

VAR.P vs. VAR.S in Excel: A Comprehensive Guide to Variance Calculation

Authored by
Mohammed loot

November 4, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *VAR.P vs. VAR.S in Excel: A Comprehensive Guide to Variance Calculation*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10039>

Understanding Variance: The Foundation of Data Dispersion

The concept of [variance](#) is one of the most fundamental metrics in descriptive statistics, serving as the cornerstone for understanding data distribution. Essentially, variance provides a quantifiable measure of how widely a set of data points deviates or spreads out from its central tendency, specifically the mean. A low variance indicates that data points tend to be very close to the mean, while a high variance suggests that the data points are spread far apart, indicating greater volatility or heterogeneity within the dataset. Mastering the accurate calculation and interpretation of variance is indispensable for fields ranging from financial modeling and risk assessment to quality control and advanced predictive data analysis.

When working within Microsoft [Excel](#), the industry-standard spreadsheet application, analysts are provided with specialized functions to calculate this crucial statistical measure. However, the seemingly minor differences between these functions carry profound statistical implications. The choice of function--whether **VAR.P** or **VAR.S**--is entirely dependent on the nature of the data being analyzed: specifically, whether the data represents the complete universe of observations (the entire statistical [population](#)) or merely a representative subset (a [sample](#)) drawn from that universe.

This guide focuses on differentiating the two modern, standardized functions: **VAR.P** (Population Variance) and **VAR.S** (Sample Variance). We will also briefly address the legacy function, **VAR**. It is imperative that analysts recognize the subtle but statistically critical distinction between population-based and sample-based calculations, as misusing these functions can lead to significantly flawed conclusions regarding data variability and risk.

VAR.P: Calculating the True Population Variance

The **VAR.P** function is specifically engineered for situations where the provided data range includes every single observation relevant to the study. In statistical parlance, this comprehensive dataset constitutes the entire [population](#). Using **VAR.P** implies complete certainty that no relevant data points have been excluded, meaning the resulting variance is the true, definitive measure of dispersion for that group, not merely an estimate. Examples of such scenarios include analyzing the test scores of every student in a single, defined classroom, or calculating the variability in the height of all employees working at a specific, small company.

When calculating population variance, the method is straightforward: the sum of the squared differences between each data point (x_i) and the true population mean (μ) is calculated. This total is then divided by the total number of observations, symbolized by N . Because we possess all data points and the true mean, this calculation yields an exact measure of the population's variability, requiring no statistical adjustments or corrections.

The calculation performed by the **VAR.P** function adheres to the following standard statistical

formula:

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

The symbols utilized in this mathematical expression represent the key components necessary to derive the exact population variance:

Σ : The Greek symbol Sigma, which universally represents the operation of summation, instructing us to sum all subsequent terms.

x_i : Represents the i th individual observation or specific data point within the comprehensive dataset.

μ : Mu, which denotes the true population mean. This is the arithmetic average calculated from all values within the defined population.

N : Represents the total count of observations, signifying the exact size of the entire statistical population.

VAR.S: Estimating Variance from Sample Data

In the overwhelming majority of practical data analysis scenarios--whether in scientific research, business intelligence, or economics--it is either impossible, prohibitively expensive, or extremely time-consuming to collect data for the entire target population. Consequently, analysts rely on analyzing a carefully selected, representative subset of the data, which is formally known as a [sample](#). The **VAR.S** function is the appropriate tool for calculating the variance of this sample data, which is used to generate an unbiased estimate of the true, unknown population variance.

Since a sample is inherently incomplete and is subject to sampling error, it usually underestimates the true variability present in the larger [population](#). If we were to calculate the variance exactly as we do for a population (dividing by n), the result would be a biased estimator--meaning it would consistently undervalue the true population variance. To correct for this inherent bias and produce an unbiased estimate, a crucial statistical adjustment must be implemented. This adjustment is known as Bessel's correction.

Bessel's correction modifies the denominator used in the calculation, substituting the sample size (n) with $(n-1)$. This slight, yet critical, change ensures that the calculated sample variance is statistically sound and provides a more conservative, robust estimate of the population's spread.

The **VAR.S** function executes this specific calculation using the following formula, incorporating Bessel's correction:

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

The definition of the formula components highlights the reliance on sample statistics rather than

true population parameters:

Σ : The Greek symbol for summation.

x_i : The i th individual observation taken from the sample dataset.

\bar{x} : X-bar, which represents the sample mean. This is the arithmetic average calculated only from the observed values within the sample.

$n-1$: This term signifies the sample size (n) minus one. This crucial adjustment accounts for the loss of one [degree of freedom](#), resulting in an unbiased estimator.

The Critical Statistical Distinction: Degrees of Freedom

The core difference between the **VAR.P** and **VAR.S** functions boils down entirely to their denominators: N for population variance and $(n-1)$ for sample variance. This distinction is not merely mathematical but reflects the fundamental challenge of statistical inference and the concept of [degrees of freedom](#). Understanding why we use $(n-1)$ in the sample calculation is essential for any serious data analyst.

When calculating the population [variance](#) using **VAR.P**, we are comparing every data point against the true, known population mean (μ). Since we know the true mean, all N deviations are independent pieces of information, thus we divide by N . The calculation is definitive and exact.

Conversely, when we calculate the variance using a [sample](#) (using **VAR.S**), we do not know the true population mean. We must first use the sample data itself to estimate the mean, resulting in the sample mean (\bar{x}). Once the sample mean is fixed, the last observation in the sample is no longer free to vary, as its value is determined by the requirement that the sum of all deviations must equal zero around the mean. Therefore, we lose one piece of independent information, or one [degree of freedom](#). Dividing by $(n-1)$ corrects for this statistical dependency, ensuring the calculated sample variance is an unbiased estimator of the true population spread.

Technical Note: Why Sample Variance is Always Larger

Because the denominator used in the population formula (the total size, N) is always larger than the denominator used in the sample formula (the sample size minus one, $n-1$), the population variance calculated by **VAR.P** will mathematically yield a smaller value than the sample variance calculated by **VAR.S** when applied to the exact same set of numbers.

This intentional inflation of the sample variance estimate is statistically necessary. When dealing with a sample, there is inherent uncertainty about how accurately it reflects the entire population. Dividing by $n-1$ slightly increases the variance value, compensating for the likely tendency of the sample data to cluster more tightly around its own mean than the population data clusters around the true population mean. This adjustment provides a necessary margin of error and yields a more

accurate, unbiased, and conservative estimate of the variability present in the larger system.

Practical Application: A Side-by-Side Comparison in Excel

To solidify the theoretical differences, examining a practical example in [Excel](#) clearly illustrates the divergence in results produced by these two functions when using identical data. Imagine a scenario where we have a small dataset of 10 observations entered into cells A2 through A11. Whether this data represents a population or a sample dictates the function choice, but for comparison, we calculate both results.

Consider the following dataset, which we will treat first as a complete population, and then as a representative sample:

	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							

To calculate the variance assuming this data represents the entire statistical [population](#), we use the formula: `=VAR.P(A2:A11)`. This calculation utilizes $N=10$ in the denominator. Conversely, if we assume these 10 observations are merely a [sample](#) intended to estimate a larger population, we must use the function: `=VAR.S(A2:A11)`, which applies $n-1=9$ in the denominator.

The outcomes of these two function calls, when executed side-by-side in Excel, provide tangible

evidence of the statistical correction:

	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 the visual comparison confirms, the sample variance calculated using **VAR.S** results in a higher value of **76.99**. In contrast, the population variance calculated using **VAR.P** yields a smaller value of **73.14**. This predictable pattern is directly attributable to Bessel's correction and the inherent statistical requirement to account for the lost [degrees of freedom](#) when working with incomplete data.

Guidelines for Function Selection in Analysis

Selecting the appropriate variance function is not a matter of preference but a critical decision that determines the statistical validity of your analytical findings. Misapplying **VAR.P** when **VAR.S** is required can lead to an artificially deflated measure of risk or variability, resulting in poor decision-making.

We should use **VAR.P** only in very specific, bounded scenarios. This function is appropriate exclusively when you are studying a finite group and have successfully collected data on every single member or unit within that defined boundary. For example, if you are analyzing the

performance of all 5 servers in your data center, or the height of every cherry tree in a small, closed orchard, you have complete census data--a [population](#)--and **VAR.P** is the correct choice.

In contrast, **VAR.S** is the default function for almost all professional statistical work. Whenever your dataset is a subset intended to generalize, infer, or extrapolate characteristics about a much larger, often theoretical, group, you are dealing with a [sample](#). This applies to estimating national unemployment rates, quality checking a small batch of manufactured goods, or conducting market research through a limited survey. Because sampling introduces uncertainty and requires unbiased estimation, the use of **VAR.S** (and its associated $(n-1)$ denominator) is mandatory.

Therefore, the general rule of thumb for robust data analysis is to always default to using **VAR.S** unless you can definitively prove that your data constitutes a complete census of the entire statistical population of interest. By employing **VAR.S**, you ensure that your measure of data [variance](#) is adjusted upward to provide a statistically sound and unbiased estimate of the true population spread.

Addressing Legacy Functions and Best Practices

Prior to the release of [Excel](#) 2010, the variance functions utilized different naming conventions, which often led to confusion among users. The legacy function **VAR** (which calculates sample variance) and **VARP** (which calculates population variance) were the standard. While the new functions, **VAR.S** and **VAR.P**, perform identical calculations, they were introduced to align Excel with modern statistical standards and to explicitly clarify the function's purpose using the suffix (.S for Sample, .P for Population).

The third function occasionally encountered, **VAR**, is retained in the current version of Excel solely for the purpose of backward compatibility. It operates exactly like **VAR.S**, calculating the sample variance using the $n-1$ denominator (Bessel's correction). However, relying on this older naming convention introduces ambiguity that modern analysis aims to eliminate.

For all new spreadsheet development and statistical modeling, the best practice is to exclusively utilize the modern, explicit functions: **VAR.S** and **VAR.P**. This ensures maximum clarity for collaborators, reduces the risk of statistical error based on function misinterpretation, and adheres to contemporary data analysis standards. Should you inherit an older spreadsheet using **VAR** or **VARP**, it is highly recommended to update those formulas to their modern equivalents during any maintenance or review cycle.

Conclusion and Further Statistical Exploration

The distinction between **VAR.P** and **VAR.S** highlights the fundamental differences between descriptive statistics (calculating true population parameters) and inferential statistics (estimating

parameters from a sample). The choice relies entirely on defining your dataset: is it a complete population, or a sample? The vast majority of real-world analysis requires the use of **VAR.S** to ensure the calculated variance is an unbiased estimate of the underlying population's variability, correctly adjusting for the loss of a [degree of freedom](#).

Understanding [variance](#) is often the first step toward deeper statistical insights. Variance is intrinsically linked to the standard deviation--its square root--which is often preferred because it returns the measure of dispersion in the original data units. For those looking to expand their quantitative skill set beyond variance, exploring related metrics is highly recommended.

For further reading on related statistical measures and their applications in data analysis, consider exploring the following essential topics:

Standard Deviation vs. Variance: Understanding the difference in units and interpretability.

The concept of Expected Value: How to calculate the average outcome of a random variable.

Analyzing Data Dispersion with Quartiles and the Interquartile Range (IQR).