MODULE: Introduction aux probabilités et statistique descriptive "PART 02" Numerical descriptive measure - Mesure of center (central tendency)

Mr. Rahmani Naceur

2025

- The last level of statistical description is the numerical summary of a statistical distribution by numerical indicators or characteristic parameters:

Indicateurs numériques

- The last level of statistical description is the numerical summary of a statistical distribution by numerical indicators or characteristic parameters:

1. Position (or central tendency) parameters."Paramètres de position"

Indicateurs numériques

- The last level of statistical description is the numerical summary of a statistical distribution by numerical indicators or characteristic parameters:

1. Position (or central tendency) parameters."Paramètres de position"
2. Dispersion parameters."Paramètres de dispersion"

Indicateurs numériques

- The last level of statistical description is the numerical summary of a statistical distribution by numerical indicators or characteristic parameters:

1. Position (or central tendency) parameters."Paramètres de position"
2. Dispersion parameters."Paramètres de dispersion"
3. Shape parameters: Paramètres de forme (kurtosis,skewness)

Position parameters

- The position parameters (mode, median, mean, etc.) allow you to know around which values the values of a statistical variable are located..

The mode of a V.S is the value which has the greatest partial frequency (or the greatest partial frequency)
denoted M_o, is the modality which admits the greatest frequency (for a qualitative variable or a discrete quantitative variable).

Note: You can have more than one mode or nothing.

Example

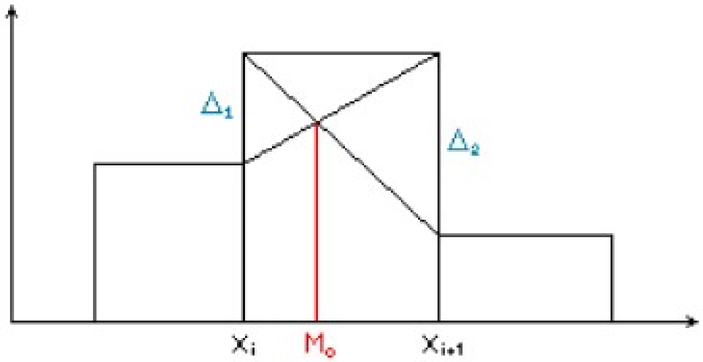
L'étude de *19 familles* a conduit à la distribution suivante (le nombre d'enfants) 2, 3 , 5, 6, 3, 7, 3, 2, 0, 1, 6, 9, 4, 3, 5, 1, 2, 3, 5.

donc $M_o = 3$.

Pour une variable quantitative continue:

For a continuous quantitative variable we speak of a modal class: it is the class whose frequency density (effectiveness) is maximum.

a- The geometric method:



- **b- The numerical method:** Given that the modal class $[x_i, x_{i+1}[$ is determined, the mode Mo satisfies:

$$Mo = x_m + \frac{\Delta_i}{\Delta_s + \Delta_i}.c$$

with:

- **b- The numerical method:** Given that the modal class $[x_i, x_{i+1}[$ is determined, the mode Mo satisfies:

$$Mo = x_m + \frac{\Delta_i}{\Delta_s + \Delta_i}.c$$

with:

- X_m : Lower boundary of the class with the highest frequency..

- **b- The numerical method:** Given that the modal class $[x_i, x_{i+1}[$ is determined, the mode Mo satisfies:

$$Mo = x_m + \frac{\Delta_i}{\Delta_s + \Delta_i}.c$$

with:

- X_m : Lower boundary of the class with the highest frequency..
- c : Width of the modal class interval $(x_{m+1} - x_m)$.

- **b- The numerical method:** Given that the modal class $[x_i, x_{i+1}[$ is determined, the mode Mo satisfies:

$$Mo = x_m + \frac{\Delta_i}{\Delta_s + \Delta_i}.c$$

with:

- X_m : Lower boundary of the class with the highest frequency..
- c : Width of the modal class interval $(x_{m+1} - x_m)$.
- Δ_i : Difference in frequency between the modal class and the closest lower class.

- **b- The numerical method:** Given that the modal class $[x_i, x_{i+1}[$ is determined, the mode Mo satisfies:

$$Mo = x_m + \frac{\Delta_i}{\Delta_s + \Delta_i}.c$$

with:

- X_m : Lower boundary of the class with the highest frequency..
- c : Width of the modal class interval $(x_{m+1} - x_m)$.
- Δ_i : Difference in frequency between the modal class and the closest lower class.
- Δ_s : Difference in frequency between the modal class and the closest upper class.

Remarks.

- When the classes adjacent to the modal class have equal frequency densities, the mode coincides with the center of the modal class.

Remarks.

- When the classes adjacent to the modal class have equal frequency densities, the mode coincides with the center of the modal class.
- The mode is highly dependent on the choice of class intervals.

Remarks.

- When the classes adjacent to the modal class have equal frequency densities, the mode coincides with the center of the modal class.
- The mode is highly dependent on the choice of class intervals.
- A statistical variable may exhibit several local modes; in this case, it is said to be **multimodal.**

Remarks.

- When the classes adjacent to the modal class have equal frequency densities, the mode coincides with the center of the modal class.
- The mode is highly dependent on the choice of class intervals.
- A statistical variable may exhibit several local modes; in this case, it is said to be **multimodal.**
- This situation is of particular interest, as it highlights the existence of several sub-populations, and thus, the heterogeneity of the studied population.

1.1.2 The median

Me

- *Me* is the value of the variable for which the cumulative frequency equals 0.5 or 50%.

1.1.2 The median

Me

- *Me* is the value of the variable for which the cumulative frequency equals 0.5 or 50%.
- It therefore corresponds to the center of the statistical series arranged in ascending order, or to the value for which 50% of the observed values are greater and 50% are smaller.

- If n is odd, then $n = 2m + 1$ and the median is the middle value.

$$Me = x_{m+1} = x_{\frac{n+1}{2}}.$$

- If n is odd, then $n = 2m + 1$ and the median is the middle value.

$$Me = x_{m+1} = x_{\frac{n+1}{2}}.$$

- If n is even, then $n = 2m$ and the median is any value between x_m and x_{m+1}. In this case, it is convenient to take the average of the two values.

$$Me = \frac{x_m + x_{m+1}}{2} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

The numerical method:

- We look for the class that contains $n/2$ individuals (the median class) of the sample.

$$Me = X_m + \frac{\frac{n}{2} - N}{n_{méd}} c = X_m + \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{méd}} c$$

with:

The numerical method:

- We look for the class that contains $n/2$ individuals (the median class) of the sample.

$$Me = X_m + \frac{\frac{n}{2} - N}{n_{méd}} c = X_m + \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{méd}} c$$

with:

- X_m: ower boundary of the median class.

The numerical method:

- We look for the class that contains $n/2$ individuals (the median class) of the sample.

$$Me = X_m + \frac{\frac{n}{2} - N}{n_{méd}}c = X_m + \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{méd}}c$$

with:

- X_m: ower boundary of the median class.
- n_{med}: Frequency of the median class.

The numerical method:

- We look for the class that contains $n/2$ individuals (the median class) of the sample.

$$Me = X_m + \frac{\frac{n}{2} - N}{n_{méd}} c = X_m + \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{méd}} c$$

with:

- X_m: ower boundary of the median class.
- n_{med}: Frequency of the median class.
- N: Cumulative frequency up to (but not including) the median class.

The numerical method:

- We look for the class that contains $n/2$ individuals (the median class) of the sample.

$$Me = X_m + \frac{\frac{n}{2} - N}{n_{méd}} c = X_m + \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{méd}} c$$

with:

- X_m: ower boundary of the median class.
- n_{med}: Frequency of the median class.
- N: Cumulative frequency up to (but not including) the median class.
- n: Sample size.

The numerical method:

- We look for the class that contains $n/2$ individuals (the median class) of the sample.

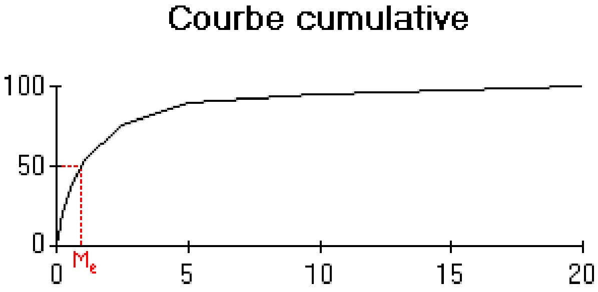
$$Me = X_m + \frac{\frac{n}{2} - N}{n_{méd}} c = X_m + \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{méd}} c$$

with:

- X_m: ower boundary of the median class.
- n_{med}: Frequency of the median class.
- N: Cumulative frequency up to (but not including) the median class.
- n: Sample size.
- c :Width of the median class interval $(x_{m+1} - x_m)$.

- **b- The geometric method:**

The median divides the entire set of observed values (arranged in ascending or descending order) into two subsets with equal frequencies.



