

Analyses of Variance and Covariance as General Linear Models

Objectives

To show how the analysis of variance can be viewed as a special case of multiple regression; to present procedures for the treatment of unequal sample sizes; to present the analysis of covariance.

Contents

- 16.1** The General Linear Model
- 16.2** One-Way Analysis of Variance
- 16.3** Factorial Designs
- 16.4** Analysis of Variance with Unequal Sample Sizes
- 16.5** The One-Way Analysis of Covariance
- 16.6** Computing Effect Sizes in an Analysis of Covariance
- 16.7** Interpreting an Analysis of Covariance
- 16.8** Reporting the Results of an Analysis of Covariance
- 16.9** The Factorial Analysis of Covariance
- 16.10** Using Multiple Covariates
- 16.11** Alternative Experimental Designs

MOST PEOPLE THINK OF MULTIPLE regression and the analysis of variance as two totally separate statistical techniques that answer two entirely different sets of questions. In fact, this is not at all the case. In the first place they ask the same kind of questions, and in the second place they return the same kind of answers, although the answers may be phrased somewhat differently. The analysis of variance tells us that three treatments (T_1 , T_2 , and T_3) have different means (\bar{X}_i). Multiple regression tells us that means (\bar{Y}_i) are related to treatments (T_1 , T_2 , and T_3), which amounts to the same thing. Furthermore, the analysis of variance produces a statistic (F) on the differences among means. The analysis of regression produces a statistic (F) on the significance of R . As we shall see shortly, these F s are equivalent.

16.1 The General Linear Model

general linear model

Just as multiple regression and the analysis of variance are concerned with the same general type of question, so are they basically the same technique. In fact, the analysis of variance is a special case of multiple linear regression, which in turn is a special case of what is commonly referred to as the **general linear model**. The fact that the analysis of variance has its own formal set of equations can be attributed primarily to good fortune. It happens that when certain conditions are met (as they are in the analysis of variance), the somewhat cumbersome multiple-regression calculations are reduced to a few relatively simple equations. If it were not for this, there probably would not be a separate set of procedures called the analysis of variance.

For the student interested solely in the application of statistical techniques, a word is in order in defense of even including a chapter on this topic. Why, you may ask, should you study what amounts to a cumbersome way of doing what you already know how to do in a simple way? Aside from the cry of “intellectual curiosity,” there are several practical (applied) answers to such a question. First, this approach represents a relatively straightforward way of handling particular cases of unequal sample sizes, and understanding this approach helps you make intelligent decisions about various options in statistical software. Second, it provides us with a simple and intuitively appealing way of running, and especially of understanding, an analysis of covariance—which is a very clumsy technique when viewed from the more traditional approach. Last, and most important, it represents a glimpse at the direction in which statistical techniques are moving. With the greatly extended use of powerful and fast computers, many of the traditional statistical techniques are giving way to what were previously impractical procedures. We saw an example when we considered the mixed models approach to repeated measures analysis of variance. Other examples are such techniques as structural equation modeling and that old and much-abused standby, factor analysis. Unless you understand the relationship between the analysis of variance and the general linear model (as represented by multiple linear regression), and unless you understand how the data for simple analysis of variance problems can be cast in a multiple-regression framework, you will find yourself in the near future using more and more techniques about which you know less and less. This is not to say that t , χ^2 , F , and so on are likely to disappear, but only that other techniques will be added, opening up entirely new ways of looking at data. The recent rise in the use of Structural Equation Modeling is a case in point, because much of what that entails builds on what you already know about regression, and what you will learn about underlying models of processes.

In the past 25 years, several excellent and very readable papers on this general topic have been written. The clearest presentation is still Cohen (1968). A paper by Overall and Spiegel (1969) is also worth reviewing. Both of these papers appeared in the *Psychological*

Bulletin and are therefore readily available. Other good discussions can be found in Overall (1972), Judd and McClelland (1989), and Cohen, Cohen, West, and Aiken (2003). Also, Cramer and Appelbaum (1980) and Howell and McConaughy (1982) provide contrasting views on the choice of the underlying model and the procedures to be followed.

There are two different ways to read this chapter, both legitimate. The first is to look for general concepts and to go lightly over the actual techniques of calculation. That is the approach I often tell my students to follow. I want them to understand where the reasoning leads, and I want them to feel that they could carry out all of the steps if they had to (with the book in front of them), but I don't ask them to commit very much of the technical material to memory. On the other hand, some instructors may want their students to grasp the material at a deeper level. There are good reasons for doing so. But I would still suggest that the first time you read the chapter, you look for general understanding. To develop greater expertise, sit down with both a computer and a calculator and work lots and lots of problems.

The Linear Model

Consider first the traditional multiple-regression problem with a criterion (Y) and three predictors (X_1 , X_2 , and X_3). We can write the usual model

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + e_i$$

or, in terms of *vector* notation

$$\mathbf{y} = \mathbf{b}_0 + b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3 + \mathbf{e}$$

where \mathbf{y} , \mathbf{x}_1 , \mathbf{x}_2 , and \mathbf{x}_3 are $(n \times 1)$ vectors (columns) of data, \mathbf{e} is a $(n \times 1)$ vector of errors, and \mathbf{b}_0 is a $(n \times 1)$ vector whose elements are the intercept. This equation can be further reduced to

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where \mathbf{X} is a $n \times (p + 1)$ matrix of predictors, the first column of which is 1s, and \mathbf{b} is a $(p + 1) \times 1$ vector of regression coefficients¹. This called a linear model because Y is the sum of a linear combination of predictor variables—nothing is raised to a power other than 1.

Now consider the traditional model for a one-way analysis of variance:

$$Y_{ij} = \mu + \tau_j + e_{ij}$$

Here the symbol τ_j is simply a shorthand way of writing $\tau_1, \tau_2, \tau_3, \dots, \tau_p$, where for any given subject we are interested in only that value of τ_j that pertains to the particular treatment in question. To see the relationship between this model and the traditional regression model, it is necessary to introduce the concept of a design matrix. Design matrices are used in a wide variety of situations, not simply the analysis of variance, so it is important to understand them.

Design Matrices

design matrix

A **design matrix** is a matrix of *coded*, or *dummy*, or *counter* variables representing group membership. The *complete* form of the design matrix (X) will have $p + 1$ columns, representing the mean (μ) and the p treatment effects. A subject is always scored 1 for μ , since

¹ Although a few equations in this chapter are laid out in matrix format, you do not need to know the rules of matrix algebra to understand the material.

μ is part of all observations. In all other columns, she is scored 1 if she is a member of the treatment associated with that column, and 0 otherwise. Thus, for three treatments with two subjects per treatment, the complete design matrix would be

$$\Sigma = \begin{matrix} S & \mu & A_1 & A_2 & A_3 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

Notice that subjects 1 and 2 (who received Treatment A_1) are scored 1 on μ and A_1 , and 0 on A_2 and A_3 , since they did not receive those treatments. Similarly, subjects 3 and 4 are scored 1 on μ and A_2 , and 0 on A_1 and A_3 .

We will now define the vector τ of treatment effects as $[\mu \ \tau_1 \ \tau_2 \ \tau_3]$. Taking \mathbf{X} as the design matrix, the analysis of variance model can be written in matrix terms as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e}$$

which can be seen as being of the same form as the traditional regression equation. The elements of $\boldsymbol{\tau}$ are the effects of each dummy treatment variable, just as the elements of b in the regression equation are the effects of each independent variable. Expanding, we obtain

$$\mathbf{y} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{21} \\ e_{12} \\ e_{22} \\ e_{13} \\ e_{23} \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X} \times \boldsymbol{\tau} + \mathbf{e}$$

which, following the rules of matrix multiplication, produces

$$\begin{aligned} Y_{11} &= \mu + \tau_1 + e_{11} \\ Y_{21} &= \mu + \tau_1 + e_{21} \\ Y_{12} &= \mu + \tau_2 + e_{12} \\ Y_{22} &= \mu + \tau_2 + e_{22} \\ Y_{13} &= \mu + \tau_3 + e_{13} \\ Y_{23} &= \mu + \tau_3 + e_{23} \end{aligned}$$

For each subject we now have the model associated with her response. Thus, for the second subject in Treatment 2, $Y_{22} = \mu + \tau_2 + e_{22}$, and for the i th subject in Treatment j , we have $Y_{ij} = \mu + \tau_j + e_{ij}$, which is the usual analysis of variance model.

The point is that the design matrix allows us to view the analysis of variance in a multiple-regression framework, in that it permits us to go from

$$Y_{ij} = \mu + \tau_j + e_{ij} \quad \text{to} \quad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

Moreover, the elements of \mathbf{b} are the values of $\mu, \tau_1, \tau_2, \dots, \tau_k$. In other words, these are the actual treatment effects in which we are interested.

The design matrix we have been using has certain technical limitations that must be circumvented. We are going to turn it into something that looks quite different but actually carries all of the same information. First, the matrix is redundant in the sense that if we are told that a subject is not in A_1 or A_2 , we know without being told that she must be in A_3 . This is another way of saying that there are only 2 *df* for treatments. For this reason we will eliminate the column headed A_3 , leaving only $a - 1$ columns for the treatment effects. (We could eliminate any one of the three columns, but we usually drop the last one.) A second change is necessary if we want to use any computer program that obtains a multiple-regression equation by way of first calculating the intercorrelation matrix. The column headed μ has no variance, and therefore cannot enter into a standard multiple-regression program—it would cause us to attempt division by 0. Thus, it too must be eliminated. This is no real loss, since our ultimate solution will not be affected. In fact, the software will sneak it back in.

One further change will be made simply for the sake of allowing us to test the desired null hypotheses using the method to be later advocated for factorial designs. Because we have omitted a column dealing with the third (or a th) level of treatments, solutions given our modified design matrix would produce estimates of treatment effects in relation to \bar{X}_3 rather than in relation to \bar{X} . In other words, b_1 would turn out to be $(\bar{X}_1 - \bar{X}_3)$ rather than $(\bar{X}_1 - \bar{X})$. This is fine if that's what you want, but I would much rather see treatment effects as deviations from the grand mean. It just seems tidier. So we will modify the design matrix to make the mean (\bar{X}_i) of each column of \mathbf{X} equal to 0. Under this new system, a subject is scored 1 in column A_i if she is a member of Treatment A_i ; she is scored -1 if she is a member of the a th (last) treatment; and she is scored 0 if neither of these conditions apply. (This restriction corresponds to the fixed-model analysis of variance requirement that $\sum \tau_i = 0$.)

These modifications have led us from

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{to} \quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{to} \quad \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{to} \quad \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix}$$

Although these look like major changes in that the last form of \mathbf{X} appears to be far removed from where we started, it actually carries all the necessary information. We have merely eliminated redundant information, removed a constant term, and then caused the treatment effects to be given as deviations from \bar{X} .

16.2 One-Way Analysis of Variance

At this point a simple example is in order. Table 16.1 contains data for three subjects in each of four treatments. Table 16.1b shows the summary table for the corresponding analysis of variance, along with the value of η^2 (discussed in Chapter 11). Table 16.1c contains the estimated treatment effects ($\hat{\tau}_i$) where $\hat{\tau}_i = \hat{\mu}_i - \hat{\mu}$. Because the fixed-model analysis of variance imposes the restriction that $\sum \tau_i = 0$, τ_4 is automatically defined by τ_1, τ_2 , and τ_3 ($\tau_4 = 0 - \sum \tau_j$).

Table 16.1 Illustrative calculations for simple one-way design with equal *n*s

(a) Data

Treatment 1	Treatment 2	Treatment 3	Treatment 4
8	5	3	6
9	7	4	4
7	3	1	9
8	5	2.667	6.333
			$\bar{X}_{..} = 5.500$

(b) Summary Table

Source	df	SS	MS	F	η^2
Treatments	3	45.667	15.222	4.46	.626
Error	8	27.333	3.417		
Total	11	73.000			

(c) Estimated Treatment Effects

$$\hat{\tau}_1 = \bar{X}_1 - \bar{X}_{..} = 8.0 - 5.5 = 2.5$$

$$\hat{\tau}_2 = \bar{X}_2 - \bar{X}_{..} = 5.0 - 5.5 = -0.5$$

$$\hat{\tau}_3 = \bar{X}_3 - \bar{X}_{..} = 2.67 - 5.5 = -2.83$$

© Cengage Learning 2013

Now let us approach the statistical treatment of these data by means of least-squares multiple linear regression. We will take as our criterion (*Y*) the raw data in Table 16.1. For the predictors we will use a design matrix *of the form*

$$\mathbf{X} = \begin{matrix} & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 \\ \text{Treatment 1} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix} \end{matrix}$$

Here the elements of any one row of the design matrix are taken to apply to *all the subjects in the treatment*. The multiple-regression solution using the design matrix **X** as the matrix of predictors is presented in Exhibit 16.1. Here the dependent variable (**Y**) is the first column of the data matrix. The next three columns together form the matrix **X**. SPSS was used to generate this solution, but any standard program would be suitable. (I have made some very minor changes in the output to simplify the discussion.)

Notice the patterns of intercorrelations among the **X** variables in Exhibit 16.1. This type of pattern with constant off-diagonal correlations will occur whenever there are equal numbers of subjects in the various groups. (The fact that we don't have constant off-diagonal correlations with unequal-*n* factorial designs is what makes our life more difficult in those situations.)

Notice that the regression coefficients are written in a column. This column can be called a *vector*, and is the vector **b**, or, in analysis of variance terms, the vector τ . Notice that $b_1 = 2.50$, which is the same as the estimated treatment effect of Treatment 1 shown in Table 16.1. In other words, $b_1 = \tau_1$. This also happens for b_2 and b_3 . This fact necessarily follows from our definition of **X** and τ . Moreover, if we were to

Tab16-1.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Ad

1: Y 8

	Y	X1	X2	X3
1	8.00	1.00	.00	.00
2	9.00	1.00	.00	.00
3	7.00	1.00	.00	.00
4	5.00	.00	1.00	.00
5	7.00	.00	1.00	.00
6	3.00	.00	1.00	.00
7	3.00	.00	.00	1.00
8	4.00	.00	.00	1.00
9	1.00	.00	.00	1.00
10	6.00	-1.00	-1.00	-1.00
11	4.00	-1.00	-1.00	-1.00
12	9.00	-1.00	-1.00	-1.00

SPSS Inc.

Correlations

		Y	X1	X2	X3
Pearson Correlation	Y	1.000	.239	-.191	-.526
	X1	.239	1.000	.500	.500
	X2	-.191	.500	1.000	.500
	X3	-.526	.500	.500	1.000

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.791(a)	.626	.485	1.848

^a Predictors: (Constant), X3, X2, X1

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.500	.534		10.307	.000
	X1	2.500	.924	.717	2.705	.027
	X2	-.500	.924	-.143	-.541	.603
	X3	-2.833	.924	-.812	-3.066	.015

^a Dependent Variable: Y

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	45.667	3	15.222	4.455	.040 ^a
	Residual	27.333	8	3.417		
	Total	73.000	11			

^aPredictors: (Constant), X3, X2, X1

^bDependent Variable: Y

Exhibit 16.1 SPSS regression analysis of data in Table 16.1

examine the significance of the b_i , given as the column of t -ratios, we would simultaneously have tests on the hypothesis ($H_0: \tau_j = \mu_i - \mu = 0$). Notice further that the intercept (b_0) is equal to the grand mean (\bar{Y}). This follows directly from the fact that we scored the a th treatment as -1 on all coded variables. Using the (-1) coding, the mean of every column of \mathbf{X} (\bar{X}_j) is equal to 0 and, as a result, $\sum b_1 \bar{X}_j = 0$ and therefore $b_0 = \bar{Y} - \sum b_1 \bar{X}_j = \bar{Y} - 0 = \bar{Y}$. This situation holds only in the case of equal n s, because otherwise \bar{X}_i would not be 0 for all i . However, in all cases, b_0 is our best estimate of μ in a least squares sense.

The value of $R^2 = .626$ is equivalent to η^2 , because they both estimate the percentage of variation in the dependent variable accounted for by variation among treatments.

If we test R^2 for significance, we have $F = 4.46$, $p = .040$. This is the F value we obtained in the analysis of variance, although this F can be found by the formula that we saw for testing R^2 in Chapter 15.

$$F(p, N - p - 1) = \frac{R^2(N - p - 1)}{(1 - R^2)p}$$

$$F(3, 8) = \frac{.626(8)}{.374(3)} = 4.46$$

Notice that the sums of squares for Regression, Error, and Total in Exhibit 16.1 are exactly equivalent to the sums of squares for Between, Error, and Total in Table 16.1. This equality makes clear that there is complete correspondence between sums of squares in regression and the analysis of variance.

The foregoing analysis has shown the marked similarity between the analysis of variance and multiple regression. This is primarily an illustration of the fact that there is no important difference between asking whether different treatments produce different means, and asking whether means are a function of treatments. We are simply looking at two sides of the same coin.

We have discussed only the most common way of forming a design matrix. This matrix could take a number of other useful forms, but we won't go into that here. For a good discussion of these, see Cohen (1968).

16.3 Factorial Designs

We can readily extend the analysis of regression to two-way and higher-order factorial designs, and doing so illustrates some important features of both the analysis of variance and the analysis of regression. (A good discussion of this approach, and the decisions that need to be made, can be found in Harris (2005).) We will consider first a two-way analysis of variance with equal n s.

The Full Model

The most common model for a two-way analysis of variance is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

As we did before, we can expand the α_i and β_j terms by using a design matrix. But then how should the interaction term be handled? The answer to this question relies on the fact

that an interaction represents a multiplicative effect of the component variables. Suppose we consider the simplest case of a 2×2 factorial design. Letting the entries in each row represent the coefficients for all subjects in the corresponding cell of the design, we can write our design matrix as

$$\mathbf{X} = \begin{matrix} & \begin{matrix} A_1 & B_1 & AB_{11} \end{matrix} \\ \begin{matrix} a_1b_1 \\ a_1b_2 \\ a_2b_1 \\ a_2b_2 \end{matrix} & \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \end{matrix}$$

The first column represents the main effect of A , and distinguishes between those subjects who received A_1 and those who received A_2 . The next column represents the main effect of B , separating B_1 subjects from B_2 subjects. The third column is the interaction of A and B . Its elements are obtained by multiplying the corresponding elements of columns 1 and 2. Thus, $1 = 1 \times 1$, $-1 = 1 \times -1$, $-1 = -1 \times 1$, and $1 = -1 \times -1$. Once again, we have as many columns per effect as we have degrees of freedom for that effect. We have no entries of 0 simply because with only two levels of each variable a subject must either be in the first or last level.

Now consider the case of a 2×3 factorial. With two levels of A and three levels of B , we will have $df_A = 1$, $df_B = 2$, and $df_{AB} = 2$. This means that our design matrix will require one column for A and two columns each for B and AB . This leads to the following matrix:

$$\mathbf{X} = \begin{matrix} & \begin{matrix} A_1 & B_1 & B_2 & AB_{11} & AB_{13} \end{matrix} \\ \begin{matrix} a_1b_1 \\ a_1b_2 \\ a_1b_3 \\ a_2b_1 \\ a_2b_2 \\ a_2b_3 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 & -1 \\ -1 & -1 & -1 & 1 & 1 \end{bmatrix} \end{matrix}$$

Column A_1 distinguishes between those subjects who are in treatment level A_1 and those in treatment level A_2 . Column 2 distinguishes level B_1 subjects from those who are not in B_1 , and Column 3 does the same for level B_2 . Once again, subjects in the first $a-1$ and first $b-1$ treatment levels are scored 1 or 0, depending on whether or not they served in the treatment level in question. Subjects in the a th or b th treatment level are scored -1 for each column related to that treatment effect. The column labeled AB_{11} is simply the product of columns A_1 and B_1 , and AB_{12} is the product of A_1 and B_2 .

The analysis for a factorial design is more cumbersome than the one for a simple one-way design, since we wish to test two or more main effects and one or more interaction effects. If we consider the relatively simple case of a two-way factorial, however, you should have no difficulty generalizing it to more complex factorial designs. The basic principles are the same—only the arithmetic is messier.

As an illustration, we will consider a case of a 2×4 factorial with four subjects per cell. Such a design is analyzed by the conventional analysis of variance in Table 16.2, which also includes means, estimated effects, and values of η^2 . From the summary

Table 16.2 Sample data and summary table for 2 × 4 factorial design

(a) Data

	B_1	B_2	B_3	B_4	Means
A_1	5	2	8	11	
	7	5	11	15	
	9	7	12	16	
	8	3	14	10	
	7.25	4.25	11.25	13.00	8.92750
A_2	7	3	9	11	
	9	8	12	14	
	10	9	14	10	
	9	11	8	12	
	8.75	7.75	10.75	11.75	9.75000
Means	8.000	6.000	11.000	12.375	9.34375

(b) Summary Table

Source	df	SS	MS	F	η^2
A	1	5.282	5.282	<1	.014
B	3	199.344	66.448	11.452*	.537
AB	3	27.344	9.115	1.571	.074
Error	24	139.250	5.802		
Total	31	371.220			

* $p < .05$

(c) Estimated Treatment Effects

$$\hat{\mu} = 9.34375$$

$$\hat{\alpha}_1 = \bar{A}_1 - \bar{X}_{..} = 8.9375 - 9.34375 = -0.40625$$

$$\hat{\beta}_1 = \bar{B}_1 - \bar{X}_{..} = 8.0000 - 9.34375 = -1.34375$$

$$\hat{\beta}_2 = \bar{B}_2 - \bar{X}_{..} = 6.0000 - 9.34375 = -3.34375$$

$$\hat{\beta}_3 = \bar{B}_3 - \bar{X}_{..} = 11.0000 - 9.34375 = 1.65625$$

$$\hat{\alpha\beta} = \overline{AB}_{11} - \bar{A}_1 - \bar{B}_1 + \bar{X}_{..} = 7.2500 - 8.9375 - 8.0000 + 9.34375 = -.34375$$

$$\hat{\alpha\beta}_{12} = \overline{AB}_{12} - \bar{A}_1 - \bar{B}_2 + \bar{X}_{..} = 4.2500 - 8.9375 - 6.0000 + 9.34375 = -1.34375$$

$$\hat{\alpha\beta}_{13} = \overline{AB}_{13} - \bar{A}_1 - \bar{B}_3 + \bar{X}_{..} = 11.2500 - 8.9375 - 11.0000 + 9.34375 = 0.65625$$

table, it is apparent that the main effect of B is significant but that the effects of A and AB are not.

To analyze these data from the point of view of multiple regression, we begin with the following design matrix. Once again, the elements of each row apply to all subjects in the corresponding treatment combination.

$$\mathbf{X} = \begin{matrix} & A_1 & B_1 & B_2 & B_3 & AB_{11} & AB_{12} & AB_{13} \\ \begin{matrix} a_1b_1 \\ a_1b_2 \\ a_1b_3 \\ a_1b_4 \\ a_2b_1 \\ a_2b_2 \\ a_2b_3 \\ a_2b_4 \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix} \end{matrix}$$

The first step in a multiple-regression analysis is presented in Exhibit 16.2 using all seven predictors (A_1 to AB_{13}). The results were obtained using SAS PROC CORR and PROC REG, although every software package should give the same answers. A program for doing this in R is presented on the Web site.

Exhibit 16.2 has several important features. First, consider the matrix of correlations among variables, often symbolized as \mathbf{R} . Suppose that we simplify this matrix by defining the following sets of predictors: $A' = [A_1]$, $B' = [B_1, B_2, B_3]$, and $AB' = [AB_{11}, AB_{12}, AB_{13}]$. If we then rewrite the intercorrelation matrix, we have

$$\begin{matrix} & A' & B' & AB' \\ \begin{matrix} A' \\ B' \\ AB' \end{matrix} & \begin{bmatrix} 1.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix} \end{matrix}$$

Notice that each of the effects is independent of the others (the intercorrelations are 0.00). Such a pattern occurs only if there are equal (or proportional) numbers of subjects in each cell; this pattern is also what makes simplified formulae for the analysis of variance possible. The fact that this structure disappears in the case of unequal n s is what makes our life more difficult when we have missing subjects.

Next notice the vector \mathbf{b} , labeled as the Parameter Estimate. The first entry (b_0) is labeled Intercept and is the grand mean of all of the observations. The subsequent entries ($b_1 \dots b_7$) are the estimates of the corresponding treatment effects. Thus $b_1 = \alpha_1$, $b_2 = \beta_1$, $b_5 = \alpha\beta_{11}$, and so on. Tests on these regression coefficients represent tests on the corresponding treatment effects. The fact that we have only the $(a - 1)(b - 1) = 3$ interaction effects presents no problem, due to the restrictions that these effects must sum to 0 across rows and down columns. Thus if $\alpha\beta_{12} = -1.34$, then $\alpha\beta_{22}$ must be $+1.34$. Similarly, $\alpha\beta_{14} = 0 - \sum \alpha\beta_{1j} = -\sum \alpha\beta_{1j} = 1.03$.

The value of $R^2 = .625$ represents the percentage of variation that can be accounted for by all the variables simultaneously. With equal n s, and therefore independent effects, it is equivalent to $\eta_A^2 + \eta_B^2 + \eta_{AB}^2 = .014 + .537 + .074 = .625$. The test on R^2 produces an F of 5.711 on 7 and 24 df , which, because it is significant ($p = .0006$), shows that there is a nonchance relationship between the treatment variables, considered together, and the dependent variable (Y).

Two more parallels can be drawn between Table 16.2, the analysis of variance, and Exhibit 16.2, the regression analysis. First, notice that $SS_{\text{regression}} = SS_{\text{Model}} = SS_Y(1 - R^2) = 231.969$. This is the variation that can be predicted by a linear combination of the predictors. This value is equal to $SS_A + SS_B + SS_{AB}$, although from Exhibit 16.2 we cannot yet partition the variation among the separate sources. Finally, notice that $SS_{\text{residual}} = SS_{\text{error}} = SS_Y(1 - R^2) = 139.250$, which is the error sum of squares in the analysis of variance. This makes sense when you recall that error is the variation that cannot be attributed to the separate or joint effects of the treatment variables.

```

Data Anova;
  infile 'Ex162.dat';
  input A1 B1 B2 B3 dv;
  AB11 = A1*B1;
  AB12 = A1*B2;
  AB13 = A1*B3;

Run;

Proc Corr Data = Anova;
  Var A1 B1 B2 B3 AB11 AB12 AB13;

Run;

Proc Reg Data = Anova;
  Model dv = A1 B1 B2 B3 AB11 AB12 AB13;

Run;

```

Pearson Correlation Coefficients, N = 32

Prob > |r| under H0: Rho=0

	A1	B1	B2	B3	AB11	AB12	AB13
A1	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
B1	0.00000	1.00000	0.50000	0.50000	0.00000	0.00000	0.00000
	1.0000		0.0036	0.0036	1.0000	1.0000	1.0000
B2	0.00000	0.50000	1.00000	0.50000	0.00000	0.00000	0.00000
	1.0000	0.0036		0.0036	1.0000	1.0000	1.0000
B3	0.00000	0.50000	0.50000	1.00000	0.00000	0.00000	0.00000
	1.0000	0.0036	0.0036		1.0000	1.0000	1.0000
AB11	0.00000	0.00000	0.00000	0.00000	1.00000	0.50000	0.50000
	1.0000	1.0000	1.0000	1.0000		0.0036	0.0036
AB12	0.00000	0.00000	0.00000	0.00000	0.50000	1.00000	0.50000
	1.0000	1.0000	1.0000	1.0000	0.0036		0.0036
AB13	0.00000	0.00000	0.00000	0.00000	0.50000	0.50000	1.00000
	1.0000	1.0000	1.0000	1.0000	0.0036	0.0036	

The REG Procedure
Dependent Variable: dv

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	231.96875	33.13839	5.71	0.0006
Error	24	139.25000	5.80208		
Corrected Total	31	371.21875			

Root MSE	2.40875	R-Square	0.6249
Dependent Mean	9.34375	Adj R-Sq	0.5155
Coeff Var	25.77928		

(continued)

Adapted from output by SAS Institute Inc.

Exhibit 16.2 Regression solutions using all predictors for data in Table 16.2

Variable	DF	Parameter Estimates		t Value	Pr > t
		Parameter Estimate	Standard Error		
Intercep	1	9.34375	0.42581	21.94	<.0001
A1	1	-0.40625	0.42581	-0.95	0.3496
B1	1	-1.34375	0.73753	-1.82	0.0809
B2	1	-3.34375	0.73753	-4.53	0.0001
B3	1	1.65625	0.73753	2.25	0.0342
AB11	1	-0.34375	0.73753	-0.47	0.6454
AB12	1	-1.34375	0.73753	-1.82	0.0809
AB13	1	0.65625	0.73753	0.89	0.3824

Exhibit 16.2 (continued)

Reduced Models

At this point we know only the amount of variation that can be accounted for by all of the predictors simultaneously. What we wish to know is how this variation can be partitioned among A , B , and AB . This information can be readily obtained by computing several reduced regression equations.

Because in the subsequent course of the analysis we must compute several multiple regression sums of squares relating to the different effects, we will change our notation and use the effect labels (α , β , and $\alpha\beta$) as subscripts. For the multiple regression just computed, the model contained variables to account for α , β , and $\alpha\beta$. Thus, we will designate the sum of squares regression in that solution as $SS_{\text{regression}_{\alpha,\beta,\alpha\beta}}$. If we dropped the last three predictors (AB_{11} , AB_{12} , and AB_{13}) we would be deleting those predictors carrying information concerning the interaction but would retain those predictors concerned with α and β . Thus, we would use the designation $SS_{\text{regression}_{\alpha,\beta}}$. If we used only A , AB_{11} , AB_{12} , and AB_{13} as predictors, the model would account for only α and $\alpha\beta$, and the result would be denoted $SS_{\text{regression}_{\alpha,\alpha\beta}}$.

I have run the individual regression solutions for our example, and the results are

$$SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} = 231.969$$

$$SS_{\text{regression}_{\alpha,\beta}} = 204.625$$

$$SS_{\text{regression}_{\beta,\alpha\beta}} = 226.687$$

$$SS_{\text{regression}_{\alpha,\alpha\beta}} = 32.635$$

Now this is the important part. If the interaction term accounts for any of the variation in Y , then removing the interaction predictors from the model should lead to a decrease in accountable variation. This decrease will be equal to the variation that can be attributable to the interaction. By this and similar reasoning,

$$SS_{AB} = SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\beta}}$$

$$SS_A = SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\beta,\alpha\beta}}$$

$$SS_B = SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\alpha\beta}}$$

Table 16.3 Regression solution for the data in Table 16.2

$SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} = 231.969$	$R^2 = .625$																														
$SS_{\text{residual}_{\alpha,\beta,\alpha\beta}} = 139.250$																															
$SS_{\text{regression}_{\alpha,\beta}} = 204.6245$	$R^2 = .551$																														
$SS_{\text{regression}_{\beta,\alpha\beta}} = 226.687$	$R^2 = .611$																														
$SS_{\text{regression}_{\alpha,\alpha\beta}} = 32.625$	$R^2 = .088$																														
$SS_{AB} = SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\beta}} = 231.969 - 204.625 = 27.344$ $SS_A = SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\beta,\alpha\beta}} = 231.969 - 226.687 = 5.282$ $SS_B = SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\alpha\beta}} = 231.969 - 32.625 = 199.344$ $SS_{\text{error}} = SS_{\text{residual}_{\alpha,\beta,\alpha\beta}} = 139.250$																															
<p>Summary Table</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Source</th> <th style="text-align: center;"><i>df</i></th> <th style="text-align: center;"><i>SS</i></th> <th style="text-align: center;"><i>MS</i></th> <th style="text-align: center;"><i>F</i></th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black;">A</td> <td style="text-align: center; border-top: 1px solid black;">1</td> <td style="text-align: center; border-top: 1px solid black;">5.282</td> <td style="text-align: center; border-top: 1px solid black;">5.282</td> <td style="text-align: center; border-top: 1px solid black;"><1</td> </tr> <tr> <td>B</td> <td style="text-align: center;">3</td> <td style="text-align: center;">199.344</td> <td style="text-align: center;">66.448</td> <td style="text-align: center;">11.452*</td> </tr> <tr> <td>AB</td> <td style="text-align: center;">3</td> <td style="text-align: center;">27.344</td> <td style="text-align: center;">9.115</td> <td style="text-align: center;">1.571</td> </tr> <tr> <td style="border-top: 1px solid black;">Error</td> <td style="text-align: center; border-top: 1px solid black;">24</td> <td style="text-align: center; border-top: 1px solid black;">139.250</td> <td style="text-align: center; border-top: 1px solid black;">5.802</td> <td></td> </tr> <tr> <td></td> <td style="text-align: center;">31</td> <td style="text-align: center;">371.220</td> <td></td> <td></td> </tr> </tbody> </table>		Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	A	1	5.282	5.282	<1	B	3	199.344	66.448	11.452*	AB	3	27.344	9.115	1.571	Error	24	139.250	5.802			31	371.220		
Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>																											
A	1	5.282	5.282	<1																											
B	3	199.344	66.448	11.452*																											
AB	3	27.344	9.115	1.571																											
Error	24	139.250	5.802																												
	31	371.220																													

* $p < .05$

© Cengage Learning 2013

The relevant calculations are presented in Table 16.3. (I leave it to you to verify that these are the sums of squares for regression that result when we use the relevant predictors.)

Looking first at the *AB* interactions, we see from Exhibit 16.2 that when the interaction terms were deleted from the model, the sum of squares that could be accounted for by the model decreased by

$$SS_{AB} = SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\beta}} = 231.969 - 204.625 = 27.344$$

This decrement can only be attributable to the predictive value of the interaction terms, and therefore

$$SS_{AB} = 27.344$$

By a similar line of reasoning, we can find the other sums of squares.²

Notice that these values agree exactly with those obtained by the more traditional procedures. Notice also that the corresponding decrements in R^2 agree with the computed values of η^2 .

As Overall and Spiegel (1969) pointed out, the approach we have taken in testing the effects of *A*, *B*, and *AB* is not the only one we could have chosen. They presented two

² A number of authors (e.g. Judd & McClelland) prefer to use the increase in the error term (rather than the decrease in $SS_{\text{regression}}$) when an effect is deleted. The result will be the same.

alternative models that might have been considered in place of this one. Fortunately, however, the different models all lead to the same conclusions in the case of equal sample sizes, since in this situation effects are independent of one another and therefore are additive. When we consider the case of unequal sample sizes, however, the choice of an underlying model will require careful consideration.

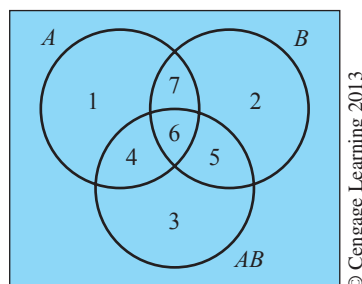
16.4 Analysis of Variance with Unequal Sample Sizes

The least-squares approach to the analysis of variance is particularly useful for the case of factorial experiments with unequal sample sizes. However, special care must be used in selecting the particular restricted models that are employed in generating the various sums of squares.

Several different models could underlie an analysis of variance. Although in the case of equal sample sizes these models all lead to the same results, in the unequal n case they do not. This is because with unequal n s, the row, column, and interaction effects are no longer orthogonal and thus account for overlapping portions of the variance. [I would strongly recommend quickly reviewing the example given in Chapter 13, Section 13.11 (pp. 444–446).] Consider the Venn diagram in Figure 16.1. The gray area enclosed by the surrounding square will be taken to represent SS_{total} . Each circle represents the variation attributable to (or accounted for by) one of the effects. The area outside the circles but within the square represents SS_{error} . Finally, the total area enclosed by the circles represents $SS_{\text{regression}_{\alpha,\beta,\alpha\beta}}$, which is the sum of squares for regression when all the terms are included in the model. If we had equal sample sizes, none of the circles would overlap, and each effect would be accounting for a separate, independent, portion of the variation. In that case, the decrease in $SS_{\text{regression}}$ resulting from deleting of an effect from the model would have a clear interpretation—it would be the area enclosed by the omitted circle and thus would be the sum of squares for the corresponding effect.

But what do we do when the circles overlap? If we were to take a model that included terms for A , B , and AB and compared it to a model containing only A and B terms, the decrement would not represent the area of the AB circle, because some of that area still would be accounted for by A and/or B . Thus, SS_{AB} , which we calculate as $SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\beta}}$, represents only the portion of the enclosed area that is *unique* to AB —the area labeled with a “3.” So far, all the models that have been seriously proposed are in agreement. SS_{AB} is that portion of the AB circle remaining after adjusting for A and B .

But now things begin to get a little sticky. Different and meaningful approaches have been put forth that differ in the way the remainder of the pie is allotted to



© Cengage Learning 2013

Figure 16.1 Venn diagram representing portions of overall variation

A and B . Overall and Spiegel (1969) put forth three models for the analysis of variance, and these models continue to generate a voluminous literature debating their proper use and interpretation, even though the discussion began 30 years ago. We will refer to these models as Type I, Type II, and Type III, from the terminology used by SPSS and SAS. (Overall and Spiegel numbered them in the reverse order, just to make things more confusing.) Basically, the choice between the three models hinges on how we see the relationship between the sample size and the treatments themselves, or, more specifically, how we want to weight the various cell means to produce row and column means. Before exploring that issue, however, we must first examine the competing methods.

Method III

Type III Sum of squares

Method III, (or **Type III Sum of squares**) is the method we used in the preceding section. In this case, each effect is adjusted for all other effects. Thus we obtain SS_{AB} as $SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\beta}}$, SS_A as $SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\beta,\alpha\beta}}$, and SS_B as $SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\alpha\beta}}$. In terms of Figure 16.1, each effect is defined as the part of the area that is unique to that effect. Thus, SS_A is represented by area “1,” SS_B by area “2,” and SS_{AB} by area “3.”

Method II

Type II SS

Method II (or **Type II SS**) breaks up the pie differently. We continue to define SS_{AB} as area “3.” But now that we have taken care of the interaction, we still have areas “1,” “2,” “4,” “5,” “6,” and “7,” which can be accounted for by the effects of A and/or B . Method II essentially redefines the full model as $SS_{\text{regression}_{\alpha,\beta}}$ and obtains $SS_A = SS_{\text{regression}_{\alpha,\beta}} - SS_{\text{regression}_{\beta}}$, and SS_B as $SS_{\text{regression}_{\alpha,\beta}} - SS_{\text{regression}_{\alpha}}$. Thus, A is allotted areas “1” and “4,” whereas B is allotted areas “2” and “5.” Methods II and III are summarized in Table 16.4.

Table 16.4 Alternative models for solution of nonorthogonal designs

Method III

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

Source	df	SS	Portion of Diagram
A	$a - 1$	$SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\beta,\alpha\beta}}$	1
B	$b - 1$	$SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\alpha\beta}}$	2
AB	$(a - 1)(b - 1)$	$SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\beta}}$	3
Error	$N - ab$	$SS_{\text{residual}_{\alpha,\beta,\alpha\beta}}$	
Total	$N - 1$	SS_Y	

Method II

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ijk}$$

and

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk}$$

Source	df	SS	Portion of Diagram
A	$a - 1$	$SS_{\text{regression}_{\alpha,\beta}} - SS_{\text{regression}_{\beta}}$	1 + 4
B	$b - 1$	$SS_{\text{regression}_{\alpha,\beta}} - SS_{\text{regression}_{\alpha}}$	2 + 5
AB	$(a - 1)(b - 1)$	$SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} - SS_{\text{regression}_{\alpha,\beta}}$	3
Error	$N - ab$	$SS_{\text{residual}_{\alpha,\beta,\alpha\beta}}$	
Total	$N - 1$	SS_Y	

Both of these methods make a certain amount of sense when looked at from the point of view of the Venn diagram in Figure 16.1. However, the diagram is only a crude approximation and we have pushed it about as far as we can go.³

As Carlson and Timm (1974) argued, a more appropriate way to compare the models is to examine the hypotheses they test. These authors point out that Method III represents an estimation of treatment effects when cell means are weighted equally and is particularly appropriate whenever we consider sample size to be independent of treatment conditions. A convincing demonstration of this is presented in Overall, Spiegel, and Cohen (1975).⁴ Carlson and Timm also showed that Method II produces estimates of treatment effects when row and column means are weighted by the sample size, but only when no interaction is present. When an interaction is present, simple estimates of row and column effects cannot be made, and, in fact, the null hypotheses actually tested are very bizarre indeed [see Carlson and Timm (1974) for a statement of the null hypotheses for Method II]. SPSS, which once relied on a method similar to Method II, finally saw the light some years ago and came around to using Method III as the default. They labeled this method “Unique SS” because each effect is assigned only that portion of the variation that it uniquely explains. SAS has always tested Type III sums of squares as the default. An excellent discussion of the hypotheses tested by different approaches is presented in Blair and Higgins (1978) and Blair (1978). Hector, von Felten, and Schmid (2010) provide an excellent discussion of the alternative analyses. As Cochran and Cox suggested, “the only complete solution of the ‘missing data’ problem is not to have them” (p. 82).

There is a third method of computing sums of squares that at first seems particularly bizarre. Just to make matters even more confusing than they need to be, this is the method that SPSS and SAS refer to as “**Type I SS**,” or **Method I**, but which I will refer to as **hierarchical sums of squares**, though it is sometimes referred to as **sequential sums of squares**, which is the term that SPSS uses. The peculiar thing about this approach is that it is dependent on the order in which you name your variables. Thus if you tell SAS or SPSS to model (predict or account for) the dependent variable on the basis of A , B , and AB , the program will first assign $SS_A = SS_{\text{regression}_A}$. Then $SS_B = SS_{\text{regression}_{A,B}} - SS_{\text{regression}_A}$, and finally $SS_{AB} = SS_{\text{regression}_{A,B,AB}} - SS_{\text{regression}_{A,B}}$. In this situation the first effect is assigned all of the sums of squares it can possibly account for. The next effect is assigned all that it can account for *over and above* what was accounted for by the first one. Finally, the interaction effect is assigned only what it accounts for over and above the two main effects. But, if you ask the software to model the dependent variable on the basis of B , A , and AB , then SS_B will equal $SS_{\text{regression}_B}$, which is quite a different thing from $SS_{\text{regression}_{A,B}} - SS_{\text{regression}_A}$. The only time I could recommend using this approach is if you have a strong reason to want to control the variables in a particular order.⁵ If you can defend the argument that Variable A is so important that it should be looked at first without controlling for any other variables, then perhaps this is a method you can use. But I have never seen a case where I would want to do that, with the possible exception of dealing

Type I SS
Method I
hierarchical sums of squares
sequential sums of squares

³ From this discussion you could easily get the impression that Method II will always account for more of the variation than Method III. This is not necessarily the case, because the degree of overlap represents the correlation between effects, and suppressor relationships might appear as “black holes,” canceling out accountable variation.

⁴ If you go back to this older literature, it is important to note that when those papers were written, what we call Method III was then called Method I and vice versa.

⁵ There is a good and honorable tradition of prioritizing variables in this way for theoretical studies using standard multiple regression with continuous variables. I have never seen a similar application in an analysis of variance framework, though I have seen a number of people talk about hypothetical examples.

with a variable as a covariate, which we will discuss shortly. The only reason that I bring the issue up here at all is to explain some of the choices you will have to make in using computer software.

Many people writing in the R community have very strong and vocal opinions about the model to be tested, and they hold strongly to the Model I solution. That is the solution provided by the “anova” command in the base version of R. Fox (2002) supports the Type II solution and his library of functions in R (called “car”) can produce either Type II or Type III solutions assuming that you specify the proper contrasts. (Fox’s main function is called Anova, with a capital A, whereas the base system uses a lower case “a” and solves for Type I sums of squares.) I prefer the Type III approach because most cases of unequal sample size in the behavioral have no reason to wish to weight some cells more than others. However, keep in mind that if there is a significant interaction you will probably not care to look at the main effects (simple effects would be more revealing), and in that case Methods II and III will bring you to the same place. However, if there is a noticeable, but nonsignificant interaction, the two models will lead you to different results. The battle over Type II and Type III sums of squares has gone on for a very long time, and there is no resolution in sight. I prefer Type III sums of squares but would not have a fit if I sat on a dissertation committee where the candidate used Type II sums of squares. Howell and McConaughy (1982) argued that there are very few instances in which one would want to test the peculiar null hypotheses tested by Method II when an interaction is present, but it is not worth falling on a sword in defense of either position. However, the conclusion to be drawn from the literature at present is that for the most common situations Method III is appropriate, because we usually want to test unweighted means. (This is the default method employed by SPSS. Method III sum of squares are the values labeled as Type III SS in SAS, and now by more recent versions of SPSS.) It is also the method that is approximated by the *unweighted means solution* discussed in Chapter 13. (You may recall that in Chapter 13 we saw that the traditional label “unweighted means solution” really should be the “**equally weighted means** solution,” if that name hadn’t been appropriated in the past for a different procedure, because, using it, we are treating all means equally, regardless of the sample sizes.) Method III essentially assumes that observations are missing completely at random, so there is no reason that a cell with more observations should carry any more weight than one with fewer observations. If this is not the case you should consider a different method.

equally weighted means

As an illustration of Method III, we will take the data used in the previous example but add six additional observations to produce unequal cell sizes. The data are given in Table 16.5, with the unweighted and weighted row and column means and the values resulting from the various regression solutions. The unweighted means are the mean of means (therefore, the mean of row₁ is the mean of the four cell means in that row). The weighted mean of row₁, for example, is just the sum of the scores in row₁ divided by the number of scores in row₁.

From Table 16.5 we see that $R^2_{\alpha,\beta,\alpha\beta} = .541$, indicating that approximately 54% of the variation can be accounted for by a linear combination of the predictor variables. We do not know, however, how this variation is to be distributed among *A*, *B*, and *AB*. For that we need to form and calculate the reduced models.

Testing the Interaction Effects

First, we delete the predictors associated with the interaction term and calculate $R^2_{\alpha,\beta}$. For these data, $R^2_{\alpha,\beta} = .483$, representing a drop in R^2 of about .05. If we examine the predictable

Table 16.5 Illustrative calculations for nonorthogonal factorial design

	B_1	B_2	B_3	B_4	Unweighted Mean	Weighted Mean
A_1	5	2	8	11	8.975	8.944
	7	5	11	15		
	9	7	12	16		
	8	3	14	10		
A_2		9		9	9.590	9.950
	7	3	9	11		
	9	8	12	14		
	10	9	14	10		
	9	11	8	12		
			7	13		
			11			
			12			
Unweighted Mean	8.000	6.475	10.625	12.029	9.282	
Weighted Mean	8.000	6.333	10.556	12.00		9.474
Full Model						
$R^2_{\alpha,\beta,\alpha\beta} = .541$						
$SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} = 216.017$						
$SS_{\text{residual}} = 183.457$						
Reduced Models						
$R^2_{\alpha,\beta} = .490$						
$SS_{\text{regression}_{\alpha,\beta}} = 195.805$						
$R^2_{\beta,\alpha\beta} = .532$						
$SS_{\text{regression}_{\beta,\alpha\beta}} = 212.552$						
$R_{\alpha,\alpha\beta} = 0.77$						
$SS_{\text{regression}_{\alpha,\alpha\beta}} = 30.845$						

sum of squares ($SS_{\text{regression}}$), we see that eliminating the interaction terms has produced a decrement in $SS_{\text{regression}}$ of

$$\begin{aligned} SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} &= 216.017 \\ SS_{\text{regression}_{\alpha,\beta}} &= \underline{195.805} \\ SS_{AB} &= 20.212 \end{aligned}$$

This decrement is the sum of squares attributable to the AB interaction (SS_{AB}).

In the case of unequal n s, it is particularly important to understand what this term represents. You should recall that $SS_{\text{regression}_{\alpha,\beta,\alpha\beta}}$, for example, equals $SS_Y(R^2_{\alpha,\beta,\alpha\beta})$. Then

$$\begin{aligned} SS_{AB} &= SS_Y(R^2_{\alpha,\beta,\alpha\beta}) - SS_Y(R^2_{\alpha,\beta}) \\ &= SS_Y(R^2_{\alpha,\beta,\alpha\beta} - R^2_{\alpha,\beta}) \\ &= SS_Y(R^2_{0(\alpha\beta,\alpha,\beta)}) \end{aligned}$$

The final term in parentheses is the squared semipartial correlation between the criterion and the interaction effects, partialling out (adjusting for) the effects of A and B . In other words, it is the squared correlation between the criterion and the part of the AB interaction that is orthogonal to A and B . Thus, we can think of SS_{AB} as really being $SS_{AB(adj)}$, where the adjustment is for the effects of A and B . (In the equal- n case, the issue does not arise because A , B , and AB are independent, and therefore there is no overlapping variation to partial out.)⁶

Testing the Main Effects

Because we are calculating Method III SS, we will calculate the main effects of A and B in a way that is directly comparable to our estimation of the interaction effect. Here, each main effect represents the sum of squares attributable to that variable after partialling out the other main effect and the interaction.

To obtain SS_A , we will delete the predictor associated with the main effect of A and calculate $SS_{\text{regression}_{\beta,\alpha\beta}}$. For these data, $R^2_{\beta,\alpha\beta} = .523$, producing a drop in R^2 of $.532 - .523 = .009$. In terms of the predictable sum of squares ($SS_{\text{regression}}$), the elimination of α from the model produces a decrement in $SS_{\text{regression}}$ of

$$\begin{aligned} SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} &= 216.017 \\ SS_{\text{regression}_{\beta,\alpha\beta}} &= \underline{212.552} \\ SS_{AB} &= 3.465 \end{aligned}$$

This decrement is the sum of squares attributable to the main effect of A .

By the same reasoning, we can obtain SS_B by comparing $SS_{\text{regression}}$ for the full model and for a model omitting β .

$$\begin{aligned} SS_{\text{regression}_{\alpha,\beta,\alpha\beta}} &= 216.017 \\ SS_{\text{regression}_{\alpha,\alpha\beta}} &= \underline{30.845} \\ SS_B &= 185.172 \end{aligned}$$

These results are summarized in Table 16.6, with the method by which they were obtained. Notice that the sums of squares do not sum to SS_{total} . This is as it should be, because the overlapping portions of accountable variation (segments “4,” “5,” “6,” and “7” of Figure 16.1) are not represented anywhere. Also notice that SS_{error} is taken as the SS_{residual} from the full model, just as in the case of equal sample sizes. Here again we define SS_{error} as the portion of the total variation that cannot be explained by any one or more of the independent variables.

⁶ Some people have trouble understanding the concept of nonindependent treatment effects. As an aid, perhaps an extreme example will help point out how a row effect could cause an *apparent* column effect, or vice versa. Consider the following two-way table. When we look at differences among means, are we looking at a difference due to A , B , or AB ? There is no way to tell.

	B_1	B_2	Means
A_1	$\bar{X} = 10$ $n = 20$	$n = 0$	10
A_2	$n = 0$	$\bar{X} = 30$ $n = 20$	30
Means	10	30	

Table 16.6 Calculation of sums of squares using Method III—the unweighted means solution

Method III (Unweighted Means)		
Source	<i>df</i>	<i>SS</i>
<i>A</i>	$a - 1$	$SS_Y(R_{\alpha,\beta,\alpha\beta}^2 - R_{\beta,\alpha\beta}^2)$
<i>B</i>	$b - 1$	$SS_Y(R_{\alpha,\beta,\alpha\beta}^2 - R_{\alpha,\alpha\beta}^2)$
<i>AB</i>	$(a - 1)(b - 1)$	$SS_Y(R_{\alpha,\beta,\alpha\beta}^2 - R_{\alpha,\beta}^2)$
Error	$N - ab$	$SS_Y(1 - R_{\alpha,\beta,\alpha\beta}^2)$
Total	$N - 1$	SS_Y

Summary Table for Analysis of Variance

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
<i>A</i>	1	3.465	3.465	<1
<i>B</i>	3	185.172	61.724	10.094
<i>AB</i>	3	20.212	6.737	<1
Error	30	183.457	6.115	
Total	37	(399.474)		

© Cengage Learning 2013

As I mentioned earlier, the unweighted-means solution presented in Chapter 13 is an approximation of the solution (Method III) given here. The main reason for discussing that solution in this chapter is so that you will understand what the computer program is giving you and how it is treating the unequal sample sizes.

A computer-based solution using SPSS is shown in Exhibit 16.3. It illustrates that the Type III sums of squares from SPSS do, in fact, produce the appropriate analysis of the data in Table 16.5.

Tests of Between-Subjects Effects

Dependent Variable: DV

Source	Type III Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	Sig.
Corrected Model	216.017 ^a	7	30.860	5.046	.001
Intercept	3163.341	1	3163.841	517.370	.000
Rows	3.464	1	3.464	.566	.458
Columns	185.172	3	61.724	10.093	.000
Rows* Columns	20.212	3	6.737	1.102	.364
Error	183.457	30	6.115		
Total	3810.000	38			
Corrected Total	399.474	37			

^aR Squared = .541 (Adjusted R Squared = .434)**Exhibit 16.3** SPSS analysis of the data in Table 16.7

Adapted from output by SPSS, Inc.

16.5 The One-Way Analysis of Covariance

analysis of covariance

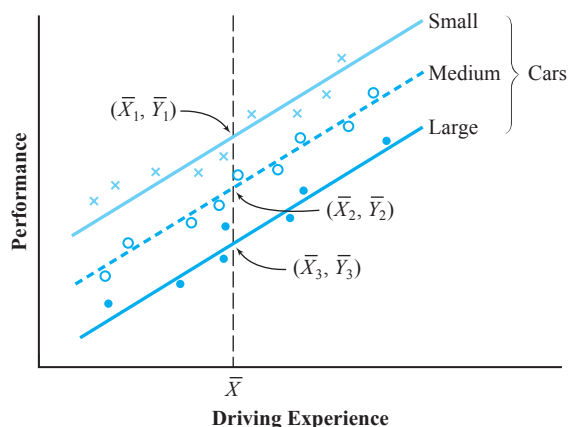
An extremely useful tool for analyzing experimental data is the **analysis of covariance**. As presented within the context of the analysis of variance, the analysis of covariance appears to be unpleasantly cumbersome, especially so when there is more than one covariate. Within the framework of multiple regression, however, it is remarkably simple, requiring little, if any, more work than does the analysis of variance. I will present the approach using coded dummy variables and multiple regression. However, once you understand the basic logic, you will understand the results from using SPSS or SAS without having to set up the design matrix yourself.

Suppose we wish to compare driving proficiency on three different sizes of cars to test the experimental hypothesis that small cars are easier to handle. We have available three different groups of drivers, but we are not able to match individual subjects on driving experience, which varies considerably within each group. Let us make the simplifying assumption, which will be discussed in more detail later, that the mean level of driving experience is equal across groups. Suppose further that using the number of steering errors as our dependent variable, we obtain the somewhat exaggerated data plotted in Figure 16.2. In this figure the data have been plotted separately for each group (size of car), as a function of driving experience (the **covariate**), and the separate regression lines have been superimposed.

covariate

One of the most striking things about Figure 16.2 is the large variability in both performance and experience within each treatment. This variability is so great that an analysis of variance on performance scores would almost certainly fail to produce a significant effect. Most of the variability in performance, however, is directly attributable to differences in driving experience, which has nothing to do with what we wish to study. If we could somehow remove (partial out) the variance that can be attributed to experience (the covariate), we would have a clearer test of our original hypothesis. This is exactly what the analysis of covariance is designed to do, and this is precisely the situation in which it does its job best—its job in this case being to reduce the error term.

A more controversial use of the analysis of covariance concerns situations in which the treatment groups have different covariate (driving experience) means. Such a situation (using the same hypothetical experiment) is depicted in Figure 16.3, in which two of the treatments



© Cengage Learning 2013

Figure 16.2 Hypothetical data illustrating error-reduction in the analysis of covariance

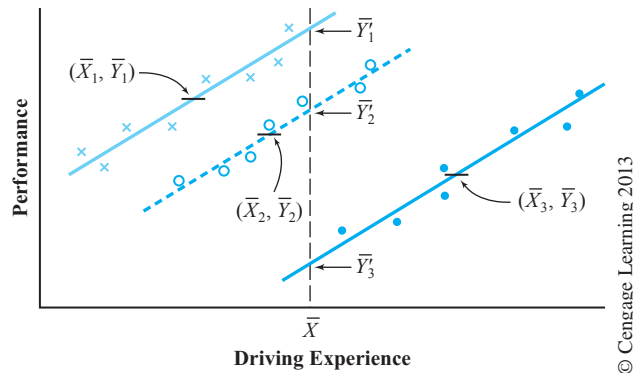


Figure 16.3 Hypothetical data illustrating mean adjustment in the analysis of covariance

© Cengage Learning 2013

adjusted Y means

have been displaced along the X axis. At the point at which the three regression lines intersect the vertical line $X = \bar{X}$, you can see the values \bar{Y}'_1 , \bar{Y}'_2 , and \bar{Y}'_3 . These are the **adjusted Y means** and represent our best guess as to what the Y values would have been *if the treatments had not differed on the covariate*. The analysis of covariance then tests whether these *adjusted* means differ significantly, again using an error term from which the variance attributable to the covariate has been partialled out. Notice that the adjusted performance means are quite different from the unadjusted means. The adjustment has increased \bar{Y}'_1 and decreased \bar{Y}'_3 .

Although the structure and procedures of the analysis of covariance are the same regardless of whether the treatment groups differ on the covariate means, the different ways of visualizing the problem as represented in Figures 16.2 and 16.3 are instructive. In the first case, we are simply reducing the error term. In the second case, we are both reducing the error term *and* adjusting the means on the dependent variable. We will have more to say about this distinction later in the chapter.

Assumptions of the Analysis of Covariance

homogeneity of regression

Aside from the usual analysis of variance assumptions of normality and homogeneity of variance, we must add two more assumptions. First, we will assume that whatever the relationship between Y and the covariate (C), this relationship is linear.⁷ Second, we will assume **homogeneity of regression**—that the regression coefficients are equal across treatments— $b_1^* = b_2^* = b_3^* = \dots = b^*$. This is merely the assumption that the three lines in Figures 16.2 or 16.3 are parallel, and it is necessary to justify our substitution of one regression line (the pooled within-groups regression line) for the separate regression lines. As we shall see shortly, this assumption is testable. Note that no assumption has been made about the nature of the covariate; it may be either a fixed or a random variable. (It can even be a categorical variable if we create dummy variables to represent the different levels of the variable, as we did in the earlier parts of this chapter.)

Calculating the Analysis of Covariance

When viewed within the framework of multiple regression, the analysis of covariance is basically no different from the analysis of variance, except that we wish to partial out the

⁷ Methods for handling nonlinear relationships are available but will not be discussed here.

effects of the covariate. As Cohen (1968) put it, “A covariate is, after all, nothing but an independent variable which, because of the logic dictated by the substantive issues of the research, assumes priority among the set of independent variables as a basis for accounting for Y variance.” (p. 439).

If we want to ask about the variation in Y after the covariate (C) has been partialled out, and if the variation in Y can be associated with only C , the treatment effect (α), and error, then $SS_{\text{regression}_{C,\alpha}}$ represents the total amount of accountable variation. If we now compare $SS_{\text{regression}_{C,\alpha}}$ with $SS_{\text{regression}_C}$, the difference will be the variation attributable to treatment effects *over and above* that attributable to the covariate.

We will take as an example a variation on the study by Conti and Musty (1984) presented in Chapter 11. As you may recall, in that study the authors were interested in examining the effects of different amounts of THC, the major active ingredient in marijuana, injected directly into the brains of rats. The dependent variable was locomotor activity, which normally increases with the administration of THC by more traditional routes. Because of the nature of the experimental setting (all animals were observed under baseline conditions and then again after the administration of THC), activity should decrease in all animals as they become familiar and more comfortable with the apparatus. If THC has its effect through the nucleus accumbens, however, the effects of moderate doses of THC should partially compensate for this anticipated decrease, leading to relatively greater activity levels in the moderate-dose groups as compared to the low- or high-dose groups.

Conti and Musty (1984) actually analyzed postinjection activity as a percentage of preinjection activity, because that is the way such data are routinely analyzed in their field. An alternative procedure would have been to run an analysis of covariance on the postinjection scores, partialling out preinjection differences. Such a procedure would adjust for the fact that much of the variability in postinjection activity could be accounted for by variability in preinjection activity. It would also control for the fact that, by chance, there were group differences in the level of preinjection activity that could contaminate postinjection scores.

As will become clear later, it is important to note here that all animals were assigned at random to groups. Therefore, we would *expect* the group means on the preinjection phase to be equal. Any differences that do appear on the preinjection phase, then, are due to chance, and, *in the absence of any treatment effect*, we would expect that postinjection means, adjusted for chance preinjection differences, would be equal. The fact that subjects were assigned at random to treatments is what allows us to expect equal adjusted group means at postinjection (if H_0 is true), and this in turn allows us to interpret group differences at postinjection to be a result of real treatment differences rather than of some artifact of subject assignment.

The data and the design matrix for the Conti and Musty (1984) study are presented in Table 16.7. The raw data have been divided by 100 simply to make the resulting sums of squares manageable.⁸ In the design matrix that follows the data, only the first and last subject in each group are represented. Columns 6 through 9 of X represent the interaction of the covariate and the group variables. These columns are used to test the hypothesis of homogeneity of regression coefficients across groups:

$$H_0: b_1^* = b_2^* = b_3^* = b_4^* = b_5^*$$

⁸ If the data had not been divided by 100, the resulting sums of squares and mean squares would be $100^2 = 10,000$ times their present size. The F and t values would be unaffected.

Table 16.7 Pre- and postinjection data from Conti and Musty (1984)

	Control		0.1 μg		0.5 μg		1 μg		2 μg	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
	4.34	1.30	1.55	0.93	7.18	5.10	6.94	2.29	4.00	1.44
	3.50	0.94	10.56	4.44	8.33	4.16	6.10	4.75	4.10	1.11
	4.33	2.25	8.39	4.03	4.05	1.54	4.90	3.48	3.62	2.17
	2.76	1.05	3.70	1.92	10.78	6.36	3.69	2.76	3.92	2.00
	4.62	0.92	2.40	0.67	6.09	3.96	4.76	1.67	2.90	0.84
	5.40	1.90	1.83	1.70	7.78	4.51	4.30	1.51	2.90	0.99
	3.95	0.32	2.40	0.77	5.08	3.76	2.32	1.07	1.82	0.44
	1.55	0.64	7.67	3.53	2.86	1.92	7.35	2.35	4.94	0.84
	1.42	0.69	5.79	3.65	6.30	3.84			5.69	2.84
	1.90	0.93	9.58	4.22					5.54	2.93
Mean	3.377	1.094	5.387	2.586	6.494	3.906	5.045	2.485	3.943	1.560

Design Matrix										
	Cov	T ₁	T ₂	T ₃	T ₄	CT ₁	CT ₂	CT ₃	CT ₄	
	4.34	1	0	0	0	4.34	0	0	0	1.30

	1.90	1	0	0	0	1.90	0	0	0	0.93
	1.55	0	1	0	0	0	1.55	0	0	0.93

	9.58	0	1	0	0	0	9.58	0	0	4.22
	7.18	0	0	1	0	0	0	7.18	0	5.10

X =	6.30	0	0	1	0	0	0	6.30	0	Y =
(47 × 9)	3.94	0	0	0	1	0	0	0	3.94	(47 × 1)

	7.35	0	0	0	1	0	0	0	7.35	2.35
	4.00	-1	-1	-1	-1	-4.00	-4.00	-4.00	-4.00	1.44

	5.54	-1	-1	-1	-1	-5.54	-5.54	-5.54	-5.54	2.93

© Cengage Learning 2013

The full model (including the interaction predictors) states that

$$Y_{ij} = \tau_j + c + c\tau_j + e_{ij}$$

where τ_j represents the treatment effect for the j th treatment, c represents the covariate, $c\tau_j$ represents our term testing homogeneity of regression, and e_{ij} represents the error associated with the i th subject in treatment j .

We can compare two models either on the basis of the change in $SS_{\text{regression}}$ between the two models (using the residual from the more complete model for our error term), or on the basis of the decrease in R^2 . In this case the latter is somewhat simpler.

The regression analysis of this model would produce

$$R^2_{\tau,c,c\tau} = .8238$$

If there is no significant difference in within-treatment regressions—that is, if the regression lines are parallel and thus the slopes of the regression lines that could be calculated for each group separately are homogeneous—called homogeneity of regression—the deletion of the interaction term should produce only a trivial decrement in the percentage of accountable variation. When we delete the CT terms, we have

$$R_{\tau,c}^2 = .8042$$

The F test on this decrement is the usual F test on the difference between two models:

$$\begin{aligned} F(f - r, N - f - 1) &= \frac{(N - f - 1)(R_{\tau,c,c\tau}^2 - R_{\tau,c}^2)}{(f - r)(1 - R_{\tau,c,c\tau}^2)} \\ &= \frac{(47 - 9 - 1)(.8238 - .8042)}{(4)(.1762)} = 1.03 \end{aligned}$$

Given an F of 1.03 on 4 and 37 degrees of freedom, we have no basis to reject the assumption of homogeneity of regression (common regression coefficients) within the five treatments. Thus, we can proceed with the analysis on the basis of the revised full model that does not include the covariate by treatment interaction:

$$Y_{ij} = \mu + \tau_j + c + e_{ij}$$

This model will serve as the basis against which we compare reduced models.

The three sets of results of the multiple-regression solutions using (1) the covariate and dummy treatment variables, (2) just the treatment variables, and then (3) just the covariates are presented in Table 16.8.

From Table 16.8 you can see that using both the covariate (Pre) and the group membership dummy variates ($T_1 \dots T_4$), the sum of squares for regression ($SS_{\text{regression}_{\tau,c}}$) is equal to 82.6435, which is the portion of the total variation that can be accounted for by these two sets of predictors. You can also see that the residual sum of squares (SS_{residual}) is 20.1254, which is the variability that cannot be predicted. In our analysis of covariance summary table, this will become the sum of squares for error.

When we remove the dummy group membership variates from the equation and use only the covariate (Pre) as a predictor, $SS_{\text{regression}}$ drops from 82.6435 to 73.4196. The difference between $SS_{\text{regression}}$ with and without the group membership predictors must be the amount of the sum of squares that can be attributable to treatment *over and above* the amount that can be explained by the covariate. For our data, this is

$$\begin{aligned} SS_{\text{treat(adj)}} &= SS_{\text{regression}_{\tau,c}} - SS_{\text{regression}_c} \\ &= 82.6435 - 73.4196 \\ &= 9.2239 \end{aligned}$$

This last value is called the *adjusted* treatment sum of squares for the analysis of covariance, because it has been adjusted for any effects of the covariate. In this case, it has been adjusted for the fact that the five groups differed on the pretest measure.

We need one additional term to form our analysis of covariance summary table, and that is the sum of squares to be attributed to the covariate. There are a number of different ways to define this term, but the most common is to define it analogously to the way the adjusted treatment effect was defined. We will attribute to the covariate that portion of the variation that cannot be defined by the treatment effect. In other words, we will take the model with both the covariate and treatment predictors and compare it to a model with only the treatment predictors. The difference in the two sums of squares due to regression will

Table 16.8 Regression analysis**(a) Full Model**

$$\hat{Y}_{ij} = 0.4347(\text{Pre}) - 0.5922(T1) + 0.0262(T2) + 0.8644(T3) \\ + 0.0738(T4) + 0.2183 \\ R^2_{\tau,c} = .8042$$

Analysis of Variance Summary Table for Regression

Source	df	SS	MS	F
Regression	5	82.6435	16.5287	33.6726
Residual	41	20.1254	0.4909	
Total	46	102.7689		

(b) Reduced Model—Omitting Treatment Predictors

$$\hat{Y}_{ij} = 0.5311(\text{Pre}) - 0.26667 \\ R^2_c = .7144$$

Analysis of Variance Summary Table for Regression

Source	df	SS	MS	F
Regression	1	73.4196	73.4196	112.5711
Residual	45	29.3493	0.6522	
Total	46	102.7689		

(c) Reduced Model—Omitting Covariate (Pre)

$$\hat{Y}_{ij} = -1.2321(T1) + 0.2599(T2) + 1.5794(T3) + 0.1589(T4) + 2.3261 \\ R^2_{\tau} = .4311$$

Analysis of Variance Summary Table for Regression

Source	df	SS	MS	F
Regression	4	44.3028	11.0757	7.9564
Residual	42	58.4661	1.3921	
Total	46	102.7689		

© Cengage Learning 2013

be the sum of squares that the covariate accounts for *over and above* what is accounted for by treatment effects.⁹ For our data, this is

$$SS_{\text{covariate}} = SS_{\text{regression}_{\tau,c}} - SS_{\text{regression}_{\tau}} \\ = 82.6435 - 44.3028 \\ = 9.2239$$

We now have all the information necessary to construct the analysis of covariance summary table. This is presented in Table 16.9. Notice that in this table the error term is SS_{residual} from the full model and the other sums of squares are as calculated before.

⁹ Not all software arranges things this way, so do not be surprised if you find printout with a different $SS_{\text{covariate}}$.

Table 16.9 Summary tables for analysis of covariance

General Summary Table for One-Way Analysis of Covariance

Source	<i>df</i>	<i>SS</i>
Covariate	<i>c</i>	$SS_{\text{regression}(r,c)} - SS_{\text{regression}(r)}$
Treat (adj)	$k - 1$	$SS_{\text{regression}(r,c)} - SS_{\text{regression}(c)}$
Error	$N - k - 1$	$SS_{\text{residual}(r,c)}$
Total	$N - 1$	

Summary Table for Data in Table 16.7

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Covariate	1	38.3407	38.3407	78.108*
Treat (adj)	4	9.2239	2.3060	4.698*
Residual	41	20.1254	0.4909	
Total	46	102.7689		

Full Model:

$$\hat{Y}_{ij} = 0.4347(\text{Pre}) - 0.5922(T_1) + 0.0262(T_2) + 0.8644(T_3) + 0.0738(T_4) + 0.2183$$

* $p < .05$

© Cengage Learning 2013

Notice also that there is one degree of freedom for the covariate, because there is one covariate; there are $(k - 1) = (5 - 1) = 4$ *df* for the adjusted treatment effect; and there are $N - k - c = 41$ *df* for error (where *k* represents the number of groups and *c* represents the number of covariates).

From the summary table we see that $SS_{\text{treat(adj)}} = 9.2239$. On 4 *df* this gives us $MS_{\text{treat(adj)}} = 2.3060$. Dividing that term by $MS_{\text{error}} = 0.4909$ we have $F = 4.698$ on (4,41) *df*, which is significant at $p < .05$. Thus, we can conclude that after we control for individual preinjection differences in activity, the treatment groups do differ on postinjection activity.

Adjusted Means

Because $F_{.05}(4,41) = 2.61 < F_{\text{obt}} = 4.698$, we have rejected $H_0: \mu_1(\text{adj}) = \mu_2(\text{adj}) = \mu_3(\text{adj}) = \mu_4(\text{adj}) = \mu_5(\text{adj})$ and conclude that there were significant differences among the treatment means after the effect of the covariate has been partialled out of the analysis. To interpret these differences, it would be useful, if not essential, to obtain the treatment means adjusted for the effects of the covariate. We are basically asking for an estimate of what the postinjection treatment means would have been had the groups not differed on the preinjection means. The adjusted means are readily obtained from the regression solution using the covariate and treatments as predictors.

From the analysis of the revised full model, we obtained (see Table 16.8)

$$\hat{Y}_{ij} = 0.4347(\text{Pre}) - 0.5922(T_1) + 0.0262(T_2) + 0.8644(T_3) + 0.0738(T_4) + 0.2183$$

Writing this in terms of means and representing adjusted means as \bar{Y}'_j , we have

$$\begin{aligned}\bar{Y}'_j &= 0.4347(\bar{\text{Pre}}) - 0.5922(T_1) + 0.0262(T_2) + 0.8644(T_3) \\ &\quad + 0.0738(T_4) + 0.2183\end{aligned}$$

where $\bar{\text{Pre}} = 4.8060$ (the mean preinjection score) and $T_1, T_2, T_3,$ and T_4 are (0, 1, -1) variables. (We substitute the mean Pre score for the individual Pre score because we are interested in the adjusted means for Y if all subjects had received the mean score on the covariate.) For our data, the adjusted means of the treatments are:

$$\begin{aligned}\bar{Y}'_1 &= 0.4347(4.8060) - 0.5922(1) + 0.0262(0) + 0.8644(0) \\ &\quad + 0.0738(0) + 0.2183 \\ &= 1.7153\end{aligned}$$

$$\begin{aligned}\bar{Y}'_2 &= 0.4347(4.8060) - 0.5922(0) + 0.0262(1) + 0.8644(0) \\ &\quad + 0.0738(0) + 0.2183 \\ &= 2.3336\end{aligned}$$

$$\begin{aligned}\bar{Y}'_3 &= 0.4347(4.8060) - 0.5922(0) + 0.0262(0) + 0.8644(1) \\ &\quad + 0.0738(0) + 0.2183 \\ &= 3.1719\end{aligned}$$

$$\begin{aligned}\bar{Y}'_4 &= 0.4347(4.8060) - 0.5922(0) + 0.0262(0) + 0.8644(0) \\ &\quad + 0.0738(1) + 0.2183 \\ &= 2.3813\end{aligned}$$

$$\begin{aligned}\bar{Y}'_5 &= 0.4347(4.8060) - 0.5922(-1) + 0.0262(-1) + 0.8644(-1) \\ &\quad + 0.0738(-1) + 0.2183 \\ &= 1.9353\end{aligned}$$

The adjusted means are plotted in Figure 16.4.

The grand mean is

$$\begin{aligned}\bar{Y}' &= 0.4347(4.8060) - 0.5922(0) + 0.0262(0) + 0.8644(0) \\ &\quad + 0.0738(0) + 0.2183 \\ &= 2.3075\end{aligned}$$

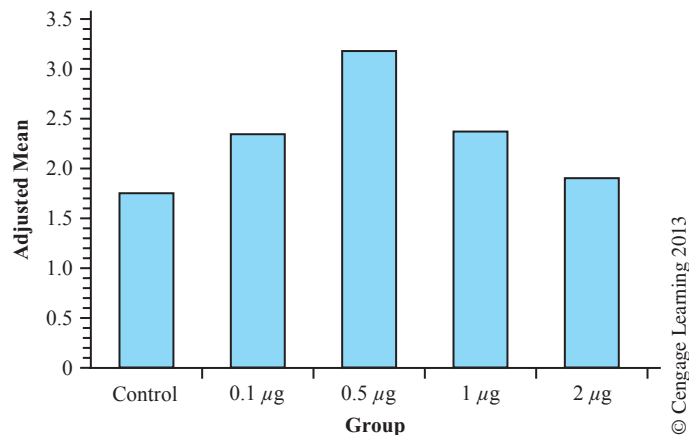


Figure 16.4 Adjusted means by group

which is the mean of the adjusted means. (In a case in which we have equal sample sizes, the adjusted grand mean will equal the unadjusted grand mean.)¹⁰

Now we are about to go into deep water in terms of formulae, and I expect eyes to start glazing over. I can't imagine that anyone is going to expect you to memorize these formulae. Just try to understand what is happening and remember where to find them when you need them. Don't expect to find them printed out by most statistical software.

Any individual comparisons among treatments would now be made using these adjusted means. In this case, however, we must modify our error term from that of the overall analysis of covariance. If we let $SS_{e(c)}$ represent the error sum of squares from an analysis of variance on the *covariate*, then Huitema (1980), in an excellent and readable book on the analysis of covariance, gives as a test of the difference between two adjusted means

$$F(1, N - a - 1) = \frac{(\bar{Y}'_j - \bar{Y}'_k)^2}{MS'_{\text{error}} \left[\left(\frac{1}{n_j} + \frac{1}{n_k} \right) + \frac{(C_j - C_k)^2}{SS_{e(c)}} \right]}$$

MS'_{error}

where MS'_{error} is the error term from the analysis of covariance. For an excellent discussion of effective error terms and comparisons among means, see Winer (1971, pp. 771ff) and, especially, Huitema (1980). Huitema has a new edition due out soon.

As an example, suppose we wish to compare \bar{Y}'_1 and \bar{Y}'_3 , which theory had predicted would show the greatest difference. From the preceding analysis, we either know or can compute

$$MS'_{\text{error}} = 0.4909$$

$$SS_{e(c)} = 202.938 \quad [\text{calculation not shown}]$$

$$\bar{C}_1 = 3.3770 \quad \bar{C}_3 = 6.4944$$

$$\bar{Y}'_1 = 1.7153 \quad \bar{Y}'_3 = 3.1719$$

$$\begin{aligned} F(1, 41) &= \frac{(1.7153 - 3.1719)^2}{0.4909 \left[\left(\frac{1}{10} + \frac{1}{9} \right) + \frac{(3.3770 - 6.4944)^2}{202.938} \right]} \\ &= \frac{2.1217}{0.1271} = 16.69 \end{aligned}$$

The critical value $F_{.05}(1,41) = 4.08$. We would thus reject the null hypothesis that the adjusted means of these two conditions are equal in the population. Even after adjusting for the fact that the groups differed by chance on the pretest, we find significant postinjection differences.

Exhibit 16.4 contains SPSS output for the analysis of variance. (The pretest and posttest means were computed using the **Compare means** procedure.) Notice that I requested a "spread versus level" plot from the options menu. This plots the group means against the group variances. If the variances are homogeneous and independent of the means, this plot should look random. The plot reveals that there is a correlation between the size of the mean and the size of the variance. Notice, however, that the relationship appears very much reduced when we plotted the relationship between the adjusted means and their standard errors.

¹⁰ An alternative approach to calculating adjusted means is to define

$$\bar{Y}'_j = \bar{Y}_j - b_w(\bar{C}_j - \bar{C})$$

where \bar{C}_j is the covariate mean for Group j , \bar{C} is the covariate grand mean, and b_w is the regression coefficient for the covariate from the complete model (here $b_w = 0.4347$). This more traditional way of calculating adjusted means makes it clear that the adjusted mean is some function of how deviant that group was on the covariate. The same values for the adjusted means will result from using either approach.

		Report	
Treatment Group		PRETEST	POSTTEST
Control	Mean	3.3770	1.0940
	N	10	10
	Std. Deviation	1.3963	.5850
0.1 µg	Mean	5.3870	2.5860
	N	10	10
	Std. Deviation	3.4448	1.5332
0.5 µg	Mean	6.4944	3.9056
	N	9	9
	Std. Deviation	2.3781	1.4768
1 µg	Mean	5.0450	2.4850
	N	8	8
	Std. Deviation	1.6876	1.1874
2 µg	Mean	3.9430	1.5600
	N	10	10
	Std. Deviation	1.2207	.8765
Total	Mean	4.8060	2.2857
	N	47	47
	Std. Deviation	2.3788	1.4947

Tests of Between-Subjects Effects

Dependent Variable: POSTTEST

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	82.644 ^a	5	16.529	33.673	.000	.804
Intercept	.347	1	.347	.707	.405	.017
PRETEST	38.341	1	38.341	78.108	.000	.656
GROUP	9.224	4	2.306	4.698	.003	.314
Error	20.125	41	.491			
Total	348.327	47				
Corrected Total	102.769	46				

^aR Squared = .804 (Adjusted R Squared = .780)

Estimated Marginal Means

Treatment Group

Dependent Variable: POSTTEST

Treatment Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Control	1.715 ^a	.232	1.246	2.185
0.1 µg	2.333 ^a	.223	1.882	2.785
0.5 µg	3.172 ^a	.248	2.671	3.672
1 µg	2.381 ^a	.248	1.880	2.882
2 µg	1.935 ^a	.226	1.480	2.391

Adapted from output by SPSS, Inc.

(continued)

Exhibit 16.4 SPSS output for analysis of Conti and Musty data

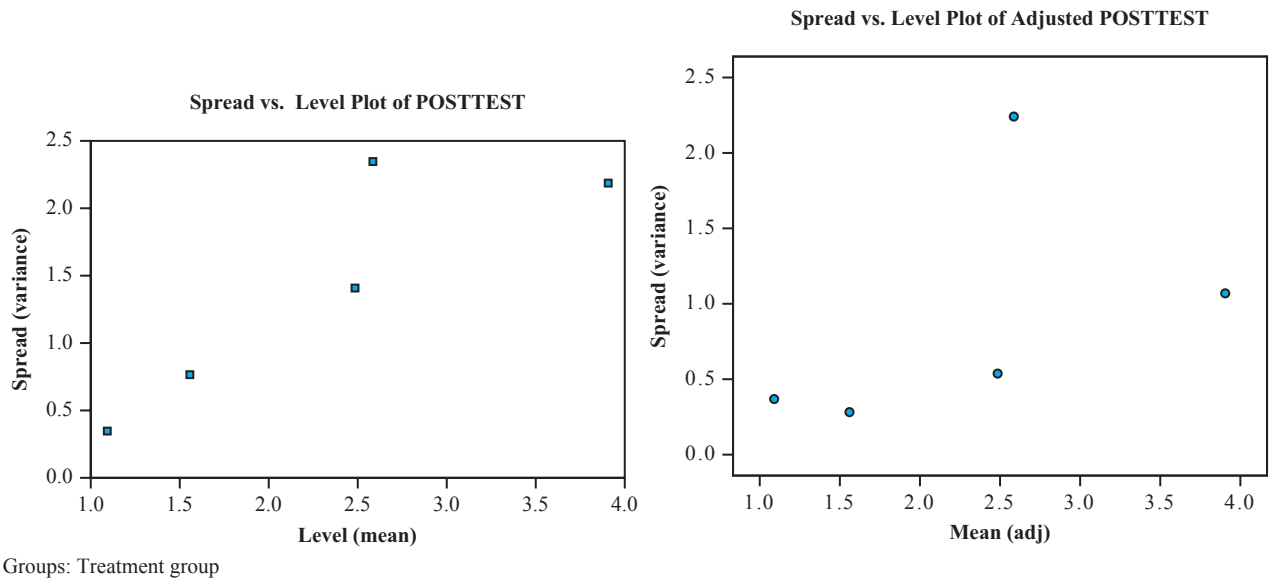


Exhibit 16.4 (continued)

16.6 Computing Effect Sizes in an Analysis of Covariance

As you might expect, computing effect sizes is a bit more complicated in analysis of covariance than it was in the analysis of variance. That is because we have choices to make in terms of the means we compare and the error term we use. You may recall that with factorial designs and repeated measures designs we had a similar problem concerning the choice of the error term for the effect size.

As before, we can look at effect size in terms of *r*-family and *d*-family measures. Normally I would suggest *r*-family measures when looking at an omnibus *F* test, and a *d*-family measure when looking at specific contrasts. We will start with an *r*-family example, and then move to the *d*-family. The example we have been using based on the study by Conti and Musty produced a significant *F* on the omnibus null hypothesis. Probably the most appropriate way to talk about this particular example would make use of the fact that Group (or Dose) was a metric variable, increasing from 0 to 2 μ g.¹¹ However, I am going to take a “second-best” approach here because the majority of the studies we run do not have the independent variable distributed as such an ordered variable.

r-Family Measure

As our *r*-family measure of association we will use η^2 , acknowledging that it is positively biased. You should recall that η^2 is defined as the treatment *SS* divided by the total *SS*. But which sums of squares for treatments should we use—the ones from an analysis of variance on the dependent variable, or the ones from the analysis of covariance? Kline (2004) offers both of those alternatives, though he uses an adjusted SS_{total} in the second,¹² without

¹¹ SPSS will test polynomial contrasts on the adjusted means. Just click on the CONTRAST button and ask for polynomial contrasts. For this example there is a significant quadratic component.

¹² SPSS uses this same adjustment if you request effect sizes, and it is simply $SS_{\text{treat}} + SS_{\text{error}}$.

suggesting a choice. If the covariate naturally varies in the population (as it does in this case, where we expect different animals to vary in their pretest score, then it makes the most sense to divide the SS_{treat} from the analysis of covariance by the SS_{total} (unadjusted) from that analysis. This will produce a value of η^2 , which is the percentage of “normal variation” accounted for by the independent variable.¹³ Then

$$\eta^2 = \frac{SS_{\text{treat(adj)}}}{SS_{\text{total}}} = \frac{9.2239}{102.7689} = .09$$

An alternative approach, which will produce the same answer, is to take η^2 as the difference between the R^2 from a model predicting the dependent variable from only the covariate (the pretest) and one predicting the dependent variable from both the covariate and the treatment. The increase in explained variation from the first of these models to the second represents what the treatment contributes after controlling for the covariate. For our example R^2 using just the covariate is .714. (You can obtain this by an analysis of variance using the covariate as the independent variable, or by a regression of the independent variable on the covariate.) When you add in the treatment effect the R^2 is .804. These values are shown in the following table.

Step	Predictors	R^2	Change in R^2	F for change
1	Pretest	.714		
2	Pretest, Treatment	.804	.090	4.689

η^2 is the difference between these two values of R^2 , which is the contribution to explained variation of the treatment after controlling for the covariate. This is the same value we obtained by the first approach.

d-Family Measure

Measures from the d -family often are more interpretable, and they are most often used for specific contrasts between two means. The example we have been using is not a very good one for a contrast of two means because the independent variable is a continuum. But I will use the contrast between the control group and the 0.5 μg group as an example, because these are the two conditions that Conti and Musty’s theory would have expected to show the greatest mean difference. Because we are working with an analysis of covariance, the appropriate means to compare are the adjusted means (\bar{Y}_i) from that analysis. In this case they are 3.1719 for the .5 μg condition and 1.7153 for the control condition. (You may recall that we performed a test on the difference between these adjusted means in the previous section, and it was significant.)

We have generally used

$$d = \frac{\hat{\psi}}{\hat{\sigma}}$$

as our effect size estimate. When we are comparing two group means, $\hat{\psi}$ is simply the difference between the two means because the coefficients are $[-1 \ 0 \ 1 \ 0 \ 0]$. For our example, $\hat{\psi}$ is $3.1719 - 1.7153 = 1.4566$. But the question of most importance is what we will use for the estimate of the standard deviation. One of several choices would be the square root of MS_{error} from an analysis of variance, because this would be an estimate of the average variability within each group, and would thus standardize the mean difference in the

¹³ If you were interested in the η^2 for the quadratic relationship between dose and the activity level, controlling for the pretest activity level, you could just divide the $SS_{\text{quadratic}}$ by SS_{total} .

metric of the original measurements. (Recall that we used SS_{total} from the analysis of variance when we calculated η^2 .) An alternative would be the square root of MS_{error} from the analysis of covariance, which would standardize the mean difference in the metric of the adjusted scores, which is a bit more difficult to understand. Cortina and Nouri (2000) have made the sensible suggestion that when the covariate normally varies in the population, as ours does, we want to include that variability in our estimate of error. This means that we would use the square root of MS_{error} from the analysis of variance on the posttest scores. In that analysis MS_{error} is 1.392, which is simply the weighted mean of the group variances. Then

$$d = \frac{\hat{\psi}}{\hat{\sigma}} = \frac{\bar{X}_3 - \bar{X}_1}{\sqrt{MS_{\text{error}}}} = \frac{3.1719 - 1.7153}{\sqrt{1.392}} = \frac{1.4566}{1.1798} = 1.23$$

Injection of the optimal dose of THC ($0.5 \mu\text{g}$) leads to an increase of postinjection activity by nearly 1 and a quarter standard deviations relative to the control group.

16.7 Interpreting an Analysis of Covariance

Interpreting an analysis of covariance can present certain problems, depending on the nature of the data and, more important, the design of the experiment. A thorough and readable discussion of most of these problems is presented by Huitema (1980). Other important sources for consideration of these problems are Anderson (1963), Evans and Anastasio (1968), Huitema (2005), Lord (1967, 1969), Maxwell and Cramer (1975), Reichardt (1979), Smith (1957), and Weisberg (1979).

The ideal application for an analysis of covariance is an experiment in which participants are randomly assigned to treatments (or cells of a factorial design). In that situation, the *expected value* of the covariate mean for each group or cell is the same, and any differences can be attributed only to chance, assuming that the covariate was measured before the treatments were applied. In this situation, the analysis of covariance will primarily reduce the error term, but it will also, properly, remove any bias in the dependent variable means caused by chance group differences on the covariate. This was the situation in the Conti and Musty (1984) study that we have been discussing.

In a randomized experiment in which the covariate is measured *after* the treatment has been applied and has affected the covariate, interpreting the results of an analysis of covariance is difficult at best. In this situation the expected values of the group covariate means are not equal, even though the subjects were assigned randomly. It is difficult to interpret the results of the analysis because you are asking what the groups would have been like had they not differed on the covariate, when in fact the covariate differences may be an integral part of the treatment effect. This problem is particularly severe if the covariate was measured with error (i.e., if it is not perfectly reliable). In this case an alternative analysis, called the **true-score analysis of covariance**, may be appropriate if the other interpretive problems can be overcome. Such an analysis is discussed in Huitema (1980, Chapter 14).

When subjects are not assigned to the treatment groups at random, interpreting the analysis of covariance can be even more troublesome. The most common example of this problem is what is called the **nonequivalent groups design**. In this design, two (or more) intact groups are chosen (e.g., schools or classrooms of children), a pretest measure is obtained from subjects in both groups, the treatment is applied to one of the groups, and the two groups are then compared on some posttest measure. Since participants are not assigned to the groups at random, we have no basis for assuming that any differences that exist on the pretest are to be attributed to chance. Similarly, we have no basis for expecting

**true-score
analysis of
covariance**

**nonequivalent
groups design**

the two groups to have the same mean on the posttest in the absence of a real treatment effect. Huitema (1980, pp. 149ff) gives an excellent demonstration that when the groups differ at the beginning of the experiment, the phenomenon of regression to the mean could lead to posttest differences even in the absence of a treatment effect. For alternative analyses that are useful under certain conditions, see Huitema (1980). Maris (1998) takes a different view of the issue.

The problems of interpreting results of designs in which subjects are not randomly assigned to the treatment groups are not easily overcome. This is one of the reasons why random assignment is even more important than random selection of subjects. It is difficult to overestimate the virtues of random assignment, both for interpreting data and for making causal statements about the relationship between variables. In what is probably only a slight overstatement of the issue, Lord (1967) remarked, “In the writer’s opinion, the explanation is that with the data usually available for such studies, there is simply no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled pre-existing differences between groups” (p. 305). (Lord was *not* referring to differences that arise by chance through random assignment.) Anderson (1963) made a similar point by stating, “One may well wonder exactly what it means to ask what the data would be like if they were not what they are” (p. 170). All of this is not to say that the analysis of covariance has no place in the analysis of data in which the treatments differ on the covariate. Anyone using covariance analysis, however, must think carefully about her data and the practical validity of the conclusions she draws.

16.8 Reporting the Results of an Analysis of Covariance

The only difference between describing the results of an analysis of covariance and an analysis of variance is that we must refer to the covariate and to adjusted means. For the experiment by Conti and Musty we could write

Conti and Musty (1984) examined the effect of THC on locomotor activity in rats. They predicted that moderate doses of THC should show the greatest increase in activity (or the least decrease due to adaptation). After a pretesting session five different groups of rats were randomly assigned to receive 0, 0.1 μg , 0.5 μg , 1 μg , or 2 μg of THC. Activity level was measured in a 10-minute postinjection interval. Because there was considerable variability in pretest activity, the pretest measure was used as a covariate in the analysis.

The analysis of covariance was significant ($F(4,41) = 4.694, p = .003$), with intermediate doses showing greater effect. Eta-squared was .09 using a SS_{total} that has not been adjusted for the covariate. A contrast of the means of the control group and the 0.5 μg group revealed a significant difference ($F(1,41) = 16.69, p < .05$), with a standardized effect size (d) of 1.23.

16.9 The Factorial Analysis of Covariance

The analysis of covariance applies to factorial designs just as well as it does to single-variable designs. Once again, the covariate may be treated as a variable that, because of methodological considerations, assumes priority in the analysis. In this section we will deal only with the case of equal cell sizes, but the generalization to unequal n s is immediate.

The logic of the analysis is straightforward and follows that used in the previous examples. $SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}}$ is the variation attributable to a linear combination of the covariate, the main effects of A and B , and the AB interaction. Similarly, $SS_{\text{regression}_{c, \alpha, \beta}}$ is the variation attributable to the linear combination of the covariate and the main effects of A and B . The difference

$$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} - SS_{\text{regression}_{c, \alpha, \beta}}$$

is the variation attributable to the AB interaction, with the covariate and the main effects partialled out. Since, with equal sample sizes, the two main effects and the interaction are orthogonal, all that is *actually* partialled out in equal n designs is the covariate.

By the same line of reasoning

$$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} - SS_{\text{regression}_{c, \alpha, \alpha\beta}}$$

represents the variation attributable to B , partialling out the covariate, and

$$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} - SS_{\text{regression}_{c, \beta, \alpha\beta}}$$

represents the variation attributable to the main effect of A , again partialling out the covariate.

The error term represents the variation remaining after controlling for A , B , and AB , and the covariate. As such it is given by

$$SS_{\text{residual}_{c, \alpha, \beta, \alpha\beta}}$$

The general structure of the analysis is presented in Table 16.10. Notice that once again the error term loses a degree of freedom for each covariate. Because the independent variable and the covariate account for overlapping portions of the variation, their sums of squares will not equal SS_{total} .

As an example, consider the study by Spilich et al. (1992) that we examined in Chapter 13 on performance as a function of cigarette smoking. In that study subjects performed either a Pattern Recognition task, a Cognitive task, or a Driving Simulation task. The subjects were divided into three groups. One group (Active Smoking) smoked during or just before the task. A second group (Delayed Smoking) were smokers who had not smoked for three hours, and a third group (NonSmoking) was composed of NonSmokers. The dependent variable was the number of errors on the task. To make this suitable for an analysis of covariance I have added an additional (hypothetical) variable, which is the subject's measured level of distractibility. (Higher distractibility scores indicate a greater ease at being distracted.)

The data are presented in Table 16.11 and represent a 3×3 factorial design with one covariate (Distract).

Table 16.12 contains an abbreviated form of the design matrix, showing only the entries for the first and last subject in each cell. Notice that the matrix contains a column for

Table 16.10 Structure of the analysis of covariance for factorial designs

Source	df	SS
A(adj)	$a - 1$	$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} - SS_{\text{regression}_{c, \beta, \alpha\beta}}$
B(adj)	$b - 1$	$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} - SS_{\text{regression}_{c, \alpha, \alpha\beta}}$
AB (adj)	$(a - 1)(b - 1)$	$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} - SS_{\text{regression}_{c, \alpha, \beta}}$
Error	$N - ab - c$	$SS_{\text{residual}_{c, \alpha, \beta, \alpha\beta}}$
Covariate	c	$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} - SS_{\text{regression}_{\alpha, \beta, \alpha\beta}}$
Total	$N - 1$	

Table 16.11 Hypothetical data on smoking and performance (modeled on Spilich et al., 1992)

		Pattern Recognition														
NS:	Errors	9	8	12	10	7	10	9	11	8	10	8	10	8	11	10
	Distract	107	133	123	94	83	86	112	117	130	111	102	120	118	134	97
DS:	Errors	12	7	14	4	8	11	16	17	5	6	9	6	6	7	16
	Distract	101	75	138	94	138	127	126	124	100	103	120	91	138	88	118
AS:	Errors	8	8	9	1	9	7	16	19	1	1	22	12	18	8	10
	Distract	64	135	130	106	123	117	124	141	95	98	95	103	134	119	123
		Cognitive Task														
NS:	Errors	27	34	19	20	56	35	23	37	4	30	4	42	34	19	49
	Distract	126	154	113	87	125	130	103	139	85	131	98	107	107	96	143
DS:	Errors	48	29	34	6	18	63	9	54	28	71	60	54	51	25	49
	Distract	113	100	114	74	76	162	80	118	99	146	132	135	111	106	96
AS:	Errors	34	65	55	33	42	54	21	44	61	38	75	61	51	32	47
	Distract	108	191	112	98	128	145	76	107	128	128	142	144	131	110	132
		Driving Simulation														
NS:	Errors	3	2	0	0	6	2	0	6	4	1	0	0	6	2	3
	Distract	130	83	91	92	109	106	99	109	136	102	119	84	68	67	114
DS:	Errors	7	0	6	0	12	17	1	11	4	4	3	5	16	5	11
	Distract	93	102	108	100	123	131	99	116	81	103	78	103	139	101	102
AS:	Errors	15	2	2	14	5	0	16	14	9	17	15	9	3	15	13
	Distract	110	96	112	114	137	125	168	102	109	111	137	106	117	101	116

© Cengage Learning 2013

the covariate (denoted C), the usual design matrix elements for the main effects of Task (T) and Group (G), and the Task \times Group interaction. I have also added columns representing the interaction of the covariate with the Task \times Group interaction. The latter will be used to test the hypothesis $H_0: b_i^* = b_j^*$ for all values of i and j , because the assumption of homogeneity of regression applies to any analysis of covariance.

It is important to consider just what the interactions involving the covariate represent. If I had included the terms CT_1 and CT_2 I could have used them to test the null hypothesis that the regression lines of Errors as a function of Distract have equivalent slopes in the three tasks. Similarly, if I had included CG_1 and CG_2 , I could have tested homogeneity of regression in each of the three smoking groups. Because I am most interested in testing the hypothesis of homogeneity of regression in each of the nine cells, I have included only the CTG_{ij} terms.

