

Six Sigma Statistics using Minitab 17

Green Belt Edition

12 Regression
Answers to Exercises

By Rehman Khan

Exercise 12.7.1

Simple Regression

Data has been collected on a process where a single predictor affects a continuous Response variable. The data has not been collected in time order. Conduct the appropriate Regression analysis and answer the questions below.

The data in File 12 Regression.xlsx worksheet Exercise1.

- 1) What order of equation best fits the data and is significant?
- 2) How much of the variation in the Response can be explained by changes in the Predictor?
- 3) Are there any unusual data points in the study and are they an issue?
- 4) Using this process could we reasonably expect a data point to appear at 60 for the predictor and 200 for the response?

Set-up-I

1. Click **Stat<<Regression<<Fitted Line Plot**
2. Enter the column headings as shown.
3. Click on the radio button to use the Cubic model.
4. Click OK to execute the procedure.

Response (Y):

Predictor (X):

Type of Regression Model

☐ Linear ☐ Quadratic ☒ Cubic

Analysis-I

Sequential Analysis of Variance				
Source	DF	SS	F	P
Linear	1	1059165	42.22	0.000
Quadratic	1	868947	381.20	0.000
Cubic	1	262	0.11	0.740

If you go to the Session Window you will see that the Cubic term is not significant and that a quadratic model should be fitted. We can do this using the Assistant.

Set-up-2

1. Click *Assistant* << *Regression*.
2. Click on the *Simple Regression* box.
3. Enter the column headings as shown.
4. Leave 'Choose for me' as the selected choice.
5. Click OK to produce the five page report.

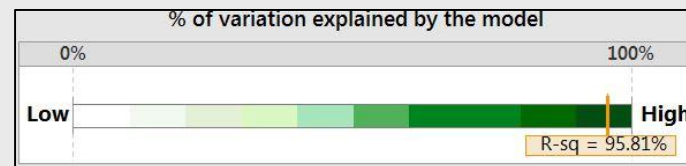
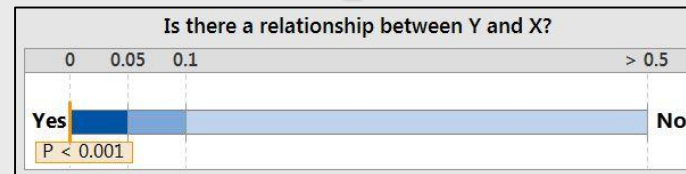
Continuous variables

Y column: Response

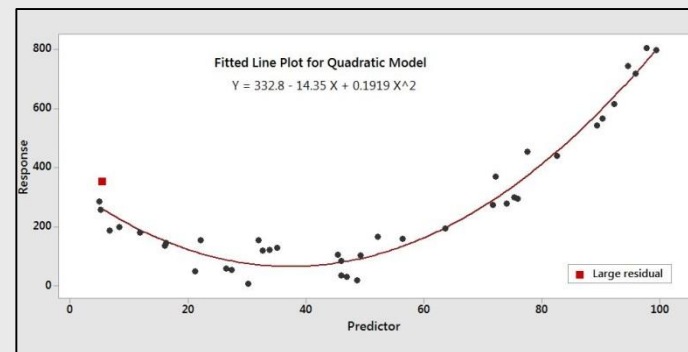
X column: Predictor



Analysis-2

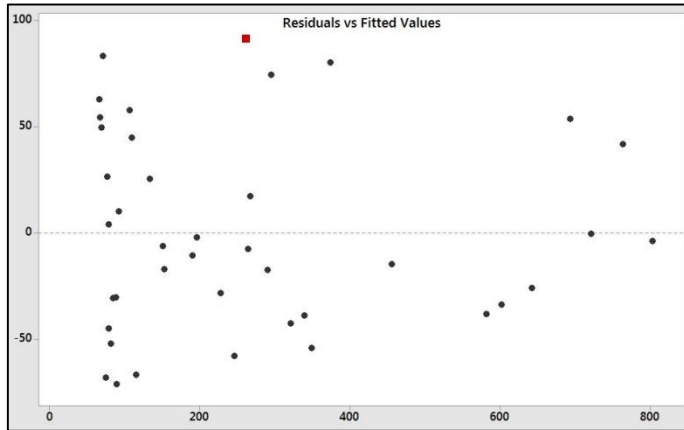


On the top left of the Summary Report we are told that the quadratic model is significant and that 95.81% of the changes in the response can be explained by changes in the predictor.



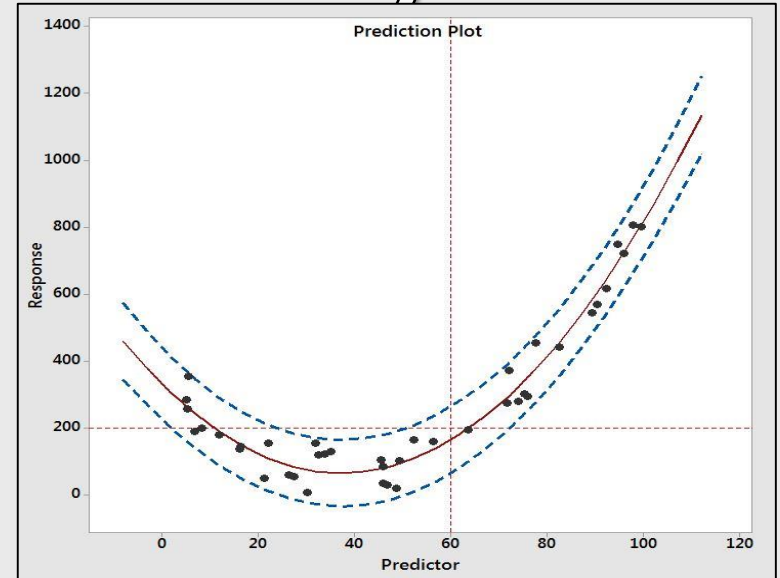
The Model Selection Report shows that one data point had a large residual.

Analysis-3



The Diagnostic Report again shows the unusual data point. It can be seen it is not too far away from the rest of the data.

Analysis-4



Reference lines have been added to the Prediction Plot on the Prediction Report. We can reasonably expect a data point to appear at 60 for the predictor and 200 for the response and this point lies within the prediction interval.

Exercise 12.7.2

Multiple Regression

Data has been collected on a process where four predictors are thought to affect a single continuous Response variable. The data has not been collected in time order. Conduct the appropriate Regression analysis and answer the questions below.

The data in File 12 Regression.xlsx worksheet Exercise2.

- 1) Are any of the predictors correlated?
- 2) Are all the predictors significant?
- 3) How much of the variation in the Response can be explained by changes in the Predictors?
- 4) What terms are present in the model?
- 5) Are there any unusual data points in the study and are they an issue?
- 6) What are the top solutions to obtain a value of 200 in the Response?

Set-up 1

1. Click **Stat**<<**Basic Stats**<<**Correlation**.
2. Enter the 4 predictors so we can check if they are correlated.
3. Click OK.
4. Go to the Session Window to check the results.

Variables:

Pred1-Pred4

Analysis- I

	Pred1	Pred2	Pred3
Pred2	-0.139 0.169		
Pred3	0.040 0.692	-0.185 0.066	
Pred4	0.016 0.871	-0.102 0.313	-0.194 0.054

Cell Contents: Pearson correlation
P-Value

The P-Values indicate that there are no significant correlations.

Set-up2

5. Click **Assistant**<<**Regression**. Click on the **Optimize Response** box.
6. Enter Response as the response.
7. Select 'Achieve a target value' from the drop-down menu and set '200' as the target.
8. Enter Pred1 to Pred4 as the four predictor variables.
9. Click OK to run the procedure and produce the 6 page report.

Response (Y) variable

Response variable: Response

What is your response goal? Achieve a target value

Enter your target: 200

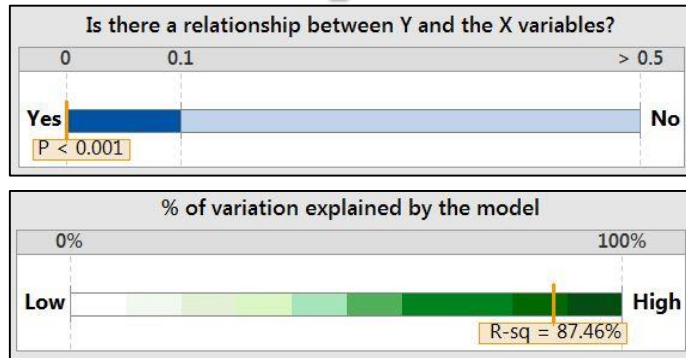
X variables

You can include up to 5 X variables.

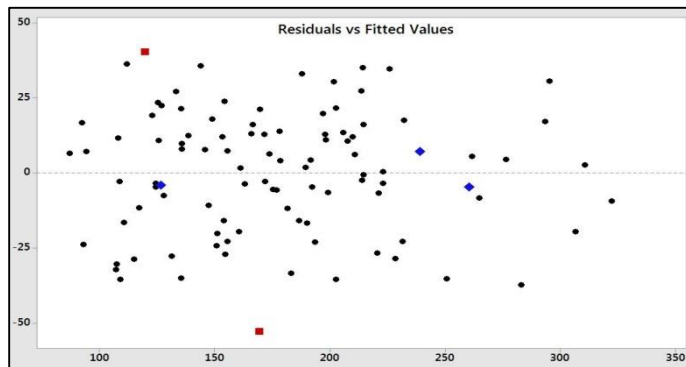
Continuous X variables:

Pred1-Pred4

Analysis-2



On the top left of the Summary Report we are told that the model is significant and that 87.46% of the changes in the response can be explained by changes in the predictor.



On the Diagnostic Report we see the Residuals vs Fitted Values plot. On this we are shown two points with large residuals and 3 points with unusual X values. We should review the data collected at these points to check for anything unusual.

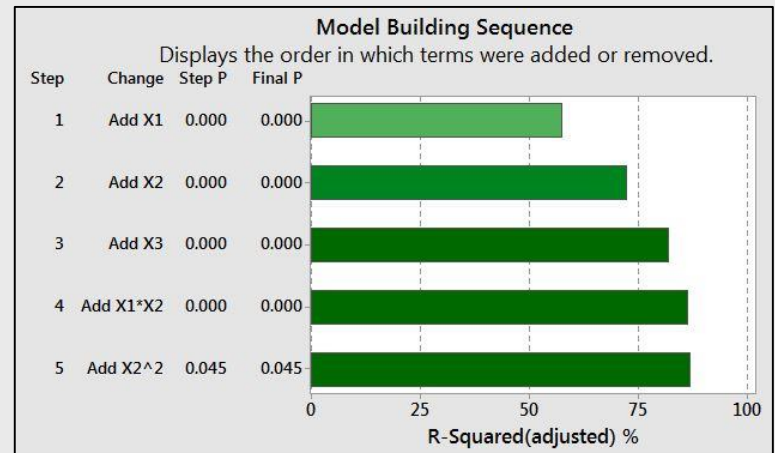
20th June 2016

Analysis-3

Final Model Equation

Response = $-93.6 + 0.030 X1 - 1.430 X2 + 6.524 X3 + 0.01305 X2^2 + 0.03284 X1 \cdot X2$

X1: Pred1 X2: Pred2 X3: Pred3 X4: Pred4



From the Model Building Report we are given the Regression Equation and we can see that it has five terms. Predictor 4 was not included as it was not significant.

Analysis-4

Solution: Optimal Settings

X1: Pred1	51.787
X2: Pred2	54.1975
X3: Pred3	36.6305

Top Five Alternative Solutions

Sample points with predicted Y values closest to the optimal solution. Evaluate these and the optimal solution to determine if any are adequate.

X1	X2	X3	Predicted Y
96.315	29.246	34.911	198.956
72.202	34.575	37.518	201.536
94.854	33.389	33.401	198.026
85.011	30.588	36.008	197.792
80.148	38.918	34.782	202.334

The Prediction and Optimisation Report shows us an optimal setting to achieve a response of 200. It gives us 5 alternate settings.

20th June 2016

Analysis-5



- Large residuals: 2 data points have large residuals and are not well fit by the equation. These points are marked in red on the Diagnostic Report.
 - Unusual X values: 4 data points have unusual X values, which can strongly influence the model equation. These points are marked on the Diagnostic Report.
- You can hover over a point or use Minitab's brushing feature to identify the worksheet rows. Because unusual data can have a strong influence on the results, try to identify the cause for their unusual nature. Correct any data entry or measurement errors. Consider removing data that are associated with special causes and redoing the analysis.

The Report Card warns us of the large residuals and Unusual X values. It would be prudent to check any history on those 5 data points that have been highlighted.