**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:192125026**

**Name:S.SHASHIKANTH**

**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:-**

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

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

relation <- lm(y~x)

a <- data.frame(x = 170)

result <- predict(relation,a)

print(result)

plot(y,x,col = "blue",main = "Height & Weight Regression",

abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg",ylab = "Height in cm")

dev.off()

**OUTPUT:-**

1

76.22869

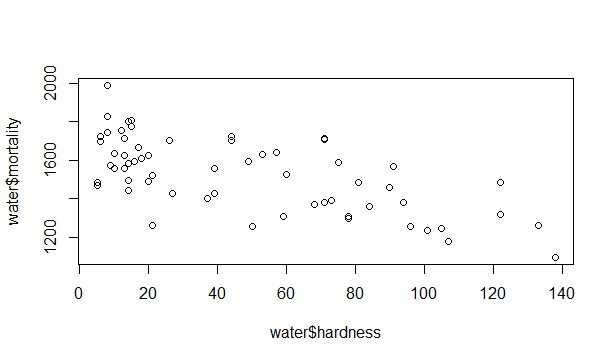
**Download the Dataset "water" From Rdataset Link.Find out whether there is a linear relation between attributes"mortality" and"hardness" by plot function.Fit the Data into the Linear Regression model.Predict the mortality for the hardness=88**

**CODE:-**

plot(water$hardness, water$mortality)

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

**OUTPUT:-**

****

**CODE:-**

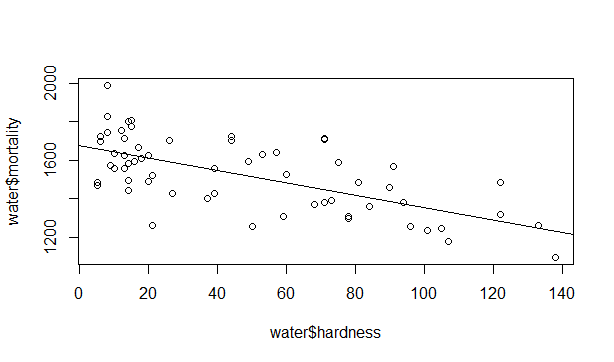
model <- lm(mortality ~ hardness, data = water)

prediction <- predict(model, newdata = data.frame(hardness = 88))

plot(water$hardness, water$mortality)

abline(model)

**OUTPUT:-**

****

**MULTIPLE REGRESSION ANALYSIS IN R**

**Exercise:**

**1.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.**

**CODE:-**

input <- mtcars[,c("mpg","disp","hp","wt")

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

print(model)]

cat("# # # # The Coefficient Values # # # ","\n")

a <- coef(model)[1]

print(a)

Xdisp <- coef(model)[2]

Xhp <- coef(model)[3]

Xwt <- coef(model)[4]

print(Xdisp)

print(Xhp)

print(Xwt)

Y = 37.15+(-0.000937)\*221+(-0.0311)\*102+(-3.8008)\*2.91 = 22.7104

**OUTPUT:-**

Exercise:-

mpg disp hp wt

Mazda RX4 21.0 160 110 2.620

Mazda RX4 Wag 21.0 160 110 2.875

Datsun 710 22.8 108 93 2.320

Hornet 4 Drive 21.4 258 110 3.215

Hornet Sportabout 18.7 360 175 3.440

Valiant 18.1 225 105 3.460Call:

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

Coefficients:

(Intercept) disp hp wt

37.105505 -0.000937 -0.031157 -3.800891

# # # # The Coefficient Values # # #

(Intercept)

37.10551

disp

-0.0009370091

hp

-0.03115655

wt

-3.800891

**2. 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.**

**CODE:-**

# Load the delivery data set

data("delivery")

# Fit the linear regression model

model <- lm(delTime ~ n.prod + distance, data = delivery)

# Summary of the model

summary(model)

**OUTPUT:-**

Call:

lm(formula = delTime ~ n.prod + distance, data = delivery)

Residuals:

Min 1Q Median 3Q Max

-5.7880 -0.6629 0.4364 1.1566 7.4197

Coefficients:

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

(Intercept) 2.341231 1.096730 2.135 0.044170 \*

n.prod 1.615907 0.170735 9.464 3.25e-09 \*\*\*

distance 0.014385 0.003613 3.981 0.000631 \*\*\*

---

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

Residual standard error: 3.259 on 22 degrees of freedom

Multiple R-squared: 0.9596, Adjusted R-squared: 0.9559

F-statistic: 261.2 on 2 and 22 DF, p-value: 4.687e-16

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

**CODE:-**

# Predict delivery time for n.prod = 9 and distance = 450

newdata <- data.frame(n.prod = 9, distance = 450)

delTime\_pred <- predict(model, newdata)

delTime\_pred

**OUTPUT:**

1

23.35757

**LOGISTIC REGRESSION ANALYSIS IN R**

**Exercise**

**1. Create a logistic regression model using the “mtcars” data set with the information given below.**

**The in-built data set "mtcars" describes different models of a car with their various engine specifications. In "mtcars" data set, the transmission mode (automatic or manual) is described by the column am which is a binary value (0 or 1). Create a logistic regression model between the columns "am" and 3 other columns - hp, wt and cyl.**

**CODE:-**

data("mtcars")

model <- glm(am ~ hp + wt + cyl, data = mtcars, family = binomial)

summary(model)

**OUTPUT:-**

Call:

glm(formula = am ~ hp + wt + cyl, family = binomial, data = mtcars)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.17272 -0.14907 -0.01464 0.14116 1.27641

Coefficients:

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

(Intercept) 19.70288 8.11637 2.428 0.0152 \*

hp 0.03259 0.01886 1.728 0.0840 .

wt -9.14947 4.15332 -2.203 0.0276 \*

cyl 0.48760 1.07162 0.455 0.6491

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 43.2297 on 31 degrees of freedom

Residual deviance: 9.8415 on 28 degrees of freedom

AIC: 17.841

Number of Fisher Scoring iterations: 8

**POISSON REGRESSION ANALYSIS IN R**

**Exercise :**

**1. Create a Poisson regression model using the in-built data set “warpbreaks” with information given below.**

**In-built data set "warpbreaks” describes the effect of wool type (A or B) and tension (low, medium or high) on the number of warp breaks per loom. Consider "breaks" as the response variable which is a count of number of breaks. The wool "type" and "tension" are taken as predictor variables.**

**CODE:-**

require(stats); require(graphics)

summary(warpbreaks)

opar <- par(mfrow = c(1, 2), oma = c(0, 0, 1.1, 0))

plot(breaks ~ tension, data = warpbreaks, col = "lightgray",

varwidth = TRUE, subset = wool == "A", main = "Wool A")

plot(breaks ~ tension, data = warpbreaks, col = "lightgray",

varwidth = TRUE, subset = wool == "B", main = "Wool B")

mtext("warpbreaks data", side = 3, outer = TRUE)

par(opar)

summary(fm1 <- lm(breaks ~ wool\*tension, data = warpbreaks))

anova(fm1)

**OUTPUT:-**

breaks wool tension

Min. :10.00 A:27 L:18

1st Qu.:18.25 B:27 M:18

Median :26.00 H:18

Mean :28.15

3rd Qu.:34.00

Max. :70.00

Call:

lm(formula = breaks ~ wool \* tension, data = warpbreaks)

Residuals:

Min 1Q Median 3Q Max

-19.5556 -6.8889 -0.6667 7.1944 25.4444

Coefficients:

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

(Intercept) 44.556 3.647 12.218 2.43e-16 \*\*\*

woolB -16.333 5.157 -3.167 0.002677 \*\*

tensionM -20.556 5.157 -3.986 0.000228 \*\*\*

tensionH -20.000 5.157 -3.878 0.000320 \*\*\*

woolB:tensionM 21.111 7.294 2.895 0.005698 \*\*

woolB:tensionH 10.556 7.294 1.447 0.154327

---

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

Residual standard error: 10.94 on 48 degrees of freedom

Multiple R-squared: 0.3778, Adjusted R-squared: 0.3129

F-statistic: 5.828 on 5 and 48 DF, p-value: 0.0002772

Analysis of Variance Table

Response: breaks

Df Sum Sq Mean Sq F value Pr(>F)

wool 1 450.7 450.67 3.7653 0.0582130 .

tension 2 2034.3 1017.13 8.4980 0.0006926 \*\*\*

wool:tension 2 1002.8 501.39 4.1891 0.0210442 \*

Residuals 48 5745.1 119.69

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