NOM : Date : 5 au 9 décembre 2011

PRENOM : Groupe : .

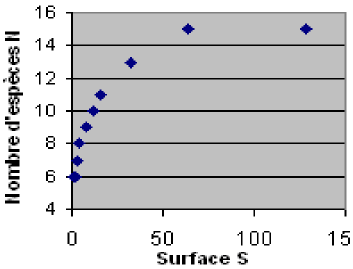
Mathématiques pour la Biologie : Feuille-réponses du TD 10
Régression linéaire : compléments

Exercice 1. : 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

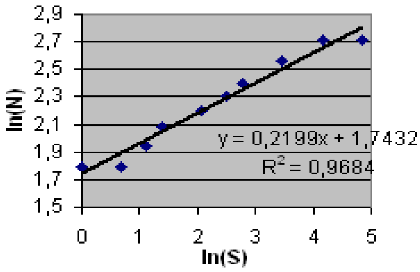
$$N = AS^B \tag{1}$$

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 régression linéaire de $\tilde{N}$ sur $\tilde{S}$ a donné l'équation $\tilde{N} = 0,2199\tilde{S} + 1,7432$, avec $R^2 = 0,9684$.

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 et a-t-on préféré transformer N en $\tilde{N}$ et S en $\tilde{S}$?

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

3. Quelle valeur $\tilde{N}$ ce modèle linéaire prédit-il pour $\tilde{S} = ln(64)$? En comparant avec la valeur de $\tilde{S}$ observée, calculer le résidu ε en ce point.

4. 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.

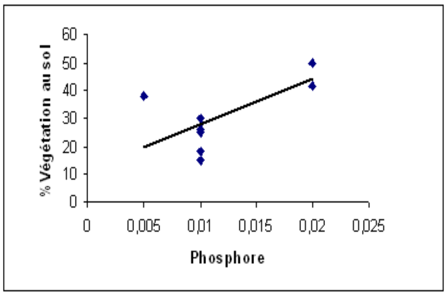
Exercice 2. : Des écologistes se sont intéressés à la répartition de la végétation dans un site aride du sud de la France, la plaine du Crau (13). Ils ont effectué 9 prélèvements de sol (S1, ..., S9) pour lesquel ils ont retenu, après analyse, 6 mesures (pH, C/N, Ca, Mg, K, P) et dans le même temps ils ont évalué dans chaque cas le pourcentage de recouvrement au sol par la végétation (%V). Ces données brutes sont regroupées dans le tableau suivant :

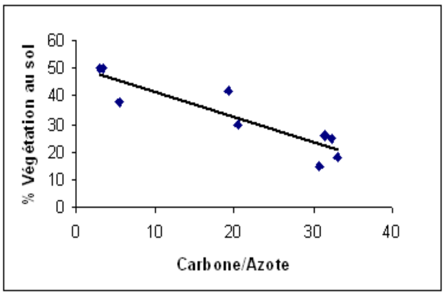
	pH	C/N	Ca	Mg	K	P	$\%V$
$S1$	$5,5$	$30,75$	$0,55$	$0,01$	$0,42$	$0,01$	15
$S2$	7	$19,29$	$1,02$	$0,07$	$0,43$	$0,02$	42
$S3$	$6,8$	$31,47$	$1,02$	$0,05$	$0,45$	$0,01$	26
$S4$	$7,3$	$2,93$	$1,82$	$0,09$	$0,44$	$0,02$	50
$S5$	$5,62$	$32,45$	$0,42$	0	$0,41$	$0,01$	25
$S6$	$6,6$	$20,05$	$0,75$	$0,01$	$0,44$	$0,01$	30
$S7$	7	$3,35$	$1,33$	$0,07$	$0,46$	$0,02$	50
$S8$	$5,8$	$33,18$	$0,48$	$0,02$	$0,43$	$0,01$	18
$S9$	7	$5,32$	$1,35$	$0,07$	$0,48$	$0,005$	38

On veut étudier de quelle façon les variables pH, C/N et P influent sur le pourcentage de recouvrement au sol par la végétation et on va pour cela chercher à expliquer par une régression linéaire la quantité %V en fonction de chacune de ces 3 variables explicatives.

1. Exprimer, par une régression linéaire, %V en fonction de pH et donner l'équation de la droite des moindres carrés ainsi que la valeur du coefficient de détermination R^2 (on pourra utiliser le fait que la variance de pH vaut $0,417$, que la variance de %V vaut $150,44$ et que leur covariance vaut $6,92$).

2. On a aussi cherché à expliquer de la même façon %V à l'aide de deux autres variables, C/N et P. On a obtenu les dessins des nuages et des droites de régression suivants :





Mais on a mélangé les résultats ; les équations $y = 1617,4x + 12$ et $y = -0,904x + 50,691$ et les coefficients $R^2 = 0,8178$ et $R^2 = 0,4937$. Sans faire de nouveaux calculs, indiquer quelle équation et quel coefficient correspond à quel dessin en justifiant vos réponses.

3. Parmi ces trois régressions, y-en-a-t-il à votre avis qui soient acceptables ? Justifier votre réponse.

4. Finalement, compte tenu de ces résultats, peut-on conclure que la chimie du sol peut expliquer le pourcentage de végétation ? On pourra se servir de la part de dispersion de $\%V$ expliquée par les régressions.

Exercice 3. : Des biologistes ont introduit au printemps 1937 sur l'ile de la Protection (Etat de Washington, USA), 8 individus d'une population de faisans qu'ils ont ensuite recencé chaque printemps jusqu'à ce qu'en 1943 l'armée débarque sur l'ile, décimant cette population de faisans. Leurs observations ont été les suivantes :

Année	1937	1938	1939	1940	1941	1942
Effectifs	8	30	81	282	705	1325

On désigne par x le nombre d'années écoulées depuis 1937, par $N(x)$ la taille de la population de faisans à la date x et par $y(x) = \ln N(x)$ le logarithme de $N(x)$.

1. Calculer la droite de régression linéaire de y par rapport à x (on pourra utiliser que $Var(x) \simeq 2,92$, $Var(y) \simeq 3,17$ et $Cox(x, y) \simeq 3,02$).

2. Calculer la valeur prédite $\hat{y}(6)$ par ce modèle pour $x = 6$ et en déduire l'effectif prédit pour l'année 1943.

3. Cette régression linéaire est-elle valide selon vous ? Expliquer.

4. A quelle expression pour $N(x)$ cette régression de $y(x)$ conduit-elle ? Comment s'appelle ce type de modèle ? Quel est son principal défaut ?

5. On opte finalement pour un modèle logistique pour $N(x)$ avec pour constante $r = 1,3$ et $K = 2000$. On a donc
$$\frac{dN(x)}{dx} = (1,3)N(x)\left(1 - \frac{N(x)}{2000}\right).$$
On rappelle que, si $N(0)$ est l'effectif en $x = 0$, la solution exacte de cette équation est donnée par la formule $N(x) = \frac{N(0)Ke^{rx}}{K+N(0)(e^{rx}-1)}$.
Calculer $N(1)$ selon ce modèle, puis calculer une valeur approchée de $N(1)$ par la méthode d'Euler en utilisant un pas $h = 1$, arrondies à l'entier le plus proche. Comparer avec l'effectif observé après un an et commenter.

6. On a préféré finalement calculer les valeurs de la solution exacte du modèle (toujours pour $N(0) = 8$) et celles de la solution approchée, mais en prenant cette fois un pas $h = 0,1$. Le tableau suivant indique les premières valeurs, exactes et approchées, arrondies à l'entier le plus proche. Compléter les trois valeurs manquantes en expliquant quels calculs vous faites pour cela. A noter que ce qui est étudié ici est un modèle mathématique : il n'est pas supposé, bien entendu que les faisans font, brusquement, 10 pontes reproductives par an !

x	0	$0,1$	$0,2$	$0,3$	$0,4$	$0,5$	$0,6$	$0,7$	$0,8$	$0,9$	1
solution exacte	8	9	10	12	13	15	17	20	....	26	29
solution approchée	8	....	10	12	13	....	17	19	21	24	27