

Understanding and Testing for Normality in Stata: A Step-by-Step Tutorial

Authored by
Mohammed looti

November 8, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding and Testing for Normality in Stata: A Step-by-Step Tutorial*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13579>

A wide array of [statistical tests](#), particularly those classified as parametric, fundamentally rely on the assumption that the variables being analyzed are distributed according to the [normal distribution](#). When this critical assumption is violated, the integrity and reliability of the resulting statistics—including effect sizes, [p-values](#), and confidence intervals—can be severely compromised, leading researchers toward potentially misleading or erroneous conclusions.

Consequently, determining whether a continuous variable adheres to a [normal distribution](#) is an indispensable preliminary step in rigorous quantitative data analysis. This comprehensive guide details four distinct and powerful methods available within the statistical software **Stata** for evaluating normality, ranging from rapid, informal visual diagnostics to sophisticated, formal hypothesis tests.

All procedures outlined in this tutorial will be demonstrated using the widely available, built-in **Stata** sample dataset, *auto*. To ensure you can replicate these steps accurately, load the dataset into your current session using the command provided below:

```
sysuse auto
```

The Critical Role of Assessing Normality in Parametric Testing

The concept of the [normal distribution](#) sits at the core of parametric statistics. Parametric procedures, such as the independent samples t-test, linear regression, and Analysis of Variance (ANOVA), assume that the sampling distribution of the test statistic, or often the residuals of the model, is normally distributed. Failure to meet this expectation—especially when data exhibit severe asymmetry or heavy tails—means the underlying mathematical models used to calculate standard errors and test statistics are inaccurate. This inaccuracy directly increases the risk of committing [Type I or Type II errors](#) in hypothesis testing.

For academic and professional researchers, confirming [normality](#) is not a trivial formality; it is a fundamental diagnostic requirement that dictates the analytical path forward. If the data significantly deviates from the classic, symmetric "bell curve," analysts must consider remedial actions. These might include applying data transformation techniques—such as logarithmic, square root, or reciprocal transformations—to normalize the distribution, or, alternatively, choosing robust non-parametric alternatives that are distribution-free and do not rely on these strict distributional assumptions.

Stata provides a comprehensive suite of tools designed for the evaluation of data distribution characteristics. These methods can be broadly categorized into two types: **visual assessment**, which offers an immediate, intuitive understanding of the data's shape; and **inferential testing**, which provides an objective, statistical decision based on a calculated test statistic and its corresponding [p-value](#). A best-practice approach involves integrating both visual and formal

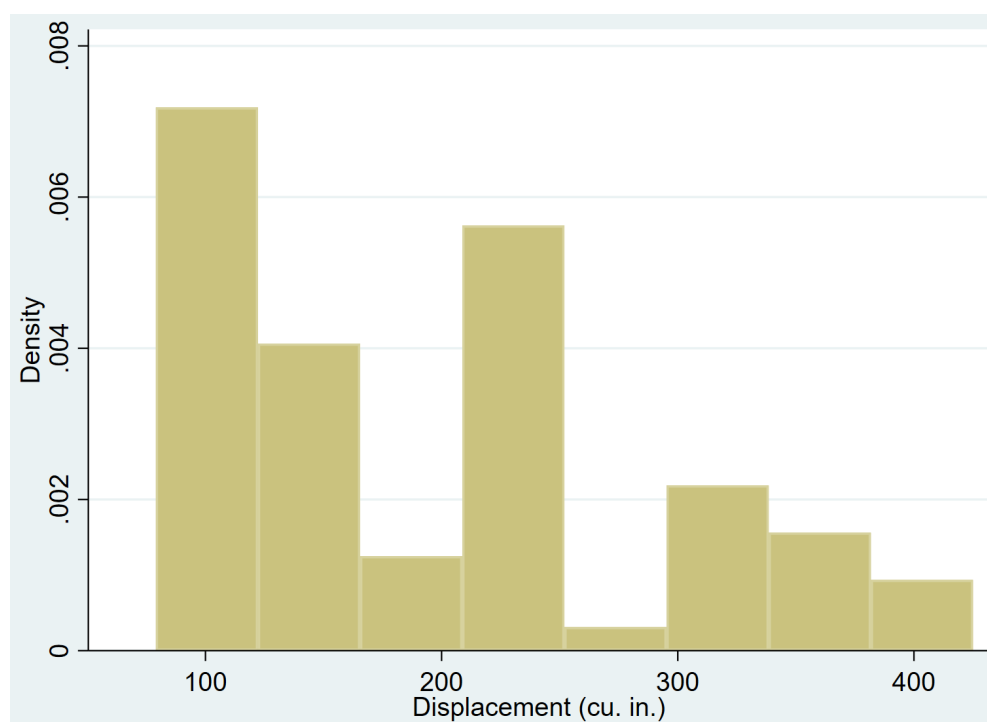
methods to ensure the most robust evidence regarding the distributional characteristics of the variables under investigation.

