Learn How to Create and Interpret Q-Q Plots in SPSS for Normality Testing

Authored by

Mohammed looti

November 8, 2025

A Q-Q plot, which stands for "quantile-quantile" plot, is a fundamental graphical tool in statistical analysis. Its primary purpose is to visually assess whether the distribution of a given variable aligns with a specified theoretical distribution, most commonly the normal distribution. Understanding the distributional properties of data is essential because many parametric statistical tests, such as t-tests and ANOVA, rely on the assumption that the data, or the residuals derived from a model, are normally distributed.

When creating a Q-Q plot, the observed data quantiles are plotted against the theoretical quantiles of the reference distribution. If the observed data perfectly matches the theoretical distribution, the points on the plot will form a straight line. Any significant deviation from this expected linear pattern signals that the data may violate the assumption of normality, potentially necessitating the use of non-parametric methods or data transformations. This comprehensive tutorial provides a detailed explanation of how to generate and correctly interpret a Q-Q plot using the powerful statistical software, SPSS.

Understanding Quantile-Quantile Plots

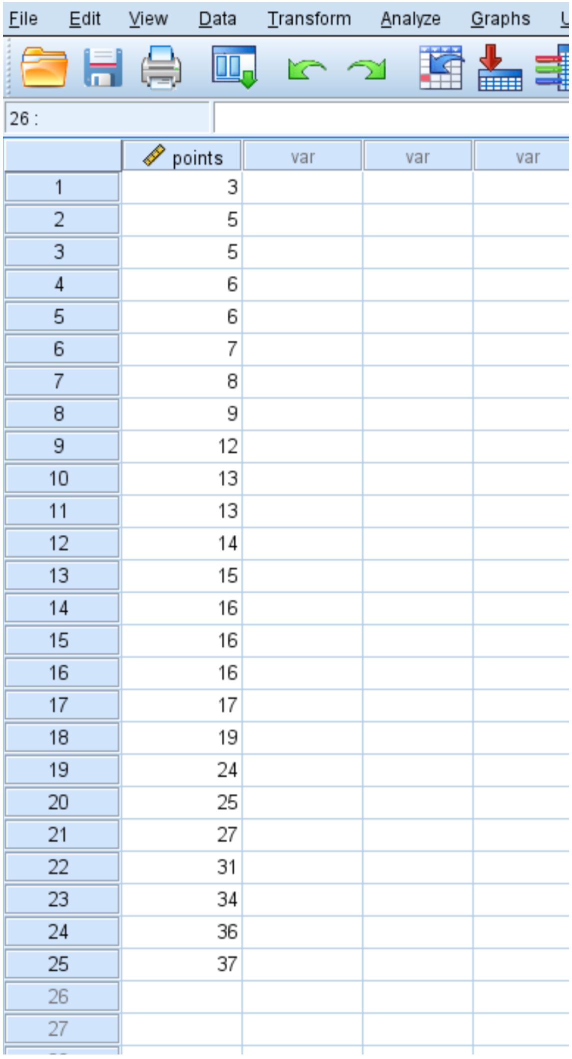
The theoretical foundation of the Q-Q plot rests on the concept of quantiles. A quantile divides the probability distribution of a random variable into continuous intervals with equal probabilities. In the context of normality testing, the plot compares the quantiles derived from the sample data against the corresponding theoretical quantiles expected if the data truly followed a standard normal distribution. This methodology provides a robust visual check, complementing formal statistical tests by allowing researchers to examine the nature and location of any non-normality--whether it occurs in the tails, center, or throughout the distribution.

While visual inspection is subjective, it often reveals aspects of the data distribution that purely numerical tests might overlook. For instance, a Q-Q plot can clearly distinguish between skewness (where points curve consistently above or below the line) and kurtosis (where points deviate significantly only at the tail ends). Therefore, the Q-Q plot is not merely a diagnostic tool; it is an interpretive lens that helps analysts understand the underlying shape of their dataset. Mastery of its interpretation is critical for ensuring the validity of subsequent statistical modeling and hypothesis testing.

Preparing Data in SPSS for Normality Assessment

Before initiating the procedure in SPSS, we must first ensure our data is loaded correctly. For this example, we will utilize a dataset containing the points per game scored by 25 distinct professional basketball players. Assessing the normality of such a variable, which we call **points**, is a common preliminary step to determine if standard parametric analyses are appropriate for comparing scoring performance across different groups or conditions.

The dataset is structured with a single quantitative variable, and the goal is to determine if this variable adheres to the assumptions required for analyses such as correlation or regression, where normality of the variables or their residuals is assumed. The following image represents the structure of the data as it appears within the SPSS Data View, showcasing the scores for each player.



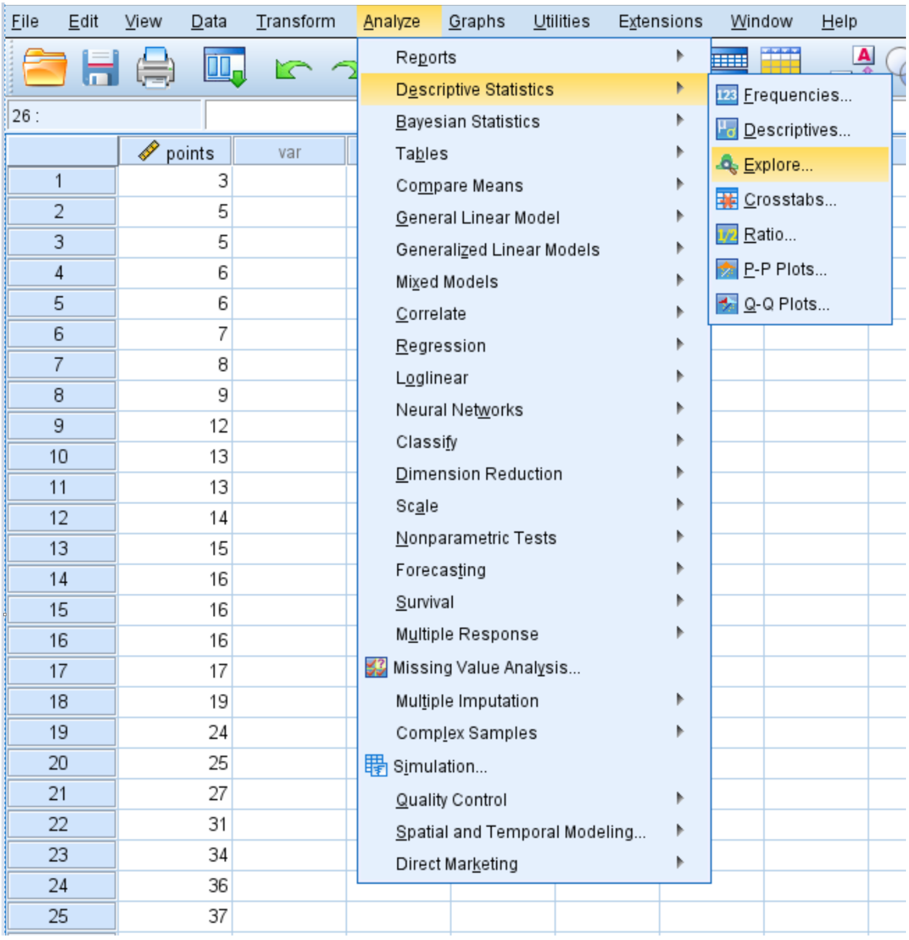
Once the data is verified, we proceed to the necessary steps within SPSS to generate the normality plots and accompanying statistical tests. This integrated approach allows for both a visual and formal assessment of whether the variable **points** is sufficiently normally distributed to proceed with parametric statistical modeling.

Step-by-Step Guide to Generating the Q-Q Plot in SPSS

Generating the Q-Q plot in SPSS is straightforward, utilizing the Explore function which is specifically designed for comprehensive data exploration and distribution diagnostics. We will follow a multi-step process to access this feature and configure the necessary output options.

