How to Perform Levene's Test for Equality of Variances in Stata

Authored by

stats writer

March 9, 2026

The Fundamental Role of Variance Analysis in Statistics

In the realm of **quantitative research**, understanding the distribution of data is as critical as identifying its central tendency. One of the most vital assumptions in parametric statistical testing is **homoscedasticity**, or the requirement that different groups within a dataset possess equal variances. When we compare the means of two or more groups--such as in an independent samples t-test or an analysis of variance (ANOVA)--we assume that the spread of data points around those means is relatively consistent. If this assumption is violated, the resulting **p-values** may be unreliable, potentially leading to Type I or Type II errors that undermine the validity of the entire study.

To address this, researchers frequently employ Levene's test, a robust inferential statistic used to assess the equality of variances across groups. Unlike some older methods, such as Bartlett's test, which are highly sensitive to departures from a **normal distribution**, Levene's test is designed to be more resilient. It provides a formal framework to test the **null hypothesis** that the population variances are equal. If the test returns a result that is statistically significant, it suggests that the groups have different levels of variability, which may require the researcher to use alternative "robust" statistical methods or data transformations.

Within the environment of Stata, a powerful integrated statistical software package, performing this test is both efficient and highly customizable. Stata provides specialized commands that not only calculate the standard version of the test but also provide variations centered on the **median** or the **trimmed mean**. This flexibility is essential for modern data science, where real-world datasets often contain outliers or exhibit skewness that can distort traditional calculations. By mastering these tools, analysts can ensure their findings are built on a solid foundation of verified assumptions.

An Introduction to Levene's Test and Its Mechanics

At its core, Levene's test operates by calculating the absolute difference between each observation and the group mean, then performing an ANOVA on those absolute differences. The logic is elegant: if the variances are truly equal across groups, the average distance of individual points from their group mean should also be equal. By transforming the variance problem into a comparison of means (of the deviations), the test leverages the well-understood properties of the F-distribution to determine **statistical significance**. This makes it a primary diagnostic tool before proceeding with more complex comparative modeling.

One of the primary reasons for the popularity of Levene's test in Stata is its "robustness." In statistical terminology, a test is robust if it performs well even when its underlying assumptions are not perfectly met. Because many real-world datasets in fields like psychology, medicine, and economics are not perfectly **normally distributed**, the standard F-test for variances can be misleading. Levene's test provides a safer alternative that minimizes the risk of incorrectly rejecting the assumption of equal variances simply because the data is slightly skewed or has heavy tails.

When implementing this in a workflow, it is important to understand the criteria for interpretation. The test yields a p-value which is compared against a pre-determined significance level, typically 0.05. If the p-value is lower than this threshold, the researcher must reject the idea of equal variances. This realization is a pivotal moment in the data analysis pipeline, as it signals that standard pooling of variances is inappropriate. In such cases, the researcher might opt for Welch's t-test or other methods that do not rely on the **homoscedasticity** assumption.

Preparing the Stata Workspace for Analysis

Before executing any complex **statistical hypothesis testing**, it is imperative to ensure that your data is correctly formatted and loaded into the Stata environment. The software requires variables to be properly typed; for instance, the measurement variable (the outcome you are testing) must be a numerical, continuous variable, while the grouping variable should be a categorical variable (often stored as an integer with value labels). Clean data preparation prevents execution errors and ensures that the **descriptive statistics** generated later are accurate and meaningful.

A common first step in any Stata session is to clear the existing memory and set the working directory. This organizational habit ensures that you are working with the correct version of your dataset and that any output or log files are saved in the appropriate location. Once the environment is ready, the data can be imported from various formats, including CSV, Excel, or Stata's native .dta format. For the purposes of learning, many researchers utilize Stata's built-in datasets or those available via official web repositories, which are designed to demonstrate specific statistical properties like variance inequality.

After loading the data, a prudent analyst will perform a preliminary check using commands like "summarize" or "codebook." This allows you to identify missing values, verify the range of your variables, and ensure that the grouping variable has at least two distinct levels. Without these checks, you might attempt to run Levene's test on a variable that contains non-numeric characters or on a group that has only one observation, both of which would cause the command to fail. Establishing a rigorous preparation routine is the hallmark of a professional statistician.

Step 1: Loading and Listing the Case Study Data

To demonstrate the practical application of variance testing, we will utilize a dataset known as "stay." This specific dataset is often used in medical research contexts to analyze the length of time patients remain in a hospital after a specific procedure. The primary goal is to determine if the variance in the length of stay differs between male and female patients. This is a classic example of a "between-groups" comparison where understanding the spread of the data is just as important as understanding the **mean** stay duration.

First, we use the **use** command to fetch the data directly from the Stata Press servers. This ensures that we are working with a standardized dataset. The command is entered as follows:

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

Once the data is loaded into memory, it is helpful to inspect the structure of the observations. We can use the **list** command to view a subset of the data, which helps us visualize how the variables "lengthstay" and "sex" are recorded. By looking at the first ten rows, we can confirm that "lengthstay" is a continuous numerical variable representing days, while "sex" identifies the gender of the patient.

list in 1/10

```
. 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 |
     +------------------+
```


In the resulting table, the first column tracks the duration of the hospital stay, while the second column categorizes the individual. This simple visual check confirms that our data is in the "long" format required for group comparisons in Stata. With 1,778 observations--divided almost equally between 884 males and 894 females--this dataset provides a robust sample size for achieving high **statistical power** in our variance analysis.

