Calculating Data Variability

by Sophia

☰ WHAT'S COVERED

This lesson discusses representing how data can vary. You will be able to identify the symbols for sample variance and sample standard deviation. You will also learn how to calculate these terms. This lesson covers:

1. Calculating Variance
2. Calculating Standard Deviation

---

1. Calculating Variance

The first step in calculating standard deviation is to calculate a related statistic, **variance**. Both statistics indicate the spread of the data, but the standard deviation is more commonly used because it has the same scale as the data from which it is calculated.

Generally speaking, a lower variance is preferable to a higher variance when trying to compare the results of two tests. A lower variance indicates we are more confident about the sample mean, and any difference in sample means is more likely to indicate a significant difference. We will discuss this more in later lessons.

IN CONTEXT

Suppose that you conducted an experiment aimed at establishing the length of time that patients had to wait to see two different doctors. Both doctors had a mean wait time of 18 minutes, but the variation in the data was significantly different.

Suppose the standard deviation of the wait time for Dr. Smith is 6 minutes, whereas for Dr. Jones, it is 2 minutes. If we were to only look at the mean, we might think patients would have a similar wait, but if we take into account variance or standard deviation, we realize that the patients of Dr. Jones have a much more consistent wait time.
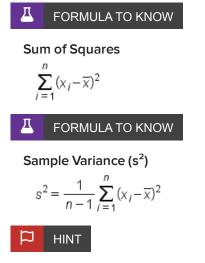
The variance can be found using the following steps:

1. Compute the mean, x̄ of the data set. Given a list of numbers, the mean is found by adding up all the numbers and then dividing by how many numbers there are. This is expressed by the equation "sum of x divided by n," where the "sum of x" refers to the sum of all the numbers and "n" is how many numbers there are in the data set.

2. Next, subtract the mean from each number in the data set and square the difference. Now, add all of the values in the previous step. This quantity is called a **sum of squares**.

3. Now we divide the sum of squares by n minus 1 (the total number of values in the data minus 1). The quantity we obtained is the **sample variance**.

🧪 FORMULA TO KNOW

Sum of Squares

$$\sum_{i=1}^{n} (x_i - \overline{x})^2$$

🧪 FORMULA TO KNOW

Sample Variance (s^2)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

🚩 HINT

Take a look at these last two formulas. Sample variance is the sum of squares divided by n minus 1

📄 TERM TO KNOW

Variance

A measure of spread for a set of data.

2. Calculating Standard Deviation

Both variance and standard deviation indicate the spread or variability of a data set, but standard deviation has the advantage of being in the same units as your data. In contrast, variance is expressed in squared units, which makes it somewhat more difficult to interpret. Standard deviation is calculated by taking the square root of the variance.

🧪 FORMULA TO KNOW

Sample Standard Deviation (s)

$$s = \sqrt{\text{variance}} = \sqrt{\frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

IN CONTEXT

This table shows the price of college textbooks, and it includes eight different observations. The left column shows the total textbook price for each individual book.

x	$x - \bar{x}$	$(x - \bar{x})^2$
145	-27	729
150	-22	484
165	-7	49
172	0	0
177	5	25
182	10	100
185	13	169
200	28	784
Σ (x) = 1376		

The first step, in this case, is to determine the sum of these values. If you add them all up together, you get 1,376. Next, divide that by 8, the number of observations, and you will obtain a mean (x̄)of 172.

The next step after that would be to subtract x̄ from each of the values of x listed in the second column. Next, square the difference as listed in the third column. Add all the squares in column three to obtain a sum of squares of 2,340.

You can calculate the variance by taking this value and dividing it by n minus 1, or in this case, 7. You get a variance of 334.28 for this sample. And in the last step to determine the standard deviation, take the square root of the sample variance, which in this case would be 18.28.

Sum of x = Σ (x) = 1,376
Sum of x/n = x̄ = 172
Sum of Squares = 2,340
Variance = s² = 334.28
Standard Deviation = s = 18.28

SUMMARY

In this lesson, you learned **how to calculate** two related measures of variability, or spread. **Variance** is equal to the sum of squares divided by n minus one observations. **Standard deviation** is the square root of variance and provides you with an indication of how close data is distributed to the mean.

Source: THIS TUTORIAL WAS AUTHORED BY DAN LAUB FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

TERMS TO KNOW

Variance

A measure of spread for a set of data.

FORMULAS TO KNOW

Sample Standard Deviation

$$s = \sqrt{variance} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Sum of Squares

$$\sum_{i=1}^{n}(x_i - \overline{x})^2$$