CHAPITRE 02 : PARAMÈTRES DE TENDANCE CENTRALE ET DE DISPERSION

INTRODUCTION

Après avoir appris comment représenter une série statistique à une seule dimension dans un tableau statistique et graphiquement, nous allons résumer et synthétiser cette série à l'aide des valeurs caractéristiques comme la moyenne, le mode, la médiane, écart-type…

Il existe 3 types de paramètres :

- Paramètres de position (ou de tendance centrale)
- Paramètres de dispersion
- Paramètres de forme (asymétrie, aplatissement, concentration)

SECTION 01 : PARAMÈTRES DE POSITION CENTRALE

I. MOYENNE ARITHMÉTIQUE

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

1. Cas d'une variable discrète ou caractère discret

Pour une variable statistique discrète X, on peut calculer la moyenne arithmétique simple ou pondérée. Pour les données détaillées, on calcule une **moyenne simple** est :

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

Avec :

$$N = \sum_{i=1}^{N} n_i$$

Pour les données agrégées et regroupées dans un tableau statistique, on calcule une **moyenne pondérée** est :

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

Avec :

$$N = \sum_{i=1}^{k} n_i$$

Et

K : le nombre de modalités statistiques.

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

$$\bar{x} = \sum_{i=1}^{k} f_i \, x_i$$

Puisque : $f_i = \frac{n_i}{N}$

Exemple : Répartition de 20 logements selon leur nombre de pièces

Nombre de pièces **xi**	nombre de logements **ni**	**ni . xi**
1	5	5
2	3	6
3	1	3
4	5	20
5	4	20
6	2	16
Total	20	66

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{N} = \frac{66}{20} = 3{,}3 \text{ pièces}$$

Le nombre moyen des pièces s'élève à 3,3 pièces.

2. Cas d'une variable continue ou caractère continu

Soit X une variable quantitative continue, ses valeurs sont incluses dans des classes $[e_i, e_{i+1}[$.

Pour pouvoir calculer la moyenne arithmétique, on doit calculer le centre de ces classes :

$$c_i = \frac{e_i + e_{i+1}}{2}$$

La moyenne est :

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

K : est le nombre de classes.

$f_i = \frac{n_i}{N}$

On peut écrire : $\bar{x} = \sum_{i=1}^{k} f_i\, x_i$

Exemple : Répartition de 425 salariés, d'une entreprise Z, selon leurs salaires nets (en dollars américains) en 2010.

Montant du salaire (en dollars Américains) **xi**	nombre de salariés **ni**	Centre de classe **ci**	**ci ni**
[750 - 1000 [	150	875	131 250
[1000 - 1500 [	100	1250	125 000
[1500- 2500 [	99	2000	198 000
[2500 - 3000 [	50	2750	137 500
[3000 - 5000 [	26	4000	104 000
Total	**425**	---	**695 750**

Source : Réalisé par nos soins

$c_1 = \frac{e_1+e_2}{2} = \frac{750+1000}{2} = 875$

$\bar{x} = \frac{\sum_{i=1}^{k} n_i c_i}{N}$ = **695 750** / **425** = 1637,06 dollars.

Le salaire moyen s'élève à 1637,06 dollars américains.

3. Changement de variable et calculs sur la moyenne arithmétique

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

$$Y = h(X) \ \ avec \ \ h : \ \mathbb{R} \rightarrow \mathbb{R}$$

Les valeurs yi sont en fonction de valeurs initiales xi :

$$y_i = h(x_i)$$

$y_i = \frac{x_i - x_0}{a}$

Les effectifs n_i restent identiques (le même échantillon) que x_i pour tout i=1,…,k. On a :

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

Démonstration :

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

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

Comment choisir a et x_0 ?

- Cas discret : Pour x_0, on choisit le mode ou la médiane et $a = 1$ si $X \in \mathbb{Z}$ 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 et $a = PGCD(c_{i+1} - c_i)$ pour tout i=1, ..., k-1.

Dans tous les cas, on y prend des petites valeurs en valeurs absolues ce qui nous facilitera les calculs.

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 observant la même variable, il est possible de calculer une moyenne générale à partir des moyennes des différents échantillons (la moyenne des moyennes).

La moyenne arithmétique présente notamment 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.

