Statistiques

I. Rappels :

1. Définitions :

Lorsque l'on réalise une enquête statistique, on étudie des **caractères** propres à chaque **individu**. L'ensemble des individus est appelé la **population**.
Le caractère peut être **qualitatif** (la couleur des yeux, les sports pratiqués ou le type de film préféré) ou **quantitatif** (la taille, l'âge, le temps passé devant la télévision, ...). Les données sont ensuite regroupées et présentées dans un **tableau de données**. Le nombre total d'individus de la population est l'**effectif total.** Le nombre d'individus qui possèdent un même caractère est appelé **effectif du caractère** (c'est à dire le nombre de fois où cette valeur apparaît dans la série).

2. Propriétés :

La fréquence d'un caractère est le quotient : $\dfrac{\textit{effectif du caractère}}{\textit{effectif total}}$ **.**

La **fréquence en pourcentage** d'un caractère est : $\dfrac{\textit{effectif du caractère} \times 100}{\textit{effectif total}}$

3. Exemples :

On demande à des élèves leur taille, et on regroupe les résultats dans un tableau.

Taille	1,20 ≤ T < 1,30	1,30 ≤ T < 1,40	1,40 ≤ T < 1,50	**1,50 ≤ T < 1,60**	1,60 ≤ T < 1,70
Effectif	5	7	13	**9**	6
Fréquence	0,125	0,175	0,325	**0,225**	0,15
Fréquence en %	12,5	17,5	32,5	**22,5**	1

La colonne grise signifie qu'il y a **9** élèves de taille comprise entre **1,50** m et **1,60** m. Ces 9 élèves représentent 22,5 % des élèves interrogés.
On regroupe ces résultats par **effectifs cumulés croissants** et **fréquence cumulées croissantes** :

Taille < ...	1,30	**1,40**	1,50	1,60	1,70
Ecc	5	**12**	25	34	40
Fcc	0,125	**0,3**	0,625	0,85	1

La colonne grise signifie qu'il y a **12** élèves de taille inférieure à **1,40** m.
On regroupe ces résultats par **effectifs cumulés décroissants** :

Taille >...	1,20	1,30	**1,40**	1,50	1,60
Ecd	40	35	**28**	15	6

La colonne grise signifie qu'il y a **28** élèves de taille supérieure à **1,40** m.

II. Caractéristiques de position d'une série statistique :

1. Moyenne d'une série statistique :

Pour calculer la **moyenne pondérée** M d'une série statistique :
- on additionne les produits des valeurs par leurs effectifs.
- on divise le résultat par l'effectif total.

$$M = \frac{\textit{effectif}_1 \times \textit{valeur}_1 + \textit{effectif}_2 \times \textit{valeur}_2 + \ldots + \textit{effectif}_p \times \textit{valeur}_p}{\textit{effectif total}}$$

Exemple 1 : Un élève a obtenu les notes suivantes au bac :

Matière	Français	Mathématiques	Histoire	Anglais	Espagnol
Note	12	10	11	8	5
Coefficient (Effectif)	4	4	2	2	1

Si on calcul la moyenne sans se soucier des coefficients (ce serait faux mais essayons...)

$M = \dfrac{12+10+11+8+5}{5} = 9{,}2$ Il n'aurait alors pas son bac...

Si on calcule la **moyenne pondérée**, on trouve : $M = \dfrac{12\times4+10\times4+11\times2+8\times2+5\times1}{13} = \dfrac{131}{13} = 10{,}1$

Ouf, il a son bac !

Remarque : Avoir une note coefficient 2 c'est « comme si » on avait 2 fois cette note.
Dans notre exemple le Français compte donc dans la moyenne 2 fois plus que l'Histoire.

Exemple 2 : On souhaite calculer la taille moyenne que l'on a regroupé dans des classes :

Taille	**1,20 ≤ T < 1,30**	1,30 ≤ T < 1,40	1,40 ≤ T < 1,50	1,50 ≤ T < 1,60	1,60 ≤ T < 1,70
Effectif	5	7	13	9	6

Pour cela il faut remplacer chaque classe par son centre : (par exemple 1,25 est le centre de 1,20 ≤ T < 1,30)

Taille	**1,25**	1,35	1,45	1,55	1,65
Effectif	5	7	13	9	6

Puis on peut calculer la moyenne pondérée : $M=\frac{1{,}25\times5+1{,}35\times7+1{,}45\times13+1{,}55\times9+1{,}65\times6}{40}=\frac{58{,}4}{40}=1{,}46\ m$

2. Médiane d'une série statistique :

Définition : La **médiane** *m* d'une série statistique est le nombre que partage le groupe de valeurs en 2 sous groupes de même effectif de sorte que:
- Toutes les valeurs du premier groupe soient inférieures à *m*.
- Toutes les valeurs du deuxième groupe soient supérieures à *m*.

Pour trouver la médiane, il faut donc en premier, ranger les valeurs par ordre croissant.
Puis prendre la valeur « centrale ».

Exemple 1 : Trouver la médiane de la série suivante : 25 ; 54 ; 32 ; 63 ; 21 ; 12, 34.
On ordonne : 12 ; 21 ; 25 ; **32** ; 34 ; 54 ; 63. Il y a en tout 7 valeurs donc la médiane est la quatrième soit **32**.

Exemple 2 : Trouver la médiane de la série suivante : 45 ; 3 ; 40 ; 69 ; 90 ; 21 ; 9 ; 71.
On ordonne les valeurs : 3 ; 9 ; 21 ; **40 ; 45** ; 69 ; 71 ; 90. Il y a en tout 8 valeurs donc la médiane est entre la quatrième (40) et la cinquième (45) valeurs on fait donc la moyenne de ces deux valeurs soit **42,5**.

On peut également se servir des effectifs cumulés pour déterminer la médiane. La médiane est la valeur à partir de laquelle l'effectif cumulé devient supérieur ou égal à la moitié de l'effectif total. (ou avec les fréquences cumulées qui égale ou dépassent les 50%).

Exemple 3 : Calculer la médiane de cette série grâce aux effectifs cumulés :32, 6, 18, 29, 6, 48, 50, 12, 32, 4, 50, 10, 18, 67, 32, 16, 16, 6, 50, 50, 4, 18, 6, 10, 29, 12, 48, 6, 32, 50.

Valeurs	4	6	10	12	16	18	29	32	48	50	67
Effectifs	2	5	2	2	2	3	2	4	2	5	1
Effectifs cumulés	2	7	9	11	13	16	18	22	24	29	30

La moitié de l'effectif total (30) est 15 donc on regarde la valeur pour laquelle l'effectif cumulé est égal à 15 ou devient supérieur. C'est le cas pour la valeur 18 : l'effectif cumulé de 18 vaut 16 et celui de la valeur 16 vaut 13 ce qui veut dire que la 14ème et la 15ème valeur de la série (la médiane) est la valeur 18.

III. Caractéristiques de dispersion d'une série statistique :

1. L'étendue :

Définition : L'**étendue** d'une série statistique est la différence entre la valeur la plus grande et la valeur la plus petite de la série.

Exemple : Dans l'exemple ci-dessus, l'étendue de la série est 67 – 4 = 63.

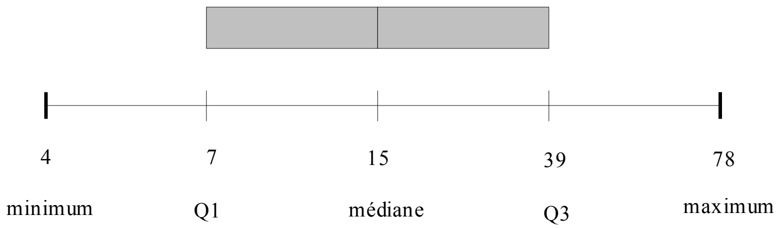
2. Les quartiles :

Définition : Les **quartiles** d'une série statistique sont les 3 valeurs qui partagent la série en 4 parties de même effectif. Chaque partie représente donc ¼ de la série. Il faut que les valeurs soient rangées dans l'ordre croissant avant le partage. On note Q1, Q2 et Q3 le premier, deuxième et troisième quartile.

Remarque : Q2 est en fait la médiane de la série.

Exemple : Trouver les quartiles de la série suivante : 20 ; 52 ; 31 ; 4 ; 78 ; 5 ; 34 ; 4 ; 9 ; 10 ; 44 ; 9
On ordonne les valeurs : 4 - 4 - 5 - 9 - 9 - 10 - 20 - 31 - 34 - 44 - 52 - 78
La série contient 12 valeurs. On divise par 4:   12 : 4 = 3
donc Q1 est "entre" la 3ème et la 4ème valeur : entre 5 et 9 donc Q1 = 7,
Q2 est "entre" la 6ème et la 7ème valeur : entre 10 et 20 donc Q2 = 15 (c'est la médiane en fait)
et Q3 est "entre" la 9ème et la 10ème valeur : entre 34 et 44 donc Q3 = 39

Toutes ces données peuvent être représentées par un diagramme en boite:



Cela permet de visualiser les valeurs de la série sachant que :

- ¼ des valeurs est au dessous de Q1
- ¼ des valeurs est au dessus de Q3
- ½ des valeurs sont entre Q1 et Q3