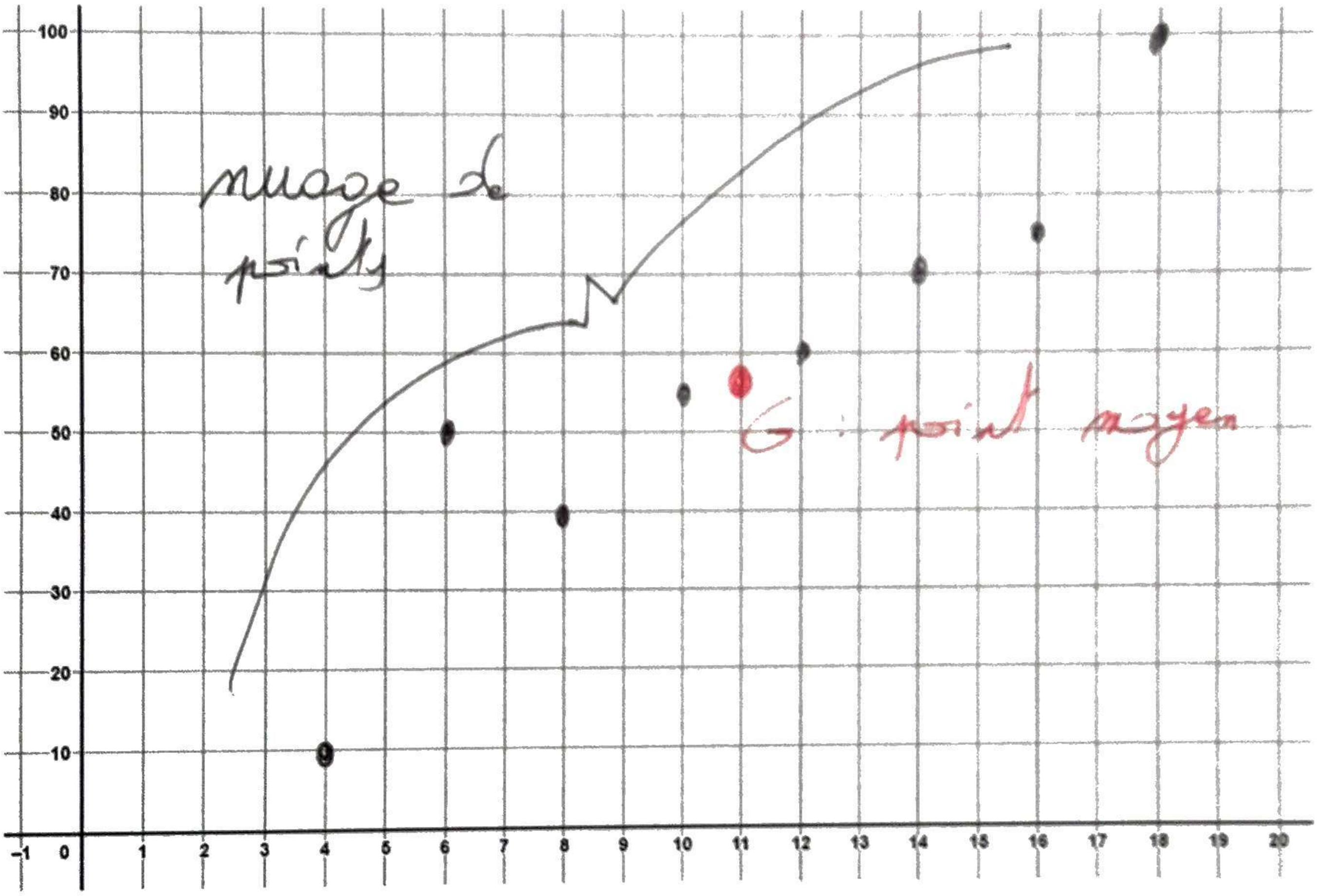
Chapitre 12 - Statistiques à 2 variables

1- Exemple d'introduction : ajustement sans changement de variable

a. Nuage de points et point moyen :

Le tableau suivant présente l'évolution du chiffre d'affaires d'une société sur 18 années :

Numéro de l'année	4	6	8	10	12	14	16	18
Chiffre d'affaires en milliers d'euros	10	50	40	55	60	70	75	100



Questions :

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

$G(\overline{x_G}\ ;\ \overline{y_G})$ et $\overline{x_G} = \dfrac{4+6+8+10+12+14+16+18}{8} = 11$

$\overline{y_G} = \dfrac{10+50+40+55+60+70+75+100}{8}$

$= 57,5$

Donc

$G(11\ ;\ 57,5)$

b. DROITE D'AJUSTEMENT PAR LA METHODE DE MAYER :

Questions :

Soit G1, le point moyen associé aux ~~trois~~ 4 premiers points du nuage et G2 le point moyen associé aux ~~trois~~ 4 derniers points du nuage.

- Calculer les coordonnées de G1 et G2.

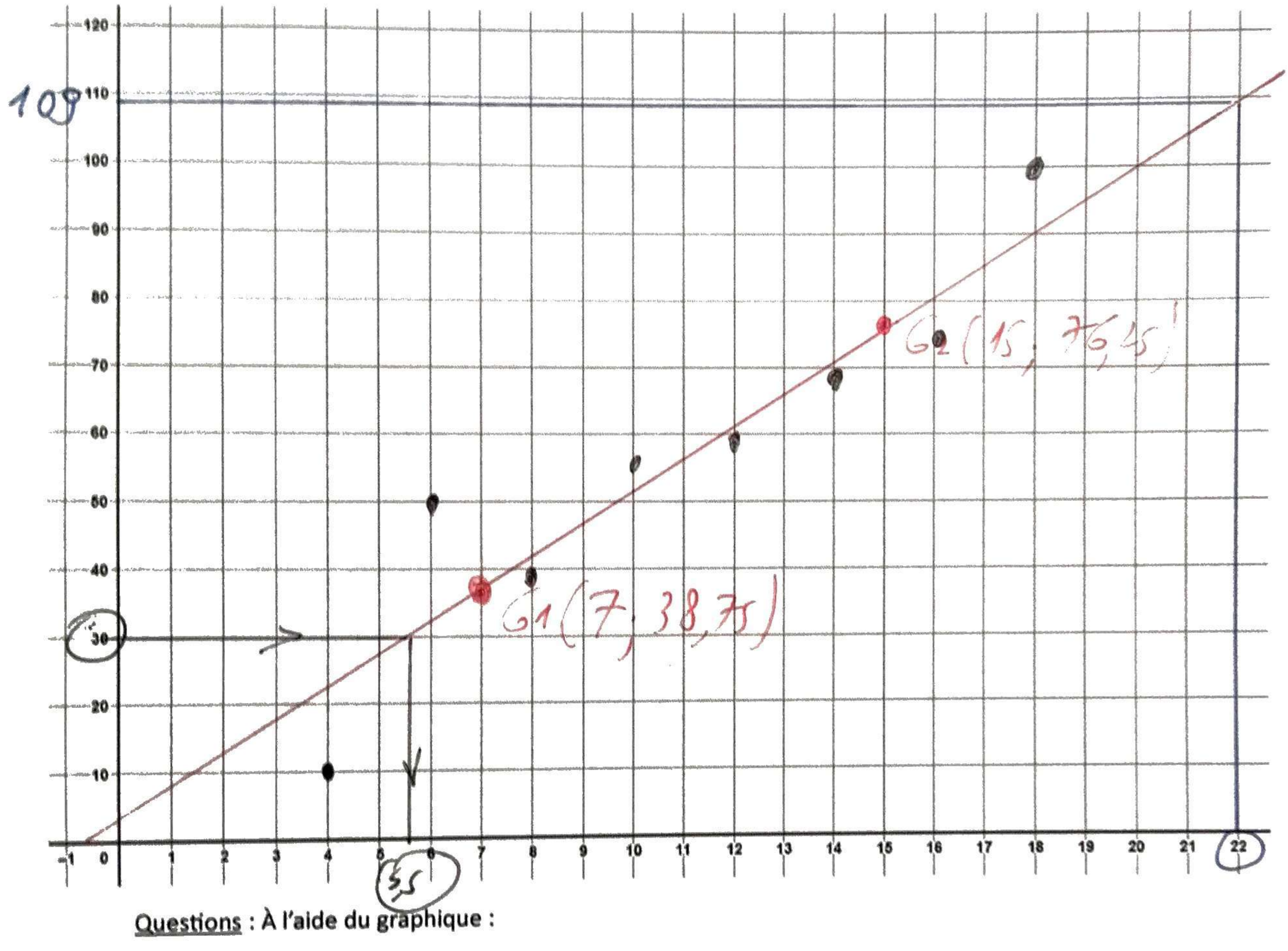
$x_{G1} = \frac{4+6+8+10}{4} = 7$

$y_{G1} = \frac{10+50+40+55}{4} = 38,75$

$x_{G2} = \frac{12+14+16+18}{4} = 15$

$y_{G2} = \frac{60+70+75+100}{4} = 76,25$

- On prend (G1G2) comme droite d'ajustement. Tracer cette droite.



Questions : À l'aide du graphique :

a) Estimer le chiffre d'affaires pour l'année 22 :

$f(22) = 108$ milliers d'euros. Pour l'année 22, le CA est de 108

b) Estimer le numéro de l'année pour lequel le chiffre d'affaires est de 30 milliers d'euros :

$f(5,5) = 30$. Le chiffre d'A est de 30 milliers d'euros sur l'année 5,5.

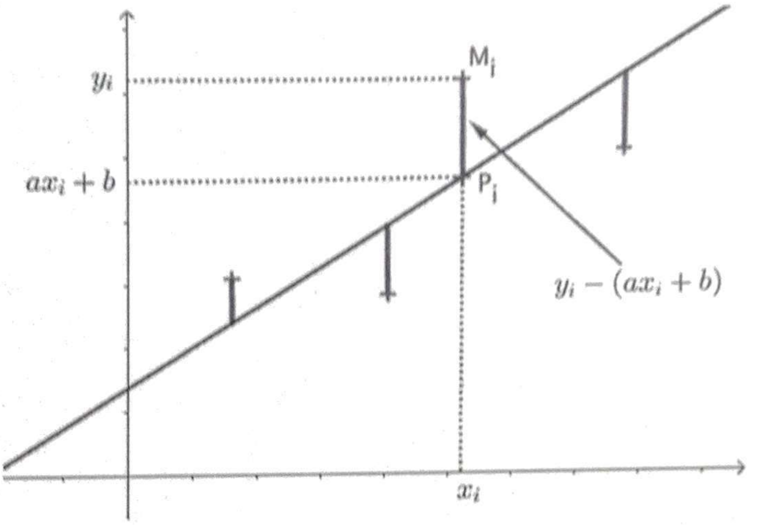
c. Droite d'ajustement par la methode des moindres CARRES :

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 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.



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

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

Pour cela, on peut appliquer la propriété suivante :

Propriété admise pour une série de n points :

La droite d'ajustement de y en x a pour équation $y = ax + b$, avec

Covariance

$$a = \frac{cov(x,y)}{var(x)} \quad \text{et} \quad b = \bar{y} - a\bar{x}$$

moyenne des y

Variance

moyenne des x

On a :

- $cov(x;y) = \frac{1}{n}\left((x_1 - \bar{x})(y_1 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})\right)$
- $var(x) = \frac{1}{n}\left((x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\right)$

On applique cette méthode à l'exemple étudié qui donne présente l'évolution du chiffre d'affaires d'une société sur 18 années :

Numéro de l'année	4	6	8	10	12	14	16	18
Chiffre d'affaires en milliers d'euros	10	50	40	55	60	70	75	100

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \times (y_i - \bar{y})$	$(x_i - \bar{x})^2$
4	10	-7	-47,5	332,5	49
6	50	-5	-7,5	37,5	25
8	40	-3	-17,5	52,5	9
10	55	-1	-2,5	2,5	1
12	60	1	2,5	2,5	1
14	70	3	12,5	37,5	9
16	75	5	17,5	87,5	25
18	100	7	42,5	297,5	49
Moyenne : $\bar{x} = 11$	Moyenne : $\bar{y} = 57,5$			Somme : 850	Somme : 168

$y = ax + b$

avec $a = \dfrac{\frac{1}{8} \times 850}{\frac{1}{8} \times 168} = \dfrac{850}{168} \simeq 5,0595$

$b = \bar{y} - a\bar{x} = 57,5 - 5,0595 \times 11$

$\simeq 1,8452$.

Donc :

$\underline{y = 5,0595\,x + 1,8452}$

À l'aide de la formule :

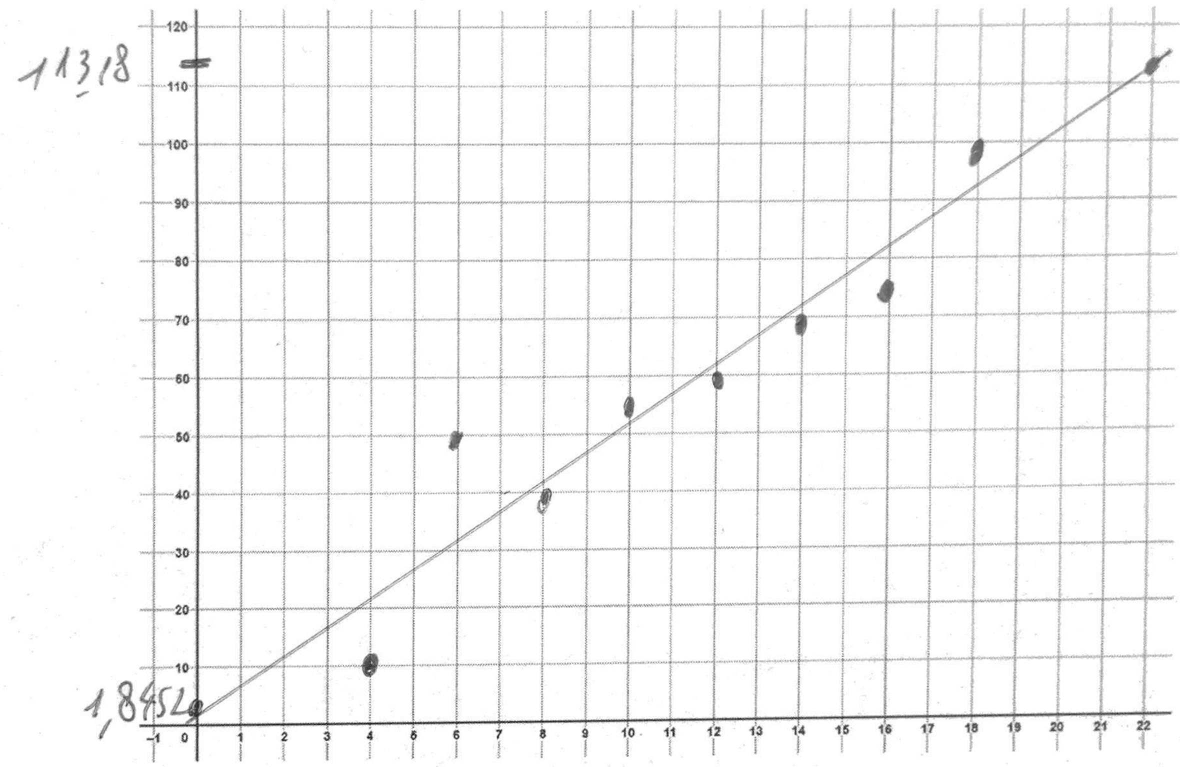
a) **Estimer le chiffre d'affaires pour l'année 22 :**

$y = 5{,}0595 \times 22 + 1{,}8452 \simeq 113{,}18$

b) **Estimer le numéro de l'année pour lequel le chiffre d'affaires est de 30 milliers d'euros :**

$5{,}0595 \times x + 1{,}8452 = 30$ d'où $x \simeq 5{,}56$

c) Tracer la droite d'ajustement :



d. COEFFICIENT DE CORRELATION :

Définition :

Le coefficient de corrélation de x et y est donné par la formule suivante :

$$\rho_{xy} = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}$$

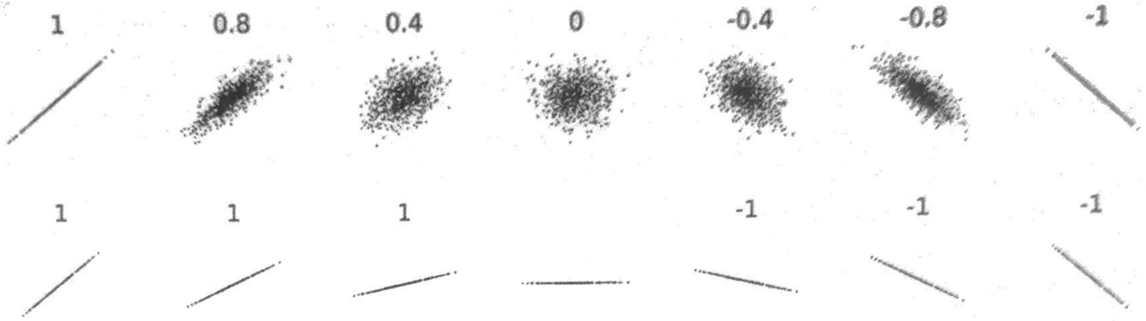
Le coefficient de corrélation ρ_{xy} est un nombre compris entre-1 et 1 qui mesure la relation entre les deux variables x et y. Plus le coefficient est proche des valeurs extrêmes-1 et 1, plus la corrélation linéaire entre les variables est forte.

- Si $\rho_{xy} > 0$, les valeurs prises par y ont tendance à croître quand les valeurs de x augmentent.

- Si $\rho_{xy} < 0$, les valeurs prises par y ont tendance à décroître quand les valeurs de x augmentent.

- Si $\rho_{xy} = 0$, les variations des variables x et y sont indépendantes.

Exemples de coefficients de corrélation :



Si la valeur de ρ_{xy} est proche de -1 ou 1, les données ont une répartition linéaire et ainsi la corrélation obtenue avec la droite d'ajustement est bonne. Par contre ce ,'est pas vrai si cette valeur est proche de 0.

Questions : En reprenant les données de l'exemple traité précédemment, calculer le coefficient de corrélation et interpréter le résultat :

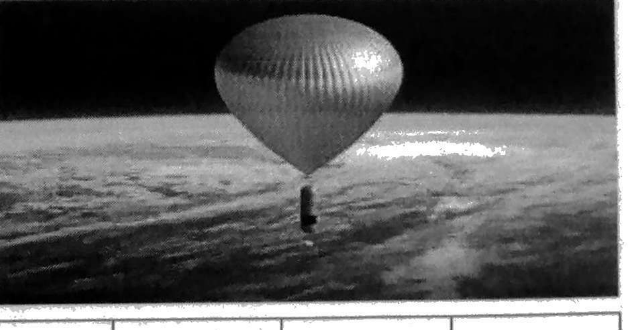
$y_i - \bar{y}$	$(y_i - \bar{y})^2$
-47,5	2256,25
-7,5	56,25
-17,5	306,25
-2,5	6,25
2,5	6,25
12,5	156,25
17,5	306,25
42,5	1806,25
	Somme : 4900

$$\rho_{xy} = \frac{\frac{1}{8} \times 850}{\sqrt{\frac{1}{8} 168 \times 4900}}$$

$$= \frac{850}{\sqrt{168 \times 4900}} \simeq 0,94$$

Interprétation : La corrélation obtenue avec la droite d'ajustement est bonne.

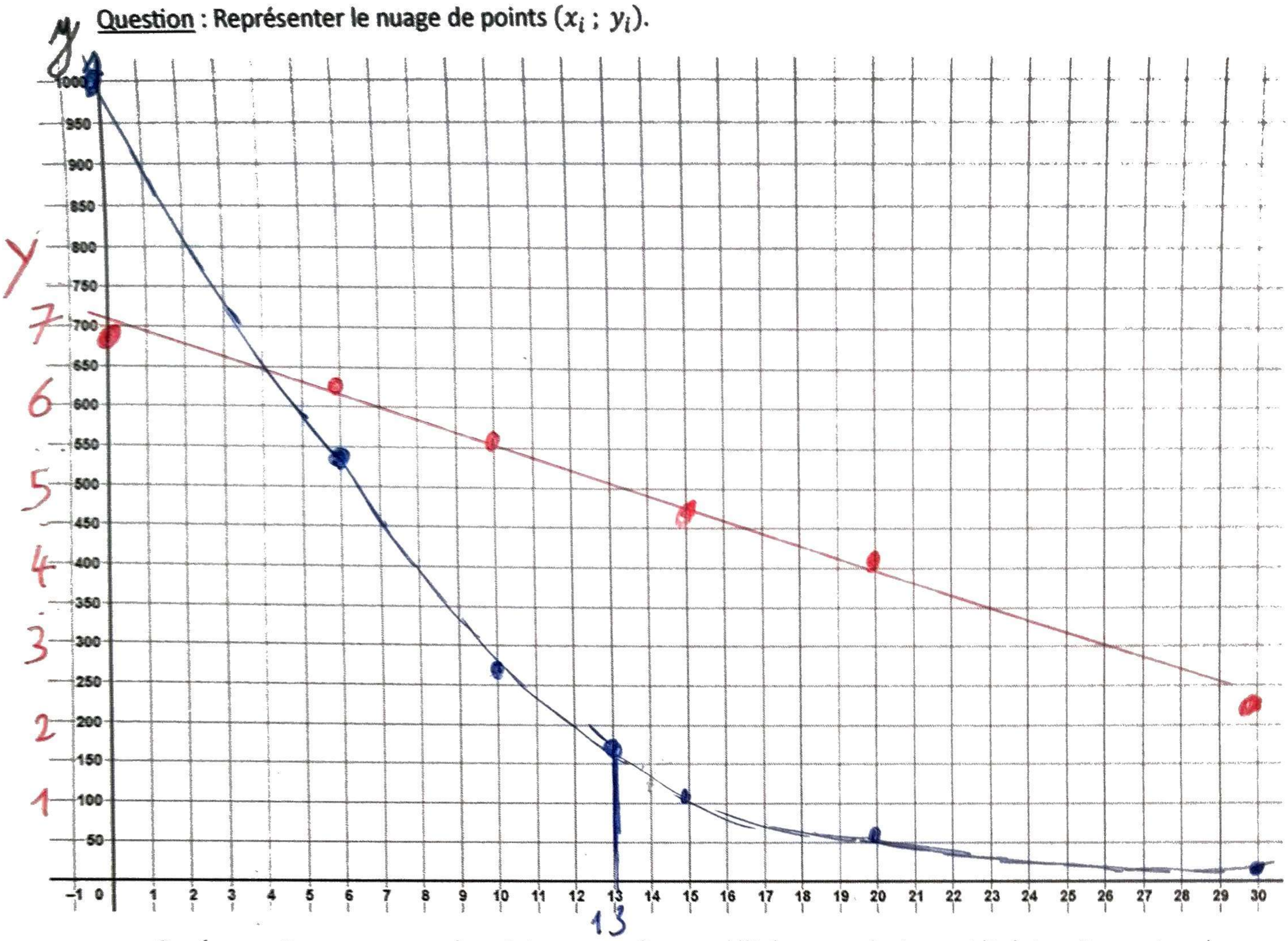
2- Autre exemple : ajustement avec changement de variable



Un ballon sonde a relevé la pression atmosphérique que l'on notera ici y en fonction de l'altitude que l'on notera ici x :

Altitude x_i en km	0	5	10	15	20	30
Pression y_i en mbar	1000	540	270	120	60	10

<u>Question</u> : Représenter le nuage de points $(x_i\ ;\ y_i)$.



On s'aperçoit que ce nuage de points ne peut être modélisé par un ajustement linéaire. Par contre, dans certains cas, il est possible de réaliser un changement de variable afin de retrouver une répartition linéaire. On propose ici de calculer pour chaque altitude x_i, la valeur du nombre $Y_i = \ln(y_i)$.

On obtient ainsi un nouveau tableau de valeurs :

Altitude x_i en km	0	5	10	15	20	30
Pression y_i en mbar	1000	540	270	120	60	10
$Y_i = \ln(y_i)$	6,91	6,29	5,60	4,79	4,09	2,30

En traçant à présent les points de coordonnées $(x_i\ , \ln(y_i))$, on constate que ce nouveau nuage de points a une forme linéaire.

On peut ainsi utiliser la méthode des moindres carrés pour déterminer les nombres a et b qui permettent d'établir la relation $Y = a\,x + b$, soit en tenant compte du changement de variable :

$$\ln(x) = a\,x + b$$

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \times (y_i - \bar{y})$	$(x_i - \bar{x})^2$
0	6,91	-13,3	1,91	-25,40	176,89
5	6,29	-8,3	1,29	-10,71	68,89
10	5,60	-3,3	0,60	-1,98	10,89
15	4,79	1,7	-0,21	-0,357	2,89
20	4,09	6,7	-0,91	-6,10	44,89
30	2,30	16,7	-2,70	-45,09	278,89
Moyenne : 13,3	Moyenne : 5,0			Somme : -89,64	Somme : 583,34

$$a = \frac{\frac{1}{6} \times (-89,64)}{\frac{1}{6} \times 583,34}$$

$$\underline{a = -0,1537}$$

$$b = \bar{y} - a\bar{x} = 5 + 0,1537 \cdot 13,3$$

$$\underline{b = 7,044}$$

Donc :

$$Y = -0,1537\, x + 7,044$$

Comme $Y = \ln(y)$

$$\ln(y) = -0,1537\, x + 7,044$$

Donc :

$$e^{\ln(y)} = e^{-0,1537\, x + 7,044}$$

$$= e^{-0,1537x} \times e^{7,044}$$

$$\boxed{y = 1146 \cdot e^{-0,1537x}}$$

Pression en mbar

altitude en km

VERIFICATION :

$x = 12$ km

$y = 1146\, e^{-0,1537 \times 12} \simeq 181$ mb.