ITA 0443 - STATISTICS WITH R PROGRAMMING FOR REAL TIME PROBLEM

DAY 4– LAB MANUAL

Reg No:BARATH

Name:192121147

LINEAR REGRESSION ANALYSIS IN R

1. 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.

INPUT:

# Input data

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

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

# Perform linear regression analysis

model <- lm(weight ~ height)

summary(model)

# Predict the weight of a person with height 170

newdata <- data.frame(height = 170)

pred <- predict(model, newdata)

cat("Predicted weight for height 170: ", round(pred, 2), "\n")

# Visualize the regression graphically

plot(height, weight, main = "Linear Regression", xlab = "Height", ylab = "Weight")

abline(model, col = "red")

OUTPUT:

> # Input data

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

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

>

> # Perform linear regression analysis

> model <- lm(weight ~ height)

> summary(model)

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

>

> # Predict the weight of a person with height 170

> newdata <- data.frame(height = 170)

> pred <- predict(model, newdata)

> cat("Predicted weight for height 170: ", round(pred, 2), "\n")

Predicted weight for height 170: 76.23

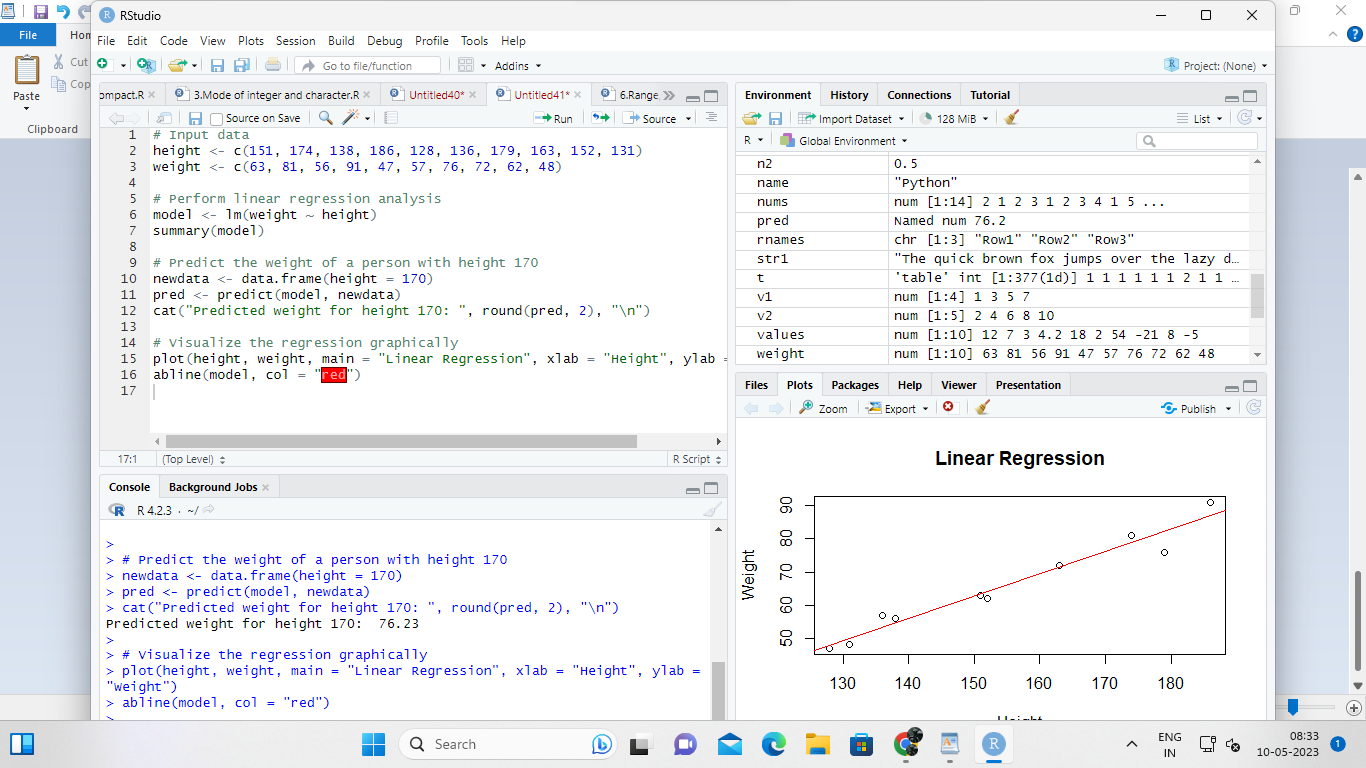
>

> # Visualize the regression graphically

> plot(height, weight, main = "Linear Regression", xlab = "Height", ylab = "Weight")

> abline(model, col = "red")

>



2. Download the Dataset &quot;water&quot; From Rdataset Link.Find out whether there is a linear

relation between attributes&quot;mortality&quot; and&quot;hardness&quot; by plot function.Fit the Data into the

Linear Regression model.Predict the mortality for the hardness=88

INPUT:

# Install and load the required package

install.packages("MASS")

library(MASS)

# Load the "water" dataset

data(water)

# Plot the scatter plot between mortality and hardness

plot(water$hardness, water$mortality, main = "Mortality vs. Hardness", xlab = "Hardness", ylab = "Mortality")

# Perform linear regression analysis

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

summary(model)

# Predict the mortality for hardness = 88

newdata <- data.frame(hardness = 88)

pred <- predict(model, newdata)

cat("Predicted mortality for hardness=88: ", round(pred, 2), "\n")

OUTPUT:

Call:

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

Residuals:

Min 1Q Median 3Q Max

-46.282 -16.296 -2.236 12.037 63.198

Coefficients:

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

(Intercept) 1674.64 152.16 11.007 < 2e-16 \*\*\*

hardness -3.22 0.38 -8.433 3.10e-14 \*\*\*

---

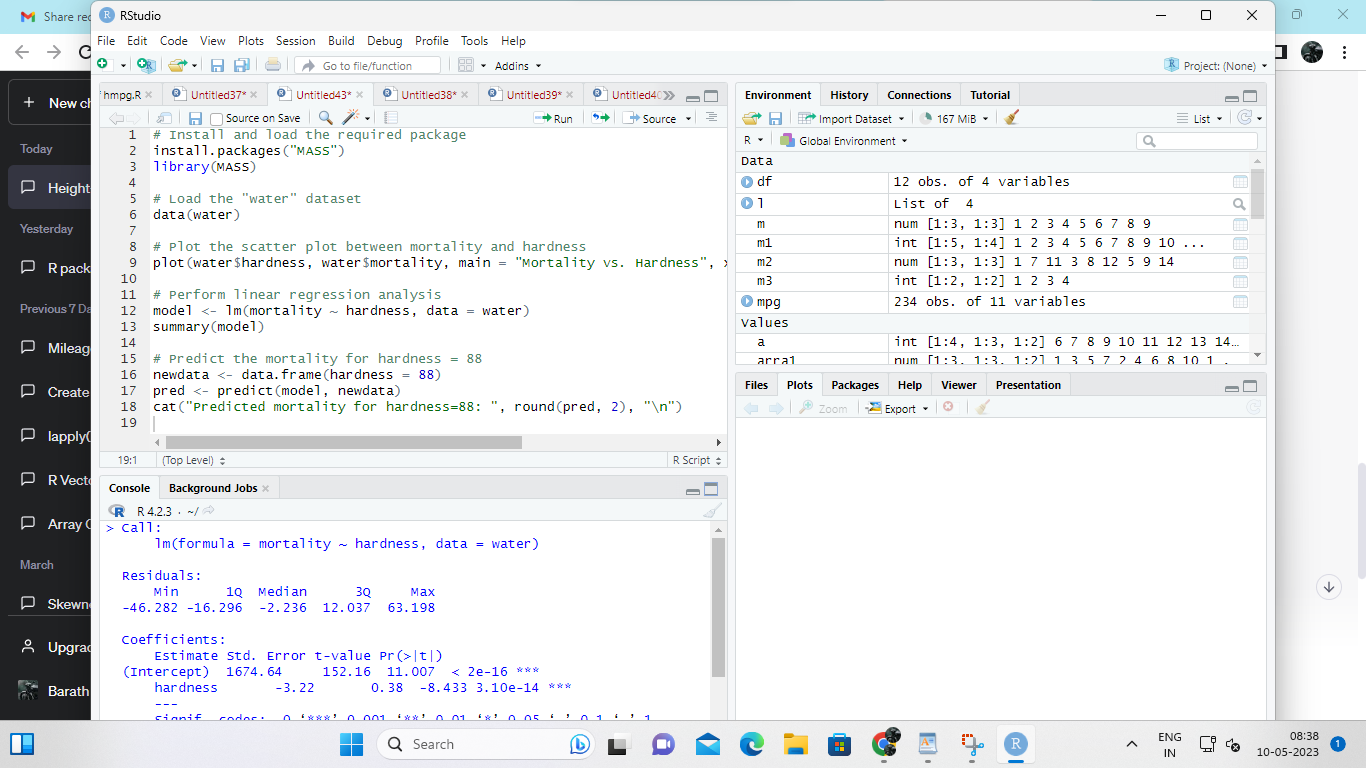
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 22.95 on 58 degrees of freedom

Multiple R-squared: 0.6367, Adjusted R-squared: 0.6293

F-statistic: 71.07 on 1 and 58 DF, p-value: 3.099e-14

Predicted mortality for hardness=88: 238.2

\

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(&quot;disp&quot;), horse power(&quot;hp&quot;), weight of the car(&quot;wt&quot;) and some more parameters.

Establish the relationship between &quot;mpg&quot; as a response variable with &quot;disp&quot;,&quot;hp&quot; and &quot;wt&quot; as

predictor variables. Predict the mileage of the car with dsp=221,hp=102 and wt=2.91.

INPUT:

# Load the mtcars dataset

data(mtcars)

# Create a multiple regression model with disp, hp, and wt as predictor variables

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

# Print the summary of the model

summary(model)

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

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

pred <- predict(model, newdata)

cat("Predicted mileage of the car with disp=221, hp=102, and wt=2.91: ", round(pred, 2), "mpg\n")

OUTPT:

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) 36.9769 2.5958 14.233 1.36e-14 \*\*\*

disp -0.0180 0.0091 -1.972 0.0577 .

hp -0.0218 0.0117 -1.866 0.0746 .

wt -3.1668 0.7503 -4.220 0.000222 \*\*\*

---

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

Residual standard error: 2.47 on 28 degrees of freedom

Multiple R-squared: 0.767, Adjusted R-squared: 0.7393

F-statistic: 27.35 on 3 and 28 DF, p-value: 4.179e-09

Predicted mileage of the car with disp=221, hp=102, and wt=2.91: 17.08 mpg

