

2.4: Measures of the Center of the Data

The "center" of a data set is also a way of describing location. The two most widely used measures of the "center" of the data are the **mean** (average) and the **median**. To calculate the **mean weight** of 50 people, add the 50 weights together and divide by 50. Technically this is the arithmetic mean. We will discuss the geometric mean later. To find the **median weight** of the 50 people, order the data and find the number that splits the data into two equal parts meaning an equal number of observations on each side. The weight 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. The mean is the most common measure of the center.

NOTE

The words "mean" and "average" are often used interchangeably. The substitution of one word for the other is common practice. The technical term is "arithmetic mean" and "average" is technically a center location. Formally, the arithmetic mean is called the first moment of the distribution by mathematicians. However, in practice among non-statisticians, "average" is commonly accepted for "arithmetic mean."

When each value in the data set is not unique, the mean can be calculated by multiplying each distinct value by its frequency and then dividing the sum by the total number of data values. The letter used to represent the **sample mean** is an x with a bar over it (pronounced " x bar"): \bar{x} .

To see that both ways of calculating the mean are the same, consider the sample:

1; 1; 1; 2; 2; 3; 4; 4; 4; 4; 4

$$\bar{x} = \frac{1+1+1+2+2+3+4+4+4+4+4}{11} = 2.7$$

$$\bar{x} = \frac{3(1)+2(2)+1(3)+5(4)}{11} = 2.7$$

In the second calculation, the frequencies are 3, 2, 1, and 5.

The Greek letter μ (pronounced "mew") represents the **population mean**. One of the requirements for the **sample mean** to be a good estimate of the **population mean** is for the sample taken to be truly random.

You can quickly find the location of the median by using the expression $\frac{n+1}{2}$.

The letter n is 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 smallest to largest). If n is an even number, the median is equal to the two middle values added together and divided by two after the data has been ordered. For example, if the total number of data values is 97, then $\frac{n+1}{2} = \frac{97+1}{2} = 49$. The median is the 49th value in the ordered data. If the total number of data values is 100, then $\frac{n+1}{2} = \frac{100+1}{2} = 50.5$. The median occurs midway between the 50th and 51st values. The location of the median and the value of the median are **not** the same. The upper case letter M is often used to represent the median. The next example illustrates the location of the median and the value of the median.

? Example 2.4.1

AIDS data indicating the number of months a patient with AIDS lives after taking a new antibody drug are as follows (smallest to largest):
3; 4; 8; 8; 10; 11; 12; 13; 14; 15; 15; 16; 16; 17; 17; 18; 21; 22; 22; 24; 24; 25; 26; 26; 27; 27; 29; 29; 31; 32; 33; 33; 34; 34; 35; 37; 40; 44; 44; 47;
Calculate the mean and the median.

Answer

The calculation for the mean is:

$$\bar{x} = \frac{3+4+(8)(2)+10+11+12+13+14+(15)(2)+\dots+35+37+40+(44)(2)+47}{40} = 23.6$$

To find the median, M , first use the formula for the location. The location is:

$$\frac{n+1}{2} = \frac{40+1}{2} = 20.5$$

Starting at the smallest value, the median is located between the 20th and 21st values (the two 24s):

3; 4; 8; 8; 10; 11; 12; 13; 14; 15; 15; 16; 16; 17; 17; 18; 21; 22; 22; 24; 24; 25; 26; 26; 27; 27; 29; 29; 31; 32; 33; 33; 34; 34; 35; 37; 40; 44; 44; 47;

$$M = \frac{24+24}{2} = 24$$

? Example 2.4.2

Suppose that in a small town 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?

Answer

$$\bar{x} = \frac{5,000,000+49(30,000)}{50} = 129,400$$

$$M = 30,000$$

(There are 49 people who earn \$30,000 and one person who earns \$5,000,000.)

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.

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 have the same frequency and that frequency is the highest. A data set with two modes is called bimodal.

? Example 2.4.3

Statistics exam scores for 20 students are as follows:

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

Find the mode.

Answer

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

? Example 2.4.4

Five real estate exam scores are 430, 430, 480, 480, 495. The data set is bimodal because the scores 430 and 480 each occur twice.

When is the mode the best measure of the "center"? Consider a weight loss program that advertises a mean weight loss of six pounds the first week of the program. The mode might indicate that most people lose two pounds the first week, making the program less appealing.

NOTE

The mode can be calculated for qualitative data as well as for quantitative data. For example, if the data set is: red, red, red, green, green, yellow, purple, black, blue, the mode is red.

Calculating the Arithmetic Mean of Grouped Frequency Tables

When only grouped data is available, you do not know the individual data values (we only know intervals and interval frequencies); therefore, you cannot compute an exact mean for the data set. What we must do is estimate the actual mean by calculating the mean of a frequency table. A frequency table is a data representation in which grouped data is displayed along with the corresponding frequencies. To calculate the mean from a grouped frequency table we can apply the basic definition of mean: $mean = \frac{\text{data sum}}{\text{number of data values}}$. We simply need to modify the definition to fit within the restrictions of a frequency table.

Since we do not know the individual data values we can instead find the midpoint of each interval. The midpoint is $\frac{\text{lower boundary} + \text{upper boundary}}{2}$. We can now modify the mean definition to be **Mean of Frequency Table** = $\frac{\sum fm}{\sum f}$ where f = the frequency of the interval and m = the midpoint of the interval.

? Example 2.4.5

A frequency table displaying professor Blount's last statistic test is shown. Find the best estimate of the class mean.

Table 2.24

Grade interval	Number of students
50–56.4	1
56.5–62.4	0
62.5–68.4	4
68.5–74.4	4
74.5–80.4	2
80.5–86.4	3
86.5–92.4	4
92.5–98.4	1

Answer

Find the midpoints for all intervals

Table 2.25

Grade interval	Midpoint
50–56.4	53.25
56.5–62.4	59.5
62.5–68.4	65.5
68.5–74.4	71.5
74.5–80.4	77.5
80.5–86.4	83.5
86.5–92.4	89.5
92.5–98.4	95.5

- Calculate the sum of the product of each interval frequency and midpoint. $\sum fm$
 $53.25(1) + 59.5(0) + 65.5(4) + 71.5(4) + 77.5(2) + 83.5(3) + 89.5(4) + 95.5(1) = 1460.25$
- $\mu = \frac{\sum fm}{\sum f} = \frac{1460.25}{19} = 76.86$

? Exercise 2.4.1

Maris conducted a study on the effect that playing video games has on memory recall. As part of her study, she compiled the following data:

Table 2.26

Hours teenagers spend on video games	Number of teenagers
0–3.4	3
3.5–7.4	7
7.5–11.4	12
11.5–15.4	7
15.5–19.4	9

What is the best estimate for the mean number of hours spent playing video games?

This page titled 2.4: Measures of the Center of the Data is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax.

- 2.3: Measures of the Center of the Data by OpenStax is licensed CC BY 4.0. Original source: <https://openstax.org/details/books/introductory-business-statistics>.