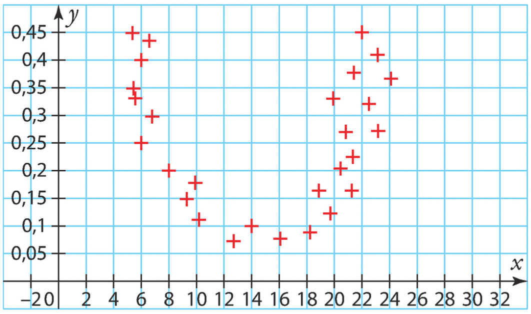
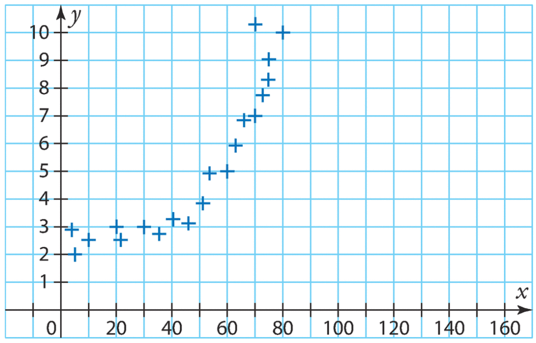
Corrigés des exercices - Statistiques

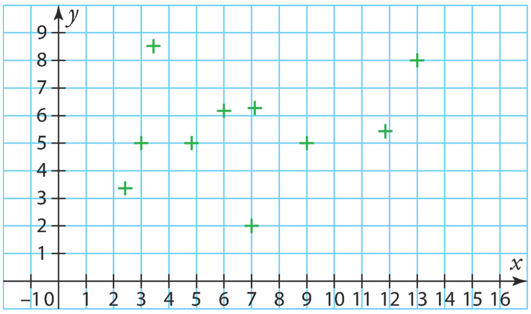
Nuage de points, ajustement

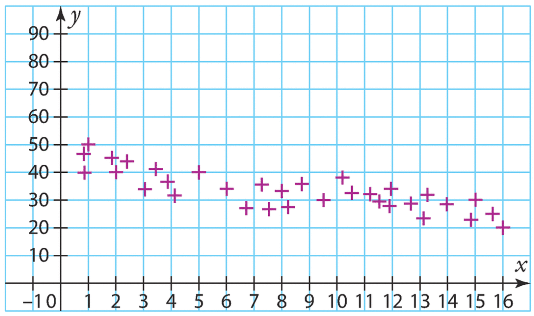
Exercice 1

Parmi les nuages de points suivants, lesquel semblent « ajustables » par une fonction reconnaissable ? Citer la fonction correspondante.

1. 

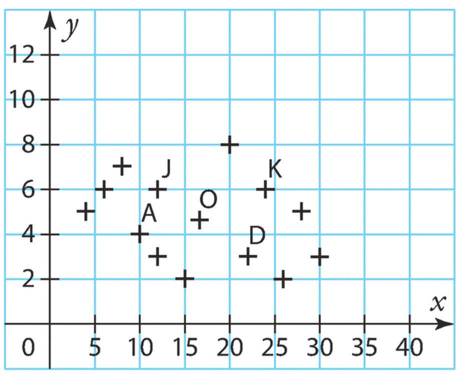
3. 

2. 

4. 

1. Le nuage de points 1 semble ajustable par une fonction du second degré.
2. Le nuage de points 2 ne ressemble pas à la courbe représentative d'une fonction connue.
3. Le nuage de points 3 semble ajustable par une fonction exponentielle.
4. Le nuage de points 4 semble ajustable par une fonction affine.

Exercice 2



1. Dans le nuage de points ci-contre, parmi les points A, J, O, D, K, quel point semble être le point moyen du nuage ?
2. Peut-on trouver une corrélation entre les deux variables x et y de la série statistique double associée à ce nuage de points ?

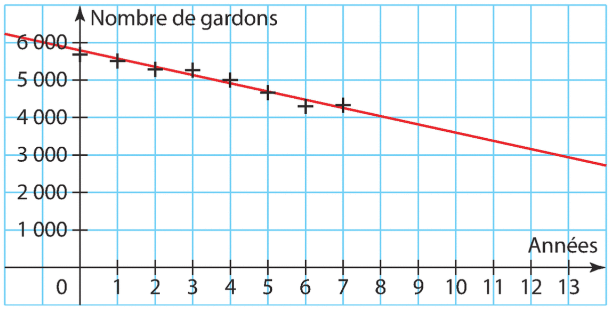
1. Le point moyen du nuage de points semble être le point O.

2. Le nuage de points n'a pas de forme particulière, il ne semble donc pas y avoir de corrélation entre les deux variables x et y.

Exploiter un ajustement affine

Exercice 3

Le graphique ci-contre présente l'évolution annuelle du nombre de gardons dans un étang depuis l'année 2017 (année 0).
Le nuage de points obtenu peut être ajusté par la droite (d) représentée sur le graphique.



Si l'évolution continue pendant encore plusieurs années selon l'ajustement modélisé par la droite (d), estimer graphiquement :

1. Le nombre de gardons dans l'étang en 2025.

2. L'année à partir de laquelle il y aura moins de 3000 gardons dans l'étang.

1. En 2025, il y aurait environ 4000 gardons dans l'étang.

2. D'après le modèle, il y aura moins de 3000 gardons dans l'étang à partir de $2017 + 13$ soit 2030.

Exercice 4

La maire d'un village a fait comptabiliser le nombre de mégots ramassés annuellement dans la rue principale de sa commune entre 2020 et 2024. La variable x_i correspond au rang de l'année d'observation et la variable y_i au nombre, en milliers, de mégots ramassés. Voici les résultats obtenus :

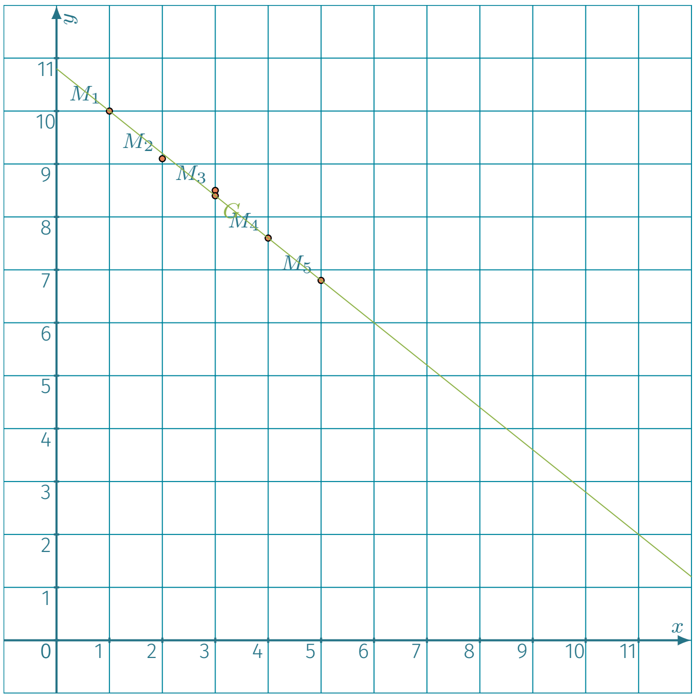
Année	2020	2021	2022	2023	2024	Total
Rang x_i	1	2	3	4	5	15
Quantité y_i	10	9,1	8,5	7,6	6,8	42

1. Dans un repère, on note M_i le point de coordonnées $M_i\,(x_i\ ;\ y_i)$ pour $i = 1, 2, ..., 5$. Représenter le nuage de points associé à cette série statistique.

2. Calculer les coordonnées du point moyen G de cette série statistique. Placer G sur le graphique.

3. On ajuste le nuage par la droite (GM_5). Déterminer l'équation réduite de (GM_5).

4. En supposant que l'évolution se poursuive ainsi, estimer le nombre de mégots qui seraient ramassés dans cette rue en 2030.

1. 

2. Moyenne des x_i : $\overline{x} = \dfrac{1+2+3+4+5}{5} = 3$

   Moyenne des y_i : $\overline{y} = \dfrac{10+9,1+8,5+7,6+6,8}{5} = 8,4$

   Le point moyen G a pour coordonnées $G\,(3\ ;\ 8,4)$.

3. L'équation de la droite (GM_5) est $y = mx + p$, avec m et p deux réels à déterminer.

   On a $m = \dfrac{y_5 - y_G}{x_5 - x_G}$ et p est solution de $y_G = m\,x_G + p$.

$$\begin{aligned} m &= \frac{y_5 - y_G}{x_5 - x_G} \\ &= \frac{6,8-8,4}{5-3} \\ &= \frac{-1,6}{2} \\ &= -0,8 \end{aligned}$$

$$\begin{aligned} y_G = m\,x_G + p &\iff 8,4 = -0,8 \times 3 + p \\ &\iff 8,4 = -2,4 + p \\ &\iff p = 8,4 + 2,4 \\ &\iff p = 10,8 \end{aligned}$$

   Donc l'équation de la droite (GM_5) est $y = -0,8x + 10,8$.

4. $2030 = 2019 + 11$.

   On calcule les coordonnées du point M_{11} de la droite (GM_5) :

$$\begin{aligned} y_{11} &= -0,8 \times 11 + 10,8 \\ &= -0,8 \times 11 + 10,8 \\ &= -8,8 + 10,8 \\ &= 2 \end{aligned}$$

   Donc en 2030, il y aurait environ 2 milliers de mégots ramassés dans cette rue d'après ce modèle.

