Table des matières

Chapitre 1

STATISTIQUES

1.1 GÉNÉRALITÉS

1.1.1 La statistique

La statistique est une science qui a pour objet de collecter, classer, analyser et présenter de façon compréhensible un ensemble de données. Ces données peuvent prévenir de plusieurs domaines comme : la médecine, l'éducation, l'économie,...

1.1.2 Les statistiques

Les statistiques sont les résultats produit par la statistique entant que science.

1.1.3 Collecte des données

Elle se fait au moyen de deux approches : le recensement et le sondage.

a) Le recensement

Il consiste à interroger chaque individu de la population cible.

b) Le sondage

Il consiste à interroger une partie de la population qu'on appelle échantillon.

1.2 Série statistique à une variable

Soit la série statistique suivante :

x_i	x_1	x_2	x_3	...	x_k
n_i	n_1	n_2	n_3	...	n_k

1.2.1 Effectif total

On appelle effectif total d'une modalité n_i le nombre d'individus ayant cette modalité. On le note souvent par N.

On a : $N = \sum_{i=0}^{k} n_i = n_1 + n_2 + n_3 + ... + n_k$; avec $\sum =$ la somme.

1.2.2 Fréquence

On appelle fréquence d'une modalité, le quotient de l'effectif de cette modalité par l'effectif total de la population. Elle est notée par f_i et définie par : $f_i = \frac{n_i}{N} \times 100$

Les fréquences sont généralement exprimées en pourcentage.

Exemple

Soit la série statistique suivante :

x_i	1	3	0	4
n_i	3	2	4	1

1) Quel est l'effectif total ?

2) Calculer les fréquences des modalités 1, 3 et 0.

Solution

1) Effectif total

   $N = 10$

2) Calculons les fréquences des modalités 1, 3 et 0.

   $f_1 = \frac{3}{10}$; $f_2 = \frac{2}{10}$ et $f_3 = \frac{4}{10}$

N.B : La fréquence est toujours un nombre compris entre 0 et 1, et la somme des fréquences est égale à 1 ou $\sum_{i=1}^{k} f_i = 1$

1.2.3 Moyenne

On a : $\overline{x} = \frac{1}{N} \sum_{i=1}^{k} n_i x_i = \frac{n_1.x_1 + n_2.x_2 + ... + n_k.x_k}{N}$ ou $\overline{x} = \sum_{i=1}^{k} x_i f_i$.

Exercice

On donne le tableau statistique suivant :

x_i	2	1	5
n_i	4	2	1

Calculer la moyenne de cette série.

Solution

Calculons la moyenne de cette série.
$\overline{X} = \dfrac{8+2+5}{7} = \dfrac{15}{7}$

1.2.4 Étendue

L'étendue est la différence entre la plus grande valeur et la plus petite valeur de la série.
On a : $E = Max - Min$.

1.2.5 Variance

On a : $V(X) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{k} n_i(x_i - \overline{x})^2$.
Théorème de Köenig
$V(X) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2$.
N.B : La variance est un nombre positif.

1.2.6 Écart-type

On a : $\rho(X) = \sqrt{V(X)}$.

1.2.7 Écart-moyen

On appelle écart moyen de la série statistique $(x_i, n_i)_{1\leq i\leq k}$ de moyenne $\overline{x}$, le nombre réel e_m défini par :
$\mathrm{e}_m = \dfrac{n_1|x_1-\overline{x}| + n_2|x_2-\overline{x}| + ... + n_k|x_k-\overline{x}|}{n_1+n_2+n_3+...+n_k}$
$\mathrm{e}_m = \dfrac{1}{N}\displaystyle\sum_{i=1}^{k} n_i|x_i - \overline{x}|$

1.2.8 Effectifs cumulés

Soit $(x_i\,;\,n_i)_{1\leq i\leq k}$ une série statistique dont les modalités sont rangées par ordre croissant.
▷ On appelle effectif cumulé croissant noté $(E.C.C)$ de modalité x_i la somme des effectifs des modalités inférieures ou égale à x_i.
▷ On appelle effectif cumulé décroissant noté $(E.C.D)$ de modalité x_i la somme des

effectifs des modalités supérieures ou égale à x_i.

Exercice 1

Compléter le tableau statistique suivant :

x_i	0	2	3	4	5	6	7	9
n_i	2	3	5	4	11	11	8	4
$E.C.C$								
$E.C.D$								

Solution

Complétons le tableau statistique suivant :

x_i	0	2	3	4	5	6	7	9
n_i	2	3	5	4	11	11	8	4
$E.C.C$	2	5	10	14	25	36	44	48
$E.C.D$	48	46	43	38	34	23	12	4

1.2.9 Fréquences Cumulées

▷ On appelle fréquence cumulée croissante notée ($F.C.C$) de modalité x_i le quotient de son effectifs croissant par l'effectif total.

$F.C.C = \dfrac{E.C.C}{N}$

▷ On appelle fréquence cumulée décroissante notée ($F.C.D$) de modalité x_i le quotient de son effectifs décroissant par l'effectif total.

$F.C.D = \dfrac{E.C.D}{N}$

1.2.10 Mode d'une série statistique

On appelle mode d'une série statistique toute modalité d'effectif maximal.

Exemple

Une enquête effectuée à la demande d'un fabriquant de chaussures et portant sur les pointures d'une population masculin a donné les résultats suivants :

Pointures	39	40	41	43
Fréquences	5%	10%	65%	20%

Le mode de cette série statistique est 43.

1.2.11 Médiane d'une série statistique

La médiane d'une série statistique ou d'une série de mesures rangées par ordre de grandeur (croissant ou décroissant) est la valeur de l'observation qui se situe au milieu de la série ; c'est-à-dire la valeur de la variable qui partage les observations en deux

effectifs égaux. Soit $x_1, x_2, x_3, ..., x_n$ cette série d'observation, n le nombre d'observation.
On distingue deux cas :

▷ Si n est pair, la médiane est la moyenne arithmétique des valeurs qui occupent les positions $\frac{n}{2}$ et $\frac{n+2}{2}$.

▷ Si n est impair, alors la médiane est la valeur qui occupe la position $\frac{n+1}{2}$.

Exemple

▷ On relève les âges d'un groupe de 15 personnes et on les range par ordre croissant :
12, 12, 13, 13, 13, 15, 16, 16, 17, 17, 17, 18, 18, 18, 19.
$n = 15$ la médiane est donc la valeur qui occupe la position $\frac{15+1}{2} = 8$. D'où $M = 16$.

▷ Une personne, âgée de 18 ans, vient se joindre au groupe précédent.
La série, ordonnée , devient : 12, 12, 13, 13, 13, 15, 16, 16, 17, 17, 17, 18, 18, 18, 18, 19.
On a : $n = 16$, la médiane est la moyenne arithmétique des valeurs qui occupent les positions $\frac{16}{2} = 8$ et $\frac{16+2}{2} = 9$. D'où $M = \frac{16+17}{2} = 16,5$.

1.3 Quartiles

1.3.1 Définition

Les quartiles d'une série statistique sont les valeurs qui partagent cette série en quatre séries de même effectif.
On distingue trois types de quartiles :

▷ le premier quartile noté Q_1, qui correspond à 25% d'effectif total.

▷ le deuxième quartile noté Q_2, qui correspond à 50% d'effectif total.
Il s'agit de la médiane.

▷ le troisième quartile noté Q_3, qui correspond à 75% d'effectif total.

N.B : 25% 50% et 75% sont lus dans les fréquences cumulées croissantes.

1.3.2 Méthode pour trouver les quartiles

Il faut commencer par classer la série dans l'ordre croissant.
On utilise une méthode approximative mais qui donner des résultats significatifs pour des séries à grands effectifs (N).
On calcule $\frac{N}{4}$ et on note a l'entier supérieur à $\frac{N}{4}$.

On calcule $\frac{3N}{4}$ et on note b l'entier supérieur à $\frac{3N}{4}$.

▷ Q_1 est a^{ieme} valeur de la série statistique.

▷ Q_3 est b^{ieme} valeur de la série statistique.

N.B : Q_1 et Q_3 sont forcement des valeurs de la série.

Exercice 2

On considère la série suivante :

x_i	3	4	5	7	8	10	11
n_i	5	7	3	8	8	6	3

1. Quel est l'effectif total de cette série statistique ?
2. Déterminer la valeur de la médiane de la série statistique.
3. Déterminer le premier et le troisième quartiles de la série statistique.

1.3.3 Écart inter-quartiles et Intervalle inter-quartiles

Définition

▷ On appelle écart inter-quartiles noté E_Q la différence entre Q_3 et Q_1.

On a : $E_Q = Q_3 - Q_1$.

▷ On appelle intervalle inter-quartiles noté I_Q l'intervalle entre Q_1 et Q_3.

On a : $I_Q = [Q_1; Q_3]$.

1.3.4 Coefficient inter-quartile relatif

On appelle coefficient inter-quartile relatif le rapport de l'écart inter-quartile par le deuxième quartile.

On a : $c = \frac{E_Q}{Q_2} = \frac{Q_3 - Q_1}{Q_2}$.

1.4 Déciles

1.4.1 Définition

Les déciles d'une série statistique sont les valeurs qui partagent cette série en dix séries de même effectif.

▷ le premier décile noté D_1, qui correspond à 10% d'effectif total.

▷ le neuvième décile noté D_9, qui correspond à 90% d'effectif total.

N.B : 10% et 90% sont lus dans les fréquences cumulées croissantes.

1.4.2 Méthode pour trouver les déciles

Il faut commencer par classer la série dans l'ordre croissant.
On utilise une méthode approximative mais qui donner des résultats significatifs pour des séries à grands effectifs (N).
On calcule $\dfrac{N}{10}$ et on note a l'entier supérieur à $\dfrac{N}{10}$.

On calculer $\dfrac{9N}{10}$ et on note b l'entier supérieur à $\dfrac{9N}{10}$.
▹ D_1 est a^{ieme} valeur de la série statistique.
▹ D_9 est b^{ieme} valeur de la série statistique.

1.4.3 Écart inter-déciles et Intervalle inter-déciles

Définition

▹ On appelle écart inter-déciles noté E_D la différence entre D_9 et D_1.
On a : $E_D = D_9 - D_1$.
▹ On appelle intervalle inter-déciles noté I_D l'intervalle entre D_1 et D_9.
On a : $I_D = [D_1; D_9]$.

Exercice 3

On considère la série statistique suivante :

x_i	1	2	4	7	8	10	12
n_i	8	15	19	31	11	10	6

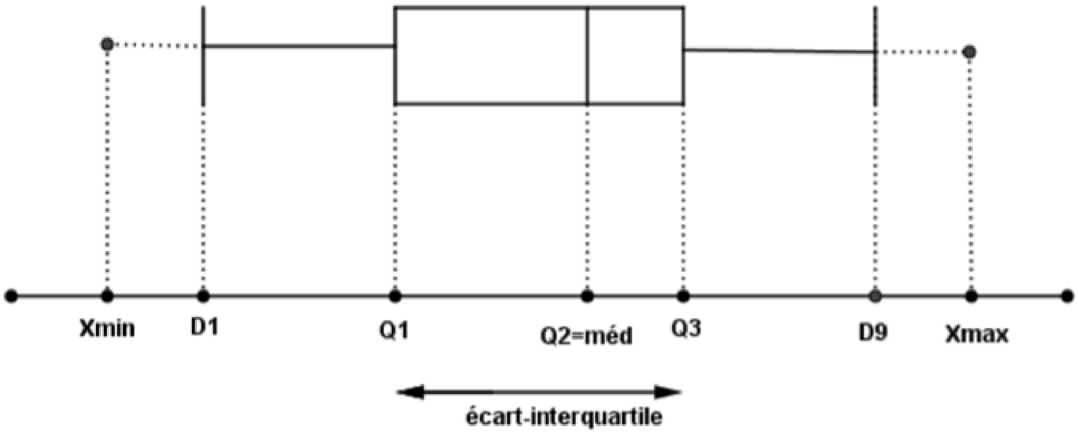
1. Quel est l'effectif total de cette série ?
2. Déterminer le premier et le neuvième déciles.
3. Déterminer écart inter-déciles.
4. Déterminer l'intervalle inter-déciles.

1.4.4 Diagramme en boîte

Définition

Un diagramme en boîte ou boîte à moustaches permet de représenter une série statistique au moyen d'une boîte
rectangulaire sur laquelle sont indiqués les informations suivantes :
▹ premier quartile Q_1, le troisième quartile Q_3.
▹ premier décile D_1, le neuvième décile D_9.

▷ la médiane Me.



