

Quartile

A **quartile** is a type of quantile which divides the number of data points into four more or less equal parts, or quarters. The first quartile (Q_1) is defined as the middle number between the smallest number and the median of the data set. It is also known as the lower quartile or the 25th empirical quartile and it marks where 25% of the data is below or to the left of it (if data is ordered on a timeline from

smallest to largest). The second quartile (Q_2) is the median of a data set and 50% of the data lies below this point. The third quartile (Q_3) is the middle value between the median and the highest value of the data set. It is also known as the upper quartile or the 75th empirical quartile and 75% of the data lies below this point.^[1]

Due to the fact that the data needs to be ordered from smallest to largest to compute quartiles, quartiles are a form of Order statistic.

Along with the minimum and the maximum of the data, which are also quartiles, the three quartiles described

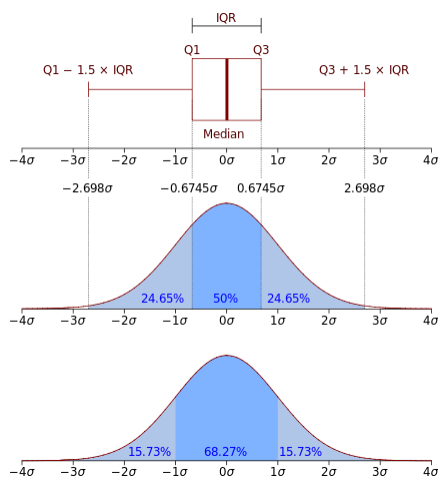
above provide a five-number summary of the data. This summary is important in statistics because it provides information about both the center and the spread of the data. Knowing the lower and upper quartile provides information on how big the spread is and if the dataset is skewed toward one side. Since quartiles divide the number of data points evenly, the range is not the same between quartiles (i.e., $Q_3 - Q_2 \neq Q_2 - Q_1$). While the maximum and minimum also show the spread of the data, the upper and lower quartiles can provide more detailed information on the location of specific data points, the presence of outliers in the data, and the

difference in spread between the middle 50% of the data and the outer data points.^[2]

For a visual representation see:

<https://www.mathsisfun.com/data/quartiles.html>

Definitions



Boxplot (with quartiles and an interquartile range) and a probability density function (pdf) of a normal

$N(0, 1\sigma^2)$ population

Symbol	Names	Definition
Q_1	first quartile lower quartile 25th percentile	splits off the lowest 25% of data from the highest 75%
Q_2	second quartile <u>median</u> 50th percentile	cuts data set in half
Q_3	third quartile upper quartile 75th percentile	splits off the highest 25% of data from the lowest 75%

Computing methods

Discrete Distributions

...

For discrete distributions, there is no universal agreement on selecting the quartile values. [3]

Method 1

...

1. Use the median to divide the ordered data set into two-halves.
 - If there is an odd number of data points in the original ordered data set, **do not include** the median (the central value in the ordered list) in either half.
 - If there is an even number of data points in the original ordered data set, split this data set exactly in half.
2. The lower quartile value is the median of the lower half of the data.

The upper quartile value is the median of the upper half of the data.

This rule is employed by the TI-83 calculator boxplot and "1-Var Stats" functions.

Method 2

...

1. Use the median to divide the ordered data set into two-halves.
 - If there are an odd number of data points in the original ordered data set, **include** the median (the central value in the ordered list) in both halves.

- If there are an even number of data points in the original ordered data set, split this data set exactly in half.
2. The lower quartile value is the median of the lower half of the data. The upper quartile value is the median of the upper half of the data.

The values found by this method are also known as "Tukey's hinges";^[4] see also midhinge.

Method 3

...

1. If there are even numbers of data points, then Method 3 is the same as

either method above

2. If there are $(4n+1)$ data points, then the lower quartile is 25% of the n th data value plus 75% of the $(n+1)$ th data value; the upper quartile is 75% of the $(3n+1)$ th data point plus 25% of the $(3n+2)$ th data point.
3. If there are $(4n+3)$ data points, then the lower quartile is 75% of the $(n+1)$ th data value plus 25% of the $(n+2)$ th data value; the upper quartile is 25% of the $(3n+2)$ th data point plus 75% of the $(3n+3)$ th data point.

