ST3131 Regression Analysis
AY 2016/2017 Semester 2
Group 32

Grace Ho A0142362N
Nicholas Foo Zi Hao A0139715Y
Tan Xin Tong A0144196B
Liu Le A0105565H

Summary

This report provides an analysis of several linear regression models that aim to predict rent paid for agricultural land planted to alfalfa in Minnesota. The dataset considered variables Rent, AllRent, Cows, Pasture, and Liming which would be further explained in the report.

In this analysis, a simple regression model with all predictors, Model 1: $\widehat{Rent} = -2.8282 + 0.8833\, AllRent + 0.4318\, Cows - 11.3804\, Pasture - 1.10117\, Liming$ is fitted. To filter out the less useful predictors and to find a more appropriate model, stepwise regression is employed and Model 2: $\widehat{Rent} = -6.11433 + 0.39255\, Cows + 0.92137\, AllRent$ is obtained. However, the test for independence revealed that this model violates the constant variance assumption hence the analysis is proceeded with transformation on the response variable. We carried out a box-cox transformation and obtained Model 3: $\widehat{\sqrt{Rent}} = 2.370239 + 0.031991\, Cows + 0.073798\, AllRent$. In order to achieve a better R^2 value, Model 4: $\widehat{\sqrt{Rent}} = 2.0534307 + 0.0724896\, Cows + 0.0735719\, AllRent - 0.0007732\, Cows^2$ is obtained from the test for quadratic and interaction effect. This analysis eventually concludes with Model 4: $\widehat{\sqrt{Rent}} = 2.0534307 + 0.0724896\, Cows + 0.0735719\, AllRent - 0.0007732\, Cows^2$ that the square-root of Average rent payable to land planted with Alfalfa (*Rent*) is directly proportional to the average rent paid to all arable land (*AllRent*) and density of cows (*Cows*). In contrast, it is inversely proportionate to the squared value of the density of cows ($Cows^2$).

1 Description of Problem

Alfalfa is one of the earliest crops domesticated by man. It has since been extensively used to feed horses, sheep, cows and many other animals. It is claimed to be one of the most important hays in world food production, and without it, many farms and ranches would fail to function.[1]

The data also reflects if liming is required to grow Alfalfa in the different counties. Liming is the application of calcium and magnesium rich substances to soil to neutralise soil acidity and increase activity of soil bacteria. Correcting acid soil conditions by adding lime has a significant impact on the yields of crops, especially alfalfa.[2]

Since Alfalfa is considered the premier forage for dairy cows, it would seem that land planted to alfalfa are better utilized than that for other agricultural purposes. At the same time, land that requires liming for the plantation of alfalfa would mean additional expense. This gives rise to a difference in rent for land in counties with different densities of dairy cows.

Therefore, we want to explore the variation in rent paid in 1977 for agricultural land planted to alfalfa.

[1] http://agric.ucdavis.edu/files/242006.pdf

[2] http://www.farmtalknewspaper.com/news/liming-prior-to-fall-seedings-of-alfalfa/article_26b15326-3079-11e5-a106-d3539ef633f8.html

2 Description of Data

The data was collected by Tiffany Douglas and used as an example in the book "Applied Linear Regression" by Sanford Weisberg (1985)[3]. The study was conducted in Minnesota to study the variation in rent of land planted with Alfalfa in 1977. Each county in Minnesota is taken as an observation hence there are 67 observations.

Response variable:

$Rent$	Average rent payable to land planted with Alfalfa

Predictor variables:

$AllRent$	Average rent paid to all arable land
$Cows$	Density of cows
$Pasture$	Proportion of pasteurized farmland
$Liming$	Was liming carried out? Yes/No

3 Regression Analysis

3.0 Checking for Multicollinearity

Before diving into regression analysis, a Multicollinearity check was done to check for any linear relationships among predictors.
From the outputs, all Condition Indexes of the predictors are below 10 indicating that there are no significant linear relationships amongst each other.

3.1 Work on Primary Model (Model 1)

From the data, we used the 4 predictor variables to conduct a full model regression analysis (Refer to Appendix B "**Running Overall Regression for Model 1"**). This model, named Model 1, is given by

Model 1 :

$$\widehat{Rent} = -2.8282 + 0.8833\,AllRent + 0.4318\,Cows - 11.3804\,Pasture - 1.10117\,Liming$$

Key findings of Model 1

[3] Dataset was obtained from Statsci.org *Rent for Land Planted to Alfalfa* http://www.statsci.org/data/general/landrent.html

(a) While conducting an F-test, the p-value of the model is <2.2e-16 which is a lot smaller than 0.05 (the 5% Significance Level). As such this model is indeed significant
(b) Adjusted R-squared is valued at 0.8301. As such Model1 explains 83.0% of the variability in the average rent payable to land planted with Alfalfa. There is thus a relatively strong linear relationship between the response variables and the predictors.

3.2 Stepwise Regression: Model 2

We then proceed to carry out stepwise regression to see if we can further improve the model by eliminating several predictors. In this procedure, each predictor variable is added or deleted from the model at each step. In this study, we carried out stepwise regression using Akaike Information Criterion (AIC), Mallows Cp and the partial F-test (see Appendix B on **"Stepwise Regression, Mallows Cp & partial F-test to obtain Model 2"**) to check for the contribution of a predictor variable while some other variables are already included in the model. At the end of all stepwise regressions, we ended up with a reduced model made up of only 2 predictor variables (*AllRent* and *Cows*). This reduced model, named Model2 is as follows:

$$\textbf{Model 2 :}\ \widehat{Rent} = -6.11433 + 0.39255\ Cows + 0.92137\ AllRent$$

Key findings of Model 2

(a) While conducting an F-test, the p-value of the model is <2.2e-16 which is a lot smaller than 0.05 (the 5% Significance Level). As such this model is indeed significant.
(b) Adjusted R-squared is valued at 0.8328. As such Model2 explains 83.28% of the variability in the average rent payable to land planted with Alfalfa. There is thus still a relatively strong linear relationship between the response variables and the predictors.

For more information on output, please refer to Appendix B **"Regression on Model 2"**.

3.2.1 Assumption Check of Model 2

After obtaining Model 2, we then decide to check on the assumptions of regression to see if any transformation of the model is needed. In regression analysis, we assume that error terms are independent and normally distributed with mean zero and variance. We thus plotted the residuals against the fitted values to examine if there is a pattern in the distribution of the residual to check on this assumption.

(a) Test for Independence

From Figure 1, it is observed that the residuals adhere to a funnel shape pattern. This violates the constant variance assumption hence a ***transformation on the response variable*** is necessary.

(b) Test for Normality

According to the Normal Q-Q Plot of Model 2 (Figure 2), it can be seen that almost all the values follow the Normal line. However, observations towards the end values are seen to be skewing away from the Normal line. In particular, observations 36,57 and 67 can be seen to be potential outliers, which we should be wary about. Outliers would be covered towards the end of the report.

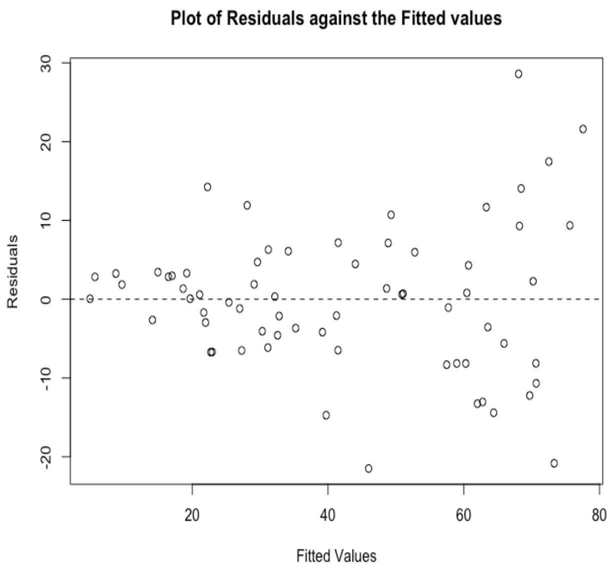


Figure 1

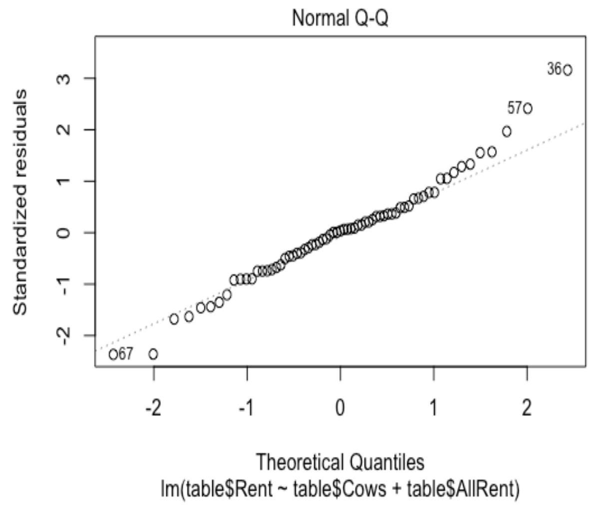


Figure 2

The predictors of Model 2 were also plotted against the residuals. With AllRent showing a more funnel shape further implying that a transformation of the response, Rent, is needed. Whereas Cows showed a curve pattern implying that the model is inadequate and a transformation of the response is needed or some other terms should be included in the model (Refer to Appendix B "**Plots of Residuals against Predictors for Model2"**).