- For a continuous real quantitative statistical variable X, he quartiles are the real numbers Q_1, Q_2, Q_3, for which the cumulative frequencies of X are respectively **0,25, 0,50, 0,75.**
- These are the values for which the ordinate of the cumulative frequency curve is respectively equal to 0,25, 0,50, 0,75.
- The quartiles divide the range into four intervals that have the same frequency.
- The second quartile, Q_2, is equal to the median.

- **Deciles**

The 9 deciles are the real numbers that divide the range into ten intervals of equal frequency.

The 99 percentiles are the real numbers that divide the range into one hundred intervals of equal frequency.

Deciles - Centiles

- **Deciles**

The 9 deciles are the real numbers that divide the range into ten intervals of equal frequency.

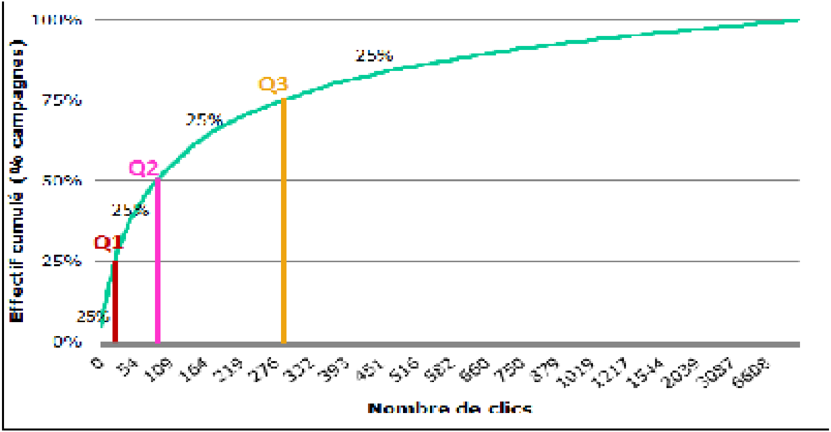
- **Percentiles**

The 99 percentiles are the real numbers that divide the range into one hundred intervals of equal frequency.

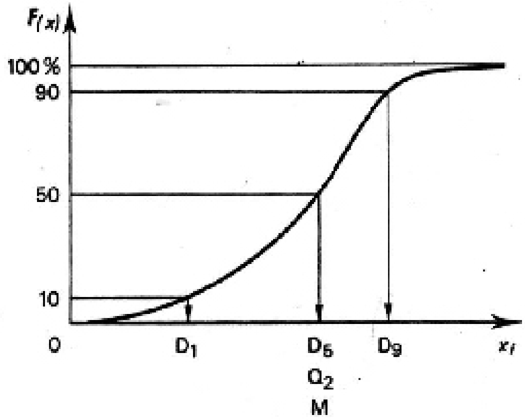
- **a- The numerical method:**

Les valeurs	**non groupées**	**sont groupées**
Quartiles	$Q_i = X_{\frac{in}{4}}$;pour $i = 1, ..., 4$	$Me = X_m + \dfrac{\frac{in}{4} - N}{n_{Q_i}} c.$
Deciles	$D_i = X_{\frac{in}{10}}$;pour $i = 1, ..., 9$	$D_i = X_m + \dfrac{\frac{in}{10} - N}{n_{D_i}} c.$
Percentiles	$P_i = X_{\frac{in}{100}}$;pour $i = 1, ..., 99$	$P_i = X_m + \dfrac{\frac{in}{100} - N}{n_{C_i}} c.$

- **b -The geometric method: (Quartiles):**



- **b - la méthode géométrique (Deciles):**



1.1.4 THE MEANS:

Calculating a mean allows us to summarize the numerical information we have, which obviously means that we lose some information (particularly about the dispersion of the values of the considered variable). The most commonly used mean is the arithmetic mean.

The arithmetic mean of a statistical series is the sum of the values divided by the total number of values.

The geometric mean of the values taken by a variable is the number that preserves the product of these values.

The harmonic mean of a series of values is the number that preserves the sum of the inverses of these values

1.1.4 THE MEANS:

Les valeurs	non groupées	groupées (classes)
The arithmetic mean	$\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{k} n_i x_i$	$\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{k} n_i c_i,$
he geometric mean	$G = \sqrt[n]{\prod\limits_{i=1}^{k} x_i^{n_i}}$	$G = \sqrt[n]{\prod\limits_{i=1}^{k} c_i^{n_i}}$
The harmonic mean	$H = \frac{n}{\sum\limits_{i=1}^{k} \frac{n_i}{x_i}}$	$H = \frac{n}{\sum\limits_{i=1}^{k} \frac{n_i}{c_i}}$