

Understanding and Applying Data Transformations: Log, Square Root, and Cube Root in Excel

Authored by
Mohammed looti

November 4, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Understanding and Applying Data Transformations: Log, Square Root, and Cube Root in Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10191>

In the realm of quantitative analysis, many powerful [statistical tests](#), such as ANOVA or t-tests, are classified as parametric. These methods rely fundamentally on the assumption that the underlying population data follows a [Normal distribution](#). When this critical assumption is violated, the reliability of the test results diminishes significantly, potentially leading to erroneous conclusions regarding hypotheses.

Unfortunately, datasets collected from the real world, particularly those dealing with counts, income, or reaction times, frequently exhibit strong skewness or unusual distributions that are decidedly non-normal. Applying standard parametric procedures to such data risks invalid inference. To mitigate this pervasive issue, data transformation techniques are employed to mathematically adjust the data values, thereby shifting the distributional shape closer to [normality](#) and stabilizing variance.

Selecting the correct transformation is a crucial step in preparing data for robust statistical modeling. The three most commonly utilized power transformations for achieving [normality](#) are detailed below, each serving a distinct purpose based on the nature and severity of the non-normality present in the raw data:

- 1. Log Transformation:** This technique replaces the original observation (denoted as y) with $\log(y)$. It is the most severe of the power transformations and is highly effective for data exhibiting pronounced positive [skewness](#), often compressing large outliers dramatically.
- 2. Square Root Transformation:** A gentler approach, this method transforms the value (y) to \sqrt{y} . It is particularly well-suited for count data (e.g., number of occurrences) where values are small, helping to stabilize the variance associated with Poisson-like distributions.
- 3. Cube Root Transformation:** This transformation converts the value (y) to $y^{1/3}$. It provides an intermediate level of adjustment between the mild square root and the strong log transformation. Crucially, unlike the standard log transformation, the cube root method can be applied successfully to data containing both positive and negative values, making it highly versatile.

By implementing these mathematical operations, researchers aim to satisfy the prerequisite conditions for parametric analysis. The following sections provide comprehensive, step-by-step instructions on how to execute each of these essential data transformations efficiently and accurately within the ubiquitous spreadsheet environment of **Microsoft Excel**.

Applying the Log Transformation in Excel

The [Log Transformation](#) is often the primary tool deployed when initial exploratory analysis reveals that the data is severely positively skewed--meaning the tail of the distribution stretches far to the right. This transformation works by reducing the proportional differences between larger values more effectively than it reduces the differences between smaller values, pulling the distribution inward toward a more centralized, normal shape.

In Excel, the base-10 logarithm is easily calculated using the built-in function **=LOG10()**. While the natural logarithm (base e , using the **=LN()** function) is also commonly used in statistics, the base-10 logarithm offers results that are often easier to interpret and communicate. To apply this transformation, select an empty column adjacent to your raw data, choose the cell corresponding to the first data point, and input the formula referencing the original value. For instance, if your raw data starts in cell A2, the formula in cell B2 would be **=LOG10(A2)**.

The subsequent image demonstrates the practical implementation of the log transformation within a standard Excel worksheet, showing the raw data column alongside the newly calculated, transformed values. Ensure that all raw data points are strictly positive, as the standard [logarithm](#) is undefined for zero and negative numbers.

	A	B	C	D	E	F	G	H
1		Data	Log(Data)					
2		1	0	=LOG10(B2)				
3		1	0					
4		1	0					
5		2	0.30103					
6		2	0.30103					
7		2	0.30103					
8		2	0.30103					
9		2	0.30103					
10		2	0.30103					
11		2	0.30103					
12		3	0.477121					
13		3	0.477121					
14		6	0.778151					
15		7	0.845098					
16		8	0.90309					
17								
18								
19								
20								
21								
22								
23								
24								
25								

Once the transformation is complete, the crucial next step is verification. A successful transformation should result in a dataset that passes a formal [normality test](#), confirming that the data is now suitable for parametric analysis.

Validating the Transformation Using the Jarque-Bera Test

A transformation is only useful if it demonstrably improves the distributional properties of the data. To quantify this improvement, statisticians rely on formal goodness-of-fit tests, such as the [Jarque-](#)

Bera (JB) test. The JB test specifically evaluates whether the sample data's [skewness](#) (symmetry) and [kurtosis](#) (tailedness) match those theoretical values expected of a perfect normal distribution (where skewness is 0 and excess kurtosis is 0).

The test statistic (JB) is calculated using the sample size (n), the sample skewness (S), and the sample kurtosis (C) according to the following mathematical relationship:

$$JB = (n/6) * (S^2 + (C-3)/4)$$

A larger JB statistic indicates a greater deviation from normality. The core of the test revolves around the hypothesis structure:

H0 (The [Null Hypothesis](#)): The data is normally distributed.

Ha (The [Alternative Hypothesis](#)): The data is not normally distributed.

Under the assumption that H0 is true, the *JB* statistic approximates a Chi-squared distribution with 2 degrees of freedom ($JB \sim \chi^2(2)$). The test generates a corresponding [p-value](#). If this [p-value](#) falls below the predetermined significance level (alpha, typically set at 0.05), we reject the null hypothesis and conclude that the data is significantly non-normal. Conversely, a [p-value](#) greater than 0.05 allows us to fail to reject H0, suggesting the data is sufficiently normal for analysis.

The following figure displays the results of the [Jarque-Bera test](#) applied to both the original, raw data and the newly created log-transformed data. This comparison is essential for validating the efficacy of the transformation.

	A	B	C	D	E	F	G	H
1		Data	Log(Data)					
2		1	0					
3		1	0					
4		1	0					
5		2	0.30103					
6		2	0.30103					
7		2	0.30103					
8		2	0.30103					
9		2	0.30103					
10		2	0.30103					
11		2	0.30103					
12		3	0.477121					
13		3	0.477121					
14		6	0.778151					
15		7	0.845098					
16		8	0.90309					
17				Formulas used				
18	n	15	15	=COUNT(C2:C16)				
19	Skewness	1.499564	0.58723	=SKEW(C2:C16)				
20	Kurtosis	1.072465	-0.23689	=KURT(C2:C16)				
21	JB Statistic	6.340593	0.89717	=(C18/6)*(C19^2+(C20^2)/4)				
22	p-value	0.041991	0.638531	=CHISQ.DIST.RT(C21, 2)				
23								
24								
25								
26								
27								
28								

Examination of the results confirms the success of the operation: the raw data's p-value is less than 0.05, confirming non-normality. However, the log-transformed data yields a p-value significantly greater than 0.05. This clear change demonstrates that the [log transformation](#) successfully achieved a satisfactory level of normality, thus validating the approach for subsequent [statistical tests](#).

Implementing the Square Root Transformation in Excel

When data exhibits only moderate positive skewness, or when dealing specifically with count data (non-negative integers), the Square Root Transformation offers a less aggressive solution than the logarithmic approach. This transformation is particularly useful because it tends to stabilize the variance in datasets where the variance is proportional to the mean, a characteristic often observed in data following a Poisson distribution.

In Microsoft Excel, the square root operation is achieved using the intuitive function **=SQRT()**.

Similar to the log transformation, you simply input the formula referencing the cell containing the original data point that requires adjustment. For instance, if you are transforming a series of counts located in column C, the formula in your new column D would be **=SQRT(C2)**, which you would then drag down to apply to the entire dataset.

The visual aid below illustrates the application of the square root transformation and includes the necessary output from the [Jarque-Bera test](#). It confirms whether this milder manipulation was sufficient to meet the assumption of normality.

	A	B	C	D	E	F	G	H
1		Data	Sqrt(Data)					
2		1	1	=SQRT(B2)				
3		1	1					
4		1	1					
5		2	1.414214					
6		2	1.414214					
7		2	1.414214					
8		2	1.414214					
9		2	1.414214					
10		2	1.414214					
11		2	1.414214					
12		3	1.732051					
13		3	1.732051					
14		6	2.44949					
15		7	2.645751					
16		8	2.828427					
17				Formulas used				
18	n	15	15	=COUNT(C2:C16)				
19	Skewness	1.499564	1.133197	=SKEW(C2:C16)				
20	Kurtosis	1.072465	0.347779	=KURT(C2:C16)				
21	JB Statistic	6.340593	3.285935	=(C18/6)*(C19^2+(C20^2)/4)				
22	p-value	0.041991	0.193405	=CHISQ.DIST.RT(C21, 2)				
23								
24								
25								
26								

As the image shows, the resulting p-value for the square root transformed data is greater than the 0.05 significance threshold. This finding confirms that, for this specific dataset, the square root transformation successfully corrected the distributional issues and made the data suitable for the intended parametric [statistical tests](#).

Executing the Cube Root Transformation in Excel

The Cube Root Transformation serves a vital niche in the data cleaning process. It is the preferred method when the square root transformation proves too weak to correct the skewness, but the [log transformation](#) proves overly aggressive, potentially leading to an inverse or negative skew. Furthermore, the cube root method holds a distinct advantage: unlike the log or square root functions, which require strictly positive inputs, the cube root of a negative number is mathematically defined and yields a real number (e.g., the cube root of -8 is -2). This makes it the only generally applicable power transformation when the dataset contains both positive and negative measurements.

Excel does not possess a dedicated **=CUBEROOT()** function. Instead, the cube root calculation must be performed by leveraging the exponentiation operator. The cube root of a number is equivalent to raising that number to the power of one-third (1/3). Therefore, the appropriate formula structure in Excel is **=DATA^(1/3)**, where "DATA" is the cell reference to the original raw value.

The following screenshot provides a visual confirmation of the correct formula setup and demonstrates the resulting transformed data column in Excel, highlighting the simple yet powerful calculation required for this technique.

	A	B	C	D	E	F	G	H
1		Data	Data^(1/3)					
2		1	1					
3		1	1					
4		1	1					
5		2	1.259921					
6		2	1.259921					
7		2	1.259921					
8		2	1.259921					
9		2	1.259921					
10		2	1.259921					
11		2	1.259921					
12		3	1.44225					
13		3	1.44225					
14		6	1.817121					
15		7	1.912931					
16		8	2					
17				Formulas used				
18	n	15	15	=COUNT(C2:C16)				
19	Skewness	1.499564	0.971302	=SKEW(C2:C16)				
20	Kurtosis	1.072465	0.12497	=KURT(C2:C16)				
21	JB Statistic	6.340593	2.368332	=(C18/6)*(C19^2+(C20^2)/4)				
22	p-value	0.041991	0.306001	=CHISQ.DIST.RT(C21, 2)				
23								
24								
25								
26								

After applying the cube root transformation, the subsequent Jarque-Bera results must be analyzed. In this particular instance, the output confirms that the [p-value](#) associated with the cube root transformed data successfully exceeds the 0.05 threshold. This achievement confirms that the cube root transformation was also effective in rendering the raw data [normally distributed](#), providing a valid option for analysis.

Selecting the Optimal Transformation: A Comparative Analysis

Based on the demonstrated examples, all three mathematical transformations--Log, Square Root, and Cube Root--proved effective in correcting the severe distributional problems of the original raw data. Each transformed dataset yielded a Jarque-Bera [p-value](#) greater than 0.05, indicating successful approximation of normality.

When multiple transformations achieve the minimum requirement of normality (i.e., $p\text{-value} > 0.05$), researchers must decide which result is superior. The accepted statistical convention in this scenario is to select the transformation that yields the largest [p-value](#) in the [Jarque-Bera test](#). A larger p-value suggests that the transformed data's [skewness](#) and [kurtosis](#) are most closely

aligned with the theoretical parameters of a perfect normal distribution, thereby maximizing the robustness of subsequent parametric modeling.

In the specific context of these illustrations, the log transformation generated the highest p-value among the three methods. This result designates the log transformation as the optimal choice for this particular dataset, providing the strongest empirical evidence of successful distributional correction.