Fahim Hoq

Ice Cream Cone Project

Data: Flour is used in baking ice cream cones. The flour content contains ash, moisture, protein etc. The percentage of these ingredients are given in the data below as well as the viscosity of the flour used in baking ice cream cones. My objective is to find a model that best describes the relationship between the response variable viscosity with the explanatory variables: moisture, protein, and ash.

$Viscosity^{\lambda} = \beta_0 + \beta_1 moisture + \beta_2 protein + \beta_3 ash.$

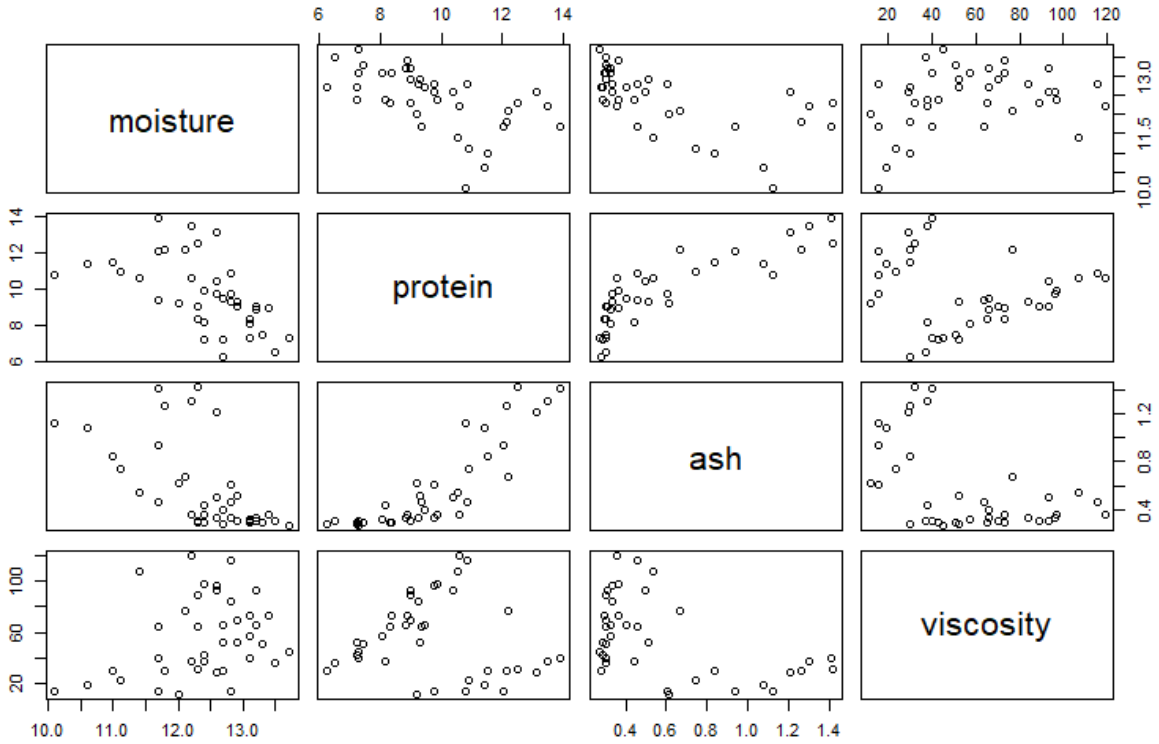
codeNum	moisture	protein	ash	viscosity
1	13.5	6.5	0.306	37
2	13.3	7.45	0.3	51
3	12.7	9.45	0.405	66
4	11.7	12.05	0.94	15
5	12.7	6.25	0.277	30
6	12.7	7.2	0.287	52
7	12.4	7.2	0.29	43
8	12.3	8.3	0.3	65
9	13.1	8.35	0.298	73
10	13.2	9	0.31	93
11	12.3	9	0.305	89
12	12.2	10.6	0.356	119
13	12.4	8.15	0.44	38
14	11.7	9.35	0.46	64
15	11.4	10.55	0.538	107
16	11.1	10.9	0.74	23
17	11	11.5	0.84	30
18	10.1	10.8	1.12	15
19	10.6	11.4	1.08	19
20	13.7	7.3	0.27	45
21	13.4	8.9	0.362	73
22	13.1	7.3	0.302	40
23	13.1	8.05	0.324	57
24	12.9	9.3	0.514	52
25	13.2	8.85	0.328	66
26	12.9	9	0.304	70
27	12.8	9.25	0.336	84
28	12.6	9.75	0.332	96
29	12.4	9.85	0.362	97
30	12.8	10.85	0.46	116
31	12	9.2	0.616	12
32	12.6	13.1	1.208	29
33	12.6	10.4	0.498	93
34	12.1	12.2	0.67	77
35	12.3	12.5	1.416	32
36	11.7	13.9	1.41	40
37	11.8	12.15	1.264	30
38	12.2	13.5	1.3	38
39	12.8	9.75	0.602	15

Descriptive Statistics

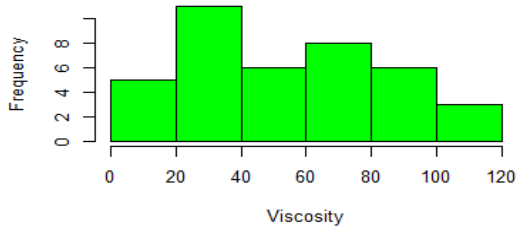
```
 moisture        protein           ash           viscosity
  Min.   :10.10   Min.   : 6.250   Min.   :0.2700   Min.   : 12.00
  1st Qu.:12.05   1st Qu.: 8.325   1st Qu.:0.3055   1st Qu.: 31.00
  Median :12.60   Median : 9.350   Median :0.4050   Median : 52.00
  Mean   :12.39   Mean   : 9.721   Mean   :0.5762   Mean   : 56.18
  3rd Qu.:12.90   3rd Qu.:10.875   3rd Qu.:0.7050   3rd Qu.: 75.00
  Max.   :13.70   Max.   :13.900   Max.   :1.4160   Max.   :119.00
```

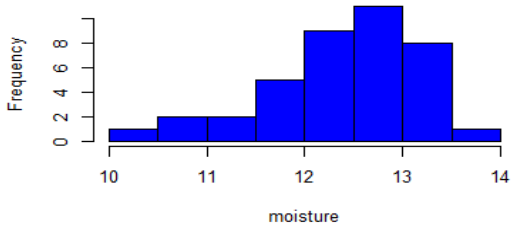
>

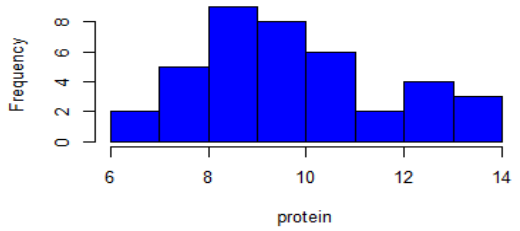
Scatter plots

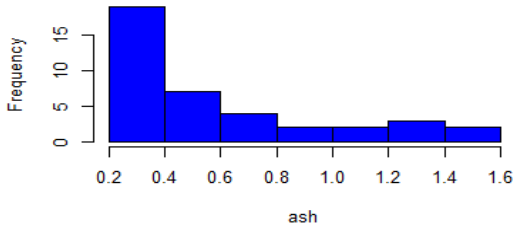


Histograms

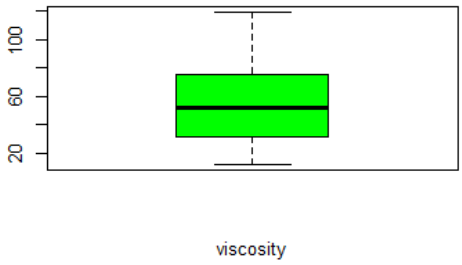

Frequency
0 2 4 6 8
0 20 40 60 80 100 120
Viscosity

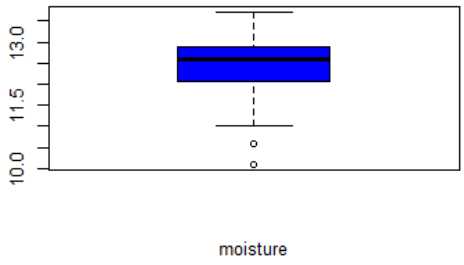

Frequency
0 2 4 6 8
10 11 12 13 14
moisture

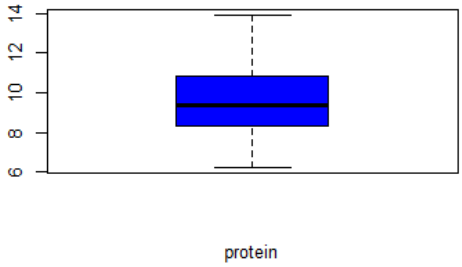

Frequency
0 2 4 6 8
6 8 10 12 14
protein

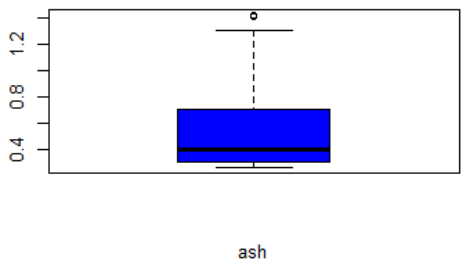

Frequency
0 5 10 15
0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6
ash

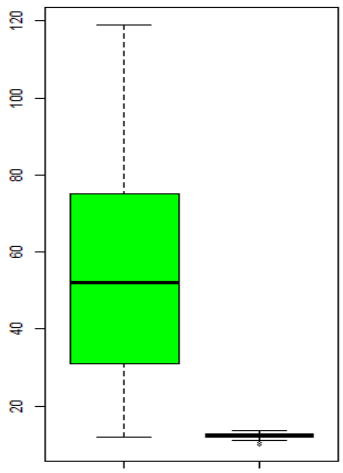
Boxplots


20 60 100
viscosity

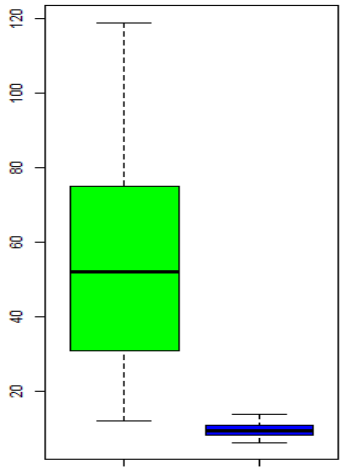

10.0 11.5 13.0
moisture


6 8 10 12 14
protein

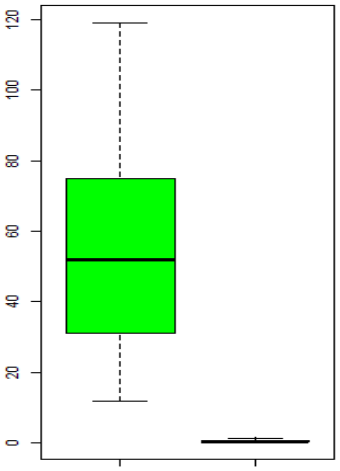

0.4 0.8 1.2
ash



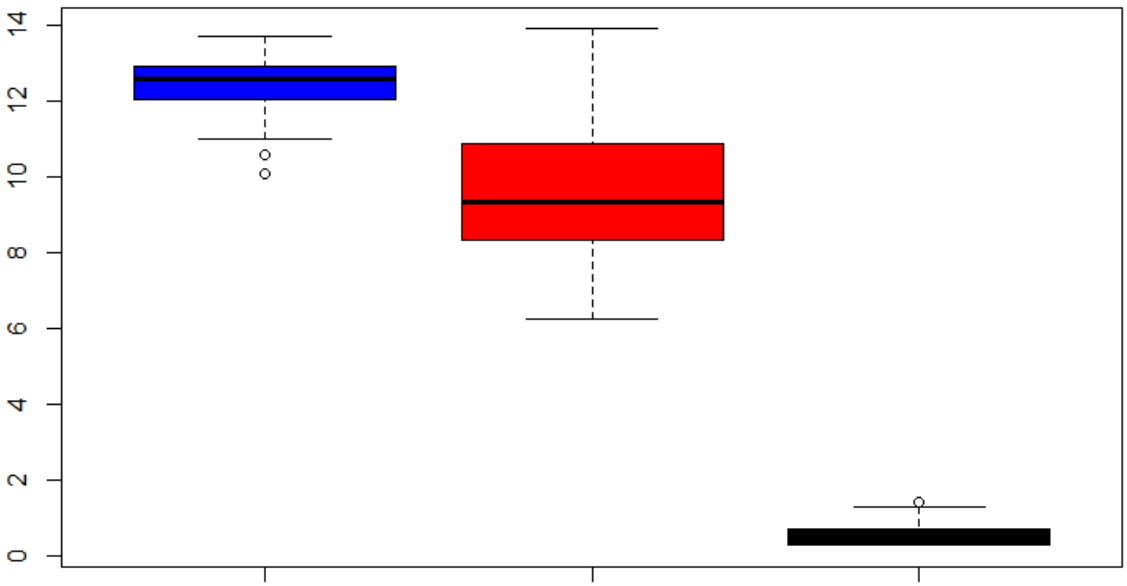
viscosity vs moisture



viscosity vs protein



viscosity vs ash



moisture vs protein vs ash

Regression Analysis:

First we use regression with viscosity and only one explanatory variables before trying the analysis of viscosity with all the other variables.

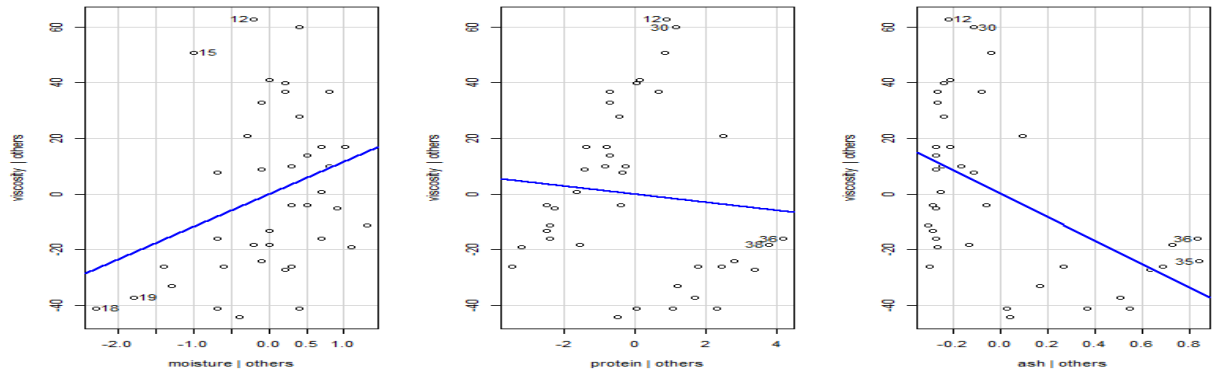
```
> m1 = lm(viscosity~moisture, data=Ice)
> summary(m1)

Call:
lm(formula = viscosity ~ moisture, data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-45.930 -18.722  -8.032  19.519  65.105

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -89.152     73.001  -1.221   0.2297
moisture      11.725      5.878   1.995   0.0535 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.91 on 37 degrees of freedom
Multiple R-squared:  0.09711,  Adjusted R-squared:  0.0727
F-statistic: 3.979 on 1 and 37 DF,  p-value: 0.05347



>
> m2 = lm(viscosity~protein, data=Ice)
> summary(m2)

Call:
lm(formula = viscosity ~ protein, data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-44.936 -22.456  -7.845  20.027  64.100

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   70.316     24.917   2.822  0.00763 **
protein       -1.454      2.514  -0.578  0.56650
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.29 on 37 degrees of freedom
Multiple R-squared:  0.008961, Adjusted R-squared:  -0.01782
F-statistic: 0.3345 on 1 and 37 DF,  p-value: 0.5665



>
> m3 = lm(viscosity~ash, data=Ice)
> summary(m3)

Call:
lm(formula = viscosity ~ ash, data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-42.501 -21.095  -0.569  18.322  54.929

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   80.444      7.947  10.123 3.28e-12 ***
```

```
ash          -42.115     11.715  -3.595 0.000941 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26.2 on 37 degrees of freedom
Multiple R-squared:  0.2589,   Adjusted R-squared:  0.2389
F-statistic: 12.92 on 1 and 37 DF,  p-value: 0.000941
```

```
>
> m4 = lm(viscosity~moisture+protein+ash,data=Ice)
> summary(m4)

Call:
lm(formula = viscosity ~ moisture + protein + ash, data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-40.120  -7.873   1.617  10.604  33.449

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -115.359     63.286  -1.823   0.0769 .
moisture       4.150      4.380   0.948   0.3498
protein       19.993      2.761   7.240 1.88e-08 ***
ash         -128.861     15.365  -8.386 6.84e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.04 on 35 degrees of freedom
Multiple R-squared:  0.7033,   Adjusted R-squared:  0.6779
F-statistic: 27.65 on 3 and 35 DF,  p-value: 2.383e-09
```

We see that only ash is significant when we do the regression with only one explanatory variables but when we combine all the variables together we find that only moisture is non-significant.

Added variable Plots:



