PSTAT 126 Regression Analysis

Analysis of Forest Fires:

Based on Meteorological Data and Fire Indices

Professor Sudeep Bapat

Nicole Grazda (Xu W 3:00PM) & Danny Baerman (Xu T 6:00PM)

Introduction

Our dataset is titled "Forest Fires." The data measures how weather and location affect the total damage caused by a forest fire. The data was collected from fires in Montesinho Park in northeast Portugal. There are 12 predictors, which are explicitly stated below with how they are measured in the experiment. The response measures the total area burned in the forest. Our research question is as follows: Is the area burned by a forest fire affected by the following predictors?

Predictors:
- X: states the horizontal spatial coordinate within Montesinho Park
    - [1, 9]
- Y: states the vertical spatial coordinate within Montesinho Park
    - [2, 9]
- month: states the month in which the data was collected
    - [jan, dec]
- day: states the day of the week in which the data was collected
    - [mon, sun]
- FFMC (Fine Fuel Moisture Code): rates amount of moisture of the litter and fuels
    - [63.5, 96.2]
- DMC (Duff Moisture Code): rates amount of moisture in surrounding organic layers
    - [3.20, 291.3]
- DC (Drought Code): rates amount of moisture in deeper, surrounding organic layers
    - [15.3, 860.6]
- ISI (Initial Spread Index): combines wind and FFMC data to rate expected spread of fire
    - [0.80, 22.7]
- temp: temperature in Celsius
    - [2.20, 33.30]
- RH: relative humidity percentage
    - [15.0, 96.0]
- wind: speed of wind, measured in km/hour
    - [0.40, 9.40]
- rain: quantity of rain, measured in mm/m^2
    - [0, 6.40]

Response:
- area: area of forest that has been burned, in hectares (1 hectare = 10,000m^2)
    - [0.09, 1090.84]

Questions of Interest

We are interested in determining the following:

1) Are higher wind speeds significantly related to the amount of burned forest fire area?
2) Is the Initial Spread Index significantly related to the amount of burned area?
3) Is a model containing at least one predictor from our data set useful in determining the burned forest area?

Regression Method

To answer question 1, we plot wind speed on burned area and determine the necessary transformation for the model to meet the LINE criteria. We run a hypothesis test to answer our question and make a decision based on the p-value.

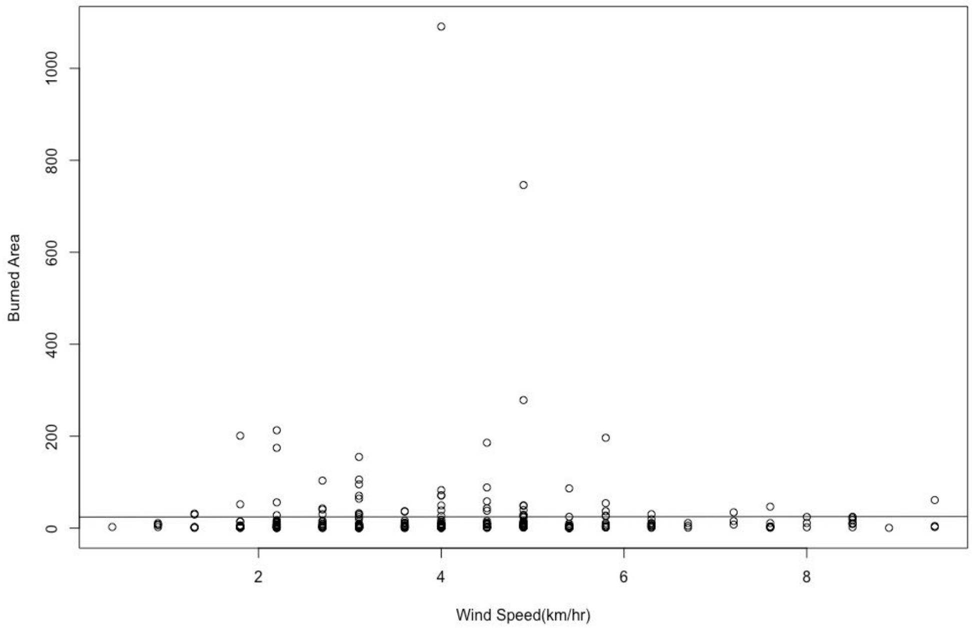
Similarly for question 2, we plot the ISI index on burned area and perform a Box-Cox transformation to improve the Residual vs. Fit and Normal Q-Q Plot. We find the p-value of this test to determine the relationship between the initial spread index and burned area.

In our final question 3, we start by transforming the response to meet the criteria for linearity, independence, normality, and equal variances. Next, we narrow down the model to a few predictors and perform a hypothesis test to obtain a p-value to answer this question. Lastly, we determine the presence of any outliers, high leverage, and/or influential points.

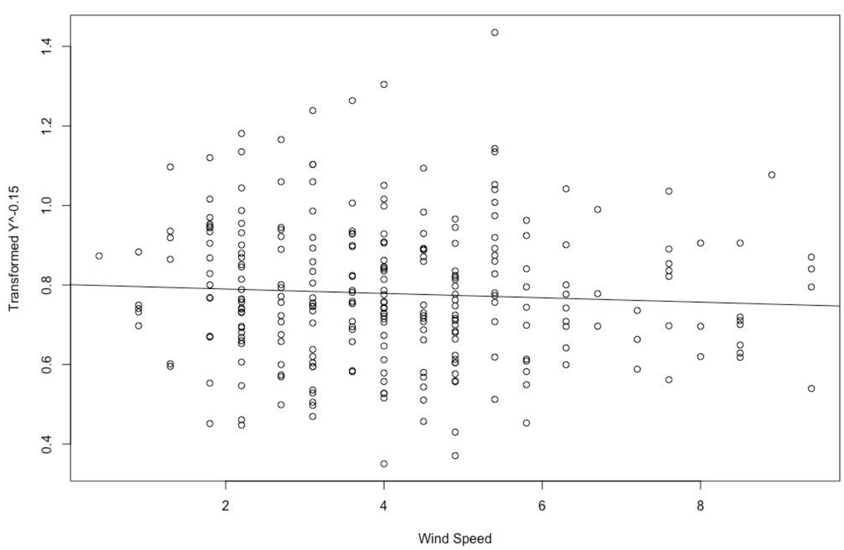
Regression Analysis

1) **Regression Analysis, Wind Speed vs. Burned Forest Area**

The scatterplot below displays burned area on wind speed. A line was also added to display the relationship between this predictor and the response.



