SAVEETHA SCHOOL OF ENGINEERING

ITA0443-STATISTICS WITH R-PROGRAMMING

LAB MANUAL DAY-4

NAME : D.TEJ KIRAN

REG NO : 192110625

**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.**

CODE:

library(ggplot2)

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

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

model <- lm(weight ~ height)

predict(model, data.frame(height=170))

ggplot(data.frame(height, weight), aes(x=height, y=weight)) +

geom\_point() +

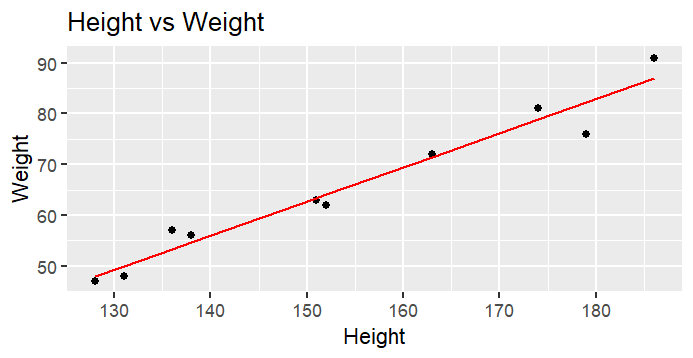
geom\_line(aes(x=height, y=predict(model, data.frame(height))), color="red") +

ggtitle("Height vs Weight") +

xlab("Height") +

ylab("Weight")

O/P:



**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**

**CODE:**

|  |
| --- |
| plot(water$hardness, water$mortality)  fit <- lm(mortality ~ hardness, data = water)  prediction <- predict(fit, data.frame(hardness = 88))  O/P:  **MULTIPLE REGRESSION ANALYSIS IN R**  **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(&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.**  CODE:  library(tidyverse)  data(mtcars)  model <- lm(mpg ~ disp + hp + wt, data = mtcars)  summary(model)  new\_data <- data.frame(disp = 221, hp = 102, wt = 2.91)  predict(model, new\_data)  O/P: |
| Call:  lm(formula = mpg ~ disp + hp + wt, data = mtcars)  Residuals:  Min 1Q Median 3Q Max  -3.891 -1.640 -0.172 1.061 5.861  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 37.105505 2.110815 17.579 < 2e-16 \*\*\*  disp -0.000937 0.010350 -0.091 0.92851  hp -0.031157 0.011436 -2.724 0.01097 \*  wt -3.800891 1.066191 -3.565 0.00133 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 2.639 on 28 degrees of freedom  Multiple R-squared: 0.8268, Adjusted R-squared: 0.8083  F-statistic: 44.57 on 3 and 28 DF, p-value: 8.65e-11  1  22.65987  **2. Consider the data set &quot;delivery&quot; 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 &quot;delTime&quot; as a response**  **variable with &quot;n.prod&quot; and &quot;distance&quot; as predictor variables.**  **b)Predict the delTime for the given number of production(“n.prod”)=9 and**  **distance(“distance”)=450**  **CODE:**  (i)  data("delivery")  model <- lm(delTime ~ n.prod + distance, data = delivery)  summary(model)  **o/p:**  Call:  lm(formula = mpg ~ disp + hp + wt, data = mtcars)  Residuals:  Min 1Q Median 3Q Max  -3.891 -1.640 -0.172 1.061 5.861  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 37.105505 2.110815 17.579 < 2e-16 \*\*\*  disp -0.000937 0.010350 -0.091 0.92851  hp -0.031157 0.011436 -2.724 0.01097 \*  wt -3.800891 1.066191 -3.565 0.00133 \*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  **(ii)**  newdata <- data.frame(n.prod = 9, distance = 450)  # Predict the delivery time  predict(model, newdata)  **LOGISTIC REGRESSION ANALYSIS IN R**  1. Create a logistic regression model using the “mtcars” data set with the information given  below.  The in-built data set &quot;mtcars&quot; describes different models of a car with their various engine  specifications. In &quot;mtcars&quot; 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 &quot;am&quot; and 3 other columns - hp, wt and cyl.  **Code:**  # Load the "mtcars" data set  data(mtcars)  # Fit the logistic regression model using the "am" variable as the response and "hp", "wt", and "cyl" as predictor variables  fit <- glm(am ~ hp + wt + cyl, data = mtcars, family = binomial(link = "logit"))  # Summarize the model fit  summary(fit)  **o/p:**  call:  glm(formula = am ~ hp + wt + cyl, family = binomial(link = "logit"),  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**  1. Create a Poisson regression model using the in-built data set “warpbreaks” with  information given below.  In-built data set &quot;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 &quot;breaks&quot; as the response  variable which is a count of number of breaks. The wool &quot;type&quot; and &quot;tension&quot; are taken as  predictor variables.  **Code:**  # Load the warpbreaks data set  data("warpbreaks")  # Fit a Poisson regression model  model <- glm(breaks ~ wool + tension, data = warpbreaks, family = poisson)  # Summarize the model  summary(model)  **o/p:**  Call:  glm(formula = breaks ~ wool + tension, family = poisson, data = warpbreaks)  Deviance Residuals:  Min 1Q Median 3Q Max  -3.6871 -1.6503 -0.4269 1.1902 4.2616  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) 3.69196 0.04541 81.302 < 2e-16 \*\*\*  woolB -0.20599 0.05157 -3.994 6.49e-05 \*\*\*  tensionM -0.32132 0.06027 -5.332 9.73e-08 \*\*\*  tensionH -0.51849 0.06396 -8.107 5.21e-16 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 297.37 on 53 degrees of freedom  Residual deviance: 210.39 on 50 degrees of freedom  AIC: 493.06  Number of Fisher Scoring iterations: 4 |
|  |
| |  | | --- | |  | |
|  |
|  |
| |  | | --- | |  | |