GLM_Assignment_1

SAGAR KUMAR
194161013
Data Science

Importing Data

```
library("readxl")   ## Including library
concrete.data<-read_excel("Concrete_Data.xls") ##Importing Data
concrete.data<-concrete.data[,-c(1)] ## Removing Id Column From data Set
str(concrete.data)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    1030 obs. of  9 variables:
##  $ Cement (component 1)(kg in a m^3 mixture)            : num  540 540 332 332 199 ...
##  $ Blast Furnace Slag (component 2)(kg in a m^3 mixture): num  0 0 142 142 132 ...
##  $ Fly Ash (component 3)(kg in a m^3 mixture)           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Water  (component 4)(kg in a m^3 mixture)            : num  162 162 228 228 192 228 228 22
8 228 228 ...
##  $ Superplasticizer (component 5)(kg in a m^3 mixture)  : num  2.5 2.5 0 0 0 0 0 0 0 0 ...
##  $ Coarse Aggregate  (component 6)(kg in a m^3 mixture) : num  1040 1055 932 932 978 ...
##  $ Fine Aggregate (component 7)(kg in a m^3 mixture)    : num  676 676 594 594 826 ...
##  $ Age (day)                                            : num  28 28 270 365 360 90 365 28 28
28 ...
##  $ Concrete compressive strength(MPa, megapascals)      : num  80 61.9 40.3 41.1 44.3 ...
```

Observation of Concrete Data :

In Concrete dataset I removed The first column (ID Column)
Looking at the output we get to know that we have 1030 samples and 9 variables.

All the predictor variables are of the numeric class. Now look at the summary of data to understand better.

```
#summary of the Data
summary(concrete.data)
```

```
##  Cement (component 1)(kg in a m^3 mixture)
##  Min.   :102.0
##  1st Qu.:192.4
##  Median :272.9
##  Mean   :281.2
##  3rd Qu.:350.0
##  Max.   :540.0
##  Blast Furnace Slag (component 2)(kg in a m^3 mixture)
##  Min.   :  0.0
##  1st Qu.:  0.0
##  Median : 22.0
##  Mean   : 73.9
##  3rd Qu.:142.9
##  Max.   :359.4
##  Fly Ash (component 3)(kg in a m^3 mixture)
##  Min.   :  0.00
##  1st Qu.:  0.00
##  Median :  0.00
##  Mean   : 54.19
##  3rd Qu.:118.27
##  Max.   :200.10
##  Water  (component 4)(kg in a m^3 mixture)
##  Min.   :121.8
##  1st Qu.:164.9
##  Median :185.0
##  Mean   :181.6
##  3rd Qu.:192.0
##  Max.   :247.0
##  Superplasticizer (component 5)(kg in a m^3 mixture)
##  Min.   : 0.000
##  1st Qu.: 0.000
##  Median : 6.350
##  Mean   : 6.203
##  3rd Qu.:10.160
##  Max.   :32.200
##  Coarse Aggregate  (component 6)(kg in a m^3 mixture)
##  Min.   : 801.0
##  1st Qu.: 932.0
##  Median : 968.0
##  Mean   : 972.9
##  3rd Qu.:1029.4
##  Max.   :1145.0
##  Fine Aggregate (component 7)(kg in a m^3 mixture)   Age (day)
##  Min.   :594.0                                      Min.   :  1.00
##  1st Qu.:731.0                                      1st Qu.:  7.00
##  Median :779.5                                      Median : 28.00
##  Mean   :773.6                                      Mean   : 45.66
##  3rd Qu.:824.0                                      3rd Qu.: 56.00
##  Max.   :992.6                                      Max.   :365.00
##  Concrete compressive strength(MPa, megapascals)
##  Min.   : 2.332
##  1st Qu.:23.707
##  Median :34.443
```

```
##  Mean   :35.818
##  3rd Qu.:46.136
##  Max.   :82.599
```

Above summary gives us information about minimum,1st quartile,median,mean,3rd quartile, and Max.

for example: Consider “Concrete compressive strength(MPa, megapascals)” Variable .

Min. : 2.332
1st Qu.:23.707
Median :34.443
Mean :35.818
3rd Qu.:46.136
Max. :82.599

To better understand the data and relationship between variable we are trying to find correlation between each variable. Correlation can have a value:

1 is a perfect positive correlation,0 is no correlation (the values (variables) don’t seem linked at all),-1 is a perfect negative correlation

```
##correlation b/w each other(All Predictors)
cor(concrete.data)
```

```
##                                                        Cement (component 1)(kg in a m^3 mixtur
e)
## Cement (component 1)(kg in a m^3 mixture)                                              1.000000
00
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)                                 -0.275193
44
## Fly Ash (component 3)(kg in a m^3 mixture)                                            -0.397475
44
## Water  (component 4)(kg in a m^3 mixture)                                             -0.081543
61
## Superplasticizer (component 5)(kg in a m^3 mixture)                                    0.092771
37
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)                                  -0.109356
04
## Fine Aggregate (component 7)(kg in a m^3 mixture)                                     -0.222720
17
## Age (day)                                                                              0.081947
26
## Concrete compressive strength(MPa, megapascals)                                        0.497832
72
##                                                        Blast Furnace Slag (component 2)(kg in
a m^3 mixture)
## Cement (component 1)(kg in a m^3 mixture)
-0.27519344
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)
1.00000000
## Fly Ash (component 3)(kg in a m^3 mixture)
-0.32356947
## Water  (component 4)(kg in a m^3 mixture)
0.10728594
## Superplasticizer (component 5)(kg in a m^3 mixture)
0.04337574
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)
-0.28399823
## Fine Aggregate (component 7)(kg in a m^3 mixture)
-0.28159326
## Age (day)
-0.04424580
## Concrete compressive strength(MPa, megapascals)
0.13482445
##                                                        Fly Ash (component 3)(kg in a m^3 mixtu
re)
## Cement (component 1)(kg in a m^3 mixture)                                             -0.397475
440
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)                                 -0.323569
468
## Fly Ash (component 3)(kg in a m^3 mixture)                                             1.000000
000
## Water  (component 4)(kg in a m^3 mixture)                                             -0.257043
997
## Superplasticizer (component 5)(kg in a m^3 mixture)                                    0.377339
559
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)                                  -0.009976
```

```
788
## Fine Aggregate (component 7)(kg in a m^3 mixture)                                 0.079076
351
## Age (day)                                                                        -0.154370
165
## Concrete compressive strength(MPa, megapascals)                                  -0.105753
348
##                                                       Water  (component 4)(kg in a m^3 mixtur
e)
## Cement (component 1)(kg in a m^3 mixture)                                        -0.081543
61
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)                             0.107285
94
## Fly Ash (component 3)(kg in a m^3 mixture)                                       -0.257044
00
## Water  (component 4)(kg in a m^3 mixture)                                         1.000000
00
## Superplasticizer (component 5)(kg in a m^3 mixture)                              -0.657464
44
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)                             -0.182311
67
## Fine Aggregate (component 7)(kg in a m^3 mixture)                                -0.450634
98
## Age (day)                                                                         0.277604
43
## Concrete compressive strength(MPa, megapascals)                                  -0.289613
48
##                                                       Superplasticizer (component 5)(kg in a
m^3 mixture)
## Cement (component 1)(kg in a m^3 mixture)
0.09277137
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)
0.04337574
## Fly Ash (component 3)(kg in a m^3 mixture)
0.37733956
## Water  (component 4)(kg in a m^3 mixture)
-0.65746444
## Superplasticizer (component 5)(kg in a m^3 mixture)
1.00000000
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)
-0.26630276
## Fine Aggregate (component 7)(kg in a m^3 mixture)
0.22250149
## Age (day)
-0.19271652
## Concrete compressive strength(MPa, megapascals)
0.36610230
##                                                       Coarse Aggregate  (component 6)(kg in a
m^3 mixture)
## Cement (component 1)(kg in a m^3 mixture)
-0.109356039
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)
-0.283998230
## Fly Ash (component 3)(kg in a m^3 mixture)
```

```
-0.009976788
## Water  (component 4)(kg in a m^3 mixture)
-0.182311668
## Superplasticizer (component 5)(kg in a m^3 mixture)
-0.266302755
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)
1.000000000
## Fine Aggregate (component 7)(kg in a m^3 mixture)
-0.178505755
## Age (day)
-0.003015507
## Concrete compressive strength(MPa, megapascals)
-0.164927821
##                                                        Fine Aggregate (component 7)(kg in a m^
3 mixture)
## Cement (component 1)(kg in a m^3 mixture)                                                    -
0.22272017
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)                                        -
0.28159326
## Fly Ash (component 3)(kg in a m^3 mixture)
0.07907635
## Water  (component 4)(kg in a m^3 mixture)                                                    -
0.45063498
## Superplasticizer (component 5)(kg in a m^3 mixture)
0.22250149
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)                                         -
0.17850575
## Fine Aggregate (component 7)(kg in a m^3 mixture)
1.00000000
## Age (day)                                                                                    -
0.15609405
## Concrete compressive strength(MPa, megapascals)                                              -
0.16724896
##                                                           Age (day)
## Cement (component 1)(kg in a m^3 mixture)               0.081947264
## Blast Furnace Slag (component 2)(kg in a m^3 mixture) -0.044245801
## Fly Ash (component 3)(kg in a m^3 mixture)             -0.154370165
## Water  (component 4)(kg in a m^3 mixture)               0.277604429
## Superplasticizer (component 5)(kg in a m^3 mixture)    -0.192716518
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)   -0.003015507
## Fine Aggregate (component 7)(kg in a m^3 mixture)      -0.156094049
## Age (day)                                               1.000000000
## Concrete compressive strength(MPa, megapascals)         0.328876976
##                                                        Concrete compressive strength(MPa, mega
pascals)
## Cement (component 1)(kg in a m^3 mixture)
0.4978327
## Blast Furnace Slag (component 2)(kg in a m^3 mixture)
0.1348244
## Fly Ash (component 3)(kg in a m^3 mixture)                                                  -
0.1057533
## Water  (component 4)(kg in a m^3 mixture)                                                   -
0.2896135
## Superplasticizer (component 5)(kg in a m^3 mixture)
```

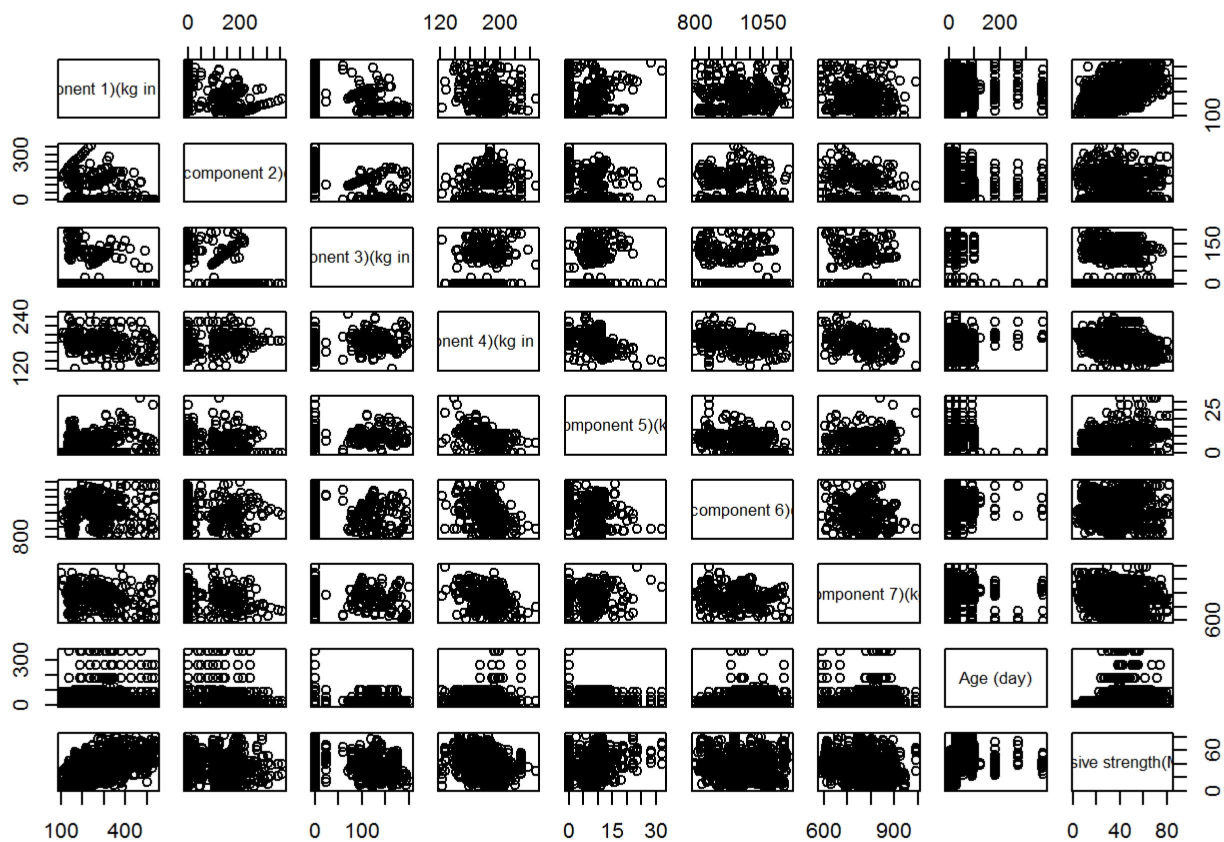
```
0.3661023
## Coarse Aggregate  (component 6)(kg in a m^3 mixture)                         -
0.1649278
## Fine Aggregate (component 7)(kg in a m^3 mixture)                             -
0.1672490
## Age (day)
0.3288770
## Concrete compressive strength(MPa, megapascals)
1.0000000
```

Observation

No. predictors are strongly correlated to other predictors No. predictors are linear Combination of others.

Pair plot between variables

```
#Plot concrete data
plot(concrete.data)
```



##Observation:

Since we didn’t specify the x and y axes,R plots all the data against Each other. **Question 1**

Fit a multiple linear regression to the above data considering all the input variables. The information of the output variable is given in the information file.?

```
#Fit the model
multiple.regression<-lm(`Concrete compressive strength(MPa, megapascals)`~.,data=concrete.data)
summary(multiple.regression)
```

```
##
## Call:
## lm(formula = `Concrete compressive strength(MPa, megapascals)` ~
##     ., data = concrete.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.653  -6.303   0.704   6.562  34.446
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                            -23.163756  26.588421
## `Cement (component 1)(kg in a m^3 mixture)`              0.119785   0.008489
## `Blast Furnace Slag (component 2)(kg in a m^3 mixture)`   0.103847   0.010136
## `Fly Ash (component 3)(kg in a m^3 mixture)`             0.087943   0.012585
## `Water  (component 4)(kg in a m^3 mixture)`             -0.150298   0.040179
## `Superplasticizer (component 5)(kg in a m^3 mixture)`    0.290687   0.093460
## `Coarse Aggregate  (component 6)(kg in a m^3 mixture)`   0.018030   0.009394
## `Fine Aggregate (component 7)(kg in a m^3 mixture)`      0.020154   0.010703
## `Age (day)`                                              0.114226   0.005427
##                                                        t value Pr(>|t|)
## (Intercept)                                             -0.871 0.383851
## `Cement (component 1)(kg in a m^3 mixture)`             14.110  < 2e-16 ***
## `Blast Furnace Slag (component 2)(kg in a m^3 mixture)`  10.245  < 2e-16 ***
## `Fly Ash (component 3)(kg in a m^3 mixture)`             6.988 5.03e-12 ***
## `Water  (component 4)(kg in a m^3 mixture)`             -3.741 0.000194 ***
## `Superplasticizer (component 5)(kg in a m^3 mixture)`    3.110 0.001921 **
## `Coarse Aggregate  (component 6)(kg in a m^3 mixture)`   1.919 0.055227 .
## `Fine Aggregate (component 7)(kg in a m^3 mixture)`      1.883 0.059968 .
## `Age (day)`                                             21.046  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.4 on 1021 degrees of freedom
## Multiple R-squared:  0.6155, Adjusted R-squared:  0.6125
## F-statistic: 204.3 on 8 and 1021 DF,  p-value: < 2.2e-16
```

Question2

Explain the results obtained from the statistical software along with the diagnostic results/plots. I expect your explanation to be detailed along with interpretations.

```
summary(multiple.regression$residuals)  #Summary of Residual
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -28.6533  -6.3027   0.7038   0.0000   6.5619  34.4462
```

```
mean(multiple.regression$residuals)   ## Expected value of the residulas.
```

```
## [1] -1.065868e-15
```

```
var(multiple.regression$residuals)   ## Variance of the residulas
```

```
## [1] 107.316
```

Residuals are essentially the difference between the actual observed response values and the response values that the model predicted.

from summary of the Residuals we can see that the distribution of the residuals is somewhat symmetry. **max=34.662 ,min=-28.65,and mean=0**

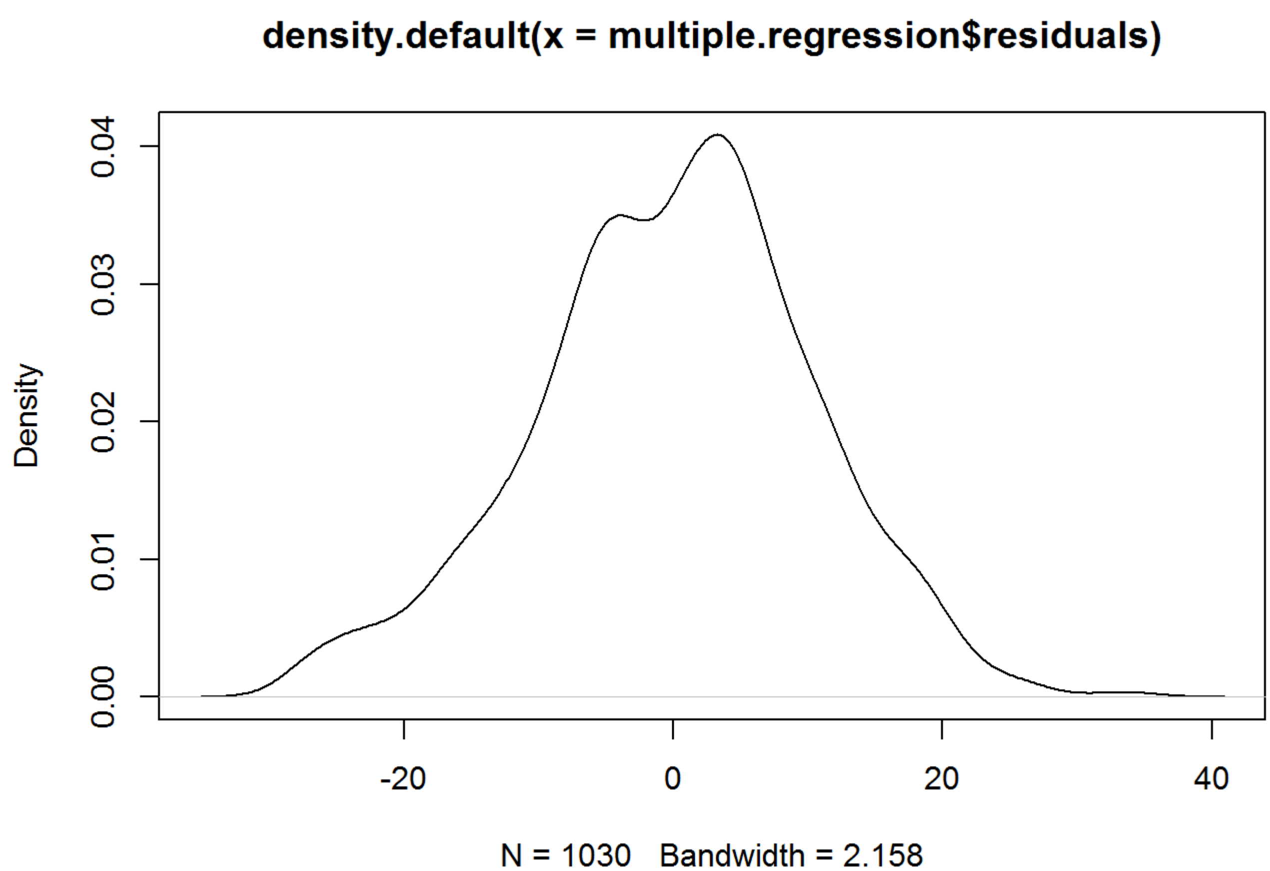
Linear regression makes several assumptions about the Residuals(Errors), such as :

1.The errors are uncorrelted with each other.Violation of this assumption can lead to very misleading assesments of the strength of the regression. 2.The expected value of errors is zero. 3.The assumption of homoscedasity.(Constant Variance) 4.Distribution of errors (Residulals) follows normal distribution.

Expected value(mean) of the residulas is: -1.065868e-15 (it means residuals expected value is almost zero) which satisfy our assumption. Variance of the residulas is: 107.316

I mentioned that the residuals follow a normal distribution. let’s check whether residuals follows normal distribution or not ?

```
plot(density(multiple.regression$residuals))   ##Density plot of residual
```



We can clearly see that the residuals follows Approximatley normal, but is a multi modal normal distribution which indicates that our model is not very good.

```
coef(multiple.regression)
```

```
##                                              (Intercept) 
##                                             -23.16375581 
##             `Cement (component 1)(kg in a m^3 mixture)` 
##                                               0.11978526 
## `Blast Furnace Slag (component 2)(kg in a m^3 mixture)` 
##                                               0.10384725 
##            `Fly Ash (component 3)(kg in a m^3 mixture)` 
##                                               0.08794308 
##             `Water  (component 4)(kg in a m^3 mixture)` 
##                                              -0.15029790 
##   `Superplasticizer (component 5)(kg in a m^3 mixture)` 
##                                               0.29068694 
##  `Coarse Aggregate  (component 6)(kg in a m^3 mixture)` 
##                                               0.01803018 
##     `Fine Aggregate (component 7)(kg in a m^3 mixture)` 
##                                               0.02015446 
##                                             `Age (day)` 
##                                               0.11422562
```

Coefficient Estimate :

This is the expected change in Y per unit change in X. In our model, one unit change in Cement contributes to 0.11978526 unit change in Concrete compressive strength(MPa, megapascals) . Same is the case for other variables as well.

If the estimate value is 0.0 ,it means it adds no significance to the model and is considered worthless.

```
summary(multiple.regression)
```

```
##
## Call:
## lm(formula = `Concrete compressive strength(MPa, megapascals)` ~
##     ., data = concrete.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.653  -6.303   0.704   6.562  34.446
##
## Coefficients:
##                                                           Estimate Std. Error
## (Intercept)                                             -23.163756  26.588421
## `Cement (component 1)(kg in a m^3 mixture)`               0.119785   0.008489
## `Blast Furnace Slag (component 2)(kg in a m^3 mixture)`   0.103847   0.010136
## `Fly Ash (component 3)(kg in a m^3 mixture)`              0.087943   0.012585
## `Water  (component 4)(kg in a m^3 mixture)`              -0.150298   0.040179
## `Superplasticizer (component 5)(kg in a m^3 mixture)`     0.290687   0.093460
## `Coarse Aggregate  (component 6)(kg in a m^3 mixture)`    0.018030   0.009394
## `Fine Aggregate (component 7)(kg in a m^3 mixture)`       0.020154   0.010703
## `Age (day)`                                               0.114226   0.005427
##                                                         t value Pr(>|t|)
## (Intercept)                                              -0.871 0.383851
## `Cement (component 1)(kg in a m^3 mixture)`              14.110  < 2e-16 ***
## `Blast Furnace Slag (component 2)(kg in a m^3 mixture)`  10.245  < 2e-16 ***
## `Fly Ash (component 3)(kg in a m^3 mixture)`              6.988 5.03e-12 ***
## `Water  (component 4)(kg in a m^3 mixture)`              -3.741 0.000194 ***
## `Superplasticizer (component 5)(kg in a m^3 mixture)`     3.110 0.001921 **
## `Coarse Aggregate  (component 6)(kg in a m^3 mixture)`    1.919 0.055227 .
## `Fine Aggregate (component 7)(kg in a m^3 mixture)`       1.883 0.059968 .
## `Age (day)`                                              21.046  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.4 on 1021 degrees of freedom
## Multiple R-squared:  0.6155, Adjusted R-squared:  0.6125
## F-statistic: 204.3 on 8 and 1021 DF,  p-value: < 2.2e-16
```

Interpreting P-Values for Variables in a Regression Model The p-values help determine whether the relationships that you observe in your sample also exist in the larger population. The p-value for each independent variable tests the null hypothesis that the variable has no correlation with the dependent variable. If there is no correlation, there is no association between the changes in the independent variable and the shifts in the dependent variable. In other words, there is insufficient evidence to conclude that there is effect at the population level.

Lower the p value allow us to reject null hypothesis.i.e the estimate of any explanatory variable is 0 or there is no relationship between the predictor and the response. Asterisks mark aside p value define significance of value, lower the value ,higher is its significance and higher is the number of asterisks. Hence, if we see a small p-value, then we can infer that there is an association between the predictor and the response. We reject the null hypothesis—that is, we declare a relationship to exist between X and Y —if the p-value is small enough

Typically, a p-value of 5% or less is a good cut-off point.

A small p-value indicates that there is a relationship between the predictor and response variables

In our dataset, Cement , Blast Furnace Slag, Fly Ash ,Water,Age and Superplasticizer are the significant variables .

Insignificant Variables are:Superplasticizer (component 5)(kg in a m^3 mixture),Fine Aggregate (component 7)(kg in a m^3 mixture)

Interpreting Regression Coefficients for Linear Relationships The sign of a regression coefficient tells you whether there is a positive or negative correlation between each independent variable the dependent variable. A positive coefficient indicates that as the value of the independent variable increases, the mean of the dependent variable also tends to increase. A negative coefficient suggests that as the independent variable increases, the dependent variable tends to decrease.

In our model only **water(predictor)** has a negative coefficient and rest all the coefficients are positive.

Standard error

The standard deviation of an estimate is called the standard error. The standard error of the coefficient measures how precisely the model estimates the coefficient's unknown value. Lower the error, better the estimate.

The coefficient Standard Error measures the average amount that the coefficient estimates vary from the actual.

We expect minimum value of standard error.

Total we have 1030 observations and 9 variable so removing 9 data points (1030-9) = 1021 degrees of freedom.

Multiple R-squared

R-square is a goodness-of-fit measure for linear regression models. It determines the percentage of variation in the response variable that is explained by variation in the explanatory variable. this is use to calculate how well the model is doing to explain the things.R-squared evaluates the scatter of the data points around the fitted regression line. Higher the values of R-square are desirables."How high is High" depends on the context.

It always lies between 0 and 1 (i.e.: a number near 0 represents a regression that does not explain the variance in the response variable well and a number close to 1 does explain the observed variance in the response variable).

In our model the Value of R-square is : 0.6155

Adjusted R-squared

The adjusted Rsquared increase only if the new terms improves the model more than would be expected by chance.

It decreases when a predictor improves the model by less then expected by chance.The adjusted Rsquared can be negative but it's usually not. It is always lower than The R-squared.

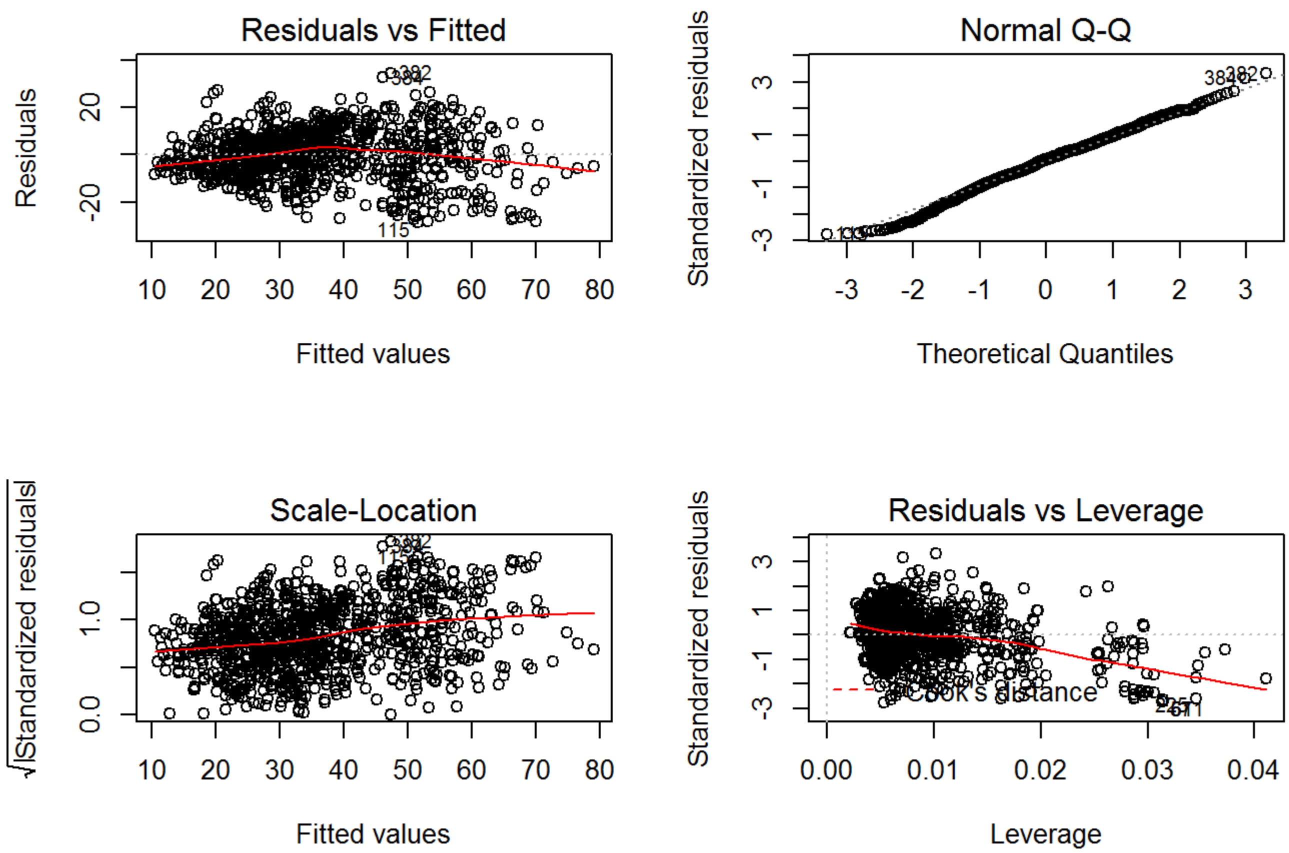
In our Model the value of Adjusted Rsquare is:0.6125

F-Statistic:

F-statistic is a good indicator of whether there is a relationship between our predictor and the response variables. We can also say that if there is greater relationship then our modelis good and not good otherwise. It tells us whether the modelis significant or not. The model is significant if any of the coefficients are non-zero.

Diagnostic Plot

```
par(mfrow=c(2,2))
plot(multiple.regression)
```



Residual vs Fitted

It is used to check the linear relationship assumptions.We expect a horizontal lines without distinct patterns.

It is also used to check the assumption of constant variance of errors.We expect the scatter should be symmetric vertically (along y-axis) about zero.

In our model As expected Scatter plot which is shown is Vertical symmetric.and there is no distinctive pattern

Normal QQ

Normal Q-Q Plot" provides a graphical way to determine the level of normality. The black line indicates the values your sample should adhere to if the distribution was normal. The dots are your actual data.

If the dots fall exactly on the black line, then your data are normal

It's good if residuals are lined well on the straight dashed line

In our models almost all the data's are normally distributed.

Scale Location Scale Location(Or Spread Location) is used to check the homoscedasticity assumption (Homogenity of Variance of the Residuals).We expect a Horizontal line with equally spread points.

Horizontal line with equally spread points is a good indication of homoscedasticity. This is not the case in our example, where we have a heteroscedasticity problem.

Residuals V/s leverage It is used to identify infulential Observations.Here we don't expect any pattern. In this plot,The dotted red lines known's as the cook's distance. We check for the points that are outside dotted line on top right or bottom right corner.If any Point falls in that region,we say that observations might influence the regression results when included or excluded from the analysis.

In our model : No points are on top right or bottom right corner of the Graph so there is no influenctial points.

Question3

Based on your diagnostics plots' interpretations, do you want to recommend any changes in the model? If so do the changes along with the reasons and fit the model again. If no change required then support your arguments.?

```
#Fit the model
multiple.regression1<-lm(`Concrete compressive strength(MPa, megapascals)`~.-1,data=concrete.dat
a)
summary(multiple.regression1)
```

```
##
## Call:
## lm(formula = `Concrete compressive strength(MPa, megapascals)` ~
##     . - 1, data = concrete.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.497  -6.461   0.708   6.551  34.453
##
## Coefficients:
##                                                          Estimate Std. Error
## `Cement (component 1)(kg in a m^3 mixture)`              0.113382   0.004249
## `Blast Furnace Slag (component 2)(kg in a m^3 mixture)`  0.096271   0.005207
## `Fly Ash (component 3)(kg in a m^3 mixture)`             0.079390   0.007873
## `Water  (component 4)(kg in a m^3 mixture)`             -0.182382   0.016064
## `Superplasticizer (component 5)(kg in a m^3 mixture)`    0.263378   0.088036
## `Coarse Aggregate  (component 6)(kg in a m^3 mixture)`   0.010292   0.003060
## `Fine Aggregate (component 7)(kg in a m^3 mixture)`      0.011360   0.003555
## `Age (day)`                                              0.114001   0.005421
##                                                         t value Pr(>|t|)
## `Cement (component 1)(kg in a m^3 mixture)`              26.687  < 2e-16 ***
## `Blast Furnace Slag (component 2)(kg in a m^3 mixture)`  18.490  < 2e-16 ***
## `Fly Ash (component 3)(kg in a m^3 mixture)`             10.084  < 2e-16 ***
## `Water  (component 4)(kg in a m^3 mixture)`             -11.354  < 2e-16 ***
## `Superplasticizer (component 5)(kg in a m^3 mixture)`     2.992 0.002841 **
## `Coarse Aggregate  (component 6)(kg in a m^3 mixture)`    3.364 0.000798 ***
## `Fine Aggregate (component 7)(kg in a m^3 mixture)`       3.196 0.001437 **
## `Age (day)`                                              21.031  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.4 on 1022 degrees of freedom
## Multiple R-squared:  0.9313, Adjusted R-squared:  0.9308
## F-statistic:  1732 on 8 and 1022 DF,  p-value: < 2.2e-16
```

Given Domain knowledge if one has zero input (predictors) then the output (response) must be zero. But as predicted by our model the intercept is -23.163756 and the p-value is 0.383851 which is insignificant. So now after removing the model has improved before removing the intercept all the variables where not significant but now each one of them are significant.

And Value of RSquare is:0.9313

Question 4

In the same email, you will get another data as “add columns”. Merge this new data with the previous data. Now fit a multiple linear regression on the merged data with all the input variables. Explain your results. Comment on the newly added variables and whether you want to keep them in the model or not. Justify your answer in either case.?

Importing Data

```
concrete.data_new<-read.csv("C:\\Users\\Sagar\\Desktop\\GLM Assignment\\Concrete_Data_Add Colum
n.csv")
concrete.data_new<-concrete.data_new[,(-c(1))] ## Removing Id Column From data Set
str(concrete.data_new)
```

```
## 'data.frame':    1030 obs. of  11 variables:
##  $ Cement..component.1..kg.in.a.m.3.mixture.           : num  540 540 332 332 199 ...
##  $ Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.: num  0 0 142 142 132 ...
##  $ Fly.Ash..component.3..kg.in.a.m.3.mixture.          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Water...component.4..kg.in.a.m.3.mixture.           : num  162 162 228 228 192 228 228 22
8 228 228 ...
##  $ Superplasticizer..component.5..kg.in.a.m.3.mixture.  : num  2.5 2.5 0 0 0 0 0 0 0 0 ...
##  $ Coarse.Aggregate...component.6..kg.in.a.m.3.mixture. : num  1040 1055 932 932 978 ...
##  $ Fine.Aggregate..component.7..kg.in.a.m.3.mixture.    : num  676 676 594 594 826 ...
##  $ Age..day.                                           : num  28 28 270 365 360 90 365 28 28
28 ...
##  $ Concrete.compressive.strength.MPa..megapascals..     : num  80 61.9 40.3 41 44.3 ...
##  $ new_input.1                                         : num  3580 3586 5904 5904 4276 ...
##  $ new_input.2                                         : num  165 165 102.8 102.8 62.6 ...
```

Observation of Concrete Data :

Looking at the output we get to know that we have 1030 samples and 11 variables.(2 new predictor added)

All the predictor variables are of the numeric class.

Now look at the summary of data to understand better.

```
summary(concrete.data_new)
```

```
##  Cement..component.1..kg.in.a.m.3.mixture.
##  Min.   :102.0
##  1st Qu.:192.4
##  Median :272.9
##  Mean   :281.2
##  3rd Qu.:350.0
##  Max.   :540.0
##  Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
##  Min.   :  0.0
##  1st Qu.:  0.0
##  Median : 22.0
##  Mean   : 73.9
##  3rd Qu.:142.9
##  Max.   :359.4
##  Fly.Ash..component.3..kg.in.a.m.3.mixture.
##  Min.   :  0.00
##  1st Qu.:  0.00
##  Median :  0.00
##  Mean   : 54.19
##  3rd Qu.:118.30
##  Max.   :200.10
##  Water...component.4..kg.in.a.m.3.mixture.
##  Min.   :121.8
##  1st Qu.:164.9
##  Median :185.0
##  Mean   :181.6
##  3rd Qu.:192.0
##  Max.   :247.0
##  Superplasticizer..component.5..kg.in.a.m.3.mixture.
##  Min.   : 0.000
##  1st Qu.: 0.000
##  Median : 6.400
##  Mean   : 6.205
##  3rd Qu.:10.200
##  Max.   :32.200
##  Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.
##  Min.   : 801.0
##  1st Qu.: 932.0
##  Median : 968.0
##  Mean   : 972.9
##  3rd Qu.:1029.4
##  Max.   :1145.0
##  Fine.Aggregate..component.7..kg.in.a.m.3.mixture.   Age..day.
##  Min.   :594.0                                      Min.   :  1.00
##  1st Qu.:731.0                                      1st Qu.:  7.00
##  Median :779.5                                      Median : 28.00
##  Mean   :773.6                                      Mean   : 45.66
##  3rd Qu.:824.0                                      3rd Qu.: 56.00
##  Max.   :992.6                                      Max.   :365.00
##  Concrete.compressive.strength.MPa..megapascals..  new_input.1
##  Min.   : 2.33                                    Min.   :2074
##  1st Qu.:23.71                                    1st Qu.:3457
##  Median :34.45                                    Median :4081
```

```
##  Mean   :35.82                                   Mean   :4013  
##  3rd Qu.:46.13                                   3rd Qu.:4419  
##  Max.   :82.60                                   Max.   :6679  
##   new_input.2    
##  Min.   : 33.60  
##  1st Qu.: 60.71  
##  Median : 84.87  
##  Mean   : 87.35  
##  3rd Qu.:108.00  
##  Max.   :165.00
```

Above summary gives us information about minimum,1st quartile,median,mean,3rd quartile, and Max.

for example: Consider “new_input.1” Variable .

Min. : 2074
1st Qu.:3457
Median :4081
Mean :4013
3rd Qu.:4419
Max. :6679

```
cor(concrete.data_new)
```

```
##                                                       Cement..component.1..kg.in.a.m.3.mixtur
e.
## Cement..component.1..kg.in.a.m.3.mixture.                                             1.000000
00
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.                                -0.275215
91
## Fly.Ash..component.3..kg.in.a.m.3.mixture.                                           -0.397467
34
## Water...component.4..kg.in.a.m.3.mixture.                                            -0.081586
75
## Superplasticizer..component.5..kg.in.a.m.3.mixture.                                   0.092386
17
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.                                 -0.109348
99
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.                                    -0.222717
85
## Age..day.                                                                             0.081946
02
## Concrete.compressive.strength.MPa..megapascals..                                      0.497831
92
## new_input.1                                                                           0.062540
16
## new_input.2                                                                           0.999999
99
##                                                       Blast.Furnace.Slag..component.2..kg.in.
a.m.3.mixture.
## Cement..component.1..kg.in.a.m.3.mixture.
-0.27521591
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
1.00000000
## Fly.Ash..component.3..kg.in.a.m.3.mixture.
-0.32357990
## Water...component.4..kg.in.a.m.3.mixture.
0.10725203
## Superplasticizer..component.5..kg.in.a.m.3.mixture.
0.04327042
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.
-0.28399861
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.
-0.28160267
## Age..day.
-0.04424602
## Concrete.compressive.strength.MPa..megapascals..
0.13482926
## new_input.1
0.06088931
## new_input.2
-0.27519683
##                                                       Fly.Ash..component.3..kg.in.a.m.3.mixtu
re.
## Cement..component.1..kg.in.a.m.3.mixture.                                            -0.397467
341
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.                                -0.323579
```

```
901
## Fly.Ash..component.3..kg.in.a.m.3.mixture.                                                     1.000000
000
## Water...component.4..kg.in.a.m.3.mixture.                                                     -0.256984
023
## Superplasticizer..component.5..kg.in.a.m.3.mixture.                                            0.377503
146
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.                                          -0.009960
828
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.                                              0.079108
491
## Age..day.                                                                                     -0.154370
516
## Concrete.compressive.strength.MPa..megapascals..                                              -0.105754
916
## new_input.1                                                                                   -0.313638
879
## new_input.2                                                                                   -0.397480
818
##                                                         Water...component.4..kg.in.a.m.3.mixtur
e.
## Cement..component.1..kg.in.a.m.3.mixture.                                                     -0.081586
75
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.                                          0.107252
03
## Fly.Ash..component.3..kg.in.a.m.3.mixture.                                                    -0.256984
02
## Water...component.4..kg.in.a.m.3.mixture.                                                      1.000000
00
## Superplasticizer..component.5..kg.in.a.m.3.mixture.                                           -0.657532
91
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.                                          -0.182293
60
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.                                             -0.450661
17
## Age..day.                                                                                      0.277618
22
## Concrete.compressive.strength.MPa..megapascals..                                              -0.289633
38
## new_input.1                                                                                    0.985808
47
## new_input.2                                                                                   -0.081583
15
##                                                         Superplasticizer..component.5..kg.in.a.
m.3.mixture.
## Cement..component.1..kg.in.a.m.3.mixture.
0.09238617
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
0.04327042
## Fly.Ash..component.3..kg.in.a.m.3.mixture.
0.37750315
## Water...component.4..kg.in.a.m.3.mixture.
-0.65753291
## Superplasticizer..component.5..kg.in.a.m.3.mixture.
```

```
1.00000000
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.
-0.26599915
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.
0.22269123
## Age..day.
-0.19270003
## Concrete.compressive.strength.MPa..megapascals..
0.36607883
## new_input.1
-0.64121506
## new_input.2
0.09238409
##                                                        Coarse.Aggregate...component.6..kg.in.
a.m.3.mixture.
## Cement..component.1..kg.in.a.m.3.mixture.
-0.109348994
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
-0.283998612
## Fly.Ash..component.3..kg.in.a.m.3.mixture.
-0.009960828
## Water...component.4..kg.in.a.m.3.mixture.
-0.182293602
## Superplasticizer..component.5..kg.in.a.m.3.mixture.
-0.265999148
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.
1.000000000
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.
-0.178480957
## Age..day.
-0.003015880
## Concrete.compressive.strength.MPa..megapascals..
-0.164934614
## new_input.1
-0.171939122
## new_input.2
-0.109357602
##                                                        Fine.Aggregate..component.7..kg.in.a.m.
3.mixture.
## Cement..component.1..kg.in.a.m.3.mixture.                                        -
0.22271785
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.                            -
0.28160267
## Fly.Ash..component.3..kg.in.a.m.3.mixture.
0.07910849
## Water...component.4..kg.in.a.m.3.mixture.                                        -
0.45066117
## Superplasticizer..component.5..kg.in.a.m.3.mixture.
0.22269123
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.                             -
0.17848096
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.
1.00000000
## Age..day.                                                                        -
```

```
0.15609470
## Concrete.compressive.strength.MPa..megapascals..                                              -
0.16724125
## new_input.1                                                                                  -
0.50161500
## new_input.2                                                                                  -
0.22272585
##                                                            Age..day.
## Cement..component.1..kg.in.a.m.3.mixture.                  0.08194602
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture. -0.04424602
## Fly.Ash..component.3..kg.in.a.m.3.mixture.             -0.15437052
## Water...component.4..kg.in.a.m.3.mixture.               0.27761822
## Superplasticizer..component.5..kg.in.a.m.3.mixture.    -0.19270003
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.   -0.00301588
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.      -0.15609470
## Age..day.                                               1.00000000
## Concrete.compressive.strength.MPa..megapascals..        0.32887300
## new_input.1                                             0.31185373
## new_input.2                                             0.08194726
##                                                        Concrete.compressive.strength.MPa..mega
pascals..
## Cement..component.1..kg.in.a.m.3.mixture.
0.4978319
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
0.1348293
## Fly.Ash..component.3..kg.in.a.m.3.mixture.                                                      -
0.1057549
## Water...component.4..kg.in.a.m.3.mixture.                                                       -
0.2896334
## Superplasticizer..component.5..kg.in.a.m.3.mixture.
0.3660788
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.                                            -
0.1649346
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.                                               -
0.1672412
## Age..day.
0.3288730
## Concrete.compressive.strength.MPa..megapascals..
1.0000000
## new_input.1                                                                                     -
0.2106030
## new_input.2
0.4978356
##                                                        new_input.1 new_input.2
## Cement..component.1..kg.in.a.m.3.mixture.               0.06254016  0.99999999
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.   0.06088931 -0.27519683
## Fly.Ash..component.3..kg.in.a.m.3.mixture.             -0.31363888 -0.39748082
## Water...component.4..kg.in.a.m.3.mixture.               0.98580847 -0.08158315
## Superplasticizer..component.5..kg.in.a.m.3.mixture.    -0.64121506  0.09238409
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.   -0.17193912 -0.10935760
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.      -0.50161500 -0.22272585
## Age..day.                                               0.31185373  0.08194726
## Concrete.compressive.strength.MPa..megapascals..       -0.21060304  0.49783558
```

```
## new_input.1                                                   1.00000000  0.06254379
## new_input.2                                                   0.06254379  1.00000000
```

