I.3.2- Les mesures de tendance centrale

Les mesures de tendance centrales sont des valeurs statistiques permettant généralement de résumer l'ensemble de valeurs prises par la variable considérée. Qu'elles soient groupées ou non groupées par valeurs ou par classes, les variables quantitatives peuvent être utilement résumées par des caractéristiques dites de « tendance centrale ».

Ces valeurs centrales sont les moyennes, la médiane et le mode. Leur détermination tient compte généralement des données ; cas des données non groupées et le cas des données regroupées (soit par valeurs, soit par classes).

I.3.2.1- La moyenne

A – La moyenne arithmétique

1) La moyenne arithmétique simple

Soit {x1, x2, …., xn} une série de chiffres. La formule de la moyenne arithmétique de cette série est donnée par :

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

Cette moyenne est appelée moyenne arithmétique simple, ou :

- Xi sont les valeurs de la variable considérée
- n est le nombre de valeurs de la variable ou taille de la série de données.

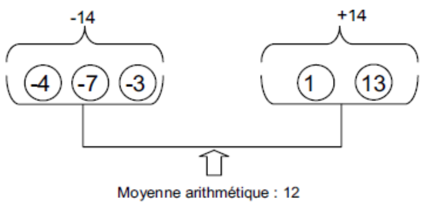
Il s'agit donc, du nombre de valeurs de la série de chiffres sans chercher de savoir s'il s'agit d'une population ou d'un échantillon.

Exemple : Soit la série de chiffres {8, 5, 9, 13, 25}. La moyenne arithmétique de cette série de chiffres se calcule ainsi :

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

$$\bar{x} = \frac{8+5+9+13+25}{5} = \frac{60}{5} = 12$$

Quand on soustrait la moyenne arithmétique à chacun des chiffres de la série, on observe la propriété suivante :



C'est-à-dire, la somme des écarts à la moyenne est nulle :

$$(-4)+(-7)+(-3)+(+1)+(+13)=0$$

Ou, ce qui revient au même, mais est plus imagé, la somme des écarts positifs est égale à la somme des écarts négatifs, au signe près.

2) La moyenne arithmétique pondérée

1) effectifs groupés par valeurs

Exemple 1 : Soit la série de chiffres {8, 13, 5, 8, 5, 9, 13, 25, 13, 9}. Certains chiffres, comme le 8, le 9 ou le 13 sont répétés. Dans ce

Soit {x1, x2, …., xn} une série de chiffres et {n1, n2, ….,nh} les effectifs correspondants.

La formule de la moyenne arithmétique dans ce cas est donnée par :

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

Cette moyenne est appelée moyenne arithmétique pondérée, ou :

- Xi sont les valeurs de la variable considérée.
- ni est le nombre de répétition ou effectif d'une valeur de la variable.
- n est le nombre de valeurs de la variable ou taille de la série de données.

1) effectifs groupés par classe de valeurs

Soit {x1, x2, …., xn} une série de chiffres et {n1, n2, ….,nh} les effectifs correspondants. Supposons que l'on regroupe les valeurs en classes, dont le centre noté ci et égale à :

$$c_i = \frac{x_1 + x_2}{2}$$

Ou

x1 et x2 sont les extrémités de classe i.

La formule de la moyenne arithmétique dans ce cas est donnée par :

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

- ci sont les centres de classes de valeurs de la variable considérée.
- ni est le nombre de répétition ou effectif d'une valeur de la variable de la classe.
- n est le nombre de valeurs de la variable ou taille de la série de données.

cas la moyenne est donnée par l'expression suivante :

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

On peut simplifier la présentation de cette série en regroupant les données par valeurs (tableau 2). La troisième ligne est le produit des deux premières lignes (xi et ni).

xi	5	8	9	13	25	Total
ni	2	2	2	3	1	10
$x_i \times n_i$	10	16	18	39	25	108

$$\bar{x} = \frac{\sum_{i=1}^{n} n_i x_i}{n} = \frac{108}{10} = 10.8$$

Exemple 2 : Soit la série de chiffres {8, 13, 5, 8, 5, 9, 13, 25, 13, 9, 35, 44, 54, 28}. Ces chiffres peuvent être regroupés en 3 catégories comme dans le tableau 3 ci-dessous.

xi	ni	ci	$c_i \times n_i$
[5 ; 13[	6	9	54
[13 ; 28[	4	20.5	82
[28 ; 54[	4	41	164
Total	**14**		**300**

Ensuite, il faut calculer le centre de chaque classe, ci, et le multiplier par son effectif c'est-à-dire $c_i \times n_i$, appliquer la formule de la moyenne pondérée.

$$\bar{x} = \frac{\sum_{i=1}^{n} n_i c_i}{n} = \frac{300}{14} = 21.43$$

B – La moyenne quadratique

1) La moyenne quadratique simple

Parfois, on souhaite obtenir une caractéristique de tendance centrale ayant une valeur positive là où le calcul de la moyenne arithmétique simple aurait donné zéro. Dans ce cas la moyenne quadratique est une autre mesure utilisée pour représenter la

Exemple : Soit la série de chiffres {-4, -2, 0, 2, 4}. Si l'on calcule la moyenne arithmétique simple on obtient zéro.

Mais si en utilise le carré des toutes les valeurs der la série et en prenant la racine carré du totale alors on aura une valeur efficace pour cette grandeur :

valeur efficace d'une grandeur qui varie au fil du temps, comme une tension électrique ou une intensité sonore. Elle permet également de comparer des ensembles de données qui peuvent avoir des échelles de mesure différentes.

Soit {x1, x2, ….,xh} une série de chiffres.

La moyenne quadratique est égale à la racine carrée de la moyenne des carrés des valeurs de l'ensemble de données.

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

1) La moyenne géométrique pondérée

Si les valeurs {x1, x2, …., xh} sont répétées et {n1, n2,…., nh} les effectifs correspondants, l'expression de la moyenne quadratique est écrite sous la forme suivant :

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (n_i x_i^2)}$$

où

- xi est chaque valeur de l'ensemble de données.
- ni est le nombre de répétition ou effectif d'une valeur de l'ensemble de données.
- n est le nombre de valeurs de la variable ou taille de la série de données.

Lorsque les valeurs sont regroupées en classes, il faut calculer les centres de classes et appliquer ensuite la formule précédente en remplaçant xi par ci.

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

$$Q = \sqrt{\frac{1}{5}((-4^2) + (-2)^2 + -(0)^2 + (2)^2 + (4)^2)}$$

$$Q = \sqrt{\frac{40}{5}} = 2.83$$

Exemple 2 : Soit le tableau de données suivant :

xi	25	8	4	12	Total
ni	10	16	25	20	71

Il suffit de rajouter deux colonnes, une pour x_i^2 et la deuxième pour $n_i \times x_i^2$

xi	ni	x_i^2	$n_i \times x_i^2$
25	10	625	6250
8	16	64	1024
4	25	16	400
12	20	144	2880
Total	**71**		**10554**

En appliquant la formule, on obtient :

$$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (n_i x_i^2)} = \sqrt{\frac{10554}{71}} = 12.19$$

C – La moyenne géométrique

1) La moyenne géométrique simple

La moyenne géométrique est une mesure de tendance centrale utilisée pour calculer les taux de croissance moyen d'une série chronologique, en particulier lorsqu'il y a des variations importantes dans les données.

Exemple : Soit la série de chiffres {8, 5, 9, 13, 25}. La moyenne géométrique de cette série est égale à :

$$G = [8 \times 5 \times 9 \times 13 \times 25]^{\frac{1}{5}}$$

$$G = \sqrt[5]{117000} = 10.32$$

Soit {x1, x2, ….,xn} une série de chiffres. La formule de la moyenne géométrique simple de cette série est donnée par :

$$G = \left[\prod_{i=1}^{n} x_i\right]^{\frac{1}{n}}$$

1) La moyenne géométrique pondérée

Si les valeurs {x1, x2, …., xh} sont répétées et {n1, n2, ….,nh} sont les effectifs correspondants. La formule de la moyenne géométrique pondérée de cette série devient :

$$G = \left[\prod_{i=1}^{n} x_i^{n_i}\right]^{\frac{1}{n}}$$

où

- xi est chaque valeur de l'ensemble de données.
- ni est le nombre de répétition ou effectif d'une valeur de l'ensemble de données.
- n est le nombre de valeurs de la variable ou taille de la série de données.

Exemple 2 : Soit le tableau de données suivant :

xi	25	8	4	12	Total
ni	10	16	25	20	71

Pour calculer la moyenne géométrique pondérée, on peut passer par les logarithmes népériens (ln).

xi	ni	$n_i \times \ln(x_i)$
25	10	32.18
8	16	32.27
4	25	34.66
12	20	49.70
Total	**71**	**149.81**

En appliquant la formule, on obtient :

$$\ln G = \ln\left[\prod_{i=1}^{n} x_i^{n_i}\right]^{\frac{1}{n}}$$

$$\ln G = \frac{1}{71}[10\ln 25 + 16\ln 8 + 25\ln 4 + 20\ln 12]$$

$$\ln G = \frac{1149.81}{71} = 2.11$$

$$G = \mathrm{e}^{2.11} = 8.24$$

D – La moyenne harmonique

1) La moyenne harmonique simple

La moyenne harmonique est une mesure de tendance centrale utilisée lorsque les taux de variation sont importants et où la moyenne arithmétique pourrait donner des résultats trompeurs. Elle est notamment utilisée en finance pour calculer des taux de rendement moyens sur une période donnée.

Soit {x1, x2, …., xn} une série de chiffres. La formule de la moyenne harmonique simple de cette série est donnée par :

$$H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

1) La moyenne harmonique pondérée

Exemple : Soit la série de chiffres {8, 5, 9, 13, 25}. La moyenne harmonique de cette série est égale à :

$$H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} = \frac{5}{\frac{1}{8}+\frac{1}{5}+\frac{1}{9}+\frac{1}{13}+\frac{1}{25}}$$

$$H = \frac{5}{0.553} = 9.04$$

Exemple 2 : Soit le tableau de données suivant :

xi	25	8	4	12	Total
ni	10	16	25	20	71

Si les valeurs {x1, x2, …., xh} sont répétées et {n1, n2, …., nh} sont les effectifs correspondants. La formule de la moyenne harmonique pondérée de cette série devient :

$$H = \frac{n}{\sum_{i=1}^{n} \frac{n_i}{x_i}}$$

où

- xi est chaque valeur de l'ensemble de données.
- ni est le nombre de répétition ou effectif d'une valeur de l'ensemble de données.
- n est le nombre de valeurs de la variable ou taille de la série de données.

Pour calculer la moyenne dans ce cas, il suffit de rajouter une colonne, pour n_i/x_i.

xi	ni	n_i/x_i
25	10	0.4
8	16	2
4	25	6.25
12	20	1.66
Total	**71**	**10.62**

En appliquant la formule, on obtient :

$$H = \frac{n}{\sum_{i=1}^{n} \frac{n_i}{x_i}} = \frac{71}{\frac{10}{25} + \frac{16}{8} + \frac{25}{4} + \frac{20}{12}}$$

$$H = \frac{71}{0.4 + 2 + 6.25 + 1.66} = \frac{71}{10.31}$$

$$H = \frac{71}{10.31} = 6.88$$

I.3.2.2- La médiane

La médiane d'une série de données, est une mesure statistique qui divise cette série, préalablement classée, en deux séries aux effectifs égaux. Plus précisément, c'est la valeur qui se situe au centre d'un ensemble de données lorsqu'elles sont triées par ordre croissant ou décroissant.

La médiane est souvent utilisée pour représenter la valeur centrale d'une série de données plutôt que la moyenne arithmétique, car elle est moins sensible aux valeurs extrêmes (outliers) qui pourraient fausser la moyenne. Elle est également utile lorsque les données sont présentées sous forme de catégories ordinales ou nominales plutôt que numériques.

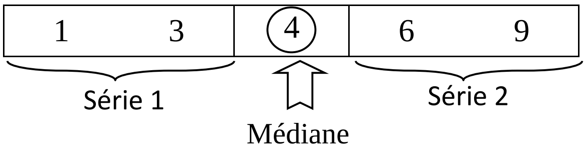
Le mode de calcul de la médiane dépend du type de données. On distinguera quatre cas :

1) Séries de valeurs non répétées (nombre de valeur impair)

Lorsque les données d'une série sont répétées le nombre d'observations est impair, la médiane est la valeur qui se trouve exactement au milieu de la série de données. Pour ce faire, il faut classer les données en ordre croissant, ensuite on localise la valeur qui partage l'effectif total en deux sous effectifs égaux en appliquant la formule: $\frac{n+1}{2}$

Exemple 01 : soit la série de données suivante : {1, 3, 4, 6, 9}.

la médiane dans ce cas est la valeur qui se trouve à la position $i = \frac{n+1}{2} = \frac{5+1}{2} = 3$:

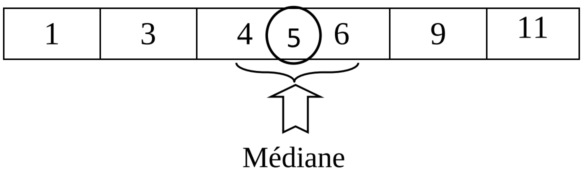


2) Série de valeurs non répétée (nombre de valeur pair)

Si le nombre d'observations est pair, la médiane est la moyenne des deux valeurs qui se situent au milieu de la série de données. Il s'agit ici, d'un intervalle médian définit par ces deux valeurs. La médiane est donc égale à la moyenne arithmétique simple de ces deux valeurs. Pour ce faire et après avoir classé les données en ordre croissant, on localise les deux valeurs définissant l'intervalle médian qui partage l'effectif total en deux sous effectifs égaux en appliquant la formule: $\frac{n+1}{2}$

Exemple 02 : soit la série de données suivante : {1, 3, 4, 6, 9, 11}.

L'effectif de cette série est pair (n = 6), la valeur médiane dans ce cas, est localisé en utilisant la formule $\frac{n+1}{2} = \frac{6+1}{2} = 3.5$, c'est a dire que la médiane sous trouve entre les valeurs 4 et 6 de la série.



Donc, la valeur de la médiane est la moyenne de 4 et 6, soit : $\frac{4+6}{2} = 5$.

3) Séries de valeurs répétées (effectifs groupés par valeurs)

Si les valeurs {x1, x2, …., xh} sont répétées et {n1, n2, …., nh} sont les effectifs correspondants. La médiane est la valeur qui se trouve exactement au milieu de la série de données. Pour la déterminer, il faut classer les données en ordre croissant, ensuite on localise la valeur qui partage l'effectif total en deux sous effectifs égaux en appliquant la condition : soit l'effectif cumulé $N_x = \frac{n}{2}$, ou la fréquence cumulé $F_x = 0{,}5$.

Cette procédure peut être réalisée graphiquement. En se basant sur la courbe cumulative des effectifs N(xi), on peut déterminer la médiane. Cette courbe « en escalier » a pour ordonnée les effectifs dont la valeur est strictement inférieure à xi.

Pour trouver la médiane, il faut localiser n/2=30/2=15 sur l'axe des ordonnées, puis tracer une flèche horizontale jusqu'à l'intersection avec la courbe. Puis il faut tracer une flèche verticale en direction de l'abscisse. On lit alors la valeur de la médiane qui, dans notre exemple, est égale à 11

Exemple 01: soit les données du tableau ci-dessous, les valeurs de la variable x sont déjà classées, ou la troisième colonne est celle des fréquences (fi) et la quatrième est celle des fréquences cumulées F(x). La cinquième colonne, est celle des effectifs cumulés N(x).

Alors, la valeur qui partage l'effectif total en deux sous effectifs égaux est celle dont l'effectif cumulé $N_x = \frac{n}{2} = \frac{30}{2} = 15$, ou la fréquence cumulé $F_x = 0{,}5$

La valeur médiane dans ces conditions est égale à $Me = 11$

	xi	ni	fi	Fx		Nx	
	2	2	0,067	0,067		2	
	8	3	0,100	0,167		5	
	9	4	0,133	0,300		9	
	10	4	0,133	0,433	0,5	13	
$Me =$	11	5	0,167	0,600	←	18	$\frac{n}{2} = 15$
	12	3	0,100	0,700		21	
	13	6	0,200	0,900		27	
	15	1	0,033	0,933		28	
	18	2	0,067	1		30	

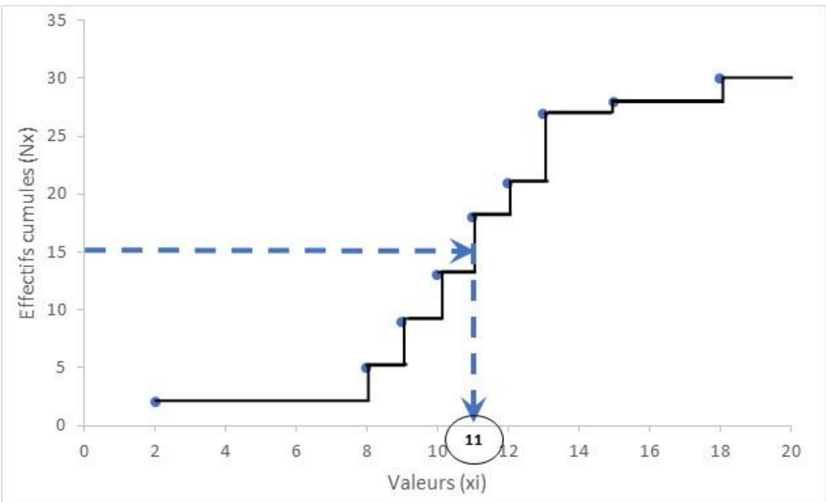


Figure : Détermination graphique de la médiane

4) Série de valeurs répétée (effectifs groupés par classes de valeurs)

Dans ce cas, le calcul de la médiane nécessite d'appliquer la formule suivante :

$$M_e = x_i^{inf} + a_i \left[\frac{\frac{N}{2} - N(x_{i-1})}{n_i} \right]$$

Où :

x_i^{inf} : Borne inférieure de la classe médiane.

$N(x_{i-1})$: Effectif cumulé strictement inférieur à la classe médiane (xi).

xi : Classe médiane

ai : Amplitude de la classe médiane

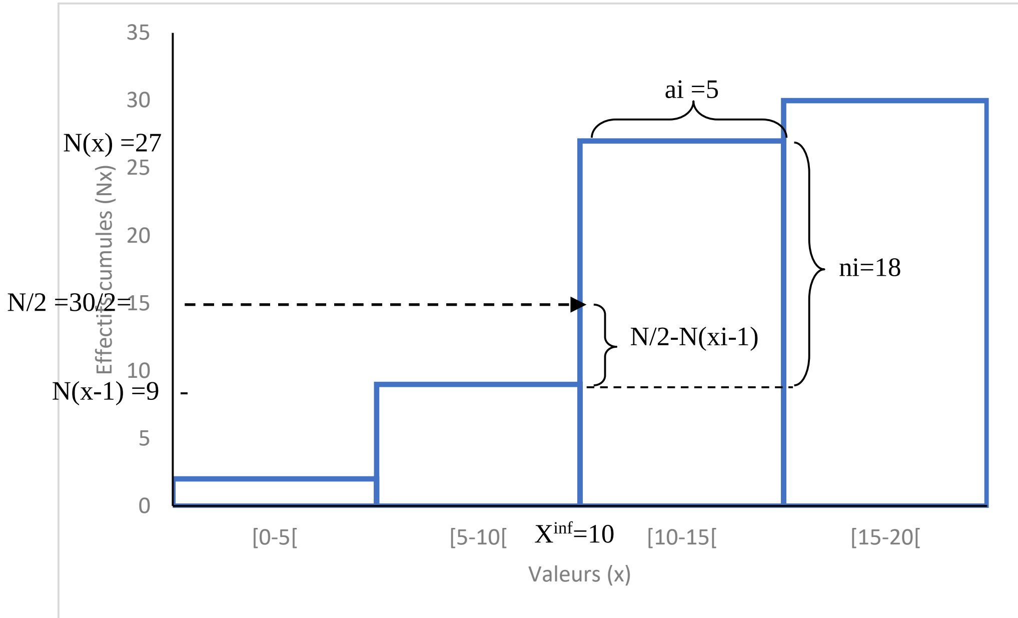
Exemple 02 : considérons les données de l'exemple précède, dont les valeurs de la variable X ont été groupées par classes de valeurs d'amplitudes égales (tableau ci-dessous).

xi	ni	N(xi)
[0-5[	2	2
[5-10[	7	9
[10-15[	18	27
[15-20[	3	30

Appliquons la formule de la médiane en l'interprétant par rapport à la figure ci-dessous qui représente le cumul des ni en ordonnée [soit N(xi)] et xi en abscisse :

$$M_e = 10 + 5.\left[\frac{\frac{30}{2} - 9}{18} \right] = 10 + 5.\left[\frac{6}{18} \right]$$

On trouve alors $M_e = 11,66$



I.3.2.3- Le mode

Le mode d'une série est défini comme étant la valeur la plus fréquente de cette série. Une série peut avoir plusieurs modes. Le calcul dépend du type de données.

1) Séries de valeurs non répétées

Lorsque les données d'une série sont non répétées, le mode n'existe pas car chaque valeur n'est répétée qu'une fois (la fréquence de chaque valeur est égale à 1).

Exemple 01 : soit la série de données suivante : {1, 3, 4, 6, 9}.

xi	1	3	4	6	9
ni	1	1	1	1	1

2) Série de valeurs répétée (effectifs groupés par valeurs)

Si les d'observations sont répétées, on peut les grouper par valeurs est le mode dans ce cas est la valeur la plus fréquente de cette série c'est a dire la valeur la plus répétée.

Un diagramme en bâtons permet aussi de confirmer la valeur modale.

3) Série de valeurs répétée (effectifs groupés par classes de valeurs)

Si les d'observations sont répétées, et sont grouper par classes valeurs. Le mode dans ce cas, est donne par la formule suivante :

$$M_o = x_i^{inf} + a_i \left[\frac{d_1}{d_1 + d_2} \right]$$

Ou,

x_i^{inf} : Borne inférieure de la classe médiane.

ai : Amplitude de la classe

$d_1 = n_i - n_{i-1}$ et $d_2 = n_i - n_{i+1}$

Dans le cas des données groupées par classes de valeurs avec des amplitudes inégales, il faut appliquer la même formule, mais la d1 et de d2 seront calcules d'une manière différente, car il faut remplacer les effectifs ni par les amplitudes corrigées hi=ni/ai.

ni est égale à 1 partout. Alors, dans ce cas il n'y a pas de mode.

Exemple 02 : Soit la série de données suivantes {8, 8, 8, 7, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6}.

xi	4	5	7	8
ni	5	4	1	3

La valeur la plus fréquente est le 4. Donc, Mo = 4

Exemple 02 : Soit les données (tableau ci-dessous) présentées par classes d'amplitudes égales.

xi	ni	N(xi)
[0-5[	2	2
[5-10[	7	9
[10-15[	18	27
[15-20[	3	30

Appliquons la formule donnant le mode en l'interprétant par rapport à la figure ci-dessous (histogramme correspondant aux données du tableau ci-dessus).

$$M_o = 10 + 5 \left[\frac{11}{11 + 15} \right] = 12{,}115$$

Dans ce cas la valeur du mode est : Mo = 12,115