

2018

# MFE 634

PART 2 OF MID-TERM  
GROUP 1

ITT CORPORATION | [Company address]

Lijiaokai Zeng  
Hargunjeet Kaur Bhatia  
Jiajing Chen  
Fang Hao

## Contents

1. DOE Analysis Introduction .....	1
2. Microsoft Excel.....	2
2.1 Full Factorial Design .....	2
2.1.1 Coding (High(+1), Low(-1)) .....	2
2.1.2 Runs.....	2
2.3 Excel Full Factorial Spread Sheet .....	3
2.4 Assign Responses .....	4
2.5 Analyze Factorial Design .....	4
2.4 Estimating Coefficients (Prediction) .....	5
3. Minitab DOE.....	5
3.1 Full Factorial Design .....	5
3.2 Minitab Full Factorial Design Framework .....	6
3.2 Assign Reponses.....	7
3.3 Analyze Factorial Design .....	8
3.3.1 Interpretation Using P-Values.....	9
3.3.2 Interpretation Using Graphs .....	10
4. Regress Data .....	11
Test of Assumptions.....	12
Normality .....	12
Equal Variance .....	12
Process Capability .....	17

## 1. DOE Analysis Introduction

Design of experiments (DOE) is used to set up and analyze a framework where multiple factors, levels and replicates can be combined and analyzed. The goal of a DOE is to determine how a response is effected by both individual factors and the interactions between each factor.

This paper analyzes three approaches to DOE using a 3 factor two level design with 2 replicates. The three methods include:

- **Microsoft Excel**

Using excels built in spread sheet, graph and intercellular calculation abilities we can both design and analyze a full factorial experiment.

- **Minitab DOE**

Minitab provides both automated set up and analysis for DOE

- **Minitab Regression**

Shows the causal relationship between output factor and input factors.

In each approach we will consider 3 components of DOE

- 1) The Design

In each case we consider the design of the experiment given the three different methods.

- 2) Analysis

We will explore the different analysis approaches provided by each method.

- 3) Prediction

We will look at each methods ability to provide a predictive equation ( $Y = F(x)$ )

At the end of this paper a comparison is done of each methods ability to satisfying the three components above.

## 2. Microsoft Excel

Using excels built in spread sheet, graph and cell calculation abilities we can both design and analyze a full factorial experiment.

### 2.1 Full Factorial Design

For this experiment we are conducting a full factorial design. That is we have 3 factors each of which have two levels and 2 replicates.

#### 2.1.1 Coding (High(+1), Low(-1))

In our setup we specify High and Low levels for each factor:

Factor	Low Level (-1)	High Level (+1)
Pressure	300	340
Temperature	50	80
Humidity	50	60

Figure 2-3 - Excel Factor Levels

The concept of coding is used to both differentiate between high and low values and to determine later values. Coding is simple taking either the High or Low value subtracting the midpoint, divided by the range and then multiplying Pressures two. This normalization is done to ensure a standardized combination of factors. See figure 2-5.

#### 2.1.2 Runs

This full factorial design has two replicates. Per the factorial design requirements this would require  $2 \times 2^3$  runs, which is 16 runs. Note the preceding equation is:

The diagram shows the equation  $2 * 2^3 = 16$ . Four boxes labeled A, B, C, and D have arrows pointing to specific parts of the equation: A points to the first '2', B points to the '2' in the exponent, C points to the '3' in the exponent, and D points to the result '16'.

- A) The number of replicates
- B) The number of levels
- C) The number of factors
- D) The total number of runs

Figure 2-1 - Excel Runs Calculation

## 2.3 Excel Full Factorial Spread Sheet

Excel can be used to create the framework for this experiment. Below we have created an experiment design where we have our 3 factors, our 2 levels and our 16 runs. Note that Y1, Y2 and Y3 represent the responses for the runs.

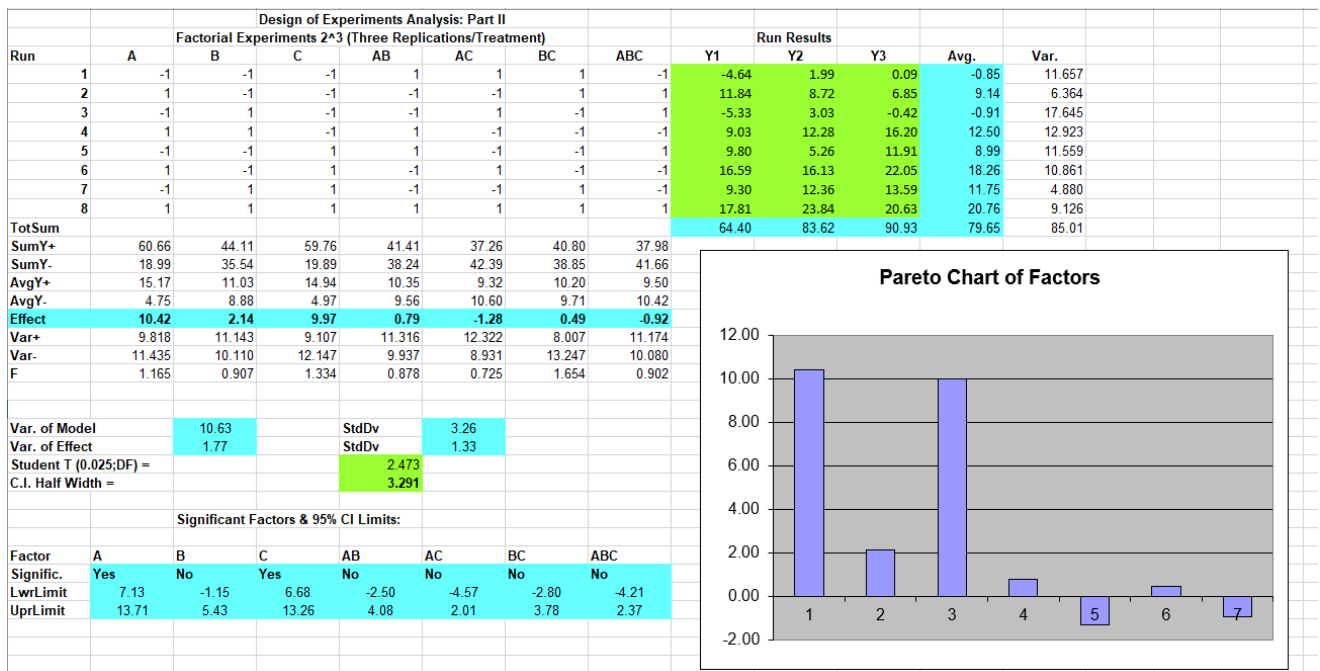


Figure 2-2 - Excel Data Entry and Analysis

To begin running experiments using the above grid we would follow the below process.

- 1) Find the first run to execute in the spreadsheet, the row numbered '1'.
- 2) Follow the row across to determine the levels each factor should be set to. In this case, run 1, in the above spread sheet has required values for the factors of
  - a. (A)Pressure = -1
  - b. (B)Temperature = -1
  - c. (C)Humidity = -1
- 3) Recall from our earlier set up the high and low levels for each factor. We set each factor to its corresponding level then run the experiments and observe the response.

- 
- a. Pressure = 300
  - b. Temperature = 50
  - c. Humidity = 50
- 4) Record the response in the “Y1” column
  - 5) Repeat step 1-4 above for each run in order as indicated.
  - 6) Populate a second runs responses using the steps above, this Pressure use ‘Y2’ for the run results and the do the same for ‘Y3’.
- Note that it assumed that all measurement systems used in the experiment are in good working order.

## 2.4 Assign Responses

As each experiment is run in the series a response is entered into the Y1,Y2 or Y3 column. Once all experiments are run there will be 16 responses ready for analysis. Since there are two replicates in the experiment we would notice that during our data entry phase that we ran each experiment setting twice. In essence through the 16 runs we gather two data points for each combination of settings Notice the excel spreadsheet is automatically calculating results in other cells as you enter data.

## 2.5 Analyze Factorial Design

Once the experiment is complete and each response has been entered we can begin the analysis phase. During the analysis phase we are primarily looking for three things

- 1) What factors have an effect on the response
- 2) What combination of factors have an effect on response.
- 3) How large is their effect

In excel we perform two calculations to determine if a factor or interaction is significant.

- 1) We first calculate the CI half width. The half width is the multiplication of the Effects stdDv Pressures the Student T value, in this case 2.473. Note the T value is based on a confidence value along with degrees of freedom.

Figure 2-3 - Excel Variance & Std Dev

- 2) To determine significant factors, we compare the factors “Effect” which is the delta between the AvgY+ and the AvgY-, see figure 2-5, factors and interactions that are significant are those that have an effect that is greater than the CI half width calculated above. The below figure 2-7 shows the significant effects.

Factor	A	B	C	AB	AC	BC	ABC
Signific.	Yes	No	Yes	No	No	No	No
LwrLimit	7.13	-1.15	6.68	-2.50	-4.57	-2.80	-4.21
UprLimit	13.71	5.43	13.26	4.08	2.01	3.78	2.37

Figure 2-4 - Excel Significant Factors

## 2.4 Estimating Coefficients (Prediction)

Figure 2-7 shows us that each factor (Pressure(A), Temperature(B) and Humidity(C)) as well as the interactions between Pressure\*Temperature and Pressure\*Humidity are significant. Determining a mathematical equation ( $Y=F(x)$ ) that we can use to make predictions about our response based on factor settings is straight forward.

- 1) Determine the constant which in this case is equal to the average of all responses.
- 2) Determine the coefficients for significant factors only, see Figure 2-7.
- 3) Coefficients are equal to the calculated Effect of the factor divided by 2.
- 4) Combine the Constant along with the coefficients to determine the predictive equation.

$$Y = 9.95 + 5.21A + 1.07B + 4.985C + 0.395AB - 0.64BC + 0.245AC - 0.46ABC$$

The above equation can be used to calculate target values for Y or target values for A, B or C given a target Y value.

- Remember that the above equation uses coded value. Therefore, calculated values of A, B and C must be decoded to get the actual setting value to use.

