C H A P T E R

14

Two-Factor Analysis of Variance (Independent Measures)

Tools You Will Need

The following items are considered essential background material for this chapter. If you doubt your knowledge of any of these items, you should review the appropriate chapter or section before proceeding.

- Introduction to analysis of variance (Chapter 12)
  - The logic of analysis of variance
  - ANOVA notation and formulas
  - Distribution of *F*-ratios

Preview

Imagine that you are seated at your desk, ready to take the final exam in statistics. Just before the exams are handed out, a television crew appears and sets up a camera and lights aimed directly at you. They explain that they are filming students during exams for a television special. You are told to ignore the camera and go ahead with your exam.

Would the presence of a TV camera affect your performance on an exam? For some of you, the answer to this question is "definitely yes" and for others, "probably not." In fact, both answers are right; whether the TV camera affects performance depends on your personality. Some of you would become terribly distressed and self-conscious, while others really could ignore the camera and go on as if everything were normal.

In an experiment that duplicates the situation we have described, Shrauger (1972) tested participants on a concept-formation task. Half of the participants worked alone (no audience), and half worked with an audience of people who claimed to be interested in observing the experiment. Shrauger also divided the participants into two groups on the basis of personality: those high in self-esteem and those low in self-esteem. The dependent variable for this experiment was the number of errors on the concept formation task. Data similar to those obtained by Shrauger are shown in Figure 14.1. Notice that the audience had no effect on the high-self-esteem participants. However, the low-self-esteem participants made nearly twice as many errors with an audience as when working alone.

The Problem: Shrauger's study is an example of research that involves two independent variables in the same study. The independent variables are:

1. Audience (present or absent)
2. Self-esteem (high or low)

The results of the study indicate that the effect of one variable (audience) *depends on* another variable (self-esteem).

You should realize that it is quite common to have two variables that interact in this way. For example, a drug may have a profound effect on some patients and have no effect whatsoever on others. Some children survive abusive environments and live normal, productive lives, while others show serious difficulties. To observe
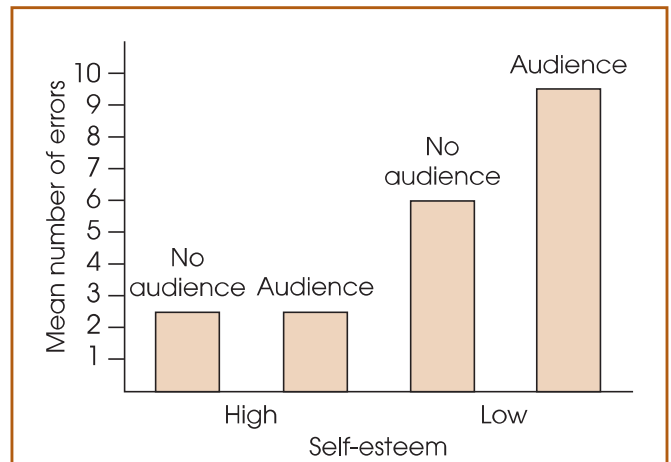

FIGURE 14.1

Results of an experiment examining the effect of an audience on the number of errors made on a concept formation task for participants who are rated either high or low in self-esteem. Notice that the effect of the audience depends on the self-esteem of the participants.

Shrauger, J. S. (1972). Self-esteem and reactions to being observed by others. *Journal of Personality and Social Psychology*, *23*, 192−200. Copyright 1972 by the American Psychological Association. Adapted by permission of the author.

how one variable interacts with another, it is necessary to study both variables simultaneously in one study. However, the analysis of variance (ANOVA) procedures introduced in Chapters 12 and 13 are limited to evaluating mean differences produced by one independent variable and are not appropriate for mean differences involving two (or more) independent variables.

The Solution: ANOVA is a very flexible hypothesis testing procedure and can be modified again to evaluate the mean differences produced in a research study with two (or more) independent variables. In this chapter we introduce the *two-factor* ANOVA, which tests the significance of each independent variable acting alone as well as the interaction between variables.

14.1 AN OVERVIEW OF THE TWO-FACTOR, INDEPENDENT-MEASURES ANOVA

In most research situations, the goal is to examine the relationship between two variables. Typically, the research study attempts to isolate the two variables to eliminate or reduce the influence of any outside variables that may distort the relationship being studied. A typical experiment, for example, focuses on one independent variable (which is expected to influence behavior) and one dependent variable (which is a measure of the behavior). In real life, however, variables rarely exist in isolation. That is, behavior usually is influenced by a variety of different variables acting and interacting simultaneously. To examine these more complex, real-life situations, researchers often design research studies that include more than one independent variable. Thus, researchers systematically change two (or more) variables and then observe how the changes influence another (dependent) variable.

An independent variable is a manipulated variable in an experiment. A quasi-independent variable is not manipulated but defines the groups of scores in a nonexperimental study.

In Chapters 12 and 13, we examined ANOVA for *single-factor* research designs—that is, designs that included only one independent variable or only one quasi-independent variable. When a research study involves more than one factor, it is called a *factorial design*. In this chapter, we consider the simplest version of a factorial design. Specifically, we examine ANOVA as it applies to research studies with exactly two factors. In addition, we limit our discussion to studies that use a separate sample for each treatment condition—that is, independent-measures designs. Finally, we consider only research designs for which the sample size (n) is the same for all treatment conditions. In the terminology of ANOVA, this chapter examines *two-factor, independent-measures, equal* n *designs.*

We use Shrauger's audience and self-esteem study described in the Chapter Preview to introduce the two-factor research design. Table 14.1 shows the structure of Shrauger's study. Note that the study involves two separate factors: One factor is manipulated by the researcher, changing from no-audience to audience, and the second factor is self-esteem, which varies from high to low. The two factors are used to create a *matrix* with the different levels of self-esteem defining the rows and the different audience conditions defining the columns. The resulting two-by-two matrix shows four different combinations of the variables, producing four different conditions. Thus, the research study would require four separate samples, one for each *cell*, or box, in the matrix. The dependent variable for the study is the number of errors on the concept-formation task for people observed in each of the four conditions.

TABLE 14.1

The structure of a two-factor experiment presented as a matrix. The two factors are self-esteem and presence/absence of an audience, with two levels for each factor.

		Factor *B*: Audience Condition	
		No Audience	Audience
Factor *A*: Self-Esteem	Low	Scores for a group of participants who are classified as low self-esteem and are tested with no audience.	Scores for a group of participants who are classified as low self-esteem and are tested with an audience.
	High	Scores for a group of participants who are classified as high self-esteem and are tested with no audience.	Scores for a group of participants who are classified as high self-esteem and are tested with an audience.

The two-factor ANOVA tests for mean differences in research studies that are structured like the audience-and-self-esteem example in Table 14.1. For this example, the two-factor ANOVA evaluates three separate sets of mean differences:

1. What happens to the mean number of errors when the audience is added or taken away?

2. Is there a difference in the mean number of errors for participants with high self-esteem compared to those with low self-esteem?

3. Is the mean number of errors affected by specific combinations of self-esteem and audience? (For example, an audience may have a large effect on participants with low self-esteem but only a small effect for those with high self-esteem.)

Thus, the two-factor ANOVA allows us to examine three types of mean differences within one analysis. In particular, we conduct three separate hypotheses tests for the same data, with a separate F-ratio for each test. The three F-ratios have the same basic structure:

$$F = \frac{\text{variance (differences) between treatments}}{\text{variance (differences) expected if there is no treatment effect}}$$

In each case, the numerator of the F-ratio measures the actual mean differences in the data, and the denominator measures the differences that would be expected if there is no treatment effect. As always, a large value for the F-ratio indicates that the sample mean differences are greater than would be expected by chance alone, and, therefore, provides evidence of a treatment effect. To determine whether the obtained F-ratios are *significant*, we need to compare each F-ratio with the critical values found in the F-distribution table in Appendix B.

14.2 MAIN EFFECTS AND INTERACTIONS

As noted in the previous section, a two-factor ANOVA actually involves three distinct hypothesis tests. In this section, we examine these three tests in more detail.

Traditionally, the two independent variables in a two-factor experiment are identified as factor A and factor B. For the study presented in Table 14.1, self-esteem is factor A, and the presence or absence of an audience is factor B. The goal of the study is to evaluate the mean differences that may be produced by either of these factors acting independently or by the two factors acting together.

MAIN EFFECTS

One purpose of the study is to determine whether differences in self-esteem (factor A) result in differences in performance. To answer this question, we compare the mean score for all of the participants with low self-esteem with the mean for those with high self-esteem. Note that this process evaluates the mean difference between the top row and the bottom row in Table 14.1.

To make this process more concrete, we present a set of hypothetical data in Table 14.2. The table shows the mean score for each of the treatment conditions (cells) as well as the overall mean for each column (each audience condition) and the overall mean for each row (each self-esteem group). These data indicate that the low self-esteem participants (the top row) had an overall mean of $M = 8$ errors. This overall mean was obtained by computing the average of the two means in the top row. In

TABLE 14.2

Hypothetical data for an experiment examining the effect of an audience on participants with different levels of self-esteem.

	No Audience	Audience	
Low	$M = 7$	$M = 9$	$M = 8$
High	$M = 3$	$M = 5$	$M = 4$
	$M = 5$	$M = 7$	

contrast, the high self-esteem participants had an overall mean of $M = 4$ errors (the mean for the bottom row). The difference between these means constitutes what is called the *main effect* for self-esteem, or the *main effect for factor* A.

Similarly, the main effect for factor B (audience condition) is defined by the mean difference between the columns of the matrix. For the data in Table 14.2, the two groups of participants tested with no audience had an overall mean score of $M = 5$ errors. Participants tested with an audience committed an overall average of $M = 7$ errors. The difference between these means constitutes the *main effect* for the audience conditions, or the *main effect for factor* B.

D E F I N I T I O N

> The mean differences among the levels of one factor are referred to as the **main effect** of that factor. When the design of the research study is represented as a matrix with one factor determining the rows and the second factor determining the columns, then the mean differences among the rows describe the main effect of one factor, and the mean differences among the columns describe the main effect for the second factor.

The mean differences between columns or rows simply *describe* the main effects for a two-factor study. As we have observed in earlier chapters, the existence of sample mean differences does not necessarily imply that the differences are *statistically significant*. In general, two samples are not expected to have exactly the same means. There are always small differences from one sample to another, and you should not automatically assume that these differences are an indication of a systematic treatment effect. In the case of a two-factor study, any main effects that are observed in the data must be evaluated with a hypothesis test to determine whether they are statistically significant effects. Unless the hypothesis test demonstrates that the main effects are significant, you must conclude that the observed mean differences are simply the result of sampling error.

The evaluation of main effects accounts for two of the three hypothesis tests in a two-factor ANOVA. We state hypotheses concerning the main effect of factor A and the main effect of factor B and then calculate two separate F-ratios to evaluate the hypotheses.

For the example we are considering, factor A involves the comparison of two different levels of self-esteem. The null hypothesis would state that there is no difference between the two levels; that is, self-esteem has no effect on performance. In symbols,

$$H_0: \quad \mu_{A_1} = \mu_{A_2}$$

The alternative hypothesis is that the two different levels of self-esteem do produce different scores:

$$H_1: \quad \mu_{A_1} \neq \mu_{A_2}$$

To evaluate these hypotheses, we compute an F-ratio that compares the actual mean differences between the two self-esteem levels versus the amount of difference that would be expected without any systematic treatment effects.

$$F = \frac{\text{variance (differences) between the means for factor } A}{\text{variance (differences) expected if there is no treatment effect}}$$

$$F = \frac{\text{variance (differences) between the row means}}{\text{variance (differences) expected if there is no treatment effect}}$$

Similarly, factor B involves the comparison of the two different audience conditions. The null hypothesis states that there is no difference in the mean number of errors between the two conditions. In symbols,

$$H_0: \quad \mu_{B_1} = \mu_{B_2}$$

As always, the alternative hypothesis states that the means are different:

$$H_1: \quad \mu_{B_1} \neq \mu_{B_2}$$

Again, the F-ratio compares the obtained mean difference between the two audience conditions versus the amount of difference that would be expected if there is no systematic treatment effect.

$$F = \frac{\text{variance (differences) between the means for factor } B}{\text{variance (differences) expected if there is no treatment effect}}$$

$$F = \frac{\text{variance (differences) between the column means}}{\text{variance (differences) expected if there is no treatment effect}}$$

INTERACTIONS   In addition to evaluating the main effect of each factor individually, the two-factor ANOVA allows you to evaluate other mean differences that may result from unique combinations of the two factors. For example, specific combinations of self-esteem and an audience acting together may have effects that are different from the effects of self-esteem or an audience acting alone. Any "extra" mean differences that are not explained by the main effects are called an *interaction,* or an *interaction between factors*. The real advantage of combining two factors within the same study is the ability to examine the unique effects caused by an interaction.

D E F I N I T I O N

> An **interaction** between two factors occurs whenever the mean differences between individual treatment conditions, or cells, are different from what would be predicted from the overall main effects of the factors.

To make the concept of an interaction more concrete, we reexamine the data shown in Table 14.2. For these data, there is no interaction; that is, there are no extra mean differences that are not explained by the main effects. For example, within each audience condition (each column of the matrix) the average number of errors for the low self-esteem participants is 4 points higher than the average for the high self-esteem participants. This 4-point mean difference is exactly what is predicted by the overall main effect for self-esteem.

Now consider a different set of data shown in Table 14.3. These new data show exactly the same main effects that existed in Table 14.2 (the column means and the row

TABLE 14.3

Hypothetical data for an experiment examining the effect of an audience on participants with different levels of self-esteem. The data show the same main effects as the values in Table 14.5 but the individual treatment means have been modified to create an interaction.

	No Audience	Audience	
Low	$M = 6$	$M = 10$	$M = 8$
High	$M = 4$	$M = 4$	$M = 4$
	$M = 5$	$M = 7$	

The data in Table 14.3 show the same pattern of results that was obtained in Shrauger's research study.

means have not been changed). But now there is an interaction between the two factors. For example, for the low self-esteem participants (top row), there is a 4-point difference in the number of errors committed with an audience and without an audience. This 4-point difference cannot be explained by the 2-point main effect for the audience factor. Also, for the high self-esteem participants (bottom row), the data show no difference between the two audience conditions. Again, the zero difference is not what would be expected based on the 2-point main effect for the audience factor. Mean differences that are not explained by the main effects are an indication of an interaction between the two factors.

To evaluate the interaction, the two-factor ANOVA first identifies mean differences that are not explained by the main effects. The extra mean differences are then evaluated by an F-ratio with the following structure:

$$F = \frac{\text{variance (mean differences) not explained by main effects}}{\text{variance (differences) expected if there is no treatment effects}}$$

The null hypothesis for this F-ratio simply states that there is no interaction:

H_0:   There is no interaction between factors A and B. All of the mean differences between treatment conditions are explained by the main effects of the two factors.

The alternative hypothesis is that there is an interaction between the two factors:

H_1:   There is an interaction between factors. The mean differences between treatment conditions are not what would be predicted from the overall main effects of the two factors.

MORE ABOUT INTERACTIONS

In the previous section, we introduced the concept of an interaction as the unique effect produced by two factors working together. This section presents two alternative definitions of an interaction. These alternatives are intended to help you understand the concept of an interaction and to help you identify an interaction when you encounter one in a set of data. You should realize that the new definitions are equivalent to the original and simply present slightly different perspectives on the same concept.

The first new perspective on the concept of an interaction focuses on the notion of independence for the two factors. More specifically, if the two factors are independent, so that one factor does not influence the effect of the other, then there is no interaction. On the other hand, when the two factors are not independent, so that the effect of one factor *depends on* the other, then there is an interaction. The notion of dependence between factors is consistent with our earlier discussion of interactions. If one factor influences the effect of the other, then unique combinations of the factors produce unique effects.

DEFINITION

> When the effect of one factor depends on the different levels of a second factor, then there is an **interaction** between the factors.

This definition of an interaction should be familiar in the context of a "drug interaction." Your doctor and pharmacist are always concerned that the effect of one medication may be altered or distorted by a second medication that is being taken at the same time. Thus, the effect of one drug (factor *A*) depends on a second drug (factor *B*), and you have an interaction between the two drugs.

Returning to Table 14.2, notice that the size of the audience effect (first column versus second column) *does not depend* on the self-esteem of the participants. For these data, adding an audience produces the same 2-point increase in errors for both groups of participants. Thus, the audience effect does not depend on self-esteem, and there is no interaction. Now consider the data in Table 14.3. This time, the effect of adding an audience *depends on* the self-esteem of the participants. For example, there is a 4-point increase in errors for the low-self-esteem participants but adding an audience has no effect on the errors for the high-self-esteem participants. Thus, the audience effect depends on the level of self-esteem, which means that there is an interaction between the two factors.

The second alternative definition of an interaction is obtained when the results of a two-factor study are presented in a line graph. In this case, the concept of an interaction can be defined in terms of the pattern displayed in the graph. Figure 14.2 shows the two sets of data we have been considering. The original data from Table 14.2, where there is no interaction, are presented in Figure 14.2(a). To construct this figure, we selected one of the factors to be displayed on the horizontal axis; in this case, the different levels of the audience factor. The dependent variable, the number of errors, is shown on the vertical axis. Note that the figure actually contains two separate graphs: The top line shows the relationship between the audience factor and errors for the low-self-esteem
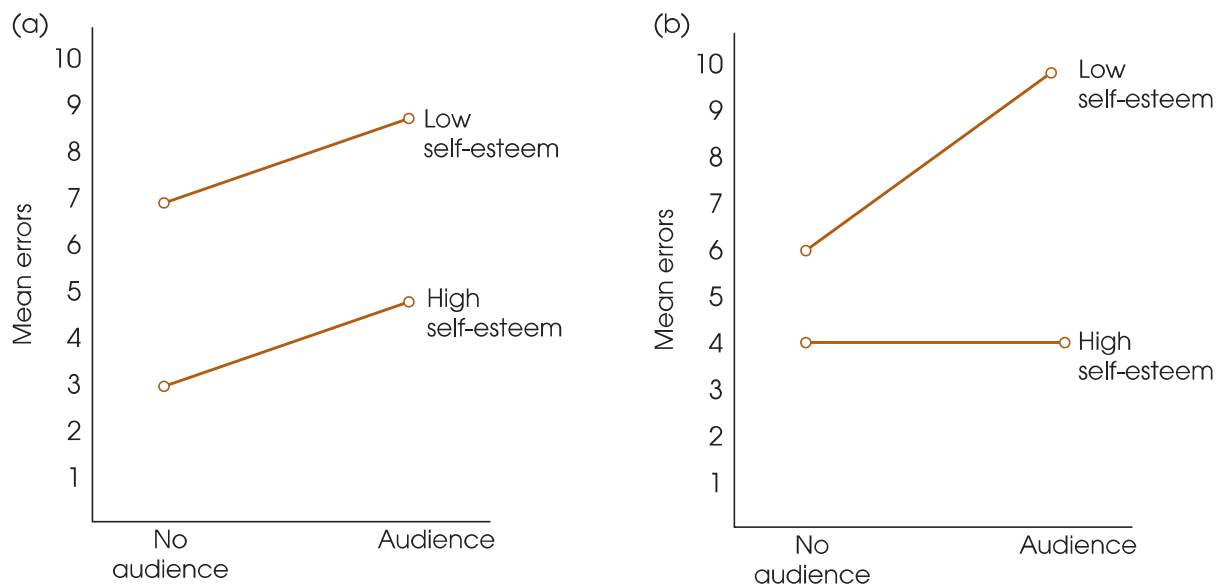

FIGURE 14.2

(a) Graph showing the treatment means from Table 14.2, for which there is no reaction. (b) Graph for Table 14.3, for which there is an interaction.

participants, and the bottom line shows the relationship for the high-self-esteem partici-pants. In general, the picture in the graph matches the structure of the data matrix; the columns of the matrix appear as values along the *X*-axis, and the rows of the matrix appear as separate lines in the graph (Box 14.1).

For the original set of data, Figure 14.2(a), note that the two lines are parallel; that is, the distance between lines is constant. In this case, the distance between lines reflects the 2-point difference in mean errors between low- and high-self-esteem participants, and this 2-point difference is the same for both audience conditions.

Now look at a graph that is obtained when there is an interaction in the data. Figure 14.2(b) shows the data from Table 14.3. This time, note that the lines in the graph are not parallel. The distance between the lines changes as you scan from left to right. For these data, the distance between the lines corresponds to the self-esteem effect—that is, the mean difference in errors for low- versus high-self-esteem partic-ipants. The fact that this difference depends on the audience condition is an indica-tion of an interaction between the two factors.

DEFINITION

When the results of a two-factor study are presented in a graph, the existence of nonparallel lines (lines that cross or converge) indicates an **interaction** between the two factors.

The $A \times B$ interaction typically is called the "*A* by *B*" interaction. If there is an interaction between an audience and self-esteem, it may be called the "audience by self-esteem" interaction.

For many students, the concept of an interaction is easiest to understand using the perspective of interdependency; that is, an interaction exists when the effects of one variable *depend* on another factor. However, the easiest way to identify an interaction within a set of data is to draw a graph showing the treatment means. The presence of nonparallel lines is an easy way to spot an interaction.

BOX 14.1 GRAPHING RESULTS FROM A TWO-FACTOR DESIGN

One of the best ways to get a quick overview of the results from a two-factor study is to present the data in a line graph. Because the graph must display the means obtained for *two* independent variables (two factors), constructing the graph can be a bit more complicated than constructing the single-factor graphs we presented in Chapter 3 (pp. 93–95).

Figure 14.3 shows a line graph presenting the results from a two-factor study with 2 levels of factor *A* and 3 levels of factor *B*. With a 2 × 3 design, there are a total of 6 different treatment means, which are shown in the following matrix.

In the graph, note that values for the dependent vari-able (the treatment means) are shown on the vertical axis. Also note that the levels for one factor (we selected factor *B*) are displayed on the horizontal axis. Directly above the B_1 value on the horizontal axis, we have placed

		Factor *B*		
		B1	*B2*	*B3*
Factor *A*	A_1	10	40	20
	A_2	30	50	30

two dots corresponding to the two means in the B_1 column of the data matrix. Similarly, we have placed two dots above B_2 and another two dots above B_3. Finally, we have drawn a line connecting the three dots corresponding to level 1 of factor *A* (the three means in the top row of the data matrix). We have also drawn a second line that connects the three dots corresponding to level 2 of factor *A*. These lines are labeled A_1 and A_2 in the figure.

FIGURE 14.3

A line graph showing the results from a two-factor experiment.
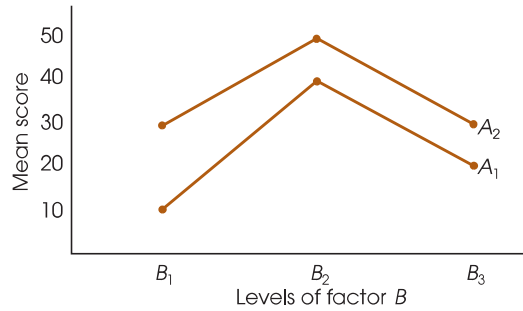
INDEPENDENCE OF MAIN EFFECTS AND INTERACTIONS

The two-factor ANOVA consists of three hypothesis tests, each evaluating specific mean differences: the A effect, the B effect, and the $A \times B$ interaction. As we have noted, these are three *separate* tests, but you should also realize that the three tests are *independent*. That is, the outcome for any one of the three tests is totally unrelated to the outcome for either of the other two. Thus, it is possible for data from a two-factor study to display any possible combination of significant and/or nonsignificant main effects and interactions. The data sets in Table 14.4 show several possibilities.

Table 14.4(a) shows data with mean differences between levels of factor A (an A effect) but no mean differences for factor B and no interaction. To identify the A effect, notice that the overall mean for A_1 (the top row) is 10 points higher than the overall mean for A_2 (the bottom row). This 10-point difference is the main effect for factor A. To evaluate the B effect, notice that both columns have exactly the same overall mean, indicating no difference between levels of factor B; hence, there is no B effect. Finally, the absence of an interaction is indicated by the fact that the overall A effect (the 10-point difference) is constant within each column; that is, the A effect *does not depend* on the levels of factor B. (Another indication is that the data indicate that the overall B effect is constant within each row.)

Table 14.4(b) shows data with an A effect and a B effect but no interaction. For these data, the A effect is indicated by the 10-point mean difference between rows, and the B effect is indicated by the 20-point mean difference between columns. The fact that the 10-point A effect is constant within each column indicates no interaction.

Finally, Table 14.4(c) shows data that display an interaction but no main effect for factor A or for factor B. For these data, there is no mean difference between rows (no A effect) and no mean difference between columns (no B effect). However, within each row (or within each column), there are mean differences. The "extra" mean differences within the rows and columns cannot be explained by the overall main effects and, therefore, indicate an interaction.

TABLE 14.4

Three sets of data showing different combinations of main effects and interaction for a two-factor study. (The numerical value in each cell of the matrices represents the mean value obtained for the sample in that treatment condition.)

(a) Data showing a main effect for factor A but no B effect and no interaction

	B_1	B_2	
A_1	20	20	A_1 mean = 20
A_2	10	10	A_2 mean = 10

B_1 mean = 15 B_2 mean = 15

10-point difference

No difference

(b) Data showing main effects for both factor A and factor B but no interaction

	B_1	B_2
A_1	10	30
A_2	20	40

A_1 mean = 20

A_2 mean = 30

10-point difference

B_1 mean = 15 B_2 mean = 35

20-point difference

(c) Data showing no main effect for either factor but an interaction

	B_1	B_2
A_1	10	20
A_2	20	10

A_1 mean = 15

A_2 mean = 15

No difference

B_1 mean = 15 B_2 mean = 15

No difference

LEARNING CHECK

1. Each of the following matrices represents a possible outcome of a two-factor experiment. For each experiment:

   a. Describe the main effect for factor A.

   b. Describe the main effect for factor B.

   c. Does there appear to be an interaction between the two factors?

Experiment I

	B_1	B_2
A_1	$M = 10$	$M = 20$
A_2	$M = 30$	$M = 40$

Experiment II

	B_1	B_2
A_1	$M = 10$	$M = 30$
A_2	$M = 20$	$M = 20$

2. In a graph showing the means from a two-factor experiment, parallel lines indicate that there is no interaction. (True or false?)

3. A two-factor ANOVA consists of three hypothesis tests. What are they?

4. It is impossible to have an interaction unless you also have main effects for at least one of the two factors. (True or false?)

ANSWERS

1. For Experiment I:

   a. There is a main effect for factor A; the scores in A_2 average 20 points higher than in A_1.

   b. There is a main effect for factor B; the scores in B_2 average 10 points higher than in B_1.

   c. There is no interaction; there is a constant 20-point difference between A_1 and A_2 that does not depend on the levels of factor B.

For Experiment II:

a. There is no main effect for factor A; the scores in A_1 and in A_2 both average 20.

b. There is a main effect for factor B; on average, the scores in B_2 are 10 points higher than in B_1.

c. There is an interaction. The difference between A_1 and A_2 depends on the level of factor B. (There is a $+10$ difference in B_1 and a -10 difference in B_2.)

2. True.

3. The two-factor ANOVA evaluates the main effect for factor A, the main effect for factor B, and the interaction between the two factors.

4. False. Main effects and interactions are completely independent.

14.3 NOTATION AND FORMULAS FOR THE TWO-FACTOR ANOVA

The two-factor ANOVA is composed of three distinct hypothesis tests:

1. The main effect for factor A (often called the A-effect). Assuming that factor A is used to define the rows of the matrix, the main effect for factor A evaluates the mean differences between rows.

2. The main effect for factor B (called the B-effect). Assuming that factor B is used to define the columns of the matrix, the main effect for factor B evaluates the mean differences between columns.

3. The interaction (called the $A \times B$ interaction). The interaction evaluates mean differences between treatment conditions that are not predicted from the overall main effects from factor A and factor B.

For each of these three tests, we are looking for mean differences between treatments that are larger than would be expected if there are no treatment effects. In each case, the significance of the treatment effect is evaluated by an F-ratio. All three F-ratios have the same basic structure:

$$F = \frac{\text{variance (mean differences) between treatments}}{\text{variance (mean differences) expected if there are no treatment effects}} \qquad (14.1)$$

The general structure of the two-factor ANOVA is shown in Figure 14.4. Note that the overall analysis is divided into two stages. In the first stage, the total variability is separated into two components: between-treatments variability and within-treatments variability. This first stage is identical to the single-factor ANOVA introduced in Chapter 12, with each cell in the two-factor matrix viewed as a separate treatment condition. The within-treatments variability that is obtained in stage 1 of the analysis is used to compute the denominator for the F-ratios. As we noted in Chapter 12, within each treatment, all of the participants are treated exactly the same. Thus, any differences that exist within the treatments cannot be caused by treatment effects. As a result,

FIGURE 14.4

Structure of the analysis for a two-factor ANOVA.
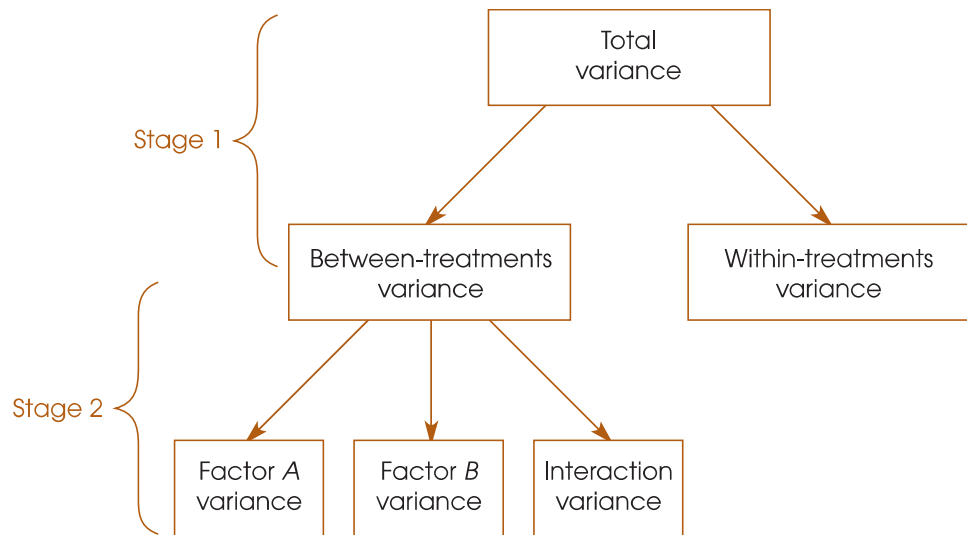
the within-treatments variability provides a measure of the differences that exist when there are no systematic treatment effects influencing the scores (see Equation 14.1).

The between-treatments variability obtained in stage 1 of the analysis combines all of the mean differences produced by factor A, factor B, and the interaction. The purpose of the second stage is to partition the differences into three separate components: differences attributed to factor A, differences attributed to factor B, and any remaining mean differences that define the interaction. These three components form the numerators for the three F-ratios in the analysis.

The goal of this analysis is to compute the variance values needed for the three F-ratios. We need three between-treatments variances (one for factor A, one for factor B, and one for the interaction), and we need a within-treatments variance. Each of these variances (or mean squares) is determined by a sum of squares value (SS) and a degrees of freedom value (df):

Remember that in ANOVA a variance is called a mean square, or MS.

$$\text{mean square} = MS = \frac{SS}{df}$$

E X A M P L E   1 4 . 1   We use the data shown in Table 14.5 to demonstrate the two-factor ANOVA. The data are representative of many studies examining the relationship between arousal and performance. The general result of these studies is that increasing the level of arousal (or motivation) tends to improve the level of performance. (You probably have tried to "psych yourself up" to do well on a task.) For very difficult tasks, however, increasing arousal beyond a certain point tends to lower the level of performance. (Your friends have probably advised you to "calm down and stay focused" when you get overanxious about doing well.) This relationship between arousal and performance is known as the Yerkes-Dodson law.

The data are displayed in a matrix with the two levels of task difficulty (factor A) making up the rows and the three levels of arousal (factor B) making up

TABLE 14.5

Data for a two-factor research study comparing two levels of task difficulty (easy and hard) and three levels of arousal (low, medium, and high). The study involves a total of six different treatment conditions with $n = 5$ participants in each condition.

		Factor B Arousal Level		
		Low	Medium	High
	Easy	3	1	10
		1	4	10
		1	8	14
		6	6	7
		4	6	9
		$M = 3$	$M = 5$	$M = 10$
		$T = 15$	$T = 25$	$T = 50$
		$SS = 18$	$SS = 28$	$SS = 26$
Factor A Task Difficulty	Difficult	0	2	1
		2		7q
		0	2	1
		0	2	6
		3	2	1
		$M = 1$	$M = 3$	$M = 2$
		$T = 5$	$T = 15$	$T = 10$
		$SS = 8$	$SS = 20$	$SS = 20$

$T_{\text{ROW1}} = 90$

$T_{\text{ROW2}} = 30$

$N = 30$
$G = 120$
$\Sigma X^2 = 860$

$T_{\text{COL1}} = 20 \quad T_{\text{COL2}} = 40 \quad T_{\text{COL3}} = 60$

the columns. For the easy task, note that performance scores increase consistently as arousal increases. For the difficult task, on the other hand, performance peaks at a medium level of arousal and drops when arousal is increased to a high level. Note that the data matrix has a total of six *cells,* or treatment conditions, with a separate sample of $n = 5$ subjects in each condition. Most of the notation should be familiar from the single-factor ANOVA presented in Chapter 12. Specifically, the treatment totals are identified by T values, the total number of scores in the entire study is $N = 30$, and the grand total (sum) of all 30 scores is $G = 120$. In addition to these familiar values, we have included the totals for each row and for each column in the matrix. The goal of the ANOVA is to determine whether the mean differences observed in the data are significantly greater than would be expected if there are no treatment effects.

STAGE 1 OF THE TWO-FACTOR ANOVA

The first stage of the two-factor ANOVA separates the total variability into two components: between-treatments and within-treatments. The formulas for this stage are identical to the formulas used in the single-factor ANOVA in Chapter 12 with the provision that each cell in the two-factor matrix is treated as a separate treatment condition. The formulas and the calculations for the data in Table 14.5 are as follows:

Total variability

$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{N} \tag{14.2}$$

For these data,

$$SS_{total} = 860 - \frac{120^2}{30}$$

$$= 860 - 480$$

$$= 380$$

This SS value measures the variability for all $N = 30$ scores and has degrees of freedom given by

$$df_{total} = N - 1 \tag{14.3}$$

For the data in Table 14.5, $df_{total} = 29$.

Within-treatments variability   To compute the variance within treatments, we first compute SS and $df = n - 1$ for each of the individual treatment conditions. Then the within-treatments SS is defined as

$$SS_{within\ treatments} = \Sigma SS_{each\ treatment} \tag{14.4}$$

And the within-treatments df is defined as

$$df_{within\ treatments} = \Sigma df_{each\ treatment} \tag{14.5}$$

For the six treatment conditions in Table 14.4,

$$SS_{within\ treatments} = 18 + 28 + 26 + 8 + 20 + 20$$

$$= 120$$

$$df_{within\ treatments} = 4 + 4 + 4 + 4 + 4 + 4$$

$$= 24$$

Between-treatments variability   Because the two components in stage 1 must add up to the total, the easiest way to find $SS_{between\ treatments}$ is by subtraction.

$$SS_{between\ treatments} = SS_{total} - SS_{within} \tag{14.6}$$

For the data in Table 14.4, we obtain

$$SS_{between\ treatments} = 380 - 120 = 260$$

However, you can also use the computational formula to calculate $SS_{between\ treatments}$ directly.

$$SS_{between\ treatments} = \Sigma \frac{T^2}{n} - \frac{G^2}{N} \tag{14.7}$$

For the data in Table 14.4, there are six treatments (six T values), each with $n = 5$ scores, and the between-treatments SS is

$$SS_{between\ treatments} = \frac{15^2}{5} + \frac{25^2}{5} + \frac{50^2}{5} + \frac{5^2}{5} + \frac{15^2}{5} + \frac{10^2}{5} - \frac{120^2}{30}$$

$$= 45 + 125 + 500 + 5 + 45 + 20 - 480$$

$$= 260$$

The between-treatments df value is determined by the number of treatments (or the number of T values) minus one. For a two-factor study, the number of treatments is equal to the number of cells in the matrix. Thus,

$$df_{\text{between treatments}} = \text{number of cells} - 1 \qquad (14.8)$$

For these data, $df_{\text{between treatments}} = 5$.

This completes the first stage of the analysis. Note that the two components, when added, equal the total for both SS values and df values.

$$SS_{\text{between treatments}} + SS_{\text{within treatments}} = SS_{\text{total}}$$

$$240 + 120 = 360$$

$$df_{\text{between treatments}} + df_{\text{within treatments}} = df_{\text{total}}$$

$$5 + 24 = 29$$

STAGE 2 OF THE TWO-FACTOR ANOVA

The second stage of the analysis determines the numerators for the three F-ratios. Specifically, this stage determines the between-treatments variance for factor A, factor B, and the interaction.

1. Factor A. The main effect for factor A evaluates the mean differences between the levels of factor A. For this example, factor A defines the rows of the matrix, so we are evaluating the mean differences between rows. To compute the SS for factor A, we calculate a between-treatment SS using the row totals in exactly the same way that we computed $SS_{\text{between treatments}}$ using the treatment totals (T values) earlier. For factor A, the row totals are 90 and 30, and each total was obtained by adding 15 scores.

   Therefore,

   $$SS_A = \Sigma \frac{T_{ROW}^2}{n_{ROW}} - \frac{G^2}{N} \qquad (14.9)$$

   For our data,

   $$SS_A = \frac{90^2}{15} + \frac{30^2}{15} - \frac{120^2}{30}$$

   $$= 540 + 60 - 480$$

   $$= 120$$

   Factor A involves two treatments (or two rows), easy and difficult, so the df value is

   $$df_A = \text{number of rows} - 1 \qquad (14.10)$$

   $$= 2 - 1$$

   $$= 1$$

2. Factor B. The calculations for factor B follow exactly the same pattern that was used for factor A, except for substituting columns in place of rows. The main

effect for factor B evaluates the mean differences between the levels of factor B, which define the columns of the matrix.

$$SS_B = \Sigma \frac{T_{COL}^2}{n_{COL}} - \frac{G^2}{N} \qquad (14.11)$$

For our data, the column totals are 20, 40, and 60, and each total was obtained by adding 10 scores. Thus,

$$SS_B = \frac{20^2}{10} + \frac{40^2}{10} + \frac{60^2}{10} - \frac{120^2}{30}$$

$$= 40 + 160 + 360 - 480$$

$$= 80$$

$$df_B = \text{number of columns} - 1 \qquad (14.12)$$

$$= 3 - 1$$

$$= 2$$

3. **The $A \times B$ Interaction.** The $A \times B$ interaction is defined as the "extra" mean differences not accounted for by the main effects of the two factors. We use this definition to find the SS and df values for the interaction by simple subtraction. Specifically, the between-treatments variability is partitioned into three parts: the A effect, the B effect, and the interaction (see Figure 14.4). We have already computed the SS and df values for A and B, so we can find the interaction values by subtracting to find out how much is left. Thus,

$$SS_{A \times B} = SS_{\text{between treatments}} - SS_A - SS_B \qquad (14.13)$$

For our data,

$$SS_{A \times B} = 260 - 120 - 80$$

$$= 60$$

Similarly,

$$df_{A \times B} = df_{\text{between treatments}} - df_A - df_B \qquad (14.14)$$

$$= 5 - 1 - 2$$

$$= 2$$

The two-factor ANOVA consists of three separate hypothesis tests with three separate F-ratios. The denominator for each F-ratio is intended to measure the variance (differences) that would be expected if there are no treatment effects. As we saw in Chapter 12, the within-treatments variance is the appropriate denominator for an independent-measures design. Remember that inside each treatment all of the individuals are treated exactly the same, which means that the differences that exist were not caused by any systematic treatment effects (see Chapter 12, p. 393). The within-treatments variance is called a *mean square,* or *MS*, and is computed as follows:

$$MS_{\text{within treatments}} = \frac{SS_{\text{within treatments}}}{df_{\text{within treatments}}}$$

For the data in Table 14.4,

$$MS_{\text{within treatments}} = \frac{120}{24} = 5.00$$

This value forms the denominator for all three F-ratios.

