Advantages & Disadvantages of Using Mean in Statistics

Authored by
stats writer

November 18, 2025

The concept of the mean, often referred to as the arithmetic average, is arguably the most foundational measure of central tendency used across all fields of statistics. It provides a single, representative value that summarizes an entire collection of data points, giving analysts and researchers a quick understanding of the typical magnitude within a dataset. Calculating the mean involves a straightforward procedure, making it accessible and universally recognized, which contributes significantly to its popularity in both academic research and practical business analysis.

Understanding how the mean is derived is crucial for appreciating its strengths and weaknesses. The calculation process ensures that every observation contributes equally to the final resulting average. When applied correctly, the mean serves as an excellent benchmark for comparing different groups or tracking changes over time, provided the underlying distribution of the data meets specific assumptions regarding symmetry and the absence of extreme values.

The formula used to calculate the arithmetic mean remains consistent, regardless of whether the data represents population characteristics or sample observations. This clear mathematical definition is central to statistical theory.

The mean is calculated using the following mathematical expression:

Mean = Σxi / n

The components of this formula are defined precisely:

Σ (Sigma): This widely recognized mathematical symbol denotes the operation of summation, indicating that all individual observations must be added together.
xi: Represents the ith individual observation or data point within the collected data series.
n: Represents the total number of observations, measurements, or elements included in the specific dataset under analysis.

Core Advantages of Employing the Arithmetic Mean

While statisticians have several measures of central tendency at their disposal, the arithmetic mean holds a favored position due to several intrinsic advantages relating to its mathematical properties and its intuitive interpretation. These strengths make it the standard choice for data summarization when the data distribution is relatively normal or symmetrical, ensuring the resulting value accurately represents the majority of the data points.

Advantage #1: The mean utilizes every observation in the dataset for its calculation. This comprehensive approach is a cornerstone of statistical rigor. Since the calculation necessitates summing every single data point before dividing by the count, no piece of information is disregarded. In the field of statistics, a measure that incorporates all available data is inherently deemed more robust and informative than measures, such as the median or mode, which rely only on positional rank or frequency. This characteristic ensures the mean is sensitive to subtle shifts across the entire distribution.

This feature is critical because it ensures that the mean is a complete summary statistic, reflecting the total magnitude of the collected values. For instance, if analyzing the performance of investment portfolios, using the mean ensures that high returns and low returns are both factored into the overall performance metric, providing a more balanced view than relying solely on the middle point. Consequently, the mean is mathematically preferred for subsequent advanced statistical analyses, as it integrates fully into concepts like variance and standard deviation.

Advantage #2: The mean is computationally simple, easily interpreted, and mathematically tractable. The calculation involves only basic arithmetic operations--addition and division--which makes it highly accessible. Even complex, large-scale data analyses rely on this simple definition, allowing for rapid computation across vast datasets, whether performed manually or via sophisticated software. Furthermore, the resulting value is highly intuitive; when someone states the mean salary is $60,000, it clearly communicates the average earning level. This ease of interpretation facilitates communication between technical experts and lay audiences, making the mean a powerful tool for reporting and decision-making.

Significant Drawbacks and Limitations of the Mean

Despite its numerous benefits, relying solely on the mean without considering the data distribution can lead to severely misleading conclusions. The very property that makes the mean desirable--the use of every observation--is also the source of its primary weaknesses when dealing with non-standard data distributions. Analysts must be aware of these limitations to select the appropriate measure of central tendency for their specific context.

Disadvantage #1: The mean is highly sensitive to outliers or extreme values. An outlier is a data point that differs significantly from other observations. Because the calculation of the mean involves summation, a single extremely large or small value can disproportionately influence the overall average, dragging the mean away from the true center of the bulk of the data. For instance, in a neighborhood where 99 houses are valued moderately and one mansion is valued extraordinarily high, including the mansion's value in the calculation will inflate the mean house price significantly, making it an unreliable representation of the typical house value on that street.

