TD II

Exercice 1. Soit θ_1 et θ_2 deux paramètres réels inconnus et soit:

- Y_1 un estimateur sans biais de $\theta_1 + \theta_2$ et de variance σ^2
- Y_2 un estimateur sans biais de $2\theta_1 - \theta_2$ et de variance $4\sigma^2$
- Y_3 un estimateur sans biais de $6\theta_1 + 3\theta_2$ et de variance $9\sigma^2$

Les estimateurs Y_1, Y_2 et Y_3 étant indépendants, nous cherchons les estimateurs sans biais de θ_1 et θ_2, linéaires en Y_1, Y_2 et Y_3, et de variance minimale.

1. Notons $\tilde{\theta} = \alpha Y_1 + \beta Y_2 + \gamma Y_3$.
   (a) Quelles sont les équations à satisfaire pour que $\tilde{\theta}$ soit un estimateur sans biais de θ_1?
   (b) Dans ce cas-là, exprimer la variance de $\tilde{\theta}$ et la minimiser.
   (c) Idem pour θ_2.
2. Notons $Z_1 = Y_1, Z_2 = Y_2/2, Z_3 = Y_3/3, Z = (Z_1, Z_2, Z_3)'$ et $\theta = (\theta_1, \theta_2)'$
   (a) Trouver la matrice X telle que $\mathbb{E}(Z) = X\theta$.
   (b) Que vaut $\mathbb{V}(Z)$?
   (c) On peut alors écrire $Z = X\theta + \varepsilon$. Retrouver les estimateurs de θ_1 et θ_2 calculés question 1.

Exercice 2.

1. Nous avons une variable Y à expliquer par une variable X. Nous avons effectué $n = 2$ mesures et trouvé
$$(x_1, y_1) = (4, 5) \text{ et } (x_2, y_2) = (1, 5)$$
   Représenter les variables, estimer β dans le modèle $y_i = \beta x_i + \varepsilon_i$ et représenter $\hat{Y}$
2. Nous avons maintenant une variable Y à expliquer par deux variables X_1 et X_2. Nous avons effectué $n = 3$ mesures et trouvé
$$(x_{1,1}, x_{1,2}, y_1) = (3, 2, 0), (x_{2,1}, x_{2,2}, y_2) = (3, 3, 5) \text{ et } (x_{3,1}, x_{3,2}, y_3) = (0, 0, 3)$$
   Représenter les variables, estimer β dans le modèle $y_i = \beta_1 x_{i.1} + \beta_2 x_{i,2} + \varepsilon_i$ et représenter $\hat{Y}$.

Exercice 3. Soit X une matrice de taille $n \times p$ composée de p vecteurs indépendants de $\mathbb{R}^n$. Nous notons X_q la matrice composée des $q (q < p)$ premiers vecteurs de X. Nous avons les deux modèles suivants :
$$Y = X\beta + \varepsilon \quad \text{et} \quad Y = X_q\beta_q + \psi$$
Comparer les R^2 dans les deux modèles.

Exercice 4. On examine l'évolution d'une variable Y en fonction de deux variables x et z. On dispose de n observations de ces variables. On note $X = \begin{pmatrix} \mathbb{1} & x & z \end{pmatrix}$ où $\mathbb{1}$ est le vecteur constant et x, z sont les vecteurs des variables explicatives.

1. Nous avons obtenu les résultats suivants:
$$X'X = \begin{pmatrix} 30 & 0 & 0 \\ ? & 10 & 7 \\ ? & ? & 15 \end{pmatrix}$$
   (a) Donner les valeurs manquantes?
   (b) Que vaut n ?
   (c) Calculer le coefficient de corrélation linéaire empirique entre x et z.

2. La régression linéaire empirique de Y sur $\mathbb{1}, x, z$ donne

$$Y = -2\mathbb{1} + x + 2z + \hat{\varepsilon}, \quad SCR = \|\hat{\varepsilon}\|^2 = 12$$

   (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 R^2 que l'on notera $R_{\mathbb{1},x,y}$.
3. (a) Calculer $X'Y$ en utilisant la valeur de $\hat{\beta}$. En déduire $\sum x_i y_i$ et $\sum z_i y_i$.
   (b) Calculer les coefficients linéaires $\rho(x, y)$ et $\rho(z, y)$. En déduire la valeur du R^2 pour le modèle de régression de y par $\mathbb{1}$ et x puis de y par $\mathbb{1}$ et z. On les note $R^2_{\mathbb{1},x}$ et $R^2_{\mathbb{1},y}$ respectivement.

Exercice 5. Régression sur données agrégées par groupes. On suppose le modèle de régression $Y = X\beta + \varepsilon, \mathbb{E}(\varepsilon) = 0, \mathbb{V}(\varepsilon) = \sigma^2 I_n$, valide mais les données individuelles $(x_{i1}, \cdots, x_{ip}, y_i)$ ne sont pas disponibles. On observe seulement les moyennes sur I groupes, notés $C_1, \cdots, C_I$, d'effectifs $n_1, \cdots, n_I$:

$$\bar{y}_k = \frac{1}{n_k} \sum_{i \in C_k} y_i \text{ et } \bar{x}_{kj} = \frac{1}{n_k} \sum_{i \in C_k} x_{ij}$$

En notant $\bar{\varepsilon}_k = \frac{1}{n_k} \sum_{i \in C_k} \varepsilon_i$, on a $\bar{Y} = \bar{X}\beta + \bar{\varepsilon}$.

1. Calculer $\mathbb{E}(\bar{\varepsilon})$ et $\mathbb{V}(\bar{\varepsilon})$.
2. Notons $M = \text{diag}\left(\sqrt{n_1}, \cdots, \sqrt{n_I}\right), Y^* = M\bar{Y}, X^* = M\bar{X}$ et $\varepsilon^* = M\bar{\varepsilon}$. Quelle est la relation entre Y^*, X^* et ε^*? Calculer $\mathbb{E}(\varepsilon^*)$ et $\mathbb{V}(\varepsilon^*)$?
3. En déduire un estimateur de β.
4. Application numérique : $I = 3$ avec $n_1 = 1$ et $n_2 = n_3 = 2$. $\bar{X}'_1 = (1, 1, 1), \bar{X}'_2 = (7, 12, 5)$ et $\bar{Y}' = (15, 25, 10)$