## 3. Minitab DOE

In addition to using excel to set up and analyze a full factorial experiment Minitab can also be used. Minitab provides both set up and analysis for DOE. The setup and analysis of the 2 level 3 factor full factorial design follows.

### 3.1 Full Factorial Design

For this experiment we are conducting a full factorial design. That is we have 3 factors (Pressure, Temperature and Humidity) each of which have two levels. A graphical representation, shown below, represents the three factors, shown as edges, as well as the two levels for each factor, shown as corners.

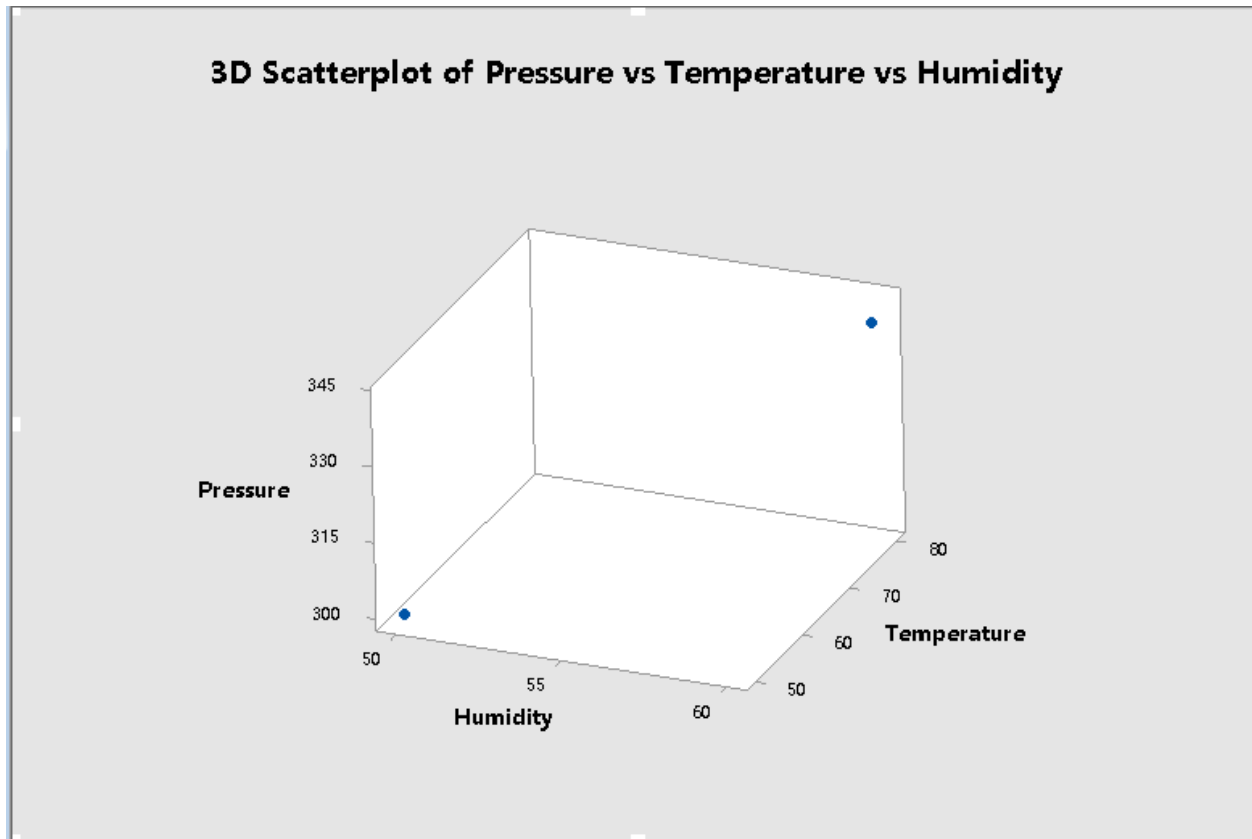


Figure 3-1 - Minitab Cube Plot

### 3.2 Minitab Full Factorial Design Framework

Minitab can be used to create the framework for this experiment. Below minitab has created a an experiment design where we have our 3 factors, our 2 levels and our 16 runs. Note that minitab has randomized the order in which each run is take place. In addition we use coded values in minitab as well.



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
	A	B	C	Y1	Y2	Y3	Y	StdOrder	RunOrder	Blocks	CenterPt				
1	-1	-1	-1	-4.6359	1.9943	0.0937	-0.8493	1	1	1	1				
2	1	-1	-1	11.8444	8.7196	6.8515	9.1385	2	2	1	1				
3	-1	1	-1	-5.3306	3.0278	-0.4176	-0.9068	3	3	1	1				
4	1	1	-1	9.0254	12.2839	16.2048	12.5047	4	4	1	1				
5	-1	-1	1	9.7995	5.2605	11.9148	8.9916	5	5	1	1				
6	1	-1	1	16.5886	16.1312	22.0542	18.2580	6	6	1	1				
7	-1	1	1	9.3019	12.3592	13.5927	11.7513	7	7	1	1				
8	1	1	1	17.8069	23.8447	20.6317	20.7611	8	8	1	1				
9															
10															
11															
12															
13															

Figure 3-2 - Minitab DOE Setup

To begin running experiments using the above grid we would follow the below process.

- 7) Find the first run to execute in the “RowOrder” Column, the row numbered ‘1’.
  - 8) Follow the row across to determine the levels each factor should be set to. In this case, run 1, in the above spread sheet has required values for the factors of
    - a. Pressure = 1
    - b. Temperature = 1
    - c. Humidity = 1
  - 9) Recall from our earlier set up the high and low levels for each factor. We set each factor to its corresponding level then run the experiments and observe the response.
    - a. Pressure = 50
    - b. Temperature = 200
    - c. Humidity = B
  - 10) Record the response in the “” column
  - 11) Repeat step 1-4 above for each run in order as indicated by the “RunOrder” column.
- Note that it assumed that all measurement systems used in the experiment are in good working order. A gage R&R analysis would show readings and settings are accurate.

### 3.2 Assign Responses

As each experiment is run in the series a response is entered into the Cost column. Once all experiments are run there will be 16 responses ready for analysis.

↓	C1	C2	C3	C4	C5	C6	C7-T	C8
	StdOrder	RunOrder	CenterPt	Blocks	Time	Temp	Catalyst (A or B)	Cost
1	14	1	1	1	1	-1	1	32.6394
2	5	2	1	1	-1	-1	1	30.5424
3	6	3	1	1	1	-1	1	33.0854
4	13	4	1	1	-1	-1	1	30.2104
5	2	5	1	1	1	-1	-1	29.3841
6	3	6	1	1	-1	1	-1	31.0513
7	4	7	1	1	1	1	-1	31.7457
8	15	8	1	1	-1	1	1	34.6241
9	8	9	1	1	1	1	1	36.8941
10	10	10	1	1	1	-1	-1	28.7501
11	7	11	1	1	-1	1	1	35.2461
12	1	12	1	1	-1	-1	-1	27.5306
13	11	13	1	1	-1	1	-1	30.7473
14	9	14	1	1	-1	-1	-1	28.0646
15	16	15	1	1	1	1	1	37.4261
16	12	16	1	1	1	1	-1	32.3437
17								

Figure 3-3 - Minitab DOE with Responses

Since there are two replicates in the experiment we would notice that during our data entry phase that we ran each experiment setting twice. For instance, rows 12 and 14 are replicates. Their settings for Pressure Temperature and Humidity are the same. In essence through the 16 runs we gather two data points for each combination of settings.

### 3.3 Analyze Factorial Design

Once the experiment is complete, each response has been entered we can begin the analysis phase. During the analysis phase we are primarily looking for three things

- 4) What factors (Pressure, Temperature and Humidity) have an effect on the response (Cost)
- 5) What combination of factors have an effect on response.
- 6) How large is their effect

We can do this in either of two ways.

- 3) We can use P-Values to determine where an effect exists. Note the Null and Alternate hypothesis:

$$H_0 : \text{effect} = 0$$

$$H_1 : \text{effect} \neq 0$$

- Remember that we fail to reject the null hypotheses when p-value > alpha ( $\alpha$ )

- 4) We can observe graphs to visual inspect the meaning of factors and how they contribute both individually and collectively.

### 3.3.1 Interpretation Using P-Values

The below minitab analysis of the DOE shows each of the three effects as well as the interactions (combinations) of effects.

```

Coded Coefficients
Term      Effect    Coef  SE Coef  T-Value  P-Value  VIF
Constant          9.956  0.655   15.20    0.000    1.00
A             10.419  5.209  0.655    7.95    0.000    1.00
B              2.143  1.071  0.655    1.64    0.120    1.00
C              9.969  4.984  0.655    7.61    0.000    1.00
A*B            0.792  0.396  0.655    0.60    0.554    1.00
A*C           -1.281 -0.640  0.655   -0.98    0.342    1.00
B*C            0.489  0.244  0.655    0.37    0.714    1.00

Regression Equation in Uncoded Units
Y = 9.956 + 5.209 A + 1.071 B + 4.984 C + 0.396 A*B - 0.640 A*C + 0.244 B*C

```

Figure 3-4 - Minitab Calculated P Values

We see from the 'P' value column that there are 4 Terms where we would fail to reject the null hypothesis. From the picture above, we could know that the p value of B, AB, AC and BC are bigger than 0.05. We make the assessment that the experiment is governed by effects from each factor as well as interactions between some factors. A summary is shown below.

Factor	Has Effect?
Pressure	Yes
Temperature	No
Humidity	Yes
Pressure * Temperature	No
Pressure * Humidity	No
Temperature * Humidity	No

Figure 3-5 - Minitab Significant Factors

### 3.3.2 Interpretation Using Graphs

Minitab graphs can give us the same information as the numerical analysis shown above. Below are graphs for each factor as well each interactions.

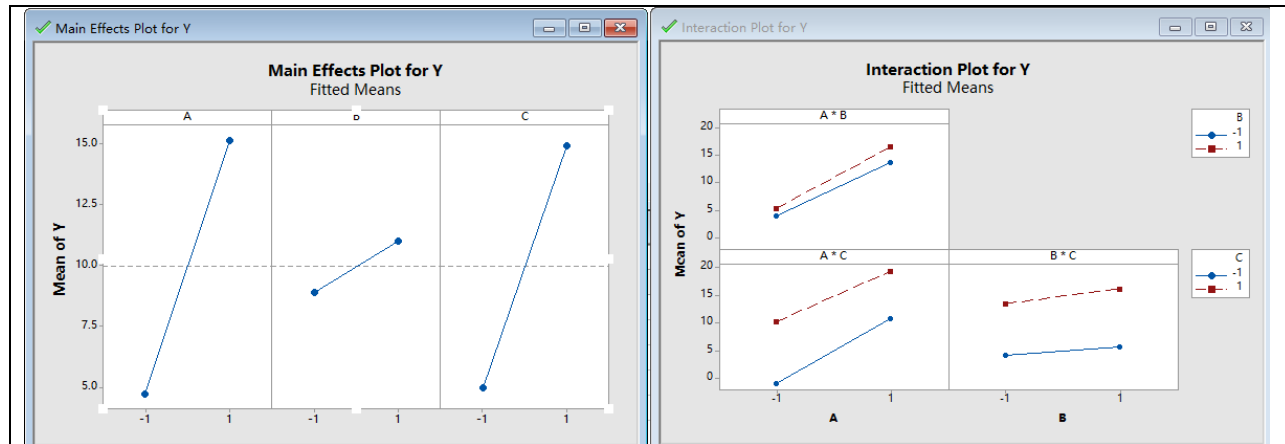


Figure 3-6 - Minitab Effects Plots

When determining if an individual factor has an effect on the experiments we look at the slope of the line. In the above figure the graph on the left shows the main effect for each of the Responses (Pressure, Temp and Humidity) note the each line shows a visible slope which indicates an effect. Only if the line were flat (i.e. slope = 0) would we conclude that the factor does not have an effect on the Responses.

Similarly, the interaction plots on the right in the above figure show the interaction between the different combinations of factors. In this case however we are looking for intersection, or the presents of possible intersection of lines (factors). It would appear from the graph of interactions that both Pressure \* Humidity and Temperature \* Catalysts do or would intersect while the interaction between Pressure \* Temperature look parallel and not likely to intersect.

The blow set of graphs can be used to tests effect in a different way. Here we look at a given value and observe if the measured values are above or below or near to or far form the reference values.

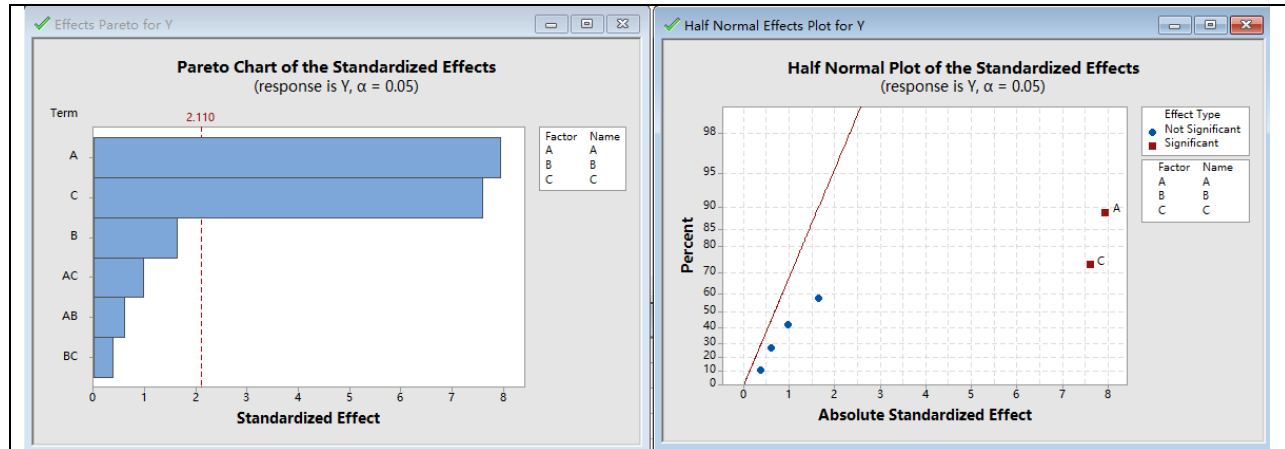


Figure 3-7 Minitab Pareto & Normal Plots For Effects

The above graphs show effects in two ways. On the left we see Pareto chart. With the pareto chart we see a boundary line (the red line labeled 2.2.6) this line is a 95% confidence boundary. Factors and interactions that go beyond this line are assumed to have an effect, factors and interactions that do not pass this line are assumed to have no effect.

Similarly in the graph to the right in the above figure there is a blue line and several dots representing the factors and interactions. The dots that are red and a distance from the line are considered to have an effect while the dot(s) that are black and near the line are considered to have no effect.

From a graphical perspective we make the same conclusions as we did in section 3.3.1.

