# First Assignment: Multiple Regression

1. (1 point) Show that the properties of least squares estimators are satisfied using the following definitions:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \hat{\mathbf{Y}} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y} \\ \hat{\boldsymbol{\varepsilon}} &= \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \end{aligned}$$

2. (1 point) Using model `modall`, check numerically that the properties of the least squares estimates are satisfied.

3. (1 point) Check that for the dataset `index.txt`, the least squares estimates of the parameters are: $\hat{\beta}_0 = 4.267$ and $\hat{\beta}_1 = 1.373$, using the results in section 2.4.1 (not using the lm() function).

4. (1 point) Given the response variable $y$ and the covariates `x2` and `x3` in the dataset `Transform_V2.txt` dataset , check if is necessary to transform any variable and the residual graphs to show that the transformed model is correct.

5. (1 point) Given the response variable $y$ and the covariates `x1` and `x2` in the datset `Transform2_V2.txt`, check if is necessary to transform any variable and the residual graphs to show that the transformed model is correct.

6. (1.25 points) In the case of ridge regression, calculate $bias\left(\hat{\beta}\right)$ and show that $Var(\hat{\beta}_{\mathbf{OLS}}) \geq Var(\hat{\beta}_{\mathbf{ridge}})$

7. (0.75 points) Calculate the value of $R^2$ and $R_a^2$ for model `fit.ridge` and compare them with the results of `modall (modall <- lm(hwfat ., data = bodyfat))`

8. (3 points) The dataset insurance.csv contains data of insurance premiums paid by people in USA depending on their personal characteristics:

   - age: age of the policy holder
   - sex: male/female

- : bmi: body mass index
- smoker: yesy/no
- children: number of children
- region: region where he/she lives
- changres: ( anual cost in dolars, of the insurance (response variable)

a) Find the model that gives the best prediction (take into account that interactions between varaibles may be present)

b) What is the profile of the people that pay more (or less) for their insurance?