Exercice 5

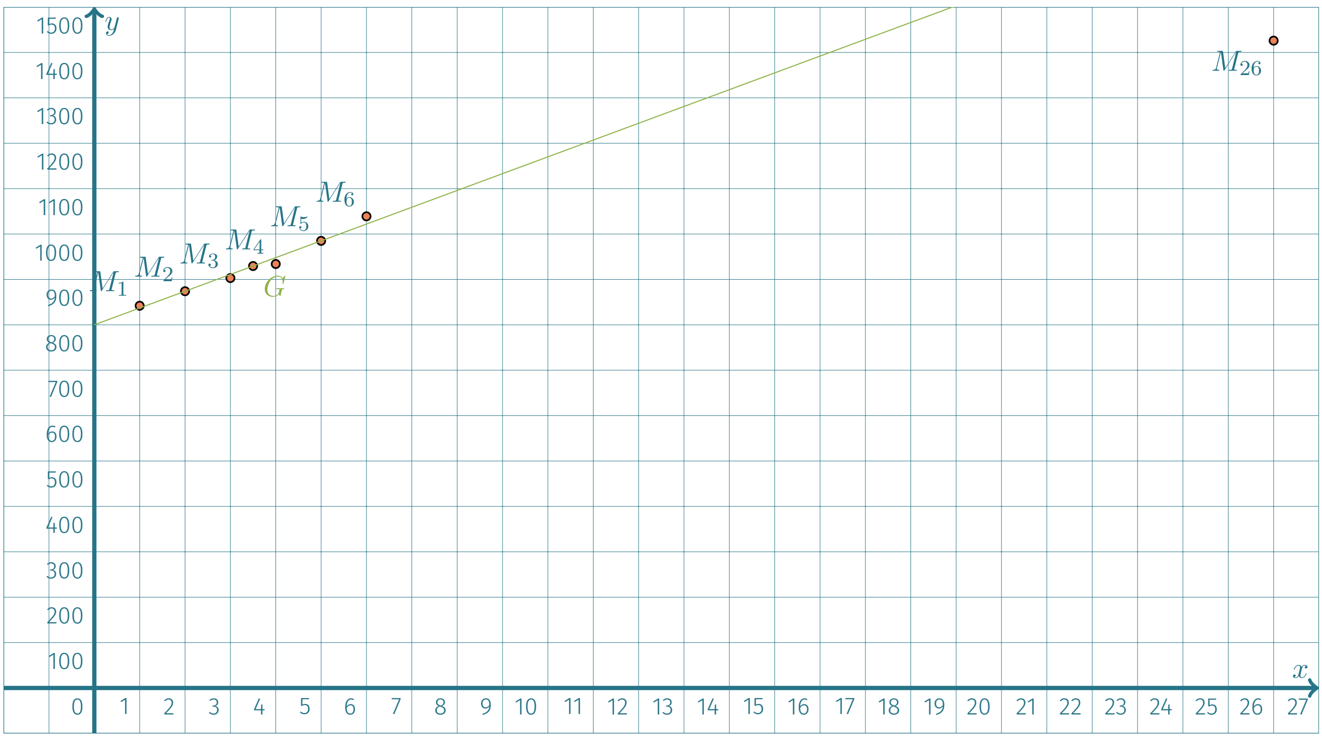
Le tableau ci-dessous donne l'évolution du montant net du SMIC (Salaire Minimum Interprofessionnel de Croissance) mensuel, en euros, entre 2000 et 2005.

Année	2000	2001	2002	2003	2004	2005
Rang x_i	1	2	3	4	5	6
Montant y_i	842	874	903	934	985	1039

1. Représenter, dans un repère, le nuage de points $M_i\,(x_i\ ;\ y_i)$ associé à cette série statistique et placer le point moyen G.

2. Déterminer l'équation de la droite (GM_5).

3. En 2025, le SMIC net mensuel est de 1426 €. Comparer ce montant à celui extrapolé avec la droite précédente.

1. 

2. Moyenne des x_i : $\overline{x} = \dfrac{1+2+3+4+5+6}{6} = \dfrac{21}{6} = 3,5$

   Moyenne des y_i : $\overline{y} = \dfrac{842+874+903+934+985+1039}{6} = \dfrac{5577}{6} = 929,5$

   Le point moyen G a pour coordonnées $G\,(3,5\ ;\ 929,5)$.

3. L'équation de la droite (GM_5) est $y = mx + p$, avec m et p deux réels à déterminer.
   On a $m = \dfrac{y_5 - y_G}{x_5 - x_G}$ et p est solution de $y_G = m\,x_G + p$.

$$
\begin{aligned}
m &= \frac{y_5 - y_G}{x_5 - x_G} \\
&= \frac{985 - 929,5}{5 - 3,5} \\
&= \frac{55,5}{1,5} \\
&= 37
\end{aligned}
\qquad
\begin{aligned}
y_G = m\,x_G + p \quad &\Longleftrightarrow \quad 929,5 = 37 \times 3,5 + p \\
&\Longleftrightarrow \quad 929,5 = 129,5 + p \\
&\Longleftrightarrow \quad p = 929,5 - 129,5 \\
&\Longleftrightarrow \quad p = 800
\end{aligned}
$$

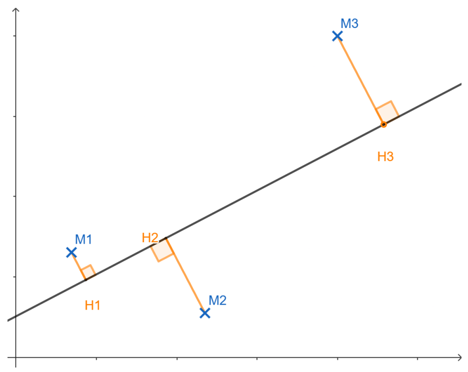
Donc l'équation de la droite (GM_5) est $y = 37x + 800$.

