

# How to Find Outliers Using the Interquartile Range

Authored by  
**stats writer**

December 13, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Find Outliers Using the Interquartile Range*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=107356>

The Interquartile Range (IQR) stands as an exceptionally robust and widely adopted technique for the systematic detection of outliers within any given data set. Outliers are characterized as observations that deviate significantly from the general pattern or distribution of the remaining data points, often skewing statistical analysis results. The fundamental principle of the IQR method involves calculating the spread of the central 50% of the data, which is defined by the difference between the 75th percentile (the third quartile, Q3) and the 25th percentile (the first quartile, Q1). By establishing a range based on this central spread, we can define upper and lower fences, mathematically identifying values that fall too far outside the expected distribution. Specifically, any data observation lying more than 1.5 times the calculated IQR below Q1 or above Q3 is statistically flagged as a potential outlier, providing a quick and effective mechanism for data cleaning and preliminary analysis.

The utility of the IQR method transcends simple visual inspection, offering a quantitative and objective standard for anomaly detection. Unlike methods that rely purely on standard deviation, the IQR is based on the median, making it highly resistant to the influence of extreme values. This property means that the very outliers you are trying to find do not distort the boundaries used to identify them. This robustness is critical in fields such as finance, quality control, and social sciences, where data integrity is paramount and extreme values, whether they are genuine anomalies or errors, must be correctly identified and managed using a statistically sound method.

## Defining the Significance of Outliers in Data Analysis

An outlier is fundamentally an observation point that deviates significantly from other observations. In simple terms, it is an element that lies abnormally far away from the majority of other values in the collected data set. Identifying and properly handling these points is not merely an optional step but a critical prerequisite for accurate statistical modeling and reliable conclusions. If left unchecked, these extreme values can disproportionately influence summary statistics, such as the mean and standard deviation, leading to biased parameter estimates and misleading interpretations of the underlying phenomena being studied. Therefore, robust methods for detection, such as those relying on the Interquartile Range, are essential tools in a statistician's toolkit to ensure data quality and integrity.

The necessity of outlier detection stems from the fact that they can represent several different scenarios, each requiring a distinct approach. In some cases, outliers are simply measurement errors, data entry mistakes, or malfunctions during data collection--in which case they should usually be removed or corrected to clean the data set. In other instances, however, outliers represent rare but genuine phenomena, providing valuable insights into extremes within the population, such as peak transaction volumes or unusual demographic characteristics. A well-structured analysis must first utilize methods like the IQR to flag these points, followed by a careful qualitative review to determine their true origin and the appropriate subsequent action, whether it is

removal, transformation, or special consideration in modeling, ensuring the analysis reflects reality.

## Understanding the Interquartile Range (IQR)

The Interquartile Range (IQR) serves as a key measure of statistical dispersion, describing the range within which the middle 50% of the data lies. It is conceptually straightforward and mathematically defined as the difference between the third quartile (Q3) and the first quartile (Q1). The Q1 represents the 25th percentile of the data, meaning 25% of the data points fall below this value. Conversely, Q3 represents the 75th percentile, meaning 75% of the data points fall below it. By focusing solely on the span between these two central boundaries, the IQR effectively ignores the most extreme upper and lower 25% of the distribution, thereby providing a measure of spread that is not unduly influenced by the data's tails, including potential outliers.

The mathematical relationship for the IQR is simple: **IQR = Q3 - Q1**. This calculated range forms the basis for the most widely accepted rule for outlier identification--the 1.5 IQR rule, also known as Tukey's Fences. Because the IQR measures the central spread, we use multiples of this spread to define how far a typical data point should be allowed to drift before being labeled as statistically unusual. This non-parametric approach is highly favored in exploratory data set analysis because it makes minimal assumptions about the underlying distribution of the data, making it suitable for both normally distributed and highly skewed datasets, offering superior performance compared to mean-based methods when extreme values are present.

## The 1.5 IQR Rule: Establishing the Outlier Boundaries

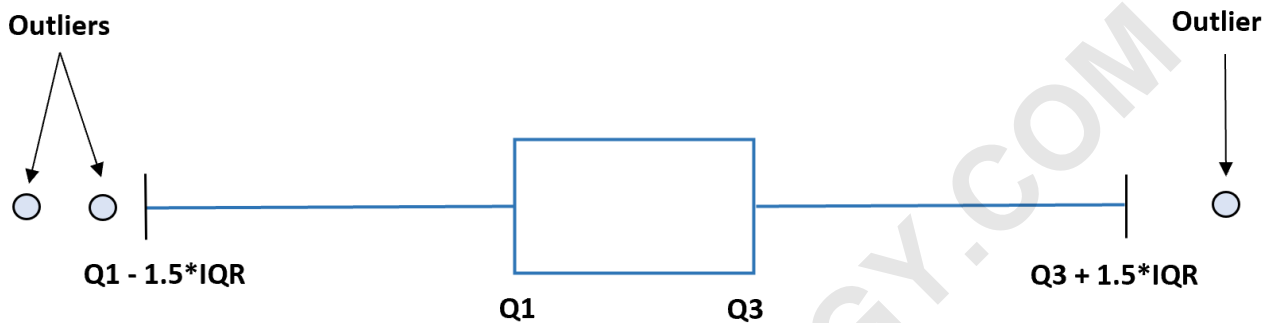
The standard methodology for detecting outliers using the IQR is known as Tukey's method. This technique defines specific fences or limits beyond which any observation is deemed an outlier. The factor 1.5 is a conventional constant established through extensive statistical practice and observation, providing a balance between sensitivity (identifying true anomalies) and specificity (avoiding false positives). The calculation involves defining a Lower Limit (or Lower Fence) and an Upper Limit (or Upper Fence) relative to Q1 and Q3, respectively. Observations falling outside these calculated limits are flagged for further investigation before finalizing the data set for modeling.

The formal calculation for these boundaries is defined by two key equations, both of which utilize the IQR and the quartile values. The **Lower Limit** is found by subtracting 1.5 times the IQR from the first quartile (Q1). Conversely, the **Upper Limit** is found by adding 1.5 times the IQR to the third quartile (Q3). Mathematically, this robust methodology is expressed as:

$$\text{Lower Limit} = Q1 - (1.5 \times \text{IQR})$$

$$\text{Upper Limit} = Q3 + (1.5 \times \text{IQR})$$

If any single data point in the entire data set has a value less than the calculated Lower Limit or greater than the calculated Upper Limit, that point is classified as an outlier. This systematic approach ensures that the identification process is objective and repeatable. The visual representation of this method is typically shown in a box plot, where the whiskers extend precisely to the most extreme data points that are still within the 1.5 IQR fences, and outliers are plotted individually beyond those whiskers, making visual identification straightforward.



This comprehensive tutorial will now walk through a step-by-step numerical example, demonstrating precisely how to apply the 1.5 IQR rule to a specific data set, thus translating the theoretical concepts into practical application for robust outlier identification and data hygiene.

## Step 1: Preparing and Examining the Sample Data Set

The initial stage in any statistical process is defining and organizing the input data. For demonstration purposes, let us assume we are working with the following sample data set, which represents a small collection of numerical observations, perhaps measurements or test scores. Before proceeding with any calculation involving quartile determination, it is absolutely essential that the data be sorted in ascending order. Sorting ensures that the subsequent identification of quartiles (Q1 and Q3) is mathematically correct and aligned with the definition of percentile ranks, as these are positional measures.

Suppose we begin with an unsorted dataset. The critical step is to arrange these observations sequentially to accurately determine the positional values of the quartiles. The sorting process establishes the true median (Q2) and the points dividing the data into four equal parts (Q1 and Q3). Once sorted, the list provides the necessary foundation for computing the IQR effectively, allowing us to move forward with confidence in the positional accuracy of our boundaries.

The sample data set for this demonstration, already presented in an ordered fashion for clarity, is:

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

## Step 2: Calculating the Quartile Values (Q1 and Q3)

Once the data set is correctly ordered, the next critical step involves identifying the first quartile (Q1) and the third quartile (Q3). These values delineate the boundaries of the middle 50% of the data. Q1 is the median of the lower half of the data, representing the 25th percentile, and Q3 is the median of the upper half of the data, representing the 75th percentile. The precise method for locating these quartile values depends slightly on the total number of observations (N) and the specific definition of the quartile used (e.g., exclusive or inclusive median calculation), but the objective remains dividing the data into four equal parts.

For the provided sample data set, detailed statistical analysis, which involves locating the position of the 25th and 75th percentiles using standard methods, yields the following precise values. The value for the first quartile (Q1) is determined to be **5**. This signifies that 25% of the data falls below this specific point. Correspondingly, the third quartile (Q3) is calculated to be **20.75**, meaning that 75% of the observations are less than this value. It is important to remember that these quartile values form the core measurement of central tendency and spread that protects the calculation from extreme values.

The visual representation below illustrates the sorted data and the position of these key quartile boundaries. Accurately determining Q1 and Q3 is the cornerstone of the IQR method, as any inaccuracy here will directly propagate into errors when calculating the final outlier fences, potentially leading to misclassification of data points.

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 is calculated by subtracting Q1 from Q3:  $20.75 - 5 = 15.75$ .

### Step 3: Determining the Boundary Threshold Multiplier

With Q1, Q3, and the Interquartile Range (IQR) successfully identified (15.75), the next step is to prepare the threshold value that will define the outlier fences. This threshold is calculated by multiplying the IQR by the conventional constant, 1.5. This factor provides the distance that an observation must exceed Q1 or Q3 to be considered statistically anomalous, based on the internal spread of the data.

Plugging in the specific numbers for our sample data, we calculate the buffer amount: 1.5 multiplied by the IQR of 15.75 yields **23.625**. This value represents the critical distance that must be exceeded. Observations must fall more than 23.625 units away from the central quartiles. This threshold distance (23.625) will now be used consistently: it will be subtracted from Q1 to define the lower fence, and added to Q3 to define the upper fence. This crucial multiplication step defines the sensitivity of the outlier detection method.

### Step 4: Finding the Lower and Upper Limits (The Outlier Fences)

The final crucial step involves establishing the two numerical fences: the Lower Limit and the Upper Limit. These fences define the acceptable range of values within the data set, based on the central spread defined by the IQR. Any data point that falls outside of this established range is

mathematically categorized as an extreme value, warranting further scrutiny as a potential outlier.

The **Lower Limit** is calculated by taking the first quartile ( $Q1 = 5$ ) and subtracting the calculated  $1.5 \times IQR$  threshold (23.625). For our example, the calculation is:

$$\text{Lower Limit} = Q1 - 1.5 \times IQR = 5 - 23.625 = \mathbf{-18.625}$$

The **Upper Limit** is calculated by taking the third quartile ( $Q3 = 20.75$ ) and adding the  $1.5 \times IQR$  threshold (23.625). This defines the maximum acceptable non-outlier value for the upper tail of the distribution.

The upper limit is calculated as:

$$\text{Upper Limit} = Q3 + 1.5 \times IQR = 20.75 + 23.625 = \mathbf{44.375}$$

We now have our defined range: acceptable data points must lie between -18.625 and 44.375. Any observation outside this interval is objectively considered an outlier.

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

} **IQR = 20.75 - 5 = 15.75**

$$\text{Lower Limit} = 5 - 1.5 \times 15.75 = \mathbf{-18.625}$$

$$\text{Upper Limit} = 20.75 + 1.5 \times 15.75 = \mathbf{44.375}$$

## Step 5: Identifying the Final Outliers

With the Lower Limit (-18.625) and the Upper Limit (44.375) firmly established, the final step involves comparing every data point in the original data set against these two boundaries. We scan the observations to determine if any value is less than -18.625 or greater than 44.375. Observations that satisfy either of these conditions are definitively classified as outliers according to the 1.5 IQR rule.

Reviewing our sample data set, we observe that most values fall well within the acceptable range. For instance, the maximum non-outlier value is 44.375. Upon inspection of the original data points, we find the observation **46**. Since 46 is strictly greater than the Upper Limit of 44.375, it is identified as an outlier. Conversely, reviewing the lower end of the data, all observations are positive, far exceeding the Lower Limit of -18.625, confirming no outliers exist in the lower tail of this distribution.

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	
8	
11	
13	
14	
15	
17	
22	
24	
26	
46	<b>OUTLIER</b>

$Q1 = 5$   
 $Q3 = 20.75$   
 $IQR = 20.75 - 5 = 15.75$

$$\text{Lower Limit} = 5 - 1.5 \times 15.75 = -18.625$$

$$\text{Upper Limit} = 20.75 + 1.5 \times 15.75 = 44.375$$

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

## Practical Implementation in Statistical Software

While the manual steps outlined above provide a clear conceptual understanding, data analysis in professional and academic settings relies heavily on robust statistical software. These programs automate the complex sorting, quartile calculation, and boundary determination required by the Interquartile Range method, especially when dealing with thousands or millions of data points. Using software ensures computational speed and minimizes human error inherent in large-scale manual computations, which is crucial for maintaining data quality.

Most common statistical packages, such as R, Python (using Pandas or NumPy libraries), SPSS, and Excel, incorporate built-in functions to calculate quartiles (Q1, Q3) and the IQR automatically. For example, in R, functions like `quantile()` are used to derive Q1 and Q3, and simple subtraction yields the IQR. Analysts then programmatically apply the 1.5 IQR rule to filter the data, often generating box plots that visually represent the outlier detection process, displaying the fences and the flagged extreme values simultaneously. Leveraging these tools is essential for maintaining efficiency and reproducibility in data hygiene and preparation for advanced statistical modeling.

## Further Exploration: Software-Specific Tutorials

The following tutorials explain how to find outliers using the Interquartile Range in different statistical software, providing the specific code and syntax required for automated analysis: