Chapter 5
(AST301) Design and Analysis of Experiments II

Md Rasel Biswas

Lecture Outline

Section 1

## 5. Introduction to Factorial Designs

Subsection 1

5.1 Basic Definitions and Principles

Factorial design

- **Factorial designs** deal with the experiments that involve the study of two or more factors.

- In factorial design, all possible combinations of the levels of the factors are investigated in each complete replication.

    - E.g. if there are a levels of factor A and b levels of factor B, each replicate contains all ab **treatment combinations**.

- The **effect of a factor** is defined to be the change in response produced by a change in the level of the factor.

- This is frequently called a **main effect** because it refers to the primary factors of interest in the experiment.

Example 5.1

- For example, consider the simple experiment in Figure 5.1.

- This is a two-factor factorial experiment with both design factors at two levels. We have called these levels "low" and "high" and denoted them "−" and "+," respectively.
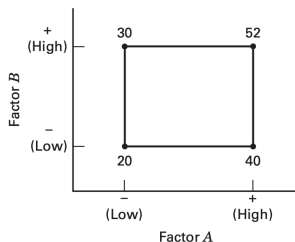
■ **FIGURE 5.1 A two-factor factorial experiment, with the response (y) shown at the corners**

Example 5.1

- The main effect of factor A in this two-level design can be thought of as the difference between the average response at the low level of A and the average response at the high level of A. Numerically, this is

$$A = \frac{40 + 52}{2} - \frac{20 + 30}{2} = 21$$

- That is, increasing factor A from the low level to the high level causes an **average response increase** of 21 units

- Similarly, the main effect of B is

$$B = \frac{30 + 52}{2} - \frac{20 + 40}{2} = 11$$

- If the factors appear at more than two levels, the above procedure must be modified because there are other ways to define the effect of a factor. This point is discussed more completely later.

Example 5.2

- In some experiments, we may find that the difference in response between the levels of one factor is not the same at all levels of the other factors.

- When this occurs, there is an **interaction** between the factors. For example, consider the two-factor factorial experiment shown in Figure 5.2
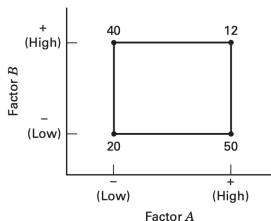
■ **FIGURE 5.2** A two-factor factorial experiment with interaction

Example 5.2

- At the low level of factor B (or $B-$), the A effect is

$$A = 50 - 20 = 30$$

- At the high level of factor B (or $B+$), the A effect is

$$A = 12 - 40 = -28$$

- Because the effect of A depends on the level chosen for factor B, we see that there is interaction between A and B.

- The magnitude of the interaction effect is the average *difference* in these two A effects, or $AB = (-28 - 30) \,/\, 2 = -29$.

- Clearly, the interaction is large in this experiment.

Example 5.2

- An interaction is the failure of one factor to produce the same effect on the response at different levels of another factor.

Illustrating interaction

- These ideas may be illustrated graphically. Figure 5.3 plots the response data in Figure 5.1 against factor A for both levels of factor B.
  - ▶ Note that the $B-$ and $B+$ lines are approximately parallel, indicating a lack of interaction between factors A and B.
- Similarly, Figure 5.4 plots the response data in Figure 5.2. Here we see that the $B-$ and $B+$ lines are not parallel.
  - ▶ This indicates an interaction between factors A and B.
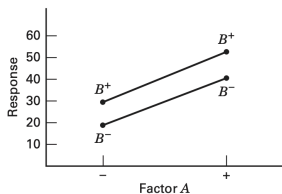
■ **FIGURE 5.3** **A factorial experiment without interaction**
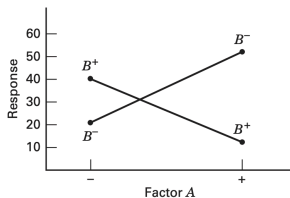
■ **FIGURE 5.4** **A factorial experiment with interaction**

Illustrating interaction

- There is another way to illustrate the concept of interaction. Suppose that both of our design factors are **quantitative** (such as temperature, pressure, time).

- Then a regression model representation of the two-factor factorial experiment could be written as

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

  ▶ where y is the response and the variables x1 and x2 are defined on a coded scale from -1 to $+1$ (the low and high levels of A and B).

Illustrating interaction

- The parameter estimates in this regression model turn out to be related to the effect estimates.

- For the experiment shown in Figure 5.1 we found the main effects of A and B to be A $= 21$ and $B = 11$.

- The least square estimates of β_1 and $\beta_2$are one-half the value of the corresponding main effect (more on this later)
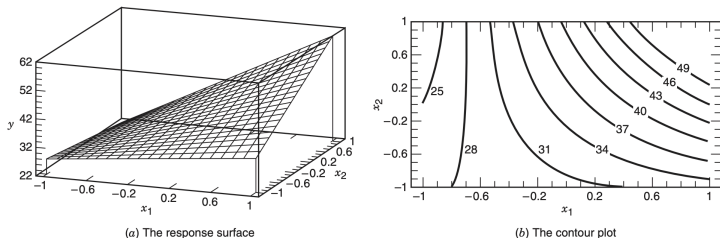
Illustrating interaction
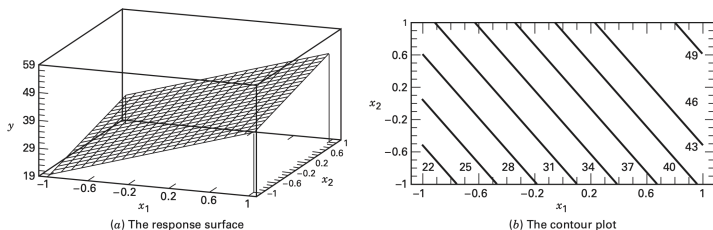
■ **FIGURE 5.5** Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2$

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 0.5x_1x_2 \equiv 35.5 + 10.5x_1 + 5.5x_2$$

Illustrating interaction

- Now suppose that the interaction contribution to this experiment was not negligible

- Figure 5.6 presents the response surface and contour plot for the model

$$\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$$

(a) The response surface

(b) The contour plot

FIGURE 5.6 Response surface and contour plot for the model $\hat{y} = 35.5 + 10.5x_1 + 5.5x_2 + 8x_1x_2$

- **Interaction is a form of curvature** in the underlying response surface model for the experiment

Subsection 2

5.2 The advantage of factorials

5.2 The advantage of factorials

- Two factors both at two levels: A^+, A^-, B^+, B^-
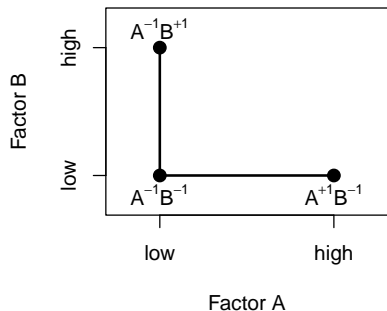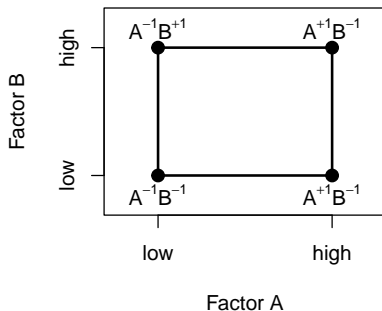- Two possible designs

Figure 1: A one-factor-at-a-time experiment

Figure 2: A two-factor factorial experiment

5.2 The advantage of factorials

One-factor-at-a-time design

- Because experimental error is present, it is desirable to take two observations, say, at each treatment combination and estimate the effects of the factors using average responses.

- Thus, a total of six observations are required.

- Effects of factors A and B can be obtained, but interaction cannot be calculated

Two-factor factorial design

- Using just four observations, two estimates of the A effect can be made. Similarly, two estimates of the B effect can be made.

- These two estimates of each main effect could be averaged to produce average main effects that are *just as precise* as those from the single-factor experiment

- Main effects of A and B and their interaction can be calculated

5.2 The advantage of factorials

- Factorial designs are more efficient than one-factor-at-a-time experiments.
  - For this example, the relative efficiency of the factorial design to the one-factor-at-a-time experiment is $(6/4) = 1.5$.
  - Generally, this relative efficiency will increase as the number of factors increases.

- A factorial design is necessary when interactions may be present to avoid misleading conclusions.

- Factorial designs allow the effects of a factor to be estimated at several levels of the other factors, yielding conclusions that are valid over a range of experimental conditions.

Subsection 3

5.3 Two-Factor factorial design

An Example

- The simplest types of factorial designs involve only two factors or sets of treatments.

- Factor A has a levels and factor B has b levels, so there will be in total ab treatment combinations and each treatment combination is replicated n times

Battery life experiment

An engineer is designing a battery for use in a device that will be subjected to some extreme variations in temperature. The only design parameter that he can select at this point is the plate material for the battery, and he has three possible choices. When the device is manufactured and is shipped to the field, the engineer has no control over the temperature extremes that the device will encounter, and he knows from experience that temperature will probably affect the effective battery life. However, temperature can be controlled in the product development laboratory for the purposes of a test.

An Example

- The engineer decides to test all three plate materials at three temperature levels - 15, 70, and 125°F - because these temperature levels are consistent with the product end-use environment.

- Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order.

- The experiment and the resulting observed battery design experiment are given below

An Example

■ TABLE 5.1
Life (in hours) Data for the Battery Design Example

Material Type	Temperature (°F)					
	15		**70**		**125**	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

$A \to$ Material type; $B \to$ Temperature (a **quantitative** variable)

Important questions:

1. What **effects** do material type and temperature have on the life of the battery?

2. Is there a choice of material that would give uniformly long life regardless of temperature?

General notations

- y_{ijk} denotes the observed response when factor A is at the ith level $(i = 1, 2, \dots, a)$ and factor B is at the jth level $(j = 1, 2, \dots, b)$ for the kth replicate $(k = 1, 2, \dots, n)$.

- The order in which the abn observations are taken is selected at random so that this design is a **completely randomized design**.

- Data layout

		B			
		1	2	$\cdots$	b
	1	$y_{111}, \dots, y_{11n}$	$y_{121}, \dots, y_{12n}$	$\cdots$	$y_{1b1}, \dots, y_{1bn}$
	2	$y_{211}, \dots, y_{21n}$	$y_{221}, \dots, y_{22n}$	$\cdots$	$y_{2b1}, \dots, y_{2bn}$
A	$\vdots$			$\ddots$	
	a	$y_{a11}, \dots, y_{a1n}$	$y_{a21}, \dots, y_{a2n}$	$\cdots$	$y_{ab1}, \dots, y_{abn}$

Modelling data

- The observations in a factorial experiment can be described by a model. There are several ways to write the model for a factorial experiment.

- The **means** model

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} \quad \left\{ \begin{array}{l} i = 1, ..., a \\ j = 1, ..., b \\ k = 1, ..., n \end{array} \right.$$

where μ_{ij} is the mean corresponding to the treatment combination i^{th} level of factor A and j^{th} level of factor B, and ϵ_{ijk} is the random error term.

- The **treatment** model (or **effects** model)

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \qquad (1)$$

where μ is the overall mean, τ_i is the effect of the ith level of factor A, β_j is the effect of the jth level of factor B, $(\tau\beta)_{ij}$ is the interaction between τ_i and β_j.

Modelling data

- We could also use a **regression model** as in Section 5.1 (particularly useful when one or more of the factors in the experiment are quantitative).

- Throughout most of this chapter we will use the effects model (Equation 1) with an illustration of the regression model in Section 5.5.

Modelling data

- In the two-factor factorial, both row and column factors (or treatments), A and B, are of equal interest.

- Specifically, we are interested in testing hypotheses about the equality of row treatment effects, say

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \text{at least one } \tau_i \neq 0$$

- and the equality of column treatment effects, say

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$$
$$H_1 : \text{at least one } \beta_j \neq 0$$

- We are also interested in determining whether row and column treatments interact:

$$H_0 : (\tau\beta)_{ij} = 0 \text{ for all } i, j$$
$$H_1 : \text{at least one } (\tau\beta)_{ij} \neq 0$$

Subsection 4

Statistical analysis of the fixed effects model

Notations

$$y_{i\cdot\cdot} = \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{i\cdot\cdot} = \frac{y_{i\cdot\cdot}}{bn} \quad i = 1, ..., a$$

$$y_{\cdot j\cdot} = \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{\cdot j\cdot} = \frac{y_{\cdot j\cdot}}{an} \quad j = 1, ..., b$$

$$y_{ij\cdot} = \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{ij\cdot} = \frac{y_{ij\cdot}}{n} \quad \begin{array}{l} i = 1, ..., a \\ j = 1, ..., b \end{array}$$

$$y_{\cdot\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \qquad \bar{y}_{\cdot\cdot\cdot} = \frac{y_{\cdot\cdot\cdot}}{abn}$$

Decomposing total variation

Total corrected sum of squares can be expressed as

$$SS_T = \sum_{i,j,k}(y_{ijk} - \bar{y}_{...})^2 = SS_A + SS_B + SS_{AB} + SS_E, \qquad (2)$$

where

$$SS_A = bn\sum_i(\bar{y}_{i..} - \bar{y}_{...})^2 \qquad SS_B = an\sum_j(\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$SS_{AB} = n\sum_{i,j}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$SS_E = \sum_{i,j,k}(y_{ijk} - \bar{y}_{ij.})^2$$

- Equation 2 is the fundamental ANOVA equation for the two-factor factorial

Decomposing total variation

The number of degrees of freedom associated with each sum of squares

Effect	Degrees of freedom
A	$a-1$
B	$b-1$
AB	$(a-1)(b-1)$
Error	$ab(n-1)$
Total	$abn-1$

- Each sum of squares divided by its degrees of freedom is a mean square, e.g.

$$MS_A = \frac{SS_A}{a-1}$$

Expected value of the mean squares

$$E(MS_A) = E\left(\frac{SS_A}{a-1}\right) = \sigma^2 + \frac{bn\sum\limits_{i=1}^{a}\tau_i^2}{a-1}$$

$$E(MS_B) = E\left(\frac{SS_B}{b-1}\right) = \sigma^2 + \frac{an\sum\limits_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MS_{AB}) = E\left(\frac{SS_{AB}}{(a-1)(b-1)}\right) = \sigma^2 + \frac{n\sum\limits_{i=1}^{a}\sum\limits_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E(MS_E) = E\left(\frac{SS_E}{ab(n-1)}\right) = \sigma^2$$

- Under the null hypotheses of no treatment effects and no interaction, the $MS_A, MS_B, MS_{AB}, MS_E$ all estimate σ^2. If there is significant treatment effect then corresponding mean squares will be larger than MS_E

Cochran's theorem

Cochran's theorem

Let Z_i be $\text{NID}(0,1)$ for $i = 1, 2, \ldots, \nu$ and

$$\sum_{i=1}^{\nu} Z_i^2 = Q_1 + Q_2 + \cdots + Q_s$$

where $s \leq v$, and Q_i has v_i degrees of freedom $(i = 1, 2, \ldots, s)$. Then $Q_1, Q_2, \ldots, Q_s$ are independent chi-square random variables with $v_1, v_2, \ldots, v_s$ degrees of freedom, respectively, if and only if

$$\nu = \nu_1 + \nu_2 + \cdots + \nu_s$$

Distributions of sum of squares

For the observed response y_{ijk}, $i = 1, \ldots, a$, $j = 1, \ldots, b$, and $k = 1, \ldots, n$, we can decompose the total sum of squares as

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$
$$(abn-1) \quad a-1 \quad b-1 \quad (a-1)(b-1) \quad ab(n-1)$$

If SS_A, SS_B, SS_{AB}, and SS_E are independent, then using Cochran's theorem we can write

$$SS_A \sim \chi^2_{(a-1)}, \ \ SS_B \sim \chi^2_{(b-1)}, \ \ SS_{AB} \sim \chi^2_{(a-1)(b-1)}, \ \ SS_E \sim \chi^2_{(ab(n-1))}$$

since

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

Distributions of sum of squares

- Since $SS_A \sim \chi^2_{(a-1)}$ and $SS_E \sim \chi^2_{(b-1)}$, and SS_A and SS_E are independent, then

$$F_A = \frac{SS_A/(a-1)}{SS_E/(ab(n-1))} = \frac{MS_A}{MS_E} \sim F_{a-1,ab(n-1)} \quad \text{under } H_0$$

