Exercice 1 :

Let consider the simple linear model defined by the matricial writting :

$$\underline{Y} = \mathbb{X}\beta + U \qquad \text{where } \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

We see that :

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum (x_i - \bar{x}_n)^2}$$

1) Prove that another expression for $\hat{\beta}_1$ is :

$$\hat{\beta}_1 = \beta_1 + \frac{\sum (x_i - \bar{x}_n)\varepsilon_i}{\sum (x_i - \bar{x}_n)^2}$$

2) Deduce that $\hat{\beta}_1$ is an unbiaised estimator of β_1

3) Compute the variance of $\hat{\beta}_1$.

4) Under the assumptions

. $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ for $i \in \{1, .., n\}$

. independance of the ε_i

prove that :

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim \mathcal{N}(\beta, \sigma^2 V)$$

$$\text{with } V = \frac{1}{\sum (x_i - \bar{x}_n)^2} \begin{pmatrix} \frac{\sum x_i^2}{n} & -\bar{x}_n \\ -\bar{x}_n & 1 \end{pmatrix}$$

5) Prove the expression of the confidence interval for y_{n+1} which is

$$\left[\hat{Y}_{n+1} \pm \hat{\sigma}_n t_{n-2;1-\alpha/2} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}} \right]$$

where $\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$, $\hat{\sigma}_n^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ and $P(\underset{T(n-2)}{T} \leq t_{n-2;1-\alpha/2}) = 1-\alpha/2$

Exercice 2

Consider the mtcars database available in the R software.

The mpg variable is the response variable, all the others are explanatory variable.

1) Write your code to compute $\hat{\beta}$, $\hat{\sigma}_n^2$ and to decide if a linear model is a good model.

2) Compare your results with the one produced by the lm function

3) Perform by yourself, without using a dedicated function, a backward procedure to perform variable selection.

4) Verify is the gaussian assumption onto the noise is verified. Why this is important to see this with respect to question 3?

Exercice 3:

Run this code:

```
n = 10 000
l = 4.7
a = -1/l * log(runif(n))
```

what kind of variable can you assign to the observations created in the object a?

Convince me with graphics and a test.