FACTORS AFFECTING THE MARKS OF STUDENTS

By SHUBHAM Umesh RAUT

Roll no= 11

OBJECTIVE:

- To study the factors affecting the marks of students in 10th and 12th grade using multivariate regression technique.
- To find the correlation between the study variables & explanatory variables.
- To visualize the data obtained graphically and to draw various conclusions from it.

METHODOLOGY:

First the independent variables were decided that might affect the marks of students. Next, a sample is obtained for the survey. The sample data used for the purpose of survey is a primary data obtained using questionnaire method. The questionnaire was conducted using google forms. The analysis of the data is done using the R-software. The questions asked were:

1) Marks obtained in 10th (In %)
2) Marks obtained in 12th (In %)
3) Time spent in school or college per day(In hours)
4) Time spent in coaching class/ tuition class per day (in hours)
5) Time given for self study per day (in hours)
6) Time given for relaxation per day (in hours) (For e.g. Reading , Listening music , Social media, etc.)
7) Time given for physical exercises per day (in hours)
8) Time given for extra events per week (in hours) (For e.g. Outing with friends, Family functions, etc)

The sample obtained is of size 79. Hence sample size, n= 79

DATA CLEANING:

In data cleaning, the first step done was to convert the questions into variables. All the variables here are quantitative variables. The two dependent variables are:

Y1: Marks obtained in 10th (In %)

Y2: Marks obtained in 12th (In %)

The six independent variables are:

X1: Time spent in school or college per day(In hours)

X2: Time spent in coaching class/ tuition class per day (in hours)

X3: Time given for self study per day (in hours)

X4: Time given for relaxation per day (in hours)

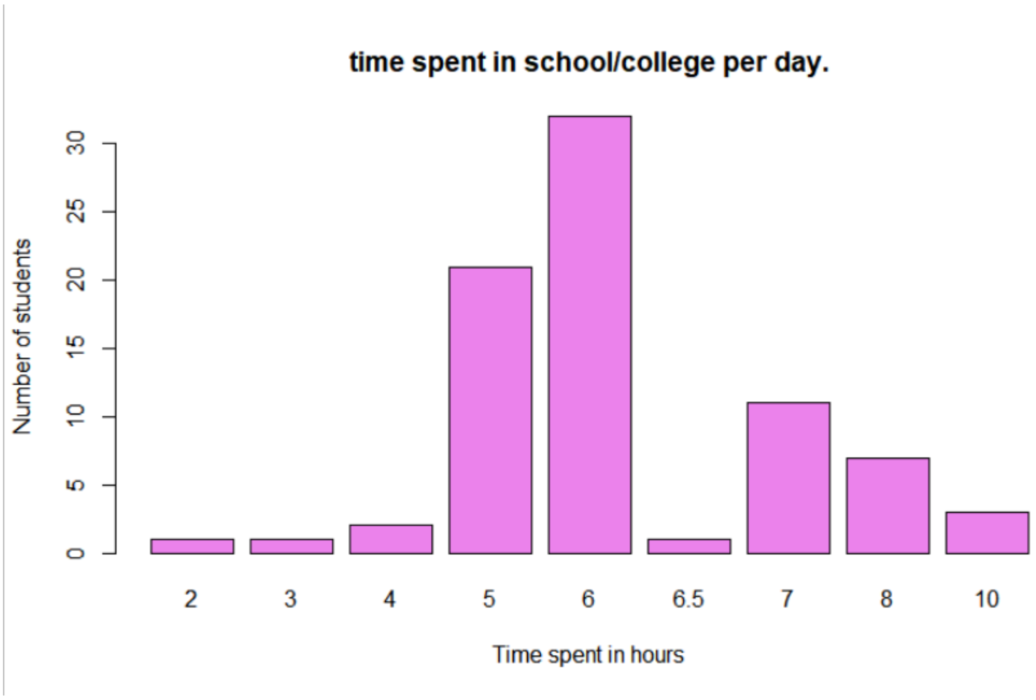
X5: Time given for physical exercises per day (in hours)

X6: Time given for extra events per week (in hours)

Next step is to check for missing values. It was first done manually and then using R-software. There are no missing values in the data. Hence we proceed for analysis.

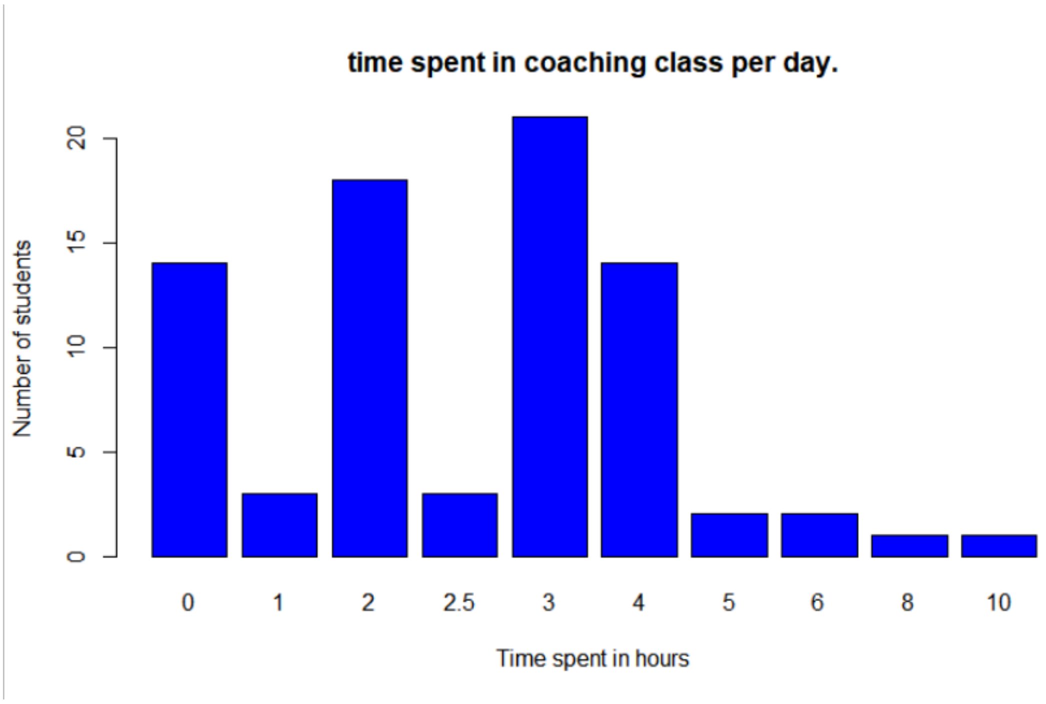
DATA VISUALIZATION:

> barplot(table(my_data$X1),xlab="Time spent in hours",ylab="Number of students", col="violet", main="time spent in school/college per day.")



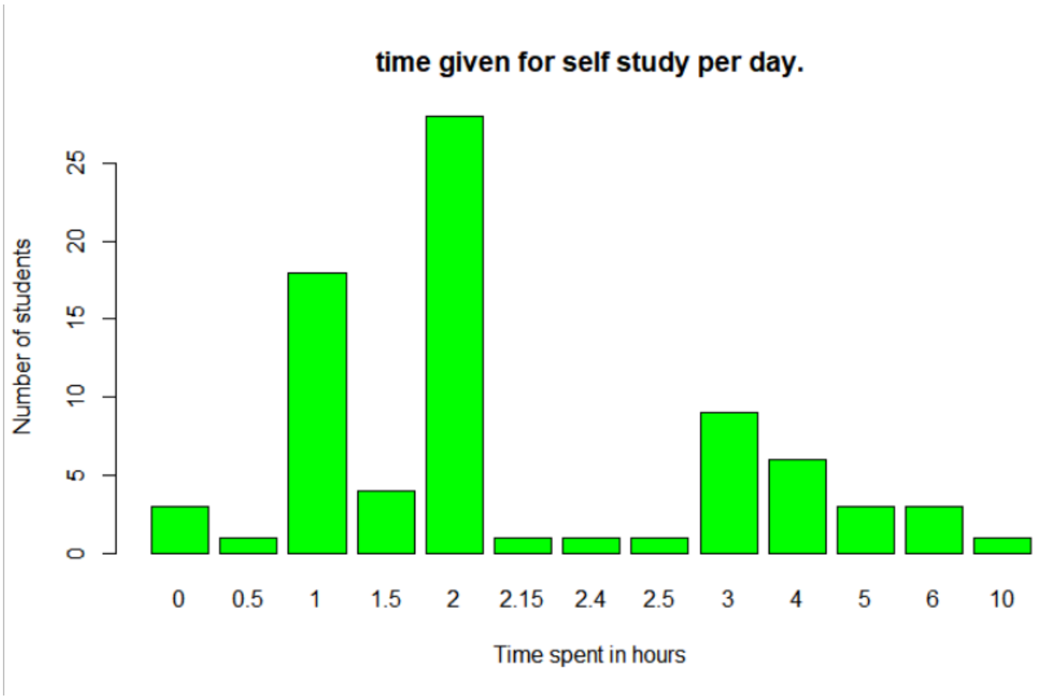
The above bar graph depicts that there are 30 students out of 79 who spent on an average 6 hours in school/college per day while only 3 students spent average 10 hours in school/college per day. Also, there are 10 students spending 7 hours of their day at school/college.

> barplot(table(my_data$X2),xlab="Time spent in hours",ylab="Number of students", col="blue", main="time spent in coaching class per day.")



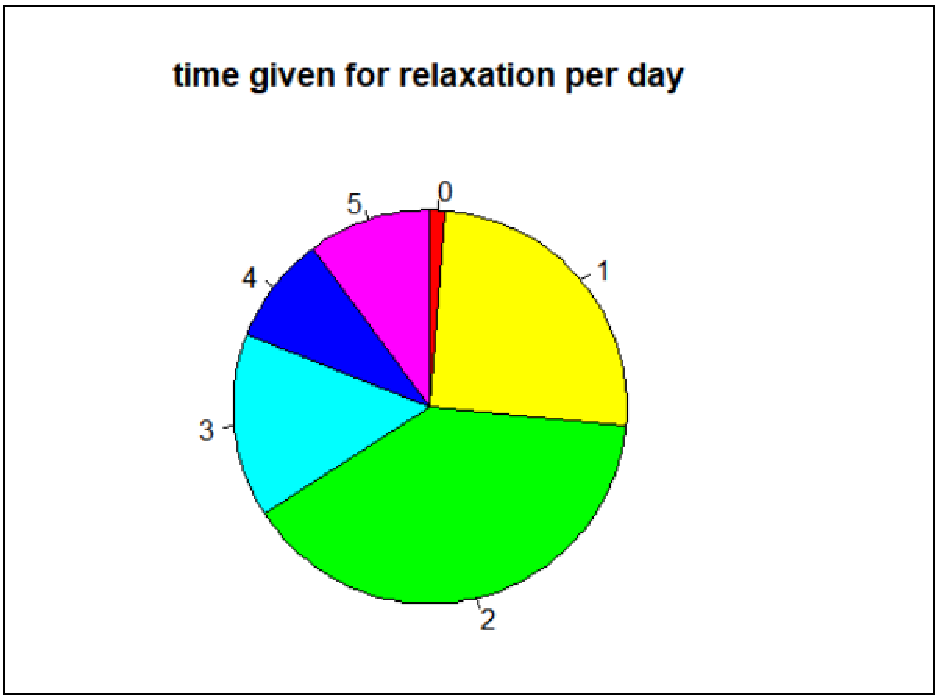
The above graph depicts that 14 students spent 0 hours in coaching classes while 20 students spent on an average 3 hours daily in coaching classes. We can also see that 2 out of 79 students went to coaching classes for 6 hours daily.

> barplot(table(my_data$X3),xlab="Time spent in hours",ylab="Number of students", col="green", main="time given for self study per day.")



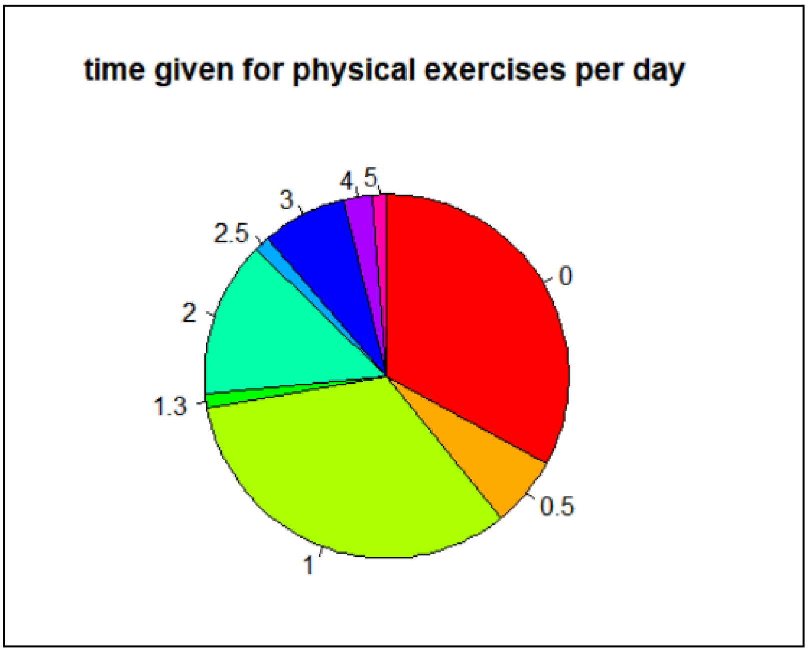
In this graph, we see 27 students giving 2 hours of their day for self-study and 4 students giving 1.5 hours per day. On the other hand, 2 students out of 79 gave on an average 6 hours of their day for self-study.

> pie(table(my_data$X4),main="time given for relaxation per day", col=rainbow(length(table(my_data$X4))), clockwise=TRUE)



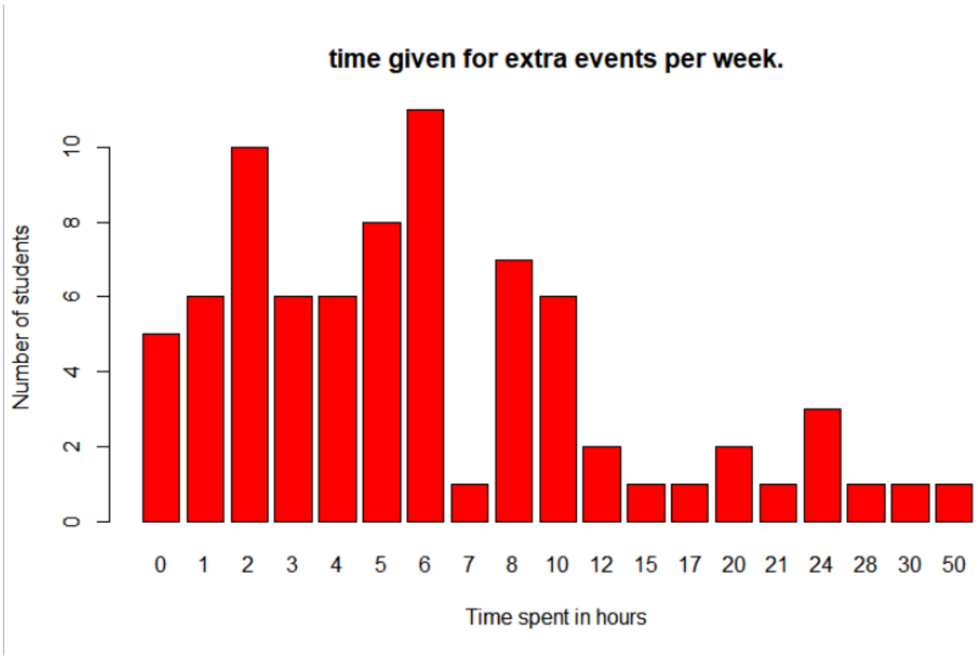
From the above pie diagram, we see maximum students prefer to give 2 hours of their day for relaxation which includes perusing hobbies and social media. While very less students give no time for relaxation. This shows us that relaxation is a vital part during studies as it calms the mind.

> pie(table(my_data$X5),main="time given for physical exercises per day", col=rainbow(length(table(my_data$X5))), clockwise=TRUE)



The above pie chart visualizes the time spent for physical exercises per day.

> barplot(table(my_data$X6),xlab="Time spent in hours",ylab="Number of students", col="red", main="time given for extra events per week.")



The above diagram gives a great description about the time given for extra social events per week. It ranges from 0-50 hours. Maximum i.e. 12 student spent 6 hours per week on an average for extra events. The next highest is 9 students spending 2 hours per week for social events.

ANALYSIS:

- IMPORTING DATASET IN R

>library(readxl)
>my_data <- read_excel("C:/Users/harsh/OneDrive/Desktop/M. Sc/WORK/Paper 2/SEM 2/dataset of paper 2.xlsx")
>head(my_data)

```
> head(my_data)
# A tibble: 6 × 8
     Y1    Y2    X1    X2    X3    X4    X5    X6
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  73.2  85.5     5   0     2       2   2       4
2  95    92       6   4     1.5     2   2.5     6
3  77    51       6   2.5   1       1   3       5
4  95    88      10   0     1       5   0       5
5  90    60       5   5     1       1   0       2
6  76    59       8   6     3       2   1      24
```

- CHECKING MISSING VALUES

```
> sum(is.na(my_data))
[1] 0
```

- PARTIAL CORRELATION MATRIX

>install.packages("ppcor")
>library(ppcor)

