Learning to Test for Normality in SPSS: A Step-by-Step Guide

Authored by

Mohammed looti

November 8, 2025

RECOMMENDED CITATION

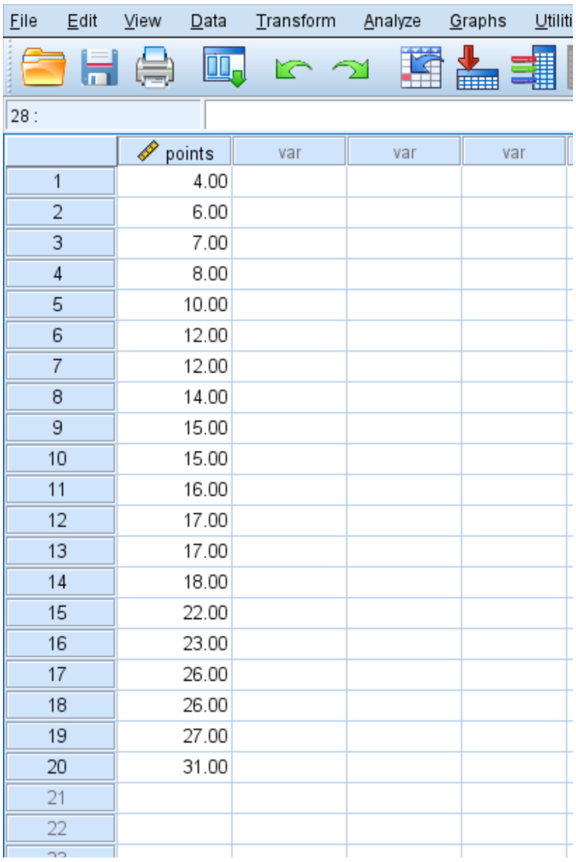
Mohammed looti (2025). *Learning to Test for Normality in SPSS: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from https://statistics.arabpsychology.com/?p=12930

Understanding the underlying distribution of data is a fundamental prerequisite for many advanced statistical tests. Specifically, numerous parametric procedures, such as the independent samples t-test or ANOVA, rely heavily on the assumption that the variables are normally distributed within the population. Failure to confirm this assumption can lead to unreliable results, inaccurate standard errors, and flawed conclusions. Therefore, rigorously testing for normality is a critical initial step in the data analysis workflow.

This comprehensive tutorial is designed to guide data analysts and researchers through two distinct, yet complementary, methods available within SPSS (Statistical Package for the Social Sciences) to effectively assess the distribution of a continuous variable. We will explore both visual assessment techniques, which provide intuitive insights, and formal statistical hypothesis testing, which offers a definitive, objective measure of compliance with the normal distribution assumption.

We will utilize a practical, real-world example throughout this guide to illustrate the process clearly. The dataset contains the average points per game scored by 20 distinct basketball players, providing a continuous variable that we need to examine for distributional conformity.

The following table displays the structure and values of the sample dataset used for this demonstration:

