

# Learning Levene's Test for Homogeneity of Variance: A Stata Tutorial

Authored by  
**Mohammed Iooti**

November 8, 2025

## RECOMMENDED CITATION

Mohammed Iooti (2025). *Learning Levene's Test for Homogeneity of Variance: A Stata Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13618>

[Levene's Test](#) is a cornerstone procedure in [inferential statistics](#), designed specifically to evaluate whether the variances of two or more independent populations are statistically equivalent. This crucial condition, known as [homoscedasticity](#), represents a foundational assumption underpinning numerous powerful parametric analyses, including the standard independent samples t-test and the general [Analysis of Variance \(ANOVA\)](#). Before drawing definitive conclusions about differences in population means, researchers must validate this assumption; its fulfillment is essential for ensuring the reliability and statistical validity of subsequent test results.

Should the variances across the comparison groups be significantly disparate--a state termed [heteroscedasticity](#)--standard parametric tests can yield substantially biased outcomes, potentially inflating the Type I error rate and leading to unwarranted conclusions about mean differences. Consequently, employing a powerful and reliable preliminary test like Levene's is indispensable for rigorous data validation. This detailed tutorial provides expert, step-by-step instructions on executing and accurately interpreting [Levene's Test](#) within the professional statistical software environment of [Stata](#).

## The Critical Role of Variance Homogeneity in Statistical Modeling

The robustness of statistical inference hinges heavily upon verifying underlying distributional assumptions. Among these, the assumption of homogeneity of variances is arguably one of the most frequently encountered and tested. This principle demands that the dispersion, or variability, of the outcome variable must remain approximately constant across all groups designated for comparison. Failing to confirm this assumption can critically compromise the interpretation of any hypothesis test specifically designed to compare central tendencies (means). For instance, if one experimental group displays a spread of scores vastly greater than another, the pooled variance estimates traditionally utilized in standard t-tests become unreliable, potentially masking or fabricating significant effects.

[Levene's Test](#) directly addresses this statistical vulnerability by formally testing the null hypothesis, which posits that the population variances are equal across all defined groups. Mechanically, the test achieves this by performing a standard one-way [ANOVA](#) on the absolute deviations of each observation from its respective group mean or median. A decision to reject the null hypothesis (typically indicated by a [p-value](#) falling below the conventional alpha level of 0.05) suggests strong evidence that the variances are significantly unequal. When this occurs, analysts must pivot to alternative, variance-unequal procedures, such as the widely accepted Welch's t-test or more sophisticated robust ANOVA methodologies, to ensure unbiased results.

## Designing the Homogeneity Test in Stata

To demonstrate the practical application of [Levene's Test](#), we will employ a widely used, authentic

medical dataset frequently featured in [Stata](#) documentation. Our primary research objective is to rigorously ascertain whether the variability in the duration of hospital stays differs significantly between male and female patients who have undergone a specific medical intervention. The dependent variable of interest is "lengthstay," which measures the duration of hospitalization in days, while the categorical factor variable is "sex."

The dataset utilized for this demonstration, designated as *stay*, comprises comprehensive records for 1,778 unique patients. This cohort is evenly distributed, consisting of 884 male patients and 894 female patients. By applying [Levene's Test](#) to this structured data, we are able to formally determine if the dispersion of hospital stay durations is consistent across the two gender groups. This preliminary analysis is not merely a formality; it is a critical requirement for validating the suitability of any subsequent mean comparison model that uses patient length of stay as the dependent variable.

## Step 1: Loading and Initial Inspection of the Data

The initial stage of any robust statistical investigation involves accessing, loading, and performing a thorough review of the dataset to confirm its structural integrity and variable types. In the [Stata](#) environment, datasets are seamlessly loaded using intuitive built-in commands. We will commence the process by fetching the specific *stay* dataset directly from the official Stata Press repository, ensuring we are using standard example data.

To successfully load the *stay* dataset into your active [Stata](#) session, you must execute the following command precisely as it appears below in the Command window:

**use <http://www.stata-press.com/data/r13/stay>**

Once the data loading process is complete, it is standard methodological practice to immediately inspect the structure and initial observations of the variables central to the analysis. A rapid method to examine the first ten observations, confirming variable types and data arrangement, is achieved using the following command:

**list in 1/10**

This command generates a precise snapshot of the dataset, which allows the analyst to verify that the structure is fully appropriate for the proposed test of variance homogeneity.

```
. use http://www.stata-press.com/data/r13/stay
. list in 1/10
```

	length~y	sex
1.	6.6	male
2.	4.4	male
3.	3.3	male
4.	3.3	male
5.	5.5	male
6.	16.5	male
7.	9.9	male
8.	6.6	male
9.	7.7	male
10.	4.4	male

As clearly illustrated in the output snapshot, the dataset is defined by two principal variables: the first column, labeled `lengthstay`, provides the continuous measurement of hospitalization duration in days for each patient. The second column, `sex`, functions as our categorical grouping variable, differentiating between male and female patients. This classic structure--one continuous measurement variable paired with one categorical grouping variable--is ideally suited for the application of [Levene's Test](#).

