Massey University

School of Engineering and Advanced Technology

228.371
Statistical Modelling for Engineers and Technologists

Design of Experiments Component

WEEK6 Lab and Tutorial

2015

A/Professor (Dr) Nigel Grigg
Dr Nihal Jayamaha

Table of Contents

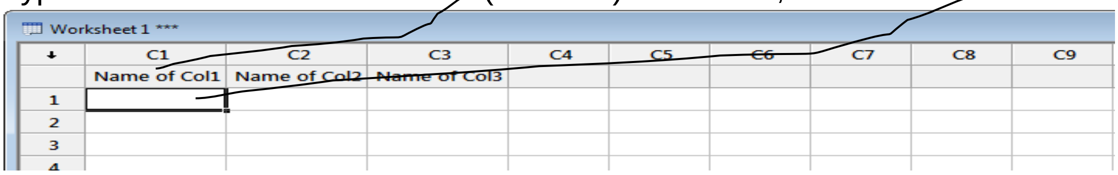
WEEK 6 Lab

Introduction to Minitab

Minitab is a user friendly statistical software package. In your first lab class you will come across Minitab terms such as the Minitab Project file *(equivalent to an Excel file)*, Minitab worksheets *(equivalent to the worksheets in an Excel file)*, *Session* and so on. Each Minitab worksheet can be named and saved individually (*.MTW) within the project file (*.MPJ). Since a Minitab project file can contain several datasheets (worksheets), you need to select the particular datasheet (to make it active) to perform the manipulation that you want to perform.

Minitab can import Excel files and worksheets as well as data files in other formats (e.g. *.txt, *.csv). If you have to enter data manually in a Minitab worksheet, depending on what you type in a particular data field (column), Minitab will detect whether a particular data field in the data sheet is “Numeric”, “Text”, or “Date/Time”. Sometimes (e.g. when you start entering text in a column that should contain numeric information) you will need to prompt Minitab to change the *type of data* (e.g. from text → numeric) that Minitab identifies by default (Data→ Change Data Type).

Type the names of the data fields (columns) in this row; enter data from row 1 onwards.



WEEK 6 Computer Lab Exercises – Two-Level Full Factorial Experiments

Exercise 1 – Creating a and Analysing a 2^4 Design

A four factor unreplicated[1] full factorial experiment was conducted in a pilot plant by a team of chemical engineers and technologists to understand how the process variables temperature (A), pressure (B), formaldehyde concentration (C), and the stirring rate (D) affect the filtration rate in a chemical plant that uses a pressure vessel.

The experiment was conducted in single block (i.e. under the same experimental conditions). Table 1.1 shows the design matrix and the response (filtration rate) values recorded by the experimenters. The actual (random) order in which the trials were conducted (run order) by the experimenters are also shown in Table 1.1.

Table 1.1: The Results of the Pilot Plant Experiment[2]

Trial Code	Run Order	Factor				Filtration Rate (gal/h)
		A	B	C	D	
(1)	3	-1	-1	-1	-1	45
a	1	+1	-1	-1	-1	71
b	16	-1	+1	-1	-1	48
ab	14	+1	+1	-1	-1	65
c	5	-1	-1	+1	-1	68
ac	13	+1	-1	+1	-1	60
bc	2	-1	+1	+1	-1	80
abc	12	+1	+1	+1	-1	65
d	4	-1	-1	-1	+1	43
ad	11	+1	-1	-1	+1	100
bd	15	-1	+1	-1	+1	45
abd	6	+1	+1	-1	+1	104
cd	10	-1	-1	+1	+1	75
acd	8	+1	-1	+1	+1	86
bcd	7	-1	+1	+1	+1	70
abcd	9	+1	+1	+1	+1	96

Notes:

(a) Here, all your factors are numeric (i.e. they are measurable).

(b) If you know what the actual (uncoded) low and high numerical values of each factor, you can tell Minitab what they are. Then Minitab will produce a regression equation based on the un-coded (actual) values.

(c) Note that often engineers and technologists also use qualitative factors (e.g. catalyst type/name, sweetener type). These factors can be introduced as "text" factors in Minitab.

You must first create a Minitab Worksheet to analyse the above data. You can efficiently create your worksheet in Minitab using two methods.

Method 1 –Prompt Minitab to create a design matrix in the standard order (i.e. prompt Minitab not to randomise the runs, using the "Options" tab). However, in this case, you will have to manually enter your response (Y) data. Thereafter, you need to manually edit the Run Order column (replace 1, 2, 3 etc. with 3, 1, 16 etc.) to get the values shown in Table 6.1. If you do not do this, your "Residual versus Observation Order" plot (one of your important residual plots) would be all wrong!

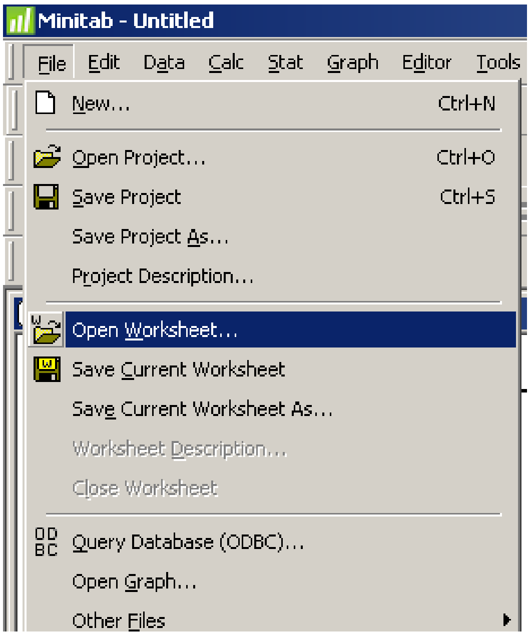
[1] In the convention used in Minitab (as well as many other software packages), an unreplicated design is treated as a design having number of replicates =1; if the 16 runs in Table 6.1 are replicated once more to end up with 32 runs altogether, the number of replicates have to be set to be = 2 in Minitab).

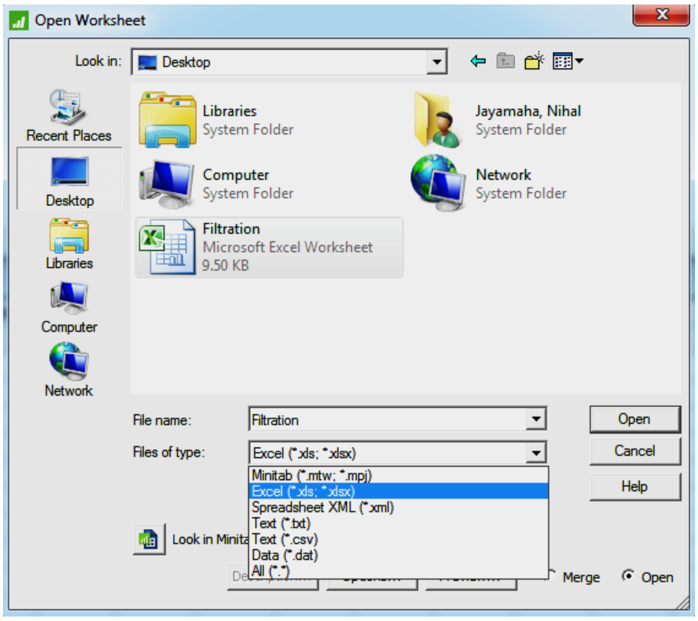
[2] Data taken from: Montgomery, D. C. (2013, p, 257). *Design and analysis of experiments* (8th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Method 2 - If you already have an Excel spreadsheet containing the data fields StdOrder, RunOrder, A, B, C, D, and Y (see below) handy, you can open your Excel spreadsheet as a Minitab Worksheet (<u>copy paste also works</u>). However, you will still have to prompt Minitab to recognise the spreadsheet before it can analyse the data. <u>Let us use this method in the lab today</u>.

Tasks

(a) In the lab, download the Excel file (this Excel file contains just one worksheet) “Filtration.xlsx” from Stream; before you need to use it, make sure that you save it (e.g. in the desktop). Open Minitab 17 and then open the worksheet (see the panel below).





Excel is just one of the formats that Minitab supports.

Save your Minitab project file and the worksheet <u>as a Minitab worksheet</u> “WK6Lab.MPJ” and “Filtration.MTW” respectively (the file extensions will automatically appear).

(b) Attempt to analyse the data! <u>S</u>tat → <u>D</u>OE→ <u>F</u>actorial → <u>A</u>nalyse Factorial Design

You should see the following display:

The current worksheet does not contain a design created by Minitab. Before Minitab can analyze your design, you will need to provide some information, such as which columns contain the factors.

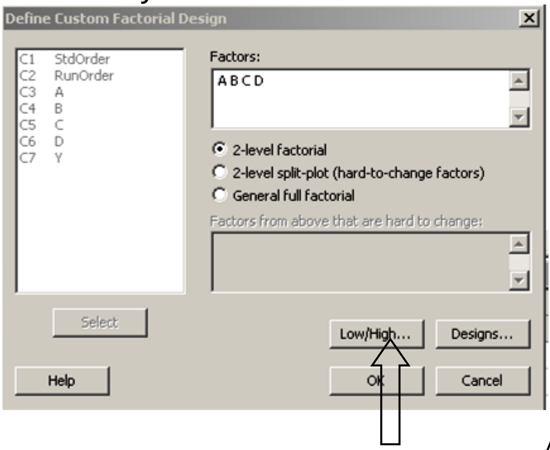
Would you like to provide this information so that Minitab can analyse the design?

Yes NO

Click “Yes”. Now we need to fully define your custom design (see the panels that follow).

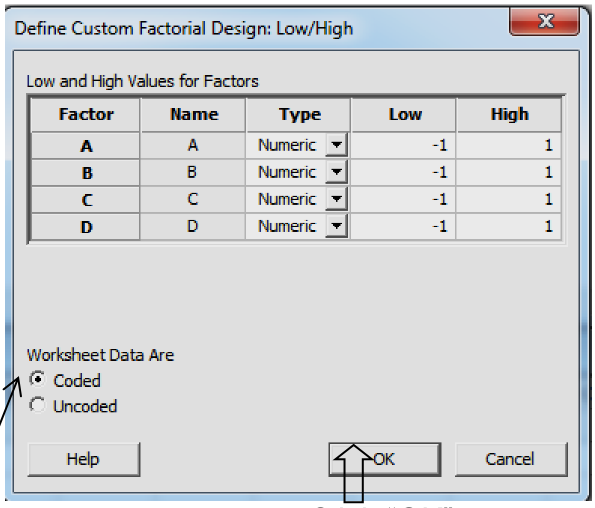
Panel 1

Introduce your Factor columns:



Then Click the "Low/High" tab

Panel 2



Click "OK"

Notes on <u>Panel 2</u>:

(a) Your worksheet data in <u>coded units</u> (i.e. -1 and +1 units instead of the actual units). For your own sanity prompt Minitab so.

(b) All your factors are numeric. So your factor type (Numeric/Text) for each factor is "numeric" (the default selection).

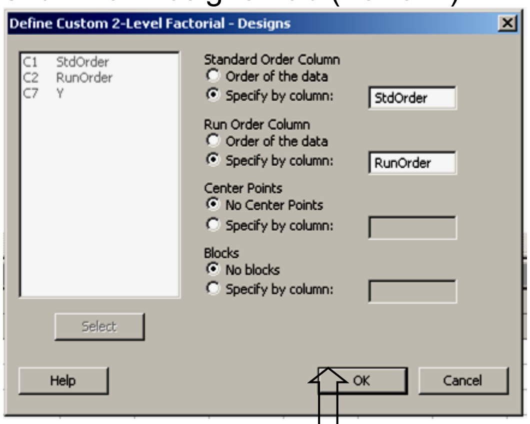
(c) If you know the actual low and high values of each factor you can replace -1 and +1 with these actual values.

(d) You can also insert the names of the factors (e.g. temperature instead of 'A') under "Name".

(e) If you have a qualitative factor, instead of -1 and +1 for low and high values you can type the actual names corresponding to -1 and +1. For example, let us say that our (qualitative) factor is "Sweetener" and you we instated in finding out whether changing from <u>saccharin</u> to <u>stevia</u> affect the Sweetness (Y1) and the bitterness (Y2) of a diet fruit punch (multi-response situations are common in experiments). Then, instead of -1 and +1 you can type names saccharin and stevia respectively.

Panel 3

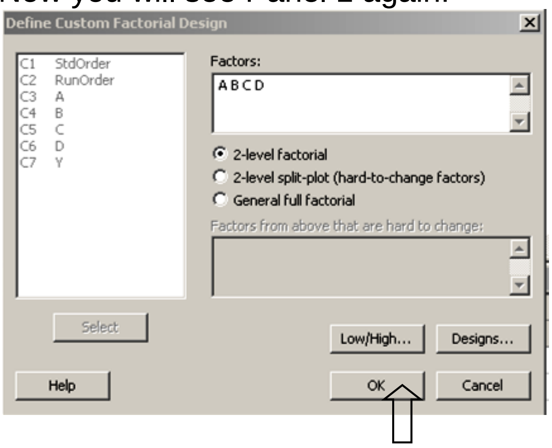
Click the "Designs" tab (Panel 1)



Specify your StdOrder and RunOrder columns and click "OK"

Panel 4

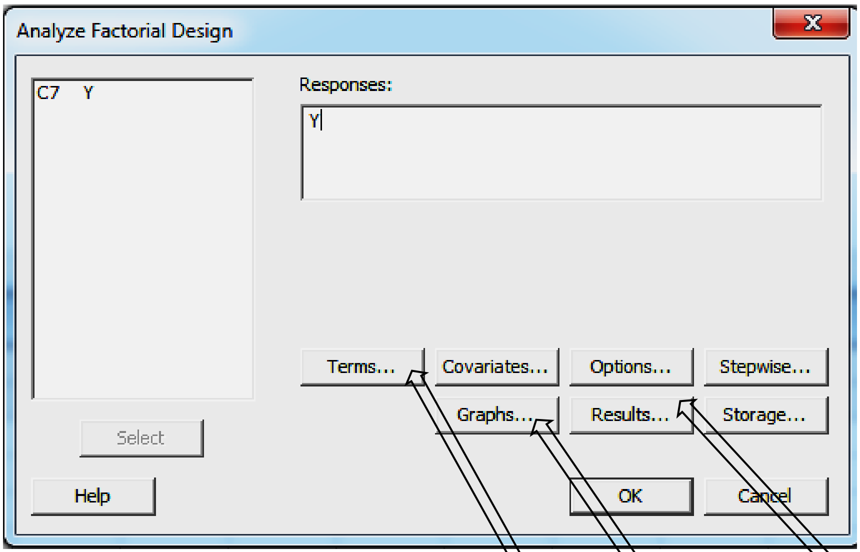
Now you will see Panel 1 again:



Click "OK"

Panel 5

Introduce your Response column (Y)

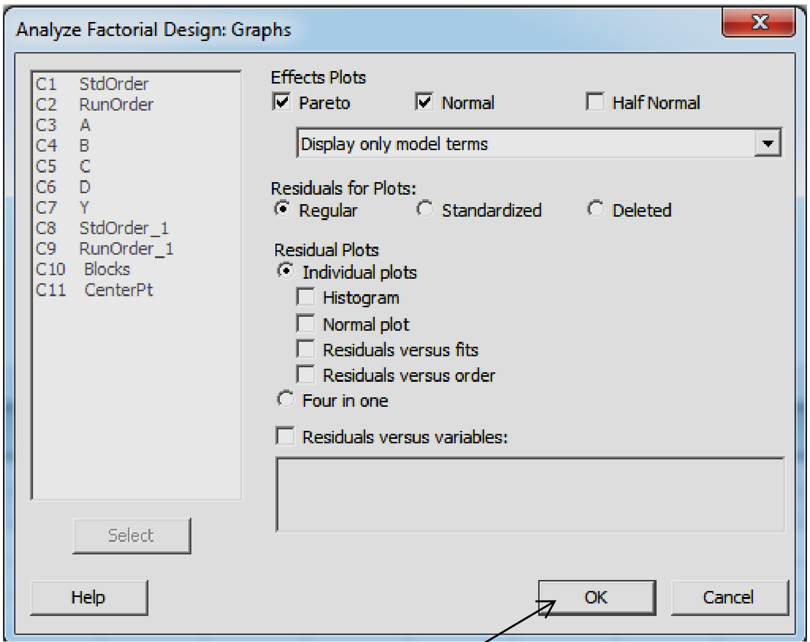


The following tabs are important to you: "Terms" , "Graphs" and "Results" (the other tabs such as the "Covariates", "Stepwise" model building would not apply in your type of factorial designs).

Open the "Terms" tab. You will see that (in the absence of your guidance) Minitab has selected all the terms to go into your model. At this stage let it be as it is. So close that tab.

Now click the "Graph" tab. Then you should see the following panel (Panel 6). (you will know that the regression equation that Minitab gives is not a good one because in reality, only few terms will be significant).

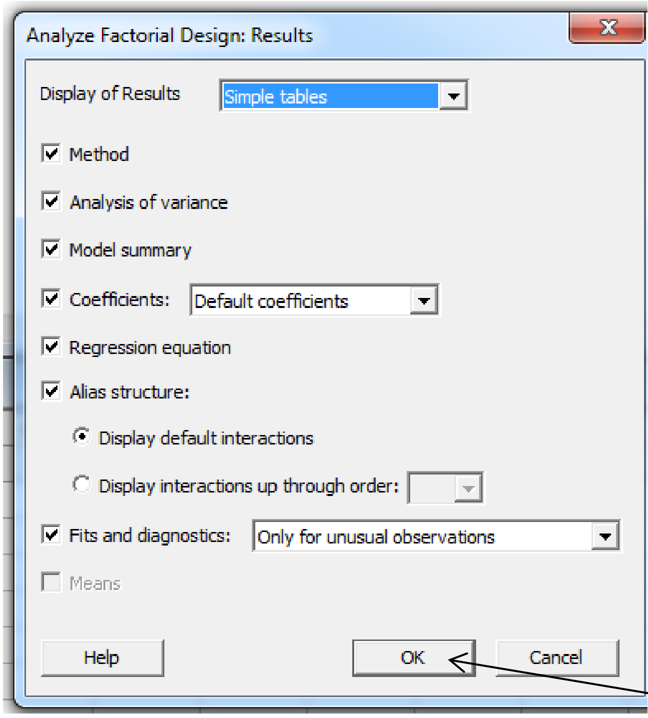
Panel 6



The residual plots are not important at this stage (so do not select any of the residual plots). In fact, you will have no residuals because your model is saturated. The only useful piece of information you need is the **normal probability plot of effects** (see next page) and of course the Pareto chart, if you have chosen it (Minitab selects Pareto chart by default). Now click OK. Then you will return back to Panel 5.

Click the Results tab (Panel 5). Now you will see the following (Panel 7).

Panel 7

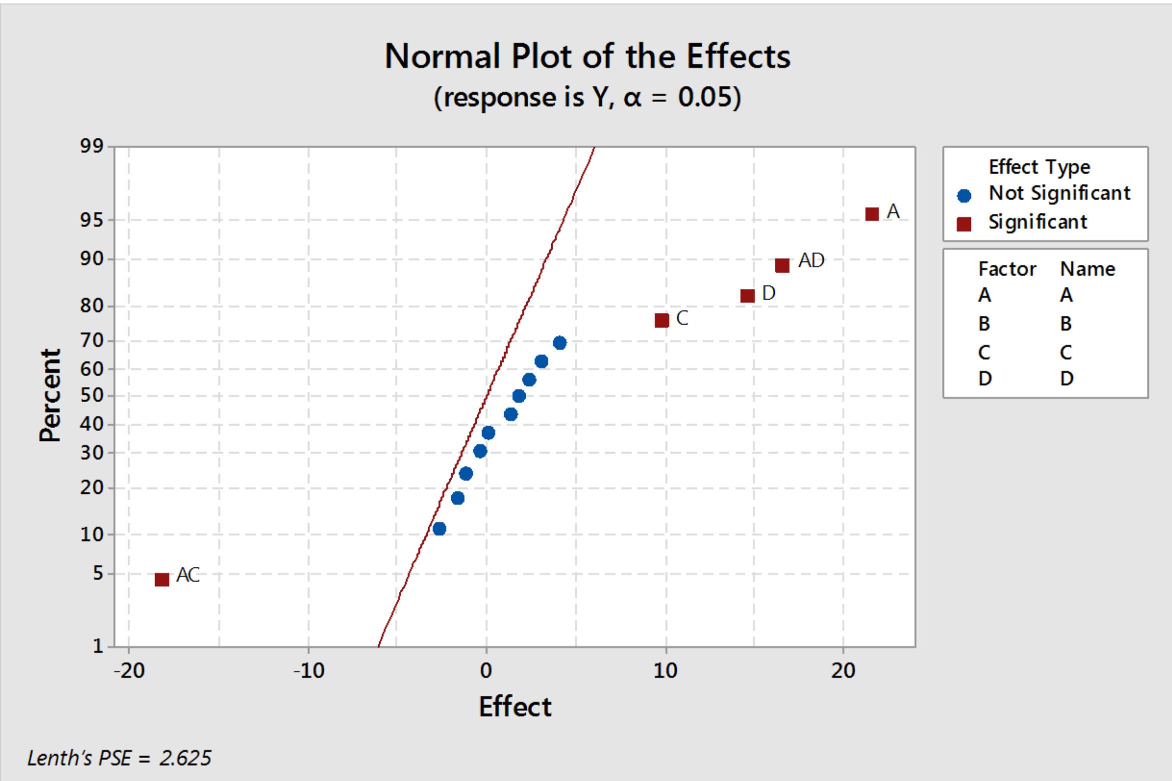


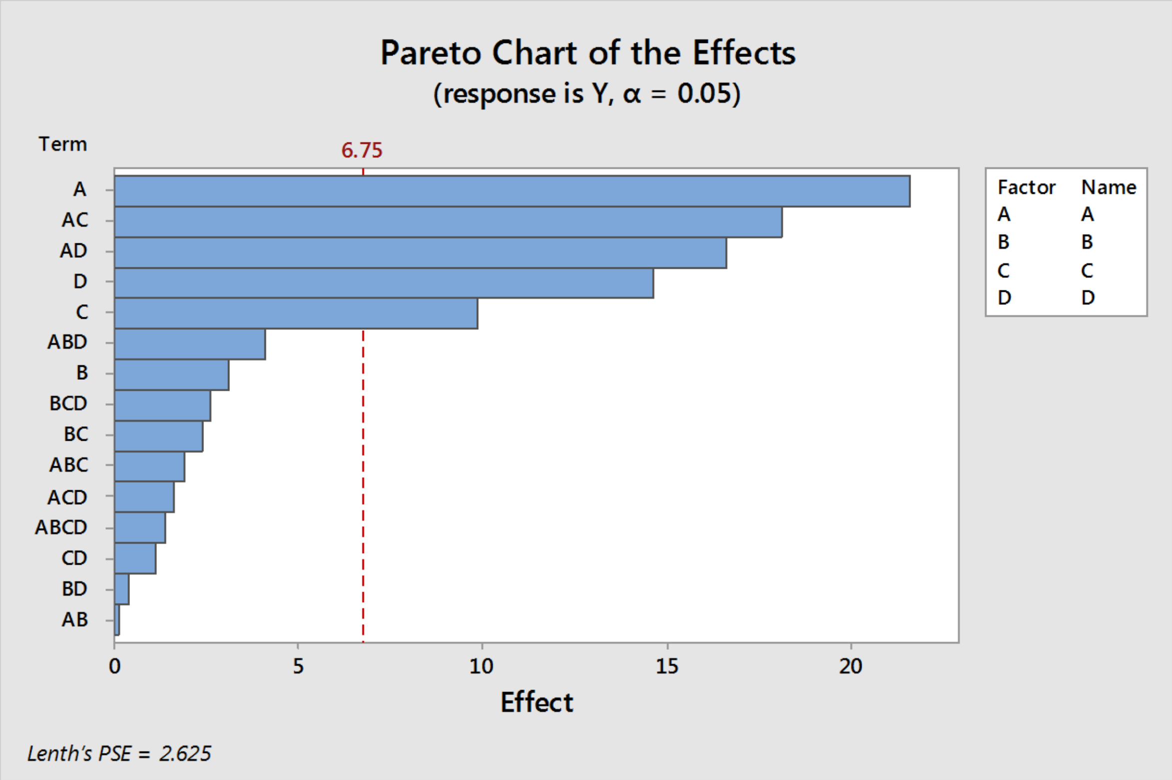
Most of the things Minitab has selected by default to display as results (in the session window) are not relevant or useful at this stage.

For example, the regression equation Minitab produces will have all the terms (because you selected all the terms to go into the model earlier). Such a model is not useful because in reality only few of these terms will be significant (both from a statistical point of view and a practical point of view).

Anyway leave all the default choices (ignore whatever that appears in the session panel at this stage). Now click "OK."

Now you are good to get the initial results. Therefore click OK in the "Analyse Factorial Design" panel (Panel 5). At this stage the only useful Minitab outputs are the normal probability plot of effects (see below) and the Pareto charts (next page).





Based on the normal probability plot of effects and the Pareto chats, which terms do appear to be significant? (Minitab uses the 0.05 level by default)

The significant terms (at 5% significance level) are: **A, C, D are the significant main effects (linear effects)** **AC and AD are the significant interaction effects. Note: We can list all significant effects together – A, C, D, AC and AD or in the descending order like the Pareto chart does.**

Although you can now try to finalise your model (i.e. remove insignificant terms from your model), at this stage you must look for the graphical plots that indicate how the factors work in explaining your response. Hence you have to look for the **factorial plots**—the main effects plots and the interaction plots—and the **cube plot** containing the response values.

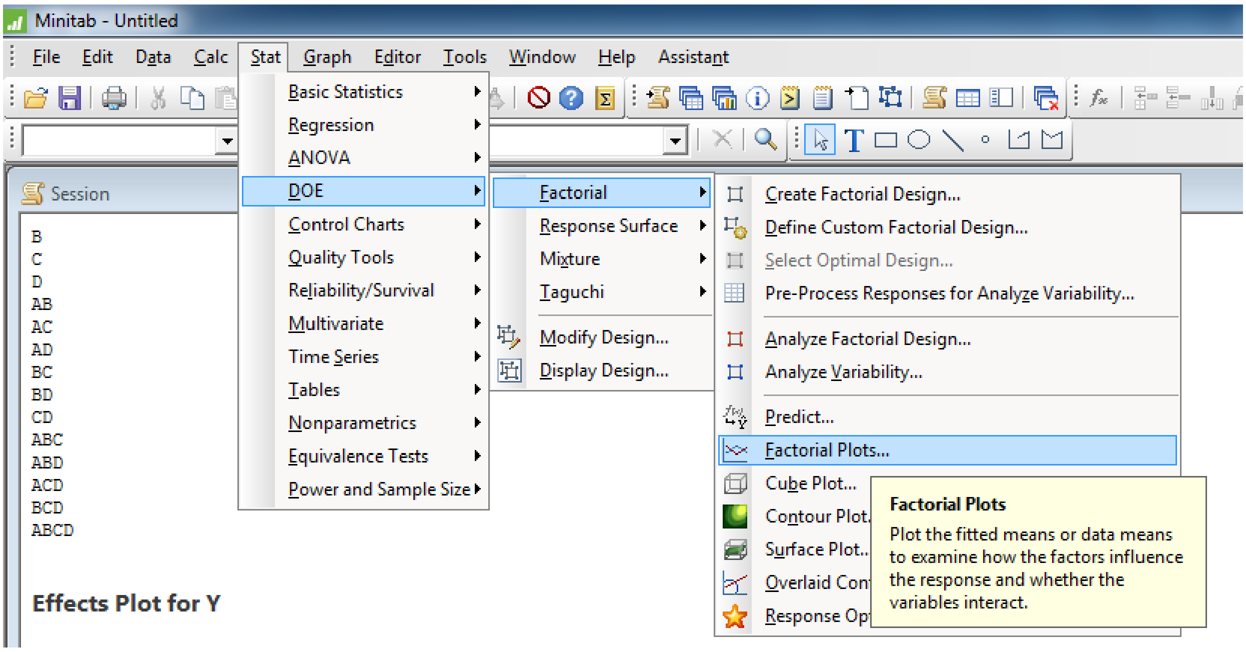
Deriving the Factorial Plots

You can prompt Minitab to show the factorial plot (the main effects plots and the interaction plots).

(3) In the lab prompt Minitab to produce all three factorial plots — the cube plot, the main effects plot and the interaction plot — and attempt to interpret the plots from a practical perspective.

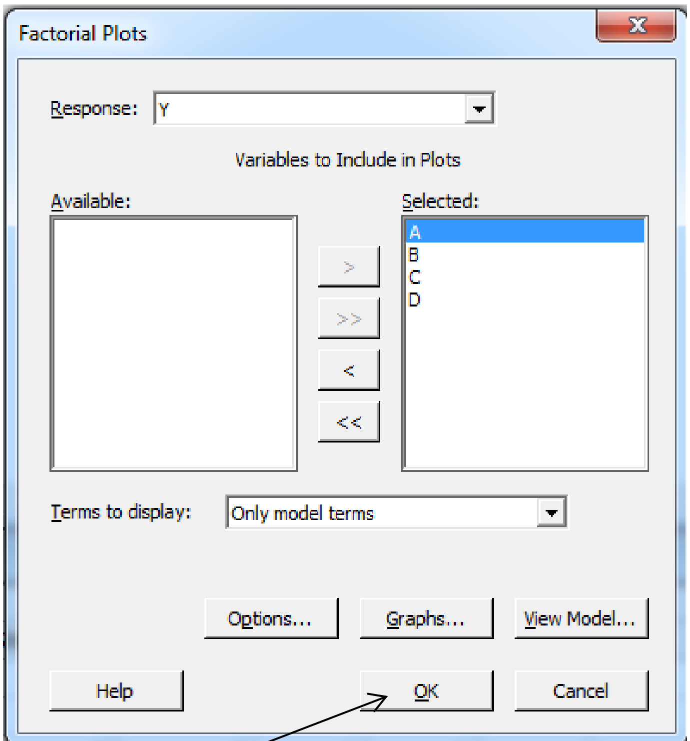
Click Stat → DOE → Factorial → Factorial Plots (see Panel 8)

Panel 8



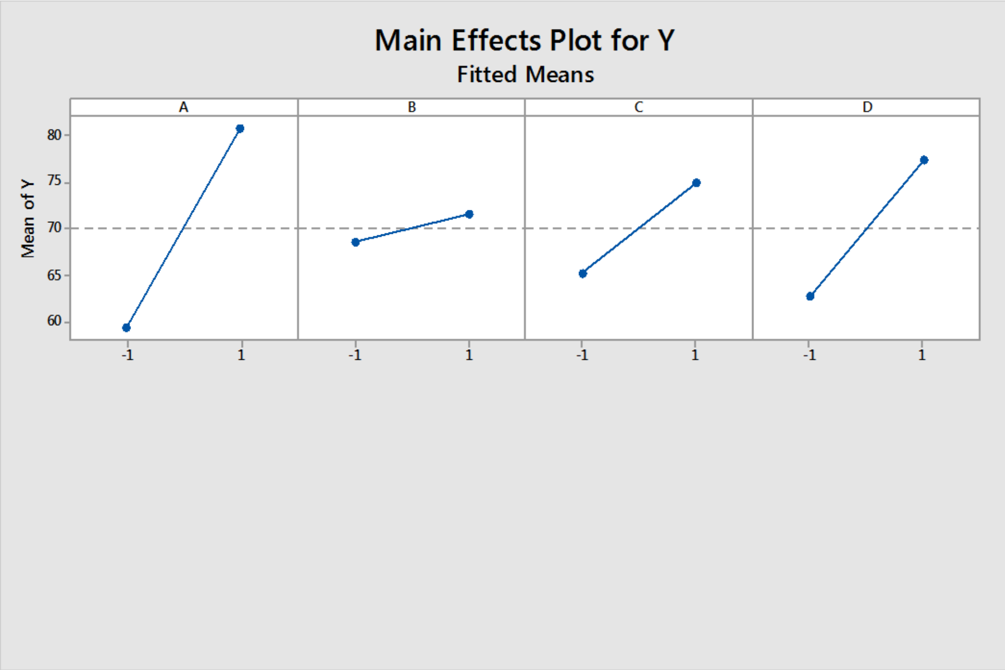
Select the factors that are of interest to you. Select all 4 factors (Panel 9) although you know that factor B does not see to play a role in predicting your response.

Panel 9



Click “OK”.

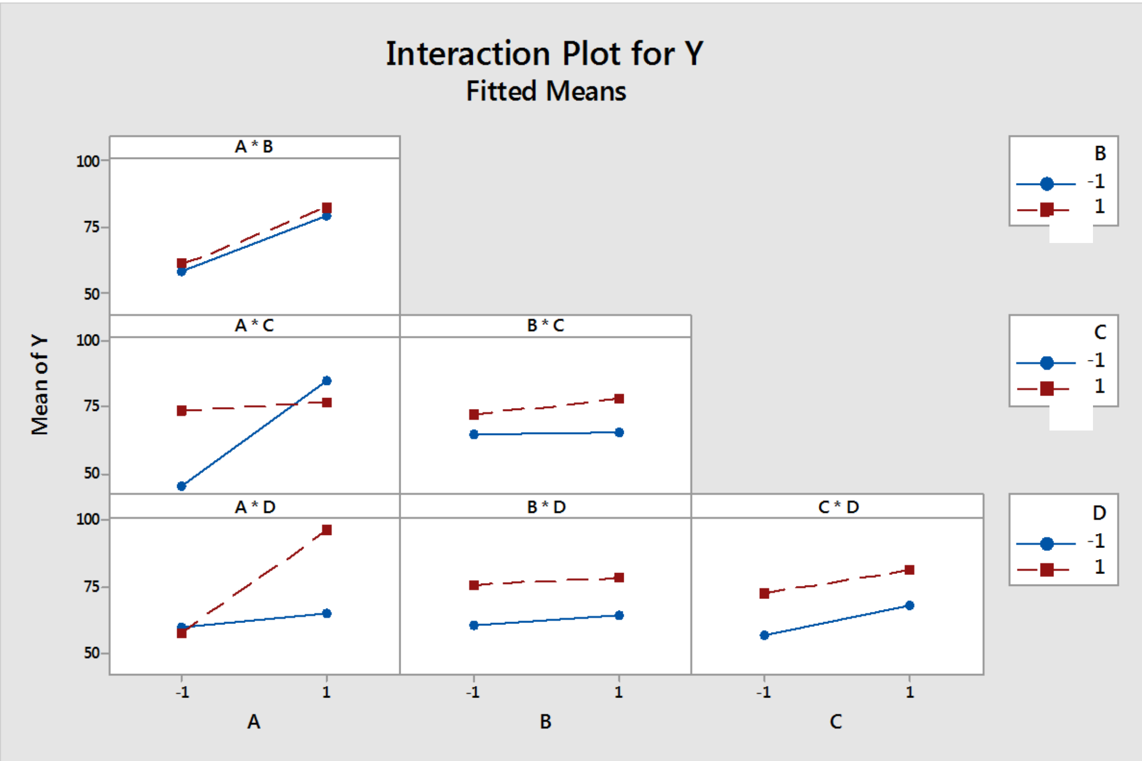
You should now see the following main effects plots and the interaction plots (next page).



Write-down what you can reasonably deduce from the main effects plots.

Factor A is the most influential factor that affects Y (because it has the highest mean change of Y, in other words the highest slope), followed by factors D and C. Factor B has no practically important effect.

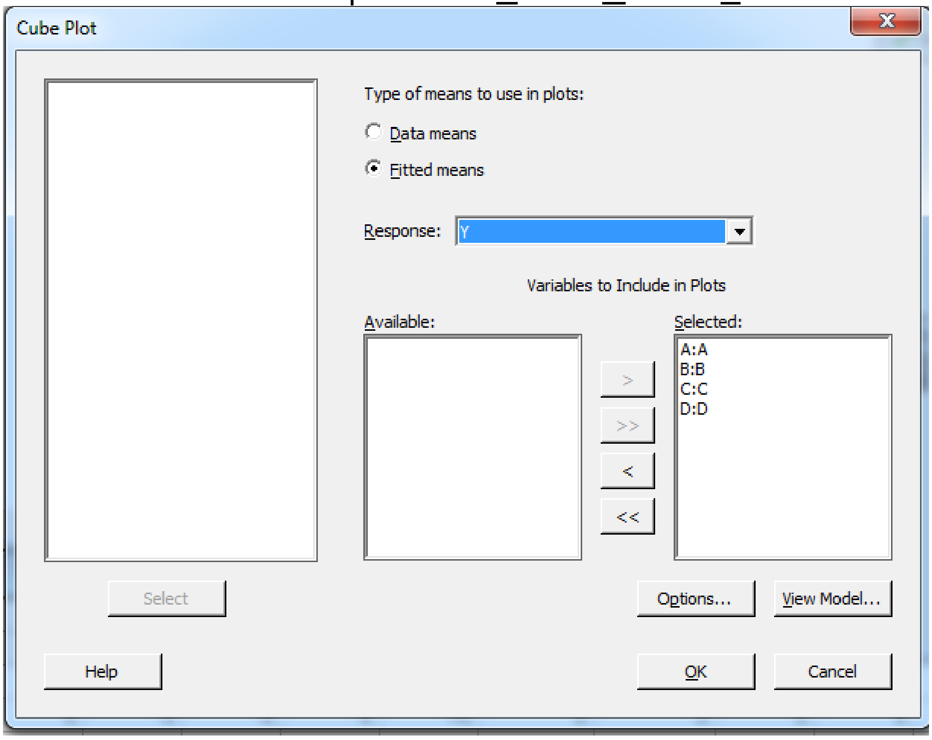
Note: Here we are examining the linear effects of each factor from a practical perspective; the tests for statistical significance of slopes (i.e. regression coefficients) are shown elsewhere.



Write-down what you can reasonably deduce from the interaction plots.

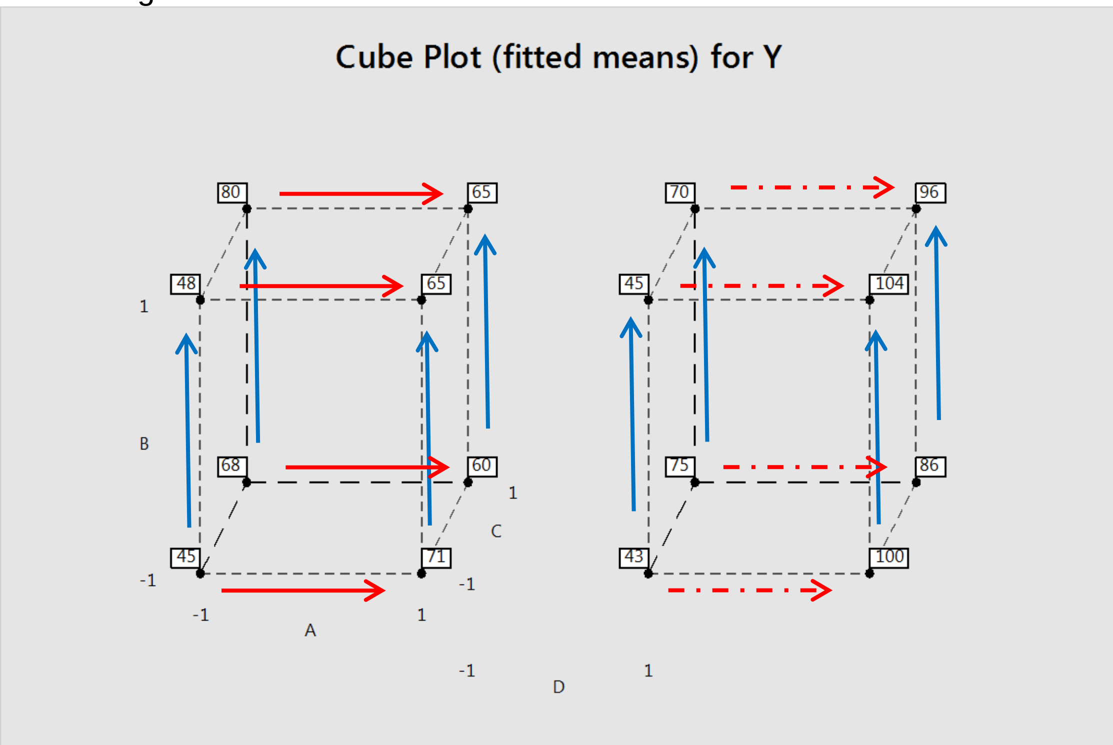
AC and AD seem to be influential interactions (everything else not) as the two lines in AC and AD interaction plots are very un-parallel.

Now call for the cube plot. Click Stat → DOE → Factorial → Cube Plot…



For the cube plot, you can select all 4 terms, although you know for the fact that Factor B does not play a role in explaining Y (if you omit B it is easier to interpret the cube plot because then, you will have only one design cube because you have only 3 factors to work with).

Fitted means are better if you have finalised the model (you have [illegible] means are better if you have not finalised the model. Let us go with the fitted means, Minitab's default selection, although we know that we are having an over-fitted model at this stage.



Write-down what the cube plot seems to be suggesting on the main effects and interactions; *a cube plot is difficult to interpret till you get used to it*; *do not worry.*

For a start, B is not important because the values hardly

change when we move upwards vertically along the vertices of the cube. For other factors the values change, suggesting that these other factors (A, C and D) are important.

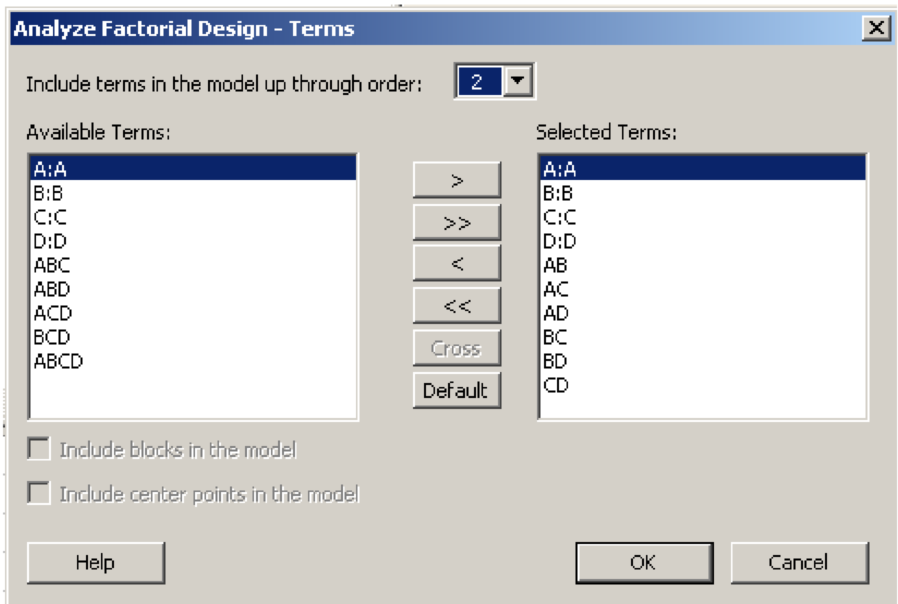
When D is at the +1 setting, changing A from -1 to +1 (follow the red hash lines in the RHS cube) will increase Y considerably. However, when D is at the -1 setting, changing A from -1 to +1 (follow the red lines in the LHS cube) does not make a significant net change in Y [= {(71-40) + (60-68) + (65-48) + (65-80)}/4]. This suggests that there is a significant AD interaction (basically this is what the AD interaction plot shows). A similar argument can be provided in favour of the AC interaction.

Note: Interactions are difficult to identify via a cube for a beginner. The fact of the matter is that the cube plot shows the same information the main effects plot and the interaction plots show (so you will know what to write under the cube plot☺).No need to show arrows and stuff!

You can finalise your model based on the normal probability plot of effects and the other graphical plots (especially the main effects plots and the interaction plots) that you used earlier.

Alternatively, you can exclude the 3-way and 4-way interactions from the default model (in other words, include "terms in the model up through order 2" as shown below) and delete terms that are not significant ($p > 0.05$) based on ANOVA results (see below).

Panel 10



Factorial Regression: Y versus A, B, C, D

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	10	5603.13	560.31	21.92	0.002
Linear	4	3155.25	788.81	30.86	0.001
A	1	1870.56	1870.56	73.18	0.000
B	1	39.06	39.06	1.53	0.271
C	1	390.06	390.06	15.26	0.011
D	1	855.56	855.56	33.47	0.002
2-Way Interactions	6	2447.88	407.98	15.96	0.004

A*B	1	0.06	0.06	0.00	0.962
A*C	1	1314.06	1314.06	51.41	0.001
A*D	1	1105.56	1105.56	43.25	0.001
B*C	1	22.56	22.56	0.88	0.391
B*D	1	0.56	0.56	0.02	0.888
C*D	1	5.06	5.06	0.20	0.675
Error	5	127.81	25.56		
Total	15	5730.94			

Note that this is not the final analysis. At this stage the R^2 value corresponds to the model containing all the terms shown in this ANOVA table).

