

8.5: Measures of Variation and Location

Learning Objectives

- Find the range of a data set
- Find and interpret the standard deviation of a data set
- Interpret percentiles and calculate quartiles of a data set

A second aspect of a distribution is how spread out it is. In other words, how much the data in the distribution vary from one another. Numbers that describe a distribution's spread or amount of variability are called measures of variation.

Consider these three sets of quiz scores:

- section A: 5 5 5 5 5 5 5 5 5
- section B: 0 0 0 0 10 10 10 10 10
- section C: 4 4 4 5 5 5 5 6 6

All three of these sets of data have a mean of 5 and median of 5, yet the sets of scores are clearly quite different. In section A, everyone had the same score; in section B half the class got no points and the other half got a perfect score, assuming this was a 10-point quiz. Section C was not as consistent as section A, but not as widely varied as section B.

In addition to the mean and median, which are measures of the "typical" or "middle" value, we also need a measure of how "spread out" or varied each data set is.

Range

There are several ways to measure this "spread" of the data. The first is the simplest and is called the **range**.

Range

The range is the difference between the maximum value and the minimum value of the data set.

✓ Example 8.5.1

Using the quiz scores from above, find the range of each section.

Solution

For section A, the range is 0 since both maximum and minimum are 5 and $5 - 5 = 0$.

For section B, the range is 10 since $10 - 0 = 10$.

For section C, the range is 2 since $6 - 4 = 2$.

In the last example, the range seems to be revealing how spread out the data is. However, suppose we add a fourth section, section D, with scores 0 5 5 5 5 5 5 5 10.

This section also has a mean and median of 5. The range is 10, yet this data set is quite different than section B. To better illuminate the differences, we'll have to turn to more sophisticated measures of variation.

Standard Deviation

Standard deviation

The **standard deviation** is a measure of variation based on measuring how far each data value deviates, or is different, from the mean. A few important characteristics:

- Standard deviation is always positive. Standard deviation will be zero if all the data values are equal, and will get larger as the data spreads out.
- Standard deviation has the same units as the original data.

- Standard deviation, like the mean, can be highly influenced by outliers.

Using the data from section D, we could compute for each data value the difference between the data value and the mean:

data value	deviation: data value - mean
0	$0 - 5 = -5$
5	$5 - 5 = 0$
5	$5 - 5 = 0$
5	$5 - 5 = 0$
5	$5 - 5 = 0$
5	$5 - 5 = 0$
5	$5 - 5 = 0$
5	$5 - 5 = 0$
5	$5 - 5 = 0$
10	$10 - 5 = 5$

We would like to get an idea of the "average" deviation from the mean, but if we find the average of the values in the second column the negative and positive values cancel each other out (this will always happen), so to prevent this we square every value in the second column:

We then add the squared deviations up to get $25 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 25 = 50$. Ordinarily we would then divide by the number of scores, n , (in this case, 10) to find the mean of the deviations. But we only do this if the data set represents a population; if the data set represents a sample (as it almost always does), we instead divide by $n - 1$ (in this case, $10 - 1 = 9$). [1]

So in our example, we would have $\frac{50}{10} = 5$ if section D represents a population and $\frac{50}{9} \approx 5.56$ if section D represents a sample. These values (5 and 5.56) are called, respectively, the **population variance** and the **sample variance** for section D.

Variance can be a useful statistical concept, but note that the units of variance in this instance would be points-squared since we squared all of the deviations. What are points-squared? Good question. We would rather deal with the units we started with (points in this case), so to convert back we take the square root and get:

$$\text{population standard deviation} = \sqrt{\frac{50}{10}} = \sqrt{5} \approx 2.2$$

or

$$\text{sample standard deviation} = \sqrt{\frac{50}{9}} \approx 2.4$$

What does this say about section D? We can say that the average score was 5 give or take 2.4. The "give or take" part is the prefix for standard deviation. Or we can say that the data values in section D are, on average, 2.4 points away from the mean 5. One data value is below the mean, some data values are the same as the mean, and one value is above the mean. But if you average out how far the data values are away from the mean, you will get about 2.4 points.

If we are unsure whether the data set is a sample or a population, we will usually assume it is a sample, and we will round answers to one more decimal place than the original data, as we have done above.

To compute standard deviation:

1. Find the deviation of each data from the mean. In other words, subtract the mean from the data value.
2. Square each deviation.
3. Add the squared deviations.
4. Divide by n , the number of data values, if the data represents a whole population; divide by $n - 1$ if the data is from a sample. The value of this answer is the variance.
5. Compute the square root of the result from the previous step. The value of this answer is the standard deviation.

✓ Example 8.5.2

Compute the standard deviation for section B above.

Solution

We first calculate the mean, which is 5. Using a table can help keep track of your computations for the standard deviation:

data value	deviation: data value - mean	deviation squared
0	$0 - 5 = -5$	$(-5)^2 = 25$
0	$0 - 5 = -5$	$(-5)^2 = 25$
0	$0 - 5 = -5$	$(-5)^2 = 25$
0	$0 - 5 = -5$	$(-5)^2 = 25$
0	$0 - 5 = -5$	$(-5)^2 = 25$
10	$10 - 5 = 5$	$(5)^2 = 25$
10	$10 - 5 = 5$	$(5)^2 = 25$
10	$10 - 5 = 5$	$(5)^2 = 25$
10	$10 - 5 = 5$	$(5)^2 = 25$
10	$10 - 5 = 5$	$(5)^2 = 25$

Assuming this data represents a population, we will add the squared deviations, divide by 10, the number of data values, and compute the square root:

$$\sqrt{\frac{25 + 25 + 25 + 25 + 25 + 25 + 25 + 25 + 25 + 25}{10}} = \sqrt{\frac{250}{10}} = 5$$

Notice that the standard deviation of this data set is much larger than that of section D since the data in this set is more spread out. On average, the data values in section B are 5 points away from the mean of 5. This is exactly what you would expect with half of the data values equal to 0 and the other half of the data values equal to 10.

For comparison, the (population) standard deviations of all four sections are:

section A: 5 5 5 5 5 5 5 5 5 5	standard deviation: 0
section B: 0 0 0 0 10 10 10 10 10 10	standard deviation: 5
section C: 4 4 4 5 5 5 5 6 6 6	standard deviation: 0.8
section D: 0 5 5 5 5 5 5 5 10	standard deviation: 2.2

Note: most scientific and graphing calculators have functions for calculating the mean and standard deviation of a data set. Check your calculator's user manual to find out how yours works.

✎ Try It 8.5.1

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 standard deviation of the prices.

Answer

Earlier we found the mean of the data was \$3.682.

data value	deviation: data value - mean	deviation squared
3.29	$3.29 - 3.682 = -0.391$	0.153664
3.59	$3.59 - 3.682 = -0.092$	0.008464
3.79	$3.79 - 3.682 = 0.108$	0.011664
3.75	$3.75 - 3.682 = 0.068$	0.004624
3.99	$3.99 - 3.682 = 0.308$	0.094864

This data is from a sample, so we will add the squared deviations, divide by $n - 1 = 4$, the number of data values minus 1, and compute the square root:

$$\sqrt{\frac{0.153664 + 0.008464 + 0.011664 + 0.004624 + 0.094864}{4}} \approx \$0.261$$

On average, the prices of peanut butter from the 5 stores are \$0.261 away from the mean of \$3.682.

Percentiles and Quartiles

There are other calculations that we can do to look at spread. One of those is called the **percentile**. This looks at what data value has a certain percent of the data at or below it. It is also known as a measure of location since it gives the location or position of a data value in the data set relative to the other data values.

Percentile

The k^{th} **percentile** is the value with $k\%$ of the data at or below this value.

For example, if a data value is in the 80th percentile, then 80% of the data values fall at or below this value, and 20% of the data values fall at or above it.

We see percentiles in many places in our lives. If you take any standardized tests, your score is given as a percentile. If you take your child to the doctor, their height and weight are given as percentiles. If your child is tested for gifted or behavior problems, the score is given as a percentile. If your child has a score on a gifted test that is in the 92nd percentile, then that means that 92% of all of the children who took the same gifted test scored the same or lower than your child. That also means that 8% scored the same or higher than your child. This may mean that your child is gifted.

✓ Example 8.5.3

Suppose you took the SAT mathematics test and received your score as a percentile.

- What does a score in the 90th percentile mean?
- What does a score in the 70th percentile mean?
- If the test was out of 800 points and you scored in the 80th percentile, what was your score on the test?
- If your score was in the 95th percentile, does that mean you passed the test?

Solution

- 90% of the scores were at or below your score and 10% of the scores were at or above your score. (You did the same as or better than 90% of the test takers.)
- 70% of the scores were at or below your score and 30% of the scores were at or above your score.
- You do not know! All you know is that you scored the same as or better than 80% of the people who took the test. If all the scores were really low, you could have still failed the test. On the other hand, if many of the scores were high you could have gotten a 95% on the test.
- No, it just means you did the same as or better than 95% of the other people who took the test. You could have failed the test, but still did the same as or better than 95% of the rest of the people.

