Advantages & Disadvantages of Using Median in Statistics

Authored by

Mohammed looti

November 10, 2025

Defining the Median: A Positional Measure of Central Tendency

In statistical analysis, the goal is often to find a single representative value that describes the center of a given dataset. The median stands out among measures of central tendency, primarily because it describes the central location based on position rather than magnitude. Unlike the arithmetic mean, which sums all values and divides by the count--thereby calculating the average magnitude--the median identifies the exact halfway point. This fundamental distinction is vital for researchers working with real-world data, which frequently exhibits significant anomalies, extreme values, or inherent asymmetry.

The calculation of the median demands a methodical approach: the dataset must first be rigorously ordered, typically from the smallest observation to the largest. Once sorted, the median is the value that perfectly bisects the data, ensuring that 50% of the observations lie below it and 50% lie above it. If the total number of observations is odd, the median is simply the single middle number. However, if the count is even, convention dictates calculating the average of the two central values to establish the precise midpoint.

Selecting the most appropriate measure of central tendency--be it the mean, median, or mode--is a critical step in achieving accurate statistical reporting. This choice is intrinsically linked to the characteristics of the data distribution, specifically its level of symmetry and the existence of extreme values. While the mean provides a holistic view by factoring in the numerical weight of every data point, the median frequently offers a more reliable and truthful depiction of the **typical value**. This makes the median indispensable in fields like socioeconomics and demographics, where metrics such as household income or property valuations are commonly analyzed and are often subject to highly irregular distributions.

Primary Advantages: Robustness Against Outliers and Skewness

The median holds a significant advantage over the mean in specific analytical contexts, rooted in its identity as a positional statistic. Its value is determined solely by its rank within the ordered data, rather than the absolute numerical magnitude of every single observation. This inherent positional characteristic yields two principal benefits that often elevate the median to the preferred measure of center.

The first compelling benefit is the median's exceptional resilience to outliers. An outlier is defined as an observation whose magnitude deviates substantially from the majority of the data points. When the arithmetic mean is calculated, these extreme values exert an outsized, gravitational pull on the average, potentially distorting the overall representation. Because the median only requires locating the middle rank, the actual numerical size of the smallest or largest observations has virtually no impact on the final calculated median value, provided the middle value itself does not change position. This characteristic solidifies the median's reputation as a highly **robust statistic**--

a measure providing inherent stability and enhanced reliability when dealing with datasets that are susceptible to measurement errors, recording mistakes, or natural but extreme variations.

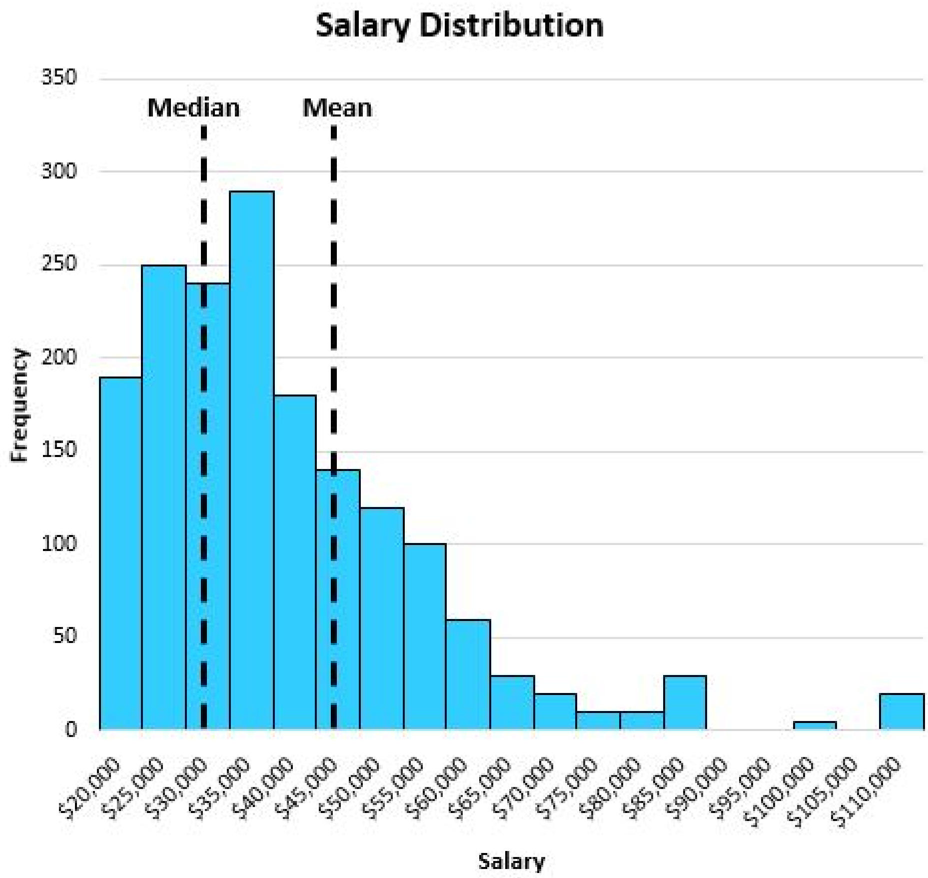
Furthermore, the median is the ideal measure of central tendency for datasets characterized by skewed distributions. A distribution is considered skewed when its data points are not symmetrically arranged around the center, resulting in a pronounced tail extending to one side. In these asymmetrical scenarios, the mean is mathematically dragged toward the elongated tail, consequently misrepresenting the true center of the bulk of the data. Take, for example, typical salary data, which is often right-skewed due to a small number of extremely highly compensated individuals. Here, the mean salary would be artificially inflated. In contrast, the median accurately identifies the point where half the population falls above and half falls below, offering a much more accurate representation of the 'typical' earning value within that population, irrespective of the distribution's asymmetry.

