A.2.3. Positional characteristics:

Position characteristics are important data for the study of statistical series.

A.2.3.1. The mode:

The mode (Mo) is a positional character in a statistical series. It represents the **most repeated value** in the data set.

- A data set can have one, two or multiple modes.
- A data set that has **two modes** is known as **bimodal** and the data set in which there are multiple modes is known as **multi-modal data set**.
- If all the elements in the data set are repeated with the same frequency or count, then the data has **no mode**.
- Mode is the only central measure that can be identified and used for both **qualitative and quantitative data.**

a) Discrete quantitative variable:

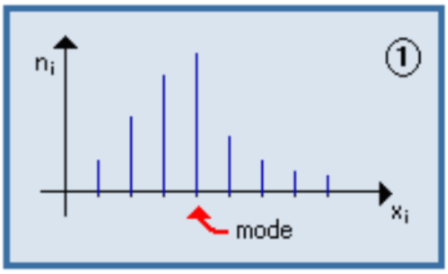

Figure. 2.11: Unimodal distribution for discrete variable.

Example 22 :

- Find the mode of the data below:

Length (m) X	**71**	**74**	**77**	**80**	**83**
Counts (ni)	6	17	41	27	9

In the above data, the most repeated value is 77 which is the mode of the data. Mo=77.

- Find the mode of the data below:

4, 5, 7, 10, 6, 4, 10, 12, 4, 5, 6, 10, 8.

In this case we have a bimodal data: Mo=4, 10

b) Continuous quantitative variable:

In the continuous case, the mode is in the class with the largest number of students (the modal class).

- **Equal class size:**
  - **Analytically**: To calculate the mode analytically, the following law is used:

$$\boldsymbol{Mo = a + c.\frac{b}{b+d}}$$

a: is the lower limit of the modal class.

c: is the length of the modal class.

b: the difference between the count (or frequency) of the modal class and the count (or frequency) of the class that precedes it.

d: the difference between the count (or frequency) of the modal class and the count (or frequency) of the class that follows it.

  - **Graphically**: The mode is calculated on the histogram represented by ni or fi:

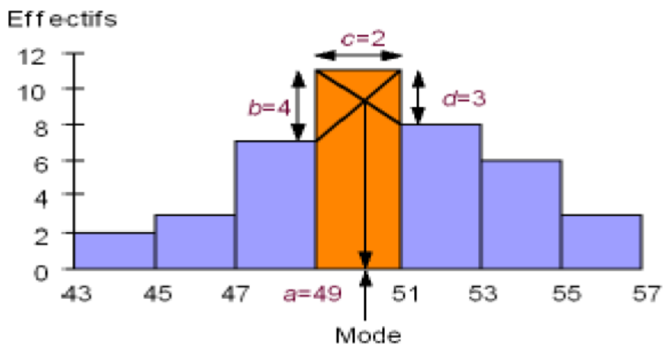

Figure. 2.12: Unimodal distribution for continuous variable.

Example 23:

We consider the age of the inhabitants in a neighborhood. Determinate the mode analytically and graphically

Classes	**[10-15[**	**[15-20[**	**[20-25[**	**[25-30[**	**[30-35]**	**Total**
ni (Counts)	5	20	10	15	10	60
fi (Frequencies)	0.08	0.33	0.17	0.25	0.17	1
ai (Class size)	5	5	5	5	5	

- **Analytically:**

The modal class here is [15-20[. Mo ϵ [15-20[

$$\mathrm{Mo} = a + c.\frac{b}{b+d} = 15 + 5.\frac{0.25}{0.25+0.16} = 18.04\sim 18$$

- **Graphically:**

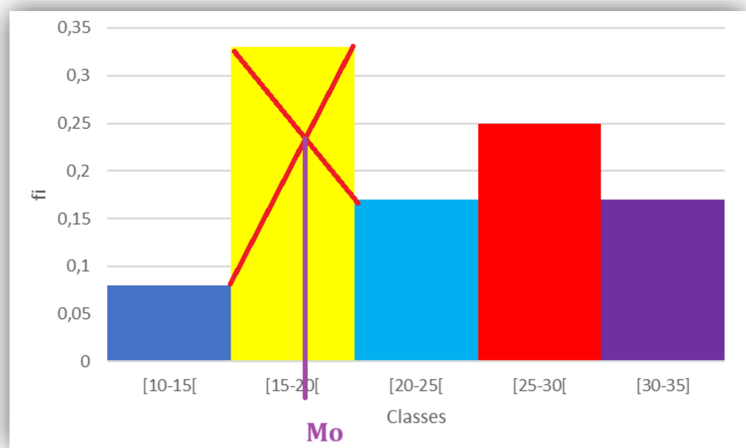

Figure. 2.13: Determination of the mode graphically.

- **Unequal class size:**

It is calculated in the same way on the histogram represented by **ni', fi'** or **di.**

Example 24:

We consider the age of the inhabitants of a neighborhood. Find the mode of these data.

Classes	[11-15[	[15-25[	[25-30[	[30-37[	[37-50]
ni (Counts)	20	10	5	15	10
fi (Frequencies)	0.33	0.16	0.08	0.25	0.16
ai (Class size)	4	10	5	7	13
ni' (Rectified ni)	$\frac{4*20}{4} = 20$	$\frac{4*10}{10} = 4$	$\frac{4*5}{5} = 4$	$\frac{4*15}{7} = 8.5$	$\frac{4*10}{13} = 3$

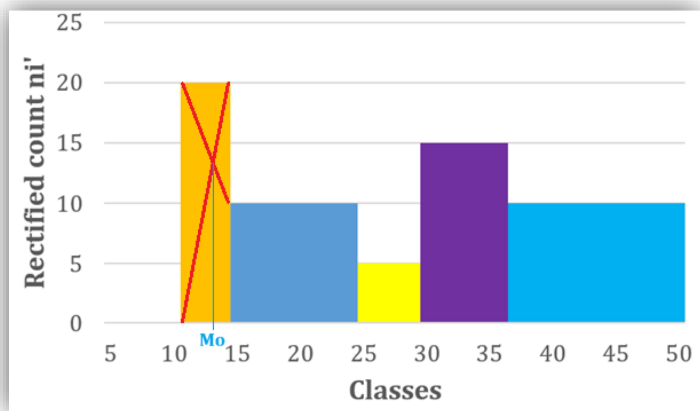

Figure. 2.13: Determination of the mode graphically.

The modal class here is [11-15[. Mo ϵ [11-15[

$$Mo = a + c.\frac{b}{b+d} = 11 + 4.\frac{20}{20+16} = 13.22$$

A.2.3.2. The median:

The median (Me or $x_{1/2}$) is the middle value of the data set. This value divides the dataset into two halves. It is a central measure that can only be calculated and used for quantitative, continuous or discrete variables.

The median is the value such that the function of the cumulative frequencies is equal to (0,5), F(Me)=0,5

The median of the data is computed after arranging the data either in ascending or descending order. The method of finding the median of the data set when the number of elements is even (pair) is different than the process of finding the median when number of elements is odd (impair). In this section, we will see how to find the median of the dataset when it has even and odd number of elements.

a) Discrete variable:

In this case it is first necessary to classify the observations of the series in ascending order, this operation is intended to assign a rank to each value and thus to determine more easily the middle of the series. Two cases can arise :

- If ***n*** is **even**, then Me equal to:

$$\boxed{x_{1/2} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}}$$

Donc la médiane est la modalité numéro (n+1)ème,

- If ***n*** is **odd**, then Me equal to:

$$\boxed{x_{1/2} = x_{(\frac{n+1}{2})}}$$

The median can be determined graphically for a discrete quantitative variable.

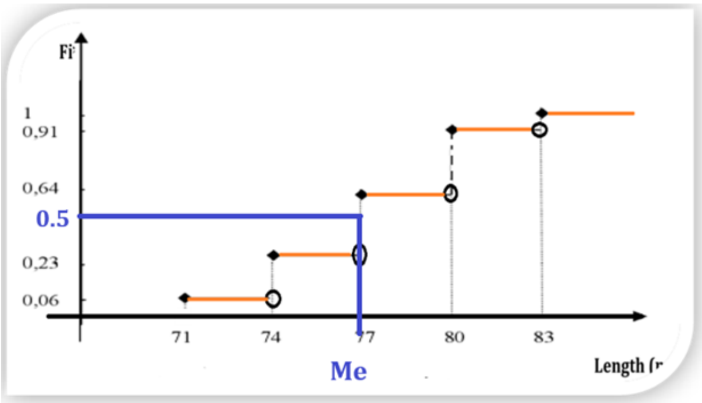

Figure. 2.14: Determination of the median graphically for DV.

Example 25:

The series of 20 marks obtained in statistics by second-year university students are:

5 6 7 7 8 8 8.5 9 9.5 10 10 10.5 11.5 11.5 12 13 13.5 13.5 14 17

We notice that this series is of size *n=20* so even, so the median is calculated as follows:

$$x_{\frac{1}{2}} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} = \frac{x_{(10)} + x_{(11)}}{2} = \frac{10 + 10}{2} = 10$$

Let us consider the following statistical series:

17 19 19 19 20 20 20 21 24 25 21 21 17 18

19 22 22 22 23 23 23 20 20 23 24 19 25

This series is of size *n=27* so odd, so the median is calculated as follows:

$$x_{1/2} = x_{\left(\frac{n+1}{2}\right)} = x_{\left(\frac{27+1}{2}\right)} = x_{\left(\frac{28}{2}\right)} = x_{(14)} = 21$$

b) Continuous variable:

To determine the median of a continuous variable, we must locate the median, i.e. find the class that contains it from the cumulative frequencies.

If the value 0.5 is between F(A) and F(B), then the median denoted ***Me*** is in the class [A , B [, this class is called: **median class.**

F(A) ≤0.5 < F(B) So ***Me*** belongs to [A , B [

$$Me = A + \frac{(0.5 - F(A))(B - A)}{F(B) - F(A)}$$

To find ***Me*** we use the **linear interpolation method.** The interpolation analytic formula is:

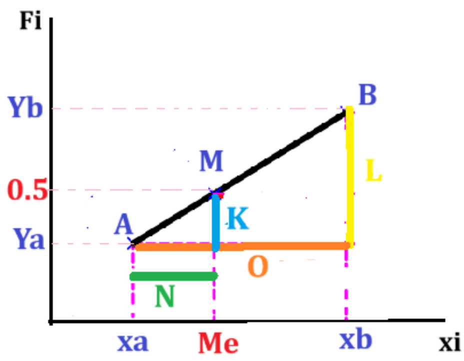

Figure. 2.15: Linear interpolation method.

$\frac{N}{K} = \frac{O}{L}$ $\frac{Me-Xa}{0.5-Ya} = \frac{Xb-Xa}{Yb-Ya}$ $Me = Xa + \frac{(0.5-Ya)(Xb-Xa)}{(Yb-Ya)}$

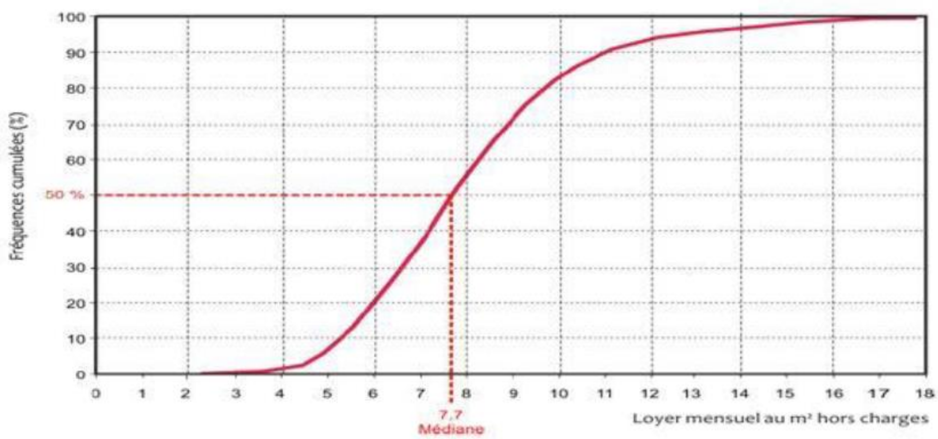

Figure. 2.16: Determination of the median graphically for CV.

