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

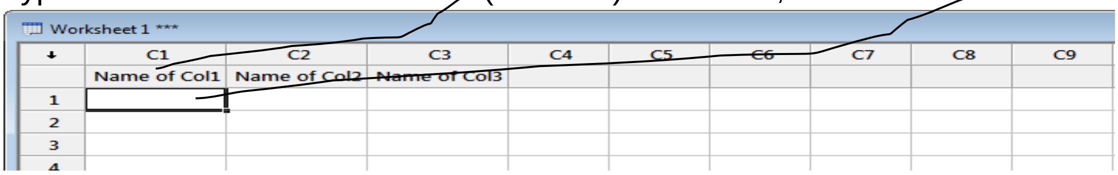
Table of Contents

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!

---

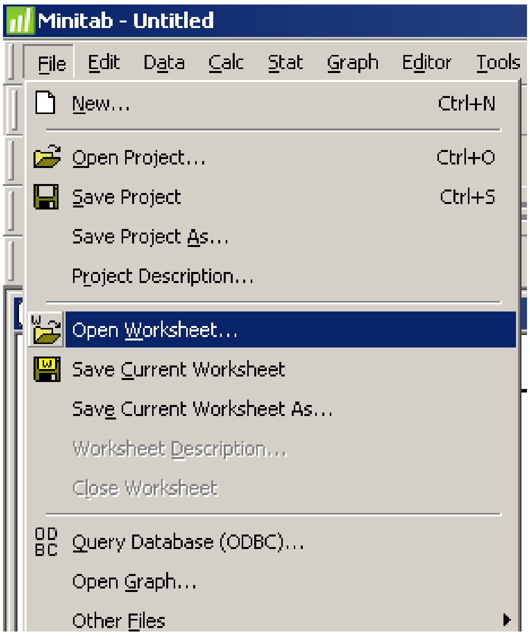
[1] In the convention used in Minitab (as well as many other software packages), an unreplicated design is treated as a design having number of replicates =1; if the 16 runs in Table 6.1 are replicated once more to end up with 32 runs altogether, the number of replicates have to be set to be = 2 in Minitab).

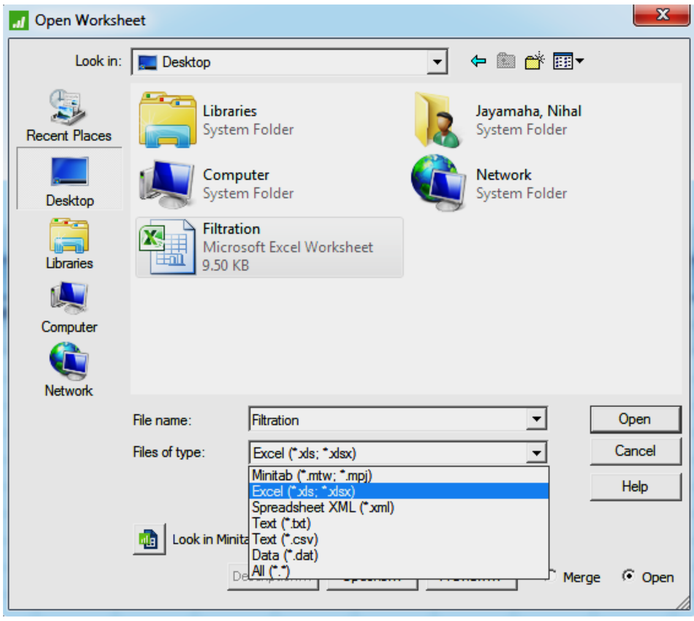
[2] Data taken from: Montgomery, D. C. (2013, p, 257). *Design and analysis of experiments* (8th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

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.

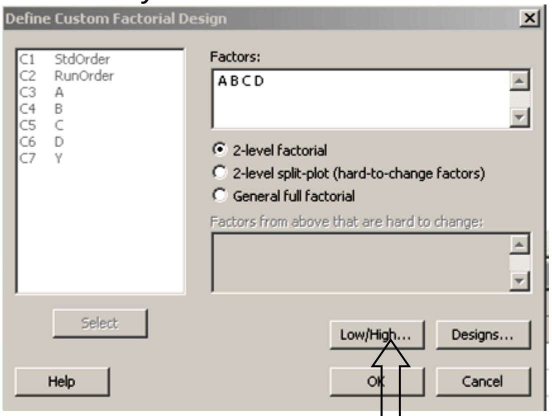
Would you like to provide this information so that Minitab can analyse the design?

Yes NO

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

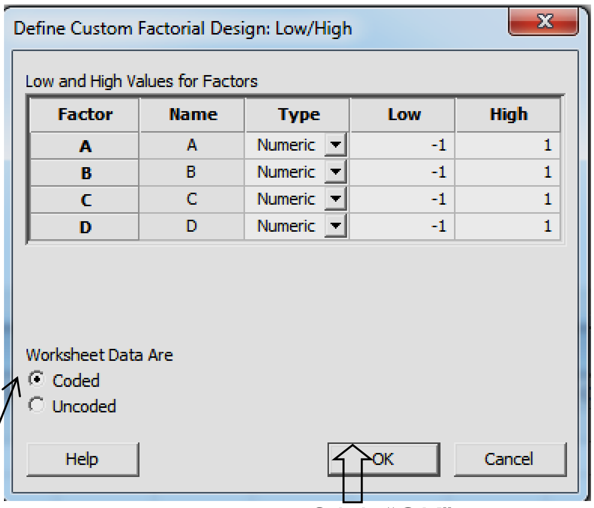
Panel 1

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).

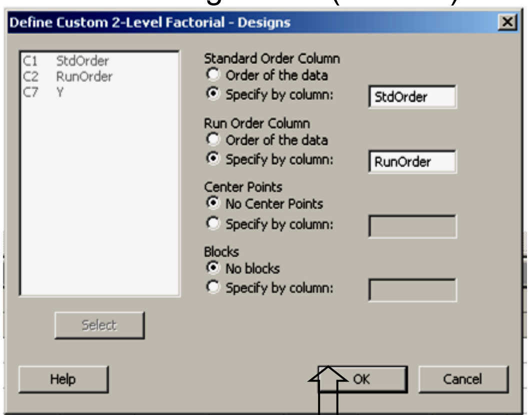
(c) If you know the actual low and high values of each factor you can replace -1 and +1 with these actual values.

(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

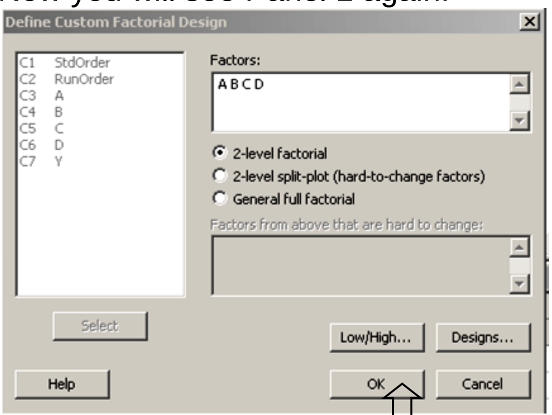
Click the "Designs" tab (Panel 1)



Specify your StdOrder and RunOrder columns and click "OK"

Panel 4

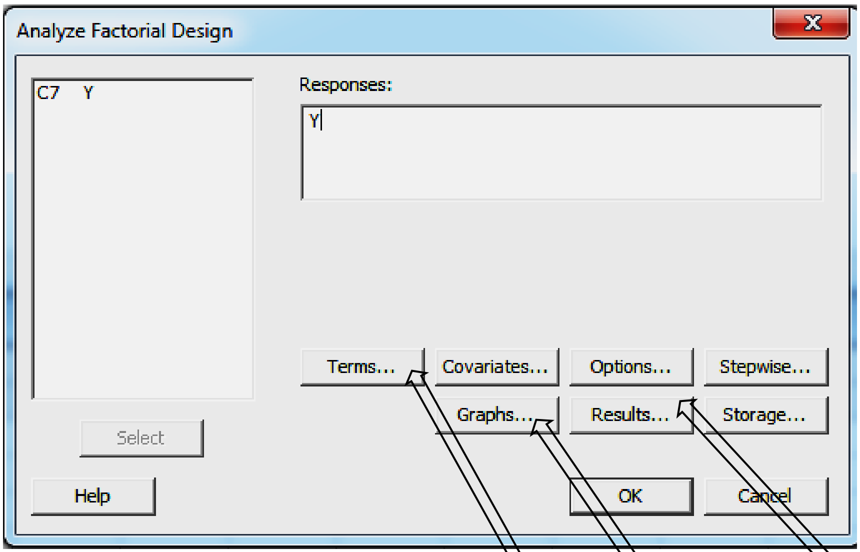
Now you will see Panel 1 again:



Click "OK"

Panel 5

Introduce your Response column (Y)

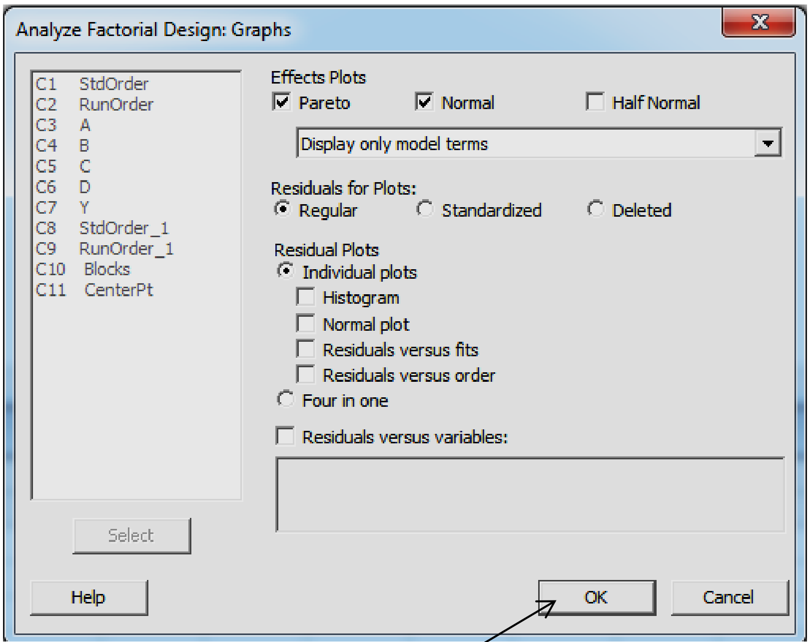


The following tabs are important to you: "Terms" , "Graphs" and "Results" (the other tabs such as the "Covariates", "Stepwise" model building would not apply in your type of factorial designs).

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).

