$$\Rightarrow \hat{a} = \frac{\sum_{i=1}^{i=N} y_i x_i - N\overline{x}.\overline{y}}{\sum_{i=1}^{i=N} x_i^{2} - N\overline{x}^{2}} = \frac{9360 - 6\times 55\times 19{,}67}{24100 - 6\times 55^{2}} \Rightarrow \hat{a} = 0{,}48$$

$$\Rightarrow \hat{b} = \overline{y} - \hat{a}\overline{x} = 19{,}67 - 0{,}48\times 55 \Rightarrow \hat{b} = -6{,}74$$

donc L'équation s'écrit : $y = 0{,}48x - 6{,}74$

3.3.3 Ajustement non linéaire d'un nuage de points

On considère N observations sur les deux variables X et Y.
Dans le cas général, la relation entre X et Y semble être plutôt non linéaire, c'est-à-dire n'est pas de la forme $y = ax + b$.
En fait lorsque le nuage de points manifeste en tendance courbe et que le coefficient de corrélation linéaire n'est pas proche de 1 en valeur absolue, l'ajustement de ce nuage par une droite est hasardeux et aboutira à des estimations de mauvaise qualité. Dans ce cas, on peut tenter d'utiliser un des modèles proposés dans ce paragraphe. En fait, chacun de ces modèles utilise le principe d'ajustement par la méthode des moindres carrés (donc ils utilisent tous une droite) mais en "transformant" au préalable les données pour obtenir un modèle linéaire à partir du modèle non-linéaire considéré.
On va voir les deux modèles d'ajustements non linéaires suivants :

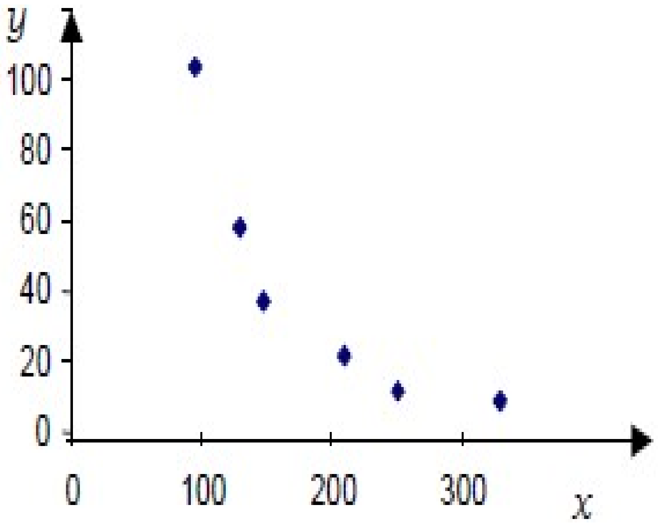
- Ajustement hyperbolique.
- Ajustement par une fonction puissance.

1er modèle : L'ajustement hyperbolique

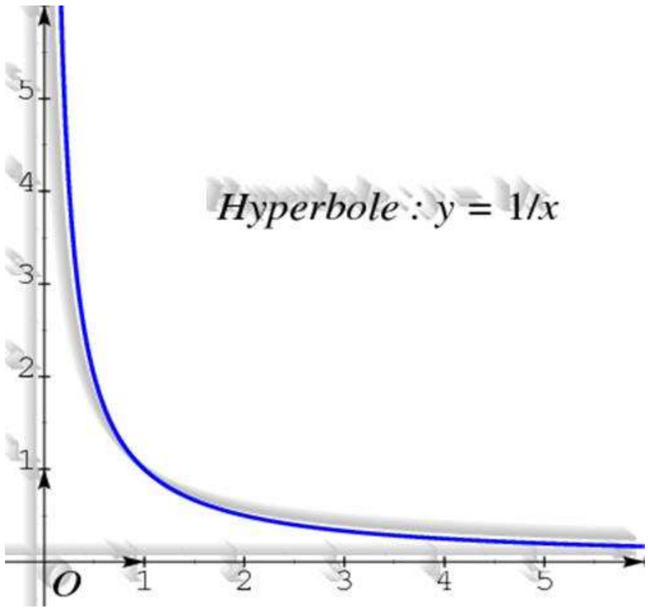
Les N points (x_i, y_i) ne sont pas alignés, mais plutôt proches d'une courbe représentant une fonction hyperbolique de la forme :

$$y = \frac{b}{x^{a}} \quad \text{avec } a > 0, \quad b > 0$$

Dans ce cas le nuage aura l'allure suivante



Qui ressemble à la courbe de la fonction $f(x)=\frac{1}{x}$ qui apparait comme suit :



<u>Comment peut-on estimer b et a ?</u>

Nous sommes en présence d'une relation non linéaire entre y et x.

Afin d'utiliser la méthode des **MCO**, il faut d'abord retrouver, moyennant une transformation, dans ce cas logarithmique, une forme linéaire :

On cherche a et b tels que :

$$y=\frac{b}{x^a}=bx^{-a}$$

En utilisant le logarithme népérien dans cette équation on trouve :

$$\Rightarrow \ln y=\ln bx^{-a}=\ln b-a\ln x$$

Et si on suppose que : $\beta=\ln b$ et $\alpha=-a$.

Le modèle linéaire est alors de la forme :

$$\ln y = \alpha \ln x + \beta$$

Donc en utilisant la méthode des **MCO**, on peut retrouver α et β :

$$\hat{\alpha} = \frac{Cov(\ln x, \ln y)}{V(\ln x)} = \frac{\sum_{i=1}^{i=N} (\ln x_i)(\ln y_i) - N\overline{\ln x}.\overline{\ln y}}{\sum_{i=1}^{i=N} (\ln x_i)^2 - N\overline{\ln x}^2}$$

$$\Rightarrow \hat{\beta} = \overline{\ln y} - \hat{\alpha}\overline{\ln x}$$

On peut maintenant retrouver la valeur de b et la valeur de a :

$\beta = \ln b \quad \Rightarrow \hat{b} = e^{\hat{\beta}}$ et $\alpha = -a \Rightarrow \hat{a} = -\hat{\alpha}$.

Exercice d'application

Une entreprise fabrique un équipement. Le prix unitaire Y (en Dollar) de ce produit est en fonction du nombre X d'unités produites. On a relevé les résultats suivants.

X	22	23	24	30	60	174
Y	120	60	25	10	4	1

1. Calculer la covariance entre le nombre d'unités produites X et le prix unitaire Y.
Que peut-on déduire sur la relation entre X et Y.

2. Calculer le coefficient de corrélation linéaire $r_{x,y}$.
Conclure sur l'intensité de la liaison entre les deux variables X et Y.

3. Représenter le nuage de points (x_i, y_i).

4. Compte tenue de cette représentation, donner la forme de l'ajustement de ce nuage de points et retrouver la relation entre les deux variables X et Y.

5. Quelle est Le prix unitaire du produit avec cette approximation pour produire 15 unités.

Solution de l'exercice

1. La covariance entre le nombre d'unités produites X et le prix unitaire Y.

$$Cov(x,y) = \left(\frac{1}{6}\sum_{i=1}^{i=6} x_i y_i\right) - (\bar{x}.\bar{y}) \text{ avec } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,\ \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

	x_i	y_i	x_i^2	$x_i \cdot y_i$
	22	120	484	2640
	23	60	529	1380
	24	25	576	600
	30	10	900	300
	60	4	3600	240
	174	1	30276	174
Total	333	220	36365	5334

$$\Rightarrow \bar{x} = \frac{333}{6} = 55{,}5; \quad \bar{y} = \frac{220}{6} = 36{,}667 \Rightarrow Cov(x,y) = \frac{5334}{6} - 55{,}5 \times 36{,}667$$

$$\Rightarrow Cov(x,y) = 889 - 2035{,}019 = -1146{,}019.$$

Conclusion

Comme $Cov(x, y) < 0$ alors alors la relation entre les deux variables est négative. Dans ce cas, ces deux variables varient en sens inverse.

2. Le coefficient de corrélation linéaire $r_{x,y}$.

$$r_{x,y} = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} \text{ avec } V(x) = \left(\frac{1}{N}\sum_{i=1}^{i=r} x_i^{\,2}\right) - \bar{x}^2 \text{ et } V(y) = \left(\frac{1}{N}\sum_{i=1}^{i=r} y_i^{\,2}\right) - \bar{y}^2$$

$$V(x) = \left(\frac{1}{N}\sum_{i=1}^{i=r} x_i^{\,2}\right) - \bar{x}^2 = \frac{36365}{6} - (55{,}5)^2 \Rightarrow V(x) = 2980{,}583.$$

Et $V(y) = \dfrac{18742}{6} - (36{,}667)^2 = 3123{,}667 - 1344{,}469 = 1779{,}198$

$$r_{x,y} = \frac{Cov(x,y)}{\sigma(x)\sigma(y)} = \frac{-1146{,}019}{\sqrt{2980{,}583 \times 1779{,}198}}$$

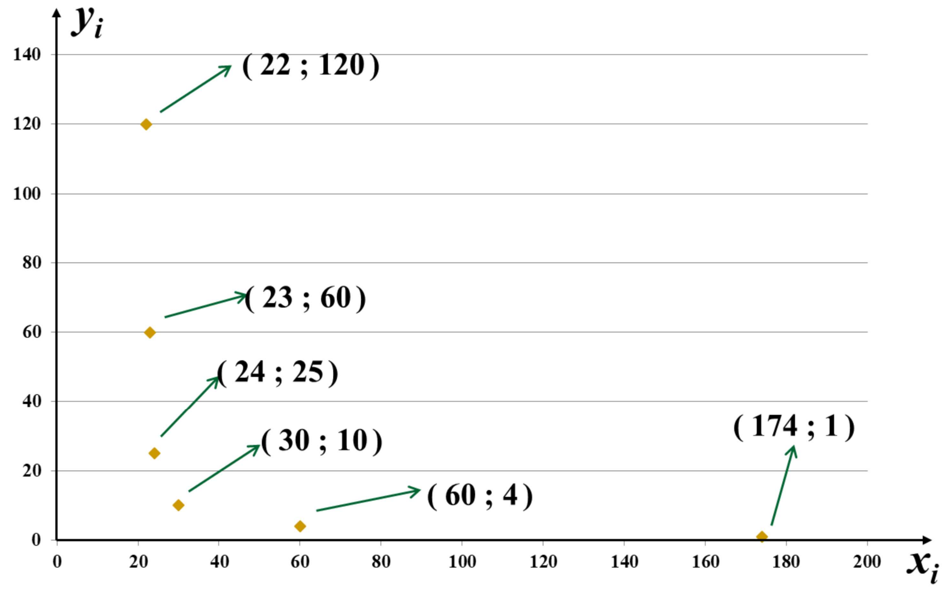
$$\Rightarrow r_{x,y} = \frac{-1146{,}019}{\sqrt{5303047{,}312}} = \frac{-1146{,}019}{2302{,}835} = -\ 0{,}498$$

Conclusion

La valeur de $r_{x,y}$ n'est pas proche ni de ± 1 ni de 0 cela traduit qu'il n'ya pas ni forte corrélation linéaire entre les deux variables ni faible corrélation linéaire.

Il y'a une juste moyenne corrélation linéaire entre les deux variables (La valeur de $r_{x,y}$ est proche de $0{,}5$).

3. Le nuage de points (x_i, y_i).



4. Compte tenue de cette représentation, donner la forme de l'ajustement de ce nuage de points et retrouver la relation entre les deux variables X et Y.

L'allure du nuage ressemble à une hyperbole, donc la forme théorique de l'ajustement de ce nuage de points est une forme hyperbolique de la forme :

$$y = \frac{b}{x^a} = bx^{-a} \text{ avec } a > 0, \quad b > 0$$

$$\Rightarrow \ln y = \ln bx^{-a} = \ln b - a\ln x$$

Posons : $\beta = \ln b$ et $\alpha = -a$.

$$\Rightarrow \ln y = \alpha \ln x + \beta$$

En considérant les deux (02) nouvelles variables et en utilisant la méthode des **MCO**, on peut retrouver α et β, tels que :

$$\Rightarrow \hat{\alpha} = \frac{\sum_{i=1}^{i=N}(\ln x_i)(\ln y_i) - N\overline{\ln x}.\overline{\ln y}}{\sum_{i=1}^{i=N}(\ln x_i)^2 - N\overline{\ln x}^2} \text{ et } \hat{\beta} = \overline{\ln y} - \hat{\alpha}\overline{\ln x}$$

	x_i	y_i	lnx_i	lny_i	$(lnx_i)(lny_i)$	$(lnx_i)^2$
	22	120	3,091	4,787	14,797	9,554
	23	60	3,135	4,094	12,835	9,828
	24	25	3,178	3,219	10,230	10,100
	30	10	3,401	2,303	7,833	11,567
	60	4	4,094	1,386	5,674	16,761
	174	1	5,159	0	0	26,615
Total	333	220	22,058	15,789	51,369	84,425

Donc $\overline{\ln x} = \frac{1}{6}\sum_{i=1}^{6} \ln x_i = \frac{22{,}058}{6} = 3{,}676$

$\overline{\ln y} = \frac{1}{6}\sum_{i=1}^{6} \ln y_i = \frac{15{,}789}{6} = 2{,}632$.

$\Rightarrow \hat{\alpha} = \frac{51{,}369 - 6 \times 3{,}676 \times 2{,}632}{84{,}425 - 6 \times (3{,}676)^2} = -1{,}9964459 \cong -2$

$\Rightarrow \hat{\beta} = \overline{\ln y} - \alpha \overline{\ln x} = 2{,}632 + 2 \times 3{,}676 = 2{,}632 + 7{,}352 = 9{,}984$

$\Rightarrow a = 2,\ \ \beta = \ln b$ et $b = e^{9{,}984} = 21676{,}847$

Donc l'ajustement est : $y = \frac{b}{x^a} = \frac{21676{,}847}{x^2}$

5. Le prix unitaire du produit avec cette approximation pour produire **15** unités.

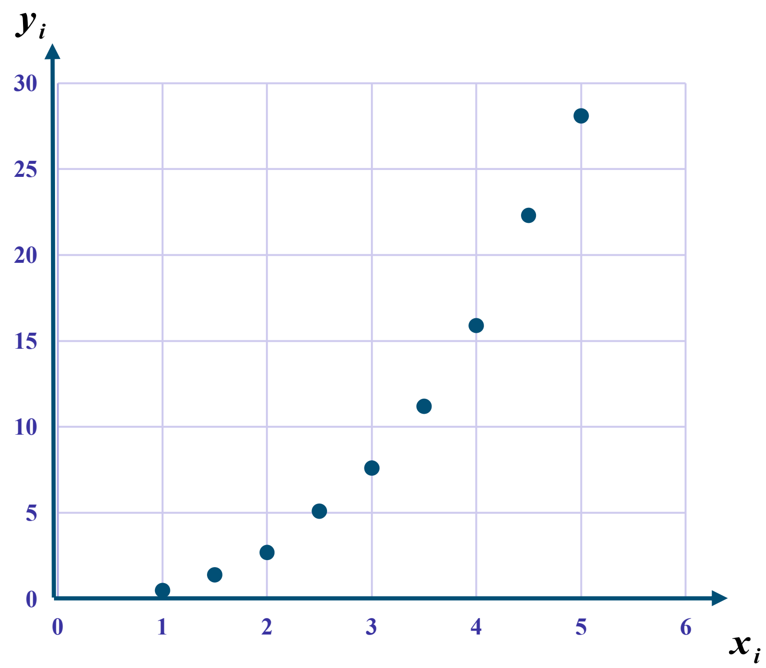
$y = \frac{b}{x^a} = \frac{21676{,}847}{x^2} = \frac{21676{,}847}{15^2} = 96{,}342$.

2eme modèle : L'ajustement par une fonction puissance

Les N points (x_i, y_i) ne sont pas alignés, mais plutôt proches d'une courbe représentant une fonction puissance de la forme :

$$y = b \cdot x^a \text{ avec } a > 0,\ \ b > 0$$

Dans ce cas le nuage aura l'allure suivante



Qui ressemble à la courbe de la fonction $f(x) = x^2$.

Comment peut-on estimer b et a ?

Nous raisonnons de la même façon que dans le cas hyperbolique et afin d'utiliser la méthode des **MCO**, moyennant une transformation, dans ce cas logarithmique, une forme linéaire :

On cherche a et b tels que :

$$y = b \cdot x^a \text{ avec } a > 0, \quad b > 0$$

En utilisant le logarithme népérien dans cette équation on trouve :

$$\Rightarrow \ln y = \ln bx^a = \ln b + a \ln x$$

Et si on suppose que : $\beta = \ln b$ et $\alpha = a$.

Le modèle linéaire est alors de la forme :

$$\ln y = \alpha \ln x + \beta$$

Donc en utilisant la méthode des **MCO**, on peut retrouver α et β :

$$\hat{\alpha} = \frac{Cov(\ln x, \ln y)}{V(\ln x)} = \frac{\sum_{i=1}^{i=N} (\ln x_i)(\ln y_i) - N \overline{\ln x}.\overline{\ln y}}{\sum_{i=1}^{i=N} (\ln x_i)^2 - N \overline{\ln x}^2}$$

$$\Rightarrow \hat{\beta} = \overline{\ln y} - \hat{\alpha} \overline{\ln x}$$

On peut maintenant retrouver la valeur de b et la valeur de a :

$\beta = \ln b \quad \Rightarrow \hat{b} = e^{\hat{\beta}}$ et $\alpha = a \Rightarrow \hat{a} = \hat{\alpha}$.

Par exemple, on obtient l'ajustement ci-dessous si on applique cette méthode aux données suivantes.

X	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
Y	0,1	0,5	1,4	2,7	5,1	7,6	11,2	15,9	22,3	28,1