CHAPTER 23

Multiple Regression Applied to Analysis of Variance Problems

23.1. INTRODUCTION

Many carefully designed experiments fall into the category of *analysis of variance problems.* Although some of these are, in fact, regression problems, they usually are not discussed in that way. Special computer programs exist for analyzing the data that are collected. The methods of analysis depend on special linear models that are traditionally adopted for particular designs.

Analysis of variance models can involve *fixed effects*, *random effects*, or a combination of both types of effects (when the model is usually referred to as a *mixed model*). We shall be concerned here only with the connections between regression and fixed effects analysis of variance models. For the other types of analysis of variance problems see, for example, Hocking (1996, Chapters 15–17).

Fixed Effects, Variable Effects

Suppose we have I groups of data defined, for example, by I different levels of a nutritional diet. Suppose the ith data group, $i = 1, 2, \ldots, I$, consisted of observations $Y_{i1}, Y_{i2}, \ldots, Y_{iJ_i}$ with average value $\overline{Y}_i$, and that this $\overline{Y}_i$ estimated the theoretical response θ_i (say). Then we could write

$$Y_{ij} = \theta_i + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma^2), \text{ iid} \tag{23.1.1}$$

to indicate that the diet levels had different effects $\theta_1, \theta_2, \ldots, \theta_I$, and that individual Y_{ij} varied about θ_i with errors that were independently and identically distributed (iid) normally, with mean zero and variance σ^2. Here, we would typically think of the θ_i as being *fixed effects*, each associated specifically with one of the diet levels.

In other applications, the θ_i of (23.1.1) could be *random effects*, that is, random variables. Suppose, for example, that the different i-values represented simply different batches of raw material. Then the specific θ_i would not be of interest in themselves, because another set of batches in a subsequent experiment would provide different θ_i. What *might* be of interest is the size of the variation in the θ_i. We would now most likely tentatively assume that $\theta_i \sim N(\mu, \sigma_\theta^2)$ and then estimate μ and σ_θ^2.

We link our standard regression methods to the fixed effects model in what follows. Analysis of variance problems that involve fixed effects *could* be handled by a general regression program. We do *not* recommend actually doing this in general, particularly

where all the data have become available, and where there was no hitch in carrying out the design. It is useful to know how to apply regression methods directly, however, for a number of reasons:

1. Using regression methods makes us think more carefully about the model being assumed. Because analysis of variance models typically have more parameters in them than can be estimated (they are *overparameterized*), the singular regressions that arise must be circumvented if regression methods are applied. Dummy variables can be selected in a variety of ways and some selections may be better than others, computationally, although all valid selections of dummies will lead to the same predictive values for the response.
2. Analysis of variance problems provide fitted values and residuals. Often, insufficient emphasis is placed on examining the residuals in analysis of variance problems.
3. If the data are incomplete, regression methods can still be applied, even if the neater analysis of variance methods cannot be. So regression provides a backup analysis for partially botched experimental designs.
4. Additional predictor variables (usually called *covariates*) can be added to the analysis of variance structure via regression. For example, in animal experiments where "final weight" is the response, the variable "initial weight" is often included as a covariate. This would be a continuous covariate addition. Or a class variable with several categories can be adjoined by using a suitable set of dummy variables to represent the discrete levels of the covariate. Orthogonal polynomials (see Chapter 22) might be useful for equally spaced classes and might lead to a low-order polynomial fit for the effects of the class variable.

In subsequent sections, we first discuss the one-way classification, and later the two-way classification, with equal numbers of observations in each cell, using a practical example with data for each case to give the reader some feel for the considerations involved and the advantages and disadvantages of the regression approach. Theoretical presentations for more general one- and equal-celled two-way classifications are also provided, together with another two-way, equal-celled example. Essentially, the procedures needed for the regression treatments are all choices of suitable dummy variable systems. As always, dummy variables can be chosen in an infinite number of ways.

23.2. THE ONE-WAY CLASSIFICATION: STANDARD ANALYSIS AND AN EXAMPLE

Suppose we have data in I groups, $i = 1, 2, \ldots, I$, with J_i observations in each group as given below:

Group 1	$Y_{11}, Y_{12}, \ldots, Y_{1J_1}$,	mean $\overline{Y}_1$
Group 2	$Y_{21}, Y_{22}, \ldots, Y_{2J_2}$,	mean $\overline{Y}_2$
$\vdots$		
Group I	$Y_{I1}, Y_{I2}, \ldots, Y_{IJ_I}$,	mean $\overline{Y}_I$

TABLE 23.1. Standard ANOVA Table for a One-Way Classification

Source	df	SS	MS
Between groups	$I - 1$	$\sum_{i=1}^{I} J_i(\overline{Y}_i - \overline{Y})^2$	s_B^2
Within groups	$\sum_{i=1}^{I} (J_i - 1) = n - I$	$\sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_i)^2$	s_W^2
Mean	1	$n\overline{Y}^2$	
Total	$\sum_{i=1}^{I} J_i = n$	$\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}^2$	

The usual fixed-effects analysis of variance model for such a situation is

$$Y_{ij} = \mu + t_i + \epsilon_{ij}, \qquad \begin{matrix} i = 1, 2, \ldots, I \\ j = 1, 2, \ldots, J_i \end{matrix} \tag{23.2.1}$$

where $t_1, t_2, \ldots, t_I$ are parameters such that

$$J_1 t_1 + J_2 t_2 + \cdots + J_I t_I = 0, \tag{23.2.2}$$

and where $\epsilon_{ij} \sim N(0, \sigma^2)$, iid. Some sort of restriction on the parameters is necessary since Eq. (23.2.1) contains more parameters than are really needed. It is usual to regard μ as the overall mean level and t_i as the difference between the ith group mean and the overall mean level. Thus the total of all differences between groups and the overall level is zero, and that is what Eq. (23.2.2) expresses. The usual analysis of variance table (where $\overline{Y} = \Sigma_{i=1}^{I} J_i \overline{Y}_i / \Sigma_{i=1}^{I} J_i$) is given in Table 23.1.

It is usual to test the hypothesis that there are no differences between means of groups; that is, H_0: $t_1 = t_2 = \cdots = t_I = 0$ by comparing the ratio $F = s_B^2/s_W^2$ to a suitable percentage point of the $F(I - 1, \Sigma_{i=1}^{I}(J_i - 1))$ distribution.

An Example

Caffeine, orally ingested, is said to be a stimulant, the amount of stimulation and its basic variability differing according to the dose ingested. In order to get some idea of the effect of caffeine on a physical task, the following simple experiment was performed.

1. *The Experiment.* Three treatment levels, 0, 100, and 200, in milligrams (mg) of caffeine were used. Thirty healthy male college students of the same age and with essentially the same physical ability were selected and trained in finger tapping. After the training was completed, ten men were randomly assigned to each treatment. Neither the men nor the physiologist knew which treatment the men received; only the statistician knew this. Two hours after the treatment was administered, each man was required to do finger tapping. The number of finger taps per minute was recorded, as shown in Table 23.2.

2. *The Analysis of Variance (ANOVA) Model for This Experiment.* Let:

Y_{ij} = the number of finger taps per minute of the "jth" man on the "ith" treatment.

μ = true value for the average number of finger taps in a population of males of which the selected 30 form a random sample.

TABLE 23.2. Number of Finger Taps Per Minute Achieved by Thirty Male Students Receiving Designated Doses of Caffeine

		Treatment (Row)		
Treatments, i	Observations	Totals $T_i = \sum_j Y_{ij}$	Means $\overline{Y}_i$	Effects $\hat{t}_i = \overline{Y}_i - \overline{Y}$
$i = 1$, 0 mg caffeine (placebo)	242, 245, 244, 248, 247, 248, 242, 244, 246, 242	2448	244.8	−1.7
$i = 2$, 100 mg caffeine	248, 246, 245, 247, 248, 250, 247, 246, 243, 244	2464	246.4	−0.1
$i = 3$, 200 mg caffeine	246, 248, 250, 252, 248, 250, 246, 248, 245, 250	2483	248.3	1.8
		$\sum_i \sum_j Y_{ij} = 7395$	246.5 = $\overline{Y}$	

t_i = the ith treatment effect, that is, the additive effect of the "ith" treatment over and above (or below) μ. To preserve μ as the true mean of the overall sample we must assume that $t_1 + t_2 + t_3 = 0$. [All J_i of Eq. (23.2.2) are equal.]

ϵ_{ij} = the random effect; this is a measure of the failure of student j tested with treatment i to have done exactly $\mu + t_i$ number of taps, due to random error.

With these definitions, the ANOVA model is

$$Y_{ij} = \mu + t_i + \epsilon_{ij}, \tag{23.2.3}$$

and we make the usual normality assumptions on the errors.

3. *The Standard ANOVA Calculations.* The usual ANOVA calculations lead to Table 23.3. The split indicated by the asterisks is explained below.

4. *Test for Equality of Treatments.* To test H_0: $t_1 = t_2 = t_3$ versus H_1: not so, we compare $F = 6.18$ with $F(2, 27, 0.95) = 3.35$ and so reject H_0. This indicates differences between the treatments.

TABLE 23.3. Analysis of Variance Table for the Finger Taps Example

Source of Variation	df	SS	MS	F
Between treatments	2	61.40	30.70	6.18
Linear*	1	61.25	61.25	12.32
Quadratic*	1	0.15	0.15	0.03
Within treatments	27	134.10	$s^2 = 4.97$	
Total corrected	29	195.50		
Mean	1	1,822,867.50		
Total	30	1,823,063.00		

* This split of the "between treatments" SS is explained in Section 23.5. The designations "among treatments," "between treatments," and "treatments" all mean the same thing.

5. *Linear and Quadratic Contrasts.* The "between treatments" sum of squares with 2 degrees of freedom can be split up in various ways. One standard way, which takes advantage of the equal spacing of the treatments, is to construct orthogonal linear and quadratic contrasts. Because of the orthogonality, the sums of squares of the two individual contrasts will add to the sum of squares between treatments. The contrasts use coefficients attached to the treatment totals $T_i = \sum_j Y_{ij}$ as follows:

	Caffeine (mg) in the Treatment		
Contrast	0	100	200
Linear (L)	−1	0	1
Quadratic (Q)	1	−2	1

These are, in fact, the linear and quadratic orthogonal polynomials shown in Section 22.2. Thus the contrasts are

$$\mathbf{c}_L'\mathbf{Y} = L = -T_1 \qquad + T_3 = 35.$$

$$\mathbf{c}_Q'\mathbf{Y} = Q = \quad T_1 - 2T_2 + T_3 = \quad 3.$$

If $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$ (for our example, $I = 3$, $J = 10$), the sums of squares attributable to L and Q are, by the general formula for individual contrasts $\mathbf{c}'\mathbf{Y}$, $\mathrm{SS}(\mathbf{c}'\mathbf{Y}) = (\mathbf{c}'\mathbf{Y})^2/\mathbf{c}'\mathbf{c}$,

$$\mathrm{SS}(L) = \frac{L^2}{J\{(-1)^2 + 0^2 + 1^2\}} = \frac{(35)^2}{10(2)} = 61.25,$$

$$\mathrm{SS}(Q) = \frac{Q^2}{J\{1^2 + (-2)^2 + 1^2\}} = \frac{(3)^2}{10(6)} = 0.15.$$

These are already entered in Table 23.3. In the denominators of these SS, J is the number of observations contributing to the totals T_i used to obtain L and Q, and the other term is the sum of squares of the coefficients that multiply the T_i. Because the contrasts $L = \mathbf{c}_L'\mathbf{Y}$ and $Q = \mathbf{c}_Q'\mathbf{Y}$ are orthogonal ($\mathbf{c}_L'\mathbf{c}_Q = 0$), the two SS add to 61.40.

Tests for linear and quadratic effects reveal a significant linear effect [because $12.32 > F(1, 27, 0.95) = 4.21$] and a nonsignificant quadratic effect. Thus we conclude that, within the range 0–200 mg of caffeine used, the true number of finger taps increases (because $\overline{Y}_1 < \overline{Y}_2 < \overline{Y}_3$) linearly with the amount of caffeine ingested.

This concludes the standard ANOVA approach to this example and we now discuss the regression approach to it.

23.3. REGRESSION TREATMENT OF THE ONE-WAY CLASSIFICATION EXAMPLE

The model given in Eq. (23.2.1) involves four parameters μ, t_1, t_2, t_3. Only two of these, μ and one t_i, are involved with any one Y_{ij}; also, $t_1 + t_2 + t_3 = 0$. A natural first step in a regression approach is to write down the model

$$Y = \mu X_0 + t_1 X_1 + t_2 X_2 + t_3 X_3 + \epsilon \tag{23.3.1}$$

and consider what choice of variables X_i would reproduce Eq. (23.2.1). A little thought shows that use of the dummies $X_0 = 1$ and

$$X_i = \begin{cases} 1, & \text{if the } i\text{th treatment is applied to provide } Y_{ij}, \\ 0, & \text{otherwise,} \end{cases} \tag{23.3.2}$$

for $i = 1, 2, 3$, will do the trick. This gives the setup:

$$\underset{(30 \times 1)}{\mathbf{Y}} = \begin{bmatrix} 242 \\ 245 \\ \vdots \\ 246 \\ 242 \\ \hline 248 \\ 246 \\ \vdots \\ 243 \\ 244 \\ \hline 246 \\ 248 \\ \vdots \\ 245 \\ 250 \end{bmatrix} \quad \underset{(30 \times 4)}{\mathbf{X}} = \begin{array}{c} \begin{matrix} \mathbf{X}_0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{matrix} \\ \left[\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{array}\right] \end{array} \quad \underset{(4 \times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ t_3 \end{bmatrix} \tag{23.3.3}$$

A drawback is immediately apparent. Because of the column dependence

$$\mathbf{X}_0 = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3 \tag{23.3.4}$$

in $\mathbf{X}$, the $\mathbf{X}'\mathbf{X}$ will necessarily be singular and so the normal equations will not have a unique solution. We have so far not taken into account the ANOVA model restriction $t_1 + t_2 + t_3 = 0$. If this is applied, the solution to the normal equations *is* unique. However, this provides an added complication (which we discuss further in Section 23.4) and means that we have not obtained a standard nonsingular regression format for our problem. How do we do that? There are many possibilities—an infinite number, technically—as there always are whenever dummy variables are used. One of these ways, which we now describe, enables us to reproduce the linear and quadratic contrasts split-up featured in Table 23.3. Define dummy variables Z_1 and Z_2 to replace X_1, X_2, and X_3 in the following correspondence:

Z_1	Z_2	for	X_1	X_2	X_3
-1	1		1	0	0
0	-2		0	1	0
1	1		0	0	1

(23.3.5)

The model can then be written in regression format as

$$Y = \beta_0 X_0 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon \tag{23.3.6}$$

with regression setup:

$$\underset{(30\times 1)}{\mathbf{Y}} = \begin{bmatrix} 242 \\ 245 \\ \vdots \\ 242 \\ \hline 248 \\ 246 \\ \vdots \\ 244 \\ \hline 246 \\ 248 \\ \vdots \\ 250 \end{bmatrix} \quad \underset{(30\times 3)}{\mathbf{X}} = \begin{array}{c} \begin{matrix} \mathbf{X}_0 & \mathbf{Z}_1 & \mathbf{Z}_2 \end{matrix} \\ \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & -1 & 1 \\ \hline 1 & 0 & -2 \\ 1 & 0 & -2 \\ \vdots & \vdots & \vdots \\ 1 & 0 & -2 \\ \hline 1 & 1 & 1 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \end{bmatrix} \end{array} \quad \underset{(3\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \tag{23.3.7}$$

Note that the vectors $\mathbf{X}_0$, $\mathbf{Z}_1$, and $\mathbf{Z}_2$ are mutually orthogonal. This makes the least squares solution extremely simple, as follows:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \frac{1}{30} & 0 & 0 \\ 0 & \frac{1}{20} & 0 \\ 0 & 0 & \frac{1}{60} \end{bmatrix} \begin{bmatrix} 7395 \\ 35 \\ 3 \end{bmatrix} = \begin{bmatrix} 246.50 \\ 1.75 \\ 0.05 \end{bmatrix}, \tag{23.3.8}$$

$$\begin{aligned} \mathrm{SS}(b_0, b_1, b_2) = \mathbf{b}'\mathbf{X}'\mathbf{Y} &= 1{,}822{,}867.50 + 61.25 + 0.15 \\ &\quad (= 1{,}822{,}928.90) \\ &= \mathrm{SS}(b_0) + \mathrm{SS}(b_1) + \mathrm{SS}(b_2). \end{aligned} \tag{23.3.9}$$

The clean split of the SS is due to the orthogonality previously mentioned. The details of Table 23.3 are thus reproduced exactly. The fitted regression model is

$$\hat{Y} = 246.50 + 1.75Z_1 + 0.05Z_2. \tag{23.3.10}$$

If we now drop the nonsignificant "$0.05Z_2$" term and note that X_1 can be regarded as a coding

$$Z_1 = (C - 100)/100, \tag{23.3.11}$$

where C = amount of caffeine ingested in milligrams, of the caffeine level, we see that the fitted equation becomes

$$\hat{Y} = 244.75 + 0.0175C. \tag{23.3.12}$$

Thus, within the ranges of C observed, the number of finger taps attained per minute by trained college students can be predicted in terms of C via Eq. (23.3.12).

(Note that, when a variable such as Z_2 is dropped, we would usually have to refit the regression equation. It would not be correct in general just to drop the term out. Here, however, because $\mathbf{Z}_2$ is orthogonal to the other columns of $\mathbf{X}$, the result is the same either way.)

A Caution

Careful choice of the levels of the dummies Z_1 and Z_2 to produce a diagonal $\mathbf{X}'\mathbf{X}$ matrix and the equality of the three sample sizes made the above example an extremely neat one. In general, we cannot always achieve such tidy results from a regression approach. Let us look again at Eq. (23.3.3). Suppose we had opted, instead, to just drop one of the parameters, say, t_3, from $\boldsymbol{\beta}$, and thus the corresponding column $\mathbf{X}_3$ from $\mathbf{X}$. The corresponding least squares solution would then have been (using $\hat{\mu}$ to denote the estimate of μ, and so on)

$$\begin{bmatrix} \hat{\mu} \\ \hat{t}_1 \\ \hat{t}_2 \end{bmatrix} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (10)^{-1}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 7395 \\ 2448 \\ 2464 \end{bmatrix}$$

$$= \tfrac{1}{10}\begin{bmatrix} 1 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}\begin{bmatrix} 7395 \\ 2448 \\ 2464 \end{bmatrix} = \begin{bmatrix} 248.3 \\ -3.5 \\ -1.9 \end{bmatrix}, \tag{23.3.13}$$

$$\mathrm{SS}(\hat{\mu}, \hat{t}_1, \hat{t}_2) = \mathbf{b}'\mathbf{X}'\mathbf{Y} = (248.3, -3.5, -1.9)\begin{bmatrix} 7395 \\ 2448 \\ 2464 \end{bmatrix}$$

$$= 1{,}822{,}928.90. \tag{23.3.14}$$

This value is exactly the same as the total of the components in Eq. (23.3.9) as, of course, it should be, but the informative split-up into three pieces achieved there does not happen here due to the lack of orthogonality, that is, because $\mathbf{X}'\mathbf{X}$ is not diagonal here. We can, of course, obtain the extra sum of squares for $\hat{t}_1$, $\hat{t}_2$ by subtraction of $n\overline{Y}^2$ to give

$$\mathrm{SS}(\hat{t}_1, \hat{t}_2 \mid \hat{\mu}) = 1{,}822{,}928.90 - 1{,}822{,}867.50 = 61.40 \tag{23.3.15}$$

in the usual way, but we would have to perform additional "extra sums of squares" calculations to proceed further. The moral is that, although many regression setups of an ANOVA problem are possible, some are more informative than others, and care in choosing a good setup will be repaid.

Proceeding further with the current (nonorthogonal) setup, we obtain the fitted model

$$\hat{Y} = \hat{Y}(X_1, X_2) = 248.3 - 3.5X_1 - 1.9X_2, \tag{23.3.16}$$

which produces predicted values as follows:

$$\hat{Y}(1, 0) = 244.8, \quad \text{when 0 mg of caffeine is ingested,}$$
$$\hat{Y}(0, 1) = 246.4, \quad \text{when 100 mg of caffeine is ingested,}$$
$$\hat{Y}(0, 0) = 248.3, \quad \text{when 200 mg of caffeine is ingested.}$$

The reader might initially be surprised to see negative coefficients in Eq. (23.3.16), because the number of finger taps *increases* with the dose level of caffeine. The results are correct, however. The dummy variable setup chosen here has caused $\hat{Y}(0, 0)$ to be the "base level" prediction and $\hat{Y}(1, 0)$ and $\hat{Y}(0, 1)$ require *negative* adjustments to achieve the proper values, that is all. As we have already said, care is needed in these applications.

In Sections 23.4 and 23.5, we discuss the one-way classification more generally. After that we turn to the two-way classification.

Relationship to the Underlying Geometry

The usual analysis of variance model, (23.2.1) in our example, is overparameterized and so the estimation space is described by more vectors, shown in (23.3.3), than are needed. An infinite number of descriptions of (the unique) $\hat{\mathbf{Y}}$ is thus possible unless extra restrictions are added. The restriction $t_1 + t_2 + t_3$ of our example simply selects one of that infinite number of descriptions. From the regression point of view it is simpler to define the estimation space by a minimal set of vectors, as in (23.3.7), or by dropping an $\mathbf{X}_i$ column, for our example. (Many alternative choices of the minimal set of vectors are, of course, feasible.) This allows us to proceed without any restrictions and leads to a unique description of $\hat{\mathbf{Y}}$. See Section 20.4 for additional discussion.

23.4. REGRESSION TREATMENT OF THE ONE-WAY CLASSIFICATION USING THE ORIGINAL MODEL

We now look at the general one-way classification without reduction of parameters and show that, while the original normal equations cannot be solved uniquely, addition of the usual restriction $J_1t_1 + J_2t_2 + \cdots + J_It_I = 0$ provides a unique solution for the parameter estimates. Essentially, the restriction can replace any one of the original normal equations. Instead of Eq. (23.3.1), write

$$E(Y) = \mu X_0 + t_1X_1 + t_2X_2 + \cdots + t_IX_I. \tag{23.4.1}$$

We want this to express the fact that if we consider an observation Y_{ij} from the ith group it must have expectation $\mu + t_i$. We define

$$\mathbf{Y}' = (Y_{11}, Y_{12}, \ldots, Y_{1J_1}; Y_{21}, Y_{22}, \ldots, Y_{2J_2}; \ldots; Y_{I1}, Y_{I2}, \ldots, Y_{IJ_I})$$

and

$$\mathbf{X} = \begin{array}{c} \begin{matrix} \mathbf{X}_0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & \cdots & \mathbf{X}_I \end{matrix} \\ \left[\begin{array}{cccccc} 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ \hline 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \hline 1 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \end{array}\right] \end{array}$$

where the dashes divide the matrix sets of rows, there being $\mathbf{J}_1, \mathbf{J}_2, \ldots, \mathbf{J}_I$ rows in successive sets. The headings show to which X the columns relate. Furthermore, define

$$\boldsymbol{\beta}' = (\mu; t_1, t_2, \ldots, t_I),$$

then

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

expresses the model in matrix notation. Now

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & J_1 & J_2 & \cdots & J_I \\ J_1 & J_1 & 0 & \cdots & 0 \\ J_2 & 0 & J_2 & \cdots & 0 \\ \vdots & & & \cdots & \\ J_I & 0 & 0 & \cdots & J_I \end{bmatrix}, \qquad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} n\overline{Y} \\ J_1\overline{Y}_1 \\ J_2\overline{Y}_2 \\ \vdots \\ J_I\overline{Y}_I \end{bmatrix} \tag{23.4.2}$$

If we write b_0, b_i for the least squares estimates of μ and t_i, we can write the normal equations $(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{Y}$ as

$$\begin{aligned} nb_0 + J_1b_1 + J_2b_2 + \cdots + J_Ib_I &= n\overline{Y}, \\ J_1b_0 + J_1b_1 \qquad\qquad\qquad\quad &= J_1\overline{Y}_1, \\ J_2b_0 \qquad + J_2b_2 \qquad\qquad &= J_2\overline{Y}_2, \\ \vdots \qquad\qquad\qquad \cdots & \\ J_Ib_0 \qquad\qquad\qquad + J_Ib_I &= J_I\overline{Y}_I. \end{aligned} \tag{23.4.3}$$

Here the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix does not exist since $\mathbf{X}'\mathbf{X}$ is singular. This is due to the fact that Eqs. (23.4.3) are *not* independent because the first equation is the sum of the other I equations. There are in fact only I equations in the $(I + 1)$ unknowns b_0, $b_1, \ldots, b_I$, because the original model (23.3.1) contained more parameters than were actually necessary. This "singularity" of the $\mathbf{X}'\mathbf{X}$ matrix is also clear from examination of the $\mathbf{X}$ matrix, where the X_0 column is equal to the sum of the $X_1, X_2, \ldots, X_I$ columns, a dependence that becomes transmitted into the normal equations of (23.4.3) as we have noted. How can we proceed then? A condition we have not so far taken into account is Eq. (23.2.2), which, if true of the parameters, must also hold for the estimates of the parameters. Hence

$$J_1 b_1 + J_2 b_2 + \cdots + J_I b_I = 0. \tag{23.4.4}$$

This provides the additional independent equation we require. We now take *any* I equations from (23.4.3) together with Eq. (23.4.4) and use these as the normal equations. It is most convenient to drop the first equation of (23.4.3) since it has the most terms. This leaves, as the equations to solve,

$$\begin{aligned}
J_1 b_1 + J_2 b_2 + \cdots + J_I b_I &= 0, \\
J_1 b_0 + J_1 b_1 &= J_1 \overline{Y}_1, \\
J_2 b_0 + J_2 b_2 &= J_2 \overline{Y}_2, \\
&\vdots \quad \cdots \\
J_I b_0 + J_I b_I &= J_I \overline{Y}_I.
\end{aligned} \tag{23.4.5}$$

To maintain symmetry we have not divided through the second to $(I + 1)$st equations by their common factors. In matrix form we can write Eq. (23.4.5) as

$$\begin{bmatrix} 0 & J_1 & J_2 & \cdots & J_I \\ J_1 & J_1 & 0 & \cdots & 0 \\ J_2 & 0 & J_2 & \cdots & \\ \vdots & & & \cdots & \\ J_I & 0 & 0 & \cdots & J_I \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_I \end{bmatrix} = \begin{bmatrix} 0 \\ J_1 \overline{Y}_1 \\ J_2 \overline{Y}_2 \\ \vdots \\ J_I \overline{Y}_I \end{bmatrix}. \tag{23.4.6}$$

Since we cannot express these in the form $(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{Y}$ it will usually be impractical to use this procedure when the work is done by a computer routine, which requires this form.

From Eq. (23.4.5),

$$b_i = \overline{Y}_i - b_0, \qquad i = 1, 2, \ldots, I.$$

Substituting in the first equation

$$\begin{aligned}
0 = \sum_{i=1}^{I} J_i b_i &= \sum_{i=1}^{I} J_i (\overline{Y}_i - b_0) \\
&= \sum_{i=1}^{I} J_i \overline{Y}_i - b_0 \sum_{i=1}^{I} J_i \\
&= n\overline{Y} - n b_0.
\end{aligned}$$

Thus

$$b_0 = \overline{Y}$$

and

$$b_i = \overline{Y}_i - \overline{Y}.$$

The sum of squares due to a vector of estimates $\mathbf{b}$ determined from equations $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$ is defined by $\mathbf{b}'\mathbf{X}'\mathbf{Y}$ even if $\mathbf{X}'\mathbf{X}$ is singular and cannot be inverted and extra conditions of the form $\mathbf{Q}\mathbf{b} = 0$ are needed to give a unique solution. No matter what form $\mathbf{Q}$ takes [here it is a vector $(0, J_1, J_2, \ldots, J_I)$], $\mathbf{b}'\mathbf{X}'\mathbf{Y}$ is invariant. This is so since if $\mathbf{b}_1$, $\mathbf{b}_2$ are two solutions arising from different "extra conditions"

$$\begin{aligned}\mathbf{b}_1'(\mathbf{X}'\mathbf{Y}) &= \mathbf{b}_1'(\mathbf{X}'\mathbf{X}\mathbf{b}_2) \\ &= (\mathbf{X}'\mathbf{X}\mathbf{b}_1)'\mathbf{b}_2\end{aligned}$$

[regrouping and using the matrix theory fact that $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$],

$$\begin{aligned}&= (\mathbf{X}'\mathbf{Y})'\mathbf{b}_2 \\ &= \mathbf{b}_2'\mathbf{X}'\mathbf{Y}.\end{aligned}$$

Of course, this is obvious geometrically. We are evaluating the squared distance from the origin O to the point $\hat{\mathbf{Y}}$ and this remains the same however we choose to describe it.

Thus the sum of squares due to regression is

$$\begin{aligned}\mathbf{b}'\mathbf{X}'\mathbf{Y} &= n\overline{Y}^2 + \sum_{i=1}^{I} J_i\overline{Y}_i(\overline{Y}_i - \overline{Y}) \\ &= n\overline{Y}^2 + \sum_{i=1}^{I} J_i(\overline{Y}_i - \overline{Y})^2\end{aligned}$$

with I degrees of freedom, since the additional term,

$$\sum_{i=1}^{I} (-\overline{Y})J_i(\overline{Y}_i - \overline{Y}),$$

added to the right-hand size, is zero by definition of the means. If the model had only a term μ in it we should have

$$\mathrm{SS}(b_0) = n\overline{Y}^2$$

with one degree of freedom. Thus

$$\begin{aligned}\mathrm{SS}(b_1, b_2, \ldots, b_I \mid b_0) &= \mathbf{b}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2 \\ &= \sum_{i=1}^{I} J_i(\overline{Y}_i - \overline{Y})^2,\end{aligned}$$

with $(I - 1)$ degrees of freedom.

These provide the sums of squares due to the "mean" and "between groups," of the analysis of variance table of Table 23.1. The "within groups" sum of squares is found by the difference $\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$ as usual and is the same as the form given in Table 23.1 if evaluated. The test for $H_0 : t_1 = t_2 = \cdots = t_I = 0$ is made exactly as in the analysis of variance case.

We have shown here that the one-way analysis of variance problem can be done formally by regression using the original model. However, to perform the calculations

on a computer, it is probably best to remove the singularity in the problem by choosing the model more carefully prior to computation.

Note. The work above shows how we must proceed in general in a regression problem, if there are more parameters to estimate than there are independent normal equations. If no natural restrictions are available, as in the analysis of variance case, we must make restrictions in an arbitrary fashion. While the choice of restrictions will influence the actual values of the regression coefficients, it will not affect the sum of squares due to regression. Usually we would select restrictions that would make the normal equations easier to solve.

Example. Suppose the normal equations were

$$\begin{aligned}
22b_1 + 10b_2 + 12b_3 + 5b_4 + 8b_5 + 9b_6 &= 34.37,\\
10b_1 + 10b_2 \qquad\quad + 3b_4 + 4b_5 + 3b_6 &= 21.21,\\
12b_1 \qquad + 12b_3 + 2b_4 + 4b_5 + 6b_6 &= 13.16,\\
5b_1 + 3b_2 + 2b_3 + 5b_4 \qquad\quad &= 10.28,\\
8b_1 + 4b_2 + 4b_3 \qquad + 8b_5 \qquad &= 14.23,\\
9b_1 + 3b_2 + 6b_3 \qquad\quad 9b_6 &= 9.86.
\end{aligned}$$

[These equations appear in Plackett (1960, p. 44).] They arise from a two-way classification with unequal numbers of observations in the cells. Such data might also arise from an intended "equal cells" analysis if several observations were missing. We discuss the "equal cells" analysis in the next section.)

Only four of the six equations are independent since the second and third equations add to give the first as do the fourth, fifth, and sixth. Thus two additional equations are required to give six equations in six unknowns. We must add two independent restrictions on $b_1, b_2, \ldots, b_6$, which must not be linear combinations of the existing equations.

Since there are actually only four independent normal equations we can drop two dependent ones, say, the first and sixth. The remaining four can be written in matrix form as follows:

$$\begin{array}{cccccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} \end{array}$$

$$\begin{bmatrix} 10 & 10 & 0 & 3 & 4 & 3 \\ 12 & 0 & 12 & 2 & 4 & 6 \\ 5 & 3 & 2 & 5 & 0 & 0 \\ 8 & 4 & 4 & 0 & 8 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} = \begin{bmatrix} 21.21 \\ 13.16 \\ 10.28 \\ 14.23 \end{bmatrix}.$$

Since the original matrix was symmetric, the row (or equation) dependence noticed earlier is also reflected in the fact that the first column is the sum of the second and third, and also the sum of the fourth, fifth, and sixth. When adding the two restrictions on the b's we must be careful on two counts. The restrictions, when placed below the four selected equations, will contribute two additional rows to the matrix and two zeros to the right-hand side vector (we would usually take the restrictions to be of the form $\sum c_i b_i = 0$). The final matrix must be such that there is no dependence between either rows *or* columns, if a unique solution is required. [There are more

elegant matrix ways of expressing this in general; see, for example, Plackett (1960), but we do not provide them in our more elementary development.] For example, we *cannot* take restrictions

$$7b_1 + 6b_2 + b_3 + b_4 + b_5 + 5b_6 = 0,$$
$$11b_1 + 9b_2 + 2b_3 + 4b_4 + 4b_5 + 3b_6 = 0,$$

since the original column dependency is preserved. Even if only one column dependency is preserved, for example, by restrictions

$$3b_4 + 4b_5 + 3b_6 = 0,$$
$$9b_1 + 5b_2 + 4b_3 = 0,$$

which allow column one to be the sum of columns two and three, the restrictions are useless. However,

$$3b_4 + 4b_5 + 3b_6 = 0,$$
$$b_2 + b_3 = 0$$

would be usable, since no dependence will occur, and we shall have six equations in six unknowns as required. (Different restrictions were used by Plackett, whose book should be consulted for the subsequent solution.)

The idea of adding arbitrary restrictions may seem somewhat peculiar at first. One must remember that it is required whenever more parameters are used than are actually necessary to express the model. At some stage this "looseness" must be removed and that is the purpose of the added restrictions.

Again this becomes obvious geometrically. When an infinite number of least squares solutions exist because of overdescription of the estimation space, we must add enough extra restrictions ("information") to particularize to one solution, *if* we wish to describe only one specific solution.

23.5. REGRESSION TREATMENT OF THE ONE-WAY CLASSIFICATION: INDEPENDENT NORMAL EQUATIONS

In this section we tackle the one-way classification yet again with a reparameterization, reducing the parameter dimensions from $I + 1$ to I. An alternative way to look at this is that we choose to drop the column of 1's associated with the parameter μ. However it is described, the underlying regression framework remains the same, with vectors $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ permanently fixed but $\hat{\mathbf{Y}}$ variously defined.

The analysis of variance model is

$$E(Y_{ij}) = \mu + t_i. \tag{23.5.1}$$

Let us write

$$\beta_i = \mu + t_i, \qquad i = 1, 2, \ldots, I.$$

Then, in regression terms, we can write

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}, \tag{23.5.2}$$

where, comparing with the definitions in Section 23.4, $\mathbf{Y}$ is as before, $\mathbf{X}$ is the matrix formed by dropping the X_0 column of the previous $\mathbf{X}$ matrix, and $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots,$

β_I). Let $\mathbf{b}' = (b_1, b_2, \ldots, b_I)$. Then

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} J_1 & & & \mathbf{0} \\ & J_2 & & \\ & & \ddots & \\ \mathbf{0} & & & J_I \end{bmatrix}, \qquad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} J_1\overline{Y}_1 \\ J_2\overline{Y}_2 \\ \vdots \\ J_I\overline{Y}_I \end{bmatrix}. \tag{23.5.3}$$

Since $\mathbf{X}'\mathbf{X}$ is a diagonal matrix with J_i in the ith diagonal position and zero elsewhere, its inverse is a diagonal matrix with $1/J_i$ in the ith diagonal position. From this it is easy to see that

$$b_i = \overline{Y}_i. \tag{23.5.4}$$

The sum of squares due to $\mathbf{b}$ is

$$\mathbf{b}'\mathbf{X}'\mathbf{Y} = \sum_{i=1}^{I} \overline{Y}_i(J_i\overline{Y}_i) = \sum_{i=1}^{I} J_i\overline{Y}_i^2 \tag{23.5.5}$$

and the residual sum of squares is

$$\begin{aligned} \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} &= \sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}^2 - \sum_{i=1}^{I} J_i\overline{Y}_i^2 \\ &= \sum_{i=1}^{I}\left\{\sum_{j=1}^{J_i} Y_{ij}^2 - J_i\overline{Y}_i^2\right\} \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_i)^2, \end{aligned} \tag{23.5.6}$$

with $(n - I)$ degrees of freedom.

The hypothesis $H_0 : t_1 = t_2 = \cdots = t_I = 0$ is expressed in our present notation as

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_I = \mu.$$

If H_0 were true, the model would be

$$E(Y_{ij}) = \mu$$

or

$$E(\mathbf{Y}) = \mathbf{j}\mu, \tag{23.5.7}$$

where $\mathbf{j}$ is a vector of ones of the same length as $\mathbf{Y}$. The (single) normal equation would be

$$n\mu = \mathbf{j}'\mathbf{j}\mu = \mathbf{j}'\mathbf{Y} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij} = n\overline{Y}.$$

Thus the estimate of μ would be

$$b_0 = \overline{Y} \tag{23.5.8}$$

and

$$\mathrm{SS}(b_0) = n\overline{Y}^2, \tag{23.5.9}$$

providing a residual sum of squares of

$$\sum_{i=1}^{I}\sum_{j=1}^{J_i} Y_{ij}^2 - n\overline{Y}^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y})^2 \tag{23.5.10}$$

as can be shown, with $(n - 1)$ degrees of freedom.

The sum of squares due to H_0 is the difference between Eqs. (23.5.10) and (23.5.6), namely,

$$\begin{aligned} \mathrm{SS}(H_0) &= \sum_{i=1}^{I}\sum_{j=1}^{J_i} \{(Y_{ij} - \overline{Y})^2 - (Y_{ij} - \overline{Y}_i)^2\} \\ &= \sum_{i=1}^{I} J_i(\overline{Y}_i - \overline{Y})^2 \end{aligned}$$

after reduction, with $(n - 1) - (n - I) = (I - 1)$ degrees of freedom. The test statistic for H_0 is thus

$$F = \frac{\mathrm{SS}(H_0)}{I-1} \Big/ \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i} (Y_{ij} - \overline{Y}_i)^2}{n-I},$$

which is exactly what we obtain from the analysis of variance. Thus if we express the model for the one-way classification as $E(Y_{ij}) = \beta_i$ and test the hypothesis $H_0 : \beta_1 = \beta_2 = \cdots = \beta_I = \mu$, we can reproduce the analysis of variance through regression analysis using standard programs. We can obtain estimates of the parameters t_i from $b_i - b_0$.

23.6. THE TWO-WAY CLASSIFICATION WITH EQUAL NUMBERS OF OBSERVATIONS IN THE CELLS: AN EXAMPLE

The principles given in Sections 23.2 and 23.3 become even more important for more complicated experimental designs. We will consider an example of a two-way classification model with main effects and an interaction term, discussed by Smith (1969).

A manufacturer was having trouble in his catalyst plant with production rates. After extensive discussion with the research unit, it was decided to investigate the effects of twelve different combinations of four reagents and three catalysts. One of the problems that the plant had been encountering was an inability to reproduce production rates under what seemed to be identical conditions. In order to obtain an estimate of this inherent variability, it was decided to do each of the experimental runs twice. Thus the experiment consisted of 24 experimental runs, and these were done in a random order. The data are given in Table 23.4; they have been coded and rounded.

TABLE 23.4. Twenty-four Production Rates (Coded and Rounded) for Twelve Combinations of Reagent and Catalyst

	Catalyst		
Reagent	1	2	3
A	4,6	11,7	5,9
B	6,4	13,15	9,7
C	13,15	15,9	13,13
D	12,12	12,14	7,9

TABLE 23.5. Analysis of Variance for Model of Eq. (23.6.1)

Source	df	SS	MS	F
Among reagents	3	120	40	10.0^a
Among catalysts	2	48	24	6.0^b
Reagents × catalysts	6	84	14	3.5^b
Pure error	12	48	4	
Total, corrected	23	200		

[a] Significant at the $\alpha = 0.05$ level.
[b] Significant at the $\alpha = 0.01$ level.

The ANOVA model for this experiment was taken as

$$Y_{ijk} = \mu + R_i + C_j + (RC)_{ij} + \epsilon_{ijk}, \tag{23.6.1}$$

where

μ = overall mean level (1 parameter).
R_i = effect of the ith reagent (4 parameters).
C_j = effect of the jth catalyst (3 parameters).
$(RC)_{ij}$ = interaction effect of reagent i and catalyst j (12 parameters).
ϵ_{ijk} = random error within the (i, j)th cell for the k observations within the cell, assumed to be distributed $N(0, \sigma^2)$. Errors are assumed to be pairwise uncorrelated.

Also,

$$i = 1, 2, \ldots, I \ (I = 4, \text{here}),$$
$$j = 1, 2, \ldots, J \ (J = 3, \text{here}),$$
$$k = 1, 2, \ldots, K \ (K = 2, \text{here}).$$

Note that this model involves 20 parameters and is overparameterized. The following restrictions are therefore assumed:

$$\sum_i R_i = \sum_j C_j = 0; \tag{23.6.2}$$

$$\sum_i (RC)_{ij} = 0, \qquad j = 1, 2, \ldots, J; \tag{23.6.3}$$

$$\sum_j (RC)_{ij} = 0, \qquad i = 1, 2, \ldots, I. \tag{23.6.4}$$

The number of independent parameters is thus reduced to $1 + 3 + 2 + 6 = 12$ [or $1 + (I - 1) + (J - 1) + (I - 1)(J - 1) = IJ$, in general], the number of cells. The usual analysis provides the ANOVA table of Table 23.5.

We now discuss how to analyze these data using a regression approach.

23.7. REGRESSION TREATMENT OF THE TWO-WAY CLASSIFICATION EXAMPLE

As we have already mentioned, twelve independent model parameters are needed for this example. Assignment of more than twelve to a regression model will thus produce a singular $\mathbf{X}'\mathbf{X}$ matrix. Of these twelve, we need one for the mean, three for reagents, two for catalysts, and six for interaction. As always, many choices are possible for the set of dummy variables used to represent the regression model. Our choice will produce

a diagonal $\mathbf{X}'\mathbf{X}$ (many choices will not, as in the one-way classification case). We shall write the regression model as

$$\begin{aligned} Y = {} & \beta_0 X_0 && \text{(constant term; } X_0 = 1\text{)} \\ & + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 && \text{(for reagents)} \\ & + \beta_4 X_4 + \beta_5 X_5 && \text{(for catalysts)} \\ & + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} && \text{(for } R \times C \text{ interaction)} \\ & + \epsilon, && \text{(error)} \end{aligned} \qquad (23.7.1)$$

and define the dummy variables as follows:

	Corresponding Values of Dummies		
Reagent Used	X_1	X_2	X_3
A	-1	0	-1
B	1	0	-1
C	0	-1	1
D	0	1	1

These are not arbitrary choices; they allow special comparisons to be made between reagents

$$\begin{aligned} A \text{ and } B, &\quad \text{via } X_1, \\ C \text{ and } D, &\quad \text{via } X_2, \\ (A + B) \text{ and } (C + D), &\quad \text{via } X_3, \end{aligned} \qquad (23.7.2)$$

and these independent comparisons or contrasts "carry" 3 degrees of freedom. Any other type of comparison can be generated from these columns. For example, to compare A and D we could use a column (dependent on X_1, X_2, and X_3 columns, of course) generated by $X_1 + X_2 + X_3$ whose entries will be $(-2, 0, 0, 2)'$. Also, because of the symmetry of the design as well, the $\mathbf{X}$ matrix columns generated by X_1, X_2, and X_3 will be orthogonal to one another. Provided they are also orthogonal to other $\mathbf{X}$ columns, as indeed they will be, individual additive sums of squares can be ascribed to each of the three contrasts generated by the X_1, X_2, and X_3 columns.

For the catalyst dummy variables, we assign as follows:

	Corresponding Values of Dummies	
Catalyst	X_4	X_5
1	-1	1
2	0	-2
3	1	1

The two orthogonal columns allow comparisons between catalysts

$$\begin{aligned} 1 \text{ and } 3, &\quad \text{via } X_4, \\ (1 + 3) \text{ and } 2, &\quad \text{via } X_5. \end{aligned} \qquad (23.7.3)$$

Because of the perfect balance of this two-way design and the fact that there are equal numbers of observations in each cell, we can construct the interaction dummies by multiplying up corresponding elements of the other dummies, taking the six combi-

nations of (three reagent dummies) times (two catalyst dummies). Thus

$$X_6 = X_1X_4, \qquad X_7 = X_1X_5,$$
$$X_8 = X_2X_4, \qquad X_9 = X_2X_5, \tag{23.7.4}$$
$$X_{10} = X_3X_4, \qquad X_{11} = X_3X_5.$$

We can now write down the **Y** vector and **X** matrix as given below.

$$\mathbf{Y} = \begin{bmatrix} 4 \\ 6 \\ 11 \\ 7 \\ 5 \\ 9 \\ 6 \\ 4 \\ 13 \\ 15 \\ 9 \\ 7 \\ 13 \\ 15 \\ 15 \\ 9 \\ 13 \\ 13 \\ 12 \\ 12 \\ 12 \\ 14 \\ 7 \\ 9 \end{bmatrix} \qquad \mathbf{X} = \begin{array}{c} \begin{matrix} \mathbf{X}_0 & \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 & \mathbf{X}_4 & \mathbf{X}_5 & \mathbf{X}_6 & \mathbf{X}_7 & \mathbf{X}_8 & \mathbf{X}_9 & \mathbf{X}_{10} & \mathbf{X}_{11} \end{matrix} \\ \begin{bmatrix} 1 & -1 & 0 & -1 & -1 & 1 & 1 & -1 & 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & -1 & -1 & 1 & 1 & -1 & 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & -1 & 0 & -2 & 0 & 2 & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -1 & 0 & -2 & 0 & 2 & 0 & 0 & 0 & 2 \\ 1 & -1 & 0 & -1 & 1 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\ 1 & -1 & 0 & -1 & 1 & 1 & -1 & -1 & 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 & 1 & -1 & 1 & 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 & -1 & 1 & -1 & 1 & 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 2 \\ 1 & 1 & 0 & -1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 2 \\ 1 & 1 & 0 & -1 & 1 & 1 & 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & 1 & 1 & 1 & 1 & 0 & 0 & -1 & -1 \\ 1 & 0 & -1 & 1 & -1 & 1 & 0 & 0 & 1 & -1 & -1 & 1 \\ 1 & 0 & -1 & 1 & -1 & 1 & 0 & 0 & 1 & -1 & -1 & 1 \\ 1 & 0 & -1 & 1 & 0 & -2 & 0 & 0 & 0 & 2 & 0 & -2 \\ 1 & 0 & -1 & 1 & 0 & -2 & 0 & 0 & 0 & 2 & 0 & -2 \\ 1 & 0 & -1 & 1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1 \\ 1 & 0 & -1 & 1 & 1 & 1 & 0 & 0 & -1 & -1 & 1 & 1 \\ 1 & 0 & 1 & 1 & -1 & 1 & 0 & 0 & -1 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 & -1 & 1 & 0 & 0 & -1 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 & 0 & -2 & 0 & 0 & 0 & -2 & 0 & -2 \\ 1 & 0 & 1 & 1 & 0 & -2 & 0 & 0 & 0 & -2 & 0 & -2 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \tag{23.7.5}$$

Examination shows that the columns of **X** are all orthogonal to one another so that $\mathbf{X}'\mathbf{X}$ is a 12×12 diagonal matrix with entries

$$\mathbf{X}'\mathbf{X} = \text{diag}\{24, 12, 12, 24, 16, 48, 8, 24, 8, 24, 16, 48\} \tag{23.7.6}$$

and $(\mathbf{X}'\mathbf{X})^{-1}$ is also diagonal with entries $\{\frac{1}{24}, \frac{1}{12}, \ldots, \frac{1}{16}, \frac{1}{48}\}$. Now

$$\mathbf{X}'\mathbf{Y} = \{240, 12, -12, 48, 0, -48, 2, -18, -6, -18, -20, 36\}'. \tag{23.7.7}$$

TABLE 23.6. Anova Table for the Two-Way Classification Example Using a Regression Approach

Source of Variation	df	SS	MS	F
Among reagents	3	120	40	10[a]
A vs B	1	12	12	3
C vs D	1	12	12	3
$(A + B)$ vs $(C + D)$	1	96	96	24[a]
Among catalysts	2	48	24	6[b]
1 vs 3	1	0[c]	0	0
(1 + 3) vs 2	1	48	48	12[a]
Reagents × catalysts	6	84[d]	14	3.5[b]
Pure error	12	48	$s^2 = 4$	
Total, corrected	23	300		
b_0	1	2400		
Total	24	2700		

[a] Significant at the $\alpha = 0.01$ level.
[b] Significant at the $\alpha = 0.05$ level.
[c] A sum of squares that is exactly zero rarely occurs with real data and is often a sign that the data are constructed. Here it happens because we coded and rounded the original numbers to simplify the arithmetic.
[d] This sum of squares can be partitioned as shown in Eq. (23.7.9) into single df sums of squares. The significant contributors are then b_{10} and b_{11}, indicating the existence of "real" interactions between (X_3 and X_4) and (X_3 and X_5).

It follows that $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ has value

$$\mathbf{b} = \{10, 1, -1, 2, 0, -1, \tfrac{1}{4}, -\tfrac{3}{4}, -\tfrac{3}{4}, -\tfrac{3}{4}, -\tfrac{5}{4}, \tfrac{3}{4}\}', \tag{23.7.8}$$

and the sum of squares for regression, $\mathbf{b}'\mathbf{X}'\mathbf{Y}$, consists of the following twelve independent contributions, grouped for convenience, each ascribable as the sum of squares due to one b-coefficient:

$$\begin{aligned}\mathbf{b}'\mathbf{X}'\mathbf{Y} &= 2400 + \{12 + 12 + 96\} + \{0 + 48\} \\ &\quad + \{\tfrac{1}{2} + 13\tfrac{1}{2} + 4\tfrac{1}{2} + 13\tfrac{1}{2} + 25 + 27\} \\ &= 2400 + 120 + 48 + 84 \\ &= 2652.\end{aligned} \tag{23.7.9}$$

These results allow us to write the ANOVA table for this example as shown in Table 23.6.

From Eq. (23.7.8) we obtain the fitted regression model as

$$\begin{aligned}\hat{Y} = 10 &+ X_1 - X_2 + 2X_3 + 0X_4 - X_5 + \tfrac{1}{4}X_6 - \tfrac{3}{4}X_7 \\ &- \tfrac{3}{4}X_8 - \tfrac{3}{4}X_9 - \tfrac{5}{4}X_{10} + \tfrac{3}{4}X_{11},\end{aligned} \tag{23.7.10}$$

and this can be used to predict the production rates in the various cells and to obtain residuals. For example, for reagent C and catalyst 2, we see that

$$\begin{aligned} &X_1 = 0, \quad X_4 = 0, \quad X_6 = 0, \quad X_7 = 0, \\ &X_2 = -1, \quad X_5 = -2, \quad X_8 = 0, \quad X_9 = 2, \\ &X_3 = 1, \quad X_{10} = 0, \quad X_{11} = -2, \end{aligned}$$

so that

$$\hat{Y} = 10 + 1 + 2 + 2 - \tfrac{3}{4}(2) + \tfrac{3}{4}(-2) = 12.$$

In that cell, the actual observations are 15 and 9, so the corresponding residuals are 3 and −3, which add to zero. The residuals add to zero in every cell, in fact, the model being essentially fitted to the means of the repeats. This happens in all regression situations in which the same number of repeat runs occur at every location.

Variances of estimated parameters and of fitted values can be obtained via the usual regression formulas. We leave further analysis of this example to the reader.

In Section 23.8, we present the standard general algebra of the two-way equal-celled analysis of variance, and in Section 23.9 we show two alternative regression approaches. One makes use of the extra sum of squares principle; the other is the "unsymmetrical dropping of parameters" approach to achieve a nonsingular $\mathbf{X}'\mathbf{X}$ matrix. The first of these methods is illustrated by an example in Section 23.10.

23.8. THE TWO-WAY CLASSIFICATION WITH EQUAL NUMBERS OF OBSERVATIONS IN THE CELLS

Suppose we have a two-way classification with I rows and J columns, with K observations Y_{ijk}, $k = 1, 2, \ldots, K$ in the cells so defined. The usual fixed-effects analysis of variance model is

$$E(Y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij}, \qquad \begin{aligned} i &= 1, 2, \ldots, I, \\ j &= 1, 2, \ldots, J, \\ k &= 1, 2, \ldots, K, \end{aligned} \tag{23.8.1}$$

subject to the restrictions

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} \gamma_{ij} \,(\text{all } j) = \sum_{j=1}^{J} \gamma_{ij} \,(\text{all } i) = 0. \tag{23.8.2}$$

These restrictions enable μ to be regarded as the overall mean level while the α_i, β_i, and γ_{ij} are differences between row level and overall mean, column level and overall mean, and cell level and the joint row plus column level, respectively. The usual analysis of variance table is as follows:

ANOVA

Source	df	SS	MS
Rows	$I - 1$	$JK \sum_{i=1}^{I} (\overline{Y}_{i\cdot\cdot} - \overline{Y})^2$	s_r^2
Columns	$J - 1$	$IK \sum_{j=1}^{J} (\overline{Y}_{\cdot j\cdot} - \overline{Y})^2$	s_c^2
Cells (interaction)	$(I - 1)(J - 1)$	$K \sum_{i=1}^{I} \sum_{j=1}^{J} (\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y})^2$	s_{rc}^2
Residual	$IJ(K - 1)$	By subtraction	s^2
Mean	1	$IJK\overline{Y}^2$	
Total	IJK	$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk}^2$	

Where $\overline{Y}_{i\cdot\cdot}$ = the mean of all observations in row i
$\overline{Y}_{\cdot j\cdot}$ = the mean of all observations in column j
$\overline{Y}_{ij\cdot}$ = the mean of all observations in the cell (i, j)
$\overline{Y}$ = the mean of all observations

The usual tests made are

$$H_0: \text{all } \alpha_i = 0 \qquad F = s_r^2/s^2 \text{ is compared with } F[(I-1), IJ(K-1)].$$

$$H_0: \text{all } \beta_j = 0 \qquad F = s_c^2/s^2 \text{ is compared with } F[(J-1), IJ(K-1)].$$

$$H_0: \text{all } \gamma_{ij} = 0 \qquad F = s_{rc}^2/s^2 \text{ is compared with } F[(I-1)(J-1), IJ(K-1)].$$

23.9. REGRESSION TREATMENT OF THE TWO-WAY CLASSIFICATION WITH EQUAL NUMBERS OF OBSERVATIONS IN THE CELLS

We could, if desired, handle this problem in a way similar to that given in Section 23.4, by writing down the *dependent* normal equations in the parameters μ, α_i, β_j, γ_{ij}, adding the equations given by the restrictions of Eq. (23.8.2), and solving a selected set of $(I+1)(J+1)$ independent equations. We shall instead deal with the situation in another way, so that nonsingular $\mathbf{X}'\mathbf{X}$ matrices will always occur in the calculations.

There are $1 + I + J + IJ = (I+1)(J+1)$ parameters in all, but these are dependent through the $1 + 1 + J + I - 1 = I + J + 1$ restrictions defined by Eq. (23.8.2). We have to take one off the (at first sight) apparent number of restrictions to allow for the fact that, if we know all the *row* sums of the γ_{ij} are zero, so is the total sum of all γ_{ij}. This means that it is necessary to specify only that $J-1$ of the column sums of the γ_{ij} be zero, in order to achieve zero for the final column sum. Thus we actually need only IJ parameters to describe the model and we can define these as

$$\delta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \qquad \begin{matrix} i = 1, 2, \ldots, I, \\ j = 1, 2, \ldots, J. \end{matrix}$$

Consider the following models:

1. $E(Y_{ijk}) = \delta_{ij}$.
2. $E(Y_{ijk}) = \delta_{\cdot j}$ independent of i.
3. $E(Y_{ijk}) = \delta_{i\cdot}$ independent of j.
4. $E(Y_{ijk}) = \delta$ independent of i and j.

We can represent all these in matrix form. Let

$$\mathbf{Y} = (Y_{111}, Y_{112}, \ldots, Y_{11K}; Y_{121}, Y_{122}, \ldots, Y_{12K}; \ldots; Y_{IJ1}, Y_{IJ2}, \ldots, Y_{IJK})',$$

where we order the cells in the sequence

$$(11), (12), \ldots, (1J); (21), (22), \ldots, (2J); \ldots; (I1), (I2), \ldots, (IJ)$$

and order the observations within the cells in the numerical order of the third subscript. Then, with the matrices indicated below, we can write all the models (1), (2), (3), and (4) in the form $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$.

Model (1)

$$\mathbf{X} = \begin{array}{c} \begin{matrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1J} & \delta_{21} & \delta_{22} & \cdots & \delta_{2J} & \cdots & \delta_{I1} & \delta_{I2} & \cdots & \delta_{IJ} \end{matrix} \\ \left[\begin{array}{cccccccccccc} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \hline \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \hline 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 1 \end{array}\right] \end{array},$$

where each row segment is of depth K.

$$\boldsymbol{\beta}' = (\delta_{11}, \delta_{12}, \ldots, \delta_{1J}; \delta_{21}, \delta_{22}, \ldots, \delta_{2J}; \ldots; \delta_{I1}, \delta_{I2}, \ldots, \delta_{IJ}).$$

Model (2)

$$\mathbf{X} = \begin{array}{c} \begin{matrix} \delta_{\cdot 1} & \delta_{\cdot 2} & \delta_{\cdot 3} & \cdots & \delta_{\cdot J} \end{matrix} \\ \left[\begin{array}{ccccc} \mathbf{j} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{j} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{j} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{j} \\ \hline \multicolumn{5}{c}{(I-1) \text{ more blocks}} \\ \multicolumn{5}{c}{\text{exactly as above}} \end{array}\right] \end{array},$$

where $\mathbf{j}$ denotes a $K \times 1$ vector of unities.

$$\boldsymbol{\beta}' = (\delta_{\cdot 1}, \delta_{\cdot 2}, \ldots, \delta_{\cdot J}).$$

Model (3)

$$\mathbf{X} = \begin{array}{c} \begin{matrix} \delta_{1\cdot} & \delta_{2\cdot} & \delta_{3\cdot} & \cdots & \delta_{I\cdot} \end{matrix} \\ \begin{bmatrix} \mathbf{j} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{j} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{j} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{j} \end{bmatrix} \end{array},$$

where $\mathbf{j}$ here denotes a $JK \times 1$ vector of unities.

$$\boldsymbol{\beta}' = (\delta_{1\cdot}, \delta_{2\cdot}, \ldots, \delta_{I\cdot}).$$

Model (4)

$$\mathbf{X} = \mathbf{j},$$

where $\mathbf{j}$ here denotes a $IJK \times 1$ vector of unities.

$$\boldsymbol{\beta} = \delta, \text{ a scalar.}$$

We can construct the standard analysis of variance table through regression methods as follows. Denote by S_1, S_2, S_3, and S_4 the regression sums of squares, which arise from the four regression models given above and let $S = \Sigma_{i=1}^{I} \Sigma_{j=1}^{J} \Sigma_{k=1}^{K} \Sigma Y_{ijk}^2$. Then using the "extra sum of squares" principle given in Section 6.1, we can construct the table below.

ANOVA

Source	df	MS
Rows	$I - 1$	$S_3 - S_4$
Columns	$J - 1$	$S_2 - S_4$
Interaction	$(I - 1)(J - 1)$	$S_1 - S_2 - S_3 + S_4$
Residual	$IJ(K - 1)$	$S - S_1$
Mean	1	S_4
Total	IJK	S

(The interaction sum of squares is actually obtained from

$$S_1 - (S_2 - S_4) - (S_3 - S_4) - S_4,$$

which reduces to the form given in the table.)

The equivalence between these sums of squares and the ones given in the analysis of variance procedure can easily be demonstrated mathematically but we shall not do this.

We are usually interested in obtaining estimates m, a_i, b_j, c_{ij} of the original parameters μ, α_i, β_j, γ_{ij} in the analysis of variance model. These estimates can be obtained from the estimated d_{ij}, $d_{\cdot j}$, $d_{i\cdot}$, d of the regression coefficients δ_{ij}, $\delta_{\cdot j}$, $\delta_{i\cdot}$, δ in the four models. They are

$$m = d,$$

$$a_i = d_{i\cdot} - d,$$

$$b_j = d_{\cdot j} - d,$$

$$c_{ij} = d_{ij} - d_{i\cdot} - d_{\cdot j} + d.$$

An Alternative Method

Our suggested method of dealing with the two-way analysis of variance classification has involved four symmetric regression analyses and the use of the extra sum of squares principle. To deal with this situation in a single regression analysis we must write an unsymmetric model that omits some of the dependent parameters of the standard analysis of variance model. We illustrate this with an example. Consider the two-way classification below, which has two observations in each cell:

	Column $j = 1$	$j = 2$
Row $i = 1$	Y_1, Y_2	Y_3, Y_4
$i = 2$	Y_5, Y_6	Y_7, Y_8
$i = 3$	Y_9, Y_{10}	Y_{11}, Y_{12}

The standard analysis of variance model is

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j + \gamma_{ij},$$

where

$$\alpha_1 + \alpha_2 + \alpha_3 = 0 \qquad \gamma_{11} + \gamma_{12} = 0$$
$$\beta_1 + \beta_2 = 0 \qquad \gamma_{21} + \gamma_{22} = 0$$
$$\gamma_{31} + \gamma_{32} = 0$$
$$\gamma_{11} + \gamma_{21} + \gamma_{31} = 0$$
$$\gamma_{12} + \gamma_{22} + \gamma_{32} = 0$$

Thus if (for example) μ, α_1, α_2, β_1, γ_{11}, and γ_{21} are known or estimated, all other parameters or their estimates, respectively, can be found from the restrictions. We can thus write the regression model

$$E(Y_{ij}) = \mu + \alpha_1 X_1 + \alpha_2 X_2 + \beta_1 X_3 + \gamma_{11} X_4 + \gamma_{21} X_5$$

or

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta},$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \\ Y_9 \\ Y_{10} \\ Y_{11} \\ Y_{12} \end{bmatrix}, \qquad \mathbf{X} = \begin{array}{c} \begin{matrix} \mu & \alpha_1 & \alpha_2 & \beta_1 & \gamma_{11} & \gamma_{21} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \end{array},$$

$$\boldsymbol{\beta}' = (\mu, \alpha_1, \alpha_2, \beta_1, \gamma_{11}, \gamma_{21}).$$

Note: The elements of the γ_{ij} column are obtained as the product of corresponding elements of the α_i and β_j columns.

Any independent subset of the parameters can be used for such a model and there are many possible alternative forms. From this point, the usual regression methods are used to estimate $\boldsymbol{\beta}$. Because of the orthogonality in $\mathbf{X}$, we can obtain separate, orthogonal sums of squares for the estimates of (1) μ, (2) α_1 and α_2, (3) β_1, (4) γ_{11} and γ_{21}. These will be the usual sums of squares for (1) mean, (2) rows, (3) columns, (4) interaction, in the standard analysis of variance setup.

23.10. EXAMPLE: THE TWO-WAY CLASSIFICATION

The data in the two-way classification below appear in Brownlee (1965, p. 475). Descriptive details are omitted.

	Column 1	Column 2	Column 3
Row 1	17, 21, 49, 54	64, 48, 34, 63	62, 72, 61, 91
Row 2	33, 37, 40, 16	41, 64, 34, 64	56, 62, 57, 72

Following the procedure given in Section 23.9 we can calculate the quantities below through regression methods.

(1) $S_1 = 65{,}863$ $\quad \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \end{pmatrix} = \begin{pmatrix} 35.25 & 52.25 & 71.50 \\ 31.50 & 50.75 & 61.75 \end{pmatrix}.$

(2) $S_2 = 65{,}640.25$ $\quad (d_{\cdot 1}, d_{\cdot 2}, d_{\cdot 3}) = (33.375, 51.5, 66.625).$

(3) $S_3 = 61{,}356$ $\quad \begin{pmatrix} d_{1\cdot} \\ d_{2\cdot} \end{pmatrix} = \begin{pmatrix} 53 \\ 48 \end{pmatrix}.$

(4) $S_4 = 61{,}206$ $\quad d = 50.5.$

Using the formulas given in the previous section we obtain an analysis of variance table as follows:

ANOVA

Source	df	SS
Rows	1	150.00
Columns	2	4,434.25
Interaction	2	72.75
Residual	18	3,495.00
Mean	1	61,206.00
Total (uncorrected)	24	69,358.00

The same table was obtained in Brownlee (1965) through the usual analysis of variance calculations. The estimates of the usual analysis of variance parameters are

given by

$$m = 50.5,$$

$$a_1 = 53 - 50.5 = 2.5, \qquad a_2 = 48 - 50.5 = -2.5.$$

Note: $a_1 + a_2 = 0$, as it should.

$$b_1 = 33.375 - 50.5 = -17.125, \qquad b_2 = 51.5 - 50.5 = 1.0,$$

$$b_3 = 66.625 - 50.5 = 16.125.$$

Note: $b_1 + b_2 + b_3 = 0$ as it should.

$$c_{11} = 35.25 - 53 - 33.375 + 50.5 = -0.625,$$

$$c_{12} = 52.25 - 53 - 51.5 + 50.5 = -1.750,$$

$$c_{13} = 71.50 - 53 - 66.625 + 50.5 = 2.375,$$

$$c_{21} = 31.50 - 48 - 33.375 + 50.5 = 0.625,$$

$$c_{22} = 50.75 - 48 - 51.5 + 50.5 = 1.750,$$

$$c_{23} = 61.75 - 48 - 66.625 + 50.5 = -2.375.$$

Note: $\Sigma_{i=1}^{2} c_{ij} = \Sigma_{j=1}^{3} c_{ij} = 0$, as they should.

The residuals from this analysis of variance model would be

$$\begin{aligned} e_{ijk} &= Y_{ijk} - m - a_i - b_j - c_{ij} \\ &= Y_{ijk} - d - (d_{i\cdot} - d) - (d_{\cdot j} - d) - (d_{ij} - d_{i\cdot} - d_{\cdot j} + d) \\ &= Y_{ijk} - d_{ij}, \end{aligned}$$

which are the residuals from the regression analysis using Model (1). These would be examined in the usual ways. Also, plots of residuals for each row and column can be examined.

23.11. RECAPITULATION AND COMMENTS

We have seen, in the specific cases discussed, that analysis of variance can, if necessary, be conducted by standard regression techniques. If the model is examined carefully and reparameterized properly, the analysis of variance table for other models could also be obtained in a similar way. The proper selection of the dummy variables is crucial to the proper presentation of the detailed results and can ease the computations considerably. However, many reparameterizations are valid, as is generally true in dummy variable situations; some reparameterizations are simply better than others for certain purposes. The more complicated the design is, the more complicated will be the regression form conversion, and the larger the **X** matrix. The effort required can be considerable and, when the design is standard and all the data are available (i.e., there are no missing observations), is usually not worthwhile. It is usually better to employ an appropriate analysis of variance computation method, or computer routine, where it exists. Nevertheless, it is useful to appreciate the connection between the two methods of analysis for several reasons:

1. It focuses attention on the fact that a model is necessary in analysis of variance problems.

2. It points up the fact that the residuals in analysis of variance models play the same role as residuals in regression models and *must* be examined for the information they contain on the possible inadequacy of the model under consideration. (There seems to be a tacit assumption in most variance analysis that the model is correct.)

3. When observations are missing from analysis of variance data, they can often be "estimated" by standard formulas. If this is inconvenient, or too many observations are missing, the data can usually be analyzed by a regression routine setting up models as illustrated above but deleting rows of the **X** matrices for which no observations are available. (The word estimated is placed in quotes because no real estimation takes place. The "estimates" are simply numbers inserted for calculation purposes, which lead to the same estimates of parameters that would otherwise have been obtained from the incomplete data via regression analysis.)

4. Adding additional predictor variables (covariates) to an analysis of variance structure is simple via regression.

EXERCISES FOR CHAPTER 23

A. Below, denoted by a and b, are two-way classification experiments. On each of these experiments, perform these analyses:

1. Analyze these data using any of the methods described in Chapter 23.

2. Evaluate the fitted values and residuals and examine the residuals in all reasonable ways. State any defects you find.

3. Confirm that the alternative regression methods given in Chapter 23 lead to the same results.

a. An experiment was conducted to determine the effect of steam pressure and blowing time on the percentage of foreign matter left in filter earth. The data are as follows:

Steam Pressure (pounds)	Percentage of Foreign Matter		
	Blowing Time (hours)		
	1	2	3
10	45.2, 46.0	40.0, 39.0	35.9, 34.1
20	41.8, 20.6	27.8, 19.0	22.5, 17.7
30	23.5, 33.1	44.6, 52.2	42.7, 48.6

b. An experiment was conducted on the effect of premixing speed and finish mixer speed on the center heights of cakes. Three different levels of speed were chosen for each of the two variables. The data collected were as follows:

(Premix speed −5)	(Finish mix speed −3.5) × 2	(Center height −2) × 100
X_1	X_2	Y
−1	−1	4, −3
−1	0	3, 2
−1	1	−1, −5
0	−1	3, 10
0	0	2, 2
0	1	0, 0
1	−1	−1, −10
1	0	1, 2
1	1	7, 9

B. A chemical experiment was performed to investigate the effect of extrusion temperature X_1 and cooling temperature X_2 on the compressibility of a finished product. Knowledge of the process suggested that a model of the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \epsilon$$

would satisfactorily explain the variation observed. Two levels of extrusion temperature and two levels of cooling temperature were chosen and all four of the combinations were performed. Each of the four experiments was carried out four times and the data yielded the following information:

ANOVA

Source of Variation	df	SS	MS
Total	16	921.0000	
Due to regression		881.2500	
b_0	1	798.0625	
b_1	1	18.0625	
b_2			
b_{12}		5.0625	
Residual			

1. a. Complete the above ANOVA table.

b. Using $\alpha = 0.05$, examine the following questions:

(1) Is the overall regression equation given b_0 statistically significant?

(2) Are all the b-coefficients significant?

2. Given the following additional information,

$$\Sigma Y_i = 113, \quad \Sigma x_{1i} y_i = 31, \quad \text{and} \quad \Sigma x_{1i} x_{2i} y_i = -9,$$

where $x_{ji} = X_{ji} - \overline{X}_j$, $j = 1, 2$, and $y_i = Y_i - \overline{Y}$:

a. Determine b_0, b_1, b_2, and b_{12}, and write out the prediction equation.

b. The predicted value of $\hat{Y}$ at $X_1 = 70°$ and $X_2 = 150°$ is 54. The variance of this predicted value is 0.6875. What is the variance of a single predicted observation at the point $X_1 = 70$, $X_2 = 150$?

c. Place 95% confidence limits on the true mean value of Y at the point $X_1 = 70$, $X_2 = 150$.

3. What conclusions can be drawn from your analysis?

C. Analyze the data given in Section 23.10 by the alternative method of Section 23.9.

D. (*Source*: *Introduction to Statistical Inference* by E. S. Keeping, Van Nostrand, Princeton, NJ, 1962, p. 216.) Apply the regression method of either Section 23.4 or Section 23.5 to the one-way classification data below. Actually carry out the calculation steps of your selected method, rather than just going to the analysis of variance table. The data were taken to test the effect of adding a small percentage of coal dust to the sand used for making concrete. Several batches were mixed under practically identical conditions except for the variation in the percentage of coal. From each batch, four cylinders were made and tested for breaking strength in pounds per square inch (lb/in.^{2}). One cylinder in the third sample was defective, so there were only three items in this sample.

Sample Number:	1	2	3	4	5
Percentage coal:	0	0.05	0.1	0.5	1.0
Breaking strengths:	1690	1550	1625	1725	1530
	1580	1445	1450	1550	1545
	1745	1645	1510	1430	1565
	1685	1545		1445	1520

(*Hint:* Work with $Y_{ij} - 1430$, it is easier; ask yourself what difference it will make to the analysis.)

E. An important step in paper manufacturing is the removal of water from the paper. In a particular process, four factors were thought to affect the amount of water removed, and it was decided to run a 2^4 factorial design to examine them. The factors and the two levels (in appropriate units) chosen for each, were as follows:

Factor	Designation	Low Level	High Level
Vacuum on pressure roll 2 (hg)	*A*	0	18
Vacuum on pressure roll 1 (hg)	*B*	0	19
Weight of the paper (lb)	*C*	10.0	13.7
Process line speed (fpm)	*D*	1700	2000

The data are given in the table below. The notation *bcd*, for example, designates the run with *A* at its low level and *B*, *C*, and *D* at their high levels; the "1" indicates all factors at their low levels.

Run Number	Factor Combination	Percentage of Water Removed, *Y*
1	*bcd*	39.7
2	*abcd*	41.1
3	*cd*	40.6
4	*acd*	40.4
5	*ad*	41.0
6	*bd*	37.6
7	*d*	38.7
8	*abd*	39.0
9	*bc*	38.9
10	*ac*	40.0
11	*abc*	41.0
12	*c*	42.9
13	*a*	40.2
14	*b*	35.4
15	*ab*	39.4
16	1	39.0

Requirements

1. Analyze these data using regression techniques.
2. Show that the factorial design approach and the analysis of variance approach to these data provide exactly the same results and/or conclusions.

F. The data below came from a one way analysis of variance problem with (overparameterized) model

$$Y_{ij} = \mu + t_i + \epsilon_{ij}, \quad i,j = 1, 2, 3.$$

Group	Y_{ij}'s	Row Sum	Row SS
1	3, 4, 5	12	50
2	4, 6, 8	18	116
3	6, 8, 10	24	200
Column sums		54	366

1. Write down the **X** matrix implied by this model, and explain why using this **X** may not be a good idea in a regression context.
2. Suggest any one way of resolving the difficulty, perform the analysis, and test the null hypothesis that all groups have the same effects.

G. (*Source*: "Obtaining a sterilized soil for the growth of *Onoclea* gametophytes," by G. Rubin and D. J. Paolillo, *The New Phytologist* **97**, 1984, 621–628. *Onoclea sensibilis* L. is a fern.) On p. 623 of the source reference, we read: "The sterilized soil within three tubes was divided into five levels, and the soil from each level was placed into individual Petri plates. Spore germination was assessed for each plate. No clear pattern emerged for the variation within and between tubes."

1. The data are given in Table G1. Use regression methods with dummy variables (any workable set will do) to confirm the results quoted below Table G1.
2. If, instead, the data had been as in Table G2, what would the conclusion have been? (*Note*: Table G2 has no basis in fact but is simply a partial rearrangement of the numbers in Table G1, made in order to provide a second exercise.)

TABLE G1. Germination of *Onoclea* Spores on Gamma-Irradiated Soil

Level Within Tube	Germination, % in Tube Number			Level Means
	1	2	3	
1	78.4	84.5	62.1	75.0
2	10.6	77.5	94.4	60.8
3	62.7	4.4	24.8	30.6
4	92 5	17.5	0.9	37.0
5	11.6	0.9	20.4	11.0
Tube means	51.2	37.0	40.5	42.9

$F^{2,8}_{\text{tube}} = 0.21$, $F^{4,8}_{\text{level}} = 1.48$. Neither is significant at the 5% level.

TABLE G2. Fake (Rearranged) Data

Level Within Tube	Germination, % in Tube Number			Level Means
	1	2	3	
1	78.4	84.5	62.1	75.0
2	92.5	77.5	94.4	88.1
3	62.7	20.4	24.8	36.0
4	10.6	17.5	11.6	13.2
5	0.9	0.9	4.4	2.1
Tube means	49.0	40.2	39.5	42.9

H. (*Source*: "Hybridization and mating behavior in *Aedes aegypti* (diptera: culicidae)," by D. F. Moore, *Journal of Medical Entomology*, **16**, 1979, No. 3, 223–226.) A study on two forms of the yellow fever mosquito, the light-colored domestic (D) and the dark-colored sylvan (S) in Kenya, examined various matings. Table H shows the cumulative mean hatchabilities of DD, DS, SD, and SS matings where the first letter refers to female and the second to male mosquitos. In the original paper, equality of population means of the four matings was tested using (nonparametric) Kruskal–Wallis tests because "the hatchabilities are not known to be normally distributed" (p. 225). (In particular, there is correlation column-wise.) Nevertheless:

1. Perform a regression analysis with model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{23} X_2 X_3 + \epsilon$ using these variables

$$Y = \text{hatchability},$$
$$X_1 = \text{number of immersions in water}$$
$$X_2 = -1 \text{ for female, } 1 \text{ for male, domestic (D)},$$
$$X_3 = 1 \text{ for female, } -1 \text{ for male, sylvan (S)}.$$

Are there differences between the four groups, do you think? Also try these regression analyses as well:

2. Replace Y by $n_i^{1/2} \sin^{-1}(Y^{1/2})$.
3. Replace X_1 by $\ln X_1$
4. Make both replacements (2) and (3).

TABLE H. Cumulative Mean Hatchabilities of Eggs Resulting from Matings Among Forms of *Aedes aegypti*

Number of Immersions in Water	Matings			
	DD	DS	SD	SS
1	0.665	0.469	0.284	0.127
2	0.736	0.504	0.354	0.183
3	0.782	0.559	0.404	0.195
8	0.899	0.770	0.577	0.297
n_i	2024	541	1100	222

n_i = total number of eggs per column.