CHAPTER

12

Introduction to Analysis of Variance

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.

- Variability (Chapter 4)
- Sum of squares
- Sample variance
- Degrees of freedom
- Introduction to hypothesis testing (Chapter 8)
- The logic of hypothesis testing
- Independent-measures *t* statistic (Chapter 10)

Preview

“But I read the chapter four times! How could I possibly have failed the exam?!”

Most of you probably have had the experience of reading a textbook and suddenly realizing that you have no idea of what was said on the past few pages. Although you have been reading the words, your mind has wandered off, and the meaning of the words has never reached memory. In an influential paper on human memory, Craik and Lockhart (1972) proposed a *levels of processing* theory of memory that can account for this phenomenon. In general terms, this theory says that all perceptual and mental processing leaves behind a memory trace. However, the quality of the memory trace depends on the level or the depth of the processing. If you superficially skim the words in a book, your memory also is superficial. On the other hand, when you think about the meaning of the words and try to understand what you are reading, the result is a good, substantial memory that should serve you well on exams. In general, deeper processing results in better memory.

Rogers, Kuiper, and Kirker (1977) conducted an experiment demonstrating the effect of levels of processing. Participants in this experiment were shown lists of words and asked to answer questions about each word. The questions were designed to require different levels of processing, from superficial to deep. In one experimental condition, participants were simply asked to judge the physical characteristics of each printed word (“Is it printed in capital letters or lowercase letters?”). A second condition asked about the sound of each word (“Does it rhyme with ‘boat’?”). In a third condition, participants were required to process the meaning of each word (“Does it have the same meaning as ‘attractive’?”). The final condition required participants to understand each word and relate its meaning to themselves (“Does this word describe you?”). After going through the complete list, all participants were given a surprise memory test. As you can see in Figure 12.1, deeper processing resulted in better memory. Remember that the participants were not trying to memorize the words; they were simply reading through the list answering questions. However, the more they processed and understood the words, the better they recalled the words on the test.

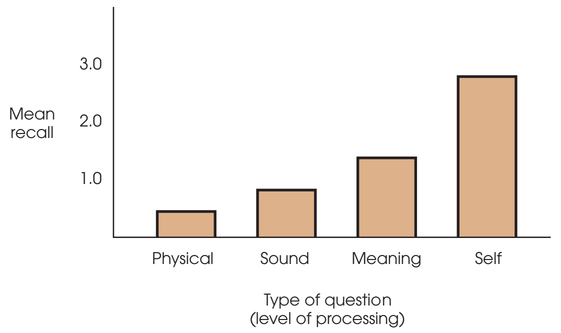


FIGURE 12.1

Mean recall as a function of the level of processing.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of personality and Social Psychology*, 35, 677–688. Copyright (1977) by the American Psychological Association. Adapted by permission of the author.

The Problem: In terms of human memory, the Rogers, Kuiper, and Kirker experiment is notable because it demonstrates the importance of “self” in memory. You are most likely to remember material that is directly related to you. In terms of statistics, however, this study is notable because it compares four different treatment conditions in a single experiment. We now have four different means and need a hypothesis test to evaluate the mean differences. Unfortunately, the *t* tests introduced in Chapter 10 and 11 are limited to comparing only two treatments. A new hypothesis test is needed for this kind of data.

The Solution: In this chapter we introduce a new hypothesis test known as *analysis of variance* that is designed to evaluate the mean differences from research studies producing two or more sample means. Although “two or more” may seem like a small step from “two,” this new hypothesis testing procedure provides researchers with a tremendous gain in experimental sophistication. In this chapter, and the two that follow, we examine some of the many applications of analysis of variance.

12.1 INTRODUCTION

Analysis of variance (ANOVA) is a hypothesis-testing procedure that is used to evaluate mean differences between two or more treatments (or populations). As with all inferential procedures, ANOVA uses sample data as the basis for drawing general conclusions about populations. It may appear that ANOVA and *t* tests are simply two different ways of doing exactly the same job: testing for mean differences. In some respects, this is true—both tests use sample data to test hypotheses about population means. However, ANOVA has a tremendous advantage over *t* tests. Specifically, *t* tests are limited to situations in which there are only two treatments to compare. The major advantage of ANOVA is that it can be used to compare *two or more treatments*. Thus, ANOVA provides researchers with much greater flexibility in designing experiments and interpreting results.

Figure 12.2 shows a typical research situation for which ANOVA would be used. Note that the study involves three samples representing three populations. The goal of the analysis is to determine whether the mean differences observed among the samples provide enough evidence to conclude that there are mean differences among the three populations. Specifically, we must decide between two interpretations:

1. There really are no differences between the populations (or treatments). The observed differences between the sample means are caused by random, unsystematic factors (sampling error) that differentiate one sample from another.
2. The populations (or treatments) really do have different means, and these population mean differences are responsible for causing systematic differences between the sample means.

You should recognize that these two interpretations correspond to the two hypotheses (null and alternative) that are part of the general hypothesis-testing procedure.

TERMINOLOGY IN ANOVA

Before we continue, it is necessary to introduce some of the terminology that is used to describe the research situation shown in Figure 12.2. Recall (from Chapter 1) that when a researcher manipulates a variable to create the treatment conditions in an experiment, the variable is called an *independent variable*. For example, Figure 12.2 could represent

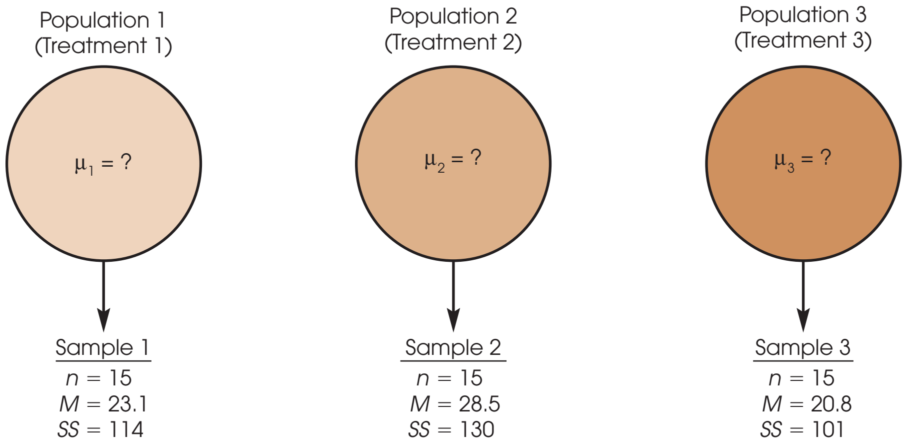


FIGURE 12.2
A typical situation in which ANOVA would be used. Three separate samples are obtained to evaluate the mean differences among three populations (or treatments) with unknown means.

a study examining driving performance under three different telephone conditions: driving with no phone, talking on a hands-free phone, and talking on a hand-held phone. Note that the three conditions are created by the researcher. On the other hand, when a researcher uses a nonmanipulated variable to designate groups, the variable is called a *quasi-independent variable*. For example, the three groups in Figure 12.2 could represent 6-year-old, 8-year-old, and 10-year-old children. In the context of ANOVA, an independent variable or a quasi-independent variable is called a *factor*. Thus, Figure 12.2 could represent an experimental study in which the telephone condition is the factor being evaluated or it could represent a nonexperimental study in which age is the factor being examined.

DEFINITION

In ANOVA, the variable (independent or quasi-independent) that designates the groups being compared is called a **factor**.

In addition, the individual groups or treatment conditions that are used to make up a factor are called the *levels* of the factor. For example, a study that examined performance under three different telephone conditions would have three levels of the factor.

DEFINITION

The individual conditions or values that make up a factor are called the **levels** of the factor.

Like the t tests presented in Chapters 10 and 11, ANOVA can be used with either an independent-measures or a repeated-measures design. Recall that an independent-measures design means that there is a separate group of participants for each of the treatments (or populations) being compared. In a repeated-measures design, on the other hand, the same group is tested in all of the different treatment conditions. In addition, ANOVA can be used to evaluate the results from a research study that involves more than one factor. For example, a researcher may want to compare two different therapy techniques, examining their immediate effectiveness as well as the persistence of their effectiveness over time. In this situation, the research study could involve two different groups of participants, one for each therapy, and measure each group at several different points in time. The structure of this design is shown in Figure 12.3. Notice that the study uses two factors, one independent-measures factor and one repeated-measures factor:

1. Factor 1: Therapy technique. A separate group is used for each technique (independent measures).
2. Factor 2: Time. Each group is tested at three different times (repeated measures).

In this case, the ANOVA would evaluate mean differences between the two therapies as well as mean differences between the scores obtained at different times. A study that combines two factors, like the one in Figure 12.3, is called a *two-factor design* or a *factorial design*.

The ability to combine different factors and to mix different designs within one study provides researchers with the flexibility to develop studies that address scientific questions that could not be answered by a single design using a single factor.

Although ANOVA can be used in a wide variety of research situations, this chapter introduces ANOVA in its simplest form. Specifically, we consider only *single-factor* designs. That is, we examine studies that have only one independent variable (or only one quasi-independent variable). Second, we consider only *independent-measures* designs; that is, studies that use a separate group of participants for each treatment condition. The basic logic and procedures that are presented in this chapter form the foundation for more complex applications of

		TIME		
		Before Therapy	After Therapy	6 Months After Therapy
THERAPY TECHNIQUE	Therapy I (Group 1)	Scores for group 1 measured before Therapy I	Scores for group 1 measured after Therapy I	Scores for group 1 measured 6 months after Therapy I
	Therapy II (Group 2)	Scores for group 2 measured before Therapy II	Scores for group 2 measured after Therapy II	Scores for group 2 measured 6 months after Therapy II

FIGURE 12.3

A research design with two factors. The research study uses two factors: One factor uses two levels of therapy technique (I versus II), and the second factor uses three levels of time (before, after, and 6 months after). Also notice that the therapy factor uses two separate groups (independent measures) and the time factor uses the same group for all three levels (repeated measures).

ANOVA. For example, in Chapter 13, we extend the analysis to single-factor, repeated-measures designs and in Chapter 14, we introduce two-factor designs. But for now, in this chapter, we limit our discussion of ANOVA to *single-factor, independent-measures* research studies.

STATISTICAL HYPOTHESES FOR ANOVA

The following example introduces the statistical hypotheses for ANOVA. Suppose that a researcher examined driving performance under three different telephone conditions: no phone, a hands-free phone, and a hand-held phone. Three samples of participants are selected, one sample for each treatment condition. The purpose of the study is to determine whether using a telephone affects driving performance. In statistical terms, we want to decide between two hypotheses: the null hypothesis (H_0), which states that the telephone condition has no effect, and the alternative hypothesis (H_1), which states that the telephone condition does affect driving. In symbols, the null hypothesis states

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

In words, the null hypothesis states that the telephone condition has no effect on driving performance. That is, the population means for the three telephone conditions are all the same. In general, H_0 states that there is no treatment effect.

The alternative hypothesis states that the population means are not all the same:

H_1: There is at least one mean difference among the populations.

In general, H_1 states that the treatment conditions are not all the same; that is, there is a real treatment effect. As always, the hypotheses are stated in terms of population parameters, even though we use sample data to test them.

Notice that we are not stating a specific alternative hypothesis. This is because many different alternatives are possible, and it would be tedious to list them all.

One alternative, for example, is that the first two populations are identical, but that the third is different. Another alternative states that the last two means are the same, but that the first is different. Other alternatives might be

H_1: $\mu_1 \neq \mu_2 \neq \mu_3$ (All three means are different.)

H_1: $\mu_1 = \mu_3$, but μ_2 is different.

We should point out that a researcher typically entertains only one (or at most a few) of these alternative hypotheses. Usually a theory or the outcomes of previous studies dictate a specific prediction concerning the treatment effect. For the sake of simplicity, we state a general alternative hypothesis rather than try to list all of the possible specific alternatives.

THE TEST STATISTIC FOR ANOVA

The test statistic for ANOVA is very similar to the independent-measures t statistic used in Chapter 10. For the t statistic, we first computed the standard error, which measures how much difference is expected between two sample means if there is no treatment effect (that is, if H_0 is true). Then we computed the t statistic with the following structure:

$$t = \frac{\text{obtained difference between two sample means}}{\text{standard error (the difference expected with no treatment effect)}}$$

For ANOVA, however, we want to compare differences among two *or more* sample means. With more than two samples, the concept of “difference between sample means” becomes difficult to define or measure. For example, if there are only two samples and they have means of $M = 20$ and $M = 30$, then there is a 10-point difference between the sample means. Suppose, however, that we add a third sample with a mean of $M = 35$. Now how much difference is there between the sample means? It should be clear that we have a problem. The solution to this problem is to use variance to define and measure the size of the differences among the sample means. Consider the following two sets of sample means:

Set 1	Set 2
$M_1 = 20$	$M_1 = 28$
$M_2 = 30$	$M_2 = 30$
$M_3 = 35$	$M_3 = 31$

If you compute the variance for the three numbers in each set, then the variance is $s^2 = 58.33$ for set 1 and the variance is $s^2 = 2.33$ for set 2. Notice that the two variances provide an accurate representation of the size of the differences. In set 1, there are relatively large differences between sample means and the variance is relatively large. In set 2, the mean differences are small and the variance is small.

Thus, we can use variance to measure sample mean differences when there are two or more samples. The test statistic for ANOVA uses this fact to compute an F-*ratio* with the following structure:

$$F = \frac{\text{variance (differences) between sample means}}{\text{variance (differences) expected with no treatment effect}}$$

Note that the F-ratio has the same basic structure as the t statistic but is based on *variance* instead of sample mean *difference*. The variance in the numerator of the F-ratio provides a single number that measures the differences among all of the sample means. The variance in the denominator of the F-ratio, like the standard error in the denominator of the *t* statistic, measures the mean differences that would be expected if there were no treatment effect. Thus, the *t* statistic and the *F*-ratio provide the same basic information. In each case, a large value for the test statistic provides evidence that the sample mean differences (numerator) are larger than would be expected if there were no treatment effects (denominator).

TYPE I ERRORS AND MULTIPLE-HYPOTHESIS TESTS

If we already have *t* tests for comparing mean differences, you might wonder why ANOVA is necessary. Why create a whole new hypothesis-testing procedure that simply duplicates what the *t* tests can already do? The answer to this question is based in a concern about Type I errors.

Remember that each time you do a hypothesis test, you select an alpha level that determines the risk of a Type I error. With $\alpha = .05$, for example, there is a 5%, or a 1-in-20, risk of a Type I error. Often a single experiment requires several hypothesis tests to evaluate all the mean differences. However, each test has a risk of a Type I error, and the more tests you do, the more risk there is.

For this reason, researchers often make a distinction between the *testwise alpha level* and the *experimentwise alpha level*. The testwise alpha level is simply the alpha level that you select for each individual hypothesis test. The experimentwise alpha level is the total probability of a Type I error accumulated from all of the separate tests in the experiment. As the number of separate tests increases, so does the experimentwise alpha level.

DEFINITIONS

The **testwise alpha level** is the risk of a Type I error, or alpha level, for an individual hypothesis test.

When an experiment involves several different hypothesis tests, the **experimentwise alpha level** is the total probability of a Type I error that is accumulated from all of the individual tests in the experiment. Typically, the experimentwise alpha level is substantially greater than the value of alpha used for any one of the individual tests.

For example, an experiment involving three treatments would require three separate *t* tests to compare all of the mean differences:

Test 1 compares treatment I with treatment II.

