

How can I easily identify outliers in Excel?

Authored by
stats writer

April 17, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I easily identify outliers in Excel?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=136295>

The process of identifying outliers in Excel can be easily achieved by utilizing various statistical methods and tools available in the software. These include using the built-in functions such as QUARTILE, IQR, and Z-SCORE, as well as creating visual representations like box plots and scatter plots. By applying these techniques, users can efficiently detect and eliminate any unusual or extreme data points that may significantly impact the overall analysis and results. This allows for more accurate and reliable data analysis in Excel.

Easily Find Outliers in Excel

An outlier is an that lies abnormally far away from other values in a dataset.

Outliers can be problematic because they can affect the results of an analysis.

We will use the following dataset in Excel to illustrate two methods for finding outliers:

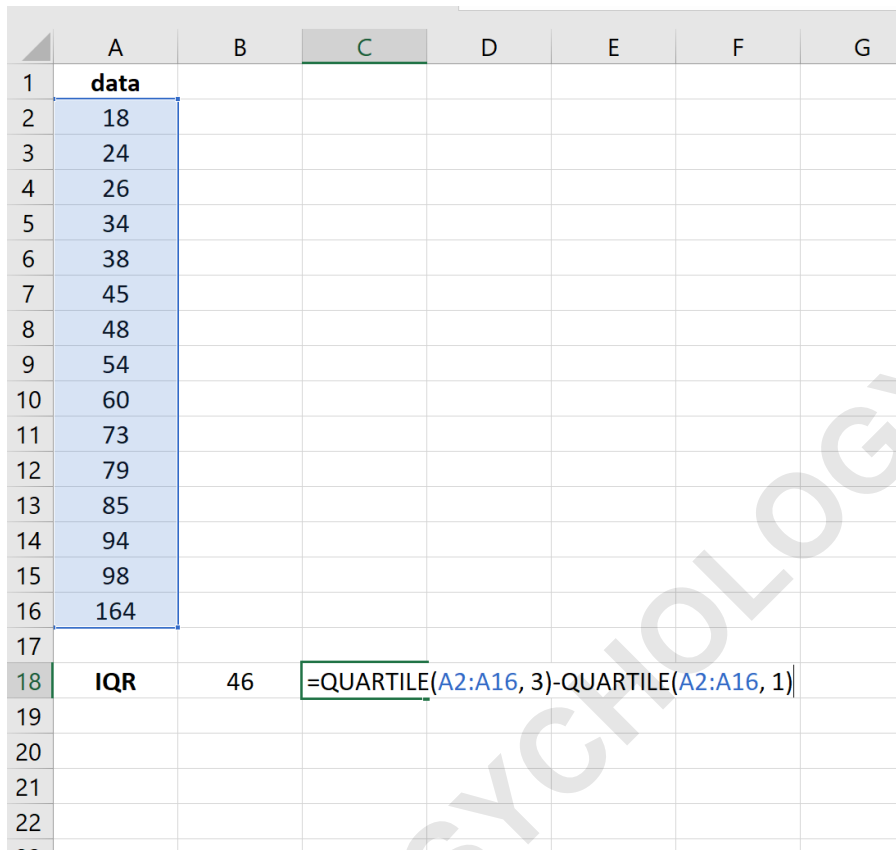
	A	B	C	D	E
1	data				
2	18				
3	24				
4	26				
5	34				
6	38				
7	45				
8	48				
9	54				
10	60				
11	73				
12	79				
13	85				
14	94				
15	98				
16	164				
17					
18					
19					
20					
21					

Method 1: Use the interquartile range

The (IQR) is the difference between the 75th percentile (Q3) and the 25th percentile (Q1) in a dataset. It measures the spread of the middle 50% of values.

We can define an observation to be an outlier if it is 1.5 times the interquartile range greater than the third quartile (Q3) or 1.5 times the interquartile range less than the first quartile (Q1).

The following image shows how to calculate the interquartile range in Excel:



	A	B	C	D	E	F	G
1	data						
2	18						
3	24						
4	26						
5	34						
6	38						
7	45						
8	48						
9	54						
10	60						
11	73						
12	79						
13	85						
14	94						
15	98						
16	164						
17							
18	IQR	46	=QUARTILE(A2:A16, 3)-QUARTILE(A2:A16, 1)				
19							
20							
21							
22							
23							

Next, we can use the formula mentioned above to assign a "1" to any value that is an outlier in the dataset:

	A	B	C	D	E	F	G	H	I	J
1	data	outlier?								
2	18	=IF(OR(A2<QUARTILE(\$A\$2:\$A\$16, 1)-1.5*\$B\$18, A2>QUARTILE(\$A\$2:\$A\$16, 3)+1.5*\$B\$18), 1, 0)								
3	24	0								
4	26	0								
5	34	0								
6	38	0								
7	45	0								
8	48	0								
9	54	0								
10	60	0								
11	73	0								
12	79	0								
13	85	0								
14	94	0								
15	98	0								
16	164	1								
17										
18	IQR	46								
19										
20										
21										
22										
23										

We see that only one value - 164 - turns out to be an outlier in this dataset.

Method 2: Use z-scores

A tells you how many standard deviations a given value is from the mean. We use the following formula to calculate a z-score:

$$z = (X - \mu) / \sigma$$

where:

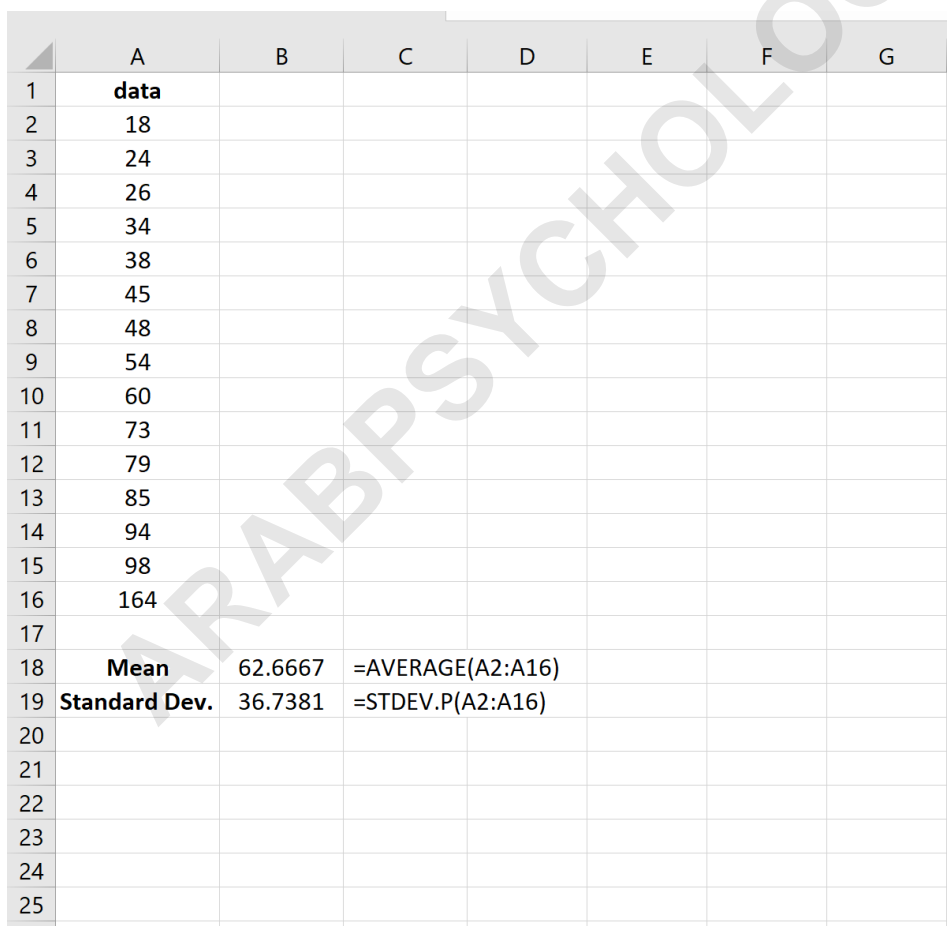
X is a single raw data value

μ is the population mean

σ is the population standard deviation

We can define an observation to be an outlier if it has a z-score less than -3 or greater than 3.

The following image shows how to calculate the mean and standard deviation for a dataset in Excel:



	A	B	C	D	E	F	G
1	data						
2	18						
3	24						
4	26						
5	34						
6	38						
7	45						
8	48						
9	54						
10	60						
11	73						
12	79						
13	85						
14	94						
15	98						
16	164						
17							
18	Mean	62.6667	=AVERAGE(A2:A16)				
19	Standard Dev.	36.7381	=STDEV.P(A2:A16)				
20							
21							
22							
23							
24							
25							

We can then use the mean and standard deviation to

find the z-score for each individual value in the dataset:

	A	B	C	D	E	F
1	data	z-score				
2	18	<code>= (A2-\$B\$18)/\$B\$19</code>				
3	24	-1.052495				
4	26	-0.998055				
5	34	-0.780298				
6	38	-0.671419				
7	45	-0.480881				
8	48	-0.399222				
9	54	-0.235904				
10	60	-0.072586				
11	73	0.28127				
12	79	0.444588				
13	85	0.607906				
14	94	0.852884				
15	98	0.961762				
16	164	2.758262				
17						
18	Mean	62.6667	<code>=AVERAGE(A2:A16)</code>			
19	Standard Dev.	36.7381	<code>=STDEV.P(A2:A16)</code>			
20						
21						
22						
23						

We can then assign a "1" to any value that has a z-score less than -3 or greater than 3:

	A	B	C	D	E
1	data	z-score	outlier?		
2	18	-1.215813	=IF(OR(B2<-3,B2>3),1,0)		
3	24	-1.052495	0		
4	26	-0.998055	0		
5	34	-0.780298	0		
6	38	-0.671419	0		
7	45	-0.480881	0		
8	48	-0.399222	0		
9	54	-0.235904	0		
10	60	-0.072586	0		
11	73	0.28127	0		
12	79	0.444588	0		
13	85	0.607906	0		
14	94	0.852884	0		
15	98	0.961762	0		
16	164	2.758262	0		
17					
18	Mean	62.6667			
19	Standard Dev.	36.7381			
20					
21					

Using this method, we see that there are no outliers in the dataset.

Note: Sometimes a z-score of 2.5 is used instead of 3. In this case, the individual value of 164 would be considered an outlier since it has a z-score greater than 2.5.

When using the z-score method, use your best judgement for which z-score value you consider to be an outlier.

How to Handle Outliers

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

1. 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.

2. 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.

3. Assign a new value to the outlier.

If the outlier is the result of a data entry error, you may decide to assign a new value to it such as of the dataset.