The numerators of the three F-ratios all measured variance or differences between treatments: differences between levels of factor A, differences between levels of factor B, and extra differences that are attributed to the $A \times B$ interaction. These three variances are computed as follows:

$$MS_A = \frac{SS_A}{df_A} \qquad MS_B = \frac{SS_B}{df_B} \qquad MS_{A \times B} = \frac{SS_{A \times B}}{df_{A \times B}}$$

For the data in Table 14.5, the three MS values are

$$MS_A = \frac{SS_A}{df_A} = \frac{120}{1} = 120 \qquad MS_B = \frac{SS_B}{df_B} = \frac{80}{2} = 40$$

$$MS_{A \times B} = \frac{SS_{A \times B}}{df_{A \times B}} = \frac{60}{2} = 30$$

Finally, the three F-ratios are

$$F_A = \frac{MS_A}{MS_{\text{within treatments}}} = \frac{120}{5} = 24.00$$

$$F_B = \frac{MS_B}{MS_{\text{within treatments}}} = \frac{40}{5} = 8.00$$

$$F_{A \times B} = \frac{MS_{A \times B}}{MS_{\text{within treatments}}} = \frac{30}{5} = 6.00$$

To determine the significance of each F-ratio, we must consult the F distribution table using the df values for each of the individual F-ratios. For this example, the F-ratio for factor A has $df = 1$ for the numerator and $df = 24$ for the denominator. Checking the table with $df = 1, 24$, we find a critical value of 4.26 for $\alpha = .05$ and a critical value of 7.82 for $\alpha = .01$. Our obtained F-ratio, $F = 24.00$ exceeds both of these values, so we conclude that there is a significant difference between the levels of factor A. That is, performance on the easy task (top row) is significantly different from performance on the difficult task (bottom row).

The F-ratio for factor B has $df = 2, 24$. The critical values obtained from the table are 3.40 for $\alpha = .05$ and 5.61 for $\alpha = .01$. Again, our obtained F-ratio, $F = 8.00$, exceeds both values, so we can conclude that there are significant differences among the levels of factor B. For this study, the three levels of arousal result in significantly different levels of performance.

Finally, the F-ratio for the $A \times B$ interaction has $df = 2, 24$ (the same as factor B). With critical values of 3.40 for $\alpha = .05$ and 5.61 for $\alpha = .01$, our obtained F-ratio of $F = 6.00$ is sufficient to conclude that there is a significant interaction between task difficulty and level of arousal.

Table 14.6 is a summary table for the complete two-factor ANOVA from Example 14.1. Although these tables are no longer commonly used in research reports, they provide a concise format for displaying all of the elements of the analysis.

LEARNING CHECK

1. Explain why the within-treatment variability is the appropriate denominator for the two-factor independent-measures F-ratios.

2. The following data summarize the results from a two-factor independent-measures experiment:

	Factor B		
	B_1	B_2	B_3
A_1	$n = 10$ $T = 0$ $SS = 30$	$n = 10$ $T = 10$ $SS = 40$	$n = 10$ $T = 20$ $SS = 50$
A_2	$n = 10$ $T = 40$ $SS = 60$	$n = 10$ $T = 30$ $SS = 50$	$n = 10$ $T = 20$ $SS = 40$

Factor A

a. Calculate the totals for each level of factor A, and compute SS for factor A.

b. Calculate the totals for factor B, and compute SS for this factor. (*Note:* You should find that the totals for B are all the same, so there is no variability for this factor.)

c. Given that the between-treatments (or between-cells) SS is equal to 100, what is the SS for the interaction?

ANSWERS

1. Within each treatment condition, all individuals are treated exactly the same. Therefore, the within-treatment variability measures the differences that exist between one score and another when there is no treatment effect causing the scores to be different. This is exactly the variance that is needed for the denominator of the F-ratios.

2. a. The totals for factor A are 30 and 90, and each total is obtained by adding 30 scores. $SS_A = 60$.

b. All three totals for factor B are equal to 40. Because they are all the same, there is no variability, and $SS_B = 0$.

c. The interaction is determined by differences that remain after the main effects have been accounted for. For these data,

$$SS_{A \times B} = SS_{\text{between treatments}} - SS_A - SS_B$$

$$= 100 - 60 - 0$$

$$= 40$$

TABLE 14.6

A summary table for the two-factor ANOVA for the data from Example 14.1.

Source	SS	df	MS	F
Between treatments	260	5		
Factor A (difficulty)	120	1	120	$F(1, 24) = 24.00$
Factor B (arousal)	80	2	40	$F(2, 24) = 8.00$
$A \times B$	60	2	30	$F(2, 24) = 6.00$
Within treatments	120	24	5	
Total	380	29		

MEASURING EFFECT SIZE FOR THE TWO-FACTOR ANOVA

The general technique for measuring effect size with an ANOVA is to compute a value for η^2, the percentage of variance that is explained by the treatment effects. For a two-factor ANOVA, we compute three separate values for eta squared: one measuring how much of the variance is explained by the main effect for factor A, one for factor B, and a third for the interaction. As we did with the repeated-measures ANOVA (p. 446), we remove any variability that can be explained by other sources before we calculate the percentage for each of the three specific effects. Thus, for example, before we compute the η^2 for factor A, we remove the variability that is explained by factor B and the variability explained by the interaction. The resulting equation is,

$$\text{for factor } A, \ \eta^2 = \frac{SS_A}{SS_{\text{total}} - SS_B - SS_{A \times B}} \tag{14.15}$$

Note that the denominator of Equation 14.15 consists of the variability that is explained by factor A and the other *unexplained* variability. Thus, an equivalent version of the equation is,

$$\text{for factor } A, \ \eta^2 = \frac{SS_A}{SS_A + SS_{\text{within treatments}}} \tag{14.16}$$

Similarly, the η^2 formulas for factor B and for the interaction are as follows:

$$\text{for factor } B, \ \eta^2 = \frac{SS_B}{SS_{\text{total}} - SS_A - SS_{A \times B}} = \frac{SS_B}{SS_B + SS_{\text{within treatments}}} \tag{14.17}$$

$$\text{for } A \times B, \ \eta^2 = \frac{SS_{A \times B}}{SS_{\text{total}} - SS_A - SS_B} = \frac{SS_{A \times B}}{SS_{A \times B} + SS_{\text{within treatments}}} \tag{14.18}$$

Because each of the η^2 equations computes a percentage that is not based on the total variability of the scores, the results are often called *partial* eta squares. For the data in Example 14.1, the equations produce the following values:

$$\eta^2 \text{ for factor } A \text{ (difficulty)} = \frac{120}{380 - 80 - 60} = \frac{120}{240} = 0.50 \ \left(50\%\right)$$

$$\eta^2 \text{ for factor } B \text{ (arousal)} = \frac{80}{380 - 120 - 60} = \frac{80}{200} = 0.40 \ \left(40\%\right)$$

$$\eta^2 \text{ for the interaction} = \frac{60}{380 - 120 - 80} = \frac{60}{180} = 0.33 \ \left(33\%\right)$$

IN THE LITERATURE
REPORTING THE RESULTS OF A TWO-FACTOR ANOVA

The APA format for reporting the results of a two-factor ANOVA follows the same basic guidelines as the single-factor report. First, the means and standard deviations are reported. Because a two-factor design typically involves several treatment conditions, these descriptive statistics usually are presented in a table or a graph.

Next, the results of all three hypothesis tests (F-ratios) are reported. The results for the study in Example 14.1 could be reported as follows:

The means and standard deviations for all treatment conditions are shown in Table 1. The two-factor analysis of variance showed a significant main effect for task difficulty, $F(1, 24) = 24.00$, $p < .01$, $\eta^2 = 0.50$; a significant main effect for arousal, $F(2, 24) = 8.00$, $p < .01$, $\eta^2 = 0.40$; and a significant interaction between difficulty and arousal, $F(2, 24) = 6.00$, $p < .01$, $\eta^2 = 0.33$.

TABLE 1
Mean performance score for each treatment condition.

		Level of Arousal		
		Low	Medium	High
Difficulty	Easy	$M = 3$ $SD = 2.12$	$M = 5$ $SD = 2.65$	$M = 10$ $SD = 2.55$
	Hard	$M = 1$ $SD = 1.41$	$M = 3$ $SD = 2.24$	$M = 2$ $SD = 2.24$

INTERPRETING THE RESULT FROM A TWO-FACTOR ANOVA

Because the two-factor ANOVA involves three separate tests, you must consider the overall pattern of results rather than focusing on the individual main effects or the interaction. In particular, whenever there is a significant interaction, you should be cautious about accepting the main effects at face value (whether they are significant or not). Remember, an interaction means that the effect of one factor *depends on* the level of the second factor. Because the effect changes from one level to the next, there is no consistent "main effect."

Figure 14.5 shows the sample means obtained from the task difficulty and arousal study. Recall that the analysis showed that both main effects and the interaction were

FIGURE 14.5

Sample means for the data in Example 14.1. The data are hypothetical results for a two-factor study examining how performance is related to task difficulty and level of arousal.
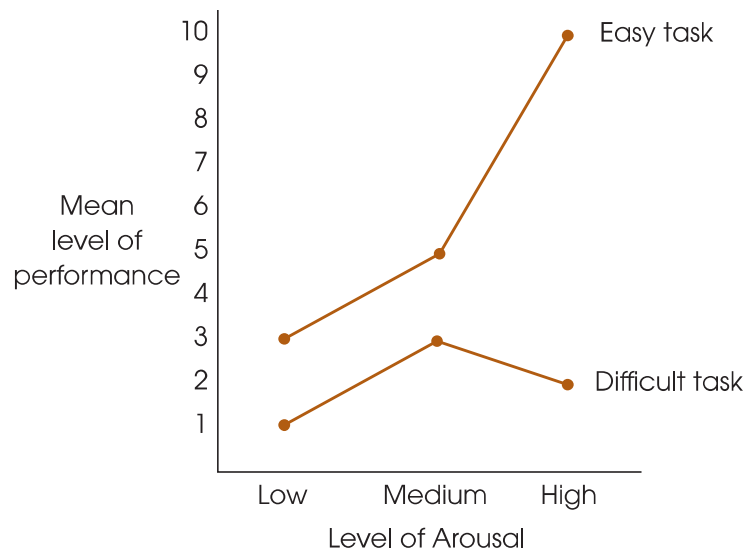
significant. The main effect for factor *A* (task difficulty) can be seen by the fact that the scores on the easy task are generally higher than scores on the difficult task.

The main effect for factor *B* (arousal) is based on the general tendency for the scores to increase as the level of arousal increases. However, this is not a completely consistent trend. In fact, the scores on the difficult task show a sharp *decrease* when arousal is increased from moderate to high. This is an example of the complications that can occur when you have a significant interaction. Remember that an interaction means that a factor does not have a consistent effect. Instead, the effect of one factor *depends on* the other factor. For the data in Figure 14.5, the effect of increasing arousal depends on the task difficulty. For the easy task, increasing arousal produces increased performance. For the difficult task, however, increasing arousal beyond a moderate level produces decreased performance. Thus, the consequences of increasing arousal *depend on* the difficulty of the task. This interdependence between factors is the source of the significant interaction.

TESTING SIMPLE MAIN EFFECTS

The existence of a significant interaction indicates that the effect (mean differences) for one factor depends on the levels of the second factor. When the data are presented in a matrix showing treatment means, a significant interaction indicates that the mean differences within one column (or row) show a different pattern than the mean differences within another column (or row). In this case, a researcher may want to perform a separate analysis for each of the individual columns (or rows). In effect, the researcher is separating the two-factor experiment into a series of separate single-factor experiments. The process of testing the significance of mean differences within one column (or one row) of a two-factor design is called testing *simple main effects*. To demonstrate this process, we once again use the data from the task-difficulty and arousal study (Example 14.1), which are summarized in Figure 14.5.

E X A M P L E   1 4 . 2

For this demonstration, we test for significant mean differences within each column of the two-factor data matrix. That is, we test for significant mean differences between the two levels of task difficulty for the low level of arousal, then repeat the test for the medium level of arousal, and once more for the high level. In terms of the two-factor notation system, we test the simple main effect of factor *A* for each level of factor *B*.

For the low level of arousal   We begin by considering only the low level of arousal. Because we are restricting the data to the first column of the data matrix, the data effectively have been reduced to a single-factor study comparing only two treatment conditions. Therefore, the analysis is essentially a single-factor ANOVA duplicating the procedure presented in Chapter 12. To facilitate the change from a two-factor to a

Low Level of Arousal		
Easy Task	Difficult Task	
$n = 5$	$n = 5$	$N = 10$
$M = 3$	$M = 1$	$G = 20$
$T = 15$	$T = 5$	

single-factor ANOVA, the data for the low level of arousal (first column of the matrix) are reproduced using the notation for a single-factor study.

S T E P   1 State the hypothesis. For this restricted set of the data, the null hypothesis would state that there is no difference between the mean for the easy task condition and the mean for the difficult task condition. In symbols,

$$H_0:\ \mu_{easy} = \mu_{difficult} \qquad \text{for the low level of arousal}$$

S T E P   2 To evaluate this hypothesis, we use an F-ratio for which the numerator, $MS_{between\ treatments}$, is determined by the mean differences between these two groups and the denominator consists of $MS_{within\ treatments}$ from the original ANOVA. Thus, the F-ratio has the structure

$$F = \frac{\text{variance (differences) for the means in column 1}}{\text{variance (differences) expected if there are no treatment effects}}$$

$$= \frac{MS_{between\ treatments}\ \text{for the two treatments in column 1}}{MS_{within\ treatments}\ \text{from the original ANOVA}}$$

To compute the $MS_{between\ treatments}$, we begin with the two treatment totals $T = 15$ and $T = 5$. Each of these totals is based on $n = 5$ scores, and the two totals add up to a grand total of $G = 20$. The $SS_{between\ treatments}$ for the two treatments is

$$SS_{between\ treatments} = \Sigma\frac{T^2}{n} - \frac{G^2}{N}$$

$$= \frac{15^2}{5} + \frac{5^2}{5} - \frac{20^2}{10}$$

$$= 45 + 5 - 40$$

$$= 10$$

$$MS_{between\ treatments} = \frac{SS}{df} = \frac{10}{1} = 10$$

Remember that the F-ratio uses $MS_{within\ treatments}$ from the original ANOVA. This $MS = 5$ with $df = 24$. Because this SS value is based on only two treatments, it has $df = 1$. Therefore,

Using $MS_{within\ treatments} = 5$ from the original two-factor analysis, the final F-ratio is

$$F = \frac{MS_{between\ treatments}}{MS_{within\ treatments}} = \frac{10}{5} = 2.00$$

Note that this F-ratio has the same df values (1, 24) as the test for factor A main effects (easy versus difficult) in the original ANOVA. Therefore, the critical value for the F-ratio is the same as that in the original ANOVA. With $df = 1, 24$ the critical value is 4.26. In this case, our F-ratio fails to reach the critical value, so we conclude that there is no significant difference between the two tasks, easy and difficult, at a low level of arousal.

For the medium level of arousal The test for the medium level of arousal follows the same process. The data for the medium level are as follows:

Medium Level of Arousal		
Easy Task	Difficult Task	
$n = 5$	$n = 5$	$N = 10$
$M = 5$	$M = 3$	$G = 40$
$T = 25$	$T = 15$	

Note that these data show a 2-point mean difference between the two conditions ($M = 5$ and $M = 3$), which is exactly the same as the 2-point difference that we evaluated for the low level of arousal ($M = 3$ and $M = 1$). Because the mean difference is the same for these two levels of arousal, the F-ratios are also identical. For the low level of arousal, we obtained $F(1, 24) = 2.00$, which was not significant. This test also produces $F(1, 24) = 2.00$ and again we conclude that there is no significant difference. (*Note:* You should be able to complete the test to verify this decision.)

For the high level of arousal The data for the high level are as follows:

High Level of Arousal		
Easy Task	Difficult Task	
$n = 5$	$n = 5$	$N = 10$
$M = 10$	$M = 2$	$G = 60$
$T = 50$	$Y = 10$	

For these data,

$$SS_{\text{between treatments}} = \Sigma \frac{T^2}{n} - \frac{G^2}{N}$$

$$= \frac{50^2}{5} + \frac{10^2}{5} + \frac{60^2}{10}$$

$$= 500 + 20 - 360$$

$$= 160$$

Again, we are comparing only two treatment conditions, so $df = 1$ and

$$MS_{\text{between treatments}} = \frac{SS}{df} = \frac{160}{1} = 160$$

Thus, for the high level of arousal, the final F-ratio is

$$F = \frac{MS_{\text{between treatments}}}{MS_{\text{within treatments}}} = \frac{160}{5} = 32.00$$

As before, this F-ratio has $df = 1, 24$ and is compared with the critical value $F = 4.26$. This time the F-ratio is far into the critical region and we conclude that

there is a significant difference between the easy task and the difficult task for the high level of arousal.

As a final note, we should point out that the evaluation of simple main effects accounts for the interaction as well as the overall main effect for one factor. In Example 14.1, the significant interaction indicates that the effect of task difficulty (factor A) depends on the level of arousal (factor B). The evaluation of the simple main effects demonstrates this dependency. Specifically, task difficulty has no significant effect on performance when arousal level is low or medium, but does have a significant effect when arousal level is high. Thus, the analysis of simple main effects provides a detailed evaluation of the effects of one factor *including its interaction with a second factor.*

The fact that the simple main effects for one factor encompass both the interaction and the overall main effect of the factor can be seen if you consider the SS values. For this demonstration,

Simple Main Effects for Arousal	Interaction and Main Effect for Arousal
$SS_{\text{low arousal}} = 10$	$SS_{A \times B} = 60$
$SS_{\text{medium arousal}} = 10$	$SS_A = 120$
$SS_{\text{high arousal}} = 160$	
Total $SS = 180$	Total $SS = 180$

Notice that the total variability from the simple main effects of difficulty (factor A) completely accounts for the total variability of factor A and the $A \times B$ interaction.

14.4 USING A SECOND FACTOR TO REDUCE VARIANCE CAUSED BY INDIVIDUAL DIFFERENCES

As we noted in Chapters 10 and 12, a concern for independent-measures designs is the variance that exists within each treatment condition. Specifically, large variance tends to reduce the size of the t statistic or F-ratio and, therefore, reduces the likelihood of finding significant mean differences. Much of the variance in an independent-measures study comes from individual differences. Recall that individual differences are the characteristics, such as age or gender, that differ from one participant to the next and can influence the scores obtained in the study.

Occasionally, there are consistent individual differences for one specific participant characteristic. For example, the males in a study may consistently have lower scores than the females. Or, the older participants may have consistently higher scores than the younger participants. For example, suppose that a researcher compares two treatment conditions using a separate group of children for each condition. Each group of participants contains a mix of boys and girls. Hypothetical data for this study are shown in Table 14.7(a), with each child's gender noted with an M or an F. While examining the results, the researcher notices that the girls tend to have higher scores than the boys, which produces big individual differences and high variance within each group. Fortunately, there is a relatively simple solution to the problem of high variance. The solution involves using the specific variable, in this case gender,

TABLE 14.7

A single-factor study comparing two treatments (a) can be transformed into a two-factor study (b) by using a participant characteristic (gender) as a second factor. This process creates smaller, more homogeneous groups, which reduces the variance within groups.

(a)

Treatment I	Treatment II
3 (M)	8 (F)
4 (F)	4 (F)
4 (F)	1 (M)
0 (M)	10 (F)
6 (F)	5 (M)
1 (M)	5 (M)
2 (F)	10 (F)
4 (M)	5 (M)
$M = 3$	$M = 6$
$SS = 50$	$SS = 68$

(b)

	Treatment I	Treatment II
Males	3	1
	0	5
	1	5
	4	5
	$M = 2$	$M = 4$
	$SS = 10$	$SS = 12$
Females	4	8
	4	4
	6	10
	2	10
	$M = 4$	$M = 8$
	$SS = 8$	$SS = 24$

as a second factor. Instead of one group in each treatment, the researcher divides the participants into two separate groups within each treatment: a group of boys and a group of girls. This process creates the two-factor study shown in Table 14.7(b), with one factor consisting of the two treatments (I and II) and the second factor consisting of the gender (male and female).

By adding a second factor and creating four groups of participants instead of only two, the researcher has greatly reduced the individual differences (gender differences) within each group. This should produce a smaller variance within each group and, therefore, increase the likelihood of obtaining a significant mean difference. This process is demonstrated in the following example.

EXAMPLE 14.3   We use the data in Table 14.7 to demonstrate how the variance caused by individual differences can be reduced by adding a participant characteristic, such as age or gender, as a second factor. For the single-factor study in Table 14.7(a), the two treatments produce $SS_{\text{within treatments}} = 50 + 68 = 118$. With $n = 8$ in each treatment, we obtain $df_{\text{within treatments}} = 7 + 7 = 14$. These values produce $MS_{\text{within treatments}} = \frac{118}{14} = 8.43$, which is the denominator of the F-ratio evaluating the mean difference between treatments. For the two-factor study in Table 14.7(b), the four treatments produce $SS_{\text{within treatments}} = 10 + 12 + 8 + 24 = 54$. With $n = 4$ in each treatment, we obtain $df_{\text{within treatments}} = 3 + 3 + 3 + 3 = 12$. These value produce $MS_{\text{within treatments}} = \frac{54}{12} = 4.50$, which is the denominator of the F-ratio evaluating the main effect for the treatments. Notice that the error term for the single-factor F is nearly twice as big as the error term for the two-factor F. Reducing the individual differences within each group has greatly reduced the within-treatment variance that forms the denominator of the F-ratio.

Both designs, single-factor and two-factor, evaluate the difference between the two treatment means, $M = 3$ and $M = 6$, with $n = 8$ in each treatment. These values produce $SS_{\text{between treatments}} = 36$ and, with $k = 2$ treatments, we obtain $df_{\text{between treatments}} = 1$. Thus, $MS_{\text{between treatments}} = \frac{36}{1} = 36$. (For the two-factor design, this is the MS for the main effect of the treatment factor.) With different denominators, however, the two designs produce very different F-ratios. For the single-factor design, we obtain

$$F = \frac{MS_{\text{between treatments}}}{MS_{\text{within treatments}}} = \frac{36}{8.43} = 4.27$$

With $df = 1, 14$, the critical value for $\alpha = .05$ is $F = 4.60$. Our F-ratio is not in the critical region, so we fail to reject the null hypothesis and must conclude that there is no significant difference between the two treatments.

For the two-factor design, however, we obtain

$$F = \frac{MS_{\text{between treatments}}}{MS_{\text{within treatments}}} = \frac{36}{4.50} = 8.88$$

With $df = 1, 14$, the critical value for $\alpha = .05$ is $F = 4.75$. Our F-ratio is well beyond this value, so we reject the null hypothesis and conclude that there is a significant difference between the two treatments.

For the single-factor study in Example 14.3, the individual differences caused by gender are part of the variance within each treatment condition. This increased variance reduces the F-ratio and results in a conclusion of no significant difference between treatments. In the two-factor analysis, the individual differences caused by gender are measured by the main effect for gender, which is a between-groups factor. Because the gender differences are now between-groups rather than within-groups, they no longer contribute to the variance.

The two-factor ANOVA has other advantages beyond reducing the variance. Specifically, it allows you to evaluate mean differences between genders as well as differences between treatments, and it reveals any interaction between treatment and gender.

14.5 ASSUMPTIONS FOR THE TWO-FACTOR ANOVA

The validity of the ANOVA presented in this chapter depends on the same three assumptions we have encountered with other hypothesis tests for independent-measures designs (the t test in Chapter 10 and the single-factor ANOVA in Chapter 12):

1. The observations within each sample must be independent (see p. 254).
2. The populations from which the samples are selected must be normal.
3. The populations from which the samples are selected must have equal variances (homogeneity of variance).

As before, the assumption of normality generally is not a cause for concern, especially when the sample size is relatively large. The homogeneity of variance assumption is more important, and if it appears that your data fail to satisfy this requirement, you should conduct a test for homogeneity before you attempt the ANOVA. Hartley's F-max test (see p. 338) allows you to use the sample variances from your data to determine whether there is evidence for any differences among the population variances. Remember, for the two-factor ANOVA, there is a separate sample for each cell in the data matrix. The test for homogeneity applies to all of these samples and the populations that they represent.

SUMMARY

1. A research study with two independent variables is called a two-factor design. Such a design can be diagramed as a matrix with the levels of one factor defining the rows and the levels of the other factor defining the columns. Each cell in the matrix corresponds to a specific combination of the two factors.

2. Traditionally, the two factors are identified as factor A and factor B. The purpose of the ANOVA is to determine whether there are any significant mean differences among the treatment conditions or cells in the experimental matrix. These treatment effects are classified as follows:

   a. The A-effect: Overall mean differences among the levels of factor A.

   b. The B-effect: Overall mean differences among the levels of factor B.

   c. The $A \times B$ interaction: Extra mean differences that are not accounted for by the main effects.

3. The two-factor ANOVA produces three F-ratios: one for factor A, one for factor B, and one for the $A \times B$ interaction. Each F-ratio has the same basic structure:

$$F = \frac{MS_{\text{treatment effect}}(\text{either } A \text{ or } B \text{ or } A \times B)}{MS_{\text{within treatments}}}$$

The formulas for the SS, df, and MS values for the two-factor ANOVA are presented in Figure 14.6.

Total
$$SS = \Sigma X^2 - \frac{G^2}{N}$$
$$df = N - 1$$

Between treatments
$$SS = \Sigma \frac{T^2}{n} - \frac{G^2}{N}$$
$$df = (\text{number of cells}) - 1$$

Within treatments
$$SS = \Sigma SS_{\text{each cell}}$$
$$df = \Sigma df_{\text{each cell}}$$

Factor A (rows)
$$SS = \Sigma \frac{T^2_{\text{ROW}}}{n_{\text{ROW}}} - \frac{G^2}{N}$$
$$df = (\text{levels of } A) - 1$$

Factor B (columns)
$$SS = \Sigma \frac{T^2_{\text{COL}}}{n_{\text{COL}}} - \frac{G^2}{N}$$
$$df = (\text{levels of } B) - 1$$

Interaction
SS is found by subtraction
df is found by subtraction

$$MS_{\text{factor}} = \frac{SS \text{ for the factor}}{df \text{ for the factor}} \qquad MS_{\text{within}} = \frac{SS_{\text{within treatments}}}{df_{\text{within treatments}}}$$
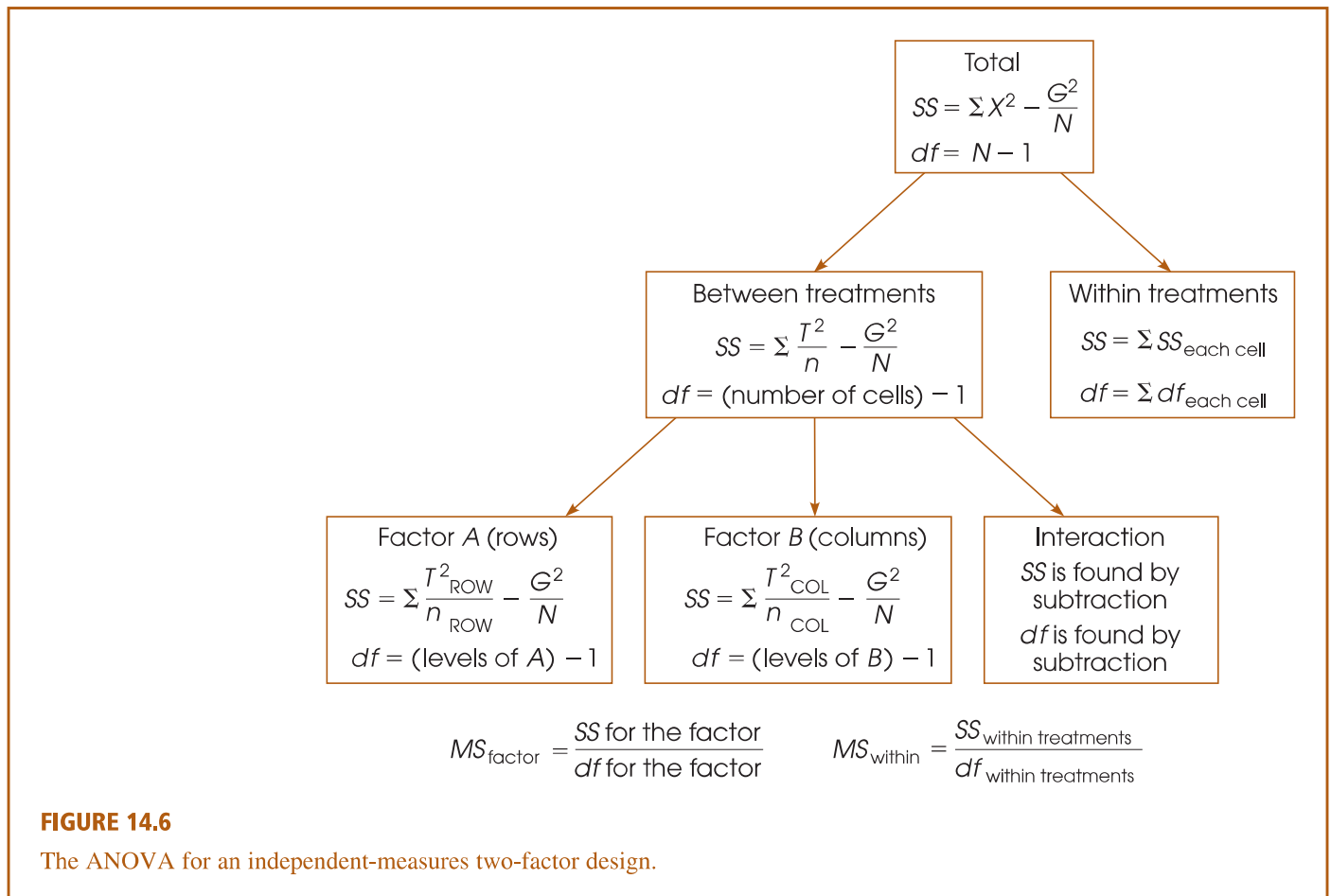
FIGURE 14.6

The ANOVA for an independent-measures two-factor design.

KEY TERMS

two-factor design (467)   main effect (469)

matrix (467)   interaction (470)

cells (467)

RESOURCES

Book Companion Website: www.cengage.com/psychology/gravetter

You can find a tutorial quiz and other learning exercises for Chapter 14 on the book companion website. The website also provides access to two workshops entitled *Two Way ANOVA* and *Factorial ANOVA* that both review the two-factor analysis presented in this chapter.

Improve your understanding of statistics with Aplia's auto-graded problem sets and immediate, detailed explanations for every question. To learn more, visit **www.aplia.com/statistics.**

Log in to CengageBrain to access the resources your instructor requires. For this book, you can access:

Psychology CourseMate brings course concepts to life with interactive learning, study, and exam preparation tools that support the printed textbook. A textbook-specific website, Psychology CourseMate includes an integrated interactive eBook and other interactive learning tools including quizzes, flashcards, and more.

Visit www.cengagebrain.com to access your account and purchase materials.

General instructions for using SPSS are presented in Appendix D. Following are detailed instructions for using SPSS to perform the **Two-Factor, Independent-Measures Analysis of Variance (ANOVA)** presented in this chapter.

Data Entry

1. The scores are entered into the SPSS data editor in a *stacked format,* which means that all of the scores from all of the different treatment conditions are entered in a single column (VAR00001).
2. In a second column (VAR00002), enter a code number to identify the level of factor *A* for each score. If factor *A* defines the rows of the data matrix, enter a 1 beside each score from the first row, enter a 2 beside each score from the second row, and so on.
3. In a third column (VAR00003), enter a code number to identify the level of factor *B* for each score. If factor *B* defines the columns of the data matrix, enter a 1 beside each score from the first column, enter a 2 beside each score from the second column, and so on.

   Thus, each row of the SPSS data editor has one score and two code numbers, with the score in the first column, the code for factor *A* in the second column, and the code for factor *B* in the third column.

Data Analysis

1. Click **Analyze** on the tool bar, select **General Linear Model,** and click on **Univariant.**
2. Highlight the column label for the set of scores (VAR0001) in the left box and click the arrow to move it into the **Dependent Variable** box.
3. One by one, highlight the column labels for the two factor codes and click the arrow to move them into the **Fixed Factors** box.
4. If you want descriptive statistics for each treatment, click on the **Options** box, select **Descriptives,** and click **Continue.**
5. Click **OK.**

SPSS Output

We used the SPSS program to analyze the data from the arousal-and-task-difficulty study in Example 14.1 and part of the program output is shown in Figure 14.7. The output begins with a table listing the factors (not shown in Figure 14.7), followed by a table showing descriptive statistics, including the mean and standard deviation for each cell, or treatment condition. The results of the ANOVA are shown in the table labeled **Tests of Between-Subjects Effects.** The top row (*Corrected Model*) presents the between-treatments *SS* and *df* values. The second row (*Intercept*) is not relevant for our purposes. The next three rows present the two main effects and the interaction (the *SS*, *df*, and *MS* values, as well as the *F*-ratio and the level of significance), with each factor identified by its column number from the SPSS data editor. The next row (*Error*) describes the error term (denominator of the *F*-ratio), and the final row (*Corrected Total*) describes the total variability for the entire set of scores. (Ignore the row labeled *Total.*)

FOCUS ON PROBLEM SOLVING

1. Before you begin a two-factor ANOVA, take time to organize and summarize the data. It is best if you summarize the data in a matrix with rows corresponding to the levels of one factor and columns corresponding to the levels of the other factor. In each cell of the matrix, show the number of scores (*n*), the total and mean for the cell, and the *SS* within the cell. Also compute the row totals and column totals that are needed to calculate main effects.

2. For a two-factor ANOVA, there are three separate *F*-ratios. These three *F*-ratios use the same error term in the denominator (MS_{within}). On the other hand, these *F*-ratios have different numerators and may have different *df* values associated with each of these numerators. Therefore, you must be careful when you look up the critical *F* values in the table. The two factors and the interaction may have different critical *F* values.

DEMONSTRATION 14.1

TWO-FACTOR ANOVA

The following data are representative of the results obtained in a research study examining the relationship between eating behavior and body weight (Schachter, 1968). The two factors in this study were:

1. The participant's weight (normal or obese)
2. The participant's state of hunger (full stomach or empty stomach)

Descriptive Statistics

Dependent Variable: VAR00001

VAR00002	VAR00003	Mean	Std. Deviation	N
1.00	1.00	3.0000	2.12132	5
	2.00	5.0000	2.64575	5
	3.00	10.0000	2.54951	5
	Total	6.0000	3.79850	15
2.00	1.00	1.0000	1.41421	5
	2.00	3.0000	2.23607	5
	3.00	2.0000	2.23607	5
	Total	2.0000	2.03540	15
1.00	1.00	2.0000	2.00000	10
	2.00	4.0000	2.53859	10
	3.00	6.0000	4.78423	10
	Total	4.0000	3.61987	30

Tests of Between-Subjects Effects

Dependent Variable: VAR00001

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	260.000	5	52.000	10.400	.000
Intercept	480.000	1	480.000	96.000	.000
VAR00002	120.000	1	120.000	24.000	.000
VAR00003	80.000	2	40.000	8.000	.002
VAR00002 * VAR00003	60.000	2	30.000	6.000	.008
Error	120.000	24	5.000		
Total	860.000	30			
Corrected Total	380.000	29			

FIGURE 14.7

Portions of the SPSS output for the two-factor ANOVA for the arousal-and-task-difficulty study in Example 14.1.

All participants were led to believe that they were taking part in a taste test for several types of crackers, and they were allowed to eat as many crackers as they wanted. The dependent variable was the number of crackers eaten by each participant. There were two specific predictions for this study. First, it was predicted that normal participants' eating behavior would be determined by their state of hunger. That is, people with empty

stomachs would eat more and people with full stomachs would eat less. Second, it was predicted that eating behavior for obese participants would not be related to their state of hunger. Specifically, it was predicted that obese participants would eat the same amount whether their stomachs were full or empty. Note that the researchers are predicting an interaction: The effect of hunger will be different for the normal participants and the obese participants. The data are as follows:

		Factor B: Hunger		
		Empty stomach	Full stomach	
Factor A: Weight	Normal	$n = 20$ $M = 22$ $T = 440$ $SS = 1540$	$n = 20$ $M = 15$ $T = 300$ $SS = 1270$	$T_{normal} = 740$
	Obese	$n = 20$ $M = 17$ $T = 340$ $SS = 1320$	$n = 20$ $M = 18$ $T = 360$ $SS = 1266$	$T_{normal} = 700$
		$T_{empty} = 780$	$T_{full} = 660$	

$G = 1440$
$N = 80$
$\Sigma X^2 = 31,836$

STEP 1   **State the hypotheses, and select alpha.**   For a two-factor study, there are three separate hypotheses, the two main effects and the interaction.

For factor $A,$ the null hypothesis states that there is no difference in the amount eaten for normal participants versus obese participants. In symbols,

$$H_0: \mu_{normal} = \mu_{obese}$$

For factor $B,$ the null hypothesis states that there is no difference in the amount eaten for full-stomach versus empty-stomach conditions. In symbols,

$$H_0: \mu_{full} = \mu_{empty}$$

For the $A \times B$ interaction, the null hypothesis can be stated two different ways. First, the difference in eating between the full-stomach and empty-stomach conditions will be the same for normal and obese participants. Second, the difference in eating between the normal and obese participants will be the same for the full-stomach and empty-stomach conditions. In more general terms,

$$H_0: \text{The effect of factor } A \text{ does not depend on the levels of factor } B \text{ (and } B \text{ does not depend on } A\text{).}$$

We use $\alpha = .05$ for all tests.

STEP 2   **The two-factor analysis.**   Rather than compute the df values and look up critical values for F at this time, we proceed directly to the ANOVA.

STAGE 1   The first stage of the analysis is identical to the independent-measures ANOVA presented in Chapter 12, where each cell in the data matrix is considered a separate treatment condition.

$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{N}$$

$$= 31{,}836 - \frac{1440^2}{80} = 5916$$

$$SS_{\text{within treatments}} = \Sigma SS_{\text{inside each treatment}} = 1540 + 1270 + 1320 + 1266 = 5396$$

$$SS_{\text{between treatments}} = \Sigma \frac{T^2}{n} - \frac{G^2}{N}$$

$$= \frac{440^2}{20} + \frac{300^2}{20} + \frac{340^2}{20} + \frac{360^2}{20} - \frac{1440^2}{80}$$

$$= 520$$

The corresponding degrees of freedom are

$$df_{\text{total}} = N - 1 = 79$$

$$df_{\text{within treatments}} = \Sigma df = 19 + 19 + 19 + 19 = 76$$

$$df_{\text{between treatments}} = \text{number of treatments} - 1 = 3$$

STAGE 2 The second stage of the analysis partitions the between-treatments variability into three components: the main effect for factor A, the main effect for factor B, and the $A \times B$ interaction.

For factor A (normal/obese),

$$SS_A = \Sigma \frac{T^2_{\text{ROWS}}}{n_{\text{ROWS}}} - \frac{G^2}{N}$$

$$= \frac{740^2}{40} + \frac{700^2}{40} - \frac{1440^2}{80}$$

$$= 20$$

For factor B (full/empty),

$$SS_B = \Sigma \frac{T^2_{\text{COLS}}}{n_{\text{COLS}}} - \frac{G^2}{N}$$

$$= \frac{780^2}{40} + \frac{660^2}{40} - \frac{1440^2}{80}$$

$$= 180$$

For the $A \times B$ interaction,

$$SS_{A \times B} = SS_{\text{between treatments}} - SS_A \times SS_B$$

$$= 520 - 20 - 180$$

$$= 320$$

The corresponding degrees of freedom are

$$df_A = \text{number of rows} - 1 = 1$$

$$df_B = \text{number of columns} - 1 = 1$$

$$df_{A \times B} = df_{\text{between treatments}} - df_A - df_B$$

$$= 3 - 1 - 1$$

$$= 1$$

The *MS* values needed for the *F*-ratios are

$$MS_A = \frac{SS_A}{df_A} = \frac{20}{1} = 20$$

$$MS_B = \frac{SS_B}{df_B} = \frac{180}{1} = 180$$

$$MS_{A \times B} = \frac{SS_{A \times B}}{df_{A \times B}} = \frac{320}{1} = 320$$

$$MS_{\text{within treatments}} = \frac{SS_{\text{within treatments}}}{df_{\text{within treatments}}} = \frac{5396}{76} = 71$$

Finally, the *F*-ratios are

$$F_A = \frac{MS_A}{MS_{\text{within treatments}}} = \frac{20}{71} = 0.28$$

$$F_B = \frac{MS_B}{MS_{\text{within treatments}}} = \frac{180}{71} = 2.54$$

$$F_{A \times B} = \frac{MS_{A \times B}}{MS_{\text{within treatments}}} = \frac{320}{71} = 4.51$$

S T E P 3   Make a decision and state a conclusion.   All three *F*-ratios have *df* = 1, 76.
With $\alpha = .05$, the critical *F* value is 3.98 for all three tests.

For these data, factor *A* (weight) has no significant effect; $F(1, 76) = 0.28$.
Statistically, there is no difference in the number of crackers eaten by normal versus
obese participants.

Similarly, factor *B* (fullness) has no significant effect; $F(1, 76) = 2.54$.
Statistically, the number of crackers eaten by full participants is no different from the
number eaten by hungry participants. (*Note:* This conclusion concerns the combined
group of normal and obese participants. The interaction concerns these two groups
separately.)

These data produce a significant interaction; $F(1, 76) = 4.51$, $p < .05$. This means
that the effect of fullness does depend on weight. A closer look at the original data shows
that the degree of fullness did affect the normal participants, but it had no effect on the
obese participants.

DEMONSTRATION 14.2

MEASURING EFFECT SIZE FOR THE TWO-FACTOR ANOVA

Effect size for each main effect and for the interaction is measured by eta squared (η^2), the percentage of variance explained by the specific main effect or interaction. In each case, the variability that is explained by other sources is removed before the percentage is computed. For the two-factor ANOVA in Demonstration 14.1,

$$\text{For factor } A, \ \eta^2 = \frac{SS_A}{SS_{\text{total}} - SS_B - SS_{A\times B}} = \frac{20}{5916 - 180 - 320} = 0.004 \left(\text{or } 0.4\%\right)$$

$$\text{For factor } B, \ \eta^2 = \frac{SS_B}{SS_{\text{total}} - SS_A - SS_{A\times B}} = \frac{180}{5916 - 20 - 320} = 0.032 \left(\text{or } 3.2\%\right)$$

$$\text{For } A \times B, \ \eta^2 = \frac{SS_{A\times B}}{SS_{\text{total}} - SS_A - SS_B} = \frac{320}{5916 - 20 - 180} = 0.056 \left(\text{or } 5.6\%\right)$$

PROBLEMS

1. Define each of the following terms:
   a. Factor
   b. Level
   c. Two-factor study

2. The structure of a two-factor study can be presented as a matrix with the levels of one factor determining the rows and the levels of the second factor determining the columns. With this structure in mind, describe the mean differences that are evaluated by each of the three hypothesis tests that make up a two-factor ANOVA.

3. Briefly explain what happens during the second stage of the two-factor ANOVA.

4. For the data in the following matrix:

	No Treatment	Treatment	
Male	$M = 5$	$M = 3$	Overall $M = 4$
Female	$M = 9$	$M = 13$	Overall $M = 11$
	overall $M = 7$	overall $M = 8$	

   a. Which two means are compared to describe the treatment main effect?
   b. Which two means are compared to describe the gender main effect?
   c. Is there an interaction between gender and treatment? Explain your answer.

5. The following matrix presents the results from an independent-measures, two-factor study with a sample of $n = 10$ participants in each treatment condition. Note that one treatment mean is missing.

	Factor B	
	B_1	B_2
Factor A A_1	$M = 20$	$M = 30$
A_2	$M = 40$	

   a. What value for the missing mean would result in no main effect for factor A?
   b. What value for the missing mean would result in no main effect for factor B?
   c. What value for the missing mean would result in no interaction?

6. The following matrix presents the results of a two-factor study with $n = 10$ scores in each of the six treatment conditions. Note that one of the treatment means is missing.

	Factor B		
	B_1	B_2	B_3
Factor A A_1	$M = 10$	$M = 20$	$M = 40$
A_2	$M = 20$	$M = 30$	

a. What value for the missing mean would result in no main effect for factor A?

b. What value for the missing mean would result in no interaction?

7. For the data in the following graph:
   a. Is there a main effect for the treatment factor?
   b. Is there a main effect for the age factor?
   c. Is there an interaction between age and treatment?

8. A researcher conducts an independent-measures, two-factor study using a separate sample of $n = 15$ participants in each treatment condition. The results are evaluated using an ANOVA and the researcher reports an F-ratio with $df = 1, 84$ for factor A, and an F-ratio with $df = 2, 84$ for factor B.
   a. How many levels of factor A were used in the study?
   b. How many levels of factor B were used in the study?
   c. What are the df values for the F-ratio evaluating the interaction?

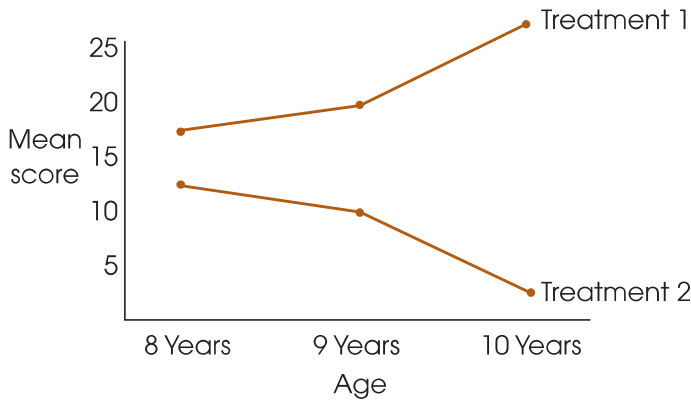9. The following results are from an independent-measures, two-factor study with $n = 10$ participants in each treatment condition.

	Factor B	
	B_1	B_2
A_1	T = 40 M = 4 SS = 50	T = 10 M = 1 SS = 30
A_2	T = 50 M = 5 SS = 60	T = 20 M = 2 SS = 40

(Factor A labels the rows)

$$N = 40$$
$$G = 120$$
$$\Sigma X^2 = 640$$

a. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the main effects and the interaction.

b. Compute η^2 to measure the effect size for each of the main effects and the interaction.

10. The following results are from an independent-measures, two-factor study with $n = 5$ participants in each treatment condition.

	Factor B		
	B_1	B_2	B_3
A_1	T = 25 M = 5 SS = 30	T = 40 M = 8 SS = 38	T = 70 M = 14 SS = 46
A_2	T = 15 M = 3 SS = 22	T = 20 M = 4 SS = 26	T = 40 M = 8 SS = 30

(Factor A labels the rows)

$$N = 40$$
$$G = 120$$
$$\Sigma X^2 = 640$$

a. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the main effects and the interaction.

b. Test the simple main effects using $\alpha = .05$ to evaluate the mean difference between treatment A_1 and A_2 for each level of factor B.

11. A researcher conducts an independent-measures, two-factor study with two levels of factor A and three levels of factor B, using a sample of $n = 12$ participants in each treatment condition.
    a. What are the df values for the F-ratio evaluating the main effect of factor A?
    b. What are the df values for the F-ratio evaluating the main effect of factor B?
    c. What are the df values for the F-ratio evaluating the interaction?

12. Most sports injuries are immediate and obvious, like a broken leg. However, some can be more subtle, like the neurological damage that may occur when soccer players repeatedly head a soccer ball. To examine long-term effects of repeated heading, Downs and Abwender (2002) examined two different age groups of soccer players and swimmers. The dependent variable was performance on a conceptual thinking task. Following are hypothetical data, similar to the research results.
    a. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the main effects and interaction.
    b. Calculate the effects size (η^2) for the main effects and the interaction.
    c. Briefly describe the outcome of the study.

	Factor B: Age	
	College	Older
Soccer	$n = 20$ $M = 9$ $T = 180$ $SS = 380$	$n = 20$ $M = 4$ $T = 80$ $SS = 390$
Swimming	$n = 20$ $M = 9$ $T = 180$ $SS = 350$	$n = 20$ $M = 8$ $T = 160$ $SS = 400$

Factor A: Sport

$$\Sigma X^2 = 6360$$

Source	SS	df	MS	
Between treatments	60	___		
Factor A	___	___	5	$F = $ ___
Factor B	___	___	___	$F = $ ___
$A \times B$ Interaction	25	___	___	$F = $ ___
Within treatments	___	___	2.5	
Total	___	___		

13. Some people like to pour beer gently down the side of the glass to preserve bubbles. Others splash it down the center to release the bubbles into a foamy head and free the aromas. Champagne, however is best when the bubbles remain concentrated in the wine. A group of French scientists recently verified the difference between the two pouring methods by measuring the amount of bubbles in each glass of champagne poured two different ways and at three different temperatures (Liger-Belair, 2010). The following data present the pattern of results obtained in the study.

	Champagne Temperature (°F)		
	40°	46°	52°
Gentle Pour	$n = 10$ $M = 7$ $SS = 64$	$n = 10$ $M = 3$ $SS = 57$	$n = 10$ $M = 2$ $SS = 47$
Splashing Pour	$n = 10$ $M = 5$ $SS = 56$	$n = 10$ $M = 1$ $SS = 54$	$n = 10$ $M = 0$ $SS = 46$

  a. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the mean differences.

  b. Briefly explain how temperature and pouring influence the bubbles in champagne according to this pattern of results.

14. The following table summarizes the results from a two-factor study with 2 levels of factor A and 3 levels of factor B using a separate sample of $n = 8$ participants in each treatment condition. Fill in the missing values. (*Hint:* Start with the df values.)

15. The following table summarizes the results from a two-factor study with 3 levels of factor A and 3 levels of factor B using a separate sample of $n = 9$ participants in each treatment condition. Fill in the missing values. (*Hint:* Start with the df values.)

Source	SS	df	MS	
Between treatments	144	___		
Factor A	___	___	18	$F = $ ___
Factor B	___	___	___	$F = $ ___
$A \times B$ Interaction	___	___	___	$F = 7.0$
Within treatments	___	___	___	
Total	360	___		

16. The Preview section for this chapter described a two-factor study examining performance under two audience conditions (factor B) for high and low self-esteem participants (factor A). The following summary table presents possible results from the analysis of that study. Assuming that the study used a separate sample of $n = 15$ participants in each treatment condition (each cell), fill in the missing values in the table. (*Hint*: Start with the df values.)

Source	SS	df	MS	
Between treatments	67	___		
Audience	___	___	___	$F = $ ___
Self-esteem	29	___	___	$F = $ ___
Interaction	___	___	___	$F = 5.50$
Within treatments	___	___	4	
Total	___	___		

17. The following table summarizes the results from a two-factor study with 2 levels of factor A and 3 levels of factor B using a separate sample of $n = 11$ participants in each treatment condition. Fill in the missing values. (*Hint:* Start with the df values.)

Source	SS	df	MS	
Between treatments	____	____		
Factor A	____	____	____	$F = 7$
Factor B	____	____	____	$F = 8$
A × B Interaction	____	____	____	$F = 3$
Within treatments	240	____	____	
Total	____	____		

18. The following data are from a two-factor study examining the effects of two treatment conditions on males and females.
    a. Use an ANOVA with $\alpha = .05$ for all tests to evaluate the significance of the main effects and the interaction.
    b. Compute η^2 to measure the size of the effect for each main effect and the interaction.

Treatments

	I	II	
Male	3	2	
	8	8	
	9	7	$T_{male} = 48$
	4	7	
	$M = 6$	$M = 6$	
	$T = 24$	$T = 24$	$N = 16$
Factor A:	$SS = 26$	$SS = 22$	$G = 96$
Gender			$\Sigma X^2 = 806$
Female	0	12	
	0	6	
	2	9	
	6	13	
	$M = 2$	$M = 10$	$T_{female} = 48$
	$T = 8$	$T = 40$	
	$SS = 24$	$SS = 30$	
	$T_I = 32$	$T_{II} = 64$	

19. The following data are from a two-factor study examining the effects of three treatment conditions on males and females.
    a. Use an ANOVA with $\alpha = .05$ for all tests to evaluate the significance of the main effects and the interaction.
    b. Test the simple main effects using $\alpha = .05$ to evaluate the mean difference between males and females for each of the three treatments.

Factor B Treatments

		I	II	III	
Male		1	7	9	
		2	2	11	
		6	9	7	$T_{male} = 54$
		$M = 3$	$M = 6$	$M = 9$	
		$T = 9$	$T = 18$	$T = 27$	
		$SS = 14$	$SS = 26$	$SS = 8$	
Factor A:					$N = 18$
Gender		3	10	16	$G = 144$
		1	11	18	$\Sigma X^2 = 1608$
		5	15	11	
Female		$M = 3$	$M = 12$	$M = 15$	$T_{female} = 90$
		$T = 9$	$T = 36$	$T = 45$	
		$SS = 8$	$SS = 14$	$SS = 26$	

20. Mathematics word problems can be particularly difficult, especially for primary-grade children. A recent study investigated a combination of techniques for teaching students to master these problems (Fuchs, Fuchs, Craddock, Hollenbeck, Hamlett, & Schatschneider, 2008). The study investigated the effectiveness of small-group tutoring and the effectiveness of a classroom instruction technique known as "hot math." The hot-math program teaches students to recognize types or categories of problems so that they can generalize skills from one problem to another. The following data are similar to the results obtained in the study. The dependent variable is a math test score for each student after 16 weeks in the study.

	No Tutoring	With Tutoring
Traditional Instruction	3	9
	6	4
	2	5
	2	8
	4	4
	7	6
Hot-Math Instruction	7	8
	7	12
	2	9
	6	13
	8	9
	6	9

a. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the significance of the main effects and the interaction.

b. Calculate the η^2 values to measure the effect size for the two main effects.

c. Describe the pattern of results. (Is tutoring significantly better than no tutoring? Is traditional classroom instruction significantly different from hot math? Does the effect of tutoring depend on the type of classroom instruction?)

21. In Chapter 12 (p. 432), we described a study reporting that college students who are on Facebook (or have it running in the background) while studying had lower grades than students who did not use the social network (Kirschner & Karpinski, 2010). A researcher would like to know if the same result extends to students in lower grade levels. The researcher planned a two-factor study comparing Facebook users with non-users for middle school students, high school students, and college students. For consistency across groups, grades were converted into six categories, numbered 0 to 5 from low to high. The results are presented in the following matrix.

a. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the mean differences.

b. Describe the pattern of results.

	Middle School	High School	College
User	3	5	5
	5	5	4
	5	2	2
	3	4	5
Non-user	5	1	1
	3	2	0
	2	3	0
	2	2	3

22. In Chapter 11, we described a research study in which the color red appeared to increase men's attraction to women (Elliot & Niesta, 2008). The same researchers have published other results showing that red also increases women's attraction to men but does not appear to affect judgments of same sex individuals (Elliot, et al., 2010). Combining these results into one study produces a two-factor design in which men judge photographs of both women and men, which are shown on both red and white backgrounds. The dependent variable is a rating of attractiveness for the person shown in the photograph. The study uses a separate group of participants for each condition. The following table presents data similar to the results from previous research.

a. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the main effects and the interaction.

		Person Shown in Photograph	
		Female	Male
Background Color for Photograph	White	$n = 10$ $M = 4.5$ $SS = 6$	$n = 10$ $M = 4.4$ $SS = 7$
	Red	$n = 10$ $M = 7.5$ $SS = 9$	$n = 10$ $M = 4.6$ $SS = 8$

b. Describe the effect of background color on judgments of males and females.

23. In the Preview section of this chapter, we presented an experiment that examined the effect of an audience on the performance of two different personality types. Data from this experiment are as follows. The dependent variable is the number of errors made by each participant.

		No Audience	Audience
	High	3	9
		6	4
		2	5
		2	8
Self-Esteem		4	4
		7	6
	Low	7	10
		7	14
		2	11
		6	15
		8	11
		6	11

a. Use an ANOVA with $\alpha = .05$ to evaluate the data. Describe the effect of the audience and the effect of self-esteem on performance.

b. Calculate the effect size (η^2) for each main effect and for the interaction.

After completing this part, you should be able to perform an ANOVA to evaluate the significance of mean differences in three research situations. These include:

1. The single-factor independent-measures design introduced in Chapter 12.
2. The single-factor repeated-measures design introduced in Chapter 13.
3. The two-factor independent-measures design introduced in Chapter 14.

In this part we introduce three applications of ANOVA that use an F-ratio statistic to evaluate the mean differences among two or more populations. In each case, the F-ratio has the following structure:

$$F = \frac{\text{variance between treatments}}{\text{variance from random unsystematic sources}}$$

The numerator of the F-ratio measures the mean differences that exist from one treatment condition to another, including any systematic differences caused by the treatments. The denominator measures the differences that exist when there are no systematic factors that cause one score to be different from another. The F-ratio is structured so that the numerator and denominator are measuring exactly the same variance when the null hypothesis is true and there are no systematic treatment effects. In this case, the F-ratio should have a value near 1.00. Thus, an F-ratio near 1.00 is evidence that the null hypothesis is true. Similarly, an F-ratio that is much larger than 1.00 provides evidence that a systematic treatment effect does exist and the null hypothesis should be rejected.

For independent-measures designs, either single-factor or two-factor, the denominator of the F-ratio is obtained by computing the variance within treatments. Inside each treatment condition all participants are treated exactly the same, so there are no systematic treatment effects that cause the scores to vary.

For a repeated-measures design, the same individuals are used in every treatment condition, so any differences between treatments cannot be caused by individual differences. Thus, the numerator of the F-ratio does not include any individual differences. Therefore, individual differences must also be eliminated from the denominator to balance the F-ratio. As a result, the repeated-measures ANOVA is a two-stage process. The first stage separates the between-treatments variance (numerator) and the within-treatments variance. The second stage removes the systematic individual differences from the within-treatments variance to produce the appropriate denominator for the F-ratio.

For a two-factor design, the mean differences between treatments can be caused by either of the two factors or by specific combinations of factors. The goal of the

ANOVA is to separate these possible treatment effects so that each can be evaluated independent of the others. To accomplish this, the two-factor ANOVA is a two-stage process. The first stage separates the between-treatments variance and the within-treatments variance (denominator). The second stage analyzes the between-treatments variance into three components: the main effect from the first factor, the main effect from the second factor, and the interaction.

Note that the repeated-measures ANOVA and the two-factor ANOVA are both two-stage processes. Both begin by separating the total variance into between-treatments and within-treatments variance. However, the second stages of these two ANOVAs serve different purposes and focus on different components. The repeated-measures ANOVA focuses on the within-treatments variance and removes the individual differences to obtain the error variance. The two-factor ANOVA separates the between-treatments variance into the two main effects and the interaction.

REVIEW EXERCISES

1. Recent research indicates that the effectiveness of anti-depressant medication is directly related to the severity of the depression (Khan, Brodhead, Kolts, & Brown, 2005). Based on pre-treatment depression scores, patients were divided into four groups by level of depression. After receiving the antidepressant medication, depression scores were measured again and the amount of improvement was recorded for each patient. The following data are similar to the results of the study.

   a. Do the data indicate significant differences among the four levels of severity? Test with $\alpha = .05$.
   b. Compute η^2, the percentage of variance explained by the group differences.
   c. Write a sentence demonstrating how the outcome of the hypothesis test and the measure of effect size would appear in a research report.

Low Moderate	High Moderate	Moderately Severe	Severe	
0	1	4	5	$N = 16$
2	3	6	6	$G = 48$
2	2	2	6	$\Sigma X^2 = 204$
0	2	4	3	
$M = 1$	$M = 2$	$M = 4$	$M = 5$	
$T = 4$	$T = 8$	$T = 16$	$T = 20$	
$SS = 4$	$SS = 2$	$SS = 8$	$SS = 6$	

2. Loss of hearing can be a significant problem for older adults. Although hearing aids can correct the physical

problem, people who have lived with hearing impairment often develop poor communication strategies and social skills. To address this problem, a home education program has been developed to help people who are receiving hearing aids for the first time. The program emphasizes communication skills. To evaluate the program, overall quality of life and satisfaction were measured before treatment, again at the end of the training program, and once more at a 6-month follow-up (Kramer, Allessie, Dondorp, Zekveld, & Kapteyn, 2005). Data similar to the results obtained in the study are shown below.

	Quality-of-Life Scores			
Person	Before	After	6 Months	
A	3	7	8	$N = 12$
B	0	5	7	$G = 60$
C	4	9	5	$\Sigma X^2 = 384$
D	1	7	4	
	$T = 8$	$T = 28$	$T = 24$	
	$SS = 10$	$SS = 8$	$SS = 10$	

a. Do the data indicate a significant improvement in the quality of life following the training program? Test at the .05 level of significance.
b. Calculate η^2 to measure the size of the effect.
c. Write a sentence demonstrating how the outcome of the hypothesis test and the measure of effect size would appear in a research report.

3. Briefly describe what is meant by an interaction between factors in a two-factor research study.
4. A recent study of driving behavior suggests that self-reported measures of high driving skills and low ratings of safety skills create a dangerous combination (Sümer, Özkan, & Lajunen, 2006). (*Note:* Those who rate themselves as highly skilled drivers are probably overly confident.) Drivers were classified as high or low in self-rated skill based on responses to a driver-skill inventory, then classified as high or low in safety skill based on responses to a driver-aggression scale. An overall measure of driving risk was obtained by combining several variables such as number of accidents, tickets, tendency to speed, and tendency to pass other cars. The following data represent results similar to those obtained in the study. Use a two-factor ANOVA with $\alpha = .05$ to evaluate the results.

		Self-Rated Driving Skill		
		Low	High	
Driving Safety	Low	$n = 8$ $M = 5$ $T = 40$ $SS = 52$	$n = 8$ $M = 8.5$ $T = 68$ $SS = 71$	$N = 32$ $G = 160$ $\Sigma X^2 = 1151$
	High	$n = 8$ $M = 3$ $T = 24$ $SS = 34$	$n = 8$ $M = 3.5$ $T = 28$ $SS = 46$	

PART

V

Correlations and Nonparametric Tests

Back in Chapter 1 we stated that the primary goal of science is to establish relationships between variables. Until this point, the statistics we have presented all attempt to accomplish this goal by comparing *groups of scores* using *means and variances* as the basic statistical measures. Typically, one variable is used to define the groups, and a second variable is measured to obtain a set of scores within each group. Means and variances are then computed for the scores, and the sample means are used to test hypotheses about population means. If the hypothesis test indicates a significant mean difference, then we conclude that there is a relationship between the variables.

However, many research situations do not involve comparing groups, and many do not produce data that allow you to calculate means and variances. For example, a researcher can investigate the relationship between two variables (for example, IQ and creativity) by measuring both variables within a single group of individuals. Also, the measurement procedure may not produce numerical scores. For example, participants can indicate their color preferences by simply picking a favorite color or by ranking several choices. Without numerical scores, it is impossible to calculate means and variances. Instead, the data consist of proportions or frequencies. For example, a research study may investigate what proportion of people select red as their favorite color and whether this proportion is different for introverted people compared with extroverted people.

Notice that these new research situations are still asking questions about the relationships between variables, and they are still using sample data to make inferences about populations. However, they are no longer comparing groups and they are no longer based on means and variances. In this part, we introduce the statistical methods that have been developed for these other kinds of research.