# Six Sigma Statistics using Minitab 17

# Green Belt Edition

## 08 ANOVA
## Answers to Exercises

By Rehman Khan

# 8.5 ANOVA Exercises

Exercise 8.5.1
One Way ANOVA

Exercise 8.5.2
StDev Test

Exercise 8.5.3
GLM

Exercise 8.5.4
GLM

.

# Exercise 8.5.1
# One Way ANOVA

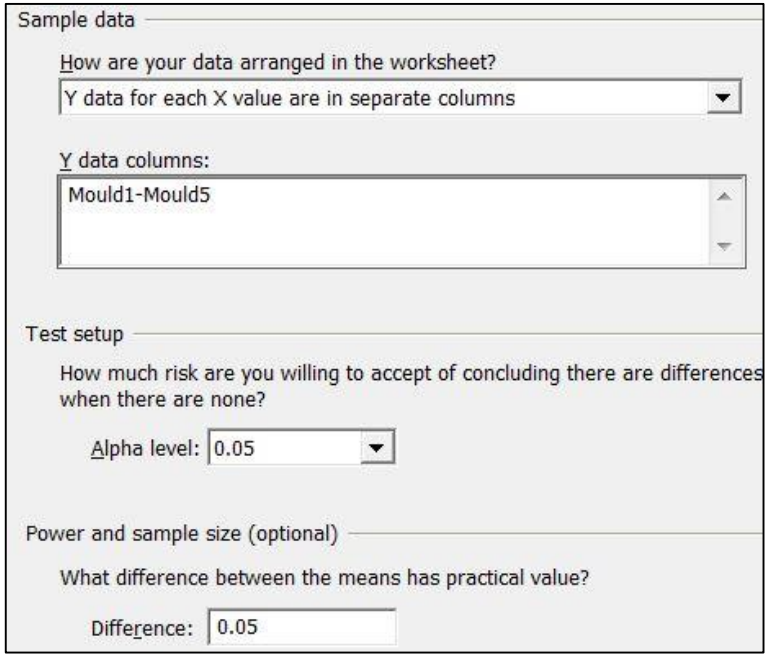
Evelee golf balls monitors the golf ball weights produced by it's 5 moulding machines. From capability tests they know that they are well below the max weight of 45.93 grams. They have hired you as a consultant and want to know if any of the moulding machines are producing golf balls at different mean weights. They feel that smallest difference that they would want to be able to detect between machines is 0.05g.

Analyse the data in File 08 ANOVA.xlsx worksheet Ex 8.5.1 and answer the questions shown below. The data was collected randomly and is recorded in time order.

1) Is there a difference in populations between the golf ball weights produced by any of the moulding machines?
2) Can the machines be grouped ?
3) Have the requirements of the test that you have used been met?
4) What was the Power of the test when you want to detect a difference of 0.05?
5) Are there any issues associated with this level of Power ?
6) Does the Report Card generate any warnings?

# Set-up

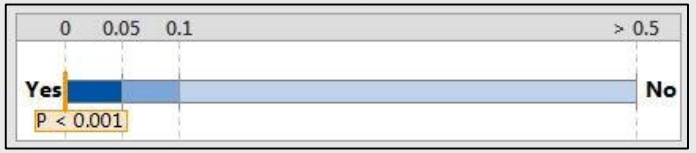
1. Click Assistant<< Hypothesis Tests
2. Click on One-Way ANOVA



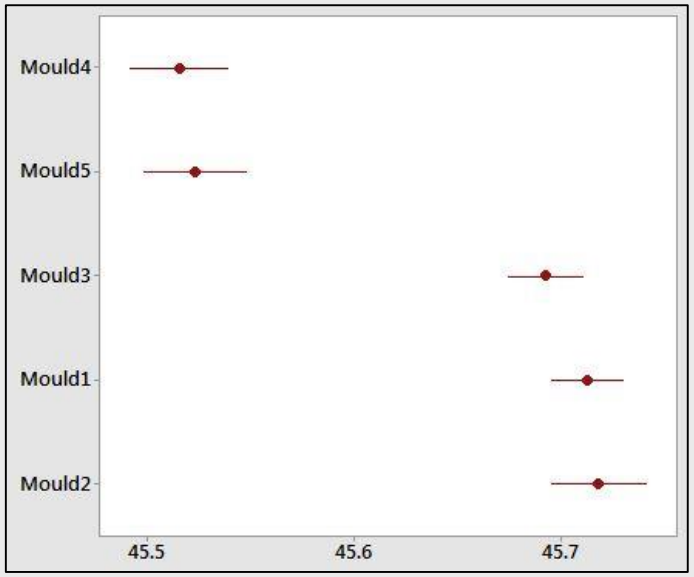

3. Complete the menu as shown and execute the procedure.

**30th Dec 2015**

# Analysis- I




Starting on the top left of the Summary Report, we can conclude that the sample data from the moulding machines came from populations where at least one pair of populations had differing mean weights.

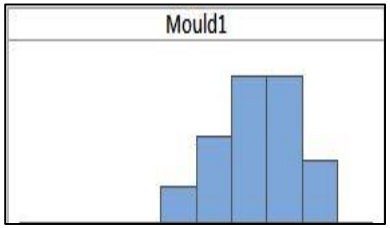


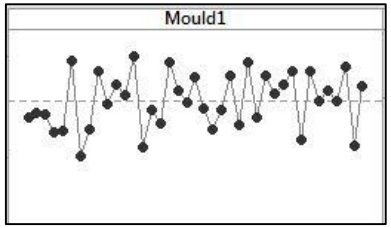

The Mean Comparison Chart shows us how the moulding machines could be grouped.

