

# EOSC 410/510 Assignment 1:

Note: Please do **not** submit your code; only submit your assignment as a PDF with the figures/results/tables embedded inside the document. **Include your name(s) in the document name**, e.g. Assignment1\_Radic.pdf

Please submit the assignment on Canvas.

## Problem 1 [total of 6 points]:

The given data file (corr\_data.mat file in Matlab format or corr.csv as a text file) consists of 6 variables: x, x2, x3, y, y2, y3, each with 40 observations. Do the following analysis on the data:

- 1) Calculate Pearson correlation for each of the following pair of variables: x and y, x2 and y2, and x3 and y3. This should produce three correlation coefficients. *[1 point for Pearson correlation coefficients]*
- 2) Make scatterplots of the data points in the x-y space, the x2-y2 space and the x3-y3 space. Also plot the linear regression line in the scatterplots. *[1 point for each plot (data and regression line), so 3 points in total]*
- 3) For each scatter plot answer the following: are there any outliers and if yes, is the correlation resistant to outliers? If not resistant to outliers, what resistant statistical measure would you need to use as an alternative to Pearson correlation? *[2 points for the answers]*

## Problem 2 [total of 9 points]:

The given data file (MLR.mat file in Matlab format or MLR.csv as a text file) consists of 6 predictors (x1, x2, x3, x4, x5 and x6 ) and one response variable y. Do the following analysis on the data:

- a) plot the data as timeseries in one plot *[1 point for the plot]*
- b) perform multiple linear regression (MLR) and show the values of regression coefficients. Plot modelled y against true y in a scatterplot, and calculate multiple correlation coefficient ( $R^2$ ) *[1 point for the regression coefficients for MLR, 1 point for the plot, 1 point for  $R^2$ ]*
- c) perform stepwise regression, and show the final model equation with coefficients' values. Plot modelled y against true y in a scatterplot, and calculate multiple correlation coefficient ( $R^2$ ). Rank the importance of the individual predictors in their influence on y. *[1 point for the final equation for the stepwise regression, 1 point for the plot, 1 point for  $R^2$ , 1 point if correct standardization of data is applied, 1 point for ranking of predictors]*