

**TITRE DE LA LEÇON : STATISTIQUE****Discipline : Mathématiques****Sous-discipline : Statistique****Cycle : Collège**

-

Niveau : Troisième**I- GENERALITES**

Activité Dans une classe de 3^{ème}, on demande à 16 élèves : « Combien de jours par semaine travaillez-vous les mathématiques ? »

Réponses recueillies : 3 ; 0 ; 0 ; 2 ; 0 ; 0 ; 2 ; 0 ; 0 ; 2 ; 0 ; 3 ; 0 ; 0 ; 1 ; 0

-Que représentent les 16 élèves ?

-Que représentent ces jours de travail dans la semaine ?

-Que représente chaque réponse recueillie ?

-Quel est le caractère étudié. Donne sa nature

-Dépouillent ces résultats en dressant un tableau regroupant tous les élèves et toutes les fréquences des réponses recueillies.

Je retiens :

- **Définition** : La statistique est un ensemble de méthodes mathématiques qui consiste à collecter, à analyser et à interpréter des observations relatives à un même phénomène.
- **Le but** de toute étude statistique, est d'obtenir une information à partir des données recueillies par des observations, des expériences scientifiques ou des enquêtes, afin de prendre une décision
- **Terminologie ou vocabulaire de base en statistique**
 - ✓ **La population statistique**, est l'ensemble des individus sur lequel porte l'étude statistique ;
 - ✓ **Un échantillon**, est une partie de la population ;
 - ✓ **Une unité statistique ou individu**, un élément de la population ;
 - ✓ **Caractères statistiques** :
 - Un **caractère**, est une propriété ou une particularité ou un aspect commun(e) qui permet d'étudier une population statistique.

Les valeurs prises par un caractère (ou les différentes situations définies sur un caractère) sont appelées : **Modalités ou valeurs du caractère**

- Un caractère est dit : **caractère qualitatif**, lorsque ses modalités sont des qualités de l'individu (ou lorsque ses modalités ne sont pas exprimées par des nombres, mais par des codes.)

Exemples : Opinion, profession, sexe, état matrimonial, comportement, couleur, catégorie,...

- Un caractère est dit : **caractère quantitatif (ou une variable statistique)**, lorsque ses modalités sont des quantités mesurables (ou lorsque ses modalités sont exprimées par des nombres.)

Exemples : Taille, durée, température, note, prix, ...

Un caractère est dit : **caractère quantitatif discret**, lorsque ses modalités prennent un nombre fini des valeurs numériques

Un caractère est dit : **caractère quantitatif continu**, lorsque ses modalités prennent une infinité de valeurs numériques, souvent regroupées en intervalles de la forme $[a; b[$, appelés :Classes.

✓ Effectifs

— L'effectif d'une modalité x_i ou effectif ou encore fréquence absolue, (dans le cas d'un caractère quantitatif discret), est le nombre de fois que cette modalité x_i apparaît dans l'ensemble des objets étudiés. On note n_i : L'effectif de la i -ième modalité x_i .

Dans le cas d'une série continue, on répartit souvent les données par classes : $[a; b[$

— L'effectif total, est le nombre d'individus de la population entière. On le note N et on écrit:

$$N = \sum_{i=1}^p n_i \text{ ou } N = n_1 + n_2 + \dots + n_p$$

— La fréquence relative ou la fréquence d'une modalité x_i est notée f_i et est définie par :

$$f_i = \frac{n_i}{N} . \text{ Avec : } 0 \leq f_i \leq 1 ; \sum_{i=1}^n f_i = 1 . f_i = \frac{n_i}{N} \times 100 \text{ est la fréquence en pourcentage}$$

— Les **effectifs cumulés croissants** (ECC) d'une modalité x_i , sont la somme des effectifs des modalités inférieures ou égales à la modalité x_i

— Les **effectifs cumulés décroissants** (ECD) d'une modalité x_i , sont la somme des effectifs des modalités supérieures ou égales à la modalité x_i

✓ **Série statistique** : Une série statistique à un seul caractère ou à une seule variable X, est l'ensemble des couples $(x_i; n_i)$, c'est-à-dire, l'ensemble des résultats d'une étude.

On représente souvent une série statistique sous forme d'un tableau ou d'une liste des valeurs du caractère prises par chaque individu.

NB : Etudier une série statistique à une seule variable, consiste à étudier un aspect des éléments d'un ensemble donné.

• Regroupement des données en classes

Pour faciliter l'étude d'une série statistique dont l'étendue des valeurs est importante, on peut regrouper les valeurs du caractère en intervalles de la forme $[a; b[$, disjoints deux à deux, appelés :Classes ou classes de modalités.

L'effectif d'une classe : $[a; b[$, est le nombre d'individus de la population étudiée, dont les modalités appartiennent à $[a; b[$.

Le centre de classe est le réel : $C_i = \frac{a+b}{2}$, l'amplitude d'une classe est le réel : $a_i = b - a$ et la densité d'une classe est le réel : $d_i = \frac{n_i}{a_i}$.

Remarque

- Les données représentent l'ensemble des individus observés
- Les statistiques sont les résultats produits par la statistique, un ensemble de données numériques.
- Collecter les données : C'est les rassembler, les dépouiller, les regrouper et les présenter sous forme de tableaux
- Analyser les données : C'est les mettre en ordre logique ou les réduire

- *Interpréter les résultats ou données : C'est déduire des conclusions à partir des résultats obtenus.*
- *Les données peuvent varier d'un moment à un autre, d'un objet à un autre, d'un lieu à un autre. On parle pour cela de variable statistique*

Exercice1 Dans une classe de 3^{ème}, on demande à 16 élèves : « Combien de jours par semaine bossez-vous les mathématiques ? »

Réponses recueillies : 3 ; 0 ; 0 ; 2 ; 0 ; 0 ; 2 ; 0 ; 0 ; 2 ; 0 ; 0 ; 3 ; 0 ; 0 ; 1 ; 0

a-Quelle est la population étudiée ? Quel est le caractère étudié. Donne sa sature

b-Dresse un tableau regroupant tous les effectifs et toutes les fréquences.

c-Détermine le pourcentage des élèves qui travaillent les mathématiques 3 fois par jour.

II- Caractéristiques ou paramètres statistiques

Activité

Je retiens :

- **Mode ou dominante:** Le **mode**, noté : M_0 (pour un caractère quantitatif discret), est la valeur (ou la modalité) du caractère qui correspond à l'effectif le plus grand.

Dans le cas d'un caractère quantitatif continu, on parle (dans ce cours) seulement de classe modale : classe qui correspond au plus grand effectif.

NB : Une série statistique peut avoir un ou plusieurs modes

- **Moyenne :** Une série statistique quantitative discret $(x_i; n_i)$, d'effectif total N, est donnée par le tableau suivant :

Modalité x_i du caractère	x_1	x_2	...	x_p
Effectif n_i	n_1	n_2	...	n_p

On appelle moyenne de la série $(x_i; n_i)$, le nombre réel noté : \bar{X} tel que :

$$\bar{X} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_p \cdot x_p}{N}, \text{ soit : } \bar{X} = \frac{1}{N} \sum_{i=1}^p n_i \cdot x_i$$

Si les valeurs sont regroupées en classes $[a; b]$, on calcule la moyenne en choisissant comme valeurs du caractère, les centres des classes : $x_i = \frac{a+b}{2}$ et comme effectifs, les effectifs des classes.

Exemple Dans une classe de 22 élèves, il y a 4 filles et 18 garçons. Lors d'un devoir, les 4 filles obtiennent 13,7 de moyenne et les 18 garçons obtiennent 12,8 de moyenne. La moyenne de la classe est donc : $\bar{X} = \frac{4 \times 13,7 + 18 \times 12,8}{4 + 18} = 12,96$.

- **Médiane :**

La médiane M_e d'une série, dont les valeurs du caractère sont discrètes et rangées dans l'ordre croissant, est la valeur partageant la population en deux groupes de même effectif.

La classe médiane, est la première classe pour laquelle l'effectif cumulé croissant dépasse $\frac{N}{2}$ ou la fréquence cumulée croissante, dépasse 50 % ($ECC \geq \frac{N}{2}$ ou $FCC \geq 50\%$)

Pour calculer la médiane M_e , dans le cas d'une variable statistique discrète, dont les données sont non condensées :

- On ordonne toutes les modalités x_i par ordre croissant et on calcule l'effectif total N ;
- Si N est impair, alors la médiane $M_e = x_i$, où x_i est la modalité de rang : $\frac{N+1}{2}$

— Si N est pair, alors la médiane $M_e = \frac{x_i + x_{i+1}}{2}$, où x_i est la modalité de rang : $\frac{N}{2}$ et x_{i+1} est la modalité de rang : $\frac{N}{2} + 1$. $[x_i; x_{i+1}[$ est appelé : Intervalle médian

Exemples a) Série 1 : 5, 12, 13, 14, 15, 16, 18, 22, 25, 27, 31

On a $N = 11$; $\frac{N+1}{2} = \frac{11+1}{2} = 6$. Donc M_e est la 6^{ème} modalité : $M_e = x_6 = 16$

b) Série 1 : 5, 12, 13, 14, 15, 16, 18, 22, 25, 27, 31, 35

On a $N = 12$; $\frac{N}{2} = \frac{12}{2} = 6$; $x_6 = 16$ et $\frac{N}{2} + 1 = \frac{12}{2} + 1 = 7$; $x_7 = 18$. Donc :

$$M_e = \frac{x_6 + x_7}{2} = \frac{16 + 18}{2} = 17$$

Pour calculer la médiane M_e dans le cas d'une variable statistique discrète, dont les données sont condensées, on calcule les effectifs cumulés croissants (ou fréquences cumulées croissantes). La médiane M_e est donc la première modalité x_k dont le premier effectif cumulé croissant (ou FCC) est supérieur ou égal à $\frac{N}{2}$

Exercice Le tableau ci-dessous indique l'âge des jeunes élèves dans un collège de Brazzaville :

Age	12	13	15	16	19	20
Jeunes élèves	2	3	2	1	1	1

- 1) Quelle est la population étudiée ? Quel est le caractère étudié ? Donner sa nature.
- 2) Déterminer : le mode, la moyenne et la médiane de cette série statistique

Solution

- 1) La population étudiée, est les 10 jeunes élèves dans un collège de Brazzaville
Le caractère étudié est l'âge. C'est un caractère quantitatif discret.
- 2) L'effectif le plus élevé étant 3, alors le mode est : $M_0 = 13$

La moyenne est : $\bar{X} = \frac{1}{N} \sum_{i=1}^6 n_i \cdot x_i = \frac{2 \times 12 + 3 \times 13 + 2 \times 15 + 1 \times 16 + 1 \times 19 + 1 \times 20}{10} \Rightarrow \bar{X} = 14,8$

Age : x_i	12	13	15	16	19	20
Effectif : n_i	2	3	2	1	1	1
ECC	2	5	7	8	9	10

$$ECC \geq \frac{N}{2} \Leftrightarrow ECC \geq 5 \Rightarrow M_e = 13$$

Exercice Un éleveur a pesé les œufs pondus par ses poules durant une semaine : Ses relevés sont indiqués dans le tableau suivant :

Masse (en g)	[44; 48[[48; 52[[52; 56[[56; 60[[60; 64[
Nombres d'œufs	18	25	30	14	8

Calculer la médiane de cette série statistique.

$$ECC \geq \frac{3N}{4} \Leftrightarrow ECC \geq 11,25 \Rightarrow Q_3 = 15$$

Remarque : Une médiane n'est toujours pas une modalité du caractère

Une médiane peut ou ne pas être unique.

Deux séries différentes peuvent avoir la même moyenne et deux médianes différentes (ou la même médiane et deux moyennes différentes).

- **Etendue**

L'étendue d'une série statistique, est la différence entre la plus grande et la plus petite valeur de la série valeur : $E = X_{max} - X_{min}$

Plus l'étendue est grande, plus les valeurs sont dispersées.

Remarque : On ne parle d'étendue que dans le cas d'un caractère quantitatif.

- **Ecart absolu moyen ou écart moyen absolu**

Une série statistique quantitative discret $(x_i; n_i)$, d'effectif total N et de moyenne \bar{X} , étant donnée.

L'écart absolu moyen de X, est la moyenne arithmétique de la valeur absolue des écarts à la moyenne

$$e_m = \frac{n_1|x_1-\bar{X}| + n_2|x_2-\bar{X}| + \dots + n_p|x_p-\bar{X}|}{N}, \text{ soit : } e_m = \frac{1}{N} \sum_{i=1}^p n_i \cdot |x_i - \bar{X}|$$

L'écart absolu moyen signifie que les valeurs du caractères s'écartent en moyenne de e_m de la moyenne.

III-Représentations graphiques

Les représentations graphiques permettent de visualiser le résumé statistique que nous donnent la distribution de fréquences absolues, de fréquences relatives, de fréquences ou effectifs cumulés.

Les graphiques utilisés, dépendent de la nature du caractère

3-1 -Cas de caractère qualitatif : Les graphiques qu'on peut utiliser sont :

3-1-1 –Le digramme circulaire ou Camembert: C'est un graphique constitué d'un cercle divisé en secteurs, dont les angles au centre sont proportionnels aux effectifs ou fréquences.

L'angle α_i d'une modalité d'effectif n_i , est donné en degrés par : $\alpha_i = \frac{n_i}{N} \times 360^\circ = f_i \times 360^\circ$

On peut aussi utiliser le diagramme semi circulaire (avec $\alpha_i = \frac{n_i}{N} \times 180^\circ = f_i \times 180^\circ$)

3-1-2 –Le digramme en tuyaux d'orgue ou diagramme en barres : C'est un graphique qui, à chaque modalité d'un caractère qualitatif, associe un rectangle de base constante, dont la hauteur est proportionnelle à l'effectif ou à la fréquence. Les rectangles sont en général disjoints (séparés), verticaux ou horizontaux.

3-2 -Cas de caractère quantitatif : Les graphiques qu'on peut utiliser sont :

3-2-1 –Le digramme en bâtons : C'est un graphique qui, à chaque variable quantitative discrète, associe un segment (bâton), dont la hauteur est proportionnelle à l'effectif ou à la fréquence.

3-2-2 –L'histogramme : C'est un diagramme composé de rectangles contigus (collés), dont les hauteurs sont proportionnelles aux effectifs ou aux fréquences et dont les bases sont déterminées par les intervalles (classes).

Lors que les classes n'ont pas la même amplitude, pour construire l'histogramme, il est nécessaire de

calculer, pour chaque classe, la densité de classe : $d_i = \frac{n_i}{a_i}$ (ou $d_i = \frac{f_i}{a_i}$ pour l'histogramme des fréquences), puis affecter à chaque rectangle, une hauteur proportionnelle à la densité.

Dans le cas d'un caractère dont les valeurs sont groupées en p classes, on adjoint deux classes $[a_0; a_1[$ et $[a_p; a_{p+1}[$ affectées de l'effectif nul (0). On trace, alors sur le même graphique que l'histogramme (ou le diagramme en bâtons, dans le cas d'un caractère quantitatif discret), une ligne

polygone joignant les points d'abscisses : $\frac{a_i + a_{i+1}}{2}$ et d'ordonnées n_i , appelée : Polygone des effectifs ou polygone des fréquences.

Le polygone des effectifs cumulés croissants (ECC), est la ligne polygonale obtenue en joignant les points ayant pour abscisses, les limites supérieures de chaque classe et pour ordonnées les ECC.
Le polygone des effectifs cumulés décroissants (ECD), est la ligne polygonale obtenue en joignant les points ayant pour abscisses, les limites inférieures de chaque classe et pour ordonnées les ECD.

Graphiquement, la médiane est l'abscisse du point $(x_i; \frac{N}{2})$, intersection de ces deux polygones : $M_e = x_i$: N : effectif total.

Exercice2 Un éleveur a pesé les œufs pondus par ses poules durant une semaine : Ses relevés sont indiqués dans le tableau suivant :

Masse (en g)	[44; 48]	[48; 52]	[52; 56]	[56; 60]	[60; 64]
Nombre s d'œufs	18	25	30	14	8

- 1) Quelle est la population étudiée ? Quel est le caractère étudié ? Donner sa nature.
- 2) Déterminer la classe modale,
- 3) Calculer la moyenne de cette série statistique.
- 4) Représente cette série statistique par un histogramme