# Analysis-2

| # | Sample | Which means differ? Differs from |
|---|---|---|
| 1 | Mould4 | 3 4 5 |
| 2 | Mould5 | 3 4 5 |
| 3 | Mould3 | 1 2 |
| 4 | Mould1 | 1 2 |
| 5 | Mould2 | 1 2 |

The Grouping Information Table confirms that in terms of mean weights produced Moulds 1,2 & 3 can be considered identical and in one group. Also, Moulds 4& 5 can be considered identical and in a different group to Mould 1,2 & 3. It's a bit confusing but Moulds 1,2,&3 are groups 3,4 &5.

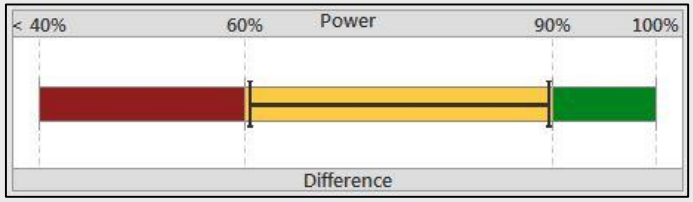


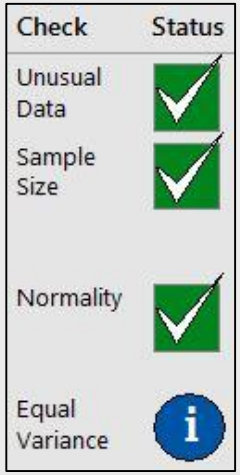

On the Diagnostic Report, the control chart shows that there were no usual data points that could affect the validity of the test.
The histograms and control charts shows the distributions were not bi-modal.

**30th Dec 2015**

# Analysis-3




The Power Report shows that a Power of between 60.6 and 89.7% % was achieved. There are no issues with this level of Power as a difference was detected.




The Report Card did not show any warnings.

# Exercise 8.5.2
# Conduct a StDev Test

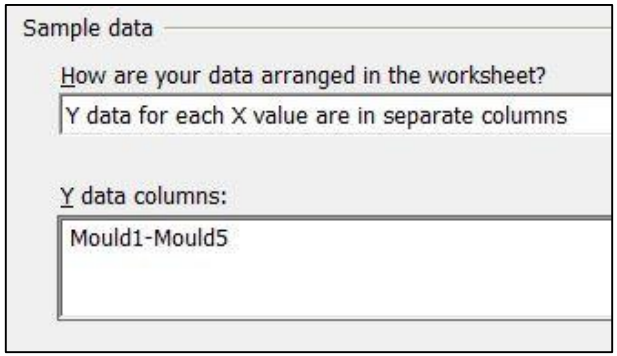
This time at Evelee golf balls someone has the bright idea that it might also be a good idea to find out if any of the moulding machines are producing golf balls at different standard deviations.

Analyse the data in File 08 ANOVA.xlsx worksheet Ex 8.5.1 and answer the questions shown below. The data was collected randomly and is recorded in time order.

1) Is there a difference in populations between the golf ball weights standard deviations produced by any of the moulding machines?
2) Can the machines be grouped ?
3) Have the requirements of the test that you have used been met?
4) Does the Report Card generate any warnings?

# Set-up

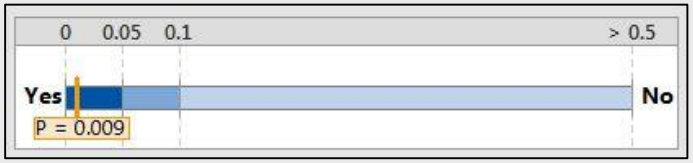
1. Click Assistant<< Hypothesis Test
2. Click on Standard Deviations Test.



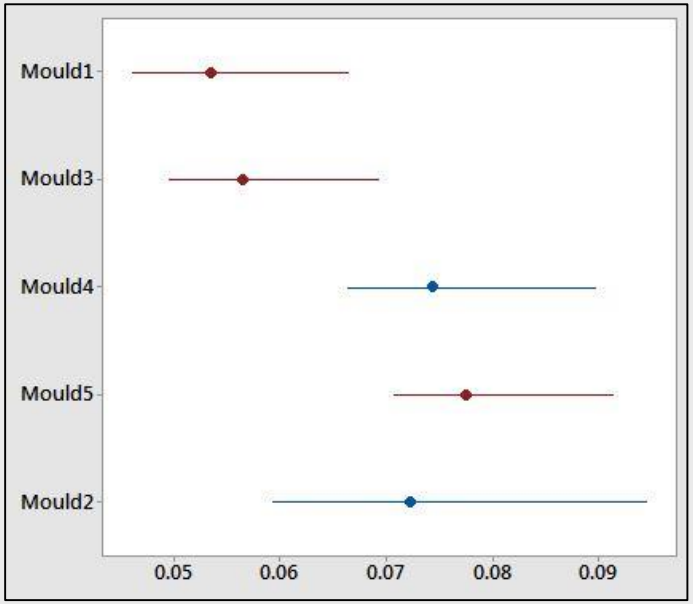

3. Complete the menu as shown and click on OK to execute the procedure.

**30th Dec 2015**

# Analysis- I




Starting on the top left of the Summary Report, we can conclude that the sample data from the moulding machines is likely to have come from populations where at least one pair of populations had differing weight StDevs.

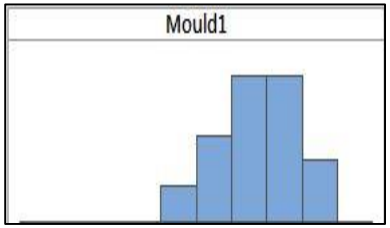


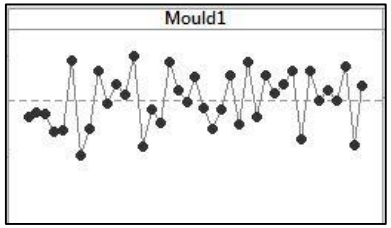

The Mean Comparison Chart shows us how the moulding machines could be grouped. The Grouping Information Table brings additional clarity.

# Analysis-2

| | | Which standard deviations differ? |
|---|---|---|
| # | Sample | Differs from |
| 1 | Mould1 | 4 |
| 2 | Mould3 | 4 |
| 3 | Mould4 | |
| 4 | Mould5 | 1 2 |
| 5 | Mould2 | |

The Grouping Information Table confirms that in terms of StDevS of weights produced Moulds 1 & 3 can be considered identical and in one group. Moulds 1 & 3 can only be considered different to 5. Moulds 2 & 4 cannot be considered different to any other mould.

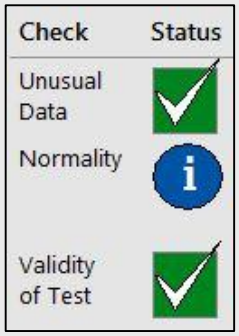



The Diagnostic Report for this example is identical to the previous example. There were no issues identified with the previous example.

**30th Dec 2015**

# Analysis-3




The Report Card did not show any warnings.

# Exercise 8.5.3
# Use the ANOVA GLM

Compo has got a new toy. It's an injection molding machine for making phone cases. Compo loads the machine with plastic granules and then sets the temperature and pressure and the molding machine forms the phone cases.

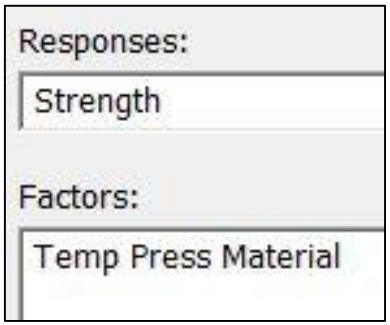
As Compo is prone to dropping things he wants to ensure the strength of the cases is maximised. He sets up an experiment where he varies three factors, the grade of the plastic granules, temperature of the heater and the pressure used for the extrusion. In this experiment **Strength** is the response and the factors are called **Temp**, **Press** and **Material**. The levels for Temp and Press are set using a dial which has fixed settings. This makes all the factors fixed. In the experiments Temp has 3 levels; 100, 110 &115 degc. Press has 3 levels; 7, 10 and 13 psi and there are two types of plastic granules called A & B.

Analyse the data in File 08 ANOVA.xlsx worksheet Ex 8.5.3 and answer the questions shown below. The data was collected randomly and is recorded in time order.

1) Form the Main Effects Plot and Interactions plot. Can you guess which factors and interactions are significant?
2) Which Factors and Interactions are significant when you use the ANOVA GLM ?
3) What is the Regression Equation for this model? And how much of the variation in Strength does it account for?
4) What settings would maximise the strength of the phone cases?
5) DO the VIFs indicate any problems with the model?
6) Do the residuals and Lack-of-Fit Test indicate any issues with your model ?

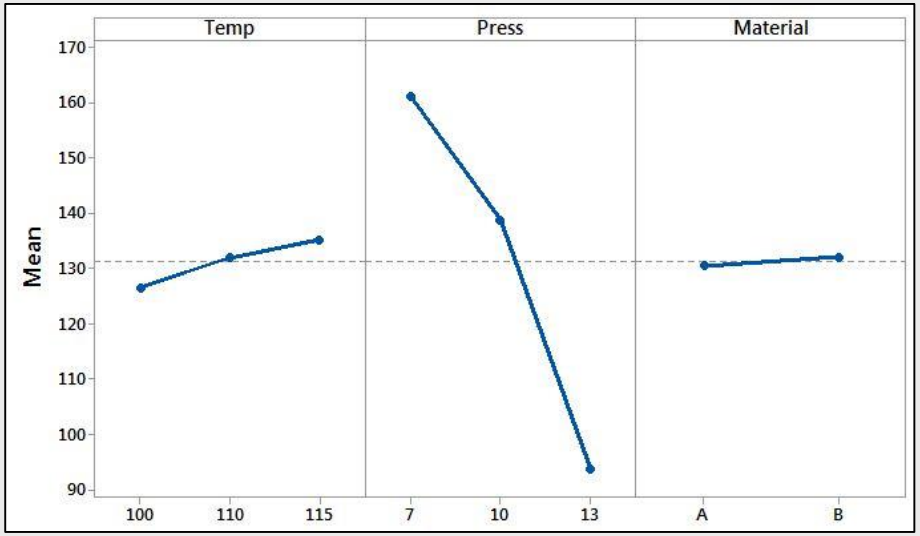
# Set-up- I

1. Click Stat<<ANOVA<< Main Effects Plot
2. Complete the menu as shown below and then click OK.




**30th Dec 2015**

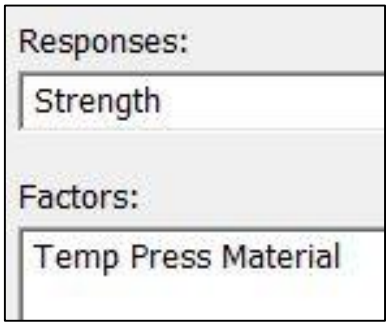
# Analysis- I




From the y-axis range exhibited between levels we can safely say that Pressure has the most effect on Strength, then Temp and then Material.
At this time we cannot say which of the factors are statistically significant.

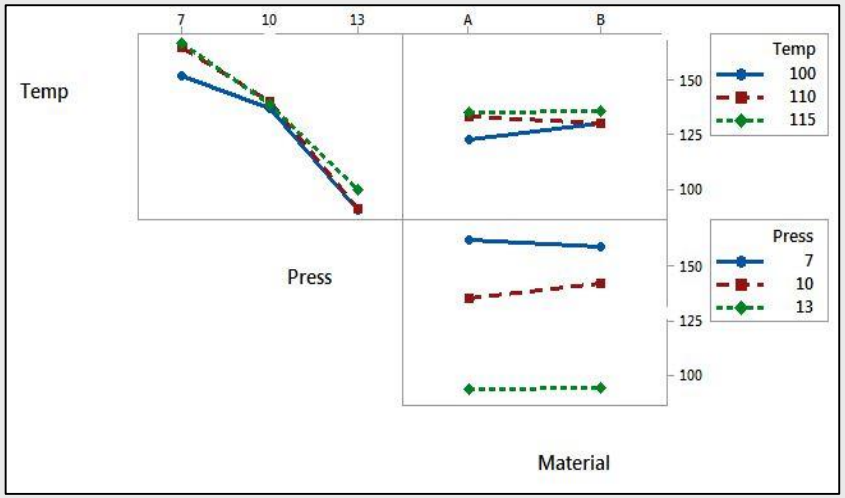
# Set-up-2

1. Click Stat<<ANOVA<< Interactions Plot.
2. Complete the menu as shown below and then click OK.




**30th Dec 2015**

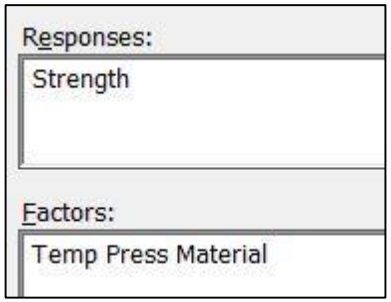
# Analysis-2



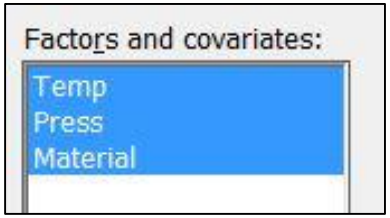

Within all three sub-plots we can see that the lines are not parallel. This indicates that there are interactions present but we cannot say if they are significant at this time.

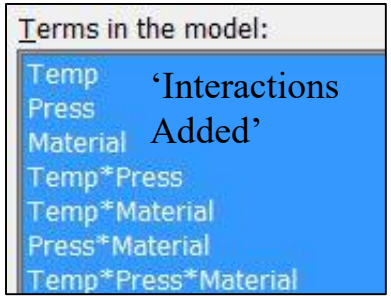
# Set-up-3

1. Click Stat<<ANOVA<<General Linear Model<<Fit General Linear Model




2. Complete the menu as shown below and then click on the **Model** button.
3. Press the Ctrl key and click on each of the factors to highlight all of them. Then go the **Interactions through order** selector and change it to **3**. Then click on the **Add** button.
4. Click OK & OK to execute the procedure.




**30th Dec 2015**

# Analysis-3

Go to the Session window and find the Analysis of Variance table. Look at the P-Value for each of the terms and consider which are significant.

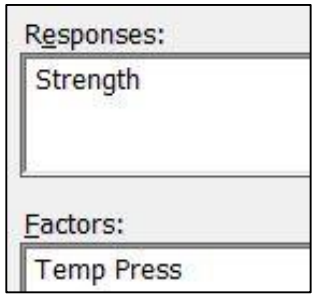
Analysis of Variance

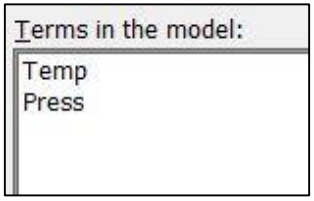
| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Temp | 2 | 925.5 | 462.7 | 4.21 | 0.020 |
| Press | 2 | 56799.3 | 28399.7 | 258.14 | 0.000 |
| Material | 1 | 47.7 | 47.7 | 0.43 | 0.513 |
| Temp*Press | 4 | 602.2 | 150.5 | 1.37 | 0.257 |
| Temp*Material | 2 | 343.6 | 171.8 | 1.56 | 0.219 |
| Press*Material | 2 | 302.0 | 151.0 | 1.37 | 0.262 |
| Temp*Press*Material | 4 | 931.6 | 232.9 | 2.12 | 0.091 |
| Error | 54 | 5940.9 | 110.0 | | |
| Total | 71 | 65892.7 | | | |

The P-Values are telling us that the Factors Press and Temp are significant. It does not matter which material we use to make the phone. Neither the 2-way or the single 3-way interactions are significant. Removing all non-significant terms from the model is the next step.

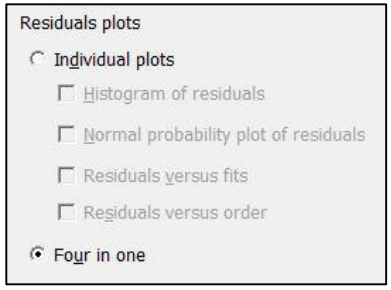
# Set-up-4

1. Click Stat<<ANOVA<<General Linear Model<<Fit General Linear Model or press Ctrl+E.
2. Remove Material as a Factor and then click on the **Model** button.
3. Minitab will remove all interactions apart from the 'Temp*Press' interaction. Remove that by pressing the Default button. Then return to the root GLM menu by clicking OK once.




4. Click on the Graphs button and select the radio button for the Four-in-One Residual plots.
5. Click OK and OK again to execute the procedure.




**30th Dec 2015**

# Analysis-4

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Temp | 2 | 925.5 | 462.7 | 3.80 | 0.027 |
| Press | 2 | 56799.3 | 28399.7 | 232.96 | 0.000 |
| Error | 67 | 8167.9 | 121.9 | | |
| Lack-of-Fit | 4 | 602.2 | 150.5 | 1.25 | 0.298 |
| Pure Error | 63 | 7565.7 | 120.1 | | |
| Total | 71 | 65892.7 | | | |

In the Session Window we find that the Analysis of Variance table only has significant terms. As we have reduced the number of terms the Lack-of-Fit p-Value has appeared. It tells us that the model does fit the data.

Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 11.0412 | 87.60% | 86.86% | 85.69% |

The R-sq value tells us that 87.6% of changes in the levels of the factors can be explained by the model.

As only Temp and Press are the significant factors we can use the Main Effects plot to establish which setting would give the greatest strength; Temp=115, Press=7.

# Analysis-5

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 131.27 | 1.30 | 100.88 | 0.000 | |
| Temp | | | | | |
| 100 | -4.71 | 1.84 | -2.56 | 0.013 | 1.33 |
| 110 | 0.73 | 1.84 | 0.39 | 0.695 | 1.33 |
| Press | | | | | |
| 7 | 29.97 | 1.84 | 16.28 | 0.000 | 1.33 |
| 10 | 7.60 | 1.84 | 4.13 | 0.000 | 1.33 |

Regression Equation

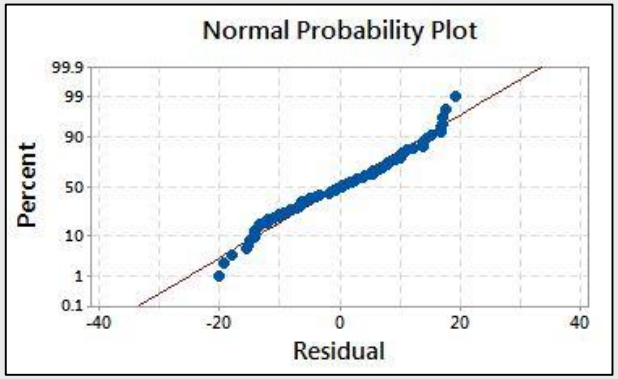
Strength = 131.27 - 4.71 Temp_100 + 0.73 Temp_110
+ 3.98 Temp_115 + 29.97 Press_7
+ 7.60 Press_10 - 37.56 Press_13

The VIFs are less than 5 which means that our model will not suffer from stability issues.

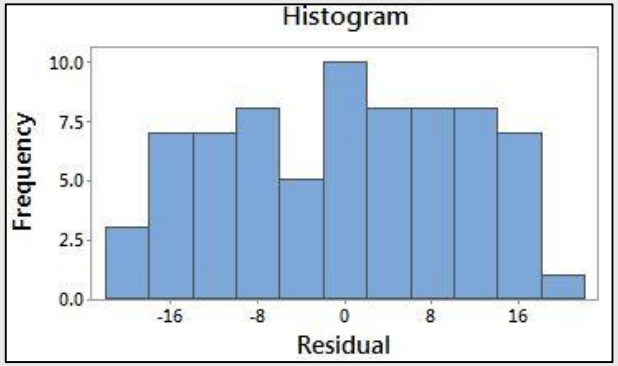
Below that we have the regression equation which we can use to predict values of strength.

**30th Dec 2015**

# Analysis-6

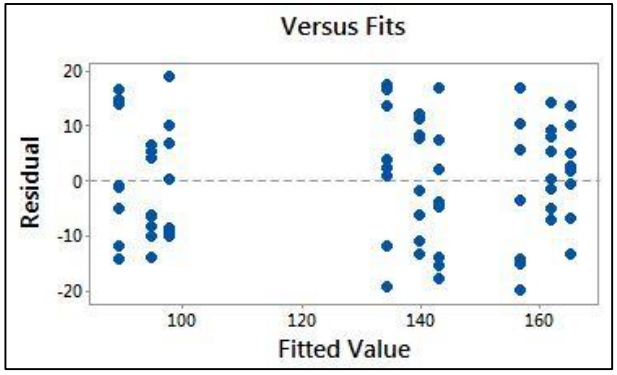



Finally, to validate the model we must check the residual plots. Find the Four-in-one residual plot in the Graph Window. Starting the with Normal Probability plot we want to know if it can be covered with a thick pencil and it can.

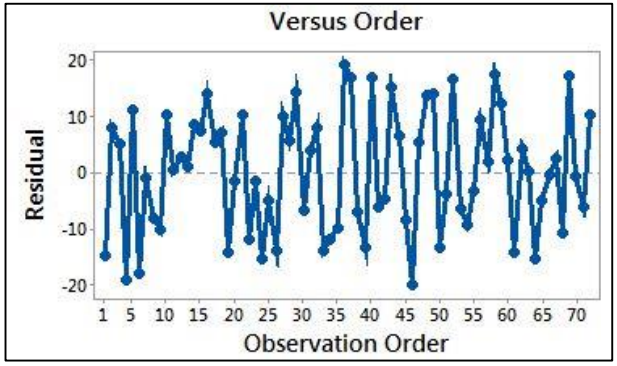



The Histogram is not extremely skewed.

# Analysis-7




The residuals are equally spaced around the zero line on the Versus Fits plot.




The data was collected in time order so using this plot is a valid check. We find no patterns in the residuals that would alert us to any problems.

**30th Dec 2015**

# Exercise 8.5.4
# Use the ANOVA GLM

A multi-media service provider wants to analyse some of their sales data. They have data on property size across 3 regions in a city. They want to know if sales are affected by property size and city region. Are there differences within the property size bands and sales? Also are there differences within the city regions and sales?

Analyse the data in File 08 ANOVA.xlsx worksheet Ex 8.5.4 and answer the questions shown above and additionally those shown below. The data was collected randomly and is recorded in time order.

In this procedure are Fixed Factor s are **Prop_size** and **Region.** The response is **Sales.**

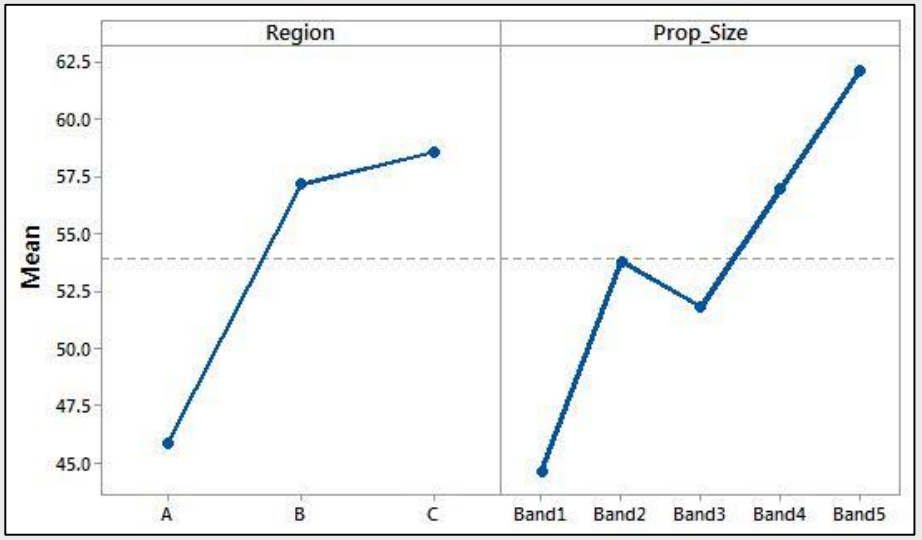
1) Form the Main Effects Plot and Interactions plot. Can you guess which factors and interactions are significant?
2) Which Factors and Interactions are significant when you use the ANOVA GLM ?
3) How much of the variation in Sales does the model account for?
4) Which Property Size and Region will generate the highest Sales?
5) DO the VIFs indicate any problems with the model?
6) Do the residuals indicate any issues with your model ?

# Set-up- I

1. Click Stat<<ANOVA<< Main Effects Plot
2. Complete the menu as shown below and then click OK.


**30th Dec 2015**

# Analysis- I

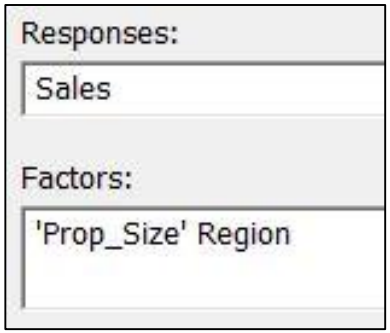



As Prop_Size has the greatest y-axis range we can say that as a Factor it is having more effect on Sales than Region. But Prop_Size has 5 levels and Region only has 3 so on a Factor level basis Region may be having more effect on Sales.

At this time we cannot say which of the factors are statistically significant.

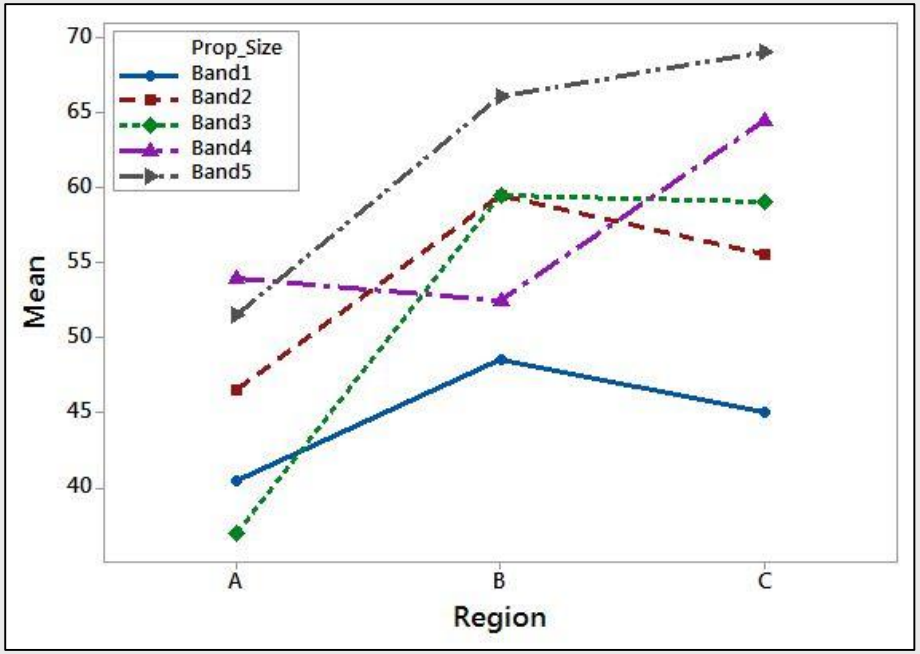
# Set-up-2

1. Click Stat<<ANOVA<< Interactions Plot.
2. Complete the menu as shown below and then click OK.




**30th Dec 2015**

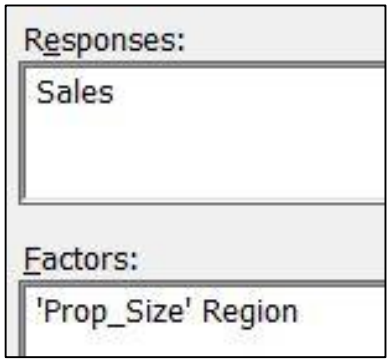
# Analysis-2



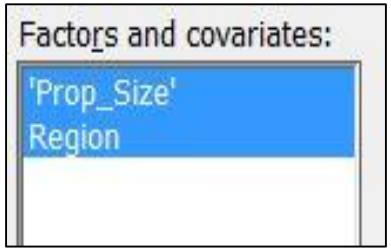

There is a single interaction plot as there are only two Factors. It appears that the trend is for Region B to have the highest sales. However, this is reversed for Band4 and this could be a significant interaction.

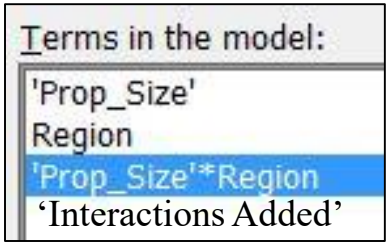
# Set-up-3

1. Click Stat<<ANOVA<<General Linear Model<<Fit General Linear Model
2. Complete the menu as shown below and then click on the **Model** button.




3. Press the Ctrl key and click on each of the factors to highlight all of them. Then go the **Interactions through order** selector and change it to **2**. Then click on the **Add** button.
4. Click OK to return to the root GLM menu.

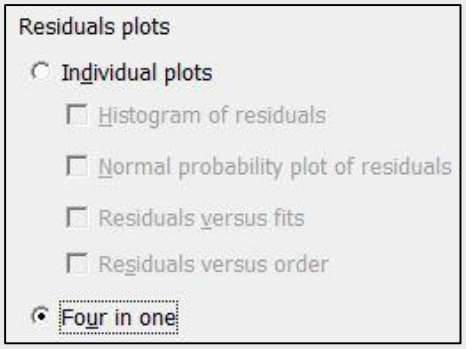



**30th Dec 2015**

# Set-up-4

5. Click on the Graphs button and select the radio button for the Four-in-One Residual plots.
6. Click OK and OK again to execute the procedure.



# Analysis-3

Go to the Session window and find the Analysis of Variance table. Look at the P-Value for each of the terms and consider which are significant.

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Prop_Size | 4 | 1004.9 | 251.22 | 14.81 | 0.000 |
| Region | 2 | 969.8 | 484.90 | 28.58 | 0.000 |
| Prop_Size*Region | 8 | 453.5 | 56.69 | 3.34 | 0.021 |
| Error | 15 | 254.5 | 16.97 | | |
| Total | 29 | 2682.7 | | | |

The P-Values tell us that all of our Factors and Interactions are significant.

**30th Dec 2015**

# Analysis-4

Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 4.11906 | 90.51% | 81.66% | 62.05% |

The R-sq value tells us that 90.51% of changes in the levels of the factors can be explained by the model.

# Analysis-5

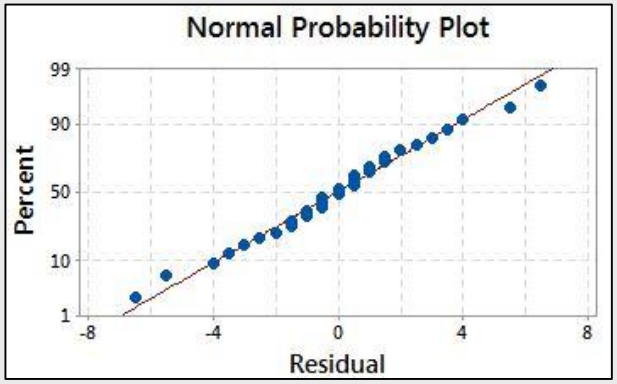
| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | 53.900 | 0.752 | 71.67 | 0.000 | |
| Prop_Size | | | | | |
| Band1 | -9.23 | 1.50 | -6.14 | 0.000 | 1.60 |
| Band2 | -0.07 | 1.50 | -0.04 | 0.965 | 1.60 |
| Band3 | -2.07 | 1.50 | -1.37 | 0.190 | 1.60 |
| Band4 | 3.10 | 1.50 | 2.06 | 0.057 | 1.60 |
| Region | | | | | |
| A | -8.00 | 1.06 | -7.52 | 0.000 | 1.33 |
| B | 3.30 | 1.06 | 3.10 | 0.007 | 1.33 |
| Prop_Size*Region | | | | | |
| Band1 A | 3.83 | 2.13 | 1.80 | 0.092 | 2.13 |
| Band1 B | 0.53 | 2.13 | 0.25 | 0.805 | 2.13 |
| Band2 A | 0.67 | 2.13 | 0.31 | 0.758 | 2.13 |
| Band2 B | 2.37 | 2.13 | 1.11 | 0.283 | 2.13 |
| Band3 A | -6.83 | 2.13 | -3.21 | 0.006 | 2.13 |
| Band3 B | 4.37 | 2.13 | 2.05 | 0.058 | 2.13 |
| Band4 A | 5.00 | 2.13 | 2.35 | 0.033 | 2.13 |
| Band4 B | -7.80 | 2.13 | -3.67 | 0.002 | 2.13 |

The VIFs are less than 5 which means that our model will not suffer from stability issues.

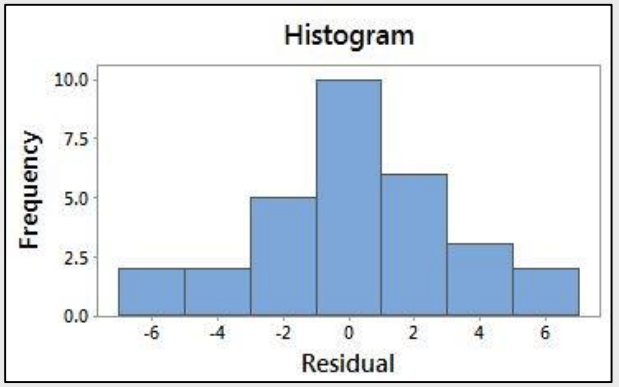
Below that we have the regression equation which we can use to predict values of Sales.

**30th Dec 2015**

# Analysis-6

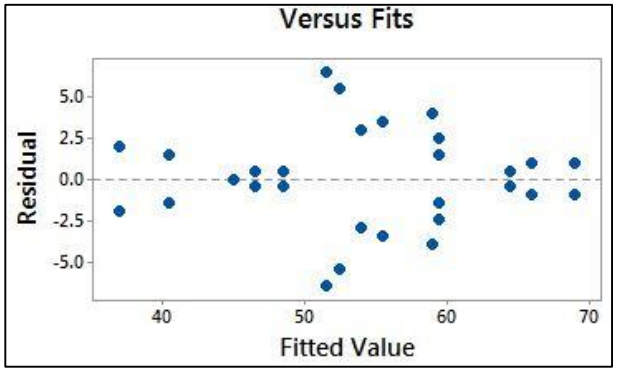



Finally, to validate the model we must check the residual plots. Find the Four-in-one residual plot in the Graph Window. Starting the with Normal Probability plot we want to know if it can be covered with a thick pencil and it can.

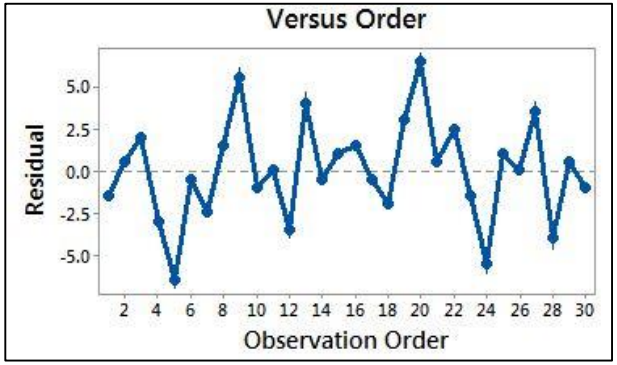



The Histogram is not extremely skewed.

# Analysis-7




The residuals are equally spaced around the zero line on the Versus Fits plot.




The data was collected in time order so using this plot is a valid check. We find no patterns in the residuals that would alert us to any problems.

**30th Dec 2015**

# Set-up-5

If we look at the Interaction Plot Band5 in Region C gives the highest Sales. But is that the full story?

1. Click Stat<<ANOVA<<General Linear Model<<Comparisons.
2. Complete the menu as shown below and then click OK. To choose the terms for comparisons click on each one to highlight it and then press the 'C=Compare…' button.

# Analysis-8

Go to the Session window and find the Grouping Information Tables generated by the pairwise comparison.

| Prop_Size | N | Mean | Grouping | | |
|---|---|---|---|---|---|
| Band5 | 6 | 62.1667 | A | | |
| Band4 | 6 | 57.0000 | A | B | |
| Band2 | 6 | 53.8333 | | B | |
| Band3 | 6 | 51.8333 | | B | C |
| Band1 | 6 | 44.6667 | | | C |

For Prop_Size, Band5 has the highest Sales but we cannot say it is different to Band4.

| Region | N | Mean | Grouping | |
|---|---|---|---|---|
| C | 10 | 58.6 | A | |
| B | 10 | 57.2 | A | |
| A | 10 | 45.9 | | B |

For Region, C has the highest Sales but we cannot say it is different to B.

**30th Dec 2015**

# Analysis-9

| Prop_Size*Region | N | Mean | Grouping | | | | | |
|---|---|---|---|---|---|---|---|---|
| Band5 C | 2 | 69.0 | A | | | | | |
| Band5 B | 2 | 66.0 | A | B | | | | |
| Band4 C | 2 | 64.5 | A | B | C | | | |
| Band3 B | 2 | 59.5 | A | B | C | D | | |
| Band2 B | 2 | 59.5 | A | B | C | D | | |
| Band3 C | 2 | 59.0 | A | B | C | D | | |
| Band2 C | 2 | 55.5 | A | B | C | D | E | |
| Band4 A | 2 | 54.0 | A | B | C | D | E | |
| Band4 B | 2 | 52.5 | | B | C | D | E | F |
| Band5 A | 2 | 51.5 | | B | C | D | E | F |
| Band1 B | 2 | 48.5 | | | C | D | E | F |
| Band2 A | 2 | 46.5 | | | | D | E | F |
| Band1 C | 2 | 45.0 | | | | D | E | F |
| Band1 A | 2 | 40.5 | | | | | E | F |
| Band3 A | 2 | 37.0 | | | | | | F |

As the interaction term was significant we need to use that to say which levels give the highest sales. Again Band5 C comes out on top. But we cannot say that it is different to the next seven groups.

However, it practise you would probably pick Band5 C as the best and then obtain more results.