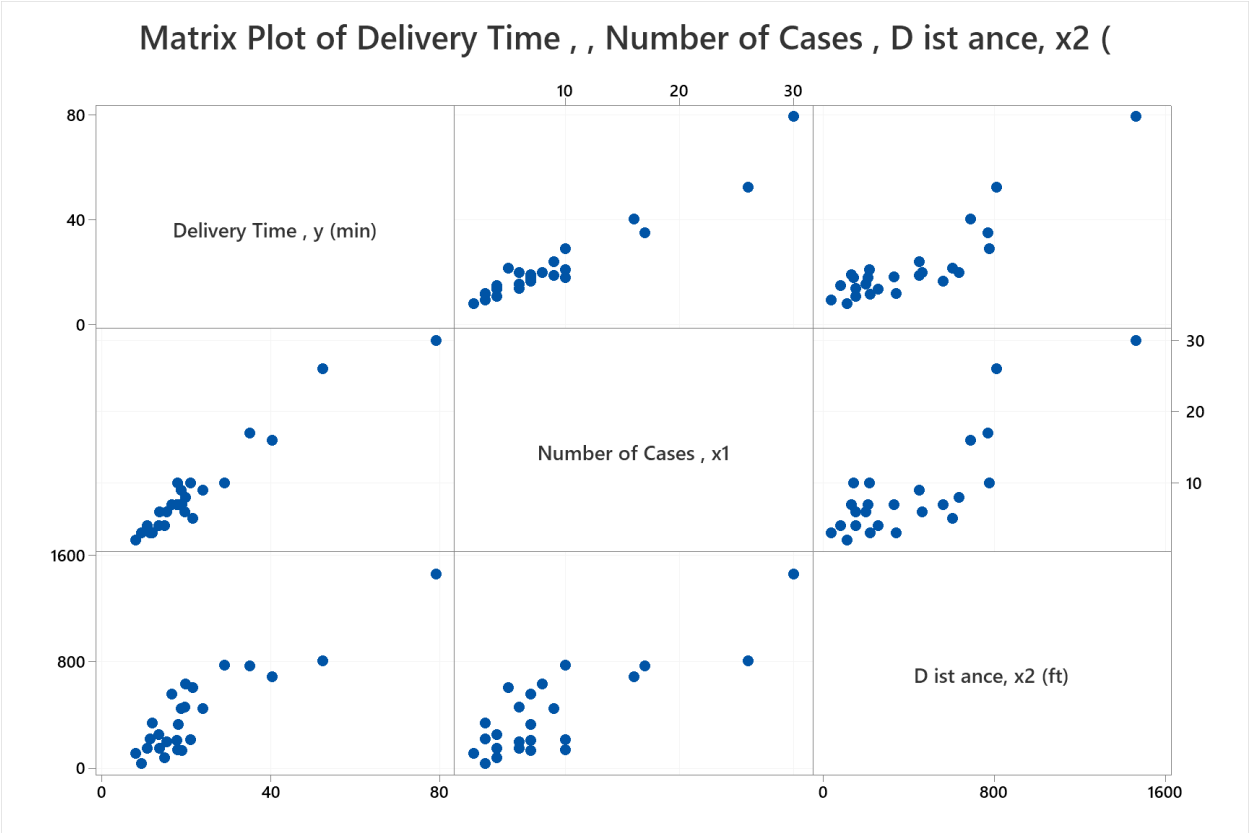
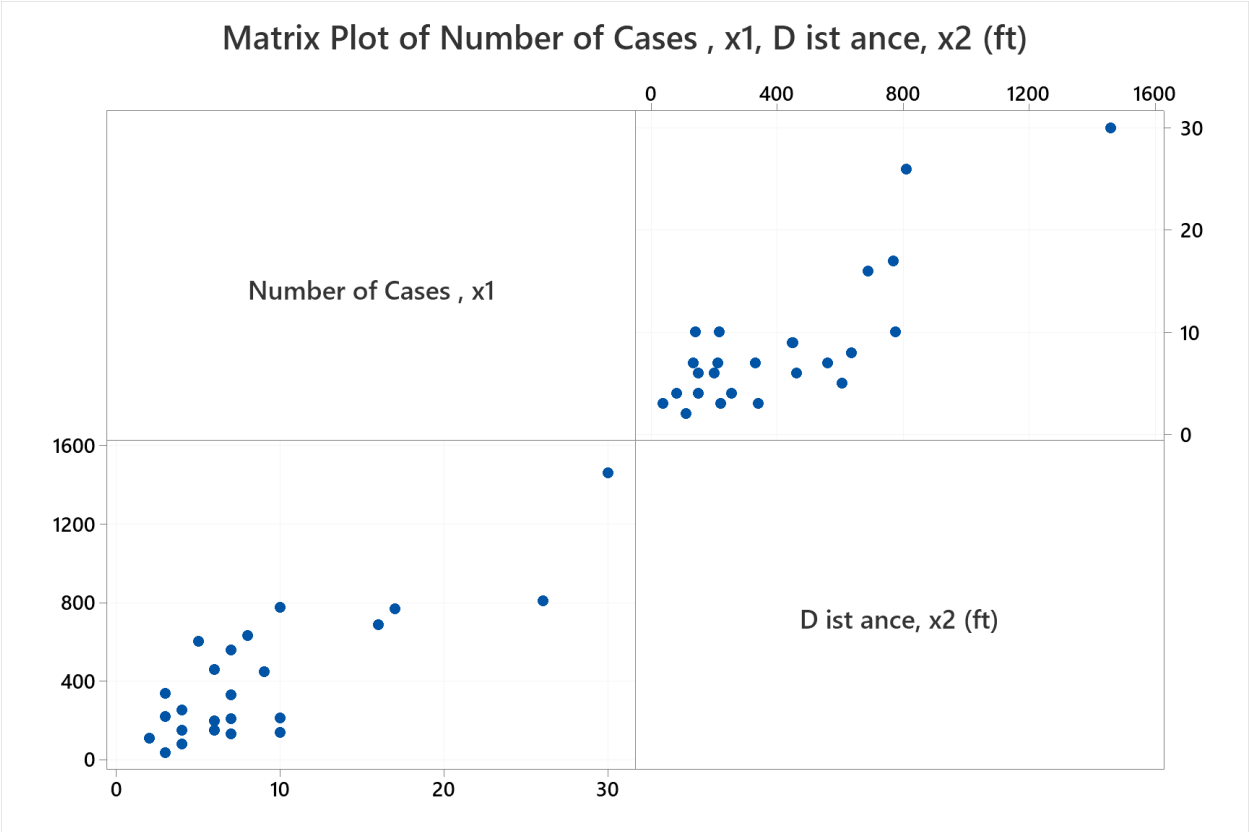
Problem-1

a.Create a matrix plot of the predictors, comment





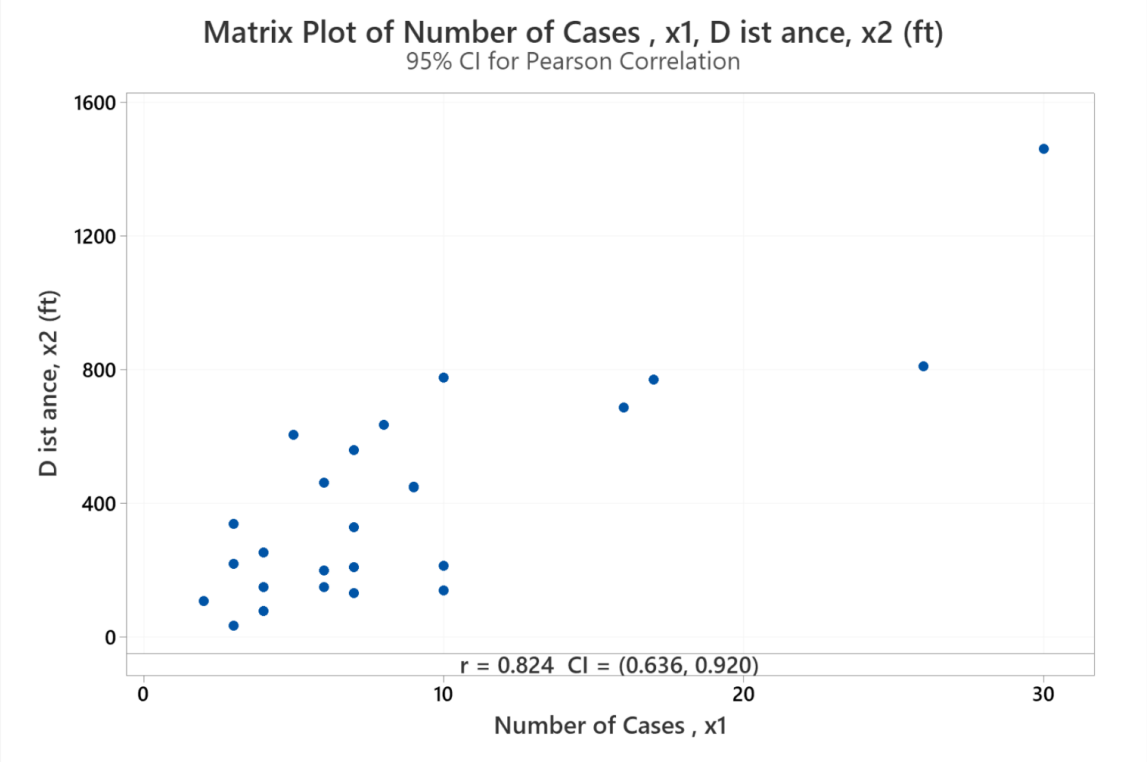
From the above graph, we can conclude that there is a linear co-relationship between two predictors and the strength of the relationship.

We also observe a strong linear relationship between the delivery time and the number of cases and delivery time and distance.

b.Create the correlation matrix for the predictors, comment

SHEET1

Correlation: Number of Cases , x1, D ist ance, x2 (ft)



Method

Correlation type	Pearson
Number of rows used	25

ρ: pairwise Pearson correlation

Correlations

	Number of Cases , x1
D ist ance, x2 (ft)	0.824

Pairwise Pearson Correlations

Sample 1	Sample 2	N	Correlation	95% CI for ρ	P-Value
D ist ance, x2 (ft)	Number of Cases , x1	25	0.824	(0.636, 0.920)	0.000

From the above correlation plot and values, we can see that the strength of the relationship between two variables is high i.e 0.824 (82%) and both variables impact each other in a positive manner.

c.Determine if there are any collinearity issues with this data using the methods that we discussed and comment.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.34	1.10	2.13	0.044	
Number of Cases , x1	1.616	0.171	9.46	0.000	3.12
D ist ance, x2 (ft)	0.01438	0.00361	3.98	0.001	3.12

From the above table and the graph we can conclude that the p-value is approximately 0 and VIF is less than 5. We can conclude that there exists no multicollinearity in the predicted variables.

d. If multicollinearity exists, apply ridge regression to the data. Use the ridge trace to select the appropriate value of

Checking for multicollinearity

W'W=

Matrix CORR

1.00000 0.82422
0.82422 1.00000

(w'w)^-1=

Matrix M4

3.11847 -2.57029
-2.57029 3.11847

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.34	1.10	2.13	0.044	
Number of Cases , x1	1.616	0.171	9.46	0.000	3.12
D ist ance, x2 (ft)	0.01438	0.00361	3.98	0.001	3.12

From the above results we can see that the VIF is 3.12.

The above values indicate that multicollinearity is not a problem in these data and that we don't need to apply bridge regression.

Problem 2:
analyze the wine quality of young red wines data in Table B.19 for multicollinearity

SHEET1(W1)

Regression Analysis: Y versus X1, X2, X3, X4, X5, X6, x7, x8, x9, X10

The following terms cannot be estimated and were removed:
x7, X10

Regression Equation

Y = -12.2 - 0.846 X1 + 7.42 X2 + 0.01046 X3 - 1.95 X4 + 4.90 X5 - 1.43 X6 - 11.43 x8
- 0.108 x9

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-12.2	14.6	-0.84	0.412	
X1	-0.846	0.586	-1.44	0.162	1.97
X2	7.42	3.51	2.11	0.046	4.09
X3	0.01046	0.00857	1.22	0.235	4.51
X4	-1.95	2.22	-0.88	0.390	603.52
X5	4.90	3.22	1.52	0.142	511.87
X6	-1.43	1.81	-0.79	0.437	33.32
x8	-11.43	7.88	-1.45	0.161	7.93
x9	-0.108	0.220	-0.49	0.629	36.17

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.17136	67.53%	56.24%	16.69%

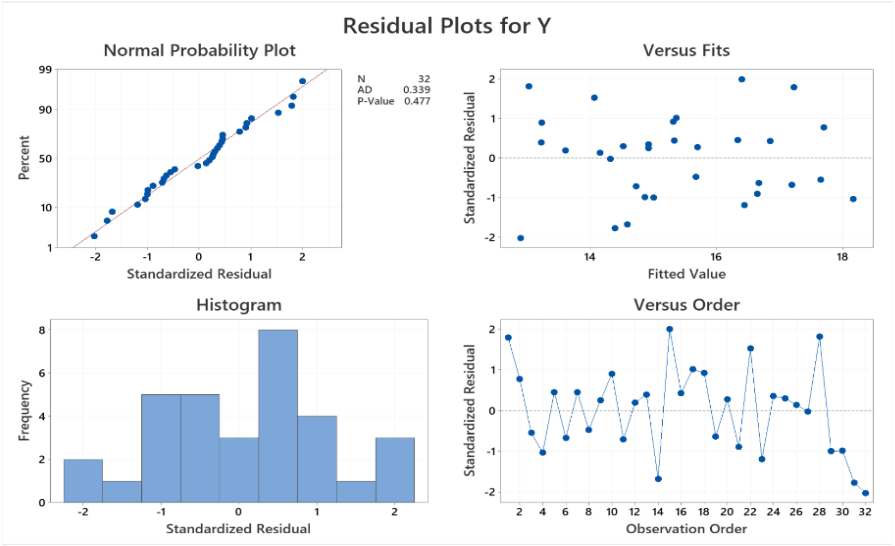
Analysis of Variance

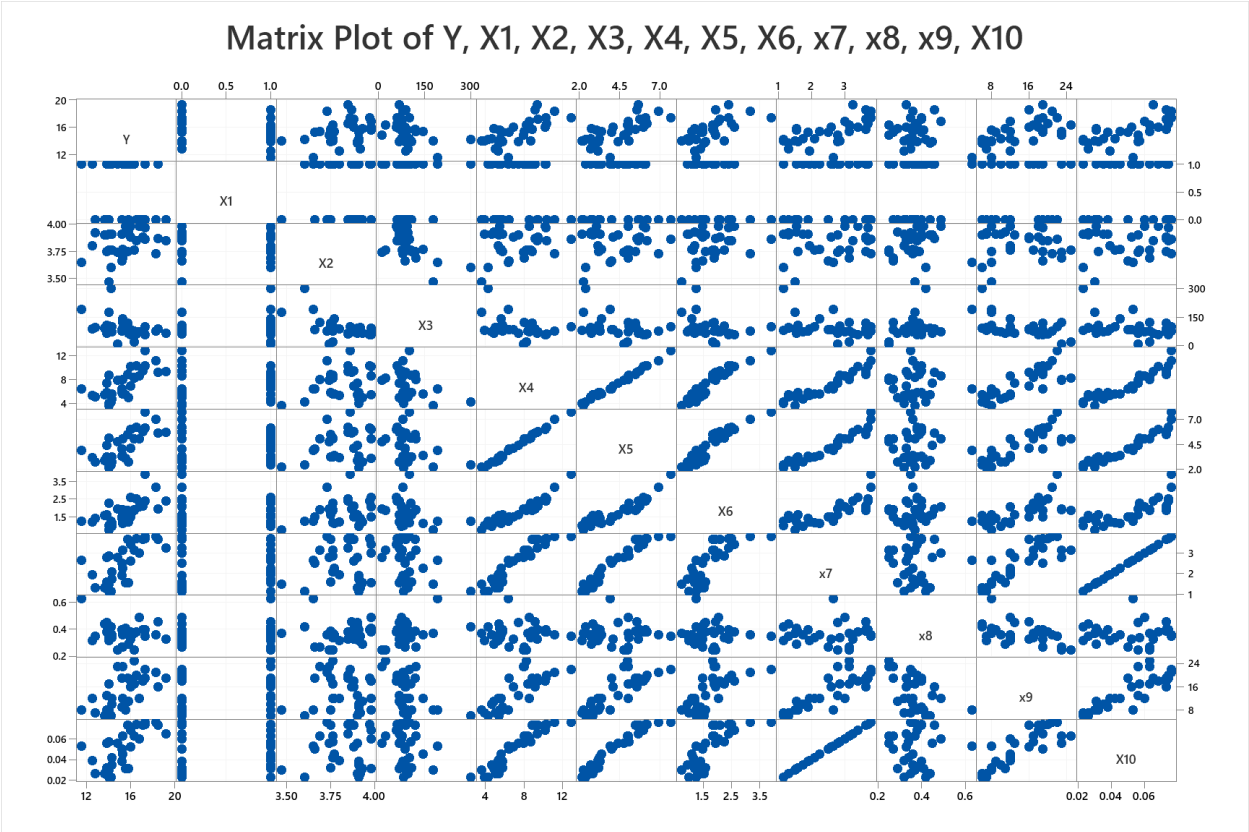
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	65.6421	8.2053	5.98	0.000
X1	1	2.8586	2.8586	2.08	0.162
X2	1	6.1207	6.1207	4.46	0.046
X3	1	2.0435	2.0435	1.49	0.235
X4	1	1.0547	1.0547	0.77	0.390
X5	1	3.1740	3.1740	2.31	0.142
X6	1	0.8585	0.8585	0.63	0.437
x8	1	2.8835	2.8835	2.10	0.161
x9	1	0.3296	0.3296	0.24	0.629
Error	23	31.5579	1.3721		
Total	31	97.2000			

Fits and Diagnostics for Unusual Observations

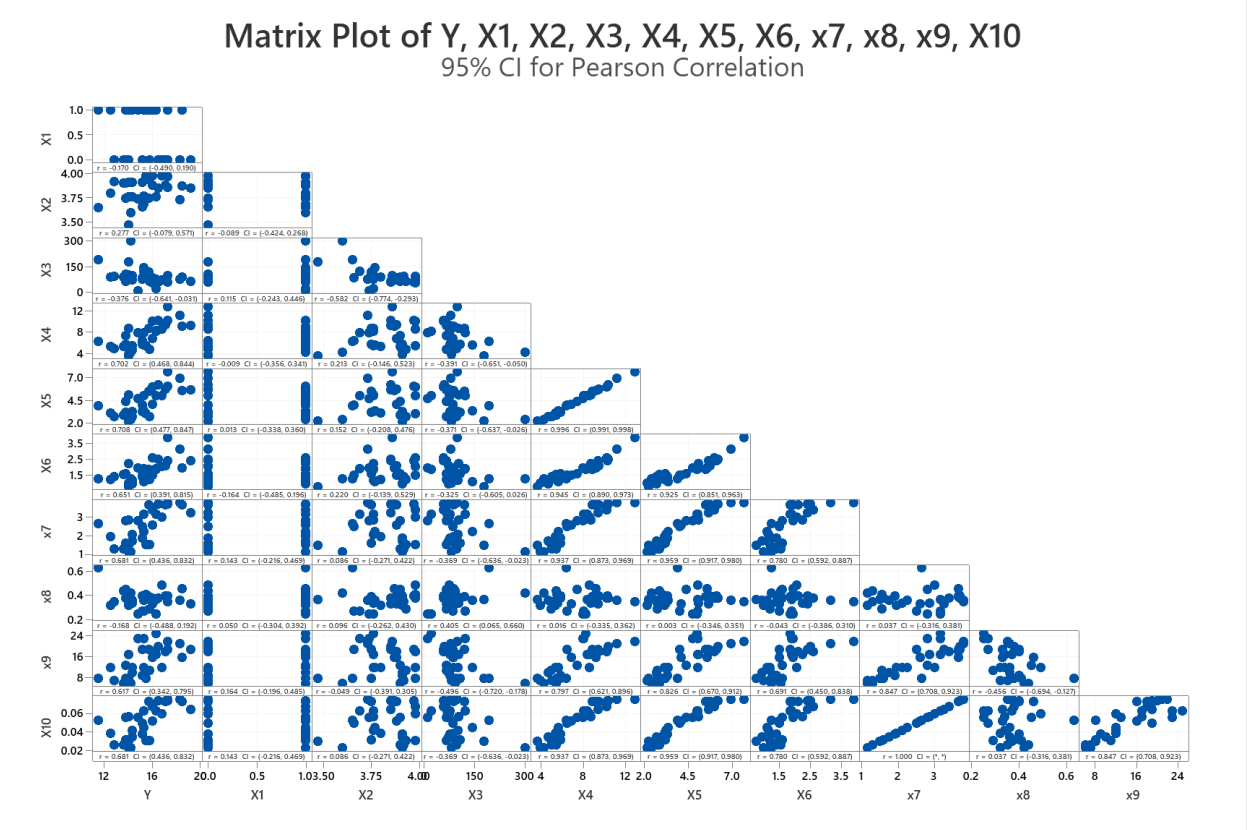
Obs	Y	Fit	Resid	Std Resid	
32	11.500	12.902	-1.402	-2.02	R

R Large residual





The matrix plot indicates strong correlation between X4, X5, X6, X9



Correlations

	Y	X1	X2	X3	X4	X5	X6	x7	x8	x9
X1	-0.170									
X2	0.277	-0.089								
X3	-0.376	0.115	-0.582							
X4	0.702	-0.009	0.213	-0.391						
X5	0.708	0.013	0.152	-0.371	0.996					
X6	0.651	-0.164	0.220	-0.325	0.945	0.925				
x7	0.681	0.143	0.086	-0.369	0.937	0.959	0.780			
x8	-0.168	0.050	0.096	0.405	0.016	0.003	-0.043	0.037		
x9	0.617	0.164	-0.049	-0.496	0.797	0.826	0.691	0.847	-0.456	
X10	0.681	0.143	0.086	-0.369	0.937	0.959	0.780	1.000	0.037	0.847

Multicollinearity is the situation when two or more independent variables in a multiple regression model are correlated with each other.
To detect multicollinearity, we use VIF (Variance Inflation Factor) to quantify its severity in the model.

Rules to analyze variance inflation factor (VIF):

1. If VIF = 1, there is no multicollinearity.
2. If $1 < VIF < 5$, there is small multicollinearity.
3. If $VIF \geq 5$, there is medium multicollinearity.
4. If $VIF \geq 10$, there is large multicollinearity.

Here we have large multicollinearity with variables X4,X5,X6,X9 (VIF>10)

In next regression model we will discard the variables X4,X5and X9 ,X10 and fit the model using X1,X2,X3,X6,X7,X8

SHEET1(W1)

Regression Analysis: Y versus X1, X2, X3, X6, x7, x8

Regression Equation

Y = -12.4 - 1.101 X1 + 6.90 X2 + 0.01382 X3 - 0.353 X6 + 1.887 x7 - 9.50 x8

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-12.4	10.2	-1.21	0.237	
X1	-1.101	0.482	-2.28	0.031	1.40
X2	6.90	2.69	2.56	0.017	2.52
X3	0.01382	0.00737	1.88	0.072	3.49
X6	-0.353	0.607	-0.58	0.566	3.90
x7	1.887	0.501	3.77	0.001	4.75
x8	-9.50	3.76	-2.53	0.018	1.89

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.14532	66.26%	58.16%	35.23%

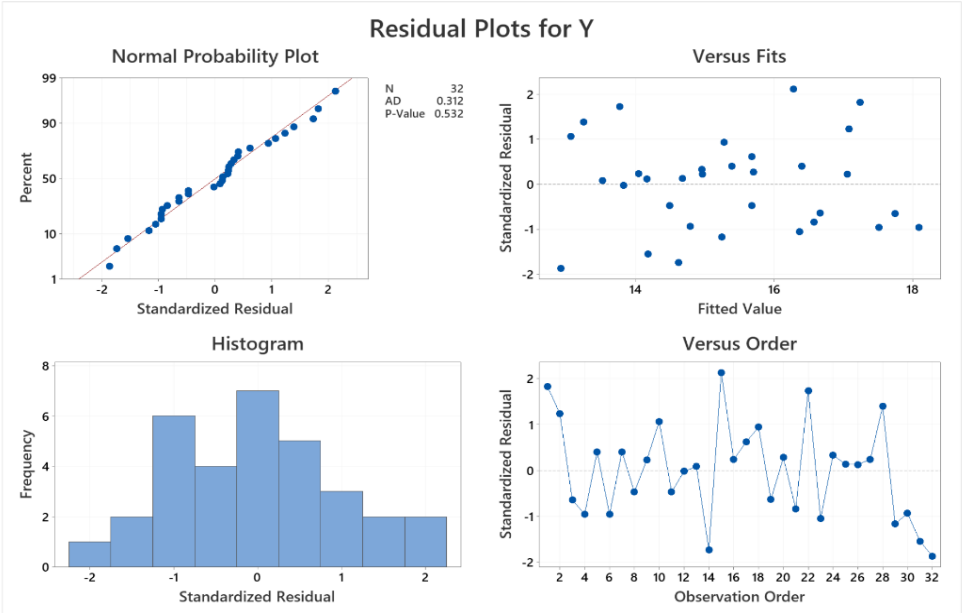
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	64.4060	10.7343	8.18	0.000
X1	1	6.8290	6.8290	5.21	0.031
X2	1	8.6137	8.6137	6.57	0.017
X3	1	4.6187	4.6187	3.52	0.072
X6	1	0.4439	0.4439	0.34	0.566
x7	1	18.6515	18.6515	14.22	0.001
x8	1	8.3670	8.3670	6.38	0.018
Error	25	32.7940	1.3118		
Total	31	97.2000			

Fits and Diagnostics for Unusual Observations

Obs	Y	Fit	Resid	Std Resid	
15	18.500	16.274	2.226	2.12	R

R Large residual



Here, all VIF are smaller than 5 hence we can conclude that the predictors have smaller collinearity.