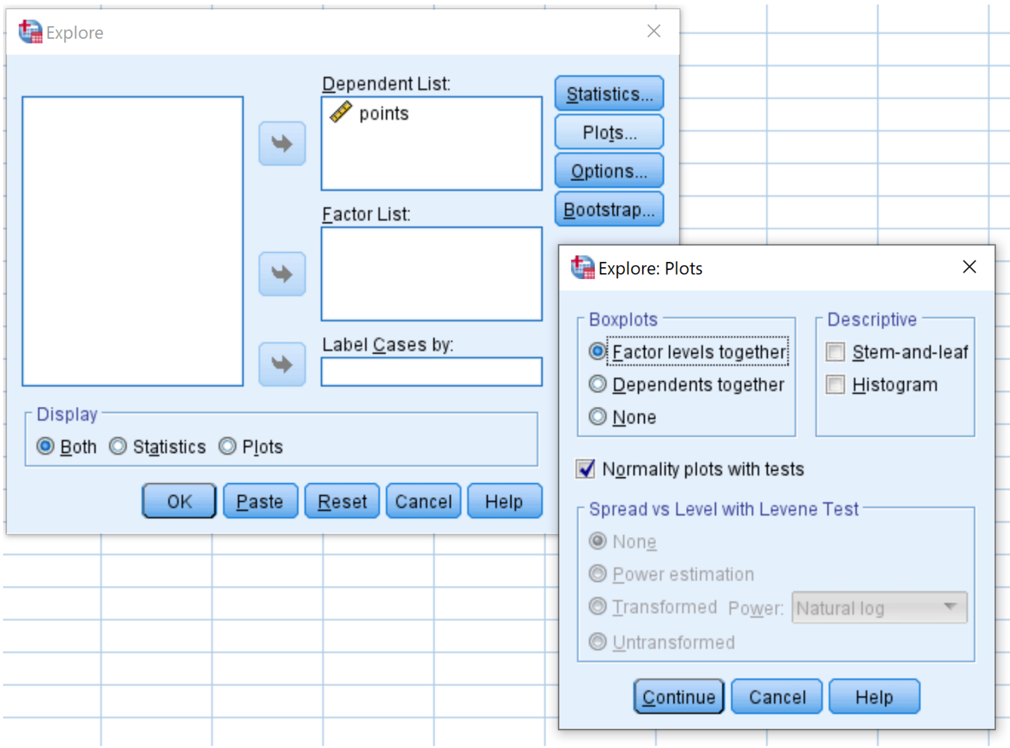
Step 1: Choose the Explore option.

To begin, navigate to the main menu bar and select the **Analyze** tab. From the subsequent dropdown menu, hover over Descriptive Statistics, and then select **Explore**. This action opens the Explore dialog box, which is the gateway to detailed distribution analysis, including plots and formal tests for normality.



Step 2: Configure the Analysis and Create the Q-Q plot.

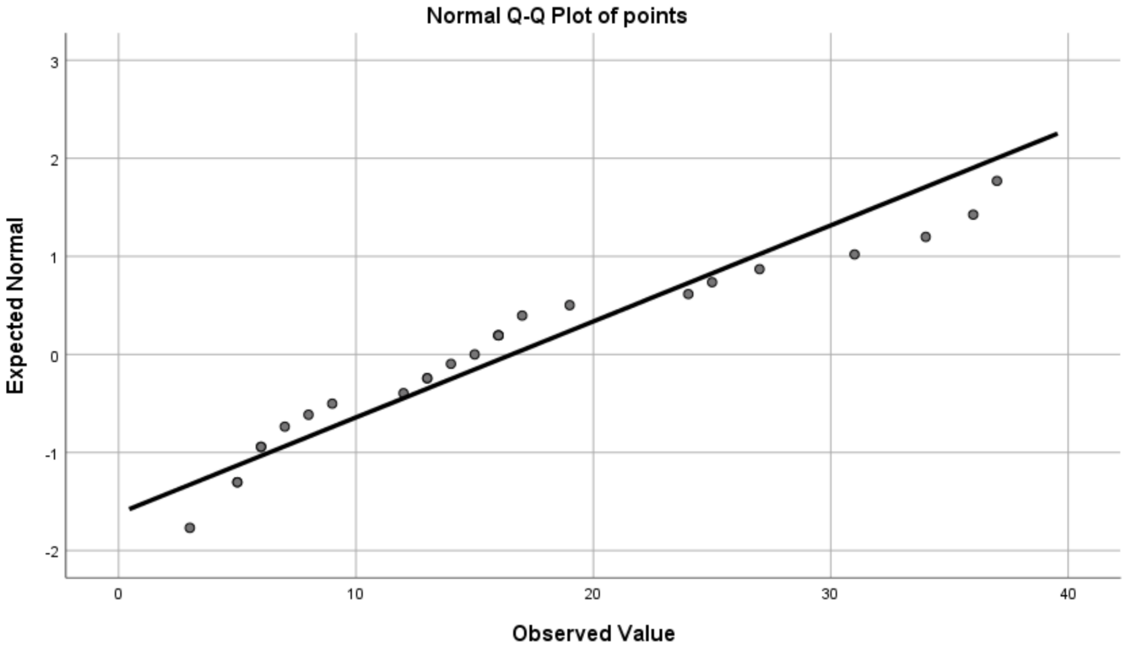
Within the Explore dialog box, you must first move the variable of interest--in this case, **points**--into the box labeled **Dependent List**. This specifies the variable whose distribution will be analyzed. Next, click the button labeled **Plots** to open the secondary dialog box for graphical options. It is absolutely essential to ensure that the checkbox next to **Normality plots with tests** is checked. This crucial selection instructs SPSS to generate the Q-Q plot as well as the accompanying numerical results from the Shapiro-Wilk and Kolmogorov-Smirnov Test. After confirming this setting, click **Continue**, and then click **OK** in the main Explore window to execute the analysis and generate the output.



Interpreting the Visual Output of the Q-Q Plot

Upon clicking **OK**, the SPSS output viewer will display the results, including the generated Q-Q plot. The fundamental principle of interpreting this plot is straightforward: if the data is perfectly normally distributed, the plotted points representing the observed data quantiles will align precisely along the diagonal reference line. This reference line typically runs at a 45-degree angle and represents the theoretical expected values under the assumption of normality.

The following graphic illustrates the resulting plot for our variable **points**:



When analyzing this specific plot, we observe that the data points do not adhere strictly to the 45-degree reference line. The most noticeable deviations occur at the extreme ends, or "tail ends," of the distribution. Points that fall significantly above the line in the upper right or below the line in the lower left suggest that the sample distribution has heavier tails than the theoretical normal distribution, which is an indication of potential kurtosis or non-normality. For the **points** variable, the visible curvature, particularly where the points cluster away from the center of the line, strongly suggests that the underlying distribution of player scores deviates from the ideal normal curve.

While minor deviations are expected due to sampling variability, especially with small sample sizes, the systematic curvature visible here raises concerns about the appropriateness of assuming normality. Therefore, based purely on visual inspection of the Q-Q plot, we have preliminary evidence suggesting that the distribution of **points** is not perfectly normally distributed. This visual evidence must now be corroborated or refuted by the accompanying formal statistical tests.

Analyzing Formal Statistical Tests (Kolmogorov-Smirnov and Shapiro-Wilk)

In addition to the visual assessment provided by the Q-Q plot, SPSS provides two primary inferential statistics for normality testing: the Kolmogorov-Smirnov Test (K-S) and the Shapiro-Wilk Test (S-W). These tests operate under a null hypothesis (H_0) that the data is drawn from a normally distributed population. Consequently, a small p-value (typically less than the chosen alpha level, commonly 0.05) leads to the rejection of H_0, indicating that the data is significantly non-normal.

The table below, which appears in the SPSS output alongside the Q-Q plot, presents the results for these two tests. Note that the Shapiro-Wilk Test is generally preferred for smaller sample sizes (N < 50), while the Kolmogorov-Smirnov Test is more appropriate for larger datasets. Given our sample size of 25, the S-W test offers a more reliable assessment.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
points	.163	25	.086	.916	25	.042

a. Lilliefors Significance Correction

We analyze the associated p-values (Sig. column) for both tests:

P-value of Kolmogorov-Smirnov Test: **.086**
P-value of Shapiro-Wilk Test: **.042**

The p-value from the preferred Shapiro-Wilk Test is 0.042. Since 0.042 is less than the standard significance level of $alpha = 0.05$, we reject the null hypothesis of normality. This formal statistical finding confirms the visual suspicion raised by the Q-Q plot: the variable **points** is statistically unlikely to be drawn from a normally distributed population. The result of the Kolmogorov-Smirnov Test (0.086) is close to the threshold but fails to reject the null hypothesis at the 0.05 level, highlighting the increased sensitivity of the Shapiro-Wilk test for small samples.

Conclusion: Synthesis of Visual and Statistical Evidence

The process of assessing normality requires the careful synthesis of both graphical evidence and formal statistical results. In the analysis of the **points** variable, the visual evidence from the Q-Q plot showed noticeable deviations from the theoretical line, particularly at the distributional tails, suggesting non-normality likely due to heavy tails (kurtosis).

This visual interpretation was subsequently validated by the statistical output, specifically the Shapiro-Wilk Test, which yielded a p-value below 0.05. When conducting advanced statistical modeling using this variable, researchers must acknowledge this non-normality. Depending on the severity of the violation and the specific analysis planned, this may necessitate using techniques such as data transformation (e.g., logarithmic transformation) to improve the distribution's shape, or shifting to non-parametric statistical methods that do not rely on the assumption of normality.

In summary, the Q-Q plot remains an indispensable tool in the data scientist's arsenal, providing immediate, intuitive insight into the distributional characteristics of a dataset, which, when combined with formal test results, ensures that the assumptions underlying subsequent statistical analysis are rigorously addressed.