Test 2 compares treatment I with treatment III.

Test 3 compares treatment II with treatment III.

If all tests use $\alpha = .05$, then there is a 5% risk of a Type I error for the first test, a 5% risk for the second test, and another 5% risk for the third test. The three separate tests accumulate to produce a relatively large experimentwise alpha level. The advantage of ANOVA is that it performs all three comparisons simultaneously in one hypothesis test. Thus, no matter how many different means are being compared, ANOVA uses one test with one alpha level to evaluate the mean differences, and thereby avoids the problem of an inflated experimentwise alpha level.

12.2 THE LOGIC OF ANOVA

The formulas and calculations required in ANOVA are somewhat complicated, but the logic that underlies the whole procedure is fairly straightforward. Therefore, this section gives a general picture of ANOVA before we start looking at the details. We introduce the logic of ANOVA with the help of the hypothetical data in Table 12.1.

TABLE 12.1

Hypothetical data from an experiment examining driving performance under three telephone conditions.*

Treatment 1: No Phone (Sample 1)	Treatment 2: Hand-Held (Sample 2)	Treatment 3: Hands-Free (Sample 3)
4	0	1
3	1	2
6	3	2
3	1	0
4	0	0
$M = 4$	$M = 1$	$M = 1$

*Note that there are three separate samples, with $n = 5$ in each sample. The dependent variable is a measure of performance in a driving simulator.

These data represent the results of an independent-measures experiment comparing performance in a driving simulator under three telephone conditions.

One obvious characteristic of the data in Table 12.1 is that the scores are not all the same. In everyday language, the scores are different; in statistical terms, the scores are variable. Our goal is to measure the amount of variability (the size of the differences) and to explain why the scores are different.

The first step is to determine the total variability for the entire set of data. To compute the total variability, we combine all of the scores from all of the separate samples to obtain one general measure of variability for the complete experiment. Once we have measured the total variability, we can begin to break it apart into separate components. The word *analysis* means dividing into smaller parts. Because we are going to analyze variability, the process is called *analysis of variance*. This analysis process divides the total variability into two basic components.

1. **Between-Treatments Variance.** Looking at the data in Table 12.1, we clearly see that much of the variability in the scores results from general differences between treatment conditions. For example, the scores in the no-phone condition tend to be much higher ($M = 4$) than the scores in the hand-held condition ($M = 1$). We calculate the variance between treatments to provide a measure of the overall differences between treatment conditions. Notice that the variance between treatments is really measuring the differences between sample means.
2. **Within-Treatment Variance.** In addition to the general differences between treatment conditions, there is variability within each sample. Looking again at Table 12.1, we see that the scores in the no-phone condition are not all the same; they are variable. The within-treatments variance provides a measure of the variability inside each treatment condition.

Analyzing the total variability into these two components is the heart of ANOVA. We now examine each of the components in more detail.

BETWEEN-TREATMENTS VARIANCE

Remember that calculating variance is simply a method for measuring how big the differences are for a set of numbers. When you see the term *variance,* you can automatically translate it into the term *differences*. Thus, the *between-treatments variance* simply measures how much difference exists between the treatment conditions. There are two possible explanations for these between-treatment differences:

1. The differences between treatments are not caused by any treatment effect but are simply the naturally occurring, random, and unsystematic differences that

exist between one sample and another. That is, the differences are the result of sampling error.

2. The differences between treatments have been caused by the *treatment effects*. For example, if using a telephone really does interfere with driving performance, then scores in the telephone conditions should be systematically lower than scores in the no-phone condition.

Thus, when we compute the between-treatments variance, we are measuring differences that could be caused by a systematic treatment effect or could simply be random and unsystematic mean differences caused by sampling error. To demonstrate that there really is a treatment effect, we must establish that the differences between treatments are bigger than would be expected by sampling error alone. To accomplish this goal, we determine how big the differences are when there is no systematic treatment effect; that is, we measure how much difference (or variance) can be explained by random and unsystematic factors. To measure these differences, we compute the variance within treatments.

WITHIN-TREATMENTS VARIANCE

Inside each treatment condition, we have a set of individuals who all receive exactly the same treatment; that is, the researcher does not do anything that would cause these individuals to have different scores. In Table 12.1, for example, the data show that five individuals were tested while talking on a hand-held phone (sample 2). Although these five individuals all received exactly the same treatment, their scores are different. Why are the scores different? The answer is that there is no specific cause for the differences. Instead, the differences that exist within a treatment represent random and unsystematic differences that occur when there are no treatment effects causing the scores to be different. Thus, the *within-treatments variance* provides a measure of how big the differences are when H_0 is true.

Figure 12.4 shows the overall ANOVA and identifies the sources of variability that are measured by each of the two basic components.

THE *F*-RATIO: THE TEST STATISTIC FOR ANOVA

Once we have analyzed the total variability into two basic components (between treatments and within treatments), we simply compare them. The comparison is made by

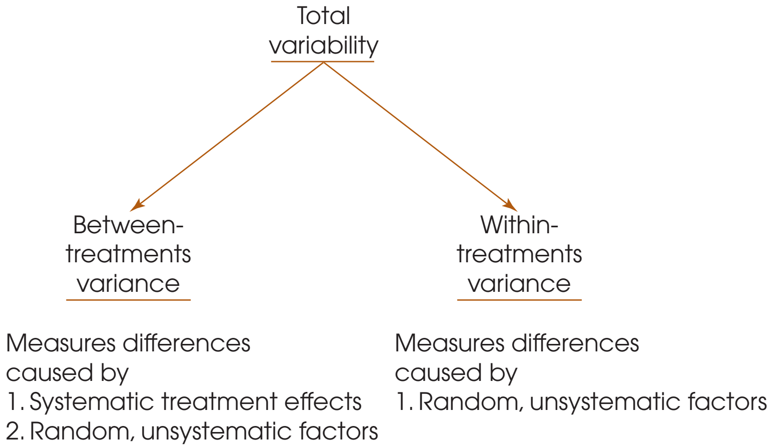

Total variability
Between-treatments variance
Within-treatments variance
Measures differences caused by
1. Systematic treatment effects
2. Random, unsystematic factors
Measures differences caused by
1. Random, unsystematic factors

FIGURE 12.4

The independent-measures ANOVA partitions, or analyzes, the total variability into two components: variance between treatments and variance within treatments.

computing an F-*ratio*. For the independent-measures ANOVA, the *F*-ratio has the following structure:

$$F = \frac{\text{variance between treatments}}{\text{variance within treatments}} = \frac{\text{differences including any treatment effects}}{\text{differences with no treatment effects}} \qquad (12.1)$$

When we express each component of variability in terms of its sources (see Figure 12.4), the structure of the *F*-ratio is

$$F = \frac{\text{systematic treatment effects} + \text{random, unsystematic differences}}{\text{random, unsystematic differences}} \qquad (12.2)$$

The value obtained for the *F*-ratio helps determine whether any treatment effects exist. Consider the following two possibilities:

1. When there are no systematic treatment effects, the differences between treatments (numerator) are entirely caused by random, unsystematic factors. In this case, the numerator and the denominator of the *F*-ratio are both measuring random differences and should be roughly the same size. With the numerator and denominator roughly equal, the *F*-ratio should have a value around 1.00. In terms of the formula, when the treatment effect is zero, we obtain

$$F = \frac{0 + \text{random, unsystematic differences}}{\text{random, unsystematic differences}}$$

   Thus, an *F*-ratio near 1.00 indicates that the differences between treatments (numerator) are random and unsystematic, just like the differences in the denominator. With an *F*-ratio near 1.00, we conclude that there is no evidence to suggest that the treatment has any effect.

2. When the treatment does have an effect, causing systematic differences between samples, then the combination of systematic and random differences in the numerator should be larger than the random differences alone in the denominator. In this case, the numerator of the *F*-ratio should be noticeably larger than the denominator, and we should obtain an *F*-ratio that is substantially larger than 1.00. Thus, a large *F*-ratio is evidence for the existence of systematic treatment effects; that is, there are consistent differences between treatments.

Because the denominator of the *F*-ratio measures only random and unsystematic variability, it is called the *error term*. The numerator of the *F*-ratio always includes the same unsystematic variability as in the error term, but it also includes any systematic differences caused by the treatment effect. The goal of ANOVA is to find out whether a treatment effect exists.

DEFINITION

For ANOVA, the denominator of the *F*-ratio is called the **error term**. The error term provides a measure of the variance caused by random, unsystematic differences. When the treatment effect is zero (H_0 is true), the error term measures the same sources of variance as the numerator of the *F*-ratio, so the value of the *F*-ratio is expected to be nearly equal to 1.00.

LEARNING CHECK

1. ANOVA is a statistical procedure that compares two or more treatment conditions for differences in variance. (True or false?)
2. In ANOVA, what value is expected, on the average, for the *F*-ratio when the null hypothesis is true?
3. What happens to the value of the *F*-ratio if differences between treatments are increased? What happens to the *F*-ratio if variability inside the treatments is increased?
4. In ANOVA, the total variability is partitioned into two parts. What are these two variability components called, and how are they used in the *F*-ratio?

ANSWERS

1. False. Although ANOVA uses variance in the computations, the purpose of the test is to evaluate differences in *means* between treatments.
2. When H_0 is true, the expected value for the *F*-ratio is 1.00 because the top and bottom of the ratio are both measuring the same variance.
3. As differences between treatments increase, the *F*-ratio increases. As variability within treatments increases, the *F*-ratio decreases.
4. The two components are between-treatments variability and within-treatments variability. Between-treatments variance is the numerator of the *F*-ratio, and within-treatments variance is the denominator.

12.3 ANOVA NOTATION AND FORMULAS

Because ANOVA typically is used to examine data from more than two treatment conditions (and more than two samples), we need a notational system to keep track of all the individual scores and totals. To help introduce this notational system, we use the hypothetical data from Table 12.1 again. The data are reproduced in Table 12.2 along with some of the notation and statistics that are described in the following list.

1. The letter k is used to identify the number of treatment conditions—that is, the number of levels of the factor. For an independent-measures study, k also specifies the number of separate samples. For the data in Table 12.2, there are three treatments, so $k = 3$.
2. The number of scores in each treatment is identified by a lowercase letter n. For the example in Table 12.2, $n = 5$ for all the treatments. If the samples are of different sizes, you can identify a specific sample by using a subscript. For example, n_2 is the number of scores in treatment 2.
3. The total number of scores in the entire study is specified by a capital letter N. When all of the samples are the same size (n is constant), $N = kn$. For the data in Table 12.2, there are $n = 5$ scores in each of the $k = 3$ treatments, so we have a total of $N = 3(5) = 15$ scores in the entire study.
4. The sum of the scores (ΣX) for each treatment condition is identified by the capital letter T (for treatment total). The total for a specific treatment can be identified by adding a numerical subscript to the T. For example, the total for the second treatment in Table 12.2 is $T_2 = 5$.
5. The sum of all of the scores in the research study (the grand total) is identified by G. You can compute G by adding up all N scores or by adding up the treatment totals: $G = \Sigma T$.
6. Although there is no new notation involved, we also have computed SS and M for each sample, and we have calculated ΣX^2 for the entire set of $N = 15$ scores

TABLE 12.2

The same data that appeared in Table 12.1 with summary values and notation appropriate for an ANOVA.

Telephone Conditions			
Treatment 1 No Phone (Sample 1)	Treatment 2 Hand-Held Phone (Sample 2)	Treatment 3 Hands-Free Phone (Sample 3)	
4	0	1	$\Sigma X^2 = 106$
3	1	2	$G = 30$
6	3	2	$N = 15$
3	1	0	$k = 3$
4	0	0	
$T_1 = 20$	$T_2 = 5$	$T_3 = 5$	
$SS_1 = 6$	$SS_2 = 6$	$SS_3 = 4$	
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	
$M_1 = 4$	$M_2 = 1$	$M_3 = 1$	

in the study. These values are given in Table 12.2 and are important in the formulas and calculations for ANOVA.

Finally, we should note that there is no universally accepted notation for ANOVA. Although we are using *G*s and *T*s, for example, you may find that other sources use other symbols.

ANOVA FORMULAS

Because ANOVA requires extensive calculations and many formulas, one common problem for students is simply keeping track of the different formulas and numbers. Therefore, we examine the general structure of the procedure and look at the organization of the calculations before we introduce the individual formulas.

Because ANOVA formulas require ΣX for each treatment and ΣX for the entire set of scores, we have introduced new notation (*T* and *G*) to help identify which ΣX is being used. Remember: *T* stands for *treatment total*, and *G* stands for *grand total*.

1. The final calculation for ANOVA is the *F*-ratio, which is composed of two variances:

$$F = \frac{\text{variance between treatments}}{\text{variance within treatments}}$$

2. Each of the two variances in the *F*-ratio is calculated using the basic formula for sample variance.

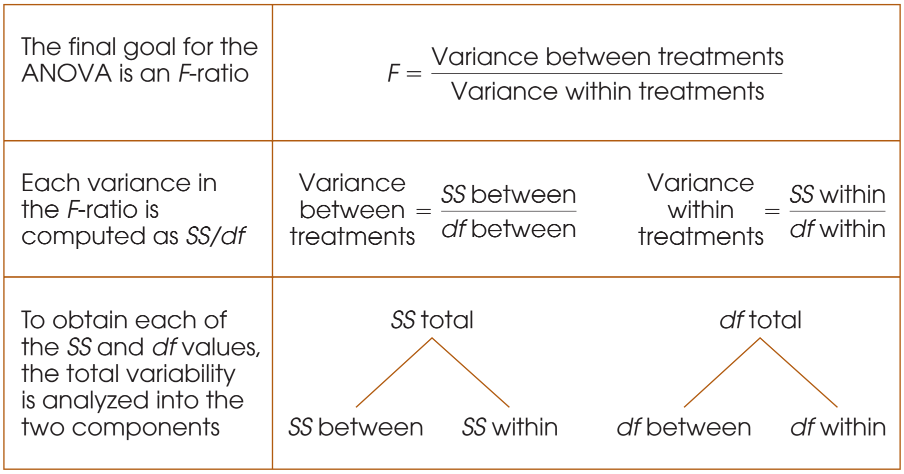
$$\text{sample variance} = s^2 = \frac{SS}{df}$$

Therefore, we need to compute an *SS* and a *df* for the variance between treatments (numerator of *F*), and we need another *SS* and *df* for the variance within treatments (denominator of *F*). To obtain these *SS* and *df* values, we must go through two separate analyses: First, compute *SS* for the total study, and analyze it in two components (between and within). Then compute *df* for the total study, and analyze it in two components (between and within).

Thus, the entire process of ANOVA requires nine calculations: three values for *SS*, three values for *df*, two variances (between and within), and a final *F*-ratio. However, these nine calculations are all logically related and are all directed toward finding the final *F*-ratio. Figure 12.5 shows the logical structure of ANOVA calculations.

FIGURE 12.5

The structure and sequence of calculations for the ANOVA.



The final goal for the ANOVA is an *F*-ratio	$F = \dfrac{\text{Variance between treatments}}{\text{Variance within treatments}}$	
Each variance in the *F*-ratio is computed as *SS*/*df*	$\text{Variance between treatments} = \dfrac{SS \text{ between}}{df \text{ between}}$	$\text{Variance within treatments} = \dfrac{SS \text{ within}}{df \text{ within}}$
To obtain each of the *SS* and *df* values, the total variability is analyzed into the two components	*SS* total *SS* between *SS* within	*df* total *df* between *df* within

ANALYSIS OF THE SUM OF SQUARES (*SS*)

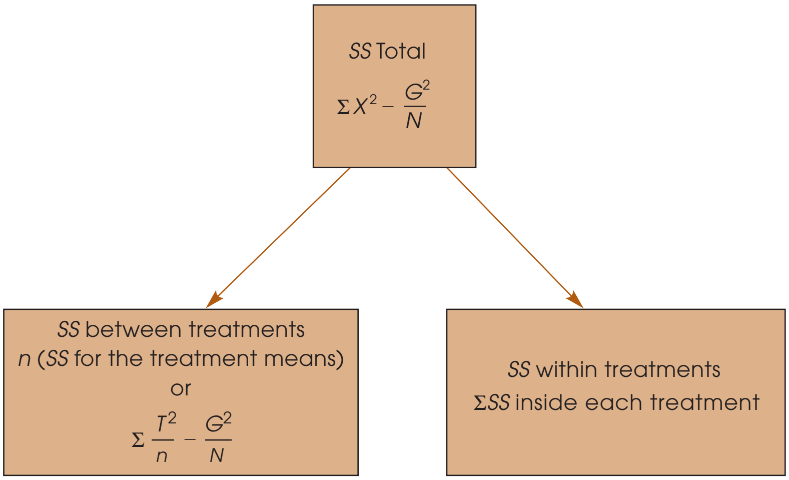
The ANOVA requires that we first compute a total sum of squares and then partition this value into two components: between treatments and within treatments. This analysis is outlined in Figure 12.6. We will examine each of the three components separately.

1. **Total Sum of Squares, SS_{total}.** As the name implies, SS_{total} is the sum of squares for the entire set of *N* scores. As described in Chapter 4 (pp. 111–112), this *SS* value can be computed using either a definitional or a computational formula. However, ANOVA typically involves a large number of scores and the mean is often not a whole number. Therefore, it is usually much easier to calculate SS_{total} using the computational formula:

$$SS = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

FIGURE 12.6

Partitioning the sum of squares (*SS*) for the independent-measures ANOVA.


SS Total
$\Sigma X^2 - \frac{G^2}{N}$
SS between treatments
n (*SS* for the treatment means)
or
$\Sigma \frac{T^2}{n} - \frac{G^2}{N}$
SS within treatments
Σ*SS* inside each treatment

To make this formula consistent with the ANOVA notation, we substitute the letter G in place of ΣX and obtain

$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{N} \qquad (12.3)$$

Applying this formula to the set of data in Table 12.2, we obtain

$$\begin{aligned} SS_{\text{total}} &= 106 - \frac{30^2}{15} \\ &= 106 - 60 \\ &= 46 \end{aligned}$$

2. **Within-Treatments Sum of Squares, $SS_{\text{within treatments}}$.** Now we are looking at the variability inside each of the treatment conditions. We already have computed the SS within each of the three treatment conditions (Table 12.2): $SS_1 = 6$, $SS_2 = 6$, and $SS_3 = 4$. To find the overall within-treatment sum of squares, we simply add these values together:

$$SS_{\text{within treatments}} = \Sigma SS_{\text{inside each treatment}} \qquad (12.4)$$

For the data in Table 12.2, this formula gives

$$\begin{aligned} SS_{\text{within treatments}} &= 6 + 6 + 4 \\ &= 16 \end{aligned}$$

3. **Between-Treatments Sum of Squares, $SS_{\text{between treatments}}$.** Before we introduce any equations for $SS_{\text{between treatments}}$, consider what we have found so far. The total variability for the data in Table 12.2 is $SS_{\text{total}} = 46$. We intend to partition this total into two parts (see Figure 12.5). One part, $SS_{\text{within treatments}}$, has been found to be equal to 16. This means that $SS_{\text{between treatments}}$ must be equal to 30 so that the two parts (16 and 30) add up to the total (46). Thus, the value for $SS_{\text{between treatments}}$ can be found simply by subtraction:

$$SS_{\text{between}} = SS_{\text{total}} - SS_{\text{within}} \qquad (12.5)$$

To simplify the notation, we use the subscripts *between* and *within* in place of *between treatments* and *within treatments.*

However, it is also possible to compute SS_{between} independently, then check your calculations by ensuring that the two components, between and within, add up to the total. Box 12.1 presents two different formulas for calculating SS_{between} directly from the data.

Computing SS_{between} Including the two formulas in Box 12.1, we have presented three different equations for computing SS_{between}. Rather than memorizing all three, however, we suggest that you pick one formula and use it consistently. There are two reasonable alternatives to use. The simplest is Equation 12.5, which finds SS_{between} simply by subtraction: First you compute SS_{total} and SS_{within}, then subtract:

$$SS_{\text{between}} = SS_{\text{total}} - SS_{\text{within}}$$

The second alternative is to use Equation 12.7, which computes SS_{between} using the treatment totals (the T values). The advantage of this alternative is that it provides a way to check your arithmetic: Calculate SS_{total}, SS_{between}, and SS_{within} separately, and then check to be sure that the two components add up to equal SS_{total}.

BOX 12.1 ALTERNATIVE FORMULAS FOR SS_{between}

Recall that the variability between treatments is measuring the differences between treatment means. Conceptually, the most direct way of measuring the amount of variability among the treatment means is to compute the sum of squares for the set of sample means, SS_{means}. For the data in Table 12.2, the samples means are 4, 1, and 1. These three values produce $SS_{\text{means}} = 6$. However, each of the three means represents a group of $n = 5$ scores. Therefore, the final value for SS_{between} is obtained by multiplying SS_{means} by n.

$$SS_{\text{between}} = n(SS_{\text{means}}) \qquad (12.6)$$

For the data in Table 12.2, we obtain

$$SS_{\text{between}} = n(SS_{\text{means}}) = 5(6) = 30$$

Unfortunately, Equation 12.6 can only be used when all of the samples are exactly the same size (equal *n*s), and the equation can be very awkward, especially when the treatment means are not whole numbers. Therefore, we also present a computational formula for SS_{between} that uses the treatment totals (*T*) instead of the treatment means.

$$SS_{\text{between}} = \Sigma\frac{T^2}{n} - \frac{G^2}{N} \qquad (12.7)$$

For the data in Table 12.2, this formula produces:

$$\begin{aligned} SS_{\text{between}} &= \frac{20^2}{5} + \frac{5^2}{5} + \frac{5^2}{5} - \frac{30^2}{15} \\ &= 80 + 5 + 5 - 60 \\ &= 90 - 60 \\ &= 30 \end{aligned}$$

Note that all three techniques (Equations 12.5, 12.6, and 12.7) produce the same result, $SS_{\text{between}} = 30$.

Using Equation 12.6, which computes *SS* for the set of sample means, is usually not a good choice. Unless the sample means are all whole numbers, this equation can produce very tedious calculations. In most situations, one of the other two equations is a better alternative.

THE ANALYSIS OF DEGREES OF FREEDOM (*DF*)

The analysis of degrees of freedom (*df*) follows the same pattern as the analysis of *SS*. First, we find *df* for the total set of *N* scores, and then we partition this value into two components: degrees of freedom between treatments and degrees of freedom within treatments. In computing degrees of freedom, there are two important considerations to keep in mind:

1. Each *df* value is associated with a specific *SS* value.
2. Normally, the value of *df* is obtained by counting the number of items that were used to calculate *SS* and then subtracting 1. For example, if you compute *SS* for a set of *n* scores, then $df = n - 1$.

With this in mind, we examine the degrees of freedom for each part of the analysis.

1. **Total Degrees of Freedom, df_{total}.** To find the *df* associated with SS_{total}, you must first recall that this *SS* value measures variability for the entire set of *N* scores. Therefore, the *df* value is

$$df_{\text{total}} = N - 1 \qquad (12.8)$$

For the data in Table 12.2, the total number of scores is $N = 15$, so the total degrees of freedom are

$$\begin{aligned} df_{\text{total}} &= 15 - 1 \\ &= 14 \end{aligned}$$

2. **Within-Treatments Degrees of Freedom, df_{within}.** To find the *df* associated with SS_{within}, we must look at how this *SS* value is computed. Remember, we first find *SS* inside of each of the treatments and then add these values together. Each of the treatment *SS* values measures variability for the *n* scores in the treatment, so each *SS* has $df = n - 1$. When all of these individual treatment values are added together, we obtain

$$df_{\text{within}} = \Sigma(n - 1) = \Sigma df_{\text{in each treatment}} \qquad (12.9)$$

For the experiment we have been considering, each treatment has $n = 5$ scores. This means there are $n - 1 = 4$ degrees of freedom inside each treatment. Because there are three different treatment conditions, this gives a total of 12 for the within-treatments degrees of freedom. Notice that this formula for *df* simply adds up the number of scores in each treatment (the *n* values) and subtracts 1 for each treatment. If these two stages are done separately, you obtain

$$df_{\text{within}} = N - k \qquad (12.10)$$

(Adding up all the *n* values gives *N*. If you subtract 1 for each treatment, then altogether you have subtracted *k* because there are *k* treatments.) For the data in Table 12.2, $N = 15$ and $k = 3$, so

$$\begin{aligned} df_{\text{within}} &= 15 - 3 \\ &= 12 \end{aligned}$$

3. **Between-Treatments Degrees of Freedom, df_{between}.** The *df* associated with SS_{between} can be found by considering how the *SS* value is obtained. This *SS* formulas measure the variability for the set of treatments (totals or means). To find df_{between}, simply count the number of treatments and subtract 1. Because the number of treatments is specified by the letter *k*, the formula for *df* is

$$df_{\text{between}} = k - 1 \qquad (12.11)$$

For the data in Table 12.2, there are three different treatment conditions (three *T* values or three sample means), so the between-treatments degrees of freedom are computed as follows:

$$\begin{aligned} df_{\text{between}} &= 3 - 1 \\ &= 2 \end{aligned}$$

Notice that the two parts we obtained from this analysis of degrees of freedom add up to equal the total degrees of freedom:

$$\begin{aligned} df_{\text{total}} &= df_{\text{within}} + df_{\text{between}} \\ 14 &= 12 + 2 \end{aligned}$$

The complete analysis of degrees of freedom is shown in Figure 12.7.

As you are computing the *SS* and *df* values for ANOVA, keep in mind that the labels that are used for each value can help you understand the formulas. Specifically,

1. The term **total** refers to the entire set of scores. We compute *SS* for the whole set of *N* scores, and the *df* value is simply $N - 1$.

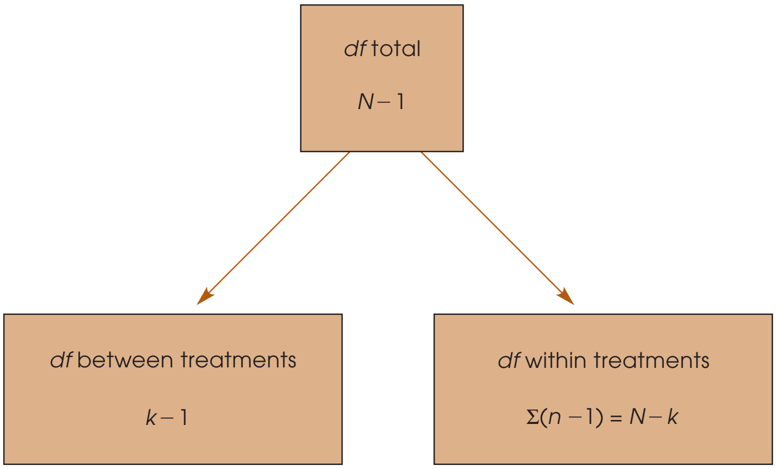


FIGURE 12.7
Partitioning degrees of freedom (*df*) for the independent-measures ANOVA.

2. The term **within treatments** refers to differences that exist inside the individual treatment conditions. Thus, we compute *SS* and *df* inside each of the separate treatments.
3. The term **between treatments** refers to differences from one treatment to another. With three treatments, for example, we are comparing three different means (or totals) and have $df = 3 - 1 = 2$.

CALCULATION OF VARIANCES (*MS*) AND THE *F*-RATIO

The next step in the ANOVA procedure is to compute the variance between treatments and the variance within treatments, which are used to calculate the *F*-ratio (see Figure 12.5).

In ANOVA, it is customary to use the term *mean square,* or simply *MS,* in place of the term *variance*. Recall (from Chapter 4) that variance is defined as the mean of the squared deviations. In the same way that we use *SS* to stand for the sum of the squared deviations, we now use *MS* to stand for the mean of the squared deviations. For the final *F*-ratio we need an *MS* (variance) between treatments for the numerator and an *MS* (variance) within treatments for the denominator. In each case

$$MS \text{ (variance)} = s^2 = \frac{SS}{df} \tag{12.12}$$

For the data we have been considering,

$$MS_{\text{between}} = s^2_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{30}{2} = 15$$

and

$$MS_{\text{within}} = s^2_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{16}{12} = 1.33$$

We now have a measure of the variance (or differences) between the treatments and a measure of the variance within the treatments. The *F*-ratio simply compares these two variances:

$$F = \frac{s^2_{\text{between}}}{s^2_{\text{within}}} = \frac{MS_{\text{between}}}{MS_{\text{within}}} \tag{12.13}$$

For the experiment we have been examining, the data give an *F*-ratio of

$$F = \frac{15}{1.33} = 11.28$$

For this example, the obtained value of $F = 11.28$ indicates that the numerator of the *F*-ratio is substantially bigger than the denominator. If you recall the conceptual structure of the *F*-ratio as presented in Equations 12.1 and 12.2, the *F* value we obtained indicates that the differences between treatments are more than 11 times bigger than what would be expected if there were no treatment effect. Stated in terms of the experimental variables: using a telephone while driving does appear to have an effect on driving performance. However, to properly evaluate the *F*-ratio, we must select an α level and consult the *F*-distribution table that is discussed in the next section.

ANOVA Summary Tables It is useful to organize the results of the analysis in one table called an *ANOVA summary table*. The table shows the source of variability (between treatments, within treatments, and total variability), *SS*, *df*, *MS*, and *F*. For the previous computations, the ANOVA summary table is constructed as follows:

Source	*SS*	*df*	*MS*	
Between treatments	30	2	15	$F = 11.28$
Within treatments	16	12	1.33	
Total	46	14		

Although these tables are no longer used in published reports, they are a common part of computer printouts, and they do provide a concise method for presenting the results of an analysis. (Note that you can conveniently check your work: Adding the first two entries in the *SS* column, 30 + 16, produces SS_{total}. The same applies to the *df* column.) When using ANOVA, you might start with a blank ANOVA summary table and then fill in the values as they are calculated. With this method, you are less likely to "get lost" in the analysis, wondering what to do next.

LEARNING CHECK

1. Calculate SS_{total}, SS_{between}, and SS_{within} for the following set of data:

Treatment 1	Treatment 2	Treatment 3	
$n = 10$	$n = 10$	$n = 10$	$N = 30$
$T = 10$	$T = 20$	$T = 30$	$G = 60$
$SS = 27$	$SS = 16$	$SS = 23$	$\Sigma X^2 = 206$

2. A researcher uses an ANOVA to compare three treatment conditions with a sample of $n = 8$ in each treatment. For this analysis, find df_{total}, df_{between}, and df_{within}.

3. A researcher reports an *F*-ratio with $df_{\text{between}} = 2$ and $df_{\text{within}} = 30$ for an independent-measures ANOVA. How many treatment conditions were compared in the experiment? How many subjects participated in the experiment?

4. A researcher conducts an experiment comparing four treatment conditions with a separate sample of $n = 6$ in each treatment. An ANOVA is used to evaluate the data, and the results of the ANOVA are presented in the following table. Complete all missing values in the table. *Hint:* Begin with the values in the *df* column.

Source	*SS*	*df*	*MS*	
Between treatments	__	__	__	$F =$ ____
Within treatments	__	__	2	
Total	58	__		

ANSWERS

1. $SS_{\text{total}} = 86$; $SS_{\text{between}} = 20$; $SS_{\text{within}} = 66$

2. $df_{\text{total}} = 23$; $df_{\text{between}} = 2$; $df_{\text{within}} = 21$

3. There were 3 treatment conditions ($df_{\text{between}} = k - 1 = 2$). A total of $N = 33$ individuals participated ($df_{\text{within}} = 30 = N - k$).

4.

Source	*SS*	*df*	*MS*	
Between treatments	18	3	6	$F = 3.00$
Within treatments	40	20	2	
Total	58	23		

12.4 THE DISTRIBUTION OF *F*-RATIOS

In ANOVA, the *F*-ratio is constructed so that the numerator and denominator of the ratio are measuring exactly the same variance when the null hypothesis is true (see Equation 12.2). In this situation, we expect the value of *F* to be around 1.00.

If the null hypothesis is false, then the *F*-ratio should be much greater than 1.00. The problem now is to define precisely which values are "around 1.00" and which are "much greater than 1.00." To answer this question, we need to look at all of the possible *F* values—that is, the *distribution of* F-*ratios*.

Before we examine this distribution in detail, you should note two obvious characteristics:

1. Because *F*-ratios are computed from two variances (the numerator and denominator of the ratio), *F* values always are positive numbers. Remember that variance is always positive.
2. When H_0 is true, the numerator and denominator of the *F*-ratio are measuring the same variance. In this case, the two sample variances should be about the same size, so the ratio should be near 1. In other words, the distribution of *F*-ratios should pile up around 1.00.

With these two factors in mind, we can sketch the distribution of *F*-ratios. The distribution is cut off at zero (all positive values), piles up around 1.00, and then tapers off to the right (Figure 12.8). The exact shape of the *F* distribution depends on the degrees

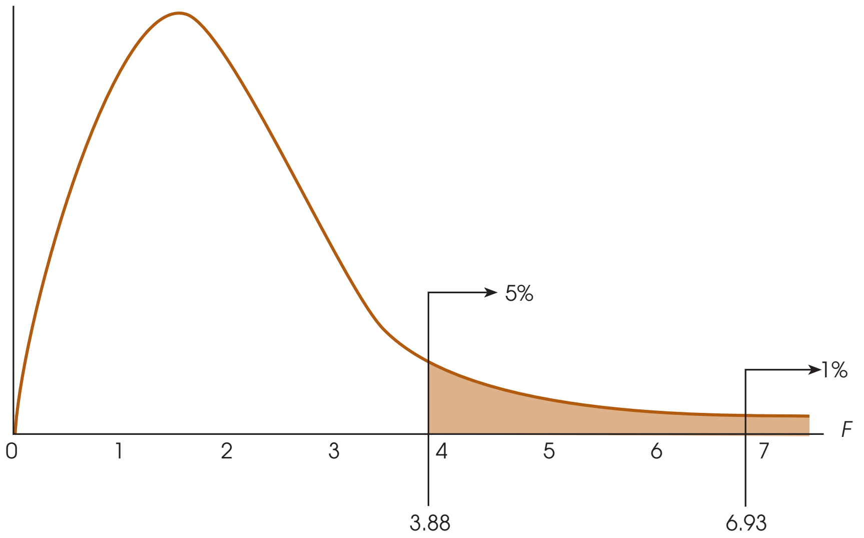


FIGURE 12.8

The distribution of *F*-ratios with $df = 2, 12$. Of all the values in the distribution, only 5% are larger than $F = 3.88$, and only 1% are larger than $F = 6.93$.

of freedom for the two variances in the *F*-ratio. You should recall that the precision of a sample variance depends on the number of scores or the degrees of freedom. In general, the variance for a large sample (large *df*) provides a more accurate estimate of the population variance. Because the precision of the *MS* values depends on *df*, the shape of the *F* distribution also depends on the *df* values for the numerator and denominator of the *F*-ratio. With very large *df* values, nearly all of the *F*-ratios are clustered very near to 1.00. With the smaller *df* values, the *F* distribution is more spread out.

THE *F* DISTRIBUTION TABLE

For ANOVA, we expect *F* near 1.00 if H_0 is true, and we expect a large value for *F* if H_0 is not true. In the *F* distribution, we need to separate those values that are reasonably near 1.00 from the values that are significantly greater than 1.00. These critical values are presented in an *F* distribution table in Appendix B, page 705. A portion of the *F* distribution table is shown in Table 12.3. To use the table, you must know the *df* values for the *F*-ratio (numerator and denominator), and you must know the alpha level for the hypothesis test. It is customary for an *F* table to have the *df* values for the numerator of the *F*-ratio printed across the top of the table. The *df* values for the denominator of *F* are printed in a column on the left-hand side. For the experiment we have been considering, the numerator of the *F*-ratio (between treatments) has $df = 2$, and the denominator of the *F*-ratio (within treatments) has $df = 12$. This *F*-ratio is said to have "degrees of freedom equal to 2 and 12." The degrees of freedom would be written as $df = 2, 12$. To use the table, you would first find $df = 2$ across the top of the table and $df = 12$ in the first column. When you line up these two values, they point to a pair of numbers in the middle of the table. These numbers give the critical cutoffs for $\alpha = .05$ and $\alpha = .01$. With $df = 2, 12$, for example, the numbers in the table are 3.88 and 6.93. Thus, only 5% of the distribution ($\alpha = .05$) corresponds to values greater than 3.88, and only1% of the distribution ($\alpha = .01$) corresponds to values greater than 6.93 (see Figure 12.8).

In the experiment comparing driving performance under different telephone conditions, we obtained an *F*-ratio of 11.28. According to the critical cutoffs in Figure 12.8, this value is extremely unlikely (it is in the most extreme 1%). Therefore, we would reject H_0 with an α level of either .05 or .01, and conclude that the different telephone conditions significantly affect driving performance.

TABLE 12.3

A portion of the F distribution table. Entries in roman type are critical values for the .05 level of significance, and bold type values are for the .01 level of significance. The critical values for df = 2, 12 have been highlighted (see text).

Degrees of Freedom: Denominator	Degrees of Freedom: Numerator					
	1	2	3	4	5	6
10	4.96	4.10	3.71	3.48	3.33	3.22
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**
11	4.84	3.98	3.59	3.36	3.20	3.09
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**
12	4.75	3.88	3.49	3.26	3.11	3.00
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**
13	4.67	3.80	3.41	3.18	3.02	2.92
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**
14	4.60	3.74	3.34	3.11	2.96	2.85
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**

LEARNING CHECK

1. A researcher obtains $F = 4.18$ with $df = 2, 15$. Is this value sufficient to reject H_0 with $\alpha = .05$? Is it big enough to reject H_0 if $\alpha = .01$?
2. With $\alpha = .05$, what value forms the boundary for the critical region in the distribution of F-ratios with $df = 2, 24$?

ANSWERS

1. For $\alpha = .05$, the critical value is 3.68 and you should reject H_0. For $\alpha = .01$, the critical value is 6.36 and you should fail to reject H_0.
2. The critical value is 3.40.

12.5 EXAMPLES OF HYPOTHESIS TESTING AND EFFECT SIZE WITH ANOVA

Although we have now seen all the individual components of ANOVA, the following example demonstrates the complete ANOVA process using the standard four-step procedure for hypothesis testing.

EXAMPLE 12.1

The data in Table 12.4 were obtained from an independent-measures experiment designed to examine people's preferences for viewing distance of a 42-inch, high-definition television. Four viewing distances were evaluated, 9 feet, 12 feet, 15 feet, and 18 feet, with a separate group of participants tested at each distance. Each individual watched a 30-minute television program from a specific distance and then completed a brief questionnaire measuring their satisfaction with the experience. One question asked them to rate the viewing distance on a scale from 1 (Very Bad—definitely need to move closer or farther away) to 7 (Excellent—perfect viewing distance). The purpose of the ANOVA is to determine whether there are any significant differences among the four viewing distances that were tested.

Before we begin the hypothesis test, note that we have already computed several summary statistics for the data in Table 12.4. Specifically, the treatment totals (T) and SS values are shown for each sample, and the grand total (G) as well as N and ΣX^2 are shown for the entire set of data. Having these summary values simplifies the

TABLE 12.4

Satisfaction with different viewing distances of a 42-inch high-definition television.

9 feet	12 feet	15 feet	18 feet	
3	4	7	6	$N = 20$
0	3	6	3	$G = 60$
2	1	5	4	$\Sigma X^2 = 262$
0	1	4	3	
0	1	3	4	
$T = 5$	$T = 10$	$T = 25$	$T = 20$	
$SS = 8$	$SS = 8$	$SS = 10$	$SS = 6$	

computations in the hypothesis test, and we suggest that you always compute these summary statistics before you begin an ANOVA.

STEP 1: State the hypotheses and select an alpha level.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (There is no treatment effect.)

H_1: At least one of the treatment means is different.

We use $\alpha = .05$.

STEP 2: Locate the critical region.

We first must determine degrees of freedom for $MS_{\text{between treatments}}$ and $MS_{\text{within treatments}}$ (the numerator and denominator of the F-ratio), so we begin by analyzing the degrees of freedom. For these data, the total degrees of freedom are

Often it is easier to postpone finding the critical region until after step 3, where you compute the *df* values as part of the calculations for the *F*-ratio.

$$\begin{aligned} df_{\text{total}} &= N - 1 \\ &= 20 - 1 \\ &= 19 \end{aligned}$$

Analyzing this total into two components, we obtain

$$df_{\text{between}} = k - 1 = 4 - 1 = 3$$

$$df_{\text{within}} = \Sigma df_{\text{inside each treatment}} = 4 + 4 + 4 + 4 = 16$$

The F-ratio for these data has $df = 3, 16$. The distribution of all the possible F-ratios with $df = 3, 16$ is presented in Figure 12.9. Note that F-ratios larger than 3.24 are extremely rare ($p < .05$) if H_0 is true and, therefore, form the critical region for the test.

STEP 3: Compute the F-ratio.

The series of calculations for computing F is presented in Figure 12.5 and can be summarized as follows:

a. Analyze the SS to obtain SS_{between} and SS_{within}.

b. Use the SS values and the df values (from step 2) to calculate the two variances, MS_{between} and MS_{within}.

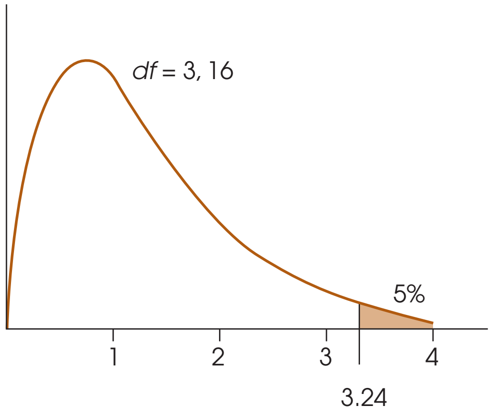
c. Finally, use the two MS values (variances) to compute the F-ratio.

Analysis of SS. First, we compute the total SS and then the two components, as indicated in Figure 12.6.

SS_{total} is simply the SS for the total set of $N = 20$ scores.

FIGURE 12.9

The distribution of F-ratios with df = 3, 16. The critical value for α = .05 is F = 3.24.



$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{N}$$
$$= 262 - \frac{60^2}{20}$$
$$= 262 - 180$$
$$= 82$$

SS_{within} combines the SS values from inside each of the treatment conditions.

$$SS_{\text{within}} = \Sigma SS_{\text{inside each treatment}} = 8 + 8 + 10 + 6 = 32$$

SS_{between} measures the differences among the four treatment means (or treatment totals). Because we have already calculated SS_{total} and SS_{within}, the simplest way to obtain SS_{between} is by subtraction (Equation 12.5).

$$SS_{\text{between}} = SS_{\text{total}} - SS_{\text{within}}$$
$$= 82 - 32$$
$$= 50$$

Calculation of mean squares. Because we already found $df_{\text{between}} = 3$ and $df_{\text{within}} = 16$ (Step 2), we now can compute the variance or MS value for each of the two components.

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{50}{3} = 16.67$$
$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{32}{16} = 2.00$$

Calculation of F. We compute the F-ratio:

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{16.67}{2.00} = 8.33$$

STEP 4: Make a decision.

The F value we obtained, $F = 8.33$, is in the critical region (see Figure 12.9). It is very unlikely ($p < .05$) that we would obtain a value this large if H_0 is true. Therefore, we reject H_0 and conclude that there is a significant treatment effect.

Example 12.1 demonstrated the complete, step-by-step application of the ANOVA procedure. There are two additional points that can be made using this example.

First, you should look carefully at the statistical decision. We have rejected H_0 and concluded that not all the treatments are the same. But we have not determined which ones are different. Is a 9-foot distance different from 12 feet? Is 12 feet different from 15 feet? Unfortunately, these questions remain unanswered. We do know that at least one difference exists (we rejected H_0), but additional analysis is necessary to find out exactly where this difference is. We address this problem in Section 12.6.

Second, as noted earlier, all of the components of the analysis (the *SS*, *df*, *MS*, and *F*) can be presented together in one summary table. The summary table for the analysis in Example 12.1 is as follows:

Source	*SS*	*df*	*MS*	
Between treatments	50	3	16.67	$F = 8.33$
Within treatments	32	16	2.00	
Total	82	19		

Although these tables are very useful for organizing the components of an ANOVA, they are not commonly used in published reports. The current method for reporting the results from an ANOVA is presented on page 409.

MEASURING EFFECT SIZE FOR ANOVA

As we noted previously, a *significant* mean difference simply indicates that the difference observed in the sample data is very unlikely to have occurred just by chance. Thus, the term significant does not necessarily mean *large,* it simply means larger than expected by chance. To provide an indication of how large the effect actually is, researchers should report a measure of effect size in addition to the measure of significance.

For ANOVA, the simplest and most direct way to measure effect size is to compute the percentage of variance accounted for by the treatment conditions. Like the r^2 value used to measure effect size for the t tests in Chapters 9, 10, and 11, this percentage measures how much of the variability in the scores is accounted for by the differences between treatments. For ANOVA, the calculation and the concept of the percentage of variance is extremely straightforward. Specifically, we determine how much of the total SS is accounted for by the $SS_{\text{between treatments}}$.

$$\text{The percentage of variance accounted for } = \frac{SS_{\text{between treatments}}}{SS_{\text{total}}} \tag{12.14}$$

For the data in Example 12.1, the percentage of variance accounted for $= \frac{50}{82} = 0.61$ (or 61%).

In published reports of ANOVA results, the percentage of variance accounted for by the treatment effect is usually called η^2 (the Greek letter *eta squared*) instead of using r^2. Thus, for the study in Example 12.1, $\eta^2 = 0.61$.

IN THE LITERATURE

REPORTING THE RESULTS OF ANOVA

The APA format for reporting the results of ANOVA begins with a presentation of the treatment means and standard deviations in the narrative of the article, a table, or a graph. These descriptive statistics are not part of the calculations for the ANOVA, but you can easily determine the treatment means from n and T ($M = T/n$) and the standard deviations from the SS and $n-1$ values for each treatment. Next, report the results of the ANOVA. For the study described in Example 12.1, the report might state the following:

The means and standard deviations are presented in Table 1. The analysis of variance indicates that there are significant differences among the four viewing distances, $F(3, 16) = 8.33$, $p < .05$, $\eta^2 = 0.61$.

