Statistique Descriptive

$3^{ème}$ Chapitre: Les caractéristiques de tendance centrale ou de position.

Suite et Fin

La médiane

III. La médiane

Définition :

La **médiane**, notée **Me,** est la valeur de la variable statistique qui partage la population en deux effectifs égaux.

Remarque importante :

-La médiane ne s'applique que lorsque les observations sont **ordonnées** : ordre croissant ou décroissant.

-Le calcul de la médiane ne concerne que les variables mesurées sur une échelle ⇒ variables qualitatives.

III. La médiane

1.Calcul de médiane pour des données non réparties en classes :

Pour calculer la Médiane, on commence par ordonner les valeurs prises par la variable statistique X .

On note $X_{(1)}$ la première valeur,$X_{(2)}$ la deuxième valeur, ..., $X_{(n)}$ la plus grande valeur. On a donc :

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

On distingue deux cas :

$$Me = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ est impair,} \\ \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}\right)+1}}{2} & \text{si } n \text{ est pair.} \end{cases}$$

III. La médiane

Exemples de données non réparties en classes :

<u>Premier Cas "n" est impair :</u> la série observée : $\{3 \;\; 2 \;\; 1 \;\; 0 \;\; 0 \;\; 1 \;\; 2\}$

Le nombre d'observations c'est $n = 7$

Après rangement on obtient $\{0 \;\; 0 \;\; 1 \;\; \mathbf{1} \;\; 2 \;\; 2 \;\; 3\}$ Donc

$Me = 1$

<u>Deuxième Cas "n" est pair :</u> la série observée : $\{0 \;\; 1 \;\; 2 \;\; 1 \;\; 2 \;\; 4 \;\; 0 \;\; 3\}$ Le nombre d'observations c'est $n = 8$

Après rangement on obtient $\{0 \;\; 0 \;\; 1 \;\; \mathbf{1} \;\; \mathbf{2} \;\; 2 \;\; 3 \;\; 4\}$ La médiane de cette série se trouve entre 1 et 2.

On calcul la moyenne arithmétique de ces deux valeurs :

$$Me = \frac{1+2}{2} = 1.5$$

III. La médiane

Exemples d'une variable discrète :

Dans le cas d'une **variable discrète**, la détermination de la médiane se fait directement à l'aide des **effectifs cumulés croissants**.

Nombre d'enfants	effectif	Effectif cumulé croissant
0	20	20
1	16	36
2	10	46
3	5	51
4	0	**51**

La médiane est la modalité "1 enfant" qui correspond au foyer 16.

III. La médiane

2.Calcul de médiane pour des données réparties en classes :

Pour une variable continue, on détermine **la classe médiane** de la même façon que pour une variable discrète en **utilisant les effectifs cumulés**. Pour calculer **la valeur de la médiane**, on applique la relation

$$Me = x_i + \frac{N/2 - N_{\{i-1;cc\}}}{n_i}(x_{i+1} - x_i)$$

Avec

* $[x_i\,; x_{i+1}[$ est la classe médiane

* n_i est la fréquence de la classe médiane

* $N_{\{i-1;cc\}}$ est la fréquence cumulée croissante située avant la classe médiane.

* N est le nombre total d'observations

III. La médiane

<u>Exemple :</u> Soit une étude sur la note d'une population de 50 étudiants.

Notes	Effectifs	Effectifs cumulés
[0 ;5[	10	10
[5 ;8[	8	18
[8 ;12[	12	30
[12 ;15[	11	41
[15 ;20[	9	**50**
Total	**50**	

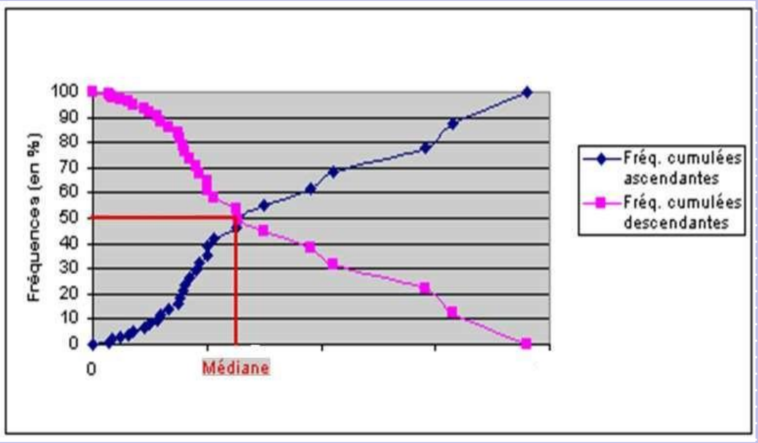
- $[x_i\,;\,x_{i+1}[$= [8; 12[est la classe médiane d'effectif n_i = 12.
-$N_{\{i-1;cc\}}$ = 18 est l' effectif cumulé croissant située avant la classe médiane.

Ainsi

$$\mathbf{Me} = \mathbf{8} + \frac{\mathbf{(50/2 - 18)}}{\mathbf{12}} \times \mathbf{4} = \mathbf{10.33}$$

III. La médiane

3. Détermination graphique de la médiane :



IV. Les quantiles

<u>Présentation :</u>

Les **quantiles** sont des mesures de position qui ne tentent pas nécessairement de déterminer le centre d'une distribution d'observations, mais de **décrire une position particulière**.

Cette notion est une extension du concept de la médiane qui divise une distribution d'observations en deux parties de même effectif.

IV. Les quantiles

Les quantiles les plus fréquemment utilisés sont :

1- Les quartiles qui divisent un ensemble d'observations en **quatre** parties égales ;

2- les déciles qui divisent un ensemble d'observations en **dix** parties égales ;

3- les centiles qui divisent un ensemble d'observations en **cent** parties égales.

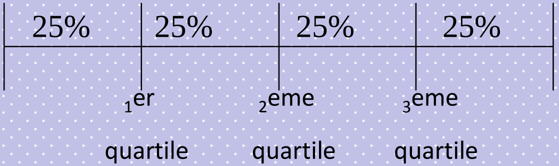
Remarque :

Le calcul des quantiles n'a de sens que pour une variable quantitative pouvant prendre des valeurs sur un intervalle déterminé.

IV. Les quantiles

1. Les quartiles :

Voici, schématiquement, une distribution partagée en quartiles. Entre chaque quartile se trouvent 25% des observations :



<u>Notations :</u> le $j^{\text{ème}}$ quartile se note **Q_j**

<u>Remarques :</u>

1- Le processus de calcul du quartile est similaire à celui de la médiane (c.-à-d. : ordonner les données).

2- les quartiles correspondent aux observations pour lesquelles la **fréquence cumulée croissante** dépasse respectivement 25 %, 50 % et 75%.

1. Les quartiles

Comment interpréter des quartiles ?

- **Le premier quartile** d'une série statistique est la plus petite valeur Q_1 telle qu'au moins 25% des valeurs sont inférieures ou égales à Q_1.
- **Le deuxième quartile** d'une série statistique c'est la **médiane**.
- **Le troisième quartile** d'une série statistique est la plus petite valeur Q_3 telle qu'au moins 75% des valeurs sont inférieures ou égales à Q_3.

1. Calcul de Q_j pour des données non réparties en classes :

On suppose que le nombre d'observations c'est n

On calcul d'abord la quantité $\frac{n}{4} = k$. Deux cas sont possibles :

1. Les quartiles

Premier Cas : *k* est un entier (la division tombe juste)

•On vérifie que les valeurs sont rangées par ordre croissant.

• Q_1 est la valeur de la modalité correspondante au k^{ieme} rang

•Q_3 est la valeur de la modalité correspondante au $3k^{ieme}$ rang.

Exemple 1 : Considérons la série suivante :

16 12 1 9 17 19 13 10 4 8 7 8 14 12 14 9

On commence d'abord par ranger les 16 termes de la série

1 4 7 8 8 9 9 10 12 12 13 14 14 16 17 19

Déterminer le premier quartile : on a **16/4** = 4 donc le premier quartile est la valeur de la modalité correspondant au $4^{ème}$ rang. Ainsi $Q_1 = 8$

Déterminer le troisième quartile : on a $3 \times 16/4 = 12$ donc le troisième quartile est la valeur de la modalité correspondant au $12^{ème}$ rang. Ainsi $Q_3 = 14$.

1. Les quartiles

Exemple 2 :On donne la répartition des notes à un contrôle dans une classe de 28 élèves.

Notes	Effectif	**Effectif cumulé**
7	2	2
8	3	5
9	5	10
10	2	12
11	1	13
12	6	19
13	3	22
14	3	25
15	3	28

on a $k = \frac{28}{4} = 7 \in \mathbb{N}$,

Cle premier quartile est au 7^e rang, soit $Q_1 = 9$

Cle troisième quartile est au $3 \times 7 = 21^e$ rang, soit $Q_3 = 13$.

1. Les quartiles

Deuxième Cas : $\frac{n}{4} = k$ n'est pas un entier (nombre décimal)

1- On vérifie que les valeurs sont rangées par ordre croissant.

2- On arrondit le décimal k à l'entier supérieur et Q_1 sera la valeur de la modalité correspondante au rang de cet entier.

3- On arrondit le décimal $3 \times k$ à l'entier supérieur et Q_3 sera la valeur de la modalité correspondante au rang de cet entier.

Exemple : Prenons **les valeurs rangées dans l'ordre croissant** :

3 5 5 6 7 8 8 9 9 10 10 10 10 11 11 12 13 13 13 14 15 16 19

On a n = 23 ; $k = \frac{n}{4} = 5,75$ et $3 \times \frac{n}{4} = 17,25$ donc

Q_1 est la 6ème valeur de la série rangée dans l'ordre croissant : $Q_1 = 8$;

Q_3 est la 18ème valeur de la série rangée dans l'ordre croissant : $Q_3 = 13$.

1. Les quartiles

2. Calcul de Q_j pour des données réparties en classes :

Selon la même analyse pour le calcul du mode ou la médiane pour des variables continues, on détermine d'abord **la classe du quartile**.
On applique la formule suivante

$$Q_j = x_i + \frac{(N \times j) - N_{i-1}^{cc}}{n_i}(x_{i+1} - x_i)$$

Avec

1- $[x_i ; x_{i+1}[$ est la classe du quartile.

2- j= 1/4 ou 1/2 ou 3/4 correspondant à Q_1, Q_2 et Q_3.

3- n_i c'est l'effectif (ou la fréquence) de la classe du quartile.

4- N_{i-1}^{cc} c'est l'effectif cumulée croissant de la classe précédente.

5- $x_{i+1} - x_i$ c'est l'amplitude de la classe du quartile.

1. Les quartiles

Exemple : 75 ouvriers que l'on classe en fonction du nombre d'heures travaillées pendant un mois.

On a $\frac{75}{4} = 18.75 = k$

La classe du premier quartile Q_1 c'est

[100 ; 130[

$$Q_1 = 100 + \frac{75 \times 0.25 - 15}{15}(130 - 100)$$

$$Q_1 = 107.5$$

- On a $3 \times k = 3 \times 18.75 =$ 56.25

La classe du troisième quartile Q_3 c'est

[150 ; 180[

$$Q_3 = 150 + \frac{75 \times 0.75 - 53}{17}(180 - 150)$$

$$Q_3 = 155.735$$

Classe de valeurs	Effectif	Effectifs cumulés Croissant
[50,70[	6	6
[70,100[	9	15
[100,130[	15	30
[130,150[	23	53
[150,180[	17	70
[180,200[	5	75
Total	*75*	

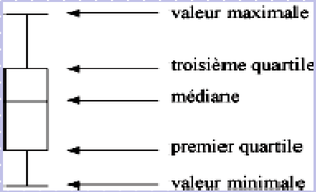
1. Les quartiles

Vocabulaires :

- $[Q_1; Q_3]$ est l'intervalle interquartile, il contient à peu prés 50% des valeurs de la série.

- $IQ = Q_3 - Q_1$ c'est l'écart interquartile.

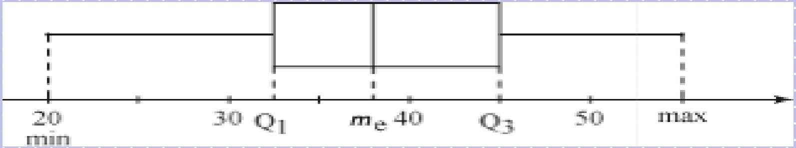
Définition : Un résumé graphique des données d'une enquête statistique, moyennant les paramètres de position, peut être fourni sous la forme d'un diagramme " **en boîte à moustaches** ". Les " moustaches " correspondent aux valeurs extrêmes de la série.



Boîte à moustaches-diagramme de Tuckey

Exemple :

Soit une variable statistique X dont le **maximum** est 55, le **minimum** 20, la **médiane** 38, le **1er quartile** 32,5 et le **3e quartile** 45. On construit alors la Boîte à moustaches comme suit :



2. Les déciles

Les mêmes procedures appliquées précédemment aux quartiles ($k = \frac{n}{4}$) resteront valables pour la suite à l'exception que $k = \frac{n}{10}$.

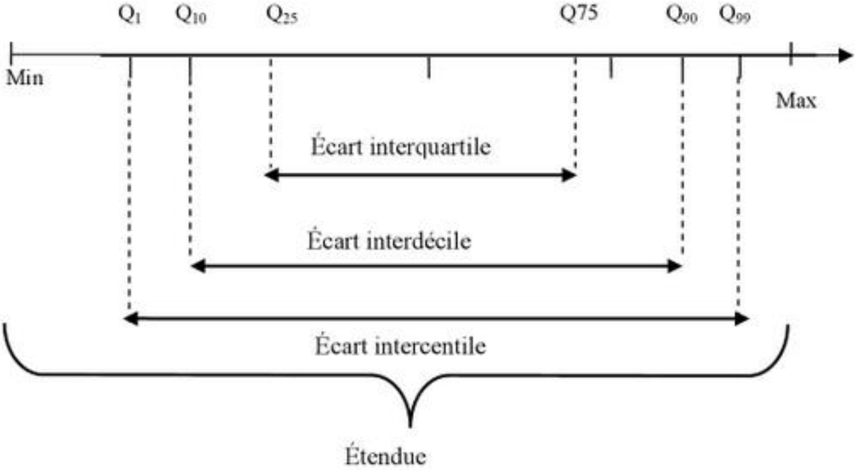
Lorsque k=10, les quantiles sont appelés déciles.

-Le premier décile d'une série la plus petite valeur D_1 des termes de la série pour laquelle au moins un dixième (10%) des données sont inférieures ou égales à D_1.

-le neuvième décile (D_9) d'une série est la plus petite valeur des termes de la série pour laquelle au moins neuf dixièmes (90%) des données sont inférieures ou égales à D_9

-l'intervalle interdécile est $[D_1; D_9]$

-l'écart interdécile est le nombre $D_9 - D_1$.



Tous ces écarts permettent de mesurer la dispersion de la série autour de la médiane.

Exercice d'application

Dans une entreprise, la répartition des individus par âge et sexe est consignée dans le tableau suivant:

Age(ans)	H	F
[20;25[	29	38
[25;30[	48	57
[30;35[	36	42
[35;40[	45	39
[40;45[	49	41
[45;50[	32	30
[50;55[	37	18
[55;60[	28	20

1- Calculer la moyenne d'âge par sexe.

2- Calculer la moyenne d'âge de la population totale.

Age(ans)	H	F	Ci	nihci	nifci	
[20;25[	29	38	22,5	652,5	855	
[25;30[	48	57	27,5	1320	1567,5	
[30;35[	36	42	32,5	1170	1365	
[35;40[	45	39	37,5	1687,5	1462,5	
[40;45[	49	41	42,5	2082,5	1742,5	
[45;50[	32	30	47,5	1520	1425	
[50;55[	37	18	52,5	1942,5	945	
[55;60[	28	20	57,5	1610	1150	
	304	285		11985	10512,5	
				mh	39,4243421	
				mf	36,8859649	
					22497,5	589
				mt	38,1960951	

Merci