Regressing Wind Speed on Burned Area would result in a model that fails to meet the linearity, equal error variances, and normality criteria as shown in the Residuals vs. Fitted Values and Normal Q-Q Plots. We employ a Box-Cox power transformation of Y raised to the negative 0.15 to resolve these issues.



The plot above demonstrates there is a linear relationship between Wind Speed and the transformed Burned Area variable. The LINE conditions are met with this new model, as shown by the updated Residuals vs. Fit and Normal Q-Q Plot.

Research Question: Are higher wind speeds significantly related amount of burned forest fire area?

No. After conducting the following test:

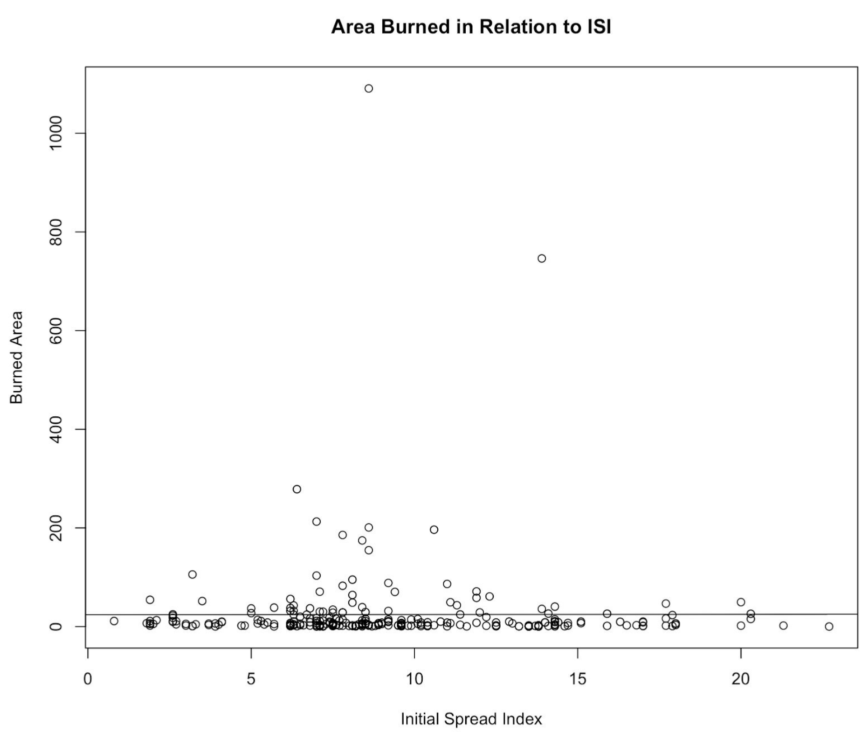
$$Ho: \beta_{wind\ speed} = 0 \ vs.\ Ha: \beta_{wind\ speed} \neq 0$$

We conclude that wind speeds have no significant effect on burned area. The t-test gives us a t-value of -0.987 corresponding to a p-value of 0.3245, which is not statistically significant at an ɑ level of 0.05. Thus, we accept the Null Hypothesis.

2) Regression Analysis, ISI vs. Burned Forest Area

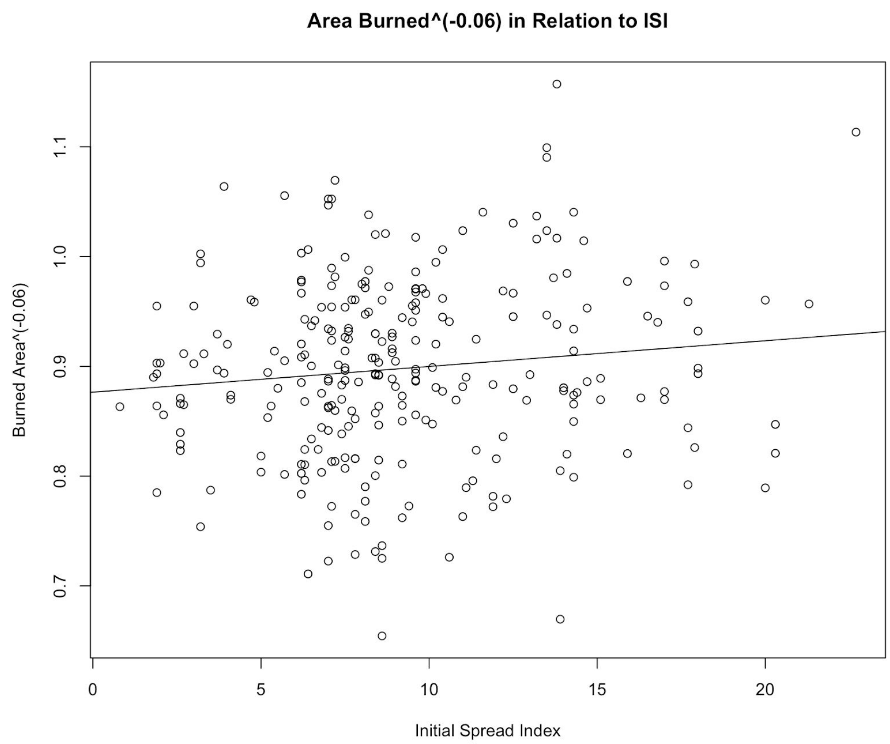
We attempt to regress our response, the total area burned, on our predictor of ISI (Initial Spread Index).

Employing no transformations, our scatterplot is as follows:



The p-value returned for our predictor, ISI, is 0.9723. Clearly, this is greater than our value of $\alpha = 0.05$, suggesting that this predictor does not influence our response variable.

We employ a Box-Cox transformation, in an attempt to create a more linear relationship between Area Burned and ISI. We raise Y to the power of $\lambda = -0.06$ and create the same model.



Our Y values appear to be more spread out, allowing us to visualize the data better and identify trends. Our updated Residuals vs Fit plot and Normal QQ plot support the idea of linear regression. Our updated Residuals vs Fit plot supports the idea of linearity and equal error variances. Our normal QQ probability plot supports the idea of the errors being normally distributed.

Research Question: Is the Initial Spread Index significantly related to the amount of burned area?

