

TD 1

Exercice 1 (Questions de cours)

1. La somme des résidus calculés d'une régression multiple vaut :

- A. 0,
- B. approximativement 0,
- C. parfois 0.

2. Le vecteur $\hat{\mathbf{Y}}$ est orthogonal au vecteur des résidus $\hat{\boldsymbol{\varepsilon}}$:

- A. oui,
- B. non,
- C. seulement si **1** fait partie des variables explicatives.

3. Un estimateur de la variance de $\hat{\boldsymbol{\beta}}$ de l'estimateur des MC de $\boldsymbol{\beta}$ vaut :

- A. $\sigma^2(\mathbf{X}^T \mathbf{X})^{-1}$,
- B. $\hat{\sigma}^2(\mathbf{X}^T \mathbf{X})^{-1}$,
- C. $\hat{\sigma}^2(\mathbf{X} \mathbf{X}^T)^{-1}$.

4. Un autre estimateur $\tilde{\boldsymbol{\beta}}$ que celui des moindres carrés (moindres valeurs absolues par exemple ou d'autres encore) a été calculé. La somme des carrés des résidus (SCR) obtenue avec cet estimateur est :

- A. plus petite que la SCR obtenue avec l'estimateur des MC classique,
- B. plus grande que la SCR obtenue avec l'estimateur des MC classique,
- C. aucun rapport.

5. Nous pouvons justifier les moindres carrés quand $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(0_n, \sigma^2 I_n)$ via l'application du maximum de vraisemblance :

- A. oui,
- B. non,
- C. aucun rapport entre les deux méthodes.

6. Les estimateurs $\hat{\boldsymbol{\beta}}$ des MC et $\tilde{\boldsymbol{\beta}}$ du maximum de vraisemblance sont-ils différents? :

- A. oui,
- B. non,
- C. pas toujours, cela dépend de la loi des erreurs.

7. Les estimateurs $\hat{\sigma}^2$ des MC et $\tilde{\sigma}$ du maximum de vraisemblance sont-ils différents? :

- A. oui,
- B. non,
- C. pas toujours, cela dépend de la loi des erreurs.

Exercice 2

Le responsable d'un centre de tri postal fait faire une enquête pour évaluer l'efficacité du centre. L'enquête porte sur un échantillon de 27 employés et rassemble des données sur les variables : productivité (nombre de lettres correctement triées par minute), expérience (nombre de mois d'expérience au sein du centre) et score d'aptitude (score obtenu au test d'entrée pour obtenir l'emploi au centre). L'équation de régression suivante a été obtenue :

$$\hat{Prod} = 2.0 + 0.5Exp + 0.2Apt$$

1. Un employé venant juste d'être recruté a obtenu un score d'aptitude de 80. Quel niveau de productivité peut-on attendre de lui?
2. A quelle augmentation de sa productivité peut-on s'attendre après avoir bénéficié de 6 mois d'expérience?
3. Donner une interprétation du terme "constant" $\hat{\beta}_0$ (2.0) dans l'équation de régression.
4. Monsieur A a 3 mois d'expérience de plus que Monsieur B mais son score d'aptitude est de 20 points en dessous de celui de Monsieur B. Lequel des deux est supposé être le plus productif?

Exercice 3

Pour déterminer les effets de différents régresseurs sur le prix du terrain à bâtir dans une région donnée, on a déterminé une équation de régression liant le prix du terrain et différents régresseurs. A partir de données relatives à $n = 316$ lots, on a obtenu l'équation suivante :

$$\hat{y} = 10.3 + 1.5x_1 - 1.1x_2 - 1.34x_3$$

où Y = prix du terrain (lot à bâtir) par unité de longueur de façade, x_1 = année de la vente ($x_1 = 1, \dots, 5$ de 1966, ..., 1970), x_2 = surface du lot (acres), x_3 = distance à la route goudronnée la plus proche (km).

1. En supposant les autres régresseurs constants, le prix d'un lot de 5 acres était-il supérieur ou inférieur à celui d'un lot de 2 acres? De combien?
2. En supposant les autres régresseurs constants, de combien augmenterait le prix d'un lot si le lot était situé à 0.5 km de la route goudronnée la plus proche par rapport à un autre situé à 1 km?
3. Le prix moyen de vente d'un lot était-il supérieur en 1970 par rapport à 1966? De combien?

Exercice 4

Supposons que sur 4 lots de terre identiques, on répande au hasard les quatre doses suivantes d'engrais, et que l'on observe le rendement en blé:

Engrais (dose en kg/ha)	Rendement (q/ha)
1	70
2	70
4	80
5	100

1. Indiquer quelles sont la variable à expliquer et la (ou les) variable(s) explicative(s).
2. Estimer les paramètres du modèle de régression linéaire en utilisant le calcul matriciel (sans l'aide d'un logiciel).
3. Utiliser l'équation de régression pour prévoir le rendement si on utilise 3kg/ha d'engrais.