Method 4

...

If we have an ordered dataset $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, we can interpolate between data points to find the p th empirical quantile if \mathbf{x}_i is in the $i/(n + 1)$ quantile. If we denote the integer part of a number a by $[a]$, then the empirical quantile function is given by,

$$q(p) = \mathbf{x}_{(k)} + \alpha(\mathbf{x}_{(k+1)} - \mathbf{x}_{(k)}),$$

where $k = [p(n + 1)]$ and

$$\alpha = p(n + 1) - [p(n + 1)].^{[1]}$$

To find the first, second, and third quartiles of the dataset we would evaluate $q(0.25)$, $q(0.5)$, and $q(0.75)$ respectively.

Example 1

...

Ordered Data Set: 6, 7, 15, 36, 39, 40, 41, 42, 43, 47, 49

	Method 1	Method 2	Method 3	Method 4
Q_1	15	25.5	20.25	15
Q_2	40	40	40	40
Q_3	43	42.5	42.75	43

Example 2

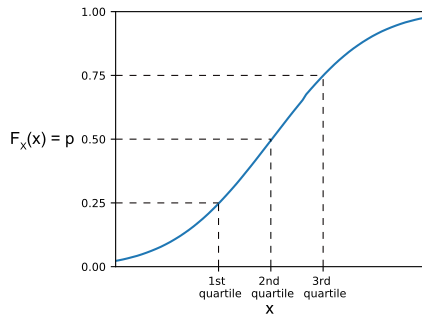
...

Ordered Data Set: 7, 15, 36, 39, 40, 41

As there are an even number of data points, the first three methods all give the same results.

	Method 1	Method 2	Method 3	Method 4
Q_1	15	15	15	13
Q_2	37.5	37.5	37.5	37.5
Q_3	40	40	40	40.25

Continuous Probability Distributions ...



Quartiles on a cumulative distribution function of a normal distribution

If we define a continuous probability distributions as $P(X)$ where X is a real valued random variable, its cumulative distribution function (CDF) is given by,

$$F_X(x) = P(X \leq x).^{[1]}$$

The CDF gives the probability that the random variable X is less than the value x . Therefore, the first quartile is the value of x when $F_X(x) = 0.25$, the second quartile is x when $F_X(x) = 0.5$, and the third quartile is x when $F_X(x) = 0.75$.^[5] The values of x can be found with the quantile function $Q(p)$ where $p = 0.25$ for the first quartile, $p = 0.5$ for the second quartile, and $p = 0.75$ for the third quartile. The quantile function is the inverse of the cumulative distribution function if the cumulative distribution function is monotonically increasing.

Outliers

There are methods by which to check for outliers in the discipline of statistics and statistical analysis. Outliers could be a result from a shift in the location (mean) or in the scale (variability) of the process of interest.^[6] Outliers could also may be evidence of a sample population that has a non-normal distribution or of a contaminated population data set.

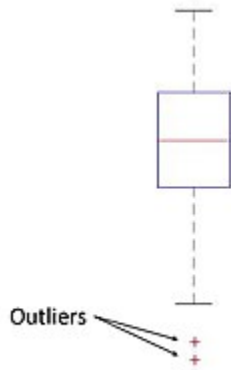
Consequently, as is the basic idea of descriptive statistics, when encountering an outlier, we have to explain this value by further analysis of the cause or origin of the outlier. In cases of extreme observations, which are not an infrequent occurrence, the typical values must be

analyzed. In the case of quartiles, the Interquartile Range (IQR) may be used to characterize the data when there may be extremities that skew the data; the interquartile range is a relatively robust statistic (also sometimes called "resistance") compared to the range and standard deviation. There is also a mathematical method to check for outliers and determining "fences", upper and lower limits from which to check for outliers.

After determining the first and third quartiles and the interquartile range as outlined above, then fences are calculated using the following formula:

$$\text{Lower fence} = Q_1 - 1.5(\text{IQR})$$

$$\text{Upper fence} = Q_3 + 1.5(\text{IQR}),$$



Boxplot Diagram with Outliers

where Q_1 and Q_3 are the first and third quartiles, respectively. The lower fence is the "lower limit" and the upper fence is the "upper limit" of data, and any data lying outside these defined bounds can be considered an outlier. Anything below the Lower fence or above the Upper fence can

be considered such a case. The fences provide a guideline by which to define an outlier, which may be defined in other ways. The fences define a "range" outside which an outlier exists; a way to picture this is a boundary of a fence, outside which are "outsiders" as opposed to outliers. It is common for the lower and upper fences along with the outliers to be represented by a boxplot. For a boxplot, only the vertical heights correspond to the visualized data set while horizontal width of the box is irrelevant. Outliers located outside the fences in a boxplot can be marked as any choice of symbol, such as an "x" or "o". The fences are sometimes

also referred to as "whiskers" while the entire plot visual is called a "box-and-whisker" plot.

When spotting an outlier in the data set by calculating the interquartile ranges and boxplot features, it might be simple to mistakenly view it as evidence that the population is non-normal or that the sample is contaminated. However, this method should not take place of a hypothesis test for determining normality of the population. The significance of the outliers vary depending on the sample size. If the sample is small, then it is more probable to get interquartile ranges that

are unrepresentatively small, leading to narrower fences. Therefore, it would be more likely to find data that are marked as outliers.^[7]

Computer Software for Quartiles

Excel:

The Excel function *QUARTILE(array, quart)* provides the desired quartile value for a given array of data. In the *Quartile* function, array is the dataset of numbers that is being analyzed and quart is any of the following 5 values depending on which quartile is being calculated. ^[8]

Quart	Output QUARTILE Value
0	Minimum value
1	Lower Quartile (25th percentile)
2	Median
3	Upper Quartile (75th percentile)
4	Maximum value

MATLAB:

In order to calculate quartiles in Matlab, the function $quantile(A,p)$ can be used.

Where A is the vector of data being analyzed and p is the percentage that relates to the quartiles as stated below. [9]

p	Output QUARTILE Value
0	Minimum value
0.25	Lower Quartile (25th percentile)
0.5	Median
0.75	Upper Quartile (75th percentile)
1	Maximum value

See also

- Five-number summary
- Range
- Box plot
- Interquartile range
- Summary statistics
- Quantile

References

1. *A modern introduction to probability and statistics : understanding why and how* . Dekking, Michel, 1946–. London: Springer. 2005. pp. 234 –238.
ISBN 978-1-85233-896-1.
OCLC 262680588 .

2. *Knoch, Jessica (February 23, 2018). "How are Quartiles Used in Statistics?" . Magoosh Statistics Blog. Retrieved December 11, 2019.*
3. *Hyndman, Rob J; Fan, Yanan (November 1996). "Sample quantiles in statistical packages" . American Statistician. **50** (4): 361–365. doi:10.2307/2684934 . JSTOR 2684934 .*
4. *Tukey, John Wilder (1977). Exploratory Data Analysis . ISBN 978-0-201-07616-5.*
5. *"6. Distribution and Quantile Functions" (PDF). math.bme.hu.*

6. *Walfish, Steven (November 2006). "A Review of Statistical Outlier Method" . Pharmaceutical Technology.*
7. *Dawson, Robert (July 1, 2011). "How Significant is a Boxplot Outlier?". Journal of Statistics Education. **19** (2): null.
doi:10.1080/10691898.2011.11889610 .*
8. *"How to use the Excel QUARTILE function | Exceljet" . exceljet.net.
Retrieved December 11, 2019.*
9. *"Quantiles of a data set – MATLAB quantile" . www.mathworks.com.
Retrieved December 11, 2019.*

External links

- [Quartile](#) – from MathWorld Includes references and compares various methods to compute quartiles
- [Quartiles](#) – From MathForum.org
- [Quartiles calculator](#) – simple quartiles calculator
- [Quartiles](#) – An example how to calculate it

Retrieved from

["https://en.wikipedia.org/w/index.php?title=Quartile&oldid=960305761"](https://en.wikipedia.org/w/index.php?title=Quartile&oldid=960305761)

Content is available under CC BY-SA 3.0 unless otherwise noted.