Statistiques

Lorsque l'on mène une enquête, on s'intéresse à une **population d'individus** *(ex : élèves d'une classe)* et on étudie une propriété commune à ces individus appelée un **caractère** *(ex : leur taille)*.
Un caractère peut prendre plusieurs **valeurs** ou **modalités**. Il peut être :

- **Quantitatif**, lorsque les valeurs possibles du caractère sont des valeurs numériques *(ex : taille, poids, nombre de frères et sœurs,…)*
- **Qualitatif**, dans le cas où les valeurs ne sont pas numériques *(ex : langues parlées, marques de voitures, …)*

I. Effectifs et Fréquences

Définition1 : Soit S une série statistique à une variable et a une modalité de S.
L'**effectif** associé à a est le nombre d'apparitions de cette modalité dans la série S, c'est-à-dire le nombre d'individus de la population pour lesquels le caractère étudié a cette modalité.
L'**effectif total** de la série est la somme des effectifs de toutes les modalités ou de toutes les classes.

Remarque : L'effectif total de la série est le nombre total d'individus de la population étudiée.

Définition2 : Soit S une série statistique à une variable et a une modalité de S.
La **fréquence** associée à a est le quotient de l'effectif de cette modalité par l'effectif total de la série.

Remarques :

- La fréquence d'une modalité ou d'une classe est un nombre compris entre 0 et 1 ; elle peut aussi être exprimée en pourcentage de l'effectif total.
- La somme des fréquences de toutes les modalités ou de toutes les classes est égale à 1 *(ou à 100% pour les fréquences exprimées en pourcentages)*.

II. Effectifs cumulés et fréquences cumulées

Soit S une série statistique à une variable de type quantitatif et a une modalité de S.

Définition3 : l'**effectif cumulé croissant** ***(décroissant)*** associé à a est la somme des effectifs de toutes les modalités inférieures ou égales *(supérieures ou égales)* à a dans la série S.

Définition4 : la **fréquence cumulée croissante** ***(décroissante)*** associée à a est la somme des fréquences de toutes les modalités inférieures ou égales *(supérieures ou égales)* à a dans la série S.

*Exemple1 : (exemple d'étude statistique à **caractère quantitatif**)*

Une étude statistique a été obtenue auprès de 25 élèves de troisième. On a relevé les notes de 25 élèves de cette classe à une interrogation écrite (sur 10).
Voici un exemple de tableau récapitulatif illustrant la méthode de détermination des effectifs et fréquences cumulés croissantes (ou décroissantes). Complétez ces deux tableaux

Notes	1	2	3	4	5	6	7	8	9	10
Effectifs	2	0	1	3	8	5	2	0	3	1
Effectifs cumulés croissants	……	……	……	6	……	……	……	……	……	……
Effectifs cumulés décroissants	……	……	……	……	……	11	……	……	……	……

Notes	1	2	3	4	5	6	7	8	9	10
Effectifs	2	0	1	3	8	5	2	0	3	1
Fréquence	0,08	0	0,04	……	0,32	0,2	0,08	0	0,12	0,04
Fréquences cumulées croissantes	……	……	0,12	……	……	……	……	……	……	……
Fréquences cumulées décroissantes	……	……	……	……	0,76	……	……	……	……	……
Fréquence (en %)	……	……	……	……	……	……	……	……	……	……

Effectif total : …… *Fréquence totale : ……* *Fréquence totale (en %) : ……*

*Exemple2 : (exemple d'étude statistique à **caractère qualitatif**)*

Dans un collège, il y a 165 élèves en 6ème, 150 en 5ème, 120 en 4ème et 105 en 3ème.
Complétez le tableau suivant :

Classes	*6ème*	*5ème*	*4ème*	*3ème*	*Total*
Effectif	……	……	……	……	……
Fréquence	……	……	……	……	……
Fréquence (en %)	……	……	……	……	……

III. Regroupement de données en classes de données

Dans le cas de nombreuses données numériques, on peut les regrouper **en classes** pour faciliter leur lecture.
Dans le cas d'une série *S* de type quantitatif dont les modalités sont regroupées en classes, l'effectif d'une classe est le nombre d'individus de la population pour lesquels le caractère étudié a une modalité appartenant à cette classe.
L'**effectif d'une classe** est le nombre de données de cette classe.
La **fréquence d'une classe** est le quotient de l'effectif de la classe par l'effectif total de la série.

Remarque : Le regroupement en classe permet donc une vision plus globale des résultats mais elles font perdre un certain nombre d'informations.

*Exemple3 : (exemple d'étude statistique à **caractère quantitatif**)*

On a relevé les distances des sauts de sportif pratiquant le saut en longueur, voici les distances relevées :

4,45 m	4,57 m	6,2 m	5,72 m	6,57 m	6,89 m	6,47 m	6,43 m	5,24 m	5,84 m
4,99 m	6,52 m	5,56 m	5,12 m	4,3 m	4,78 m	6,9 m	6,13 m	5,78 m	4,89 m
5,66 m	4,62 m	6,66 m	6,28 m	4,02 m	6,12 m	5,14 m	5,25 m	5,48 m	4,91 m

On décide de regrouper ces résultats en classes, Complétez le tableau suivant :

Longueur	$4 \leq n < 4{,}5$	$4{,}5 \leq n < 5$	$5 \leq n < 5{,}5$	$5{,}5 \leq n < 6$	$6 \leq n < 6{,}5$	$6{,}5 \leq n < 7$
effectif	……	……	……	……	……	……

IV. Diagrammes statistiques

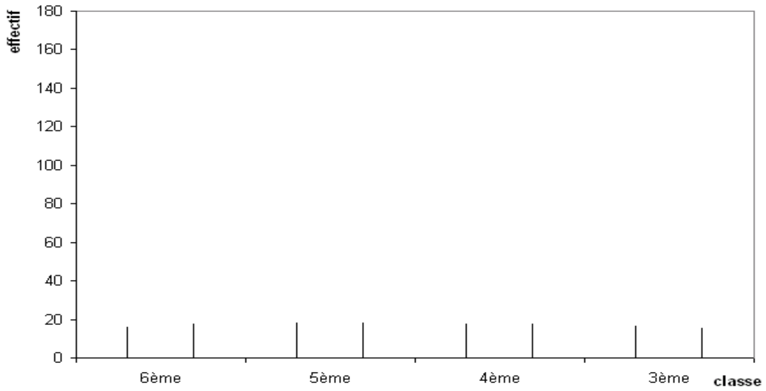
1) Lorsque le **caractère est qualitatif :**

a. Diagrammes en barres (horizontales ou verticales)

Dans un diagramme en barres, les valeurs sont représentées par des rectangles :
- ces rectangles ont la même largeur
- **la hauteur de chaque rectangle est proportionnelle** à l'effectif de la valeur qu'il représente.

Retour exemple2 :

Classes	*6ème*	*5ème*	*4ème*	*3ème*	*choix d'unité*
Effectif	165	150	120	105	40
Hauteur associée (cm)	……	……	……	……	1



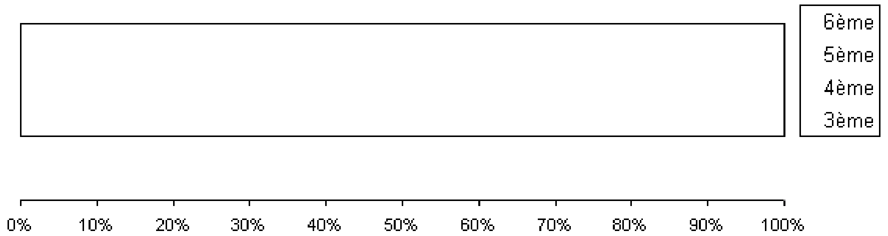
b. Diagrammes en bandes

Dans un diagramme en bandes, on représente la totalité des effectifs par une bande rectangulaire. Cette bande est « découpée » proportionnellement à l'effectif de chaque valeur représentée.

Retour exemple2 : On choisit une longueur pour représenter l'effectif total, complétez alors le tableau suivant

	$6^{ème}$	$5^{ème}$	$4^{ème}$	$3^{ème}$	*Choix d'unité*
Effectif	165	150	120	105	540
Longueur (cm)	……	……	……	……	10

A l'aide du tableau précédent complétez le diagramme en bandes suivant :



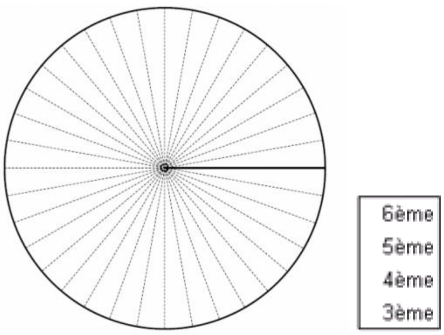
c. Diagrammes circulaires

Un diagramme circulaire résulte du découpage d'un disque en secteurs circulaires dont les mesures d'angle sont proportionnelles aux effectifs des valeurs qu'ils représentent.
L'effectif total est représenté par un angle de mesure égale à 360°

Retour exemple2 :

	$6^{ème}$	$5^{ème}$	$4^{ème}$	$3^{ème}$	*Total*
Effectif	165	150	120	105	540
Angles associés (en degrés)	……	……	……	……	360

A l'aide de ce tableau, représenter ces données sur le diagramme circulaire suivant :



Remarque : Pour un diagramme semi-circulaire, l'effectif total est représenté par un angle de mesure égale à 180°.

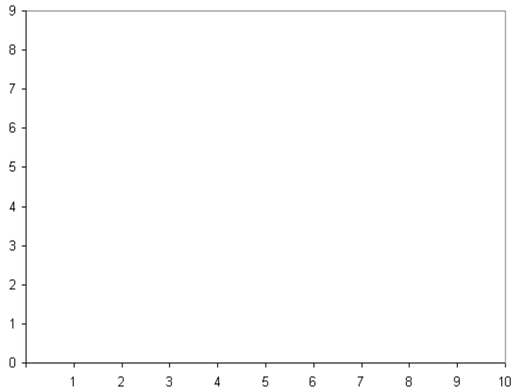
2) Lorsque le **caractère est quantitatif :**

a. Diagrammes en bâtons : (variable discrète)

Un diagramme en bâtons est constitué de segments de droite verticaux dont les hauteurs sont égales aux effectifs ou aux fréquences de chaque modalité.
Sur l'axe des abscisses sont reportées les modalités de la série.

Retour exemple1 :

Notes	1	2	3	4	5	6	7	8	9	10	*Choix d'unité*
Effectifs	2	0	1	3	8	5	2	0	3	1	8
Hauteur associée (en cm)	……	……	……	……	……	……	……	……	……	……	4



b. Histogrammes : (variable continue)

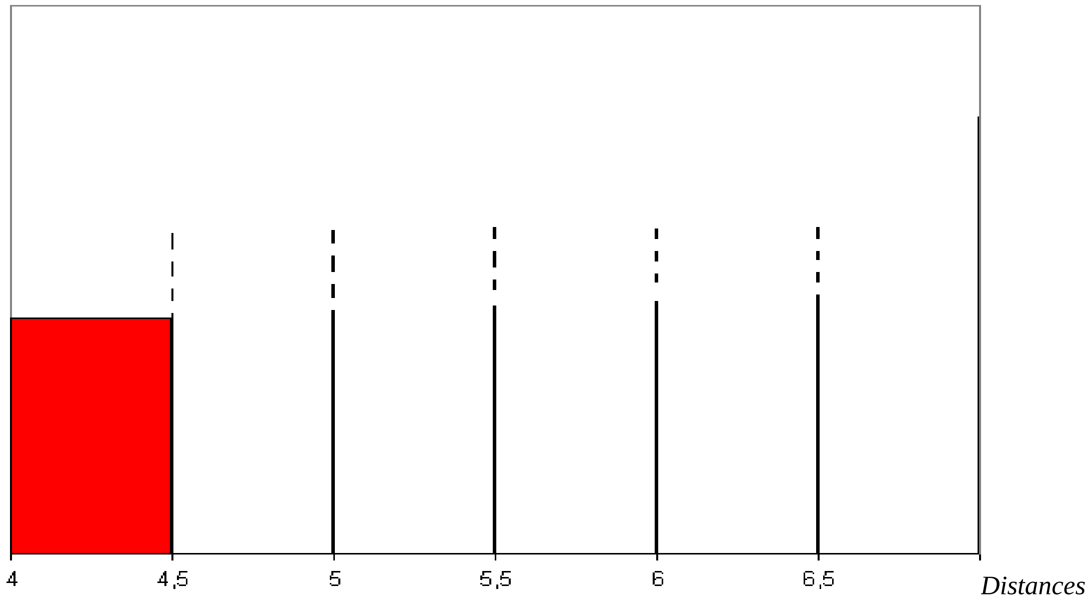
On utilise un **histogramme** pour représenter des données numériques regroupées en classes : *(Les données des abscisses doivent être continues)*
- sur l'axe des abscisses, on repère les classes
- sur l'axe des ordonnées, on repère les effectifs ou les fréquences

Dans le cas d'un histogramme, c'est **l'aire du rectangle qui est proportionnelle** à l'effectif. C'est pourquoi, il est souhaitable que toutes les classes soient de même longueur

Remarque : lorsque les valeurs sont réparties en classes d'égale amplitude, ce sont les hauteurs des rectangles qui sont proportionnelles aux effectifs des classes représentées.

Retour exemple3 : complétez le tableau suivant, et s'appuyer sur cela pour construire l'histogramme suivant.

Longueur	$4 \leq n < 4{,}5$	$4{,}5 \leq n < 5$	$5 \leq n < 5{,}5$	$5{,}5 \leq n < 6$	$6 \leq n < 6{,}5$	$6{,}5 \leq n < 7$	Maxi
effectif	3	6	5	5	5	6	6
Aire associée (en cm²)	……	……	……	……	……	……	12



V. Caractéristiques de position

1) Mode :

Définition5 : Soit *S* une série statistique quantitative discrète à une variable.
Un mode est une modalité de *S* de plus grand effectif.

2) Moyenne :

Définition5 : Pour obtenir la moyenne d'une série statistique :
- ✓ On multiplie chaque valeur (ou centre de classe) par l'effectif correspondant
- ✓ On additionne les produits ainsi obtenus
- ✓ On divise cette somme par l'effectif total

On parle dans ce cas de **moyenne pondérée** par les effectifs.

Remarques :

- Lorsque toutes les valeurs ont le même poids, la moyenne est le quotient de la somme de toutes les valeurs de la série par l'effectif total.
- Attribuer le coefficient 4 aux devoirs surveillés revient à les compter 4 fois dans la moyenne pondérée.
- Lorsque l'on calcule la moyenne de données regroupées en classes, c'est en fait une approximation de la moyenne qui est obtenue.

Exemple4 :

Un élève a obtenu les notes suivantes en mathématiques :

Devoirs surveillés *(coefficient 4)* : 17 ; 15 ; 8

Interrogations écrites *(coefficient 2)* : 15 ; 20 ; 5

Devoirs non surveillés *(coefficient 1)* : 17 ; 18 ; 20

Calculer la moyenne brute de cet élève en mathématiques *(c'est-à-dire sans tenir compte des coefficients)*, puis déterminez sa moyenne pondérée par les coefficients.

