

TD29 – Correction

Je m'échauffe avec les compétences de base!

Exercice 1:

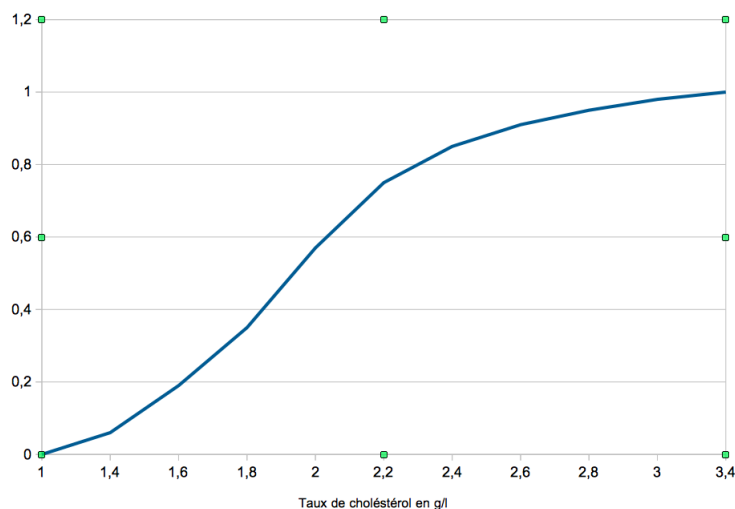
1. Pour chaque classe on note x_i le centre de chaque classe. On a alors

$$\bar{x} = \sum_{i=1}^{10} \frac{n_i}{100} x_i = 1,974.$$

2.

Taux	[1; 1,4[	[1,4; 1,6[	[1,6; 1,8[	[1,8; 2[	[2; 2,2[	[2,2; 2,4[	[2,4; 2,6[	[2,6; 2,8[	[2,8; 3[	[3; 3,4[
Eff	6	13	16	22	18	10	6	4	3	2
Fréq	0,06	0,19	0,35	0,57	0,75	0,85	0,91	0,95	0,98	1

polygone des fréquences cumulées croissantes



3.

4. Graphiquement, la médiane est l'antécédant de 0,5, soit environ 1,95.

Soient $C(x_c; 0,5)$, $A(1,8; 0,35)$ et $B(2; 0,57)$.

On cherche x_c tel que les points A , B , C soient alignés.

$$\begin{aligned} \overrightarrow{AC} \text{ et } \overrightarrow{AB} \text{ sont colinéaires} &\iff \begin{vmatrix} x_c - 1,8 & 0,2 \\ 0,15 & 0,22 \end{vmatrix} = 0 \\ &\iff 0,22(x_c - 1,8) - 0,2(0,15) = 0 \\ &\iff x_c = 1,93636 \end{aligned}$$

5. Graphiquement, on obtient $Q_1 \simeq 1,7$ et $Q_3 = 2,2$

6. On a une valeur exact pour Q_3 . Il reste à trouver une valeur exact pour Q_1 .

Soient $D(Q_1; 0,25)$, $A(1,8; 0,35)$ et $E(1,6; 0,19)$.

On cherche Q_1 tel que les points E , A , D soient alignés.

$$\begin{aligned} \overrightarrow{EQ_1} \text{ et } \overrightarrow{EA} \text{ sont colinéaires} &\iff \begin{vmatrix} Q_1 - 1,6 & 0,2 \\ 0,06 & 0,16 \end{vmatrix} = 0 \\ &\iff 0,16(Q_1 - 1,6) - 0,2(0,06) = 0 \\ &\iff Q_1 = 1,675 \end{aligned}$$

