How to Find Quartiles in Even and Odd Length Datasets

Authored by
stats writer

November 21, 2025

In the process of analyzing quantitative data, the calculation of quartiles is a crucial step for summarizing the distribution and variability of a dataset. To accurately determine Q1 (the first quartile) and Q3 (the third quartile), the data must first be meticulously sorted in ascending order, ranging from the smallest observation to the largest. For datasets with an **even** number of entries, the median is derived by averaging the two central values. Q1 is then defined as the median of the lower half of the data, and Q3 is the median of the upper half. Conversely, for datasets containing an **odd** number of entries, the median is the single middle value, which is then excluded when segmenting the data. Q1 remains the median of the remaining lower half, and Q3 is the median of the remaining upper half.

Introduction to Quartiles and Data Segmentation

In the field of descriptive statistics, Quartiles are fundamental measures used to divide a numerical dataset into four equal parts, each segment containing 25% of the values. These divisions are essential for understanding the spread and central tendency of the data, offering a robust measure of variability that is less sensitive to outliers than traditional metrics like the standard deviation. A complete set of quartiles includes the first quartile (Q1), the second quartile (Q2), and the third quartile (Q3). Q1 marks the 25th percentile, Q2 represents the 50th percentile (the median), and Q3 designates the 75th percentile. Analyzing these values provides statisticians and researchers with crucial insights into the distribution's shape, including skewness and concentration.

Calculating these quartiles requires slightly different methodologies depending on whether the total count of observations in the dataset is an **even** or an **odd** number. Regardless of the size, the initial and most critical step remains consistent: the data must first be sorted in ascending order, from the smallest value to the largest. This foundational arrangement ensures that the positional metrics--Q1, Q2, and Q3--are accurately determined based on their rank within the ordered sequence. Miscalculating quartiles due to unsorted data is one of the most common errors in introductory statistics, underscoring the importance of this preliminary step.

This comprehensive guide details the precise steps needed to identify Q1 and Q3, commonly known as the lower and upper quartiles, utilizing a standardized method that treats the calculation differently based on data parity. We will first locate Q2 (the median), which serves as the central demarcation point, effectively splitting the ordered data into a lower half and an upper half. The subsequent quartiles, Q1 and Q3, are then defined as the medians of those respective halves. This procedure ensures a clear and repeatable calculation, especially when working with discrete, finite datasets, which often present ambiguity depending on the chosen statistical formula.

Defining the Median (Q2) and Data Halves

The second quartile, or Q2, is synonymous with the median of the entire dataset. It is the central value that separates the lower 50% of the observations from the upper 50%. Establishing the median is paramount because it dictates how the remaining data is partitioned for the subsequent calculation of Q1 and Q3. The calculation of the median itself is the first point where the parity (even or odd length) of the data becomes relevant, thereby initiating the distinct calculation paths.

For a dataset containing an **even** number of observations (e.g., 10, 12, or 20 values), the median does not correspond to a single physical data point within the set. Instead, it is calculated as the average of the two central values. If a dataset has N values, the median is the average of the values located at the (N/2) position and the (N/2 + 1) position. Once this calculated median value is found, the lower half of the data consists of all observations preceding this calculated midpoint, and the upper half consists of all observations following it. Importantly, neither the calculated median nor the two values used to create it are included in the subsequent lower and upper subsets when determining Q1 and Q3 using this specific methodology.

When the dataset contains an **odd** number of observations (e.g., 9, 11, or 21 values), the median is simply the single, central value located precisely at the position. This value is physically present within the data sequence. Since this central value serves as the Q2 point, it is crucial to note that it must be **excluded** entirely when defining the lower and upper halves of the data used for Q1 and Q3 calculations. The lower half comprises all values strictly below the median, and the upper half comprises all values strictly above the median. This exclusion principle is characteristic of the calculation method used in many educational and statistical environments.

Methodology for Even Length Datasets

To find the first and third quartile for a dataset with an **even** number of values, we must first establish the central split point. Since the median (Q2) in this case is a calculated average--and therefore not a specific physical data point--the original data cleanly separates into two equal-sized subsets, which we will use to find the lower and upper quartiles.

The procedure for even length datasets is as follows:

Identify the median value (Q2): Calculate the average of the two middle values in the sorted data. This value is used only for defining the center point.
Split the dataset into two equal halves (Lower Half and Upper Half) based on the calculated median point. All original data points are included in either the lower or upper half.
Q1 is the median value found within the **Lower Half** of the dataset. This half is treated as a new, smaller dataset, and its median is calculated using the standard method (single middle value or average of two middle values, depending on the half's parity).
Q3 is the median value found within the **Upper Half** of the dataset, calculated identically to Q1 using the standard median rules applied to that specific sub-set of values.

It is important to emphasize that for even-length datasets, the resulting lower and upper halves will always be of equal size. If the halves themselves are odd-sized (e.g., N=14 yielding two halves of 7), Q1 and Q3 will be single data points. If the halves are even-sized (e.g., N=12 yielding two halves of 6), Q1 and Q3 will be the average of the two middle values of those respective sub-sets.

Methodology for Odd Length Datasets

The process for determining quartiles in an odd length dataset requires careful handling of the central data point. Since the median (Q2) is a specific observation within the dataset, including it in either the lower or upper half would result in unequal sub-sets and distort the quartile calculation. Therefore, the exclusion principle is mandatory for this method.

The standard steps for odd length datasets are clearly defined below:

Identify the median value (Q2): This is the single, physical middle value in the sorted data located at the (N+1)/2 position.
Split the dataset into the Lower Half and Upper Half, ensuring that the identified median (Q2) is **excluded** entirely from both halves.
Q1 is the median of the **Lower Half** of the dataset. Since the remaining halves will always have an equal number of values, Q1 is calculated using the standard median rules applied to this sub-set.
Q3 is the median of the **Upper Half** of the dataset, calculated identically to Q1.

For instance, if a dataset has 21 values, the median is the 11th value. Excluding this 11th value leaves 10 values in the lower half (1st through 10th) and 10 values in the upper half (12th through 21st). Both resulting halves are even-sized, meaning Q1 and Q3 will be determined by averaging the two central values in their respective sub-sets.

Example 1: Calculating Quartiles for Even Length Dataset

To illustrate the application of these rules, consider a concrete example involving an even-length dataset. This demonstration adheres strictly to the procedure outlined for sets where the median is determined by averaging the two central values.

Suppose we analyze the following sorted dataset, containing ten values (N=10):

Data: 3, 3, 6, 8, 10, 14, 16, 16, 19, 24

First, we determine the median (Q2). Since N=10, the median lies between the 5th value (10) and the 6th value (14). The median is calculated as the average of these two central observations: (10 + 14) / 2 = **12**.

Since the median (12) is a calculated midpoint and not a physical data point, the dataset is cleanly split into two halves of five values each. The subsequent calculation of Q1 and Q3 uses the standard median formula applied to these two halves:

Lower Half: 3, 3, 6, 8, 10 (N=5).
Upper Half: 14, 16, 16, 19, 24 (N=5).

The first quartile (Q1) is the median of the Lower Half. Since the lower half has 5 values (an odd number), its median is the single middle value (the 3rd value): Q1 = **6**.

The third quartile (Q3) is the median of the Upper Half. Since the upper half also has 5 values, its median is the single middle value (the 3rd value of the upper set): Q3 = **16**.

Thus, the first and third quartiles for this dataset are 6 and 16, respectively. This example clearly demonstrates how Q1 and Q3 are found by treating the sub-sets as independent datasets for median calculation.

Example 2: Calculating Quartiles for Odd Length Dataset

This example demonstrates the procedure required for an odd-length dataset, where the necessity of excluding the central median value is paramount to ensure the two resulting sub-sets are equal in size and maintain statistical balance.

Consider the following sorted dataset, containing nine values (N=9):

Data: 3, 3, 6, 8, 10, 14, 16, 16, 19

First, we identify the median (Q2). Since N=9, the median is the single value located at the = 5th position. The median is **10**.

Crucially, this median value (10) is now **excluded** when partitioning the data. The remaining values form two sub-sets of four values each (N=4):

Lower Half (values below 10): 3, 3, 6, 8 (4 values).
Upper Half (values above 10): 14, 16, 16, 19 (4 values).

To find the first quartile (Q1), we calculate the median of the Lower Half (3, 3, 6, 8). Since this half has an even number of values, Q1 is the average of the two middle terms (3 and 6): Q1 = (3 + 6) / 2 = **4.5**.

To find the third quartile (Q3), we calculate the median of the Upper Half (14, 16, 16, 19). Similarly, Q3 is the average of the two middle terms (16 and 16): Q3 = (16 + 16) / 2 = **16**.

Thus, the first and third quartiles for this odd-length dataset are 4.5 and 16, respectively. This demonstrates the necessity of the exclusion principle for odd-length datasets to achieve the correct Q1 and Q3 calculations.

Understanding Methodological Variations

When calculating quartiles, especially for discrete distributions, it is essential to understand that there is no single, universally agreed-upon formula. The method for calculating quartiles for discrete, finite datasets can vary based on the statistical software, textbook, or calculator being utilized. The choice of whether to include or exclude the median (Q2) when calculating the lower and upper halves is the primary difference between these established methods.

The approach detailed throughout this article--the method of **excluding the median** regardless of whether the dataset is even or odd--is the specific algorithm employed by popular educational tools like the **TI-84 calculators**. We have chosen to focus on this approach due to its prevalence in academic settings, ensuring that users can easily replicate these results using widely accessible technology. However, users should be aware that other methodologies exist, such as those used by certain statistical programming languages, which might yield slightly different quartile values for the same dataset.

Further Resources and Computational Tools

While mastering the manual calculation steps is vital for foundational understanding, large-scale data analysis necessitates the use of computational tools. The principles of sorting, median identification, and subsequent subset median calculation remain central even when utilizing advanced software.

The following tutorials explain how to find the quartiles of a dataset using different statistical software, providing resources for replicating these calculations in a programmatic environment: