

Leçon
7

STATISTIQUE À DEUX VARIABLES



SITUATION D'APPRENTISSAGE

Contexte : Des élèves ont découvert sur internet des informations relatives à la lutte contre le réchauffement climatique par la réduction de l'émission de gaz carboniques :

- Un premier « plan climat » a été appliqué par l'Europe dès 2006 pour baisser les émissions de Gaz à Effet de Serre (GES), lequel a donné les résultats suivants :

Rang de l'année à partir de 2006	1	2	3	4	5
Emissions de GES (en Mteq CO ₂)	540	530	527	520	514

- Un nouveau « plan climat » a été élaboré dès 2010.

Circonstance : Les élèves sont soucieux du danger que constitue le phénomène des émissions de gaz à effet de serre.

Tâche(s) : Les élèves souhaitent étudier en classe les données statistiques découvertes relatives à l'évolution des émissions de gaz à effet de serre.

Proposition d'un questionnaire de compréhension :

- Dis ce que les élèves ont découvert sur internet.
- Décris le tableau découvert.
- De quoi sont soucieux les élèves ?
- Que décident faire les élèves ?

Commentaire : Cette leçon vise à faire acquérir aux apprenants des moyens pour étudier les séries statistiques à deux caractères.

INSTALLATION DES HABILITÉS

Dans ce chapitre, on n'étudiera que des séries statistiques doubles dont les deux caractères étudiés sont quantitatifs. Hormis le paragraphe 1-1, cette étude exclura le cas des séries double pondérées.

Activité 1 Série statistique à deux caractères discrets

1.1. Définition

Réponse de l'activité :

- 1) Caractères :
 - Moyenne de Français ;
 - Moyenne de Philosophie.
- 2)

Notes en français	6	7	8	9	10	10,5	11	12	13	15
Notes en philosophie	9	8	10	14	12	15	9	11	8	13

Exercices de fixation

1.1-1) 1 F ; 2 V.

1.1-2)

Nombre d'enfants de la famille	1	2	3	5	7
Nombre de chambre de leur maison	2	4	2	2	3

1.2. Séries marginales d'un tableau à double entrée (Série A1 seulement)

Réponse de l'activité :

- 1) - Caractère 1 :
 - (x_i) : Nombre de filles redoublantes.
- Caractère 2 :
 - (y_i) : Nombre de garçons redoublants.
- 2) 2-a)
 - Le couple (5;3) correspond à la 1^{ère} et 4^e colonne du tableau ;
 - Le couple (2;7) correspond à la 2^e, 3^e, 8^e et 10^e colonne du tableau ;
 - Le couple (6;4) correspond à la 7^e et 9^e colonne du tableau.

- 2-b) • Il y a 2 couples (5;3) ;
 • Il n'y a aucun couple (3;4).

2-c)

Tableau 1 :

y_i	3	4	5	7	10
x_i					
2	0	0	0	4	1
3	0	0	1	0	0
5	2	0	0	0	0
6	0	2	0	0	0

- 3) 3-a) En utilisant le tableau 1, calcule :

- le nombre de couples ayant pour première composante 3 est : $0+0+1+0+0=1$.
- le nombre de couples ayant pour première composante 2 est : $0+0+0+4+1=5$.
- le nombre de couples ayant pour deuxième composante 7 est : $4+0+0+0+0=4$.

3-b)

Tableau 2 :

x_i	2	3	5	6
Effectifs	5	1	2	2

Tableau 3 :

y_i	3	4	5	7	10
Effectifs	2	2	1	4	1

3-b) Rappel : $Fréquence\ de\ x_i = \frac{Effectif\ de\ x_i}{Effectif\ total\ des\ x_i}$.

Tableau 4 :

Effectif total des $x_i = 10$.

y_i	2	3	5	6	2
Fréquences (en %)	50 %	10 %	20 %	20 %	50 %

Tableau 5 :

Effectif total des $y_i=10$.

y_i	3	4	5	7	10
Fréquences (en %)	20 %	20 %	10 %	40 %	10 %

Exercices de fixation

1.2-1

$x_i \backslash y_i$	10	12	20	23	35
3	1	0	0	0	0
4	0	1	0	0	0
5	3	0	0	0	0
7	0	0	0	2	0
9	0	2	0	0	0
10	0	0	0	0	1
11	0	0	1	0	0

1.2-2

1) Série marginale associé à x :

$$0+3+2+11=16 ; 0+1+1+0=2 ; 0+1+0+2=3 ; 1+0+2+3=5.$$

x_i	100	120	150	200
Effectifs	16	2	3	5

2) Série marginale associé à y :

$$0+0+0+1=1 ; 3+1+1+0=5 ; 2+1+0+2=5 ; 11+0+2+3=16.$$

y_i	1,5	3,5	8,2	11
Effectifs	1	5	5	16

3) le tableau fréquences marginales du caractère x :

$$\text{Rappel : Fréquence de } x_i = \frac{\text{Effectif de } x_i}{\text{Effectif total des } x_i}.$$

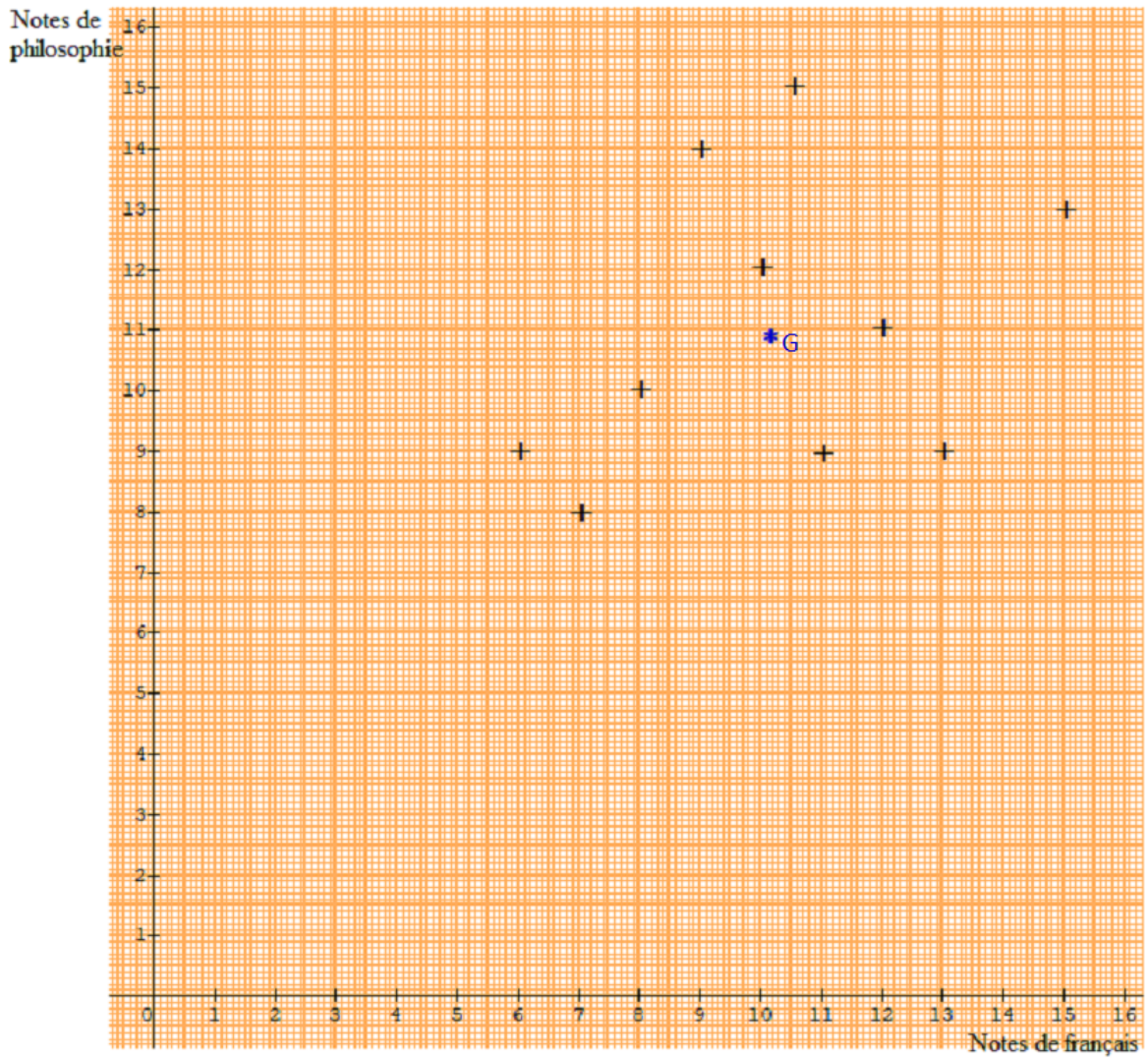
$$\text{Effectif total des } x_i = 16+2+3+5=26.$$

x_i	100	120	150	200
Fréquences	11,54 %	19,23 %	11,54 %	19,23 %

1.3. Nuage de points, point moyen

Réponse de l'activité :

1)



2) 2-a) $\bar{X} = \frac{6+7+8+9+10+10,5+11+12+13+15}{10} = 10,15$.

2-b) $\bar{Y} = \frac{9+8+10+14+12+15+9+11+8+13}{10} = 10,9$.

2-b) G (10,15;10,9) : voir graphique.

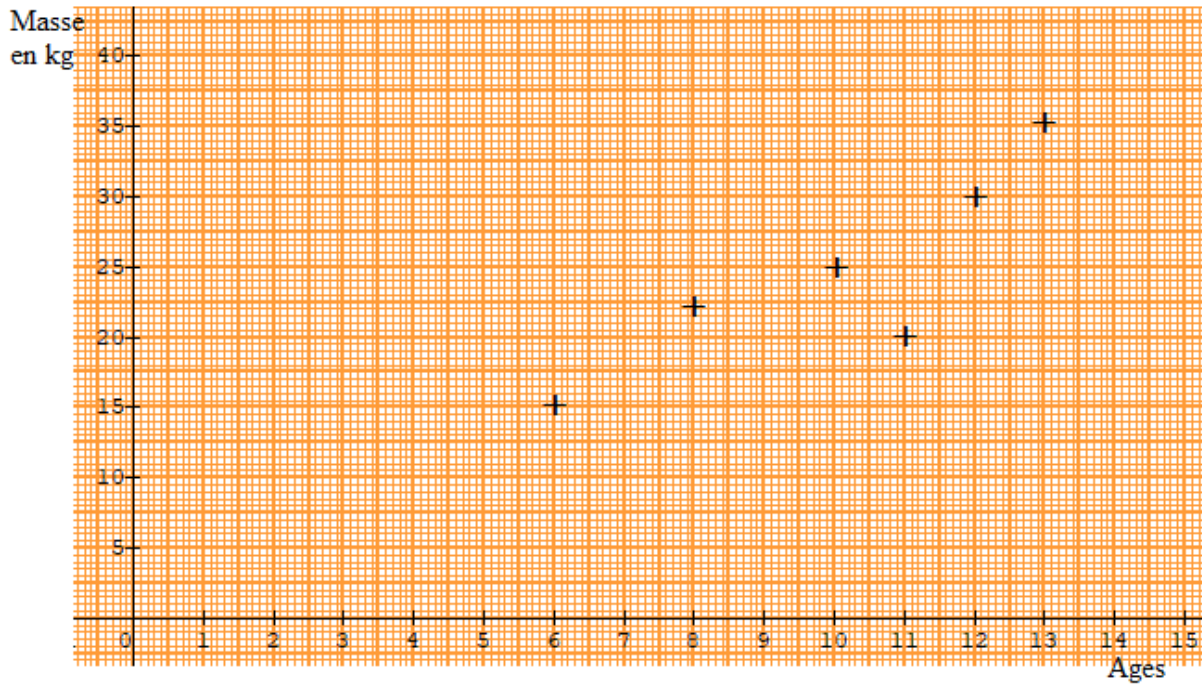
Exercices de fixation

1.3-1 : C)

1.3-2

$\bar{X} = \frac{6+8+10+11+12+13}{6} = 10$ et $\bar{Y} = \frac{15+22+25+20+30+35}{6} = 24,5$ donc G(10 ; 24,5).

1.3-3 Nuage de points :



Activité 2 Ajustement affine d'une série statistique à deux variables

2.1. Ajustement linéaire

NB :

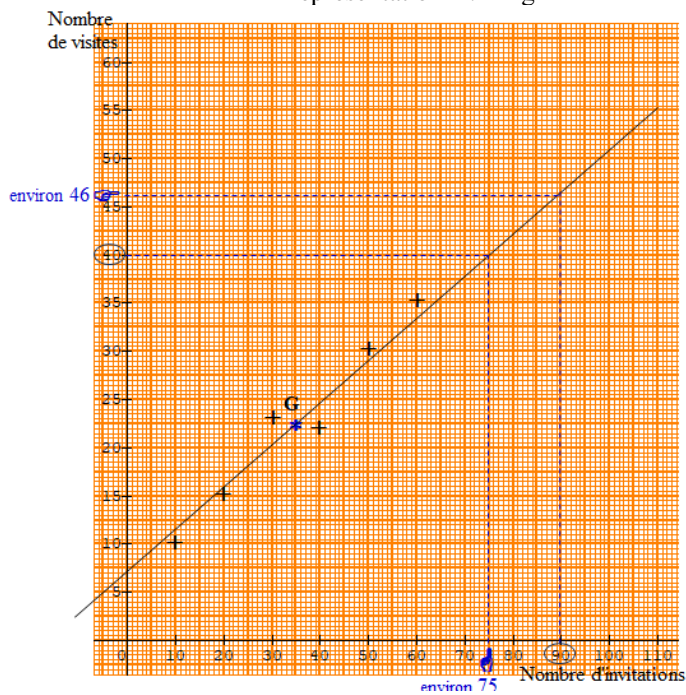
Supprimer la consigne 5).

Dans la consigne 3.b), lire « conjecture » au lieu de « donne ».

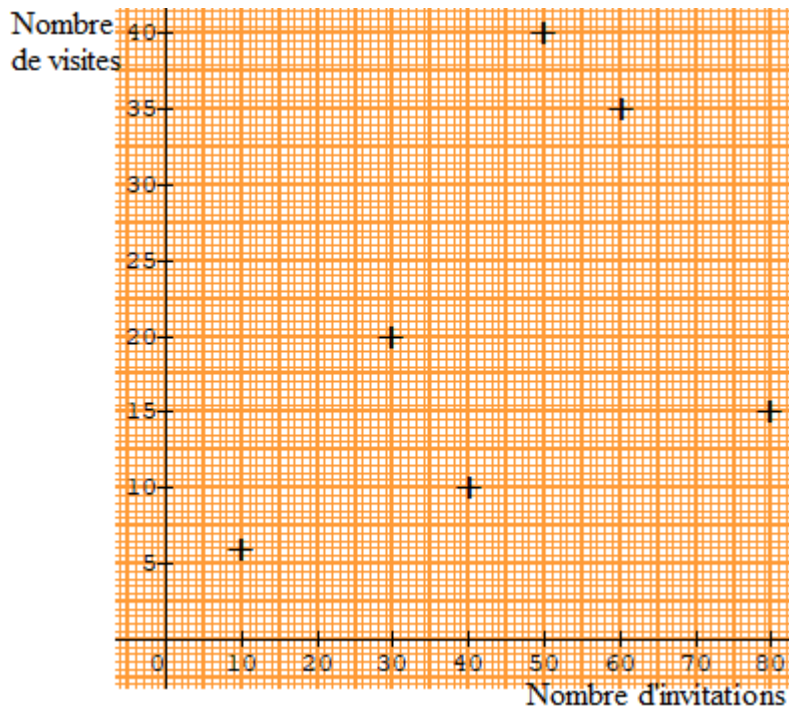
Réponse de l'activité :

1) 1-a)

Représentation 1 : Blog d'Henri



Représentation 2 : Blog d'Angela



1-b)

• Série statistique d'Henri

$$\bar{X} = \frac{10+20+30+40+50+60}{6} = \frac{210}{6} = 35 \text{ et } \bar{Y} = \frac{10+15+23+22+30+35}{6} = \frac{135}{6} = 22,5$$

donc $G(35 ; 22,5)$.

• Série statistique d'Angela

$$\bar{X}' = \frac{10+30+40+50+60+80}{6} = \frac{270}{6} = 45 \text{ et } \bar{Y}' = \frac{6+20+10+40+35+15}{6} = \frac{126}{6} = 21$$

donc $G'(45 ; 21)$.

2) 2-a)

NB : On pourra marquer G et G' ; et essayer la possibilité de tracer une telle droite (D).

C'est le nuage de points correspondant à la série statistique d'Henri qui offre la possibilité de tracer une droite (D) qui passe par G et le plus près possible des points du nuage.

Explication :

Dans le nuage de points correspondant à la série statistique d'Angela, les points du nuage sont très dispersés.

Dans le nuage de points correspondant à la série statistique d'Henri, le nuage de point est allongé et les points semble alignés ; c'est pourquoi ici, il est possible de tracer une droite (D) qui passe par G et le plus près possible des points du nuage.

2-b) Voir représentation 1 ci-dessus.

3) 3-a) C'est dans le cas du nuage de points correspondants à la série statistique du blog d'Henri que l'on peut conjecturer les valeurs pour compléter le tableau proposé.

3-b) Sur la représentation 1, on utilise la droite (D) pour conjecturer les valeurs manquantes du tableau :

- Pour 90 invitations ; on lit environ 46 (ou 45 ou 47 selon le tracé de (D)) visites.
- Pour 40 visites ; on lit environ 75 (ou 74 ou 76 selon le tracé de (D)) invitations.

3-c) Sens des valeurs trouvées :

- Pour 90 invitations envoyés, Henri aura environ **46 visites** sur son blog.
- Pour constater 40 visites sur son blog, Henri doit envoyer environ **75 invitations**.

4) (D) : $y = \frac{13}{30}x + \frac{22}{3}$.

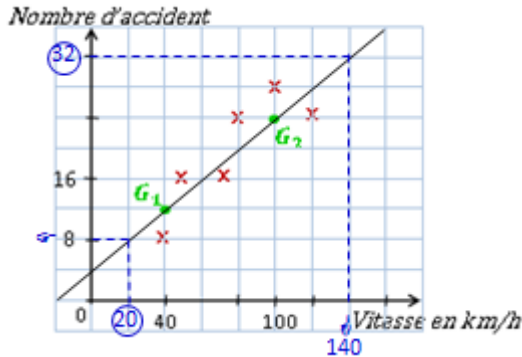
Ici, x désigne le nombre d'invitation et par y le nombre de visites.

- Pour $x=90$, on a : $y = \frac{13}{30} \times 90 + \frac{22}{3} = \frac{139}{3}$ et donc $y \approx 46$.
- Pour $y=40$, on a : $40 = \frac{13}{30}x + \frac{22}{3}$ c'est-à-dire $\frac{13}{30}x = 980$; et donc $x = \frac{980}{13} \approx 75$.

Exercices de fixation

2.1-1

1) Démarche : Le nuage de points est allongé et les points rapprochés ; c'est pourquoi, il est possible de tracer une droite d'ajustement. Or le point moyen G appartient évidemment à la droite (G_1G_2) . On peut donc prendre la droite (G_1G_2) pour droite d'ajustement comme le suggère la donnée des points G_1 et G_2 .



- 1) C'est 8 accidents pour une vitesse de 20 km/h.
- 2) C'est la vitesse 140km/h qui occasionne 32 accidents.

2.1-2

NB : Dans l'énoncé de l'exercice, lire « **la droite (L) d'ajustement** » au lieu de « la droite (L) de Mayer ». (L) : $y = 0,5x + 6$.

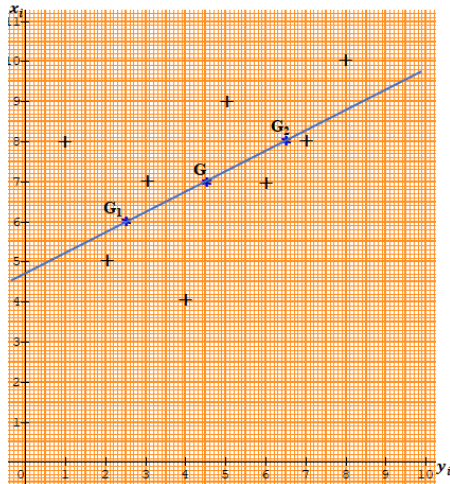
- Pour $y=8$, on a : $8 = 0,5x + 6$ c'est-à-dire $x = 4$.
- Pour $x=16$, on a : $y = 0,5 \times 16 + 6$ c'est-à-dire $y=14$.
- Pour $y=15,5$, on a : $15,5 = 0,5x + 6$ c'est-à-dire $x = 19$.

Superficie exploitée x_i (en hectares)	1	4	6	9	12	14	16	19
Bénéfice y_i (en centaines de milliers de francs CFA.)	7	8	8,9	10,1	12	13	13,5	15,5

2.2. Droite de Mayer

Réponse de l'activité :

1) 1-a)



1-b) Calcule les coordonnées du point moyen G .

$$\bar{X} = \frac{1+2+3+4+5+6+7+8}{8} = \frac{36}{8} = 4,5 \quad \text{et} \quad \bar{Y} = \frac{8+5+7+4+9+7+6+4}{8} = \frac{56}{8} = 7.$$

donc $G(4,5 ; 7)$.

2) 2-a)

Série S_1

x_i	1	2	3	4
y_i	8	5	7	4

Série S_2

x_i	5	6	7	8
y_i	9	7	6	10

2-b)

$$G_1 \left(\frac{1+2+3+4}{4}; \frac{8+5+7+4}{4} \right) \quad \text{et} \quad G_2 \left(\frac{5+6+7+8}{4}; \frac{9+7+6+10}{4} \right)$$

$G_1(2,5; 6)$ et $G_2(6,5; 8)$.

2-c) Voir représentation ci-dessus.

3) 3-a) $\frac{8-6}{6,5-2,5} = \frac{2}{4} = \frac{1}{2}$.

3-b) La droite $(G_1 G_2)$ passe par $G_1(2,5 ; 6)$ et a pour coefficient directeur $\frac{1}{2}$.

Une équation de $(G_1 G_2)$ est :

$$y = \frac{1}{2}(x - x_{G_1}) + y_{G_1} ;$$

$$y = \frac{1}{2}(x - 2,5) + 6 ;$$

$$y = 0,5x + 4,75.$$

4) 4-a) Voir représentation ci-dessus.

4-b) On a $G(4,5 ; 7)$ et $(G_1 G_2)$ a pour équation $y = 0,5x + 4,75$.

L'égalité $7 = 0,5 \times 4,5 + 4,75$ est vraie ; donc le point $G(4,5 ; 7)$ appartient à la droite $(G_1 G_2)$.

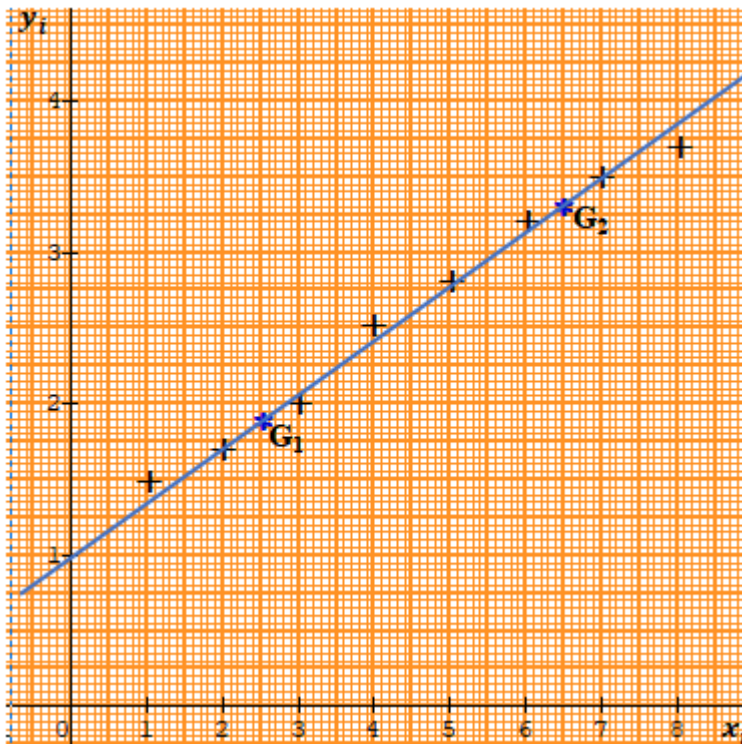
Exercices de fixation

2.2-1

x_i	1	2	3	4	x_i	5	6	7	8
y_i	1,5	1,7	2	2,5	y_i	2,8	3,2	3,5	3,7

$$G_1 \left(\frac{1+2+3+4}{4} ; \frac{1,5+1,7+2+2,5}{4} \right) \text{ et } G_2 \left(\frac{5+6+7+8}{4} ; \frac{2,8+3,2+3,5+3,7}{4} \right)$$

$$G_1(2,5; 1,925) \text{ et } G_2(6,5; 3,3).$$



2.2-2

La droite de Mayer est la droite $(G_1 G_2)$ avec $G_1(2,5; 1,925)$ et $G_2(6,5; 3,3)$.

La droite $(G_1 G_2)$ a pour coefficient directeur : $\frac{3,3-1,925}{6,5-2,5} = \frac{1,375}{4} = 0,34375$.

La droite $(G_1 G_2)$ passe par $G_2(6,5 ; 3,3)$ et a pour coefficient directeur $0,34375$.

Une équation de $(G_1 G_2)$ est :

$$y = 0,34375(x - x_{G_2}) + y_{G_2} ;$$

$$y = 0,34375(x - 6,5) + 3,3 ;$$

$$y = 0,34375x + 1,065625.$$

2.3. Ajustement affine par la méthode des moindres carrés

2.3.1. Variance, Covariance d'une série statistique à deux variables (Série A1 seulement)

NB : Pour que l'activité soit simple et plus compréhensive, l'enseignant fera lire dans l'énoncé « $\bar{y} = 5,10$ » au lieu de « $\bar{y} = 5,08$ » et dans les tableaux « 3,6 » au lieu de « 3,5 ».

Réponse de l'activité :

Taille (en cm) : x_i	50	55	60	65	70	75
Masse (en kg) y_i	3,6	4	5	5,5	6	6,5

NB : Les résultats seront arrondis à 10^{-2} près.

$\bar{x} = 62,5$ et $\bar{y} = 5,10$.

1)

x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
50	3,6	156,25	2,25	18,75
55	4	56,25	1,21	8,25
60	5	6,25	0,01	0,25
65	5,5	6,25	0,16	1
70	6	56,25	0,81	6,75
75	6,5	156,25	1,96	17,5

2) 2-a) Moyenne des $(x_i - \bar{x})^2$.

$$\frac{156,25+56,25+6,25+6,25+56,25+156,25}{6} = \frac{437,5}{6} = 72,92.$$

2-b) Moyenne des $(y_i - \bar{y})^2$.

$$\frac{2,25+1,21+0,01+0,16+0,81+1,96}{6} = \frac{6,4}{6} = 1,07.$$

2-c) Moyenne des $(x_i - \bar{x})(y_i - \bar{y})$.

$$\frac{18,75+8,25+0,25+1+6,75+17,5}{6} = \frac{52,5}{6} = 8,75.$$

3) 3-a)

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
50	3,6	2500	12,96	180
55	4	3025	16	220
60	5	3600	25	300
65	5,5	4225	30,25	357,5
70	6	4900	36	420
75	6,5	5625	42,25	487,5

3-b)

$$\bullet \frac{x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3 + x_4 \cdot y_4 + x_5 \cdot y_5 + x_6 \cdot y_6}{6} - \bar{x} \bar{y} = \frac{180+220+300+357,5+420+487,5}{6} - 62,5 \times 5,10$$

$$= \frac{1965}{6} - 316,71$$

$$= 8,75.$$

$$\bullet \frac{x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2}{6} - \bar{x}^2 = \frac{2500+3025+3600+4225+4900+5625}{6} - 62,5^2$$

$$= \frac{23875}{6} - 3906,25$$

$$= 72,92.$$

3-c) Compare les résultats des questions 2-c) et 3-b), puis 2-a) et 3-b),

La moyenne des $(x_i - \bar{x})(y_i - \bar{y})$ est égale à $\frac{x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3 + x_4 \cdot y_4 + x_5 \cdot y_5 + x_6 \cdot y_6}{6} - \bar{x} \bar{y}$.

