

**TD n° 1 : Régression linéaire simple 1**

**Exercice 1.** On souhaite expliquer le nombre de cellules végétales au millimètre carré (variable  $Y$ ) à partir du temps d'exposition au soleil en jours (variable  $X$ ). Pour  $n = 7$  expériences indépendantes, on observe les valeurs  $(x_1, y_1), \dots, (x_n, y_n)$  de  $(X, Y)$  suivantes :

$x_i$	2	4	8	10	24	40	52
$y_i$	6	11	15	20	39	62	85

1. Représenter le nuage de points  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ . À partir de celui-ci, expliquer pourquoi on peut envisager l'existence d'une liaison linéaire entre  $Y$  et  $X$ .  
On adopte le modèle de *rls* :  $Y = \beta_0 + \beta_1 X + \epsilon$ , avec  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ . Les paramètres  $\beta_0, \beta_1$  et  $\sigma$  sont des réels inconnus.
2. Calculer  $\bar{x}, \bar{y}, s_x, s_y, s_{ce_x}, s_{ce_y}, spe_{x,y}$  et  $r_{x,y}$ .
3. Calculer le  $R^2$  et le  $R^2$  ajusté. Que peut-on en déduire ?
4. Donner l'*emco* ponctuel  $b_1$  de  $\beta_1$  et l'*emco* ponctuel  $b_0$  de  $\beta_0$ .
5. Tracer la droite de régression.
6. Donner une estimation ponctuelle de  $\sigma$ . Calculer  $ete_1$  et  $ete_0$ .
7. Est-ce que la régression est significative ?
8. Donner un intervalle de confiance pour  $\beta_1$  au niveau 95%, et un intervalle de confiance pour  $\beta_0$  au niveau 95%.
9. Combien de cellules végétales peut-on prédire pour 30 jours d'exposition au soleil ?
10. Donner un intervalle de confiance pour le nombre moyen de cellules végétales pour 30 jours d'exposition au soleil au niveau 95%.
11. Représenter le graphique des résidus. Que peut-on en dire ?

**Exercice 2.** On souhaite expliquer le chemin de freinage en mètres d'un véhicule (distance parcourue entre le début du freinage et l'arrêt total) (variable  $Y$ ) à partir de sa vitesse en kilomètres heure (variable  $X$ ). Pour  $n = 12$  expériences indépendantes, on observe les valeurs  $(x_1, y_1), \dots, (x_n, y_n)$  de  $(X, Y)$  suivantes :

$x_i$	40	50	60	70	80	90	100	110	120	130	140	150
$y_i$	9	11	20	27	39	45	58	78	79	93	108	124

1. Représenter le nuage de points  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ . À partir de celui-ci, expliquer pourquoi on peut envisager l'existence d'une liaison linéaire entre  $Y$  et  $X$ .  
On adopte le modèle de *rls* :  $Y = \beta_0 + \beta_1 X + \epsilon$ , avec  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ . Les paramètres  $\beta_0, \beta_1$  et  $\sigma$  sont des réels inconnus.

2. Calculer  $\bar{x}$ ,  $\bar{y}$ ,  $s_x$ ,  $s_y$ ,  $sce_x$ ,  $sce_y$ ,  $spe_{x,y}$  et  $r_{x,y}$ .
3. Calculer le  $R^2$  et le  $R^2$  ajusté. Que peut-on en déduire ?
4. Donner l'emco ponctuel  $b_1$  de  $\beta_1$  et l'emco ponctuel  $b_0$  de  $\beta_0$ .
5. Tracer la droite de régression.
6. Donner une estimation ponctuelle de  $\sigma$ . Calculer  $ete_1$  et  $ete_0$ .
7. Est-ce que la régression est significative ?
8. Quel chemin de freinage peut-on prédire pour un véhicule roulant à 160 kilomètres heure ?
9. On considère les hypothèses :

$$H_0 : \beta_0 \leq -50 \quad \text{contre} \quad H_1 : \beta_0 > -50.$$

Peut-on rejeter  $H_0$  au risque 5% ?

10. On considère les hypothèses :

$$H_0 : \beta_1 = 1 \quad \text{contre} \quad H_1 : \beta_1 \neq 1.$$

Peut-on rejeter  $H_0$  au risque 5% ?

11. Donner un ellipsoïde de confiance pour  $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$  au niveau 95%.

**Exercice 3.** On souhaite expliquer le salaire annuel d'un homme de 30 ans (variable  $Y$ ) à partir du nombre d'années d'étude qu'il a suivi (variable  $X$ ). Pour un échantillon de  $n = 10$  hommes âgés de 30 ans, on observe des valeurs  $(x_1, y_1), \dots, (x_n, y_n)$  de  $(X, Y)$ . Celles-ci vérifient :

$\sum_{i=1}^n x_i$	$\sum_{i=1}^n y_i$	$\sum_{i=1}^n x_i y_i$	$\sum_{i=1}^n x_i^2$	$\sum_{i=1}^n y_i^2$
151	96.3	1526.3	2357	1036.13

Une modélisation possible est le modèle de *rls* :  $Y = \beta_0 + \beta_1 X + \epsilon$ , avec  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ . Les paramètres  $\beta_0$ ,  $\beta_1$  et  $\sigma$  sont des réels inconnus.

1. Calculer  $\bar{x}$ ,  $\bar{y}$ ,  $sce_x$ ,  $sce_y$ ,  $spe_{x,y}$  et  $r_{x,y}$ .
2. Calculer le  $R^2$  et le  $R^2$  ajusté. Que peut-on en déduire ?
3. Donner l'emco ponctuel  $b_1$  de  $\beta_1$  et l'emco ponctuel  $b_0$  de  $\beta_0$ . Donner l'équation de la droite de régression.
4. Donner une estimation ponctuelle de  $\sigma$ .
5. Calculer  $ete_1$ . Est-ce que la régression est significative ?