

Multivariate Statistical Methods for Engineering and Management (MEMEG, 1st Semester, 2022/2023)

Handed out on 5 of November, 2022.

To be handed back on 30 of November, 2022.

Group 4:

Maria Pereira	89499
Margarida de Abreu	93656
Ana Marques	93424
Francisco da Silva	93639
António Andion	90537

Consider for the **Boston** data frame, available in R library **MASS**, the subset with observations **101 to 150** for all variables except **zn**, **chas** and **ptratio**.

1. Make an exploratory analysis, using plots and summary statistics (e. g. mean, covariance, generalized/total variance and Mahalanobis distances) to describe the data.
2. One researcher has rudimentary knowledge about multiple linear regression analysis and wants your help to find a way to explain the response variable **medv** with some predictor variables.
 - (a) With the aim of applying a multiple regression model, make a preliminary analysis of the data and discuss what you have learned from this analysis.
 - (b) Fit a regression model to the dataset.
 - (c) Test for significance of the regression. Discuss the results in terms of the test p-value. Compare the test results with the coefficient of multiple determination and its adjusted version. Is there any evidence that a subset of the original variables should be excluded from the model? Proceed in order to find the best subset of regressors.
 - (d) For the subset of regressors selected in (c), obtain 94% confidence intervals (CI) for the mean response with the observation **101** and the observation **146**. For the same values of the regressors, and the same confidence level, compute the two prediction intervals (PI). Compare and discuss the results.
 - (e) Check model adequacy, investigate possible influential/leverage observations and outliers.

About the report:

- The report should not exceed 20 pages (including Annexes).
- Do not forget to include in the report: introduction, the dataset in study, objectives of the study, methodology used, decisions, conclusions and bibliography.
- The R code and the report must be send to me: `irodrig@math.tecnico.ulisboa.pt` and also a print copy.