4. $2025 = 1999 + 26$.
On calcule les coordonnées du point M_{26} de la droite (GM_5) :

$$
\begin{aligned}
y_{26} &= 37 \times 26 + 800 \\
&= 962 + 800 \\
&= 1762
\end{aligned}
$$

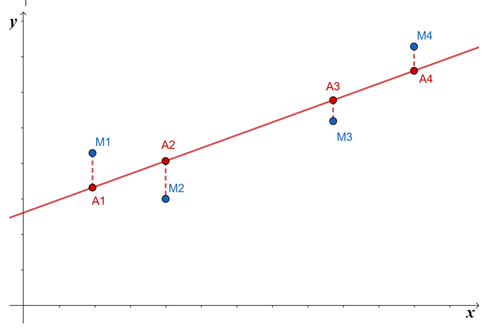
Donc en 2025, d'après ce modèle, le SMIC net mensuel devrait être de 1762 €, c'est-à-dire 336 € de plus que le montant réel de 1426 €.

Méthode des moindres carrés



Exercice 6

Edwin affirme « La méthode des moindres carrés minimise la somme des $M_iH_i^2$, comme représenté ci-contre. »
A-t-il raison ?



Edwin a tort.
La méthode des moindres carrés minimise la somme des $M_iA_i^2$.

Exercice 7

Dans un repère orthogonal, le point $G\,(5\ ;\ 4)$ est le point moyen d'un nuage de points.
Martha affirme « La droite d'ajustement affine associée à ce nuage obtenue par la méthode des moindres carrés a pour équation $y = -2x + 1$. »
Peut-elle avoir raison ?

La droite d'ajustement affine obtenue par la méthode des moindres carrés passe par le point moyen G.

$$
\begin{aligned}
-2x_G+1 &= -2\times 5+1\\
&= -10+1\\
&= -9\\
&\neq y_G
\end{aligned}
$$

Donc Martha a tort, le point moyen G n'est pas sur la droite.

Exercice 8

Le tableau ci-dessous donne la masse x (en kg) et la taille y (en cm) d'un enfant, relevées par son médecin lors de six visites médicales.

Masse x_i (en kg)	4	5,4	10,2	11	12,6	19,8
Taille y_i (en cm)	53	61	72	78	94	113

On veut réaliser un ajustement affine de cette série par la méthode des moindres carrés.

1. Calculer les valeurs moyennes $\overline{x}$ et $\overline{y}$ des deux séries (x_i) et (y_i).
2. Compléter le tableau suivant :

Masse x_i (en kg)	4	5,4	10,2	11	12,6	19,8
Taille y_i (en cm)	53	61	72	78	94	113
Écarts $x_i-\overline{x}$						
Écarts $y_i-\overline{y}$						
Produit des écarts $(x_i-\overline{x})(y_i-\overline{y})$						
Carré des écarts $(x_i-\overline{x})^2$						

   En déduire covariance entre les deux séries Cov(x, y) et la variance Var(x) de la série (x_i).
3. En déduire l'équation de la droite de d'ajustement affine (d) de la série (x_i, y_i) par la méthode des moindres carrés.
4. Représenter le nuage de points $M_i\,(x_i\ ;\ y_i)$ et la droite (d) dans un repère orthonormé.
5. Quelle estimation peut-on faire de la taille de cette enfant lorsqu'il pèsera 25 kg ?

1. Moyenne des x_i : $\overline{x}=\dfrac{4+5,4+10,2+11+12,6+19,8}{6}=\dfrac{63}{6}=10,5$

   Moyenne des y_i : $\overline{y}=\dfrac{53+61+72+78+94+113}{6}=\dfrac{471}{6}=78,5$
2. On complète le tableau :

Masse x_i (en kg)	4	5,4	10,2	11	12,6	19,8
Taille y_i (en cm)	53	61	72	78	94	113
Écarts $x_i - \overline{x}$	$-6,5$	$-5,1$	$-0,3$	$0,5$	$2,1$	$9,3$
Écarts $y_i - \overline{y}$	$-25,5$	$-17,5$	$-6,5$	$-0,5$	$15,5$	$34,5$
Produit des écarts $(x_i - \overline{x})(y_i - \overline{y})$	$165,75$	$89,25$	$1,95$	$-0,25$	$32,55$	$321,35$
Carré des écarts $(x_i - \overline{x})^2$	$42,25$	$26,01$	$0,09$	$0,25$	$4,41$	$86,49$

On en déduit :

$$\begin{aligned}\text{Cov}(x,y) &= \frac{165,75+89,25+1,95-0,25+32,55+321,35}{6}\\ &= \frac{610,1}{6}\\ &\approx 101,68\end{aligned}$$

$$\begin{aligned}\text{Var}(x) &= \frac{42,25+26,01+0,09+0,25+4,41+86,49}{6}\\ &= \frac{159,5}{6}\\ &= 26,58\end{aligned}$$

