How to Easily Calculate Pearson's Coefficient of Skewness in Excel

Authored by

stats writer

December 6, 2025

The Pearson's coefficient of skewness is a fundamental measure used in descriptive statistics to quantify the asymmetry of a given probability distribution. A perfectly symmetrical distribution, such as the normal distribution, yields a skewness value of zero. However, real-world data often deviates from perfect symmetry, displaying either a positive or negative skew.

To accurately calculate this coefficient within a tool like **Excel**, a foundational understanding of several core statistical metrics is required. Specifically, one must first determine the mean, the standard deviation, and either the median or the mode for the data set. While the original method sometimes involves calculating the third moment, the two practical methods most commonly employed in applied statistics leverage these central tendency measures to derive the final skewness value.

Understanding how to implement these calculations in **Excel** provides analysts with a powerful tool for quickly assessing the characteristics of their data. Although Excel provides a built-in function for related calculations (`SKEW`), using Pearson's coefficient allows for a more intuitive interpretation derived directly from the relationships between the dataset's core statistical properties, offering an accessible measure of distribution symmetry.

The metric known as the **Pearson's coefficient of skewness** was developed by the renowned biostatistician, Karl Pearson. This coefficient serves as an essential tool for quantifying the degree of asymmetry, or lack thereof, within a sample dataset. It helps statisticians determine if the data points are distributed evenly around the central value or if they are clustered to one side, pulling the distribution's tail in the opposite direction.

Pearson established two distinct methods for calculating this coefficient. Both approaches utilize the data's central tendency measures (mean, median, or mode) in conjunction with the data's spread (standard deviation), ensuring a robust measure that accounts for both the location and variability of the data points. These two methods are known as the first and second coefficients of skewness.

We will explore both primary calculation methodologies in detail, providing the practical context necessary for implementation in **Excel**. While mathematically distinct, both aim to achieve the same goal: providing a numerical value that summarizes the shape of the data distribution.

Understanding the Two Pearson Skewness Formulas

The two coefficients of skewness developed by Pearson are based on slightly different foundational statistics, leading to subtle variations in the resulting value, especially in highly non-normal distributions.

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

This method relies on the difference between the arithmetic Mean and the Mode, which is the most frequently occurring value in the dataset. The concept here is that in a perfectly symmetrical distribution, the Mean and the Mode are identical, leading to zero skewness.

The formula for the first coefficient is:

Skewness = (Mean - Mode) / Sample standard deviation

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

The second method uses the relationship between the Mean and the Median. This formula is particularly useful because the Median is less susceptible to the influence of extreme outliers compared to the Mode. This coefficient is derived from the empirical relationship that, for moderately skewed distributions, the difference between the Mean and the Median is approximately one-third of the difference between the Mean and the Mode.

The formula for the second coefficient is:

Skewness = 3(Mean - Median) / Sample standard deviation

Why the Median-Based Method is Preferred

While both formulas are valid ways to calculate Pearson's coefficient, the second method, which utilizes the **Median**, is generally preferred by statisticians for several important reasons related to data reliability and distribution characteristics.

The primary reason for preferring the second method lies in the inherent limitations of the **Mode**. The Mode, defined as the most frequent observation, can be highly unstable in smaller datasets or those with many unique values. A dataset can also be multimodal (having more than one mode), which complicates the calculation and interpretation of Pearson's first coefficient. Because of its sensitivity to minor data fluctuations, the Mode is often not considered a reliable indicator of the dataset's true "central" value.

Conversely, the **Median** (the middle value when data is ordered) is a robust measure of central tendency. It remains unaffected by extreme outliers, making it a much better indicator of the true center of a skewed distribution. Therefore, utilizing the Median ensures that the resulting skewness coefficient provides a more stable and representative measure of the distribution's asymmetry, making the second method the default choice in many statistical analyses.

