

How to Calculate Modified Z-Scores in Excel: A Step-by-Step Guide

Authored by
stats writer

December 6, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Calculate Modified Z-Scores in Excel: A Step-by-Step Guide*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=106215>

The Importance of Robust Statistics and the Modified Z-Score

The field of statistics relies heavily on metrics that accurately describe the spread and central tendency of a data set. While the traditional Z-score (or standard score) is widely utilized, it suffers from a significant drawback: its reliance on the arithmetic mean and standard deviation, which are extremely sensitive to outliers. This sensitivity can skew the perception of data dispersion and lead to inaccurate conclusions regarding whether a specific observation is truly anomalous. To address this limitation, statisticians often turn to more robust methods, such as those employing the median rather than the mean.

The concept of the modified z-score provides a powerful alternative for identifying potential outliers in scenarios where the data may not follow a perfect normal distribution, or when the presence of extreme values is suspected. By substituting the mean with the median and the standard deviation with the Median Absolute Deviation (MAD), the modified z-score ensures that the resulting measure of deviation is less influenced by unusual or extreme observations. This statistical robustness makes the modified z-score an indispensable tool in data cleaning and preliminary analysis across various scientific disciplines.

The calculation process, which we will detail using Microsoft Excel, demonstrates a practical application of these robust statistical principles. By learning how to implement this calculation in a common software environment, analysts can efficiently screen large datasets for observations that warrant closer scrutiny. This approach is particularly valuable when analyzing real-world data, where errors, measurement noise, or genuine but rare events can produce values that severely distort measures based on the mean.

Defining the Modified Z-Score Formula

In statistical terms, the modified z-score provides a standardized measure of how far a specific data point is from the center of the distribution, relative to the typical deviation observed in the dataset. Unlike the standard Z-score, which assumes normality and uses the standard deviation as its scale, the modified z-score scales the difference between the data point and the median using the Median Absolute Deviation (MAD).

The formal definition of the modified z-score is calculated using the following equation:

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

This formula uses several critical components that ensure its robustness against outliers. It is essential to understand the meaning of each variable before attempting the calculation in Excel:

x_i : Represents a single observation or data value from the dataset under analysis.

$x?$: Denotes the median of the entire dataset, which serves as the robust measure of central tendency.

MAD: Stands for the Median Absolute Deviation, which is the median of the absolute deviations from the median, serving as the robust measure of statistical dispersion.

The constant 0.6745 is included as a scaling factor. This value is derived from the fact that, for a large, normally distributed dataset, the MAD is approximately 0.6745 times the standard deviation. Multiplying the entire expression by this constant ensures that, in the case of normally distributed data, the modified z-score is comparable in magnitude to the standard Z-score, allowing for similar interpretation thresholds for identifying potential outliers.

Thresholds for Identifying Potential Outliers

One of the primary applications of calculating the modified z-score is the identification of potential outliers. Unlike the standard Z-score, which often uses a threshold of ± 3 , the modified z-score uses a slightly tighter, yet more conservative, threshold due to its reliance on the median and MAD.

Many expert practitioners recommend that any observation yielding a modified z-score less than **-3.5** or greater than **3.5** should be flagged as a potential outlier. This specific threshold is critical because it balances the need to detect genuinely anomalous data points without mistakenly identifying values that are merely part of the expected variance in a non-normal distribution.

The subsequent step-by-step example walks through the practical implementation of these calculations in Microsoft Excel, providing a detailed guide on how to derive the modified z-score for every data point within a sample dataset. Following these steps ensures accurate and robust outlier detection capabilities.

Step 1: Setting Up the Sample Data in Excel

The first essential step in calculating the modified z-scores is organizing the raw data efficiently within an Excel spreadsheet. This foundational arrangement sets the stage for all subsequent calculations, ensuring that formulas can be applied consistently across the entire range of observations. We begin by inputting the raw data into a single column, starting from cell A1.

For this example, we will utilize a dataset comprising 16 distinct numerical values. Organizing the data in this manner--a vertical column--simplifies the calculation of descriptive statistics and facilitates the use of Excel's built-in array functions.

The dataset should look like the following image structure after entry:

	A	B	C	D	E	F
1	Data					
2	6					
3	7					
4	7					
5	8					
6	12					
7	14					
8	15					
9	16					
10	16					
11	19					
12	22					
13	24					
14	26					
15	26					
16	29					
17	46					
18						
19						
20						
21						
22						

By structuring the data first, we prepare the necessary inputs for calculating the median and the subsequent absolute deviations, which are the two core elements needed to compute the Median Absolute Deviation (MAD) and, ultimately, the final modified z-scores. Proper initial setup is crucial for avoiding errors in later, more complex formulas.

Step 2: Calculating the Median of the Dataset

The second step involves determining the median of the dataset, denoted as $x?$. The median is defined as the middle value when all the observations are sorted in ascending order. Since it is based on the rank position rather than the numerical magnitude of the values, the median is inherently robust against extreme outliers. Excel makes this calculation straightforward using a dedicated function.

To calculate the median, select an empty cell (e.g., cell B1) and input the following formula: `=MEDIAN(A2:A17)`, assuming your data occupies cells A2 through A17. This function instructs Excel to find the median value among all 16 observations.

	A	B	C	D	E	F	G	H
1	Data				Median	16	=MEDIAN(A2:A17)	
2	6							
3	7							
4	7							
5	8							
6	12							
7	14							
8	15							
9	16							
10	16							
11	19							
12	22							
13	24							
14	26							
15	26							
16	29							
17	46							
18								
19								
20								
21								
22								
23								
24								

Upon executing the formula for the sample dataset, the calculated median value should be **16**. This value, 16, now represents the robust measure of central tendency for our data, replacing the traditional mean that would have been used in a standard Z-score calculation. This median will be used in subsequent steps to determine the deviation of every single observation.

Step 3: Determining the Absolute Difference from the Median

The third critical phase involves calculating the absolute deviation of each individual data value (x_i) from the robust measure of central tendency (the median, $x?$). This step is essential because the Median Absolute Deviation (MAD) is calculated based on these absolute differences. The use of the absolute difference ensures that both positive and negative deviations from the median are treated equally in terms of magnitude.

To achieve this in Excel, we must create a new column, typically column B, labeled "Absolute Difference." For the first data point in cell A2, we enter a formula in cell B2 that finds the absolute difference between A2 and the median value (which we assume is stored in B1, anchored using absolute referencing). The required Excel formula is: `=ABS(A2-B1)`. The dollar signs around B1 are crucial; they fix the reference to the median value so that when the formula is copied down, it

always refers back to the single median cell.

	A	B	C	D	E	F	G
1	Data	Abs Diff			Median	16	
2	6	10	=ABS(A2-\$F\$1)				
3	7						
4	7						
5	8						
6	12						
7	14						
8	15						
9	16						
10	16						
11	19						
12	22						
13	24						
14	26						
15	26						
16	29						
17	46						
18							
19							
20							
21							
22							
23							

Once the formula is correctly entered in cell B2, it must be applied to all remaining data points. To efficiently copy this formula down the column, click on cell B2 and hover over the bottom right corner until a small cross icon (+) appears. Double-clicking this cross automatically copies and pastes the formula to every corresponding row in the dataset, calculating the absolute deviation for all 16 values instantaneously.

	A	B	C	D	E	F	G
1	Data	Abs Diff			Median	16	
2	6	10					
3	7	9					
4	7	9					
5	8	8					
6	12	4					
7	14	2					
8	15	1					
9	16	0					
10	16	0					
11	19	3					
12	22	6					
13	24	8					
14	26	10					
15	26	10					
16	29	13					
17	46	30					
18							
19							
20							
21							
22							

This resulting column of absolute differences forms the basis for the next calculation, where we will determine the median of these deviation values to arrive at the Median Absolute Deviation.

Step 4: Calculating the Median Absolute Deviation (MAD)

The fourth step involves calculating the Median Absolute Deviation (MAD), which is the robust measure of variability required for the modified z-score. The MAD is defined as the median of the list of absolute differences calculated in Step 3. It essentially measures the typical distance of data points from the center of the distribution.

To compute the MAD, select an empty cell (e.g., cell C1) and apply the `MEDIAN` function again, but this time referencing the column containing the absolute differences (Column B, cells B2 through B17). The required formula is: `=MEDIAN(B2:B17)`.

	A	B	C	D	E	F	G	H
1	Data	Abs Diff			Median	16		
2	6	10			MAD	8	=MEDIAN(B2:B17)	
3	7	9						
4	7	9						
5	8	8						
6	12	4						
7	14	2						
8	15	1						
9	16	0						
10	16	0						
11	19	3						
12	22	6						
13	24	8						
14	26	10						
15	26	10						
16	29	13						
17	46	30						
18								
19								
20								
21								
22								
23								
24								
25								
26								

For the sample dataset utilized in this demonstration, the calculated MAD turns out to be **8**. This value, 8, is the Median Absolute Deviation, and it serves as the denominator in the final modified z-score formula. Since the MAD is calculated using medians twice (once for the center, once for the spread), it remains highly resilient to the influence of extreme outliers, providing a much more stable measure of spread than the standard deviation would.

Having successfully calculated both the median ($x?$) and the MAD, we now possess all the necessary components to calculate the final modified z-score for every single observation in the dataset, leading directly into Step 5.

Step 5: Final Calculation of Modified Z-Scores

The final and most crucial step is to calculate the modified z-score for each individual data point. This calculation standardizes the deviation using the MAD, enabling us to determine which points lie far outside the typical range defined by robust statistics. We apply the defined formula, which incorporates the constant 0.6745.

Recall the core formula: **Modified z-score = 0.6745 * (xi - x?) / MAD**. In Excel, we will create a new column, Column C, to hold these scores. For the first data value in cell A2, we reference the

original value (A2), the calculated median (B1), and the calculated MAD (C1). Remember to use absolute references (dollar signs) for the median and MAD values so they remain constant when the formula is copied.

The specific Excel formula for the first data point in cell C2 is: $= (A2 - \$B\$1) * 0.6745 / \$C\1 . This formula translates the calculation directly into Excel logic. For instance, the calculation for the first data point (24) would be: $(24 - 16) * 0.6745 / 8$.

	A	B	C	D	E	F	G
1	Data	Abs Diff	Modified z-score		Median	16	
2	6	10	$=0.6745*(A2-\$F\$1)/\$F\2		MAD	8	
3	7	9					
4	7	9					
5	8	8					
6	12	4					
7	14	2					
8	15	1					
9	16	0					
10	16	0					
11	19	3					
12	22	6					
13	24	8					
14	26	10					
15	26	10					
16	29	13					
17	46	30					
18							
19							
20							
21							
22							

To complete the process, click on cell C2 and utilize the fill handle (the small cross, +, appearing in the bottom right corner) to swiftly copy and paste this formula down the entire column. This action calculates the modified z-score for all 16 data values in the dataset, providing a complete picture of how standardized each point is relative to the robust median.

	A	B	C	D	E	F	G
1	Data	Abs Diff	Modified z-score		Median	16	
2	6	10	-0.843125		MAD	8	
3	7	9	-0.7588125				
4	7	9	-0.7588125				
5	8	8	-0.6745				
6	12	4	-0.33725				
7	14	2	-0.168625				
8	15	1	-0.0843125				
9	16	0	0				
10	16	0	0				
11	19	3	0.2529375				
12	22	6	0.505875				
13	24	8	0.6745				
14	26	10	0.843125				
15	26	10	0.843125				
16	29	13	1.0960625				
17	46	30	2.529375				
18							
19							
20							
21							
22							
23							
24							
25							

Upon completion, we can analyze the resulting z-scores. In this specific example, upon reviewing the calculated scores, we observe that no value exceeds the absolute threshold of 3.5 (i.e., less than -3.5 or greater than 3.5). Therefore, based on the modified z-score criteria, we would not label any observation in this dataset as a potential outlier, confirming the robustness and relatively tight distribution of the sample data.

Best Practices for Handling Identified Outliers

If the calculation of the modified z-score successfully flags one or more values as potential outliers (scores outside the +/- 3.5 range), the analyst must then decide on the appropriate course of action. Handling outliers is a critical decision that requires careful judgment and subject-matter expertise, as inappropriate handling can lead to biased or misleading results.

One of the first and most crucial steps is to **verify the integrity of the data entry**. Many apparent outliers are simply the result of human error--a misplaced decimal point, a typo, or an incorrect unit conversion during data recording. If an observation appears anomalous, the analyst should trace the data back to its original source to confirm whether the value was entered correctly. If a data entry error is confirmed, the value should be corrected to its true measurement, which will

subsequently recalculate the modified z-score and likely bring the data point back within acceptable limits.

If the value is confirmed to be a true measurement but is still flagged as an outlier, two common methods are often considered: ****imputation or removal****. If the outlier is minor, or if the analyst fears the impact of its removal on the sample size, they might choose to assign a new, less extreme value to it--a process known as imputation. A common imputation technique is assigning the value of the dataset's median to the outlier, as the median is already known to be robust and representative of the central tendency.

Finally, if the outlier is highly influential and represents a genuine but extremely rare event, and if it significantly distorts the overall analysis, the analyst may choose to **remove the outlier** entirely. This option should be exercised with caution. If removal is deemed necessary, it is paramount that the analyst clearly documents and justifies this decision in their final report or publication, detailing the criteria used (e.g., the modified z-score threshold) and the resulting impact on the final statistics. Failing to report the removal of outliers can severely undermine the transparency and trustworthiness of the research findings.

Conclusion: Advantages of Using Robust Metrics

The application of the modified z-score highlights a significant shift towards utilizing robust statistical methods in modern data analysis. By favoring the median and the Median Absolute Deviation (MAD) over the mean and standard deviation, analysts gain a powerful tool that resists the undue influence of extreme values. This resistance is crucial for maintaining the fidelity of descriptive statistics, particularly in datasets derived from complex, real-world processes.

Mastering the step-by-step process of calculating the modified z-score in Excel empowers researchers and data scientists to conduct more accurate and reliable outlier detection. This structured approach ensures that potential anomalies are identified based on sound, robust statistical principles, providing a strong foundation for subsequent modeling and inference.

Ultimately, the careful identification and appropriate handling of outliers, informed by the modified z-score, ensures that statistical conclusions are reflective of the underlying data patterns rather than distorted by noise or error. This rigorous methodology is a hallmark of high-quality data analysis.