QUANTILE

A quantile is a value (in a sense, a limit) that divides a set of observations into parts of a certain size. The four most commonly used quantiles are:

- 1/4, 2/4, 3/4 quantiles are otherwise called quartiles.
- 1/5, 2/5, 3/5, 4/5 quantiles are otherwise quintiles.
- 1/10, 2/10, ..., 9/10 quantiles are deciles.
- 1/100, 2/100, ..., 99/100 quantiles are otherwise percentiles.

Computing quantiles can be compared to cutting a cake. We can cut it into four equal parts - each is 25% of the cake. Alternatively, we can divide the cake into five equal parts or ten.

As already mentioned, quantiles are values that separate parts of the whole. So, the quantiles are at the "cuts".

For example, a 2/5 quantile, or quintile, divides the data set into two parts: 40% of the values below this quintile and 60% of the observations above it.

Quantiles are expressed in the same units as the variable.

Quantiles have their parameters: q, k.
The q number indicates how many equal (or nearly equal) parts the set of values will be divided into. Quartiles have q = 4, i.e. the dataset will be divided into four parts with 25% of the values in each part.

The k parameter represents the number of the specified q-quantile. It can be described as k-th q-quantile. Therefore, the upper quartile or third quartile (Q3) has k = 3 and q = 4 (3/4).

FORMULAS

There are various ways of calculating quantiles. And to tell the truth, there is no universal agreement on calculating the quartile values. Four methods are used. You can read more about it on Wikipedia: https://en.wikipedia.org/wiki/Quartile#Computing_methods.

Here I will depict two ways of computing quantiles. Let's focus specifically on the quartiles. The first is the graphical method which is the simplest one. However, it cannot be used effectively on large data sets.

The second method is to use Python. For those interested in Data Science, it will be the best choice.

1) graphic way (see graphic):
   a) set the values in ascending order.
   b) calculate the median - for sets with an even number of elements, the median is the arithmetic mean of the two middle values. For odd numbers, the median is the middle value. This gave us the middle quartile (Q2).
   c) the median divides the set of values into halves. To calculate the lower quartile (Q1), calculate the median for the first half of the set of values.
   d) to calculate the upper quartile (Q3), calculate the median for the second half of the value.

2) the python way (you can find this in the file quantiles.py)

PERCENTILE
Percentiles divide the set of values into 100 parts, and the values fall below or above the selected percentile. If the set of values had 100 elements (containing all integers from 1 to 100) and the percentile 5/100 were chosen, then the point separating the intervals would be 5. This means that 5% of the value is below or equal to this percentile and the remaining 95% is above or equal to it.

WHAT IS A QUARTILE?
A quartile is a measure of the position of observation. For example, the Q1 quartile separates the lowest 25% of observations from the 75% highest. The second quartile (median) divides the set of values into two halves - 50% of values are below the median and 50% of values are above.

IQR: DEFINITION
The interquartile range, like variance and standard deviation, is a measure of variability, and dispersion. In the case of variance and standard deviation, the problem is their susceptibility to outliers. The solution to this problem is the interquartile range, which is the difference between the 75th and 25th percentiles or, in other words, the upper quartile and lower quartile.

By definition, 50% of all observations are between these quartiles. You could say that the IQR is the middle 50% of the values.

IQR: FORMULA
IQR = Q3 - Q1

IQR - interquartile range
Q3 - upper quartile
Q1 - lower quartile

APPLICATIONS
The above knowledge may find its application in quantile regression. It is an extension of linear regression and is used when linear regression cannot be utilized because its assumptions are not met (homoscedasticity, linearity, etc.).

In contrast to linear regression which estimates the conditional mean, quantile regression calculates the conditional median (or other quantiles) of the response variable.