

Learning Guide: Calculating Quartiles for Grouped Data Sets

Authored by
Mohammed loot

October 28, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *Learning Guide: Calculating Quartiles for Grouped Data Sets*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=4684>

Introduction to Quartiles and Grouped Data

In the expansive field of [statistics](#), achieving a comprehensive understanding of data distribution requires more than just measures of central tendency, such as the mean or median. [Quartiles](#) are powerful positional measures that reveal the internal structure and spread of a dataset by dividing it into four segments, each containing 25% of the observations. By pinpointing the 25th, 50th, and 75th percentiles, analysts gain crucial insight into data variability, potential skewness, and the concentration of values across the entire distribution.

Calculating quartiles for raw, ungrouped data is typically straightforward. However, practical data analysis often involves working with large datasets summarized as [grouped data](#). This involves organizing observations into predefined class intervals along with their corresponding frequencies. While this aggregation simplifies visualization and reporting, it introduces a challenge: the exact values of individual data points are lost. Therefore, specialized statistical methods are essential to estimate measures of position accurately when working with frequency distributions.

This comprehensive guide is designed to clarify the methodology for calculating quartiles specifically for [grouped data](#). We will meticulously break down the required estimation formula, define each component variable, and demonstrate the process through a detailed, step-by-step practical example. Mastery of this technique is fundamental for enhancing your quantitative analysis skills and deriving meaningful conclusions from aggregated information.

The Estimation Formula for Quartiles in Grouped Data

Since we cannot identify the precise location of every observation within a frequency distribution, we must utilize a formula that estimates the quartile value based on its position relative to the class boundaries. This formula is adapted from the calculation of the median for grouped data, ensuring that the estimated quartile accurately reflects its expected location within the designated interval. The standard formula for the i th quartile is:

$$Q_i = L + (C/F) * (iN/4 - M)$$

A thorough understanding of each element in this formula is critical for successful calculation. These variables collectively allow us to interpolate the precise value of the quartile within its identified class interval:

L: Represents the [lower bound](#) of the class interval containing the i th quartile. This boundary marks the starting value of the class where the quartile is estimated to lie.

C: Denotes the [class width](#) of the quartile class. This is the difference between the upper and lower boundaries of that specific interval.

F: Refers to the **frequency** of the quartile class. This count indicates the number of observations

contained within the interval that houses the desired quartile.

N: Is the **total frequency** of the entire dataset. It is calculated as the sum of all frequencies across all class intervals, representing the total number of observations analyzed.

M: Stands for the **cumulative frequency** of the class interval immediately preceding the quartile class. This value accounts for all observations that fall below the lower boundary (L) of the quartile class.

i: Indicates the specific quartile being calculated: 1 for the first quartile (Q1), 2 for the second quartile (Q2, or the median), and 3 for the third quartile (Q3).

Systematic Guide to Calculating Quartiles

Determining quartiles for [grouped data](#) is a multi-step process that demands accuracy and sequential execution. Following these steps ensures that the calculation correctly applies the estimation formula:

Construct the Cumulative Frequency Distribution Table: The initial step requires transforming the raw frequency data into a complete [frequency distribution](#) table. A mandatory addition to this table is a column dedicated to [cumulative frequency](#). This column is essential for rapidly locating the interval in which the quartile resides.

Determine the Quartile Position: Calculate the theoretical position of the desired quartile using the formula $(iN/4)$. For Q1, use $(1N/4)$; for Q2, use $(2N/4)$ or $(N/2)$; and for Q3, use $(3N/4)$. This result identifies the specific observation number that corresponds to the quartile if the data were arranged linearly.

Identify the Quartile Class Interval: Locate the "quartile class" by referencing the cumulative frequency column. The quartile class is the first class interval whose [cumulative frequency](#) is greater than or equal to the calculated quartile position $(iN/4)$. This interval contains the actual estimated value of the quartile.

Extract Formula Components: Once the quartile class is definitively identified, extract the five crucial parameters required by the formula from the table:

L: The [lower bound](#) of the identified quartile class.

C: The [class width](#) of the quartile class.

F: The **frequency** of the quartile class.

N: The **total frequency** of the dataset.

M: The **cumulative frequency** of the class interval immediately preceding the quartile class.

Apply the Quartile Formula: Substitute all extracted numerical values (L, C, F, N, M, and i) into the formula: $Q_i = L + (C/F) * (iN/4 - M)$.

Compute and Interpret the Result: Execute the arithmetic calculations to arrive at the final quartile value. Finally, state what this numerical result signifies in the context of the original

dataset, providing a clear interpretation of the data distribution.

Practical Example: Determining the Third Quartile (Q3)

To solidify the understanding of this process, let us apply the steps to a practical scenario. Consider the following [frequency distribution](#) table, which summarizes the scores achieved by students on a recent examination:

Class Interval	Frequency	Cumulative Frequency
1-5	6	6
6-10	19	25
11-15	13	38
16-20	20	58
21-25	12	70
26-30	11	81
31-35	6	87
36-40	5	92

Our objective is to calculate the **third quartile (Q3)**. This measure will define the score below which 75% of the students fall. First, we must ascertain the position of the third quartile within the dataset. Given that the total frequency (N) is 92, we use the position formula with $i=3$:

$$\text{Quartile Position (Q3)} = (3 * N) / 4 = (3 * 92) / 4 = 69.$$

The 69th observation, if the data were ranked, would be the third quartile. We now use the extended frequency table, which includes the cumulative frequency column, to locate the class containing this 69th score:

Class Interval	Frequency	Cumulative Frequency
1-5	6	6
6-10	19	25
11-15	13	38
16-20	20	58
21-25	12	70
26-30	11	81
31-35	6	87
36-40	5	92

} 69 is between 58 and 70

Upon reviewing the table, the cumulative frequency reaches 58 at the end of the 16-20 interval. This indicates that 58 scores are 20 or less. The subsequent cumulative frequency is 70, corresponding to the 21-25 interval. Since our desired position (69) falls between 58 and 70, the **21-25 interval** is definitively identified as the quartile class. This crucial step narrows the search for Q3 to this specific range of scores.

Extracting Values and Performing the Calculation

Having successfully identified the quartile class as 21-25, we proceed to extract the required variables for interpolation:

L (Lower bound of the quartile class): The [lower bound](#) of the 21-25 interval is **21**.

C (Class width of the quartile class): The [class width](#) (Upper Limit - Lower Limit) is $25 - 21 = 4$.

F (Frequency of the quartile class): The frequency for the 21-25 class is **12**.

N (Total frequency): The total number of students is **92**.

M (Cumulative frequency of the preceding class): The [cumulative frequency](#) of the class immediately before 21-25 (i.e., 16-20) is **58**.

We now substitute these extracted values into the quartile formula to calculate Q3:

$$Q_i = L + (C/F) * (iN/4 - M)$$

$$Q_3 = 21 + (4/12) * ((3 * 92) / 4 - 58)$$

$$Q_3 = 21 + (0.333333...) * (69 - 58)$$

$$Q_3 = 21 + (0.333333...) * (11)$$

$$Q_3 = 21 + 3.666663...$$

$$Q_3 = 24.666663...$$

Rounding the result to two decimal places, the calculated value for the **third quartile (Q3)** is **approximately 24.67**. This outcome signifies that 75% of the students achieved a score of 24.67 or less, while the top 25% of students scored above this value. This score provides a critical benchmark for evaluating the upper performance range of the student group.

It is important to note that this same methodical approach is used for calculating the **first quartile (Q1)** by setting $i=1$, and the **second quartile (Q2)** (the median) by setting $i=2$. The only variable that changes in the formula is the required quartile position ($iN/4$), making the formula robust and highly adaptable across all quartile calculations.

Applications and Significance of Quartiles in Analysis

While the calculation process for [quartiles](#) in grouped data is complex, the resulting values are essential [measures of position](#) in [statistics](#). They offer far more information than simple averages, providing deep insights into data distribution, variability, and quality control.

Understanding Data Spread: Quartiles effectively delineate the central 50% of the data (the range between Q1 and Q3). This reveals how tightly or loosely the majority of observations are clustered around the median, aiding in the assessment of data homogeneity.

Identifying Distribution Skewness: Quartiles help analysts quickly visualize the shape of the distribution. If the distance between the median (Q2) and Q3 is significantly larger than the distance between Q1 and Q2, the distribution is likely positively skewed (skewed to the right), and vice versa.

Constructing [Box Plots](#): Quartiles are the foundational elements for creating box plots (or box-and-whisker plots). These plots graphically summarize the dataset's five-number summary: minimum, Q1, Q2 (median), Q3, and maximum. The box itself spans the interquartile range, offering an immediate visual representation of central variability.

Calculating the [Interquartile Range \(IQR\)](#): The IQR, calculated as Q3 minus Q1, is a robust measure of statistical dispersion. Unlike standard deviation, the IQR is resistant to the influence of extreme outliers, providing a reliable measure of variability for the middle half of the data.

Benchmarking and Comparison: In professional contexts, quartiles are frequently used for benchmarking. For instance, in financial performance analysis or quality control, metrics might be categorized by quartile to quickly determine if a specific data point falls within the top 25% or the bottom 25% relative to its peers.

Mastering the accurate calculation of quartiles, particularly for [grouped data](#), equips the data analyst with powerful tools for descriptive statistics, leading to more nuanced interpretations and better-informed strategic decisions.

Additional Resources for Grouped Data Analysis

To further advance your proficiency in working with frequency distributions and related statistical concepts, we recommend exploring these resources focusing on other measures of central tendency and position for grouped data:

[Calculating the Mean for Grouped Data](#)

[Finding the Median for Grouped Data](#)

[Determining the Mode for Grouped Data](#)

[Understanding Percentiles for Grouped Data](#)