

ST 3082 / DS 3003 Lab worksheet

14th November 2025

1. ‘advertising.csv’ contains **Sales** data of a product in 200 different markets, along with advertising budgets for the product in each of those markets for three different media: **TV**, **Radio**, and **Newspaper**.
 - a. Import the dataset into R. Split the data into training and test set using 80%-20% ratio.
 - b. Draw a scatter plot for the variables **Sales** and **TV** using training data. Add test set data points also to the same graph using a different colour/shape.
 - c. Fit an OLS regression model to predict **Sales** using only **TV** budget. Represent the fitted regression line on the graph drawn in b).
 - d. Assess the model accuracy using suitable metrics.
 - e. Will the regression model improve if **Radio** and **Newspaper** budgets are also included as predictors?
 - f. How would you provide statistical evidence to determine whether at least one of the predictors **TV**, **Radio** or **Newspaper** is useful in predicting **Sales**?
 - g. Find the best models using best subset selection, forward selection and backward elimination methods.