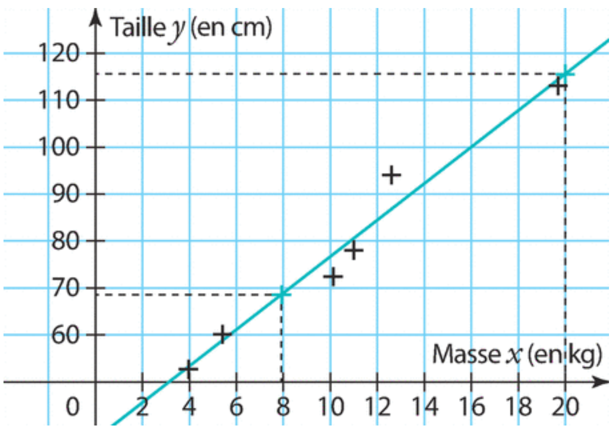
3. L'équation de la droite (d) d'ajustement affine de la série (x_i, y_i) par la méthode des moindres carrés est : $y = m(x - \overline{x}) + \overline{y}$ avec $m = \dfrac{\text{Cov}(x,y)}{\text{Var}(x)}$.

$$\begin{aligned}m &= \frac{\text{Cov}(x,y)}{\text{Var}(x)}\\ &= \frac{101,68}{26,58}\\ &\approx 3,83\end{aligned}$$

Donc l'équation de la droite (d) d'ajustement affine de la série (x_i, y_i) est :

$$\begin{aligned}y &= 3,83(x-10,5)+78,5\\ &= 3,83x - 40,165 + 78,5\\ &= 3,83x + 38,335\end{aligned}$$

4. On représente la série (x_i, y_i) et la droite (d) dans un repère orthonormé :



5. On cherche à estimer la taille de l'enfant lorsqu'il pèsera 25 kg.
   On calcule l'ordonnée du point M_{25} d'abscisse 25 de la droite (d) :

$$\begin{aligned} y_{25} &= 3,83 \times 25 + 38,335 \\ &= 95,75 + 38,335 \\ &= 134,085 \end{aligned}$$

   Donc d'après ce modèle, la taille de l'enfant lorsqu'il pèsera 25 kg serait d'environ 134 cm.

Exercice 9

On a relevé à certaines latitudes x_i (en degrés) la température y_i (en °C) à la surface de la mer. Voici les résultats obtenus :

Latitude x_i (en °)	-45	-20	0	30
Température y_i (en °C)	12,2	12,9	13,4	14,1

1. Calculer la covariance entre les deux séries Cov(x, y) et la variance Var(x) de la série (x_i).
2. En déduire l'équation de la droite de d'ajustement affine (d) de la série (x_i, y_i) par la méthode des moindres carrés.
3. Représenter le nuage de points $M_i\,(x_i\ ;\ y_i)$ et la droite (d) dans un repère orthonormé.

1. Moyenne des x_i : $\overline{x} = \dfrac{-45-20+0+30}{4} = \dfrac{-35}{4} = -8,75$
   Moyenne des y_i : $\overline{y} = \dfrac{12,2+12,9+13,4+14,1}{4} = \dfrac{52,6}{4} = 13,15$
   On complète le tableau :

Latitude x_i (en °)	-45	-20	0	30
Température y_i (en °C)	12,2	12,9	13,4	14,1
Écarts $x_i - \overline{x}$	$-36,25$	$-11,25$	$8,75$	$38,75$
Écarts $y_i - \overline{y}$	$-0,95$	$-0,25$	$0,25$	$0,95$
Produit des écarts $(x_i - \overline{x})(y_i - \overline{y})$	$34,4375$	$2,8125$	$2,1875$	$36,8125$
Carré des écarts $(x_i - \overline{x})^2$	$1314,0625$	$126,5625$	$76,5625$	$1501,5625$

   On en déduit :

$$\begin{aligned} \text{Cov}(x, y) &= \frac{34,4375+2,8125+2,1875+36,8125}{4} \\ &= \frac{76,25}{4} \\ &= 19,0625 \end{aligned}$$

$$\begin{aligned}\text{Var}(x) &= \frac{1314,0625 + 126,5625 + 76,5625 + 1501,5625}{4} \\ &= \frac{3018,75}{4} \\ &= 754,6875\end{aligned}$$

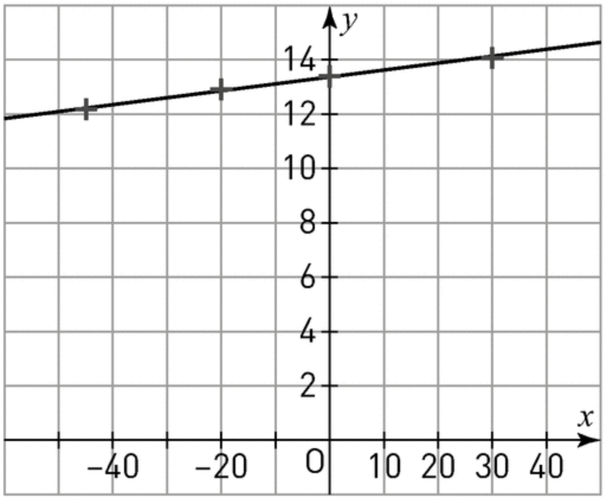
2. L'équation de la droite (d) d'ajustement affine de la série (x_i, y_i) par la méthode des moindres carrés est : $y = m(x - \overline{x}) + \overline{y}$ avec $m = \dfrac{\text{Cov}(x,y)}{\text{Var}(x)}$.