Method 1: Visual Assessment Using Histograms

The most accessible and intuitive initial step in checking for [normality](#) involves a visual inspection, typically achieved by constructing a [histogram](#). A histogram graphically displays the frequency distribution of a continuous variable, offering immediate insights into its overall shape, central tendency, symmetry, and spread. For a variable that is perfectly normally distributed, its histogram should closely mirror the symmetric, unimodal bell curve, with the highest frequencies concentrated near the mean and observations smoothly tapering off toward the extreme tails.

To generate a histogram for the variable *displacement* (engine displacement in cubic inches) from the *auto* dataset, we simply use the **hist** command in Stata. This preliminary visual check is highly effective at revealing gross violations of normality, such as substantial skewness, the presence of multiple modes (bimodality), or extreme outliers, often before any formal statistical tests are required.

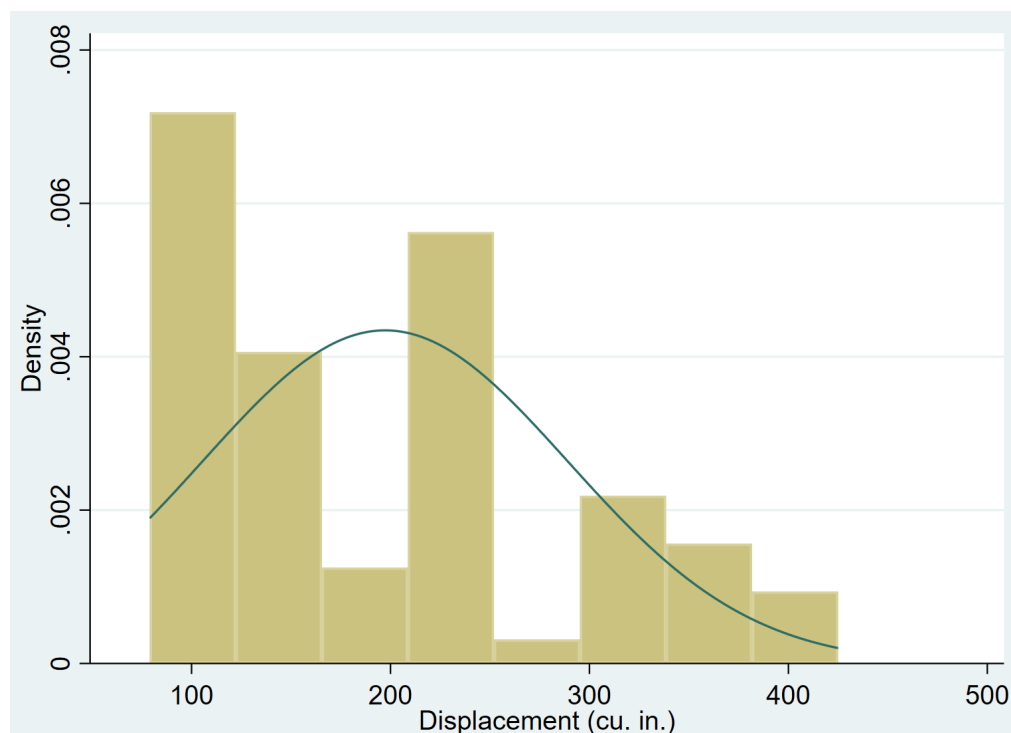
hist displacement



To facilitate a clearer comparison against the theoretical ideal, it is standard practice to superimpose a [normal density curve](#) onto the histogram. This curve illustrates the exact shape the distribution would assume if the variable were perfectly normal, possessing the same mean and

standard deviation as the sample data. In Stata, this enhancement is achieved by adding the **normal** option to the **hist** command, providing a benchmark against which the observed data can be measured.

hist displacement, normal



The enhanced histogram above clearly illustrates that the variable *displacement* deviates noticeably from the theoretical normal curve. Specifically, the distribution exhibits a pronounced **positive skew** (or right skew), indicated by the concentration of most observations toward the lower values on the left side, and a long, drawn-out tail extending toward the higher values. This compelling visual evidence strongly suggests that the assumption of normality is violated for this variable, warranting confirmation through more rigorous formal statistical testing.

