How to Create a Q-Q Plot in Excel to Check for Normal Distribution

Authored by

stats writer

March 10, 2026

How can I create a Q-Q plot in Excel?

Understanding the Fundamentals of a Q-Q Plot

A **Q-Q plot**, which stands for a quantile-quantile plot, serves as a powerful diagnostic tool in **data analysis**. Its primary function is to allow researchers and analysts to visually compare two probability distributions by plotting their **quantiles** against one another. By doing so, one can effectively determine if a sample dataset aligns with a specific theoretical model, providing a more intuitive understanding than numerical tests alone.

In most practical scenarios, this **graphical method** is utilized to verify the assumption of normality. Many **statistical tests**, such as the t-test or ANOVA, rely on the premise that the underlying data follows a **normal distribution**. The **Q-Q plot** provides a clear visual representation of how closely the observed data matches this bell-shaped curve, highlighting potential outliers or skewness in the process.

When constructing these plots in **Microsoft Excel**, the process involves comparing the empirical quantiles of your sorted data against the theoretical quantiles of a **standard normal distribution**. If the data points follow a linear trend, it suggests the distributions are similar. This tutorial provides a comprehensive guide to executing this technique within a spreadsheet environment, ensuring accuracy and clarity for your analytical projects.

The Importance of Normality in Statistical Analysis

Establishing whether a dataset follows a **normal distribution** is a critical first step in many workflows. This distribution is characterized by its symmetrical shape, where the mean, median, and mode coincide at the center. In **inferential statistics**, the validity of your conclusions often depends on this distributional assumption. A **Q-Q plot** acts as a gatekeeper, helping you decide whether to proceed with parametric tests or opt for non-parametric alternatives.

Unlike a histogram, which can be sensitive to the choice of bin width, a **Q-Q plot** provides a more precise look at the behavior of the data in the tails. The tails of a distribution are often where the most significant deviations from normality occur, such as heavy-tailedness or extreme outliers. By plotting **quantiles**, you gain a granular view of how each segment of your data behaves relative to the expected theoretical values.

Furthermore, the **Q-Q plot** is highly versatile. While this guide focuses on the **normal distribution**, the same methodology can be applied to compare data against other distributions, such as the Exponential or Uniform distributions. Mastery of this tool in **Microsoft Excel** empowers you to perform robust **exploratory data analysis** without needing specialized statistical software.

Step 1: Data Entry and Preliminary Sorting

The first stage in creating a **Q-Q plot** involves organizing your raw observations. Open **Microsoft Excel** and enter your values into a single vertical column. For the plot to be meaningful, the data must be organized in a way that reflects its cumulative probability, which necessitates sorting the values from the lowest to the highest.

	A	B	C	D
1	**Data**			
2	-3.4			
3	-2.9			
4	-2.8			
5	-2.3			
6	-1.5			
7	-0.4			
8	0.4			
9	1.7			
10	2.4			
11	2.9			
12				
13				
14				
15				

To ensure your data is properly sequenced, highlight the range containing your numbers and navigate to the **Data** tab on the Excel ribbon. Locate the **Sort & Filter** group and select the **Sort A to Z** icon. This ensures that the empirical **quantiles** are calculated in the correct order, which is a prerequisite for the mathematical steps that follow.

Maintaining a clean dataset is vital. Ensure there are no empty cells or non-numeric characters within your selection, as these can disrupt the **data analysis** functions. Once the data is sorted, you have established the foundation for calculating the relative rank and percentile of each observation.

Step 2: Calculating the Rank of Each Observation

Once your data is sorted, the next objective is to assign a numerical rank to every data point. This rank represents the position of each value within the ordered set. In **Microsoft Excel**, this can be automated using the **RANK** function, although since the data is already sorted, you could also use a simple incrementing list (1, 2, 3...).