- *La moyenne brute est égale à :* ……………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
- *La moyenne pondérée est égale à :* ……………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

Retour exemple3 : Donner une approximation de la distance moyenne d'un saut

Longueur	$4 \leq n < 4{,}5$	$4{,}5 \leq n < 5$	$5 \leq n < 5{,}5$	$5{,}5 \leq n < 6$	$6 \leq n < 6{,}5$	$6{,}5 \leq n < 7$	Total
effectif	3	6	5	5	5	6	30
Centre de classe	……	……	……	……	……	……	

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

3) Une médiane

Définition2 : Une médiane d'une série statistique ordonnée est une valeur m telle qu'il y ait autant de valeurs supérieures ou égales à m que de valeurs inférieures ou égales à m.

- Lorsque la série est ordonnée et a un nombre impair de valeurs, alors il existe une unique valeur centrale : la médiane.

Exemple : La série suivante a un nombre impair de valeurs : 5

3 – 5,5 – 12,5 – 14 – 35

12,5 est la médiane

- Lorsque la série est ordonnée et a un nombre pair de valeurs, alors une médiane peut ne pas faire partie de la série.

Exemple : La série suivante a un nombre pair de valeurs : 6

3 – 4 – 6 – 7 – 10 – 12

On peut prendre comme médiane tout nombre compris entre 6 et 7, par exemple 6,5

Remarques :

- On peut déterminer une médiane à partir d'un graphique représentant les effectifs cumulés.
- Pour déterminer plus facilement une médiane d'une série statistique donnée, il faut ordonner les valeurs de la série.
- Deux séries peuvent avoir la même moyenne et deux médianes différentes, et réciproquement.

Retour exemple2 :

Notes	1	2	3	4	5	6	7	8	9	10
Effectifs	2	0	1	3	8	5	2	0	3	1

Déterminer une médiane de cette série statistique.

……………………………………………………………………………………………………
……………………………………………………………………………………………………

4) Les quartiles :

Définition :
Soit S une série statistique à une variable quantitative discrète ordonnée dans l'ordre croissant.

- Le **premier quartile Q_1** de S est la plus petite modalité de la série telle qu'au moins 25 % *(le quart)* des données lui soit inférieures ou égales.
- Le **troisième quartile Q_3** de S est la plus petite valeur de la série telle qu'au moins 75 % *(les trois quarts)* des données lui soient inférieures ou égales.

Activité d'introduction à la notion de dispersion :
Soient deux séries de notes :

Série A : 9 – 8 – 10 – 10 – 13
Série B : 0 – 1 – 10 – 19 – 20

1/ Déterminer la moyenne de chacune de ces séries
2/ Déterminer une médiane de chacune de ces séries
3/ Le travail semble-t-il aussi régulier chez ces deux élèves ?

Ces deux séries admettent même moyenne et même médiane pourtant, la répartition des notes semble bien différente dans ces deux séries. Ce sont les caractéristiques de dispersion qui vont permettre de donner une indication quant à la répartition des valeurs de part et d'autre des caractéristiques de position.

VI. Caractéristiques de dispersion

1) L'étendue :

Définition7 : L'**étendue** d'une série statistique est la différence entre la plus grande et la plus petite des modalités de cette série statistique.

Retour à l'activité d'introduction à la notion de dispersion :
Déterminer l'étendue des séries A et B de l'activité précédente
L'étendue de la série A est égale à……, celle de la série B est égale à……

Bilan : ……………………………………………………………………………………………………….

2) L'écart interquartile :

Définition : Si l'on désigne par Q_1 *(resp. Q_3)* le premier *(resp. le troisième quartile)* d'une série alors l'**écart interquartile** est obtenu par la différence $Q_3 - Q_1$

Retour à l'activité d'introduction à la notion de dispersion :
Déterminer l'écart interquartile des séries A et B de l'activité précédente.
Le premier quartile de la série A est ……, le troisième quartile de la série A est ……, et donc l'écart interquartile de la série A est égale à ………………………
Le premier quartile de la série B est ……, le troisième quartile de la série B est ……, et donc l'écart interquartile de la série B est égale à ………………………

Bilan : ……………………………………………………………………………………………………….
……

Remarque : il existe des séries ayant toutes ces caractéristiques identiques, c'est pourquoi d'autres paramètres seront étudiés dans les années futures.