3.3 Transformation of Model 2 to Model 3

Box Cox Transformation

In this study, the two conditions for the Box-Cox transformation which are (1) y is always positive and (2) y_{max} /y_{min}>10 are met. Therefore, the Box-Cox transformation on Y is carried out. The value $\boldsymbol{\lambda}$=0.5 is chosen using the maximum likelihood method. The response variable Y is square rooted. The new fitted model, Model3 becomes

$$\textbf{Model 3 :} \widehat{\sqrt{Rent}} = 2.370239 + 0.031991\, Cows + \ 0.073798\, AllRent$$

<u>Key findings of Model3</u>

(a) While conducting an F-test, the p-value of the model is <2.2e-16 which is a lot smaller than 0.05 (the 5% Significance Level). As such Model3 is significant.

(b) Adjusted R-squared is valued at 0.8685 (a slight improvement from Model 2). As such Model1 explains 86.85% of the variability in the average rent payable to land planted with Alfalfa. There is thus a stronger linear relationship between the response variables and the predictors as compared to Model 2.

For more information on output, please refer to Appendix B **"Regression on Model 3"**.

Checking of Independence and Normality for Model 3

(a) Test for Independence

With the transformed model, the residual plot against the fitted values (Figure 3) is observed to be more randomised hence the variance of the error terms in the model is stabilised.

(b) Test for Normality

According to the Normal Q-Q plot (Figure 4) , it can be seen that there is a slightly improved Normal relation with observations as the points are closer to the dotted line, with the largest improvements being at the end points as compared to the Normal Q-Q plot prior. However, the usual suspected outliers or influential points can still be seen.

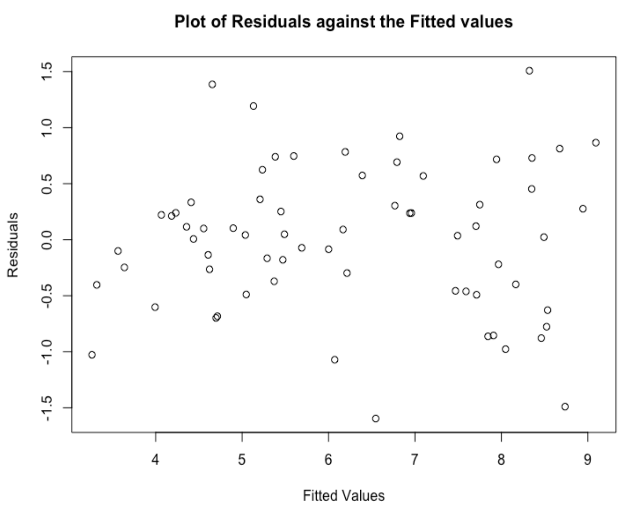


Figure 3

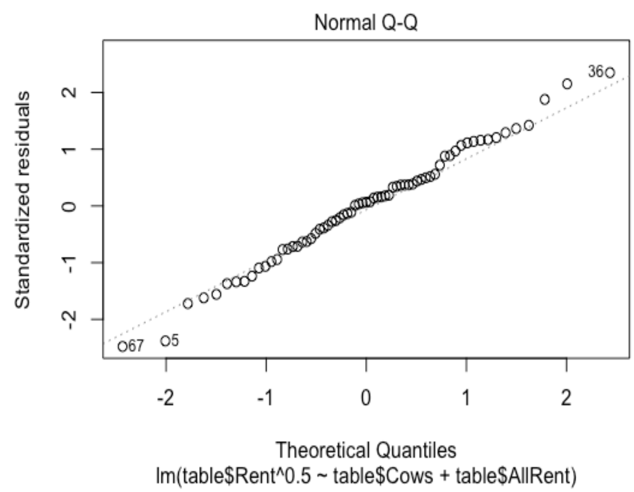


Figure 4

3.4 Testing Quadratic Terms to Obtain Model 4

We went on to check if the following quadratic terms, $AllRent^2$ and $Cows^2$, could be potential terms to be added into the model.

Similar to before, a stepwise regression based on partial F-test as well as forward and backward integration based on AIC was done. Mallows Cp was also conducted.(Refer to Appendix B Stepwise for

Quadratic Terms Model) After these procedures, $Cows^2$ was retained in the end, yielding a model, named Model4 as follows:

Model 4 : $\widehat{\sqrt{Rent}} = 2.0534307 + 0.0724896\, Cows + \; 0.0735719\, AllRent \; - \; 0.0007732\, Cows^2$

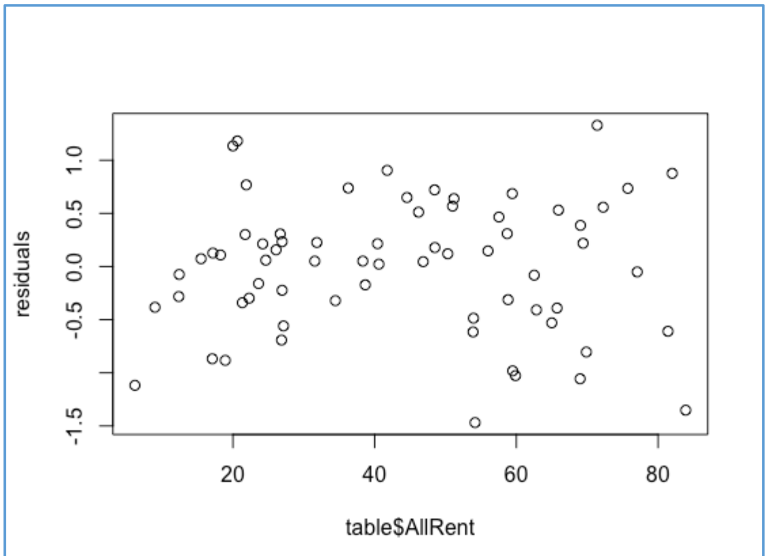
<u>Key findings of Model4</u>

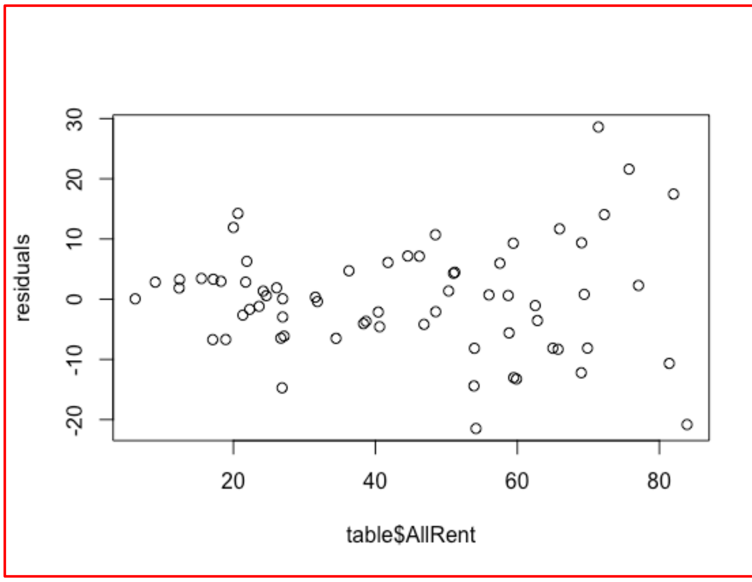
(a) While conducting an F-test, the p-value of the model is <2.2e-16 which is a lot smaller than 0.05 (the 5% Significance Level). As such Model4 is significant.
(b) Adjusted R-squared is valued at 0.8726 (a slight improvement from Model3). As such Model4 explains 87.26% of the variability in the average rent payable to land planted with Alfalfa. There is thus a stronger linear relationship between the response variables and the predictors as compared to Model3 signifying that Model4 is a better model.
(c) The residual plots seem to have improved as well, this would be covered in depth in the next section

For more information on output, please refer to Appendix B **"Regression on Model 4"**.

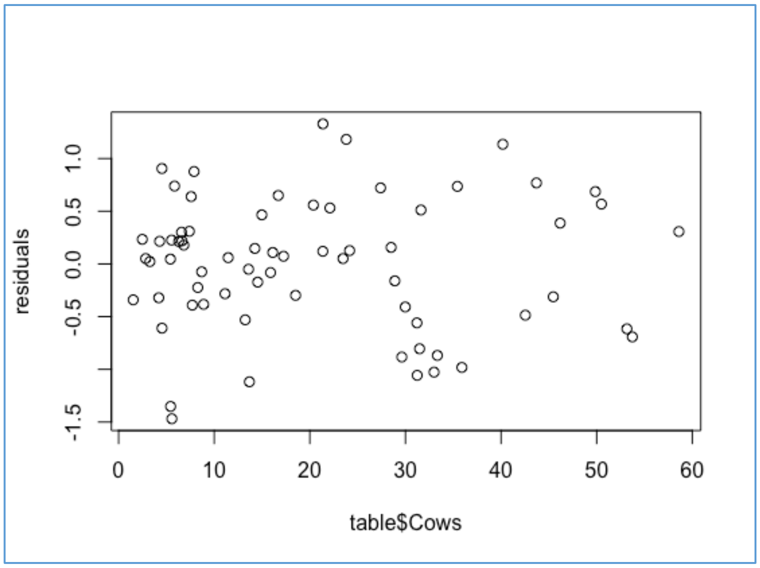
3.4.1 Residuals Vs Predictors Plots for Model 4

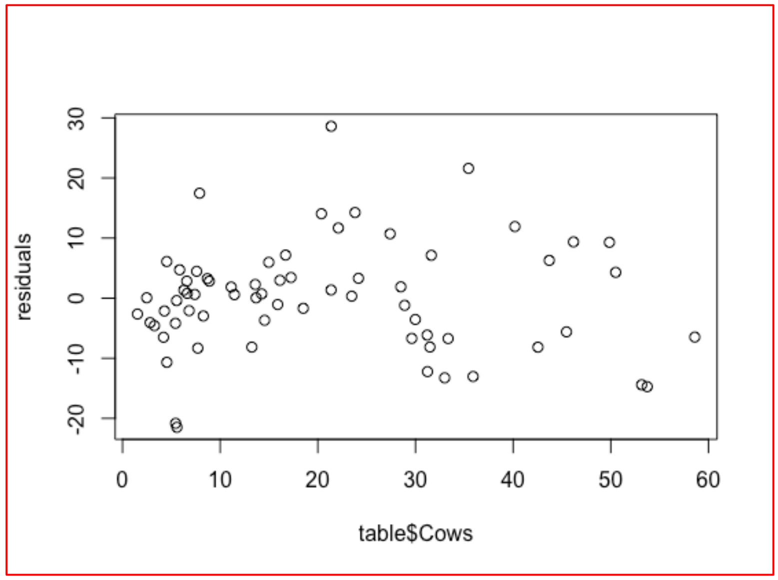
After adding in quadratic predictors, we then conducted a Residuals against Predictors plot and noticed an improvement in these patterns as mentioned below.





From the plots, it can be seen that the scatters have improved from a funnel shaped pattern before (in Red belonging to Model2) to a more random scatter (after in Blue belonging to Model4) indicating no observable or distinct pattern.





From the plot, it can be seen that scatters have also improved from a curved pattern before (in Red in Model 2) to a random scatter after transformation (in Blue in Model 4).

The plots of Cows2 against Residuals do not show any apparent pattern as well (Reference Appendix B **"Plots of Residuals against Predictors for Model4"**).

3.5 Test for Interaction Effects

We conduct a partial F-test to check the significance of interaction term, $AllRent \times Cows$. From the output (Refer to Appendix B Test for Interaction Effects), p-value is at 0.9863 thus rejecting the null hypothesis, the interaction term is thus not significant in this model.

3.6 Identification of Potential Outliers

i	h_{ii}	e_i^*	$DFBETAS_{i,0}$	$DFBETAS_{i,1}$	$DFBETAS_{i,2}$	$DFBETAS_{i,3}$	$DFFITS_i$
5	**0.0968**	**-2.31509130**	0.018420	**-0.592025**	**0.314720**	-0.23505	**-0.75796**
12	0.0583	-1.74139350	**0.285814**	**-0.251537**	**-0.253049**	0.22300	-0.43335
19	0.0542	1.95528517	0.064252	**-0.279663**	**0.293629**	**-0.28662**	**0.46820**
33	0.0625	1.88246285	0.079262	**-0.284738**	0.162554	-0.07809	**0.48595**
36	0.0567	**2.21964332**	**-0.314245**	**0.361191**	**0.274048**	**-0.28584**	**0.54414**
52	0.0685	-1.85884472	**-0.312146**	**0.418466**	-0.094146	0.12237	**0.50412**
56	**0.2989**	0.57411535	0.150925	-0.070105	**-0.219865**	0.28917	0.37488
64	0.0795	1.45075213	-0.057042	**0.347171**	-0.111531	0.06746	0.42624

66	**0.1490**	-1.05120153	-0.046297	-0.061272	0.195255	**-0.28700**	**-0.43988**
67	0.0418	**-2.45415027**	-0.243869	-0.174521	**0.299089**	-0.22287	**-0.51276**

A total of 10 observations were tagged to be potential influential or outlier points with values that exceeded various benchmarks to be in **red**. Some observations are suspected of high leverage (far exceeding h_{ii} value) or considerable influence (exceeding DFFITS, DFBETAS or Studentized Residuals). Unsurprisingly, observations 5, 36 and 67 (mentioned prior) are amongst the suspected outliers.

However, further investigation would have to be conducted to check if there were any mistakes in obtaining the data to these observations or if it could be telling of another more suitable model instead.

4 Interpretation

The final model (i.e. Model4), suggests a strong relationship between average rent paid to all arable land and the density of cows. Rather surprisingly, it did not matter if liming was required on the plots or the proportion of pasteurised land. This could be seen from the Stepwise Regressions done in the first step where both liming and pastures were predictors that were eliminated. This falls out of our expectation as intuitively, if a land required liming, rental of land should technically be lower as land is not as fertile. Similarly, low pasteurised land proportion meant less forage for the dairy cows. These variables were thought to be significant factors affecting rent. Further investigation would have to be conducted to find out if there is really no association.

5 Conclusion

The relationship of Average rent payable to land planted with Alfalfa can be expressed as:

$$\widehat{\sqrt{Rent}} = 2.0534307 + 0.0724896\, Cows + 0.0735719\, AllRent - 0.0007732\, Cows^2$$

From the equation, the square-root of Average rent payable to land planted with Alfalfa (*Rent*) is directly proportional to the average rent paid to all arable land (*AllRent*) and density of cows (*Cows*). In contrast, it is inversely proportionate to the squared value of the density of cows ($Cows^2$). We derived this formula as this model has the strongest R-squared reading and fulfils all assumptions on the error terms.

6 Limitations

Due to lack of background knowledge, we are unsure of the sufficiency of the number of predictors and would recommend even deeper study to uncover more potential predictors that could affect the average rent payable for land planted with Alfalfa.

R-squared for our final model has a value of 0.8784 (a slight improvement from Model1). Despite its strong value, more predictors could be added to the model and could possibly improve the model even further.

Since the data was collected in 1977, the information may not be representative of the rent prices in the current market. Also, we only know the origin of the data but are unclear of the method used to collect the data. This could undermine the accuracy of the model in predicting present prices. Perhaps, new data could be collected and further study can be undertaken.

References

StatSci.org **Rent for Land Planted to Alfalfa** Retrieved 5 April, 2017, from
http://www.statsci.org/data/general/landrent.html

Putnam D. et al. (2001). Alfalfa, Wildlife, and the Environment: **The Importance and Benefits of Alfalfa in the 21st Century** Retrieved 9 April, 2017, from
http://agric.ucdavis.edu/files/242006.pdf

Mengel D. (2015). **Liming prior to fall seedings of Alfalfa** Retrieved 9 April, 2017, from
http://www.farmtalknewspaper.com/news/liming-prior-to- fall-seedings-of-alfalfa/article_26b15326-3079-11e5-a106-d3539ef633f8.html

Appendix

A Dataset

```
Rent AllRent  Cows Pasture  Liming
18.38     15.50     17.25     0.24 No
20.00     22.29     18.51     0.20 Yes
11.50     12.36     11.13     0.12 No
25.00     31.84      5.54     0.12 Yes
52.50     83.90      5.44     0.04 No
82.50     72.25     20.37     0.05 Yes
25.00     27.14     31.20     0.27 No
30.67     40.41      4.29     0.10 Yes
12.00     12.42      8.69     0.41 No
61.25     69.42      6.63     0.04 Yes
60.00     48.46     27.40     0.12 No
57.50     69.00     31.23     0.08 No
31.00     26.09     28.50     0.21 Yes
60.00     62.83     29.98     0.17 No
72.50     77.06     13.59     0.05 No
60.33     58.83     45.46     0.16 No
49.75     59.48     35.90     0.32 No
 8.50      9.00      8.89     0.08 No
36.50     20.64     23.81     0.24 No
60.00     81.40      4.54     0.05 Yes
16.25     18.92     29.62     0.72 No
50.00     50.32     21.36     0.19 Yes
11.50     21.33      1.53     0.10 Yes
35.00     46.85      5.42     0.08 Yes
75.00     65.94     22.10     0.09 No
31.56     38.68     14.55     0.17 Yes
48.50     51.19      7.59     0.13 Yes
77.50     59.42     49.86     0.13 No
21.67     24.64     11.46     0.21 Yes
19.75     26.94      2.48     0.10 Yes
56.00     46.20     31.62     0.26 No
25.00     26.86     53.73     0.43 No
40.00     20.00     40.18     0.56 No
56.67     62.52     15.89     0.05 No
51.79     56.00     14.25     0.15 Yes
96.67     71.41     21.37     0.05 No
50.83     65.00     13.24     0.08 Yes
34.33     36.28      5.85     0.10 Yes
```

