Experiment - 4

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

Aim: To understand the multiple linear regression model with computation and interpretation using R

Introduction

Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y.
Multiple linear regression models are defined by the equation

$$Y=\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta pXp+\epsilon$$

Procedure:

- Import the data set
- Determine the multiple linear regression using R functions
- Visualize the multiple linear regression using R functions

Note: Please make sure that the following package is already installed.

"Scatterplot 3d"

Codes and Results:

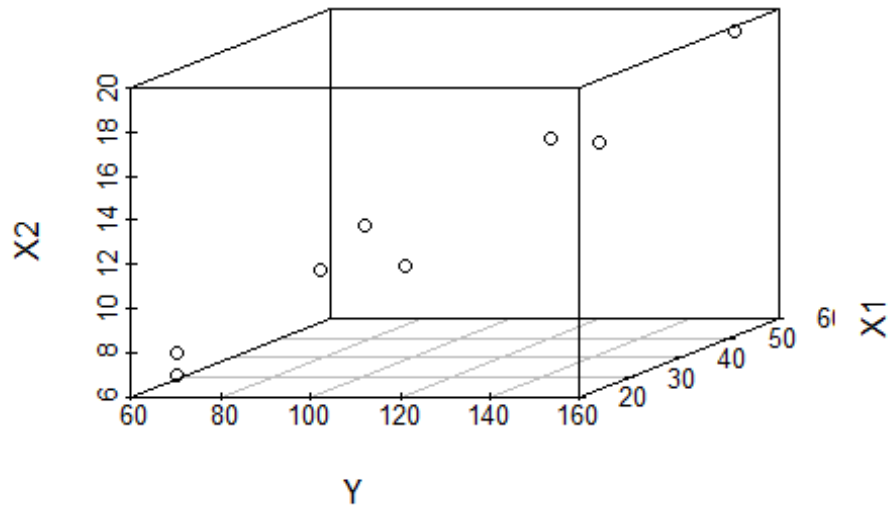
Problem 1: *The sale of a Product in lakhs of rupees(Y) is expected to be influenced by two variables namely the advertising expenditure X1 (in'OOORs) and the number of sales persons(X2) in a region. Sample data on 8 Regions of a state has given the following results*

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

```
# Input the variables
Y=c(110,80,70,120,150,90,70,120)
Y

## [1] 110  80  70 120 150  90  70 120

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

## [1] 30 40 20 50 60 40 20 60

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

## [1] 11 10  7 15 19 12  8 14

# Linear regression model of Y on X1 and X2
RegModel=lm(Y~X1+X2)
RegModel

##
## Call:
## lm(formula = Y ~ X1 + X2)
##
## Coefficients:
## (Intercept)           X1           X2
##     16.8314      -0.2442       7.8488

# Summary of the data
summary(RegModel)

##
## Call:
## lm(formula = Y ~ X1 + X2)
```

```
##
## 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
```

```
# install.packages("scatterplot3d")
library(scatterplot3d)
# Plot the data set
scatterplot3d(Y,X1,X2)
```



Interpretation :

Now the regression the regression model is

$$Y = 16.834 - 0.2442 * X1 + 7.8488 * X2$$

Since R^2 is 0.9593 and the ANOVA shows that the F-ratio is significant, this model can be taken as good-fit in explaining the sales in terms of the other two variables

```
#Problem 2
data=mtcars
data
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

```
X=mtcars$mpg
X
```

```
##  [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2
10.4
```

```
## [16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8
19.7
## [31] 15.0 21.4
```

```
Y=mtcars$disp
Y
```

```
##  [1] 160.0 160.0 108.0 258.0 360.0 225.0 360.0 146.7 140.8 167.6 167.6
275.8
## [13] 275.8 275.8 472.0 460.0 440.0  78.7  75.7  71.1 120.1 318.0 304.0
350.0
## [25] 400.0  79.0 120.3  95.1 351.0 145.0 301.0 121.0
```

```
Z=mtcars$hp
Z
```

```
##  [1] 110 110  93 110 175 105 245  62  95 123 123 180 180 180 205 215 230
66  52
## [20]  65  97 150 150 245 175  66  91 113 264 175 335 109
```

```
RegModel<- lm(Z~X+Y)
RegModel
```

```
##
## Call:
## lm(formula = Z ~ X + Y)
##
## Coefficients:
## (Intercept)            X            Y
##    172.2204      -4.2732       0.2614
```

```
summary(RegModel)
```

```
##
## Call:
## lm(formula = Z ~ X + Y)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -48.70 -17.67 -10.16  10.12 148.19
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 172.2204    69.9014   2.464   0.0199 *
## X            -4.2732     2.3027  -1.856   0.0737 .
## Y             0.2614     0.1120   2.335   0.0267 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41.01 on 29 degrees of freedom
## Multiple R-squared:  0.6653, Adjusted R-squared:  0.6423
## F-statistic: 28.83 on 2 and 29 DF,  p-value: 1.279e-07
```

```
library(scatterplot3d)
graph=scatterplot3d(X,Y,Z)
# Visualize the plane
graph$plane3d(RegModel)
```

Conclusion:

Multiple linear regression model has been explored and visualized.