Step 2: Executing the robvar Command in Stata

While some software packages have a command explicitly named after Levene, Stata uses a more descriptive command: **robvar**. The name stands for "robust variance," reflecting the test's ability to handle various data distributions. The syntax for this command is straightforward but requires specific placement of variables. You must specify the continuous measurement variable first, followed by the grouping variable within the "by" option.

For our patient stay analysis, the command is structured as follows:

robvar lengthstay, by(sex)

This command instructs Stata to calculate Levene's test (and its variants) for the variable "lengthstay," partitioned by the categories found in the "sex" variable. Upon execution, Stata processes the data and generates a comprehensive output window containing descriptive statistics and three distinct test statistics. This multi-faceted approach is what makes **robvar** superior to simpler variance tests, as it allows the researcher to choose the most appropriate version of the test based on the data's characteristics.

```
. robvar lengthstay, by(sex)

            |  Summary of Length of stay in days
        sex |        Mean   Std. Dev.       Freq.
------------+-------------------------------------
       male |   9.0874434   9.7884747         884
     female |    8.800671   9.1081478         894
------------+-------------------------------------
      Total |   8.9432508   9.4509466       1,778

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
```


The resulting output provides an immediate summary of the groups being compared. Before diving into the hypothesis test results, the summary table at the top of the output displays the **mean**, standard deviation, and sample size (N) for each group. In our case, we can observe that the standard deviation for males is approximately 9.79, while for females it is roughly 9.11. While these numbers differ numerically, the **robvar** command will determine if this difference is large enough to be considered "statistically significant" or if it is merely the result of random sampling variation.

Understanding the Summary Statistics and Group Variability

The initial section of the **robvar** output is crucial because it provides the context needed to interpret the test statistics. The mean length of stay tells us the average duration, but the standard deviation is the star of the show here. The standard deviation is the square root of the **variance**, representing the average distance of the data points from the mean. A higher standard deviation indicates that the data is more "spread out," whereas a lower value suggests that the observations are clustered more tightly around the average.

In our example, the male group shows a slightly higher standard deviation (9.788) than the female group (9.108). This suggests that there is more variability in how long men stay in the hospital compared to women. Some men might have very short stays while others have very long ones, creating a wider distribution. However, looking at these numbers in isolation is not enough to conclude that the populations are fundamentally different. We must rely on the **p-value** generated by Levene's test to make a scientific determination.

It is also worth noting the total number of observations (N) for each group. With nearly 900 subjects in each category, the test is very sensitive to even small differences in variance. In smaller datasets, a difference of 0.68 in standard deviation might be ignored, but in a large sample, it could potentially reach the threshold of **statistical significance**. The summary table ensures that you have a clear understanding of your sample size before you move on to the final interpretation of the test results.

Interpreting W0, W50, and W10 Test Statistics

Stata's **robvar** output provides three different versions of the test, labeled as W0, W50, and W10. These represent different ways of calculating the "center" of the data from which deviations are measured. The first, **W0**, is the standard Levene's test centered at the **mean**. In our results, W0 is 0.555 with a p-value of 0.456. Since this p-value is significantly higher than the 0.05 threshold, this version of the test fails to reject the **null hypothesis** of equal variances.

The second statistic, **W50**, is known as the Brown-Forsythe test, which centers the calculations on the median rather than the mean. This version is generally considered more robust when the data does not follow a **normal distribution** or contains outliers. Our output shows a W50 of 0.427 and a p-value of 0.513. This further reinforces the conclusion that there is no significant difference in variance between the two groups, as the p-value remains well above the 0.05 mark.

Finally, **W10** uses a trimmed mean, specifically removing the top and bottom 5% of the data to calculate the center. This is useful when the distribution has "heavy tails" or extreme values that might pull the mean in one direction. With a W10 of 0.445 and a p-value of 0.504, all three methods align in their conclusion. Regardless of the specific centering method used, we can confidently state that the assumption of **homoscedasticity** holds for this dataset, allowing us to proceed with standard parametric tests.

Final Conclusions and Methodological Recommendations

Performing Levene's test in Stata is a vital diagnostic step that ensures the integrity of subsequent statistical analyses. By using the **robvar** command, researchers can move beyond simple visual inspections and apply a rigorous mathematical framework to check for equal variances. In our case study of hospital stays, the consistent p-values across all three test variations (W0, W50, and W10) provided clear evidence that gender does not significantly impact the variability of stay duration, despite minor differences in the sample standard deviations.

It is important for researchers to remember that the choice between W0, W50, and W10 should be guided by the nature of their data. As suggested by Conover, Johnson, and Johnson (1981), the **median**-based test (W50) is often the superior choice for asymmetric or skewed data, as it provides more accurate results in the presence of non-normality. If your data is perfectly symmetric and **normally distributed**, the mean-based test (W0) and the median-based test will likely yield very similar results, as seen in our hospital stay example.

In summary, if your p-value is greater than 0.05, you can assume that your groups have equal variances and proceed with standard tests like the ANOVA. However, if the test is significant ($p < 0.05$), you should consider using **robust statistics** or transforming your data (e.g., using a logarithmic transformation) to stabilize the variance. By integrating these checks into your standard Stata workflow, you improve the reliability, reproducibility, and overall quality of your scientific research.