Séries statistiques à deux variables quantitatives

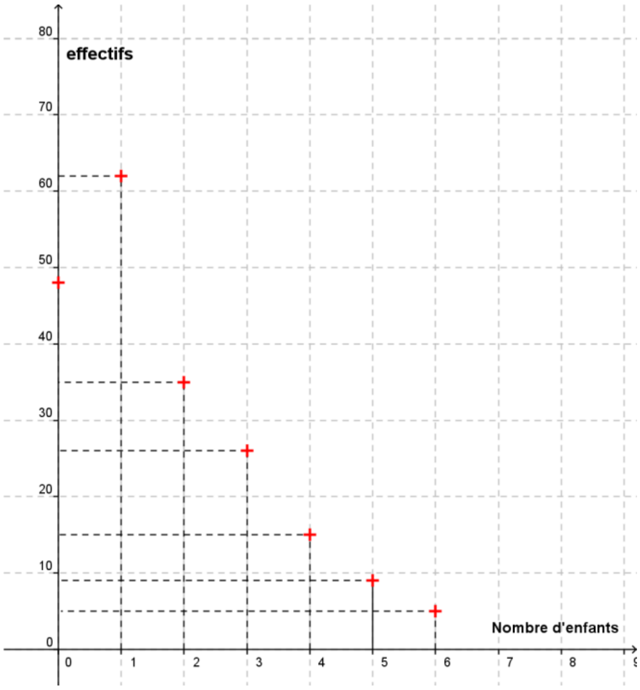
I) Nuage de points

1) Définition

Dans un repère orthogonal du plan, l'ensemble des points Mi , pour i variant de 1 à n , de coordonnées $(x_i\,;\,y_i)$, est appelé le nuage de points associé à la série statistiques $(x_1\,;\,y_1), (x_2\,;\,y_2), \ldots, (x_n\,;\,y_n)$ à deux variables.

Exemple : On étudie le nombre d'enfants dans 200 familles d'un village.

Nombre d'enfants par famille	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Effectifs (nombre de familles)	48	62	35	26	15	9	5



Méthode :

- On légende les axes du graphique : En abscisse le nombre d'enfants par famille, en ordonnée les effectifs correspondants
- On place les points de coordonnées (nombre d'enfants par familles ; effectif correspondant)

Attention : On construit un nuage de points. **Il ne faut pas relier les points.**

Le premier point a pour coordonnées (0 ;48), le deuxième a pour coordonnées (1 ;62) ………

2) Point moyen

Le point G de coordonnées $(\overline{x}\,;\overline{y})$, où $\overline{x}$ et $\overline{y}$ sont les moyennes respectives des x_i et des y_i, est appelé le point moyen du nuage de points associé à la série.

$\overline{x} = \frac{x_1+x_2+\cdots+x_n}{n}$ **et** $\overline{y} = \frac{y_1+y_2+\cdots+y_n}{n}$

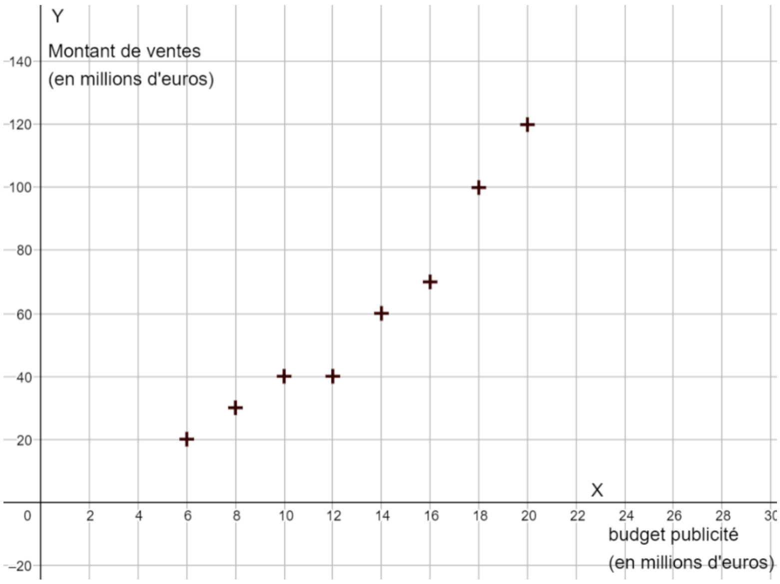
Exemple : Le tableau suivant présente simultanément une série de budgets publicité (en millions d'euros) et le montant de ventes correspondant (en millions).

Budget publicité X (en millions d'euros)	6	8	10	12	14	16	18	20
Montant de ventes Y (en millions d'euros)	20	30	40	40	60	70	100	120

1.Tracer le nuage des points dans un repère orthogonal du plan.

2. Calculer les coordonnées du point moyen G et le placer sur le graphique

Réponse :
1.

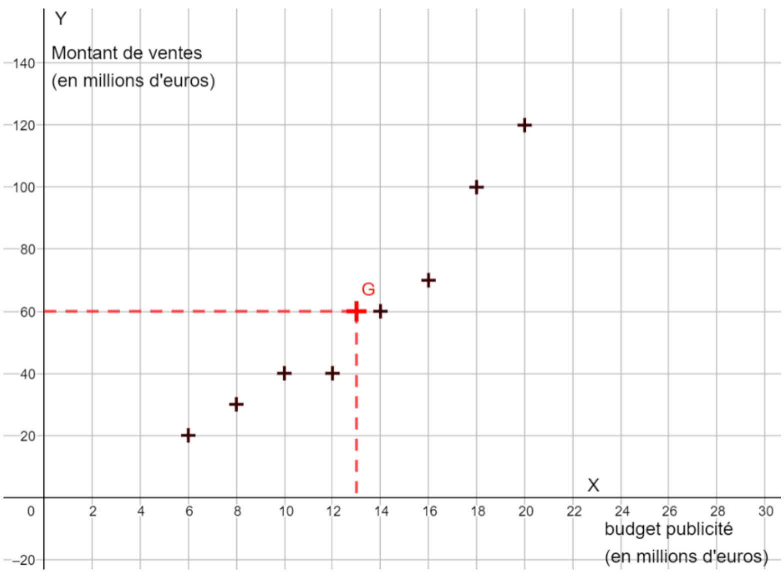


2.

$$\bar{x} = \frac{6+8+10+12+14+16+18+20}{8} = 13$$

$$\bar{y} = \frac{20+30+40+40+60+70+100+120}{8} = 60$$

Le point moyen G a pour coordonnées $G(13\,;60)$

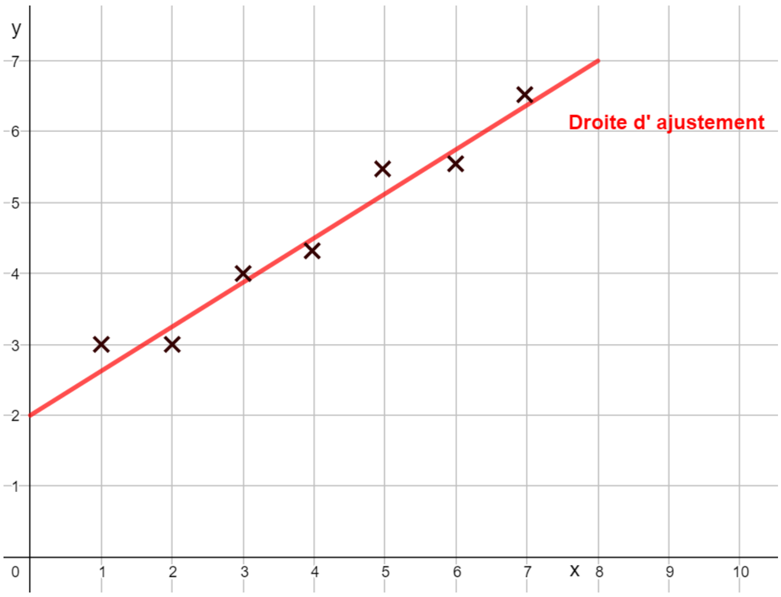


II) Ajustement affine

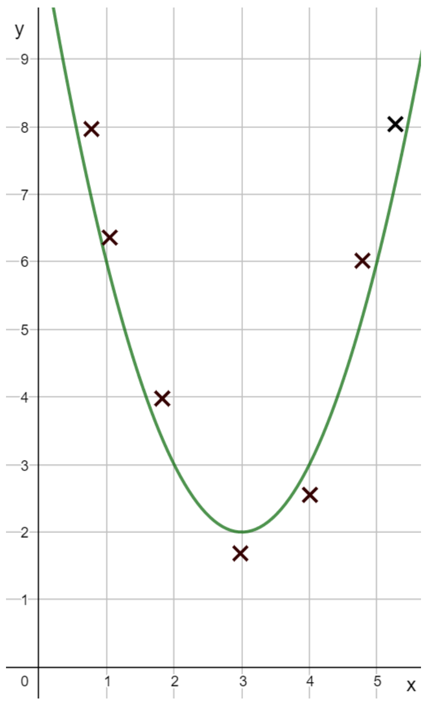
On cherche s'il existe une relation, qui est approximative, entre les deux variables x et y d'une série statistique. L'allure du nuage de points représentant la série statistique, en fonction de leur disposition, peut parfois suggérer la courbe d'une fonction f. Dans ce cas le lien approximatif entre x et y est de la forme $y = f(x)$.

On est donc amené à chercher une courbe « passant au plus près » des points du nuage. **On parle d'ajustement du nuage.**

Dans le cas où les points du nuage sont proches de l'alignement, la courbe cherchée est une droite on dit alors que cette droite réalise un **ajustement affine** du nuage de points.



Ajustement affine



Ajustement par une parabole

L'objectif est, à partir des valeurs d'une série statistique à deux variables, d'obtenir des approximations pour des valeurs inconnues de cette série utile par exemple pour faire des prévisions.

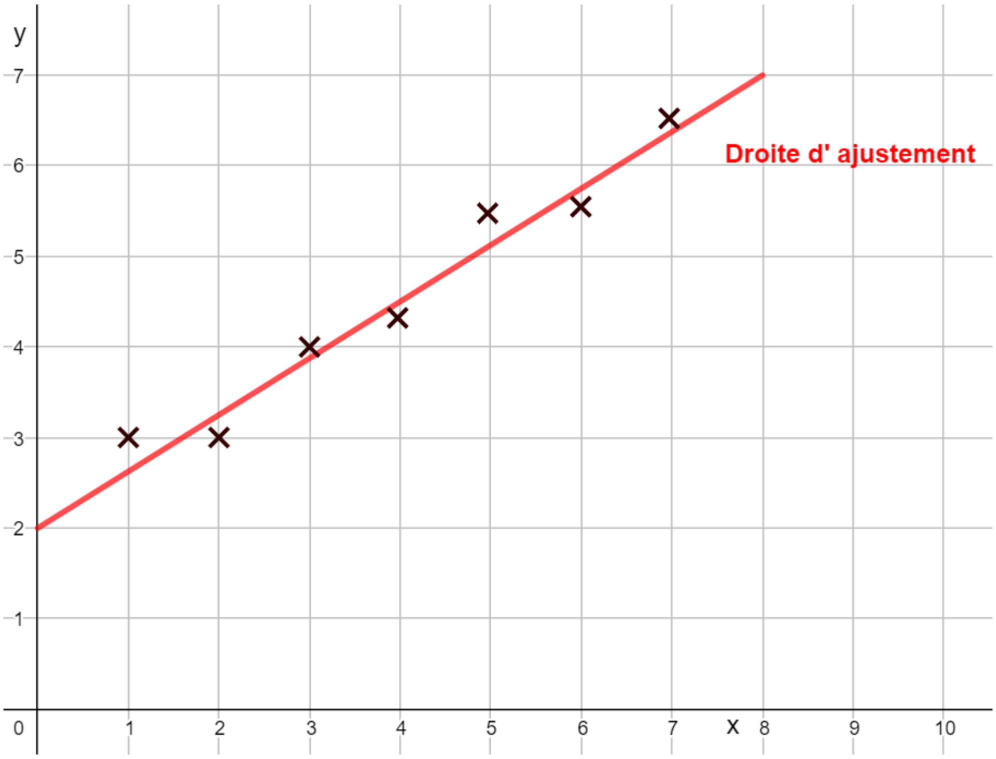
Par exemple, faire des prévisions pour les années à venir à partir d'une série exprimant la population d'une ville, d'un département, d'un pays en fonction des années.