To calculate the rank accurately using a formula, enter the following into the cell adjacent to your first data point:

=RANK(A2, $A$2:$A$11, 1)

B2 =RANK(A2, $A$2:$A$11, 1)

	A	B	C	D	E	F
1	**Data**	**Rank**				
2	-3.4	1				
3	-2.9					
4	-2.8					
5	-2.3					
6	-1.5					
7	-0.4					
8	0.4					
9	1.7					
10	2.4					
11	2.9					
12						
13						
14						
15						

After entering the formula, use the fill handle to copy it down to the remaining cells in the column. This step is essential because the rank (denoted as i) is a primary variable in the formula used to calculate the **cumulative probability** of each observation.

	A	B	C	D	E
1	**Data**	**Rank**			
2	-3.4	1			
3	-2.9	2			
4	-2.8	3			
5	-2.3	4			
6	-1.5	5			
7	-0.4	6			
8	0.4	7			
9	1.7	8			
10	2.4	9			
11	2.9	10			
12					
13					
14					
15					

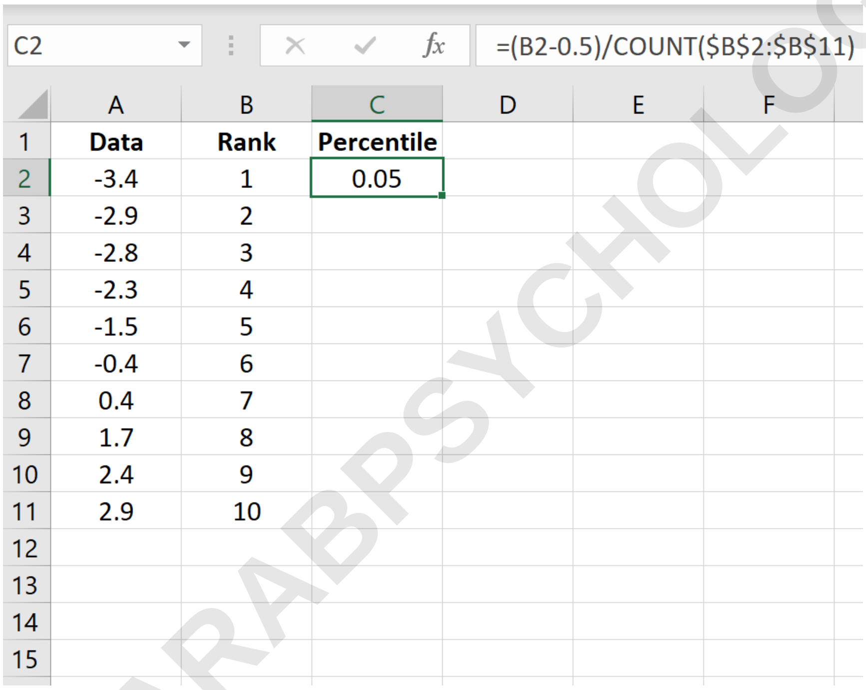
By defining the rank, you are essentially quantifying the "order statistic" of your sample. This allows the **Q-Q plot** to map the specific location of each data point relative to the total number of observations in your dataset, which is the next logical step in our **data analysis** process.

Step 3: Determining the Percentile and Cumulative Probability

With the ranks established, we must now calculate the percentile for each data point. This value represents the **cumulative distribution function** (CDF) for the sample. A common formula used for **Q-Q plots** is the **Blom's plotting position** or a similar variation like (i - 0.5) / n, where i is the rank and n is the sample size.

In **Microsoft Excel**, apply the following formula to your third column:

=(B2-0.5)/COUNT($B$2:$B$11)


C2 =(B2-0.5)/COUNT($B$2:$B$11)

	A	B	C	D	E	F
1	Data	Rank	Percentile			
2	-3.4	1	0.05			
3	-2.9	2				
4	-2.8	3				
5	-2.3	4				
6	-1.5	5				
7	-0.4	6				
8	0.4	7				
9	1.7	8				
10	2.4	9				
11	2.9	10				
12						
13						
14						
15						

The "0.5" subtraction is a continuity correction that prevents the cumulative probability from reaching exactly 1.0, which would result in an undefined value during the next step. Drag this formula down to populate the entire column, ensuring each data point has a corresponding percentile value between 0 and 1.

	A	B	C	D	E
1	**Data**	**Rank**	**Percentile**		
2	-3.4	1	0.05		
3	-2.9	2	0.15		
4	-2.8	3	0.25		
5	-2.3	4	0.35		
6	-1.5	5	0.45		
7	-0.4	6	0.55		
8	0.4	7	0.65		
9	1.7	8	0.75		
10	2.4	9	0.85		
11	2.9	10	0.95		
12					
13					
14					
15					

These percentiles indicate where each data point falls on a scale of 0 to 100% within the dataset. By calculating these values, you are preparing to translate the sample's empirical distribution into a format that can be compared directly to the **standard normal distribution**.

Step 4: Computing Theoretical Quantiles (Z-Scores)

The next phase involves converting the cumulative probabilities into theoretical **quantiles**. These are expressed as **z-scores**, which represent the number of standard deviations a point is from the mean in a **normal distribution**. This transformation is what allows the **Q-Q plot** to function as a comparative tool.

To perform this in **Microsoft Excel**, use the **NORM.S.INV** function. This function takes a probability as an input and returns the corresponding value from the **standard normal distribution**. Enter the following formula:

=NORM.S.INV(C2)

D2 =NORM.S.INV(C2)

	A	B	C	D	E	F
1	**Data**	**Rank**	**Percentile**	**Z-Score**		
2	-3.4	1	0.05	-1.64485		
3	-2.9	2	0.15			
4	-2.8	3	0.25			
5	-2.3	4	0.35			
6	-1.5	5	0.45			
7	-0.4	6	0.55			
8	0.4	7	0.65			
9	1.7	8	0.75			
10	2.4	9	0.85			
11	2.9	10	0.95			
12						
13						
14						
15						

As with the previous steps, copy this formula down for all rows. The resulting **z-scores** represent where your data points should lie if they were perfectly normally distributed. These values will eventually serve as the coordinates for the x-axis of our scatter plot.

	A	B	C	D	E
1	**Data**	**Rank**	**Percentile**	**Z-Score**	
2	-3.4	1	0.05	-1.64485	
3	-2.9	2	0.15	-1.03643	
4	-2.8	3	0.25	-0.67449	
5	-2.3	4	0.35	-0.38532	
6	-1.5	5	0.45	-0.12566	
7	-0.4	6	0.55	0.125661	
8	0.4	7	0.65	0.38532	
9	1.7	8	0.75	0.67449	
10	2.4	9	0.85	1.036433	
11	2.9	10	0.95	1.644854	
12					
13					
14					
15					

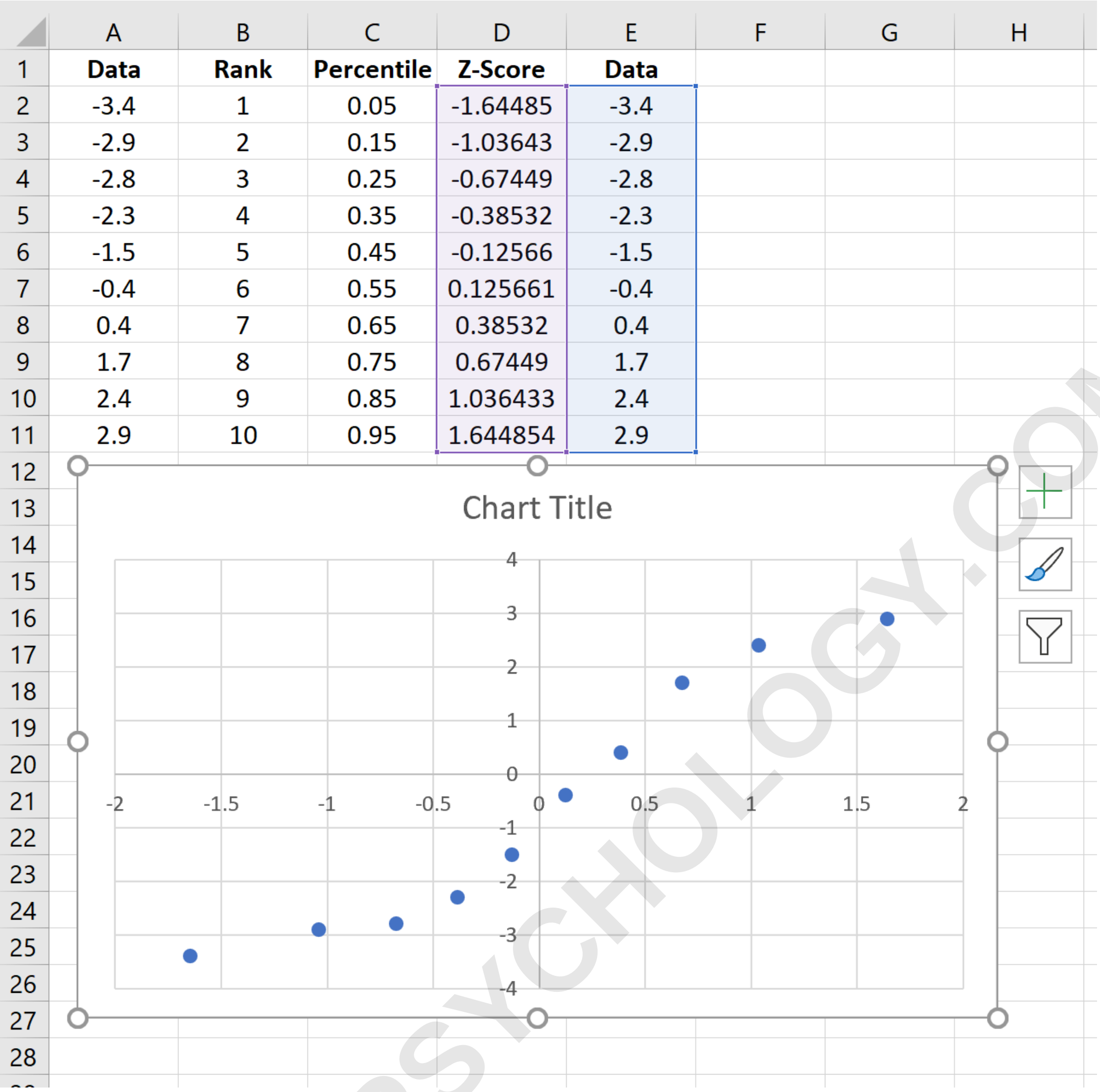
By generating these theoretical values, you have completed the mathematical preparation required for the visualization. You now have two sets of values: the actual observed data (empirical) and the expected **z-scores** (theoretical), which are ready to be plotted against each other.

Step 5: Visualizing the Data with a Scatter Plot

With all necessary calculations finished, it is time to generate the **scatter plot**. To make the plotting process easier in **Microsoft Excel**, it is often helpful to copy your original data (from Column A) into a new column (Column E) next to your **z-scores** (Column D). This ensures the x and y values are adjacent.

	A	B	C	D	E	F
1	Data	Rank	Percentile	Z-Score	Data	
2	-3.4	1	0.05	-1.64485	-3.4	
3	-2.9	2	0.15	-1.03643	-2.9	
4	-2.8	3	0.25	-0.67449	-2.8	
5	-2.3	4	0.35	-0.38532	-2.3	
6	-1.5	5	0.45	-0.12566	-1.5	
7	-0.4	6	0.55	0.125661	-0.4	
8	0.4	7	0.65	0.38532	0.4	
9	1.7	8	0.75	0.67449	1.7	
10	2.4	9	0.85	1.036433	2.4	
11	2.9	10	0.95	1.644854	2.9	
12						
13						
14						
15						

Highlight the data in both columns D and E, then navigate to the **Insert** tab on the ribbon. Within the **Charts** group, select **Insert Scatter (X, Y)** and choose the basic **Scatter** option. This will create the initial **Q-Q plot**, where the theoretical quantiles are on the x-axis and the observed data is on the y-axis.



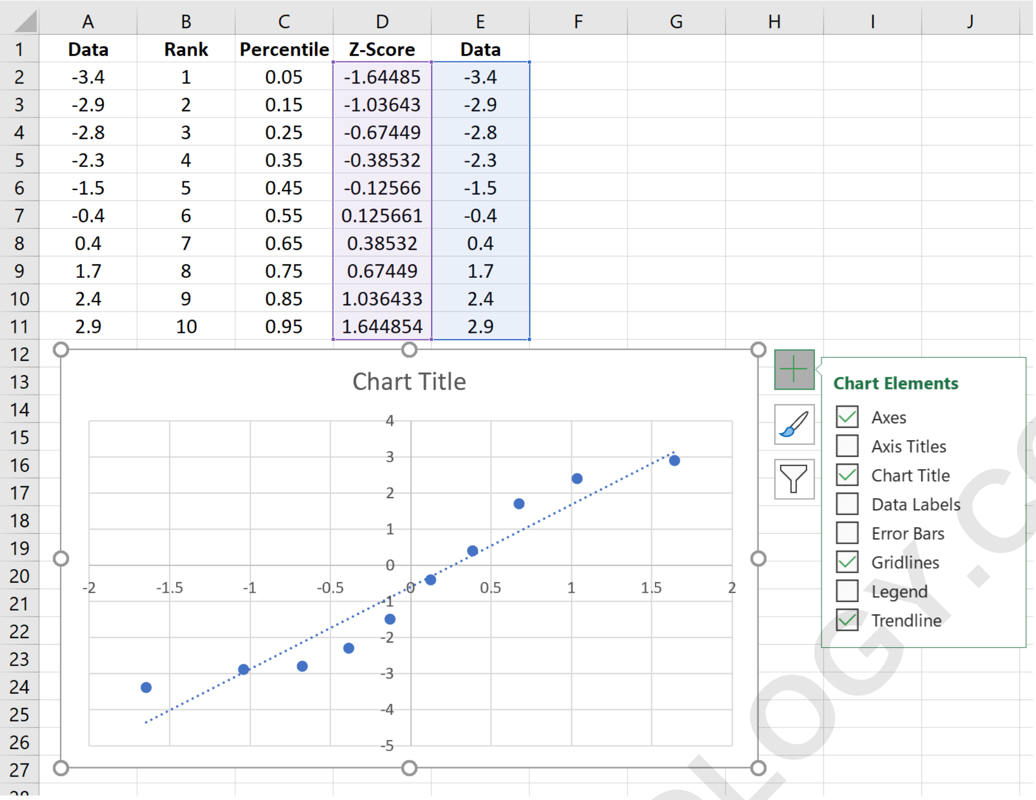
	A	B	C	D	E	F	G	H
1	Data	Rank	Percentile	Z-Score	Data			
2	-3.4	1	0.05	-1.64485	-3.4			
3	-2.9	2	0.15	-1.03643	-2.9			
4	-2.8	3	0.25	-0.67449	-2.8			
5	-2.3	4	0.35	-0.38532	-2.3			
6	-1.5	5	0.45	-0.12566	-1.5			
7	-0.4	6	0.55	0.125661	-0.4			
8	0.4	7	0.65	0.38532	0.4			
9	1.7	8	0.75	0.67449	1.7			
10	2.4	9	0.85	1.036433	2.4			
11	2.9	10	0.95	1.644854	2.9			

The resulting **scatter plot** provides the first visual evidence of your data's distribution. If the points appear to form a straight line, it is a strong indicator that your data follows the **normal distribution**. However, to make this comparison more rigorous, we need to add a reference line.