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.05594	97.77%	93.31%	77.16%

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		70.06	1.26	55.43	0.000	
A	21.62	10.81	1.26	8.55	0.000	1.00
B	3.13	1.56	1.26	1.24	0.271	1.00
C	9.87	4.94	1.26	3.91	0.011	1.00
D	14.62	7.31	1.26	5.79	0.002	1.00
A*B	0.13	0.06	1.26	0.05	0.962	1.00
A*C	-18.13	-9.06	1.26	-7.17	0.001	1.00
A*D	16.63	8.31	1.26	6.58	0.001	1.00
B*C	2.37	1.19	1.26	0.94	0.391	1.00
B*D	-0.37	-0.19	1.26	-0.15	0.888	1.00
C*D	-1.13	-0.56	1.26	-0.45	0.675	1.00

Regression Equation in Uncoded Units

```
Y = 70.06 + 10.81 A + 1.56 B + 4.94 C + 7.31 D + 0.06 A*B - 9.06 A*C + 8.31 A*D
+ 1.19 B*C
    - 0.19 B*D - 0.56 C*D
```

Note that the above regression equation is not the final model as it contains insignificant terms. Finalise the model by eliminating insignificant terms from the “selected terms” box (Panel 10).

Analysis of Variance for Y (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	3116.19	3116.19	1038.73	53.23	0.000
A	1	1870.56	1870.56	1870.56	95.86	0.000
C	1	390.06	390.06	390.06	19.99	0.001
D	1	855.56	855.56	855.56	43.85	0.000
2-Way Interactions	2	2419.62	2419.62	1209.81	62.00	0.000
A*C	1	1314.06	1314.06	1314.06	67.34	0.000
A*D	1	1105.56	1105.56	1105.56	56.66	0.000
Residual Error	10	195.12	195.12	19.51		
Lack of Fit	2	15.62	15.62	7.81	0.35	0.716
Pure Error	8	179.50	179.50	22.44		
Total	15	5730.94				

Let us write-down the final model:

$$Y = 70.06 + 10.81*A + 4.94*C + 7.31*D - 9.06*AC + 8.31*AD + \varepsilon$$

Note that the equation is valid in coded units (-1, +1) only. Poor Minitab thinks that you have entered the actual uncoded (actual) values in Panel 2, under low and high values.

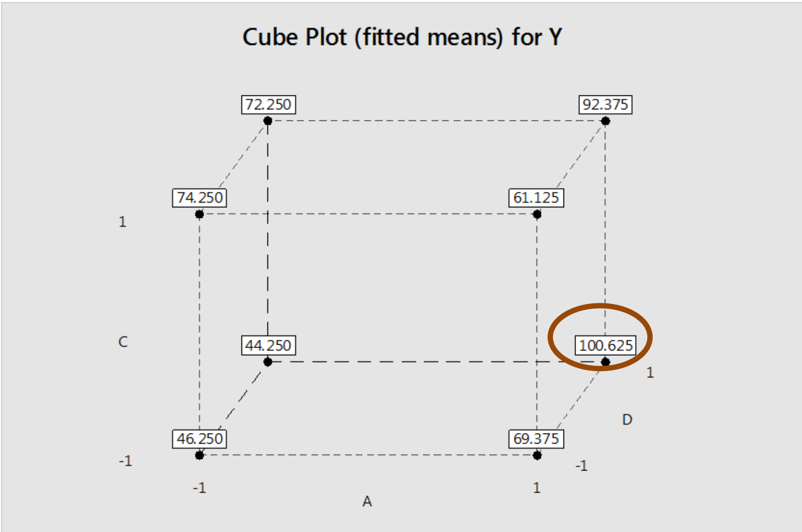
By looking at the above equation, can you can you work out that A, C and D should be set at +1, -1 and +1 respectively, to maximise the filtration rate Y? B is statistically and practically insignificant.

Are the above settings consistent with what is shown in the cube plot? It is better to get the cube plot with factors A, C and D only (with fitted means).

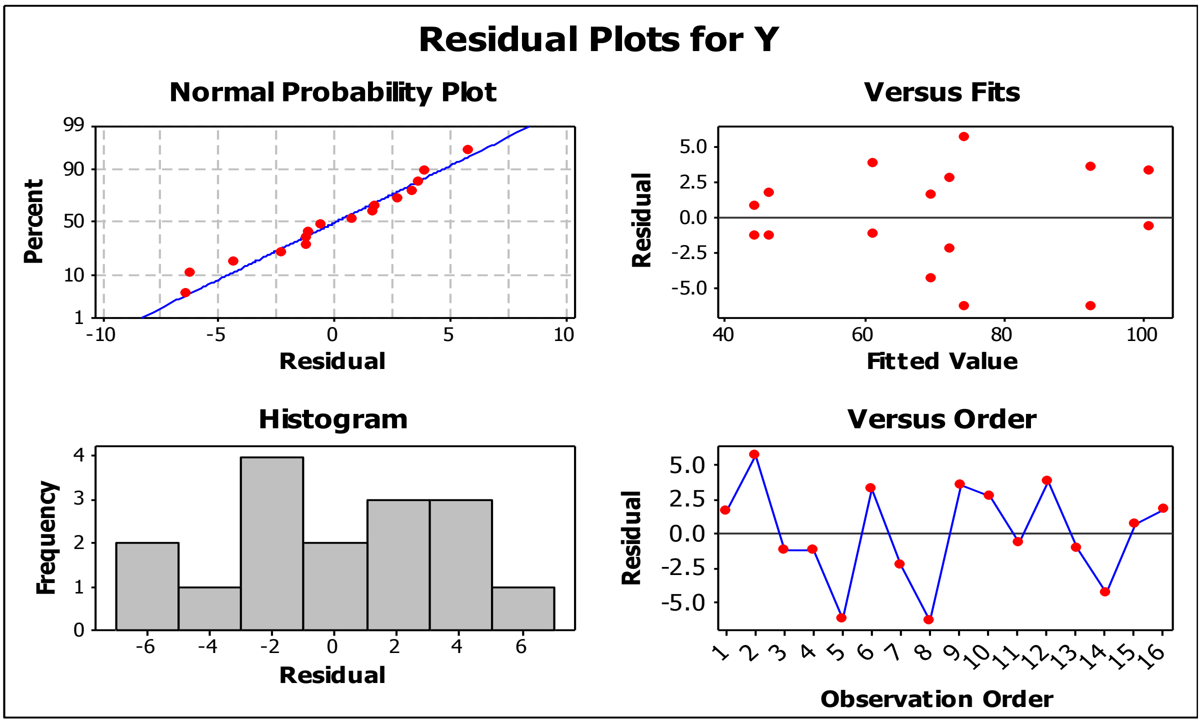
Calculate the fitted Y at the optimum parameter setting: A = +1, C = -1 and D = +1. Note that the fitted Y is nothing but the expected value of Y (Minitab calls this the fitted mean).

The calculated fitted mean is: **70.06 + 10.81*(1) + 4.94*(-1) + 7.31*(1) - 9.06*(1)(-1) + 8.31*(1)(1) = 100.61**

Minitab's fitted mean estimate (shown in the cube plot) for the above optimum setting is: ...**100.625**.... (both should be the same). **Note: You Need to reconstruct the cube plot (see below) after removing factor B as well as all other insignificant terms (if you have done it already remove factor B as well as all interactions except AD and AC) from the terms box to re-estimate the fitted values.**



Now, click the "Graphs" tab (see Panel 5 in page 5) to call for the good old four-in-one residual plot for the final model (by this stage you will only have the significant terms in the terms selection box, so you know that what Minitab throws-out is the residual plot for the final model).



Finally, discuss the model adequacy

The normal probability plot of residuals shows that… **Residuals seem to have come from a normal population (slight crookedness is fine. So normality √**

The histogram shows that…. **The distribution is roughly symmetric (not skewed) and somewhat bell shaped, thus showing further support for normality.**

The versus fit graph shows that…. **The variance of the residuals are roughly equal throughout the X-Y relationship (I am using X to mean the two factors) except in the case of small values of Y. Can say that homoscedasticity assumption is not seriously threatened. So homoscedasticity assumption √**

The versus order graph shows that…. **The residuals are independent (no apparent non-random pattern such as a trend). So independence assumption √**

Some important properties of 2-level multi factor designs came into play in this exercise. Since all terms in the design are un-correlated adding or removing one term does not affect the contribution of the other terms in the model (i.e. their effects and regression coefficients remained unchanged).

On the same token the shape of the main effects plots and interaction plots remain the same irrespective of what you do to your model (this is the reason why it does not matter at what stage you have a look at these plots; they remain the same always).

Likewise, the sequence in which you add or remove terms does not affect the *p* values of the terms in the model.

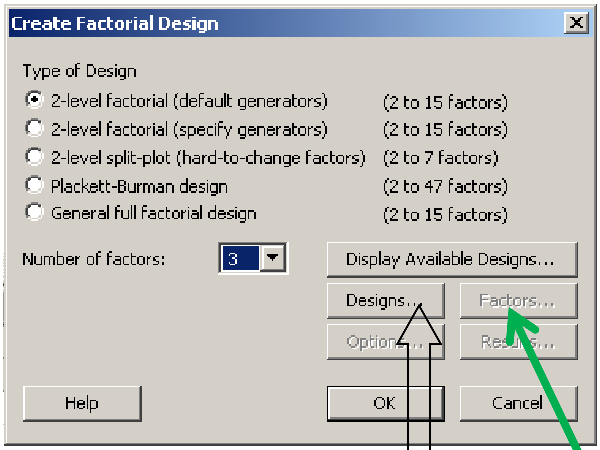
Exercise 2_Part 1: Creating a Replicated Full Factorial Design

Create a 3 factor (2-level) full factorial design with two replicates in Minitab to plug in response data (allow Minitab to use its random number generator to randomise the runs).

Start from Stat → DOE → Factorial → Create Factorial Design

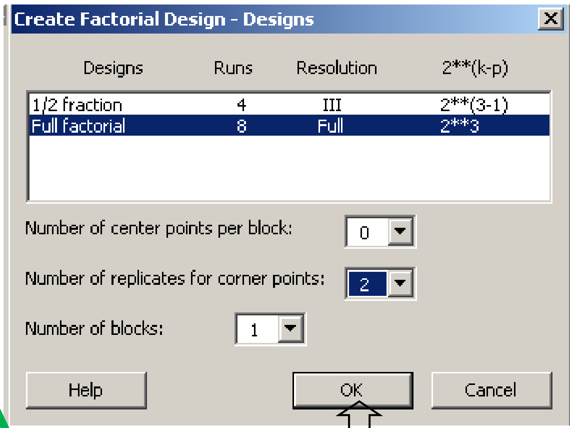
Some useful screen shots are given below

Select number of factors as "3"



Click "Designs"

Select "Full Factorial" and **Number of replicates as "2";** assume that you will run the experiment in one block and that there are no centre points.



Click "OK" and then click OK (previous slide). You are done!

You should see the following in the session panel:

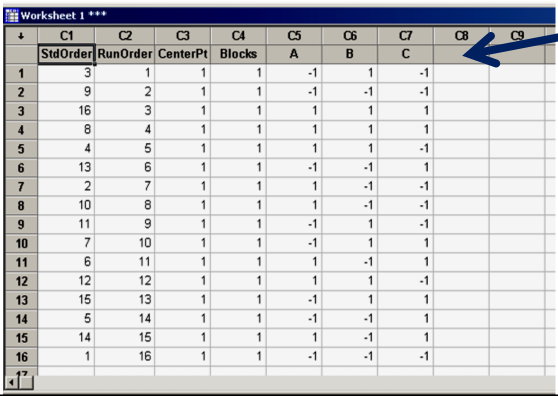
```
Welcome to Minitab, press F1 for help
```

Full Factorial Design

```
Factors:    3   Base Design:         3, 8
Runs:      16   Replicates:             2
Blocks:     1   Center pts (total):     0

All terms are free from aliasing.
```

Note: In doing Part 1, it is better for you to do this exercise keeping Part 2 of the exercise in mind and type the Factor Names and their actual settings (uncoded values) at the setup stage (so click Factors) instead of leaving the default -1 and +1 (coded) settings as they are.

Your Minitab worksheet should see something like this:


Worksheet 1 ***

↓	C1	C2	C3	C4	C5	C6	C7	C8	C9
	StdOrder	RunOrder	CenterPt	Blocks	A	B	C		
1	3	1	1	1	-1	1	-1		
2	9	2	1	1	-1	-1	-1		
3	16	3	1	1	1	1	1		
4	8	4	1	1	1	1	1		
5	4	5	1	1	1	1	-1		
6	13	6	1	1	-1	-1	1		
7	2	7	1	1	1	-1	-1		
8	10	8	1	1	1	-1	-1		
9	11	9	1	1	-1	1	-1		
10	7	10	1	1	-1	1	1		
11	6	11	1	1	1	-1	1		
12	12	12	1	1	1	1	-1		
13	15	13	1	1	-1	1	1		
14	5	14	1	1	-1	-1	1		
15	14	15	1	1	1	-1	1		
16	1	16	1	1	-1	-1	-1		

Type the label "Y" (your response) in column C8 and plug in the Y values when they become available.

Make sure that you save your worksheet (and the project if you do not want to use your previous *.MPJ file).

You can replace the labels A, B, C and Y with their actual names if you wish.

Exercise 2_Part 2: Analysing a Replicated Full Factorial Design

A researcher is interested in investigating the effects of three factors on the yield (% of base material converted) of monomer for an adhesive formulation. The primary objective of the experiment is to identify the influential factors that affect the yield and to determine the factor combination that maximises the yield, given the range in which the factors are manipulated. Identifying the factor combination that minimises the process variation was the secondary objective. The researcher has resources to run 16 trials; s/he ran a 2^3 factorial experiment (i.e. a three factor full factorial experiment) with two replicates. Factors and levels are shown in Table 2 below.

Table 2: Factors and levels

Factor	Factor name	Low (-)	High (+)
A	Temperature	160^{o}C	180^{o}C
B	Catalyst level	20%	40%
C	Catalyst type	vendor E	vendor J

The researcher identified few background variables that could potentially affect the results. However, in the laboratory, the researcher was able to hold the levels of the background variables 'constant' during the course of the experiment. As such the researcher was able to run all 16 runs in one block. The Yield values recorded by the researcher are shown in Table 3 below.

Table 3: The Design Matrix and Response Data

Trial Code	Factors			Yield (Y)	
	A	B	C	Replicate 1	Replicate 2
(1)	-1	-1	-1	58	60
a	+1	-1	-1	74	70
b	-1	+1	-1	43	51
ab	+1	+1	-1	75	73
c	-1	-1	+1	56	60
ac	+1	-1	+1	74	78
bc	-1	+1	+1	46	44
abc	+1	+1	+1	78	80
Note: The rows containing data correspond to StdOrder 1 through to 8 (Replicate 1) and 9 through to 16 (Replicate 2).					

Tasks to be done in the Lab:

- Type the response values (Table 3) in column C8 of the Minitab worksheet that you created in <u>Part 1 of Exercise 2</u> (see the previous page). Make sure that you select the right factor combination corresponding to a particular Y value.

- **Analyse the data** as fully as you can (show all relevant Minitab outputs along with your findings) and make actionable conclusions.

  Note that you will have to create a separate worksheet containing only 8 runs, if you are also analysing the statistical significance of the process variation — for example, the range of Y (ΔY), being the absolute difference of the Y value between replicate 1 and replicate 2, for a given factor setting.

PREPARING YOU FOR EXERCISE 2_Part 2

So you would have set up your design matrix by typing in the factor name, data type (numeric vs text) and actual low and high levels as shown below. It is always nice to have Factor Labels and actual units in your graphical outputs.

Create Factorial Design: Factors

Factor	Name	Type	Low	High
A	Temperature	Numeric	160	180
B	Catalyst Lev	Numeric	0.20	0.40
C	Catalyst Typ	Text	E	J

Help OK Cancel

Unless you tell Minitab otherwise, she'd give you the experimental runs in a random order (randomisation is highly desired). Now chuck in the Y values (do this for your DOE project). Because the factor levels are now shown in actual units, you may wonder how I chucked in the Y (Yield) values so easily. Here is the thing. Table 3 of this exercise gives Y values in the standard order (StdOrder), which starts from 1 to 8 and then stars recounting from 9 to 16 (replicate 2). So you will know what value to plug in for a given row by looking at the StdOrder column.

Session

4/14/2015 4:54:51 PM

Welcome to Minitab, press F1 for help.

Full Factorial Design

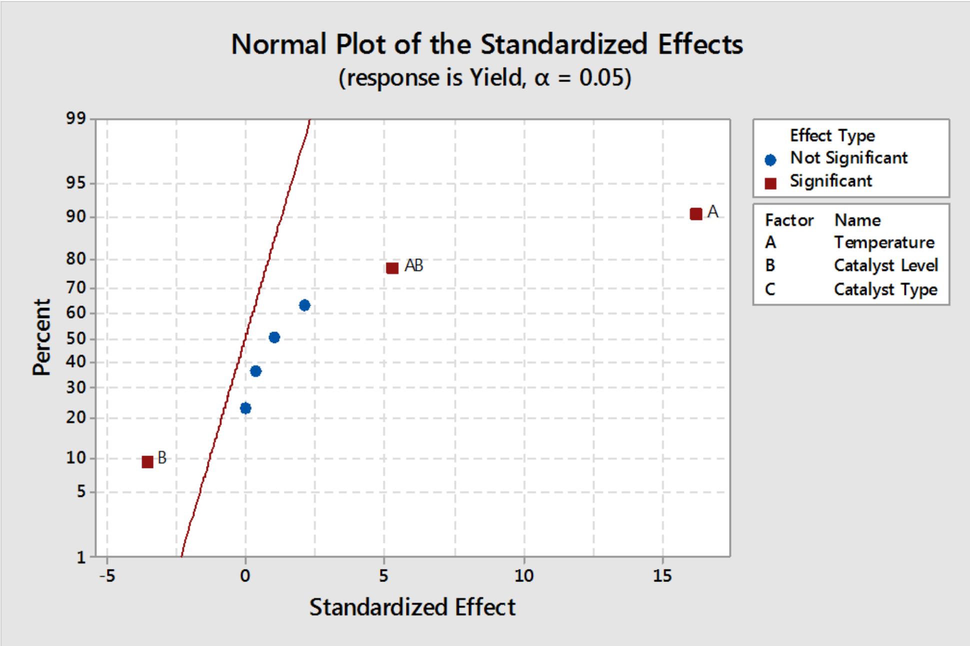
```
Factors:   3   Base Design:          3, 8
Runs:     16   Replicates:              2
Blocks:    1   Center pts (total):      0

All terms are free from aliasing.
```

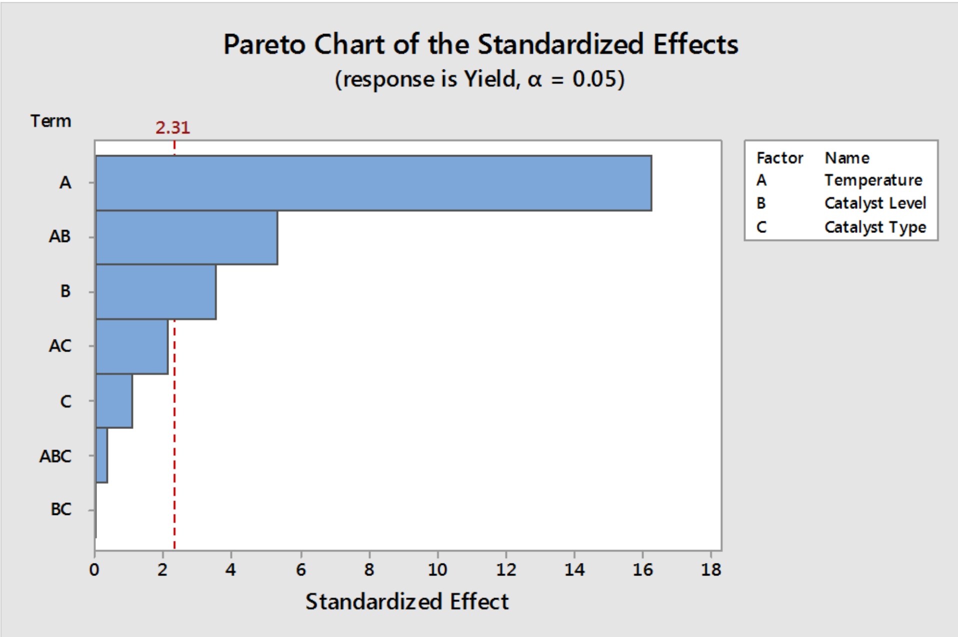
Worksheet 1 ***

↓	C1	C2	C3	C4	C5	C6	C7-T	C8	C9	C10
	StdOrder	RunOrder	CenterPt	Blocks	Temperature	Catalyst Level	Catalyst Type	Yield		
2	6	2	1	1	180	0.2	J	74		
3	16	3	1	1	180	0.4	J	80		
4	1	4	1	1	160	0.2	E	58		
5	12	5	1	1	180	0.4	E	73		
6	15	6	1	1	160	0.4	J	44		
7	4	7	1	1	180	0.4	E	75		
8	3	8	1	1	160	0.4	E	43		
9	7	9	1	1	160	0.4	J	46		
10	8	10	1	1	180	0.4	J	78		
11	9	11	1	1	160	0.2	E	60		
12	2	12	1	1	180	0.2	E	74		
13	11	13	1	1	160	0.4	E	51		
14	13	14	1	1	160	0.2	J	60		
15	10	15	1	1	180	0.2	E	70		
16	14	16	1	1	180	0.2	J	78		

Analysing the Data:

Part A: Understanding how the Yield (Y) relates to the factors





In this case (because of the replicated data set) Minitab can estimate the error variance even if it is asked to include all the terms in the model. Minitab does the job faithfully and reports the standardised effects (the *T* statistics of the model coefficients (= Effect/2) instead of the raw effects for the X axis. It is very clear from the NPP and the Pareto Charts (no need to have both, the NPP is enough) that A, B and AB are significant terms that affect the yield. ANOVA (below) also confirms this.

Factorial Regression: Yield versus Temperature, Catalyst Level, Catalyst Type

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	7	2487.00	355.29	44.41	0.000
Linear **(as a whole)**	3	2225.00	741.67	92.71	0.000
Temperature	1	2116.00	2116.00	264.50	0.000
Catalyst Level	1	100.00	100.00	12.50	0.008
Catalyst Type	1	9.00	9.00	1.13	0.320
2-Way Interactions **(as a whole)**	3	261.00	87.00	10.87	0.003
Temperature*Catalyst Level	1	225.00	225.00	28.12	0.001
Temperature*Catalyst Type	1	36.00	36.00	4.50	0.067
Catalyst Level*Catalyst Type	1	0.00	0.00	0.00	1.000
3-Way Interactions **(as a whole; Coz we have only 1 term)**	1	1.00	1.00	0.12	0.733
Temperature*Catalyst Level*Catalyst Type	1	1.00	1.00	0.12	0.733
Error	8	64.00	8.00		
Total	15	2551.00			

Here the model coefficients are given by Minitab to suit coded units (i.e. -1 and +1 units). Down below, in the equation, the coefficients are given by Minitab to uncoded units (actual units).

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.82843	97.49%	95.30%	89.96%

Coded Coefficients

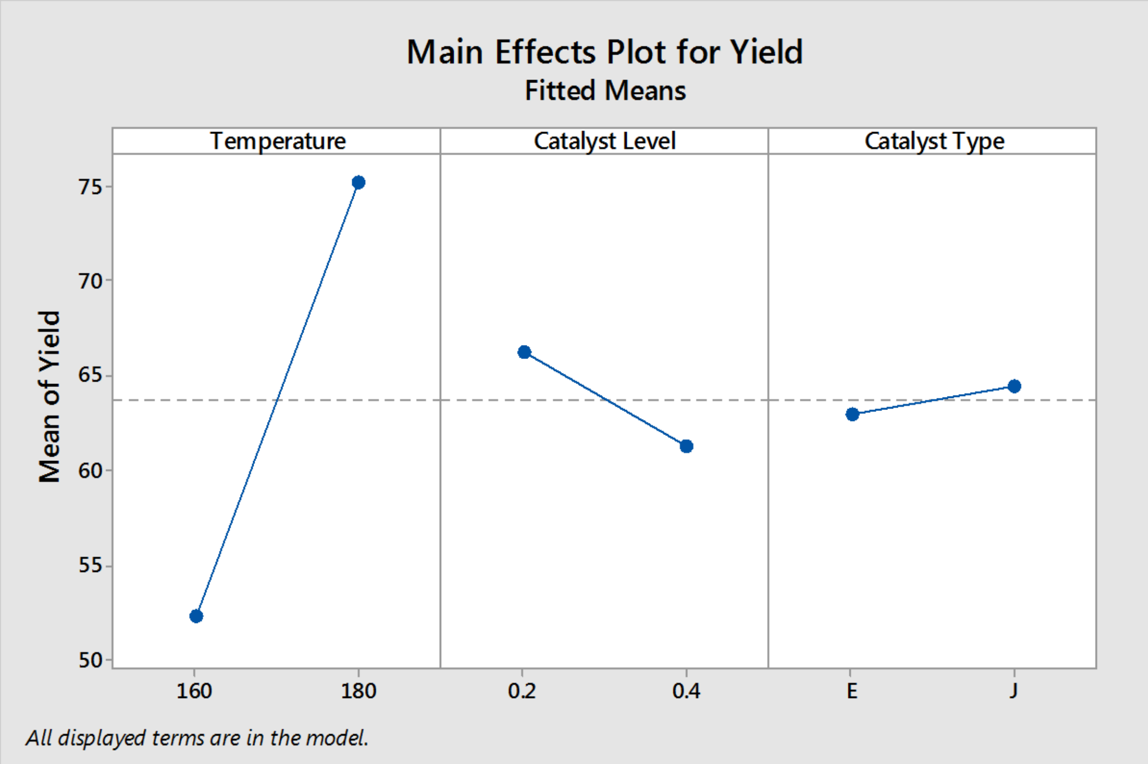
Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		63.750	0.707	90.16	0.000	
Temperature	23.000	11.500	0.707	16.26	0.000	1.00
Catalyst Level	-5.000	-2.500	0.707	-3.54	0.008	1.00
Catalyst Type	1.500	0.750	0.707	1.06	0.320	1.00
Temperature*Catalyst Level	7.500	3.750	0.707	5.30	0.001	1.00
Temperature*Catalyst Type	3.000	1.500	0.707	2.12	0.067	1.00
Catalyst Level*Catalyst Type	0.000	0.000	0.707	0.00	1.000	1.00
Temperature*Catalyst Level*Catalyst Type	0.500	0.250	0.707	0.35	0.733	1.00

Regression Equation in Uncoded Units

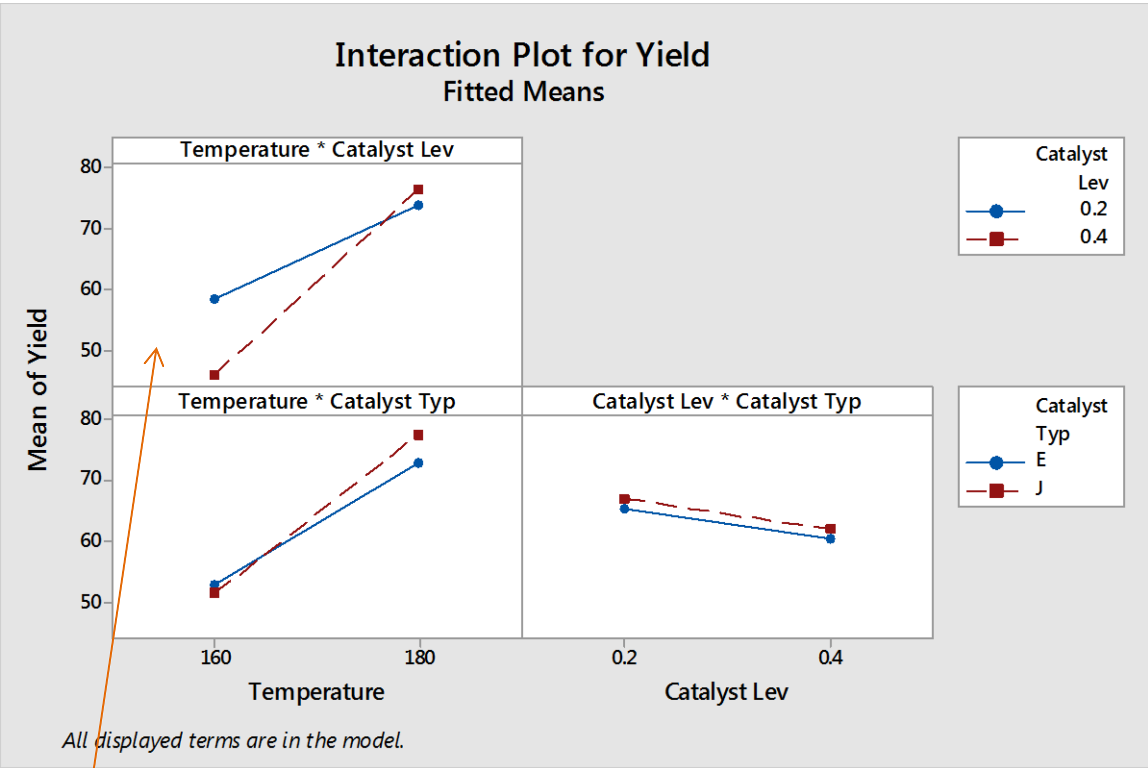
Yield = 67.0 + 0.025 Temperature - 662 Catalyst Level - 12.0 Catalyst Type
+ 3.750 Temperature*Catalyst Level + 0.075 Temperature*Catalyst Type
- 43 Catalyst Level*Catalyst Type
+ 0.250 Temperature*Catalyst Level*Catalyst Type

Lecturer's Comment: The above equation is no good at this stage (it is not parsimonious and contains several terms that are not predictors in the true/population model)

Now, let us have a look at the effects plots and the cube plots (I have asked Minitab to show the data means, that is Y averages at the corner points).



