The 2^k Factorial Design

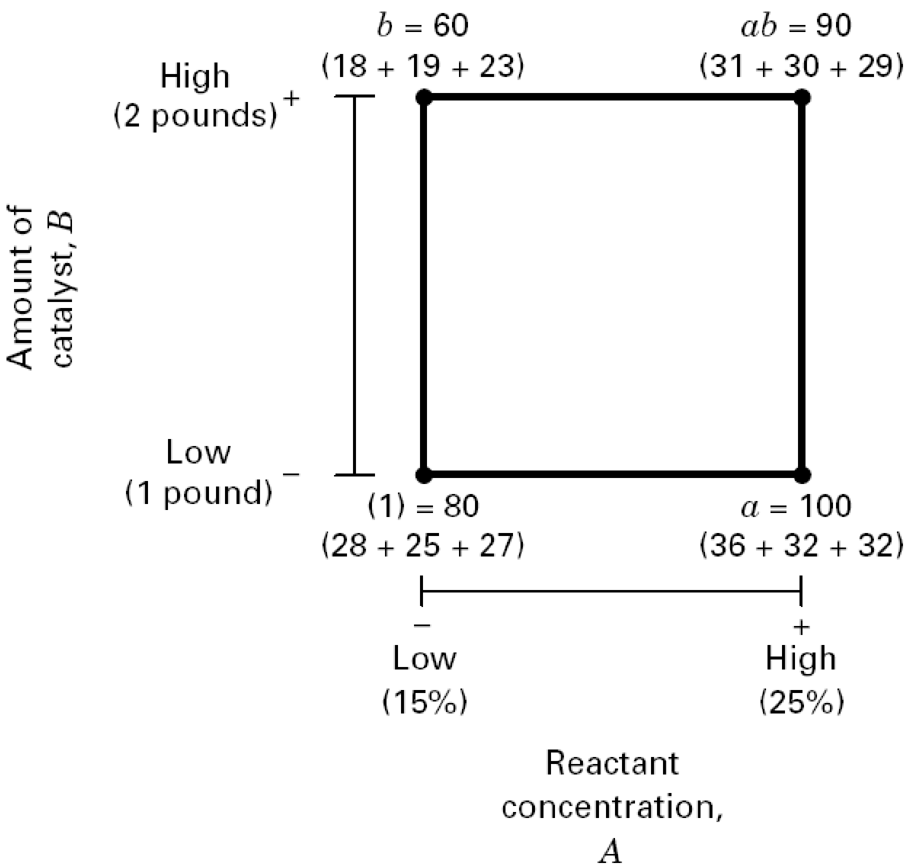
- In a factorial design, all possible combinations of the levels of the factors are investigated. Study how each factor affects the response
- **Special case** of the general factorial design; *k* factors, all at two levels
- The two levels are usually called **low** and **high** (they could be either quantitative or qualitative)
- Very widely used in industrial experimentation
- ANOVA for 2^k factorial designs
- Linear regression for 2^k factorial designs

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

The Simplest Case: The 2^2 Design



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

ANOVA

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1{,}2 \\ j = 1{,}2 \\ k = 1{,}2, \ldots, n \end{cases}$$

$$\sum \hat{\tau}_i = 0 \text{ , } \sum \hat{\beta}_j = 0 \text{ , } \sum \left(\widehat{\tau\beta}\right)_{ij} = 0 \quad \sum \left(\widehat{\tau\beta}\right)_{ij} = 0 \quad i = 1, 2, \quad j = 1, 2,$$

$$\begin{aligned} \widehat{\mu} &= \bar{y}_{...} \\ \hat{\tau}_i &= \bar{y}_{i..} - \bar{y}_{...} \quad i = 1, 2, \\ \hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...} \quad j = 1, 2, \\ \left(\widehat{\tau\beta}\right)_{ij} &= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \quad \begin{cases} i = 1, 2, \\ j = 1, 2, \end{cases} \end{aligned}$$

