STAT9006   Sample Assessment 04   Dr Seán Lacey

A catalog company, interested in developing a forecasting model, has collected data on quarterly *Sales of men's clothing (€)* along with several measurements (predictors) that might be used to explain some of the variation in sales. Possible predictors include the *Number of pages in catalogs* that were mailed, the *Number of phone lines open for ordering*, the *Amount spent on print advertising (€)*, and the *Number of customer service representatives*. Are any of these predictors useful for forecasting?

Open *Sample Assessment 04: Dataset* on Canvas. Report and interpret your findings using 0.05 level of significance.

Requirements: In your assessment, you are tasked with compiling a report investigating whether or not the sales of men's clothing is dependent on the number of pages in a catalog, phone lines open for ordering, customer service representatives and amount spent on print advertising. Your report should use *R* to compile the appropriate tables/graphs/tests/etc. You are expected to **explain all concepts and procedures** used in the analysis of the data - i.e., all the **descriptive statistics and statistical inference should be fully explained!** Your report should have the following sections:

Breakdown of marks (200 marks)

- 20 marks for a clear introduction to the assessment with reference to the:

  - **Type of data** presented, **type of study**, **sample size** and the **main questions** to be answered in the assessment.

- 20 marks for presentation of descriptive statistics (sample size, mean, standard deviation, median, 1st quartile, 3rd quartile, min, max) and determination of normality on all measurements (mean, median, skewness, verdict, p-value). Note: **no explanation required on statistics outputted for this section**.

- 40 marks for the **scatterplots**:

  - 20 marks for presentation of four plots describing the relationship between the response variable and each of the predictors (explanatory variables);

  - 20 marks for the appropriate **correlation coefficient** (with explanations) for each scatterplot.

- 20 marks for presentation and interpretation of the **correlation matrix** (predictors only).

  - 10 marks for clearly interpreting the correlation matrix;

  - 10 marks for a general comment on **multi-collinearity**.

- 50 marks for using **model selection** to develop a regression model:

  - 40 marks for presentation and explanation of the full model (adjusted R^2, regression coefficients, standardised regression coefficients, p-values);
  - 10 marks for applying and interpreting the Variance Inflation Factor (VIF) to determine whether or not multi-collinearity is present.

- 20 marks for **presentation of the final model** [adjusted R^2, regression coefficients (with 95% confidence interval), standardised regression coefficients, p-values]. Note: **no explanation required on statistics outputted for this section**.

- 30 marks for determining whether or not the conditions placed on the **residuals** are violated for the final model.

  - 15 marks for presenting the relevant plots;
  - 15 marks for correct conclusions (with corresponding p-values).