Step 1: Preparing the Dataset in Excel

To demonstrate the calculation process, we must first establish a representative dataset within our

Excel spreadsheet. This dataset will serve as the basis for calculating all necessary statistical measures: the mean, median, mode, and standard deviation.

Begin by opening a new worksheet and entering the sample data points into a single column, starting perhaps in cell A1. For this example, we will use a small dataset to illustrate the calculations clearly.

First, let's create the following dataset in Excel:

	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							

Step 2: Calculating Necessary Statistical Components

Before calculating the final skewness coefficients, we must determine the core statistical values for the dataset (A1:A10). We will use built-in **Excel** functions to ensure accuracy. These components are essential as they form the numerator and denominator of both Pearson formulas.

The required components are:

Mean: Calculated using the formula `=AVERAGE(A1:A10)`. This provides the arithmetic average of all values.
Median: Calculated using the formula `=MEDIAN(A1:A10)`. This finds the 50th percentile of the data.
Mode: Calculated using the formula `=MODE.SNGL(A1:A10)`. This finds the single most frequently occurring number.
Sample Standard Deviation (SD): Calculated using the formula `=STDEV.S(A1:A10)`. We use the 'S' version as we are typically analyzing a sample, not the entire population.

Calculating these values first simplifies the final skewness calculation immensely, making the process less error-prone and easier to audit.

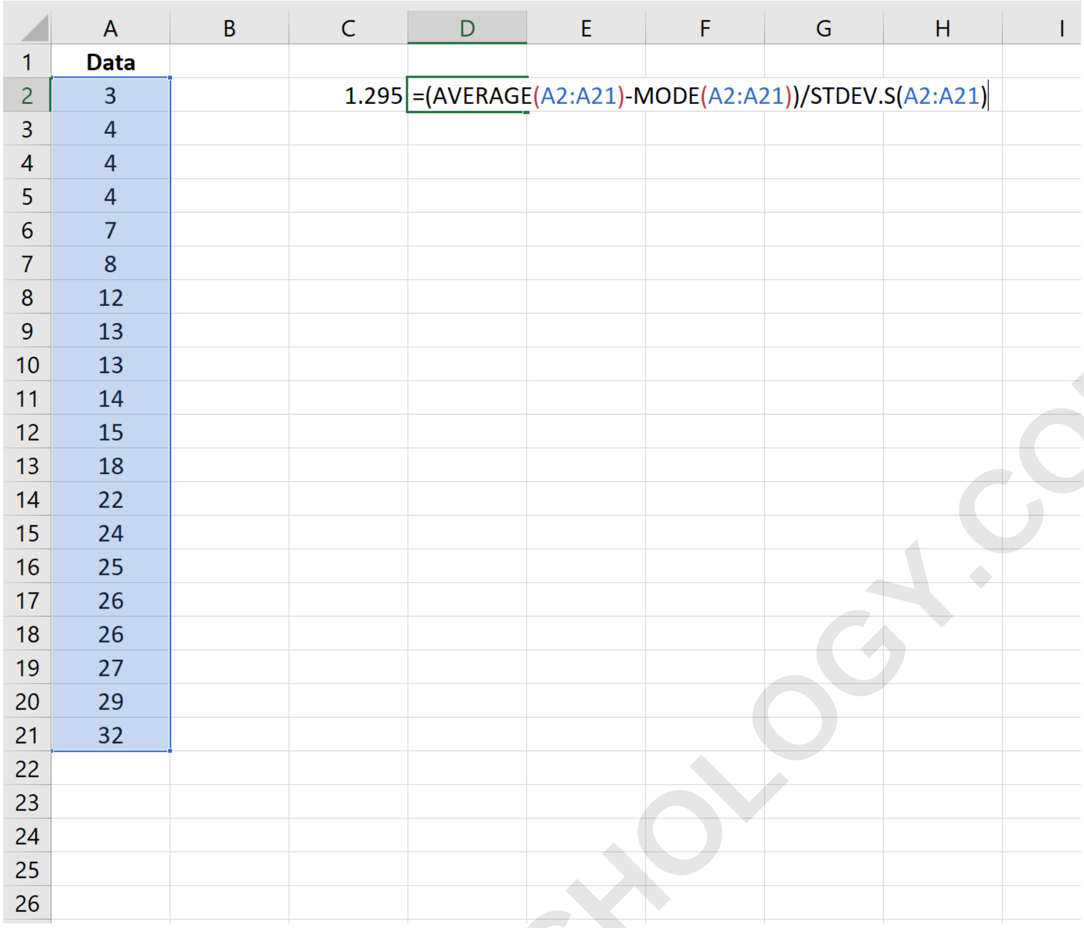
Step 3: Calculating Skewness using the Mode (Method 1)

We will now proceed with calculating Pearson's First Coefficient of Skewness, utilizing the calculated **Mean** and **Mode** values, and dividing the difference by the **Sample Standard Deviation**. This method gives us a rapid assessment of asymmetry.

In **Excel**, we reference the cells containing the pre-calculated statistical values. If, for instance, the Mean is in cell B1, the Mode is in B3, and the Standard Deviation is in B4, the formula would be structured according to the mathematical definition:

```
=(B1 - B3) / B4
```

Executing this calculation yields the first skewness coefficient for our dataset. Next, we can use the following formula (displayed visually in the image below) to calculate the Pearson Coefficient of Skewness using the mode:



	A	B	C	D	E	F	G	H	I
1	Data								
2	3		1.295	=(AVERAGE(A2:A21)-MODE(A2:A21))/STDEV.S(A2:A21)					
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									

Based on the example data provided, the skewness turns out to be **1.295**. This high positive value suggests a pronounced level of right skewness in the distribution, where the Mean is significantly greater than the Mode.

Step 4: Calculating Skewness using the Median (Method 2)

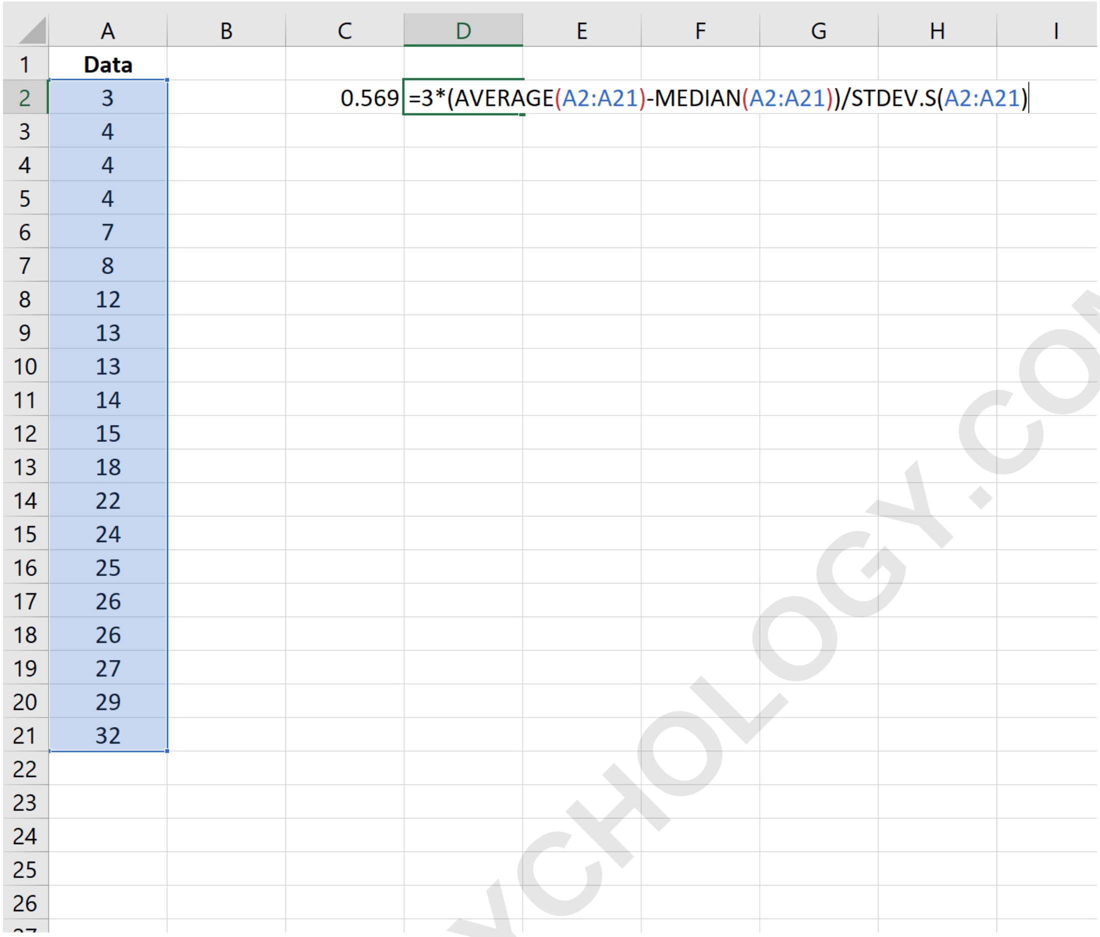
Next, we calculate the generally preferred measure: Pearson's Second Coefficient of Skewness, which uses the **Median**. This calculation is considered more stable and reliable, particularly when dealing with distributions that are heavily skewed or contain outliers.