## Step 2: Executing Levene's Test using the `robvar` Command

In the [Stata](#) statistical environment, the implementation of [Levene's Test](#) is streamlined through the use of the `robvar` command, which is an abbreviation for "robust variance test." This command is highly favored because it efficiently computes several variations of Levene's Test concurrently, providing necessary flexibility depending on the underlying characteristics and distribution of the data. The general syntax required for the `robvar` command is straightforward, mandating only the specification of the measurement variable followed by the grouping variable.

The generalized syntax structure that must be adhered to is:

**`robvar measurement_variable, by(grouping_variable)`**

Applying this exact syntax to our specific research context, where `lengthstay` serves as the continuous measurement variable and `sex` is the binary grouping variable, we proceed by executing the following precise command in the Stata Command window:

**robvar lengthstay, by(sex)**

Upon successful execution, [Stata](#) immediately generates a comprehensive output. This output begins with a summary table containing descriptive statistics for each group, followed by the specific test statistics and their corresponding [p-values](#) for the three primary variations of Levene's Test. This detailed numerical output forms the foundation upon which our formal statistical conclusion regarding variance homogeneity will be established.

**. robvar lengthstay, by(sex)**

sex	Summary of Length of stay in days		
	Mean	Std. Dev.	Freq.
male	<b>9.0874434</b>	<b>9.7884747</b>	<b>884</b>
female	<b>8.800671</b>	<b>9.1081478</b>	<b>894</b>
Total	<b>8.9432508</b>	<b>9.4509466</b>	<b>1,778</b>

W0 = **0.55505315** df(1, 1776) Pr > F = **0.45635888**

W50 = **0.42714734** df(1, 1776) Pr > F = **0.51347664**

W10 = **0.44577674** df(1, 1776) Pr > F = **0.50443411**

**Detailed Interpretation of the Stata Output**

The output generated by the `robvar` command is methodically structured to provide both the essential descriptive context and the crucial inferential results necessary to reach a conclusion on variance homogeneity. A meticulous examination of each section of this output is essential for accurate reporting and decision-making.

The initial segment of the output displays the descriptive **Summary table**. This table furnishes key descriptive statistics for the length of stay variable, meticulously stratified by gender. We can observe the calculated mean length of stay, the standard deviation, and the total count of observations for both males and females. Although the standard deviation for males (9.7884747) is numerically larger than the standard deviation reported for females (9.1081478), this observed descriptive difference alone is insufficient to confirm a statistically significant variation in the underlying population variances. This scenario precisely highlights why the inferential component provided by [Levene's Test](#) is statistically necessary.

The subsequent sections present the calculated results for the three main variations of the test statistic (W0, W50, and W10). Each variation is calculated by centering the data using a different

measure of central tendency:

**W0 (Centered at the Mean):** This represents the classic and most traditional form of [Levene's Test](#). The calculated test statistic is reported as **0.55505315**, which corresponds to a [p-value](#) of **0.45625888**.

**W50 (Centered at the Median):** This version is commonly known as the Brown-Forsythe test. It is generally considered statistically more [robust](#) to deviations from normality and severe skewness when compared to the mean-centered version (W0). The resulting test statistic is **0.42714734**, with an associated [p-value](#) of **0.51347664**.

**W10 (Centered using the 10% Trimmed Mean):** This advanced version utilizes a 10% trimmed mean (which involves removing the top 5% and bottom 5% of extreme values) to mitigate the undue influence of severe outliers on the final variance estimate. The test statistic generated is **0.44577674**, yielding a [p-value](#) of **0.50443411**.

## Selecting the Most Robust Test Statistic

To reach a definitive statistical conclusion, we must compare the calculated [p-values](#) against our predetermined significance level, typically  $\alpha = 0.05$ . In this specific analysis, regardless of which version of Levene's Test is utilized, all resulting p-values (0.456, 0.513, and 0.504) are substantially greater than 0.05. Consequently, based on the evidence, we formally fail to reject the null hypothesis of equal variances. This outcome confirms that there is no statistically significant evidence to suggest a difference in the variance of the length of hospital stay between male and female patients. We can therefore confidently conclude that the assumption of homogeneity of variances is upheld for this dataset, paving the way for standard parametric mean comparisons.

While the choice of centering method--mean (W0), median (W50), or trimmed mean (W10)--did not alter the final conclusion in this relatively symmetric example, this selection becomes critically important when dealing with datasets that exhibit significant asymmetry or heavy skewness. Statistical methodology, notably based on influential research by Conover, Johnson, and Johnson (1981), strongly recommends the routine adoption of the median-centered test (W50) when the underlying data distribution is suspected to be non-normal, asymmetric, or highly skewed.

The median-centered test (W50) is universally regarded as a more [robust](#) approach because the median is inherently less susceptible to distortion by extreme values (outliers) compared to the mean. When data distributions are truly symmetric, the mean and median are nearly identical, and subsequently, all three versions of Levene's Test (W0, W50, W10) will produce closely aligned results, exactly as demonstrated in our patient stay example. However, in instances where severe non-normality is present, relying on the W50 statistic ensures that the test maintains appropriate control over the critical Type I error rate, guaranteeing a more statistically reliable assessment of

variance homogeneity. Analysts are therefore strongly advised to conduct a visual and numerical examination of their measurement variable's distribution before finalizing which test statistic result is reported.