Lab 02: Bivariate regression

**Due date:** Wednesday, Feb 5, 2025 submitted as Word document to Canvas ***Lab01***  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. 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 (3 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.
2. Get detailed information about the linear model you constructed in 1.a. Interpret the intercept, slope and .
3. Compute the 99% confidence interval for the estimated regression parameters. Does the conclusion align with the results obtained from the t-test in part 1.a? yes or no, please explain.

# Task 2: Transformation for variable (3 pts)

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

1. Remove NA value in UN dataset using na.omit(object)
2. Create a scatterplot of **infant mortality rate (infantMortality)** versus **GDP per capita (ppgdp) using scatterplot()**. By visually inspecting the box plots and the LOESS curve, determine whether data transformation is advisable for dependent variable and independent variable.
3. If a transformation appears necessary for independent variable, estimate the optimal **lambda** value using the **Box-Cox transformation (summary(car::powerTransform(lm(*varName*~1))))**.
4. Transformed the data using
   1. Construct the histogram for the transformed distribution with different value
   2. Evaluate the skewness and test whether the variables are approximately normal distribution
5. If you do the transformation, based on result from 1.b and 1.a, which (0,, -1)you should use? Please elaborate on the reason.
6. If a transformation appears necessary for dependent variable, estimate the optimal lambda value.
7. Create the scatterplot using transformed variables and please explain the change in box plot and scatterplot.