48.75	59.88	32.99	0.21	No
25.80	23.62	28.89	0.24	Yes
20.00	24.20	6.29	0.06	Yes
16.00	17.09	33.34	0.66	No
48.67	44.56	16.70	0.15	Yes
20.78	34.46	4.20	0.03	Yes
32.50	31.55	23.47	0.19	Yes
19.00	26.94	8.28	0.10	Yes
51.50	58.71	7.40	0.04	Yes
49.17	65.74	7.71	0.02	Yes
85.00	69.05	46.18	0.22	Yes
58.75	57.54	14.98	0.11	Yes
19.33	21.73	6.58	0.06	No
5.00	6.17	13.68	0.18	No
65.00	51.00	50.50	0.24	No
20.00	18.25	16.12	0.32	No
62.50	69.88	31.48	0.07	No
35.00	26.68	58.60	0.23	No
99.17	75.73	35.43	0.05	No
40.25	41.77	4.53	0.08	Yes
39.17	48.50	6.82	0.08	Yes
37.50	21.89	43.70	0.36	No
26.25	38.33	2.83	0.04	Yes
52.14	53.95	42.54	0.25	No
22.50	17.17	24.16	0.36	No
90.00	82.00	7.89	0.03	Yes
28.00	40.60	3.27	0.02	Yes
50.00	53.89	53.16	0.24	No
24.50	54.17	5.57	0.06	Yes

B Selected R Codes & Outputs

Reading Data from Online "txt" file

```
table <- read.table("http://www.statsci.org/data/general/landrent.txt", header = T)
```

Changing Categorical Variable "Liming" Into Dummy 1's & 0's

```
table$Liming <- factor(table$Liming, levels=c("Yes","No"), labels=c(1,0))
table$Liming <- as.character(table$Liming)
table$Liming <- as.numeric(table$Liming)
```

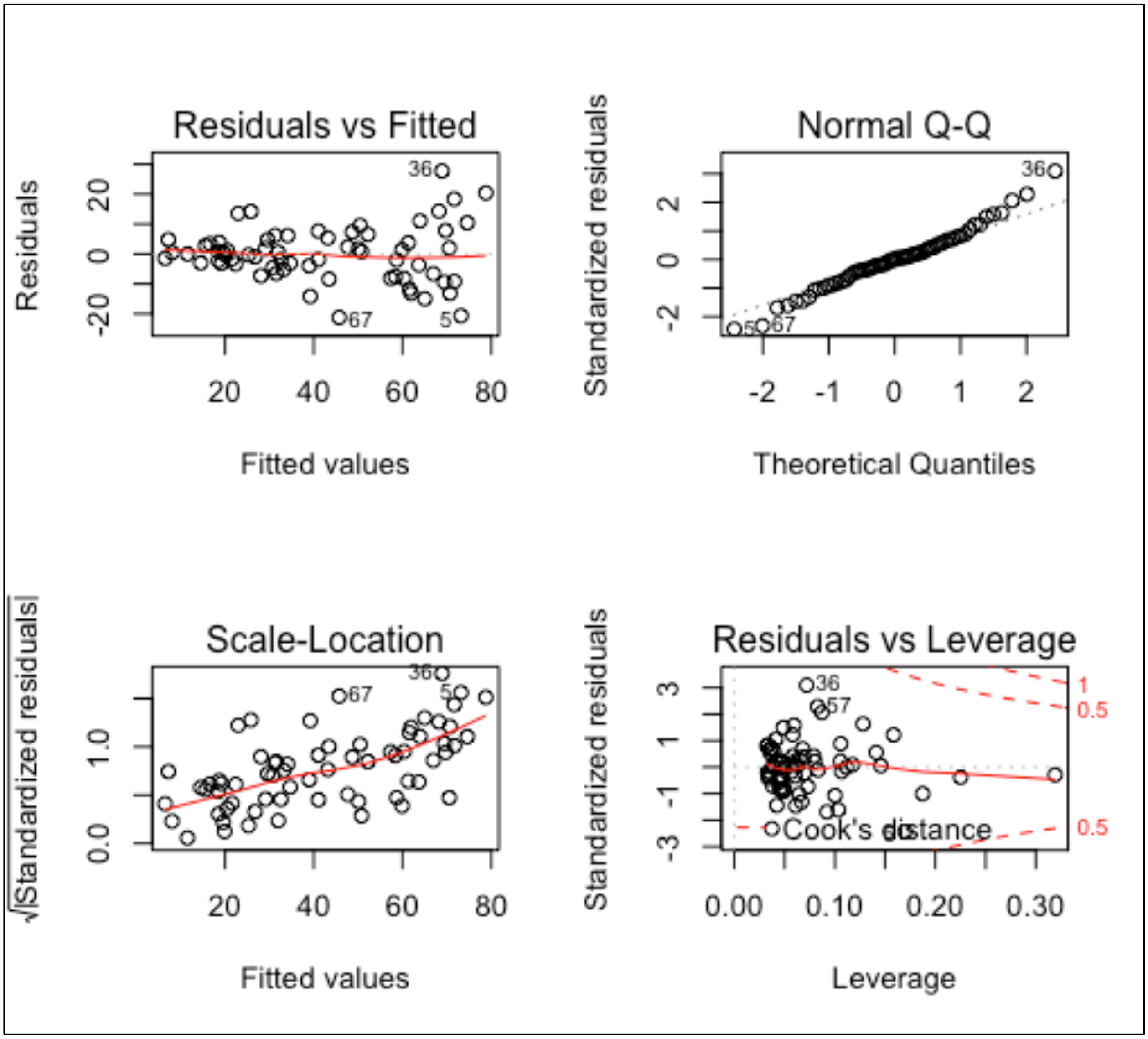
Multicollinearity Check

```
> library("perturb", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")
> colldiag(Model1)
Condition
Index   Variance Decomposition Proportions
          intercept table$AllRent table$Cows table$Pasture table$Liming
1  1.000 0.004     0.007         0.010      0.009         0.011
2  2.073 0.000     0.006         0.038      0.049         0.201
3  3.359 0.000     0.130         0.037      0.218         0.178
4  6.084 0.069     0.143         0.903      0.167         0.466
5  9.655 0.927     0.713         0.014      0.556         0.144
```

Running Overall Regression for Model 1

```
> Model1 <- lm(table$Rent ~ table$AllRent + table$Cows + table$Pasture + table$Liming)
> summary(Model1)

Call:
lm(formula = table$Rent ~ table$AllRent + table$Cows + table$Pasture +
    table$Liming)

Residuals:
     Min       1Q   Median       3Q      Max
-21.2287  -4.8686  -0.0287   4.7547  27.7666

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -2.8282     4.6749  -0.605 0.547399
table$AllRent   0.8833     0.0690  12.801  < 2e-16 ***
table$Cows      0.4318     0.1080   3.999 0.000172 ***
table$Pasture -11.3804    11.8937  -0.957 0.342359
table$Liming   -1.0117     2.8490  -0.355 0.723706
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.311 on 62 degrees of freedom
Multiple R-squared:  0.8404,    Adjusted R-squared:  0.8301
F-statistic:  81.6 on 4 and 62 DF,  p-value: < 2.2e-16
```

Diagnostic Plots for Model 1



Stepwise AIC to obtain Model 2

```
> nullmodel <- lm(table$Rent ~ 1, data = table)
> fullmodel <- lm(table$Rent ~ ., data = table)
> step(nullmodel,data=table, scope=list(upper= fullmodel,lower=nullmodel),direction = "both", k
= 2, test = "F")
Start:  AIC=418.72
table$Rent ~ 1

          Df Sum of Sq   RSS    AIC  F value    Pr(>F)    
+ AllRent  1   25824.2  7846 323.12 213.9425 < 2.2e-16 ***
+ Pasture  1    3521.0 30149 413.32   7.5911  0.007601 ** 
+ Cows     1    3205.9 30464 414.01   6.8404  0.011068 *  
<none>                 33670 418.72                       
+ Liming   1     266.4 33404 420.19   0.5184  0.474115    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=323.12
table$Rent ~ AllRent

          Df Sum of Sq   RSS    AIC  F value    Pr(>F)    
+ Cows     1    2386.3  5460 300.83  27.9736 1.593e-06 ***
+ Liming   1     944.9  6901 316.53   8.7634  0.004306 ** 
+ Pasture  1     586.9  7259 319.92   5.1746  0.026279 *  
<none>                  7846 323.12                       
- AllRent  1   25824.2 33670 418.72 213.9425 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=300.83
table$Rent ~ AllRent + Cows

          Df Sum of Sq     RSS    AIC  F value    Pr(>F)    
<none>                  5459.6 300.83                       
+ Pasture  1      73.9  5385.7 301.92   0.8640    0.3562    
+ Liming   1       5.4  5454.2 302.76   0.0626    0.8032    
- Cows     1    2386.3  7845.9 323.12  27.9736 1.593e-06 ***
- AllRent  1   25004.6 30464.2 414.01 293.1160 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = table$Rent ~ AllRent + Cows, data = table)

Coefficients:
(Intercept)      AllRent         Cows  
    -6.1143       0.9214       0.3925  
```

