Problem-1:
Solution:

Source Variation	Sum of Square	Degree of freedom	Mean Square	F_0
Regression	5550.8166	**2**	**2775.4083**	
Residual	**233.726**	**22**	**10.6239**	**261.241**
Total	5784.5426	**24**		

Here,
n=25, k=3

Residual=Total-Regression
Residual= 5784.5426 - 5550.8166
Residual= 233.726

df_total=df_regression+df_residual
df_Total=n-1=25-1=24
df_regression=k-1=2
df_residual=n-k=24-2=22

$MS_{Regression}$= SS_R/k =5550.8166/2=2775.4083

$MS_{Residual}$= SS_R/df_residual = 233.726/22= 10.6239

F_0= $MS_{Regression}$ / $MS_{Residual}$ = 2775.4083 / 10.6239
F_0=261.241

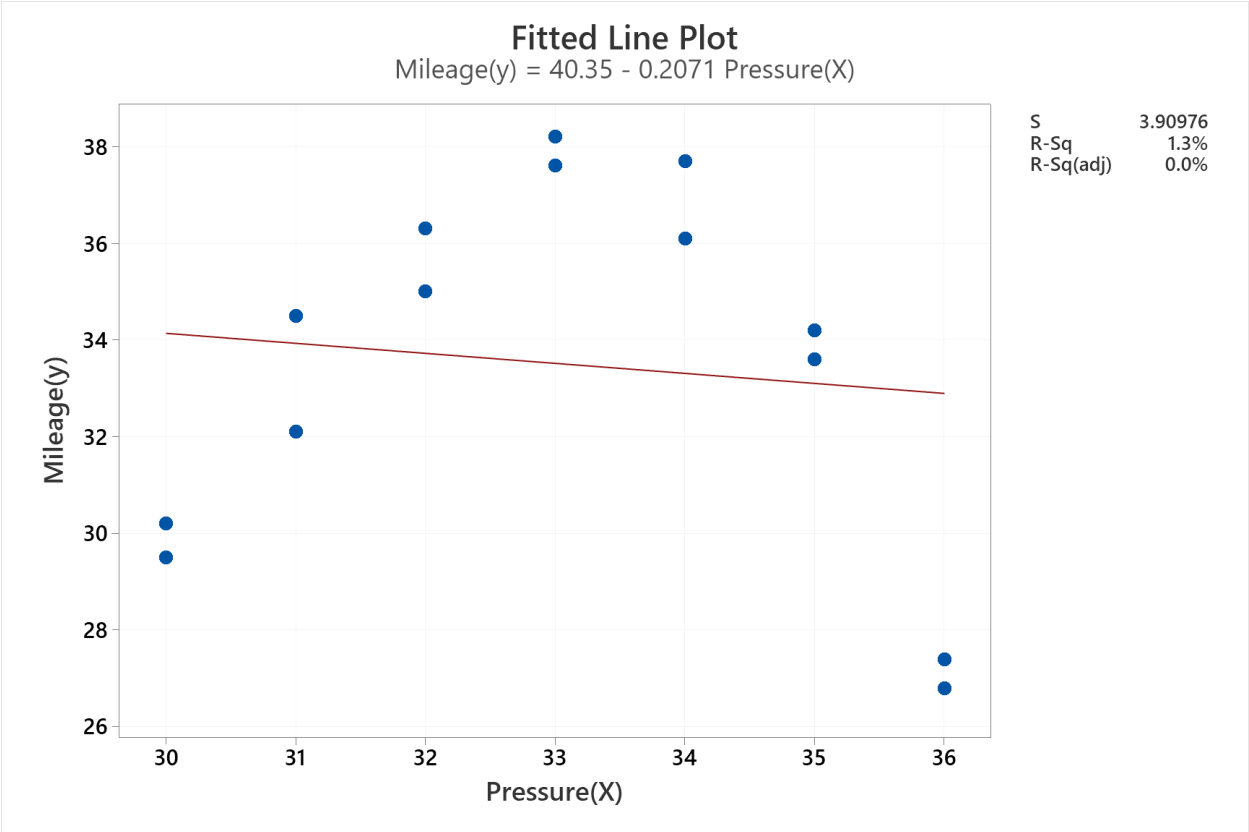
$F_{\alpha,k-1,(n-k-1)}$= $F_{0.05,2,22}$ = 3.44

Here, $F_0 > F_{\alpha,k-1,(n-k-1)}$

Conclusion: We reject the null hypothesis, H_0

Problem-2:
Solution:

a. Plot the data and describe the result.



There is colinearity between the variables.

b. Compute the least-squares regression line:

Regression Equation

Mileage(y) = 40.3 - 0.207 Pressure(X)

c. Construct an ANOVA table in the form shown below:

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	2.403	1.29%	2.403	2.4029	0.16	0.699
Pressure(X)	1	2.403	1.29%	2.403	2.4029	0.16	0.699
Error	12	183.434	98.71%	183.434	15.2862		
Lack-of-Fit	5	177.644	95.59%	177.644	35.5289	42.95	0.000
Pure Error	7	5.790	3.12%	5.790	0.8271		
Total	13	185.837	100.00%				

d. Test for significant regression:
The null hypothesis, Ho:β1=0
An alternative hypothesis, H1:β1≠0

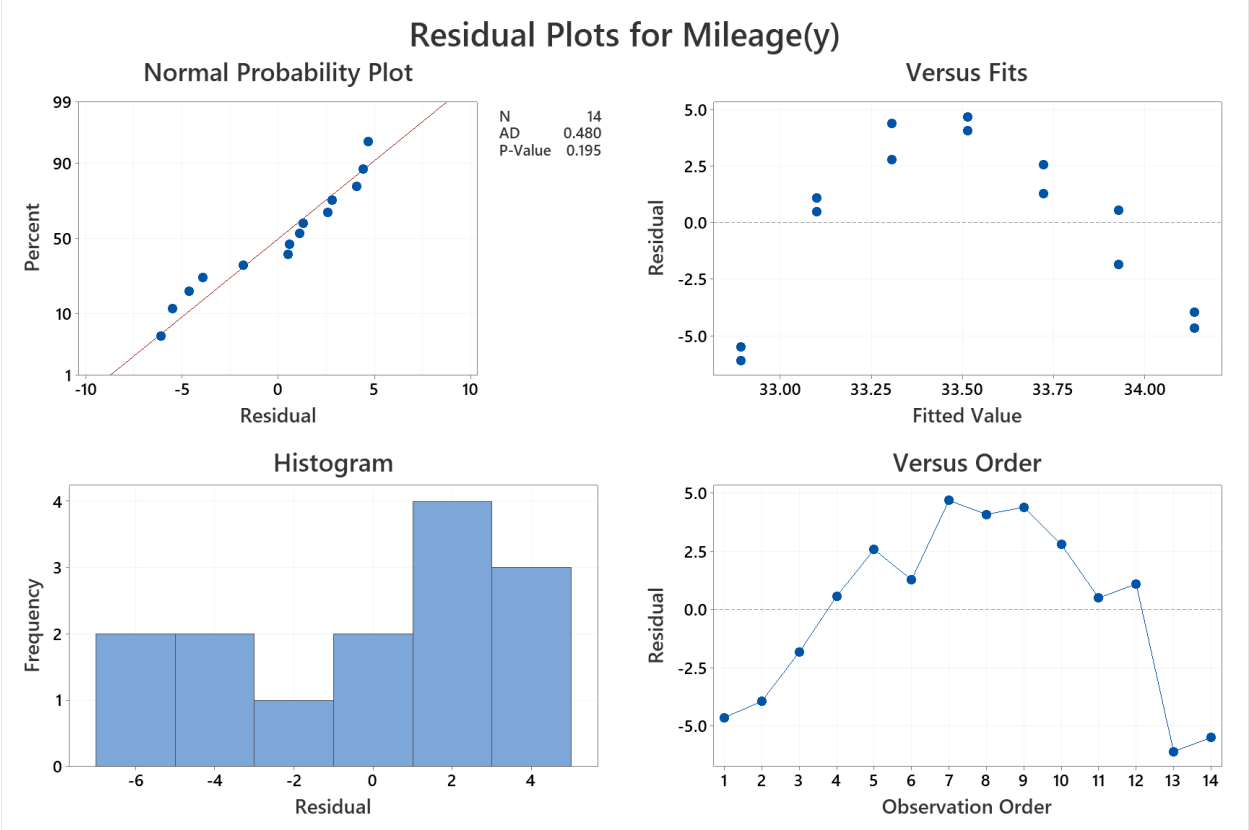
Since the p-value= 0.699, we failed to reject the null hypothesis.
Therefore, we can conclude that the regression is not significant.

e. Test that linear regression is appropriate (perform the appropriate test for lack of fit):
To perform the appropriate test for lack of fit, we specify the null and alternative hypotheses:

H_0: The linear regression is appropriate, i.e. there is no lack of fit in the model.
H_A: The linear regression is not appropriate, i.e there is a lack of fit in the model.
From the above table, we can see that the p-value is smaller than the significant level i.e 0.05.

Since the p-value is greater than 0.05, we cannot conclude that the regression is appropriate and linear.

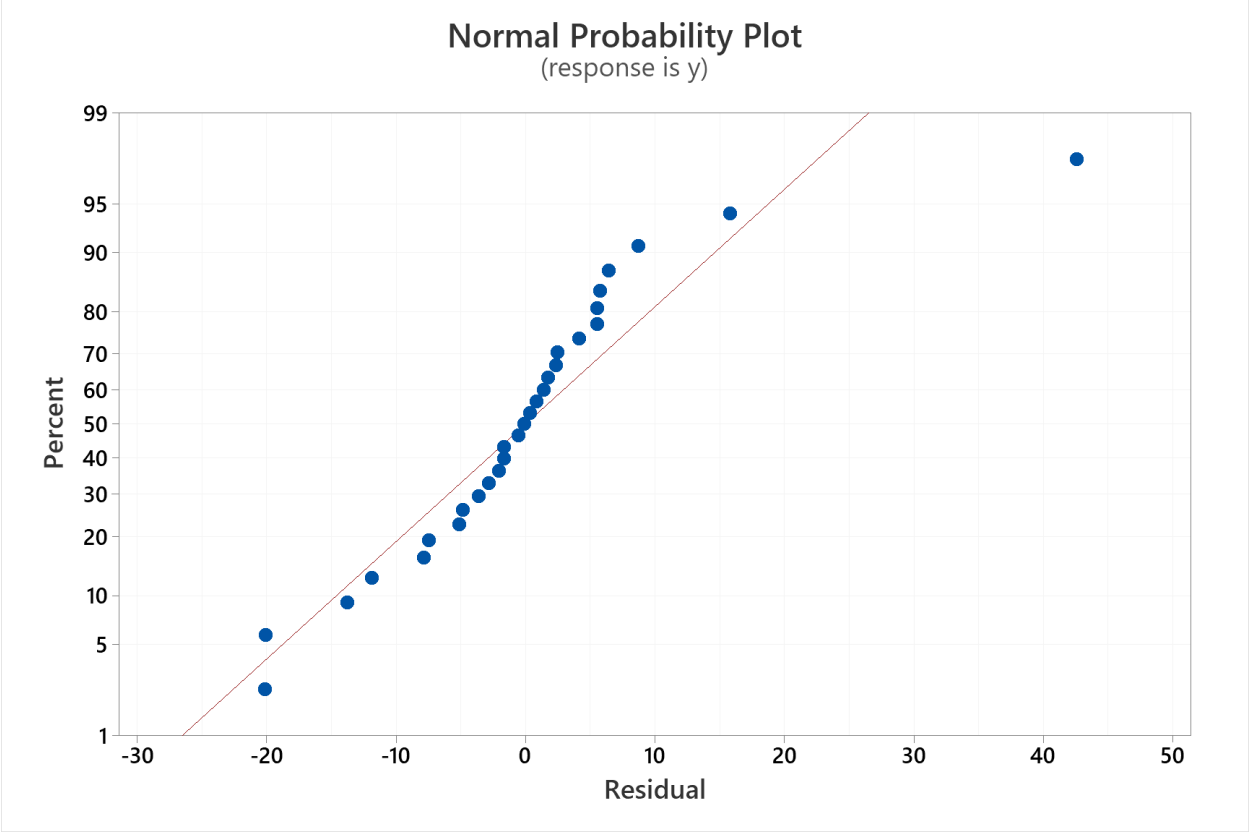
f. Did you make any assumptions in your analysis? If so, how could you check that they are valid?
Yes, we did make assumptions about the analysis.



Assumptions are violated.
The test of lack of fit assumes normality, independence is not made and is not valid.
The above graphs show that the assumptions are not valid since the p-value is fail to reject we cannot conclude that the regression is significant at the level of significance.

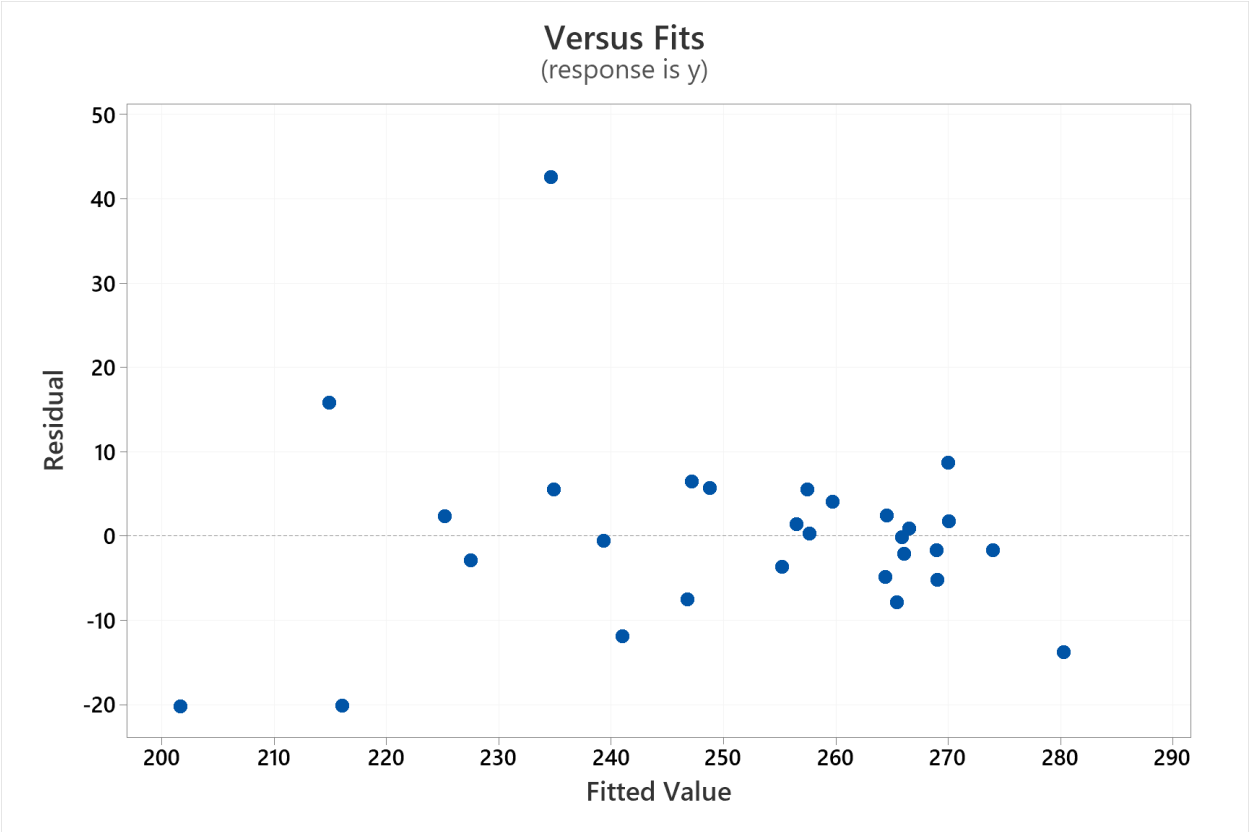
Problem-3
Solution:

a.Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?
Solution:



A normal probability graph is used to test the normality assumption. Normal probability graph residuals should approximately follow the same line. There is not seem to be any problem with the normality assumption.

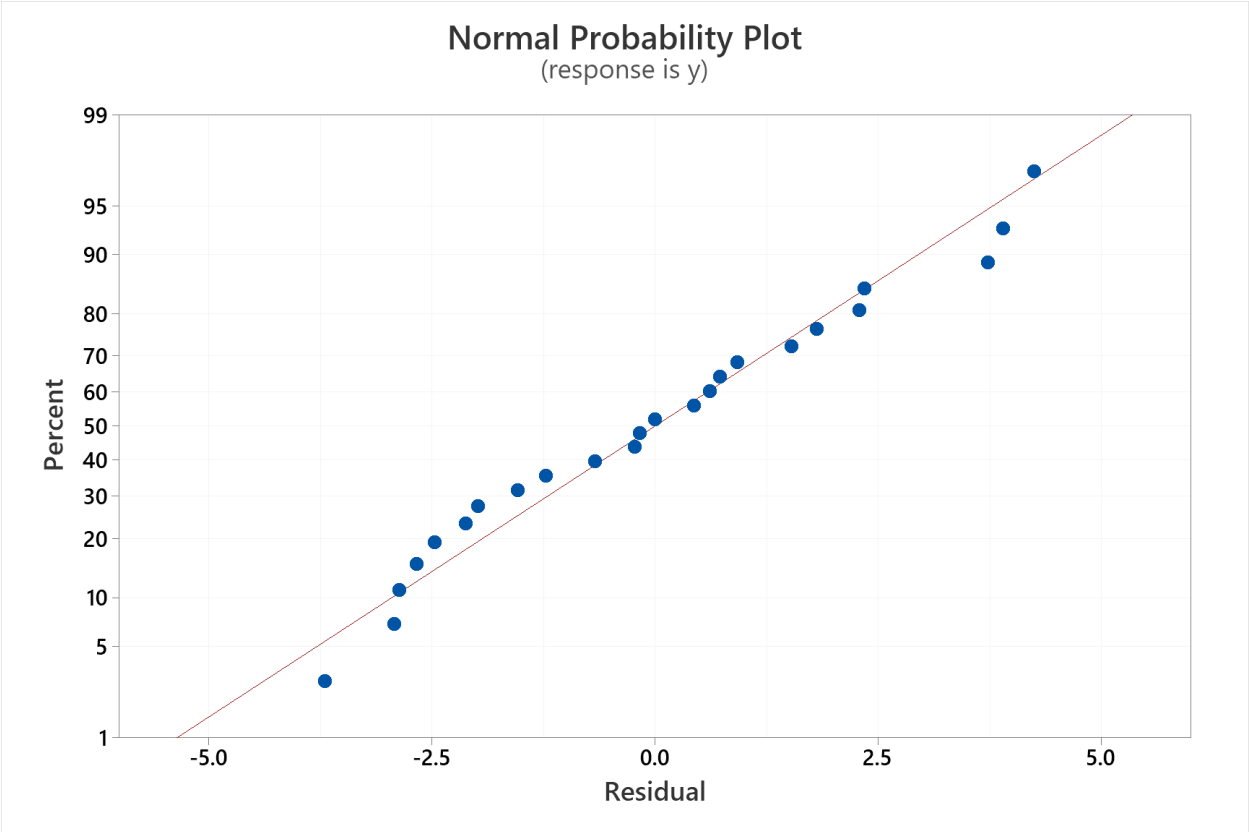
b. Construct and interpret a plot of the residuals versus the predicted response



The residual can be obtained in the horizontal band and it is used to predict the model inequalities in fitted value.

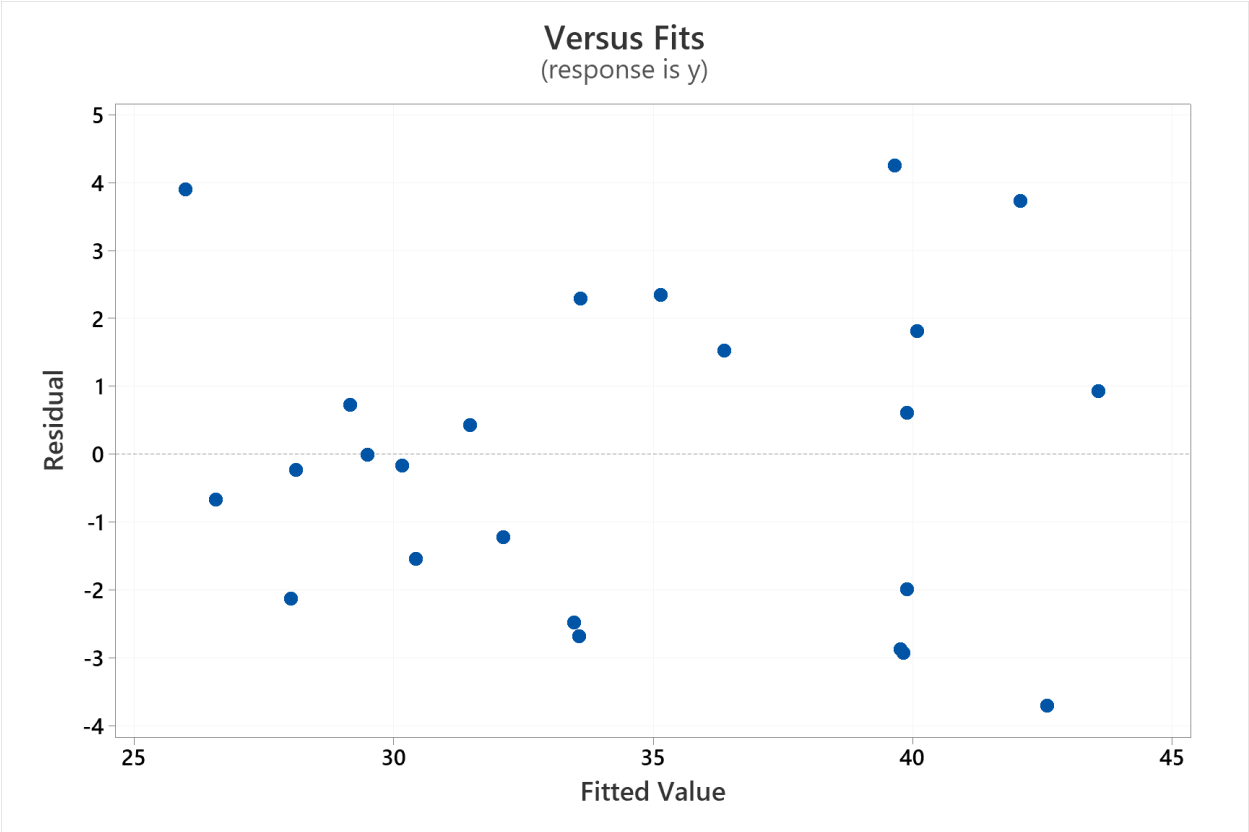
Problem:4.5
Solution:

a.Construct a normal probability plot of the residuals. Does there seem to be any problem with the normality assumption?



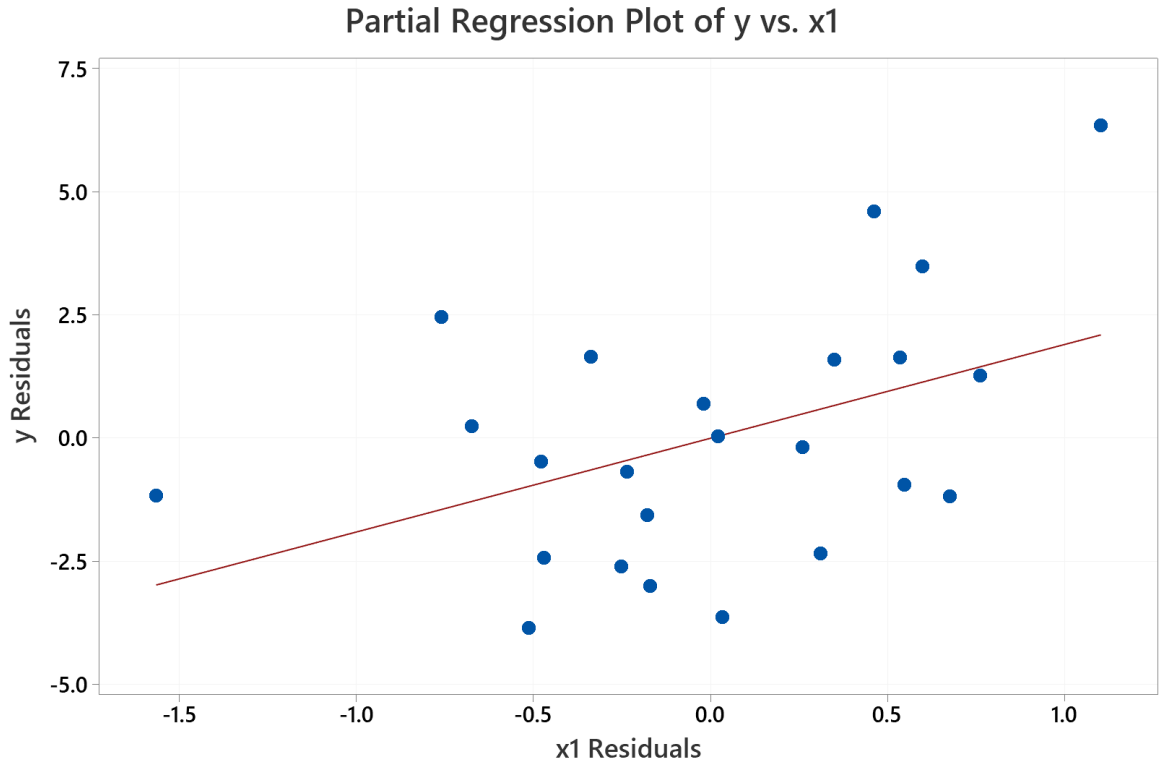
A normal probability graph is used to test the normality assumption. Normal probability graph residuals should approximately follow the same line.

b. Construct and interpret a plot of the residuals versus the predicted response.

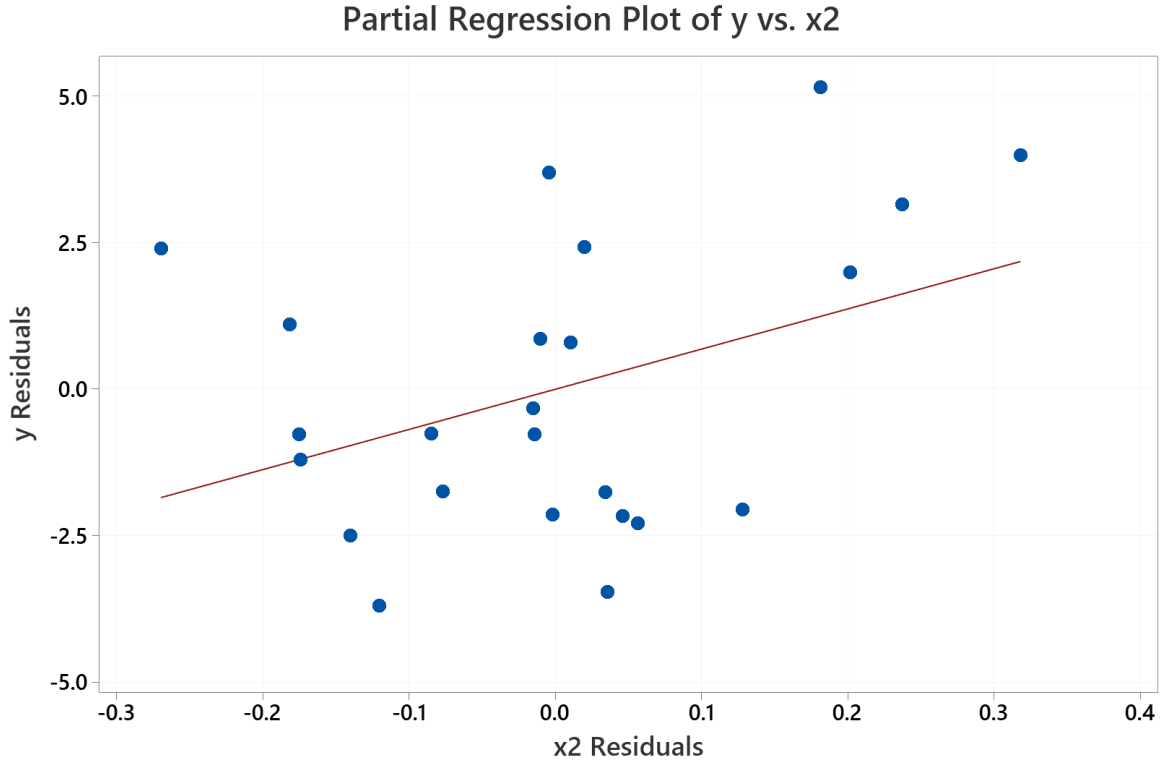


The residual can be obtained in the horizontal band and it is used to predict the model inequalities in fitted value.

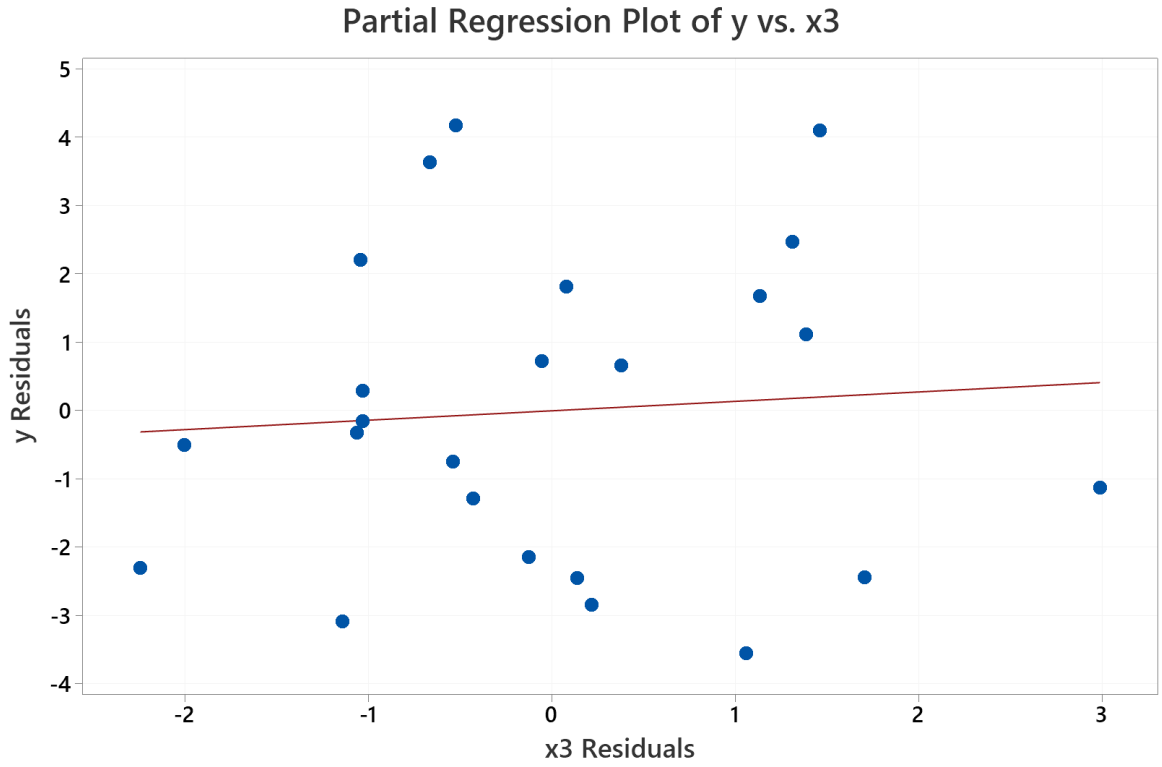
c. Construct the partial regression plots for this model. Does it seem that some variables currently in the model are not necessary?



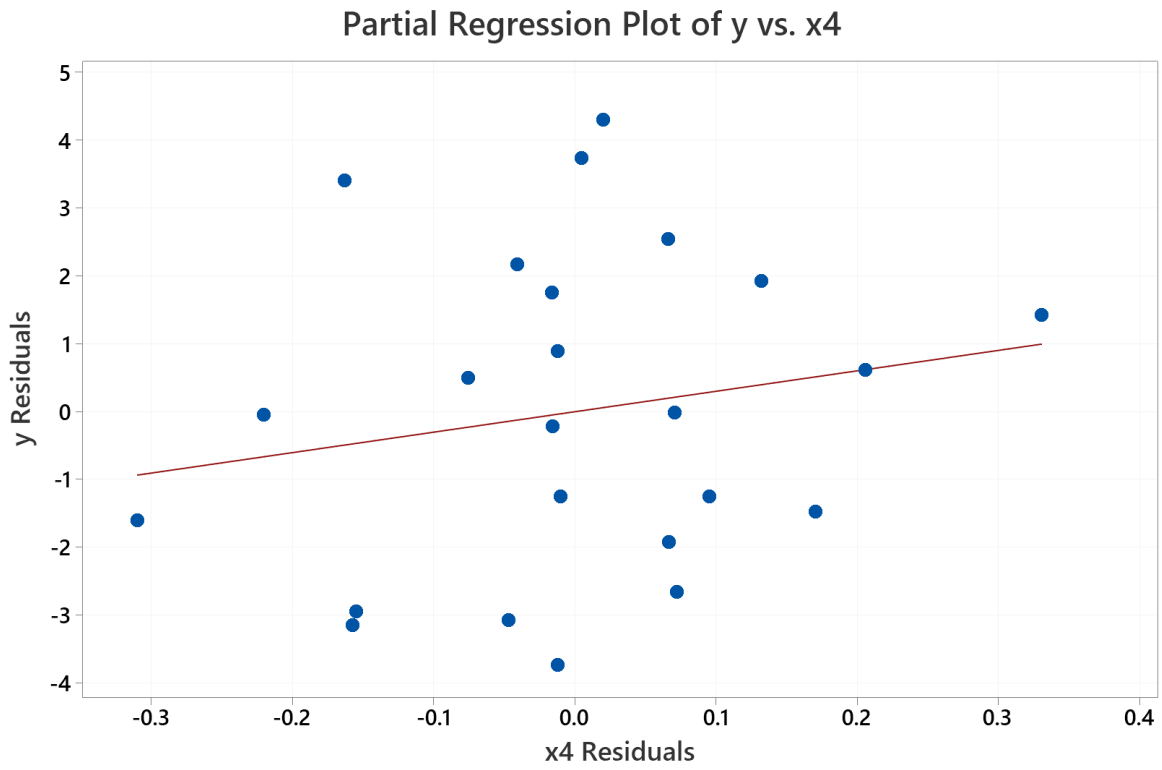
Estimated Slope of the Least Squares Line = 1.90139



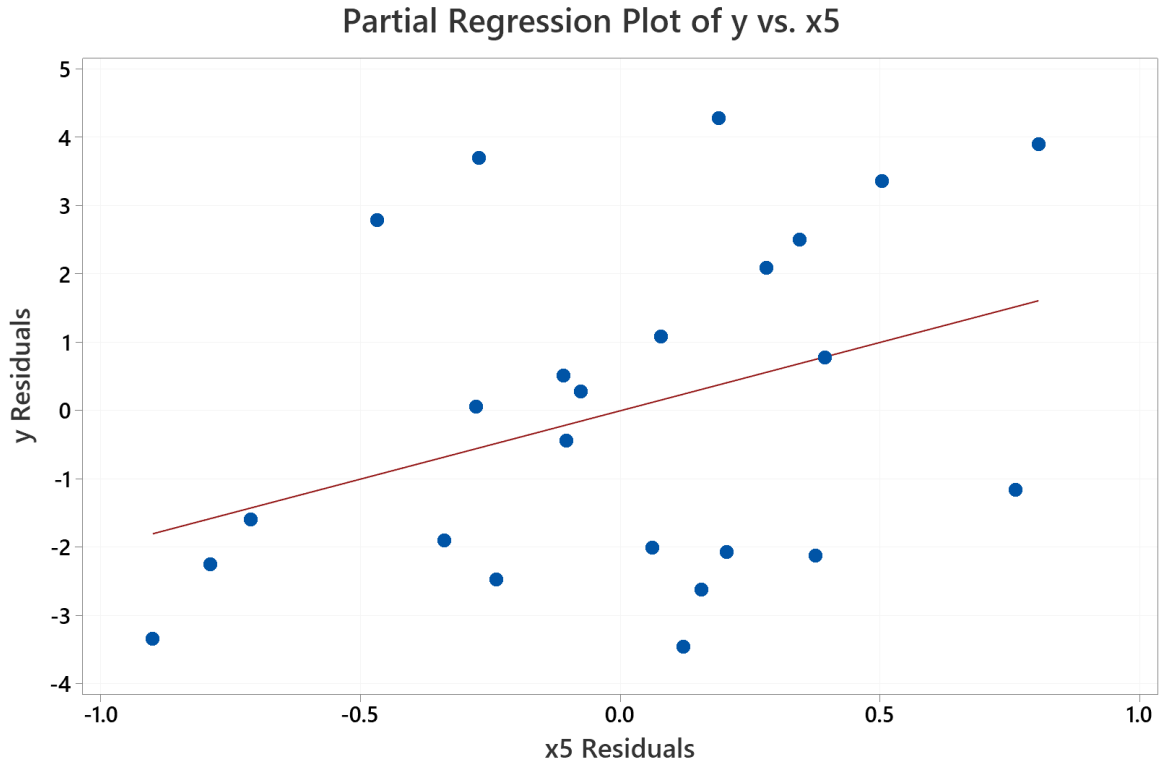
Estimated Slope of the Least Squares Line = 6.85133



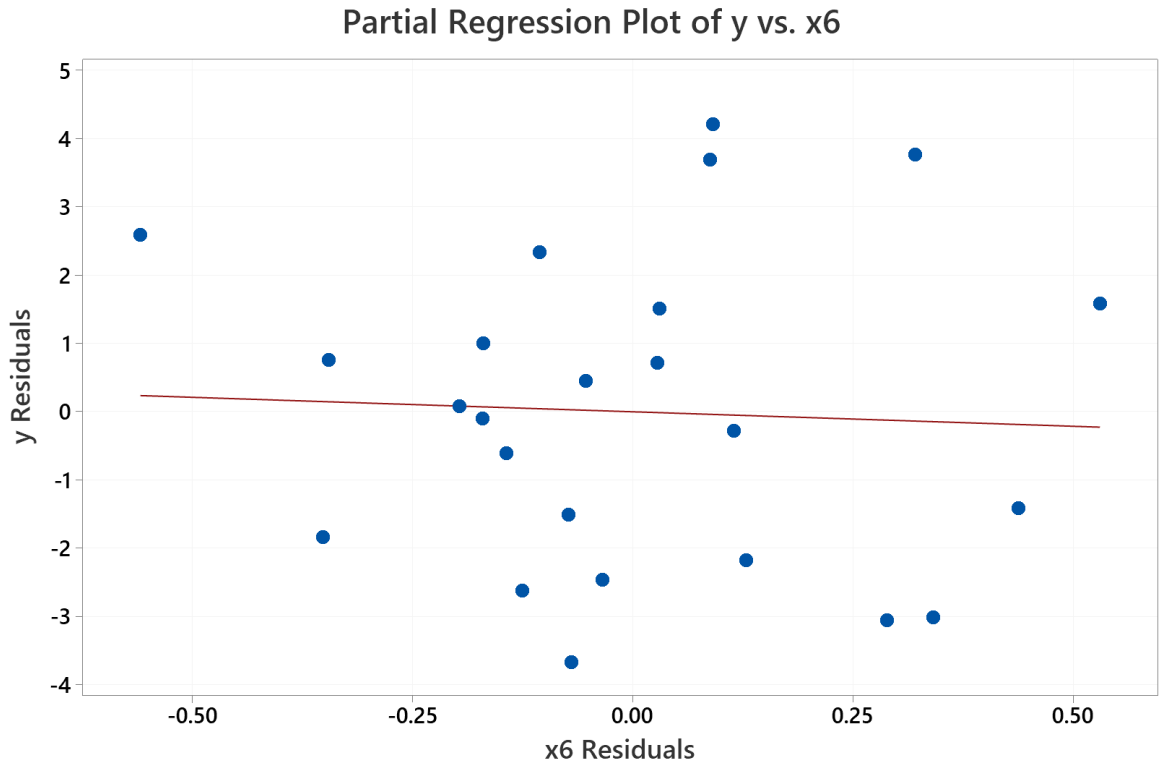
Estimated Slope of the Least Squares Line = 0.137766



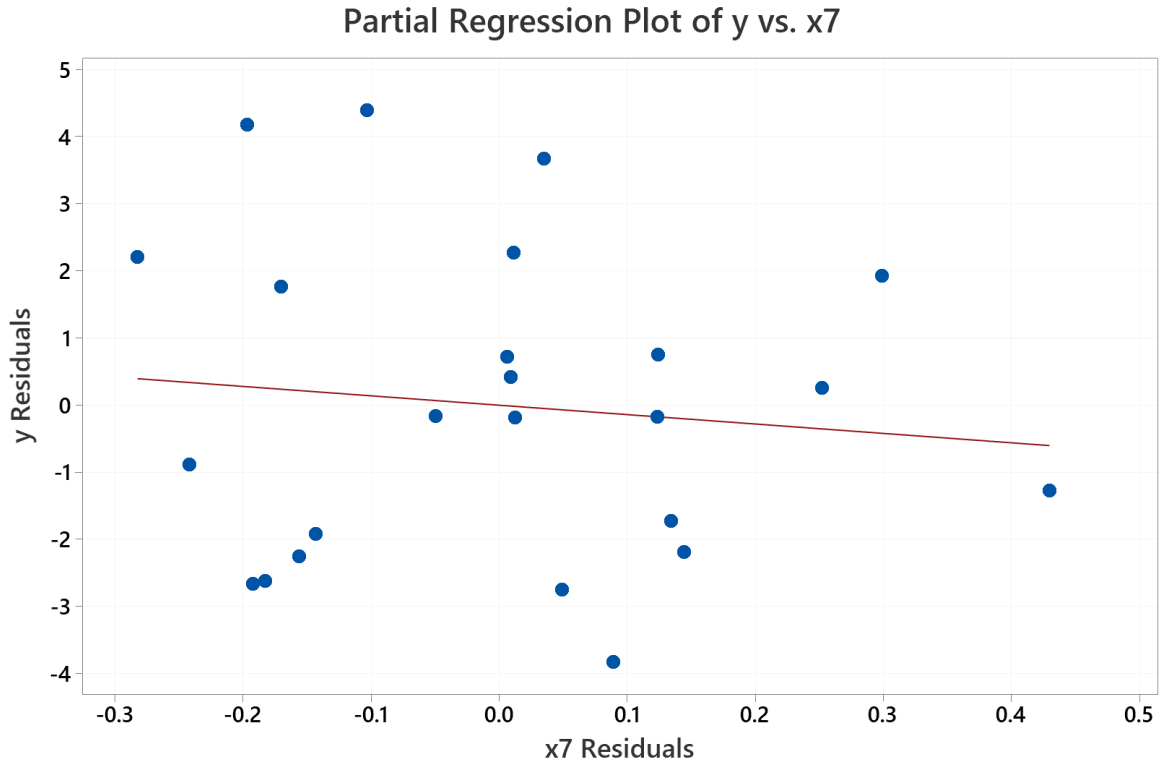
Estimated Slope of the Least Squares Line = 3.01209



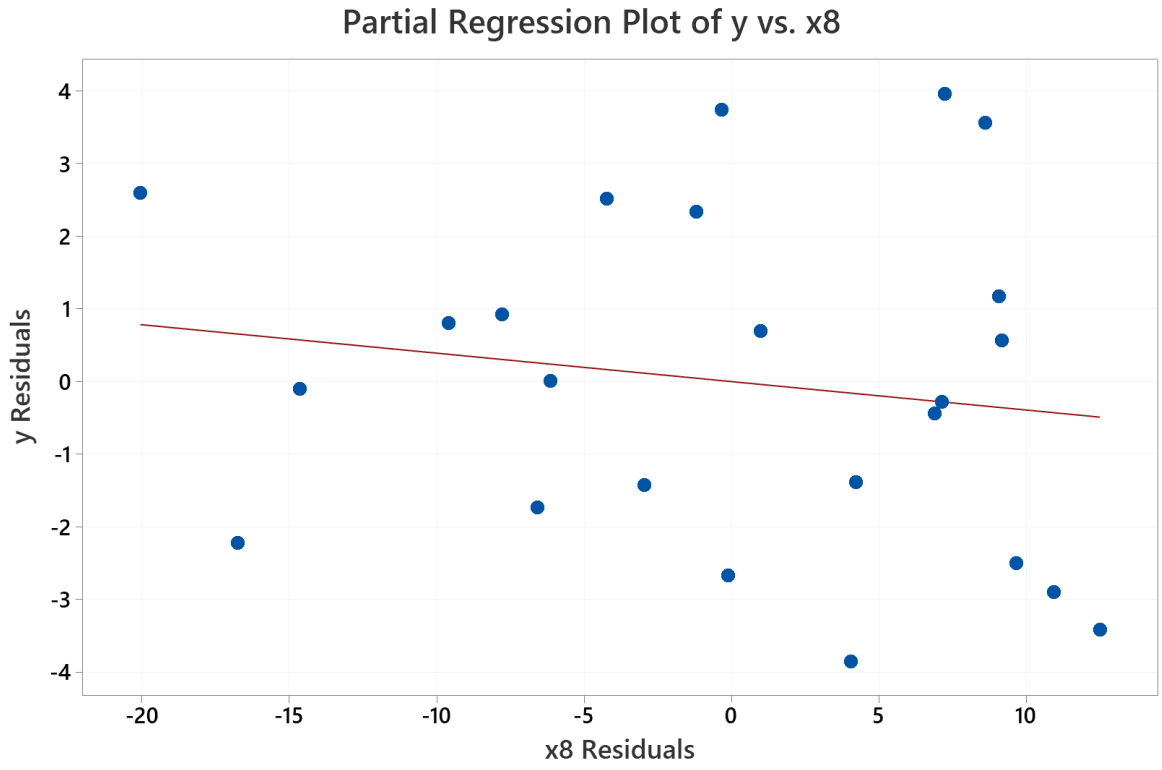
Estimated Slope of the Least Squares Line = 2.00009



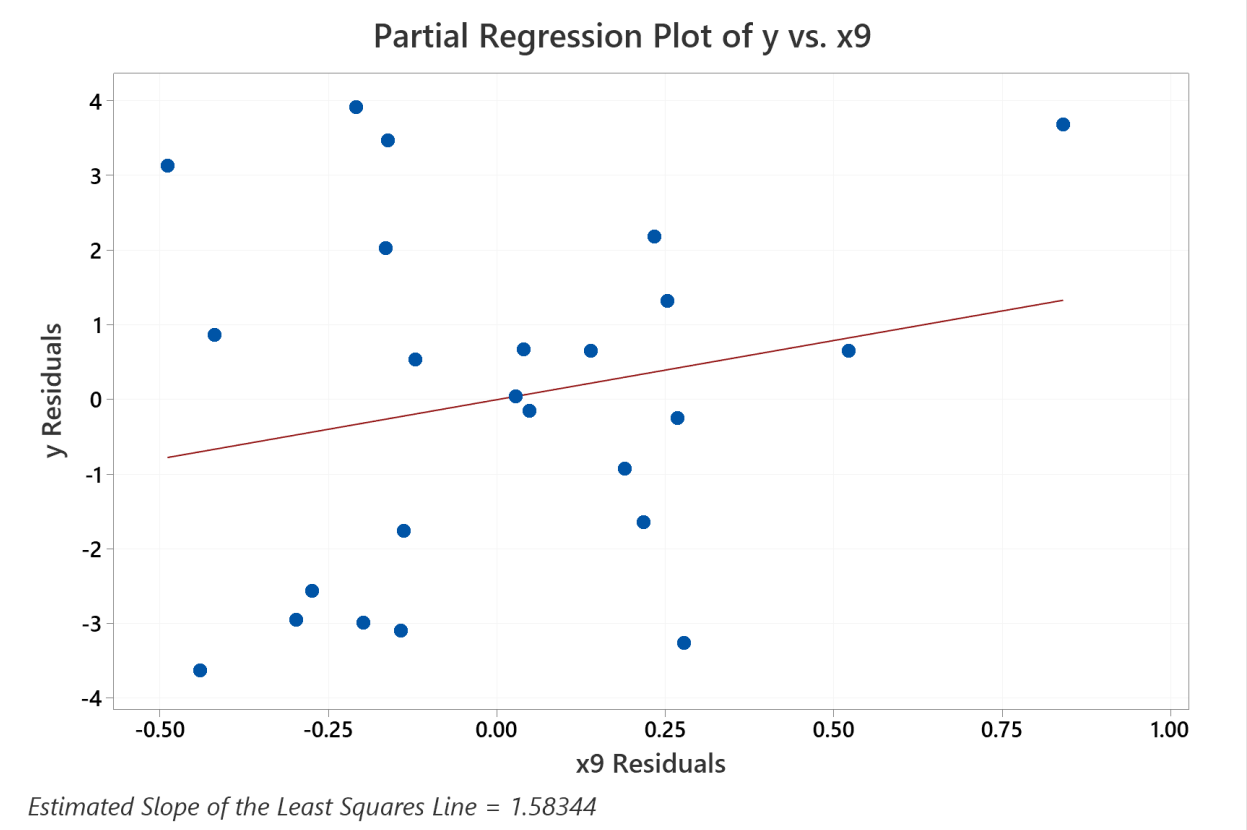
Estimated Slope of the Least Squares Line = -0.42395



Estimated Slope of the Least Squares Line = -1.39848



Estimated Slope of the Least Squares Line = -0.03911



Yes, it seems that some variables currently in the model are not necessary. Variables in the models are far from each other and sparse.

d. Compute the studentized residuals and the R - student residuals for this model. What information is conveyed by these scaled residuals?

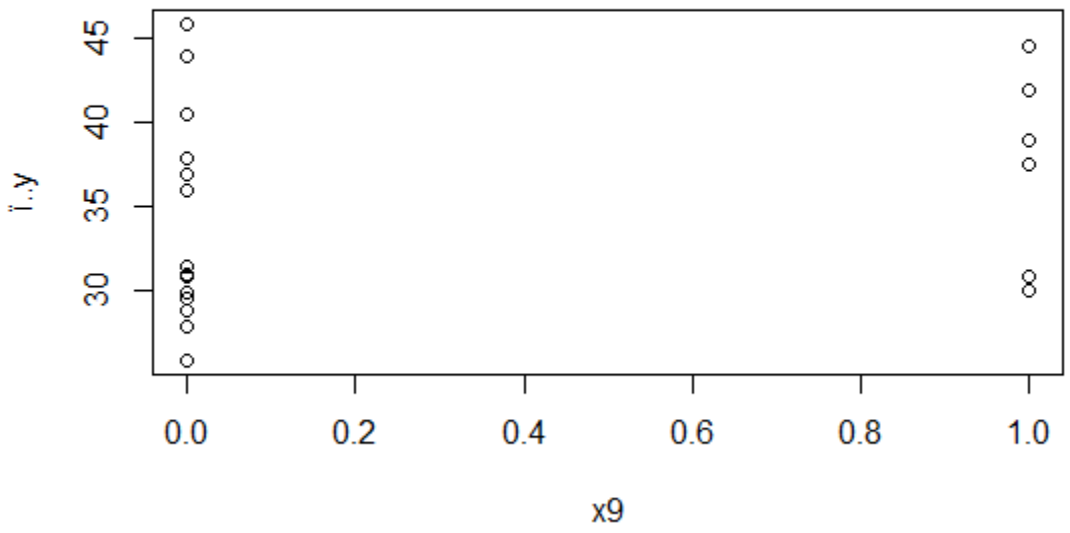
Coefficients:

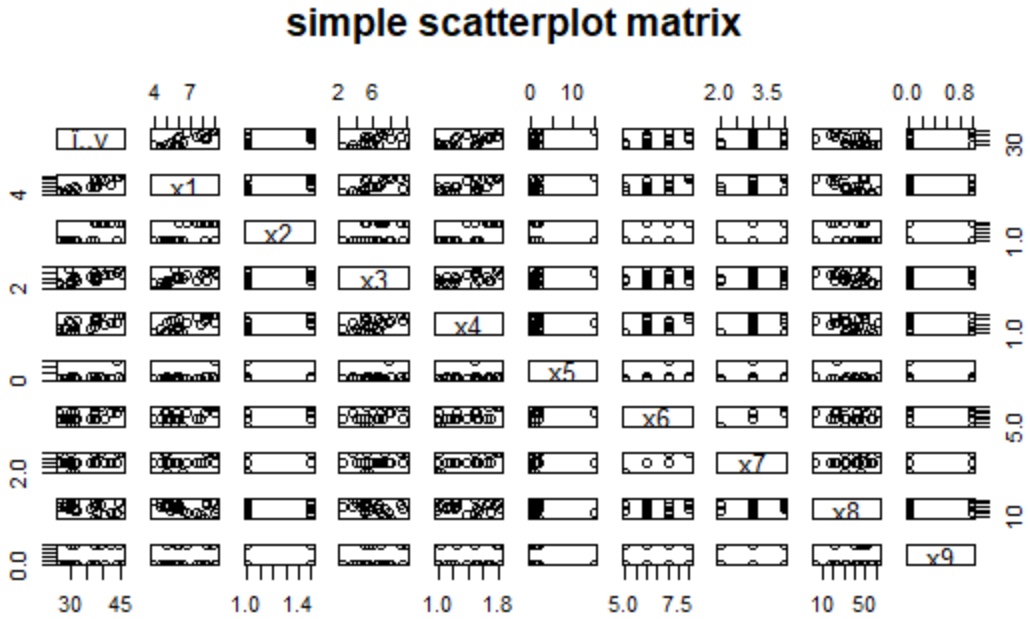
(Intercept)	x1	x2	x3	x4	x5
15.30351	1.23514	10.47104	0.16571	1.88436	0.62622
x6	**x7**	**x8**	**x9**		
-1.05667	1.07720	-0.07584	2.79492		

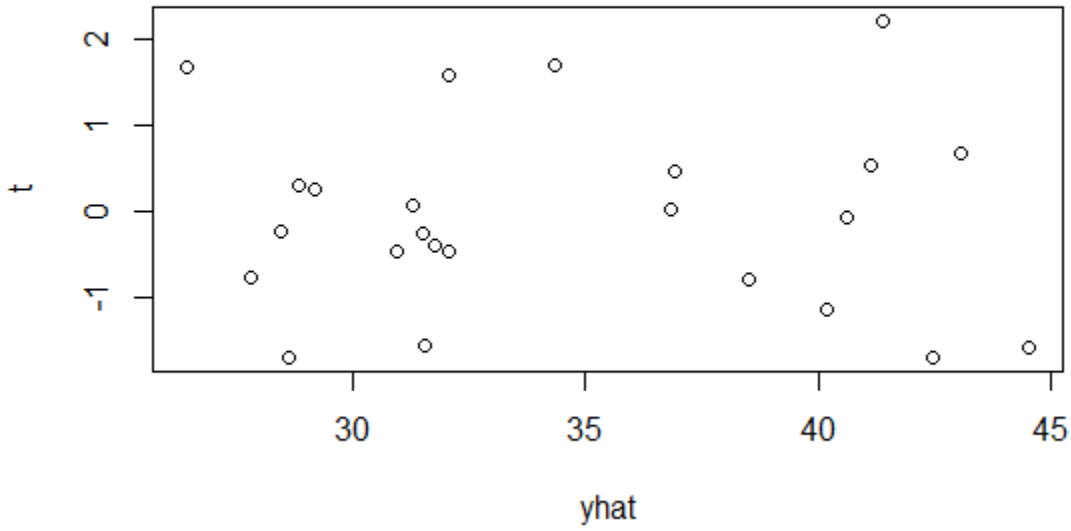
```
> res <- resid(delivery.lm)
> res
          1           2           3           4           5           6
-2.75762720  0.64968017 -0.56113936 -1.91876127  0.69272682  3.44855445
          7           8           9          10          11          12
-0.85381527 -2.66693754  3.83486999  0.18499149 -0.49450185 -1.16923815
         13          14          15          16          17          18
-0.94010395 -1.60057761  0.76991382 -0.12284063 -0.64127891  3.14074144
         19          20          21          22          23          24
-2.28825279  1.44737067  0.96783837 -3.56410671  0.05183914  4.39065487
> 
```

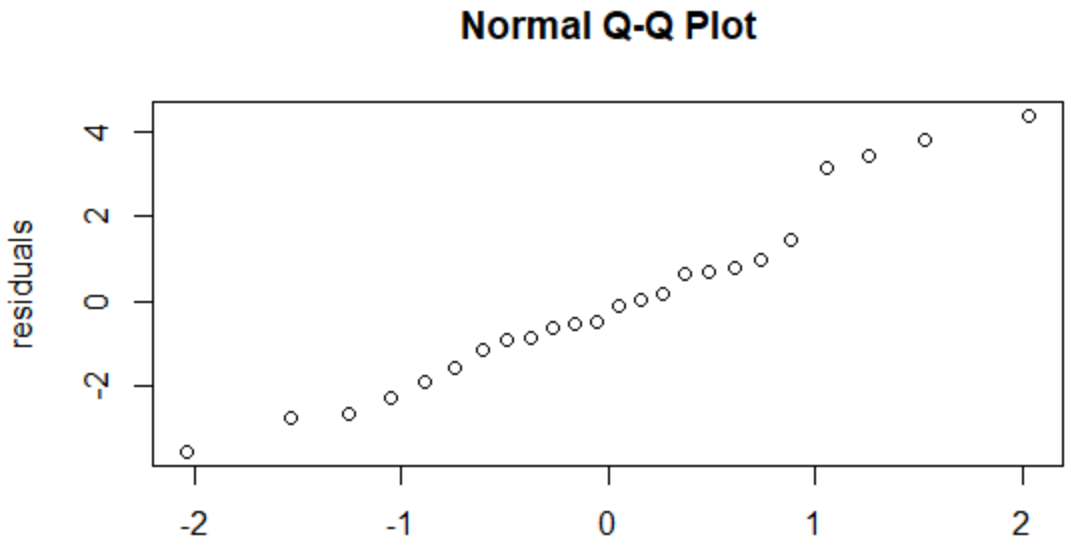
```
> # calculate the residuals
> # the regular residuals (= yi - yi_hat)
> res
          1           2           3           4           5           6
-2.75762720  0.64968017 -0.56113936 -1.91876127  0.69272682  3.44855445
          7           8           9          10          11          12
-0.85381527 -2.66693754  3.83486999  0.18499149 -0.49450185 -1.16923815
         13          14          15          16          17          18
-0.94010395 -1.60057761  0.76991382 -0.12284063 -0.64127891  3.14074144
         19          20          21          22          23          24
-2.28825279  1.44737067  0.96783837 -3.56410671  0.05183914  4.39065487
>
```

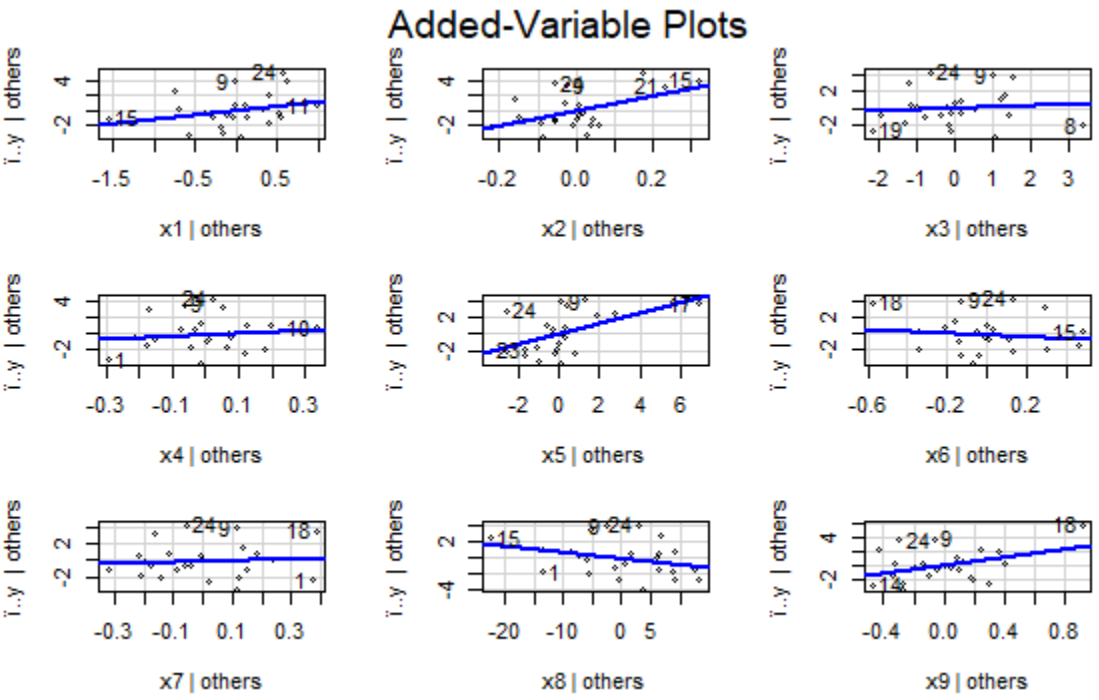
```
> stud_res
          1           2           3           4           5           6           7           8
-2.47025157  0.40479567 -0.27297949 -0.86222346  0.28274068  1.97965101 -0.52379561 -2.22092221
          9          10          11          12          13          14          15          16
 1.60492239  0.10671003 -0.33934684 -0.53228946 -0.61777185 -1.06366833  1.13319712 -0.07885813
         17          18          19          20          21          22          23          24
-9.35889236  2.20904324 -1.49302043  0.88952242  0.63529627 -1.91090850  0.04934293  2.37347711
> Rstudent_res
          1           2           3           4           5           6           7           8
-1.68493310  0.30097310 -0.22936622 -0.76919615  0.25950709  1.68717804 -0.39344627 -1.54922822
          9          10          11          12          13          14          15          16
 1.58495881  0.08219462 -0.24011477 -0.46522494 -0.44915686 -0.78081046  0.55267391 -0.05757073
         17          18          19          20          21          22          23          24
-1.56138739  1.70388946 -1.13315932  0.67515593  0.46236348 -1.68472951  0.02958067  2.21616580
```




simple scatterplot matrix
y
x1
x2
x3
x4
x5
x6
x7
x8
x9


t
yhat





We can see that the value in studentized residual for observation 17 and 1 far from other values which could be the outlier value.