

# Understanding Standard Deviation in Excel: A Guide to STDEV.P and STDEV.S

Authored by  
**Mohammed looti**

November 6, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Understanding Standard Deviation in Excel: A Guide to STDEV.P and STDEV.S*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11537>

In the realm of statistics and quantitative [data analysis](#), the concept of [standard deviation](#) is paramount. It serves as the definitive measure of data dispersion, illustrating how widely individual values deviate from the mean. While spreadsheet tools like **Microsoft Excel** offer powerful functions to compute this metric, analysts frequently encounter confusion regarding the appropriate choice between **STDEV.P** and **STDEV.S**. The distinction between these two functions is not merely semantic; it fundamentally depends on whether your dataset represents the **entire statistical population** or merely a **representative sample** drawn from that population.

To ensure the statistical validity of your results, understanding Excel's offerings is critical. The software provides three primary functions for calculating standard deviation, each with a specific application:

**STDEV.P:** This function is exclusively used when the data includes the **entire population** of interest.

**STDEV.S:** This function is the standard choice when the data available is only a **sample** used to estimate the characteristics of the larger, unseen population.

**STDEV:** This is a legacy function retained for backward compatibility, which is mathematically identical to **STDEV.S**.

## Calculating Dispersion for the Entire Population: STDEV.P

The **STDEV.P** function calculates the true standard deviation when the input data range comprises the entirety of the [population](#) under study. Utilizing this function is an explicit assertion that you possess every single possible observation relevant to your analysis. Because all data points are present, there is no inherent uncertainty or need for estimation regarding the population's true mean ( $\mu$ ) or its overall variability.

This formula yields the precise, unbiased measure of spread around the true population mean. Crucially, the calculation utilizes the full population size, **N**, in the denominator. This ensures the resulting standard deviation is an exact measure of the population's dispersion, rather than an estimate.

The mathematical expression that **STDEV.P** employs is defined as follows, dividing the sum of squared differences by the population size (N):

$$\text{Population standard deviation} = \sqrt{\sum (x_i - \mu)^2 / N}$$

The variables within this formula represent the following statistical components:

$\Sigma$ : The Greek symbol indicating the operation of summation (the sum of all terms).

**$x_i$** : An individual data point or observation within the full population dataset.

$\mu$ : The true **population mean** (the calculated average of all data points).

**N:** The total count of observations, representing the full size of the population.

## Estimating Variability from a Subset: STDEV.S

In most practical applications of [statistical inference](#), gathering data for the entire population is infeasible. Analysts typically rely on a smaller, representative subset, known as a [sample](#). The **STDEV.S** function is specifically engineered for this common scenario, calculating the standard deviation based on the sample data to provide an unbiased estimate of the variability present in the larger source population.

The fundamental difference in the **STDEV.S** computation lies in its denominator: it uses  $n-1$  instead of the full sample size,  $N$  (or  $n$ ). This critical adjustment is known as [Bessel's Correction](#), and its purpose is to counteract the inherent tendency of sample variability to underestimate the true population variability. By reducing the denominator, the resulting standard deviation is slightly inflated, yielding a statistically more accurate estimate.

The structure of the formula used by **STDEV.S** reflects this correction, utilizing the sample size minus one in the denominator:

$$\text{Sample standard deviation} = \sqrt{\sum (x_i - \bar{x})^2 / (n-1)}$$

When calculating the sample standard deviation, the components are defined as:

$\Sigma$ : The summation symbol.

$x_i$ : The  $i$ th observed value in the sample dataset.

$\bar{x}$ : The **sample mean** (the average of the observed values, used as an estimate for the population mean).

$n$ : The total number of observations contained within the [sample](#).

It is important to recognize that the use of the sample mean ( $\bar{x}$ ) in the numerator introduces the uncertainty that Bessel's Correction (the  $n-1$  denominator) is designed to address.

## A Note on Backward Compatibility: The Legacy STDEV Function

Before Microsoft standardized its statistical functions to align with modern best practices, the primary function for standard deviation calculation was simply **STDEV**. To promote clarity and ensure users explicitly define their data context (population or sample), Excel introduced the explicit names **STDEV.P** and **STDEV.S**.

The older **STDEV** function is mathematically equivalent to **STDEV.S**, meaning it calculates the sample standard deviation using the  $n-1$  denominator adjustment. Although **STDEV** is maintained in Excel for backward compatibility, its use is strongly discouraged in new analyses. Modern

statistical practice dictates using **STDEV.S** to unambiguously signal that the calculated result is an estimate based on a [sample](#), thus improving the transparency and accuracy of the spreadsheet model.

Running a calculation over the same range of data using both **STDEV** and **STDEV.S** will always produce the exact same numerical output, confirming their statistical identity.

## Technical Deep Dive: The Statistical Necessity of N vs. N-1

The choice between dividing by **N** (population size) or **n-1** (sample size minus one) is the heart of the statistical difference between these functions. This distinction is rooted in the concept of statistical [bias](#) and the calculation of [degrees of freedom](#).

When the standard deviation is calculated using the population formula (dividing by  $N$ ), the resulting value will inherently be mathematically smaller than the value derived from the sample formula (dividing by  $n-1$ ) for the same dataset. This difference highlights the fundamental uncertainty introduced when working with partial data.

### Understanding Bias and Degrees of Freedom

When we analyze the entire [population](#), we know the **true mean** ( $\mu$ ). Since every possible data point is accounted for, the calculation of dispersion is exact and unbiased. There is no element of estimation required, and the measurement of variance is precise.

Conversely, when we use a sample, two key factors introduce uncertainty. Firstly, we must estimate the true population mean using the **sample mean** ( $\bar{x}$ ). Because the sample mean is calculated directly from the observed data points, those points naturally cluster closer to their own sample mean than they would to the (usually unknown) true population mean. If we were to calculate the standard deviation using  $N$ , this effect would cause the standard deviation to be systematically underestimated, resulting in a biased estimate.

To correct for this inherent downward bias and account for the uncertainty associated with estimating the mean, statisticians utilize  $n-1$ , which represents the [degrees of freedom](#). By dividing the sum of squares by a smaller number ( $n-1$ ), we artificially inflate the calculated standard deviation. This adjustment yields a statistically **unbiased estimator** that more accurately reflects the true variability of the larger population.

### Practical Application: A Side-by-Side Excel Example

To solidify the theoretical differences, examining a practical example in Excel clearly demonstrates

the numerical divergence between these two functions. Consider a small, defined dataset of five numerical values. We will apply both the population and sample standard deviation functions to this identical range of cells.

Assume the following dataset has been entered into an Excel column:

	A	B	C	D	E
1	<b>Dataset</b>				
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					
27					

When we input the respective formulas, the resulting values clearly illustrate the mathematical impact of the denominator (N vs. n-1) adjustment:

	A	B	C	D	E	F
1	<b>Dataset</b>				Formula	
2	4		<b>STDEV.P</b>	8.896067	=STDEV.P(A2:A21)	
3	5		<b>STDEV.S</b>	9.127172	=STDEV.S(A2:A21)	
4	5		<b>STDEV</b>	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 results confirm the statistical theory: the sample standard deviation (calculated via **STDEV.S**) is **9.127**, whereas the population standard deviation (calculated via **STDEV.P**) is **8.896**. This minor yet significant difference is entirely attributable to the use of [Bessel's Correction](#) in the sample calculation, which ensures the estimate is unbiased.

## The Decision Matrix: When to Select STDEV.P or STDEV.S

Selecting the correct standard deviation function is paramount to accurate statistical reporting. The decision hinges entirely on how your data was collected and what it purports to represent:

**Choose STDEV.P (Population):** Use this function if your dataset includes absolutely every single element or observation relevant to your study scope. Examples include: total sales figures for a specific, completed fiscal year; the ages of every employee in a small company; or the final scores of every student in a single defined class.

**Choose STDEV.S (Sample):** Use this function if your data is a selection, subset, or random draw used to extrapolate or infer characteristics about a larger, often inaccessible, group. Examples include: using a market survey of 500 individuals to estimate the opinions of a million potential customers, or analyzing a small batch of manufactured parts to assess the quality of an entire production run.

In the vast majority of real-world [data science](#), modeling, and business analysis tasks, analysts are dealing with samples. Collecting comprehensive data on an entire [population](#) is often too time-consuming, expensive, or physically impossible.

Consequently, **STDEV.S** is typically the default, statistically appropriate function you should rely upon. It provides the most robust and unbiased estimate of population variability when working exclusively with sample data.