

# How do you calculate sample and population variance in Python?

Authored by  
**stats writer**

May 6, 2024

## RECOMMENDED CITATION

stats writer (2024). *How do you calculate sample and population variance in Python?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=143557>

Calculating sample and population variance in Python involves using specific formulas and methods to measure the variability of a set of data points. The sample variance is computed by taking the sum of squared deviations from the mean of a sample, divided by the number of observations minus one. On the other hand, the population variance is calculated by taking the sum of squared deviations from the mean of a population, divided by the total number of observations. These calculations can be done using built-in functions in Python, such as the `numpy.var()` function, which takes in a dataset as an input and returns the variance value. Alternatively, the `var()` function from the `statistics` module can also be used to calculate the sample variance. By understanding these concepts and utilizing the appropriate functions, one can effectively measure the spread of data in Python.

## Calculate Sample & Population Variance in Python

**The variance is a way to measure of values in a dataset.**

**The formula to calculate population variance is:**

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

**where:**

**$\Sigma$ : A symbol that means "sum"**

**$\mu$ : Population mean**

**$x_i$ : The  $i$ th element from the population**

**$N$ : Population size**

**The formula to calculate sample variance is:**

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

**where:**

**x:** Sample mean  
 **$x_i$ :** The  $i$ th element from the sample  
**n:** Sample size

We can use the `variance` and `pvariance` functions from the `statistics` library in Python to quickly calculate the sample variance and population variance (respectively) for a given array.

```
from statistics import variance, pvariance
```

```
#calculate sample variance  
variance(x)
```

```
#calculate population variance  
pvariance(x)
```

The following examples show how to use each function in practice.

Example 1: Calculating Sample Variance in Python

The following code shows how to calculate the sample variance of an array in Python:

```
from statistics import variance
```

```
#define data
```

```
data =
```

```
#calculate sample variance
```

```
variance(data)
```

```
22.067
```

**The sample variance turns out to be 22.067.**

**Example 2: Calculating Population Variance in Python**

```
from statistics import pvariance
```

```
#define data
```

```
data =
```

```
#calculate sample variance
```

```
pvariance(data)
```

```
20.596
```

**The population variance turns out to be 20.596.**

**Notes on Calculating Sample & Population Variance**

**Keep in mind the following when calculating the sample and population variance:**

You should calculate the population variance when the dataset you're working with represents an entire population, i.e. every value that you're interested in. You should calculate the sample variance when the dataset you're working with represents a sample taken from a larger population of interest. The sample variance of a given array of data will always be larger than the population variance for the same array of data because there is more uncertainty when calculating the sample variance, thus our estimate of the variance will be larger.

The following tutorials explain how to calculate other measures of spread in Python: