

Learn How to Perform a Brown-Forsythe Test in Python: Assessing Homoscedasticity for ANOVA

Authored by
Mohammed loot

November 6, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learn How to Perform a Brown-Forsythe Test in Python: Assessing Homoscedasticity for ANOVA*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11526>

The Critical Role of Homoscedasticity in Statistical Analysis

The [one-way ANOVA](#) (Analysis of Variance) stands as a cornerstone statistical method, fundamentally designed to test for statistically significant differences between the means of three or more independent groups. This powerful technique is indispensable across diverse scientific disciplines, ranging from agricultural research and biological studies to sophisticated engineering applications and social science research, providing clarity on how various treatments or factors affect an outcome variable.

The reliability of any conclusion drawn from the [one-way ANOVA](#) is predicated upon satisfying several core statistical assumptions. Among these, the assumption of homogeneity of variances--universally known as [homoscedasticity](#)--is exceptionally critical. This means that the population variances from which the different samples are collected must be statistically equal. Essentially, the spread or variability of the data around the mean should be consistent across all groups being compared.

When this fundamental assumption is violated, a condition known as [heteroscedasticity](#) occurs. If the variances are unequal, the resulting F-statistic calculated by the standard ANOVA becomes distorted, leading to inflated Type I error rates (false positives) or, conversely, reduced statistical power. Consequently, researchers might draw erroneous conclusions regarding the true differences, or lack thereof, among the group means. Therefore, establishing the presence of [homoscedasticity](#) through a formal preliminary test is a mandatory step before committing to the main ANOVA analysis.

Why Choose the Brown-Forsythe Test?

To rigorously assess the validity of the [homoscedasticity](#) assumption, statisticians employ specific tests designed to compare variances across multiple groups. While the standard Levene's test is common, the [Brown-Forsythe test](#) is frequently the preferred choice, particularly when dealing with data sets that are suspected of deviating significantly from a normal distribution. Its preference stems from its superior robustness under conditions of non-normality.

The [Brown-Forsythe test](#) is structurally similar to Levene's test but incorporates a crucial modification: instead of calculating the absolute deviations from the group means (as Levene's test does), it calculates the absolute deviations from the group medians. Because the median is less sensitive to extreme outliers than the mean, this modification makes the Brown-Forsythe test significantly more reliable and robust when analyzing data from non-normal distributions, ensuring a more trustworthy assessment of variance equality.

The test evaluates the variances against the standard statistical hypotheses that govern all inferential tests:

H0 (Null Hypothesis): The population variances across all groups are equal (the assumption of homoscedasticity holds true).

HA (Alternative Hypothesis): At least one population variance is different from the others (heteroscedasticity exists).

The decision criterion is based on the resulting [p-value](#). If the [p-value](#) (the probability of observing the data given that H0 is true) falls below the predetermined significance level (alpha, typically set at **0.05**), we must reject the null hypothesis. Rejecting H0 means concluding that the variances are significantly unequal, confirming a violation of the primary assumption required for a standard [one-way ANOVA](#).

Step 1: Setting Up the Data Environment in Python

This section initiates the practical application of the Brown-Forsythe methodology using the Python programming language, leveraging its extensive ecosystem of statistical libraries. By following these steps, practitioners can efficiently integrate this crucial assumption check into their data analysis workflow.

To illustrate the process, consider a common research scenario in experimental design. A team of agricultural scientists is investigating the comparative efficacy of three unique chemical fertilizers (designated Group 1, Group 2, and Group 3) on maximizing plant height. They established an experiment where thirty identical plant specimens were randomly allocated into three treatment groups, each receiving one distinct fertilizer type. After a standardized growth period of one month, the final height of every plant was meticulously measured in consistent units.

The collected plant height measurements for each of the three experimental treatment groups are structured below as standard Python lists, ready for import and analysis:

```
group1 =
```

```
group2 =
```

```
group3 =
```

Step 2: Preliminary Variance Assessment using NumPy

Before proceeding directly to the formal statistical test, it is highly recommended to conduct an exploratory data analysis (EDA). This preliminary step involves calculating and visually inspecting descriptive statistics, particularly the sample variances, which provides immediate context for the formal Brown-Forsythe test. We employ the robust numerical capabilities of the [NumPy](#) library for this task.

By calculating the sample variance for the plant height measurements within each of the three

groups, we gain an initial quantitative measure of the data spread. While numerical differences in variance are expected, this context helps us anticipate the scope of the variability. If the variances appear dramatically different at this stage, it increases the likelihood that the formal Brown-Forsythe test will reject the null hypothesis.

Import the NumPy library for numerical operations

```
import numpy as np
```

```
# Calculate the variance of plant measurements for each group
```

```
print(np.var(group1), np.var(group2), np.var(group3))
```

```
8.69 7.81 7.0
```

The output reveals the calculated variances: **8.69** for Group 1, **7.81** for Group 2, and **7.0** for Group 3. Though these values exhibit some degree of variation, the differences are numerically small. It is imperative to remember that descriptive statistics alone cannot confirm statistical significance. We must proceed with the rigorous [Brown-Forsythe test](#) to ascertain whether these observed differences are statistically significant enough to warrant rejecting the assumption of [homoscedasticity](#).

Step 3: Executing the Brown-Forsythe Test using SciPy

The statistical functions necessary for performing the [Brown-Forsythe test](#) are integrated within the comprehensive `scipy.stats` module in Python, which is the cornerstone library for scientific computing and advanced statistics. The specific function utilized for variance testing is `scipy.stats.levene`.

It is absolutely critical to understand that the `levene` function is flexible and can execute both Levene's test and the Brown-Forsythe variant. To correctly execute the more robust Brown-Forsythe test, the user must explicitly set the optional `center` parameter to the string value `'median'`. If this parameter were omitted or set to the default value of `'mean'`, the function would revert to performing the standard, less robust Levene's test. By specifying `center='median'`, we ensure the calculation uses deviations from the median, adhering to the Brown-Forsythe methodology.

```
import scipy.stats as stats
```

```
stats.levene(group1, group2, group3, center='median')
```

```
LeveneResult(statistic=0.17981072555205047, pvalue=0.8364205218185946)
```

Interpreting the Statistical Output and Conclusion

The execution of the `stats.levene` function returns a `LeveneResult` object, which concisely presents the two vital pieces of information required for hypothesis testing: the calculated test statistic (often denoted as W) and its corresponding [p-value](#).

Test Statistic (W): 0.1798. This value represents the magnitude of the variance differences observed in the sample data.

P-value: 0.8364. This is the probability of observing a test statistic as extreme as 0.1798 (or more extreme) if the null hypothesis of equal variances were true.

The formal decision process requires comparing the derived [p-value](#) (0.8364) against the predetermined significance threshold ($\alpha = 0.05$). The core rule of statistical inference dictates that if $p < \alpha$, we reject H_0 . Conversely, since 0.8364 is profoundly larger than 0.05, we **fail to reject the null hypothesis**.

This outcome signifies that the differences in variance observed among the three fertilizer groups (8.69, 7.81, and 7.0) are not statistically significant at the 5% level. Consequently, we conclude that there is insufficient evidence to suggest that the population variances are unequal. The critical assumption of [homoscedasticity](#) has been successfully met, allowing the researchers to confidently proceed to perform the primary [one-way ANOVA](#) to analyze the mean effectiveness of the fertilizers.

Handling Variance Heterogeneity (If H_0 is Rejected)

It is crucial for any expert data analyst to understand the appropriate recourse if the Brown-Forsythe test were to indicate a significant violation of the equal variances assumption. Had the test result led to a rejection of the null hypothesis, indicating statistically unequal variances, two primary options would be available for continuing the statistical investigation:

Proceeding with the One-Way ANOVA with Caution:

Despite the formal violation, the [one-way ANOVA](#) is remarkably robust to minor infractions of the [homoscedasticity](#) assumption, particularly when the sample sizes across all groups are equal (balanced design). A widely accepted rule of thumb suggests that ANOVA results remain reasonably reliable as long as the ratio between the largest sample variance and the smallest sample variance does not exceed 4:1.

Applying this check to our example data from Step 2: the largest variance was 8.69 (Group 1) and the smallest was 7.0 (Group 3). The variance ratio is calculated as $8.69 / 7.0$, yielding a ratio of approximately **1.24**. Since 1.24 is significantly below the threshold of 4, even a test result

suggesting inequality might be considered minor, allowing the researcher to proceed with the standard ANOVA while noting the potential for slight inflation of the Type I error rate. This contextual assessment is vital for interpreting test sensitivity.

Performing a Non-Parametric Alternative: The Kruskal-Wallis Test:

If the ratio of the largest to smallest variance substantially exceeds the 4:1 threshold, or if the data distributions are highly non-normal alongside unequal variances, the standard ANOVA is inappropriate. In such cases, the most statistically sound alternative is the [Kruskal-Wallis Test](#). This test is the non-parametric analogue to the one-way ANOVA, designed to test for differences in medians across groups without making stringent assumptions about the underlying data distribution or the equality of variances.

For situations where the variance assumption is severely violated, transitioning to the [Kruskal-Wallis Test](#) in Python (also available within `scipy.stats`) ensures the integrity and validity of the final conclusions regarding group differences.