find the mean of her exam scores.

Solution

$$\frac{79 + 86 + 82 + 94}{4} = \frac{341}{4} = 85.25.$$

Typically we round the mean to one more decimal place than the original data had. In this case, we would round 85.25 to 85.3. Marci's mean exam score was 85.3.

✓ Example 8.4.2

The number of touchdown (TD) passes thrown by each of the 32 teams in the National Football League (NFL) in the 2021 season are shown below [1].

40 43 38 36 37 39 36 27 41 27 24 20 26 22 21 12

30 34 21 29 21 21 21 20 23 20 23 23 16 20 14 15

Find the mean number of touchdown passes thrown in the NFL in the 2021 season.

Solution

Adding these values, we get 840 total TDs. Dividing by $n = 32$, the number of data values, we get $\frac{840}{32} = 26.25$. It would be appropriate to round this to 26.3.

It would be most correct for us to report that "The mean number of touchdown passes thrown per team in the NFL in the 2021 season was 26.3 passes," but it is not uncommon to see the more casual word "average" used in place of "mean."

Try It 8.4.1

The price of a jar of peanut butter at 5 stores was: \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99. Find the mean price.

Answer

$$\frac{\$3.29 + \$3.59 + \$3.79 + \$3.75 + \$3.99}{5} = \frac{\$18.41}{5} = \$3.682$$

Let's look at an example for calculating the mean given a frequency table.

✓ Example 8.4.3

The 100 families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars. The results are summarized in a frequency table below.

Income (thousands of dollars)	Frequency
15	6
20	8
25	11
30	17
35	19
40	20
45	12
50	7

Find the mean annual household income.

Solution

Calculating the mean by hand could get tricky if we try to type in all 100 values:

$$\frac{15 + 15 + 15 + 15 + 15 + 15 + 20 + 20 + \dots + 50 + 50 + 50 + 50 + 50 + 50 + 50}{100}$$

We could calculate this more easily by noticing that adding 15 to itself 6 times is the same as $15 \cdot 6 = 90$. Using this simplification, we get

$$\frac{15 \cdot 6 + 20 \cdot 8 + 25 \cdot 11 + 30 \cdot 17 + 35 \cdot 19 + 40 \cdot 20 + 45 \cdot 12 + 50 \cdot 7}{100} = \frac{3390}{100} = 33.9$$

The mean household income of our sample is 33.9 thousand dollars (\$33,900).

✓ Example 8.4.4

Extending off the last example, suppose a new family moves into the neighborhood example that has a household income of \$5 million (\$5000 thousand). What is the new mean household income?

Solution

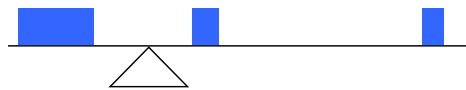
Adding this to our sample, our mean is now:

$$\frac{15 \cdot 6 + 20 \cdot 8 + 25 \cdot 11 + 30 \cdot 17 + 35 \cdot 19 + 40 \cdot 20 + 45 \cdot 12 + 50 \cdot 7 + 5000 \cdot 1}{101} = \frac{8390}{101} = 83.069$$

While 83.1 thousand dollars (\$83,069) is the correct mean household income, it no longer represents a "typical" value.

Median

Imagine the data values on a see-saw or balance scale. The mean is the value that keeps the data in balance, like in the picture below.



If we graph our household data, the \$5 million data value is so far out to the right that the mean has to adjust up to keep things in balance



For this reason, when working with data that have **outliers** – values far outside the primary grouping – it is common to use a different measure of center, the **median**.

Median

The **median** of a set of data is the value in the middle when the data is in order.

To find the median:

- List the data in order from smallest to largest, or largest to smallest
- If the number of data values, n , is odd, then the median is the middle data value. This is the data value in the $\frac{n+1}{2}$ position.
- If the number of data values is even, there is no one middle value, so we find the mean of the 2 middle values (values in the positions of $\frac{n}{2}$ and $\frac{n}{2} + 1$).

We can interpret the median as “half of the data is less than or equal to the median and the other half is more than or equal to the median.” Of course, we can rewrite this in context of the problem.

✓ Example 8.4.5

Find the median of these quiz scores: 5 10 8 6 4 8 2 5 7 7 9

Solution

We start by listing the data in order: 2 4 5 5 6 7 7 8 8 9 10

Since there are $n = 11$ data values, an odd number, the middle data value will be in the $\frac{11+1}{2} = \frac{12}{2} = 6^{\text{th}}$ position. So the median quiz score is 7. Half of the quiz scores are 7 or less, and half of the quiz scores are 7 or more.

✓ Example 8.4.6

Returning to the football touchdown data, find the median number of touchdown passes thrown per team in the NFL in the 2021 season.

Solution

We would start by listing the data in order.

12 14 15 16 20 20 20 20 21 21 21 21 21 22 23 23

23 24 26 27 27 29 30 34 36 36 37 38 39 40 41 43

Since there are $n = 32$ data values, an even number, the median will be the mean of the 2 middle values, the 16th and 17th data values. The 16th and 17th data values are both 23, so the mean of them is also 23 (since $\frac{23+23}{2} = 23$). The median number of touchdown passes per team in the 2021 season was 23 passes. Half of the teams had 23 touchdown passes thrown or less for the season, and half of the teams had 23 or more.

Try It 8.4.2

The price of a jar of peanut butter at 5 stores were: \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99. Find the median price.

Answer

First we put the data in order: \$3.29, \$3.59, \$3.75, \$3.79, \$3.99. Since there are an odd number of data, the median will be the middle value, the 3rd data value, \$3.75.

✓ Example 8.4.7

Let us return now to our original household income data. Find the median household income.

Income (thousands of dollars)	Frequency
15	6
20	8
25	11
30	17
35	19
40	20
45	12
50	7

Solution

Here we have $n = 100$ data values. If we didn't already know that, we could find it by adding the frequencies. Since 100 is an even number, we need to find the mean of the middle two data values - the 50th and 51st data values. To find these, we start counting up from the bottom:

There are 6 data values of 15, so	values 1 to 6 are \$15 thousand
The next 8 data values are 20, so	values 7 to $(6 + 8) = 14$ are \$20 thousand
The next 11 data values are 25, so	values 15 to $(14 + 11) = 25$ are \$25 thousand
The next 17 data values are 30, so	values 26 to $(25 + 17) = 42$ are \$30 thousand
The next 19 data values are 35, so	values 43 to $(42 + 19) = 61$ are \$35 thousand

From this we can tell that values 50 and 51 will be \$35 thousand, and the mean of these two values is \$35 thousand. The median income in this neighborhood is \$35 thousand. Thus, half of the households' annual income is \$35,000 or less and the other half is \$35,000 or more.

✓ Example 8.4.8

If we add in the new neighbor with a \$5 million household income, what is the median household income?

Solution

There will be $n = 101$ data values, and the 51st value will be the median. As we discovered in the last example, the 51st value is \$35 thousand. Notice that the new neighbor did not affect the median in this case. The median is not swayed as much by outliers as the mean is.

Let's think about the previous example. When we added the 101st family's income, the mean was \$81,069 from \$31,900. That's a big difference in the average household income. We see that the mean is influenced by the values of the data, i.e., the mean could get larger or smaller depending on the values of the data. However, when calculating the median including the 101st family's income, the median wasn't influenced at all. In fact, in general, the median is known as a better statistic for household income since there is a wide spread of income among families. Thus, the values of the data influence the mean, but not the median.

Mode

In addition to the mean and the median, there is one other common measurement of the "typical" value of a data set: the **mode**.

Mode

The **mode** is the value of the data set that occurs most frequently.

The mode is most commonly used for categorical data, for which the median and mean cannot be computed. Also, the mode is the only measure of central tendency that is used for both categorical and quantitative data. The mean and median are only used with quantitative data.

✓ Example 8.4.8

In our vehicle color survey, we collected the data below. Find the mode.

Color	Frequency
Blue	3
Green	5
Red	4
White	3
Black	2
Grey	3

Solution

For this data, Green is the mode, since it is the data value that occurred the most frequently. This is often called the *modal class* when referring to categorical data.

It is possible for a data set to have more than one mode if several categories have the same frequency, or no modes if each every category occurs only once.

Try It 8.4.3

Reviewers were asked to rate a product on a scale of 1 to 5. The results of the survey are below. Find:

- The mean rating
- The median rating
- The mode rating

Rating	Frequency
1	4
2	8
3	7
4	3
5	1

Answer

- a. The mean rating is $\frac{1 \cdot 4 + 2 \cdot 8 + 3 \cdot 7 + 4 \cdot 3 + 5 \cdot 1}{23} = \frac{58}{23} \approx 2.5$. Note that this is actually categorical data, so it is not appropriate to calculate the mean for this data. This number is meaningless since the rating values are based on a customer's opinion. Think about restaurant ratings or ratings for products on Amazon: some customers will rate the product itself, while others will rate the customer service. There is no clear criteria for what makes a product a "1" rating versus a "5" rating. This erroneous type of "mean rating" is used quite extensively when shopping online.

- b. There are 23 data values, so the median will be the 12th data value. Ratings of 1 are the first 4 values, while a rating of 2 are the next 8 values, so the 12th value will be a rating of 2. The median rating is 2. Half of the ratings are 2 or less, and the other half of the ratings are 2 or more.
- c. The mode is the most frequent rating. The mode rating is 2.

[1] <https://www.statmuse.com/nfl/ask/mos...-nfl-2021-team>

This page titled [8.4: Measures of Central Tendency](#) is shared under a [CC BY-SA 3.0](#) license and was authored, remixed, and/or curated by [David Lippman & Jeff Eldridge \(The OpenTextBookStore\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [11.5: Measures of Central Tendency](#) by David Lippman & Jeff Eldridge is licensed [CC BY-SA 3.0](#). Original source: <http://www.opentextbookstore.com/mathinsociety>.