

A Comprehensive Guide to Skewness and Kurtosis Calculations in SAS for Statistical Analysis

Authored by
Mohammed looti

November 14, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *A Comprehensive Guide to Skewness and Kurtosis Calculations in SAS for Statistical Analysis*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=1414>

In the realm of [statistics](#) and advanced data analysis, acquiring a deep understanding of the inherent characteristics of a dataset's [distribution](#) is paramount for generating robust and reliable insights. Beyond the basic measures of central tendency (like the mean) and measures of variability (like standard deviation), analysts must assess the shape of the data. Two fundamental metrics used to quantify this shape are [skewness](#) and [kurtosis](#). These metrics are not mere academic additions; they provide critical diagnostic insights that determine whether a dataset conforms to the underlying assumptions required by many [parametric statistical models](#), ultimately guiding the selection of appropriate analytical techniques.

The Importance of Measuring Distribution Shape in Statistical Analysis

Before commencing any serious modeling or inferential testing, statisticians are obligated to evaluate whether their data is approximately normally distributed. Significant deviations from perfect symmetry or ideal tail behavior can severely compromise the validity of subsequent analyses. **Skewness** is the metric dedicated to measuring the asymmetry of the data, indicating the direction in which the bulk of the observations are concentrated relative to the mean. Conversely, [kurtosis](#) focuses on the distribution's "tailedness," signaling the likelihood and extremity of [outliers](#) when compared to a standard normal curve.

Accurate assessment of these distributional properties is crucial because many standard inferential tests, including the ubiquitous t-tests and Analysis of Variance (ANOVA), rely heavily on the assumption of an underlying [normal distribution](#). If a dataset exhibits substantial [skewness](#) or high [kurtosis](#), these assumptions are violated, which can lead to inflated Type I error rates, inaccurate p-values, and potentially misleading conclusions. Therefore, quantifying these parameters during the exploratory data analysis phase is a prerequisite for making informed decisions, such as applying specific data transformation techniques (e.g., logarithmic or square root) to achieve symmetry or pivoting to non-parametric methods entirely.

Within the robust [SAS](#) statistical software environment, calculating these descriptive statistics is remarkably straightforward and highly efficient. The primary tool for this task is the powerful [PROC MEANS](#) procedure, which is designed to handle large volumes of data quickly. This comprehensive article will guide you through leveraging **SAS** to rapidly compute and accurately interpret both **skewness** and **kurtosis** for numeric variables within any given dataset, ensuring your data preparation is methodologically sound.

Understanding Skewness: Measuring Asymmetry

[Skewness](#) serves as a critical measure that precisely quantifies the degree to which a distribution deviates from perfect symmetry. It essentially reflects the relative length and weight of the distribution's tails. For instance, a perfectly symmetrical distribution, such as the ideal bell curve of

the standard [normal distribution](#), will mathematically yield a **skewness** value of exactly zero. Any deviation from zero indicates that the data is concentrated more heavily on one side of the mean than the other.

The interpretation of the **skewness** coefficient provides clear guidance on the shape of the data:

A **negative skew** (often referred to as left skew) signifies that the elongated tail of the distribution extends towards the left side (more negative values). This pattern suggests that the majority of the data mass and the mode are concentrated on the right side of the mean, while a few low values are dragging the average down.

A **positive skew** (or right skew) occurs when the tail is stretched out towards the right side (more positive values) of the distribution. In this common scenario, the bulk of the data observations are clustered toward the lower values, and a few unusually high values are pulling the mean toward the right tail.

A value approaching zero indicates that the distribution is highly symmetrical, implying minimal **skewness**. While absolute zero is rare in empirical data, a small value suggests that the assumption of symmetry is reasonably met.

Understanding both the direction and magnitude of [skewness](#) is vital. It helps analysts identify potential underlying issues, such as ceiling or floor effects in data collection, or the presence of non-linear relationships that require advanced transformation techniques to normalize the data before applying powerful [parametric statistical models](#).

Understanding Kurtosis: Analyzing Tail Extremity

While [skewness](#) addresses horizontal symmetry, [kurtosis](#) provides insight into the vertical distribution characteristics--specifically, the shape of the tails and the peak (peakedness or flatness) relative to a benchmark distribution, typically the standard [normal distribution](#). This measure is crucial because it helps determine whether the data generates more or fewer extreme [outliers](#) than would be expected under the assumption of normality.

It is important to note that **SAS** and most contemporary statistical software packages calculate what is known as **excess kurtosis**. In this standard calculation, the kurtosis of a perfect normal distribution is standardized to 0, making the interpretations significantly more intuitive and comparative:

A **kurtosis** value of 0 indicates a **mesokurtic** distribution. This distribution possesses tails and a peak structure that are statistically similar to those of the **normal distribution**.