La moyenne des $(x_i - \bar{x})^2$ est égale à $\frac{x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2}{6} - \bar{x}^2$.

Exercices de fixation

2.3.2-1

1)

Série 1

x_i	12	20	30	100
y_i	7	1	30	4

Moyenne en x : $\bar{x} = \frac{172}{4} = 43$;

Moyenne en y : $\bar{y} = \frac{42}{4} = 10,5$.

Méthode 1	Méthode 2																																																
<p>Variance de x :</p> $V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$ <p>Variance de y :</p> $V(y) = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n}$ <table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>$(x_i - \bar{x})^2$</th> <th>$(y_i - \bar{y})^2$</th> </tr> </thead> <tbody> <tr> <td>12</td> <td>7</td> <td>961</td> <td>12,25</td> </tr> <tr> <td>30</td> <td>1</td> <td>169</td> <td>90,25</td> </tr> <tr> <td>30</td> <td>30</td> <td>169</td> <td>380,25</td> </tr> <tr> <td>100</td> <td>4</td> <td>3249</td> <td>42,25</td> </tr> <tr> <td>Total</td> <td>172</td> <td>4548</td> <td>525</td> </tr> </tbody> </table> <p>$V(x) = \frac{4548}{4} = 1137$; $V(y) = \frac{525}{4} = 131,25$.</p>	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	12	7	961	12,25	30	1	169	90,25	30	30	169	380,25	100	4	3249	42,25	Total	172	4548	525	<p>Variance de x :</p> $V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$ <p>Variance de y :</p> $V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2$ <table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>x_i^2</th> <th>y_i^2</th> </tr> </thead> <tbody> <tr> <td>12</td> <td>7</td> <td>144</td> <td>49</td> </tr> <tr> <td>30</td> <td>1</td> <td>900</td> <td>1</td> </tr> <tr> <td>30</td> <td>30</td> <td>900</td> <td>900</td> </tr> <tr> <td>100</td> <td>4</td> <td>10000</td> <td>16</td> </tr> <tr> <td>Total</td> <td>172</td> <td>11944</td> <td>966</td> </tr> </tbody> </table> <p>$V(x) = \frac{11944}{4} - 43^2 = 1137$; $V(y) = \frac{966}{4} - 10,5^2 = 131,25$.</p>	x_i	y_i	x_i^2	y_i^2	12	7	144	49	30	1	900	1	30	30	900	900	100	4	10000	16	Total	172	11944	966
x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$																																														
12	7	961	12,25																																														
30	1	169	90,25																																														
30	30	169	380,25																																														
100	4	3249	42,25																																														
Total	172	4548	525																																														
x_i	y_i	x_i^2	y_i^2																																														
12	7	144	49																																														
30	1	900	1																																														
30	30	900	900																																														
100	4	10000	16																																														
Total	172	11944	966																																														

2)

Série 2

x_i	y_i
1	5
2	12
3	7

Moyenne en x : $\bar{x} = \frac{6}{3} = 2$;

Moyenne en y : $\bar{y} = \frac{24}{3} = 8$;

Méthode 1	Méthode 2																																								
<p>Variance de x :</p> $V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$ <p>Variance de y :</p> $V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2$ <table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>x_i^2</th> <th>y_i^2</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>5</td> <td>1</td> <td>25</td> </tr> <tr> <td>2</td> <td>12</td> <td>4</td> <td>144</td> </tr> <tr> <td>3</td> <td>7</td> <td>9</td> <td>49</td> </tr> <tr> <td>Total</td> <td>6</td> <td>14</td> <td>218</td> </tr> </tbody> </table> <p>$V(x) = \frac{14}{3} - 2^2 \approx 0,67$; $V(y) = \frac{218}{3} - 8^2 \approx 8,67$.</p>	x_i	y_i	x_i^2	y_i^2	1	5	1	25	2	12	4	144	3	7	9	49	Total	6	14	218	<p>Variance de x :</p> $V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$ <p>Variance de y :</p> $V(y) = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n}$ <table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>$(x_i - \bar{x})^2$</th> <th>$(y_i - \bar{y})^2$</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>5</td> <td>1</td> <td>9</td> </tr> <tr> <td>2</td> <td>12</td> <td>0</td> <td>16</td> </tr> <tr> <td>3</td> <td>7</td> <td>1</td> <td>1</td> </tr> <tr> <td>Total</td> <td>7</td> <td>2</td> <td>26</td> </tr> </tbody> </table> <p>$V(x) = \frac{2}{3} \approx 0,67$; $V(y) = \frac{26}{3} \approx 8,67$.</p>	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	1	5	1	9	2	12	0	16	3	7	1	1	Total	7	2	26
x_i	y_i	x_i^2	y_i^2																																						
1	5	1	25																																						
2	12	4	144																																						
3	7	9	49																																						
Total	6	14	218																																						
x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$																																						
1	5	1	9																																						
2	12	0	16																																						
3	7	1	1																																						
Total	7	2	26																																						

2.3.2-2

1)

Série 1

x_i	3	5	7	14
y_i	10	2	10	3

Moyenne en x : $\bar{x} = \frac{29}{4} = 7,25$;

Moyenne en y : $\bar{y} = \frac{25}{4} = 6,25$.

Méthode 1	Méthode 2
$\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$	$\text{Cov}(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y}$

	x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$		x_i	y_i	$x_i y_i$
	3	10	-15,938		3	10	30
	5	2	9,563		5	2	10
	7	10	-0,938		7	10	70
	14	3	-21,938		14	3	42
Total	29	25	-29,251	Total	29	25	152

$\text{Cov}(x ; y) = \frac{-29,251}{4} \approx -7,313.$
 $\text{Cov}(x ; y) = \frac{152}{4} - 7,25 \times 6,25 \approx -7,313.$

2)

Série 2

x_i	y_i
2	2
5	8
1	12

Moyenne en x : $\bar{x} = \frac{8}{3} \approx 2,667$;
Moyenne en y : $\bar{y} = \frac{22}{3} \approx 7,333.$

<p>Méthode 1</p> <p>$\text{Cov}(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y} .$</p> <table border="1" style="margin-left: 40px;"> <tr><td>x_i</td><td>y_i</td><td>$x_i y_i$</td></tr> <tr><td>2</td><td>2</td><td>4</td></tr> <tr><td>5</td><td>8</td><td>40</td></tr> <tr><td>1</td><td>12</td><td>12</td></tr> <tr><td style="border: 1px solid black;">Total</td><td style="border: 1px solid black;">8</td><td style="border: 1px solid black;">22</td><td style="border: 1px solid black;">56</td></tr> </table> <p>$\text{Cov}(x ; y) = \frac{56}{3} - 2,667 \times 7,333 \approx -0,890.$</p>	x_i	y_i	$x_i y_i$	2	2	4	5	8	40	1	12	12	Total	8	22	56	<p>Méthode 2</p> <p>$\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$</p> <table border="1" style="margin-left: 40px;"> <tr><td>x_i</td><td>y_i</td><td>$(x_i - \bar{x})(y_i - \bar{y})$</td></tr> <tr><td>2</td><td>2</td><td>3,557</td></tr> <tr><td>5</td><td>8</td><td>1,556</td></tr> <tr><td>1</td><td>12</td><td>-7,78</td></tr> <tr><td style="border: 1px solid black;">Total</td><td style="border: 1px solid black;">8</td><td style="border: 1px solid black;">22</td><td style="border: 1px solid black;">-2,667</td></tr> </table> <p>$\text{Cov}(x ; y) = \frac{-2,667}{3} \approx -0,889.$</p>	x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$	2	2	3,557	5	8	1,556	1	12	-7,78	Total	8	22	-2,667
x_i	y_i	$x_i y_i$																															
2	2	4																															
5	8	40																															
1	12	12																															
Total	8	22	56																														
x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$																															
2	2	3,557																															
5	8	1,556																															
1	12	-7,78																															
Total	8	22	-2,667																														

2.3.2. Équation de la droite de régression (Série A1 seulement)

Réponse de l'activité (Série A1 seulement):

1) 1-a) $a = 0,12$ (coefficient directeur de la droite d'ajustement (D)).

1-b) Justifions que $b = \bar{y} - a\bar{x}$:

a et b désignant respectivement le coefficient directeur et l'ordonnée à l'origine de (D), une équation de (D) est $y = ax + b$.

Désignons par G le point moyen du nuage de points ; on a : $G(\bar{x} ; \bar{y})$.

Comme toute droite d'ajustement passe par le point moyen du nuage de points, la droite d'ajustement (D) passe par G. Ainsi, on a : $b = \bar{y} - a\bar{x}$.

2) Sachant que $V(x) = 72,92$ et $\text{Cov}(x ; y) = 8,86$, on a : $\frac{\text{Cov}(x,y)}{V(x)} = \frac{8,86}{72,92} \approx 0,12.$

3) Il semble que : le coefficient directeur de la droite d'ajustement (D) est égal au quotient $\frac{\text{Cov}(x,y)}{V(x)}$.

Exercices de fixation

2.3.3-1

1) La droite de régression (L_1) d'équation de la forme $x = \alpha y + \beta$ est la droite de régression de x en y ; elle est telle que : $\alpha = \frac{\text{Cov}(x;y)}{V(y)}$ et $\beta = \bar{x} - \alpha\bar{y}$.

On a : $\alpha = \frac{18}{200} = 0,09$ et $\beta = 4 - 0,09 \times 30 = 1,3.$

2) La droite de régression (L_2) d'équation de la forme $y = ux + v$ est la droite de régression de y en x ; elle est telle que : $u = \frac{\text{Cov}(x;y)}{V(x)}$ et $v = \bar{y} - u\bar{x}$.

On a : $u = \frac{18}{2} = 9$ et $v = 30 - 9 \times 4 = -6.$

2.3.3-2

1) Calcul de a coefficient directeur de la droite de régression de y en x :

On a : $a = \frac{Cov(x;y)}{V(x)}$.

Méthode 1					Méthode 2				
	x_i	y_i	x_i^2	$x_i y_i$		x_i	y_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
	0	1	0	0		0	1	4	4,8
	1	3	1	3		1	3	1	0,4
	2	4	4	8		2	4	0	0
	3	5	9	15		3	5	1	1,6
	4	4	16	16		4	4	4	1,2
Total	10	17	30	42	Total	10	17	10	8

Moyenne en x : $\bar{x} = \frac{10}{5} = 2$;
Moyenne en y : $\bar{y} = \frac{17}{5} = 3,4$.
Variance de x :
 $V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$.
 $V(x) = \frac{30}{5} - 2^2 = 2$.
Covariance :
 $Cov(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y}$
 $Cov(x ; y) = \frac{42}{5} - 2 \times 3,4 = 1,6$.

Moyenne en x : $\bar{x} = \frac{10}{5} = 2$;
Moyenne en y : $\bar{y} = \frac{17}{5} = 3,4$.
Variance de x :
 $V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$
 $V(x) = \frac{10}{5} = 2$.
Covariance :
 $Cov(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$
 $Cov(x ; y) = \frac{8}{5} = 1,6$.

Ainsi : $a = \frac{1,6}{2} = 0,8$.

2) Calcul de b ordonnée à l'origine de la droite de régression de y en x :

On a : $b = \bar{y} - a\bar{x}$;

$b = 3,4 - 0,8 \times 2$

$b = 1,8$.

2.3.3-3

Cet exercice vise déterminer les droites de régressions : (L) de y en x et (Δ) de x en y .

Démarche :

Pour (L) droite de régression de y en x , il suffit de calculer a son coefficient directeur et b son ordonnée à l'origine.

Pour (Δ) droite de régression de x en y , il suffit de calculer a' son coefficient directeur et b' son ordonnée à l'origine.

En effet, on a :

$a = \frac{Cov(x;y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$; $a' = \frac{Cov(x;y)}{V(y)}$ et $b' = \bar{x} - a'\bar{y}$.

Moyenne en x : $\bar{x} = \frac{120}{4} = 30$;

Moyenne en y : $\bar{y} = \frac{2\,000}{4} = 500$.

Variance de x : $V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$.

$V(x) = \frac{4350}{4} - 30^2 = 187,5$.

Variance de y : $V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2$.

$V(y) = \frac{1\,225\,000}{4} - 500^2 = 56\,250$.

Covariance : $Cov(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y}$.

$Cov(x ; y) = \frac{47\,250}{4} - 30 \times 500 = -3187,5$.

1) On a : $a = \frac{Cov(x;y)}{V(x)} = \frac{-3\,187,5}{187,5} = -17$ et $b = \bar{y} - a\bar{x} = 500 - (-17) \times 30 = 1\,010$.

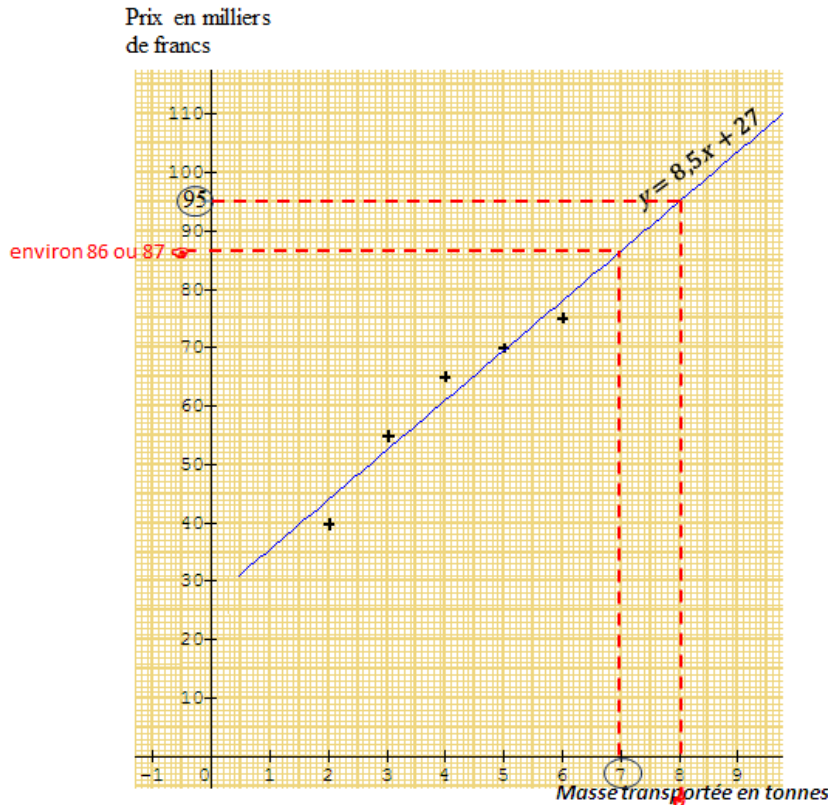
Une équation de la droite de régression (L) de y en x est :
 $y = ax + b$ c'est-à-dire $y = -17x + 1\,010$.

2) On a : $a' = \frac{Cov(x;y)}{V(y)} = \frac{-3\,187,5}{56\,250} \approx -0,057$ et $b' = \bar{x} - a'\bar{y} = 30 - (-0,057) \times 500 = 58,5$.

Une équation de la droite de régression (Δ) de x en y est :
 $x = a'y + b'$ c'est-à-dire $x = -0,057y + 58,5$.

2.3.3-4

1)



1-a) Environ 86 ou 87.

1-b) 8 :

2) 2-a) Pour $x = 10$ t, on a : $y = 8,5 \times 10 + 27 = 112$ en milliers de francs.

Le coût du transport de 10 tonnes de marchandise est 112 000 francs.

2-b) Pour $y = 146$ en milliers de francs, on a : $146 = 8,5x + 27$, c'est-à-dire $x = 14$ en tonnes.

La masse de marchandise à transporter pour un coût de 146 000 francs est 14 tonnes.

2.3.3-5

Construction (Δ) : $y = -17x + 1\,010$.

Pour $x = 10$, on a $y = -17 \times 10 + 1\,010 = 840$.

Pour $x = 50$, on a $y = -17 \times 50 + 1\,010 = 160$.

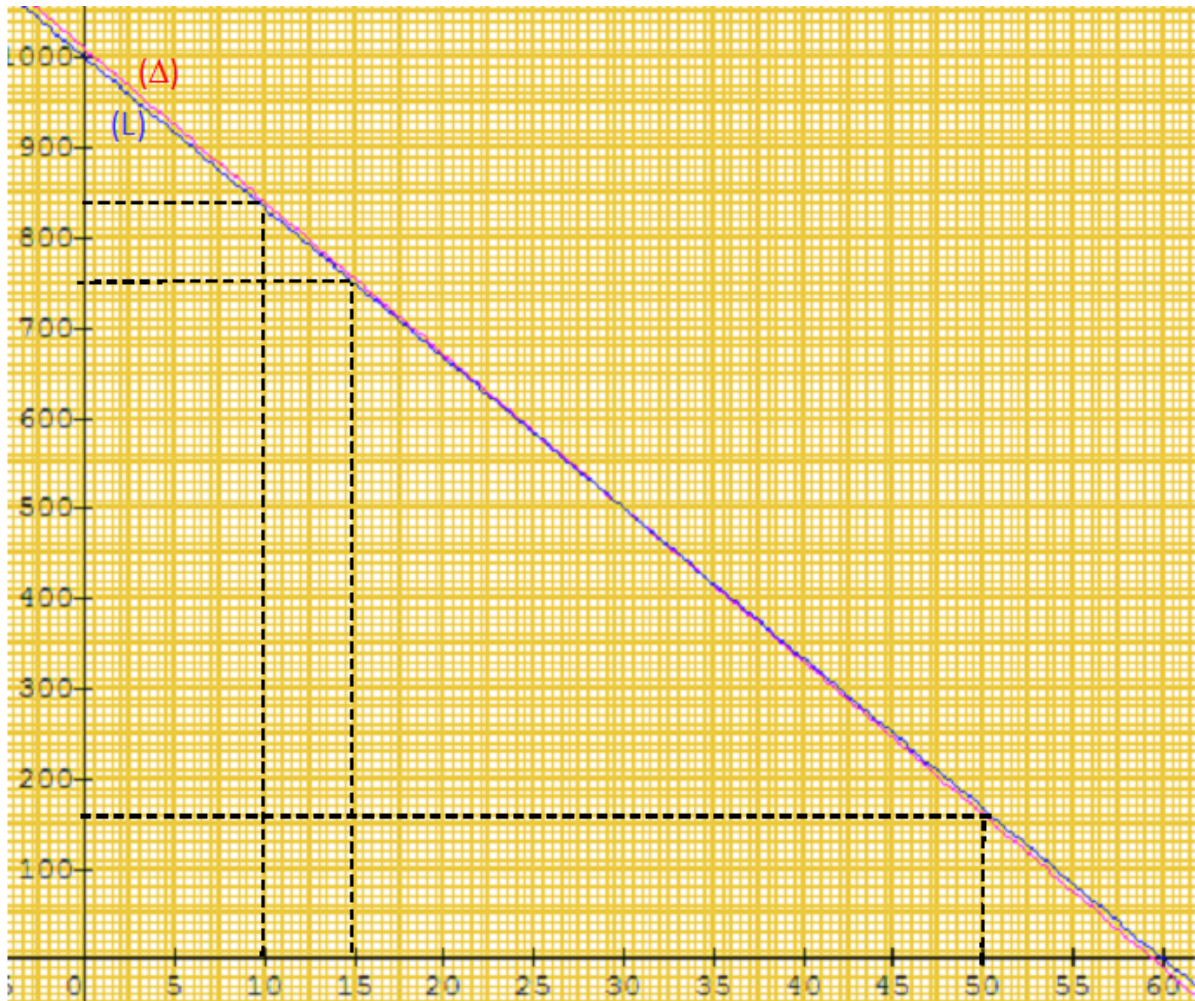
(Δ) passe par les points de coordonnées (10 ; 840) et (50 ; 160),

Construction (L) : $x = -0,06y + 60$.

Pour $x = 15$, on a : $15 = -0,06y + 60$, c'est-à-dire $y = 750$.

Pour $x = 60$, on a : $60 = -0,06y + 60$, c'est-à-dire $y = 0$.

(L) passe par les points de coordonnées (15 ; 750) et (60 ; 0),



Activité 3 Coefficient de corrélation (Série A1 seulement)

Réponse de l'activité :

1) $a = \frac{\text{Cov}(x,y)}{V(x)}$ et avec $a' = \frac{\text{Cov}(x,y)}{V(y)}$.

2) $aa' = \frac{\text{Cov}(x,y)}{V(x)} \times \frac{\text{Cov}(x,y)}{V(y)} = \frac{(\text{Cov}(x,y))^2}{V(x) \cdot V(y)} = \frac{(\text{Cov}(x,y))^2}{\sqrt{V(x) \cdot V(y)}^2} = \left(\frac{\text{Cov}(x,y)}{\sqrt{V(x) \cdot V(y)}} \right)^2$.

3) Posons $r = \frac{\text{Cov}(x,y)}{\sqrt{V(x) \cdot V(y)}}$;

3-a) Pour $r=1$:

- $aa'=1$.

- La droite (D) d'équation $y = ax + ba$ pour coefficient directeur a .

L'équation de (D') peut s'écrire $x = a'y + b'$ ou encore $y = \frac{1}{a'}x - \frac{b'}{a'}$.

La droite (D') a pour coefficient directeur $\frac{1}{a'}$.

De $aa'=1$, on tire que $\frac{1}{a'} = a$.

Donc La droite (D') a pour coefficient directeur a .

Ainsi (D) et (D') ayant le même coefficient directeur, sont parallèles.

De plus les droites d'ajustement (D) et (D') passe par le point moyen G.

(D) et (D') sont finalement confondues.

- Appréciation de l'ajustement affine : l'ajustement affine est parfait.

3-b) On prend r très proches de 1, par exemple dans l'intervalle $[0,87 ; 1[$.

Dans ce cas, les droites (D) et (D') sont très proches. Le lien entre les deux caractères est très fort.

L'ajustement linéaire peut être qualifié bon.

Exercices de fixation

3-1 Coefficient de corrélation linéaire :

$$r = \frac{\text{Cov}(x;y)}{\sqrt{V(x)V(y)}} = \frac{-5,25}{\sqrt{7,5 \times 3,6875}} = \frac{-5,25}{\sqrt{27,65625}} \approx -0,998.$$

3-2 Coefficient de corrélation linéaire : $r = \frac{\text{Cov}(x;y)}{\sqrt{V(x)V(y)}}$.

On a :

$$\text{Variance de } x : V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

$$V(x) = \frac{7,5}{4} = 1,875.$$

$$\text{Variance de } y : V(y) = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n}$$

$$V(y) = \frac{130,75}{4} = 32,6875.$$

$$\text{Covariance: } \text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$$

$$\text{Cov}(x ; y) = \frac{24,5}{4} = 6,125.$$

$$\text{Donc le coefficient de corrélation linéaire vaut : } r = \frac{6,125}{\sqrt{1,875 \times 32,6875}} = \frac{6,125}{\sqrt{61,2890625}} \approx 0,782.$$

3-3

Série A : on a $r = \frac{4}{5} = 0,8$; ainsi $|r| < 0,87$.

Interprétation du coefficient de corrélation linéaire :

« Il y a une **mauvaise corrélation** entre l'évolution de l'âge et la taille de la plante A. »

Série B : on a $r = 1$.

Interprétation du coefficient de corrélation linéaire :

L'ajustement affine est dite parfaite, on dit qu'il y a une parfaite corrélation linéaire entre l'évolution de l'âge et la taille de la plante B.

Série C : on a $r = -\frac{8}{9} \approx -0,888$; ainsi $0,87 \leq |r| \leq 1$.

Interprétation du coefficient de corrélation linéaire :

« il y a une **bonne corrélation** ou **une forte corrélation** entre l'évolution de l'âge et la taille de la plante C »

Exercices de renforcement

1

n°3

2

1-C ; 2-A ; 3-C ; 4-B .

3

1) Caractères étudiés et leurs effectifs :

• Altitude atteint : 6.

• Température extérieure : 06.

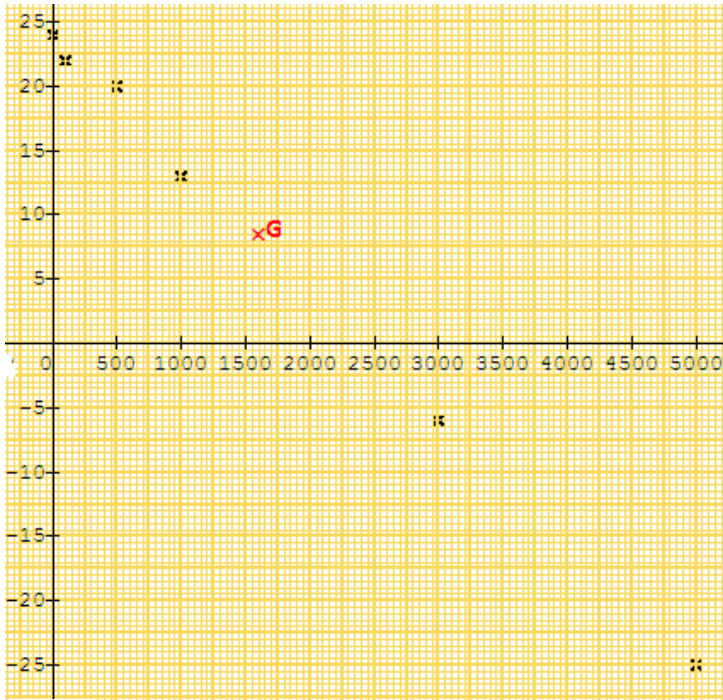
2) Désignons par x le caractère « Altitude atteint » et par y le caractère « Température extérieure ».

Les moyennes sont :

$$\bar{x} = \frac{0+100+500+1\,000+3\,000+5\,000}{6} = \frac{9\,600}{6} = 1\,600.$$

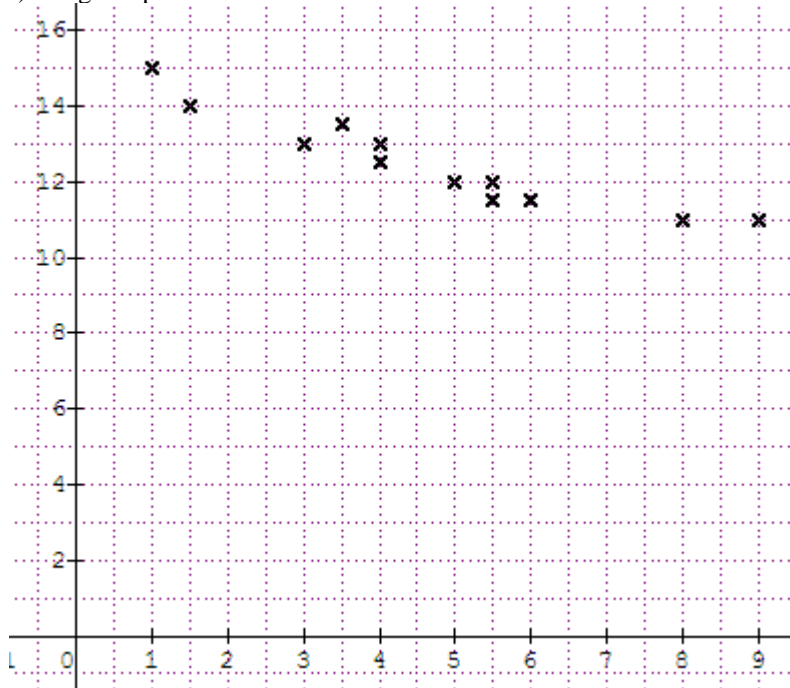
$$\bar{y} = \frac{24+22+20+13+(-6)+(-25)}{6} = \frac{48}{6} = 8.$$

3) Nuage de points et point moyen G.



4

1) Nuage de points :



2)

	Série 1						Total
x	1	2,5	3	3,5	4	4	18
y	15	14	13	13,5	13	12,5	81

	Série 2						Total
x	5	5,5	5,5	6	8	9	39
y	12	11,5	12	11,5	11	11	69

$$x_{G_1} = \frac{18}{6} = 3 \text{ et } y_{G_1} = \frac{81}{6} = 13,5.$$

Donc $G_1(3; 13,5)$.

$$x_{G_2} = \frac{39}{6} = 6,5 \text{ et } y_{G_2} = \frac{69}{6} = 11,5.$$

Donc $G_2(6,5; 11,5)$.

3) Une équation de la droite (G_1G_2) est de la forme : $y = ax + b$.

La droite (G_1G_2) a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{11,5 - 13,5}{6,5 - 3} = \frac{-2}{3,5} = -\frac{4}{7} \approx -0,57$.

La droite (G_1G_2) passe par G_1 , donc : $13,5 = -\frac{4}{7} \times 3 + b$; d'où $b = \frac{213}{14}$.

Une équation de la droite (G_1G_2) est $y = -\frac{4}{7}x + \frac{213}{14}$; avec les arrondis d'ordre 2, on obtient $y = -0,57x + 15,21$.

4) Droite de Mayer.

5

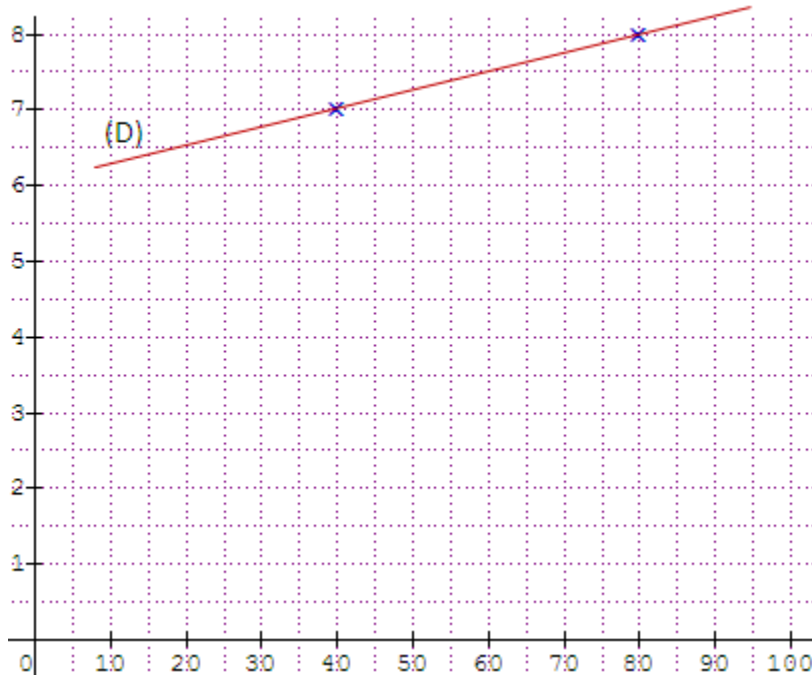
1) Construction de la droite de Mayer.

Série 1				Total
x_i	20	40	60	120
y_i	4	12	5	21

$x_{G_1} = \frac{120}{3} = 40$ et $y_{G_1} = \frac{21}{3} = 7$.
Donc $G_1(40; 7)$.

Série 2				Total
x_i	70	80	90	240
y_i	8	10	7	24

$x_{G_2} = \frac{240}{3} = 80$ et $y_{G_2} = \frac{24}{3} = 8$.
Donc $G_2(80; 8)$.



2) Une équation de la droite de Mayer, c'est à de la droite (G_1G_2) est de la forme : $y = ax + b$.

La droite (G_1G_2) a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{8 - 7}{80 - 40} = \frac{1}{40} = 0,025$.

La droite (G_1G_2) passe par $G_1(40; 7)$, donc : $7 = 0,025 \times 40 + b$; d'où $b = 6$.

Une équation de la droite (G_1G_2) est $y = 0,025x + 6$.

6

1) Calcul des coordonnées du point moyen G.

											Total
x_i	2	3	5	7	9	10	12	14	16	20	98
y_i	2	5	3	7	10	10	4	8	5	1	55

$\bar{x} = \frac{98}{10} = 9,8$ et $\bar{y} = \frac{55}{10} = 5,5$; donc $G(9,8 ; 5,5)$.

2) Equation de la droite de Mayer :

Série 1						Total
x_i	2	3	5	7	9	26
y_i	2	5	3	7	10	27

Série 2						Total
x_i	10	12	14	16	20	72
y_i	10	4	8	5	1	28

$x_{G_1} = \frac{26}{5} = 5,2$ et $y_{G_1} = \frac{27}{5} = 5,4$; $x_{G_2} = \frac{72}{5} = 14,4$ et $y_{G_2} = \frac{28}{5} = 5,6$;

donc $G_1(5,2; 5,4)$.

donc $G_2(14,4 ; 5,6)$.

Une équation de la droite de Mayer, c'est à de la droite (G_1G_2) est de la forme : $y = ax + b$.

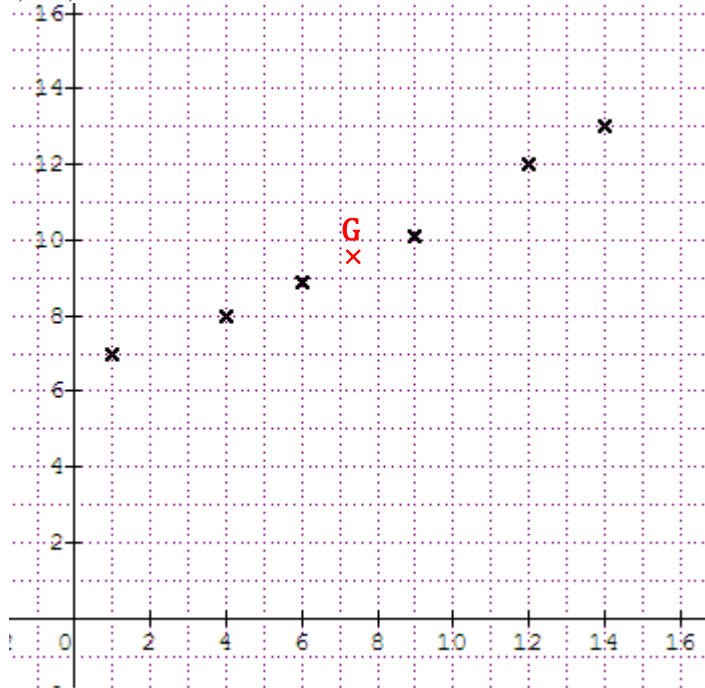
La droite (G_1G_2) a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{5,6 - 5,4}{14,4 - 5,2} = \frac{0,2}{9,2} = \frac{1}{46} \approx 0,02$.

La droite (G_1G_2) passe par $G_1(5,2; 5,4)$, donc : $5,4 = 0,02 \times 5,2 + b$; d'où $b = -5,296$.

Une équation de la droite (G_1G_2) est $y = 0,02x - 5,296$.

7

1) 1-a)



1-b) Démarche : On calcule d'abord les coordonnées du point G, puis on le place dans le repère.

$$\bar{x} = \frac{1+4+6+9+12+14}{6} = \frac{46}{6} = \frac{23}{3} \text{ et } \bar{y} = \frac{7+8+8,9+10,1+12+13}{6} = \frac{59}{6}; \text{ donc } G\left(\frac{23}{3}, \frac{59}{6}\right).$$

2) 2-a) Le nuage de point est allongé et les points semblent alignés, c'est pourquoi il est possible de faire un ajustement affine de la série statistique.

2-b) Equation de la droite d'ajustement (D) de Mayer :

Série 1			
x_i	1	4	6
y_i	7	8	8,9

Série 2			
x_i	9	12	14
y_i	10,1	12	13

Série 1 : $x_{G_1} = \frac{1+4+6}{3} = \frac{11}{3} = 3,67$ et $y_{G_1} = \frac{7+8+8,9}{3} = \frac{23,9}{3} = 7,97$; donc $G_1\left(\frac{11}{3}, \frac{23,9}{3}\right)$.

Série 2 : $x_{G_2} = \frac{9+12+14}{3} = \frac{35}{3} = 11,67$ et $y_{G_2} = \frac{10,1+12+13}{3} = \frac{35,1}{3} = 11,7$; donc $G_2\left(\frac{35}{3}, \frac{117}{10}\right)$.

Une équation de la droite de Mayer, c'est-à-dire de la droite (G_1G_2) est de la forme : $y = ax + b$.

La droite (G_1G_2) a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{\frac{117}{10} - \frac{23,9}{3}}{\frac{35}{3} - \frac{11}{3}} = \frac{\frac{117}{30} - \frac{239}{30}}{\frac{24}{3}} = \frac{\frac{112}{30}}{8} = \frac{112}{30} \times \frac{1}{8} = \frac{7}{15}$.

La droite (G_1G_2) passe par $G\left(\frac{23}{3}, \frac{59}{6}\right)$, donc b vérifie : $\bar{y} = a\bar{x} + b$, c'est-à-dire $b = \bar{y} - a\bar{x}$.

On obtient $b = \frac{59}{6} - \frac{7}{15} \times \frac{23}{3} = \frac{885}{90} - \frac{7 \times 23 \times 2}{15 \times 3 \times 2} = \frac{563}{90}$.

Une équation de la droite d'ajustement (G_1G_2) , c'est-à-dire de la droite (D) de Mayer, est :

$$y = \frac{7}{15}x + \frac{563}{90}$$

8

(A1 seulement):

Expression	Objet calculé
$\frac{8 + 12 + 6 + 10}{4}$	moyenne de y
$\frac{6 + 10 + 5 + 6}{4}$	moyenne de x
$\frac{0,75 + 9,75 + 5,25 + 0,75}{4}$	Covariance
$\frac{48+120+30+60}{4} - 6,75 \times 9$	Covariance

$\frac{0,5625 + 14,0625 + 3,0625 + 0,5625}{4}$	variance de x
$\frac{1 + 9 + 9 + 1}{4}$	variance de y
$\frac{64+144+34+100}{4} - 9^2$	variance de y
$\frac{36+100+25+36}{4} - 6,75^2$	variance de x

9

(A1 seulement)

1) Utilisons la formule : $V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$.

Moyenne de x : $\bar{x} = \frac{1+2+8+9}{4} = \frac{20}{4} = 5$;

Variance de x : $V(x) = \frac{1^2+2^2+8^2+9^2}{4} - 5^2 = \frac{150}{4} - 25 = 5$.

2) Utilisons la formule : $Cov(x ; y) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} - \bar{x}\bar{y}$.

Moyenne de y : $\bar{y} = \frac{10+(-1)+(-5)+5}{4} = \frac{9}{4} = 2,25$;

$Cov(x ; y) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} - \bar{x}\bar{y}$
 $= \frac{1 \times 10 + 2 \times (-1) + 8 \times (-5) + 9 \times 5}{4} - 5 \times 2,25$
 $= \frac{13}{4} - 5 \times 2,25$

$Cov(x ; y) = -8$.

10

(A1 seulement)

1) $\bar{x} = \frac{8+9+10+12}{4} = \frac{39}{4} = 9,75$ et $\bar{y} = \frac{3,5+5+5+3,5}{4} = \frac{17}{4} = 4,25$.

2)

x_i	y_i	x_i^2	y_i^2	$x_i y_i$	
8	3,5	64	12,25	28	
9	5	81	25	45	
10	5	100	25	50	
12	3,5	144	12,25	42	
Total	39	17	389	74,5	165

Moyenne de x : $\bar{x} = \frac{39}{4} = 9,75$;

Moyenne de y : $\bar{y} = \frac{17}{4} = 4,25$;

Variance de x : $V(x) = \frac{389}{4} - 9,75^2 = 2,1875$;

Variance de y : $V(y) = \frac{74,5}{4} - 4,25^2 = 0,5625$.

3) $Cov(x ; y) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} - \bar{x}\bar{y}$

$= \frac{165}{4} - 9,75 \times 4,25$

$= -0,1875$.

4) Le coefficient de corrélation linéaire : $r = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} = \frac{-0,1875}{\sqrt{2,1875 \times 0,5625}} \approx -0,169$.

$|r| \notin [0,87 ; 1]$: il y a une mauvaise corrélation entre les caractères x et y.

11

(A1 seulement)

1) Détermination d'une équation de la droite de régression de y en x.

Démarche : Nous allons déterminer par la méthode des moindres carrés, l'équation sous la forme : $y = ax + b$

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

On a :

• $a = \frac{Cov(x,y)}{V(x)} = \frac{1935081,01}{1387360,55} \approx 1,39$.

• $b = \bar{y} - a\bar{x} = 3105,44 - 1,39 \times 5564,81 = -4629,6459 \approx -4629,65$.

• Equation de la droite de régression de y en x : $x = 1,39y - 4629,65$

2) Détermination d'une équation de la droite de régression de x en y.

Démarche : Nous allons déterminer par la méthode des moindres carrés, l'équation de (Δ) sous la forme : $x = a'y + b'$ avec $a' = \frac{Cov(x;y)}{V(y)}$ et $b' = \bar{x} - a'\bar{y}$.

On a :

• $a' = \frac{Cov(x;y)}{V(y)} = \frac{1571718,61}{1387360,55} \approx 1,13$.

• $b' = \bar{x} - a'\bar{y} = 5564,81 - 1,13 \times 3105,44 = 2055,6628 \approx 2055,66$.

• Equation de la droite de régression de x en y : $x = 1,13y + 2055,66$

3) Le coefficient de corrélation linéaire :

$r = \frac{Cov(x;y)}{\sqrt{V(x)V(y)}} = \frac{1571718,61}{\sqrt{1935081,01 \times 1387360,55}} \approx 0,96$.

12

(A1 seulement):

1)

x_i	y_i	x_i^2	y_i^2	$x_i y_i$	
1	4,4	1	19,36	4,4	
2	5,2	4	27,04	10,4	
3	4,3	9	18,49	12,9	
4	3,2	16	10,24	12,8	
5	3,3	25	10,89	16,5	
6	2,8	36	7,84	16,8	
Total	21	23,2	91	93,86	73,8

Moyenne de x : $\bar{x} = \frac{21}{6} = 3,5$;

Moyenne de y : $\bar{y} = \frac{23,2}{6} = 3,8667$;

Variance de x : $V(x) = \frac{91}{6} - 3,5^2 = 2,9167$.

Variance de y : $V(y) = \frac{93,86}{6} - 3,8667^2 = 0,692$.

Covariance : $Cov(x,y) = \frac{73,8}{6} - 3,5 \times 3,8667 = -1,2335$.

Le coefficient de corrélation linéaire : $r = \frac{Cov(x;y)}{\sqrt{V(x)V(y)}} = \frac{-1,2335}{\sqrt{2,9167 \times 0,692}} \approx -0,8680$.

2) $|r| \notin [0,87 ; 1]$: il y a une mauvaise corrélation entre les caractères x et y .

13

NB : Pour chacun des tableaux proposés dans l'énoncé, lire la consigne : « Détermine les coordonnées du point moyen G ».

On peut présenter la rédaction de la solution de deux manières.

a)

1^{ère} façon

$\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$

et $\bar{y} = \frac{24,5+24,9+25+25,7+25,7}{5} = \frac{125,8}{5} = 25,16$;

donc $G(3; 25,16)$.

2^e façon

x_i	1	2	3	4	5	Total
y_i	24,5	24,9	25	25,7	25,7	125,8

$\bar{x} = \frac{15}{5} = 3$ et $\bar{y} = \frac{125,8}{5} = 25,16$;

donc $G(3 ; 25,16)$.

b)

1^{ère} façon

$\bar{x} = \frac{40+36+32+28+24+20+16+12}{8} = \frac{208}{8} = 26$

et $\bar{y} = \frac{6,5+7+13+21+24+34+39+42}{8} = \frac{186}{8} = 23,25$;

donc $G(26; 23,25)$.

2^e façon

x_i	40	36	32	28	24	20	16	12	Total
y_i	6	7	13	21	24	34	39	42	186

$\bar{x} = \frac{208}{8} = 26$ et $\bar{y} = \frac{186}{8} = 23,25$;

donc $G(26; 23,25)$.

c)

1^{ère} façon

2^e façon

$$\bar{x} = \frac{40+36+32+28+24+20+16+12}{5} = \frac{1050}{6} = 175$$

$$\text{et } \bar{y} = \frac{6,5+7+13+21+24+34+39+42}{5} = \frac{105}{6} = 17,5;$$

donc $G(175; 17,5)$.

								Total
x_i	50	100	150	200	250	300		1050
y_i	6	7	13	21	24	34		105

$\bar{x} = \frac{1050}{6} = 175$ et $\bar{y} = \frac{105}{6} = 17,5$;
donc $G(175; 17,5)$.

14

$$\bar{x} = \frac{10,2+9,4+x+8,1+11}{5} = \frac{x+38,7}{5} \quad \text{et} \quad \bar{y} = \frac{16,3+13,4+17,6+y+13,3}{5} = \frac{y+60,6}{5}$$

Par définition, le point moyen G a pour coordonnées $(\bar{x}; \bar{y})$. Comme il est donné que le point moyen a pour coordonnées $(175; 17,5)$, on a :

$$\begin{aligned} \bar{x} &= 175 \text{ et } \bar{y} = 17,5 ; \\ \frac{x+38,7}{5} &= 175 \text{ et } \frac{y+60,6}{5} = 17,5 ; \\ x + 38,7 &= 875 \text{ et } y + 60,6 = 87,5 ; \\ x &= 836,3 \text{ et } y = 26,9. \end{aligned}$$

15

NB : Dans l'énoncé, lire pour la consigne 3-a) « $y = \frac{9}{70}x - \frac{188}{7}$ » au lieu de « $y = 1\,124x - 22,01$ ».

1) Calcul des coordonnées du point moyen G.

									Total
x_i	260	260	280	300	310	340	350		2100
y_i	7	8	8,9	10,1	12	13	23		82

$$\bar{x} = \frac{2100}{7} = 300 \quad \text{et} \quad \bar{y} = \frac{82}{7}; \quad \text{donc } G(300; \frac{82}{7}).$$

2)

Série 1					Série 2			
x_i	260	260	280	300	x_i	310	340	350
y_i	7	8	8,9	10,1	y_i	12	13	23

2) a) Equation de la droite (D) de Mayer :

$$\text{Série 1 : } x_{G_1} = \frac{260+260+280+300}{4} = \frac{1100}{4} = 275 \quad \text{et} \quad y_{G_1} = \frac{7+8+8,9+10,1}{4} = \frac{34}{4} = \frac{17}{2}; \quad \text{donc } G_1(275; \frac{17}{2}).$$

$$\text{Série 2 : } x_{G_2} = \frac{310+340+350}{3} = \frac{1000}{3} \quad \text{et} \quad y_{G_2} = \frac{12+13+23}{3} = \frac{48}{3} = 16; \quad \text{donc } G_2(\frac{1000}{3}; 16).$$

Une équation de la droite de Mayer, c'est-à-dire de la droite (G_1G_2) est de la forme :
 $y = ax + b$.

$$\text{La droite } (G_1G_2) \text{ a pour coefficient directeur } a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{16 - \frac{17}{2}}{\frac{1000}{3} - 275} = \frac{\frac{15}{2}}{\frac{175}{3}} = \frac{45}{350} = \frac{9}{70}.$$

La droite (G_1G_2) passe par $G_1(275; \frac{17}{2})$, donc b vérifie : $y_{G_1} = \frac{9}{70}x_{G_1} + b$, c'est-à-dire
 $\frac{17}{2} = \frac{9}{70} \times 275 + b$; d'où $b = \frac{17}{2} - \frac{9}{70} \times 275 = -\frac{188}{7}$.

Une équation de la droite (G_1G_2) , c'est-à-dire de la droite (D) de Mayer est $y = \frac{9}{70}x - \frac{188}{7}$.

b) La droite (D) a pour équation $y = \frac{9}{70}x - \frac{188}{7}$ et le point moyen $G(300; \frac{82}{7})$.

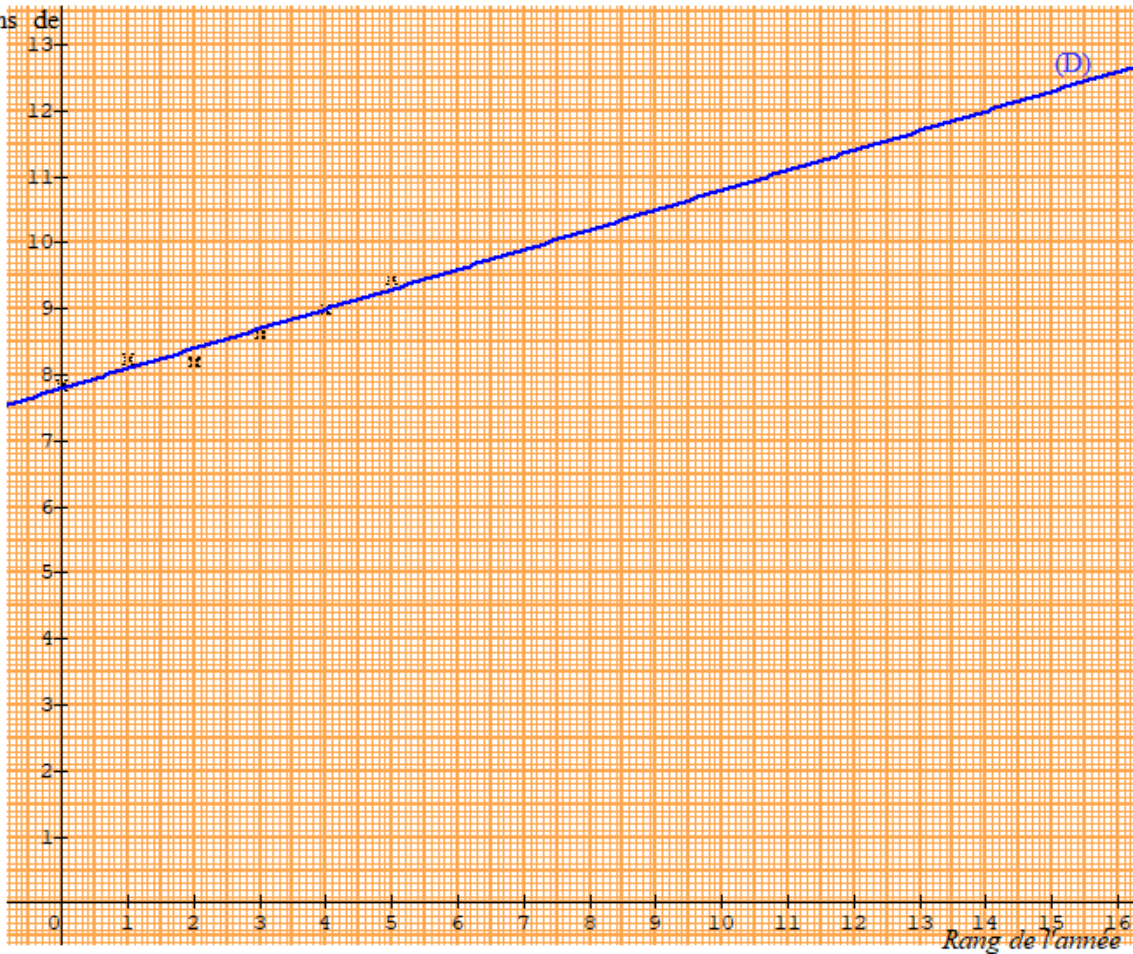
On a : $\frac{9}{70} \times 300 - \frac{188}{7} = \frac{2700}{70} - \frac{188}{7} = \frac{270}{7} - \frac{188}{7} = \frac{282}{7}$; donc G appartient effectivement à la droite (D).

16

NB : L'enseignant informera les apprenants de donner les résultats sous forme décimale tronquée à l'ordre 4.

a)

Chiffre
d'affaire en
millions de
francs



b) La droite de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x;y)}{V(x)}, b = \bar{y} - a\bar{x}, V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2 \text{ et } \text{Cov}(x;y) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} - \bar{x}\bar{y}.$$

x_i	y_i	$x_i y_i$	x_i^2
0	7,85	0	0
1	8,23	8,23	1
2	8,19	16,38	4
3	8,62	25,86	9
4	8,98	35,92	16
5	9,46	47,3	25
Total	15	51,33	133,69

Moyenne en x : $\bar{x} = \frac{15}{6} = 2,5$;

Moyenne en y : $\bar{y} = \frac{51,33}{6} \approx 8,56$;

Variance de x : $V(x) = \frac{55}{6} - 2,5^2 \approx 2,92$;

Covariance : $\text{Cov}(x;y) = \frac{133,69}{6} - 2,5 \times 8,56 \approx 0,88$.

• On a $a = \frac{0,88}{2,92} \approx 0,30$ et $b = 8,56 - 0,30 \times 2,5 \approx 7,81$.

La droite de régression de y en x admet pour équation : $y = 0,30x + 7,81$.

c) Construction de la droite (L) : Voir représentation.

(L) : $y = 0,3x + 7,8$.

x	3	9
y	8,7	10,5

Si $x=3$, alors $y=0,3 \times 3 + 7,8 = 8,7$.

Si $x=9$, alors $y=0,3 \times 9 + 7,8 = 10,5$.

d) Méthode algébrique :

La droite d'ajustement à utiliser a pour équation $y = 0,3x + 7,8$.

L'an 2030 correspond à l'année derang $y=14$.

Pour $y = 14$, on obtient $14 = 0,3x + 7,8$; c'est à dire $x \approx 20,666\ 666$.

Le chiffre d'affaires de l'année 2030 correspond à $x \approx 20,666\ 666$ (exprimé en millions de francs). Ainsi, le chiffre d'affaires de l'année 2030 peut s'estimer à 20 666 666 francs.

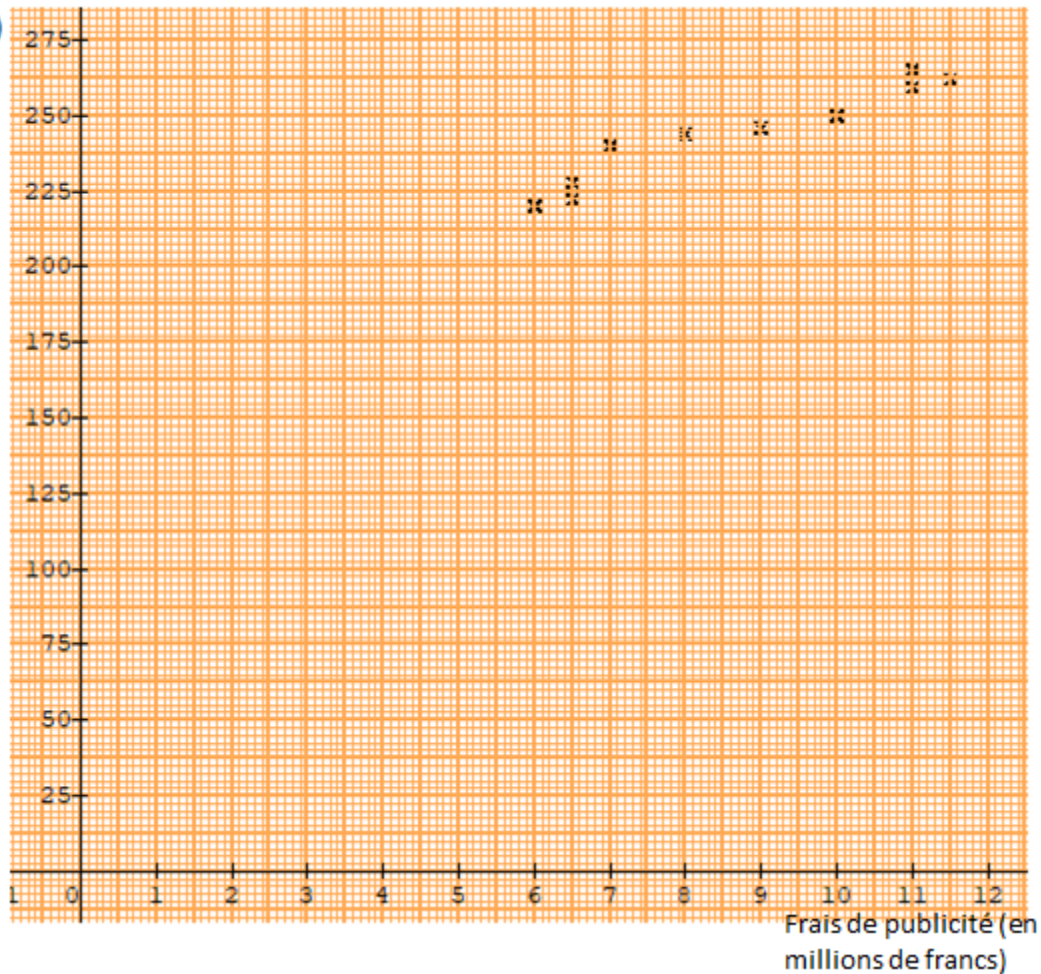
17

NB : Dans l'énoncé :

- lire « millions de francs » au lieu de « centaine d'Euros ».
- lire dans la deuxième ligne du tableau « 265 » au lieu de « 4 ».

1) Nuage de points

Chiffre d'affaires
(en millions de francs)



2) Coordonnées du point moyen G :

$$\bar{X} = \frac{10 + 6 + 6,5 + 11,5 + 11 + 8 + 7 + 6,5 + 11 + 9}{10} = \frac{86,5}{10} = 8,65$$

$$\text{et } \bar{Y} = \frac{250+220+228+262+265+244+240+222+259+246}{10} = \frac{2\ 436}{10} = 243,6.$$

donc G(8,65 ; 243,6).

3) On commence à classer les x_i par ordre croissant :

x_i	6	6,5	6,5	7	8	9	10	11	11	11,5
y_i	220	222	228	240	244	246	250	265	259	262

Série 1						Total
x	6	6,5	6,5	7	8	34
y	220	222	228	240	244	1154

$$x_{G_1} = \frac{34}{5} = 6,8 \text{ et } y_{G_1} = \frac{1154}{5} = 230,8.$$

Donc $G_1(6,8; 230,8)$.

Une équation de la droite de Mayer c'est-à-dire la droite (G_1G_2) est de la forme :

$$y = ax + b.$$

Elle a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{256,4 - 230,8}{10,5 - 6,8} = \frac{25,6}{3,7} \approx 6,92$ et pour ordonnée à l'origine $b = \bar{y} -$

$$a \times \bar{x} = 243,6 - 6,92 \times 8,65 = 183,74.$$

Une équation de la droite de Mayer est donc : $y = 6,92x + 183,74$.

Série 2						Total
x	9	10	11	11	11,5	52,5
y	246	250	265	259	262	1282

$$x_{G_2} = \frac{52,5}{5} = 10,5 \text{ et } y_{G_2} = \frac{1282}{5} = 256,4.$$

Donc $G_2(10,5; 256,4)$.

3) Les estimation se feront à l'aide d'une droite d'ajustement, celle de Mayer précisément.

a) Si le chiffre d'affaires est $y = 250$ (exprimé en en millions de francs), alors la valeur x (exprimée en millions de francs) des frais publicitaires vérifie :

$$250 = 6,92x + 183,74; \text{ c'est à dire } x = 9,5751445 \text{ (exprimé en en millions de francs).}$$

Dans ce cas, la valeur des frais publicitaires s'estime à 9 575 144,5 francs.

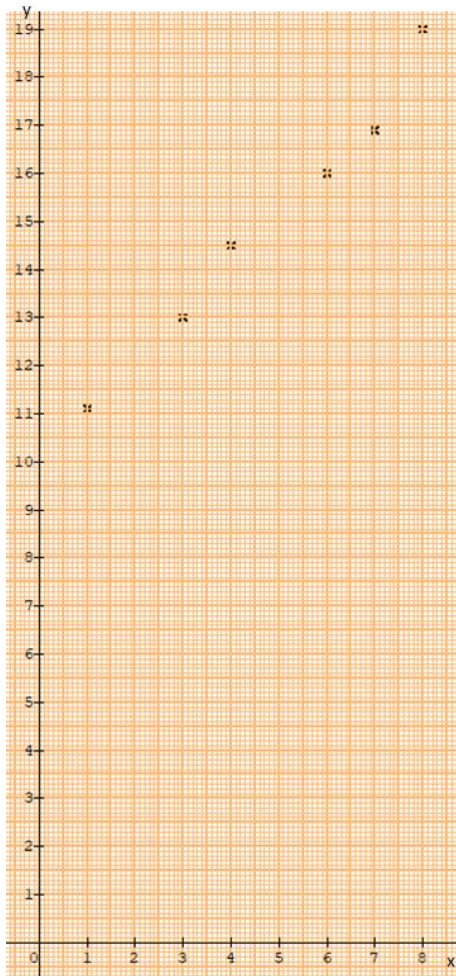
b) Si la valeur des frais publicitaires est $x = 13$ (exprimée en millions de francs), alors la valeur y (exprimée en millions de francs) du chiffre d'affaires vérifie :

$$y = 6,92 \times 13 + 183,74; \text{ c'est à dire } y = 273,7 \text{ (exprimé en en millions de francs).}$$

Dans ce cas, la valeur du chiffre d'affaires s'estime à 273 700 000 francs.

18

1) Nuage de points



2) Coordonnées du point moyen G :

$$\bar{X} = \frac{1+3+4+6+7+8}{6} = \frac{29}{6} = 4,83$$

et $\bar{Y} = \frac{11,1+13+14,5+16+16,9+19}{6} = \frac{90,5}{6} = 15,08$;

donc $G(4,83 ; 15,08)$.

3) a)

Variance de x , $V(x) = \frac{x_1^2+x_2^2+\dots+x_n^2}{n} - \bar{x}^2$

et Variance de y , $V(y) = \frac{y_1^2+y_2^2+\dots+y_n^2}{n} - \bar{y}^2$.

	x_i	y_i	x_i^2	y_i^2
	1	11,1	1	123,21
	3	13	9	169
	4	14,5	16	210,25
	6	16	36	256
	7	16,9	49	285,61
	8	19	64	361
Total	29	90,5	175	1405,07

Moyenne en x : $\bar{x} = \frac{29}{6} \approx 4,83$;

Moyenne en y : $\bar{y} = \frac{90,5}{6} \approx 15,08$;

$V(x) = \frac{175}{6} - 4,83^2 \approx 5,84$;

$V(y) = \frac{1405,07}{6} - 15,08^2 \approx 6,77$.

b) Covariance : $Cov(x ; y) = \frac{x_1y_1+x_2y_2+\dots+x_ny_n}{n} - \bar{x}\bar{y}$.

	x_i	y_i	x_iy_i
	1	11,1	11,1
	3	13	39
	4	14,5	58
	6	16	96
	7	16,9	118,3
	8	19	152
Total	29	90,5	474,4

$$Cov(x ; y) = \frac{11,1+39+58+96+118,3+152}{6} - 4,83 \times 15,08$$

$$= \frac{474,4}{6} - 72,8364$$

$Cov(x ; y) \approx 6,23$.

4) Le coefficient de corrélation : $r = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} = \frac{6,23}{\sqrt{5,84 \times 6,77}} \approx 0,99$.

$0,87 \leq r \leq 1$, donc il y a une **bonne corrélation** ou **une forte corrélation** entre les deux variables.

5) La droite de régression de y en x admet une équation de la forme $y = ax + b$

avec $a = \frac{Cov(x;y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$.

On a : $a \approx \frac{6,23}{5,84} \approx 1,07$ et $b \approx 15,08 - 1,07 \times 4,83 \approx 9,91$.

Une équation de la droite de régression de y en x est : $y = 1,07x + 9,91$.

19

1)

	Groupe 1					
x	1	2,5	3	3,5	4	4
y	15	14	13	13,5	13	12,5

	Groupe 2					
x	5	5,5	5,5	6	8	9
y	12	11,5	12	11,5	11	11

2)

Groupe 1						Total	
x	1	2,5	3	3,5	4	4	18
y	15	14	13	13,5	13	12,5	81

$$x_{G_1} = \frac{18}{6} = 3 \text{ et } y_{G_1} = \frac{81}{6} = 13,5.$$

Donc $G_1(3 ; 13,5)$.

Groupe 2						Total	
x	5	5,5	5,5	6	8	9	39
y	12	11,5	12	11,5	11	11	69

$$x_{G_2} = \frac{39}{6} = 6,5 \text{ et } y_{G_2} = \frac{69}{6} = 11,5.$$

Donc $G_2(6,5 ; 11,5)$.

3) Une équation de la droite de Mayer c'est-à-dire la droite (G_1G_2) est de la forme :

$$y = ax + b.$$

Elle a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{11,5 - 13,5}{6,5 - 3} = \frac{-2}{3,5} \approx -0,57$ et comme elle passe par $G_1(3 ; 13,5)$, elle a pour ordonnée à l'origine :

$$b = y_{G_1} - a \times x_{G_1} = 13,5 - (-0,57) \times 3 \approx 15,21.$$

Une équation de la droite de Mayer est donc : $y = -0,57x + 15,21$.

20

1)

Groupe 1					Total	
x	3	5	6	6	9	29
y	5	8	10	8	10	41

$$x_{G_1} = \frac{29}{5} = 5,8 \text{ et } y_{G_1} = \frac{41}{5} = 8,2.$$

Donc $G_1(5,8 ; 8,2)$.

Groupe 2					Total	
x	9	12	12	14	14	61
y	13	13	16	17	16	75

$$x_{G_2} = \frac{61}{5} = 12,2 \text{ et } y_{G_2} = \frac{75}{5} = 15.$$

Donc $G_2(12,2 ; 15)$.

2) Une équation de la droite de Mayer c'est-à-dire la droite (G_1G_2) est de la forme :

$$y = ax + b.$$

Elle a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{15 - 8,2}{12,2 - 5,8} = \frac{6,8}{6,4} = 1,0625$ et comme elle passe par $G_2(12,2 ; 15)$, elle a pour ordonnée à l'origine :

$$b = y_{G_2} - a \times x_{G_2} = 15 - 1,0625 \times 12,2 = 2,0375.$$

Une équation de la droite (G_1G_2) est donc : $y = 1,0625x + 2,0375$.

21

(A1 seulement)

NB :

- L'enseignant pourra demander aux apprenants de donner les résultats des consignes de cet exercice avec des fractions irréductibles.
- Lire la consigne suivante « Détermine une équation de la droite de régression de y en x par la méthode des moindres carrés » à la place de la consigne 2).

1) Calcul de la covariance de x et y :

Méthode 1	Méthode 2																																																												
$\text{Cov}(x ; y) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} - \bar{x}\bar{y}$	$\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$																																																												
<table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>$x_i y_i$</th> </tr> </thead> <tbody> <tr><td>1</td><td>41</td><td>41</td></tr> <tr><td>2</td><td>68</td><td>136</td></tr> <tr><td>3</td><td>55</td><td>165</td></tr> <tr><td>4</td><td>80</td><td>320</td></tr> <tr><td>5</td><td>95</td><td>475</td></tr> <tr><td>6</td><td>104</td><td>624</td></tr> <tr><td>7</td><td>100</td><td>700</td></tr> <tr><td>8</td><td>122</td><td>976</td></tr> <tr> <td>Total</td> <td>36</td> <td>665</td> </tr> </tbody> </table>	x_i	y_i	$x_i y_i$	1	41	41	2	68	136	3	55	165	4	80	320	5	95	475	6	104	624	7	100	700	8	122	976	Total	36	665	<table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>$(x_i - \bar{x})(y_i - \bar{y})$</th> </tr> </thead> <tbody> <tr><td>1</td><td>41</td><td>147,4375</td></tr> <tr><td>2</td><td>68</td><td>37,8125</td></tr> <tr><td>3</td><td>55</td><td>42,1875</td></tr> <tr><td>4</td><td>80</td><td>1,5625</td></tr> <tr><td>5</td><td>95</td><td>5,9375</td></tr> <tr><td>6</td><td>104</td><td>31,3125</td></tr> <tr><td>7</td><td>100</td><td>42,1875</td></tr> <tr><td>8</td><td>122</td><td>136,0625</td></tr> <tr> <td>Total</td> <td>36</td> <td>665</td> </tr> </tbody> </table>	x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$	1	41	147,4375	2	68	37,8125	3	55	42,1875	4	80	1,5625	5	95	5,9375	6	104	31,3125	7	100	42,1875	8	122	136,0625	Total	36	665
x_i	y_i	$x_i y_i$																																																											
1	41	41																																																											
2	68	136																																																											
3	55	165																																																											
4	80	320																																																											
5	95	475																																																											
6	104	624																																																											
7	100	700																																																											
8	122	976																																																											
Total	36	665																																																											
x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$																																																											
1	41	147,4375																																																											
2	68	37,8125																																																											
3	55	42,1875																																																											
4	80	1,5625																																																											
5	95	5,9375																																																											
6	104	31,3125																																																											
7	100	42,1875																																																											
8	122	136,0625																																																											
Total	36	665																																																											
<p>Moyenne en x : $\bar{x} = \frac{36}{8} = \frac{9}{2}$,</p>	<p>Moyenne en x : $\bar{x} = \frac{36}{8} = \frac{9}{2}$,</p>																																																												
	<p>Moyenne en y : $\bar{y} = \frac{665}{8}$.</p>																																																												

Moyenne en y : $\bar{y} = \frac{665}{8}$. Covariance : $\text{Cov}(x ; y) = \frac{3437}{8} - \frac{9}{2} \times \frac{665}{8} = \frac{889}{16}$.	Covariance: $\text{Cov}(x ; y) = \frac{444,5}{8} = \frac{4445}{80} = \frac{889}{16}$.
--	---

2) Détermination d'une équation de la droite de régressions de y en x par la méthode des moindres carrés :

La droite de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x;y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x}.$$

Calculons d'abord la variance de x :

Méthode 1	Méthode 2																																																												
Variance de x : $V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$ <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>x_i^2</th> </tr> </thead> <tbody> <tr><td>1</td><td>41</td><td>1</td></tr> <tr><td>2</td><td>68</td><td>4</td></tr> <tr><td>3</td><td>55</td><td>9</td></tr> <tr><td>4</td><td>80</td><td>16</td></tr> <tr><td>5</td><td>95</td><td>25</td></tr> <tr><td>6</td><td>104</td><td>36</td></tr> <tr><td>7</td><td>100</td><td>49</td></tr> <tr><td>8</td><td>122</td><td>64</td></tr> <tr> <td>Total</td> <td>36</td> <td>665</td> </tr> </tbody> </table> $V(x) = \frac{204}{8} - \left(\frac{9}{2}\right)^2 = \frac{102}{4} - \frac{81}{4} = \frac{21}{4}$	x_i	y_i	x_i^2	1	41	1	2	68	4	3	55	9	4	80	16	5	95	25	6	104	36	7	100	49	8	122	64	Total	36	665	Variance de x : $V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$ <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>$(x_i - \bar{x})^2$</th> </tr> </thead> <tbody> <tr><td>1</td><td>41</td><td>12,25</td></tr> <tr><td>2</td><td>68</td><td>6,25</td></tr> <tr><td>3</td><td>55</td><td>2,25</td></tr> <tr><td>4</td><td>80</td><td>0,25</td></tr> <tr><td>5</td><td>95</td><td>0,25</td></tr> <tr><td>6</td><td>104</td><td>2,25</td></tr> <tr><td>7</td><td>100</td><td>6,25</td></tr> <tr><td>8</td><td>122</td><td>12,25</td></tr> <tr> <td>Total</td> <td>36</td> <td>665</td> </tr> </tbody> </table> Variance de x : $V(x) = \frac{42}{8} = \frac{21}{4}$.	x_i	y_i	$(x_i - \bar{x})^2$	1	41	12,25	2	68	6,25	3	55	2,25	4	80	0,25	5	95	0,25	6	104	2,25	7	100	6,25	8	122	12,25	Total	36	665
x_i	y_i	x_i^2																																																											
1	41	1																																																											
2	68	4																																																											
3	55	9																																																											
4	80	16																																																											
5	95	25																																																											
6	104	36																																																											
7	100	49																																																											
8	122	64																																																											
Total	36	665																																																											
x_i	y_i	$(x_i - \bar{x})^2$																																																											
1	41	12,25																																																											
2	68	6,25																																																											
3	55	2,25																																																											
4	80	0,25																																																											
5	95	0,25																																																											
6	104	2,25																																																											
7	100	6,25																																																											
8	122	12,25																																																											
Total	36	665																																																											

• On obtient : $a = \frac{\text{Cov}(x;y)}{V(x)} = \frac{\frac{889}{16}}{\frac{21}{4}} = \frac{889}{16} \times \frac{4}{21} = \frac{889}{84} = \frac{127}{12}$

et $b = \bar{y} - a\bar{x} = \frac{665}{8} - \frac{127}{12} \times \frac{9}{2} = \frac{665}{8} - \frac{127}{4} \times \frac{3}{2} = \frac{665}{8} - \frac{381}{8} = \frac{284}{8} = \frac{71}{2}$.

Une équation de la droite de régression de y en x est : $y = \frac{127}{12}x + \frac{71}{2}$.

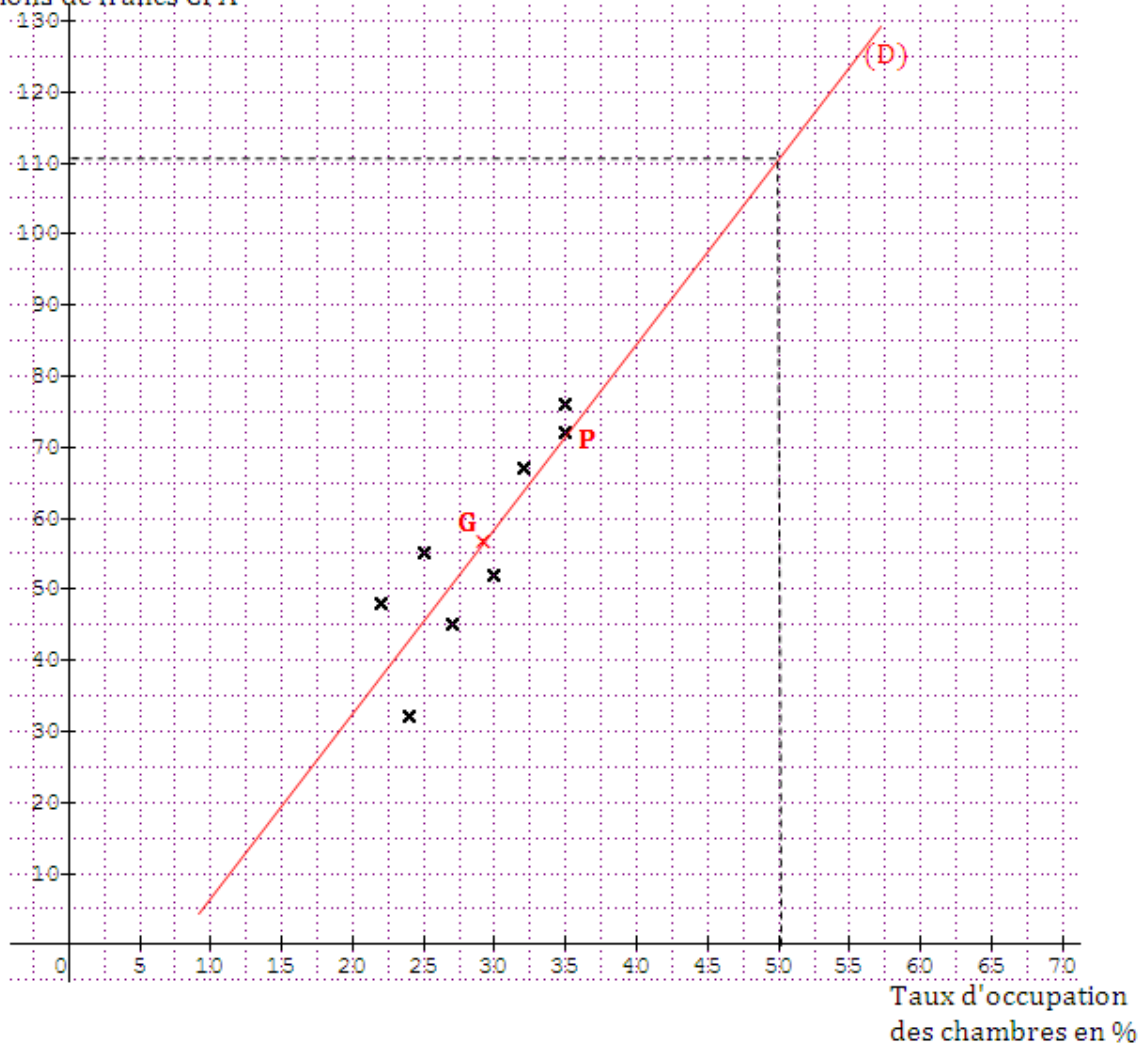
Exercices d'approfondissement

22*

NB : Pour la première consigne, lire « Prendre 1 cm pour 10 millions de francs et 1 cm pour 5 % ». au lieu de « Prendre 1 cm pour 5 millions de francs et 1 cm pour 10 % ». »

1) Nuage de points.

Frais de publicité en millions de francs CFA.



2) Calcul des coordonnées du point moyen G :

									Total
x_i	30	27	32	25	35	22	24	35	230
y_i	52	45	67	55	76	48	32	72	447

$\bar{x} = \frac{230}{8} = 28,75$ et $\bar{y} = \frac{447}{8} = 55,875$; donc $G(28,75 ; 55,875)$.

3) 3-a) Voir le graphique.

3-b) Une équation de la droite d'ajustement (D), c'est-à-dire de la droite (GP) est de la forme : $y = ax + b$.

La droite (D) a pour coefficient directeur $a = \frac{y_P - y_G}{x_P - x_G} = \frac{72 - 55,875}{35 - 28,75} = \frac{16,125}{6,25} = 2,58$.

La droite (D) passe par $P(35; 72)$, donc : $72 = 2,58 \times 35 + b$; d'où $b = -18,3$.

Une équation de la droite (D) est $y = 2,58x - 18,3$.

3-c) On acceptera 110 millions ou 111 millions.

4) 15 000 000 = 15 millions.

Pour $y=15$, on a $15 = 2,58 \times 15 - 18,3$, c'est à dire $x=20,4$.

Ainsi pour 15 000 000 de frais de publicité engagés, le taux d'occupation des chambres s'estime à 20,4 millions (20 400 000) de francs CFA.

23

(A1 seulement)

1) Calcul de la covariance de x et y :

Méthode 1	Méthode 2
$\text{Cov}(x ; y) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} - \bar{x}\bar{y}$	$\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$

x_i	y_i	$x_i y_i$	
2	83	166	
5	70	350	
6	70	420	
10	54	540	
12	49	588	
Total	35	326	2064

Moyenne en x : $\bar{x} = \frac{35}{5} = 7$;
Moyenne en y : $\bar{y} = \frac{326}{5}$.
Covariance :
Cov(x ; y) = $\frac{2064}{5} - 7 \times \frac{326}{5} = -\frac{218}{5}$.

x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$	
2	83	-89	
5	70	-9,6	
6	70	-4,8	
10	54	-33,6	
12	49	-81	
Total	35	326	-218

Moyenne en x : $\bar{x} = \frac{35}{5} = 7$;
Moyenne en y : $\bar{y} = \frac{326}{5}$.
Covariance :
Cov(x ; y) = $-\frac{218}{5}$.

2) Le coefficient de corrélation: $r = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}}$.
Calculons d'abord les variances de x et de y :

Méthode 1	Méthode 2																																																										
<p>Variances :</p> $V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2$ $V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2$ <table border="1" style="margin: 10px auto;"> <thead> <tr><th>x_i</th><th>y_i</th><th>x_i^2</th><th>y_i^2</th></tr> </thead> <tbody> <tr><td>2</td><td>83</td><td>4</td><td>6889</td></tr> <tr><td>5</td><td>70</td><td>25</td><td>4900</td></tr> <tr><td>6</td><td>70</td><td>36</td><td>4900</td></tr> <tr><td>10</td><td>54</td><td>100</td><td>2916</td></tr> <tr><td>12</td><td>49</td><td>144</td><td>2401</td></tr> <tr><td>Total</td><td>35</td><td>326</td><td>309</td><td>22006</td></tr> </tbody> </table> $V(x) = \frac{309}{5} - 7^2 = \frac{64}{5}$ $V(y) = \frac{22006}{5} - \left(\frac{326}{5}\right)^2 = \frac{3754}{25}$	x_i	y_i	x_i^2	y_i^2	2	83	4	6889	5	70	25	4900	6	70	36	4900	10	54	100	2916	12	49	144	2401	Total	35	326	309	22006	<p>Variances :</p> $V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$ $V(y) = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n}$ <table border="1" style="margin: 10px auto;"> <thead> <tr><th>x_i</th><th>y_i</th><th>$(x_i - \bar{x})^2$</th><th>$(x_i - \bar{x})^2$</th></tr> </thead> <tbody> <tr><td>2</td><td>83</td><td>25</td><td>316,84</td></tr> <tr><td>5</td><td>70</td><td>4</td><td>23,04</td></tr> <tr><td>6</td><td>70</td><td>1</td><td>23,04</td></tr> <tr><td>10</td><td>54</td><td>9</td><td>125,44</td></tr> <tr><td>12</td><td>49</td><td>25</td><td>262,44</td></tr> <tr><td>Total</td><td>35</td><td>326</td><td>64</td><td>750,8</td></tr> </tbody> </table> $V(x) = \frac{64}{5}$ $V(y) = \frac{750,8}{5} = \frac{7508}{50} = \frac{3754}{25}$	x_i	y_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2$	2	83	25	316,84	5	70	4	23,04	6	70	1	23,04	10	54	9	125,44	12	49	25	262,44	Total	35	326	64	750,8
x_i	y_i	x_i^2	y_i^2																																																								
2	83	4	6889																																																								
5	70	25	4900																																																								
6	70	36	4900																																																								
10	54	100	2916																																																								
12	49	144	2401																																																								
Total	35	326	309	22006																																																							
x_i	y_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2$																																																								
2	83	25	316,84																																																								
5	70	4	23,04																																																								
6	70	1	23,04																																																								
10	54	9	125,44																																																								
12	49	25	262,44																																																								
Total	35	326	64	750,8																																																							

• Coefficient de corrélation : $r = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} = \frac{-\frac{218}{5}}{\sqrt{\frac{64}{5} \times \frac{3754}{25}}} \approx -0,9945$.

• $-1 \leq r \leq -0,87$, donc il y a une **bonne corrélation** ou **une forte corrélation** entre les deux variables.

3) Détermination d'une équation de la droite de régressions de y en x par la méthode des moindres carrés :

La droite de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$a = \frac{Cov(x;y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$.