This vulnerability means that the mean is not considered a robust statistic. If the presence of an outlier cannot be justified (i.e., it is an error or an anomaly that should be excluded), the mean becomes a poor descriptor of central tendency. In such scenarios, the median is often preferred, as it is resistant to extreme values, focusing only on the middle position rather than the total magnitude. This sensitivity necessitates careful data cleaning and validation before reporting the mean as the primary summary measure.

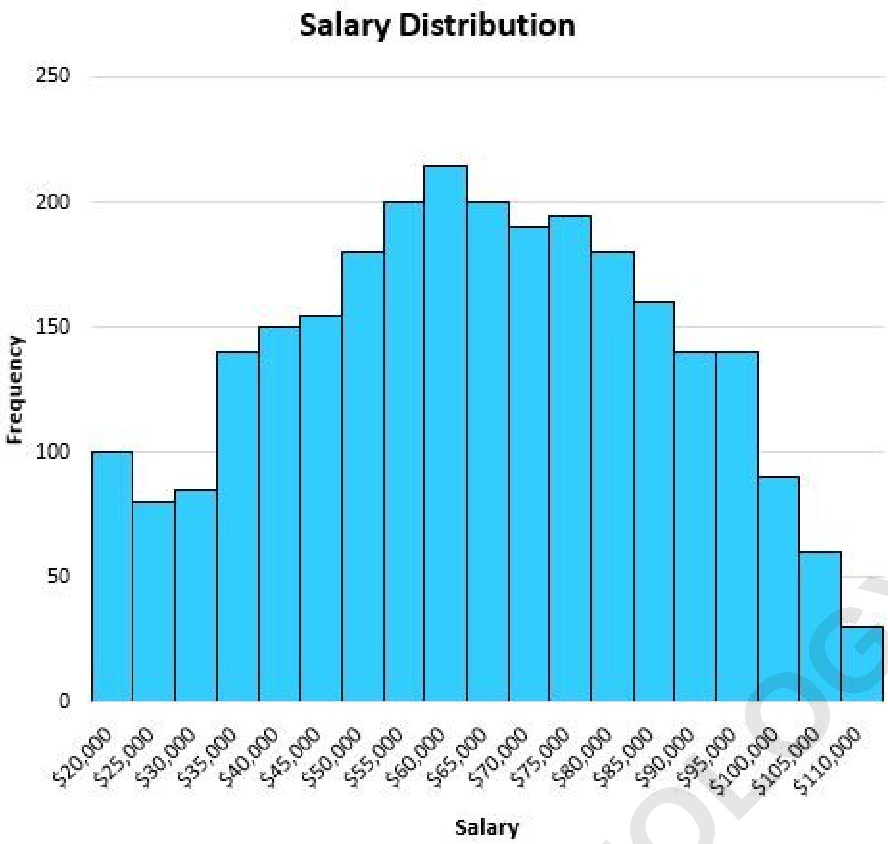
Disadvantage #2: The mean can be misleading when summarizing heavily skewed datasets. Data skewness occurs when data points cluster towards one side of the distribution, resulting in a long tail extending in the opposite direction (either right-skewed or left-skewed). When a distribution is skewed, the mean tends to be pulled in the direction of the tail, making it appear larger or smaller than the majority of the observations. The original text mentioned right skewness, where the mean is pulled towards higher values, failing to represent the experience of the typical individual in the dataset.

In practical terms, if we examine income distribution, which is typically right-skewed (most people earn less, but a small percentage earns extremely high salaries), the arithmetic mean will overestimate the typical income. The long tail created by high earners pulls the average upward, suggesting a higher central value than what the median (the true middle point) would indicate. Therefore, when faced with non-symmetrical, skewed datasets, statisticians often report both the mean and the median to provide a complete picture of the data's center and spread.

Case Study 1: The Reliability of the Mean in Symmetrical Distributions

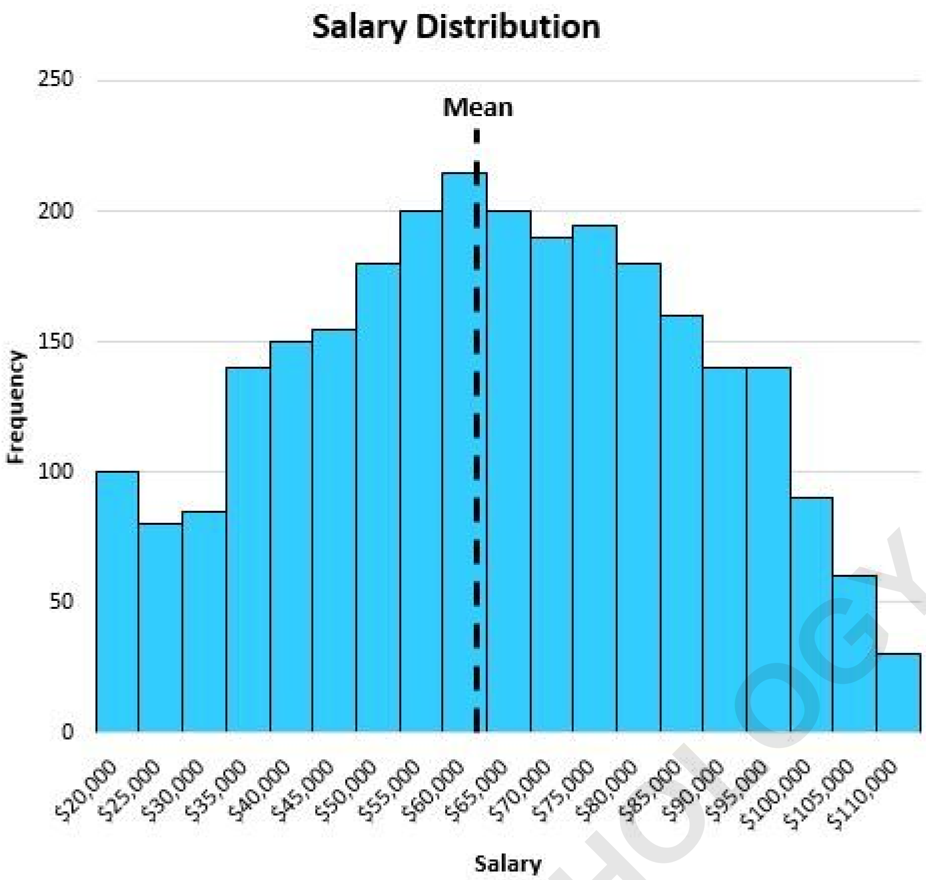
To truly appreciate the strengths of the mean, we must examine its performance within a data environment where its assumptions hold true--specifically, a relatively symmetrical distribution free from significant outliers. Suppose we are analyzing the distribution of hourly wages for non-management employees in a stable manufacturing industry within a specific city. We expect this distribution to be approximately normal, or at least highly symmetrical, because wages are tightly regulated and standardized across the sector.

The following histogram visually represents the salaries of residents in a particular city, demonstrating a classic, bell-shaped distribution:



Because this distribution is highly symmetrical--meaning that if you were to fold the distribution down the middle, the left half would look nearly identical to the right half--and crucially, because there are no extreme salary figures acting as outliers, the mean serves as an extraordinarily effective and reliable measure of central tendency. In this context, the mean, median, and mode will all converge at roughly the same point, affirming the data's well-behaved nature and justifying the use of the mean.

When we calculate the mean for this specific dataset, it is found to be $63,000. As illustrated below, this value is located precisely in the heart of the distribution, confirming that it accurately reflects the typical salary earned by the residents:

