

How to Calculate a Modified Z-Score to Easily Identify Outliers

Authored by
stats writer

December 6, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Calculate a Modified Z-Score to Easily Identify Outliers*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106230>

The Modified Z-Score is an essential tool in robust statistics, specifically designed for the identification of potential outliers within a dataset. Unlike the standard Z-Score, which is highly susceptible to extreme values, the Modified Z-Score leverages the median and the Median Absolute Deviation (MAD), making it significantly more resistant to influence by anomalous data points. This measure is calculated based on the deviation from the dataset's center, standardized by its variability.

While the initial constant mentioned (1.4826) relates to estimating the standard deviation from the MAD for normally distributed data (where $1/0.6745 \approx 1.4826$), the core formula for the modified approach is centered around the median. For instance, if the data point is 30, the median is 20, and the MAD is 2, the calculation using the scaling factor (0.6745) would yield a standardized result, allowing data analysts to confidently flag observations that fall significantly outside the expected range.

Understanding the Standard Z-Score

In the field of statistics, the standard Z-Score--also known as the standard score--is a fundamental measure. It quantifies the distance and direction of a raw score from the dataset's mean, expressing this distance in units of standard deviation. Essentially, it tells us precisely how many standard deviations away a particular data value lies from the average value of the entire distribution.

The formula used globally for calculating the standard Z-Score is relatively straightforward, relying on the population or sample mean and standard deviation. This method assumes that the data is normally or nearly normally distributed for effective interpretation.

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

where the variables are defined as:

x_i : Represents a single observed data value.

μ : Represents the Mean (average) of the entire dataset.

σ : Represents the Standard Deviation of the dataset.

This traditional Z-Score is frequently employed for initial outlier detection. A common rule of thumb suggests that observations that yield a Z-Score less than -3 or greater than +3 are potentially extreme values and warrant further investigation as outliers.

The Need for Robust Statistics: Introducing the Modified Z-Score

While the standard Z-Score is useful, it suffers from a critical vulnerability: it is highly sensitive to

the presence of outliers itself. Since the calculation relies on the mean and standard deviation, and both of these metrics are easily inflated or deflated by extreme values, a single large outlier can significantly shift the mean and standard deviation, thereby masking other, less extreme outliers or making non-outliers appear significant.

To address this limitation, statisticians utilize a more robust technique: the Modified Z-Score. This calculation replaces the sensitive metrics (mean and standard deviation) with resistant metrics, namely the median and the Median Absolute Deviation (MAD). The use of the median ensures that the measure of central tendency is not skewed by extreme values.

The formula for calculating the Modified Z-Score incorporates a scaling factor (0.6745) designed to make the Modified Z-Score comparable to the standard Z-Score, assuming the data is normally distributed. This provides a robust alternative for anomaly detection, especially in datasets where normality cannot be guaranteed or where pre-existing outliers are suspected.

Modified z-score = $0.6745(x_i - x?) / MAD$

In this robust formulation, the variables represent:

x_i : The specific data value being tested.

$x?$: The Median of the complete dataset.

MAD: The Median Absolute Deviation of the dataset.

Because this measure relies on the median rather than the mean, the Modified Z-Score is inherently more robust to the presence of unusual observations. This resilience makes it a preferred method when the integrity of the center and spread measurements is paramount.

For identification purposes, standard statistical guidelines recommend that data points exhibiting modified z-scores less than -3.5 or greater than +3.5 should be flagged and investigated as potential outliers. This threshold is slightly wider than the standard Z-Score threshold (± 3) to account for the inherent robustness of the calculation.

Step-by-Step Calculation: Applying the Modified Z-Score

To fully appreciate the practical application of this statistical measure, we will walk through a detailed, step-by-step example demonstrating how to calculate the modified z-scores for a specific dataset.

Step 1: Define the Dataset

We begin by establishing a sample dataset containing 16 individual values. It is important to organize the data, typically by sorting it, although the calculation of the median (Step 2) requires

this step implicitly.

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

This organized list of values forms the basis for all subsequent calculations, particularly those revolving around the median and absolute deviations.

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

The next crucial step is finding the median of the dataset. The median represents the exact middle point of the ordered data distribution, dividing the data into two equal halves. Since our dataset has 16 values (an even number), the median is the average of the 8th and 9th values in the sorted list.

Upon calculation, the median for this specific dataset is found to be **16**. This value will serve as the central reference point ($x?$) in the Modified Z-Score formula.

Step 3: Calculate the Absolute Difference from the Median

We now calculate the absolute difference between every individual data value (x_i) and the median ($x? = 16$). This step determines how far each observation lies from the center of the distribution, regardless of direction (positive or negative).

For instance, considering the first data value, 6, the absolute difference calculation is:

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

This process is repeated for all 16 data values to generate a new column representing all absolute deviations from the central median:

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: Calculate the Median Absolute Deviation (MAD)

The Median Absolute Deviation (MAD) is the measure of statistical dispersion used in the Modified Z-Score. To find the MAD, we must calculate the median of the newly created column of absolute differences (the second column in the previous table).

By finding the middle value of these absolute differences, we obtain a highly robust measure of variability. For this dataset, the MAD is calculated to be **8**. This robust measure of spread is far less affected by extreme values than the standard deviation would be.

Step 5: Compute the Modified Z-Score for Each Value

The final step involves integrating the median ($x? = 16$) and the MAD (8) into the Modified Z-Score formula for every data point. We use the established formula:

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

To illustrate, the calculation for the first data point ($x_i = 6$) is executed as follows:

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

We then apply this calculation iteratively across the entire dataset to determine the final modified z-score for every observation:

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

After reviewing the resulting modified z-scores, we compare them against the recommended threshold of ± 3.5 . In this specific example, none of the calculated values fall below -3.5 or exceed $+3.5$. Consequently, based on the robust Modified Z-Score criteria, we would not label any observation in this particular dataset as a potential outlier.

Strategies for Handling Identified Outliers

If the Modified Z-Score calculation does successfully identify one or more outliers, data analysts must then decide on the appropriate course of action. Handling outliers correctly is critical, as arbitrary removal or manipulation can skew research findings. There are generally three established methods for managing extreme values:

Verification and Correction of Data Errors: The first and most critical step is to ensure that the flagged observation is not simply the result of a data entry error, a measurement mistake, or a transcription fault. Sometimes, a simple typo can introduce an extreme value. If an error is detected, the value should be corrected to its true measurement.

Imputation or Value Assignment: If the outlier turns out to be a result of a data entry error, you may decide to assign a new value to it such as the median of the dataset. If an outlier is confirmed to be a genuine measurement error (e.g., equipment malfunction) but its true value cannot be recovered, or if it represents a missing value, analysts may choose to assign a new, less extreme value to it.

Removal of the Observation: If the value is a true outlier, you may choose to remove it if it will have a significant impact on your overall analysis. This should be done judiciously and only if the impact on the final conclusions is deemed significant. Crucially, any removal of data points must be clearly documented and justified within the final statistical report or academic analysis.

ARABPSYCHOLOGY.COM