The first regression is based on all predictors in X . From this analysis we obtain

$$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta, \alpha\beta c}} = 36724.771$$

$$MS_{\text{residual}_{c, \alpha, \beta, \alpha\beta, \alpha\beta c}} = 71.134$$

If we drop the interaction terms representing the interaction of the covariate (Distract) with the Task \times Group interaction, we have

$$SS_{\text{regression}_{c, \alpha, \beta, \alpha\beta}} = 36389.60175$$

The difference between these two sums of squares = 335.16925. The most complete model had 13 degrees of freedom, while the second had 9 df , meaning that the above sum of squares is based on $13 - 9 = 4$ df . Converting to mean squares we have

$$MS_{\text{difference}} = 335.16925/4 = 83.79231$$

Table 16.12 Design matrix for the analysis of covariance for smoking data

	<i>C</i>	<i>T</i> ₁	<i>T</i> ₂	<i>G</i> ₁	<i>G</i> ₂	<i>TG</i> ₁₁	<i>TG</i> ₁₂	<i>TG</i> ₂₁	<i>TG</i> ₂₂	<i>cTG</i> ₁₁	<i>cTG</i> ₁₂	<i>cTG</i> ₂₁	<i>cTG</i> ₂₂	
<i>X</i>	107	1	0	1	0	1	0	0	0	107	0	0	0	9
=
	97	1	0	1	0	1	0	0	0	97	0	0	0	10
	101	1	0	0	1	0	1	0	0	0	101	0	0	12

	118	1	0	0	1	0	1	0	0	0	118	0	0	16
	64	1	0	-1	-1	-1	-1	0	0	-64	-64	0	0	8

	123	1	0	-1	-1	-1	-1	0	0	-123	-123	0	0	10
	126	0	1	1	0	0	0	1	0	0	0	126	0	27

	143	0	1	1	0	0	0	1	0	0	0	143	0	49
	113	0	1	0	1	0	0	0	1	0	0	0	113	48

	96	0	1	0	1	0	0	0	1	0	0	0	96	49
	108	0	1	-1	-1	0	0	-1	-1	0	0	-108	-108	34

	132	0	1	-1	-1	0	0	-1	-1	0	0	-132	-132	47
	130	-1	-1	1	0	-1	0	1	0	-130	0	-130	0	3

	114	-1	-1	1	0	-1	0	1	0	-114	0	-114	0	3
	93	-1	-1	0	1	0	-1	0	-1	0	-93	0	-93	7

	102	-1	-1	0	1	0	-1	0	-1	0	-102	0	-102	11
	110	-1	-1	-1	-1	1	1	1	1	110	110	110	110	15

	116	-1	-1	-1	-1	1	1	1	1	116	116	116	116	13

© Cengage Learning 2013

We can test the difference between these two models by using $MS_{residual}$ from the more complete model and computing

$$F = \frac{MS_{difference}}{MS_{residual}} = \frac{83.792}{71.134} = 1.18$$

This is an F on $[(f - r), (N - f - 1)] = 4$ and 121 df . The critical value is $F_{.05}(4, 121) = 2.45$, so we will not reject the null hypothesis of homogeneity of regression. We will conclude that we have no reason to doubt that the regression lines relating Errors to Distract have the same slope in the nine cells. This will allow us to adopt a much simpler full model against which to compare subsequent reduced models. Our revised full model is

$$\hat{Y} = b_0 + b_1C + b_2T_1 + b_3T_2 + b_4G_1 + b_5G_2 + b_6TG_{11} + b_7TG_{12} + b_8TG_{21} + b_9TG_{22}$$

or, in more traditional analysis of variance terms,

$$Y_{ijk} = \mu + C_k + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

The results of the several multiple regression solutions needed for the analysis of covariance are shown in Table 16.13. By calculating and testing the differences between full and reduced models, you will be able to compute the complete analysis of covariance.

Exhibit 16.5 contains the results of an SPSS analysis of these data. You should compare your results with the results in that exhibit.

Table 16.13 Regression results for various models for data in Table 16.11

Model	SS _{regression}	MS _{residual}	R ²
C, T, G, TG	36,389.602	71.539	.803
C, T, G	35,154.816		.775
C, G, TG	12,519.117		.276
C, T, TG	35,416.789		.781
T, G, TG	31,744.726		.700

© Cengage Learning 2013

Tests of Between-Subjects Effects

Dependent Variable: ERRORS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Noncent. Parameter	Observed Power ^a
Corrected Model	36389.602 ^b	9	4043.289	56.519	.000	.803	508.671	1.000
Intercept	892.395	1	892.395	12.474	.001	.091	12.474	.939
DISTRACT	4644.876	1	4644.876	64.928	.000	.342	64.928	1.000
TASK	23870.485	2	11935.243	166.836	.000	.727	333.673	1.000
SMKGRP	563.257	2	281.629	3.937	.022	.059	7.873	.699
TASK * SMKGRP	1626.511	4	406.628	5.684	.000	.154	22.736	.977
Error	8942.324	125	71.539					
Total	90341.000	135						
Corrected Total	45331.926	134						

^a Computed using alpha = .05

^b R Squared = .803 (Adjusted R Squared = .789)

1. Task * Smoking Group

Dependent Variable: ERRORS

Task	Smoking Group	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Patrecog	NonSmokers	9.805 ^a	2.184	5.482	14.128
	Delayed smokers	9.732 ^a	2.184	5.410	14.054
	Active Smokers	9.558 ^a	2.184	5.235	13.882
Cognitive	NonSmokers	27.770 ^a	2.188	23.440	32.101
	Delayed smokers	40.436 ^a	2.185	36.112	44.760
	Active Smokers	43.785 ^a	2.233	39.366	48.204
Driving	NonSmokers	8.505 ^a	2.191	4.169	12.842
	Delayed smokers	8.921 ^a	2.200	4.568	13.275
	Active Smokers	5.820 ^a	2.226	1.414	10.226

^a Evaluated at covariates appeared in the model: DISTRACT = 112.52.

(continued)

Exhibit 16.5 SPSS analysis of covariance of Spilich data

Adapted from output by SPSS, Inc.

2. Task

Dependent Variable: ERRORS

Task	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Patrecog	9.699 ^a	1.261	7.203	12.194
Cognitive	37.330 ^a	1.274	34.810	39.851
Driving	7.749 ^a	1.273	5.230	10.268

^aEvaluated at covariates appeared in the model: DISTRACT = 112.52.**3. Smoking Group**

Dependent Variable: ERRORS

Smoking Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
NonSmokers	15.360 ^a	1.264	12.859	17.862
Delayed smoker	19.696 ^a	1.266	17.191	22.202
Active Smokers	19.721 ^a	1.261	17.225	22.217

^aEvaluated at covariates appeared in the model: DISTRACT = 112.52.**Exhibit 16.5** (continued)

For purposes of comparison I have presented the analysis of variance from Exhibit 13.1. This is the analysis on the same data, but without the covariate.

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Task	2	28,661.526	14,330.763	132.895*
Group	2	1813.748	906.874	8.410*
Task × Group	4	1269.452	317.363	2.943*
Error	126	13,587.200	107.835	
Total	134	45,331.926		

 $p < .05$

© Cengage Learning 2013

Notice that in this analysis we have a significant effect due to Task, which is uninteresting because the tasks were quite different and we would expect that some tasks would lead to more errors than others. We have a significant effect due to Group, but the interaction better addresses our interests. We have a Task × Group interaction, which was what we were seeking because it tells us that smoking makes a difference in certain kinds of situations (which require a lot of cognitive processing) but not in others. Notice also that our MS_{error} was 107.835, whereas in the analysis of covariance it was 71.539.

When we look at our analysis of covariance, one of the first things we see is that MS_{error} (71.539) is about one-third smaller than it was in the analysis of variance. This is due to the fact that the covariate (Distract) was able to explain much of the variability in Errors that had been left unexplained in the analysis of variance.

Notice that Exhibit 16.5 presents *partial* eta-squared for the effects. These effect-size measures can be calculated as the difference between two R^2 values, divided by $(1 - R^2_{\text{reduced}})$. For example, the model without the dummy variables for Task has an $R^2 = .276$. This leaves $1 - .276 = 72.4\%$ of the variation unexplained. When we add in the Task variables (going to the full model) we have $R^2 = .803$. This is an increase of $.803 - .276 = .527$, which accounts for $.527/.724 = 72.8\%$ of the variation *that had been left unexplained*. This is the value given in Exhibit 16.5 for Task, although SPSS doesn't mention that this is a partial η^2 . Similar calculations will reproduce the other values.

Adjusted Means

The method of obtaining adjusted means is simply an extension of the method employed in the Conti and Musty example. We want to know what the cell means would have been if the treatment combinations had not differed on the covariate.

From the full model we have

$$\hat{Y} = b_0 + b_1C + b_2T_1 + b_3T_2 + b_4G_1 + b_5G_2 + b_6TG_{11} + b_7TG_{12} + b_8TG_{21} + b_9TG_{22}$$

which equals

$$\begin{aligned} \hat{Y} = & -14.654 - 8.561T_1 + 19.071T_2 - 3.794G_1 + 1.437G_2 + 3.901TG_{11} - 1.404TG_{12} \\ & - 5.766TG_{21} + 1.668TG_{22} + 0.293\text{Distract} \end{aligned}$$

Because we want to know what the Y means would be if the treatments did not differ on the covariate, we will set $C = \bar{C} = 112.518$ for all treatments.

For all observations in Cell₁₁ the appropriate row of the design matrix, with C replaced by \bar{C} , is

$$1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 112.518$$

Applying the regression coefficients and taking the intercept into account, we have

$$\begin{aligned} \hat{Y} = & -14.654 - 8.561(1) + 19.071(0) - 3.794(1) + 1.437(0) + 3.901(1) \\ & - 1.404(0) - 5.766(0) + 1.668(0) + 0.293(112.518) \\ = & 5.860 \end{aligned}$$

Applying this procedure to all cells we obtain the following adjusted cell means

	Pattern Rec	Cognitive	Driving	Row Means
NonSmokers	9.805	27.770	5.820	14.465
Delayed	9.732	40.436	8.921	19.696
Active	9.558	43.785	8.505	20.616
Column Means	9.699	37.330	7.749	18.259

These are the cell means given in Exhibit 16.5, and the row and column means can be found as the mean of the cells in that row or column.

Testing Adjusted Means

The adjusted means are plotted in Figure 16.5. They illustrate the interaction and also the meaning that may be attached to the main effects. Further analyses of these data are probably unnecessary because differences due to smoking seemed to be confined to the condition that requires high levels of cognitive processing. However, for the sake of completeness we

will assume that you wish to make a comparison between the mean of the NonSmoking group and the combined means of the Active and Delayed groups. In this case you want to compare \bar{X}'_1 , with \bar{X}'_2 , and \bar{X}'_3 , combined. This comparison requires some modification of the error term, to account for differences in the covariate. This adjustment is given by Winer (1971) as

$$MS''_{\text{error}} = MS'_{\text{error}} \left[1 + \frac{\frac{SS_{g(c)}}{g-1}}{SS_{e(c)}} \right]$$

where $SS_{g(c)}$ and $SS_{e(c)}$ represent the sum of squares attributable to Groups and Error (respectively) in an analysis of variance on the *covariate*, and MS'_{error} is the error term from the overall analysis of covariance. This is not a very memorable formula, and it is one that I can see no reason to remember.

$$\begin{aligned} MS'_{\text{error}} &= 71.538 \\ SS_{g(c)} &= 2701.215 \\ SS_{e(c)} &= 54285.867 \end{aligned}$$

Thus

$$MS''_{\text{error}} = 71.538 \left[1 + \frac{\frac{2701.215}{2-1}}{54285.867} \right] = 75.098$$

To compare the adjusted means we have

$$\begin{aligned} \psi &= 2(14.465) - 1(19.696) - 1(20.616) = -11.382 \\ F(1,125) &= \frac{n\psi^2}{\sum a_i^2 MS''_{\text{error}}} = \frac{45(-11.382)^2}{6(75.098)} = 12.938 \end{aligned}$$

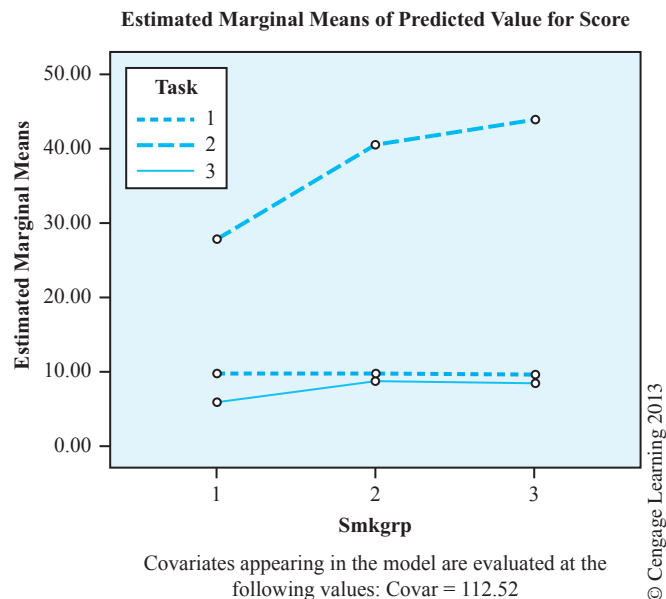


Figure 16.5 Adjusted cell means as a function of Group and Task

Because $F_{.05}(1,125) = 3.92$, we can reject H_0 and conclude that the Active Smoking group performs more poorly (overall) than the average of the other two groups.

Another experimenter might be interested in examining the effects of Group only for the Cognitive task. If we want to examine these simple effects, we will again need to modify our error term in some way. This is necessary because we will be looking at Groups for only some of the data, and the covariate mean of the Cognitive task subjects may differ from the covariate mean for all subjects. Probably the safest route here would be to run a separate analysis of covariance for only those subjects performing the cognitive task. Although this method has the disadvantage of costing us degrees of freedom for error, it has the advantage of simplicity and eliminates the need to make discomfiting assumptions in the adjustment of our error term.

To complete our discussion of the tests we might wish to conduct, consider the experimenter who wants to compare two particular adjusted cell means (whether or not they are in the same row or column). The adjusted error term for this comparison was given by Winer (1971) as

$$MS''_{\text{error}} = \frac{2MS'_{\text{error}}}{n} \left[1 + \frac{\frac{SS_{\text{cells}(c)}}{tg - 1}}{SS_{e(c)}} \right]$$

where $SS_{\text{cells}(c)}$ is the sum of squares from an analysis of variance on the covariate.

You may wonder why we continually worry about adjusting the error term in making comparisons. The general nature of the answer is apparent when you recall what the confidence limits around the regression line looked like in Chapter 9. (They were curved—in fact, they were elliptical.) For $X_i = \bar{X}$, we were relatively confident about \hat{Y} . However, as X_i departed more and more from \bar{X} we became less and less confident of our prediction, and consequently the confidence limits widened. If you now go back to Figure 16.3, you will see that the problem applies directly to the case of adjusted means. In that figure, \bar{Y}'_1 is a long way from \bar{Y}_1 , and we would probably have relatively little confidence that we have estimated it correctly. On the other hand, we can probably have a reasonable degree of confidence in our estimate of \bar{Y}'_2 . It is just this type of consideration that causes us constantly to adjust our error term.

The example we have used points up an important feature of the analysis of covariance—the fact that the covariate is just another variable that happens to receive priority. In designing the study, we were concerned primarily with evaluating the effects of smoking. However, we had two variables that we considered it necessary to control: type of task and distractibility. The first one (Task) we controlled by incorporating it into our design as an independent variable. The second (Distractibility) we controlled by measuring it and treating it as a covariate. In many respects, these are two ways of treating the same problem. Although there are obvious differences in the way these two variables are treated, there are also important similarities. In obtaining SS_{group} , we are actually partialling out *both* Task and the covariate. It is true that in the case of equal ns Task is orthogonal to Group, leaving nothing to partial out; but that is merely a technicality. In the case of unequal ns , the partialling out of both variables is a very real procedure. Although it is important not to lose sight of the fact that the analysis of covariance is a unique technique with its own additional assumptions, it is equally important to keep in mind that a covariate is just another variable.

16.10 Using Multiple Covariates

We have been concerned with the use of a single covariate. There is no theoretical or practical reason, however, why we must restrict ourselves in this way. For example, a study on the effectiveness of several different teaching methods might wish to treat IQ, Age, and

Type of School (progressive or conservative) as covariates. When viewed from the point of view of multiple regression, this presents no particular problem, whereas when viewed within the traditional framework of the analysis of variance, the computational complexities for only a very few covariates would be overwhelming.

In the expression $R^2_{c, \alpha, \beta, \alpha\beta}$, β is really only a shorthand way of representing a set of predictors (e.g., B_1, B_2, \dots, B_b). By the same token, c can be used to stand for a set of covariates (C_1, C_2, \dots, C_k). Thus, in terms of the more specific notation, $R^2_{c, \alpha, \beta, \alpha\beta}$ might really represent

$$R^2_{0, \text{IQ}, \text{Age}, \text{School}, A_1, B_1, B_2, AB_{11}, AB_{12}}$$

When seen in this light, the use of multiple covariates is no different from that of single covariates. If C represents the covariates IQ, Age, and School, then $SS_{AB(\text{adj})}$ remains

$$SS_{AB(\text{adj})} = SS_{\text{regression}(\text{IQ}, \text{Age}, \text{School}, A_1, B_1, B_2, AB_{11}, AB_{12})} - SS_{\text{regression}(\text{IQ}, \text{Age}, \text{School}, A_1, B_1, B_2)}$$

It should be apparent from the previous example that no restriction is placed on the nature of the covariate, other than that it is assumed to be linearly related to the criterion. It can be a continuous variable, as in the case of IQ and Age, or a discrete variable, as in the dichotomous classification of Schools as progressive and conservative.

A word of warning: Just because it is possible (and in fact easy) to use multiple covariates is not a good reason for adopting this procedure. Interpreting an analysis of covariance may be difficult enough (if not impossible) with only one covariate. The problems increase rapidly with the addition of multiple covariates. Thus, it might be easy to *say*, in evaluating several methods of teaching English, that such and such a method is better if groups are equated for age, IQ, type of school, parents' occupation, and so on. But the experimenter must then ask himself if such equated groups actually exist in the population. If they do not, he has just answered a question about what would happen in groups that could never exist, and it is unlikely that he will receive much applause for his efforts. Moreover, even if it is possible to form such groups, will they behave in the expected manner? The very fact that the students are now in homogeneous classes may itself have an effect on the dependent variable that could not have been predicted.

16.11 Alternative Experimental Designs

stratification

The analysis of covariance is not the only way to handle data in which a covariate is important. Two common alternative procedures are also available: **stratification** (matched samples) and difference scores.

If we have available measures on the covariate and are free to assign subjects to treatment groups, then we can form subsets of subjects who are homogeneous with respect to the covariate, and then assign one member of each subset to a different treatment group. In the resulting analysis of variance, we can then pull out an effect due to blocks (subsets) from the error term.

The use of matched samples and the analysis of covariance are almost equally effective when the regression of Y on C is linear. If ρ equals the correlation in the population between Y and C , and σ_e^2 represents the error variance in a straight analysis of variance on Y , then the use of matched samples reduces the error variance to

$$\sigma_e^2(1 - \rho^2)$$

The reduction due to the analysis of covariance in this situation is given by

$$\sigma_e^2(1 - \rho^2) \frac{(f_e)}{(f_e - 1)}$$

where f_e is the degrees of freedom for the error variance. Obviously, for any reasonable value of f_e , the two procedures are almost equally effective, assuming linearity of

difference scores

regression. If the relationship between Y and C is not linear, however, matching will be more effective than covariance analysis.

A second alternative to the analysis of covariance concerns the use of **difference scores**. If the covariate (C) represents a test score before the treatment is administered and Y a score on the same test after the administration of the treatment, the variable $C - Y$ is sometimes used as the dependent variable in an analysis of variance to control for initial differences on C . Obviously, this approach will work only if C and Y are comparable measures. We could hardly justify subtracting a driving test score (Y) from an IQ score (C). If the relationship between C and Y is linear and if $b_{CY} = 1.00$, which is rarely true, the analysis of difference scores and the analysis of covariance will give the same estimates of the treatment effects. When b_{CY} is not equal to 1, the two methods will produce different results, and in this case it is difficult to justify the use of difference scores. In fact, for the Conti and Musty (1984) data on THC, if we took the *difference* between the Pre and Post scores as our dependent variable, the results would be decidedly altered ($F_{4,42} = 0.197$). In this case, the analysis of covariance was clearly a more powerful procedure. Exercise 16.24 at the end of the chapter illustrates this view of the analysis of covariance. For a more complete treatment of this entire problem, see Harris (1963) and Huitema (1980, 2005).

The thing to keep in mind here is that a slope of 1.0 on the relationship between pre- and post-test scores implies that the intervention led to a similar increase in scores, regardless of where people started. But it might be that the change is *proportional* to where people started out. Someone who is very poor in math may have much more to gain by an intervention program than someone who was doing well, and thus the gain score will be directly (and negatively) related to the pretest score. In the example from Conti and Musty (1984), more active animals were likely to change more than less active animals, which may be why they took as their dependent variable the posttest score as a percentage of the pretest score, rather than just the difference between their two scores.

Key Terms

General linear model (16.1)	Type I SS (16.4)	Homogeneity of regression (16.5)
Design matrix (16.1)	Hierarchical sums of squares (16.4)	MS'_{error} (16.5)
Method III (16.4)	Sequential sums of squares (16.4)	True-score analysis of covariance (16.6)
Type III SS (16.4)	Equally weighted means (16.4)	Nonequivalent groups design (16.6)
Method II (16.4)	Analysis of covariance (16.5)	Stratification (16.11)
Type II SS (16.4)	Covariate (16.5)	Difference scores (16.11)
Method I (16.4)	Adjusted Y means (16.5)	

Exercises

- 16.1 The following hypothetical data were obtained from poor, average, and good readers on the number of eye fixations per line of text.

Poor	Average	Good
10	5	3
7	8	5
8	4	2
11	6	3
5	5	4

- a. Construct the design matrix for these data.
 - b. Use any standard regression program to calculate a least-squares analysis of variance.
 - c. Run the analysis of variance in the traditional manner and compare your answers.
- 16.2 For the data in Exercise 16.1,
- a. Calculate treatment effects and show that the regression model reproduces these treatment effects.
 - b. Demonstrate that R^2 for the regression model is equal to η^2 for the analysis of variance.
- 16.3 Taking the data from Exercise 16.1, add the scores 5 and 8 to the Average group and the scores 2, 3, 3, and 5 to the Good group. Rerun the analysis for Exercise 16.1 using the more complete data.
- 16.4 Rerun the analysis of Exercise 16.2 for the amended data from Exercise 16.3.
- 16.5 A psychologist was concerned with the relationship between Gender, Socioeconomic Status (SES), and perceived Locus of Control. She took eight adults (age = 25 to 30 years) in each Gender–SES combination and administered a scale dealing with Locus of Control (a high score indicates that the individual feels in control of his or her everyday life).

	SES		
	Low	Average	High
Male	10	16	18
	12	12	14
	8	19	17
	14	17	13
	10	15	19
	16	11	15
	15	14	22
	13	10	20
Female	8	14	12
	10	10	18
	7	13	14
	9	9	21
	12	17	19
	5	15	17
	8	12	13
	7	8	16

© Cengage Learning 2013

- a. Run a traditional analysis of variance on these data.
 - b. The following sums of squares have been computed on the data using the appropriate design matrix (α = Gender, β = SES).
- $$SS_Y = 777.6667 \quad SS_{\text{reg}(\alpha, \beta, \alpha\beta)} = 422.6667$$
- $$SS_{\text{reg}(\alpha, \beta)} = 404.0000 \quad SS_{\text{reg}(\beta, \alpha\beta)} = 357.333$$
- $$SS_{\text{reg}(\alpha, \alpha\beta)} = 84.000$$
- Compute the summary table for the analysis of variance using these sums of squares.
- 16.6 Using the SES portion of the design matrix as our predictor, we find that $SS_{\text{reg}(\beta)} = 338.6667$.
- a. Why is this value the same as SS_{SES} in the answer to Exercise 16.5?
 - b. Will this be the case in all analyses of variance?
- 16.7 When we take the data in Exercise 16.5 and delete the last two low-SES males, the last three average-SES males, and the last two high-SES females, we obtain the following sums of squares:

$$\begin{aligned}
 SS_Y &= 750.1951 & SS_{\text{reg}(\alpha, \beta, \alpha\beta)} &= 458.7285 \\
 SS_{\text{reg}(\alpha, \beta)} &= 437.6338 & SS_{\text{reg}(\beta, \alpha\beta)} &= 398.7135 \\
 SS_{\text{reg}(\alpha, \alpha\beta)} &= 112.3392 & SS_{\text{reg}(\alpha)} &= 95.4511 \\
 & & SS_{\text{reg}(\beta)} &= 379.3325 \\
 & & SS_{\text{reg}(\alpha\beta)} &= 15.8132
 \end{aligned}$$

Compute the analysis of variance using these sums of squares.

- 16.8 Using only the SES predictors for the data in Exercise 16.7, we find $SS_{\text{reg}(\beta)} = 379.3325$. Why is this not the same as SS_{SES} in Exercise 16.7?
- 16.9 For the data in Exercise 16.5, the complete model is $1.1667A_1 - 3.1667B_1 - 0.1667B_2 + 0.8333AB_{11} - 0.1667AB_{12} + 13.4167$
- Show that this model reproduces the treatment and interaction effects as calculated by the method shown in Table 16.2.
- 16.10 For the data in Exercise 16.7, the complete model is $1.2306A_1 - 3.7167B_1 - 0.3500B_2 + 0.4778AB_{11} - 0.5444AB_{12} + 13.6750$
- Show that this model reproduces the treatment and interaction effects as calculated in Table 16.3.
- 16.11 Using the following data, demonstrate that Method III (the method advocated in this chapter) really deals with unweighted means.

	B_1	B_2	
	5	11	
	3	9	
A_1		14	
		6	
		11	
		9	
	10	6	
A_2	11	2	
	12		
	7		

© Cengage Learning 2013

- 16.12 Draw a Venn diagram representing the sums of squares in Exercise 16.5.
- 16.13 Draw a Venn diagram representing the sums of squares in Exercise 16.7.
- 16.14 If you have access to SAS, use that program to analyze the data in Exercise 16.7. Add /SS1 SS2 SS3 SS4 to the end of your Model command and show that
- Type I sums of squares adjust each term in the model only for those that come earlier in the model statement.
 - Type II sums of squares adjust main effects only for other main effect variables, while adjusting the interaction for each of the main effects.
 - Type III sums of squares adjust each term for all other terms in the model.
 - Type IV sums of squares in this case are equal to the Type II sums of squares.
- 16.15 In studying the energy consumption of families, we have broken them into three groups. Group 1 consists of those who have enrolled in a time-of-day electrical-rate system (the charge per kilowatt-hour of electricity is higher during peak demand times of the day). Group 2 is made up of those who inquired into such a system but did not use it. Group 3 represents those who have shown no interest in the system. We record the amount of the electrical bill per month for each household as our dependent variable (Y). As a covariate,

we take the electrical bill for that household for the same month last year (C). The data follow:

Group 1		Group 2		Group 3	
Y	C	Y	C	Y	C
58	75	60	70	75	80
25	40	30	25	60	55
50	68	55	65	70	73
40	62	50	50	65	61
55	67	45	55	55	65

© Cengage Learning 2013

- Set up the design matrix.
 - Run the analysis of covariance.
- 16.16 To refine the experiment described in Exercise 16.15, a psychologist added an additional set of households to each group. This group had a special meter installed to show them exactly how fast their electric bill was increasing. (The amount-to-date was displayed on the meter.) The data follow; the nonmetered data are the same as those in Exercise 16.15.

	Y	C	Y	C	Y	C
Nonmetered	58	75	60	70	75	80
	25	40	30	25	60	55
	50	68	55	65	70	73
	40	62	50	50	65	61
	55	67	45	55	55	65
Metered	25	42	40	55	55	56
	38	64	47	52	62	74
	46	70	56	68	57	60
	50	67	28	30	50	68
	55	75	55	72	70	76

© Cengage Learning 2013

- Run the analysis of covariance on these data—after first checking the assumption of homogeneity of regression.
 - Draw the appropriate conclusions.
- 16.17 Compute the adjusted means for the data in Exercise 16.16.
- 16.18 Compute the energy savings per household for the data in Exercise 16.16 by subtracting this year's bill from last year's bill. Then run an analysis of variance on the savings scores and compare that to the analysis of covariance.
- 16.19 Klemchuk, Bond, and Howell (1990) examined role taking in children. Children were administered a battery of role-taking tasks. They were classified as being in daycare or not being in daycare, and as ages 2–3 or ages 4–5. The hypothesis was that children with daycare experience would perform better on role-taking tasks. The data are available at the book's Web site as Ex16-19.dat. Run the appropriate analysis.

Computer Exercises

- 16.20 Use the data set named in Epinuneq.dat on the Web site to examine the results of the study by Introini-Collison and McGaugh (1986) described prior to Exercises 11.29–11.31. Using any statistical package, run a two-way analysis of variance with unequal sample sizes. What would you conclude from this analysis?

- 16.21 Use the data from Mireault and Bond (1992) in the file named Mireault.dat referred to in Exercise 7.6 to run a two-way analysis of variance on the Global Symptom Index T score (GSIT) using Gender and Group as independent variables. Plot out the cell means and interpret the results.
- 16.22 Using the same data as in Exercise 16.21, run an analysis of covariance instead, using year in college (YearColl) as the covariate.
- Why would we want to consider YearColl as a covariate?
 - How would you interpret the results?
- 16.23 In Exercise 16.22 we used YearColl as the covariate. Run an analysis of variance on YearColl, using Gender and Group as the independent variables. What does this tell us that is relevant to the preceding analysis of covariance?
- 16.24 Everitt reported data on a study of three treatments for anorexia in young girls. One treatment was cognitive behavior therapy, a second was a control condition with no therapy, and a third was a family therapy condition. These are the same data we examined in Chapter 14. The data follow and are available on the Web site.

Group	Pretest	Posttest	Gain	Group	Pretest	Posttest	Gain
1	80.5	82.2	1.7	2	80.7	80.2	-.5
1	84.9	85.6	.7	2	89.4	80.1	-9.3
1	81.5	81.4	-.1	2	91.8	86.4	-5.4
1	82.6	81.9	-.7	2	74.0	86.3	12.3
1	79.9	76.4	-3.5	2	78.1	76.1	-2.0
1	88.7	103.6	14.9	2	88.3	78.1	-10.2
1	94.9	98.4	3.5	2	87.3	75.1	-12.2
1	76.3	93.4	17.1	2	75.1	86.7	11.6
1	81.0	73.4	-7.6	2	80.6	73.5	-7.1
1	80.5	82.1	1.6	2	78.4	84.6	6.2
1	85.0	96.7	11.7	2	77.6	77.4	-0.2
1	89.2	95.3	6.1	2	88.7	79.5	-9.2
1	81.3	82.4	1.1	2	81.3	89.6	8.3
1	76.5	72.5	-4.0	2	78.1	81.4	3.3
1	70.0	90.9	20.9	2	70.5	81.8	11.3
1	80.4	71.3	-9.1	2	77.3	77.3	0.0
1	83.3	85.4	2.1	2	85.2	84.2	-1.0
1	83.0	81.6	-1.4	2	86.0	75.4	-10.6
1	87.7	89.1	1.4	2	84.1	79.5	-4.6
1	84.2	83.9	-.3	2	79.7	73.0	-6.7
1	86.4	82.7	-3.7	2	85.5	88.3	2.8
1	76.5	75.7	-.8	2	84.4	84.7	0.3
1	80.2	82.6	2.4	2	79.6	81.4	1.8
1	87.8	100.4	12.6	2	77.5	81.2	3.7
1	83.3	85.2	1.9	2	72.3	88.2	15.9
1	79.7	83.6	3.9	2	89.0	78.8	-10.2
1	84.5	84.6	.1	3	83.8	95.2	11.4
1	80.8	96.2	15.4	3	83.3	94.3	11.0
1	87.4	86.7	-.7	3	86.0	91.5	5.5

3	82.5	91.9	9.4	3	81.6	77.8	-3.8
3	86.7	100.3	13.6	3	82.1	95.5	13.4
3	79.6	76.7	-2.9	3	77.6	90.7	13.1
3	76.9	76.8	-0.1	3	83.5	92.5	9.0
3	94.2	101.6	7.4	3	89.9	93.8	3.9
3	73.4	94.9	21.5	3	86.0	91.7	5.7
3	80.5	75.2	-5.3	3	87.3	98.0	10.7

- Run an analysis of variance on group differences in Gain scores. (You may have already done this in Chapter 14.)
 - Now run the analysis on Posttest scores, ignoring Pretest scores.
 - Finally, run the analysis on Posttest scores using Pretest as the covariate.
 - How do these three answers relate to one another, and what do they show about the differences and similarities between analysis of covariance and the treatment of gain (or change) scores?
 - Calculate η^2 on Groups for the analysis of covariance.
 - Calculate d for the contrast on the two therapy groups (ignoring the control group) using adjusted means.
- 16.25 Write up the results of Everitt's experiment, including effect sizes.

Discussion Questions

- 16.26 I initially thought of creating an analysis of variance example from the example in Chapter 14, Section 14.7. I could have used Sex and Group as the independent variables, posttest scores as the dependent variable, and pretest scores as the covariate (ignoring FU6 and FU12 entirely). This would have made a very bad example for the analysis of covariance. Why would that be? Is there any way in which we might be able to salvage the study as an analysis of covariance example?
- 16.27 I said that in any experiment where we have pretest and posttest scores we could either look at the difference scores (compared across groups) or use the pretest as a covariate. These two analyses will be equivalent only when the slope relating posttest to pretest scores is 1.00. How likely do you think it is that such a condition would be met (or at least approximated)? What does $b = 1.00$ actually imply?
- 16.28 Make up or find an example with respect to Exercise 16.25 where the slope is not nearly 1.0. Analyze it using both the analysis of covariance and a t test on difference scores. Do either of these analyses make sense?