

## TD03

### Exercice 1 (régression linéaire simple)

dans le cadre de travaux de recherche sur la *biomasse* ( $mg$ ), d'un certain type de plante, en fonction de la concentration de l'azote  $nh_4^+$  ( $\mu mol$ ), nous avons réalisé des expériences dont la biomasse moyenne ( $y$ ) ainsi que la concentration du l'azote ( $x$ ) en question sont données dans le tableau ci-dessus :

concentration $\mu mol$	0	100	200	400	600
biomasse $mg$	305	378	458	540	565

on donne :  $\sum x_i = 1300$ ;  $\sum y_i = 2246$ ;  $\sum x_i^2 = 570000$ ;  $\sum y_i^2 = 1056498$ ;  $\sum x_i y_i = 684400$ ;

afin de modéliser ces données, nous avons proposé le modèle linéaire suivant :

$$y = a + bx.$$

1. présenter graphiquement le nuage des points  $(x_i, y_i)$ . que peut-on conclure sur le modèle proposer?
2. calculer les estimations des paramètres  $a$  et  $b$  et donner la droite de régression.
3. calculer le coefficient de corrélation linéaire. que peut-on conclure?
4. pour un seuil de risque  $\alpha = 5\%$ , le modèle proposé est-il pertinent?
5. quelle biomasse prévoyez-vous à une concentration  $500 \mu mol$ ?

### Exercice 2 (Régression linéaire simple)

Dans le cadre de travaux de recherche sur la durée de la saison de végétation en montagne, des stations météorologiques sont installées à différentes altitudes. La température moyenne (variable  $Y$  en degrés Celsius) ainsi que l'altitude (variable  $X$  en mètres) de chaque station données dans le tableau ci-dessous :

altitude	1040	1230	1500	1600	1740	1950	2200	2530	2800	3100
température	7.4	6	4.5	3.8	2.9	1.9	1	-1.2	-1.5	-4.5

On donne :  $\sum x_i = 19690$ ;  $\sum y_i = 20.3$ ;  $\sum x_i^2 = 42925500$ ;  $\sum y_i^2 = 162.41$ ;  $\sum x_i y_i = 17671$ ;

1. Calculer le coefficient de corrélation linéaire.
2. Calculer les estimations des paramètres  $a$ ,  $b$  et  $\sigma^2$  pour la régression linéaire de  $Y$  sur  $X$ .
3. Quelle température moyenne prévoyez-vous à 1100 m? à 2300 m?

### Exercice 3 (Régression linéaire simple)

Dans le cadre d'une enquête visant à comparer, selon certains critères, différents *Sandwich* vendus dans les fast-foods, nous avons retenu les informations se trouvant dans le tableau ci-dessous.

Sandwich	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$
Poids (g)	150	92	193	90	135	169
Prix (DA)	190	140	270	90	180	130

On donne :

$\sum Poids = 829$ ;  $\sum Poids^2 = 123099$ ;  $\sum Prix = 1000$ ;  $\sum Prix^2 = 186000$ ; et  
 $\sum Poids * Prix = 147860$ ;

- Y-a-t-il une relation linéaire entre les variables Poids et Prix ? Pour répondre à cette question, faire un graphique, calculer le coefficient de corrélation des deux variables et l'équation de la droite de régression.
- Supposons qu'on désir d'augmenter le poids du Sandwich  $S_6$  à 180 g, alors quelle sera son nouveau prix?
- Le modèle linéaire proposé est-il adéquat pour la description de la relation entre les variables Poids et Prix?

**Exercice 4** (*Régression linéaire simple et transformation des variables*)

On veut prédire la hauteur  $H$  d'un arbre en fonction de son diamètre  $D$ . Pour faire une régression linéaire, on effectue un changement de variable en posant  $X = \ln(D)$  et  $Y = \ln(H)$ . Voici les mesures faites sur 5 arbres :

$D$	0.1999	0.3012	0.3791	0.6005	0.6570
$H$	9.2073	9.6794	10.8049	13.4637	14.1540

- Donner le coefficient de corrélation linéaire entre  $X$  et  $Y$ .
- Donner l'équation de la droite de régression de  $Y$  par rapport à  $X$ .
- Tester la pertinence de la régression au seuil de 5%.
- Donner la hauteur prévue d'un arbre de diamètre 0.7.

**Exercice 5** (*Régression linéaire simple et changement des variables*)

Dans le cadre de travaux de recherche sur l'absorbance, d'un produit en fonction de sa concentration, par une certaine plante, nous avons réalisé des expériences dont l'absorbance moyenne ( $Y$ ) ainsi que la concentration ( $x$ ) en question sont données dans le tableau ci-dessus :

X $\mu g/\mu l$	0	20	40	60	80	100	Somme
Y	0	0.205	0.331	0.515	0.584	0.671	2.3060

a) Afin de modéliser ces données, nous avons proposé le modèle linéaire suivant :

$$Y = a_1 x + b_1.$$

- Calculer les estimations des paramètres  $a_1$  et  $b_1$  et donner la droite de régression.
- Quelle absorbance prévoyez-vous à une concentration  $40 \mu g/\mu l$ ? Que peut-on conclure?
- Calculer le coefficient de corrélation linéaire, ce résultat confirme t-il les résultats obtenus en 3)?
- Pour un seuil de risque  $\alpha = 5\%$ , le modèle proposé est-il pertinent?

b) Vue les doutes qu'on a sur le modèle précédent, nous avons proposé le modèle suivant :

$$Z = e^Y = a_2 x + b_2.$$

- Complétez le tableau suivant :

X $\mu g/\mu l$	0	20	40	60	80	100	Somme
Z	1.0000						

2. Calculer les estimations des paramètres  $a_2$  et  $b_2$  pour la régression linéaire de  $Z$  sur  $X$ .
3. Quelle absorbance prévoyez-vous à une concentration  $40 \mu\text{g}/\mu\text{l}$ . Que peut-on conclure par rapport au premier modèle?
4. Calculer le coefficient de corrélation linéaire de ce nouveau modèle.
5. Indiquer quel est le meilleur modèle parmi les deux proposés (avec justification).