What is the difference between STDEV.P vs. STDEV.S in Excel?

Authored by

stats writer

December 16, 2025

Understanding the fundamental difference between the standard deviation functions in Excel--specifically **STDEV.P** and **STDEV.S**--is essential for accurate statistical analysis. These two functions calculate the variability of a dataset, but they differ critically based on whether the data represents a complete statistical population or merely a sample drawn from that larger group. The **STDEV.P** function calculates the standard deviation assuming that the data range includes every single value within the entire population under study. In contrast, **STDEV.S** calculates the standard deviation when the data is only a subset, or sample, of the total population, which is the scenario most commonly encountered in empirical research and data analysis.

This distinction leads to a crucial variation in their underlying mathematical formulas. While both rely on the sum of squared differences from the mean, **STDEV.P** uses the total number of observations, N, as the divisor in its variance calculation. Conversely, **STDEV.S** employs N-1 as the divisor, a statistical adjustment known as Bessel's correction. This subtle yet powerful difference ensures that the sample standard deviation provides a less biased estimate of the true population variability, acknowledging the inherent uncertainty introduced by using incomplete data. We will delve into these concepts to clarify exactly when and why you should choose one function over the other.

Understanding Standard Deviation and Data Spread

Before exploring the specific functions, it is vital to firmly grasp what the concept of standard deviation represents in statistics. Standard deviation is one of the most widely used measures of dispersion, quantifying the amount of variation or spread of a set of data values. A low standard deviation indicates that the data points tend to be close to the dataset's mean, while a high standard deviation suggests that the data points are spread out over a wider range of values. It provides essential context to the average, allowing analysts to understand how reliable or representative that average truly is when interpreting performance metrics or drawing conclusions.

Calculating the standard deviation involves several sequential steps rooted in the concept of variance. The process begins with calculating the mean (average) of the dataset. Subsequently, the deviation of each data point from that mean is determined. These deviations are then squared to remove negative values and, importantly, to give disproportionately more weight to extreme deviations. Finally, the variance is calculated by averaging these squared deviations, and the square root of the variance is taken to return the measure back to the original units of the data. This rigorous process provides a clear, interpretable measure of volatility or consistency within the data, essential for applications ranging from financial risk assessment to quality control.

The choice between **STDEV.P** and **STDEV.S** hinges entirely on the scope of your data collection. If you have collected every possible data point relevant to your study, you use the population function (P). If, as is typical, you have only managed to collect a partial set of observations, you

must use the sample function (S). Using the wrong function will systematically introduce error into your analysis, potentially skewing conclusions about data volatility and assessment of parameters.

Statistical Population Versus Sample Data

The core statistical context driving the Excel functions is the differentiation between a statistical population and a sample. The population refers to the entire group of items or individuals that is the focus of a statistical investigation. If we are studying the sales performance of all retail stores operated by a particular chain in the United States, the population is every single one of those stores. Crucially, a population dataset is complete; it contains every data point that could possibly exist within the defined scope, allowing for the calculation of true population parameters, often denoted by Greek letters such as mu for the mean and $sigma$ for the standard deviation.

A sample, conversely, is a manageable subset of the population, selected for analysis. Since analyzing an entire population is often impossible due to time, cost, or logistical constraints, researchers rely on samples to make inferences about the whole population. For instance, instead of surveying every customer who purchased a product, a researcher might analyze a random selection of 500 purchase records. Statistics derived from a sample, such as the sample mean ($bar{x}$) and sample standard deviation (s), are estimates of the true population parameters. Since these statistics are based on incomplete information, they are subject to sampling error, meaning they might not perfectly reflect the true population values.

The decision to use **STDEV.P** or **STDEV.S** must be based on whether the data analyst genuinely believes that the provided dataset represents the total universe of possible values. If you are calculating the standard deviation of the height of every single tree planted in a designated urban park, and your interest is only in that park, you have the population. If you are using the heights of those trees to estimate the average height of all similar trees planted in the entire city, those measured trees constitute a sample.

STDEV.P: Calculating Population Standard Deviation

The **STDEV.P** function in Excel is specifically designed to calculate the true population standard deviation ($sigma$). This function should only be utilized when the range of values supplied to Excel encompasses every single observation belonging to the defined statistical population. When this strict condition is met, the resulting value is the most accurate measure of dispersion possible, as there is no uncertainty regarding missing data points. The population standard deviation is inherently descriptive--it perfectly describes the variability within that finite, complete set of numbers, and requires no adjustments for inference.

