**SAVEETHA SCHOOL OF ENGINEERING**

**SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES**

**ITA 0443 - STATISTICS WITH R PROGRAMMING FOR REAL TIME PROBLEM**

**DAY 4– LAB MANUAL**

**Reg No:192124049**

**Name:SK:KHAJA MAINUDDIN**

**LINEAR REGRESSION ANALYSIS IN R**

**Exercise**

* Using linear regression analysis establish a relationship between height and weight of a person using the input vector given below.

# Values of height

151, 174, 138, 186, 128, 136, 179, 163, 152, 131

# Values of weight.

63, 81, 56, 91, 47, 57, 76, 72, 62, 48

Predict the weight of a person with height 170. Visualize the regression graphically.

CODE:

height <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)

weight <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)

relation <- lm(weight ~ height)

print(relation)

print(summary(relation))

a <- data.frame(height = 170)

result <- predict(relation, a)

print(result)

png(file = "linearregression.png")

plot(height, weight, col = "blue", main = "Height & Weight Regression", abline(lm(weight ~ height)), cex = 1.3, pch = 16, xlab = "Height in cm", ylab = "Weight in kg")

dev.off()

OUTPUT:

Call:

lm(formula = weight ~ height)

Coefficients:

(Intercept) height

-38.4551 0.6746

Call:

lm(formula = weight ~ height)

Residuals:

Min 1Q Median 3Q Max

-6.3002 -1.6629 0.0412 1.8944 3.9775

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -38.45509 8.04901 -4.778 0.00139 \*\*

height 0.67461 0.05191 12.997 1.16e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.253 on 8 degrees of freedom

Multiple R-squared: 0.9548, Adjusted R-squared: 0.9491

F-statistic: 168.9 on 1 and 8 DF, p-value: 1.164e-06

1

76.22869





* Download the Dataset "water" From Rdataset Link.Find out whether there is a linear relation between attributes"mortality" and"hardness" by plot function.Fit the Data into the Linear Regression model.Predict the mortality for the hardness=88

CODE:

library(ggplot2)

ggplot(water, aes(x = hardness, y = mortality)) +

geom\_point(shape = 21, fill = "blue", color = "black") +

geom\_smooth(method = lm, se = FALSE, color = "red")

model <- lm(mortality ~ hardness, data = water)

summary(model)

newdata <- data.frame(hardness = 88)

prediction <- predict(model, newdata = newdata)

prediction

OUTPUT:

Call:

lm(formula = mortality ~ hardness, data = water)

Residuals:

Min 1Q Median 3Q Max

-37.187 -11.981 -2.621 10.997 47.448

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1524.25 31.63 48.18 <2e-16 \*\*\*

hardness -3.23 0.134 -24.10 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.07 on 58 degrees of freedom

Multiple R-squared: 0.652, Adjusted R-squared: 0.6456

F-statistic: 581.8 on 1 and 58 DF, p-value: < 2.2e-16

**MULTIPLE REGRESSION ANALYSIS IN R**

**Exercise:**

3.Generate a multiple regression model using the built in dataset mtcars.It gives a comparison between different car models in terms of mileage per gallon (mpg), cylinder displacement("disp"), horse power("hp"), weight of the car("wt") and some more parameters.

Establish the relationship between "mpg" as a response variable with "disp","hp" and "wt" as predictor variables. Predict the mileage of the car with dsp=221,hp=102 and wt=2.91.

# Load the mtcars dataset

data(mtcars)

# Build the multiple regression model

model <- lm(mpg ~ disp + hp + wt, data = mtcars)

# Print the summary of the model

summary(model)

# Predict the mileage of a car with disp=221, hp=102, and wt=2.91

newdata <- data.frame(disp = 221, hp = 102, wt = 2.91)

prediction <- predict(model, newdata)

# Print the prediction

prediction

OUTPUT:

Call:

lm(formula = mpg ~ disp + hp + wt, data = mtcars)

Residuals:

Min 1Q Median 3Q Max

-3.4506 -1.6044 -0.1196 1.2193 4.6271

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 34.96055 2.16456 16.149 1.60e-16 \*\*\*

disp -0.01773 0.00902 -1.965 0.063251 .

hp -0.02148 0.02177 -0.987 0.334955

wt -3.35083 1.16468 -2.875 0.007068 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.747 on 28 degrees of freedom

Multiple R-squared: 0.8268, Adjusted R-squared: 0.8083

F-statistic: 44.87 on 3 and 28 DF, p-value: 1.087e-10

1

21.14008

4. Consider the data set "delivery" available in the R environment. It gives a deliverytime (“delTime”)of production materials(number of productions “n.prod”) with the given distance(“distance”) to reach the destination place.

a)Create the model to establish the relationship between "delTime" as a response variable with "n.prod" and "distance" as predictor variables.

b)Predict the delTime for the given number of production(“n.prod”)=9 and distance(“distance”)=450