By observing the predictor Variables we can see that there is strong coorelation between cement (component 1) (kg in a m^3 mixture) and new_input.2 (0.99999) almost one.

```
multiple.regression_new<-lm(Concrete.compressive.strength.MPa..megapascals..~.,data=concrete.dat
a_new)
summary(multiple.regression_new)
```

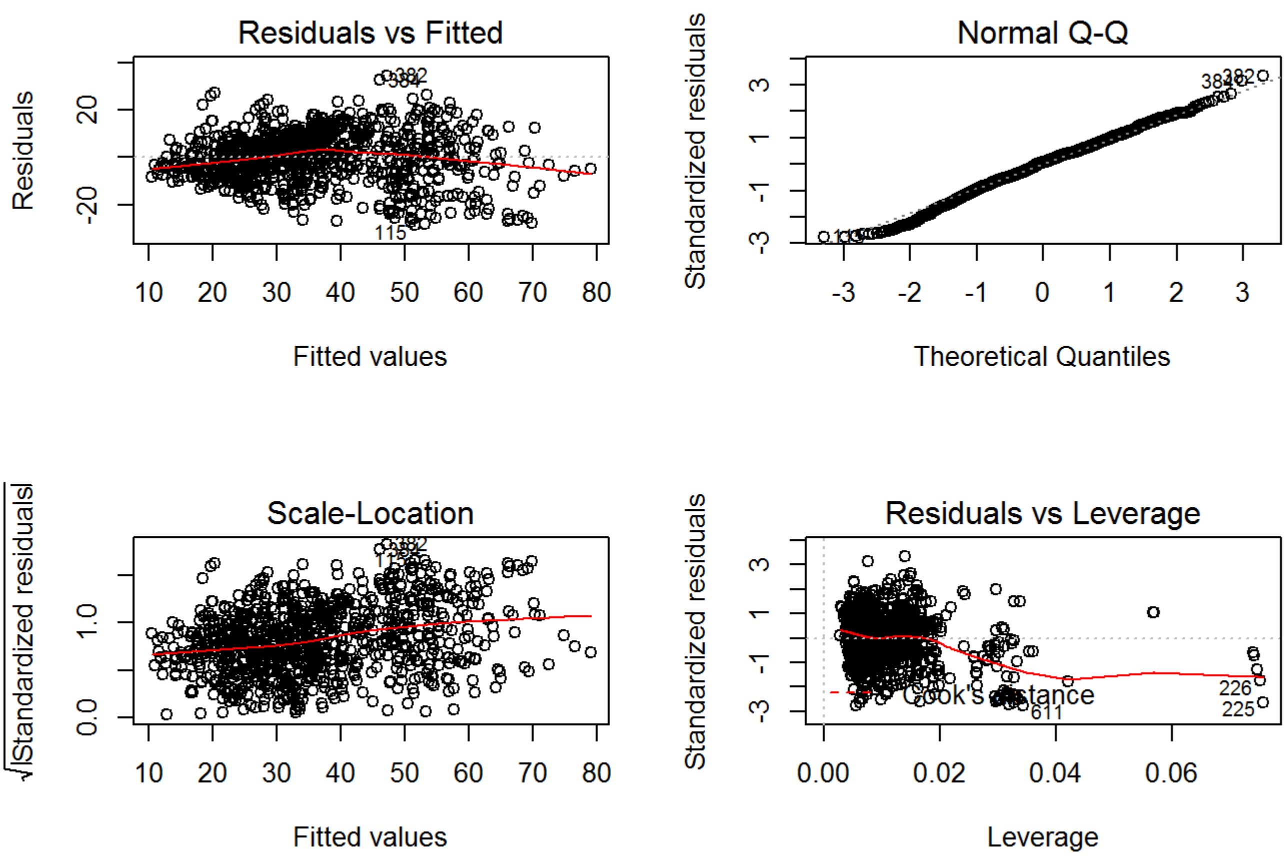
```
##
## Call:
## lm(formula = Concrete.compressive.strength.MPa..megapascals.. ~
##     ., data = concrete.data_new)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.668  -6.309   0.693   6.609  34.527
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                            -1.131e+01  2.339e+02
## Cement..component.1..kg.in.a.m.3.mixture.               1.463e+00  2.296e+01
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.  1.036e-01  1.033e-02
## Fly.Ash..component.3..kg.in.a.m.3.mixture.              8.758e-02  1.287e-02
## Water...component.4..kg.in.a.m.3.mixture.              -1.266e-01  1.998e-01
## Superplasticizer..component.5..kg.in.a.m.3.mixture.     2.940e-01  9.501e-02
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.    1.806e-02  9.445e-03
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.       1.983e-02  1.104e-02
## Age..day.                                               1.144e-01  5.629e-03
## new_input.1                                            -6.618e-04  5.508e-03
## new_input.2                                            -4.474e+00  7.653e+01
##                                                        t value Pr(>|t|)
## (Intercept)                                             -0.048  0.96146
## Cement..component.1..kg.in.a.m.3.mixture.                0.064  0.94922
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.  10.027  < 2e-16 ***
## Fly.Ash..component.3..kg.in.a.m.3.mixture.               6.807  1.7e-11 ***
## Water...component.4..kg.in.a.m.3.mixture.               -0.634  0.52655
## Superplasticizer..component.5..kg.in.a.m.3.mixture.      3.095  0.00202 **
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.     1.912  0.05612 .
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.        1.796  0.07273 .
## Age..day.                                               20.325  < 2e-16 ***
## new_input.1                                             -0.120  0.90439
## new_input.2                                             -0.058  0.95339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.41 on 1019 degrees of freedom
## Multiple R-squared:  0.6155, Adjusted R-squared:  0.6118
## F-statistic: 163.1 on 10 and 1019 DF,  p-value: < 2.2e-16
```

When we are adding Two new Variables(new_input.1,new_input.2) then significant variables are Blast Furnance, Fly ash ,super Plasticizer,Age.

We don't want to keep newely added Variables(new_input.1,new_input.2) because both variables are insignificant. when we are adding to the model.

we can see that Value of R-square is : 0.6155. and value of Adjusted Rsquare is:0.6118 . we can see that by adding this two new variable value of adjusted RSquare decreses.

```
par(mfrow=c(2,2))
plot(multiple.regression_new)
```



Question 5

You have decided that you will only allow three input variables in the model to make it simple. Choose the most appropriate three input variables and justify your answer along with results.?

subsets selection

```
library(leaps)
best.fit <-regsubsets(Concrete.compressive.strength.MPa..megapascals..~.,data=concrete.data_new,
nvmax=3)
reg.summary=summary(best.fit)
reg.summary
```

```
## Subset selection object
## Call: regsubsets.formula(Concrete.compressive.strength.MPa..megapascals.. ~
##     ., data = concrete.data_new, nvmax = 3)
## 10 Variables  (and intercept)
##                                                        Forced in Forced out
## Cement..component.1..kg.in.a.m.3.mixture.                  FALSE      FALSE
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.      FALSE      FALSE
## Fly.Ash..component.3..kg.in.a.m.3.mixture.                 FALSE      FALSE
## Water...component.4..kg.in.a.m.3.mixture.                  FALSE      FALSE
## Superplasticizer..component.5..kg.in.a.m.3.mixture.        FALSE      FALSE
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.       FALSE      FALSE
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.          FALSE      FALSE
## Age..day.                                                  FALSE      FALSE
## new_input.1                                                FALSE      FALSE
## new_input.2                                                FALSE      FALSE
## 1 subsets of each size up to 3
## Selection Algorithm: exhaustive
##          Cement..component.1..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Fly.Ash..component.3..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Water...component.4..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Superplasticizer..component.5..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) "*"
## 3  ( 1 ) "*"
##          Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Fine.Aggregate..component.7..kg.in.a.m.3.mixture. Age..day.
## 1  ( 1 ) " "                                               " "
## 2  ( 1 ) " "                                               " "
## 3  ( 1 ) " "                                               "*"
##          new_input.1 new_input.2
## 1  ( 1 ) " "         "*"
## 2  ( 1 ) " "         "*"
## 3  ( 1 ) " "         "*"
```

Subset selection refers to the task of finding a small number of the available independant variables that does a good job of predicting the dependant variables.

Exhaustive search of good model is possible when we have upto 15 independant variables. After that we need to use searching algorithm like forward selection and backward elimination.

We are making best subset selection on the entire dataset with a maximum of 3 variables and fitting the model for different subsets.

In the summary of the model we can see Forced In and Forced Out for all 10 variables except Response Variable (output variable).This tells us if are setting any variable to forcefully include into the model or forcefully exculde out of the model.

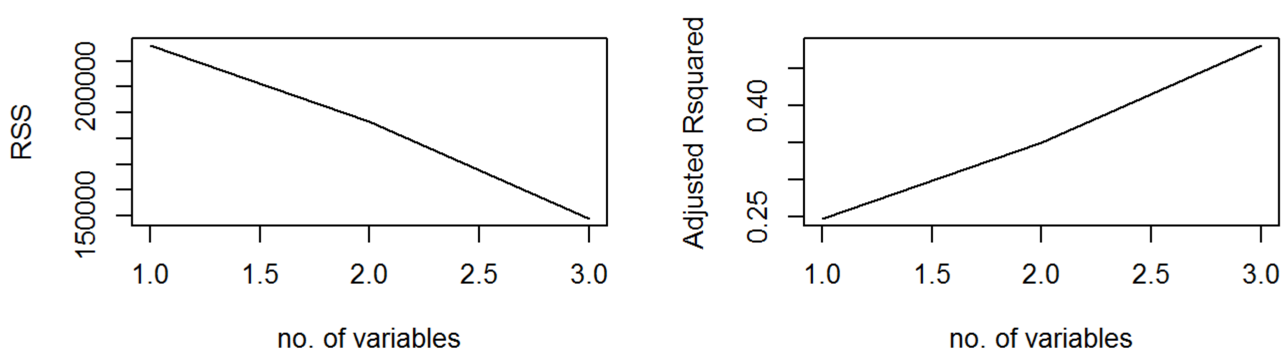
If we set any variable Force In as TRUE that will definetely be included into the model and setting Force Out to TRUE will exclude the variable from the model.

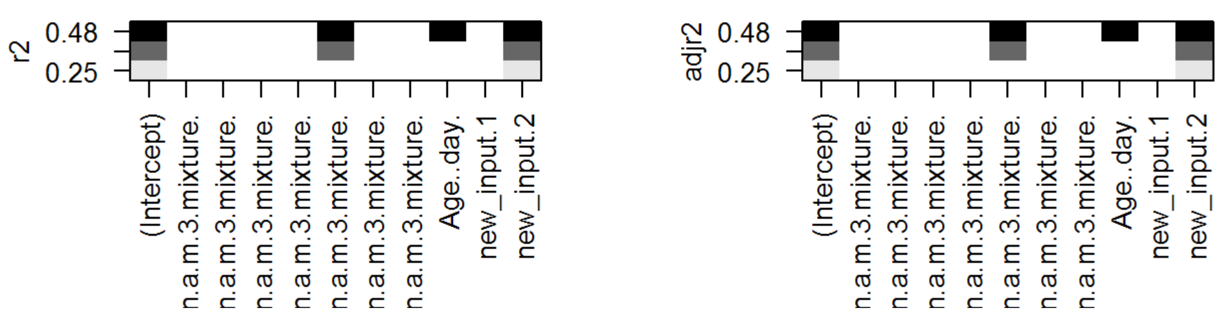
Three best predictor variable : **new_input.2,superplasticizer,Age(day).**

```
par(mfrow=c(2,2))
plot(reg.summary$rss,xlab="no. of variables",ylab="RSS",type="l")
plot(reg.summary$adjr2,xlab="no. of variables",ylab="Adjusted Rsquared",type="l")
which.max(reg.summary$adjr2)
```

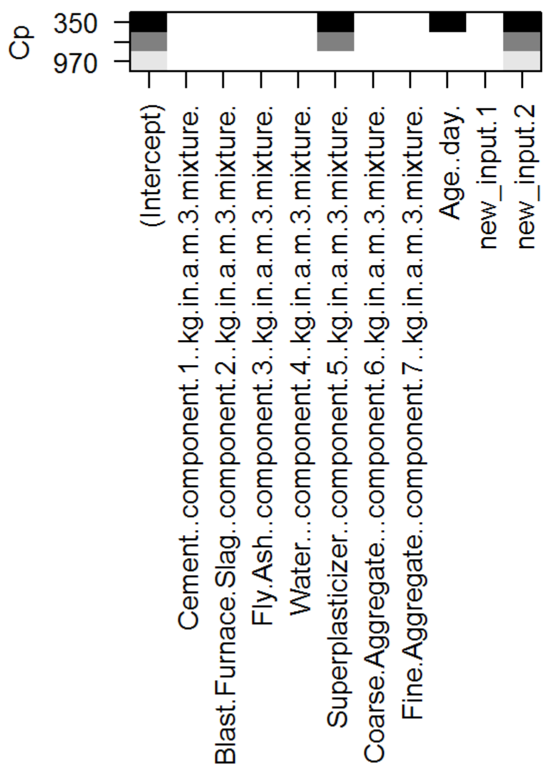
```
## [1] 3
```

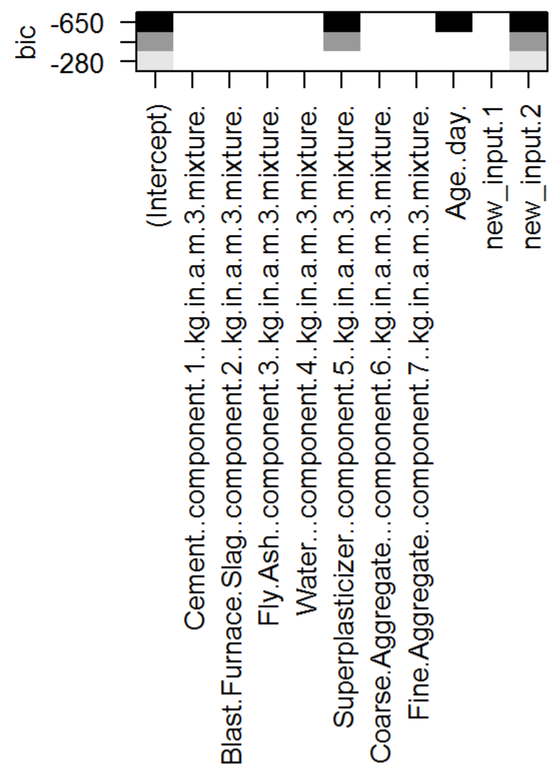
```
plot(best.fit,scale="r2")
plot(best.fit,scale="adjr2")
```





```
plot(best.fit,scale="Cp")
plot(best.fit,scale="bic")
```





Forward Selection

```
library(leaps)
forward.fit <-regsubsets(Concrete.compressive.strength.MPa..megapascals..~.,data=concrete.data_n
ew,method="forward",nvmax=3)
regfit<-summary(forward.fit)
regfit
```

```
## Subset selection object
## Call: regsubsets.formula(Concrete.compressive.strength.MPa..megapascals.. ~
##     ., data = concrete.data_new, method = "forward", nvmax = 3)
## 10 Variables  (and intercept)
##                                                        Forced in Forced out
## Cement..component.1..kg.in.a.m.3.mixture.                  FALSE      FALSE
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.     FALSE      FALSE
## Fly.Ash..component.3..kg.in.a.m.3.mixture.                 FALSE      FALSE
## Water...component.4..kg.in.a.m.3.mixture.                  FALSE      FALSE
## Superplasticizer..component.5..kg.in.a.m.3.mixture.       FALSE      FALSE
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.      FALSE      FALSE
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.         FALSE      FALSE
## Age..day.                                                  FALSE      FALSE
## new_input.1                                                FALSE      FALSE
## new_input.2                                                FALSE      FALSE
## 1 subsets of each size up to 3
## Selection Algorithm: forward
##          Cement..component.1..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Fly.Ash..component.3..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Water...component.4..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Superplasticizer..component.5..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) "*"
## 3  ( 1 ) "*"
##          Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Fine.Aggregate..component.7..kg.in.a.m.3.mixture. Age..day.
## 1  ( 1 ) " "                                               " "
## 2  ( 1 ) " "                                               " "
## 3  ( 1 ) " "                                               "*"
##          new_input.1 new_input.2
## 1  ( 1 ) " "         "*"
## 2  ( 1 ) " "         "*"
## 3  ( 1 ) " "         "*"
```

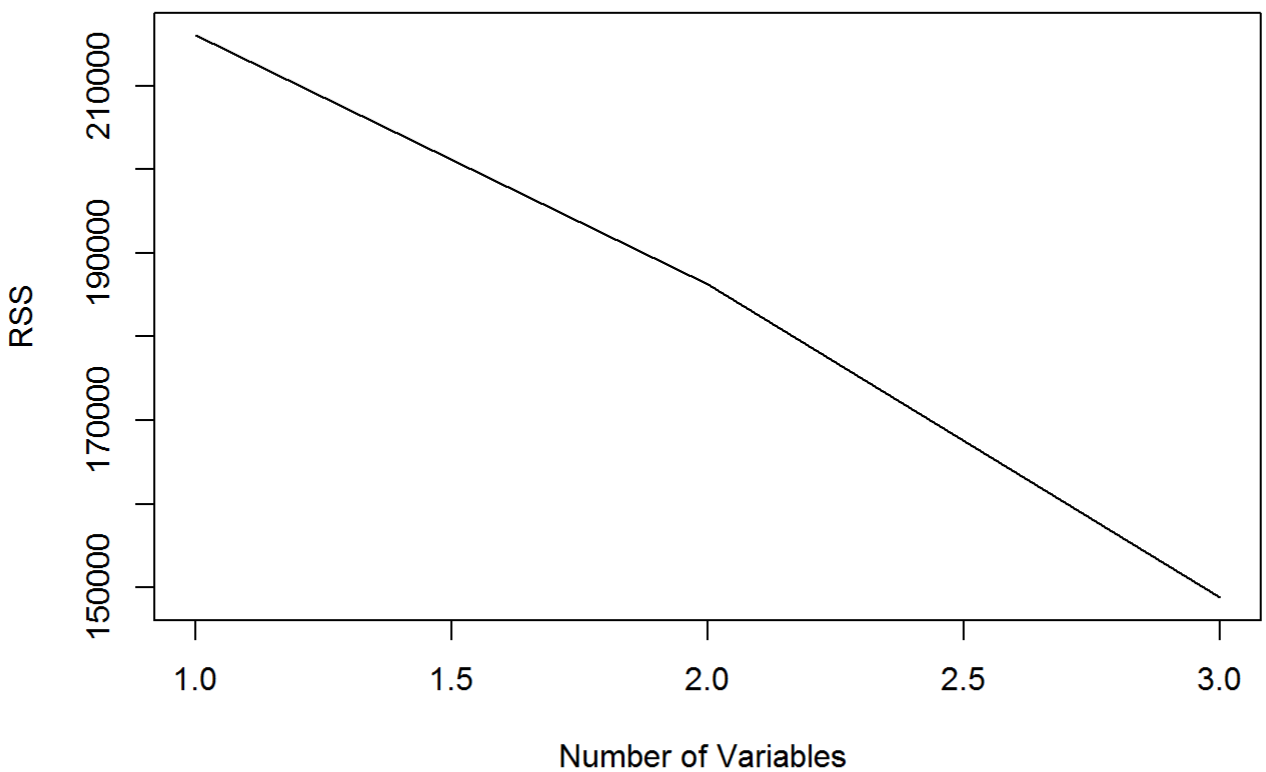
In this approach we start with a null model i.e. it does not have any input variables. Then, at every step we add a one new independant variable which helps us to best fit the data. This process is repeated multiple times.

Three best predictor variable : **new_input.2,superplasticizer,Age(day).**

```
regfit$rsq
```

```
## [1] 0.2478403 0.3511778 0.4817573
```

```
plot(regfit$rss ,xlab="Number of Variables ",ylab="RSS", type="l")
```



As no of predictor increases RSS decreases(As Expected)

Backward Regression

```
library(leaps)
backward.fit <-regsubsets(Concrete.compressive.strength.MPa..megapascals..~.,data=concrete.data_
new,method="backward",nvmax=3)
summary(backward.fit)
```

```
## Subset selection object
## Call: regsubsets.formula(Concrete.compressive.strength.MPa..megapascals.. ~
##     ., data = concrete.data_new, method = "backward", nvmax = 3)
## 10 Variables  (and intercept)
##                                                        Forced in Forced out
## Cement..component.1..kg.in.a.m.3.mixture.                  FALSE      FALSE
## Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.      FALSE      FALSE
## Fly.Ash..component.3..kg.in.a.m.3.mixture.                 FALSE      FALSE
## Water...component.4..kg.in.a.m.3.mixture.                  FALSE      FALSE
## Superplasticizer..component.5..kg.in.a.m.3.mixture.        FALSE      FALSE
## Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.       FALSE      FALSE
## Fine.Aggregate..component.7..kg.in.a.m.3.mixture.          FALSE      FALSE
## Age..day.                                                  FALSE      FALSE
## new_input.1                                                FALSE      FALSE
## new_input.2                                                FALSE      FALSE
## 1 subsets of each size up to 3
## Selection Algorithm: backward
##          Cement..component.1..kg.in.a.m.3.mixture.
## 1  ( 1 ) "*"
## 2  ( 1 ) "*"
## 3  ( 1 ) "*"
##          Blast.Furnace.Slag..component.2..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Fly.Ash..component.3..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Water...component.4..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) "*"
##          Superplasticizer..component.5..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Coarse.Aggregate...component.6..kg.in.a.m.3.mixture.
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
##          Fine.Aggregate..component.7..kg.in.a.m.3.mixture. Age..day.
## 1  ( 1 ) " "                                               " "
## 2  ( 1 ) " "                                               "*"
## 3  ( 1 ) " "                                               "*"
##          new_input.1 new_input.2
## 1  ( 1 ) " "         " "
## 2  ( 1 ) " "         " "
## 3  ( 1 ) " "         " "
```

In this approach we start with a full model i.e. it does fits all input variables. Then, at every step we remove a one independant variable which is least important in determining the output variable. This process is repeated multiple times.

Three best predictor variable : **cement,water,Age(day).**

Let’s fit the model with the best three input variables.

```
#Forward and Best Subset
best1<-lm(Concrete.compressive.strength.MPa..megapascals..~new_input.2+Age..day.+Superplasticize
r..component.5..kg.in.a.m.3.mixture.,data=concrete.data_new)
summary(best1)
```

```
##
## Call:
## lm(formula = Concrete.compressive.strength.MPa..megapascals.. ~
##     new_input.2 + Age..day. + Superplasticizer..component.5..kg.in.a.m.3.mixture.,
##     data = concrete.data_new)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.281  -8.073  -0.541   7.546  45.670
##
## Coefficients:
##                                                       Estimate Std. Error t value
## (Intercept)                                           4.399058   1.176771   3.738
## new_input.2                                           0.229530   0.012090  18.985
## Age..day.                                             0.097901   0.006089  16.078
## Superplasticizer..component.5..kg.in.a.m.3.mixture. 1.111936   0.064446  17.254
##                                                       Pr(>|t|)
## (Intercept)                                           0.000196 ***
## new_input.2                                            < 2e-16 ***
## Age..day.                                              < 2e-16 ***
## Superplasticizer..component.5..kg.in.a.m.3.mixture.  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.04 on 1026 degrees of freedom
## Multiple R-squared:  0.4818, Adjusted R-squared:  0.4802
## F-statistic: 317.9 on 3 and 1026 DF,  p-value: < 2.2e-16
```

Rsquare is 0.4818

Adjusted Rsquare is 0.4802

All the Predictors are Significant.