

How to Check for Normal Data Distribution in SPSS

Authored by
stats writer

March 14, 2026

RECOMMENDED CITATION

stats writer (2026). *How to Check for Normal Data Distribution in SPSS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=135777>

The determination of whether a dataset follows a **normal distribution** is a fundamental prerequisite for many forms of quantitative analysis. In the context of the **SPSS** (Statistical Package for the Social Sciences) software, evaluating normality involves a multi-faceted approach that examines the **shape**, symmetry, and **kurtosis** of the data. This process is essential because many **parametric statistics**, such as t-tests and ANOVA, rely on the assumption that the underlying population is normally distributed. By performing rigorous checks, researchers can ensure the **reliability** and **validity** of their findings, ultimately leading to more accurate **statistical inferences**.

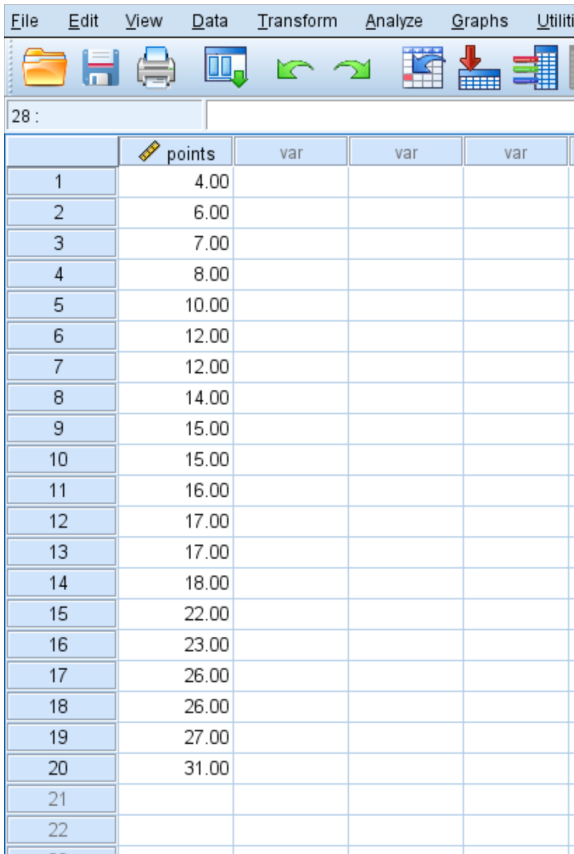
Test for Normality in SPSS

The Critical Role of Normality in Statistical Testing

In the realm of **quantitative research**, the assumption of normality is often the gatekeeper for selecting appropriate analytical techniques. Many **statistical tests** require that at least one variable follows a bell-shaped curve to ensure the results are theoretically sound and representative of the population. When data deviates significantly from this **probability distribution**, the **standard error** of the estimate may be compromised, leading to an increased risk of Type I or Type II errors. Therefore, before conducting any high-level modeling, a researcher must rigorously assess the distribution of their variables within **SPSS**.

This comprehensive tutorial is designed to guide you through two distinct and highly effective methods for evaluating normality among variables. Whether you are a student working on a thesis or a professional data scientist, understanding these workflows is vital for maintaining **data integrity**. By the end of this guide, you will be proficient in both visual inspections and formal mathematical assessments, providing you with a holistic view of your data's characteristics. We will explore how to identify **skewness** and other distributional properties that might influence your research outcomes.

To provide a clear and practical demonstration, each method discussed in this guide will utilize a sample dataset. This dataset tracks the average points per game scored by a cohort of 20 different basketball players. This specific metric is ideal for illustrating **continuous variables** that often exhibit near-normal tendencies but frequently contain slight deviations that require careful interpretation. Below is a representation of the raw data as it appears in the **SPSS** Data View window.