Step 6: Adding a Trendline and Finalizing Labels

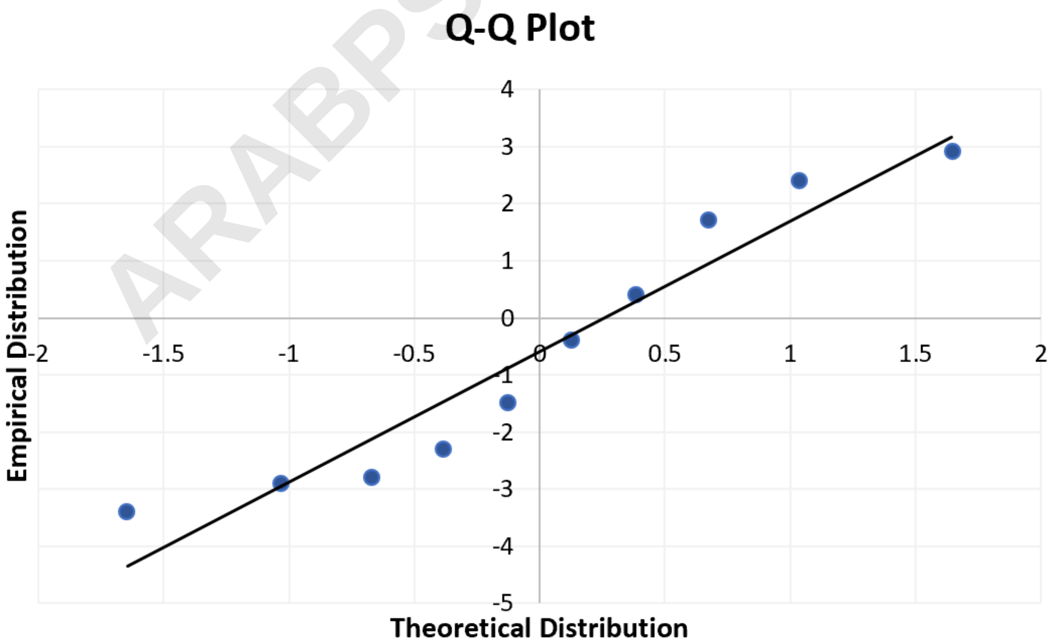
To properly interpret the **Q-Q plot**, you must add a **trendline**. This line serves as a reference for what the data would look like if it perfectly matched the theoretical distribution. In **Microsoft Excel**, click anywhere on the chart area to reveal the **Chart Elements** button (the plus sign at the top-right corner).

Check the box labeled **Trendline**. This will insert a linear **trendline** through your data points. If the data is normally distributed, the points should hug this line closely across the entire range of the plot.



	Data	Rank	Percentile	Z-Score	Data
2	-3.4	1	0.05	-1.64485	-3.4
3	-2.9	2	0.15	-1.03643	-2.9
4	-2.8	3	0.25	-0.67449	-2.8
5	-2.3	4	0.35	-0.38532	-2.3
6	-1.5	5	0.45	-0.12566	-1.5
7	-0.4	6	0.55	0.125661	-0.4
8	0.4	7	0.65	0.38532	0.4
9	1.7	8	0.75	0.67449	1.7
10	2.4	9	0.85	1.036433	2.4
11	2.9	10	0.95	1.644854	2.9

For a professional finish, you should also add descriptive labels to your axes and a clear title. Label the x-axis as "Theoretical Quantiles" and the y-axis as "Observed Values." This ensures that anyone viewing the **Q-Q plot** can immediately understand the context of the **data analysis**.



Step 7: Interpreting the Q-Q Plot Results

The interpretation of a **Q-Q plot** is centered on the linearity of the plotted points. If the data points lie approximately on a 45-degree diagonal line, we conclude that the sample data is likely drawn from a **normal distribution**. Minor fluctuations are common, especially with smaller sample sizes, but significant patterns of deviation are noteworthy.

Specific shapes in the plot reveal different distributional characteristics. For example, if the points curve upward at both ends, it may indicate that the distribution has "heavy tails" (more extreme values than a normal distribution). Conversely, an S-shaped curve might suggest that the data is skewed or has "thin tails." Identifying these patterns is a vital skill in **exploratory data analysis**.

In our current example, if the data points deviate significantly from the **trendline** at the extremities, it suggests that the assumption of normality might be violated. While the **Q-Q plot** is not a formal **statistical test** like the Shapiro-Wilk or Kolmogorov-Smirnov tests, its visual nature often makes it easier to communicate findings to stakeholders.

Ultimately, **Microsoft Excel** provides a robust environment for creating these plots. By following these steps, you can quickly assess the distributional properties of your data, leading to more informed decisions in your statistical modeling and overall research strategy. Whether you are a student or a professional analyst, the **Q-Q plot** remains an indispensable part of the toolkit.