

C H A P T E R

13

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.

- Independent-measures analysis of variance (Chapter 12)
- Repeated-measures designs (Chapter 11)
- Individual differences

Repeated-Measures Analysis of Variance

Preview

- 13.1 Overview of Repeated-Measures Designs
- 13.2 The Repeated-Measures ANOVA
- 13.3 Hypothesis Testing and Effect Size with the Repeated-Measures ANOVA
- 13.4 Advantages and Disadvantages of the Repeated-Measures Design
- 13.5 Repeated-Measures ANOVA and Repeated-Measures t test

Summary

Focus on Problem Solving

Demonstrations 13.1 and 13.2

Problems

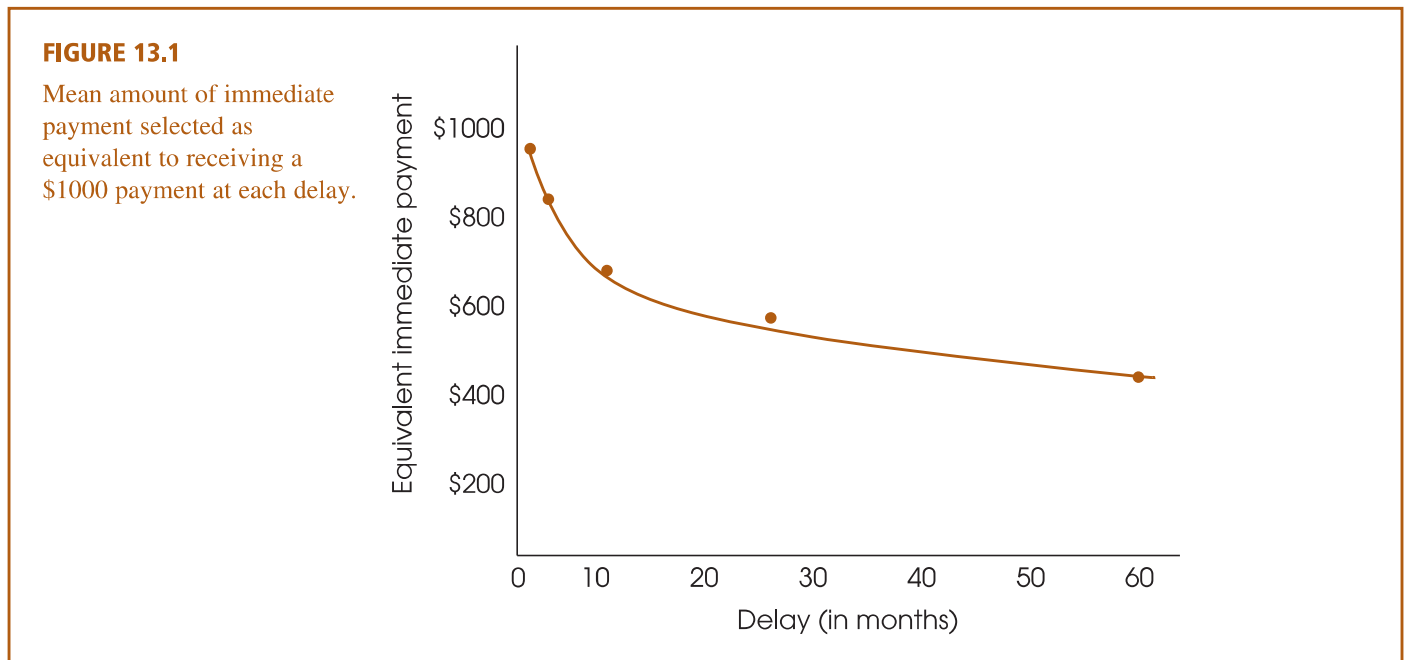
Preview

Suppose that you were offered a choice between receiving \$1000 in 5 years or a smaller amount today. How much would you be willing to take today to avoid waiting 5 years to get the full \$1000 payment?

The general result for this kind of decision is that the longer the \$1000 payment is delayed, the smaller the amount that people will accept today. For example, you may be willing to take \$300 today rather than waiting 5 years for the \$1000. However, you might be willing to settle for \$100 today if you have to wait 10 years for the \$1000. This phenomenon is known as delayed discounting because people discount the value of a future reward depending on how long it is delayed (Green,

Fry, & Myerson, 1994). In a typical study examining delayed discounting, people are asked to place a value on a future reward for several different delay periods. For example, how much would you accept today instead of waiting for a future reward of \$1000 if you had to wait 1 month before receiving the payment? How about waiting 6 months? 12 months? 24 months? 60 months?

Typical results for a sample of college students are shown in Figure 13.1. Note that the average value declines regularly as the delay period increases. The statistical question is whether the mean differences from one delay period to another are significant.



The Problem: You should recognize that evaluating mean differences for more than two sample means is a job for analysis of variance (ANOVA). However, the discounting study is a repeated-measures design with five scores for each individual, and the ANOVA introduced in Chapter 12 is intended for independent-measures studies. Once again, a new hypothesis test is needed.

The Solution: In this chapter we introduce the *repeated-measures ANOVA*. As the name implies, this new procedure is used to evaluate the differences between two or more sample means obtained from a repeated-measures research study. As you will see, many of the notational symbols and computations are the same as those used for the independent-measures ANOVA. In fact, your best preparation for this chapter is a good understanding of the basic ANOVA procedure presented in Chapter 12.

13.1 OVERVIEW OF REPEATED-MEASURES DESIGNS

In the preceding chapter, we introduced ANOVA as a hypothesis-testing procedure for evaluating differences among two or more sample means. The specific advantage of ANOVA, especially in contrast to t tests, is that ANOVA can be used to evaluate the significance of mean differences in situations in which there are more than two sample means being compared. However, the presentation of ANOVA in Chapter 12 was limited to single-factor, independent-measures research designs. Recall that *single factor* indicates that the research study involves only one independent variable (or only one quasi-independent variable), and the term *independent-measures* indicates that the study uses a separate sample for each of the different treatment conditions being compared.

In this chapter, we extend the ANOVA procedure to single-factor, repeated-measures designs. The defining characteristic of a repeated-measures design is that one group of individuals participates in all of the different treatment conditions. The repeated-measures ANOVA is used to evaluate mean differences in two general research situations:

1. An experimental study in which the researcher manipulates an independent variable to create two or more treatment conditions, with the same group of individuals tested in all of the conditions.
2. A nonexperimental study in which the same group of individuals is simply observed at two or more different times.

Examples of these two research situations are presented in Table 13.1. Table 13.1(a) shows data from a study in which the researcher changes the type of distraction to

TABLE 13.1

Two sets of data representing typical examples of single-factor, repeated-measures research designs.

(a) Data from an experimental study evaluating the effects of different types of distraction on the performance of a visual detection task.

Participant	Visual Detection Scores		
	No Distraction	Visual Distraction	Auditory Distraction
A	47	22	41
B	57	31	52
C	38	18	40
D	45	32	43

(b) Data from a nonexperimental design evaluating the effectiveness of a clinical therapy for treating depression.

Participant	Depression Scores		
	Before Therapy	After Therapy	6-Month Follow-Up
A	71	53	55
B	62	45	44
C	82	56	61
D	77	50	46
E	81	54	55

create three treatment conditions. One group of participants is then tested in all three conditions. In this study, the factor being examined is the type of distraction.

Table 13.1(b) shows a study in which a researcher observes depression scores for the same group of individuals at three different times. In this study, the time of measurement is the factor being examined. Another common example of this type of design is found in developmental psychology when the participants' age is the factor being studied. For example, a researcher could study the development of vocabulary skill by measuring vocabulary for a sample of 3-year-old children, then measuring the same children again at ages 4 and 5.

13.2 THE REPEATED-MEASURES ANOVA

HYPOTHESES FOR THE REPEATED-MEASURES ANOVA

The hypotheses for the repeated-measures ANOVA are exactly the same as those for the independent-measures ANOVA presented in Chapter 12. Specifically, the null hypothesis states that, for the general population, there are no mean differences among the treatment conditions being compared. In symbols,

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

The null hypothesis states that, on average, all of the treatments have exactly the same effect. According to the null hypothesis, any differences that may exist among the sample means are not caused by systematic treatment effects but rather are the result of random and unsystematic factors.

The alternative hypothesis states that there are mean differences among the treatment conditions. Rather than specifying exactly which treatments are different, we use a generic version of H_1 , which simply states that differences exist:

$$H_1: \text{At least one treatment mean } (\mu) \text{ is different from another.}$$

Notice that the alternative says that, on average, the treatments do have different effects. Thus, the treatment conditions may be responsible for causing mean differences among the samples. As always, the goal of the ANOVA is to use the sample data to determine which of the two hypotheses is more likely to be correct.

THE *F*-RATIO FOR REPEATED-MEASURES ANOVA

The F -ratio for the repeated-measures ANOVA has the same structure that was used for the independent-measures ANOVA in Chapter 12. In each case, the F -ratio compares the actual mean differences between treatments with the amount of difference that would be expected if there were no treatment effect. The numerator of the F -ratio measures the actual mean differences between treatments. The denominator measures how big the differences should be if there is no treatment effect. As always, the F -ratio uses variance to measure the size of the differences. Thus, the F -ratio for both ANOVAs has the general structure

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

A large value for the F -ratio indicates that the differences between treatments are greater than would be expected without any treatment effect. If the F -ratio is larger than

the critical value in the F distribution table, then we can conclude that the differences between treatments are *significantly* larger than would be caused by chance.

Individual Differences in the F -ratio Although the structure of the F -ratio is the same for independent-measures and repeated-measures designs, there is a fundamental difference between the two designs that produces a corresponding difference in the two F -ratios. Specifically, individual differences are a part of one ratio but are eliminated from the other.

You should recall that the term *individual differences* refers to participant characteristics such as age, personality, and gender that vary from one person to another and may influence the measurements that you obtain for each person. Suppose, for example, that you are measuring reaction time. The first participant in your study is a 19-year-old female with an IQ of 136 who is on the college varsity volleyball team. The next participant is a 42-year-old male with an IQ of 111 who returned to college after losing his job and comes to the research study with a head cold. Would you expect to obtain the same reaction time score for these two individuals?

Individual differences are a part of the variance in the numerator and in the denominator of the F -ratio for the independent-measures ANOVA. However, individual differences are eliminated or removed from the variances in the F -ratio for the repeated measures ANOVA. The idea of removing individual differences was first presented in Chapter 11 when we introduced the repeated-measures design (p. 367), but we review it briefly now.

In a repeated-measures study, exactly the same individuals participate in all of the treatment conditions. Therefore, if there are any mean differences between treatments, they cannot be explained by individual differences. Thus, individual differences are automatically eliminated from the numerator of the repeated-measures F -ratio.

A repeated-measures design also allows you to remove individual differences from the variance in the denominator of the F -ratio. Because the same individuals are measured in every treatment condition, it is possible to measure the size of the individual differences. In Table 13.1(a), for example, participant A has scores that are consistently 10 points lower than the scores for participant B. Because the individual differences are systematic and predictable, they can be measured and separated from the random, unsystematic differences in the denominator of the F -ratio.

Thus, individual differences are automatically eliminated from the numerator of the repeated-measures F -ratio. In addition, they can be measured and removed from the denominator. As a result, the structure of the final F -ratio is as follows:

$$F = \frac{\text{variance/differences between treatments} \\ \text{(without individual differences)}}{\text{variance/differences with no treatment effect} \\ \text{(with individual differences removed)}}$$

The process of removing individual differences is an important part of the procedure for a repeated-measures ANOVA.

**THE LOGIC OF THE
REPEATED-MEASURES
ANOVA**

The general purpose of the repeated-measures ANOVA is to determine whether the differences that are found between treatment conditions are significantly greater than would be expected if there is no treatment effect. In the numerator of the F -ratio, the *between-treatments variance* measures the actual mean differences between the treatment conditions. The variance in the denominator is intended to measure how much

difference is reasonable to expect if there are no systematic treatment effects and no systematic individual differences. In other words, the denominator measures variability caused entirely by random and unsystematic factors. For this reason, the variance in the denominator is called the *error variance*. In this section we examine the elements that make up the two variances in the repeated-measures F -ratio.

The numerator of the F -ratio: between-treatments variance Logically, any differences that are found between treatments can be explained by only two factors:

1. **Systematic Differences Caused by the Treatments.** It is possible that the different treatment conditions really do have different effects and, therefore, cause the individuals' scores in one condition to be higher (or lower) than in another. Remember that the purpose for the research study is to determine whether a *treatment effect* exists.
2. **Random, Unsystematic Differences.** Even if there is no treatment effect, it is possible for the scores in one treatment condition to be different from the scores in another. For example, suppose that I measure your IQ score on a Monday morning. A week later I come back and measure your IQ again under exactly the same conditions. Will you get exactly the same IQ score both times? In fact, minor differences between the two measurement situations would probably cause you to end up with two different scores. For example, for one of the IQ tests you might be more tired, or hungry, or worried, or distracted than you were on the other test. These differences can cause your scores to vary. The same thing can happen in a repeated-measures research study. The same individuals are measured at two or more different times and, even though there may be no difference between the two treatment conditions, you can still end up with different scores. However, these differences are random and unsystematic and are classified as error variance.

Thus, it is possible that any differences (or variance) found between treatments could be caused by treatment effects, and it is possible that the differences could simply be the result of chance. On the other hand, it is *impossible* that the differences between treatments are caused by individual differences. Because the repeated-measures design uses exactly the same individuals in every treatment condition, individual differences are *automatically eliminated* from the variance between treatments in the numerator of the F -ratio.

The denominator of the F -ratio: error variance The goal of the ANOVA is to determine whether the differences that are observed in the data are greater than would be expected without any systematic treatment effects. To accomplish this goal, the denominator of the F -ratio is intended to measure how much difference (or variance) is reasonable to expect from random and unsystematic factors. This means that we must measure the variance that exists when there are no treatment effects or any other systematic differences.

We begin exactly as we did with the independent-measures F -ratio; specifically, we calculate the variance that exists within treatments. Recall from Chapter 12 that within each treatment all of the individuals are treated in exactly the same way. Therefore, any differences that exist within treatments cannot be caused by treatment effects.

In a repeated-measures design, however, it is also possible that individual differences can cause systematic differences between the scores within treatments. For example, one individual may score consistently higher than another. To eliminate the

individual differences from the denominator of the F -ratio, we measure the individual differences and then subtract them from the rest of the variability. The variance that remains is a measure of pure *error* without any systematic differences that can be explained by treatment effects or by individual differences.

In summary, the F -ratio for a repeated-measures ANOVA has the same basic structure as the F -ratio for independent measures (Chapter 12) except that it includes no variability caused by individual differences. The individual differences are automatically eliminated from the variance between treatments (numerator) because the repeated-measures design uses the same individuals in all treatments. In the denominator, the individual differences are subtracted during the analysis. As a result, the repeated-measures F -ratio has the following structure:

$$F = \frac{\text{between-treatments variance}}{\text{error variance}} \\ = \frac{\text{treatment effects} + \text{random, unsystematic differences}}{\text{random, unsystematic differences}} \quad (13.1)$$

Note that this F -ratio is structured so that there are no individual differences contributing to either the numerator or the denominator. When there is no treatment effect, the F -ratio is balanced because the numerator and denominator are both measuring exactly the same variance. In this case, the F -ratio should have a value near 1.00. When research results produce an F -ratio near 1.00, we conclude that there is no evidence of a treatment effect and we fail to reject the null hypothesis. On the other hand, when a treatment effect does exist, it contributes only to the numerator and should produce a large value for the F -ratio. Thus, a large value for F indicates that there is a real treatment effect and, therefore, we should reject the null hypothesis.

LEARNING CHECK

1. Explain why individual differences do not contribute to the between-treatments variability in a repeated-measures study.
2. What sources of variability contribute to the within-treatment variability for a repeated-measures study?
3. Describe the structure of the F -ratio for the repeated-measures ANOVA.

ANSWERS

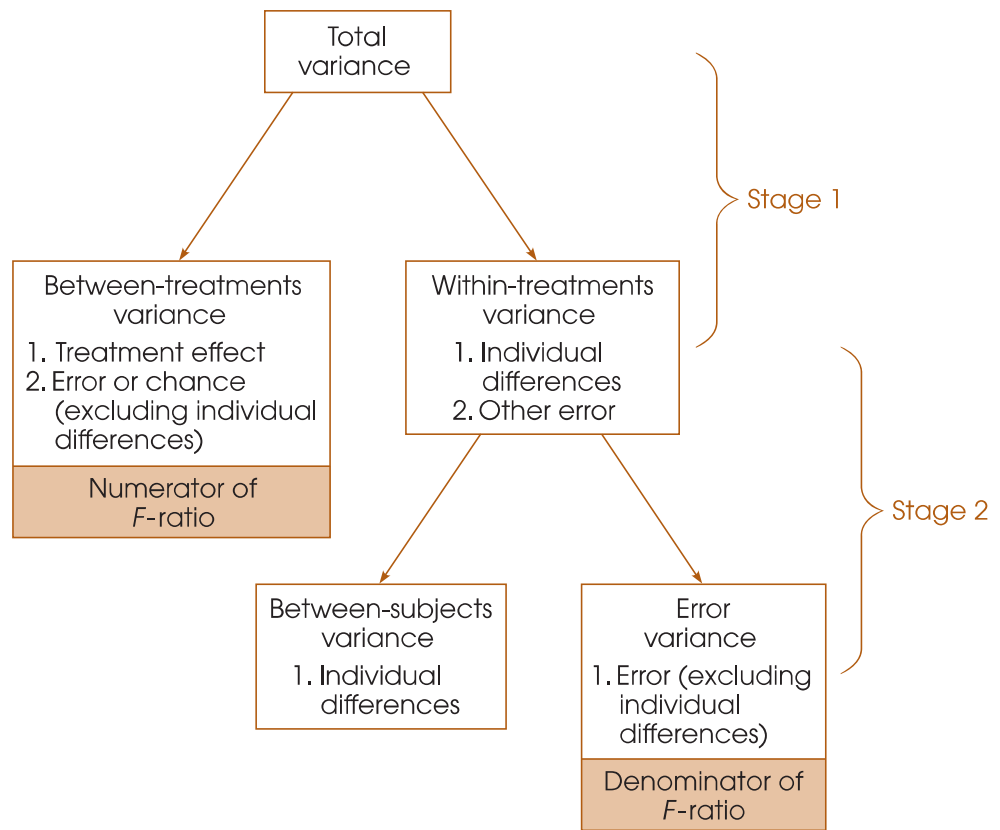
1. Because the individuals in one treatment are exactly the same as the individuals in every other treatment, there are no individual differences from one treatment to another.
2. Variability (differences) within treatments is caused by individual differences and random, unsystematic differences.
3. The numerator of the F -ratio measures between-treatments variability, which consists of treatment effects and random, unsystematic differences. The denominator measures variability that is exclusively caused by random, unsystematic differences.

13.3 HYPOTHESIS TESTING AND EFFECT SIZE WITH THE REPEATED-MEASURES ANOVA

The overall structure of the repeated-measures ANOVA is shown in Figure 13.2. Note that the ANOVA can be viewed as a two-stage process. In the first stage, the total variance is partitioned into two components: *between-treatments variance* and *within-treatments*

FIGURE 13.2

The partitioning of variance for a repeated-measures experiment.



variance. This stage is identical to the analysis that we conducted for an independent-measures design in Chapter 12.

The second stage of the analysis is intended to remove the individual differences from the denominator of the F -ratio. In the second stage, we begin with the variance within treatments and then measure and subtract out the *between-subject variance*, which measures the size of the individual differences. The remaining variance, often called the *residual variance*, or *error variance*, provides a measure of how much variance is reasonable to expect after the treatment effects and individual differences have been removed. The second stage of the analysis is what differentiates the repeated-measures ANOVA from the independent-measures ANOVA. Specifically, the repeated-measures design requires that the individual differences be removed.

DEFINITION

In a repeated-measures ANOVA, the denominator of the F -ratio is called the **residual variance**, or the **error variance**, and measures how much variance is expected if there are no systematic treatment effects and no individual differences contributing to the variability of the scores.

NOTATION FOR THE REPEATED-MEASURES ANOVA

We use the data in Table 13.2 to introduce the notation for the repeated-measures ANOVA. The data represent the results of a study comparing different viewing distances for a 42-inch high-definition television. Four viewing distances were evaluated,

TABLE 13.2

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

Note: For comparison, the scores are identical to the values in Example 12.1 on page 405.

Person	Viewing Distance				Person Totals	
	9 Feet	12 Feet	15 Feet	18 Feet		
A	3	4	7	6	$P = 20$	$n = 5$
B	0	3	6	3	$P = 12$	$k = 4$
C	2	1	5	4	$P = 12$	$N = 20$
D	0	1	4	3	$P = 8$	$G = 60$
E	0	1	3	4	$P = 8$	$\Sigma X^2 = 262$
	$T = 5$	$T = 10$	$T = 25$	$T = 20$		
	$SS = 8$	$SS = 8$	$SS = 10$	$SS = 6$		

9 feet, 12 feet, 15 feet, and 18 feet. Each participant was free to move back and forth among the four distances while watching a 30-minute video on the television. The only restriction was that each person had to spend at least 2 minutes watching from each of the four distances. At the end of the video, each participant rated the all of the viewing distances on a scale from 1 (Very Bad, definitely need to move closer or farther away) to 7 (excellent, perfect viewing distance). You may notice that this research study and the numerical values in the table are identical to those used to demonstrate the independent-measures ANOVA in the previous chapter (Example 12.1, page 405). In this case, however, the data represent a repeated-measures study in which the same group of $n = 5$ individuals is tested in all four treatment conditions.

You should recognize that most of the notation in Table 13.2 is identical to the notation used in an independent-measures analysis (Chapter 12). For example, there are $n = 5$ participants who are tested in $k = 4$ treatment conditions, producing a total of $N = 20$ scores that add up to a grand total of $G = 60$. Note, however, that $N = 20$ now refers to the total number of scores in the study, not the number of participants.

The repeated-measures ANOVA introduces only one new notational symbol. The letter P is used to represent the total of all of the scores for each individual in the study. You can think of the P values as “Person totals” or “Participant totals.” In Table 13.2, for example, participant A had scores of 3, 4, 6, and 7 for a total of $P = 20$. The P values are used to define and measure the magnitude of the individual differences in the second stage of the analysis.

EXAMPLE 13.1

We use the data in Table 13.2 to demonstrate the repeated-measures ANOVA. Again, the goal of the test is to determine whether there are any significant differences among the four distances being compared. Specifically, are any of the mean differences in the data greater than would be expected if there are no systematic differences among the four viewing distances?

STAGE 1 OF THE REPEATED-MEASURES ANOVA

The first stage of the repeated-measures analysis is identical to the independent-measures ANOVA that was presented in Chapter 12. Specially, the SS and df for the total set of scores are analyzed into within-treatments and between-treatments components.

Because the numerical values in Table 13.2 are the same as the values used in Example 12.1 (p. 405), the computations for the first stage of the repeated-measures analysis are identical to those in Example 12.1. Rather than repeating the same

arithmetic, the results of the first stage of the repeated-measures analysis can be summarized as follows:

Total:

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

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

Within treatments:

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

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

Between treatments: For this example we use the computational formula for $SS_{\text{between treatments}}$.

$$SS_{\text{between treatments}} = \sum \frac{T^2}{n} - \frac{G^2}{N} = \frac{5^2}{5} + \frac{10^2}{5} + \frac{25^2}{5} + \frac{20^2}{5} - \frac{60^2}{20} = 50$$

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

For more details on the formulas and calculations, see Example 12.1, pages 405–407.

This completes the first stage of the repeated-measures ANOVA. Note that the two components, between and within, add up to the total for the SS values and for the df values. Also note that the between-treatments SS and df values provide a measure of the mean differences between treatments and are used to compute the variance in the numerator of the final F -ratio.

STAGE 2 OF THE REPEATED-MEASURES ANOVA

The second stage of the analysis involves removing the individual differences from the denominator of the F -ratio. Because the same individuals are used in every treatment, it is possible to measure the size of the individual differences. For the data in Table 13.2, for example, participant A tends to have the highest scores and participants D and E tend to have the lowest scores. These individual differences are reflected in the P values, or person totals, in the right-hand column. We use these P values to create a computational formula for $SS_{\text{between subjects}}$ in much the same way that we used the treatment totals, the T values, in the computational formula for $SS_{\text{between treatments}}$. Specifically, the formula for the between-subjects SS is

$$SS_{\text{between subjects}} = \sum \frac{P^2}{k} - \frac{G^2}{N} \quad (13.2)$$

Notice that the formula for the between-subjects SS has exactly the same structure as the computational formula for the between-treatments SS (see the calculation above). In this case we use the person totals (P values) instead of the treatment totals (T values). Each P value is squared and divided by the number of scores that were added to obtain the total. In this case, each person has k scores, one for each treatment. Box 13.1 presents another demonstration of the similarity

of the formulas for SS between subjects and SS between treatments. For the data in Table 13.2,

$$\begin{aligned} SS_{\text{between subjects}} &= \frac{20^2}{4} + \frac{12^2}{4} + \frac{12^2}{4} + \frac{8^2}{4} + \frac{8^2}{4} - \frac{60^2}{20} \\ &= 100 + 36 + 36 + 16 + 16 - 180 \\ &= 24 \end{aligned}$$

The value of $SS_{\text{between subjects}}$ provides a measure of the size of the individual differences—that is, the differences between subjects. In the second stage of the analysis, we simply subtract the individual differences to obtain the measure of error that forms the denominator of the F -ratio. Thus, the final step in the analysis of SS is

$$SS_{\text{error}} = SS_{\text{within treatments}} - SS_{\text{between subjects}} \quad (13.3)$$

We have already computed $SS_{\text{within treatments}} = 32$ and $SS_{\text{between subjects}} = 24$, therefore

$$SS_{\text{error}} = 32 - 24 = 8$$

The analysis of degrees of freedom follows exactly the same pattern that was used to analyze SS . Remember that we are using the P values to measure the magnitude of the individual differences. The number of P values corresponds to the number of subjects, n , so the corresponding df is

$$df_{\text{between subjects}} = n - 1 \quad (13.4)$$

For the data in Table 13.2, there are $n = 5$ subjects and

$$df_{\text{between subjects}} = 5 - 1 = 4$$

BOX 13.1

$SS_{\text{between subjects}}$ AND $SS_{\text{between treatments}}$

The data for a repeated-measures study are normally presented in a matrix, with the treatment conditions determining the columns and the participants defining the rows. The data in Table 13.2 demonstrate this normal presentation. The calculation of $SS_{\text{between treatments}}$ provides a measure of the differences between treatment conditions—that is, a measure of the mean differences between the *columns* in the data matrix. For the data in Table 13.2, the column totals are 5, 10, 20, and 25. These values are variable, and $SS_{\text{between treatments}}$ measures the amount of variability.

The following table reproduces the data from Table 13.2, but now we have turned the data matrix on its side so that the participants define the columns and the treatment conditions define the rows.

In this new format, the differences between the columns represent the between-subjects variability. The

column totals are now P values (instead of T values) and the number of scores in each column is now identified by k (instead of n). With these changes in notation, the formula for $SS_{\text{between subjects}}$ has exactly the same structure as the formula for $SS_{\text{between treatments}}$. If you examine the two equations, the similarity should be clear.

	Participant					
	A	B	C	D	E	
9 feet	3	0	2	0	0	$T = 5$
12 feet	4	3	1	1	1	$T = 10$
15 feet	7	6	5	4	3	$T = 25$
18 feet	6	3	4	3	4	$T = 20$
	$P = 20$	$P = 12$	$P = 12$	$P = 8$	$P = 8$	

Next, we subtract the individual differences from the within-subjects component to obtain a measure of error. In terms of degrees of freedom,

$$df_{\text{error}} = df_{\text{within treatments}} - df_{\text{between subjects}} \quad (13.5)$$

For the data in Table 13.2,

$$df_{\text{error}} = 16 - 4 = 12$$

An algebraically equivalent formula for df_{error} uses only the number of treatment conditions (k) and the number of participants (n):

$$df_{\text{error}} = (k - 1)(n - 1) \quad (13.6)$$

The usefulness of equation 13.6 is discussed in Box 13.2.

Remember: The purpose for the second stage of the analysis is to measure the individual differences and then remove the individual differences from the denominator of the F -ratio. This goal is accomplished by computing SS and df between subjects (the individual differences) and then subtracting these values from the within-treatments values. The result is a measure of variability resulting from error with the individual differences removed. This error variance (SS and df) is used in the denominator of the F -ratio.

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

The final calculation in the analysis is the F -ratio, which is a ratio of two variances. Each variance is called a *mean square*, or MS , and is obtained by dividing the appropriate SS by its corresponding df value. The MS in the numerator of the F -ratio measures the size of the differences between treatments and is calculated as

$$MS_{\text{between treatments}} = \frac{SS_{\text{between treatments}}}{df_{\text{between treatments}}} \quad (13.7)$$

For the data in Table 13.2,

$$MS_{\text{between treatments}} = \frac{50}{3} = 16.67$$

BOX 13.2

USING THE ALTERNATIVE FORMULA FOR df_{error}

The statistics presented in a research report not only describe the significance of the results but typically provide enough information to reconstruct the research design. The alternative formula for df_{error} is particularly useful for this purpose. Suppose, for example, that a research report for a repeated-measures study includes an F -ratio with $df = 2, 10$. How many treatment conditions were compared in the study, and how many individuals participated?

To answer these questions, begin with the first df value, which is $df_{\text{between treatments}} = 2 = k - 1$. From this value, it is clear that $k = 3$ treatments. Next, use the

second df value, which is $df_{\text{error}} = 10$. Using this value and the fact that $k - 1 = 2$, use equation 13.6 to find the number of participants.

$$df_{\text{error}} = 10 = (k - 1)(n - 1) = 2(n - 1)$$

If $2(n - 1) = 10$, then $n - 1$ must equal 5. Therefore, $n = 6$.

Therefore, we conclude that a repeated-measures study producing an F -ratio with $df = 2, 10$ must have compared 3 treatment conditions using a sample of 6 participants.

The denominator of the F -ratio measures how much difference is reasonable to expect if there are no systematic treatment effects and the individual differences have been removed. This is the error variance, or the residual variance, obtained in stage 2 of the analysis.

$$MS_{\text{error}} = \frac{SS_{\text{error}}}{df_{\text{error}}} \quad (13.8)$$

For the data in Table 13.2,

$$MS_{\text{error}} = \frac{8}{12} = 0.67$$

Finally, the F -ratio is computed as

$$F = \frac{MS_{\text{between treatments}}}{MS_{\text{error}}} \quad (13.9)$$

For the data in Table 13.2,

$$F = \frac{16.67}{0.67} = 24.88$$

Once again, notice that the repeated-measures ANOVA uses MS_{error} in the denominator of the F -ratio. This MS value is obtained in the second stage of the analysis, after the individual differences have been removed. As a result, individual differences are completely eliminated from the repeated-measures F -ratio, so that the general structure is

$$F = \frac{\text{treatment effects} + \text{unsystematic differences (without individual diffs)}}{\text{unsystematic differences (without individual diffs)}}$$

For the data we have been examining, the F -ratio is $F = 24.88$, indicating that the differences between treatments (numerator) are almost 25 times bigger than you would expect without any treatment effects (denominator). A ratio this large provides clear evidence that there is a real treatment effect. To verify this conclusion you must consult the F distribution table to determine the appropriate critical value for the test. The degrees of freedom for the F -ratio are determined by the two variances that form the numerator and the denominator. For a repeated-measures ANOVA, the df values for the F -ratio are reported as

$$df = df_{\text{between treatments}}, df_{\text{error}}$$

For the example we are considering, the F -ratio has $df = 2, 12$ (“degrees of freedom equal two and twelve”). Using the F distribution table (p. 705) with $\alpha = .05$, the critical value is $F = 3.88$, and with $\alpha = .01$ the critical value is $F = 6.93$. Our obtained F -ratio, $F = 24.88$, is well beyond either of the critical values, so we can conclude that the differences between treatments are *significantly* greater than expected by chance using either $\alpha = .05$ or $\alpha = .01$.

The summary table for the repeated-measures ANOVA from Example 13.1 is presented in Table 13.3. Although these tables are no longer commonly used in research reports, they provide a concise format for displaying all of the elements of the analysis.

MEASURING EFFECT SIZE FOR THE REPEATED-MEASURES ANOVA

The most common method for measuring effect size with ANOVA is to compute the percentage of variance that is explained by the treatment differences. In the context of ANOVA, the percentage of variance is commonly identified as η^2 (eta squared). In Chapter 12, for the independent-measures analysis, we computed η^2 as

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

The intent is to measure how much of the total variability is explained by the differences between treatments. With a repeated-measures design, however, there is another component that can explain some of the variability in the data. Specifically, part of the variability is caused by differences between individuals. In Table 13.2, for example, person A consistently scored higher than person B. This consistent difference explains some of the variability in the data. When computing the size of the treatment effect, it is customary to remove any variability that can be explained by other factors, and then compute the percentage of the remaining variability that can be explained by the treatment effects. Thus, for a repeated-measures ANOVA, the variability from the individual differences is removed before computing η^2 . As a result, η^2 is computed as

$$\eta^2 = \frac{SS_{\text{between treatments}}}{SS_{\text{total}} - SS_{\text{between subjects}}} \tag{13.10}$$

Because Equation 13.10 computes a percentage that is not based on the total variability of the scores (one part, $SS_{\text{between subjects}}$, is removed), the result is often called a *partial eta squared*.

The general goal of Equation 13.10 is to calculate a percentage of the variability that has not already been explained by other factors. Thus, the denominator of Equation 13.10 is limited to variability from the treatment differences and variability that is exclusively from random, unsystematic factors. With this in mind, an equivalent version of the η^2 formula is

$$\eta^2 = \frac{SS_{\text{between treatments}}}{SS_{\text{between treatments}} + SS_{\text{error}}} \tag{13.11}$$

TABLE 13.3

A summary table for the repeated-measures ANOVA for the data from Example 13.1.

Source	SS	df	MS	F
Between treatments	50	3	16.67	$F(3,12) = 24.88$
Within treatments	32	16		
Between subjects	24	4		
Error	8	12	0.67	
Total	82	19		

In this new version of the eta-squared formula, the denominator consists of the variability that is explained by the treatment differences plus the other *unexplained* variability. Using either formula, the data from Example 13.1 produce

$$\eta^2 = \frac{50}{58} = 0.862 \text{ (or 86.2\%)}$$

This result means that 86.2% of the variability in the data (except for the individual differences) is accounted for by the differences between treatments.



IN THE LITERATURE

REPORTING THE RESULTS OF A REPEATED-MEASURES ANOVA

As described in Chapter 12 (p. 409), the format for reporting ANOVA results in journal articles consists of

1. A summary of descriptive statistics (at least treatment means and standard deviations, and tables or graphs as needed)
2. A concise statement of the outcome of the ANOVA

For the study in Example 13.1, the report could state:

The means and variances for the four television viewing distances are shown in Table 1. A repeated-measures analysis of variance indicated significant mean differences in the participants' ratings of the four distances, $F(3, 12) = 24.88$, $p < .01$, $\eta^2 = 0.862$.

TABLE 1

Ratings of satisfaction with different television-viewing distances

	9 Feet	12 Feet	15 Feet	18 Feet
<i>M</i>	1.00	2.00	5.00	4.00
<i>SD</i>	1.41	1.41	1.58	1.22

POST HOC TESTS WITH REPEATED-MEASURES ANOVA

Recall that ANOVA provides an overall test of significance for the mean differences between treatments. When the null hypothesis is rejected, it indicates only that there is a difference between at least two of the treatment means. If $k = 2$, it is obvious which two treatments are different. However, when k is greater than 2, the situation becomes more complex. To determine exactly where significant differences exist, the researcher must follow the ANOVA with post hoc tests. In Chapter 12, we used Tukey's HSD and the Scheffé test to make these multiple comparisons among treatment means. These two procedures attempt to control the overall alpha level by making adjustments for the number of potential comparisons.

For a repeated-measures ANOVA, Tukey's HSD and the Scheffé test can be used in the exact same manner as was done for the independent-measures ANOVA, *provided* that you substitute MS_{error} in place of $MS_{\text{within treatments}}$ in the formulas and use df_{error} in place of $df_{\text{within treatments}}$ when locating the critical value in a statistical table. Note that statisticians are not in complete agreement about the appropriate error term in post hoc tests for repeated-measures designs (for a discussion, see Keppel, 1973, or Keppel & Zedeck, 1989).

ASSUMPTIONS OF THE REPEATED-MEASURES ANOVA

The basic assumptions for the repeated-measures ANOVA are identical to those required for the independent-measures ANOVA.

1. The observations within each treatment condition must be independent (see p. 254).
2. The population distribution within each treatment must be normal. (As before, the assumption of normality is important only with small samples.)
3. The variances of the population distributions for each treatment should be equivalent.

For the repeated-measures ANOVA, there is an additional assumption, called homogeneity of covariance. Basically, it refers to the requirement that the relative standing of each subject be maintained in each treatment condition. This assumption is violated if the effect of the treatment is not consistent for all of the subjects or if order effects exist for some, but not other, subjects. This issue is very complex and is beyond the scope of this book. However, methods do exist for dealing with violations of this assumption (for a discussion, see Keppel, 1973).

If there is reason to suspect that one of the assumptions for the repeated-measures ANOVA has been violated, an alternative analysis known as the Friedman test can be used. The Friedman test is presented in Appendix E. It requires that the original scores be transformed into ranks before evaluating the differences between treatment conditions.

LEARNING CHECK

1. Explain how SS_{error} is computed in the repeated-measures ANOVA.
2. A repeated-measures study is used to evaluate the mean differences among three treatment conditions using a sample of $n = 8$ participants. What are the df values for the F -ratio?
3. For the following data, compute $SS_{\text{between treatments}}$ and $SS_{\text{between subjects}}$.

Subject	Treatment					
	1	2	3	4		
A	2	2	2	2	$G = 32$ $\Sigma X^2 = 96$	
B	4	0	0	4		
C	2	0	2	0		
D	4	2	2	4		
		$T = 12$	$T = 4$	$T = 6$	$T = 10$	
		$SS = 4$	$SS = 4$	$SS = 3$	$SS = 11$	

4. A research report includes a repeated-measures F -ratio with $df = 3, 24$. How many treatment conditions were compared, and how many individuals participated in the study? (See Box 13.2.)

ANSWERS

1. $SS_{\text{error}} = SS_{\text{within}} - SS_{\text{between subjects}}$ Variability from individual differences is subtracted from the within-treatments variability.
2. $df = 2, 14$
3. $SS_{\text{between treatments}} = 10, SS_{\text{between subjects}} = 8$
4. There were 4 treatment conditions ($k - 1 = 3$) and 9 participants ($n - 1 = 8$).

13.4 ADVANTAGES AND DISADVANTAGES OF THE REPEATED-MEASURES DESIGN

When we first encountered the repeated-measure design (Chapter 11), we noted that this type of research study has certain advantages and disadvantages (pp. 366–369). On the bright side, a repeated-measures study may be desirable if the supply of participants is limited. A repeated-measures study is economical in that the research requires relatively few participants. Also, a repeated-measures design eliminates or minimizes most of the problems associated with individual differences. However, disadvantages also exist. These take the form of order effects, such as fatigue, that can make the interpretation of the data difficult.

Now that we have introduced the repeated-measures ANOVA, we can examine one of the primary advantages of this design—namely, the elimination of variability caused by individual differences. Consider the structure of the F -ratio for both the independent-and the repeated-measures designs.

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

In each case, the goal of the analysis is to determine whether the data provide evidence for a treatment effect. If there is no treatment effect, then the numerator and denominator are both measuring the same random, unsystematic variance and the F -ratio should produce a value near 1.00. On the other hand, the existence of a treatment effect should make the numerator substantially larger than the denominator and result in a large value for the F -ratio.

For the independent-measures design, the unsystematic differences include individual differences as well as other random sources of error. Thus, for the independent-measures ANOVA, the F -ratio has the following structure:

$$F = \frac{\text{treatment effect} + \text{individual differences and other error}}{\text{individual differences and other error}}$$

For the repeated-measures design, the individual differences are eliminated or subtracted out, and the resulting F -ratio is structured as follows:

$$F = \frac{\text{treatment effect} + \text{error (excluding individual differences)}}{\text{error (excluding individual differences)}}$$

The removal of individual differences from the analysis becomes an advantage in situations in which very large individual differences exist among the participants being studied.

When individual differences are large, the presence of a treatment effect may be masked if an independent-measures study is performed. In this case, a repeated-measures design would be more sensitive in detecting a treatment effect because individual differences do not influence the value of the F -ratio.

This point will become evident in the following example. Suppose that we know how much variability is accounted for by the different sources of variance. For example,

treatment effect = 10 units of variance

individual differences = 10 units of variance

other error = 1 unit of variance

Notice that a large amount of the variability in the experiment is caused by individual differences. By comparing the F -ratios for an independent- and a repeated-measures analysis, we are able to see a fundamental difference between the two types of experimental designs. For an independent-measures experiment, we obtain

$$\begin{aligned} F &= \frac{\text{treatment effect} + \text{individual differences} + \text{error}}{\text{individual differences} + \text{error}} \\ &= \frac{10 + 10 + 1}{10 + 1} = \frac{21}{11} = 1.91 \end{aligned}$$

Thus, the independent-measures ANOVA produces an F -ratio of $F = 1.91$. Recall that the F -ratio is structured to produce $F = 1.00$ if there is no treatment effect whatsoever. In this case, the F -ratio is near to 1.00 and strongly suggests that there is little or no treatment effect. If you check the F -distribution table in Appendix B, you will find that it is almost impossible for an F -ratio as small as 1.91 to be significant. For the independent-measures ANOVA, the 10-point treatment effect is overwhelmed by all of the other variance.

Now consider what happens with a repeated-measures ANOVA. With the individual differences removed, the F -ratio becomes:

$$\begin{aligned} F &= \frac{\text{treatment effect} + \text{error}}{\text{error}} \\ &= \frac{10 + 1}{1} = \frac{11}{1} = 11 \end{aligned}$$

For the repeated-measures ANOVA, the numerator of the F -ratio (which includes the treatment effect) is 11 times larger than the denominator (which has no treatment effect). This result strongly indicates that there is a substantial treatment effect. In this example, the F -ratio is much larger for the repeated-measures study because the individual differences, which are extremely large, have been removed. In the independent-measures ANOVA, the presence of a treatment effect is obscured by the influence of individual differences. This problem is eliminated by the repeated-measures design, in which variability caused by individual differences is partitioned out of the analysis. When the individual differences are large, a repeated-measures experiment may provide a more sensitive test for a treatment effect. In statistical terms, a repeated-measures test has more *power* than an independent-measures test; that is, it is more likely to detect a real treatment effect.

**INDIVIDUAL DIFFERENCES
AND THE CONSISTENCY
OF THE TREATMENT EFFECTS**

As we have demonstrated, one major advantage of a repeated-measures design is that it removes individual differences from the denominator of the F -ratio, which usually increases the likelihood of obtaining a significant result. However, removing individual differences is an advantage only when the treatment effects are reasonably consistent for all of the participants. If the treatment effects are not consistent across participants, the individual differences tend to disappear and value in the denominator is not noticeably reduced by removing them. This phenomenon is demonstrated in the following example.

EXAMPLE 13.2

Table 13.4 presents hypothetical data from a repeated-measures research study. We constructed the data specifically to demonstrate the relationship between consistent treatment effects and large individual differences.

TABLE 13.4

Data from a repeated-measures study comparing three treatments. The data show consistent treatment effects from one participant to another, which produce consistent and relatively large differences in the individual P totals.

Person	Treatment			
	I	II	III	
A	0	1	2	$P = 3$
B	1	2	3	$P = 6$
C	2	4	6	$P = 12$
D	3	5	7	$P = 15$
	$T = 6$	$T = 12$	$T = 18$	
	$SS = 5$	$SS = 10$	$SS = 17$	

First, notice the consistency of the treatment effects. Treatment II has the same effect on every participant, increasing everyone's score by 1 or 2 points compared to treatment I. Also, treatment III produces a consistent increase of 1 or 2 points compared to treatment II. One consequence of the consistent treatment effects is that the individual differences are maintained in all of the treatment conditions. For example, participant A has the lowest score in all three treatments, and participant D always has the highest score. The participant totals (P values) reflect the consistent differences. For example, participant D has the largest score in every treatment and, therefore, has the largest P value. Also notice that there are big differences between the P totals from one individual to the next. For these data, $SS_{\text{between subjects}} = 30$ points.

Now consider the data in Table 13.5. To construct these data we started with the same numbers within each treatment that were used in Table 13.4. However, we scrambled the numbers within each column to eliminate the consistency of the treatment effects. In Table 13.5, for example, two participants show an increase in scores as they go from treatment I to treatment II, and two show a decrease. The data also show an inconsistent treatment effect as the participants go from treatment II to treatment III. One consequence of the inconsistent treatment effects is that there are no consistent individual differences between participants. Participant C, for example, has the lowest score in treatment II and the highest score in treatment III. As a result, there are no longer consistent differences between the individual participants. All of the P totals are about the same. For these data, $SS_{\text{between subjects}} = 3.33$ points. Because the two sets of data (Tables 13.4 and 13.5) have the same treatment totals (T values) and SS values, they have the same $SS_{\text{between treatments}}$ and $SS_{\text{within treatments}}$. For both sets of data,

$$SS_{\text{between treatments}} = 18 \text{ and } SS_{\text{within treatments}} = 32$$

However, there is a huge difference between the two sets of data when you compute SS_{error} for the denominator of the F -ratio. For the data in Table 13.4, with consistent treatment effects and large individual differences,

$$\begin{aligned} SS_{\text{error}} &= SS_{\text{within treatments}} - SS_{\text{between subjects}} \\ &= 32 - 30 \\ &= 2 \end{aligned}$$

For the data in Table 13.5, with no consistent treatment effects and relatively small differences between the individual P totals,

$$\begin{aligned} SS_{\text{error}} &= SS_{\text{within treatments}} - SS_{\text{between subjects}} \\ &= 32 - 3.33 \\ &= 28.67 \end{aligned}$$

TABLE 13.5

Data from a repeated-measures study comparing three treatments. The data show treatment effects that are inconsistent from one participant to another and, as a result, produce relatively small differences in the individual P totals. Note that the data have exactly the same scores within each treatment as the data in Table 13.5, however, the scores have been scrambled to eliminate the consistency of the treatment effects.

Person	Treatment			
	I	II	III	
A	0	4	3	$P = 7$
B	1	5	2	$P = 8$
C	2	1	7	$P = 10$
D	3	2	6	$P = 11$
	$T = 6$	$T = 12$	$T = 18$	
	$SS = 5$	$SS = 10$	$SS = 17$	

Thus, consistent treatment effects tend to produce a relatively small error term for the F -ratio. As a result, consistent treatment effects are more likely to be statistically significant (reject the null hypothesis). For the examples we have been considering, the data in Table 13.4 produce an F -ratio of $F = 27.0$. With $df = 2, 6$, this F -ratio is well into the critical region for $\alpha = .05$ or $.01$ and we conclude that there are significant differences among the three treatments. On the other hand, the same mean differences in Table 13.5 produce $F = 1.88$. With $df = 2, 6$, this value is not in the critical region for $\alpha = .05$ or $.01$, and we conclude that there are no significant differences.

In summary, when treatment effects are consistent from one individual to another, the individual differences also tend to be consistent and relatively large. The large individual differences get subtracted from the denominator of the F -ratio producing a larger value for F and increasing the likelihood that the F -ratio will be in the critical region.

13.5 REPEATED-MEASURES ANOVA AND REPEATED-MEASURES t TEST

As we noted in Chapter 12 (pp. 420–421), whenever you are evaluating the difference between two sample means, you can use either a t test or ANOVA. In Chapter 12, we demonstrated that the two tests are related in many respects, including:

1. The two tests always reach the same conclusion about the null hypothesis.
2. The basic relationship between the two test statistics is $F = t^2$.
3. The df value for the t statistic is identical to the df value for the denominator of the F -ratio.
4. If you square the critical value for the two-tailed t test, you obtain the critical value for the F -ratio. Again, the basic relationship is $F = t^2$.

In Chapter 12, these relationships were demonstrated for the independent-measures tests, but they are also true for repeated-measures designs comparing two treatment conditions. The following example demonstrates the relationships.

EXAMPLE 13.3

The following table shows the data from a repeated-measures study comparing two treatment conditions. We have structured the data in a format that is compatible with the repeated-measures *t* test. Note that the calculations for the *t* test are based on the difference scores (*D* values) in the final column.

Participant	Treatment		
	I	II	D
A	3	5	2
B	4	14	10
C	5	7	2
D	4	6	2

$M_D = 4$
 $SS_D = 48$

The repeated-measures *t* test The null hypothesis for the *t* test states that, for the general population, there is no mean difference between the two treatment conditions.

$$H_0: \mu_D = 0$$

With $n = 4$ participants, the test has $df = 3$ and the critical boundaries for a two-tailed test with $\alpha = .05$ are $t = \pm 3.182$.

For these data, the sample mean difference is $M_D = 4$, the variance for the difference scores is $s^2 = 16$, and the standard error is $S_{M_D} = 2$ points. These values produce a *t* statistic of

$$t = \frac{M_D - \mu_D}{s_{M_D}} = \frac{4 - 0}{2} = 2.00$$

The *t* value is not in the critical region so we fail to reject H_0 and conclude that there is no significant difference between the two treatments.

The repeated-measures ANOVA Now we reorganize the data into a format that is compatible with a repeated-measures ANOVA. Notice that the ANOVA uses the original scores (not the difference scores) and requires the *P* totals for each participant.

Participant	Treatment			<i>P</i>
	I	II	<i>P</i>	
A	3	5	8	$G = 48$
B	4	14	18	$\Sigma X^2 = 372$
C	5	7	12	$N = 8$
D	4	6	10	

Again, the null hypothesis states that, for the general population, there is no mean difference between the two treatment conditions.

$$H_0: \mu_1 = \mu_2$$

For this study, $df_{\text{between treatments}} = 1$, $df_{\text{within treatments}} = 6$, $df_{\text{between subjects}} = 3$, which produce $df_{\text{error}} = (6 - 3) = 3$. Thus, the F -ratio has $df = 1, 3$ and the critical value for $\alpha = .05$ is $F = 10.13$. Note that the denominator of the F -ratio has the same df value as the t statistic ($df = 3$) and that the critical value for F is equal to the squared critical value for t ($10.13 = 3.182^2$).

For these data, $SS_{\text{total}} = 84$,

$$SS_{\text{within}} = 52$$

$$SS_{\text{between treatments}} = (84 - 52) = 32$$

$$SS_{\text{between subjects}} = 28$$

$$SS_{\text{error}} = (52 - 28) = 24$$

The two variances in the F -ratio are

$$MS_{\text{between treatments}} = \frac{SS_{\text{between treatments}}}{df_{\text{between treatments}}} = \frac{32}{1} = 32$$

$$\text{and } MS_{\text{error}} = \frac{SS_{\text{error}}}{df_{\text{error}}} = \frac{24}{3} = 8$$

$$\text{and the } F\text{-ratio is } F = \frac{MS_{\text{between treatments}}}{MS_{\text{error}}} = \frac{32}{8} = 4.00$$

Notice that the F -ratio and the t statistic are related by the equation $F = t^2$ ($4 = 2^2$). The F -ratio (like the t statistic) is not in the critical region so, once again, we fail to reject H_0 and conclude that there is no significant difference between the two treatments.

SUMMARY

1. The repeated-measures ANOVA is used to evaluate the mean differences obtained in a research study comparing two or more treatment conditions using the same sample of individuals in each condition. The test statistic is an F -ratio, in which the numerator measures the variance (differences) between treatments and the denominator measures the variance (differences) that is expected without any treatment effects or individual differences.

$$F = \frac{MS_{\text{between treatments}}}{MS_{\text{error}}}$$

2. The first stage of the repeated-measures ANOVA is identical to the independent-measures ANOVA and separates the total variability into two components: between-treatments and within-treatments. Because a

repeated-measures design uses the same subjects in every treatment condition, the differences between treatments cannot be caused by individual differences. Thus, individual differences are automatically eliminated from the between-treatments variance in the numerator of the F -ratio.

3. In the second stage of the repeated-measures analysis, individual differences are computed and removed from the denominator of the F -ratio. To remove the individual differences, you first compute the variability between subjects (SS and df) and then subtract these values from the corresponding within-treatments values. The residual provides a measure of error excluding individual differences, which is the appropriate denominator for the repeated-measures F -ratio. The equations for analyzing SS and df for the repeated-measures ANOVA are presented in Figure 13.3.

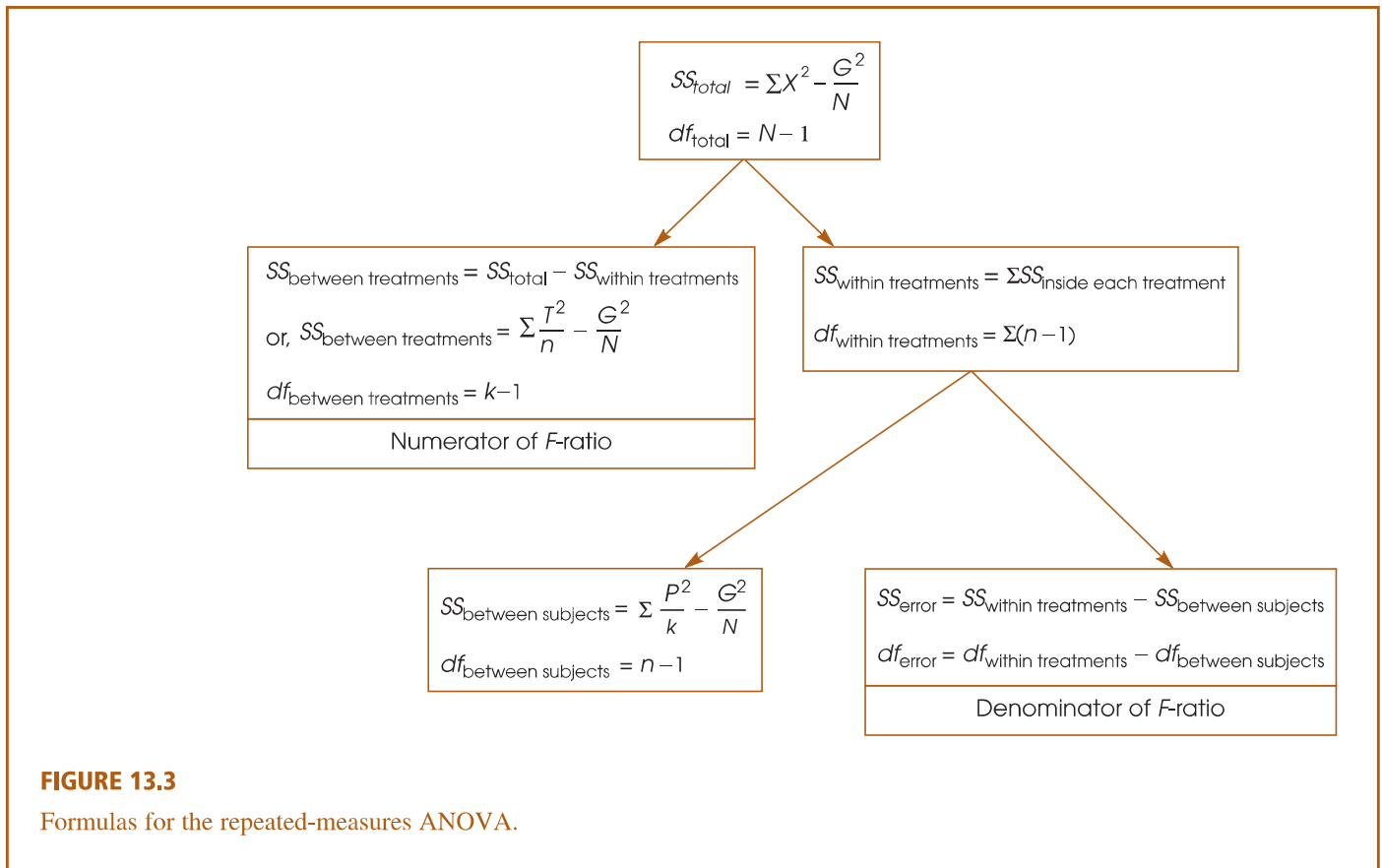


FIGURE 13.3

Formulas for the repeated-measures ANOVA.

4. Effect size for the repeated-measures ANOVA is measured by computing eta squared, the percentage of variance accounted for by the treatment effect. For the repeated-measures ANOVA

$$\eta^2 = \frac{SS_{\text{between treatments}}}{SS_{total} - SS_{\text{between subjects}}}$$

$$= \frac{SS_{\text{between treatments}}}{SS_{\text{between treatments}} + SS_{\text{error}}}$$

Because part of the variability (the SS caused by individual differences) is removed before computing η^2 , this measure of effect size is often called a partial eta squared.

5. When the obtained F -ratio is significant (that is, H_0 is rejected), it indicates that a significant difference lies between at least two of the treatment conditions. To determine exactly where the difference lies, post hoc comparisons may be made. Post hoc tests, such as Tukey's HSD, use MS_{error} rather than $MS_{\text{within treatments}}$ and df_{error} instead of $df_{\text{within treatments}}$.
6. A repeated-measures ANOVA eliminates the influence of individual differences from the analysis. If individual differences are extremely large, then a treatment effect might be masked in an independent-measures experiment. In this case, a repeated-measures design might be a more sensitive test for a treatment effect.

KEY TERMS

individual differences (437)

between-treatments variance(437)

error variance (438)

between-subjects variance (440)

RESOURCES

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

You can find a tutorial quiz and other learning exercises for Chapter 13 on the book companion website.



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

CENGAGE ^{brain}.com

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

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

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

SPSS

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

Data Entry

Enter the scores for each treatment condition in a separate column, with the scores for each individual in the same row. All of the scores for the first treatment go in the VAR00001 column, the second treatment scores go in the VAR00002 column, and so on.