Mallows Cp to obtain Model 2

```
> step(fullmodel,data=table,scope = list(upper=fullmodel,lower=nullmodel),direction = "both",sc
ale = (summary(fullmodel)$sigma)^2,k=2,test = "F")
Start:  AIC=5
table$Rent ~ AllRent + Cows + Pasture + Liming

          Df Sum of Sq     RSS       Cp  F value    Pr(>F)    
- Liming   1      10.9  5385.7   3.1261   0.1261  0.723706    
- Pasture  1      79.4  5454.2   3.9156   0.9156  0.342359    
<none>                  5374.8   5.0000                       
- Cows     1    1386.3  6761.1  18.9910  15.9910  0.000172 ***
- AllRent  1   14204.8 19579.6 166.8570 163.8570 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=3.13
table$Rent ~ AllRent + Cows + Pasture

          Df Sum of Sq     RSS       Cp  F value    Pr(>F)    
- Pasture  1      73.9  5459.6   1.9781   0.8640    0.3562    
<none>                  5385.7   3.1261                       
+ Liming   1      10.9  5374.8   5.0000   0.1261    0.7237    
- Cows     1    1873.3  7259.0  22.7348  21.9127 1.562e-05 ***
- AllRent  1   14199.0 19584.7 164.9157 166.0935 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=1.98
table$Rent ~ AllRent + Cows

          Df Sum of Sq     RSS       Cp  F value    Pr(>F)    
<none>                  5459.6   1.9781                       
+ Pasture  1      73.9  5385.7   3.1261   0.8640    0.3562    
+ Liming   1       5.4  5454.2   3.9156   0.0626    0.8032    
- Cows     1    2386.3  7845.9  27.5051  27.9736 1.593e-06 ***
- AllRent  1   25004.6 30464.2 288.4139 293.1160 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = table$Rent ~ AllRent + Cows, data = table)

Coefficients:
(Intercept)      AllRent         Cows  
    -6.1143       0.9214       0.3925  
```

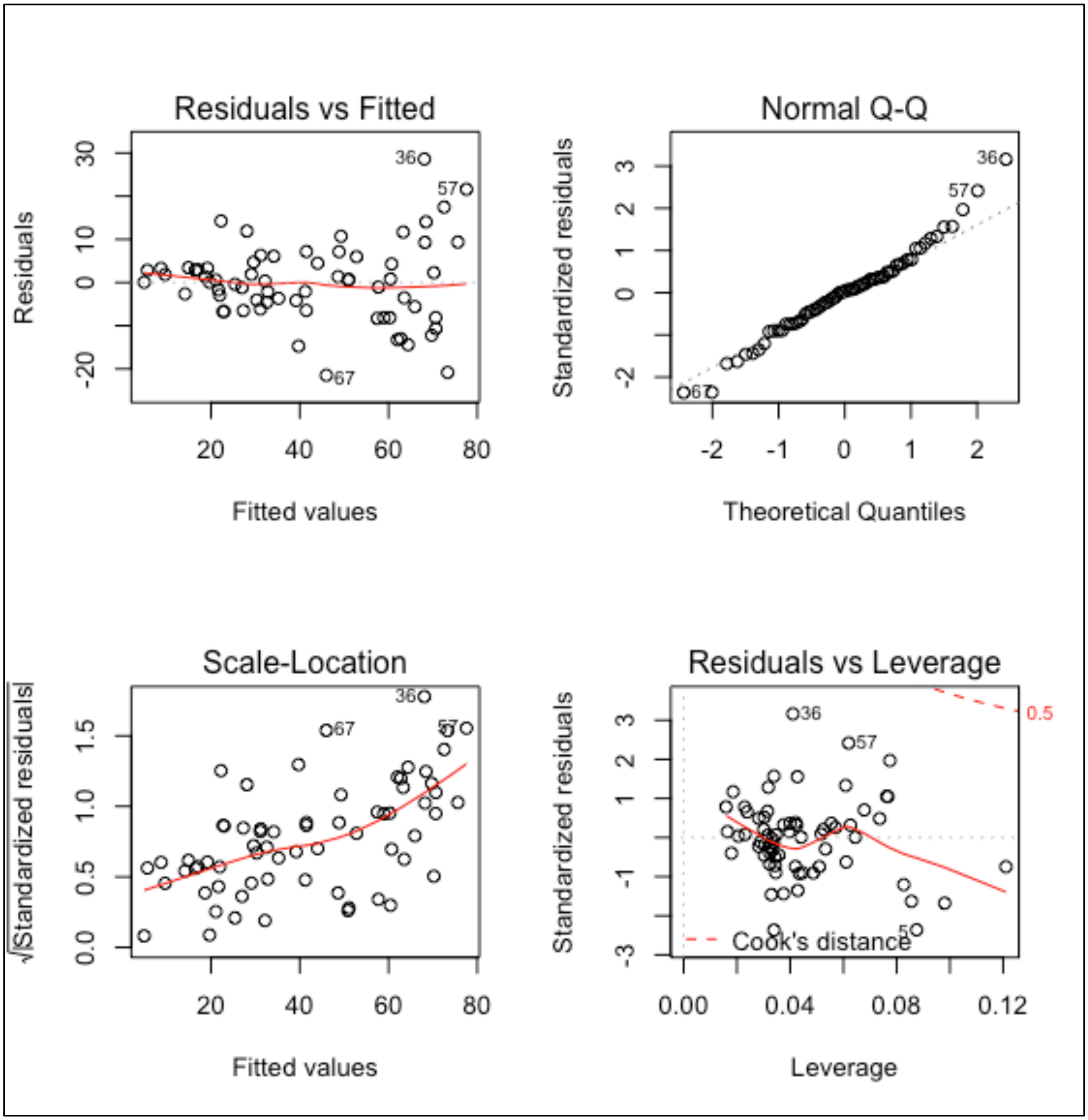
Stepwise Regression via Partial F-test to obtain Model 2

```
#Stepwise regression step by step Partial F-test
cor(table2)
#R= 0.878 for Allrent is highest so pick that first
#decide that 10% inclusion and 5% stay
modelAllrent <- lm(table2$Rent ~ table2$AllRent)
summary(modelAllrent)
modelAllrentCows = lm(table2$Rent ~ table2$AllRent + table2$Cows)
modelAllrentPasture = lm(table2$Rent ~ table2$AllRent + table2$Pasture)
modelAllrentLiming = lm(table2$Rent ~ table2$AllRent + table2$Liming)
anova(modelAllrent,modelAllrentCows)
anova(modelAllrent,modelAllrentPasture)
anova(modelAllrent,modelAllrentLiming)
#find the maximum F stat out of all of them -> choose that predictor
#now check if AllRent can be removed from the model with AllRent and Cows
modelCows = lm(table$Rent ~ table$Cows)
anova(modelCows,modelAllrentCows)
#now check if pasture or liming can be included in the model w Allrent and cows
modelAllrentCowsPasture = lm(table2$Rent ~ table2$AllRent + table2$Cows + table2$Pasture)
modelAllrentCowsLiming = lm(table2$Rent ~ table2$AllRent + table2$Cows + table2$Liming)
anova(modelAllrentCows,modelAllrentCowsLiming)
anova(modelAllrentCows,modelAllrentCowsPasture)
#both cant be included by the 10% inclusion -> final fitted model is table$Rent ~ table$Cows + table$AllRent
```

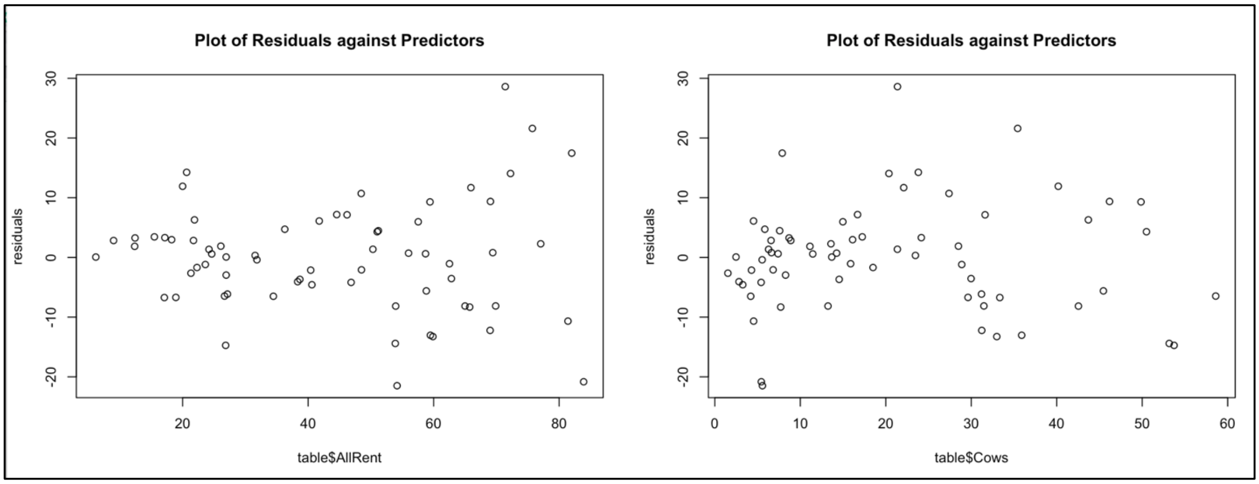
Regression on Model2

```
> Model2 <- lm(table$Rent ~ table$Cows + table$AllRent)
> summary(Model2)

Call:
lm(formula = table$Rent ~ table$Cows + table$AllRent)

Residuals:
     Min       1Q   Median       3Q      Max 
-21.4827  -5.8720   0.3321   4.3855  28.6007 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -6.11433    2.96123  -2.065    0.043 *  
table$Cows     0.39255    0.07422   5.289 1.59e-06 ***
table$AllRent  0.92137    0.05382  17.121  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.236 on 64 degrees of freedom
Multiple R-squared:  0.8379,	Adjusted R-squared:  0.8328 
F-statistic: 165.3 on 2 and 64 DF,  p-value: < 2.2e-16
```

