

## Séquence 3 : STATISTIQUE DESCRIPTIVE. ANALYSE DE DONNEES

### I/ Représentation d'une série statistique :

#### a) Vocabulaire des statistiques :

##### Définition 1 :

L'ensemble sur lequel porte l'étude statistique est appelé la population.

Si cet ensemble est trop vaste, on en restreint l'étude à une partie appelée échantillon.

Un élément de cet ensemble est appelé individu.

##### Définition 2 :

L'étude statistique rend compte d'un aspect des individus de la population appelé caractère.

Si l'on peut mesurer cet aspect, le caractère est dit quantitatif. Il prend alors différentes valeurs.

Si les valeurs sont isolées, on dit que le caractère est discrêt.

Si les valeurs appartiennent à un intervalle de IR, on dit que le caractère est continu.

Si l'aspect ne se traduit pas par des nombres, le caractère est dit qualitatif. Il prend alors différentes modalités.

##### Exemples :

1) Le nombre de frères et sœurs d'un élève de seconde 7 est un caractère .....

Il peut prendre les valeurs .....

2) La taille des élèves de seconde 7 est un caractère ..... Il peut prendre toutes les valeurs entre .....

3) Dans une population, être marié est un caractère ..... à deux modalités : ..... et .....

##### Définition 3 :

On appelle effectif le nombre d'individus pour lesquels le caractère prend une valeur donnée.

La somme des effectifs de toutes les valeurs est l'effectif total. C'est le nombre d'individus de la population.

On appelle fréquence d'une valeur, le quotient  $\frac{\text{effectif de la valeur}}{\text{effectif total}}$ .

##### Remarque :

- On résume en général une étude statistique dans un tableau.

Valeur	x <sub>1</sub>	x <sub>2</sub>	.....	x <sub>p</sub>	Total
Effectif	n <sub>1</sub>	n <sub>2</sub>	.....	n <sub>p</sub>	N
Fréquence	f <sub>1</sub>	f <sub>2</sub>	.....	f <sub>p</sub>	1

- La fréquence est un nombre toujours compris entre 0 et 1.

- La somme des fréquences f<sub>i</sub> est toujours égale à 1.

**b) Effectifs cumulés, fréquences cumulées :**

**Définition 3 :**

On note  $x_i$  une valeur prise par un caractère quantitatif.

L'**effectif cumulé croissant** (resp. **décroissant**) de  $x_i$  est la somme des effectifs des valeurs inférieures (resp. supérieures) ou égales à  $x_i$ .

La **fréquence cumulée croissante** (resp. **décroissante**) de  $x_i$  est la somme des fréquences des valeurs inférieures (resp. supérieures) ou égales à  $x_i$ .

**Exemple :**

1) Ce tableau donne le nombre d'enfants âgés de 0 à 16 ans dans un échantillon de 100 familles.

On a fabriqué la série des **effectifs cumulés croissants**.

On peut lire par exemple que ..... familles ont **au plus** 1 enfant (soit 0, soit 1 enfant).

Nombre d'enfants	0	1	2	3	4	5
Effectif	50	23	10	14	2	1
Effectifs cumulés croissants	50	73				

2) Ce tableau donne la répartition des masses des nouveaux-nés dans un hôpital.

On a fabriqué la série des **fréquences cumulées décroissantes**.

On peut lire par exemple que ..... de ces nouveaux-nés pèsent **au moins** 3 kg.

Masse (en kg)	Fréquence	Fréquences cumulées décroissantes
[2,5 ; 3[	0,15	
[3 ; 3,5[	0,32	
[3,5 ; 4[	0,40	0,53
[4 ; 4,5[	0,13	0,13

**II/ Paramètres de position :**

On considère la liste des N données suivantes :  $x_1, x_2, \dots, x_N$ .

**a) La moyenne :**

La **moyenne** notée  $\bar{x}$  est égale à  $\bar{x} = \frac{\text{somme des valeurs}}{\text{l'effectif total}}$ .

**Exemple :**

Anaïs a obtenu 3 notes en mathématiques au 1<sup>er</sup> trimestre :

15,50 de coefficient 1 ; 12 de coefficient 3 et 16 de coefficient 2.

Quelle est sa moyenne ? .....

.....

Remarque : On peut aussi calculer la moyenne à partir de la distribution des fréquences.

Exemple :

Au dernier contrôle, 12 élèves ont obtenu 14,50 ; 7 ont obtenu 12 et 5 ont obtenu 10,25.

Quelle est la moyenne de la classe ? .....

**b) La médiane :** La liste des N données est rangée par ordre croissant

- Si la série est de taille impaire ( $N = 2n+1$ ), la médiane est la donnée de rang  $n+1$ ;

- Si la série est de taille paire ( $N = 2n$ ), la médiane est la demi-somme des données de rang  $n$  et de rang  $n+1$ .

Remarque-Méthode :

Pour trouver la médiane, on range les valeurs de la variable dans l'ordre croissant et on choisit la valeur « centrale » si l'effectif est impair, ou la moyenne des valeurs « centrales » si l'effectif total est pair.

Exemples :

1) Considérons la série suivante : 5 ; 9 ; 7 ; 4 ; 15

On commence par ranger les valeurs de la série dans l'ordre croissant : .....

La médiane de la série est alors ..... car la valeur « centrale » (la 3<sup>ème</sup>) est .....

2) Considérons la série suivante : 8 ; 9 ; 5 ; 14

On commence par ranger les valeurs de la série dans l'ordre croissant : .....

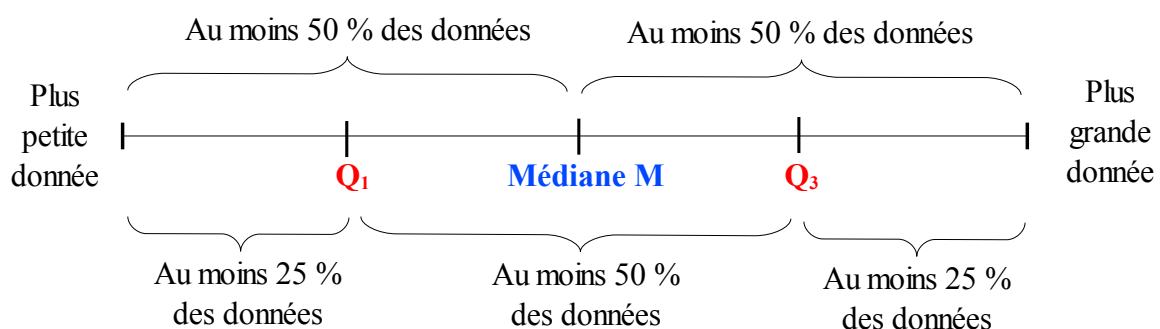
La médiane de la série est alors ..... car les valeurs « centrales » (les 2<sup>ème</sup> et 3<sup>ème</sup>) sont ..... et ..... et leur moyenne est .....

**c) Les quartiles :** La liste des N données est rangée par ordre croissant

Le premier quartile est la plus petite donnée  $Q_1$  de la liste telle qu'au moins un quart des données de la liste sont inférieures ou égales à  $Q_1$ .

Le troisième quartile est la plus petite donnée  $Q_3$  de la liste telle qu'au moins les trois quart des données de la liste sont inférieures ou égales à  $Q_3$ .

Représentation :



Application :

Dans une maternité, on a référencé les périmètres crâniens à la naissance de 290 nouveaux-nés.

Périmètre (en cm)	32	32,5	33	33,5	34	34,5	35	35,5	36	36,5	37	37,5
Effectif	4	19	17	20	59	62	43	20	18	18	4	6

Déterminer la médiane M et les quartiles Q<sub>1</sub> et Q<sub>3</sub> de cette série :

L'effectif total est 290. Or  $290 = 2 \times 145$ , donc la médiane est la demi-somme des 145<sup>e</sup> et 146<sup>e</sup> périmètres crâniens.

On cumule les effectifs :  $4 + 19 + 17 + 20 + 59 = 119$  et  $119 + 62 = 181$ .

Donc les 145<sup>e</sup> et 146<sup>e</sup> périmètres crâniens sont égaux à 34,5 donc  $M = 34,5$  cm.

$\frac{290}{4} = 72,5$  donc Q<sub>1</sub> est le 73<sup>e</sup> périmètre crânien.

On cumule les effectifs :  $4 + 19 + 17 + 20 = 60$  et  $60 + 59 = 119$ . Donc Q<sub>1</sub> = 34 cm.

$\frac{3 \times 290}{4} = 217,5$  donc Q<sub>3</sub> est le 218<sup>e</sup> périmètre crânien.

On cumule les effectifs :  $181 + 43 = 224$ . Donc Q<sub>3</sub> = 35 cm.

III/ Paramètres de dispersion :

Définition 4 :

On appelle **étendue** d'une série statistique, la différence entre la plus grande valeur du caractère et la plus petite valeur.

Exemple :

Au cours des 20 premiers jours du mois d'août 2009, Météo France a relevé les températures suivantes à Moulins :

25°, 29°, 21°, 31°, 32°, 27°, 28°, 27°, 29°, 31°, 38°, 25°, 26°, 25°, 34°, 35°, 35°, 29°, 28°, 31°.

Quelle est l'étendue de cette série de températures ? .....

Définition 5 :

On appelle **écart interquartile** d'une série statistique, la différence entre le 3<sup>ème</sup> quartile et le 1<sup>er</sup> quartile c'est-à-dire Q<sub>3</sub> - Q<sub>1</sub>.

Exemple :

En utilisant la série obtenue dans l'application sur les périmètres crâniens, déterminer l'écart interquartile : .....