

NOM :  
PRENOM :

*Compte*

Date :  
Groupe :

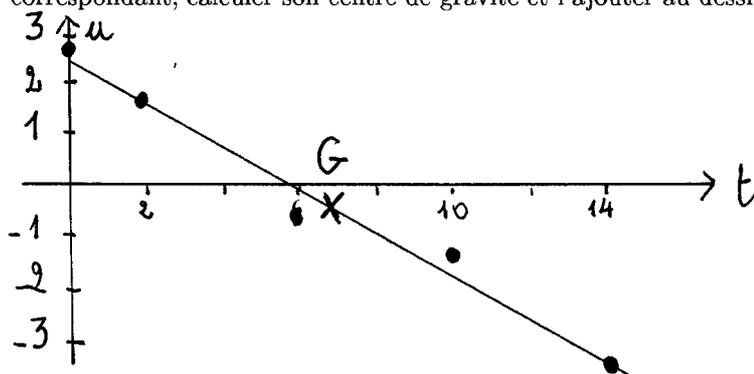
**Mathématiques pour la Biologie : Feuille-réponses du TD 10**  
**Régression linéaire : suite**

**Exercice 1.** : On étudie la croissance d'une plante à partir d'un instant considéré comme instant initial. On effectue des mesures du diamètre de la tige principale et on obtient les résultats ci-dessous.

1. Pour tout  $t \geq 0$ , on pose  $u(t) = \ln\left(\frac{8}{d(t)} - 1\right)$ . Compléter les valeurs manquantes de  $u_i$ .

|   |      |      |       |       |       |
|---|------|------|-------|-------|-------|
| Temps $t_i$ , en semaines                 | 0    | 2    | 6     | 10    | 14    |
| Diamètre $d_i$ , en centimètres           | 0,4  | 1,2  | 5,4   | 6,4   | 7,8   |
| $u_i = \ln\left(\frac{8}{d_i} - 1\right)$ | 2,94 | 1,73 | -0,73 | -1,39 | -3,66 |

2. On considère les données  $(t_i, u_i)$  comme une série de 5 points. Représenter le nuage de points correspondant, calculer son centre de gravité et l'ajouter au dessin.



Centre de gravité  $G$  :

$$\frac{0+2+6+10+14}{5} = 6,4$$

$$\frac{2,94+1,73-0,73-1,39-3,66}{5} \approx -0,22$$

$G : (6,4 ; -0,22)$

3. Compléter le tableau suivant et en déduire les valeurs des variances et covariance demandées :

| $i$    | $t_i$ | $u_i$ | $(t_i)^2$ | $(u_i)^2$ | $(t_i)(u_i)$ |
|--------|-------|-------|-----------|-----------|--------------|
| 1      | 0     | 2,94  | 0         | 8,67      | 0,00         |
| 2      | 2     | 1,73  | 4         | 3,01      | 3,47         |
| 3      | 6     | -0,73 | 36        | 0,53      | -4,39        |
| 4      | 10    | -1,39 | 100       | 1,93      | -13,86       |
| 5      | 14    | -3,66 | 196       | 13,42     | -51,29       |
| total: | 32    | -1,10 | 336       | 27,56     | -66,07       |
| $\mu$  | 6,4   | -0,22 | 67,2      | 5,52      | -13,214      |

$Var(t) = 67,2 - 6,4^2$   
 $\approx 26,24$

$Var(u) = 5,52 - (-0,22)^2$   
 $\approx 5,46$

$Cov(t,u) = -13,214 - (6,4)(-0,22)$   
 $\approx -11,806$

$Var(t) = 26,24$     $Var(u) = 5,46$     $cov(t,u) = -11,806$

4. Déterminer, par la méthode des moindres carrés ordinaires une équation de la droite de régression de  $u$  en  $t$ .  $u = -0,4498t + 2,6586$  =  $at + b$

5. Tracer cette droite sur le graphique de la deuxième question.

6. Calculer le coefficient de corrélation linéaire; commenter.  $\rho(t,u) = -0,986$

7. En déduire que, pour cette plante, le diamètre de sa tige principale est donné par une relation de la forme  $d(t) = \frac{8}{1+Ce^{-at}}$  dans laquelle on précisera  $C$  et  $a$ .

Comme  $u(t) = \ln\left(\frac{8}{d(t)} - 1\right)$ ,  $\exp(u(t)) = \frac{8}{d(t)} - 1$ , d'où

$d(t) = \frac{8}{1 + e^{u(t)}}$  - Mais comme  $u(t) = -0,4498t + 2,6586$ , on en déduit

que  $e^{u(t)} = e^{2,6586} e^{-0,4498t}$  - D'où  $a = 0,4498$  et  $C = 14,276$

$= 14,276$

**Exercice 2.** : *Posidonia oceanica* est une plante à fleurs poussant sur les fonds de faibles profondeurs (inférieures à 40 mètres) en mer méditerranée. En 1989, des individus de posidonies provenant de différentes prairies méditerranéennes ont été transplantés dans une baie du parc national de Port-Cros (Var) pour étudier la croissance des transplants. Un suivi a été mis en place : régulièrement, le nombre de plantes vivantes a été décompté par des plongeurs et on a obtenu les données suivantes :

| Année (t)               | 1989 | 1990 | 1991 | 1992 | 1993 | 1995 | 1997 |
|-------------------------|------|------|------|------|------|------|------|
| Nb de plantes ( $N_t$ ) | 12   | 17   | 27   | 36   | 54   | 71   | 129  |

1. Le tracé du nuage de points correspondant ( $t, N_t$ ) suggère une relation linéaire de  $y(t) = \ln \frac{N_t}{N_{1989}}$  sur  $t$ . Expliquer pourquoi.

On peut penser que, partant de l'effectif  $N_{1989}$  en 1989, le nombre de plantes  $N_t$  présente une croissance exponentielle de la forme  $N_t = N_{1989} e^{at}$ , avec  $y(t) = at + b$ . D'où une relation linéaire de  $\ln \frac{N_t}{N_{1989}}$  par rapport à  $t$ .

2. Dans le tableau ci-dessous, on a procédé à la régression linéaire de  $y(t) = \ln(N_t/N_{1989})$  sur  $t$  selon la méthode MCO. Compléter les 8 cellules vides puis justifier vos réponses.

| Année (t) | $y = \ln(N_t / N_0)$ | $t^2$   | $y^2$     | $t \cdot y$ |
|-----------|----------------------|---------|-----------|-------------|
| 1989      | 0                    | 3956121 | 0,00      | 0,00        |
| 1990      | 0,35                 | 3960100 | 0,12      | 693,13      |
| 1991      | 0,81                 | 3964081 | 0,66      | 1614,56     |
| 1992      | 1,10                 | 3968064 | 1,21      | 2188,44     |
| 1993      | 1,50                 | 3972049 | 2,26      | 2997,63     |
| 1995      | 1,78                 | 3980025 | 3,16      | 3546,66     |
| 1997      | 2,37                 | 3988009 | 5,64      | 4742,69     |
| Somme     | 13947                | 8       | 27788449  | 13,05       |
| Moyenne   | 1992,43              | 1,13    | 3969778,4 | 1,86        |

| Var(t) | Var(y) | Cov(t,y) | a chapeau | b chapeau | $R^2$  |
|--------|--------|----------|-----------|-----------|--------|
| 6,82   | 0,59   | 1,97     | 0,2895    | -575,64   | 0,9752 |

$$\rho = \frac{\text{Cov}(t, y)}{\sqrt{\text{Var}(t) \text{Var}(y)}} = \frac{1,97}{\sqrt{(6,82)(0,59)}} \approx 0,989 \quad R^2 = \rho^2 \approx 0,9752$$

$$\hat{a} = \frac{\text{Cov}(t, y)}{\text{Var}(t)} = \frac{1,97}{6,82} \approx 0,2895$$

$$\hat{b} = \mu(y) - \hat{a} \mu(t) = 1,13 - (0,2895) 1992,43 \approx -575,64$$

3. Quelle est l'équation de la droite de régression de  $y(t)$  sur  $t$ ?

$$y(t) = \hat{a} t + \hat{b} \quad \text{donc, ici, } \boxed{y(t) = 0,2895 t - 575,64}$$

4. Selon vous, le modèle linéaire obtenu est-il justifié?

Le  $R^2$  indique que plus de 97% de la variabilité du nuage est expliquée par la régression. A priori, le modèle linéaire semble donc adapté mais il faudrait également étudier les résidus pour confirmation.

**Exercice 3. : Calibration d'un modèle par MCO** Lorsqu'on choisit un modèle dynamique, par exemple un modèle logistique, pour étudier la dynamique d'une population donnée, l'une des premières difficultés est de déterminer les constantes (dans le cas du modèle logistique il y en a deux  $r$  et  $K$ ) qui correspondent le mieux à la population particulière que l'on étudie. Choisir ces constantes s'appelle la *calibration* (en anglais *fitting*) du modèle.

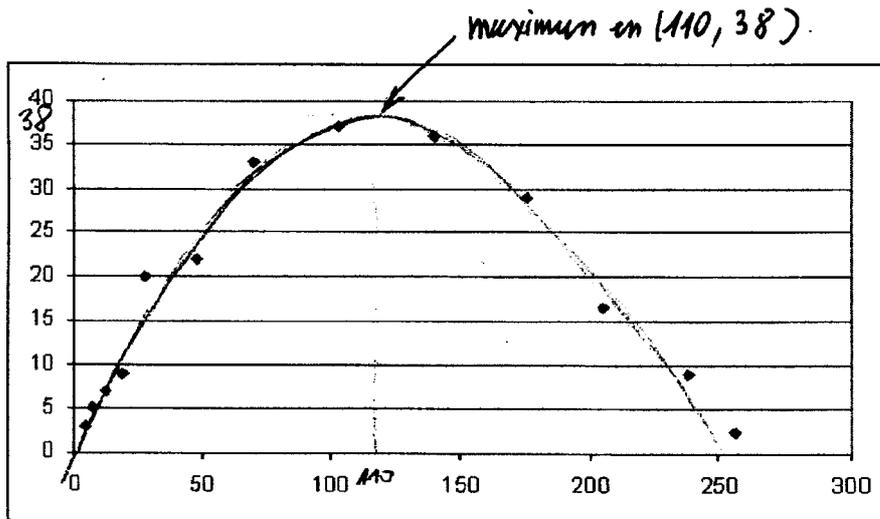
Une première approche peut être faite en comparant le *graphe théorique* de  $f$  et un *graphe empirique* que l'on peut construire à partir des observations recueillies sur la population que l'on étudie. Dans l'exemple d'un modèle logistique, le graphe théorique est celui de la parabole  $f(N) = rN(1 - \frac{N}{K})$ , de sommet  $(\frac{K}{2}, \frac{rK}{4})$  et le graphe empirique est obtenu en traçant en abscisse la suite des effectifs  $N_1, N_2, \dots, N_n$  mesurés en des instants successifs  $t_1, t_2, \dots, t_n$  et en ordonnée il conviendrait de tracer la dérivée  $N'_i$  par rapport au temps pour chaque valeur  $N_i$  de l'effectif (puisque l'on a la relation  $N' = f(N)$ ). En pratique, on remplace cette dérivée qui ne peut être mesurée par les taux de variations  $Z$  de ces effectifs par unité de temps  $Z_1 = \frac{N_2 - N_1}{t_2 - t_1}, Z_2 = \frac{N_3 - N_2}{t_3 - t_2}, \dots, Z_{n-1} = \frac{N_n - N_{n-1}}{t_n - t_{n-1}}$ . S'il est raisonnable de supposer que la population a bien un comportement de type logistique, ce graphe empirique de la fonction  $f$  doit avoir l'allure d'une parabole et les coordonnées de son sommet doivent être égales à  $(\frac{K}{2}, \frac{rK}{4})$ . Ceci permet d'estimer les deux constantes  $r$  et  $K$  approximativement à partir des coordonnées observées du sommet.

Voici un exemple d'application :

En 1927, Pearl a étudié la dynamique d'une culture de cellules de levure et il a obtenu les mesures suivantes (la taille de la levure est exprimée en biomasse ( $mg\ 100ml^{-1}$ )) :

| t=Heures   | 0   | 1   | 2   | 3   | 4  | 5  | 6  | 7   | 8   | 9   | 10   | 12  | 14   | 18  |
|------------|-----|-----|-----|-----|----|----|----|-----|-----|-----|------|-----|------|-----|
| N=Biomasse | 4   | 7   | 12  | 19  | 28 | 48 | 70 | 103 | 140 | 176 | 205  | 238 | 256  | 265 |
| Z=taux     | ..3 | ..5 | ..7 | ..9 | 20 | 22 | 33 | 37  | 36  | 29  | 16,5 | 9   | 2,25 | --- |

Calculer les taux de variation (compléter le tableau) et tracer le graphe empirique de  $f$  correspondant (en utilisant votre calculatrice pour tracer le graphe empirique de  $f$ , si vous avez une calculatrice graphique, ou approximativement à la main) Puis utiliser ce graphe pour calibrer un modèle logistique aux données de Pearl (en proposant des valeurs de  $K$  et  $r$ ). Expliquez vos choix.



On observe un "maximum" pour cette courbe situé auprès du point  $(103, 37)$ , peut-être à  $(110, 38)$  approximativement.

Comme le maximum théorique est situé en  $(\frac{K}{2}, \frac{rK}{4})$ , on en déduit que  $\begin{cases} \frac{K}{2} = 110 \\ \frac{rK}{4} = 38 \end{cases}$  et donc que  $K = 220$  et  $r = \frac{4(38)}{220}$

d'où  $r = 0,69$ . On sent bien que ce choix est plutôt arbitraire compte tenu du petit nombre de points et de la forme irrégulière de la courbe.

La méthode précédente est en fait assez grossière. En pratique, une bonne calibration relève plutôt de méthodes statistiques, la plus simple de ces méthodes statistiques étant la *méthode des moindres carrés* que nous venons d'introduire. Voici comment on procède.

On note que si  $N(t)$  désigne la taille de la population à l'instant  $t$  l'équation différentielle logistique  $N' = rN(1 - \frac{N}{K})$  peut se ré-écrire  $\frac{N'}{N} = r - \frac{r}{K}N$ . On en déduit que le *taux de biomasse*,  $\frac{N'}{N}$ , est en réalité une *fonction linéaire* (ou plutôt *affine*) de la biomasse. Donc si l'on peut calculer ce taux à partir des données, une simple régression linéaire pourra permettre de calculer les deux constantes  $\frac{r}{K}$  (pente de la droite de régression) et  $r$  (ordonnée à l'origine de la droite de régression). En pratique on approxime ce taux de biomasse par le rapport  $\tau_i = \frac{N_{i+1} - N_i}{(t_{i+1} - t_i)N_i}$ .

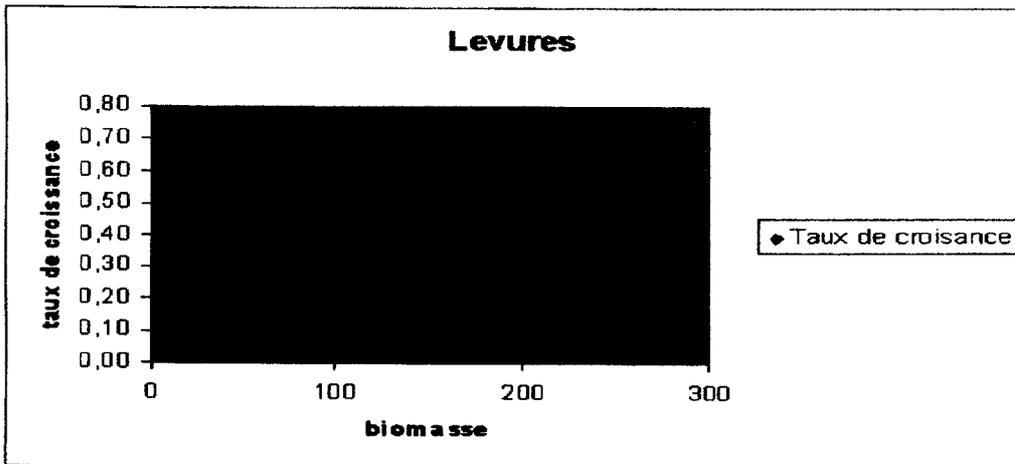
1. Expliquer pourquoi  $\tau_i$  est une approximation de  $\frac{N'}{N}$

$N'(t)$  est par définition la limite quand  $h \rightarrow 0$  de  $\frac{N(t+h) - N(t)}{(t+h) - t}$   
 Si l'on prend  $h = t_{i+1} - t_i$  comme "pas de temps" et si c'est assez petit, le rapport  $\frac{N_{i+1} - N_i}{t_{i+1} - t_i}$  sera proche de  $N'$  et donc  $\frac{N_{i+1} - N_i}{(t_{i+1} - t_i)N_i}$  proche de  $\frac{N'}{N_i}$ .

2. Compléter le tableau suivant :

| t=Heures   | 0 | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 12   | 14   | 18   |     |
|------------|---|------|------|------|------|------|------|------|------|------|------|------|------|------|-----|
| N=Biomasse | 4 | 7    | 12   | 19   | 28   | 48   | 70   | 103  | 140  | 176  | 205  | 238  | 256  | 265  |     |
| $\tau$     |   | 0,75 | 0,71 | 0,58 | 0,47 | 0,71 | 0,46 | 0,47 | 0,36 | 0,26 | 0,16 | 0,08 | 0,04 | 0,01 | --- |

3. Voici le résultat de la régression linéaire de  $\tau$  sur  $N$  obtenu sous Excel. Calculer les valeurs de deux constantes  $r$  et  $K$  en expliquant vos calculs.



En déduire les valeurs de  $K$  et  $r$  et le modèle logistique suivi par la biomasse de levure  $N(t)$ . Comparer avec les résultats précédents et commenter.

L'équation  $\frac{N'}{N} = r - \frac{r}{K}N$  est de la forme  $\frac{N'}{N} = aN + b$  avec  $a = -\frac{r}{K}$  et  $b = r$ . Le tableau précédent indique que le taux de croissance est égal à  $-0,0026$  fois la biomasse, plus  $0,6588$ .  
 Donc  $r = 0,6588$  et  $-\frac{r}{K} = -0,0026$  - Donc  $K = \frac{0,6588}{0,0026} \approx 253$ .