The formula requires multiplying the difference between the Mean and the Median by three before dividing by the Standard Deviation. Assuming the Mean is in B1, the Median is in B2, and the Standard Deviation is in B4, the structure of the **Excel** formula is:

```
=3 * (B1 - B2) / B4
```

This adjustment (multiplying by three) is crucial as it accounts for the typical statistical relationship between the mean, median, and mode in moderately skewed datasets. We can also use the following formula (displayed visually in the image below) to calculate the Pearson Coefficient of

Skewness using the median:



For our sample data, the skewness turns out to be **0.569**. This value is lower than the first coefficient, which is common. The fact that it is still positive indicates right skewness, confirming the asymmetry found by the first method, but providing a more normalized, robust estimation of the magnitude of that skewness.

Interpreting the Results of the Skewness Coefficient

Interpreting the numerical result of the Pearson coefficient is straightforward and critical for understanding the shape of the underlying distribution. The coefficient indicates both the direction (positive or negative) and the magnitude of the asymmetry relative to the data's standard deviation.

We interpret the Pearson coefficient of skewness in the following ways:

A **value of 0** indicates no skewness. If we created a histogram to visualize the distribution of values in a dataset, it would be perfectly symmetrical, meaning the Mean, Median, and Mode would all coincide. This is characteristic of ideal distributions like the Normal distribution.

A **positive value** indicates positive skew or "right" skew. This occurs when the tail of the distribution extends further to the right. Statistically, this means the **Mean** is greater than the **Median** (and often the Mode), as the higher values pull the average upward.
A **negative value** indicates a negative skew or "left" skew. Here, the tail extends further to the left, indicating that lower values are pulling the average down. The **Mean** is less than the **Median** (and often the Mode).

In our previous example, both Pearson coefficients calculated (1.295 using the Mode and 0.569 using the Median) were positive, which strongly indicates that the distribution of data values was positively skewed or "right" skewed. This suggests that while most data points cluster on the lower end, there are some extreme, high-value outliers present in the dataset.

For a more detailed visual breakdown of how positive (right) and negative (left) skewness appear when plotted, it is helpful to consult dedicated statistical resources. Understanding this visual representation solidifies the interpretation of the numerical coefficient.