

How to Understand and Use Standard Deviation for Data Analysis

Authored by
stats writer

February 20, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Understand and Use Standard Deviation for Data Analysis*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=131761>

Standard deviation is a commonly used statistical measure that describes the spread or variability of a set of data from its mean. It has several advantages and disadvantages that should be considered when using it for data analysis.

One of the main advantages of using standard deviation is that it provides a precise and objective measure of the dispersion of data. It takes into account all the values in a dataset and provides a single value that can be compared across different datasets. This makes it a useful tool for understanding and comparing the variability of data.

Another advantage is that it is a well-established and widely used measure in statistics, making it easy to interpret and apply in various fields of study. It can also be used to identify outliers or extreme values in a dataset, which can help in detecting errors or unusual patterns in the data.

However, there are also some limitations or disadvantages to using standard deviation. One of the main drawbacks is that it is highly influenced by extreme values, also known as outliers. This means that a few extreme values can significantly affect the value of the standard deviation and may not accurately represent the overall variability of the data.

Another disadvantage is that it assumes a normal distribution of data, which may not always be the case in real-world situations. This can lead to misleading interpretations of the data if the underlying distribution is not normal. Additionally, calculating standard deviation requires a large sample size, which may not always be feasible or practical.

In conclusion, standard deviation has its advantages and disadvantages when used for data analysis. While it is a useful and widely accepted measure of variability, it is important to consider its limitations and potential biases when interpreting the results.

Advantages & Disadvantages of Using Standard Deviation

The standard deviation of a dataset is a way to measure the typical deviation of individual values from the mean value.

The formula to calculate a sample standard deviation,

denoted s , is:

$$s = \sqrt{\sum (x_i - \bar{x})^2 / (n - 1)}$$

where:

Σ : A symbol that means "sum"
 x_i : The i th value in a dataset
 $\bar{x}$: The sample mean
 n : The sample size

There are two main advantages of using the standard deviation to describe the spread of values in a dataset:

Advantage #1: The standard deviation uses all observations in a dataset in its calculation. In statistics, we generally say it's a good thing when we are able to use all observations in a dataset to perform some calculation because we are using all possible "information" available in the dataset.

Advantage #2: The standard deviation is easy to interpret. The standard deviation is a single value that gives us a good idea of how far the "typical" observation in a dataset lies from the mean value.

However, there is one main disadvantage of using the standard deviation:

Disadvantage #1: The standard deviation can be affected by outliers. When extreme outliers are present in a dataset, this can inflate the value of the standard deviation and thus give a misleading idea of the spread of values in a dataset.

The following examples provide more information about these advantages and disadvantages of using the standard deviation.

Advantage #1: The standard deviation uses all observations

Suppose we have the following dataset that shows the distribution of exam scores for students in a class:

Scores: 68, 70, 71, 75, 78, 82, 83, 83, 85, 90, 91, 91, 92

We can use a calculator or statistical software to find that the sample standard deviation of this dataset is 8.46.

The nice thing about using the standard deviation in this example is that we use all possible observations in the dataset to find the typical "spread" of values.

By contrast, we could use another metric such as the

interquartile range to measure the spread of values in this dataset.

Now suppose we change the lowest value in the dataset to be much lower:

Scores: 22, 70, 71, 75, 78, 82, 83, 83, 85, 90, 91, 91, 92

We can use a calculator to find that the sample standard deviation is 18.37.

However, the interquartile range is still 17.5 because none of the middle 50% of values were affected.

This shows that the sample standard deviation considers all observations in the dataset in its calculation while other do not.

Advantage #2: The standard deviation is easy to interpret

Recall the following dataset that shows the distribution of exam scores for students in a class:

Scores: 68, 70, 71, 75, 78, 82, 83, 83, 85, 90, 91, 91, 92

We used a calculator to find that the sample standard deviation of this dataset was 8.46.

This is easy to interpret because it simply means the deviation of a "typical" exam score is about 8.46 away from the mean exam score.

By contrast, other measures of dispersion are not so straightforward to interpret.

For example, a coefficient of variation is another measure of dispersion that represents the ratio of the standard deviation to the sample mean.

Coefficient of Variation: $s / \bar{x}$?

In this example, the mean exam score is 81.46 so the coefficient of variation is calculated as $8.46 / 81.46 = 0.104$.

This represents the ratio of the sample standard deviation to the sample mean, which can be useful for comparing the spread of values between multiple datasets but it isn't very straightforward to interpret as a metric by itself.

Disadvantage #1: The standard deviation can be affected by outliers

Suppose we have the following dataset that contains information about the salaries of 10 employees (in thousands of dollars) at some company:

Salaries: 44, 48, 57, 68, 70, 71, 73, 79, 84, 94

The sample standard deviation of salaries is about 15.57.

Now suppose we have the exact same dataset but the largest salary is *much* larger:

Salaries: 44, 48, 57, 68, 70, 71, 73, 79, 84, 895

The sample standard deviation of salaries in this dataset is about 262.47.

By including just one extreme outlier, the standard deviation is highly affected and now provides a misleading idea of the "typical" spread of salaries.

Note: When outliers are present in a dataset, the interquartile range can provide a better measure of dispersion because it is unaffected by outliers.

The following tutorials provide additional information

about using the standard deviation in statistics:

ARABPSYCHOLOGY.COM