Exercice 4

On a demandé à 50 personnes prenant l'auto bus pendant une semaine, le nombre de fois où chacune de ces personnes a utilisé ce type de transport pendant la semaine écoulée. Voici les résultats :

x_i (nombre de voyageur)	1	2	3	4	5	6	7	8	9	10
n_i	3	3	5	7	6	9	5	4	5	3
E.C.C										
$f_i(\%)$										
F.C.C (%)										

1. Recopier et compléter le tableau.
2. Quel est l'effectif total de cette série statistique ?
3. Déterminer la médiane.
4. Déterminer les quartiles de cette série.
5. Déterminer les déciles de cette série.
6. Déterminer Écart inter-quartiles et Intervalle inter-quartiles.
7. Déterminer Écart inter-déciles et Intervalle inter-déciles.
8. Construire un diagramme en boite.

Solution 2

1. Effectif total de cette série statistique.
$N = \sum_{i=1}^{7} n_i = 5 + 7 + 3 + 8 + 8 + 6 + 3 = 40$
D'où $N = 40$

2. Déterminons la valeur de la médiane de la série statistique

   Comme $N = 40$ est pair, alors la médiane est donc la moyenne arithmétique des valeurs qui occupent les positions $\frac{N}{2} = 20$ et $\frac{N+2}{2} = 21$, alors $M_e = \frac{7+7}{2} = 7$
   D'où $M_e = 7$

3. Déterminons le premier et le troisième quartiles de la série statistique

   On a : $\begin{cases} \frac{N}{4} = \frac{40}{4} = 10 \\ \frac{3N}{4} = \frac{120}{4} = 30 \end{cases} \Longrightarrow \begin{cases} Q_1 = 4 \\ Q_3 = 8 \end{cases}$

Solution 3

1. Effectif total de cette série statistique.

   $N = \sum_{i=1}^{7} n_i = 100$
   D'où $N = 100$

2. Déterminons le premier et le neuvième déciles de la série statistique

   On a : $\begin{cases} \frac{N}{10} = \frac{100}{10} = 10 \\ \frac{9N}{10} = \frac{900}{10} = 90 \end{cases} \Longrightarrow \begin{cases} D_1 = 2 \\ D_9 = 10 \end{cases}$

3. Déterminons l'écart inter-déciles

   $E_D = D_9 - D_1 = 10 - 2 = 8 \Longrightarrow E_D = 8$

4. Déterminons l'intervalle inter-déciles

   $I_D = [D_1; D_9] \Longrightarrow I_D = [2; 10]$

Solution 4

1. Recopions et complétons le tableau

x_i (nombre de voyageur)	1	2	3	4	5	6	7	8	9	10
n_i	3	3	5	7	6	9	5	4	5	3
E.C.C	3	6	11	18	24	33	38	42	47	50
$f_i(\%)$	6%	6%	10%	14%	12%	18%	10%	8%	10%	6%
F.C.C (%)	6%	12%	22%	36%	48%	66%	76%	84%	94%	100%

2. Donnons effectif total de cette série statistique.

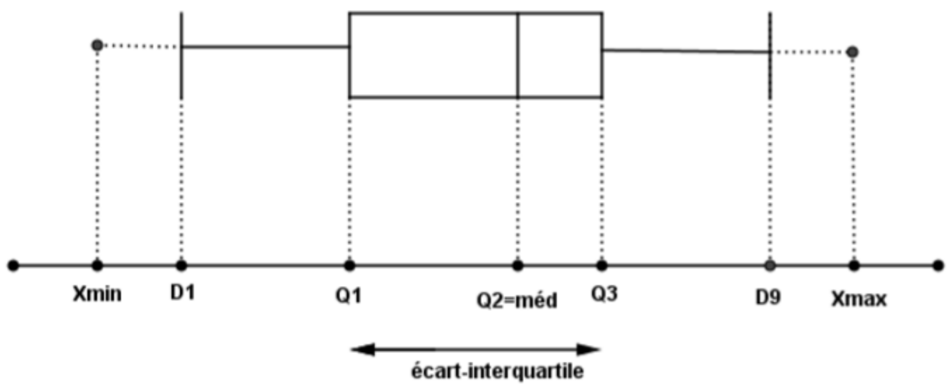
   D'après l'énoncé de l'exercice, on a : $N = 50$

3. Déterminons la médiane de la série statistique

   Comme $N = 50$ est pair, alors la médiane est donc la moyenne arithmétique des

valeurs qui occupent les positions $\frac{N}{2} = 25$ et $\frac{N+2}{2} = 26$, alors $M_e = \frac{6+6}{2} = 6$
D'où $M_e = 6$.

4. Déterminons le premier et le troisième quartiles de la série statistique
   On a : $\begin{cases} \frac{N}{4} = \frac{50}{4} = 12?5 \\ \frac{3N}{4} = \frac{150}{4} = 37,5 \end{cases} \Longrightarrow \begin{cases} Q_1 = 4 \\ Q_3 = 7 \end{cases}$
5. Déterminons le premier et le neuvième déciles de la série statistique
   On a : $\begin{cases} \frac{N}{10} = \frac{50}{10} = 5 \\ \frac{9N}{10} = \frac{450}{10} = 45 \end{cases} \Longrightarrow \begin{cases} D_1 = 2 \\ D_9 = 9 \end{cases}$
6. Déterminons l'écart inter-quartiles et intervalle inter-quartiles
   $E_Q = Q_3 - Q_1 = 7 - 4 \Longrightarrow E_Q = 3$
   $I_D = [Q_1; Q_3] \Longrightarrow I_Q = [4; 7]$
7. Déterminons l'écart inter-déciles et l'intervalle inter-déciles
   $E_D = D_9 - D_1 = 9 - 2 \Longrightarrow E_D = 7$ et $I_D = [D_1; D_9] \Longrightarrow I_D = [2; 9]$
8. Construisons un diagramme en boite



1.5 Série statistique à deux variables

1.5.1 Définition

Soit Ω une population de N individus.
On appelle série statistique à deux variables, l'ensemble des triplets (x_i, y_j, n_{ij}).
n_{ij} est l'ensemble d'individus de Ω prenant simultanément les valeurs x_i et y_j de deux caractères statistiques X et Y avec $i \in \{1, 2, 3, 4, ..., p\}$ et $j \in \{1, 2, 3, 4, ..., q\}$.

1.5.2 Tableau linéaire

Soit le tableau statistique suivant :

x_i	x_1	x_2	x_3	...	x_p
y_i	y_1	y_2	y_3	...	y_p

a) Lois marginales

▷ Pour X

x_i	x_1	x_2	x_3	...	x_p
n_i	1	1	1	...	1

▷ Pour Y

y_i	y_1	y_2	y_3	...	y_p
n_i	1	1	1	...	1

Remarque

Dans le cas d'un tableau linéaire l'effectif total est égal au nombres de carreaux des valeurs de x_i et y_i.

Exemple :

x_i	2	3	5	7
y_i	-1	0	1	2

On a $N = 4$

b) Nuage des points

L'ensemble des points M_{ij} de coordonnées (x_i, y_j) est appelé nuage des points d'une série statistique.

Exemple :

x_i	2	3	4	5
y_i	0	1	2	3

Représenter le nuage des points de cette série statistique.

c) Point moyen

On appelle point moyen du nuage, le point G de coordonnées $(\overline{x}, \overline{y})$ où $\overline{x}$ et $\overline{y}$ désignent respectivement les moyennes des séries marginales (x_i, n_i) et (y_j, n_j).

On note $G\left(\overline{x}, \overline{y}\right)$

1.5.3 Ajustement linéaire

Ajuster un nuage de points consiste à déterminer une courbe simple passant " le plus prés possible " par des points du nuage. Si la courbe cherchée est une droite, on dit que l'ajustement est linéaire.

a) Droite de régression par la Méthode de MAYER

La méthode de MAYER consiste à diviser le nuage des points en deux sous-nuages d'effectifs égaux. Soit $G_1(\overline{x}_1, \overline{y}_1)$ et $G_2(\overline{x}_2, \overline{y}_2)$ les points moyens respectifs des sous-nuages ainsi obtenus. La droite (G_1G_2) est appelée droite de MAYER.

b) Équation de la droite (G_1G_2)

les points moyens $G_1(\overline{x}_1, \overline{y}_1)$ et $G_2(\overline{x}_2, \overline{y}_2)$

On a : l'équation de la droite passant par G_1 et G_2 est : $\dfrac{x - \overline{x}_1}{\overline{x}_2 - \overline{x}_1} = \dfrac{y - \overline{y}_1}{\overline{y}_2 - \overline{y}_1}$.

N.B : cette droite est de la forme : $y = ax + b$ avec $a = \dfrac{\overline{y}_1 - \overline{y}_2}{\overline{x}_1 - \overline{x}_2}$ et $b = \overline{y}_1 - a\overline{x}_1$ ou $b = \overline{y}_2 - a\overline{x}_2$.

APPLICATIONS

Exercice 1

Le tableau ci-dessous indique l'évolution de la taille en mètre (m) d'un manguier en fonction du nombre de mois depuis son plantage.

x_i (mois)	1	2	3	4	5	6
y_i (taille en m)	3	4,66	6,32	7,98	9,64	11,30

1. Représenter dans un repère orthonormé $(O; \overrightarrow{i}, \overrightarrow{j})$, le nuage des points $M_i(x_i, y_i)$.
2. La série statistique est partagée en deux séries S_1 et S_2 d'effectifs égaux.
   a) Déterminer les coordonnées des points moyens G_1 et G_2 respectifs des séries S_1 et S_2.
   b) Montrer que l'équation cartésienne de la droite (G_1G_2) est :
   $y = 1,66x + 1,34$.
3. a) Quelle est la taille de l'arbre au moment du plantage.
   b) En combien de mois, l'arbre atteindra-t-il $34,54m$?

Exercice 2

Le tableau suivant donne la statistique portant sur les pourcentage x de des jeunes filles et y des jeunes garçons qui ne poursuit plus les études après le Baccalauréat pendant les dix dernière années au Congo Brazzaville :

x	4	7	8	10	11	12	14	15	9	20
y	2	3	4	6	15	5	10	11	13	16

1. Représenter le nuage des points dans le plan muni d'un repère orthonormé.
2. Donner l'effectif total de cette série.
3. Donner la série marginale de x et y puis calculer $\overline{x}$ et $\overline{y}$.
4. En déduire les coordonnées du point moyen G de ce nuage.
5. Soit (D) : $5x - 6y - 4 = 0$ une droite trouvée par la méthode de Mayer. Estimer le pourcentage des jeunes filles si le pourcentage des jeunes garçons qui ne poursuit plus les études après le Baccalauréat est de 31%

***Exercice* 3**

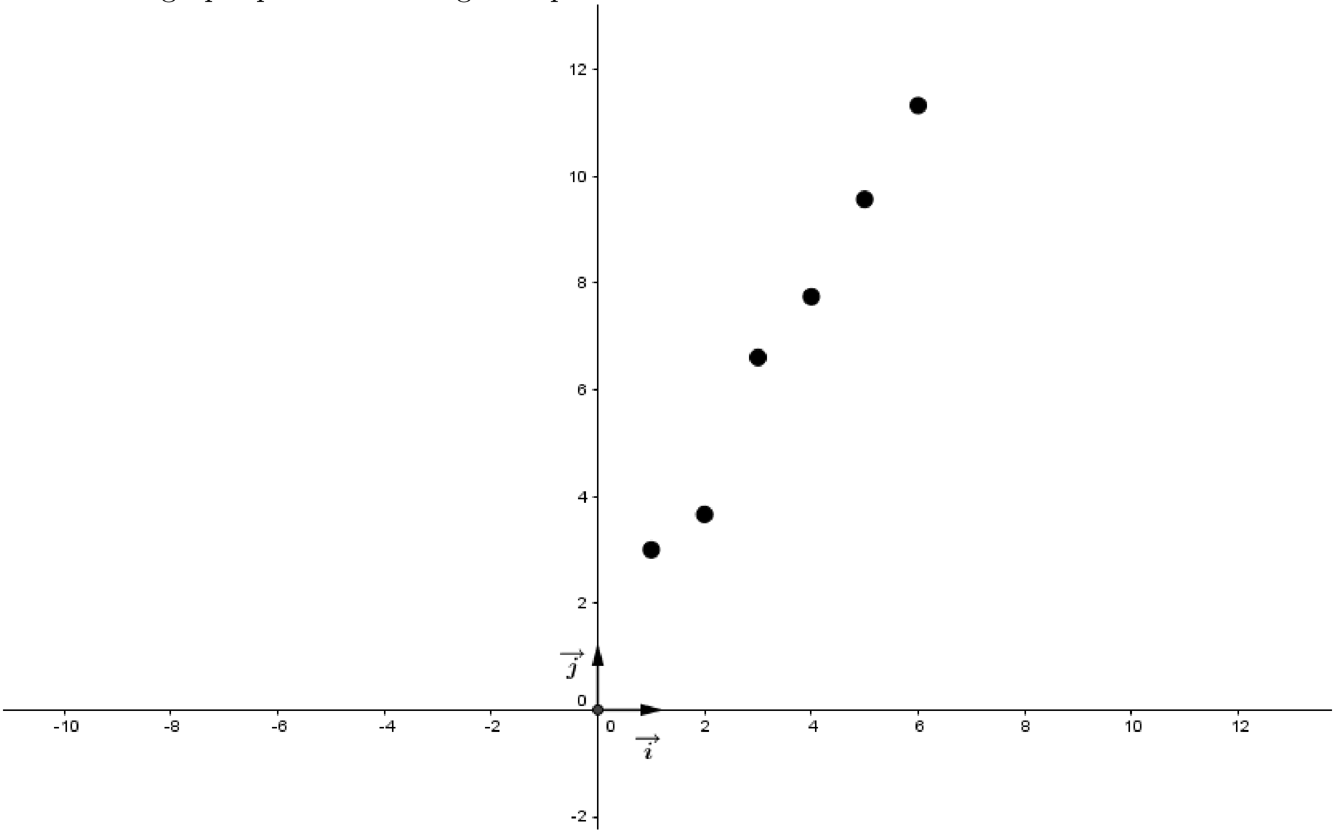
Le tableau suivant donne le poids x en grammes et la taille y en centimètre en fonction du poids d'une population donnée :

Poids x	10	25	40	50	55	60	65	70	75	80
Taille y	11	20	35	45	50	53	60	63	73	75

1. Représenter le nuage des points dans le plan muni d'un repère orthonormé.
   Échelle : $1cm$ pour $10g$ et $1cm$ pour $10cm$.
2. Déterminer le point moyen G de ce nuage.
3. La série ci-dessus est divisée en deux sous-séries S_1 et S_2 de même effectif.
   a) Calculer les coordonnées de G_1 et G_2, points moyens respectifs des sous-séries
   b) Placer les points G_1 et G_2, puis tracer la droite (G_1G_2) dans le repère orthonormé précédent.
   c) Que représente cette droite pour la série étudier ?
   d) Déterminer une équation cartésienne de la droite (G_1G_2).
4. A l'aide de l'équation de la droite (G_1G_2) obtenue, estimer :
   a) La taille d'une personne ayant un poids de $97g$.
   b) Le poids d'une personne de taille $151cm$.

Solution 1

1. Présentons graphiquement le nuage des points.



2. (a) Déterminons les coordonnées des points moyens G_1 et G_2.

▷ Sous série S_1

x_1	1	2	3
n_1	3	4,66	6,32

On a : $\begin{cases} \overline{x_1} = \dfrac{1+2+3}{3} = 2 \\ \overline{y_1} = \dfrac{3+4,66+6,32}{3} = 4,66 \end{cases}$

D'où $G_1(2; 4,66)$

▷ Sous série S_2

x_1	4	5	6
n_1	7,98	9,64	11,30

On a : $\begin{cases} \overline{x_2} = \dfrac{4+5+6}{3} = 5 \\ \overline{y_2} = \dfrac{7,98+9,64+11,30}{3} = 9,64 \end{cases}$

D'où $G_2(5; 9,64)$

(b) Montrons que l'équation cartésienne de la droite (G_1G_2) est : $y = 1,66x + 1,34$

On a : $(G_1G_2) : \dfrac{x-\overline{x_1}}{x_2-\overline{x_1}} = \dfrac{y-\overline{y_1}}{y_2-\overline{y_1}} \Leftrightarrow \dfrac{x-2}{5-2} = \dfrac{y-4,66}{9,64-4,66}$
$\Longrightarrow \dfrac{x-2}{3} = \dfrac{y-4,66}{4,98} \Longrightarrow y = \dfrac{4,98}{3}x + \dfrac{4,02}{3}$
D'où $y = 1,66x + 1,34$

3. (a) Taille de l'arbre au moment du plantage

   $x = 0\,;\ y = 1,34cm$

   (b) Déterminons le mois où l'arbre atteindra $34,54cm$
   $y = 34,54\,;\ x = \dfrac{y-1,34}{1,66} = \dfrac{34,54-1,34}{1,66} = 20$
   D'où $x = 20$ mois

   C'est au bout de 20 mois que l'arbre atteindra $34,54cm$