Multiple Linear Regression

PRANAVCHENDUR T K

15BCE1097

Faculty: ARUN PRASATH G M

Aim:

Model Fitting and investigation of relationships between more than two variables within a regression framework

Applying multiple linear regression model to real dataset; computing and interpreting the multiple coefficient of determination

Program:

Problem 1) The sale of a product in lakhs of rupees (Y) is expected to be influenced by two variables namely the advertising expenditure X1 (in '000Rs) and the number of sales persons (X2) in a region. Sample data on 8 Regions of a state has given the following results :

Area	Y	X1	X2
1	110	30	11
2	80	40	10
3	70	20	7
4	120	50	15
5	150	60	19
6	90	40	12
7	70	20	8
8	120	60	14

```
> Y=c(110,80,70,120,150,90,70,120)

> X1=c(30,40,20,50,60,40,20,60)

> X2=c(11,10,7,15,19,12,8,14)
```

```
> input_data = data.frame(Y,X1,X2)

> input_data
    Y X1 X2
1 110 30 11
2  80 40 10
3  70 20  7
4 120 50 15
5 150 60 19
6  90 40 12
7  70 20  8
8 120 60 14

> regModel<-lm(Y~X1+X2,data=input_data)

> regModel

Call:
lm(formula = Y ~ X1 + X2, data = input_data)

Coefficients:
(Intercept)           X1           X2
    16.8314      -0.2442       7.8488


> lm(formula = Y~X1+X2,data=input_data)

Call:
lm(formula = Y ~ X1 + X2, data = input_data)

Coefficients:
(Intercept)           X1           X2
    16.8314      -0.2442       7.8488


> summary(regModel)

Call:
lm(formula = Y ~ X1 + X2, data = input_data)

Residuals:
      1       2       3       4       5       6       7       8
 14.157  -5.552   3.110  -2.355  -1.308 -11.250  -4.738   7.936

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)  16.8314    11.8290   1.423   0.2140
X1          -0.2442     0.5375  -0.454   0.6687
X2           7.8488     2.1945   3.577   0.0159 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.593 on 5 degrees of freedom
Multiple R-squared:  0.9191,  Adjusted R-squared:  0.8867
F-statistic:  28.4 on 2 and 5 DF,  p-value: 0.001862
```

Problem 2)

```
> Ex3d <- read.csv("~/15BCE97/03-21 R PC/Ex3d.csv")

> View(Ex3d)

> regmodel<-lm(Productive.Capital~Number.of.Factories+No..of.Employees,data=Ex3d)

> regmodel

Call:
lm(formula = Productive.Capital ~ Number.of.Factories + No..of.Employees,
    data = Ex3d)

Coefficients:
        (Intercept)  Number.of.Factories     No..of.Employees
          1767.6100              -1.9790               0.1885
```

Problem 3) The table below found in the built-in dataset stackloss in R, shows observations of 21 Days of “operation of a plant for the oxidation of ammonia to nitric acid”, where airflow is the “ow of cooling air”, Water Temp is “the temperature of cooling water”, Acid Coc. is the”Concentration of acid [per 1000, minus 500]”, and stack.loss is “an (inverse) measure of the overall efficiency of the plant”. Find the equations that best predict stack.loss.

```
> stackloss
   Air.Flow Water.Temp Acid.Conc. stack.loss
1        80         27         89         42
2        80         27         88         37
3        75         25         90         37
4        62         24         87         28
5        62         22         87         18
6        62         23         87         18
7        62         24         93         19
8        62         24         93         20
9        58         23         87         15
```

10	58	18	80	14
11	58	18	89	14
12	58	17	88	13
13	58	18	82	11
14	58	19	93	12
15	50	18	89	8
16	50	18	86	7
17	50	19	72	8
18	50	19	79	8
19	50	20	80	9
20	56	20	82	15
21	70	20	91	15

```
> par(mex=0.5)

> pairs(stackloss,gap=0,cex.labels = 1,col="blue")
```