Topic 13: 2^K factorial design (I)

Montgomery: chapter 6

Prof. Lingling An
University of Arizona

Outline

- Intro to two level factorial design
  - 2^2 design
  - 2^3 design

2^k Factorial Design

- Involving k factors
- Each factor has two levels (often labeled $+$ and $-$)
- Factor screening experiment (preliminary study)
- 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 Process Example

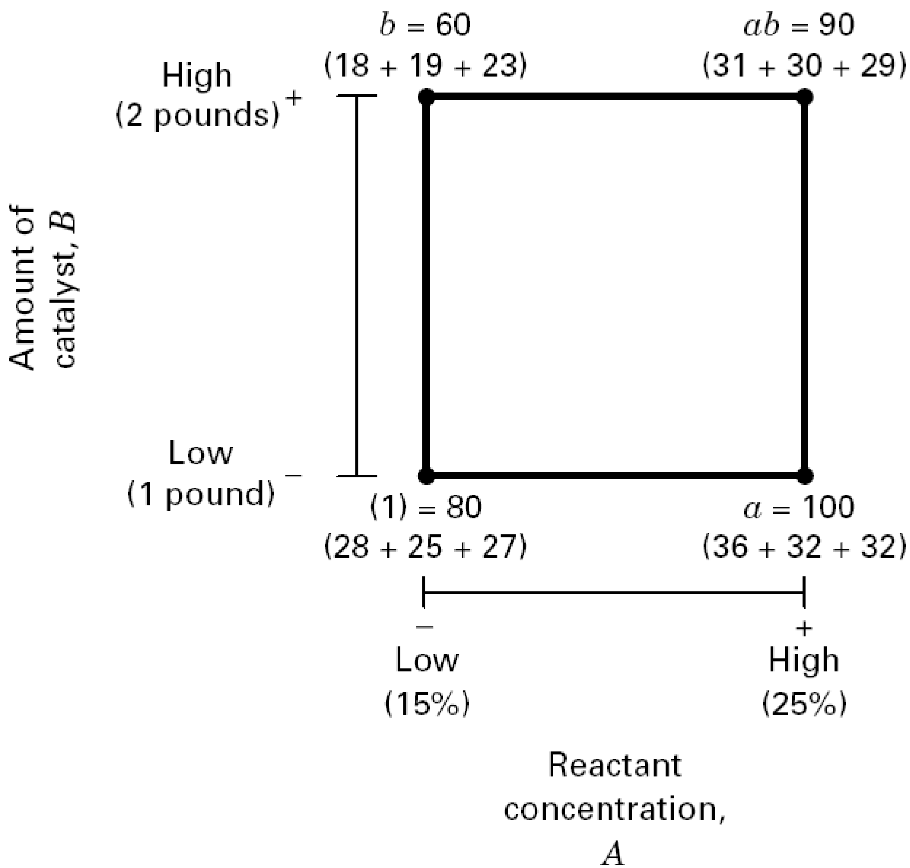
Factor		Treatment	Replicate			
A	B	Combination	I	II	III	Total
−	−	A low, B low	28	25	27	80
+	−	A high, B low	36	32	32	100
−	+	A low, B high	18	19	23	60
+	+	A high, B high	31	30	29	90

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

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

The Simplest Case: The 2^2



■ **FIGURE 6.1** **Treatment combinations in the 2^2 design**

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

- Low and high are arbitrary terms
- Factors can be quantitative or qualitative, although their treatment in the final model will be different
- Four treatment combinations are represented by lowercase letters.

Estimation of Factor Effects

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

$$= (\bar{y}(A_+) - \bar{y}_{..}) - (\bar{y}(A_-) - \bar{y}_{..}) = \hat{\tau}_2 - \hat{\tau}_1$$

Main effect is defined in a different way than Chapter 5. But they are connected and equivalent.

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

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

See textbook, pg. 209-210 For **manual** calculations

The effect estimates are:

$A = 8.33$, $B = -5.00$, $AB = 1.67$

The quantities in brackets are **contrasts** in the treatment combinations.

Effects and Contrasts

factor				effect (contrast)			
A	B	total	mean	I	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.
- For a effect corresponding to contrast $c = (c_1, c_2, \ldots)$ 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.

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 $c = (c_1, c_2, \ldots)$, its sum of squares is

$$\mathrm{SS}_{\mathrm{Contrast}} = \frac{\left(\sum c_i \bar{y}_i\right)^2}{\sum c_i^2 / n}$$

So

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

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

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

Sum of Squares and ANOVA

- Total sum of squares: $\mathrm{SS}_T = \sum_{i,j,k} y_{ijk}^2 - \frac{y_{\cdots}^2}{N}$
- Error sum of squares: $\mathrm{SS}_E = \mathrm{SS}_T - \mathrm{SS}_A - \mathrm{SS}_B - \mathrm{SS}_{AB}$
- ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A	SS_A	1	MS_A	
B	SS_B	1	MS_B	
AB	SS_{AB}	1	MS_{AB}	
Error	SS_E	$N-4$	MS_E	
Total	SS_T	$N-1$		

SAS file and output

```
option nocenter;
data one;
input A B resp;
datalines;
-1 -1 28
-1 -1 25
-1 -1 27
 1 -1 36
 1 -1 32
 1 -1 32
-1  1 18
-1  1 19
-1  1 23
 1  1 31
 1  1 30
 1  1 29
;
```

```
proc glm;
class A B;
model resp=A|B;
run;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	291.6666667	97.2222222	24.82	0.0002
Error	8	31.3333333	3.9166667		
Cor Total	11	323.0000000			
A	1	208.3333333	208.3333333	53.19	<.0001
B	1	75.0000000	75.0000000	19.15	0.0024
A*B	1	8.3333333	8.3333333	2.13	0.1828

The *F*-test for the "model" source is testing the significance of the overall model; that is, is either *A*, *B*, or *AB* or some combination of these effects important?

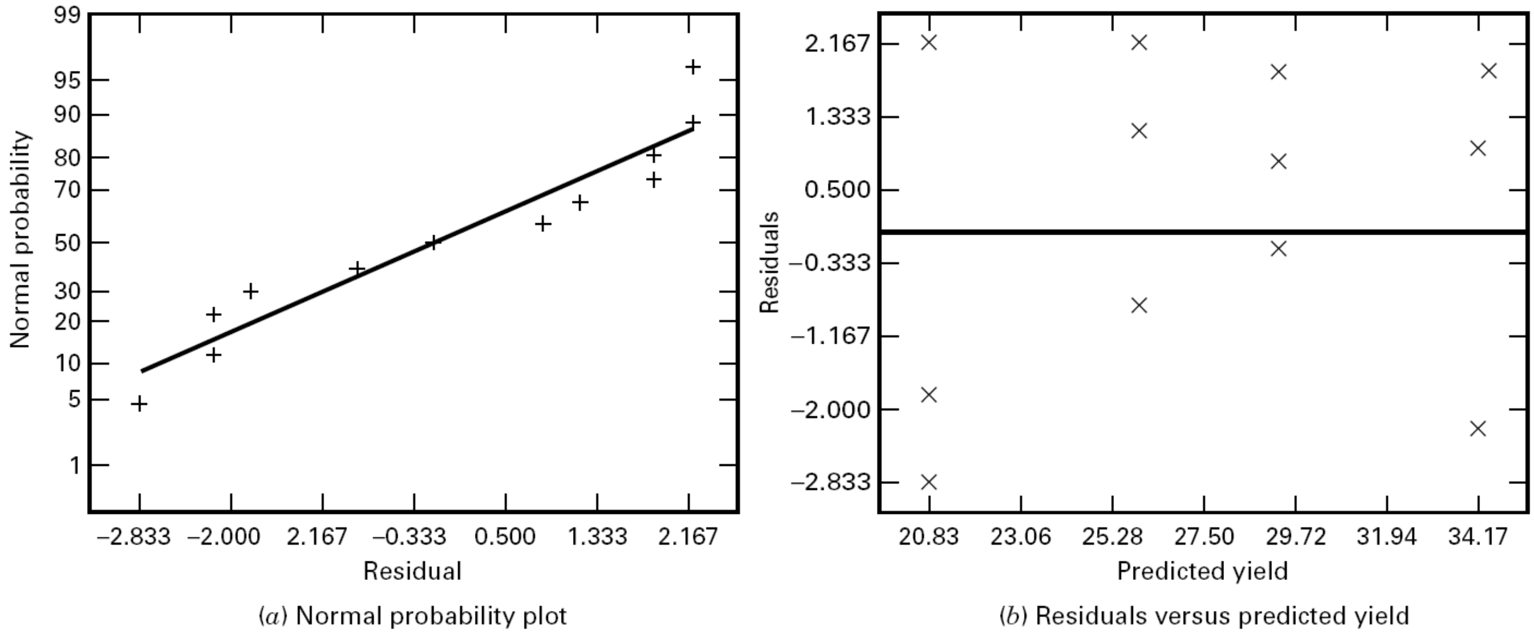
Statistical Testing - ANOVA

■ **TABLE 6.1**

Analysis of Variance for the Experiment in Figure 6.1

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_q	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

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 follows

A	x1	B	x2
−	−1	−	−1
+	1	+	1

Fit regression model

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

The fitted model should be

$$y = \bar{y}_{..} + \frac{A}{2} x_1 + \frac{B}{2} x_2 + \frac{AB}{2} x_1 x_2$$

i.e. the estimated coefficients are half of the effects, respectively.

SAS Code and Output

```
option nocenter;
data one;
input x1 x2 resp;
x1x2=x1*x2;
datalines;
-1 -1 28
-1 -1 25
-1 -1 27
........
 1  1 31
 1  1 30
 1  1 29
;
proc reg;
model resp=x1 x2 x1x2;
run
```

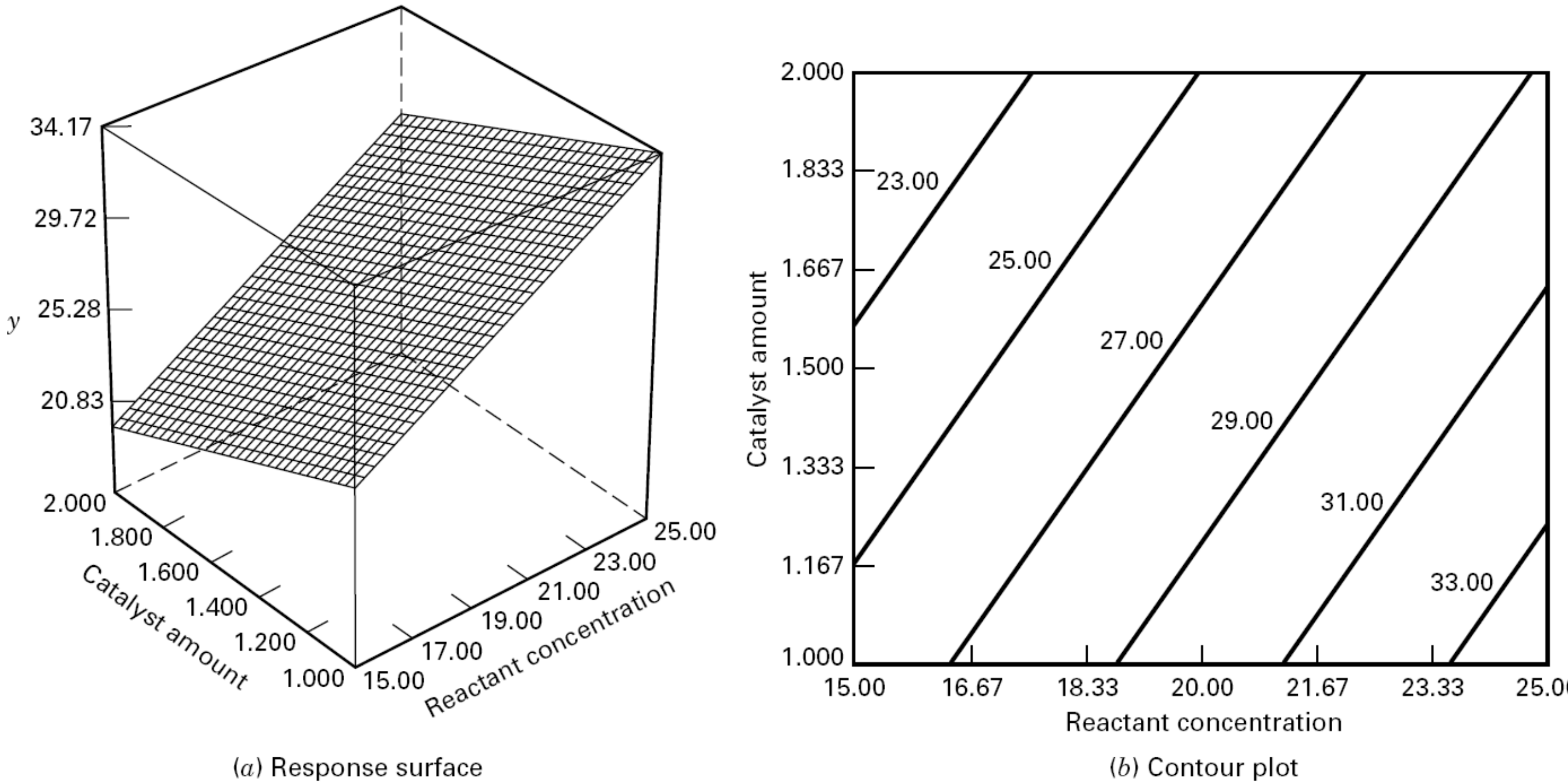
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	291.66667	97.22222	24.82	0.0002
Error	8	31.33333	3.91667		
Corrected Total	11	323.00000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	27.50000	0.57130	48.14	<.0001
x1	1	4.16667	0.57130	7.29	<.0001
x2	1	-2.50000	0.57130	-4.38	0.0024
x1x2	1	0.83333	0.57130	1.46	0.1828

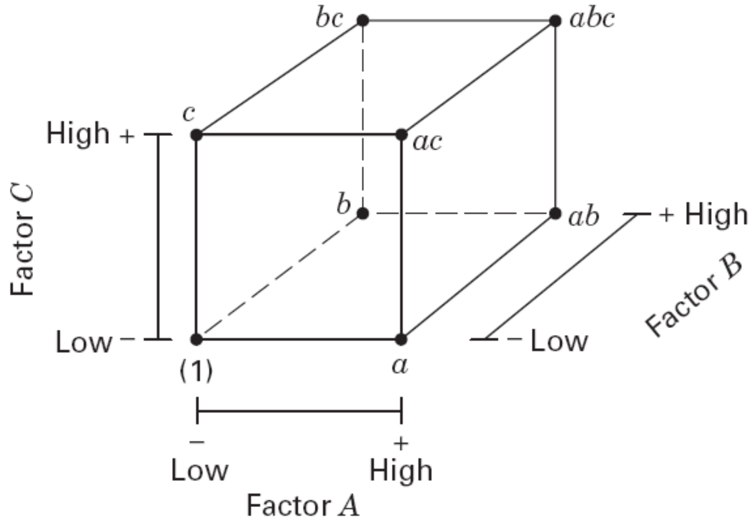
The Response Surface



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

The 2^3 Factorial Design

■ FIGURE 6.4
The 2^3 factorial design



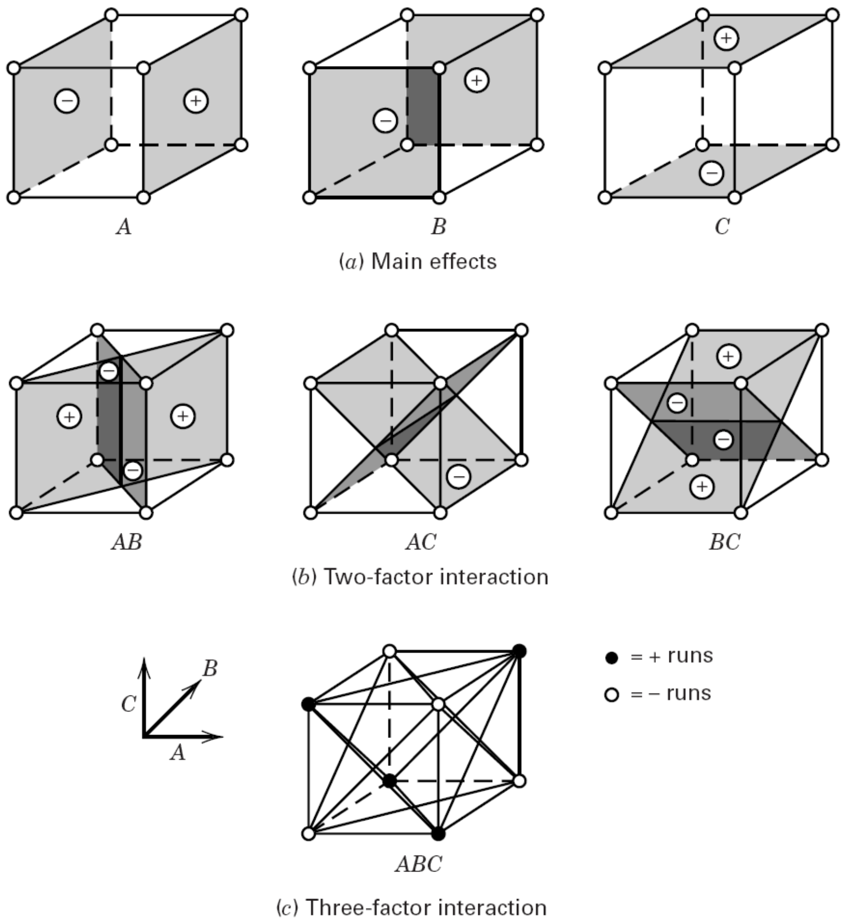
(a) Geometric view

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

(b) Design matrix

Effects in The 2^3 Factorial Design

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



$$A = \bar{y}_{A^+} - \bar{y}_{A^-}$$

$$B = \bar{y}_{B^+} - \bar{y}_{B^-}$$

$$C = \bar{y}_{C^+} - \bar{y}_{C^-}$$

etc, etc, ...

Analysis done via computer

An Example of a 2^3 Factorial Design

■ **TABLE 6.4**
The Plasma Etch Experiment, Example 6.1

	Coded Factors			Etch Rate			Factor Levels		
Run	***A***	***B***	***C***	**Replicate 1**	**Replicate 2**	**Total**		**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

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	+	+	+	+	+	+	+	+

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:

$$A \times B = AB$$

$$AB \times BC = AB^2C = AC$$

- **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

Design matrix

Estimates:

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

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

Contrast Sum of Squares:

$$\text{SS}_{\text{effect}} = \frac{(\sum c_i \bar{y}_{i.})^2}{2^3/n} = 2n(\text{effect})^2$$

Variance of Estimate

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

t-test for effects (confidence interval approach)

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

Estimation of Factor Effects

■ **TABLE 6.5**

Effect Estimate Summary for Example 6.1

Factor	Effect Estimate	Sum of Squares	Percent Contribution
A	−101.625	41,310.5625	7.7736
B	7.375	217.5625	0.0409
C	306.125	374,850.0625	70.5373
AB	−24.875	2475.0625	0.4657
AC	−153.625	94,402.5625	17.7642
BC	−2.125	18.0625	0.0034
ABC	5.625	126.5625	0.0238

ANOVA Summary – Full Model

■ **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			

Model Coefficients – 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

Refine Model – Remove Nonsignificant Factors

■ **TABLE 6.7** (*Continued*)

Response: Etch rate
ANOVA for Selected Factorial Model
Analysis of variance table [Partial sum of squares]

Source	Sum of Squares	DF	Mean Square	*F* Value	Prob > *F*
Model	5.106E+005	3	1.702E+005	97.91	<0.0001
A	*41310.56*	*1*	*41310.56*	*23.77*	*0.0004*
C	*3.749E+005*	*1*	*3.749E+005*	*215.66*	*<0.0001*
AC	*94402.56*	*1*	*94402.56*	*54.31*	*<0.0001*
Residual	20857.75	12	1738.15		
Lack of Fit	*2837.25*	*4*	*709.31*	*0.31*	*0.8604*
Pure Error	*18020.50*	*8*	*2252.56*		
Cor Total	5.314E+005	15			

Std. Dev.	41.69	R-Squared	0.9608
Mean	776.06	Adj R-Squared	0.9509
C.V.	5.37	Pred R-Squared	0.9302
PRESS	37080.44	Adeq Precision	22.055

Factor	Coefficient Estimate	DF	Standard Error	95% CI Low	95% CI High	VIF
Intercept	776.06	1	10.42	753.35	798.77	
A-Gap	−50.81	1	10.42	−73.52	−28.10	1.00
C-Power	153.06	1	10.42	130.35	175.77	1.00
AC	−76.81	1	10.42	−99.52	−54.10	1.00

Final Equation in Terms of Coded Factors:

Etch rate	=
+776.06	
−50.81	* *A*
+153.06	* *C*
−76.81	* *A* * *C*

Final Equation in Terms of Actual Factors:

Etch rate	=
−5415.37500	
+4354.68750	* Gap
+21.48500	* Power
−15.36250	* Gap * Power

Model Coefficients – 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

Model Summary Statistics for Reduced Model

- R^2 and adjusted R^2

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

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

- R^2 for prediction (based on PRESS)

$$R^2_{\text{Pred}} = 1 - \frac{PRESS}{SS_T} = 1 - \frac{37080.44}{5.314\times 10^5} = 0.9302$$

- PRESS: a statistic of measuring how well the model will predict new data
- A model with a small value of PRESS indicates that the model is likely to be a good predictor.

Model Summary Statistics

- **Standard error** of model coefficients (full model)

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

- **Confidence interval** on model coefficients

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

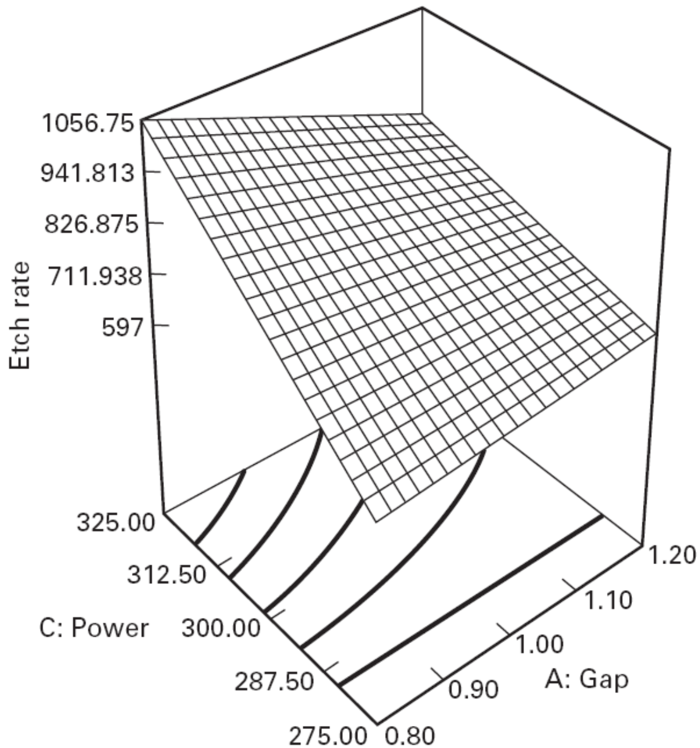
The Regression Model

Final Equation in Terms of Coded Factors:

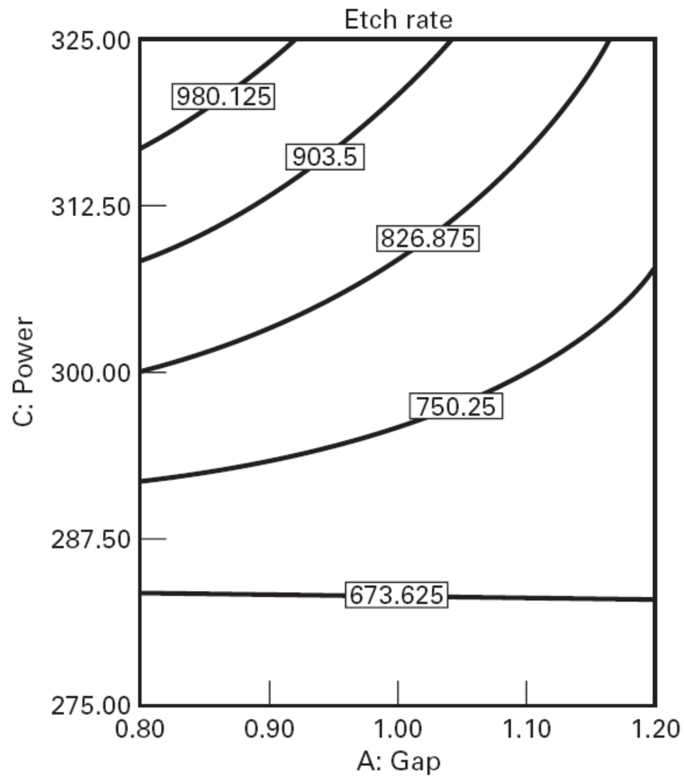
Etch rate	=
+776.06	
−50.81	* A
+153.06	* C
−76.81	* A * C

Final Equation in Terms of Actual Factors:

Etch rate	=
−5415.37500	
+4354.68750	* Gap
+21.48500	* Power
−15.36250	* Gap * Power



(*a*) The response surface



(*b*) The contour plot

■ **FIGURE 6.7** **Response surface and contour plot of etch rate for Example 6.1**

Last slide

- Read Sections: 6.1-6.3