Data Analysis

1. Click **Analyze** on the tool bar, select **General Linear Model**, and click on **Repeated-Measures**.
2. SPSS presents a box entitled **Repeated-Measures Define Factors**. Within the box, the Within-Subjects Factor Name should already contain **Factor 1**. If not, type in Factor 1.
3. Enter the **Number of levels** (number of different treatment conditions) in the next box.
4. Click **Add**.
5. Click **Define**.
6. One by one, move the column labels for your treatment conditions into the **Within Subjects Variables** box. (Highlight the column label on the left and click the arrow to move it into the box.)
7. If you want descriptive statistics for each treatment, click on the **Options** box, select **Descriptives**, and click **Continue**.
8. Click **OK**.

SPSS Output

We used the SPSS program to analyze the data from the television viewing study in Example 13.1 and portions of the program output are shown in Figure 13.4. Note that large portions of the SPSS output are not relevant for our purposes and are not included in Figure 13.1. The first item of interest is the table of **Descriptive Statistics**, which presents the mean, standard deviation, and number of scores for each treatment. Next, we skip to the table showing **Tests of Within-Subjects Effects**. The top line of the factor 1 box (Sphericity Assumed) shows the between-treatments sum of squares, degrees of freedom, and mean square that form the numerator of the F -ratio. The same line reports the value of the F -ratio and the level of significance (the p value or alpha level). Similarly, the top line of the Error (factor 1) box shows the sum of squares, the degrees of freedom, and the mean square for the error term (the denominator of the F -ratio). The final box in the output (not shown in Figure 13.4) is labeled **Tests of Between-Subjects Effects** and the bottom line (Error) reports the between-subjects sum of squares and degrees of freedom (ignore the mean square and F -ratio, which are not part of the repeated-measures ANOVA).

Descriptive Statistics

	Mean	Std. Deviation	N
VAR00001	1.0000	1.41421	5
VAR00002	2.0000	1.41421	5
VAR00003	5.0000	1.58114	5
VAR00004	4.0000	1.22474	5

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor 1	Sphericity Assumed	50.000	3	16.667	25.000	.000
	Greenhouse-Geisser	50.000	1.600	31.250	25.000	.001
	Huynh-Feldt	50.000	2.500	20.000	25.000	.000
	Lower-bound	50.000	1.000	50.000	25.000	.007
Error (factor 1)	Sphericity Assumed	8.000	12	.667		
	Greenhouse-Geisser	8.000	6.400	1.250		
	Huynh-Feldt	8.000	10.000	.800		
	Lower-bound	8.000	4.000	2.000		

FIGURE 13.4

Portions of the SPSS output for the repeated-measures ANOVA for the television viewing study in Example 13.1.

FOCUS ON PROBLEM SOLVING

1. Before you begin a repeated-measures ANOVA, complete all of the preliminary calculations needed for the ANOVA formulas. This requires that you find the total for each treatment (*T*s), the total for each person (*P*s), the grand total (*G*), the *SS* for each treatment condition, and ΣX^2 for the entire set of *N* scores. As a partial check on these calculations, be sure that the *T* values add up to *G* and that the *P* values have a sum of *G*.
2. To help remember the structure of repeated-measures ANOVA, keep in mind that a repeated-measures experiment eliminates the contribution of individual differences. There are no individual differences contributing to the numerator of the *F*-ratio ($MS_{\text{between treatments}}$) because the same individuals are used for all treatments. Therefore, you must also eliminate individual differences in the denominator. This is accomplished by partitioning within-treatments variability into two components: between-subjects variability and error variability. It is the *MS* value for error variability that is used in the denominator of the *F*-ratio.

DEMONSTRATION 13.1

REPEATED-MEASURES ANOVA

The following data were obtained from a research study examining the effect of sleep deprivation on motor-skills performance. A sample of five participants was tested on a motor-skills task after 24 hours of sleep deprivation, tested again after 36 hours, and tested once more after 48 hours. The dependent variable is the number of errors made on the motor-skills task. Do these data indicate that the number of hours of sleep deprivation has a significant effect on motor skills performance?

Participant	24 Hours	36 Hours	48 Hours	P totals	
A	0	0	6	6	<i>N</i> = 15
B	1	3	5	9	<i>G</i> = 45
C	0	1	5	6	$\Sigma X^2 = 245$
D	4	5	9	18	
E	0	1	5	6	
$T = 5$ $T = 10$ $T = 30$ $SS = 12$ $SS = 16$ $SS = 12$					

STEP 1 State the hypotheses, and specify alpha. The null hypothesis states that, for the general population, there are no differences among the three deprivation conditions. Any differences that exist among the samples are simply the result of chance or error. In symbols,

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

The alternative hypothesis states that there are differences among the conditions.

$$H_1: \text{At least one of the treatment means is different.}$$

We use $\alpha = .05$.

STEP 2 The repeated-measures analysis. Rather than compute the df values and look for a critical value for F at this time, we proceed directly to the ANOVA.

STAGE 1 The first stage of the analysis is identical to the independent-measures ANOVA presented in Chapter 12.

$$SS_{\text{total}} = \sum X^2 - \frac{G^2}{N} = 245 - \frac{45^2}{15} = 110$$

$$SS_{\text{within}} = \sum SS_{\text{inside each treatment}} = 12 + 16 + 12 = 40$$

$$SS_{\text{between}} = \sum \frac{T^2}{n} - \frac{G^2}{N} = \frac{5^2}{5} + \frac{10^2}{5} + \frac{30^2}{5} - \frac{45^2}{15} = 70$$

and the corresponding degrees of freedom are

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

$$df_{\text{within}} = \sum df = 4 + 4 + 4 = 12$$

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

STAGE 2 The second stage of the repeated-measures analysis removes the individual differences from the denominator of the F -ratio.

$$\begin{aligned} SS_{\text{between subjects}} &= \sum \frac{P^2}{k} - \frac{G^2}{N} \\ &= \frac{6^2}{3} + \frac{9^2}{3} + \frac{6^2}{3} + \frac{18^2}{3} + \frac{6^2}{3} - \frac{45^2}{15} \\ &= 36 \end{aligned}$$

$$\begin{aligned} SS_{\text{error}} &= SS_{\text{within}} - SS_{\text{between subjects}} \\ &= 40 - 36 \\ &= 4 \end{aligned}$$

and the corresponding df values are

$$df_{\text{between subjects}} = n - 1 = 4$$

$$\begin{aligned} df_{\text{error}} &= df_{\text{within}} - df_{\text{between subjects}} \\ &= 12 - 4 \\ &= 8 \end{aligned}$$

The mean square values that form the F -ratio are as follows:

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

$$MS_{\text{error}} = \frac{SS_{\text{error}}}{df_{\text{error}}} = \frac{4}{8} = 0.50$$

Finally, the F -ratio is

$$F = \frac{MS_{\text{between}}}{MS_{\text{error}}} = \frac{35}{0.50} = 70.00$$

STEP 3 Make a decision and state a conclusion. With $df = 2, 8$ and $\alpha = .05$, the critical value is $F = 4.46$. Our obtained F -ratio ($F = 70.00$) is well into the critical region, so our decision is to reject the null hypothesis and conclude that there are significant differences among the three levels of sleep deprivation.

DEMONSTRATION 13.2

EFFECT SIZE FOR THE REPEATED-MEASURES ANOVA

We compute η^2 , the percentage of variance explained by the treatment differences, for the data in Demonstration 13.1. Using Equation 13.11 we obtain

$$\eta^2 = \frac{SS_{\text{between treatments}}}{SS_{\text{between treatments}} + SS_{\text{error}}} = \frac{70}{70 + 4} = \frac{70}{74} = 0.95 \quad (\text{or } 95\%)$$

PROBLEMS

- How does the denominator of the F -ratio (the error term) differ for a repeated-measures ANOVA compared to an independent-measures ANOVA?
- The repeated-measures ANOVA can be viewed as a two-stage process. What is the purpose of the second stage?
- A researcher conducts an experiment comparing three treatment conditions with $n = 10$ scores in each condition.
 - If the researcher uses an independent-measures design, how many individuals are needed for the study and what are the df values for the F -ratio?
 - If the researcher uses a repeated-measures design, how many individuals are needed for the study and what are the df values for the F -ratio?
- A researcher conducts a repeated-measures experiment using a sample of $n = 8$ subjects to evaluate the differences among four treatment conditions. If the results are examined with an ANOVA, what are the df values for the F -ratio?
- A researcher uses a repeated-measures ANOVA to evaluate the results from a research study and reports an F -ratio with $df = 2, 30$.
 - How many treatment conditions were compared in the study?
 - How many individuals participated in the study?
- A published report of a repeated-measures research study includes the following description of the statistical analysis. “The results show significant differences among the treatment conditions, $F(2, 20) = 6.10, p < .01$.”
 - How many treatment conditions were compared in the study?
 - How many individuals participated in the study?
- The following data were obtained from a repeated-measures study comparing three treatment conditions. Use a repeated-measures ANOVA with $\alpha = .05$ to determine whether there are significant mean differences among the three treatments.

Person	Treatments			Person Totals	
	I	II	III		
A	0	4	2	$P = 6$	
B	1	5	6	$P = 12$	$N = 18$
C	3	3	3	$P = 9$	$G = 48$
D	0	1	5	$P = 6$	$\Sigma X^2 = 184$
E	0	2	4	$P = 6$	
F	2	3	4	$P = 9$	
$M = 1$ $M = 3$ $M = 4$ $T = 6$ $T = 18$ $T = 24$ $SS = 8$ $SS = 10$ $SS = 10$					

8. The following data were obtained from a repeated-measures study comparing two treatment conditions. Use a repeated-measures ANOVA with $\alpha = .05$ to determine whether there are significant mean differences between the two treatments.

Person	Treatments		Person Totals	
	I	II		
A	3	5	$P = 8$	
B	5	9	$P = 14$	$N = 16$
C	1	5	$P = 6$	$G = 80$
D	1	7	$P = 8$	$\Sigma X^2 = 500$
E	5	9	$P = 14$	
F	3	7	$P = 10$	
G	2	6	$P = 8$	
H	4	8	$P = 12$	

$M = 3$	$M = 7$
$T = 24$	$T = 56$
$SS = 18$	$SS = 18$

9. The following data were obtained from a repeated-measures study comparing three treatment conditions.
- Use a repeated-measures ANOVA with $\alpha = .05$ to determine whether there are significant mean differences among the three treatments.
 - Compute η^2 , the percentage of variance accounted for by the mean differences, to measure the size of the treatment effects.
 - Write a sentence demonstrating how a research report would present the results of the hypothesis test and the measure of effect size.

