

# How to Easily Perform a Shapiro-Wilk Test in SAS

Authored by  
**stats writer**

December 1, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Easily Perform a Shapiro-Wilk Test in SAS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103269>

To accurately perform a Shapiro-Wilk test in SAS, analysts must utilize the powerful statistical procedure known as PROC UNIVARIATE. This procedure is specifically designed for comprehensive analysis of single variables, including detailed assessments of data normality. By invoking this procedure and specifying the appropriate options, you can input your dataset and initiate a thorough examination of its distribution characteristics, which is a foundational requirement for many parametric statistical models. The test is crucial for determining if your sample data originates from a population that follows a normal distribution.

The output generated by the test is highly informative, providing both a test statistic ( $W$ ) and the corresponding p-value. These metrics are the key outputs used to formally assess the dataset's adherence to normality. The interpretation of the p-value is straightforward: it dictates whether we should reject the hypothesis that the data is normally distributed. Understanding how to correctly execute and interpret this test is fundamental for ensuring the validity and reliability of subsequent statistical inferences.

## Understanding the Importance of Normality Testing

Before executing any parametric tests, such as t-tests or ANOVA, it is absolutely essential to confirm that the underlying data satisfies the assumption of normality. The Shapiro-Wilk test is universally recognized as one of the most powerful and reliable tools for this purpose, particularly when dealing with smaller sample sizes ( $N < 50$ ). This test compares the distribution of the sample data against a theoretical normal distribution with the same mean and variance, providing a quantitative measure of similarity.

When data deviates significantly from a normal distribution, applying parametric methods can lead to incorrect conclusions and inflated Type I or Type II errors. Therefore, researchers often rely on the results of the Shapiro-Wilk test to guide their choice between parametric and non-parametric statistical procedures. If the normality assumption is violated, transformation methods or robust non-parametric alternatives must be considered to ensure statistical rigor.

This comprehensive guide details the precise, step-by-step methodology required to perform the Shapiro-Wilk test within the SAS environment. We will cover data preparation, the necessary procedure syntax, and the critical interpretation of the resulting statistical output to determine if your dataset adheres to the requirements of normality.

The Shapiro-Wilk test is used to determine whether or not a dataset follows a normal distribution. The following step-by-step example shows how to perform this crucial test for a dataset in SAS.

## Defining the Shapiro-Wilk Hypotheses

Every statistical hypothesis test is built upon a formal set of statements--the null hypothesis and the alternative hypothesis. For the Shapiro-Wilk test, these hypotheses are specifically structured to assess the assumption of population normality. Understanding these formal definitions is paramount to correctly interpreting the test's outcome, as the final decision hinges on whether there is sufficient evidence to reject the baseline assumption.

The testing framework is defined as follows, where the test statistic is calculated under the assumption that the null hypothesis is true. The test is designed to find evidence against the null hypothesis:

**H<sub>0</sub>**: The population from which the sample data was drawn is normally distributed. This is the assumption of normality.

**H<sub>A</sub>**: The population from which the sample data was drawn is *not* normally distributed. This suggests a deviation requiring non-parametric methods.

If the calculated p-value is sufficiently small (typically less than the predetermined significance level,  $\alpha = 0.05$ ), we reject H<sub>0</sub> and conclude that the data is not normally distributed. Conversely, if the p-value is large, we lack sufficient evidence to reject the null hypothesis, thus assuming the data is acceptably normal.

## Step 1: Preparing the Sample Dataset in SAS

The first practical step in conducting any analysis in SAS involves creating or importing the necessary dataset. For demonstrative purposes, we will construct a simple dataset named `my_data` containing a single variable, `x`, which represents 15 distinct numerical observations. This small sample size is ideal for illustrating the strengths of the Shapiro-Wilk procedure, although the test is applicable across a range of sample sizes.

Data creation in SAS is achieved using the standard `DATA` step, followed by the `DATALINES` statement to input the raw values directly into the program. Following the data input, it is always recommended practice to view the newly created dataset using `PROC PRINT` to ensure the data was read correctly and accurately reflects the intended observations. This step verifies the integrity of the data before proceeding to the statistical analysis.

The following code block executes the data creation and subsequent printing of the dataset, confirming the 15 observations are ready for analysis. Note the use of specialized SAS keywords such as `data`, `input`, and `datalines`:

```
/*create dataset*/  
data my_data;  
input x;
```

```
datalines;
```

```
3
```

```
3
```

```
4
```

```
6
```

```
7
```

```
8
```

```
8
```

```
9
```

```
12
```

```
14
```

```
15
```

```
15
```

```
17
```

```
20
```

```
21
```

```
;
```

```
run;
```

```
/*view dataset*/
```

```
proc print data=my_data;
```

ARABPSYCHOLOGY.COM

Obs	x
1	3
2	3
3	4
4	6
5	7
6	8
7	8
8	9
9	12
10	14
11	15
12	15
13	17
14	20
15	21

## Step 2: Executing the Normality Test with PROC UNIVARIATE

Next, we'll use **proc univariate** with the **normal** command to perform a Shapiro-Wilk test for normality. Once the data is successfully loaded, the normality assessment is performed using the versatile **PROC UNIVARIATE** procedure. This procedure provides detailed descriptive statistics, graphs, and, critically, tests for distribution fitting. To specifically request the Shapiro-Wilk test and other goodness-of-fit tests for normality, we must include the **NORMAL** option in the procedure statement.

The syntax is concise but powerful. The **PROC UNIVARIATE** statement identifies the dataset to be analyzed (**data=my\_data**), and the subsequent keyword **NORMAL** instructs SAS to calculate the specific statistics required for assessing the normality assumption. It is important to note that when the **NORMAL** option is specified, **PROC UNIVARIATE** automatically performs a battery of normality checks, not just the Shapiro-Wilk test, providing a comprehensive view of the data's distribution profile.

Executing the following code will generate extensive output, which includes the descriptive statistics, moments, and, most importantly, the table containing the results of the normality tests required for interpretation:

```
/*perform Shapiro-Wilk test*/
```

```
proc univariate data=my_data normal;
run;
```

The UNIVARIATE Procedure  
Variable: x

Moments			
N	15	Sum Weights	15
Mean	10.8	Sum Observations	162
Std Deviation	5.96657356	Variance	35.6
Skewness	0.30348727	Kurtosis	-1.1151814
Uncorrected SS	2248	Corrected SS	498.4
Coeff Variation	55.2460514	Std Error Mean	1.54056267

Basic Statistical Measures			
Location		Variability	
Mean	10.80000	Std Deviation	5.96657
Median	9.00000	Variance	35.60000
Mode	3.00000	Range	18.00000
		Interquartile Range	9.00000

Note: The mode displayed is the smallest of 3 modes with a count of 2.

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	7.010426	Pr >  t	<.0001
Sign	M	7.5	Pr >=  M	<.0001
Signed Rank	S	60	Pr >=  S	<.0001

Tests for Normality				
Test		Statistic	p Value	
Shapiro-Wilk	W	0.936921	Pr < W	0.3452
Kolmogorov-Smirnov	D	0.151886	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.052851	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.331213	Pr > A-Sq	>0.2500

## Reviewing the Tests for Normality Table

The output provides us with a ton of information, but the only table we need to look at is the one

titled **Tests for Normality**. This table is where PROC UNIVARIATE aggregates the results of various goodness-of-fit assessments. Focusing on this section allows for rapid identification of the key metrics needed to evaluate the distributional characteristics of the variable  $x$ .

This table provides the test statistics and p-values for several normality tests including: While the Shapiro-Wilk test is our main concern, this comprehensive output ensures we have multiple checks for confirmation:

The Shapiro-Wilk Test

The Kolmogorov-Smirnov Test

The Cramer-von Mises Test

The Anderson-Darling Test

Each of these tests is based on comparing the observed data distribution to the theoretical normal distribution. The results presented in the table are essential for drawing a formal conclusion about the data's suitability for parametric analysis.

## Interpreting the Shapiro-Wilk P-Value

From this table we can see that the p-value for the Shapiro-Wilk test is **.3452**. This value represents the probability of observing our data (or data more extreme) if the null hypothesis ( $H_0$ : data is normal) were true. We must recall the fundamental structure of the Shapiro-Wilk test:

**$H_0$** : The data is normally distributed.

**$H_A$** : The data is *not* normally distributed.

The critical decision rule in hypothesis testing involves comparing this calculated p-value to the chosen level of significance ( $\alpha$ ), conventionally set at 0.05. Since the p-value (**.3452**) is not less than the threshold of 0.05, we fail to reject the null hypothesis. This lack of evidence against  $H_0$  means we conclude that the data does not significantly deviate from a normal distribution.

In other words, because  $0.3452 > 0.05$ , it is statistically safe to assume that the dataset `my_data` is normally distributed. This validates the use of parametric statistical methods for further analysis on the variable  $x$ .

## Conclusion: Validating Data for Parametric Analysis

The successful execution of the Shapiro-Wilk test using PROC UNIVARIATE in SAS confirms that the sample data analyzed here adheres to the critical assumption of the normal distribution. A high p-value led us to retain the null hypothesis, certifying the data as suitable for subsequent parametric modeling. This procedure is an indispensable first step in rigorous quantitative analysis.

The ability to reliably test and confirm distributional assumptions is paramount for any data scientist or statistician working in the SAS environment. Mastery of the `NORMAL` option within `PROC UNIVARIATE` not only facilitates accurate normality checks but also opens the door to numerous other descriptive and exploratory techniques available within the same procedure.

The following tutorials explain how to perform other common statistical tests in SAS:

ARABPSYCHOLOGY.COM