The mathematical formula utilized by **STDEV.P** involves dividing the sum of the squared differences from the population mean (mu) by the total number of observations, N. The

formula is fundamentally based on the calculation of variance divided by N, followed by taking the square root. The structure of the population standard deviation calculation is illustrated below, where x_i represents the individual data points:

There are three different functions you can use to calculate standard deviation in Excel:

1. STDEV.P: This function calculates the population standard deviation. Use this function when the range of values represents the entire population.

This function uses the following formula:

Population standard deviation = $\sqrt{\Sigma}$ (xi - μ)2 / N

where:

Σ: A greek symbol that means "sum"
xi: The ith value in the dataset
μ: The population mean
N: The total number of observations

Notice the denominator: N. Because we possess all members of the population, dividing by N (the population size) provides the precise average squared deviation, or variance, before taking the square root. If you are certain that your data is complete--such as analyzing the historical revenue generated by all customers who subscribed to a service during a single calendar year--then **STDEV.P** is the statistically correct function to choose.

STDEV.S: Calculating Sample Standard Deviation

The **STDEV.S** function is the more frequently utilized function in inferential statistics, as most real-world data collection involves taking a sample from a much larger, often unbounded, population. This function calculates the sample standard deviation, s, which serves as an unbiased estimate of the true population standard deviation ($sigma$). Since a sample, by definition, does not contain all possible data points, calculating the variability must account for the added uncertainty and potential bias inherent in partial data collection.

The key differentiator for **STDEV.S** lies in its implementation of Bessel's correction. Instead of dividing the sum of squared deviations by the sample size n, it divides by $n-1$, which represents the degrees of freedom. This adjustment is crucial because the sample mean ($bar{x}$) used in the calculation is itself derived from the sample data. The data points used to calculate the deviations tend to cluster more closely around their own sample mean than they would around the true population mean, leading to an underestimation of the true population variance if n were used. Dividing by $n-1$ systematically inflates the resulting estimate, correcting for this inherent

downward bias.

The mathematical representation used by **STDEV.S** in Excel is shown below, illustrating the use of $n-1$ in the denominator and the sample mean ($bar{x}$):

2. STDEV.S: This function calculates the sample standard deviation. Use this function when the range of values represents a sample of values, rather than an entire population.

This function uses the following formula:

Sample standard deviation = √Σ (xi - x)2 / (n-1)

where:

Σ: A greek symbol that means "sum"
xi: The ith value in the dataset
x: The sample mean
N: The total number of observations (representing the sample size, n)

If you are unsure whether your data represents the entire population, the statistically conservative and correct choice is almost always **STDEV.S**. It provides a more robust and less biased estimate when projecting sample statistics onto a larger population context, which is the primary goal of inferential statistical analysis.

The Statistical Justification: Bessel's Correction

The difference between dividing by N (for population variance) and $N-1$ (for sample variance, known as Bessel's correction) is arguably the most critical statistical concept distinguishing **STDEV.P** and **STDEV.S**. This correction addresses the mathematical reality that when the true population mean (mu) is unknown, and we must instead rely on the sample mean ($bar{x}$) to calculate variance, we inevitably introduce a systematic underestimation of the true population variability. This happens because the sum of squared deviations is minimized when calculated relative to its own mean.

Using N as the divisor results in a biased estimator of the population variance--it systematically underestimates the true spread. To correct this downward bias and ensure that the sample variance is an unbiased estimator of the population variance, we divide by the degrees of freedom, $N-1$. The degrees of freedom signify the number of values in the final calculation of a statistic that are free to vary. Since one parameter (the sample mean, $bar{x}$) must be estimated from the data itself before calculating the variance, one degree of freedom is lost, leaving $N-1$ degrees of freedom for the variance estimation.

This technical difference directly translates into the observed results: the population standard deviation calculated via **STDEV.P** will always be mathematically smaller than the sample standard deviation calculated via **STDEV.S** for the exact same dataset, provided $N>1$. When we know the entire population, our uncertainty is zero, resulting in the smallest possible measure of dispersion. Conversely, when we only have a sample, our uncertainty regarding the true population parameters is higher. The larger value returned by **STDEV.S** reflects this increased uncertainty, providing a more conservative and appropriate estimate for statistical inference.

Technical Note:

Since the formula for the population standard deviation divides by *N* instead of *n-1*, the population standard deviation will always be smaller than the sample standard deviation.

The reason the population standard deviation will be smaller is because if we know every value in the population, then we know the exact standard deviation.

However, when we only have a sample of the population then we have more uncertainty around the exact standard deviation of the overall population, so our estimate for the standard deviation needs to be larger.