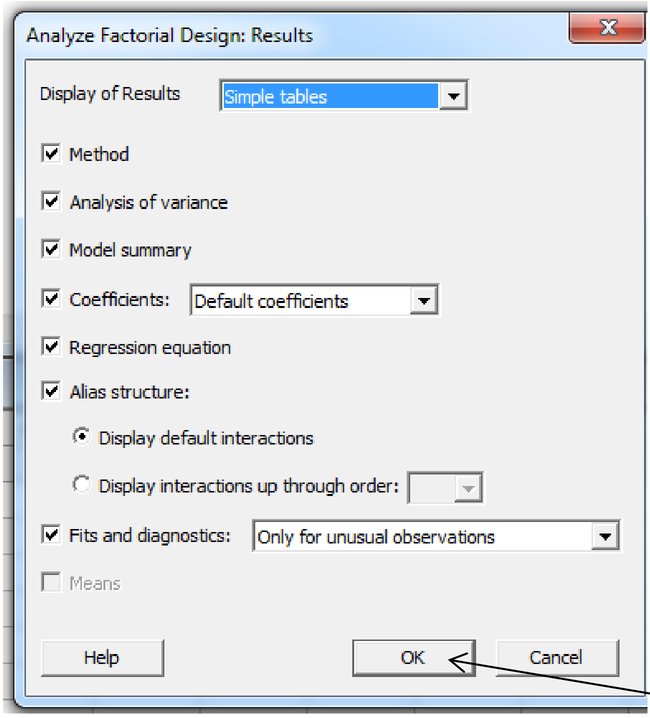
Panel 6



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

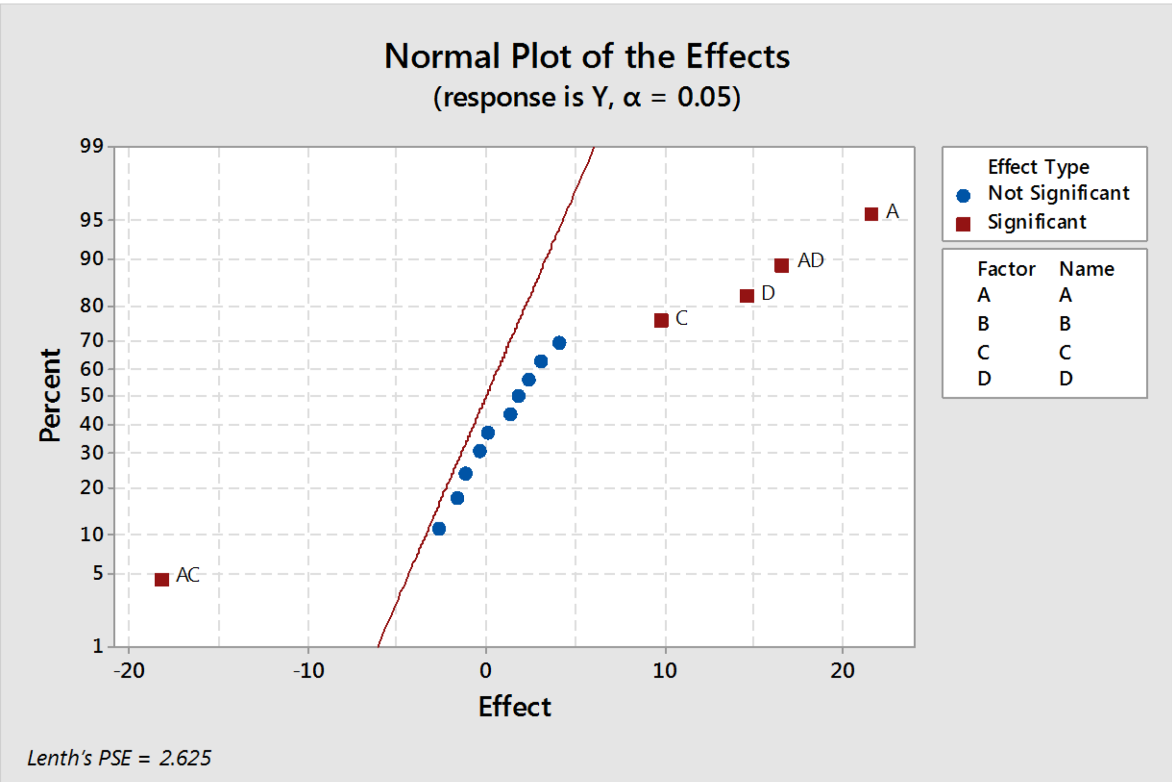


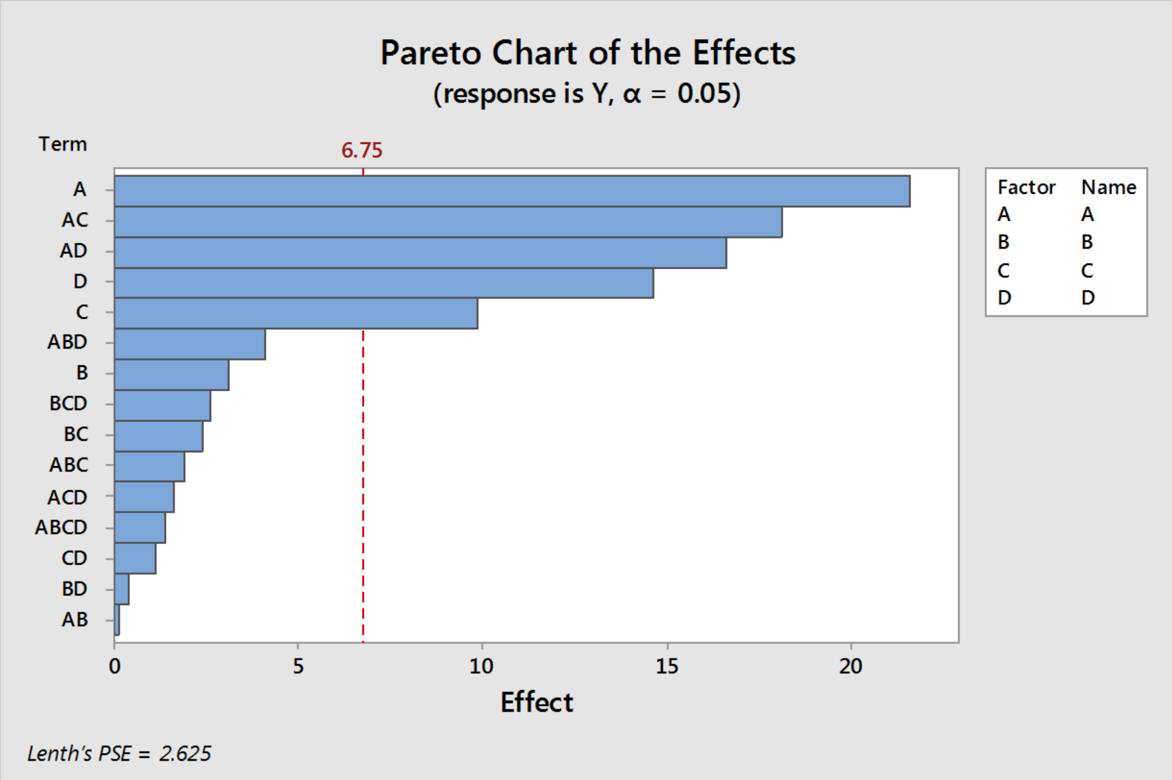
Most of the things Minitab has selected by default to display as results (in the session