On obtient : $a = \frac{Cov(x;y)}{V(x)} = \frac{-\frac{218}{5}}{\frac{64}{5}} = -\frac{109}{32} = -3,40625$

et $b = \bar{y} - a\bar{x} = \frac{326}{5} - (-3,40625) \times 7 = 89,04375$.

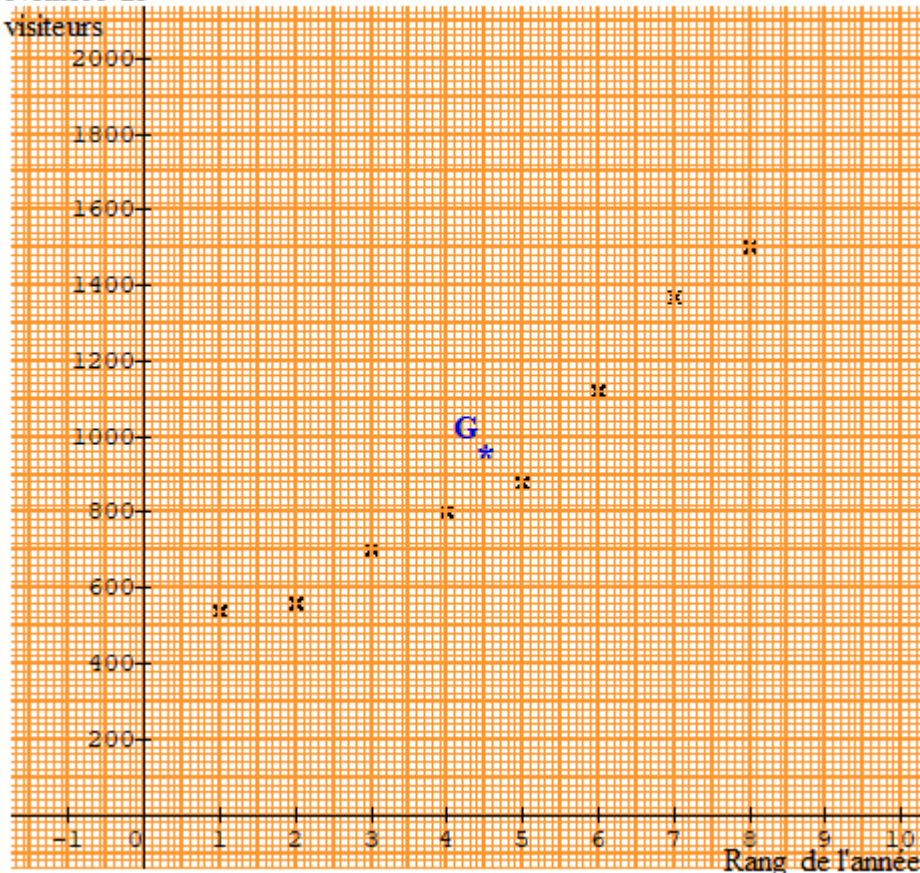
Une équation de la droite de régression de y en x est : $y = -3,40625x + 89,04375$.

NB : Dans l'énoncé, lire « 3 370 » dans la deuxième ligne du tableau « 265 » au lieu de « 31 370 ».

1) a)

Nombre de

visiteurs



b) Calcul des coordonnées du point moyen G :

									Total
x_i	1	2	3	4	5	6	7	8	36
y_i	540	560	700	800	875	1120	1370	1500	7465

$$\bar{x} = \frac{36}{8} = 4,5 \text{ et } \bar{y} = \frac{7465}{8} = 933,125 ; \text{ donc } G(4,5; 933,125).$$

2) a)

	S1				Total
x	1	2	3	4	10
y	540	560	700	800	2600

$$x_{G_1} = \frac{10}{4} = 2,5 \text{ et } y_{G_1} = \frac{2600}{4} = 650.$$

Donc $G_1(2,5; 650)$.

	S2				Total
x	5	6	7	8	26
y	875	1120	1370	1500	4865

$$x_{G_2} = \frac{26}{4} = 6,5 \text{ et } y_{G_2} = \frac{4865}{4} = 1216,25.$$

Donc $G_2(6,5; 1216,25)$.

b) Une équation de la droite (G_1G_2) est de la forme $y = ax + b$.

Elle a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{1216,25 - 650}{6,5 - 2,5} = \frac{566,25}{4} = 141,5625$ et comme elle passe par $G(4,5;$

$933,125)$, elle a pour ordonnée à l'origine :

$$b = \bar{y} - a\bar{x} = 933,125 - 141,5625 \times 4,5 = 296,09375.$$

Une équation de la droite (G_1G_2) est donc : $y = 141,5625x + 296,09375$.

c) La droite (G_1G_2) est la droite de Mayer ; on peut donc l'utiliser comme droite d'ajustement du nuage de points de la série statistique proposée.

Ainsi, pour l'année de rang $x=10$, le nombre y de visiteurs vérifie :

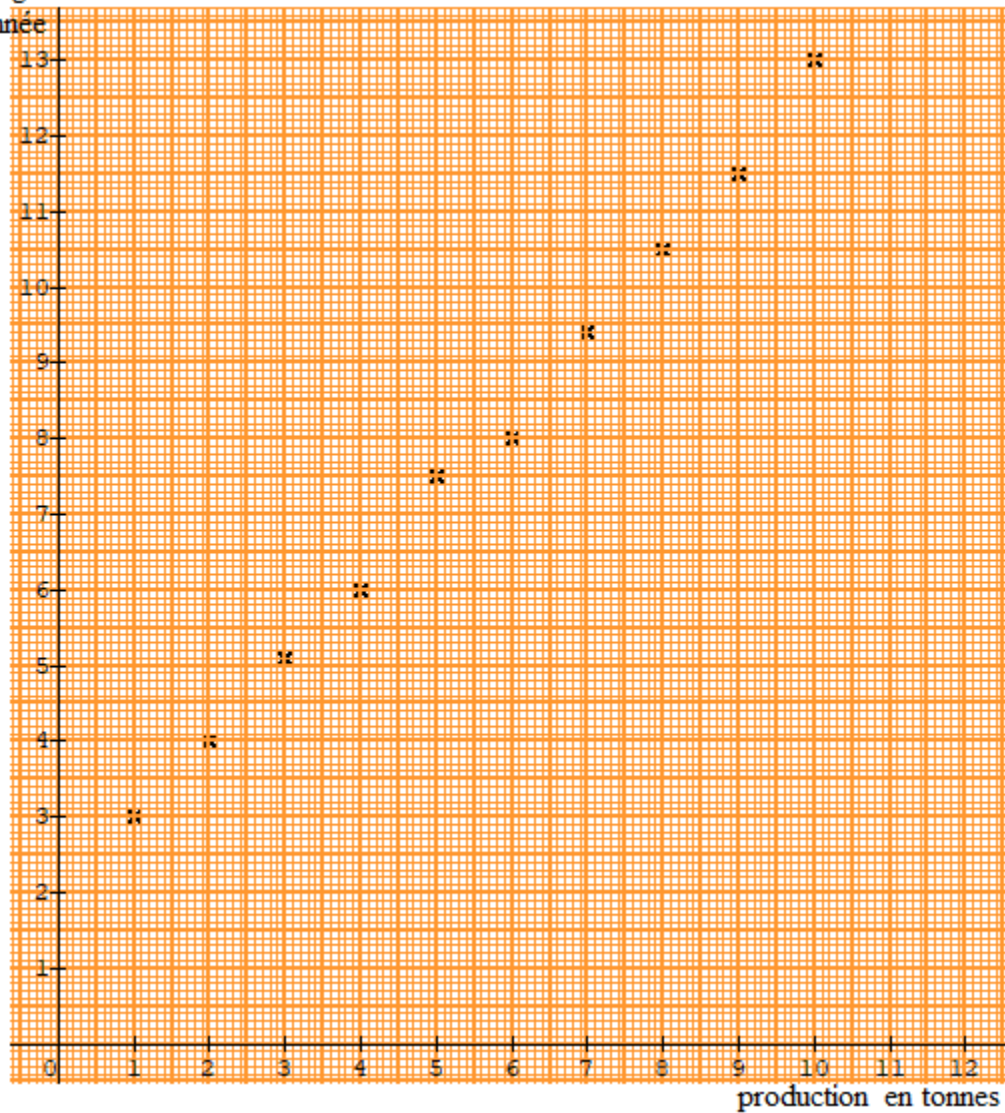
$$y = 141,5625 \times 10 + 296,09375 \approx 1712.$$

Pour l'année de rang 10, le nombre de visiteurs s'estime à 1712.

25*

1)

Rang de l'année



2) Calcul des coordonnées du point moyen G :

											Total
x_i	1	2	3	4	5	6	7	8	9	10	55
y_i	3	4	5,1	6	7,5	8	10,5	11,5	13	13	78

$\bar{x} = \frac{55}{10} = 5,5$ et $\bar{y} = \frac{78}{10} = 7,8$; donc $G(5,5; 7,8)$.

3) a)

	S1					Total
x	1	2	3	4	5	15
y	3	4	5,1	6	7,5	25,6

$x_{G_1} = \frac{15}{5} = 3$ et $y_{G_1} = \frac{25,6}{5} = 5,12$.

Donc $G_1(3; 5,12)$.

Une équation de la droite (G_1G_2) est de la forme : $y = ax + b$.

Elle a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{10,48 - 5,12}{8 - 3} = \frac{5,36}{5} = 1,072$; comme elle passe par $G(5,5; 7,8)$, elle

a pour ordonnée à l'origine :

$b = \bar{y} - a\bar{x} = 7,8 - 1,072 \times 5,5 = 1,904$.

Une équation de la droite (G_1G_2) est donc : $y = 1,072x + 1,904$.

	S2					Total
x	6	7	8	9	10	40
y	8	10,5	11,5	13	13	52,4

$x_{G_2} = \frac{40}{5} = 8$ et $y_{G_2} = \frac{52,4}{5} = 10,48$

Donc $G_2(8; 10,48)$.

La droite (G₁G₂) est la droite de Mayer ; une équation de cette droite d'ajustement du nuage de points de la série statistique proposée est donc : $y = 1,072x + 1,904$.

b) Pour l'année de rang $x=12$, la production yde la société vérifie :

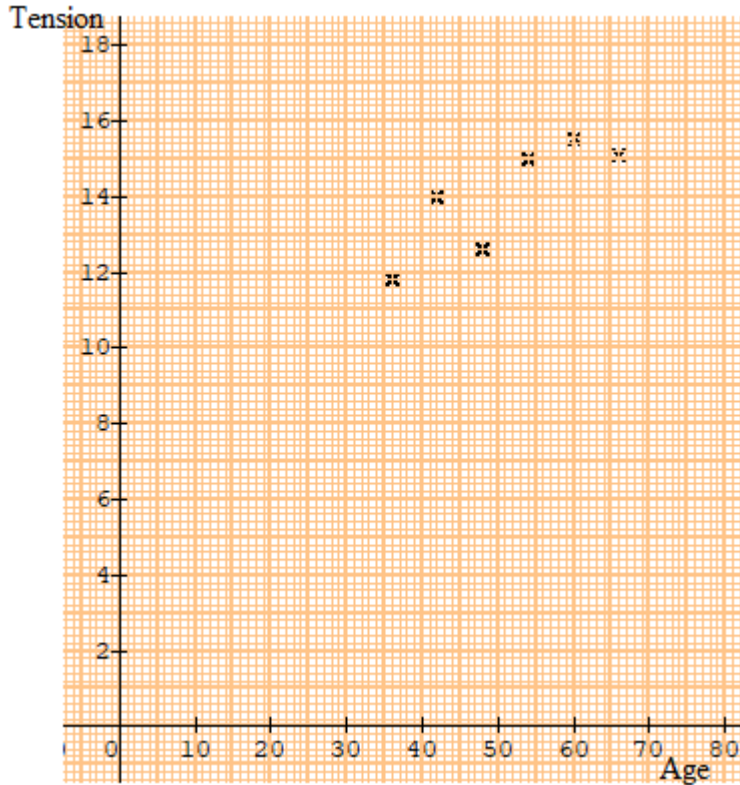
$$y = 1,072 \times 12 + 1,904 = 14,768 .$$

Pour la douzième année, la production yde la société s'estime à 14,768 tonnes.

26

(A1 seulement)

1)



2) Equation de la droite de régressions de y en x par la méthode des moindres carrés :

La droite de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x;y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x} .$$

Calcul des variance et de la covariance:

Méthode 1						Méthode 2					
$V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2 ;$						$V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} ;$					
$V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2 ;$						$V(y) = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n} ;$					
$\text{Cov}(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y} .$						$\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n} .$					
	x_i	y_i	x_i^2	y_i^2	$x_i y_i$		x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
	36	11,8	1296	139,24	424,8		36	11,8	225	4,84	33
	42	14	1764	196	588		42	14	81	0	0
	48	12,6	2304	158,76	604,8		48	12,6	9	1,96	4,2
	54	15	2916	225	810		54	15	9	1	3
	60	15,5	3600	240,25	930		60	15,5	81	2,25	13,5
	66	15,1	4356	228,01	996,6		66	15,1	225	1,21	16,5
Total	306	84	16 236	1 187,26	4 354,2	Total	306	84	630	11,26	70,2
Moyenne en $x : \bar{x} = \frac{306}{6} = 51 ;$						Moyenne en $x : \bar{x} = \frac{306}{6} = 51 ;$					
Moyenne en $y : \bar{y} = \frac{84}{6} = 14 .$						Moyenne en $y : \bar{y} = \frac{84}{6} = 14 .$					
$V(x) = \frac{16 236}{6} - 51^2 = 105 ;$						$V(x) = \frac{630}{6} = 105 ;$					

$$V(y) = \frac{1187,26}{6} - 14^2 = 1,876667;$$

$$\text{Cov}(x ; y) = \frac{4354,2}{6} - 51 \times 14 = 11,7.$$

$$V(y) = \frac{11,26}{6} \approx 1,876667 ;$$

$$\text{Cov}(x ; y) = \frac{70,2}{6} = 11,7.$$

On obtient : $a = \frac{\text{Cov}(x;y)}{V(x)} = \frac{11,7}{105} \approx 0,111429$

et $b = \bar{y} - a\bar{x} = 14 - 0,111429 \times 51 = 8,317121$.

Une équation de la droite de régression de y en x est : $y = 0,111429x + 8,317121$.

3) a) Coefficient de corrélation : $r = \frac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}} = \frac{11,7}{\sqrt{105 \times 1,876667}} \approx 0,833485$.

b) $|r| \notin [0,87; 1]$, donc il y a une **mauvaise corrélation** entre les deux variables.

4) A l'aide de la droite de régression de y en x , pour $x=70$ (ans), on obtient :

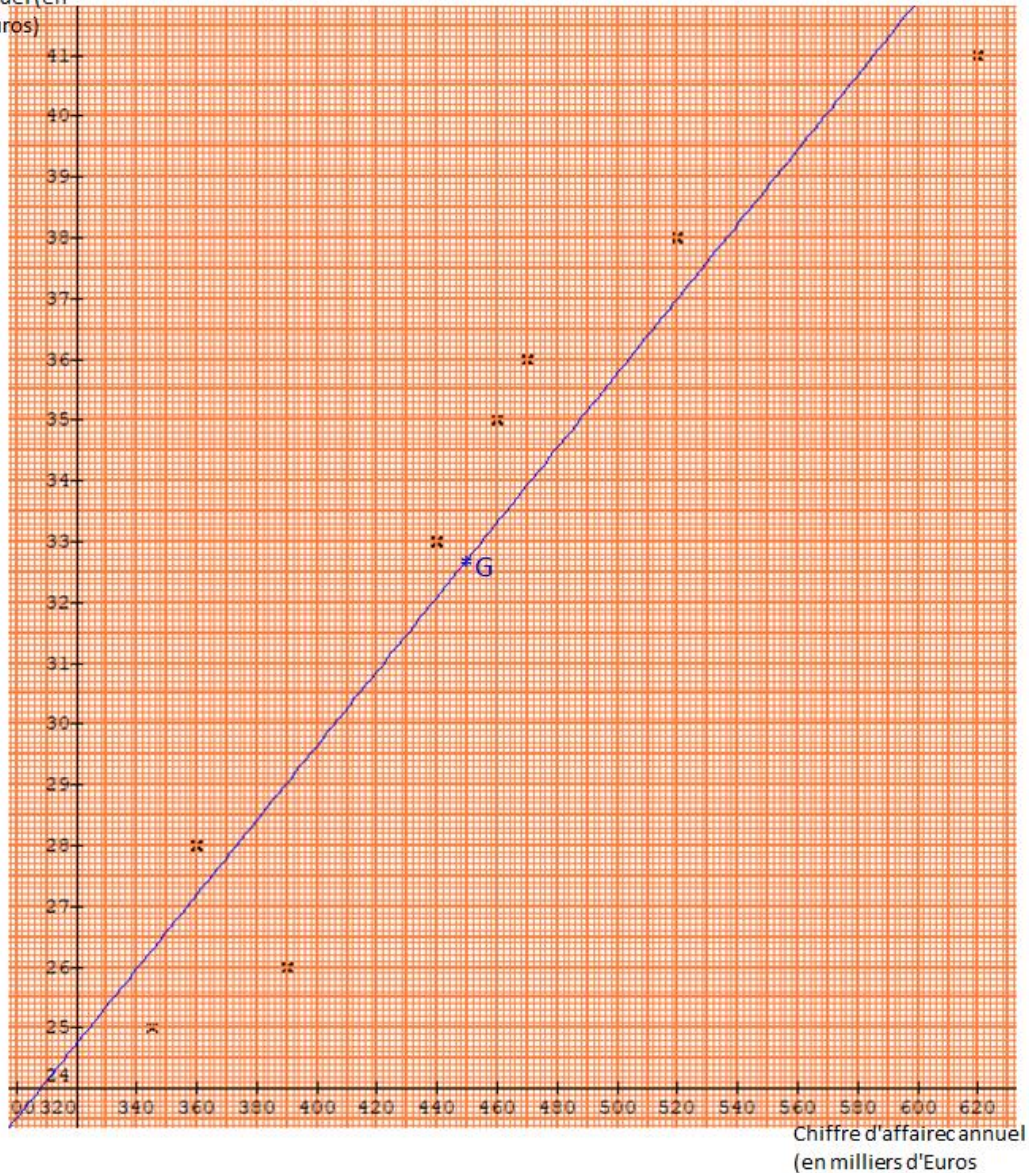
$$y = 0,111429 \times 70 + 8,317121 = 16,117151 \approx 16,1.$$

Avoir 70 ans et une tension artérielle 16,2 semble proche de la normale ; mais la **mauvaise** qualité de la corrélation entre les caractères âge et tension artérielle exclue cela.

27

1) a)

Revenu annuel (en
milliers d'Euros)



b) Calcul des coordonnées du point moyen G :

Total

x_i	340	360	390	440	460	470	520	620	3600
y_i	25	28	26	33	35	36	38	41	262

$\bar{x} = \frac{3600}{8} = 450$ et $\bar{y} = \frac{262}{8} = 32,75$; donc $G(450; 32,75)$.

2)

	S1				Total
x	340	360	390	440	1530
y	25	28	26	33	112

$x_{G_1} = \frac{1530}{4} = 382,5$ et $y_{G_1} = \frac{112}{4} = 28$.

Donc $G_1(382,5; 28)$.

	S2				Total
x	460	470	520	620	2070
y	35	36	38	41	150

$x_{G_2} = \frac{2070}{4} = 517,5$ et $y_{G_2} = \frac{150}{4} = 37,5$

Donc $G_2(517,5; 37,5)$.

3) Une équation de la droite de Mayer, c'est-à-dire de la droite (G_1G_2) est de la forme :

$y = ax + b$.

Elle a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{37,5 - 28}{517,5 - 382,5} = \frac{9,5}{135} = \frac{95}{1350} = \frac{19}{270}$;

comme elle passe par $G(450; 32,75)$, elle a pour ordonnée à l'origine :

$b = \bar{y} - a\bar{x} = 32,75 - \frac{19}{270} \times 450 = \frac{3275}{100} - \frac{19 \times 45}{27} = \frac{131}{4} - \frac{19 \times 5}{3} = \frac{13}{12}$.

Une équation de la droite (G_1G_2) est donc : $y = \frac{19}{270}x + \frac{13}{12}$.

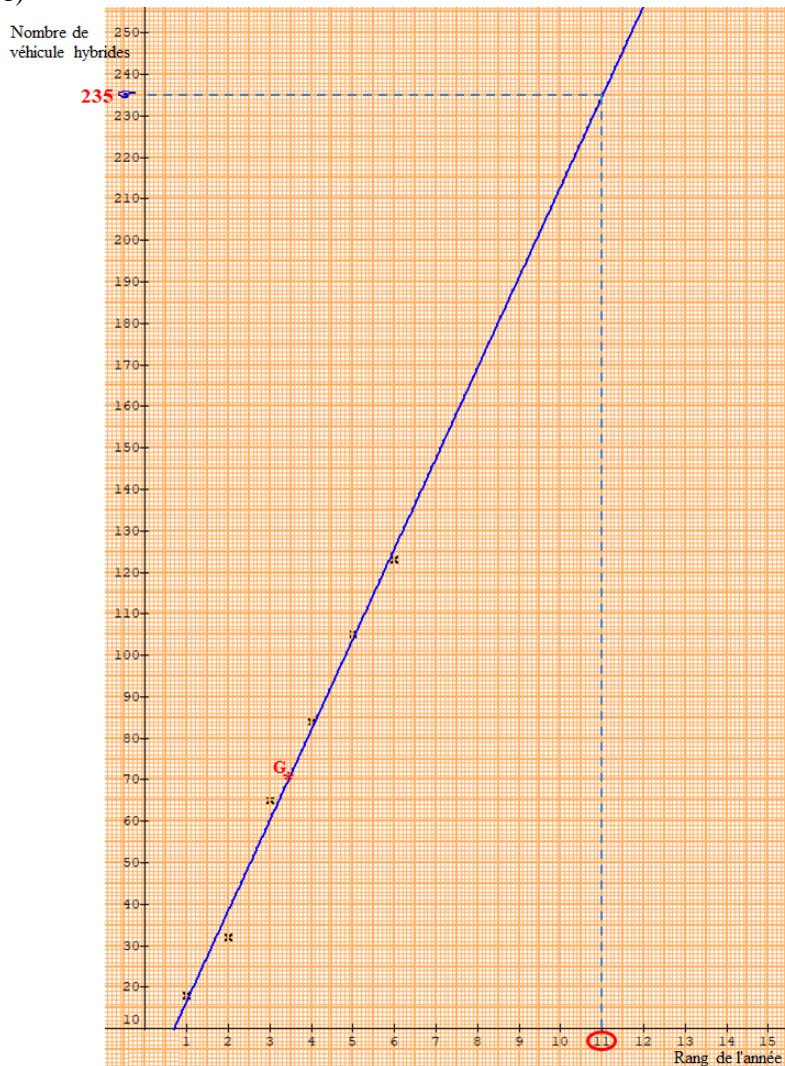
La droite (G_1G_2) est la droite de Mayer ; une équation de cette droite d'ajustement du nuage de points de la série statistique proposée est donc : $y = \frac{19}{270}x + \frac{13}{12}$.

28

(A1 seulement)

NB : On pourra au besoin arrondir les résultats à l'ordre 4.

1)



2) Vérifions qu'un ajustement linéaire est justifié :

Le coefficient de corrélation : $r = \frac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}}$.

Calculons d'abord les variances et la variance :

Méthode 1						Méthode 2					
$V(x) = \frac{x_1^2+x_2^2+\dots+x_n^2}{n} - \bar{x}^2 ;$						$V(x) = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\dots+(x_n-\bar{x})^2}{n} ;$					
$V(y) = \frac{y_1^2+y_2^2+\dots+y_n^2}{n} - \bar{y}^2 ;$						$V(y) = \frac{(y_1-\bar{y})^2+(y_2-\bar{y})^2+\dots+(y_n-\bar{y})^2}{n} ;$					
$\text{Cov}(x ; y) = \frac{x_1y_1+x_2y_2+\dots+x_ny_n}{n} - \bar{x}\bar{y}.$						$\text{Cov}(x ; y) = \frac{(x_1-\bar{x})(y_1-\bar{y})+(x_2-\bar{x})(y_2-\bar{y})+\dots+(x_n-\bar{x})(y_n-\bar{y})}{n}.$					
	x_i	y_i	x_i^2	y_i^2	x_iy_i		x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
	1	18	1	324	18		1	18	6,25	2826,698	132,9168
	2	32	4	1024	64		2	32	2,25	1534,0304	58,7501
	3	65	9	4225	195		3	65	0,25	38,0282	3,0834
	4	84	16	7056	336		4	84	0,25	164,6936	6,4167
	5	105	25	11025	525		5	105	2,25	1144,6922	50,75
	6	123	36	15129	738		6	123	6,25	2686,691	129,5833
Total	21	427	91	38783	1876	Total	21	427	17,5	8394,8334	381,5003
Moyenne en x : $\bar{x} = \frac{21}{6} = 3,5;$						Moyenne en x : $\bar{x} = \frac{21}{6} = 3,5;$					
Moyenne en y : $\bar{y} = \frac{427}{6} = 71,1667.$						Moyenne en y : $\bar{y} = \frac{427}{6} = 71,1667.$					
$V(x) = \frac{91}{6} - 3,5^2 \approx 2,9167 ;$						$V(x) = \frac{17,5}{6} \approx 2,9167 ;$					
$V(y) = \frac{38783}{6} - 71,1667^2 \approx 1399,1341 ;$						$V(y) = \frac{8394,8334}{6} \approx 1399,1389 ;$					
$\text{Cov}(x ; y) = \frac{1876}{6} - 3,5 \times 71,1667 \approx 63,5832.$						$\text{Cov}(x ; y) = \frac{381,5003}{6} \approx 63,5834.$					

Dans tous les cas, on obtient $\frac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}} \approx 0,9953$

C'est-à-dire : $r \approx 0,9953.$

$0,87 \leq r \leq 1$, donc il y a une bonne corrélation ou une forte corrélation entre les deux variables.

Ainsi, un ajustement affine est approprié.

3) Détermination d'une équation de la droite de régressions de y en x par la méthode des moindres carrés :

La droite de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x,y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x}.$$

Selon la méthode utilisée dans le traitement de la consigne précédente, on trouve les résultats suivants :

$$y = 21,7997x - 5,13225 \text{ ou } y = 21,7998x - 5,1326.$$

4) a) On utilise une droite de régression construite dans le même repère que le nuage de points, par exemple la droite d'équation $y = 21,7997x - 5,13225$. Cette droite passe évidemment par le point moyen G(3,5 ; 71,1667).

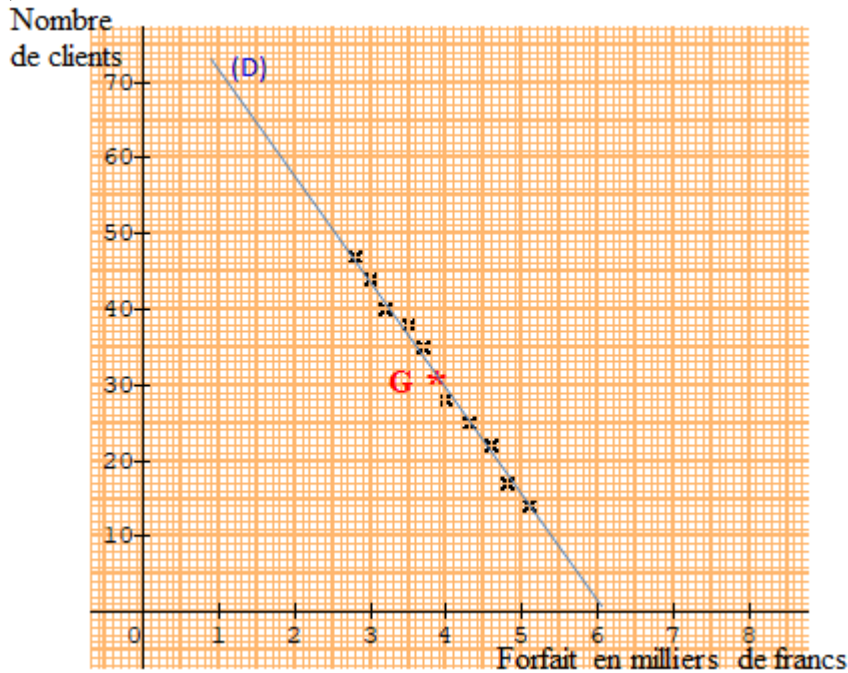
Par lecture graphique, pour l'an 2025 c'est-à-dire pour $x=11$, le nombre de véhicule s'estime à 135.

b) Vérification par le calcul :

L'an 2025 correspond à l'année de rang 11, c'est-à-dire à $x=11$; le nombre y de véhicule vérifie : $y = 21,7997 \times 11 - 5,13225$; c'est-à-dire $y = 234,66445 \approx 235$.

Ainsi, en 2025, le nombre de véhicule à vendre s'estime à 135.

1)



2) Calcul des coordonnées du point moyen G :

											Total
x_i	2,8	3	3,2	3,5	3,7	4	4,3	4,6	4,8	5,1	39
y_i	47	44	40	38	35	28	25	22	17	14	310

$$\bar{x} = \frac{39}{10} = 3,9 \text{ et } \bar{y} = \frac{310}{10} = 31; \text{ donc } G(3,9; 31).$$

3) a) Voir représentation : il s'agit d'un tracer approximatif. Toutefois, on veillera à faire passer cette droite tracée par le point moyen G.

b) Détermination d'une équation de la droite de régressions de y en x par la méthode des moindres carrés :

La droite de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x;y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x}.$$

Calcul des variances et de la covariance :

Méthode 1						Méthode 2					
$V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2 ;$						$V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} ;$					
$V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2 ;$						$V(y) = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n} ;$					
$\text{Cov}(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y}.$						$\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n} .$					
x_i	y_i	x_i^2	y_i^2	$x_i y_i$		x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	
2,8	47	7,84	2209	131,6		2,8	47	1,21	256	-17,6	
3	44	9	1936	132		3	44	0,81	169	-11,7	
3,2	40	10,24	1600	128		3,2	40	0,49	81	-6,3	
3,5	38	12,25	1444	133		3,5	38	0,16	49	-2,8	
3,7	35	13,69	1225	129,5		3,7	35	0,04	16	-0,8	
4	28	16	784	112		4	28	0,01	9	-0,3	
4,3	25	18,49	625	107,5		4,3	25	0,16	36	-2,4	
4,6	22	21,16	484	101,2		4,6	22	0,49	81	-6,3	
4,8	17	23,04	289	81,6		4,8	17	0,81	196	-12,6	
5,1	14	26,01	196	71,4		5,1	14	1,44	289	-20,4	
Total	39	310	157,72	10792	1127,8	Total	39	310	5,62	1182	-81,2
$V(x) = \frac{157,72}{10} - 3,9^2 = 0,562;$						$V(x) = \frac{5,62}{10} = 0,562;$					

$$V(y) = \frac{10792}{10} - 31^2 = 118,2;$$

$$\text{Cov}(x ; y) = \frac{1127,8}{10} - 3,9 \times 31 = -8,12.$$

$$V(y) = \frac{1182}{10} = 118,2;$$

$$\text{Cov}(x ; y) = \frac{-81,2}{10} = -8,12.$$

On obtient : $a = \frac{\text{Cov}(x;y)}{V(x)} = \frac{-8,12}{0,562} \approx -14,4484$

et $b = \bar{y} - a\bar{x} = 31 - (-14,4484) \times 3,9 = 87,34876$.

Une équation de la droite de régression de y en x est : $y = -14,4484x + 87,34876$.

4) Pour un forfait de 5 400 francs, c'est-à-dire pour $x=5,4$ (exprimé en milliers de francs), le nombre y de clients vérifie : $y = -14,4484 \times 5,4 + 87,34876 = 9,3274$.

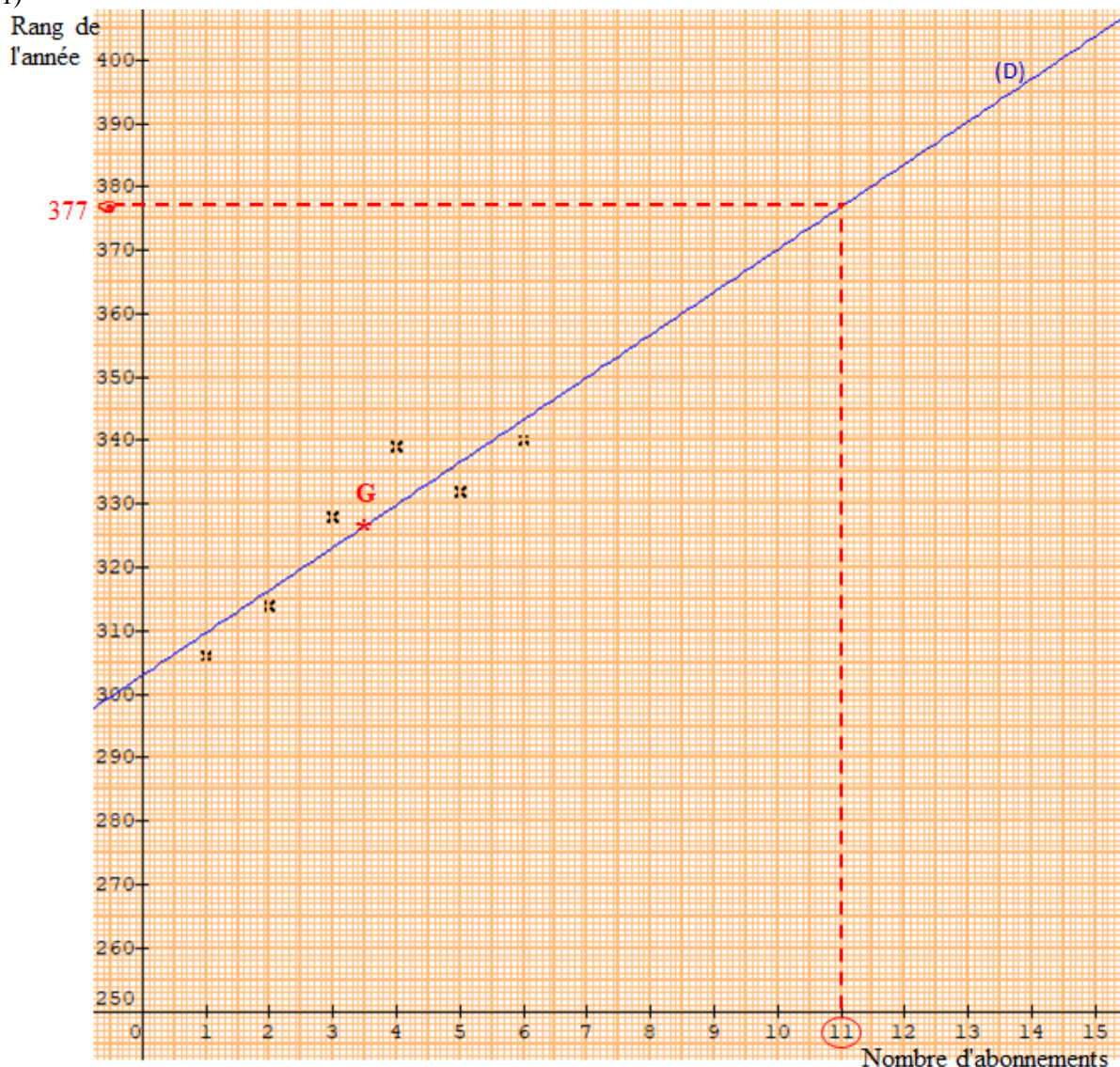
Ainsi, on peut estimer à 9 le nombre de client(e)s intéressé(e)s par un forfait de 5 400 francs.

30 (A1 seulement)

NB : Dans l'énoncé, lire pour la consigne 3) « $y = \frac{47}{7}x + 303$ »

au lieu de « $y = 6,71x + 303$ ».

1)



2) Calcul des coordonnées du point moyen G :

	Total						
x_i	1	2	3	4	5	6	21
y_i	306	314	328	339	332	340	1959

$\bar{x} = \frac{21}{6} = \frac{7}{2}$ et $\bar{y} = \frac{1959}{6} = \frac{653}{2}$; donc $G(3,5; 326,5)$.

3) Détermination d'une équation de la droite de régressions de y en x par la méthode des moindres carrés :

La droite de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x;y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x}.$$

Calcul des variances et de la covariance :

Méthode 1	Méthode 2																																																																
$V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2 ;$ $\text{Cov}(x ; y) = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} - \bar{x}\bar{y}.$ <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>x_i^2</th> <th>x_iy_i</th> </tr> </thead> <tbody> <tr><td>1</td><td>306</td><td>1</td><td>306</td></tr> <tr><td>2</td><td>314</td><td>4</td><td>628</td></tr> <tr><td>3</td><td>328</td><td>9</td><td>984</td></tr> <tr><td>4</td><td>339</td><td>16</td><td>1356</td></tr> <tr><td>5</td><td>332</td><td>25</td><td>1660</td></tr> <tr><td>6</td><td>340</td><td>36</td><td>2040</td></tr> <tr> <td>Total</td> <td>21</td> <td>1959</td> <td>6974</td> </tr> </tbody> </table> $V(x) = \frac{91}{6} - 3,5^2 = \frac{35}{12};$ $\text{Cov}(x ; y) = \frac{6974}{6} - \frac{7}{2} \times \frac{653}{2} = \frac{6974}{6} - \frac{4571}{4}$ $\text{Cov}(x ; y) = \frac{235}{12}.$	x_i	y_i	x_i^2	x_iy_i	1	306	1	306	2	314	4	628	3	328	9	984	4	339	16	1356	5	332	25	1660	6	340	36	2040	Total	21	1959	6974	$V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n};$ $\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}.$ <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>$(x_i - \bar{x})^2$</th> <th>$(x_i - \bar{x})(y_i - \bar{y})$</th> </tr> </thead> <tbody> <tr><td>1</td><td>306</td><td>6,25</td><td>51,25</td></tr> <tr><td>2</td><td>314</td><td>2,25</td><td>18,75</td></tr> <tr><td>3</td><td>328</td><td>0,25</td><td>-0,75</td></tr> <tr><td>4</td><td>339</td><td>0,25</td><td>6,25</td></tr> <tr><td>5</td><td>332</td><td>2,25</td><td>8,25</td></tr> <tr><td>6</td><td>340</td><td>6,25</td><td>33,75</td></tr> <tr> <td>Total</td> <td>21</td> <td>17,5</td> <td>117,5</td> </tr> </tbody> </table> $V(x) = \frac{17,5}{6} = \frac{35}{12};$ $\text{Cov}(x ; y) = \frac{117,5}{6} = \frac{235}{12};$ $\text{Cov}(x ; y) = \frac{235}{12}.$	x_i	y_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	1	306	6,25	51,25	2	314	2,25	18,75	3	328	0,25	-0,75	4	339	0,25	6,25	5	332	2,25	8,25	6	340	6,25	33,75	Total	21	17,5	117,5
x_i	y_i	x_i^2	x_iy_i																																																														
1	306	1	306																																																														
2	314	4	628																																																														
3	328	9	984																																																														
4	339	16	1356																																																														
5	332	25	1660																																																														
6	340	36	2040																																																														
Total	21	1959	6974																																																														
x_i	y_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$																																																														
1	306	6,25	51,25																																																														
2	314	2,25	18,75																																																														
3	328	0,25	-0,75																																																														
4	339	0,25	6,25																																																														
5	332	2,25	8,25																																																														
6	340	6,25	33,75																																																														
Total	21	17,5	117,5																																																														

On obtient : $a = \frac{\text{Cov}(x;y)}{V(x)} = \frac{\frac{235}{12}}{\frac{35}{12}} = \frac{235}{35} = \frac{47}{7}$

et $b = \bar{y} - a\bar{x} = \frac{653}{2} - \frac{47}{7} \times \frac{7}{2} = \frac{653}{2} - \frac{47}{2} = \frac{606}{2} = 303.$

Une équation de la droite de régression de y en x est : $y = \frac{47}{7}x + 303.$

4) a) G : voir représentation.

Construction de la droite (L) : Voir représentation.

Si $x=0$, alors $y=6,71 \times 0 + 303 = 303.$

Si $x=10$, alors $y=6,71 \times 10 + 303 = 370,1.$

La droite de régression passe par les points de coordonnées (0 ; 303) et (10 ; 370,1).

b) Le point moyen $G(3,5 ; 326,5)$ appartient évidemment à la droite de régression.

5) L'an 2025 correspond à l'année de rang 11.

Pour $x=11$; on lit graphiquement $y = 377.$

Ainsi, le nombre d'abonnement annuels s'estime graphiquement à 377

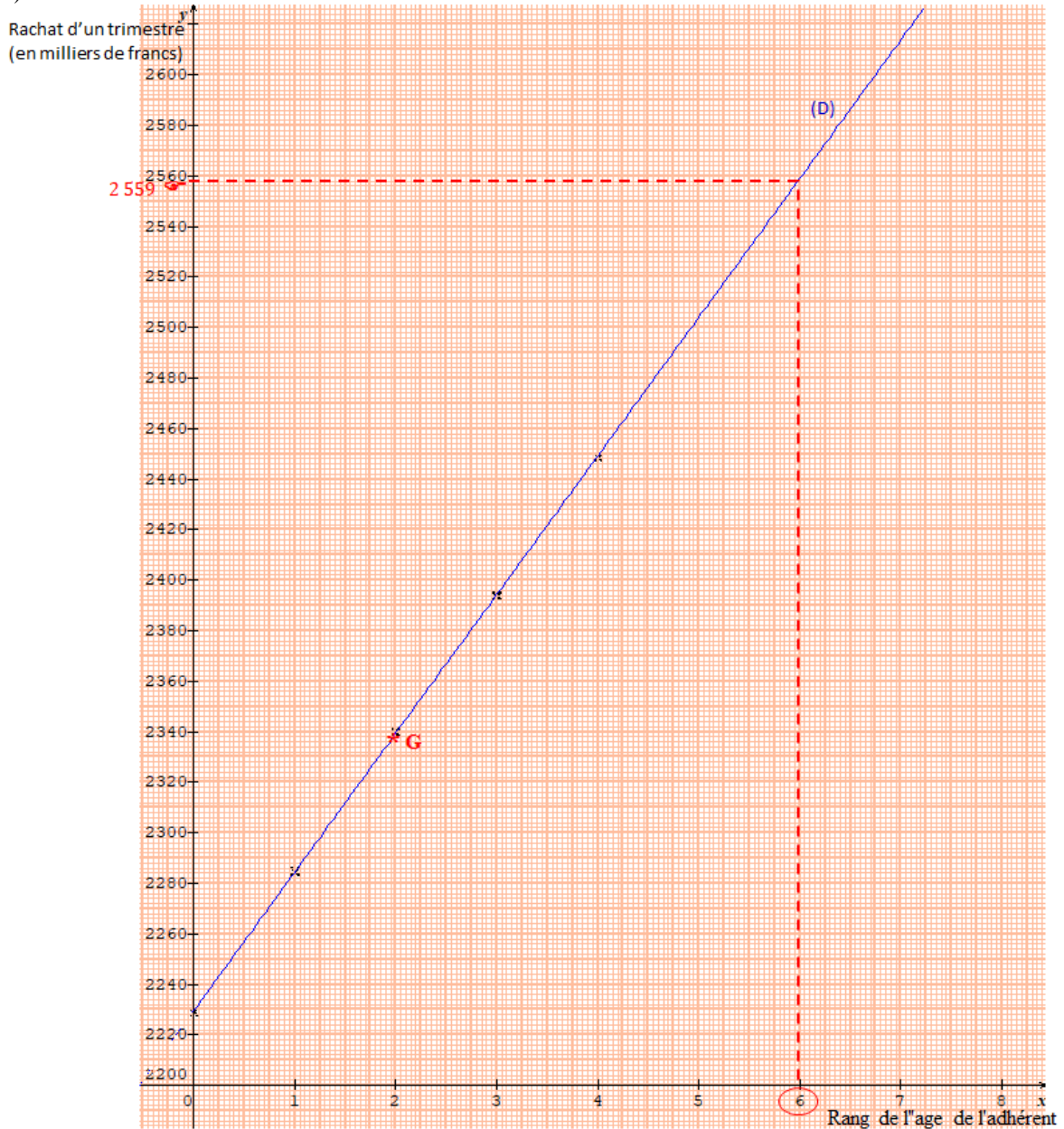
Vérification par le calcul :

Pour l'année de rang $x=11$, le nombre y d'abonnement annuels vérifie :

$$y = \frac{47}{7} \times 11 + 303 ; \text{ c'est-à-dire } y \approx 377.$$

Ainsi, en 2025, le nombre d'abonnement annuels s'estime effectivement à 377.

1)



2) a) Les points du nuage sont presque alignés ; donc un ajustement par une droite se justifie.

b) Vérifions qu'un ajustement linéaire est justifié :

Le coefficient de corrélation : $r = \frac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}}$.

Calculons d'abord les variances et la variance :

Méthode 1	Méthode 2																				
$V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2 ;$	$V(x) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} ;$																				
$V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2 ;$	$V(y) = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n} ;$																				
$\text{Cov}(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y}.$	$\text{Cov}(x ; y) = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}.$																				
<table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>x_i^2</th> <th>y_i^2</th> <th>$x_i y_i$</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>2229</td> <td>0</td> <td>4968441</td> <td>0</td> </tr> </tbody> </table>	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	0	2229	0	4968441	0	<table border="1"> <thead> <tr> <th>x_i</th> <th>y_i</th> <th>$(x_i - \bar{x})^2$</th> <th>$(y_i - \bar{y})^2$</th> <th>$(x_i - \bar{x})(y_i - \bar{y})$</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>2229</td> <td>4</td> <td>12188,16</td> <td>220,8</td> </tr> </tbody> </table>	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	0	2229	4	12188,16	220,8
x_i	y_i	x_i^2	y_i^2	$x_i y_i$																	
0	2229	0	4968441	0																	
x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$																	
0	2229	4	12188,16	220,8																	

	1	2285	1	5221225	2285		1	2285	1	2959,36	54,4
	2	2340	4	5475600	4680		2	2340	0	0,36	0
	3	2394	9	5731236	7182		3	2394	1	2981,16	54,6
	4	2449	16	5997601	9796		4	2449	4	12012,16	219,2
Total	10	11697	30	27394103	23943	Total	10	11697	10	30141,2	549

$$\text{Moyenne en } x : \bar{x} = \frac{10}{5} = 2;$$

$$\text{Moyenne en } y : \bar{y} = \frac{11\,697}{5} = 2339,4.$$

$$V(x) = \frac{30}{5} - 2^2 = 2;$$

$$V(y) = \frac{27\,394\,103}{5} - 2\,339,4^2 = 6028,24;$$

$$\text{Cov}(x ; y) = \frac{23\,943}{5} - 2 \times 2339,4 = 109,8.$$

$$\text{Moyenne en } x : \bar{x} = \frac{10}{5} = 2;$$

$$\text{Moyenne en } y : \bar{y} = \frac{11\,697}{5} = 2339,4.$$

$$V(x) = \frac{10}{5} = 2;$$

$$V(y) = \frac{30\,141,2}{5} = 6028,24;$$

$$\text{Cov}(x ; y) = \frac{549}{5} = 109,8.$$

$$\text{On obtient : } r = \frac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}} = \frac{109,8}{\sqrt{2 \times 6\,028,24}} \approx 0,99998175.$$

$0,87 \leq r \leq 1$, donc il y a une bonne corrélation ou une forte corrélation entre les deux variables. Ainsi, un ajustement affine est approprié.

c) Détermination d'une équation de la droite (D) de régressions de y en x par la méthode des moindres carrés :

La droite (D) de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x,y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x}.$$

$$\text{On obtient : } a = \frac{109,8}{2} = 54,9 \text{ et } b = \bar{y} - a\bar{x} = 2\,339,4 - 54,9 \times 2 = 2\,229,6.$$

Une équation de la droite (D) de régression de y en x est : $y = 54,9x + 2\,229,6$.

d) (D) passe évidemment par le point moyen G(2 ; 2339,4).

Si $x=0$, alors $y=54,9 \times 0 + 2\,229,6 = 2\,229,6$.

(D) passe également par le point de coordonnées (0 ; 2 229,6).

Construction de la droite (D) : Voir représentation.

3) L'âge 64 ans correspond au rang $x=6$.

Méthode graphique :

Pour $x=6$; on lit graphiquement $y = 2\,559$ exprimé en milliers de francs.

Ainsi, pour un salarié âgé de 64 ans, le montant du rachat d'un trimestre de cotisation s'estime à 2 559 000 francs.

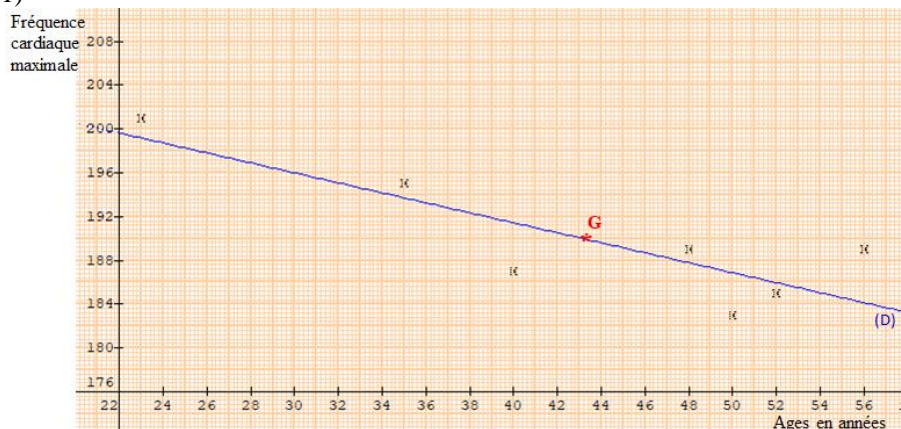
Méthode algébrique :

Pour l'année de rang $x=6$, le montant y du rachat d'un trimestre de cotisation en milliers de francs, vérifie : $y = 54,9 \times 6 + 2\,229,6$; c'est-à-dire $y = 2\,559$.

Ainsi, pour un salarié âgé de 64 ans, le montant du rachat d'un trimestre de cotisation s'estime à 2 559 000 francs.

32 (A1 seulement)

1)



2) a) Le coefficient de corrélation : $r = \frac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}}$.

Calculons d'abord les variances et la variance :

$$V(x) = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} - \bar{x}^2 ;$$

$$V(y) = \frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} - \bar{y}^2 ;$$

$$\text{Cov}(x ; y) = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} - \bar{x} \bar{y}.$$

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
23	201	529	40401	4623
35	195	1225	38025	6825
40	187	1600	34969	7480
48	189	2304	35721	9072
50	183	2500	33489	9150
52	185	2704	34225	9620
56	189	3136	35721	10584
Total	304	1329	13998	57354

$$\text{Moyennes : } \bar{x} = \frac{304}{7} \text{ et } \bar{y} = \frac{1329}{7}.$$

$$V(x) = \frac{13998}{7} - \left(\frac{304}{7}\right)^2 = \frac{5570}{49} ;$$

$$V(y) = \frac{252551}{7} - \left(\frac{1329}{7}\right)^2 = \frac{1616}{49} ;$$

$$\text{Cov}(x ; y) = \frac{57354}{7} - \frac{304}{7} \times \frac{1329}{7} = \frac{401478}{49} - \frac{404016}{49} = -\frac{2538}{49}.$$

$$\text{Dans tous les cas, on obtient } \frac{\text{Cov}(x,y)}{\sqrt{V(x)V(y)}} = \frac{-\frac{2538}{49}}{\sqrt{\frac{5570}{49} \times \frac{1616}{49}}} = \frac{-2538}{\sqrt{5570 \times 1616}}.$$

C'est-à-dire : $r \approx -0,845947$.

b) $0,87 \leq |r| \leq 1$, donc on a l'interprétation suivante :

« Il y a une bonne corrélation ou une forte corrélation entre l'âge et la fréquence cardiaque maximale ».

c) Détermination d'une équation de la droite de régressions de y en x par la méthode des moindres carrés :

La droite (D) de régression de y en x admet une équation de la forme $y = ax + b$ avec :

$$a = \frac{\text{Cov}(x;y)}{V(x)} \text{ et } b = \bar{y} - a\bar{x}.$$

$$\text{On obtient : } a = \frac{-\frac{2538}{49}}{\frac{5570}{49}} = -\frac{2538}{5570} = -\frac{1269}{2785} \text{ et } b = \bar{y} - a\bar{x} = \frac{1329}{7} - \left(-\frac{1269}{2785}\right) \times \frac{304}{7} = \frac{4087041}{19495}.$$

$$\text{Une équation de la droite (D) de régression de y en x est : } y = -\frac{1269}{2785}x + \frac{4087041}{19495}.$$

3) (D) passe évidemment par le point moyen $G\left(\frac{304}{7} ; \frac{1329}{7}\right)$.

$$\text{Si } x=0, \text{ alors } y = -\frac{1269}{2785} \times 0 + \frac{4087041}{19495} = \frac{4087041}{19495} \approx 209,64560.$$

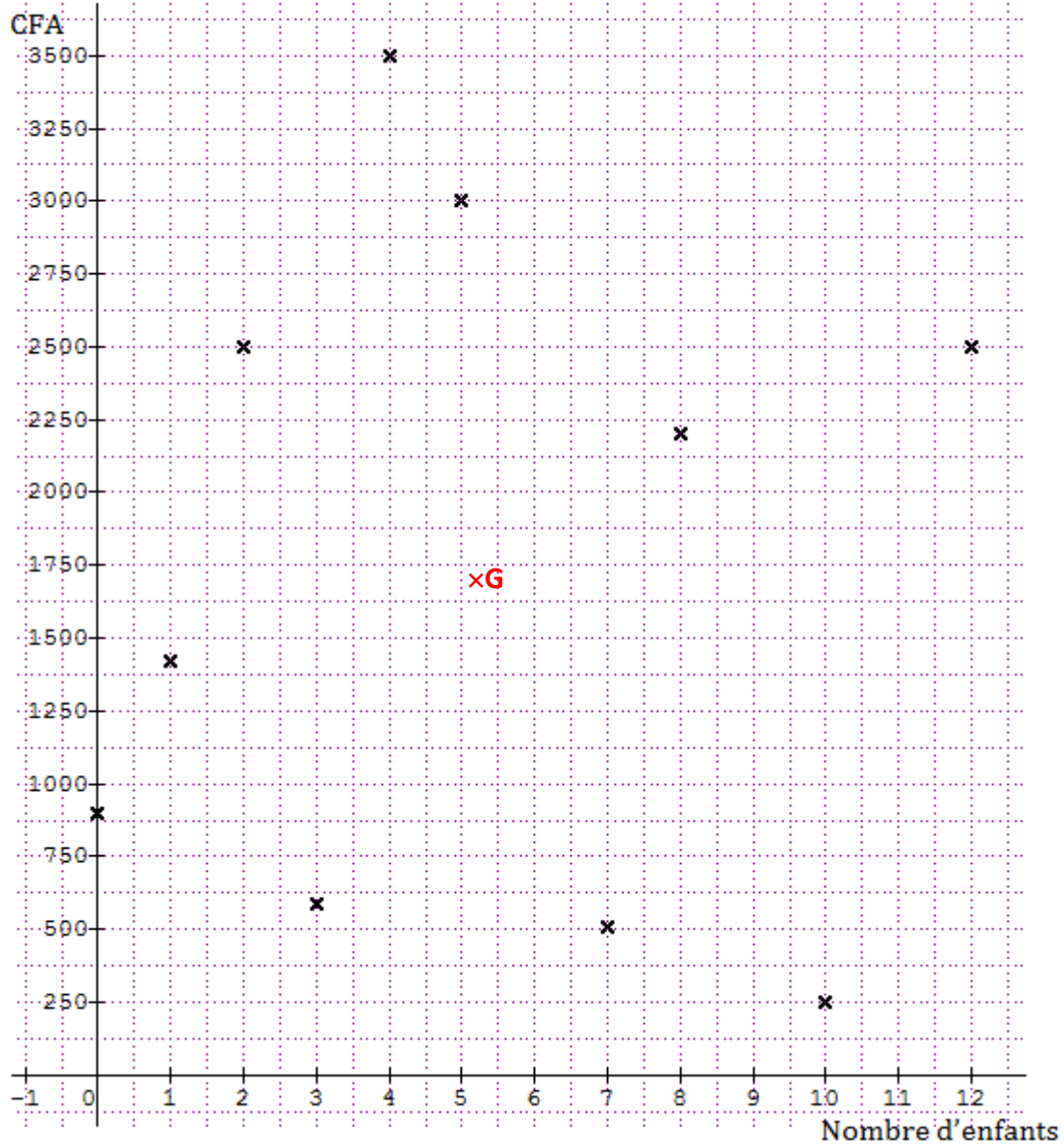
(D) passe également par le point de coordonnées $\left(0 ; \frac{4087041}{19495}\right)$.

