

How can I identify outliers using the interquartile range?

Authored by
stats writer

April 24, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I identify outliers using the interquartile range?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138712>

The interquartile range is a statistical measure that can be used to identify outliers in a dataset. It is calculated by finding the difference between the upper and lower quartiles of the data. Outliers, which are data points that fall significantly outside of the range of the rest of the data, can be identified by comparing them to the interquartile range. Any data points that are more than 1.5 times the interquartile range above the upper quartile or below the lower quartile can be considered outliers. This method is useful for identifying extreme values in a dataset and can help in detecting errors or unusual data points that may skew the overall analysis.

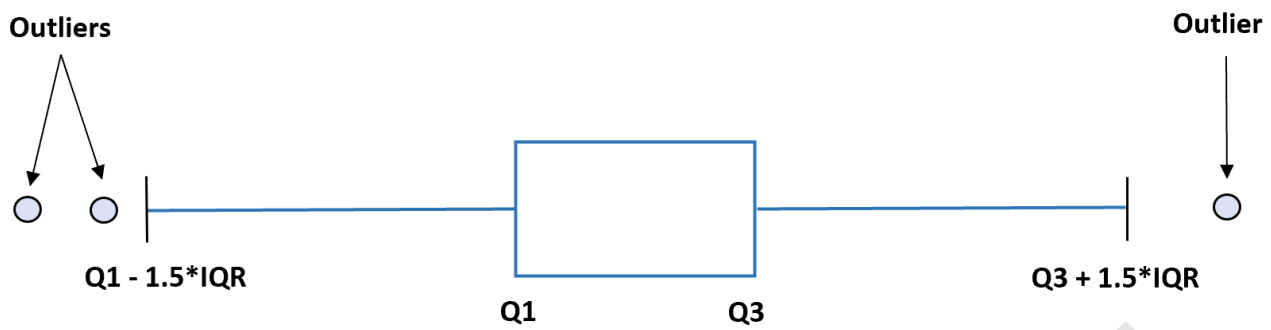
Find Outliers Using the Interquartile Range

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.

One common way to find outliers in a dataset is to use the interquartile range.

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

One popular method is to declare an observation to be an outlier if it has a value 1.5 times greater than the IQR or 1.5 times less than the IQR.



This tutorial provides a step-by-step example of how to find outliers in a dataset using this method.

Step 1: Create the Data

Suppose we have the following dataset:

Data
1
3
3
4
8
11
13
14
15
17
22
24
26
46

Step 2: Identify the First and Third Quartile

The first quartile turns out to be 5 and the third quartile turns out to be 20.75.

Data	
1	
3	
3	
4	Q1 = 5
8	
11	
13	
14	
15	
17	Q3 = 20.75
22	
24	
26	
46	

IQR = 20.75 - 5 = 15.75

Thus, the interquartile range turns out to be 20.75 - 5 = 15.75.

Step 3: Find the Lower and Upper Limits

The lower limit is calculated as:

$$\text{Lower limit} = Q1 - 1.5 \times \text{IQR} = 5 - 1.5 \times 15.75 = -18.625$$

And the upper limited is calculated as:

Data	
1	
3	
3	
4	Q1 = 5
8	
11	
13	
14	
15	Q3 = 20.75
17	
22	
24	
26	
46	

IQR = 20.75 - 5 = 15.75

Lower Limit = $5 - 1.5 \times 15.75 = -18.625$

Upper Limit = $20.75 + 1.5 \times 15.75 = 44.375$

Step 4: Identify the Outliers

The only observation in the dataset with a value less than the lower limit or greater than the upper limit is 46. Thus, this is the only outlier in this dataset.

Data	
1	
3	
3	
4	Q1 = 5
8	
11	
13	
14	
15	Q3 = 20.75
17	
22	
24	
26	
46	OUTLIER

IQR = 20.75 - 5 = 15.75

Lower Limit = $5 - 1.5 \times 15.75 = -18.625$

Upper Limit = $20.75 + 1.5 \times 15.75 = 44.375$

Note: You can use this to automatically find the upper and lower boundaries for outliers in a given dataset.

How to Find Outliers in Practice

The following tutorials explain how to find outliers using the interquartile range in different statistical software: