Rachel Prokopius

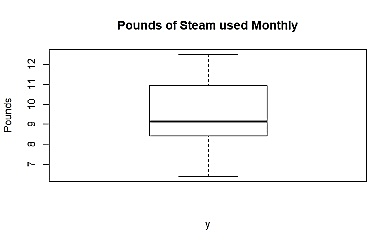
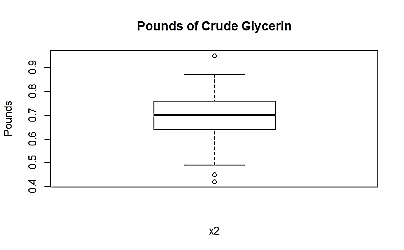
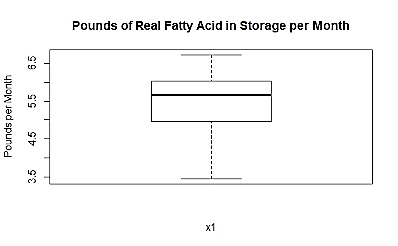
Regression Analysis

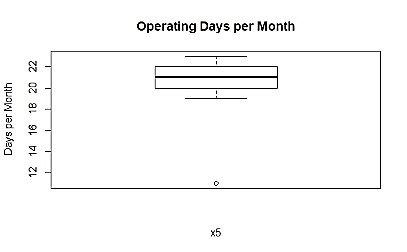
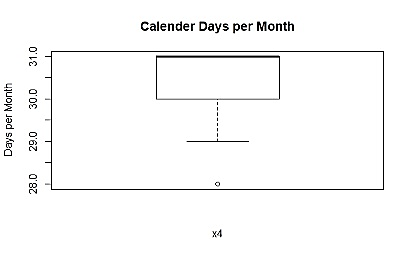
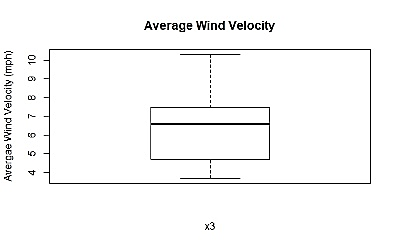
Final Exam

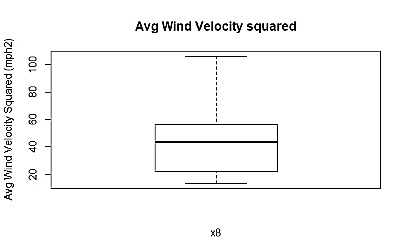
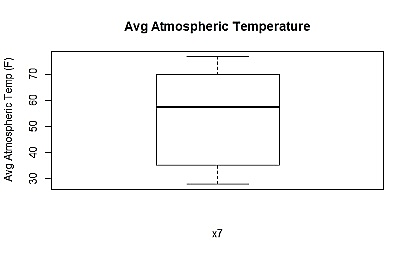
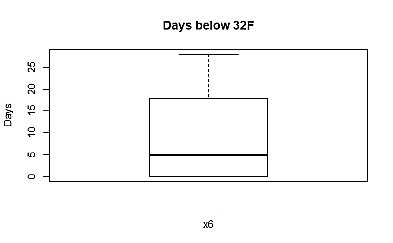
21 April 2020

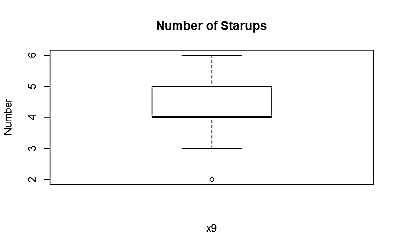
1. The following are box plots of the y variable and x variables for observations at a steam plant. Each boxplot is separate because each variable has different units.

**Figure 1**: Boxplots depicting the spread of the y variable and the regressor variables of observations at a steam plant.

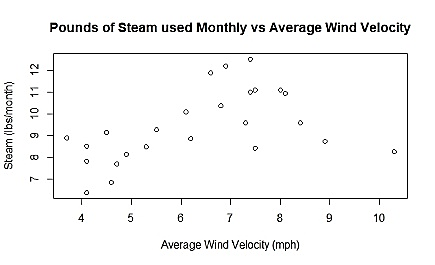
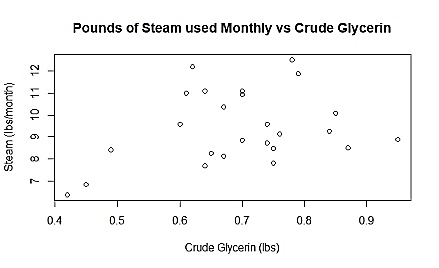
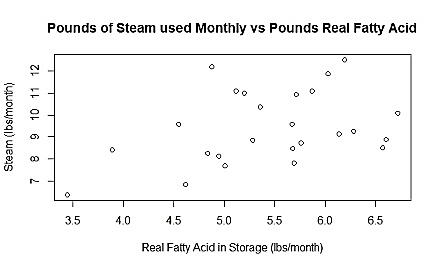


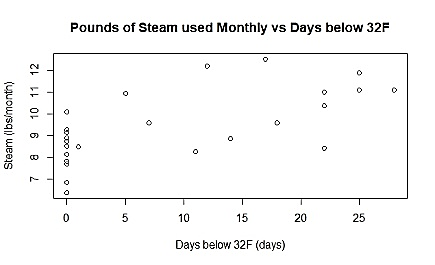
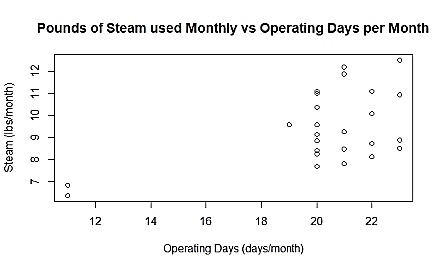
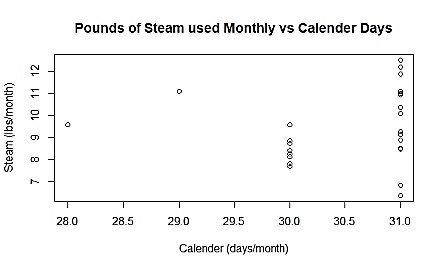


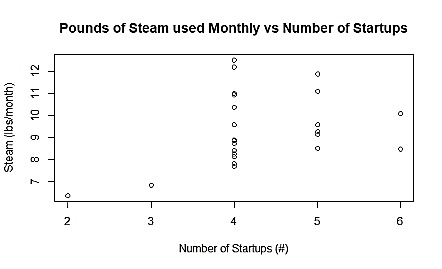
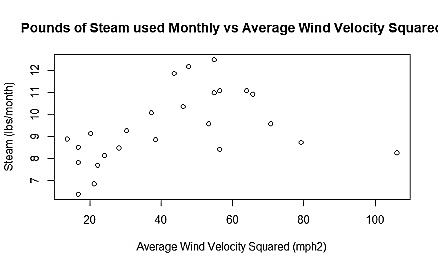
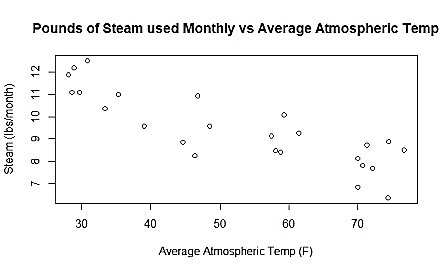


One of the requirements for a data set so that statistical analyses are valid is normal distribution, which means equal distribution of data points around the mean that have roughly 95% of the data points within two standard deviations of the mean and no outliers. Based on the box plots from this data set, certain variables appear to fulfill this requirement and others appear not to. The pounds of steam used monthly (y), pounds of real fatty acid in storage per month (x1), average wind velocity (x3), average atmospheric temperature (x7) and average wind velocity squared (x8) appear to have a normal distribution and no outliers. On the other hand, pounds of crude glycerin (x2) has outliers at both extremes, calendar days per month (x4), operating days per month (x5) and number of startups (x9) have high extreme outliers, and calendar days per month (x4) and days below 32F (x6) are missing extremes for the boxplot in the data set. Therefore, certain variables appear to fulfill one of the requirements for statistical analysis while others do not.

1. **Figure 2**: The following are scatterplots relating steam (lbs/month) to each regressor variable accounted for in observations from a steam plant.







Based on the scatter plots, certain regressor variables seem to have a linear relationship with steam (lbs/month) while others do not. Real fatty acid (lbs/month) and average atmospheric temperature (F) seem to have a fairly strong linear relationship with steam (lbs/month). Crude glycerin (lbs), average wind velocity (mph) and average wind velocity squared (mph2) seem to have more of a parabolic relationship with steam (lbs/month). Days below 32F may have a linear relationship with steam (lbs/month) but there are quite a few data points where there is a decent spread of steam (lbs/month) with zero days below 32F, so this regressor does not seem particularly linear. The same could be said for operating days per month for slightly different reasons; though fewer operating days seem to correlate with lower steam in lbs/month, more operating days correlates with a wide range of steam in lbs. Finally, calendar days and number of startups do not really seem to have a linear relationship with steam in lbs/month, but it is difficult to tell because, like operating days and number of days below 32F, it is a discrete numerical variable compared to a continuous numerical variable. Further analysis for these variables are warranted.

1. The full model for relating the y variable and regressors for the steam model is as follows (**note: coefficients that correspond to regressors that are measured in the same units as the y variable (lbs/month) remain unitless in order to correctly relate the regressor to the y variable**):

**Steam (lbs/month) = 0.700(x1) -1.868[month-1](x2) +1.140[lbs\*hr\*month-1\*miles-1](x3) + 0.123[lbs\*days-1](x4) +0.180[lbs\*days-1](x5) -0.018[lbs\*days-1\*month-1](x6) -0.077[lbs\*ºF-1\*month-1](x7) -0.086[lbs\*hr2\*miles-2\*month-1](x8) -0.346[lbs\*number of startups-1\*month-1](x9) +1.761 (lbs/month)**

The summary statistics for the full model area as follows:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.76116 6.96637 0.253 0.803847

x1 0.70084 0.56248 1.246 0.231880

x2 -1.86794 4.12852 -0.452 0.657421

x3 1.14038 0.74289 1.535 0.145591

x4 0.12253 0.20374 0.601 0.556546

x5 0.17957 0.08060 2.228 0.041619 \*

x6 -0.01831 0.02440 -0.751 0.464557

x7 -0.07734 0.01652 -4.681 0.000295 \*\*\*

x8 -0.08626 0.05178 -1.666 0.116445

x9 -0.34610 0.20979 -1.650 0.119777

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

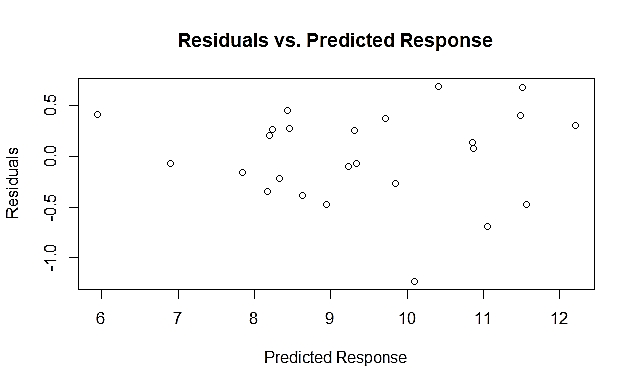
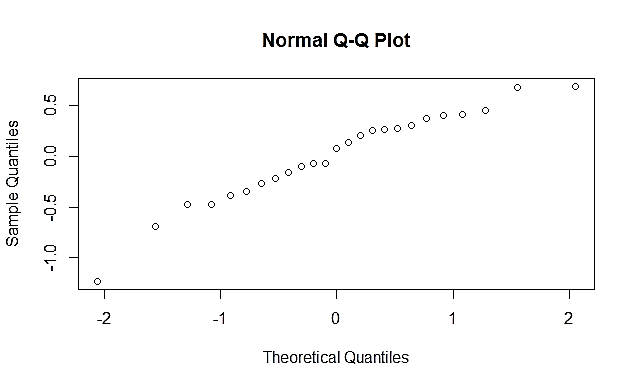
Residual standard error: 0.5673 on 15 degrees of freedom

Multiple R-squared: 0.9242, Adjusted R-squared: 0.8788

F-statistic: 20.33 on 9 and 15 DF, p-value: 7.576e-07

Though the adjusted R2 value is fairly high (0.8788), only two of the regressors (operating days per month and average atmospheric temperature) are significant, so the model likely could use some alteration.

**Figure 3.1:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model and a plot of the residuals from the full model versus the y-values the constructed model predicts.



**Table 3.1**: The following is a table of the standardized residuals for the full model relating steam (lbs/month) to all of the provided regressor variables.

**1 2 3 4 5 6**

0.2542561 1.4748225 0.6363896 0.6141057 -0.1488749 0.6723499

**7 8 9 10 11 12**

1.1986003 0.5678766 -0.7316112 -0.2035156 -1.7552012 1.6696907

**13 14 15 16 17 18**

0.8503303 0.6887397 0.1596926 -0.5628649 0.8309716 -0.4543152

**19 20 21 22 23 24**

-0.2101913 1.1567576 -0.3102859 -1.1720115 -2.4195457 -1.3487168

**25**

-1.0444930

**Table 3.2:** The following is a table of the studentized residuals for the full model relating steam (lbs/month) to all of the provided regressor variables.

**1 2 3 4 5 6**

0.2461658 1.5409084 0.6232825 0.6008842 -0.1439332 0.6595664

**7 8 9 10 11 12**

1.2177407 0.5546151 -0.7197621 -0.1968867 -1.9022436 1.7877397

**13 14 15 16 17 18**

0.8420427 0.6761632 0.1544091 -0.5496144 0.8219363 -0.4419615

**19 20 21 22 23 24**

-0.2033638 1.1709832 -0.3007313 -1.1879701 -2.9935535 -1.3899885

**25**

-1.0479030

**Table 3.3:** The following is a table of the VIF values for the full model relating steam (lbs/month) to all of the provided regressor variables.

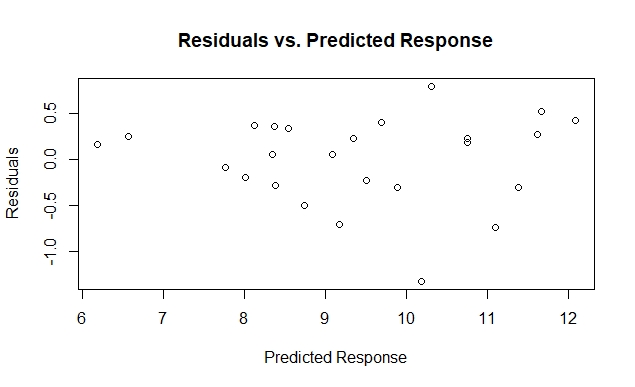
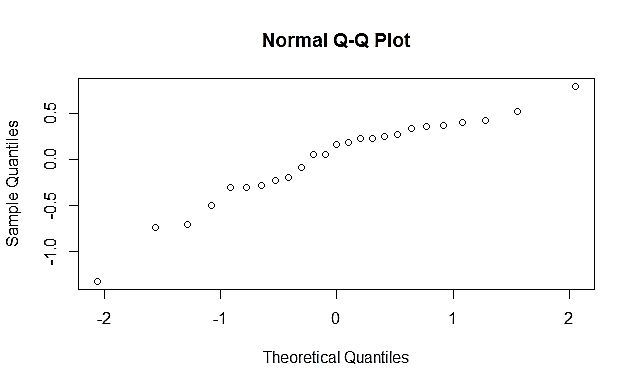
**x1 x2 x3 x4 x5 x6**

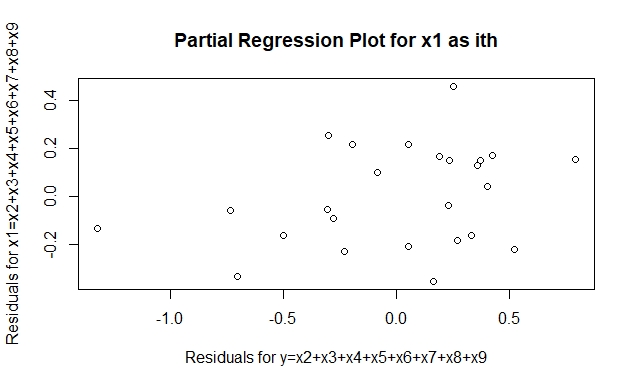
15.746595 20.137114 126.625618 1.836626 4.411920 4.695013

**x7 x8 x9**

6.067426 107.590891 2.385046

**Figure 3.2:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except pounds of real fatty acid per month (x1), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x1 versus the residuals for the x1 explained by the other regressor variables.





**Table 3.4**: The following is a table of the standardized residuals for the partial model without x1 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.4441977 1.6510694 0.8543183 0.1531863 -0.4543402 0.8720314

**7 8 9 10 11 12**

0.4103772 0.7668158 -0.3919954 0.1078524 -2.0623675 1.2142951

**13 14 15 16 17 18**

0.5545728 0.6053296 0.4001498 -0.6318953 0.8814951 -0.5757276

**19 20 21 22 23 24**

0.6058804 0.8247769 -0.1650782 -1.5598131 -2.5312539 -1.4006357

**25**

-0.6138719

**Table 3.5**: The following is a table of the studentized residuals for the partial model without x1 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.4327693 1.7551351 0.8467278 0.1484309 -0.4427785 0.8651502

**7 8 9 10 11 12**

0.3994538 0.7564971 -0.3813837 0.1044656 -2.3305308 1.2339697

**13 14 15 16 17 18**

0.5421991 0.5929368 0.3893967 -0.6196102 0.8750158 -0.5633113

**19 20 21 22 23 24**

0.5934890 0.8161244 -0.1599726 -1.6401235 -3.1652623 -1.4478207

**25**

-0.6015045

**Table 3.6**: The following is a table of the VIF values for the partial model without x1 relating steam (lbs/month) to all of the remaining regressor variables.

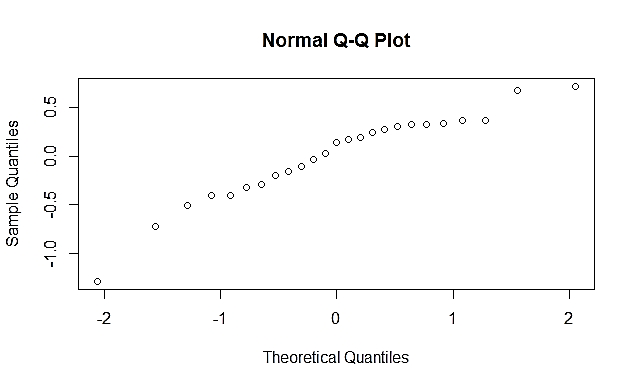
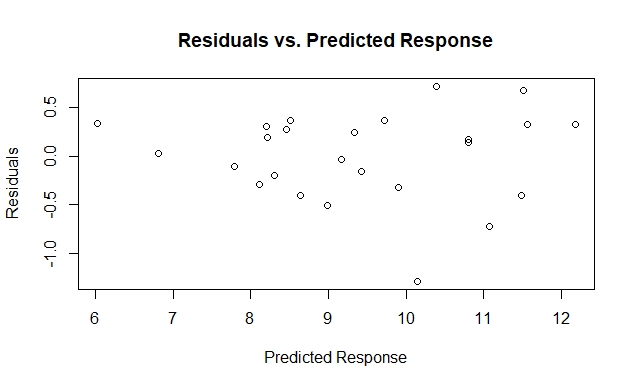
**x2 x3 x4 x5 x6 x7**

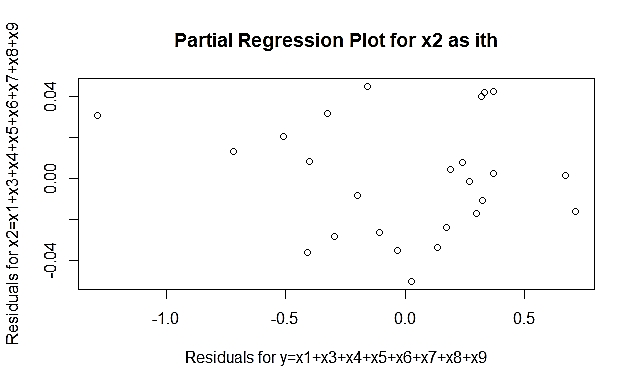
5.180428 117.669441 1.370684 4.024388 4.667488 6.064969

**x8 x9**

100.986695 2.232906

**Figure 3.3:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except pounds of crude glycerin (x2), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x2 versus the residuals for the x2 explained by the other regressor variables.



**Table 3.7**: The following is a table of the standardized residuals for the partial model without x2 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.34458364 1.56370197 0.69349876 0.60394268 -0.30824780 0.69710608

**7 8 9 10 11 12**

0.89015998 0.64587856 -0.61682186 -0.06514238 -1.85042861 1.70507243

**13 14 15 16 17 18**

0.66716441 0.66500086 0.29151391 -0.67997466 0.84178824 -0.43011399

**19 20 21 22 23 24**

0.07162692 0.89155551 -0.21381134 -1.27219570 -2.52041442 -1.42484640

**25**

-0.87446700

**Table 3.8**: The following is a table of the studentized residuals for the partial model without x2 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.33488660 1.64494915 0.68180255 0.59154650 -0.29934982 0.68545982

**7 8 9 10 11 12**

0.88406284 0.63368463 -0.60446533 -0.06308221 -2.02091565 1.82504421

**13 14 15 16 17 18**

0.65515642 0.65297132 0.28300970 -0.66810683 0.83372911 -0.41888477

**19 20 21 22 23 24**

0.06936359 0.88552121 -0.20731833 -1.29926334 -3.14275077 -1.47644878

**25**

-0.86768771

**Table 3.9**: The following is a table of the VIF for the partial model without x2 relating steam (lbs/month) to all of the remaining regressor variables.

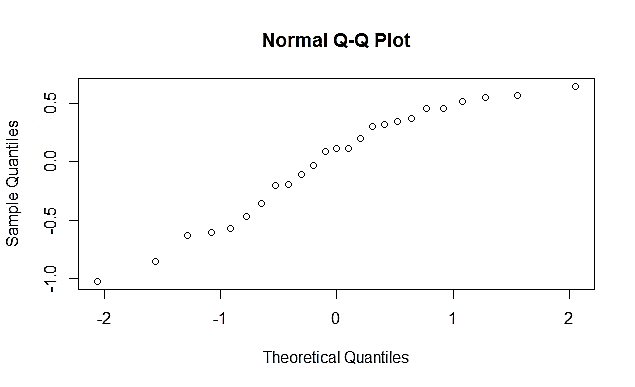
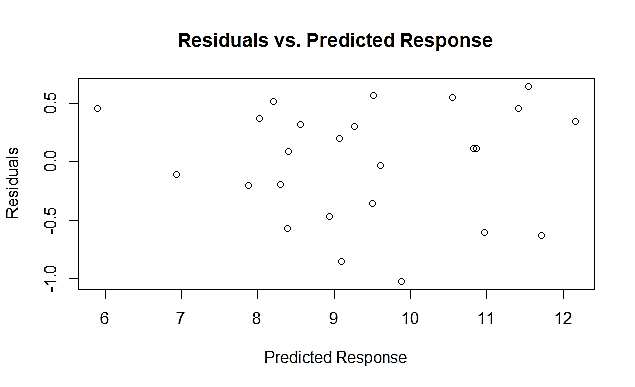
**x1 x3 x4 x5 x6 x7**

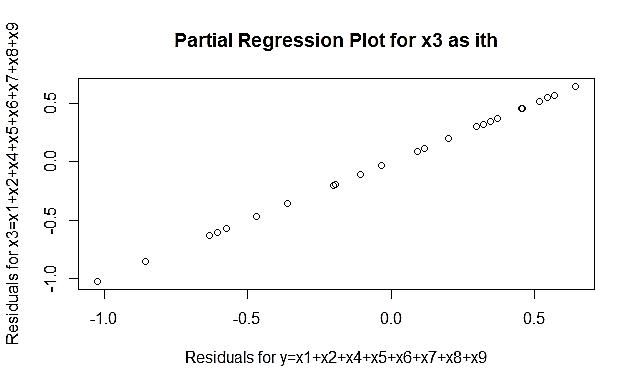
4.050933 103.260596 1.650794 2.704032 4.665612 5.999323

**x8 x9**

90.287806 2.384648

**Figure 3.4:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except average wind velocity in miles per hour (x3), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x3 versus the residuals for the x3 explained by the other regressor variables.



**Table 3.10**: The following is a table of the standardized residuals for the partial model without x3 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.21978832 1.11306319 0.69462526 1.04461250 0.37127291 1.15551286

**7 8 9 10 11 12**

1.26741150 0.18096145 -1.10497637 -0.67945367 -2.17941076 1.53195823

**13 14 15 16 17 18**

0.94586367 0.77027276 0.24323331 -0.06193097 1.17340759 -0.39424347

**19 20 21 22 23 24**

-0.32270450 0.78449261 -0.38018695 -1.12298900 -1.87259191 -1.12532471

**25**

-1.29792337

**Table 3.11**: The following is a table of the studentized residuals for the partial model without x3 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.2131311 1.1220345 0.6829445 1.0478032 0.3610420 1.1686447

**7 8 9 10 11 12**

1.2938309 0.1753948 -1.1132064 -0.6675797 -2.5165479 1.6057460

**13 14 15 16 17 18**

0.9425598 0.7600386 0.2359463 -0.0599716 1.1884327 -0.3835923

**19 20 21 22 23 24**

-0.3134791 0.7746255 -0.3697885 -1.1328924 -2.0518635 -1.1354510

**25**

-1.3285962

**Table 3.12**: The following is a table of the VIF values for the partial model without x3 relating steam (lbs/month) to all of the remaining regressor variables.

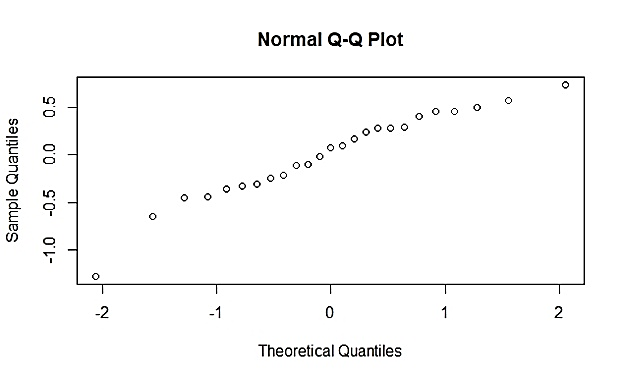
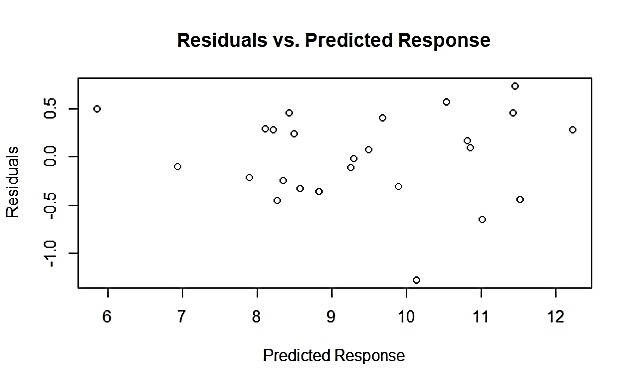
**x1 x2 x4 x5 x6 x7 x8**

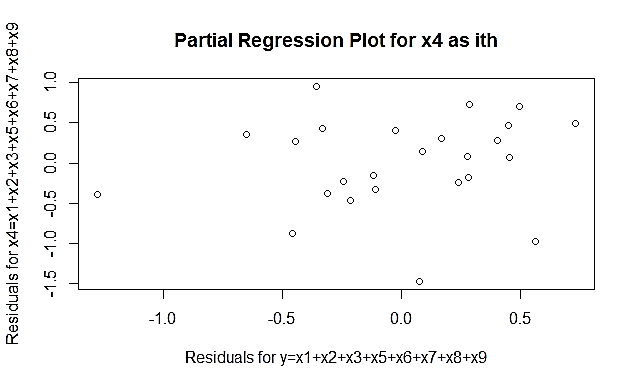
14.632845 16.421404 1.824200 3.856707 4.694497 5.056370 1.867641

**x9**

2.256168

**Figure 3.5:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except calendar days per month (x4), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x4 versus the residuals for the x4 explained by the other regressor variables.



**Table 3.13**: The following is a table of the standardized residuals for the partial model without x4 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.33290107 1.14265177 0.59932418 0.82460700 -0.04766458 0.60577340

**7 8 9 10 11 12**

1.36514236 0.60142111 -0.91375395 -0.24626395 -1.43178824 1.80107286

**13 14 15 16 17 18**

0.97282224 0.16106991 0.20167227 -0.66457396 0.91734269 -0.52126578

**19 20 21 22 23 24**

-0.33453219 1.20181466 -0.42317941 -0.81231332 -2.53402695 -1.27840162

**25**

-0.98515588

**Table 3.14**: The following is a table of the studentized residuals for the partial model without x4 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.32345220 1.15447460 0.58691850 0.81594885 -0.04615431 0.59338168

**7 8 9 10 11 12**

1.40622356 0.58901944 -0.90876778 -0.23889723 -1.48469617 1.95306785

**13 14 15 16 17 18**

0.97108823 0.15608186 0.19551699 -0.65254024 0.91253467 -0.50905441

**19 20 21 22 23 24**

-0.32504817 1.22002141 -0.41205416 -0.80325686 -3.17105230 -1.30632039

**25**

-0.98418962

**Table 3.15**: The following is a table of the VIF values for the partial model without x4 relating steam (lbs/month) to all of the remaining regressor variables.

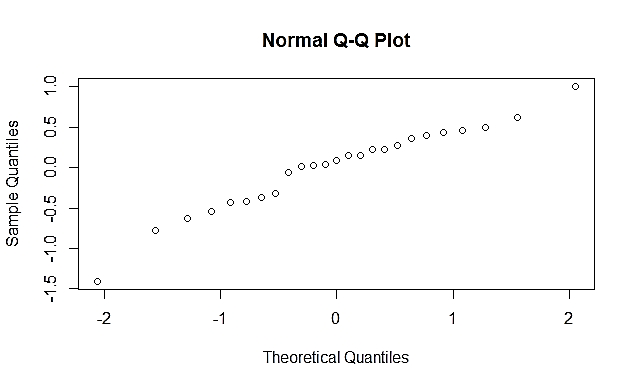
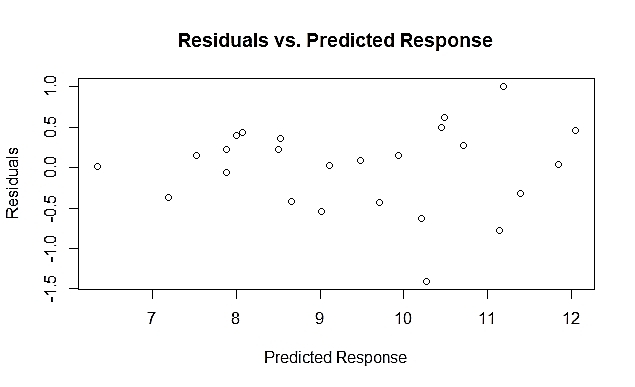
**x1 x2 x3 x5 x6 x7**

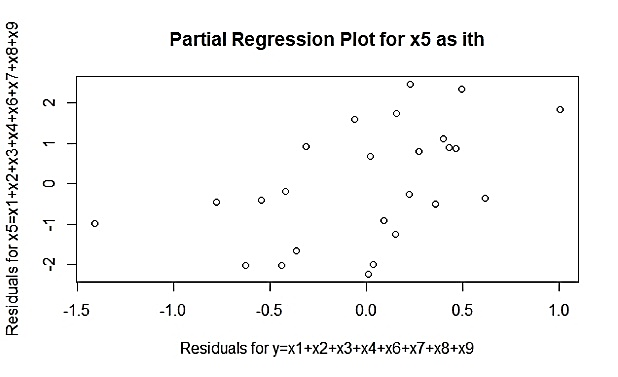
11.751771 18.099620 125.768897 4.392236 4.634851 5.862202

**x8 x9**

107.419983 2.080472

**Figure 3.6:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except operating days per month (x5), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x5 versus the residuals for the x5 explained by the other regressor variables.



**Table 3.16**: The following is a table of the standardized residuals for the partial model without x5 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.47361134 1.18993889 0.85685321 1.06250175 -0.76465028 0.49306366

**7 8 9 10 11 12**

0.02267385 0.80864384 -0.10868656 0.04194676 -1.70069400 2.09083363

**13 14 15 16 17 18**

0.06206378 0.21576074 0.89235610 -1.12637653 0.29228189 0.39875392

**19 20 21 22 23 24**

-0.94037856 0.81973694 0.26525590 -1.20661706 -2.44826380 -1.34668784

**25**

-0.60626942

**Table 3.17**: The following is a table of the studentized residuals for the partial model without x5 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.46182082 1.20678884 0.84936084 1.06709696 -0.75427958 0.48107568

**7 8 9 10 11 12**

0.02195422 0.79947330 -0.10527418 0.04061701 -1.81932195 2.37467866

**13 14 15 16 17 18**

0.06010023 0.20921402 0.88635800 -1.13660364 0.28375928 0.38802469

**19 20 21 22 23 24**

-0.93677305 0.81091801 0.25739951 -1.22538351 -2.99759874 -1.38476469

**25**

-0.59387897

**Table 3.18**: The following is a table of the VIF residuals for the partial model without x5 relating steam (lbs/month) to all of the remaining regressor variables.

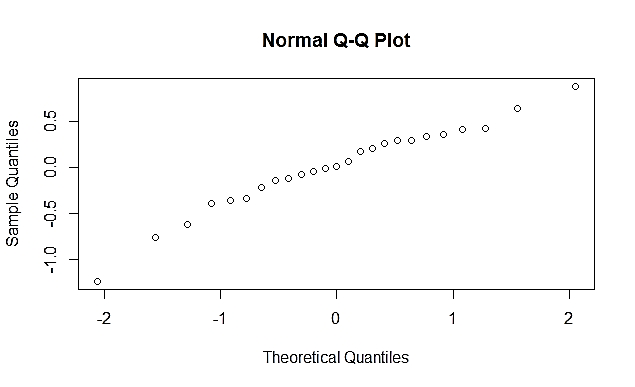
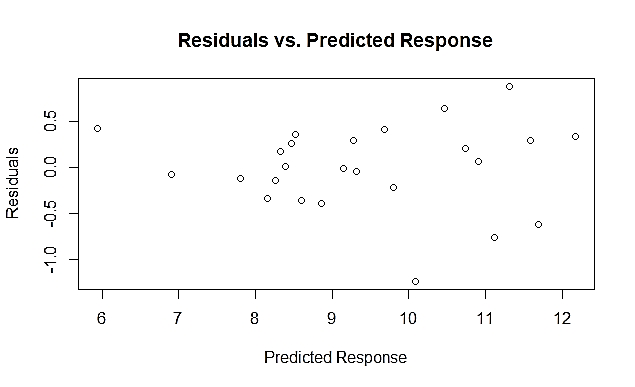
**x1 x2 x3 x4 x6 x7**

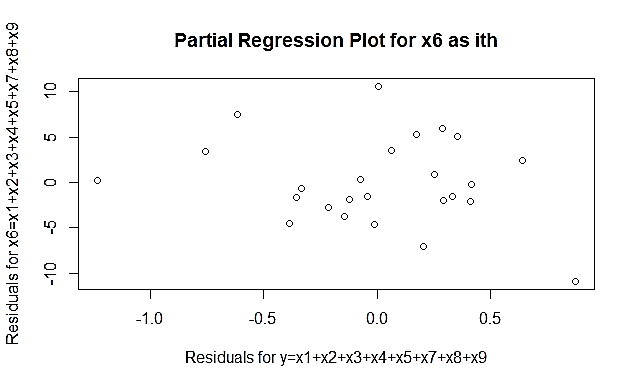
14.363452 12.341883 110.690568 1.828431 4.530146 5.840032

**x8 x9**

97.334598 2.354898

**Figure 3.7:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except days below 32F (x6), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x6 versus the residuals for the x6 explained by the other regressor variables.



**Table 3.19**: The following is a table of the standardized residuals for the partial model without x6 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.12579090 1.38574126 0.70218304 0.01215781 -0.08966443 0.63946578

**7 8 9 10 11 12**

1.22778901 0.35485753 -0.71384817 -0.02629473 -1.61570740 1.83234109

**13 14 15 16 17 18**

0.59561943 0.79049787 0.42062397 -0.45870909 0.92144628 -0.30746771

**19 20 21 22 23 24**

-0.22933923 0.88660864 -0.24655628 -0.94537103 -2.45962806 -1.47255796

**25**

-1.26552346

**Table 3.20**: The following is a table of the studentized residuals for the partial model without x6 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.12185679 1.43031190 0.69061010 0.01177180 -0.08683903 0.62722706

**7** 8 9 10 11 12

1.24909999 0.34494943 -0.70245721 -0.02546032 -1.71011900 1.99588273

**13 14 15 16 17 18**

0.58320791 0.78079527 0.40953799 -0.44709273 0.91684497 -0.29858774

**19 20 21 22 23 24**

-0.22242264 0.88035321 -0.23918189 -0.94203979 -3.01994138 -1.53349456

**25**

-1.29168898

**Table 3.21**: The following is a table of the VIF values for the partial model without x6 relating steam (lbs/month) to all of the remaining regressor variables.

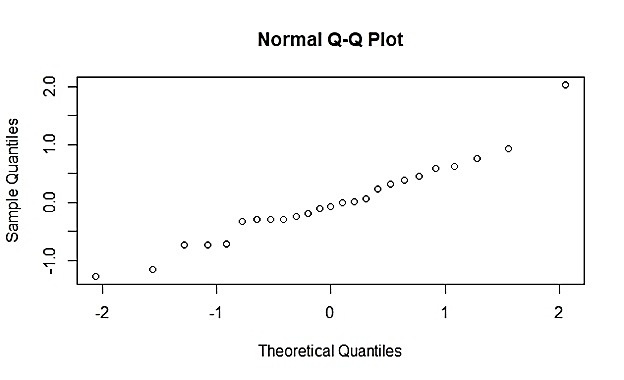
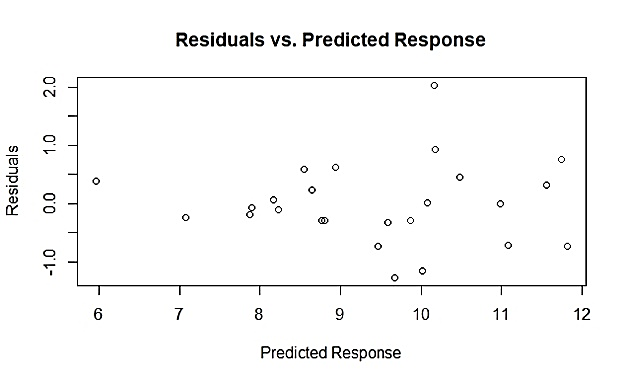
**x1 x2 x3 x4 x5 x7**

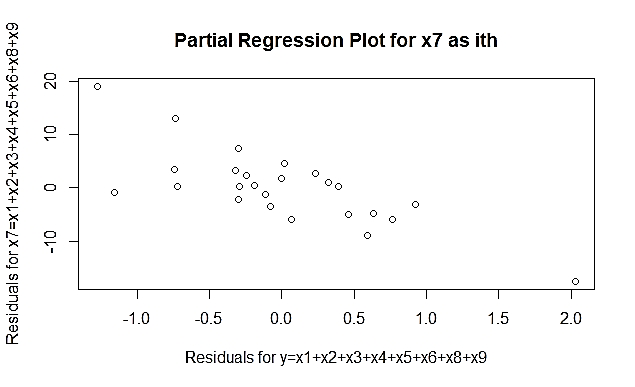
15.654279 20.011010 126.611716 1.813091 4.256994 2.707750

**x8 x9**

107.589948 2.384853

**Figure 3.8:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except average atmospheric temperature (x7), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x7 versus the residuals for the x7 explained by the other regressor variables.



**Table 3.22**: The following is a table of the standardized residuals for the partial model without x7 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5**

-0.004280291 1.308122679 1.027464977 -1.854430998 -0.435022623

**6 7 8 9 10**

-1.064714863 0.757797127 -0.404796745 -0.102563672 0.775142622

**11 12 13 14 15**

0.192660293 2.690707046 0.453713494 1.098732540 0.652335753

**16 17 18 19 20**

-0.402602580 0.024830698 -0.156057688 -0.495848853 0.398504864

**21 22 23 24 25**

-0.244553018 -0.486886978 -1.495983602 -0.917094485 -1.057002079

**Table 3.23**: The following is a table of the studentized residuals for the partial model without x7 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5**

-0.004144376 1.340281420 1.029377425 -2.026481852 -0.423722146

**6 7 8 9 10**

-1.069488942 0.747266530 -0.393965308 -0.099339509 0.765030619

**11 12 13 14 15**

0.186759282 3.520931882 0.442159823 1.106401147 0.640192143

**16 17 18 19 20**

-0.391807939 0.024042683 -0.151217336 -0.483835440 0.387779906

**21 22 23 24 25**

-0.237231229 -0.474957952 -1.561820663 -0.912274062 -1.061157694

**Table 3.24**: The following is a table of the VIF values for the partial model without x7 relating steam (lbs/month) to all of the remaining regressor variables.

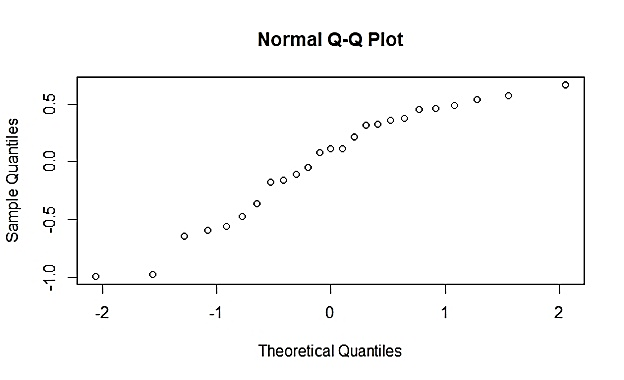
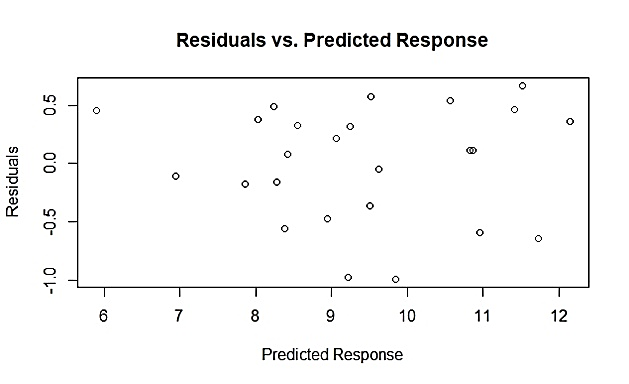
**x1 x2 x3 x4 x5 x6**

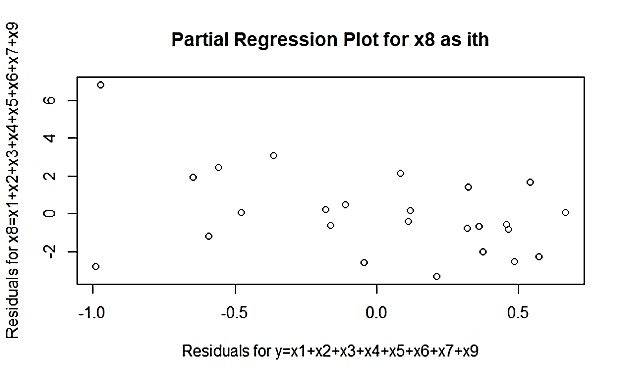
15.740218 19.911086 105.525135 1.774504 4.246571 2.095274

**x8 x9**

93.393373 2.377952

**Figure 3.9:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except average wind velocity squared (x8), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x8 versus the residuals for the x8 explained by the other regressor variables.



**Table 3.25**: The following is a table of the standardized residuals for the partial model without x8 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.21729176 1.08816727 0.71558471 1.04173314 0.39239001 1.09680110

**7 8 9 10 11 12**

1.26167965 0.16465417 -1.07432000 -0.68097205 -2.22223837 1.56615747

**13 14 15 16 17 18**

0.94789723 0.81749603 0.22764192 -0.08670867 1.16616172 -0.32307847

**19 20 21 22 23 24**

-0.32605162 0.77911432 -0.33327258 -1.12071787 -1.77808752 -1.08372175

**25**

-1.30830038

**Table 3.26**: The following is a table of the studentized residuals for the partial model without x8 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.21070296 1.09490739 0.70422246 1.04470461 0.38177135 1.10429795

**7 8 9 10 11 12**

1.28733150 0.15956096 -1.07988362 -0.66911603 -2.58777304 1.64799950

**13 14 15 16 17 18**

0.94470665 0.80860444 0.22077115 -0.08397504 1.18041000 -0.31384478

**19 20 21 22 23 24**

-0.31675218 0.76910468 -0.32381570 -1.13040583 -1.92195507 -1.09007920

**25**

-1.34048530

**Table 3.27**: The following is a table of the VIF values for the partial model without x8 relating steam (lbs/month) to all of the remaining regressor variables.

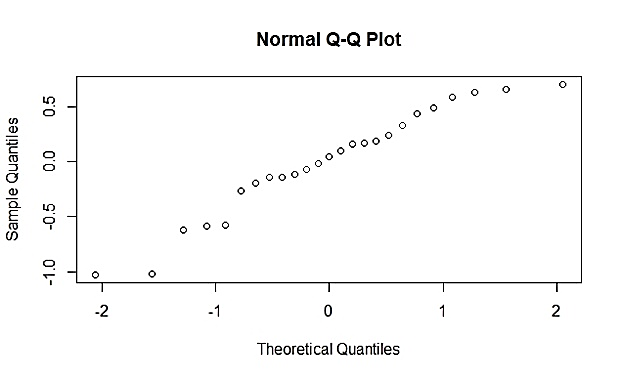
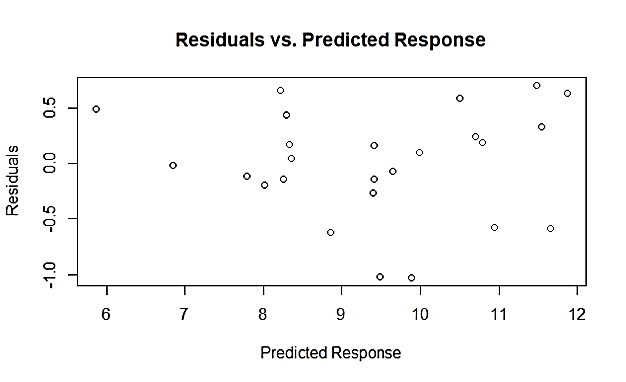
**x1 x2 x3 x4 x5 x6 x7**

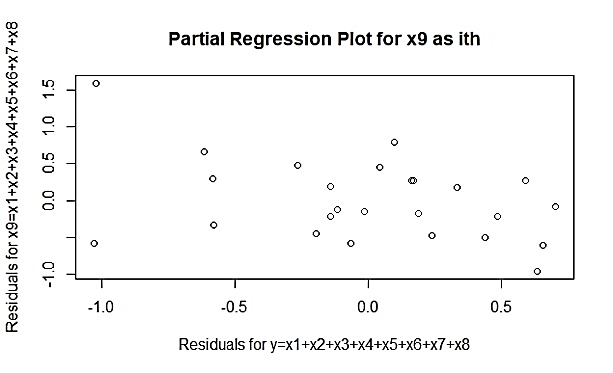
14.780030 16.898604 2.198060 1.833708 3.991346 4.694972 5.266779

**x9**

2.224572

**Figure 3.10:** The following is a normal probability plot relating steam (lbs/month) to the regressor variables in the full model except number of starups (x9), a plot of the residuals from this model versus the y-values the constructed model predicts, and a partial regression plot of residuals for steam (lbs/month) explained by the all regressors except x9 versus the residuals for the x9 explained by the other regressor variables.



**Table 3.28**: The following is a table of the standardized residuals for the partial model without x9 relating steam (lbs/month) to all of the remaining regressor variables.

**1 2 3 4 5 6**

0.35337483 1.20330596 1.16282131 0.12237207 -0.27203442 1.01269469

**7 8 9 10 11 12**

1.33820146 0.34627126 -0.37873464 -0.50954470 -2.24821617 1.65498827

**13 14 15 16 17 18**

0.67871459 0.41607185 0.48517966 -0.12808026 0.19531462 -0.28329663

**19 20 21 22 23 24**

-0.04110146 1.53347520 -0.21573584 -1.85710423 -1.86770605 -1.05948298

**25**

-1.19638859

**Table 3.29**: The following is a table of the studentized residuals for the partial model without x9 relating steam (lbs/month) to all of the remaining regressor variables.

1 2 3 4 5 6

0.34349676 1.22168583 1.17671617 0.11854173 -0.26400744 1.01355829

7 8 9 10 11 12

1.37493465 0.33653909 -0.36836314 -0.49741690 -2.63187596 1.76016016

13 14 15 16 17 18

0.66683201 0.40505710 0.47326755 -0.12407680 0.18933841 -0.27499134

19 20 21 22 23 24

-0.03979841 1.60760985 -0.20918981 -2.03020507 -2.04501474 -1.06383604

25

-1.21397061

**Table 3.30**: The following is a table of the VIF values for the partial model without x9 relating steam (lbs/month) to all of the remaining regressor variables.

**x1 x2 x3 x4 x5 x6**

14.742133 20.133751 119.783271 1.602086 4.356152 4.694632

**x7 x8**

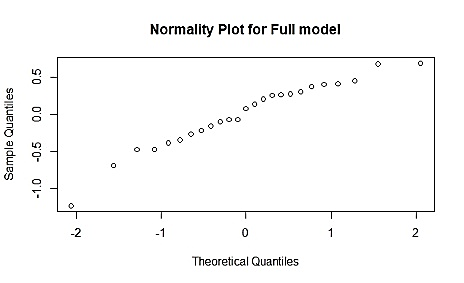
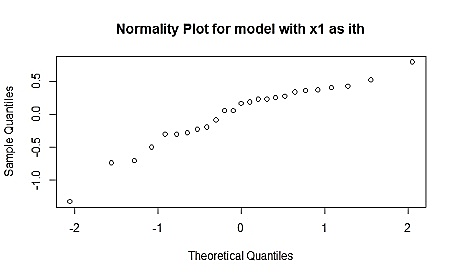
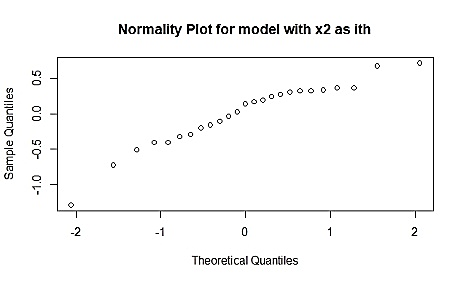
6.049379 100.351820

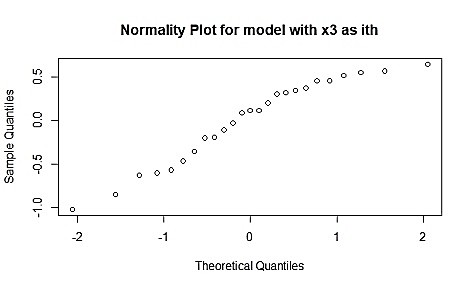
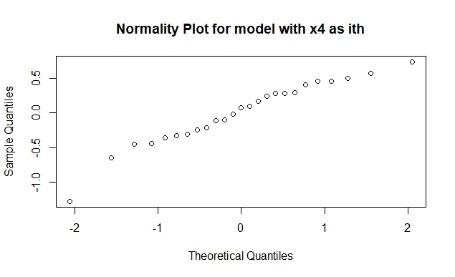
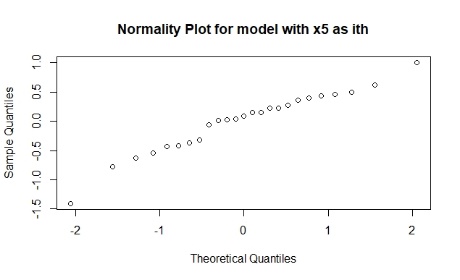
**Table 3.31**: The following is a table of the PRESS statistics for all of the models represented above.

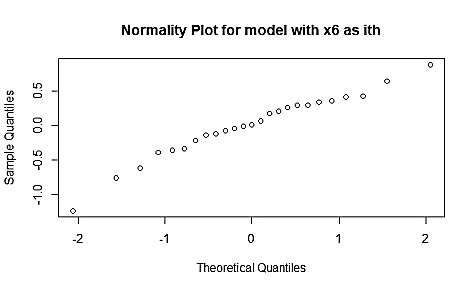
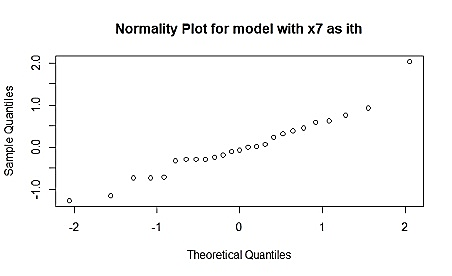
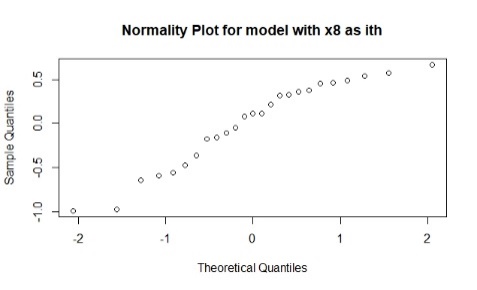
|  |  |
| --- | --- |
| **Model** | **PRESS Statistic** |
| Full model | 18.785 |
| x1 ith | 18.605 |
| x2 ith | 17.347 |
| x3 ith | 16.504 |
| x4 ith | 14.913 |
| x5 ith | 22.269 |
| x6 ith | 15.784 |
| x7 ith | 25.748 |
| x8 ith | 16.202 |
| x9 ith | 20.962 |

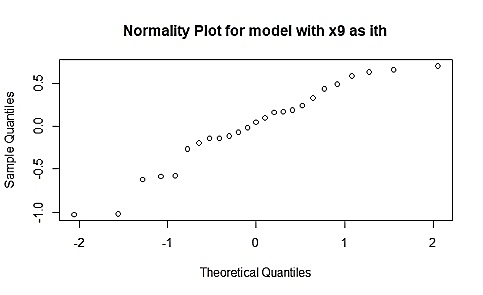
1. For analyzing model adequacy, figures have been provided with side-by-side comparison plots.

