

Find Outliers Using the Interquartile Range

Authored by
Mohammed loot

November 6, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Find Outliers Using the Interquartile Range*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=11364>

Maintaining the integrity of findings is a **fundamental goal** in all forms of [data analysis](#). Central to this effort is the accurate identification and careful handling of anomalous observations, commonly known as [outliers](#). An **outlier** is formally defined as an observation that resides an abnormal distance from other values within a given [dataset](#). While sometimes these values represent legitimate extreme events, they frequently indicate measurement errors or unusual variations that can severely distort fundamental statistical measures, such as the mean and [standard deviation](#), thereby compromising the reliability and validity of subsequent analytical conclusions.

The detection of these anomalous data points is a necessary prerequisite for developing **robust statistical models**. Although numerous detection methodologies exist, one of the most reliable, non-parametric, and widely adopted techniques is the use of the [Interquartile Range](#) (IQR) method, often credited to the renowned statistician [John Tukey](#). This technique is inherently resistant to extreme values because it relies solely on the median and the [quartiles](#) rather than the mean and standard deviation, making it exceptionally effective for identifying the very outliers that might otherwise skew the detection process itself.

Understanding the Interquartile Range Method

The **Interquartile Range** (IQR) serves as a powerful measure of statistical dispersion, specifically designed to quantify the spread of the central 50% of values within any data distribution. Its calculation is straightforward: it is the difference between the third **quartile** (Q3), which corresponds to the 75th percentile, and the first **quartile** (Q1), corresponding to the 25th percentile. Mathematically, the formula is expressed simply as $IQR = Q3 - Q1$. By focusing strictly on the middle half of the data, the IQR delivers a highly resilient assessment of central variability, effectively minimizing the distorting influence of extreme values.

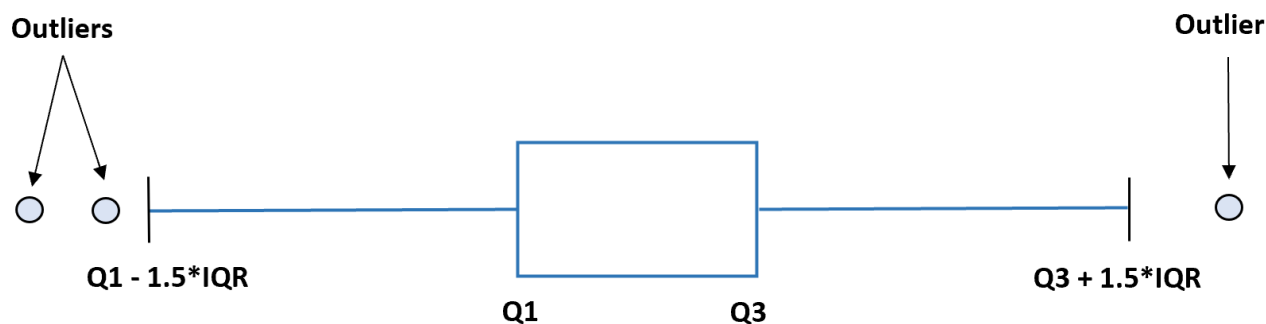
The IQR method utilizes a specific and crucial principle, known globally as the **1.5 × IQR rule**, to establish formal boundaries or "fences." Any observation falling outside these calculated fences is immediately flagged for review as a potential [outlier](#). These statistical fences define the acceptable range of values based on the data's inherent central spread. Observations that lie beyond these fences are considered statistically improbable under a standard distribution assumption and thus necessitate dedicated investigation to determine their cause and appropriate handling.

This established approach classifies an observation as an outlier if its value deviates significantly from the central body of the data. Specifically, a value is flagged if it is 1.5 times the IQR greater than the third quartile (Q3) or 1.5 times the IQR less than the first quartile (Q1). The formal numerical definitions for calculating these outlier limits are as follows:

Lower Limit (Lower Fence): $Q1 - (1.5 \times IQR)$

Upper Limit (Upper Fence): $Q3 + (1.5 \times IQR)$

The visual aid below provides a clear illustration of how the **IQR** defines the central spread and how the subsequent application of the $1.5 \times \text{IQR}$ factor expands this range to establish the precise lower and upper fences required for definitive outlier detection.



Step 1: Data Preparation and Sorting

To demonstrate the robust application of the IQR method, we will work through a step-by-step example using a sample [dataset](#) containing 13 observations. The initial and most critical preparatory step for accurate quartile calculation is to sort the entire data series in **ascending order**. This mandatory ordering ensures that the subsequent identification of the 25th and 75th percentiles (Q1 and Q3) is mathematically sound and statistically reliable.

The raw data points we are analyzing are presented below:

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

In practical, high-volume data environments, this essential sorting step is often performed automatically by [statistical software](#) packages like R or Python. However, when performing manual calculations or mastering the underlying methodology, confirming that the data is correctly arranged is paramount before proceeding to the next analytical stage.

Step 2: Calculating the Quartiles and the IQR

With the dataset now properly ordered, the next objective is to precisely locate the first [quartile](#) (Q1) and the third [quartile](#) (Q3). Q1 identifies the point below which exactly 25% of the data falls, while Q3 identifies the point below which 75% of the data falls. The median, sometimes referred to as the second quartile (Q2), naturally resides at the exact center of the distribution.

For our specific 13-observation dataset, standard statistical calculation methods yield the following quartile values: The first quartile (Q1) is determined to be **5**, and the third quartile (Q3) is calculated as **20.75**. It is important to note that the specific method employed for calculating quartiles (e.g., whether using an inclusive or exclusive median approach) can lead to minor variations based on the statistical platform being used, but these figures represent the typical and reliable outcome for this distribution.

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

Once Q1 and Q3 are established, we can calculate the [Interquartile Range](#) (IQR). This value, representing the span of the middle 50% of the data, is derived by subtraction:

$IQR = Q3 - Q1 = 20.75 - 5 = \mathbf{15.75}$. This figure, 15.75, quantifies the total spread of the central, most reliable half of our sample [dataset](#).

Step 3: Determining the Outlier Fences (Limits)

The next critical phase involves applying the $1.5 \times IQR$ rule to precisely calculate the numerical boundaries--the lower and upper fences--that demarcate the acceptable range of non-outlying observations. Observations that fall outside these statistically determined fences are definitively classified as potential anomalies that require further scrutiny.

First, we must calculate the necessary adjustment factor based on the [IQR](#): 1.5 multiplied by the calculated IQR: $1.5 \times 15.75 = \mathbf{23.625}$. This factor is the essential buffer zone; it is subtracted from Q1 to set the lower fence and added to Q3 to establish the upper fence.

The **lower limit**, or the lower fence, is determined by subtracting this adjustment factor from the first quartile (Q1):

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

Subsequently, the **upper limit**, or the upper fence, is calculated by adding the factor to the third

quartile (Q3):

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

Based on these calculations, the resultant acceptable range for non-outliers in our distribution is rigorously defined as the interval between -18.625 and 44.375. Any data point outside this span represents a statistical anomaly.

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

Q1 = 5

Q3 = 20.75

IQR = 20.75 - 5 = 15.75

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

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

Step 4: Final Identification of Outliers

With the lower limit established at -18.625 and the upper limit at 44.375, the final step involves systematically comparing every observation in the original [dataset](#) against these calculated boundaries. Any data point that is strictly less than the lower limit or strictly greater than the upper limit is formally designated as an [outlier](#) according to the rigorous IQR methodology.

Upon reviewing our sorted data points, it becomes clear that the vast majority of values fall comfortably within the calculated range. However, one specific observation violates the upper

boundary condition: the value **46**. Since 46 is unambiguously greater than the upper limit of 44.375, it is conclusively flagged as the sole high-end outlier in this specific sample dataset.

Furthermore, no observations in our sample fall below the calculated lower limit of -18.625. Therefore, the value 46 is the only data point requiring immediate statistical investigation, context-specific validation, or potential treatment (such as transformation or removal), depending entirely on the goals and requirements of the ongoing analysis.

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 on Automation: While mastering the manual calculation provides crucial clarity regarding the underlying methodology, practitioners handling large-scale production data almost universally rely on automated tools. Utilizing specialized web-based calculators or computational scripts within dedicated [statistical software](#) environments significantly streamlines the process of finding the upper and lower boundaries for outliers, ensuring efficiency and accuracy.

Practical Implementation and Conclusion

The [IQR method](#) offers a robust, non-parametric, and distribution-free approach to [outlier](#) detection, making it an indispensable tool across diverse domains, including financial risk

modeling, quality control monitoring, and rigorous scientific research. Once these statistical anomalies are successfully identified, they demand careful, contextual consideration. The subsequent decision regarding their treatment--whether through data transformation, winsorization, or complete removal--must be guided by expert domain knowledge and a clear understanding of the data generation process.

For analysts routinely working with massive volumes of data, implementing the $1.5 \times \text{IQR}$ rule is most efficiently accomplished within specialized programming or statistical environments. The following resources detail how to execute this precise outlier detection process using common statistical platforms: