Chapitre ⑦ : Statistiques

I - Rappel :

1) Etude statistique : étude d'un phénomène ou d'un caractère qui caractérise les membres d'un groupe.

2) Population statistique : c'est l'échantillon ou le groupe qui fait l'objet de l'étude statistique et chaque membre s'appelle individu ou unité statistique.

3) Caractère : c'est le phénomène étudié qui est une propriété qu'on peut observer ou mesurer et est de deux types :

a - caractère quantitatif : caractère qu'on peut exprimer par des nombres (nombres d'enfants, les notes, âge, poids, longueur ...)

b - caractère qualitatif : caractère qu'on ne peut pas l'exprimer par des nombres (sexe, couleur, type de voiture ...)

4) Effectif : c'est le nombre des unités qui prennent une des valeurs du caractère, on le symbolise par n_i (nombre de répétitions)

5) Effectif total : c'est la somme de tous les effectifs symbolisé par N

6) Effectif cumulé d'une valeur du caractère est la somme ~~des~~ des effectifs de tous les valeurs qui sont inférieurs ou égale à cette valeur.

7) Fréquence : la fréquence d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total N $\quad f_i = \frac{n_i}{N}$

8) Fréquence cumulée d'une valeur est le quotient de l'effectif cumulé de cette valeur par l'effectif total N

9) Pourcentage : est le produit de la fréquence par 100

$$p_i = \frac{\text{Effectif}}{\text{Effectif total}} \times 100 = f_i \times 100$$

* Remarques importantes :

→ La somme de toutes les ~~valeurs~~ fréquences d'une série statistique est égale à 1

→ La fréquence cumulée d'une valeur est la somme de la fréquence de cette valeur et des fréquences des valeurs précédentes.

→ On a deux types de série statistique.

discrèt en valeurs	continu en classes
si le nombre des valeurs est petit et on les range dans l'ordre croissant	Si le nombre des valeurs est élevé et on les regroupe en classes de même amplitude $a \leq x < b$

II - Tableau des effectifs, des effectifs cumulés, des fréquences et des fréquences cumulés

1) Série statistique discrète : en valeurs

Si le caractère est quantitatif et le nombre des valeurs est petit, on les range dans l'ordre croissant.

* Application ① :

Le tableau suivant représente une série statistique qui exprime la répartition de 24 adhérent dans un club sportif selon leur âge :

Caractère	12	13	14	15	16
Effectif	5	6	x	8	4
Effectif cumulée		11			24
fréquence					
fréquence cumulée					

1) La population statistique est 24 adhérent dans un club

2) L'unité statistique est un adhérent.

3) Le caractère étudié est l'âge de l'adhérent qui est un caractère quantitatif discret.

4) Le nombre d'adhérents x dont leurs âges est 14 ans :

On a l'effectif total est $N=24$

On a : $5+6+x+8+4=24$

$23+x=24$

$x=24-23 \Rightarrow \boxed{x=1}$

5) Remplissons le tableau précédent :

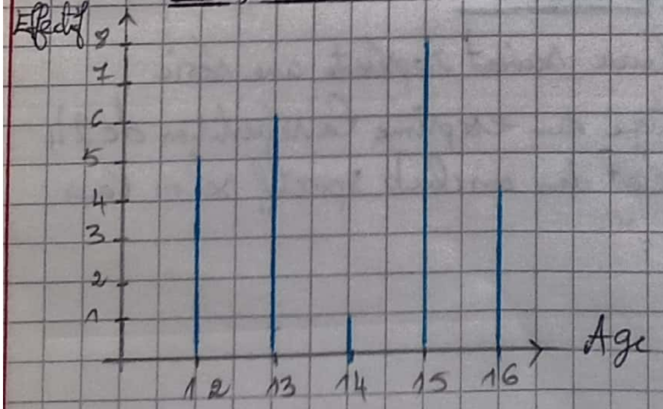
Caractère	12	13	14	15	16
Effectif	5	6	1	8	4
Effectif cumulé	5	11	12	20	24
fréquence	$\frac{5}{24}=0,21$	$\frac{6}{24}=0,25$	$\frac{1}{24}=0,04$	$\frac{8}{24}=0,33$	0,17
fréquence cumulé	0,21	0,46	0,50	0,83	1