Example 26:

Classes	**[10-15[**	**[15-20[**	**[20-25[**	**[25-30[**	**[30-35]**	**Total**
ni	5	20	10	15	10	60
fi	0.08	0.33	0.17	0.25	0.17	1
Fi	0.08	0.41	0.58	0.83	1	
F(x) **0**	0.08	0.41	0.58	0.83	1	

$$Me = Xa + \frac{(0.5 - Ya)(Xb - Xa)}{(Yb - Ya)}$$

Me ϵ [20, 25[

$$Me = 20 + \frac{(0.5 - 0.41)(25 - 20)}{(0.58 - 0.41)} = 22.6$$

A.2.3.3. The quantiles (Quartiles):

Different quartiles reflect the three values that divide the data into four equal parts. Quantiles **are** called **quartiles**. There are 3 quartiles Q1, Q2, Q3.

- We denote the first quartile as **Q1** and it represents 25% of the values less than it and 75% of the values greater than it. **F(Q1)=0.25**
- The second quartile is also known as median as it reflects 50% of the values above it and 50% below it. It is denoted as **Q2**. **F(Q2)=0.5**
- The third quartile has 25% values greater than it and 75% values smaller than it. It is represented as **Q3**. **F(Q3)=0.75**
- **Deciles**: The deciles divide the ordered series into 10 equal parts.
- **Percentiles**: The percentiles divide the ordered series into 100 equal parts.

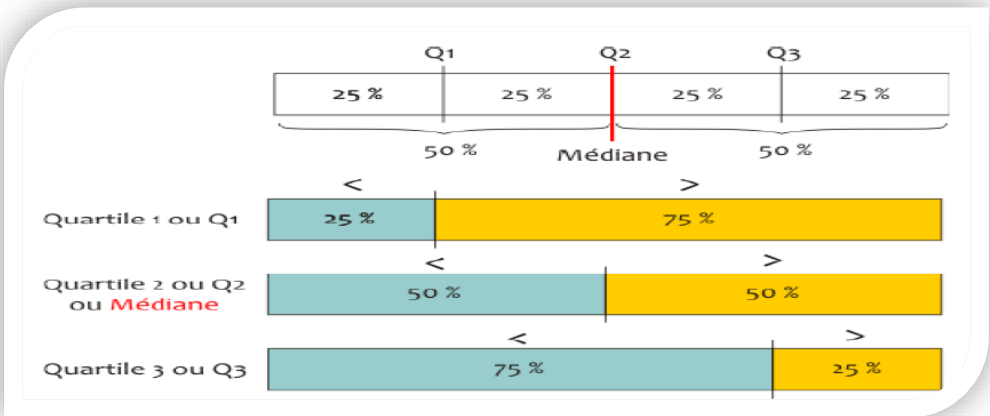

Figure. 2.17: The quartiles.

❖ **Case of a discrete variable:**

- **Analytically:**

– ***First method:***

- If ***n*** is **even**, then Me equal to:

$$Q1 = \frac{x_{\left(\frac{n}{4}\right)} + x_{\left(\frac{n}{4}+1\right)}}{2} \qquad Q2 = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} \qquad Q3 = \frac{x_{\left(\frac{3n}{4}\right)} + x_{\left(\frac{3n}{4}+1\right)}}{2}$$

- If ***n*** is **odd**, then Me equal to:

$$\boldsymbol{Q1 = x_{\left(\frac{n+1}{4}\right)}} \qquad \boldsymbol{Q2 = x_{\left(\frac{n+1}{2}\right)}} \qquad \boldsymbol{Q3 = x_{\left(\frac{3n}{4}+1\right)}}$$

