Factorial ANOVA

Models with Multiple Categorical Predictors and Product Terms

9

In the previous chapter we developed models with one categorical predictor. In this chapter we expand our consideration to models with two or more categorical predictor variables. Our reasons for wanting to include more than one categorical variable as a predictor in our model are the same as those that motivated us to expand from simple regression models with one predictor to multiple regression models with two or more predictors in Chapter 6. Models in which predictions are conditional on two or more categorical variables may be required by our data and, more importantly, by the underlying process that generates the data. Just as in multiple regression, controlling for one categorical variable by including it in the model often allows us to have a better look at the effects of other categorical variables. Also, as in multiple regression, we are often interested in modeling the joint effect of two or more categorical variables. With categorical variables we will be especially interested in whether the effect of a given categorical variable depends on the levels of the other categorical variables; that is, we are interested in whether or not there is an interaction between the categorical variables analogous to the interactions of continuous variables in multiple regression considered in Chapter 7.

The generalization of models with one categorical predictor (one-way ANOVA of the previous chapter) to models with two categorical predictors (two-way ANOVA) and to models with more than two categorical predictors (q-way factorial ANOVA, where q is the number of categorical predictors) is straightforward. As we shall demonstrate, classical analysis of variance with two or more categorical predictors is nothing more than a simple one-way ANOVA with a specific, clever set of contrast codes. In other words, the only new thing to learn is how to generate the appropriate set of contrast codes; fitting the model and testing hypotheses are *exactly* the same as for one-way ANOVA in Chapter 8.

FACTORIAL ANOVA AS ONE-WAY ANOVA

We begin by considering the hypothetical dataset in Figure 9.1. In this hypothetical experiment clinically depressed patients either receive psychotherapy (treatment) or not (control) and receive one of three drugs (A, B, or placebo). After six months each patient completes a mood questionnaire on which higher scores mean improved mood or decreased depression.

FIGURE 9.1 Hypothetical data (mood scores) in a Drug (3) by Psychotherapy (2) experimental design

	Psychotherapy					
Drug	*Treatment*			*Control*		
A	31	31	34	17	15	19
B	25	25	28	23	18	16
Placebo	17	18	16	9	10	8

FIGURE 9.2 Hypothetical data of Figure 9.1 arrayed as a one-way design

	Group 1 (A,T)	*Group 2 (B,T)*	*Group 3 (P,T)*	*Group 4 (A,C)*	*Group 5 (B,C)*	*Group 6 (P,C)*
	31	25	17	17	23	9
	31	25	18	15	18	10
	34	28	16	19	16	8
Mean	32	26	17	17	19	9

		Group						
Contrast codes		*Predictor*	*1*	*2*	*3*	*4*	*5*	*6*
λ_1	1 vs. 2, 3, 4, 5, 6	Z_1	5	−1	−1	−1	−1	−1
λ_2	2 vs. 3, 4, 5, 6	Z_2	0	4	−1	−1	−1	−1
λ_3	3 vs. 4, 5, 6	Z_3	0	0	3	−1	−1	−1
λ_4	4 vs. 5, 6	Z_4	0	0	0	2	−1	−1
λ_5	5 vs. 6	Z_5	0	0	0	0	1	−1

The experimental design depicted in Figure 9.1 is known as a *factorial* design because every level of one categorical variable or factor is combined with every level of the other categorical variable or factor. There are three levels of the Drug variable and two levels of the Psychotherapy variable, so there are a total of $3 \times 2 = 6$ different combinations, each defining a group or cell in the design. We often refer to this as a 3×2 *design*. In this hypothetical study, from a total of 18 patients, three patients are randomly assigned to each of the six groups.

Although it is natural to display these data in a table with three rows and two columns as in Figure 9.1, we can also display the data as a one-way layout in terms of the six groups or cells as in Figure 9.2. Seeing the data in the one-way layout makes it clear that we can use any of the sets of contrast codes developed in the previous chapter to analyze these data. To illustrate this, we will first do an analysis with one-way contrast codes that are statistically correct but that do not ask questions that are usually interesting. In this first set of contrast codes we form the first contrast λ_1 by comparing Group 1 (Drug A combined with psychotherapy Treatment: A,T) with all the other groups; the corresponding contrast-coded predictor is Z_1.[1] The contrast λ_2 compares Group 2 against all the remaining groups except Group 1, and so on. These contrast codes are also displayed in Figure 9.2. We can verify that these codes are orthogonal by checking that the sum of each set of crossproducts is zero. For example:

$$\sum_{k=1}^{6} \lambda_{1k}\lambda_{2k} = 5(0) - 1(4) - 1(-1) - 1(-1) - 1(-1) - 1(-1)$$
$$= 0 - 4 + 1 + 1 + 1 + 1$$
$$= 0$$

We can regress Y_i, the mood scores, on the predictor variables Z_1, Z_2, . . ., Z_5 using a standard multiple regression program or, equivalently, we could use the one-way ANOVA formulas from Chapter 8. Figure 9.3 shows the actual data matrix we would use to regress Y_i on Z_1, Z_2, . . ., Z_5. Note that the mean for each contrast-coded predictor is zero because there are equal numbers of observations in each group.

We can ask whether all six group means are equal by comparing these models:

MODEL A: $Y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \beta_5 Z_{5i} + \varepsilon_i$

MODEL C: $Y_i = \beta_0 + \varepsilon_i$

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

Regressing the mood scores on the five contrast-coded predictors yields the analysis in Figure 9.4. The two estimated models are:

MODEL A: $\hat{Y}_i = 20 + 2.4Z_{1i} + 2.1Z_{2i} + 0.5Z_{3i} + 1Z_{4i} + 5Z_{5i}$

MODEL C: $Y_i = 20$

FIGURE 9.3 Data matrix for analyzing mood scores using multiple regression and one-way coded predictors

Group		*Mood scores*	*Predictors*				
Drug	*Psychotherapy*	*Y*	Z_1	Z_2	Z_3	Z_4	Z_5
A	T	31	5	0	0	0	0
A	T	31	5	0	0	0	0
A	T	34	5	0	0	0	0
B	T	25	−1	4	0	0	0
B	T	25	−1	4	0	0	0
B	T	28	−1	4	0	0	0
P	T	17	−1	−1	3	0	0
P	T	18	−1	−1	3	0	0
P	T	16	−1	−1	3	0	0
A	C	17	−1	−1	−1	2	0
A	C	15	−1	−1	−1	2	0
A	C	19	−1	−1	−1	2	0
B	C	23	−1	−1	−1	−1	1
B	C	18	−1	−1	−1	−1	1
B	C	16	−1	−1	−1	−1	1
P	C	9	−1	−1	−1	−1	−1
P	C	10	−1	−1	−1	−1	−1
P	C	8	−1	−1	−1	−1	−1
Mean		20	0	0	0	0	0

FIGURE 9.4 ANOVA table of the hypothetical mood scores using the model: $Y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \beta_5 Z_{5i} + \varepsilon_i$

Source	b_j	*SS*	*df*	*MS*	*F*	*p*	*PRE*
Between groups		960.0	5	192.0	46.1	.0001	.95
Z_1	2.4	518.4	1	518.4	124.4	.0001	.91
Z_2	2.1	264.6	1	264.6	63.5	.0001	.84
Z_3	0.5	9.0	1	9.0	2.2	.17	.15
Z_4	1.0	18.0	1	18.0	4.3	.06	.26
Z_5	5.0	150.0	1	150.0	36.0	.0001	.75
Within groups (MSE)		50.0	12	4.2			
Total		1010.0	17				

The augmented model using all five predictors reduces the error of the compact model using the mean by 960/1010 = .9505. We calculate F in the usual way:

$$F_{5,12} = \frac{.9505/5}{(1 - .9505)/12} = 46.1$$

This omnibus tests rejects Model C in favor of Model A because the large value of PRE (.95) is statistically surprising ($F_{5,12} = 46.1$, $p < .0001$), which implies, as it did in one-way ANOVA, that the means for the six groups are not all equal to one another. That is, the six groups cannot be adequately represented by the overall mean.

But which groups are different from each other? The test of the Z_1 contrast indicates that the estimate $b_1 = 2.4$ is significantly different from zero and it reduces 91% of the error remaining after all the other codes are in the equation.[2] Remember that this test is obtained by comparing an augmented model including all the Z values with a compact model including all the Z values except Z_1. In particular:

MODEL A: $Y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \beta_5 Z_{5i} + \varepsilon_i$

MODEL C: $Y_i = \beta_0 + \quad \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \beta_5 Z_{5i} + \varepsilon_i$

H_0: $\beta_1 = 0$

In this case, we can reject Model C in favor of Model A and therefore reject the null hypothesis that $\beta_1 = 0$. The corresponding contrast λ_1 codes the comparison between Group 1 (Drug A with psychotherapy treatment) and the average of all the other groups; thus, we can conclude that Group 1 is significantly different from the average of all other cells in the study.

We interpret the value of the coefficient the same as we always have. Specifically, the coefficient $b_1 = 2.4$ means that for a one-unit increase in Z_1 the predicted mood score increases, on average, by 2.4 points. To verify this, reexamine Figure 9.2 to see that the value of Z_1 for all the other groups is –1 but for Group 1 the value of Z_1 is 5. Thus, there is a change of six units on Z_1 between the two comparison groups. Model A therefore predicts the mean for Group 1 to be $6 \times 2.4 = 14.4$ higher than the average in all the other groups. This is indeed the case: the mean for Group 1 is 32 and the mean for all the other groups combined is (26 + 17 + 17 + 19 + 9)/5 = 17.6, thus 32 – 17.6 = 14.4. But why is Group 1 better, on average, than all the other groups? Is it because of Drug A? Or is it because patients receiving psychotherapy score higher regardless of which

drug they receive? Or is it because Drug A is especially effective for patients who also receive psychotherapy treatment? The contrast λ_1 cannot tell us. It simply indicates that there is a difference.

We can similarly interpret the results for the contrast-coded predictor Z_2. That the estimate $b_2 = 2.1$ is significantly different from zero indicates that Group 2 (Drug B administered to those receiving psychotherapy treatment) is different from the average of all the subsequent groups (i.e., excluding Group 1). The difference in means equals $5 \times 2.1 = 10.5$. But again we do not know whether the higher average mood scores of Group 2 are due to Drug B or psychotherapy treatment or their combination.

The other contrast-coded predictor whose coefficient is significantly different from zero (using $\alpha = .05$) is Z_5; the corresponding contrast λ_5 compares Groups 5 and 6 or the difference between taking Drug B versus placebo for those in the Control condition. We know in this case that the higher average score for Group 5 ($2 \times 5 = 10$) is due to Drug B relative to the placebo because all patients in this comparison are in the control group not receiving psychotherapy. But is Drug B also better than the placebo for those patients who do receive psychotherapy treatment? None of the Z contrasts help to answer that question.

A BETTER SET OF CONTRAST CODES

Even though the one-way analysis of variance using the Z contrast-coded predictors is statistically correct, it has failed to answer important questions we want to ask of the data. One solution would be to use the multiple comparison procedures of Chapter 8 to address some of the unanswered questions. However, a much more efficient strategy is to begin with a set of contrast codes that do ask many of the questions we would naturally want to consider. Given the two-way layout of the data (as in Figure 9.1), there are several sets of codes that are more natural than the Z codes. "Two-way" ANOVA is nothing more than one-way ANOVA using one of these natural sets of codes.

To develop a more natural set of codes, we begin by considering each of the two categorical variables separately. The strategy is to develop contrast codes identical to the ones that we would use if each categorical variable were considered alone in its own one-way ANOVA. That is, our first step is to code each categorical variable as if the other one did not exist. For the Drug variable an interesting question is whether the drugs (either A or B), on average, do better than the placebo. The contrast code λ_1 (and its corresponding contrast-coded predictor X_1) in Figure 9.5 makes precisely that comparison with the pattern (1, 1, –2). This pattern is repeated for each level of the other categorical variable (in this case, Psychotherapy). The contrast code λ_2 (and its corresponding contrast-coded predictor X_2) then asks whether there is any difference between Drugs A and B. We can verify that λ_1 and λ_2 are orthogonal by checking whether the sum of their crossproducts equals zero. That is:

$$\begin{aligned}\sum_{k=1}^{6} \lambda_{1k}\lambda_{2k} &= 1(1) + 1(-1) - 2(0) + 1(1) + 1(-1) - 2(0) \\ &= 1 - 1 + 0 + 1 - 1 + 0 \\ &= 0\end{aligned}$$

FIGURE 9.5 Hypothetical data of Figure 9.1 coded for two-way ANOVA

	Group 1 (A,T)	*Group 2 (B,T)*	*Group 3 (P,T)*	*Group 4 (A,C)*	*Group 5 (B,C)*	*Group 6 (P,C)*
	31	25	17	17	23	9
	31	25	18	15	18	10
	34	28	16	19	16	8
Mean	32	26	17	17	19	9

			Group					
Contrast codes		*Predictor*	*1*	*2*	*3*	*4*	*5*	*6*
λ_1	Drugs vs. Placebo	X_1	1	1	–2	1	1	–2
λ_2	Drug A vs. Drug B	X_2	1	–1	0	1	–1	0
λ_3	Treatment vs. Control	X_3	1	1	1	–1	–1	–1
λ_4	Interaction: $\lambda_1 \times \lambda_3$	X_4	1	1	–2	–1	–1	2
λ_5	Interaction: $\lambda_2 \times \lambda_3$	X_5	1	–1	0	–1	1	0

The orthogonality combined with the equal numbers of observations in each cell ensures that the coded predictors X_1 and X_2 are uncorrelated or not redundant. With three levels of the Drug categorical variable, we can only have two orthogonal codes so λ_1 and λ_2 are sufficient for the one-way analysis of that variable.

There are only two levels of the Psychotherapy categorical variable, so we need only one code. Hence, λ_3 codes the contrast between receiving psychotherapy treatment versus being in the control condition and provides the complete one-way ANOVA for that variable.

So far we have only three codes, λ_1, λ_2, and λ_3, for the two separate one-way analyses of the categorical variables, but five orthogonal contrast codes are required for the complete analysis of six groups. To generate the other two necessary codes, we simply multiply the contrast codes between the two categorical variables; that is, $\lambda_4 = \lambda_1 \times \lambda_3$ and $\lambda_5 = \lambda_2 \times \lambda_3$. These are the same kind of product terms we considered in Chapter 7 when we introduced the concept of interactions between variables. In the context of two-way ANOVA these product terms ask especially interesting questions. The code λ_4, the product of λ_1 and λ_3, asks whether the difference coded by λ_1 (Drugs vs. Placebo) depends on the level of psychotherapy (Treatment vs. Control). In other words, might there be one drug effect for those receiving psychotherapy treatment and a different drug effect for those in the control group. Similarly, λ_5, the product of λ_2 and λ_3, asks whether the comparison coded by λ_2 (Drug A vs. Drug B) depends on the level of Psychotherapy (Treatment vs. Control). The order of multiplication is arbitrary, so either statement could be reversed. That is, λ_4 asks whether the effectiveness of psychotherapy (the Treatment vs. Control difference) depends on whether the patient also received a drug or the placebo. We consider the interpretation of interactions like these in greater detail later.

Note that we do not form a code by multiplying $\lambda_1 \times \lambda_2$ because that would yield another code for just the Drug categorical variable that could not be orthogonal to the other two codes for that variable. However, the two codes formed from products, λ_4 and λ_5, are orthogonal to each other and to the other codes. You may want to compute some of the crossproducts to verify this claim. This gives a total of five orthogonal contrast codes, precisely the number we need for the analysis of six groups.

FIGURE 9.6 Data matrix for analyzing mood scores using multiple regression and two-way coded predictors

Group		*Mood scores*	*Predictors*				
Drug	*Psychotherapy*	*Y*	X_1	X_2	X_3	X_4	X_5
A	T	31	1	1	1	1	1
A	T	31	1	1	1	1	1
A	T	34	1	1	1	1	1
B	T	25	1	–1	1	1	–1
B	T	25	1	–1	1	1	–1
B	T	28	1	–1	1	1	–1
P	T	17	–2	0	1	–2	0
P	T	18	–2	0	1	–2	0
P	T	16	–2	0	1	–2	0
A	C	17	1	1	–1	–1	–1
A	C	15	1	1	–1	–1	–1
A	C	19	1	1	–1	–1	–1
B	C	23	1	–1	–1	–1	1
B	C	18	1	–1	–1	–1	1
B	C	16	1	–1	–1	–1	1
P	C	9	–2	0	–1	2	0
P	C	10	–2	0	–1	2	0
P	C	8	–2	0	–1	2	0
Mean		20	0	0	0	0	0

Figure 9.6 shows the data matrix that we can analyze with a multiple regression program to model mood scores in terms of the X variables. That is, we can regress Y_i on X_1 to X_5. Note again that the mean for each predictor is zero because we are using contrast codes and there are an equal number of observations for each group.

Figure 9.7 presents the results of the regression analysis using the X contrast-coded predictors. Note that the sum of squares for the augmented model including all the predictors, and the F and PRE for the omnibus test of the complete model are *exactly* the same as when we used the Z predictors. This must be the case because in each analysis we used a complete set of codes so that, as in one-way ANOVA, the predicted value $\hat{Y}$ for each group is the mean $\overline{Y}$ for that group. If the predictions of the two models are

FIGURE 9.7 ANOVA table of the hypothetical mood scores using the model: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

Source	b_j	*SS*	*df*	*MS*	*F*	*p*	*PRE*
Between groups		960	5	192.0	46.1	.0001	.95
Drug		453	2	226.5	53.9	.0001	.90
X_1	3.5	441	1	441	105.8	.0001	.90
X_2	1.0	12	1	12	2.9	.11	.19
Psychotherapy		450	1	450	108.0	.0001	.90
X_3	5.0	450	1	450	108.0	.0001	.90
Drug x Psychotherapy		57	2	28.5	6.8	.01	.53
X_4	0.5	9	1	9	2.2	.16	.15
X_5	2.0	48	1	48	11.5	.005	.49
Within groups (MSE)		50.0	12	4.2			
Total		1010.0	17				

the same, then the total error and the total error reduced must be the same. The only difference in the two analyses is how that total error reduction is divided into separate components. The X predictors ask different questions than the Z predictors, so the individual error reduction associated with each of the X predictors differs from that of the Z predictors, but the total error reduction must be the same.

There are nine PRE and F values in Figure 9.7. Each one corresponds to a comparison between a particular Model C and a Model A. To understand the meaning of each test, it is important to be precise about those models for each test. The omnibus test is reported in the row labeled "Between groups" and represents the comparison between the following two models:

MODEL A: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

MODEL C: $Y_i = \beta_0 + \varepsilon_i$

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

This corresponding null hypothesis, equivalent to assuming that all the group means are equal, is rejected by the large values of PRE and F.

The rows for each of the X_j compare a Model A that uses all the contrast-coded predictors to a Model C that includes all the predictors except X_j. For example, the PRE and F in the row labeled X_1 compare:

MODEL A: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

MODEL C: $Y_i = \beta_0 + \quad \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

H_0: $\beta_1 = 0$

We generated the predictors X_1 and X_2 as one-way codes for the Drug categorical variable for the data matrix in Figure 9.1. The row labeled "Drug" in Figure 9.7 reports the results from testing a model including all the contrast codes to one that omits *both* the drug codes X_1 and X_2. That is, we are comparing:

MODEL A: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

MODEL C: $Y_i = \beta_0 + \quad \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

H_0: $\beta_1 = \beta_2 = 0$

In other words, we are asking whether the predictions would suffer if we were to ignore the Drug categorical variable. The large PRE and F indicate that we cannot ignore which drug a patient received. In the traditional language of ANOVA, this is known as the test of the *main effect* of the Drug categorical variable. In this case, we would conclude that overall there is a statistically significant main effect for the Drug categorical variable. Note that because X_1 and X_2 are uncorrelated in this instance, their sums of squares add to produce the sum of squares attributable to both of them.

We prefer the single-degree-of-freedom tests of the focused comparisons X_1 and X_2 to this global test of the drug effect because rejection of the hypothesis $\beta_1 = \beta_2 = 0$ is ambiguous. We do not know which part of this multiple-degree-of-freedom hypothesis is at fault. Maybe the global hypothesis is rejected because $\beta_1 \neq 0$ or because $\beta_2 \neq 0$ or because both β_1 and β_2 are not equal to zero. We present the omnibus test for the Drug variable in Figure 9.7 not as a recommended practice, but only to show what hypothesis is being tested by the "main effect" in the traditional approach to analysis of variance.

Unfortunately, in the traditional approach usually *only* the omnibus tests in Figure 9.7 are presented in a source table. Doing so ignores readily available detailed information about the data revealed by the single-degree-of-freedom contrast codes.

We can similarly assess the main effect for the Psychotherapy variable. In this case there is only one code so that test is the same as the test of the null hypothesis that $\beta_3 = 0$. Finally, we can group together the two codes X_4 and X_5, which we constructed from products of other codes. Each of these two products involved one code for the Drug categorical variable and one code for the Psychotherapy categorical variables. So, in the traditional language of ANOVA, they are known together as the "Drug × Psychotherapy Interaction." Testing the Drug × Psychotherapy interaction is equivalent to comparing:

MODEL A: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$

MODEL C: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$

H_0: $\beta_4 = \beta_5 = 0$

A traditional ANOVA source table similar to the one in Figure 9.7 would contain rows only for Drug, Psychotherapy, Drug × Psychotherapy, Within groups (or Error), and Total. Such a source table, because it aggregates individual contrasts into more global effects, omits information that is readily available in the individual contrasts. In other words, the traditional ANOVA table fails to analyze the variance as much as is possible using the focused comparisons of contrast codes.

Any standard multiple regression program can produce the analysis in Figure 9.7. Equivalently, one can use the equations from Chapter 8 expressing the parameter estimates and their corresponding sums of squares as a function of the contrast codes and the cell means. As an illustration, we use the formulas to calculate the estimate b_1 and its associated SSR. The coefficient that estimates β_1 is:

$$b_1 = \frac{\sum_k \lambda_{1k} \bar{Y}_k}{\sum_k \lambda_k^2}$$

$$= \frac{1(32) + 1(26) - 2(17) + 1(17) + 1(19) - 2(9)}{1^2 + 1^2 + (-2)^2 + 1^2 + 1^2 + (-2)^2}$$

$$= \frac{42}{12} = 3.5$$

Then the sum of squared error reduced by including X_1 in the model along with all the other predictors is:

$$\mathrm{SSR}_{b_1} = \frac{\left(\sum_k \lambda_{1k} \bar{Y}_k\right)^2}{\sum_k \lambda_{1k}^2 / n_k}$$

$$= \frac{[1(32) + 1(26) - 2(17) + 1(17) + 1(19) - 2(9)]^2}{(1^2 + 1^2 + (-2)^2 + 1^2 + 1^2 + (-2)^2)/3}$$

$$= \frac{42^2}{12/3} = 441$$

Note that because two-way ANOVA is exactly the same as a one-way ANOVA with a clever set of codes, the formulas and all other details from Chapter 8 apply unaltered to two-way ANOVA. Thus, in terms of parameter estimates and statistical tests there is nothing new that needs to be learned to do two-way ANOVA—it is *exactly* the same as one-way ANOVA.

INTERPRETATION OF COEFFICIENTS

The one new difficulty introduced by the generalization of one-way ANOVA to two-way and higher ANOVA is the interpretation of the coefficients. But it is simply an interpretation problem and not a difficulty involving statistical procedures. And the strategy for interpreting coefficients in two-way and higher ANOVA is the same as the strategies we have developed in Chapters 7 and 8. The new difficulty is interpretation of the coefficients for the predictor variables that code the interactions—the products between one-way codes. Newcomers to ANOVA, whether using our approach or more traditional approaches, often have initial difficulty understanding the concept of an interaction. We therefore devote considerable attention in this section to the interpretation of interactions. We do so by considering two different approaches to interpreting the coefficients in a model with two-way codes such as those in Figure 9.5. Those two approaches are (a) re-expression of the codes in terms of the cell means, as was done in Chapter 8, and (b) interpretation of product predictors as changes in simple slopes, as was done in Chapter 7. In each approach we also consider the interpretation of the coefficients for predictors not involving products because interactions are best understood in comparison.

Interpretation in Terms of Cell Means

The interpretation of the parameter estimates associated with the X predictors proceeds in the same manner as for any set of predictors in the one-way analyses of Chapter 8. The surprisingly large values, relative to Model C being correct, of $F_{1,12} = 105.8$ and PRE = .95 for X_1 reject the null hypothesis that $\beta_1 = 0$. To see exactly what the rejection of this null hypothesis means, we express β_1 in terms of the contrast code using the equation from Chapter 8 that expresses the parameter estimate as a function of the codes and the cell means. That is, an equivalent statement of the null hypothesis $\beta_1 = 0$ is:

$$\beta_1 = \frac{\sum_k \lambda_{1k}\mu_k}{\sum_k \lambda_{1k}^2} = \frac{\mu_{AT} + \mu_{BT} - 2\mu_{PT} + \mu_{AC} + \mu_{BC} - 2\mu_{PC}}{12} = 0$$

where μ_{AT} represents the true but unknown mean for the group that received Drug A and participated in Psychotherapy treatment, etc. Doing a little algebra (multiplying by 12 and moving the terms for the placebo conditions to the right side of the equation) yields the following equivalent statement of the hypothesis:

$$\mu_{AT} + \mu_{AC} + \mu_{BT} + \mu_{BC} = 2(\mu_{PT} + \mu_{PC})$$

Dividing each side of this equation by 2 gives:

$$\frac{\mu_{AT} + \mu_{AC}}{2} + \frac{\mu_{BT} + \mu_{BC}}{2} = 2\left(\frac{\mu_{PT} + \mu_{PC}}{2}\right)$$

If we let $\mu_{A.} = (\mu_{AT} + \mu_{AC})/2$, which is the mean of the Drug A conditions averaged (indicated by the dot) across both the Treatment and Control levels of the Psychotherapy variable, $\mu_{B.} = (\mu_{BT} + \mu_{BC})/2$ and $\mu_{P.} = (\mu_{PT} + \mu_{PC})/2$, then the null hypothesis reduces to:

$$H_0\text{: } \mu_{A.} + \mu_{B.} = 2\mu_{P.}$$

Dividing each side of this equation by 2 gives:

$$\frac{\mu_{A.} + \mu_{B.}}{2} = \mu_{P.}$$

That is, concluding that $\beta_1 \neq 0$ is equivalent to concluding that the average of the drug conditions combined is not equal to the average for the placebo groups, ignoring whether the patients received psychotherapy treatment or not. Thus, unlike the ambiguous conclusions using the Z contrast codes, rejection of the null hypothesis for β_1 clearly implies that the mood scores for the drug groups, on average, differed from the mood scores of the placebo groups. With experience it is usually easy to go directly from the contrast codes to a statement of the null hypothesis without doing the formal derivation as above. For example, the (1, –1, 0) code for X_2 compares the Drug A groups with the Drug B groups so that the null hypothesis $\beta_2 = 0$ is equivalent to the following null hypothesis:

$$H_0\text{: } \mu_{A.} - \mu_{B.} = 0 \text{ or } H_0\text{: } \mu_{A.} = \mu_{B.}$$

The direction and magnitude of a difference revealed by rejecting the null hypothesis is given by b_j, the estimate of β_j. For X_1, this estimate is $b_1 = 3.5$. As with all regression coefficients, this means that our model prediction $\hat{Y}_i$ increases, on average, by 3.5 units for each unit increase in X_1. The change on X_1 from the placebo groups ($X_1 = –2$) to the drug groups ($X_1 = 1$) is a change of 3 units, so the predicted difference between the placebo and drug groups is $3 \times 3.5 = 10.5$. Indeed, the average of the four drug groups, $(32 + 26 + 17 + 19)/4 = 23.5$, exceeds by 10.5 the average of the two placebo groups, $(17 + 9)/2 = 13$. Figure 9.8(a) depicts graphically the means compared by the first contrast.

We cannot conclude in this analysis that the estimate $b_2 = 1.0$ for β_2 is reliably different from zero. Nevertheless, it is useful to interpret it to see which means are not significantly different from one another. On X_2 there is a two-unit difference between those taking Drug B and those taking Drug A, regardless of whether they were also receiving psychotherapy treatment or not. Thus, the mood scores of those taking Drug A are $2 \times 1 = 2$ points higher than the mood scores of those taking Drug B, but this difference is not statistically significant. Figure 9.8(b) depicts this comparison graphically.

The coefficient $b_3 = 5.0$ is significantly different from zero. The predictor X_3 codes the difference between the two psychotherapy groups, so we can conclude that, on average, there is a significant difference between the mood scores of those receiving treatment compared to those in the control group. That is, we can reject the null hypothesis:

$$H_0\colon \mu_{.T} = \mu_{.C}$$

where the dots in the subscript indicate that we averaged across levels of drug. To be specific about the amount of difference in mood scores between receiving psychotherapy versus not, we need to interpret the meaning of the estimate $b_3 = 5.0$ as before. There is a two-unit difference on X_3 between receiving Treatment ($X_3 = 1$) and receiving the Control ($X_3 = -1$). Thus, the predicted average difference between the two groups is $2 \times 5 = 10$. Indeed, the average of the three groups receiving Treatment $(32 + 26 + 17)/3 = 25$, exceeds by 10 the average of the three Control groups $(17 + 19 + 9)/3 = 15$. Figure 9.8(c) depicts these means and their comparison.

Even though the coefficient for $b_4 = 0.5$ is not significantly different from zero, its interpretation provides a good introduction to understanding interactions: products of contrast codes. We begin by considering the full expression for β_4 in terms of the unknown means:

$$\beta_4 = \frac{\sum_k \lambda_{4k}\mu_k}{\sum_k \lambda_{4k}^2} = \frac{\mu_{AT} + \mu_{BT} - 2\mu_{PT} - \mu_{AC} - \mu_{BC} + 2\mu_{PC}}{12} = 0$$

Except for the signs on the last three terms, this is very similar to the expression for β_1 above. Putting the T and C terms on opposite sides of the equation gives:

$$\frac{\mu_{AT} + \mu_{BT} - 2\mu_{PT}}{12} = \frac{\mu_{AC} + \mu_{BC} - 2\mu_{PC}}{12}$$

Dividing the numerators and denominators on both sides by 2 yields:

$$\frac{1}{6}\left[\frac{\mu_{AT} + \mu_{BT}}{2} - \mu_{PT}\right] = \frac{1}{6}\left[\frac{\mu_{AC} + \mu_{BC}}{2} - \mu_{PC}\right]$$

Note that the terms in each of the two brackets are identical to the question asked by the first code: Is there a difference in mood scores between those taking either drug compared to those taking the placebo? The first coded predictor X_1 asked whether there was a difference regardless of whether the patient was also receiving psychotherapy treatment or not. The interaction code $X_4 = X_1 \times X_3$ asks whether that difference, whatever it might be, is the same for those receiving psychotherapy treatment:

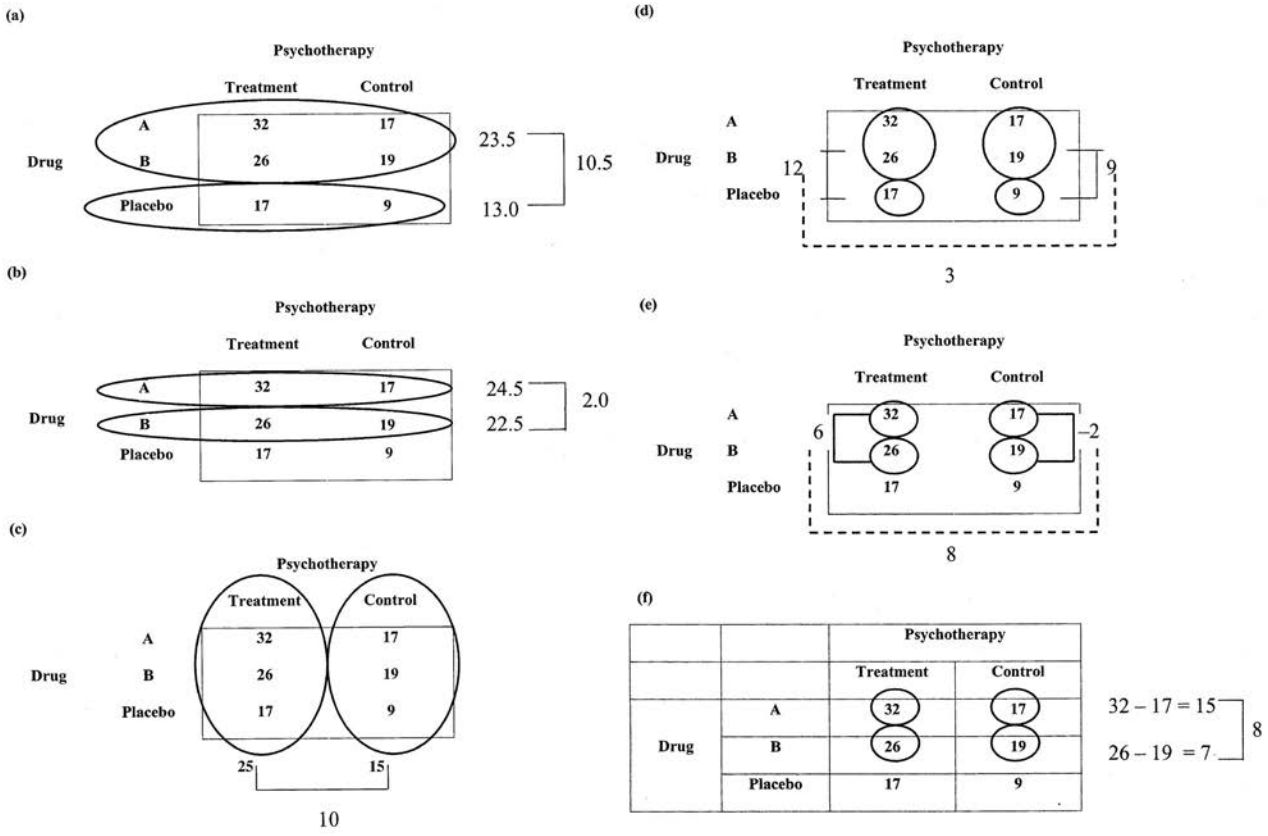
$$\frac{\mu_{AT} + \mu_{BT}}{2} - \mu_{PT}$$