Plots on Model2



Plots of Residuals against Predictors for Model2



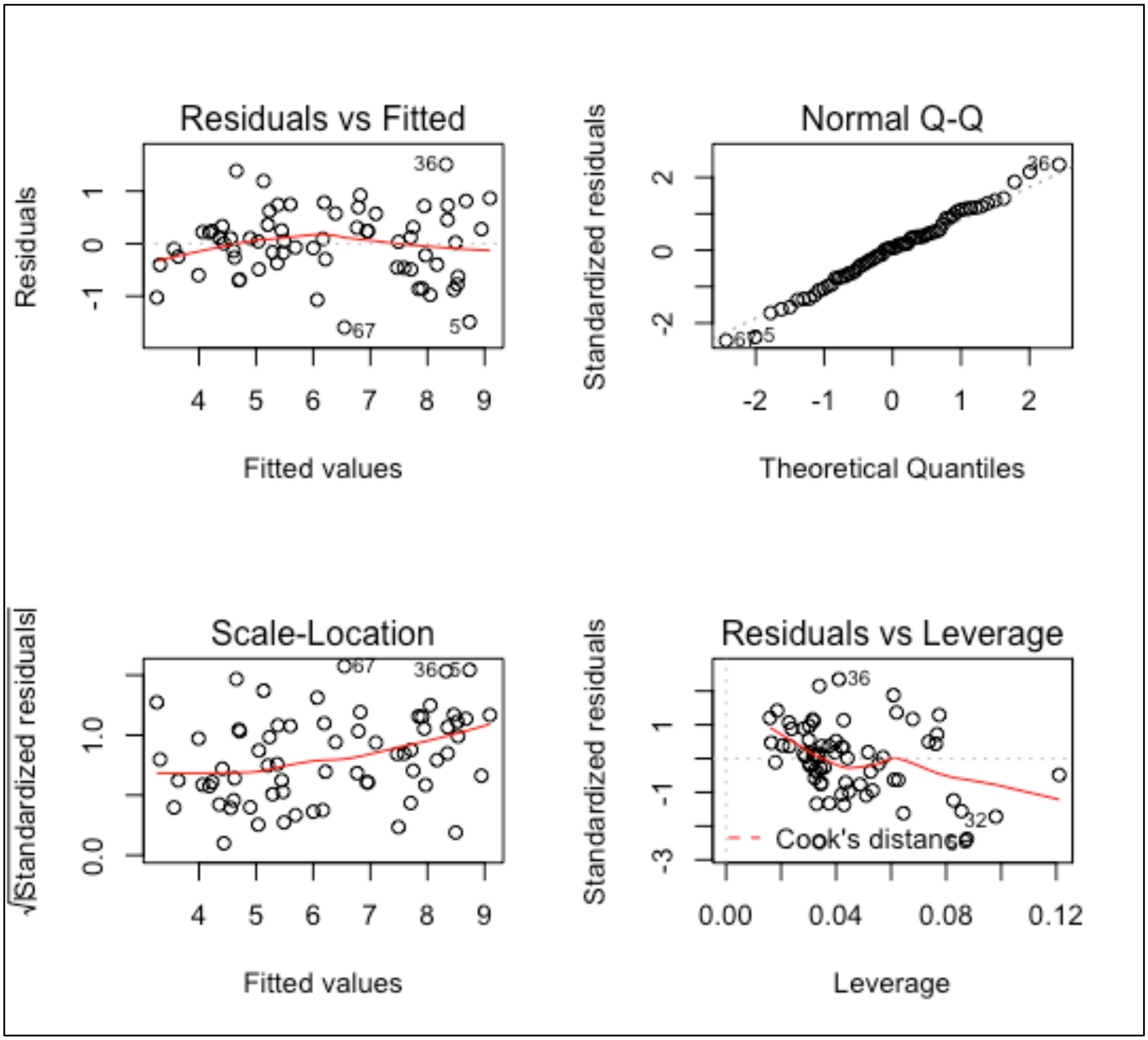
Box Cox Transformation on Model2 to Model3

```
#boxcox transofrmation
library(MASS)
modelbc = boxcox(Model2,lambda = seq(-1,1,0.01))
modelbc$x[modelbc$y==max(modelbc$y)] #0.46, lambda approx 0.5
```

Regression on Model3

```
> Model3<- lm(table$Rent ^ 0.5 ~ table$Cows + table$AllRent )
> summary(Model3)

Call:
lm(formula = table$Rent^0.5 ~ table$Cows + table$AllRent)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.5963 -0.4298  0.0418  0.3468  1.5083 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.370239   0.210117  11.281  < 2e-16 ***
table$Cows    0.031991   0.005266   6.075 7.55e-08 ***
table$AllRent 0.073798   0.003819  19.326  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6554 on 64 degrees of freedom
Multiple R-squared:  0.8685,	Adjusted R-squared:  0.8644 
F-statistic: 211.4 on 2 and 64 DF,  p-value: < 2.2e-16
```

Plots on Model3



Trying out Quadratic Terms to get Model4

```
#trying out quadratic terms (square to predictors)
cows2 <- table$Cows^2
AllRent2 <- table$AllRent^2
nullmodel <- lm(table$Rent^0.5 ~ 1, data = table)
fullmodel <- lm(table$Rent^0.5 ~ table$AllRent + table$Cows + cows2 + AllRent2 )
```

Stepwise AIC for Quadratic Terms Model

```
> step(nullmodel,data=table2, scope=list(upper= fullmodel,lower=nullmodel),direction = "both", k = 2, test = "F")
Start:  AIC=78.25
table$Rent^0.5 ~ 1

                 Df Sum of Sq     RSS     AIC F value    Pr(>F)    
+ table$AllRent  1   165.760  43.336 -25.192 248.622 < 2.2e-16 ***
+ AllRent2       1   152.504  56.592  -7.311 175.162 < 2.2e-16 ***
+ table$Cows     1    21.193 187.903  73.093   7.331  0.008651 ** 
+ cows2          1    15.538 193.558  75.079   5.218  0.025629 *  
<none>                       209.096  78.253                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=-25.19
table$Rent^0.5 ~ table$AllRent

                 Df Sum of Sq     RSS     AIC  F value    Pr(>F)    
+ table$Cows     1    15.849  27.488 -53.694  36.9007 7.550e-08 ***
+ cows2          1    11.810  31.527 -44.509  23.9741 6.946e-06 ***
+ AllRent2       1     1.426  41.911 -25.433   2.1773     0.145    
<none>                        43.336 -25.192                       
- table$AllRent  1   165.760 209.096  78.253 248.6221 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=-53.69
table$Rent^0.5 ~ table$AllRent + table$Cows

                 Df Sum of Sq     RSS     AIC  F value    Pr(>F)    
+ cows2          1     2.057  25.431 -56.904   5.0950   0.02747 *  
<none>                        27.488 -53.694                       
+ AllRent2       1     0.568  26.920 -53.092   1.3286   0.25340    
- table$Cows     1    15.849  43.336 -25.192  36.9007  7.55e-08 ***
- table$AllRent  1   160.416 187.903  73.093 373.4979 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=-56.9
table$Rent^0.5 ~ table$AllRent + table$Cows + cows2

                 Df Sum of Sq     RSS     AIC  F value    Pr(>F)    
+ AllRent2       1     0.981  24.450 -57.539   2.4869 0.1198852    
<none>                        25.431 -56.904                       
- cows2          1     2.057  27.488 -53.694   5.0950 0.0274708 *  
- table$Cows     1     6.096  31.527 -44.509  15.1006 0.0002475 ***
- table$AllRent  1   159.316 184.747  73.958 394.6707 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Step:  AIC=-57.54
table$Rent^0.5 ~ table$AllRent + table$Cows + cows2 + AllRent2
```

Continued Stepwise AIC

```
                  Df Sum of Sq    RSS     AIC F value    Pr(>F)    
<none>                         24.450 -57.539                      
- AllRent2        1    0.9807 25.431 -56.904  2.4869 0.1198852    
- cows2           1    2.4697 26.920 -53.092  6.2625 0.0149824 *  
- table$Cows      1    6.6062 31.056 -43.515 16.7516 0.0001253 ***
- table$AllRent   1   12.5140 36.964 -31.848 31.7324 4.596e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = table$Rent^0.5 ~ table$AllRent + table$Cows + cows2 + 
    AllRent2, data = table)

Coefficients:
  (Intercept)  table$AllRent     table$Cows          cows2       AllRent2  
    1.5523161      0.1013519      0.0760176     -0.0008579     -0.0003106  
```

Mallows Cp for Quadratic Terms Model

```
> library(leaps)
Warning message:
package ‘leaps’ was built under R version 3.3.2 
> x<- model.matrix(fullmodel)[,-1]
> allcp <- leaps(x,table$Rent,method = "Cp")
> for(i in 2:5){
+ mincp <- min(allcp$Cp[allcp$size==i])
+ whichmodel <- allcp$which[allcp$Cp==mincp,]
+ namemodel <-names(whichmodel)[whichmodel==T]
+ cat(namemodel,"\n",mincp,"\n")
+ }
1 
 34.87669 
1 2 
 7.107673 
1 2 3 
 3.508327 
1 2 3 4 
 5 
> #based on the output stopping at cp=3.50 for 1 2 3 would be the closest to number of parameters = 4
```

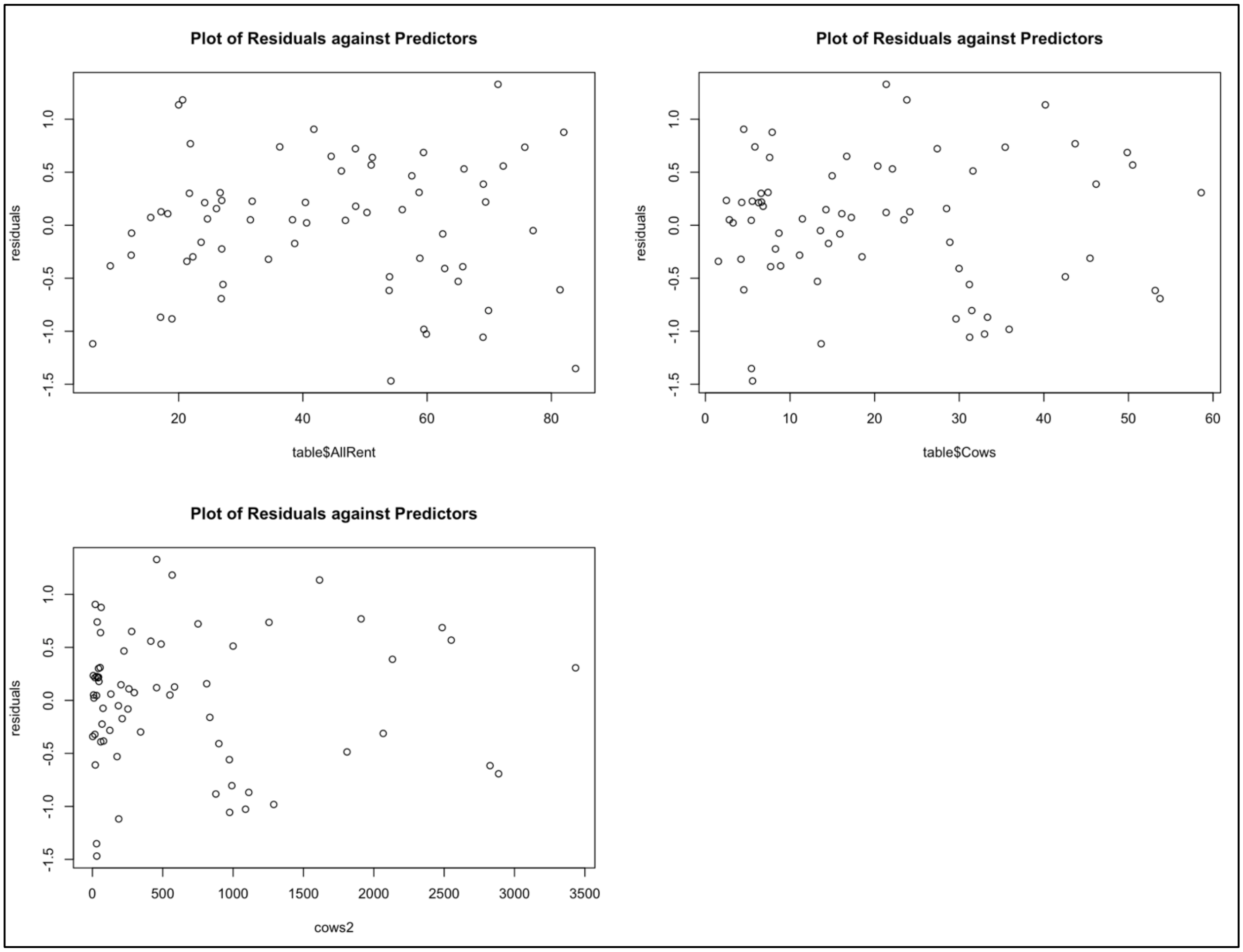
Stepwise Regression via F-test

```
#Stepwise regression Partial F-test
#10% entry 5% stay
fullmodel <- lm(table$Rent^0.5 ~ table$AllRent + table$Cows + cows2 + AllRent2 )
#AR - 1 , C -2 , AR2 - 3 , C2 - 4
model123 <- lm(table$Rent^0. ~ table$AllRent + table$Cows + AllRent2)
model134 <- lm(table$Rent^0.5 ~ table$AllRent + cows2 + AllRent2)
model124 <- lm(table$Rent^0.5 ~ table$AllRent + table2$Cows + cows2)
model234 <- lm(table$Rent^0.5 ~ table$Cows + cows2 + AllRent2)
#10% entry 5% stay
anova(model123,fullmodel)
anova(model134,fullmodel)
anova(model124,fullmodel)
anova(model234,fullmodel)
#look at minimum F stat -> take out AllRent2 (3)
#model124 chosen , now see if can take out anymore?
model24 <- lm(table2$Rent^0.5 ~  table2$Cows + cows2)
model14 <- lm(table2$Rent^0.5 ~ table2$AllRent +  cows2)
model12 <- lm(table2$Rent^0.5 ~ table2$AllRent + table2$Cows)
anova(model24,model124)
anova(model14,model124)
anova(model12,model124)
#conclusion: final model still only requires Allrent , cows and cows2
```

Regression on Model4

```
> Model4 <- lm(table$Rent^0.5 ~ table$AllRent + table$Cows + cows2)
> summary(Model4)

Call:
lm(formula = table$Rent^0.5 ~ table$AllRent + table$Cows + cows2)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.46885 -0.38713  0.05968  0.42654  1.32890 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    2.0534307  0.2473727   8.301 1.06e-11 ***
table$AllRent  0.0735719  0.0037033  19.866  < 2e-16 ***
table$Cows     0.0724896  0.0186542   3.886 0.000247 ***
cows2         -0.0007732  0.0003426  -2.257 0.027471 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6353 on 63 degrees of freedom
Multiple R-squared:  0.8784,	Adjusted R-squared:  0.8726 
F-statistic: 151.7 on 3 and 63 DF,  p-value: < 2.2e-16
```

Plots of Residuals against Predictors for Model 4



Testing for Interaction Effects

```
> interactionmodel <- lm(table$Rent^0.5 ~ table$AllRent + table$Cows + cows2 +table$AllRent*table$Cows)
> summary(interactionmodel)

Call:
lm(formula = table$Rent^0.5 ~ table$AllRent + table$Cows + cows2 +
    table$AllRent * table$Cows)

Residuals:
     Min       1Q   Median       3Q      Max
-1.46952 -0.38693  0.06029  0.42796  1.32903

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              2.049e+00  3.453e-01   5.934 1.44e-07 ***
table$AllRent            7.366e-02  6.489e-03  11.352  < 2e-16 ***
table$Cows               7.272e-02  2.312e-02   3.145  0.00255 **
cows2                   -7.736e-04  3.461e-04  -2.235  0.02901 *
table$AllRent:table$Cows -4.792e-06  2.783e-04  -0.017  0.98632
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6404 on 62 degrees of freedom
Multiple R-squared:  0.8784,	Adjusted R-squared:  0.8705
F-statistic: 111.9 on 4 and 62 DF,  p-value: < 2.2e-16

> anova(interactionmodel,Model4)
Analysis of Variance Table

Model 1: table$Rent^0.5 ~ table$AllRent + table$Cows + cows2 + table$AllRent *
    table$Cows
Model 2: table$Rent^0.5 ~ table$AllRent + table$Cows + cows2
  Res.Df    RSS Df   Sum of Sq     F Pr(>F)
1     62 25.431
2     63 25.431 -1 -0.00012159 3e-04 0.9863
```

Test for Outliers and Influential Points

```
> influence.measures(Model4)
Influence measures of
         lm(formula = table$Rent^0.5 ~ table$AllRent + table$Cows + cows2) :

      dfb.1_  dfb.t.AR  dfb.tb.C dfb.cws2    dffit cov.r   cook.d    hat inf
1   0.011298 -0.019914  0.011353 -0.01244  0.02766 1.124 1.94e-04 0.0528
2  -0.028782  0.061718 -0.051843  0.05535 -0.10150 1.098 2.61e-03 0.0434
3  -0.079070  0.083647 -0.002521  0.01151 -0.10763 1.111 2.93e-03 0.0532
4   0.065615 -0.022306 -0.042340  0.03171  0.07447 1.102 1.41e-03 0.0408
5   0.018420 -0.592025  0.314720 -0.23505 -0.75796 0.847 1.34e-01 0.0968   *
6  -0.124757  0.152081  0.104192 -0.11089  0.22200 1.073 1.24e-02 0.0570
7   0.008111  0.095979 -0.139335  0.12115 -0.20558 1.065 1.06e-02 0.0494
8   0.058901 -0.003272 -0.051775  0.04083  0.07513 1.109 1.43e-03 0.0456
9  -0.024828  0.021811  0.005082 -0.00224 -0.02915 1.128 2.16e-04 0.0558
10  0.009449  0.057071 -0.035657  0.02477  0.08603 1.121 1.88e-03 0.0563
11 -0.095523  0.025572  0.177294 -0.16628  0.23174 1.017 1.34e-02 0.0383
12  0.285814 -0.251537 -0.253049  0.22300 -0.43335 0.935 4.55e-02 0.0583
```

```
13 -0.000653 -0.028179  0.040079 -0.03662  0.05784 1.118 8.49e-04 0.0502
14  0.091880 -0.070352 -0.097858  0.08830 -0.14980 1.091 5.66e-03 0.0495
15  0.007702 -0.016441 -0.002620  0.00429 -0.02049 1.133 1.07e-04 0.0597
16  0.032224 -0.041934  0.002487 -0.03165 -0.13275 1.121 4.46e-03 0.0647
17  0.194553 -0.137445 -0.192979  0.14526 -0.34915 0.950 2.97e-02 0.0454
18 -0.135002  0.126117  0.023287 -0.00890 -0.16059 1.110 6.51e-03 0.0628
19  0.064252 -0.279663  0.293629 -0.28662  0.46820 0.887 5.25e-02 0.0542
20 -0.016918 -0.243467  0.159632 -0.12384 -0.32937 1.105 2.71e-02 0.0963
21 -0.032388  0.226004 -0.233079  0.20842 -0.37030 0.995 3.37e-02 0.0615
22 -0.011932  0.006801  0.024145 -0.02486  0.03514 1.100 3.14e-04 0.0329
23 -0.165803  0.068622  0.125794 -0.10643 -0.17171 1.144 7.45e-03 0.0865
24  0.009334  0.002008 -0.008979  0.00675  0.01456 1.109 5.38e-05 0.0386
25 -0.107076  0.110277  0.110599 -0.11349  0.19320 1.069 9.37e-03 0.0486
26 -0.009928  0.007835 -0.014946  0.01922 -0.04164 1.086 4.40e-04 0.0228
27  0.080275  0.052175 -0.075824  0.04743  0.18048 1.028 8.14e-03 0.0303
28 -0.024104  0.102439 -0.112527  0.19633  0.38771 1.094 3.74e-02 0.1031
29  0.011671 -0.010390  0.000809 -0.00270  0.01737 1.101 7.67e-05 0.0325
30  0.096842 -0.033704 -0.074890  0.06217  0.10324 1.135 2.70e-03 0.0692
31 -0.064594  0.006001  0.121185 -0.10504  0.16444 1.062 6.80e-03 0.0387
32 -0.191199  0.136064  0.236512 -0.34733 -0.53845 1.169 7.20e-02 0.1679
33  0.079262 -0.284738  0.162554 -0.07809  0.48595 0.910 5.67e-02 0.0625
34  0.008335 -0.014518 -0.008869  0.01084 -0.02497 1.104 1.58e-04 0.0355
35 -0.005465  0.017182  0.010846 -0.01492  0.03883 1.092 3.83e-04 0.0270
36 -0.314245  0.361191  0.274048 -0.28584  0.54414 0.832 6.97e-02 0.0567
37  0.038586 -0.108934 -0.024731  0.04169 -0.16615 1.057 6.93e-03 0.0369
38  0.190406 -0.042328 -0.131572  0.09676  0.23183 1.011 1.33e-02 0.0366
39  0.215907 -0.148843 -0.233965  0.19668 -0.36865 0.936 3.30e-02 0.0460
40 -0.001517  0.032816 -0.041302  0.03746 -0.06148 1.122 9.59e-04 0.0537
41  0.067482 -0.036832 -0.033320  0.02383  0.07330 1.107 1.36e-03 0.0443
42 -0.042090  0.239923 -0.212485  0.17472 -0.37563 1.002 3.47e-02 0.0650
43 -0.006083  0.004557  0.085134 -0.09786  0.16465 1.021 6.77e-03 0.0246
44 -0.100431  0.023161  0.078131 -0.06188 -0.11579 1.101 3.39e-03 0.0481
45 -0.000468 -0.006232  0.011736 -0.01157  0.01658 1.110 6.98e-05 0.0402
46 -0.056001  0.033198  0.019167 -0.01025 -0.06676 1.095 1.13e-03 0.0340
47  0.026708  0.047421 -0.039683  0.02551  0.09712 1.090 2.39e-03 0.0373
48 -0.014962 -0.086584  0.047372 -0.02926 -0.13690 1.089 4.73e-03 0.0456
49 -0.060915  0.092615 -0.013349  0.05125  0.18963 1.132 9.08e-03 0.0823
50 -0.027193  0.061077  0.042077 -0.05420  0.12819 1.060 4.14e-03 0.0290
51  0.097874 -0.059463 -0.043712  0.03054  0.10619 1.101 2.85e-03 0.0463
52 -0.312146  0.418466 -0.094146  0.12237 -0.50412 0.921 6.12e-02 0.0685
53  0.017646  0.035901 -0.105584  0.17750  0.32380 1.125 2.63e-02 0.1053
54  0.016697 -0.026374  0.014204 -0.01621  0.03802 1.115 3.67e-04 0.0459
55  0.220051 -0.196758 -0.189689  0.16627 -0.33116 1.016 2.71e-02 0.0598
56  0.150925 -0.070105 -0.219865  0.28917  0.37488 1.489 3.55e-02 0.2989   *
57 -0.222934  0.223265  0.146980 -0.11357  0.33125 1.045 2.72e-02 0.0702
58  0.238455 -0.002434 -0.212824  0.16637  0.31451 0.972 2.43e-02 0.0437
59  0.028357  0.010182 -0.025736  0.01759  0.05150 1.096 6.73e-04 0.0320
```

```
60  0.074077 -0.176350  0.043584  0.02599  0.34075 1.034 2.88e-02 0.0682
61  0.016923 -0.001685 -0.015606  0.01283  0.02052 1.132 1.07e-04 0.0587
62  0.053602 -0.041351 -0.035189 -0.00305 -0.17701 1.077 7.88e-03 0.0487
63  0.009032 -0.033611  0.031277 -0.03034  0.05192 1.132 6.84e-04 0.0607
64 -0.057042  0.347171 -0.111531  0.06746  0.42624 1.013 4.46e-02 0.0795
65  0.006765 -0.000262 -0.006369  0.00518  0.00856 1.127 1.86e-05 0.0541
66 -0.046297 -0.061272  0.195255 -0.28700 -0.43988 1.167 4.83e-02 0.1490
67 -0.243869 -0.174521  0.299089 -0.22287 -0.51276 0.768 6.09e-02 0.0418   *
```

Studentized Residuals for Model4

```
> rstudent(Model4)
          1           2           3           4           5           6           7           8           9          10 
 0.11717622 -0.47671686 -0.45428433  0.36097198 -2.31509130  0.90335308 -0.90134997  0.34368414 -0.11994170  0.35208318 
         11          12          13          14          15          16          17          18          19          20 
 1.16123467 -1.74139350  0.25156119 -0.65611727 -0.08134318 -0.50454656 -1.60138154 -0.62033186  1.95528517 -1.00907470 
         21          22          23          24          25          26          27          28          29          30 
-1.44704807  0.19043260 -0.55791073  0.07266203  0.85519094 -0.27246301  1.02162306  1.14329128  0.09474515  0.37853708 
         31          32          33          34          35          36          37          38          39          40 
 0.81949333 -1.19855379  1.88246285 -0.13011097  0.23296217  2.21964332 -0.84862450  1.18886642 -1.67817979 -0.25807896 
         41          42          43          44          45          46          47          48          49          50 
 0.34035842 -1.42454269  1.03596394 -0.51486506  0.08103411 -0.35583897  0.49338574 -0.62663266  0.63339161  0.74122403 
         51          52          53          54          55          56          57          58          59          60 
 0.48205067 -1.85884472  0.94403006  0.17333874 -1.31361224  0.57411535  1.20517139  1.47041502  0.28347615  1.25908244 
         61          62          63          64          65          66          67 
 0.08221151 -0.78230971  0.20425357  1.45075213  0.03577094 -1.05120153 -2.45415027 
```