Lab 02: Bivariate regression and Transformation

**Due date:** Wednesday, Feb 12, 2025 submitted as Word document to Canvas ***Lab02***  link

This lab counts 9 % toward your total grade.

**Objectives:** In this lab, you will practice your skills in

1. Explore bivariate regression
2. Confidence interval
3. Accuracy of the model
4. Variable Transformation

**Format of answer:** Submit your answers as a **Word document** with graphs and verbal descriptions, properly labeled in the task sequence, with answers in red text and only relevant content included

# Task 1: Bivariate regression (2.5 pts)

The **MASS** library contains the **Boston** data set, which records the attribute information about house in suburbs of Boston. We will use **rm** (average number rooms per house) to predict **medv** (median value of owner-occupied homes in $1000s).

1. Using the **lm()** function to fit a bivariate linear regression model, with **medv** as the dependent variable and **rm** as the independent variable. (0.5 pts)
2. Get detailed information about the linear model you constructed in 1.a using **summary()**. Interpret the intercept, slope and . (1 pts)
3. Compute the 95% confidence interval for the estimated regression parameters. Does the conclusion align with the results obtained from the t-test in part 1.b? yes or no, please interpret. (1pts)

# Task 2: Bivariate Regression Model and Variable Transformation (6.5 pts)

The **UN** dataset from the **carData** package contains various global development indicators. We will analyze the relationship between **infant mortality rate (infantMortality) (**dependent variable) and **GDP per capita (**ppgdp**)** as (independent variable).

1. Remove NA value in UN dataset using **na.omit(UN).** (0.5 pts)
2. Create a scatterplot of **infant mortality rate (infantMortality)** versus **GDP per capita (ppgdp) using car::scatterplot()**. By visually inspecting the box plots and the LOESS curve, determine whether data transformation is advisable for dependent variable and independent variable. (1pts)
3. If a transformation is needed for the independent variable, find the using:

**Box-Cox transformation (summary (car::powerTransform(lm(*varName*~1))))** Please justify whether log-transformation () should be used or if is more appropriate. (1pts)

1. Transformed the independent variable using , , . (1pts)
   1. Construct the histogram for the transformed distribution with different value.
   2. Evaluate the skewness and test whether the variables are approximately normal distribution.
2. If a transformation appears necessary for dependent variable, find the . (1pts)
3. Create a scatterplot with transformed variables **using car::scatterplot()**. Visually inspecting the box plots and the LOESS curve and describe how the transformation affects the relationship compared to the scatterplot in Task 2.b. (1pts)
4. Estimate the bivariate regression model using the transformed variables and interpret the estimated coefficients. (1pts)