TABLE 1

Ratings of satisfaction with different television viewing distances.

	9 Feet	12 Feet	15 Feet	18 Feet
M	1.00	2.00	5.00	4.00
SD	1.41	1.41	1.58	1.22

Note how the F-ratio is reported. In this example, degrees of freedom for between and within treatments are $df = 3, 16$, respectively. These values are placed in parentheses immediately following the symbol F. Next, the calculated value for F is reported, followed by the probability of committing a Type I error (the alpha level) and the measure of effect size.

When an ANOVA is done using a computer program, the F-ratio is usually accompanied by an exact value for p. The data from Example 12.1 were analyzed using the SPSS program (see Resources at the end of this chapter) and the computer output included a significance level of $p = .001$. Using the exact p value from the computer output, the research report would conclude, "The analysis of variance revealed significant differences among the four viewing distances, $F(3, 16) = 8.33$, $p = .001$, $\eta^2 = 0.61$."

A CONCEPTUAL VIEW OF ANOVA

Because ANOVA requires relatively complex calculations, students encountering this statistical technique for the first time often tend to be overwhelmed by the formulas and arithmetic and lose sight of the general purpose for the analysis. The following two examples are intended to minimize the role of the formulas and shift attention back to the conceptual goal of the ANOVA process.

EXAMPLE 12.2 The following data represent the outcome of an experiment using two separate samples to evaluate the mean difference between two treatment conditions. Take a minute to look at the data and, without doing any calculations, try to predict the outcome of an ANOVA for these values. Specifically, predict what values should be obtained for the between-treatments variance (*MS*) and the *F*-ratio. If you do not "see" the answer after 20 or 30 seconds, try reading the hints that follow the data.

Treatment I	Treatment II	
4	2	$N = 8$
0	1	$G = 16$
1	0	$\Sigma X^2 = 56$
3	5	
$T = 8$ $SS = 10$	$T = 8$ $SS = 14$	

If you are having trouble predicting the outcome of the ANOVA, read the following hints, and then go back and look at the data.

Hint 1: Remember: SS_{between} and MS_{between} provide a measure of how much difference there is between treatment conditions.

Hint 2: Find the mean or total (*T*) for each treatment, and determine how much difference there is between the two treatments.

You should realize by now that the data have been constructed so that there is zero difference between treatments. The two sample means (and totals) are identical, so $SS_{\text{between}} = 0$, $MS_{\text{between}} = 0$, and the *F*-ratio is zero.

Conceptually, the numerator of the *F*-ratio always measures how much difference exists between treatments. In Example 12.2, we constructed an extreme set of scores with zero difference. However, you should be able to look at any set of data and quickly compare the means (or totals) to determine whether there are big differences or small differences between treatments.

Being able to estimate the magnitude of between-treatment differences is a good first step in understanding ANOVA and should help you to predict the outcome of an ANOVA. However, the *between-treatment* differences are only one part of the analysis. You must also understand the *within-treatment* differences that form the denominator of the *F*-ratio. The following example is intended to demonstrate the concepts underlying SS_{within} and MS_{within}. In addition, the example should give you a better understanding of how the between-treatment differences and the within-treatment differences act together within the ANOVA.

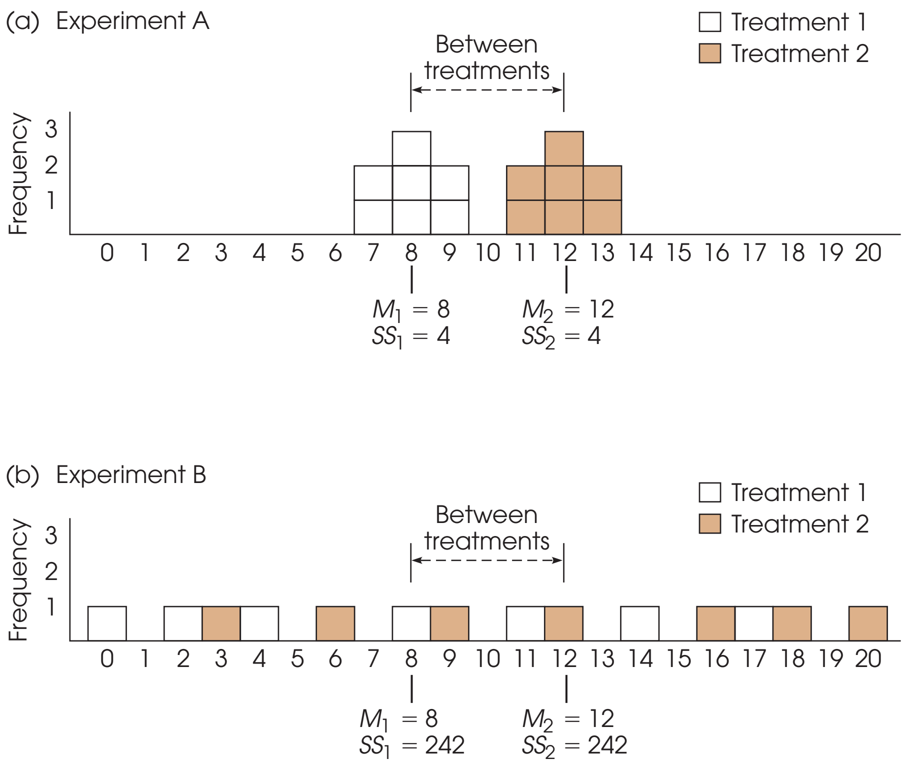
EXAMPLE 12.3 The purpose of this example is to present a visual image for the concepts of between-treatments variability and within-treatments variability. In this example, we compare two hypothetical outcomes for the same experiment. In each case, the experiment uses two separate samples to evaluate the mean difference between two treatments. The following data represent the two outcomes, which we call experiment A and experiment B.

Experiment A		Experiment B	
Treatment		Treatment	
I	II	I	II
8	12	4	12
8	13	11	9
7	12	2	20
9	11	17	6
8	13	0	16
9	12	8	18
7	11	14	3
$M = 8$ $s = 0.82$	$M = 12$ $s = 0.82$	$M = 8$ $s = 6.35$	$M = 12$ $s = 6.35$

The data from experiment A are displayed in a frequency distribution graph in Figure 12.10(a). Notice that there is a 4-point difference between the treatment means ($M_1 = 8$ and $M_2 = 12$). This is the *between-treatments* difference that contributes to the numerator of the F-ratio. Also notice that the scores in each treatment are clustered closely around the mean, indicating that the variance inside each treatment is relatively small. This is the *within-treatments* variance that contributes to the denominator of the F-ratio. Finally, you should realize that it is easy to see the mean difference between the two samples. The fact that there is a clear mean difference between the two treatments is confirmed by computing the F-ratio for experiment A.

FIGURE 12.10

A visual representation of the between-treatments variability and the within-treatments variability that form the numerator and denominator, respectively, of the F-ratio. In (a), the difference between treatments is relatively large and easy to see. In (b), the same 4-point difference between treatments is relatively small and is overwhelmed by the within-treatments variability.



$$F = \frac{\text{between-treatments difference}}{\text{within-treatments differences}} = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{56}{0.667} = 83.96$$

An F-ratio of $F = 83.96$ is sufficient to reject the null hypothesis, so we conclude that there is a significant difference between the two treatments.

Now consider the data from experiment B, which are shown in Figure 12.10(b) and present a very different picture. This experiment has the same 4-point difference between treatment means that we found in experiment A ($M_1 = 8$ and $M_2 = 12$). However, for these data the scores in each treatment are scattered across the entire scale, indicating relatively large variance inside each treatment. In this case, the large variance within treatments overwhelms the relatively small mean difference between treatments. In the figure it is almost impossible to see the mean difference between treatments. For these data, the F-ratio confirms that there is no clear mean difference between treatments.

$$F = \frac{\text{between-treatments difference}}{\text{within-treatments differences}} = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{56}{40.33} = 1.39$$

For experiment B, the F-ratio is not large enough to reject the null hypothesis, so we conclude that there is no significant difference between the two treatments. Once again, the statistical conclusion is consistent with the appearance of the data in Figure 12.10(b). Looking at the figure, we see that the scores from the two samples appear to be intermixed randomly with no clear distinction between treatments.

As a final point, note that the denominator of the F-ratio, MS_{within}, is a measure of the variability (or variance) within each of the separate samples. As we have noted in previous chapters, high variability makes it difficult to see any patterns in the data. In Figure 12.10(a), the 4-point mean difference between treatments is easy to see because the sample variability is small. In Figure 12.10(b), the 4-point difference gets lost because the sample variability is large. In general, you can think of variance as measuring the amount of "noise" or "confusion" in the data. With large variance, there is a lot of noise and confusion and it is difficult to see any clear patterns.

Although Examples 12.2 and 12.3 present somewhat simplified demonstrations with exaggerated data, the general point of the examples is to help you *see* what happens when you perform an ANOVA. Specifically:

1. The numerator of the F-ratio (MS_{between}) measures how much difference exists between the treatment means. The bigger the mean differences, the bigger the F-ratio.
2. The denominator of the F-ratio (MS_{within}) measures the variance of the scores inside each treatment; that is, the variance for each of the separate samples. In general, larger sample variance produces a smaller F-ratio.

We should note that the number of scores in the samples also can influence the outcome of an ANOVA. As with most other hypothesis tests, if other factors are held constant, increasing the sample size tends to increase the likelihood of rejecting the null hypothesis. However, changes in sample size have little or no effect on measures of effect size such as η^2.

Finally, we should note that the problems associated with high variance often can be minimized by transforming the original scores to ranks and then conducting an alternative statistical analysis known as the *Kruskal-Wallis test,* which is designed specifically for ordinal data. The Kruskal-Wallis test is presented in Appendix E, which also discusses the general purpose and process of converting numerical scores into ranks. The Kruskal-Wallis test also can be used if the data violate one of the assumptions for the independent-measures *ANOVA*, which are outlined at the end of section 12.7.

MS_{within} AND POOLED VARIANCE

You may have recognized that the two research outcomes presented in Example 12.3 are similar to those presented earlier in Example 10.5 in Chapter 10. Both examples are intended to demonstrate the role of variance in a hypothesis test. Both examples show that large values for sample variance can obscure any patterns in the data and reduce the potential for finding significant differences between means.

For the independent-measures *t* statistic in Chapter 10, the sample variance contributed directly to the standard error in the bottom of the *t* formula. Now, the sample variance contributes directly to the value of MS_{within} in the bottom of the *F*-ratio. In the *t*-statistic and in the *F*-ratio the variances from the separate samples are pooled together to create one average value for sample variance. For the independent-measures *t* statistic, we pooled two samples together to compute

$$\text{pooled variance} = s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

Now, in ANOVA, we are combining two or more samples to calculate

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{\Sigma SS}{\Sigma df} = \frac{SS_1 + SS_2 + SS_3 + \cdots}{df_1 + df_2 + df_3 + \cdots}$$

Notice that the concept of pooled variance is the same whether you have exactly two samples or more than two samples. In either case, you simply add the *SS* values and divide by the sum of the *df* values. The result is an average of all of the different sample variances.

AN EXAMPLE WITH UNEQUAL SAMPLE SIZES

In the previous examples, all of the samples were exactly the same size (equal *n*s). However, the formulas for ANOVA can be used when the sample size varies within an experiment. You also should note, however, that the general ANOVA procedure is most accurate when used to examine experimental data with equal sample sizes. Therefore, researchers generally try to plan experiments with equal *n*s. However, there are circumstances in which it is impossible or impractical to have an equal number of subjects in every treatment condition. In these situations, ANOVA still provides a valid test, especially when the samples are relatively large and when the discrepancy between sample sizes is not extreme.

The following example demonstrates an ANOVA with samples of different sizes.

EXAMPLE 12.4

A researcher is interested in the amount of homework required by different academic majors. Students are recruited from Biology, English, and Psychology to participant in the study. The researcher randomly selects one course that each student is currently taking and asks the student to record the amount of out-of-class work required each week for the course. The researcher used all of the volunteer participants, which resulted in unequal sample sizes. The data are summarized in Table 12.5.

STEP 1: State the hypotheses, and select the alpha level.

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

H_1: At least one population is different.

$$\alpha = .05$$

STEP 2: Locate the critical region.

To find the critical region, we first must determine the *df* values for the *F*-ratio:

$$df_{\text{total}} = N - 1 = 20 - 1 = 19$$

$$df_{\text{between}} = k - 1 = 3 - 1 = 2$$

$$df_{\text{within}} = N - k = 20 - 3 = 17$$

The *F*-ratio for these data has $df = 2, 17$. With $\alpha = .05$, the critical value for the *F*-ratio is 3.59.

STEP 3: Compute the *F*-ratio.

First, compute the three *SS* values. As usual, SS_{total} is the *SS* for the total set of $N = 20$ scores, and SS_{within} combines the *SS* values from inside each of the treatment conditions.

$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{N} = 3377 - \frac{250^2}{20} = 3377 - 3125 = 252$$

$$SS_{\text{within}} = \Sigma SS_{\text{inside each treatment}} = 37 + 90 + 60 = 187$$

SS_{between} can be found by subtraction (Equation 12.5).

$$SS_{\text{between}} = SS_{\text{total}} - SS_{\text{within}} = 252 - 187 = 65$$

TABLE 12.5

Average hours of homework per week for one course for students in three academic majors.

Biology	English	Psychology	
$n = 4$	$n = 10$	$n = 6$	$N = 20$
$M = 9$	$M = 13$	$M = 14$	$G = 250$
$T = 36$	$T = 130$	$T = 84$	$\Sigma X^2 = 3377$
$SS = 37$	$SS = 90$	$SS = 60$	

Or, SS_{between} can be calculated using the computation formula (Equation 12.7). If you use the computational formula, be careful to match each treatment total (T) with the appropriate sample size (n) as follows:

$$\begin{aligned} SS_{\text{between}} &= \Sigma\frac{T^2}{n} - \frac{G^2}{N} \\ &= \frac{36^2}{4} + \frac{130^2}{10} + \frac{84^2}{6} - \frac{250^2}{20} \\ &= 324 + 1690 + 1176 - 3125 \\ &= 65 \end{aligned}$$

Finally, compute the MS values and the F-ratio:

$$MS_{\text{between}} = \frac{SS}{df} = \frac{65}{2} = 32.5$$

$$MS_{\text{within}} = \frac{SS}{df} = \frac{187}{17} = 11$$

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{32.5}{11} = 2.95$$

STEP 4: Make a decision.

Because the obtained F-ratio is not in the critical region, we fail to reject the null hypothesis and conclude that there are no significant differences among the three populations of students in terms of the average amount of homework each week.

LEARNING CHECK

1. A researcher used ANOVA and computed $F = 4.25$ for the following data.

Treatments		
I	II	III
$n = 10$	$n = 10$	$n = 10$
$M = 20$	$M = 28$	$M = 35$
$SS = 1005$	$SS = 1391$	$SS = 1180$

a. If the mean for treatment III were changed to $M = 25$, what would happen to the size of the F-ratio (increase or decrease)? Explain your answer.

b. If the SS for treatment I were changed to $SS = 1400$, what would happen to the size of the F-ratio (increase or decrease)? Explain your answer.

