

What is a Modified Z-Score? (Definition & Example)

Authored by
Mohammed looti

November 5, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *What is a Modified Z-Score? (Definition & Example)*.
PSYCHOLOGICAL STATISTICS. Retrieved from
<https://statistics.arabpsychology.com/?p=10403>

In the field of [statistics](#), the [Z-Score](#), often referred to as the **standard score**, is a fundamental statistical measure used to quantify the relationship between an individual data point and the mean of a dataset. Essentially, a **Z-Score** tells us how many [standard deviations](#) a specific observation is above or below the population average. This powerful standardization technique enables analysts and researchers to compare observations from disparate datasets, even if those datasets possess radically different means and variances.

The core utility of calculating a standard **Z-Score** lies in providing crucial context regarding the probability of observing a particular value. For example, within a perfectly normal distribution, a Z-Score of 1.5 signifies that the data point is 1.5 standard deviations greater than the mean. The formula utilized for this classical statistical calculation is direct and relies on the population parameters:

$$\mathbf{Z\text{-Score}} = (\mathbf{x_i} - \mu) / \sigma$$

Understanding the variables involved in the classical Z-Score calculation is essential for its correct application across various analyses:

x_i: Represents the single, specific data value or observation being evaluated.

μ: Denotes the population **mean**, which is the arithmetic average of the entire dataset.

σ: Stands for the population **standard deviation**, which quantifies the typical spread or dispersion of the data points around the mean.

Despite its ubiquity and utility, the traditional **Z-Score** suffers from a critical weakness: its reliance on the mean (μ) and standard deviation (σ). Both of these measures are highly susceptible to distortion by **extreme values**, commonly known as [outliers](#). When an outlier is present, it exerts a disproportionate pull on the mean and simultaneously inflates the standard deviation. This sensitivity can lead to two problematic outcomes: masking genuine outliers that are less extreme, or incorrectly labeling typical values as significant deviations. Consequently, using rigid Z-Score thresholds (e.g., $|Z| > 3$) for outlier detection can be unreliable, especially in datasets that are heavily skewed or contaminated by influential points.

Introducing the Modified Z-Score: A Robust Alternative

To address the inherent sensitivity issues of the standard **Z-Score**, statisticians developed methodologies rooted in [robust statistics](#). The **modified Z-score** emerged as a primary solution, specifically engineered to maintain reliability in the presence of **outliers** and when the underlying data distribution deviates significantly from normality. The fundamental innovation of the modified Z-score is the strategic replacement of the sensitive moment-based statistics (mean and standard deviation) with their more robust, positional counterparts: the [median](#) and the Median Absolute

Deviation (MAD).

This modification ensures that the resulting measure is significantly less vulnerable to the distorting influence of extreme values. The median, representing the 50th percentile, is resistant to the magnitude of scores at the extremes of the distribution, providing a much stabler measure of central tendency than the mean. Similarly, the Median Absolute Deviation replaces the standard deviation as a reliable measure of dispersion. The formula for the **modified Z-score** also incorporates a crucial constant, 0.6745, which serves to normalize the measure, ensuring that for data that is perfectly normally distributed, the modified Z-score is consistent with the traditional Z-Score, thus allowing for comparable interpretation.

The formula for calculating this robust measure of deviation is structured as follows:

$$\text{Modified z-score} = 0.6745(x_i - x?) / \text{MAD}$$

Due to its increased robustness, statisticians generally recommend adopting a slightly stricter cutoff criterion for flagging potential [outliers](#) when using the modified Z-score, compared to the standard Z-score. It is widely suggested that observations yielding modified Z-scores less than -3.5 or greater than 3.5 should be flagged as potential anomalies requiring immediate and thorough investigation.

Key Components of the Modified Z-Score Formula

A deep understanding of the core components of the modified Z-score is paramount for its accurate application and meaningful interpretation. By strategically employing positional statistics--measures based on the data's rank or position--rather than statistics based on moments (like variance), we ensure the resulting measure of deviation remains stable and reliable, regardless of data skewness or the presence of significant anomalies.

The variables utilized in this robust calculation are defined as follows:

x_i : This remains the specific data value or observation currently under evaluation.

$x?$: Represents the **median** of the entire dataset. The median is the value separating the higher half from the lower half of a data sample, and unlike the mean, it is highly resistant to the influence of extreme data points.

MAD: Stands for the [Median Absolute Deviation \(MAD\)](#) of the dataset. This measure serves as the robust replacement for the standard deviation, quantifying the spread of the data based on absolute distances from the median.

The function and calculation of the **Median Absolute Deviation (MAD)** require specific attention.

MAD is calculated in a sequential manner: first, determine the median of the dataset ($x?$); second, calculate the absolute difference between every data point (x_i) and that median ($|x_i - x?|$); and finally, find the median of those resulting absolute differences. Because the MAD itself is defined as the median of deviations, it inherits the median's insensitivity to extreme data points, making it a highly superior and stable measure of statistical dispersion when robustness against anomalies is a primary requirement. The constant 0.6745 is included because in a normal distribution, the standard deviation (σ) is theoretically approximated by the MAD divided by 0.6745 ($\sigma \approx \text{MAD} / 0.6745$). Multiplying by 0.6745 thus scales the MAD so that the resulting modified Z-score behaves comparably to the standard Z-score when the dataset is reasonably well-behaved.

Calculating the Modified Z-Score: A Step-by-Step Example

To fully grasp the calculation process and appreciate the utility of the modified Z-score, let us walk through a detailed, concrete example. For this illustration, assume we are analyzing a dataset comprising 16 observed values, which could represent anything from manufacturing tolerances to observed financial returns. We will systematically calculate the modified Z-score for each observation.

Step 1: Define the Data

Our initial dataset, consisting of $N=16$ values, is presented below. Before any calculation can proceed, it is always necessary to sort the data in ascending order to facilitate the finding of the median ($x?$).

The initial dataset values are:

Data
6
7
7
8
12
14
15
16
16
19
22
24
26
26
29
46

Step 2: Find the Median ($x?$)

The first critical step in computing the modified Z-score is establishing the **median** ($x?$) of the dataset, replacing the outlier-sensitive mean. After sorting the data in ascending order, we must locate the central value. Since our dataset contains an even number of observations ($N=16$), the median is calculated as the average of the two middle values (the 8th and 9th values in the sorted list). Upon ordering these 16 observations, the central point, or median, is determined to be **16**. This value will serve as our stable, robust measure of central tendency for all subsequent calculations.

Step 3: Calculate the Absolute Deviation from the Median

Next, we determine the extent to which every single data value (x_i) deviates from the newly calculated median ($x? = 16$). Because we are interested exclusively in the magnitude of this deviation, irrespective of direction (positive or negative), we calculate the **absolute difference**. This process involves subtracting the median (16) from each data point and then taking the absolute value of the result. For instance, calculating the absolute difference for the first data value ($x_1 = 6$) yields the following result:

$$\text{Absolute Difference} = |6 - 16| = 10$$

This calculation is meticulously repeated for all 16 values in the dataset. This step generates a new column of absolute deviation values, which are indispensable for the next step: determining the MAD.

Data	Abs. Difference from Median
6	10
7	9
7	9
8	8
12	4
14	2
15	1
16	0
16	0
19	3
22	6
24	8
26	10
26	10
29	13
46	30

Step 4: Find the Median Absolute Deviation (MAD)

The fourth step requires us to find the [Median Absolute Deviation \(MAD\)](#). As defined previously, the MAD is simply the median of the new dataset created in Step 3--the column containing the absolute differences. We must treat these differences as a new distribution and find their central value. Similar to Step 2, we sort the absolute differences and identify the middle value(s). For our running example, the median of these absolute differences is calculated to be **8**. This value, the MAD, provides the robust measure of the spread or variability within our original data sample.

Step 5: Find the Modified Z-Score for Each Data Value

Finally, having determined the robust central tendency ($x? = 16$) and the robust measure of dispersion ($MAD = 8$), we possess all the necessary components to calculate the **modified Z-score** for every observation using the complete formula:

$$\text{Modified z-score} = 0.6745(x_i - x?) / MAD$$

Using the first data value ($x_i = 6$) as an example, the calculation proceeds as follows:

$$\text{Modified z-score} = 0.6745 * (6 - 16) / 8 = 0.6745 * (-10) / 8 = \mathbf{-0.843}$$

This rigorous process is repeated for all values in the dataset. The resulting modified Z-scores quantify how far each value deviates from the median, measured in units of MAD, and scaled to be comparable to standard deviation units for normally distributed data.

Data	Abs. Difference from Median	Modified Z-Score
6	10	-0.843
7	9	-0.759
7	9	-0.759
8	8	-0.675
12	4	-0.337
14	2	-0.169
15	1	-0.084
16	0	0.000
16	0	0.000
19	3	0.253
22	6	0.506
24	8	0.675
26	10	0.843
26	10	0.843
29	13	1.096
46	30	2.529

Upon reviewing the final column of modified Z-scores, we look for any values that exceed the recommended robust threshold of $|3.5|$. In this specific dataset, the maximum absolute modified Z-score observed is 2.36. Since no value is less than -3.5 or greater than 3.5 , we conclude that, based on this robust method, none of the observations in this dataset would be flagged as a potential [outlier](#).

Interpreting Results and Handling Outliers

The primary advantage derived from employing the modified Z-score method is the confidence it provides in accurately identifying truly anomalous data points--those that are disconnected from the central tendency and inherent spread of the majority of the data. Once an observation is reliably flagged as a potential **outlier** (i.e., its modified Z-score exceeds 3.5 in magnitude), the subsequent critical phase involves making an informed, contextual decision regarding its treatment within the overall statistical analysis framework.

It is crucial to maintain the perspective that not every outlier represents an error; some reflect rare

but entirely genuine events or measurements. The appropriate approach for handling a flagged data point must therefore be tailored to the nature of the data and the specific objectives of the study. The following outlines standard procedures for managing observations identified as potential anomalies:

Verify Data Integrity: The mandatory first action must always be to confirm that the extreme value is not a consequence of a simple **data entry error**, a transcription mistake, or a measurement malfunction. A misplaced decimal point can instantly generate a significant outlier. If an error is substantiated, the value must be corrected to reflect the true measurement.

Assign a New Imputed Value: If the outlier is confirmed to be an error but the true original value cannot be determined, or if the value is deemed non-representative and removed, analysts may choose to assign a new imputed value. Common imputation techniques involve replacing the outlier with a robust measure, such as the **median** of the remaining dataset, or utilizing a calculated value derived from neighboring non-outlier data points.

Remove the Outlier: If the observation is confirmed to be a genuine, yet highly influential and detrimental, value that significantly skews the underlying data distribution, its removal may be statistically justified. This action is typically taken when the outlier severely violates core assumptions required for specific statistical tests or if its inclusion leads to dangerously misleading analytical results. If removal is performed, **transparency** is paramount: the decision, rationale, and impact on the final analysis must be thoroughly documented in the report.

Utilize Non-Parametric Methods: If the dataset exhibits numerous outliers, or if the data distribution remains persistently non-normal even after verification, an effective alternative strategy is to pivot toward **non-parametric statistical methods**. These methods are inherently distribution-free, meaning they do not rely on assumptions about the shape of the data, and are therefore naturally more resistant to the influence of extreme values.

Choosing the most suitable technique for managing outliers constitutes a vital component of sound statistical practice. The modified Z-score furnishes analysts with a powerful, robust initial tool, ensuring that the identification of anomalies is accurate, leading to more reliable and representative conclusions about the underlying phenomena being studied.