PRACTICAL 7 (Regression)

In this lab, we will explore the foundational concepts of regression analysis, a powerful statistical technique used to model relationships between a dependent variable and one or more independent variables. Regression allows us to understand how changes in predictor variables affects the outcome and to make predictions based on these relationships.

Model 1

```
# Load libraries
library(ggplot2)

# Simple linear regression
Model1 <- lm(mtcars$wt ~ mtcars$mpg)
Model1
summary(Model1)

# Plot with regression line
ggplot(mtcars, aes(x=mpg, y=wt)) +
  geom_point() +
  geom_smooth(method=lm)

# Plot without confidence interval
ggplot(mtcars, aes(x=mpg, y=wt)) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE)

# Customize point color, shape, line type, and color
ggplot(mtcars, aes(x=mpg, y=wt)) +
  geom_point(shape=18, color="blue") +
  geom_smooth(method=lm, se=FALSE, linetype="dashed", color="darkred")
```

Model 2 (multiple regression)

```
Model2 <- lm(mtcars$wt ~ mtcars$mpg + mtcars$hp)
Model2
summary(Model2)
```

Interpretation

1. State the regression equation for model 1 & 2.
2. How much variation in weight is explained by dependent variables in model 1 & model 2?
   (Hint : use multiple R2)
3. Assess whether the overall model1 and model 2 are statistically significant?
   (Hint : p value of ANOVA)
4. Is the variable hp contributing significantly in estimation of wt?
   (Hint : compare adjusted R2 values of both models)
5. Which of the independent variables in botht the models are statistically significant?
   (hint: p values of each independent variables)

6. Predict wt if mpg=3 using model1.
7. Predict wt If mpg=2, hp=5 using model2.

Exercise

1. Find the most likely price in Mumbai corresponding to the price of Rs. 70.00 at Calcutta from the following using correlation coefficient as 0.75:

Calcutta(X)		Mumbai(Y)
Mean	$\bar{x} = 65$	$\bar{y} = 67$
S.D.	$\sigma_x = 2.5$	$\sigma_y = 3.5$

   Write a R program for above problem.

2. The curb weight x in hundreds of pounds and braking distance y in feet, at 50 miles per hour on dry pavement, were measured for five vehicles, with the results shown in the table.
   X 25 27.5 32.5 35 45
   Y 105 125 140 140 150
3. Compute the linear correlation coefficient for these sample data and interpret its meaning. Also write a R program.
4. For auto dataset, Use the lm() function to perform a multiple linear regression with mpg as the response and all other variables except name as the predictors. Use the summary() function to print the results.
5. For carseats dataset, Fit a multiple regression model to predict Sales using Price,Urban, and US.
6. If the lines of regression of sample are x + 6y = 6 and 3x + 2y = 10.
   Find i) mean of x & y and ii) correlation coefficient between x and y
   Write a R program for the above example.