

Statistique

OBJECTIFS :

Ce thème vise à :

- Etablir une statistique à savoir à collecter des informations sur des individus d'une population puis de les résumés dans un tableau appelé tableau statistique ;
- Représente des données statistiques dans un repère ;
- Elargir le champ des statistiques étudiées les caractéristiques de position.
- Utiliser les séries statistiques à double variables pour étudier le nuage des points.

Commentaires

Andreï Andreïevitch Markov (2 Juin 1856 – 20 Juillet 1922) Mathématicien Russe fut l'un des piliers des premières notions de la statistique.

Consolider les connaissances acquises en première sur le dénombrement des ensembles finis à l'occasion d'exercices de probabilité.

Les calculs statistiques sur une population n'est pas un chapitre nouveau pour les élèves. En classe de terminale, on approfondira cette notion par la statistique à doubles variables. On fera le lien entre les fréquences en statistique et la probabilité de l'événement correspondant.

La notion de statistique à doubles variables permet d'étudier deux caractères numériques d'une population étudiée.

Les ajustements linéaires sont des notions très importantes qui doivent être maîtrisée par les élèves.

Volume horaire : 16 heures

SAVOIRS	SAVOIR-FAIRE
• Vocabulaire de la statistique • Définition des caractéristiques de position • Définition des caractéristiques de dispersion. • Utiliser les ajustements linéaires. • Nuages de points pour une série statistique double • Régression linéaire et corrélation	• Définir : - population et individu, - Caractères et modalités, - Echantillon • Calcule le mode ou dominante, la classe modale ou dominante, la médiane quartiles et moyenne. • Calcule l'étendue, interquartile, variance et écart-type. • Représente le nuage des points, Détermine les coordonnées du point moyen. • Détermine les droite d'ajustement par la : - Méthode des moindres carrés - Méthode de Mayer

Remarques et suggestions

On introduira le vocabulaire de la statistique au travers de situations concrètes. Les ajustements d'une série pondérée et par changement de variable probabilités conditionnelles sont hors programme, ce qui n'empêche pas de les définir.

On se placera dans des situations ayant du sens ; en particulier on présentera des applications de la statistique notamment dans le domaine de la biologie et en économie.

I) Définition et langages de la statistique

1- Définition :

Définition : La statistique est la branche de la mathématique qui consiste à regrouper les données ; à les analyser et à tirer des conclusions. Elle peut être définie également comme un ensemble de données chiffrées ou non chiffrées.

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

2- Langages statistiques :

a- Notion de population et d'individu:

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

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

b- Caractères et modalités :

Dans une étude statistique, 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 (discontinu) ou (continu)

- Un caractère quantitatif est dit **discontinu** si ses valeurs prises sont isolées.

Exemple : le nombre de frère propre à un individu, le nombre de femme, etc.....

- 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, etc....

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 il est nécessaire dans une première phase de les grouper, de les présenter sous forme de tableau et de les Représente graphiquement.

1- Cas d'un caractère qualitatif :

a- Tableau statistique :

Activité

Une étude faite sur la situation matrimoniale de 10 personnes a donnée le résultat suivant : Marié ; Divorcé ; Célibataire ; Marié ; Veuf ; Célibataire ; Divorcé ; Célibataire ; Célibataire ; Marié.

Dresse un tableau justifiant les résultats donnés ci-dessus.

Synthèse

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

Les valeurs du caractère étant rangées dans l'ordre croissant, on a :

- La 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

$N = \sum n_i$ 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 (x_i)	Nombre d'éléments (n_i)	Fréquences (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	$10/10 = 1$

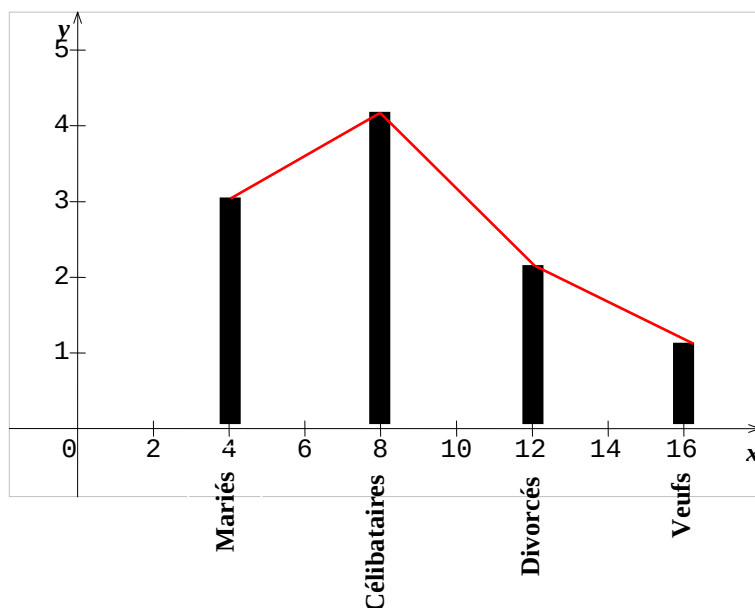
NB : la somme des fréquences est toujours égale à 1 , c'est-à-dire : $\sum f_i = 1$

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



NB : Si les fréquences f_i étaient en ordonnée, le diagramme 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 $B_i(0 ; n_i)$

Diagramme en secteur circulaire :

Soit le demi – cercle (C). Représentons la série précédente sur (C). Chaque modalité est représentée par un secteur angulaire α_i proportionnel à sa fréquence f_i tel que $\alpha_i = f_i \times 180^\circ$.

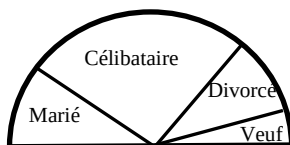
Mariés : $\alpha_1 = 0,3 \times 180^\circ \Rightarrow \alpha_1 = 54^\circ$

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

Divorcés : $\alpha_3 = 0,2 \times 180^\circ \Rightarrow \alpha_3 = 36^\circ$

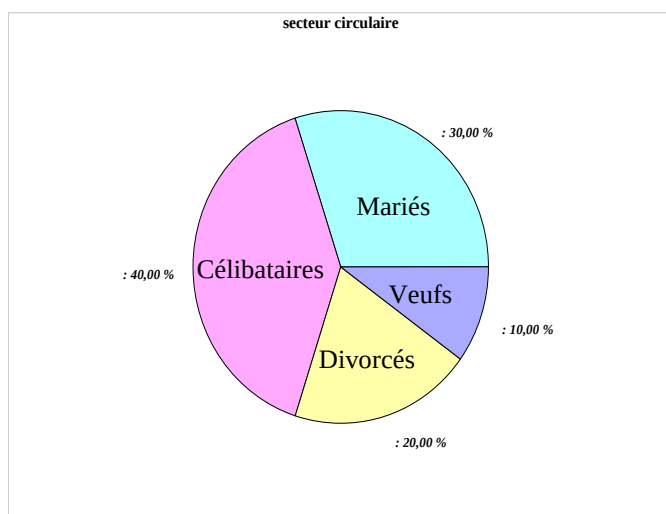
Veufs : $\alpha_4 = 0,1 \times 180^\circ \Rightarrow \alpha_4 = 18^\circ$

Modalités (x_i)	Nombre d'éléments (n_i)	Fréquences (f_i)	Angles (α_i)
Marié	3	$3/10 = 0,3$	54°
Célibataire	4	$4/10 = 0,4$	72°
Divorcé	2	$2/10 = 0,2$	36°
Veuf	1	$1/10 = 0,1$	18°
Total	10	$10/10 = 1$	180°



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

$$\alpha_1 = 108^\circ ; \quad \alpha_2 = 144^\circ ; \quad \alpha_3 = 72^\circ ; \quad \alpha_4 = 36^\circ$$



2- Cas d'un caractère quantitatif discret :

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

Activité :

Une étude statistique faite sur le nombre de frère d'un groupe d'individu a donné les résultats suivants : 4 ; 2 ; 2 ; 1 ; 3 ; 2 ; 3 ; 3 ; 2 ; 1.

Cosigne :

- Dresse un tableau statistique correspondant à cette étude.
- Complète le tableau en y insérant la colonne des Effectifs cumulés croissants et celui des effectifs cumulés décroissants.

Synthèse

- Dressons un tableau statistique correspondant à cette étude.

Modalités x_i	Effectif n_i
1	2
2	4
3	3
4	1
Total	10

- Complétons- le tableau en y insérant la colonne des effectifs cumulés croissants et celui des effectifs cumulés décroissants.

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 la colonne des effectifs cumulés croissants et des effectifs cumulés décroissants ainsi que les fréquences cumulées croissantes et des fréquences cumulées décroissantes. 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 :

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

$$n_1 = n_1 = 2$$

$$n_2 = n_1 + n_2 = 2 + 4 = 6$$

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

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

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 : $f_i = \frac{n_i}{N}$

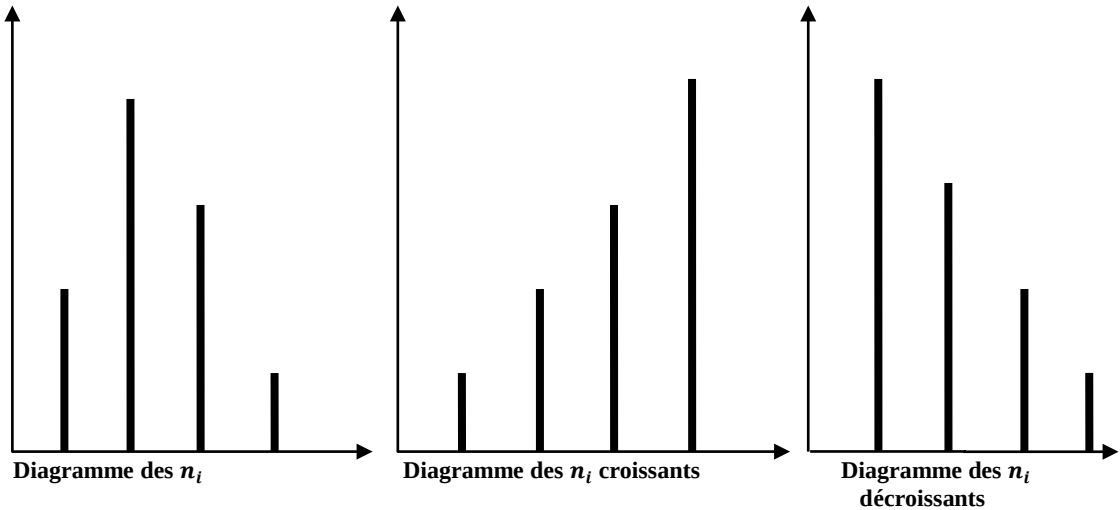
D'où le tableau suivant :

Modalités x_i	Effectif n_i	Effectif cumulé croissant ECC	Effectif cumulé décroissant ECD
1	2	2	10
2	4	6	8
3	3	9	4
4	1	10	1
Total	10	27	23

a- 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 Construis 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 $[x_i ; x_{i+1}[$ ou encore appelées des classes.

Exemple :

Une étude faite sur la taille de 30 personnes 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

Dresse le tableau des effectifs et des fréquences cumulées.

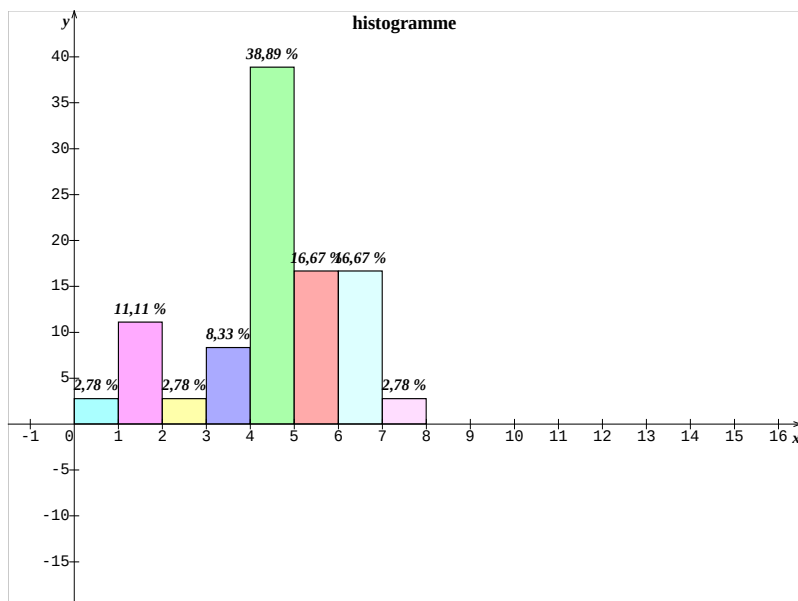
On prendra des classes d'amplitude 6.

Classes $[x_i ; x_{i+1}[$	Effectif n_i	Centre des classes C_i	Effectif cumulé croissant ECC	Fréquences cumulées f_i
[144 ; 150]	1	147		
[150 ; 156]	4	153	30	0,633
[156 ; 162]	1	159	29	0,166
[162 ; 168]	3	165	25	0,2
[168 ; 174]	14	171	24	0,3
[174 ; 180]	6	177	21	0,76
[180 ; 186]	1	183	7	0,966
			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 conviendrons de prendre la valeur c_i commune à tous les individus de la classe (x_i, x_{i+1})

Histogramme des effectifs (représentation graphique)



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

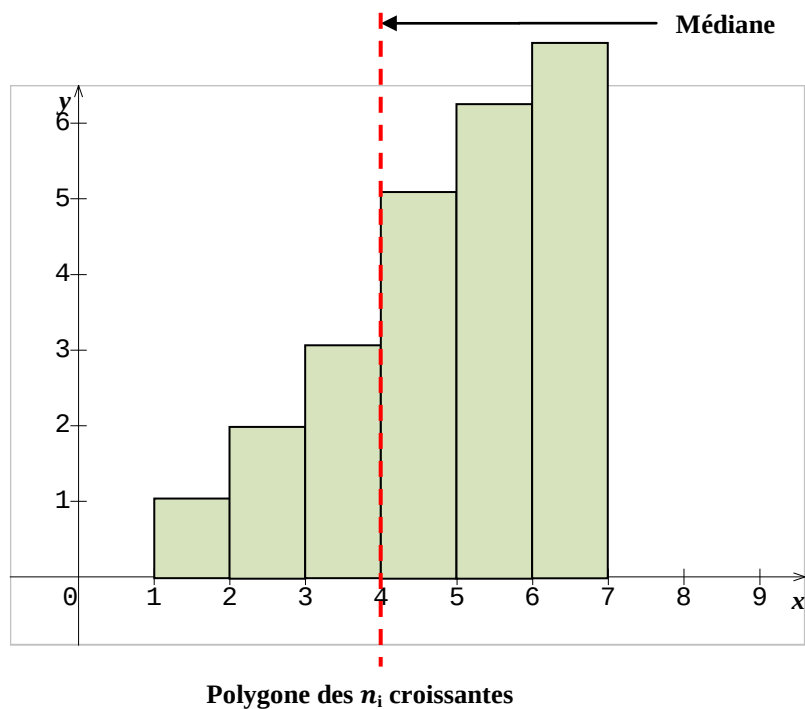
Les polygones :

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 bisée 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- Éléments caractéristiques d'une série statistique:

1- Les caractéristiques de position :

a- Le mode ou dominante :

Le mode d'une série statistique est la valeur du caractère qui a le plus grand effectif (ou fréquence). On le note M_0 .

Exemple : Détermine le mode de la série suivante : 10 ; 7 ; 8 ; 5 ; 2 ; 7 ; 4 ; 4 ; 5 ; 7 ; 10

Réponse : le mode est 7 (car 7 compte le plus grand effectif parmi les autres chiffres énumérés)

b- Classe modale

Pour une série continue, lorsque les classes sont de même largeur, la classe modale est la classe ayant l'effectif le plus important.

Exemple :

Voici le tableau des effectifs des notes des élèves lors d'un examen :

Notes	[0 ; 4[	[4 ; 8[	[8 ; 12[	[12 ; 16[	[16 ; 20[
Effectifs	5	32	61	80	15

Détermine la classe modale de cette série statistique.

Solution

La classe modale de cette série statistique est la classe ayant le plus grand effectif c'est-à-dire : [12 ; 16[.

c- 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

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

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

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

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 $n_1, n_2, \dots, n_k$ les effectifs alors on a :

La moyenne est notée $\bar{x}$ tel que : $\bar{x} = \frac{\sum n_i \times x_i}{N}$ ou $\bar{x} = \frac{\sum n_i}{N}$

NB : Dans le cas d'un caractère continu, les x_i sont remplacés par les centres C_i de classe $[x_i ; x_{i+1}]$ tel que : $C_i = \frac{x_i + x_{i+1}}{2}$

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.

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 par : $Q_1 ; Q_2 ; Q_3$ ces 3 valeurs.

NB : Q_1 : est la valeur du caractère qui correspond au rang $\frac{N}{4}$

- Q_3 : est la valeur du caractère qui correspond au rang $\frac{3N}{4}$

- $Q_3 - Q_1$: 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 mesure de la dispersion des échantillons autour de la moyenne.

On la note $V(X)$ tel que si $x_1, x_2, \dots, x_k$ sont les valeurs du caractère et $n_1 ; n_2 ; \dots, n_k$ les effectifs

alors on a : $V(X) = \frac{\sum n_i \times x_i^2}{N} - \bar{x}^2$

d- L'écart type :

C'est la racine carrée de la variance, on la note : $\sigma = \sqrt{V(X)}$

IV- Série statistique à double 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 $(x_i ; y_j)$.

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

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

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 :

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

4- Détermination des droites de régression

Méthode des moindres carrées ordinaire:

a- 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 : $y(x) = ax + b$ avec : $a = \frac{Cov(X,Y)}{V(X)}$ et $b = \bar{Y} - a\bar{X}$.

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

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

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

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

Ainsi différentes méthodes sont prévues :

Méthode de Mayer:

Cette méthode consiste à :

- Attribuer à chaque année du tableau, un rang compté à partir de 1.
- Partager le nuage en deux nuages N_1 et N_2 dont les effectifs sont égaux.
- Détermine les points moyens G_1 et G_2 de chacun des deux nuages N_1 et N_2 .
- Ajuster le nuage par la droite ($G_1; G_2$) dont on Déterminea une équation.

Inconvénient de la méthode de Mayer :

La droite de régression obtenue à partir de la méthode de **Mayer** peut être différente de celle obtenue à partir de la méthode des **Moindres carrés ordinaires** et avoir des écarts non négligeables sur les prévisions.

Activité

Soit le tableau statistique suivant, portant sur l'évolution des ventes de médicaments des six dernières années d'une pharmacie :

Années	2005	2006	2007	2008	2009	2010
Ventes y_i	100	130	150	175	210	240

Détermine les prévisions de vente de cette pharmacie en 2011 puis en 2012 en utilisant :
La méthode des **Moindres carrés ordinaire** et la méthode de **Mayer**

Synthèse**Méthode des moindres carrés ordinaire.**- **Première étape :**

On attribue à chaque année du tableau, un rang compté à partir de 1 puis on complète le tableau en ajoutant la colonne des $x_i \cdot y_i$ et celui des x_i^2 :

							Total
Années	2005	2006	2007	2008	2009	2010	
Rang de l'année x_i	1	2	3	4	5	6	21
Ventes y_i	100	130	150	175	210	240	1005
$x_i \cdot y_i$	100	260	450	700	1050	1440	4000
x_i^2	1	4	9	16	25	36	91

- **Deuxième étape :**

On calcule les coordonnées $\bar{x}$ et $\bar{y}$ du point moyen G tel que :

$$\bar{x} = \frac{\sum x_i}{n} = \frac{21}{6} = 3,5 \quad \text{et} \quad \bar{y} = \frac{\sum y_i}{n} = \frac{1005}{6} = 167,5$$

- **Troisième étape :**

On détermine la droite de régression (D) de y en x en calculant les réels a et b tel que :

$$a = \frac{\sum x_i \cdot y_i - n\bar{x} \cdot \bar{y}}{\sum x_i^2 - n(\bar{x})^2} = \frac{4000 - 6 \times 3,5 \times 167,5}{91 - 6 \times (3,5)^2} = 27,57 \quad \text{et} \quad b = \bar{y} - a\bar{x} = 167,5 - 27,57 \times 3,5$$

$$\Rightarrow b = 71,005 \approx 71$$

Alors on a (D): $y(x) = ax + b \Rightarrow y(x) = 27,57x + 71$

- **Quatrième étape :**

On calcule les prévisions de vente de cette pharmacie en 2011 et en 2012.

• **Prévisions de vente de cette pharmacie en 2011**

L'année 2011 correspond au rang $x = 7 \Rightarrow y(7) = 27,57 \times 7 + 71 = 263,99 \approx 264$

• **Prévisions de vente de cette pharmacie en 2012**

L'année 2012 correspond au rang $x = 8 \Rightarrow y(8) = 27,57 \times 8 + 71 = 291,56 \approx 292$

Méthode de Mayer

- **Première étape :**

On attribue à chaque année du tableau, un rang compté à partir de 1 puis on partage le nuage en deux nuages N_1 et N_2 dont les effectifs sont égaux.

	Groupe 1			Groupe 2		
Années	2005	2006	2007	2008	2009	2010
Rang de l'année x_i	1	2	3	4	5	6
Ventes y_i	100	130	150	175	210	240

Méthode de partage des points en deux groupes.

Nombre de point x_i	Premier groupe	Deuxième groupe
3 points : $x_1 ; x_2 ; x_3$	$x_1 ; x_2$	x_3
4 points : $x_1 ; x_2 ; x_3 ; x_4$	$x_1 ; x_2$	$x_3 ; x_4$
5 points : $x_1 ; x_2 ; x_3 ; x_4 ; x_5$	$x_1 ; x_2 ; x_3$	$x_4 ; x_5$
6 points : $x_1 ; x_2 ; x_3 ; x_4 ; x_5 ; x_6$	$x_1 ; x_2 ; x_3$	$x_4 ; x_5 ; x_6$

NB : les périodes peuvent être en année ; en mois ; en semaine ou en jour

- **Deuxième étape :**

On calcule un point moyen pour chacun des deux groupes :

Premier groupe

Soit $\bar{y} = a\bar{x} + b$ la droite de tendance relative au premier groupe.

Calculons les coordonnées du point moyen $G_1 \left(\bar{x} ; \bar{y} \right)$ relative premier groupe

Rang de l'année x_i	1	2	3
Ventes y_i	100	130	150

$$\bar{y} = \frac{\sum y_i}{n} = \frac{100+130+150}{3} = \frac{380}{3} = 126,66 \quad \text{et} \quad \bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

Alors le point moyen pour le premier groupe est : $G_1 \left(\begin{smallmatrix} 2 \\ 126,66 \end{smallmatrix} \right)$

Or la droite de tendance relative au premier groupe est : $\bar{y} = a\bar{x} + b \Leftrightarrow 126,66 = 2a + b \Leftrightarrow 2a + b = 126,66$ (1)

Deuxième groupe

Soit $\bar{y} = a\bar{x} + b$ la droite de tendance relative au premier groupe.

Calculons les coordonnées du point moyen $G_2 \left(\begin{smallmatrix} \bar{x} \\ \bar{y} \end{smallmatrix} \right)$ relative premier groupe

Rang de l'année x_i	4	5	6
Ventes y_i	175	210	240

$$\bar{y} = \frac{\sum y_i}{n} = \frac{175+210+240}{3} = \frac{625}{3} = 208,33$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{4+5+6}{3} = \frac{15}{3} = 5$$

Alors le point moyen pour le deuxième groupe est : $G_2 \left(\begin{smallmatrix} 5 \\ 208,33 \end{smallmatrix} \right)$

Or la droite de tendance relative au premier groupe est : $\bar{y} = a\bar{x} + b \Leftrightarrow 208,33 = 5a + b$

$$\Leftrightarrow 5a + b = 208,33 \text{ (2)}$$

- Troisième étape :

On résout le système formé par les équations (1) et (2) pour Détermine les réels a et b :

$$\begin{cases} 2a + b = 126,66 \\ 5a + b = 208,33 \end{cases}$$

La résolution de ce système donne : $a = 27,22$ et $b = 72,23$

Or la droite de tendance de la série statistique est: $y(x) = ax + b \Leftrightarrow y(x) = 27,22x + 72,23$

- Quatrième et dernière étape :

On calcule les prévisions de vente de cette pharmacie en 2011 et en 2012.

• **Prévisions de vente de cette pharmacie en 2011**

L'année 2011 correspond au rang $x = 7$

$$\Rightarrow y(7) = 27,22 \times 7 + 72,23 = 262,77 \approx 263$$

- **Prévisions de vente de cette pharmacie en 2012**

L'année 2012 correspond au rang $x = 8$

$$\Rightarrow y(8) = 27,22 \times 8 + 72,23 = 289,99 \approx 290.$$

c- 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.

d- Coefficient de corrélation linéaire :

Définition :

On appelle coefficient de corrélation linéaire d'une série statistique double caractère (X ; Y), le nombre réel r définie par $r = \frac{Cov(X;Y)}{\sqrt{V(X).V(Y)}}$

Propriétés de r :

P_1 : Si $0,87 \leq |r| \leq 1$, on considère que la corrélation linéaire entre les deux variables est forte.

P_2 : si $r = 1$, on dit que la corrélation est parfaite.

e- 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

Série statistiques simples

1 Voici les notes de quatre groupes d'élèves au baccalauréat malien.

1) Remplis les cases des différents indicateurs ci-dessous :

	Notes Groupe 1	Notes Groupe 2	Notes Groupe 3	Notes Groupe 4
	5 – 6 – 10 – 10 – 11 – 12 – 12 – 14	6 – 8 – 8 – 8 – 10 11 – 14 – 15	8 – 8,5 – 8,5 – 9 11 – 11 – 12 – 12	6 – 6 – 7 – 8 – 10 11 – 11 – 15
Moyenne				
Etendue				
Médiane				

2) Compare d'un point de vue qualitatif à la lueur des indicateurs calculés précédemment :

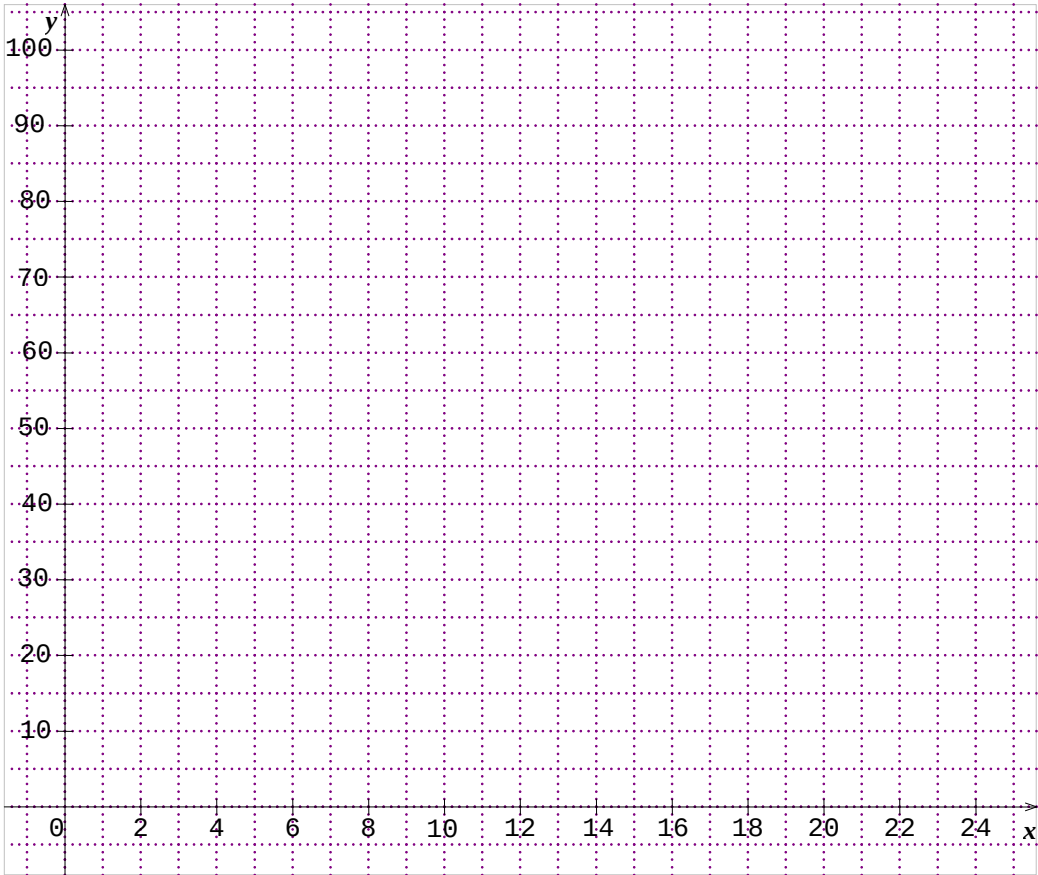
a- Le groupe 1 et le groupe 2 ; **b-** Le groupe 2 et le groupe 4 ; **c-** Le groupe 1 et le groupe 3.

2 Voici le tableau des effectifs des notes des élèves lors d'un examen :

Note	[0 ; 4[	[4 ; 8[	[8 ; 12[	[12 ; 16[	[16 ; 20[
Effectifs	5	32	61	80	15
Effectifs cumulés Croissants					
Fréquences cumulées Croissantes en %					

- 1) Quelle est la classe modale de cette série statistique.
2) Calcule la moyenne de l'établissement lors de cet examen.
(on donnera le résultat au centième près).

- 3) a- Remplir la ligne des effectifs cumulés croissant, puis des fréquences cumulées croissantes en pourcentage.
b- Construis, dans le repère ci-dessous, le polygone des fréquences cumulées croissantes en pourcentage.



- c- A l'aide du polygone des fréquences cumulées croissantes en pourcentage, donne la valeur de la médiane, du premier quartile et du troisième quartile.
(laissez les traits de constructions apparants).

- 3** Donne l'étendue, la médiane, le premier et le troisième quartile de la série statistique suivante :

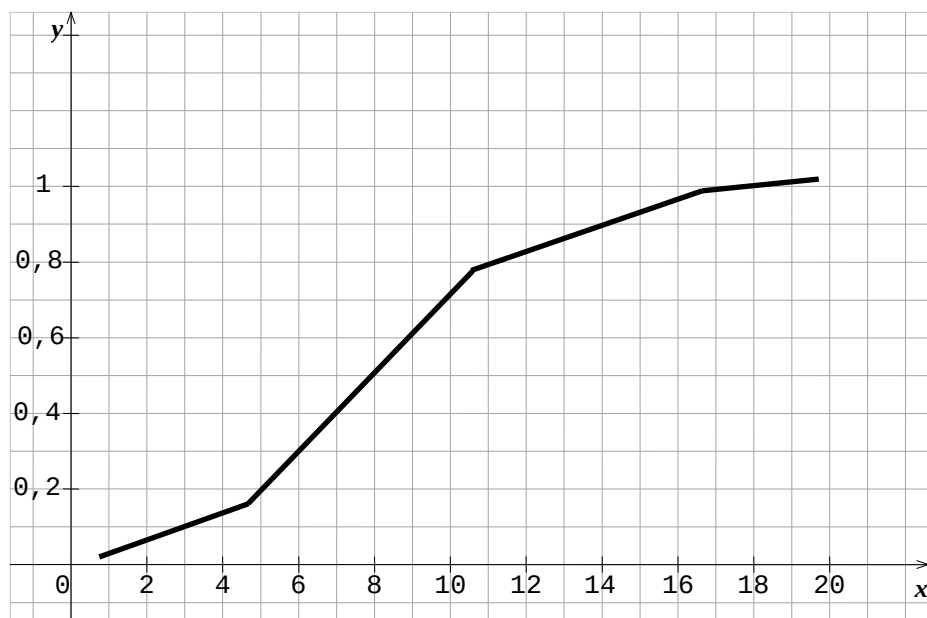
34	38	39	41	42	43	44	45	45	47
47	48	49	50	51	51	52	52	53	54
55	55	55	55	55	55	55	56	56	57
58	58	58	59	59	59	60	62	62	62
62	63	64	65	66	66	66	66	67	68
68	73	74	74	75	75	79	81	81	85

- 4** L'IMC est l'indice de masse corporelle. Dans une étude statistique portant sur 400 femmes, voici le tableau des effectifs de l'étude portant sur L'IMC de cette population :

IMC	19	20	21	22	23	24	25	26	27	28	29
Effectifs											
Effectifs cumulés croissants											

- 1) Complète, dans le tableau précédent la ligne des Effectifs cumulés croissants.
- 2) Détermine le premier quartile, la médiane, le troisième quartile et l'étendue de cette série.

- 5** Ci-dessous est représenté, le polygone des fréquences cumulées croissantes des notes obtenues par une classe de 28 élèves :



- 1) a- Détermine la médiane et les quartiles de cette série statistiques.
b- Donne la valeur de l'écart interquartile.
- 2) Tracer le diagramme en boîtes associé à cette série statistiques.

- 6** Le tableau ci-dessous représente les températures maximales dans la ville de Sikasso au cours d'une semaine :

Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
26,2	27	27,4	27,4	25,5	26	26,5

Les résultats seront arrondis au centième de degré Celsius.

- 1) Détermine la température maximale moyenne au cours de cette semaine.
- 2) Sachant que sur les deux semaines précédentes la moyenne de ces températures maximales étaient de 25,64. Détermine la moyenne des températures maximales sur ces trois semaines.

- 7** Une étude statistique sur une population a permis d'obtenir le tableau des effectifs suivants :

Taille	[150 ; 160[	[160 ; 170[	[170 ; 180[	[180 ; 190[
Effectif	3	23	79	7

Tous les calculs seront arrondis au centième près.

- 1) Calcule la moyenne de cette série statistique à partir du tableau des effectifs.
- 2) Complète le tableau suivant :

Classes	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
[150 ; 160[			
[160 ; 170[			
[170 ; 180[			
[180 ; 190[			

- 3) En déduire la valeur de la variance $V = \frac{1}{N} \sum_{i=1}^k n_i(x_i - \bar{x})^2$

- 4) Donne la valeur de l'écart-type : $\sigma = \sqrt{V}$

- 8** Soit tableau de la série statistique correspondant aux pointures des chemises de 48 militaires.

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

- 1) Détermine la population et le caractère étudié.
- 2) Détermine le mode des chemises
- 3) Calcule la pointure moyenne des chemises

- 9** L'étude statistique faite sur la taille des voitures de la société Renault a donné les résultats suivants :

Classes $[x_i; x_{i+1}[$	[144 ; 150[	[150 ; 156[	[156 ; 162[	[162 ; 168[	[168 ; 174[	[174 ; 180[	[180 ; 186[
Effectif	1	4	1	3	14	6	1

s	n_i						
-----	-------	--	--	--	--	--	--

- 1) Dresse le tableau des centres C_i .
- 2) Dresse le tableau des effectifs et des fréquences cumulées (croissants et décroissants).
- 3) En déduire l'histogramme des effectifs n_i puis celle effectif (n_i croissants et décroissants).

10 Soit la série suivante : 3 ; 8 ; 7 ; 11 ; 5 ; 13 ; 15 ; 10 ; 2 ; 17 ; 1 ; 20 ; 12.

Détermine la médiane et les quartiles.

11 Une agence de voyage propose 21 voyages organisés dont voici les tarifs en euros :

152 ; 320 ; 250 ; 320 ; 137 ; 183 ; 137 ; 152 ; 244 ; 122 ; 190 ;
139 ; 202 ; 285 ; 137 ; 180 ; 143 ; 190 ; 137 ; 180 ; 190.

- 1) Calcule la moyenne et l'écart type de la série des prix. Quel est la médiane, le premier et le troisième quartile?
- 2) Ne rentrant pas dans ses frais, l'agence hésite entre deux solutions : augmenter tous les tarifs de 190 euros ou doubler tous les tarifs. Donne pour chacune des solutions la moyenne et l'écart – type de la nouvelle série des prix. Que deviennent les quartiles et la médiane ?
- 3) Une des deux solutions est – elle plus avantageuse financièrement pour l'agence ? Tracer les deux diagrammes en boîte correspondants.

12 Lors d'un concours d'entrée à l'école statistique, la moyenne générale des 600 candidats a été de 11,4 sur 20.

Il y a 300 places au concours. Fanta a eu une moyenne de 12. Pourtant elle a échoué. Expliquer ce résultat.

13 Une série statistique a pour effectif 40. Sa moyenne est égale à 15,75 et son écart – type à 11,438. On ajoute à cette série une 41^{ème} valeur égale à 18. Calcule la moyenne et l'écart – type de la nouvelle série.

14 On a relevé le montant en euros de 100 chèques déposés dans un commerce dans une journée. Voici le tableau obtenu : les chèques ont été regroupés en fonction de leur montant par classes dont l'amplitude est 15 euros.

Montant des chèques	Effectif	Fréquence	Fréquence cumulée croissante
[0,15[	20		
[15,30[	9		
[30,45[	19		
[45,60[	27		
[60,75[	18		
[75,90[	6		
[90,105[	1		
Total			

- 1) Complète le tableau
- 2) Soit S le montant de l'ensemble des 100 chèques. Donne un encadrement de S . En déduire un intervalle I tel que le montant moyen des chèques appartienne à I .
- 3) Pour estimer la moyenne d'une série statistique où les valeurs ont été regroupées en classes de même amplitude, on procède souvent de la manière suivante :

On remplace chaque classe par son centre et on calcule la moyenne de la série des centres (avec les mêmes effectifs).

Faire le calcul pour la série ci-dessus. Vérifier que la moyenne obtenue est le centre de l'intervalle I trouvé à la question précédente.

4) Quelle est la classe pour laquelle la fréquence cumulée croissante passe la « barre » des 50 % ? Justifier pourquoi on appelle cette classe la classe médiane.

Série statistiques à doubles variables

- 15** Soit le tableau statistique suivant, portant sur l'évolution des ventes de médicaments des six dernières années d'une pharmacie :

Années x_i	2005	2006	2007	2008	2009	2010
Ventes y_i	100	130	150	175	210	240

Quelles sont les prévisions de vente de cette pharmacie en 2011 ? En 2012 ?

NB : On utilisera La méthode de **Mayer** et la méthode des **moindres carrés ordinaire**.

- 16** 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 x_i	9	10	11	12	14	15	16	17
Nombre d'acheteurs potentiels y_i	180	160	150	130	100	90	80	70

1) Représente graphiquement le nuage de points de cette série statistique. On prendra pour unité :

1 cm pour 1000 francs en abscisse et 1cm pour 10 acheteurs en ordonnée.

2) Détermine les coordonnées du point G

3) Détermine le coefficient de corrélation linéaire r puis interpréter le résultat obtenu.

4) Détermine 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 Détermine :

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.

- 17** 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- Calcule $\bar{X}$ et $\bar{Y}$ les moyennes des variables X et Y .
- 2- Représente 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- Calcule 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), Calcule une estimation du chiffre d'affaires de cette pharmacie à la fin du 7^{ème} mois.

Solutions

Série statistiques simples

- 1** Voici les notes de quatre groupes d'élèves au baccalauréat malien.

1) Remplissons les cases des différents indicateurs ci-dessous :

	Groupe 1	Groupe 2	Groupe 3	Groupe 4
Notes	5 – 6 – 10 – 10 – 11 12 – 12 – 14	6 – 8 – 8 – 8 – 10 11 – 14 – 15	8 – 8,5 – 8,5 – 9 11 – 11 – 12 – 12	6 – 6 – 7 – 8 – 10 11 – 11 – 15
Moyenne	10	10	10	9,25
Etendue	9	9	4	9
Médiane	10,5	9	10	9

Pour Complète le tableau ci-dessus, on à procédé de la manière suivante :

Exemple pour le groupe 1

Moyenne : $\bar{x} = \frac{\sum n_i}{N} = \frac{5 + 6 + 10 + 10 + 11 + 12 + 14}{8} = \frac{80}{8} = 10$

Etendue : c'est la différence entre la plus grande valeur et la plus petite valeur du caractère :
 $14 - 5 = 9$

Médiane : c'est la valeur du caractère qui divise la série en deux séries de même effectif

Ici l'effectif des notes pour le **groupe 1** est : 8 alors on pose $N = 8$ qui est un nombre pair.

Donc la médiane correspond à la valeur de rang : $\frac{N}{2}$ et $\frac{N}{2} + 1$ c'est-à-dire que la médiane correspond au : $\left(\frac{N}{2} \text{ et } \frac{N}{2} + 1\right)^{\text{ième}} \text{rang} = \left(\frac{8}{2} \text{ et } \frac{8}{2} + 1\right)^{\text{ième}} \text{rang} = (4 \text{ et } 4 + 1)^{\text{ième}} \text{rang} = (4 \text{ et } 5)^{\text{ième}} \text{rang}$

Or dans le groupe 1, le **4^{ième} rang** correspond à **10** et le **5^{ième} rang** correspond à **11**
Alors la médiane $Me = \frac{10 + 11}{2} = 10,5$

Ainsi le reste du tableau a été Complète avec la même procédure.

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 $\frac{N}{2} + 1$

2) Comparons d'un point de vue qualitatif à la lueur des indicateurs calculés précédemment :

a- Le **groupe 1** et le **groupe 2** ont la même moyenne et la même étendue. La médiane du **groupe 1** indique que ce groupe à la moitié de ces élèves ont eu des notes plus élevées que le **groupe 2**.

b- Le **groupe 2** et le **groupe 4** ont la même étendue et la même médiane. La moyenne plus grande du **groupe 2** indique que leurs notes dans la première moitié sont d'avantages regroupés vers la médiane 10 et la seconde moitié vers sa borne supérieure.

c- La médiane du groupe 3 peut aussi être choisie pour 10,5. Ainsi le **groupe 1** et le **groupe 3** ont la même moyenne et la même médiane. L'étendue du **groupe 3** étant plus faible indique une classe plus homogène.

2 1) Déterminons la classe modale de cette série statistique.

La classe modale de cette série statistique est la classe ayant le plus grand effectif c'est-à-dire : [12 ; 16[.

2) Calcule la moyenne de l'établissement lors de cet examen. (on donnera le résultat au centième près).

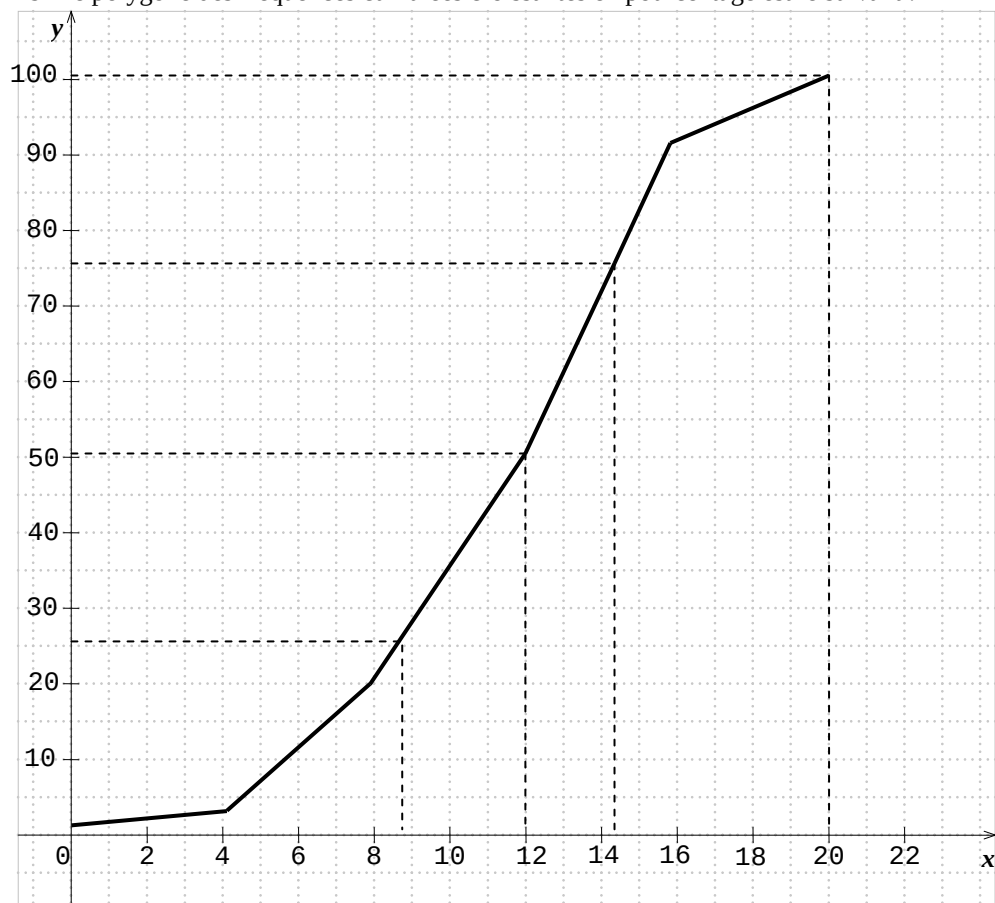
$$\bar{x} = \frac{\left(\frac{4+0}{2}\right) \times 5 + \left(\frac{8+4}{2}\right) \times 32 + \left(\frac{12+8}{2}\right) \times 61 + \left(\frac{16+12}{2}\right) \times 80 + \left(\frac{20+16}{2}\right) \times 15}{5 + 32 + 61 + 80 + 15}$$

$$= \frac{2 \times 5 + 6 \times 32 + 10 \times 61 + 14 \times 80 + 18 \times 15}{5 + 32 + 61 + 80 + 15} = \frac{2202}{193} \approx 11,41$$

3) a- On a le tableau complété ci-dessous.

Note	$[0 ; 4[$	$[4 ; 8[$	$[8 ; 12[$	$[12 ; 16[$	$[16 ; 20[$
Effectifs	5	32	61	80	15
Effectifs cumulés Croissants	5	37	98	178	193
Fréquences cumulées Croissantes en %	2,6	19,2	50,8	92,2	100

b- Le polygone des fréquences cumulées croissantes en pourcentage est le suivant :



c- La lecture graphique permet d'obtenir les valeurs suivantes :

Médiane : $Me = 12$; 1^{er} quartile : $Q_1 \approx 9$; 3^{ième} quartile : $Q_3 \approx 14,5$

- 3** Donnons : l'étendue, la médiane, le premier et le troisième quartile de la série statistique suivante :

34	38	39	41	42	43	44	45	45	47
47	48	49	50	51	51	52	52	53	54
55	55	55	55	55	55	55	56	56	57
58	58	58	59	59	59	60	62	62	62
62	63	64	65	66	66	66	66	67	68
68	73	74	74	75	75	79	81	81	85

Ce tableau comporte 6 lignes et 10 colonnes : il comporte donc 60 valeurs. Ainsi, on a les valeurs suivantes :

- Médiane : elle est comprise entre la 30^{ième} et la 31^{ième} valeur. Donc $Me = 57,5$.
- Le premier quartile est compris entre la 15^{ième} et la 16^{ième} valeur. Donc $Q_1 = 51$.
- Le troisième quartile est compris entre la 45^{ième} et la 46^{ième} valeur. Donc $Q_3 = 66$

- 4** 1) Le tableau Complète est le suivant :

IMC	19	20	21	22	23	24	25	26	27	28	29
Effectifs	25	37	106	92	38	39	16	12	15	13	7
Effectifs cumulés croissants	25	62	168	260	298	337	353	365	380	393	400

2) Déterminons le premier quartile, la médiane, le troisième quartile et l'étendue de cette série.

On a les indicateurs de positions suivants :

L'effectif étant 400 individus, la médiane se situe entre la 200^{ième} et la 201^{ième} valeur.
Donc $M_e = 22$.

- 5** 1) a- D'après le graphique, on a :

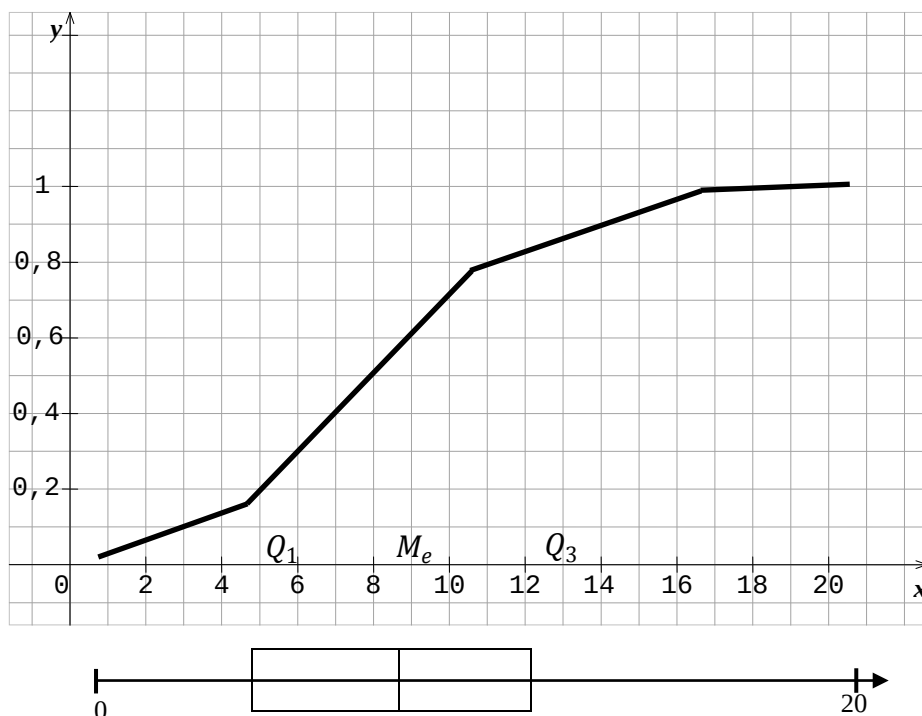
- Médiane : $Me = 8,78$.
- Le premier : $Q_1 = 5,75$.
- Le troisième : $Q_3 = 12$

b- Donnons la valeur de l'écart interquartile.

La valeur de l'écart interquartile correspond à : $Q_3 - Q_1 = 12 - 5,75 = 6,25$

2) Traçons le diagramme en boîtes associé à cette série statistiques.

(voir figure)

**6**

1) La température maximale moyenne au cours de cette semaine est :

$$\bar{x} = \frac{\sum n_i}{N} = \frac{26,2 + 27 + 27,4 + 24,7 + 25,5 + 26 + 26,5}{7} = \frac{183,3}{7} \approx 26,19^\circ \text{C}$$

2) Sachant que sur les deux semaines précédentes la moyenne de ces températures maximales étaient de 25,64.

Déterminons la moyenne des températures maximales sur ces trois semaines On a :

- 7 jours avec une moyenne de 26,19°C.
- 14 jours avec une moyenne de 25,64°C.

La moyenne sur ces trois semaines est obtenue par le calcul :

$$\bar{x} = \frac{7 \times 26,19 + 14 \times 25,64}{21} = \frac{542,29}{21} \approx 25,82^\circ \text{C}$$

7

1) Calculons la moyenne de cette série statistique à partir du tableau des effectifs.

$$\begin{aligned} \bar{x} &= \frac{\left(\frac{160+150}{2}\right) \times 3 + \left(\frac{170+160}{2}\right) \times 23 + \left(\frac{170+180}{2}\right) \times 79 + \left(\frac{190+180}{2}\right) \times 7}{3 + 23 + 79 + 7} \\ &= \frac{3 \times 155 + 23 \times 165 + 79 \times 175 + 7 \times 185}{112} = \frac{2202}{112} \approx 173,04 \end{aligned}$$

2) Le tableau complété est le suivant :

Classes	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
[150 ; 160[	-18,04	325,44	976,32
[160 ; 170[	-8,04	64,64	1486,76
[170 ; 180[	1,96	3,84	303,49
[180 ; 190[	11,96	143,04	1001,29

3) La valeur de la variance V est donnée par :

$$V = \frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = 33,64$$

4) Donnons la valeur de l'écart-type : σ

$$\sigma = \sqrt{V} = \sqrt{33,64} = 5,80$$

8 Soit tableau de la série statistique correspondant aux pointures des chemises de 48 militaires.

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

1) Déterminons la population et le caractère étudié.

La population désigne ici les militaires et le caractère étudié est la pointure des chemises.

2) Déterminons le mode des chemises

En observant le tableau, on remarque que la pointure 38 est portée par 12 personnes. Or 12 représente le plus grand nombre des effectifs n_i .

Par conséquent, le mode $M_0 = 38$

3) Calculons la pointure moyenne des chemises

Pour cela complétons le tableau statistique

x_i	36	37	38	39	40	41	42	43	Total
n_i	5	11	12	8	4	3	4	1	48
$n_i \times x_i$	180	407	456	312	160	123	168	43	1849

La moyenne des pointures de chemises est donnée par la formule suivante :

$$\bar{x} = \frac{\sum n_i \times x_i}{N} = \frac{1849}{48} \Rightarrow \bar{x} = 38,52$$

9 L'étude statistique faite sur la taille des voitures de la société Renault a donné les résultats suivants :

Classes $[x_i; x_{i+1}[$	[144 ; 150[	[150 ; 156[	[156 ; 162[	[162 ; 168[	[168 ; 174[	[174 ; 180[	[180 ; 186[
Effectifs n_i	1	4	1	3	14	6	1

1) Dressons le tableau des centres C_i .

Classes $[x_i ; x_{i+1}]$	Effectifs n_i	Centre des classes C_i	Effectifs n_i croissante	Effectifs n_i cumulé décroissante	Fréquence f_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				

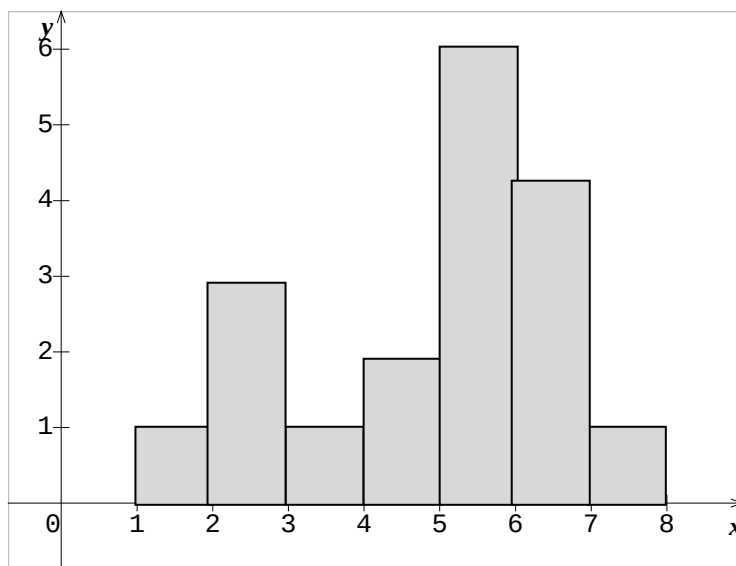
NB : Le tableau des effectifs et des fréquences cumulées a été dressé pour des classes d'amplitude 6.

Ainsi le centre C_i de la classe $[x_i ; x_{i+1}]$ est le nombre défini par $C_i = \frac{x_i + x_{i+1}}{2}$

Pour les calculs, nous conviendrons de prendre la valeur C_i commune à tous les individus de la classe $[x_i ; x_{i+1}]$

2) Dressons le tableau des effectifs et des fréquences cumulées (croissants et décroissants).
(Voir tableau ci-dessus)

3) En déduisons l'histogramme des effectifs n_i puis celle effectif (n_i croissants et décroissants).



10 Soit la série suivante : 3 ; 8 ; 7 ; 11 ; 5 ; 13 ; 15 ; 10 ; 2 ; 17 ; 1 ; 20 ; 12.

Déterminons la médiane et les quartiles.

Pour cela écrivons la liste dans l'ordre croissant :

1 ; 2 ; 3 ; 5 ; 7 ; 8 ; 10 ; 11 ; 12 ; 13 ; 15 ; 17 ; 20 ; 23 ; 26

N.B :

La médiane M_e est la valeur de rang $\frac{N+1}{2}$ si N est impair.

De même M_e est la valeur de rang $\left(\frac{N}{2} \text{ et } \frac{N}{2} + 1\right)$ si N est pair

Le premier quartile Q_1 de la série est la valeur dont le rang est le plus petit entier supérieur ou égal à $\frac{N}{4}$

Le troisième quartile Q_3 de la série est la valeur dont le rang est le plus petit entier supérieur ou égal à $\frac{3N}{4}$

Ainsi sur la liste on a 15 chiffres $\Rightarrow N = 15$ (chiffre impair)

- M_e correspond donc au $\left(\frac{N+1}{2}\right)^{ième}$ rang c'est-à-dire au $\left(\frac{15+1}{2}\right)^{ième} = 8^{ième}$ rang.

Or le $8^{ième}$ rang correspond à 11. D'où la médiane $M_e = 11$.

- Q_1 correspond donc au $\left(\frac{N}{4}\right)^{ième}$ rang c'est-à-dire au $\left(\frac{15}{4}\right)^{ième} = 3,75^{ième} \approx 4^{ième}$ rang.

Or le $4^{ième}$ rang correspond à 5. D'où le 1^{er} quartile $Q_1 = 5$

- Q_3 correspond donc au $\left(\frac{3N}{4}\right)^{ième}$ rang c'est-à-dire au $\left(\frac{3 \times 15}{4}\right)^{ième} = 11,75^{ième} \approx 12^{ième}$ rang.

Or le $12^{ième}$ rang correspond à 17. D'où le $3^{ième}$ quartile $Q_3 = 17$

- $Q_2 = M_e = 11$

Conclusion : $M_e = 11$; $Q_1 = 5$; $Q_2 = M_e = 11$; $Q_3 = 17$

11 Une agence de voyage propose 21 voyages organisés dont voici les tarifs en euros :

152 ; 320 ; 250 ; 320 ; 137 ; 183 ; 137 ; 152 ; 244 ; 122
190 ; 139 ; 202 ; 285 ; 137 ; 180 ; 143 ; 190 ; 137 ; 180 ; 190.

Dressons d'abord le tableau statistique des 21 voyages organisés.

x_i	122	137	139	143	152	180	183
n_i	1	4	1	1	2	2	1
$n_i \times x_i$	122	548	139	143	304	360	183
$n_i \times x_i^2$	14884	75076	19321	20449	46208	64800	33489

x_i	190	202	244	250	285	320	Total
n_i	3	1	1	1	1	2	21
$n_i \times x_i$	570	202	244	250	285	640	3990
$n_i \times x_i^2$	108300	40804	59536	62500	81225	204800	831392

1) Calculons la moyenne et l'écart type de la série des prix

La moyenne des prix est donnée par $\bar{x} = \frac{\sum n_i \times x_i}{N} = \frac{3990}{21} = 190$

Pour Détermine l'écart-type, nous allons Détermine d'abord la variance.

La variance est donnée par Variance : $V = \frac{\sum n_i \times x_i^2}{N} - (\bar{X})^2$

$$\Rightarrow V = \frac{831392}{21} - (190)^2 = 3490,09$$

D'où l'écart-type est : $\delta = \sqrt{3490,09} = 59,07$

Déterminons : la médiane le premier et le troisième quartile

De même que l'exercice 3 ; on écrit d'abord les valeurs de la série dans l'ordre croissante.
Ainsi on a :

122 ; 137; 137; 137; 137 ; 139 ; 143 ; 152; 152 ; 180; 180 ;
183 ; 190; 190; 190 ; 202 ; 244 ; 250 ; 285 ; 320; 320

La série ainsi rangée dans l'ordre, elle comprend 21 termes $\Rightarrow N = 21$

- M_e correspond donc au $\left(\frac{N+1}{2}\right)^{ième}$ rang c'est-à-dire au $\left(\frac{21+1}{2}\right)^{ième} = 11^{ième}$ rang.

Or le $11^{ième}$ rang correspond à 180. D'où la médiane $M_e = 180$.

- Q_1 correspond donc au $\left(\frac{N}{4}\right)^{ième}$ rang c'est-à-dire au $\left(\frac{21}{4}\right)^{ième} = 5,25^{ième} \approx 6^{ième}$ rang.

Or le $6^{ième}$ rang correspond à 139. D'où le 1^{er} quartile $Q_1 = 139$

- Q_3 correspond donc au $\left(\frac{3N}{4}\right)^{ième}$ rang c'est-à-dire au $\left(\frac{3 \times 21}{4}\right)^{ième} = 15,75^{ième} \approx 16^{ième}$ rang.

Or le $16^{ième}$ rang correspond à 202. D'où le $3^{ième}$ quartile $Q_3 = 202$

- $Q_2 = M_e = 180$

Conclusion : $M_e = 180$; $Q_1 = 139$; $Q_2 = M_e = 180$; $Q_3 = 202$

2) Ne rentrant pas dans ses frais, l'agence hésite entre deux solutions : augmenter tous les tarifs de 190 euros ou doubler tous les tarifs.

Donnons pour chacune des solutions la moyenne et l'écart – type de la nouvelle série des prix.

Si on ajoute 190 à chacun des termes de la série, la moyenne, la médiane et les quartiles augmentent de 190, l'écart – type ne change pas.

La deuxième série a une moyenne de 380, un écart – type de 59, une médiane de 370 et des quartiles de 329 et 392.

Si on double chacun des termes de la série, la moyenne, la médiane, les quartiles et l'écart – type doublent donc la moyenne devient 380, la médiane 360, les quartiles 278 et 404 et l'écart – type 108.

3) On remarque que la moyenne est la même dans les deux solutions envisagées, ce qui fait que ces solutions sont aussi avantageuses l'une que l'autre pour l'agence.

- 12** Lors d'un concours d'entrée à l'école statistique, la moyenne générale des 600 candidats a été de 11,4 sur 20. Il y a 300 places au concours. Fanta a eu une moyenne de 12. Pourtant elle a échoué.

Explication de l'échec de Fanta

Cet exercice illustre la différence entre la moyenne qui (est de 11,4) et la médiane qui (ici est strictement supérieur à 12).

Ainsi il y a 300 places au concours pour 600 candidats. Fanta n'a pas été reçue car son rang de classement était strictement supérieur à 300. Cela signifie donc qu'il existait plus de 300 candidats ayant plus de 12 de moyenne. Dans de nombreux cas, la médiane d'une série statistique est plus significative que sa moyenne.

- 13** Une série statistique a pour effectif 40. Sa moyenne est égale à 15,75 et son écart – type à 11,438. On ajoute à cette série une 41^{ème} valeur égale à 18. Calculons la moyenne et l'écart – type de la nouvelle série : la série statistique a pour effectif 40. Sa moyenne est égale à 15,75 et son écart – type à 11,438.

Donc la somme des 40 valeurs est $40 \times 15,75 = 630$.

On ajoute à cette série une 41^{ème} valeur égale à 18, la somme devient alors 648. La nouvelle moyenne est $\frac{648}{41}$ soit environs 15,8.

L'écart – type étant de 11,468 ; la variance est $V = 11,438^2$

On désigne par S la somme des carrés des valeurs de la série. D'après la deuxième formule de la variance : $V = \frac{S}{40} - 15,75^2$. On en déduit que $S = 40 (V + 248,0625)$.

Si on ajoute la valeur 18, la variance devient : $V' = \frac{S+18^2}{41} - 15,8^2$

Soit en remplaçant par les valeurs numériques, on a : $V' = 127,91156$.

Le nouvel écart – type est donc : $\sqrt{V'} = \sqrt{127,91156} = 11,31$

- 14** 1) complétons le tableau

Montant des chèques	Effectifs	Fréquences	Fréquences cumulées croissantes
[0 ; 15[	20	20%	20%
[15 ; 30[	9	9%	29%
[30 ; 45[	19	19%	48%
[45 ; 60[	27	27%	75%
[60 ; 75[	18	18%	93%
[75 ; 90[	6	6%	99%
[90 ; 105[	1	1%	100%
Total	100	100%	

- 2) Donnons un encadrement de S ; S étant le montant de l'ensemble des 100 chèques.

Il y a 20 chèques entre 0 et 15 euros. Le montant total de ces chèques est compris entre 0 et $20 \times 15 = 300$ euros.

De même, le montant total des 9 chèques suivants est compris entre 15×9 et 30×9 .

En raisonnant de même pour les autres catégories de chèques puis en additionnant les encadrements obtenus, on obtient un encadrement de S .

$$15 \times 9 + 30 \times 19 + 45 \times 27 + 60 \times 18 + 75 \times 6 + 90 \leq S < 15 \times 20 + 30 \times 9 + 45 \times 19 + 60 \times 27 + 75 \times 18 + 90 \times 6 + 105. \text{ D'où } 3540 \leq S < 5040$$

S est le montant total des 100 chèques. Le montant moyen des chèques est : $\frac{S}{100}$.

De l'encadrement de S , on déduit que : $35,40 \leq \frac{S}{100} < 50,40$.

L'intervalle I est $[35,4 ; 50,4[$

- 3) On commence par Détermine le centre de chaque classe, qui est égal à la demi – somme des bornes de l'intervalle.

Montant des chèques	Centre de la classe
$[0 ; 15[$	7,5
$[15 ; 30[$	22,5
$[30 ; 45[$	37,5
$[45 ; 60[$	52,5
$[60 ; 75[$	67,5
$[75 ; 90[$	82,5
$[90 ; 105]$	97,5

On calcule alors la moyenne de la série des centres (avec les mêmes effectifs).

Ainsi, on obtient comme moyenne : $m = 42,9$.

On calcule le centre de l'intervalle $I = \frac{5,4 + 50,4}{2} = 42,9$

Donc la moyenne obtenue est le centre de l'intervalle I trouvé à la question précédente.

4) La fréquence cumulée croissante passe la barre des 50% à partir de la classe $[45 ; 60[$. On appelle cette classe la classe médiane par analogie avec la définition de la médiane : celle-ci partage la série en deux classes de même effectif, c'est-à-dire que 50% des valeurs sont inférieures à la médiane.

Série statistiques à doubles variables

15

Soit le tableau statistique suivant, portant sur l'évolution des ventes de médicaments des six dernières années d'une pharmacie :

Années x_i	2005	2006	2007	2008	2009	2010
Ventes y_i	100	130	150	175	210	240

Déterminons les prévisions de vente de cette pharmacie en 2011 puis en 2012 en utilisant : La méthode de **Mayer** et la méthode des **moindres carrés ordinaire**.

Méthode des moindres carrés ordinaire.

- Première étape :

On attribue à chaque année du tableau, un rang compté à partir de 1 puis on complète le

							Total
Années	2005	2006	2007	2008	2009	2010	
Rang de l'année x_i	1	2	3	4	5	6	21
Ventes y_i	100	130	150	175	210	240	1005
$x_i \cdot y_i$	100	260	450	700	1050	1440	4000
x_i^2	1	4	9	16	25	36	91

tableau en ajoutant la colonne des $x_i \cdot y_i$ et celui des x_i^2 :

- Deuxième étape :

On calcule les coordonnées $\bar{x}$ et $\bar{y}$ du point moyen G tel que :

$$\bar{x} = \frac{\sum x_i}{n} = \frac{21}{6} = 3,5 \quad \text{et} \quad \bar{y} = \frac{\sum y_i}{n} = \frac{1005}{6} = 167,5$$

- Troisième étape :

On détermine la droite de régression (D) de y en x en calculant les réels a et b tel que :

$$a = \frac{\sum x_i \cdot y_i - n \bar{x} \cdot \bar{y}}{\sum x_i^2 - n(\bar{x})^2} = \frac{4000 - 6 \times 3,5 \times 167,5}{91 - 6 \times (3,5)^2} = 27,57 \quad \text{et} \quad b = \bar{y} - a \bar{x} = 167,5 - 27,57 \times 3,5$$

$$\Rightarrow b = 71,005 \approx 71$$

Alors on a (D): $y(x) = ax + b \Rightarrow y(x) = 27,57x + 71$

- Quatrième étape :

On calcule les prévisions de vente de cette pharmacie en 2011 et en 2012.

• **Prévisions de vente de cette pharmacie en 2011**

L'année 2011 correspond au rang $x = 7 \Rightarrow y(7) = 27,57 \times 7 + 71 = 263,99 \approx 264$

• **Prévisions de vente de cette pharmacie en 2012**

L'année 2012 correspond au rang $x = 8 \Rightarrow y(8) = 27,57 \times 8 + 71 = 291,56 \approx 292$

Méthode de Mayer

- Première étape :

On attribue à chaque année du tableau, un rang compté à partir de 1 puis on partage le nuage en deux nuages N_1 et N_2 dont les effectifs sont égaux.

	Groupe 1			Groupe 2		
Années	2005	2006	2007	2008	2009	2010
Rang de l'année x_i	1	2	3	4	5	6
Ventes y_i	100	130	150	175	210	240

- Deuxième étape :

On calcule un point moyen pour chacun des deux groupes :

Premier groupe

Soit $\bar{y} = a\bar{x} + b$ la droite de tendance relative au premier groupe.

Calculons les coordonnées du point moyen $G_1 \left(\bar{x}, \bar{y} \right)$ relative premier groupe

Rang de l'année x_i	1	2	3
Ventes y_i	100	130	150

$$\bar{y} = \frac{\sum y_i}{n} = \frac{100+130+150}{3} = \frac{380}{3} = 126,66 \quad \text{et} \quad \bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3}{3} = \frac{6}{3} = 2$$

Alors le point moyen pour le premier groupe est : $G_1 \left(2, 126,66 \right)$

Or la droite de tendance relative au premier groupe est : $\bar{y} = a\bar{x} + b \Leftrightarrow 126,66 = 2a + b \Leftrightarrow 2a + b = 126,66$ (1)

Deuxième groupe

Soit $\bar{y} = a\bar{x} + b$ la droite de tendance relative au premier groupe.

Calculons les coordonnées du point moyen $G_2 \left(\bar{x}, \bar{y} \right)$ relative premier groupe

Rang de l'année x_i	4	5	6
Ventes y_i	175	210	240

$$\bar{y} = \frac{\sum y_i}{n} = \frac{175+210+240}{3} = \frac{625}{3} = 208,33$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{4+5+6}{3} = \frac{15}{3} = 5$$

Alors le point moyen pour le deuxième groupe est : $G_2 \left(\frac{5}{208,33} \right)$

Or la droite de tendance relative au premier groupe est : $\bar{y} = a\bar{x} + b \Leftrightarrow 208,33 = 5a + b$

$$\Leftrightarrow 5a + b = 208,33 \text{ (2)}$$

- **Troisième étape :**

On résout le système formé par les équations (1) et (2) pour Détermine les réels a et b :

$$\begin{cases} 2a + b = 126,66 \\ 5a + b = 208,33 \end{cases}$$

La résolution de ce système donne : $a = 27,22$ et $b = 72,23$

Or la droite de tendance de la série statistique est: $y(x) = ax + b \Leftrightarrow y(x) = 27,22x + 72,23$

- **Quatrième et dernière étape :**

On calcule les prévisions de vente de cette pharmacie en 2011 et en 2012.

- **Prévisions de vente de cette pharmacie en 2011**

L'année 2011 correspond au rang $x = 7$

$$\Rightarrow y(7) = 27,22 \times 7 + 72,23 = 262,77 \approx 263$$

- **Prévisions de vente de cette pharmacie en 2012**

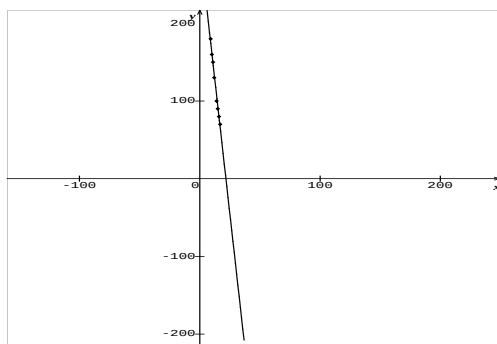
L'année 2012 correspond au rang $x = 8$

$$\Rightarrow y(8) = 27,22 \times 8 + 72,23 = 289,99 \approx 290.$$

16 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 x_i	9	10	11	12	14	15	16	17
Nombre d'acheteurs potentiels y_i	180	160	150	130	100	90	80	70

- 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.

$$\bar{X} = \frac{\sum_{i=1}^8 x_i}{8} = \frac{9+10+11+12+14+15+16+17}{8} = 13 \quad \Rightarrow G(13 ; 120)$$

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

3-Détermination du coefficient de corrélation linéaire r puis interprétons le résultat obtenu.

$$\begin{aligned} \text{Cov}(X ; Y) &= \frac{\sum_{i=1}^n x_i \cdot y_i}{n} - \bar{X} \cdot \bar{Y} \\ &= \frac{9 \times 180 + 10 \times 160 + 11 \times 150 + 12 \times 130 + 14 \times 100 + 15 \times 90 + 16 \times 80 + 17 \times 70}{8} - 13 \times 120 \end{aligned}$$

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

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

$$V(X) = \frac{\sum x_i^2}{N} - \bar{X}^2 = \frac{9^2+10^2+11^2+12^2+14^2+15^2+16^2+17^2}{8} - 13^2 = 7,499$$

$$V(Y) = \frac{\sum y_i^2}{N} - \bar{Y}^2 = \frac{180^2+160^2+150^2+130^2+100^2+90^2+80^2+70^2}{8} - 120^2 = 1450,003$$

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

$0,87 \leq |r| \leq 1$: 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.

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

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

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

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

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 $x = 13$

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

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

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

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

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

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

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

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

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

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 Détermine x pour $y = 250$

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

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

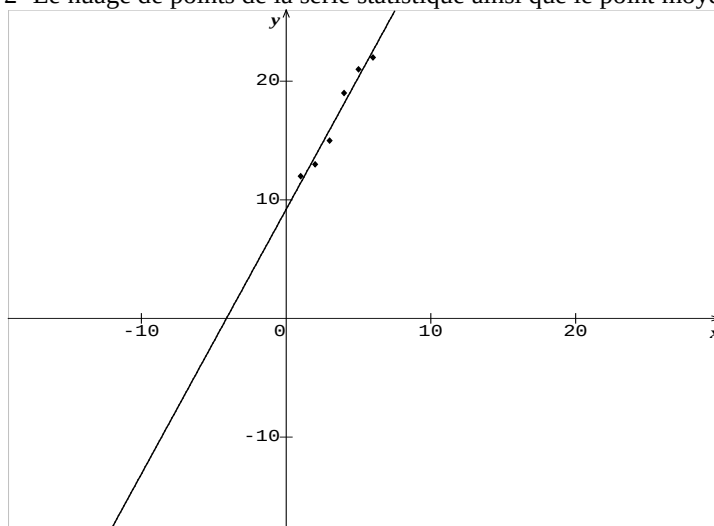
17 1- Calcul de $\bar{X}$ et $\bar{Y}$ les moyennes des variables x et y .

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

	X_i	Y_i	$X_i \cdot Y_i$	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
Total	21	102	396	91	2 331

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

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



3) Calcul de la variance $V(x)$ de x et la covariance $Cov(x, y)$ de x et y .

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

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

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

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

5) 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 $A(0 ; 9,2)$.

6) 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 Détermine 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.