

How to Easily Measure Skewness and Kurtosis for Data Analysis

Authored by
stats writer

December 1, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Measure Skewness and Kurtosis for Data Analysis*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=103227>

Skewness and kurtosis are fundamental quantitative measures used in descriptive statistics to characterize the shape of a probability distribution. While measures of central tendency (mean, median, mode) and dispersion (standard deviation, range) describe the location and spread of the data, skewness and kurtosis provide crucial insights into the overall structural form of the dataset. Understanding these metrics is essential for selecting appropriate statistical tests and building accurate predictive models, as many statistical methods assume data follows a specific shape, most often the Normal distribution.

The core distinction between these two measures lies in the specific aspect of the distribution shape they quantify. **Skewness** focuses on the degree of asymmetry, indicating whether data points are clustered more heavily on one side of the mean, resulting in a tail extending in the opposite direction. Conversely, **kurtosis** provides information regarding the "tailedness" of the distribution, quantifying the frequency and extremity of outliers compared to a standard bell curve. Together, these statistics allow researchers to move beyond simple averages and assess the true underlying structure of the data, identifying whether a distribution is normal, skewed, or leptokurtic.

Understanding Skewness: Asymmetry in Data Distributions

Skewness, statistically defined as the third standardized moment, measures the extent to which a probability distribution deviates from symmetry. A perfectly symmetrical distribution, such as the Normal distribution, has a skewness value of zero, meaning the data values are evenly balanced around the mean, and the mean, median, and mode are approximately equal. When the distribution is asymmetric, the mean is pulled away from the center of the distribution towards the longer tail, influencing the skewness coefficient.

There are three primary categories of skewness, each offering distinct insights into the concentration of data. A **negative skew**, also known as left-skewness, occurs when the tail extends towards more negative values, meaning the mass of the distribution is concentrated on the right. In such cases, the mean is typically less than the median. Conversely, a **positive skew**, or right-skewness, occurs when the tail extends towards more positive values, and the majority of the data is clustered on the left side. Here, the mean is typically greater than the median. Understanding the direction of the skew is vital, as it often reflects real-world phenomena; for example, income data is typically positively skewed because a few individuals earn extremely high salaries, pulling the mean upwards.

Quantifying skewness often involves the use of Pearson's coefficients or the calculation based on the third moment. The formulas calculate a numerical value that ranges from negative infinity to positive infinity. The magnitude of this value indicates the strength of the asymmetry. For instance, a skewness value of -1.5 suggests a more substantial asymmetry than a value of -0.5. Interpretation guidelines often suggest that skewness values between -0.5 and 0.5 indicate near-

symmetry, while values outside the range of -1 and 1 suggest highly skewed data requiring careful consideration before applying parametric statistical tests.

Interpreting the Direction and Magnitude of Skew

The sign of the skewness coefficient dictates the orientation of the distribution's tail, which is the key feature defining the asymmetry. The coefficient provides a precise metric for assessing the degree of non-normality in terms of symmetry, which is crucial for statistical modeling assumptions.

Negative Skew (Left-Skewed): The left tail is longer or heavier than the right side. This indicates that there are relatively few extremely low values pulling the mean down. The bulk of the observations are positioned towards the higher end of the scale.

Positive Skew (Right-Skewed): The right tail is longer or heavier than the left side. This signifies the presence of a few extremely high values that disproportionately increase the mean. The majority of observations are concentrated near the lower end of the scale.

Zero Skew (Symmetrical): The distribution is perfectly balanced, and the shape on one side of the center mirrors the shape on the other. In practical data analysis, a coefficient very close to zero is usually sufficient to assume symmetry.

Proper interpretation of skewness is critical for assumptions validation. If a distribution is heavily skewed, it violates the assumption of normality required by tests like t-tests or ANOVA. In such scenarios, researchers must consider transforming the data (e.g., using logarithmic or square root transformation) or employing non-parametric statistical methods that do not rely on assumptions about the distributional shape.

Measuring Kurtosis: Analyzing Tail Extremity

Kurtosis, derived from the fourth standardized moment, is a statistical measure that quantifies the shape of the probability distribution, specifically focusing on the peakedness and the heaviness of the tails relative to the Normal distribution. While skewness measures horizontal deviation from symmetry, kurtosis measures vertical deviation, reflecting how much of the variance is due to infrequent, extreme deviations (outliers) versus frequent, moderate deviations.

It is crucial to understand that kurtosis is fundamentally a measure of the weight of the tails. High kurtosis indicates that the distribution has heavy tails, meaning it is prone to producing more frequent and/or more extreme outliers than a normal curve. Conversely, low kurtosis suggests light tails, implying fewer extreme outliers. This metric is especially important in fields like finance and quality control, where the occurrence of rare but high-impact events must be accurately modeled for risk assessment.

Statisticians utilize two main definitions for kurtosis: the standard definition (where the kurtosis of a Normal distribution is 3) and the excess kurtosis (Fisher's definition). The standard definition dictates that the kurtosis of a perfectly Normal distribution is exactly 3. This benchmark value is critical for comparative analysis. However, many statistical software packages report **Excess Kurtosis**, which is the standard kurtosis minus 3. The use of excess kurtosis simplifies comparisons: a value of zero indicates a normal distribution, positive values indicate heavier tails (leptokurtic), and negative values indicate lighter tails (platykurtic). Researchers must always confirm which definition their software uses when reporting results.

Classifying Distributions Based on Kurtosis Values

Based on the comparison to the normal distribution (where excess kurtosis is 0 or standard kurtosis is 3), distributions are categorized into three distinct types:

Mesokurtic: These distributions exhibit an excess kurtosis value of 0 (or a standard kurtosis of 3). The term "mesokurtic" implies that the distribution's tails are similar in weight and extremity to those of a standard Normal distribution. This represents the necessary baseline against which other distributions are measured.

Leptokurtic: These distributions have a positive excess kurtosis (standard kurtosis greater than 3). Leptokurtic distributions are characterized by heavier tails and a sharper peak than the normal curve. The presence of these heavy tails signifies a higher probability of observing extreme values or outliers.

Platykurtic: These distributions have a negative excess kurtosis (standard kurtosis less than 3). Platykurtic distributions are often described as having "light tails" and a flatter peak compared to the normal curve. They tend to produce fewer and less extreme outliers than the mesokurtic distribution, meaning variance is more evenly distributed.

The interpretation of kurtosis provides crucial context for estimating risks. For instance, if data--such as financial returns--are highly leptokurtic, relying solely on variance measures like standard deviation will underestimate the probability of extreme deviations, necessitating specialized risk modeling techniques that explicitly account for the heavy tails.

Mathematical Foundations: Calculating Moments

While modern statistical software abstracts the complex computations, understanding the underlying mathematical basis--the moments of a distribution--is vital. Skewness is based on the third standardized moment, and kurtosis is based on the fourth standardized moment. The standardization process ensures that the resulting coefficients are independent of the units of measurement and scale of the data, allowing for meaningful comparison across diverse datasets.

The general structure for calculating these coefficients involves deviation from the mean raised to the power corresponding to the moment being calculated (3 for skewness, 4 for kurtosis), standardized by the standard deviation raised to the same power. This normalization ensures the coefficients reflect the shape, not the scale, of the data.

For calculating sample Skewness (γ_1), we rely on the third standardized moment ($k=3$), which sums the cubed deviations from the mean:

$$\gamma_1 = M_3 = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{s^3}$$

For calculating sample Kurtosis (γ_2), we rely on the fourth standardized moment ($k=4$), which sums the deviations raised to the fourth power:

$$\gamma_2 = M_4 = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{s^4}$$

It is important to note that many software packages apply complex adjustments to these fundamental formulas to correct for bias, particularly when working with small sample sizes, leading to what are often called "unbiased estimators." When interpreting statistical outputs, verifying the precise formula used by the software is always recommended to ensure correct assessment, particularly in distinguishing between population and sample estimates.

