Understanding Percentiles, Quartiles, and Quantiles: A Guide to Data Division

Authored by

Mohammed looti

November 4, 2025

Understanding Quantiles: The Foundation of Data Division

In the rigorous field of statistics, the structured division of data is a fundamental technique employed to analyze distributions, measure variability, and identify critical data points. Analysts frequently encounter three interrelated terms: **percentiles**, **quartiles**, and **quantiles**. Although these terms are often used interchangeably by novices, they possess a hierarchical relationship; the quantile serves as the generalized, overarching category from which the others are derived.

A **quantile** is formally defined as a point that partitions the range of a probability distribution into continuous intervals, ensuring that the probability within each interval is equal, or similarly divides the observations in a sample. Quantiles are essential because they generalize standard positional measures, encompassing the concepts of the median, quartiles, and percentiles. The primary function of any quantile is to precisely locate a specific value relative to the rest of the dataset, providing immediate context regarding its positional significance within the overall distribution. This is a core component of descriptive statistics.

To clarify the scope and specificity of these terms, we can outline their distinct division properties:

Quantiles: These represent the most comprehensive measure, capable of dividing a dataset into any specified number of equal parts, defined by the variable 'q'. They provide maximal flexibility in data segmentation.
Percentiles: These are specialized quantiles that divide the data into exactly 100 equal parts, indexing from 1 to 100.
Quartiles: These are a specific, standardized type of quantile, designed to divide the data into four equal parts, ranging from 1 to 4.

Ultimately, understanding positional statistics requires acknowledging that both percentiles and quartiles are merely standardized manifestations, or specific *types*, of the generalized quantile measure, each serving a unique yet predictable role in analyzing how data is spread.

Percentiles: Granular Division of Data

A percentile, often referred to as a 100-quantile, specifies a value below which a given percentage of observations in a group falls. Because percentiles segment the dataset into 100 distinct parts, they offer the most granular view of a data distribution possible using standard quantile measures. For example, stating that a score falls in the 90th percentile means that the score exceeds 90% of all other scores recorded in that particular distribution.

The primary utility of percentiles is found in contexts requiring precise benchmarking, standardization, and comparative ranking. They are indispensable when dealing with exceptionally large datasets or standardized measurement systems, such as standardized educational testing

(SAT scores), tracking health metrics (like pediatric growth charts), or detailed economic reporting. Since they are indexed from 1 to 100, percentiles provide an immediate and intuitively interpretable measure of an observation's relative standing within the larger population.

When leveraging percentiles, analysts are typically focused on identifying a specific numerical cutoff point--the exact value corresponding to a desired percentage threshold. This fine-tuning capability is crucial for establishing precise eligibility criteria or performance benchmarks, such as determining the value associated with the top 5% of earners, or defining the threshold for the bottom 10% of patient blood pressure readings. Percentiles allow for targeted analysis beyond the fixed 25% increments offered by quartiles.

Quartiles: Focusing on Four Key Segments

In direct contrast to the fine-grained precision of percentiles, quartiles divide a dataset into four large, equal parts, or quarters. These divisions are defined by three crucial data points: the first quartile (Q1), the second quartile (Q2), and the third quartile (Q3). These three points, combined with the minimum and maximum values of the dataset, constitute the five-number summary--a foundational tool used in exploratory data analysis and visualization, most notably through box plots.

Each quartile point serves a specific function. The first quartile (Q1) marks the boundary of the lowest 25% of the data, meaning 25% of observations fall at or below this value. The second quartile (Q2) represents the 50th percentile and is mathematically synonymous with the median--the exact midpoint of the dataset. Finally, the third quartile (Q3) delineates the 75th percentile, separating the highest 25% of values from the remaining 75% of the distribution.

The most frequent application of quartiles is the calculation of the Interquartile Range (IQR). The IQR is computed by taking the difference between the third and first quartiles (Q3 - Q1). This calculation yields a highly robust measure of statistical dispersion, as it captures the spread of the central 50% of the data. Because the IQR ignores the extreme tails of the distribution, it is significantly less sensitive to the distorting effects of outliers compared to measures like the standard range.

The Fundamental Relationship Between Percentiles and Quartiles

The key to fully mastering these positional statistics is recognizing that quartiles are simply fixed, standardized markers within the percentile system. Since quartiles inherently divide the distribution into four segments of 25%, their boundaries align perfectly with specific percentile values. Grasping this direct correspondence streamlines statistical interpretation and ensures consistent results regardless of whether the analysis is framed in terms of quartiles or percentiles.

The following list explicitly details the overlap between these two measures, confirming that quartiles provide universally understood reference points within the more flexible percentile framework:

The 0th percentile is equivalent to the 0th quartile, commonly identified as the **minimum** value of the dataset.
The 25th percentile is equivalent to the 1st quartile (Q1).
The 50th percentile is equivalent to the 2nd quartile (Q2), which is also the **median**.
The 75th percentile is equivalent to the 3rd quartile (Q3).
The 100th percentile is equivalent to the 4th quartile (Q4), commonly identified as the **maximum** value of the dataset.

This established relationship confirms that while quartiles serve as a simplified, standardized means of reporting key positional statistics, percentiles offer the necessary flexibility to create customized and precise divisions of the data distribution for tailored analytical needs.

Practical Application: Calculating and Interpreting Quantiles

To illustrate the practical application of these concepts, let us consider a hypothetical dataset. The following list consists of 20 numerical values, which have been pre-ordered from the smallest observation to the largest:

Data
3
4
4
6
7
9
12
13
14
16
17
19
22
23
23
25
28
29
34
37

Attempting to calculate percentiles and quartiles manually, particularly with large datasets, can be cumbersome and error-prone due to the various complex interpolation methods employed by statisticians. Consequently, data professionals overwhelmingly rely on robust statistical software packages, such as Microsoft Excel, the R programming language, or Python libraries, to accurately and efficiently determine these crucial values.

After inputting this sample data into a statistical tool, the resulting key quantile measures for this specific distribution are calculated as follows:

Data
3
4
4
6
7
9
12
13
14
16
17
19
22
23
23
25
28
29
34
37

Percentile	Quartile	Value
0	0	3
25	1	8.5
50	2	16.5
75	3	23.5
100	4	37

Based on these computed results, we can interpret the positional statistics of the dataset:

The 0th percentile and 0th quartile (Minimum) is **3**.
The 25th percentile and 1st quartile (Q1) is **8.5**. This means 25% of all values in the dataset are 8.5 or lower.
The 50th percentile and 2nd quartile (Median, Q2) is **16.5**. This represents the central point of the distribution.
The 75th percentile and 3rd quartile (Q3) is **23.5**. Only 25% of the observed values are greater than 23.5.
The 100th percentile and 4th quartile (Maximum) is **37**.

When to Choose Percentiles Versus Quartiles

The strategic decision of whether to employ percentiles or quartiles hinges entirely on the required level of detail for the analysis and the specific question the researcher is attempting to answer. Generally, quartiles are the preferred tool for generating rapid summaries and standard measures of data spread (such as the IQR), whereas percentiles are essential for defining specific performance benchmarks or establishing arbitrary, non-standardized cutoff points.

Percentiles are optimally utilized when the research demands a precise boundary or when the required relative position within the distribution does not align with the fixed 25% increments. Consider these applications:

Determining a high-performance cutoff: If the goal is to identify the score a student must achieve to be ranked in the **top 10%** of scores, the analyst must calculate the 90th percentile of the distribution. This value is the exact cutoff that separates the bottom 90% of scores from the top 10%.
Defining an arbitrary central range: If we need to find the specific range of weights that encompasses the **middle 40%** of employees at a company, we must calculate two specific percentiles: the 70th percentile (the upper boundary) and the 30th percentile (the lower boundary).

Quartiles, conversely, are the superior choice when the primary objective is to quickly summarize the data distribution, focusing particularly on central tendency and standardized measures of dispersion. Key application examples include:

Identifying the top quarter: To determine the score required for a student to be ranked in the **top quarter** (top 25%) of test scores, the analyst simply needs to find the 3rd quartile (Q3). This single point serves as the boundary separating the top 25% from the bottom 75%.
Measuring variability: The most common application is calculating the **interquartile range (IQR)** of a dataset. This measure of spread is directly and simply calculated using the difference between the 3rd quartile and the 1st quartile.

Other Specialized Quantiles for Distribution Analysis

As established, quartiles and percentiles are specialized instances of the overarching term, quantiles. Depending on the level of detail necessary for the analysis, quantiles adopt specific, named identities based on the number of equal divisions they create (known as k-quantiles). These specialized quantiles are frequently employed in specific disciplines, such as economics, demography, and sociology, where categorizing populations or financial distributions into equal segments is essential for equitable comparison.

The following list details common examples of specialized quantiles, based on the number of equal divisions (k) they produce:

4-quantiles are universally called **quartiles**.
5-quantiles are called **quintiles**, which are frequently used in income or wealth studies to divide populations into fifths for comparative analysis.

8-quantiles are called **octiles**.
10-quantiles are called **deciles**, commonly used in market analysis or ranking systems to separate data into ten 10% segments.
100-quantiles are called **percentiles**.

While quartiles and percentiles remain the most frequent measures encountered in general statistical practice, understanding the full spectrum of quantiles reinforces the conceptual framework of dividing data based on equal probability or observation counts, providing comprehensive flexibility for diverse analytical requirements.

Additional Resources for Further Study

For those interested in delving deeper into positional statistics, descriptive measures, and the mathematical methods behind data distribution, the following resources provide comprehensive coverage.