Temperature is the most influential factor (or can say that temperature has the strongest linear effect on Y). The Catalyst Level (Factor B) is the other factor that can be manipulated (to change Y) but the Catalyst Type (Factor C has) no practically significant effect on Y. The main effect plot of Temperature (factor A) shows that increasing the temperature increases the yield (on average). The plot also shows that increasing the catalyst level (Factor B) decreases the yield (on average).



The interaction plot suggests that the Temperature*Catalyst interaction is an influential 2-way interaction. The Temperature*Catalyst plot also shows that the temperature has

to be at the higher setting (180) to get a higher yield. Also the plot shows that the catalyst level should be at the lower setting to get the maximum yield. This is what the main effects plot also says.

We can also conclude the above from the final regression model shown below (or the cube plot with the fitted values).

Finalising the model and verifying model adequacy (4 in 1 plot, R^2 bla bla)

Factorial Regression: Yield versus Temperature, Catalyst Level

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	3	2441.00	813.67	88.76	0.000
Linear	2	2216.00	1108.00	120.87	0.000
Temperature	1	2116.00	2116.00	230.84	0.000
Catalyst Level	1	100.00	100.00	10.91	0.006
2-Way Interactions	1	225.00	225.00	24.55	0.000
Temperature*Catalyst Level	1	225.00	225.00	24.55	0.000
Error	12	110.00	9.17		
Lack-of-Fit	4	46.00	11.50	1.44	0.306
Pure Error	8	64.00	8.00		
Total	15	2551.00			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.02765	95.69%	94.61%	92.33%

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		63.750	0.757	84.22	0.000	
Temperature	23.000	11.500	0.757	15.19	0.000	1.00
Catalyst Level	-5.000	-2.500	0.757	-3.30	0.006	1.00
Temperature*Catalyst Level	7.500	3.750	0.757	4.95	0.000	1.00

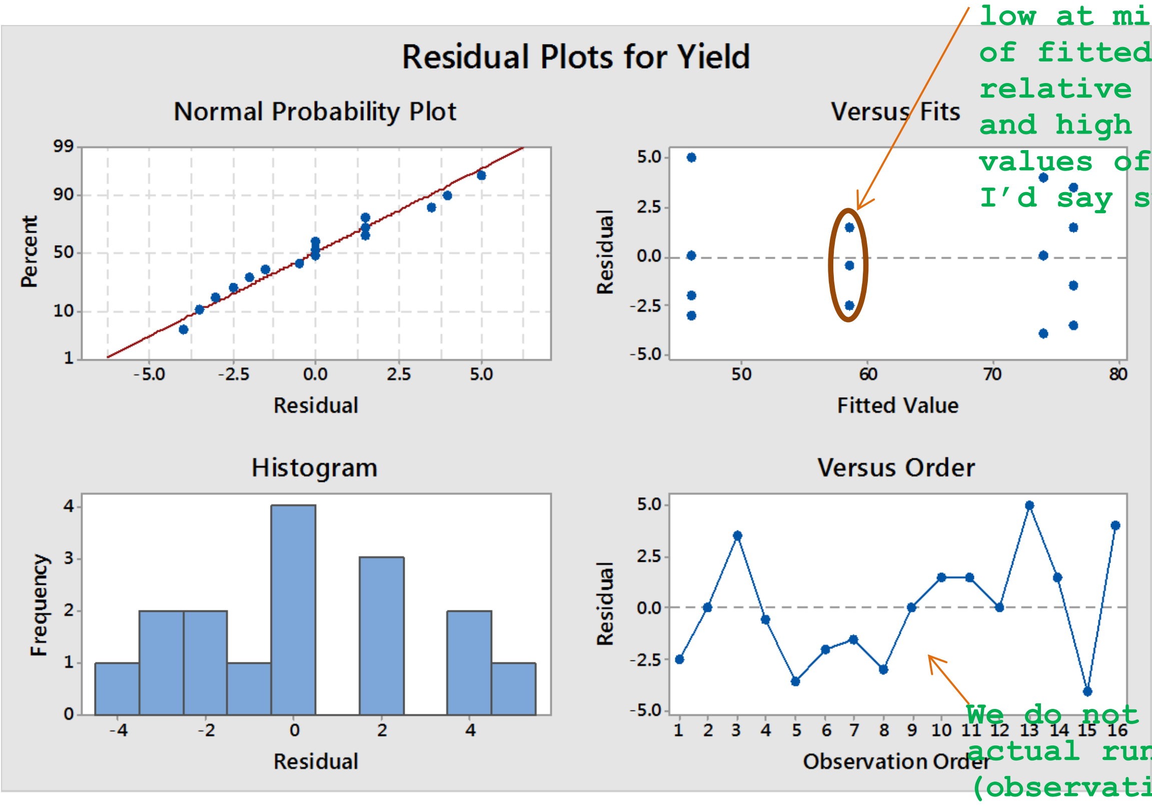
Regression Equation in Uncoded Units **<< The Final Model**

Yield = 67.0 + 0.025 Temperature - 662 Catalyst Level + 3.750 Temperature*Catalyst Level

The above equation shows that we get the max Y when the temperature at the higher setting, the catalyst level at the lower setting (= 0.4). Factor C is unimportant (within the context of the experiment it is just a source of noise).

The R^2 of the model is high are the residual plots (below) are OK. So all hunky dory ☺. Residuals versus observation order plot has to be probably ignored because we do not know the actual run order (observation order).

Variance is slight low at mid values of fitted *Y* relative to low and high fitted values of *Y* but I'd say sweet as!



We do not have the actual run order (observation order) to write a thing or two about this plot.

Part B: Understanding how ΔY (the range of Y) relates to the factors

Creating the new design matrix (worksheet #2); this time I've asked Minitab not to randomise the runs (of course there is no reason for me to do that ☺)

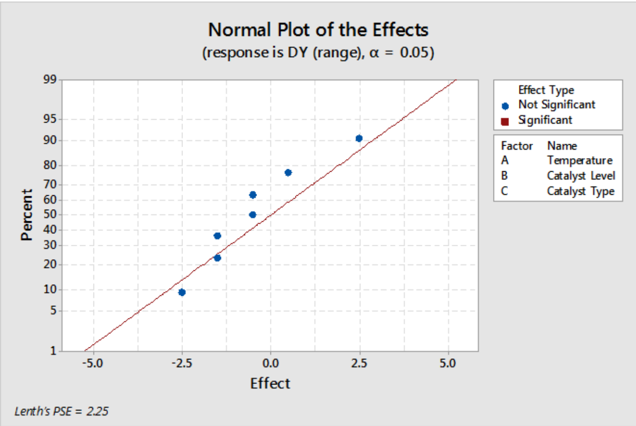
Results for: Worksheet 2

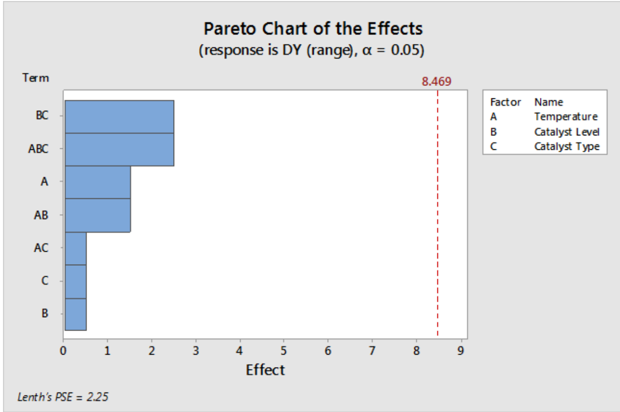
Full Factorial Design

```
Factors:  3   Base Design:         3, 8
Runs:     8   Replicates:             1
Blocks:   1   Center pts (total):     0
```

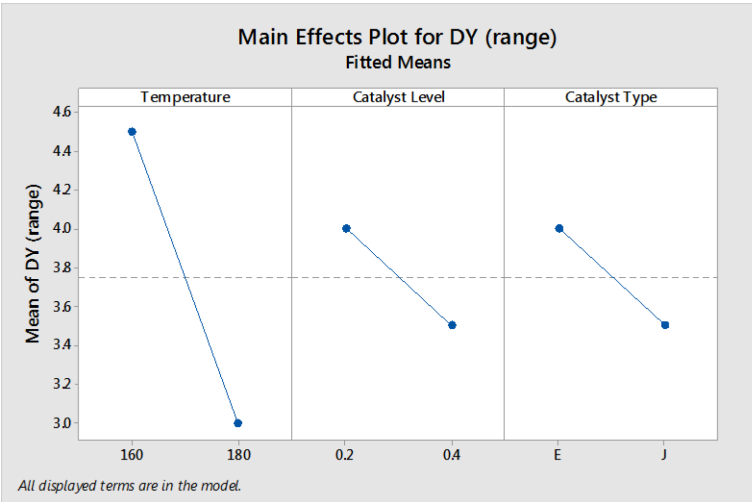
Worksheet 2 ***

↓	C1	C2	C3	C4	C5	C6	C7-T	C8	C9	C
	StdOrder	RunOrder	CenterPt	Blocks	Temperature	Catalyst Level	Catalyst Type	DY (range)		
1	1	1	1	1	160	0.2	E	2		
2	2	2	1	1	180	0.2	E	4		
3	3	3	1	1	160	0.4	E	8		
4	4	4	1	1	180	0.4	E	2		
5	5	5	1	1	160	0.2	J	6		
6	6	6	1	1	180	0.2	J	4		
7	7	7	1	1	160	0.4	J	2		
8	8	8	1	1	180	0.4	J	2		
9										

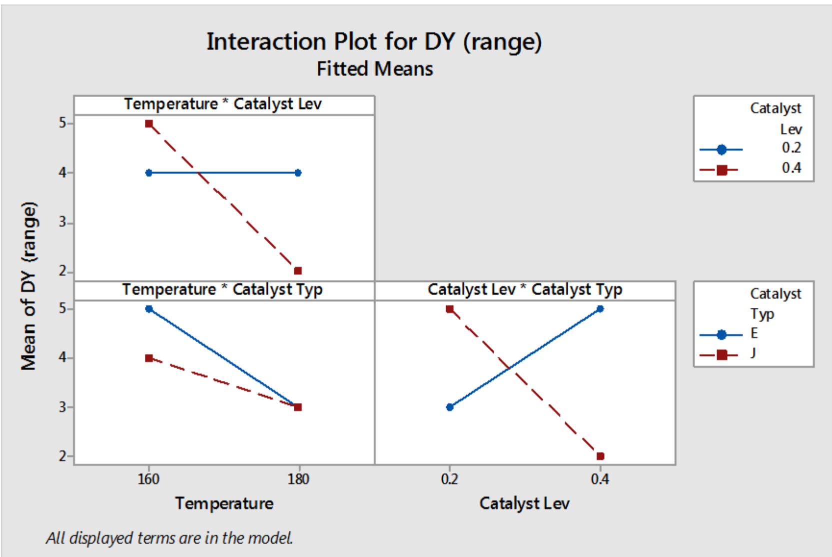




Nothing is significant and hence it is not possible to build a model. We can look at the main effects and interaction plots to see which terms could potentially be the predictors of Δy. However, this is not good science because when we do not have a model (i.e. all terms with p values > 0.05) the results (graphical plots such as the main effects plot for example) could be due to a fluke.



Do not get carried away by the slopes of these lines☺.



Temperature (factor A) is the factor that seems to have the highest effect on ΔY. The other two factors seem to have an equal (but substantially smaller) effect on ΔY. Of course none of these effects may be significant. We can try including temperature only (I've done that for you as shown below), if you want, you can include some other terms also. No matter what terms you include, you will find that none of the terms become significant at 5% level of significance.

Factorial Regression: DY (range) versus Temperature

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	1	4.500	4.500	0.87	0.387
Linear	1	4.500	4.500	0.87	0.387
Temperature	1	4.500	4.500	0.87	0.387
Error	6	31.000	5.167		

Note: Part B of the exercise is touches upon a very important product design concept known as "robust parameter design" (RPD) invented by engineer Genichi Taguchi but being reshaped by other scholars such as George Box. Design of experiments to minimize the variability of Y is beyond the scope of your DOE project; RPD is a PhD level topic. Also in our example, it could be quite possible that the engineers have failed to list the real factors—through brain storming and what not—that affect the variability of Y (i.e. ΔY).

Exercise 3: Full Factorial Designs with Blocks (Self Practice)

In the theory class you learnt that blocking is one of the three good features of an experimental design (the other two being replication and randomisation). This does not mean that all good designs should involve blocking. George Box once said: "block what you can and randomize what you cannot block". This statement epitomises what blocking (and randomisation) does and when it can (and should) be used.

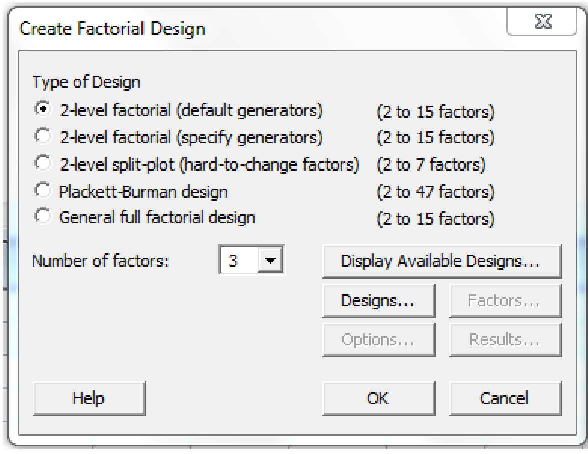
The objective of this self-practice Minitab exercise, which contains two parts, is to help you understand how to design and analyse factorial designs containing two or more blocks, using Minitab17.

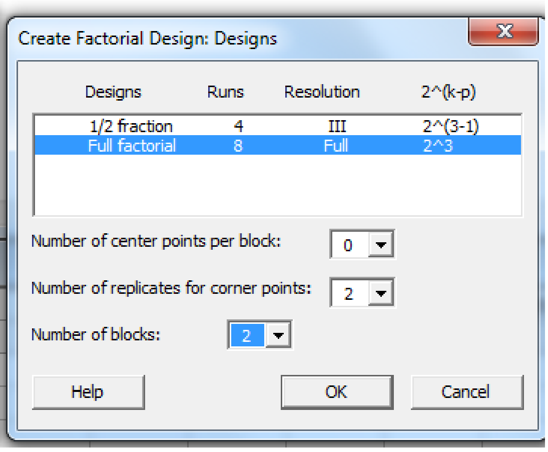
(1) Prompt Minitab to produce design matrices for the following factorial designs. Try to make sense out of each Minitab output (design matrix). Pay particular attention to the randomisation patterns and any confounding.

- 2^3 design with two replicates, the experiment being run in two blocks.
- 2^3 unreplicated design (no of replicates = 1), the experiment being run in two blocks.
- 2^4 unreplicated design (no of replicates = 1), the experiment being run in two blocks.

Justify the designs based on what you learnt on blocking in the theory class.

Prompting Minitab 17 to give me a 2^3 design with 2 replicates but also having 2 blocks.





StdOrder	RunOrder	CenterPt	Blocks	A	B	C
7	1	1	1	-1	1	1
8	2	1	1	1	1	1
2	3	1	1	1	-1	-1
4	4	1	1	1	1	-1
5	5	1	1	-1	-1	1
1	6	1	1	-1	-1	-1
3	7	1	1	-1	1	-1
6	8	1	1	1	-1	1
11	9	1	2	-1	1	-1
12	10	1	2	1	1	-1
14	11	1	2	1	-1	1
13	12	1	2	-1	-1	1
16	13	1	2	1	1	1
10	14	1	2	1	-1	-1
15	15	1	2	-1	1	1
9	16	1	2	-1	-1	-1

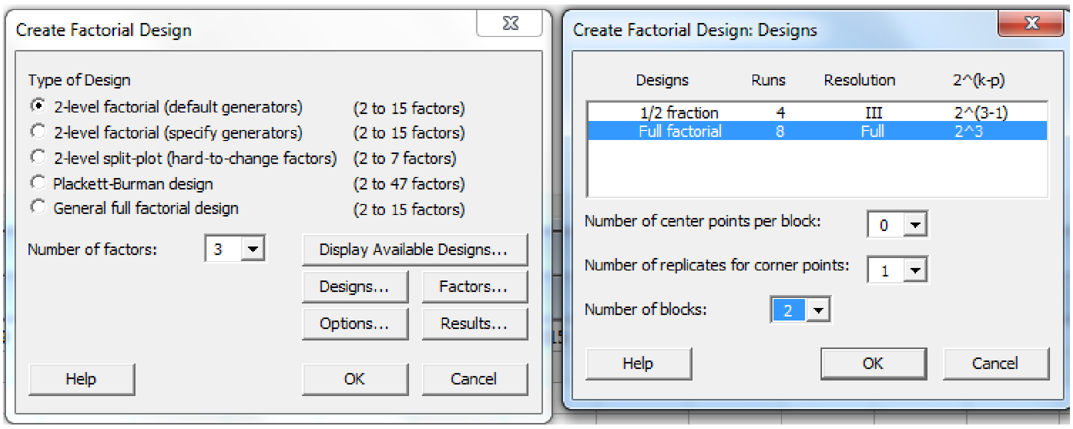
```
Welcome to Minitab, press F1 for help.
```

Full Factorial Design

```
Factors:    3   Base Design:          3, 8
Runs:      16   Replicates:              2
Blocks:     2   Center pts (total):      0


Block Generators:   replicates


All terms are free from aliasing.
```

Note: Minitab has allowed you to run each replicate in each block in order to give you a design in which where all would become free from aliasing (meaning no term would be represented by another term). <u>A good strategy for your DOE Project</u>!

Prompting Minitab 17 to give me an un-replicated 2^3 design to be run in <u>two blocks</u>



StdOrder	RunOrder	CenterPt	Blocks	A	B	C
1	1	1	1	-1	-1	-1
4	2	1	1	-1	1	1
2	3	1	1	1	1	-1
3	4	1	1	1	-1	1
7	5	1	2	-1	-1	1
8	6	1	2	1	1	1
6	7	1	2	-1	1	-1
5	8	1	2	1	-1	-1

```
Welcome to Minitab, press F1 for help.
```

Full Factorial Design

```
Factors:   3   Base Design:         3, 8   Resolution with blocks:  IV
Runs:      8   Replicates:             1
Blocks:    2   Center pts (total):     0


Block Generators: ABC

Alias Structure

Blk = ABC
```

Now what?

Again Minitab has given you a useful design but in this design (since Minitab has no other way to do it) your block column is actually being represented by the column ABC (this is what Blk = ABC means). This means that you lose the opportunity of estimating the effect of ABC independently. Whatever you estimate as the effect of ABC will be the effect of ABC + Blk. For all other terms no aliases exist. However we can argue (using factor sparsity principles) that the effect of ABC→ 0.So this aliasing is not a great tragedy!

Work out the other case (unreplicated 2^4 design in 2 blocks; your WK6 tutorial question belongs to this type.

(2) A full factorial experiment was conducted by a Massey University student to study the effects of three factors on compressive strength of concrete. Two replicates were done. However, each replicate was treated as a separate block because two different laboratories had to be used to test concrete specimens. Table 4 below shows the factors and levels used while Table 5 shows the design matrix along with the responses (Y).

Table 4: Factors and Levels

Factor Label	Factor Name	Setting	
		Block1	Block2
A	Aggregate type	12mm aggregate	18mm aggregate
B	Curing time	7 days	28 days
C	Concrete temperature	16^0C	22^0C

Table 5: Design Matrix and Results

Label	A	B	C	Y = Compressive Strength of Concrete (MPa)	
				Block1 (Replicate 1)	Block2 (Replicate 2)
(1)	-1	-1	-1	5.5	4.1
a	+1	-1	-1	8.3	8.3
b	-1	+1	-1	24.2	20.7
ab	+1	+1	-1	42.8	38.0
c	-1	-1	+1	13.1	10.4
ac	+1	-1	+1	10.4	13.8
bc	-1	+1	+1	27.6	23.5
abc	+1	+1	+1	44.9	46.9

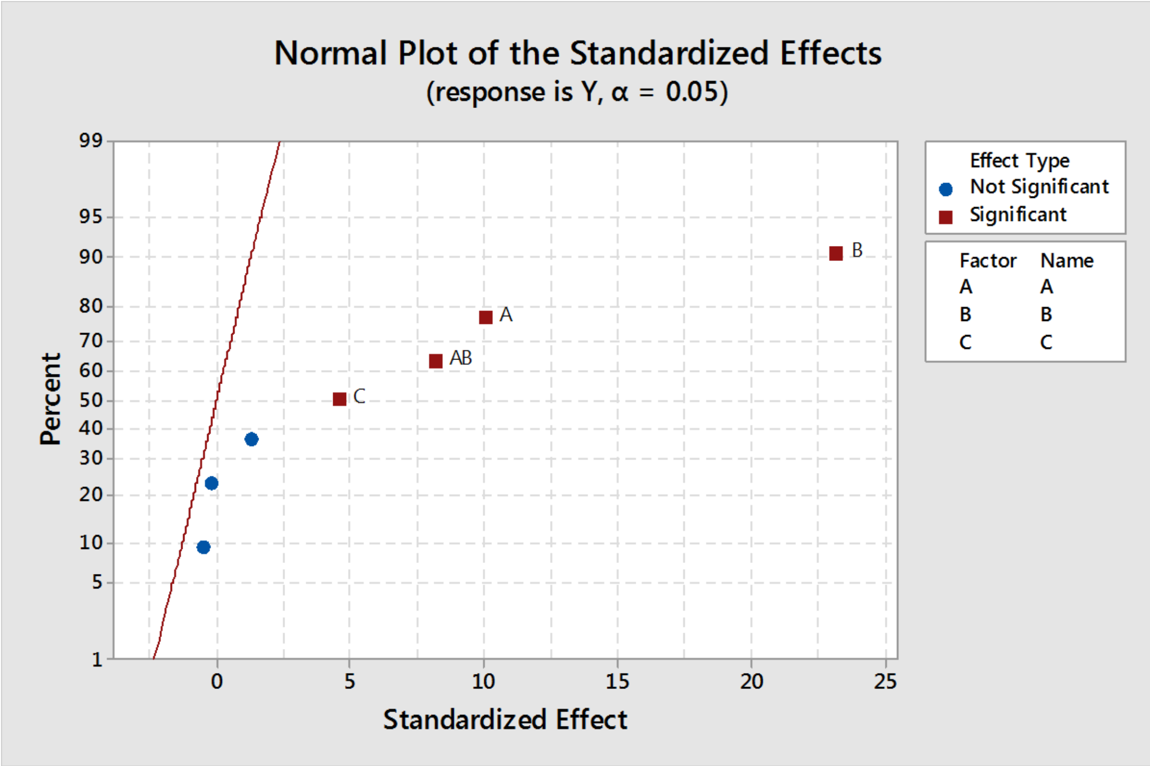
Using Minitab 17, analyse the above results as fully as you can and make actionable conclusions.

Discuss your results with the instructor when you can.

The Minitab file is on Stream (under Labs folder): Exercise3.MPJ

StdOrder	RunOrder	CenterPt	Blocks	A	B	C	Y
1	1	1	1	-1	-1	-1	5.5
2	2	1	1	1	-1	-1	8.3
3	3	1	1	-1	1	-1	24.2
4	4	1	1	1	1	-1	42.8
5	5	1	1	-1	-1	1	13.1
6	6	1	1	1	-1	1	10.4
7	7	1	1	-1	1	1	27.6
8	8	1	1	1	1	1	44.9
9	9	1	2	-1	-1	-1	4.1
10	10	1	2	1	-1	-1	8.3
11	11	1	2	-1	1	-1	20.7
12	12	1	2	1	1	-1	38
13	13	1	2	-1	-1	1	10.4
14	14	1	2	1	-1	1	13.8
15	15	1	2	-1	1	1	23.5
16	16	1	2	1	1	1	46.9

I asked Minitab not to randomise the runs (using the Options tab) so that I could copy paste the Y values from Table 3 (Cheeky!).



A, B, AB and C are significant terms (ANOVA shows the same thing and so does the Pareto chart).

Let us have quick look at: (a) The ANOVA results and (b) the effects estimates (or the regression coefficient estimates) and their significance.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	8	3220.42	402.55	91.62	0.000
Blocks	1	7.70	7.70	1.75	0.227
Linear	3	2907.02	969.01	220.56	0.000
A	**1**	**444.16**	**444.16**	**101.09**	**0.000**
B	**1**	**2369.26**	**2369.26**	**539.27**	**0.000**
C	**1**	**93.61**	**93.61**	**21.31**	**0.002**
2-Way Interactions	3	298.00	99.33	22.61	0.001
A*B	**1**	**296.70**	**296.70**	**67.53**	**0.000**
A*C	1	0.14	0.14	0.03	0.863
B*C	1	1.16	1.16	0.26	0.624
3-Way Interactions	1	7.70	7.70	1.75	0.227
A*B*C	1	7.70	7.70	1.75	0.227
Error	7	30.75	4.39		
Total	15	3251.17			

Model Summary

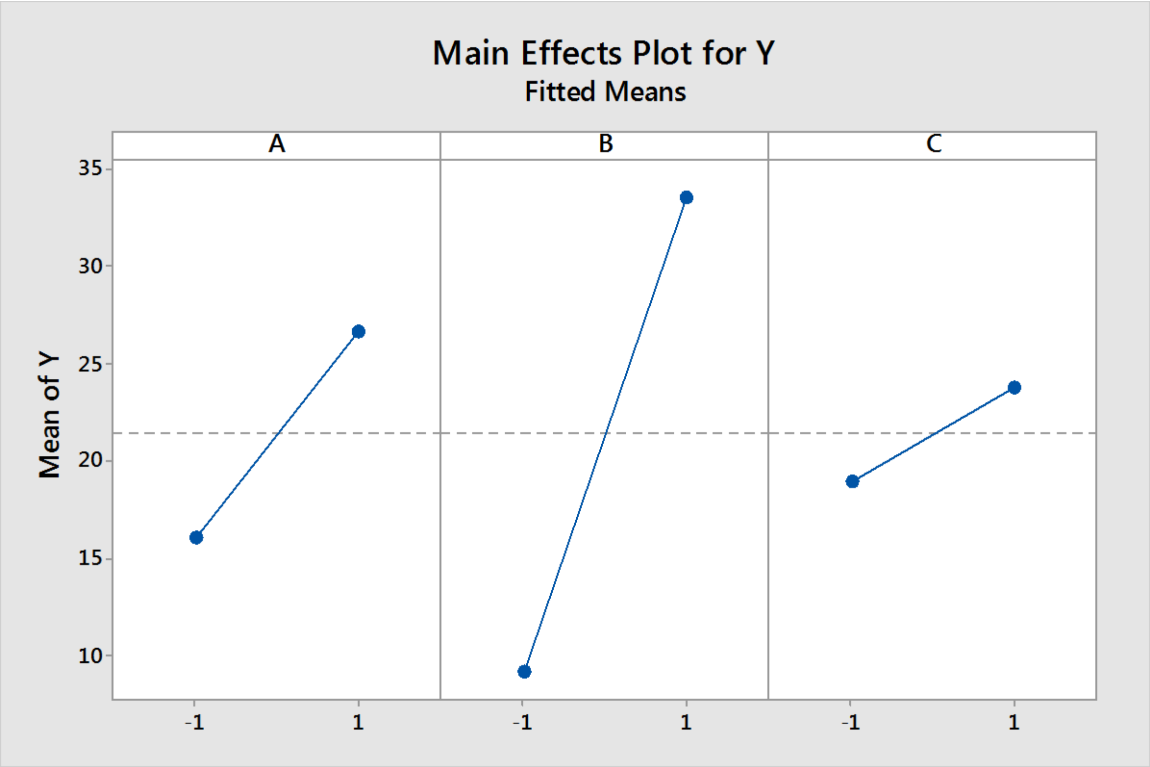
S	R-sq	R-sq(adj)	R-sq(pred)
2.09606	99.05%	97.97%	95.06%

A, B, C and AB are significant. The block effect (thankfully it was not confounded with any other effect) is not significant. He can interpret the latter to mean that there is not much difference between the two labs (generally the case in dry engineering disciplines) and why should there be a difference in the first place? May be the engineer would have thought that one lab uses older equipment than the other, which is a reasonable basis to have 2 blocks.

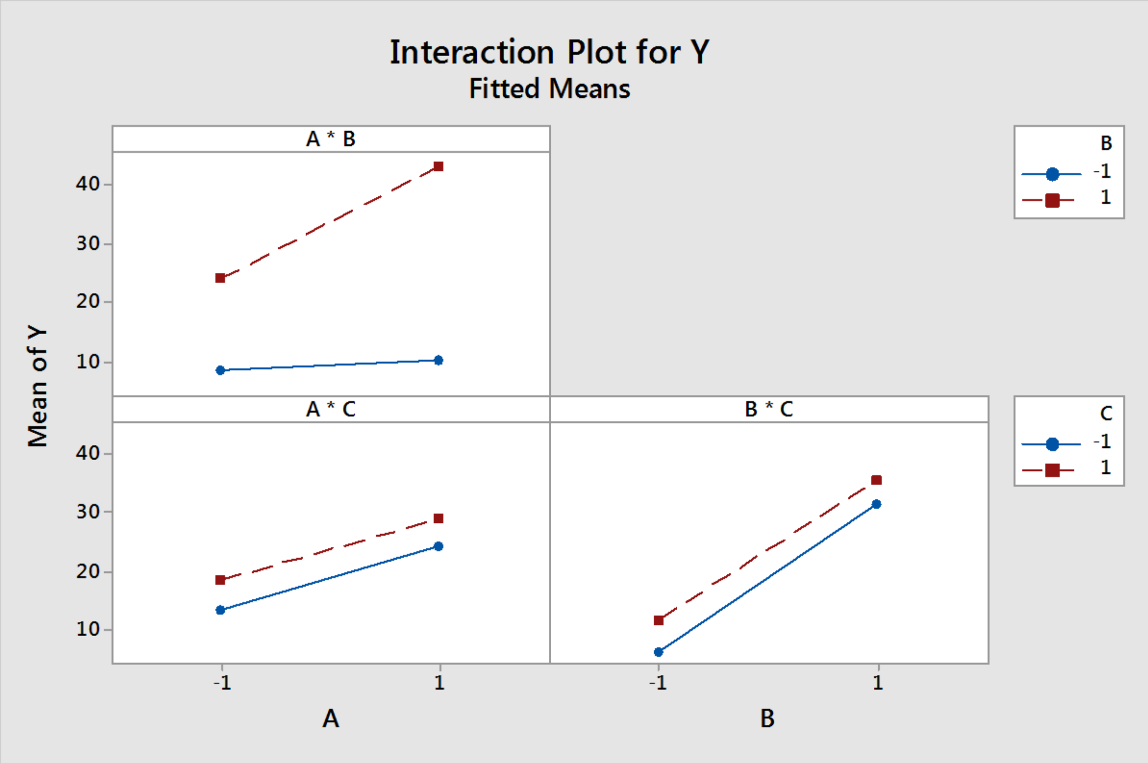
Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		21.406	0.524	40.85	0.000	
Blocks						
1		0.694	0.524	1.32	0.227	1.00
A	**10.538**	**5.269**	0.524	**10.05**	**0.000√**	**1.00**
B	**24.338**	**12.169**	0.524	**23.22**	**0.000√**	**1.00**
C	**4.838**	**2.419**	0.524	**4.62**	**0.002√**	**1.00**
A*B	**8.612**	**4.306**	0.524	**8.22**	**0.000√**	**1.00**
A*C	-0.188	-0.094	0.524	-0.18	0.863	1.00
B*C	-0.538	-0.269	0.524	-0.51	0.624	1.00
A*B*C	1.388	0.694	0.524	1.32	0.227	1.00

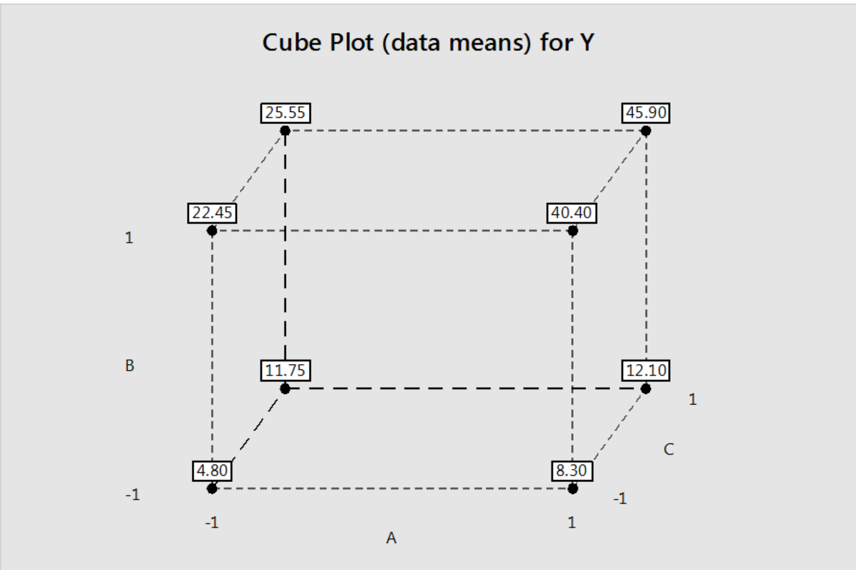
Let us look at the factorial plots and the cube plot to interpret the results (effects) from a practical perspective.



B is the most important factor in manipulating the response. A is the next influential factor and C is the least influential factor. Keep all factors at high level to maximise Y.



AB interaction seems to be a practically important one (the lines are not parallel). When B is at -1 (7 day curing time), it does not matter what aggregate type you choose (the 12mm one or the 18mm one) you get roughly the same Y (strength). However, when you B (curing time) is at +1 (28 day curing time) changing your aggregate type from 12mm to 18mm makes a big difference: you will increase the strength of the concrete considerably. Note: It pays you to name your factors and define their levels rather than leaving them with default options: A,B,C and -1 and +1 (when you are setting up your design matrix, under factors you can do these changes).



We know what to write because the cube plot shows basically the same things the factorial plots show! However, under the cube plot, you need to comment only on the most important 2 or 3 things—I'd say you could comment on the role factor B plays and about the AB interaction.

I have asked Minitab to get me the corner points based on data means because I know that the fitted values (based on the model) is not trustworthy because I have allowed Minitab to include all the terms (at this stage) in the model.

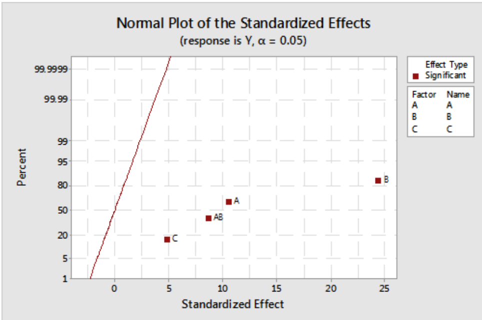
Finalising the model:

Remove all unimportant terms from the terms box and re-analyse. Your regression model should now be good for use.

Factorial Regression: Y versus Blocks, A, B, C

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	5	3211.42	642.28	161.58	0.000
Blocks	1	7.70	7.70	1.94	0.194
Linear	3	2907.02	969.01	243.77	0.000
A	1	444.16	444.16	111.73	0.000
B	1	2369.26	2369.26	596.02	0.000
C	1	93.61	93.61	23.55	0.001
2-Way Interactions	1	296.70	296.70	74.64	0.000
A*B	1	296.70	296.70	74.64	0.000
Error	10	39.75	3.98		
Total	15	3251.17			



Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.99377	98.78%	98.17%	96.87%

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		21.406	0.498	42.95	0.000	
Blocks						
1		0.694	0.498	1.39	0.194	1.00
A	10.538	5.269	0.498	10.57	0.000	1.00
B	24.338	12.169	0.498	24.41	0.000	1.00
C	4.838	2.419	0.498	4.85	0.001	1.00
A*B	8.612	4.306	0.498	8.64	0.000	1.00

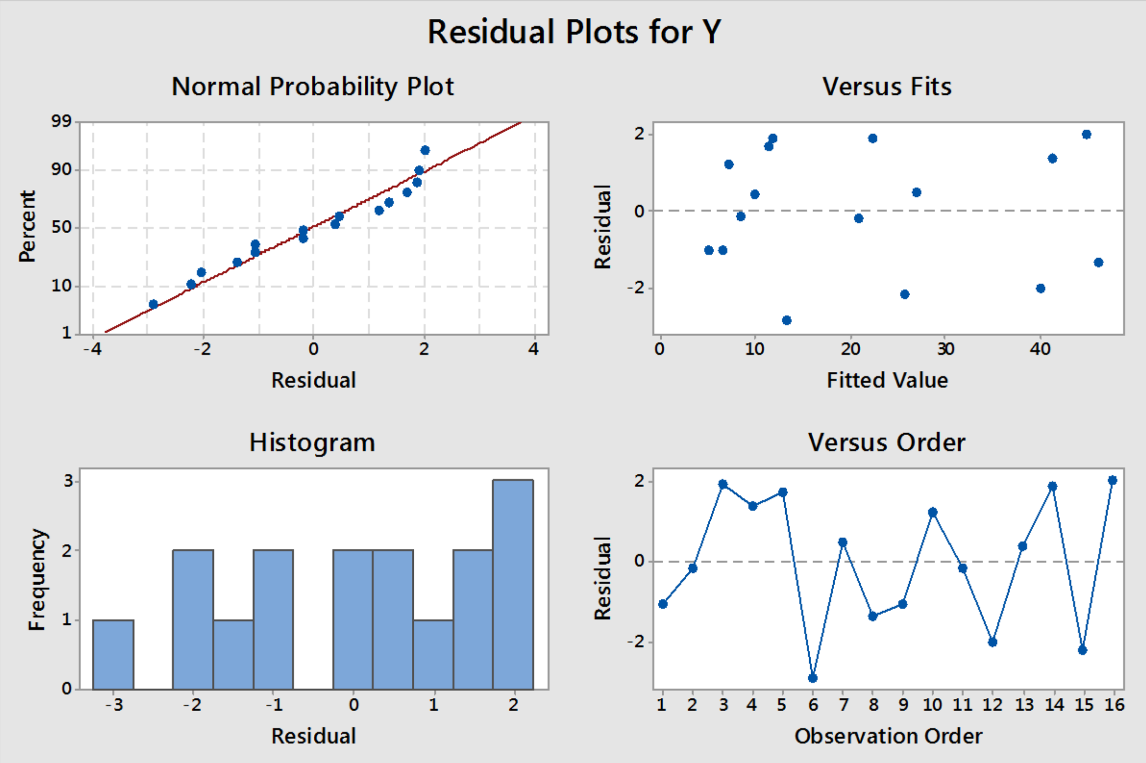
Regression Equation in Uncoded Units

Y = 21.406 + 5.269 A + 12.169 B + 2.419 C + 4.306 A*B

The model has high R^2. As much as 98.78% of the variability of Y is explained by the terms in the model. The model also suggests that all factors must be at =1 setting to get the highest Y.

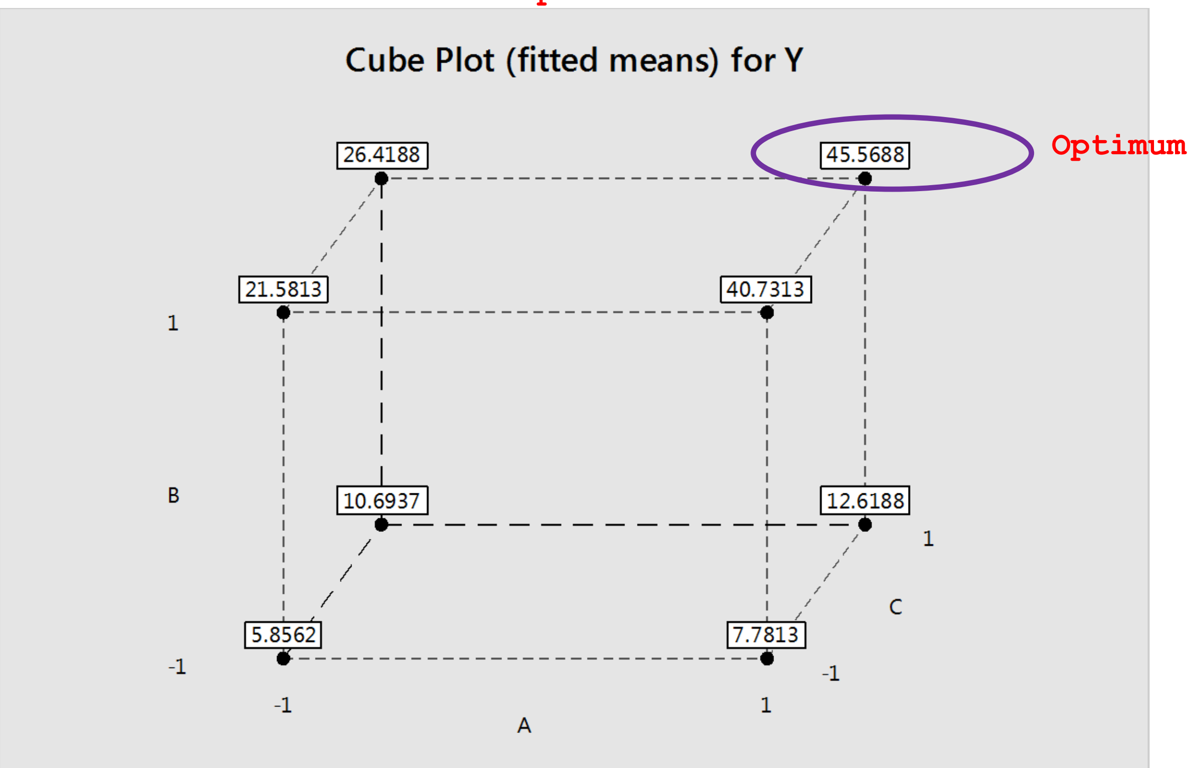
Finally, check the adequacy of the model and ensure that your model meets the standards of an adequate model (normality, equal variance, and independence). So ask for the 4 in 1 plot (make sure that you only have A, B, C and AB in the terms box).

After you are satisfied with the residual plots show the results in a cube plot with <u>fitted (estimated) values</u>; your model will do that for you.



Residuals do not give us the confidence that the data have come from a normal population. May be we have to consider a data transformation (no need to these stuff for your DOE project). The other plots are OK but not useful if we have to consider converting your Y scale to say ln(Y). Also the residuals Vs observation order plot is of no use because the runs are not in the actual run order (i.e. the random order).This would not be the case for your DOE project!

Notwithstanding the concerns we have on normality, let us show the final results on a cube plot once and for all.



Conclusions:

All factors were found to be important but factor B was found to be the most important factor. An interaction between factors A and B was also found. It was found that when B is at -1 (7 day curing time), it did not matter what aggregate type is used—12mm one or the 18mm one—one is expected to get approximately the same Y (concrete strength). However, when factor B (curing time) is at +1 (28 day curing time) changing the aggregate type from 12mm to 18mm makes a big difference: the strength of the concrete did increase considerably.

The following model that explains the relationship between the factors and the concrete strength Y

$$Y = 21.406 + 5.269\,A + 12.169\,B + 2.419\,C + 4.306\,A*B$$

Note that the model is true for coded units (-1, +1 stuff)only.

The maximum expected yield could be obtained when all factors are set at their high settings (for your DOE project try to use actual settings rather than the high/low thing). The maximum expected compressive strength of concrete is 45.57 (<<Give the Figure No of the cube plot in the previous page>>).

The residual plot shows non-normality of data, implying that the results are perhaps, not generalisable. Residuals being normally distributed or otherwise (more so for the latter) it is highly recommended that some confirmation runs be carried out at +1 settings of all 3 factors (the optimum settings) to verify that the results do stay close to the predicted values.

Note: For your DOE project, do not be afraid to conduct conformation runs, whatever the outcome might be. What you need to remember is that great discoveries in science, technology, and engineering (although this is unlikely to be the case in regards to your DOE project☺) are made when the experimenter finds something that they did not expect. You being able to find that your model is not predicting future results very well is a significant achievement! This is because that will lead you to discover more things about the phenomenon that you studied sometime later, may be after you graduate, just kidding ☺.