

Understanding Quartiles: Calculation Using Mean and Standard Deviation

Authored by
Mohammed loot

November 2, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Understanding Quartiles: Calculation Using Mean and Standard Deviation*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=8535>

Understanding the spread and location of data within a set is a cornerstone of [statistics](#). While the [mean](#) (μ) provides the average central location and the [standard deviation](#) (σ) quantifies variance, [quartiles](#) offer a robust perspective on the distribution profile. When analyzing a dataset that adheres to a [normal distribution](#) (often referred to as the bell curve), calculating the first (Q_1) and third (Q_3) quartiles simplifies dramatically. This powerful technique leverages a fixed constant derived directly from the properties of the standard normal curve, allowing statisticians to quickly estimate the boundaries that contain the central 50% of the data using only μ and σ .

The following precise formulas are used to determine the first and third quartiles for any dataset confirmed to follow a **normal distribution**. This method is highly valued for its computational efficiency and accuracy under the Gaussian assumption.

$$Q_1 \text{ (25th Percentile)} = \mu - (0.675)\sigma$$

$$Q_3 \text{ (75th Percentile)} = \mu + (0.675)\sigma$$

It is essential to recall the definitions of the input variables: μ represents the population [mean](#) (the true average value of the population), and σ represents the population [standard deviation](#) (the average distance data points deviate from the mean). These formulas provide an elegant way to translate measures of central tendency and dispersion into measures of data segmentation.

The Statistical Significance of Quartiles and Percentiles

Quartiles serve as critical measures of location, segmenting a dataset into four equal parts. This concept is intrinsically linked to [percentiles](#), which define the values below which a specific percentage of observations fall. By definition, the **first quartile (Q_1)** is equivalent to the 25th percentile, meaning 25% of the data points lie below this value. Conversely, the **third quartile (Q_3)** corresponds to the 75th percentile, indicating that 75% of observations fall below it. The second quartile (Q_2) is the median, or the 50th percentile, marking the exact center of the distribution.

In a perfectly symmetrical distribution, such as the **normal distribution**, the [mean](#) (μ) and the median (Q_2) coincide. This inherent symmetry is the mathematical foundation that allows Q_1 and Q_3 to be equidistant from the central mean. They differ only by the sign applied to the standard deviation multiplier in the formula, reflecting their balanced positions on either side of the average.

This method of calculation is remarkably efficient because it distills the comprehensive task of data sorting and counting down to a simple algebraic operation. By requiring only the mean and the standard deviation, we can quickly and accurately locate the 25th and 75th [percentiles](#), providing a

rapid and informative summary of the data spread without needing access to every single raw data point.

Deriving the Constant: Why 0.675 is Used

The constant value of 0.675, which acts as the multiplier for the [standard deviation](#), is not an arbitrary figure. It is derived directly from the properties of the Standard Normal Distribution, often called the Z-distribution. Specifically, 0.675 represents the [Z-score](#) that corresponds precisely to the 25th and 75th percentiles of the standardized curve.

A [Z-score](#) quantifies how many standard deviations a particular data point lies above or below the mean. To locate the 25th percentile (Q_1), one consults a standard Z-table to find the Z-value where the cumulative probability (area under the curve) is 0.25. The exact theoretical value is approximately $Z = -0.6745$. Symmetrically, the 75th percentile (Q_3) corresponds to $Z = +0.6745$. For ease of calculation and general statistical practice, this figure is consistently rounded to 0.675.

The relationship between a raw score (X) and its standardized Z-score is defined by the formula $X = \mu + Z\sigma$. By substituting the quartile Z-scores into this equation, we precisely generate the formulas used above: $Q_1 = \mu + (-0.675)\sigma$ and $Q_3 = \mu + (+0.675)\sigma$. This direct mathematical connection confirms the validity and accuracy of this approximation specifically when applied to **normally distributed** data.

Example 1: Calculating Quartiles for Population Data

To illustrate this methodology, let us consider a practical scenario involving a large dataset, such as the heights of a population, which are known to follow a [normal distribution](#). Suppose the population [mean](#) (μ) is 300 units, and the population [standard deviation](#) (σ) is 45 units. Our objective is to determine the first and third [quartiles](#), which will define the range containing the central half of the population's heights.

We begin by substituting these known values into the established quartile formulas:

Calculate Q_1 : $Q_1 = \mu - (0.675)\sigma = 300 - (0.675) \text{ times } 45$

Calculate the Multiplier: $0.675 \text{ times } 45 = 30.375$

Q_1 Result: $300 - 30.375 = \mathbf{269.625}$

For the third quartile, we follow the same steps, adding the standard deviation multiplier to the mean:

Calculate Q_3 : $Q_3 = \mu + (0.675)\sigma = 300 + (0.675) \text{ times } 45$

Q₃ Result: $300 + 30.375 = \mathbf{330.375}$

The interpretation of these results is straightforward: 25% of the population's values fall below 269.625 units, and 75% of the values fall below 330.375 units. This provides a clear bracket of the most typical observations within the distribution.

Measuring Data Dispersion: The Interquartile Range (IQR)

Once Q₁ and Q₃ have been successfully calculated, we can determine a highly informative measure of dispersion known as the [Interquartile Range](#) (IQR). The IQR defines the span of the middle 50% of all data points, making it robust against extreme outliers that might skew the standard deviation. It is calculated by simply subtracting the first quartile from the third quartile.

Using the results from Example 1 (Q₁ = 269.625 and Q₃ = 330.375), we calculate the IQR:

IQR Formula: $IQR = Q_3 - Q_1$

Substitution: $IQR = 330.375 - 269.625$

IQR Result: $\mathbf{60.75}$

The resulting IQR of 60.75 units signifies the range within which the central half of the population's observations are concentrated. This measure is crucial not only for understanding data spread but also for identifying potential outliers in subsequent statistical analysis, such as constructing box plots.

Example 2: Analyzing a Tighter Distribution and its IQR

In our second example, we explore a dataset characterized by significantly lower variance, which results in a tighter clustering of data points around the mean. Consider a set of test scores confirmed to be normally distributed, where the [mean](#) (μ) is 50 and the [standard deviation](#) (σ) is 2. The small σ value immediately suggests that the quartiles will be very close to the mean.

We proceed with the calculations for the [quartiles](#) using the constant 0.675:

Q₁ Calculation: $Q_1 = \mu - (0.675)\sigma = 50 - (0.675) \text{ times } 2 = 50 - 1.35 = \mathbf{48.65}$

Q₃ Calculation: $Q_3 = \mu + (0.675)\sigma = 50 + (0.675) \text{ times } 2 = 50 + 1.35 = \mathbf{51.35}$

These results reveal that 50% of the test scores fall within the narrow window between 48.65 and 51.35. This narrow range is a direct mathematical consequence of the small standard deviation, confirming the tight clustering of the data.

Calculating the [Interquartile Range](#) (IQR) further emphasizes the minimal variance in this dataset:

$$\text{IQR} = Q_3 - Q_1$$

$$\text{IQR} = 51.35 - 48.65$$

$$\text{IQR} = \mathbf{2.7}$$

An IQR of only 2.7 confirms that the middle half of the test scores spans less than three units, providing strong evidence that the scores are highly consistent and clustered closely around the average of 50.

Critical Limitations and Assumptions of the Method

While highly efficient, this method is entirely dependent upon one fundamental assumption: that the underlying data is **normally distributed**. If the data exhibits significant skewness, is multimodal, or possesses heavy tails (i.e., it is not Gaussian), these formulas will yield estimates that are statistically inaccurate. In such cases, relying on the mean and standard deviation to estimate quartiles is inappropriate.

For data that does not conform to the normal distribution, it is imperative to employ nonparametric methods. This typically involves sorting the entire dataset and directly identifying the values corresponding to the 25th and 75th [percentiles](#). Furthermore, it must be acknowledged that the constant 0.675 is a rounded approximation of the theoretical [Z-score](#) (0.6745). Although this rounding has negligible impact on most statistical applications, the precision of the final quartile calculation is ultimately contingent upon the accuracy of the estimated population [mean](#) and standard deviation. Nonetheless, for voluminous datasets confirmed to be Gaussian, this method remains the most straightforward and powerful computational shortcut.