

How can I calculate modified Z-scores in Excel?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate modified Z-scores in Excel?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=140317>

Modified Z-scores in Excel refer to a statistical measure used to identify outliers in a dataset. To calculate modified Z-scores in Excel, you can use the built-in function "MDETERM" which calculates the determinant of a matrix. First, arrange your dataset in a column and then use the function to find the median and median absolute deviation (MAD) of the dataset. Next, subtract the median from each data point and divide it by the MAD. This will give you the Z-score for each data point. Finally, apply a threshold value, typically 3 or 3.5, to identify and remove the outliers from your dataset. This method is useful for detecting extreme values that may skew the analysis of your data.

Calculate Modified Z-Scores in Excel

In statistics, a modified z-score 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 than an ordinary z-score because it uses the median in its formula 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 in Excel.

Step 1: Create the Data

First, we'll create the following dataset that contains 16 values:

	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						

Step 2: Calculate the Median

Next, we'll calculate the median of the dataset:

	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								

The median turns out to be 16.

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

Next, we'll calculate the absolute difference between each value and the median:

	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							

Next, click on cell B2. Then hover over the bottom right corner of the cell until a little cross (+) appears.

Double click the cross to copy and paste this formula to all remaining cells in the column:

	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							

Step 4: Calculate the Median Absolute Deviation

Next, we will use the following formula to calculate the median absolute deviation of the dataset:

	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								

The median absolute deviation 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:

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

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

	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							

Next, click on cell C2. Then hover over the bottom right corner of the cell until a little cross (+) appears.

Double click the cross to copy and paste this formula to all remaining cells in the column:

	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							

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.