

What is the difference between COVARIANCE.P and COVARIANCE.S in Excel?

Authored by
stats writer

November 24, 2025

RECOMMENDED CITATION

stats writer (2025). *What is the difference between COVARIANCE.P and COVARIANCE.S in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=100374>

Understanding the distinction between statistical functions in spreadsheet software is crucial for accurate data analysis. In Excel, two key functions are used to measure the relationship between two datasets: **COVARIANCE.P** and **COVARIANCE.S**. While both calculate covariance (the degree to which two variables change together), their underlying statistical assumptions dictate which one you should use.

The core difference lies in their approach to the dataset: **COVARIANCE.P** calculates the covariance assuming the data represents the entire population (P for Population), taking every value into account. Conversely, **COVARIANCE.S** calculates the covariance assuming the data is merely a sample (S for Sample), utilizing a subset of values to estimate the true population covariance. This subtle but critical distinction impacts the divisor used in the final calculation, leading to different numerical results.

Defining Covariance and Its Interpretation

In the field of statistics, covariance serves as a fundamental metric for assessing the directional linear relationship between two variables. It quantifies how much two random variables change together, often forming the basis for more advanced analyses like correlation and regression.

The sign of the resulting covariance value provides direct insight into the nature of this relationship. A **positive covariance** indicates that as one variable increases, the other variable tends to increase as well, demonstrating a positive linear association. Conversely, a **negative covariance** suggests an inverse relationship: as one variable increases, the other tends to decrease.

It is important to note that while covariance indicates the direction of the relationship, its magnitude is dependent on the units of the variables involved, making it difficult to interpret strength directly. This is why analysts often normalize covariance to obtain the correlation coefficient, which is a standardized, unitless measure of linear strength.

COVARIANCE.P: Calculation for the Entire Population

The **COVARIANCE.P** function in Excel is specifically designed to calculate the true covariance of a dataset when the provided data ranges represent the complete population under study. This calculation is appropriate only in situations where all possible data points for the variable pairs are known, such as analyzing fixed, finite datasets where no further extrapolation is required.

Because it assumes complete knowledge of the data, **COVARIANCE.P** uses the population covariance formula. This formula involves summing the products of the deviations of each point from its respective mean and then dividing the result by the total number of observations, n . This divisor is used because we are calculating the actual population parameter, not an estimate.

The mathematical formula utilized by the **COVARIANCE.P** function is defined as follows:

$$\text{Population covariance} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Where the components of the formula represent:

Σ : The statistical symbol denoting the sum (or aggregation) of the subsequent elements.

x_i : The i th individual observation for the variable X.

\bar{x} : The mean (average) value of the variable X across the entire population.

y_i : The i th individual observation for the variable Y.

\bar{y} : The mean (average) value of the variable Y across the entire population.

n : The total number of paired observations (the size of the population).

COVARIANCE.S: Estimating from a Sample Dataset

The **COVARIANCE.S** function is the function applied in the vast majority of statistical analyses. It calculates the covariance based on the critical assumption that the input data ranges constitute only a sample drawn from a much larger, often theoretical population. Using a sample to infer characteristics about a large population is standard practice when data collection limitations exist.

When calculating sample covariance, the primary objective is to produce an unbiased estimate of the true population covariance. To achieve this, a statistical adjustment, sometimes referred to in this context as Bessel's correction, is employed. This correction requires dividing the sum of the products of the deviations not by n , but by $n-1$, which accounts for the loss of one degree of freedom.

The resulting sample covariance formula utilized by **COVARIANCE.S** is:

$$\text{Sample covariance} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$$

The statistical definitions for the variables are as follows:

Σ : The summation operator.

x_i : The i th observation for variable X within the sample.

\bar{x} : The sample mean for variable X.

y_i : The i th observation for variable Y within the sample.

\bar{y} : The sample mean for variable Y.

$n-1$: The degrees of freedom, used as the divisor to ensure the sample covariance is an unbiased estimator of the population covariance.

Analyzing the Difference in Calculation: n vs. $n-1$

The fundamental structural difference between the two statistical formulas is the denominator: **COVARIANCE.P** divides by n (the total number of observations), while **COVARIANCE.S** divides by $n-1$. This distinction is critically important because when we use sample data, we must estimate the population mean using the sample mean. The sample mean minimizes the deviations for that specific sample, meaning the deviations calculated from the sample mean are inherently smaller than they would be if calculated from the (unknown) true population mean.

By dividing by the smaller value, $n-1$, the result is slightly inflated, mathematically compensating for the inherent tendency of sample data to underestimate the population's true variability. This adjustment, Bessel's correction, ensures that **COVARIANCE.S** provides an unbiased estimate of the population parameter.

Consequently, for any identical dataset, the calculation derived from the **COVARIANCE.S** formula will always yield a value greater than that derived from the **COVARIANCE.P** formula. This consistent disparity highlights the differing assumptions about the scope and completeness of the data being analyzed.

Example: Applying COVARIANCE.P and COVARIANCE.S in Excel

To fully grasp the operational difference between these two functions, let us apply them to a practical dataset within Excel. Consider a dataset charting the Points scored and Assists made for 15 basketball players. Regardless of whether we treat these 15 players as an entire population or a representative sample, we can compare the results directly.

The initial dataset, containing the paired observations for Points and Assists, is structured as follows:

	A	B	C	D	E	F
1	Points	Assists				
2	7	2				
3	7	4				
4	8	4				
5	10	3				
6	12	2				
7	14	3				
8	14	2				
9	15	4				
10	17	6				
11	18	5				
12	22	4				
13	22	5				
14	24	7				
15	29	11				
16	32	8				
17						
18						
19						
20						
21						
22						

We then calculate both the sample covariance and the population covariance using their respective built-in Excel functions, `=COVARIANCE.S(Array1, Array2)` and `=COVARIANCE.P(Array1, Array2)`:

	A	B	C	D	E	F
1	Points	Assists				Formula
2	7	2		COVARIANCE.P	14.64444	=COVARIANCE.P(A2:A16, B2:B16)
3	7	4		COVARIANCE.S	15.69048	=COVARIANCE.S(A2:A16, B2:B16)
4	8	4				
5	10	3				
6	12	2				
7	14	3				
8	14	2				
9	15	4				
10	17	6				
11	18	5				
12	22	4				
13	22	5				
14	24	7				
15	29	11				
16	32	8				
17						
18						
19						
20						
21						
22						
23						

As illustrated by the results, the calculated **sample covariance (COVARIANCE.S)** yields a value of **15.69**, while the corresponding **population covariance (COVARIANCE.P)** results in **14.64**. This numerical outcome confirms the theoretical expectation: the sample covariance is greater than the population covariance due to the utilization of the degrees of freedom adjustment ($n-1$ versus n).

Determining the Correct Function: Population vs. Sample Context

The choice between **COVARIANCE.P** and **COVARIANCE.S** hinges entirely on how the collected data relates to the overall scope of your statistical inquiry. It is essential to correctly define whether your data constitutes an exhaustive population or a representative sample.

In most real-world analytical scenarios, gathering data for the entire population is practically impossible due to constraints like time, cost, and accessibility. Consequently, researchers rely on a smaller, representative subset--the sample. Therefore, if your goal is to infer characteristics about a large, unobserved group, you must use **COVARIANCE.S**, as it provides the statistically unbiased estimate necessary for accurate generalization.

The use of **COVARIANCE.P** is generally reserved for niche situations where the data range demonstrably covers every single member of the group being analyzed. Unless you are absolutely certain that your dataset is comprehensive of the target group, the robust statistical choice for estimation is always the sample function, **COVARIANCE.S**.

Conclusion and Further Reading

Choosing the correct covariance function--be it **COVARIANCE.P** or **COVARIANCE.S**--is paramount for maintaining statistical integrity in your Excel analysis. Misapplying the population function to sample data can lead to a slight but systematic underestimation of the true population covariance, potentially skewing subsequent regression or correlation studies.

Always review the origin and scope of your data: if it's a complete dataset (P), use **COVARIANCE.P**; if it's an estimate derived from a subset (S), use **COVARIANCE.S**. This diligence ensures that your statistical conclusions are grounded in the appropriate mathematical principles.

For those interested in exploring similar distinctions in other commonly used Excel functions, the following resources provide further tutorials on population versus sample calculations:

[QUARTILE.EXC vs. QUARTILE.INC in Excel: What's the Difference?](#)