- Similarly,

$$F_B = \frac{SS_B/(b-1)}{SS_E/(ab(n-1))} = \frac{MS_B}{MS_E} \sim F_{b-1,ab(n-1)} \quad \text{under } H_0$$

$$F_{AB} = \frac{SS_{AB}/(a-1)(b-1)}{SS_E/(ab(n-1))} = \frac{MS_{AB}}{MS_E} \sim F_{(a-1)(b-1),ab(n-1)} \quad \text{under } H$$

Analysis of variance table

■ **TABLE 5.3**
The Analysis of Variance Table for the Two-Factor Factorial, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	$a - 1$	$MS_A = \dfrac{SS_A}{a - 1}$	$F_0 = \dfrac{MS_A}{MS_E}$
B treatments	SS_B	$b - 1$	$MS_B = \dfrac{SS_B}{b - 1}$	$F_0 = \dfrac{MS_B}{MS_E}$
Interaction	SS_{AB}	$(a - 1)(b - 1)$	$MS_{AB} = \dfrac{SS_{AB}}{(a - 1)(b - 1)}$	$F_0 = \dfrac{MS_{AB}}{MS_E}$
Error	SS_E	$ab(n - 1)$	$MS_E = \dfrac{SS_E}{ab(n - 1)}$	
Total	SS_T	$abn - 1$		

Analysis of variance table

Manual computing formulas for sum of squares

$$SS_T = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 = \sum_i \sum_j \sum_k y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$SS_A = bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \frac{1}{bn} \sum_i y_{i..}^2 - \frac{y_{...}^2}{abn}$$

$$SS_B = an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = \frac{1}{an} \sum_j y_{.j.}^2 - \frac{y_{...}^2}{abn}$$

Analysis of variance table

Manual computing formulas for sum of squares

$$SS_{AB} = n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\dots})^2$$

$$= n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{\dots})^2 - SS_A - SS_B \quad (how?)$$

$$= SS_{\text{subtotals}} - SS_A - SS_B,$$

where

$$SS_{\text{subtotals}} = n \sum_i \sum_j (\bar{y}_{ij\cdot} - \bar{y}_{\dots})^2 = \frac{1}{n} \sum_i \sum_j y_{ij\cdot}^2 - \frac{y_{\dots}^2}{abn}$$

Then

$$SS_E = SS_T - SS_{AB} - SS_A - SS_B$$

or

$$SS_E = SS_T - SS_{subtotals}$$

Subsection 5

The battery design experiment

The battery design experiment

- Let y_{ijk} denote the observed lifetime of the battery corresponding to the kth replication of the treatment combination i^{th} **material type** (treatment A) and j^{th} **temperature** (treatment B) $(i = 1, 2, 3; \ j = 1, 2, 3; \ k = 1, \dots, 4)$.

- Consider the effects model

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

where μ is the overall mean, τ_i and β_j are the effects of the ith level of factor A and j^{th} level of factor B, $(\tau\beta)_{ij}$ is the interction between the ith level of factor A and j^{th} level of factor B.

- Random error term ϵ_{ijk} is assumed to be normally distributed with mean 0 and a constant variance σ^2.

The battery design experiment

■ **TABLE 5.4**
Life Data (in hours) for the Battery Design Experiment

Material Type	Temperature (°F)									$y_{i..}$
	15			**70**			**125**			
1	130	155	⑤③⑨	34	40	②②⑨	20	70	②③⓪	998
	74	180		80	75		82	58		
2	150	188	⑥②③	136	122	④⑦⑨	25	70	①⑨⑧	1300
	159	126		106	115		58	45		
3	138	110	⑤⑦⑥	174	120	⑤⑧③	96	104	③④②	1501
	168	160		150	139		82	60		
$y_{.j.}$	1738			1291			770			$3799 = y_{...}$

$$SS_T = \sum_i \sum_j \sum_j y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$= (130)^2 + \cdots + (60)^2 - \frac{(3799)^2}{36} = 77646.97$$

The battery design experiment

$$SS_A = \frac{1}{bn} \sum_i y_{i\cdot\cdot}^2 - \frac{y_{\cdot\cdot\cdot}^2}{abn}$$

$$= \frac{1}{(3)(4)} \{(998)^2 + (1300)^2 + (1501)^2\} - \frac{(3799)^2}{36} = 10683.72$$

$$SS_B = \frac{1}{an} \sum_j y_{\cdot j\cdot}^2 - \frac{y_{\cdot\cdot\cdot}^2}{abn}$$

$$= \frac{1}{(3)(4)} \{(1738)^2 + (1291)^2 + (770)^2\} - \frac{(3799)^2}{36} = 39118.72$$

$$SS_{\text{Subtotals}} = \frac{1}{n} \sum_i \sum_j y_{ij\cdot}^2 - \frac{y_{\cdot\cdot\cdot}^2}{abn}$$

$$= \frac{1}{4} \{(539)^2 + \cdots + (342)^2\} - \frac{(3799)^2}{36} = 49515373$$

The battery design experiment

$$SS_T = 77,646.97$$
$$SS_A = 10,683.72$$
$$SS_B = 39118.72$$
$$SS_{AB} = SS_{\text{subtotals}} - SS_A - SS_B$$
$$= 49515373 - 10683.72 - 39118.72$$
$$= 9,613.78$$
$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$
$$= 77646.97 - 10683.72 - 39118.72 - 9613.78$$
$$= 18,230.78$$

The battery design experiment

## The battery design experiment (anova table)

■ **TABLE 5.5**
Analysis of Variance for Battery Life Data

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Material types	10,683.72	2	5,341.86	7.91	0.0020
Temperature	39,118.72	2	19,559.36	28.97	< 0.0001
Interaction	9,613.78	4	2,403.44	3.56	0.0186
Error	18,230.75	27	675.21		
Total	77,646.97	35			

- There is a significant interaction between material type and temperature
- Main effects of material type and temperature are also significant

The battery design experiment
The battery design experiment (interpreting results)
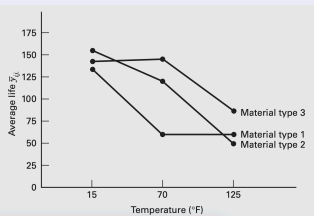
1. The significant interaction is indicated by the lack of parallelism of the lines
2. In general, longer life is attained at the low temperature, regardless of material type.
3. Changing from low to intermediate temperature, battery life with material type 3 increases, whereas it decreases for types 1 and 2.
4. From intermediate to high temperature, battery life decreases for material types 2 and 3, and is essentially unchanged for type 1.
5. Material type 3 seems to give the best result if we want less loss of effective life as the temperature changes.

The battery design experiment

The battery design experiment (multiple comparisons)

One of the goals of the experiment is to identify the best treatment combination. In two-factor factorial experiment, significance of interaction plays an important role in selecting the best treatment combination.

- **When interaction is not significant**, multiple comparison methods can be used to identify the best level for each factor separately.
- **When interaction is significant**, the best level of one factor need to be identified at each level of the other factor. e.g. comparisons between the means of factor A can be obtained for a specific level of factor B applying Tukey's test.

The battery design experiment

- Suppose we are interested in detecting differences among the means of the three material types.

- Because interaction is significant, we make this comparison at just one level of temperature, say level 2 (70°F)

- The three material type averages at $70°\mathrm{F}$ arranged in ascending order are

$$\bar{y}_{12.} = 57.25 \quad (\text{ material type } 1)$$
$$\bar{y}_{22.} = 119.75 \quad (\text{ material type } 2)$$
$$\bar{y}_{32.} = 145.75 \quad (\text{ material type } 3)$$

The battery design experiment

$$T_{0.05} = q_{0.05}(3, 27)\sqrt{\frac{MS_E}{n}}$$

$$= 3.50\sqrt{\frac{675.21}{4}}$$

$$= 45.47$$

where we obtained $q_{0.05}(3, 27) \simeq 3.50$ by interpolation in Appendix Table VII. The pairwise comparisons yield

3 vs. 1:	$145.75 - 57.25 = 88.50 > T_{0.05} = 45.47$	
3 vs. 2:	$145.75 - 119.75 = 26.00 < T_{0.05} = 45.47$	
2 vs. 1:	$119.75 - 57.25 = 62.50 > T_{0.05} = 45.47$	

This analysis indicates that at the temperature level 70°F, the mean battery life is the same for material types 2 and 3, and that the mean battery life for material type 1 is significantly lower in comparison to both types 2 and 3.

Model adequacy checking

- Before making the conclusions from the analysis of variance, the adequacy of the underlying model should be checked using **residual analysis** (e.g. checking normality, independence, constant variance, etc.).

- The residuals for the two-factor factorial model

$$e_{ijk} = y_{ijk} - \widehat{y}_{ijk}$$
$$= y_{ijk} - \bar{y}_{ij\cdot}$$

- **Different tools of residual analysis:**
  - ▶ q-q normal plot of residuals
  - ▶ Plot of residuals against fitted values
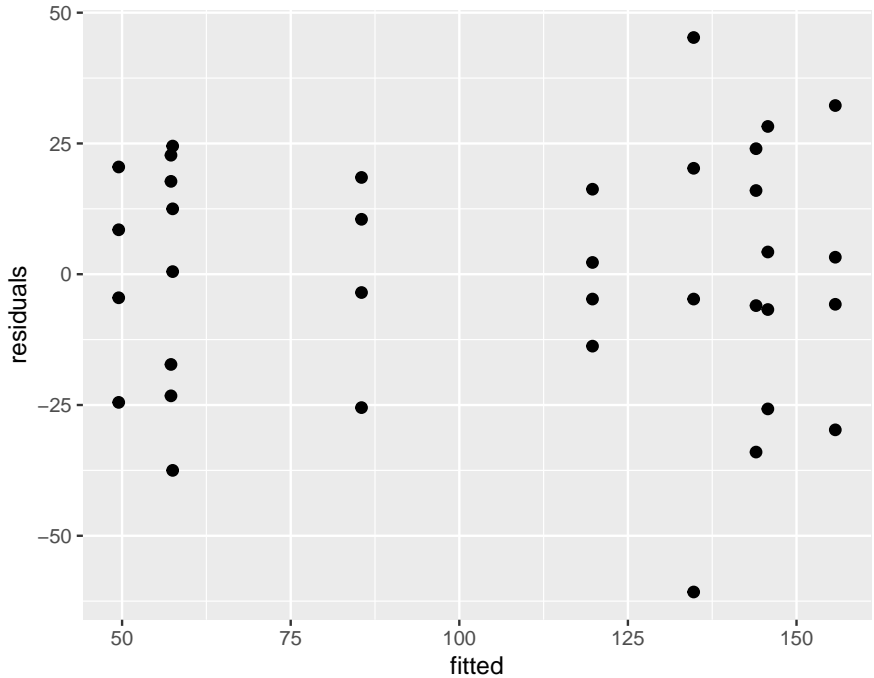 - ▶ Plot of residuals against different factors separately

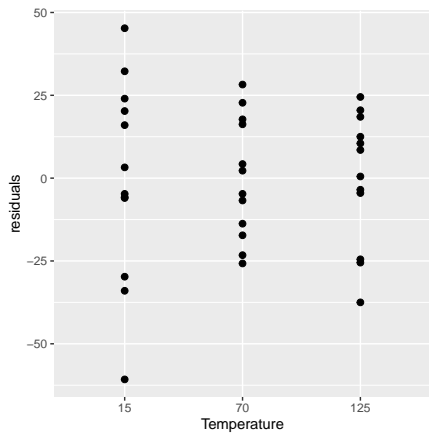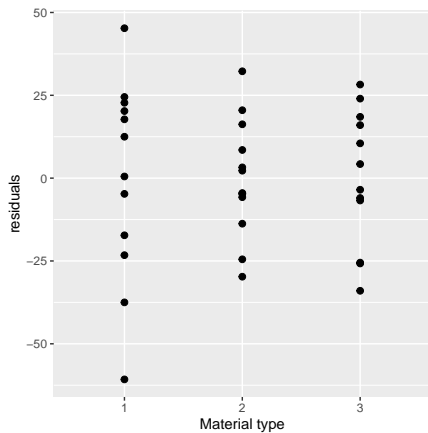# Model adequacy checking

■ **TABLE 5.6**
Residuals for Example 5.1

Material Type	Temperature (°F)					
	15		**70**		**125**	
1	−4.75	20.25	−23.25	−17.25	−37.50	12.50
	−60.75	45.25	22.75	17.75	24.50	0.50
2	−5.75	32.25	16.25	2.25	−24.50	20.50
	3.25	−29.75	−13.75	−4.75	8.50	−4.50
3	−6.00	−34.00	28.25	−25.75	10.50	18.50
	24.00	16.00	4.25	−6.75	−3.50	−25.50
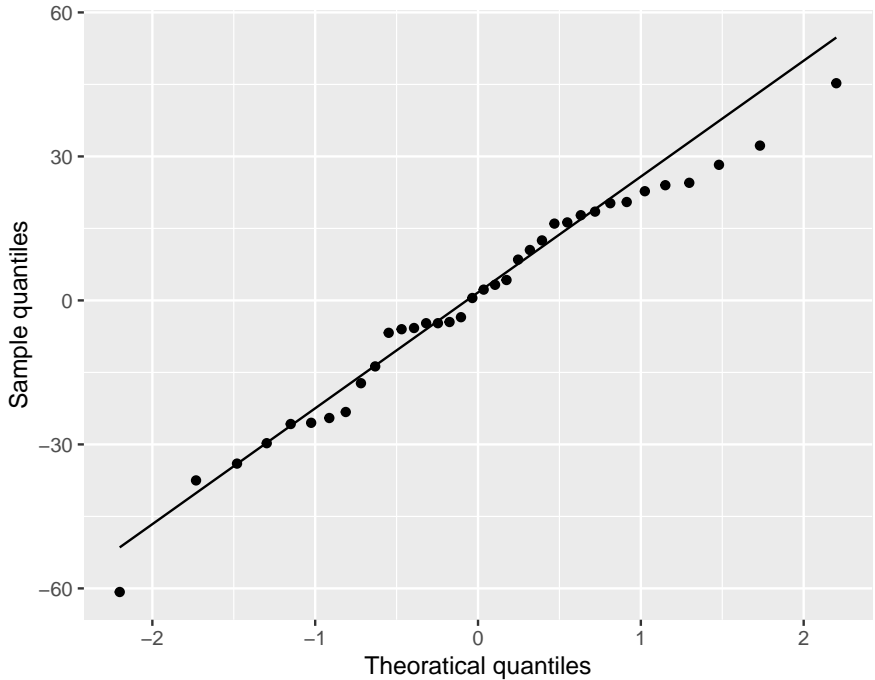
Model adequacy checking

Subsection 6

Estimating model parameters

Estimating model parameters

The effects model for two-factor factorial is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}, \tag{3}$$

where μ, τ_i, β_j, and $(\tau\beta)_{ij}$ are parameters of interest, and random error term is assumed to be normally distributed, i.e.

$$\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \;\; \Rightarrow \;\; y_{ijk} \sim \mathcal{N}(\mu_{ij}, \sigma^2)$$

- They may be estimated by least squares. Because the model has $1 + a + b + ab$ parameters to be estimated, and there $1 + a + b + ab$ normal equations.

Estimating model parameters

Using the method of Section 3.9, we find that it is not difficult to show that the normal equations are

$$\mu: abn\hat{\mu} + bn\sum_{i=1}^{a}\hat{\tau}_i + an\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{(\tau\beta)}_{ij} = y_{...} \tag{5.14a}$$

$$\tau_i: bn\hat{\mu} + bn\hat{\tau}_i + n\sum_{j=1}^{b}\hat{\beta}_j + n\sum_{j=1}^{b}\widehat{(\tau\beta)}_{ij} = y_{i..} \qquad i = 1, 2, \ldots, a \tag{5.14b}$$

$$\beta_j: an\hat{\mu} + n\sum_{i=1}^{a}\hat{\tau}_i + an\hat{\beta}_j + n\sum_{i=1}^{a}\widehat{(\tau\beta)}_{ij} = y_{.j.} \qquad j = 1, 2, \ldots, b \tag{5.14c}$$

$$(\tau\beta)_{ij}: n\hat{\mu} + n\hat{\tau}_i + n\hat{\beta}_j + n\widehat{(\tau\beta)}_{ij} = y_{ij.} \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases} \tag{5.14d}$$

Estimating model parameters

- The effects model (Equation 3) is an overparameterized model.

- Notice that the a equations in Equation 5.14b sum to Equation 5.14a and that the b equations of Equation 5.14c sum to Equation 5.14a.

- Also summing Equation 5.14d over j for a particular i will give Equation 5.14b, and summing Equation 5.14d over i for a particular j will give Equation 5.14c.

- Therefore, there are $a + b + 1$ linear dependencies in this system of equations, and no unique solution will exist.

- In order to obtain a solution, we impose the constraints

Estimating model parameters

$$\sum_{i=1}^{a} \hat{\tau}_i = 0 \tag{5.15a}$$

$$\sum_{j=1}^{b} \hat{\beta}_j = 0 \tag{5.15b}$$

$$\sum_{i=1}^{a} \widehat{(\tau\beta)}_{ij} = 0 \qquad j = 1, 2, \ldots, b \tag{5.15c}$$

$$\sum_{j=1}^{b} \widehat{(\tau\beta)}_{ij} = 0 \qquad i = 1, 2, \ldots, a \tag{5.15d}$$

Equations 5.15a and 5.15b constitute two constraints, whereas Equations 5.15c and 5.15d form $a + b - 1$ independent constraints. Therefore, we have $a + b + 1$ total constraints, the number needed.

Estimating model parameters

Applying these constraints, the normal equations (Equations 5.14) simplify considerably, and we obtain the solution

$$\hat{\mu} = \bar{y}_{...}$$
$$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \qquad i = 1, 2, \ldots, a$$
$$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} \qquad j = 1, 2, \ldots, b$$
$$(\widehat{\tau\beta})_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \qquad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \end{cases}$$

The MLE

The log-likelihood function

$$l(\theta) = \sum_i \sum_j \sum_k \left\{ -\frac{\log \sigma^2}{2} - \frac{1}{2\sigma^2} \left(y_{ijk} - \mu_{ij} \right)^2 + constant \right\},$$

where $\theta = \{\mu, \tau_i, \beta_j, (\tau\beta)_{ij}\}$ is the vector of parameters of order $1 + a + b + ab$.

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \epsilon_{ijk}^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(y_{ijk} - \mu_{ij} \right)^2.$$

- Here

$$\hat{\theta}_{MLE} = arg\,max\ l(\theta)$$

$$\hat{\theta}_{LSE} = arg\,min\ SSE = arg\,max\ l(\theta)$$

- Therefore, the solutions obtained (through LSE) are also MLE.

Subsection 7

Choice of sample size

Choice of sample size

- In any experimental design problem, a critical decision is the choice of sample size — that is, determining the number of replicates to run.

- Generally, if the experimenter is interested in detecting small effects, more replicates are required than if the experimenter is interested in detecting large effects.

OC (Operating Characteristic) curve

- An operating characteristic (OC) curve is a plot of the type II error probability (β) of a statistical test for a particular sample size versus a parameter Φ that reflects the extent to which the null hypothesis is false.

- These curves can be used to guide the experimenter in selecting the number of replicates so that the design will be sensitive to important potential differences in the treatments.

- The operating characteristic curves in Appendix Chart V (Montgomery book) can be used to assist the experimenter in determining an appropriate sample size (number of replicates, n) for a two-factor factorial design.

- Curves are available for $\alpha = 0.05$ and $\alpha = 0.01$ and a range of degrees of freedom for numerator and denominator.

OC (Operating Characteristic) curve

V   Operating Characteristic Curves for the Fixed Effects Model Analysis of Variance[a]
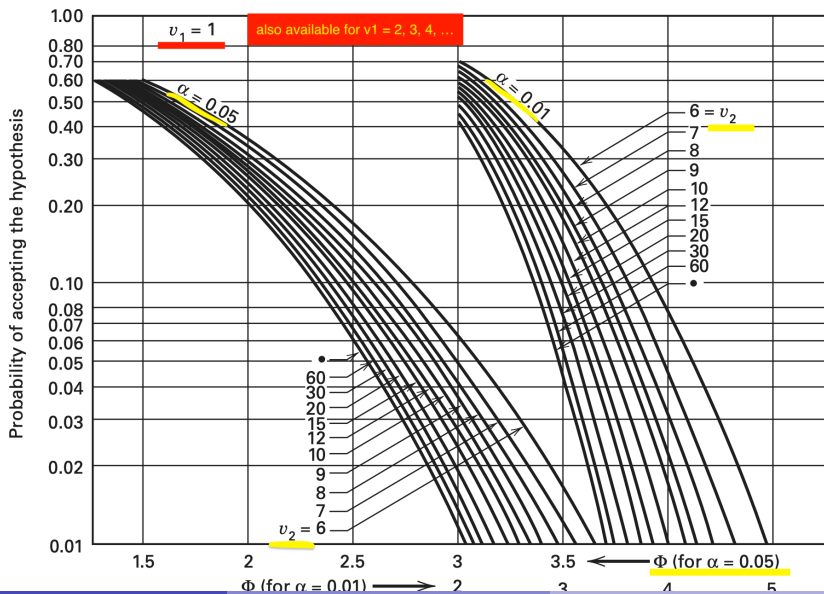
OC curve for two-factor factorial design

OC Curve parameters for chart V of the Appendix for the two-factor factorial, fixed effects model

Factor	Φ^2	Numerator DF	Denominator DF
A	$\frac{bn\sum_{i=1}^{a}\tau_i^2}{a\sigma^2}$	$a-1$	$ab(n-1)$
B	$\frac{an\sum_{j=1}^{b}\beta_j^2}{b\sigma^2}$	$b-1$	$ab(n-1)$
AB	$\frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{\sigma^2[(a-1)(b-1)+1]}$	$(a-1)(b-1)$	$ab(n-1)$

- For two-factor factorial design, the appropriate value of the parameter Φ and the numerator and denominator degrees of freedom are shown in the table.

OC curve for two-factor factorial design

- To determine Φ, we need to know the actual values of the treatment means on which the sample size decision should be based. If we know that, we can use the formulas of the table and proceed as the one-factor design. But a set of actual treatment means is not available most of the time.

- An alternate approach is to select a sample size such that if the difference between any two treatment means exceeds a specified value, the null hypothesis should be rejected.

OC curve for two-factor factorial design

- For example, if the difference in any two row (Factor A) means is D, then the minimum value of Φ^2 is

$$\Phi^2 = \frac{nbD^2}{2a\sigma^2}$$

- If the difference in any two column (Factor B) means is D, then the minimum value of Φ^2 is

$$\Phi^2 = \frac{naD^2}{2b\sigma^2}$$

- Finally, the minimum value of Φ^2 corresponding to a difference of D between any two interaction effects is

$$\Phi^2 = \frac{nD^2}{2\sigma^2[(a-1)(b-1)+1]}.$$

OC curve…(Battery design experiment)

- To illustrate the use of these equations, consider the battery design experiment.

- Suppose that before running the experiment we decide that the null hypothesis should be rejected with a high probability if the difference in mean battery life between any two temperatures is as great as 40 hours.

- Thus a difference of $D = 40$ has engineering significance, and if we assume that the standard deviation of battery life is approximately 25 , then the corresponding equation gives

$$\Phi^2 = \frac{naD^2}{2b\sigma^2}$$
$$= \frac{n(3)(40)^2}{2(3)(25)^2}$$
$$= 1.28n$$

as the minimum value of Φ.

OC curve…(Battery design experiment)

Assuming that $\alpha = 0.05$, we can now use Appendix Table V to construct the following display:

n	Φ^2	Φ	$\nu_1 =$ Numerator Degrees of Freedom	$\nu_2 =$ Error Degrees of Freedom	β
2	2.56	1.60	2	9	0.45
3	3.84	1.96	2	18	0.18
4	5.12	2.26	2	27	0.06

OC curve…(Battery design experiment)

- Note that $n = 4$ replicates give a β risk of about 0.06, or approximately a 94 percent chance of rejecting the null hypothesis if the difference in mean battery life at any two temperature levels is as large as 40 hours.