If a distribution yields a **kurtosis** coefficient less than 0 (i.e., negative), it is classified as [platykurtic](#). Platykurtic distributions are characterized by lighter tails and often a flatter peak. Critically, this suggests that the distribution tends to produce fewer and less extreme [outliers](#)

compared to the normal curve.

If a distribution has a **kurtosis** coefficient greater than 0 (i.e., positive), it is classified as **leptokurtic**. Leptokurtic distributions are marked by heavier tails and often a higher, sharper peak. This signifies a greater concentration of probability in the tails, meaning the distribution is prone to generating more extreme **outliers** than the normal distribution.

Understanding **kurtosis** is particularly vital in fields like finance and risk management. High positive **kurtosis** (leptokurtosis) serves as a potent warning sign, as it statistically indicates a significantly greater probability of observing rare, extreme events (often called "black swans") that lie far away from the expected mean.

Implementing Skewness and Kurtosis Calculation using SAS PROC MEANS

In the **SAS** programming environment, the most efficient, direct, and widely used method for calculating both **skewness** and **kurtosis** simultaneously for numerous variables is through the versatile **PROC MEANS** procedure. While this procedure is most commonly associated with basic descriptive statistics like counts, means, and standard deviations, its functionality can be easily extended to include shape measures using specific keywords.

To instruct **PROC MEANS** to compute these critical shape measures, the analyst must simply include the **SKEWNESS** and **KURTOSIS** keywords directly within the PROC statement. By default, **SAS** is designed to automatically calculate these values for all numeric variables present in the specified input dataset. For cases where only a subset of variables is relevant for analysis, a ``VAR`` statement can be optionally utilized to restrict the calculation scope; however, omitting the ``VAR`` statement ensures a comprehensive initial screening of all numeric fields.

This streamlined implementation methodology is highly beneficial for initial data screening, providing a quick and standardized output table that summarizes the distributional characteristics of the variables. The following practical example will clearly illustrate the minimal syntax required to implement these statements within a standard **SAS** program, setting the stage for a detailed analysis of a sample basketball performance dataset. This approach emphasizes efficiency and accuracy in assessing data distribution.

Case Study: Calculating Distribution Shape for Basketball Player Data

To provide a tangible demonstration of **PROC MEANS** in action, we will analyze a synthetic dataset containing performance metrics for a group of basketball players across different teams. This dataset, named `my_data`, includes categorical variables (`team`) and two key numeric performance variables: `points` scored and `assists` recorded.

The first step in any **SAS** analysis involves creating and verifying the integrity of the sample data.

We execute a standard **SAS** data step to input the observations, followed by a proc print procedure to visually inspect the resulting table and ensure the data structure is correct before proceeding to the statistical calculations:

```
/*create dataset*/  
data my_data;  
input team $ points assists;  
datalines;  
A 10 2  
A 17 5  
A 17 6  
A 18 3  
A 15 0  
B 10 2  
B 14 5  
B 13 4  
B 29 0  
B 25 2  
C 12 1  
C 30 1  
C 34 3  
C 12 4  
C 11 7  
;  
run;  
  
/*view dataset*/  
proc print data=my_data;
```

Obs	team	points	assists
1	A	10	2
2	A	17	5
3	A	17	6
4	A	18	3
5	A	15	0
6	B	10	2
7	B	14	5
8	B	13	4
9	B	29	0
10	B	25	2
11	C	12	1
12	C	30	1
13	C	34	3
14	C	12	4
15	C	11	7

Once the dataset integrity is confirmed, we proceed directly to the statistical calculation phase. We execute **PROC MEANS**, appending the essential **SKEWNESS** and **KURTOSIS** keywords to the procedure call. This generates a concise output table that summarizes the shape metrics for all numeric variables within the dataset--in this case, `points` and `assists`:

```
/*calculate skewness and kurtosis for each numeric variable*/
proc means data=my_data SKEWNESS KURTOSIS;
run;
```

The MEANS Procedure

Variable	Skewness	Kurtosis
points	1.0093179	-0.2991564
assists	0.3035599	-0.7824519

Interpreting SAS Output for Skewness and Kurtosis

The resulting summary table generated by **PROC MEANS** provides the calculated **skewness** and

kurtosis coefficients for both the `points` and `assists` variables. Analyzing these numerical values allows the analyst to objectively assess the distributional shape of each performance metric:

(1) Analysis of the Points Variable

The `points` variable exhibits a **skewness** coefficient of **1.009**. Since this coefficient is positive and substantially greater than zero, we confidently conclude that the distribution of points is positively skewed (right-tailed). This indicates that the vast majority of basketball players scored a moderate to low number of points, but a small subset of high-performing players achieved exceptionally high scores, which effectively pulls the distribution's tail toward the right.

The `points` variable has an excess **kurtosis** value of **-0.299**. As this value is negative, the distribution is classified as **platykurtic**. This means the distribution is slightly flatter than the standard **normal distribution**, suggesting that the data tends to produce marginally fewer and less extreme **outliers** than a theoretical normal curve.

(2) Analysis of the Assists Variable

The `assists` variable shows a **skewness** of **0.304**. Similar to the points data, this positive value indicates the presence of a slight positive skew. Although less severe than the skew in the points variable, it suggests that most players recorded a lower number of assists, with the distribution's tail extending modestly toward higher assist counts due to a few players with excellent passing skills.

The `assists` variable possesses a **kurtosis** of **-0.782**. This value is also negative and has a larger absolute magnitude than the kurtosis for points. This confirms the distribution is distinctly **platykurtic**. The data is characterized by tails that are notably lighter than those of a normal distribution, implying that extreme assist records are relatively rare within this sample.

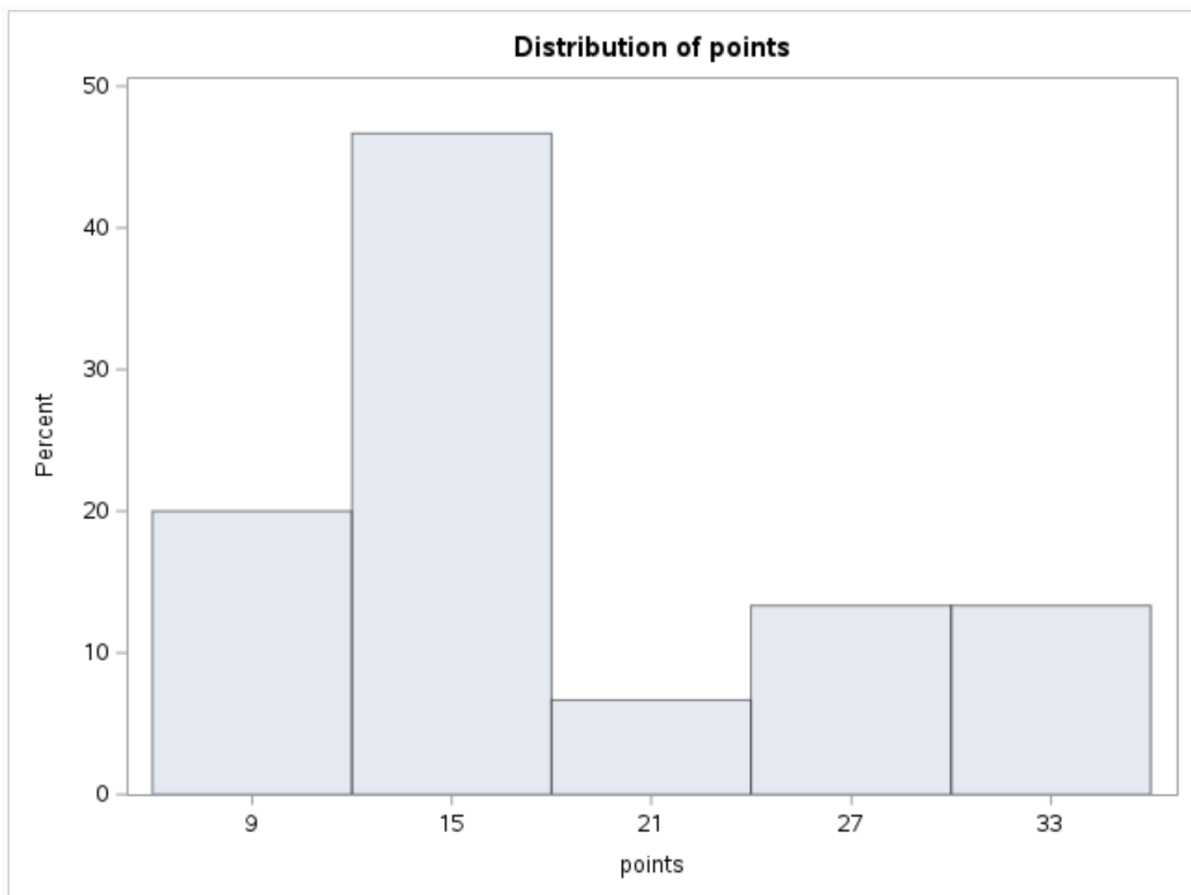
Visualizing Distributions with SAS PROC UNIVARIATE

While numerical coefficients offer precision in measuring **skewness** and **kurtosis**, visualization remains an indispensable tool for confirming and intuitively understanding these interpretations. The human visual system is highly effective at detecting asymmetry and assessing tail weight when data is presented in a graphical format, such as a histogram. In the **SAS** system, the **PROC UNIVARIATE** procedure is specifically designed for detailed univariate analysis and is the ideal tool for generating high-quality graphical representations.