Formal Reporting Standards for Statistical Results

When presenting the findings of skewness and kurtosis in academic papers, research reports, or formal documentation, adherence to established reporting standards is necessary to maintain clarity and professionalism. The coefficients themselves are dimensionless, but their interpretation regarding the shape of the distribution must be explicitly stated.

The standard practice involves two critical steps for presentation: precision and formatting. First, results should generally be reported to **two decimal places**, which is standard practice across most quantitative fields for descriptive statistics. Secondly, conventional statistical reporting often dictates dropping the leading zero for values between -1 and 1 (e.g., using .79 instead of 0.79), although this rule can vary slightly based on specific journal or institutional guidelines. This streamlined approach ensures the report is concise and focuses on the significant digits.

The descriptive narrative accompanying the numerical results must clearly state what the coefficient implies about the distribution's shape. The following structure is widely accepted when formally documenting these measures, ensuring clear communication of both the quantitative measure and its qualitative interpretation:

The skewness of was found to be -.89, indicating that the distribution was left-skewed.

The kurtosis of was found to be 4.26, indicating that the distribution was more heavy-tailed compared to the Normal distribution.

Guidelines for Formatting Numerical Outputs

To ensure consistency and clarity in professional statistical reports, specific formatting rules must be followed when documenting the coefficients for skewness and kurtosis. These rules primarily concern precision and the presentation of the leading digit.

Keep in mind the following crucial rules when reporting the calculated values:

Precision: Round the values for skewness and kurtosis to **two decimal places**. This level of precision is typically sufficient for descriptive statistical reporting.

Leading Zero: Drop the leading 0 when reporting values between -1 and 1 (e.g., use **.79**, not 0.79). Ensure the negative sign is retained if the value is negative (e.g., **-.35**).

These conventions contribute significantly to the overall professionalism and ease of reading in statistical documentation, minimizing clutter and aligning the report with standard academic publishing practices.

Example Application: Reporting Exam Scores Distribution

Suppose we are analyzing the distribution of exam scores among students at a certain university. Using standard statistical software, we calculate the values for the skewness and kurtosis of the score distribution. This preliminary analysis is necessary to determine if subsequent parametric tests (like t-tests) are appropriate.

The raw calculated values provided by the software are:

Skewness Coefficient: **-1.391777**

Kurtosis Coefficient (Excess Kurtosis): **4.170865**

We must now translate these raw figures into the required formal reporting format by applying rounding and interpretation.

First, we round the values to two decimal places, resulting in a skewness of -1.39 and a kurtosis of 4.17. Since both absolute values are greater than one, the leading digits are maintained. Based on these values, we determine the distributional shape: the negative skewness suggests left-skewing, and the positive kurtosis (significantly greater than 0) suggests a leptokurtic, heavy-tailed distribution.

We would formally report these values as follows:

The skewness of the exam scores was found to be -1.39, indicating that the distribution was **significantly left-skewed**. This suggests that the majority of scores were concentrated at the upper end of the scale.

The kurtosis of the exam scores was found to be 4.17, indicating that the distribution was **leptokurtic**, possessing heavy tails and a higher propensity for extreme values compared to the Normal distribution.

This detailed analysis confirms that the data significantly deviates from normality in both symmetry and tailedness. For subsequent inferential statistical analysis, researchers would need to apply appropriate transformations or switch to non-parametric tests to ensure the validity of their conclusions.

Further Statistical Resources

Understanding the calculation and reporting of distribution shape metrics often requires practical application within specific statistical software environments. The methods for obtaining skewness and kurtosis coefficients vary slightly across different platforms, such as R, Python, SPSS, or specialized statistical packages.

The following tutorials explain how to calculate skewness and kurtosis in different statistical software:

The following tutorials explain how to report other statistical results: