

Understanding and Calculating Pearson's Coefficient of Skewness in Excel: A Step-by-Step Guide

Authored by
Mohammed Iooti

November 5, 2025

RECOMMENDED CITATION

Mohammed Iooti (2025). *Understanding and Calculating Pearson's Coefficient of Skewness in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=10609>

The analysis of data distribution is foundational to statistical inference. A key characteristic that statisticians must quantify is the lack of symmetry, or **skewness**, within a dataset. The most widely recognized method for this quantification is the [Pearson's coefficient of skewness](#), developed by the distinguished mathematician and statistician, [Karl Pearson](#). This coefficient provides a simple, interpretable measure of how far a distribution deviates from a perfectly symmetrical bell curve. Understanding this measure is essential because the skewness of a distribution dictates the appropriate choice of statistical tests and models used in subsequent analysis.

Unlike the moment-based calculations of skewness, Pearson's approach relies on comparing the central tendency measures--the mean, median, and mode--relative to the data's overall dispersion, typically represented by the standard deviation. Statisticians employ two main formulas derived from Pearson's framework, with the selection largely depending on which measure of central tendency--the mode or the median--is deemed most representative and stable for the specific dataset under investigation. Both methods aim to provide a robust estimate of the distribution's asymmetry.

The Foundation: Quantifying Asymmetry with Pearson's Coefficients

Pearson's framework offers two distinct methods for calculating skewness, both designed to capture the directional bias of the data. When a dataset is perfectly symmetrical, the mean, median, and mode will be identical, resulting in a coefficient of zero. Any deviation from this perfect alignment indicates asymmetry, and the magnitude and sign of the coefficient reveal the degree and direction of the skew. These coefficients are particularly useful in practical settings, such as business intelligence or scientific research, where assumptions about data normality must often be tested before hypothesis testing can proceed.

The core principle behind both formulas is the measurement of the distance between the distribution's center (the mean) and its peak (either the mode or the median), scaled by the [standard deviation](#) to standardize the result. This normalization allows for meaningful comparison of skewness across different datasets, regardless of their original scale or units of measurement. Mastering these calculations in a versatile environment like Microsoft Excel enables rapid and reliable data diagnostics, serving as a critical first step in exploratory data analysis.

Formulaic Approaches: Pearson's First and Second Coefficients

Pearson's two coefficients of skewness represent slightly different statistical perspectives on central tendency, leading to distinct calculation methods. While both are valid, the choice between them often comes down to the reliability and stability of the median versus the mode in the face of complex or outlier-heavy data. The simplicity and straightforwardness of the first coefficient contrast with the statistical robustness of the second, which is generally favored in modern

statistical practice.

Method 1: Utilizing the Mode (Pearson's First Skewness Coefficient)

This formula is conceptually simple, comparing the difference between the arithmetic [mean](#) and the [mode](#), and then normalizing this difference by the sample standard deviation. It assumes that the mode provides a stable, accurate representation of the distribution's peak. However, this assumption frequently breaks down in real-world data, especially if the distribution is multimodal (has multiple modes) or if the distribution is highly irregular, making the mode difficult to determine accurately or reliably.

The formula is expressed as:

$$\text{Skewness} = (\text{Mean} - \text{Mode}) / \text{Sample standard deviation}$$

While this method is easy to calculate, especially when the mode is clearly defined, its dependency on a single point estimate (the mode) often makes it less preferred than the median-based approach for general analytical purposes.

Method 2: Utilizing the Median (Pearson's Second Skewness Coefficient)

The second coefficient, often referred to as the preferred or robust method, utilizes the [median](#). This formula capitalizes on the relationship that exists in moderately skewed distributions where the distance between the mean and the median is approximately one-third of the distance between the mean and the mode. Crucially, the median is a more resilient measure of central tendency than the mode, particularly in the presence of extreme outliers or when dealing with highly skewed data, where the mode can be highly volatile.

The formula is:

$$\text{Skewness} = 3(\text{Mean} - \text{Median}) / \text{Sample standard deviation}$$

Statisticians generally recommend using this second coefficient because the median's resistance to extreme values provides a more stable and reliable estimate of the distribution's true asymmetry, ensuring that the skewness measure is not unduly influenced by a few anomalous data points. For calculations performed in software environments like Excel, where efficiency and reliability are paramount, the median-based approach offers superior analytical rigor.

Step 1: Structuring the Data Environment in Excel

Before applying either of Pearson's formulas, the raw data must be properly organized within an Excel worksheet. This foundational step involves not only entering the raw measurements but also

establishing dedicated cells for the calculation of the necessary statistical components: the mean, median, mode, and standard deviation. Careful preparation ensures accuracy and provides a clear structure for tracing the calculations, which is vital for auditing statistical results.

We begin by inputting our raw data into a continuous range, such as Column A. This hypothetical dataset, representing continuous measurements, will serve as the basis for calculating all central tendencies and dispersion measures required by Pearson's coefficients.

Let us input the following hypothetical data points into Column A of a new worksheet, representing any continuous measurement:

	A	B	C	D	E	F	G
1	Data						
2	3						
3	4						
4	4						
5	4						
6	7						
7	8						
8	12						
9	13						
10	13						
11	14						
12	15						
13	18						
14	22						
15	24						
16	25						
17	26						
18	26						
19	27						
20	29						
21	32						
22							
23							
24							
25							
26							

Once the data range is defined (e.g., A1:A15), the next step is to use Excel's powerful built-in functions to compute the required statistical parameters: `AVERAGE()` for the mean, `MODE.SNGL()` for the mode, `MEDIAN()` for the median, and `STDEV.S()` for the sample standard deviation. Using the sample standard deviation (which divides by n-1) is appropriate when analyzing a sample dataset to estimate the population skewness.

Step 2: Calculating Pearson's First Coefficient (Mode-Based) in Excel

The calculation of Pearson's First Coefficient of Skewness requires the mean, the mode, and the standard deviation. We implement the formula **Skewness = (Mean - Mode) / Sample standard deviation** by nesting the relevant Excel statistical functions within a single calculation cell. This approach minimizes error and makes the calculation highly reproducible.

To achieve this calculation efficiently in Excel, we use the following combination of functions, assuming our data is in range A1:A15:

The combined formula structure should first calculate the numerator (Mean - Mode) and then divide it by the denominator (Standard Deviation). In Excel terminology, this translates to: `= (AVERAGE(A1:A15) - MODE.SNGL(A1:A15)) / STDEV.S(A1:A15)`. The `MODE.SNGL()` function is used to handle single-mode scenarios, which is appropriate for this calculation.

The implementation looks like this in the spreadsheet:

	A	B	C	D	E	F	G	H	I
1	Data								
2	3		1.295	<code>= (AVERAGE(A2:A21) - MODE.SNGL(A2:A21)) / STDEV.S(A2:A21)</code>					
3	4								
4	4								
5	4								
6	7								
7	8								
8	12								
9	13								
10	13								
11	14								
12	15								
13	18								
14	22								
15	24								
16	25								
17	26								
18	26								
19	27								
20	29								
21	32								
22									
23									
24									
25									
26									

Upon executing the formula shown in the image above, the resulting value for the Pearson's First Coefficient of Skewness is calculated to be **1.295**. This initial result, being significantly greater than

zero, immediately indicates a substantial degree of positive skewness in the distribution. However, given the known instability of the mode, we must proceed to the preferred median-based calculation to confirm this finding.

Step 3: Calculating Pearson's Second Coefficient (Median-Based) in Excel

The next step involves calculating the Pearson's Second Coefficient of Skewness, which is the gold standard for measuring asymmetry due to the median's robustness. The formula, **Skewness = 3(Mean - Median) / Sample standard deviation**, only requires the introduction of the median calculation, as the mean and standard deviation components remain consistent from the previous step. This calculation is vital because it often provides a more conservative and reliable estimate of the true distributional shape.

The revised calculation leverages Excel's `MEDIAN()` function, multiplying the difference between the mean and median by three before normalizing by the standard deviation. The formula structure is therefore: `= (3 * (AVERAGE(A1:A15) - MEDIAN(A1:A15))) / STDEV.S(A1:A15)`.

The setup in Excel is as follows:

	A	B	C	D	E	F	G	H	I
1	Data								
2	3		0.569	<code>=3*(AVERAGE(A2:A21)-MEDIAN(A2:A21))/STDEV.S(A2:A21)</code>					
3	4								
4	4								
5	4								
6	7								
7	8								
8	12								
9	13								
10	13								
11	14								
12	15								
13	18								
14	22								
15	24								
16	25								
17	26								
18	26								
19	27								
20	29								
21	32								
22									
23									
24									
25									
26									
27									

Using the preferred median approach, the calculated skewness value is determined to be **0.569**. The significant difference between this result (0.569) and the mode-based result (1.295) underscores the critical importance of selecting the appropriate formula. In scenarios where the mode is unstable or poorly defined, the median-based coefficient offers a superior assessment of the underlying distribution.

Interpreting the Results: What the Coefficient Value Means

Once the Pearson coefficient of skewness is calculated, its value serves as a numerical summary of the distribution's shape. Interpreting this value involves understanding three primary outcomes: zero, positive, or negative. Furthermore, the magnitude of the coefficient helps determine if the skewness is negligible, moderate, or severe. A common rule of thumb suggests that coefficients between -0.5 and +0.5 indicate reasonably symmetrical data, which may be close enough to assume normality for some statistical tests.

The coefficient's sign is directly tied to the relationship between the mean, median, and mode, offering a visual cue about the distribution, often represented by a [histogram](#):

A Value of 0: This signifies zero [skewness](#). In a perfectly symmetrical distribution, the mean, median, and mode coincide at the exact center, meaning the data points are equally distributed on both sides.

A Positive Value (e.g., 0.569 or 1.295): This indicates a positive skew, often called "right skew." In this scenario, the tail of the distribution extends towards the higher values (the right side). Statistically, the mean is greater than the median, as the extreme high values pull the mean towards the right.

A Negative Value: This indicates a negative skew, or "left skew." Here, the distribution's tail extends towards the lower values (the left side). Consequently, the mean is typically less than the median, as the extreme low values pull the mean towards the left.

Based on our Excel example, the preferred median-based calculation yielded a positive result of 0.569. This leads to the confident conclusion that the dataset exhibits a **positive skew**. Although the skewness is not severe (falling outside the tightest definition of symmetry but close to the moderate range), it confirms that the data points are clustered more densely on the left side of the distribution, resulting in a longer, thinner tail trailing out to the right. This information is crucial for subsequent statistical modeling, as highly skewed data may require transformation before parametric tests can be reliably applied.

Conclusion: Reliability and Application in Statistical Analysis

The ability to quickly and accurately calculate Pearson's coefficient of skewness in Excel is an indispensable skill for any data analyst or researcher. It provides immediate, quantifiable evidence

regarding the symmetry of a dataset, influencing crucial decisions regarding data preprocessing and statistical methodology. While Pearson offered two valid coefficients, the median-based approach (the Second Coefficient) is overwhelmingly recommended for its inherent stability and resistance to the complications arising from multimodal or outlier-laden distributions.

The determination of whether data is positively or negatively skewed is vital. If significant skewness is detected, it may violate the assumption of normality required by powerful statistical tools like t-tests or ANOVA, necessitating the use of non-parametric alternatives or requiring the application of data transformation techniques (such as log or square root transformations) to mitigate the asymmetry. Therefore, always prioritize the Pearson's Second Coefficient for stability, and use the resulting value to inform your selection of appropriate statistical tests and models, ensuring the rigor and validity of your entire analytical process.