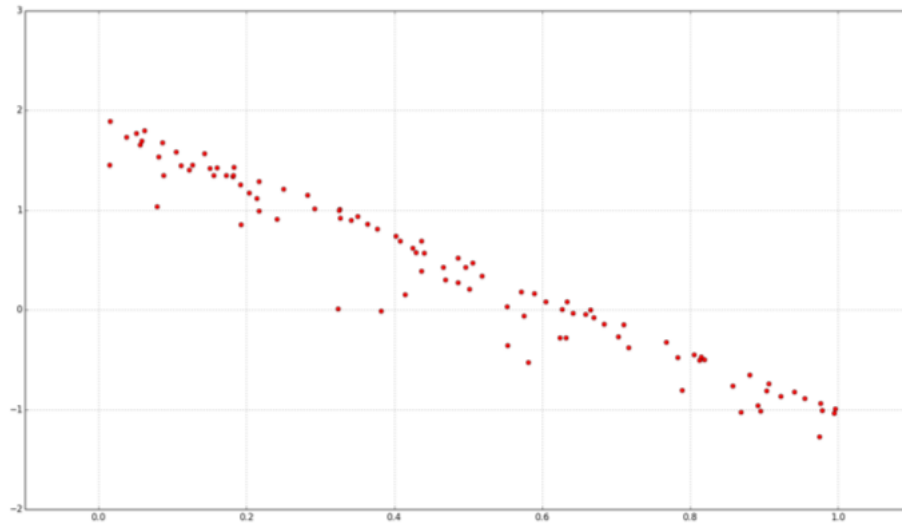
On obtient l'écart interquartile suivant :

$$Q_3 - Q_1 = 2,2 - 1,675 = 0,525.$$

Exercice 2: Figure 1

1. On cherche la droite d'équation $y = ax + b$ passant par la point moyen et de coefficient directeur $a = \frac{\text{cov}(x,y)}{\sigma_x^2}$.

On trouve $a = \frac{\text{cov}(x,y)}{\sigma_x^2} = \frac{-0,2544}{0,0850} \simeq -3$ et $b = \bar{y} - a\bar{x} = 0,3816 - a0,4757 \simeq 1,73$.



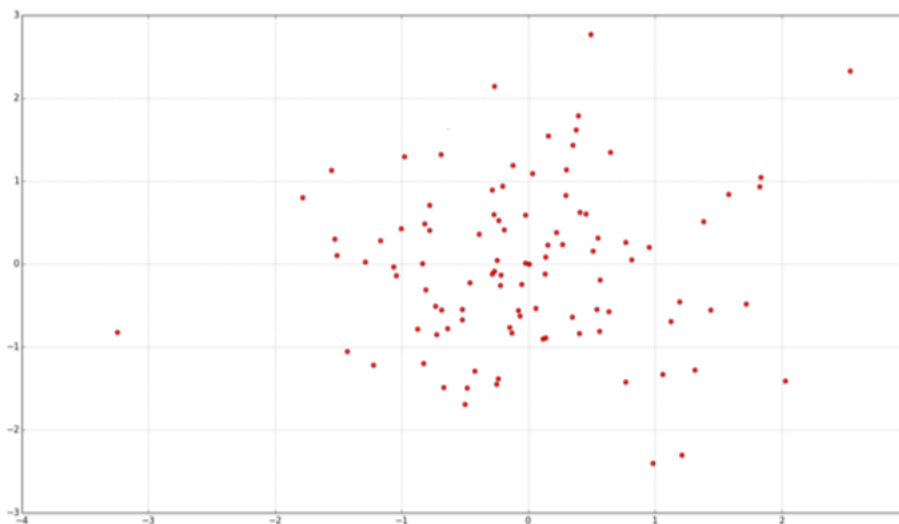
2.

3. Le coefficient de corrélation vaut

$$r_{x,y} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} = \frac{-0,2544}{\sqrt{0,0850}\sqrt{0,7727}} \simeq -0,9926$$

Figure 2

1. On cherche la droite d'équation $y = ax + b$ passant par la point moyen et de coefficient directeur $a = \frac{\text{cov}(x,y)}{\sigma_x^2}$. On trouve $a = \frac{\text{cov}(x,y)}{\sigma_x^2} = \frac{0,05568}{0,8292} \simeq 0,07$ et $b = \bar{y} - a\bar{x} \simeq -0,029$.



2.

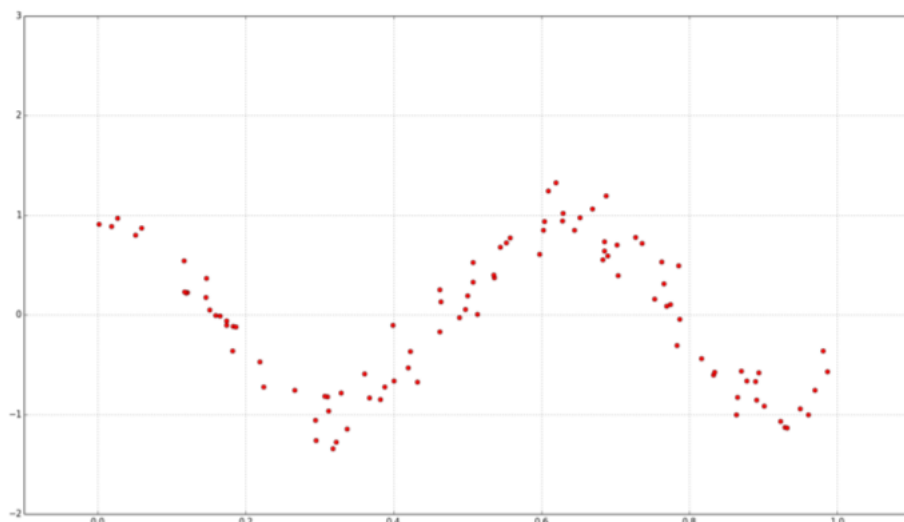
3. Le coefficient de corrélation vaut

$$r_{x,y} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} = \frac{0,05568}{\sqrt{0,8292}\sqrt{0,953}} \simeq 0,063.$$

Figure 3

1. On cherche la droite d'équation $y = ax + b$ passant par la point moyen et de coefficient directeur $a = \frac{\text{cov}(x,y)}{\sigma_x^2}$.

On trouve $a = \frac{\text{cov}(x,y)}{\sigma_x^2} = \frac{-0,0254}{0,07631} \simeq -0,33$ et $b = \bar{y} - a\bar{x} = -0,04186 - a0,5242 \simeq -0,21$.



2.

3. Le coefficient de corrélation vaut

$$r_{x,y} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} = \frac{-0,0254}{\sqrt{0,07631} \sqrt{0,5019}} \simeq -0,13$$

Exercice 3: On utilise la calculatrice. On obtient $\sigma_l = 2874$ et $\sigma_B = 2,99$.

On obtient alors d'après le théorème de Koenig Huygens :

$$\text{cov}(L, B) = \frac{1}{8} \sum_{i=1}^n L_i B_i - \overline{LB} = 8411,265.$$

Et on trouve le coefficient de corrélation

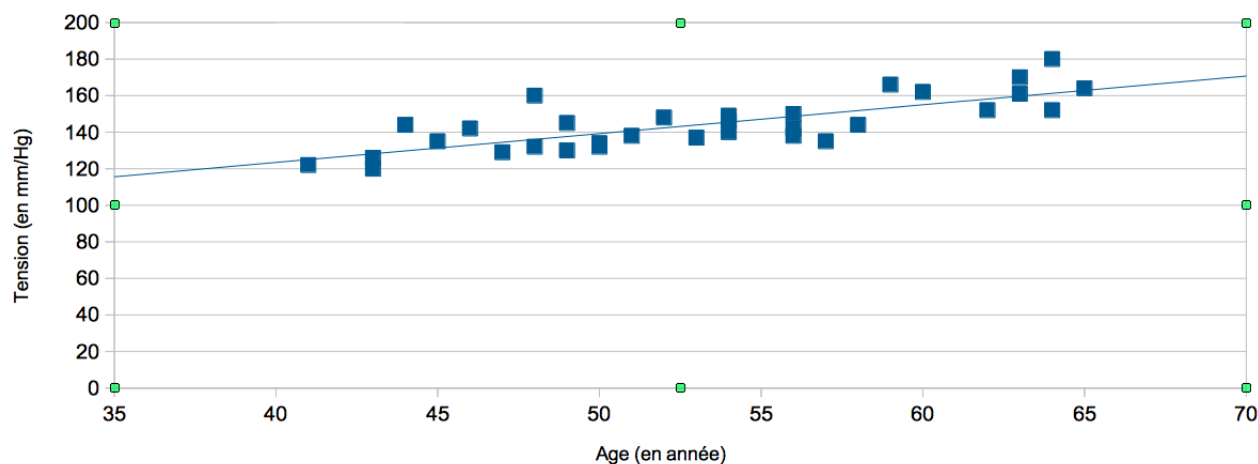
$$r = \frac{\text{cov}(L, B)}{\sigma_l \sigma_B} \simeq 0,98.$$

Les variables statistiques sont fortement corrélées. Attention la corrélation ne signifie pas qu'il y a une relation de cause à effet. Cela peut être une coïncidence.

Exercice 4:

1.

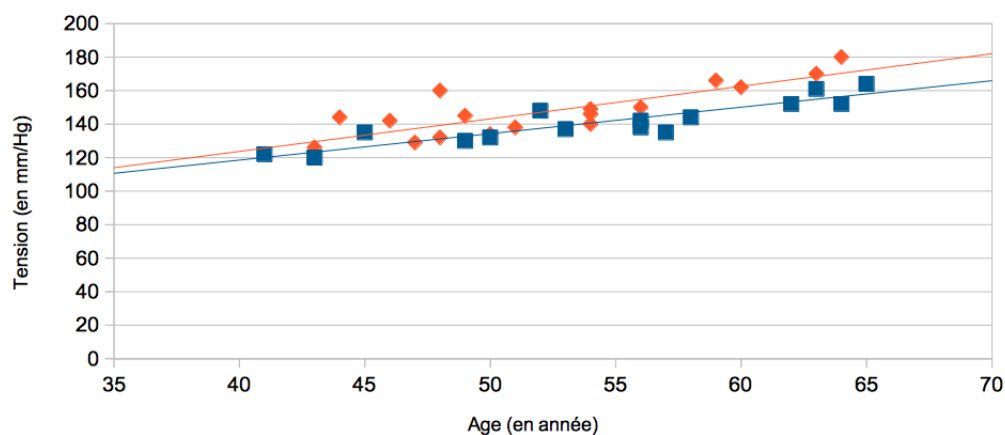
Régression linéaire



2. On cherche la droite d'équation $y = ax + b$ passant par la point moyen et de coefficient directeur $a = \frac{\text{cov}(x,y)}{\sigma_x^2}$. On trouve $a \simeq 1,57$ et $b \simeq 60,4$.

3. Le coefficient de corrélation vaut $r_{x,y} = 0,67$. Le coefficient de corrélation, nous informe que la régression linéaire n'est pas bonne.

Régression linéaire Non fumeur



4.

5. L'équation de la droite de régression linéaire pour les fumeurs est :

$$y = 1,95x + 45,728.$$

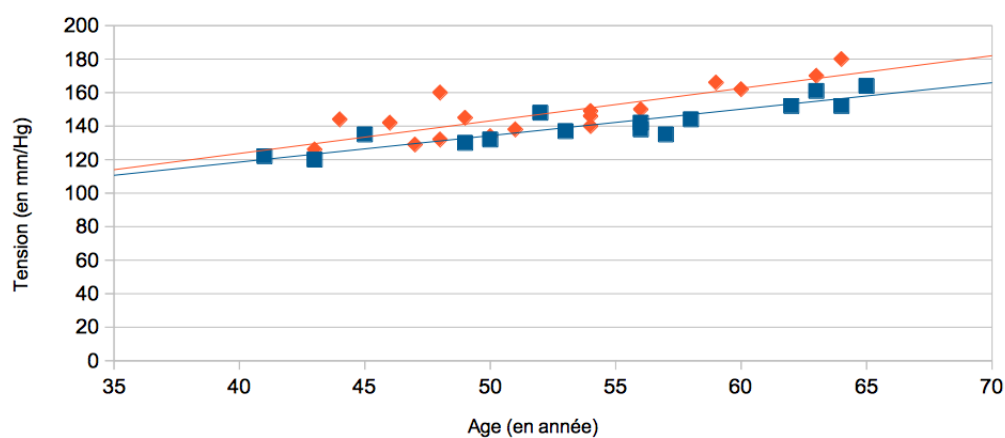
Le coefficient de corrélation vaut $r_{x,y} = 0,82$.

L'équation de la droite de régression linéaire pour les non fumeurs est :

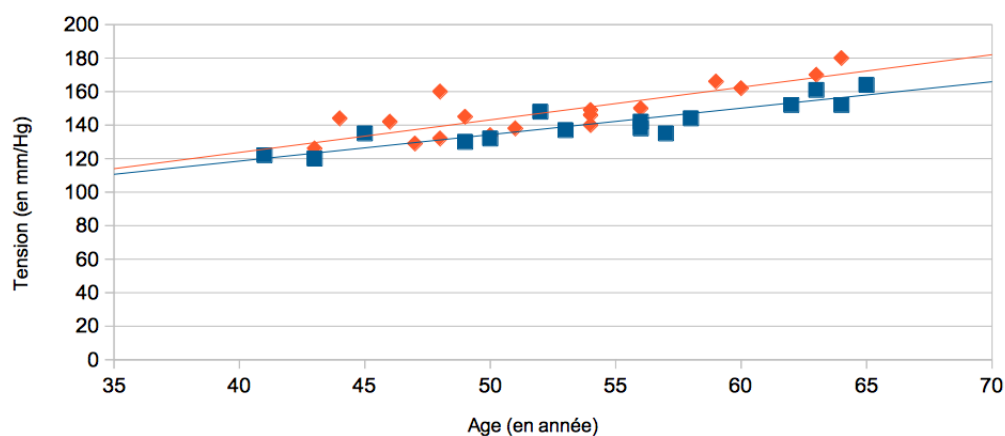
$$y = 1,58x + 55,329.$$

Le coefficient de corrélation vaut $r_{x,y} = 0,91$.

Régression linéaire Non fumeur



Régression linéaire Non fumeur

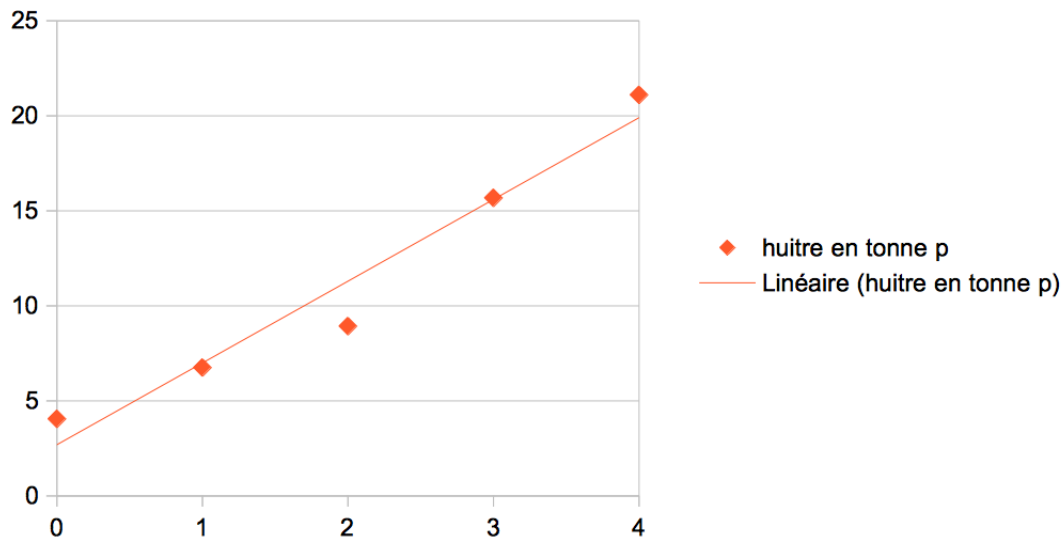


6. En séparant en deux groupe le nuage de points (fumeur/non fumeur), on obtient une meilleur approximation. Les coefficients de corrélation des 2 régressions linéaires sont plus proche de 1. Cela permet d'obtenir un meilleur suivi de la tension par rapport à l'âge des patient.

Je me perfectionne !

Exercice 5:

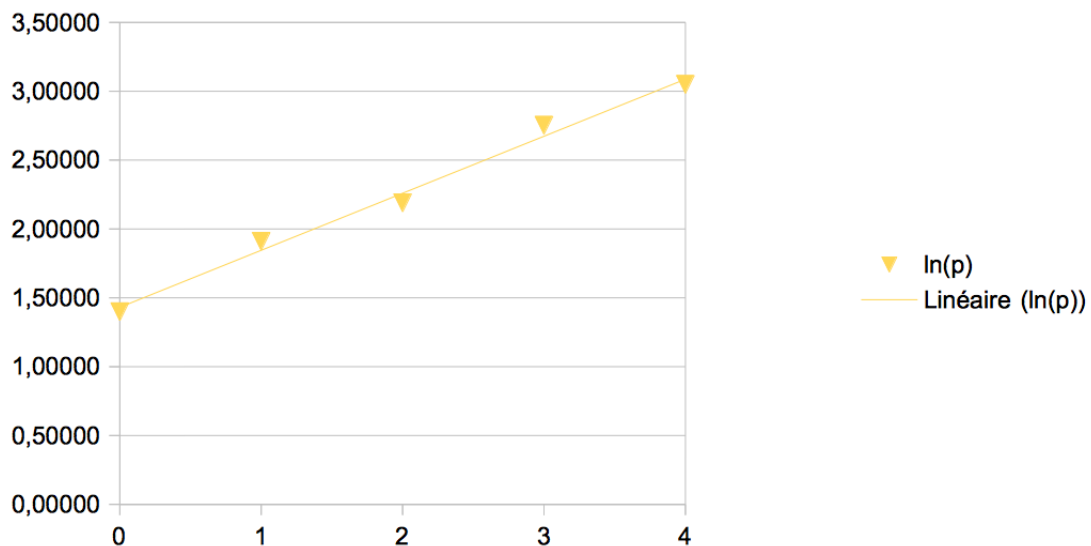
En représentant le nuage de points en posant en abscisse le temps et en ordonnée la quantité d'huitre en tonnes on obtient :



L'équation de la droite de régression linéaire est

$$y = 4,3x + 2,696.$$

Le coefficient de corrélation est $r_{t,p} = 0,976$. Le coefficient de corrélation est correct. Mais la droite ne semble pas approcher au mieux les points. On va chercher une meilleur approximation.



1.

2. Le coefficient de corrélation est $r_{t,z} = \frac{\text{cov}(t,z)}{\sigma_t \sigma_z} = 0,995$.
3. L'équation de la droite de régression linéaire est

$$y = 0,41x + 1,43.$$

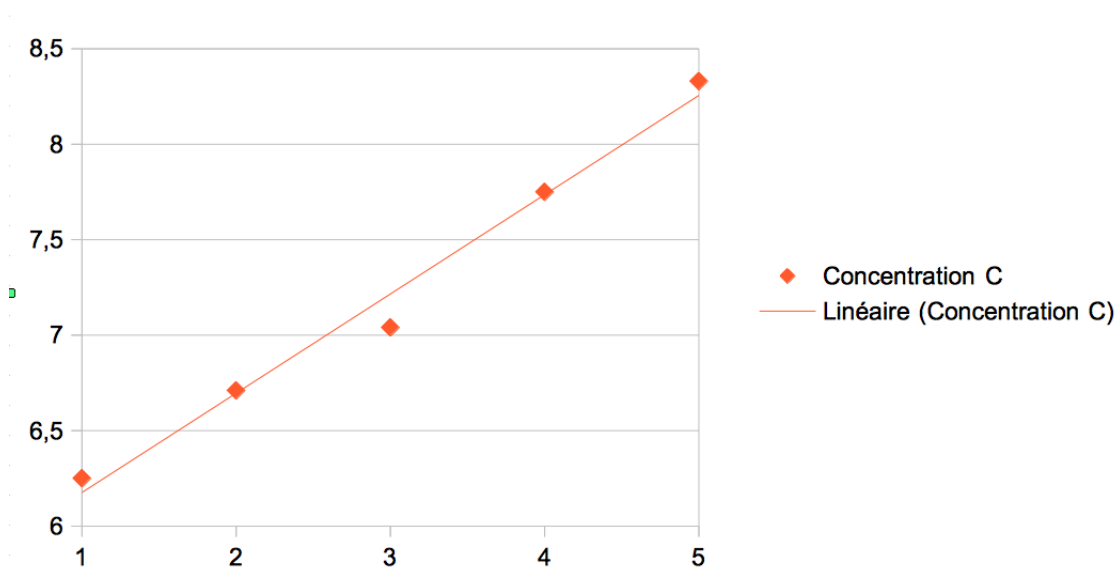
4. On peut donc conjecturer $\ln(p(t)) = 0,41t + 1,43$.

D'où $p(t) = e^{0,41t+1,43}$

5. L'année 2015 correspond à $t=6$. $p(6) = e^{0,41 \times 6 + 1,43} = 46,91$. Selon ce modèle on peut prévoir une récolte de 46,91 tonnes.

Exercice 6:

En représentant le nuage de points en posant en abscisse le temps et en ordonnée la concentration C , on obtient :



L'équation de la droite de régression linéaire est

$$y = 0,52x + 5,656.$$

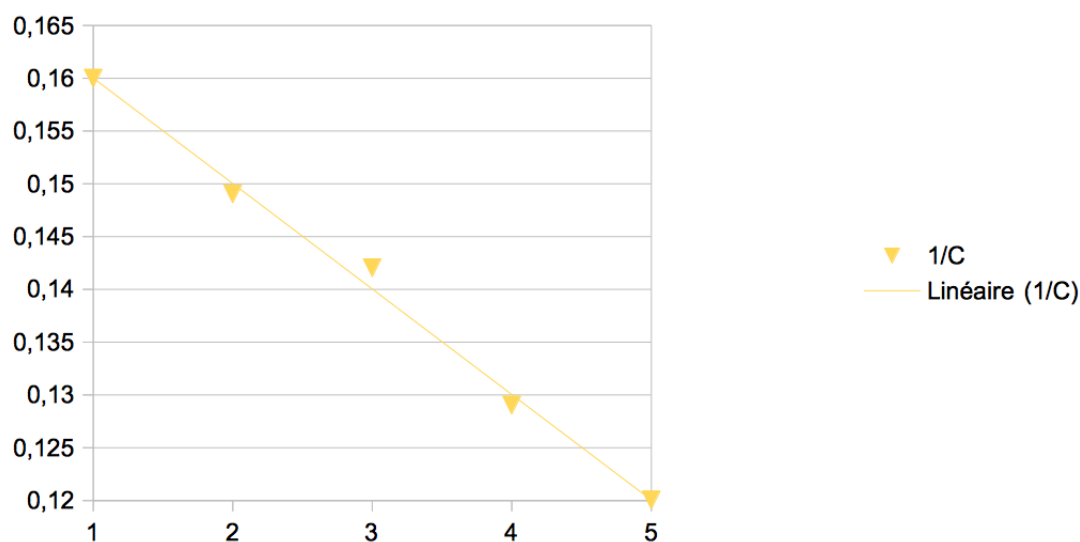
Le coefficient de corrélation est $r_{x,y} = 0,9845$.

Suivons maintenant l'indication.

$$C = \frac{1}{at + b} \iff \frac{1}{C} = at + b.$$

On cherche la droite de régression linéaire associée à $\frac{1}{C}$.

temps t en heure	1	2	3	4	5
Concentration C en g/l	6,25	6,71	7,04	7,75	8,33
$\frac{1}{C}$	0,16	0,149	0,142	0,129	0,12



On obtient comme droite de régression linéaire

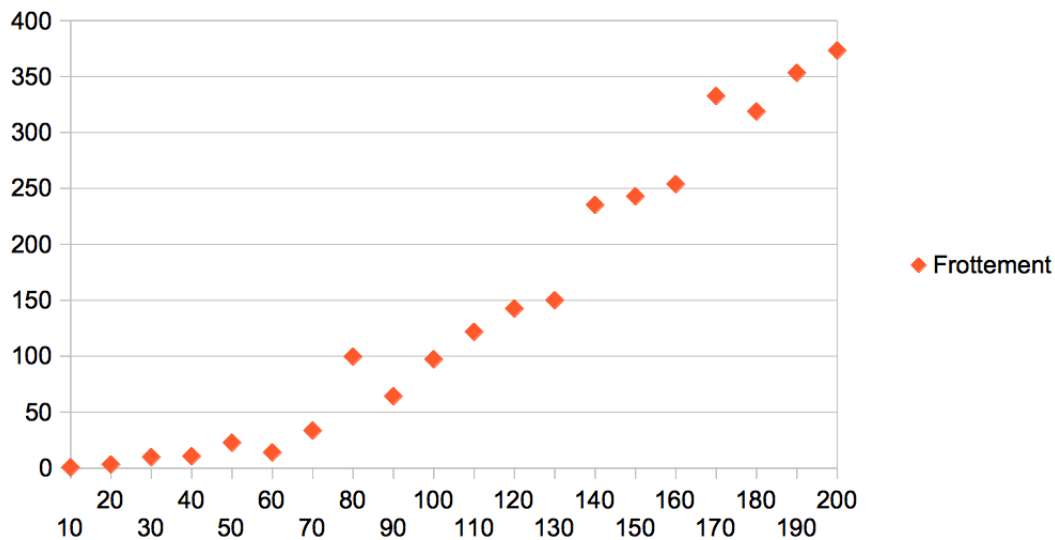
$$y = -0,01x + 0,1704$$

avec un coefficient de régression linéaire $r_{x,y} = -0,998$. Cette approximation est meilleur que la première.

En revenant à C , on obtient :

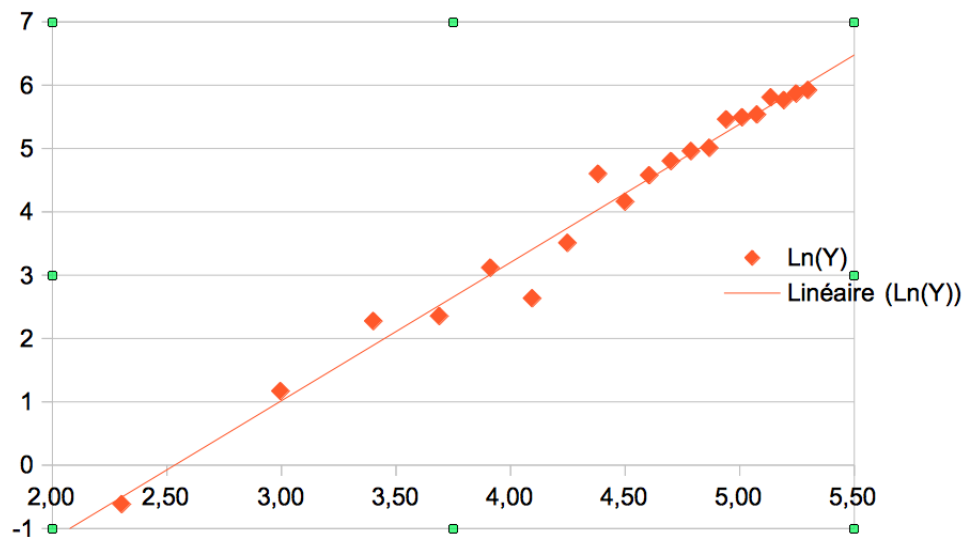
$$C(t) = \frac{1}{-0,01t + 0,1704}.$$

Exercice 7:



On soupçonne que Y et X sont reliés par une relation du type $Y = CX^k$ avec $C > 0$ et $k \in \mathbb{N}$.

- Si $Y = CX^k$ alors $\ln(Y) = k \ln(X) + \ln(C)$. On cherche la droite de regression linéaire associée au nuage de points $(\ln(y_i), \ln(x_i))$. On obtient



La droite de régression linéaire a pour équation :

$$y = 2,18x - 5,53.$$

On trouve donc $k = 2,18$ et $C = \exp(-5,53) = 0,004$.

- Le coefficient de corrélation vaut

$$r_{\ln(X), \ln(Y)} = 0,99.$$

Il est très proche de 1. Les résultats obtenus ne contredisent pas la conjecture.