

# How to Easily Perform Bartlett's Test in Python

Authored by  
**stats writer**

December 6, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform Bartlett's Test in Python*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106160>

The ability to confidently assess the underlying assumptions of statistical models is fundamental to rigorous data analysis. One of the most critical assumptions for parametric tests, such as ANOVA, is the concept of homoscedasticity—the equality of population variance across different samples or groups.

The Bartlett's Test provides a robust statistical method for verifying this assumption. This detailed guide demonstrates how to implement and interpret Bartlett's Test efficiently using the SciPy library in Python, ensuring your subsequent statistical inferences are valid and reliable.

To perform this test in Python, you must first import the necessary modules from the **scipy.stats** library. The core function is `scipy.stats.bartlett()`. When executed with your sample data, this function returns two key metrics: the test statistic and the associated p-value. The comparison of the p-value against a chosen significance level dictates whether the assumption of homoscedasticity holds.

If the calculated p-value is less than the predetermined significance level (commonly  $\alpha = 0.05$ ), then the null hypothesis of equal variances is rejected. Rejecting the null hypothesis implies that the data exhibits heteroscedasticity (unequal variances), which may necessitate using alternative statistical methods or data transformations.

## Understanding Bartlett's Test: A Foundation for Statistical Analysis

Bartlett's test, named after British statistician Maurice Bartlett, is a classical procedure utilized specifically to check for the equality of population variances across three or more groups. It is highly sensitive to non-normality; therefore, before applying Bartlett's test, it is usually assumed that the data within each group is sampled from a normal distribution. If the data severely violates the normality assumption, alternative tests like Levene's test or Brown-Forsythe test might be more appropriate, as they are less sensitive to distributional deviations.

The primary utility of Bartlett's test lies in validating the prerequisites for powerful parametric tests like the Analysis of Variance (ANOVA). ANOVA relies fundamentally on the stability of variance across all groups being compared. If variances are unequal (heteroscedastic), the standard F-test in ANOVA may produce misleading results, potentially leading to incorrect conclusions about the equality of population means. By confirming homoscedasticity, we ensure the validity of the chosen statistical framework.

The core objective of the test is to consolidate the evidence from multiple samples to determine if they could reasonably originate from populations sharing the same variance. This makes Bartlett's test an essential screening step in multivariate analysis and experimental design, especially when dealing with data derived from controlled experiments.

## The Core Hypotheses and Distribution of Bartlett's Test

Like all hypothesis tests, Bartlett's test operates based on a pair of competing statements regarding the population parameters. Understanding these hypotheses is crucial for correctly interpreting the test output. The test structure is defined by the null hypothesis ( $H_0$ ) and the alternative hypothesis ( $H_A$ ):

**Null Hypothesis ( $H_0$ ):** The variances among all  $k$  observed groups are statistically equal ( $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ ). This is the assumption of homoscedasticity.

**Alternative Hypothesis ( $H_A$ ):** At least one group possesses a variance that is significantly different from the others. This indicates heteroscedasticity.

The test statistic, denoted typically as  $B$ , is calculated based on the sample variances and sample sizes of the groups. The formula involves calculating a weighted average of the logarithms of the sample variances, which is then adjusted to approximate a known distribution. This adjustment ensures that the resultant statistic provides a reliable measure of deviation from the null assumption.

Under the assumption that the null hypothesis is true and the data are normally distributed, this test statistic approximately follows a Chi-Square distribution ( $\chi^2$ ). The degrees of freedom for this distribution are calculated as  $k-1$ , where  $k$  represents the number of independent groups being compared. The final decision relies on comparing the calculated  $B$  value to the critical value from the  $\chi^2$  distribution or, more commonly in computational environments like Python, using the corresponding p-value.

### Prerequisites: When and Why to Use This Test

The decision to use Bartlett's test should be guided by specific analytical needs and data characteristics. It is most often employed when an analyst needs to verify the stability of variation across experimental conditions. For instance, in clinical trials, researchers might want to ensure that a new drug affects patient outcomes (like recovery time) with similar variability across different dosage levels before comparing the mean recovery times using ANOVA.

The fundamental requirements for the reliability of Bartlett's test are rigorous:

**Independence:** The samples for each group must be drawn independently from the others.

**Normality:** The data in each group must be approximately normally distributed. Bartlett's test is highly sensitive to violations of this assumption. If normality is questionable, Levene's test is usually preferred as a more robust non-parametric alternative for variance equality assessment.

**Measurement Scale:** The dependent variable must be measured on an interval or ratio scale to allow for meaningful variance calculation.

By confirming these conditions, the statistician gains high confidence that a non-significant result from Bartlett's Test truly supports the homogeneity of variance, thereby justifying the use of subsequent parametric procedures like one-way or two-way ANOVA. Ignoring the homoscedasticity assumption when it is violated can lead to inflated Type I error rates, resulting in spurious findings of significant mean differences.