$$\begin{aligned} m &= \frac{\text{Cov}(x,y)}{\text{Var}(x)} \\ &= \frac{19,0625}{754,6875} \\ &\approx 0,025\end{aligned}$$

Donc l'équation de la droite (d) d'ajustement affine de la série (x_i, y_i) est :

$$\begin{aligned} y &\approx 0,025(x + 8,75) + 13,15 \\ &\approx 0,025x + 0,21875 + 13,15 \\ &\approx 0,025x + 13,37\end{aligned}$$

3. On représente la série (x_i, y_i) et la droite (d) dans un repère orthonormé :



Exercice 10

Sous des conditions de température et de volume constants, on étudie la pression et la quantité de matière d'un gaz. Les résultats sont donnés dans le tableau ci-dessous :

Nombre de moles x_i	0	10	20	30
Pression y_i (en kPa)	0	46	98	145

1. Dans un reprère orthonormé, représenter le nuage de points $M_i\,(x_i\ ;\ y_i)$ et vérifier qu'il peut être ajusté par une droite.

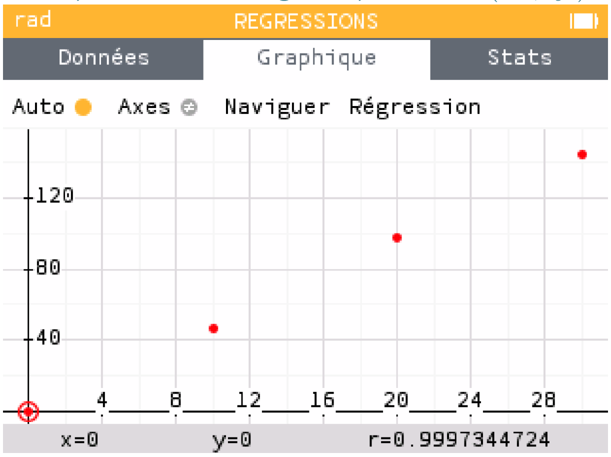
2. Déterminer l'équation réduite de la droite (d) d'ajustement affine de la série (x_i, y_i) par la méthode des moindres carrés et le coefficient de corrélation linéaire.

3. Représenter la droite (d) dans le repère.

4. Quel serait la pression du gaz si l'on avait 50 moles ?

5. Combien de moles aurait-on si le gaz avait une pression de 200 kPa ?

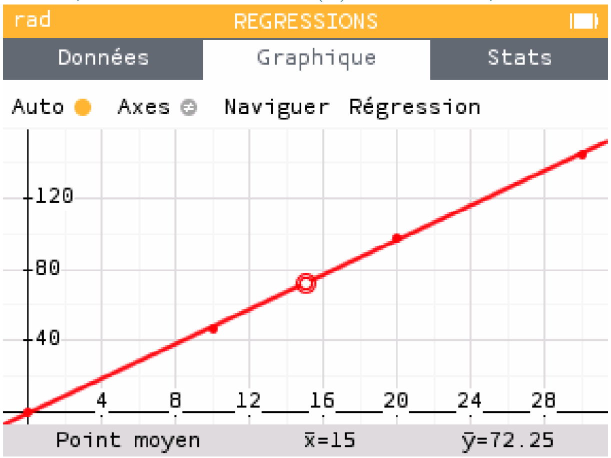
1. On représente le nuage de points $M_i\,(x_i\ ;\ y_i)$ à l'aide de la calculatrice :



Les points semblent alignés, donc le nuage de points peut être ajusté par une droite.

2. On utilise la calculatrice pour déterminer l'équation de la droite (d) d'ajustement affine de la série (x_i, y_i) par la méthode des moindres carrés.
   On trouve : $y = 4,87x - 0,8$ et le coefficient de corrélation linéaire est $r \approx 0,9997$.

3. On représente la droite (d) dans le repère :



4. On cherche à estimer la pression du gaz si l'on avait 50 moles.
   On calcule l'ordonnée du point M_{50} d'abscisse 50 de la droite (d) :

$$\begin{aligned} y_{50} &= 4,87 \times 50 - 0,8 \\ &= 243,5 - 0,8 \\ &= 242,7 \end{aligned}$$

Donc d'après ce modèle, la pression du gaz serait de 242,7 kPa si l'on avait 50 moles.

5. On cherche à estimer le nombre de moles si le gaz avait une pression de 200 kPa.
   On calcule l'abscisse du point M' d'ordonnée 200 de la droite (d) :

$$\begin{aligned} 200 = 4,87x - 0,8 \quad &\Longleftrightarrow \quad 4,87x = 200 + 0,8 \\ &\Longleftrightarrow \quad 4,87x = 200,8 \\ &\Longleftrightarrow \quad x = \frac{200,8}{4,87} \end{aligned}$$

D'après ce modèle, le nombre de moles serait d'environ $\frac{200,8}{4,87} \approx 41$ moles si le gaz avait une pression de 200 kPa.

