Chapter 7
Multifactor Analysis of Variance

Chapter 7 presents the analysis of multifactor ANOVA models. The first three sections deal with the balanced two-way ANOVA model. Section 1 examines the no interaction model. Section 2 examines the model with interaction. Section 3 discusses the relationship between polynomial regression and the balanced two-way ANOVA model. Sections 4 and 5 discuss unbalanced two-way ANOVA models. Section 4 treats the special case of proportional numbers. Section 5 examines the general case. Finally, Section 6 extends the earlier results of the chapter to models with more than two factors. A review of the tensor concepts in Appendix B may aid the reader of this chapter.

7.1 Balanced Two-Way ANOVA Without Interaction

The balanced two-way ANOVA without interaction model is generally written

$$y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}, \tag{1}$$

$i = 1,\ldots,a,\ j = 1,\ldots,b,\ k = 1,\ldots,N.$

R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-9816-3_7,

EXAMPLE 7.1.1. Suppose $a = 3$, $b = 2$, $N = 4$. In matrix terms write

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{124} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{222} \\ y_{223} \\ y_{224} \\ y_{311} \\ y_{312} \\ y_{313} \\ y_{314} \\ y_{321} \\ y_{322} \\ y_{323} \\ y_{324} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \eta_1 \\ \eta_2 \end{bmatrix} + e.$$

In general, we can write the model as

$$Y = [X_0, X_1, \ldots, X_a, X_{a+1}, \ldots, X_{a+b}] \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \\ \eta_1 \\ \vdots \\ \eta_b \end{bmatrix} + e.$$

Write $n = abN$ and the observation vector as $Y = [y_{ijk}]$, where the three subscripts i, j, and k denote a row of the vector. (Multiple subscripts denoting the rows and columns of matrices were introduced in Chapter 4.) The model matrix of the balanced two-way ANOVA model has

$$\begin{aligned} X_0 &= J, \\ X_r &= [t_{ijk}], \qquad t_{ijk} = \delta_{ir}, \qquad r = 1, \ldots, a, \\ X_{a+s} &= [t_{ijk}], \qquad t_{ijk} = \delta_{js}, \qquad s = 1, \ldots, b, \end{aligned}$$

where $\delta_{gh} = 1$ if $g = h$, and 0 otherwise. This is just a formal way of writing down model matrices that look like the one in Example 7.1.1. For example, an observation y_{rst} is subject to the effects of α_r and η_s. The rst row of the column X_r needs to be 1 so that α_r is added to y_{rst} and the rst row of X_{a+s} needs to be 1 so that η_r is added to y_{rst}. The rst rows of the columns X_j, $j = 1,\ldots,a$, $j \neq r$ and X_{a+j}, $j = 1,\ldots,b$, $j \neq s$ need to be 0 so that none of the α_js other than α_r, nor η_js other than η_s, are added to y_{rst}. The definition of the columns X_r and X_{a+s} given above ensures that this occurs.

The analysis of this model is based on doing two separate one-way ANOVAs. It is frequently said that this can be done because the treatments α and η are orthogonal. This is true in the sense that after fitting μ, the column space for the αs is orthogonal to the column space for the ηs. (See the discussion surrounding Proposition 3.6.3.) To investigate this further, consider a new matrix

$$Z = [Z_0, Z_1, \ldots, Z_a, Z_{a+1}, \ldots, Z_{a+b}],$$

where

$$Z_0 = X_0 = J$$

and

$$Z_r = X_r - \frac{X_r'J}{J'J}J,$$

for $r = 1,\ldots,a+b$. Here we have used Gram–Schmidt to eliminate the effect of J (the column associated with μ) from the rest of the columns of X. Since $J'J = abN$, $X_r'J = bN$ for $r = 1,\ldots,a$, and $X_{a+s}'J = aN$ for $s = 1,\ldots,b$, we have

$$Z_r = X_r - \frac{1}{a}J, \quad r = 1,\ldots,a,$$

$$Z_{a+s} = X_{a+s} - \frac{1}{b}J, \quad s = 1,\ldots,b.$$

Observe that

$$\begin{aligned} C(X) = C(X_0, X_1, \ldots, X_a, X_{a+1}, \ldots, X_{a+b}) \\ = C(Z_0, Z_1, \ldots, Z_a, Z_{a+1}, \ldots, Z_{a+b}) = C(Z), \end{aligned}$$

$$C(X_0, X_1, \ldots, X_a) = C(Z_0, Z_1, \ldots, Z_a),$$

$$C(X_0, X_{a+1}, \ldots, X_{a+b}) = C(Z_0, Z_{a+1}, \ldots, Z_{a+b}),$$

$$Z_0 \perp Z_r, \quad r = 1,\ldots,a+b,$$

and

$$C(Z_1, \ldots, Z_a) \perp C(Z_{a+1}, \ldots, Z_{a+b}).$$

To see the last of these, observe that for $r = 1,\ldots,a$ and $s = 1,\ldots,b$,

$$Z_{a+s}'Z_r = \sum_{ijk}(\delta_{js} - 1/b)(\delta_{ir} - 1/a)$$

$$
\begin{aligned}
&= \sum_{ijk} \delta_{js}\delta_{ir} - \sum_{ijk} \delta_{js}\frac{1}{a} - \sum_{ijk} \delta_{ir}\frac{1}{b} + \sum_{ijk}\frac{1}{ab} \\
&= \sum_{ij} N\delta_{js}\delta_{ir} - \sum_{j} \delta_{js}\frac{aN}{a} - \sum_{i} \delta_{ir}\frac{bN}{b} + \frac{abN}{ab} \\
&= N - aN/a - bN/b + N = 0.
\end{aligned}
$$

We have decomposed $C(X)$ into three orthogonal parts, $C(Z_0)$, $C(Z_1,\ldots,Z_a)$, and $C(Z_{a+1},\ldots,Z_{a+b})$. M, the perpendicular projection operator onto $C(X)$, can be written as the matrix sum of the perpendicular projection matrices onto these three spaces. By appealing to the one-way ANOVA, we can actually identify these projection matrices.

$C([X_0,X_1,\ldots,X_a])$ is the column space for the one-way ANOVA model

$$y_{ijk} = \mu + \alpha_i + e_{ijk}, \tag{2}$$

where the subscripts j and k are both used to indicate replications. Similarly, $C([X_0,X_{a+1},\ldots,X_{a+b}])$ is the column space for the one-way ANOVA model

$$y_{ijk} = \mu + \eta_j + e_{ijk}, \tag{3}$$

where the subscripts i and k are both used to indicate replications. If one actually writes down the matrix $[X_0,X_{a+1},\ldots,X_{a+b}]$, it looks a bit different from the usual form of a one-way ANOVA model matrix because the rows have been permuted out of the convenient order generally used.

Let M_α be the projection matrix used to test for no treatment effects in model (2). M_α is the perpendicular projection matrix onto $C(Z_1,\ldots,Z_a)$. Similarly, if M_η is the projection matrix for testing no treatment effects in model (3), then M_η is the perpendicular projection matrix onto $C(Z_{a+1},\ldots,Z_{a+b})$. It follows that

$$M = \frac{1}{n}J_n^n + M_\alpha + M_\eta.$$

By comparing models, we see, for instance, that the test for $H_0 : \alpha_1 = \cdots = \alpha_a$ is based on

$$\frac{Y'M_\alpha Y/r(M_\alpha)}{Y'(I-M)Y/r(I-M)}.$$

It is easy to see that $r(M_\alpha) = a-1$ and $r(I-M) = n-a-b+1$. $Y'M_\alpha Y$ can be found as in Chapter 4 by appealing to the analysis of the one-way ANOVA model (2). In particular, since pairs jk identify replications,

$$M_\alpha Y = [t_{ijk}], \qquad \text{where } t_{ijk} = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot} \tag{4}$$

and

$$SS(\alpha) = Y'M_\alpha Y = [M_\alpha Y]'[M_\alpha Y] = bN\sum_{i=1}^{a}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2.$$

Expected mean squares can also be found by appealing to the one-way ANOVA

$$\mathrm{E}(Y'M_\alpha Y) = \sigma^2(a-1) + \beta'X'M_\alpha X\beta.$$

Substituting $\mu + \alpha_i + \eta_j$ for y_{ijk} in (4) gives

$$M_\alpha X\beta = [t_{ijk}], \qquad \text{where } t_{ijk} = \alpha_i - \bar{\alpha}_\cdot$$

and thus

$$\mathrm{E}\left[Y'M_\alpha Y/(a-1)\right] = \sigma^2 + \frac{bN}{a-1}\sum_{i=1}^{a}(\alpha_i - \bar{\alpha}_\cdot)^2.$$

Similar results hold for testing $H_0 : \eta_1 = \cdots = \eta_b$.

The *SSE* can be found using (4) and the facts that

$$M_\eta Y = [t_{ijk}], \qquad \text{where } t_{ijk} = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}$$

and

$$\frac{1}{n}J_n^n Y = [t_{ijk}], \qquad \text{where } t_{ijk} = \bar{y}_{\cdots}\,.$$

Because $(I-M)Y = Y - (1/n)J_n^n Y - M_\alpha Y - M_\eta Y$,

$$(I-M)Y = [t_{ijk}]$$

where

$$\begin{aligned} t_{ijk} &= y_{ijk} - \bar{y}_{\cdots} - (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) - (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\ &= y_{ijk} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}. \end{aligned}$$

Finally,

$$\begin{aligned} SSE &= Y'(I-M)Y = [(I-M)Y]'[(I-M)Y] \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{N}(y_{ijk} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2. \end{aligned}$$

The analysis of variance table is given in Table 7.1.

7.1.1 Contrasts

We wish to show that estimation and testing of contrasts in a balanced two-way ANOVA is done exactly as in the one-way ANOVA: by ignoring the fact that a second type of treatment exists. This follows from showing that, say, a contrast in the α_is involves a constraint on $C(M_\alpha)$, and $C(M_\alpha)$ is defined by the one-way ANOVA without the η_js.

Table 7.1 Balanced Two-Way Analysis of Variance Table with No Interaction.

Matrix Notation		
Source	df	SS
Grand Mean	1	$Y'\left(\frac{1}{n}J_n^n\right)Y$
Treatments(α)	$a-1$	$Y'M_\alpha Y$
Treatments(η)	$b-1$	$Y'M_\eta Y$
Error	$n-a-b+1$	$Y'(I-M)Y$
Total	$n=abN$	$Y'Y$

Source	SS	$\mathrm{E}(MS)$
Grand Mean	$SSGM$	$\sigma^2+\beta'X'\left(\frac{1}{n}J_n^n\right)X\beta$
Treatments(α)	$SS(\alpha)$	$\sigma^2+\beta'X'M_\alpha X\beta/(a-1)$
Treatments(η)	$SS(\eta)$	$\sigma^2+\beta'X'M_\eta X\beta/(b-1)$
Error	SSE	σ^2
Total	$SSTot$	

Algebraic Notation		
Source	df	SS
Grand Mean	$dfGM$	$n^{-1}y_{\cdots}^2=n\bar{y}_{\cdots}^2$
Treatments(α)	$df(\alpha)$	$bN\sum_{i=1}^a(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdots})^2$
Treatments(η)	$df(\eta)$	$aN\sum_{j=1}^b(\bar{y}_{\cdot j\cdot}-\bar{y}_{\cdots})^2$
Error	dfE	$\sum_{ijk}\left(y_{ijk}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}+\bar{y}_{\cdots}\right)^2$
Total	$dfTot$	$\sum_{ijk}y_{ijk}^2$

Source	MS	$\mathrm{E}(MS)$
Grand Mean	$SSGM$	$\sigma^2+abN(\mu+\bar{\alpha}_\cdot+\bar{\eta}_\cdot)^2$
Treatments(α)	$SS(\alpha)/(a-1)$	$\sigma^2+bN\sum_{i=1}^a(\alpha_i-\bar{\alpha}_\cdot)^2/(a-1)$
Treatments(η)	$SS(\eta)/(b-1)$	$\sigma^2+aN\sum_{j=1}^b(\eta_j-\bar{\eta}_\cdot)^2/(b-1)$
Error	$SSE/(n-a-b+1)$	σ^2

Theorem 7.1.2. Let $\lambda'\beta$ be estimable and $\rho'X = \lambda'$. Then $\lambda'\beta$ is a contrast in the α_is if and only if $\rho'M = \rho'M_\alpha$. In this case, $\lambda'\hat{\beta} = \rho'MY = \rho'M_\alpha Y$, which is the estimate from the one-way ANOVA ignoring the η_js.

PROOF. Let $\lambda'\beta = \sum_{i=1}^a c_i\alpha_i$ with $\sum_{i=1}^a c_i = 0$. Thus, $\lambda' = (0, c_1, \ldots, c_a, 0, \ldots, 0)$ and $\lambda' J_{a+b+1} = 0$. To have such a λ is to have ρ with

$$\rho' X_i = 0, \quad i = 0, a+1, a+2, \ldots, a+b,$$

which happens if and only if ρ is orthogonal to $C(Z_0, Z_{a+1}, \ldots, Z_{a+b}) = C(M - M_\alpha)$. In other words, $\rho'(M - M_\alpha) = 0$ and $\rho'M = \rho'M_\alpha$. □

One interpretation of this result is that having a contrast in the α_is equal to zero puts a constraint on $C(X)$ that requires $\mathrm{E}(Y) \in C(X)$ and $\mathrm{E}(Y) \perp M_\alpha\rho$. Clearly this constitutes a constraint on $C(M_\alpha)$, the space for the α treatments. Another interpretation is that estimation or testing of a contrast in the α_is is done using M_α, which is exactly the way it is done in a one-way ANOVA ignoring the η_js.

Specifically, if we have a contrast $\sum_{i=1}^a c_i\alpha_i$, then the corresponding vector $M_\alpha\rho$ is

$$M_\alpha\rho = [t_{ijk}], \qquad \text{where } t_{ijk} = c_i/bN.$$

The estimated contrast is $\rho'M_\alpha Y = \sum_{i=1}^a c_i\bar{y}_{i\cdot\cdot}$ having a variance of $\sigma^2\rho'M_\alpha\rho = \sigma^2\sum_{i=1}^a c_i^2/bN$ and a sum of squares for testing $H_0 : \sum_{i=1}^a c_i\alpha_i = 0$ of

$$\left(\sum_{i=1}^a c_i\bar{y}_{i\cdot\cdot}\right)^2 \Big/ \left(\sum_{i=1}^a c_i^2/bN\right).$$

To get two orthogonal constraints on $C(M_\alpha)$, as in the one-way ANOVA, take ρ_1 and ρ_2 such that $\rho_1'X\beta$ and $\rho_2'X\beta$ are contrasts in the α_is and $\rho_1'M_\alpha\rho_2 = 0$. If $\rho_j'X\beta = \sum_{i=1}^a c_{ji}\alpha_i$, then, as shown for the one-way ANOVA, $\rho_1'M_\alpha\rho_2 = 0$ if and only if $\sum_{i=1}^a c_{1i}c_{2i} = 0$.

In the balanced two-way ANOVA without interaction, if N is greater than 1, we have a row structure to the model matrix with ab distinct rows. This allows estimation of pure error and lack of fit. The balanced two-way ANOVA with interaction retains the row structure and is equivalent to a large one-way ANOVA with ab treatments. Thus, the interaction model provides one parameterization of the model developed in Subsection 6.6.1 for testing lack of fit. The sum of squares for interaction is just the sum of squares for lack of fit of the no interaction model.

7.2 Balanced Two-Way ANOVA with Interaction

The balanced two-way ANOVA with interaction model is written

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk},$$

$i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, N$.

EXAMPLE 7.2.1. Suppose $a = 3$, $b = 2$, $N = 4$. In matrix terms we write

$$\begin{bmatrix} y_{111} \\ y_{112} \\ y_{113} \\ y_{114} \\ y_{121} \\ y_{122} \\ y_{123} \\ y_{124} \\ y_{211} \\ y_{212} \\ y_{213} \\ y_{214} \\ y_{221} \\ y_{222} \\ y_{223} \\ y_{224} \\ y_{311} \\ y_{312} \\ y_{313} \\ y_{314} \\ y_{321} \\ y_{322} \\ y_{323} \\ y_{324} \end{bmatrix} = \begin{bmatrix} 1&1&0&0&1&0&1&0&0&0&0&0 \\ 1&1&0&0&1&0&1&0&0&0&0&0 \\ 1&1&0&0&1&0&1&0&0&0&0&0 \\ 1&1&0&0&1&0&1&0&0&0&0&0 \\ 1&1&0&0&0&1&0&1&0&0&0&0 \\ 1&1&0&0&0&1&0&1&0&0&0&0 \\ 1&1&0&0&0&1&0&1&0&0&0&0 \\ 1&1&0&0&0&1&0&1&0&0&0&0 \\ 1&0&1&0&1&0&0&0&1&0&0&0 \\ 1&0&1&0&1&0&0&0&1&0&0&0 \\ 1&0&1&0&1&0&0&0&1&0&0&0 \\ 1&0&1&0&1&0&0&0&1&0&0&0 \\ 1&0&1&0&0&1&0&0&0&1&0&0 \\ 1&0&1&0&0&1&0&0&0&1&0&0 \\ 1&0&1&0&0&1&0&0&0&1&0&0 \\ 1&0&1&0&0&1&0&0&0&1&0&0 \\ 1&0&0&1&1&0&0&0&0&0&1&0 \\ 1&0&0&1&1&0&0&0&0&0&1&0 \\ 1&0&0&1&1&0&0&0&0&0&1&0 \\ 1&0&0&1&1&0&0&0&0&0&1&0 \\ 1&0&0&1&0&1&0&0&0&0&0&1 \\ 1&0&0&1&0&1&0&0&0&0&0&1 \\ 1&0&0&1&0&1&0&0&0&0&0&1 \\ 1&0&0&1&0&1&0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \eta_1 \\ \eta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{bmatrix} + e.$$

In general, the model matrix can be written

$$X = [X_0, X_1, \ldots, X_a, X_{a+1}, \ldots, X_{a+b}, X_{a+b+1}, \ldots, X_{a+b+ab}].$$

The columns $X_0, \ldots, X_{a+b}$ are exactly the same as for (7.1.1), the model without interaction. The key fact to notice is that

$$C(X) = C(X_{a+b+1}, \ldots, X_{a+b+ab});$$

this will be shown rigorously later. We can write an equivalent model using just $X_{a+b+1}, \ldots, X_{a+b+ab}$, say

$$y_{ijk} = \mu_{ij} + e_{ijk}.$$

This is just a one-way ANOVA model with ab treatments and is sometimes called the *cell means model*.

We can decompose $C(X)$ into four orthogonal parts based on the identity

$$M = \frac{1}{n}J_n^n + M_\alpha + M_\eta + M_\gamma,$$

where

$$M_\gamma \equiv M - \frac{1}{n}J_n^n - M_\alpha - M_\eta .$$

M_α and M_η come from the no interaction model. Thus, as discussed earlier, M_α and M_η each come from a one-way ANOVA model. Since M also comes from a one-way ANOVA model, we can actually find M_γ. The interaction space is defined as $C(M_\gamma)$. The sum of squares for interaction is $Y'M_\gamma Y$. It is just the sum of squares left over after explaining as much as possible with μ, the α_is, and the η_js. The degrees of freedom for the interaction are

$$\begin{aligned} r(M_\gamma) &= r\left(M - \frac{1}{n}J_n^n - M_\alpha - M_\eta\right) \\ &= ab - 1 - (a-1) - (b-1) = (a-1)(b-1). \end{aligned}$$

The algebraic formula for the interaction sum of squares can be found as follows:

$$SS(\gamma) = Y'M_\gamma Y = [M_\gamma Y]'[M_\gamma Y], \tag{1}$$

where

$$M_\gamma Y = \left(M - \frac{1}{n}J_n^n - M_\alpha - M_\eta\right)Y = MY - \frac{1}{n}J_n^n Y - M_\alpha Y - M_\eta Y.$$

With M the projection operator for a one-way ANOVA, all of the terms on the right of the last equation have been characterized, so

$$\left(M - \frac{1}{n}J_n^n - M_\alpha - M_\eta\right)Y = [t_{ijk}],$$

where

$$\begin{aligned} t_{ijk} &= \bar{y}_{ij\cdot} - \bar{y}_{\cdots} - (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) - (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\ &= \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}. \end{aligned}$$

It follows immediately from (1) that

$$\begin{aligned} SS(\gamma) &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{N}[\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}]^2 \\ &= N\sum_{i=1}^{a}\sum_{j=1}^{b}[\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}]^2. \end{aligned}$$

The expected value of $Y'M_\gamma Y$ is $\sigma^2(a-1)(b-1) + \beta'X'M_\gamma X\beta$. The second term is a quadratic form in the γ_{ij}s because $(M - M_\alpha - M_\eta - [1/n]J_n^n)X_r = 0$ for $r = 0, 1, \ldots, a+b$. The algebraic form of $\beta'X'M_\gamma X\beta$ can be found by substituting $\mu +$

$\alpha_i + \eta_j + \gamma_{ij}$ for y_{ijk} in $SS(\gamma)$. Simplification gives

$$\beta' X' M_\gamma X\beta = N \sum_{ij} [\gamma_{ij} - \bar{\gamma}_{i\cdot} - \bar{\gamma}_{\cdot j} - \bar{\gamma}_{\cdot\cdot}]^2.$$

The expected values of $Y'M_\alpha Y$, $Y'M_\eta Y$, and $Y'(1/n)J_n^n Y$ are now different from those found for the no interaction model. As above, algebraic forms for the expected values can be computed by substituting for the y_{ijk}s in the algebraic forms for the sums of squares. For instance,

$$\mathrm{E}(Y'M_\alpha Y) = \sigma^2(a-1) + \beta' X' M_\alpha X\beta = \sigma^2(a-1) + bN \sum_{i=1}^{a} (\alpha_i + \bar{\gamma}_{i\cdot} - \bar{\alpha}_{\cdot} - \bar{\gamma}_{\cdot\cdot})^2,$$

which depends on the γ_{ij}s and not just the α_is. This implies that the standard test is not a test that the α_is are all equal; it is a test that the $(\alpha_i + \bar{\gamma}_{i\cdot})$s are all equal. In fact, since the column space associated with the γ_{ij}s spans the entire space, i.e., $C(X_{a+b+1}, \ldots, X_{a+b+ab}) = C(X)$, all estimable functions of the parameters are functions of the γ_{ij}s. To see this, note that if $\lambda'\beta$ is not a function of the γ_{ij}s, but is estimable, then $\rho' X_i = 0$, $i = a+b+1, \ldots, a+b+ab$, and hence $\lambda' = \rho' X = 0$; so $\lambda'\beta$ is identically zero.

If we impose the "usual" side conditions, $\alpha_{\cdot} = \eta_{\cdot} = \gamma_{i\cdot} = \gamma_{\cdot j} = 0$, we obtain, for example,

$$\mathrm{E}(Y'M_\alpha Y) = \sigma^2(a-1) + bN \sum_{i=1}^{a} \alpha_i^2,$$

which looks nice but serves no purpose other than to hide the fact that these new α_i terms are averages over any interactions that exist.

As we did for a two-way ANOVA without interaction, we can put a single degree of freedom constraint on, say, $C(M_\alpha)$ by choosing a function $\lambda'\beta$ such that $\lambda' = \rho' X$ and $\rho' M = \rho' M_\alpha$. However, such a constraint no longer yields a contrast in the α_is. To examine $\rho' M_\alpha X\beta$, we examine the nature of $M_\alpha X\beta$. Since $M_\alpha Y$ is a vector whose rows are made up of terms like $(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})$, algebraic substitution gives $M_\alpha X\beta$ as a vector whose rows are terms like $(\alpha_i + \bar{\gamma}_{i\cdot} - \bar{\alpha}_{\cdot} - \bar{\gamma}_{\cdot\cdot})$. $\lambda'\beta = \rho' M_\alpha X\beta$ will be a contrast in these terms or, equivalently, in the $(\alpha_i + \bar{\gamma}_{i\cdot})$s. Such contrasts are generally hard to interpret. A contrast in the terms $(\alpha_i + \bar{\gamma}_{i\cdot})$ will be called a contrast in the α space. Similarly, a contrast in the terms $(\eta_j + \bar{\gamma}_{\cdot j})$ will be called a contrast in the η space.

There are two fundamental approaches to analyzing a two-way ANOVA with interaction. In both methods, the test for whether interaction adds to the two-way without interaction model is performed. If this is not significant, the interactions are tentatively assumed to be zero. If the effect of the interaction terms is significant, the easiest approach is to do the entire analysis as a one-way ANOVA. The alternative approach consists of trying to interpret the contrasts in $C(M_\alpha)$ and $C(M_\eta)$ and examining constraints in the interaction space.

7.2.1 Interaction Contrasts

We now consider how to define and test constraints on the interaction space. The hypothesis $H_0 : \lambda'\beta = 0$ puts a constraint on the interaction space if and only if $\lambda' = \rho'X$ has the property $\rho'M = \rho'(M - M_\alpha - M_\eta - [1/n]J_n^n)$. To find hypotheses that put constraints on the interaction space, it suffices to find $\rho \perp C(M_\alpha + M_\eta + [1/n]J_n^n)$ or, alternately, $\rho'X_i = 0$, $i = 0, \ldots, a+b$.

The goal of the following discussion is to characterize vectors in the interaction space, i.e., to characterize the vectors $M\rho$ that have the property $\rho'X_i = 0$ for $i = 0, \ldots, a+b$. A convenient way to do this is to characterize the vectors ρ that have two properties: 1) $M\rho = \rho$, and 2) $\rho'X_i = 0$ for $i = 0, \ldots, a+b$. The second property ensures that $M\rho$ is in the interaction space.

First we find a class of vectors that are contained in the interaction space. From this class of vectors we will get a class of orthogonal bases for the interaction space. The class of vectors and the class of orthogonal bases are found by combining a contrast in the α space with a contrast in the η space. This method leads naturally to the standard technique of examining interactions. Finally, a second class of vectors contained in the interaction space will be found. This class contains the first class as a special case. The second class is closed under linear combinations, so the second class is a vector space that contains a basis for the interaction space but which is also contained in the interaction space. It follows that the second class is precisely the interaction space.

At this point a problem arises. It is very convenient to write down ANOVA models as was done in Example 7.2.1, with indices to the right in y_{ijk} changing fastest. It is easy to see what the model looks like and that it can be written in a similar manner for any choices of a, b, and N. In the example it would be easy to find a vector that is in the interaction space, and it would be easy to see that the technique could be extended to cover any two-way ANOVA problem. Although it is easy to see how to write down models as in the example, it is awkward to develop a notation for it. It is also less than satisfying to have a proof that depends on the way in which the model is written down. Consequently, the material on finding a vector in the interaction space will be presented in three stages: an application to Example 7.2.1, a comment on how that argument can be generalized, and finally a rigorous presentation.

EXAMPLE 7.2.2. In the model of Example 7.2.1, let $d' = (d_1, d_2, d_3) = (1, 2, -3)$ and $c' = (c_1, c_2) = (1, -1)$. The d_is determine a contrast in the α space and the c_js determine a contrast in the η space. Consider

$$\begin{aligned}[d' \otimes c'] &= [c_1d_1, c_2d_1, c_1d_2, c_2d_2, c_1d_3, c_2d_3] \\ &= [1, -1, 2, -2, -3, 3];\end{aligned}$$

and since $N = 4$, let

$$\rho' = \frac{1}{4}[d' \otimes c'] \otimes J_1^4 = \frac{1}{4}[c_1d_1J_1^4, c_2d_1J_1^4, \ldots, c_2d_3J_1^4]$$

$$= \frac{1}{4}\big[1,1,1,1,-1,-1,-1,-1,2,2,2,2,-2,-2,-2,-2,$$
$$-3,-3,-3,-3,3,3,3,3\big].$$

It is easily seen that ρ is orthogonal to the first six columns of X; thus $M\rho$ is in the interaction space. However, it is also easily seen that $\rho \in C(X)$, so $M\rho = \rho$. The vector ρ is itself a vector in the interaction space.

Extending the argument to an arbitrary two-way ANOVA written in standard form, let $d' = (d_1, \ldots, d_a)$ with $\sum_{i=1}^{a} d_i = 0$ and $c' = (c_1, \ldots, c_b)$ with $\sum_{j=1}^{b} c_j = 0$. The d_is can be thought of as determining a contrast in the α space and the c_js as determining a contrast in the η space. Let $\rho' = (1/N)[d' \otimes c'] \otimes J_1^N$, i.e.,

$$\rho' = \frac{1}{N}(c_1 d_1 J_1^N, c_2 d_1 J_1^N, \ldots, c_b d_1 J_1^N, c_1 d_2 J_1^N, \ldots, c_b d_a J_1^N).$$

It is clear that $\rho' X_i = 0$ for $i = 0, \ldots, a+b$ and $\rho \in C(X_{a+b+1}, \ldots, X_{a+b+ab})$ so $\rho' M = \rho'$ and $\rho' X\beta = 0$ puts a constraint on the correct space.

To make the argument completely rigorous, with d and c defined as above, take

$$\rho = [\rho_{ijk}], \qquad \text{where } \rho_{ijk} = \frac{1}{N}(d_i c_j).$$

Using the characterizations of $X_0, \ldots, X_{a+b}$ from Section 1 we get

$$\rho' X_0 = \sum_i \sum_j \sum_k \frac{1}{N}(d_i c_j) = \sum_i d_i \sum_j c_j = 0;$$

for $r = 1, \ldots, a$, we get

$$\rho' X_r = \sum_i \sum_j \sum_k \frac{1}{N}(d_i c_j)\delta_{ir} = d_r \sum_j c_j = 0;$$

and for $s = 1, \ldots, b$, we get

$$\rho' X_{a+s} = \sum_i \sum_j \sum_k \frac{1}{N}(d_i c_j)\delta_{js} = c_s \sum_i d_i = 0.$$

This shows that $M\rho$ is in the interaction space.

To show that $M\rho = \rho$, we show that $\rho \in C(X)$. We need to characterize $C(X)$. The columns of X that correspond to the γ_{ij}s are the vectors $X_{a+b+1}, \ldots, X_{a+b+ab}$. Reindex these as $X_{(1,1)}, X_{(1,2)}, \ldots, X_{(1,b)}, \ldots, X_{(2,1)}, \ldots, X_{(a,b)}$. Thus $X_{(i,j)}$ is the column of X corresponding to γ_{ij}. We can then write

$$X_{(r,s)} = [t_{ijk}], \qquad \text{where } t_{ijk} = \delta_{(i,j)(r,s)}$$

and $\delta_{(i,j)(r,s)} = 1$ if $(i,j) = (r,s)$, and 0 otherwise. It is easily seen that

$$X_0 = \sum_{i=1}^{a}\sum_{j=1}^{b} X_{(i,j)},$$

$$X_r = \sum_{j=1}^{b} X_{(r,j)}, \quad r = 1,\ldots,a,$$

$$X_{a+s} = \sum_{i=1}^{a} X_{(i,s)}, \quad s = 1,\ldots,b.$$

This shows that $C(X) = C(X_{(1,1)}, X_{(1,2)}, \ldots, X_{(a,b)})$. (In the past, we have merely claimed this result.) It is also easily seen that

$$\rho = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{d_i c_j}{N} X_{(i,j)},$$

so that $\rho \in C(X)$.

We have found a class of vectors contained in the interaction space. If we find $(a-1)(b-1)$ such vectors, say ρ_r, $r = 1,\ldots,(a-1)(b-1)$, where

$$\rho_r'\,(M - M_\alpha - M_\eta - [1/n]J_n^n)\,\rho_s = 0$$

for any $r \neq s$, then we will have an orthogonal basis for $C(M - M_\alpha - M_\eta - [1/n]J_n^n)$. Consider now another pair of contrasts for α and η, say $d^* = (d_1^*,\ldots,d_a^*)'$ and $c^* = (c_1^*,\ldots,c_b^*)'$, where one, say $c^* = (c_1^*,\ldots,c_b^*)'$, is orthogonal to the corresponding contrast in the other pair. We can write

$$\rho_* = [\rho_{ijk}^*], \qquad \text{where } \rho_{ijk}^* = d_i^* c_j^*/N,$$

and we know that $\sum_{i=1}^{a} d_i^* = \sum_{j=1}^{b} c_j^* = \sum_{j=1}^{b} c_j c_j^* = 0$. With our choices of ρ and ρ_*,

$$\begin{aligned}
\rho'\,(M - M_\alpha - M_\eta - [1/n]J_n^n)\,\rho_* &= \rho'\rho_* \\
&= N^{-2}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{N} d_i d_i^* c_j c_j^* \\
&= N^{-1}\sum_{i=1}^{a} d_i d_i^* \sum_{j=1}^{b} c_j c_j^* \\
&= 0.
\end{aligned}$$

Since there are $(a-1)$ orthogonal ways of choosing $d' = (d_1,\ldots,d_a)$ and $(b-1)$ orthogonal ways of choosing $c' = (c_1,\ldots,c_b)$, there are $(a-1)(b-1)$ orthogonal vectors ρ that can be chosen in this fashion. This provides the desired orthogonal breakdown of the interaction space.

To actually compute the estimates of these parametric functions, recall that M is the perpendicular projection operator for a one-way ANOVA. With ρ chosen as above,

$$\rho' X\beta = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{N} \frac{d_i c_j}{N} \gamma_{ij} = \sum_{i=1}^{a} \sum_{j=1}^{b} d_i c_j \gamma_{ij}.$$

Its estimate reduces to

$$\rho' MY = \sum_{i=1}^{a} \sum_{j=1}^{b} d_i c_j \bar{y}_{ij\cdot}$$

and its variance to

$$\sigma^2 \rho' M\rho = \sigma^2 \rho'\rho = \sigma^2 \sum_{i=1}^{a} \sum_{j=1}^{b} d_i^2 c_j^2 / N.$$

A handy method for computing these is to write out the following two-way table:

	c_1	c_2	$\cdots$	c_b	h_i
d_1	$\bar{y}_{11\cdot}$	$\bar{y}_{12\cdot}$	$\cdots$	$\bar{y}_{1b\cdot}$	h_1
d_2	$\bar{y}_{21\cdot}$	$\bar{y}_{22\cdot}$	$\cdots$	$\bar{y}_{2b\cdot}$	h_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
d_a	$\bar{y}_{a1\cdot}$	$\bar{y}_{a2\cdot}$	$\cdots$	$\bar{y}_{ab\cdot}$	h_a
g_j	g_1	g_2	$\cdots$	g_b	

where $h_i = \sum_{j=1}^{b} c_j \bar{y}_{ij\cdot}$ and $g_j = \sum_{i=1}^{a} d_i \bar{y}_{ij\cdot}$. We can then write

$$\rho' MY = \sum_{j=1}^{b} c_j g_j = \sum_{i=1}^{a} d_i h_i.$$

Unfortunately, not all contrasts in the interaction space can be defined as illustrated here. The $(a-1)(b-1)$ orthogonal vectors that we have discussed finding form a basis for the interaction space, so any linear combination of these vectors is also in the interaction space. However, not all of these linear combinations can be written with the method based on two contrasts.

Let $Q = [q_{ij}]$ be any $a \times b$ matrix such that $J_a' Q = 0$ and $QJ_b = 0$ (i.e., $q_{i\cdot} = 0 = q_{\cdot j}$). If the model is written down in the usual manner, the vectors in the interaction space are the vectors of the form $\rho = (1/N)\text{Vec}(Q') \otimes J_N$. In general, we write the vector ρ with triple subscript notation as

$$\rho = [\rho_{ijk}], \qquad \text{where } \rho_{ijk} = q_{ij}/N \quad \text{with } q_{i\cdot} = q_{\cdot j} = 0. \tag{2}$$

First, note that linear combinations of vectors with this structure retain the structure; thus vectors of this structure form a vector space. Vectors ρ with this structure are in $C(X)$, and it is easily seen that $\rho' X_i = 0$ for $i = 0, 1, \ldots, a+b$. Thus, a vector with this structure is contained in the interaction space. Note also that the first method of finding vectors in the interaction space using a pair of contrasts yields a vector of the structure that we are currently considering, so the vector space alluded to above is both contained in the interaction space and contains a basis for the interaction space. It follows that the interaction space is precisely the set of all vectors with the form (2).

Exercise 7.1 Prove the claims of the previous paragraph. In particular, show that linear combinations of the vectors presented retain their structure, that the vectors are orthogonal to the columns of X corresponding to the grand mean and the treatment effects, and that the vectors based on contrasts have the same structure as the vector given above.

For estimation and tests of single-degree-of-freedom hypotheses in the interaction space, it is easily seen, with ρ taken as above, that

$$\rho' X\beta = \sum_{i=1}^{a}\sum_{j=1}^{b} q_{ij}\gamma_{ij}\,,$$

$$\rho' MY = \sum_{i=1}^{a}\sum_{j=1}^{b} q_{ij}\bar{y}_{ij\cdot}\,,$$

$$\mathrm{Var}(\rho' MY) = \sigma^2 \sum_{i=1}^{a}\sum_{j=1}^{b} q_{ij}^2/N\,.$$

Table 7.2 gives the ANOVA table for the balanced two-way ANOVA with interaction. Note that if $N = 1$, there will be no pure error term available. In that case, it is often *assumed* that the interactions add nothing to the model, so that the mean square for interactions can be used as an estimate of error for the two-way ANOVA without interaction. See Example 13.2.4 for a graphical procedure that addresses this problem.

Exercise 7.2 Does the statement "the interactions add nothing to the model" mean that $\gamma_{11} = \gamma_{12} = \cdots = \gamma_{ab}$? If it does, justify the statement. If it does not, what does the statement mean?

Two final comments on exploratory work with interactions. If the $(a-1)(b-1)$ degree-of-freedom F test for interactions is not significant, then neither Scheffé's method nor the LSD method will allow us to claim significance for any contrast in the interactions. Bonferroni's method may give significance, but it is unlikely. Nevertheless, if our goal is to explore the data, there may be *suggestions* of possible interactions. For example, if you work with interactions long enough, you begin to think that some interaction contrasts have reasonable interpretations. If such a contrast exists that accounts for the bulk of the interaction sum of squares, and if the corresponding F test approaches significance, then it would be unwise to ignore this possible source of interaction. (As a word of warning though, recall that there always exists an interaction contrast, usually uninterpretable, that accounts for the entire sum of squares for interaction.) Second, in exploratory work it is very useful

Table 7.2 Balanced Two-Way Analysis of Variance Table with Interaction.

Matrix Notation		
Source	df	SS
Grand Mean	1	$Y'\left(\frac{1}{n}J_n^n\right)Y$
Treatments(α)	$a-1$	$Y'M_\alpha Y$
Treatments(η)	$b-1$	$Y'M_\eta Y$
Interaction(γ)	$(a-1)(b-1)$	$Y'\left(M-M_\alpha-M_\eta-\frac{1}{n}J_n^n\right)Y$
Error	$n-ab$	$Y'(I-M)Y$
Total	$n=abN$	$Y'Y$

Source	SS	$\mathrm{E}(MS)$
Grand Mean	$SSGM$	$\sigma^2+\beta'X'\left(\frac{1}{n}J_n^n\right)X\beta$
Treatments(α)	$SS(\alpha)$	$\sigma^2+\beta'X'M_\alpha X\beta/(a-1)$
Treatments(η)	$SS(\eta)$	$\sigma^2+\beta'X'M_\eta X\beta/(b-1)$
Interaction(γ)	$(a-1)(b-1)$	$\sigma^2+\beta'X'M_\gamma X\beta/(a-1)(b-1)$
Error	SSE	σ^2
Total	$SSTot$	

Algebraic Notation		
Source	df	SS
Grand Mean	$dfGM$	$n^{-1}y_{\cdot\cdot\cdot}^2=n\bar{y}_{\cdot\cdot\cdot}^2$
Treatments(α)	$df(\alpha)$	$bN\sum_{i=1}^a\left(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot\cdot\cdot}\right)^2$
Treatments(η)	$df(\eta)$	$aN\sum_{j=1}^b\left(\bar{y}_{\cdot j\cdot}-\bar{y}_{\cdot\cdot\cdot}\right)^2$
Interaction(γ)	$df(\gamma)$	$N\sum_{ij}\left(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}+\bar{y}_{\cdot\cdot\cdot}\right)^2$
Error	dfE	$\sum_{ijk}\left(y_{ijk}-\bar{y}_{ij\cdot}\right)^2$
Total	$dfTot$	$\sum_{ijk}y_{ijk}^2$

Source	MS	$\mathrm{E}(MS)$
Grand Mean	$SSGM$	$\sigma^2+abN(\mu+\bar{\alpha}_\cdot+\bar{\eta}_\cdot+\bar{\gamma}_{\cdot\cdot})^2$
Treatments(α)	$SS(\alpha)/(a-1)$	$\sigma^2+\frac{bN}{a-1}\sum_i\left(\alpha_i+\bar{\gamma}_{i\cdot}-\bar{\alpha}_\cdot-\bar{\gamma}_{\cdot\cdot}\right)^2$
Treatments(η)	$SS(\eta)/(b-1)$	$\sigma^2+\frac{aN}{b-1}\sum_j\left(\eta_j+\bar{\gamma}_{\cdot j}-\bar{\eta}_\cdot-\bar{\gamma}_{\cdot\cdot}\right)^2$
Interaction(γ)	$SS(\gamma)/df(\gamma)$	$\sigma^2+\frac{N}{df(\gamma)}\sum_{ij}\left(\gamma_{ij}-\bar{\gamma}_{i\cdot}-\bar{\gamma}_{\cdot j}+\bar{\gamma}_{\cdot\cdot}\right)^2$
Error	$SSE/(n-ab)$	σ^2

to plot the cell means. For example, one can plot the points $(i, \bar{y}_{ij\cdot})$ for each value of j, connecting the points to give b different curves, one for each j. If the α treatments correspond to levels of some quantitative factor x_i, plot the points $(x_i, \bar{y}_{ij\cdot})$ for each value of j. If there are no interactions, the plots for different values of j should be (approximately) parallel. (If no interactions are present, the plots estimate plots of the points $(i, \mu + \alpha_i + \eta_j)$. These plots are parallel for all j.) Deviations from a parallel set of plots can suggest possible sources of interaction. The data are suggesting possible interaction contrasts, so if valid tests are desired, use Scheffé's method. Finally, it is equally appropriate to plot the points $(j, \bar{y}_{ij\cdot})$ for all i or a corresponding set of points using quantitative levels associated with the η treatments.

7.3 Polynomial Regression and the Balanced Two-Way ANOVA

Consider first the balanced two-way ANOVA without interaction. Suppose that the ith level of the α treatments corresponds to some number w_i and that the jth level of the η treatments corresponds to some number z_i. We can write vectors taking powers of w_i and z_i. For $r = 1, \ldots, a-1$ and $s = 1, \ldots, b-1$, write

$$W^r = [t_{ijk}], \qquad \text{where } t_{ijk} = w_i^r\,,$$

$$Z^r = [t_{ijk}], \qquad \text{where } t_{ijk} = z_i^s\,.$$

Note that $W^0 = Z^0 = J$.

EXAMPLE 7.3.1. Consider the model $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$, $i = 1,2,3$, $j = 1,2$, $k = 1,2$. Suppose that the α treatments are 1, 2, and 3 pounds of fertilizer and that the η treatments are 5 and 7 pounds of manure. Then, if we write $Y = (y_{111}, y_{112}, y_{121}, y_{122}, y_{211}, \ldots, y_{322})'$, we have

$$W^1 = (1,1,1,1,2,2,2,2,3,3,3,3)',$$

$$W^2 = (1,1,1,1,4,4,4,4,9,9,9,9)',$$

$$Z^1 = (5,5,7,7,5,5,7,7,5,5,7,7)'.$$

From the discussion of Section 6.7 on polynomial regression and one-way ANOVA, we have

$$C(J, W^1, \ldots, W^{a-1}) = C(X_0, X_1, \ldots, X_a),$$
$$C(J, Z^1, \ldots, Z^{b-1}) = C(X_0, X_{a+1}, \ldots, X_{a+b}),$$

and thus

$$C(J, W^1, \ldots, W^{a-1}, Z^1, \ldots, Z^{b-1}) = C(X_0, X_1, \ldots, X_{a+b}).$$

Fitting the two-way ANOVA is the same as fitting a joint polynomial in w_i and z_j. Writing the models out algebraically, the model

$$y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk},$$

$i = 1,\ldots,a,\ j = 1,\ldots,b,\ k = 1,\ldots,N$, is equivalent to

$$y_{ijk} = \beta_{0,0} + \beta_{1,0}w_i + \cdots + \beta_{a-1,0}w_i^{a-1} + \beta_{0,1}z_j + \cdots + \beta_{0,b-1}z_j^{b-1} + e_{ijk},$$

$i = 1,\ldots,a,\ j = 1,\ldots,b,\ k = 1,\ldots,N$. The correspondence between contrasts and orthogonal polynomials remains valid.

We now show that the model

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk}$$

is equivalent to the model

$$y_{ijk} = \sum_{i=0}^{a-1}\sum_{j=0}^{b-1} \beta_{rs} w_i^r z_j^s + e_{ijk}.$$

EXAMPLE 7.3.1 CONTINUED. The model $y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk}$ is equivalent to $y_{ijk} = \beta_{00} + \beta_{10}w_i + \beta_{20}w_i^2 + \beta_{01}z_j + \beta_{11}w_i z_j + \beta_{21}w_i^2 z_j + e_{ijk}$, where $w_1 = 1$, $w_2 = 2$, $w_3 = 3$, $z_1 = 5$, $z_2 = 7$.

The model matrix for this polynomial model can be written

$$S = \big[J_n^1, W^1, W^2, \ldots, W^{a-1}, Z^1, \ldots, Z^{b-1}, \\ W^1Z^1, \ldots, W^1Z^{b-1}, W^2Z^1, \ldots, W^{a-1}Z^{b-1}\big],$$

where for any two vectors in $\mathbf{R}^n$, say $U = (u_1,\ldots,u_n)'$ and $V = (v_1,\ldots,v_n)'$, we define VU to be the vector $(v_1u_1, v_2u_2, \ldots, v_nu_n)'$.

To establish the equivalence of the models, it is enough to notice that the row structure of $X = [X_0, X_1, \ldots, X_{a+b+ab}]$ is the same as the row structure of S and that $r(X) = ab = r(S)$. $C(X)$ is the column space of the one-way ANOVA model that has a separate effect for each distinct set of rows in S. From our discussion of pure error and lack of fit in Section 6.6, $C(S) \subset C(X)$. Since $r(S) = r(X)$, we have $C(S) = C(X)$; thus the models are equivalent.

We would like to characterize the test in the interaction space that is determined by, say, the quadratic contrast in the α space and the cubic contrast in the η space. We want to show that it is the test for W^2Z^3, i.e., that it is the test for $H_0 : \beta_{23} = 0$ in the model $y_{ijk} = \beta_{00} + \beta_{10}w_i + \beta_{20}w_i^2 + \beta_{01}z_j + \beta_{02}z_j^2 + \beta_{03}z_j^3 + \beta_{11}w_iz_j + \beta_{12}w_iz_j^2 + \beta_{13}w_iz_j^3 + \beta_{21}w_i^2z_j + \beta_{22}w_i^2z_j^2 + \beta_{23}w_i^2z_j^3 + e_{ijk}$. In general, we want to be able to identify the columns W^rZ^s, $r \geq 1, s \geq 1$, with vectors in the interaction space. Specifically, we would like to show that the test of W^rZ^s adding to the model based on

$C([W^i Z^j : i = 0, \ldots, r, j = 0, \ldots, s])$ is precisely the test of the vector in the interaction space defined by the rth degree polynomial contrast in the α space and the sth degree polynomial contrast in the η space. Note that the test of the rth degree polynomial contrast in the α space is a test of whether the column W^r adds to the model based on $C(J, W^1, \ldots, W^r)$, and that the test of the sth degree polynomial contrast in the η space is a test of whether the column Z^s adds to the model based on $C(J, Z^1, \ldots, Z^s)$.

It is important to remember that the test for $W^r Z^s$ adding to the model is not a test for $W^r Z^s$ adding to the full model. It is a test for $W^r Z^s$ adding to the model spanned by the vectors $W^i Z^j$, $i = 0, \ldots, r$, $j = 0, \ldots, s$, where $W^0 = Z^0 = J_n^1$. As discussed above, the test of the vector in the interaction space corresponding to the quadratic contrast in the α space and the cubic contrast in the η space should be the test of whether $W^2 Z^3$ adds to a model with $C(J_n^1, W^1, W^2, Z^1, Z^2, Z^3, W^1 Z^1, W^1 Z^2, W^1 Z^3, W^2 Z^1, W^2 Z^2, W^2 Z^3)$. Intuitively, this is reasonable because the quadratic and cubic contrasts are being fitted after all terms of smaller order.

As observed earlier, if $\lambda' = \rho' X$, then the constraint imposed by $\lambda' \beta = 0$ is $M\rho$ and the test of $\lambda' \beta = 0$ is the test for dropping $M\rho$ from the model. Using the Gram–Schmidt algorithm, find $R_{0,0}, R_{1,0}, \ldots, R_{a-1,0}, R_{0,1}, \ldots, R_{0,b-1}, R_{1,1}, \ldots, R_{a-1,b-1}$ by orthonormalizing, in order, the columns of S. Recall that the polynomial contrasts in the α_is correspond to the vectors $R_{i,0}$, $i = 1, \ldots, a-1$, and that $C(R_{1,0}, \ldots, R_{a-1,0}) = C(M_\alpha)$. Similarly, the polynomial contrasts in the η_js correspond to the vectors $R_{0,j}$ $j = 1, \ldots, b-1$ and $C(R_{0,1}, \ldots, R_{0,a-1}) = C(M_\eta)$. If the test of $\lambda' \beta = 0$ is to test whether, say, Z^s adds to the model after fitting Z^j, $j = 1, \ldots, s-1$, we must have $C(M\rho) = C(R_{0,s})$. Similar results hold for the vectors W^r.

First, we need to examine the relationship between vectors in $C(M_\alpha)$ and $C(M_\eta)$ with vectors in the interaction space. A contrast in the α space is defined by, say, $(d_1, \ldots, d_a)$, and a contrast in the η space by, say, $(c_1, \ldots, c_b)$. From Chapter 4, if we define

$$\rho_1 = [t_{ijk}], \qquad \text{where } t_{ijk} = d_i / Nb$$

and

$$\rho_2 = [t_{ijk}], \qquad \text{where } t_{ijk} = c_j / Na,$$

then $\rho_1 \in C(M_\alpha)$, $\rho_2 \in C(M_\eta)$, $\rho_1 X \beta = \sum_{i=1}^a d_i(\alpha_i + \bar{\gamma}_{i\cdot})$, and $\rho_2 X \beta = \sum_{j=1}^b c_j(\eta_j + \bar{\gamma}_{\cdot j})$. The vector $\rho_1 \rho_2$ is

$$\rho_1 \rho_2 = [t_{ijk}] \qquad \text{where } t_{ijk} = N^{-2}(ab)^{-1} d_i c_j.$$

This is proportional to a vector in the interaction space corresponding to $(d_1, \ldots, d_a)$ and $(c_1, \ldots, c_b)$.

From this argument, it follows that since $R_{r,0}$ is the vector (constraint) in $C(M_\alpha)$ for testing the rth degree polynomial contrast and $R_{0,s}$ is the vector in $C(M_\eta)$ for testing the sth degree polynomial contrast, then $R_{r,0} R_{0,s}$ is a vector in the interaction space. Since the polynomial contrasts are defined to be orthogonal, and since $R_{r,0}$ and $R_{0,s}$ are defined by polynomial contrasts, our discussion in the previous section about orthogonal vectors in the interaction space implies that the set $\{R_{r,0}R_{0,s}|r=1,\ldots,a-1,s=1,\ldots,b-1\}$ is an orthogonal basis for the interaction space. Moreover, with $R_{r,0}$ and $R_{0,s}$ orthonormal,

$$[R_{r,0}R_{0,s}]'[R_{r,0}R_{0,s}] = 1/abN.$$

We now check to see that $(abN)Y'[R_{r,0}R_{0,s}][R_{r,0}R_{0,s}]'Y/MSE$ provides a test of the correct thing, i.e., that W^rZ^s adds to a model containing all lower order terms. Since, by Gram–Schmidt, for some a_is and b_js we have $R_{r,0} = a_0W^r + a_1W^{r-1} + \cdots + a_{r-1}W^1 + a_rJ_n^1$ and $R_{0,s} = b_0Z^s + b_1Z^{s-1} + \cdots + b_{s-1}Z^1 + b_sJ_n^1$, we also have

$$R_{r,0}R_{0,s} = a_0b_0W^rZ^s + \sum_{j=1}^{s} b_jZ^{s-j}W^r + \sum_{i=1}^{r} a_iW^{r-i}Z^s + \sum_{j=1}^{s}\sum_{i=1}^{r} a_ib_jZ^{s-j}W^{r-i}.$$

Letting $R_{1,0} = R_{0,1} = J$, it follows immediately that

$$C(R_{i,0}R_{0,j} : i=0,\ldots,r, j=0,\ldots,s) \subset C(W^iZ^j : i=0,\ldots,r, j=0,\ldots,s).$$

However, the vectors listed in each of the sets are linearly independent and the number of vectors in each set is the same, so the ranks of the column spaces are the same and

$$C(R_{i,0}R_{0,j} : i=0,\ldots,r, j=0,\ldots,s) = C(W^iZ^j : i=0,\ldots,r, j=0,\ldots,s).$$

The vectors $R_{i,0}R_{0,j}$ are orthogonal, so $abN[R_{r,0}R_{0,s}][R_{r,0}R_{0,s}]'$ is the projection operator for testing if W^rZ^s adds to the model after fitting all terms of lower order. Since $R_{r,0}R_{0,s}$ was found as the vector in the interaction space corresponding to the rth orthogonal polynomial contrast in the α_is and the sth orthogonal polynomial contrast in the η_js, the technique for testing if W^rZ^s adds to the appropriate model is a straightforward test of an interaction contrast.

7.4 Two-Way ANOVA with Proportional Numbers

Consider the model

$$y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk},$$

$i=1,\ldots,a$, $j=1,\ldots,b$, $k=1,\ldots,N_{ij}$. We say that the model has *proportional numbers* if, for $i,i'=1,\ldots,a$ and $j,j'=1,\ldots,b$,

$$N_{ij}/N_{ij'} = N_{i'j}/N_{i'j'}.$$

The special case of $N_{ij} = N$ for all i and j is the balanced two-way ANOVA.

The analysis with proportional numbers is in the same spirit as the analysis with balance presented in Section 1. After fitting the mean, μ, the column space for the

α treatments and the column space for the η treatments are orthogonal. Before showing this we need a result on the N_{ij}s.

Lemma 7.4.1. If the N_{ij}s are proportional, then for any $r = 1,\ldots,a$, and $s = 1,\ldots,b$,

$$N_{rs} = N_{r\cdot}N_{\cdot s}/N_{\cdot\cdot}\,.$$

PROOF. Because the numbers are proportional

$$N_{ij}N_{rs} = N_{rj}N_{is}\,.$$

Summing over i and j yields

$$N_{rs}N_{\cdot\cdot} = N_{rs}\sum_{i=1}^{a}\sum_{j=1}^{b}N_{ij} = \sum_{i=1}^{a}\sum_{j=1}^{b}N_{rj}N_{is} = N_{r\cdot}N_{\cdot s}\,.$$

Dividing by $N_{\cdot\cdot}$ gives the result. □

As in Section 1, we can write the model matrix as $X = [X_0, X_1, \ldots, X_{a+b}]$, where $X_0 = J_n^1$,

$$\begin{aligned} X_r &= [t_{ijk}], & t_{ijk} &= \delta_{ir}, & r &= 1,\ldots,a; \\ X_{a+s} &= [u_{ijk}], & u_{ijk} &= \delta_{js}, & s &= 1,\ldots,b. \end{aligned}$$

Orthogonalizing with respect to J_n^1 gives

$$\begin{aligned} Z_r &= X_r - \frac{N_{r\cdot}}{N_{\cdot\cdot}}J, & r &= 1,\ldots,a; \\ Z_{a+s} &= X_{a+s} - \frac{N_{\cdot s}}{N_{\cdot\cdot}}J, & s &= 1,\ldots,b. \end{aligned}$$

It is easily seen that for $r = 1,\ldots,a$, $s = 1,\ldots,b$,

$$Z_{a+s}'Z_r = N_{rs} - N_{\cdot s}\frac{N_{r\cdot}}{N_{\cdot\cdot}} - N_{r\cdot}\frac{N_{\cdot s}}{N_{\cdot\cdot}} + N_{\cdot\cdot}\frac{N_{r\cdot}}{N_{\cdot\cdot}}\frac{N_{\cdot s}}{N_{\cdot\cdot}} = 0.$$

Exercise 7.3 Find the ANOVA table for the two-way ANOVA without interaction model when there are proportional numbers. Find the least squares estimate of a contrast in the α_is. Find the variance of the contrast and give a definition of orthogonal contrasts that depends only on the contrast coefficients and the N_{ij}s. If the α_is correspond to levels of a quantitative factor, say x_is, find the linear contrast.

The analysis when interaction is included is similar. It is based on repeated use of the one-way ANOVA.

Exercise 7.4 Using proportional numbers, find the ANOVA table for the two-way ANOVA with interaction model.

7.5 Two-Way ANOVA with Unequal Numbers: General Case

Without balance or proportional numbers, there is no simplification of the model, so, typically, $R(\alpha|\mu,\eta) \neq R(\alpha|\mu)$ and $R(\eta|\mu,\alpha) \neq R(\eta|\mu)$. We are forced to analyze the model on the basis of the general theory alone.

First consider the two-way ANOVA without interaction

$$y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk},$$

$i = 1,\dots,a,\ j = 1,\dots,b,\ k = 1,\dots,N_{ij}$. Note that we have not excluded the possibility that $N_{ij} = 0$.

One approach to analyzing a two-way ANOVA is by model selection. Consider $R(\alpha|\mu,\eta)$ and $R(\eta|\mu,\alpha)$. If both of these are large, the model is taken as

$$y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}.$$

If, say, $R(\alpha|\mu,\eta)$ is large and $R(\eta|\mu,\alpha)$ is not, then the model

$$y_{ijk} = \mu + \alpha_i + e_{ijk}$$

should be appropriate; however, a further test of α based on $R(\alpha|\mu)$ may give contradictory results. If $R(\alpha|\mu)$ is small and $R(\alpha,\eta|\mu)$ is large, we have a problem: no model seems to fit. The treatments, α and η, are together having an effect, but neither seems to be helping individually. A model with μ and α is inappropriate because $R(\alpha|\mu)$ is small. A model with μ and η is inappropriate because if μ and η are in the model and $R(\alpha|\mu,\eta)$ is large, we need α also. However, the model with μ, α, and η is inappropriate because $R(\eta|\mu,\alpha)$ is small. Finally, the model with μ alone is inappropriate because $R(\alpha,\eta|\mu)$ is large. Thus, every model that we can consider is inappropriate by some criterion. If $R(\alpha,\eta|\mu)$ had been small, the best choice probably would be a model with only μ; however, some would argue that all of μ, α, and η should be included on the basis of $R(\alpha|\mu,\eta)$ being large.

Fortunately, it is difficult for these situations to arise. Suppose $R(\alpha|\mu,\eta) = 8$ and $R(\alpha|\mu) = 6$, both with 2 degrees of freedom. Let $R(\eta|\mu,\alpha) = 10$ with 4 degrees of freedom, and $MSE = 1$ with 30 degrees of freedom. The 0.05 test for α after μ and η is

$$[8/2]/1 = 4 > 3.32 = F(0.95,2,30).$$

The test for α after μ is

$$[6/2]/1 = 3 < 3.32 = F(0.95,2,30).$$

The test for η after μ and α is

$$[10/4]/1 = 2.5 < 2.69 = F(0.95, 4, 30).$$

The test for α and η after μ is based on $R(\alpha, \eta|\mu) = R(\alpha|\mu) + R(\eta|\mu, \alpha) = 6 + 10 = 16$ with $2 + 4 = 6$ degrees of freedom. The test is

$$[16/6]/1 = 2.67 > 2.42 = F(0.95, 6, 30).$$

Although the tests are contradictory, the key point is that the P values for all four tests are about 0.05. For the first and last tests, the P values are just below 0.05; for the second and third tests, the P values are just above 0.05. The real information is not which tests are rejected and which are not rejected; the valuable information is that all four P values are approximately 0.05. All of the sums of squares should be considered to be either significantly large or not significantly large.

Because of the inflexible nature of hypothesis tests, we have chosen to discuss sums of squares that are either large or small, without giving a precise definition of what it means to be either large or small. The essence of any definition of large and small should be that the sum of two large quantities should be large and the sum of two small quantities should be small. For example, since $R(\alpha, \eta|\mu) = R(\alpha|\mu) + R(\eta|\mu, \alpha)$, it should be impossible to have $R(\alpha|\mu)$ small and $R(\eta|\mu, \alpha)$ small, but $R(\alpha, \eta|\mu)$ large. This consistency can be achieved by the following definition that exchanges one undefined term for another: *Large means significant or nearly significant*.

With this approach to the terms large and small, the contradiction alluded to above does not exist. The contradiction was based on the fact that with $R(\alpha|\mu, \eta)$ large, $R(\eta|\mu, \alpha)$ small, $R(\alpha|\mu)$ small, and $R(\alpha, \eta|\mu)$ large, no model seemed to fit. However, this situation is impossible. If $R(\eta|\mu, \alpha)$ is small and $R(\alpha|\mu)$ is small, then $R(\alpha, \eta|\mu)$ must be small; and the model with μ alone fits. (Note that since $R(\alpha|\mu, \eta)$ is large but $R(\alpha, \eta|\mu)$ is small, we must have $R(\eta|\mu)$ small; otherwise, two large quantities would add up to a small quantity.)

I find the argument of the previous paragraph convincing. Having built a better mousetrap, I expect the world to beat a path to my door. Unfortunately, I suspect that when the world comes, it will come carrying tar and feathers. It is my impression that in this situation most statisticians would prefer to use the model with all of μ, α, and η. In any case, the results are sufficiently unclear that further data collection would probably be worthwhile.

Table 7.3 contains some suggested model choices based on various sums of squares. Many of the suggestions are potentially controversial; these are indicated by asterisks.

The example that has been discussed throughout this section is the case in the fourth row and second column of Table 7.3. The entry in the second row and fourth column is similar.

The entry in the second row and second column is the case where each effect is important after fitting the other effect, but neither is important on its own. My inclination is to choose the model based on the results of examining $R(\alpha, \eta|\mu)$. On

Table 7.3 Suggested Model Selections: Two-Way Analysis of Variance Without Interaction (Unequal Numbers).

Table Entries Are Models as Numbered Below						
		$R(\alpha\|\mu,\eta)$	L		S	
$R(\eta\|\mu,\alpha)$	$R(\alpha\|\mu)$	$R(\eta\|\mu)$	L	S	L	S
L	L		1	1	3*	I
	S		1	4,1*	3	4,1*
S	L		2*	2	1,2,3*	2
	S		I	4,1*	3	4

L indicates that the sum of squares is large.
S indicates that the sum of squares is small.
I indicates that the sum of squares is impossible.
* indicates that the model choice is debatable.

Models:
1 $y_{ijk} = \mu + \alpha_i + \eta_j + e_{ijk}$
2 $y_{ijk} = \mu + \alpha_i + e_{ijk}$
3 $y_{ijk} = \mu + \eta_j + e_{ijk}$
4 $y_{ijk} = \mu + e_{ijk}$

the other hand, it is sometimes argued that since the full model gives no basis for dropping either α or η individually, the issue of dropping them both should not be addressed.

The entry in the third row and third column is very interesting. Each effect is important on its own, but neither effect is important after fitting the other effect. The corresponding models (2 and 3 in the table) are not hierarchical, i.e., neither column space is contained in the other, so there is no way of testing which is better. From a testing standpoint, I can see no basis for choosing the full model; but since both effects have been shown to be important, many argue that both effects belong in the model. One particular argument is that the structure of the model matrix is hiding what would otherwise be a significant effect. As will be seen in Chapter 15, with collinearity in the model matrix that is quite possible.

The entries in the third row, first column and first row, third column are similar. Both effects are important by themselves, but one of the effects is not important after fitting the other. Clearly, the effect that is always important must be in the model. The model that contains only the effect for which both sums of squares are large is an adequate substitute for the full model. On the other hand, the arguments in favor of the full model from the previous paragraph apply equally well to this situation.

Great care needs to be used in all the situations where the choice of model is unclear. With unequal numbers, the possibility of collinearity in the model matrix (see Chapter 15) must be dealt with. If collinearity exists, it will affect the conclusions that can be made about the models. Of course, when the conclusions to be drawn are questionable, additional data collection would seem to be appropriate.

The discussion of model selection has been predicated on the assumption that there is no interaction. Some of the more bizarre situations that come up in model selection are more likely to arise if there is an interaction that is being ignored. A test for interaction should be made whenever possible. Of course, just because the test gives no evidence of interaction does not mean that interaction does not exist, or even that it is small enough so that it will not affect model selection.

Finally, it should be recalled that model selection is not the only possible goal. One may accept the full model and only seek to interpret it. For the purpose of interpreting the full model, $R(\alpha|\mu)$ and $R(\eta|\mu)$ are not very enlightening. In terms of the full model, the hypotheses that can be tested with these sums of squares are complicated functions of both the αs and the ηs. The exact nature of these hypotheses under the full model can be obtained from the formulae given below for the model with interaction.

We now consider the two-way model with interaction. The model can be written

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk}, \tag{1}$$

$i = 1,\ldots,a,\ j = 1,\ldots,b,\ k = 1,\ldots,N_{ij}$. However, for most purposes, we do not recommend using this parameterization of the model. The full rank cell means parameterization

$$y_{ijk} = \mu_{ij} + e_{ijk} \tag{2}$$

is much easier to work with. The interaction model has the column space of a one-way ANOVA with unequal numbers.

Model (1) lends itself to two distinct orthogonal breakdowns of the sum of squares for the model. These are

$$R(\mu),\ R(\alpha|\mu),\ R(\eta|\mu,\alpha),\ R(\gamma|\mu,\alpha,\eta)$$

and

$$R(\mu),\ R(\eta|\mu),\ R(\alpha|\mu,\eta),\ R(\gamma|\mu,\alpha,\eta).$$

If $R(\gamma|\mu,\alpha,\eta)$ is small, one can work with the reduced model. If $R(\gamma|\mu,\alpha,\eta)$ is large, the full model must be retained. Just as with the balanced model, the F tests for α and η test hypotheses that involve the interactions. Using the parameterization of model (2), the hypothesis associated with the test using $R(\alpha|\mu)$ is that for all i and i',

$$\sum_{j=1}^{b} N_{ij}\mu_{ij}/N_{i\cdot} = \sum_{j=1}^{b} N_{i'j}\mu_{i'j}/N_{i'\cdot}\,.$$

The hypothesis associated with the test using $R(\alpha|\mu,\eta)$ is that for all i,

$$\sum_{j=1}^{b} N_{ij}\mu_{ij} - \sum_{i'=1}^{a}\sum_{j=1}^{b} N_{ij}N_{i'j}\mu_{i'j}/N_{\cdot j} = 0.$$

Since $\mu_{ij} = \mu + \alpha_i + \eta_j + \gamma_{ij}$, the formulae above can be readily changed to involve the alternative parameterization. Similarly, by dropping the γ_{ij}, one can get the hypotheses associated with the sums of squares in the no interaction model. Neither of these hypotheses appears to be very interesting. The first of them is the hypothesis of equality of the weighted averages, taken over j, of the μ_{ij}s. It is unclear why one should be interested in weighted averages of the μ_{ij}s when the weights are the sample sizes. For many purposes, a more reasonable hypothesis would seem to be that the simple averages, taken over j, are all equal. The second hypothesis above is almost uninterpretable. In terms of model selection, testing for main effects when there is interaction does not make much sense. The reduced model corresponding to such a test is not easy to find, so it is not surprising that the parametric hypotheses corresponding to these tests are not very interpretable.

A better idea seems to be to choose tests based directly on model (2). For example, the hypothesis of equality of the simple averages $\bar{\mu}_{i\cdot}$ (i.e., $\bar{\mu}_{i\cdot} = \bar{\mu}_{i'\cdot}$ for $i,i' = 1,\ldots,a$) is easily tested using the fact that the interaction model is really just a one-way ANOVA with unequal numbers.

At this point we should say a word about computational techniques, or rather explain why we will not discuss computational techniques. The difficulty in computing sums of squares for a two-way ANOVA with interaction and unequal numbers lies in computing $R(\alpha|\mu,\eta)$ and $R(\eta|\mu,\alpha)$. The sums of squares that are needed for the analysis are $R(\gamma|\mu,\alpha,\eta)$, $R(\alpha|\mu,\eta)$, $R(\eta|\mu,\alpha)$, $R(\alpha|\mu)$, $R(\eta|\mu)$, and SSE. SSE is the error computed as in a one-way ANOVA. $R(\alpha|\mu)$ and $R(\eta|\mu)$ are computed as in a one-way ANOVA, and the interaction sum of squares can be written as $R(\gamma|\mu,\alpha,\eta) = R(\gamma|\mu) - R(\eta|\mu,\alpha) - R(\alpha|\mu)$, where $R(\gamma|\mu)$ is computed as in a one-way ANOVA. Thus, only $R(\alpha|\mu,\eta)$ and $R(\eta|\mu,\alpha)$ present difficulties.

There are formulae available for computing $R(\alpha|\mu,\eta)$ and $R(\eta|\mu,\alpha)$, but the formulae are both nasty and of little use. With the advent of high speed computing, the very considerable difficulties in obtaining these sums of squares have vanished. Moreover, the formulae add very little to one's understanding of the method of analysis. The key idea in the analysis is that for, say $R(\eta|\mu,\alpha)$, the η effects are being fitted after the α effects. The general theory of linear models provides methods for finding the sums of squares and there is little simplification available in the special case.

Of course, there is more to an analysis than just testing treatment effects. As mentioned above, if there is evidence of interactions, probably the simplest approach is to analyze the μ_{ij} model. If there are no interactions, then one is interested in testing contrasts in the main effects. Just as in the one-way ANOVA, it is easily seen that all estimable functions of the treatment effects will be contrasts; however, there is no assurance that all contrasts will be estimable. To test a contrast, say $\alpha_1 - \alpha_2 = 0$, the simplest method is to fit a model that does not distinguish between α_1 and α_2, see Example 3.2.0 and Section 3.4. In the two-way ANOVA with unequal numbers, the model that does not distinguish between α_1 and α_2 may or may not have a different column space from that of the unrestricted model.

EXAMPLE 7.5.1. Scheffé (1959) (cf. Bailey, 1953) reports on an experiment in which infant female rats were given to foster mothers for nursing, and the weights, in grams, of the infants were measured after 28 days. The two factors in the experiment were the genotypes of the infants and the genotypes of the foster mothers. In the experiment, an entire litter was given to a single foster mother. The variability within each litter was negligible compared to the variability between different litters, so the analysis was performed on the litter averages. Table 7.4 contains the data.

Table 7.4 Infant Rats' Weight Gain with Foster Mothers.

Genotype of Litter	Genotype of Foster Mother A	F	I	J
A	61.5	55.0	52.5	42.0
	68.2	42.0	61.8	54.0
	64.0	60.2	49.5	61.0
	65.0		52.7	48.2
	59.7			39.6
F	60.3	50.8	56.5	51.3
	51.7	64.7	59.0	40.5
	49.3	61.7	47.2	
	48.0	64.0	53.0	
		62.0		
I	37.0	56.3	39.7	50.0
	36.3	69.8	46.0	43.8
	68.0	67.0	61.3	54.5
			55.3	
			55.7	
J	59.0	59.5	45.2	44.8
	57.4	52.8	57.0	51.5
	54.0	56.0	61.4	53.0
	47.0			42.0
				54.0

We use the two-way ANOVA model

$$y_{ijk} = G + L_i + M_j + [LM]_{ij} + e_{ijk}, \tag{3}$$

where L indicates the effect of the litter genotype and M indicates the effect of the foster mother genotype. Tables 7.5 and 7.6 contain, respectively, the sums of squares for error for a variety of submodels and the corresponding reductions in sums of squares for error and F tests.

Some percentiles of the F distribution that are of interest in evaluating the statistics of Table 7.6 are $F(0.9,9,45) = 1.78$, $F(0.99,3,45) = 4.25$, and $F(0.95,6,45) = 2.23$. Clearly, the litter–mother interaction and the main effect for litters can be dropped from model (3). However, the main effect for mothers is important. The smallest model that fits the data is

$$y_{ijk} = G + M_j + e_{ijk}.$$

Table 7.5 Sums of Squares Error for Fitting Models to the Data of Table 7.4.

Model	Model	SSE	df
$G+L+M+LM$	$[LM]$	2441	45
$G+L+M$	$[L][M]$	3265	54
$G+L$	$[L]$	4040	57
$G+M$	$[M]$	3329	57
G	$[G]$	4100	60

Table 7.6 F Tests for Fitting Models to the Data of Table 7.4.

Reduction in SSE		df	MS	F*
$R(LM\|L,M,G)$	$= 3265 - 2441 = 824$	9	91.6	1.688
$R(M\|L,G)$	$= 4040 - 3265 = 775$	3	258.3	4.762
$R(L\|G)$	$= 4100 - 4040 = 60$	3	20.0	0.369
$R(L\|M,G)$	$= 3329 - 3265 = 64$	3	21.3	0.393
$R(M\|G)$	$= 4100 - 3329 = 771$	3	257.	4.738
$R(L,M\|G)$	$= 4100 - 3265 = 835$	6	139.2	2.565

*All F statistics calculated using $MSE([LM]) = 2441/45$ $= 54.244$ in the denominator.

This is just a one-way ANOVA model and can be analyzed as such. By analogy with balanced two-factor ANOVAs, tests of contrasts might be best performed using $MSE([LM])$ rather than $MSE([M]) = 3329/57 = 58.40$.

Table 7.7 Mean Values for Foster Mother Genotypes.

Foster Mother	Parameter	$N_{\cdot j}$	Estimate
A	$G+M_1$	16	55.400
F	$G+M_2$	14	58.700
I	$G+M_3$	16	53.363
J	$G+M_4$	15	48.680

The mean values for the foster mother genotypes are reported in Table 7.7, along with the number of observations for each mean. It would be appropriate to continue the analysis by comparing all pairs of means. This can be done with either Scheffé's method, Bonferroni's method, or the LSD method. The LSD method with $\alpha = 0.05$ establishes that genotype J is distinct from genotypes A and F. (Genotypes F and I are almost distinct.) Bonferroni's method with $\alpha = 0.06$ establishes that J is distinct from F and that J is almost distinct from A.

Exercise 7.5 Analyze the following data as a two-factor ANOVA where the subscripts i and j indicate the two factors.

		y_{ijk}s		
	i	1	2	3
j	1	0.620	1.228	0.615
		1.342	3.762	2.245
		0.669	2.219	2.077
		0.687	4.207	3.357
		0.155		
		2.000		
	2	1.182	3.080	2.240
		1.068	2.741	0.330
		2.545	2.522	3.453
		2.233	1.647	1.527
		2.664	1.999	0.809
		1.002	2.939	1.942
		2.506		
		4.285		
		1.696		

The dependent variable is a mathematics ineptitude score. The first factor (i) identifies economics majors, anthropology majors, and sociology majors, respectively. The second factor (j) identifies whether the student's high school background was rural (1) or urban (2).

Exercise 7.6 Analyze the following data as a two-factor ANOVA where the subscripts i and j indicate the two factors.

		y_{ijk}s		
	i	1	2	3
j	1	1.620	2.228	2.999
		1.669	3.219	1.615
		1.155	4.080	
		2.182		
		3.545		
	2	1.342	3.762	2.939
		0.687	4.207	2.245
		2.000	2.741	1.527
		1.068		0.809
		2.233		1.942
		2.664		
		1.002		

The dependent variable is again a mathematics ineptitude score and the levels of the first factor identify the same three majors as in Exercise 7.5. In these data, the second factor identifies whether the student is lefthanded (1) or righthanded (2).

7.6 Three or More Way Analyses

With balanced or proportional numbers, the analyses for more general models follow the same patterns as those of the two-way ANOVA. (Proportional numbers can be defined in terms of complete independence in higher dimensional tables, see Fienberg (1980) or Christensen (1997).) Consider, for example, the model

$$y_{ijkm} = \mu + \alpha_i + \eta_j + \gamma_k + (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + (\eta\gamma)_{jk} + (\alpha\eta\gamma)_{ijk} + e_{ijkm},$$

$i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$, $m = 1, \ldots, N$. The sums of squares for the main effects of α, η, and γ are based on the one-way ANOVA ignoring all other effects, e.g.,

$$SS(\eta) = acN \sum_{j=1}^{b} (\bar{y}_{\cdot j \cdot\cdot} - \bar{y}_{\cdot\cdot\cdot\cdot})^2. \tag{1}$$

The sums of squares for the two-way interactions, $\alpha\eta$, $\alpha\gamma$, and $\eta\gamma$, are obtained as in the two-way ANOVA by ignoring the third effect, e.g.,

$$SS(\alpha\gamma) = bN \sum_{i=1}^{a} \sum_{k=1}^{c} (\bar{y}_{i\cdot k\cdot} - \bar{y}_{i\cdot\cdot\cdot} - \bar{y}_{\cdot\cdot k\cdot} + \bar{y}_{\cdot\cdot\cdot\cdot})^2. \tag{2}$$

The sum of squares for the three-way interaction $\alpha\eta\gamma$ is found by subtracting all of the other sums of squares (including the grand mean's) from the sum of squares for the full model. Note that the full model is equivalent to the one-way ANOVA model

$$y_{ijkm} = \mu_{ijk} + e_{ijkm}.$$

Sums of squares and their associated projection operators are defined from reduced models. For example, M_η is the perpendicular projection operator for fitting the ηs after μ in the model

$$y_{ijkm} = \mu + \eta_j + e_{ijkm}.$$

The subscripts i, k, and m are used to indicate the replications in this one-way ANOVA. $SS(\eta)$ is defined by

$$SS(\eta) = Y'M_\eta Y.$$

The algebraic formula for $SS(\eta)$ was given in (1). Similarly, $M_{\alpha\gamma}$ is the projection operator for fitting the $(\alpha\gamma)$s after μ, the αs, and the γs in the model

$$y_{ijkm} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + e_{ijkm}.$$

In this model, the subscripts j and m are used to indicate replication. The sum of squares $SS(\alpha\gamma)$ is

$$SS(\alpha\gamma) = Y'M_{\alpha\gamma}Y.$$

The algebraic formula for $SS(\alpha\gamma)$ was given in (2). Because all of the projection operators (except for the three-factor interaction) are defined on the basis of reduced models that have previously been discussed, the sums of squares take on the familiar forms indicated above.

The one new aspect of the model that we are considering is the inclusion of the three-factor interaction. As mentioned above, the sum of squares for the three-factor interaction is just the sum of squares that is left after fitting everything else, i.e.,

$$SS(\alpha\eta\gamma) = R((\alpha\eta\gamma)|\mu, \alpha, \eta, \gamma, (\alpha\eta), (\alpha\gamma), (\eta\gamma)).$$

The space for the three-factor interaction is the orthogonal complement (with respect to the space for the full model) of the design space for the model that includes all factors except the three-factor interaction. Thus, the space for the three-factor interaction is orthogonal to everything else. (This is true even when the numbers are not balanced.)

In order to ensure a nice analysis, we need to show that the spaces associated with all of the projection operators are orthogonal and that the projection operators add up to the perpendicular projection operator onto the space for the full model. First, we show that $C(M_\mu, M_\alpha, M_\eta, M_\gamma, M_{\alpha\eta}, M_{\alpha\gamma}, M_{\eta\gamma})$ is the column space for the model

$$y_{ijkm} = \mu + \alpha_i + \eta_j + \gamma_k + (\alpha\eta)_{ij} + (\alpha\gamma)_{ik} + (\eta\gamma)_{jk} + e_{ijkm},$$

and that all the projection operators are orthogonal. That the column spaces are the same follows from the fact that the column space of the model without the three-factor interaction is precisely the column space obtained by combining the column spaces of all of the two-factor with interaction models. Combining the spaces of the projection operators is precisely combining the column spaces of all the two-factor with interaction models. That the spaces associated with all of the projection operators are orthogonal follows easily from the fact that all of the spaces come from reduced models. For the reduced models, characterizations have been given for the various spaces. For example,

$$C(M_\eta) = \{v|v = [v_{ijkm}], \text{ where } v_{ijkm} = d_j \text{ for some } d_1, \ldots, d_b \text{ with } d_\cdot = 0\}.$$

Similarly,

$$C(M_{\alpha\gamma}) = \{w|w = [w_{ijkm}], \text{ where } w_{ijkm} = r_{ik} \text{ for some } r_{ik}$$
$$\text{with } r_{i\cdot} = r_{\cdot k} = 0 \text{ for } i = 1, \ldots, a, k = 1, \ldots, c\}.$$

With these characterizations, it is a simple matter to show that the projection operators define orthogonal spaces.

Let M denote the perpendicular projection operator for the full model. The projection operator onto the interaction space, $M_{\alpha\eta\gamma}$, has been defined as

$$M_{\alpha\eta\gamma} = M - [M_\mu + M_\alpha + M_\eta + M_\gamma + M_{\alpha\eta} + M_{\alpha\gamma} + M_{\eta\gamma}];$$

thus,

$$M = M_\mu + M_\alpha + M_\eta + M_\gamma + M_{\alpha\eta} + M_{\alpha\gamma} + M_{\eta\gamma} + M_{\alpha\eta\gamma},$$

where the spaces of all the projection operators on the right side of the equation are orthogonal to each other.

Exercise 7.7 Show that $C(M_\eta) \perp C(M_{\alpha\gamma})$ and that $C(M_{\eta\gamma}) \perp C(M_{\alpha\gamma})$. Give an explicit characterization of a typical vector in $C(M_{\alpha\eta\gamma})$ and show that your characterization is correct.

If the α, η, and γ effects correspond to quantitative levels of some factor, the three-way ANOVA corresponds to a polynomial in three variables. The main effects and the two-way interactions can be dealt with as before. The three-way interaction can be broken down into contrasts such as the linear-by-linear-by-quadratic.

For unequal numbers, the analysis can be performed by comparing models.

EXAMPLE 7.6.1. Table 7.8 below is derived from Scheffé (1959) and gives the moisture content (in grams) for samples of a food product made with three kinds of salt (A), three amounts of salt (B), and two additives (C). The amounts of salt, as measured in moles, are equally spaced. The two numbers listed for some treatment combinations are replications. We wish to analyze these data.

Table 7.8 Moisture Content of a Food Product.

A (salt)		1			2			3		
B (amount salt)		1	2	3	1	2	3	1	2	3
	1	8	17	22	7	26	34	10	24	39
			13	20	10	24		9		36
C (additive)										
	2	5	11	16	3	17	32	5	16	33
		4	10	15	5	19	29	4		34

We will consider these data as a three-factor ANOVA. From the structure of the replications, the ANOVA has unequal numbers. The general model for a three-factor ANOVA with replications is

$$y_{ijkm} = G + A_i + B_j + C_k + [AB]_{ij} + [AC]_{ik} + [BC]_{jk} + [ABC]_{ijk} + e_{ijkm}.$$

Our first priority is to find out which interactions are important. Table 7.9 contains the sum of squares for error and the degrees of freedom for error for all models that include all of the main effects. Each model is identified in the table by the highest order terms in the model (cf. Table 7.5, Section 5). Readers familiar with methods for fitting log-linear models (cf. Fienberg, 1980 or Christensen, 1997) will notice a correspondence between Table 7.9 and similar displays used in fitting three-dimensional contingency tables. The analogies between selecting log-linear models and selecting models for unbalanced ANOVA are pervasive.

All of the models have been compared to the full model using F statistics in Table 7.9. It takes neither a genius nor an F table to see that the only models that fit the data are the models that include the $[AB]$ interaction. A number of other comparisons can be made among models that include $[AB]$. These are $[AB][AC][BC]$ versus $[AB][AC]$, $[AB][AC][BC]$ versus $[AB][BC]$, $[AB][AC][BC]$ versus $[AB][C]$, $[AB][AC]$ versus $[AB][C]$, and $[AB][BC]$ versus $[AB][C]$. None of the comparisons show any lack of fit. The last two comparisons are illustrated below.

Table 7.9 Statistics for Fitting Models to the Data of Table 7.8.

Model	SSE	df	F*
$[ABC]$	32.50	14	—
$[AB][AC][BC]$	39.40	18	0.743
$[AB][AC]$	45.18	20	0.910
$[AB][BC]$	40.46	20	0.572
$[AC][BC]$	333.2	22	16.19
$[AB][C]$	45.75	22	0.713
$[AC][B]$	346.8	24	13.54
$[BC][A]$	339.8	24	13.24
$[A][B][C]$	351.1	26	11.44

*The F statistics are for testing each model against the model with a three-factor interaction, i.e., $[ABC]$. The denominator of each F statistic is $MSE([ABC]) = 32.50/14 = 2.3214$.

$$[AB][AC] \text{ versus } [AB][C]:$$
$$R(AC|AB,C) = 45.75 - 45.18 = 0.57,$$
$$F = (0.57/2)/2.3214 = 0.123.$$

$$[AB][BC] \text{ versus } [AB][C]:$$
$$R(BC|AB,C) = 45.75 - 40.46 = 5.29,$$
$$F = (5.29/2)/2.3214 = 1.139.$$

Note that, by analogy to the commonly accepted practice for balanced ANOVAs, all tests have been performed using $MSE([ABC])$, that is, the estimate of pure error from the full model.

The smallest model that seems to fit the data adequately is $[AB][C]$. The F statistics for comparing $[AB][C]$ to the larger models are all extremely small. Writing out the model $[AB][C]$, it is

$$y_{ijkm} = G + A_i + B_j + C_k + [AB]_{ij} + e_{ijkm}.$$

We need to examine the $[AB]$ interaction. Since the levels of B are quantitative, a model that is equivalent to $[AB][C]$ is a model that includes the main effects for C but, instead of fitting an interaction in A and B, fits a separate regression equation in the levels of B for each level of A. Let x_j, $j = 1,2,3$, denote the levels of B. There are three levels of B, so the most general polynomial we can fit is a second-degree polynomial in x_j. Since the levels of salt were equally spaced, it does not matter much what we use for the x_js. The computations were performed using $x_1 = 1$, $x_2 = 2$, $x_3 = 3$. In particular, the model $[AB][C]$ was reparameterized as

$$y_{ijkm} = A_{i0} + A_{i1}x_j + A_{i2}x_j^2 + C_k + e_{ijkm}. \tag{3}$$

With a notation similar to that used in Table 7.9, the *SSE* and the *dfE* are reported in Table 7.10 for model (3) and three reduced models.

Table 7.10 Additional Statistics for Fitting Models to the Data of Table 7.8.

Model	*SSE*	*df*
$[A_0][A_1][A_2][C]$	45.75	22
$[A_0][A_1][C]$	59.98	25
$[A_0][A_1]$	262.0	26
$[A_0][C]$	3130.	28

Note that the *SSE* and *df* reported in Table 7.10 for $[A_0][A_1][A_2][C]$ are identical to the values reported in Table 7.9 for $[AB][C]$. This, of course, must be true if the models are merely reparameterizations of one another. First we want to establish whether the quadratic effects are necessary in the regressions. To do this we test

$$[A_0][A_1][A_2][C] \text{ versus } [A_0][A_1][C]:$$

$$R(A_2|A_1, A_0, C) = 59.98 - 45.75 = 14.23,$$
$$F = (14.23/3)/2.3214 = 2.04.$$

Since $F(0.95, 3, 14) = 3.34$, there is no evidence of any nonlinear effects.

At this point it might be of interest to test whether there is any linear effect. This is done by testing $[A_0][A_1][C]$ against $[A_0][C]$. The statistics needed for this test are given in Table 7.10. Instead of actually doing the test, recall that no models in Table 7.9 fit the data unless they included the $[AB]$ interaction. If we eliminated the linear effects, we would have a model that involved none of the $[AB]$ interaction. (The model $[A_0][C]$ is identical to the ANOVA model $[A][C]$.) We already know that such models do not fit.

Finally, we have never explored the possibility that there is no main effect for C. This can be done by testing

$$[A_0][A_1][C] \text{ versus } [A_0][A_1]:$$

$$R(C|A_1, A_0) = 262.0 - 59.98 = 202,$$
$$F = (202/1)/2.3214 = 87.$$

Obviously, there is a substantial main effect for C, the type of food additive.

Our conclusion is that the model $[A_0][A_1][C]$ is the smallest model yet considered that adequately fits the data. This model indicates that there is an effect for the type of additive and a linear relationship between amount of salt and moisture content. The slope and intercept of the line may depend on the type of salt. (The intercept of the line also depends on the type of additive.) Table 7.11 contains parameter estimates and standard errors for the model. All estimates in the example use the side condition $C_1 = 0$.

Table 7.11 Parameter Estimates and Standard Errors for the Model $y_{ijkm} = A_{i0} + A_{i1}x_j + C_k + e_{ijkm}$.

Parameter	Estimate	S.E.
A_{10}	3.350	1.375
A_{11}	5.85	0.5909
A_{20}	−3.789	1.237
A_{21}	13.24	0.5909
A_{30}	−4.967	1.231
A_{31}	14.25	0.5476
C_1	0.	none
C_2	−5.067	0.5522

Note that, in lieu of the F test, the test for the main effect C could be performed by looking at $t = -5.067/0.5522 = -9.176$. Moreover, we should have $t^2 = F$. The t statistic squared is 84, while the F statistic reported earlier is 87. The difference is due to the fact that the S.E. reported uses the MSE for the model being fitted, while in performing the F test we used the $MSE([ABC])$.

Are we done yet? No! The parameter estimates suggest some additional questions. Are the slopes for salts 2 and 3 the same, i.e., is $A_{21} = A_{31}$? In fact, are the entire lines for salts 2 and 3 the same, i.e., are $A_{21} = A_{31}$, $A_{20} = A_{30}$? We can fit models that incorporate these assumptions.

Model	SSE	df
$[A_0][A_1][C]$	59.98	25
$[A_0][A_1][C]$, $A_{21} = A_{31}$	63.73	26
$[A_0][A_1][C]$, $A_{21} = A_{31}$, $A_{20} = A_{30}$	66.97	27

It is a small matter to check that there is no lack of fit displayed by any of these models. The smallest model that fits the data is now $[A_0][A_1][C]$, $A_{21} = A_{31}$, $A_{20} = A_{30}$. Thus there seems to be no difference between salts 2 and 3, but salt 1 has a different regression than the other two salts. (We did not actually test whether salt 1 is different, but if salt 1 had the same slope as the other two, then there would be no interaction, and we know that interaction exists.) There is also an effect for the food additives. The parameter estimates and standard errors for the final model are given in Table 7.12.

Figure 7.1 shows the fitted values for the final model. The two lines for a given additive are shockingly close at $B = 1$, which makes me wonder if $B = 1$ is the condition of no salt being used. Scheffé does not say.

Are we done yet? Probably not. We have not even considered the validity of the assumptions. Are the errors normally distributed? Are the variances the same for every treatment combination? Some methods for addressing these questions are discussed in Chapter 13. Technically, we need to ask whether $C_1 = C_2$ in this new

Table 7.12 Parameter Estimates and Standard Errors for the Model $[A_0][A_1][C]$, $A_{21} = A_{31}$, $A_{20} = A_{30}$.

Parameter	Estimate	S.E.
A_{10}	3.395	1.398
A_{11}	5.845	0.6008
A_{20}	−4.466	0.9030
A_{21}	13.81	0.4078
C_1	0.	none
C_2	−5.130	0.5602

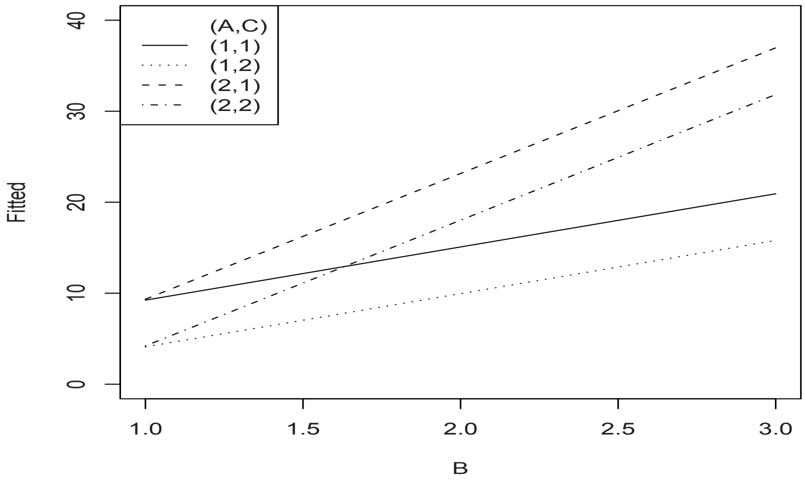

Fig. 7.1 Fitted values for Scheffé's (1959) moisture content data.

model. A quick look at the estimate and standard error for C_2 answers the question in the negative. We also have not asked whether $A_{10} = A_{20}$. Personally, I find this last question so uninteresting that I would be loath to examine it. However, a look at the estimates and standard errors suggests that the answer is no. A more interesting question is whether $A_{10} + A_{11} = A_{20} + A_{21}$, but it is pretty clear from Figure 7.1 that there will be no evidence against this hypothesis that was suggested by the data, cf. Exercise 7.7.6.

As mentioned in Example 7.6.1, the correspondences between log-linear models and unbalanced ANOVAs are legion. Often these correspondences have been overlooked. We have just considered a three-factor unbalanced ANOVA. What would we do with a four-factor or five-factor ANOVA? There is a considerable amount of literature in log-linear model theory about how to select models when there are a large number of factors. In particular, Benedetti and Brown (1978) and Christensen (1997) have surveyed strategies for selecting log-linear models. Those strategies can be applied to unbalanced ANOVAs with equal success.

I might also venture a personal opinion that statisticians tend not to spend enough time worrying about what high-dimensional ANOVA models actually mean. Log-linear model theorists do worry about interpreting their models. Wermuth (1976) has developed a method of searching among log-linear models that have nice interpretations, cf. also Christensen (1997). I believe that her method could be applied equally well to ANOVA models.

Exercise 7.8 Analyze the following three-way ANOVA: The treatments (amount of flour, brand of flour, and brand of shortening) are indicated by the subscripts i, j, and k, respectively. The dependent variable is a "chewiness" score for chocolate chip cookies. The amounts of flour correspond to quantitative factors,

i	1	2	3
x_i	3.2	4.4	6.8

The data are:

j	k	i	1 ($y_{ijk}s$)	2	3	j	k	i	1 ($y_{ijk}s$)	2	3
1	1		1.620	3.228	6.615	2	1		2.282	5.080	8.240
			1.342	5.762	8.245				2.068	4.741	6.330
					8.077				3.545	4.522	9.453
											7.727
1	2		2.669	6.219	11.357	2	2		4.233	4.647	7.809
			2.687	8.207					4.664	4.999	8.942
			2.155						3.002	5.939	
			4.000						4.506		
									6.385		
									3.696		

7.7 Additional Exercises

Exercise 7.7.1 In the mid-1970s, a study on the prices of various motor oils was conducted in (what passes for) a large town in Montana. The study consisted of pricing 4 brands of oil at each of 9 stores. The data follow.

	Brand			
Store	P	H	V	Q
1	87	95	95	82
2	96	104	106	97
3	75	87	81	70
4	81	94	91	77
5	70	85	87	65
6	85	98	97	83
7	110	123	128	112
8	83	98	95	78
9	105	120	119	98

Analyze these data.

Exercise 7.7.2 An experiment was conducted to examine thrust forces when drilling under different conditions. Data were collected for four drilling speeds and three feeds. The data are given below.

	Speed			
Feed	100	250	400	550
	121	98	83	58
	124	108	81	59
0.005	104	87	88	60
	124	94	90	66
	110	91	86	56
	329	291	281	265
	331	265	278	265
0.010	324	295	275	269
	338	288	276	260
	332	297	287	251
	640	569	551	487
	600	575	552	481
0.015	612	565	570	487
	620	573	546	500
	623	588	569	497

Analyze these data.

Exercise 7.7.3 Consider the model

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk},$$

$i = 1,2,3,4$, $j = 1,2,3$, $k = 1,\ldots,N_{ij}$, where for $i \neq 1 \neq j$, $N_{ij} = N$, and $N_{11} = 2N$. This model could arise from an experimental design having α treatments of No Treatment (NT), a_1, a_2, a_3 and η treatments of NT, b_1, b_2. This gives a total of 12 treatments: NT, a_1, a_2, a_3, b_1, a_1b_1, a_2b_1, a_3b_1 b_2, a_1b_2, a_2b_2, and a_3b_2. Since NT is a control, it might be of interest to compare all of the treatments to NT. If NT is to play such an important role in the analysis, it is reasonable to take more observations on NT than on the other treatments. Find sums of squares for testing

(a) no differences between a_1, a_2, a_3,
(b) no differences between b_1, b_2,
(c) no $\{a_1, a_2, a_3\} \times \{b_1, b_2\}$ interaction,
(d) no differences between NT and the averages of a_1, a_2, and a_3 when there is interaction,
(e) no differences between NT and the average of a_1, a_2, and a_3 when there is no interaction present,
(f) no differences between NT and the average of b_1 and b_2 when there is interaction,
(g) no differences between NT and the average of b_1 and b_2 when there is no interaction present.

Discuss the orthogonality relationships among the sums of squares. For parts (e) and (g), use the assumption of no interaction. Do not just repeat parts (d) and (f)!

Exercise 7.7.4 Consider the linear model $y_{ij} = \mu + \alpha_i + \eta_j + e_{ij}$, $i = 1, \ldots, a$, $j = 1, \ldots, b$. As in Section 1, write $X = [X_0, X_1, \ldots, X_a, X_{a+1}, \ldots, X_{a+b}]$. If we write the observations in the usual order, we can use Kronecker products to write the model matrix. Write $X = [J, X_*, X_{**}]$, where $X_* = [X_1, \ldots, X_a]$, and $X_{**} = [X_{a+1}, \ldots, X_{a+b}]$. Using Kronecker products, $X_* = [I_a \otimes J_b]$, and $X_{**} = [J_a \otimes I_b]$. In fact, with $n = ab$, $J = J_n = [J_a \otimes J_b]$. Use Kronecker products to show that $X_*'(I - [1/n]J_n^n)X_{**} = 0$. In terms of Section 1, this is the same as showing that $C(Z_1, \ldots, Z_a) \perp C(Z_{a+1}, \ldots, Z_{a+b})$. Also show that $[(1/a)J_a^a \otimes I_b]$ is the perpendicular projection operator onto $C(X_{**})$ and that $M_\eta = [(1/a)J_a^a \otimes (I_b - (1/b)J_b^b)]$.

Exercise 7.7.5 Consider the balanced two-way ANOVA with interaction model $y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + e_{ijk}$, $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, N$, with e_{ijk}s independent $N(0, \sigma^2)$. Find $\mathrm{E}[Y'(\frac{1}{n}J_n^n + M_\alpha)Y]$ in terms of μ, the α_is, the η_js, and the γ_{ij}s.

Exercise 7.7.6 For Example 7.6.1, develop a test for $H_0: A_{10} + A_{11} = A_{20} + A_{21}$.