Person	Treatments			Person Totals	
	I	II	III		
A	1	1	4	$P = 6$	
B	3	4	8	$P = 15$	$N = 15$
C	0	2	7	$P = 9$	$G = 45$
D	0	0	6	$P = 6$	$\Sigma X^2 = 231$
E	1	3	5	$P = 9$	

$M = 1$	$M = 2$	$M = 6$
$T = 5$	$T = 10$	$T = 30$
$SS = 6$	$SS = 10$	$SS = 10$

10. For the data in problem 9,
- Compute SS_{total} and $SS_{\text{between treatments}}$.
 - Eliminate the mean differences between treatments by adding 2 points to each score in treatment I, adding 1 point to each score in treatment II, and

subtracting 3 points from each score in treatment III. (All three treatments should end up with $M = 3$ and $T = 15$.)

- Calculate SS_{total} for the modified scores. (*Caution:* You first must find the new value for ΣX^2 .)
 - Because the treatment effects were eliminated in part b, you should find that SS_{total} for the modified scores is smaller than SS_{total} for the original scores. The difference between the two SS values should be exactly equal to the value of $SS_{\text{between treatments}}$ for the original scores.
11. The following data were obtained from a repeated-measures study comparing three treatment conditions.

Subject	Treatment			P	
	I	II	III		
A	6	8	10	24	$G = 48$
B	5	5	5	15	$\Sigma X^2 = 294$
C	1	2	3	6	
D	0	1	2	3	

$T = 12$	$T = 16$	$T = 20$
$SS = 26$	$SS = 30$	$SS = 38$

Use a repeated-measures ANOVA with $\alpha = .05$ to determine whether these data are sufficient to demonstrate significant differences between the treatments.

12. In Problem 11 the data show large and consistent differences between subjects. For example, subject A has the largest score in every treatment and subject D always has the smallest score. In the second stage of the ANOVA, the large individual differences are subtracted out of the denominator of the F -ratio, which results in a larger value for F .

The following data were created by using the same numbers that appeared in Problem 11. However, we eliminated the consistent individual differences by scrambling the scores within each treatment.

Subject	Treatment			P	
	I	II	III		
A	6	2	3	11	$G = 48$
B	5	1	5	11	$\Sigma X^2 = 294$
C	0	5	10	15	
D	1	8	2	11	

$T = 12$	$T = 16$	$T = 20$
$SS = 26$	$SS = 30$	$SS = 38$

- Use a repeated-measures ANOVA with $\alpha = .05$ to determine whether these data are sufficient to demonstrate significant differences between the treatments.

- b. Explain how the results of this analysis compare with the results from Problem 11.
13. One of the primary advantages of a repeated-measures design, compared to an independent-measures design, is that it reduces the overall variability by removing variance caused by individual differences. The following data are from a research study comparing three treatment conditions.
- Assume that the data are from an independent-measures study using three separate samples, each with $n = 6$ participants. Ignore the column of P totals and use an independent-measures ANOVA with $\alpha = .05$ to test the significance of the mean differences.
 - Now assume that the data are from a repeated-measures study using the same sample of $n = 6$ participants in all three treatment conditions. Use a repeated-measures ANOVA with $\alpha = .05$ to test the significance of the mean differences.
 - Explain why the two analyses lead to different conclusions.

Treatment 1	Treatment 2	Treatment 3	P	
6	9	12	27	
8	8	8	24	$N = 18$
5	7	9	21	$G = 108$
0	4	8	12	$\Sigma X^2 = 800$
2	3	4	9	
3	5	7	15	

$M = 4$	$M = 6$	$M = 8$
$T = 24$	$T = 36$	$T = 48$
$SS = 42$	$SS = 28$	$SS = 34$

14. The following data are from an experiment comparing three different treatment conditions:

A	B	C	
0	1	2	$N = 15$
2	5	5	$\Sigma X^2 = 354$
1	2	6	
5	4	9	
2	8	8	

$T = 10$	$T = 20$	$T = 30$
$SS = 14$	$SS = 30$	$SS = 30$

- If the experiment uses an *independent-measures design*, can the researcher conclude that the treatments are significantly different? Test at the .05 level of significance.

- If the experiment is done with a *repeated-measures design*, should the researcher conclude that the treatments are significantly different? Set alpha at .05 again.
 - Explain why the analyses in parts a and b lead to different conclusions.
15. A researcher is evaluating customer satisfaction with the service and coverage of two phone carriers. Each individual in a sample of $n = 25$ uses one carrier for two weeks and then switches to the other. Each participant then rates the two carriers. The following table presents the results from the repeated-measures ANOVA comparing the average ratings. Fill in the missing values in the table. (*Hint: Start with the df values.*)

Source	SS	df	MS
Between treatments	___	___	2 $F =$ ___
Within treatments	___	___	
Between subjects	___	___	
Error	12	___	___
Total	23	___	

16. The following summary table presents the results from a repeated-measures ANOVA comparing three treatment conditions with a sample of $n = 11$ subjects. Fill in the missing values in the table. (*Hint: Start with the df values.*)

Source	SS	df	MS
Between treatments	___	___	___ $F = 5.00$
Within treatments	80	___	
Between subjects	___	___	
Error	60	___	___
Total	___	___	

17. The following summary table presents the results from a repeated-measures ANOVA comparing four treatment conditions, each with a sample of $n = 12$ participants. Fill in the missing values in the table. (*Hint: Start with the df values.*)

Source	SS	df	MS
Between treatments	54	___	20 $F =$ ___
Within treatments	___	___	
Between subjects	___	___	
Error	___	___	3
Total	194	___	

18. A recent study indicates that simply giving college students a pedometer can result in increased walking (Jackson & Howton, 2008). Students were given pedometers for a 12-week period, and asked to record the average number of steps per day during weeks 1, 6, and 12. The following data are similar to the results obtained in the study.

Participant	Number of steps (x1000)			P		
	Week					
	1	6	12			
A	6	8	10	24		
B	4	5	6	15		
C	5	5	5	15	$G = 72$	
D	1	2	3	6	$\Sigma X^2 = 400$	
E	0	1	2	3		
F	2	3	4	9		
		$T = 18$	$T = 24$	$T = 30$		
		$SS = 28$	$SS = 32$	$SS = 40$		

- Use a repeated-measures ANOVA with $\alpha = .05$ to determine whether the mean number of steps changes significantly from one week to another.
 - Compute η^2 to measure the size of the treatment effect.
 - Write a sentence demonstrating how a research report would present the results of the hypothesis test and the measure of effect size.
19. A repeated-measures experiment comparing only two treatments can be evaluated with either a t statistic or an ANOVA. As we found with the independent-measures design, the t test and the ANOVA produce equivalent conclusions, and the two test statistics are related by the equation $F = t^2$. The following data are from a repeated-measures study:

Subject	Treatment 1	Treatment 2	Difference
1	2	4	+2
2	1	3	+2
3	0	10	+10
4	1	3	+2

- Use a repeated-measures t statistic with $\alpha = .05$ to determine whether the data provide evidence of a significant difference between the two treatments. (*Caution:* ANOVA calculations are done with the X values, but for t you use the difference scores.)
- Use a repeated-measures ANOVA with $\alpha = .05$ to evaluate the data. (You should find $F = t^2$.)

20. For either independent-measures or repeated-measures designs comparing two treatments, the mean difference can be evaluated with either a t test or an ANOVA. The two tests are related by the equation $F = t^2$. For the following data,

- Use a repeated-measures t test with $\alpha = .05$ to determine whether the mean difference between treatments is statistically significant.
- Use a repeated-measures ANOVA with $\alpha = .05$ to determine whether the mean difference between treatments is statistically significant. (You should find that $F = t^2$.)

Person	Treatment 1	Treatment 2	Difference	
A	4	7	3	
B	2	11	9	
C	3	6	3	
D	7	10	3	
		$M = 4$	$M = 8.5$	$M_D = 4.5$
		$T = 16$	$T = 34$	
		$SS = 14$	$SS = 17$	$SS = 27$

21. In the Preview section for this chapter, we presented an example of a delayed discounting study in which people are willing to settle for a smaller reward today in exchange for a larger reward in the future. The following data represent the typical results from one of these studies. The participants are asked how much they would take today instead of waiting for a specific delay period to receive \$1000. Each participant responds to all 5 of the delay periods. Use a repeated-measures ANOVA with $\alpha = .01$ to determine whether there are significant differences among the 5 delay periods for the following data:

Participant	1 month	6 months	1 year	2 years	5 years
A	950	850	800	700	550
B	800	800	750	700	600
C	850	750	650	600	500
D	750	700	700	650	550
E	950	900	850	800	650
F	900	900	850	750	650

22. The endorphins released by the brain act as natural painkillers. For example, Gintzler (1970) monitored endorphin activity and pain thresholds in pregnant rats during the days before they gave birth. The data showed an increase in pain threshold as the pregnancy progressed. The change was gradual until 1 or 2 days

before birth, at which point there was an abrupt increase in pain threshold. Apparently a natural painkilling mechanism was preparing the animals for the stress of giving birth. The following data represent pain-threshold scores similar to the results obtained by Gintzler. Do these data indicate a significant change in pain threshold? Use a repeated-measures ANOVA with $\alpha = .01$.

Subject	Days Before Giving Birth			
	7	5	3	1
A	39	40	49	52
B	38	39	44	55
C	44	46	50	60
D	40	42	46	56
E	34	33	41	52



Improve your statistical skills with ample practice exercises and detailed explanations on every question. Purchase www.aplia.com/statistics