Interaction terms: Next we analyze the interaction terms.

```
> n1 =  lm(viscosity~moisture+protein+ash + I(moisture*protein)+I(protein*ash)+
I(ash*moisture)+I(moisture*ash*protein) ,data=Ice)
> summary(n1)

Call:
lm(formula = viscosity ~ moisture + protein + ash + I(moisture *
    protein) + I(protein * ash) + I(ash * moisture) + I(moisture *
    ash * protein), data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-29.935  -8.674   0.443   5.719  41.050

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                 404.219    702.811   0.575    0.569
moisture                    -31.481     55.963  -0.563    0.578
protein                     -55.387     74.444  -0.744    0.462
ash                         777.824   1335.926   0.582    0.565
I(moisture * protein)         5.613      5.933   0.946    0.351
I(protein * ash)            -44.588    119.145  -0.374    0.711
I(ash * moisture)           -90.394    111.824  -0.808    0.425
I(moisture * ash * protein)   4.929      9.909   0.497    0.622

Residual standard error: 16.52 on 31 degrees of freedom
Multiple R-squared:  0.753,    Adjusted R-squared:  0.6972
F-statistic:  13.5 on 7 and 31 DF,  p-value: 7.432e-08



>
> n2 =  lm(viscosity~moisture+protein+ash +I(ash*moisture) ,data=Ice)
> summary(n2)

Call:
lm(formula = viscosity ~ moisture + protein + ash + I(ash * moisture),
    data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-38.928  -8.939   2.525   9.813  35.485

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -198.670    146.775  -1.354    0.185
moisture           10.298     10.708   0.962    0.343
protein            20.878      3.119   6.693  1.1e-07 ***
ash               -27.043    162.265  -0.167    0.869
I(ash * moisture)  -8.631     13.691  -0.630    0.533
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.19 on 34 degrees of freedom
Multiple R-squared:  0.7067,   Adjusted R-squared:  0.6722
F-statistic: 20.48 on 4 and 34 DF,  p-value: 1.143e-08

> n3 = lm(viscosity~moisture+protein+ash +I(protein*ash) ,data=Ice)
```

```
> summary(n3)

Call:
lm(formula = viscosity ~ moisture + protein + ash + I(protein *
    ash), data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-36.637  -7.205   0.299   7.424  32.563

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        -7.5763    95.1763  -0.080 0.937020
moisture           -0.5296     5.3207  -0.100 0.921294
protein            16.1701     3.7275   4.338 0.000122 ***
ash              -258.5075    87.9637  -2.939 0.005883 **
I(protein * ash)   10.0253     6.7011   1.496 0.143859
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.75 on 34 degrees of freedom
Multiple R-squared:  0.7216,   Adjusted R-squared:  0.6889
F-statistic: 22.03 on 4 and 34 DF,  p-value: 4.802e-09

> n4 = lm(viscosity~moisture+protein+ash +I(protein*moisture) ,data=Ice)
> summary(n4)

Call:
lm(formula = viscosity ~ moisture + protein + ash + I(protein *
    moisture), data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-40.875  -7.137   1.783  11.170  33.677

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           -1.6066   379.0595  -0.004    0.997
moisture              -4.7358    29.5193  -0.160    0.873
protein                8.4699    37.9490   0.223    0.825
ash                 -126.4906    17.4062  -7.267 2.06e-08 ***
I(protein * moisture)    0.8948     2.9387   0.304    0.763
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.27 on 34 degrees of freedom
Multiple R-squared:  0.7041,   Adjusted R-squared:  0.6693
F-statistic: 20.23 on 4 and 34 DF,  p-value: 1.325e-08

> n5= lm(viscosity~moisture+protein+ash +I(moisture*ash*protein) ,data=Ice)
> summary(n5)

Call:
lm(formula = viscosity ~ moisture + protein + ash + I(moisture *
    ash * protein), data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-36.827  -6.377   0.495   8.099  31.174

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                11.6128   133.1746   0.087 0.931024
```

```
moisture                      -2.9882     7.9085  -0.378 0.707896
protein                       16.7725     4.0539   4.137 0.000218 ***
ash                         -211.2295    77.5978  -2.722 0.010157 *
I(moisture * ash * protein)    0.5404     0.4991   1.083 0.286509
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17 on 34 degrees of freedom
Multiple R-squared:  0.7132,   Adjusted R-squared:  0.6794
F-statistic: 21.14 on 4 and 34 DF,  p-value: 7.891e-09



>
```

Here, we see that whether we take all the interaction terms together or separately, they are always non-significant.

Finally, we try using quadratic terms to see if we can increase R-squared value.

```
>
> m5 = lm(viscosity~moisture+protein+ash + I(moisture^2) + I(protein^2) + I(ash^2),
data=Ice)
> summary(m5)

Call:
lm(formula = viscosity ~ moisture + protein + ash + I(moisture^2) +
    I(protein^2) + I(ash^2), data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-20.339  -7.395  -0.687   5.288  34.704

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    238.2963   428.0931   0.557  0.58164
moisture       -51.4417    68.7103  -0.749  0.45952
protein         41.9744    13.5628   3.095  0.00407 **
ash           -393.4947    51.8234  -7.593 1.19e-08 ***
I(moisture^2)    1.9179     2.8434   0.675  0.50483
I(protein^2)    -0.9354     0.7316  -1.279  0.21020
I(ash^2)       152.5965    28.9773   5.266 9.16e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.01 on 32 degrees of freedom
Multiple R-squared:  0.8418,   Adjusted R-squared:  0.8122
F-statistic: 28.39 on 6 and 32 DF,  p-value: 1.699e-11



>
> m6 = lm(viscosity~ protein+ash + I(protein^2) + I(ash^2), data=Ice)
> summary(m6)
```

```
Call:
lm(formula = viscosity ~ protein + ash + I(protein^2) + I(ash^2),
    data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-24.021  -6.894  -0.768   5.389  38.161

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -135.7863    57.7020  -2.353  0.02453 *
protein        46.9602    13.1215   3.579  0.00106 **
ash          -353.8118    43.6456  -8.106 1.88e-09 ***
I(protein^2)   -1.2578     0.6945  -1.811  0.07897 .
I(ash^2)      136.6221    26.4727   5.161 1.06e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.02 on 34 degrees of freedom
Multiple R-squared:  0.8317,   Adjusted R-squared:  0.8118
F-statistic: 41.99 on 4 and 34 DF,  p-value: 1.061e-12
```

```
>
> anova(m6,m5)
Analysis of Variance Table

Model 1: viscosity ~ protein + ash + I(protein^2) + I(ash^2)
Model 2: viscosity ~ moisture + protein + ash + I(moisture^2) + I(protein^2) +
    I(ash^2)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     34 5767.6
2     32 5418.5  2    349.08 1.0308 0.3683


>
```

The R-square value increases but, the moisture and its quadratic terms is still not significant, so we remove it from the model, we then compare the full model with the reduced model and find that it is not significant.