window) are not relevant or useful at this stage.

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:

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.

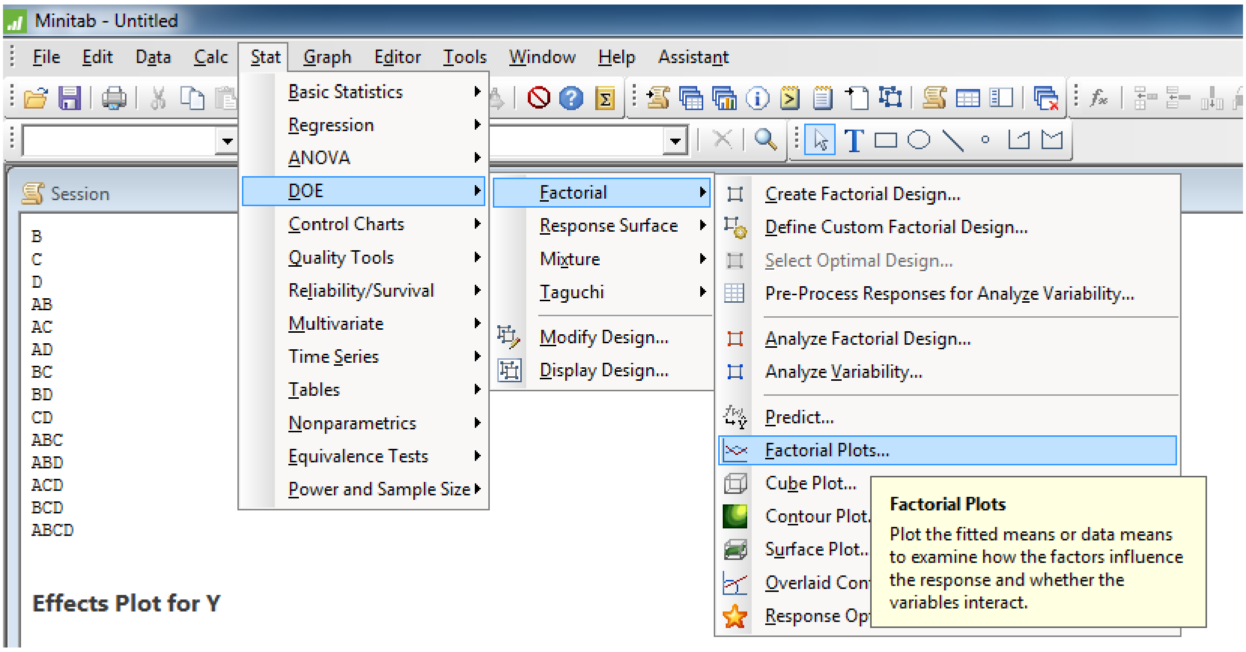
Deriving the Factorial Plots

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.

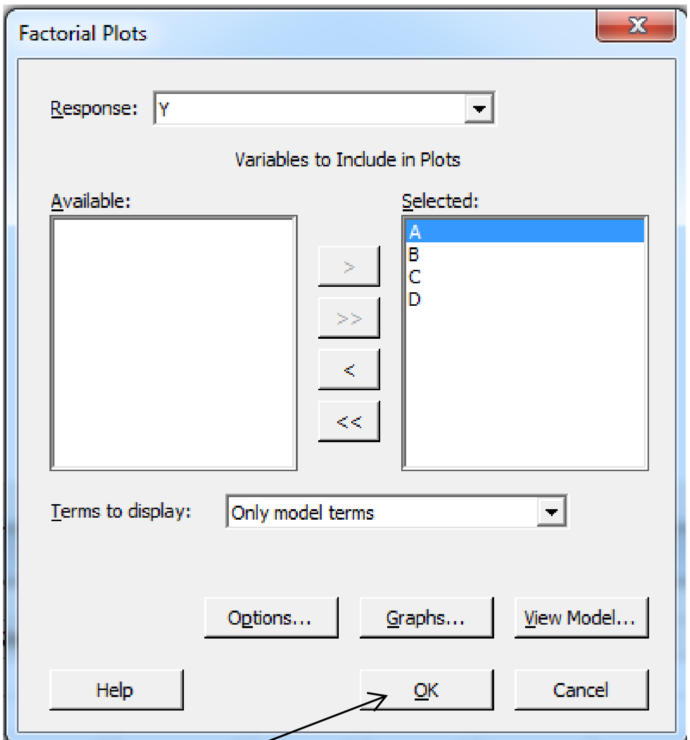
Click Stat → DOE → Factorial → Factorial Plots (see Panel 8)

Panel 8



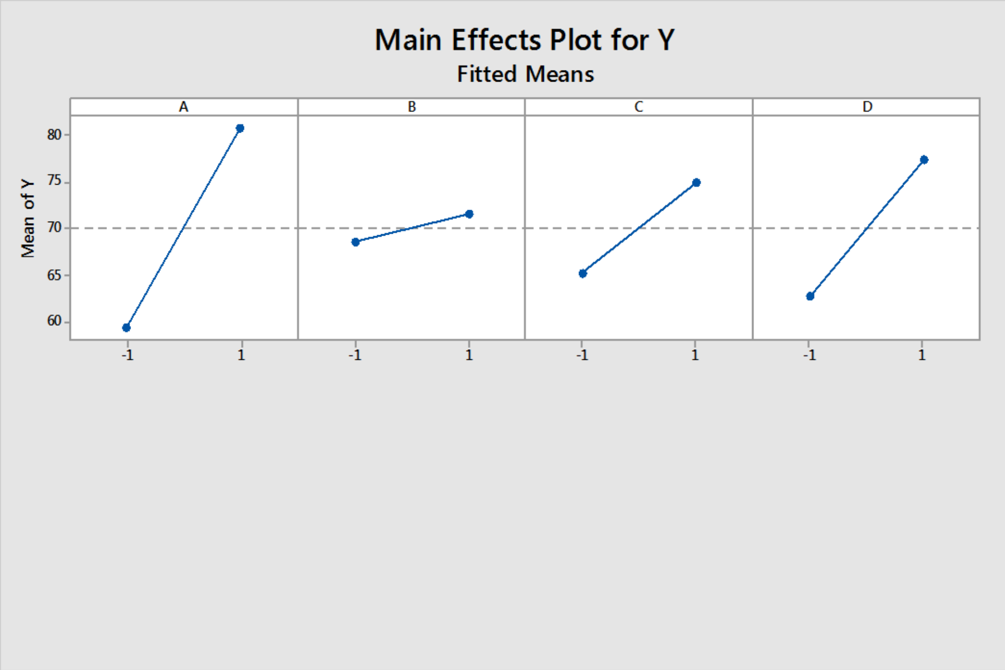
Select the factors that are of interest to you. Select all 4 factors (Panel 9) although you know that factor B does not see to play a role in predicting your response.

Panel 9

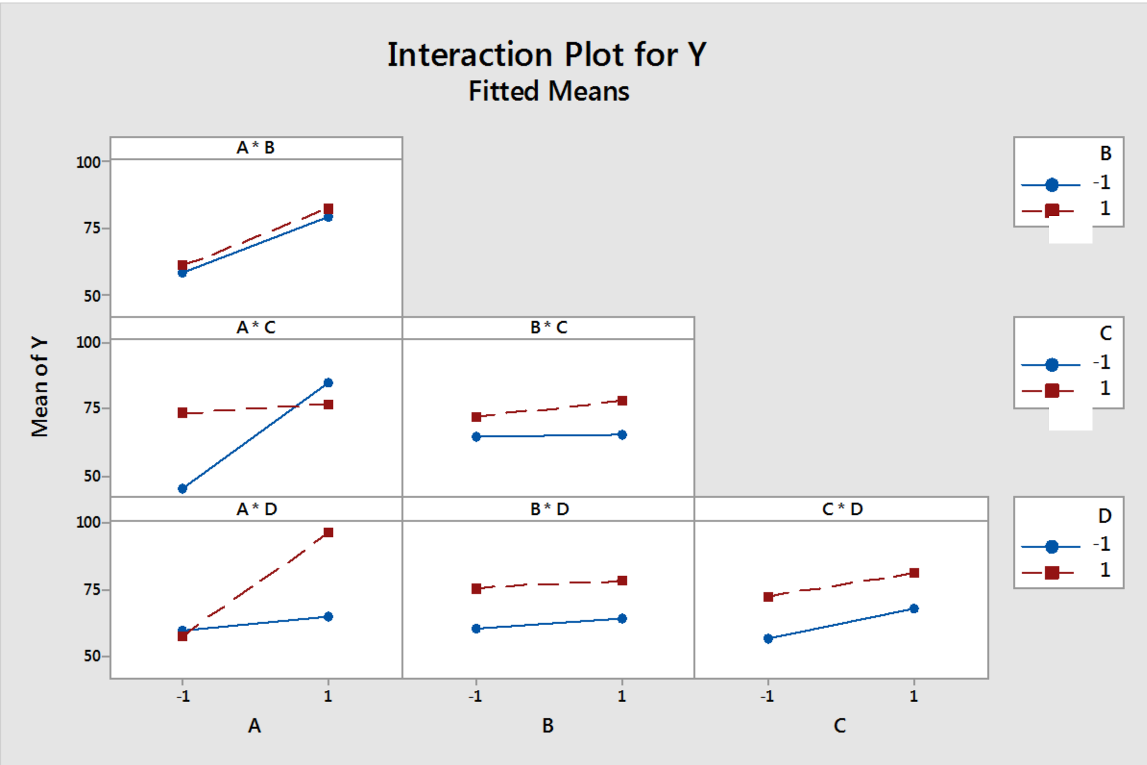


Click "OK".

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

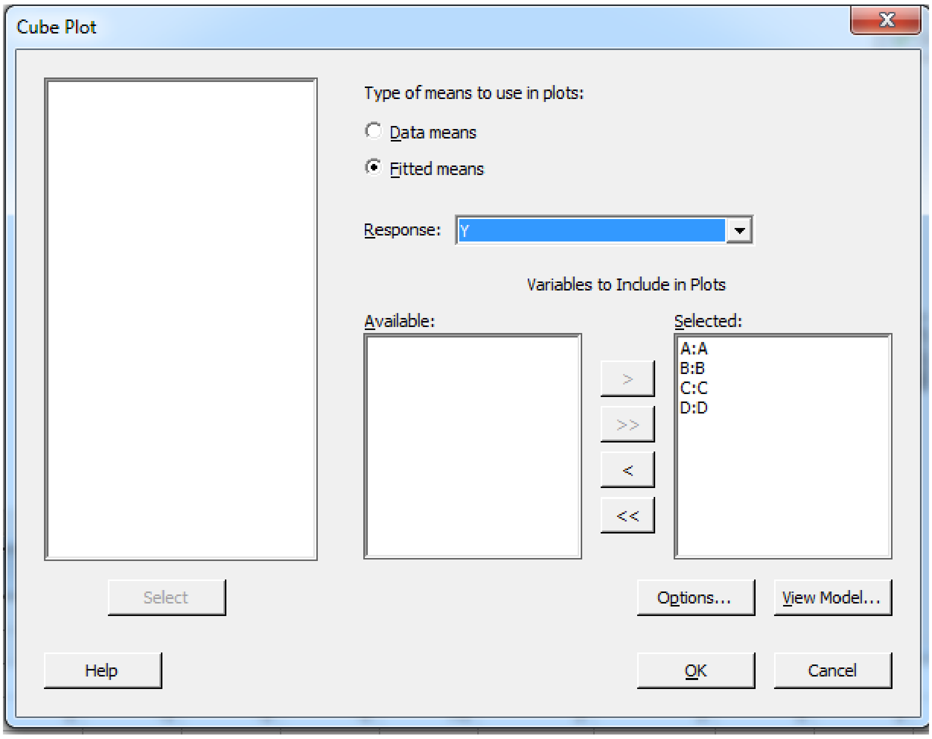


Write-down what you can reasonably deduce from the main effects plots.



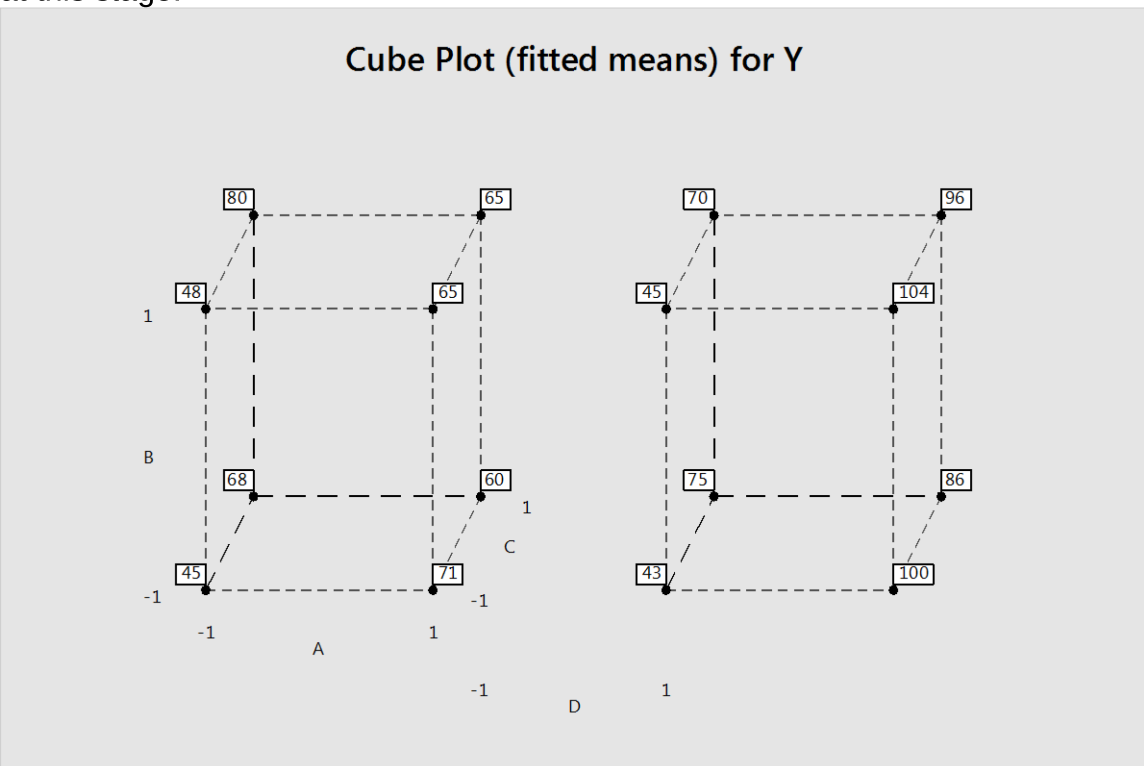
Write-down what you can reasonably deduce from the interaction plots.

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 not ☺), while data 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.

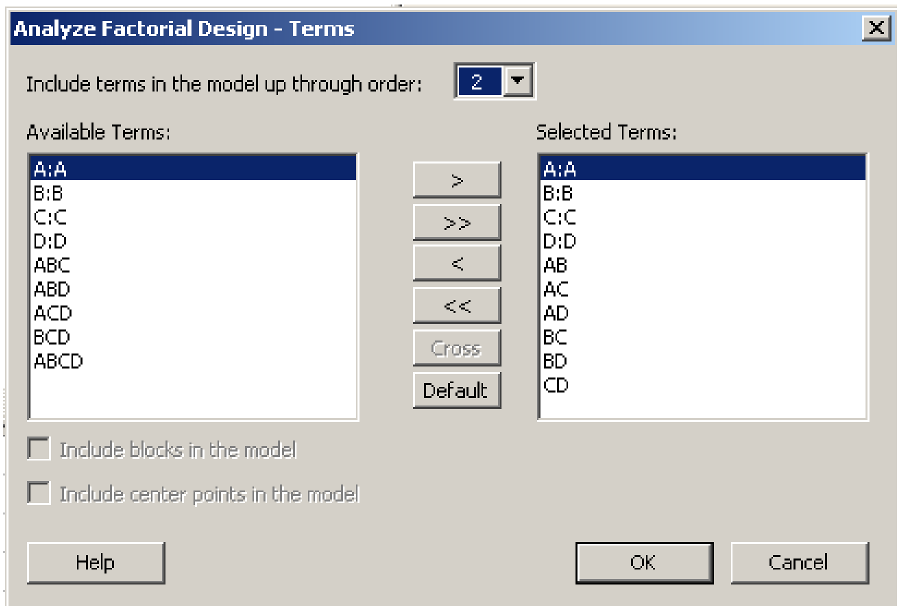


Write-down what the cube plot seems to be suggesting on the main effects and interactions; *a cube plot is difficult to interpret till you get used to it*; *do not worry.*

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 + ε

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 <u>*maximise*</u> *the filtration rate Y? B is statistically and practically insignificant.*

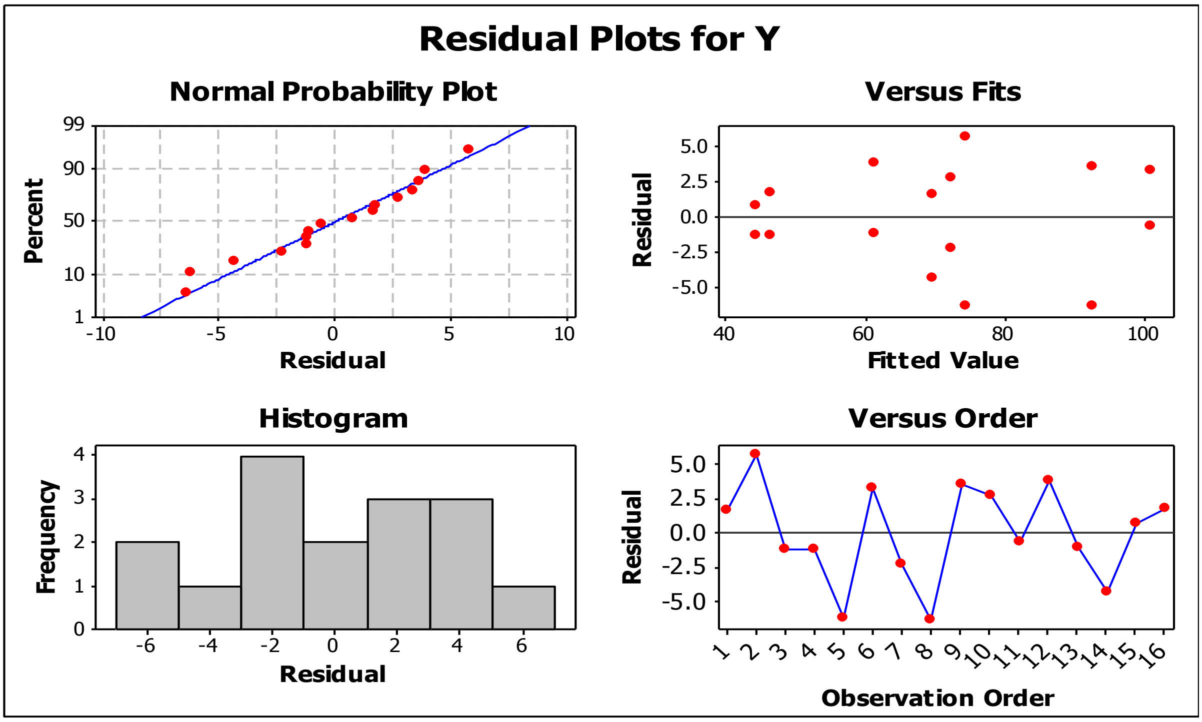
Are the above settings consistent with what is shown in the cube plot? It is better to get the cube plot with factors A, C and D only (with fitted means).

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: ………………….

Minitab’s fitted mean estimate (shown in the cube plot) for the above optimum setting is: ……………………………….. (both should be the same).

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…   The histogram shows that….   The versus fit graph shows that….   The versus order graph shows that….

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.

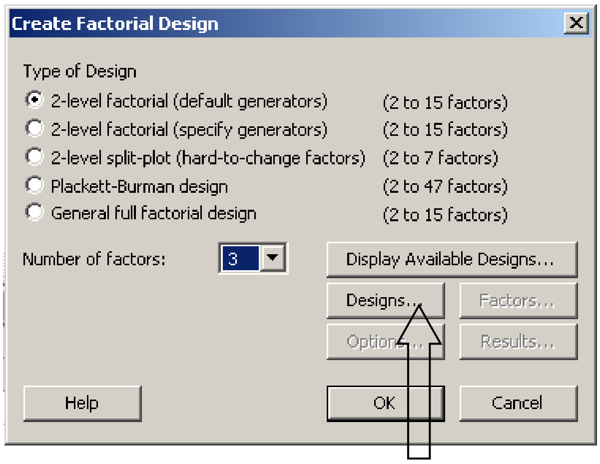
Exercise 2_Part 1: Creating a Replicated Full Factorial Design

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

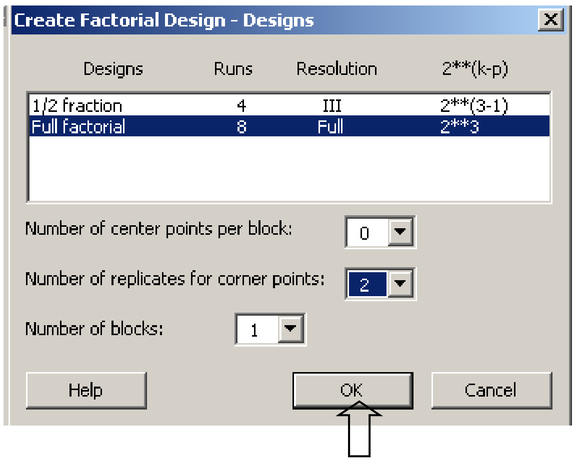
Some useful screen shots are given below

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:

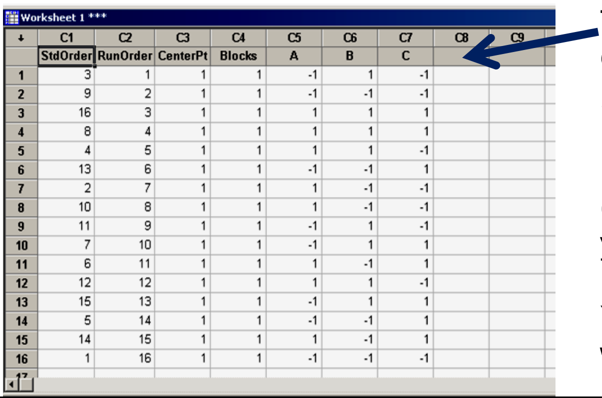
```
Welcome to Minitab, press F1 for help.
```

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.
```

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°C	180°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 Part 1 of Exercise 2 (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.

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 Minitab16.

(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.

(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	$16^{0}C$	$22^{0}C$

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.

WEEK 6 Tutorial

(a) A food technologist is planning to conduct a two-level factorial experiment to study the effect of four process variables—Temperature (Factor A), Pressure (Factor B), Speed (Factor C), and Time (Factor D)—on three response variables related to a processed food: Bulk Density (Y_1), Stickiness (Y_2), and Porosity (Y_3). Apart from understanding how the process behaves, in terms of the relationships between the experimental factors and the three response variables, the researcher is interested in determining which factor combination achieves the target values of Y_1, Y_2 and Y_3 as close as possible.

The technologist wants to conduct the experiment in the two food processing plants (P1 and P2). They have resources to conduct 16 experimental runs altogether. They know from prior experience that factors A and C as well as B and C could interact and that other than the AC and BC interactions no other significant interaction would likely exist.

In groups, perform the following tasks and formulate brief responses to the eight discussion points given below.

1. Design an experimental plan (design matrix) that could investigate the four factors A, B, C, D independently using only 16 trials.

Note 1: *Do not include in your design matrix the (11) columns containing the interactions between factors.*

Note 2: *Do not include the processing plant variable (P1/P2) in your design (yet!).*

2. Since there are thought to be no 3 factor or higher order interactions, take any 3 or 4 factors from your matrix, and generate an additional column of - / + for your design that would be used to measures the interaction between your chosen factors if it did exist.

3. Discuss how this new column in the design matrix could instead be used to measure the effect of processing plants (P1 against P2).

4. Discuss possible reasons for conducting the experiment in both processing plants (P1 and P2).

5. Why the processing plant variable should not be treated as a fifth factor (E)?

6. Now discuss how you will randomise your design and what (if any) restrictions apply to your ability to completely randomise the design trials.

7. What problems might the addition of the processing plants create in your design, analysis or conclusions?

8. If you were to present your findings on the optimum factor setting that simultaneously achieves the target values of Y_1, Y_2 and Y_3 as close as possible, which graphical plot would you most likely choose to present your findings to your manager? Justify your choice in one or two sentences.

****