

Understanding Skewness: A Step-by-Step Guide Using Microsoft Excel

Authored by
Mohammed loot

November 8, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Understanding Skewness: A Step-by-Step Guide Using Microsoft Excel*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=13884>

In the critical field of [descriptive statistics](#), analyzing the shape of a dataset is often as crucial as determining its central tendency or variability. The statistical measure specifically designed to quantify this shape is known as **Skewness**. Fundamentally, skewness measures the degree of [asymmetry](#) exhibited by a dataset or a [probability distribution](#) around its mean. This powerful coefficient, which can result in a positive, negative, or near-zero value, offers essential insights into how data points are clustered and distributed across the range of possible values. A significant deviation from zero skewness signals non-normality in the data, alerting analysts that standard parametric models may not be appropriate without preliminary data transformation.

Understanding the sign of the skewness coefficient is key to interpreting the data's orientation. When a dataset displays **negative skew**, the tail of the [distribution](#) extends prominently toward the left, meaning the bulk of the data is concentrated toward higher values. Conversely, **positive skew** is characterized by a tail stretching toward the right, indicating that most observations are clustered at the lower end of the scale. While a value of exactly zero for [skewness](#) implies perfect symmetry (a theoretical ideal often associated with the [Normal distribution](#)), real-world empirical data rarely achieves this perfect balance. Therefore, values close to zero are generally interpreted as evidence of near-symmetry, validating the use of many standard statistical procedures.

The Core Concept of Skewness: Definition and Crucial Role in Analysis

[Skewness](#) is formally derived from the third standardized moment of a [distribution](#). Unlike measures of location (mean) or spread ([standard deviation](#)), skewness isolates the direction and magnitude of the data's density orientation. In practical terms, calculating this coefficient helps practitioners determine whether the majority of observations are clustered on the high or low side relative to the central average. This understanding is foundational across various disciplines, from quantitative finance--where skewed returns influence risk modeling--to industrial quality control, where biased measurements might suggest systemic issues in manufacturing processes.

The necessity of evaluating skewness becomes paramount when preparing data for inferential statistical modeling. Many powerful analytical techniques, including correlation analyses, t-tests, and Analysis of Variance (ANOVA), are predicated on the assumption that the underlying data follows a [Normal distribution](#). If the data exhibits high [skewness](#), this fundamental assumption is violated, potentially leading to inaccurate or misleading conclusions from the tests performed. Consequently, computing and interpreting the skewness coefficient acts as a vital diagnostic step, informing the analyst whether corrective measures, such as log or square root transformations, are required to normalize the data before proceeding to advanced statistical inference.

Furthermore, skewness fundamentally dictates the relationship between the three main measures of central tendency: the mean, the median, and the mode. In a perfectly symmetrical [distribution](#), these three measures are identical. However, when [asymmetry](#) is present, the mean is pulled disproportionately toward the direction of the extended tail due to the influence of extreme values (outliers). For instance, in a positively skewed scenario (right tail), the mean is typically greater than the median. Conversely, in a negatively skewed scenario (left tail), the mean is pulled below the median. Analyzing this comparative relationship provides a rapid, intuitive assessment of the data's inherent shape and bias.

Interpreting Skewness Values: Distinguishing Positive, Negative, and Zero Skew

A clear interpretation of the sign and magnitude of the skewness coefficient is essential for deriving meaningful conclusions about the structure of the data. A **Negative Skew** (or left-skewed distribution) is defined by its elongated tail on the left side, with the majority of the data mass concentrated on the right. This pattern often emerges in measurements that have a natural upper bound but no corresponding lower limit, such as ceiling-effect test scores where most participants score highly, but a few perform poorly. Statistically, the mean is lower than the median because the small, extreme values in the left tail drag the average downward.

In contrast, a **Positive Skew** (or right-skewed distribution) is characterized by a long tail extending toward the right, indicating that the bulk of the observations are concentrated on the left side of the [distribution](#). The classic example is the distribution of individual wealth or income; most people fall into the moderate salary range (clustering on the left), but a small number of ultra-high earners extend the tail significantly to the right. In this configuration, the mean will be greater than the median. Grasping this distinction is crucial for accurate modeling; for instance, understanding that a financial asset's returns are positively skewed suggests a higher likelihood of extreme gains, while negative skew suggests a higher probability of extreme losses.

A **Zero Skew** represents perfect symmetry, where data values are distributed identically on both sides of the central point. While true zero skew is statistically rare in observed data, a value close to zero indicates that the [distribution](#) can be reliably approximated by a symmetrical model, such as the Gaussian model. A widely accepted rule of thumb suggests that skewness coefficients falling between -0.5 and +0.5 indicate a moderately symmetrical distribution. If the absolute value of the coefficient exceeds 1, the distribution is generally considered highly skewed, necessitating careful analysis and potentially the utilization of non-parametric statistical methods.

Automating Calculation: Implementing the SKEW Function in Microsoft Excel

Calculating [skewness](#) manually, especially for large datasets, is an algebraically complex and time-consuming task. Fortunately, [Microsoft Excel](#) provides a powerful and efficient built-in function to handle this complexity automatically. The primary function for this purpose is simply **SKEW()**. This function is specifically engineered to calculate the sample skewness of a range of values, applying the standard formula that includes adjustments to correct for the inherent bias found when estimating population parameters from a limited sample.

The syntax for the SKEW function is straightforward, requiring only the contiguous range of numerical data as its single argument. The function takes the following simple form, where "array of values" refers to the cell range containing the numbers you intend to analyze:

=SKEW(array of values)

It is crucial to emphasize that the SKEW function in [Excel](#) calculates **sample skewness**. This involves a crucial correction factor that accounts for the fact that the data represents only a subset of a broader population. While Excel offers an alternative, **SKEW.P()**, for calculating population skewness (which omits this bias correction), the standard SKEW function remains the appropriate and recommended choice when dealing with empirical data samples intended to generalize findings to a wider population.

Deconstructing the Mathematical Formula for Sample Skewness

While the SKEW function simplifies execution, statistical proficiency requires familiarity with the underlying mathematical definition. The formula utilized by Excel's **SKEW()** function calculates the adjusted sample skewness coefficient, serving as an unbiased estimator of population skewness. The inclusion of the adjustment factor compensates precisely for the increased variability introduced when working with a limited [sample size](#). This inherent complexity confirms why automated software like Excel is indispensable for modern data analysis tasks.

The precise formula employed by the SKEW function is detailed below:

Skewness = $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$

Each element of this equation plays a distinct and critical role. The term inside the summation

symbol (Σ) represents the deviation of each data point from the mean, standardized by the [standard deviation](#) (effectively the Z-score), and then cubed. Cubing this value ensures that extreme deviations (outliers) far from the mean are given greater weight, accurately reflecting the [asymmetry](#). The bracketed term, $\frac{1}{n}$, is the essential factor applied specifically for sample estimation to ensure the result is unbiased.

The variables utilized in the sample skewness formula are defined as follows:

n = The total [sample size](#), representing the count of data points in the array.

Σ = The summation operator, directing that the standardized cubed difference must be calculated for every observation and aggregated.

x_i = The value of the *i*th individual observation within the dataset.

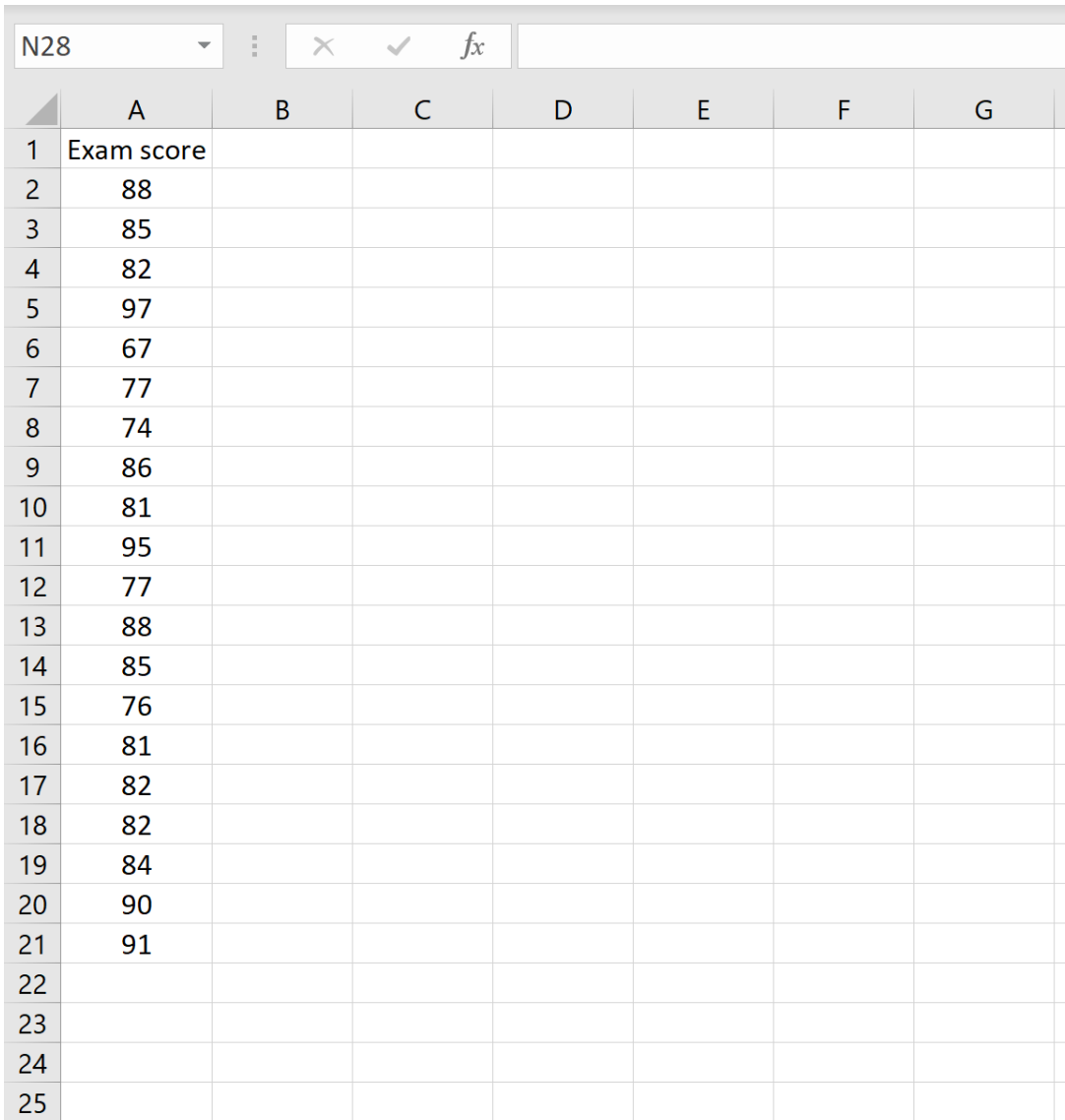
\bar{x} = The arithmetic mean of the entire sample dataset.

s = The [standard deviation](#) of the sample.

Practical Application: Step-by-Step Calculation in Microsoft Excel

To demonstrate the efficiency of the SKEW function, consider a hypothetical dataset comprising 20 observations, stored in the cell range A2 through A21 within an [Excel](#) worksheet. The objective is to swiftly ascertain the [asymmetry](#) of this data.

The initial step requires setting up the data structure. The visual representation below illustrates the dataset prepared for calculation:



	A	B	C	D	E	F	G
1	Exam score						
2	88						
3	85						
4	82						
5	97						
6	67						
7	77						
8	74						
9	86						
10	81						
11	95						
12	77						
13	88						
14	85						
15	76						
16	81						
17	82						
18	82						
19	84						
20	90						
21	91						
22							
23							
24							
25							

To calculate the sample skewness for this specific [distribution](#), we select any empty cell (for example, cell C2) and input the SKEW function, ensuring the reference encompasses the entire data range. Since the data spans from A2 to A21, the exact formula entered is **=SKEW(A2:A21)**. Excel immediately executes the complex sample skewness calculation across all 20 data points and returns the resulting coefficient.

The result yielded by executing the function is displayed below, confirming the automated nature of the process:

	A	B	C	D	E	F
1	Exam score		Skewness	Formula used		
2	88		-0.1849	=SKEW(A2:A21)		
3	85					
4	82					
5	97					
6	67					
7	77					
8	74					
9	86					
10	81					
11	95					
12	77					
13	88					
14	85					
15	76					
16	81					
17	82					
18	82					
19	84					
20	90					
21	91					
22						
23						
24						

The computed skewness for this particular dataset is **-0.1849**. Because this value is negative, we can deduce immediately that the distribution is slightly left-skewed, meaning the tail extends toward the left and the mean is marginally less than the median. Crucially, since the absolute value (-0.1849) is extremely close to zero (comfortably within the accepted -0.5 to +0.5 range), we can confidently conclude that the dataset is highly symmetrical and is suitable for most parametric statistical tests without requiring advanced data transformation techniques.

Troubleshooting: Addressing Common Technical Errors in Skewness

Calculation

While the SKEW function is robust, users must be cognizant of specific data conditions that can lead to calculation failures. Since the skewness formula involves division by expressions derived from the [sample size](#) (n) and the sample [standard deviation](#) (s), certain scenarios can cause the calculation to halt, resulting in the notorious **#DIV/0!** error. This error fundamentally signifies that the denominator in the underlying mathematical formula has evaluated to zero, rendering the computation mathematically impossible.

The two primary scenarios outlined below detail when the SKEW() function will return this division error:

Technical Note: SKEW() Error Scenarios

The SKEW() function will return the error #DIV/0! in the following two scenarios, both of which result in a division by zero in the underlying formula:

If there are fewer than three data points ($n < 3$). The sample adjustment factor requires that both $n-1$ and $n-2$ are non-zero. A minimum of three data points is statistically necessary to reliably estimate sample skewness.

If the sample standard deviation (s) is zero. This situation occurs exclusively when all values in the dataset are identical (e.g., 10, 10, 10, 10...). If there is no variation, the denominator related to the standardization step becomes zero, preventing the calculation.

To mitigate these technical issues, analysts must always confirm that their dataset contains at least three unique observations. Furthermore, in cases of entirely uniform data, while skewness is theoretically zero (due to perfect symmetry), the SKEW function cannot numerically compute this because of the division by zero in the standardization process. In such uniform instances, a simple visual inspection or confirmation of zero [standard deviation](#) provides sufficient, alternative evidence of symmetry.

Alternative Methods: Utilizing Dedicated Online Statistical Calculators

Although [Excel](#) is the standard tool for embedded data analysis, dedicated online statistical calculators offer valuable alternatives for result verification or for simultaneously calculating multiple statistics, such as both skewness and [kurtosis](#). These external tools typically require only the raw data input and provide instantaneous results, often accompanied by comprehensive statistical summaries. This methodology is particularly useful for students or researchers seeking

rapid, cross-validated results without the need for spreadsheet formatting.

A typical online skewness and [kurtosis](#) calculator allows the user to paste or type the raw data values directly into an input field. Upon clicking the "Calculate" button, the necessary statistical algorithms are executed server-side. This streamlined approach bypasses the need for careful cell referencing and formula input within a spreadsheet environment, accelerating the calculation process for singular analytical tasks.

To illustrate this verification process, we use the same dataset from the previous Excel example to generate an external calculation. The data input results in the following output:

Skewness and Kurtosis Calculator

Skewness is a measure of the asymmetry of a dataset or distribution. This value can be positive or negative. A negative skew typically indicates that the *tail* is on the left side of the distribution. A positive value typically indicates that the tail is on the right.

Kurtosis is simply a measure of the "tailedness" of a dataset or distribution. The kurtosis formula used by this calculator is identical to the formula used in Excel, which finds what is known as *excess kurtosis*.

To find the skewness and kurtosis of a dataset, simply enter the comma-separated values in the box below, then click the "Calculate" button.

88, 85, 82, 97, 67, 77, 74, 86, 81, 95, 77, 88, 85, 76,
81, 82, 82, 84, 90, 91

CALCULATE

Skewness: -0.18490

As confirmed by the calculator output, the skewness value is precisely **-0.1849**. This confirms that fundamental statistical measures are consistent regardless of the computational environment utilized. Whether employing the integrated SKEW function in [Excel](#) or a sophisticated external tool, the interpretation of the resulting [skewness](#) coefficient--determining the direction and degree of [asymmetry](#)--remains the crucial cornerstone of effective statistical data analysis.