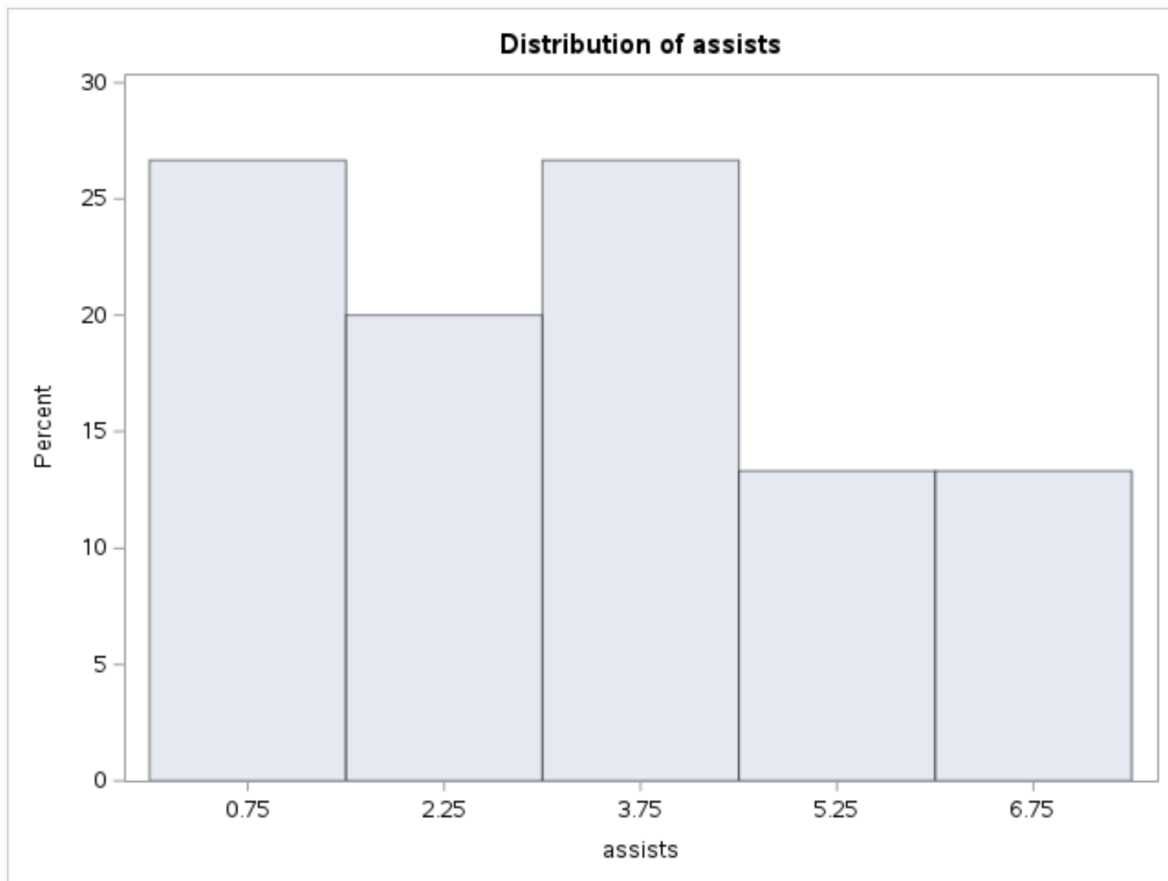
To visually assess the distributions of our performance variables, we execute **PROC UNIVARIATE** and explicitly request histograms for both the `points` and `assists` variables. The resulting graphical output will serve to confirm the numerical conclusions derived from **PROC MEANS**, providing a complete picture of the data shape:

```
/*create histograms for points and assists variables*/  
proc univariate data=my_data;  
var points assists;  
histogram points assists;  
run;
```

The first output image below visually confirms the strong positive skew (tail extending right) that was numerically observed for the **points** variable:



The second output provides a clear visual confirmation of the mild positive skew and the relatively short, light tails (platykurtic nature) for the **assists** variable, reinforcing the conclusion that extreme values are infrequent:



By meticulously combining the precise numerical output from [PROC MEANS](#) with the robust visual confirmation provided by [PROC UNIVARIATE](#), analysts achieve a comprehensive and reliable understanding of their data's distributional properties. This dual-approach is essential for making informed methodological decisions regarding subsequent statistical modeling and hypothesis testing.

Additional Resources for SAS Procedures

To further enhance your skills in **SAS** data exploration and the generation of descriptive statistics, consider reviewing tutorials that cover other common data analysis tasks and procedures. Mastering these fundamental procedures is key to becoming proficient in statistical programming using [SAS](#).