as it is for those in the control condition:

$$\frac{\mu_{AC} + \mu_{BC}}{2} - \mu_{PC}$$

In other words, the null hypothesis for this and all interactions is whether two differences are equal. Different differences imply an interaction. Figure 9.8(d) depicts the means and differences involved in this comparison. For the Treatment groups, the means for those receiving a drug (either A or B) are 32 and 26 (the top-left oval) and their mean

FIGURE 9.8 Means compared by contrast codes



of 29 differs from the mean of the Treatment group taking the placebo (17) by 12 points. For the Control group, the mean of those taking either Drug A or Drug B (the top-right oval) is (17 + 19)/2 = 18, which exceeds the mean of those taking the placebo (9) by 9 points. The statistical test for β_4 is therefore equivalent to asking whether these two differences—12 for the Treatment group and 9 for the Control group—are significantly different from each other. Or, equivalently, whether the difference of the differences (12 – 9 = 3) is significantly different from zero. In this case, because we could not reject the null hypothesis that $\beta_4 = 0$, we also cannot reject the null hypothesis that the Drug versus Placebo difference is the same for the Treatment group as it is for the Control group. Note that the algebra above shows that the estimated coefficient $b_4 = 0.5$ equals one-sixth of the difference of the differences. Indeed, $0.5 \times 6 = 3$, which is the difference of the differences depicted in Figure 9.8(d).

The algebra to determine which differences are being compared in an interaction may seem complicated at first, but again with experience it is usually easy to jump from the code to a specification of the differences being compared. For example, $X_5 = X_2 \times X_3$ so the null hypothesis:

$$H_0: \beta_5 = 0$$

is equivalent to testing whether the difference coded by X_2—the difference between those taking Drug A versus Drug B—depends on the level coded by X_3, whether the patient is in the Treatment or Control groups. Thus, the equivalent null hypothesis is:

$$H_0: \mu_{AT} - \mu_{BT} = \mu_{AC} - \mu_{BC}$$

In words, the null hypothesis is that the difference between Drugs A and B is the same for those receiving Psychotherapy treatment as it is for those in the Control group. Knowing that $b_5 = 2$ is reliably different from zero allows us to conclude that the drug effect is different for those receiving treatment than for those in the control group. Figure 9.8(e) depicts this difference and shows its magnitude. For those receiving Treatment, the difference in mood scores for those taking Drug A compared to those taking Drug B equals 32 – 26 = 6, whereas for those in the Control group the difference is in the other direction: 17 – 19 = –2. The difference of the differences is 8 and is significantly different from zero. Note that although it appears that Drug A is better for those receiving Treatment but Drug B is better for those in the Control group, we have no statistical test of these two individual differences. In other words, we do not know whether the drug difference of 6 for the Treatment group or the difference of –2 for the Control group differs significantly *from zero*. We simply know that 6 is significantly different from –2. We return later to address the specific question of whether there are significant differences between the drugs within each treatment condition.

In Chapter 7 we noted that we can always interpret an interaction with either variable in the product as the focal variable. The same is true for products of categorical predictors. Thus, we can focus on whether the Treatment versus Control difference is the same for those taking Drug A as it is for those taking Drug B. That is, the null hypothesis is:

$$H_0: \mu_{AT} - \mu_{AC} = \mu_{BT} - \mu_{BC}$$

Figure 9.8(f) depicts this way of viewing the same interaction depicted in Figure 9.8(e) and displays the differences being compared. For those receiving Drug A, the

Treatment—Control difference equals 32 – 17 = 15, but for those receiving Drug B the difference is only 26 – 19 = 7. The difference of the differences is 15 – 7 = 8, the same as before, as it must be. In other words, Treatment is more effective if one takes Drug A instead of Drug B. We therefore have two equivalent interpretations for the interaction implied by rejecting the null hypothesis that $\beta_5 = 0$. Indeed, for any interaction there will be multiple ways to express it. It will often help to understand the interaction if all the possible interpretations are considered, but it is important to realize that they are all equivalent statements of the same effect.

Interpretation in Terms of Slopes

It is also useful to examine the coefficients for two-way contrast-coded predictors using the same procedures we developed in Chapter 7 for multiple regression models involving product terms. We begin with the estimated equation for Model A:

$$\hat{Y} = 20 + 3.5X_1 + X_2 + 5X_3 + 0.5X_4 + 2X_5$$

Then we substitute for the product definitions ($X_4 = X_1X_3$ and $X_5 = X_2X_3$) to get:

$$\hat{Y} = 20 + 3.5X_1 + X_2 + 5X_3 + 0.5X_1X_3 + 2X_2X_3$$

Next we regroup the terms to express the "simple" linear relationship between $\hat{Y}$ and one of the three predictors X_1, X_2, or X_3. We will begin with X_3; regrouping terms yields:

$$\hat{Y} = (20 + 3.5X_1 + X_2) + (5 + 0.5X_1 + 2X_2)X_3$$

which expresses the "simple" linear relationship between $\hat{Y}$ and X_3. The term in the first set of parentheses represents the intercept (given specific values of X_1 and X_2), and the term in the second set of parentheses represents the slope (i.e., as always, the change in $\hat{Y}$ given a unit change in X_3, given specific values of X_1 and X_2). Remember that X_3 codes whether or not the patient received psychotherapy treatment, so this linear relationship reflects the effectiveness of treatment versus control—the steeper the slope, the more effective the treatment relative to the control.

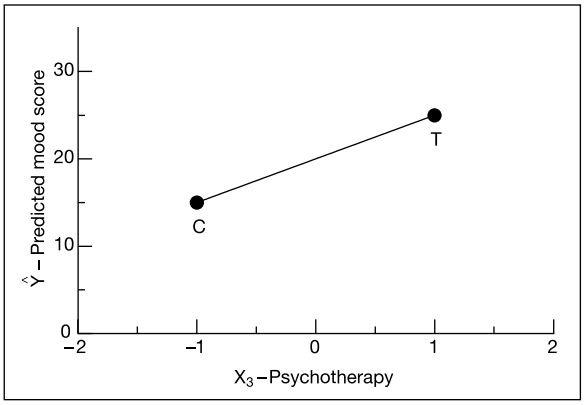
To understand the equation expressing $\hat{Y}$ as a linear function of X_3, we begin by examining Figure 9.9, which displays the relationship between $\hat{Y}$ and X_3 when $X_1 = 0$ and $X_2 = 0$.

Both X_1 and X_2 are contrast-coded predictors with an equal number of observations in each cell, so average values of X_1 and X_2 are zero; hence, the line in Figure 9.9 depicts the average relationship between $\hat{Y}$ and X_3. For $X_1 = 0$ and $X_2 = 0$, the relationship between $\hat{Y}$ and X_3 reduces to:

$$\hat{Y} = 20 + 5 \times 3$$

The intercept of 20 is, as always, the predicted value when $X_3 = 0$. Although X_3 is never zero (it is either +1 or –1), its average value is zero. Hence, we can interpret the intercept as the average value across levels of X_3. Indeed, 20 is the grand mean for these data. The slope for X_3 is 5, which implies that for every unit change in X_3 the predicted value $\hat{Y}$ changes by 5. The two-unit change from $X_3 = –1$ (Control) to $X_3 = +1$ (Treatment) predicts a change in $\hat{Y}$ (mood scores) of 2(5) = 10 points. This, as it must be, is the same interpretation we obtained for $b_3 = 5$ in terms of cell means. On average, ignoring the Drug condition, those receiving Treatment had mood scores 10 points higher than those in the Control condition.

FIGURE 9.9 Simple relationship between $\hat{Y}$ and X_3 when $X_1 = 0$ and $X_2 = 0$



What happens to the simple relationship between $\hat{Y}$ and X_3 if we do not focus on the average relationship, but instead allow X_1 and X_2 to be different from zero? The intercept term in the "simple" relationship is:

$$(20 + 3.5X_1 + X_2)$$

so if $X_1 = 1$ (indicating Drug A or B) then the intercept will increase by 3.5 points and if $X_1 = -2$ (indicating Placebo) then the intercept will decrease by 7 points. These represent the changes in $\hat{Y}$ due to X_1 when X_3 equals zero (which it does equal on average). On average, then, $\hat{Y}$ for the four drug groups ought to be 3.5 + 7 = 10.5 points higher than $\hat{Y}$ for the two placebo groups. That is, on average, mood scores for those receiving Drug A or B are 10.5 points higher than for those in the Control group. This is necessarily the same interpretation we obtained in terms of cell means. Similarly, for $X_2 = +1$ (Drug A), $X_2 = 0$ (Placebo), and $X_2 = -1$ (Drug B), the intercept increases, stays the same, or decreases. These changes in $\hat{Y}$ (mood scores) are similarly interpreted.

Now we turn to the slope in the "simple" relationship between $\hat{Y}$ and X_3 and observe what happens when X_1 and X_2 are not equal to zero. The slope term is:

$$(5 + 0.5X_1 + 2X_2)$$

so the slope relating $\hat{Y}$ and X_3 *changes* as a function of both X_1 and X_2. If $X_2 = +1$ (Drug A), then the slope for the prediction increases by $2 \times 1 = 2$ units. If $X_2 = -1$ (Drug B) then the slope decreases by $2 \times -1 = -2$ units. The steepness of the slope for X_3 represents the relative effectiveness of Treatment versus Control. Hence, the steeper slope for Drug A indicates that the relative effectiveness of Treatment versus Control is greater for Drug A than for Drug B. This is the essence of any ANOVA interaction. If there is an interaction, *the effect of one categorical variable depends on the level of another variable*. In this case, the relative effectiveness of Treatment depends on whether the patient is taking Drug A or Drug B.

Similarly, different values of X_1 also yield different slopes. If $X_1 = +1$ (either Drug A or B), then the slope for X_3 increases slightly by $b_4 = 0.5$ units, and if $X_1 = -2$ (Placebo) then the slope decreases by $0.5(-2) = -1$ units. However, $b_4 = 0.5$ is not statistically different from zero so there is no evidence that the Treatment versus Control difference depends on whether any Drug was administered ($X_1 = -1$) as compared to the Placebo.

Figure 9.10 depicts the changes in the intercept and the slope for X_3 as X_1 and X_2 change. Clearly, the different values of X_1 (Drug versus Placebo) yield noticeably different intercepts, but the changes in the slope for X_3 (Treatment versus Control) due to changes in X_1 are minimal. Although it may not appear to be the case, the slope for the bottom line ($X_1 = -2$, or Placebo) is slightly flatter than the average of the slopes of the top two lines ($X_1 = +1$, Drugs). In contrast, the intercept changes due to X_2 (Drug A versus Drug B) are noticeable but much smaller than those due to X_1 (any Drug versus Placebo). However, the slope changes due to X_2 are obvious. Indeed, the lines cross within the range of values used in this study. That is, mood scores are higher for Drug A ($X_2 = +1$) than for Drug B ($X_2 = -1$) for those receiving Treatment ($X_3 = +1$). For those in the Control condition ($X_3 = -1$) the reverse is true—mood scores for Drug B are higher than those for Drug A. These differences in slopes are the essence of an interaction.

Similar to the above interpretation of the augmented model in terms of the simple linear relationship between $\hat{Y}$ and X_3, we could also do the interpretation in terms of $\hat{Y}$ and X_2. Although the interpretation of the interaction between X_2 and X_3 must necessarily be equivalent, the alternative interpretation in terms of the relationship between $\hat{Y}$ and X_2 may produce different insights. We do not provide that interpretation here, but it is recommended as an exercise for the reader to construct and interpret the relevant graph similar to Figure 9.10. Finally, we could do the interpretation in terms of the simple linear relationship between $\hat{Y}$ and X_1. However, that is unlikely to be useful for these data because the only interaction involving X_1 is $X_4 = X_1 \times X_3$ and the coefficient b_4 for that interaction was not significantly different from zero.

Summary of Interpretation

Although both interpretation strategies yield the same interpretation, each is useful for providing a slightly different perspective. For interpreting a given ANOVA, you should use as many of the strategies as necessary until a clear interpretation of the model emerges. Below we suggest a summary interpretation that one might include in a research report:

> On average, both drug treatments produced higher mood scores ($M = 23.5$) than did the placebo ($M = 13$), PRE = .90, $F(1, 12) = 105.8$, $p < .0001$. There was not a statistically significant difference in the average mood scores produced by Drug A ($M = 24.5$) versus Drug B ($M = 22.5$), PRE = .19, $F(1, 12) = 2.9$, $p = .11$. Averaged across all three drug conditions, patients receiving psychotherapy treatment had higher mood scores ($M = 25$) than did those in the control group ($M = 15$), PRE = .90, $F(1, 12) = 108$, $p < .0001$. However there was a significant interaction between type of drug administered (A versus B) and whether or not the patient received psychotherapy treatment, PRE = .49, $F(1, 12) = 11.5$, $p < .005$, such that Drug A, relative to Drug B, produced an increase in mood scores of 6 points for those receiving treatment, but a slight decrease of 2 points for those in the control group.

Useful adjuncts to the above journal summary are graphs of the cell means (which of course are also the predicted values for the full model) as a function of the drug and psychotherapy variables. Figure 9.10 (which we developed in the section on interpreting slopes) and Figure 9.11 present two views of these data. In Figure 9.10, the differences between the lines for the drug treatments represent the "Drug" differences. In particular,

FIGURE 9.10 Simple relationship between $\hat{Y}$ and X_3 for different values of X_1 and X_2 (filled circles represent cell means)

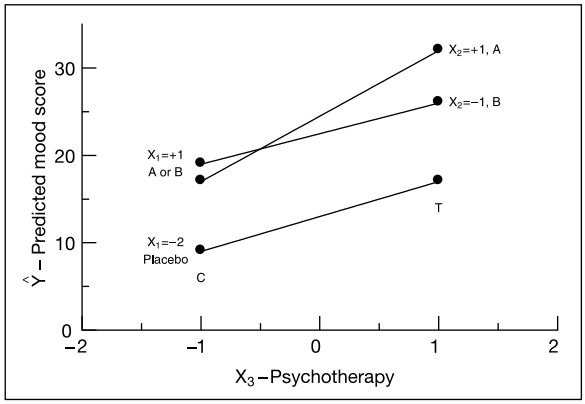
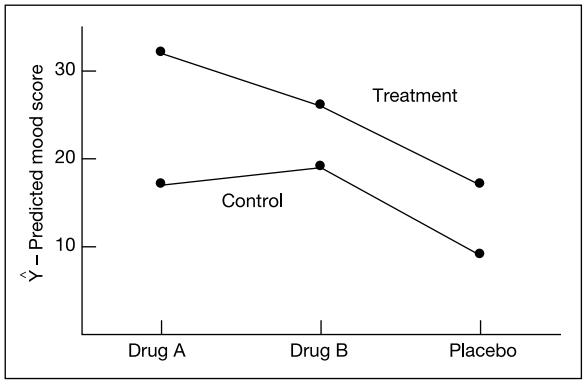


FIGURE 9.11 Cell means for mood scores by drug and psychotherapy treatment



the relatively large difference between the two lines for Drug A and Drug B and the line for Placebo corresponds to the large value for b_1, and the small difference between the Drug A and Drug B lines corresponds to the small value for b_2. That the three lines for the drug treatment groups are not parallel—the differences between lines are not constant— indicates the interaction between the drug and psychotherapy variables. In particular, the crossing of the Drug A and Drug B lines corresponds to the statistically significant value of b_5 ($X_5 = X_2X_3$), which asks whether the difference between the Drug A and Drug B lines is constant or dependent on whether the patient was in the Treatment or Control groups. The relatively large difference between the two treatment lines in Figure 9.11 similarly corresponds to the large value of b_3 and the nonparallelism corresponds to the statistically significant value of b_5. Each of these graphs depicts the interaction, but from different perspectives, just as we were able to examine the interaction from different perspectives in each of the interpretation strategies presented above.

