15 Analysis-of-Variance: The Cell Means Model for Unbalanced Data

15.1 INTRODUCTION

The theory of linear models for ANOVA applications was developed in Chapter 12. Although all the examples used in that and the following chapters have involved balanced data (where the number of observations is equal from one cell to another), the theory also applies to unbalanced data.

Chapters 13 and 14 show that simple and intuitive results are obtained when the theory is applied to balanced ANOVA situations. Intuitive marginal means are informative in analysis of the data [e.g., see (14.69) and (14.70)]. When applied to unbalanced data, however, the general results of Chapter 12 do not simplify to intuitive formulas. Even worse, the intuitive marginal means one is tempted to use can be misleading and sometimes paradoxical. This is especially true for two-way or higher-way data. As an example, consider the unbalanced two-way data in Figure 15.1. The data follow the two-way additive model (Section 12.1.2) with no error

$$y_{ij} = \mu + \alpha_i + \beta_j, \quad i = 1, 2, \quad j = 1, 2,$$

where $\mu = 25$, $\alpha_1 = 0$, $\alpha_2 = -20$, $\beta_1 = 0$, $\beta_2 = 5$. Simple marginal means of the data are given to the right and below the box.

The true effects of factors A and B are, respectively, $\alpha_2 - \alpha_1 = -20$ and $\beta_2 - \beta_1 = 5$. Even for error-free unbalanced data, however, naive estimates of these effects based on the simple marginal means are highly misleading. The effect of factor A appears to be $8.75 - 25.125 = -16.375$, and even more surprisingly the effect of factor B appears to be $15 - 20 = -5$.

Still other complications arise in the analysis of unbalanced data. For example, it was mentioned in Section 14.4.2.1 that many texts discourage testing for main effects in the presence of interactions. But little harm or controversy results from doing so when the data are balanced. The numerators for the main effect F tests are exactly

---

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

		B: 1	B: 2	
A	1	25, 25, 25	30	25.125
A	2	5	10, 10, 10	8.750
		20	15	

Figure 15.1 Hypotetical error-free data from an unbalanced two-way model.

the same whether the model with or without interactions is being entertained as the full model. Such is not the case for unbalanced data. The numerator sums of squares in these F tests depend greatly on which model is used as the full model, and, obviously, conclusions can be affected. Several types of sums of squares [usually types I, II, and III; see Milliken and Johnson (1984, pp. 138–158)] have been suggested to help clarify this issue.

The issues involved in choosing the appropriate full model for a test are subtle and often confusing. The use of different full models results in different weightings in the sums of squares calculations and expected mean squares. But some of the same weightings also arise for other reasons. For example, the weights might arise because the data are based on "probability proportional to size" (pps) sampling of populations (Cochran 1977, pp. 250–251).

Looking at this complex issue from different points of view has led to completely contradictory conclusions. For example, Milliken and Johnson (1984, p. 158) wrote that "in almost all cases, type III sums of squares will be preferred," whereas Nelder and Lane (1995) saw "no place for types III and IV sums of squares in making inferences from the use of linear models."

Further confusion regarding the analysis of unbalanced data has arisen from the interaction of computing advances with statistical practice. Historically, several different methods for unbalanced data analysis were developed as approximate methods, suitable for the computing resources available at the time. Looking back, however, we simply see a confusing array of alternative methods. Some such methods include weighted squares of means (Yates 1934; Morrison 1983, pp. 407–412), the method of unweighted means (Searle 1971; Winer 1971), the method of fitting constants (Rao 1965, pp. 211–214; Searle 1971, p. 139; Snedecor and Cochran 1967), and various methods of imputing data to make the dataset balanced (Hartley 1956; Healy and Westmacott 1969; Little and Rubin 2002, pp. 28–30).

The overparameterized (non-full rank) model (Sections 12.2, 12.5, 13.1, and 14.1) has some advantages in the analysis of unbalanced data, while the cell means approach (Section 12.1.1) has other advantages. The non-full rank approach builds the structure (additive two-way, full two-way, etc.) of the dataset into the model from the start, but relies on the subtle concepts of estimability, testability, and generalized inverses. The cell means model has the advantages of being a full-rank model, but the structure of the dataset is not an explicit part of the model. Whichever model is used, hard questions about the exact hypotheses of interest have to be faced. Many of the complexities are a matter of statistical practice rather than mathematical statistics.

The most extreme form of imbalance is that in which one or more of the cells have no observations. In this "empty cells" situation, even the cell means model is an overparameterized model. Nonetheless, the cell means approach allows one to deal specifically with nonestimability problems arising from the empty cells. Such an approach is almost impossible using the overparameterized approach.

In the remainder of this chapter we discuss the analysis of unbalanced data using the cell means model. Unbalanced one-way and two-way models are covered in Sections 15.2 and 15.3. In Section 15.4 we discuss the empty-cell situation.

15.2 ONE-WAY MODEL

The non-full-rank and cell means versions of the one-way unbalanced model are

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} && (15.1) \\ &= \mu_i + \varepsilon_{ij}, && (15.2) \\ i = 1, 2, \ldots, k,& \quad j = 1, 2, \ \ldots, n_i. \end{aligned}$$

For making inferences, we assume the ε_{ij}'s are independently distributed as $N(0, \sigma^2)$.

15.2.1 Estimation and Testing

To estimate the μ_i's, we begin by writing the $N = \sum_i n_i$ observations for the model (15.2) in the form

$$\mathbf{y} = \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\varepsilon}, \tag{15.3}$$

where

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}.$$

The normal equations are given by

$$\mathbf{W}'\mathbf{W}\hat{\boldsymbol{\mu}} = \mathbf{W}'\mathbf{y},$$

where $\mathbf{W}'\mathbf{W} = \text{diag}(n_1, n_2, \ldots, n_k)$ and $\mathbf{W}'\mathbf{y} = (y_{1.}, y_{2.}, \ldots, y_{k.})'$, with $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$. Since the matrix $\mathbf{W}$ is full rank, we have, by (7.6)

$$\hat{\boldsymbol{\mu}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} \tag{15.4}$$

$$= \bar{\mathbf{y}} = \begin{pmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \vdots \\ \bar{y}_{k.} \end{pmatrix}, \tag{15.5}$$

where $\bar{y}_{i.} = \sum_{j=1}^{n_i} y_{ij}/n_i$.

To test $H_0\colon \mu_1 = \mu_2 = \cdots = \mu_k$, we compare the full model in (15.2) and (15.3) with the reduced model $y_{ij} = \mu + \varepsilon_{ij}^*$, where μ is the common value of $\mu_1, \mu_2, \ldots, \mu_k$ under H_0. (We do not use the notation μ^* in the reduced model because there is no μ in the full model $y_{ij} = \mu_i + \varepsilon_{ij}$.) In matrix form, the N observations in the reduced model become $\mathbf{y} = \mu\mathbf{j} + \boldsymbol{\varepsilon}^*$, where $\mathbf{j}$ is $N \times 1$. For the full model, we have $\text{SS}(\mu_1, \mu_2, \ldots, \mu_k) = \hat{\boldsymbol{\mu}}'\mathbf{W}'\mathbf{y}$, and for the reduced model, we have $\text{SS}(\mu) = \hat{\mu}\mathbf{j}'\mathbf{y} = N\bar{y}_{..}^2$, where $N = \sum_i n_i$ and $\bar{y}_{..} = \sum_{ij} y_{ij}/N$. The difference $\text{SS}(\mu_1, \mu_2, \ldots, \mu_k) - \text{SS}(\mu)$ is equal to the regression sum of squares SSR in (8.6), which we denote by SSB for "between" sum of squares

$$\text{SSB} = \hat{\boldsymbol{\mu}}'\mathbf{W}'\mathbf{y} - N\bar{y}_{..}^2 = \sum_{i=1}^{k} \bar{y}_{i.}y_{i.} - N\bar{y}_{..}^2 \tag{15.6}$$

$$= \sum_{i=1}^{k} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}, \tag{15.7}$$

where $y_{..} = \sum_{ij} y_{ij}$ and $\bar{y}_{..} = y_{..}/N$. From (15.7), we see that SSB has $k - 1$ degrees of freedom. The error sum of squares is given by (7.24) or (8.6) as

$$\text{SSE} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\mu}}'\mathbf{W}'\mathbf{y}$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k} \frac{y_{i.}^2}{n_i}, \tag{15.8}$$

which has $N - k$ degrees of freedom. These sums of squares are summarized in Table 15.1.

TABLE 15.1 One-Way Unbalanced ANOVA

Source	Sum of Squares	df
Between	$\text{SSB} = \sum_i y_{i.}^2/n_i - y_{..}^2/N$	$k-1$
Error	$\text{SSE} = \sum_{ij} y_{ij}^2 - \sum_i y_{i.}^2/n_i$	$N-k$
Total	$\text{SST} = \sum_{ij} y_{ij}^2 - y_{..}^2/N$	$N-1$

The sums of squares SSB and SSE in Table 15.1 can also be written in the form

$$\text{SSB} = \sum_{i=1}^{k} n_i(\bar{y}_{i.} - \bar{y}_{..})^2, \tag{15.9}$$

$$\text{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2. \tag{15.10}$$

If we assume that the y_{ij}'s are independently distributed as $N(\mu_i, \sigma^2)$, then by Theorem 8.1d, an F statistic for testing H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ is given by

$$F = \frac{\text{SSB}/(k-1)}{\text{SSE}/(N-k)}. \tag{15.11}$$

If H_0 is true, F is distributed as $F(k-1, N-k)$.

Example 15.2.1. A sample from the output of five filling machines is given in Table 15.2 (Ostle and Mensing 1975, p. 359).

The analysis of variance is given in Table 15.3. The F is calculated by (15.11). There is no significant difference in the average weights filled by the five machines. □

15.2.2 Contrasts

A contrast in the population means is defined as $\delta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$, where $\sum_{i=1}^{k} c_i = 0$. The contrast can be expressed as $\delta = \mathbf{c}'\boldsymbol{\mu}$, where

TABLE 15.2 Net Weight of Cans Filled by Five Machines (A–E)

A	B	C	D	E
11.95	12.18	12.16	12.25	12.10
12.00	12.11	12.15	12.30	12.04
12.25		12.08	12.10	12.02
12.10				12.02

TABLE 15.3 ANOVA for the Fill Data in Table 15.2

Source	df	Sum of Squares	Mean Square	F	p Value
Between	4	.05943	.01486	1.9291	.176
Error	11	.08472	.00770		
Total	15	.14414			

$\text{SST} = \sum_{ij} y_{ij}^2 - y_{..}^2/N$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)'$. The best linear unbiased estimator of δ is given by $\hat{\delta} = c_1\bar{y}_{1.} + c_2\bar{y}_{2.} + \cdots + c_k\bar{y}_{k.} = \mathbf{c}'\hat{\boldsymbol{\mu}}$ [see (15.5) and Corollary 1 to Theorem 7.3d]. By (3.42), $\text{var}(\hat{\delta}) = \sigma^2\mathbf{c}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{c}$, which can be written as $\text{var}(\hat{\delta}) = \sigma^2 \sum_{i=1}^k c_i^2/n_i$, since $\mathbf{W}'\mathbf{W} = \text{diag}(n_1, n_2, \ldots, n_k)$. By (8.38), the F statistic for testing H_0: $\delta = 0$ is

$$F = \frac{(\mathbf{c}'\hat{\boldsymbol{\mu}})'\left[\mathbf{c}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{c}\right]^{-1}\mathbf{c}'\hat{\boldsymbol{\mu}}}{s^2}, \tag{15.12}$$

$$= \frac{\left(\sum_{i=1}^k c_i\bar{y}_{i.}\right)^2 / \left(\sum_{i=1}^k c_i^2/n_i\right)}{s^2}, \tag{15.13}$$

where $s^2 = \text{SSE}/(N-k)$ with SSE given by (15.8) or (15.10). We refer to the numerator of (15.13) as the sum of squares for the contrast. If H_0 is true, the F statistic in (15.12) or (15.13) is distributed as $F(1, N-k)$, and we reject H_0: $\delta = 0$ if $F \geq F_{\alpha, 1, N-k}$ or if $p \leq \alpha$, where p is the p value.

Two contrasts, say, $\hat{\delta} = \sum_{i=1}^k a_i\bar{y}i.$ and $\hat{\gamma} = \sum_{i=1}^k b_i\bar{y}_{i.}$, are said to be *orthogonal* if $\sum_{i=1}^k a_ib_i = 0$. However, in the case of unbalanced data, two orthogonal contrasts of this type are not independent, as they were in the balanced case (Theorem 13.6a).

Theorem 15.2. If the y_{ij}'s are independently distributed as $N(\mu_i, \sigma^2)$ in the unbalanced model (15.2), then two contrasts $\hat{\delta} = \sum_{i=1}^k a_i\bar{y}_{i.}$ and $\hat{\gamma} = \sum_{i=1}^k b_i\bar{y}_{i.}$ are independent if and only if $\sum_{i=1}^k a_ib_i/n_i = 0$.

Proof. We express the two contrasts in vector notation as $\hat{\delta} = \mathbf{a}'\bar{\mathbf{y}}$ and $\hat{\gamma} = \mathbf{b}'\bar{\mathbf{y}}$, where $\bar{\mathbf{y}} = (\bar{y}_{1.}, \bar{y}_{2.}, \ldots, \bar{y}_{k.})'$. By (7.14), we obtain

$$\text{cov}(\bar{\mathbf{y}}) = \sigma^2(\mathbf{W}'\mathbf{W})^{-1} = \sigma^2 \begin{pmatrix} 1/n_1 & 0 & \ldots & 0 \\ 0 & 1/n_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1/n_k \end{pmatrix} = \sigma^2\mathbf{D}.$$

Then by (3.43), we have

$$\begin{aligned}\text{cov}(\hat{\delta}, \hat{\gamma}) &= \text{cov}(\mathbf{a}'\bar{\mathbf{y}}, \mathbf{b}'\bar{\mathbf{y}}) = \mathbf{a}'\text{cov}(\bar{\mathbf{y}})\mathbf{b} = \sigma^2\mathbf{a}'\mathbf{D}\mathbf{b} \\ &= \sigma^2 \sum_{i=1}^{k} \frac{a_i b_i}{n_i}. \end{aligned} \tag{15.14}$$

Hence, by Theorem 4.4c, $\hat{\delta}$ and $\hat{\gamma}$ are independent if and only if $\sum_i a_i b_i / n_i = 0$. □

We refer to contrasts whose coefficients satisfy $\sum_i a_i b_i / n_i = 0$ as *weighted orthogonal contrasts*. If we define $k - 1$ contrasts of this type, they partition the treatment sum of squares SSB into $k - 1$ independent sums of squares, each with 1 degree of freedom. Unweighted orthogonal contrasts that satisfy only $\sum_i a_i b_i = 0$ are not independent (see Theorem 15.2), and their sums of squares do not add up to the treatment sum of squares (as they do for balanced data; see Theorem 13.6a).

In practice, weighted orthogonal contrasts are often of less interest than unweighted orthogonal contrasts because we may not wish to choose the a_i's and b_i's on the basis of the n_i's in the sample. The n_i's seldom reflect population characteristics that we wish to take into account. However, it is not necessary that the sums of squares be independent in order to proceed with the tests. If we use unweighted orthogonal contrasts with $\sum_i a_i b_i = 0$, the general linear hypothesis test based on (15.12) or (15.13) tests each contrast adjusted for the other contrasts (see Theorem 8.4d).

Example 15.2.2a. Suppose that we wish to compare the means of three treatments and that the coefficients of the orthogonal contrasts $\delta = \mathbf{a}'\boldsymbol{\mu}$ and $\gamma = \mathbf{b}'\boldsymbol{\mu}$ are given by $\mathbf{a}' = (2 - 1 - 1)$ and $\mathbf{b}' = (0\ 1\ {-1})$ with corresponding hypotheses

$$H_{01} : \mu_1 = \frac{1}{2}(\mu_2 + \mu_3), \quad H_{02} : \mu_2 = \mu_3.$$

If the sample sizes for the three treatments are, for example, $n_1 = 10$, $n_2 = 20$, and $n_3 = 5$, then the two estimated contrasts

$$\hat{\delta} = 2\bar{y}_{1.} - \bar{y}_{2.} - \bar{y}_{3.} \quad \text{and} \quad \hat{\gamma} = \bar{y}_{2.} - \bar{y}_{3.}$$

are not independent, and the corresponding sums of squares do not add to the treatment sum of squares.

The following two vectors provide an example of contrasts whose coefficients satisfy $\sum_i a_i b_i / n_i = 0$ for $n_1 = 10$, $n_2 = 20$, and $n_3 = 5$:

$$\mathbf{a}' = (25 - 20 - 5) \quad \text{and} \quad \mathbf{b}' = (0\ 1 {-1}). \tag{15.15}$$

However, $\mathbf{a}'$ leads to the comparison

$$H_{03} : 25\mu_1 = 20\mu_2 + 5\mu_3 \quad \text{or} \quad H_{03} : \mu_1 = \frac{4}{5}\mu_2 + \frac{1}{5}\mu_3,$$

which is not the same as the hypothesis $H_{01} : \mu_1 = \frac{1}{2}(\mu_2 + \mu_3)$ that we were initially interested in. □

Example 15.2.2b. We illustrate both weighted and unweighted contrasts for the fill data in Table 15.2. Suppose that we wish to make the following comparisons of the five machines:

A, D	versus	B, C, E
B, E	versus	D
A	versus	D
B	versus	E

Orthogonal (unweighted) contrast coefficients that provide these comparisons are given as rows of the following matrix:

$$\begin{pmatrix} 3 & -2 & -2 & 3 & -2 \\ 0 & 1 & -2 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{pmatrix}.$$

We give the sums of squares for these four contrasts and the F values [see (15.13)] in Table 15.4.

Since these are unweighted contrasts, the contrast sums of squares do not add up to the between sum of squares in Table 15.3. None of the p values is less than .05, so we do not reject $H_0: \sum_i c_i\mu_i = 0$ for any of the four contrasts. In fact, the p values should be less than .05/4 for familywise significance (see the Bonferroni approach in Section 8.5.2), since the overall test in Table 15.3 did not reject $H_0 : \mu_1 = \mu_2 \cdots = \mu_5$.

TABLE 15.4 Sums of Squares and *F* Values for Contrasts for the Fill Data in Table 15.2

Contrast	df	Contrast SS	F	p Value
A, D versus B, C, E	1	.005763	0.75	.406
B, E versus C	1	.002352	0.31	.592
A versus D	1	.034405	4.47	.0582
B versus E	1	.013333	1.73	.215

As an example of two weighted orthogonal contrasts that satisfy $\sum_{i} a_i b_i / n_i$, we keep the first contrast above and replace the second contrast with (0 2 −6 0 4). Then, for these two contrasts, we have

$$\sum_i \frac{a_i b_i}{n_i} = \frac{3(0)}{4} - \frac{2(2)}{2} - \frac{2(-6)}{3} + \frac{3(0)}{3} - \frac{2(4)}{4} = 0.$$

The sums of squares and F values [using (15.13)] for the two contrasts are as follows:

Contrast	df	Contrast SS	F	p Value
A, D versus B, C, E	1	.005763	.75	.406
B, E versus C	1	.005339	.69	.423

□

15.3 TWO-WAY MODEL

The unbalanced two-way model can be expressed as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \tag{15.16}$$

$$= \mu_{ij} + \varepsilon_{ijk}, \tag{15.17}$$

$$i = 1, 2, \ldots, a, \quad j = 1, 2, \ldots, b, \quad k = 1, 2, \ldots, n_{ij}.$$

The ε_{ijk}'s are assumed to be independently distributed as $N(0, \sigma^2)$. In this section we consider the case in which all $n_{ij} > 0$.

The cell means model for analyzing unbalanced two-way data was first proposed by Yates (1934). The cell means model has been advocated by Speed (1969), Urquhart et al. (1973), Nelder (1974), Hocking and Speed (1975), Bryce (1975), Bryce et al. (1976, 1980b), Searle (1977), Speed et al. (1978), Searle et al. (1981), Milliken and Johnson (1984, Chapter 11), and Hocking (1985, 1996). Turner (1990) discusses the relationship between (15.16) and (15.17). In our development we follow Bryce et al. (1980b) and Hocking (1985, 1996).

15.3.1 Unconstrained Model

We first consider the *unconstrained model* in which the μ_{ij}'s are unrestricted. To accommodate a no-interaction model, for example, we must place constraints on the μ_{ij}'s. The constrained model is discussed in Section.

To illustrate the cell means model (15.17), we use $a = 2$ and $b = 3$ with the cell counts n_{ij} given in Figure 15.2. This example with $N = \sum_{ij} n_{ij} = 11$ will be referred to throughout the present section and Section 15.3.2.

$$
\begin{array}{cc|c|c|c|}
 & & & B & \\
 & & 1 & 2 & 3 \\
\cline{3-5}
 & 1 & n_{11}=2 & n_{12}=1 & n_{13}=2 \\
\cline{3-5}
A & 2 & n_{21}=1 & n_{22}=3 & n_{23}=2 \\
\cline{3-5}
\end{array}
$$

Figure 15.2 Cell counts for unbalanced data illustration.

For each of the 11 observations in Figure 15.2, the model $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ is

$$
\begin{aligned}
y_{111} &= \mu_{11} + \varepsilon_{111} \\
y_{112} &= \mu_{11} + \varepsilon_{112} \\
y_{121} &= \mu_{12} + \varepsilon_{121} \\
&\vdots \\
y_{231} &= \mu_{23} + \varepsilon_{231} \\
y_{232} &= \mu_{23} + \varepsilon_{232},
\end{aligned}
$$

or in matrix form

$$\mathbf{y} = \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\varepsilon}, \tag{15.18}$$

where

$$
\mathbf{y} = \begin{pmatrix} y_{111} \\ y_{112} \\ \vdots \\ y_{232} \end{pmatrix}, \quad
\mathbf{W} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix},
$$

$$
\boldsymbol{\mu} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{pmatrix}, \quad
\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \vdots \\ \varepsilon_{232} \end{pmatrix}.
$$

Each row of $\mathbf{W}$ contains a single 1 that corresponds to the appropriate μ_{ij} in $\boldsymbol{\mu}$. For example, the fourth row gives $y_{131} = (001000)\boldsymbol{\mu} + \varepsilon_{131} = \mu_{13} + \varepsilon_{131}$. In this illustration, $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are 11×1, and $\mathbf{W}$ is 11×6. In general, $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $N\times 1$, and $\mathbf{W}$ is $N\times ab$, where $N = \sum_{ij} n_{ij}$.

Since $\mathbf{W}$ is full-rank, we can use the results in Chapters 7 and 8. The analysis is further simplified because $\mathbf{W}'\mathbf{W} = \text{diag}(n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23})$. By (7.6), the least-squares estimator of $\boldsymbol{\mu}$ is given by

$$\hat{\boldsymbol{\mu}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} = \bar{\mathbf{y}}, \tag{15.19}$$

where $\bar{\mathbf{y}} = (\bar{y}_{12.}, \bar{y}_{13.}, \bar{y}_{14.}, \bar{y}_{21.}, \bar{y}_{22.}, \bar{y}_{23.})'$ contains the sample means of the cells, $\bar{y}_{ij.} = \sum_k y_{ijk}/n_{ij}$. By (7.14), the covariance matrix for $\hat{\boldsymbol{\mu}}$ is

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\mu}}) = \sigma^2(\mathbf{W}'\mathbf{W})^{-1} &= \sigma^2 \text{diag}\left(\frac{1}{n_{11}}, \frac{1}{n_{12}}, \cdots, \frac{1}{n_{23}}\right) \\ &= \text{diag}\left(\frac{\sigma^2}{n_{11}}, \frac{\sigma^2}{n_{12}}, \cdots, \frac{\sigma^2}{n_{23}}\right). \end{aligned} \tag{15.20}$$

For general a, b, and N, an unbiased estimator of σ^2 [see (7.23)] is given by

$$s^2 = \frac{\text{SSE}}{\nu_E} = \frac{(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\mu}})'(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\mu}})}{N - ab}, \tag{15.21}$$

where $\nu_E = \sum_{i=1}^{a} \sum_{j=1}^{b} (n_{ij} - 1) = N - ab$, with $N = \sum_{ij} n_{ij}$. In our illustration with $a = 2$ and $b = 3$, we have $N - ab = 11 - 6 = 5$. Two alternative forms of SSE are

$$\text{SSE} = \mathbf{y}'[\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}']\mathbf{y} \quad \text{[see (7.26)]}, \tag{15.22}$$

$$\text{SSE} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2 \quad \text{[see (14.48)]}. \tag{15.23}$$

Using (15.23), we can express s^2 as the pooled estimator

$$s^2 = \frac{\sum_{i=1}^{a} \sum_{j=1}^{b} (n_{ij} - 1)s_{ij}^2}{N - ab}, \tag{15.24}$$

where s_{ij}^2 is the variance estimator in the (ij)th cell, $s_{ij}^2 = \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2/(n_{ij} - 1)$.

The overparameterized model (15.16) includes parameters representing main effects and interactions, but the cell means model (15.17) does not have such parameters. To carry out tests in the cell means model, we use contrasts to express the main effects and the interaction as functions of the μ_{ij}'s in $\boldsymbol{\mu}$. We begin with the main effect of A.

In the vector $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{21}, \mu_{22}, \mu_{23})'$, the first three elements correspond to the first level of A and the last three to the second level, as seen in Figure 15.3. Thus, for the main effect of A, we could compare the average of μ_{11},

		B		
		1	2	3
A	1	μ_{11}	μ_{12}	μ_{13}
	2	μ_{21}	μ_{22}	μ_{23}

Figure 15.3 Cell means corresponding to Figure 15.1.

μ_{12}, and μ_{13} with the average of μ_{21}, μ_{22}, and μ_{23}. The difference between these averages (sums) can be conveniently expressed by the contrast

$$\begin{aligned}\mathbf{a}'\boldsymbol{\mu} &= \mu_{11} + \mu_{12} + \mu_{13} - \mu_{21} - \mu_{22} - \mu_{23}, \\ &= (1,\ 1,\ 1,\ -1,\ -1,\ -1)\boldsymbol{\mu}.\end{aligned}$$

To compare the two levels of A, we can test the hypothesis H_0: $\mathbf{a}'\boldsymbol{\mu} = 0$, which can be written as $H_0 : (\mu_{11} - \mu_{21}) + (\mu_{12} - \mu_{22}) + (\mu_{13} - \mu_{23}) = 0$. In this form, H_0 states that the effect of A averaged (summed) over the levels of B is 0. This corresponds to a common main effect definition in the presence of interaction; see comments following (14.62). Note that this test is not useful in model selection. It simply tests whether the interaction is "symmetric" such that the effect of A, averaged over the levels of B, is zero.