Construction de la droite (D) : Voir représentation.

33

1)

Salaire
en milliers de francs



2) Point moyen G :

x	y
4	3500
3	590
2	2500
1	1420
0	900
5	3000
10	250
7	510
12	2500
8	2200
Total	52
	17370

$\bar{x} = \frac{52}{10} = 5,2$ et $\bar{y} = \frac{17370}{10} = 1737$; donc $G(5,2 ; 1737)$.

3) Les points du nuage sont tellement dispersés qu'on ne peut ajuster le nuage de points ; il n'y a aucune dépendance entre le nombre d'enfants d'un couple et son salaire.

1) 1-a) Nuage de points : Voir ci-après.

1-b) Calcul des coordonnées du point moyen G :

											Total
x_i	1400	1600	1800	2200	1800	2000	2500	3200	3000	4000	23500
y_i	4,1	4,2	4,3	4,6	4,3	4,5	4,9	5,4	5,3	5,9	47,5

$\bar{x} = \frac{23\ 500}{10} = 2\ 350$ et $\bar{y} = \frac{47,5}{10} = 4,75$; donc $G(2\ 350 ; 4,75)$.

2) Le nuage de point est allongé et les points semblent alignés, c'est pourquoi il est possible de faire un ajustement affine de la série statistique.

3) 3-a) Construction cette droite (D) de Mayer :

	Série 1					Total
x_i	1400	1600	1800	2200	1800	8 800
y_i	4,1	4,2	4,3	4,6	4,3	21,5

	Série 2					Total
x_i	2000	2500	3200	3000	4000	14 700
y_i	4,5	4,9	5,4	5,3	5,9	26

$x_{G_1} = \frac{8\ 800}{5} = 1\ 760$ et $y_{G_1} = \frac{21,5}{5} = 4,3$; $x_{G_2} = \frac{14\ 700}{5} = 2\ 940$ et $y_{G_2} = \frac{26}{5} = 5,2$;

Donc $G_1(1\ 760; 4,3)$.

Donc $G_2(2\ 940; 5,2)$.

La droite (D) passe par les points G_1 et G_2 .

3-b) Détermine une équation de la droite (D).

Une équation de la droite d'ajustement (D), c'est à dire de la droite (G_1G_2) est de la forme : $y = ax + b$.

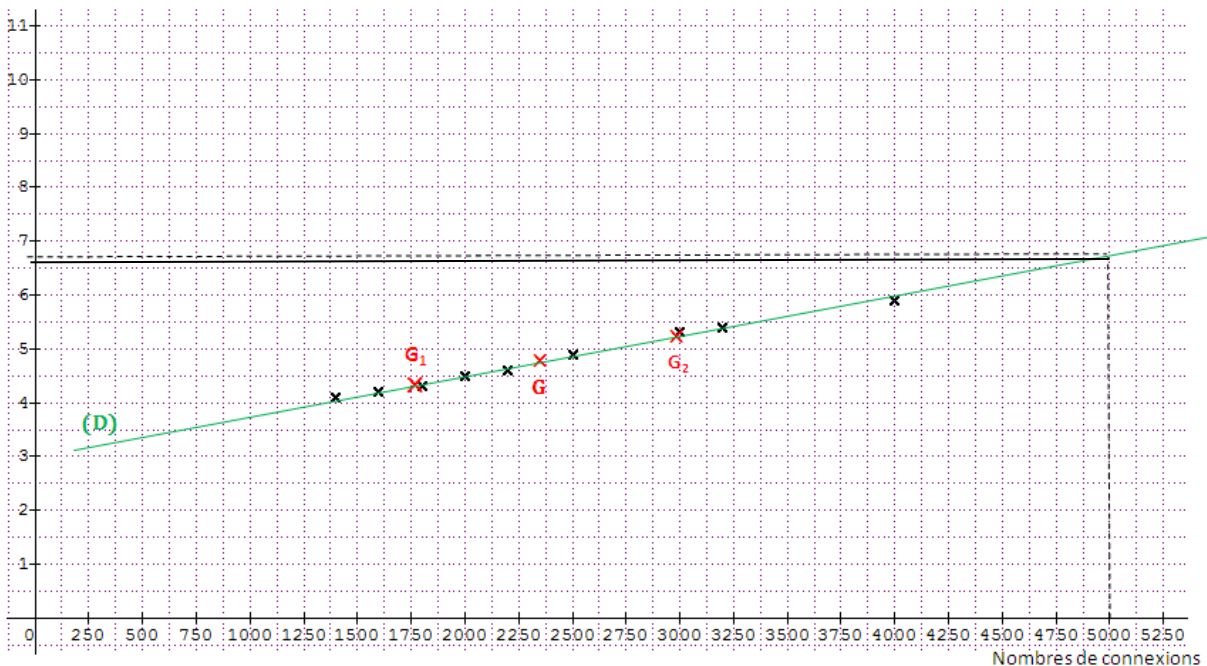
La droite (D) a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{5,2 - 4,3}{2\ 940 - 1\ 760} = \frac{0,9}{1\ 180} = 0,00076$.

La droite (D) passe par $G(2\ 350 ; 4,75)$, donc : $4,75 = 0,00076 \times 2\ 350 + b$; d'où $b = 2,964$.

Une équation de la droite (D) est $y = 0,000\ 76x + 2,964$.

4)4-a)

Chiffre d'affaires
en millions de F CFA



On acceptera 6,7 millions ou 6,8 millions.

4-b) La feuille ne permettant pas de faire une estimation graphique, le calcul est le moyen qui convient ici :

Dix millions de francs CFA de chiffre d'affaire correspond à $y=10$.

Pour $y=10$, on obtient $10 = 0,00076x + 2,964$; c'est-à-dire $x = 9257,89$

Ainsi, Pour espérer un chiffre d'affaire de dix millions de F CFA, le nombre de connexion nécessaire peut s'estimer à 9 258 connexions.

35 (A1 seulement):

1) Tableau de la série marginale associé à y :

Y	Moins de 18 ans	18 ans	19 ans	Plus de 19 ans	Total
Effectifs	133	420	182	62	797

2) Tableau de la série marginale associé à x :

X	2 ans	3 ans	4 ans	Total
Effectifs	400	274	123	797

3) Tableau fréquences marginales du caractère y :

En utilisant les données du tableau de la question 1 et la formule

« Fréquence d'une modalité = $\frac{\text{Effectif de la modalité}}{\text{Effectif total}}$ », on effectue les calcul et on établit le tableau :

Y	Moins de 18 ans	18 ans	19 ans	Plus de 19 ans	Total
Fréquences	16,7 %	52,7 %	22,8 %	7,8 %	100 %

36 (A1 seulement):

1) Nuage de points : Voir graphique ci-après.

2) Calcul des coordonnées du point moyen G :

	3	4	5	6	7	8	9	10	Total
x_i	3	4	5	6	7	8	9	10	52
y_i	10	12	20	23	26	30	32	35	188

$\bar{x} = \frac{52}{8} = 6,5$ et $\bar{y} = \frac{188}{8} = 23,5$; donc $G(6,5 ; 23,5)$.

3) 3-a) Calcul du coefficient de corrélation linéaire r :

	9	16	25	36	49	64	81	100	Total
x_i^2	9	16	25	36	49	64	81	100	380
y_i^2	100	144	400	529	676	900	1024	1225	4998
$x_i y_i$	30	48	100	138	182	240	288	350	1376

Variance de x : $V(x) = \frac{380}{8} - 6,5^2 = 5,25$.

Variance de y : $V(y) = \frac{4998}{8} - 23,5^2 = 72,5$.

Covariance : $Cov(x, y) = \frac{1376}{8} - 6,5 \times 23,5 = 19,25$.

Le coefficient de corrélation linéaire : $r = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} = \frac{19,25}{\sqrt{5,25 \times 72,5}} \approx 0,987$.

3-b) $0,87 \leq r < 1$: il y a une bonne corrélation ou une forte corrélation entre la puissance d'une voiture et sa consommation d'essence.

4) Equation de la droite de régression de y en x :

Nous allons déterminer par la méthode des moindres carrés, l'équation sous la forme : $y = ax + b$ avec $a = \frac{Cov(x,y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$.

On a :

• $a = \frac{Cov(x,y)}{V(x)} = \frac{19,25}{5,25} = \frac{11}{3} \approx 3,667$.

• $b = \bar{y} - a\bar{x} \approx 23,5 - 3,667 \times 6,5 \approx -0,336$.

• Equation de la droite de régression de y en x : $y = 3,667x - 0,336$.

5) 5-a) • Méthode graphique :

On construit d'abord la droite de régression de y en x.

Pour $x = 12$, on peut accepter l'une des valeurs lues graphiquement : $y \approx 43$ L ; $y \approx 44$ L.

• Méthode algébrique :

Pour $x = 12$, on obtient $y = 3,667 \times 12 - 0,336 = 43,668$.

5-b) • Méthode graphique :

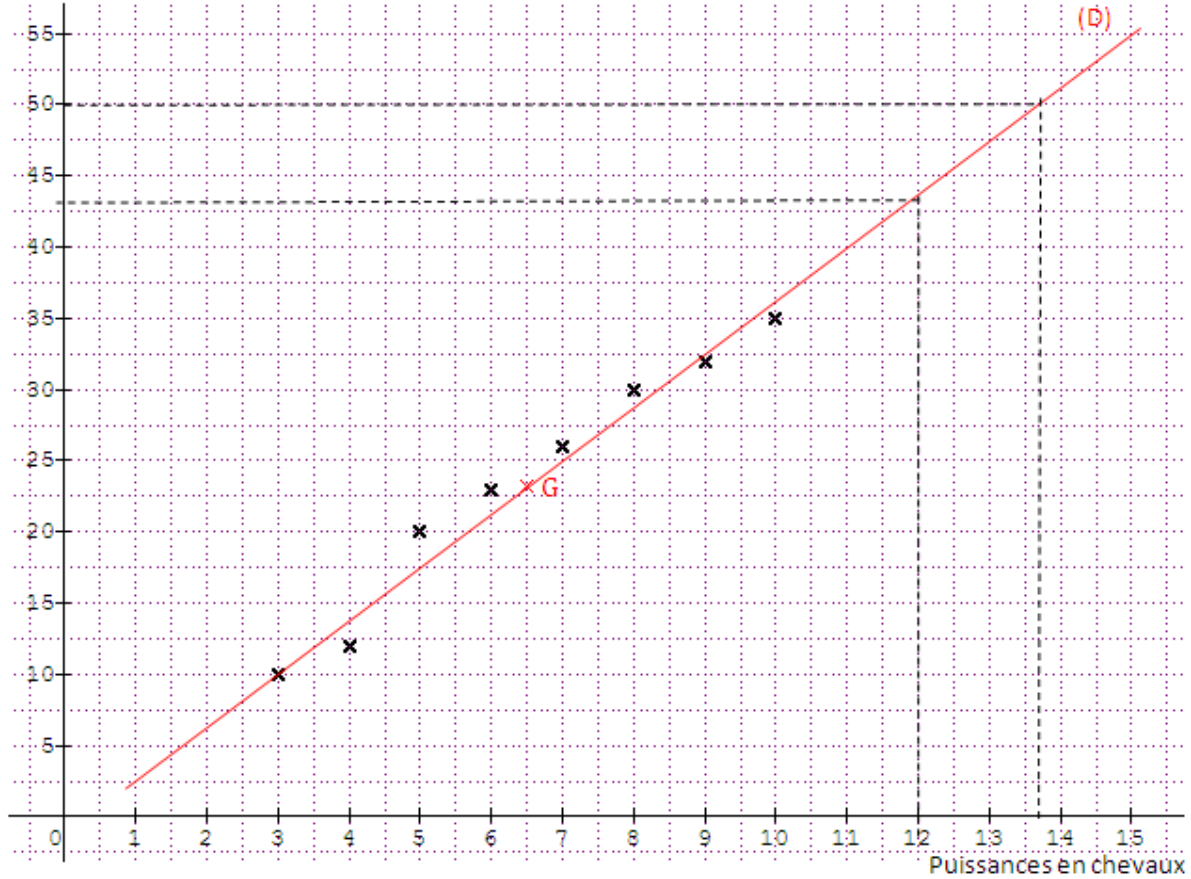
Pour $y = 50$, on peut accepter le résultat graphique : $x \approx 13,7$, soit 14 chevaux..

• Méthode algébrique :

Pour $y = 50$, on obtient $50 = 3,667x - 0,336$,
c'est-à-dire $x \approx 13,73$, soit 14 chevaux.

Quantité d'essence

en litres



37 (A1 seulement):

1) 1- a) Nuage de points : Voir ci-après.

1-b) Calcul des coordonnées du point moyen G :

	Total								
x_i	2	4	6	8	10	12	14	16	72
y_i	10	15	10	20	21	40	32	35	183

$\bar{x} = \frac{72}{8} = 9$ et $\bar{y} = \frac{183}{8} = 22,875$; donc $G(9 ; 22,875)$.

2) Calcul du coefficient de corrélation linéaire r :

	Total								
x_i^2	4	16	36	64	100	144	196	256	816
y_i^2	100	225	100	400	441	1600	1024	1225	5 115
$x_i y_i$	20	60	60	160	210	480	448	560	1 998

Variance de x : $V(x) = \frac{816}{8} - 9^2 = 21$.

Variance de y : $V(y) = \frac{5 115}{8} - 22,875^2 = 116,109$.

Covariance : $Cov(x ; y) = \frac{1 998}{8} - 9 \times 22,875 = 43,875$.

Le coefficient de corrélation linéaire : $r = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} = \frac{43,875}{\sqrt{21 \times 116,109}} \approx 0,889$.

3) 3-a) Equation de la droite (L) de régression de y en x :

Nous allons déterminer par la méthode des moindres carrés, l'équation sous la forme : $y = ax + b$ avec $a =$

$\frac{Cov(x;y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$.

• $a = \frac{Cov(x;y)}{V(x)} = \frac{43,875}{21} \approx 2,089$.

• $b = \bar{y} - a\bar{x} \approx 22,875 - 2,089 \times 9 \approx 4,074$.

• Equation de la droite de régression de y en x : $y = 2,089x + 4,074$

3-b) Construis la droite (L) dans le même repère.

x	20	5,230
y	45,854	15

Si $x=20$, alors $y=2,089 \times 20 + 4,074 = 8,252$.

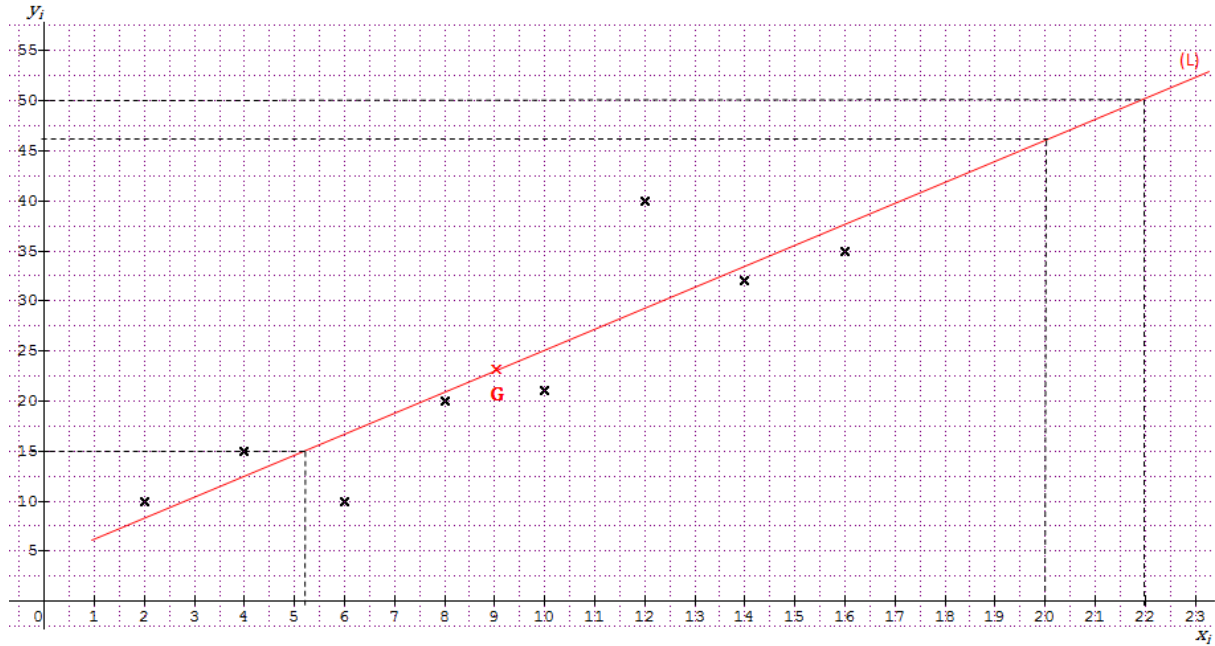
Si $y=15$, alors $15=2,089 \times x + 4,074$, c'est à dire $x \approx 5,230$.

4) • Méthode graphique :

Pour $y = 50$, on acceptera le résultat graphique : $x \approx 21$ ou $x \approx 22$.

• Méthode algébrique :

Pour $y = 50$, on obtient $50 = 2,089 \times x + 4,074$, c'est à dire $x \approx 21,985$.



38

1) Nuage de points : Voir ci-après.

2) Point moyen G :

x	-20°	-15°	-4°	0°	5°	12°	20°	14°	9°	-2°	-10°	-18°	Total
y (mm)	10	10	16	22	31	55	90	65	40	18	16	8	381

$\bar{x} = \frac{-9}{12} = -0,75^\circ$ et $\bar{y} = \frac{381}{12} = 31,75$ mm ; donc $G(-0,75 ; 31,75)$.

3) Construction cette droite (D) de Mayer :

Série 1							Total
x_i	-20°	-15°	-4°	0°	5°	12°	-22
y_i	10	10	16	22	31	55	144

$x_{G_1} = \frac{-22}{6} \approx -3,66$ et $y_{G_1} = \frac{144}{6} = 24$;

Donc $G_1(-3,66 ; 24)$.

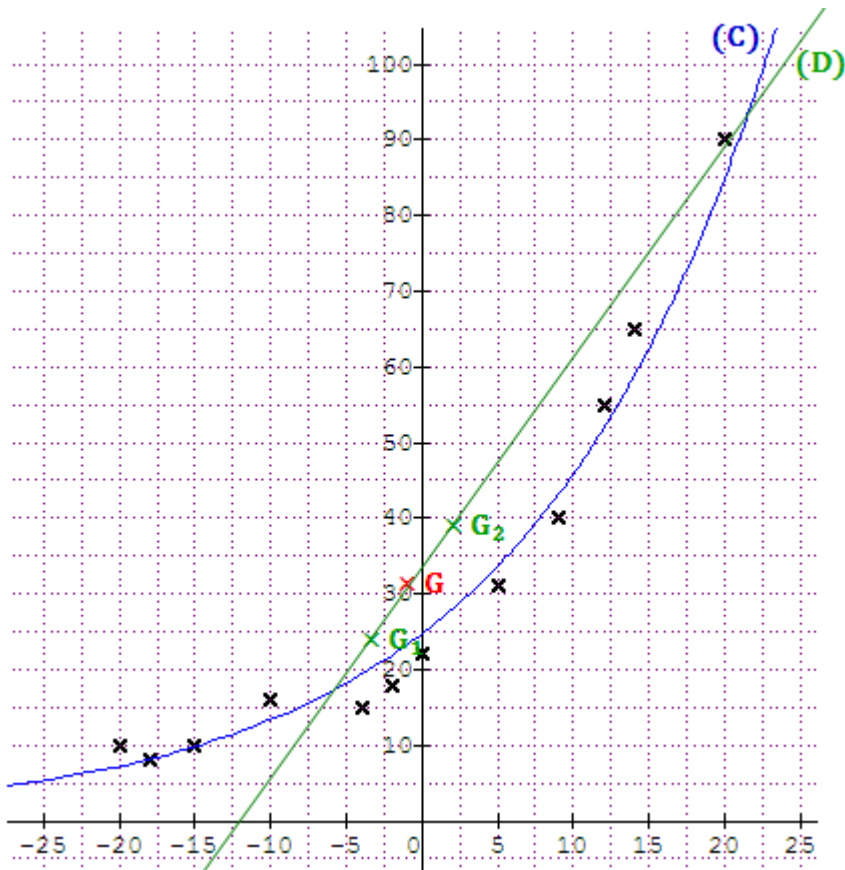
La droite (D) passe par les points G_1 et G_2 .

4)

Série 2							Total
x_i	20°	14°	9°	-2°	-10°	-18°	13
y_i	90	65	40	18	16	8	237

$x_{G_2} = \frac{13}{6} \approx 2,17$ et $y_{G_2} = \frac{237}{6} = 39,5$;

Donc $G_2(2,17 ; 39,5)$.



5) C'est la courbe (C) qui semble constituée le meilleur ajustement du nuage de points. En effet, c'est elle qui passe le plus près des points du nuage ; la disposition des points forme l'allure de cette courbe.

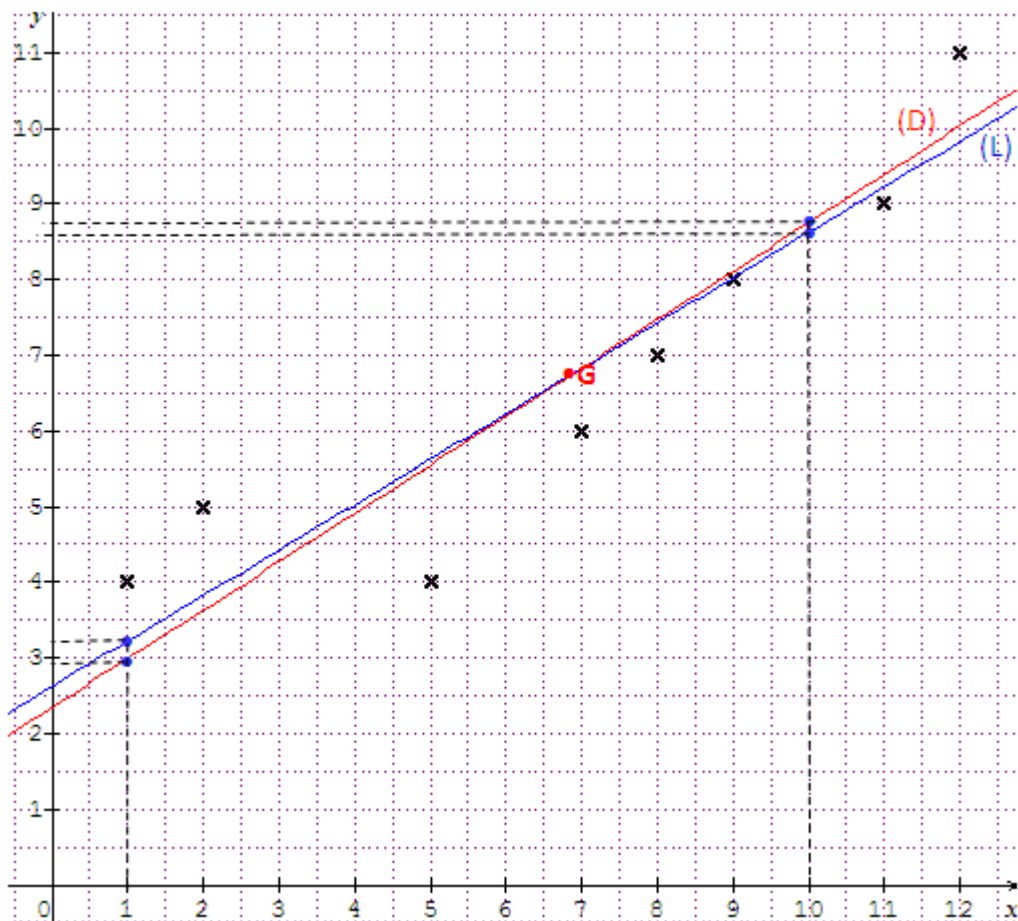
39

1)

(D) : $y = 0,64x + 2,35$		
x	1	10
y	2,99	8,75

(L) : $y = 0,6x + 2,625$		
x	1	10
y	3,225	8,625

G est l'intersection des droites (L) et (D).



2)

Tableau 1

x_i	y_i	Valeur de : $0,64x_i+2,35$	Valeur de : $y_i-(0,64x_i+2,35)$
1	4	2,99	1,01
2	5	3,63	1,37
5	4	5,55	-1,55
7	6	6,83	-0,83
8	7	7,47	-0,47
9	8	8,11	-0,11
11	9	9,39	-0,39
Total			1,41

Tableau 2 :

x_i	y_i	Valeur de : $0,6x_i+2,625$	Valeur de : $y_i-(0,6x_i+2,625)$
1	4	3,225	0,775
2	5	3,825	1,175
5	4	5,625	-1,625
7	6	6,825	-0,825
8	7	7,425	-0,425
9	8	8,025	-0,025
11	9	9,225	-0,225
Total			1,275

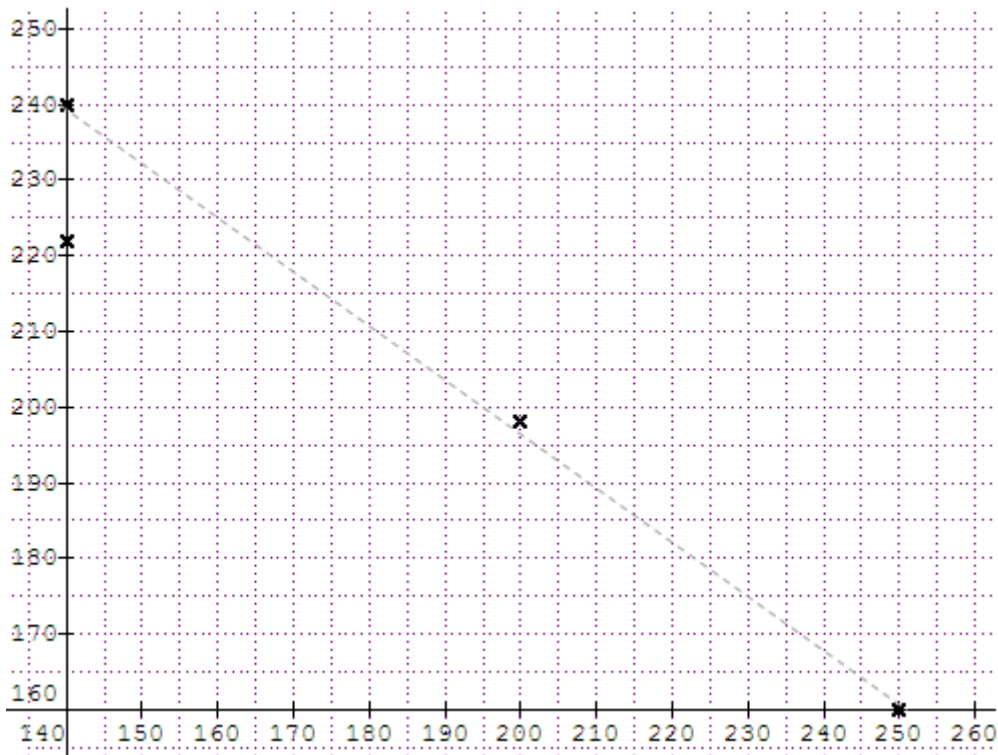
3) 3-a) -0,11 .

