Experimental Design
Note 7
2^K Factorial Design

Keunbaik Lee

Sungkyunkwan University

2^k *FactorialDesign*

- Involving k factors to detect the important factors in the process with a minimum of experimental units.
- Each factor has two levels (often labeled $+$ and $-$)
- Factor screening experiment (preliminary study) often used to at the early stage of experimentation to detect potential candidate factors for more detailed investigation.
- Identify important factors and their interactions
- Interaction (of any order) has **ONE** degree of freedom
- Factors need not be on numeric scale
- Ordinary regression model can be employed

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

where β_1, β_2, and β_{12} are related to main effects, interaction effects defined later.

Chemical Processes Example

Factor		Treatment	Replicate			
A	B	Combination	I	II	III	Total
-	-	A low, B low	28	25	27	80
+	-	A high, B low	36	32	32	100
-	+	A low, B high	18	19	23	60
+	+	A high, B high	31	30	29	90

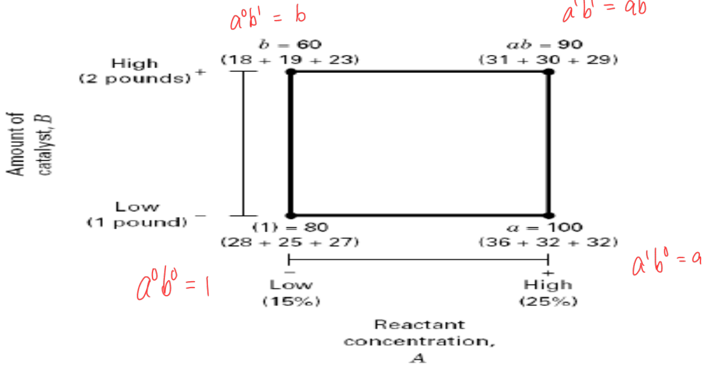
A=reactant concentration, B=catalyst amount, y=recovery

		B	
		−	+
A	−	1,2,3	7,8,9
	+	4,5,6	10,11,12

Analysis Procedure for a Factorial Design

- Estimate factor effects
- Formulate model
  - With replication, use full model
  - With an unreplicated design, use normal probability plots
- Statistical testing (ANOVA)
- Refine the model
- Analyze residuals (graphical)
- Interpret results

Figure 6.1 Treatment combinations in the 2^2 design



- "-" and "+" denote the low and high levels of a factor, respectively.

Estimation of Factor Effects I

$$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{ab+a}{2n} - \frac{b+(1)}{2n} = \frac{1}{2n}\left[ab+a-b-(1)\right]$$

$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{ab+b}{2n} - \frac{a+(1)}{2n} = \frac{1}{2n}\left[ab+b-a-(1)\right]$$

$$AB = \frac{ab+(1)}{2n} - \frac{a+b}{2n} = \frac{1}{2n}\left[ab+(1)-a-b\right]$$

The effect estimates are:
$A = 8.33$, $B = -5.00$, $AB = 1.67$
The quantities in brackets are **contrasts** in the treatment combinations.

Estimation of Factor Effects II

Effects and Contrasts

factor				effect (contrast)			
A	B	total	mean	1	A	B	AB
-	-	80	80/3	1	-1	-1	1
+	-	100	100/3	1	1	-1	-1
-	+	60	60/3	1	-1	1	-1
+	+	90	90/3	1	1	1	1

- There is a one-to-one correspondence between effects and contrasts, and contrasts can be directly used to estimate the effects.

Estimation of Factor Effects III

- For a effect corresponding to contrast $c = (c_1, c_2, \cdots)$ in 2^2 design

$$\text{effect} = \frac{1}{2} \sum_i c_i \bar{y}_i$$

where i is an index for treatments and the summation is over all treatments.
For example,

$$\begin{aligned}
\text{effect}(A) &= \frac{1}{2n} \{(y_{++\cdot} + y_{+-\cdot}) - (y_{-+\cdot} + y_{--\cdot})\} \\
&= \frac{1}{2} \{\bar{y}_{++} + \bar{y}_{+-} - \bar{y}_{-+} - \bar{y}_{--}\}, \\
\text{effect}(B) &= \frac{1}{2} \{\bar{y}_{++} + \bar{y}_{-+} - \bar{y}_{+-} - \bar{y}_{--}\}, \\
\text{effect}(AB) &= \frac{1}{2} \{\bar{y}_{++} + \bar{y}_{--} - \bar{y}_{+-} - \bar{y}_{-+}\},
\end{aligned}$$