II. MÉDIANE

La médiane $\boldsymbol{Me}$ est telle que l'effectif des observations dont les modalités sont inférieures à $\boldsymbol{Me}$ est égal à l'effectif des observations dont les modalités sont supérieures à $\boldsymbol{Me}$. Autrement dit, 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.

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 : Répartition des 9 personnes selon âge exact en années (N = 9).

10 11 8 9 15 17 7 14 13

Étape 1 : On classe les valeurs par ordre croissant : 7 8 9 10 11 13 14 15 17

Étape 02 : On cherche le rang de la médiane $\frac{\mathrm{N+1}}{2} = 5$

Étape 03 : la valeur de la médiane est $\mathbf{Me = 11}$ **ans**

Exemple 2 : Répartition 10 immeubles selon leurs nombre d'appartements (N =10) :

10 11 8 9 15 17 7 14 12 16

Étape 1 : On classe les valeurs par ordre croissant : 7 8 9 10 11 12 14 15 16 17

Étape 02 : On cherche le rang de la médiane $\frac{\mathrm{N}}{2} = 5$ et $\frac{\mathrm{N}}{2} + 1 = 6$

Étape 03 : la valeur de la médiane est $\mathbf{Me} =$ (11 + 12)/2 = 11,5 appartements

Remarque

On peut trouver la médiane $\boldsymbol{Me}$ à partir du diagramme cumulatif (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 $A(Me, F(Me))$. C'est aussi le point d'intersection entre F et G et

$F(Me) = G(\ Me) = \ 0{,}5$.

2. Le 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 :

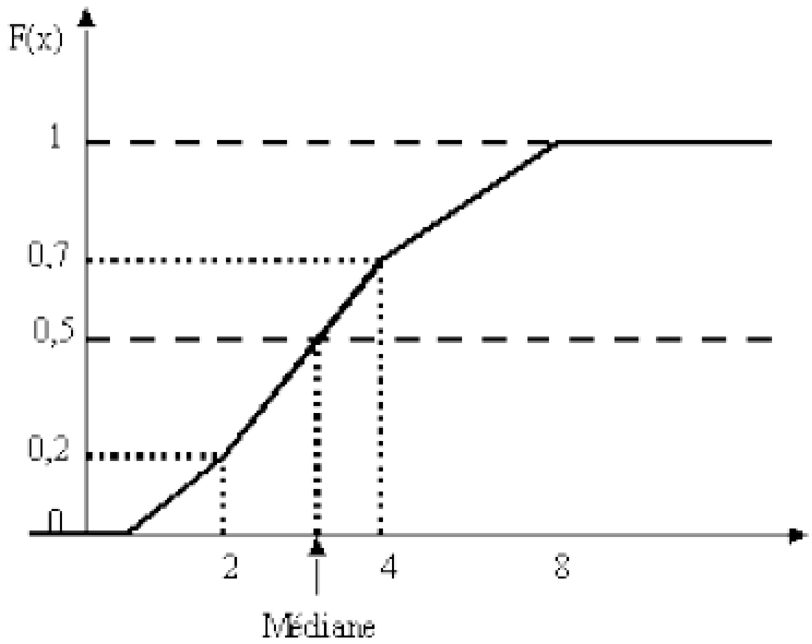
$$Me = (e_{i+1} - e_i).\frac{0.5 - F(e_i)}{F(e_{i+1}) - F(e_i)} + \ e_i$$

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

La médiane correspond à la valeur du caractère ayant un effectif cumulé croissant $\frac{N}{2}$. La classe à laquelle appartient la médiane est appelée classe médiane.

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

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



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.

III. LE MODE STATISTIQUE

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

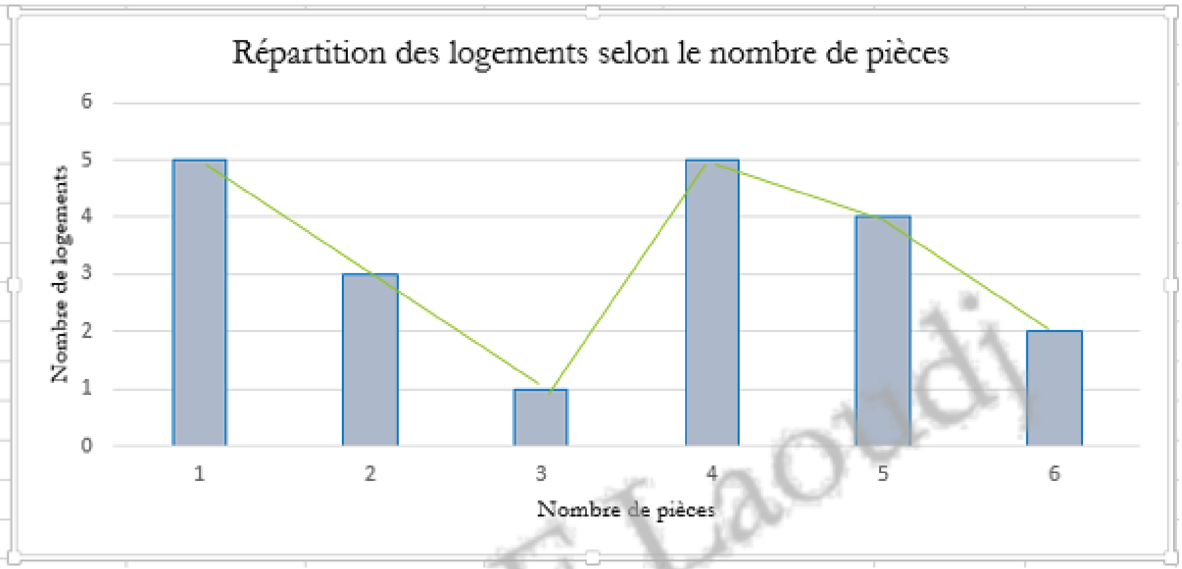
$$f(Mo) = Max(f_i);\ \ i \in [1, k]$$

Ou bien $$f(Mo) = Max(n_i);\ \ i \in [1, k]$$

1. Le cas d'une variable discrète

Le mode est la valeur qui correspond au plus grand effectif.

Exemple : Répartition de 20 logements selon leur nombre de pièces



Source : Réalisé par nos soins

Mo = 1 pièce ; modalité a 5 logements .

M o = 5 pièces ; modalité a 5 logements.

Remarque : On constate qu'une série statistique peut avoir plusieurs modes.

2. Le cas d'une variable continue

Pour une variable quantitative continue, on cherche d'abord la classe modale qui correspond à l'effectif le plus élevé (si les amplitudes sont égales) sinon la classe modale est celle qui correspond à la densité maximale.

Nous définissons le **mode**, pour une variable quantitative continue, en tenant compte des densités de fréquence des 2 classes adjacentes par la méthode ci-dessous.

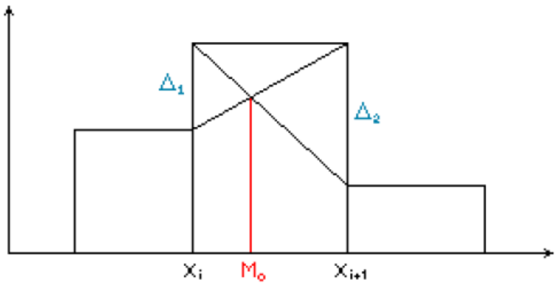
Si les amplitudes des classes sont égales :

La classe modale $[e_i, e_{i+1}[= [x_i, x_{i+1}[$ étant déterminée, le mode Mo vérifie :

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

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

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



d'où : $Mo = x_i + \frac{\Delta_1}{\Delta_1 + \Delta_2}(x_{i+1} - x_i)$

avec $(x_{i+1} - x_i)$ est l'amplitude de la classe modale.

$\Delta_1 = n_0 - n_{[x_{i-1}, x_i]}$ est la différence entre l'effectif de la classe modale et l'effectif de la classe qui la précède.

$\Delta_2 = n_0 - n_{[x_{i+1}, x_{i+2}]}$ est la différence entre l'effectif de la classe modale et l'effectif de la classe qui la suit :

$$Mo = x_i + \frac{n_0 - n_{[x_{i-1}, x_i]}}{2n_0 - n_{[x_{i-1}, x_i]} - n_{[x_{i+1}, x_{i+2}]}}(x_{i+1} - x_i)$$

Si les amplitudes des classes sont inégales :

Prenons l'exemple précédent de la répartition de 425 salariés, d'une entreprise Z, selon leurs salaires nets (en dollars américains) en 2010.

L'idée est d'utiliser **l'amplitude de base** , ici est **250 et le rapport de correction :**

pour la première classe : 250*1 = 250 , rapport de correction est 1 = 250/250

deuxième classe : 250*2 = 500 , rapport de correction est 2 = 500/250

troisième classe : 250* 4 = 1000 , rapport de correction est 4 = 1000/250

….

Les rapports obtenus vont servir à corriger les effectifs (deuxième méthode de corriger les effectifs)

Montant du salaire (en dollars Américains) **xi**	nombre de salariés **ni**	amplitude **ai**	Rapport de correction amplitude	Effectifs corrigés (2ème méthode)	Première méthode de correction n_i^c
[750 - 1000 [	150	250	1	150	0.6
[1000 - 1500 [	100	500	2	50	0.2
[1500- 2500 [	99	1000	4	24.75	0.099
[2500 - 3000 [	50	500	2	25	0.1
[3000 - 5000 [	26	2000	8	3.25	0.013
Total	425				

Source : Réalisé par nos soins

La classe modale est [750 , 1000[.

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

$$Mo = 750 \quad + \frac{150 - 0}{(150 - 0) \quad + (150 - 50)}(1000 - 750)$$

$Mo = 900$ dollars américans

Deuxième méthode avec les n_i^c

$$Mo = 750 \quad + \frac{0.6 - 0}{(0.6 - 0) \quad + (0.6 - 0.2)}(1000 - 750)$$

$Mo = 900$ dollars américans

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.

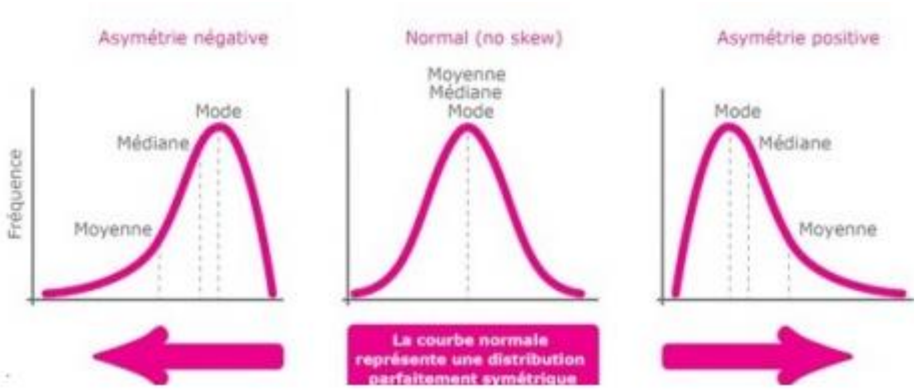
Remarques:

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 statistiques 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.

- Si : $\bar{x} < Me < Mo$, la distribution est alors étalée à gauche (asymétrie positive).
- Si : $Mo < Me < \bar{x}$, la distribution est étalée à droite (asymétrie négative)
- Si $\bar{x} = Me = Mo$, la distribution est symétrique.



Source : https://www.ecom-store.fr/469-panier-moyen-median-e-commerce/

Remarque 4 :

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

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

IIII . AUTRES MOYENNES

1. Moyenne géométrique

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

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

$$\bar{X}_g = exp\left(\frac{\sum_{i=1}^{k} \text{ni} \ln x_i}{N}\right)$$

Cas continu : : $\bar{X}_g = \sqrt[N]{\prod_{i=1}^{k} c_i^{n_i}}$

Avec $N = \sum_{i=1}^{k} n_i$

2. Moyenne harmonique

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

d'où $\frac{1}{\overline{X}_H} = \frac{1}{N}\sum_{i=1}^{k}\frac{ni}{x_i}$

Cas continu : $\frac{1}{\overline{X}_H} = \frac{1}{N}\sum_{i=1}^{k}\frac{n_i}{c_i}$

3. Moyenne quadratique

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

$$\overline{\boldsymbol{x^2}}_{\boldsymbol{q}} = \frac{\boldsymbol{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{k}} \boldsymbol{n_i\, x_i^2}$$

4. Relation entre les différentes moyennes

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

$$\overline{X_H} < \overline{X_g} < \bar{X} < \overline{X_q}$$

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.