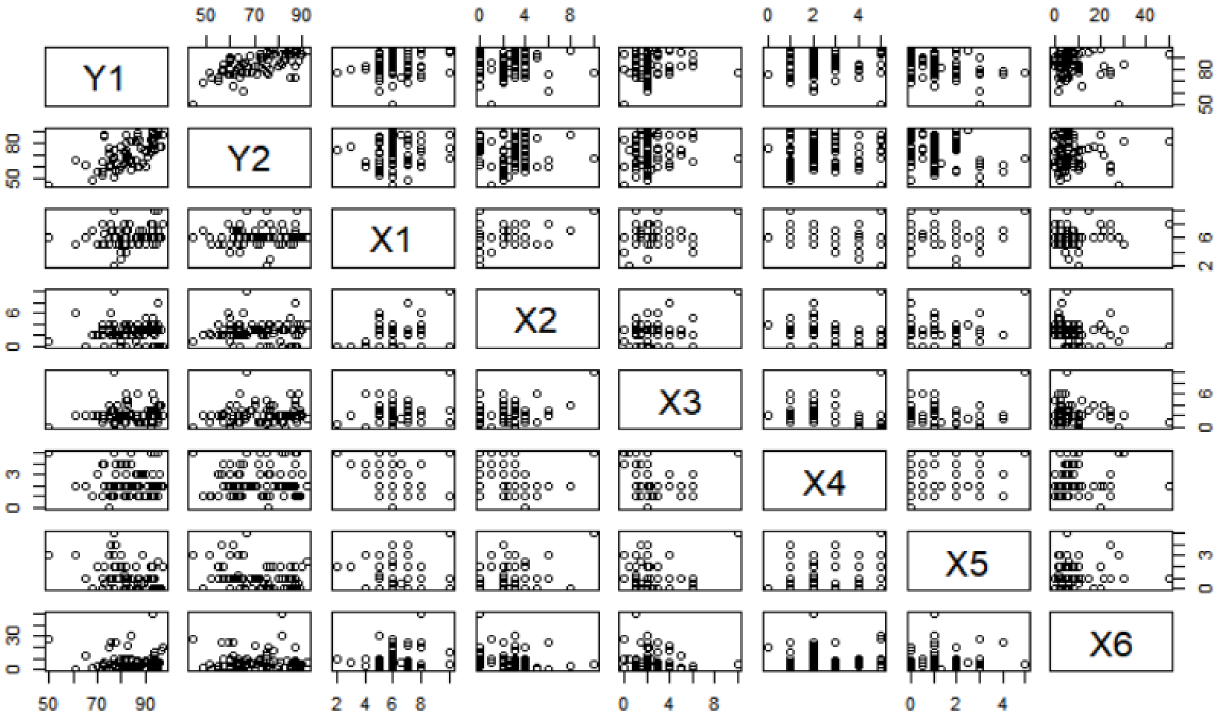
```
> pcor(my_data)$estimate
            Y1          Y2          X1         X2          X3          X4          X5          X6
Y1  1.00000000  0.59216931  0.08643813  0.1083257 -0.02101029 -0.02741925 -0.23357915  0.02865989
Y2  0.59216931  1.00000000  0.06638745 -0.1459437  0.16776020  0.13436733 -0.11549700 -0.03780373
X1  0.08643813  0.06638745  1.00000000  0.1348030  0.17832154 -0.12741116  0.01550496  0.25006243
X2  0.10832569 -0.14594368  0.13480302  1.0000000  0.27021764 -0.10226463  0.13248464 -0.22531787
X3 -0.02101029  0.16776020  0.17832154  0.2702176  1.00000000 -0.09915971  0.16340383 -0.10158904
X4 -0.02741925  0.13436733 -0.12741116 -0.1022646 -0.09915971  1.00000000  0.24882575  0.09166261
X5 -0.23357915 -0.11549700  0.01550496  0.1324846  0.16340383  0.24882575  1.00000000  0.20008500
X6  0.02865989 -0.03780373  0.25006243 -0.2253179 -0.10158904  0.09166261  0.20008500  1.00000000
```

- MULTIPLE CORRELATION MATRIX

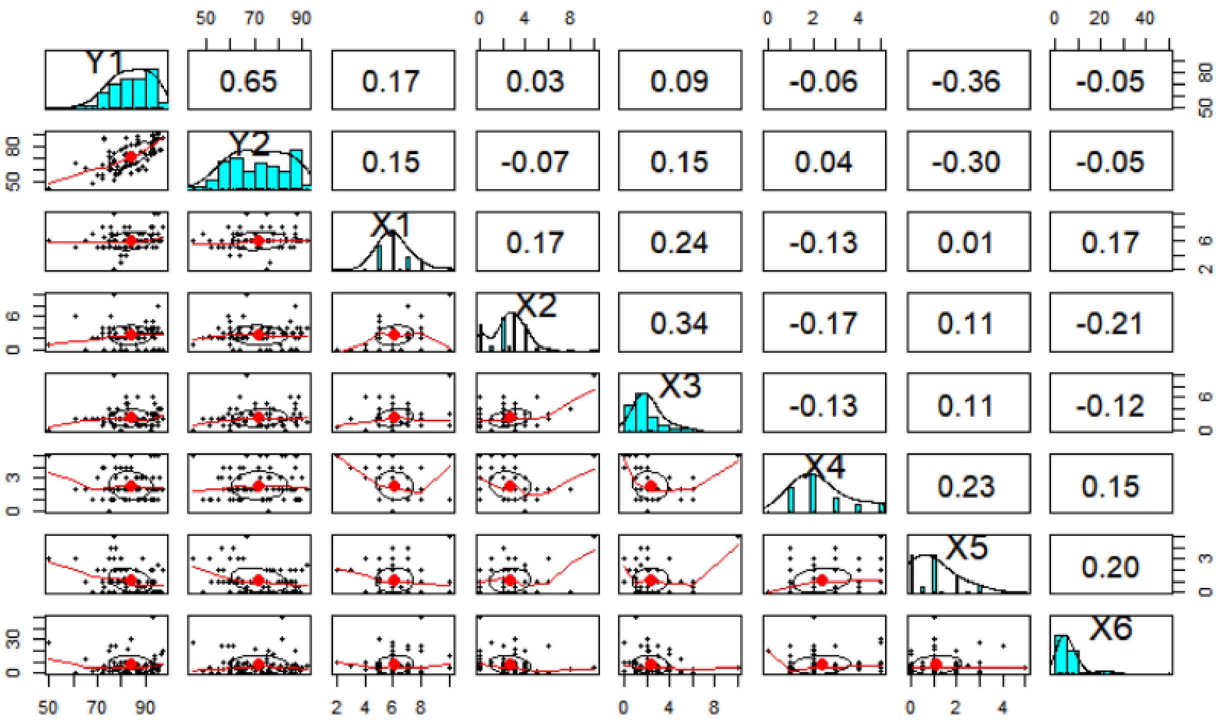
```
> cor(my_data)
            Y1          Y2           X1          X2          X3          X4           X5          X6
Y1  1.00000000  0.64614944  0.172513253  0.03369130  0.09407933 -0.06407799 -0.364702436 -0.05244174
Y2  0.64614944  1.00000000  0.149061511 -0.06630044  0.14536681  0.03582817 -0.296092101 -0.04616088
X1  0.17251325  0.14906151  1.000000000  0.16880589  0.24216121 -0.13468780  0.008604148  0.17423115
X2  0.03369130 -0.06630044  0.168805888  1.00000000  0.34082769 -0.16813751  0.105015267 -0.21356568
X3  0.09407933  0.14536681  0.242161215  0.34082769  1.00000000 -0.13172445  0.107254133 -0.11520025
X4 -0.06407799  0.03582817 -0.134687796 -0.16813751 -0.13172445  1.00000000  0.227212404  0.15359129
X5 -0.36470244 -0.29609210  0.008604148  0.10501527  0.10725413  0.22721240  1.000000000  0.20092756
X6 -0.05244174 -0.04616088  0.174231150 -0.21356568 -0.11520025  0.15359129  0.200927562  1.00000000
```

- CHECKING LINEARITY USING SCATTER PLOT MATRIX:

```
> pairs(my_data)
```




```
>install.packages("psych")
>library(psych)
>pairs.panels(my_data, method="pearson",density= TRUE, ellipses = TRUE)
```



1. MODEL BUILDING USING TWO MLR MODELS:

- MODEL BUILDING BEFORE FORWARD SELECTION METHOD:

```
> z1= lm(Y1~ X1+X2+X3+X4+X5+X6, data=my_data)
> z1

Call:
lm(formula = Y1 ~ X1 + X2 + X3 + X4 + X5 + X6, data = my_data)

Coefficients:
(Intercept)          X1          X2          X3          X4          X5          X6
  78.104733    1.048584    0.138947    0.565051    0.452553   -3.217113    0.008479
```

```
> z2= lm(Y2~X1+X2+X3+X4+X5+X6, data=my_data)
> z2

Call:
lm(formula = Y2 ~ X1 + X2 + X3 + X4 + X5 + X6, data = my_data)

Coefficients:
(Intercept)          X1          X2          X3          X4          X5          X6
   63.44846     1.29714    -0.68656     1.48237     1.35635    -3.57769    -0.03725
```

- FORWARD SELECTION PROCEDURE FOR Y1:

```
> newdata= my_data[,-2]
> head(newdata)
# A tibble: 6 x 7
     Y1    X1    X2    X3    X4    X5    X6
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  73.2     5   0     2       2   2       4
2  95       6   4     1.5     2   2.5     6
3  77       6   2.5   1       1   3       5
4  95      10   0     1       5   0       5
5  90       5   5     1       1   0       2
6  76       8   6     3       2   1      24
```

```
> intercept_only= lm(Y1~1, data=newdata)
> forward= step(intercept_only, direction= "forward", scope= formula(z1))
Start:  AIC=350.31
Y1 ~ 1

       Df Sum of Sq    RSS    AIC
+ X5    1    863.58 5629.1 341.03
+ X1    1    193.23 6299.5 349.92
<none>              6492.7 350.31
+ X3    1     57.47 6435.3 351.61
+ X4    1     26.66 6466.1 351.99
+ X6    1     17.86 6474.9 352.09
+ X2    1      7.37 6485.4 352.22
```

```
Step:  AIC=341.03
Y1 ~ X5

       Df Sum of Sq    RSS    AIC
+ X1    1   200.337 5428.8 340.17
<none>              5629.1 341.03
+ X3    1   116.528 5512.6 341.38
+ X2    1    34.025 5595.1 342.56
+ X6    1     2.938 5626.2 342.99
+ X4    1     2.416 5626.7 343.00
```

```
Step:  AIC=340.17
Y1 ~ X5 + X1

       Df Sum of Sq    RSS    AIC
<none>              5428.8 340.17
+ X3    1    57.564 5371.2 341.33
+ X4    1    12.787 5416.0 341.99
+ X2    1    12.204 5416.6 341.99
+ X6    1     0.626 5428.2 342.16
> forward

Call:
lm(formula = Y1 ~ X5 + X1, data = newdata)

Coefficients:
(Intercept)           X5           X1
     79.763       -2.979        1.193
```

Hence the fitted model built for Y1 after forward selection method is:
Y1= 79.763 – 2.979X5 + 1.193X1

- FORWARD SELECTION PROCEDURE FOR Y2:

```
> newdata1= my_data[,-1]
> head(newdata1)
# A tibble: 6 x 7
     Y2    X1    X2    X3    X4    X5    X6
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1  85.5     5   0     2       2   2       4
2  92       6   4     1.5     2   2.5     6
3  51       6   2.5   1       1   3       5
4  88      10   0     1       5   0       5
5  60       5   5     1       1   0       2
6  59       8   6     3       2   1      24
```

```
> intercept_only1= lm(Y2~1, data=newdata1)
> forward1= step(intercept_only1, direction= "forward", scope= formula(z2))
Start:  AIC=393.24
Y2 ~ 1

       Df Sum of Sq   RSS    AIC
+ X5    1    980.11 10199 387.99
<none>              11180 393.24
+ X1    1    248.40 10931 393.46
+ X3    1    236.24 10943 393.55
+ X2    1     49.14 11130 394.89
+ X6    1     23.82 11156 395.07
+ X4    1     14.35 11165 395.14
```

```
Step:  AIC=387.99
Y2 ~ X5

       Df Sum of Sq     RSS    AIC
+ X3    1    354.82  9844.6 387.19
+ X1    1    256.98  9942.4 387.97
<none>              10199.4 387.99
+ X4    1    125.31 10074.1 389.01
+ X2    1     14.01 10185.4 389.88
+ X6    1      2.07 10197.3 389.97
```

```
Step:  AIC=387.19
Y2 ~ X5 + X3

       Df Sum of Sq    RSS    AIC
<none>              9844.6 387.19
+ X4    1   207.927 9636.7 387.51
+ X1    1   139.550 9705.0 388.07
+ X2    1   112.989 9731.6 388.28
+ X6    1    17.013 9827.6 389.06
> forward1

Call:
lm(formula = Y2 ~ X5 + X3, data = newdata1)

Coefficients:
(Intercept)           X5           X3
     72.538       -3.366        1.344
```

Hence the fitted model built for Y2 after forward selection method is:

Y2= 72.538 – 3.366X5 + 1.344X3

- FITTING MODEL AFTER VARIABLE SELECTION:

```
> model1= lm(Y1~X1 + X5, data=my_data)
> model1

Call:
lm(formula = Y1 ~ X1 + X5, data = my_data)

Coefficients:
(Intercept)           X1           X5
     79.763        1.193       -2.979

> model2= lm(Y2~X3 + X5, data=my_data)
> model2

Call:
lm(formula = Y2 ~ X3 + X5, data = my_data)

Coefficients:
(Intercept)           X3           X5
     72.538        1.344       -3.366
```

Thus the 2 models fitted are:

Y1= 79.763 – 2.979X5 + 1.193X1

Y2= 72.538 – 3.366X5 + 1.344X3

- TESTING SIGNIFICANCE OF PARAMETERS OF MODEL 1:

```
> anova(model1)
Analysis of Variance Table

Response: Y1
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1  193.2  193.23  2.7051 0.1041589
X5         1  870.7  870.69 12.1892 0.0008041 ***
Residuals 76 5428.8   71.43
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$H_{o:}$ $\beta_1 = \beta_2 = 0$

H_1: At least one is non zero

Interpretation: p-value of X1= 0.1041589 > 0.05 and p-value of X5= 0.0008<0.05, hence we do not reject H_o and conclude that X1 is insignificant while X5 is significant at 5% l.o.s.

```
> summary(model1)

Call:
lm(formula = Y1 ~ X1 + X5, data = my_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-27.980  -3.281   1.253   6.181  15.530 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  79.7633     4.5130  17.674  < 2e-16 ***
X1            1.1926     0.7121   1.675 0.098105 .  
X5           -2.9795     0.8534  -3.491 0.000804 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.452 on 76 degrees of freedom
Multiple R-squared:  0.1639,	Adjusted R-squared:  0.1419 
F-statistic: 7.447 on 2 and 76 DF,  p-value: 0.001113
```

$H_{o1:}$ $\beta_1 = 0$ against H_{11} : $\beta_1 \neq 0$
H_{o2} : $\beta_2 = 0$ against H_{12} : $\beta_2 \neq 0$
Interpretation: Since p-value of X1 is 0.098 >0.05, we reject H_{01} at 5% l.o.s. and conclude that β_1 is individually insignificant. p-value of X2 is 0.0008<0.05, we do not reject H_{02} at 5% l.o.s. and conclude that β_2 is individually significant.
<u>R- squared obtained in 16.4%</u>

- TESTING SIGNIFICANCE OF PARAMETERS OF MODEL 2:

```
> anova(model2)
Analysis of Variance Table

Response: Y2
          Df Sum Sq Mean Sq F value   Pr(>F)   
X3         1  236.2  236.24  1.8238 0.180873   
X5         1 1098.7 1098.69  8.4819 0.004707 **
Residuals 76 9844.6  129.53                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$H_{o:}$ $\beta_1 = \beta_2 = 0$
H_1: At least one is non zero
Interpretation: p-value of X3= 0.180873 > 0.05 and p-value of X5= 0.004707<0.05, hence we do not reject H_o and conclude that X3 is insignificant while X5 is significant at 5% l.o.s.

```
> summary(model2)

Call:
lm(formula = Y2 ~ X3 + X5, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-27.2266  -8.7075  -0.8604  10.0645  25.8610 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  72.5383     2.4862  29.176  < 2e-16 ***
X3            1.3442     0.8122   1.655  0.10204    
X5           -3.3662     1.1558  -2.912  0.00471 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.38 on 76 degrees of freedom
Multiple R-squared:  0.1194,	Adjusted R-squared:  0.09624 
F-statistic: 5.153 on 2 and 76 DF,  p-value: 0.007969
```

$H_{o1:}$ β_1 =0 against H_{11} : $\beta_1 \neq 0$

H_{o2} : β_2 =0 against H_{12} : $\beta_2 \neq 0$

Interpretation: Since p-value of X3 is 0.10204 >0.05, we reject H_{01} at 5% l.o.s. and conclude that β_1 is individually insignificant. p-value of X2 is 0.00471<0.05, we do not reject H_{02} at 5% l.o.s. and conclude that β_2 is individually significant.

R- squared obtained in 11.94%

2. CHECKING THE ASSUMPTIONS OF REGRESSION:

- Normality-

H_0: Errors are normally distributed

H_1 : Not H_0

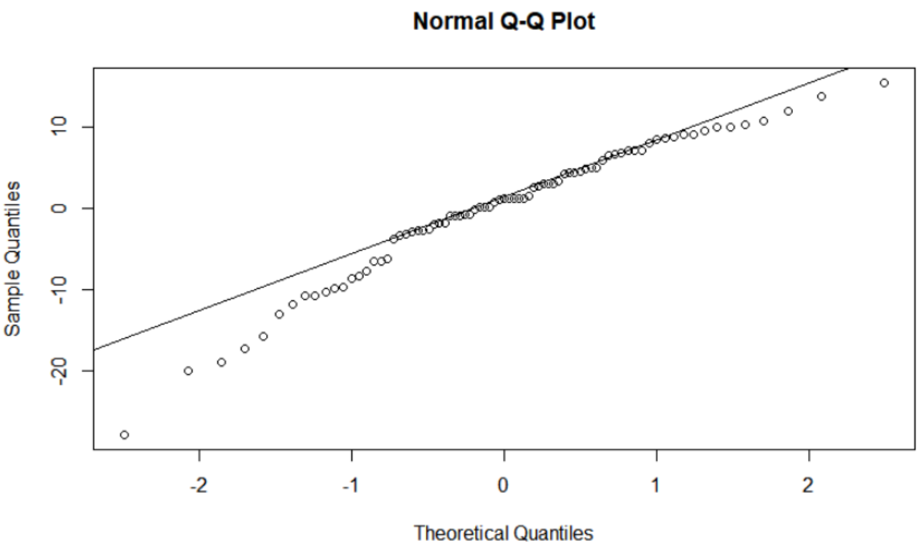
```
> resi_1= residuals(model1)
> shapiro.test(resi_1)

	Shapiro-Wilk normality test

data:  resi_1
W = 0.95583, p-value = 0.008067
```

As p-value = 0.008067< 0.05, we reject H0 at 5% l.o.s and conclude that errors are not normally distributed for model 1.

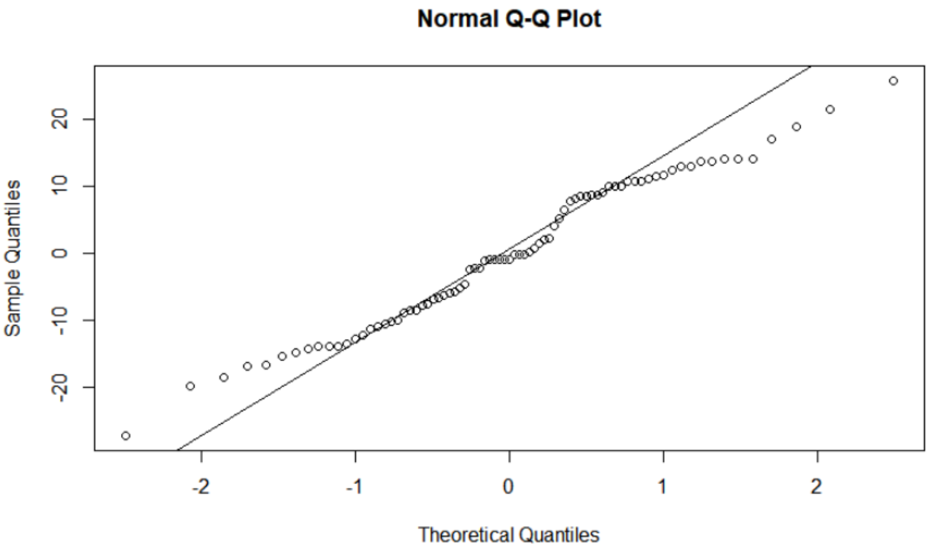
```
> resi_2= residuals(model2)
> shapiro.test(resi_2)

	Shapiro-Wilk normality test

data:  resi_2
W = 0.97682, p-value = 0.16
```

As p-value = 0.16> 0.05, we do not reject H0 at 5% l.o.s and conclude that errors are normally distributed for model 2.

> qqnorm(resi_1)
> qqline(resi_1)



> qqline(resi_2)
> qqnorm(resi_2)



- Autocorrelation-

>install.packages("car")
>install.packages("carData")
>install.packages("lmtest")
> library(car)
> library(carData)
> library(lmtest)

$H_0: \rho = 0$
$H_1: \rho \neq 0$

```
> dwtest(model1)

        Durbin-Watson test

data:  model1
DW = 2.342, p-value = 0.9351
alternative hypothesis: true autocorrelation is greater than 0
```

Since p-value = 0.9351> 0.05, we do not reject H_0 at 5% l.o.s. and conclude that the errors are independently distributed for model 1.

```
> dwtest(model2)

        Durbin-Watson test

data:  model2
DW = 1.8938, p-value = 0.3131
alternative hypothesis: true autocorrelation is greater than 0
```

Since p-value = 0.3131> 0.05, we do not reject H_0 at 5% l.o.s. and conclude that the errors are independently distributed for model 2.

- Heteroscedasticity-

> install.packages("tseries")
>library(tseries)

H_0: constant variance
H_1: Not H_0

```
> bptest(model1)

        studentized Breusch-Pagan test

data:  model1
BP = 2.0062, df = 2, p-value = 0.3667
```

Since p-value = 0.3667> 0.05, we do not reject H_0 at 5% l.o.s. and conclude that the errors have constant variance i.e. heteroscedasticity is absent for model 1.

```
> bptest(model2)

        studentized Breusch-Pagan test

data:  model2
BP = 1.5206, df = 2, p-value = 0.4675
```

Since p-value = 0.4675> 0.05, we do not reject H_0 at 5% l.o.s. and conclude that the errors have constant variance i.e. heteroscedasticity is absent for model 2.

- Multicollinearity-

>install.packages("olsrr")
>library(olsrr)

```
> ols_vif_tol(model1)
  Variables Tolerance      VIF
1        X1  0.999926 1.000074
2        X5  0.999926 1.000074
> ols_vif_tol(model2)
  Variables Tolerance      VIF
1        X3 0.9884966 1.011637
2        X5 0.9884966 1.011637
```

We can see that the VIF values in both the models are very close to 1. Hence we can conclude that multicollinearity is absent in the data.

3. FITTING A MULTIVARIATE REGRESSION MODEL:

```
> model3= lm(cbind(Y1,Y2)~. ,data= my_data)
> model3

Call:
lm(formula = cbind(Y1, Y2) ~ ., data = my_data)

Coefficients:
             Y1         Y2        
(Intercept)  78.104733  63.448459
X1            1.048584   1.297142
X2            0.138947  -0.686562
X3            0.565051   1.482369
X4            0.452553   1.356348
X5           -3.217113  -3.577685
X6            0.008479  -0.037252
```

Hence the two fitted models are:

Y1= 78.104733 +1.048584X1 +0.138947X2 +0.565051X3 +0.452553X4 –3.217113X5 +0.008479X6

Y2= 63.448459 + 1.297142X1 -0.686562X2 +1.482369X3 + 1.356348X4 -3.577685X5 -0.037252X6

```
> result= manova(model3)
> result
Call:
   manova(model3)

Terms:
                      X1       X2       X3       X4       X5       X6 Residuals
Y1               193.229    0.140   19.813    9.245  924.142    0.325  5345.829
Y2               248.401   96.265  249.060   30.396 1191.053    6.264  9358.078
Deg. of Freedom        1        1        1        1        1        1        72

Residual standard errors: 8.616706 11.40058
Estimated effects may be unbalanced
```

```
> summary(model3)
Response Y1 :

Call:
lm(formula = Y1 ~ X1 + X2 + X3 + X4 + X5 + X6, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max
-27.3840  -4.4210   0.8477   5.8584  16.2872

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 78.104733   5.236992  14.914  < 2e-16 ***
X1           1.048584   0.780412   1.344 0.183288
X2           0.138947   0.595784   0.233 0.816254
X3           0.565051   0.672040   0.841 0.403244
X4           0.452553   0.814906   0.555 0.580381
X5          -3.217113   0.930701  -3.457 0.000921 ***
X6           0.008479   0.128256   0.066 0.947472
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.617 on 72 degrees of freedom
Multiple R-squared:  0.1766,	Adjusted R-squared:  0.108
F-statistic: 2.574 on 6 and 72 DF,  p-value: 0.02573
```

```
Response Y2 :

Call:
lm(formula = Y2 ~ X1 + X2 + X3 + X4 + X5 + X6, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max 
-27.4019  -8.4044  -0.8855   9.3933  27.7464 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 63.44846    6.92895   9.157 1.07e-13 ***
X1           1.29714    1.03255   1.256  0.21308    
X2          -0.68656    0.78827  -0.871  0.38666    
X3           1.48237    0.88916   1.667  0.09983 .  
X4           1.35635    1.07818   1.258  0.21246    
X5          -3.57769    1.23139  -2.905  0.00487 ** 
X6          -0.03725    0.16969  -0.220  0.82686    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.4 on 72 degrees of freedom
Multiple R-squared:  0.1629,	Adjusted R-squared:  0.09317 
F-statistic: 2.336 on 6 and 72 DF,  p-value: 0.04069
```

4. CHECKING ASSUMPTIONS FOR MODEL 3:

- Normality-

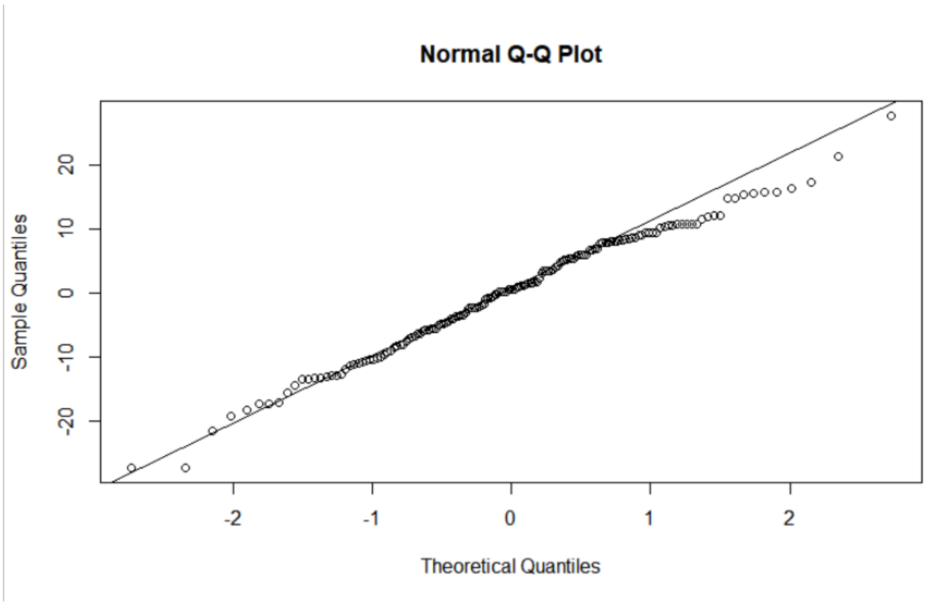
H_0: Errors are normally distributed

H_1 : Not H_0

```
> resi3= residuals(model3)
> shapiro.test(resi3)

	Shapiro-Wilk normality test

data:  resi3
W = 0.98999, p-value = 0.3269
```

As p-value = 0.3269> 0.05, we do not reject H_0 at 5% l.o.s and conclude that errors are normally distributed for model 3.

```
> qqnorm(resi3)
> qqline(resi3)
```



- Autocorrelation-

$H_0: \rho = 0$

$H_1: \rho \neq 0$

```
> dwtest(model3)

        Durbin-Watson test

data:  model3
DW = 2.0483, p-value = 0.5867
alternative hypothesis: true autocorrelation is greater than 0
```

Since p-value = 0.5867> 0.05, we do not reject H_0 at 5% l.o.s. and conclude that the errors are independently distributed for model 3.

- Heteroscedasticity-

H_0: constant variance

H_1: Not H_0

```
> bptest(model3)

        studentized Breusch-Pagan test

data:  model3
BP = 31.482, df = 6, p-value = 2.051e-05
```

Since p-value = 0.0138< 0.05, we do reject H_0 at 5% l.o.s. and conclude that the errors don't have constant variance i.e. heteroscedasticity is present for model 3.