

Statistique

I- Vocabulaire de la statistique

Evaluation diagnostique

Activité

Voici une liste de notes attribuées à 30 élèves d'une terminale :

- 13 personnes ont une note comprise entre 0 et 3.
- 9 personnes ont une note comprise entre 3 et 8.
- 4 personnes ont une note comprise entre 8 et 10.
- 3 personnes ont une note comprise entre 10 et 15.
- 2 personnes ont une note comprise entre 15 et 18.

Utilise le tableau ci-dessous pour classer ses notes dans des intervalles bien déterminés.

Nombre d'élève	Intervalles de notes attribuées

Solution

Nombre d'élève	Intervalles de notes attribuées
----------------	---------------------------------

13	[0 ; 3 [
9	[3 ; 8 [
4	[8 ; 10 [
3	[10 ; 15 [
1	[15 ; 18 [

Le fait que tu es classé tes notes donne la notion de statistique.

1- Définition :

D₁ : La statistique est un ensemble de données chiffrées ou non.

D₂ : La statistique est la branche de la mathématique qui consiste à regrouper les données ; à les analyser et à tirer des conclusions.

Comme toutes les sciences, la statistique fait appel à un langage spécialisé.

2- Langage statistique :

a- Notion de population et d'individu :

L'ensemble sur lequel porte l'étude statistique est appelé la population, un élément de cet ensemble est l'individu.

N.B : En statistique, les mots (population – individu) peuvent désigner des êtres ou des objets animés ou inanimés.

b- Caractères et modalités :

Dans l'étude de statistique d'une population, l'attention est portée généralement sur un événement commun à tous les membres de la population : c'est le caractère.

Exemple : Le caractère situation matrimoniale à 4 modalités : marié, divorcé, célibataire, veuf.

Il existe deux sortes de caractères : **Les caractères qualitatifs et les caractères quantitatifs.**

- **Caractères qualitatifs :**

Un caractère est dit qualitatif si ses différentes modalités ne sont pas mesurables.

Exemple : La situation matrimoniale ; le sexe, la profession, la nationalité....

- **Caractères quantitatifs :**

Un caractère est dit quantitatif si ses différentes modalités sont mesurables c'est à dire des nombres.

Exemple : l'âge, la taille, le poids.....

Un caractère quantitatif peut être discret (discontinue) ou (continue)

Un caractère quantitatif est dit discret si ses valeurs prises sont isolées.

Exemple : le nombre de frère de chaque élève, le nombre de femme.....

Un caractère quantitatif est dit continu si ses valeurs prises sont en nombre infini dans un intervalle.

Exemple : la température du corps, la taille, l'âge...;

c- **Echantillon :**

L'étude d'un caractère sur une grande population est parfois sensible et coûteux.

Exemple : recensement de population, enquête etc... c'est pourquoi on étudie ce caractère sur un petit groupe de cette membre de cette population : c'est l'échantillon.

II- Présentation des données statistiques :

Plusieurs méthodes sont utilisées par les organismes de la statistique pour la collecte des données statistiques. Les données ayant été recueillis et est nécessaire dans une première phase de les grouper, de les présenter sous forme de tableau et de représentation graphique.

1- Cas d'un caractère qualitatif :

a- Tableau statistique :

Exemple : l'étude du caractère situation matrimoniale sur 10 élèves d'une classe a donné le résultat suivant :

A : marié

B : Divorcé

C : Célibataire

D : Célibataire

E : Marié

F : Veuf

G : Divorcé

H : Célibataire

I : Célibataire

J : Marié

Modalité (x_i)	Nombre d'élément (n_i)
Marié	3
Célibataire	4
Divorcé	2
Veuf	1
Total	10

Le caractère situation matrimoniale à 4 modalités; on dresse un tableau à deux colonnes.

- 1^{ère} Colonne contient la nomenclature des différentes modalités. On la note x_i
- La 2^{ème} Colonne contient le nombre d'élément de chaque modalité. On la note n_i .

n_i est l'effectif relatif de la modalité x_i . C'est le nombre d'élément présentant la modalité x_i .

Exemple : $n_1 = 3$, $n_2 = 4$, $n_3 = 2$, $n_4 = 1$ (d'après le tableau).

Alors $N = \sum n_i = n_1 + n_2 + n_3 + n_4 = 10$

N est appelé somme des n_i .

A ce tableau, on peut ajouter une 3^{ème} colonne ; celle des **fréquences notés f_i** tel que $f_i =$

$$\frac{n_i}{N} = \frac{\text{effectif relatif}}{\text{effectif total}}$$

Modalités	Nombre d'éléments	Fréquences
(x_i)	(n_i)	(f_i)
Marié	3	$3/10 = 0,3$
Célibataire	4	$4/10 = 0,4$
Divorcé	2	$2/10 = 0,2$
Veuf	1	$1/10 = 0,1$
Total	10	$10/10 = 1$

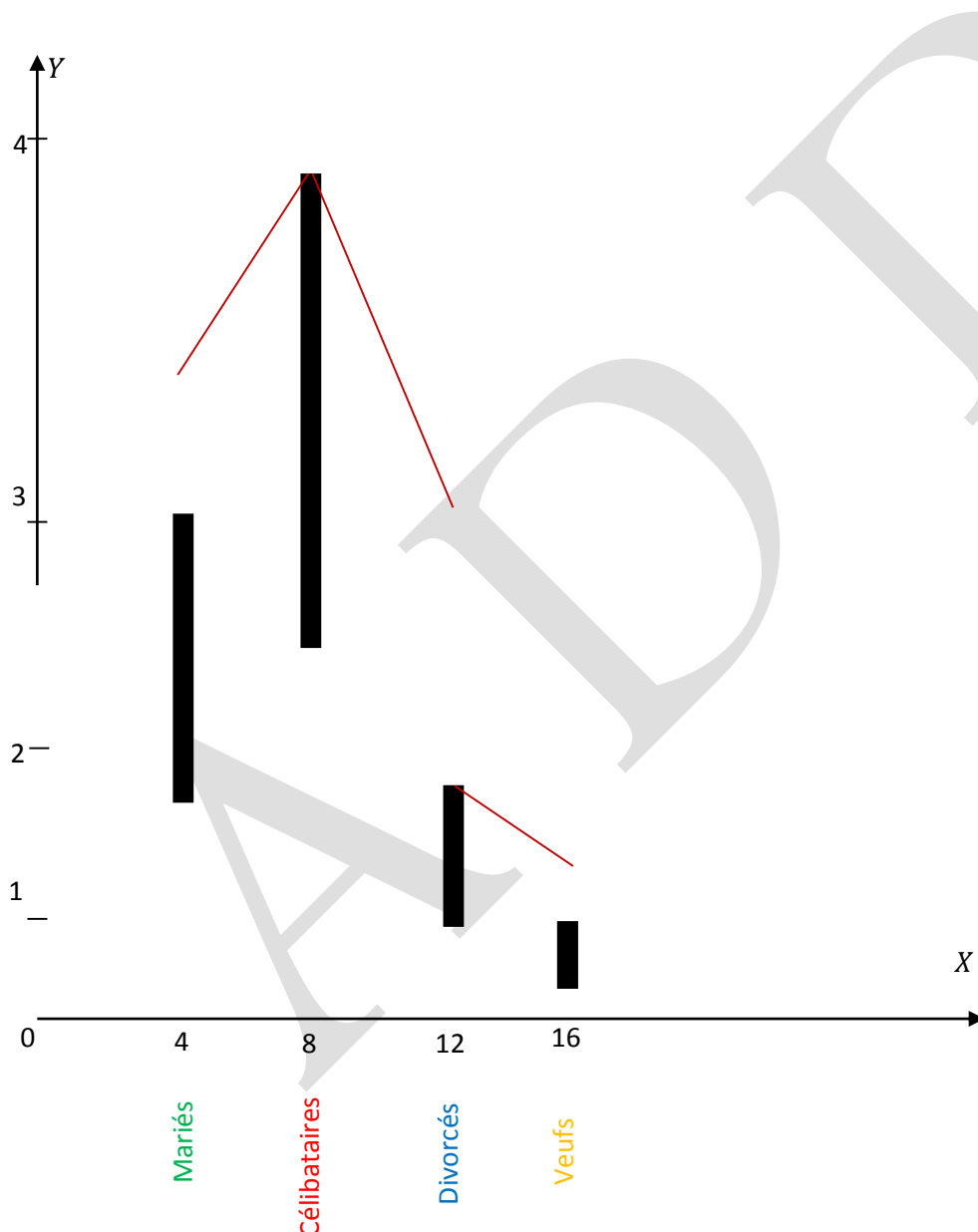
NB : la somme des fréquences est toujours égale à 1 : $\sum f_i = 1$

b- Représentation graphique :

Diagramme en bâton :

Dans un repère orthogonal, on place sur l'axe des abscisses les x_i (modalités) ; et sur l'axe des ordonnées les effectifs n_i ou les fréquences f_i .

Pour construire le diagramme en bâton des effectifs, on place les points $A_i(x_i ; 0)$ et $B_i(x_i ; n_i)$ et on construit en trait plein les segments $[A_i B_i]$.



N.B :

Si les fréquences f_i étaient en ordonnée, le digramme est appelé diagramme en bâton des fréquences. Le polygone des effectifs est la ligne brisée reliant les extrémités des bâtons dans le diagramme (Voir figure ci-dessus).

D'une manière générale, la ligne polygonale obtenue en joignant les différents points $B_i(x_i ; n_i)$

Diagramme en secteur circulaire :

Soit le demi – cercle (C) ; représentons la série précédente sur (C). Chaque modalité est représenté par un secteur angulaire proportionnel à son effectif α_i tel que $\alpha_i = f_i \times 180^\circ$.

Marié : $\alpha_1 = 0,3 \times 180^\circ \Rightarrow \alpha_1 = 54^\circ$

Célibataire : $\alpha_2 = 0,4 \times 180^\circ \Rightarrow \alpha_2 = 72^\circ$

Divorcé : $\alpha_3 = 0,2 \times 90^\circ \Rightarrow \alpha_3 = 18^\circ$

Veuf : $\alpha_4 = 0,1 \times 180^\circ \Rightarrow \alpha_4 = 18^\circ$

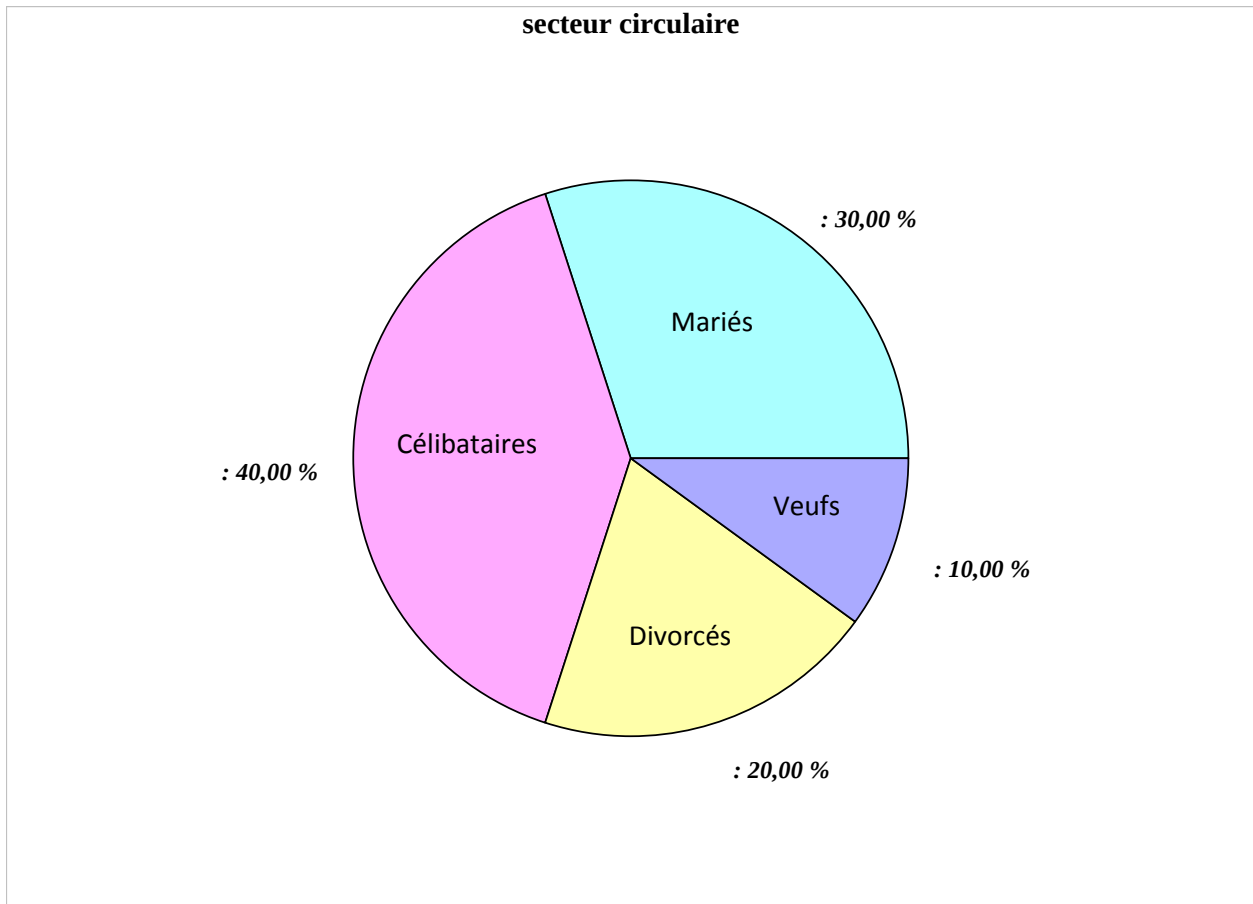
NB : pour le cercle tout entier le secteur angulaire α_i est : $\alpha_i = f_i \times 360$

$$\alpha_1 = 108^\circ$$

$$\alpha_2 = 144^\circ$$

$$\alpha_3 = 72^\circ$$

$$\alpha_4 = 36^\circ$$



2- Cas d'un caractère quantitatif discret :

a- Etude du caractère nombre de frère dans un tableau statistique

L'étude du caractère nombre de frère dans une classe de 12^e a donné les résultats suivants : 4 ; 2 ; 2 ; 1 ; 3 ; 2 ; 3 ; 3 ; 2 ; 1.

Regrouper ces données dans un tableau statistique.

Modalités x_i	Effectif n_i	Fréquence f_i	Effectif cumulé croissant	Effectif cumulé décroissant	Fréquence croissante	Fréquence Décroissante
1	2	0,2	2	10	0,2	1
2	4	0,4	6	8	0,6	0,8
3	3	0,3	9	4	0,9	0,4
4	1	0,1	10	1	1	0,1
Total	10	1	/	/	/	/

Le regroupement en tableau se fait de la même manière que les caractères qualitatifs. La seule différence est que les modalités sont des nombres s'écrivant dans l'ordre croissant. A ce tableau, on peut ajouter les colonnes, les effectifs cumulés croissants et décroissants ainsi que les fréquences cumulés croissantes et de croissantes. Les valeurs du caractère étant rangées dans l'ordre croissant, on appelle :

- **Effectif cumulé croissant de x_i :**

La somme des modalités inférieures ou égales à x_i c'est-à-dire :

$$n_i = (n_1 + n_2) + \dots + n_i$$

$$n_1 = n_1 = 2$$

$$n_2 = n_1 + n_2 + n_3 = 2 + 4 = 0$$

$$n_3 = n_1 + n_2 + n_3 = 2 + 4 + 3 = 9$$

$$n_4 = n_1 + n_2 + n_3 + n_4 = 2 + 4 + 3 + 1 = 10 \text{ et la fréquence croissante cumulé est } f_i = \frac{n_i}{N = 10}$$

- **Effectif cumulé décroissant de x_i :**

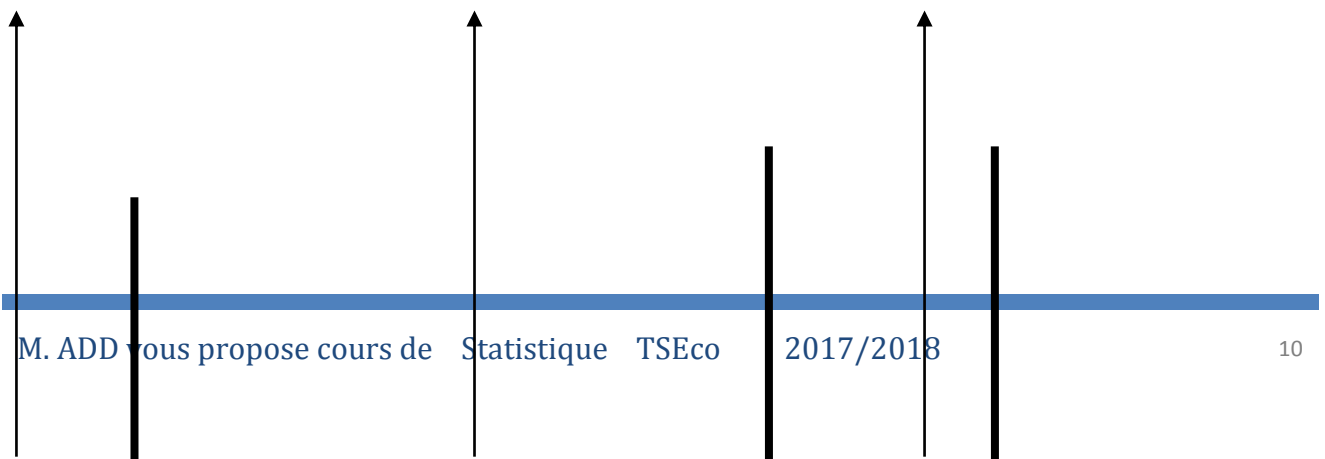
On appelle effectif cumulé décroissant de la modalité x_i , la somme des modalités qui le succède ; c'est-à-dire $n_i = n_i + n_{i+1} + \dots + n_k$ et la fréquence cumulée décroissante est : $f_i = \frac{n_i}{N}$

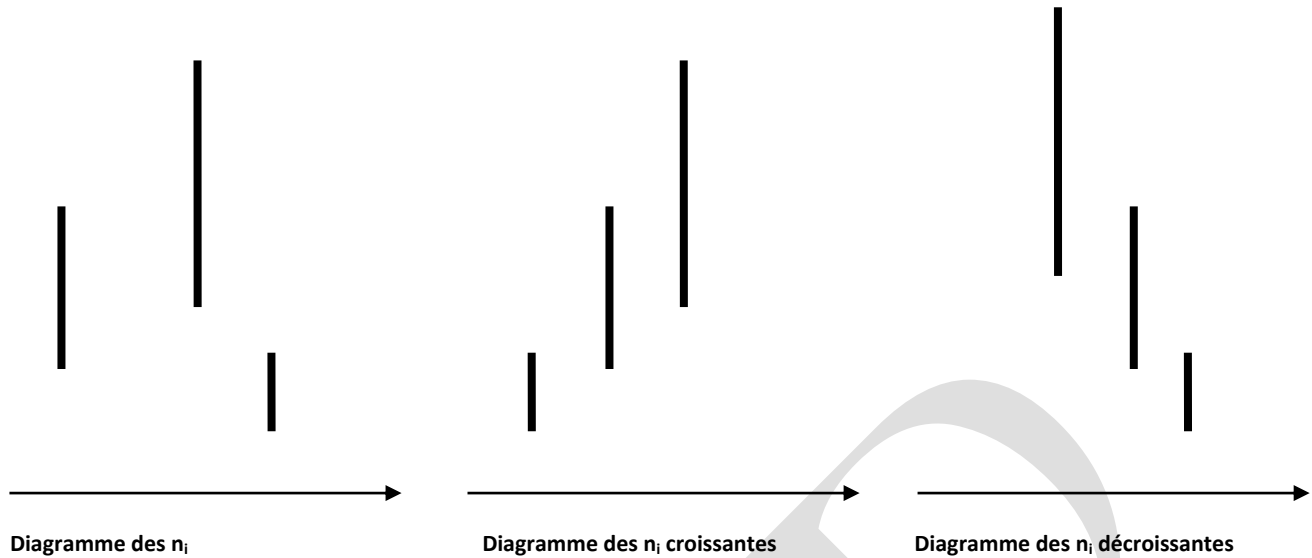


b- Diagramme en bâton des effectifs cumulés :

Dans un repère orthogonal, on place sur l'axe des abscisses les x_i (modalités) ; et sur l'axe des ordonnées les effectifs n_i cumulés croissants ou décroissants.

Pour construire le diagramme en bâton des effectifs cumulés croissants ou décroissants, on place les points $A_i(x_i ; 0)$ et $B_i(x_i ; n_i)$ et on construit en trait plein les segments $[A_i B_i]$.





3- Cas des caractères quantitatifs continus :

a- Tableau statistique :

Dans le cas d'un caractère quantitatif continu, les valeurs du caractère sont regroupées dans un intervalle de même amplitude $[x_i; x_{i+1}]$ soit appelées des classes.

Exemple : l'étude du caractère taille sur 30 élèves d'une classe de 12^e a donné les résultats suivants en cm :

175	165	172	171	168
170	178	170	171	173
170	170	168	150	160
165	175	179	180	170
173	165	168	176	150
174	158	153	144	172

Dresser le tableau des effectifs et des fréquences cumulées.

Prenons des classes d'amplitude 6.

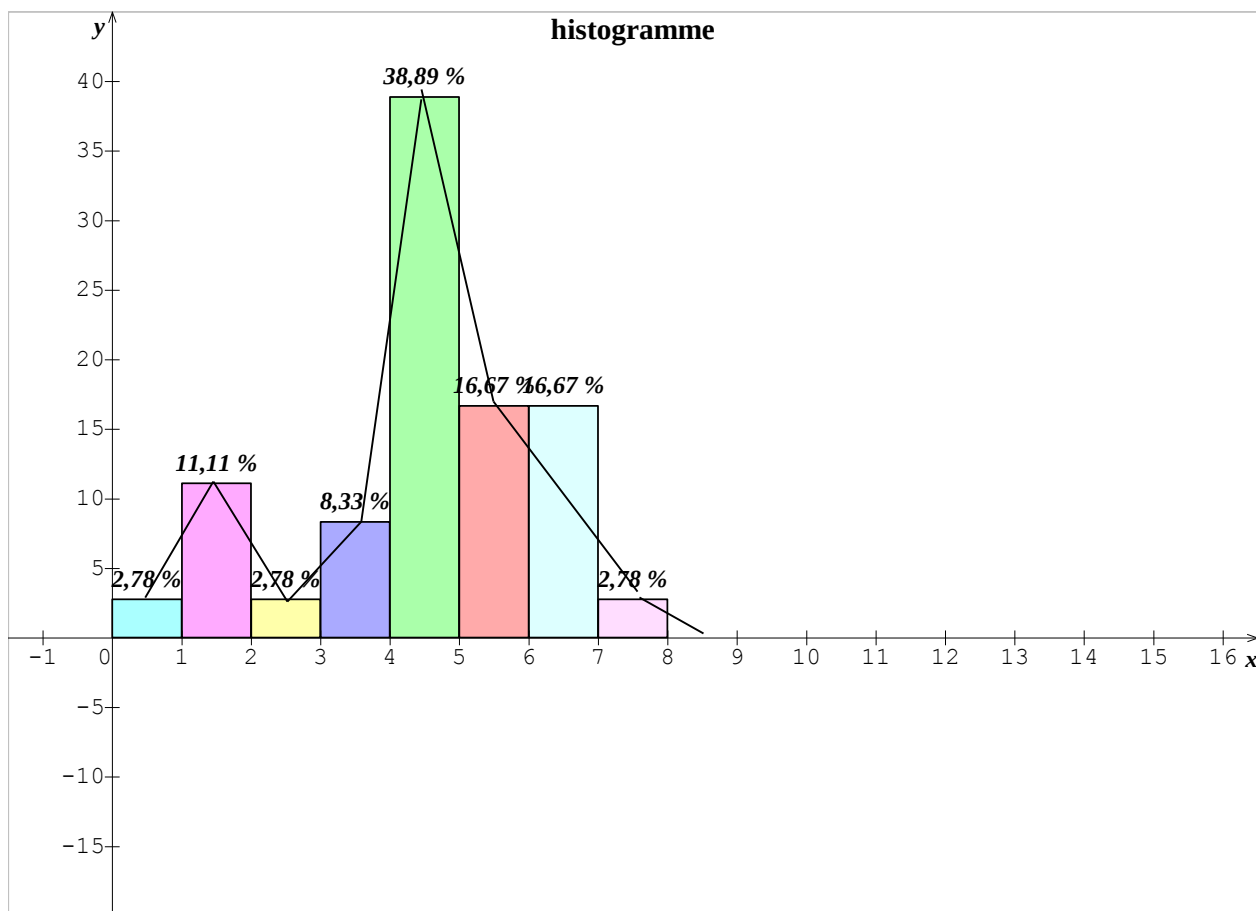
Classes	Effectif	Centre des classes	n_i	n_i croissante	f_i
$[x_i ; x_{i+1}]$	n_i	c_i			
[144 ; 150]	1	147	1	30	0,633
[150 ; 156]	4	153	5	29	0,166
[156 ; 162]	1	159	6	25	0,2
[162 ; 168]	3	165	9	24	0,3
[168 ; 174]	14	171	23	21	0,76
[174 ; 180]	6	177	29	7	0,966
[180 ; 186]	1	183	30	1	1
Total	30	—	—	—	—

On appelle centre (c_i) de la classe (x_i, x_{i+1}) le nombre $c_i = \frac{x_i + x_{i+1}}{2}$

Pour les calculs, nous prendrons de prendre la valeur c_i commune à tous les individus de la classe (x_i, x_{i+1})

ADD

Histogramme des effectifs (représentation graphique)



Dans un repère cartésien, les valeurs des bornes de classe sont portées en abscisse et les effectifs (ou fréquence) sont portés en ordonnées.

Pour chaque classe $[x_i ; x_{i+1}[$, on trace le rectangle de base $[x_i ; x_{i+1}[$, et de hauteur (n_i ou f_i). La représentation graphique obtenue est appelée : **Histogramme des effectifs ou des fréquences**.

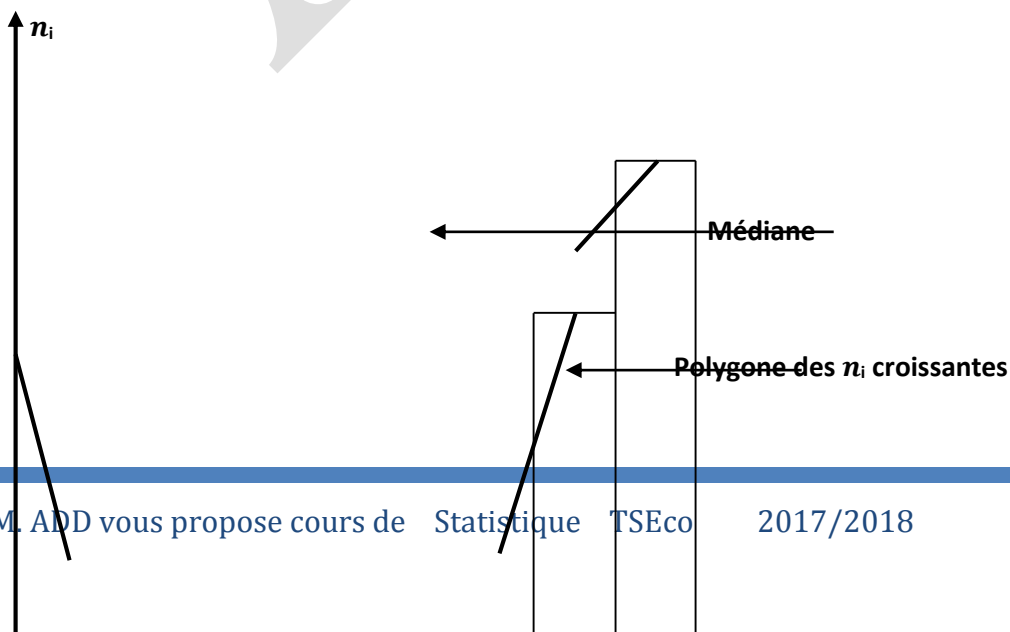
Les polygones :

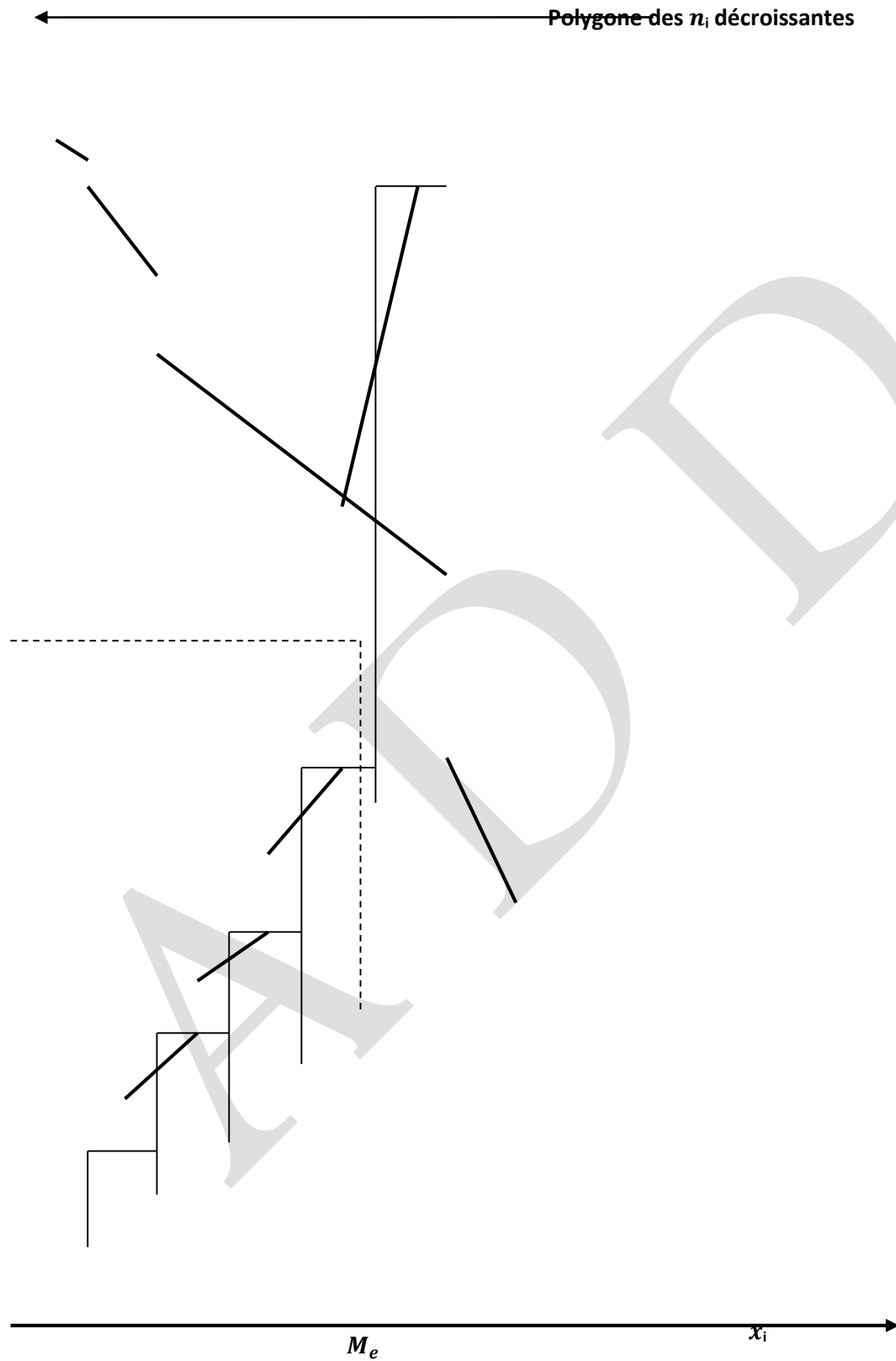
On appelle polygones les effectifs, la ligne brisée joignant les points $(c_i ; n_i)$.

On appelle polygone des effectifs cumulés croissants, la ligne brisée passant par les points d'abscisse et d'ordonnées $(x_{i+1} ; n_i)$.

On appelle polygone des effectifs cumulés décroissants, la ligne brisée joignant les points de coordonnées $(x_i ; n_i)$ ➤

Les deux polygones des effectifs cumulés croissant et décroissant se coupent en un point d'abscisse M_e : (M_e est la médiane).





III- Les éléments caractéristiques d'une série :

1- Les caractéristiques de position :

a- Le mode ou dominantes :

Le mode d'une série statistique est la valeur du caractère qui a le plus grand effectif (ou fréquence). On le note M_o .

Exemple : Déterminer le mode de la série suivante :

10, 7, 8 ; 5, 2, 7 ; 4, 4, 5 ; 7, 10

Ici : le mode est 7 (car 7 est compté plus que les autres)

b- La médiane d'une série :

Définition : La médiane d'une série est la valeur du caractère qui divise la série en deux séries de même effectif. On la note M_e .

Exemple 1 : soit la série suivante :

11, 7, 8, 3, 5, 6, 2

Déterminer la médiane. Pour cela écrivons la série dans l'ordre croissant :

2, 3, 5, 6, 7, 8, 11

La médiane est donc $Me = 6$

Exemple 2 : soit la série suivante : 12, 3, 5, 7, 8, 6, 11, 4

Déterminer la médiane de cette série :

3, 4, 5, 6, 7, 8, 11, 12. Par convention.

Ici la médiane est $Me = \frac{6+7}{2} = 6,5$

La médiane d'une série statistique est la valeur du caractère qui correspond à l'effectif cumulé

$$n_i = \frac{N}{2}$$

NB :

N désignant les valeurs de la série statistique rangée dans l'ordre croissant, on a :

-Si N est impaire, la médiane correspond à la valeur de rang $\frac{N+1}{2}$

-Si N est paire, la médiane correspond à la valeur de rang $\frac{N}{2}$ et $\frac{N}{2} + 1$

Détermination de la médiane dans le cas d'un caractère continu.

Exemple :

Déterminer la médiane dans la distribution des tailles de classe de 12^{ème} dont les données sont dans le tableau en 3°).

La médiane Me correspond à l'effectif cumulé $\frac{N}{2} = 15$

Me $\in$ [168 ; 174[est la classe médiane.

$$x_i : 168 < M_e < 174$$

$$n_i : 9 < 15 < 23$$

$$\frac{M_e - 168}{174 - 168} = \frac{15 - 9}{23 - 9} \Rightarrow M_e = 170,57$$

c- Moyenne arithmétique d'une série :

Définition : La moyenne arithmétique d'une série est le quotient de la somme de ces valeurs par l'effectif total. Si $x_1, x_2, \dots, x_k$ sont les valeurs et $n_1, n_2, \dots, n_k$ sont les effectifs alors la moyenne est notée :

$$\bar{x} = \frac{\sum n_i x_i}{N} = \frac{1}{N} \sum_i^k n_i x_i$$

NB : Dans le cas d'un caractère continu, les x_i sont remplacés par les c_i

(centre de classe $[x_i, x_{i+1})$)

Exemple : Soit le tableau statistique donné par :

x_i	n_i	$n_i \times x_i$
7	1	7
8	3	24
9	5	45
10	4	40
11	2	22
12	2	24
13	4	52
14	1	14

15	2	30
16	0	0
17	1	17
Total	25	275

2- Les caractéristiques de dispersion :

a- Amplitude ou étendue d'une série :

L'amplitude d'une série est la différence entre la plus grande valeur et la plus petite valeur du caractère.

Exemple : Déterminer l'amplitude des séries suivantes :

- a) $-5 ; 4 ; 3 ; 1 ; 10 ; 12$ / Rep. : $(12) - (-5) = 12 + 5 = 17$
- b) $5 ; 10 ; 9 ; 20 ; 40 ; 30 ; 8$ / Rep : $(40) - (5) = 40 - 5 = 35$

b- Les quartiles

Les valeurs du caractère étant rangées dans l'ordre croissant ou décroissant, on appelle quartiles les 3 valeurs qui divisent la série en 4 groupes de même effectifs : on note : Q_1 ; Q_2 ; Q_3 ces 3 valeurs

Exemple : Déterminer les quantiles de la série suivante :

9 ; 12 ; 5 ; 11 ; 7 ; 15 ; 1 ; 2 ; 18 ; 20 ; 16 ; 19 ; 30 ; 40 ; 34

Rep : 1, 2; 5 ; **7** ; 9 ; 11 ; 12 ; **15** ; 16 ; 18 ; 19 ; **20** ; 30 ; 34 ; 40

Q_1

Q_2

Q_3

Déterminer les quantiles de la série suivante : ci-dessus. On a donc $Q_1 = 7$; $Q_2 = 15$; $Q_3 = 20$

Q_2 est la médiane, il est appelé quantile centrale

Q_1 et Q_3 sont les quantiles extrémités

NB : Q_1 est la valeur du caractère qui correspond à l'effectif cumulé $n_i = \frac{N}{4}$

Q_3 est la valeur du caractère qui correspond à l'effectif cumulé $n_i = \frac{3N}{4}$

$Q_3 - Q_1$ est l'écart interquartile

$\frac{Q_3 - Q_1}{Q_2}$ Est l'écart interquartile relatif.

c- Variance ou fluctuation :

La variance d'une série est la moyenne des carrés des écarts à la moyenne c'est-à-dire :

Si

$x_1, x_2, \dots, x_k$ sont les valeurs du caractère et $n_1 ; n_2 ; \dots ; n_k$ les effectifs.

La variance est

notée : $V = \frac{1}{N} \sum_i^k n_i x_i^2 - (\bar{x})^2$ ou encore on la note : $V = \frac{1}{N} \sum_i^k (n_i x_i - \bar{x})^2$

d- L'écart type :

C'est la racine carrée de la variance, on la note σ tel que $\sigma = \sqrt{V} = \sqrt{\frac{1}{N} \sum_i^k n_i x_i^2 - (\bar{x})^2}$

EXERCICE 1 :

Soit le tableau de la série statistique correspondant aux pointures des chaussures de 48 élèves.

x_i	n_i
36	5
37	11
38	12
39	8
40	4
41	3
42	4
43	1

- 1) Quelle est la population ?
- 2) le caractère étudié ?
- 3) Déterminer le mode
- 4) Déterminer la pointure moyenne

Solution

- 1- La population est un ensemble d'élèves.
- 2- Le caractère est : la pointure des chaussures.
- 3- Le mode est : 38

4- La moyenne des pointures : $\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i$

$$\Rightarrow \bar{x} = \frac{1849}{48} \Rightarrow \bar{x} = 38,52$$

x_i	n_i	$n_i x_i$	x_i^2	$n_i x_i^2$
36	5	180	1296	6480
37	11	407	1369	15059
38	12	456	1444	17328
39	8	312	1521	12168
40	4	160	1600	6400
41	3	123	1681	5043
42	4	168	1761	7056
43	1	43	1849	1849
Total		1849		71383

EXERCICE 2 :

A l'occasion d'un concours de pêche, les 250 poissons pêchés ont été mesurés et leurs tailles ont été regroupées dans le tableau suivant :

Classe des tailles (cm)	Effectifs	n_i	c_i	$n_i \cdot c_i$
[10, 20[	16	16	15	240
[20, 30[	26	42	25	1050
[30, 40[	36	78	35	2730
[40, 50[	52	130	45	5850
[50, 60[	38	168	55	9240
[60, 70[	30	198	65	12870
[70, 80[	20	218	75	16350
[80, 90[	12	230	85	19530
[90, 100[	10	240	95	22800
[100, 110[	6	246	105	25830
[110, 120[	4	250	115	28750

N.B : Les parties complétées sont :

n_i : $n_1 = 16$; $n_2 = 16 + 26 = 42$, $n_3 = 16 + 26 + 36 = 78$

c_i : $c_1 = \frac{10 + 20}{2} = 15$; $c_2 = \frac{20 + 30}{2} = 25$; $c_3 = \frac{30 + 40}{2} = 35$

$n_i \cdot c_i$: $n_1 \cdot c_1 = 16 \times 15 = 240$; $n_2 \cdot c_2 = 42 \times 25 = 1050$; $n_3 \cdot c_3 = 78 \times 35 = 2730$

Évaluation

- a- compléter le tableau :
- b- déterminer la population, le caractère et la nature du caractère
- c- déterminer les quartiles
- d- déterminer la moyenne et l'écart type.

Solution

a- Complétons le tableau (voir tableau)

b- La population : les poissons

- le caractère : la taille
- nature du caractère : caractère quantitatif continu

c- Déterminons les quartiles :

$$Q_1 n_i = \frac{N}{4} = \frac{250}{4} = 62,5 \quad \text{alors} \quad Q_1 \in [30, 40[\Leftrightarrow 42 < 62,5 < 78$$

$$\frac{Q_1 - 30}{40 - 30} = \frac{62,5 - 42}{78 - 42} \Leftrightarrow \frac{Q_1 - 30}{10} = 0,56 \Rightarrow Q_1 = 35,6 \quad \text{or le } n_i = \frac{N}{2} = \frac{250}{2} = 125 \Leftrightarrow 40 < Q_2 < 50$$

$$n_i = \frac{3N}{4} = \frac{3 \times 250}{4} = 187,5$$

$$Q_3 \in [60 - 70[$$

d- Déterminons la moyenne et l'écart type

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i \Rightarrow \bar{x} = \frac{1}{250} \times 13090 = 52,36$$

L'écart type :

$$\begin{aligned} - \text{Variance } V &= \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - (\bar{x})^2 = \frac{1}{250} \times (1063064) - (52,36)^2 \\ - \text{Ecart type : } \sigma &= \sqrt{V} = \sqrt{1510,6864} = 38,86 \end{aligned}$$

$$\sigma = 38,86$$

IV- Série statistique à deux caractères x et y

1- Nuage de points associé :

Le plan étant muni d'un repère orthogonal, on appelle nuage de points associé, l'ensemble des points M_{ij} de coordonnées (x_i, y_i) .

2- Point moyen d'un nuage de points :

On appelle point moyen d'un nuage de points représentant une série, le point G de coordonnées $(X_G ; Y_G)$ où

$$\begin{cases} X_G = \bar{X} = \frac{\sum x_i}{N} \\ \text{et} \\ Y_G = \bar{Y} = \frac{\sum y_i}{N} \end{cases}$$

3- Covariance :

La covariance d'une série statistique à deux caractères x et y, de moyennes respectives $\bar{x}$ et $\bar{y}$ et d'effectif total N est le nombre réel noté Cov (X, Y) ou $\sigma_{x,y}$ tel que :

$$\text{Cov (x,y)} = \frac{\sum n_{ij}(x_i - \bar{X})(y_i - \bar{Y})}{\text{Effectif total}}$$

ou

$$\text{Cov (x,y)} = \frac{\sum n_{ij}x_i y_i}{\text{Effectif total}} - \bar{X} \cdot \bar{Y}$$

4- Droite de régression de y en x :

La droite (D) de régression de y en fonction de x passe par le point G $(\bar{X}, \bar{Y})$, point moyen du nuage de point son équation est $y = ax + b$ avec :

$$a = \frac{\text{Cov}(X,Y)}{V(X)} \text{ et } b = \bar{Y} - a\bar{X}$$

où V(X) est la variance et a pour formule : $V(X) = \frac{\sum x_i^2}{N} - \bar{X}^2$

5- Droite de régression de X en Y :

La droite (D') de régression de X en fonction de Y passe par le point $G(\bar{X} ; \bar{Y})$, point moyen du nuage de point. Son équation est $x = a'y + b'$ avec :

$$a' = \frac{\text{Cov}(X,Y)}{V(Y)} \text{ et } b' = \bar{X} - a'\bar{Y}$$

où $V(Y)$ est la variance et a pour formule : $V(Y) = \frac{\sum Y_i^2}{N} - \bar{Y}^2$

Remarque :

$G(\bar{X} ; \bar{Y})$ appartient aux deux droites de régression pour tracer ces droites, il suffit donc de déterminer un autre point vérifiant l'équation.

6- **Corrélation linéaire**

Deux variables statistiques X et Y sont dites en corrélation linéaire lorsque la courbe de régression de Y en X et la courbe de régression de X en y sont des droites.

7- **Coefficient de corrélation linéaire :**

➤ **Définition :**

On appelle coefficient de corrélation linéaire d'une série statistique double de caractère (X ; Y) , le

nombre réel **r** définie par $\mathbf{r} = \frac{\text{Cov}(X;Y)}{\sqrt{V(X).V(Y)}}$

➤ **Propriétés de r :**

P₁ : Si $0,87 \leq |r| \leq 1$, on considère que la corrélation linéaire entre les deux variables est forte.

P₂ : si $r = 1$, on dit que la corrélation est parfaite.

8- **Estimation et prévision**

Lorsqu'il existe une bonne corrélation linéaire, il est possible de prévoir la valeur de y connaissant celle de x, en utilisant l'équation de la droite de régression de y en x ; et vice – versa.

Exercices d'application :

Exercice 1 :

Un pharmacien observe, durant les 6 premiers mois de l'ouverture de son officine, le chiffre d'affaire en millions de francs CFA. Le résultat de l'observation est résumé dans le tableau suivant où x désigne le numéro du mois et y le chiffre d'affaires – correspondant.

X	1	2	3	4	5	6
Y	12	13	15	19	21	22

- 1- Calculer $\bar{X}$ et $\bar{Y}$ les moyennes des variables X et Y.
- 2- Représenter graphiquement le nuage de points de cette série statistique double ainsi que le point G. on prendra : (Unités : 2 cm en abscisses et 1 cm pour 2 unités en ordonnées).
- 3- Calculer la variance $V(X)$ de x et la covariance $Cov(X, Y)$ de x et y.

N.B : Les résultats seront donnés sous forme de fractions irréductibles.

- 4- Démontrer qu'une équation de la droite de régression (D) de y en x est : $y = \frac{78}{35}x + 9,2$
- 5- Tracer la droite (D).
- 6- En utilisant la droite (D), calculer une estimation du chiffre d'affaires de cette pharmacie à la fin du 7^{ème} mois.

Exercice 2 :

En prévision du lancement d'un nouveau produit, une société a effectué une enquête auprès de clients éventuels pour fixer le prix de vente (en milliers de francs) de ce produit. Les résultats sont donnés dans le tableau ci – dessous.

Prix de vente en milliers de francs xi	9	10	11	12	14	15	16	17
--	---	----	----	----	----	----	----	----

Nombre d'acheteurs potentiels y_i	180	160	150	130	100	90	80	70
-------------------------------------	-----	-----	-----	-----	-----	----	----	----

1- Représenter graphiquement le nuage de points de cette série statistique. On prendra pour unité :

1 cm pour 1000 francs en abscisse

1 cm pour 10 acheteurs en ordonnée.

2- Déterminer les coordonnées du point G

3- Déterminer le coefficient de corrélation linéaire r puis interpréter le résultat obtenu.

4- Déterminer une équation de la droite (D) de régression de y en x . Tracer (D)

5- Estimer graphiquement le prix maximum à fixer pour qu'il y ait au moins 50 acheteurs potentiels.

6- En utilisant l'équation de la droite (D) de régression de y en x , puis déterminer :

a) Le nombre d'acheteurs à prévoir si le prix est fixé à 13 000 F

b) Le prix à fixer pour que le nombre d'acheteurs potentiels soit supérieur à 250.

Solution

Exercice 1

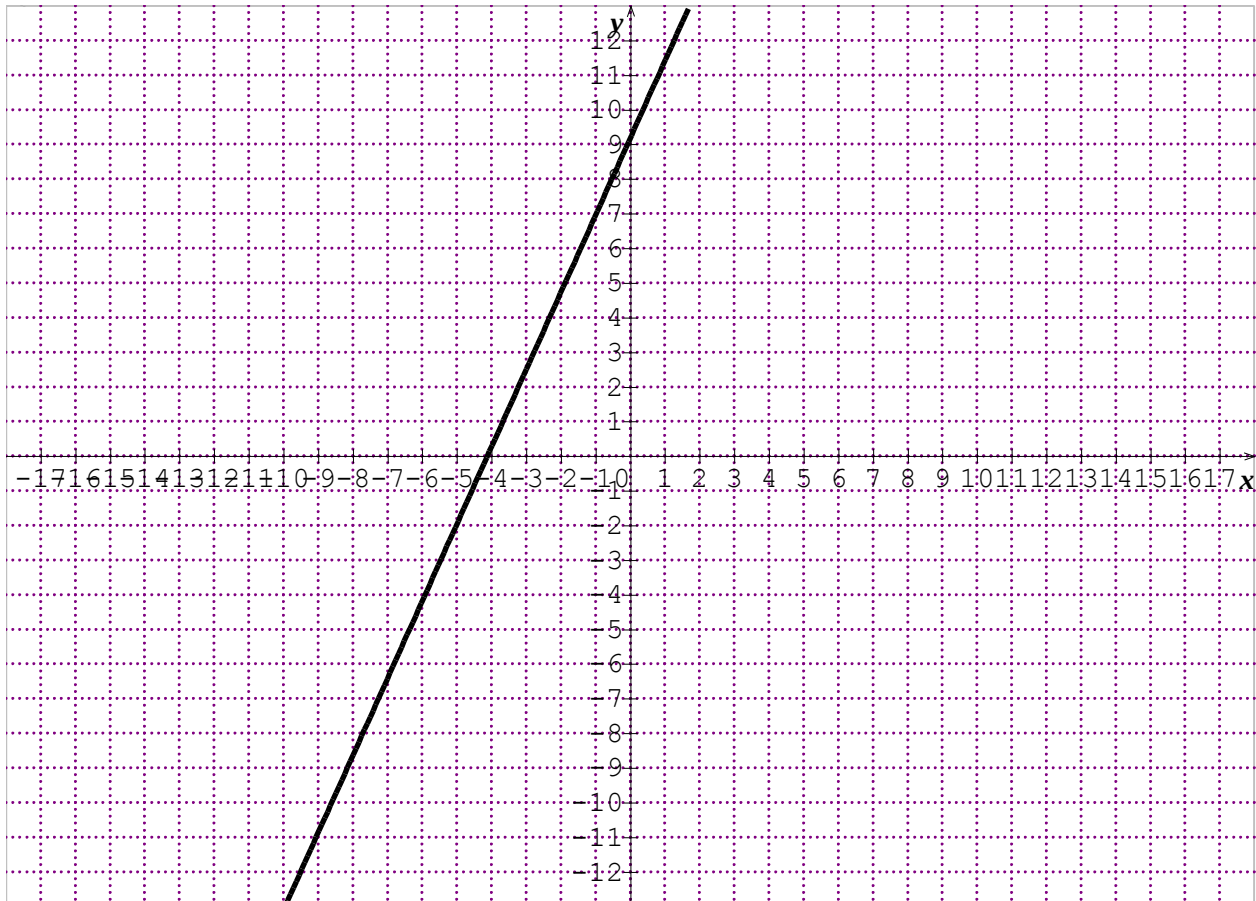
Pour le calcul, on pourra s'aider du tableau ci-dessous :

X_i	Y_i	X_i, y_i	X_i^2	Y_i^2	
1	12	12	1	144	
2	13	26	4	169	
3	15	45	9	225	
4	19	76	16	361	
5	21	105	25	441	
6	22	132	36	484	
21	102	396	91	2 331	Total

1- Calcul de $\bar{X}$ et $\bar{Y}$ les moyennes des variables x et y.

$$\bar{X} = \frac{\sum_{i=1}^6 x_i}{6} = \frac{21}{6} = \frac{7}{2} = 3,5 ; \bar{Y} = \frac{\sum_{i=1}^6 y_i}{6} = \frac{102}{6} = 17$$

2- Le nuage de points de la série statistique ainsi que le point moyen G.



1- Calcul de la variante $V(x)$ de x et la covariance $\text{Cov}(x, y)$ de x et y.

$$V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{70}{24} = \frac{35}{12} ; \text{cov}(X; Y) = \frac{396}{6} - \left(\frac{7}{2} \times 17\right) = \frac{39}{6} = \frac{13}{2}$$

2- Montons qu'une équation de la droite de régression (D) de y en x est : $y = \frac{78}{35}x + 9,2$

$$a = \frac{\text{cov}(X,Y)}{V(X)} = \frac{\frac{13}{2}}{\frac{35}{12}} = \frac{13}{2} \times \frac{12}{35} = \frac{156}{70} = \frac{78}{35} \text{ et } b = \bar{Y} - a\bar{X} = 17 - \frac{78}{35} \times \frac{7}{2} = 9,2$$

donc l'équation de la droite de régression est : $y = \frac{78}{35}x + 9,2$

3- Tracé de la droite (D). (Voir nuage de points)

La droite (D) passe par le point moyen point G (3,5 ; 17) et par un autre point que l'on détermine en remplaçant dans l'équation de la droite de régression, x par une valeur que l'on choisira.

Par exemple, si $x = 0$, on obtient $y = 9,2$. (D) passe donc également par le point A(0 ; 9,2).

4- Calcul d'une estimation du chiffre d'affaires de cette pharmacie à la fin du 7^{ème} mois. Le 7^{ème} mois correspond à $x = 7$. Il s'agit de déterminer la valeur de y associée à cette valeur de x.

Pour $x = 7$, on a : $y = \frac{78}{35} \times 7 + 9,2 = 24,8$ millions de francs CFA.

3°) Déterminer le coefficient de corrélation linéaire r. interpréter le résultat obtenu.

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{\sum X_i \cdot Y_i}{n} - \bar{X} \cdot \bar{Y} \\ &= \frac{9 \times 180 + 10 \times 160 + 11 \times 150 + 12 \times 130 + 14 \times 100 + 15 \times 90 + 16 \times 80 + 17 \times 70}{8} - 13 \times 120 \end{aligned}$$

$$\text{Cov}(X, Y) = -103,75$$

$$r = \frac{\text{cov}(X,Y)}{\sqrt{V(X) \cdot V(Y)}}$$

$$V(X) = \frac{\sum X_i^2}{N} - \bar{X}^2 = \frac{9^2 + 10^2 + 11^2 + 12^2 + 14^2 + 15^2 + 16^2 + 17^2}{8} - 13^2$$

$$V(X) = 7,499$$

$$V(Y) = \frac{\sum Y_i^2}{N} - \bar{Y}^2 = \frac{180^2 + 160^2 + 150^2 + 130^2 + 100^2 + 90^2 + 80^2 + 70^2}{8} - 120^2$$

$$V(X) = 1450,003$$

$$r = \frac{\text{cov}(X,Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{-103,75}{\sqrt{7,499 \times 1450,003}} = -0,995$$

$0,87 \leq |r| \leq 1$: il existe alors une forte corrélation linéaire entre les variables X et Y. L'on peut donc faire un ajustement linéaire entre X et Y.

4°) Equation de la droite (D) de régression de y en x.

$$(D) : y = ax + b$$

$$a = \frac{\text{cov}(X,Y)}{V(X)} = \frac{-103,75}{7,499} = -13,88$$

$$\bar{Y} = a\bar{x} + b \Rightarrow b = \bar{y} - a\bar{x} = 120 - (-13,88) \times 13 = 199,83$$

$$(D) : y = -13,88x + 199,83$$

Tracé de (D) : (D) passe par le point moyen G(13,120), déterminons un second point en remplaçant par exemple x par 10. On trouve y = 161,03. On obtient : A (10 ; 161,03)

5°) Pour qu'il y ait au moins 50 acheteurs potentiels, on lit graphiquement que le prix à fixer ne doit pas dépasser 18 000 francs.

6°) a°) Le nombre d'acheteurs à prévoir, si le prix est fixé à 13 000 F, correspond à x = 13

$$y = -13,88 \times 13 + 199,83 = 119,39 \simeq 119$$

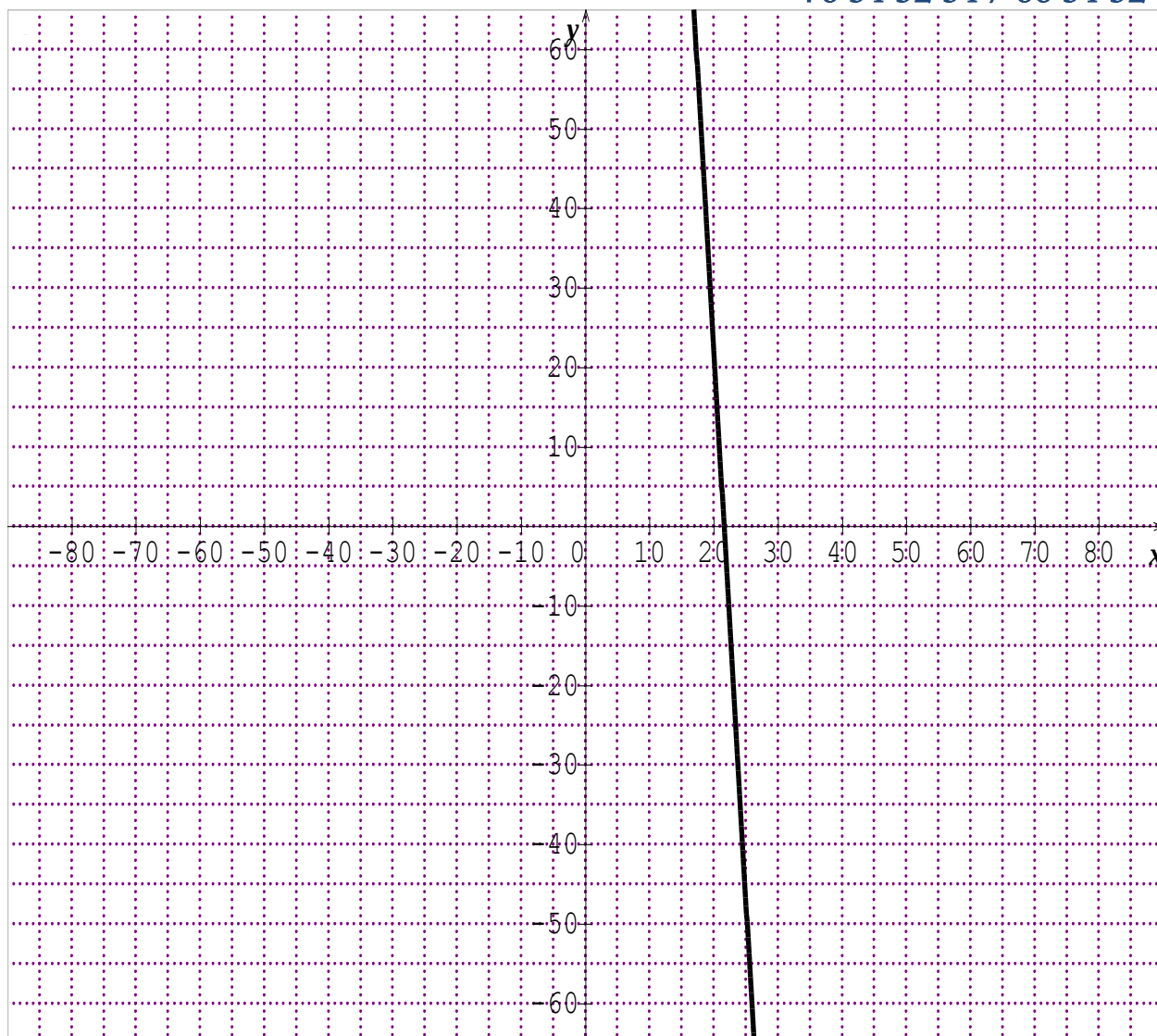
b°) Il s'agit de déterminer x pour $y = 250$.

$$y = -13,88x + 299,83 \Rightarrow x = \frac{y - 299,83}{-13,88}$$

$$\text{Pour } y = 250, \text{ on a : } x = \frac{250 - 299,83}{-13,88} = 3,59 \text{ milliers de francs} = 3590 \text{ F}$$

Exercice 2 :

- 1- Représentation graphique du nuage de points de cette série statistique (unité : 1 cm pour 1000 francs en abscisse, 1 cm pour 10 acheteurs en ordonnée).



2- Déterminons les coordonnées du point G.

$$\bar{X} = \frac{\sum_{i=1}^8 x_i}{8} = \frac{9+10+11+12+14+15+16+17}{8} = 13$$

$$\bar{Y} = \frac{\sum_{i=1}^8 y_i}{8} = \frac{180+160+150+130+100+90+80+70}{8} = 120$$

$\Rightarrow G(13; 120)$

3-Détermination du coefficient de corrélation linéaire r .

$$\text{Cov}(X, Y) = \text{Cov} = \frac{\sum X_i \cdot Y_i}{N} - \bar{X} \cdot \bar{Y}$$

$$\Rightarrow \text{Cov}(X, Y) = \frac{9 \cdot 180 + 10 \cdot 160 + 11 \cdot 150 + 12 \cdot 130 + 14 \cdot 100 + 15 \cdot 90 + 16 \cdot 80 + 17 \cdot 70}{8} - 13 \cdot 120$$

$$\text{Cov}(X, Y) = -103,75$$

$$r = \frac{\text{cov}(X, Y)}{\sqrt{V(X) \cdot V(Y)}}$$

$$V(X) = \frac{\sum X_i^2}{N} - \bar{X}^2 = \frac{9^2 + 10^2 + 11^2 + 12^2 + 14^2 + 15^2 + 16^2 + 17^2}{8} - 13^2$$

$$V(X) = 7,499$$

$$V(Y) = \frac{\sum Y_i^2}{N} - \bar{Y}^2 = \frac{180^2 + 160^2 + 150^2 + 130^2 + 100^2 + 90^2 + 80^2 + 70^2}{8} - 120^2$$

$$V(Y) = 1450,003$$

$$r = \frac{\text{cov}(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{-103,75}{\sqrt{7,499 \cdot 1450,003}} = -0,995$$

$0,87 \leq |r| \leq 1$: il existe alors une forte corrélation linéaire entre les variations X et Y. linéaire

L'on peut donc faire un ajustement linéaire entre X et Y.

4-Equation de la droite (D) de régression de Y en X

(D) : $y = ax + b$

$$A = \frac{\text{cov}(X,Y)}{V(X)} = \frac{-103,75}{7,499} = -13,88$$

Alors on a $\bar{y} = a\bar{x} + b \Rightarrow b = \bar{y} - a\bar{x} = 120(-13,88) \times 13 = 299,83$

(D) : $y = -13,88x + 299,83$

Tracé de (D) : (D) passe par le point moyen G (13 ; 120). Déterminons un second en remplaçant par exemple x par 10. On trouve y = 161,03. On obtient : A (10 ; 161,03)

5- Pour qu'il ait au moins 50 acheteurs potentiels, on lit graphiquement que le prix à fixer ne doit pas dépasser 18000 francs.

6- a) le nombre d'acheteurs à prévoir, si le prix est fixé à 13000 F, correspond à $x = 13$.

$$Y = -13,88 \times 13 + 299,83 = 119,39$$

b) il s'agit de déterminer x pour y = 250

$$Y = -13,88x + 299,83 \text{ alors } x = \frac{y - 299,83}{-13,88}$$

$$\text{Pour } y = 240, \text{ on a : } x = \frac{250 - 299,83}{-13,88} = 3,59 \text{ milliers de franc} = 3590 \text{ F}$$