2. A research study comparing three treatment conditions produces $T = 20$ with $n = 4$ for the first treatment, $T = 10$ with $n = 5$ for the second treatment, and $T = 30$ with $n = 6$ for the third treatment. Calculate $SS_{\text{between treatments}}$ for these data.

ANSWERS

1. a. If the mean for treatment III were changed to $M = 25$, it would reduce the size of the mean differences (the three means would be closer together). This would reduce the size of MS_{between} and would reduce the size of the F-ratio.

b. If the SS in treatment I were increased to $SS = 1400$, it would increase the size of the variability within treatments. This would increase MS_{within} and would reduce the size of the F-ratio.

2. With $G = 60$ and $N = 15$, $SS_{\text{between}} = 30$.

12.6 POST HOC TESTS

As noted earlier, the primary advantage of ANOVA (compared to t tests) is that it allows researchers to test for significant mean differences when there are *more than two* treatment conditions. ANOVA accomplishes this feat by comparing all the individual mean differences simultaneously within a single test. Unfortunately, the process of combining several mean differences into a single test statistic creates some difficulty when it is time to interpret the outcome of the test. Specifically, when you obtain a significant F-ratio (reject H_0), it simply indicates that somewhere among the entire set of mean differences there is at least one that is statistically significant. In other words, the overall F-ratio only tells you that a significant difference exists; it does not tell exactly which means are significantly different and which are not.

Consider, for example, a research study that uses three samples to compare three treatment conditions. Suppose that the three sample means are $M_1 = 3$, $M_2 = 5$, and $M_3 = 10$. In this hypothetical study there are three mean differences:

1. There is a 2-point difference between M_1 and M_2.
2. There is a 5-point difference between M_2 and M_3.
3. There is a 7-point difference between M_1 and M_3.

If an ANOVA were used to evaluate these data, a significant F-ratio would indicate that at least one of the sample mean differences is large enough to satisfy the criterion of statistical significance. In this example, the 7-point difference is the biggest of the three and, therefore, it must indicate a significant difference between the first treatment and the third treatment ($\mu_1 \neq \mu_3$). But what about the 5-point difference? Is it also large enough to be significant? And what about the 2-point difference between M_1 and M_2? Is it also significant? The purpose of *post hoc tests* is to answer these questions.

DEFINITION

Post hoc tests (or **posttests**) are additional hypothesis tests that are done after an ANOVA to determine exactly which mean differences are significant and which are not.

As the name implies, post hoc tests are done after an ANOVA. More specifically, these tests are done after ANOVA when

1. You reject H_0 and
2. There are three or more treatments ($k \geq 3$).

Rejecting H_0 indicates that at least one difference exists among the treatments. If there are only two treatments, then there is no question about which means are different and, therefore, no need for posttests. However, with three or more treatments (k $\geq$ 3), the problem is to determine exactly which means are significantly different.

POSTTESTS AND TYPE I ERRORS

In general, a post hoc test enables you to go back through the data and compare the individual treatments two at a time. In statistical terms, this is called making *pairwise comparisons*. For example, with $k = 3$, we would compare μ_1 versus μ_2, then μ_2 versus μ_3, and then μ_1 versus μ_3. In each case, we are looking for a significant mean difference. The process of conducting pairwise comparisons involves performing a series of separate hypothesis tests, and each of these tests includes the risk of a Type I error. As you do more and more separate tests, the risk of a Type I error accumulates and is called the *experimentwise alpha level* (see p. 391).

We have seen, for example, that a research study with three treatment conditions produces three separate mean differences, each of which could be evaluated using a post hoc test. If each test uses $\alpha = .05$, then there is a 5% risk of a Type I error for the first posttest, another 5% risk for the second test, and one more 5% risk for the third test. Although the probability of error is not simply the sum across the three tests, it should be clear that increasing the number of separate tests definitely increases the total, experimentwise probability of a Type I error.

Whenever you are conducting posttests, you must be concerned about the experimentwise alpha level. Statisticians have worked with this problem and have developed several methods for trying to control Type I errors in the context of post hoc tests. We consider two alternatives.

TUKEY'S HONESTLY SIGNIFICANT DIFFERENCE (HSD) TEST

The first post hoc test we consider is *Tukey's HSD test.* We selected Tukey's HSD test because it is a commonly used test in psychological research. Tukey's test allows you to compute a single value that determines the minimum difference between treatment means that is necessary for significance. This value, called the *honestly significant difference,* or HSD, is then used to compare any two treatment conditions. If the mean difference exceeds Tukey's HSD, then you conclude that there is a significant difference between the treatments. Otherwise, you cannot conclude that the treatments are significantly different. The formula for Tukey's HSD is

$$HSD = q\sqrt{\frac{MS_{\text{within}}}{n}} \tag{12.15}$$

The q value used in Tukey's HSD test is called a Studentized range statistic.

where the value of q is found in Table B.5 (Appendix B, p. 708), MS_{within} is the within-treatments variance from the ANOVA, and n is the number of scores in each treatment. Tukey's test requires that the sample size, n, be the same for all treatments. To locate the appropriate value of q, you must know the number of treatments in the overall experiment (k), the degrees of freedom for MS_{within} (the error term in the F-ratio), and you must select an alpha level (generally the same α used for the ANOVA).

EXAMPLE 12.5

To demonstrate the procedure for conducting post hoc tests with Tukey's HSD, we use the hypothetical data shown in Table 12.6. The data represent the results of a study comparing scores in three different treatment conditions. Note that the table displays summary statistics for each sample and the results from the overall ANOVA. With $k = 3$ treatments, $df_{\text{within}} = 24$, and $\alpha = .05$, you should find that the value of q for the test is $q = 3.53$ (see Table B.5). Therefore, Tukey's HSD is

$$HSD = q\sqrt{\frac{MS_{\text{within}}}{n}} = 3.53\sqrt{\frac{4.00}{9}} = 2.36$$

TABLE 12.6

Hypothetical results from a research study comparing three treatment conditions. Summary statistics are presented for each treatment along with the outcome from the ANOVA.

Treatment A	Treatment B	Treatment C
$n = 9$	$n = 9$	$n = 9$
$T = 27$	$T = 49$	$T = 63$
$M = 3.00$	$M = 5.44$	$M = 7.00$

Source	*SS*	*df*	*MS*
Between	73.19	2	36.60
Within	96.00	24	4.00
Total	169.19	26	
Overall $F(2, 24) = 9.15$			

Thus, the mean difference between any two samples must be at least 2.36 to be significant. Using this value, we can make the following conclusions:

1. Treatment A is significantly different from treatment B ($M_A - M_B = 2.44$).
2. Treatment A is also significantly different from treatment C ($M_A - M_C = 4.00$).
3. Treatment B is not significantly different from treatment C ($M_B - M_C = 1.56$).

THE SCHEFFÉ TEST

Because it uses an extremely cautious method for reducing the risk of a Type I error, the *Scheffé test* has the distinction of being one of the safest of all possible post hoc tests (smallest risk of a Type I error). The Scheffé test uses an F-ratio to evaluate the significance of the difference between any two treatment conditions. The numerator of the F-ratio is an MS_{between} that is calculated using *only the two treatments you want to compare*. The denominator is the same MS_{within} that was used for the overall ANOVA. The "safety factor" for the Scheffé test comes from the following two considerations:

1. Although you are comparing only two treatments, the Scheffé test uses the value of k from the original experiment to compute df between treatments. Thus, df for the numerator of the F-ratio is $k - 1$.
2. The critical value for the Scheffé F-ratio is the same as was used to evaluate the F-ratio from the overall ANOVA. Thus, Scheffé requires that every posttest satisfy the same criterion that was used for the complete ANOVA. The following example uses the data from Table 12.6 to demonstrate the Scheffé posttest procedure.

EXAMPLE 12.6

Remember that the Scheffé procedure requires a separate SS_{between}, MS_{between}, and F-ratio for each comparison being made. Although Scheffé computes SS_{between} using the regular computational formula (Equation 12.7), you must remember that all of the numbers in the formula are entirely determined by the two treatment conditions being compared. We begin by comparing treatment A (with $T = 27$ and $n = 9$) and treatment B (with $T = 49$ and $n = 9$). The first step is to compute SS_{between} for these two groups. In the formula for SS, notice that the grand total for the two groups is $G = 27 + 49 = 76$, and the total number of scores for the two groups is $N = 9 + 9 = 18$.

$$\begin{aligned} SS_{\text{between}} &= \Sigma\frac{T^2}{n} - \frac{G^2}{N} \\ &= \frac{27^2}{9} + \frac{49^2}{9} - \frac{76^2}{18} \\ &= 81 + 266.78 - 320.89 \\ &= 26.89 \end{aligned}$$

Although we are comparing only two groups, these two were selected from a study consisting of $k = 3$ samples. The Scheffé test uses the overall study to determine the degrees of freedom between treatments. Therefore, $df_{\text{between}} = 3 - 1 = 2$, and the MS_{between} is

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{26.89}{2} = 13.45$$

Finally, the Scheffé procedure uses the error term from the overall ANOVA to compute the F-ratio. In this case, $MS_{\text{within}} = 4.00$ with $df_{\text{within}} = 24$. Thus, the Scheffé test produces an F-ratio of

$$F_{\text{A verus B}} = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{13.45}{4.00} = 3.36$$

With $df = 2, 24$ and $\alpha = .05$, the critical value for F is 3.40 (see Table B.4). Therefore, our obtained F-ratio is not in the critical region, and we must conclude that these data show no significant difference between treatment A and treatment B.

The second comparison involves treatment B ($T = 49$) and treatment C ($T = 63$). This time the data produce $SS_{\text{between}} = 10.89$, $MS_{\text{between}} = 5.45$, and $F(2, 24) = 1.36$ (check the calculations for yourself). Once again the critical value for F is 3.40, so we must conclude that the data show no significant difference between treatment B and treatment C.

The final comparison is treatment A ($T = 27$) and treatment C ($T = 63$). This time the data produce $SS_{\text{between}} = 72$, $MS_{\text{between}} = 36$, and $F(2, 24) = 9.00$ (check the calculations for yourself). Once again the critical value for F is 3.40, and this time we conclude that the data show a significant difference.

Thus, the Scheffé posttest indicates that the only significant difference is between treatment A and treatment C.

There are two interesting points to be made from the posttest outcomes presented in the preceding two examples. First, the Scheffé test was introduced as being one of the safest of the posttest techniques because it provides the greatest protection from Type I errors. To provide this protection, the Scheffé test simply requires a larger difference between sample means before you may conclude that the difference is significant. For example, using Tukey's test in Example 12.5, we found that the difference between treatment A and treatment B was large enough to be significant. However, this same difference failed to reach significance according to the Scheffé test (Example 12.6). The discrepancy between the results is an example of the Scheffé test's extra demands: The Scheffé test simply requires more evidence and, therefore, it is less likely to lead to a Type I error.

The second point concerns the pattern of results from the three Scheffé tests in Example 12.6. You may have noticed that the posttests produce what are apparently contradictory results. Specifically, the tests show no significant difference between A and B and they show no significant difference between B and C. This combination of outcomes might lead you to suspect that there is no significant difference between A and C. However, the test did show a significant difference. The answer to this apparent contradiction lies in the criterion of statistical significance. The differences between A and B and between B and C are too small to satisfy the criterion of significance. However, when these differences are combined, the total difference between A and C is large enough to meet the criterion for significance.

LEARNING CHECK

1. With $k = 2$ treatments, are post hoc tests necessary when the null hypothesis is rejected? Explain why or why not.
2. An ANOVA comparing three treatments produces an overall F-ratio with $df = 2, 27$. If the Scheffé test was used to compare two of the three treatments, then the Scheffé F-ratio would also have $df = 2, 27$. (True or false?)

3. Using the data and the results from Example 12.1,
   - **a.** Use Tukey's HSD test to determine whether there is a significant mean difference between a 12-foot and a 15-foot distance. Use $\alpha = .05$.
   - **b.** Use the Scheffé test to determine whether there is a significant mean difference between 12 feet and 15 feet. Use $\alpha = .05$.

ANSWERS

1. No. Post hoc tests are used to determine which treatments are different. With only two treatment conditions, there is no uncertainty as to which two treatments are different.
2. True
3. **a.** For this test, $q = 4.05$ and HSD $= 2.55$. There is a 3-point mean difference between 12 feet and 15 feet, which is large enough to be significant.

   b. The Scheffé $F = 3.75$, which is greater than the critical value of 3.24. Conclude that the mean difference between 12 feet and 15 feet is significant.

12.7 THE RELATIONSHIP BETWEEN ANOVA AND *t* TESTS

When you are evaluating the mean difference from an independent-measures study comparing only two treatments (two separate samples), you can use either an independent-measures *t* test (Chapter 10) or the ANOVA presented in this chapter. In practical terms, it makes no difference which you choose. These two statistical techniques always result in the same statistical decision. In fact the two methods use many of the same calculations and are very closely related in several other respects. The basic relationship between *t* statistics and *F*-ratios can be stated in an equation:

$$F = t^2$$

This relationship can be explained by first looking at the structure of the formulas for *F* and *t*. The *t* statistic compares *distances*: the distance between two sample means (numerator) and the distance computed for the standard error (denominator). The *F*-ratio, on the other hand, compares *variances*. You should recall that variance is a measure of squared distance. Hence, the relationship: $F = t^2$.

There are several other points to consider in comparing the *t* statistic to the *F*-ratio.

1. It should be obvious that you are testing the same hypotheses whether you choose a *t* test or an ANOVA. With only two treatments, the hypotheses for either test are

$$H_0\colon \mu_1 = \mu_2$$
$$H_1\colon \mu_1 \neq \mu_2$$

2. The degrees of freedom for the *t* statistic and the *df* for the denominator of the *F*-ratio (df_{within}) are identical. For example, if you have two samples, each with six scores, the independent-measures *t* statistic has $df = 10$, and the *F*-ratio has $df = 1, 10$. In each case, you are adding the *df* from the first sample $(n - 1)$ and the *df* from the second sample $(n - 1)$.
3. The distribution of *t* and the distribution of *F*-ratios match perfectly if you take into consideration the relationship $F = t^2$. Consider the *t* distribution with $df = 18$ and the corresponding *F* distribution with $df = 1, 18$ that are presented in Figure 12.11. Notice the following relationships:
   - **a.** If each of the *t* values is squared, then all of the negative values become positive. As a result, the whole left-hand side of the *t* distribution (below

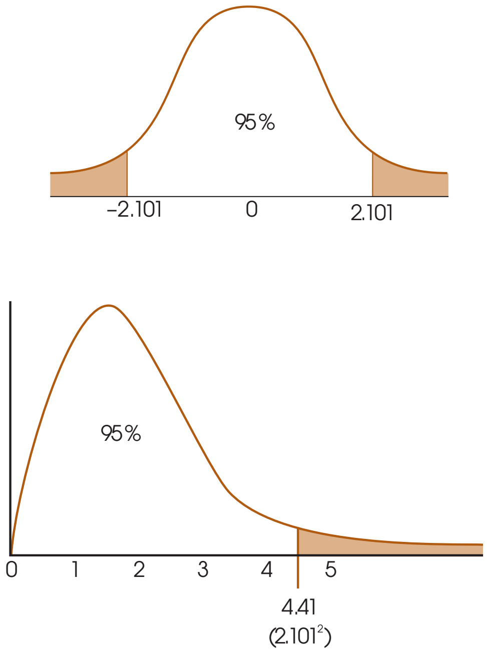


FIGURE 12.11

The distribution of *t* statistics with *df* = 18 and the corresponding distribution of *F*-ratios with *df* = 1, 18. Notice that the critical values for α = .05 are $t = \pm 2.101$ and that $F = 2.101^2 = 4.41$.

zero) is flipped over to the positive side. This creates an asymmetrical, positively skewed distribution—that is, the *F* distribution.

b. For α = .05, the critical region for *t* is determined by values greater than +2.101 or less than −2.101. When these boundaries are squared, you get $\pm 2.101^2 = 4.41$

Notice that 4.41 is the critical value for α = .05 in the *F* distribution. Any value that is in the critical region for *t* ends up in the critical region for *F*-ratios after it is squared.