	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				

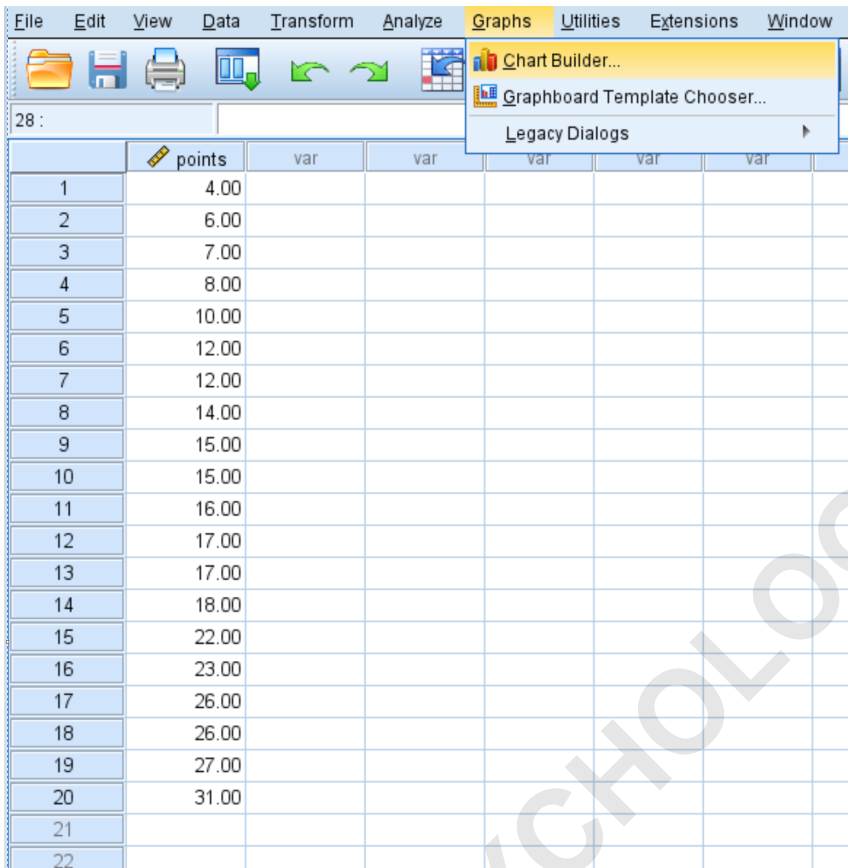
Utilizing Histograms for Visual Assessment

One of the most intuitive ways to evaluate whether a variable is normally distributed is to generate a **histogram**. A histogram provides a graphical representation of the **frequency distribution** of a variable, allowing the researcher to see where values are concentrated. In a perfectly **normal distribution**, the histogram should exhibit a distinctive "bell" shape. This means that the majority of the data points are clustered around the **mean**, with the frequency of values gradually tapering off as they move toward the upper and lower **tails**.

Visual inspection is often the first line of defense against **outliers** and extreme non-normality. While it lacks the mathematical precision of a formal test, a histogram can reveal patterns that numbers alone might hide, such as bimodal distributions or heavy-tailedness. For our basketball dataset, creating a histogram allows us to quickly gauge if the scoring averages are balanced or if there is a significant **skewness** toward exceptionally high or low scorers. It serves as a qualitative check that complements quantitative findings.

To initiate the creation of a histogram for the basketball scoring dataset in **SPSS**, you should navigate to the top navigation bar. Click on the **Graphs** tab, which houses a variety of visualization tools. From the dropdown menu, select **Chart Builder**. This tool provides a highly customizable

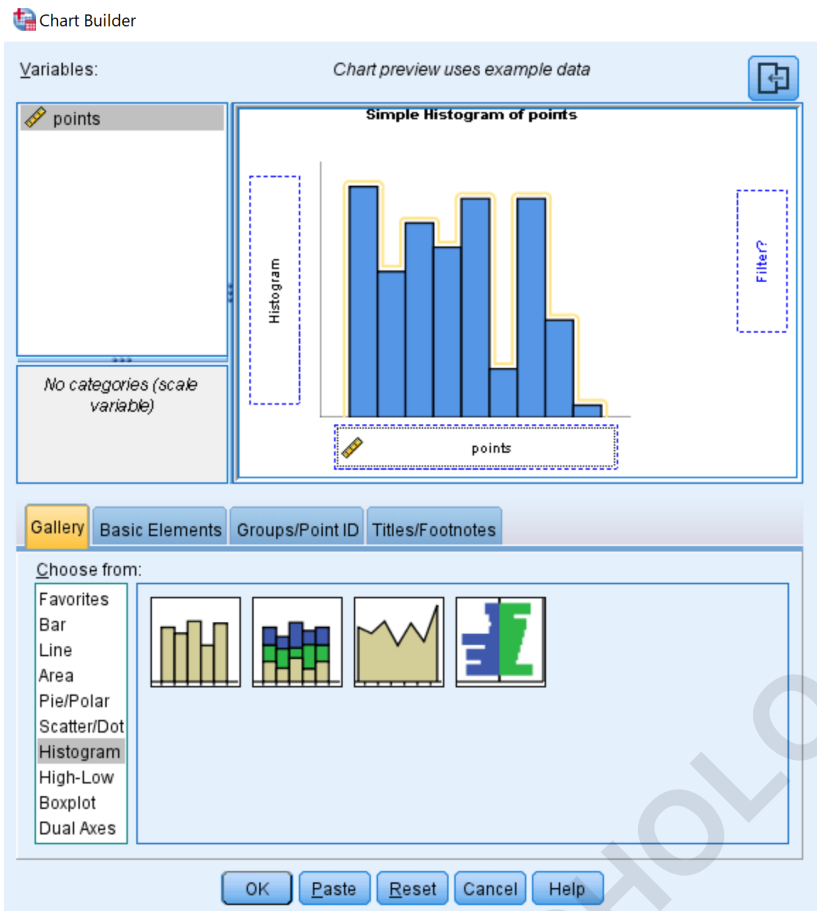
interface for generating professional-grade graphics that are suitable for academic publications or business reports.



Step-by-Step Configuration in Chart Builder

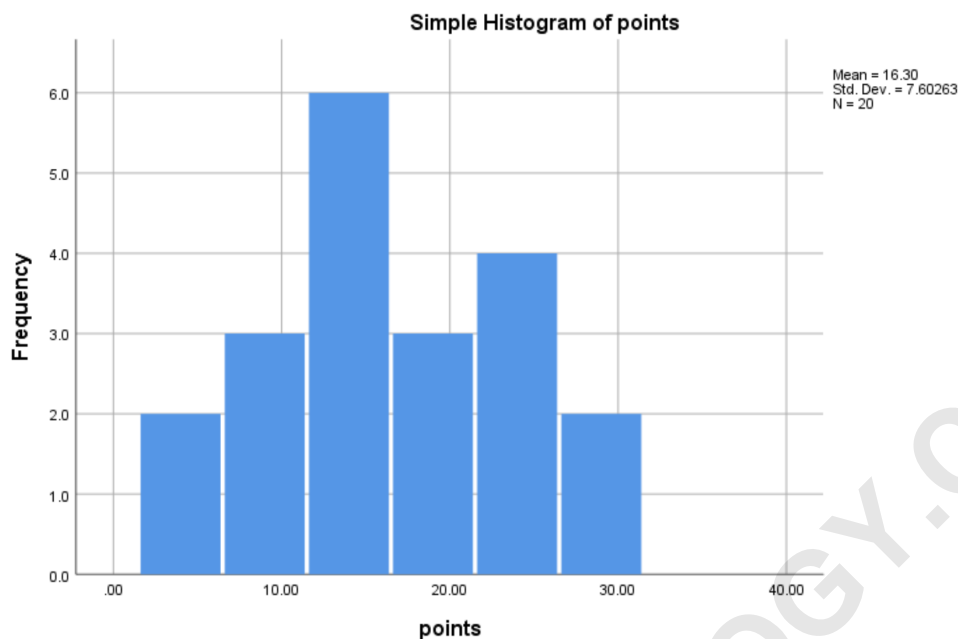
Once the Chart Builder interface is active, you will see a variety of chart types categorized by their structure. In the **Choose from** list, locate and select the **Histogram** category. From the available templates, drag the simple histogram icon into the main editing canvas. This action prepares the software to receive the specific **variables** you wish to analyze. The interface is designed to be user-friendly, utilizing a drag-and-drop mechanism that simplifies complex graphing tasks.

Next, you must identify your target variable in the variable list, which in this case is **points**. Click on this variable and drag it into the placeholder for the x-axis. The x-axis represents the **measurement** (points per game), while the y-axis will automatically calculate the **count** or frequency of those scores within specified intervals or bins. Ensuring that the variable is correctly placed is essential for generating a meaningful visual output.



After configuring the axes, you may choose to adjust the bin size or add a normal curve overlay within the Chart Builder options to enhance the **visualization**. Once you are satisfied with the settings, click the **OK** button. The software will process the request and display the resulting histogram in the **SPSS** Output Viewer, where you can further edit or export the graphic for use in your documentation.

Upon clicking **OK**, the following histogram will appear in your output window, providing a clear view of how the points are distributed across the twenty players:

→ GGraph

Interpreting the Visual Distribution

Analyzing the generated histogram reveals that while the variable "points" is not perfectly symmetrical, it does approximate a **normal distribution**. Most of the players in our sample score between 10 and 20 points per game, creating a central peak. As we move away from this central range, the number of players scoring very high or very low amounts decreases, which is characteristic of the bell curve. There is no evidence of extreme **multimodality** or severe outliers that would immediately invalidate the use of parametric tests.

It is important to acknowledge that real-world data is rarely "perfectly" normal. Minor fluctuations in the heights of the bars are expected, especially with a relatively small **sample size** of 20 observations. The goal of this visual check is not to demand perfection, but to ensure that the data is "normal enough" to not violate the underlying assumptions of your chosen **statistical model**. If the histogram were heavily skewed to the left or right, a data transformation might be necessary.

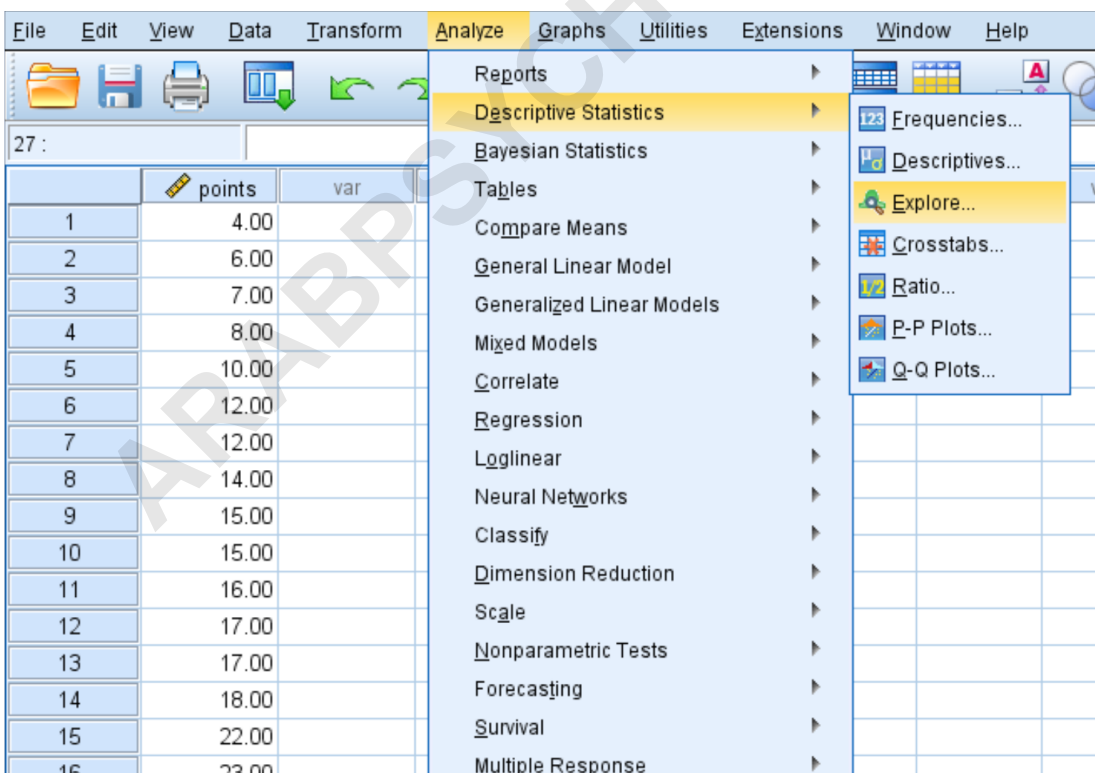
While the histogram method is incredibly useful for a quick **exploratory data analysis**, it remains a subjective tool. Different researchers might interpret the same histogram differently based on their visual perception. To achieve a more rigorous and objective conclusion, it is highly recommended to move beyond visualization and apply **formal statistical tests**. These tests provide specific numerical values that can be compared against established thresholds to determine normality with greater confidence.

Formal Statistical Tests for Normality

To eliminate the subjectivity of visual charts, we use formal hypothesis tests to determine if a variable follows a **normal distribution**. Within the **SPSS** environment, the two most frequently cited and utilized tests are the **Shapiro-Wilk test** and the **Kolmogorov-Smirnov test**. These tests compare the scores in your sample to a normally distributed set of scores with the same mean and standard deviation.

The **null hypothesis** for these tests is that the data is normally distributed. Therefore, if the resulting **p-value** is greater than a predetermined alpha level (typically 0.05), we fail to reject the null hypothesis, suggesting that the data does not significantly deviate from normality. Conversely, a p-value less than 0.05 indicates that the distribution is significantly different from a normal curve. Each test has its own strengths; for instance, the Shapiro-Wilk test is generally considered more powerful for smaller sample sizes.

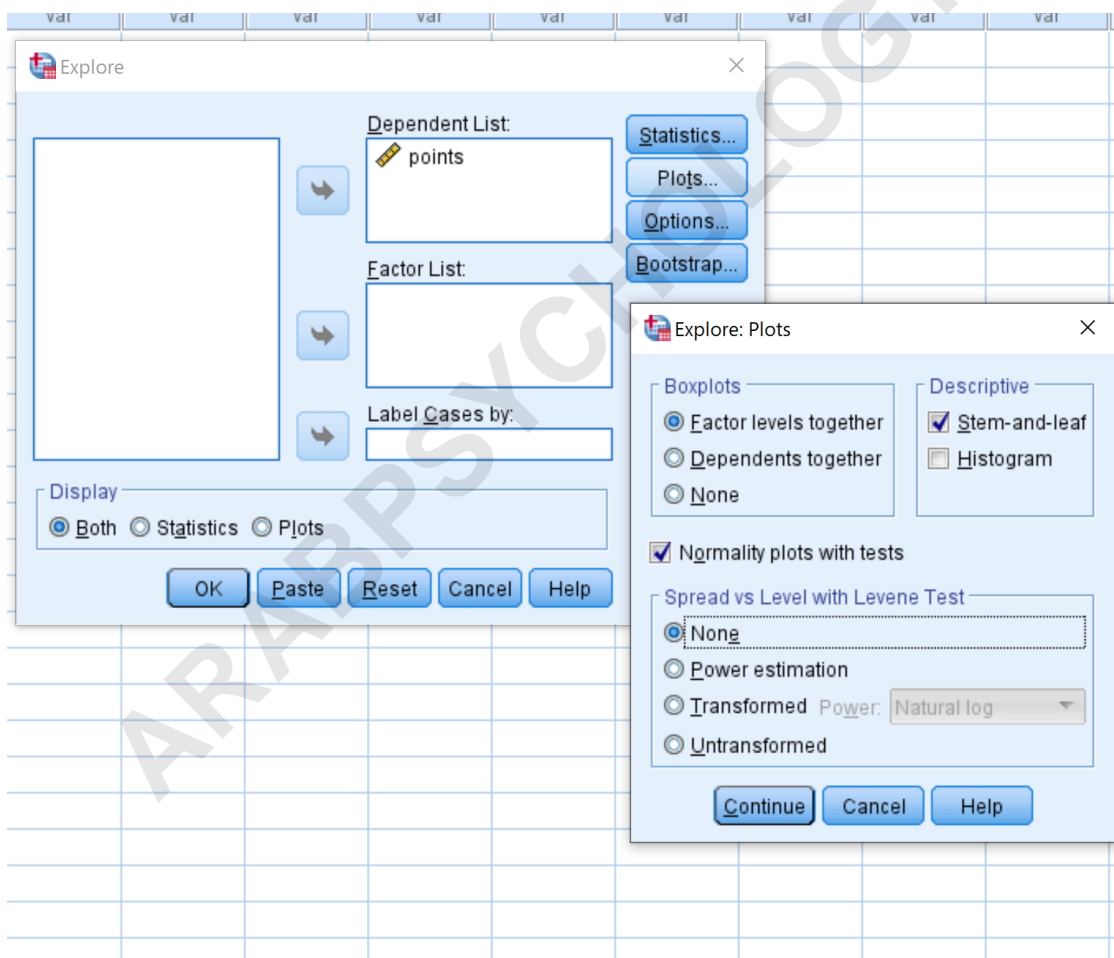
To perform these tests simultaneously in **SPSS**, you will need to access the **Descriptive Statistics** suite. Click on the **Analyze** tab in the main menu, select **Descriptive Statistics**, and then choose **Explore**. This specific path is the most efficient way to generate comprehensive normality statistics alongside other essential descriptive metrics.



Configuring the Explore Menu for Normality

When the Explore dialog box opens, you will need to specify which variables are being tested. Locate the variable **points** in the source list and move it into the **Dependent List** box. This tells **SPSS** that "points" is the primary outcome variable you wish to analyze for its distributional properties. If you had multiple variables, you could move them all into this box to test them all at once.

Next, you must specify the type of output you require by clicking the **Plots** button on the right side of the dialog box. In the new window that appears, ensure you check the box labeled **Normality plots with tests**. This is a critical step, as it instructs the software to run the **Shapiro-Wilk** and **Kolmogorov-Smirnov** calculations. You may also want to keep the "Stem-and-leaf" or "Histogram" options selected for additional visual context.



Once your selections are finalized, click **Continue** to return to the main Explore window, and then click **OK** to execute the analysis. The **SPSS** processor will then perform the necessary **algorithms** and generate a series of tables in the Output Viewer. The most important table for our purposes

will be titled "Tests of Normality."

Upon execution, the results will be presented in a concise table similar to the one shown below, which displays the **test statistic**, the degrees of freedom, and the significance levels for both tests:

	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

Interpreting the Numerical Output

When analyzing the "Tests of Normality" output, we focus on the **p-value**, which is labeled as "Sig." in **SPSS**. For the basketball dataset, the results are as follows:

Kolmogorov-Smirnov Test: The test statistic is **.113** with a p-value of **.200**.

Shapiro-Wilk Test: The test statistic is **.967** with a p-value of **.699**.

In both instances, the p-values are significantly higher than the standard **significance level** of 0.05. This leads us to a clear conclusion: we do not have enough evidence to reject the **null hypothesis**. Therefore, we can confidently assume that the variable "points" follows a **normal distribution**. This finding validates our initial visual assessment from the histogram and provides a mathematical foundation for further analysis.

With this confirmation, the researcher is now free to proceed with **parametric tests**. For example, if you wanted to compare the mean points of this group to another group using an **independent samples t-test**, you now know that the assumption of normality is satisfied. This ensures that the **confidence intervals** and p-values generated by subsequent tests will be accurate and dependable. Understanding and verifying these assumptions is what separates rigorous scientific research from mere data observation.