Contrast

In general, a **contrast** is a linear combination of parameters of the form

$$\Gamma = \sum_{i=1}^{2^k} c_i \mu_i \qquad \text{with} \qquad \sum_{i=1}^{2^k} c_i = 0$$

$$SS_\Gamma = \frac{\left(\sum_{i=1}^{2^k} c_i \bar{y}_{i\cdot}\right)^2}{\frac{1}{n}\sum_{i=1}^{2^k} c_i^2} = \frac{\left(\sum_{i=1}^{2^k} c_i y_{i\cdot}\right)^2}{n\sum_{i=1}^{2^k} c_i^2}$$

Contrast

TABLE 6.2

Algebraic Signs for Calculating Effects in the 2^2 Design

	Factorial Effect			
Treatment Combination	I	A	B	AB
(1)	+	−	−	+
a	+	+	−	−
b	+	−	+	−
ab	+	+	+	+

ANOVA

$$SS_A = \frac{[ab + a - b - (1)]^2}{4n} = \frac{(50)^2}{4(3)} = 208.33$$

$$SS_B = \frac{[ab + b - a - (1)]^2}{4n} = \frac{(-30)^2}{4(3)} = 75.00$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{4n} = \frac{(10)^2}{4(3)} = 8.33$$

ANOVA

TABLE 6.1

Analysis of Variance for the Experiment in Figure 6.1

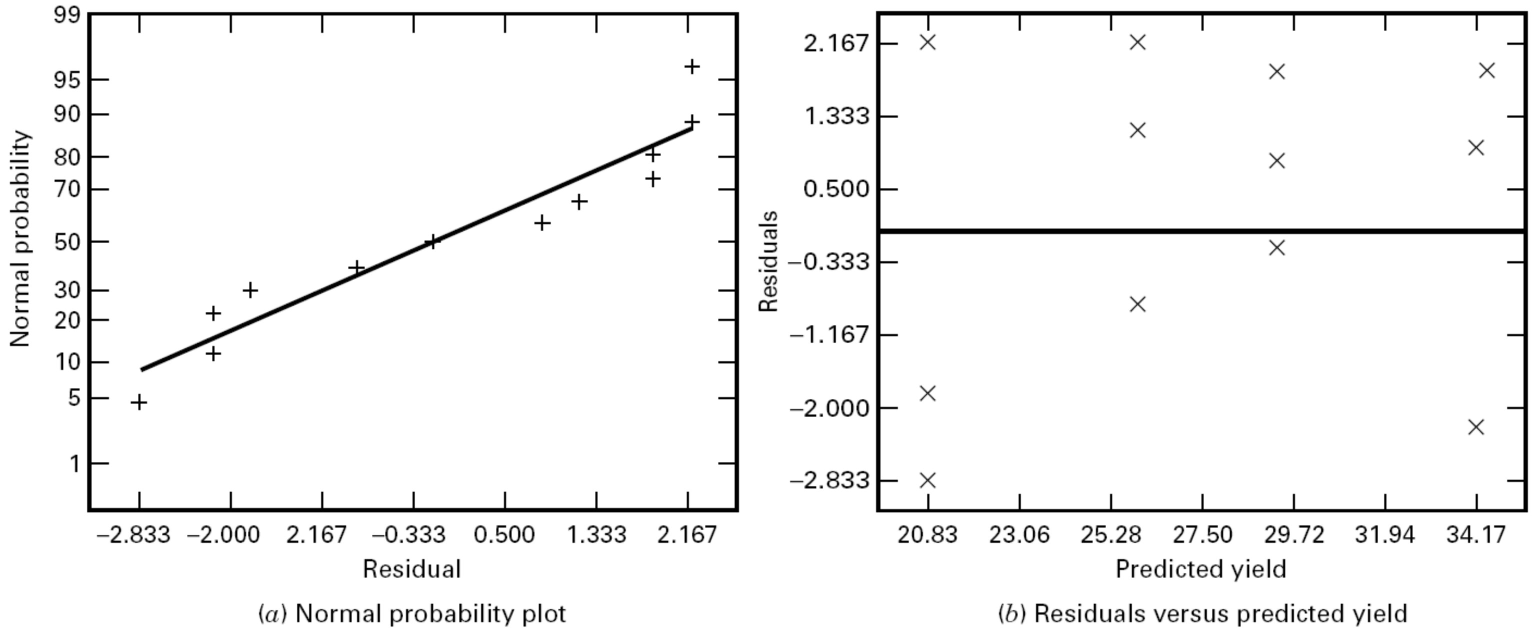
Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_q	*P*-Value
A	208.33	1	208.33	53.15	0.0001
B	75.00	1	75.00	19.13	0.0024
AB	8.33	1	8.33	2.13	0.1826
Error	31.34	8	3.92		
Total	323.00	11			

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?

df breakdown:

$$abn - 1 = a - 1 + b - 1 + (a-1)(b-1) + ab(n-1)$$

Residuals and Diagnostic Checking



■ **FIGURE 6.2** **Residual plots for the chemical process experiment**

Linear Regression with Coded Variables

Linear regression and ANOVA are equivalent for 2^k factorial designs

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

Coded variables: $x_1 = -1$ for A^-, $x_1 = 1$ for A^+

$$x_1 = \frac{A - (A^- + A^+)/2}{(A^+ - A^-)/2}$$

$$x_2 = \frac{B - (B^- + B^+)/2}{(B^+ - B^-)/2}$$

The 2^k design and design optimality

The model parameter estimates in a 2^k design (and the effect estimates) are least squares estimates. For example, for a 2^2 design the model is

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

$$\begin{aligned}
(1) &= \beta_0 + \beta_1(-1) + \beta_2(-1) + \beta_{12}(-1)(-1) + \varepsilon_1 \\
a &= \beta_0 + \beta_1(1) + \beta_2(-1) + \beta_{12}(1)(-1) + \varepsilon_2 \\
b &= \beta_0 + \beta_1(-1) + \beta_2(1) + \beta_{12}(-1)(1) + \varepsilon_3 \\
ab &= \beta_0 + \beta_1(1) + \beta_2(1) + \beta_{12}(1)(1) + \varepsilon_4
\end{aligned}$$

← The four observations from a 2^2 design

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \mathbf{y} = \begin{bmatrix} (1) \\ a \\ b \\ ab \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

The least squares estimate of β is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$= \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}^{-1} \begin{bmatrix} (1)+a+b+ab \\ a+ab-b-(1) \\ b+ab-a-(1) \\ (1)-a-b+ab \end{bmatrix}$$

The "usual" contrasts

The $\mathbf{X}'\mathbf{X}$ matrix is diagonal – consequences of an orthogonal design

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_{12} \end{bmatrix} = \frac{1}{4}\mathbf{I}_4 \begin{bmatrix} (1)+a+b+ab \\ a+ab-b-(1) \\ b+ab-a-(1) \\ (1)-a-b+ab \end{bmatrix} = \begin{bmatrix} \dfrac{(1)+a+b+ab}{4} \\ \dfrac{a+ab-b-(1)}{4} \\ \dfrac{b+ab-a-(1)}{4} \\ \dfrac{(1)-a-b+ab}{4} \end{bmatrix}$$

The regression coefficient estimates are exactly half of the 'usual" effect estimates if all treatments are examined once

The matrix $\mathbf{X}'\mathbf{X}$ has interesting and useful properties:

$$V(\hat{\beta}) = \sigma^2 \, (\mathbf{X}'\mathbf{X})^{-1}$$

$$V(\widehat{\beta_1}) = V(\widehat{\beta_2}) = V(\widehat{\beta_{12}}) = \frac{\sigma^2}{4}$$

Minimum possible value for a four-run design

$$|(\mathbf{X}'\mathbf{X})| = 256$$

Maximum possible value for a four-run design

Notice that these results depend on both the design that you have chosen and the model.

Factorial designs are optimal designs for a linear model w/o interactions: minimize the variance of estimators