CHAPTER 12

Hypothesis Testing With Three or More Population Means

Analysis of Variance

In Chapter 11, you learned how to determine whether a two-class categorical variable exerts an impact on a continuous outcome measure: This is a case in which a two-population *t* test for differences between means is appropriate. In many situations, though, a categorical independent variable (IV) has more than two classes. The proper hypothesis-testing technique to use when the IV is categorical with three or more classes and the dependent variable (DV) is continuous is **analysis of variance (ANOVA)**.

As its name suggests, ANOVA is premised on variance. Why do we care about variance when testing for differences between means? Consider the hypothetical distributions displayed in Figure 12.1. The distributions have the same mean but markedly disparate variances—one curve is wide and flat, indicating substantial variance, whereas the other is tall and thin, indicating relatively little variance.

In any analysis of differences between means, the variance associated with each mean must be accounted for. This is what ANOVA does. It combines means and variances into a single test for significant differences between means. A rejected null indicates the presence of a relationship between an IV and a DV.

You might be wondering why, if we have a categorical IV and a continuous DV, we do not just use a series of *t* tests to find out if one or more of the means are different from the others. **Familywise error** is the primary reason that this is not a viable analytic strategy. Every time that you run a *t* test, there is a certain probability that the null is true (i.e., that there is no relationship between the IV and the DV) but will be rejected erroneously. This probability, as we saw in Chapter 9, is alpha, and the mistake is called a Type I error. Alpha (the probability of incorrectly rejecting a true null) attaches to each *t* test, so, in a series of *t* tests, the Type I error rate increases exponentially until the likelihood of mistake reaches an unacceptable level. This is the familywise error rate, and it is the reason that you should not run multiple *t* tests on a single sample.

Another problem is that multiple *t* tests get messy. Imagine a categorical IV with classes A, B, C, and D. You would have to

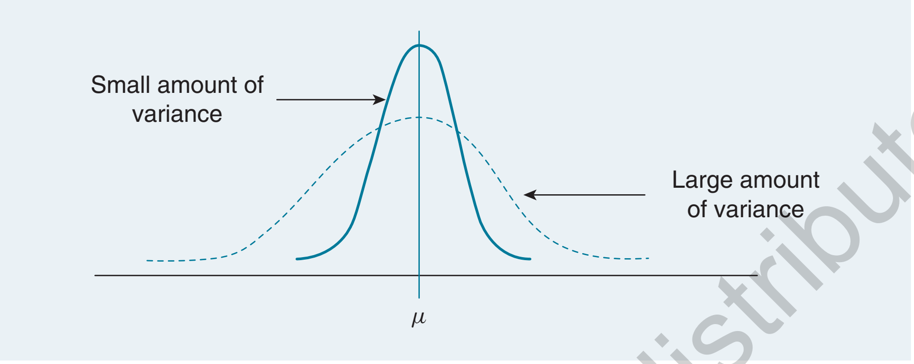
Learning Objectives

- Identify situations in which, based on the levels of measurement of the independent and dependent variables, analysis of variance is appropriate.
- Explain between- and within-group variances and how they can be compared to make a judgment about the presence or absence of group effects.
- Explain the *F* statistic conceptually.
- Explain what the null and alternative hypotheses predict.
- Use raw data to solve equations and conduct five-step hypothesis tests.
- Explain measures of association and why they are necessary.
- Use SPSS to run analysis of variance and interpret the output.

Analysis of variance (ANOVA): The analytic technique appropriate when an independent variable is categorical with three or more classes and a dependent variable is continuous.

Familywise error: The increase in the likelihood of a Type I error (i.e., erroneous rejection of a true null hypothesis) that results from running repeated statistical tests on a single sample.

Figure 12.1 Hypothetical Distributions With the Same Mean and Different Variances



Between-group variance: The extent to each group or classes is similar to or different from the others in a sample. This is a measure of true group effect, or a relationship between the independent and dependent variables.

Within-group variance: The amount of diversity that exists among the people or objects in a single group or class. This is a measure of random fluctuation, or error.

***F* statistic:** The statistic used in ANOVA; a ratio of the amount of between-group variance present in a sample relative to the amount of within-group variance.

***F* distribution:** The sampling distribution for ANOVA. The distribution is bounded at zero on the left and extends to positive infinity; all values in the *F* distribution are positive.

run a separate *t* test for each combination (AB, AC, AD, BC, BD, CD). That is a lot of *t* tests! The results would be cumbersome and difficult to interpret.

The ANOVA test solves the problems of familywise error and overly complicated output because ANOVA analyzes all classes on the IV simultaneously. One test is all it takes. This simplifies the process and makes for cleaner results.

ANOVA: Different Types of Variances

There are two types of variance analyzed in ANOVA. Both are based on the idea of groups, which are the classes on the IV. If an IV was *political orientation* measured as *liberal, moderate,* or *conservative*, then liberals would be a group, moderates would be a group, and conservatives would be a group. Groups are central to ANOVA.

