Quantiles:

The quantile of order p or pth quantile ($0<p<1$) is a value of the variable which divides the whole frequency distribution in two parts such that p-proportion of the total number of observations are less than or equal to it and (1-p) proportion of the total number of observations are greater than it. p= 0.5 refers to the median.

Some *p*-quantiles have special names:

- The 2-quantile is called the **median**
- The 3-quantiles are called tertiles or terciles
- The 4-quantiles are called **quartiles**
- The 5-quantiles are called quintiles
- The 6-quantiles are called sextiles
- The 10-quantiles are called **deciles**
- The 12-quantiles are called duo-deciles
- The 20-quantiles are called vigintiles
- The 100-quantiles are called **percentiles**
- The 1000-quantiles are called permilles

Calculation of quantiles:

A. **Quantiles for ungrouped data /simple series:**
   Let n be the total no. of observations and we want to find the pth quantile, Z_p. The steps for calculating Z_p are:
   1) Arrange the observations in an ascending order $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$
   2) Calculate Q=(n+1)p. Suppose Q is such that, $r \le Q < r+1$, where r is the largest integer not exceeding Q.
   3) Let $x_{(r)}$ *and* $x_{(r+1)}$ be the rth and (r+1)th order observations. Then, we have (using interpolation)

$$\frac{Z_p - x_{(r)}}{x_{(r+1)} - x_{(r)}} = \frac{(n+1)p - r}{(r+1) - r}$$

$$\Rightarrow Z_p = x_{(r)} + \{(n+1)p - r\}\{x_{(r+1)} - x_{(r)}\}$$

   When, (n+1)p is an integer r the the 2nd term of the above vanished and $Z_p = x_{(r)}$

B. **Quantiles for ungrouped frequency distribution:**
   Step 1: Prepare cumulative frequency (less than type) distribution table
   Step 2: Calculate Np, where N is the total frequency
   Step 3: Identify the class corresponding to the cumulative frequency just greater than or equal to Np i.e., $F_{k-1} < Np \le F_k$, where F_i is the cumulative frequency of ith class. The class value corresponding to this class (i.e., kth class) gives pth quantile.

C. **Quantiles for grouped frequency distribution:**

Here first identify the pth quantile class as the class for which the cumulative frequency (less than type) just exceeds Np, N=Total frequency. The pth quantile, Z_p is given by,

$$Z_p = l + \frac{h(Np - F)}{f}$$

where,

l → lower class boundary of the pth quantile class
h →width of the pth quantile class
f → frequency of the pth quantile class
F → cumulative frequency (less than type) of the class previous to the pth quantile class

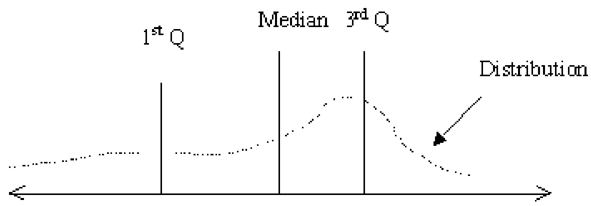
Special cases:

Quartiles:

Quartiles are the points which divide the whole distribution in four equal parts. There are 3 quartiles, viz. 1st quartile, 2nd quartile and 3rd quartile.

1st quartile divides the whole frequency distribution in 1:3 ratio. It is a value of the variable such that 25% (i.e., one-fourth) of the total observations fall below it and 75% (i.e., three-fourth) above. 2nd quartile is nothing but the median.

3rd quartile divides the whole frequency distribution in 3:1 ratio i.e., 75% of the total observations fall below it and 25% above.



Calculation of quartile:

For ungrouped data:

Step 1: Arrange the data in ascending order

Step 2: For 1st quartile, obtain (n+1)/4, n being the total no. of observations. If (n+1)/4 is an integer then (n+1)/4th ordered value gives Q_1, otherwise we have to interpolate.

In that case, say, (n+1)/4 = I+F (I is the integral part and F is the fractional part).

Then, Q_1=Ith value + F.((I+1)th value - Ith value)

Similarly, for 3rd quartile we have to obtain 3(n+1)/4 and proceed as above.

For ungrouped frequency distribution:

In a cumulative frequency (less than type) distribution table, the variate value corresponding to smallest of the cumulative frequencies >= (N+1)/4 gives 1st quartile Q_1 and smallest of the cumulative frequencies >=3(N+1)/4 gives 3rd quartile Q_3, where N is the total frequency.

For grouped frequency distribution:

Step 1: Obtain cumulative frequency distribution (less than type).

Step 2: Identify the classes containing quartiles as follows:

The class corresponding to the cumulative frequency just >= (N+1)/4 contains Q_1 and the class corresponding to the cumulative frequency just >= 3(N+1)/4 contains Q_3, N being the total frequency.

Step 3: Now Q_1 and Q_3 are given by,

$$Q_1 = l + \frac{h(\frac{N}{4} - F)}{f}$$

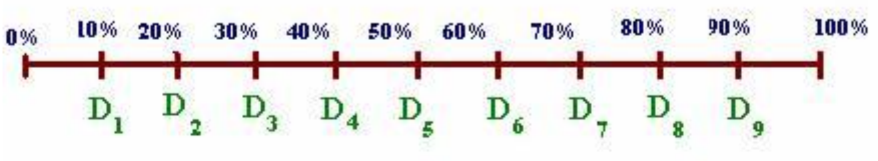
$$Q_3 = l' + \frac{h'(\frac{3N}{4} - F')}{f'}$$

where,

l (l') → *lower class boundary of the* Q_1(Q_3) *containing class*
h (h') → *width of the* Q_1(Q_3) *containing class*
f (f') → *frequency of the* Q_1(Q_3) *containing class*
F (F') → *cumulative frequency (less than type) of the class previous to the* Q_1(Q_3) *containing class*

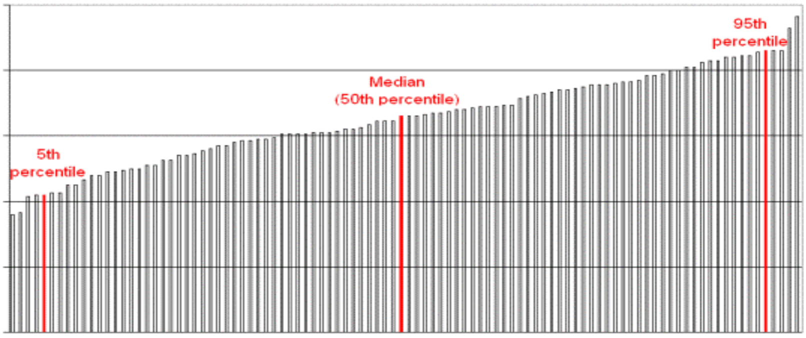
Deciles:

Deciles are the points which divide the whole distribution in 10 equal parts. There are 9 deciles of which 5th decile is nothing but the median.



Percentiles:

Percentiles are the summary measures that divide a ranked dataset into 100 equal parts. Each ranked dataset has 99 percentiles. Percentiles are usually denoted by P_1, P_2, ...,P_{99}. Clearly 25th percentile, $P_{25}=Q_1$, the 1st quartile, 50th percentile, $P_{50}=Q_2$, the median and 75th percentile, $P_{75}=Q_3$, the 3rd quartile.



Calculation of percentiles:

For ungrouped data:

Let n be the total number of observations. We want to obtain k^{th} percentile P_k.

Step 1: Arrange the observations in ascending order

Step 2: Obtain (n+1)k/100. If this is an integer then P_k is $(n+1)k/100^{th}$ order observation.

If (n+1)k/100 is not an integer then suppose, (n+1)k/100 = I + F (I be the Integral part and F be the fractional part). Then using interpolation,

$$P_k = I^{th} \text{ observation} + F.((I+1)^{th} \text{ observation} - I^{th} \text{ observation})$$

For frequency distribution:

The calculation is same as the p^{th} quantile where p = k/100 for k^{th} percentile. So replacing Np by Nk/100 (in the formula of pth quantile) the k^{th} percentile is given by,

$$P_k = l + \frac{h(\frac{Nk}{100} - F)}{f}$$

Percentile rank:

Percentile rank of an observation is the percentage of observations lying below or equal to it. It is obtained from the above formula where P_k is known and k is to be obtained.

Percentile rank for a series data can be obtained by the following simple formula:

$$Percentage\ rank\ of\ x_i = \frac{Number\ of\ data\ values\ less\ than\ x_i\ (i.e., cumulative\ frequency\ of\ less\ than\ type\ of\ x_i)}{N(Total\ no.of\ values\ in\ the\ dataset)}$$

Determination of quantile using graphical method:

Determination of quantile by using less than type ogive is similar to the determination of median. The pth quantile would be the value of the variable corresponding to the cumulative frequency Np in y-axis of less than type ogive. In case of more than type ogive, pth quantile is the the variate value corresponding to N(1-p) in y-axis.

For 1st quartile the variate value (along x-axis) corresponding to N/4 (on y-axis) gives Q_1 and 3N/4 gives Q_3 (from less than type ogive).

For kth percentile, P_k would be the value of the variable corresponding to the cumulative frequency Nk/100 from less than type ogive.