McMaster University

Chemical Engineering

ChE 4C03

Spring 2008

Design of Experiments

John MacGregor

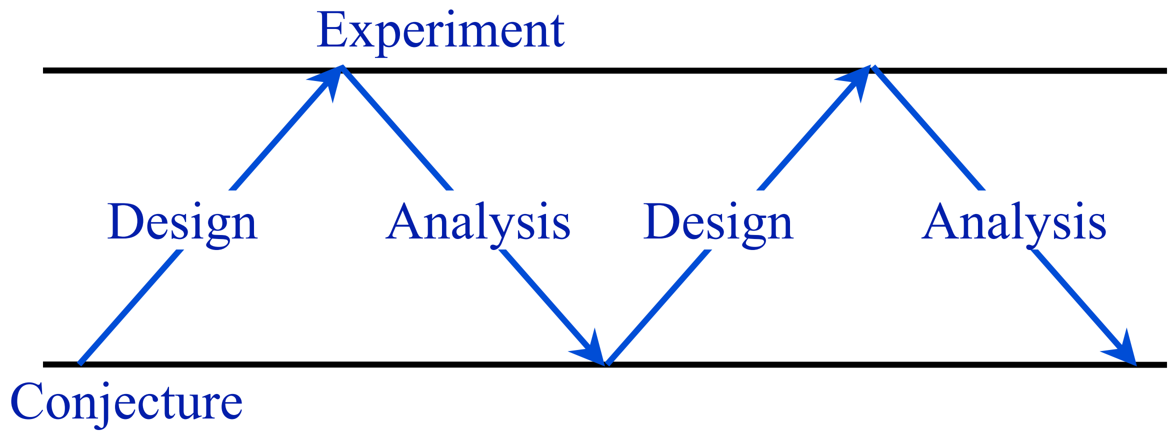
Two aspects of statistics

1. **Analysis of Data**
   - How to extract the most information out of given set of data
2. **Design of Experiments (DOE)**
   - How to ensure that data contains information

DOE is the most important:

- If there is little information in the data, then no amount of analysis will help
- If there is a lot of information in the data, then even simple analysis will reveal it

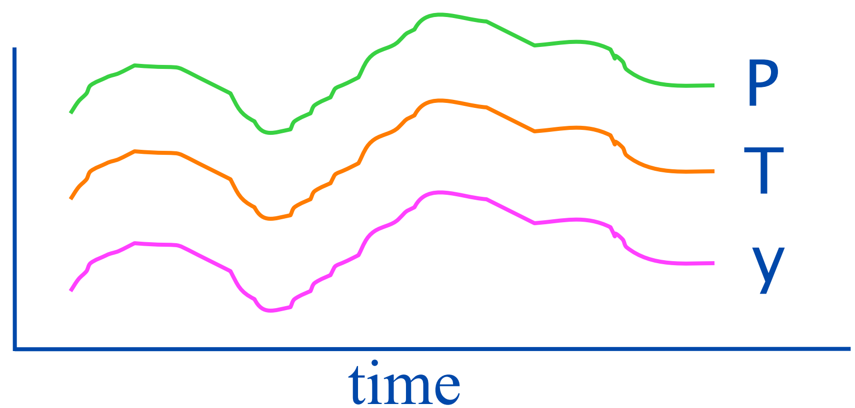
Design of Experiments



Iterative Nature of Experimentation

Why Design?

1. Reduces amount of experimentation needed.
2. Ensures adequate range of variation in all x’s
3. Minimizes confounding of effects



Why Design?

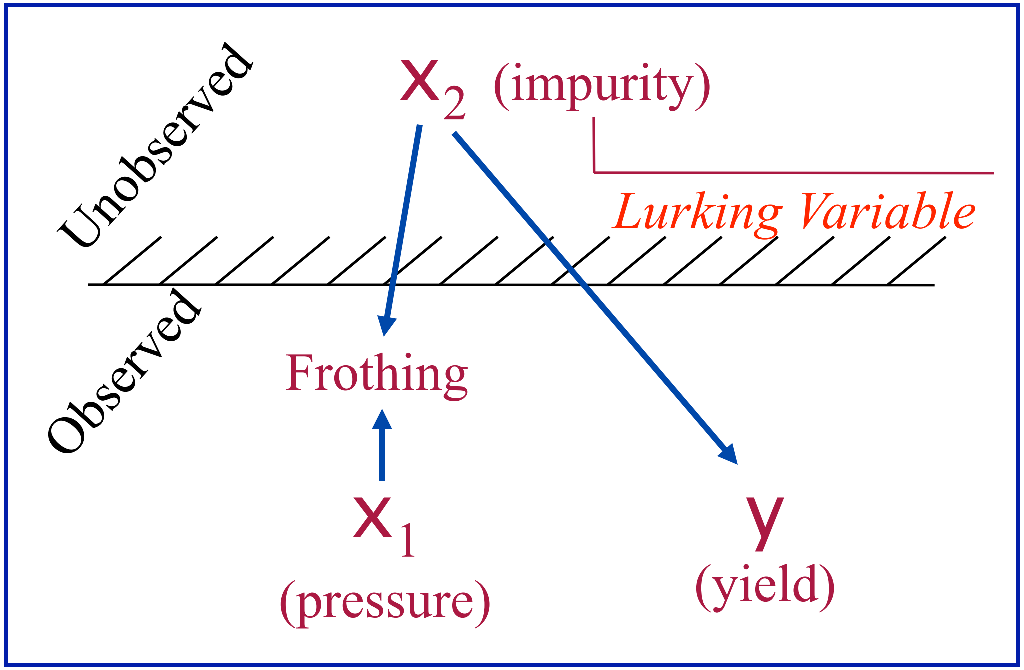
4. Ensures that one finds causal relationships rather than just correlations. Examples.

Example: Chemical Process (BH^2, pg. 487)

Observed that undesirable frothing in reactor could be reduced by increasing Pressure (x_1)

Operating Procedure:

Increase x_1 when frothing



Truth (unknown):

i. High impurity (x_2) causes frothing

ii. High x_2 lowers yield (y)

iii. Press (x_1) has no effect on y

Get non-causal model between $x1$ and y

Randomization + Blocking

Simple Comparative Experiment:

- Effect of two treatments on strength of rubber

(i)

A	A	A	A	- - - - -	A	B	B	B	- - - - -	B	

$$\begin{array}{cc} \underline{A} & \underline{B} \\ y_{A_1} & y_{B_1} \\ \vdots & \vdots \\ \underline{y_{A_{n_A}}} & \underline{y_{B_{n_B}}} \\ \bar{y}_A & \bar{y}_B \end{array}$$

$$H_0 : E(y_A) = E(y_B)$$
$$H_1 : E(y_A) \neq E(y_B)$$

t-Test: $$\frac{\bar{y}_A - \bar{y}_B}{s_p^2 \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \sim t_{n_A+n_B-2}$$

Problems with this?

Randomization + Blocking

(ii) What if strip of rubber had variations (eg. thickness) along its length?

Then $\bar{y}_A - \bar{y}_B$ might just be reflecting this difference
ie. thickness = lurking variable

One solution → Randomize allocation of rubber pieces to treatments (A +B)
eg. Flipping a coin

A	B	B	A	B	A	B	B	B	A	- - - -	A	B	B	

$\therefore$

- Randomly allocates any lurking variable effect to A and B
- Ensures validity of hypothesis test
- Randomized design

Problems?

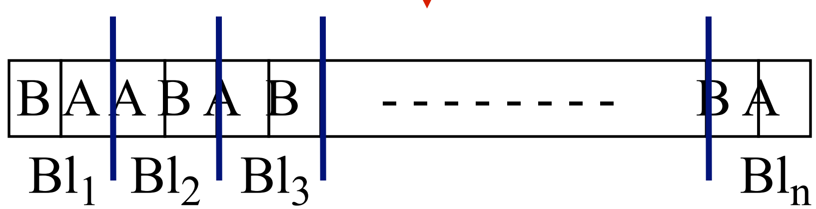
Randomization + Blocking

(iii) Suppose we expect variation in rubber to be progressive along length of the strip*!*

Then two adjacent pieces will be much more similar than two distant ones

$\therefore$

- Block into pairs of adjacent pieces
- Assign treatments (A,B) RANDOMLY within blocks
- Randomized block design



Randomization + Blocking

(iii)

Only compare within blocks →

Block	A	B	difference $d = y_A - y_B$
Bl_1	y_{A_1}	y_{B_1}	$d_1 = y_{A_1} - y_{B_1}$
Bl_2	y_{A_2}	y_{B_2}	$d_1 = y_{A_2} - y_{B_2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
Bl_n	y_{A_n}	y_{B_n}	$d_1 = y_{A_n} - y_{B_n}$
			$\bar{d}$

Blocking removes effect of possible uncontrolled variations along length of strip

$\therefore$ $\bar{d}$ better measure of $\mu_A - \mu_B$ than $\bar{y}_A - \bar{y}_B$

H_0: E(d)=0

Paired t-Test: $\dfrac{\bar{d} - 0}{s_{\bar{d}}} \sim t_{n-1}$

$$s_{\bar{d}}^2 = \frac{s_d^2}{n}$$

Concepts in Design of Experiments

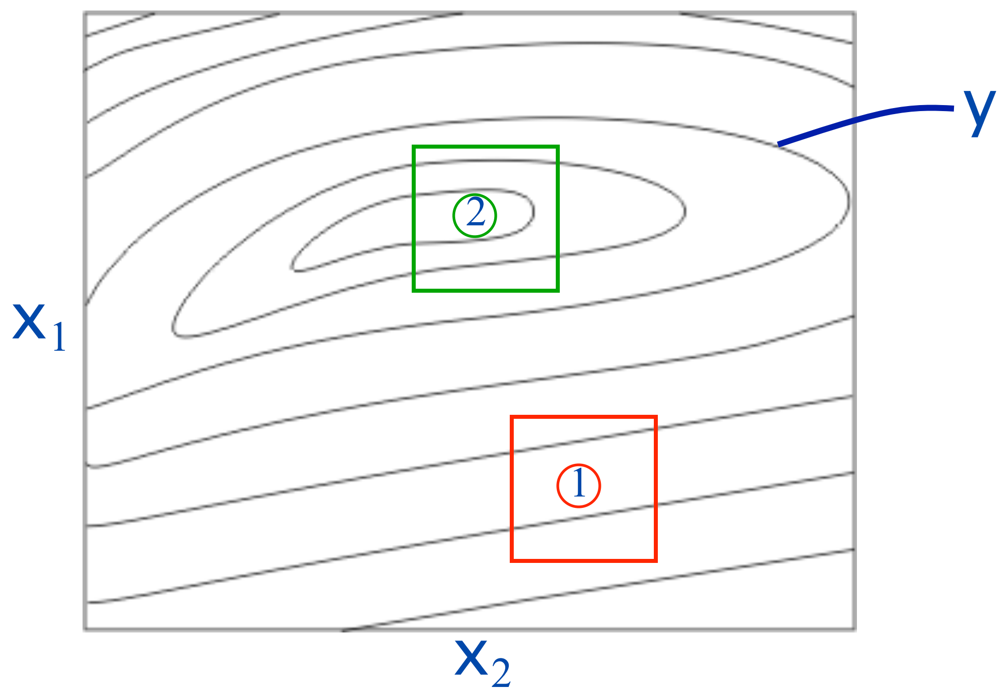
Designs for Empirical Studies

1. **Screening Studies**: Discovering which of a large number of variables affect response.
2. **Empirical model building studies**

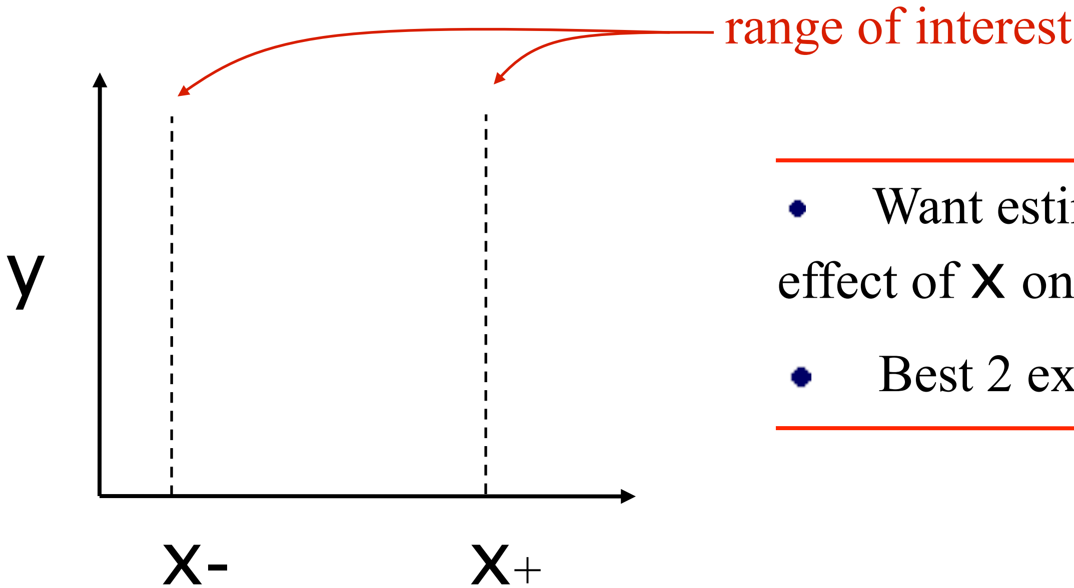
$$y = f(x_1, x_2 \ldots x_k)$$

True model unknown. Use approximate models.

- Region 1: Linear model OK
- Region 2: Need model quadratic in X's



2^k Factorial Designs



- Want estimate of linear effect of x on y.
- Best 2 experiments?

If fit LS model: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{\beta}_1 =$ Effect of changing x by one unit

Effect on y of changing x from x- to x+ is $(y_2 - y_1)$ ← Main effect of x

- Linear effect only (two level experiment)

2^2 Factorial Design

- 2 independent variables:

	Range
Temperature (T)	160°C - 180°C
Concentration (C)	20% - 40%

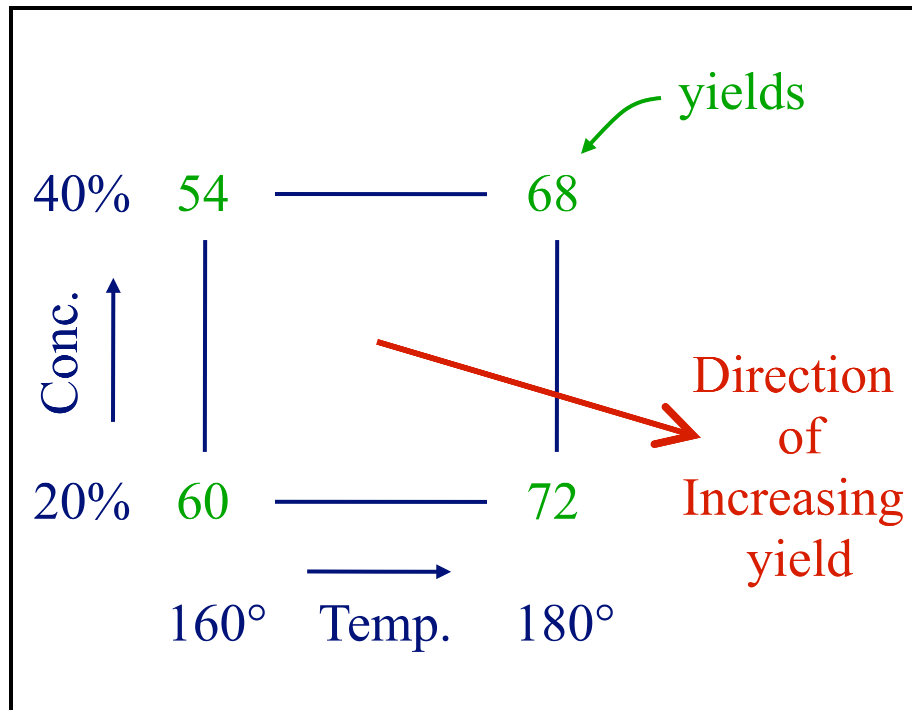
Study effect of $T + C$ on yield **y**

two variables

Design: 2^2 Factorial in $2^2 = 4$ runs

two levels

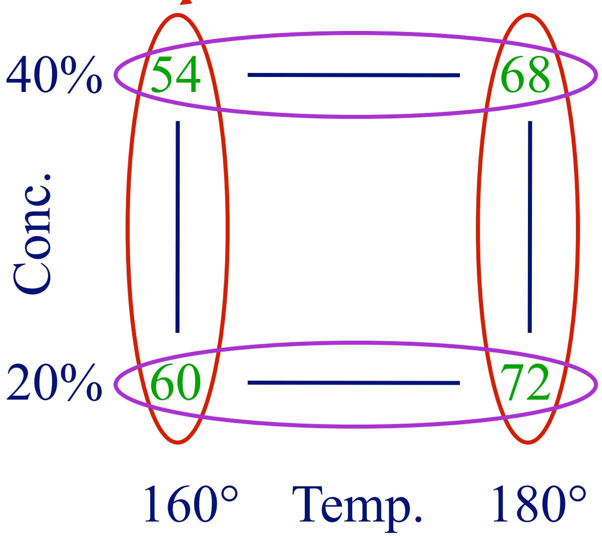
all possible combination
of 2 levels of 2 variables



2^k Factorial Designs

Main Effects of *T* + C

Two measures of effect of *C*



$54 - 60 = -6$

$68 - 72 = -4$

Avg. = - 5

Main effect of *C*

Two measures of main effect of *T*

$68 - 54 = 14$

$72 - 60 = 12$

Avg. = 13 % / 20°C change in *T*

Main effect of *T*

2^k Factorial Designs

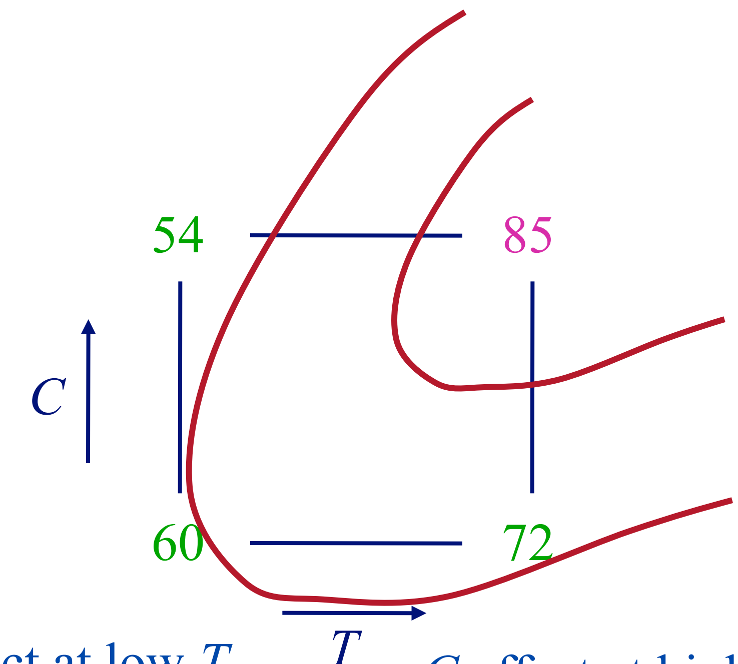
Interaction between T + C

- Do variables T + C act independently on y?
- Is effect of T same at both levels of C?
- Is effect of C same at both levels of T?

If effect is different $\Rightarrow$ T x C interaction

Above example $\longrightarrow$ very little interaction

But change 68 $\longrightarrow$ 85



T effect at high C

85 – 54 = 31

T effect at low C

72 – 60 = 12

C effect at low T

54 – 60 = – 6

C effect at high T

85 – 72 = +13

Large T x C interaction

2^k Factorial Designs

Design

T	C
160°	20
180°	20
160°	40
180°	40

Centre condition

$T = 170°$

$C = 30\%$

Transform $T + C$ to scaled variables

$$x_i = \frac{\text{Variable} - \text{Centre point}}{\text{Range}/2}$$

$$x_1 = \frac{T - 170°}{10}$$

$$x_2 = \frac{C - 30}{10}$$

Range of x_i's

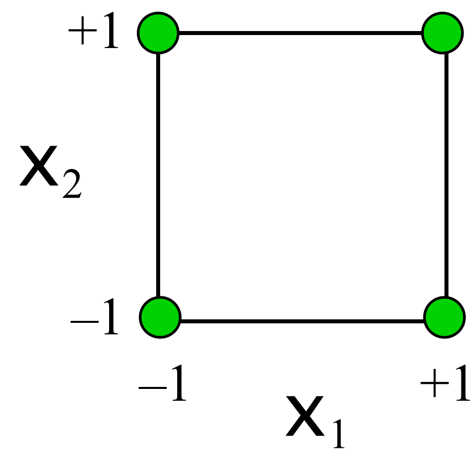
–1 to +1

–1 to +1

2^k Factorial Designs

Design matrix becomes

X_1	X_2
−1	−1
+1	−1
−1	+1
+1	+1



Fit model : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$

Use least squares regression to estimate parameters (effects)

Note: I will denote $\hat{\beta}_i =$ effect of variable X_i

($\hat{\beta}_i =$ effect on y of changing X_i from 0 $\longrightarrow$ +1)

Most texts denote "effect of X_i" = change in y due to changing X_i from −1 to +1 $\longrightarrow$ ie $= 2\hat{\beta}_i$

95% Confidence intervals on β_i's

$$\hat{\beta}_i \pm t_{\nu,0.025} \sqrt{\frac{s^2}{\sum x_i^2}}$$

Does the interval include "zero"**?** $\Longrightarrow$ Not significant

2^k Factorial Designs

2^3 Factorial Design

2^3 ← 3 variables

↑ 2 levels

($2^3 = 8$ runs)

Variables: T, C, Catalyst type (eg. A, B)

qualitative variable

Denote: $x_3 = -1$ for catalyst A

$x_3 = +1$ for catalyst B

2^3 factorial = All combinations of the 2 levels of the 3 variables

2^3 Factorial Design

Run order	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
6	+1	–1	–1	–1	+1	+1	+1	–1
3	+1	+1	–1	–1	–1	–1	+1	+1
1	+1	–1	+1	–1	–1	+1	–1	+1
7	+1	+1	+1	–1	+1	–1	–1	–1
2	+1	–1	–1	+1	+1	–1	–1	+1
8	+1	+1	–1	+1	–1	+1	–1	–1
5	+1	–1	+1	+1	–1	–1	+1	–1
4	+1	+1	+1	+1	+1	+1	+1	+1

Randomized

Design Matrix

x (indep. var. matrix)

2^3 Factorial Design

$$
\begin{aligned}
y = \beta_0 &+ \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\
&+ \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon
\end{aligned}
$$

Again fit by Least Squares Regression $\longrightarrow$ $\hat{\beta}_i = \dfrac{\sum x_i y}{\sum x_i^2}$

Note that all columns of **X** are orthogonal

- Implies that can estimate all effects independent of the others

---

- 2^k Factorial in **k** variables can easily be written down in standard form

2^k Factorial Designs

Desirable Features of Factorial Designs

i. Orthogonal → easy calculations
→ uncorrelated estimates

ii. Good variation in all variables

iii. Efficient use of all data points

iv. Well patterned design → Good visual appreciation

v. Allows experiments to be performed in blocks
(Fractional Factorials)

vi. Allows designs of increasing order to be built up sequentially

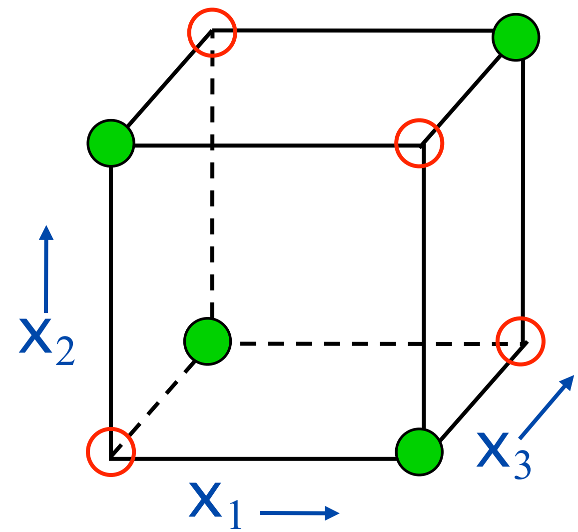
Blocking of a 2^3 Factorial

Want to examine 3 factors in a $2^3 = 8$ run design. But material to be used in experiment comes in batches sufficient for only 4 runs, and differences may exist between batches of material.

Can we split the design so that differences in the material will not affect the results?

Split on this column

x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	
–	–	–	+	+	+	–	○
+	–	–	–	–	+	+	●
–	+	–	–	+	–	+	●
+	+	–	+	–	–	–	○
–	–	+	+	–	–	+	●
+	–	+	–	+	–	–	○
–	+	+	–	–	+	–	○
+	+	+	+	+	+	+	●



Run all experiments with + sign in $x_1x_2x_3$ column in one block and all – signs in other!

(Randomize within blocks)

Blocking of a 2^3 Factorial

$\therefore$ Any block effect (ie. differences in material) will be CONFOUNDED with 3-factor interaction term $X_1X_2X_3$.

$\therefore$ Can't tell whether $\hat{\beta}_{123}$ is due to a real $X_1X_2X_3$ interaction or a block effect (material)

$$\left(\text{ie.} \quad \hat{\beta}_{123} = \underbrace{x_1 x_2 x_3 \text{ effect}}_{\text{expected to be small anyway}} + \text{block effect} \right)$$

Since $X_1X_2X_3$ column is orthogonal to all other columns, any block effect will have no influence on them !

Designs for 2^{nd} Order Models

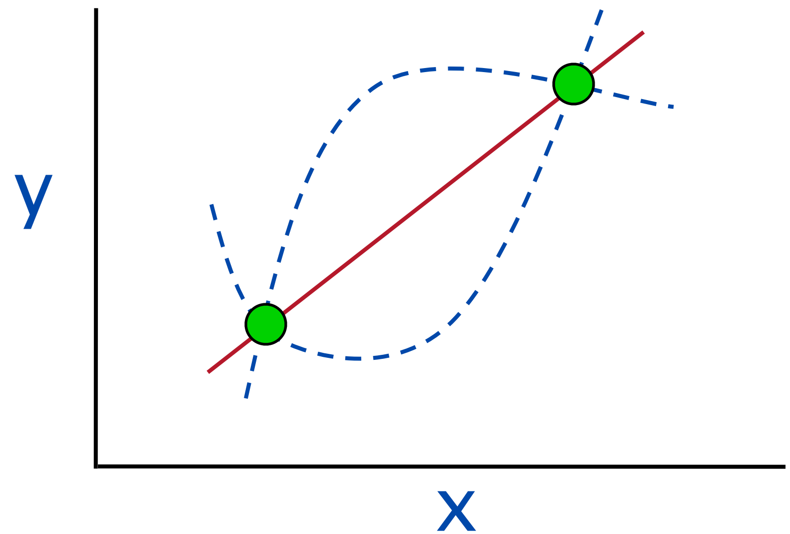
First order + interaction model may exhibit Lack of Fit

or

Prior knowledge may tell us we need second order terms

$$x_1^2,\ x_2^2,\ \ldots$$

Need more than 2 levels designs



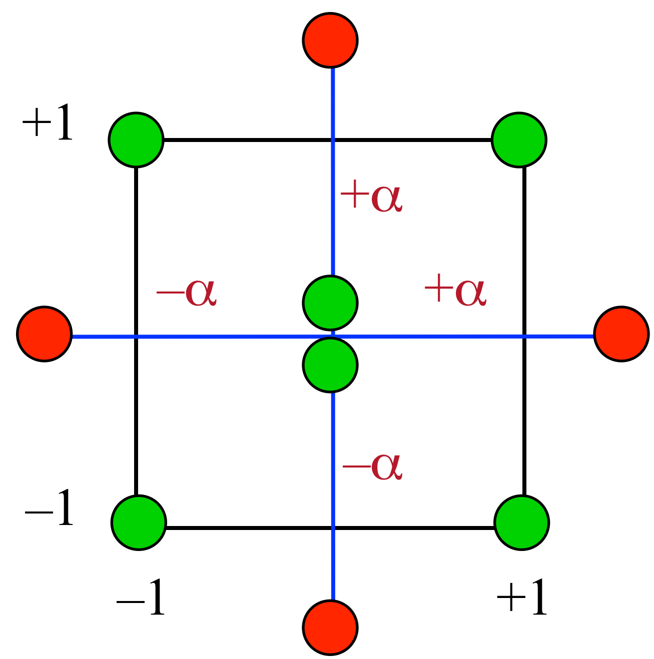
2^k Factorial Designs

Central Composite Designs

1. Start with 2^k or 2^{k-p} design with centre points

2. Add vertices of star

x_1	x_2	
-1	-1	2^2
$+1$	-1	
-1	$+1$	
$+1$	$+1$	
0	0	cp
0	0	
$-\alpha$	0	star
$+\alpha$	0	
0	$-\alpha$	
0	$+\alpha$	



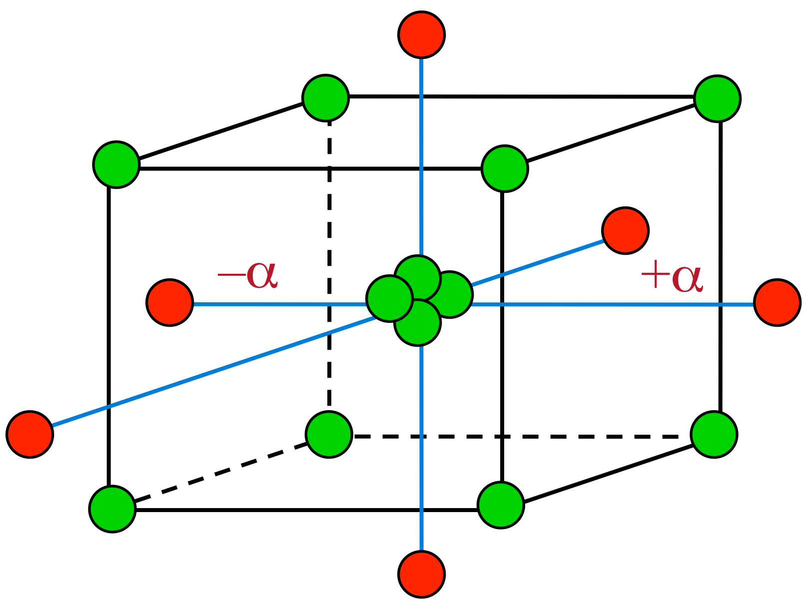
For 2 variable design

$\alpha = \sqrt{2} = 1.414$ is good choice

2^k Factorial Designs

Central Composite Designs

3 variables: 2^3 + cp + star



$\alpha = 1.68$

	x_1	x_2	x_3
2^3	-1	-1	-1
	$+1$	-1	-1
	-1	$+1$	-1
	$+1$	$+1$	-1
	-1	-1	$+1$
	$+1$	-1	$+1$
	-1	$+1$	$+1$
	$+1$	$+1$	$+1$
cp	0	0	0
	0	0	0
	0	0	0
	0	0	0
star	$-\alpha$	0	0
	$+\alpha$	0	0
	0	$-\alpha$	0
	0	$+\alpha$	0
	0	0	$-\alpha$
	0	0	$+\alpha$

2^k Factorial Designs

Central Composite Designs

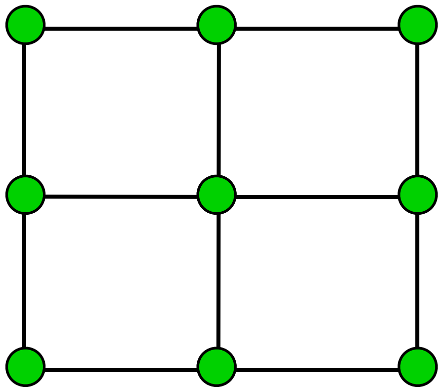
For $k = 4$: $2^4 + cp + star$

For $k > 4$: $2^{k-p} + cp + star$

k	Design	α (for rotatability)
2	2^2	1.414
3	2^3	1.68
4	2^4	2.0
5	2^{5-1}	2.0
6	2^{6-1}	2.38

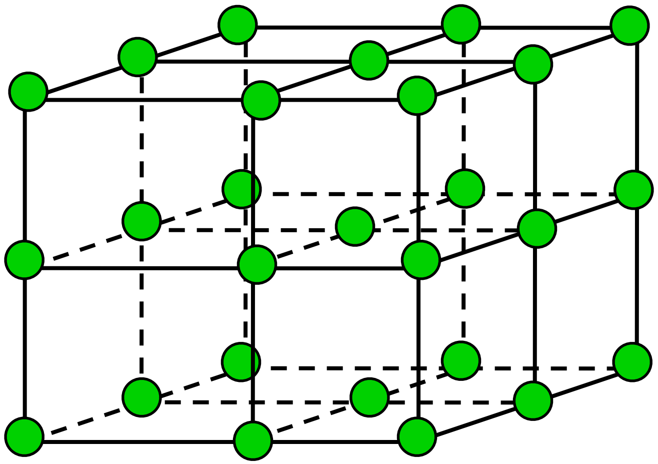
3 – Level Factorials

3^2 design: 2 variables at all combinations of 3 levels



3^3 design:

27 runs



3 – Level Factorials

Fit full quadratic model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$+ \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

$$+ \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \epsilon$$

(10 parameters)

Allows for approximation of many responses

Most statistical software provides 2-D and 3-D plotting to examine response surface.

Response Surface Methods (RSM)

BH^2 Chapter 15
M + R Chapter 12.9

Empirical (data – driven) approach to process optimization

1. Design experiment in region of interest
2. Build model : $\hat{y} = f(x_1, x_2 \ldots x_k)$
3. Use model to find new conditions $x_1, x_2 \ldots x_k$ that will improve a single response $\hat{y}_i$ or give good region for several responses
4. Repeat steps 1, 2 and 3 until attain optimal conditions

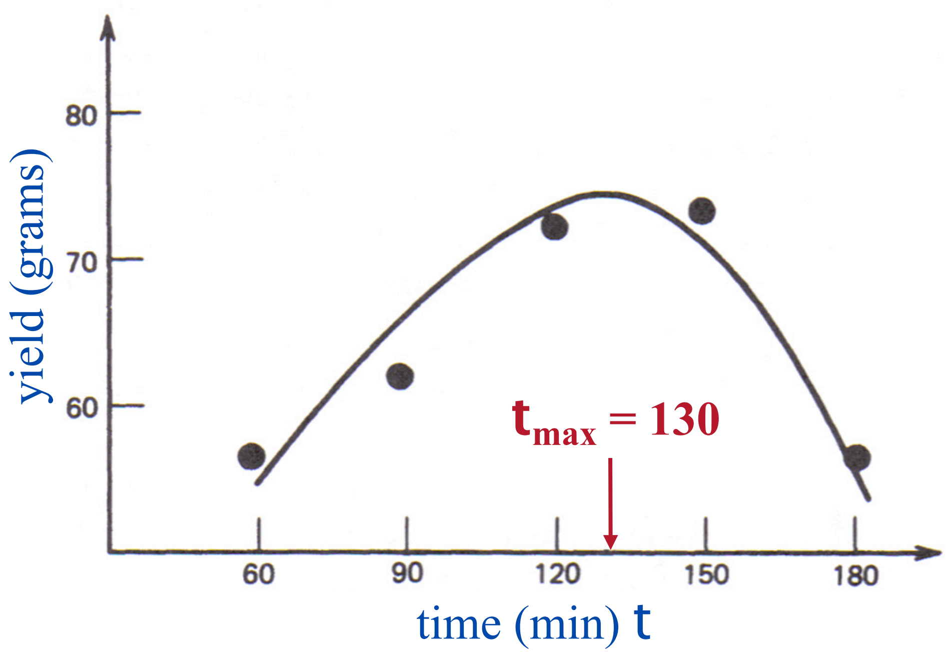
Response Surface Methods (RSM)

Problem with COST approach

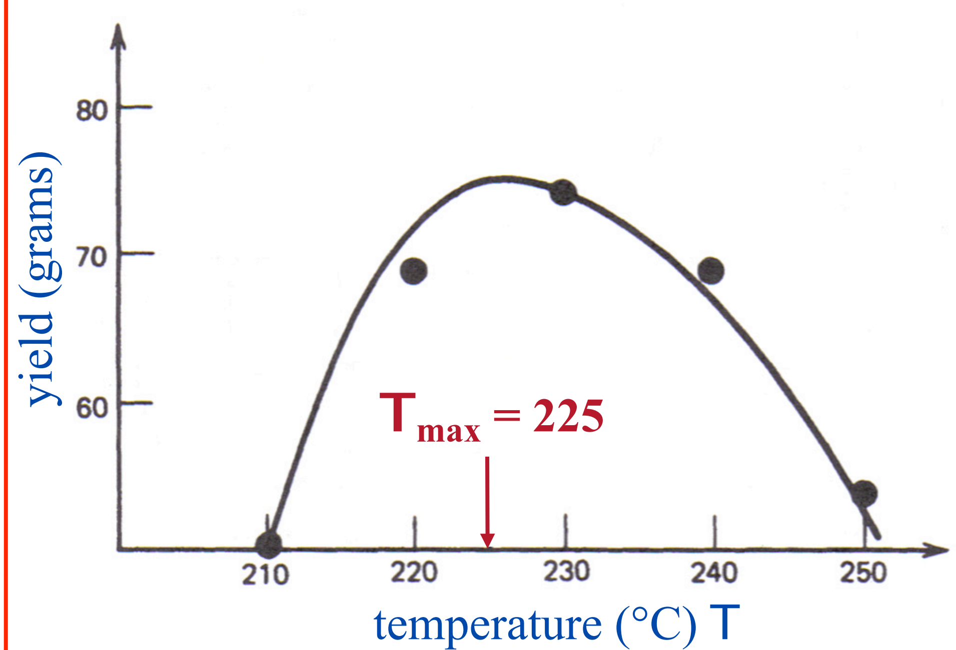
Ex. Maximize yield of a reaction by choice of :

- reaction time (t)
- reaction temperature (T)

Fix T = 225°C
Vary t (60 min → 180 min)

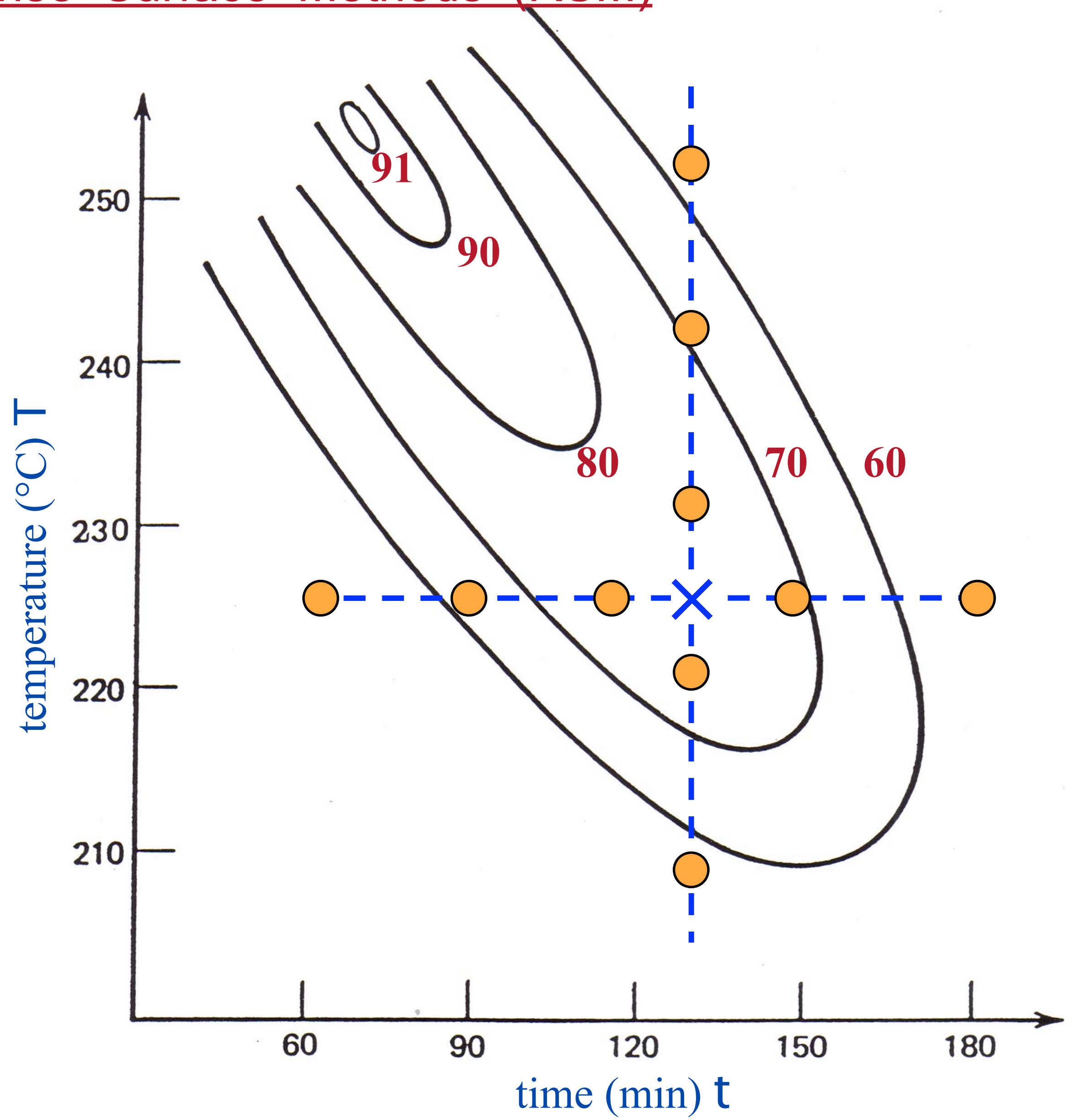


Fix t = 130 min
Vary (210°C → 250°C)



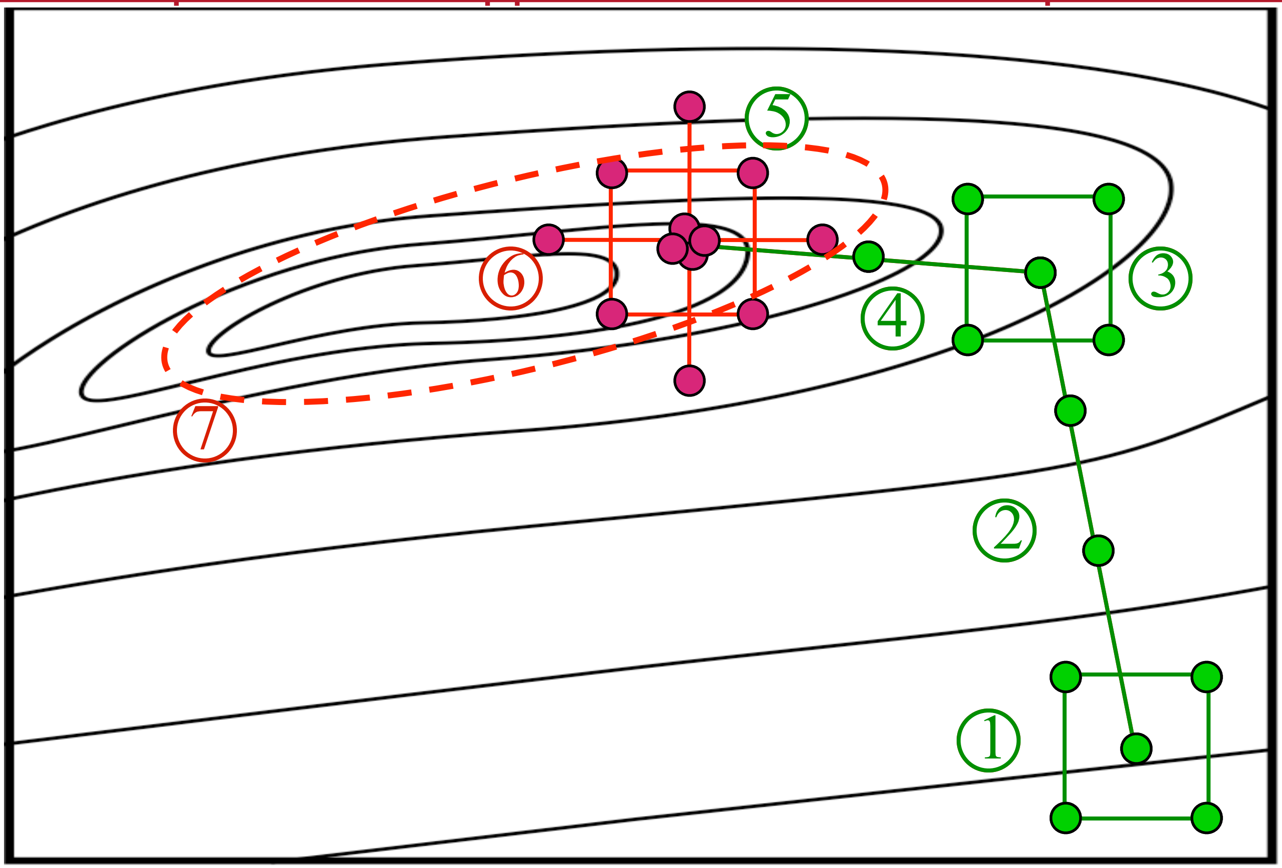
Response Surface Methods (RSM)

Possible true response surface representing yield versus reaction time and temperature, with points shown for one-variable-at-a-time approach



2^k Factorial Designs

RSM – Experimental Approach to Process Optimization



x_1

x_2

⑥ If curvature and/or interaction large relative to main effects then add star points ⟶ 2nd order central composite design.

⑦ Plot response contours.

2^k Factorial Designs

RSM – Experimental Approach to Process Optimization

① Perform factorial (or fractional factorial) design about current operating conditions

Fit linear model : $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2$

significant small

② Calculate direction of Steepest Ascent + perform experiments along this direction until response doesn't improve.

Path of SA :

$$\frac{\partial \hat{y}}{\partial x_1} = \hat{\beta}_1$$

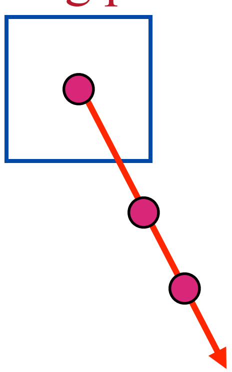
$$\frac{\partial \hat{y}}{\partial x_2} = \hat{\beta}_2$$

(If $\hat{\beta}_{12}$ term is small ie model is linear)

∴ Move x_1 $\hat{\beta}_1$ units in direction of x_1 for every $\hat{\beta}_2$ units in direction of x_2

eg. $\hat{y} = 3.5 + 1.5x_1 - 3.0x_2$

Points along path of SA:



x_1	x_2
0	0
1.0	–2.0
1.5	–3.0
3.0	–6.0

RSM – Experimental Approach to Process Optimization

3 Lay down a new Factorial design:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2$$

4 If linear terms still large + $\hat{\beta}_{12}$ small again calculate direction of SA + perform experiments along it.

5 3rd Factorial design

$\longrightarrow$ Clear lack of Fit of Linear Model

- Linear effects $\hat{\beta}_1$, $\hat{\beta}_2$ small
- Interaction term $\hat{\beta}_{12}$ large

RSM – Experimental Approach to Process Optimization

5 • Check on curvature or quadratic effect ($\beta_{11}x_1^2 + \beta_{22}x_2^2$ terms)? Can if have center points !

$$\begin{array}{rcc l}
 & x_1 & x_2 & \\
\hline
 & - & - & \\
 & + & - & \bar{y}_f = \dfrac{y_1 + y_2 + y_3 + y_4}{4} \\
 & - & + & \\
 & + & + & \\
\text{cp} & 0 & 0 & \bar{y}_{cp} = \dfrac{y_{\text{cp},1} + y_{\text{cp},2}}{2} \\
 & 0 & 0 &
\end{array}$$

If model were linear (ie $\hat{\beta}_{11} = \hat{\beta}_{22} = 0$) then $\hat{\beta}_0 = \bar{y}_f$ would be estimate of response at centre of design ($\hat{y}(x = 0)$)

$\therefore \quad \bar{y}_f - \bar{y}_{\text{cp}} \longrightarrow$ Estimate of ($\hat{\beta}_{11} + \hat{\beta}_{22}$) (curvature)

RSM – Experimental Approach to Process Optimization

⑥ If curvature and/or interaction large relative to main effects then add star points → 2^{nd} order central composite design.

Fit full second order model:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2 + \hat{\beta}_{11} x_1^2 + \hat{\beta}_{22} x_2^2$$

⑦
- Plot response contours. 2-D or 3-D plot routines in all statistical design software (MINITAB, MODDE)
- Examine response surface and move towards best conditions

For k > 3 variables, use fractional factorials

Fractional Factorial Designs

- With many variables ($k > 4$) factorial designs need many runs
- High order interaction terms (eg. $x_1x_2x_3x_4$) are not of interest
- Perform a fraction of the full design
  - fewer runs
  - but enable estimates of terms of interest

Fractional Factorial Designs

Half Fractions – 2^{k-1} Designs

Example : → ½ Fraction of a full 2^3 factorial
→ $2^{3-1} = 2^2 = 4$ runs (but which 4?)

—— Full 2^3 factorial (8 runs) ——

Run #	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	–1	–1	–1	+1	+1	+1	–1
2	+1	–1	–1	–1	–1	+1	+1
3	–1	+1	–1	–1	+1	–1	+1
4	+1	+1	–1	+1	–1	–1	–1
5	–1	–1	+1	+1	–1	–1	+1
6	+1	–1	+1	–1	+1	–1	–1
7	–1	+1	+1	–1	–1	+1	–1
8	+1	+1	+1	+1	+1	+1	+1

- As with blocking the 2^3 factorial into 2 blocks choose 4 runs with $x_1x_2x_3 = +1$ (–1 gives other half fraction)

Fractional Factorial Designs

Half Fractions – 2^{k-1} Designs

2^{3-1} fraction with $x_1 x_2 x_3 = +1$

Run #	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
5	–1	–1	+1	+1	–1	–1	+1
2	+1	–1	–1	–1	–1	+1	+1
3	–1	+1	–1	–1	+1	–1	+1
8	+1	+1	+1	+1	+1	+1	+1

But ! have only 4 runs
can't estimate all 8 terms in the model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

Fractional Factorial Designs

Half Fractions – 2^{k-1} Designs

Note:

x_1 column = x_2x_3 column

x_2 column = x_1x_3 column

x_3 column = x_1x_2 column

$x_1x_2x_3$ column = I column

These effects are CONFOUNDED

Can only fit the 4 parameter model :

$$y = l_0 + l_1 x_1 + l_2 x_2 + l_3 x_3 + \epsilon$$

where l_1 = estimate of $(\beta_1+\beta_{23})$

l_2 = estimate of $(\beta_2+\beta_{13})$

l_3 = estimate of $(\beta_3+\beta_{12})$

l_0 = estimate of $(\beta_0+\beta_{123})$

Fractional Factorial Designs

How to determine what Effects are Confounded ?

1) Can always examine columns of all effects and look for identical columns

$\longrightarrow$ tedious for $k > 3$

2) Systematic method

- Chose runs because $x_1 x_2 x_3 = +1$
- Defining relation : $I = x_1 x_2 x_3$ (columns of $+1$'s)
- Note the following:
  - Any column multiplied by identity column (I) is equal to itself

    $$x_1 \cdot I = x_1$$

  - Any column multiplied by itself produces the identity column (I)

    $$x_1 \cdot x_1 = I$$

- Use defining relation to unravel confounding pattern

$$I = x_1 x_2 x_3$$

$$x_1 \cdot I = x_1 \cdot x_1 x_2 x_3 = x_2 x_3$$

$$x_2 \cdot I = x_1 x_2 \cdot x_2 x_3 = x_1 x_3$$

$\vdots$

Fractional Factorial Designs

Construction of 2^{3-1} Fractional Factorial ($I = x_1 x_2 x_3$)

Write down full 2^2 design in $x_1\ x_2$

Note that $x_3 = x_1 x_2$.

Therefore assign variable x_3 to $x_1 x_2$ column :

x_1	x_2	$x_3 = x_1 x_2$
-1	-1	$+1$
$+1$	-1	-1
-1	$+1$	-1
$+1$	$+1$	$+1$

Other half fraction of 2^3 design given by defining relation $I = -x_1 x_2 x_3$

Assign x_3 with $-x_1 x_2$ column

x_1	x_2	$x_3 = -x_1 x_2$
-1	-1	$+1$
$+1$	-1	-1
-1	$+1$	-1
$+1$	$+1$	$+1$

Combining these two fractions gives full 2^3 factorial in 2 blocks

2^{4-1} Fraction Factorial Design

Write down full 2^3 factorial (3 variables, 8 runs)

Associate variable x_4 with the $x_1x_2x_3$ column

Run #	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_4 = x_1x_2x_3$
1	–1	–1	–1	+1	+1	+1	–1
2	+1	–1	–1	–1	–1	+1	+1
3	–1	+1	–1	–1	+1	–1	+1
4	+1	+1	–1	+1	–1	–1	–1
5	–1	–1	+1	+1	–1	–1	+1
6	+1	–1	+1	–1	+1	–1	–1
7	–1	+1	+1	–1	–1	+1	–1
8	+1	+1	+1	+1	+1	+1	+1

2^{4-1} Fraction Factorial Design

Confounding of effects ?

Defining relation $I = x_1x_2x_3x_4$

$$x_1 \cdot I = (x_1 \cdot x_1)\, x_2x_3x_4 = x_2x_3x_4$$

$$x_2 \cdot I = x_1\,(x_2 \cdot x_2)\, x_3x_4 = x_1x_3x_4$$

$$x_3 \cdot I = x_1x_2\,(x_3 \cdot x_3)\, x_4 = x_1x_2x_4$$

$$x_4 \cdot I = x_1x_2x_3\,(x_4 \cdot x_4) =$$

$$x_1x_2x_3$$

$$x_1x_2 \cdot I = (x_1 \cdot x_1)\,(x_2 \cdot x_2)\, x_3x_4 = x_3x_4$$

$$x_1x_3 \cdot I = (x_1 \cdot x_1)\, x_2\,(x_3 \cdot x_3)\, x_4 = x_2x_4$$

$$x_2x_3 \cdot I = x_1\,(x_2 \cdot x_2)\,(x_3 \cdot x_3)\, x_4 = x_1x_4$$

2^{4-1} Fraction Factorial Design

Fit 8 parameter model:

$$y = l_0 + l_1 x_1 + l_2 x_2 + l_3 x_3 + l_4 x_4 + l_{12} x_1 x_2 + l_{13} x_1 x_3 + l_{23} x_2 x_3 + \epsilon$$

$l_0 \longrightarrow$ estimate of $(\beta_0+\beta_{1234})$

$l_1 \longrightarrow$ estimate of $(\beta_1+\beta_{234})$

$\vdots$

$l_{12} \longrightarrow$ estimate of $(\beta_{12}+\beta_{34})$

$l_{13} \longrightarrow$ estimate of $(\beta_{13}+\beta_{24})$

$\vdots$

- Often 3 factor interactions are small
- If we ignore 3 factor interactions then 2^{4-1} fractional factorial will give estimates of
  - All main effects β_1, β_2, β_3, β_4
  - 3 combinations of 2 factor interactions $(\beta_{12}+\beta_{34})$, $(\beta_{13}+\beta_{24})$, $(\beta_{14}+\beta_{23})$

Fractional Factorial Designs

Saturated Fractional Factorial Designs

- Useful for screening studies
- Study N variables in N – 1 runs

Consider 2^{7-4} design for 7 variables in 8 runs (a $2^{-4} = 1/16$ fraction of a full 2^7 design)

- Write down full 2^3 design (3 variables in 8 runs)
- Associate additional variables with all the interaction columns

x_1	x_2	x_3	$x_4 = x_1x_2$	$x_5 = x_1x_3$	$x_6 = x_2x_3$	$x_7 = x_1x_2x_3$
–1	–1	–1	+1	+1	+1	–1
+1	–1	–1	–1	–1	+1	+1
–1	+1	–1	–1	+1	–1	+1
+1	+1	–1	+1	–1	–1	–1
–1	–1	+1	+1	–1	–1	+1
+1	–1	+1	–1	+1	–1	–1
–1	+1	+1	–1	–1	+1	–1
+1	+1	+1	+1	+1	+1	+1

Saturated Fractional Factorial Designs

- Many effects are now confounded
- Fit model :

$$y = l_0 + l_1 x_1 + l_2 x_2 + l_3 x_3 + \cdots + l_7 x_7 + \epsilon$$

$l_1 \longrightarrow$ estimate of $(\beta_1 + \beta_{24} + \beta_{35} + \beta_{67}$ + higher order interaction)

$l_2 \longrightarrow$ estimate of $(\beta_2 + \beta_{14} + \beta_{36} + \beta_{57}$ + higher order interaction)

$\vdots$

Fractional Factorial Designs

Resolve Confounding by Addition of other Fractions

- Many confounded effects after running first 2^{7-4} design
- Run another fraction to resolve conflicts Switch signs of all variables in first design (fold over)
- Adding these two 8 runs 1/16 fractions together gives a 16 run 1/8 fraction of the full 2^3 design

x_1	x_2	x_3	x_4	x_5	x_6	x_7
+	+	+	−	−	−	+
−	+	+	+	+	−	−
+	−	+	+	−	+	−
−	−	+	−	+	+	+
+	+	−	−	+	+	−
−	+	−	+	−	+	+
+	−	−	+	+	+	+
−	−	−	+	−	+	−

- Can estimate the following effects:

$$l_1 = \beta_1 \qquad l_{24} = (\beta_{24} + \beta_{35} + \beta_{67})$$

$$l_2 = \beta_2 \qquad l_{14} = (\beta_{14} + \beta_{36} + \beta_{57})$$

$$l_3 = \beta_3$$

$$l_4 = \beta_4 \qquad \vdots$$

$$l_5 = \beta_5$$

$$l_6 = \beta_6$$

$$l_7 = \beta_7 \qquad l_{34} = (\beta_{34} + \beta_{25} + \beta_{16})$$

Optimal Designs

"Optimal Designs" are designs which optimize some objective function via the choice of design conditions (**<u>X</u>**)

Consider a chosen model form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon$$

Var($\hat{\beta}$) = $(X^T X)\sigma_e^2$

Where X = matrix of settings for the independent variables ($x_1,\ x_2\ \cdots\ x_1^2$) for N experiments

D-Optimal Design

$$\max_{x_{1i},\, x_{2i}(i=1,\, 2\, \cdots\, N)} |X^T X|$$

$\longrightarrow$ N experiments yielding the smallest uncertainty for $\hat{\beta}$'s

Important Areas for Optimal Designs

(1) Constrained Design Regions

Want to fit model $\eta = x\beta$ in a constrained experimental region

Ex. Mixture Designs

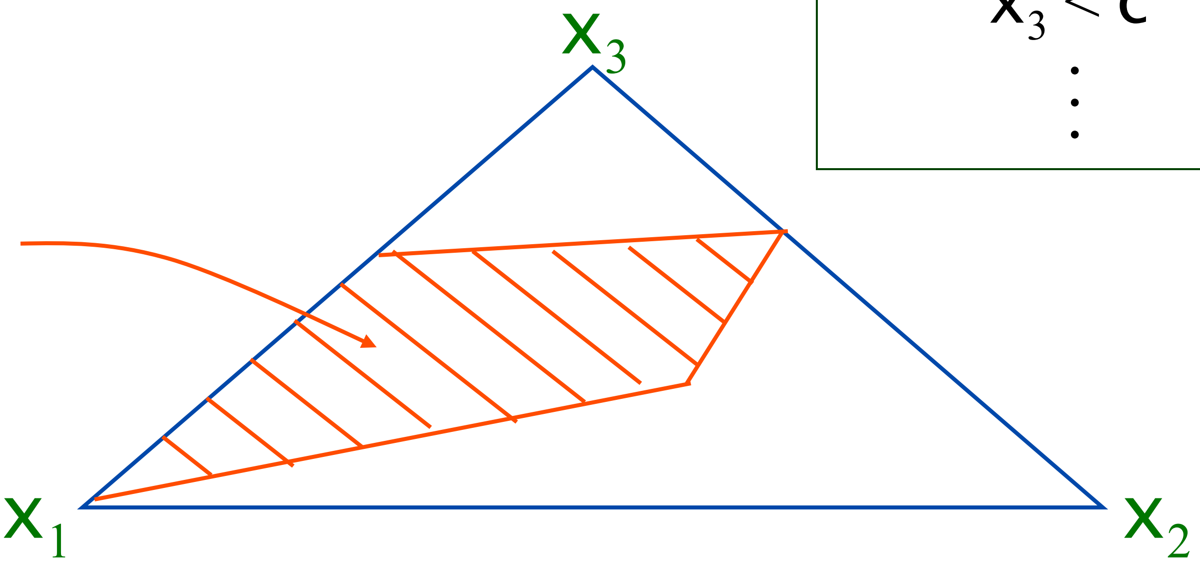
Constraints

$$x_1 + x_2 + x_3 = 1$$

$$x_3 < c$$

$$\vdots$$

Want best design for this region



Important Areas for Optimal Designs

(2) Sequential Designs

- Have already run n experiments

- Find another m experiments which, when add to the first n will give the most information on parameters

- Allows sequential experimentation
  - Perform some runs
  - Analyze data
  - Plan new runs

- Fix up a set of existing, poorly designed experiments
  - Places a XXX new experiments in important regions overlooked in existing data