Chapitre 2

Représentation numérique des données

2.1 Caractéristiques de tendance centrale (Paramètres de position)

En statistiques, un paramètre de tendance centrale est une valeur résumant une série statistique pour une variable quantitative ou ordinale. Les principales caractéristiques sont : le mode, la médiane, la moyenne (arithmétique, géométrique, harmonique, quadratique) et les quartiles.

2.1.1 Le mode

Le mode est la valeur de la variable statistique X qui correspond à l'effectif maximal ou à la fréquence maximale, elle est notée Mo

Exemple 2.1.1 *Distribution de 200 employés selon la nationalité*

Modalité x_i	*Allemande*	*Russe*	*Chinoise*	*Francaise*	*Alérienne*
effectif n_i	20	30	30	40	80

d'où Mo =Alérienne

Exemple 2.1.2 *Soient les valeurs suivants*

$6, 18, 2, 2, 4, 4, 5, 5, 5$

Le mode est $Mo = 5$.

Remarque 2.1.3 1) *Le mode peut-être calcule pour tous les types de variable quantitative et qualitative.*

2) *Une variable statistique peut avoir plusieurs modes.*

Exemple 2.1.4 *soit la série suivante :*

Modalité x_i	*1*	*3*	*9*	*11*	*12*
effectif n_i	5	17	1	17	3

cette série possède deux modes : $Mo_1 = 3$ *et* $Mo_2 = 11$

Cas d'une série continue

Si X est continue on parle de la classe modale. Si la classe modale $[a, b[$, le mode Mo de calcule par la formule suivante :

$$Mo = a + L\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$$

Où :

a : La borne inférieure de la classe modale

Δ_1 : Différence entre l'effectif de la classe modale et la classe précédente.

Δ_2 : Différence entre l'effectif de la classe modale et la classe suivante.

L : L' amplitude de la classe modale *i.e.*, $L = b - a$.

Exemple 2.1.5 *Soit le tableau suivant*

Age	$[0, 4[$	$[4, 8[$	$[8, 12[$	$[12, 16[$	$[16, 20[$
effectif n_i	4	3	12	6	2

La classe modale est $[8, 12[$. *Donc* $Mo = 8 + 4\left(\frac{(12-3)}{(12-3)+(12-6)}\right) = 10.4$.

2.1.2 La médiane

La médiane est la valeur de la série (qui n'est pas toujours une valeur de la série) qui partage la série en deux groupes de même effectif (ou de même fréquence), elle est notée Me.

Cas d'une variable discrète

Soit la série ordonnée croissante de taille N

1) Si $N = 2p + 1$ ($i.e.$, N est impair), alors $Me = x_{p+1} = x_{\frac{N+1}{2}}$.

Exemple 2.1.6 $2, 3, 7, 10, 12, 12, 19$ *est une série statistique croissante,* $N = 7$ *est impair* $N = 2 \times 3 + 1$ *alors* $Me = x_{3+1} = x_4 = 10$.

2) Si $N = 2p$ ($i.e.$, N est pair), alors $Me = \frac{x_p + x_{p+1}}{2} = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$.

Exemple 2.1.7 $1, 3, 4, 6, 10, 10, 10, 12$ *est une série statistique croissante,* $N = 8$ *est pair* $N = 2 \times 4$ *alors* $Me = \frac{x_4 + x_5}{2} = \frac{6+10}{2} = 8$.

Cas d'une variable continue

Les données de la série statistique sont regroupées dans un tableau exprimé en classes. Si la classe médiane $[a, b[$. Le calcul de la médiane se fait en appliquant la formule suivante :

$$Me = a + L\left(\frac{\frac{N}{2} - ECC_{e-1}}{n_e}\right)$$

Où :

a : La borne inférieure de la classe médiane.

ECC_{e-1} : L'effectif cumulé croissant de la classe qui précède de la classe médiane

N : L'effectif total.

n_e : L'effectif de la classe médiane.

L : L' amplitude de la classe médiane. $i.e.$, $L = b - a$.

Exemple 2.1.8 *Soit le tableau suivant*

classes	$[5,10[$	$[10,15[$	$[15,20[$	$[20,25[$	$[25,30[$	*Totale*
effectif n_i	6	12	15	21	6	$N=60$
ECC	6	18	33	54	60	

On a $\frac{N}{2}=30$, *alors la classe médiane est* $[15,20[$, *donc* $Me=15+5\left(\frac{30-18}{15}\right)=19$.

2.1.3 La moyenne arithmétique

La moyenne arithmétique simple

La moyenne arithmétique d'une variable statistique X notée par $\overline{X}$ est définie par :
$\overline{X}=\frac{x_1+x_2+..+x_n}{n}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i$.

Exemple 2.1.9 *Si on a la série* $3,5,10,11,15$ *alors* $\overline{X}=\frac{3+5+10+12+15}{5}=\frac{45}{5}=9$.

La moyenne arithmétique pondérée

La moyenne arithmétique d'une série statistique (x_i,n_i) est la quantité, elle est notée $\overline{X}$

$\overline{X}=\frac{1}{N}\sum\limits_{i=1}^{n}n_ix_i$ ou $\overline{X}=\sum\limits_{i=1}^{n}f_ix_i$.

Exemple 2.1.10 *Soit le tableau suivant*

modalité x_i	1	2	3	4	*Total*
effectif n_i	4	6	2	8	20

La moyenne arithmétique $\overline{X}=\frac{1}{N}\sum\limits_{i=1}^{n}n_ix_i=\frac{1}{20}\left(4\left(1\right)+6\left(2\right)+2\left(3\right)+8\left(4\right)\right)=2.7$.

Remarque 2.1.11 *Dans le cas continu, on choisit la valeur* x_i *égale au centre de la classe correspondante* c_i *c'est-à-dire*

$\overline{X}=\frac{1}{N}\sum\limits_{i=1}^{n}n_ic_i$ ou $\overline{X}=\sum\limits_{i=1}^{n}f_ic_i$.

Exemple 2.1.12 *Soit le tableau suivant*

Age	$[3,4[$	$[4,5[$	$[5,6[$	$[6,7[$	*Total*
c_i	3.5	4.5	5.5	6.5	20
effectif n_i	19	7	1	3	30

$\overline{X} = \frac{1}{N}\sum\limits_{i=1}^{4} n_i c_i = \frac{1}{30}\left(4\left(3.5\right)+6\left(4.5\right)+2\left(5.5\right)+8\left(6.5\right)\right) = 3.46.$

Propriétés de la moyenne arithmétique

1) La somme des écarts à la moyenne est égale à 0
$\sum\limits_{i=1}^{n} n_i\left(x_i - \overline{X}\right) = 0.$
2) Si $x_i = \alpha y_i$, $\alpha \in \mathbb{R}$, alors $\overline{X} = \alpha\overline{Y}$

3) X variable statistique de modalités $x_1, x_2, .., x_n$. Si X' Variable statistique de modalités $x_1 - m, x_2 - m, , .., x_k - m$, Alors : $\overline{X'} = \overline{X} - m$. où $m > 0$.

Exemple 2.1.13 *calculer* $\overline{X}$ *de la série :* $106, 111, 112, 115$. *Si on prend* $m = 100$ *, on trouve*

$X' : 6, 11, 12, 15,$ *d'où* $\overline{X'} = \frac{6+11+12+15}{4} = 11$. *Alors :* $\overline{X} = \overline{X'} + 100 = 11 + 100 = 11.$

2.1.4 La moyenne quadratique

La moyenne quadratique est la racine carée de la moyenne arithmétique des carés des valeurs x_1, notée MQ est définie par :$MQ = \sqrt{\overline{X}^2}$, *i.e.*, $MQ = \sqrt{\frac{1}{N}\sum\limits_{i=1}^{n} n_i x_i^2}$ ou $MQ = \sqrt{\sum\limits_{i=1}^{n} f_i x_i^2}$.

Remarque 2.1.14 *Dans le cas continu, on choisit la valeur* x_i *égale au centre de la classe correspondante* c_i *c'est-à-dire*

$MQ = \sqrt{\frac{1}{N}\sum\limits_{i=1}^{n} n_i c_i^2}$ ou $MQ = \sqrt{\sum\limits_{i=1}^{n} f_i c_i^2}$.

2.1.5 La moyenne géométrique

La moyenne géométrique simple

La moyenne géométrique G est la racine nième du produit de n valeurs positives $x_1, x_2, .., x_n$ est

$G = \sqrt[n]{x_1 \times x_2 \times .. \times x_n}$

La moyenne géométrique pondérée

La moyenne géométrique d'une série statistique (x_i, n_i) est la quantité, elle est notée G

$G = \sqrt[N]{x_1^{n_1} \times x_2^{n_2} \times .. \times x_k^k}$

Dans le cas continu, on a

$G = \sqrt[N]{c_1^{n_1} \times c_2^{n_2} \times .. \times c_k^k}$

2.1.6 La moyenne harmonique

La moyenne harmonique simple

La moyenne harmonique notée H d'une série des valeurs non nulles $x_1, x_2, .., x_n$ est

$H = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n}\frac{1}{x_i}} = \frac{n}{\sum\limits_{i=1}^{n}\frac{1}{x_i}}$, par exemple pour $n = 2$, on écrit $H = \frac{2}{\frac{1}{x_1}+\frac{1}{x_2}} = \frac{2x_1x_2}{x_1+x_2}$.

La moyenne harmonique pondérée

La moyenne géométrique d'une série statistique (x_i, n_i) est la quantité, elle est notée H

$H = \frac{N}{\sum\limits_{i=1}^{n}\frac{n_i}{x_i}} = \frac{1}{\sum\limits_{i=1}^{n}\frac{f_i}{x_i}}$

Dans le cas continu, on a

$H = \frac{N}{\sum\limits_{i=1}^{n}\frac{n_i}{c_i}} = \frac{1}{\sum\limits_{i=1}^{n}\frac{f_i}{c_i}}$.

La fonction de répartition

Définition 2.1.15 *La fonction de répartition satisfait, pour $i \in \{1, .., n\}$, l'égalité, $F_x(x_i) = F_i$. L'expression,*

$$F_x(x_i) = \begin{cases} 0 & si \quad x < x_1 \\ F_1 & si \quad x_1 \leq x < x_2 \\ F_i & si \quad x_i \leq x < x_{i+1} \\ 1 & si \quad x \geq x_n \end{cases}$$

2.1.7 Les quartiles

Les quartiles généralisent la notion de la médiane, ils s'agissent des valeurs de la variable qui partagent la série statistique en quatre parties égales et sont définies par :

Le premier quartile Q_1 est la valeur au dessous de laquelle on trouve 25% de série statistique.

Le troisième quartile Q_3 est la valeur au dessus de laquelle on trouve 75% de série statistique.

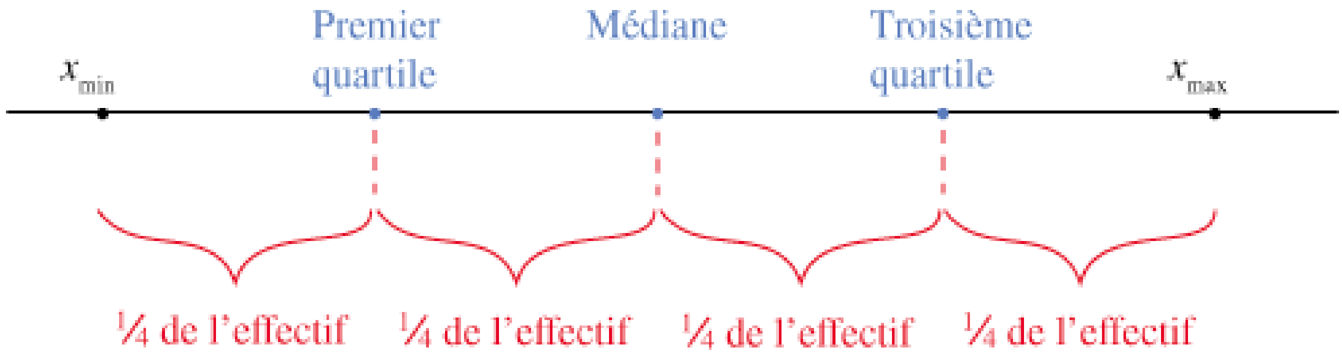
Remarque 2.1.16 *Q_1 et Q_3 sont deux valeurs de la série statistique, opposées à la médiane.*

Exemple 2.1.17 *Soit le tableau suivant*

modalité x_i	3	4	5	7	8	10	11	*total*
effectif n_i	5	7	3	8	8	6	3	40

1- Calculer ECC, les fréquences f_i et les fréquences cumulées croissantes f_iCC.

2 -Calculer la médiane Me et les quartiles Q_1, Q_3.



Solution 2.1.18

modalité x_i	3	4	5	7	8	10	11	*total*
effectif n_i	5	7	3	8	8	6	3	40
ECC	5	12	15	23	31	37	40	
f_i	0.125	0.175	0,075	0.2	0.2	0.15	0,075	1
f_iCC	0.125	0.3	0,375	0.575	0.775	0,925	1	

$N = 40 = 2 \times 20$ *est pair, alors* $Me = \frac{x_{20}+x_{21}}{2} = \frac{7+7}{2} = 7.$

pour Q_1*, on a* $\frac{N}{4} = 10$, Q_1 *est la valeur* $10^{ièm}$*, d'où* $Q_1 = x_{10} = 4$

pour $Q3$*, on a* $\frac{3N}{4} = 30$, Q_3 *est la valeur* $30^{ièm}$*, d'où* $Q_3 = x_{30} = 8$

2.2 Caractéristiques de dispersion (variabilité)

2.2.1 L'étendue

L'étendue est la différence entre la plus grande et la plus petite modalité du caractère, notée E et on écrit $E = x_{\max} - x_{\min}$.

2.2.2 La variance

La variance de la variable X notée $V(X)$ est définie par :

$V(x) = \frac{1}{N}\sum_{i=1}^{n} n_i \left(x_i - \overline{X}\right)^2$ ou $V(x) = \sum_{i=1}^{n} f_i \left(x_i - \overline{X}\right)^2$.

Théorème 2.2.1 (Théorème de Ko■nig-Huygens) *Soit (x_i, n_i) une série statistique de moyenne $\overline{X}$ et de variance $V(X)$. Alors $V(x) = \frac{1}{N}\sum_{i=1}^{n} n_i x_i^2 - \overline{X}^2$ ou $V(x) = \sum_{i=1}^{n} f_i x_i^2 - \overline{X}^2$.*

Démonstration. Exercice ■

Remarque 2.2.2 *Dans le cas continu, on choisit la valeur x_i égale au centre de la classe correspondante c_i c'est-à-dire*
$V(x) = \frac{1}{N}\sum_{i=1}^{n} n_i c_i^2 - \overline{X}^2$ *ou* $V(x) = \sum_{i=1}^{n} f_i c_i^2 - \overline{X}^2$.

2.2.3 L'écart-type

L'écart-type de X est donnée par $\sigma_x = \sqrt{V(x)}$.

2.2.4 L'écart absolu moyen

L'écart absolu moyen noté e_{abc} est défini par : $e_{abc} = \frac{1}{N}\sum_{i=1}^{n} n_i \left|x_i - \overline{X}\right|$.

Exercice 2.2.3 *Soit le tableau suivant*

modalité x_i	7	8	9	10	11	12	13	14	15	*total*
effectif n_i	1	1	2	4	5	2	3	1	1	20
ECC	5	12	15	23	31	37	40			
f_i	0.125	0.175	$0,075$	0.2	0.2	0.15	$0,075$	1		
f_iCC	0.125	0.3	$0,375$	0.575	0.775	$0,925$	1			

$1-$ *calculer l'étendue E.*

$2-$*calculer la moyenne arithmétique $\overline{X}$ et l'écart absolu moyen e_{abc}.*

$3-$*calculer La variance $V(X)$ et l'écart-type σ_x.*

Solution 2.2.4 *On a*

modalité x_i	7	8	9	10	11	12	13	14	15	*total*
effectif n_i	1	1	2	4	5	2	3	1	1	20
$n_i x_i$	7	8	18	40	55	24	39	14	15	220
$\left\|x_i - \overline{X}\right\|$	4	3	2	1	0	1	2	3	4	
$n_i \left\|x_i - \overline{X}\right\|$	4	3	4	4	0	2	6	3	4	30
$n_i \left(x_i - \overline{X}\right)^2$	16	9	8	4	0	2	12	9	16	76

1− *Etendue* $E = x_{\max} - x_{\min} = 15 - 7 = 8$

2−*calculer la moyenne arithmétique* $\overline{X} = \frac{1}{20}\sum_{i=1}^{9} n_i x_i = \frac{220}{20} = 11.$

l'écart absolu moyen $e_{abc} = \frac{1}{N}\sum_{i=1}^{n} n_i \left|x_i - \overline{X}\right|. = \frac{1}{20}\sum_{i=1}^{9} n_i \left|x_i - \overline{X}\right| = \frac{30}{20} = 1.5.$

3−*La variance* $V(X) = \frac{1}{N}\sum_{i=1}^{n} n_i \left(x_i - \overline{X}\right)^2 = \frac{1}{20}\sum_{i=1}^{9} n_i \left(x_i - \overline{X}\right)^2 = \frac{76}{20} = 3.8.$

l'écart-type $\sigma_x = \sqrt{V(X)} = \sqrt{3.8} = 1.95$

2.2.5 L'écart médian absolu

L'écart médian absolu notée e_{med} est défini par : $e_{med} = \frac{1}{N}\sum_{i=1}^{n} n_i \left|x_i - Me\right|.$

2.2.6 Coefficient de variation

Le coefficient de variation est une mesure relative de la dispersion des données autour de la moyenne. Le coefficient de variation est donnée par $CV = \frac{\sigma_x}{\overline{X}}$

2.2.7 Intervalle(distance) interquartile

La distance interquartile notée I_Q est la différence entre le troisième et le premier quartile $I_Q = Q_3 - Q_1$. L'intervalle interquartile est $[Q_1, Q_3]$.

Exercice 2.2.5 *Soit le tableau suivant*

modalité x_i	4	7	10	13	16	*total*
effectif n_i	6	5	7	4	1	23

1− *Calculer* ECC*, les fréquences* f_i*.*

2−*Calculer la médiane* Me *puis déduire l'écart médian absolu* e_{med}

3−*Calculer les quartiles* Q_1, Q_3 *puis déduire la distance interquartile* I_Q *et l'intervalle interquartile est* $[Q_1, Q_3]$.

Solution 2.2.6

modalité x_i	4	7	10	13	16	*total*
effectif n_i	6	5	7	4	1	23
ECC	6	11	18	22	23	
f_i	$\frac{6}{23}$	$\frac{7}{23}$	$\frac{10}{23}$	$\frac{4}{23}$	$\frac{1}{23}$	1

1- On a $N = 23 = 2 \times 11 + 1$ *est impair, alors* $Me = x_{12} = 10$. *On déduit* $e_{med} = \frac{1}{23} \sum_{i=1}^{5} n_i |x_i - 10|$

d'où

$$\begin{aligned} e_{med} &= \frac{1}{23} (6|4-10| + 5|7-10| + 7|10-10| + 4|13-10| + 1|16-10|) \\ &= \frac{1}{23} (6(6) + 5(3) + 7(0) + 4(3) + 1(6)) \\ &= \frac{1}{23} (6(6) + 5(3) + 7(0) + 4(3) + 1(6)) \\ &= \frac{69}{23} = 3 \end{aligned}$$

pour Q_1*, on a* $\frac{N}{4} = 5.75$, Q_1 *est la valeur* $6^{ièm}$*, d'où* $Q_1 = x_6 = 4$

pour $Q3$*, on a* $\frac{3N}{4} = 17.25$, Q_3 *est la valeur* $18^{ièm}$*, d'où* $Q_3 = x_{18} = 10$.

On déduit la distance interquartile $I_Q = Q_3 - Q_1 = 10 - 4 = 6$ *et l'intervalle interquartile est* $[4, 10]$.