STATISTIQUES

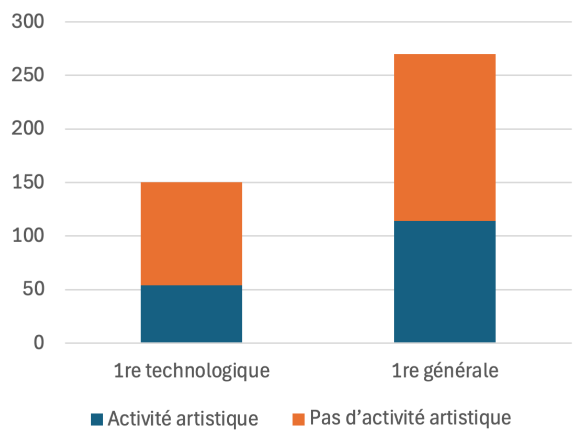
Partie 1 : Tableau croisé d'effectifs

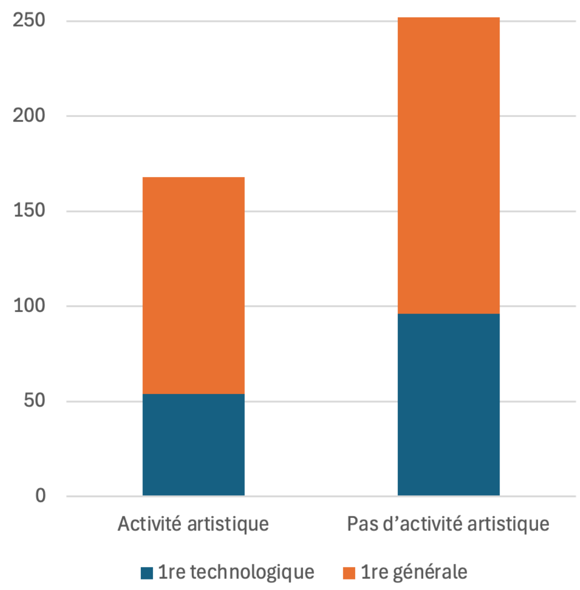
Exemple :
Dans un lycée, une enquête a été menée auprès des 420 élèves de niveau première concernant leur participation à des activités artistiques (théâtre, musique, arts plastiques…).
Le lycée compte 150 élèves en première technologique, les autres étant en terminale générale.
L'étude a montré que 168 élèves participent à une activité artistique. Parmi eux, 54 sont en filière technologique.
On peut présenter les données de cette enquête dans un tableau.

On commence d'abord par inscrire les **données de l'énoncé**.

	1re technologique	1re générale	Total
Activité artistique	**54**	168 – 54 = 114	**168**
Pas d'activité	150 – 54 = 96	270 – 114 = 156	420 – 168 = 252
Total	**150**	420 – 150 = 270	**420**

On peut également représenter les données du tableau dans des diagrammes en barres :





Partie 2 : Nuage de points

Définition : L'ensemble des points M_i de coordonnées $(x_i\ ;\ y_i)$, est appelé le **nuage de points** associé à la série statistique $(x_1\ ;\ y_1), (x_2\ ;\ y_2), \dots, (x_n\ ;\ y_n)$ à deux variables.

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

Méthode : Représenter un nuage de points

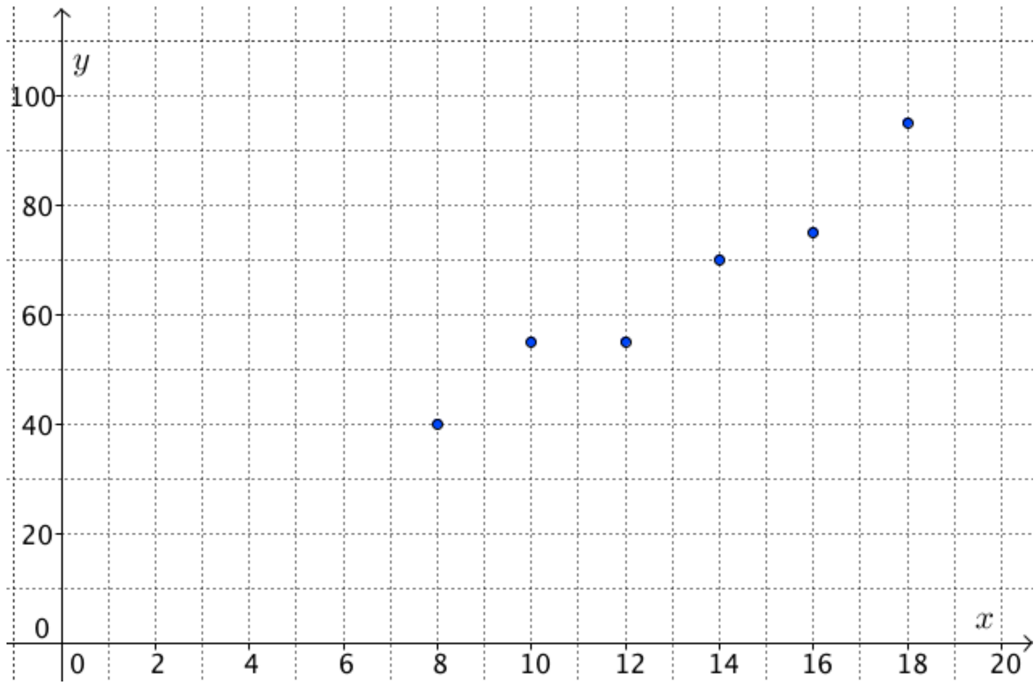
Vidéo https://youtu.be/Nn6uckb3RvE

Le tableau suivant présente l'évolution du budget publicitaire et du chiffre d'affaires d'une société au cours des 6 dernières années :

Budget publicitaire en milliers d'euros x_i	8	10	12	14	16	18
Chiffre d'affaire en milliers d'euros y_i	40	55	55	70	75	95

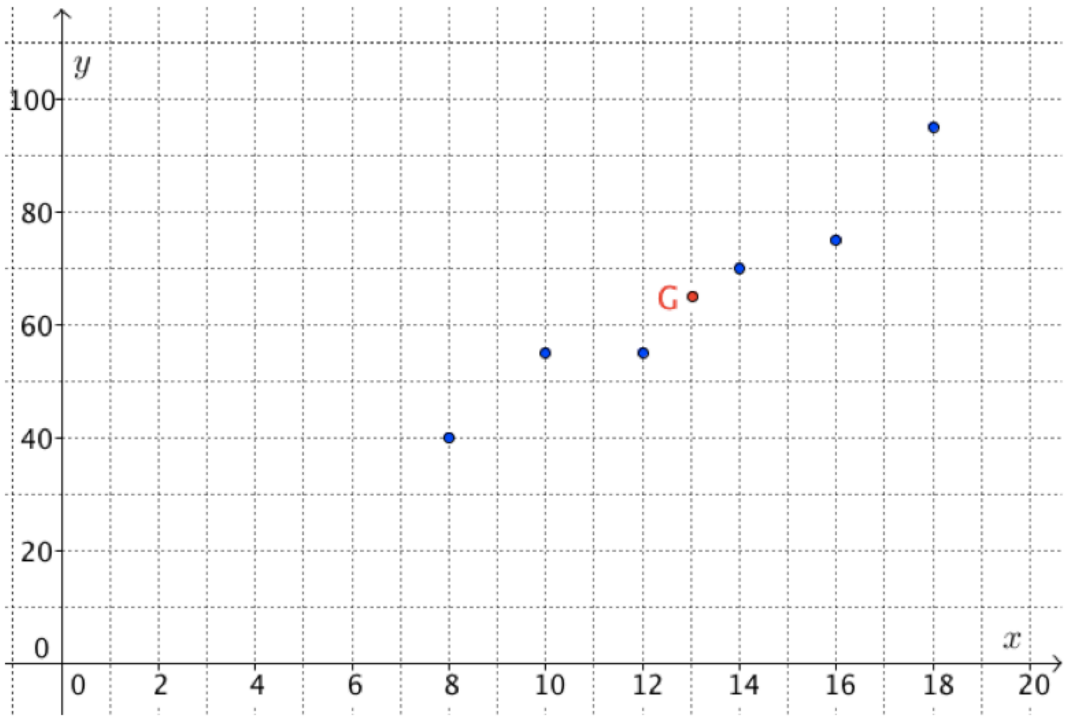
a) Dans un repère, représenter le nuage de points $(x_i\ ;\ y_i)$.
b) Déterminer les coordonnées du point moyen G du nuage de points.

Correction
a)



On a représenté ci-dessus le nuage de points de la série $(x_i\ ;\ y_i)$.

b) $\bar{x}$ = (8 + 10 + 12 + 14 + 16 + 18) : 6 = 13
$\bar{y}$ = (40 + 55 + 55 + 70 + 75 + 95) : 6 = 65.
Le point moyen G du nuage de points a pour coordonnées (13 ; 65). On peut placer ce point dans le repère.



Partie 3 : Ajustement affine

1) Interpolation, extrapolation

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.

Exemples :
- On donne une série exprimant la population d'une ville en fonction des années et on souhaite faire des prévisions pour les années à venir.
Les prévisions sortent du domaine d'étude de la série, on parle dans ce cas d'**extrapolation**.
- On donne une série exprimant la température extérieure et la consommation électrique correspondante. Les températures étudiées s'échelonnent entre -10°C et 10°C avec un pas de 4°C.
Sans faire de nouveaux relevés, on souhaite estimer la consommation électrique pour toutes les températures entières comprises entre -10°C et 10°C.
Les calculs sont dans le domaine d'étude de la série, on parle dans ce cas d'**interpolation**.

Définitions : L'**interpolation** et l'**extrapolation** sont des méthodes qui consistent à estimer une valeur inconnue dans une série statistique.
- Pour une interpolation, le calcul est réalisé dans le domaine d'étude fourni par les valeurs de la série.
- Pour une extrapolation, le calcul est réalisé en dehors du domaine d'étude.

La méthode d'extrapolation est parfois contestable car en dehors du domaine d'étude fourni par les valeurs de la série. Rien ne nous assure en effet que le modèle mathématique mis en œuvre soit encore valable.

2) Droite d'ajustement

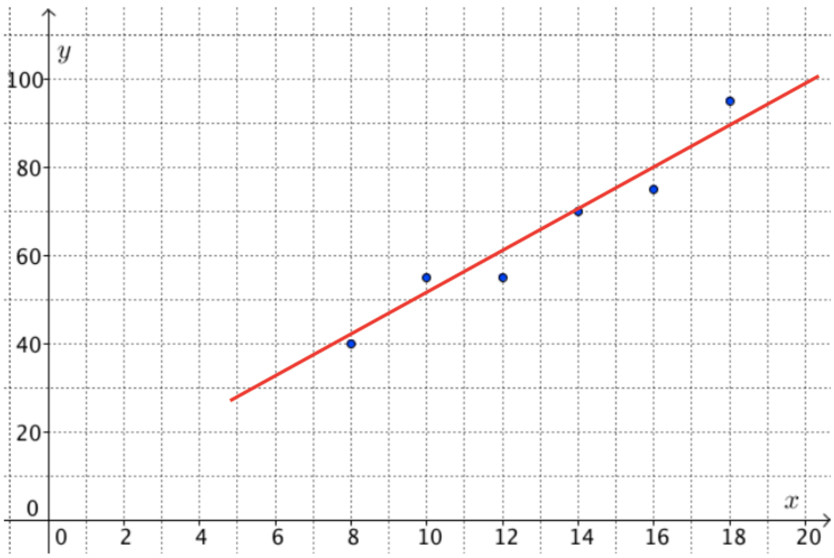
Pour obtenir de telles estimations, il faudra déterminer une droite passant « le plus près possible » des points du nuage.
L'interpolation ou l'extrapolation consiste à effectuer l'estimation par lecture graphique sur la droite ou par calcul à l'aide de l'équation de la droite.

Définition : Lorsque les points d'un nuage sont sensiblement alignés, on peut construire une droite, appelé **droite d'ajustement (ou droite de régression)**, passant « au plus près » de ces points.

Dans la suite, nous allons étudier différentes méthodes permettant d'obtenir une telle droite.

3) Méthode « au jugé »

Sans technique particulière, on trace « au jugé » une droite passant « au plus près » des points du nuage.
Cette méthode a le mérite d'être rapide mais pour être « juste », il faut un peu d'expérience.



On a tracé ci-dessus « au jugé » la droite d'ajustement du nuage de points de la méthode précédente.

4) Méthode des points moyens

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

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

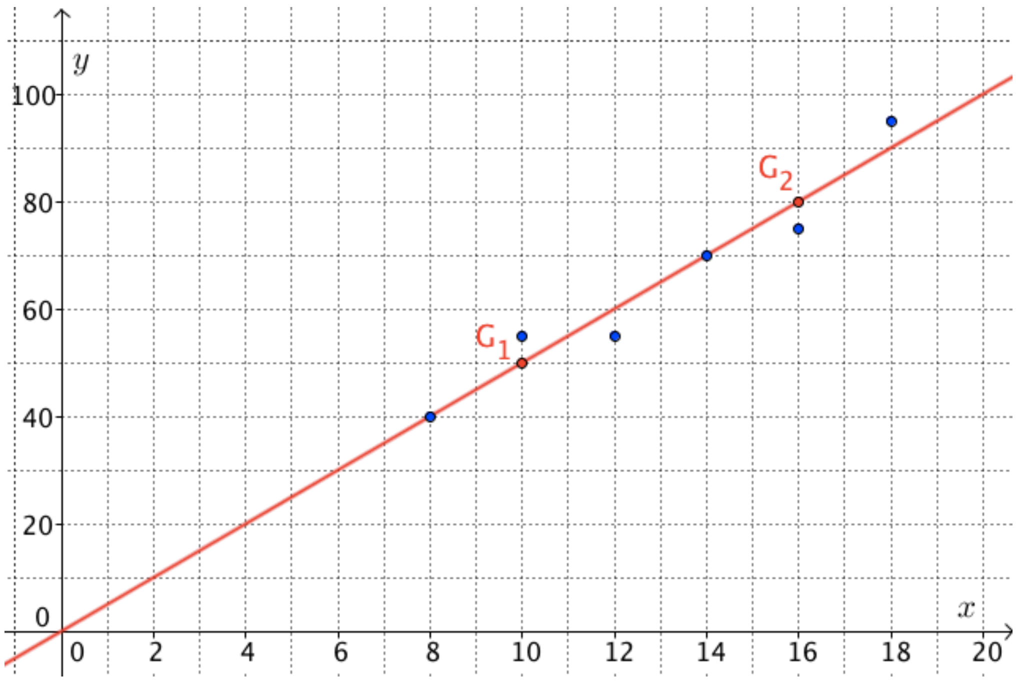
Vidéo https://youtu.be/ESHY4QPgriw

On reprend les données de la méthode de la partie 1.
1) Soit G_1, le point moyen associé aux trois premiers points du nuage et G_2 le point moyen associé aux trois derniers points du nuage.
  a) Calculer les coordonnées de G_1 et G_2.
  b) On prend (G_1G_2) comme droite d'ajustement. Tracer cette droite.
2) À l'aide du graphique :
  a) Estimer le chiffre d'affaires à prévoir pour un budget publicitaire de 22 000 €.
  b) Estimer le budget publicitaire qu'il faudrait prévoir pour obtenir un chiffre d'affaires de 100 000 €.
  c) La méthode utilisée dans les questions 2a et 2b consiste-t-elle en une interpolation ou une extrapolation ?

Correction
1) a) $\overline{x_1} = (8 + 10 + 12) : 3 = 10$
$\overline{y_1} = (40 + 55 + 55) : 3 = 50.$
Le point moyen G_1 a pour coordonnées (10 ; 50).

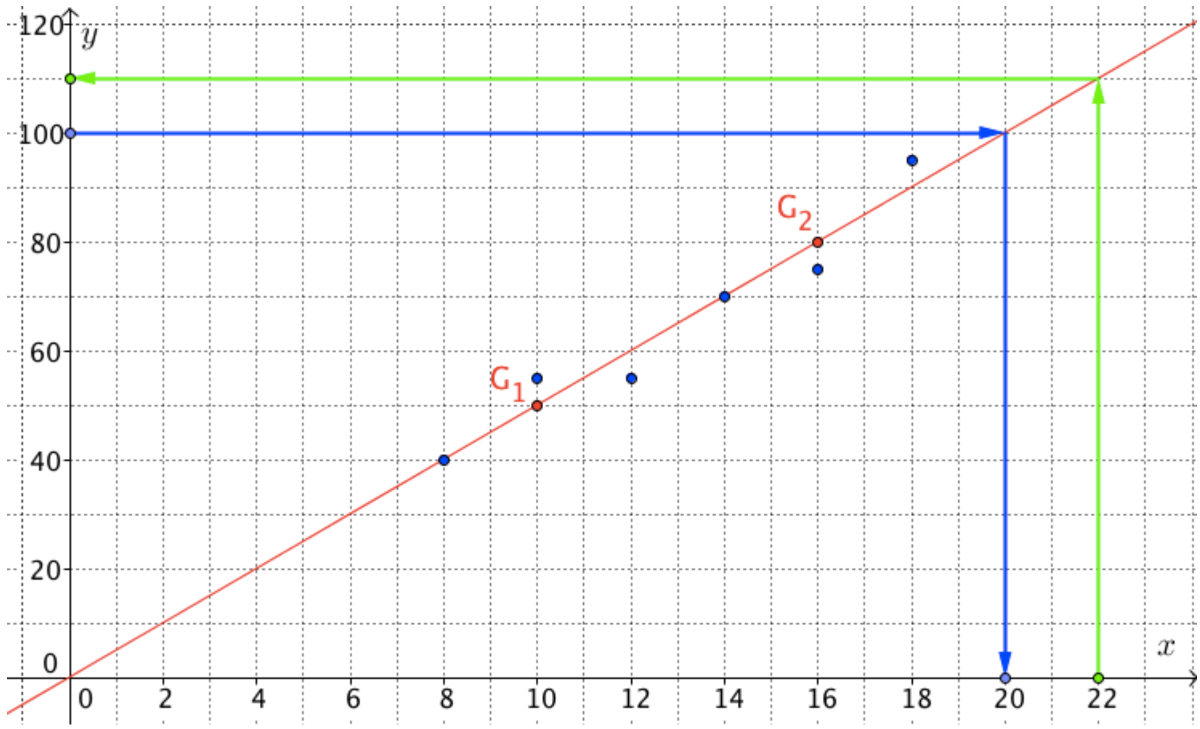
$\overline{x_2} = (14 + 16 + 18) : 3 = 16$
$\overline{y_2} = (70 + 75 + 95) : 3 = 80.$
Le point moyen G_2 a pour coordonnées (16 ; 80).

b)



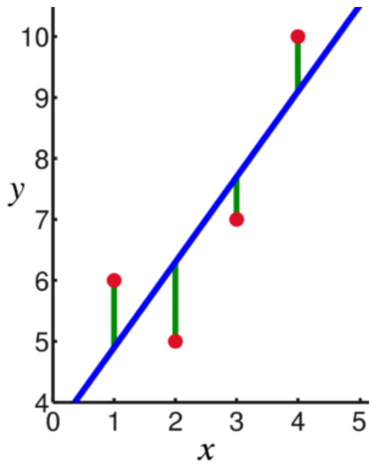
2) On lit graphiquement :
  a) Le chiffre d'affaires à prévoir pour un budget publicitaire de 22 000 € est de 110 000 €.
  b) Le budget publicitaire qu'il faudrait prévoir pour obtenir un chiffre d'affaires de

100 000 € est de 20 000€.



c) Les lectures graphiques sont réalisées ici en dehors du domaine d'étude, on parle donc d'extrapolation.

5) Méthode des moindres carrés



Cette méthode porte le nom de « moindres carrés » car elle consiste à rechercher la position de la droite d'ajustement tel que la somme des carrés des longueurs donnant les distances respectives (en vert) entre la droite et les points soit minimale.

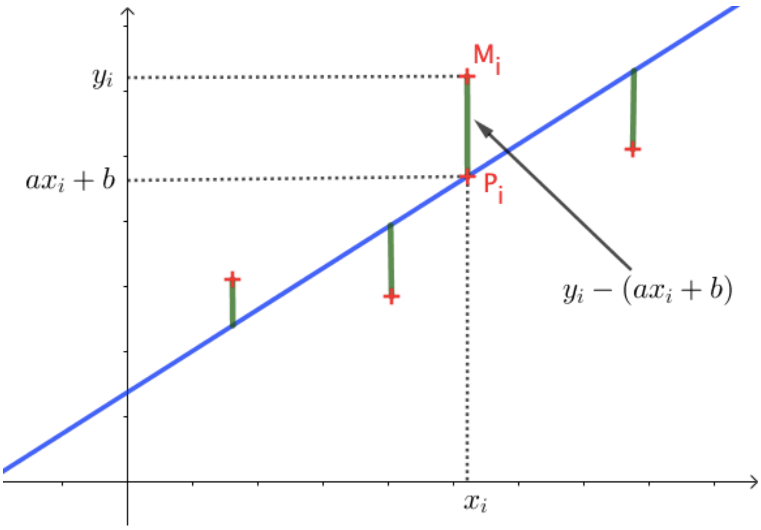
Le principe consiste donc à déterminer les coefficients a et b d'une droite d'équation $y = ax + b$ de sorte qu'elle passe le « plus près possible » des points du nuage.

Pour chaque abscisse x_i, on calcule la distance M_iP_i entre le point du nuage et le point de la droite, soit :

$$M_iP_i = |y_i - (ax_i + b)|$$

Il s'agit dans ce cas, de la droite d'ajustement de y en x.

A noter : Il existe également une droite d'ajustement de x en y en calculant les distances obtenues par projection horizontale.



Dans la méthode des moindres carrés, on recherche a et b pour lesquels la somme des carrés des distances est minimale, soit :

$$M_1P_1{}^2 + \cdots + M_nP_n{}^2 = \left(y_1 - (ax_1 + b)\right)^2 + \cdots + \left(y_n - (ax_n + b)\right)^2 \text{ est minimale.}$$

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

Vidéo https://youtu.be/vdEL0MOKAIg

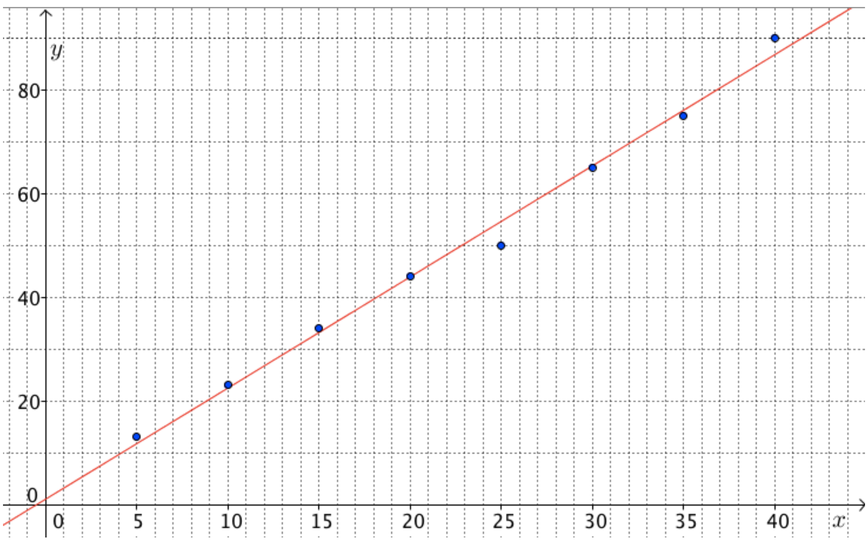
On considère la série statistique à deux variables données dans le tableau suivant :

x_i	5	10	15	20	25	30	35	40
y_i	13	23	34	44	50	65	75	90

1) Dans un repère, représenter le nuage de points $(x_i\ ;\ y_i)$.
2) a) À l'aide de la calculatrice, déterminer une équation de la droite d'ajustement par la méthode des moindres carrés.
   b) Représenter la droite d'ajustement de y en x.
3) Estimer graphiquement la valeur de x pour y = 70. Retrouver ce résultat par calcul. S'agit-il d'une interpolation ou d'une extrapolation ?

Correction

1)



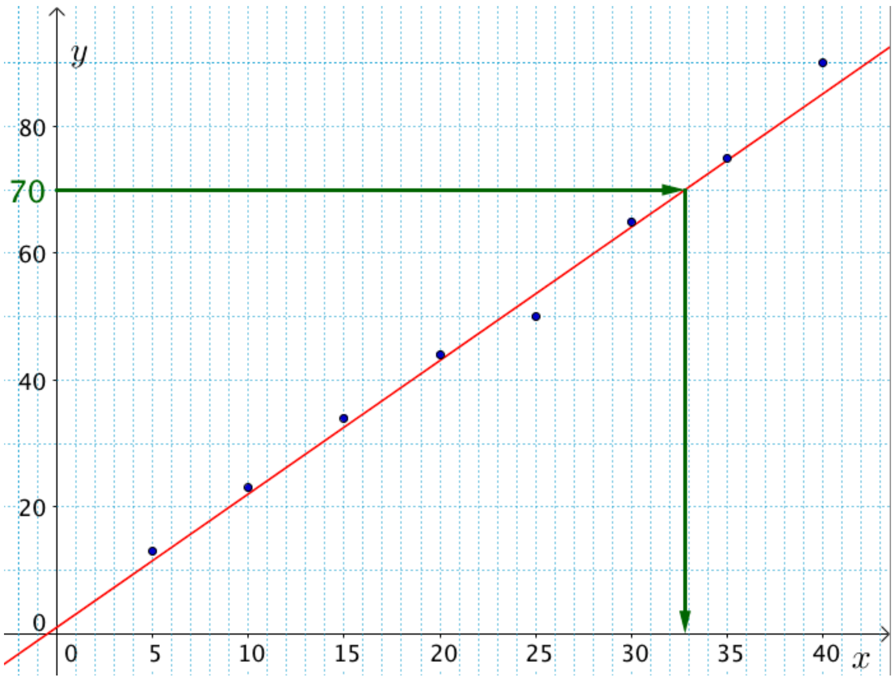
2) Avec TI :
- Appuyer sur « **STATS** » puis « **Edite** » et saisir les valeurs de x_i dans L1 et les valeurs de y_i dans L2.
- Appuyer à nouveau sur « **STATS** » puis « **CALC** » et « **RegLin(ax+b)** ».
- Saisir **L1,L2**

Avec CASIO :
- Aller dans le menu « **STAT** ».
- Saisir les valeurs de x_i dans *List1* et les valeurs de y_i dans *List2*.
- Sélectionner « **CALC** » puis « **SET** ».
- Choisir *List1* pour *2Var XList* et *List2* pour *2Var YList* puis « **EXE** ».
- Sélectionner « **REG** » puis « **X** » et « **aX+b** ».

La calculatrice nous renvoie : $a = 2.138095238$ et $b = 1.142857143$
Une équation de la droite d'ajustement est : $y = 2{,}1x + 1{,}1$

Pour tracer la droite, il suffit de calculer les coordonnées de deux points de la droite d'ajustement :

- Si $x = 0$ alors $y = 2{,}1 \times 0 + 1{,}1 = 1{,}1$ donc le point de coordonnées $(0\,;1{,}1)$ appartient à la droite d'ajustement.
- Si $x = 10$ alors $y = 2{,}1 \times 10 + 1{,}1 = 22{,}1$ donc le point de coordonnées $(10\,;22{,}1)$ appartient à la droite d'ajustement.



3) - Pour $y = 70$, on lit graphiquement $x \approx 33$.

- Par calcul, si $y = 70$, alors $70 = 2{,}1x + 1{,}1$

Soit $2{,}1x = 70 - 1{,}1$
$2{,}1x = 68{,}9$

$$x = \frac{68,9}{2,1} \approx 32,8$$

- Les calculs sont réalisés dans le domaine d'étude, on parle donc d'interpolation.