ASSUMPTIONS FOR THE INDEPENDENT-MEASURES ANOVA

The independent-measures ANOVA requires the same three assumptions that were necessary for the independent-measures *t* hypothesis test:

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).

Ordinarily, researchers are not overly concerned with the assumption of normality, especially when large samples are used, unless there are strong reasons to suspect that the assumption has not been satisfied. The assumption of homogeneity of variance is an important one. If a researcher suspects that it has been violated, it can be tested by Hartley's *F*-max test for homogeneity of variance (Chapter 10, p. 338).

Finally, if you suspect that one of the assumptions for the independent-measures ANOVA has been violated, you can still proceed by transforming the original scores into ranks and then using an alternative statistical analysis known as the Kruskal-Wallis test, which is designed specifically for ordinal data. The Kruskal-Wallis test is presented in Appendix E. As noted earlier, the Kruskal-Wallis test also can be useful if large sample variance prevents the independent-measures ANOVA from producing a significant result.

LEARNING CHECK

1. A researcher uses an independent-measures *t* test to evaluate the mean difference obtained in a research study, and obtains a *t* statistic of $t = 3.00$. If the researcher had used an ANOVA to evaluate the results, the *F*-ratio would be $F = 9.00$. (True or false?)
2. An ANOVA produces an *F*-ratio with $df = 1, 34$. Could the data have been analyzed with a *t* test? What would be the degrees of freedom for the *t* statistic?

ANSWERS

1. True. $F = t^2$
2. If the *F*-ratio has $df = 1, 34$, then the experiment compared only two treatments, and you could use a *t* statistic to evaluate the data. The *t* statistic would have $df = 34$.

SUMMARY

1. Analysis of variance (ANOVA) is a statistical technique that is used to test the significance of mean differences among two or more treatment conditions. The null hypothesis for this test states that, in the general population, there are no mean differences among the treatments. The alternative states that at least one mean is different from another.
2. The test statistic for ANOVA is a ratio of two variances called an *F*-ratio. The variances in the *F*-ratio are called mean squares, or *MS* values. Each *MS* is computed by

$$MS = \frac{SS}{df}$$

3. For the independent-measures ANOVA, the *F*-ratio is

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

The MS_{between} measures differences between the treatments by computing the variability of the treatment means or totals. These differences are assumed to be produced by

a. Treatment effects (if they exist)
b. Random, unsystematic differences (chance)

The MS_{within} measures variability inside each of the treatment conditions. Because individuals inside a treatment condition are all treated exactly the same, any differences within treatments cannot be caused by treatment effects. Thus, the within-treatments *MS* is produced only by random, unsystematic differences. With these factors in mind, the *F*-ratio has the following structure:

$$F = \frac{\text{treatment effect} + \text{differences due to chance}}{\text{differences due to chance}}$$

When there is no treatment effect (H_0 is true), the numerator and the denominator of the *F*-ratio are measuring the same variance, and the obtained ratio should be near 1.00. If there is a significant treatment effect, then the numerator of the ratio should be larger than the denominator, and the obtained *F* value should be much greater than 1.00.

4. The formulas for computing each *SS*, *df*, and *MS* value are presented in Figure 12.12, which also shows the general structure for the ANOVA.
5. The *F*-ratio has two values for degrees of freedom, one associated with the *MS* in the numerator and one associated with the *MS* in the denominator. These *df* values are used to find the critical value for the *F*-ratio in the *F* distribution table.
6. Effect size for the independent-measures ANOVA is measured by computing eta squared, the percentage of variance accounted for by the treatment effect.

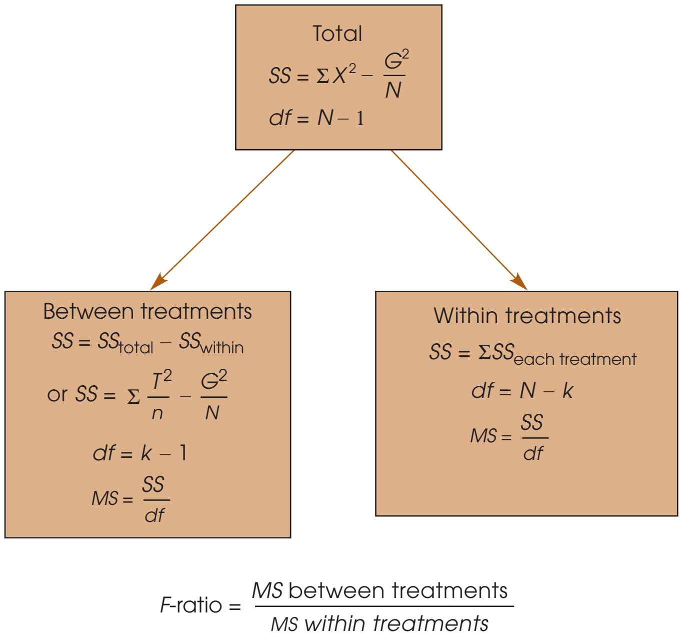

Total
$SS = \Sigma X^2 - \frac{G^2}{N}$
$df = N - 1$
Between treatments
$SS = SS_{total} - SS_{within}$
or $SS = \Sigma \frac{T^2}{n} - \frac{G^2}{N}$
$df = k - 1$
$MS = \frac{SS}{df}$
Within treatments
$SS = \Sigma SS_{each\ treatment}$
$df = N - k$
$MS = \frac{SS}{df}$
$F\text{-ratio} = \frac{MS\ \text{between treatments}}{MS\ \text{within treatments}}$

FIGURE 12.12
Formulas for ANOVA.

$$\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{between}} + SS_{\text{within}}} = \frac{SS_{\text{between}}}{SS_{\text{total}}}$$

7. When the decision from an ANOVA is to reject the null hypothesis and when the experiment contains more than two treatment conditions, it is necessary to continue the analysis with a post hoc test, such as Tukey's HSD test or the Scheffé test. The purpose of these tests is to determine exactly which treatments are significantly different and which are not.

KEY TERMS

analysis of variance (ANOVA) (386)
factor (388)
levels (388)
testwise alpha level (391)
experimentwise alpha level (391)
between-treatments variance (392)
treatment effect (393)
within-treatments variance (393)
F-ratio (394)
error term (394)
mean square (*MS*) (401)
ANOVA summary table (402)
distribution of *F*-ratios (403)
eta squared (η^2) (408)
Kruskal-Wallis test (413)
post hoc tests (416)
pairwise comparisons (416)
Tukey's HSD test (417)
Scheffé test (418)

RESOURCES

Book Companion Website: www.cengage.com/psychology/gravetter.
You can find a tutorial quiz and other learning exercises for Chapter 12 on the book companion website.

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

Data Entry

1. The scores are entered in a *stacked format* in the data editor, which means that all of the scores from all of the different treatments are entered in a single column (VAR00001). Enter the scores for treatment #2 directly beneath the scores from treatment #1 with no gaps or extra spaces. Continue in the same column with the scores from treatment #3, and so on.
2. In the second column (VAR00002), enter a number to identify the treatment condition for each score. For example, enter a 1 beside each score from the first treatment, enter a 2 beside each score from the second treatment, and so on.

Data Analysis

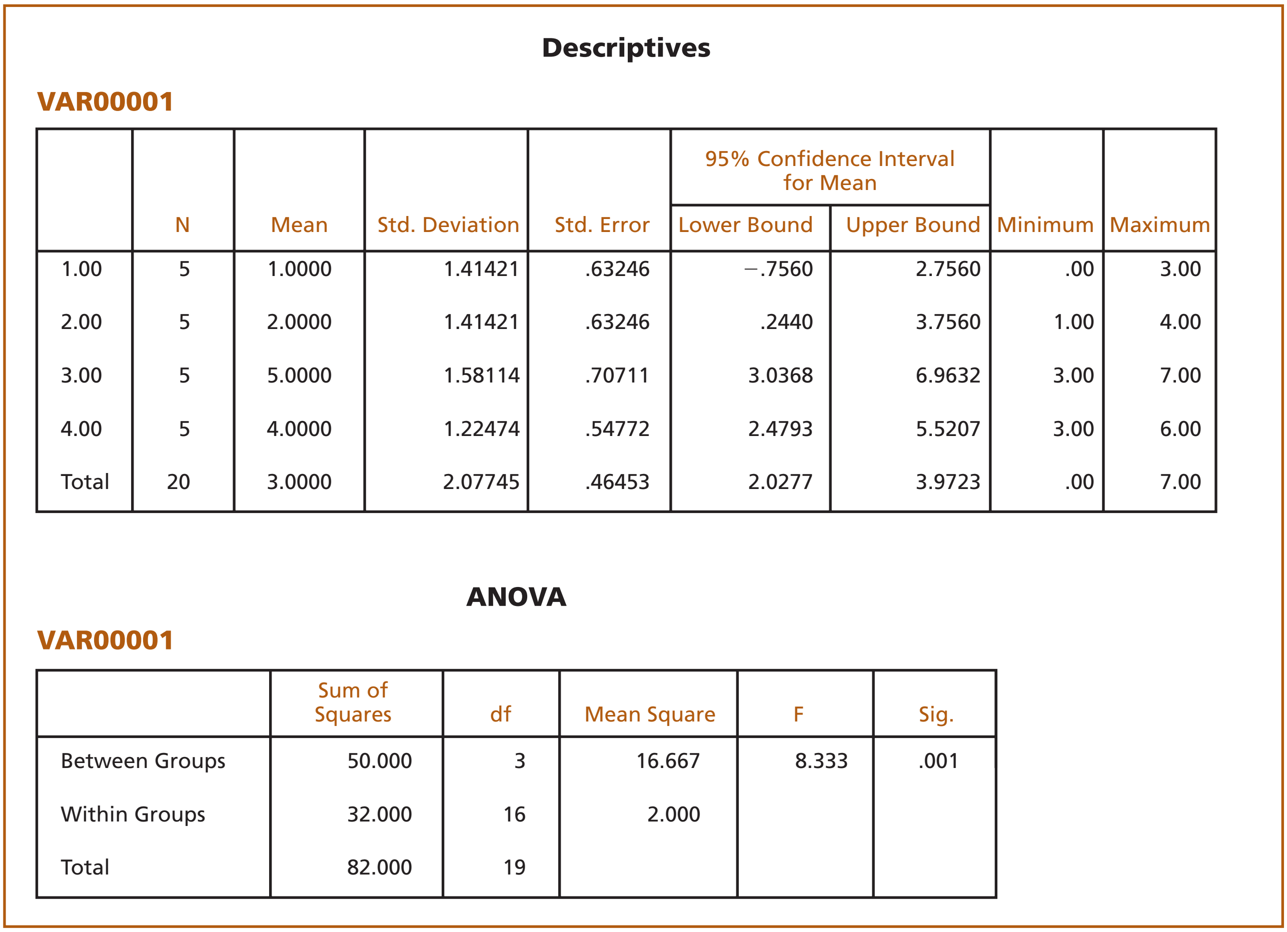
1. Click **Analyze** on the tool bar, select **Compare Means,** and click on **One-Way ANOVA.**
2. Highlight the column label for the set of scores (VAR0001) in the left box and click the arrow to move it into the **Dependent List** box.
3. Highlight the label for the column containing the treatment numbers (VAR0002) in the left box and click the arrow to move it into the **Factor** 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 television viewing study in Example 12.1 and the program output is shown in Figure 12.13. The output begins with a table showing descriptive statistics (number of scores, mean, standard deviation, standard error for the mean, a 95% confidence interval for the mean, maximum and minimum scores) for each sample. The second part of the output presents a summary table showing the results from the ANOVA.

FOCUS ON PROBLEM SOLVING

1. It can be helpful to compute all three *SS* values separately, then check to verify that the two components (between and within) add up to the total. However, you can



Descriptives

VAR00001

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	5	1.0000	1.41421	.63246	−.7560	2.7560	.00	3.00
2.00	5	2.0000	1.41421	.63246	.2440	3.7560	1.00	4.00
3.00	5	5.0000	1.58114	.70711	3.0368	6.9632	3.00	7.00
4.00	5	4.0000	1.22474	.54772	2.4793	5.5207	3.00	6.00
Total	20	3.0000	2.07745	.46453	2.0277	3.9723	.00	7.00

ANOVA

VAR00001

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	50.000	3	16.667	8.333	.001
Within Groups	32.000	16	2.000		
Total	82.000	19			

FIGURE 12.13
SPSS output of the ANOVA for the television-viewing distance study in Example 12.1.

greatly simplify the calculations if you simply find SS_{total} and SS_{within}, then obtain SS_{between} by subtraction.

2. Remember that an F-ratio has two separate values for *df*: a value for the numerator and one for the denominator. Properly reported, the df_{between} value is stated first. You will need both *df* values when consulting the F distribution table for the critical F value. You should recognize immediately that an error has been made if you see an F-ratio reported with a single value for *df*.

3. When you encounter an F-ratio and its *df* values reported in the literature, you should be able to reconstruct much of the original experiment. For example, if you see "$F(2, 36) = 4.80$," you should realize that the experiment compared $k = 3$ treatment groups (because $df_{\text{between}} = k - 1 = 2$), with a total of $N = 39$ subjects participating in the experiment (because $df_{\text{within}} = N - k = 36$).

DEMONSTRATION 12.1

ANALYSIS OF VARIANCE

A human-factors psychologist studied three computer keyboard designs. Three samples of individuals were given material to type on a particular keyboard, and the number of errors committed by each participant was recorded. The data are as follows:

Keyboard A	Keyboard B	Keyboard C	
0	6	6	$N = 15$
4	8	5	$G = 60$
0	5	9	$\Sigma X^2 = 356$
1	4	4	
0	2	6	
$T = 5$ $SS = 12$	$T = 25$ $SS = 20$	$T = 30$ $SS = 14$	

Are these data sufficient to conclude that there are significant differences in typing performance among the three keyboard designs?

STEP 1 **State the hypotheses, and specify the alpha level.** The null hypothesis states that there is no difference among the keyboards in terms of number of errors committed. In symbols, we would state

H_0: $\mu_1 = \mu_2 = \mu_3$ (Type of keyboard used has no effect.)

As noted previously in this chapter, there are a number of possible statements for the alternative hypothesis. Here we state the general alternative hypothesis:

H_1: At least one of the treatment means is different.

We set alpha at $\alpha = .05$.

STEP 2 **Locate the critical region.** To locate the critical region, we must obtain the values for df_{between} and df_{within}.

$$df_{\text{between}} = k - 1 = 3 - 1 = 2$$

$$df_{\text{within}} = N - k = 15 - 3 = 12$$

The F-ratio for this problem has $df = 2, 12$, and the critical F value for $\alpha = .05$ is $F = 3.88$.

STEP 3 **Perform the analysis.** The analysis involves the following steps:

1. Perform the analysis of *SS*.
2. Perform the analysis of *df*.
3. Calculate mean squares.
4. Calculate the *F*-ratio.

Perform the analysis of SS. We compute SS_{total} followed by its two components.

$$\begin{aligned} SS_{\text{total}} &= \Sigma X^2 - \frac{G^2}{N} = 356 - \frac{60^2}{15} = 356 - \frac{3600}{15} \\ &= 356 - 240 = 116 \end{aligned}$$

$$\begin{aligned} SS_{\text{within}} &= \Sigma SS_{\text{inside each treatment}} \\ &= 12 + 20 + 14 \\ &= 46 \end{aligned}$$

$$\begin{aligned} SS_{\text{between}} &= \Sigma\frac{T^2}{n} - \frac{G^2}{N} \\ &= \frac{5^2}{5} + \frac{25^2}{5} + \frac{30^2}{5} - \frac{60^2}{15} \\ &= \frac{25}{5} + \frac{625}{5} + \frac{900}{5} - \frac{3600}{15} \\ &= 5 + 125 + 180 - 240 \\ &= 70 \end{aligned}$$

Analyze degrees of freedom. We compute df_{total}. Its components, df_{between} and df_{within}, were previously calculated (see step 2).

$$df_{\text{total}} = N - 1 = 15 - 1 = 14$$

$$df_{\text{between}} = 2$$

$$df_{\text{within}} = 12$$

Calculate the MS values. We determine the values for MS_{between} and MS_{within}.

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{df_{\text{between}}} = \frac{70}{2} = 35$$

$$MS_{\text{within}} = \frac{SS_{\text{within}}}{df_{\text{within}}} = \frac{46}{12} = 3.83$$

Compute the F-ratio. Finally, we can compute F.

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{35}{3.83} = 9.14$$

STEP 4 **Make a decision about H_0, and state a conclusion.** The obtained F of 9.14 exceeds the critical value of 3.88. Therefore, we can reject the null hypothesis. The type of keyboard used has a significant effect on the number of errors committed, $F(2, 12) = 9.14$, $p < .05$. The following table summarizes the results of the analysis:

Source	*SS*	*df*	*MS*	
Between treatments	70	2	35	$F = 9.14$
Within treatments	46	12	3.83	
Total	116	14		

DEMONSTRATION 12.2

COMPUTING EFFECT SIZE FOR ANOVA

We compute eta squared (η^2), the percentage of variance explained, for the data that were analyzed in Demonstration 12.1. The data produced a between-treatments *SS* of 70 and a total *SS* of 116. Thus,

$$\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}} = \frac{70}{116} = 0.60 \text{ (or 60\%)}$$

PROBLEMS

1. Explain why the F-ratio is expected to be near 1.00 when the null hypothesis is true.

2. Describe the similarities between an F-ratio and a t statistic.

3. Several factors influence the size of the F-ratio. For each of the following, indicate whether it would influence the numerator or the denominator of the F-ratio, and indicate whether the size of the F-ratio would increase or decrease.
   a. Increase the differences between the sample means.
   b. Increase the size of the sample variances.

4. Why should you use ANOVA instead of several t tests to evaluate mean differences when an experiment consists of three or more treatment conditions?

5. Posttests are done after an ANOVA.
   a. What is the purpose of posttests?
   b. Explain why you do not need posttests if the analysis is comparing only two treatments.
   c. Explain why you do not need posttests if the decision from the ANOVA is to fail to reject the null hypothesis.

6. An independent-measures research study compares three treatment conditions with a sample of $n = 10$ in each condition. The sample means are $M_1 = 2$, $M_2 = 3$, and $M_3 = 7$.
   a. Compute *SS* for the set of 3 treatment means. (Use the three means as a set of $n = 3$ scores and compute *SS*.)
   b. Using the result from part a, compute $n(SS_{\text{means}})$. Note that this value is equal to SS_{between} (see Equation 12.6).
   c. Now, compute SS_{between} with the computational formula using the T values (Equation 12.7). You should obtain the same result as in part b.

7. The following data summarize the results from an independent-measures study comparing three treatment conditions.

I	II	III	
$n = 6$	$n = 6$	$n = 6$	
$M = 1$	$M = 5$	$M = 6$	$N = 18$
$T = 6$	$T = 30$	$T = 36$	$G = 72$
$SS = 30$	$SS = 35$	$SS = 40$	$\Sigma X^2 = 477$

a. Use an ANOVA with $\alpha = .05$ to determine whether there are any significant differences among the three treatment means.

b. Calculate η^2 to measure the effect size for this study.

c. Write a sentence demonstrating how a research report would present the results of the hypothesis test and the measure of effect size.

8. For the preceding problem you should find that there are significant differences among the three treatments. The primary reason for the significance is that the mean for treatment I is substantially smaller than the means for the other two treatments. To create the following data, we started with the values from problem 7 and added 3 points to each score in treatment I. Recall that adding a constant causes the mean to change but has no influence on the variability of the sample. In the resulting data, the mean differences are much smaller than those in problem 7.

I	II	III	
$n = 6$	$n = 6$	$n = 6$	
$M = 4$	$M = 5$	$M = 6$	$N = 18$
$T = 24$	$T = 30$	$T = 36$	$G = 90$
$SS = 30$	$SS = 35$	SS = 40	$\Sigma X^2 = 567$

a. Before you begin any calculations, predict how the change in the data should influence the outcome of the analysis. That is, how will the F-ratio and the value of η^2 for these data compare with the values obtained in problem 7?

b. Use an ANOVA with $\alpha = .05$ to determine whether there are any significant differences among the three treatment means. (Does your answer agree with your prediction in part a?)

c. Calculate η^2 to measure the effect size for this study. (Does your answer agree with your prediction in part a?)

9. The following data summarize the results from an independent-measures study comparing three treatment conditions.

I	II	III	
$n = 5$	$n = 5$	$n = 5$	
$M = 2$	$M = 5$	$M = 8$	$N = 15$
$T = 10$	$T = 25$	$T = 40$	$G = 75$
$SS = 16$	$SS = 20$	$SS = 24$	$\Sigma X^2 = 525$

a. Calculate the sample variance for each of the three samples.

b. Use an ANOVA with $\alpha = .05$ to determine whether there are any significant differences among the three treatment means.

10. For the preceding problem you should find that there are significant differences among the three treatments. One reason for the significance is that the sample variances are relatively small. To create the following data, we started with the values from problem 9 and increased the variability (the SS values) within each sample.

I	II	III	
$n = 5$	$n = 5$	$n = 5$	
$M = 2$	$M = 5$	$M = 8$	$N = 15$
$T = 10$	$T = 25$	$T = 40$	$G = 75$
$SS = 64$	$SS = 80$	$SS = 96$	$\Sigma X^2 = 705$

a. Calculate the sample variance for each of the three samples. Describe how these sample variances compare with those from problem 9.

b. Predict how the increase in sample variance should influence the outcome of the analysis. That is, how will the F-ratio for these data compare with the value obtained in problem 9?

c. Use an ANOVA with $\alpha = .05$ to determine whether there are any significant differences among the three treatment means. (Does your answer agree with your prediction in part b?)

11. Binge drinking on college campuses has been a hot topic in the popular media and in scholarly research. Flett, Goldstein, Wall, Hewitt, Wekerle, and Azzi (2008) report the results of a study relating perfectionism to binge drinking. In the study, students were classified into three groups based on the number of binge drinking episodes they experienced during the past month (0, 1, 2 or more). The students then completed a perfectionism questionnaire including one scale measuring parental criticism. One sample item is "I never felt that I could meet my parents' standards." Students rated their level of agreement with each item, and the total score was calculated for each student. The following results are similar to those obtained by the researchers.

Binge Drinking Episodes in Past Month			
0	1	2 or more	
8	10	13	$N = 15$
8	12	14	
10	8	12	$G = 165$
9	9	15	
10	11	16	$\Sigma X^2 = 1909$
$M = 9$	$M = 10$	$M = 14$	
$T = 45$	$T = 50$	$T = 70$	
$SS = 4$	$SS = 10$	$SS = 10$	

a. Use an ANOVA with $\alpha = .05$ to determine whether there are any significant differences among the three treatment means.

b. Calculate η^2 to measure the effect size for this study.

c. Write a sentence demonstrating how a research report would present the results of the hypothesis test and the measure of effect size.

12. A researcher reports an F-ratio with $df = 3, 36$ from an independent-measures research study.

a. How many treatment conditions were compared in the study?

b. What was the total number of participants in the study?

13. A research report from an independent-measures study states that there are significant differences between treatments, $F(2, 54) = 3.58$, $p < .05$.

a. How many treatment conditions were compared in the study?

b. What was the total number of participants in the study?

14. There is some evidence that high school students justify cheating in class on the basis of poor teacher skills or low levels of teacher caring (Murdock, Miller, and Kohlhardt, 2004). Students appear to rationalize their illicit behavior based on perceptions of how their teachers view cheating. Poor teachers are thought not to know or care whether students cheat, so cheating in their classes is okay. Good teachers, on the other hand, do care and are alert to cheating, so students tend not to cheat in their classes. Following are hypothetical data similar to the actual research results. The scores represent judgments of the acceptability of cheating for the students in each sample.

Poor Teacher	Average Teacher	Good Teacher	
$n = 6$	$n = 8$	$n = 10$	$N = 24$
$M = 6$	$M = 2$	$M = 2$	$G = 72$
$SS = 30$	$SS = 33$	$SS = 42$	$\Sigma X^2 = 393$

a. Use an ANOVA with $\alpha = .05$ to determine whether there are significant differences in student judgments depending on how they see their teachers.

b. Calculate η^2 to measure the effect size for this study.

c. Write a sentence demonstrating how a research report would present the results of the hypothesis test and the measure of effect size.

15. The following summary table presents the results from an ANOVA comparing three treatment conditions with $n = 8$ participants in each condition. Complete all missing values. (*Hint:* Start with the *df* column.)

Source	*SS*	*df*	*MS*	
Between treatments	____	____	15	$F =$ ____
Within treatments	____	____	____	
Total	93	____		

16. A pharmaceutical company has developed a drug that is expected to reduce hunger. To test the drug, two samples of rats are selected with $n = 20$ in each sample. The rats in the first sample receive the drug every day and those in the second sample are given a placebo. The dependent variable is the amount of food eaten by each rat over a 1-month period. An ANOVA is used to evaluate the difference between the two sample means and the results are reported in the following summary table. Fill in all missing values in the table. (*Hint:* Start with the *df* column.)

Source	*SS*	*df*	*MS*	
Between treatments	____	____	20	$F = 4.00$
Within treatments	____	____	____	
Total	____	____		

17. A developmental psychologist is examining the development of language skills from age 2 to age 4. Three different groups of children are obtained, one for each age, with $n = 16$ children in each group. Each child is given a language-skills assessment test. The resulting data were analyzed with an ANOVA to test for mean differences between age groups. The results of the ANOVA are presented in the following table. Fill in all missing values.

Source	*SS*	*df*	*MS*	
Between treatments	20	____	____	$F =$ ____
Within treatments	____	____	____	
Total	200	____		

18. The following data were obtained from an independent-measures research study comparing three treatment conditions. Use an ANOVA with α = .05 to determine whether there are any significant mean differences among the treatments.

Treatment			
I	II	III	
2	5	7	$N = 14$
5	2	3	$G = 42$
0	1	6	$\Sigma X^2 = 182$
1	2	4	
2			
2			
$T = 12$	$T = 10$	$T = 20$	
$SS = 14$	$SS = 9$	$SS = 10$	

19. The following values summarize the results from an independent-measures study comparing two treatment conditions.

a. Use an independent-measures *t* test with α = .05 to determine whether there is a significant mean difference between the two treatments.

b. Use an ANOVA with α = .05 to determine whether there is a significant mean difference between the two treatments.

Treatment		
I	II	
$n = 8$	$n = 4$	
$M = 4$	$M = 10$	$N = 12$
$T = 32$	$T = 40$	$G = 72$
$SS = 45$	$SS = 15$	$\Sigma X^2 = 588$

20. The following data represent the results from an independent-measures study comparing two treatment conditions.

a. Use an independent-measures *t* test with α = .05 to determine whether there is a significant mean difference between the two treatments.

b. Use an ANOVA with α = .05 to determine whether there is a significant mean difference between the two treatments.

Treatment		
I	II	
8	2	$N = 10$
7	3	$G = 50$
6	3	$\Sigma X^2 = 306$
5	5	
9	2	
$M = 7$	$M = 3$	
$T = 35$	$T = 15$	
$SS = 10$	$SS = 6$	

21. One possible explanation for why some birds migrate and others maintain year round residency in a single location is intelligence. Specifically, birds with small brains, relative to their body size, are simply not smart enough to find food during the winter and must migrate to warmer climates where food is easily available (Sol, Lefebvre, & Rodriguez-Teijeiro, 2005). Birds with bigger brains, on the other hand, are more creative and can find food even when the weather turns harsh. Following are hypothetical data similar to the actual research results. The numbers represent relative brain size for the individual birds in each sample.

Non-Migrating	Short-Distance Migrants	Long-Distance Migrants	
18	6	4	$N = 18$
13	11	9	$G = 180$
19	7	5	$\Sigma X^2 = 2150$
12	9	6	
16	8	5	
12	13	7	
$M = 15$	$M = 9$	$M = 6$	
$T = 90$	$T = 54$	$T = 36$	
$SS = 48$	$SS = 34$	$SS = 16$	

a. Use an ANOVA with α = .05 to determine whether there are any significant mean differences among the three groups of birds.

b. Compute η^2, the percentage of variance explained by the group differences, for these data.

c. Write a sentence demonstrating how a research report would present the results of the hypothesis test and the measure of effect size.

d. Use the Tukey HSD posttest to determine which groups are significantly different.

22. There is some research indicating that college students who use Facebook while studying tend to have lower grades than non-users (Kirschner & Karpinski, 2010). A representative study surveys students to determine the amount of Facebook use during the time they are studying or doing homework. Based on the amount of time spent on Facebook, students are classified into three groups and their grade point averages are recorded. The following data show the typical pattern of results.

Facebook Use While Studying		
Non-User	Rarely Use	Regularly Use
3.70	3.51	3.02
3.45	3.42	2.84
2.98	3.81	3.42
3.94	3.15	3.10
3.82	3.64	2.74
3.68	3.20	3.22
3.90	2.95	2.58
4.00	3.55	3.07
3.75	3.92	3.31
3.88	3.45	2.80

a. Use an ANOVA with $\alpha = .05$ to determine whether there are significant mean differences among the three groups.

b. Compute η^2 to measure the size of the effect.

c. Write a sentence demonstrating how the result from the hypothesis test and the measure of effect size would appear in a research report.

23. New research suggests that watching television, especially medical shows such as *Grey's Anatomy* and *House* can result in more concern about personal health (Ye, 2010). Surveys administered to college students measure television viewing habits and health concerns such as fear of developing the diseases and disorders seen on television. For the following data, students are classified into three categories based on their television viewing patterns and health concerns are measured on a 10-point scale with 0 indicating "none."

Television Viewing		
Little or None	Moderate	Substantial
4	5	5
2	7	7
5	3	6
1	4	6
3	8	8
7	6	9
4	2	6
4	7	4
8	3	6
2	5	8

a. Use an ANOVA with $\alpha = .05$ to determine whether there are significant mean differences among the three groups.

b. Compute η^2 to measure the size of the effect.

c. Use Tukey's HSD test with $\alpha = .05$ to determine which groups are significantly different.