6) On a $0,21+0,25+0,04+0,33+0,17=1$

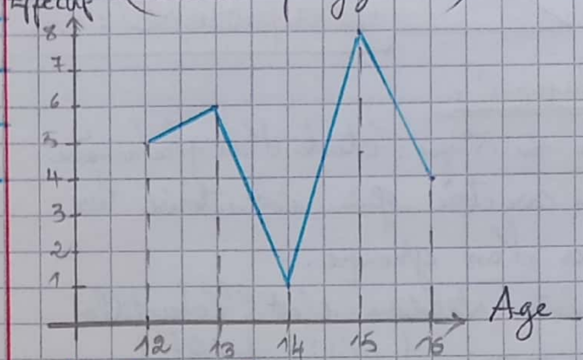
Alors la somme de tous les fréquences est égale à 1

7) Représentation graphique

a – Diagramme en bâtons :



b – Diagramme à ligne brisée (Courbe polygonale)



* Application ② :

Une étude statistique a été menée sur le nombre d'enfants dans 20 familles et a donné les résultats suivantes ;

2 - 3 - 4 - 3 - 0 - 4 - 3 - 2 - 1 - 1 - 2 - 1 - 0 - 2 - 3 - 4 - 1 - 3 - 0 - 1

1) Donner le tableau des effectifs et effectifs cumulées de cette série.

2) Calculer la fréquence de la valeur 0

3) Calculer le pourcentage de la valeur 0

4) ~~Dés~~ Calculer le pourcentage des nombre de familles dont le nombre d'enfant dépasse 2

5) Représenter cette série en colonnes

⇒ Solution :

1) Le tableau des effectifs et effectifs cumulées.

Caractère : nombre d'enfants	0	1	2	3	4
Effectif : nombre de familles	3	5	4	5	3
Effectif cumulée	3	8	12	17	20

2) Soit f la fréquence de la valeur 0, donc

$f=\frac{n}{N}=\frac{3}{20}=0,15$

3) Le pourcentage de la valeur 0 est :

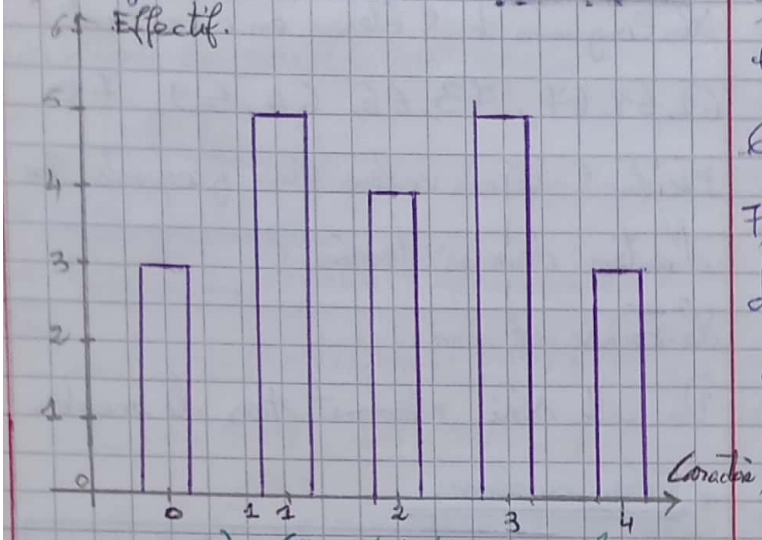
$p=f\times100=0,15\times100=15\%$

4) Le nombre de familles dont le nombre d'enfants dépasse 2 est ; $n=5+3=8$

donc le pourcentage est : $p=f\times100=\frac{n}{N}\times100$

$=\frac{8}{}\times100=40\%$

5) Diagramme en colonnes.



2) Série statistique en classes :

Si le caractère est quantitatif et le nombre de ses valeurs est élevé, au lieu d'étudier toutes les valeurs, on les regroupe dans des intervalles de même amplitude $[a,b[$ appelés des classes. et le centre de la classe est $\frac{a+b}{2}$

* Application ③ :

Le bilan suivant donne la répartition d'âges des ouvriers dans une ferme.

16 _ 26 _ 34 _ 17 _ 22 _ 45 _ 36 _ 27 _ 29 _ 25 _ 19 _ 48 _ 32 _ 42 _ 21 _ 33 _ 35 _ 16 _ 26 _ 34 _ 17 _ 22 _ 38 _ 36 _ 27 _ 29 _ 38 _ 13 _ 18 _ 32 _ 30 _ 39.

1) Déterminer la population de cette série

2) Déterminer le caractère étudié et son type.

3) Compléter le tableau suivant :

Age en année	$[10,20[$	$[20,30[$	$[30,40[$	$[40,50[$
centre de classe				
Effectif : nombre d'ouvriers				

4) Combien d'ouvriers dans la ferme ?

5) Calculer le pourcentage des ouvriers dont l'âge est inférieur à 20 ans.

6) Calculer la fréquence de la classe $[30,40[$

7) Construire l'histogramme de la répartition des ouvriers de la ferme selon les classes de leurs âges.

⟹ Solution :

1) La population est les ouvriers de la ferme

2) La caractère statistique est l'âge d'ouvrier et c'est une caractère quantitatif continu.

3)

Age en année	$[10,20[$	$[20,30[$	$[30,40[$	$[40,50[$
Centre de classe	$\frac{10+20}{2}=15$	25	35	45
Effectif : nombre d'ouvriers	8	10	12	2

4) Le nombre d'ouvrier dans la ferme est :

$N = 8+10+12+2 = 32$

5) Le nombre d'ouvriers dont l'âge est inférieur à 20 ans est $n = 8$

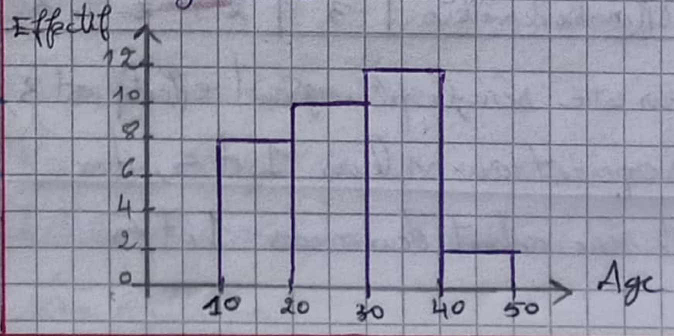
donc leur pourcentage est :

$p = \frac{n}{N} \times 100 = \frac{8}{32} \times 100 = 25\%$

6) La fréquence f de la classe $[30,40[$ est :

$f = \frac{n}{N} = \frac{12}{32} = 0,375$

7) Si la répartition est sous forme de classes $[a,b[$ ou d'intervalles $a \leq x < b$, on emploie un histogramme.



⇒ Diagramme circulaire :

Convenable pour les caractères qualitatifs.

Un diagramme circulaire est un diagramme qui a la forme d'un disque décomposé en secteurs dont les mesures des angles sont proportionnelles aux effectifs (et également aux fréquences)

$$\text{Mesure de l'angle} = \frac{\text{Effectif}}{\text{Effectif total}} \times 360°$$
$$= \text{fréquence} \times 360°$$
$$= \frac{\text{pourcentage}}{100} \times 360°$$

III _ Les paramètres de position :

1) Le mode :

a_ Définition :

Le mode d'une série statistique est la valeur du caractère qui a le plus grand effectif.

b_ Exemples :

* Série en valeurs :

→ Exemple ①: Regarde Application ①

Caractère	12	13	14	15	16
Effectif	5	6	1	8	4

On le plus grand effectif est 8 et ~~sa~~ sa valeur est 15

donc le mode de cette série est 15

→ Exemple ②:

Caractère (coefficient)	1	2	3	5
Effectif (nombre de matières)	3	2	2	3

Pour cette série, le plus grand effectif est 3, correspondant aux valeurs 1 et 5, donc cette série admet deux modes 1 et 5

→ Exemple ③:

Les longueurs de 8 élèves en cm sont :

61, 68, 67, 73, 66, 64, 59, 70

Existe-t-il une valeur plus fréquente que d'autres dans ces données ?

La réponse est non.

Donc cette série n'admet pas de mode.

* Série en classes :

Définition : On appelle classe modale d'une série statistique groupées en classe, toute classe qui a le plus grand effectif.

⇒ Exemple : On considère la série suivante :

Classe	[120, 130[	[130, 140[	[140, 150[	[150, 160[
Effectif	9	11	12	18

Le plus grand effectif est 18, correspond à la classe [150, 160[, donc la classe modale est [150, 160[

Remarques

Une série statistique peut ne pas avoir un mode (classe modale), comme elle peut avoir plusieurs modes (classe modal)

2) La moyenne arithmétique :

a_ Définition :

La moyenne arithmétique est le rapport de la somme de toutes les produits de chaque valeurs (centre de classe) fois son effectif ~~par~~ sur l'effectif total. On le note m.

Remarque: La moyenne arithmétique m est égale au rapport de la somme de toutes les valeurs du caractère sur l'effectif totale c'est à dire c'est la valeur obtenue si toutes les valeurs du caractère sont égaux

b- Exemples:

*** Série en valeurs:**

→ Exemple ①: Regarde l'application ②:

On a:

$$m = \frac{(0\times3)+(1\times5)+(2\times4)+(3\times5)+(4\times3)}{20}$$

$$= \frac{0+5+8+15+12}{20} = 2$$

Donc 2 est la moyenne du nombre d'enfants dans chaque famille.

→ Exemple ②: On considère la série suivante:

Caractère (coefficient)	1	2	3	5
Effectif (nombre de matières)	3	2	2	3

$$m = \frac{(1\times3)+(2\times2)+(3\times2)+(5\times3)}{10}$$

$$= \frac{3+4+6+15}{10} = \frac{28}{10}$$

$$\boxed{m = 2,8}$$

*** Série en classes**

→ Règle: Si $a \leq x < b$ est une classe d'une série, alors son milieu est $\frac{a+b}{2}$

* Pour calculer la moyenne d'une série statistique en classes, on utilise la définition précédente en remplaçant les valeurs par les centres des classes.

→ Exemple: Regarde l'application ③

Age (années)	[10,20[	[20,30[	[30,40[	[40,50[
Centre de classe	$\frac{10+20}{2}=15$	25	35	45
Effectif: nombre d'ouvriers	8	10	12	2

$$m = \frac{(15\times8)+(25\times10)+(35\times12)+(45\times2)}{32}$$

$$\boxed{m = 27,5}$$

Alors la moyenne d'âges des ouvriers est 27,5 et cela signifie que si on a supposé que tous les ouvriers ont le même âge, l'âge de chaque ouvrier va être 27,5 ans

3) La médiane:

a- Définition:

La médiane d'une série statistique est la plus petite valeur du caractère dont l'effectif cumulé est supérieur ou égale à la moitié de l'effectif total.

b- Exemples:

*** Série en valeurs:**

Regarde l'application ①

Caractère	12	13	14	15	16
Effectif	5	6	1	8	4
Effectif cumulé	5	11	12	20	24

- La moitié de l'effectif total est $\frac{24}{2} = 12$
- La plus petite valeur dont l'effectif cumulé est supérieur ou égale à 12 est 14

La médiane est donc égale à 14.

*** Série en classes:**

Classe	[120,130[	[130,140[	[140,150[	[150,160[
Effectif	9	11	12	18
Effectif cumulé	9	20	32	50

- La moitié de l'effectif totale est $\frac{50}{2} = 25$
- Le premier effectif cumulé supérieur ou égale à 25 est 32, correspond à la classe [140, 150[

Donc la médiane de cette série se trouve dans la classe [140, 150[

Remarque: On peut dire que 145 (centre de la classe [140,150[) est la médiane de cette série statistique.

* **autre définition : médiane :**

La médiane d'une série statistique, dont les valeurs sont ordonnées, est la valeur du caractère qui partage la série en deux parties de même effectif.

→ **Cas ①** : effectif total N est impair

* **Exemple :** Nombre d'absences pendant 7 jours d'ouvriers d'une société sont :

3 - 1 - 2 - 0 - 4 - 2 - 3

On les range dans l'ordre croissant :

$$\underbrace{0 - 1 - 2}_{3 \text{ valeurs}} - 2 - \underbrace{3 - 3 - 4}_{3 \text{ valeurs}}$$

Donc la médiane est 2

→ **Cas ②** : N est pair

* **Exemple :** Nombre d'absences pendants 8 jours ordonné.

$$\underbrace{0 - 1 - 1 - 2}_{4 \text{ valeurs}} - \underbrace{3 - 4 - 4 - 5}_{4 \text{ valeurs}}$$

$$M = \frac{2+3}{2} = \frac{5}{2} = 2,5$$

On peut prendre comme médiane tout nombre compris entre 14 et 16

d'où la médiane de cette série est 2,5

VI - La dispersion :

1) Définition :

On considère deux séries statistiques S_1 et S_2 qui ont la même moyenne arithmétique m. On dit que la série S_1 est moins dispersée que S_2 si les valeurs de S_1 sont plus proches de m que celles de S_2.

2) Exemple :

On considère le tableau suivant :

Devoir	D_1	D_2	D_3	D_4	D_5
Notes de Hassan	9	14	10	13	14
Notes de Khaled	8	16	10	17	9

Moyenne de Hassan

$$m_1 = \frac{9+14+10+13+14}{5} = \frac{60}{5} = 12$$

Moyenne de Khaled

$$m_2 = \frac{8+16+10+17+9}{5} = \frac{60}{5} = 12$$

donc $m_1 = m_2$

C à d Hassan et Khaled ont même moyenne

On remarque que les notes de Hassan sont plus proches de la moyenne 12 que celles de Khaled

* On dit que : Les notes de Hassan sont moins dispersés que les notes de Khaled.