

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

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *What is a Modified Z-Score? (Definition & Example)*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140298>

A Modified Z-Score is a statistical measure used to identify extreme values or outliers in a dataset. It is similar to a regular Z-Score, but it uses a different formula to calculate the standard deviation, making it more robust to extreme values. This score is often used in data analysis and can help in understanding the distribution of a dataset. An example of a Modified Z-Score would be in a population study where the weight of individuals is being measured. The Modified Z-Score can help identify individuals with significantly higher or lower weight compared to the rest of the population, indicating potential outliers. This can aid in further investigation and analysis of the data. Overall, a Modified Z-Score is a useful tool for identifying and understanding extreme values in a dataset.

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

In statistics, a z-score tells us how many standard deviations away a value is from . We use the following formula to calculate a z-score:

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

where:

x_i : A single data value
 μ : The mean of the dataset
 σ : The standard deviation of the dataset

Z-scores are often used to detect outliers in a dataset. For example, observations with a z-score less than -3 or greater than 3 are often deemed to be outliers.

However, z-scores can be affected by unusually large or small data values, which is why a more robust way to

detect outliers is to use a modified z-score, which is calculated as:

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

where:

x_i : A single data value
 $x?$: The median of the dataset
MAD: The median absolute deviation of the dataset

A modified z-score is more robust because it uses the median to calculate z-scores as opposed to the mean, .

recommend that values with modified z-scores less than -3.5 or greater than 3.5 be labeled as potential outliers.

The following step-by-step example shows how to calculate modified z-scores for a given dataset.

Step 1: Create the Data

Suppose we have the following dataset with 16 values:

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

Step 2: Find the Median

Next, we will find the median. This represents the middle point in the dataset, which turns out to be 16.

Step 3: Find the Absolute Difference Between Each Value & the Median

Next, we will find the absolute difference between each individual data value and the median. For example, the absolute difference between the first data value and the median is calculated as:

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

We can use the same formula to calculate the absolute difference between each individual data value and the 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: Find the Median Absolute Deviation

Next, we'll find the median absolute deviation. This is the median of the second column, which turns out to be 8.

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

Lastly, we can calculate the modified z-score for each data value using the following formula:

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

For example, the modified z-score for the first data value is calculated as:

Modified z-score = $0.6745*(6-16) / 8 = -0.843$

We can repeat this formula for every value in the dataset:

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

We can see that no value in the dataset has a modified z-score less than -3.5 or greater than 3.5, thus we wouldn't label any value in this dataset as a potential outlier.

How to Handle Outliers

If an outlier is present in your dataset, you have a few options:

Make sure the outlier is not the result of a data entry error. Sometimes an individual simply enters the wrong data value when recording data. If an outlier is present, first verify that the value was entered correctly and that it wasn't an error. Assign a new value to the outlier. 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 of the dataset. Remove the outlier. If the value is a true outlier, you may choose to remove it if it will have a significant impact on your overall analysis. Just make sure to mention in your final report or analysis that you removed an outlier.