Illustrative Example: Analyzing Skewed Income Distributions

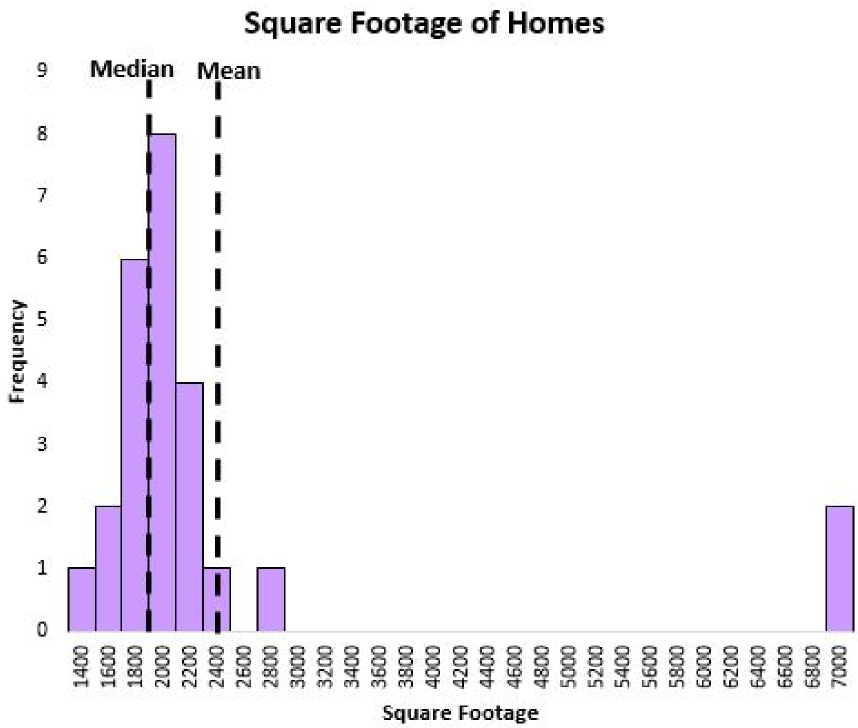
To fully appreciate the median's advantage, we can examine a common scenario: the distribution of annual employee salaries within a corporate setting. This financial data inherently tends to be right-skewed, characterized by a dense concentration of employees earning moderate salaries and a sparse tail of highly paid executives or specialists earning disproportionately higher incomes. When the objective is to communicate the "typical" or **central earning power**, the distinction between the mean and the median becomes a matter of ethical and analytical integrity.

Imagine a situation where calculations reveal a mean salary of approximately $47,000, yet the corresponding median salary is only $32,000. In this instance, the median value of $32,000 provides a far more honest representation of what the majority of employees actually earn. The mean is misleadingly high, having been inflated by the extreme financial magnitudes of the few high earners located in the distribution's right tail. This case study confirms that the median resists the pull of these wealthy outliers, maintaining its integrity as a measure of the middle position.

The following visual chart provides a clear demonstration of this principle, showing how the arithmetic average (mean) shifts significantly due to the presence of higher values, while the median remains anchored close to the true center of the majority of observations. This utility extends beyond income data to other metrics that frequently contain outliers, such as real-estate valuations or the square footage of residential properties, where a single mansion can drastically skew the calculated average for an entire neighborhood.



The subsequent illustration further reinforces this concept, demonstrating how even a small number of extraordinarily large properties can dramatically influence the mean square footage. By comparison, the median remains stable, reflecting the size of the typical house on that street. This capability to ignore the disproportionate influence of extremes is the defining feature that establishes the median as the superior metric for descriptive statistics whenever anomalies are present.



Inherent Limitations: Sacrificing Efficiency and Algebraic Utility

While the median excels in providing stability against extremes, it is essential to acknowledge its inherent drawbacks. These limitations often compel statisticians to revert to the mean or employ more complex statistical techniques. The core issue lies in the median's focus purely on position, which requires sacrificing valuable numerical information regarding the overall magnitude and spread of the dataset.