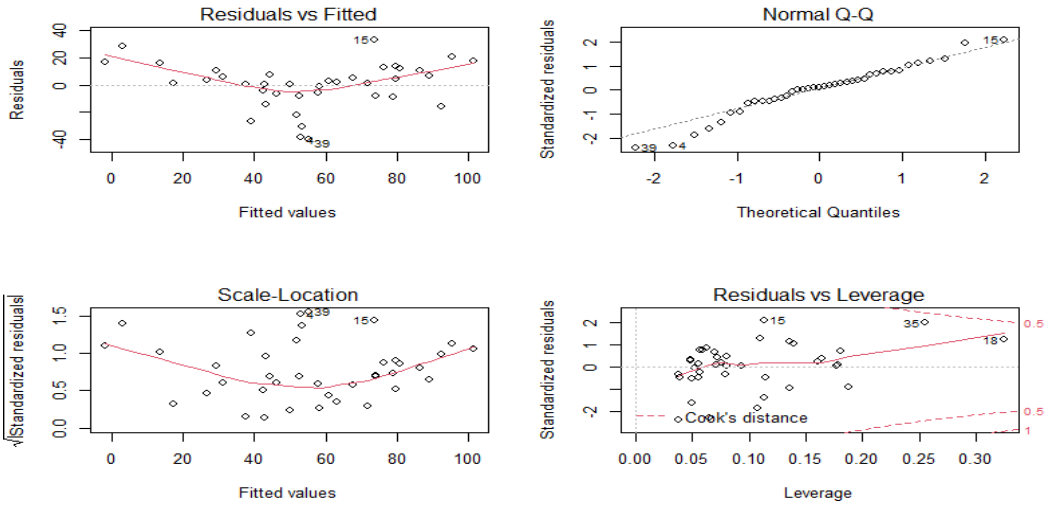
Multicollinearity

```
> vif(m4)
moisture  protein      ash
1.598264 3.811027 4.065172
```

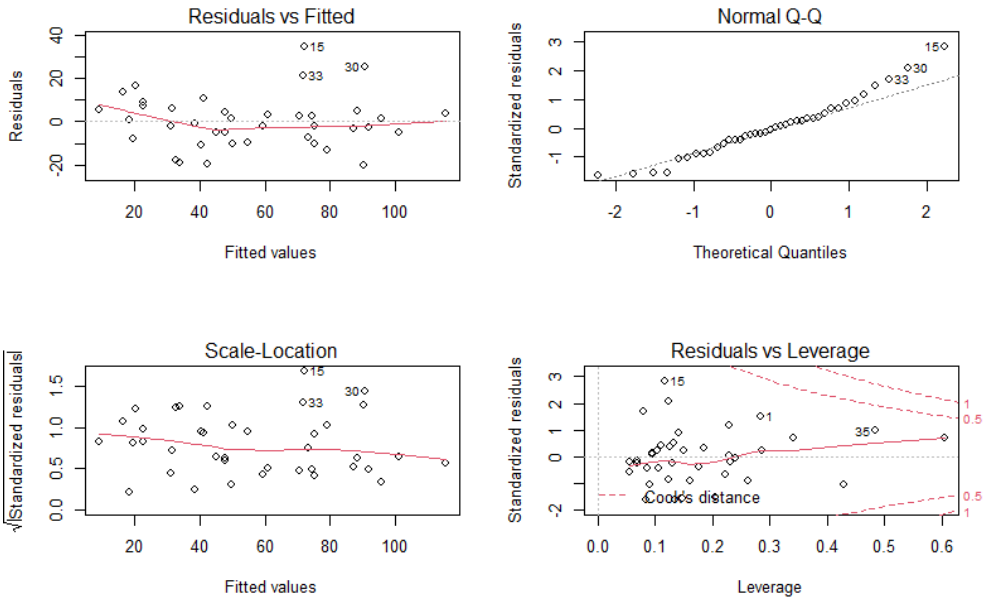
As all the values are below 5, there is no multicollinearity present in the data.

Plots: We then do the plots for all the models described above.

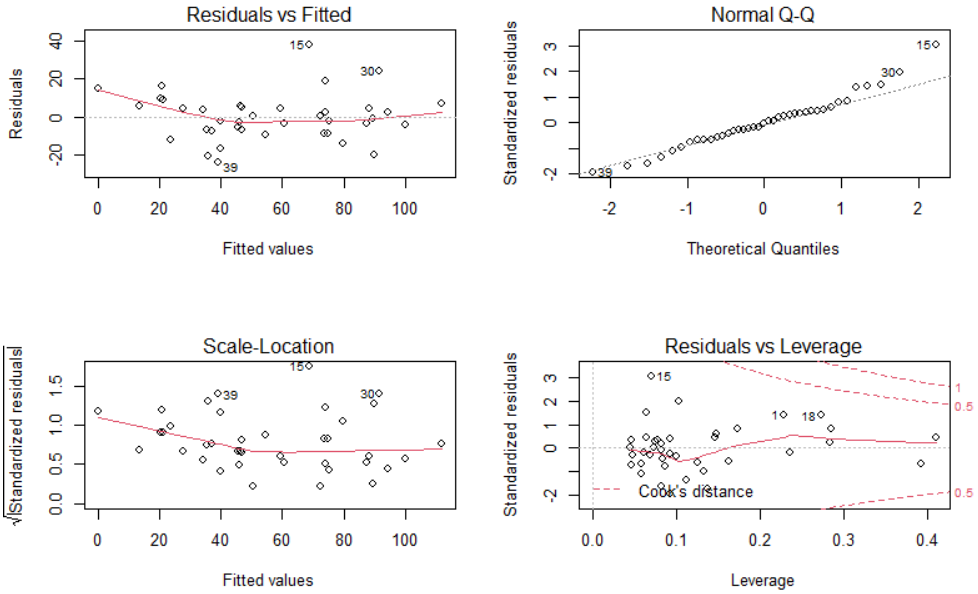
```
> plot(m4)
```




```
> plot(m5)
```

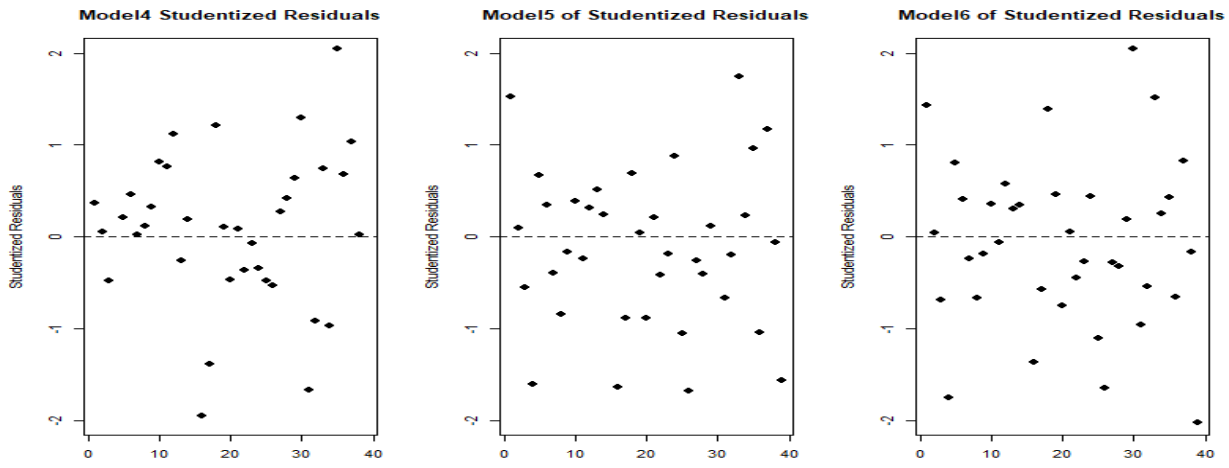



```
> plot(m6)
```



All the residuals vs leverage shows no outliers.

Studentized Residuals:



The studentized residuals does not show any outliers either.

```
> outlierTest(m6)
No Studentized residuals with Bonferroni p < 0.05
```

```
Largest |rstudent|:
   rstudent unadjusted p-value Bonferroni p
15 3.506785          0.0013306     0.051893
```

Moreover, I detected no outliers using the Bonferroni approach as well.

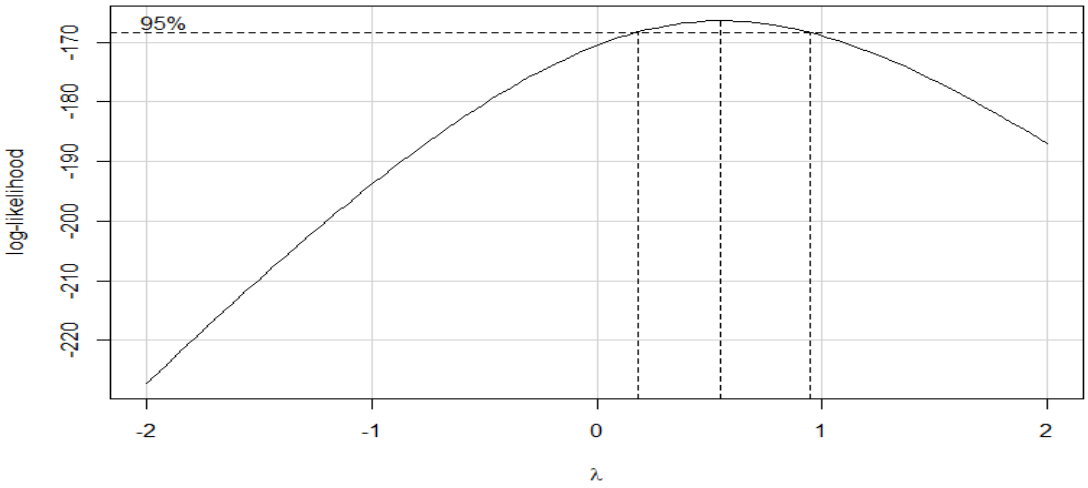
AIC/BIC:

```
> AIC(m4,m5,m6)
   df      AIC
m4  5 337.6392
m5  8 319.1039
m6  6 317.5388
> BIC(m4,m5,m6)
   df      BIC
m4  5 345.9570
m5  8 332.4124
m6  6 327.5202
```

The AIC and BIC values are very similar for the models without the quadratic terms, with the quadratic full model and the reduced models.

Box-Cox Transformations:

We then decide to apply the box cox in our reduced model and we find that $\lambda = 0.5$, so we use 0.25, 0.5, and 0.75 as possible transforations.




```
> m7 = lm(as.numeric(viscosity)^0.25 ~ protein+ash + I(protein^2) + I(ash^2), data=Ice)
> summary(m7)

Call:
lm(formula = as.numeric(viscosity)^0.25 ~ protein + ash + I(protein^2) +
    I(ash^2), data = Ice)

Residuals:
     Min       1Q   Median       3Q      Max
-0.39969 -0.07134  0.00497  0.09530  0.45117

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.960401   0.771194   1.245  0.22152
protein       0.481275   0.175371   2.744  0.00961 **
ash          -5.114680   0.583328  -8.768 3.02e-10 ***
I(protein^2) -0.010076   0.009282  -1.086  0.28532
I(ash^2)      2.074405   0.353810   5.863 1.30e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1741 on 34 degrees of freedom
Multiple R-squared:  0.8273,   Adjusted R-squared:  0.807
F-statistic: 40.73 on 4 and 34 DF,  p-value: 1.623e-12



>


> m8 = lm(as.numeric(viscosity)^0.5 ~ protein+ash + I(protein^2) + I(ash^2), data=Ice)
> summary(m8)

Call:
lm(formula = as.numeric(viscosity)^0.5 ~ protein + ash + I(protein^2) +
    I(ash^2), data = Ice)

Residuals:
     Min       1Q   Median       3Q      Max
```

```
-1.94132 -0.40263 -0.06008  0.44027  2.45256

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -3.36268    3.92579  -0.857  0.39769
protein        2.81035    0.89273   3.148  0.00341 **
ash          -26.03720    2.96945  -8.768 3.02e-10 ***
I(protein^2)  -0.06645    0.04725  -1.406  0.16874
I(ash^2)      10.39429    1.80109   5.771 1.71e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8861 on 34 degrees of freedom
Multiple R-squared:  0.836,    Adjusted R-squared:  0.8167
F-statistic: 43.33 on 4 and 34 DF,  p-value: 6.831e-13



>
> m9 = lm(as.numeric(viscosity)^0.75 ~ protein+ash + I(protein^2) + I(ash^2), data=Ice)
> summary(m9)

Call:
lm(formula = as.numeric(viscosity)^0.75 ~ protein + ash + I(protein^2) +
    I(ash^2), data = Ice)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1850 -1.7427 -0.2945  1.4895 10.1799

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   -28.2129    15.6492  -1.803  0.08028 .
protein        12.2122     3.5587   3.432  0.00159 **
ash          -101.0215    11.8370  -8.534 5.72e-10 ***
I(protein^2)   -0.3115     0.1884  -1.654  0.10739
I(ash^2)       39.6719     7.1796   5.526 3.56e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.532 on 34 degrees of freedom
Multiple R-squared:  0.8372,   Adjusted R-squared:  0.8181
F-statistic: 43.72 on 4 and 34 DF,  p-value: 6.024e-13



>
```

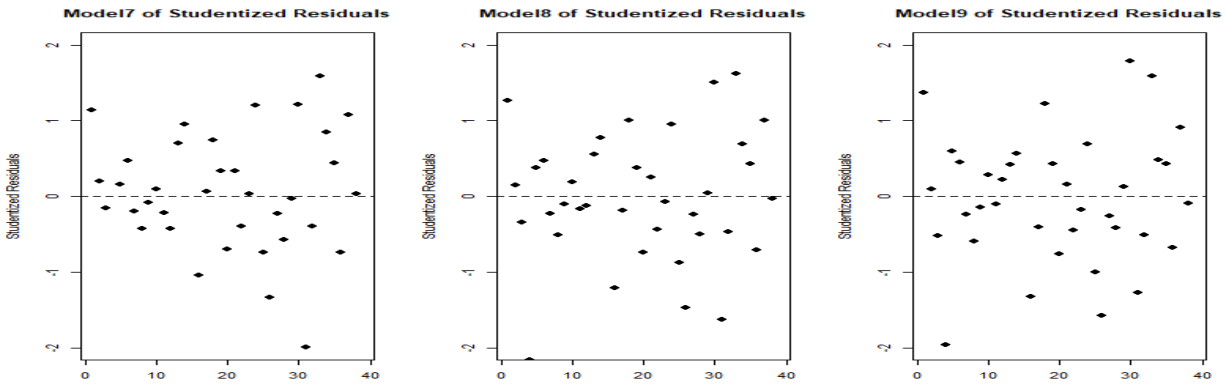
In all three cases we find that the protein^2 is not significant and thus, we reduce our model further.

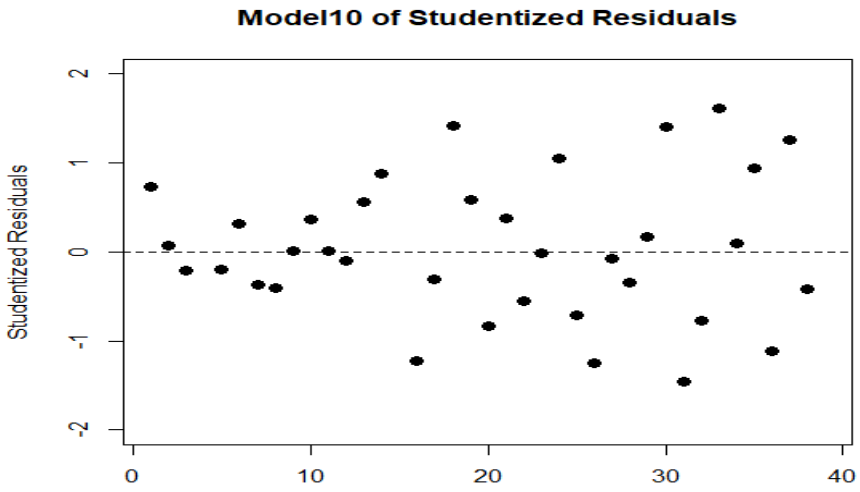
```
> m10 = lm(as.numeric(viscosity)^0.5 ~ protein+ash + I(ash^2), data=Ice)
> summary(m10)

Call:
lm(formula = as.numeric(viscosity)^0.5 ~ protein + ash + I(ash^2),
    data = Ice)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-1.88513 -0.41702 -0.01735  0.49718  2.44292

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.9410     1.1046   1.757   0.0876 .
protein       1.5734     0.1543  10.194 5.12e-12 ***
ash         -24.6818     2.8476  -8.668 3.11e-10 ***
I(ash^2)      9.1204     1.5783   5.779 1.51e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8984 on 35 degrees of freedom
Multiple R-squared:  0.8265,   Adjusted R-squared:  0.8116
F-statistic: 55.56 on 3 and 35 DF,  p-value: 2.155e-13



>

> anova(m10,m8)
Analysis of Variance Table

Model 1: as.numeric(viscosity)^0.5 ~ protein + ash + I(ash^2)
Model 2: as.numeric(viscosity)^0.5 ~ protein + ash + I(protein^2) + I(ash^2)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     35 28.250
2     34 26.697  1    1.5527 1.9774 0.1687


>
```

We again use anova between the model with protein^2 and the one without it, and we see it is again non-significant.

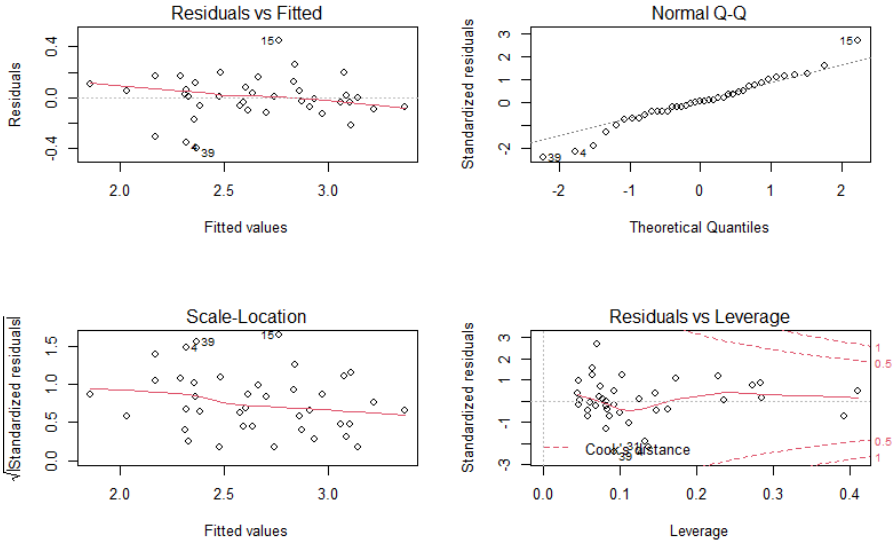
Studentized Residuals after the box cox transformations

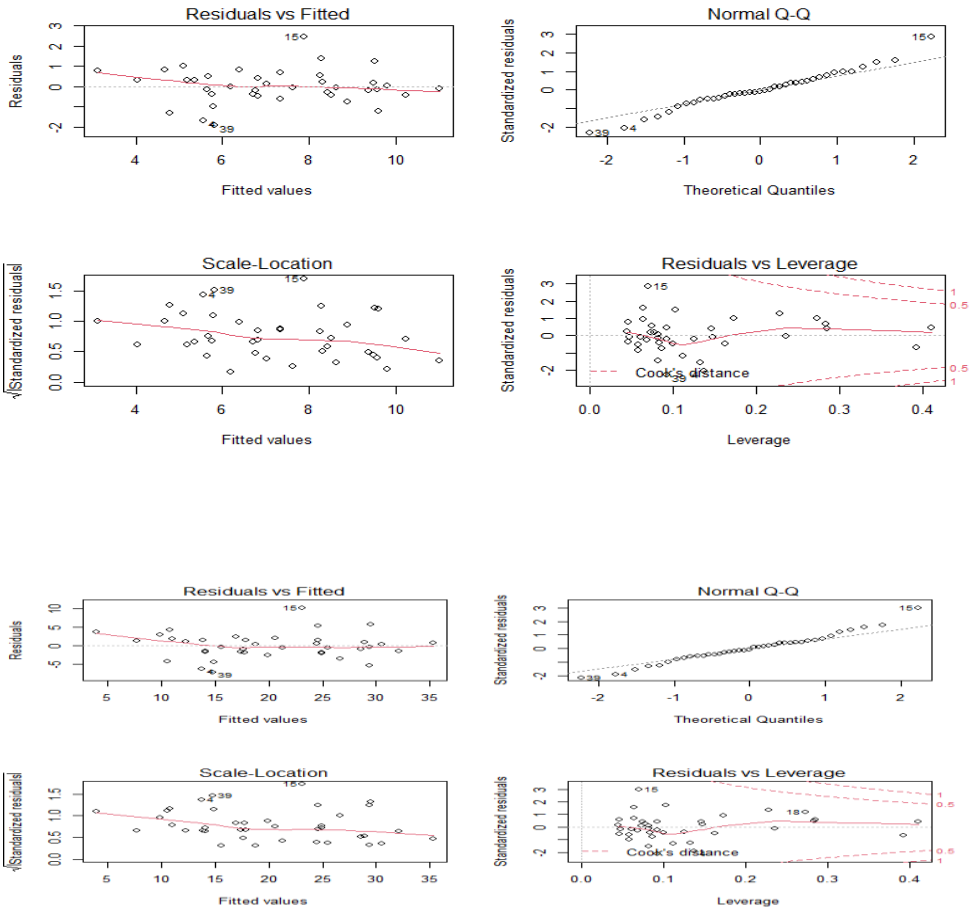




The studentized residuals does not show any outliers either.

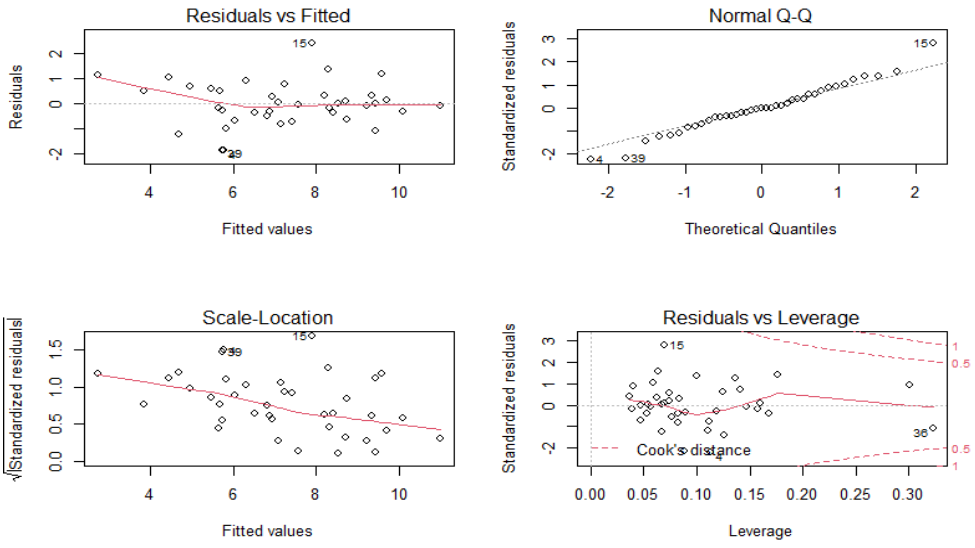
Plots after box-cox transformations:




Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

Plots of m10.


Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

None of the model shows any outliers. The quantile-quantile plot is almost linear.

```
> outlierTest(m10)
No Studentized residuals with Bonferroni p < 0.05
Largest |rstudent|:
   rstudent unadjusted p-value Bonferroni p
15 3.160946          0.0032984      0.12864
```

Moreover, I detected no outliers using the Bonferroni approach as well.

Conclusion:

So the best 2 models so far are m6 (before box-cox transformation) and m10 (after box-cox transfromation). The R-square value is similar in both cases, so I did model validation technique for both.

Model Validation: Model10, m10

```
> spr
[1] 4151.542 2598.319 2792.369
> msetr
[1] 0.8071438 0.8071438 0.8071438
> mean(spr)
[1] 3180.743
> mean(msetr)
[1] 0.8071438


>
```

Model Validation: Model6, m6

```
> spr
[1] 132.6175 134.3920 176.6529
> msetr
[1] 169.6356 169.6356 169.6356
> mean(spr)
[1] 147.8874
> mean(msetr)
[1] 169.6356


>
```

We see that the mean(spr) and mean(msetr) are much closer for m6 than compare to m10.

Thus, our final model chosen is m6:

Viscosity= -135.79 + 46.96protein – 353.81ash – 1.26protein^2 – 136.62ash^2

Reference

V.T. Huang, J.B. Lindamood, P.M.T. Hansen (1988). "Ice-Cream Cone Baking: Dependence of Baking Performance on Flour and Batter Viscosity," Food Hydrocolloids, Vol. 2, #6, pp. 451-466.