Chapitre 02 :

Les caractéristiques d'une série statistique à une variable : mesures de tendance centrale et de dispersion

Nous avons vu comment mettre en forme et présenter une série statistique dans un tableau statistique et comment la présenter graphiquement, nous allons dans cette section voir comment la résumer à l'aide des valeurs caractéristiques comme la moyenne, le mode, la médiane, écart-type,….car le but d'une étude statistique est aussi de résumer des données par des paramètres ou synthétiseurs.
Il existe 3 types de paramètres :

- Paramètres de position (ou de tendance centrale)
- paramètres de dispersion
- paramètres de forme : Asymétrie, aplatissement, concentration (3eme année).

1. Paramètres de tendance centrale

1.1 La moyenne arithmétique

La moyenne $\bar{x}$ ne se définit que pour une variable statistique quantitative.
Elle se définit comme le rapport de la somme des valeurs de la série statistique par l'effectif.

1.1.1 Cas d'une variable discrète

Pour une variable statistique discrète X, on peut calculer la moyenne arithmétique simple et pondérée.
La moyenne arithmétique simple est :

$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$ et $N = \sum_{i=1}^{k} n_i$

La moyenne arithmétique pondérée est :

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \sum_{i=1}^{k} f_i \, x_i \quad \text{puisque } f_i = \frac{n_i}{N}$$

Exemple :

La répartition 20 ménages selon le nombre d'enfants est :

0 1 3 1 4 5 5 5 6 7
2 1 4 4 4 5 2 1 1 2

Que l'on peut présenter dans le un tableau statistique suivant :

xi	ni	xi ni
0	1	0
1	5	5
2	3	6
3	1	3
4	4	16
5	4	20
6	1	6
7	1	7
total	20	63

$\bar{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N} =$ 63/20 = 3,15 enfants

Le nombre moyen d'enfants, pour cet échantillon, est de 3.15 enfants.

1.1.2 Le cas d'une variable continue

Soit X une variable quantitative continue, ses valeurs sont incluses dans des classes **[**$\mathbf{e_i}$, $\mathbf{e_{i+1}}$ **[.** Pour pouvoir calculer la moyenne arithmétique, on doit calculer le centre de ces classes **ci** = ($\mathbf{e_i}$+ $\mathbf{e_{i+1}}$)/2

La moyenne arithmétique pondérée est :

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \sum_{i=1}^{k} f_i \, c_i \quad \text{puisque} \quad f_i = \frac{n_i}{N}$$

Exemple :

salaire en euros	ni	fi	ci	ai	Ci ni
[0,1000[	200	0,5	500	1000	100 000
[1000,1500[	120	0,3	1250	500	150 000
[1500,2500[	60	0,15	2000	1000	120 000
[2500,3000[	20	0,05	2750	500	55 000
total	400	1			425000

$\bar{x} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \frac{425000}{400}$ =1062,5 euros

1.1.3 Calculs sur la moyenne arithmétique

Soit Y une nouvelle variable statistique et Y est une fonction de X.

Y=h (X) avec :

h : IR $-> IR$

$y_i = h(x_i)$

On a donc les valeurs de la nouvelle variable Y : $y_{i=} \frac{x_i - x_0}{a}$

y_i a les mêmes effectif ni que xi pour tout i=1,..k

$\bar{y} = h(\overline{x})$ donc $\bar{y} = \frac{\bar{x} - x_0}{a}$

Démonstration :

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{k} ni\ yi = \frac{1}{N}\sum_{i=1}^{k} ni\ (\frac{x_i - x_0}{a}) = \frac{1}{a}\frac{1}{N}(\sum_{i=1}^{k} n_i\ x_i\ -\ x_0 \sum_{i=1}^{k} n_i)$$

$$\bar{y} = \frac{1}{a}\frac{1}{N}(\sum_{i=1}^{k} n_i\ x_i\ -\ x_0 N) = \frac{1}{a}\ (\bar{x} -\ x_0)$$

Comment choisir a et x_0 ?

- Cas discret : Pour x_0 on choisi le mode ou la médiane
  et a=1 si X ϵ ZI sinon a= PGCD($x_{i+1} - x_i$) pour tout i=1,..k-1

- Cas continu : x_0 est égal au centre de classe médiane (si le nombre de classes est impaire) et a = PGCD($c_{i+1} - c_i$) pour tout i=1,..k-1

  Dans tous les cas, on y prend des petite valeurs en valeurs absolue ce qui facilite les calculs.

1.1.4 Avantages et inconvénients de la moyenne

La moyenne arithmétique présente plusieurs avantages :

- Elle est relativement facile à déterminer ;
- Son calcul fait intervenir toutes les observations ;
- Elle est unique car chaque série n'a qu'une et une seule moyenne ;
- Quand on dispose de plusieurs échantillons qui observent la même variable, il est possible de calculer une moyenne générale à partir des moyennes des différents échantillons

La moyenne arithmétique présente aussi des inconvénients :

- Elle est sensible aux valeurs extrêmes, ce qui en fait un paramètre moins stable que la médiane ;
- Elle ne prend son sens que si elle est accompagnée d'une estimation de sa précision

1.2 La médiane

La médiane *Me* est telle que l'effectif des observations dont les modalités sont inférieures à *Me* est égal à l'effectif des observations dont les modalités sont supérieures à *Me*. Les valeurs étant rangées par ordre croissant, c'est la valeur de la variable qui sépare les observations en deux groupes d'effectifs égaux.

Cette définition n'a de sens que si les modalités sont toutes ordonnées.

Dans le cas d'une variable qualitative il est parfois possible de choisir un ordre. Exemple : niveau d'études scolaires : école primaire < 1er cycle < CAP < BEP < Bac < BTS < DEUG < ....

Une variable quantitative *X* doit être définie dans IR.

1.2.1 Cas d'une variable discrète

Pour déterminer la médiane, on doit ranger la série statistique des N données par ordre croissant ou décroissant.

- si la série est de taille N impaire, la médiane est de rang $\frac{N+1}{2}$ et la médiane est la valeur qui lui correspond.
- si la série est de taille N paire, la médiane est la demi-somme des données de rang $\frac{N}{2}$ et $\frac{N}{2}+1$.

Exemple 1 : N = 9

Trouvez la médiane des notes de 9 étudiants en statistique :

10 11 8 9 15 17 7 14 13

On la range : 7 8 9 10 11 13 14 15 17

Puis on a le rang de la médiane est $\frac{N+1}{2} = 5$ donc *Me = 11*

Exemple 2 : N = 10

Trouvez la médiane des notes de 10 étudiants en mathématiques :

10 11 8 9 15 17 7 14 12 16

On la range : 7 8 9 10 11 12 14 15 16 17

Le rang de la médiane est entre 5 eT 6 et *Me = (11+12)/2 = 11,5*

Remarque :

On peut trouver la médiane Me à partir de la courbe des fréquences cumulées de la variable X. Puisque la médiane divise la série en deux parties égales donc F(Me) = 0,5. On repère la valeur 0,5 sur l'axe des ordonnées et on détermine le point A ayant les coordonnées A(Me, F(Me)).

1.2.2 Cas d'une variable continue

Pour une série regroupée en classes c'est-à-dire à caractère continu, la médiane correspond à la valeur du caractère ayant une fréquence cumulée croissante 0,5. La classe à laquelle appartient la médiane est appelée classe médiane.

Après avoir déterminé la classe médiane, on calcule la Me :

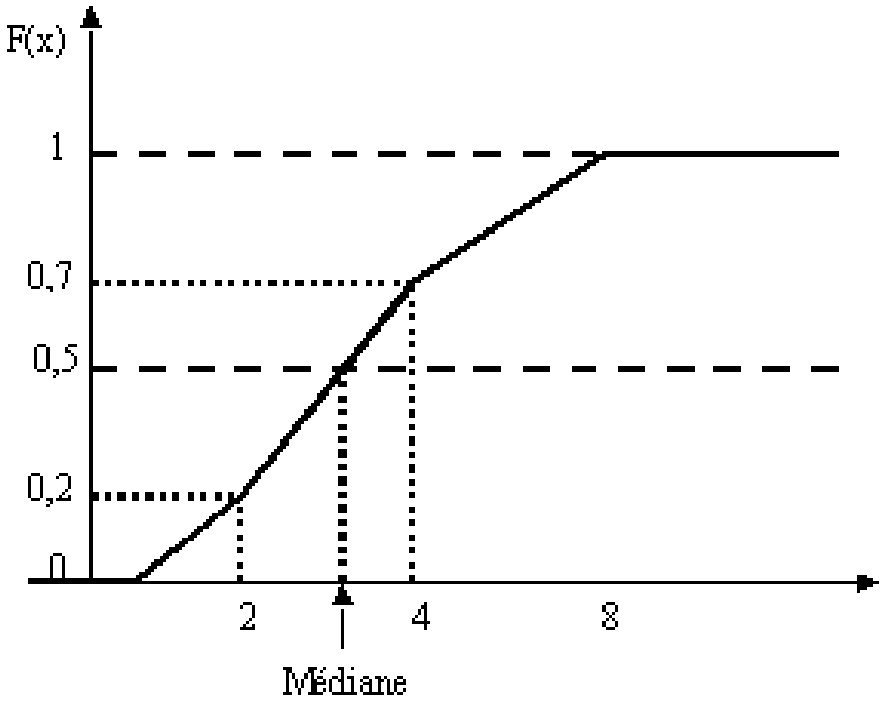
$$Me = (e_{i+1} - e_i).\frac{0{,}5 - F(ei)}{F(e_{i+1}) - F(e_i)} + e_i$$

On peut aussi la calculer en utilisant les effectifs cumulés croissant :

la médiane correspond à la valeur du caractère ayant un effectif cumulé croissant N/2 . La classe à laquelle appartient la médiane est appelée classe médiane.

Après avoir déterminé la classe médiane, on calcule la Me :

$$Me = (e_{i+1} - e_i).\frac{\frac{N}{2} - \vec{N}(ei)}{\vec{N}(e_{i+1}) - \vec{N}(e_i)} + e_i$$



Polygone cumulatif

Remarque :

On peut trouver la médiane graphiquement :

Une fois que la classe médiane est trouvée, on applique la théorie de THALES pour calculer Me :

$$\frac{\boldsymbol{F(Me)-F(ei)}}{\boldsymbol{F(ei+1)-F(ei)}} = \frac{\boldsymbol{Me-ei}}{\boldsymbol{ei+1-ei}}$$

On déduit :

$$\text{Me} = (e_{i+1} - e_i).\frac{0{,}5-F(ei)}{F(e_{i+1})-F(e_i)} + e_i$$

1.2.3 Avantages et inconvénients de la médiane

- Elle est facile à déterminer ;

- Elle part du classement de toutes les observations, donc elle est représentative de l'ensemble ;
- Elle est unique pour une série statistique ;
- Elle est insensible aux valeurs extrêmes, ce qui en fait un paramètre remarquablement stable

Elle présente un inconvénient ; lorsqu'on dispose de dispose de plusieurs échantillons qui observent la même variable, il n'est possible de calculer une médiane générale à partir des médianes partielles

1.3 Le Mode Mo

Le mode d'une variable statistique est la modalité ayant le plus grand effectif ou la plus grande fréquence :

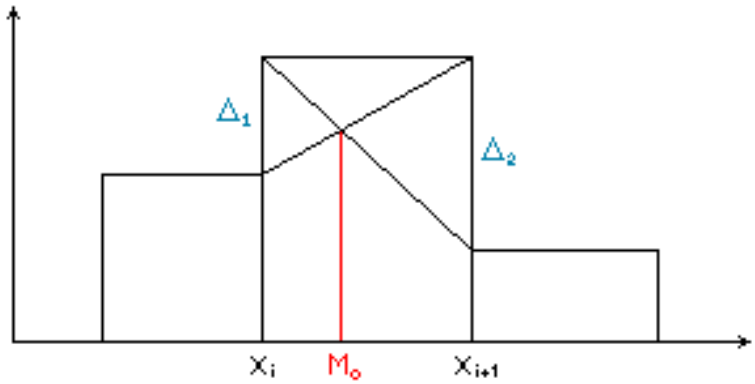
$f(Mo) = Max(fi)$;i ϵ[1, k]

1.3.1 Cas d'une variable discrète

Il est parfaitement défini pour une variable qualitative ou une variable quantitative discrète. Le mode est la valeur qui correspond au plus grand effectif.

1.3.2 Cas d'une variable continue

Pour une variable quantitative continue nous parlons de **classe modale** : c'est la classe dont la densité de fréquence est maximum.
Si les classes ont même amplitude la densité est remplacée par l'effectif ou la fréquence et nous retrouvons la définition précédente.
Nous définissons le **mode**, pour une variable quantitative continue, en tenant compte des densités de fréquence des 2 classes adjacentes par la méthode suivante.



La classe modale [e*i*, e*i* + 1 [= [*xi*, *xi* + 1 [étant déterminée, le mode *Mo* vérifie :

$$\frac{Mo-xi}{\Delta_1}=\frac{x_{i+1}-Mo}{\Delta_2}$$

Dans une proportion, on ne change pas la valeur du rapport en additionnant les numérateurs et en additionnant les dénominateurs :

$$\frac{Mo-xi}{\Delta_1}=\frac{x_{i+1}-Mo}{\Delta_2}=\frac{x_{i+1}-xi}{\Delta_{1+\Delta_2}}$$ d'où :

$$Mo = x_i + \frac{\Delta_1}{\Delta_1+\Delta_2}(x_{I+1}-x_i)$$

Avec : $\Delta_1 = n_o - n_{[xi-1,xi]}$ c'est la différence entre l'effectif de la classe modale et l'effectif de la classe qui la précède ;

Et $\Delta_2 = n_o - n_{[xi+1,xi+2]}$ c'est la différence entre l'effectif de la classe modale et l'effectif de la classe qui la suit :

$$Mo = x_i + \frac{n_o - n_{[xi-1,xi]}}{2n_o - n_{[xi-1,xi]} - n_{[xi+1,xi+2]}}(x_{I+1}-x_i)$$

Remarque 1 :

Le choix entre le mode, la moyenne et la médiane est fonction du type de distribution que l'on observe :

•le mode fait apparaître le comportement le plus fréquent. Dans le cas d'une distribution fortement asymétrique, cette valeur peut être peu représentative du comportement de l'ensemble des observations effectuées.

•la moyenne donne la même importance à chaque observation. Elle est donc sensible aux observations extrêmes.

•la médiane, qui partage en deux la distribution, n'est pas sensible à l'importance de l'éloignement des données extrêmes. Elle est en ce sens plus robuste.

Remarque 2 :

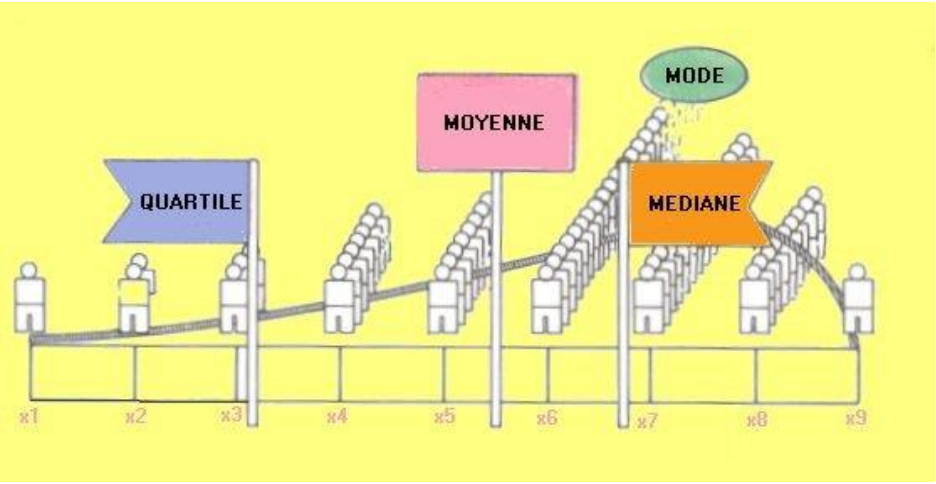
Il peut être intéressant de connaître ces trois indicateurs à la fois, car, ensemble, ils permettent déjà d'obtenir certaines précisions sur une distribution. En particulier, comme la moyenne est « tirée » vers les valeurs extrêmes, on sait que :

- si la moyenne est beaucoup plus basse que la médiane, quelques individus ont des valeurs de caractère beaucoup plus basses que l'ensemble des autres.
- si la moyenne est beaucoup plus haute que la médiane, quelques individus ont des valeurs de caractère beaucoup plus hautes que l'ensemble des autres.

Remarque 3 :

La distribution asymétrique :

Si $\bar{x}$ < Me < Mo , la distribution est étalée à gauche :



Si Mo< Me < $\bar{x}$, la distribution est étalée à droite

Si $\bar{x}$ = Me = Mo, la distribution est symétrique

Remarque 4 :

Il existe une relation mathématique entre la moyenne arithmétique, le mode et la médiane :

Me - $\bar{x}$ = (Mo - $\bar{x}$)/3

1.4 La moyenne géométrique

À côté de la moyenne arithmétique que nous avons vue dans ce cours, il existe d'autres moyennes.

Cas discret : $\bar{X}_g = \sqrt[N]{\prod_{i=1}^{N} x_i}$ et

$\log \bar{x}_g = \frac{\log x_1 + \log x_1 + \ldots . \log x_N +}{N} = \frac{1}{N} \sum_{i=1}^{N} log x_i$

Cas continu :

$\bar{X}_g = \sqrt[N]{\prod_{i=1}^{k} c_i^{n_i}}$ avec N=$\sum_{i=1}^{K} n_i$

1.5 La moyenne harmonique

Cas discret : $\bar{X}_H = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i}} = \frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}$

Où

$\frac{1}{\overline{X_H}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i}$

Cas continu :

$\frac{1}{\overline{X_H}} = \frac{1}{N}\sum_{i=1}^{N}\frac{n_i}{c_i}$

1.6 La moyenne quadratique

La moyenne quadratique est une moyenne d'une série de valeurs, définie comme la racine de la moyenne des carrés des valeurs :

$$\overline{x_q^2} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2$$

1.7 Relation entre les différentes moyennes

Pour une variable statistique X, les différentes moyennes, harmonique, géométrique, arithmétique, quadratique, sont liées par la relation :

$$\bar{X}_H \leq \bar{X}_g \leq \bar{X} \quad \leq \bar{X}_q$$

Il y a égalité si, et seulement si, toutes les valeurs de X sont égales.
La moyenne géométrique est bien adaptée à l'étude des phénomènes de croissance.
La moyenne harmonique est utilisée pour les calculs d'indices économiques.

La moyenne géométrique est bien adaptée à l'étude des phénomènes de croissance.
La moyenne harmonique est utilisée pour les calculs d'indices économiques.

2. Paramètres de dispersion
3. paramètres de forme