While standard deviation is a measure of variation based on the mean, **quartiles** are a measure of location based on the median. Quartiles are a type of percentile.

Quartiles

Quartiles are values that divide the data in quarters.

The **first quartile** (Q_1) is the value so that 25% of the data values are less than or equal to it; the **third quartile** (Q_3) is the value so that 75% of the data values are less than or equal to it. You may have guessed that the **second quartile** is the same as the median, since the median is the value so that 50% of the data values are less than or equal to it.

This divides the data into quarters; 25% of the data is between the minimum and Q_1 , 25% is between Q_1 and the median, 25% is between the median and Q_3 , and 25% is between Q_3 and the maximum value.

To find the first quartile, we need to find the median of the first half of the data set. Similarly, to find the third quartile, we need to find the median of the second half of the data set. It is easiest to find the median first to divide the data set into the first half and second half.

Finding the Quartiles

1. Begin by ordering the data from smallest to largest.
2. Find the median. This is the second quartile.
3. Separate the data set into 2 data sets: the half before the median and the half after the median. Do not include the median in either of the half sets.
4. Find the median of the first half of the data set. This is the first quartile.
5. Find the median of the second half of the data set. This is the third quartile.

Examples should help make this clearer.

✓ Example 8.5.4

Suppose we have measured 9 females and their heights (in inches), sorted from smallest to largest are:

59 60 62 64 66 67 69 70 72

Find the first and third quartiles.

Solution

Since there are $n = 9$ data values, the median will be the 5th data value, 66. (This is the second quartile.)

To find the first quartile we find the median of the first half of the data set: 59 60 62 64. Since there are now $n = 4$ data values, the median will be the average of the 2nd and 3rd data values: $\frac{60 + 62}{2} = 61$. The first quartile is 61 inches. We can say that 25% of females are 61 inches or shorter and 75% of females are 61 inches or taller.

To find the third quartile, we find the median of the second half of the data set: 67 69 70 72. The median will again be the average of the 2nd and 3rd data values: $\frac{69 + 70}{2} = 69.5$. The third quartile is 69.5 inches. We can say that 75% of females are 69.5 inches or shorter and 25% of females are 69.5 inches or taller.

Note the locations of the quartiles denoted by the red lines and the median shaded in red:

59 60 | 62 64 **66** 67 69 | 70 72

The median separates the data set into 2 halves. The first quartile is the median of the first half of the data set and the third quartile is the median of the second half of the data set.

It is also worth noting that if there were only 8 females with the following data: 59 60 62 64 66 67 69 70, the median would be 65 inches (the average of the 4th and 5th data values) but the first quartile would be the same since the first half of the data set would be the same. The third quartile would be 68 inches. It would look like:

59 60 | 62 64 | 66 67 | 69 70

Try It 8.5.2

The prices of a jar of peanut butter at 5 stores were: \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99. Find the first and third quartiles.

Answer

The data in order are: \$3.29, \$3.59, \$3.75, \$3.79, \$3.99.

The median is the middle value, \$3.75.

The first quartile is the median of the first half of the data, \$3.29 and \$3.59. Since there are only 2 values, the first quartile will be the average of the data values: $\frac{3.29 + 3.59}{2} = 3.44$. The first quartile is \$3.44. 25% of peanut butter costs \$3.44 or less and 75% of peanut butter costs \$3.44 or more.

The third quartile is the median of the second half of the data, \$3.79 and \$3.99. It will be the average of the 2 data values: $\frac{3.79 + 3.99}{2} = 3.89$. The third quartile is \$3.89. 75% of peanut butter costs \$3.89 or less and 25% of peanut butter costs \$3.89 or more.

[1] The reason we do this is highly technical, but we can see how it might be useful by considering the case of a small sample from a population that contains an outlier, which would increase the average deviation: the outlier very likely won't be included in the sample, so the mean deviation of the sample would underestimate the mean deviation of the population; thus we divide by a slightly smaller number to get a slightly bigger average deviation.

This page titled [8.5: Measures of Variation and Location](#) 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.

- **11.6: Measures of Variation** by David Lippman & Jeff Eldridge is licensed [CC BY-SA 3.0](#). Original source: <http://www.opentextbookstore.com/mathinsociety>.
- **2.3: Measures of Spread** by Maxie Inigo, Jennifer Jameson, Kathryn Kozak, Maya Lanzetta, & Kim Sonier is licensed [CC BY-SA 4.0](#). Original source: <https://www.coconino.edu/open-source-textbooks#college-mathematics-for-everyday-life-by-inigo-jameson-kozak-lanzetta-and-sonier>.