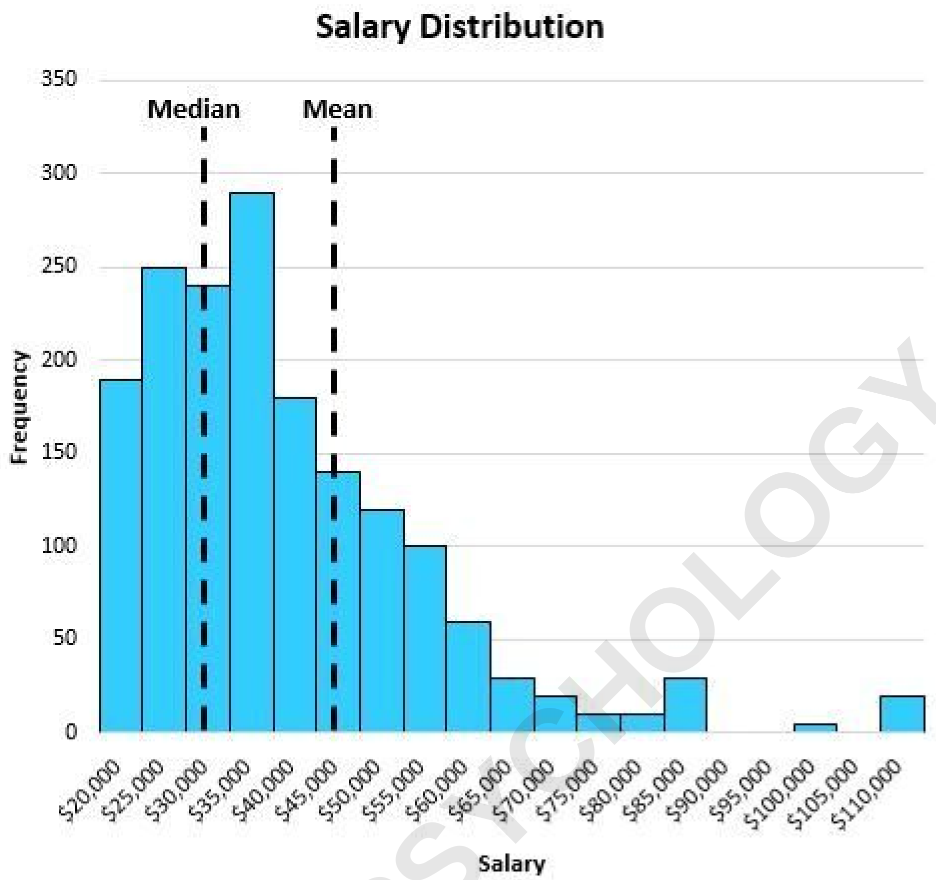


This example perfectly demonstrates the two key advantages discussed earlier. Firstly, by leveraging **Advantage #1: Comprehensive Data Inclusion**, we were able to use every recorded salary to calculate the mean. Since the distribution was symmetrical and lacked extreme values, every data point contributed positively to defining the center. This provided a statistically sound and complete picture of the "average" salary in this specific population. Secondly, **Advantage #2: Computational Simplicity and Interpretability** is highlighted, as the value of $63,000 is straightforward to understand and communicate. It represents the typical earning level, providing a clear benchmark for financial planning and comparison across different demographic groups within the city, confirming the mean's utility when data behaves predictably.

Case Study 2: When the Mean Misrepresents Skewed and Outlier-Affected Data

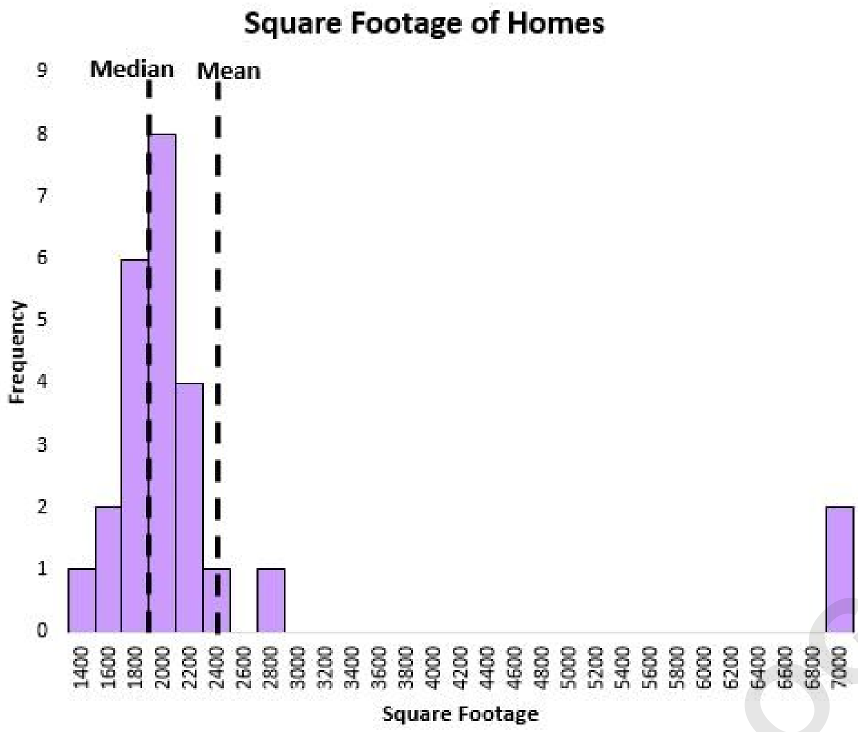
The true test of the mean's robustness comes when analyzing distributions characterized by asymmetry or the presence of extreme data points. Consider a hypothetical scenario involving the annual income distribution in a rapidly growing tech hub. In such an area, the vast majority of workers earn moderate salaries, but a small cluster of highly compensated executives or startup founders pulls the upper end of the income scale dramatically outward. This results in a right-skewed distribution.

When we analyze such a distribution, calculating both the mean and the median salary reveals a significant disparity, as shown in the visualization below:



In this scenario, the higher values extending into the long, right-hand tail of the skewed dataset exert a strong gravitational pull on the mean, shifting it away from the peak where most of the data resides. The calculation reveals that the mean annual income is approximately $47,000. Conversely, the median, which only measures the position of the 50th percentile, is $32,000. For the typical individual in this city, $32,000 is a much more representative income value. This illustrates **Disadvantage #2: Misleading results with skewed datasets**, proving that the mean provides an inflated and unrepresentative summary when asymmetry exists.

Furthermore, let us consider a separate example demonstrating the impact of outliers. Suppose we collect information regarding the square footage of residential homes on a specific suburban street, where most houses are standardized sizes, but two or three newly constructed luxury estates possess exceptionally large footprints.



As clearly visualized, the mean square footage value is substantially influenced by those few extremely large houses, illustrating **Disadvantage #1: Sensitivity to Outliers**. The resulting mean is pulled toward a much higher value than the typical home size, making it a poor measure of the "average" or "typical" square footage on this street. In this case, the median would accurately reflect the size of most houses, remaining unaffected by the anomalous data points. This highlights why understanding the shape of the dataset is paramount before selecting the appropriate measure of central tendency.

Selecting the Right Measure: Mean vs. Median

The decision of whether to use the mean or alternative measures like the median fundamentally depends on the underlying structure of the dataset and the analytical goal. The mean is mathematically superior, utilizing all data points, making it indispensable for advanced statistics that rely on concepts like moments and derivatives. However, this precision comes with the fragility of being easily distorted by external factors such as extreme values or inherent skewness.

In practice, a rule of thumb dictates that if the data distribution is symmetrical or approximately normal, the mean is the preferred measure due to its efficiency and completeness. In these ideal circumstances, the mean is the most accurate estimator of the population center. Conversely, when dealing with financial data, environmental measurements, or any collection of observations known to produce outliers or show signs of asymmetry, the median often provides a more truthful and robust summary of the typical value.

Ultimately, expert data analysts often report both measures. Reporting both the mean and the median allows the audience to instantly gauge the distribution's shape. If the mean and median are close, symmetry is indicated; if they diverge significantly (mean > median indicates positive skew, mean < median indicates negative skew), the analyst knows to exercise caution when interpreting the average and perhaps recommend data transformation or non-parametric statistical methods.

The following tutorials provide additional information about the mean and median in statistics: