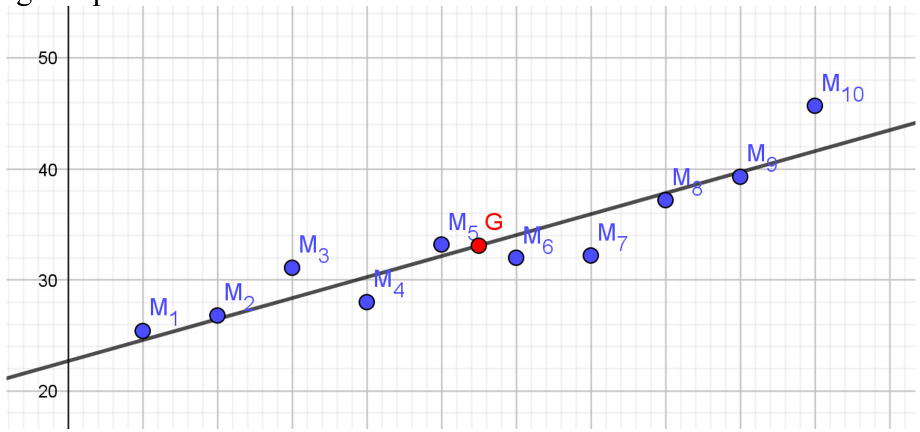
***Ex 1** : Nuage de points & ajustement affine (*) - 6 pts*

Le tableau suivant donne l'évolution du nombre de nuitées réservées dans les gîtes ruraux d'un département touristique, au cours de dix années :

Année	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Rang x_i	1	2	3	4	5	6	7	8	9	10
Nb de nuitées y_i (en milliers)	25,4	26,8	31,1	28,0	33,2	32,0	32,2	37,2	39,3	45,7

Le nuage de points est donné ci-dessous



La forme du nuage (points assez alignés) suggère un ajustement affine de y en x
les coordonnées du point moyen sont $G(5,5\,;33,09)$

car $\overline{x}=\frac{55}{10}=5,5$ et $\overline{y}=\frac{330,9}{10}=33,09$

$V_x=\frac{385}{10}-5,5^2=8,25$ donc $\sigma_x=\sqrt{8,25}\simeq 2,87$

$V_y=\frac{11294,51}{10}-33,09^2=34,5029$ donc $\sigma_y=\sqrt{34,5029}\simeq 5,87$

$cov(x,y)=\frac{1976}{10}-5,5\times 33,09=15,625$

l'équation de la droite des moindres carrés est $y=a\,x+b$

avec $a=\frac{15,625}{8,25}\simeq 1,89$ et $b=33,09-1,89\times 5,5\simeq 22,7$

d'où : $y=1,89\,x+22,7$

le coefficient de corrélation linéaire du couple $(x\,;y)$ est :

$r=\frac{15,625}{2,87\times 5,87}\simeq 0,927$; ainsi $r>0,9$ donc la corrélation est forte !

De plus le coefficient de détermination vaut $r^2\simeq 0,855$ ce qui signifie que le modèle « affine » reste fiable

Enfin l'erreur quadratique moyenne sur ce type d'ajustement est $MSe\simeq 6,23$ ainsi ce MSe est assez faible ce qui traduit une bonne fiabilité !

Le nombre de nuitées prévisibles en 2010 est : $y=1,89\times 20+22,7\simeq 60,5\,mille$
de plus $y=65$ donne $1,89\,x+22,7=65$ soit $x=22,4$
donc on peut espérer obtenir 65000 réservations dans les gîtes ruraux en 2023

<u>BONUS</u> : un autre ajustement possible de y en x est :
$y=0,067\,x^3-0,93\,x^2+5\,x+20,97$ avec un MSe de 2,9

***Ex 2** : Comparaison de deux ajustements non affines (**) - 6 pts*

On étudie le prix de revente d'un appareil électro-ménager au cours du temps
On dispose de l'étude statistique ci-dessous :

Nb d'années écoulées x_i	1	2	3	4	5	6
Prix de revente y_i (en €)	3000	2400	1920	1356	1229	983

<u>Ajustement affine :</u>
l'équation de la droite d'ajustement affine obtenue par la méthode des moindres carrés est $y=-404,63\,x+3230,9$
on a : $r\simeq -0,977$, $r^2\simeq 0,953$ et $MSe\simeq 34650$
donc le modèle d'ajustement reste correct mais d'une fiabilité très modérée !
Le prix (théorique) de revente de ce produit après 8 ans est de -6,14 € (soit 0 €)
cela signifie que ce produit sera devenu obsolète dans 8 ans !

<u>Ajustement non affine :</u>
l'expression d'un ajustement exponentiel de y en x est : $y=3719\,e^{-0,227\,x}$
on a : $r\simeq -0,993$, $r^2\simeq 0,986$ et $MSe\simeq 0,003$
donc le MSe est quasi parfait, ainsi cet ajustement reste très fiable !
Le prix (théorique) de revente de ce produit après 8 ans est de 605 €
(on observe que ce prix est beaucoup plus cohérent)

on calcule que le prix de ce produit sera aux alentours de 100 € dans 15 ans
cela signifie que ce produit sera devenu obsolète dans 15 ans !

BONUS : un 3eme type d'ajustement possible est : $g(x)=3341\times x^{-0,626\,x}$ avec $r\simeq -0,972$, $r^2\simeq 0,943$ et $MSe\simeq 0,01$

Ex 3** : **Étude d'une série « inconnue » (***) - 4 pts

Soit une série statistiques à deux variables $(x\,;y)$ définie par le tableau suivant :

x_i	1	2	5	x_4	11	13
y_i	24	y_2	29	31	34	39

Le point moyen de la série $(x_i; y_i)$ est $G(3{,}5\,;30{,}5)$

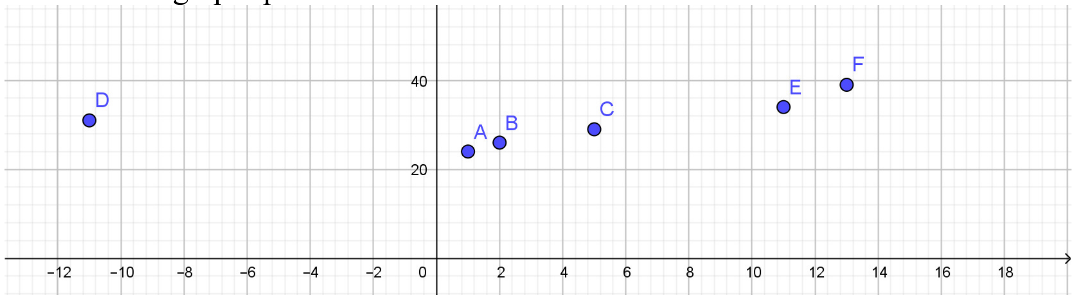
on sait que $\overline{x}=3{,}5$ donc $\dfrac{1+2+5+x_4+11+13}{6}=3{,}5$ donc $x_4=-11$

de même $\overline{y}=30{,}5$ donc $\dfrac{24+y_2+29+31+34+39}{6}=30{,}5$ donc $y_2=26$

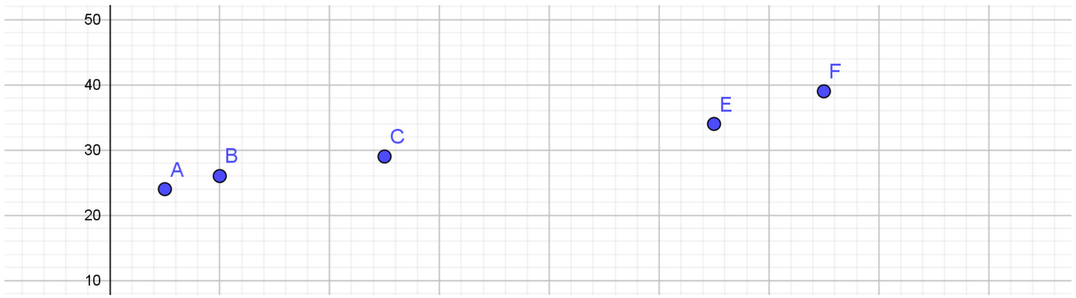
donc le tableau des données est :

x_i	1	2	5	-11	11	13
y_i	24	26	29	31	34	39

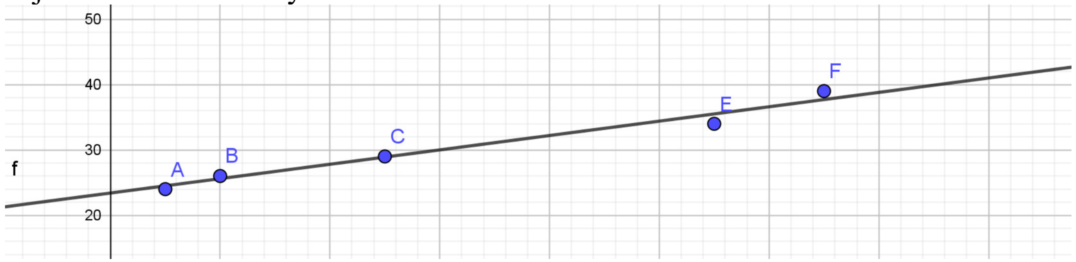
on obtient le graphique ci-dessous :



Ainsi, on observe clairement que le point $M_4(-11\,;31)$ est aberrant !
Donc on décide de supprimer cette valeur afin d'élaborer un ajustement de y en x



l'ajustement affine de y en x est le suivant :



à l'aide de la calculatrice on obtient : $y=1{,}12\,x+23{,}22$
les critères de fiabilité sont les suivants :
$r\simeq 0{,}985$, $r^2\simeq 0{,}971$ et $MSe\simeq 1{,}43$

l'ajustement obtenu par la fonction affine f définie par $f(x)=1{,}1\,x+23{,}4$ donne le tableau suivant :

x_i	1	2	5	11	13
y_i	24	26	29	34	39
$(y_i-f(x_i))^2$	0,25	0,16	0,01	2,25	1,69

Donc on déduit $MSe=\dfrac{1}{2}\times(0{,}25+0{,}16+0{,}01+2{,}25+1{,}69)=2{,}18$

on rappelle que $MSE=\dfrac{1}{n-ddl}\times\sum_{i=1}^{n}(y_i-a\,x_i-b)^2$ (ici $ddl=2$)

on en conclut que cet ajustement est moins fiable que celui obtenu par la méthode des moindres carrés
On choisit donc l'équation $y=1{,}12\,x+23{,}22$ comme ajustement le plus fiable

une estimation de y si $x=18$ est : $y=43{,}38$
une estimation de x si $y=52$ est : $x=25{,}7$