

# Learn How to Calculate Sample and Population Variance in Excel

Authored by  
**Mohammed looti**

November 2, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *Learn How to Calculate Sample and Population Variance in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=8696>

The **variance** is a foundational metric within descriptive statistics, serving as an indispensable tool for quantifying the spread, dispersion, or variability of values contained within a **dataset**. Fundamentally, variance calculates the average of the squared differences between each individual data point and the **mean** of the entire distribution. By squaring these deviations, the calculation ensures that both positive and negative deviations from the mean contribute equally to the measure of spread, and simultaneously gives greater weight to extreme outliers. Understanding variance is critical because it provides analysts with insight into how tightly clustered or widely scattered the data points are relative to their central tendency. A high variance value signals that data points are widely distributed across the range, suggesting high volatility or heterogeneity, whereas a low variance indicates that the values cluster closely around the mean, implying high consistency and homogeneity.

While the concept of variance is singular, its calculation must differentiate based on whether the data represents a complete collection or merely a subset. Statistical rigor demands precision in calculating variance, which means selecting the correct methodology--either for a statistical population or a sample--is paramount. Fortunately, powerful computational tools like **Excel** provide specialized functions that automate these complex statistical computations, allowing users to rapidly determine accurate variance figures without the need for manual, error-prone calculations, provided the user understands the critical distinction between population and sample data.

## Understanding Population Variance Calculation

The first step in accurate statistical analysis is recognizing the scope of your data. The calculation for **population variance** (often symbolized as  $\sigma^2$  or  $\sigma^2$ ) is employed exclusively when the available dataset includes every single element or observation that constitutes the entire group of interest. In this ideal scenario, the resulting variance is a true parameter, reflecting the exact spread of the whole **population**, rather than an estimate. This calculation is used when studying specific, finite groups, such as the height of every student in a single classroom or the sales figures for every store owned by a company in a specific quarter. Since all possible values are included, there is no need to adjust for sampling error or uncertainty.

The fundamental mathematical formula used to calculate the **population variance** is derived directly from the definition of variance itself. It involves three primary steps: finding the deviation of each point from the population mean, squaring those deviations, and finally averaging the squared results across all observations. This averaging process is achieved by dividing the total sum of squared deviations by the total count of elements within the population.

The formula used to calculate the **population variance** ( $\sigma^2$ ) is expressed formally as follows:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

To properly apply and interpret this formula, it is essential to define the role of each component, which together represent the rigorous process of quantifying the population's dispersion:

$\Sigma$ : This Greek symbol (Sigma) represents the operation of [summation](#), indicating the requirement to sum or aggregate all the calculated squared differences across the entire population.

$\mu$ : Denotes the [Population mean](#), which is the true arithmetic average of all values in the population. It acts as the central reference point from which all deviations are measured.

$x_i$ : Refers to the  $i$ th individual element or data point taken from the total population, representing a specific observation within the set.

$N$ : Represents the total size or count of elements within the entire [population](#) being studied. This denominator is used because we possess all data points, providing an exact measure of spread.

## Deriving Sample Variance

In the vast majority of real-world statistical applications, it is impractical, if not impossible, to collect data for every single element of the population. When researchers must work with a smaller, manageable subset of a larger group, they must calculate the **sample variance**. This calculation is used specifically when the data collected is only a [sample](#) drawn from a much larger, often theoretical or inaccessible, population. The primary objective of calculating sample variance ( $s^2$ ) is to use the limited data available to produce the best possible estimate of the true, but unknown, population variance ( $\sigma^2$ ).

The structure of the sample variance formula closely mirrors that of the population variance, as both measure the average squared deviation from the mean. However, a crucial statistical adjustment must be made to the denominator to ensure the estimate is unbiased. If we were to divide by the sample size ( $n$ )--similar to the population calculation--the resulting variance would consistently underestimate the true population variance, especially for small samples. This phenomenon occurs because the sample mean ( $\bar{x}$ ) is calculated from the sample itself, inherently fitting the data better than the true population mean ( $\mu$ ) would, thus minimizing the deviations.

To correct this inherent bias and provide an unbiased estimate, the denominator is adjusted to  $(n-1)$ . This value, known as the [degrees of freedom](#), accounts for the constraint imposed by using the sample mean in the calculation. Conceptually, losing one degree of freedom means that once the sample mean is fixed, only  $n-1$  values in the sample are truly free to vary. Using this adjusted divisor ensures that the resulting sample variance is slightly larger, making it a more statistically robust and reliable estimate of the population's true spread.

The formula used to calculate the **sample variance** ( $s^2$ ) is:

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

The variables used in the sample variance calculation are defined carefully to reflect that they are statistics derived from a subset of data:

**x**: Represents the Sample mean, calculated from the specific subset of data collected. This serves as the temporary central point for deviation measurement.

**xi**: Refers to the *i*th individual element or observation within the specific [sample](#) being analyzed.

**n**: Denotes the total Sample size, which is the count of observations included in the subset.

**(n-1)**: This is the critical [degrees of freedom](#) term, used as the denominator to ensure the resulting sample variance is an unbiased estimator of the population variance.

## Leveraging Excel Functions for Variance Calculation

While understanding the statistical formulas is essential, the manual calculation of variance--especially involving large [datasets](#)--is highly time-consuming and inherently prone to computational errors. Recognizing the need for efficiency in data analysis, [Excel](#), the industry standard spreadsheet software, provides robust, built-in functions specifically designed to streamline the variance calculation process. These functions eliminate the need for users to manually calculate deviations, square differences, perform [summation](#), or manage the complex division steps required by the underlying statistical equations.

The key to utilizing Excel effectively is selecting the appropriate function based on the data's nature. Excel clearly separates the functions corresponding to the two types of variance we have discussed:

We utilize the [VAR.S\(\)](#) function to calculate the sample variance. The 'S' explicitly denotes 'Sample,' confirming that this function employs the critical  $(n-1)$  in its denominator, thereby producing an unbiased estimate suitable for subsets of data. Conversely, the [VAR.P\(\)](#) function is used to calculate the population variance. The 'P' stands for 'Population,' and this function uses the total count  $(N)$  in its denominator, yielding the true variance for the entire dataset when all elements are known. It is absolutely critical for statistical validity to select the correct function; using VAR.S when VAR.P is appropriate, or vice versa, will lead to systematically inaccurate results.

The structure for invoking these functions is remarkably simple: the user only needs to specify the range of cells containing the numeric data. Excel handles the rest of the complex mathematics behind the scenes, providing the result almost instantaneously. The subsequent practical examples illustrate the precise application of both VAR.S() and VAR.P() within a standard Excel worksheet environment using identical input data to highlight the numerical difference resulting from the change in the denominator.

## Example 1: Calculating Sample Variance using VAR.S()

In this inaugural demonstration, we establish a foundational assumption: the collection of numerical values listed within Column A represents only a [sample](#), meaning it is a statistically relevant subset drawn from a hypothetical, much larger population whose parameters we wish to estimate. Because the data is merely a sample, we must employ the function specifically designed for sample estimation, **VAR.S()**. This choice ensures that the calculation correctly incorporates the crucial [degrees of freedom](#) adjustment, which involves dividing the sum of squared deviations by  $(n-1)$  to achieve an unbiased estimate of the population variance.

The following screenshot provides a visual confirmation of the precise syntax required to invoke the function. If the data points occupy cells A1 through A10, the required formula is straightforward:

```
=VAR.S(A1:A10)
```

. This syntax references the entire range of data points required for the computation, effectively instructing [Excel](#) to execute all the required steps--calculating the sample mean, finding squared deviations, summing them, and finally dividing by  $(10-1)$ , or 9, in this instance.

	A	B	C	D	E	F	G
1	<b>Data</b>		<b>Sample Variance</b>				
2	4		35.2079				
3	8						
4	12						
5	15						
6	9						
7	6						
8	14						
9	18						
10	12						
11	9						
12	16						
13	17						
14	17						
15	20						
16	14						
17	19						
18	22						
19	24						
20	26						
21	19						
22							
23							
24							
25							

Upon successful execution of the **VAR.S()** function, the calculated **sample variance** for this specific dataset is determined to be **35.2079**. This quantified value represents the average squared deviation of the sample points from their calculated sample [mean](#). This figure is interpreted as the best statistical estimate of the true variability within the larger population from which the sample was drawn.

### Example 2: Calculating Population Variance using VAR.P()

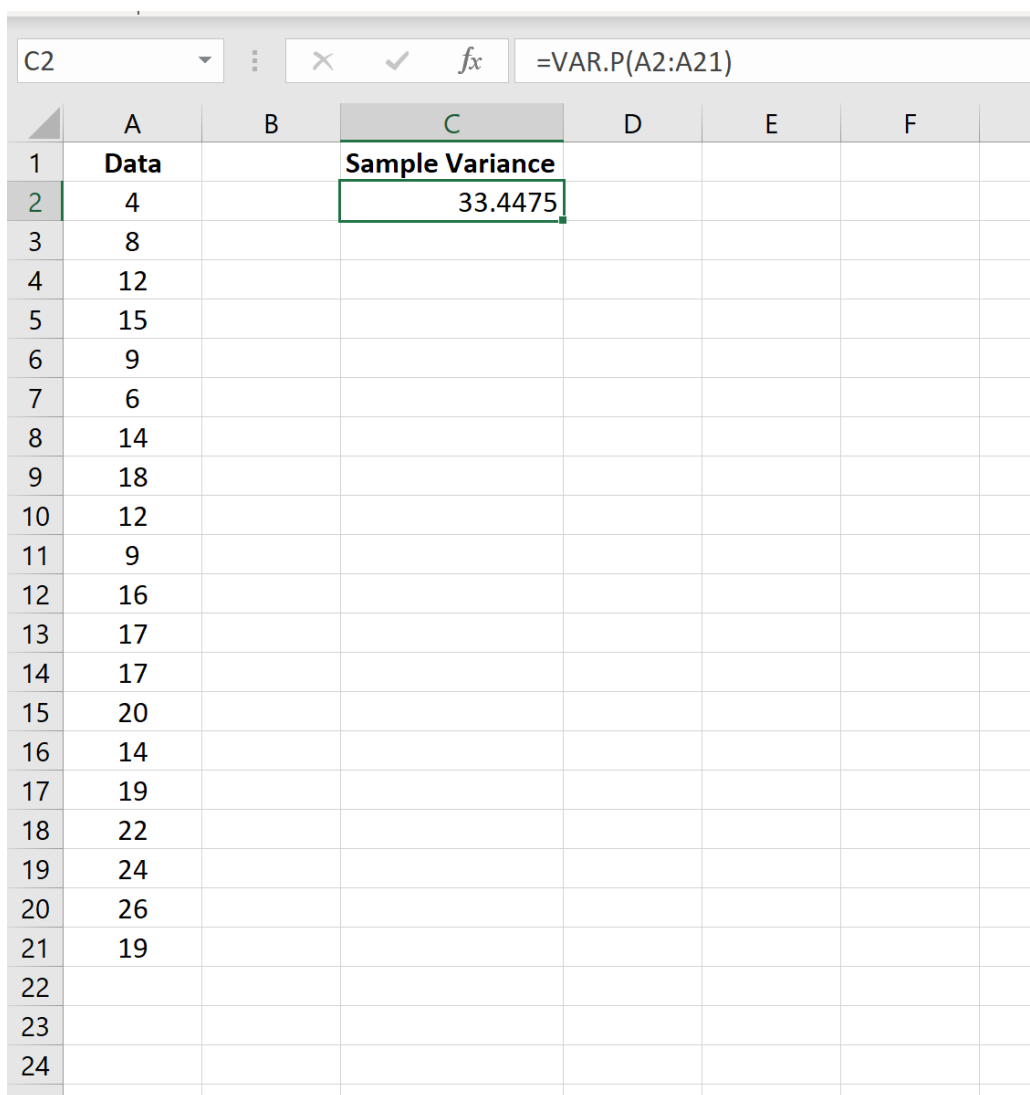
In stark contrast to the preceding example, this scenario fundamentally alters the statistical assumption: we now proceed under the premise that the data contained within Column A represents the complete and total [population](#)--implying that we possess every single data point relevant to our inquiry. Given this comprehensive scope, we are calculating a population parameter rather than estimating a statistic. Consequently, we must utilize the **VAR.P()** function, which correctly divides the sum of squared differences by the total population size  $(N)$ , providing an

exact measure of the population's true [variance](#) without any adjustments for estimation uncertainty.

The screenshot below illustrates the application of the **VAR.P()** function to the exact same range of values (A1:A10) used in the previous example. The formula syntax,

=VAR.P(A1:A10)

, maintains an identical structure to VAR.S(), emphasizing that the primary difference lies entirely in the underlying mathematical operation performed by [Excel](#)--specifically, the divisor used in the final step. In this case, the sum of squared deviations is divided by  $N=10$ , not  $n-1=9$ .



The screenshot shows an Excel spreadsheet with the following data and formula:

	A	B	C	D	E	F
1	<b>Data</b>		<b>Sample Variance</b>			
2	4		33.4475			
3	8					
4	12					
5	15					
6	9					
7	6					
8	14					
9	18					
10	12					
11	9					
12	16					
13	17					
14	17					
15	20					
16	14					
17	19					
18	22					
19	24					
20	26					
21	19					
22						
23						
24						

The formula bar at the top shows the formula: =VAR.P(A2:A21)

The resulting calculation yields a **population variance** of **33.4475**. When compared directly to the sample variance of 35.2079 derived from the same data, this figure is noticeably lower. This subtle but significant numerical difference is a direct result of the statistical definitions: the population

variance calculation averages the squared deviations over the total count  $(N)$ , while the sample variance applies the [degrees of freedom](#) correction  $(n-1)$ , inflating the result slightly to account for the uncertainty inherent in sampling.

## Critical Considerations for Choosing Variance Type

The most pivotal decision in calculating variance is the initial determination of whether the input data represents a [population](#) or a [sample](#). Misidentifying the scope of your data source--for instance, treating a sample as if it were the entire population--leads to systematically inaccurate estimations of data spread and can undermine the validity of subsequent statistical inferences, such as confidence intervals or hypothesis testing. Therefore, before inputting data into [Excel](#), analysts must rigorously confirm the definition and scope of their [dataset](#).

To maintain statistical integrity, adhere strictly to the following fundamental rules of thumb when applying these calculations in Excel:

You must calculate the **population variance** using the [VAR.P\(\)](#) function exclusively when the dataset encompasses every single value that is relevant to your inquiry. This implies that the data itself constitutes the complete population of interest, and the goal is to define its exact parameter.

You must calculate the **sample variance** using the [VAR.S\(\)](#) function when the data you are analyzing represents a smaller, representative subset or sample taken from a much larger, often inaccessible population. The output is an unbiased statistic intended to estimate the true population parameter.

It is a statistical certainty, resulting from the use of the [degrees of freedom](#) adjustment, that the sample variance calculated for a given set of numbers will always be marginally larger than the population variance calculated for the exact same set of numbers. This intentional inflation in the sample calculation inherently builds in a correction factor, accounting for the increased uncertainty and variability inherent in using a sample to estimate a larger population's characteristics.

## Additional Resources for Statistical Analysis in Excel

Variance is but one component of a complete statistical picture. For those interested in exploring related measures of spread and central tendency that complement the [variance](#) calculation, the following tutorials provide guidance on other essential statistical calculations within [Excel](#), enabling a deeper dive into data distribution and analysis: