**LAB 03**

**LINEAR REGRESSION & COVARIANCE**

**Aim: To find the linear regression and covariance of a given data.**

**New-term used:**

1. **summary(): This function is used to compute summary statistics of a vector or data frame. It is used to display summary statistics of the data data frame.**
2. **var(): This function is used to calculate the variance of a numeric vector. It is used to calculate the variances of the "speed" and "dist" variables in the data data frame.**
3. **cov(): This function is used to calculate the covariance between two numeric vectors. It is used to calculate the covariance between the "speed" and "dist" variables in the data data frame.**
4. **sd(): This function is used to calculate the standard deviation of a numeric vector. It is used to calculate the standard deviations of the "speed" and "dist" variables in the data data frame.**
5. **cor.test(): This function is used to perform a correlation test between two numeric vectors. It is used to perform correlation tests between the "speed" and "dist" variables in the data data frame using different methods (Pearson, Spearman, and Kendall).**
6. **plot(): This function is used to create various types of plots. It is used to create scatter plots of the "speed" and "dist" variables in the data data frame and the "weight" and "BMI" variables in the second example.**
7. **lm(): This function is used to perform linear regression analysis. It is used to perform linear regression analyses with "speed" as the independent variable and "dist" as the dependent variable, and vice versa, in the data data frame. It is also used to perform a linear regression analysis with "weight" as the independent variable and "BMI" as the dependent variable in the second example.**
8. **abline(): This function is used to add lines to plots. It is used to add a line representing the linear regression model to the scatter plots of the "speed" and "dist" variables in the data data frame, and the "weight" and "BMI" variables in the second example.**
9. **library(): This function is used to load R packages. In this code, it is used to load the "scatterplot3d" package for creating 3D scatter plots in the second example.**

**Input:**

**#The following R code is used to analyze a data set called "cars"**

**#which contains information about the speed and stopping distance of cars.**

**#Assigning the data set "cars" to a variable called "data"**

**data = cars**

**data**

**summary(data)**

**#Calculating the variance of the "speed" variable in the "data" data frame**

**v1 = var(data$speed)**

**v1**

**#VARIANCE OF "dist"**

**v2 = var(data$dist)**

**v2**

**#covariance between "speed" and "dist"**

**covariance = cov(data$speed,data$dist)**

**covariance**

**#correlation coefficient between speed and distance**

**corr = covariance/(sd(data$speed)\*sd(data$dist))**

**corr**

**#Performing a correlation test between the "speed" and "dist" variables**

**#in the "data" data frame using the different methods and printing the results**

**#default method - pearson**

**cor.test(data$speed,data$dist)**

**cor.test(data$speed,data$dist,method="pearson")**

**cor.test(data$speed,data$dist,method="spearman")**

**cor.test(data$speed,data$dist,method="kendel")**

**#Creating a scatter plot of the "speed" and "dist" variables in the "data" data frame**

**plot(data$speed, data$dist)**

**#erforming a linear regression analysis with**

**#"speed" as the independent variable and "dist" as the dependent variable**

**regression1 = lm(data$speed ~ data$dist)**

**regression1**

**#Adding a line representing the linear regression model to the scatter plot**

**#of the "speed" and "dist" variables in the "data" data frame**

**abline(regression1)**

**#Performing a linear regression analysis with**

**#"dist" as the independent variable and "speed" as the dependent variable**

**regression2 = lm(data$dist~data$speed)**

**regression2**

**abline(regression2)**

**#EXAMPLE – 2 : Weight, BMI, and height in a class**

**weight= c( 15, 26, 27, 25 ,25.5, 27, 32, 18, 22,20 ,26, 24);**

**weight**

**BMI= c(15.35, 16.12, 16.74, 16.00, 13.59, 15.73, 15.65, 13.85, 16.07, 12.80, 13.65, 14.42);**

**BMI**

**height = c(175, 168, 170, 171, 169, 165, 165, 160, 180, 186);**

**height**

**plot(weight, BMI)**

**lm(weight ~ BMI)**

**abline(lm(weight ~ BMI))**

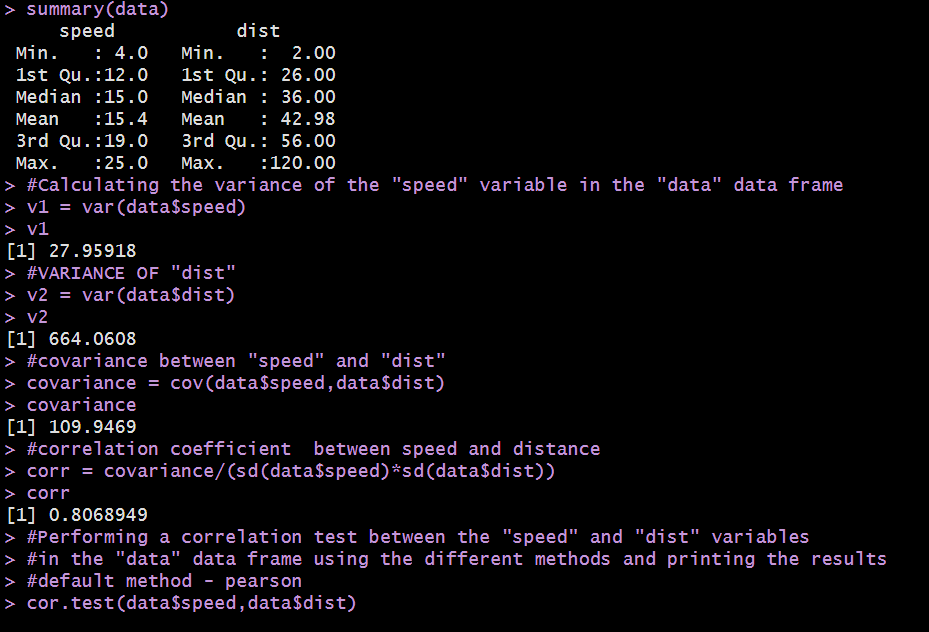
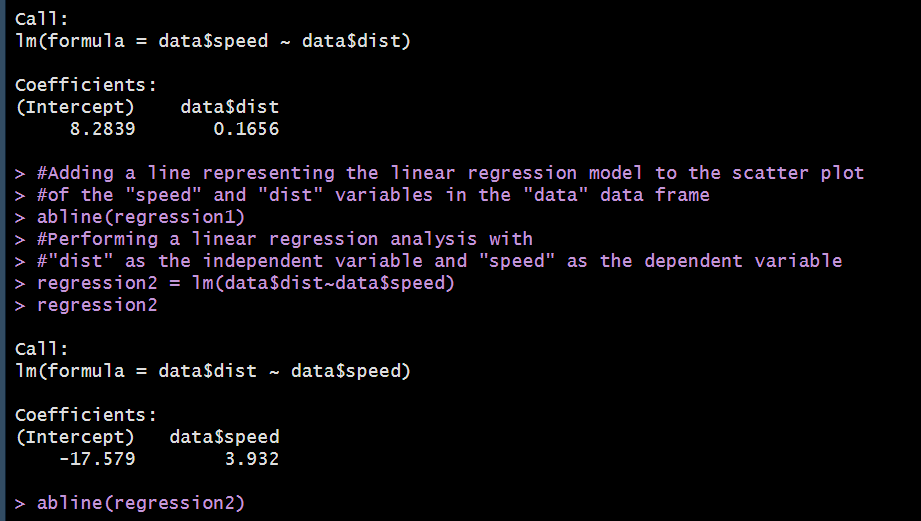
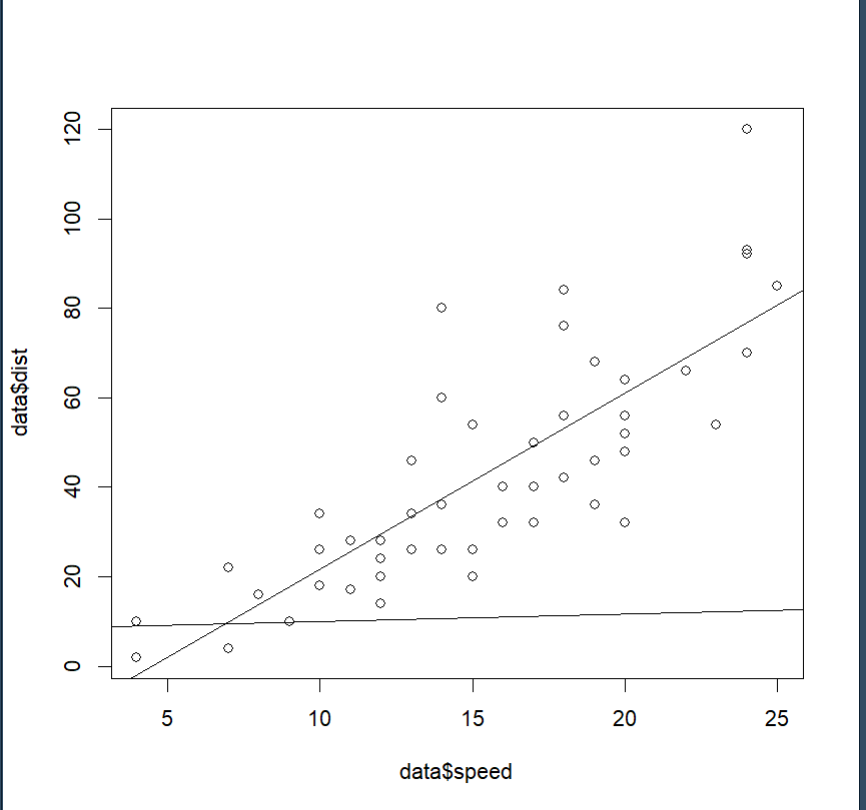
**plot(weight, BMI, main = "Scatterplot")**

**# Add a regression line**

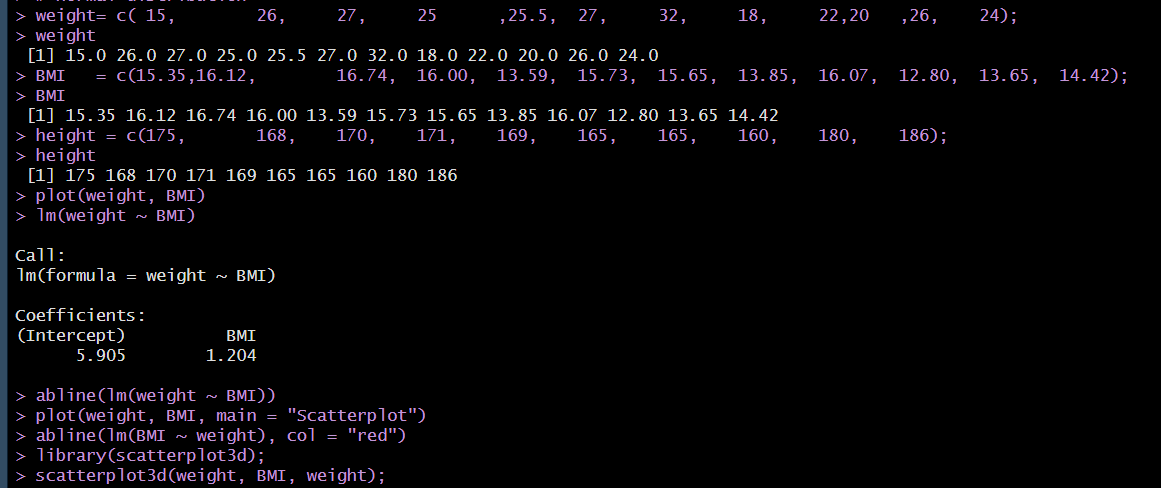
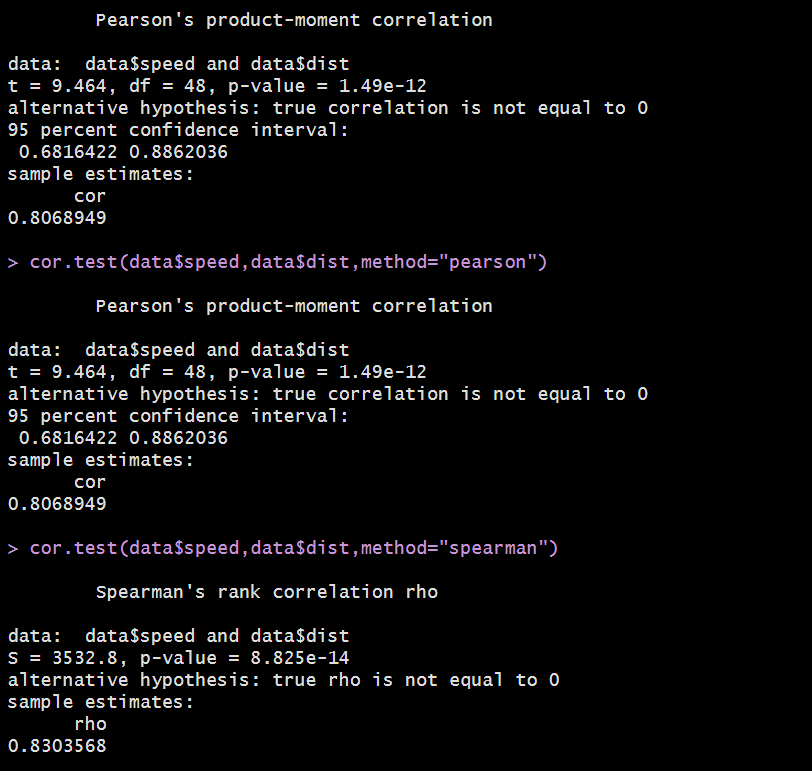
**abline(lm(BMI ~ weight), col = "red")**

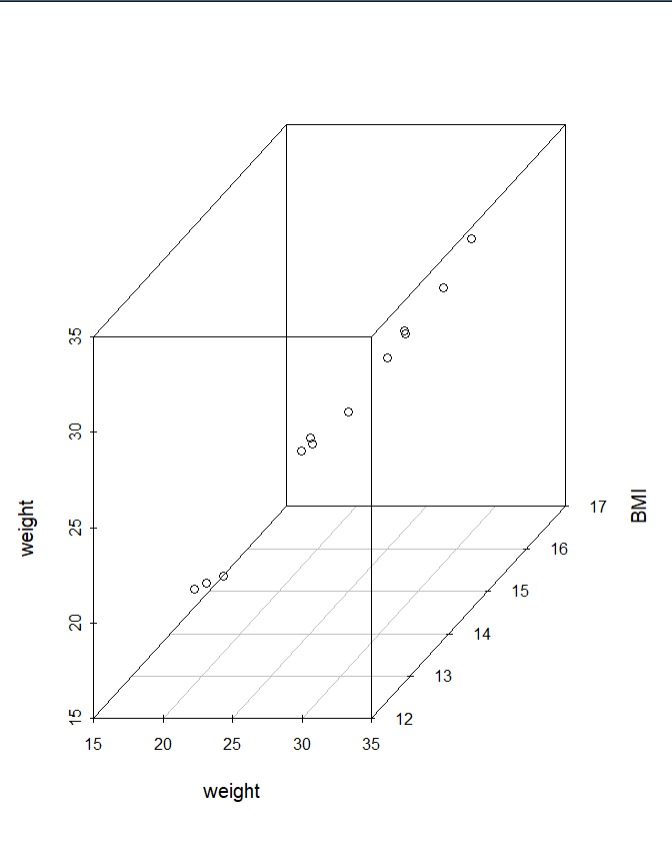
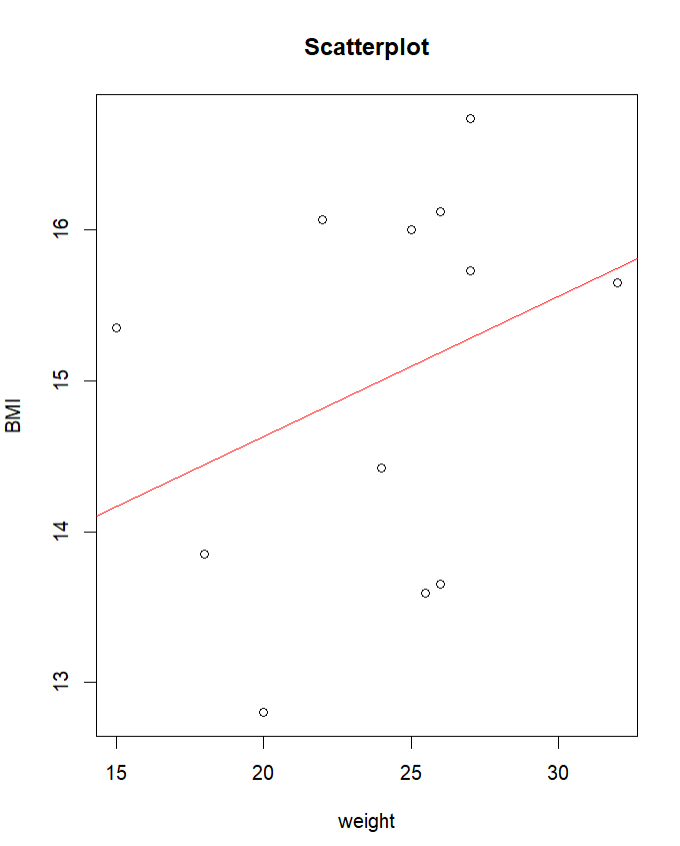
**library(scatterplot3d);**

**scatterplot3d(weight, BMI, weight);**

**OUTPUT:**

**OUTPUT – Example 2**

****

****

**INFERENCE:  
Hence linear regression of a data can be plotted and can be drawn in R programming**

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***