A primary limitation of the median is its inherent lack of statistical efficiency, stemming from the fact that it does not incorporate the numerical values of every observation. Although the entire dataset must be ordered, the specific magnitudes of the extreme high and low values--those far from the center--are effectively disregarded once the middle position is identified. For rigorous statistical analysis, the conventional preference is to utilize all available data information to construct the most comprehensive summary possible. By ignoring the magnitude contained in the periphery, the median loses discriminatory power; two datasets with radically different ranges and spreads can easily possess the exact same median value, leading to incomplete or potentially misleading conclusions about data variability.

The second critical disadvantage is the median's lack of algebraic utility, specifically its inability to be used to calculate the aggregate sum of all observations. This structural limitation severely restricts its application in advanced statistical modeling, forecasting, and calculations requiring total values. For instance, if we know the arithmetic mean and the total sample size, determining the total aggregate value is a simple algebraic step: mean multiplied by sample size. This property is fundamental in fields such as finance, economics, and logistics planning. Conversely, no such straightforward algebraic relationship exists for the median; knowing only the median and the sample size is insufficient to reliably infer the total aggregate sum of the underlying values.

Demonstrating Insensitivity to Numerical Magnitude

The limitations of the median are most clearly demonstrated when the values at the extremes of the dataset are altered. This illustrates the core drawback: the median ignores peripheral numerical information. Consider an initial dataset representing exam scores for 13 students, already sorted:

Scores: 68, 70, 71, 75, 78, 82, **83**, 83, 85, 90, 91, 91, 92

In this sequence, the median score is **83**, corresponding to the seventh value. Now, let us hypothesize a scenario where the three lowest scores are significantly worse, perhaps reflecting students who missed a substantial portion of the course material. The updated, ordered dataset is:

Scores: 22, 35, 38, 75, 78, 82, **83**, 83, 85, 90, 91, 91, 92

Despite the drastic reduction in the lower tail (from 68, 70, and 71 down to 22, 35, and 38), the median score in this profoundly changed distribution remains precisely **83**. This experiment confirms that the median is purely a measure of position. It successfully achieves stability against massive changes in the magnitude of peripheral outliers, but at the cost of being completely insensitive to critical shifts in the data's overall magnitude and range.

We also must revisit the algebraic restriction. Consider a business dataset tracking quarterly sales for 11 employees, with the following results:

Sales (in thousands): 12, 12, 15, 19, 22, **24**, 28, 30, 32, 35, 38

If we only know that the median sales figure is 24 and the sample size is 11, we possess insufficient information to calculate the total aggregate sales volume for the company. Contrast this with the mean: if the arithmetic mean were 24, we could effortlessly calculate the total sum of sales as 24 multiplied by 11, equaling 264 (thousand). This ability to derive the aggregate total is why the mean is an indispensable tool for financial budgeting, predictive modeling, and resource allocation where knowing the grand total is often paramount.

Strategic Choice: When to Prioritize the Mean or the Median

The decision regarding whether to employ the mean or the median as the primary measure of central tendency is a strategic one, contingent upon two fundamental criteria: the underlying shape of the data distribution and the specific analytical objectives of the study. Choosing the wrong metric can lead to misinterpretation and flawed conclusions.

The median is definitively the superior metric under specific conditions, primarily when descriptive accuracy is prioritized over algebraic efficiency:

When the data distribution is markedly skewed, whether positively (right-tailed) or negatively (left-tailed), the median provides a non-inflated representation of the center.
When the dataset contains genuine, high-impact outliers--such as extreme wealth figures, catastrophic accident costs, or peak performance metrics--which must not unduly influence the measure of typicality.
If the principal goal is to accurately identify the **typical position** where half the data points reside above and half reside below, rather than calculating the arithmetic center of total magnitude.

Conversely, the arithmetic mean is the preferred and often required choice in statistical computation when:

The data is known to be symmetrically distributed, ideally approximating a normal distribution, ensuring the mean, median, and mode are closely aligned.
The analysis requires subsequent algebraic manipulation, such as finding the aggregate total, performing statistical hypothesis testing, or utilizing inferential models where the magnitude contribution of every observation is crucial.
The researcher must maximize statistical efficiency by leveraging all numerical information embedded within the dataset, including the precise values of the extremes.

As a best practice in rigorous statistical reporting, it is highly recommended to calculate and present both the mean and the median simultaneously. A comparison between these two values offers immediate, powerful insight into the symmetry of the distribution and confirms the extent of outlier influence, providing a far richer and more complete summary of the data than relying on either measure in isolation.

Additional Resources for Comprehensive Statistical Study

For readers seeking to deepen their understanding of descriptive statistics and the relationship between the mean, median, and mode, the following resources offer valuable supplementary information and practical tutorials:

Normal distribution
Central tendency
Sample size