Measurements of Disperion

Quintiles

Quantiles describe location of individual values (within the variable scope) and are resistant to outlying observations similarly like the mode. Generally the quantile is defined as a value that divides the sample into two parts. The first one contains values that are smaller than given quantile and the second one with values larger or equal than the given quantile. The data must be sorted ascendingly from the lowest to the highest value.

Quantile of variable x that separates 100% smaller values from the rest of the samples (i.e. from 100(1-p)% values) will be called 100p % quantile and marked x_p.

In real life you most often come across the following quantiles:

- **QUARTILES**

In case of the four-part division the values of the variate corresponding to 25%, 50%, and 75% of the total distribution are called quartiles.

1- **Lower quartile** :

Lower quartile x0,25 = 25% quantile - divides a sample of data in a way that 25% of the values are smaller than the quartile, i.e. 75% are bigger (or equal).

2- Median:

Median x0,5 = 50% quantile - divides a sample of data in a way that 50% of the values are smaller than the median and 50% of values are bigger (or equal).

3- Upper quartile:

Upper quartile x0,75 = 75% quantile - divides a sample of data in a way that 75% of values are smaller than the quartile, i.e. 25% are bigger (or equal).

Example [1]

Data	6 47 49 15 43 41 7 39 43 41 36
Data in ascending order	6 7 15 36 39 41 41 43 43 47 49
Median	41
Upper quartile	43
Lower quartile	15

The difference between the 1st and 3rd quartile is called the **Inter-Quartile Range (IQR)**.

$$IQR = x_{0.75} - x_{0.25}$$

- **DECILES** – $x_{0.1}$; $x_{0.2}$; ... ; $x_{0.9}$

The deciles divide the data into 10 equal regions.

I] **Percentiles** – $x_{0.01}$; $x_{0.02}$; …; $x_{0.99}$

The percentiles divide the data into 100 equal regions.

For example, the 80th percentile is the number that has 80% of values below it and 20% above it. Rather than counting 80% from the bottom, count 20% from the top.

Note: The 50th percentile is the median.

ii] · **Minimum x_{min} and Maximum x_{max}**

$x_{min} = x_o$, i.e. 0% of values are less than minimum

$x_{max} = x_1$, i.e. 100% of values are less than maximum

1. The sample population needs to be ordered by size
2. The individual values are sequenced so that the smallest value is at the first place and the highest value is at n-th place (n is the total number of values)
3. 100p% quantile is equal to a variable value with the sequence z_p where:

$$z = n \times p + 0.5$$

z_p has to be rounded to integer is the following process to determine quantiles:

REMEMBER

In case of a data set with an even number of values the median is not uniquely defined. *Any number between two middle values (including these values) can be accepted as the median. Most often it is the middle value.*

We are now going to discuss the **relation** between **quantiles** and **the cumulative relative frequency**. The value ***p*** denotes cumulative relative frequency of quantile x_p i.e. relative frequency of those variable values that are smaller than quantile xp.

Quantile and cumulative relative frequency are inverse concepts.

Graphical or tabular representation of the ordered variable and appropriate cumulative frequencies is known as **distribution function of the cumulative frequency** or **empirical distribution function**.

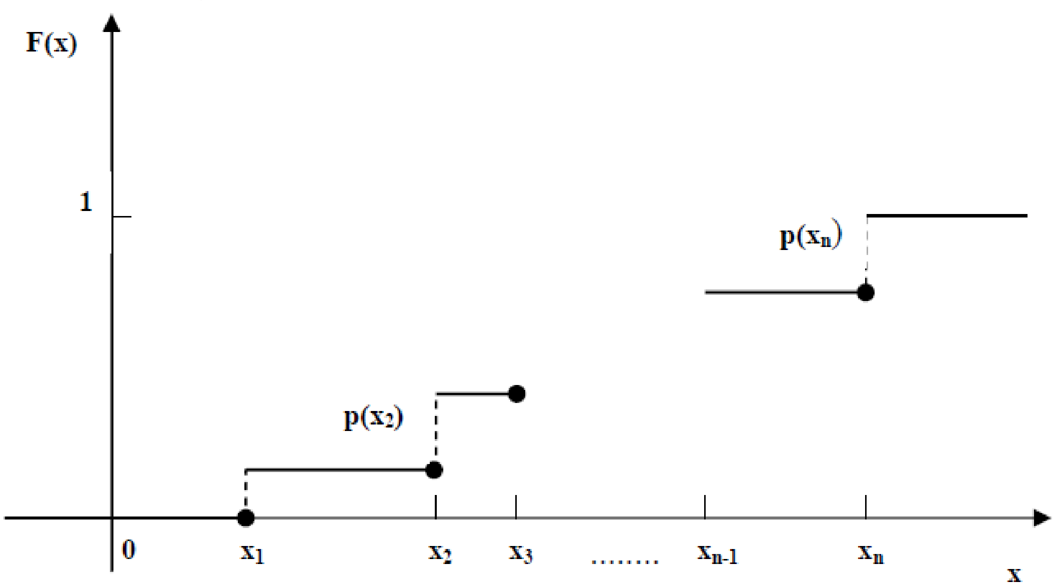
- **EMPIRICAL DISTRIBUTION FUNCTION F(X) FOR THE QUANTITATIVE VARIABLE**

We put the sample population in ascending order ($x1<x_2< \dots <x_n$) and we denote $p(x_i)$ as relative frequency of the value x_i. For empirical distribution function F(x) it must then be true that:

$$F(x)=\begin{cases} 0 & \text{for } x \le x_1 \\ \sum_{i=1}^{j} p(x_i) & \text{for } x_j < x \le x_{j+1},\ 1 \le j \le n-1 \\ 1 & \text{for } x_n < x \end{cases}$$

The empirical distribution function is a monotonous, increasing function and it runs from the left.

$$p(x_i)=\lim_{x \to x_{i+}} F(x)-F(x_i)$$



- **MAD**

MAD is a short for **Median Absolute Deviation** from the median.

MAD is determined as follows:

1. Order the sample population by size.
2. Determine the median of the sample population.
3. For each value determine absolute value of its deviation from the median.
4. Put absolute deviations from the median in ascending order by size.
5. Determine the median of the absolute deviations from the median i.e. MAD