3-b) C° et la droite (L).

Justification : Lorsqu'on choisit la droite (L) comme droite d'ajustement, le bilan des erreurs engendrées est 1,275, qui est plus petit que le bilan des erreurs (1,41) lorsqu'on choisit (D).

40

1) 1-a) Nuage de points M_i de coordonnées $(x_i; y_i)$:



1-b) Le nuage de point a une forme allongée et la majorité des points semblent alignés ; un ajustement affine est donc possible.

2)

	Série 1		Total
x_i	200	140	340
y_i	198	240	438

$$x_{G_1} = \frac{340}{2} = 170 \text{ et } y_{G_1} = \frac{438}{2} = 219.$$

Donc $G_1(170; 219)$.

	Série 2		Total
x_i	140	250	390
y_i	222	160	382

$$x_{G_2} = \frac{390}{2} = 195 \text{ et } y_{G_2} = \frac{382}{2} = 191.$$

Donc $G_2(195; 191)$.

Une équation de la droite de Mayer, c'est à dire de la droite (G_1G_2) est de la forme : $y = ax + b$.

La droite (G_1G_2) a pour coefficient directeur $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = \frac{191 - 219}{195 - 170} = \frac{-28}{25} = -1,12$.

La droite (G_1G_2) passe par $G_1(170; 219)$, donc : $219 = -1,12 \times 170 + b$; d'où $b = 409,4$.

Une équation de la droite (G_1G_2) c'est à dire (L) , est $y = -1,12x + 409,4$.

3) 3-a) Chiffre d'affaire = Recette issue des ventes de machine

= Nombre de machines vendues \times Prix d'une machine.

Donc : $s(x) = y \times x$

$$s(x) = (-1,12x + 409,4) \times x$$

$$s(x) = -1,12x^2 + 409,4x$$

3-b) s est dérivable et pour tout x appartenant à $[140; 250]$, on a :

$$s'(x) = -2,24x + 409,4$$

$$s'(x) = 0 \text{ équivaut à : } x \approx 182,768.$$

x	140	$\frac{409,4}{2,24}$	250
Signe de $s'(x)$	+	0	-

s est croissante sur l'intervalle $[140; \frac{409,4}{2,24}]$; s est décroissante sur l'intervalle $[\frac{409,4}{2,24}; 250]$.

4)

• $s(x)$ est maximal pour $s'(x) = 0$, c'est-à-dire : $x \approx 182,768$.

Ainsi, le chiffre d'affaire $s(x)$ est maximal pour le prix estimatif de vente d'une machine fixé à $182,768 \times 1\,000$, c'est-à-dire 182 768 francs CFA.

• Ce chiffre d'affaire vaut :

$$\begin{aligned} s(182,768) &= -1,12 \times 182,768^2 + 409,4 \times 182,768 \\ &= -33\,404,141\,824 + 74\,825,219\,2 \end{aligned}$$

= 41 421,077 376 (exprimé en milliers de francs CFA).

Ainsi, le chiffre d'affaire maximal s'estime à : 41 421 077,376 F CFA.

• Le nombre d'exemplaires vendus correspondant est :

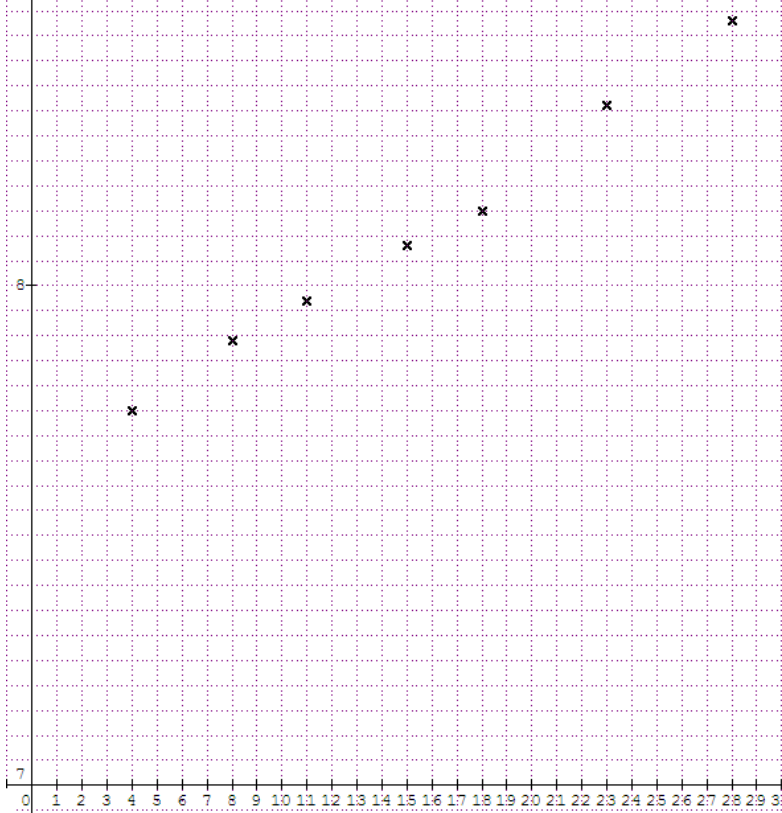
$y = -1,12 \times 182,768 + 409,4$, c'est-à-dire 205 exemplaires.

41

1) $y_i = \ln(N_i)$.

x_i	4	8	11	15	18	23	28
$\ln(N_i)$	7,75	7,89	7,97	8,08	8,15	8,36	8,53

2) 2-a)



2-b) Point moyen G :

									Total
x	4	8	11	15	18	23	28		107
$\ln(N_i)$	7,75	7,89	7,97	8,08	8,15	8,36	8,53		56,73

$\bar{x} = \frac{107}{7} \approx 15,28$ et $\bar{y} = \frac{56,73}{7} \approx 8,10$; donc $G(15,28; 8,10)$.

2-c) Le nuage de point a une forme allongée et la majorité des points semblent alignés ; un ajustement affine est donc possible.

3) 3-a) Pour $x = 38$, on a : $y = 0,0325 \times 38 + 7,62$, c'est-à-dire $y = 8,855$.

3-b) $x = 38$ jours en comptant à partir du 1^{er} avril nous situe au 8 mai.

Ainsi, au 8 mai, on a : $y = 8,855$.

c'est-à-dire : $\ln(N) = 8,855$

$e^{\ln(N)} = e^{8,855}$

$N \approx 7009$.

Le nombre de cas prévisible le 8 mai s'estime à 7 009.

3-c) On a : $y = 0,0325x + 7,62$; donc $\ln(N) = 0,0325x + 7,62$;

$$N = e^{0,0325x+7,62}$$

$$N = e^{0,0325x+7,62}$$

3-d) $N > 10\,000$ équivaut à $e^{0,0325x+7,62} > 10\,000$

$$\ln(e^{0,0325x+7,62}) > \ln(10\,000)$$

$$0,0325x + 7,62 > \ln(10^4)$$

$$x > \frac{4 \ln(10) - 7,62}{0,0325}$$

$$x > 48,933 \dots$$

Comme les valeurs de x sont entières, sa valeur qui convient est 49 ; ce qui correspond au 19 mai : C'est à partir du 19 mai que le nombre prévisible N de cas déclarés excède 10 000.

42

1) 1-a) Abscisse du point moyen $G : x_G = \bar{x} = \frac{0+1+2+3+4}{5} = 2$.

1-b) Ordonnée de G :

La droite d'ajustement (D) passe par le point moyen $G(x_G ; y_G)$; donc :

$$y_G = 57,3x_G + 516,2$$

$$y_G = 57,3 \times 2 + 516,2$$

$$y_G = 630,8.$$

1-c) Valeur du chiffre d'affaire y_3 :

$$y_G = 630,8 \text{ se traduit également par : } \frac{504+580+644+y_3+735}{5} = 630,8, \text{ c'est-à-dire}$$

$$\frac{2463+y_3}{5} = 630 ; \text{ d'où } y_3 = 691.$$

2) Le chiffre d'affaire « un milliard quatre cent trente-trois millions » peut s'écrire : « 1 433 millions », ce qui correspond à $y = 1\,433$.

$$\text{Le rang } x \text{ de l'année vérifie : } 1\,433 = 57,3x + 516,2.$$

$$\text{Donc : } x = 16.$$

L'année où le chiffre d'affaire de l'entreprise atteindra un milliard quatre cent trente-trois millions de francs CFA est 2016+16, c'est-à-dire 2 032.

Situations complexes

43

Nous allons utiliser les connaissances sur les statistiques pour traiter cette situation.

Tâche à réaliser : Estimation du bénéfice de l'entreprise à la fin de l'année en cours.

Donnée :

Tableau statistique donnant l'évolution des bénéfices de l'entreprise sur la période des six dernières années. (Voir tableau)

Démarche :

Nous allons vérifier qu'un ajustement affine de la série est possible, puis déterminer une droite d'ajustement, enfin estimer le bénéfice de l'entreprise pour l'année de rang 7.

1^{ère} méthode : Nous allons utiliser la méthode de Mayer.

• Nuage de points : Voir représentation ci-dessous.

Les points du nuage de points semblent alignés, c'est pourquoi il est possible de faire un ajustement affine de la série statistique.

• Droite (D) de Mayer :

Série 1				Total
x_i	1	2	3	6
y_i	5 200 000	5 500 000	5 800 000	16 500 000

$$x_{G_1} = \frac{6}{3} = 2 \text{ et } y_{G_1} = \frac{16\,500\,000}{3} = 5\,500\,000 ;$$

Donc $G_1(2 ; 5\,500\,000)$.

Série 2				Total
x_i	4	5	6	15
y_i	6 000 000	6 250 000	6 500 000	18 750 000

$$x_{G_2} = \frac{15}{3} = 5 \text{ et } y_{G_2} = \frac{18\,750\,000}{3} = 6\,250\,000 ;$$

Donc $G_2(5 ; 6\,250\,000)$.

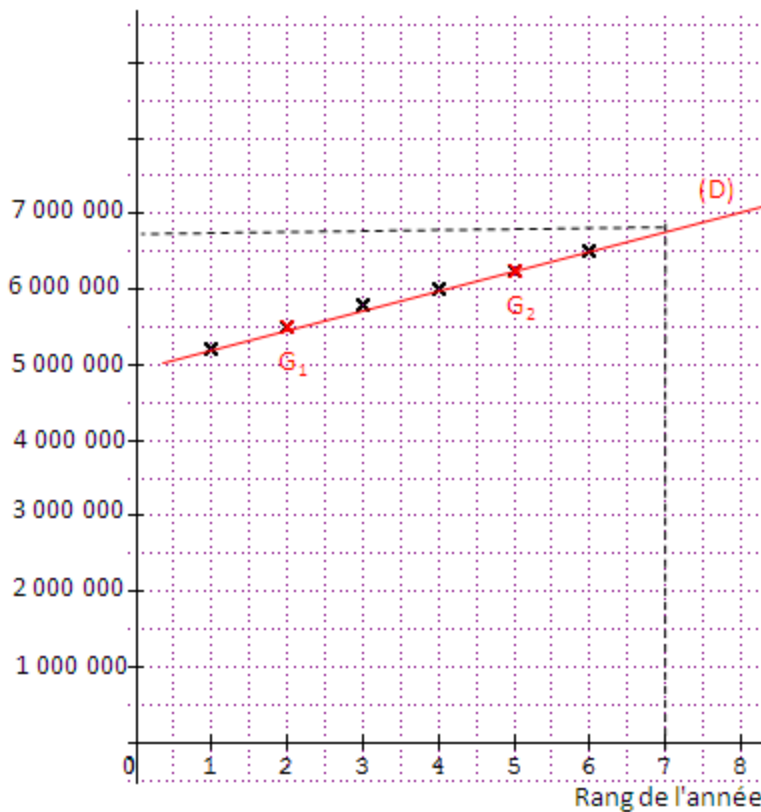
La droite (D) passe par les points G_1 et G_2 .

On construit cette droite dans un repère ou bien on en détermine une équation, et on trouve : $y = 250\,000x + 5\,000\,000$.

• Estimation du bénéfice probable de l'entreprise à la fin de l'année en cours :

- Estimation graphique :

Bénéfice en F CFA



Pour $x = 7$, on acceptera l'un des résultats de lecture graphique suivants : 6 750 000 ; 6 700 000 ; 6 800 000.

- Estimation algébrique : $y = 250\,000x + 5\,000\,000$

Pour $x = 7$, on obtient $y = 250\,000 \times 7 + 5\,000\,000$, c'est à dire $y \approx 6\,750\,000$.

2^{ème} méthode : Nous allons utiliser la méthode des moindres carrés.

• Calcul du coefficient de corrélation linéaire :

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	1	5200000	1	2,7040 000 000 000	5 200 000
	2	5500000	4	3,0250 000 000 000	11 000 000
	3	5800000	9	3,3640 000 000 000	17 400 000
	4	6000000	16	36 000 000 000 000	24 000 000
	5	6250000	25	39062500 000 000	31 250 000
	6	6500000	36	42250 000 000 000	39 000 000
Total	21	35250000	91	208243 000 000 000	127850000

Moyenne de x : $\bar{x} = \frac{21}{6} = 3,5$;

Moyenne de y : $\bar{y} = \frac{35\,250\,000}{6} = 5\,875\,000$;

Variance de x : $V(x) = \frac{91}{6} - 3,5^2 \approx 2,917$.

Variance de y : $V(y) = \frac{208\,243\,000\,000\,000}{6} - 5\,875\,000^2 \approx 191\,458\,000\,000$.

Covariance : $Cov(x, y) = \frac{127\,850\,000}{6} - 3,5 \times 5\,875\,000 \approx 745\,833,333$.

Le coefficient de corrélation linéaire : $r = \frac{Cov(x, y)}{\sqrt{V(x)V(y)}} = \frac{745\,833,333}{\sqrt{2,917 \times 191\,458\,000\,000}} \approx 0,998$.

• $0,87 \leq r < 1$, donc il y a une bonne corrélation ou une forte corrélation entre le rang de l'année et le bénéfice.

• Droite (D) de régression de y en x :

L'équation de la droite (D) sous la forme : $y = ax + b$ avec $a = \frac{Cov(x, y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$.

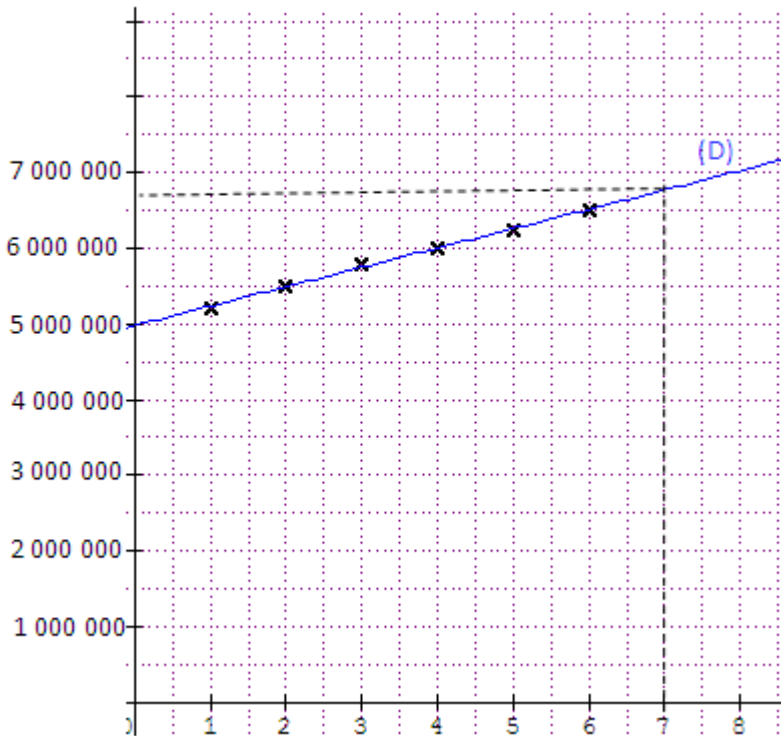
$a = \frac{Cov(x, y)}{V(x)} = \frac{745\,833,333}{2,917} \approx 255\,685,064$;

$b = \bar{y} - a\bar{x} \approx 5\,875\,000 - 255\,685,064 \times 3,5 \approx 4\,980\,102,276$.

Equation de la droite de régression de y en x : $y = 255\,685,064x + 4\,980\,102,276$.

- Estimation du bénéfice probable de l'entreprise à la fin de l'année en cours :
 - Estimation graphique :
 - On construit la droite (D).

Bénéfice en F CFA



Pour $x = 7$, on acceptera l'un des résultats de lecture graphique suivants pour y :
6 750 000 ; 6 700 000 ; 6 800 000.

- Estimation algébrique :

Pour $x = 7$, on obtient $y = 255685,064 \times 7 + 4\,980\,102,276$, c'est à dire $y \approx 6\,769\,898$.

Conclure selon la valeur et la méthode adoptée.

44

NB : Le texte « Son protégé, élève ... de maman. » est répété dans l'énoncé.

Nous allons utiliser les connaissances sur les statistiques pour traiter cette situation.

Tâche à réaliser : Estimer la taille qu'atteindra l'enfant à son premier anniversaire de naissance.

Donnée :

Nuage de points donnant l'évolution de la taille de l'enfant en fonction de son âge. (Voir représentation)

Démarche :

Comme le premier anniversaire est atteint au bout du 12^{ème} mois, nous allons vérifier qu'un ajustement affine de la série est possible, puis déterminer une droite d'ajustement, enfin estimer la taille qu'atteindra l'enfant au 12^{ème} mois.

- Les points du nuage de points semblent alignés, c'est pourquoi il est possible de faire un ajustement affine de la série statistique.

1^{ère} méthode : Nous allons utiliser la méthode de Mayer.

• Droite (D) de Mayer :

Par lecture graphique, on traduit le nuage de points par les tableaux suivants :

Série 1					Total
Age x_i en mois	1	2	3	4	10
Taille y_i en centimètres	52	53	54,5	56	215,5

$$x_{G_1} = \frac{10}{4} = 2,5 \quad \text{et} \quad y_{G_1} = \frac{215,5}{4} = 53,875.$$

Série 2					Total
Age x_i en mois	5	6	7	8	26
Taille y_i en centimètres	57,5	59	60,5	62	239

$$x_{G_2} = \frac{26}{4} = 6,5 \text{ et } y_{G_2} = \frac{239}{4} = 59,75.$$

Donc le point moyen de la série 1 est $G_1(2,5; 53,875)$ et celui de la série 2 est $G_2(6,5; 59,75)$.

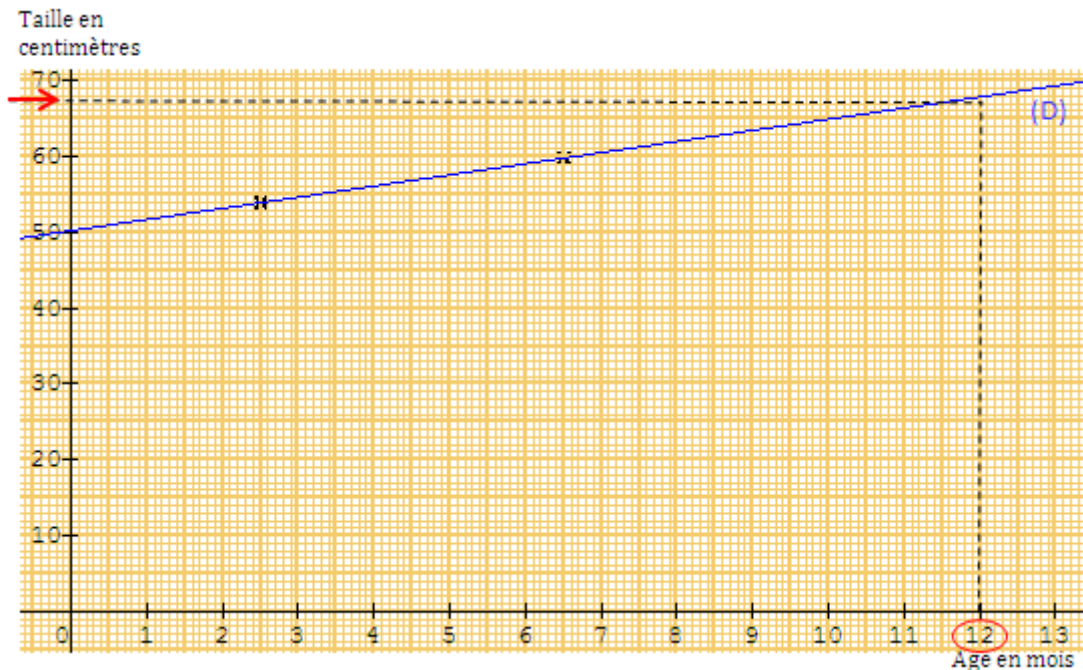
La droite de Mayer est la droite (D) passant par les points G_1 et G_2 .

On construit cette droite dans un repère ou bien on en détermine une équation, et on trouve : $y = 1,46875x + 50,203125$ (ou une équation avec les coefficients arrondis).

• Estimation de la taille qu'atteindra l'enfant à son premier anniversaire :

- Estimation graphique :

On construit sur une feuille de papier millimétré, la droite d'ajustement, celle de Mayer :



et on lit pour $x = 12$, la valeur correspondante de y , on acceptera $y \approx 67 \text{ cm}$ ou $y \approx 68 \text{ cm}$.

- Estimation algébrique : $y = 1,46875x + 50,203125$.

Pour $x = 12$, on obtient $y = 1,46875 \times 12 + 50,203125$, c'est à dire $y \approx 67,828125 \text{ cm}$.

2^e méthode : Nous allons utiliser la méthode des moindres carrés.

• Calcul du coefficient de corrélation linéaire :

Par lecture graphique, on traduit le nuage de points en tableau statistique et on obtient :

x_i	y_i	x_i^2	y_i^2	$x_i y_i$	
1	52	52	1	2704	
2	53	106	4	2809	
3	54,5	163,5	9	2970,25	
4	56	224	16	3136	
5	57,5	287,5	25	3306,25	
6	59	354	36	3481	
7	60,5	423,5	49	3660,25	
8	62	496	64	3844	
Total	36	454,5	2106,5	204	25910,75

$$\text{Moyenne de } x : \bar{x} = \frac{36}{8} = 4,5;$$

$$\text{Moyenne de } y : \bar{y} = \frac{454,5}{8} = 56,8125;$$

$$\text{Variance de } x : V(x) = \frac{2106,5}{8} - 4,5^2 \approx 5,25.$$

Variance de y : $V(y) = \frac{204}{8} \cdot 56,8125^2 \approx 11,1836$.

Covariance : $Cov(x, y) = \frac{25910,75}{8} \cdot 4,5 \times 56,8125 \approx 7,6563$.

Le coefficient de corrélation linéaire : $r = \frac{Cov(x, y)}{\sqrt{V(x)V(y)}} = \frac{7,6563}{\sqrt{5,25 \times 11,1836}} \approx 0,9992$.

- $0,87 \leq r < 1$, donc il y a une bonne corrélation ou une forte corrélation entre l'âge et la taille du nourrisson.
- Droite (D) de régression de y en x :

L'équation de la droite (D) sous la forme : $y = ax + b$ avec $a = \frac{Cov(x, y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$.

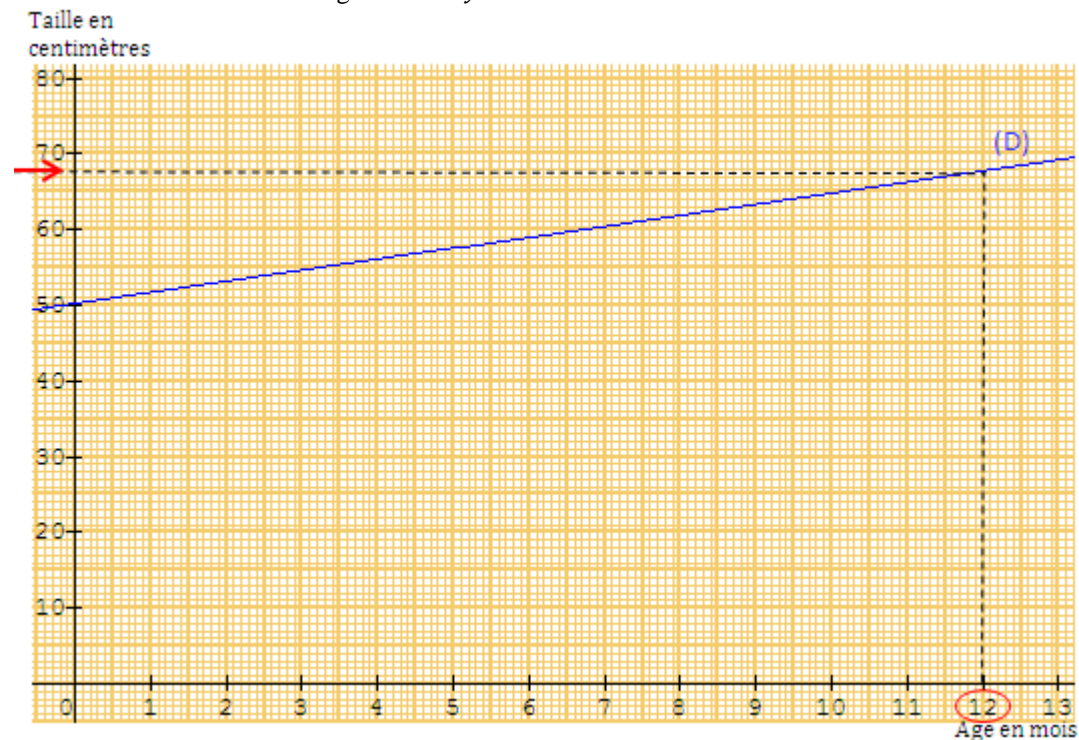
$a = \frac{Cov(x, y)}{V(x)} = \frac{7,6563}{5,25} \approx 1,4583$; $b = \bar{y} - a\bar{x} \approx 56,8125 - 1,4583 \times 4,5 \approx 50,2502$.

Equation de la droite de régression de y en x : $y = 1,4583x + 50,2502$.

- Estimation de la taille qu'atteindra l'enfant à son premier anniversaire :

- Estimation graphique :

On construit la droite de régression de y en x :



et on lit pour $x = 12$, la valeur correspondante de y ; on acceptera $y \approx 67$ cm ou $y \approx 68$ cm .

- Estimation algébrique :

Pour $x = 12$, on obtient $y = 1,4583 \times 12 + 50,2502$, c'est à dire $y \approx 67,7498$ cm.

Conclusion : A son premier anniversaire de naissance, l'enfant atteindra 67,7498 cm. (NB : on acceptera toute approximation décimale de cette valeur).