

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	$10/10 = 1$

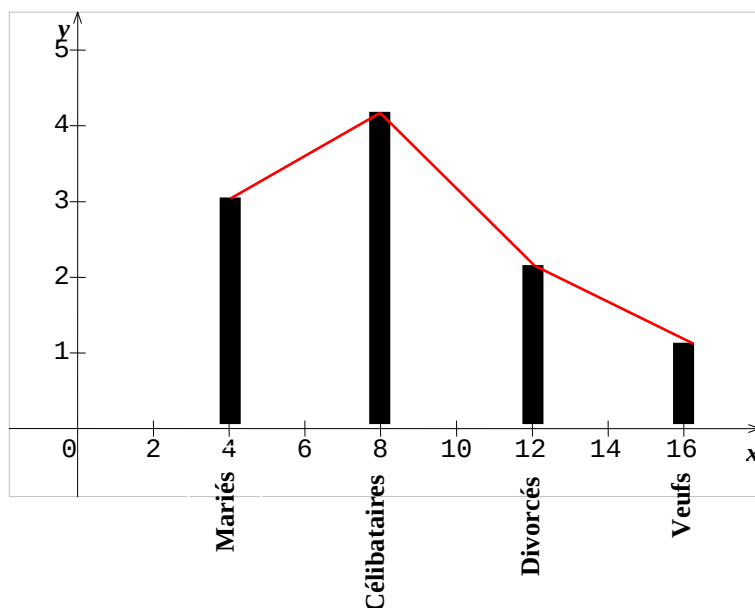
NB : la somme des fréquences est toujours égale à 1 , c'est-à-dire : $\sum f_i = 1$

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 $\alpha_i = f_i \times 180^\circ$.

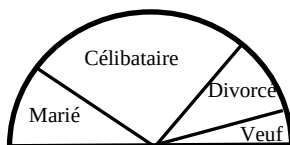
Mariés : $\alpha_1 = 0,3 \times 180^\circ \Rightarrow \alpha_1 = 54^\circ$

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

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

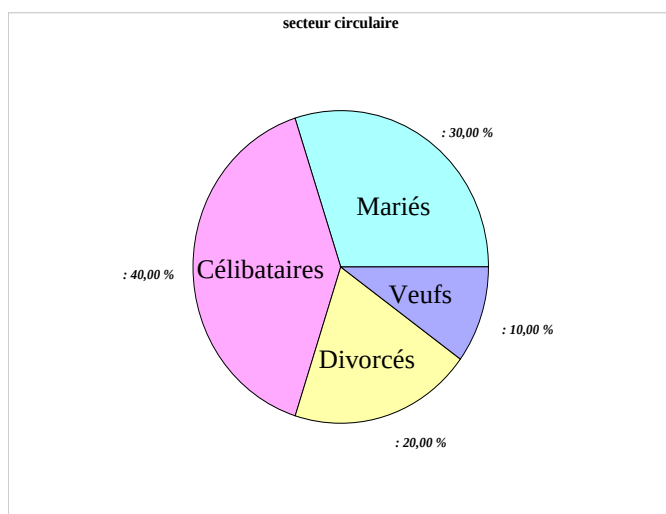
Veufs : $\alpha_4 = 0,1 \times 180^\circ \Rightarrow \alpha_4 = 18^\circ$

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

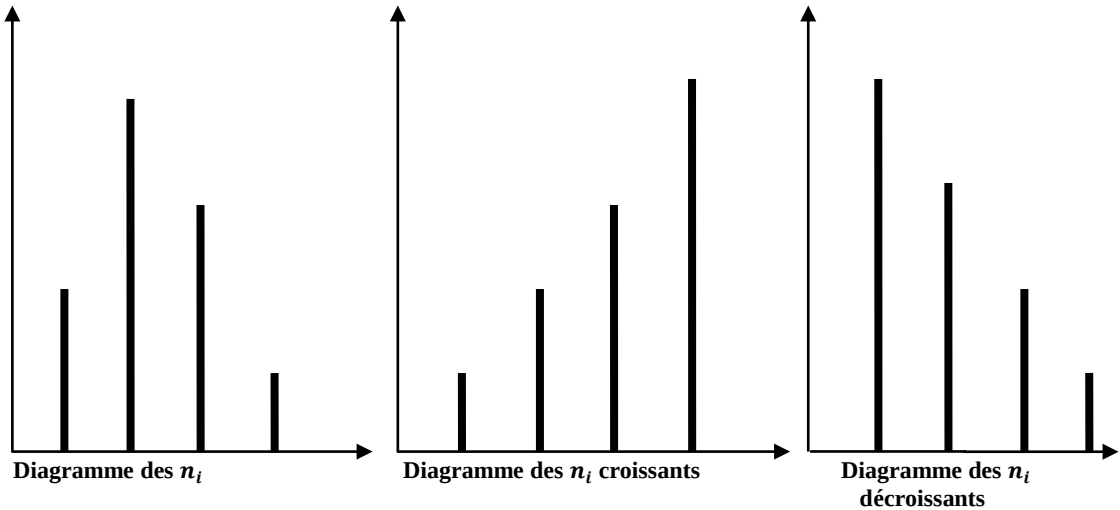
D'où le tableau suivant :

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.

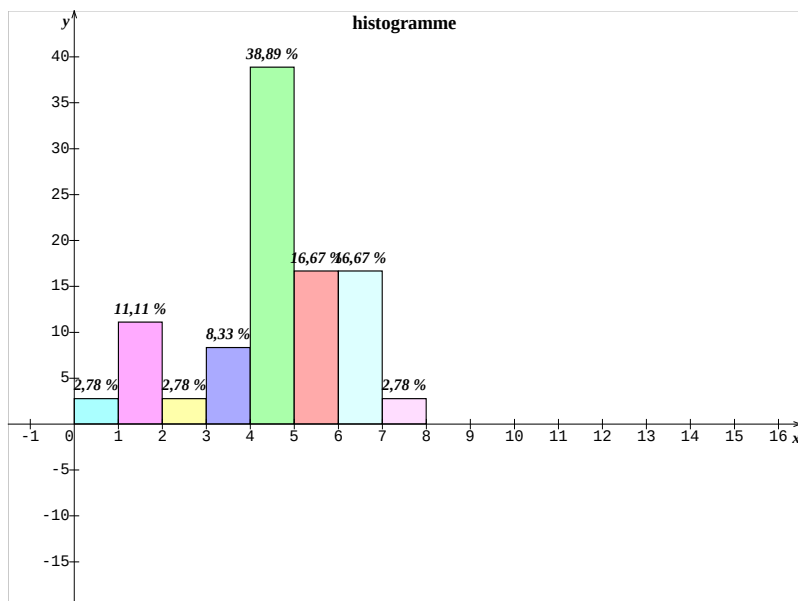
On prendra des classes d'amplitude 6.

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.**

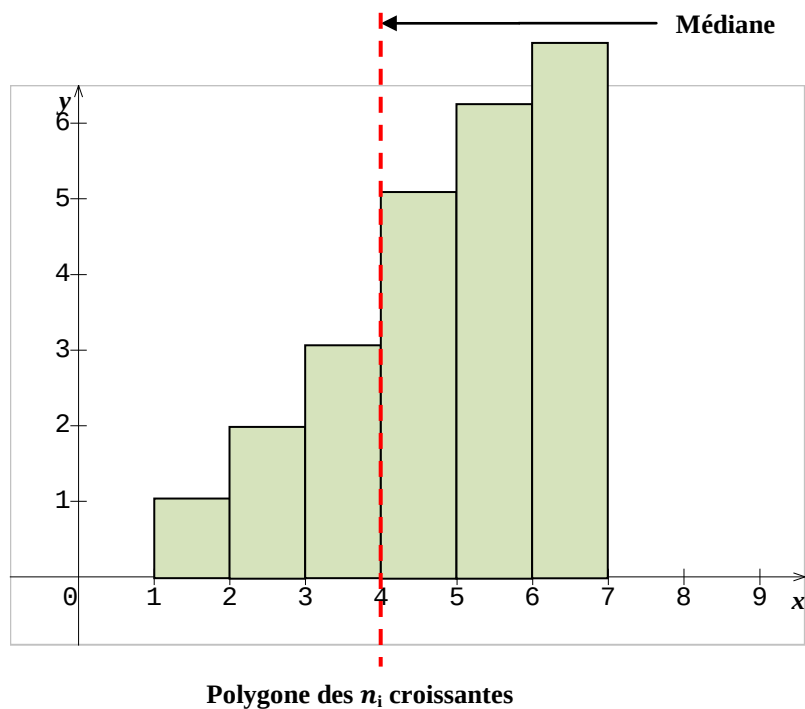
Les polygones :

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 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- 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.