- Thus, we conclude that four replicates are enough to provide the desired sensitivity as long as our estimate of the standard deviation of battery life is not seriously in error.

- If in doubt, the experimenter could repeat the above procedure with other values of σ to determine the effect of mis-estimating this parameter on the sensitivity of the design.

Subsection 8

Assumption of no interaction in a two-factor model

Assumption of no interaction in a two-factor model

Occasionally, an experimenter feels that a two-factor model without interaction is appropriate, say

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk} \left\{ \begin{array}{l} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., n \end{array} \right.$$

The statistical analysis of a two-factor factorial model without interaction is straightforward. The following table presents the analysis of the battery design experiment, assuming no interaction.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Material types	10,683.72	2	5,341.86	5.95
Temperature	39,118.72	2	19,559.36	21.78
Error	27,844.52	31	898.21	
Total	77,646.96	35		

Assumption of no interaction in a two-factor model

As noted previously, both main effects are significant. However, as soon as a residual analysis is performed for these data, it becomes clear that the no-interaction model is inadequate.

For the two-factor model without interaction, the fitted values are $\hat{y}_{ijk} = \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}$. A plot of fitted versus residuals is shown below.
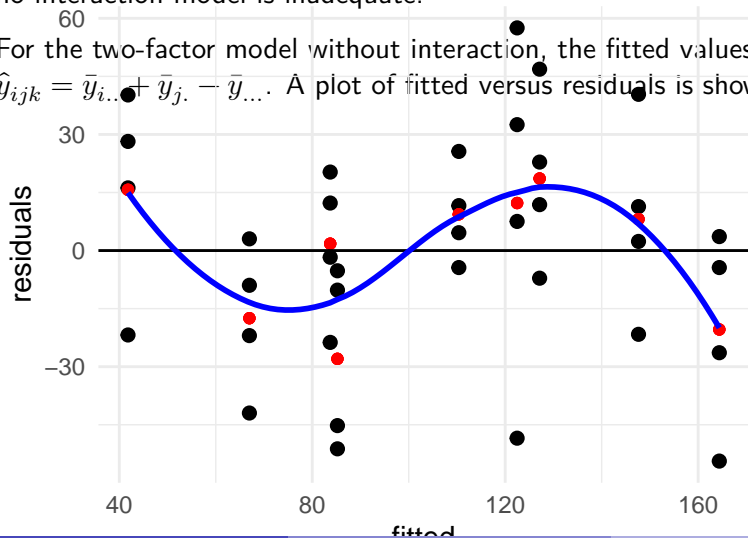
Subsection 9

One observation per cell

One observation per cell

Occasionally, one encounters a two-factor experiment with only a single replicate, that is, only one observation per cell. If there are two factors and only one observation per cell, the effects model is

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij} \quad \left\{ \begin{array}{l} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{array} \right.$$

The analysis of variance for this situation is shown in Table 5.9, assuming that both factors are fixed.

One observation per cell

■ **TABLE 5.9**

Analysis of Variance for a Two-Factor Model, One Observation per Cell

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square
Rows (A)	$\sum_{i=1}^{a} \dfrac{y_{i.}^2}{b} - \dfrac{y_{..}^2}{ab}$	$a - 1$	MS_A	$\sigma^2 + \dfrac{b \sum \tau_i^2}{a - 1}$
Columns (B)	$\sum_{j=1}^{b} \dfrac{y_{.j}^2}{a} - \dfrac{y_{..}^2}{ab}$	$b - 1$	MS_B	$\sigma^2 + \dfrac{a \sum \beta_j^2}{b - 1}$
Residual or AB	Subtraction	$(a - 1)(b - 1)$	MS_{Residual}	$\sigma^2 + \dfrac{\sum \sum (\tau\beta)_{ij}^2}{(a - 1)(b - 1)}$
Total	$\sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij}^2 - \dfrac{y_{..}^2}{ab}$	$ab - 1$		

One observation per cell

From examining the expected mean squares, we see that the error variance σ^2 is not estimable; that is, the two-factor interaction effect $(\tau\beta)_{ij}$ and the experimental error cannot be separated in any obvious manner. Consequently, there are no tests on main effects unless the interaction effect is zero. If there is no interaction present, then $(\tau\beta)_{ij} = 0$ for all i and j, and a plausible model is

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \left\{ \begin{array}{l} i = 1, 2, ..., a \\ j = 1, 2, ..., b \end{array} \right.$$

If this later model is appropriate, then the residual mean square in Table 5.9 is an unbiased estimator of σ^2, and the main effects may be tested by comparing MS_A and MS_B to MS_{Residual} .

Subsection 10

5.4 The general factorial design

General factorial design

The two-factor factorial design may be extended to the general case where there are a levels of factor A, b levels of factor B, c levels of factor C, and so on.

In general, there will be $abc \ldots n$ total observations if there are n replicates of the complete experiment.

Note that we must have at least two replicates $(n \geq 2)$ to determine a sum of squares due to error if all possible interactions are included in the model.

General factorial design

For example, consider the three-factor analysis of variance model:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

$$\begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, n \end{cases}$$

General factorial design

Assuming A, B, and C are fixed, the ANOVA table is

■ **TABLE 5.12**
The Analysis of Variance Table for the Three-Factor Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Expected Mean Square	F_0
A	SS_A	$a-1$	MS_A	$\sigma^2 + \dfrac{bcn\sum \tau_i^2}{a-1}$	$F_0 = \dfrac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$\sigma^2 + \dfrac{acn\sum \beta_j^2}{b-1}$	$F_0 = \dfrac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$\sigma^2 + \dfrac{abn\sum \gamma_k^2}{c-1}$	$F_0 = \dfrac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\sigma^2 + \dfrac{cn\sum\sum (\tau\beta)_{ij}^2}{(a-1)(b-1)}$	$F_0 = \dfrac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$\sigma^2 + \dfrac{bn\sum\sum (\tau\gamma)_{ik}^2}{(a-1)(c-1)}$	$F_0 = \dfrac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$\sigma^2 + \dfrac{an\sum\sum (\beta\gamma)_{jk}^2}{(b-1)(c-1)}$	$F_0 = \dfrac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$\sigma^2 + \dfrac{n\sum\sum\sum (\tau\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$	$F_0 = \dfrac{MS_{ABC}}{MS_E}$
Error	SS_E	$abc(n-1)$	MS_E	σ^2	
Total	SS_T	$abcn-1$			

General factorial design

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{abcn}$$

$$SS_A = \frac{1}{bcn} \sum_{i=1}^{a} y_{i...}^2 - \frac{y_{....}^2}{abcn}$$

$$SS_B = \frac{1}{acn} \sum_{j=1}^{b} y_{.j..}^2 - \frac{y_{....}^2}{abcn}$$

$$SS_C = \frac{1}{abn} \sum_{k=1}^{c} y_{..k.}^2 - \frac{y_{....}^2}{abcn}$$

General factorial design

$$SS_{AB} = \frac{1}{cn} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B$$

$$= SS_{\text{Subtotals}(AB)} - SS_A - SS_B$$

$$SS_{AC} = \frac{1}{bn} \sum_{i=1}^{a} \sum_{k=1}^{c} y_{i.k.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_C$$

$$= SS_{\text{Subtotals}(AC)} - SS_A - SS_C$$

$$SS_{BC} = \frac{1}{an} \sum_{j=1}^{b} \sum_{k=1}^{c} y_{.jk.}^2 - \frac{y_{....}^2}{abcn} - SS_B - SS_C$$

$$= SS_{\text{Subtotals}(BC)} - SS_B - SS_C$$

$$SS_{ABC} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} y_{ijk.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$= SS_{\text{Subtotals}(ABC)} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

$$SS_E = SS_T - SS_{\text{Subtotals}(ABC)}$$

Soft drink bottling problem

A soft drink bottler is interested in obtaining more uniform fill heights in the bottles produced by his manufacturing process. The filling machine theoretically fills each bottle to the correct target height, but in practice, there is variation around this target, and the bottler would like to understand the sources of this variability better and eventually reduce it.