HIGHER ORDER ANOVA

Our strategy for two-way ANOVA was to generate an appropriate set of contrast codes reflecting that two-way structure and then do the statistical analysis exactly as we did for one-way ANOVA. We can use that same strategy when we have three or more categorical predictor variables. Once we have an appropriate set of contrast codes for an experimental design with more than two categorical variables, the statistical analysis proceeds just as before. We therefore only need to learn a procedure for generating the appropriate codes. That procedure is a generalization of the same procedure we used for two-way ANOVA. For a factorial design of q categorical variables, the appropriate contrast codes are constructed according to the following rules:

1. For each of the q categorical variables, develop a set of orthogonal one-way contrast codes.
2. For each *pair* of categorical variables, construct additional contrast codes by multiplying all possible pairs of contrast codes, one from each categorical variable.
3. For each *triple* of categorical variables, construct additional contrast codes by multiplying all possible triads of contrast codes, one from each categorical variable.
4. Continue this procedure for each *quadruple, quintuple*, etc., until products are formed using codes from all q categorical variables simultaneously.

If $k_1, k_2, \ldots, k_q$ are the respective number of categories for each of the q categorical variables, then the above algorithm for generating contrast codes will produce the complete set of $k_1 \times k_2 \times \ldots \times k_q - 1$ orthogonal contrast codes.

As an example, suppose we also wanted to use the patient's gender as a predictor variable in the study of mood scores. Figure 9.12 shows the design of such a study with 12 groups, for which we will need 11 orthogonal contrast codes. We can use the five contrast codes we have already specified for the Drug by Psychotherapy design. Let λ_6 be the code for gender, with –1 indicating males and +1 indicating females. Then we need to form the products between that code and each code from the other variables. For the Drug by Gender interactions:

$$\lambda_7 = \lambda_1 \times \lambda_6$$

$$\lambda_8 = \lambda_2 \times \lambda_6$$

and for the Psychotherapy by Gender interaction:

$$\lambda_9 = \lambda_3 \times \lambda_6$$

Finally, for the Drug by Psychotherapy by Gender three-way interactions, we form the products of each triple of categorical codes:

$$\lambda_{10} = \lambda_1 \times \lambda_3 \times \lambda_6$$

$$\lambda_{11} = \lambda_2 \times \lambda_3 \times \lambda_6$$

This provides the 11 orthogonal contrast codes required by the 12 groups of the three-way design. Figure 9.13 shows the codes for each group. Select some codes and verify that they sum to zero as required for contrasts. Select several pairs of codes and verify that their crossproducts sum to zero as required by orthogonality.

FIGURE 9.12 A three-way factorial design: Drug (3) by Psychotherapy (2) by Gender (2)

	Gender			
	Male		*Female*	
	Psychotherapy		*Psychotherapy*	
Drug	*Control*	*Treatment*	*Control*	*Treatment*
A				
B				
Placebo				

FIGURE 9.13 Orthogonal contrast codes for the three-way design of Figure 9.12

Group												
	A,C,M	*B,C,M*	*P,C,M*	*A,T,M*	*B,T,M*	*P,T,M*	*A,C,F*	*B,C,F*	*P,C,F*	*A,T,F*	*B,T,F*	*P,T,F*
λ_1	1	1	−2	1	1	−2	1	1	−2	1	1	−2
λ_2	1	−1	0	1	−1	0	1	−1	0	1	−1	0
λ_3	−1	−1	−1	1	1	1	−1	−1	−1	1	1	1
$\lambda_4 = \lambda_1\lambda_3$	−1	−1	2	1	1	−2	−1	−1	2	1	1	−2
$\lambda_5 = \lambda_2\lambda_3$	−1	1	0	1	−1	0	−1	1	0	1	−1	0
λ_6	−1	−1	−1	−1	−1	−1	1	1	1	1	1	1
$\lambda_7 = \lambda_1\lambda_6$	−1	−1	2	−1	−1	2	1	1	−2	1	1	−2
$\lambda_8 = \lambda_2\lambda_6$	−1	1	0	−1	1	0	1	−1	0	1	−1	0
$\lambda_9 = \lambda_3\lambda_6$	1	1	1	−1	−1	−1	−1	−1	−1	1	1	1
$\lambda_{10} = \lambda_1\lambda_3\lambda_6$	1	1	−2	−1	−1	2	−1	−1	2	1	1	−2
$\lambda_{11} = \lambda_2\lambda_3\lambda_6$	1	−1	0	−1	1	0	−1	1	0	1	−1	0

Interpretation of the one-way codes λ_1,λ_2,λ_3, and λ_6 and the two-way codes λ_4,λ_5,λ_7,λ_8, and λ_9 use the same strategy illustrated for interpreting codes in two-way ANOVA. The three-way codes λ_{10} and λ_{11} ask whether a given two-way interaction depends on the level of the third variable. For example, we can represent $\lambda_{10} = \lambda_1\lambda_3\lambda_6$ as $\lambda_1\lambda_3 \times \lambda_6$, asking whether the two-way interaction of λ_1 (Drug versus Placebo) by λ_3 (Treatment versus Control) is the same for each level of λ_6 (Female versus Male). Equivalently, the representation $\lambda_{10} = \lambda_1\lambda_6 \times \lambda_3$ shows that the three-way interaction equally asks whether the two-way interaction of λ_1 (Drug versus Placebo) by λ_6 (Female versus Male) depends on the level of λ_3 (Treatment versus Control). Finally, the representation $\lambda_{10} = \lambda_3\lambda_6 \times \lambda_1$ asks whether the two-way interaction of λ_3 (Treatment versus Control) by λ_6 (Female versus Male) depends on the level of λ_1 (Drug versus Placebo). At a fundamental level, these three questions are the same. However, depending on the focus of the study (is it about drug effects, psychotherapy effects, or gender differences?), one representation of each three-way interaction will be most appropriate to report.

The global tests, as for one-way ANOVA, are not generally as useful as the focused questions represented by the one-degree-of-freedom contrasts. However, the testing of interactions is one instance for which global tests are sometimes useful. The proliferation of interaction contrasts in factorial designs makes testing and interpreting individual contrasts unwieldy. Interpretations of three-way and higher order interactions are often so theoretically ad hoc that some data analysts recommend examining only those interactions having a priori theoretical predictions. If the higher order interactions are

not expected and not to be examined, then it might be better to eliminate them from the analysis, using only those contrasts that will be considered. This effectively includes those eliminated interaction contrast codes in the "Error" term of the source table. Rather than eliminating the higher order interactions outright, a more conservative strategy that is frequently recommended is to do a global test of the higher order interactions and then pool those contrasts in the Error term only if the null hypothesis is not rejected for the global test. This strategy protects against missing something very unusual in the data while greatly simplifying the data analysis and its presentation. An extra benefit is that the statistical power of the other contrasts will be increased if the F for the higher order interactions is < 1, because then more will be gained by adding degrees of freedom to the within-group (Error) term than will be lost by adding extra SS in that term.

When eliminating interaction terms one must remember the rule from Chapter 7 that products represent interactions only if all the components of the product are also included in the model. This implies, for example, that if the three-way interactions are retained, then none of the one-way contrasts can be eliminated. Less obviously, it also implies that none of the two-way contrasts can be eliminated because, as illustrated above, any three-way interaction can be represented as the product of any of the two-way interactions and one of the one-way contrasts.

OTHER DETAILS IN FACTORIAL ANOVA

In all other details factorial ANOVA with two or more categorical variables is exactly the same as one-way ANOVA with one categorical variable. Specifically, confidence intervals, problems with nonorthogonal codes, handling unequal numbers of observations in each group, source tables, computational formulas for b_j and SSR_j for contrast-coded predictors, planned and post hoc comparisons, and statistical power are exactly the same. However, we provide additional details on two issues—asking other questions and assessing statistical power—in order to highlight some issues that frequently arise in factorial ANOVA.

Asking Other Questions

It is sometimes not possible to generate a set of orthogonal contrast codes addressing all the theoretical questions that one might want to ask of a given set of data. This is often true for questions that span two or more categorical variables. The strategy above for generating codes for a standard factorial design allows choice of the one-way codes for each categorical variable but then determines the remaining interaction codes as products of those one-way codes. Those products may not ask the most relevant theoretical questions. For example, in a 2 × 2 design there is only one choice (+1, –1) for each one-way code, so there is no choice for the interaction code, which must be the product of the two one-way codes. Thus, the necessary codes for two variables A and B are:

	A_1B_1	A_1B_2	A_2B_1	A_2B_2
λ_1	1	1	–1	–1
λ_2	1	–1	1	–1
λ_3	1	–1	–1	1

The interaction code λ_3 asks whether the average of the A_1B_1 and A_2B_2 groups equals the average of the other two groups. However, the theoretical question of interest might be whether the mean of one group, say A_2B_2, differs from the average of the other three groups. For example, the research hypothesis might be that both A_2 and B_2 must be present for an effect to exist. If so, the code of interest is [–1, –1, –1, 3]. This is not an interaction because it cannot be represented as the product of two contrast codes, one for each categorical variable. Nevertheless, if it is the question of interest, then it should be asked. We can compute its SSR with the usual formula and then compare it as a planned comparison (as in Chapter 8) to the mean within-group error (MSE) resulting from the analysis of variance using λ_1 through λ_3. Or, if that question is the focus of the study, one may simply ignore the factorial structure of the design and analyze the data using a set of one-way codes for four groups.

Other questions may also present themselves after the initial analysis. We can use the same post hoc comparison procedures that we used in Chapter 8. For example, in the mood score data, whether Drug A was better than Drug B depended on whether the patient was also receiving psychotherapy. Examining the means, it appears that Drug A is more effective for those in Treatment but Drug B is more effective for those in the Control group. This suggests the recommendation that Drug A, even though it was superior on average, only be administered to those concurrently receiving psychotherapy. But before making such a recommendation we ought to test with post hoc comparisons whether the difference between the means within levels of Psychotherapy are statistically significant. Comparisons such as this between categories (Drug A versus Drug B) of one variable within a single level (Treatment) are known as *simple effects*. We do the calculation for this comparison as an example. The codes for the two relevant simple effects are:

	A,T	*B,T*	*P,T*	*A,C*	*B,C*	*P,C*
λ_6	1	–1	0	0	0	0
λ_7	0	0	0	1	–1	0

The code λ_6 asks whether there is a difference between Drug A and Drug B for those receiving Treatment and the code λ_7 asks the same drug difference question for those in the Control condition. Then:

$$SSR_6 = \frac{\left(\sum_k \lambda_k \bar{Y}_k\right)^2}{\sum_k \lambda_k^2/n_k} = \frac{(32 - 26)^2}{1^2/3 + (-1)^2/3} = 54$$

The MSE (from Figure 9.4 or 9.7) is 4.2, so:

$$F = \frac{\text{SSR}}{\text{MSE}} = \frac{54}{4.2} = 12.86$$

If this were not a theoretically motivated, planned comparison, but only became a relevant question after examining the data, it would be advisable to compare F with the Scheffé adjusted critical value of:

$$(m - 1)F_{m-1;n-PA;\alpha} = 5F_{5,12;.05} = 5(3.11) = 15.55$$

$F = 12.86$ is below the adjusted critical value, so we would not be able to conclude for a post hoc comparison that Drug A was significantly better than Drug B for those patients receiving Treatment. The difference for λ_7 is even smaller, so we know without calculation that it too would not be statistically significant as a post hoc comparison.

Note that we have the odd situation in which we know that the effects of the drugs were different in the two conditions (the interaction coded by λ_5 was significant, so the Drug A vs. Drug B difference of 6 for Treatment is significantly different from the difference of –2 for Control), but neither simple effect is significant (for Treatment, 6 is not statistically different from 0; and for Control, –2 is not statistically different from 0). They are different questions, so in any analysis they certainly can have different answers. The remedy would be to conduct a replication study with more statistical power or with the simple effects as now legitimate planned comparisons.

Power in Factorial ANOVA

The methods we have presented before for estimating power work equally well for estimating statistical power in factorial ANOVA. It is interesting to examine how the power is affected for questions we ask about a given variable when additional categorical variables are added to the analysis. As an example, again consider the mood score data of Figure 9.1. If we were mainly interested in the questions about the effects of the drugs, we could have performed a one-way analysis of variance on these data, completely ignoring the psychotherapy variable. If we ignored whether patients were in the Treatment or Control groups, we would have three groups—Drug A, Drug B, and Placebo—each with six observations. The respective means would be (32 + 17)/2 = 24.5, (26 + 19)/2 = 22.5, and (17 + 9)/2 = 13. The two contrast codes would be (1, 1, –2) and (1, –1, 0). Applying the usual formulas from Chapter 8, we obtain $b_1 = 3.5$ with SSR = 441 and $b_2 = 1$ with SSR = 12, exactly the same values as before. However, PRE and F would be different because we are now testing the null hypothesis $\beta_1 = 0$ by comparing:

$$\text{MODEL A: } Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$$

$$\text{MODEL C: } Y_i = \beta_0 \qquad\quad + \beta_2 X_2 + \varepsilon_i$$

This is the same basic question asked when testing $\beta_1 = 0$ in the complete two-way analysis. However, the statistical power is not the same because the question is being tested in the context of a different Model A/C comparison.

To see the likely changes in power due to ignoring the other categorical variable, let us consider Figure 9.14, which presents the complete source table for the one-way ANOVA ignoring the psychotherapy variable. The parameter estimates, the sums of squares, the degrees of freedom, and the mean square errors reduced are exactly as they were in Figure 9.7. However, the error sum of squares and the error degrees of freedom have increased substantially. The reason is that the sums of squares that were reduced by X_3, X_4, and X_5 in the two-way analysis are now included in the error because those predictors are not used in this analysis. Given the same SSRs for X_1 and X_2 and increased SSE, then the F and PRE values must be considerably less than they were before. For example, in the complete two-way analysis of Figure 9.7, for X_1 $F_{1,12} = 105.8$ and PRE = .90, but in the one-way analysis only considering the Drug treatment variable, $F_{1,15} = 11.9$ and PRE = .44. The differences between the means for Drug A, Drug B, and Placebo

have not changed, so clearly the two-way analysis provided considerably more statistical power for testing $\beta_1 = 0$ and $\beta_2 = 0$. This increase in power is generally the case in multiway factorial ANOVA. Using one or more other categorical variables to reduce SSE allows a more powerful test of hypothesis concerning another categorical variable. The only situation for which there would be a decrease in statistical power would be if the F values for the codes of the additional variables and the codes for the interactions were < 1.0. If $F < 1.0$ for a contrast, then the proportional reduction in error associated with that contrast is less than we would expect for a randomly chosen parameter. Hence, including that contrast-coded predictor would not be worth the loss of a degree of freedom due to the extra parameter, and so power goes down. Therefore, one should add other categorical variables for the purpose of increasing statistical power only if one expects those variables and their interactions to themselves reduce error in the criterion variable. Interactions proliferate rapidly as other categorical variables with many levels are added. One must be cautious that those many interactions do not thereby reduce statistical power.

FIGURE 9.14 One-way ANOVA of mood scores ignoring the psychotherapy variable

Source	b_j	*SS*	*df*	*MS*	*F*	*p*	*PRE*
Between groups (Drug)		453	2	226.5	6.1	.012	.45
X_1	3.5	441	1	441	11.9	.004	.44
X_2	1.0	12	1	12	0.3	.58	.02
Within groups (MSE)		557	15	37.1			
Total		1010.0	17				

SUMMARY

In a factorial design, there are two or more categorical predictor variables, and all levels of each categorical variable are combined with all levels of other categorical variables. We analyze data from a two-way or higher order factorial design by applying the one-way ANOVA techniques from Chapter 8 to an appropriate set of contrast codes. We generate this set of codes by developing separate sets of one-way contrast codes for each categorical variable and then forming appropriate products among these codes to represent the interaction between categorical variables. Interpreting the interaction contrasts can be difficult, but the same interpretative techniques from Chapters 7 and 8 apply.

Notes

1 We use Z here instead of the customary X because subsequently we will do an analysis of these same data using a different set of contrast codes. The Z values are the contrast-coded predictor variables for the first analysis and the X values will be the contrast-coded predictor variables for the second analysis.

2 Note that the Z values are independent of each other, because they are based on orthogonal contrast codes and there are equal numbers of observations in each group. Hence, the sums of squares in Figure 9.4 for the individual contrast-coded predictors sum to the overall sum of squares for the complete model.