Designs for experimental studies

※Objectives

- Screening studies

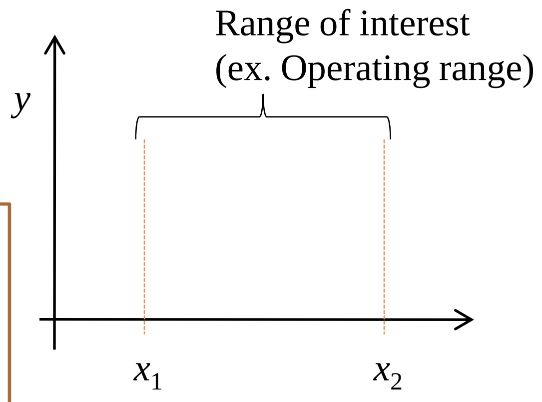
  : discovering which of a large number of variations affect response

- Empirical model building studies

  : true model unknown. Use approximate models, $y = f(x_1, x_2, \ldots, x_k)$

2^k factorial designs

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



2^2 factorial design

- Two independent variables:

	range
Temperature (T)	160°C ~ 180°C
Concentration (C)	20% ~ 40%

  - Study effect of T & C on yield y.

- Design: 2^2 factorial in $2^2 = \underline{4 \text{ runs}}$

Two variables

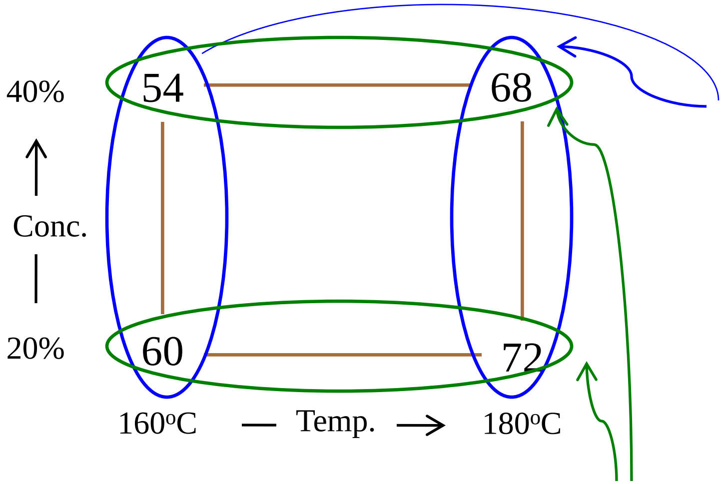
Two levels

All possible combination
of two levels of two variables

- & run the experiments:

2^2 factorial design (Cont.)

- Main effects of T & C



Main effect of C

Two measures of effect of C

54 – 60 = -6

68 – 72 = -4

avg. = -5

Main effect of T

Two measures of effect of T

68 – 54 = 14

72 – 60 = 12

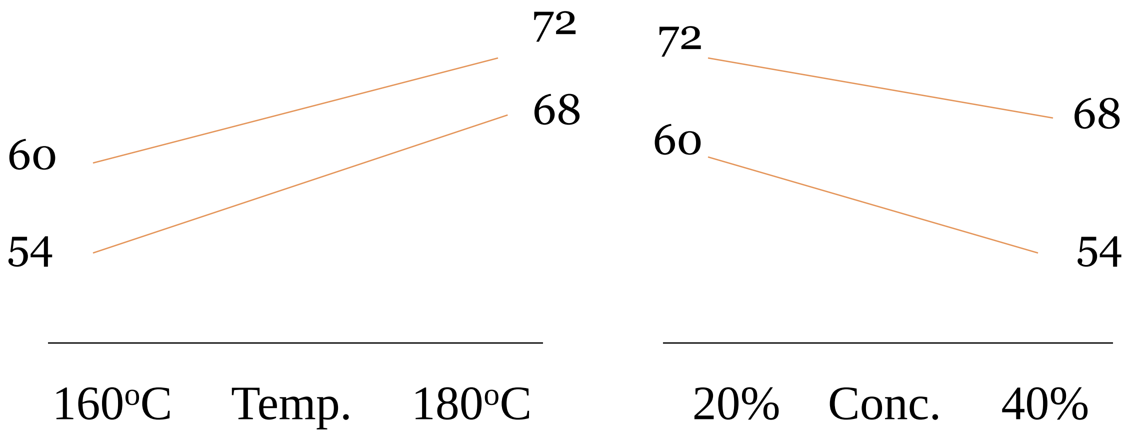
avg. = 13

※ 13 % yield/20 ^{o}C change in T

2^2 factorial design (Cont.)

- Interaction between T & C
  - Do variables T & C act independent on *y*?
  - Or, is effect of T (or C) same at both levels of C (or T)?
  - If effect is different → T x C interaction.

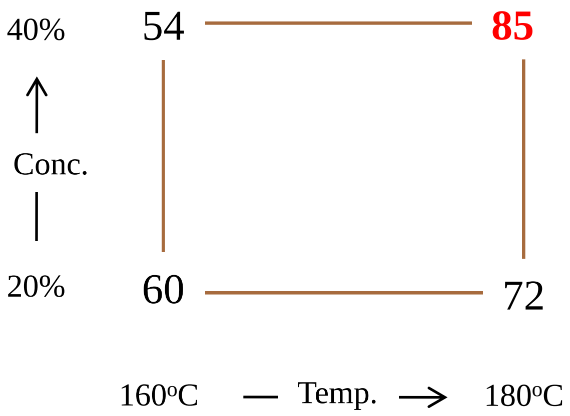
Visualize this with an interaction plot.



Lines are roughly parallel.

2^2 factorial design (Cont.)

- But change 68 to 85



Main effect of C

Two measures of effect of C

54 – 60 = -6

85 – 72 = 13

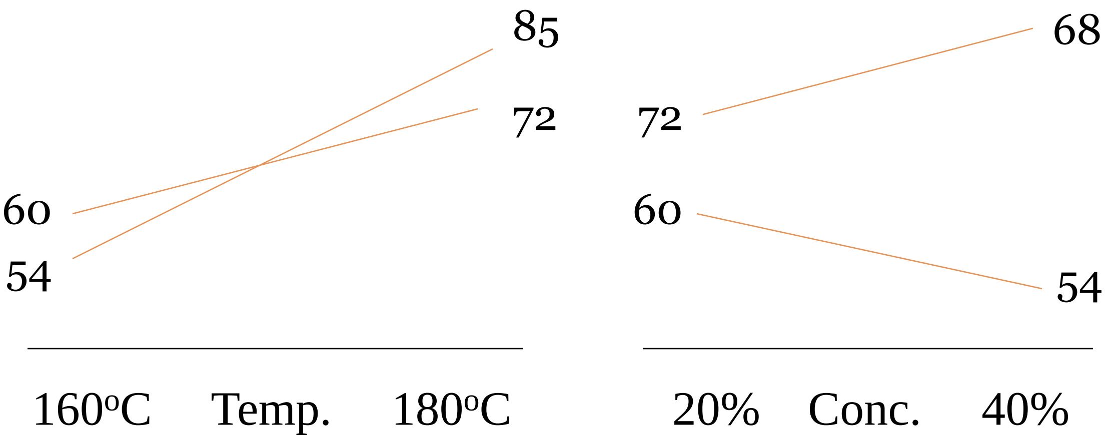
Lines in interaction plots will be far from being parallel → **Large interaction** between T & C

Main effect of T

Two measures of effect of T

85 – 54 = 31

72 – 60 = 12



2^2 factorial design (Cont.)

- Analysis by least squares
  - Design matrix (condition) & experimental results

T	C	y
160	20	60
180	20	72
160	40	54
180	40	68

※Center: usually current condition

  - Transform x variable (T & C) to scaled variables

※why?: remove scale effect

$$x_i = \frac{\text{variable} - \text{centerpoint}}{\text{Range}/2}$$

$$x_1 = \frac{\text{T} - 170°C}{10}$$

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

Range of x_i's

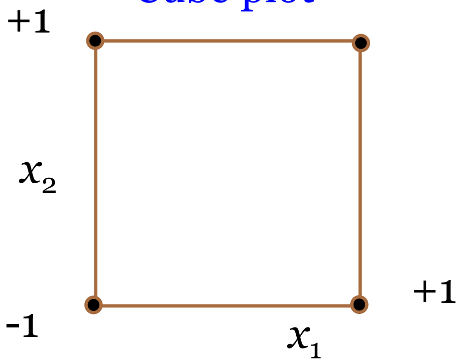
-1	to	+1
-1	to	+1

2^2 factorial design (Cont.)

- Design matrix becomes

x_1	x_2
-1	-1
+1	-1
-1	+1
+1	+1

“Cube plot”



- Fit model:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_{12}\boxed{x_1 x_2}$$

Interaction term

In matrix-vector notation,

$$\mathbf{y} = \mathbf{X}\mathbf{a}$$

$$\mathbf{y} = \begin{bmatrix} 60 \\ 72 \\ 54 \\ 68 \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_{12} \end{bmatrix} \qquad \mathbf{X} = \begin{array}{c} \begin{matrix} 1 & x_1 & x_2 & x_1x_2 \end{matrix} \\ \left[\boxed{\begin{matrix} & & & \\ & & & \\ & & & \\ & & & \end{matrix}}\right] \end{array}$$

2^2 factorial design (Cont.)

- Regression coefficients (usually from S/W)

$$\mathbf{a} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

Columns of $\mathbf{X}$: orthogonal (i.e., $\mathbf{x}_i \cdot \mathbf{x}_j = \mathbf{x}_i^T\mathbf{x}_j = 0$)

$$\rightarrow \sum x_0x_1 = \sum x_0x_2 = \sum x_0(x_1x_2) = \sum x_1x_2 \sum x_1(x_1x_2) = \sum x_2(x_1x_2) = 0$$

$$\mathbf{X}^T\mathbf{X} = \left[\quad\right] = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \qquad \mathbf{X}^T\mathbf{y} = \left[\quad\right]$$

$$\mathbf{a} = \begin{bmatrix} 1/\sum x_0^2 & 0 & 0 & 0 \\ 0 & 1/\sum x_1^2 & 0 & 0 \\ 0 & 0 & 1/\sum x_2^2 & 0 \\ 0 & 0 & 0 & 1/\sum (x_1x_2)^2 \end{bmatrix} \begin{bmatrix} \sum x_0y \\ \sum x_1y \\ \sum x_2y \\ \sum (x_1x_2)y \end{bmatrix}$$

2^2 factorial design (Cont.)

- i.e., $$a_i = \frac{\sum x_i y_i}{\sum x_i^2}$$

  Each a_i can be calculated independently.

  e.g., $a_0 = \dfrac{y_1 + y_2 + y_3 + y_4}{4}$

  a_i = effect of changing variable x_i from 0 to +1.

- Confidence interval of a_i

  $$\mathrm{var}(\mathbf{a}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\sigma^2 \Rightarrow \mathrm{var}(a_i) = \frac{\sigma^2}{\sum x_i^2}$$

  a_i are uncorrelated due to orthogonality of design

  95% C.I $\quad a_i \pm z_{0.025}\sqrt{\sigma^2 / \sum x_i^2}$

  95% C.I $\quad a_i \pm t_{\nu,0.025}\sqrt{s^2 / \sum x_i^2}$ (when σ unknown)

Estimating s^2 : (1) from historical database

(2) replicates (corner points or center points)

2^3 factorial design

2^3 → 3 variables

→ 2 levels

Qualitative variable

- Three variables: T, C, and catalyst type (A and B)
  - Denote: x_3 = -1 for catalyst A

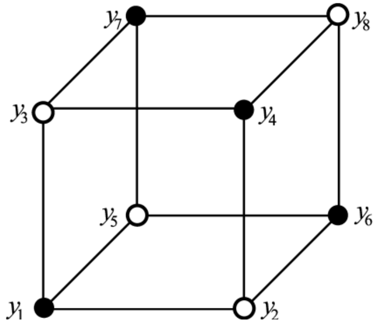
    = +1 for catalyst B
  - 2^3 factorial (= 8 runs): all combination of the 2 levels of the 3 variables.

Design

Matrix, **X**

x_0	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
+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

"Cube plot"



2^3 factorial design (cont.)

- Analysis by least squares
  - Fit model:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_{12} x_1 x_2 + a_{13} x_1 x_3 + a_{23} x_2 x_3 + a_{123} x_1 x_2 x_3$$

In matrix-vector notation,

$\mathbf{y} = \mathbf{X}\mathbf{a}$

  - Again, by least squares

$$\mathbf{a} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y} \Rightarrow a_i = \frac{\sum x_i y}{\sum x_i^2}$$

C.I of a_i $\quad \mathrm{var}(\mathbf{a}) = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \sigma^2 \Rightarrow \mathrm{var}(a_i) = \frac{\sigma^2}{\sum x_i^2}$

95% C.I $\quad a_i \pm z_{0.025} \sqrt{\sigma^2 / \sum x_i^2}$

95% C.I $\quad a_i \pm t_{\nu,0.025} \sqrt{s^2 / \sum x_i^2}$ (when σ unknown)

2^k factorial design

- Desirable features of factorial designs
  - Othogonal → easy calculations

    → uncorrelated estimates a_i
  - Good variation in all variables
  - Efficient use of all data points
  - The only way to discover interactions between variables
  - Allows experiments to be performed in blocks
  - Allows designs of increasing order to be build up sequentially

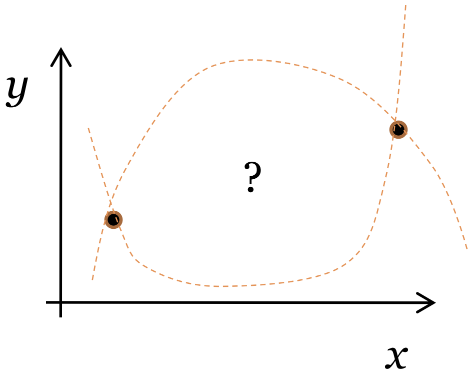
Design for 2nd order models

- If 1st order + interaction model exhibits "Lack of fit"

  → Include $x_1^2, x_2^2, \cdots$ terms

  But we need more than 2 level designs.

  → Central composite design or 3 level factorials



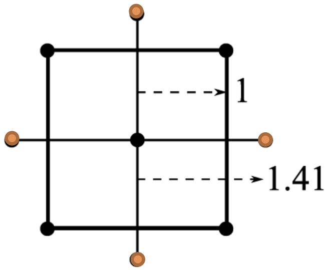
- Central composite design (k = 2)

  (1) Start with 2^k design with center points

  (2) Add vertices of star (for k=2, $\alpha = \sqrt{2}$)

  (3) Run experiments & analysis

"Cube plot"



	x_1	x_2
(1)	-1	-1
	+1	-1
	-1	+1
	+1	+1
	0	0
(2)	$-\alpha$	0
	$+\alpha$	0
	0	$-\alpha$
	0	$+\alpha$

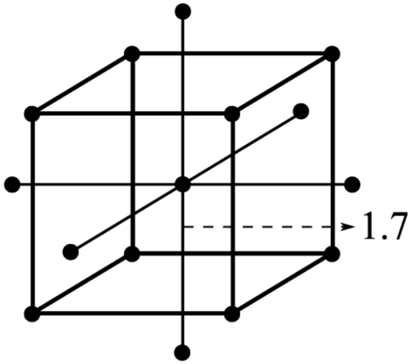
9 runs
For central composite design (k = 2)

Design for 2nd order models (cont.)

- Values of α

k	design	α
2	2^2	$\sqrt{2}$
3	2^3	$\sqrt{3}$
4	2^4	$\sqrt{4}$

Cube plot for 3 variables (factors)

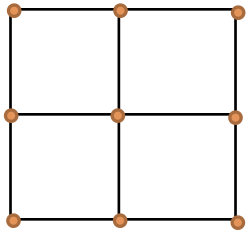


15 runs
For central
composite
design (k = 3)

- 3 level factorial

3^2 2 variables at all combinations of 3 levels

3^3 **27** runs for 3 variables



※Full quadratic model (assume 123 interaction is negligible.)

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2$$

Allows for approximation of many response.

Design for 2^{nd} order models (cont.)

※A t-statistic for curvature

Average y of corner points

Average y of center points

$$t_{curvature} = \frac{\bar{y}_F - \bar{y}_C}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_F} + \frac{1}{n_C}\right)}}$$

Pure error calculated
from center points

# of center points

# of corner points

Minitab uses ANOVA for testing curvature when center point replicates exist.