## Implementing Bartlett's Test in Python: Setup and Data Preparation

The practical application of Bartlett's test in Python relies entirely on the powerful statistical functions provided by the SciPy library. SciPy is the standard scientific computing library in Python, providing optimized numerical routines, including specific tools for statistical inference contained within the `scipy.stats` submodule. Before execution, data must be structured such that the samples from each group are provided as separate, distinct arrays or lists. This organization is necessary because the `bartlett()` function takes each group's data as an individual positional argument.

Consider a scenario where a university professor is investigating whether three distinct studying techniques (Technique A, B, and C) result in different performance variability on a standardized exam. The professor randomly assigned 10 students to each technique, resulting in 30 total observations. We are primarily concerned here with checking if the spread (variance) of scores is consistent across the techniques, a required check before comparing the average effectiveness of the techniques using an ANOVA.

### Step 1: Create the Data

The data below represents the exam scores achieved by the 30 students across the three study techniques. We organize these scores into three separate lists, A, B, and C, corresponding to the respective techniques. This structure is essential for input into the **scipy.stats** function. Data preparation is the critical first step; any errors in assigning observations to the correct group lists will render the subsequent statistical results meaningless.

#### #create data for three study techniques

A =

B =

C =

It is important to note that while this example uses small sample sizes for clarity, Bartlett's test assumes that the underlying population distribution is normal. Larger sample sizes generally mitigate some concerns about non-normality, but careful inspection of the data distribution via histograms or Q-Q plots remains a best practice before proceeding with this test. If these

preliminary checks reveal substantial deviations from normality, the assumption for Bartlett's test is violated, and an alternative test should be considered.

## Executing the Test using `scipy.stats.bartlett()`

### Step 2: Perform Bartlett's Test

Once the data is prepared, performing the test is straightforward using Python. We import `scipy.stats` and then pass all three data arrays (A, B, and C) directly to the `bartlett()` function. The function handles the entire calculation internally, providing the test statistic and the resultant p-value. The simplicity of the execution masks the complexity of the underlying log-likelihood ratio calculations used in the test.

The `scipy.stats.bartlett()` function requires at least two data arrays as input. It calculates the pooled standard deviation and uses the logarithm of the variances to compute the final test statistic  $B$ . The function is highly efficient and outputs a specific tuple-like object containing the necessary statistical information.

#### **import scipy.stats as stats**

```
#perform Bartlett's test on the scores
stats.bartlett(A, B, C)
```

```
BartlettResult(statistic=3.30243757, pvalue=0.191815983)
```

The output object, `BartlettResult`, packages the calculated statistic and the p-value. In this specific execution, the function calculated a test statistic of approximately 3.3024 and an associated p-value of approximately 0.1918. These two values are the sole input required for making the statistical decision regarding the equality of variances across the three techniques.

## Interpreting the Results

The interpretation phase is where the statistical output informs the practical conclusion. We must compare the calculated p-value against our pre-selected significance level ( $\alpha$ ). For most social and scientific research,  $\alpha$  is set at 0.05, meaning we are willing to accept a 5% chance of falsely rejecting the true null hypothesis (Type I error). Our decision rule is straightforward:

If  $p\text{-value} \leq \alpha$  (e.g., 0.05), we reject  $H_0$ .

If  $p\text{-value} > \alpha$ , we fail to reject  $H_0$ .

The key results extracted from the test are:

Test statistic  $\chi^2$ : **3.3024** (This value follows the  $\chi^2$  distribution with  $k-1=2$  degrees of freedom).

P-value: **0.1918** (The probability of observing a test statistic as extreme as 3.3024, assuming the variances are truly equal).

Since the calculated p-value (0.1918) is substantially greater than the conventional significance level ( $\alpha = 0.05$ ), we fail to reject the null hypothesis. The professor does not have sufficient statistical evidence to conclude that the three groups have significantly different variances. This result strongly suggests that the population variances for the exam scores across the three study techniques (A, B, and C) are likely equal, thereby satisfying the homoscedasticity assumption.

If, conversely, the p-value had been smaller than 0.05 (e.g., 0.01), we would have rejected  $H_0$ , concluding that the techniques exhibit significantly different score variabilities (heteroscedasticity). In that scenario, the professor would need to employ a robust ANOVA method, such as Welch's ANOVA, or a non-parametric test, that does not rely on the assumption of equal variance.

## Conclusion: Applying Homoscedasticity Findings

The successful confirmation of homoscedasticity using the Bartlett's Test provides essential validation for the professor's planned next step. Because the variances are deemed equal, the professor can proceed confidently with the primary analysis, typically a one-way ANOVA, to determine if there are significant differences in the average exam scores (means) resulting from the three studying techniques. The integrity of the subsequent ANOVA results is thus preserved.

In summary, Bartlett's Test serves as a crucial gatekeeper in parametric statistical analysis. Its implementation through the `scipy.stats` module in Python is efficient and straightforward, offering a necessary check for data compliance with model assumptions. While powerful, researchers must always remember its high sensitivity to deviations from normality. If the normality assumption is highly suspect, using the Levene's test alongside or instead of Bartlett's test is a prudent practice to ensure the robustness of the homoscedasticity assessment before drawing final inferences.