Method 2: The Powerful Shapiro-Wilk Test

While visual methods offer essential preliminary insight, researchers require formal statistical procedures to determine objectively whether observed deviations from normality are statistically significant. The [Shapiro-Wilk Test](#) stands out as one of the most respected and powerful tests for assessing normality, especially effective for small to moderately sized samples (N generally between 4 and 2,000 observations). The test quantifies how well the observed data correlates with a theoretical perfectly normal distribution.

The foundation of the [Shapiro-Wilk Test](#) is its articulation of the [null hypothesis](#) (H_0), which posits

that the sample data originated from a normally distributed population. To execute this test in Stata, one uses the **swilk** command followed by the name of the variable. Due to its computational complexity, it is important to remember the limitation that this test is most appropriate when the total number of observations (N) is within the specified range of 4 to 2,000.

We apply the [Shapiro-Wilk Test](#) to the *displacement* variable using the following Stata syntax:

swilk displacement

. swilk displacement

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
displacement	74	0.92542	4.803	3.423	0.00031

Interpreting the output involves comparing the calculated [p-value](#) (labeled Prob>z) against the predetermined significance level (alpha, conventionally set at 0.05). Let's dissect the key components:

Obs: 74. The sample size used for the test.

W: 0.92542. The test statistic W. Values closer to 1 suggest a higher degree of conformity to the [normal distribution](#).

Prob>z: 0.00031. The associated [p-value](#).

Since the [p-value](#) (0.00031) is significantly less than the 0.05 threshold, we have sufficient statistical evidence to reject the [null hypothesis](#). We therefore conclude definitively that the variable *displacement* is statistically non-normally distributed, confirming the initial visual assessment from the histogram. A notable advantage of the **swilk** command is its ability to test multiple variables simultaneously by listing them after the command, facilitating rapid comparative assessment, as demonstrated below.

swilk displacement mpg length

```
. swilk displacement mpg length
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
displacement	74	0.92542	4.803	3.423	0.00031
mpg	74	0.94821	3.335	2.627	0.00430
length	74	0.97165	1.825	1.313	0.09461

Using an alpha level of 0.05 for this multiple variable test, we observe that *displacement* ($p=0.00031$) and *mpg* ($p=0.00001$) are both highly non-normally distributed, as their p-values are far below 0.05. Conversely, the p-value for *length* ($p=0.17646$) is greater than 0.05, meaning we fail to reject the [null hypothesis](#). Therefore, based on the [Shapiro-Wilk Test](#), we cannot conclude that the variable *length* significantly deviates from a normal distribution.

Method 3: The Shapiro-Francia Test for Larger Samples

The [Shapiro-Francia Test](#) provides another respected formal statistical method for evaluating normality. This test was developed as a simplified variation of the Shapiro-Wilk test, making it particularly advantageous for application to larger datasets where the original Shapiro-Wilk test might become computationally inefficient or inappropriate due to its strict upper limit. The Shapiro-Francia test is suitable for sample sizes ranging from 10 up to 5,000 observations.

The underlying logic of the [Shapiro-Francia Test](#) is identical to that of its predecessor: the [null hypothesis](#) asserts that the variable is drawn from a normally distributed population. Consequently, the decision rule remains unchanged: if the resulting [p-value](#) falls below the chosen significance level (e.g., $\alpha = 0.05$), the null hypothesis is rejected, providing strong evidence of non-normality.

In Stata, the command used to execute this test is **sfrancia**. We will apply this command to the *displacement* variable to compare its findings directly with the Shapiro-Wilk results and the visual evidence:

