

# Learning Guide: Calculating Mean, Median, and Mode with SAS

Authored by  
**Mohammed loot**

October 30, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Learning Guide: Calculating Mean, Median, and Mode with SAS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=6245>

## The Foundation of Data Insight: Understanding Central Tendency with SAS

In the rigorous domain of [data analysis](#), mastering the methods to accurately summarize and characterize the fundamental properties of a [dataset](#) is absolutely essential. Measures of [central tendency](#) represent the core statistical metrics that condense a distribution into a single, representative value, effectively describing the midpoint or typical value of the data. These statistics are indispensable for initial data exploration, enabling analysts to quickly identify underlying patterns and establish a robust foundation for subsequent quantitative decision-making.

The trifecta of fundamental measures of central location includes the [mean](#), the [median](#), and the [mode](#). Each metric offers a unique perspective on where the data points cluster within the distribution. The [mean](#), commonly recognized as the arithmetic average, is computed by dividing the sum of all observations by the total count of values. Conversely, the [median](#) is the absolute middle value in a numerically ordered dataset, ensuring that 50% of the observations fall above and 50% fall below it. Finally, the [mode](#) identifies the value or category that occurs with the greatest frequency. Understanding the relationship among these three measures is critical for assessing the shape, symmetry, and potential skewness of your data distribution.

For statistical researchers and analysts who rely on [SAS](#), calculating these vital descriptive statistics is both remarkably efficient and highly precise. The [PROC UNIVARIATE](#) procedure is a cornerstone tool within the [SAS](#) system, specifically engineered to furnish comprehensive descriptive statistics, inherently including the mean, median, and mode, for any designated numeric [variables](#). This detailed guide will walk you through the practical steps required to leverage this powerful procedure to extract meaningful statistical insights from your raw data.

## The Power Tool: Utilizing PROC UNIVARIATE for Descriptive Statistics

The [PROC UNIVARIATE](#) procedure transcends a mere calculation of [central tendency](#); it serves as a cornerstone of descriptive statistics in [SAS](#), offering a holistic statistical summary. Its robust output includes quantiles, moments (such as variance and standard deviation), and sophisticated tests for distributional assumptions. This comprehensive output makes it the definitive choice for performing a thorough initial exploration of a variable's characteristics and distribution. Analysts can quickly gain a deep, multi-faceted view of their numeric [variables](#) with minimal coding overhead.

The foundational syntax for executing [PROC UNIVARIATE](#) is elegantly straightforward: you simply state the procedure name followed by the mandatory `data=` option to specify the input [dataset](#) you wish to analyze. A crucial feature of this procedure is its default behavior: when no specific [variables](#) are explicitly designated for analysis, [PROC UNIVARIATE](#) intelligently processes all numeric variables found within the specified dataset. This automatic analysis capability is highly

advantageous for generating a swift, overarching statistical summary.

```
proc univariate data=my_data;  
run;
```

This concise [SAS](#) code block is sufficient to initiate a complete univariate analysis. The `data=my_data;` clause directs the procedure to focus its computational resources on the dataset titled `my_data`. The subsequent `run;` statement signals to the [SAS](#) system that the block is complete and execution should commence, displaying the resulting statistical report. This streamlined approach highlights the efficiency inherent in [SAS](#) for foundational data exploration tasks, which we will now demonstrate with a practical example.

## Structuring the Input: A Hands-On Data Preparation Example

To provide a clear, functional demonstration of [PROC UNIVARIATE](#), we must first construct a sample [dataset](#). This example will simulate a small collection of sports performance statistics, featuring key numeric [variables](#) such as `points` scored, `rebounds` collected, and `assists` made, alongside a character identifier for each team. Creating this structured dataset within [SAS](#) is the vital initial step, guaranteeing that our data is correctly formatted and ready for sophisticated statistical operations.

The data creation process in [SAS](#) commences with a [DATA step](#), initiated by the `data` statement, which assigns the name `my_data` to our new dataset. Following this, the critical [input statement](#) defines the names and types for our variables: `team` is designated as a character variable (indicated by the `$` suffix), while `points`, `rebounds`, and `assists` are defined as numeric variables. The `datalines` statement then serves as the instruction to [SAS](#) that the raw, line-by-line data input will immediately follow.

```
/*create dataset*/  
data my_data;  
input team $ points rebounds assists;  
datalines;  
A 25 10 8  
B 18 4 5  
C 18 7 10  
D 24 12 4  
E 27 11 5  
F 30 8 7  
G 12 8 5  
;
```

```
run;
```

```
/*view dataset*/
```

```
proc print data=my_data;
```

Once the [DATA step](#) has been successfully executed and the dataset created, the subsequent `proc print` statement is deployed to visualize the contents of the newly formed `my_data` [dataset](#). This verification step is fundamental, as it allows the analyst to confirm visually that the raw data has been correctly imported, variables have been properly defined, and the structure is sound before proceeding with computationally intensive statistical analysis.

Obs	team	points	rebounds	assists
1	A	25	10	8
2	B	18	4	5
3	C	18	7	10
4	D	24	12	4
5	E	27	11	5
6	F	30	8	7
7	G	12	8	5

## Broad Statistical Overview: Calculating All Measures of Central Tendency

With the `my_data` [dataset](#) successfully prepared and validated, we are now poised to conduct a comprehensive statistical analysis. Our primary goal is to calculate the three essential measures of [central tendency](#) - the mean, median, and mode - for every relevant numeric [variable](#) within the dataset. As previously noted, the inherent design of [PROC UNIVARIATE](#) simplifies this task immensely by automatically analyzing all numeric columns when no specific variables are designated, providing an ideal starting point for initial data exploration.

The elegance and efficiency of [SAS](#) procedures are clearly demonstrated by the minimal code required for this comprehensive statistical summary. By simply invoking [PROC UNIVARIATE](#) and specifying our dataset name, we are instructing the system to generate a highly detailed and exhaustive statistical report for each numeric column. This default functionality proves exceptionally useful when dealing with new or large datasets, as it allows for a rapid, all-encompassing assessment of the data's statistical characteristics without the need for repetitive coding.

```
/*calculate mean, median, mode for each variable in my_data*/
```

```
proc univariate data=my_data;
run;
```

Upon execution, this code generates an extensive output, systematically segmented into dedicated tables for each numeric [variable](#) present in the `my_data` [dataset](#). Each section of the report will feature a summary table densely populated with descriptive statistics. Crucially, this includes the mean, median, and mode, alongside other important measures such as standard deviation, variance, quartiles, and extreme values. This comprehensive statistical portrait provides a profound understanding of the underlying distributions, laying a solid foundation for more complex modeling and hypothesis testing.

## Interpreting Distribution Shape: Deciphering Mean, Median, and Mode

The output produced by [PROC UNIVARIATE](#) is highly structured, presenting a wealth of statistical information in a clear, digestible format. To derive meaningful insights from our sports team data, it is imperative to understand how to interpret the key measures of [central tendency](#) for each variable: Points, Rebounds, and Assists. The comparison between these three measures often reveals critical information about the data's distribution shape.

### 1. Analysis for the Points Variable

The UNIVARIATE Procedure			
Variable: points			
Moments			
N	7	Sum Weights	7
Mean	22	Sum Observations	154
Std Deviation	6.244998	Variance	39
Skewness	-0.4368625	Kurtosis	-0.6965155
Uncorrected SS	3622	Corrected SS	234
Coeff Variation	28.3863545	Std Error Mean	2.36038738

  

Basic Statistical Measures			
Location		Variability	
Mean	22.00000	Std Deviation	6.24500
Median	24.00000	Variance	39.00000
Mode	18.00000	Range	18.00000
		Interquartile Range	9.00000

Examining the detailed summary statistics provided for the `Points` variable yields the following crucial metrics:

The **mean** score is calculated as **22**. This value represents the standardized arithmetic average of points scored across all teams included in the dataset.

The **median** points value is **24**. This central value signifies that half of the teams scored 24 points or less, and half scored 24 points or more. The observation that the median is slightly higher than the mean often suggests a modest left-skew in the distribution.

The **mode** points value is **18**. This is the single score that occurred most frequently in our sample. When the mean, median, and mode diverge significantly, it provides strong evidence of a non-normal distribution, guiding the analyst toward specific data transformation or modeling choices.

## 2. Analysis for the Rebounds Variable

**The UNIVARIATE Procedure**  
Variable: rebounds

Moments			
<b>N</b>	7	<b>Sum Weights</b>	7
<b>Mean</b>	8.57142857	<b>Sum Observations</b>	60
<b>Std Deviation</b>	2.69920623	<b>Variance</b>	7.28571429
<b>Skewness</b>	-0.5012367	<b>Kurtosis</b>	0.05221069
<b>Uncorrected SS</b>	558	<b>Corrected SS</b>	43.7142857
<b>Coeff Variation</b>	31.4907394	<b>Std Error Mean</b>	1.02020406

  

Basic Statistical Measures			
Location		Variability	
<b>Mean</b>	8.571429	<b>Std Deviation</b>	2.69921
<b>Median</b>	8.000000	<b>Variance</b>	7.28571
<b>Mode</b>	8.000000	<b>Range</b>	8.00000
		<b>Interquartile Range</b>	4.00000

A dedicated review of the statistics for the `Rebounds` variable offers further insight into team performance dynamics:

The **mean** rebounds value is calculated to be approximately **8.57**, establishing the average number of rebounds achieved per team.

The **median** rebounds value is precisely **8**, indicating the exact center point of the ordered rebound counts.