Close, but no. After conducting the following test:

$$Ho: \beta_{ISI} = 0 \text{ vs. } Ha: \beta_{ISI} \neq 0$$

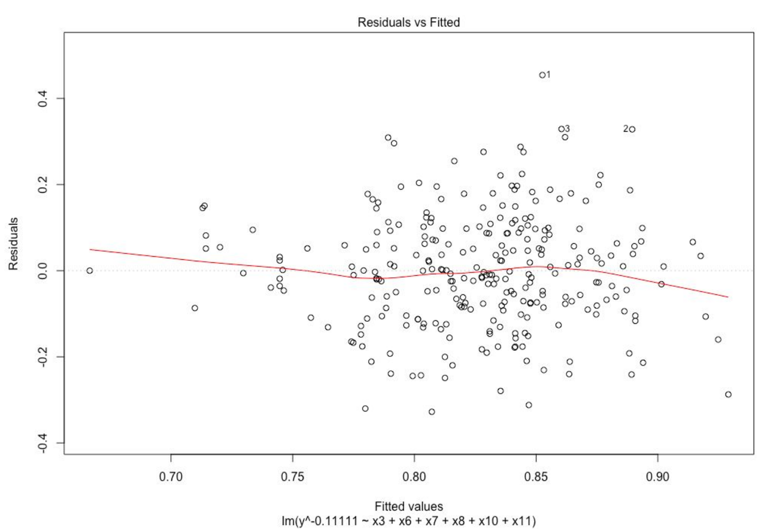
We conclude that an ISI value is not a valuable predictor for the area burned. The t-test returns a value of 1.951, with a p-value of 0.0521. Although it's close, it's greater than our α level of 0.05.
However, if we were to raise our α level to 0.1, then ISI would be a significant predictor of area burned!

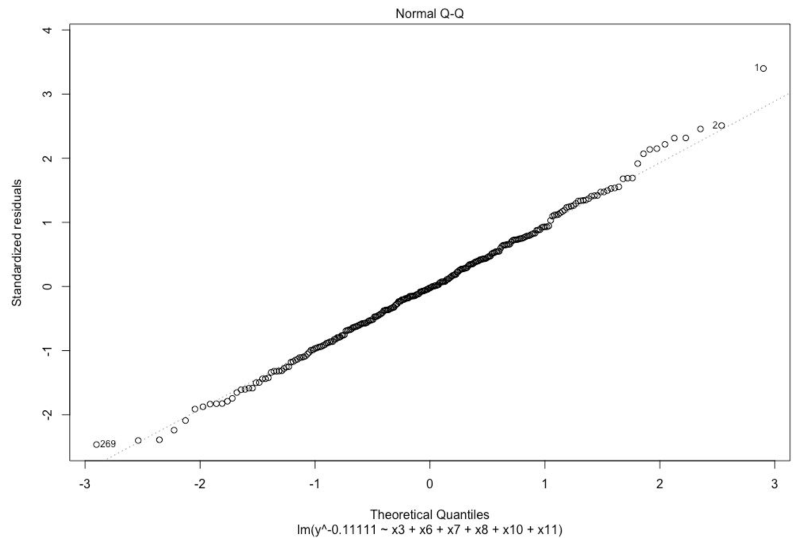
3) Regression Analysis, Building a Model for Burned Forest Area

We regressed our unaltered predictor, Y, on all 12 predictors, x1… x12. We named this model *ufit*. This model returns a p-value of 0.5897, which is not statistically significant. Some kind of transformation is necessary, especially since this model fails to meet LINE criteria.

We conducted a Box-Cox transformation on *ufit*. We obtained a lambda value of $\lambda = -0.11111$. We raise our predictor, Y, to the power of $\lambda = -0.11111$. After performing diagnostic checks, we conclude that this model meets the LINE criteria. We then regress our newly transformed Y on the 12 predictors, creating a model called *bfit*. This model returns a p-value of 0.1217. This is slightly more statistically significant, as it's less than the p-value returned by *ufit*. It's still greater than our alpha value of $\alpha = 0.05$, however.

We attempt to reduce our newly transformed model, by means of best subsets regression. We obtained values of adjusted R^2 and MSE, which allowed us to identify a reliable reduced model. Our reduced model, *ffit*, contains 6 predictors: month, DMC, DC, ISI, RH and wind. After performing diagnostic checks, we conclude that this model meets the LINE criteria. The Residuals vs. Fit and Normal Q-Q Plots below display that these requirements are met.





Research Question: Is a model containing at least one predictor useful in determining the burned forest area?

Yes. We conduct the following hypothesis test to answer our third question of interest:

Ho: $\beta_{\text{month}} = \beta_{\text{DMC}} = \beta_{\text{DC}} = \beta_{\text{ISI}} = \beta_{\text{RH}} = \beta_{\text{wind}} = 0$ *vs. Ha: At least one* $\beta k \neq 0$

This model returns a p-value of 0.03143. This is statistically significant as it is lower than our alpha value of $\alpha = 0.05$! Thus, we make a decision to reject the Null Hypothesis and conclude that this model containing at least one predictor is useful in determining the burned forest area. It appears that the model with the month, DMC index, DC index, ISI index, RH index, and wind speed is statistically significant.

Furthermore, analyzing this model in more depth we find that this model contains one outlier by the studentized deleted residual criterion. In addition, by calculating the leverages of this reduced model we find there are 14 points with high leverages. The DFFITS test claims there are 7 influential points.

Conclusion

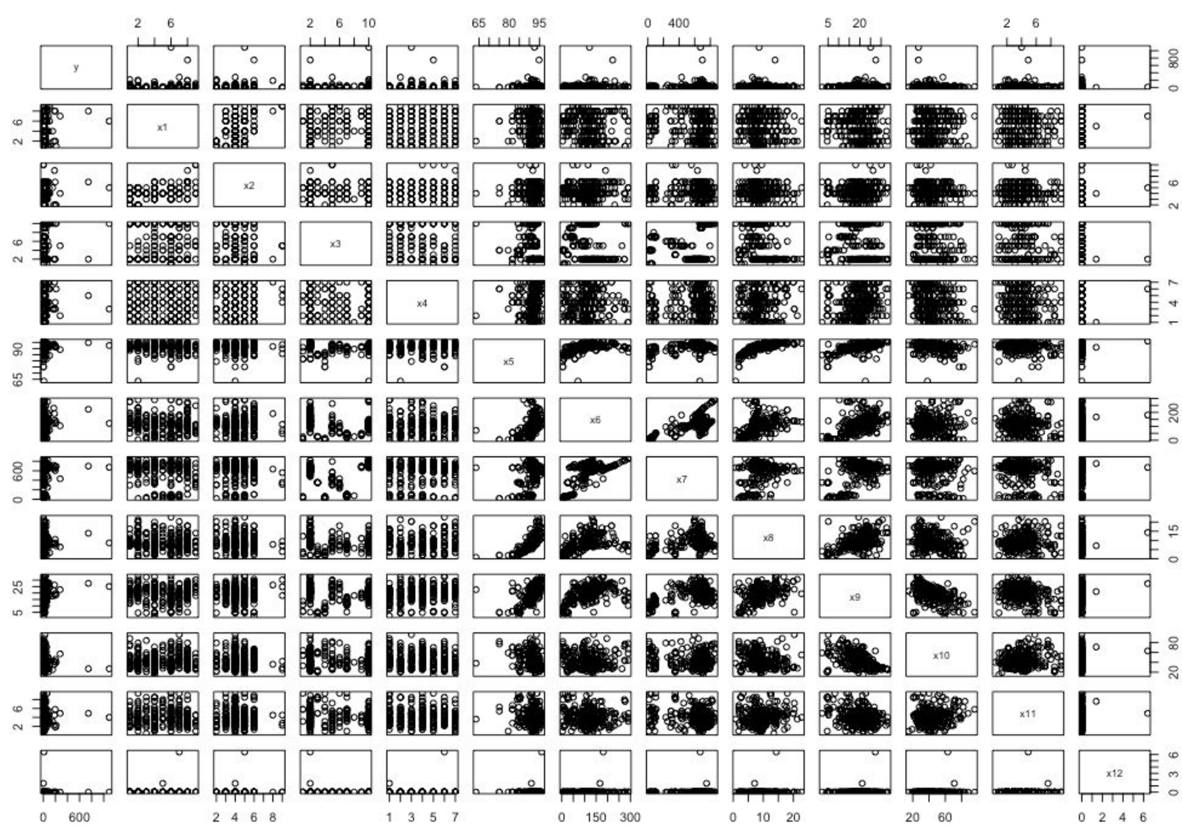
Our results suggest that there exists a relationship between various factors and the amount of damage a wildfire can cause. Our final model combines the time of year, the Duff Moisture Code, the Drought Code, the Initial Spread Index, the relative humidity, and the wind speed into six useful predictors that allow us to estimate the total burned area a forest fire will create. The Duff Moisture Code, which measures the amount of moisture found in the loosely compacted organic layers near the ground, appears to have an inverse relationship with the total area burned. In other words, more moisture allows us to confidently predict that less land will be burned. The Drought Code, which measures the amount of moisture found in the more tightly compacted organic layers deep underground, appears to have a positive correlation with the total area burned. In other words, a higher DC rating allows us to confidently predict that more land will be burned.

Our final model could be improved, however. For instance, the data is taken exclusively from Montesinho Park, in northeastern Portugal. The geography conditions certainly may have influenced the total amount of damage caused by these fires, in ways that may not be pertinent in other countries. Our model is limited by its predictors, as well. Our predictors are indices from the Canadian Forest Fire Weather Index (FWI) System. However, there exist many different ways to measure fires, such as the National Fire Danger Rating System, used in the U.S.

Appendix

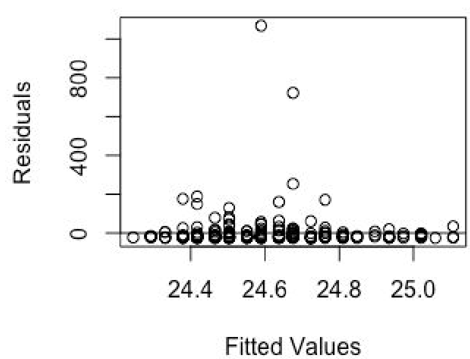
```
# Creation of Variables
x1 <- forestfires1$X
x2 <- forestfires1$Y
x3 <- factor(forestfires1$month)
x4 <- factor(forestfires1$day)
x5 <- forestfires1$FFMC
x6 <- forestfires1$DMC
x7 <- forestfires1$DC
x8 <- forestfires1$ISI
x9 <- forestfires1$temp
x10 <- forestfires1$RH
x11 <- forestfires1$wind
x12 <- forestfires1$rain
y <- forestfires1$area
```

```
# Scatterplot Matrix of Burned Area vs. All Predictors
n <- data.frame(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,y)
pairs(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12)
cor(n)
```

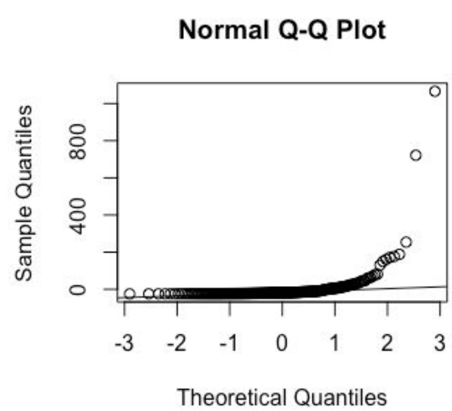


Code for Research Question #1:

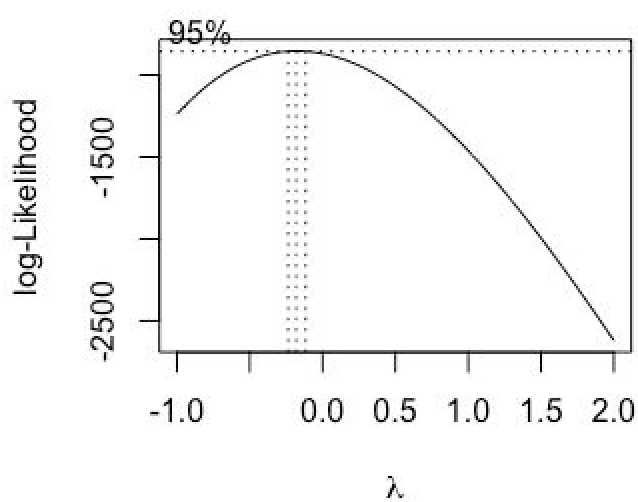
```
> # Relationship between Wind and Burned Area
> plot(x11,y, xlab = "Wind Speed(km/hr)",ylab = "Burned Area")
> fit <- lm(y~x11)
> fit

Call:
lm(formula = y ~ x11)

Coefficients:
(Intercept)          x11
   24.20641      0.09574

> abline(24.20641,0.09574)

> # Residuals vs. Fitted Values
> yhat <- 24.20641+.09574*x11
> e <- y-yhat
> plot(yhat,e,ylab="Residuals",xlab="Fitted Values")
> abline(0,0)
```



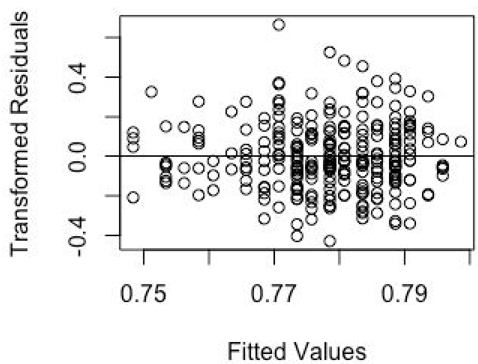

```
#Normal Q-Q Plot
qqnorm(e)
qqline(e)
```

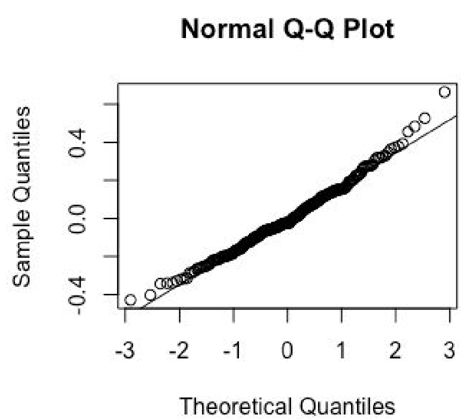


```
> # Boxcox Transformation, with Code for Revised Plot
> boxcox(fit,lambda= seq(-1,2))
> plot(x11,y^-.15,xlab= "Wind Speed",ylab = "Transformed Y^-0.15")
> fitnew <- lm(y^-.15~x11)
> fitnew

Call:
lm(formula = y^-0.15 ~ x11)

Coefficients:
(Intercept)          x11
   0.800867    -0.005585

> abline(0.800867,-0.005585)
> summary(fitnew)

Call:
lm(formula = y^-0.15 ~ x11)

Residuals:
     Min       1Q   Median       3Q      Max
-0.42831 -0.11827 -0.02367  0.11537  0.66433

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.800867   0.025589  31.297   <2e-16 ***
x11         -0.005585   0.005658  -0.987    0.324
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1749 on 268 degrees of freedom
Multiple R-squared:  0.003622,  Adjusted R-squared:  -9.54e-05
F-statistic: 0.9743 on 1 and 268 DF,  p-value: 0.3245
```



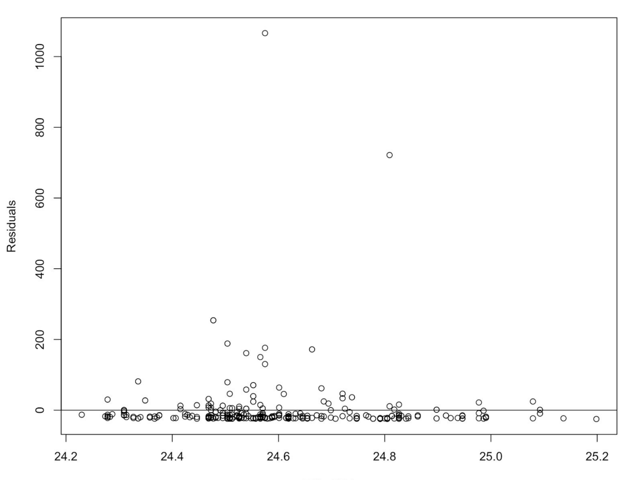
```
> # Transformed Residuals vs. Fitted Values
> # and Normal Q-Q Plot
> yhat<- .800867-0.005585*x11
> e <- y^-.15-yhat
> plot(yhat,e,ylab="Residuals",xlab="Fitted Values")
> abline(0,0)
> qqnorm(e)
> qqline(e)
```

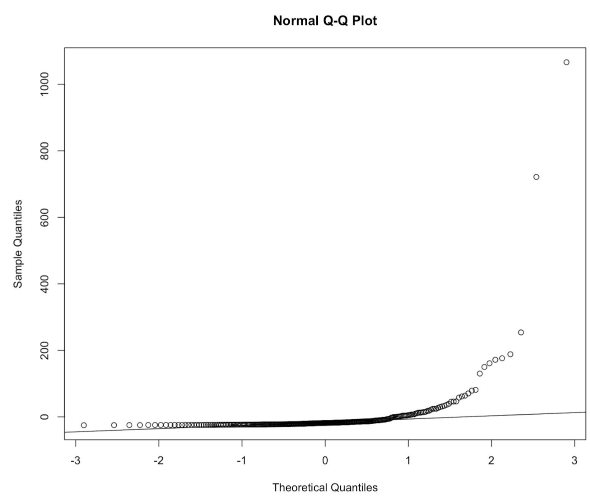




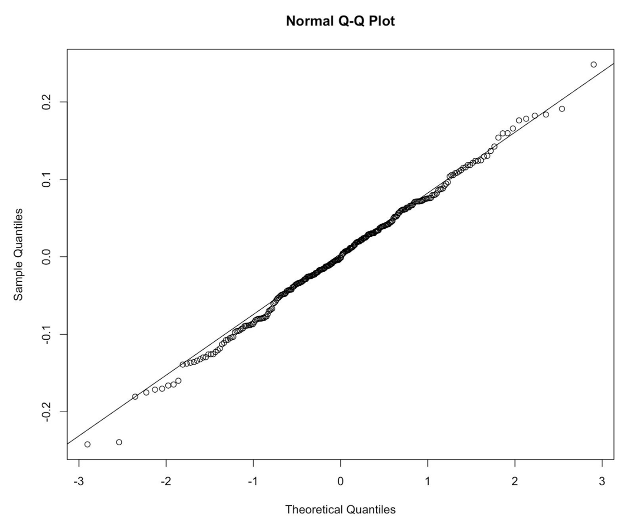
Code for Research Question #2:

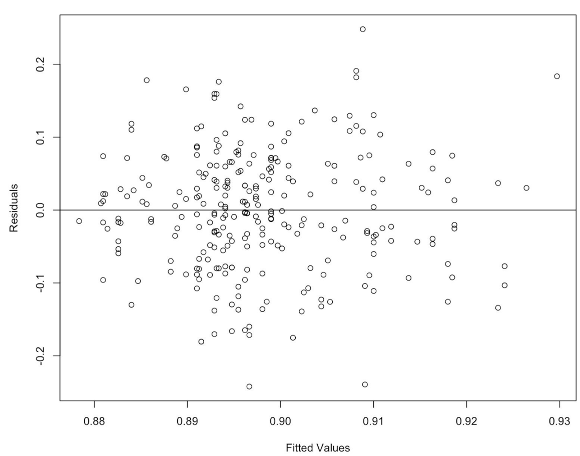
```
> fit8<-lm(y~x8)
> summary(fit8)

Call:
lm(formula = y ~ x8)

Residuals:
    Min      1Q  Median      3Q     Max 
 -25.03  -22.41  -18.24   -9.39 1066.27 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 24.19420   12.82648   1.886   0.0603 .
x8           0.04424    1.27405   0.035   0.9723  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 86.66 on 268 degrees of freedom
Multiple R-squared:  4.499e-06, Adjusted R-squared:  -0.003727 
F-statistic: 0.001206 on 1 and 268 DF,  p-value: 0.9723
> coef(fit8)
(Intercept)          x8 
24.19420015  0.04423923 
> plot(x8,y,xlab = "Initial Spread Index",ylab = "Burned Area",main = "Area Burned in Relation to ISI")
> abline(24.19420015,0.04423923)
> yhat<-24.19420015+0.04423923*x8
> e=y-yhat
> plot(yhat,e,ylab = "Residuals",xlab = "Fitted Values")
> abline(0,0)
> qqnorm(e)
> qqline(e)
```

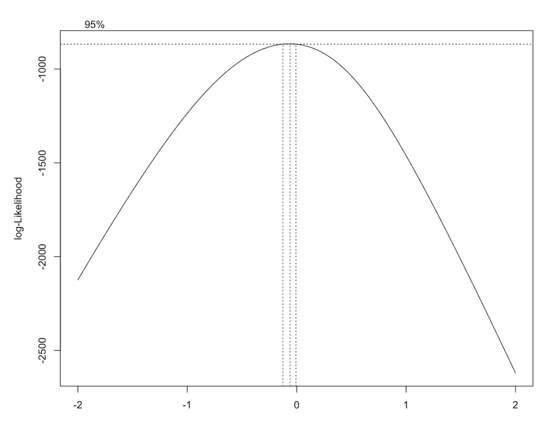




```
> bc8<-boxcox(fit8)
> lam8<-bc8$x[which(bc8$y==max(bc8$y))]
> newy8<-y^lam8
> plot(x8,newy8,xlab = "Initial Spread Index",ylab = "Burned Area^(-0.06)",main = "Area Burned^(-0.06) in Relation
to ISI")
> newfit8<-lm(newy8~x8)
> summary(newfit8)

Call:
lm(formula = newy8 ~ x8)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.242175 -0.048877 -0.001426  0.056908  0.248278 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.876484   0.012098  72.448   <2e-16 ***
x8          0.002345   0.001202   1.951   0.0521 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08174 on 268 degrees of freedom
Multiple R-squared:  0.01401,	Adjusted R-squared:  0.01033 
F-statistic: 3.808 on 1 and 268 DF,  p-value: 0.05206

> coef(newfit8)
(Intercept)          x8 
0.876484292 0.002344889 
> abline(0.876484292,0.002344889)
> y8hat<-0.876484292+0.002344889*x8
> e8<-newy8-y8hat
> plot(y8hat,e8,ylab = "Residuals",xlab = "Fitted Values")
> abline(0,0)
> qqnorm(e8)
> qqline(e8)
```

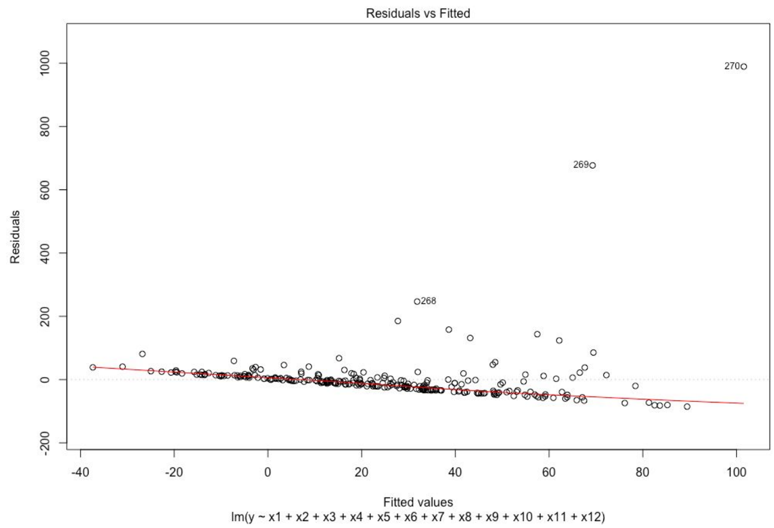


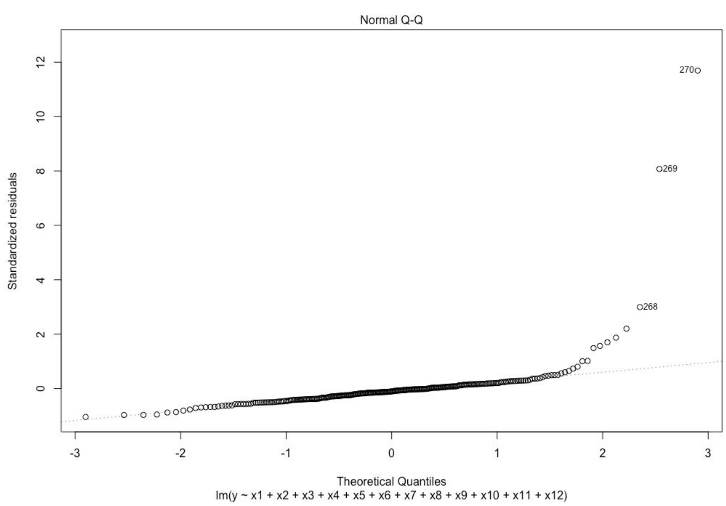




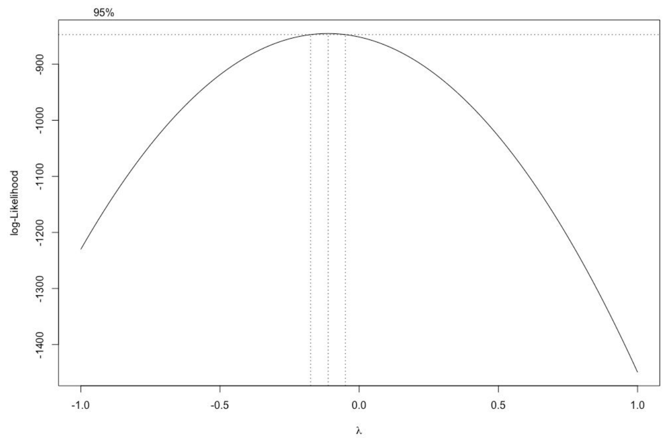
Code for Research Question #3:

```
> # Finding Full Model p-value
> ufit <- lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12)
> summary(ufit)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 +
    x10 + x11 + x12)

Residuals:
   Min     1Q Median     3Q    Max
-85.51 -29.17  -8.10  10.63 989.28

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -72.8651   229.6151  -0.317   0.7513
x1            4.4445     2.7409   1.622   0.1062
x2           -1.3338     5.5634  -0.240   0.8107
x3aug       113.8268    79.1494   1.438   0.1517
x3dec        95.0408    67.4670   1.409   0.1602
x3feb         7.7780    52.9044   0.147   0.8832
x3jul        64.2299    67.6623   0.949   0.3434
x3jun        19.4600    63.6391   0.306   0.7600
x3mar       -12.7067    49.4486  -0.257   0.7974
x3may        23.8829   100.6820   0.237   0.8127
x3oct       162.4368    96.5643   1.682   0.0938 .
x3sep       170.3176    88.8641   1.917   0.0565 .
x4mon         7.0362    20.8770   0.337   0.7364
x4sat        36.8533    19.7802   1.863   0.0636 .
x4sun         6.2859    19.5138   0.322   0.7476
x4thu        19.4611    21.7411   0.895   0.3716
x4tue         7.9825    20.3898   0.391   0.6958
x4wed         6.6694    21.4715   0.311   0.7564
x5            0.4093     2.5127   0.163   0.8707
x6            0.3766     0.1642   2.293   0.0227 *
x7           -0.3008     0.1203  -2.501   0.0131 *
x8           -1.4656     2.1599  -0.679   0.4981
x9           2.7029     2.0204   1.338   0.1822
x10         -0.1591     0.5802  -0.274   0.7841
x11          3.9190     3.3890   1.156   0.2487
x12         -4.3321    14.0259  -0.309   0.7577
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 86.86 on 244 degrees of freedom
Multiple R-squared:  0.0854,	Adjusted R-squared:  -0.008304
F-statistic: 0.9114 on 25 and 244 DF,  p-value: 0.5897

> # Attempt to meet LINE conditions
> plot(ufit, which = c(1,2))
```



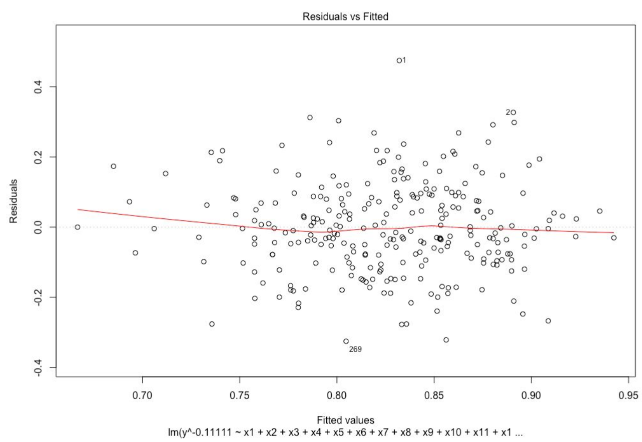


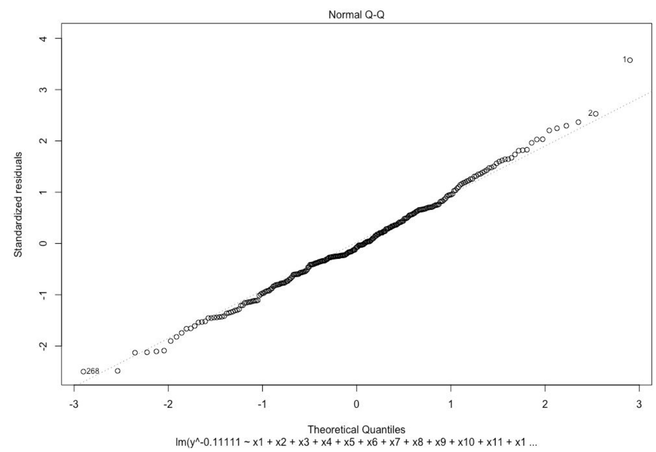

```
> # Finding boxcox transformation
> bc<-boxcox(ufit, lambda=seq(-1,1))
> best.lam<-bc$x[which(bc$y==max(bc$y))]
> best.lam
[1] -0.1111111
```




```
> # Finding Full Model p-value with transformation
> bfit<-lm(y^-.11111~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11+x12)
> summary(bfit)

Call:
lm(formula = y^-0.11111 ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 +
    x8 + x9 + x10 + x11 + x12)

Residuals:
     Min       1Q   Median       3Q      Max
-0.32520 -0.07917 -0.00924  0.08334  0.47468

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.810e-01  3.584e-01   2.458  0.01466 *
x1          -3.467e-03  4.278e-03  -0.810  0.41847
x2           7.674e-03  8.683e-03   0.884  0.37771
x3aug       -3.168e-02  1.235e-01  -0.256  0.79785
x3dec       -1.557e-01  1.053e-01  -1.479  0.14045
x3feb        2.264e-02  8.257e-02   0.274  0.78419
x3jul        1.991e-02  1.056e-01   0.188  0.85065
x3jun        5.400e-02  9.933e-02   0.544  0.58717
x3mar        3.711e-02  7.718e-02   0.481  0.63103
x3may       -1.088e-01  1.571e-01  -0.692  0.48942
x3oct       -2.599e-01  1.507e-01  -1.725  0.08586 .
x3sep       -1.418e-01  1.387e-01  -1.022  0.30764
x4mon        1.234e-03  3.259e-02   0.038  0.96982
x4sat       -5.167e-02  3.087e-02  -1.674  0.09549 .
x4sun       -3.464e-02  3.046e-02  -1.137  0.25657
x4thu       -2.261e-02  3.393e-02  -0.666  0.50580
x4tue       -2.645e-02  3.182e-02  -0.831  0.40668
x4wed        1.844e-03  3.351e-02   0.055  0.95616
x5          -7.060e-04  3.922e-03  -0.180  0.85728
x6          -8.081e-04  2.563e-04  -3.153  0.00182 **
x7           3.987e-04  1.877e-04   2.124  0.03471 *
x8           2.734e-03  3.371e-03   0.811  0.41814
```

```
x9           -3.228e-03  3.153e-03  -1.024  0.30702
x10          -2.161e-05  9.055e-04  -0.024  0.98098
x11          -4.250e-03  5.290e-03  -0.803  0.42253
x12          -6.512e-03  2.189e-02  -0.297  0.76637
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1356 on 244 degrees of freedom
Multiple R-squared:  0.1226,    Adjusted R-squared:  0.03269
F-statistic: 1.364 on 25 and 244 DF,  p-value: 0.1217
```






```
> # Finding the Reduced Model using various criteria
> mod <- regsubsets(cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12),y^-.11111)
> summary.mod <- summary(mod)
> summary.mod$which
  (Intercept)    x1    x2    x3    x4    x5    x6    x7    x8    x9   x10   x11
1        TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
2        TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
3        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
4        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE
5        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE
6        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
7        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
8        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
    x12
1 FALSE
2 FALSE
3 FALSE
4 FALSE
5 FALSE
6 FALSE
7  TRUE
8  TRUE
> summary.mod$rsq
[1] 0.01558636 0.02276688 0.03426588 0.04166039 0.04850822 0.05388437 0.05489087
[8] 0.05547298
> summary.mod$adjr2
[1] 0.01191317 0.01544678 0.02337414 0.02719488 0.03048754 0.03229998 0.02963987
[8] 0.02652196
> summary.mod$rss
[1] 5.031599 4.994897 4.936123 4.898328 4.863327 4.835848 4.830703 4.827728
```

```
> mod<-regsubsets(cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12),y^-.11111)
> summary.mod<-summary(mod)
> n=270
> rss<-summary.mod$rss
> mses<-c(rss[1]/(n-2),rss[2]/(n-3),rss[3]/(n-4),rss[4]/(n-5),rss[5]/(n-6),rss[6]/(n-7),rss[7]/
+         (n-8),rss[8]/(n-9),rss[9]/(n-10),rss[10]/(n-11),rss[11]/(n-12),rss[12]/(n-13))
> mses
 [1] 0.01877462 0.01870748 0.01855685 0.01848426 0.01842169 0.01838725 0.01843780 0.01849704
> summary.mod$which
  (Intercept)    x1    x2    x3    x4    x5    x6    x7    x8    x9   x10   x11   x12
1        TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
2        TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
3        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
4        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE
5        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE
6        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE
7        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
8        TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
```

```
> # We choose the adjusted R^2 model
> ffit <- lm(y^-.11111~x3+x6+x7+x8+x10+x11)
> summary(ffit)

Call:
lm(formula = y^-0.11111 ~ x3 + x6 + x7 + x8 + x10 + x11)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.32765 -0.08525 -0.00245  0.08689  0.45418 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.7584728  0.0786178   9.648  < 2e-16 ***
x3aug       -0.0507184  0.1183013  -0.429 0.668487    
x3dec       -0.1063938  0.0979805  -1.086 0.278563    
x3feb        0.0201773  0.0804293   0.251 0.802117    
x3jul       -0.0113114  0.0992439  -0.114 0.909346    
x3jun        0.0255494  0.0937115   0.273 0.785352    
x3mar        0.0299064  0.0751107   0.398 0.690842    
x3may       -0.1171833  0.1512896  -0.775 0.439316    
x3oct       -0.2538513  0.1458088  -1.741 0.082892 .  
x3sep       -0.1465011  0.1339665  -1.094 0.275178    
x6          -0.0008493  0.0002417  -3.514 0.000521 ***
x7           0.0003736  0.0001821   2.051 0.041251 *  
x8           0.0022957  0.0025642   0.895 0.371474    
x10          0.0005589  0.0005845   0.956 0.339894    
x11         -0.0040235  0.0050750  -0.793 0.428623    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1349 on 255 degrees of freedom
Multiple R-squared:  0.09252,	Adjusted R-squared:  0.0427 
F-statistic: 1.857 on 14 and 255 DF,  p-value: 0.03143

> # Diagnostic check
> plot(ffit,which = c(1,2))
```

```r
# Check for Outliers
r <- rstudent(ffit)
m <- c()
for (i in r){
  if (abs(i) >= 3){
    m <- c(m,i)
  } else {
    m <- c(m)
  }
  m
}
m
```

```
> m
[1] 3.471583
```

```
> # Check leverages
> h <- hatvalues(ffit)
> p <- c()
> for (i in h){
+   if (i > 3*(13/270)){
+     p <- c(p,i)
+   } else {
+     p <- c(p)
+   }
+   p
+ }
> p
 [1] 0.1693098 0.1665829 0.1720016 0.2664046 0.2526552 0.2097565 0.2030739 0.2068989
 [9] 0.2526552 0.2024344 0.2028824 1.0000000 0.2066723 0.2811382
```

```r
# Check for influential points
d <- dffits(ffit)
t <- c()
for (i in d){
  if (i > 2*sqrt((13)/(270-13-1))){
    t <- c(t,i)
  } else {
    t <- c(t)
  }
  t
}
t
```

```
> t
[1] 0.4732632 0.6391002 0.6600535 0.4761728 0.5166490 0.5501014 0.4672992
```