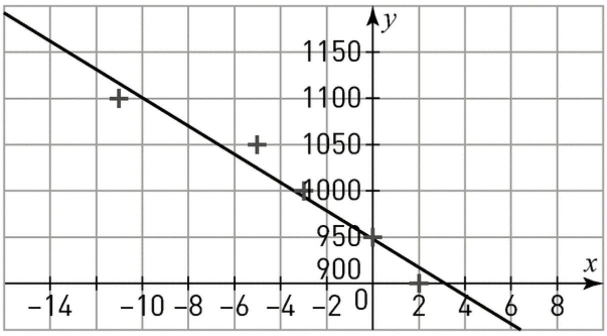
Exercice 11

On considère la série statistique à deux variables (x, y) donnée par le tableau suivant :

Valeurs x_i	-11	-3	2	0	-5
Valeurs y_i	1100	1000	900	950	1050

1. Dans un repère orthogonal, représenter le nuage de points $M_i\,(x_i\ ;\ y_i)$ et vérifier qu'il peut être ajusté par une droite.
2. Déterminer l'équation réduite de la droite (d) d'ajustement affine de la série (x_i, y_i) par la méthode des moindres carrés et le coefficient de corrélation linéaire.
3. Représenter la droite (d) dans le repère.

1. On représente le nuage de points $M_i\,(x_i\ ;\ y_i)$ dans un repère orthonormé :



Les points semblent alignés, donc le nuage de points peut être ajusté par une droite.

2. On utilise la calculatrice pour déterminer l'équation de la droite (d) d'ajustement affine de la série (x_i, y_i) par la méthode des moindres carrés.
On trouve : $y \approx -15,32x + 971,92$ et le coefficient de corrélation linéaire est $r \approx -0,97$.

Ajustement se ramenant par changement de variable à un ajustement affine

Exercice 12

On injecte un médicament dans le sang d'une patiente en lui faisant une piqûre en intraveineuse. La concentration y_i de ce médicament dans le sang, en micogrammes par millilitres est relevée à différents instants x_i (en heures) après la piqûre. Voici les résultats obtenus :

Temps x_i (en h)	0	1	2	4	6	10
Concentration y_i (en $\mu g/mL$)	90	65	45	23	12	4

1. Représenter le nuage de points $M_i\,(x_i\ ;\ y_i)$ sur la calculatrice et vérifier qu'il peut être ajusté par une fonction exponentielle décroissante.

2. On pose $y'_i = \ln(y_i)$ pour tout entier i de 1 à 6.
Compléter le tableau suivant :

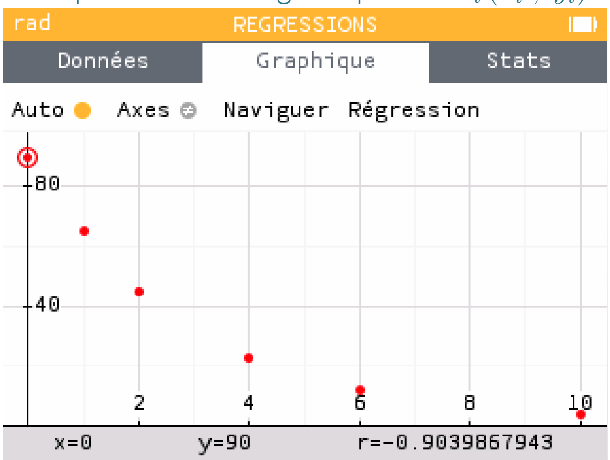
x_i	0	1	2	4	6	10
y_i	90	65	45	23	12	4
y'_i						

3. Représenter le nuage de points $M_i\,(x_i\ ;\ y'_i)$ sur la calculatrice et vérifier qu'il peut être ajusté par une droite.

4. Déterminer l'équation de la droite (d) d'ajustement affine de la série (x_i, y'_i) par la méthode des moindres carrés.

5. En déduire une approximation de la concentration y en fonction du temps x.

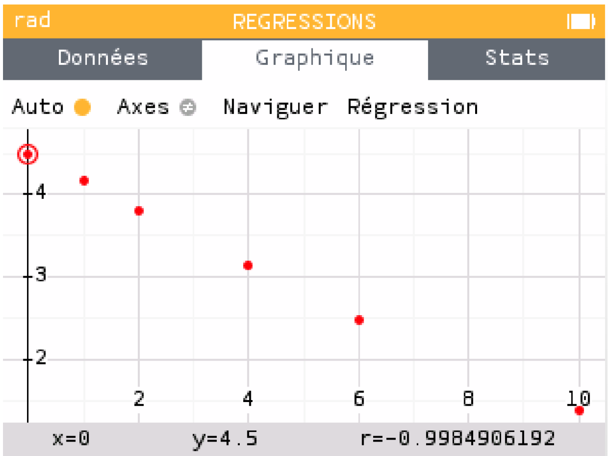
1. On représente le nuage de points $M_i\,(x_i\ ;\ y_i)$ à l'aide de la calculatrice :



2. On complète le tableau :

x_i	0	1	2	4	6	10
y_i	90	65	45	23	12	4
y'_i	4,5	4,17	3,81	3,14	2,48	1,39

3. On représente le nuage de points $M_i\,(x_i\ ;\ y'_i)$ à l'aide de la calculatrice :



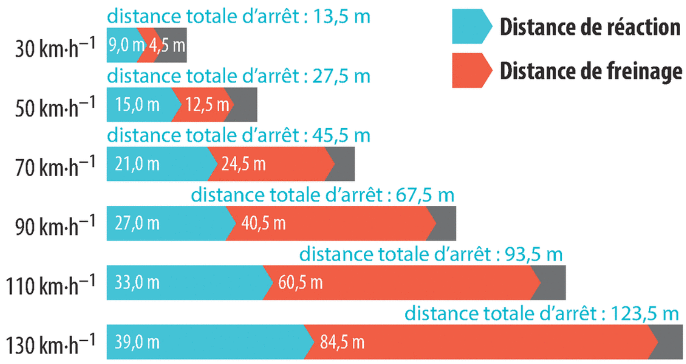
Les points semblent alignés, donc le nuage de points peut être ajusté par une droite.

4. On utilise la calculatrice pour déterminer l'équation de la droite (d) d'ajustement affine de la série (x_i, y'_i) par la méthode des moindres carrés.
   On trouve : $y' \approx -0,25x + 4,5$.

5. En déduire une approximation de la concentration y en fonction du temps x.
   On a : $\ln(y) \approx -0,25x + 4,5$.
   Donc : $y \approx e^{-0,25x+4,5} \approx e^{4,5}e^{-0,25x} \approx 90e^{-0,25x}$.
   Donc d'après ce modèle, la concentration y en fonction du temps x est :
   $y \approx 90e^{-0,25x}$.

Exercice 13

Le document ci-contre donne les valeurs moyennes de la distance de réaction et de la distance de freinage selon la vitesse d'un véhicule, sur route sèche, de jour et dans des conditions de visibilité normales.



Partie A : Étude de la série statistique à deux variables : vitesse x et distance de réaction y

1. Construire dans un repère orthonormé le nuage de points correspondant à la série statistique (x, y) où x représente la vitesse du véhicule en km.h^{-1} et y la distance de réaction en mètres. Vérifier que la forme du nuage de points invite à réaliser un ajustement affine.

2. Déterminer à l'aide de la calculatrice l'équation de la droite d'ajustement affine du nuage obtenu par la méthode des moindres carrés.

3. Selon cet ajustement :

   a. Estimer la distance de réaction d'un véhicule roulant, en excès de vitesse, à 140 km.h^{-1}.

   b. Estimer la vitesse à laquelle un véhicule doit rouler pour que la distance de réaction soit inférieure à 3 m.

Partie B : Étude de la série statistique à deux variables : vitesse x et distance de freinage y'

1. Construire dans un repère orthonormé le nuage de points correspondant à la série statistique (x, y') où x représente la vitesse du véhicule en km.h^{-1} et y' la distance de freinage en mètres. Vérifier que la forme du nuage de points invite à réaliser un ajustement affine.

2. On considère que ce nuage de points peut être ajusté par une parabole d'équation $y' = ax^2$ où a est un réel non nul.

   a. On pose $z = \sqrt{y'}$. Compléter le tableau suivant :

Vitesse x (en km.h^{-1})	30	50	70	90	110	130
Distance de freinage y' (en m)	4,5	12,5	24,5	40,5	60,5	84,5
$z = \sqrt{y'}$						

   b. À l'aide de la calculatrice, déterminer l'équation de la droite de regression de z en x.

   c. En déduire une approximation de la distance de freinage y' en fonction de la vitesse x.

3. Pour les questions suivantes, on admet que la fonction définie sue $[0\ ;\ +\infty[$ par $f(x) = \dfrac{x^2}{200}$ est une bonne modélisation de la distance de freinage y' en fonction de la vitesse x.

   a. Estimer la distance de freinage d'un véhicule roulant, en excès de vitesse, à 140 km.h^{-1}.

   b. À l'aide de la partie A, estimer la distane totale d'arrêt pour un véhicule roulant en excès de vitesse à 140 km.h^{-1}.

Exercice 14

En 2020, un lycée décide de créer un club d'échecs.
Le nombre d'adhérents chaque année depuis 2020 est donné dans le tableau ci-dessous :

Année	2020	2021	2022	2023
Rang x_i	1	2	3	4
Nombre d'adhérents y_i	9	15	22	31

1. On considère la série statistique à deux variables (x, y) donnée par le tableau ci-dessus. Déterminer le coefficient de corrélation linéaire r_1 de cette série statistique (x, y).

2. On pose $z = \ln(y)$ et on considère la série statistique à deux variables (x, z). Déterminer le coefficient de corrélation linéaire r_2 de cette série statistique (x, z).

3. Depuis la création du club, la présidente espère une progression exponentielle du nombre d'adhérents. Pour ces quatre premières années, peut-on dire que ses espérances sont satisfaites ?

4. En 2024, une grande campage de publicité a été organisée au lycée. Le nombre de participants totalisés à la fin de l'année est de 44.
   La présidente du club, après avoir observé l'ensemble des données de 2020 à 2024, affirme dans son discours de fin d'année : « Si on continue sur cette lancée, nous serons au moins 67 adhérents l'année prochaine ! »
   Qu'en pensez-vous ?