

One-Way, Repeated-Measures ANOVA



LEARNING OBJECTIVES

- Describe what repeated-measures ANOVA does.
- Complete a one-way, repeated-measures ANOVA.*
- Interpret a one-way, repeated-measures ANOVA.

CHAPTER OVERVIEW

The last chapter introduced between-subjects, one-way ANOVA, an extension of the independent-samples t test to situations in which more than two independent samples are being compared. Just as the chapter on independent-samples t tests was followed by a chapter on paired-samples t tests, the chapter on *between-subjects*, one-way ANOVA is followed by a chapter on *within-subjects*, measures one-way ANOVA. Said another way, the chapter on ANOVA for independent samples is followed by this chapter on ANOVA for dependent samples. Repeated-measures, one-way ANOVA is an extension of paired-samples t tests to situations in which there are more than two dependent samples.

11.1 Introduction to Repeated-Measures ANOVA

11.2 Calculating One-Way, Repeated-Measures ANOVA

11.3 Interpreting One-Way, Repeated-Measures ANOVA

11.1 Introduction to Repeated-Measures ANOVA

In this chapter, we learn about *within-subjects* ANOVA, commonly called **repeated-measures ANOVA**. Repeated-measures ANOVA is an extension of the paired-samples t test. The paired-samples t test can only be used when comparing the means of two dependent samples. But, a repeated-measures ANOVA can be used when there are two or more dependent samples.

There are many different types of experiments that use repeated measures. What they have in common is a connection between the cases in one cell and the cases in the other cells. Repeated measures may involve:

- *The same cases measured at multiple points of time.* For example, to examine how weight changes over college, a group of first-year college students might be weighed at the start and the end of each of the four years of college.

*Note: Formulas for calculating sums of squares appear in an appendix at the end of the chapter.

- *The same cases measured under multiple conditions.* For example, to examine how alcohol affects one's ability to walk a straight line, 21-year-olds might be asked to walk a straight line before drinking anything, after drinking a placebo beer, after drinking one beer, after two, after four, and after six.
- *Different cases, matched so that they are similar in some dimension(s).* For example, a researcher comparing murder rates for cities with high, medium, and low numbers of sunny days per year might be concerned that population density and poverty level could affect the results. To account for these other factors, he could group cities with the same populations and unemployment rates into sets of three: one with a high number of sunshine days, one low on the number of sunshine days, and one in the middle.

In all these examples of dependent samples, we're looking at differences within the same subjects (or matched subjects) over time or across conditions (the treatment). Thus, it is important to make sure that the subjects are arrayed in the same order from cell to cell to cell. If a participant's data point is listed first in one cell, then it should be first in the other cells.

The beauty of analysis of variance is that it partitions the variability in the dependent variable into the components that account for the variability. Between-subjects, one-way ANOVA, covered in the last chapter, distinguishes variability due to how the groups are treated (between-group variability) from variability due to individual differences among subjects (within-group variability). With between-subjects ANOVA, it doesn't matter how subjects in a cell are arrayed. In our example, the three rats that ran the maze to reach low-calorie food ran it in 30, 31, and 32 seconds. And that is how the data showed up in the first cell of Table 10.3. But, the data could just as easily have been listed as 31, 32, and 30 or as 31, 30, and 32. These cases are independent of each other. No matter the order that the results are listed, the results of the between-subjects ANOVA will be the same. This is not true of repeated-measures ANOVA. Because the samples in repeated-measures ANOVA are dependent samples, the order in which they are arranged matters very much. Changing the order, unless it is changed consistently in all cells, will change the outcome.

In repeated-measures ANOVA, the levels of the independent variable, for example, the different times or conditions, make up the columns and the subjects make up the rows. By having each subject on a row by itself, subjects can be treated as a second explanatory variable, as a second factor. Treating subjects as a factor allows repeated-measures ANOVA to measure—to account for—variability due to them. By partitioning out this variability, repeated-measures ANOVA obtains a purer measure of the effect of treatment and is a more statistically powerful test than is between-subjects ANOVA. What does it mean to be a statistically more powerful test? It means that this test gives us a greater likelihood of being able to reject the null hypothesis.

Figure 11.1 shows the similarities and differences between the sources of variability for between-subjects and repeated-measures ANOVA. With between-subjects ANOVA, total variability is partitioned into between and within. For repeated-measures ANOVA, total variability is also partitioned into between and within. But, within-group variability is then broken down further. First, variability due to subjects is removed. Then, the variability left over, labeled “residual,” is used as the denominator in the *F* ratio.

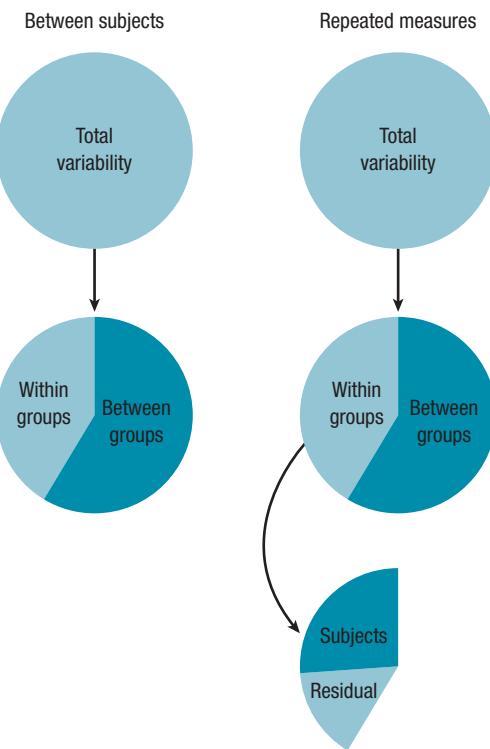


Figure 11.1 Partitioning Variability for One-Way ANOVA Variability in the dependent variable in one-way ANOVA is divided into between-groups and within-groups factors. For between-subjects, one-way ANOVA, the partitioning stops there. For a within-subjects design, the within-groups variability is further divided into that due to subjects and that which still remains.

The statistical power of repeated-measures ANOVA comes from the fact that it partitions out the variability due to subjects. Between-subjects ANOVA divides between-group variability by within-group variability to find the F ratio. Repeated-measures ANOVA calculates between-group variability the same way, so it will have the same numerator for the F ratio. But the denominator, which has variability due to subjects removed, will be smaller. This means that the F ratio will be larger—meaning it is more likely to fall in the rare zone, and therefore more likely to result in a rejection of the null hypothesis than for between-subjects, one-way ANOVA.

To see what this means, let's analyze the same set of data two ways—one way that doesn't partition out the subject variability and one that does. Let's compare two neighboring cities, Erie, PA, and Cleveland, OH, to see if they differ in how hot they get. From weather archives, we have obtained the average high temperature for each city for each month of the year. The data are seen in [Figure 11.2](#). Two things are readily apparent in this graph:

- The high temperatures in the two cities are very similar to each other.
- Both cities share a strong seasonal trend to temperature—it is hottest in the summer and coldest in the winter.

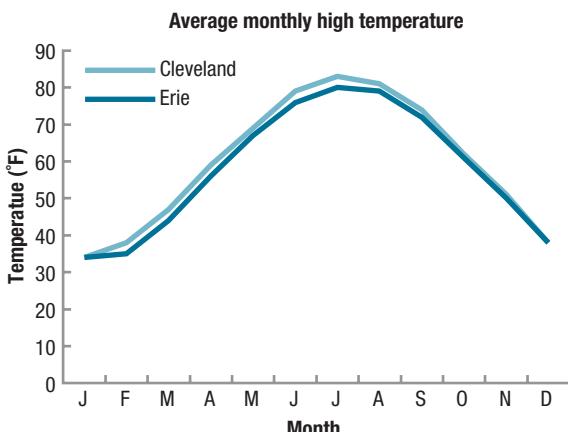


Figure 11.2 Average High Temperature per Month in Cleveland, OH, and Erie, PA Notice how little difference there is in temperature between the two cities each month and how much variability exists from month to month.

The paired-samples t test is appropriate to use to see if there is a statistically significant difference in temperature between the two cities. The t value turns out to be 5.38 and the difference between the two cities is a statistically significant one—Cleveland with a mean high temperature of 59.6°F is statistically significantly hotter than Erie with a mean of 57.7°F. A difference of

1.9°F doesn't sound like much, but it is enough to be statistically significant when analyzed with a paired-samples t test.

The reasonable next question to ask is what the size of the effect is. How much of the variability in the dependent variable of temperature is explained by the grouping variable of city? Judging by Figure 11.2, it shouldn't be much, right? Well, be surprised—72% of the variability in temperature is accounted for by city status! (This book does NOT advocate calculating either d or r^2 for a paired-samples t test. In this one instance, however, and just for this one example, I calculated r^2 .)

An r^2 of 72% is a huge effect, a whopping effect. Look at Figure 11.2 again. How can we account for the variability in temperature? Is there more variability in temperature due to differences between the cities each month or is there more variability in temperature due to changes from month to month? Clearly, more variability exists from month to month, which is due to individual differences among our cases, which are months. Regardless of city, months like July and August differ by 40 to 50 degrees F from months like January and February.

Does it make sense that the variability in temperature accounted for by the between-groups factor, city, is 72%? No! And it can't be so. If we just agreed that more of the variability in Figure 11.2 is due to the within-group factor (month) than the between-group factor, and the between-group factor accounted for 72% of the variability, then the within-group factor must explain at least 73% of the variability. That's impossible—we have just accounted for at least 145% of the variability!

So, what is going on? Because a paired-samples t test doesn't partition out the variability due to subjects, it mixes in variability due to individual differences with variability due to the grouping variable when calculating r^2 . Thus, r^2 is an overestimate of the percentage of variability due to the grouping variable and should not be used with paired-samples t tests. The same is true for Cohen's d for a paired-samples t test—it also overestimates the size of the effect.

How much of the variability in temperature is really due to the grouping variable of city? To answer that, we need a repeated-measures ANOVA, which treats variability due to the independent variable and that due to the subjects separately. ANOVA, which can be used when comparing *two* or more means, can be used in place of a t test. As we are about to see, there is a distinct advantage to using a repeated-measures ANOVA over a paired-samples t test.

When these temperature data are analyzed with a repeated-measures ANOVA, the difference between the mean temperatures due to location is still a statistically significant one. But now, because the variance due to subjects has been removed, the

effect size will be a purer measure of the effect of the independent variable. By removing individual subject variability, the explanatory power of the city where the measurement was taken is reduced from 72% to 1%. Yes, 1%. Look at Figure 11.2 again. Doesn't a measure of 1% of the variability due to location reflect reality a lot better than 72%? If you are comparing the means of two dependent samples, use repeated-measures ANOVA, not a paired-samples t test.

We'll use the following example throughout the chapter to learn how to complete a repeated-measures ANOVA. Imagine that Dr. King wanted to determine if a new treatment for ADHD was effective. He took four children with ADHD and measured their level of distraction on an interval-level scale where higher scores mean a higher level of distraction. Next, each child received individual behavior therapy for three months, after which the child's level of distraction was assessed again. Finally, to determine if the effects of treatment are long-lasting, each child had his or her level of distraction measured again six months later.

Table 11.1 contains the data, each case in its own row and each phase of treatment in a separate column. The effects of treatment are found in the column means. **Figure 11.3** shows that the mean level of distraction is highest before the treatment starts, lowest at the end of treatment, and somewhere in the middle at follow-up. This suggests that treatment has an effect on reducing distraction and that the effect lingers, though in a weakened form, for six months. Of course, Dr. King needs to do a statistical test to determine if the treatment effect is statistically significant or if the differences can be accounted for by sampling error.

TABLE 11.1 Level of Distraction Scores Pre-Treatment, Post-Treatment, and at Six-Month Follow-Up for Four Children with ADHD

	Pre-Treatment	Post-Treatment	Follow-Up
Participant 1	13	10	12
Participant 2	30	20	26
Participant 3	20	13	17
Participant 4	26	17	20
<i>M (s)</i>	22.25 (7.41)	15.00 (4.40)	18.75 (5.85)

Looking at the column means, treatment improves the level of distraction from a mean of 22.25 before treatment to a mean of 15.00 after treatment. Further, some improvement lingers for at least six months: after treatment, the distraction score gets worse, but it doesn't return to pre-treatment levels.

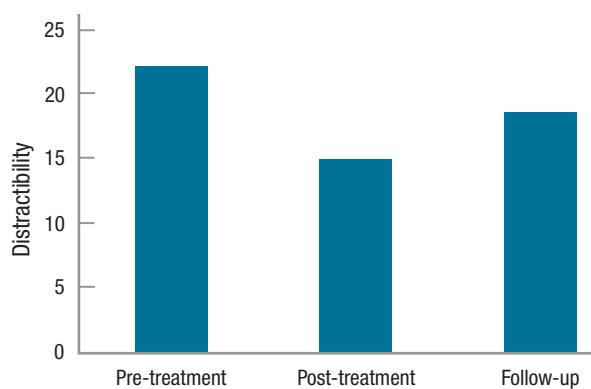


Figure 11.3 ADHD Study Pre-Treatment, Post-Treatment, and Six-Month Follow-Up Distraction Scores The mean level of distraction decreases with treatment and increases when treatment is over. To determine if these changes in level of distraction are statistically significant, a one-way, repeated-measures ANOVA needs to be completed.

The differences between rows are due to individual differences, uncontrolled characteristics that vary from case to case. For example, case 1 has the lowest level of distraction pre-treatment. This could be due to the participant's having a less severe form of ADHD, having eaten a good breakfast on the day of the test, having been taught by a teacher who stressed the skills on the test, or a whole host of other uncontrolled factors.

To reiterate the message of this introduction, the effect of individual differences is intertwined with the effect of treatment. Is the first case's low score at the end of treatment due to the effect of treatment or due to individual differences like a less severe form of ADHD? One-way, repeated-measures ANOVA is statistically powerful because it divides the variability in scores into these two factors. Repeated-measures ANOVA determines (1) how much variability is due to the column effect (the effect of treatment), and (2) how much variability is due to the row effect (individual differences). Let's see it in action in the next section.

Practice Problems 11.1

Review Your Knowledge

- 11.01** When is repeated-measures ANOVA used?
11.02 Into what two factors does one-way, repeated-measures ANOVA divide variability in a set of scores?

- 11.03** When comparing the means of two dependent samples, what test should be used? Why?

11.2 Calculating One-Way, Repeated-Measures ANOVA

Step 1 Pick a Test

Dr. King's ADHD/distraction study compares the means of three samples, so it calls for an ANOVA to analyze the data. Because there is only one explanatory variable (time), it will be a one-way ANOVA. There are three levels of the explanatory variable—pre-treatment, post-treatment, and follow-up—so there are three groups. This is a longitudinal study in which the same participants are being assessed at three points in time. This means the samples are dependent. The dependent variable (level of distraction) is measured at the interval level, so a mean for each group can be calculated. Comparing the means of three dependent samples calls for a one-way, repeated-measures ANOVA.

Step 2 Check the Assumptions

Assumptions for the one-way, repeated-measures ANOVA are similar to those for the paired-samples t test: (1) each sample is a random sample from its population, (2) cases within samples are independent and don't influence each other, and (3) the dependent variable is normally distributed in the population. There is a fourth assumption, which is beyond the scope of this book to assess. It is called the sphericity assumption, and it is a form of the homogeneity of variance assumption: If difference scores were calculated, would variability be the same in each set of difference scores?



Don't forget: Tom and Harry despise crabby infants is a mnemonic to remember the six steps of hypothesis testing.

Dr. King makes the following evaluation of the first three assumptions.

1. *Random samples.* There is no evidence that the sample is a random sample of children with ADHD, so this assumption is violated. However, the assumption is robust to violation, so the child psychologist can proceed. He will need to be careful about the population to which the results are generalized.
2. *Independence of observations.* There is no evidence that the cases within a group influence each other, so this assumption is not violated.
3. *Normality.* Dr. King assumes that distraction scores are normally distributed in the larger population of people with ADHD.

Only the random samples assumption is violated (and it's robust to violation), so Dr. King can proceed with the one-way, repeated-measures ANOVA.

Step 3 List the Hypotheses

For a one-way, repeated-measures ANOVA, the null hypothesis states that all the population means are the same. The alternative hypothesis states that at least one population mean is different from at least one other. As with between-subjects one-way ANOVA, the null hypothesis is easy to state using mathematical symbols, but not the alternative hypothesis, so words are used to express the alternative hypothesis. For the ADHD/level of distraction data with three samples, the null and alternative hypotheses are

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

$H_1:$ At least one population mean is different from at least one of the others.

Step 4 Set the Decision Rule

To set the decision rule with repeated-measures ANOVA, Dr. King needs to find F_{cv} , the critical value of F .

As with the other tests we've covered up to this point, to find F_{cv} , a researcher needs (a) to set the alpha level on the basis of willingness to make a Type I error and (b) to know the degrees of freedom for the numerator and the denominator of the F ratio. These are then used in Appendix Table 4 to find F_{cv} . (One thing a researcher doesn't need to worry about is whether to do a one-tailed or a two-tailed test—ANOVA is always nondirectional.)

In a one-way, repeated-measures ANOVA, the numerator, where the effect of the independent variable is isolated, is called the treatment effect and the denominator is called the residual effect. The formulas for finding degrees of freedom treatment ($df_{Treatment}$) and degrees of freedom residual ($df_{Residual}$) for a one-way, repeated-measures ANOVA are given in Equation 11.1. Equation 11.1 also shows how to calculate two other degrees of freedom that will be necessary to complete the ANOVA summary table for repeated-measures ANOVA: (1) degrees of freedom subjects, $df_{Subjects}$, and (2) degrees of freedom total, df_{Total} .

Equation 11.1 Degrees of Freedom for a One-Way, Repeated-Measures ANOVA

$$df_{Subjects} = n - 1$$

$$df_{Treatment} = k - 1$$

$$df_{Residual} = df_{Subjects} \times df_{Treatment}$$

$$df_{Total} = N - 1$$

where $df_{Subjects}$ = degrees of freedom for variability due to subjects

$df_{Treatment}$ = degrees of freedom for the treatment effect

$df_{Residual}$ = degrees of freedom for residual variability

df_{Total} = degrees of freedom total

n = sample size per group

k = number of groups

N = total number of observations ($n \times k$)

Dr. King sets alpha at .05. He is willing to run a 5% chance of making a Type I error. To calculate degrees of freedom for the level of distraction/ADHD data, Dr. King needs to know $n = 4$, $k = 3$, and $N = 12$. That is, there are four cases per group and three groups, for a total of 12 cases. These three values are then used in Equation 11.1 to calculate the 4 degrees of freedom:

$$\begin{aligned} df_{Subjects} &= n - 1 \\ &= 4 - 1 \\ &= 3 \end{aligned}$$

$$\begin{aligned} df_{Treatment} &= k - 1 \\ &= 3 - 1 \\ &= 2 \end{aligned}$$

$$\begin{aligned} df_{\text{Residual}} &= df_{\text{Subjects}} \times df_{\text{Treatment}} \\ &= 3 \times 2 \\ &= 6 \end{aligned}$$

$$\begin{aligned} df_{\text{Total}} &= N - 1 \\ &= 12 - 1 \\ &= 11 \end{aligned}$$

In this example, note that $3 + 2 + 6 = 11$. For one-way, repeated-measures ANOVA, $df_{\text{Subjects}} + df_{\text{Treatment}} + df_{\text{Residual}} = df_{\text{Total}}$. This means that the total degrees of freedom in a repeated-measures ANOVA are divided into subcomponents for variability due to individual differences (df_{Subjects}), due to the explanatory variable ($df_{\text{Treatment}}$), and remaining variability (df_{Residual}).

To find the critical value of F , the only degrees of freedom needed are $df_{\text{Treatment}}$ and df_{Residual} , the numerator and the denominator, respectively, of the F ratio. For the level of distraction data, those degrees of freedom are 2 and 6. Using Appendix Table 4 with $\alpha = .05$, Dr. King finds the intersection of the column for 2 degrees of freedom and the row with 6 degrees of freedom and arrives at $F_{cv} = 5.143$. Hence, the decision rule for the ADHD/level of distraction data for a one-way, repeated-measures ANOVA is:

- If $F \geq F_{cv}$ of 5.143, reject H_0 .
- If $F < F_{cv}$ of 5.143, fail to reject H_0 .

Figure 11.4 uses F_{cv} to show the sampling distribution of F with the rare and common zones marked. If the value of F that Dr. King will calculate in the next step falls in the rare zone or on the line that separates the rare zone from the common zone, then (a) the null hypothesis is rejected, (b) the alternative hypothesis is accepted, (c) the results are called statistically significant, and (d) there is reason to believe that at least one difference exists among the population means. If F falls in the common zone, then (a) Dr. King will have failed to reject the null hypothesis, (b) the results will be called not statistically significant, and (c) he concludes there is not enough evidence to conclude any difference exists among any of these population means.

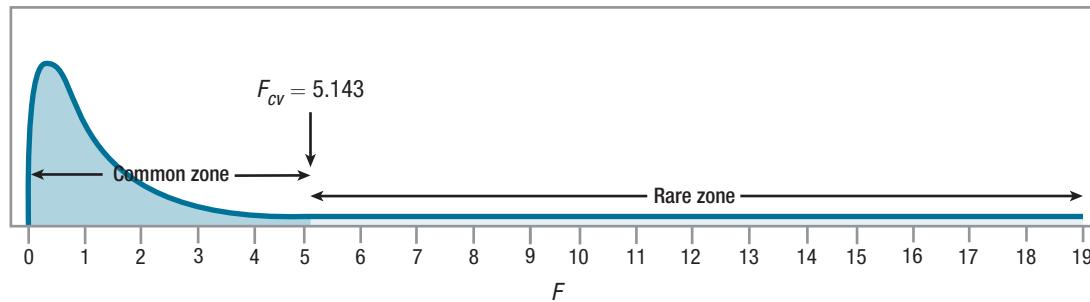


Figure 11.4 Sampling Distribution of F for ADHD Distractibility Study This is the sampling distribution of F at the .05 level with 2 and 6 degrees of freedom. If the observed value of F falls in the rare zone, then the null hypothesis is rejected.

**Step 5 Calculate the Test Statistic**

To calculate the F ratio, a researcher completes an ANOVA summary table. **Table 11.2** is a guide to completing the cells in an ANOVA summary table for one-way, repeated-measures ANOVA. Though column labels in this table are identical to the ones for between-subjects, one-way ANOVA, the sources of variability (the rows) are different.

TABLE 11.2

Template for ANOVA Summary Table for One-Way, Repeated-Measures ANOVA

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	See Chapter Appendix	$n - 1$		
Treatment	See Chapter Appendix	$k - 1$	$\frac{SS_{Treatment}}{df_{Treatment}}$	$\frac{MS_{Treatment}}{MS_{Residual}}$
Residual	See Chapter Appendix	$(n - 1)(k - 1)$	$\frac{SS_{Residual}}{df_{Residual}}$	
Total	See Chapter Appendix	$N - 1$		

The sums of squares for subjects, treatment, and residual will be provided in order to complete the ANOVA summary table. n = the number of cases per cell, k = the number of cells, and N = the total number of observations.

- *First column: Source of variability.* In the summary table, the sources of variability are listed in the first column in this order: variability due to subjects, variability due to treatment, residual variability, and total variability. Residual variability is the portion of total variability that is not accounted for by variability due to subjects or treatment.
- *Second column: Sum of squares.* Each source of variability is represented by a sum of squares in the second column. In order to focus on understanding repeated-measures ANOVA and not get slowed down by calculations, the necessary sums of squares— $SS_{Subjects}$, $SS_{Treatment}$, $SS_{Residual}$, and SS_{Total} —will be supplied in this chapter. (For those who wish to learn how to calculate sums of squares, or who have instructors who wish them to learn, formulas are given in an appendix at the end of the chapter.)
- *Third column: Degrees of freedom.* The third column shows the degrees of freedom for each source of variability. The degrees of freedom are calculated through Equation 11.1.
- *Fourth column: Mean squares.* In the fourth column, mean squares for the treatment effect and the residual effect are calculated. These are abbreviated $MS_{Treatment}$ and $MS_{Residual}$. They are calculated by dividing a sum of squares by its degrees of freedom.
- *Fifth column: Fratio.* In the fifth column, the F ratio for the effect of treatment is calculated by dividing the numerator term ($MS_{Treatment}$) by the denominator term ($MS_{Residual}$).



The sums of squares for the ADHD/distraction data follow:

- $SS_{\text{Subjects}} = 308.67$
- $SS_{\text{Treatment}} = 105.17$
- $SS_{\text{Residual}} = 16.83$
- $SS_{\text{Total}} = 430.67$

Then Dr. King retrieves the degrees of freedom he calculated earlier and enters them in the summary table:

- $df_{\text{Subjects}} = 3$
- $df_{\text{Treatment}} = 2$
- $df_{\text{Residual}} = 6$
- $df_{\text{Total}} = 11$

The next step is to calculate the two mean squares, $MS_{\text{Treatment}}$ and MS_{Residual} . To do so, Dr. King divides each sum of squares by its respective degrees of freedom:

$$MS_{\text{Treatment}} = \frac{SS_{\text{Treatment}}}{df_{\text{Treatment}}} = \frac{105.17}{2} = 52.5850 = 52.59$$

$$MS_{\text{Residual}} = \frac{SS_{\text{Residual}}}{df_{\text{Residual}}} = \frac{16.83}{6} = 2.8050 = 2.81$$

Finally, Dr. King divides the mean square for the numerator ($MS_{\text{Treatment}}$) by the mean square for the denominator (MS_{Residual}) to find the F ratio:

$$\begin{aligned} F &= \frac{MS_{\text{Treatment}}}{MS_{\text{Residual}}} \\ &= \frac{52.59}{2.81} \\ &= 18.7153 \\ &= 18.72 \end{aligned}$$

At this point, the calculation phase for one-way, repeated-measures ANOVA is over. The complete ANOVA summary table for the ADHD/distraction study is shown in **Table 11.3**. The next section of the chapter, after more practice with calculation, focuses on interpreting the results.

TABLE 11.3 ANOVA Summary Table for Level of Distraction/ADHD Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	308.67	3		
Treatment	105.17	2	52.59	18.72
Residual	16.83	6	2.81	
Total	430.67	11		

Worked Example 11.1

Imagine that Dr. Agosto, an orthopedic surgeon, decided to test the effectiveness of the standard surgical treatment for back pain. She found 30 people with chronic back pain and matched them into groups of three on the basis of age, sex, number of years of pain, and degree of physical impairment. She then randomly assigned members of each group to receive one of three different levels of treatment: (1) standard surgery and physical therapy, (2) sham surgery and physical therapy, or (3) physical therapy alone. (In sham surgery, the patient undergoes anesthesia and has an incision made in his or her back, but is sewn up without any real surgical intervention.) This was a double-blind study for the patients who received surgery—they did not know if they received standard surgery or sham surgery, and neither did the medical staff who provided follow-up care. After receiving surgery or no surgery, all patients received physical therapy for six months. At this point, the outcome was measured on a quality-of-life scale. Higher scores mean that a person has a better quality of life and is more able to enjoy the regular activities of daily living.

A Common Question

Q Is sham surgery ever really used as a control group?

A Yes. Sham surgery controls for the placebo effect of having a surgical procedure. As long as informed consent is used, there is nothing unethical about using it in a study.

Figure 11.5 shows the outcome of the experiment. Patients in Group 1, real surgery, had a better outcome than the sham surgery patients in Group 2, or the physical therapy patients in Group 3. But, the differences are slight: $M_{\text{RealSurgery}} = 51.00$, $M_{\text{ShamSurgery}} = 48.00$, and $M_{\text{PhysicalTherapy}} = 45.00$. Dr. Agosto suspects that the differences between groups could be explained with sampling error and don't represent an advantage of surgery. She'll need a statistical test to find out.

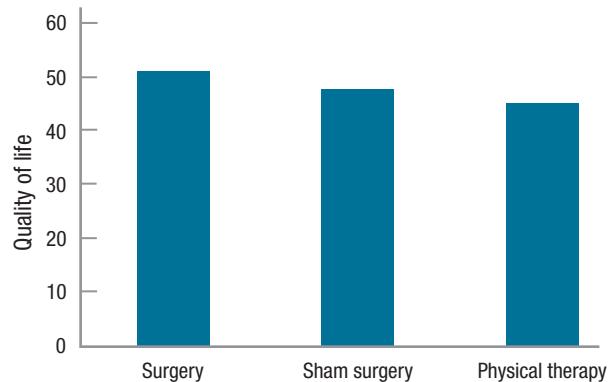


Figure 11.5 Results of Surgery, Sham Surgery, and Physical Therapy on the Quality of Life for People with Chronic Back Pain The surgery patients reported the highest mean for quality of life. However, the degree of improvement seems small and unlikely to be statistically significant. One-way, repeated-measures ANOVA can be used to determine this.

Step 1 Pick a Test

Dr. Agosto is comparing the means of three groups (real surgery vs. sham surgery vs. physical therapy). This calls for an ANOVA. There is only one independent variable (type of treatment), so this is a one-way ANOVA. The cases are made up of matched participants—each one assigned to one type of treatment, so the groups are dependent. That means Dr. Agosto is doing a one-way, repeated-measures ANOVA.

Step 2 Check the Assumptions

- *Random samples.* It was not stated that the 30 participants were a random sample from the population of people with back pain, so it is unlikely they are. Almost certainly the sample is a convenience sample and the random samples assumption has been violated. The random samples assumption is robust, though, so Dr. Agosto can proceed with the one-way, repeated-measures ANOVA. She will just need to be cautious about the population to which she generalizes the results.
- *Independence of observations.* This assumption is about independence within a group, not between groups. Within groups, there is no connection between participants and how one participant performs does not influence the performance of others. Each participant participates only once. This assumption is not violated.
- *Normality.* Dr. Agosto, from her review of the literature, knows that quality-of-life scores are normally distributed when quality of life is measured in the population of people receiving treatment for back pain.

Only the random samples assumption is violated, which is robust, so Dr. Agosto can proceed with the planned one-way, repeated-measures ANOVA.

Step 3 List the Hypotheses

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

$H_1:$ At least one of the three population means is different from at least one of the others.

Step 4 Set the Decision Rule

To determine the decision rule, Dr. Agosto needs to find the degrees of freedom for the numerator in the F ratio (the treatment effect) and for the denominator (the residual effect). Later on, to complete the ANOVA summary table, she'll need to know the other 2 degrees of freedom—subjects and total—so she might as well use Equation 11.1 to calculate all the degrees of freedom now. To do so, she'll need to know the number of cases per group ($n = 10$), the number of groups ($k = 3$), and the total number of observations ($N = 30$):

$$\begin{aligned} df_{\text{Subjects}} &= n - 1 \\ &= 10 - 1 \\ &= 9 \end{aligned}$$

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

$$\begin{aligned} df_{\text{Residual}} &= df_{\text{Subjects}} \times df_{\text{Treatment}} \\ &= 9 \times 2 \\ &= 18 \end{aligned}$$

$$\begin{aligned} df_{\text{Total}} &= N - 1 \\ &= 30 - 1 \\ &= 29 \end{aligned}$$

Knowing that $df_{\text{Treatment}}$ (the numerator degrees of freedom) = 2 and df_{Residual} (the denominator degrees of freedom) = 18, Dr. Agosto can look in Appendix Table 4 to find the critical value of F . But first, she has to decide how willing she is to make a Type I error.

A Type I error occurs when the null hypothesis is mistakenly rejected. In this study, that would mean concluding there is a treatment effect when such is not the case. Dr. Agosto wants to avoid this error, so she sets alpha at .01, not .05. This means that there is a 1% chance of this error occurring, not a 5% chance.

Using Appendix Table 4, Dr. Agosto finds $F_{cv} = 6.013$. **Figure 11.6** shows the sampling distribution of F with the rare zone marked. The decision rule is:

- If $F \geq 6.013$, reject H_0 .
- If $F < 6.013$, fail to reject H_0 .

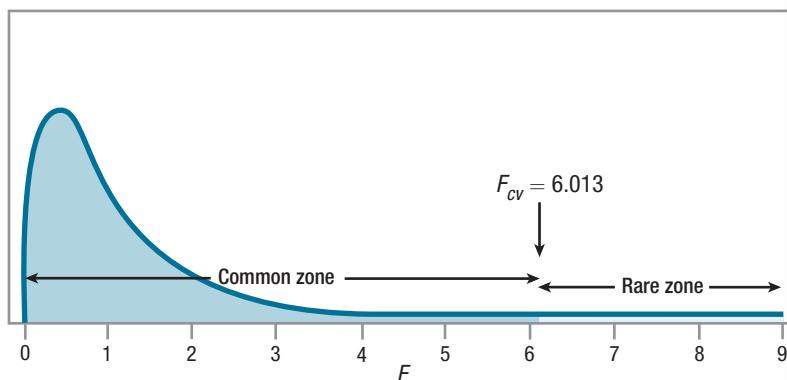


Figure 11.6 Sampling Distribution of F for Effect of Surgery and Physical Therapy on Back Pain. F_{cv} of 6.013 is the boundary between the common zone and the rare zone for this sampling distribution of F with 2 and 18 degrees of freedom and $\alpha = .01$. The null hypothesis is not rejected if the results fall in the common zone; it is rejected if results fall on the line or in the rare zone.

Step 5 Calculate the Test Statistic

To calculate the test statistic, Dr. Agosto calculated the sums of squares using the formulas provided in the appendix at the end of this chapter. She calculated:

- $SS_{\text{Subjects}} = 2,193.33$
- $SS_{\text{Treatment}} = 180.00$
- $SS_{\text{Residual}} = 50.67$
- $SS_{\text{Total}} = 2,424.00$

Next, she uses the degrees of freedom for treatment, 2, and for the residual, 18, to calculate $MS_{\text{Treatment}}$ and MS_{Residual} .

$$MS_{\text{Treatment}} = \frac{SS_{\text{Treatment}}}{df_{\text{Treatment}}} = \frac{180.00}{2} = 90.0000 = 90.00$$

$$MS_{\text{Residual}} = \frac{SS_{\text{Residual}}}{df_{\text{Residual}}} = \frac{50.67}{18} = 2.8150 = 2.82$$

Finally, to find the *F* ratio, the surgeon divides the numerator term ($MS_{\text{Treatment}}$) by the denominator term (MS_{Residual}):

$$\begin{aligned} F &= \frac{MS_{\text{Treatment}}}{MS_{\text{Residual}}} \\ &= \frac{90.00}{2.82} \\ &= 31.9149 \\ &= 31.91 \end{aligned}$$

The complete ANOVA summary table for the back surgery/quality-of-life one-way, repeated-measures ANOVA can be seen in **Table 11.4**. In the next section, we'll see how Dr. Agosto might interpret her results.

TABLE 11.4 ANOVA Summary Table for Back Surgery/Quality-of-Life Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	2,193.33	9		
Treatment	180.00	2	90.00	31.91
Residual	50.67	18	2.82	
Total	2,424.00	29		

Practice Problems 11.2

Apply Your Knowledge

- 11.04** What test should be used to analyze these two studies? Select from the single-sample *z* test; single-sample *t* test; independent-samples *t* test; paired-samples *t* test; between-subjects, one-way ANOVA; and one-way, repeated-measures ANOVA.

- High school seniors who plan to (i) not go to college, (ii) go to a community college, or (iii) go to a four-year college are matched in terms of intelligence. Ten years later, the three matched groups are compared in terms of mean annual income.
- A random sample of people with a family history of Alzheimer's disease is compared to a random sample of people

from the general population in terms of their mean score on an interval-level measure of fear of dementia.

- 11.05** List the assumptions for a one-way, repeated-measures ANOVA.
- 11.06** State the hypotheses for a one-way, repeated-measures ANOVA.
- 11.07** There's a sample of 12 cases that is measured on an interval-level variable at four points in time. The data will be analyzed with a one-way, repeated-measures ANOVA.
- What are the numerator degrees of freedom for the *F* ratio?
 - What are the denominator degrees of freedom for the *F* ratio?



■ **396** Chapter 11 One-Way, Repeated-Measures ANOVA

- 11.08** If $df_{\text{Treatment}} = 3$, $df_{\text{Residual}} = 21$, and $\alpha = .05$, what is the decision rule?
- 11.09** Given $SS_{\text{Subjects}} = 123.00$, $SS_{\text{Treatment}} = 216.00$, $SS_{\text{Residual}} = 410.40$, $SS_{\text{Total}} = 749.40$,
- $df_{\text{Subjects}} = 19$, $df_{\text{Treatment}} = 3$, $df_{\text{Residual}} = 57$, and $df_{\text{Total}} = 79$, complete a one-way, repeated-measures ANOVA summary table.

11.3 Interpreting One-Way, Repeated-Measures ANOVA

Statistical tests are used to help answer questions and make decisions. It is in the sixth step of hypothesis testing, the interpretation of results, that clear language is used to tell what was found and what it means. In the ADHD/level of distraction study, for example, repeated-measures ANOVA is being used to determine if treatment decreases the level of distraction and if the improvement is maintained when treatment stops.

For a one-way, repeated-measures ANOVA, just as with a between-subjects, one-way ANOVA, there are three tasks involved in interpretation:

1. Determine whether the null hypothesis is rejected.
2. Determine how big the *overall* effect is. That is, how much of the variability in the dependent variable is explained by the independent variable.
3. Determine where the effect is found. All a statistically significant ANOVA reveals is that at least one sample mean differs statistically from at least one other sample mean. Post-hoc tests are needed to find out which pair(s) of means differ and what the direction of the difference is.

Step 6 Interpret the Results

Was the Null Hypothesis Rejected?

As with previous tests, a researcher determines if the null hypothesis was rejected by checking the decision rule generated in Step 4 against the value of the test statistic calculated in Step 5. For the ADHD/level of distraction data, $F_{cv} = 5.143$ and $F = 18.72$. (The F value can be found in the ANOVA summary table, Table 11.3. There will be several other values in the ANOVA summary table needed for interpretation, so it is reprinted here as **Table 11.5**.)

In Step 4, the decision rule Dr. King generated was:

- If $F \geq 5.143$, reject H_0 .
- If $F < 5.143$, fail to reject H_0 .

TABLE 11.5 ANOVA Summary Table for Level of Distraction/ADHD Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	308.67	3		
Treatment	105.17	2	52.59	18.72
Residual	16.83	6	2.81	
Total	430.67	11		

This ANOVA summary table is reprinted here, so values can be taken from it to be used in interpretation.

Because $18.72 \geq 5.143$, the first statement is true and Dr. King rejects the null hypothesis. **Figure 11.7** shows that the *F* ratio fell in the rare zone. This means that the results are called statistically significant and the alternative hypothesis is accepted. The psychologist can conclude that at least one of the population means—either the pre-treatment level of distraction, or the post-treatment level of distraction, or the six-month follow-up level of distraction—probably differs from at least one other population mean. It isn't possible yet to talk about the direction of the difference because which pair(s) differs is still unknown. Discussing the direction of the difference needs to wait until post-hoc tests have been completed.

To report the analysis of variance results in APA format, Dr. King would write $F(2, 6) = 18.72, p < .05$. This gives five pieces of information:

1. What test was done (an *F* test).
2. The degrees of freedom for the numerator ($df_{\text{Treatment}} = 2$) and the denominator ($df_{\text{Residual}} = 6$) of the *F* ratio. From these, it is possible to figure out n , how many cases there are in each group by using this equation: $n = \frac{df_{\text{Residual}}}{df_{\text{Treatment}}} + 1$. For the ADHD study, this would be $n = \frac{6}{2} + 1 = 3 + 1 = 4$.
3. The observed value of the test statistic (18.72).
4. What alpha level was used (.05). (This indicates how likely a Type I error is.)
5. Whether the null hypothesis was rejected ($p < .05$) or not ($p > .05$).

How Big Is the Effect?

The next step in interpretation involves how big the effect is. This should be calculated whether or not the null hypothesis is rejected.

- If the null hypothesis is rejected, the researcher needs to calculate the size of the effect to quantify how much impact the independent variable has on the dependent variable. In the ADHD/level of distraction example, the question is to what degree time/treatment affects the level of distraction.
- If the null hypothesis is not rejected, it is still important to calculate the effect size. If the effect of the independent variable on the dependent variable is more than small, that alerts the researcher to the possibility of a Type II error. He or she will likely end up suggesting a replication with a larger sample size.

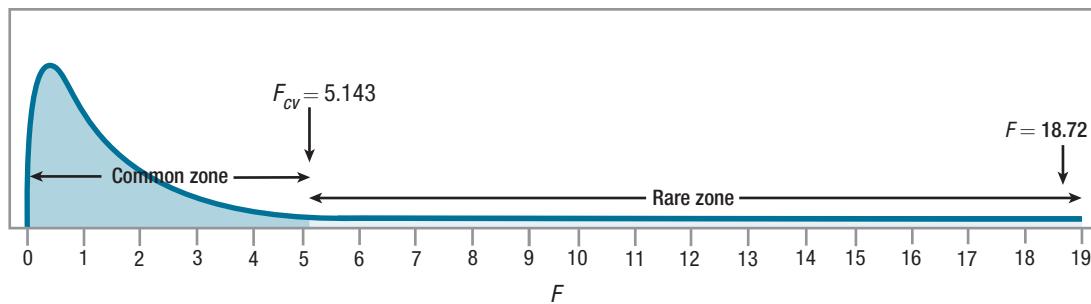


Figure 11.7 Results of ANOVA for Level of Distraction/ADHD Study This sampling distribution shows that the observed *F* value of 18.72 falls in the rare zone. The null hypothesis is rejected and the results are called statistically significant.

Whereas r^2 was used to quantify the size of the treatment effect for a between-subjects, one-way ANOVA, we will use eta squared (η^2) for a one-way, repeated-measures ANOVA. The formula is given in Equation 11.2.

Equation 11.2 Formula for Calculating Eta Squared for the Effect of Treatment for a One-Way, Repeated-Measures ANOVA

$$\eta^2 = \frac{SS_{\text{Treatment}}}{SS_{\text{Total}}} \times 100$$

where η^2 = eta squared for the treatment effect, the percentage of variability in the dependent variable that is explained by the explanatory (treatment) variable

$SS_{\text{Treatment}}$ = sum of squares treatment from the ANOVA summary table

SS_{Total} = sum of squares total from the ANOVA summary table

For the ADHD data, where it is known that $SS_{\text{Treatment}} = 105.17$ and $SS_{\text{Total}} = 430.67$ (see Table 11.5), Dr. King calculated η^2 this way:

$$\eta^2 = \frac{SS_{\text{Treatment}}}{SS_{\text{Total}}} \times 100 = \frac{105.17}{430.67} \times 100 = 0.2442 \times 100 = 24.42\%$$

Eta squared for the treatment effect in the level of distraction data is 24.42%. **Eta squared** tells the percentage of the variability in the dependent variable that is accounted for by the explanatory variable. It can range from 0% (meaning the explanatory variable has no impact on the dependent variable) to 100% (meaning the explanatory variable wholly determines the dependent variable).

For the ADHD/level of distraction data, eta squared shows how well the explanatory variable of treatment status—pre-treatment, post-treatment, or follow-up—predicts outcome as measured by the dependent variable of level of distractibility. If $\eta^2 = 0\%$, then knowing the treatment status doesn't provide any information about a child's level of distraction. If $\eta^2 = 100\%$, then knowing the treatment status precisely predicts a child's level of distraction.

Explaining 100% of the variability never happens and it is all too common to explain only a small percentage of the variability. As a guide to interpreting eta squared, use Cohen's guidelines for r^2 :

- $\eta^2 \approx 1\%$ is a small effect.
- $\eta^2 \approx 9\%$ is a medium effect.
- $\eta^2 \approx 25\%$ is a large effect.

Figure 11.8 shows graphically what small, medium, and large effect sizes look like and how much variability they leave unexplained.

As $\eta^2 = 24.42\%$ in the ADHD study, Dr. King can conclude that treatment status has a large effect on the level of distraction. It is now time to find out specifically which group—pre-treatment, post-treatment, or follow-up—has the effect.

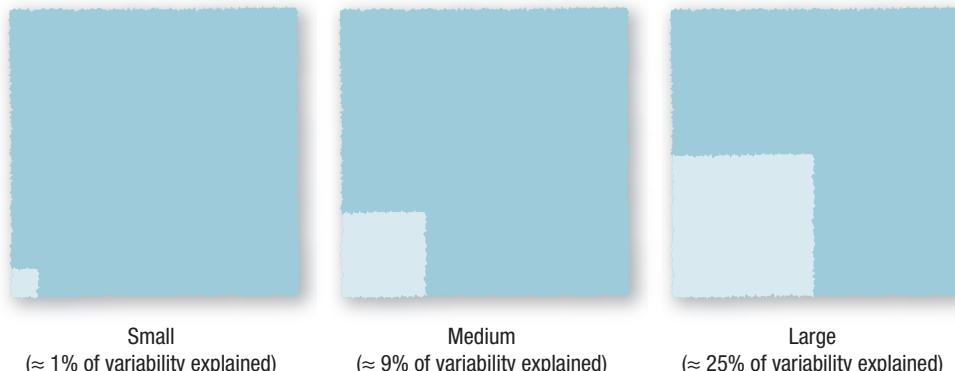


Figure 11.8 Graphic Representations of Cohen's Small, Medium, and Large Effect Sizes Cohen (1988) considers a small effect one that explains about 1% of the variability in a dependent variable. A medium effect explains about 9% and a large effect about 25%. Each figure represents all 100% of the variability that is available to be explained in a dependent variable. The shaded portions represent the percentage of variability that is explained by the different effect sizes and the unshaded portions are left unexplained.

Where Are the Effects and What Is Their Direction?

Once again, the Tukey *HSD* will be used as the post-hoc test. *HSD* stands for “honestly significant difference” and any two sample means that differ by at least the *HSD* value represent two populations that are honestly likely to have different means. The formula for the *HSD* value is given in Equation 11.3. Please remember, post-hoc tests are only performed when the results of the ANOVA are statistically significant!

Equation 11.3 Formula for Calculating Tukey *HSD* Values for Post-hoc Tests for One-Way, Repeated-Measures ANOVA

$$HSD = q \sqrt{\frac{MS_{\text{Residual}}}{n}}$$

where $HSD = HSD$ value being calculated

$q = q$ value from Appendix Table 5, where $k =$ the number of groups and $df = df_{\text{Residual}}$

$MS_{\text{Residual}} =$ mean square residual, from the ANOVA summary table

$n =$ number of cases per group

Dr. King needs three values in order to apply Equation 11.3: q , MS_{Residual} , and n .

- To find the q value, Dr. King uses Appendix Table 5. $k = 3$ as there are three groups in the ADHD study, so he looks in that column. The row is based on the residual degrees of freedom, which are 6 for the ADHD data. At the intersection of the column for $k = 3$ and the row for $df = 6$, he finds the q value, $q = 4.34$.
- In the ANOVA summary table (Table 11.3 and Table 11.5), he finds $MS_{\text{Residual}} = 2.81$.
- There are four cases in each group, so $n = 4$.

Given those three values, here are Dr. King's calculations to find the *HSD* value using Equation 11.3:

$$HSD = q\sqrt{\frac{MS_{\text{Residual}}}{n}} = 4.34\sqrt{\frac{2.81}{4}} = 4.34\sqrt{0.7025} = 4.34 \times 0.8382 = 3.6378 = 3.64$$

The Tukey *HSD* value is 3.64. Any pair of sample means for the ADHD/level of distraction study that differs by this much or more has a statistically significant difference. It doesn't matter which mean is subtracted from which in order to compare the difference to the *HSD* value, so Dr. King has arranged the comparisons to give positive numbers. The post-hoc test results are summarized in **Table 11.6** and explained in detail below.

TABLE 11.6 Post-Hoc Test Results for Mean Level of Distraction/ADHD Study

	Pre-Treatment	End of Treatment	6-Month Follow-Up
Pre-treatment	—	7.25*	3.50
End of treatment	7.25*	—	3.75*
6-month follow-up	3.50	3.75*	—

The values in the cells represent mean differences between pairs of means. Values followed by an asterisk (*) represent statistically significant differences. The dashes (—) represent comparisons that can't be made. This table contains redundant information as the values below the diagonal are the same as those above.

- Pre-treatment mean level of distraction vs. post-treatment mean level of distraction:
 - $22.25 - 15.00 = 7.25$.
 - $7.25 \geq 3.64$, so Dr. King concludes that there is a statistically significant difference between the level of distraction from pre-treatment to post-treatment.
 - The sample means can be used to comment on the direction of the difference: the *reduction* in level of distraction from pre-treatment to post-treatment is a statistically significant one. Phrased simply: treatment works.
- Pre-treatment mean level of distraction vs. six-month follow-up mean level of distraction:
 - $22.25 - 18.75 = 3.50$.
 - $3.50 < 3.64$, so this difference is not statistically significant. For this population of children with ADHD, there's not enough evidence to conclude that any difference exists in the level of distraction between the start of treatment and six months after treatment. Phrased simply: there's not enough evidence to show that treatment has a long-term impact.
 - However, be aware of two things: (1) the difference was close to being statistically significant, and (2) the sample size was small. These values alert Dr. King to the risk of making a Type II error, so he's going to end up recommending replication with a larger sample size.
- Post-treatment mean level of distraction vs. six-month follow-up mean level of distraction:
 - $18.75 - 15.00 = 3.75$.
 - $3.75 \geq 3.64$, so this difference is statistically significant. For this population of children with ADHD, one can conclude that the level of distraction is statistically

significantly higher six months after treatment than it is at the end of treatment. Phrased simply: some relapse seems to occur when treatment is over.

Putting It All Together

Here is Dr. King's four-point interpretation in which he (1) states what was done, (2) presents the main findings, (3) explains the results, and (4) makes suggestions for future research:

This study was conducted to evaluate the long-term effectiveness of individual behavior therapy in treating children with ADHD. Four children with ADHD had their levels of distraction assessed before treatment started ($M = 22.25$), at the end of treatment ($M = 15.00$), and again six months later ($M = 18.75$). Using a one-way, repeated-measures ANOVA, there was a large and statistically significant effect on distractibility as a function of when the outcome was assessed [$F(2, 6) = 18.72, p < .01$]. Post-hoc tests showed that the mean level of distraction significantly improved from pre-treatment to the end of treatment. Unfortunately, a statistically significant amount of relapse occurred over the next six months, with the mean distractibility level getting worse from the end of treatment to the six-month follow-up. The change in the mean level of distraction from pre-treatment to six-month follow-up was not statistically significant, meaning there was insufficient evidence to show that ADHD symptoms were any better six months after treatment than they were before treatment started. Individual behavior therapy appears effective in treating ADHD, but the results don't remain fully in effect for six months when treatment is over. However, the present study only had four participants and it is possible that the small sample size prevented some effects from being found. To get a better sense of the long-term impact of individual behavior therapy on distractibility, the study should be replicated with a larger sample size.

Worked Example 11.2

For practice interpreting a one-way, repeated-measures ANOVA, let's use the results from the back pain surgery study. In that study, Dr. Agosto, an orthopedic surgeon, took 30 people with chronic back pain and divided them into groups of three that were matched on age, sex, years of pain, and degree of impairment. One of the three people in each group was randomly assigned to receive real surgery, one sham surgery, and one no surgery. Each participant then received physical therapy for six months, at which point his or her quality of life was measured (with higher scores indicating better functioning). The mean was 51.00 for the 10 surgery patients, 48.00 for the sham surgery patients, and 45.00 for the physical therapy only patients. F_{cv} was 3.555 and the ANOVA summary table for the one-way, repeated-measures ANOVA is presented here again as **Table 11.7**.

TABLE 11.7 Representation of ANOVA Summary Table for Back Surgery/Quality-of-Life Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	2,193.33	9		
Treatment	180.00	2	90.00	31.91
Residual	50.67	18	2.82	
Total	2,424.00	29		

Was the Null Hypothesis Rejected?

The decision rule was:

- If $F \geq 3.555$, reject H_0 .
- If $F < 3.555$, fail to reject H_0 .

The observed value of F (see Table 11.5) is 31.91, which is greater than 3.555.

Figure 11.9 shows how the F value landed in the rare zone of the sampling distribution. As a result, the null hypothesis is rejected, the alternative hypothesis is accepted, and the results are called statistically significant. This means it is concluded that at least one of these three treatments is different from at least one other for this population of patients with back pain. In APA format, the results are written $F(2, 18) = 31.91, p < .05$.

The three means (51.00, 48.00, and 45.00) all seemed close to one another and Dr. Agosto had speculated that the ANOVA would not be statistically significant. The fact that the results are statistically significant shows the power of a repeated-measures design. By isolating the variability due to treatment from variability due to individual differences, it is easier to reject the null hypothesis with a repeated-measures design. If one had, erroneously, used a one-way, between-subjects ANOVA to analyze the ADHD data, the results would not have been statistically significant. Repeated-measures tests are more powerful than between-subjects tests.

Repeated-measures tests are more powerful than between-subjects tests.

How Big Is the Effect?

To determine effect size, η^2 , Dr. Agosto used Equation 11.2. Eta squared quantifies the percentage of variability in quality of life (the dependent variable) that is accounted for by which treatment cases receive (the independent variable.) To calculate eta squared, a researcher needs to know sum of squares treatment and sum of squares total. Looking in Table 11.7, Dr. Agosto finds $SS_{\text{Treatment}} = 180.00$ and $SS_{\text{Total}} = 2,424.00$. Substituting those into Equation 11.2 gives

$$\eta^2 = \frac{SS_{\text{Treatment}}}{SS_{\text{Total}}} \times 100 = \frac{180.00}{2,424.00} \times 100 = 0.0743 \times 100 = 7.43\%$$

An η^2 of 7.43% qualifies as a medium effect.

Where Are the Effects and What Is Their Direction?

The ANOVA was statistically significant, so Dr. Agosto will need post-hoc tests to pinpoint where the effect occurred. To use Equation 11.3 to calculate the Tukey HSD, she will need to know q , MS_{Residual} , and n .

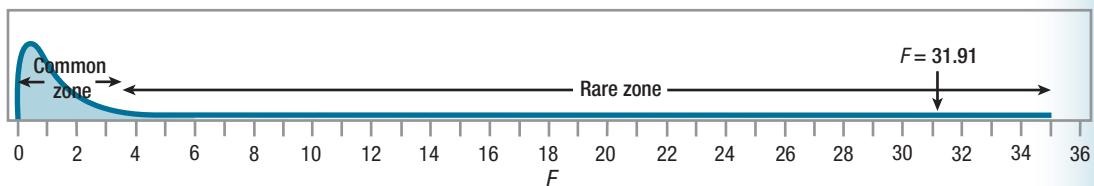


Figure 11.9 ANOVA Results for the Effect of Surgery and Physical Therapy on Back Pain The observed value of F , 31.91, falls in the rare zone of the sampling distribution. The null hypothesis is rejected and the results are called statistically significant. At least one treatment group had an outcome statistically different from at least one other.

- With $k = 3$ and $df_{\text{Residual}} = 18$, $q = 4.70$. (Note: Dr. Agosto had set alpha at .01 for the repeated-measures ANOVA to reduce the risk of Type I error. To be consistent, she is using the Appendix 5 q table with $\alpha = .01$.)
- $MS_{\text{Residual}} = 2.82$.
- $n = 10$.

Now she can calculate Tukey HSD using Equation 11.3:

$$\begin{aligned} HSD &= q \sqrt{\frac{MS_{\text{Residual}}}{n}} = 4.70 \sqrt{\frac{2.82}{10}} = 4.70 \sqrt{0.2820} = 4.70 \times 0.5310 \\ &= 2.4957 = 2.50 \end{aligned}$$

Dr. Agosto has three comparisons to make. Any pairs of means that differ by at least 2.50 points have a statistically significant difference.

- Group 1 vs. Group 2: Real surgery plus physical therapy vs. sham surgery plus physical therapy.
 - $51.00 - 48.00 = 3.00$.
 - $3.00 \geq 2.50$, so the difference is statistically significant. For this population of people with back pain, Dr. Agosto can conclude that people who receive real surgery have a statistically better outcome, in terms of quality of life, than do people who receive sham surgery.
- Group 1 vs. Group 3: Real surgery plus physical therapy vs. physical therapy alone.
 - $51.00 - 45.00 = 6.00$.
 - $6.00 \geq 2.50$, so the difference is statistically significant. For this population of people with back pain, Dr. Agosto can conclude that people who receive real surgery and physical therapy have a statistically better outcome, in terms of quality of life, than do people who receive physical therapy alone.
- Group 2 vs. Group 3: Sham surgery plus physical therapy vs. physical therapy alone.
 - $48.00 - 45.00 = 3.00$.
 - $3.00 \geq 2.50$, so the difference is statistically significant. For this population of people with back pain, Dr. Agosto can conclude that people who receive sham surgery plus physical therapy have a statistically better outcome, in terms of quality of life, than do people who receive physical therapy alone.

There's a lot here to keep track of, so Dr. Agosto uses inequality signs to keep the effects straight: Real surgery $>$ Sham surgery $>$ Physical therapy.

Putting It All Together

Here is Dr. Agosto's four-point interpretation:

I conducted a study investigating how effective surgery is as a treatment for chronic back pain. Thirty people with chronic back pain were matched into groups of three on the basis of age, sex, length of illness, and severity of illness. One member of each group was randomly assigned to receive real surgery, one

to receive sham surgery, and one to receive no surgery. Each participant then received six months of physical therapy. Using a one-way, repeated-measures ANOVA, there was a small to modest, statistically significant effect for type of treatment on quality of life [$F(2, 18) = 31.91, p < .05$]. People receiving real surgery had a statistically higher quality of life ($M = 51.00$) than people receiving sham surgery ($M = 48.00$), and people receiving sham surgery had a statistically better quality of life than people just receiving physical therapy ($M = 45.00$). These results show that though there is a placebo effect for having surgery for back pain, there is a benefit to receiving surgery that is above and beyond the placebo effect. Future research should investigate which types of back problems are helped by which type of treatment.

Practice Problems 11.3

Apply Your Knowledge

- 11.10** Given $F_{cv} = 2.866$, $df_{\text{Treatment}} = 3$, $df_{\text{Residual}} = 36$, $F = 4.36$, and $\alpha = .05$, (a) write the results in APA format. (b) Are the results statistically significant?
- 11.11** Given $SS_{\text{Treatment}} = 35.76$ and $SS_{\text{total}} = 124.64$, (a) calculate η^2 . (b) Is this considered a small, medium, or large effect?
- 11.12** Given $k = 5$, $n = 16$, $MS_{\text{Residual}} = 12.98$, and $\alpha = .05$, (a) find q and (b) calculate Tukey's HSD .
- 11.13** A software designer was curious as to how comfort with technology varied across generations. He obtained 10 families and brought in a teenager, a parent, and a grandparent from each family. Then he had each of them, individually, install a new piece of software. For each one, he timed how long

it took to click on the “Agree to Terms of Installation” button. The mean time for the teens was 1.50 seconds, for the parents 4.50 seconds, and for the grandparents 20.00 seconds. No assumptions were violated and he conducted a one-way, repeated-measures ANOVA with $\alpha = .05$. He calculated $\eta^2 = 94.53\%$ and $HSD = 2.42$. Below is the one-way, repeated-measures ANOVA summary table. Write a four-point interpretation for the results.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	33.03	9		
Treatment	1,971.67	2	985.84	219.08
Residual	81.05	18	4.50	
Total	2,085.75	29		

Application Demonstration

In the most famous example of classical conditioning, a conditioned stimulus (a bell) was repeatedly paired with an unconditioned stimulus (meat) until a dog learned to salivate to the sound of the bell. Salivation to the bell is called a conditioned response. The process of then teaching the dog *not* to salivate to the bell is called extinction. In extinction, the dog hears the bell but doesn't see the meat and eventually the bell no longer elicits salivation. If a researcher then waits a few days after extinction has occurred and rings the bell around the dog again, the dog will salivate to the sound of the bell again. This is called spontaneous recovery.

Let's imagine a researcher who wanted to see if extinction and spontaneous recovery occurred in humans with naturally occurring classical conditioning.

That is, he wanted to see if these phenomena occurred when the conditioned response hadn't been taught in a laboratory, but had naturally developed.

Dr. Brian had noticed that many people salivate when they see a lemon, so he decided to use this conditioned reaction. (Try this thought experiment: Think of someone cutting open a juicy yellow lemon, right in front of you. You watch as she takes a slice of lemon, bites into it, and grimaces a little at the sourness of it. Just thinking of this, do you feel salivation happening in your mouth? That's a conditioned response.)

Dr. Brian brought 35 participants from the experiment participation pool into his lab, one at a time, and measured how many grams of saliva each one produced in 2 minutes (as a baseline). How is salivation measured? Weigh some cotton balls, put them in a mouth for 2 minutes, and then weigh them again. The gain in weight reflects the amount of salivation.

After this, Dr. Brian had the participants smell a lemon for a 2-minute period, during which salivation was measured again. This was the first extinction trial. There was then a 3-minute break, after which the second extinction trial occurred (sniffing a lemon but not tasting it). This continued for seven more extinction trials. Salivation during extinction wasn't measured again until the tenth (and last) extinction trial. The next day, participants came back to the lab and, for the last time, sniffed a lemon for 2 minutes while salivation was measured. This was the spontaneous recovery trial.

The mean grams of saliva for the four trials are shown in [Figure 11.10](#).

- It looks as if there is a conditioned response of salivation to a lemon. The amount of saliva increased from the first baseline measurement to when a lemon is smelled on the first extinction trial.
- It looks as if extinction works. The amount of saliva measured decreases from the first extinction trial to the last trial.
- It looks as if some spontaneous recovery occurs. The amount of saliva increases from the last extinction trial to the next day.

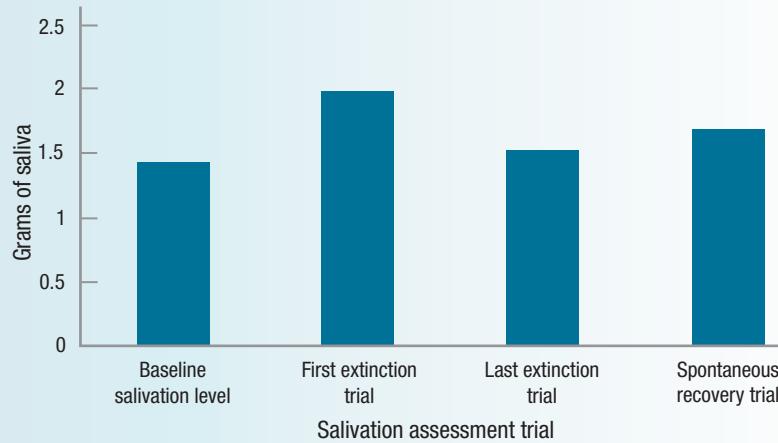


Figure 11.10 Extinction and Spontaneous Recovery of a Conditioned Response in Humans This figure appears to show that there's a conditioned response to smelling a lemon, that this conditioned response can be extinguished, and that spontaneous recovery occurs.

Of course, Dr. Brian can't know if these effects indicate an effect in the larger population until he conducts a statistical test. The first question is, "What test?"

Step 1 Pick a Test. The same participants are measured at four points in time, so the samples are dependent samples. There are more than two groups being compared on a ratio-level variable. Comparing the means of four dependent samples calls for a one-way, repeated-measures ANOVA.

Step 2 Check the Assumptions. As is common in psychological research, the random samples assumption was violated. This did not concern Dr. Brian much, because repeated-measures ANOVA is robust to violations of this assumption and because he has no reason to believe that the response of these participants to a lemon would be much different from anyone else's response to a lemon. The other two assumptions, independence of observations and normality, were not violated. So, Dr. Brian can proceed.

Step 3 List the Hypotheses.

$$H_0: \mu_{\text{Baseline}} = \mu_{\text{Lemon}} = \mu_{\text{Extinction}} = \mu_{\text{Spontaneous Recovery}}$$

H_1 : At least one of the four population means differs from at least one other population mean.

Step 4 Set the Decision Rule. First, calculate the degrees of freedom:

$$\begin{aligned} df_{\text{Subjects}} &= n - 1 \\ &= 35 - 1 \\ &= 34 \end{aligned}$$

$$\begin{aligned} df_{\text{Treatment}} &= k - 1 \\ &= 4 - 1 \\ &= 3 \end{aligned}$$

$$\begin{aligned} df_{\text{Residual}} &= df_{\text{Subjects}} \times df_{\text{Treatment}} \\ &= 34 \times 3 \\ &= 102 \end{aligned}$$

$$\begin{aligned} df_{\text{Total}} &= N - 1 \\ &= 140 - 1 \\ &= 139 \end{aligned}$$

The numerator degrees of freedom, $df_{\text{Numerator}}$, for the F ratio is $df_{\text{Treatment}}$, which is 3, and the denominator degrees of freedom, $df_{\text{Denominator}}$, is df_{Residual} , which is 102. In order to find the critical value of F , F_{cv} , Dr. Brian has to decide how willing he is to make a Type I error. He can accept having a 5% chance of a Type I error, so $\alpha = .05$. Looking in Appendix Table 4 for the column with 3 degrees of freedom and the row with 102 degrees of freedom, Dr. Brian discovers that there is no such row. In such a situation, the row with the degrees of freedom that is closest without going over is used. (Remember, this is called *The Price Is Right* rule.) This is the row with 100 degrees of freedom. $F_{cv} = 2.696$ with $\alpha = .05$. The decision rule is:

- If $F \geq 2.696$, reject H_0 .
- If $F < 2.696$, fail to reject H_0 .

Step 5 Calculate the Test Statistic. Using the formulas from the appendix to this chapter, Dr. Brian calculated $SS_{\text{Subjects}} = 83.25$, $SS_{\text{Treatment}} = 6.28$, $SS_{\text{Residual}} = 24.35$, and $SS_{\text{Total}} = 113.88$. With the sums of squares and with the degrees of freedom calculated in Step 4, Dr. Brian can complete the ANOVA summary table seen in **Table 11.8** by following the template in Table 11.2.

ANOVA Summary Table for Extinction and Spontaneous Recovery of a Naturally Occurring Conditioned Response in Humans				
Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	83.25	34		
Treatment	6.28	3	2.09	8.71
Residual	24.35	102	0.24	
Total	113.88	139		

Step 6 Interpret the Results. *Was the null hypothesis rejected?* The first interpretation question is answered by implementing the decision rule. The observed F ratio of 8.71 is greater than or equal to the critical value of F , 2.696, so the null hypothesis is rejected. The alternative hypothesis is accepted—it is probably true that at least one of the population means differs from at least one other population mean. Dr. Brian knows that there is a difference in the level of salivation between at least two of these conditions: baseline, the first extinction trial, the last extinction trial, or the spontaneous recovery trial. There may be more than two means that differ. To learn which mean or means differ will have to await post-hoc tests. For now, though, Dr. Brian can write the results in APA format:

$$F(3, 102) = 8.71, p < .05$$

How big is the effect? The second interpretation question for one-way, repeated-measures ANOVA is answered by using Equation 11.2 to calculate η^2 . Eta squared provides information about how much of the variability in the amount of salivation is explained by the four different conditions (baseline, first extinction trial, last extinction trial, and spontaneous recovery trial). Below are the calculations, resulting in $\eta^2 = 5.51\%$. Following Cohen's guidelines, Dr. Brian considers explaining 5.51% of the effect in salivation by the treatment condition to be a small to medium effect:

$$\eta^2 = \frac{SS_{\text{Treatment}}}{SS_{\text{Total}}} \times 100 = \frac{6.28}{113.88} \times 100 = .0551 \times 100 = 5.51\%$$

Where are the effects and what is their direction? This is the question Dr. Brian has been waiting to answer. With four groups, there are six possible comparisons, as shown in **Table 11.9**. In fact, arranging the comparisons this way is very helpful for a researcher thinking about the results, because different comparisons address different questions.

Dr. Brian has three questions he wants to address:

1. Is there a conditioned response of salivation to the sight and smell of a lemon? This is addressed by Comparison 1, which compares baseline salivation to the first time the participants smelled a lemon, which was the first extinction trial.

TABLE 11.9

Possible Comparisons Among the Four Groups in the Lemon Salivation Extinction Study

	First Lemon Extinction Trial	Last Lemon Extinction Trial	Spontaneous Recovery Trial
Baseline Salivation Trial	#1	#2	#3
First Extinction Trial		#4	#5
Last Extinction Trial			#6

When there are four groups, there are six unique paired comparisons.

2. Does extinction occur? This question is broken down into two subquestions. Is there a decrease in salivation from the first extinction trial to the last extinction trial (Comparison 4)? If there is a decrease in salivation, does it go all the way back to baseline levels (Comparison 2)?
3. Does spontaneous recovery occur? This also has two subquestions. Is there more salivation on the spontaneous recovery trial than at the last extinction trial (Comparison 6)? If so, does the increase go all the way back to the level of salivation at the first extinction trial (Comparison 5)?

Note that Comparison 3 is not mentioned. Dr. Brian is not interested in comparing the amount of spontaneous recovery salivation to baseline salivation, so he'll ignore it.

In order to complete the comparisons he has planned, Dr. Brian needs to find the *HSD* value for the Tukey post-hoc test. The first step in using Equation 11.3 is to find the *q* value in Appendix Table 5. With $k = 4$ and $df_{\text{Residual}} = 102$, $q = 3.74$. (Note that *The Price Is Right* rule was applied.) After he retrieves $MS_{\text{Residual}} = 0.24$ from the ANOVA summary table (Table 11.8) and notes that as he had 35 participants $n = 35$, he can complete Equation 11.3:

$$\begin{aligned}
 HSD &= q \sqrt{\frac{MS_{\text{Residual}}}{n}} \\
 &= 3.74 \sqrt{\frac{0.24}{35}} \\
 &= 3.74 \sqrt{.0069} \\
 &= 3.74 \times 0.0831 \\
 &= 0.3108 \\
 &= 0.31
 \end{aligned}$$

Any pair of means that differs by at least 0.31 grams is a statistically significant difference. Table 11.10 shows the differences between the means, with the statistically significant ones circled. Dr. Brian can now answer his three questions.

TABLE 11.10 Lemon Salivation Study: Differences Between Sample Means

	First Lemon Extinction Trial ($M = 1.99$)	Last Lemon Extinction Trial ($M = 1.53$)	Spontaneous Recovery Trial ($M = 1.68$)
Baseline Salivation Trial ($M = 1.44$)	$1.99 - 1.44 = 0.55$	$1.53 - 1.44 = 0.09$	Not calculated
First Lemon Extinction Trial ($M = 1.99$)		$1.99 - 1.53 = 0.46$	$1.99 - 1.68 = 0.31$
Last Lemon Extinction Trial ($M = 1.53$)			$1.68 - 1.53 = 0.13$

Each cell presents the difference between two sample means. Circled differences are statistically significant ones.

1. Is there a conditioned response of salivation to the sight and smell of a lemon? Yes. Smelling a lemon causes a statistically significant increase in mean salivation from baseline.
2. Does extinction occur? Yes. There is a statistically significant decrease in mean salivation from the first extinction trial to the last. In fact, by the last extinction trial, the mean amount of salivation can't be differentiated from baseline salivation, suggesting that extinction was complete. (This is one of the rare instances where nonsignificant results are informative.)
3. Does spontaneous recovery occur? There is not much evidence for spontaneous recovery. There is not statistically significantly more salivation at the spontaneous recovery trial than there was on the last extinction trial.

Putting it all together.

This study investigated whether humans show a naturally occurring response of salivation to the smell of a lemon and whether such a conditioned response can be extinguished and then spontaneously recovered. Thirty-five volunteers measured their baseline salivation and then had 10 extinction trials during which they held and smelled a lemon. Amount of salivation to a lemon was measured on the first and last extinction trials and again, at a spontaneous recovery trial, 24 hours later.

There were statistically significant differences among the means $F(3, 102) = 8.71, p < .05$. Post-hoc tests showed that there was a naturally occurring response of salivation to the sight/smell of a lemon and that this response could be extinguished. In this way a naturally occurring conditioned response behaves like an experimentally induced conditioned response. However, the naturally occurring conditioned response did not behave like an experimentally induced conditioned response in terms of spontaneous recovery. Whether naturally occurring conditioned responses really don't show spontaneous recovery, or not a long enough period of time was given for spontaneous recovery is a fruitful area for future research.

**DIY**

Do you have three grocery stores in your area: maybe a big chain grocery; a smaller, locally owned one; and a discount super-store? Any three will do. Make a list of four or five items you buy regularly—every week or month, such as milk or bread. Go to each store and record the prices of those items. Make sure you find exactly

the same items (brand and size/quantity) at each store. Then calculate the sums of squares (following the directions in the appendix to this chapter), complete an ANOVA summary table, and interpret your results. In terms of price, does it matter where you shop?

SUMMARY**Identify what repeated-measures ANOVA does.**

- Within-subjects, one-way ANOVA, called repeated-measures ANOVA, is a one-way ANOVA for dependent samples. It divides variability in a set of scores into two factors: (1) variability due to individual differences and (2) variability due to treatment. Thus, repeated-measures ANOVA provides a more pure measure of the effect of treatment than a paired-samples t test because it separates out variability due to individual differences.

Complete a one-way, repeated-measures ANOVA.

- Check the assumptions (random samples, independence of observations, normality) and form the null and alternative hypotheses.

- Calculate degrees of freedom, find the critical value of F , and set the decision rule.
- Given sums of squares, complete an ANOVA summary table.

Interpret a one-way, repeated-measures ANOVA.

- Determine if the null hypothesis was rejected and write the results in APA format.
- Calculate an effect size, eta squared.
- If the null hypothesis was rejected, complete post-hoc tests (Tukey HSD) to determine where effects are and what their direction is.
- Write a four-point interpretation (What was done? What was found? What does it mean? What suggestions are there for future research?).

KEY TERMS

eta squared (η^2) – an effect size that calculates the percentage of variability in the dependent variable accounted for by the independent variable.

repeated-measures ANOVA – a statistical test used to compare three or more dependent samples on

an interval- or ratio-level-dependent variable; also called within-subjects ANOVA, dependent-samples ANOVA, or related-samples ANOVA.

CHAPTER EXERCISES**Review Your Knowledge**

11.01 Between-subjects, one-way ANOVA extends the ___ to situations in which there are more than two independent samples.

11.02 Within-subjects, one-way ANOVA is like between-subjects, one-way ANOVA but is for ___ samples.

- 11.03** ____ ANOVA is the common name for within-subjects, one-way ANOVA.
- 11.04** Repeated-measures ANOVA is used to compare the ____ of ____ or more dependent samples.
- 11.05** The data used in repeated-measures ANOVA are arranged with cases on ____ and levels of the explanatory variable in ____.
- 11.06** In repeated-measures ANOVA, the cases should be in the ____ order in each cell.
- 11.07** Because a repeated-measures ANOVA is more ____ than a between-subjects, one-way ANOVA, there is a ____ likelihood of being able to reject the null hypothesis.
- 11.08** What is called the between-groups effect in between-subjects, one-way ANOVA is called the ____ effect in repeated-measures ANOVA.
- 11.09** Because variability due to subjects is removed from within-groups variability in repeated-measures ANOVA, the *numerator/denominator* of the *F* ratio is smaller than it is in between-subjects, one-way ANOVA.
- 11.10** If r^2 is calculated for a paired-samples *t* test, it will *over/underestimate* the effect size.
- 11.11** When comparing the means of two dependent samples, the author advocates using ____, not ____.
- 11.12** The random samples assumption says that the samples in a repeated-measures ANOVA are ____ samples from the population to which one wishes to generalize the results.
- 11.13** The random samples assumption is ____ to violation.
- 11.14** If cases within a sample influence each other's scores on the dependent variable, then the ____ assumption is ____.
- 11.15** It is often assumed that the dependent variable, if it is of a psychological attribute, is ____ in the ____.
- 11.16** "All population means are equal." This is the ____ hypothesis.
- 11.17** $\mu_1 \neq \mu_2 \neq \mu_3$ is / is not an accurate statement of the alternative hypothesis.
- 11.18** The abbreviation for the critical value of *F* is ____.
- 11.19** The null hypothesis is rejected if *F* is ____ the critical value of *F*.
- 11.20** The null hypothesis is not rejected if *F* is ____ the critical value of *F*.
- 11.21** If the results are called statistically significant, then the null hypothesis was / was not rejected.
- 11.22** If the conclusion of a repeated-measures ANOVA is that there is no evidence of a difference among population means, then the null hypothesis was / was not rejected.
- 11.23** If one was forced to accept the alternative hypothesis, then the null hypothesis was / was not rejected.
- 11.24** Values of F_{cv} depend on the numerator and denominator ____.
- 11.25** df_{num} are the numerator degrees of freedom for a one-way, repeated-measures ANOVA and df_{den} are the denominator degrees of freedom.
- 11.26** df_{num} and df_{den} are not needed to find F_{cv} for a one-way, repeated-measures ANOVA.
- 11.27** To apply Equation 11.1 to calculate all 4 degrees of freedom, one needs to know ____, ____, and ____.
- 11.28** To calculate *N*, one multiplies together ____ and ____.
- 11.29** If one adds together df_{Subjects} , $df_{\text{Treatment}}$, and df_{Residual} , this equals ____.
- 11.30** One-way, repeated-measures ANOVA divides df_{Total} into subcomponents for two different sources of ____.
- 11.31** The effect of individual differences in repeated-measures ANOVA is called the effect of ____ in the ANOVA summary table.
- 11.32** After df_{Subjects} and $df_{\text{Treatment}}$ are removed from df_{Total} , what is left is called ____.

■ 412 Chapter 11 One-Way, Repeated-Measures ANOVA

- 11.33** If the observed value of F falls in the ____ zone of the sampling distribution, the null hypothesis is rejected.
- 11.34** If F falls on the line that separates the rare zone from the common zone, the null hypothesis *is / is not* rejected.
- 11.35** The first column in an ANOVA summary table lists the sources of ____.
- 11.36** If one knows SS_{Subjects} , $SS_{\text{Treatment}}$, and SS_{Residual} , one can calculate SS ____.
- 11.37** To calculate a mean square, one divides a ____ by its ____.
- 11.38** The only mean squares one needs to calculate for a one-way, repeated-measures ANOVA are MS ____ and MS ____.
- 11.39** The *Fratio* for one-way, repeated-measures ANOVA is ____ divided by ____.
- 11.40** The first question to be addressed in an interpretation of a one-way, repeated-measures ANOVA is ____.
- 11.41** By implementing the ____ from Step 4, one can determine if the results of a statistical test are statistically significant.
- 11.42** If one knows that the null hypothesis for a one-way, repeated-measures ANOVA was rejected, one *does / does not* know which pairs of sample means had statistically significant differences.
- 11.43** Results were reported in APA format as $F(2, 12) = 7.89, p < .05$. The null hypothesis *was / was not* rejected.
- 11.44** Results were reported in APA format as $F(2, 22) = 0.89, p > .01$. The alpha level was ____.
- 11.45** If the null hypothesis is not rejected, one *should / should not* calculate an effect size.
- 11.46** Eta squared provides information about the size of the ____.
- 11.47** A researcher ends up suggesting replication of study with a larger sample size. The null hypothesis probably *was / was not* rejected.
- 11.48** If a researcher suggests replication with a larger sample size, then he or she is probably worried about Type ____ error.
- 11.49** The effect size used with one-way, repeated-measures ANOVA is ____.
- 11.50** Eta squared tells the percentage of variability in the ____ variable that is explained by the ____ variable.
- 11.51** Eta squared, like ____, ranges from 0% to 100%.
- 11.52** The closer eta squared is to ____%, the bigger the effect.
- 11.53** If the treatment effect explained 10% of the variability in a set of scores, Cohen would consider this a ____ effect.
- 11.54** The ____ is a post-hoc test for use with one-way, repeated-measures ANOVA.
- 11.55** Post-hoc tests for ANOVA should only be calculated when the null hypothesis is ____.
- 11.56** To calculate an *HSD* value, one needs to find a ____ value in Appendix Table 5.
- 11.57** If a pair of sample means differ by less than the *HSD* value, the difference is considered ____ significant.
- 11.58** If a difference in a post-hoc test is statistically significant, the two sample means can be used to determine the ____ of the difference for the two population means.

Apply Your Knowledge

Selecting the appropriate statistical test.

(For 11.59–11.62, select from the *single-sample z test; single-sample t test; independent-samples t test; paired-samples t test; between-subjects, one-way ANOVA; and one-way, repeated-measures ANOVA*.)

- 11.59** People who traveled between Philadelphia and New York City by different vehicles (train, car, bus, or plane) were surveyed to see how pleasant the experience had been. Pleasantness was measured on an interval scale. What statistical test should be used to analyze these data to see if different modes of travel were associated with different mean levels of pleasantness?

- 11.60** First-year college students were surveyed about how much they liked their roommates (a) within five minutes of meeting them, (b) after the first week of classes, and (c) at the end of the semester. An interval measure of liking was used. What statistical test should be used to see if the mean degree of liking changed over the course of the semester?
- 11.61** When making purchases with cash, some people drop their pennies in the “penny cup” next to the cash register and some don’t. A social psychologist wondered if those who did were more altruistic. She obtained a sample of people who dropped their pennies into penny cups and a sample who didn’t. To each person, she administered an interval-level measure of altruism. What statistical test should she use to see if the groups differ on the mean level of altruism?
- 11.62** A recreational therapist knows the U.S. population mean and standard deviation for scores on an interval-level risk-taking scale. She obtains a random sample of people who enjoy riding roller coasters at amusement parks and has them complete the risk-taking scale. What statistical test should she use to see if roller coaster riders differ in the mean level of sensation-seeking from the general population?
- 11.63** The dean of retention at a college wanted to find out if academic problems or social problems caused students to drop out at her college. She took a random sample of first-year students at her college and, on the basis of SAT scores, matched them into groups of three. One member of each group was then randomly assigned to be in (a) the control group, (b) the academic enhancement group, or (c) the social enhancement group. Nothing was done to the control participants. The academic enhancement participants met together as a group 10 times over the course of the semester, to cover study skills, time management, test anxiety, and so on. The social enhancement participants also met together as a group 10 times over the course of the semester, though the focus was on social skills, dating, alcohol safety, and the like. At the end of the semester, she compared the three group means on the interval-level Adjustment to College Inventory.
- For each of the three assumptions for a one-way, repeated-measures ANOVA, determine if it was violated.
 - Decide if the dean can proceed with the one-way, repeated-measures ANOVA.
- 11.64** A clinical psychologist wanted to compare three treatments for generalized anxiety disorder (GAD). She put an ad in the local paper to find people with GAD. She matched them into groups of three, based on severity of symptoms, and randomly assigned each of the matched cases to one of the three treatments. Treatment was administered individually. Outcome was assessed at the end of treatment using an interval-level measure of generalized anxiety.
- For each of the three assumptions for a one-way, repeated-measures ANOVA, determine if it was violated.
 - Decide if the psychologist can proceed with the one-way, repeated-measures ANOVA.

Stating the hypotheses for a one-way, repeated-measures ANOVA

- 11.65** a. What is the null hypothesis for a one-way, repeated-measures ANOVA?
 b. What is the alternative hypothesis for a one-way, repeated-measures ANOVA?
- 11.66** a. What is the null hypothesis for Exercise 11.64?
 b. What is the alternative hypothesis for Exercise 11.64?

Calculating degrees of freedom for one-way, repeated-measures ANOVA

- 11.67** If $k = 3$, $n = 8$, and $N = 24$, what are df_{Subjects} , $df_{\text{Treatment}}$, df_{Residual} , and df_{Total} ?
- 11.68** If $k = 5$, $n = 20$, and $N = 100$, what are df_{Subjects} , $df_{\text{Treatment}}$, df_{Residual} , and df_{Total} ?
- 11.69** If $k = 4$ and $n = 15$, what are df_{Subjects} , $df_{\text{Treatment}}$, df_{Residual} , and df_{Total} ?

■ 414 Chapter 11 One-Way, Repeated-Measures ANOVA

- 11.70** If $k = 3$ and $n = 30$, what are df_{Subjects} , $df_{\text{Treatment}}$, df_{Residual} , and df_{Total} ?

Setting the decision rule for one-way, repeated-measures ANOVA (assume $\alpha = .05$)

- 11.71** Given $df_{\text{Treatment}} = 2$ and $df_{\text{Residual}} = 10$, (a) what is F_{cv} ? (b) Draw the sampling distribution of F , being sure to label the common zone, the rare zone, and F_{cv} .

- 11.72** Given $df_{\text{Treatment}} = 3$ and $df_{\text{Residual}} = 36$, (a) what is F_{cv} ? (b) Draw the sampling distribution of F , being sure to label the common zone, the rare zone, and F_{cv} .

- 11.73** If $df_{\text{Treatment}} = 3$ and $df_{\text{Residual}} = 15$, state the decision rule.

- 11.74** If $df_{\text{Treatment}} = 2$ and $df_{\text{Residual}} = 76$, state the decision rule.

Calculating F

- 11.75** Given the following, complete an ANOVA summary table for a one-way, repeated-measures ANOVA: $df_{\text{Subjects}} = 11$, $df_{\text{Treatment}} = 2$, $df_{\text{Residual}} = 22$, $df_{\text{Total}} = 35$, $SS_{\text{Subjects}} = 137.50$, $SS_{\text{Treatment}} = 48.48$, $SS_{\text{Residual}} = 115.50$, and $SS_{\text{Total}} = 301.48$.

- 11.76** Given the following, complete an ANOVA summary table for a one-way, repeated-measures ANOVA: $df_{\text{Subjects}} = 19$, $df_{\text{Treatment}} = 3$, $df_{\text{Residual}} = 57$, $df_{\text{Total}} = 79$, $SS_{\text{Subjects}} = 101.27$, $SS_{\text{Treatment}} = 99.81$, $SS_{\text{Residual}} = 957.48$, and $SS_{\text{Total}} = 1,158.56$.

Rejecting the null hypothesis?

- 11.77** If $F_{cv} = 3.259$ and $F = 3.259$, was H_0 rejected?

- 11.78** If $F_{cv} = 2.310$ and $F = 1.96$, was H_0 rejected?

- 11.79** Using the information in this ANOVA summary table, determine if the null hypothesis is rejected if alpha is set at .05:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	5.00	20		
Treatment	5.00	2	2.50	5.00
Residual	20.00	40	0.50	
Total	30.00	62		

- 11.80** Using the information in this summary table, determine if the null hypothesis is rejected if alpha is set at .05:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	12.50	17		
Treatment	5.00	3	1.67	1.78
Residual	48.00	51	0.94	
Total	65.50	71		

Writing results in APA format

- 11.81** Given $\alpha = .05$, $df_{\text{Treatment}} = 5$, $df_{\text{Residual}} = 45$, $F = 7.84$, and $F_{cv} = 2.579$, (a) write the results in APA format, and (b) state if the results are statistically significant or not.

- 11.82** Given $\alpha = .05$, $df_{\text{Treatment}} = 3$, $df_{\text{Residual}} = 24$, $F = 2.76$, and $F_{cv} = 3.009$, (a) write the results in APA format, and (b) state if the results are statistically significant or not.

- 11.83** Given $\alpha = .05$ and the information in this summary table, (a) write the results in APA format, and (b) state if the results are statistically significant or not:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	88.00	9		
Treatment	60.00	2	30.00	11.24
Residual	48.00	18	2.67	
Total	196.00	29		

- 11.84** Given $\alpha = .05$ and the information in this summary table, (a) write the results in APA format, and (b) state if the results are statistically significant or not:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	120.00	21		
Treatment	32.00	2	16.00	1.00
Residual	674.00	42	16.05	
Total	826.00	65		

Interpreting results

- 11.85** Given the results of a one-way, repeated-measures ANOVA in APA format, $F(2, 40) = 1.52, p > .05$, interpret them based only on this.
- 11.86** Given the results of a one-way, repeated-measures ANOVA in APA format, $F(3, 36) = 12.56, p < .05$, interpret them based only on this.

Calculating effect size

- 11.87** a. Given $SS_{\text{Treatment}} = 9.89$ and $SS_{\text{Total}} = 86.98$, what is η^2 ?
b. Classify it as small, medium, or large.
- 11.88** a. Given $SS_{\text{Treatment}} = 45.55$ and $SS_{\text{Total}} = 1,893.44$, what is η^2 ?
b. Classify it as small, medium, or large.
- 11.89** a. Given the information in this summary table, what is η^2 ?
b. Classify it as small, medium, or large.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	76.00	14		
Treatment	46.00	2	23.00	3.01
Residual	214.00	28	7.64	
Total	336.00	44		

- 11.90** a. Given the information in this summary table, what is η^2 ?
b. Classify it as small, medium, or large.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	233.00	16		
Treatment	111.00	2	55.50	5.35
Residual	332.00	32	10.38	
Total	676.00	50		

Worrying about Type II error

- 11.91** If the results of a one-way, repeated-measures ANOVA are $F(2, 42) = 0.63, p > .05$ and if $\eta^2 = 2.12\%$, (a) how worried should the researcher be about having made a Type II error? (b) Should he or she recommend replication with a larger sample size?

- 11.92** If the results of a one-way, repeated-measures ANOVA are $F(2, 58) = 3.00, p > .05$ and if $\eta^2 = 8.11\%$, (a) how worried should the researcher be about having made a Type II error? (b) Should he or she recommend replication with a larger sample size?

Finding q

- 11.93** If $k = 3$ and $df_{\text{Residual}} = 16$, what is q if $\alpha = .01$?
11.94 If $k = 4$ and $df_{\text{Residual}} = 87$, what is q if $\alpha = .05$?

Calculating HSD

- 11.95** If $q = 2.90$, $MS_{\text{Residual}} = 12.27$, and $n = 8$, what is HSD ?
11.96 If $q = 3.70$, $MS_{\text{Residual}} = 122.98$, and $n = 30$, what is HSD ?

Interpreting HSD

- 11.97** If $M_1 = 12.83, M_2 = 14.98, M_3 = 8.22$, and $HSD = 3.78$, (a) determine for each possible pair of means if the difference is statistically significant, and (b) comment on the direction of the difference for the populations.

- 11.98** If $M_1 = 115.54, M_2 = 98.98, M_3 = 118.22$, and $HSD = 14.78$, (a) determine for each possible pair of means if the difference is statistically significant, and (b) comment on the direction of the difference for the populations.

Writing a complete interpretation

- 11.99** Researchers from PETA and NIH collaborated on a study to examine the effect, on rats, of being reared in laboratory settings. Sets of 3 rats from 8 litters were randomly assigned to three conditions: (1) being reared in a standard laboratory setting, (2) being reared in an enriched laboratory setting, and (3) being reared in a setting that mimics the wild (e.g., the rats have to forage for their own food; food is occasionally sparse and of poor quality). When the rats reached adulthood, they were given a series of behavioral tasks from which their rat IQ (RIQ) scores were calculated. RIQ scores are interval-level and are scored like human IQ scores. The three means were 95.00 (normal lab), 108.00



■ **416** Chapter 11 One-Way, Repeated-Measures ANOVA

(enriched lab), and 94.00 (mimicked wild). No nonrobust assumptions were violated and a one-way, repeated-measures ANOVA was completed with $\alpha = .05$. The results are shown below. F_{cv} was 3.739, η^2 was calculated as 37.14%, and HSD was found to be 9.32. Write a four-point interpretation.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	941.33	7		
Treatment	976.00	2	488.00	9.61
Residual	710.67	14	50.76	
Total	2,628.00	23		

- 11.100** A women's studies professor was curious about the long-term effect, on men, of taking a women's studies class. She thought that it would make them more open-minded and less sexist. She obtained a random sample of 10 first-year male students at her university and assigned them to take Introduction to Women's Studies. She administered a scale that measures sexist beliefs before the class started, again at the end of the semester, and again 10 years later. The scale measures at the interval level and has a mean of 50, with a standard deviation of 10. Scores above 50 indicate a person has more sexist beliefs than average; scores below 50 indicate a person has fewer sexist beliefs than average. Here are the mean scores for her 10 participants: 54.00 (pre-class), 46.00 (post-class), 38.00 (10 years later). No assumptions were violated and a one-way, repeated-measures ANOVA, with the alpha set at .05, was completed. Results are shown below. $F_{cv} = 3.555$, $\eta^2 = 33.21\%$, $HSD = 5.10$. Given this information, complete a four-point interpretation.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	2,215.33	9		
Treatment	1,280.00	2	640.00	32.11
Residual	358.67	18	19.93	
Total	3,854.00	29		

Completing all six steps of hypothesis testing

- 11.101** A psychologist studying addictions investigated the effectiveness of three different treatments for alcoholism on the number of strong urges to drink alcohol that were experienced at the end of treatment. He compared three treatments: (1) Alcoholics Anonymous, (2) individual psychotherapy, and (3) a medication that is supposed to reduce urges. He matched 30 alcoholics into groups of three based on the severity of their addiction and then randomly assigned them to the three different treatments. (Each person assigned to Alcoholics Anonymous attended a different group.) At the end of treatment, he had the participants keep a diary of how many strong urges to drink alcohol they experienced each day. Here are the results:

	Alcoholics Anonymous	Individual Psychotherapy	Medication
M	8.00	6.00	6.00
s	5.42	4.52	4.37

Given $SS_{\text{Subjects}} = 532.00$, $SS_{\text{Treatment}} = 26.67$, $SS_{\text{Residual}} = 88.00$, and $SS_{\text{Total}} = 646.67$, complete the appropriate statistical test and interpret the results. (Don't forget to follow all six steps of the hypothesis test.)

- 11.102** A psychologist teaching introductory psychology wanted to demonstrate the effects of mere exposure on liking. On the first day of class, each of her 20 students went to a private booth to view a very abstract piece of modern art. Each was asked to rate how much he or she liked it on a scale ranging from -10 (extremely dislike) to +10 (extremely like). She followed the same procedure once a week for the rest of the semester. At the middle, and again at the end, of the semester, she had the students individually rate their liking of the picture. Here are the results:

	First Class	Mid-Semester	Last Class
M	-6.30	-2.40	1.70
s	2.85	2.35	1.75

Given $SS_{\text{Subjects}} = 176.67$, $SS_{\text{Treatment}} = 640.13$, $SS_{\text{Residual}} = 140.53$, and $SS_{\text{Total}} = 957.33$, complete the appropriate statistical test and interpret the results. (Don't forget to follow all six steps of the hypothesis test.)

Expand Your Knowledge

- 11.103** If the results of a one-way, repeated-measures ANOVA are $F(3, 45) = 12.56$, $p < .05$, what is N ?
- 11.104** Is it possible, in a one-way, repeated-measures ANOVA, for $df_{\text{Treatment}}$ to be 2 and df_{Residual} to be 75?
- 11.105** If the results of a one-way, repeated-measures ANOVA are $F(3, 45) = 12.50$, $p < .05$, and $\eta^2 = 28.30$, how concerned should one be about Type II error?

- 11.106** A group of students had their IQs measured at the start, middle, and end of a school year. The three means, respectively, were 125.5, 127.5, and 130.5. A one-way, repeated-measures ANOVA was used to analyze the data and the results were $F(2, 18) = 0.12$, $p > .05$. Use the summary table below to conduct post-hoc tests as appropriate.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Subjects	4,917.50	9		
Treatment	126.67	2	63.34	0.12
Residual	9,590.00	18	532.78	
Total	14,634.17	29		

- 11.107** Which *HSD* value will be larger: one for $\alpha = .01$ or one for $.05$? Why?

SPSS

Data entry in SPSS for one-way, repeated-measures ANOVA is similar to data entry for a paired-samples *t* test—each case has a row to itself and the repeated measures each have a column. **Figure 11.11** shows the data from the ADHD/level of distraction study in the SPSS Data Editor.

	V1	pre	post	f_u
1	1	13	10	12
2	2	30	20	26
3	3	20	13	17
4	4	26	17	20

Figure 11.11 Data Entry in SPSS for One-Way, Repeated-Measures ANOVA Data for a repeated-measures ANOVA in SPSS are arranged with each case on a separate line and each level of the independent variable in a separate column. (Source: SPSS)

The repeated-measures ANOVA commands are not easy to find in SPSS. They are found under “Analyze” and then “General Linear Model” (see **Figure 11.12**).

Clicking on “Repeated Measures” opens up a new dialog box, seen in **Figure 11.13**. “Treatment” is named as the “Within-Subject Factor.” There are three levels of treatment, so the “Number of Levels” is “3.”

After clicking on the “Add” button (see Figure 11.13), the commands seen in **Figure 11.14** open by clicking the “Define” button. Notice that two of the three levels of the “Within-Subjects Variables” are already defined (the variables “pre” and “post”). The third level, “f_u,” is highlighted prior to using the arrow key to send it over.

■ 418 Chapter 11 One-Way, Repeated-Measures ANOVA

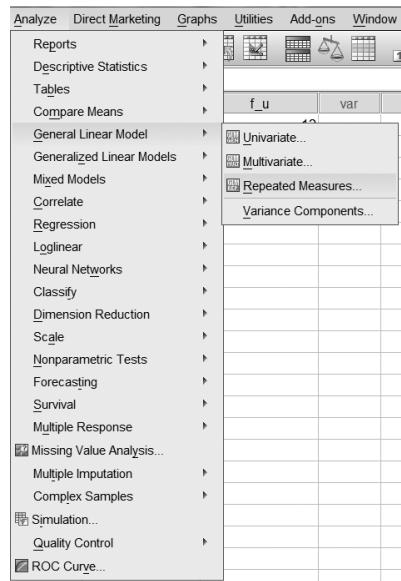


Figure 11.12 Finding Repeated-Measures ANOVA in SPSS The commands for conducting a repeated-measures ANOVA in SPSS are initiated within “General Linear Model.” (Source: SPSS)

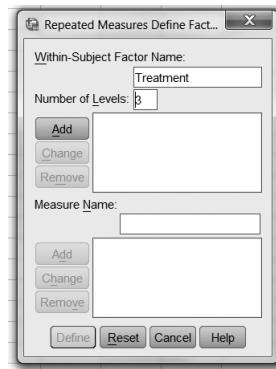


Figure 11.13 Defining the Within-Subjects Factor in SPSS In this step, the effect is given a name and the number of levels of the independent variable is indicated. (Source: SPSS)

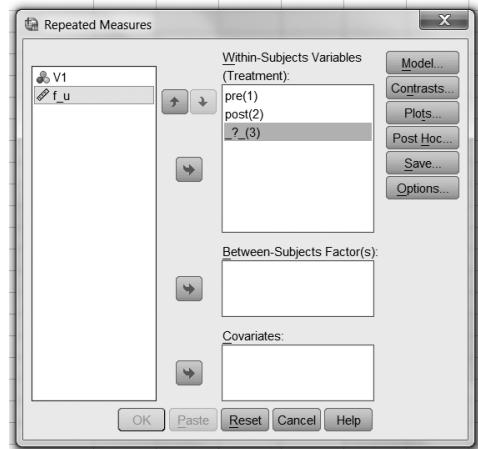


Figure 11.14 Defining the Within-Subjects Variables for a Repeated-Measures ANOVA in SPSS Once the third level of the within-subjects variable, “follow,” is moved over to the box where the other two are, the “OK” button will become active and the ANOVA can be completed. (Source: SPSS)

When all three variables have been defined, the “OK” button on the bottom right becomes active. Clicking the OK button causes SPSS to complete the analysis. The printout that SPSS generates for a repeated-measures ANOVA is fairly detailed, but most of the data is not relevant for introductory statistics. **Figure 11.15** shows the printout that is most similar to the ANOVA summary table found in Table 11.5.

Tests of Within-Subjects Effects						
Measure: MEASURE_1		Type III Sum of Squares	df	Mean Square	F	Sig.
Treatment	Sphericity Assumed	105.167	2	52.583	18.743	.003
	Greenhouse-Geisser	105.167	1.314	80.053	18.743	.011
	Huynh-Feldt	105.167	1.930	54.485	18.743	.003
	Lower-bound	105.167	1.000	105.167	18.743	.023
Error(Treatment)	Sphericity Assumed	16.833	6	2.806		
	Greenhouse-Geisser	16.833	3.941	4.271		
	Huynh-Feldt	16.833	5.791	2.907		
	Lower-bound	16.833	3.000	5.611		

Figure 11.15 ANOVA Summary Table for One-Way, Repeated-Measures ANOVA in SPSS The *F* ratio for the effect of treatment is displayed on the first line, the line for the effect with “Sphericity Assumed.” (Source: SPSS)

In this summary table, SPSS reports only two sources of variability, the one called treatment and the one called residual. Here, treatment is called “treatment” because it was labeled that way in Figure 11.13. Residual is called “error.” For a basic repeated-measures ANOVA, only pay attention to the first row (of the four rows) for each source, the one labeled “Sphericity Assumed.” The first rows are the only lines with integer values for the degrees of freedom. (The slight differences between values calculated in the text and by SPSS are due to the number of decimal places carried.)

Remember, SPSS reports exact significance levels. Here, the significance level for the F ratio of 18.743 is .003. And $.003 \leq .05$, so the results are statistically significant. APA format prefers the use of exact significance levels, so the results would be reported as $F(2, 6) = 18.74, p = .003$.

Appendix

Calculating Sums of Squares for One-Way Repeated-Measures ANOVA

Calculating sums of squares for a one-way repeated-measures ANOVA is quite similar to calculating sums of squares for a between-subjects, one-way ANOVA. Here are the pre-treatment, post-treatment, and follow-up distraction scores, squared and summed just as they were in Chapter 10 for a between-subjects, one-way ANOVA:

	Pre-Treatment		Post-Treatment		Follow-Up			
	X	X ²	X	X ²	X	X ²		
Case 1	13	169	10	100	12	144		
Case 2	30	900	20	400	26	676		
Case 3	20	400	13	169	17	289	Grand	
Case 4	26	676	17	289	20	400	X	X ²
Sum	89.00	2,145.00	60.00	958.00	75.00	1,509.00	224.00	4,612.00
n	4		4		4		12	

Once this is done, the values can be plugged into Equations 10.2, 10.3, and 10.4. SS_{Total} is calculated using Equation 10.2:

$$\begin{aligned} SS_{\text{Total}} &= \Sigma X^2 - \frac{(\Sigma X)^2}{N} \\ &= 4,612.00 - \frac{224.00^2}{12} \\ &= 430.67 \end{aligned}$$

Next, we obtain SS_{Between} via Equation 10.3. This is called SS_{Treatment} for repeated-measures ANOVA.

$$\begin{aligned} SS_{\text{Treatment}} &= \sum \left(\frac{(\Sigma X_{\text{Group}})^2}{n_{\text{Group}}} \right) - \frac{(\Sigma X)^2}{N} \\ &= \left(\frac{89^2}{4} + \frac{60^2}{4} + \frac{75^2}{4} \right) - \frac{224^2}{12} \\ &= 105.17 \end{aligned}$$

■ 420 Chapter 11 One-Way, Repeated-Measures ANOVA

Then, using Equation 10.4, calculate SS_{Within} :

$$\begin{aligned} SS_{\text{Within}} &= \sum \left(\Sigma X_{\text{Group}}^2 - \frac{(\Sigma X_{\text{Group}})^2}{n_{\text{Group}}} \right) \\ &= \left(2145.00 - \frac{89.00^2}{4} \right) + \left(958.00 - \frac{60.00^2}{4} \right) + \left(1509.00 - \frac{75.00^2}{4} \right) \\ &= 325.50 \end{aligned}$$

So far, all has been the same as for a between-subjects, one-way ANOVA. Now, it is time to diverge and calculate SS_{Between} . To do so, we need to add two new columns to the data table:

	Pre-Treatment		Post-Treatment		Follow-Up			Case Total	T^2/k
	X	X^2	X	X^2	X	X^2			
Case 1	13	169	10	100	12	144		35	408.33
Case 2	30	900	20	400	26	676		76	1,925.33
Case 3	20	400	13	169	17	289	Grand		833.33
Case 4	26	676	17	289	20	400	X	X^2	1,323.00
Sum	89.00	2,145.00	60.00	958.00	75.00	1,509.00	224.00	4,612.00	224.00
n	4		4		4		12		

The two new columns are on the far right. The first, labeled “Case Total,” is the sum of all the scores for each case. For case 1, this is $T = 13 + 10 + 12$. The next column is that case total squared and then divided by k , the number of conditions. For case 1, this is $35^2/3$. Note that all of these values are summed. SS_{Subjects} is calculated:

$$\begin{aligned} SS_{\text{Subjects}} &= \sum \left(\frac{T^2}{k} \right) - \frac{(\sum X)^2}{N} \\ &= 4490.00 - \frac{224.00^2}{12} \\ &= 308.67 \end{aligned}$$

Finally, SS_{Residual} is needed. SS_{Residual} is what remains after SS_{Subjects} is removed from SS_{Within} :

$$\begin{aligned} SS_{\text{Residual}} &= SS_{\text{Within}} - SS_{\text{Subjects}} \\ &= 325.50 - 308.67 \\ &= 16.83 \end{aligned}$$