The **mode** rebounds value is also **8**.

In this specific case, the median and mode are identical, and the mean is extremely close to these values. This strong convergence among the three measures of **central tendency** is a key indicator of a nearly symmetrical distribution for the rebound statistics within our sample, suggesting that the

data points are distributed quite evenly around the central axis.

### 3. Analysis for the Assists Variable

The UNIVARIATE Procedure			
Variable: assists			
Moments			
N	7	Sum Weights	7
Mean	6.28571429	Sum Observations	44
Std Deviation	2.13808994	Variance	4.57142857
Skewness	0.91495216	Kurtosis	-0.1585938
Uncorrected SS	304	Corrected SS	27.4285714
Coeff Variation	34.0150672	Std Error Mean	0.80812204

  

Basic Statistical Measures			
Location		Variability	
Mean	6.285714	Std Deviation	2.13809
Median	5.000000	Variance	4.57143
Mode	5.000000	Range	6.00000
		Interquartile Range	3.00000

The calculated statistics for the `assists` variable provide a final perspective on the data's central distribution:

The **mean** assists value is approximately **6.28**, representing the average playmaking performance.

The **median** assists value is **5**, marking the middle observation in the ordered list.

The **mode** assists value is also **5**.

In this scenario, while the median and mode are both 5, the mean is noticeably higher at 6.28. This classic pattern where the mean is greater than both the median and the mode typically signifies a **right-skewed distribution**. This means that a cluster of lower assist scores exists, but a few teams recorded significantly higher assist values, pulling the **mean** toward the upper tail of the distribution. Recognizing these subtle differences is paramount for accurately understanding and modeling the data.

### Targeted Precision: Limiting Analysis with the VAR Statement

While executing a comprehensive analysis across all numeric **variables** within a **dataset** is excellent for initial discovery, many analytical objectives demand a more focused and streamlined approach. Data professionals frequently need to concentrate their statistical review on a single

variable or a select subset of variables, ensuring that the resulting output is concise and directly addresses the primary research questions without the distraction of extraneous statistics.

To achieve this high level of targeted analysis within **PROC UNIVARIATE**, analysts employ the powerful [VAR statement](#). The [VAR statement](#) enables the explicit listing of only those variables for which descriptive statistics are desired. This capability is exceptionally valuable when working with vast datasets containing numerous unrelated variables, where a full univariate report would prove unnecessarily cumbersome and diminish focus on the primary analytical targets.

For instance, if the specific mandate is to exclusively examine the distribution and [central tendency](#) of the `points` [variable](#), we can effortlessly modify the **PROC UNIVARIATE** code to incorporate the [VAR statement](#), thereby limiting the scope of the statistical computation to only that variable of interest.

```
/*calculate mean, median, and mode only for points variable*/  
proc univariate data=my_data;  
var points;  
run;
```

The execution of this refined code produces output solely dedicated to the `points` [variable](#), providing its mean, median, mode, and other related descriptive statistics in a highly focused report. This method is strongly recommended for optimizing analytical concentration, ensuring maximum clarity in statistical reporting, and directly addressing specific research queries.

**Note:** For an exhaustive understanding of all available options, report customizations, and advanced capabilities, analysts are strongly encouraged to consult the official documentation for **PROC UNIVARIATE**, which is maintained on the official [SAS](#) website.

## Conclusion: Mastering Descriptive Statistics in SAS

The capability to accurately calculate and critically interpret measures of [central tendency](#) - specifically the mean, median, and mode - constitutes a fundamental competency for every professional engaged in [data analysis](#). These foundational statistics offer immediate and crucial insights into the typical values and the overall shape of your data's distribution. They serve as a vital initial step, essential for identifying preliminary patterns, spotting potential outliers, and assessing the quality and characteristics of a [dataset](#) before proceeding to more intricate statistical modeling.

As thoroughly demonstrated throughout this guide, [SAS](#) significantly simplifies and accelerates this crucial analytical process through its highly robust and intuitive **PROC UNIVARIATE** procedure. Whether your analytical goals necessitate a broad, comprehensive overview of all numeric

[variables](#) or require a precise, targeted analysis of a specific subset of variables using the `VAR` statement, **PROC UNIVARIATE** remains an efficient, reliable, and exceptionally versatile solution for fulfilling all descriptive statistical needs within the [SAS](#) environment.

By mastering the practical application of this foundational procedure, you empower yourself to rapidly extract valuable, actionable insights from your data, thereby establishing a solid analytical bedrock for conducting advanced statistical investigations and supporting robust data-driven decision-making processes. We strongly encourage readers to explore the extensive documentation and myriad of options available within [PROC UNIVARIATE](#) to further enhance their statistical analysis proficiency in [SAS](#).

## Additional Resources

The following tutorials explain how to perform other common tasks in SAS: