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

Mr. Rahmani Naceur

14 -03- 2024

Indicateurs numériques

- 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.

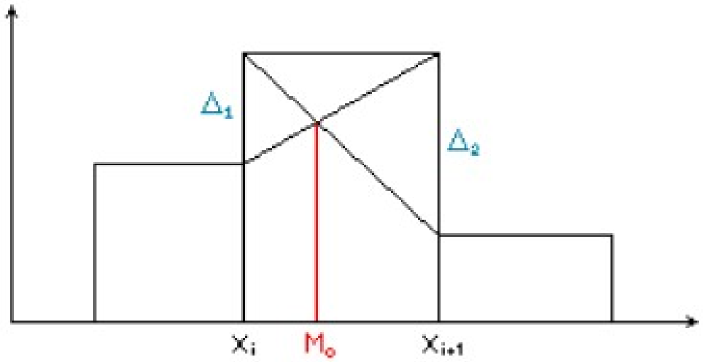
Example

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- la méthode numérique:** La classe modale $[x_i, x_{i+1}[$ étant déterminée, le mode *Mo* vérifie:

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

avec:

- **b- la méthode numérique:** La classe modale $[x_i, x_{i+1}[$ étant déterminée, le mode Mo vérifie:

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

avec:

- X_m : limite inférieure de la classe d'effectif maximal.

- **b- la méthode numérique:** La classe modale $[x_i, x_{i+1}[$ étant déterminée, le mode Mo vérifie:

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

avec:

- X_m : limite inférieure de la classe d'effectif maximal.
- c : intervalle de classe modale $(x_{m+1} - x_m)$.

1.1.1:
The mode:

- **b- la méthode numérique:** La classe modale $[x_i, x_{i+1}[$ étant déterminée, le mode Mo vérifie:

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

avec:

- X_m : limite inférieure de la classe d'effectif maximal.
- c : intervalle de classe modale $(x_{m+1} - x_m)$.
- Δ_i : Ecart d'effectif entre la classe modale et la classe inférieure la plus proche.

- **b- la méthode numérique:** La classe modale $[x_i, x_{i+1}[$ étant déterminée, le mode Mo vérifie:

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

avec:

- X_m : limite inférieure de la classe d'effectif maximal.
- c : intervalle de classe modale $(x_{m+1} - x_m)$.
- Δ_i : Ecart d'effectif entre la classe modale et la classe inférieure la plus proche.
- Δ_s : Ecart d'effectif entre la classe modale et la classe supérieure la plus proche.

Remarques.

- Lorsque les classes adjacentes à la classe modale ont des densités de fréquences égales, le mode coïncide avec le centre de la classe modale.

Remarques.

- Lorsque les classes adjacentes à la classe modale ont des densités de fréquences égales, le mode coïncide avec le centre de la classe modale.
- Le mode dépend beaucoup de la répartition en classes.

Remarques.

- Lorsque les classes adjacentes à la classe modale ont des densités de fréquences égales, le mode coïncide avec le centre de la classe modale.
- Le mode dépend beaucoup de la répartition en classes.
- Une variable statistique peut présenter plusieurs modes locaux : on dit alors qu'elle est plurimodale. Cette situation est intéressante : elle met en évidence l'existence de plusieurs sous-populations, donc l'hétérogénéité de la population étudiée.

1.1.2 La médiane

- La médiane, *Me*, est la valeur du caractère pour laquelle la fréquence cumulée est égale à 0,5 ou 50%. Elle correspond donc au centre de la série statistique classée par ordre croissant, ou à la valeur pour laquelle 50% des valeurs observées sont supérieures et 50% sont inférieures.

1.1.2 La médiane

Cas des données non groupées:

- si n est impair, alors $n = 2m + 1$ et la médiane est la valeur du milieu

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

1.1.2 La médiane

Cas des données non groupées:

- si n est impair, alors $n = 2m + 1$ et la médiane est la valeur du milieu

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

- si n est pair, alors $n = 2m$ et une médiane est une valeur quelconque entre x_m et x_{m+1}. Dans ce cas il peut être commode de prendre le milieu

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

- **a-la méthode numérique:**

- **a-la méthode numérique:**
- On cherche la classe contenant le $n/2$ individu (la classe médiane) de l'échantillon.

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

avec:

1.1.2 La médiane

Cas des données groupées:

- **a-la méthode numérique:**
- On cherche la classe contenant le $n/2$ individu (la classe médiane) de l'échantillon.

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

avec:

- X_m: limite inférieure de la classe médiane.

1.1.2 La médiane

Cas des données groupées:

- **a-la méthode numérique:**
- On cherche la classe contenant le $n/2$ individu (la classe médiane) de l'échantillon.

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

avec:

- X_m: limite inférieure de la classe médiane.
- n_{med}: effectif de la classe médiane

- **a-la méthode numérique:**
- On cherche la classe contenant le $n/2$ individu (la classe médiane) de l'échantillon.

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

avec:

- X_m: limite inférieure de la classe médiane.
- n_{med}: effectif de la classe médiane
- N: effectif cumulé croissante de la classe avant la classe médiane.

- **a-la méthode numérique:**
- On cherche la classe contenant le $n/2$ individu (la classe médiane) de l'échantillon.

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

avec:

- X_m: limite inférieure de la classe médiane.
- n_{med}: effectif de la classe médiane
- N: effectif cumulé croissante de la classe avant la classe médiane.
- n: taille de l'échantillon

- **a-la méthode numérique:**
- On cherche la classe contenant le $n/2$ individu (la classe médiane) de l'échantillon.

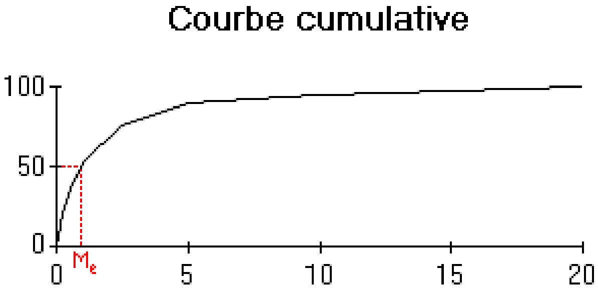
$$Me = X_m + \frac{\frac{n}{2} - N}{n_{méd}} c = X_m + \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{méd}} c$$

avec:

- X_m: limite inférieure de la classe médiane.
- n_{med}: effectif de la classe médiane
- N: effectif cumulé croissante de la classe avant la classe médiane.
- n: taille de l'échantillon
- c : intervalle de classe médiane $(x_{m+1} - x_m)$.

- **b- la méthode géométrique:**

La médiane partage l'ensemble des valeurs observées (classées par valeurs croissantes ou décroissantes) en deux sous- ensembles d'effectifs égaux.



- Pour une variable statistique quantitative réelle continue X, on appelle quartiles les nombres réels Q_1, Q_2, Q_3, pour lesquels les fréquences cumulées de X sont respectivement 0,25, 0,50, 0,75.
- Ce sont les valeurs pour lesquelles l'ordonnée de la courbe cumulative des fréquences est respectivement égale à 0,25, 0,50, 0,75.
- Les quartiles partagent l'étendue en quatre intervalles qui ont le même effectif.
- Le deuxième quartile, Q_2, est égal à la médiane.

Déciles - Centiles

- **Les Déciles**

Les 9 déciles sont les nombres réels qui partagent l'étendue en dix intervalles de même effectif.

Les 99 percentiles sont les nombres réels qui partagent l'étendue en cent intervalles de même effectif.

Déciles - Centiles

- **Les Déciles**

Les 9 déciles sont les nombres réels qui partagent l'étendue en dix intervalles de même effectif.

- **LesCentiles**

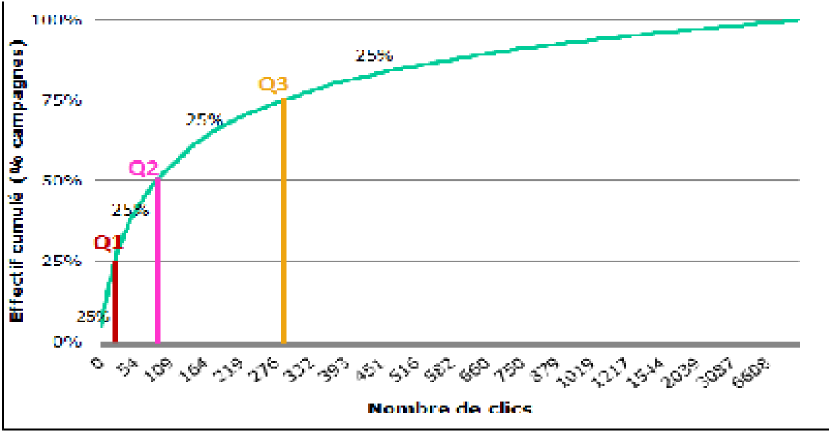
Les 99 percentiles sont les nombres réels qui partagent l'étendue en cent intervalles de même effectif.

1.1.3 Quartiles, Déciles, Centiles

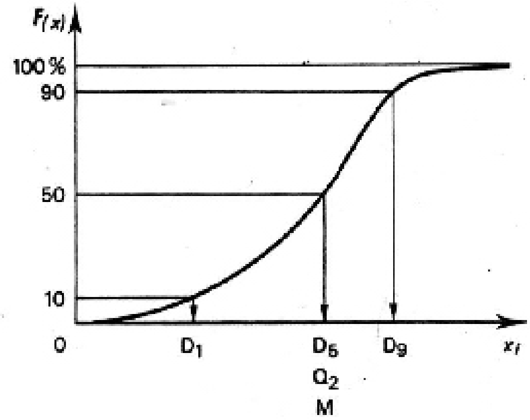
- **a-la méthode numérique:**

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

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



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



1.1.4 LES MOYENNES:

Le calcul d'une moyenne permet de résumer l'information chiffrée dont on dispose, ce qui signifie évidemment que l'on perd en même temps de l'information (notamment sur la dispersion des valeurs de la variable considérée).
La moyenne la plus communément utilisée est la moyenne arithmétique.

- **La moyenne arithmétique** d'une série statistique est la somme des valeurs divisée par le nombre total des valeurs.

1.1.4 LES MOYENNES:

Le calcul d'une moyenne permet de résumer l'information chiffrée dont on dispose, ce qui signifie évidemment que l'on perd en même temps de l'information (notamment sur la dispersion des valeurs de la variable considérée).
La moyenne la plus communément utilisée est la moyenne arithmétique.

- **La moyenne arithmétique** d'une série statistique est la somme des valeurs divisée par le nombre total des valeurs.
- **La moyenne géométrique** des valeurs prises par une variable est le nombre qui conserve le produit de ces valeurs.

1.1.4 LES MOYENNES:

Le calcul d'une moyenne permet de résumer l'information chiffrée dont on dispose, ce qui signifie évidemment que l'on perd en même temps de l'information (notamment sur la dispersion des valeurs de la variable considérée).
La moyenne la plus communément utilisée est la moyenne arithmétique.

- **La moyenne arithmétique** d'une série statistique est la somme des valeurs divisée par le nombre total des valeurs.
- **La moyenne géométrique** des valeurs prises par une variable est le nombre qui conserve le produit de ces valeurs.
- **La moyenne harmonique** d'une série de valeurs est le nombre qui conserve la somme des inverses de ces valeurs.

1.1.4 LES MOYENNES:

Les valeurs	non groupées	groupées (classes)
La moyenne arithmétique	$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$	$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i,$
La moyenne géométrique	$G = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}}$	$G = \sqrt[n]{\prod_{i=1}^{k} c_i^{n_i}}$
La moyenne harmonique	$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$	$H = \frac{n}{\sum_{i=1}^{k} \frac{n_i}{c_i}}$