

## TD 2

**Exercice 1**

On examine l'évolution d'une variable  $Y$  en fonction de deux variables explicatives  $x$  et  $z$ . On dispose de  $n$  observations de ces variables. On note  $\mathbf{X} = (\mathbf{1} \ \mathbf{x} \ \mathbf{z})$  où  $\mathbf{1}$  est le vecteur constant de 1 et  $\mathbf{x}, \mathbf{z}$  sont les vecteurs des variables explicatives (de taille  $n$ ).

1. Nous avons obtenu les résultats suivants :

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 30 & 0 & 0 \\ ? & 10 & 7 \\ ? & ? & 15 \end{pmatrix}$$

- (a) Donner les valeurs manquantes.
  - (b) Que vaut  $n$ ?
  - (c) Calculer le coefficient de corrélation linéaire entre  $\mathbf{x}$  et  $\mathbf{z}$ .
2. La régression linéaire de  $\mathbf{Y}$  sur  $\mathbf{1}, \mathbf{x}, \mathbf{z}$  donne

$$\mathbf{y} = -2\mathbf{1} + \mathbf{x} + 2\mathbf{z} + \hat{\boldsymbol{\varepsilon}}, \quad SCR = \|\hat{\boldsymbol{\varepsilon}}\|^2 = 12$$

- (a) Déterminer la moyenne arithmétique  $\bar{y}$ .
- (b) Calculer la somme des carrés expliquée (SCE), la somme des carrés totale (SCT) et le coefficient de détermination.

**Exercice 2**

On a calculé l'équation de régression suivante à partir des données relatives à une classe de 66 élèves infirmières:

$$\hat{y} = 3.1 + 0.021x_1 + 0.075x_2 + 0.043x_3$$

où  $Y$  = note de l'examen théorique de l'étudiante,

$x_1$  = classement (à partir du bas de l'échelle) de l'étudiante dans le secondaire,

$x_2$  = note d'expression orale de l'étudiante,

$x_3$  = mesure psychométrique du caractère de l'étudiante.

D'autre part, les écart-types estimés associés aux estimations  $\hat{\beta}_j$ ,  $j = 1, 2, 3$ , sont respectivement de 0.019, 0.034 et 0.018.

- 1. Donner l'intervalle de confiance à 95% pour chacun des paramètres  $\beta_j$ ,  $j = 1, 2, 3$ .
- 2. Donner la valeur  $t_j$  de la statistique du test de Student pour chacun des paramètres  $\beta_j$ .

3. Quelle hypothèse sur la loi des erreurs faites-vous en répondant aux 2 questions précédentes?
4. Quel régresseur vous semble-t-il pouvoir conduire aux conclusions statistiquement les plus sûrs (c'est aussi le régresseur susceptible d'influer le plus possible sur la note de l'étudiante)?
5. Si vous aviez à écrire un rapport final, garderiez-vous le premier régresseur ou le supprimeriez-vous? Pourquoi?

### Exercice 3

Dans un pays européen, on a mesuré la perte de de voix du parti au pouvoir (variable  $Y$ ) lors de 6 élections intermédiaires, ayant eu lieu au cours des 30 dernières années.  $Y$  dépend de plusieurs variables explicatives dont deux semblent particulièrement importantes et faciles à mesurer :  $x_1$  = résultats des sondages favorables au dirigeant du gouvernement (issu du parti au pouvoir) au moment de l'élection (pourcentage d'électeurs approuvant la manière dont le dirigeant mène sa politique) et  $x_2$  = modification du revenu par tête (en euros) l'année précédant l'élection.

$y$	$x_1$	$x_2$
6.3%	35%	-30
2.5	44	90
1.3	62	-10
4.9	53	-10
-0.8	65	50
1.5	47	90

1. A partir des données ci-dessus, calculer l'équation de la régression multiple de  $Y$  en  $x_1$  et  $x_2$  en indiquant bien toutes les étapes intermédiaires du calcul (utiliser le logiciel  $R$  pour calculer  $\hat{\beta}$  à l'aide des commandes de produit et d'inversion de matrices).
2. Estimer la perte de voix  $y$  pour un sondage d'élection donnant 50% de satisfaits et pour une augmentation du revenu par tête de 60 euros.
3. Tester la nullité des paramètres  $\beta_1$  et  $\beta_2$  simultanément au seuil de 5%.
4. Donner les bornes de l'intervalle de confiance à 90% du paramètre  $\beta_1$  du modèle.

### Exercice 4

En reprenant l'exemple 4.4 du cours :

1. Tester l'hypothèse  $\beta_1 = \beta_2 = 0$ .
2. Calculer un intervalle de confiance des prévisions du sentiment de bien-être à partir des valeurs des régresseurs données à la page 19 du cours.