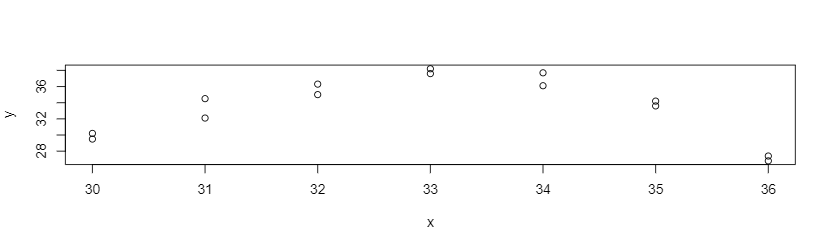
STAT641 Regression Analysis

Homework #4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source Variation | Sum of Square | Degree of Freedom | Mean Square | F0 |
| Regression | 5550.8166 | 2 | 2775.408 | 261.2417 |
| Residual | 233.7260 | 22 | 10.6239 |  |
| Total | 5784.5426 | 24 |  |  |

p-value=4.6862e-16

Because the p-value is <0.05, we reject the null hypothesis; there is enough evidence to conclude that at least one of the regressors contributes significantly to the model.

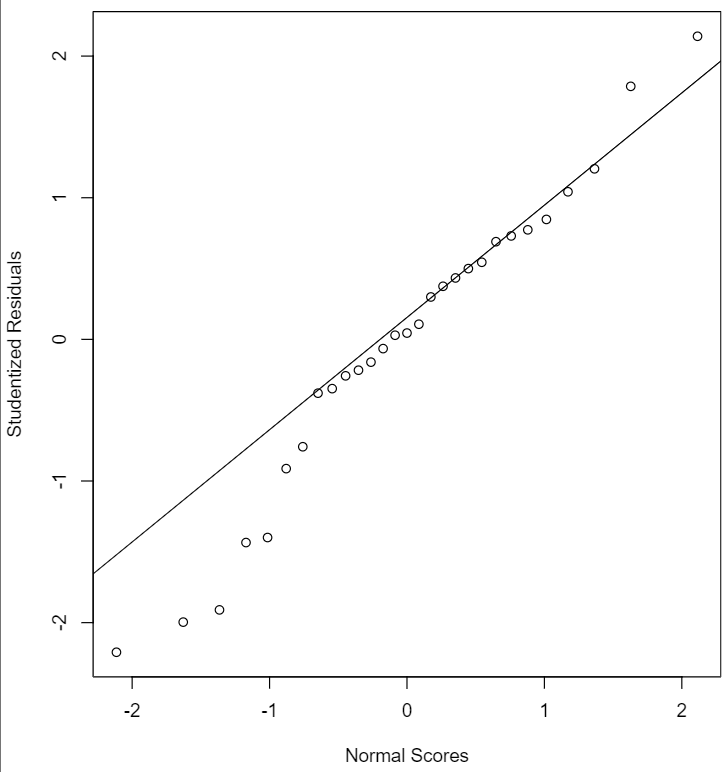
* 1. 

The plot of the data appears to be parabolic with a maximum around 33 or so.

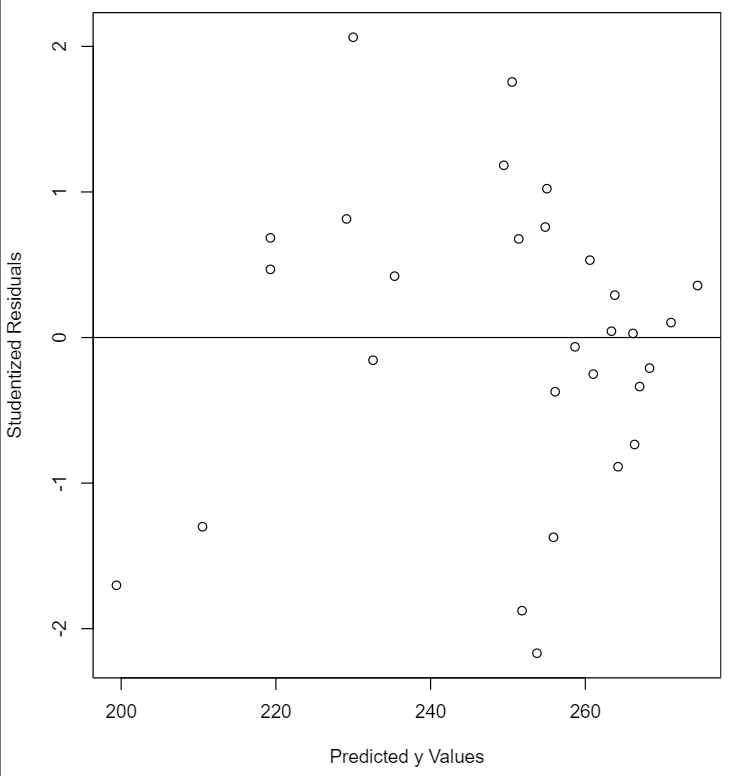
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F | p-value |
| Regression | 1 | 2.403 | 2.403 |  |  |
| Error | 12 | 183.434 | 15.2862 |  |  |
| Lack of Fit | 5 | 177.644 | 35.529 | 42.954 | 4.177e-05 |
| Pure Error | 7 | 5.790 | 0.827 |  |  |
| Total | 13 | 185.837 |  |  |  |

* 1. 

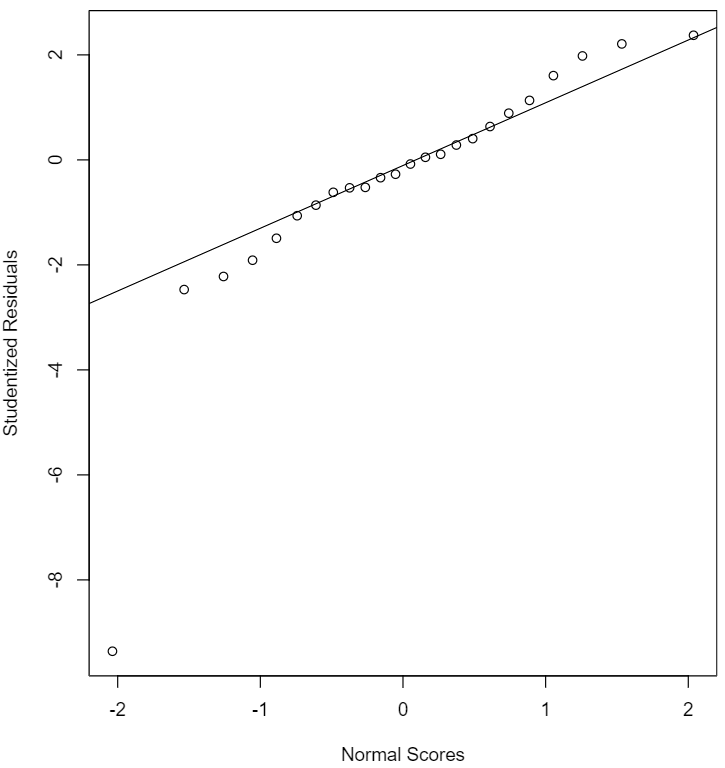
Because 0.6987>0.05, we fail to reject the null hypothesis; there is not enough evidence to conclude that pressure is a significant determiner of mileage in tires.

* 1. Based on the ANOVA table from part C, we find a p-value of 4.177e-05, which is <0.05; from this, we can conclude that the regression function is not linear.
  2. I assumed that the relationship between response and regressors is linear (already checked and refuted), the error term has zero mean, the error term has constant variance, errors are uncorrelated, and errors are normally distributed. I can check the rest of these assumptions via residual analysis, particularly graphical analysis. I could plot the residuals against their corresponding fitted values to check for nonconstant variance. I could plot the residuals in time order to check for correlated errors. I could create a normal probability plot of the residuals to visually ensure that they are normally distributed.
  3. 

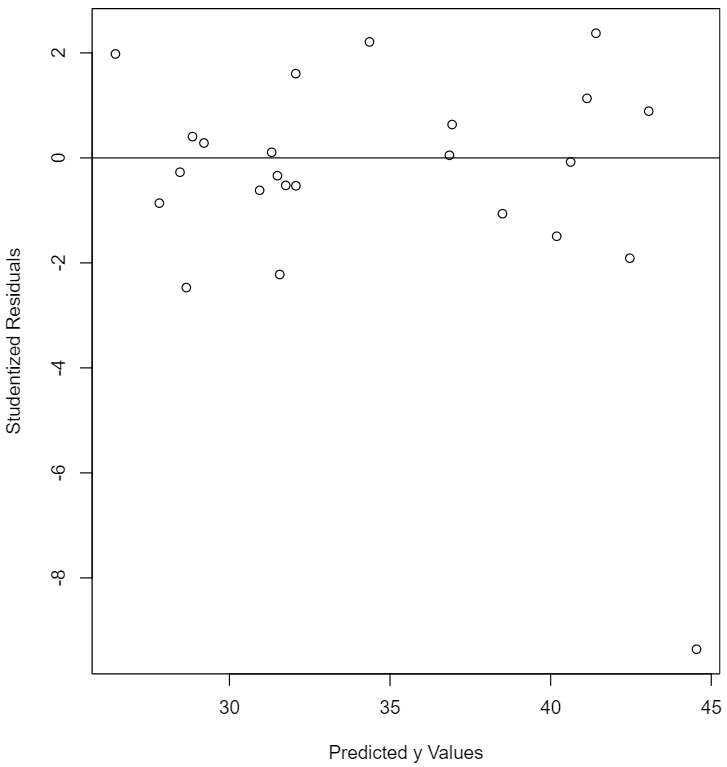
The residuals seem to be a bit heavy-tailed, which may indicate some errors that pull the least-squares fit too much in their direction. However, the nonlinearity of the plot seems relatively minor, so I would say that the residuals are basically normal.

* 1. 

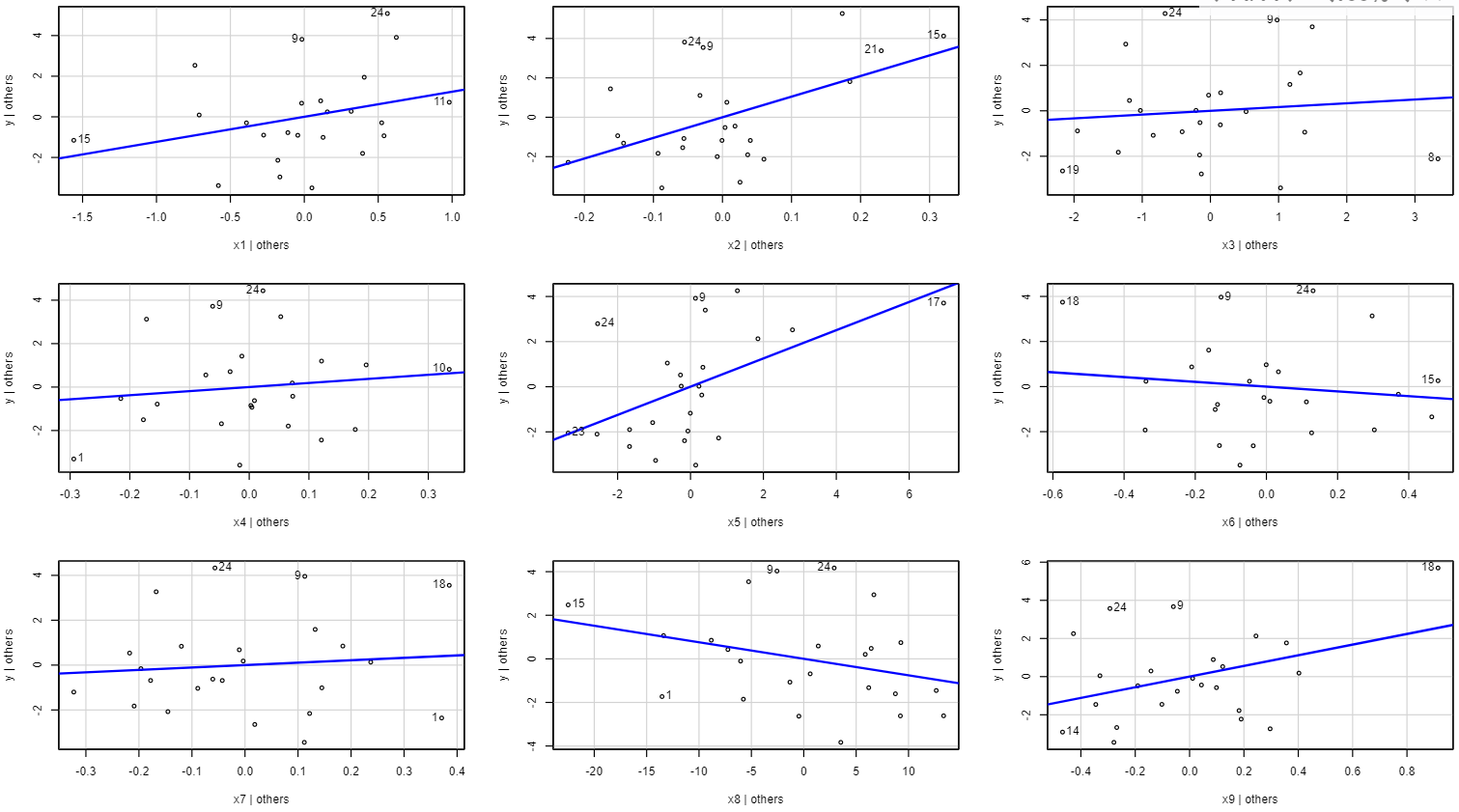
This plot suggests a funnel pattern to me, indicating nonconstant variance and that, in this case, variance is decreasing as a function of y.

* 1. 

Based on this plot, the normality of most of the residuals seems to hold. However, there is clearly one point that has a much larger error than the rest that may have an outsized impact on the model.

* 1. 

Most of the residuals plotted here fall within a band centered around 0, suggesting a relatively satisfactory model. However, we again see a point that is far removed from the rest which warrants further investigation.

* 1. 

Yes, the lack of any clear linear relationship in many of these plots, as well as the near horizontal relationship in some of them, suggests that several of the variables currently in the model are not necessary.

* 1. Computing the studentized residuals allows us to identify the outlier discussed in earlier parts of this question, as point 17 has a much larger studentized residual than the others. However, its corresponding R-student residual is much smaller and falls within the range of the other R-student residuals, suggesting that this point is not so influential on the model.