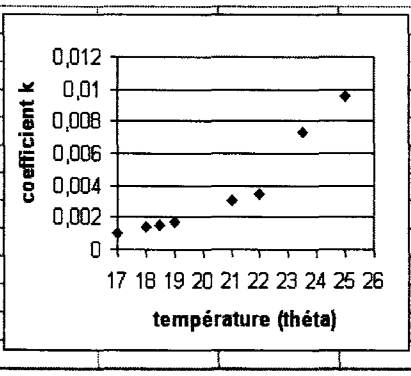
Corrigé

NOM : Date :
PRENOM : Groupe :

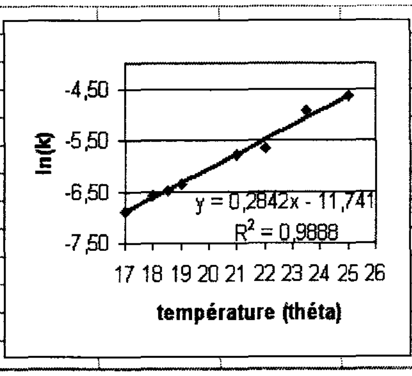
Mathématiques pour la Biologie : Feuille-réponses du TD 9
Régression linéaire

Exercice 1. : On cherche à modéliser la croissance de la taille d'*Elysia subornata*, petite limace de mer de $2,5cm$ de long - ordre des Ascoglosses opistobranches. On pense que sa taille à l'instant t (comptée en jours) peut s'écrire $y(t) = 55(1 - \exp(-kt))$ où 55 est la longueur maximale théorique de l'animal et k un coefficient de croissance. On observe que ce coefficient de croissance k dépend de la température de l'eau, notée θ pour *théta*. Expérimentalement on a pu obtenir les valeurs de k et θ rapportées dans le tableau de gauche. Sur la première figure on a représenté le nuage de point correspondant et sur la figure de droite le nuage $(\theta, \ln(k))$ pour lequel on a effectué une régression linéaire.

theta	k
17	0,001
18	0,0014
18,5	0,0016
19	0,0017
21	0,0031
22	0,0035
23,5	0,0073
25	0,0096



theta	ln(k)
17	-6,91
18	-6,57
18,5	-6,47
19	-6,37
21	-5,78
22	-5,65
23,5	-4,92
25	-4,64



1. Pourquoi n'a-t-on pas effectué directement une régression linéaire de k sur θ ?

Les points sont plutôt alignés sur une courbe exponentielle plutôt qu'une droite. La régression sur ce nuage de points ne serait pas bonne.

2. Expliquer comment a été calculée la pente 0.2842 de la droite de régression linéaire.

La pente de la droite est donnée par la formule $\hat{a} = \dfrac{cov(\theta, k)}{var(\theta)}$ avec $cov(\theta, k) = \frac{1}{8}\sum \theta k - \mu(\theta)\mu(k)$
et $var(\theta) = \frac{1}{8}\sum \theta^2 - [\mu(\theta)]^2$.

3. Que représente R^2 et que peut-on déduire de sa valeur ?

R^2 est le carré du coefficient de corrélation linéaire ρ. Il est toujours compris entre 0 et 1 et lorsqu'il est proche de 1, c'est le signe que les points du nuage sont en principe bien alignés sur une droite.

4. De la relation linéaire obtenue entre $\ln(k)$ et θ, on peut déduire que k est une fonction exponentielle de θ de la forme $k = Ae^{B\theta}$. Calculer les constantes A et B.

$k = Ae^{B\theta} \Rightarrow \ln k = \ln A + \ln(e^{B\theta})$
$\Rightarrow \underbrace{\ln k}_{y} = \underbrace{\ln A}_{-11,741} + \underbrace{B}_{0,2842}\underbrace{\theta}_{x}$

On en déduit que $\boxed{A = \exp(-11,741) \simeq 7,96.10^{-6}}$ et $\boxed{B = 0,2842}$

5. Calculer la valeur de k prédite par ce modèle pour une température $\theta = 24°C$. En déduire la taille atteinte à cette température par Elysia subornata après 20 jours.

On a $\ln k = -11,741 + 0,2842(24) \simeq -4,9202$
d'où $k \simeq 7,297.10^{-3}$ Donc $y(t) = 55(1 - \exp(-20k)) \simeq 7,47$
A 24°C, l'animal a atteint une longueur d'environs 7,47cm après 20 jours.

Exercice 2. :

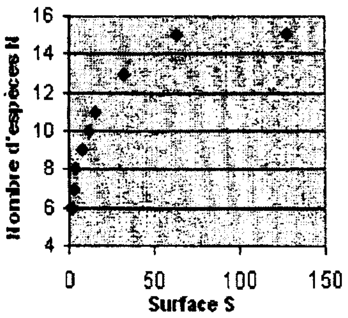
L'une des rares lois que l'on a pu mettre en évidence en Ecologie est la relation existant entre le nombre N d'espèces présentes dans un habitat donné (bien délimité) et la surface S de cet habitat. On considère généralement que cette relation est de la forme

$$N = AS^B \qquad (1)$$

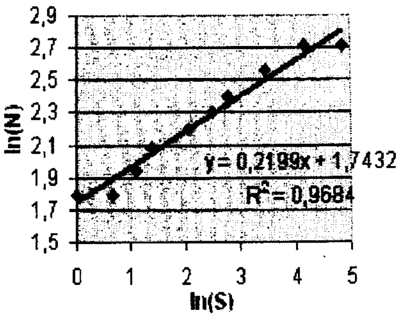
où A et B sont deux constantes. Afin de vérifier cette relation pour les plantes présentes dans une prairie (pissenlit, paquerettes, orties, boutons d'or, ...), on a effectué les mesures indiquées dans le premier tableau ci-dessous. On a représenté sur la première figure ci-dessous les valeurs de N en fonction de celles de S et sur la deuxième les valeurs de $\tilde{N} = \ln(N)$ en fonction de celles de $\tilde{S} = \ln(S)$. On voit que la regression linéaire de $\tilde{N}$ sur $\tilde{S}$ a donné :

$$\tilde{N} = 0,2199\tilde{S} + 1,7432 \text{ avec } R^2 = 0,9684 \qquad (2)$$

S	N
1	6
2	6
3	7
4	8
8	9
12	10
16	11
32	13
64	15
128	15



ln(S)	ln(N)
0,00	1,79
0,69	1,79
1,10	1,95
1,39	2,08
2,08	2,20
2,48	2,30
2,77	2,40
3,47	2,56
4,16	2,71
4,85	2,71



1. Pourquoi n'a-t-on pas effectué directement une régression linéaire de N sur S ? Expliquez l'intéret de cette transformation des données.

La fonction $f(S) = AS^B$ n'est pas de la forme $aS + b$; en revanche $\tilde{N} = \ln(N) = \ln(AS^B) = \ln(A) + B\ln(S) = b + a\tilde{S}$ est bien une fonction linéaire-affine de $\tilde{S} = \ln(S)$, avec $b = \ln(A)$ et $a = B$

2. Que représente R^2 et que peut-on déduire de sa valeur ?

R^2 est le carré de la corrélation linéaire entre $\tilde{N}$ et $\tilde{S}$. On a $R^2 = \frac{DR}{DT}$ où DT est la dispersion totale de $\tilde{N}$ et DR est la dispersion expliquée par la regression. On voit qu'avec $R^2 = 0.9684.. \approx 97\%$ la dispersion totale est très bien expliquée par la regression linéaire.

3. A partir de la régression linéaire (2), calculer les constantes A et B de la relation (1).

On lit sur l'équation de la droite de regression imprimée sur le graphe que $a = 0.2199$ et $b = 1.7432$. Or $B = a = 0.2199$ et $A = e^b$ (puisque $b = \ln(A)$) et donc $A = \exp(1.7432) = 5.7156..$ et $B = 0.2199$

4. Quelle valeur $\tilde{N}$ ce modèle linéaire prédit-il pour $\tilde{S} = ln(128)$? En comparant avec la valeur de $\tilde{S}$ observée, calculer le résidu ε en ce point.

La valeur prédite $\tilde{N}$ est $a\tilde{S} + b$, et donc $\tilde{N} = 0.2199 \ln(128) + 1.7432 = 2.810..$
Le résidu ε en ce point vaut $\varepsilon = 2.71.. - 2.81.. = -0.10..$

5. Quelle valeur $\tilde{N}$ ce modèle linéaire prédit-il pour $\tilde{S} = ln(100)$? En déduire le nombre d'espèces pouvant coexister dans un habitat de surface $S = 100$, selon ce modèle.

Pour $\tilde{S} = \ln(100)$ le modèle prédit $\tilde{N} = a\tilde{S} + b = 0.2199 \cdot 4.60.. + 1.74..$ donc $\tilde{N} = 2.75..$ et donc $N = \exp(\tilde{N}) = e^{2.75..} = 15.73..$ c'est-à-dire 15 à 16 espèces différentes sur un terrain de surface 100.