Factor B has three levels corresponding to the three columns of Figure 15.3. In a comparison of three levels, there are 2 degrees of freedom, which will require two contrasts. Suppose that we wish to compare the first level of B with the other two levels and then compare the second level of B with the third. To do this, we compare the average of the means in the first column of Figure 15.3 with the average in the second and third columns and similarly compare the second and third columns. We can make these comparisons using $H_0 : \mathbf{b}_1'\boldsymbol{\mu} = 0$ and $\mathbf{b}_2'\boldsymbol{\mu} = 0$, where $\mathbf{b}_1'\boldsymbol{\mu}$ and $\mathbf{b}_2'\boldsymbol{\mu}$ are the following two orthogonal contrasts:

$$\begin{aligned}\mathbf{b}_1'\boldsymbol{\mu} &= 2(\mu_{11} + \mu_{21}) - (\mu_{12} + \mu_{22}) - (\mu_{13} + \mu_{23}) \qquad (15.25)\\ &= 2\mu_{11} - \mu_{12} - \mu_{13} + 2\mu_{21} - \mu_{22} - \mu_{23} \\ &= (2, -1, -1, 2, -1, -1)\boldsymbol{\mu},\end{aligned}$$

$$\begin{aligned}\mathbf{b}_2'\boldsymbol{\mu} &= (\mu_{12} + \mu_{22}) - (\mu_{13} + \mu_{23}) \qquad (15.26)\\ &= \mu_{12} - \mu_{13} + \mu_{22} - \mu_{23} \\ &= (0, 1,\ -1, 0, 1,\ -1)\boldsymbol{\mu}.\end{aligned}$$

We can combine $\mathbf{b}_1{}'$ and $\mathbf{b}_2{}'$ into the matrix

$$\mathbf{B} = \begin{pmatrix}\mathbf{b}_1' \\ \mathbf{b}_2'\end{pmatrix} = \begin{pmatrix}2 & -1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1\end{pmatrix}, \qquad (15.27)$$

and the hypothesis becomes H_0: $\mathbf{B}\boldsymbol{\mu} = \mathbf{0}$, which, by (15.25) and (15.26), is equivalent to

$$H_0 : \mu_{11} + \mu_{21} = \mu_{12} + \mu_{22} = \mu_{13} + \mu_{23} \tag{15.28}$$

(see Problem15.9). In this form, H_0 states that the interaction is symmetric such that the three levels of B do not differ when averaged over the two levels of A. Note that other orthogonal or linearly independent contrasts besides those in $\mathbf{b}_1{}'$ and $\mathbf{b}_2{}'$ would lead to (15.28) and to the same F statistic in (15.33) below.

By analogy to (14.30), the interaction hypothesis can be written as

$$H_0\colon \mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} = \mu_{13} - \mu_{23},$$

which is a comparison of the "A effects" across the levels of B. If these A effects differ, we have an interaction. We can express the two equalities in H_0 in terms of orthogonal contrasts similar to those in (15.25) and (15.26):

$$\begin{aligned}\mathbf{c}_1'\boldsymbol{\mu} &= 2(\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22}) - (\mu_{13} - \mu_{23}) = 0,\\ \mathbf{c}_2'\boldsymbol{\mu} &= (\mu_{12} - \mu_{22}) - (\mu_{13} - \mu_{23}) = 0.\end{aligned}$$

Thus H_0 can be written as $H_0\colon \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, where

$$\mathbf{C} = \begin{pmatrix}\mathbf{c}_1'\\ \mathbf{c}_2'\end{pmatrix} = \begin{pmatrix}2 & -1 & -1 & -2 & 1 & 1\\ 0 & 1 & -1 & 0 & -1 & 1\end{pmatrix}.$$

Note that $\mathbf{c}_1$ can be found by taking products of corresponding elements of $\mathbf{a}$ and $\mathbf{b}_1$ and $\mathbf{c}_2$ can be obtained similarly from $\mathbf{a}$ and $\mathbf{b}_2$, where $\mathbf{a}$, $\mathbf{b}_1$, and $\mathbf{b}_2$ are the coefficient vectors in $\mathbf{a}'\boldsymbol{\mu}$, $\mathbf{b}_1{}'\boldsymbol{\mu}$ and $\mathbf{b}_2{}'\boldsymbol{\mu}$. Thus

$$\begin{aligned}\mathbf{c}_1' &= [(1)(2), (1)(-1), (1)(-1), (-1)(2), (-1)(-1), (-1)(-1)]\\ &= (2, -1, -1, -2, 1, 1),\\ \mathbf{c}_2' &= [(1)(0), (1)(1), (1)(-1), (-1)(0), (-1)(1), (-1)(-1)]\\ &= (0, 1, -1, 0, -1, 1).\end{aligned}$$

The elementwise multiplication of these two vectors (the *Hadamard product* — see Section 2.2.4) produces interaction contrasts that are orthogonal to each other and to the main effect contrasts.

We now construct tests for the general linear hypotheses H_0: $\mathbf{a}'\boldsymbol{\mu} = 0$, H_0: $\mathbf{B}\boldsymbol{\mu} = \mathbf{0}$, and H_0: $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ for the main effects and interaction. The hypothesis H_0: $\mathbf{a}'\boldsymbol{\mu} = 0$ for the main effect of A, is easily tested using an F statistic similar to (8.38) or (15.12):

$$F = \frac{(\mathbf{a}'\hat{\boldsymbol{\mu}})'[\mathbf{a}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{a}]^{-1}(\mathbf{a}'\hat{\boldsymbol{\mu}})}{s^2} = \frac{\text{SSA}}{\text{SSE}/\nu_E}, \tag{15.29}$$

where s^2 is given by (15.21) and $\nu_E = N - ab$. [For our illustration, $N - ab = 11 - (2)(3) = 5$.]
If H_0 is true, F in (15.29) is distributed as $F(1, N - a_b)$.

The F statistic in (15.29) can be written as

$$F = \frac{(\mathbf{a}'\hat{\boldsymbol{\mu}})^2}{s^2\mathbf{a}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{a}} \tag{15.30}$$

$$= \frac{\left(\sum_{ij} a_{ij}\bar{y}_{ij.}\right)^2}{s^2 \sum_{ij} a_{ij}^2/n_{ij}}, \tag{15.31}$$

which is analogous to (15.13). Since $t^2(\nu_E) = F(1, \nu_E)$ (see Problem 5.16), a t statistic for testing H_0: $\mathbf{a}'\boldsymbol{\mu} = 0$ is given by the square root of (15.30)

$$t = \frac{\mathbf{a}'\hat{\boldsymbol{\mu}}}{s\sqrt{\mathbf{a}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{a}}} = \frac{\mathbf{a}'\hat{\boldsymbol{\mu}} - 0}{\sqrt{\widehat{\text{var}}(\mathbf{a}'\hat{\boldsymbol{\mu}})}}, \tag{15.32}$$

which is distributed as $t(N - ab)$ when H_0 is true. Note that the test based on either of (15.29) or (15.32) is a full–reduced-model test (see Theorem 8.4d) and therefore tests for factor A "above and beyond" (adjusted for) factor B and the interaction.

By Theorem 8.4b, a test statistic for the factor B main effect hypothesis H_0: $\mathbf{B}\boldsymbol{\mu}=\mathbf{0}$ is given by

$$F = \frac{(\mathbf{B}\hat{\boldsymbol{\mu}})'[\mathbf{B}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{B}']^{-1}\mathbf{B}\hat{\boldsymbol{\mu}}/\nu_B}{\text{SSE}/\nu_E} = \frac{\text{SSB}/\nu_B}{\text{SSE}/\nu_E}, \tag{15.33}$$

where $\nu_E = N - ab$ and ν_B is the number of rows of $\mathbf{B}$. (For our illustration, $\nu_E = 5$ and $\nu_B = 2$.) When H_0 is true, F in (15.33) is distributed as $F(\nu_B, \nu_E)$.

A test statistic for the interaction hypothesis H_0: $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ is obtained similarly:

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\mu}})'[\mathbf{C}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{C}']^{-1}\mathbf{C}\hat{\boldsymbol{\mu}}/\nu_{AB}}{\text{SSE}/\nu_E} = \frac{\text{SSAB}/\nu_{AB}}{\text{SSE}/\nu_E}, \tag{15.34}$$

which is distributed as $F(\nu_{AB}, \nu_E)$, where ν_{AB}, the degrees of freedom for interaction, is the number of rows of $\mathbf{C}$. (In our illustration, $\nu_{AB} = 2$.)

Because of the unequal n_{ij}'S, the three sums of squares SSA, SSB, and SSAB do not add to the overall sum of squares for treatments and are not statistically independent, as in the balanced case [see (14.40) and Theorem 14.4b]. Each of SSA, SSB, and SSAB is adjusted for the other effects; that is, the given effect is tested "above and beyond" the others (see Theorem 8.4d).

Example 15.3a. Table 15.5 contains dressing percentages of pigs in a two-way classification (Snedecor and Cochran 1967, p. 480). Let factor A be gender and factor B be breed.

TABLE 15.5 Dressing Percentages (Less 70%) of 75 Swine Classified by Breed and Gender

Breed									
1		2		3		4		5	
Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
13.3	18.2	10.9	14.3	13.6	12.9	11.6	13.8	10.3	12.8
12.6	11.3	3.3	15.3	13.1	14.4	13.2	14.4	10.3	8.4
11.5	14.2	10.5	11.8	4.1		12.6	4.9	10.1	10.6
15.4	15.9	11.6	11.0	10.8		15.2		6.9	13.9
12.7	12.9	15.4	10.9			14.7		13.2	10.0
15.7	15.1	14.4	10.5			12.4		11.0	
13.2		11.6	12.9					12.2	
15.0		14.4	12.5					13.3	
14.3		7.5	13.0					12.9	
16.5		10.8	7.6					9.9	
15.0		10.5	12.9						
13.7		14.5							
		10.9							
		13.0							
		15.9							
		12.8							

We arrange the elements of the vector $\boldsymbol{\mu}$ to correspond to a row of Table 15.5, that is

$$\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, \ldots, \mu_{52})',$$

where the first subscript represents breed and the second subscript is associated with gender.

The vector $\boldsymbol{\mu}$ is 10×1, the matrix $\mathbf{W}$ is 75×10, the vector $\mathbf{a}$ is 10×1, and the matrices $\mathbf{B}$ and $\mathbf{C}$ are each 4×10. We show $\mathbf{a}$, $\mathbf{B}$, and $\mathbf{C}$:

$$\mathbf{a}' = (1, -1, 1, -1, 1, -1, 1, -1, 1, -1),$$

$$\mathbf{B} = \begin{pmatrix} 3 & 3 & 3 & 3 & -2 & -2 & -2 & -2 & -2 & -2 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -2 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 3 & -3 & 3 & -3 & -2 & 2 & -2 & 2 & -2 & 2 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -2 & 2 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

TABLE 15.6 ANOVA for Unconstrained Model

Source	df	Sum of Squares	Mean Square	F	p Value
A (gender)	1	1.984	1.984	0.303	.584
B (breed)	4	90.856	22.714	3.473	.0124
AB	4	24.876	6.219	0.951	.440
Error	65	425.089	6.540		
Total	74	552.095			

(Note that other sets of othogonal contrasts could be used in **B**, and the value of F_B below would be the same.) By (15.19), we obtain

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{y}} = (14.08,\ 14.60,\ 11.75,\ 12.06,\ 10.40,\ 13.65,\ 13.28,\ 11.03,\ 11.01,\ 11.14)'.$$

By (15.22) or (15.23) we obtain $\mathbf{SSE} = 425.08895$, with $\nu_E = 65$. Using (15.29), (15.33), and (15.34), we obtain

$$F_A = .30337, \quad F_B = 3.47318, \quad F_C = .95095.$$

The sums of squares leading to these Fs are given in Table 15.6. Note that the sums of squares for A, B, AB, and error do not add up to the total sum of squares because the data are unbalanced. (These are the type III sums of squares referred to in Section 15.1.) □

15.3.2 Constrained Model

To allow for additivity or other restrictions, constraints on the μ_{ij}'s must be added to the cell means model (15.17) or (15.18). For example, the model

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

cannot represent the no-interaction model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \tag{15.35}$$

unless we specify some relationships among the μ_{ij}'s.

In our 2×3 illustration in Section 15.3.1, the two interaction contrasts are expressible as

$$\mathbf{C}\boldsymbol{\mu} = \begin{pmatrix} 2 & -1 & -1 & -2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix} \boldsymbol{\mu}.$$

If we wish to use a model without interaction, then $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ is not a hypothesis to be tested but an assumption to be included in the statement of the model.

In general, for constraints $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$, the model can be expressed as

$$\mathbf{y} = \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \text{ subject to } \mathbf{G}\boldsymbol{\mu} = \mathbf{0}. \tag{15.36}$$

We now consider estimation and testing in this constrained model. [For the case $\mathbf{G}\boldsymbol{\mu} = \mathbf{h}$, where $\mathbf{h} \neq \mathbf{0}$, see Bryce et al. (1980b).]

To incorporate the constraints $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$ into $\mathbf{y} = \mathbf{W}\boldsymbol{\mu} + \varepsilon$, we can use the Lagrange multiplier method (Section 2.14.3). Alternatively, we can reparameterize the model using the matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix}, \tag{15.37}$$

where $\mathbf{K}$ specifies parameters of interest in the constrained model. For the no-interaction model (15.35), for example, $\mathbf{G}$ would equal $\mathbf{C}$, the first row of $\mathbf{K}$ could correspond to a multiple of the overall mean, and the remaining rows of $\mathbf{K}$ could include the contrasts for the A and B main effects. Thus, we would have

$$\mathbf{K} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 2 & -1 & -1 & 2 & -1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{pmatrix},$$

$$\mathbf{G} = \mathbf{C} = \begin{pmatrix} 2 & -1 & -1 & -2 & 1 & 1 \\ 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}.$$

The second row of $\mathbf{K}$ is $\mathbf{a}'$ and corresponds to the average effect of A. The third and fourth rows are from $\mathbf{B}$ and represent the average B effect.

If the rows of $\mathbf{G}$ are linearly independent of the rows of $\mathbf{K}$, then the matrix $\mathbf{A}$ in (15.37) is of full rank and has an inverse. This holds true in our example, in which we have $\mathbf{G} = \mathbf{C}$. In our example, in fact, the rows of $\mathbf{G}$ are orthogonal to the rows of $\mathbf{K}$. We can therefore insert $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ into (15.36) to obtain the reparameterized model

$$\begin{aligned} \mathbf{y} &= \mathbf{W}\mathbf{A}^{-1}\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad \text{subject to} \quad \mathbf{G}\boldsymbol{\mu} = \mathbf{0} \\ &= \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \qquad\qquad \text{subject to} \quad \mathbf{G}\boldsymbol{\mu} = \mathbf{0}, \end{aligned} \tag{15.38}$$

where $\mathbf{Z} = \mathbf{W}\,\mathbf{A}^{-1}$ and $\boldsymbol{\delta} = \mathbf{A}\boldsymbol{\mu}$.

In the balanced two-way model, we obtained a no-interaction model by simply inserting $\gamma_{ij}^* = 0$ into $y_{ijk} = \mu + \alpha_i^* + \beta_j^* + \gamma_{ij}^* + \varepsilon_{ij}$ [(see 14.37) and (14.38)]. To analogously incorporate the constraint $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$ directly into the model in the unbalanced case, we partition $\boldsymbol{\delta}$ into

$$\boldsymbol{\delta} = \mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix}\boldsymbol{\mu} = \begin{pmatrix} \mathbf{K}\boldsymbol{\mu} \\ \mathbf{G}\boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix}.$$

With a corresponding partitioning on the columns of $\mathbf{Z}$, the model can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} = (\mathbf{Z}_1, \mathbf{Z}_2)\begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix} + \boldsymbol{\varepsilon} \\ &= \mathbf{Z}_1\boldsymbol{\delta}_1 + \mathbf{Z}_2\boldsymbol{\delta}_2 + \boldsymbol{\varepsilon} \quad \text{subject to} \quad \mathbf{G}\boldsymbol{\mu} = \mathbf{0}. \end{aligned} \tag{15.39}$$

Since $\delta_2 = \mathbf{G}\boldsymbol{\mu}$, the constraint $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$ gives $\boldsymbol{\delta}_2 = \mathbf{0}$ and the constrained model in (15.39) simplifies to

$$\mathbf{y} = \mathbf{Z}_1\boldsymbol{\delta}_1 + \boldsymbol{\varepsilon}. \tag{15.40}$$

An estimator of $\boldsymbol{\delta}_1$ [see (7.6)] is given by

$$\hat{\boldsymbol{\delta}}_1 = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{y}.$$

To obtain an expression for $\boldsymbol{\mu}$ subject to the constraints, we multiply

$$\mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ \mathbf{0} \end{pmatrix}$$

by

$$\mathbf{A}^{-1} = (\mathbf{K}^*, \mathbf{G}^*).$$

If the rows of $\mathbf{G}$ are orthogonal to the rows of $\mathbf{K}$, then

$$(\mathbf{K}^*, \mathbf{G}^*) = [\mathbf{K}'(\mathbf{K}\mathbf{K}')^{-1}, \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}] \tag{15.41}$$

(see Problem15.13). If the rows of $\mathbf{G}$ are linearly independent of (but not necessarily orthogonal to) the rows of $\mathbf{K}$, we obtain

$$\mathbf{K}^* = \mathbf{H}_G\mathbf{K}'(\mathbf{K}\mathbf{H}_G\mathbf{K}')^{-1}, \tag{15.42}$$

where

$$\mathbf{H}_G = \mathbf{I} - \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{G},$$

and $\mathbf{G}^*$ is similarly defined (see Problem15.14). In any case, we denote the product of $\mathbf{K}^*$ and $\boldsymbol{\delta}_1$ by $\boldsymbol{\mu}_c$:

$$\boldsymbol{\mu}_c = \mathbf{K}^*\boldsymbol{\delta}_1.$$

We estimate $\boldsymbol{\mu}_c$ by

$$\hat{\boldsymbol{\mu}}_c = \mathbf{K}^*\hat{\boldsymbol{\delta}}_1 = \mathbf{K}^*(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{y}, \tag{15.43}$$

which has covariance matrix

$$\mathrm{cov}(\hat{\boldsymbol{\mu}}_c) = \sigma^2 \mathbf{K}^*(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{K}^{*\prime}. \tag{15.44}$$

To test for factor B in the constrained model, the hypothesis is H_0: $\mathbf{B}\boldsymbol{\mu}_c = \mathbf{0}$. The covariance matrix of $\mathbf{B}\hat{\mu}_c$ is obtained from (3.44) and (15.44) as

$$\mathrm{cov}(\mathbf{B}\hat{\boldsymbol{\mu}}_c) = \sigma^2 \mathbf{B}\mathbf{K}^*(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{K}^{*\prime}\mathbf{B}'.$$

Then, by Theorem 8.4b, the test statistic for H_0: $\mathbf{B}\boldsymbol{\mu}_c = \mathbf{0}$ in the constrained model becomes

$$F = \frac{(\mathbf{B}\hat{\boldsymbol{\mu}}_c)'[\mathbf{B}\mathbf{K}^*(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{K}^{*\prime}\mathbf{B}']^{-1}\mathbf{B}\hat{\boldsymbol{\mu}}_c/\nu_B}{\mathrm{SSE}_c/\nu_{E_c}}, \tag{15.45}$$

where SSE_c (subject to $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$) is obtained using $\hat{\mu}_c$ [from (15.43)] in (15.21). (In our example, where $\mathbf{G} = \mathbf{C}$ for interaction, SSE_c effectively pools SSE and SSAB from the unconstrained model.) The degrees of freedom ν_{E_c} is obtained as $\nu_{E_c} = \nu_E + \mathrm{rank}(\mathbf{G})$, where $\nu_E = N - ab$ is for the unconstrained model, as defined following (15.21). [In our example, $\mathrm{rank}(\mathbf{G}) = 2$ since there are 2 degrees of freedom for SSAB.] We reject H_0: $\mathbf{B}\boldsymbol{\mu}_c = \mathbf{0}$ if $F \geq F_{\alpha,\nu_B,\nu_{E_c}}$, where F_α is the upper α percentage point of the central F distribution.

For H_0: $\mathbf{a}'\boldsymbol{\mu}_c = 0$, the F statistic becomes

$$F = \frac{(\mathbf{a}'\hat{\boldsymbol{\mu}}_c)'[\mathbf{a}'\mathbf{K}^*(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{K}^{*\prime}\mathbf{a}]^{-1}(\mathbf{a}'\hat{\boldsymbol{\mu}}_c)}{\mathrm{SSE}_c/\nu_{E_c}}, \tag{15.46}$$

which is distributed as $F(1, \nu_{E_c})$ if H_0 is true.

Example 15.3b. For the pigs data in Table 15.5, we test for factors A and B in a no-interaction model, where factor A is gender and factor B is breed. The matrix $\mathbf{G}$ is the same as $\mathbf{C}$ in Example 15.3a. For $\mathbf{K}$ we have

$$\mathbf{K} = \begin{pmatrix} \mathbf{j}' \\ \mathbf{a}' \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 3 & 3 & 3 & 3 & -2 & -2 & -2 & -2 & -2 & -2 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -2 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 \end{pmatrix}$$

By (15.43), we obtain

$$\hat{\boldsymbol{\mu}}_c = (14.16, 14.42, 11.77, 12.03, 11.40, 11.65, 12.45, 12.70, 10.97, 11.22).'$$

TABLE 15.7 ANOVA for Constrained Model

Source	df	Sum of Squares	Mean Square	F	p Value
A (gender)	1	1.132	1.132	0.17	.678
B (breed)	4	101.418	25.355	3.89	.00660
Error	69	449.965	6.521		
Total	74	552.0955			

For SSE_c, we use $\hat{\boldsymbol{\mu}}_c$ in place of $\hat{\boldsymbol{\mu}}$ in (15.21) to obtain $\mathrm{SSE}_c = 449.96508$. For ν_{E_c}, we have

$$\nu_{E_c} = \nu_E + \mathrm{rank}(\mathbf{G}) = 65 + 4 = 69.$$

Then by (15.45), we obtain $F_{B_c} = 3.8880003$. The sums of squares leading to F_{B_c} and F_{A_c} are given in Table 15.7. □

15.4 TWO-WAY MODEL WITH EMPTY CELLS

Possibly the greatest advantage of the cell means model in the analysis of unbalanced data is that extreme situations such as empty cells can be dealt with relatively easily. The cell means approach allows one to deal specifically with nonestimability problems arising from the empty cells (as contrasted with nonestimability arising from overparameterization of the model). Much of our discussion here follows that of Bryce et al. (1980a).

Consider the unbalanced two-way model in (15.17), but allow n_{ij} to be equal to 0 for one or more (say m) *isolated* cells; that is, the empty cells do not constitute a whole row or whole column. Assume also that the empty cells are *missing at random* (Little and Rubin 2002, p. 12); that is, the emptiness of the cells is independent of the values that would be observed in those cells.

In the empty cells model, $\mathbf{W}$ is non-full-rank in that it has m columns equal to $\mathbf{0}$. To simplify notation, assume that the columns of $\mathbf{W}$ have been rearranged with the columns of $\mathbf{0}$ occurring last. Hence

$$\mathbf{W} = (\mathbf{W}_1, \mathbf{O}),$$

where $\mathbf{W}_1$ is an $n \times (ab - m)$ matrix and $\mathbf{O}$ is $n \times m$. Correspondingly

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu}_e \end{pmatrix},$$

where $\boldsymbol{\mu}_o$ is the vector of cell means for the occupied cells while $\boldsymbol{\mu}_e$ is the vector of cell means for the empty cells. The model is thus the non-full-rank model

$$\mathbf{y} = (\mathbf{W}_1, \mathbf{O})\begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu}_e \end{pmatrix} + \boldsymbol{\varepsilon}. \tag{15.47}$$

The first task in the analysis of two-way data with empty cells is to test for the interaction between the factors A and B. To test for the interaction when there are isolated empty cells, care must be exercised to ensure that a testable hypothesis is being tested (Section 12.6). The full–reduced-model approach [see (8.31)] is useful here. A sensible full model is the unconstrained cell means model in (15.47). Even though $\mathbf{W}$ is not full-rank

$$\mathrm{SSE}_u = \mathbf{y}'[\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}']\mathbf{y} \tag{15.48}$$

is invariant to the choice of a generalized inverse (Theorem 12.3e). The reduced model is the additive model, given by

$$\mathbf{y} = \mathbf{W}\mathbf{A}^{-1}\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad \text{subject to} \quad \mathbf{G}\boldsymbol{\mu} = \mathbf{0},$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix},$$

in which $\mathbf{K}$ is a matrix specifying the overall mean and linearly independent main effect contrasts for factors A and B, and the rows of $\mathbf{G}$ are linearly independent interaction contrasts (see Section 15.3.2) such that $\mathbf{A}$ is full-rank. We define $\mathbf{Z}_1$ as $\mathbf{W}\mathbf{K}^*$ [see (15.41)]. Because the empty cells are isolated, $\mathbf{Z}_1$ is full-rank even though some of the constraints in $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$ are nonestimable. The error sum of squares for the additive model is then

$$\mathrm{SSE}_a = \mathbf{y}'[\mathbf{I} - \mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1']\mathbf{y}, \tag{15.49}$$

and the test statistic for the interaction is

$$F = \frac{(\mathrm{SSE}_a - \mathrm{SSE}_u)/[(a-1)(b-1)-m]}{\mathrm{SSE}_u/(n-ab+m)}. \tag{15.50}$$

Equivalently the interaction could be tested by the general linear hypothesis approach in (8.27). However, a maximal set of nonestimable interaction side conditions involving $\boldsymbol{\mu}_e$ must first be imposed on the model. For example, the side conditions could be specified as

$$\mathbf{T}\boldsymbol{\mu} = \mathbf{0}, \tag{15.51}$$

where $\mathbf{T}$ is an $m \times ab$ matrix with rows corresponding to the contrasts $\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}$ for all m empty cells (Henderson and McAllister 1978). Using (12.37), we obtain

$$\hat{\boldsymbol{\mu}} = (\mathbf{W}'\mathbf{W} + \mathbf{T}'\mathbf{T})^{-1}\mathbf{W}'\mathbf{y} \tag{15.52}$$

and

$$\operatorname{cov}(\hat{\boldsymbol{\mu}}) = \sigma^2(\mathbf{W}'\mathbf{W} + \mathbf{T}'\mathbf{T})^{-1}\mathbf{W}'\mathbf{W}(\mathbf{W}'\mathbf{W} + \mathbf{T}'\mathbf{T})^{-1}. \tag{15.53}$$

The interaction can then be tested using the general linear hypothesis test of H_0: $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ where $\mathbf{C}$ is the full matrix of $(a-1)(b-1)$ interaction contrasts. Even though some of the rows of $\mathbf{C}\boldsymbol{\mu}$ are not estimable, the test statistic can be computed using a generalized inverse in the numerator as

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\mu}})'\{\mathbf{C}[\operatorname{cov}(\hat{\boldsymbol{\mu}})/\sigma^2]\mathbf{C}'\}^{-}(\mathbf{C}\hat{\boldsymbol{\mu}})/[(a-1)(b-1)-m]}{\mathrm{SSE}/(n-ab+m)}. \tag{15.54}$$

The error sum of squares for this model, SSE, turns out to be the same as SSE_u in (15.48). By Theorem 2.8c(v), the numerator of this F statistic is invariant to the choice of a generalized inverse (Problem 15.16).

Both versions of this additivity test involve the unverifiable assumption that the means of the empty cells follow the additive pattern displayed by the means of the occupied cells. If there are relatively few empty cells, this is usually a reasonable assumption.

If the interaction is not significant and is deemed to be negligible, the additive model can be used as in Section 15.3.2 without any modifications. The isolated empty cells present no problems for the use of the additive model.

If the interaction is significant, it may be possible to partially constrain the interaction in an attempt to render all cell means (including those in $\boldsymbol{\mu}_e$) estimable. This is not always possible, because it requires a set of constraints that are both a priori reasonable and such that they render $\boldsymbol{\mu}$ estimable. Nonetheless, it is often advisable to make this attempt because no new theoretical results are needed. The greatest challenges are practical, in that sensible constraints must be used. Many constraints will do the job mathematically, but the results are meaningless unless the constraints are reasonable. Unlike many other methods associated with linear models, the validity of this procedure depends on the parameterization of the model and the specific constraints that are chosen.

We proceed in this attempt by proposing partial interaction constraints

$$\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$$

for the empty cells model in (15.47). We choose $\mathbf{K}$ such that its rows are linearly independent of the rows of $\mathbf{G}$ so that

$$\mathbf{A} = \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix}$$

is nonsingular. Thus $\mathbf{A}^{-1} = (\mathbf{K}^* \ \mathbf{G}^*)$ as in the comments following (15.41). Suppose that the constraints are realistic, and that they are such that the constrained model is not the additive model; that is, at least a portion of the interaction is unconstrained. Then, if $\mathbf{Z}_1 = \mathbf{W}\mathbf{K}^*$ is full-rank, all the cell means (including $\boldsymbol{\mu}_e$) can be estimated as

$$\hat{\boldsymbol{\mu}} = \mathbf{K}^*(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{y}, \tag{15.55}$$

and cov($\hat{\boldsymbol{\mu}}$) is given by (15.44). Further inferences about linear combinations of the cell means can then be readily carried out. If $\mathbf{Z}_1$ is not full-rank, care must be exercised to ensure that only estimable functions of $\boldsymbol{\mu}$ are estimated and that testable hypotheses involving $\boldsymbol{\mu}$ are tested (see Section 12.2).

A simple way to quickly check whether $\mathbf{Z}_1$ is full-rank (and thus all cell means are estimable) is given in the following theorem.

Theorem 15.4. Consider the constrained empty cells model in (15.47) with m empty cells. Partition $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix} = \begin{pmatrix} \mathbf{K}_1 & \mathbf{K}_2 \\ \mathbf{G}_1 & \mathbf{G}_2 \end{pmatrix}$$

conformal with the partitioned vector

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu}_e \end{pmatrix}.$$

The elements of $\boldsymbol{\mu}$ are estimable (equivalently $\mathbf{Z}_1$ is full-rank) if and only if rank($\mathbf{G}_2$) $= m$.

Proof. We prove this theorem for the special case in which $\mathbf{G}$ has m rows so that $\mathbf{G}_2$ is $m \times m$. We partition $\mathbf{A}^{-1}$ as

$$\begin{pmatrix} \mathbf{K}_1^* & \mathbf{G}_1^* \\ \mathbf{K}_2^* & \mathbf{G}_2{}^* \end{pmatrix},$$

with submatrices conforming to the partitioning of $\mathbf{A}$. Then

$$\mathbf{Z}_1 = (\mathbf{W}_1, \mathbf{O})\begin{pmatrix} \mathbf{K}_1^* \\ \mathbf{K}_2^* \end{pmatrix} = \mathbf{W}_1\mathbf{K}_1^*.$$

Since $\mathbf{W}_1$ is full-rank and each of its rows consists of one 1 with several 0s, $\mathbf{W}_1\mathbf{K}_1^*$ contains one or more copies of all of the rows of $\mathbf{W}_1$. Thus rank($\mathbf{Z}_1$) = rank($\mathbf{K}_1^*$). Since $\mathbf{A}^{-1}$ is nonsingular, $\mathbf{K}_1^{*-1}$ exists if $\mathbf{K}_1^*$ is full rank. If so, the product

$$\begin{pmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{K}_2^*\mathbf{K_1}^{*-1} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{K}_1^* & \mathbf{G}_1^* \\ \mathbf{K}_2^* & \mathbf{G}_2^* \end{pmatrix} = \begin{pmatrix} \mathbf{K}_1^* & \mathbf{G}_1^* \\ \mathbf{O} & \mathbf{G}_2^* - \mathbf{K}_2^*\mathbf{K}_1^{*-1}\mathbf{G}_1^* \end{pmatrix}$$

is defined and is nonsingular by Theorem 2.4(ii). By Corollary 1 to Theorem 2.9b, $\mathbf{G}_2^* - \mathbf{K}_2^*\mathbf{K}_1^{*-1}\mathbf{G}_1^*$ is also nonsingular. But by equation (2.50), $(\mathbf{G}_2^* - \mathbf{K}_2^*\mathbf{K}_1^{*-1}\mathbf{G}_1^*)^{-1} = \mathbf{G}_2$. Thus, if $\mathbf{A}^{-1}$ is nonsingular, nonsingularity of $\mathbf{K}_1^*$ implies nonsingularity of $\mathbf{G}_2$. Analogous reasoning leads to the converse. Thus $\mathbf{K}_1^*$ is full-rank if and only if $\mathbf{G}_2$ is full-rank. Furthermore, $\mathbf{Z}_1$ is full-rank if and only if rank($\mathbf{G}_2$) $= m$. □

Example 15.4a. For the second-language data of Table 15.8, we test for the interaction of native language and gender. There are two empty cells, and thus $\mathbf{W}$ is a

TABLE 15.8 Comfort in Using English as a Second Language for Students at BYU-Hawaii[a]

Native Language	Gender: Male	Gender: Female
Samoan	24	28
	3.20	3.38
	0.66	0.68
Tongan	25	39
	3.03	3.10
	0.69	0.61
Hawaiian	4	2
	3.47	3.13
	0.68	0.47
Fijian	1	—
	3.79	—
	—	—
Pacific Islands English	26	49
	3.71	3.13
	0.58	0.73
Maori	3	1
	4.07	3.04
	0.061	—
Mandarin	15	43
	3.33	3.14
	0.74	0.61
Cantonese	—	21
	—	3.00
	—	0.54

[a] Brigham Young University–Hawaii; data classified by gender and native language. *Key to table entries*: number of observations, mean, and standard deviation.

281×16 matrix with two columns of $\mathbf{0}$. For the unconstrained model we use (15.48) to obtain

$$\mathrm{SSE}_u = 113.235.$$

Numbering the cells of Table 15.8 from 1 to 8 for the first column and from 9 to 16 for the second column, we now define

$$\mathbf{A} = \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix} \tag{15.56}$$

where

$$\mathrm{K} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 2 & 2 & -1 & -1 & -1 & -1 & 0 & 0 & 2 & 2 & -1 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -3 & -3 & 1 & 1 & 1 & 1 & 1 & 1 & -3 & -3 \end{pmatrix}$$

and

$$\mathbf{G} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 2 & 2 & -1 & -1 & -1 & -1 & 0 & 0 & -2 & -2 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -3 & -3 & -1 & -1 & -1 & -1 & -1 & -1 & 3 & 3 \end{pmatrix}$$

The overall mean and main effect contrasts are specified by **K** while interaction contrasts are specified by **G**. Using (15.49), $\mathrm{SSE}_a = 119.213$. The full–reduced F test for additivity (15.50) yields the test statistic

$$F = \frac{(119.213 - 113.235)/5}{119.213/267} = 2.82,$$

which is larger than the critical value of $F_{.05,\,5,\,267} = 2.25$.

As an alternative approach to testing additivity, we impose the nonestimable side conditions $\mu_{8,1} - \mu_{8.} - \mu_{.1} + \mu_{..} = 0$ and $\mu_{4,\,2} - \mu_{4.} - \mu_{.2} + \mu_{..} = 0$ on the model by setting

$$\mathbf{T} = \begin{pmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & 7 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -7 \\ -1 & -1 & -1 & 7 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & -7 & 1 & 1 & 1 & 1 \end{pmatrix}$$

in (15.51) and

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

in (15.54). The F statistic for the general linear hypothesis test of additivity (15.54) is again equal to 2.82.

Since the interaction is significant for this dataset, we partially constrain the interaction with contextually sensible estimable constraints in an effort to make all of the cell means estimable. We use $\mathbf{A}$ as defined in (15.56), but repartition it so that

$$\mathbf{K} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 2 & 2 & -1 & -1 & -1 & -1 & 0 & 0 & 2 & 2 & -1 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -3 & -3 & 1 & 1 & 1 & 1 & 1 & 1 & -3 & -3 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 2 & 2 & -1 & -1 & -1 & -1 & 0 & 0 & -2 & -2 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & -3 & -3 & -1 & -1 & -1 & -1 & -1 & -1 & 3 & 3 \end{pmatrix}$$

and

$$\mathbf{G} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The partial interaction constraints specified by $\mathbf{G}\boldsymbol{\mu} = \mathbf{0}$ seem sensible in that they specify that the male–female difference is the same for Samoan and Tongan speakers, for Fijian and Hawaiian speakers, and for Mandarin and Cantonese speakers. Because the empty cells are the eighth and twelfth cells, we have

$$\mathbf{G}_2 = \begin{pmatrix} 0 & 0 \\ 0 & -1 \\ -1 & 0 \end{pmatrix}$$

which obviously has rank = 2. Thus, by Theorem 15.4, all the cell means are estimable. Using (15.55) to compute the constrained estimates and (15.44) to compute their standard errors, we obtain the results in Table 15.9. □

TABLE 15.9 Estimated Mean Comfort in Using English as Second Language (with Standard Error) for Students at BYU-Hawaii[a]

Native Language	Gender	
	Male	Female
Samoan	3.23 (.11)	3.35 (.11)
Tongan	3.00 (.11)	3.12 (.09)
Hawaiian	3.47 (.33)	3.13 (.46)
Fijian	3.79 (.65)	3.20 (.67)
Pacific Islands English	3.71 (.03)	3.13 (.09)
Maori	4.07 (.38)	3.04 (.65)
Mandarin	3.33 (.17)	3.14 (.10)
Cantonese	3.19 (.24)	3.00 (.14)

[a] On the basis of a constrained empty-cells model.

PROBLEMS

15.1 For the model $\mathbf{y} = \mathbf{W}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$ in (15.2.1), find $\mathbf{W}'\mathbf{W}$ and $\mathbf{W}'\mathbf{y}$ and show that $(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} = \bar{\mathbf{y}}$ as in (15.5).

15.2 **(a)** Show that for the reduced model $y_{ij} = \mu + \varepsilon_{ij}^*$ in Section 15.3, $\text{SS}(\mu) = N\bar{y}_{..}^2$ as used in (15.6).

(b) Show that $\text{SSB} = \sum_{i=1}^k \bar{y}_{i.}y_{i.} - N\bar{y}_{..}^2$ as in (15.6).

(c) Show that (15.6) is equal to (15.7), that is, $\text{SSB} = \sum_i \bar{y}_{i.}y_{i.} - N\bar{y}_{..}^2 = \sum_i y_{i.}^2/n_i - y_{..}^2/N$.

15.3 **(a)** Show that SSB in (15.9) is equal to SSB in (15.7), that is, $\sum_{i=1}^k n_i(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^k y_{i.}^2/n_i - y_{..}^2/N$.

(b) Show that SSE in (15.10) is equal to SSE in (15.8), that is, $\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^k y_{i.}^2/n_i$.

15.4 Show that $F = (\Sigma_i c_i \bar{y}_{i.})^2/(s^2 \sum_i c_i^2/n_i)$ in (15.13) follows from (15.12).

15.5 Show that $\mathbf{a}'$ and $\mathbf{b}'$ in (15.15) provide contrast coefficients that satisfy the property $\sum_i a_i b_i/n_i = 0$.

15.6 Show that $\hat{\boldsymbol{\mu}} = \bar{\mathbf{y}}$ as in (15.19).

15.7 Obtain (15.23) from (15.21); that is, show that $(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\mu}})'(\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\mu}}) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2$.

15.8 Obtain (15.24) from (15.23); that is, show that $\sum_{k=1}^{n_{ij}} (y_{ijk} - \bar{y}_{ij.})^2 = (n_{ij} - 1)s_{ij}^2$.

15.9 Show that H_0: $\mathbf{B}\boldsymbol{\mu} = \mathbf{0}$, where $\mathbf{B}$ is given in (15.27), is equivalent to H_0: $\mu_{11} + \mu_{21} = \mu_{12} + \mu_{22} = \mu_{13} + \mu_{23}$ in (15.28).

15.10 Obtain $F = \left(\sum_{ij} a_{ij}\bar{y}_{ij.}\right)^2 / \left(s^2 \sum_{ij} a_{ij}^2/n_{ij}\right)$ in (15.31) from $F = (\mathbf{a}'\hat{\boldsymbol{\mu}})^2/[s^2\mathbf{a}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{a}]$ in (15.30).

15.11 Evaluate $\mathbf{a}'(\mathbf{W}'\mathbf{W})^{-1}\mathbf{a}$ in (15.29) or (15.30) for $\mathbf{a}' = (1, 1, 1, -1, -1, -1)$. Use the $\mathbf{W}$ matrix for the 11 observations in the illustration in Section 15.3.1.

15.12 Evaluate $\mathbf{B}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{B}'$ in (15.33) for the matrices $\mathbf{B}$ and $\mathbf{W}$ used in the illustration in Section 15.3.1.

15.13 Show that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, where $\mathbf{A} = \begin{pmatrix} \mathbf{K} \\ \mathbf{G} \end{pmatrix}$ as in (15.37) and $\mathbf{A}^{-1} = [\mathbf{K}'(\mathbf{K}\mathbf{K}')^{-1}, \mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}]$ as in (15.41).

15.14 Obtain $\mathbf{G}^*$ analogous to $\mathbf{K}^*$ in (15.42).

15.15 Show that $\text{cov}(\hat{\boldsymbol{\mu}}_c) = \sigma^2\mathbf{K}'(\mathbf{K}\mathbf{K}')^{-1}(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}(\mathbf{K}\mathbf{K}')^{-1}\mathbf{K}$, thus verifying (15.44).

15.16 Show that the numerator of the F statistic in (15.54) is invariant to the choice of a generalized inverse.

15.17 In a feeding trial, chicks were given five protein supplements. Their final weights at 6 weeks are given in Table 15.10 (Snedecor 1948, p. 214).

(a) Calculate the sums of squares in Table 15.1 and the F statistic in (15.11).

(b) Compare the protein supplements using (unweighted) orthogonal contrasts whose coefficients are the rows in the matrix

$$\begin{pmatrix} 3 & -2 & -2 & -2 & 3 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Thus we are making the following comparisons:

L, C	versus	So, Su, M
So, M	versus	Su
So	versus	M
L	versus	C

(c) Replace the second contrast with a weighted contrast whose coefficients satisfy $\sum_i a_i b_i / n_i = 0$ when paired with the first contrast. Find sums of squares and F statistics for these two contrasts.

TABLE 15.10 Final Weights (g) of Chicks at 6 Weeks

Protein Supplement				
Linseed	Soybean	Sunflower	Meat	Casein
309	243	423	325	368
229	230	340	257	390
181	248	392	303	379
141	327	339	315	260
260	329	341	380	404
203	250	226	153	318
148	193	320	263	352
169	271	295	242	359
213	316	334	206	216
257	267	322	344	222
244	199	297	258	283
271	177	318		332
	158			
	248			

15.18 **(a)** Carry out the computations to obtain $\hat{\mu}$, **SSE**, F_{A}, F_{B}, and F_{C} in Example 15.3a.

(b) Carry out the computations to obtain $\hat{\boldsymbol{\mu}}_c$, SSE_c, F_{A_c}, and F_{B_c} in Example 15.3b.

(c) Carry out the tests in parts (a) and (b) using a software package such as SAS GLM.

15.19 Table 15.11 lists weight gains of male rats under three types of feed and two levels of protein.

(a) Let factor A be level of protein and factor B be type of feed. Define a vector $\mathbf{a}$ corresponding to factor A and matrices $\mathbf{B}$ and $\mathbf{C}$ for factor B and interaction AB, respectively, as in Section 15.3.1. Use these to construct general linear hypothesis tests for main effects and interaction as in (15.29), (15.33), and (15.34).

(b) Test the main effects in the no-interaction model (15.35) using the constrained model (15.36). Define $\mathbf{K}$ and $\mathbf{G}$ and find $\hat{\boldsymbol{\mu}}_c$ in (15.43), $\mathrm{SSE}_{\mathrm{c}}$, and F for H_0: $\mathbf{a}'\boldsymbol{\mu}_c = 0$ and H_0: $\mathbf{B}\boldsymbol{\mu}_c = \mathbf{0}$ in (15.45).

(c) Carry out the tests in parts (a) and (b) using a software package such as SAS GLM.

15.20 Table 15.12 lists yields when five varieties of plants and four fertilizers were tested. Test for main effects and interaction.

TABLE 15.11 Weight Gains (g) of Rats under Six Diet Combinations

High Protein			Low Protein		
Beef	Cereal	Pork	Beef	Cereal	Pork
73	98	94	90	107	49
102	74	79	76	95	82
118	56	96	90	97	73
104	111	98	64	80	86
81	95	102	86	98	81
107	88	102	51	74	97
100	82		72		106
87	77		90		
	86		95		
	92		78		

Source: Snedecor and Cochran (1967, p. 347).

TABLE 15.12 Yield from Five Varieties of Plants Treated with Four Fertilizers

	Variety				
Fertilizer	1	2	3	4	5
1	57	26	39	23	48
	46	38	—	36	35
	—	20	—	18	—
2	67	44	57	74	61
	72	68	61	47	—
	66	64	—	69	—
3	95	92	91	98	78
	90	89	82	85	89
	89	—	—	—	95
4	92	96	98	99	99
	88	95	93	90	—
	—	—	98	98	—

Source: Ostle and Mensing (1975, p. 368).