- ***<u>Second method:</u>***
  - For Q1, we calculate N/4, then we determine the first integer *p* greater than or equal to N/4. The integer p is the rank of Q1.
  - For Q3, we do the same with 3N/4

Example 27:

17 19 19 19 20 20 20 21 24 25 21 21 17 18
19 22 22 22 23 23 23 20 20 23 24 19 25

For N=27, we have:

N/4=27/4=6.75, So Q1 is the seventh value in the series (Q1=19)

N/2=13.5, So Q2 is the fourteenth value in the series (Q2=21).

3N/4 = 20.25. and Q3 is the value number 21 (Q3=23).

- **Graphically:**

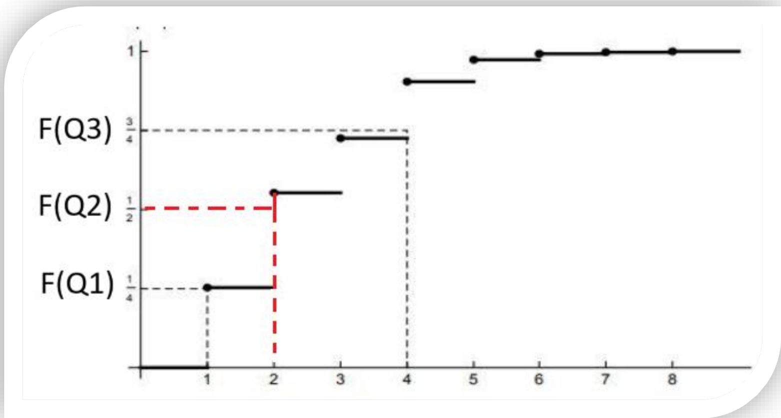

Figure. 2.18: Determination of the quartiles graphically for CD.

❖ **Case of a continuous variable:**

- **Analytically:**

$Q1 = A + \frac{(0.25 - F(A))(B-A)}{F(B)-F(A)}$ F(Q1) = 0.25 Q1 ϵ [A, B [

$Q2 = Me = B + \frac{(0.5-F(B))(C-B)}{F(C)-F(B)}$ F(Q2) = 0.5 Q2 ϵ [B, C [

$Q3 = C + \frac{(0.75-F(C))(D-C)}{F(D)-F(C)}$ F(Q3) = 0.75 Q3 ϵ [C, D [

- **Graphically:**

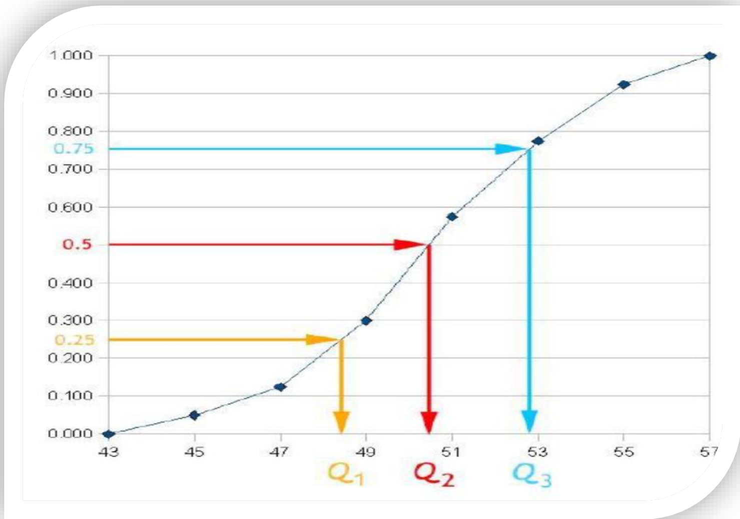

Figure. 2.19: Determination of the quartiles graphically for CV.

❖ **Interquartile Range**

The **interquartile range (IQR)** is the difference between the upper and lower quartile of a given data set and is also called a **midspread**. The IQR formula is given by;

$$IQR = Q3 - Q1$$