Estimation of Factor Effects IV

Sum of Squares due to Effect

- Because effects are defined using contrasts, their sum of squares can also be calculated through contrasts.
- Recall for contrast $c = (c_1, c_2, \cdots)$, its sum of squares is

$$SS_{Contrast} = \frac{(\sum_i c_i \bar{y}_i)^2}{\sum_i c_i^2 / n}$$

So

$$SS_A = \frac{(-\bar{y}(A_-B_-) + \bar{y}(A_+B_-) - \bar{y}(A_-B_+) + \bar{y}(A_+B_+))^2}{4/n} = 208.33$$

$$SS_B = \frac{(-\bar{y}(A_-B_-) - \bar{y}(A_+B_-) + \bar{y}(A_-B_+) + \bar{y}(A_+B_+))^2}{4/n} = 75.00$$

$$SS_{AB} = \frac{(\bar{y}(A_-B_-) - \bar{y}(A_+B_-) - \bar{y}(A_-B_+) + \bar{y}(A_+B_+))^2}{4/n} = 8.33$$

Estimation of Factor Effects V

Sum of Squares and ANOVA

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
A	SS_A	1	MS_A	
B	SS_B	1	MS_B	
AB	SS_{AB}	1	MS_{AB}	
Error	SSE	$N-4$	MSE	
Total	SS_T	$N-1$		

where $SS_T = \sum_{i,j,k} y_{ijk}^2 - y_{...}/N$, $SSE = SS_T - SS_A - SS_B - SS_{AB}$.

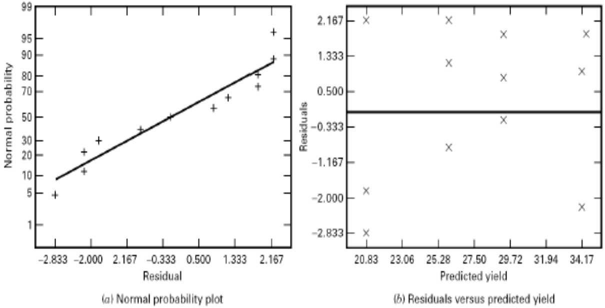
Revisit Chemical Process Example I

ANOVA Table

See Chemical-Process.SAS.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	*P*-value
A	208.33	1	208.33	53.15	0.0001
B	75.00	1	75.00	19.13	0.0024
AB	8.33	1	8.33	2.13	0.1826
Error	31.34	8	3.92		
Total	323.00	11			

Residuals and Diagnostic Checking



■ FIGURE 6.2 Residual plots for the chemical process experiment

Analyzing 2^2 Experiment using Regression Model I

Because every effect in 2^2 design, or its sum of squares, has one degree of freedom, it can be equivalently represented by a numerical variable, and regression analysis can be directly used to analyze the data. The original factors are not necessarily continuous.

Code the levels of factor A and B as follow

A	X_1	B	X_2
-	-1	-	-1
+	1	+	1

$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$

$\sum \tau_i = 0$, $\sum \beta_j = 0$

Fit regression model

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

Analyzing 2^2 Experiment using Regression Model II

The fitted model should be

$$\underbrace{y = \overset{\hat{\beta}_0}{\bar{y}_{\cdot\cdot}} + \overset{\hat{\beta}_1}{\frac{A}{2}}x_1 + \overset{\hat{\beta}_2}{\frac{B}{2}}x_2 + \overset{\hat{\beta}_3}{\frac{AB}{2}}x_1x_2}$$

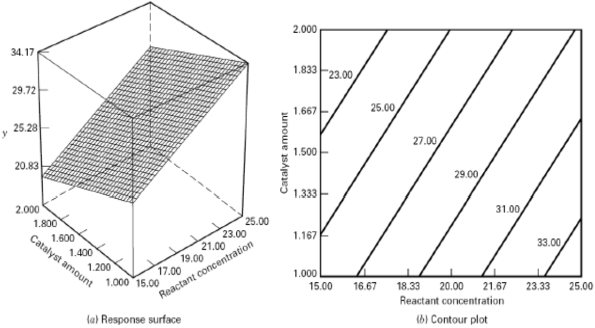
i.e., the estimated coefficients are half of the effects, respectively. See Chemical-Process.SAS.

$$y_{ijk} = \beta_0 + \beta_1 x_i + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{bmatrix}$$

$\ast$ $\hat{\beta} = (X^T X)^{-1} X^T y$

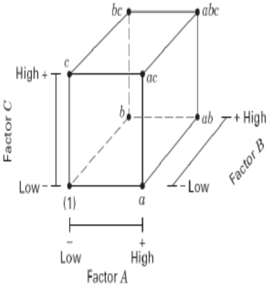
Response Surface



■ **FIGURE 6.3** Response surface plot and contour plot of yield from the chemical process experiment

The 2^3 Factorial Design I

FIGURE 6.4 The 2^3 factorial design



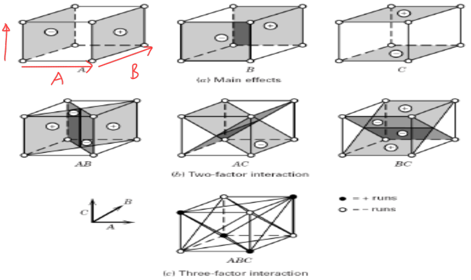
(a) Geometric view

Run	Factor A	Factor B	Factor C
1	−	−	−
2	+	−	−
3	−	+	−
4	+	+	−
5	−	−	+
6	+	−	+
7	−	+	+
8	+	+	+

(b) Design matrix

The 2^3 Factorial Design II

■ **FIGURE 6.5** Geometric presentation of contrasts corresponding to the main effects and interactions in the 2^3 design



$$A = \bar{y}_{A^+} - \bar{y}_{A^-}\,, \quad B = \bar{y}_{B^+} - \bar{y}_{B^-}\,,$$
$$C = \bar{y}_{C^+} - \bar{y}_{C^-}\,, \quad \text{etc...}$$

Analysis is done via computer.

The 2^3 Factorial Design III

Table of - and + Signs for the 2^3 Factorial Design

TABLE 6.3

Algebraic Signs for Calculating Effects in the 2^3 Design

Treatment Combination	Factorial Effect							
	I	A	B	AB	C	AC	BC	ABC
(1)	+	−	−	+	−	+	+	−
a	+	+	−	−	−	−	+	+
b	+	−	+	−	−	+	−	+
ab	+	+	+	+	−	−	−	−
c	+	−	−	+	+	−	−	+
ac	+	+	−	−	+	+	−	−
bc	+	−	+	−	+	−	+	−
abc	+	+	+	+	+	+	+	+

The 2^3 Factorial Design IV

Properties of the Table

- Except for column I, every column has an equal number of $+$ and - signs
- The sum of the product of signs in any two columns is zero
- Multiplying any column by I leaves that column unchanged (identity element)
- The product of any two columns yields a column in the table:

$$A \times B = AB, \quad AB \times BC = AB^2C = AC$$

- Orthogonal design
- Orthogonality is an important property shared by all factorial designs

Contrasts for Calculating Effects in 2^3 Design

				factorial effects							
A	B	C	treatment	I	A	B	AB	C	AC	BC	ABC
$-$	$-$	$-$	(1)	1	-1	-1	1	-1	1	1	-1
$+$	$-$	$-$	a	1	1	-1	-1	-1	-1	1	1
$-$	$+$	$-$	b	1	-1	1	-1	-1	1	-1	1
$+$	$+$	$-$	ab	1	1	1	1	-1	-1	-1	-1
$-$	$-$	$+$	c	1	-1	-1	1	1	-1	-1	1
$+$	$-$	$+$	ac	1	1	-1	-1	1	1	-1	-1
$-$	$+$	$+$	bc	1	-1	1	-1	1	-1	1	-1
$+$	$+$	$+$	abc	1	1	1	1	1	1	1	1

The 2^3 Factorial Design VI

$$\hat{\mu} = \bar{y}_{....} = \frac{1}{8n}\left(y_{+++\cdot} + y_{++-\cdot} + y_{+--\cdot} + y_{+-+\cdot} + y_{-++\cdot} + y_{-+-\cdot} + y_{--+\cdot} + y_{---\cdot}\right) = \frac{1}{8}\left(\bar{y}_{++++} + \cdots + \bar{y}_{----}\right)$$

2^k

Estimates:

$$\text{grand mean: } \frac{\sum_i \bar{y}_{i\cdot}}{2^3}$$

$$\text{effect: } \frac{\sum_i c_i \bar{y}_{i\cdot}}{2^{3-1}}$$

Contrast Sum of Squares:

$$SS_{effect} = \frac{(\sum_i c_i \bar{y}_{i\cdot})^2}{2^3/n} = 2n(\text{effect})^2$$

Variance of Estimate:

$$var(\text{effect}) = \frac{\sigma^2}{n2^{3-2}}$$

The 2^3 Factorial Design VII

t-test for effects (confidence interval approach):

$$\text{effect} \pm t_{\alpha/2, 2^k(n-1)} SE(\text{effect})$$

Model Coefficients-Full Model

TABLE 6.4
The Plasma Etch Experiment, Example 6.1

Run	Coded Factors A	B	C	Etch Rate Replicate 1	Replicate 2	Total	Factor Levels	Low (−1)	High (+1)
1	−1	−1	−1	550	604	(1) = 1154	A (Gap, cm)	0.80	1.20
2	1	−1	−1	669	650	a = 1319	B (C_2F_6 flow, SCCM)	125	200
3	−1	1	−1	633	601	b = 1234	C (Power, W)	275	325
4	1	1	−1	642	635	ab = 1277			
5	−1	−1	1	1037	1052	c = 2089			
6	1	−1	1	749	868	ac = 1617			
7	−1	1	1	1075	1063	bc = 2138			
8	1	1	1	729	860	abc = 1589			

A=Gap, B=Flow, C=Power, y=Etch Rate

Example II

Full Model

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

Model Coefficients for Full Model

Factor	Coefficient Estimated	DF	Standard Error	95% CI Low	95% CI High
Intercept	776.06	1	11.87	748.70	803.42
A-Gap	−50.81	1	11.87	−78.17	−23.45 ✓
B-Gas flow	3.69	1	11.87	−23.67	31.05
C-Power	153.06	1	11.87	125.70	180.42 ✓
AB	−12.44	1	11.87	−39.80	14.92
AC	−76.81	1	11.87	−104.17	−49.45 ✓
BC	−1.06	1	11.87	−28.42	26.30
ABC	2.81	1	11.87	−24.55	30.17

Example III

ANOVA Table

■ TABLE 6.6

Analysis of Variance for the Plasma Etching Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	*P*-Value
Gap (*A*)	41,310.5625	1	41,310.5625	18.34	0.0027
Gas flow (*B*)	217.5625	1	217.5625	0.10	0.7639
Power (*C*)	374,850.0625	1	374,850.0625	166.41	0.0001
AB	2475.0625	1	2475.0625	1.10	0.3252
AC	94,402.5625	1	94,402.5625	41.91	0.0002
BC	18.0625	1	18.0625	0.01	0.9308
ABC	126.5625	1	126.5625	0.06	0.8186
Error	18,020.5000	8	2252.5625		
Total	531,420.9375	15			

Example IV

Reduced Model

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_1 x_3 + \epsilon$$

Model Coefficients for Reduced Model

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High
Intercept	776.06	1	10.42	753.35	798.77
A-Gap	−50.81	1	10.42	−73.52	28.10
C-Power	153.06	1	10.42	130.35	175.77
AC	−76.81	1	10.42	−99.52	−54.10

Example V

Model Summary Statistics

- R^2 and adjusted R^2 for reduced model

$$R^2 = \frac{SS_{Model}}{SS_T} = \frac{5.106 \times 10^5}{5.314 \times 10^5} = 0.9608$$

$$R^2_{Adj} = 1 - \frac{SSE/df_E}{SS_T/df_T} = 1 - \frac{20857.75/12}{5.314 \times 10^5/15} = 0.9509$$

- Standard error of full model coefficients

$$se(\hat{\beta}) = \sqrt{\frac{MSE}{n2^k}} = \sqrt{\frac{2252.56}{2 \times 8}} = 11.87$$

Example VI

- Confidence interval on model coefficients

$$\hat{\beta} - t_{\alpha/2, df_E} SE(\hat{\beta}) \leq \beta \leq \hat{\beta} + t_{\alpha/2, df_E} SE(\hat{\beta})$$