

How to Easily Calculate Variance in Excel: Understanding VAR.P vs. VAR.S

Authored by
stats writer

December 5, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Calculate Variance in Excel: Understanding VAR.P vs. VAR.S*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=105826>

Calculating statistical variance is a fundamental step in data analysis, providing a measure of how spread out a set of data points is relative to its mean. In Excel, the process is streamlined through specialized functions. However, users often encounter confusion when choosing between the two primary variance formulas: **VAR.P** and **VAR.S**. While both yield a variance figure, they are designed for fundamentally different statistical scenarios, and using the wrong one can lead to significant inaccuracies in your analytical conclusions.

The distinction between **VAR.P** and **VAR.S** hinges entirely on whether your dataset represents the entire measurable **population** or merely a **sample** drawn from that population. **VAR.P** (Population Variance) is appropriate when you have every possible data point (e.g., the test scores of all 100 students in a specific course). Conversely, **VAR.S** (Sample Variance) is the function you should almost always default to, as most real-world datasets--like surveying 500 customers out of a million--are inherently samples, not entire populations.

Understanding this statistical dichotomy is not just academic; it dictates which mathematical formula Excel applies, specifically impacting the denominator in the variance calculation. This seemingly small difference--dividing by N (total observations) versus $n-1$ (degrees of freedom)--is crucial for ensuring that your statistical estimate is an unbiased estimator of the true population variance. This guide will meticulously dissect these differences, provide the underlying formulas, and illustrate when each function must be employed for robust data analysis.

The Fundamental Difference: Population vs. Sample

The statistical theory underpinning the choice between **VAR.P** and **VAR.S** relies on defining the scope of your data collection. A **population** includes all items or individuals of interest, constituting the complete set of data you wish to draw conclusions about. When you possess data for the entire population, the calculated variance is the precise and true measure of dispersion for that group. There is no uncertainty or estimation involved, as every data point is accounted for.

In contrast, a sample is a manageable subset of observations selected from the larger population. Because the sample does not contain all data points, any variance calculated from it is merely an **estimate** of the true population variance. Due to this inherent uncertainty, the formula must be adjusted to account for the fact that a sample mean is highly likely to be closer to the sample data points than the true population mean would be. This necessary adjustment leads directly to the core mathematical difference between the two Excel functions.

When analyzing data, it is imperative to first categorize your dataset accurately. If your dataset represents every single observation relevant to your study goal--whether it's the heights of all employees in a small company or the total sales figures for a specific quarter--then you are dealing with a population. If, however, your dataset is a random selection used to infer properties of a

larger, inaccessible group (such as market research data or quality control testing on a batch of manufactured goods), it is a sample, necessitating the use of the appropriate sample statistics.

Deep Dive into the VAR.P Function (Population Variance)

The **VAR.P** function is specifically designed to calculate the population variance. You should utilize this function exclusively when you are certain that the array of values you are analyzing encompasses the entire statistical population of interest. Using **VAR.P** implies that you have perfect knowledge of the data set, meaning your calculated variance is definitive, not an estimation.

This function calculates the average of the squared differences from the population mean. Because we know the entire population, we divide the sum of squared deviations by the total count of observations, N . This calculation provides the exact measure of dispersion for the entire group.

The **VAR.P** function employs the following precise mathematical formula:

$$\text{Population variance} = \frac{\sum(x_i - \mu)^2}{N}$$

where the statistical components are defined as follows:

Σ : This is the Greek symbol Sigma, which signifies the operation of summation across all observations.

x_i : Represents the i th individual value, or observation, within the dataset.

μ : Denotes the **population mean** (μ), the arithmetic average of all values in the population.

N : Represents the **total number of observations** in the complete population.

Analyzing the VAR.S Function (Sample Variance and Unbiased Estimation)

The **VAR.S** function is used to calculate the sample variance. This is the far more common scenario in practical analysis, where data is collected from a subset of a larger group. Because a sample cannot perfectly represent the entire population, the variance derived from it must be adjusted to serve as an **unbiased estimator** of the true population variance.

If we were to simply use the population formula (dividing by n), the resulting variance would systematically underestimate the true population variance. This occurs because the sample mean (\bar{x}) naturally minimizes the sum of squared deviations for that specific sample, making the calculated spread artificially smaller than the population's true spread. To correct this bias, the formula uses $n-1$ in the denominator, a concept known as the correction for **degrees of freedom**.

The **VAR.S** function employs the following formula, which incorporates the adjustment necessary for sample data:

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

where the key statistical elements are defined as follows:

Σ : The summation symbol, indicating the total sum of the squared deviations.

x_i : The i th individual value in the analyzed sample dataset.

\bar{x} : Denotes the **sample mean**, the arithmetic average of the values within the sample.

n : Represents the **total number of observations** in the specific sample.

The Critical Role of Degrees of Freedom ($n-1$)

The use of $n-1$ in the denominator of the sample variance formula is a crucial statistical concept often referred to as the degrees of freedom (df). When calculating the variance for a sample, we must first estimate the population mean using the sample mean (\bar{x}). Because the sample mean is derived from the same data used to calculate the deviations, one degree of freedom is lost.

To visualize this, imagine a sample of 5 numbers ($n=5$). If you know the sample mean and 4 of the 5 numbers, the value of the final fifth number is mathematically constrained--it must be whatever value makes the mean correct. Since one observation is constrained by the calculation of the mean, only $n-1$ observations are truly free to vary. Dividing by $n-1$ instead of n effectively increases the estimated variance, compensating for the fact that the sample mean inherently minimizes the squared deviations, thereby providing a more accurate (unbiased) estimate of the larger population's true variance.

This adjustment ensures that, over many repeated samples, the average of the sample variances calculated using **VAR.S** will converge precisely on the true population variance. If **VAR.P** were incorrectly used on a sample dataset, the resulting variance estimate would be systematically biased downwards, leading to overly optimistic conclusions about the precision or consistency of the population data.

Technical Note on Magnitude:

Since the formula for the population variance (**VAR.P**) divides by N , and the sample variance (**VAR.S**) divides by the smaller value $n-1$ (assuming $N=n$), the population variance will mathematically always yield a result that is smaller than the sample variance calculated on the same dataset.

This difference in magnitude reflects the certainty of the calculation. When we possess data for the entire population, we have absolute certainty about the true center (μ) and spread of the data. However, when working only with a sample, there is greater uncertainty surrounding the

population's true parameters. To reflect this lack of perfect knowledge, the unbiased estimate of the sample variance must be slightly larger, acknowledging the potential for greater spread in the overall population we did not observe.

Comparing Excel's Variance Functions: VAR, VAR.S, and VAR.P

In addition to **VAR.P** and **VAR.S**, older versions of Excel included the function **VAR**. It is important to understand the relationship between these three functions for compatibility and statistical accuracy.

The **VAR** function is a legacy command retained in modern Excel primarily for backward compatibility with spreadsheets created in Excel 2007 and earlier. Crucially, the **VAR** function calculates the sample variance, meaning it uses the $n-1$ adjustment. Therefore, the output of **VAR** is mathematically identical to the output of **VAR.S** when applied to numerical data. Microsoft introduced **VAR.S** and **VAR.P** (and other statistical functions like STDEV.S and STDEV.P) to clearly differentiate between sample-based and population-based calculations, enhancing clarity and reducing potential statistical errors.

For all new analysis and reporting, it is strongly recommended to abandon the use of the generic **VAR** function. Instead, explicitly use **VAR.P** if you are certain you have the entire population, or **VAR.S** if your data is a sample--which, as previously noted, is the standard practice in most analytical contexts. This naming convention ensures transparency in your methodology and confirms whether you are calculating a parameter (population variance) or an unbiased estimator (sample variance).

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

To illustrate the practical effect of the N versus $n-1$ denominator, consider a small, representative dataset. Suppose we have the following five values entered into cells in an Excel sheet, and we calculate both the population variance and the sample variance using the respective functions.

	A	B	C	D	E	F	G
1	Dataset						
2	3						
3	5						
4	5						
5	6						
6	6						
7	7						
8	8						
9	10						
10	12						
11	12						
12	13						
13	14						
14	15						
15	19						
16	22						
17	24						
18	27						
19	27						
20	28						
21	29						
22							
23							
24							

When applying the two functions to this set of numbers, the results will visibly demonstrate the predictable difference in magnitude. The calculation of **VAR.S**, designed as the unbiased estimate, will always yield a higher result than **VAR.P**, which assumes perfect knowledge of the entire group.

	A	B	C	D	E	F
1	Dataset				Formula	
2	3		VAR.P	73.14	=VAR.P(A2:A21)	
3	5		VAR.S	76.99	=VAR.S(A2:A21)	
4	5		VAR	76.99	=VAR(A2:A21)	
5	6					
6	6					
7	7					
8	8					
9	10					
10	12					
11	12					
12	13					
13	14					
14	15					
15	19					
16	22					
17	24					
18	27					
19	27					
20	28					
21	29					
22						
23						
24						

As depicted in the final results of this calculation, the sample variance (using **VAR.S**) is calculated to be **76.99**, while the population variance (using **VAR.P**) yields a smaller result of **73.14**. This numerical difference is entirely attributable to the use of $n-1$ (which is 4 in this case) versus N (which is 5) in the denominator. The **VAR.S** result is the statistically rigorous figure to report if these five observations were taken as a subset of a much larger group.

Determining the Right Function for Your Data

The crucial decision point for any analyst is determining whether their dataset constitutes a true population or a sample. Statistically speaking, if the data gathering is exhaustive and includes every single member of the group you are studying, you must use **VAR.P**. Examples include the inventory of a small store, the grades of all students in a single classroom, or the temperature readings from all sensors in a controlled environment over a defined period.

However, in nearly all scenarios involving large-scale data, inferential statistics, or predictive modeling, we are dealing with a sample. It is often impossible or impractical to collect data from the entire population, such as all potential customers, all manufactured parts, or all possible outcomes

of an event. Because the overwhelming majority of statistical analysis aims to draw conclusions about a larger group based on a subset, **VAR.S** is the standard default function used in professional analysis and academic research utilizing Excel.

Therefore, unless you have explicit confirmation that your dataset is exhaustive of the target population, the recommended best practice is always to use **VAR.S**. This choice ensures that you are utilizing the unbiased estimator for variance, maximizing the robustness and statistical validity of your analysis and conclusions regarding the larger population parameters. Remember that the older **VAR** function is functionally identical to **VAR.S**, but **VAR.S** provides superior clarity regarding the methodology employed.

ARABPSYCHOLOGY.COM