```
sfrancia displacement
```

```
. sfrancia displacement
```

Shapiro-Francia W' test for normal data

Variable	Obs	W'	V'	z	Prob>z
displacement	74	0.93011	4.975	3.110	0.00094

The structure of this output closely resembles the Shapiro-Wilk results, though the test statistic is denoted as W':

Obs: 74. The sample size.

W': 0.93011. The [Shapiro-Francia Test](#) statistic.

Prob>z: 0.00094. The corresponding p-value.

With a [p-value](#) of 0.00094, which is substantially below 0.05, we once again reject the null hypothesis. This confirms the consistency across the methods: the visual inspection, the Shapiro-Wilk test, and the Shapiro-Francia test all converge on the conclusion that the variable *displacement* is not normally distributed. Similar to the other formal tests, **sfrancia** supports the simultaneous assessment of multiple variables by listing them in the command line.

Method 4: The Skewness and Kurtosis Test

A third widely used formal method for normality assessment is the [Skewness and Kurtosis Test](#), which is often implemented as the D'Agostino-Pearson omnibus test in statistical software. This test uniquely focuses on whether the variable's measured skewness (a measure of symmetry) and kurtosis (a measure of the peak's sharpness and the thickness of the tails) significantly differ from the values expected under a standard [normal distribution](#). Under perfect normality, the skewness coefficient should be zero, and the excess kurtosis (kurtosis minus 3) should also be zero.

The [Skewness and Kurtosis Test](#) integrates the deviations found in both symmetry and tail weight into a single omnibus test statistic, which is distributed as a chi-squared distribution. The [null hypothesis](#) remains that the variable is normally distributed. If the test yields a [p-value](#) below the significance level, we reject H0, concluding that the distribution is non-normal due to either excessive skewness, problematic kurtosis, or a combination of both. This test requires a minimum sample size of 8 observations.

In Stata, we utilize the **sktest** command to perform this joint assessment on the *displacement* variable:

```
sktest displacement
```

```
. sktest displacement
```

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
displacement	74	0.0337	0.2048	5.81	0.0547

Interpretation relies on the combined test result found at the bottom of the output table:

Obs: 74. The number of observations.

adj chi(2): 5.81. The Chi-Square test statistic for the combined assessment, with 2 degrees of freedom.

Prob>chi2: 0.0547. The p-value associated with the overall test.

In a divergence from the highly significant results of the Shapiro-Wilk and Shapiro-Francia tests, the overall p-value here is 0.0547. Since 0.0547 is marginally greater than the conventional alpha level of 0.05, we must technically **fail to reject** the [null hypothesis](#) based solely on the [Skewness and Kurtosis Test](#). This means that, according to this specific test, we lack sufficient statistical evidence to conclude that *displacement* is non-normally distributed at the 5% significance level. This outcome underscores the necessity of employing multiple diagnostic methods for a comprehensive and cautious assessment of normality. The **sktest** command, like the others, also allows for simultaneous testing of multiple variables.

Summary and Best Practices for Normality Testing

The four methods reviewed--visual inspection via histograms and the formal [Shapiro-Wilk Test](#), [Shapiro-Francia Test](#), and [Skewness and Kurtosis Test](#)--constitute a robust toolkit for diagnosing distributional assumptions in Stata. In practical statistical analysis, relying on a single test is discouraged; it is highly recommended to integrate multiple techniques. Visual diagnostics are essential for understanding the nature and source of the deviation (e.g., identifying severe outliers or bimodality), while formal tests provide the necessary objective statistical confirmation for decision-making.

As a general guideline, the [Shapiro-Wilk Test](#) is often the preferred choice due to its superior power across a broad range of distributions, particularly for small to medium sample sizes ($N < 2000$). For significantly larger datasets, the [Shapiro-Francia Test](#) offers a computationally efficient and equally powerful alternative. The Skewness and Kurtosis Test is valuable for specifically pinpointing whether asymmetry or tail thickness is the primary issue, but analysts should be mindful that it can occasionally produce conflicting results compared to the rank-based tests, as demonstrated in our example.

When the assumption of [normal distribution](#) is unequivocally rejected by multiple tests, researchers must proceed with caution. The appropriate subsequent analytical step is determined by the severity of the violation and the size of the sample. Options typically include attempting data transformations (such as log or inverse transformations), employing robust statistical methods that are less sensitive to non-normality, or pivoting completely to non-parametric [statistical tests](#) that do not require distributional assumptions, thereby ensuring the validity and integrity of the study's conclusions.