How Can I Easily Find Outliers in Google Sheets?

Authored by

stats writer

May 5, 2024

Google Sheets is a powerful tool for organizing and analyzing data. However, when dealing with large datasets, it can be difficult to identify outliers, or data points that are significantly different from the rest of the data. In order to easily find outliers in Google Sheets, there are a few simple steps that can be followed. By using the built-in functions and tools, such as the conditional formatting and the outlier detection add-on, users can efficiently identify and highlight outliers in their data. This allows for a more thorough analysis and better understanding of the data set. With these tools, users can easily and accurately identify outliers in Google Sheets, making it a valuable tool for data analysis.

Easily Find Outliers in Google Sheets

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

We often define an observation to be an outlier if it is 1.5 times the interquartile range greater than the third quartile or 1.5 times the interquartile range less than the first quartile.

Note: The interquartile range is the difference between the third quartile (75th percentile) and the first quartile (25th percentile) in a dataset. It measures the spread of the middle 50% of values.

The following step-by-step example shows how to use this formula to find outliers in a dataset in Google Sheets.

Step 1: Enter the Data

First, let's enter the values for the following dataset into Google Sheets:

	A	B	C	D	E	F
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						
22						
23						

Step 2: Calculate the Interquartile Range

Next, let's calculate the first quartile, third quartile, and interquartile range of the dataset:

	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			*Formula*		
18	**Q1**	36	=QUARTILE(A2:A16, 1)		
19	**Q3**	82	=QUARTILE(A2:A16, 3)		
20	**IQR**	46	=B19-B18		
21					
22					
23					
24					
25					

Step 3: Identify Outliers

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

=IF(A2<$B$18-$B$20*1.5,1,IF(A2>$B$19+$B$20*1.5,1,0))

This formula checks to see if an observation is 1.5 times the interquartile range greater than the third quartile or 1.5 times the interquartile range less than the

first quartile.

If either is true, the observation is assigned a "1" to designate it as an outlier.

The following screenshot shows how to use this formula in practice:

B2 fx =IF(A2<$B$18-$B$20*1.5, 1, IF(A2>$B$19+$B$20*1.5, 1, 0))

	A	B	C	D	E
1	**Data**	**Outlier?**			
2	18	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			*Formula*		
18	**Q1**	36	=QUARTILE(A2:A16, 1)		
19	**Q3**	82	=QUARTILE(A2:A16, 3)		
20	**IQR**	46	=B19-B18		
21					
22					
23					

How to Deal With 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 data simple gets recorded incorrectly. If an outlier is present, first check that the value was entered correctly and that it wasn't an error.

2. 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 the mean or the median of the dataset.

3. 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 that you removed an outlier.

The following tutorials explain how to remove outliers in other statistical software: