# STATISTICAL PARAMETERS

There are a set of parameters that help us understand better the distribution of a given variable. They can be of one of the following types:

**Measures of central tendency:** There are three of them:

-) **Mean:** It is the arithmetic average of the data. We calculate by adding all the values of the variable and then dividing by the total number of values. This can be written as:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{N}$$

So we add another column to the table, with each cell showing the value of the pair $x_i f_i$

-) **Median:** for an ordered set of data, the median is the value in the middle of the distribution. If we have two middle values then we calculate the average of them. If we have many different values of the variable then it can sometimes be really bothersome to order them all into a line in order to find the one in the middle, so we will use this other method:

-) Divide the total number of values, N, by 2

-) Calculate the absolute cumulative frequency column in your data table

-) Find the first number in that column that's greater than $N/2$. The corresponding value of the variable will be the median.

-) **Mode:** It is the value of the variable that occurs more often, that is, the one with a higher absolute frequency. The mode can be a single value or we can have several modes at the same time if they all have the same absolute frequency.

**Measures of dispersion:** There are four of them:

-) **Range:** It is the difference between the highest and the lowest values of the variable.

-) **Variance:** It is determined using the formula

$$\sigma^2 = \frac{\sum x_i^2 \cdot f_i}{N} - \bar{x}^2$$

So we will add yet another column to our table in order to calculate $x_i^2 f_i$. The rest of the parameters are already known from previous calculations.

-) **Standard deviation:** It is the positive square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$

A small standard deviation shows that the data values tend to be very close to the mean, while a high standard deviation indicates that the data values are widely spread.

-) **Pearson's coefficient of variation:** It is used to compare several distributions, especially when they don't use the same units, since this parameter is dimensionless.

$$CV = \sigma / \bar{x}$$