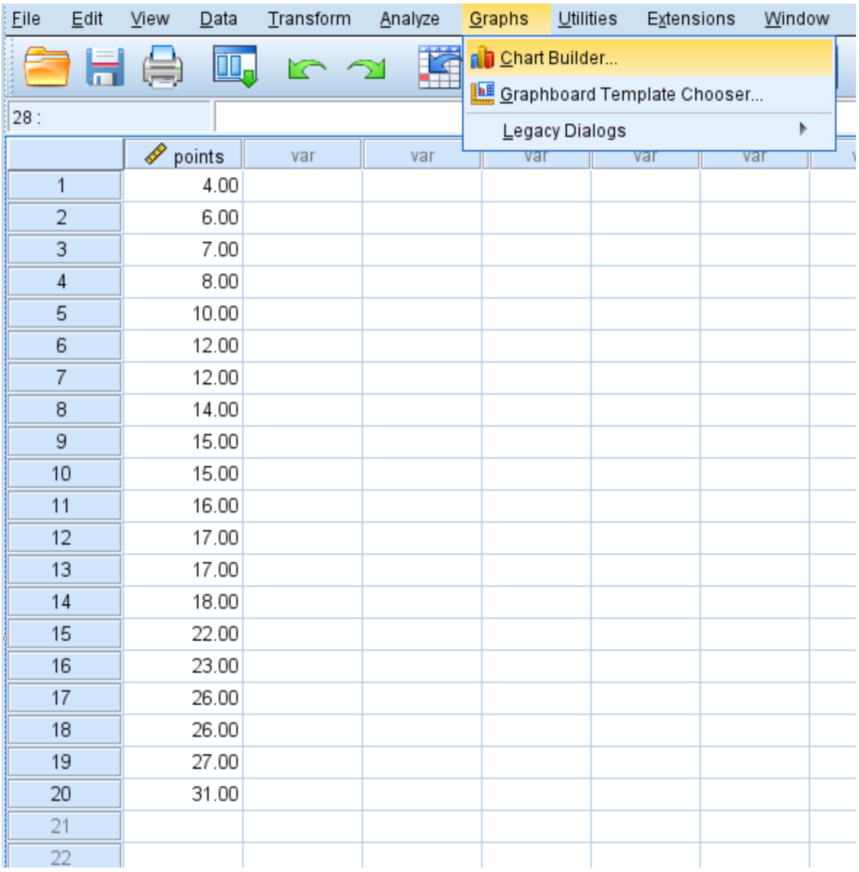


	points	var	var	var
1	4.00			
2	6.00			
3	7.00			
4	8.00			
5	10.00			
6	12.00			
7	12.00			
8	14.00			
9	15.00			
10	15.00			
11	16.00			
12	17.00			
13	17.00			
14	18.00			
15	22.00			
16	23.00			
17	26.00			
18	26.00			
19	27.00			
20	31.00			
21				
22				

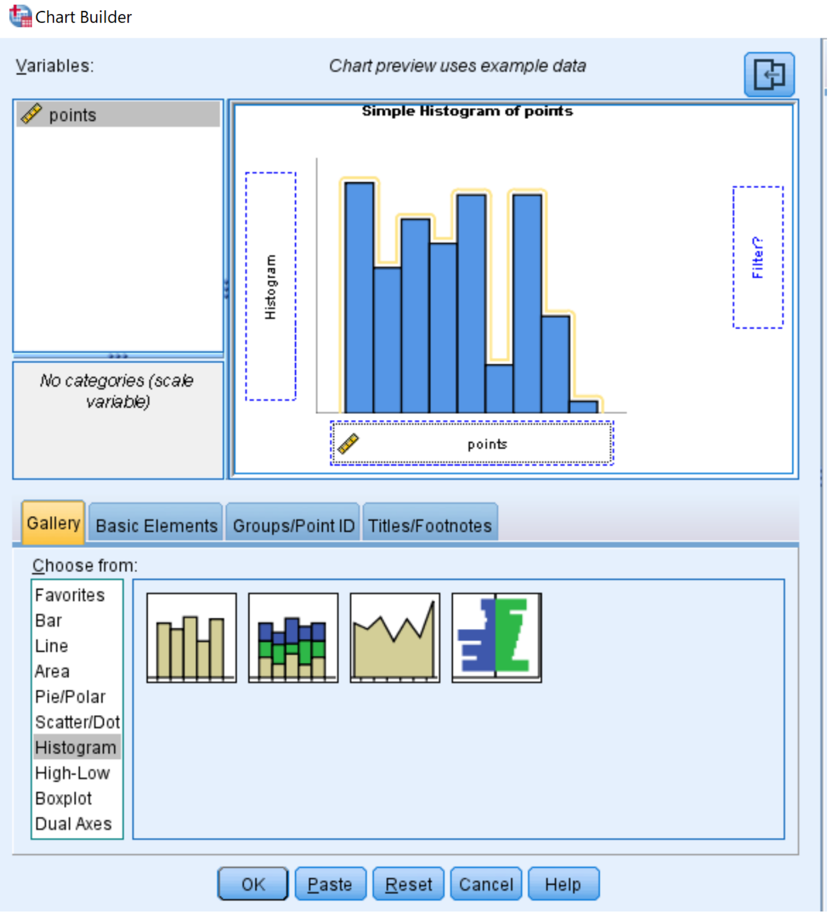
Method 1: Visual Inspection Using Histograms in SPSS

The most immediate and intuitive method for assessing whether a variable follows a normal distribution is through visual inspection. Creating a histogram allows us to graphically represent the frequency distribution of the data. If the variable is truly normally distributed, the resultant histogram should approximate the classic "bell" shape, characterized by a concentration of values near the center (the mean) and a symmetrical tapering off toward the extremities (the tails).

To generate a histogram for the variable **points** in our basketball dataset, we initiate the process by navigating through the Chart Builder interface within SPSS. Begin by selecting the **Graphs** tab from the main menu, followed by clicking on **Chart Builder**. This tool provides a drag-and-drop interface for creating various graphical representations of data distributions.



In the subsequent dialog box that appears, look for the **Choose from** list located at the bottom left. Select the **Histogram** option from this menu. Once selected, click and drag the histogram icon into the main editing window, which is the large canvas area. The next critical step is to specify the variable whose distribution we wish to analyze. Locate the variable **points** in the list of available variables and drag it directly onto the x-axis placeholder within the editing window, thereby defining the data range for the plot:

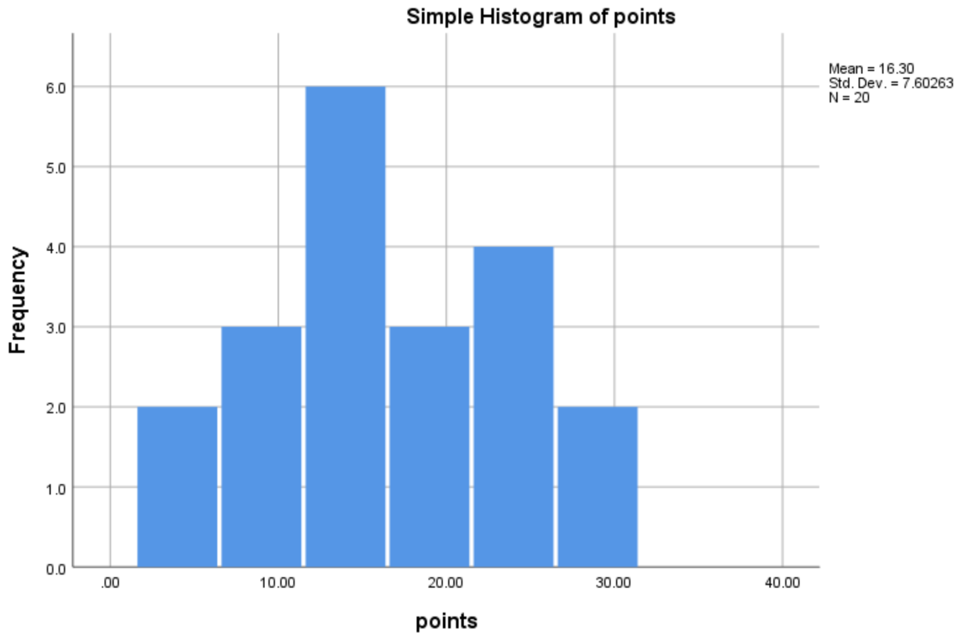


Upon completing the setup and clicking **OK**, SPSS will execute the command and generate the visual output. The resulting histogram will then appear in the SPSS Output Viewer, allowing for immediate visual assessment of the distribution shape.

Interpreting the Histogram for Distribution Shape

Once the histogram is displayed, as shown below, the analyst must subjectively evaluate its shape against the ideal normal distribution curve. This visual evaluation is based on symmetry, kurtosis (peakedness), and the presence of outliers or multiple modes. A perfectly normally distributed variable is rarely found in empirical data; instead, we look for a reasonable approximation.

➔ **GGraph**



Analyzing the visual representation above, we observe that the variable **points** does not exhibit perfect symmetry. However, the distribution generally concentrates the majority of observations--specifically, most players' scores--in the central region, approximately between 10 and 20 points per game. The frequencies decrease gradually as the scores move toward the lower and upper bounds of the data range, giving the plot a rough bell shape.

While visual methods like the histogram are excellent for gaining an initial understanding of the data's shape and identifying potential skewness or modality issues, it is important to recognize their limitations. Visual assessment is inherently subjective and lacks the statistical rigor necessary for formal hypothesis testing. It is a quick and effective diagnostic tool, but it should typically be complemented by objective, quantitative measures, particularly when sample sizes are either very small or very large.

Method 2: Employing Formal Normality Tests

To move beyond subjective visual assessment, we must employ formal statistical tests designed to quantify the probability that a sample was drawn from a population that follows a normal distribution. These tests establish a null hypothesis (H_0) stating that the data are normally distributed. Consequently, rejecting the null hypothesis indicates a significant deviation from normality.

SPSS provides output for two of the most widely recognized formal tests for normality: the Shapiro-Wilk Test and the Kolmogorov-Smirnov Test. Both tests compare the observed cumulative distribution function of the sample data against the expected cumulative distribution function if the data were perfectly normal.

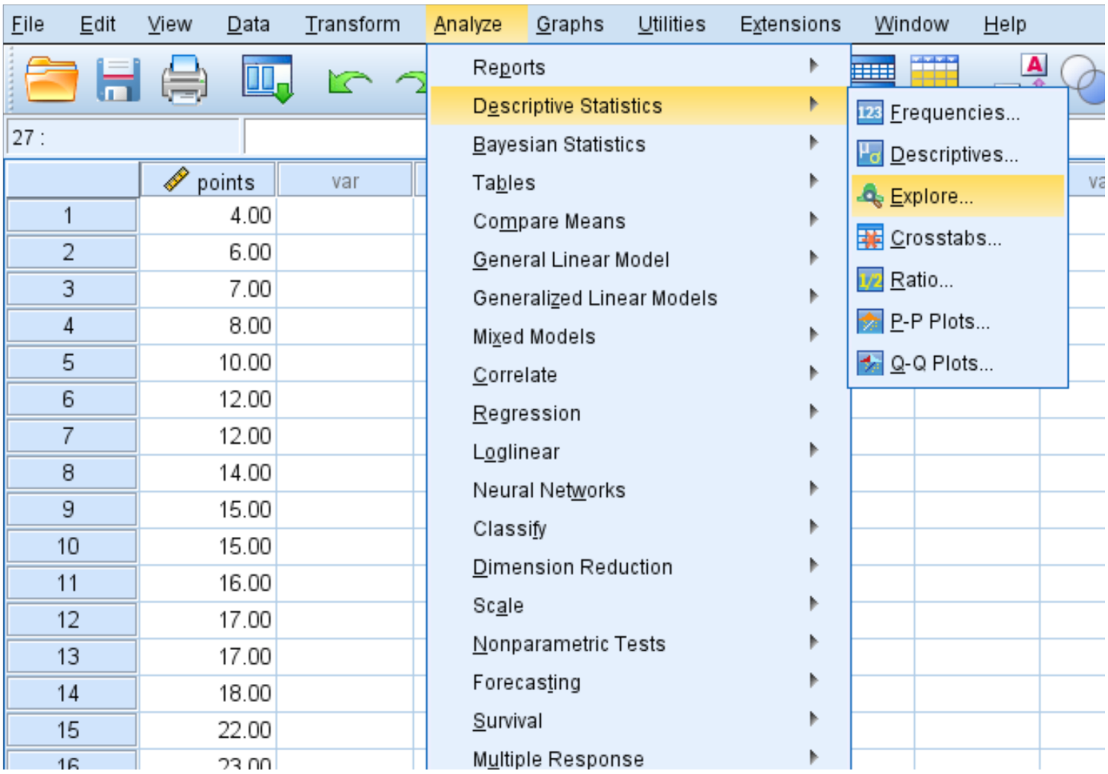
The **Shapiro-Wilk Test**: This test is generally preferred, especially for smaller sample sizes (typically N < 50). It is considered to have greater statistical power in detecting departures from normality compared to the Kolmogorov-Smirnov test.

The **Kolmogorov-Smirnov Test** (often reported with the Lilliefors correction): This test is often recommended for larger sample sizes (N > 50) and compares the largest absolute difference between the empirical distribution function and the theoretical normal distribution function.

The choice between these two often depends on the sample size, though SPSS conveniently calculates both simultaneously, allowing the researcher to consider both results in their interpretation.

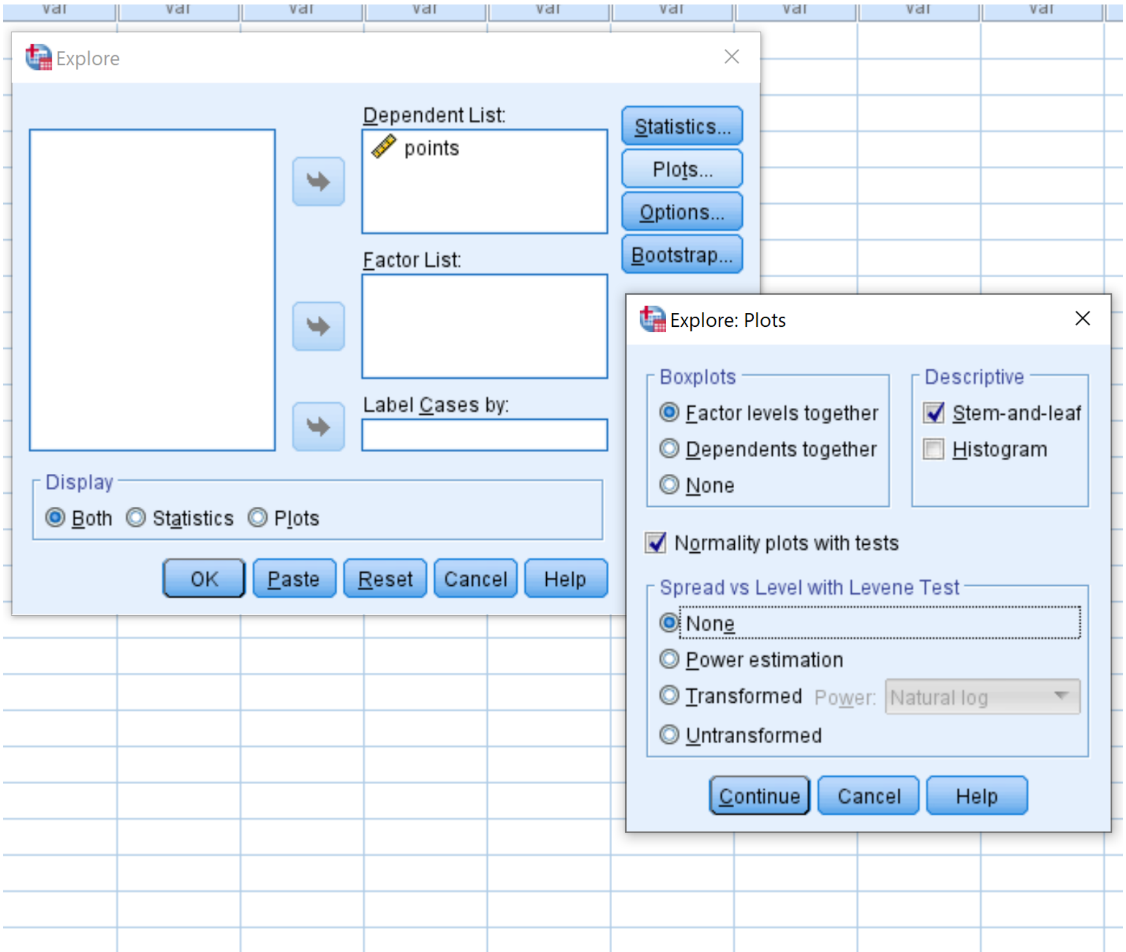
Executing the Kolmogorov-Smirnov and Shapiro-Wilk Tests in SPSS

To execute both formal tests simultaneously within SPSS, we use the Explore function, which is typically utilized for comprehensive descriptive analyses and initial assumption testing. Begin by navigating the menu system: click the **Analyze** tab, hover over **Descriptive Statistics**, and then select **Explore**.



In the subsequent dialog box titled "Explore," the variable **points** must be moved into the box labeled **Dependent List**. This designates **points** as the variable whose distribution we intend to analyze. Once the dependent variable is specified, click the **Plots** button located on the right side of the dialog box. This action opens the "Explore: Plots" sub-menu.

Within the Plots sub-menu, it is crucial to ensure that the checkbox corresponding to **Normality plots with tests** is explicitly selected. This command instructs SPSS to generate the Q-Q plots (another visual check) and, more importantly for this method, the quantitative results for the Shapiro-Wilk Test and the Kolmogorov-Smirnov Test. After confirming this selection, click **Continue** to close the Plots menu, and finally, click **OK** in the main "Explore" dialog box to run the analysis.



Analyzing and Concluding from Test Results

After executing the steps, the SPSS Output Viewer will display the results, including a table titled "Tests of Normality." This table contains the essential information needed to make a formal determination regarding the distribution of the **points** variable.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
points	.113	20	.200*	.967	20	.699

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

The interpretation of these formal tests centers on the corresponding significance value, or the p-value, for each test. As established earlier, the null hypothesis (H_0) assumes that the data are normally distributed. We typically compare the reported p-value against a predetermined significance level ($alpha$), conventionally set at 0.05. If the p-value is less than $alpha$ (e.g., $p < 0.05$), we reject H_0, concluding that the data are significantly non-normal. Conversely, if the p-value is greater than or equal to $alpha$ ($p ge 0.05$), we fail to reject H_0, meaning there is insufficient evidence to conclude that the distribution is non-normal.

Let us examine the specific output for the variable **points**:

Kolmogorov-Smirnov Test (Lilliefors Significance Correction):

Test statistic: **.113**
p-value (Sig.): **.200**

Shapiro-Wilk Test:

Test statistic: **.967**
p-value (Sig.): **.699**

In both cases, the reported significance values (p-values) are substantially greater than the standard threshold of 0.05. For the Kolmogorov-Smirnov Test, $p = 0.200$, and for the Shapiro-Wilk Test, $p = 0.699$. Since neither p-value falls below 0.05, we must conclude that we lack sufficient statistical evidence to reject the null hypothesis of normality.

This finding is critical for subsequent inferential statistical tests. If the researcher intended to perform a parametric test--one that assumes the dependent variable is normally distributed--the results of the formal normality tests confirm that the variable **points** satisfies this distributional assumption. Combining this quantitative evidence with the initial visual assessment from the histogram provides a robust basis for proceeding with parametric statistical analysis.