The Legacy STDEV Function

In addition to **STDEV.P** and **STDEV.S**, users of Excel may encounter the older function, **STDEV**. This function was the primary standard deviation calculator in versions of Excel prior to Excel 2010. It is crucial to note that the **STDEV** function operates identically to **STDEV.S**; it calculates the sample standard deviation using the $N-1$ denominator (Bessel's correction). The purpose of its retention in modern Excel is purely for backward compatibility, ensuring older spreadsheets continue to calculate correctly.

Microsoft introduced the explicit `.P` and `.S` suffixes in Excel 2010 to dramatically improve clarity regarding the underlying statistical assumption of the calculation. This change aligns Excel with industry standards, which clearly separate population parameters from sample estimates. While **STDEV** is still functional, analysts are strongly encouraged to adopt the use of **STDEV.S** or **STDEV.P** to make their spreadsheet logic transparent and statistically robust. Mathematically, the calculation performed by the legacy **STDEV** function yields the exact same numerical result as **STDEV.S**.

3. STDEV: This function calculates the sample standard deviation as well. It will return the exact same value as the **STDEV.S** function.

Practical Application Example and Results Comparison

To demonstrate the numerical impact of choosing the correct function, let us examine a practical example using a defined dataset in Excel. We apply both **STDEV.P** and **STDEV.S** to the exact same range of data. The resulting difference highlights the mathematical consequence of the N versus $N-1$ divisor. The following example shows how to use these functions in practice.

Example: STDEV.P vs. STDEV.S in Excel

Suppose we have the following dataset in Excel:

	A	B	C	D	E
1	**Dataset**				
2	4				
3	5				
4	5				
5	6				
6	8				
7	9				
8	12				
9	12				
10	13				
11	15				
12	16				
13	17				
14	19				
15	21				
16	22				
17	24				
18	27				
19	28				
20	32				
21	33				
22					
23					
24					
25					
26					

As demonstrated in the visual output below, applying the respective functions to the same data range results in two distinct values. The calculation using **STDEV.S** consistently yields a higher result than the calculation using **STDEV.P**, confirming the theoretical difference introduced by Bessel's correction, which inflates the sample estimate to account for uncertainty.

	A	B	C	D	E	F
1	**Dataset**				Formula	
2	4		**STDEV.P**	8.896067	=STDEV.P(A2:A21)	
3	5		**STDEV.S**	9.127172	=STDEV.S(A2:A21)	
4	5		**STDEV**	9.127172	=STDEV(A2:A21)	
5	6					
6	8					
7	9					
8	12					
9	12					
10	13					
11	15					
12	16					
13	17					
14	19					
15	21					
16	22					
17	24					
18	27					
19	28					
20	32					
21	33					
22						
23						
24						
25						
26						
27						

The sample standard deviation (calculated using **STDEV.S**) turns out to be **9.127**, and the population standard deviation (calculated using **STDEV.P**) turns out to be **8.896**.

This numerical outcome perfectly illustrates the statistical principle. Since the population function assumes complete knowledge of the data, its estimate of dispersion is lower. The sample function, by using $N-1$, provides a slightly larger, more cautious estimate, which is generally more accurate for inferring the true variability of the larger population from which the data was drawn. As mentioned earlier, the population standard deviation will always be smaller than the sample standard deviation when calculated from the same finite set of observations.

When to Use STDEV.P vs. STDEV.S

The definitive decision between **STDEV.P** and **STDEV.S** should be guided by a clear understanding of the scope and context of your data collection. Statistically, you must use

STDEV.P only in cases where you have irrefutably collected data from every single member of the target group and your findings are intended to describe only that specific group. Examples include calculating the variability of final scores for all students currently enrolled in a small, defined course, or calculating the standard deviation of all transactions logged in a corporate database for a closed quarter.

However, in the overwhelming majority of statistical studies, particularly in fields relying on generalization--such as market forecasting, scientific experimentation, or large-scale sociological surveys--it is often impractical or impossible to collect data for the entire population. Consequently, the dataset you are analyzing almost always represents a sample. When the objective is to draw reliable inferences or conclusions about the larger population from which the data was collected, using **STDEV.S** is the appropriate and methodologically sound choice, as it provides an unbiased estimate of the population variability.

Thus, we almost always use **STDEV.S** to calculate the standard deviation of a dataset because our dataset typically represents a sample. Unless you are absolutely certain that your data constitutes the complete universe of observations relevant to your inquiry, you should default to **STDEV.S**. This function acknowledges the uncertainty of generalization and is the standard practice for inferential statistics. Note that **STDEV** and **STDEV.S** return the exact same values, so we can use either function to calculate the sample standard deviation of a given dataset, though **STDEV.S** is the preferred modern nomenclature.