Il peut ne pas y avoir d'ajustement possible.

1) Détermination de la droite d'ajustement « au jugé »

Sans technique particulière, on trace « au jugé » une droite passant « au plus près » des points du nuage.

Exemple :



Sans technique particulière, on trace « au jugé » une droite passant « au plus près » des points du nuage.

2) Méthode des points moyens

Cet ajustement consiste à déterminer la droite passant par deux points moyens du nuage de point.

Méthode et exemple : Déterminer la droite d'ajustement par la méthode des points moyens.

Le tableau suivant présente simultanément une série de budgets publicité (en millions d'euros) et le montant de ventes correspondant (en millions).

Budget publicité X (en millions d'euros)	6	8	10	12	14	16	18	20
Montant de ventes Y (en millions d'euros)	20	30	40	40	60	70	100	120

1. Soit G_1, le point moyen associé aux **quatre premiers points** du nuage et G_2 le point moyen associé aux **quatre derniers points** du nuage.
Calculer les coordonnées de G_1 et G_2.

2. On prend (G_1G_2) comme droite d'ajustement. Tracer cette droite.

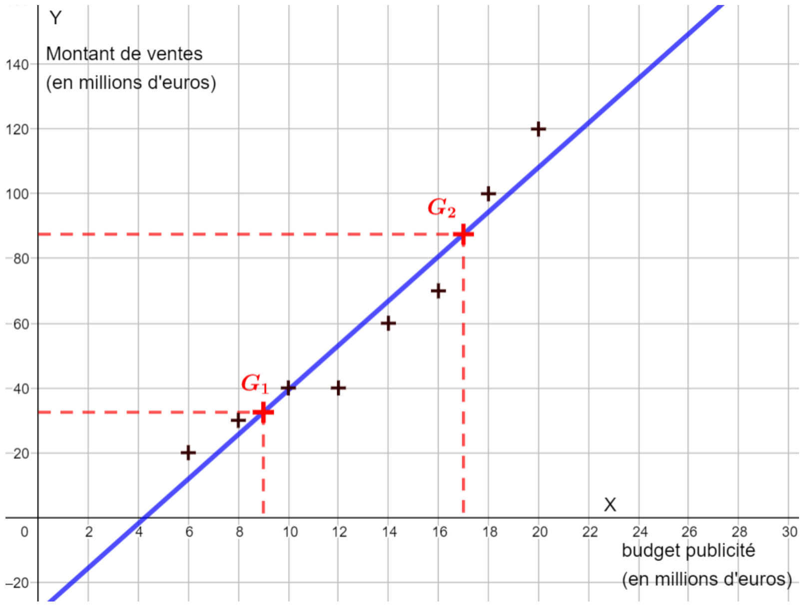
3. À l'aide du graphique, estimer le montant des ventes à prévoir pour un budget publicité de 22 000 000 €.

Réponse :

1. $\overline{x_1} = \frac{6+8+10+12}{4} = 9$ et $\overline{y_1} = \frac{20+30+40+40}{4} = 32{,}5$ donc $G_1(9\,;32{,}5)$

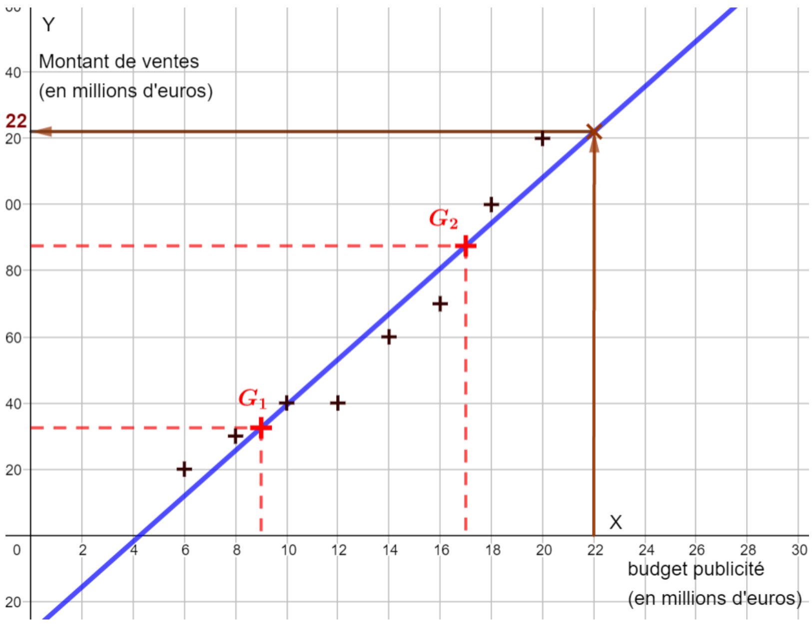
$\overline{x_2} = \frac{14+16+18+20}{4} = 17$ et $\overline{y_2} = \frac{60+70+100+120}{4} = 87{,}5$ donc $G_2(17\,;87{,}5)$

2.



3. On détermine graphiquement l'image de 22, on obtient environ 122

On estime que le montant des ventes est d'environ 122 millions pour un budget publicité de 22 millions d'euros :



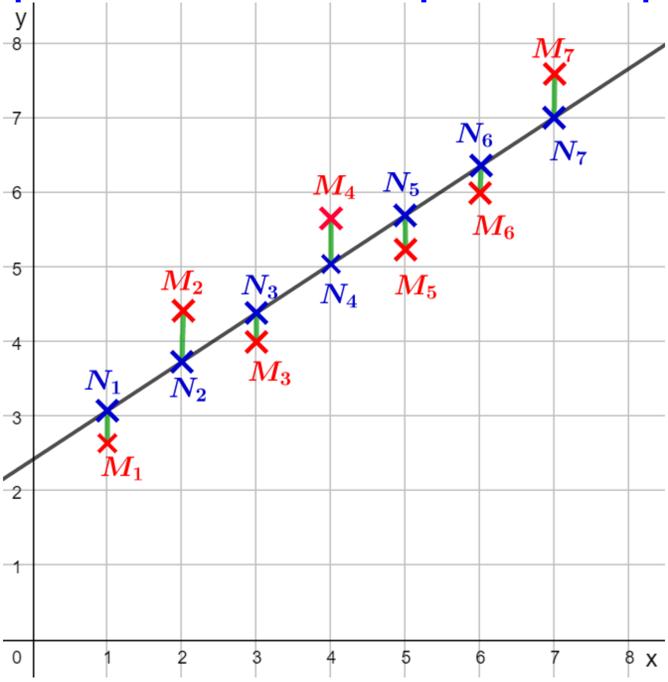
.

Cette méthode est possible lorsque les points sont proches de l'alignement, la droite tracée est appelée droite d'ajustement (ou droite de régression), passant au plus près de ces points. Si les points ne sont pas du tout alignés et loin de l'être alors cette méthode ne sera pas judicieuse.

3) Méthode des moindres carrés

a) Définition

Comme nous l'avons vu précédemment, si la forme d'un nuage de points $M_i(x_i; y_i)$ suggère un ajustement affine, on peut trouver un grand nombre de droites ajustant ce nuage de points.
La méthode des moindres carrés consiste à déterminer la droite rendant la somme $(M_1N_1)^2 + (M_2N_2)^2 + \cdots + (M_nN_n)^2$ la plus petite possible. On admettra que cette droite existe et qu'elle est unique.



L'équation de cette droite nous sera donnée par la calculatrice, en utilisant la régression linéaire du menu « statistiques » : la droite ayant une équation du type $y = a\,x + b$, la calculatrice nous fournira les coefficients a et b .

b) Exemple et utilisation de la calculatrice

Exemple : Déterminer la droite d'ajustement par la méthode des moindres carrés

On considère la série statistique à deux variables données dans le tableau suivant :

x_i	5	10	15	20	25	30	35	40
y_i	14	24	35	48	55	68	80	95

1. Dans un repère, représenter le nuage de points $(x_i\,;\,y_i)$.

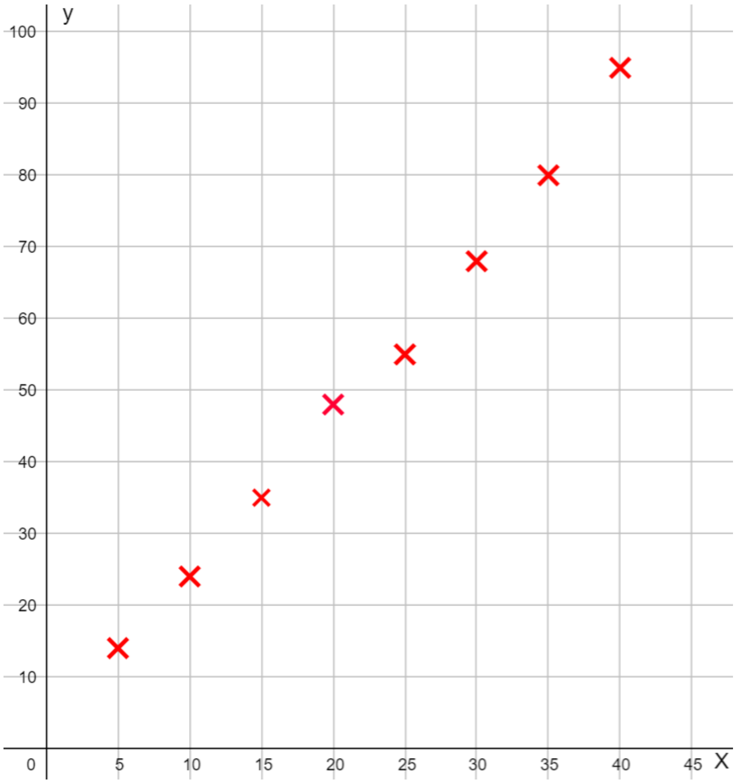
2. À l'aide de la calculatrice, déterminer une équation de la droite d'ajustement par la méthode des moindres carrés.

3. Représenter la droite d'ajustement de y en x.

4. Estimer graphiquement la valeur de x pour y = 60. Retrouver ce résultat par calcul.

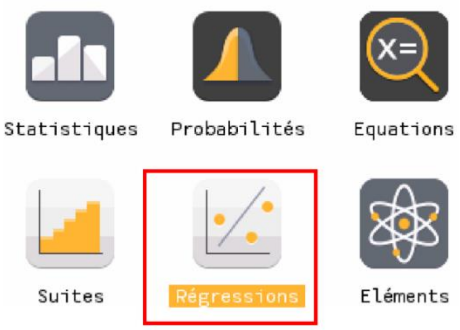
Réponse :

1.

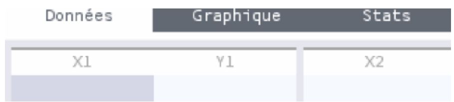


2. En utilisant la calculatrice Numworks :

• Aller dans le menu « Régressions » puis cliquer sur EXE.



Il s'affiche :

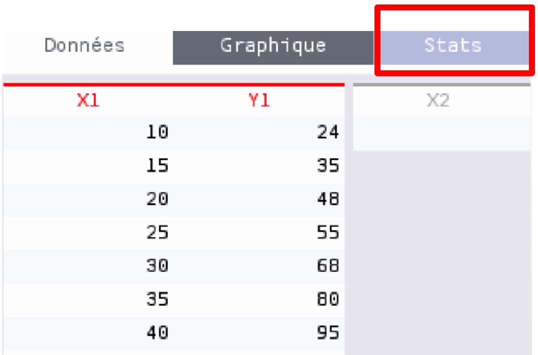


• On rentre les valeurs de notre série : les x_i dans la première colonne et les y_i dans la deuxième colonne comme ci-dessous

Données | Graphique | Stats

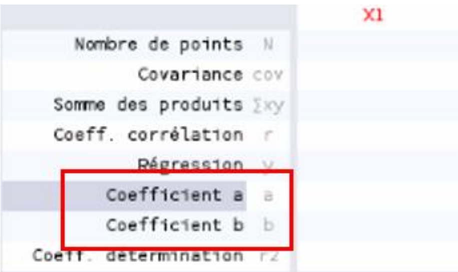
X1	Y1	X2
10	24	
15	35	
20	48	
25	55	
30	68	
35	80	
40	95	

● Puis on va dans l’onglet Stats puis on clique sur EXE :

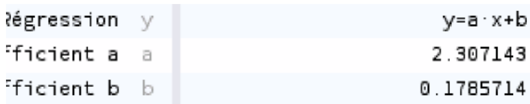


Il s’affiche

Avec les flèches on descend jusqu’aux lignes « coefficient a » et « coefficient b » :



Avec la flèche on va vers la droite et on obtient les résultats

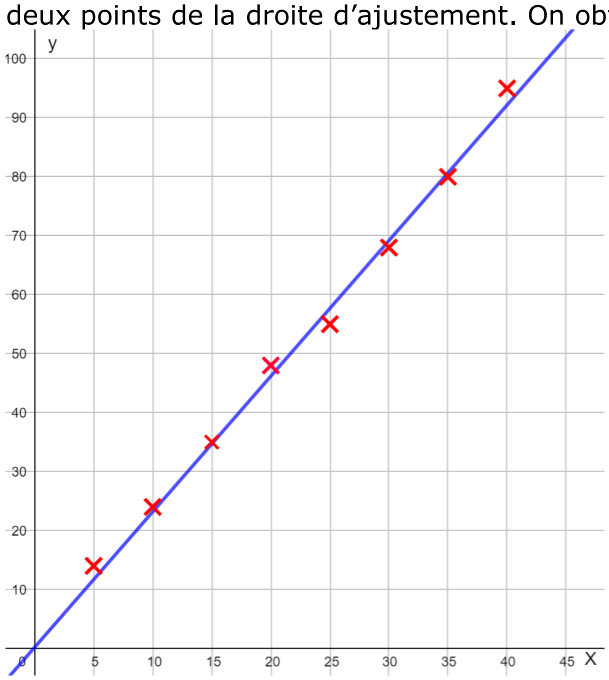


:

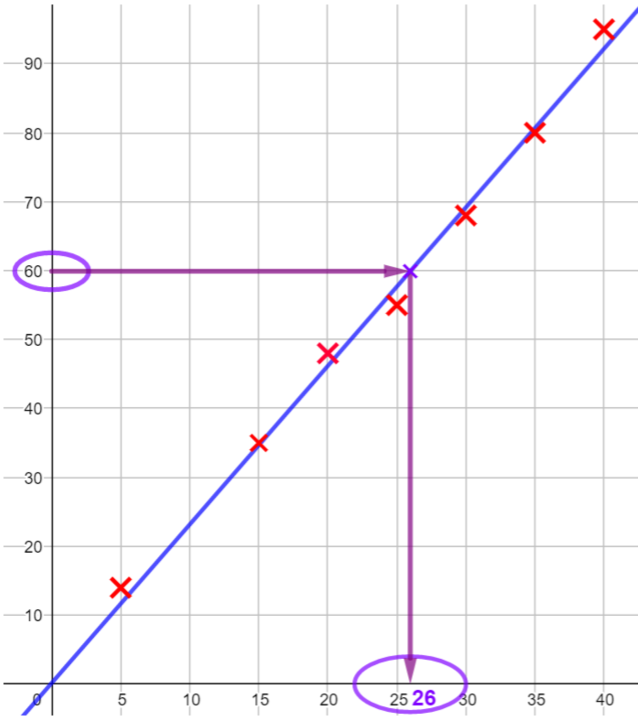
$a = 2{,}307143$ et $b = \ 0{,}1785714$

Une équation de la droite d’ajustement est : $y = 2{,}3x + 0{,}2$ (on arrondit)

Pour tracer la droite (ci-dessous tracée en bleu) il suffit de calculer les coordonnées de deux points de la droite d’ajustement. On obtient le graphique :



4. Pour $y = 60$ on lit graphiquement $x \approx 26$:



Par le calcul, la droite d'ajustement est : $y = 2,3x + 0,2$
si $y = 60$, alors $60 = 2,3x + 0,2$
Soit $2,3x = 60 - 0,2$
$2,3x = 59,8$
$x = \frac{59,8}{2,3} = 26$
Donc $x = 26$ lorsque $y = 60$

Remarque : Lorsque les calculs sont réalisés dans le domaine d'étude, on parle d'interpolation, dans notre cas 60 ∈ [14 ;95], on est bien dans le domaine d'étude. Dans le cas où l'on cherche des valeurs en dehors du domaine d'étude on parle d'extrapolation. Par exemple si on cherche la valeur de x pour $y = 120$, là on sort du domaine d'étude et dans ce cas on parle d'extrapolation.