**Figure 4.1:** The following is a side-by-side comparison of all of the normality plots created in Question 3:

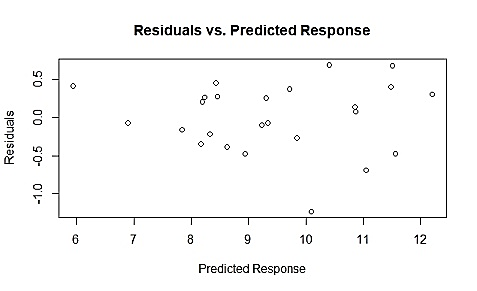
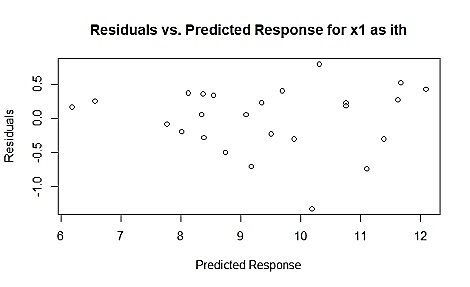
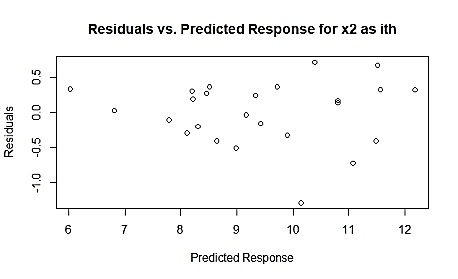
  

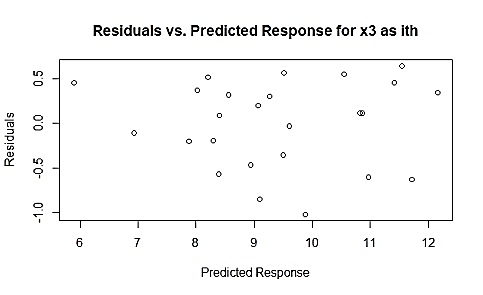
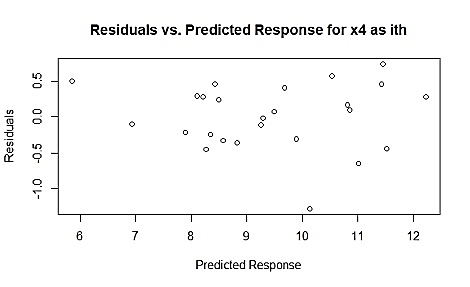
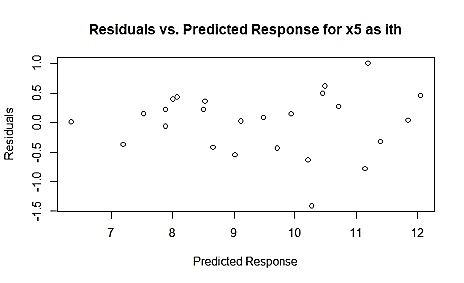
  

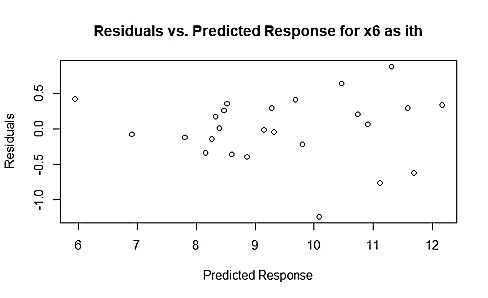
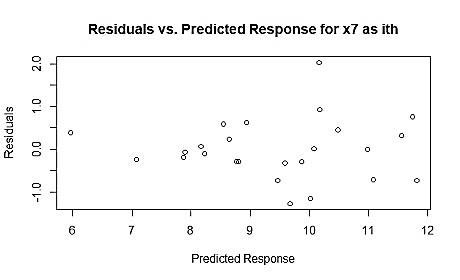
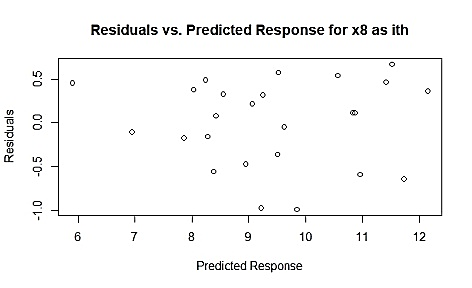
  

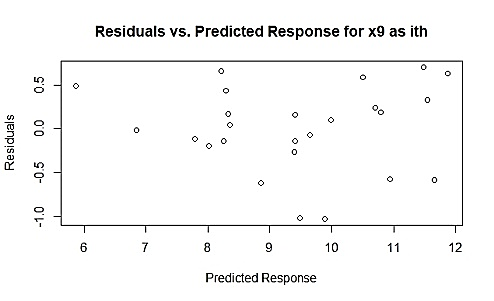


**Figure 4.2:** The following is a side-by-side comparison of all of the residual plots created in Question 3:

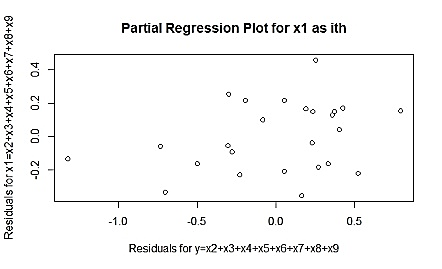
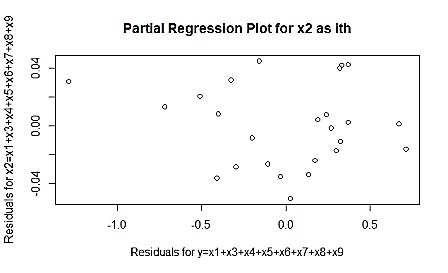
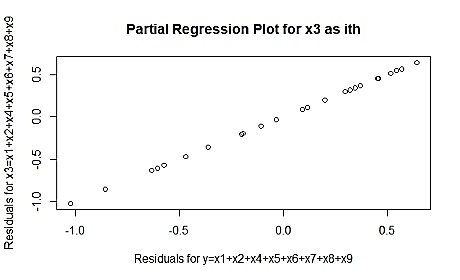
  

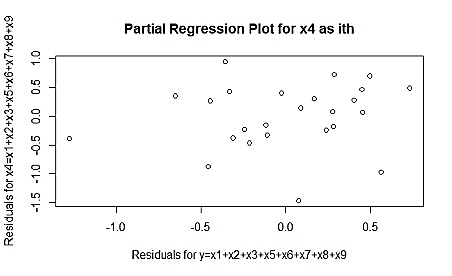
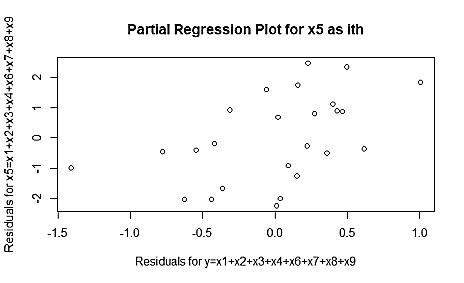
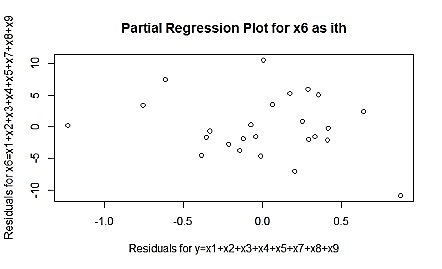
  

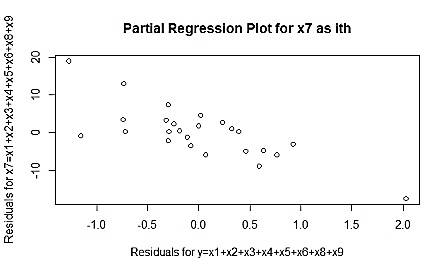
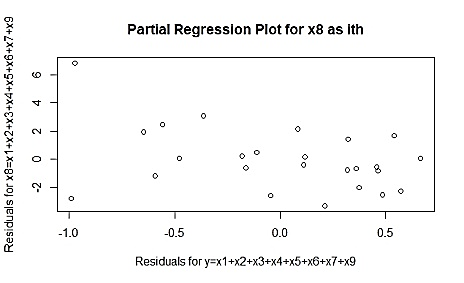
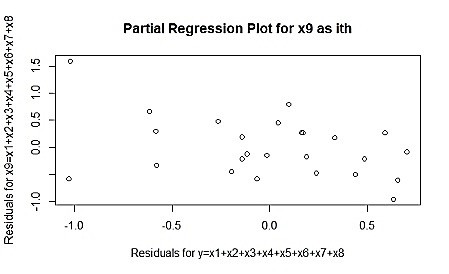
  



**Figure 4.3:** The following is a side-by-side comparison of all of the partial plots created in Question 3:

The normal probability plot of residuals, when the model follows the rules of normal distribution, independent distribution and equal variance, should have all the residuals fall on a straight line. When comparing the normality plots for the full model to the plots with each regressor removed individually, it appears that the full model normality plot (upper left corner) deviates more from a straight line when average wind velocity in miles per hour (x3), average wind velocity squared (x8) and number of startups (x9) are removed (Figure 4.1). This suggests that the model has a more-normal distribution when these regressors are present if they are taken out individually.

If the data set is normally distributed, independently distributed and has a constant variance, there will be no pattern in the data points in the constructed residuals vs. predicted response plot. The original model has a fairly randomly-distributed residual plot, suggesting little if any transformation to the model must be done (Figure 4.2, upper left corner). Removing regressors for partial regressions also does not really alter the residual plot, suggesting removal of regressors will not yield a model with a more-constant variance.

The standardized and studentized residuals are scaled residuals that can indicate if the data set is normally distributed and has a constant variance. If the absolute value of any of these residuals is greater than 3, it could indicate an outlier in the data set. An absolute value of a studentized residual or a R-student residual that is greater than 3 indicates the associated data point’s y-coordinate is likely an outlier. None of the standardized and studentized residuals from the full model have an absolute value greater than 3 (though the studentized residual for the 23rd y value is very close, at 2.994) (Table 3.2). It is likely that there are no glaring outliers created by this model. The partial plots do not drastically reduce the standardized and studentized residuals either, though the removal of certain regressors individually are accompanied by a spike in certain residuals over 3 when the absolute value is taken and suggest the regressor should remain in the model to control for outliers (Tables 3.5, 3.8, 3.14, 3.20). However, these spikes are seen in the studentized residuals, which are much more conservative than the standardized residuals and are therefore more likely to consider a residual an outlier than the corresponding standardized residuals.

When considering PRESS residuals, the model that has the lowest PRESS residual when one regressor is removed is when the average wind velocity in miles per hour (x3) regressor is removed. However, the partial plot for removing this regressor shows that it is the regressor that, when removed individually, explains a great amount of the variance of the model after the other regressor variables are accounted for (Figures 3.4 and 4.3). Therefore, it should remain in the model. Other regressors, when removed individually, increase the PRESS statistic compared to the original model, such as calendar days per month (x4) and average atmospheric temperature in degrees F (x7) (Table 3.31). The higher PRESS statistic means that, when the individual regressor is removed, influential points in the data set affect the model more than when the regressor was present, and suggests that the model would be benefited by the regressor in question remaining in the model.

Finally, removal of regressor variables alters the VIF values differently depending on the regressor variable that is removed. All regressors have moderate to low collinearity in the full model except pounds of real fatty acid in storage per month (x1), pounds of crude glycerin made per month (x2), average wind velocity (x4) and average wind velocity squared (x8) (Table 3.3). These collinearites make sense when considering the regressors; if more real fatty acid is stored per month, then more glycerin can be made because glycerin is comprised primarily of fatty acids. When x1 is removed from the model, x2’s VIF value drops from 20.137 to 5.180 (Tables 3.3 and 3.6). Similarly, when x2 is removed from the model, x1’s VIF value drops from 15.747 to 4.051 (Tables 3.3 and 3.9). The collinearity between average wind velocity and average wind velocity squared is obvious; average wind velocity squared is simply average wind velocity multiplied by itself. When x8 is removed, x3’s VIF value drops from 126.626 to 2.198 (Tables 3.3 and 3.27). Similarly, when x3 is removed, x8’s VIF value drops from 107.591 to 1.868 (Tables 3.3 and 3.12). However, multicollinearity is not a good enough reason to remove a regressor from a model.

Based on these analyses, there is no definitive indication that any particular regressor should be removed from the model for the model to perform better. It could be argued that certain regressors must stay in the model for it to perform well, such as average wind velocity (x3) based on its partial regression plot, but there is no evidence that any regressor should absolutely be removed from a model (Figures 3.4 and 4.3).

Based on the original scatter plots, the regressors that show a pattern with steam (lbs/month) that is nonlinear are average wind velocity (x3) and average wind velocity squared (x8). Both plots show a slight parabolic curve, suggesting that squaring these regressors may be warranted. There is a slight parabolic curve to the pounds of crude glycerin made per month (x2) as well, so this may also warrant squaring as well. Finally, because of the large number of data points at zero days for the number of days below 32F (x6), there is a slight square-root-like curve to the scatterplot and could perhaps benefit from square-rooting that regressor. The transformed model that is a better fit for the data set is as follows: (**note: coefficients that correspond to regressors that are measured in the same units as the y variable (lbs/month) remain unitless in order to correctly relate the regressor to the y variable**):

**Steam (lbs/month) = 0.022(x1) +1.724[lbs-1\*month](x2^2) +0.070[lbs\*hr2\*month-1\*miles-2](x3^2) + 0.206[lbs\*days-1](x4) +0.147[lbs\*days-1](x5) -0.246[lbs\*days-0.5\*month-1](sqrt[x6]) -0.091[lbs\*ºF-1\*month-1](x7) -0.00063[lbs\*hr2\*miles-2\*month-1](x8^2) -0.291[lbs\*number of startups-1\*month-1](x9) + 4.232 (lbs/month)**

The summary statistics for the new model are as follows:

Residuals:

Min 1Q Median 3Q Max

-1.07899 -0.20368 0.05293 0.26547 0.60527

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.2319065 5.5446609 0.763 0.4572

x1 0.0221917 0.4810566 0.046 0.9638

sqx2 1.7240965 2.2457111 0.768 0.4546

sqx3 0.0698773 0.0269608 2.592 0.0204 \*

x4 0.2060570 0.1731591 1.190 0.2525

x5 0.1473051 0.0631305 2.333 0.0340 \*

sqrtx6 -0.2455904 0.1387612 -1.770 0.0971 .

x7 -0.0911187 0.0170326 -5.350 8.1e-05 \*\*\*

sqx8 -0.0006255 0.0002186 -2.861 0.0119 \*

x9 -0.2908944 0.1795418 -1.620 0.1260

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4971 on 15 degrees of freedom

Multiple R-squared: 0.9418, Adjusted R-squared: 0.9069

F-statistic: 26.98 on 9 and 15 DF, p-value: 1.106e-07

Compared to the statistics for the original model, the adjusted R2 value for the transformed model has increased, and more regressor coefficients have become significant for the model (x3 and x8 have become significant to an alpha level of 0.05, and x6 has become significant to an alpha value of 0.1).

1. Based on the results from the stepwise regression in R, the model of best-fit is as follows: (**note: coefficients that correspond to regressors that are measured in the same units as the y variable (lbs/month) remain unitless in order to correctly relate the regressor to the y variable**):

**Steam (lbs/month) = 0.488(x1) +0.108[lbs\*days-1](x5) -0.076[lbs\*ºF-1\*month-1](x7) + 8.556 (lbs/month)**

The output and summary statistics for this model are as follows:

Stepwise Selection Summary

------------------------------------------------------------------------------------

Added/ Adj.

Step Variable Removed R-Square R-Square C(p) AIC RMSE

------------------------------------------------------------------------------------

1 x7 addition 0.714 0.701 52.7850 69.0538 0.8903

2 x1 addition 0.860 0.847 17.1190 53.1954 0.6369

3 x5 addition 0.880 0.862 14.0670 51.4289 0.6046

Call:

lm(formula = y ~ x1 + x5 + x7)

Residuals:

Min 1Q Median 3Q Max

-1.33205 -0.30490 0.05466 0.35996 1.16505

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.55609 1.03675 8.253 4.98e-08 \*\*\*

x1 0.48842 0.21162 2.308 0.0313 \*

x5 0.10827 0.05859 1.848 0.0788 .

x7 -0.07572 0.00746 -10.150 1.49e-09 \*\*\*

---

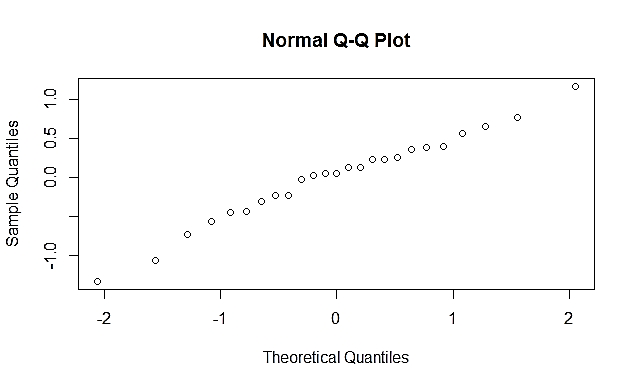
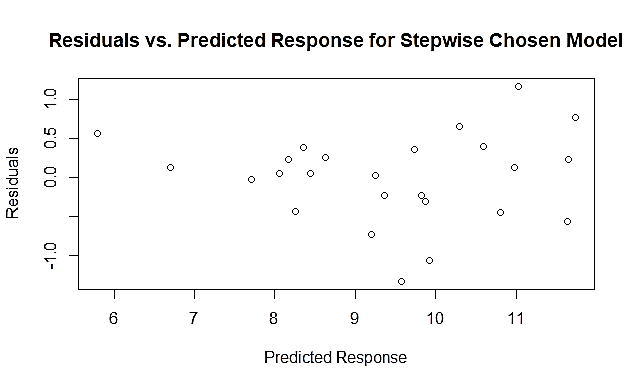
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6046 on 21 degrees of freedom

Multiple R-squared: 0.8795, Adjusted R-squared: 0.8623

F-statistic: 51.1 on 3 and 21 DF, p-value: 7.988e-10

**Figure 5.1:** The following is a normal probability plot relating steam (lbs/month) to the pounds of real fatty acid per month (x1), operating days per month (x5) and average atmospheric temperature (F) in the stepwise model and a plot of the residuals from the full model versus the y-values the constructed model predicts.

Though the adjusted R2 value is lower than the original model and the transformed model, it is still quite high and only uses three of the nine regressors of the model, which suggests it is a much more efficient model and uses the regressors that best explain the y variable. The normality plot is more linear than the original normality plot, and the residual plot is still fairly randomly distributed, suggesting a normally-distributed data set with independence and constant variance (Figure 5.1). To summarize, the computer is much better at predicting regressors to use in a model than I am.

**Rcode with output:**

library(car)

library(qpcR)

#Add Data

#Subset

> y=Data[,1]

> x1=Data[,2]

> x2=Data[,3]

> x3=Data[,4]

> x4=Data[,5]

> x5=Data[,6]

> x6=Data[,7]

> x7=Data[,8]

> x8=Data[,9]

> x9=Data[,10]

>

> #Question 1: Box Plots

> boxplot(y, main = "Pounds of Steam used Monthly", xlab = "y", ylab = "Pounds")

> boxplot(x1, main = "Pounds of Real Fatty Acid in Storage per Month", xlab= "x1", ylab = "Pounds per Month")

> boxplot(x2, main = "Pounds of Crude Glycerin", xlab = "x2", ylab = "Pounds")

> boxplot(x3, main = "Average Wind Velocity", xlab = "x3", ylab = "Avergae Wind Velocity (mph)")

> boxplot(x4, main = "Calender Days per Month", xlab = "x4", ylab = "Days per Month")

> boxplot(x5, main = "Operating Days per Month", xlab = "x5", ylab= "Days per Month")

> boxplot(x6, main = "Days below 32F", xlab = "x6", ylab = "Days")

> boxplot(x7, main = "Avg Atmospheric Temperature", xlab = "x7", ylab = "Avg Atmospheric Temp (F)")

> boxplot(x8, main = "Avg Wind Velocity squared", xlab = "x8", ylab = "Avg Wind Velocity Squared (mph2)")

> boxplot(x9, main = "Number of Starups", xlab = "x9", ylab = "Number")

> #Question 2: Scatter Plots

> plot(x1, y, main = "Pounds of Steam used Monthly vs Pounds Real Fatty Acid", xlab= "Real Fatty Acid in Storage (lbs/month)", ylab = "Steam (lbs/month)")

> plot(x2, y, main = "Pounds of Steam used Monthly vs Crude Glycerin", xlab= "Crude Glycerin (lbs)", ylab = "Steam (lbs/month)")

> plot(x3, y, main = "Pounds of Steam used Monthly vs Average Wind Velocity", xlab= "Average Wind Velocity (mph)", ylab = "Steam (lbs/month)")

> plot(x4, y, main = "Pounds of Steam used Monthly vs Calender Days", xlab= "Calender (days/month)", ylab = "Steam (lbs/month)")

> plot(x5, y, main = "Pounds of Steam used Monthly vs Operating Days per Month", xlab= "Operating Days (days/month)", ylab = "Steam (lbs/month)")

> plot(x6, y, main = "Pounds of Steam used Monthly vs Days below 32F", xlab= "Days below 32F (days)", ylab = "Steam (lbs/month)")

> plot(x7, y, main = "Pounds of Steam used Monthly vs Average Atmospheric Temp", xlab= "Average Atmospheric Temp (F)", ylab = "Steam (lbs/month)")

> plot(x8, y, main = "Pounds of Steam used Monthly vs Average Wind Velocity Squared", xlab= "Average Wind Velocity Squared (mph2)", ylab = "Steam (lbs/month)")

> plot(x9, y, main = "Pounds of Steam used Monthly vs Number of Startups", xlab= "Number of Startups (#)", ylab = "Steam (lbs/month)")

#Question 3: Linear Model

> lm.full=lm(y~x1+x2+x3+x4+x5+x6+x7+x8+x9)

> lm.full

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

1.76116 0.70084 -1.86794 1.14038 0.12253

x5 x6 x7 x8 x9

0.17957 -0.01831 -0.07734 -0.08626 -0.34610

> summary(lm.full)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.22921 -0.26565 0.07307 0.30513 0.68420

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.76116 6.96637 0.253 0.803847

x1 0.70084 0.56248 1.246 0.231880

x2 -1.86794 4.12852 -0.452 0.657421

x3 1.14038 0.74289 1.535 0.145591

x4 0.12253 0.20374 0.601 0.556546

x5 0.17957 0.08060 2.228 0.041619 \*

x6 -0.01831 0.02440 -0.751 0.464557

x7 -0.07734 0.01652 -4.681 0.000295 \*\*\*

x8 -0.08626 0.05178 -1.666 0.116445

x9 -0.34610 0.20979 -1.650 0.119777

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5673 on 15 degrees of freedom

Multiple R-squared: 0.9242, Adjusted R-squared: 0.8788

F-statistic: 20.33 on 9 and 15 DF, p-value: 7.576e-07

>

> residfull=resid(lm.full)

> residfull

1 2 3 4 5 6

0.12971046 0.68420417 0.30512860 0.19897269 -0.07212726 0.26888236

7 8 9 10 11 12

0.41150181 0.26680258 -0.34906745 -0.09772008 -0.38567331 0.67409485

13 14 15 16 17 18

0.39507038 0.25510806 0.07307058 -0.26564789 0.37327741 -0.21481118

19 20 21 22 23 24

-0.06727046 0.44771277 -0.15717316 -0.47223264 -1.22921459 -0.69490449

25

-0.47769422

>

> predictlmfull=predict(lm.full)

> predictlmfull

1 2 3 4 5 6 7

10.850290 10.415796 12.204871 8.201027 9.342127 8.461118 5.948498

8 9 10 11 12 13 14

8.233197 8.169067 9.237720 8.625673 11.515905 11.484930 9.314892

15 16 17 18 19 20 21

10.866929 9.845648 9.716723 8.324811 6.897270 8.432287 7.837173

22 23 24 25

8.942233 10.089215 11.054904 11.557694

> plot(predictlmfull,residfull, main= "Residuals vs. Predicted Response", xlab ="Predicted Response", ylab="Residuals")

> qqnorm(residfull, main = "Normality Plot for Full model")

>

> standfull=rstandard(lm.full)

> standfull

1 2 3 4 5 6

0.2542561 1.4748225 0.6363896 0.6141057 -0.1488749 0.6723499

7 8 9 10 11 12

1.1986003 0.5678766 -0.7316112 -0.2035156 -1.7552012 1.6696907

13 14 15 16 17 18

0.8503303 0.6887397 0.1596926 -0.5628649 0.8309716 -0.4543152

19 20 21 22 23 24

-0.2101913 1.1567576 -0.3102859 -1.1720115 -2.4195457 -1.3487168

25

-1.0444930

> studentfull=rstudent(lm.full)

> studentfull

1 2 3 4 5 6

0.2461658 1.5409084 0.6232825 0.6008842 -0.1439332 0.6595664

7 8 9 10 11 12

1.2177407 0.5546151 -0.7197621 -0.1968867 -1.9022436 1.7877397

13 14 15 16 17 18

0.8420427 0.6761632 0.1544091 -0.5496144 0.8219363 -0.4419615

19 20 21 22 23 24

-0.2033638 1.1709832 -0.3007313 -1.1879701 -2.9935535 -1.3899885

25

-1.0479030

> vif(lm.full)

x1 x2 x3 x4 x5 x6

15.746595 20.137114 126.625618 1.836626 4.411920 4.695013

x7 x8 x9

6.067426 107.590891 2.385046

> PRESS(lm.full)

.........10.........20.....

$stat

[1] 18.78545

$residuals

[1] 0.16039785 1.02311518 0.42716331 0.60999105 -0.09889503

[6] 0.54107699 1.12358967 0.38899908 -0.49349421 -0.13640873

[11] -2.57078019 1.33101141 0.58902199 0.59843466 0.11232023

[16] -0.38382449 0.59534866 -0.30923483 -0.21136629 0.96186910

[21] -0.19714075 -0.93613495 -1.53274768 -0.84245581 -0.73500760

$P.square

[1] 0.7051611

>

> #Partial x1 ith

> lm.x1=lm(y~x2+x3+x4+x5+x6+x7+x8+x9)

> lm.x1

Call:

lm(formula = y ~ x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x2 x3 x4 x5

-1.91505 2.56535 1.38655 0.25039 0.14980

x6 x7 x8 x9

-0.02064 -0.07775 -0.10224 -0.28008

> lm.x1ith=lm(x1~x2+x3+x4+x5+x6+x7+x8+x9)

> lm.x1ith

Call:

lm(formula = x1 ~ x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x2 x3 x4 x5

-5.2454368 6.3256737 0.3512527 0.1824399 -0.0424695

x6 x7 x8 x9

-0.0033219 -0.0005911 -0.0228058 0.0942014

> resid.x1=resid(lm.x1)

> resid.x1

1 2 3 4 5 6

0.23354461 0.79268140 0.42492442 0.05365364 -0.23156537 0.36067629

7 8 9 10 11 12

0.16533581 0.37234436 -0.19640345 0.05434779 -0.49858432 0.52116620

13 14 15 16 17 18

0.26828176 0.22843196 0.19011488 -0.30397607 0.40333156 -0.27849061

19 20 21 22 23 24

0.25294612 0.33364620 -0.08559584 -0.70416688 -1.32193831 -0.73547262

25

-0.29923352

> resid.x1ith=resid(lm.x1ith)

> resid.x1ith

1 2 3 4 5 6

0.14815643 0.15478143 0.17093144 -0.20734944 -0.22749529 0.13097676

7 8 9 10 11 12

-0.35124356 0.15059297 0.21782961 0.21697903 -0.16110780 -0.21820724

13 14 15 16 17 18

-0.18090915 -0.03806297 0.16700542 -0.05468881 0.04288295 -0.09086140

19 20 21 22 23 24

0.45690310 -0.16275662 0.10213056 -0.33093687 -0.13230344 -0.05788491

25

0.25463780

> plot(resid.x1,resid.x1ith,main="Partial Regression Plot for x1 as ith", xlab= "Residuals for y=x2+x3+x4+x5+x6+x7+x8+x9", ylab= "Residuals for x1=x2+x3+x4+x5+x6+x7+x8+x9")

> qqnorm(resid.x1, main = "Normality Plot for model with x1 as ith")

>

> summary(lm.x1)

Call:

lm(formula = y ~ x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.3219 -0.2785 0.1653 0.3337 0.7927

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.91505 6.41875 -0.298 0.769273

x2 2.56535 2.12985 1.204 0.245922

x3 1.38655 0.72840 1.904 0.075113 .

x4 0.25039 0.17902 1.399 0.180993

x5 0.14980 0.07830 1.913 0.073782 .

x6 -0.02064 0.02475 -0.834 0.416490

x7 -0.07775 0.01680 -4.628 0.000279 \*\*\*

x8 -0.10224 0.05102 -2.004 0.062304 .

x9 -0.28008 0.20647 -1.357 0.193759

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.577 on 16 degrees of freedom

Multiple R-squared: 0.9164, Adjusted R-squared: 0.8746

F-statistic: 21.92 on 8 and 16 DF, p-value: 3.126e-07

> predictlm.x1=predict(lm.x1)

> predictlm.x1

1 2 3 4 5 6 7

10.746455 10.307319 12.085076 8.346346 9.501565 8.369324 6.194664

8 9 10 11 12 13 14

8.127656 8.016403 9.085652 8.738584 11.668834 11.611718 9.341568

15 16 17 18 19 20 21

10.749885 9.883976 9.686668 8.388491 6.577054 8.546354 7.765596

22 23 24 25

9.174167 10.181938 11.095473 11.379234

> plot(predictlm.x1,resid.x1, main= "Residuals vs. Predicted Response for x1 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx1=rstandard(lm.x1)

> standx1

1 2 3 4 5 6

0.4441977 1.6510694 0.8543183 0.1531863 -0.4543402 0.8720314

7 8 9 10 11 12

0.4103772 0.7668158 -0.3919954 0.1078524 -2.0623675 1.2142951

13 14 15 16 17 18

0.5545728 0.6053296 0.4001498 -0.6318953 0.8814951 -0.5757276

19 20 21 22 23 24

0.6058804 0.8247769 -0.1650782 -1.5598131 -2.5312539 -1.4006357

25

-0.6138719

> studentx1=rstudent(lm.x1)

> studentx1

1 2 3 4 5 6

0.4327693 1.7551351 0.8467278 0.1484309 -0.4427785 0.8651502

7 8 9 10 11 12

0.3994538 0.7564971 -0.3813837 0.1044656 -2.3305308 1.2339697

13 14 15 16 17 18

0.5421991 0.5929368 0.3893967 -0.6196102 0.8750158 -0.5633113

19 20 21 22 23 24

0.5934890 0.8161244 -0.1599726 -1.6401235 -3.1652623 -1.4478207

25

-0.6015045

> vif(lm.x1)

x2 x3 x4 x5 x6 x7

5.180428 117.669441 1.370684 4.024388 4.667488 6.064969

x8 x9

100.986695 2.232906

> PRESS(lm.x1)

.........10.........20.....

$stat

[1] 18.60529

$residuals

[1] 0.28129156 1.14500099 0.57187560 0.14561780 -0.29679908

[6] 0.70197309 0.33913559 0.52578859 -0.26048733 0.07126088

[11] -2.84032033 0.94198950 0.38168032 0.53407329 0.28041606

[16] -0.43734548 0.64143325 -0.39627567 0.48319193 0.67882932

[21] -0.10599872 -1.15038550 -1.61374196 -0.88809147 -0.41929439

$P.square

[1] 0.7079888

>

> #Partial x2 ith

> lm.x2=lm(y~x1+x3+x4+x5+x6+x7+x8+x9)

> lm.x2

Call:

lm(formula = y ~ x1 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x3 x4 x5

0.65648 0.48151 1.28476 0.15185 0.15688

x6 x7 x8 x9

-0.01744 -0.07655 -0.09566 -0.34487

> lm.x2ith=lm(x2~x1+x3+x4+x5+x6+x7+x8+x9)

> lm.x2ith

Call:

lm(formula = x2 ~ x1 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x3 x4 x5

0.5913909 0.1174171 -0.0772954 -0.0156973 0.0121469

x6 x7 x8 x9

-0.0004677 -0.0004240 0.0050293 -0.0006567

> resid.x2=resid(lm.x2)

> resid.x2

1 2 3 4 5 6

0.17457402 0.71451258 0.32557851 0.19103415 -0.15583568 0.27180123

7 8 9 10 11 12

0.33342933 0.29923543 -0.29556999 -0.03186179 -0.40111895 0.67114083

13 14 15 16 17 18

0.32057940 0.24096869 0.13588405 -0.32450889 0.36871177 -0.19880378

19 20 21 22 23 24

0.02663444 0.36888969 -0.10798310 -0.51070158 -1.28678190 -0.71968254

25

-0.41012592

> resid.x2ith=resid(lm.x2ith)

> resid.x2ith

1 2 3 4 5

-0.024017670 -0.016225579 -0.010947847 0.004249893 0.044813234

6 7 8 9 10

-0.001562614 0.041796044 -0.017362898 -0.028639822 -0.035257186

11 12 13 14 15

0.008268808 0.001581436 0.039878686 0.007569498 -0.033627143

16 17 18 19 20

0.031511193 0.002444214 -0.008569550 -0.050271917 0.042197878

21 22 23 24 25

-0.026333860 0.020594313 0.030818615 0.013264913 -0.036172639

> plot(resid.x2,resid.x2ith,main="Partial Regression Plot for x2 as ith", xlab= "Residuals for y=x1+x3+x4+x5+x6+x7+x8+x9", ylab= "Residuals for x2=x1+x3+x4+x5+x6+x7+x8+x9")

> qqnorm(resid.x2, main = "Normality Plot for model with x2 as ith")

>

> summary(lm.x2)

Call:

lm(formula = y ~ x1 + x3 + x4 + x5 + x6 + x7 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.2868 -0.2956 0.1359 0.3206 0.7145

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.65648 6.36028 0.103 0.919074

x1 0.48151 0.27811 1.731 0.102620

x3 1.28476 0.65398 1.965 0.067079 .

x4 0.15185 0.18829 0.806 0.431801

x5 0.15688 0.06151 2.550 0.021388 \*

x6 -0.01744 0.02371 -0.735 0.472698

x7 -0.07655 0.01601 -4.780 0.000205 \*\*\*

x8 -0.09566 0.04624 -2.069 0.055118 .

x9 -0.34487 0.20450 -1.686 0.111103

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.553 on 16 degrees of freedom

Multiple R-squared: 0.9232, Adjusted R-squared: 0.8848

F-statistic: 24.04 on 8 and 16 DF, p-value: 1.616e-07

> predictlm.x2=predict(lm.x2)

> predictlm.x2

1 2 3 4 5 6 7

10.805426 10.385487 12.184421 8.208966 9.425836 8.458199 6.026571

8 9 10 11 12 13 14

8.200765 8.115570 9.171862 8.641119 11.518859 11.559421 9.329031

15 16 17 18 19 20 21

10.804116 9.904509 9.721288 8.308804 6.803366 8.511110 7.787983

22 23 24 25

8.980702 10.146782 11.079683 11.490126

> plot(predictlm.x2,resid.x2, main= "Residuals vs. Predicted Response for x2 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx2=rstandard(lm.x2)

> standx2

1 2 3 4 5 6

0.34458364 1.56370197 0.69349876 0.60394268 -0.30824780 0.69710608

7 8 9 10 11 12

0.89015998 0.64587856 -0.61682186 -0.06514238 -1.85042861 1.70507243

13 14 15 16 17 18

0.66716441 0.66500086 0.29151391 -0.67997466 0.84178824 -0.43011399

19 20 21 22 23 24

0.07162692 0.89155551 -0.21381134 -1.27219570 -2.52041442 -1.42484640

25

-0.87446700

> studentx2=rstudent(lm.x2)

> studentx2

1 2 3 4 5 6

0.33488660 1.64494915 0.68180255 0.59154650 -0.29934982 0.68545982

7 8 9 10 11 12

0.88406284 0.63368463 -0.60446533 -0.06308221 -2.02091565 1.82504421

13 14 15 16 17 18

0.65515642 0.65297132 0.28300970 -0.66810683 0.83372911 -0.41888477

19 20 21 22 23 24

0.06936359 0.88552121 -0.20731833 -1.29926334 -3.14275077 -1.47644878

25

-0.86768771

> vif(lm.x2)

x1 x3 x4 x5 x6 x7

4.050933 103.260596 1.650794 2.704032 4.665612 5.999323

x8 x9

90.287806 2.384648

> PRESS(lm.x2)

.........10.........20.....

$stat

[1] 17.34658

$residuals

[1] 0.20801686 1.04661496 0.45177739 0.58394141 -0.18647556

[6] 0.54680839 0.72681049 0.42636110 -0.39368424 -0.04073296

[11] -2.61072029 1.32483215 0.42463811 0.56127105 0.19126651

[16] -0.43576034 0.58777020 -0.28459768 0.05891118 0.65900561

[21] -0.12947757 -0.96923591 -1.50982920 -0.86274804 -0.57024107

$P.square

[1] 0.7277444

>

> #Partial x3 ith

> lm.x3=lm(y~x1+x2+x4+x5+x6+x7+x8+x9)

> lm.x3

Call:

lm(formula = y ~ x1 + x2 + x4 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x4 x5

6.356212 0.930473 -4.590272 0.096805 0.223461

x6 x7 x8 x9

-0.018707 -0.087690 -0.007475 -0.271240

> lm.x3ith=lm(x3~x1+x2+x4+x5+x6+x7+x8+x9)

> lm.x3ith

Call:

lm(formula = x3 ~ x1 + x2 + x4 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x4 x5

4.0293949 0.2013638 -2.3872116 -0.0225582 0.0384888

x6 x7 x8 x9

-0.0003442 -0.0090780 0.0690877 0.0656458

> resid.x3=resid(lm.x3)

> resid.x3

1 2 3 4 5 6

0.11679845 0.54762053 0.34745389 0.37361302 0.19944411 0.52000836

7 8 9 10 11 12

0.45471574 0.09113951 -0.57452882 -0.36078770 -0.85678713 0.64488688

13 14 15 16 17 18

0.45957264 0.29799958 0.11613558 -0.03198664 0.57114445 -0.19422652

19 20 21 22 23 24

-0.10793624 0.32311824 -0.20089488 -0.47127023 -1.02443718 -0.60755376

25

-0.63324186

> resid.x3ith=resid(lm.x3ith)

> resid.x3

1 2 3 4 5 6

0.11679845 0.54762053 0.34745389 0.37361302 0.19944411 0.52000836

7 8 9 10 11 12

0.45471574 0.09113951 -0.57452882 -0.36078770 -0.85678713 0.64488688

13 14 15 16 17 18

0.45957264 0.29799958 0.11613558 -0.03198664 0.57114445 -0.19422652

19 20 21 22 23 24

-0.10793624 0.32311824 -0.20089488 -0.47127023 -1.02443718 -0.60755376

25

-0.63324186

> plot(resid.x3,resid.x3,main="Partial Regression Plot for x3 as ith", xlab= "Residuals for y=x1+x2+x4+x5+x6+x7+x8+x9", ylab= "Residuals for x3=x1+x2+x4+x5+x6+x7+x8+x9")

> qqnorm(resid.x3, main = "Normality Plot for model with x3 as ith")

>

> summary(lm.x3)

Call:

lm(formula = y ~ x1 + x2 + x4 + x5 + x6 + x7 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.0244 -0.3608 0.1161 0.3736 0.6449

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.356212 6.551668 0.970 0.3464

x1 0.930473 0.564739 1.648 0.1189

x2 -4.590272 3.883029 -1.182 0.2544

x4 0.096805 0.211479 0.458 0.6533

x5 0.223461 0.078489 2.847 0.0117 \*

x6 -0.018707 0.025415 -0.736 0.4723

x7 -0.087690 0.015708 -5.583 4.12e-05 \*\*\*

x8 -0.007475 0.007105 -1.052 0.3084

x9 -0.271240 0.212519 -1.276 0.2201

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5909 on 16 degrees of freedom

Multiple R-squared: 0.9123, Adjusted R-squared: 0.8685

F-statistic: 20.81 on 8 and 16 DF, p-value: 4.516e-07

> predictlm.x3=predict(lm.x3)

> predictlm.x3

1 2 3 4 5 6 7

10.863202 10.552379 12.162546 8.026387 9.070556 8.209992 5.905284

8 9 10 11 12 13 14

8.408860 8.394529 9.500788 9.096787 11.545113 11.420427 9.272000

15 16 17 18 19 20 21

10.823864 9.611987 9.518856 8.304227 6.937936 8.556882 7.880895

22 23 24 25

8.941270 9.884437 10.967554 11.713242

> plot(predictlm.x3,resid.x3, main= "Residuals vs. Predicted Response for x3 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx3=rstandard(lm.x3)

> standx3

1 2 3 4 5 6

0.21978832 1.11306319 0.69462526 1.04461250 0.37127291 1.15551286

7 8 9 10 11 12

1.26741150 0.18096145 -1.10497637 -0.67945367 -2.17941076 1.53195823

13 14 15 16 17 18

0.94586367 0.77027276 0.24323331 -0.06193097 1.17340759 -0.39424347

19 20 21 22 23 24

-0.32270450 0.78449261 -0.38018695 -1.12298900 -1.87259191 -1.12532471

25

-1.29792337

> studentx3=rstudent(lm.x3)

> studentx3

1 2 3 4 5 6

0.2131311 1.1220345 0.6829445 1.0478032 0.3610420 1.1686447

7 8 9 10 11 12

1.2938309 0.1753948 -1.1132064 -0.6675797 -2.5165479 1.6057460

13 14 15 16 17 18

0.9425598 0.7600386 0.2359463 -0.0599716 1.1884327 -0.3835923

19 20 21 22 23 24

-0.3134791 0.7746255 -0.3697885 -1.1328924 -2.0518635 -1.1354510

25

-1.3285962

> vif(lm.x3)

x1 x2 x4 x5 x6 x7 x8

14.632845 16.421404 1.824200 3.856707 4.694497 5.056370 1.867641

x9

2.256168

> PRESS(lm.x3)

.........10.........20.....

$stat

[1] 16.50452

$residuals

[1] 0.14439181 0.78982402 0.48481314 1.01966816 0.24128799

[6] 0.89641612 1.23329143 0.12543995 -0.74193281 -0.44672309

[11] -1.93542198 1.27051773 0.67963126 0.69509449 0.17784892

[16] -0.04186178 0.84163201 -0.27937715 -0.33683186 0.66494643

[21] -0.25118631 -0.93422484 -1.19500953 -0.72768164 -0.92875018

$P.square

[1] 0.7409605

>

> #Partial x4 ith

> lm.x4=lm(y~x1+x2+x3+x5+x6+x7+x8+x9)

> lm.x4

Call:

lm(formula = y ~ x1 + x2 + x3 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x5

5.53799 0.87123 -2.65774 1.10363 0.18281

x6 x7 x8 x9

-0.01998 -0.07916 -0.08502 -0.39119

> lm.x4ith=lm(x4~x1+x2+x3+x5+x6+x7+x8+x9)

> lm.x4ith

Call:

lm(formula = x4 ~ x1 + x2 + x3 + x5 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x5

30.82361 1.39056 -6.44574 -0.29993 0.02643

x6 x7 x8 x9

-0.01356 -0.01491 0.01013 -0.36797

> resid.x4=resid(lm.x4)

> resid.x4

1 2 3 4 5 6

0.16767898 0.56490838 0.28243712 0.28759656 -0.02294737 0.23917935

7 8 9 10 11 12

0.49733412 0.27705107 -0.45576115 -0.11609789 -0.33238735 0.73349946

13 14 15 16 17 18

0.45195940 0.07515008 0.09060310 -0.31127565 0.40688469 -0.24265573

19 20 21 22 23 24

-0.10713394 0.45607275 -0.21361358 -0.35580336 -1.27636840 -0.65165858

25

-0.44465205

> resid.x4ith=resid(lm.x4ith)

> resid.x4ith

1 2 3 4 5 6

0.30987078 -0.97360354 -0.18519099 0.72328205 0.40136968 -0.24241391

7 8 9 10 11 12

0.70049944 0.08364055 -0.87075463 -0.14998599 0.43488037 0.48481625

13 14 15 16 17 18

0.46428591 -1.46868320 0.14308741 -0.37237985 0.27427764 -0.22724648

19 20 21 22 23 24

-0.32533609 0.06822794 -0.46062476 0.95020923 -0.38483431 0.35294090

25

0.26966562

> plot(resid.x4,resid.x4ith,main="Partial Regression Plot for x4 as ith", xlab= "Residuals for y=x1+x2+x3+x5+x6+x7+x8+x9", ylab= "Residuals for x4=x1+x2+x3+x5+x6+x7+x8+x9")

> qqnorm(resid.x4, main = "Normality Plot for model with x4 as ith")

>

> summary(lm.x4)

Call:

lm(formula = y ~ x1 + x2 + x3 + x5 + x6 + x7 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.27637 -0.31128 0.07515 0.28760 0.73350

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.53799 2.95468 1.874 0.079266 .

x1 0.87123 0.47613 1.830 0.085973 .

x2 -2.65774 3.83522 -0.693 0.498257

x3 1.10363 0.72546 1.521 0.147702

x5 0.18281 0.07880 2.320 0.033895 \*

x6 -0.01998 0.02376 -0.841 0.412833

x7 -0.07916 0.01591 -4.975 0.000138 \*\*\*

x8 -0.08502 0.05069 -1.677 0.112933

x9 -0.39119 0.19199 -2.038 0.058487 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5559 on 16 degrees of freedom

Multiple R-squared: 0.9224, Adjusted R-squared: 0.8836

F-statistic: 23.77 on 8 and 16 DF, p-value: 1.75e-07

> predictlm.x4=predict(lm.x4)

> predictlm.x4

1 2 3 4 5 6 7

10.812321 10.535092 12.227563 8.112403 9.292947 8.490821 5.862666

8 9 10 11 12 13 14

8.222949 8.275761 9.256098 8.572387 11.456501 11.428041 9.494850

15 16 17 18 19 20 21

10.849397 9.891276 9.683115 8.352656 6.937134 8.423927 7.893614

22 23 24 25

8.825803 10.136368 11.011659 11.524652

> plot(predictlm.x4,resid.x4, main= "Residuals vs. Predicted Response for x4 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx4=rstandard(lm.x4)

> standx4

1 2 3 4 5 6

0.33290107 1.14265177 0.59932418 0.82460700 -0.04766458 0.60577340

7 8 9 10 11 12

1.36514236 0.60142111 -0.91375395 -0.24626395 -1.43178824 1.80107286

13 14 15 16 17 18

0.97282224 0.16106991 0.20167227 -0.66457396 0.91734269 -0.52126578

19 20 21 22 23 24

-0.33453219 1.20181466 -0.42317941 -0.81231332 -2.53402695 -1.27840162

25

-0.98515588

> studentx4=rstudent(lm.x4)

> studentx4

1 2 3 4 5 6

0.32345220 1.15447460 0.58691850 0.81594885 -0.04615431 0.59338168

7 8 9 10 11 12

1.40622356 0.58901944 -0.90876778 -0.23889723 -1.48469617 1.95306785

13 14 15 16 17 18

0.97108823 0.15608186 0.19551699 -0.65254024 0.91253467 -0.50905441

19 20 21 22 23 24

-0.32504817 1.22002141 -0.41205416 -0.80325686 -3.17105230 -1.30632039

25

-0.98418962

> vif(lm.x4)

x1 x2 x3 x5 x6 x7

11.751771 18.099620 125.768897 4.392236 4.634851 5.862202

x8 x9

107.419983 2.080472

> PRESS(lm.x4)

.........10.........20.....

$stat

[1] 14.91339

$residuals

[1] 0.20422160 0.71416687 0.39296308 0.73056738 -0.03059204

[6] 0.47407455 1.15786400 0.40341071 -0.56607133 -0.16140882

[11] -1.90573718 1.36650958 0.64701966 0.10667173 0.13870727

[16] -0.43842106 0.63906021 -0.34600136 -0.32277408 0.97856750

[21] -0.25904195 -0.57304276 -1.55452008 -0.77493348 -0.67443410

$P.square

[1] 0.7659334

#Partial x5 ith

> lm.x5=lm(y~x1+x2+x3+x4+x6+x7+x8+x9)

> lm.x5

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

-0.368080 0.329448 3.854711 1.727504 0.152849

x6 x7 x8 x9

-0.008127 -0.070212 -0.121875 -0.293553

> lm.x5ith=lm(x5~x1+x2+x3+x4+x6+x7+x8+x9)

> lm.x5ith

Call:

lm(formula = x5 ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

-11.85754 -2.06825 31.86884 3.26962 0.16884

x6 x7 x8 x9

0.05673 0.03968 -0.19833 0.29264

> resid.x5=resid(lm.x5)

> resid.x5

1 2 3 4 5

0.271993846 0.617901205 0.463949832 0.399031314 -0.436784619

6 7 8 9 10

0.220578892 0.009821334 0.429469371 -0.060027689 0.022639821

11 12 13 14 15

-0.418351147 1.004493185 0.034111634 0.091016307 0.493330485

16 17 18 19 20

-0.628092036 0.150255019 0.228490638 -0.364155846 0.356378118

21 22 23 24 25

0.155748369 -0.544836425 -1.406229673 -0.777027849 -0.313704086

> resid.x5ith=resid(lm.x5ith)

> resid.x5ith

1 2 3 4 5 6

0.7923612 -0.3692343 0.8844588 1.1141054 -2.0307385 -0.2689970

7 8 9 10 11 12

-2.2369164 0.9058743 1.6096321 0.6702717 -0.1819794 1.8399536

13 14 15 16 17 18

-2.0101413 -0.9138097 2.3403833 -2.0184134 -1.2419883 2.4687013

19 20 21 22 23 24

-1.6533235 -0.5086331 1.7426271 -0.4043228 -0.9857784 -0.4573364

25

0.9132438

> plot(resid.x5,resid.x5ith,main="Partial Regression Plot for x5 as ith", xlab= "Residuals for y=x1+x2+x3+x4+x6+x7+x8+x9", ylab= "Residuals for x5=x1+x2+x3+x4+x6+x7+x8+x9")

> qqnorm(resid.x5, main = "Normality Plot for model with x5 as ith")

>

> summary(lm.x5)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.40623 -0.36416 0.09102 0.35638 1.00449

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.368080 7.707917 -0.048 0.96250

x1 0.329448 0.600066 0.549 0.59057

x2 3.854711 3.610297 1.068 0.30150

x3 1.727504 0.775848 2.227 0.04069 \*

x4 0.152849 0.227068 0.673 0.51047

x6 -0.008127 0.026775 -0.304 0.76540

x7 -0.070212 0.018105 -3.878 0.00133 \*\*

x8 -0.121875 0.055009 -2.216 0.04157 \*

x9 -0.293553 0.232855 -1.261 0.22550

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6337 on 16 degrees of freedom

Multiple R-squared: 0.8992, Adjusted R-squared: 0.8487

F-statistic: 17.83 on 8 and 16 DF, p-value: 1.331e-06

> predictlm.x5=predict(lm.x5)

> predictlm.x5

1 2 3 4 5 6 7

10.708006 10.482099 12.046050 8.000969 9.706785 8.509421 6.350179

8 9 10 11 12 13 14

8.070531 7.880028 9.117360 8.658351 11.185507 11.845888 9.478984

15 16 17 18 19 20 21

10.446670 10.208092 9.939745 7.881509 7.194156 8.523622 7.524252

22 23 24 25

9.014836 10.266230 11.137028 11.393704

> plot(predictlm.x5,resid.x5, main= "Residuals vs. Predicted Response for x5 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx5=rstandard(lm.x5)

> standx5

1 2 3 4 5 6

0.47361134 1.18993889 0.85685321 1.06250175 -0.76465028 0.49306366

7 8 9 10 11 12

0.02267385 0.80864384 -0.10868656 0.04194676 -1.70069400 2.09083363

13 14 15 16 17 18

0.06206378 0.21576074 0.89235610 -1.12637653 0.29228189 0.39875392

19 20 21 22 23 24

-0.94037856 0.81973694 0.26525590 -1.20661706 -2.44826380 -1.34668784

25

-0.60626942

> studentx5=rstudent(lm.x5)

> studentx5

1 2 3 4 5 6

0.46182082 1.20678884 0.84936084 1.06709696 -0.75427958 0.48107568

7 8 9 10 11 12

0.02195422 0.79947330 -0.10527418 0.04061701 -1.81932195 2.37467866

13 14 15 16 17 18

0.06010023 0.20921402 0.88635800 -1.13660364 0.28375928 0.38802469

19 20 21 22 23 24

-0.93677305 0.81091801 0.25739951 -1.22538351 -2.99759874 -1.38476469

25

-0.59387897

> vif(lm.x5)

x1 x2 x3 x4 x6 x7

14.363452 12.341883 110.690568 1.828431 4.530146 5.840032

x8 x9

97.334598 2.354898

> PRESS(lm.x5)

.........10.........20.....

$stat

[1] 22.26887

$residuals

[1] 0.33115331 0.92018314 0.63545650 1.13604707 -0.53752957

[6] 0.44257415 0.02101959 0.61140130 -0.07902128 0.03120814

[11] -2.77622969 1.74757554 0.04534380 0.20538543 0.64816031

[16] -0.81112521 0.22830679 0.27943804 -0.97512917 0.75715009

[21] 0.18140557 -1.07304215 -1.71160743 -0.93721945 -0.47049515

$P.square

[1] 0.6504887

>

> #Partial x6 ith

> lm.x6=lm(y~x1+x2+x3+x4+x5+x7+x8+x9)

> lm.x6

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

0.41533 0.73316 -1.62274 1.14622 0.13984

x5 x7 x8 x9

0.16823 -0.06811 -0.08615 -0.34468

> lm.x6ith=lm(x6~x1+x2+x3+x4+x5+x7+x8+x9)

> lm.x6ith

Call:

lm(formula = x6 ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

73.483795 -1.764852 -13.388023 -0.318968 -0.945086

x5 x7 x8 x9

0.618941 -0.503772 -0.006281 -0.077402

> resid.x6=resid(lm.x6)

> resid.x6

1 2 3 4 5

0.064210634 0.639295090 0.333045549 0.004969973 -0.042981268

6 7 8 9 10

0.252588594 0.415787678 0.170457276 -0.336130958 -0.012793301

11 12 13 14 15

-0.355830336 0.873768227 0.285973270 0.291198968 0.202940240

16 17 18 19 20

-0.215633735 0.410744196 -0.146047434 -0.072406531 0.355191105

21 22 23 24 25

-0.123652956 -0.389488667 -1.232450231 -0.758263112 -0.614492270

> resid.x6ith=resid(lm.x6ith)

> resid.x6ith

1 2 3 4 5 6

3.5763422 2.4520714 -1.5242876 10.5927016 -1.5913944 0.8896527

7 8 9 10 11 12

-0.2340117 5.2605297 -0.7063425 -4.6370693 -1.6294501 -10.9023238

13 14 15 16 17 18

5.9567881 -1.9705922 -7.0909861 -2.7308123 -2.0457159 -3.7545547

19 20 21 22 23 24

0.2804335 5.0517561 -1.8302296 -4.5178863 0.1766686 3.4594309

25

7.4692814

> plot(resid.x6,resid.x6ith,main="Partial Regression Plot for x6 as ith", xlab= "Residuals for y=x1+x2+x3+x4+x5+x7+x8+x9", ylab= "Residuals for x6=x1+x2+x3+x4+x5+x7+x8+x9")

> qqnorm(resid.x6, main = "Normality Plot for model with x6 as ith")

>

> summary(lm.x6)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.23245 -0.21563 0.00497 0.29120 0.87377

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.41533 6.63911 0.063 0.9509

x1 0.73316 0.55312 1.326 0.2036

x2 -1.62274 4.05901 -0.400 0.6946

x3 1.14622 0.73264 1.565 0.1373

x4 0.13984 0.19965 0.700 0.4937

x5 0.16823 0.07809 2.154 0.0468 \*

x7 -0.06811 0.01088 -6.257 1.14e-05 \*\*\*

x8 -0.08615 0.05106 -1.687 0.1110

x9 -0.34468 0.20690 -1.666 0.1152

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5595 on 16 degrees of freedom

Multiple R-squared: 0.9214, Adjusted R-squared: 0.8821

F-statistic: 23.44 on 8 and 16 DF, p-value: 1.937e-07

> predictlm.x6=predict(lm.x6)

> predictlm.x6

1 2 3 4 5 6 7

10.915789 10.460705 12.176954 8.395030 9.312981 8.477411 5.944212

8 9 10 11 12 13 14

8.329543 8.156131 9.152793 8.595830 11.316232 11.594027 9.278801

15 16 17 18 19 20 21

10.737060 9.795634 9.679256 8.256047 6.902407 8.524809 7.803653

22 23 24 25

8.859489 10.092450 11.118263 11.694492

> plot(predictlm.x6,resid.x6, main= "Residuals vs. Predicted Response for x6 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx6=rstandard(lm.x6)

> standx6

1 2 3 4 5 6

0.12579090 1.38574126 0.70218304 0.01215781 -0.08966443 0.63946578

7 8 9 10 11 12

1.22778901 0.35485753 -0.71384817 -0.02629473 -1.61570740 1.83234109

13 14 15 16 17 18

0.59561943 0.79049787 0.42062397 -0.45870909 0.92144628 -0.30746771

19 20 21 22 23 24

-0.22933923 0.88660864 -0.24655628 -0.94537103 -2.45962806 -1.47255796

25

-1.26552346

> studentx6=rstudent(lm.x6)

> studentx6

1 2 3 4 5 6

0.12185679 1.43031190 0.69061010 0.01177180 -0.08683903 0.62722706

7 8 9 10 11 12

1.24909999 0.34494943 -0.70245721 -0.02546032 -1.71011900 1.99588273

13 14 15 16 17 18

0.58320791 0.78079527 0.40953799 -0.44709273 0.91684497 -0.29858774

19 20 21 22 23 24

-0.22242264 0.88035321 -0.23918189 -0.94203979 -3.01994138 -1.53349456

25

-1.29168898

> vif(lm.x6)

x1 x2 x3 x4 x5 x7

15.654279 20.011010 126.611716 1.813091 4.256994 2.707750

x8 x9

107.589948 2.384853

> PRESS(lm.x6)

.........10.........20.....

$stat

[1] 15.78359

$residuals

[1] 0.077144164 0.940317496 0.463456124 0.009310396 -0.058556181

[6] 0.506795168 1.134978047 0.231261904 -0.474585846 -0.016918683

[11] -2.296645938 1.202895727 0.388350491 0.671773968 0.272917547

[16] -0.305470714 0.647113105 -0.202636032 -0.227400044 0.692808845

[21] -0.153900214 -0.718325600 -1.536671644 -0.895233773 -0.815897434

$P.square

[1] 0.7522756

>

> #Partial x7 ith

> lm.x7=lm(y~x1+x2+x3+x4+x5+x6+x8+x9)

> lm.x7

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

-14.10165 0.75383 0.17962 2.56000 0.29794

x5 x6 x8 x9

0.10652 0.06669 -0.17431 -0.29254

> lm.x7ith=lm(x7~x1+x2+x3+x4+x5+x6+x8+x9)

> lm.x7ith

Call:

lm(formula = x7 ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

205.1118 -0.6852 -26.4756 -18.3561 -2.2680

x5 x6 x8 x9

0.9445 -1.0992 1.1385 -0.6926

> resid.x7=resid(lm.x7)

> resid.x7

1 2 3 4 5

-0.003321882 0.927552635 0.763735527 -1.270856030 -0.322066598

6 7 8 9 10

-0.733645272 0.395191250 -0.298255126 -0.074886766 0.591294949

11 12 13 14 15

0.070336809 2.031485858 0.320375384 0.632635230 0.460747028

16 17 18 19 20

-0.288626950 0.017183373 -0.112198126 -0.242689145 0.235870612

21 22 23 24 25

-0.188174654 -0.299230653 -1.154959970 -0.717748923 -0.739748559

> resid.x7ith=resid(lm.x7ith)

> resid.x7ith

1 2 3 4 5 6

1.7201549 -3.1465811 -5.9299488 19.0054019 3.2318034 12.9630346

7 8 9 10 11 12

0.2109013 7.3063947 -3.5452525 -8.9092065 -5.8963711 -17.5515427

13 14 15 16 17 18

0.9658325 -4.8815590 -5.0127927 0.2971273 4.6044211 -1.3268228

19 20 21 22 23 24

2.2682252 2.7391936 0.4008602 -2.2369765 -0.9601383 0.2953866

25

3.3884547

> plot(resid.x7,resid.x7ith,main="Partial Regression Plot for x7 as ith", xlab= "Residuals for y=x1+x2+x3+x4+x5+x6+x8+x9", ylab= "Residuals for x7=x1+x2+x3+x4+x5+x6+x8+x9")

> qqnorm(resid.x7, main = "Normality Plot for model with x7 as ith")

>

> summary(lm.x7)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.27086 -0.29923 -0.07489 0.39519 2.03149

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -14.10165 9.24517 -1.525 0.1467

x1 0.75383 0.85419 0.883 0.3906

x2 0.17962 6.23559 0.029 0.9774

x3 2.56000 1.03010 2.485 0.0244 \*

x4 0.29794 0.30418 0.979 0.3419

x5 0.10652 0.12011 0.887 0.3883

x6 0.06669 0.02476 2.693 0.0160 \*

x8 -0.17431 0.07327 -2.379 0.0302 \*

x9 -0.29254 0.31818 -0.919 0.3715

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8617 on 16 degrees of freedom

Multiple R-squared: 0.8135, Adjusted R-squared: 0.7203

F-statistic: 8.726 on 8 and 16 DF, p-value: 0.0001402

> predictlm.x7=predict(lm.x7)

> predictlm.x7

1 2 3 4 5 6 7

10.983322 10.172447 11.746264 9.670856 9.592067 9.463645 5.964809

8 9 10 11 12 13 14

8.798255 7.894887 8.548705 8.169663 10.158514 11.559625 8.937365

15 16 17 18 19 20 21

10.479253 9.868627 10.072817 8.222198 7.072689 8.644129 7.868175

22 23 24 25

8.769231 10.014960 11.077749 11.819749

> plot(predictlm.x7,resid.x7, main= "Residuals vs. Predicted Response for x7 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx7=rstandard(lm.x7)

> standx7

1 2 3 4 5

-0.004280291 1.308122679 1.027464977 -1.854430998 -0.435022623

6 7 8 9 10

-1.064714863 0.757797127 -0.404796745 -0.102563672 0.775142622

11 12 13 14 15

0.192660293 2.690707046 0.453713494 1.098732540 0.652335753

16 17 18 19 20

-0.402602580 0.024830698 -0.156057688 -0.495848853 0.398504864

21 22 23 24 25

-0.244553018 -0.486886978 -1.495983602 -0.917094485 -1.057002079

> studentx7=rstudent(lm.x7)

> studentx7

1 2 3 4 5

-0.004144376 1.340281420 1.029377425 -2.026481852 -0.423722146

6 7 8 9 10

-1.069488942 0.747266530 -0.393965308 -0.099339509 0.765030619

11 12 13 14 15

0.186759282 3.520931882 0.442159823 1.106401147 0.640192143

16 17 18 19 20

-0.391807939 0.024042683 -0.151217336 -0.483835440 0.387779906

21 22 23 24 25

-0.237231229 -0.474957952 -1.561820663 -0.912274062 -1.061157694

> vif(lm.x7)

x1 x2 x3 x4 x5 x6

15.740218 19.911086 105.525135 1.774504 4.246571 2.095274

x8 x9

93.393373 2.377952

> PRESS(lm.x7)

.........10.........20.....

$stat

[1] 25.74788

$residuals

[1] -0.004095079 1.369803834 1.026339519 -2.009208436 -0.436292743

[6] -1.147309267 1.078943193 -0.407930424 -0.104299454 0.754499674

[11] 0.391833680 2.646177786 0.477094316 1.416871913 0.685772100

[16] -0.416980954 0.026642160 -0.161170192 -0.752225871 0.499911826

[21] -0.235985283 -0.588234144 -1.438754598 -0.870072825 -1.121418061

$P.square

[1] 0.5958855

>

> #Partial x8 ith

> lm.x8=lm(y~x1+x2+x3+x4+x5+x6+x7+x9)

> lm.x8

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

6.17909 0.93302 -4.62632 -0.08652 0.10900

x5 x6 x7 x9

0.22103 -0.01819 -0.08734 -0.25544

> lm.x8ith=lm(x8~x1+x2+x3+x4+x5+x6+x7+x9)

> lm.x8ith

Call:

lm(formula = x8 ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x9)

Coefficients:

(Intercept) x1 x2 x3 x4

-51.215802 -2.691535 31.977132 14.223104 0.156832

x5 x6 x7 x9

-0.480643 -0.001395 0.115910 -1.051031

> resid.x8=resid(lm.x8)

> resid.x8

1 2 3 4 5 6

0.11685567 0.54114034 0.36257195 0.37399549 0.21246790 0.48632345

7 8 9 10 11 12

0.45807282 0.08373865 -0.55875840 -0.36311343 -0.97301299 0.66650186

13 14 15 16 17 18

0.46614894 0.32097243 0.10991827 -0.04480286 0.57111765 -0.16138290

19 20 21 22 23 24

-0.11034738 0.32366967 -0.17799235 -0.47597095 -0.98956739 -0.59269648

25

-0.64584993

> resid.x8ith=resid(lm.x8ith)

> resid.x8ith

1 2 3 4 5 6

0.14902204 1.65849950 -0.66592491 -2.02899093 -3.29923318 -2.52073448

7 8 9 10 11 12

-0.53988489 2.12220971 2.43088939 3.07663190 6.80886683 0.08802353

13 14 15 16 17 18

-0.82399412 -0.76354745 -0.42716518 -2.56019546 -2.29350726 -0.61937926

19 20 21 22 23 24

0.49937879 1.43799751 0.24135113 0.04333723 -2.77816387 -1.18486926

25

1.94938271

> plot(resid.x8,resid.x8ith,main="Partial Regression Plot for x8 as ith", xlab= "Residuals for y=x1+x2+x3+x4+x5+x6+x7+x9", ylab= "Residuals for x8=x1+x2+x3+x4+x5+x6+x7+x9")

> qqnorm(resid.x8, main = "Normality Plot for model with x8 as ith")

>

> summary(lm.x8)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x9)

Residuals:

Min 1Q Median 3Q Max

-0.9896 -0.3631 0.1099 0.3740 0.6665

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.17909 6.79000 0.910 0.3763

x1 0.93302 0.57439 1.624 0.1238

x2 -4.62632 3.98634 -1.161 0.2629

x3 -0.08652 0.10317 -0.839 0.4140

x4 0.10900 0.21457 0.508 0.6184

x5 0.22103 0.08081 2.735 0.0147 \*

x6 -0.01819 0.02572 -0.707 0.4895

x7 -0.08734 0.01622 -5.383 6.09e-05 \*\*\*

x9 -0.25544 0.21356 -1.196 0.2491

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.598 on 16 degrees of freedom

Multiple R-squared: 0.9102, Adjusted R-squared: 0.8653

F-statistic: 20.27 on 8 and 16 DF, p-value: 5.432e-07

> predictlm.x8=predict(lm.x8)

> predictlm.x8

1 2 3 4 5 6 7

10.863144 10.558860 12.147428 8.026005 9.057532 8.243677 5.901927

8 9 10 11 12 13 14

8.416261 8.378758 9.503113 9.213013 11.523498 11.413851 9.249028

15 16 17 18 19 20 21

10.830082 9.624803 9.518882 8.271383 6.940347 8.556330 7.857992

22 23 24 25

8.945971 9.849567 10.952696 11.725850

> plot(predictlm.x8,resid.x8, main= "Residuals vs. Predicted Response for x8 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx8=rstandard(lm.x8)

> standx8

1 2 3 4 5 6

0.21729176 1.08816727 0.71558471 1.04173314 0.39239001 1.09680110

7 8 9 10 11 12

1.26167965 0.16465417 -1.07432000 -0.68097205 -2.22223837 1.56615747

13 14 15 16 17 18

0.94789723 0.81749603 0.22764192 -0.08670867 1.16616172 -0.32307847

19 20 21 22 23 24

-0.32605162 0.77911432 -0.33327258 -1.12071787 -1.77808752 -1.08372175

25

-1.30830038

> studentx8=rstudent(lm.x8)

> studentx8

1 2 3 4 5 6

0.21070296 1.09490739 0.70422246 1.04470461 0.38177135 1.10429795

7 8 9 10 11 12

1.28733150 0.15956096 -1.07988362 -0.66911603 -2.58777304 1.64799950

13 14 15 16 17 18

0.94470665 0.80860444 0.22077115 -0.08397504 1.18041000 -0.31384478

19 20 21 22 23 24

-0.31675218 0.76910468 -0.32381570 -1.13040583 -1.92195507 -1.09007920

25

-1.34048530

> vif(lm.x8)

x1 x2 x3 x4 x5 x6 x7

14.780030 16.898604 2.198060 1.833708 3.991346 4.694972 5.266779

x9

2.224572

> PRESS(lm.x8)

.........10.........20.....

$stat

[1] 16.20219

$residuals

[1] 0.14446878 0.78238173 0.50496957 1.03748996 0.25910742

[6] 0.88443891 1.24251307 0.11575956 -0.73855111 -0.45661855

[11] -1.81467954 1.31585123 0.68918388 0.74445915 0.16856674

[16] -0.06000074 0.85139145 -0.23125749 -0.34446730 0.67056005

[21] -0.22311827 -0.94351637 -1.14234738 -0.70850110 -0.94759032

$P.square

[1] 0.7457056

>

> #Partial x9 ith

> lm.x9=lm(y~x1+x2+x3+x4+x5+x6+x7+x8)

> lm.x9

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8)

Coefficients:

(Intercept) x1 x2 x3 x4

-1.10958 0.46648 -1.77992 0.85549 0.24264

x5 x6 x7 x8

0.16462 -0.01795 -0.07585 -0.06410

> lm.x9ith=lm(x9~x1+x2+x3+x4+x5+x6+x7+x8)

> lm.x9ith

Call:

lm(formula = x9 ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8)

Coefficients:

(Intercept) x1 x2 x3 x4

8.294526 0.677157 -0.254322 0.823146 -0.347039

x5 x6 x7 x8

0.043195 -0.001047 -0.004295 -0.064016

> resid.x9=resid(lm.x9)

> resid.x9

1 2 3 4 5 6

0.19021737 0.59177636 0.63604388 0.04345789 -0.13919279 0.44038936

7 8 9 10 11 12

0.48789363 0.17247904 -0.19383881 -0.26303752 -0.61588656 0.70389124

13 14 15 16 17 18

0.33295606 0.16398528 0.23923609 -0.06568201 0.09847997 -0.14159158

19 20 21 22 23 24

-0.01391441 0.65747717 -0.11515386 -1.02178799 -1.02720299 -0.57974810

25

-0.58124672

> resid.x9ith=resid(lm.x9ith)

> resid.x9ith

1 2 3 4 5 6

-0.17482443 0.26705447 -0.95612353 0.44933362 0.19377444 -0.49554033

7 8 9 10 11 12

-0.22072119 0.27253186 -0.44850677 0.47765666 0.66516209 -0.08609159

13 14 15 16 17 18

0.17946877 0.26328379 -0.48010703 -0.57776746 0.79398055 -0.21155559

19 20 21 22 23 24

-0.15416325 -0.60607861 -0.12140764 1.58784685 -0.58367821 -0.33272483

25

0.29919738

> plot(resid.x9,resid.x9ith,main="Partial Regression Plot for x9 as ith", xlab= "Residuals for y=x1+x2+x3+x4+x5+x6+x7+x8", ylab= "Residuals for x9=x1+x2+x3+x4+x5+x6+x7+x8")

> qqnorm(resid.x9, main = "Normality Plot for model with x9 as ith")

>

> summary(lm.x9)

Call:

lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8)

Residuals:

Min 1Q Median 3Q Max

-1.02720 -0.19384 0.04346 0.33296 0.70389

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.10958 7.09917 -0.156 0.877754

x1 0.46648 0.57278 0.814 0.427364

x2 -1.77992 4.34460 -0.410 0.687469

x3 0.85549 0.76042 1.125 0.277174

x4 0.24264 0.20026 1.212 0.243246

x5 0.16462 0.08429 1.953 0.068536 .

x6 -0.01795 0.02568 -0.699 0.494563

x7 -0.07585 0.01736 -4.369 0.000477 \*\*\*

x8 -0.06410 0.05263 -1.218 0.240831

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.597 on 16 degrees of freedom

Multiple R-squared: 0.9105, Adjusted R-squared: 0.8657

F-statistic: 20.34 on 8 and 16 DF, p-value: 5.306e-07

> predictlm.x9=predict(lm.x9)

> predictlm.x9

1 2 3 4 5 6 7

10.789783 10.508224 11.873956 8.356542 9.409193 8.289611 5.872106

8 9 10 11 12 13 14

8.327521 8.013839 9.403038 8.855887 11.486109 11.547044 9.406015

15 16 17 18 19 20 21

10.700764 9.645682 9.991520 8.251592 6.843914 8.222523 7.795154

22 23 24 25

9.491788 9.887203 10.939748 11.661247

> plot(predictlm.x9,resid.x9, main= "Residuals vs. Predicted Response for x9 as ith", xlab ="Predicted Response", ylab="Residuals")

> standx9=rstandard(lm.x9)

> standx9

1 2 3 4 5 6

0.35337483 1.20330596 1.16282131 0.12237207 -0.27203442 1.01269469

7 8 9 10 11 12

1.33820146 0.34627126 -0.37873464 -0.50954470 -2.24821617 1.65498827

13 14 15 16 17 18

0.67871459 0.41607185 0.48517966 -0.12808026 0.19531462 -0.28329663

19 20 21 22 23 24

-0.04110146 1.53347520 -0.21573584 -1.85710423 -1.86770605 -1.05948298

25

-1.19638859

> studentx9=rstudent(lm.x9)

> studentx9

1 2 3 4 5 6

0.34349676 1.22168583 1.17671617 0.11854173 -0.26400744 1.01355829

7 8 9 10 11 12

1.37493465 0.33653909 -0.36836314 -0.49741690 -2.63187596 1.76016016

13 14 15 16 17 18

0.66683201 0.40505710 0.47326755 -0.12407680 0.18933841 -0.27499134

19 20 21 22 23 24

-0.03979841 1.60760985 -0.20918981 -2.03020507 -2.04501474 -1.06383604

25

-1.21397061

> vif(lm.x9)

x1 x2 x3 x4 x5 x6

14.742133 20.133751 119.783271 1.602086 4.356152 4.694632

x7 x8

6.049379 100.351820

> PRESS(lm.x9)

.........10.........20.....

$stat

[1] 20.96213

$residuals

[1] 0.23401019 0.87218425 0.75779652 0.12283148 -0.18951548

[6] 0.83010679 1.30837298 0.24780513 -0.26378071 -0.35185247

[11] -2.92542604 1.38706879 0.49317512 0.37630979 0.35074557

[16] -0.08902904 0.13808144 -0.20205011 -0.04327762 1.27493061

[21] -0.14407208 -1.20316580 -1.21052703 -0.69017998 -0.87780434

$P.square

[1] 0.6709981

#Partial x8 and x1 removed

> lm.nox8x1=lm(y~x2+x3+x4+x5+x6+x7+x9)

> lm.nox8x1

Call:

lm(formula = y ~ x2 + x3 + x4 + x5 + x6 + x7 + x9)

Coefficients:

(Intercept) x2 x3 x4 x5

2.12600 0.93669 -0.05979 0.28680 0.18971

x6 x7 x9

-0.02146 -0.09055 -0.13797

> summary(lm.nox8x1)

Call:

lm(formula = y ~ x2 + x3 + x4 + x5 + x6 + x7 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.2875 -0.2377 0.1293 0.3050 0.6737

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.12600 6.61225 0.322 0.7517

x2 0.93669 2.13621 0.438 0.6666

x3 -0.05979 0.10664 -0.561 0.5823

x4 0.28680 0.19325 1.484 0.1561

x5 0.18971 0.08217 2.309 0.0338 \*

x6 -0.02146 0.02685 -0.799 0.4351

x7 -0.09055 0.01686 -5.370 5.09e-05 \*\*\*

x9 -0.13797 0.21040 -0.656 0.5208

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6261 on 17 degrees of freedom

Multiple R-squared: 0.8954, Adjusted R-squared: 0.8523

F-statistic: 20.79 on 7 and 17 DF, p-value: 3.65e-07

> resid.nox8x1=resid(lm.nox8x1)

> resid.nox8x1

1 2 3 4 5 6

0.26074962 0.65740043 0.54757955 0.21388003 0.06112276 0.67366270

7 8 9 10 11 12

0.12116450 0.18532312 -0.39733621 -0.21717550 -1.28751392 0.44760142

13 14 15 16 17 18

0.30499911 0.30044597 0.28561069 -0.04112656 0.66573767 -0.23766093

19 20 21 22 23 24

0.33250817 0.12928548 -0.08194265 -0.80591523 -1.05810124 -0.62337536

25

-0.43692363

> qqnorm(resid.nox8x1, main = "Normality Plot for model without x1 and x8")

> predictlm.nox8x1=predict(lm.nox8x1)

> plot(predictlm.nox8x1,resid.nox8x1, main= "Residuals vs. Predicted Response for x9 as ith", xlab ="Predicted Response", ylab="Residuals")

> PRESS(lm.nox8x1)

.........10.........20.....

$stat

[1] 14.71888

$residuals

[1] 0.31384925 0.93118232 0.72595561 0.55172715 0.07239647

[6] 1.14750510 0.24779006 0.25237291 -0.50669009 -0.26556137

[11] -2.00850073 0.80314827 0.43329598 0.69612929 0.41707495

[16] -0.05507631 0.97860166 -0.33757857 0.62948305 0.24732428

[21] -0.10147349 -1.30016411 -1.21448235 -0.74428660 -0.60030286

$P.square

[1] 0.7689863

> standnox8x1=rstandard(lm.nox8x1)

> standnox8x1

1 2 3 4 5 6

0.45690070 1.24963099 1.00699677 0.54865170 0.10624527 1.40426153

7 8 9 10 11 12

0.27674454 0.34541056 -0.71663780 -0.38356352 -2.56839252 0.95761960

13 14 15 16 17 18

0.58061874 0.73042828 0.55124379 -0.07601391 1.28915155 -0.45239288

19 20 21 22 23 24

0.73070590 0.28559966 -0.14564002 -1.63490755 -1.81054219 -1.08791346

25

-0.81797006

> vif(lm.nox8x1)

x2 x3 x4 x5 x6 x7 x9

4.426137 2.142163 1.356569 3.764093 4.666206 5.188386 1.969501

>

> lm.nox8x2=lm(y~x1+x3+x4+x5+x6+x7+x9)

> lm.nox8x2

Call:

lm(formula = y ~ x1 + x3 + x4 + x5 + x6 + x7 + x9)

Coefficients:

(Intercept) x1 x3 x4 x5

4.33882 0.36033 -0.05475 0.19119 0.16739

x6 x7 x9

-0.01558 -0.08821 -0.22268

> summary(lm.nox8x2)

Call:

lm(formula = y ~ x1 + x3 + x4 + x5 + x6 + x7 + x9)

Residuals:

Min 1Q Median 3Q Max

-1.2074 -0.2540 0.1206 0.3071 0.6553

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.33882 6.66926 0.651 0.5240

x1 0.36033 0.29695 1.213 0.2415

x3 -0.05475 0.10048 -0.545 0.5929

x4 0.19119 0.20461 0.934 0.3632

x5 0.16739 0.06696 2.500 0.0229 \*

x6 -0.01558 0.02588 -0.602 0.5552

x7 -0.08821 0.01637 -5.388 4.9e-05 \*\*\*

x9 -0.22268 0.21383 -1.041 0.3123

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.604 on 17 degrees of freedom

Multiple R-squared: 0.9027, Adjusted R-squared: 0.8626

F-statistic: 22.52 on 7 and 17 DF, p-value: 2.018e-07

> resid.nox8x2=resid(lm.nox8x2)

> resid.nox8x2

1 2 3 4 5 6

0.24513144 0.58460682 0.44139017 0.40682238 0.05689118 0.56482843

7 8 9 10 11 12

0.24262338 0.12061815 -0.46826859 -0.25404659 -1.20738237 0.65534296

13 14 15 16 17 18

0.26914702 0.30041263 0.30714582 -0.14753715 0.62123314 -0.09696659

19 20 21 22 23 24

0.15295199 0.05116576 -0.03950754 -0.59070741 -1.08244011 -0.63297300

25

-0.50048192

> qqnorm(resid.nox8x2, main = "Normality Plot for model without x1 and x8")

> predictlm.nox8x2=predict(lm.nox8x2)

> plot(predictlm.nox8x2,resid.nox8x2, main= "Residuals vs. Predicted Response for x9 as ith", xlab ="Predicted Response", ylab="Residuals")

> PRESS(lm.nox8x2)

.........10.........20.....

$stat

[1] 14.88712

$residuals

[1] 0.29077314 0.84045833 0.60389323 1.12159180 0.06537262

[6] 1.00384602 0.52170771 0.16609298 -0.60533935 -0.30984152

[11] -1.85679086 1.29316046 0.35555923 0.69535739 0.41911121

[16] -0.19195028 0.91895625 -0.13725586 0.32942392 0.08033740

[21] -0.04716908 -1.11076678 -1.22424611 -0.75361250 -0.68988981

$P.square

[1] 0.7663458

> standnox8x2=rstandard(lm.nox8x2)

> standnox8x2

1 2 3 4 5 6

0.4419997 1.1604740 0.8547454 1.1183177 0.1009638 1.2466289

7 8 9 10 11 12

0.5890136 0.2343297 -0.8814392 -0.4644852 -2.4788445 1.5240735

13 14 15 16 17 18

0.5121488 0.7566735 0.5939944 -0.2786060 1.2508928 -0.1909949

19 20 21 22 23 24

0.3716212 0.1061436 -0.0714684 -1.3410440 -1.9058191 -1.1434364

25

-0.9728131

> vif(lm.nox8x2)

x1 x3 x4 x5 x6 x7 x9

3.871233 2.043285 1.633958 2.685582 4.658877 5.255365 2.185702

> #Conclusion: don't remove based on multicollinearity

>

> #Transformed model attempts

> sqx3=x3^2

> sqx8=x8^2

> lm.sqx3x8=lm(y~x1+x2+sqx3+x4+x5+x6+x7+sqx8+x9)

> summary(lm.sqx3x8)

Call:

lm(formula = y ~ x1 + x2 + sqx3 + x4 + x5 + x6 + x7 + sqx8 +

x9)

Residuals:

Min 1Q Median 3Q Max

-1.16318 -0.13067 -0.00757 0.31610 0.64030

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.0609004 6.0664240 0.505 0.621196

x1 0.5253783 0.5389080 0.975 0.345074

x2 -0.8008825 3.8840706 -0.206 0.839410

sqx3 0.0555070 0.0292446 1.898 0.077107 .

x4 0.1700484 0.1926556 0.883 0.391349

x5 0.1607023 0.0759639 2.116 0.051528 .

x6 -0.0232787 0.0228982 -1.017 0.325454

x7 -0.0771157 0.0148893 -5.179 0.000112 \*\*\*

sqx8 -0.0005262 0.0002384 -2.207 0.043311 \*

x9 -0.2897835 0.1908812 -1.518 0.149773

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5302 on 15 degrees of freedom

Multiple R-squared: 0.9338, Adjusted R-squared: 0.8941

F-statistic: 23.52 on 9 and 15 DF, p-value: 2.831e-07

> plot(sqx3,y)

> plot(sqx8,y)

> residsqx3x8=resid(lm.sqx3x8)

> predictsqx3x8=predict(lm.sqx3x8)

> plot(predictsqx3x8,residsqx3x8, main= "Residuals vs. Predicted Response for x9 as ith", xlab ="Predicted Response", ylab="Residuals")

>

> sqx2=x2^2

> sqx9=x9^2

> lm.sqx2x3x8=lm(y~x1+sqx2+sqx3+x4+x5+x6+x7+sqx8+x9)

> summary(lm.sqx2x3x8)

Call:

lm(formula = y ~ x1 + sqx2 + sqx3 + x4 + x5 + x6 + x7 + sqx8 +

x9)

Residuals:

Min 1Q Median 3Q Max

-1.20665 -0.17438 0.09663 0.24083 0.66091

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.7667156 5.7807599 0.479 0.639119

x1 0.2371022 0.4821657 0.492 0.630018

sqx2 1.0970955 2.3331518 0.470 0.644962

sqx3 0.0638924 0.0282053 2.265 0.038734 \*

x4 0.2035612 0.1838486 1.107 0.285638

x5 0.1363155 0.0663938 2.053 0.057928 .

x6 -0.0243351 0.0228971 -1.063 0.304682

x7 -0.0777092 0.0148649 -5.228 0.000102 \*\*\*

sqx8 -0.0005895 0.0002303 -2.560 0.021758 \*

x9 -0.2757266 0.1901496 -1.450 0.167632

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.527 on 15 degrees of freedom

Multiple R-squared: 0.9346, Adjusted R-squared: 0.8954

F-statistic: 23.82 on 9 and 15 DF, p-value: 2.598e-07

> plot(sqx2,y)

> lm.sqx2x3x8x9=lm(y~x1+sqx2+sqx3+x4+x5+x6+x7+sqx8+sqx9)

> summary(lm.sqx2x3x8x9)

Call:

lm(formula = y ~ x1 + sqx2 + sqx3 + x4 + x5 + x6 + x7 + sqx8 +

sqx9)

Residuals:

Min 1Q Median 3Q Max

-1.20393 -0.16998 0.06469 0.24623 0.67286

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.7198552 5.6227538 0.306 0.763905

x1 0.1728404 0.4742948 0.364 0.720634

sqx2 1.3275457 2.3375694 0.568 0.578494

sqx3 0.0654148 0.0284245 2.301 0.036130 \*

x4 0.2302134 0.1798594 1.280 0.220000

x5 0.1247320 0.0658305 1.895 0.077569 .

x6 -0.0254494 0.0230791 -1.103 0.287534

x7 -0.0775042 0.0149685 -5.178 0.000112 \*\*\*

sqx8 -0.0006003 0.0002325 -2.582 0.020838 \*

sqx9 -0.0272822 0.0200343 -1.362 0.193372

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5309 on 15 degrees of freedom

Multiple R-squared: 0.9336, Adjusted R-squared: 0.8938

F-statistic: 23.45 on 9 and 15 DF, p-value: 2.89e-07

> sqrtx6=sqrt(x6)

> lm.sqx2x3x8sqrtx6=lm(y~x1+sqx2+sqx3+x4+x5+sqrtx6+x7+sqx8+x9)

> summary(lm.sqx2x3x8sqrtx6)

Call:

lm(formula = y ~ x1 + sqx2 + sqx3 + x4 + x5 + sqrtx6 + x7 + sqx8 +

x9)

Residuals:

Min 1Q Median 3Q Max

-1.07899 -0.20368 0.05293 0.26547 0.60527

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.2319065 5.5446609 0.763 0.4572

x1 0.0221917 0.4810566 0.046 0.9638

sqx2 1.7240965 2.2457111 0.768 0.4546

sqx3 0.0698773 0.0269608 2.592 0.0204 \*

x4 0.2060570 0.1731591 1.190 0.2525

x5 0.1473051 0.0631305 2.333 0.0340 \*

sqrtx6 -0.2455904 0.1387612 -1.770 0.0971 .

x7 -0.0911187 0.0170326 -5.350 8.1e-05 \*\*\*

sqx8 -0.0006255 0.0002186 -2.861 0.0119 \*

x9 -0.2908944 0.1795418 -1.620 0.1260

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4971 on 15 degrees of freedom

Multiple R-squared: 0.9418, Adjusted R-squared: 0.9069

F-statistic: 26.98 on 9 and 15 DF, p-value: 1.106e-07

> plot(sqrtx6,y)

> lm.sqx2x3x8x9sqrtx6=lm(y~x1+sqx2+sqx3+x4+x5+sqrtx6+x7+sqx8+sqx9)

> summary(lm.sqx2x3x8x9sqrtx6)

Call:

lm(formula = y ~ x1 + sqx2 + sqx3 + x4 + x5 + sqrtx6 + x7 + sqx8 +

sqx9)

Residuals:

Min 1Q Median 3Q Max

-1.07500 -0.20284 0.06192 0.27059 0.61688

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.2166860 5.3840022 0.597 0.5591

x1 -0.0472108 0.4730110 -0.100 0.9218

sqx2 1.9828891 2.2462257 0.883 0.3913

sqx3 0.0716932 0.0271209 2.643 0.0184 \*

x4 0.2323285 0.1689150 1.375 0.1892

x5 0.1355870 0.0624261 2.172 0.0463 \*

sqrtx6 -0.2547064 0.1398612 -1.821 0.0886 .

x7 -0.0913240 0.0171467 -5.326 8.48e-05 \*\*\*

sqx8 -0.0006385 0.0002203 -2.898 0.0110 \*

sqx9 -0.0296213 0.0189163 -1.566 0.1382

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4995 on 15 degrees of freedom

Multiple R-squared: 0.9413, Adjusted R-squared: 0.906

F-statistic: 26.7 on 9 and 15 DF, p-value: 1.188e-07

> lm.sqx

Error: object 'lm.sqx' not found

>

> #Chosen transformed model

> lm.sqx2x3x8sqrtx6=lm(y~x1+sqx2+sqx3+x4+x5+sqrtx6+x7+sqx8+x9)

> summary(lm.sqx2x3x8sqrtx6)

Call:

lm(formula = y ~ x1 + sqx2 + sqx3 + x4 + x5 + sqrtx6 + x7 + sqx8 +

x9)

Residuals:

Min 1Q Median 3Q Max

-1.07899 -0.20368 0.05293 0.26547 0.60527

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.2319065 5.5446609 0.763 0.4572

x1 0.0221917 0.4810566 0.046 0.9638

sqx2 1.7240965 2.2457111 0.768 0.4546

sqx3 0.0698773 0.0269608 2.592 0.0204 \*

x4 0.2060570 0.1731591 1.190 0.2525

x5 0.1473051 0.0631305 2.333 0.0340 \*

sqrtx6 -0.2455904 0.1387612 -1.770 0.0971 .

x7 -0.0911187 0.0170326 -5.350 8.1e-05 \*\*\*

sqx8 -0.0006255 0.0002186 -2.861 0.0119 \*

x9 -0.2908944 0.1795418 -1.620 0.1260

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4971 on 15 degrees of freedom

Multiple R-squared: 0.9418, Adjusted R-squared: 0.9069

F-statistic: 26.98 on 9 and 15 DF, p-value: 1.106e-07

> lm.sqx2x3x8sqrtx6

Call:

lm(formula = y ~ x1 + sqx2 + sqx3 + x4 + x5 + sqrtx6 + x7 + sqx8 +

x9)

Coefficients:

(Intercept) x1 sqx2 sqx3 x4

4.2319065 0.0221917 1.7240965 0.0698773 0.2060570

x5 sqrtx6 x7 sqx8 x9

0.1473051 -0.2455904 -0.0911187 -0.0006255 -0.2908944

>

> #Question 5: Stepwise

> library(olsrr)

> Stepwise=ols\_step\_both\_p(lm.sqx2x3x8sqrtx6)

> Stepwise

Stepwise Selection Summary

------------------------------------------------------------------------------------

Added/ Adj.

Step Variable Removed R-Square R-Square C(p) AIC RMSE

------------------------------------------------------------------------------------

1 x7 addition 0.714 0.701 52.7850 69.0538 0.8903

2 x1 addition 0.860 0.847 17.1190 53.1954 0.6369

3 x5 addition 0.880 0.862 14.0670 51.4289 0.6046

------------------------------------------------------------------------------------

>

> lm.stepwise=lm(y~x1+x5+x7)

> lm.stepwise

Call:

lm(formula = y ~ x1 + x5 + x7)

Coefficients:

(Intercept) x1 x5 x7

8.55609 0.48842 0.10827 -0.07572

> summary(lm.stepwise)

Call:

lm(formula = y ~ x1 + x5 + x7)

Residuals:

Min 1Q Median 3Q Max

-1.33205 -0.30490 0.05466 0.35996 1.16505

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.55609 1.03675 8.253 4.98e-08 \*\*\*

x1 0.48842 0.21162 2.308 0.0313 \*

x5 0.10827 0.05859 1.848 0.0788 .

x7 -0.07572 0.00746 -10.150 1.49e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6046 on 21 degrees of freedom

Multiple R-squared: 0.8795, Adjusted R-squared: 0.8623

F-statistic: 51.1 on 3 and 21 DF, p-value: 7.988e-10

> residste=resid(lm.stepwise)

> qqnorm(residste)

> predictste=predict(lm.stepwise)

> plot(predictste,residste, main= "Residuals vs. Predicted Response for Stepwise Chosen Model", xlab ="Predicted Response", ylab="Residuals")