University of Montpellier — Year 2024-2025
HAX912X - Generalized Linear Model / Machine Learning

# Additional work on multiple linear regression

**Exercise 1**

Consider the model

$$y_i = bx_i + u_i, \quad \mathbb{E}[u_i] = 0, \quad \mathbb{E}[u_i^2] = \sigma^2, \quad \mathbb{E}[u_i u_j] = 0$$

where $x_i$ is a scalar.
Give the least squares estimator $\widehat{b}$. Let $b^{\#} = \sum y_i / \sum x_i$ be the estimator.
Compare these two estimators by calculating their and their variance (check that the BLUE property applies : the variance of $\widehat{b}$ is less than that of $b^{\#}$).

**Exercise 2**

We regress $y$ on two explanatory variables $x$ and $z$, i.e. $X = (\mathbf{1}, x, z)$. There are $n$ individuals in all, we obtained the following result :

$$X^T X = \begin{pmatrix} 5 & 3 & 0 \\ 3 & 3 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

1. What is the value of $n$ ? What is the empirical correlation between $x$ and $z$ ?

Linear regression provides the results :

$$y = -1 + 3x + 4z + \widehat{u}, \quad RSS = 3.$$

2 What is the value of $\bar{y}$ ?

3 Calculate $\|\widehat{y}\|^2$ and deduce ESS, TSS and the coefficient of determination $R^2$.

**Exercise 3**

Consider the regression model

$$y_i = ax_i + u_i, \quad i = 1, \ldots N$$

with $E[u_i] = 0, \quad \mathbb{V}(u_i) = \sigma_i^2, \quad \mathbb{C}(u_i, u_j) = 0, \quad i \neq j.$
$x_i$ and $a$ are scalars. Give the expression for the OLS and GLS estimators of $a$ and compare their variance.