

# STATISTIQUE

## Partie 1 : 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 statistiques  $(x_1 ; y_1), (x_2 ; y_2), \dots, (x_n ; y_n)$  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>

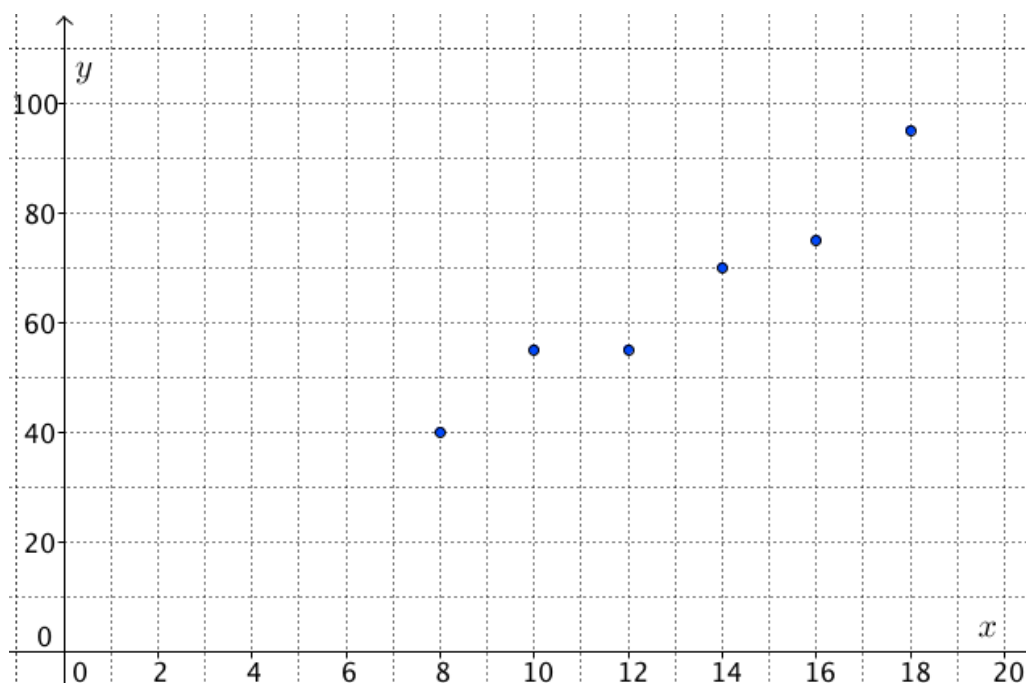
Le tableau suivant présente l'évolution du budget publicitaire et du chiffre d'affaire d'une société au cours des 6 dernières années :

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

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

### Correction

a)

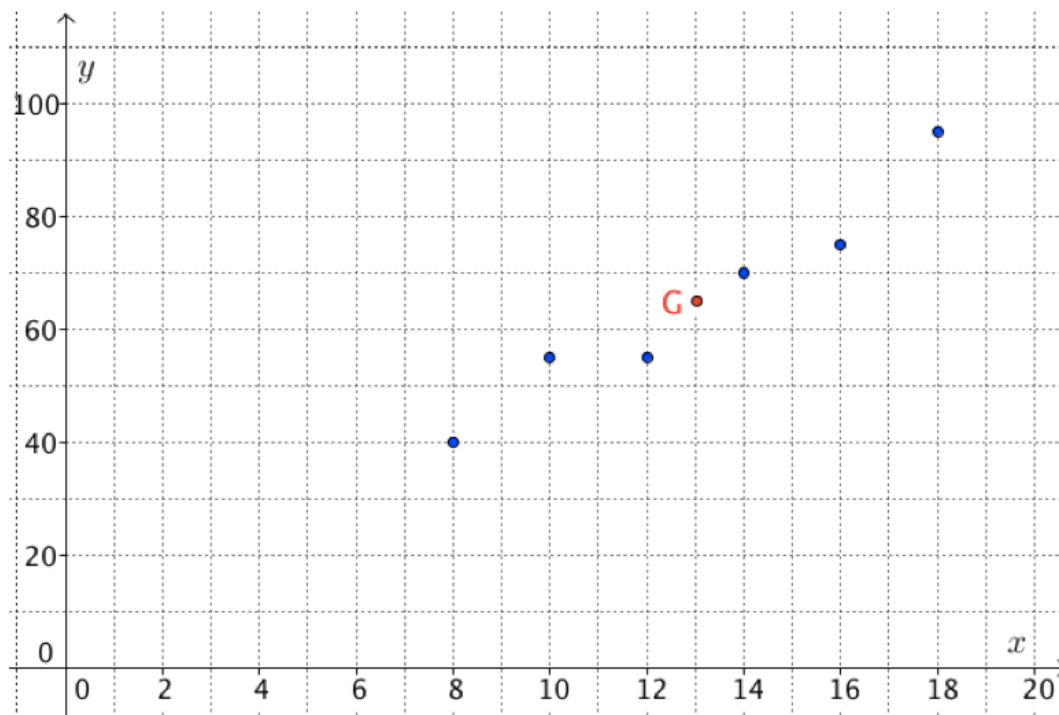


On a représenté ci-dessus le nuage de points de la série  $(x_i ; y_i)$ .

$$b) \bar{x} = (8 + 10 + 12 + 14 + 16 + 18) : 6 = 13$$

$$\bar{y} = (40 + 55 + 55 + 70 + 75 + 95) : 6 = 65.$$

Le point moyen G du nuage de points a pour coordonnées (13 ; 65). On peut placer ce point dans le repère.



## Partie 2 : 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  $-10^{\circ}\text{C}$  et  $10^{\circ}\text{C}$  avec un pas de  $4^{\circ}\text{C}$ .

Sans faire de nouveaux relevés, on souhaite estimer la consommation électrique pour toutes les températures entières comprises entre  $-10^{\circ}\text{C}$  et  $10^{\circ}\text{C}$ .

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.

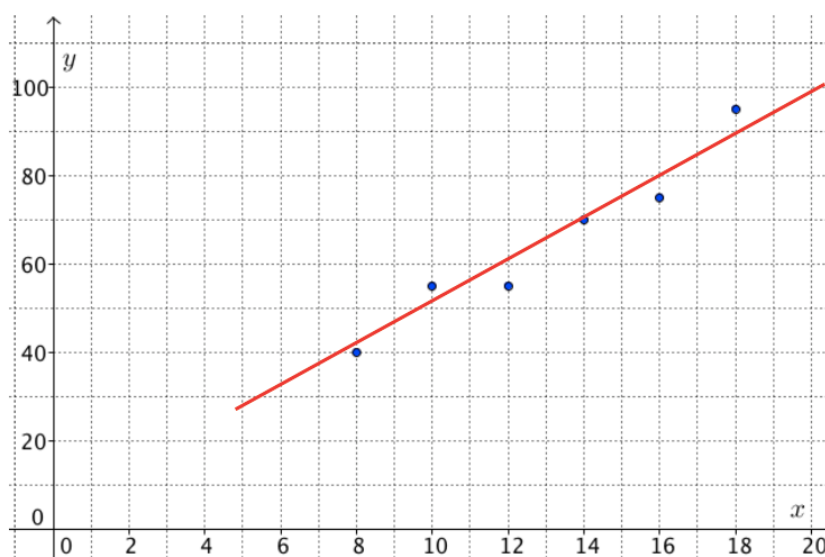
**Définition :** Lorsque les points d'un nuage sont sensiblement alignés, on peut construire une droite, appelé **droite d'ajustement (ou droite de régression)**, passant « au plus près » de ces points.

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/ESHY4QPgrw>

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'affaire à prévoir pour un budget publicitaire de 22 000 €.

b) Estimer le budget publicitaire qu'il faudrait prévoir pour obtenir un chiffre d'affaire de 100 000 €.

c) La méthode utilisée dans les questions 2a et 2b consiste-t-elle en une interpolation ou une extrapolation ?

#### Correction

$$1) a) \bar{x}_1 = (8 + 10 + 12) : 3 = 10$$

$$\bar{y}_1 = (40 + 55 + 55) : 3 = 50.$$

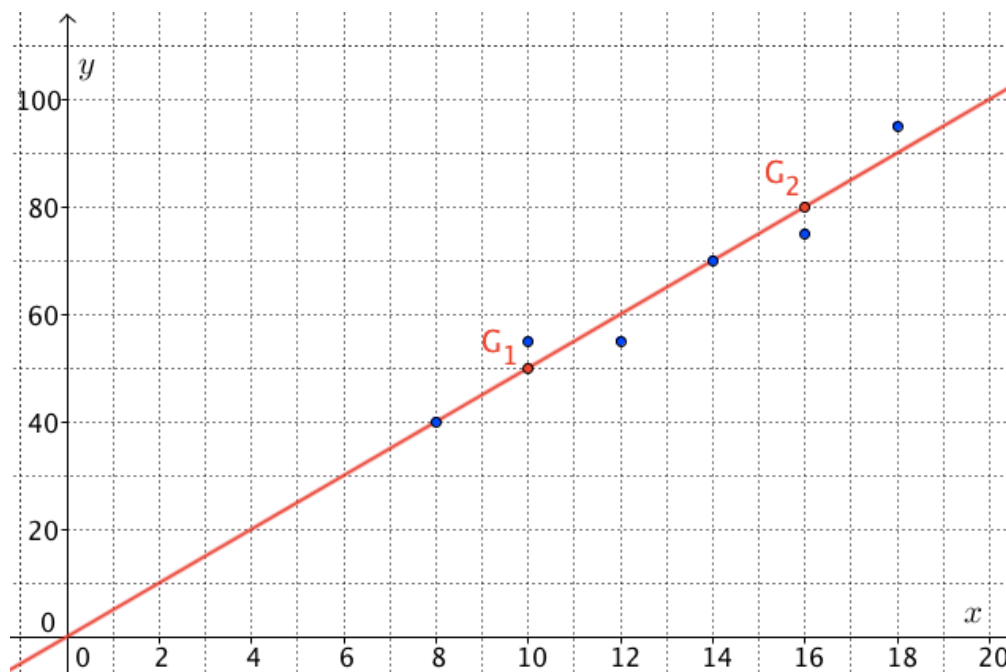
Le point moyen  $G_1$  a pour coordonnées (10 ; 50).

$$\bar{x}_2 = (14 + 16 + 18) : 3 = 16$$

$$\bar{y}_2 = (70 + 75 + 95) : 3 = 80.$$

Le point moyen  $G_2$  a pour coordonnées (16 ; 80).

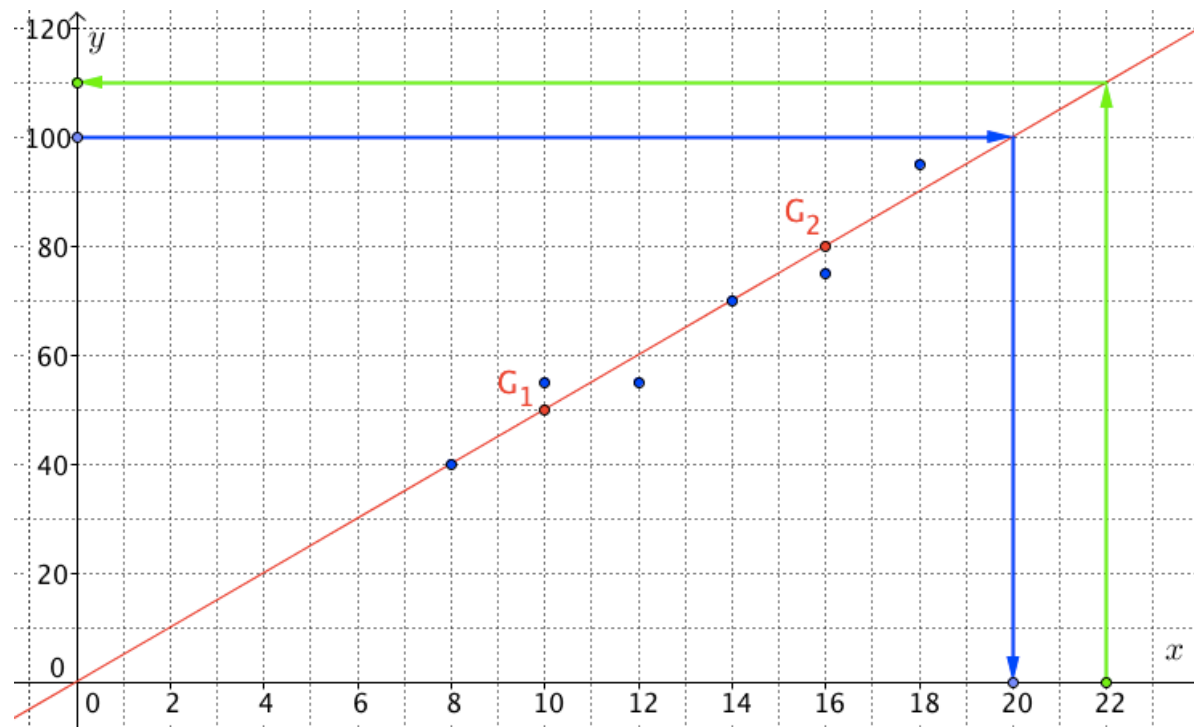
b)



2) On lit graphiquement :

a) Le chiffre d'affaire 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'affaire 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

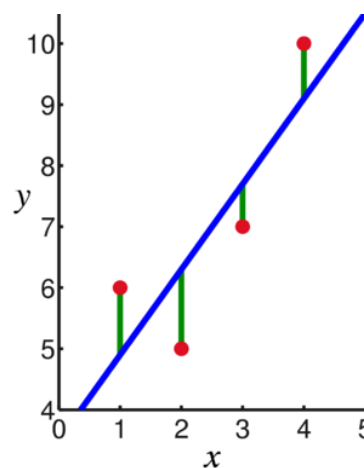
Cette méthode porte le nom de « moindre carrés » car elle consiste à rechercher la position de la droite d'ajustement tel que la somme des carrés des longueurs donnant les distances respectives (en vert) entre la droite et les points soit minimale.

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.

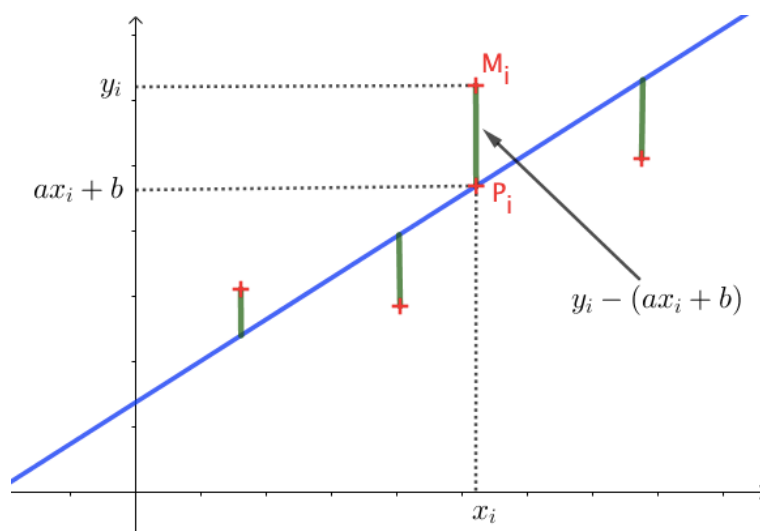
Pour chaque abscisse  $x_i$ , on calcule la distance  $M_i P_i$  entre le point du nuage et le point de la droite, soit :

$$M_i P_i = |y_i - (ax_i + b)|$$

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 :

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

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

**Vidéo** <https://youtu.be/vdELOMOKAlg>

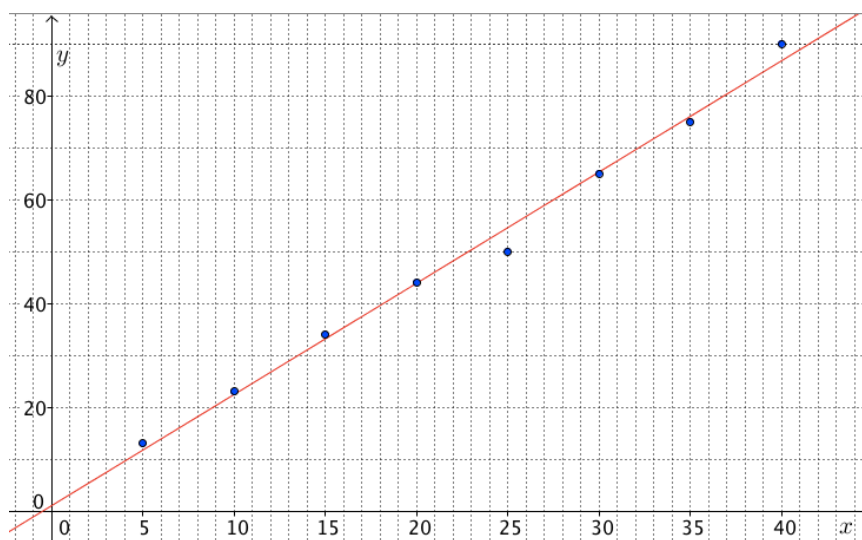
On considère la série statistique à deux variables données dans le tableau suivant :

$x_i$	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 :

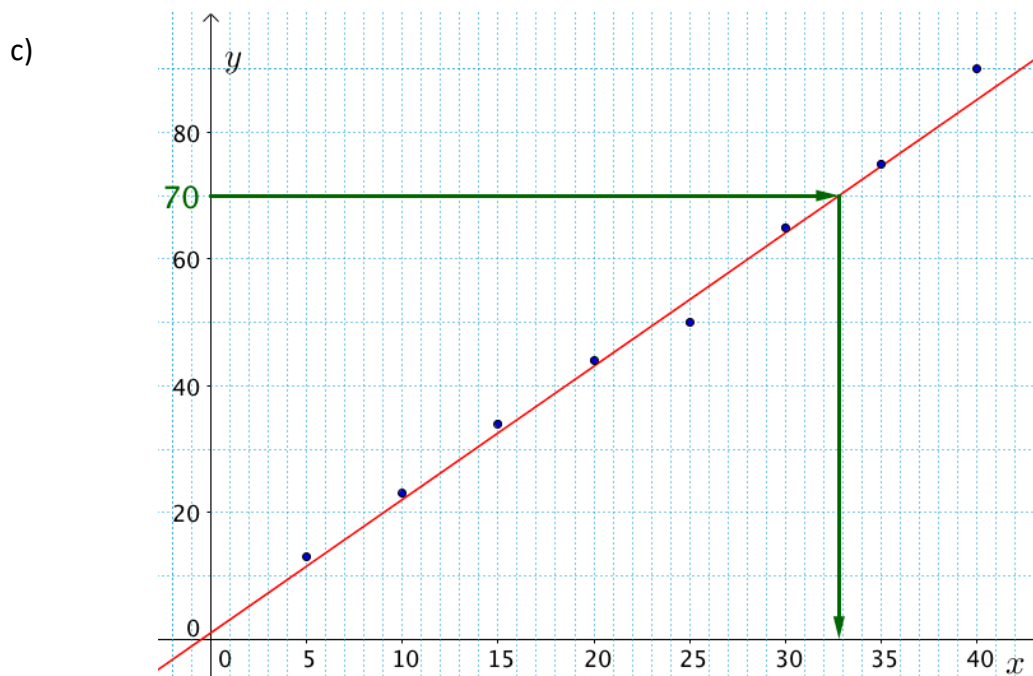
- Aller dans le menu « **STAT** ».
- Saisir les valeurs de  $x_i$  dans List1 et les valeurs de  $y_i$  dans List2.
- Sélectionner « **CALC** » puis « **SET** ».
- Choisir List1 pour 2Var XList et List2 pour 2Var YList puis « **EXE** ».
- Sélectionner « **REG** » puis « **X** » et « **aX+b** ».

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 domaine d'étude, on parle donc d'interpolation.

## Partie 3 : Ajustement par changement de variable

Lorsque le nuage de points n'est à priori pas modélisable par une droite, on peut réaliser un ajustement linéaire en effectuant un changement de variable.

**Méthode :** Effectuer un ajustement se ramenant par changement de variable à un ajustement affine

**Vidéo** <https://youtu.be/98L7oKSb0nA>

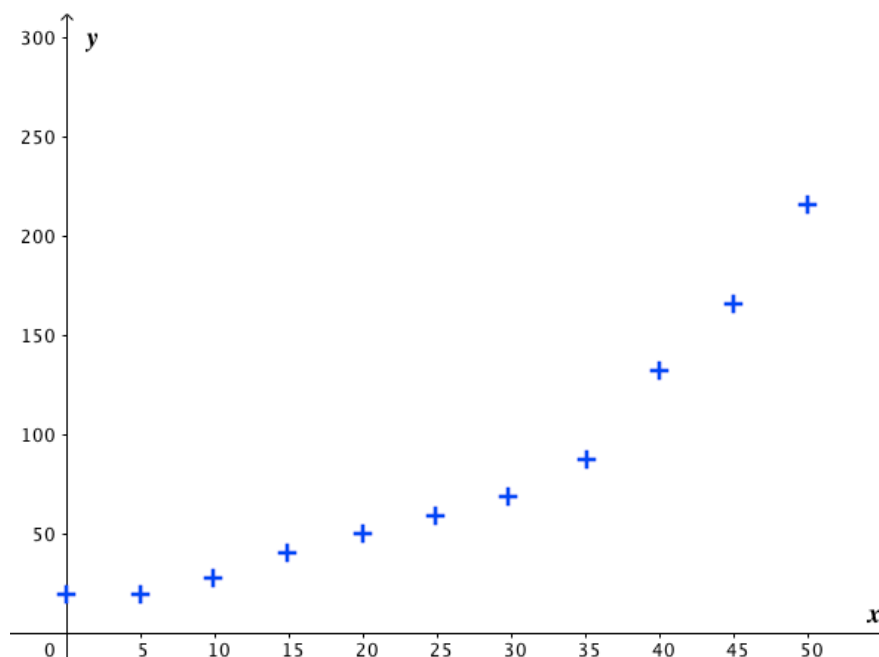
On a relevé la population d'une grande métropole sur 50 ans tous les 5 ans. Les résultats sont présentés dans le tableau suivant :

Année $x_i$	0	5	10	15	20	25	30	35	40	45	50
Population en milliers $y_i$	19,4	19,4	27,6	40,3	50	59	69	87	132	166	216

- 1) Représenter le nuage de points dans un repère.
- 2) a) On effectue le changement de variable  $z = \log y$ . Réaliser un nouveau tableau présentant les valeurs prises par les variables  $x$  et  $z$ .
  - b) Représenter un nouveau nuage de points à partir des données des variables  $x$  et  $z$ .
  - c) A l'aide la calculatrice, déterminer une équation de la droite d'ajustement de  $Z$  en  $X$  par la méthode des moindres carrés. Représenter la droite d'ajustement.
- 3) a) En déduire la relation qui lie  $y$  et  $x$  puis tracer la courbe représentative de la fonction  $f$  définie par  $y = f(x)$  dans le repère contenant le premier nuage de points.
  - b) En admettant que le modèle mathématique reste valable en dehors du domaine d'étude, extrapoler le nombre d'habitant 5 ans après l'étude.

### Correction

1)

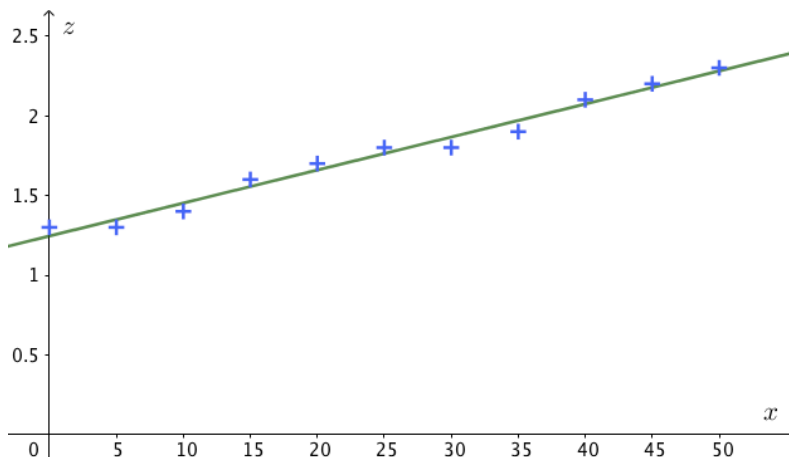




2) a)

$x_i$	0	5	10	15	20	25	30	35	40	45	50
$z_i = \log y_i$	1,29	1,29	1,44	1,61	1,7	1,77	1,84	1,94	2,12	2,22	2,33

b)



c) Une équation de la droite d'ajustement est :

$$z = 0,0214x + 1,2427$$

On trace la droite : voir ci-dessus.

L1	L2
0	1.29
5	1.29
10	1.44
15	1.61
20	1.7
25	1.77
30	1.84
35	1.94
40	2.12
45	2.22
50	2.33

**Réglin**

$$y = ax + b$$

$$a = 0.0213818182$$

$$b = 1.242727273$$

$$r^2 = 0.9883906748$$

$$r = 0.9941783918$$

3) a) On a :  $z = 0,0214x + 1,2427$

et :  $z = \log y$ , soit :

$$\log y = 0,0214x + 1,2427$$

$$10^{\ln y} = 10^{0,0214x + 1,2427}$$

$$y = 10^{0,0214x + 1,2427}$$

$$y = 10^{0,0214x} 10^{1,2427}$$

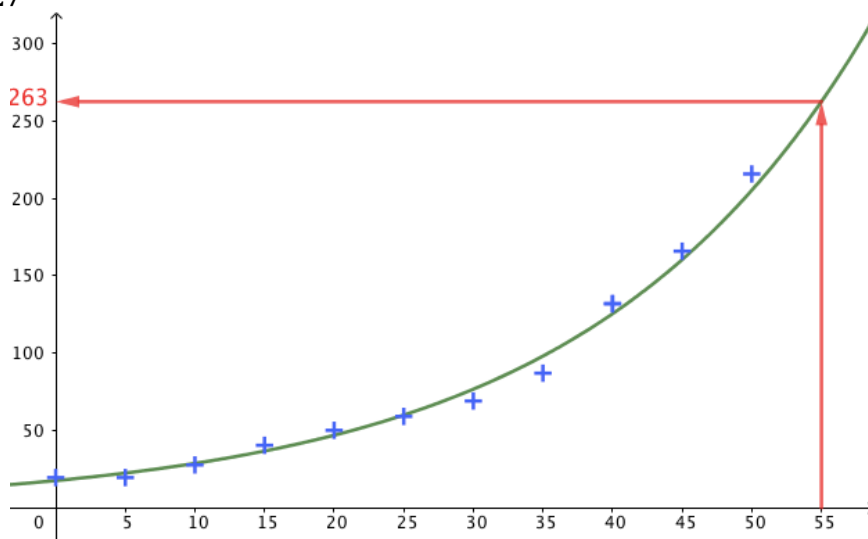
$$y = 17,4864 \times 10^{0,0214x}$$

Donc :

$$f(x) = 17,4864 \times 10^{0,0214x}$$

est l'expression de la fonction permettant d'ajuster le nuage de points  $(x_i ; y_i)$ .

On trace la courbe de  $f$  ci-contre.



b)  $y = 17,4864 \times 10^{0,0214 \times 55} \approx 263$

On peut supposer que 5 années après la fin de l'étude, la population de la ville sera proche de 263 000 habitants.

© Copyright

Hors du cadre de la classe, aucune reproduction, même partielle, autres que celles prévues à l'article L 122-5 du code de la propriété intellectuelle, ne peut être faite de ce site sans l'autorisation expresse de l'auteur.

[www.maths-et-tiques.fr/index.php/mentions-legales](http://www.maths-et-tiques.fr/index.php/mentions-legales)

Yvan Monka – Académie de Strasbourg – [www.maths-et-tiques.fr](http://www.maths-et-tiques.fr)