The process engineer can control three variables during the filling process: the percent carbonation (A), the operating pressure in the filler (B), and the bottles produced per minute or the line speed (C). For purposes of an experiment, the engineer can control carbonation at three levels: 10, 12, and 14 percent. She chooses two levels for pressure (25 and 30 psi) and two levels for line speed (200 and 250 bpm)

Soft drink bottling problem

The engineer decides to run two replicates of a factorial design in these three factors, with all 24 runs taken in random order. The response variable observed is the average deviation from the target fill height observed in a production run of bottles at each set of conditions. Positive deviations are fill heights above the target, whereas negative deviations are fill heights below the target. The circled numbers are the three-way cell totals y_{ijk}.

Soft drink bottling problem

■ TABLE 5.13
Fill Height Deviation Data for Example 5.3

Percent Carbonation (A)	Operating Pressure (B)				$y_{i...}$
	25 psi		30 psi		
	Line Speed (C)		Line Speed (C)		
	200	250	200	250	
10	-3 -1 (−4)	-1 0 (−1)	-1 0 (−1)	1 1 ②	-4
12	0 1 ①	2 1 ③	2 3 ⑤	6 5 ⑪	20
14	5 4 ⑨	7 6 ⑬	7 9 ⑯	10 11 ㉑	59
$B \times C$ Totals $y_{.jk.}$	6	15	20	34	$75 = y_{....}$
$y_{.j..}$	21		54		

$A \times B$ Totals $y_{ij..}$			$A \times C$ Totals $y_{i.k.}$		
A	B		A	C	
	25	30		200	250
10	−5	1	10	−5	1
12	4	16	12	6	14
14	22	37	14	25	34

Soft drink bottling problem

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} \sum_{l=1}^{n} y_{ijkl}^2 - \frac{y_{....}^2}{abcn}$$

$$= 571 - \frac{(75)^2}{24} = 336.625$$

$$SS_{\text{Carbonation}} = \frac{1}{bcn} \sum_{i=1}^{a} y_{i...}^2 - \frac{y_{....}^2}{abcn}$$

$$= \frac{1}{8} [(-4)^2 + (20)^2 + (59)^2]$$

$$- \frac{(75)^2}{24} = 252.750$$

$$SS_{\text{Pressure}} = \frac{1}{acn} \sum_{j=1}^{b} y_{.j..}^2 - \frac{y_{....}^2}{abcn}$$

$$= \frac{1}{12} [(21)^2 + (54)^2] - \frac{(75)^2}{24} = 45.375$$

Soft drink bottling problem

$$SS_{\text{Speed}} = \frac{1}{abn} \sum_{k=1}^{c} y_{..k.}^2 - \frac{y_{....}^2}{abcn}$$

$$= \frac{1}{12}[(26)^2 + (49)^2] - \frac{(75)^2}{24} = 22.042$$

$$SS_{AB} = \frac{1}{cn} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij..}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B$$

$$= \frac{1}{4}[(-5)^2 + (1)^2 + (4)^2 + (16)^2 + (22)^2 + (37)^2]$$

$$- \frac{(75)^2}{24} - 252.750 - 45.375$$

$$= 5.250$$

Soft drink bottling problem

$$
\begin{aligned}
SS_{AC} &= \frac{1}{bn} \sum_{i=1}^{a} \sum_{k=1}^{c} y_{i.k.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_C \\
&= \frac{1}{4}[(-5)^2 + (1)^2 + (6)^2 + (14)^2 + (25)^2 + (34)^2] \\
&\quad - \frac{(75)^2}{24} - 252.750 - 22.042 \\
&= 0.583
\end{aligned}
$$

$$
\begin{aligned}
SS_{BC} &= \frac{1}{an} \sum_{j=1}^{b} \sum_{k=1}^{c} y_{.jk.}^2 - \frac{y_{....}^2}{abcn} - SS_B - SS_C \\
&= \frac{1}{6}[(6)^2 + (15)^2 + (20)^2 + (34)^2] - \frac{(75)^2}{24} \\
&\quad - 45.375 - 22.042 \\
&= 1.042
\end{aligned}
$$

Soft drink bottling problem

$$SS_{ABC} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} y_{ijk.}^2 - \frac{y_{....}^2}{abcn} - SS_A - SS_B - SS_C$$

$$- SS_{AB} - SS_{AC} - SS_{BC}$$

$$= \frac{1}{2} [(-4)^2 + (-1)^2 + (-1)^2 + \cdots + (16)^2 + (21)^2]$$

$$- \frac{(75)^2}{24} - 252.750 - 45.375 - 22.042$$

$$- 5.250 - 0.583 - 1.042$$

$$= 1.083$$

$$SS_{\text{Subtotals}(ABC)} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} y_{ijk.}^2 - \frac{y_{....}^2}{abcn} = 328.125$$

$$SS_E = SS_T - SS_{\text{Subtotals}(ABC)}$$

$$= 336.625 - 328.125$$

$$= 8.500$$

Soft drink bottling problem

■ TABLE 5.14
Analysis of Variance for Example 5.3

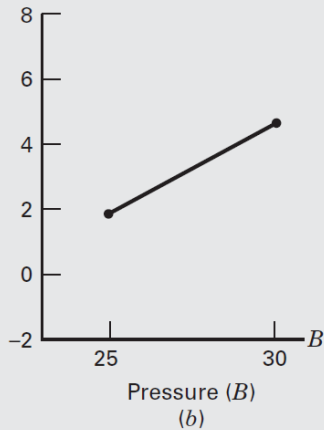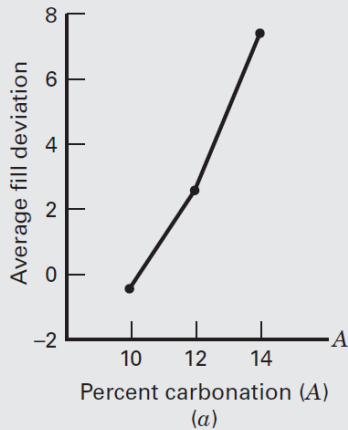| Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Percentage of carbonation (A)	252.750	2	126.375	178.412	<0.0001
Operating pressure (B)	45.375	1	45.375	64.059	<0.0001
Line speed (C)	22.042	1	22.042	31.118	0.0001
AB	5.250	2	2.625	3.706	0.0558
AC	0.583	2	0.292	0.412	0.6713
BC	1.042	1	1.042	1.471	0.2485
ABC	1.083	2	0.542	0.765	0.4867
Error	8.500	12	0.708		
Total	336.625	23			

Soft drink bottling problem

From the ANOVA we see that the percentage of carbonation, operating pressure, and line speed significantly affect the fill volume.
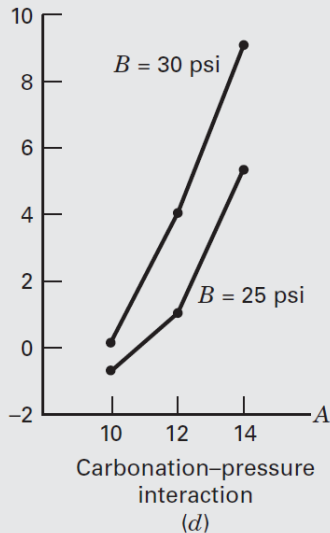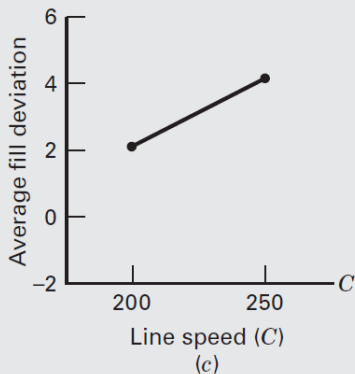
The carbonation pressure interaction F ratio has a p-value of 0.0558, indicating some interaction between these factors.

To assist in the practical interpretation of this experiment, the following figure presents plots of the three main effects and the AB (carbonation–pressure) interaction.

Soft drink bottling problem

(a) Percent carbonation (A)

(b) Pressure (B)

Soft drink bottling problem

(c) Average fill deviation vs Line speed (C), with C axis from 200 to 250.

(d) Carbonation–pressure interaction: Average fill deviation vs A, with A axis at 10, 12, 14. Lines for $B = 30$ psi and $B = 25$ psi.

Soft drink bottling problem

The main effect plots are just graphs of the marginal response averages at the levels of the three factors. Notice that all three variables have positive main effects; that is, increasing the variable moves the average deviation from the fill target upward.

The interaction between carbonation and pressure is fairly small, as shown by the similar shape of the two curves.

Subsection 11

5.5 Fitting Response Curves and Surfaces

5.5 Fitting Response Curves and Surfaces

The ANOVA always treats all of the factors in the experiment as if they were qualitative or categorical. Many experiments involve at least one quantitative factor. It can be useful to fit a **response curve** to the levels of a quantitative factor so that the experimenter has an equation that relates the response to the factor

This equation might be used for interpolation, that is, for predicting the response at factor levels between those actually used in the experiment. When at least two factors are quantitative, we can fit a **response surface** for predicting y at various combinations of the design factors. In general, **linear regression methods** are used to fit these models to the experimental data

The battery design experiment

Consider the battery life experiment described previously. The factor temperature is quantitative, and the material type is qualitative.

Furthermore, there are three levels of temperature.

Consequently, we can compute a linear and a quadratic temperature effect to study how temperature affects the battery life.

Because material type is a qualitative factor there is an equation for predicted life as a function of temperature for each material type.

The battery design experiment

Final Equation in Terms of Actual Factors:

Material Type 1

$$Life = +169.38017 - 2.48860 * Temp + 0.012851 * Temp^2$$

Material Type 2

$$Life = +159.62397 - 0.17901 * Temp + 0.41627 * Temp^2$$

Material Type 3

$$Life = +132.76240 + 0.89264 * Temp - 0.43218 * Temp^2$$

The battery design experiment

Figure 5.18 shows the response curves generated by these three prediction equations.
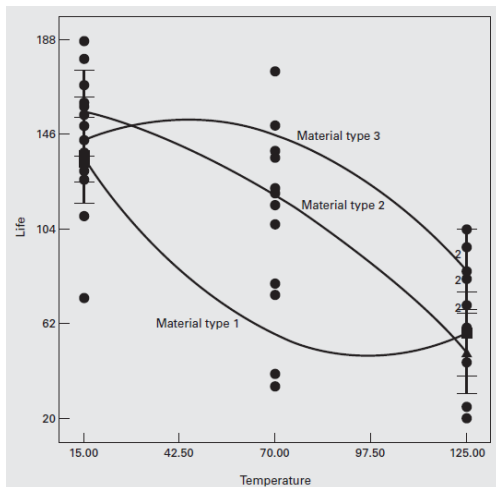
Subsection 12

5.6 Blocking in a facorial design

5.6 Blocking in a facorial design

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \epsilon_{ijk} \quad \left\{ \begin{array}{l} i = 1, 2, ..., a \\ j = 1, 2, ..., b \\ k = 1, 2, ..., n \end{array} \right.$$

where δ_k is the effect of the k th block. Of course, within a block the order in which the treatment combinations are run is completely randomized.

5.6 Blocking in a facorial design

■ TABLE 5.20

Analysis of Variance for a Two-Factor Factorial in a Randomized Complete Block

Source of Variation	Sum of Squares	Degrees of Freedom	Expected Mean Square	F_0
Blocks	$\dfrac{1}{ab}\sum_k y_{..k}^2 - \dfrac{y_{...}^2}{abn}$	$n - 1$	$\sigma^2 + ab\sigma_\delta^2$	
A	$\dfrac{1}{bn}\sum_i y_{i..}^2 - \dfrac{y_{...}^2}{abn}$	$a - 1$	$\sigma^2 + \dfrac{bn\sum \tau_i^2}{a - 1}$	$\dfrac{MS_A}{MS_E}$
B	$\dfrac{1}{an}\sum_j y_{.j.}^2 - \dfrac{y_{...}^2}{abn}$	$b - 1$	$\sigma^2 + \dfrac{an\sum \beta_j^2}{b - 1}$	$\dfrac{MS_B}{MS_E}$
AB	$\dfrac{1}{n}\sum_i\sum_j y_{ij.}^2 - \dfrac{y_{...}^2}{abn} - SS_A - SS_B$	$(a - 1)(b - 1)$	$\sigma^2 + \dfrac{n\sum\sum (\tau\beta)_{ij}^2}{(a - 1)(b - 1)}$	$\dfrac{MS_{AB}}{MS_E}$
Error	Subtraction	$(ab - 1)(n - 1)$	σ^2	
Total	$\sum_i\sum_j\sum_k y_{ijk}^2 - \dfrac{y_{...}^2}{abn}$	$abn - 1$		

Example 5.6

An engineer is studying methods for improving the ability to detect targets on a radar scope. Two factors she considers to be important are the amount of background noise, or **ground clutter,** on the scope and the **type of filter** placed over the screen. An experiment is designed using 3 levels of ground clutter and 2 filter types.

Because of operator availability, it is convenient to select an operator and keep him or her at the scope until all the necessary runs have been made. Furthermore, operators differ in their skill and ability to use the scope. Consequently, it seems logical to use the operators as blocks.

Four operators are randomly selected. Once an operator is chosen, the order in which the six treatment combinations are run is randomly determined.

Example 5.6

Thus, we have a 3×2 factorial experiment run in a randomized complete block. The data are shown below

■ **TABLE 5.21**
Intensity Level at Target Detection

Operators (blocks) Filter Type	1		2		3		4	
	1	**2**	**1**	**2**	**1**	**2**	**1**	**2**
Ground clutter								
Low	90	86	96	84	100	92	92	81
Medium	102	87	106	90	105	97	96	80
High	114	93	112	91	108	95	98	83

Example 5.6

The linear model for this experiment is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \epsilon_{ijk} \qquad \begin{cases} i = 1, 2, 3 \\ j = 1, 2 \\ k = 1, 2, 3, 4 \end{cases}$$

where τ_i represents the ground clutter effect, β_j represents the filter type effect, $(\tau\beta)_{ij}$ is the interaction, δ_k is the block effect, and ϵ_{ijk} is the NID$(0, \sigma^2)$ error component. The sums of squares for ground clutter, filter type, and their interaction are computed in the usual manner. The sum of squares due to blocks is found from the operator totals $\{y_{..k}\}$ as follows:

Example 5.6

$$SS_{\text{Blocks}} = \frac{1}{ab} \sum_{k=1}^{n} y_{..k}^2 - \frac{y_{...}^2}{abn}$$

$$= \frac{1}{(3)(2)} [(572)^2 + (579)^2 + (597)^2 + (530)^2]$$

$$- \frac{(2278)^2}{(3)(2)(4)}$$

$$= 402.17$$

Example 5.6

■ TABLE 5.22
Analysis of Variance for Example 5.6

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Ground clutter (G)	335.58	2	167.79	15.13	0.0003
Filter type (F)	1066.67	1	1066.67	96.19	<0.0001
GF	77.08	2	38.54	3.48	0.0573
Blocks	402.17	3	134.06		
Error	166.33	15	11.09		
Total	2047.83	23			

Example 5.6

Both ground clutter level and filter type are significant at the 1 percent level, whereas their interaction is significant only at the 10 percent level.

Thus, we conclude that both ground clutter level and the type of scope filter used affect the operator's ability to detect the target, and there is some evidence of mild interaction between these factors.

Subsection 13

Exercise

Exercise

The following output was obtained from a computer program that performed a two-factor ANOVA on a factorial experiment.

```
Two-way ANOVA: y versus A, B

Source        DF      SS        MS       F      P
A             1      _____     0.0002   ____   ____
B             _      180.378   _____   ____   ____
Interaction   3      8.479     _____   ____   0.932
Error         8      158.797   _____
Total         15     347.653
```

- (a) Fill in the blanks in the ANOVA table. You can use bounds on the p-values.
- (b) How many levels were used for factor B?
- (c) How many replicates of the experiment were performed?
- (d) What conclusions would you draw about this experiment?

Exercise

Practice exercises from Montgomery book:

5.41, 5.42, 5.43, 5.44, 5.45, 5.46