Factor	Has Effect?
Pressure	Yes
Temperature	No
Humidity	Yes
Pressure * Temperature	No
Pressure * Humidity	No
<b>Temperature * Humidity</b>	No

Figure 3-8 - Minitab Significatn Effects

From the result above, we could know that the result of the Minitab is the same as the result of Excel.

## 4. Regress Data

In order to regress the data, the log10 was taken of both the variance and the standard deviation through Minitab. Below is a screen shot of the Minitab work sheet:

↓	C1	C2	C3	C4 <input checked="" type="checkbox"/>	C5	C6	C7	C8	C9	C10	C11
	A	B	C	Y	StdOrder	RunOrder	Blocks	CenterPt	variance	std	logstd
1	-1	-1	-1	-4.6359	1	1	1	1	11.6568	3.41421	0.533290
2	1	-1	-1	11.8444	2	2	1	1	6.3639	2.52267	0.401861
3	-1	1	-1	-5.3306	3	3	1	1	17.6449	4.20059	0.623310
4	1	1	-1	9.0254	4	4	1	1	12.9228	3.59483	0.555679
5	-1	-1	1	9.7995	5	5	1	1	11.5594	3.39991	0.531467
6	1	-1	1	16.5886	6	6	1	1	10.8605	3.29553	0.517926
7	-1	1	1	9.3019	7	7	1	1	4.8799	2.20904	0.344204
8	1	1	1	17.8069	8	8	1	1	9.1262	3.02096	0.480146

Figure 9 - Minitab work sheet for LogVar and LogStdev

Regression analysis was run of LogVar vs A, B, and C to determine the source of the variance.

## Test of Assumptions

### Normality

From the following picture, we could conclude that all the data submit to the normal distribution.

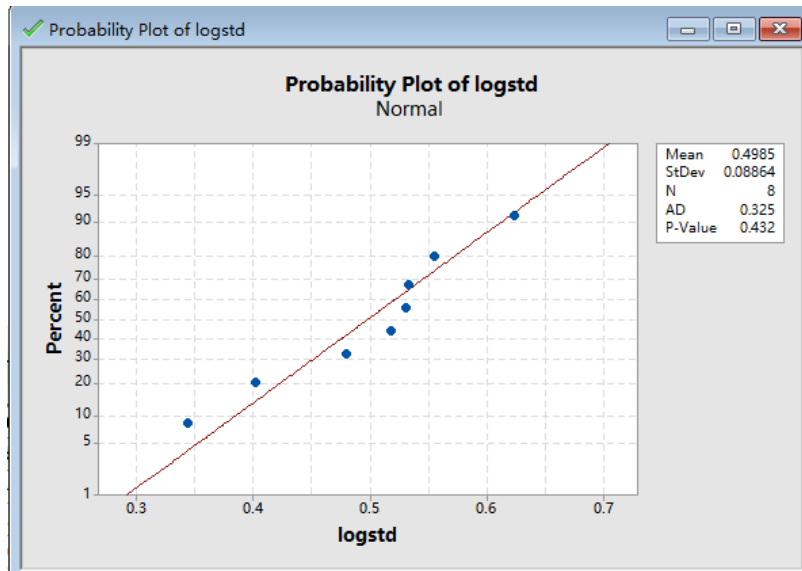


Figure 10 - Plots for check of assumptions

### Equal Variance

In order to test whether the yield has equal variance with the factors, the null hypothesis is:

$$H_0: \sigma^2_A = \sigma^2_B = \sigma^2_C = \sigma^2_{\text{Yield}}$$

$H_a$ : At least one variance is not equal

An ANOVA test for equal variances was run in Minitab for each of the factors with respect to the yield. Below is a screen capture of the Minitab work sheet.

<b>C14</b>	<b>C15</b>	<b>C16</b>	<b>C17</b>
<b>A_1</b>	<b>B_1</b>	<b>C_1</b>	<b>var</b>
300	50	50	-4.6359
340	50	50	11.8444
300	80	50	-5.3306
340	80	50	9.0254
300	50	60	9.7995
340	50	60	16.5886
300	80	60	9.3019
340	80	60	17.8069
300	50	50	1.9943
340	50	50	8.7196
300	80	50	3.0278
340	80	50	12.2839
300	50	60	5.2605
340	50	60	16.1312
300	80	60	12.3592
340	80	60	23.8447
300	50	50	0.0937
340	50	50	6.8515
300	80	50	-0.4176
340	80	50	16.2048
300	50	60	11.9148
340	50	60	22.0542
300	80	60	13.5927
340	80	60	20.6317

Figure 11 - Minitab work sheet for test of equal variance

Below are the figures created by Minitab with respect to factors A, B, and C:

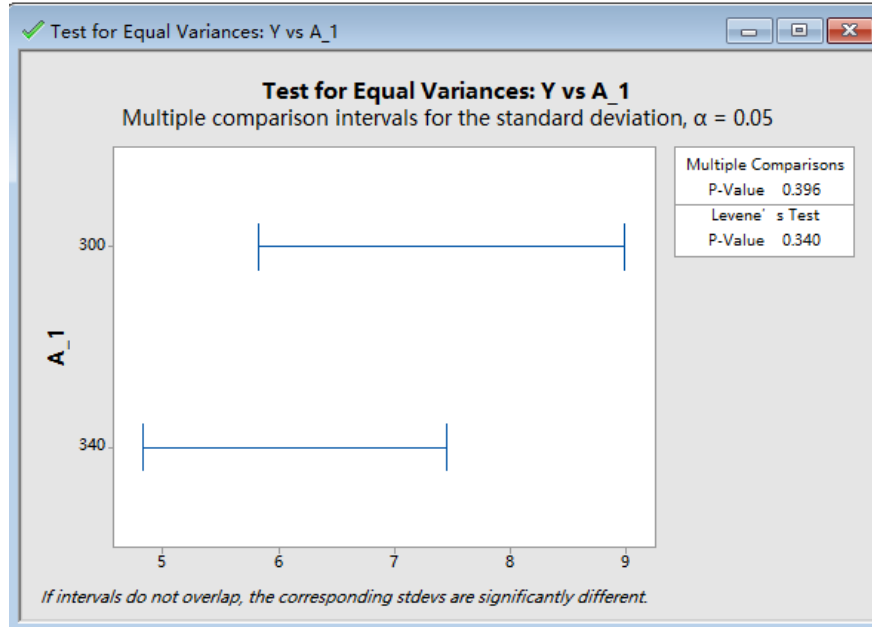


Figure 12 - Test for equal variance with factor A

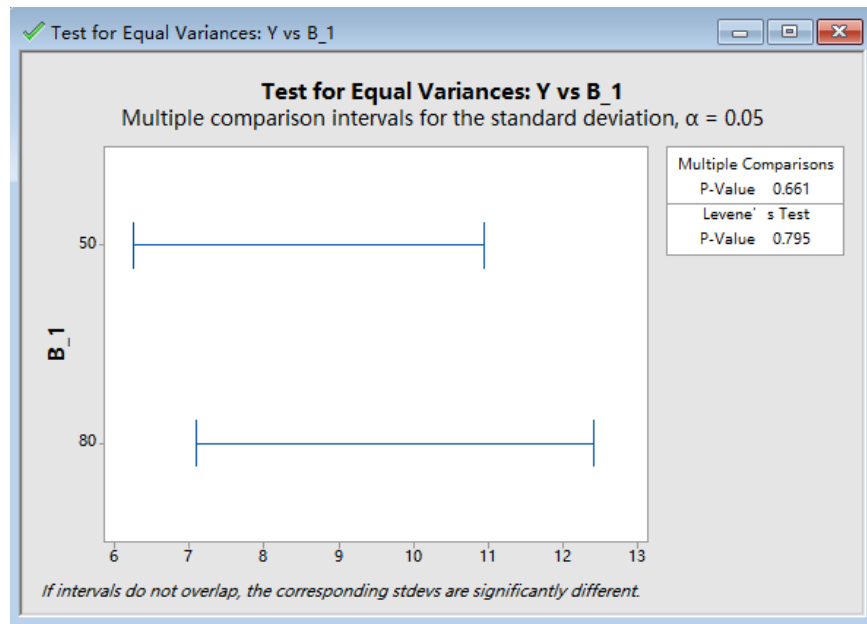


Figure 13 - Test for equal variance with factor B



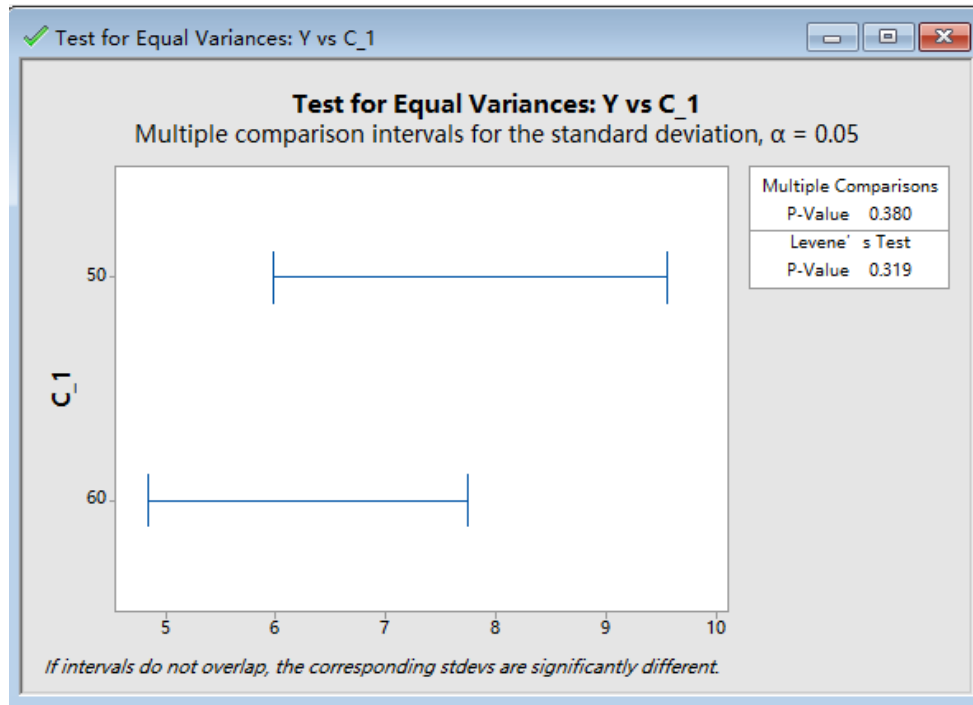


Figure 14 - Test for equal variance with factor C

In conclusion, we fail to reject the  $H_0$ , because all p-value are greater than 0.05.

## .Regression Analysis: logstd\_1 versus A\_1, B\_1, C\_1

### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	0.024008	0.008003	1.14	0.359
A_1	1	0.002204	0.002204	0.31	0.582
B_1	1	0.000132	0.000132	0.02	0.892
C_1	1	0.021672	0.021672	3.07	0.095
Error	20	0.141006	0.007050		
Lack-of-Fit	4	0.141006	0.035252	*	*
Pure Error	16	0.000000	0.000000		
Total	23	0.165014			

### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0839661	14.55%	1.73%	0.00%

### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.972	0.341	2.85	0.010	

---

A_1	-0.000479	0.000857	-0.56	0.582	1.00
B_1	0.00016	0.00114	0.14	0.892	1.00
C_1	-0.00601	0.00343	-1.75	0.095	1.00

Regression Equation

$$\text{logstd}_1 = 0.972 - 0.000479 A_1 + 0.00016 B_1 - 0.00601 C_1$$

**Figure 15 – Minitab screen output for Regression of LogVar vs A, B, and C**

From the Minitab output, we can see that all the Factor Effect p-values are greater than 0.05, which means that the factor effects have no significant effect on the variance of the response. And the result gotten from Minitab is identical to the result in the Excel

## Process Capability

The LSL is 4 and USL is 10. The process capability was determined through Minitab and is shown below:

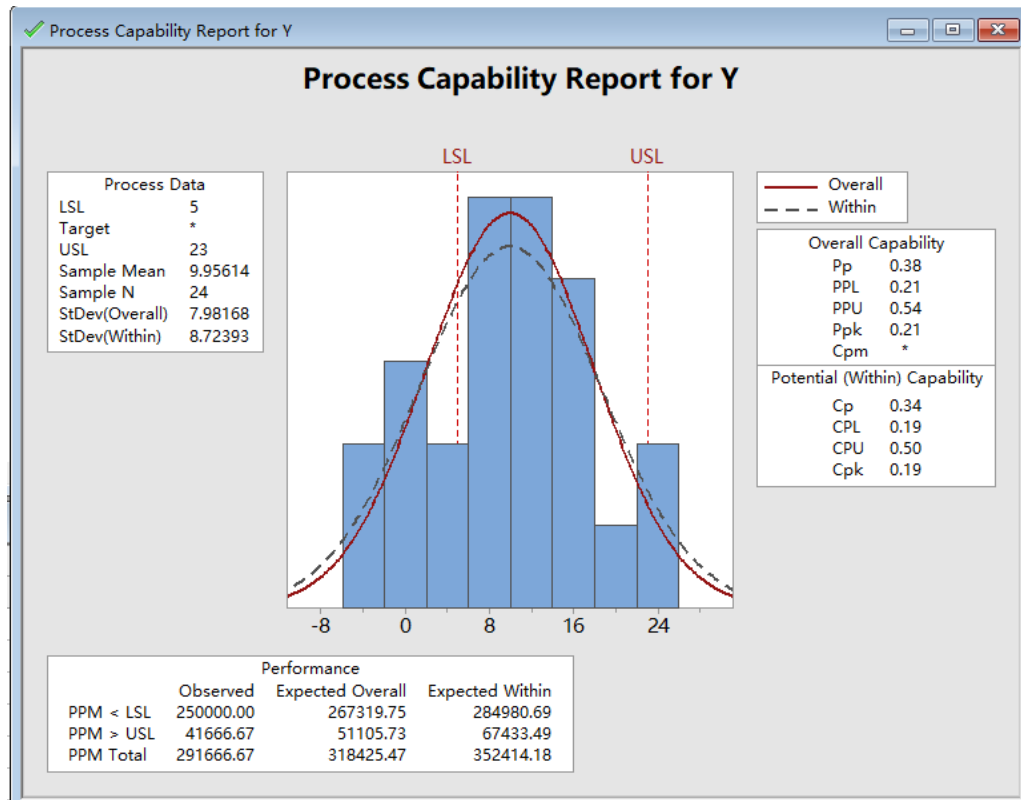


Figure 16 - Current process capability

From the picture above, we could see that the  $C_p$  is 0.34, which is great less than 1.33, so it is still hard to satisfy the customers requirement. Besides,  $C_{pk}$  is 0.19, which means that there exist a great difference between  $C_p$  and  $C_{pk}$ , so the u has a great bias./

From the picture above, we could see that the u is very close to the LSL, so we have to move whole the curve right. From the analysis before, we could know that A and B has a significant effect on the response. In order to increase the u, we could increase A or B, then the curve will move right, which will better satisfy the requirement.