The first type of variance is **between-group variance**. This is a measure of the similarity among or difference between the groups. It assesses whether groups are markedly different from one another or whether the differences are trivial and meaningless. This is a measure of true group effect. Figure 12.2 illustrates the concept of between-group variance. The groups on the left cluster closely together, while those on the right are distinctly different from one another.

The second kind of variance is **within-group variance** and measures the extent to which people or objects differ from their fellow group members. Within-group variance is driven by random variations between people or objects and is a measure of error. Figure 12.3 depicts the conceptual idea behind within-group variance. The cases in the group on the left cluster tightly around their group's mean, whereas the cases in the right-hand group are scattered widely around their mean. The left-hand group, then, would be said to have much smaller within-group variability than the right-hand group.

The ANOVA test statistic—called the ***F* statistic** because the theoretical probability distribution for ANOVA is the ***F* distribution**—is a ratio that compares the amount of variance between groups to that within groups. When true differences between groups

Figure 12.2 Small and Large Between-Group Variability

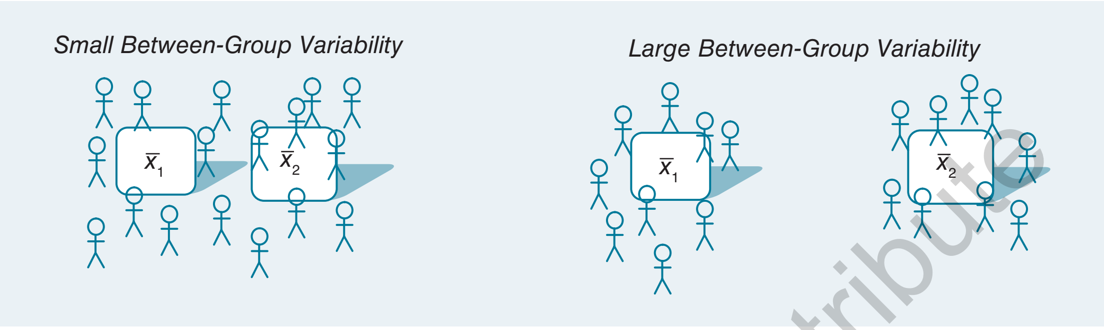
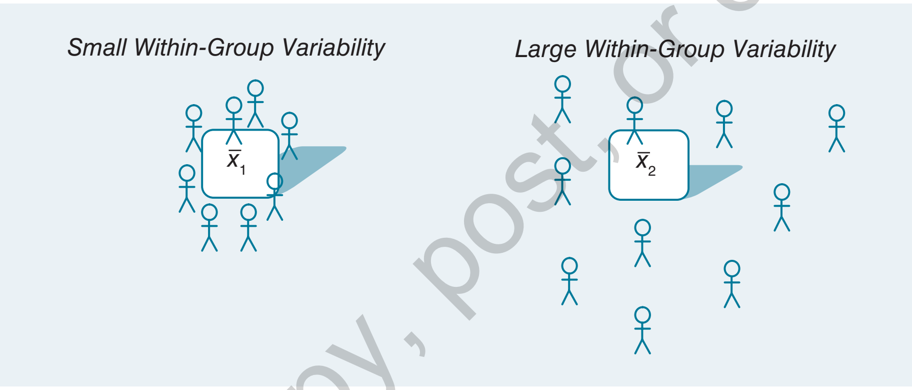


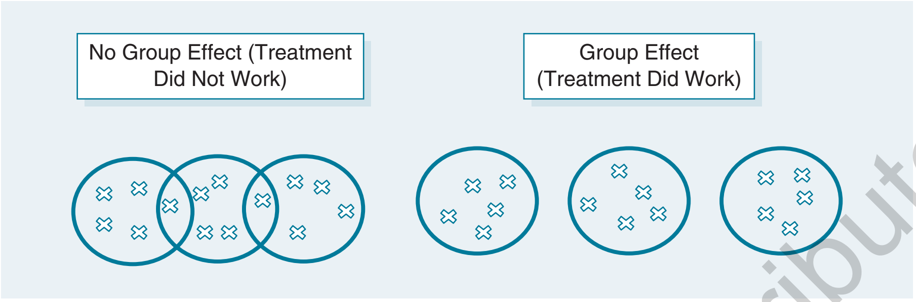
Figure 12.3 Small and Large Within-Group Variability



substantially outweigh the random fluctuations present within each group, the *F* statistic will be large and the null hypothesis that there is no IV–DV relationship will be rejected in favor of the alternative hypothesis that there is an association between the two variables. When between-group variance is small relative to within-group variance, the *F* statistic will be small, and the null will be retained.

An example might help illustrate the concept behind the *F* statistic. Suppose we wanted to test the effectiveness of a mental-health treatment program on recidivism rates in a sample of probationers. We gather three samples: treatment-program completers, people who started the program and dropped out, and those who did not participate in the program at all. Our DV is the number of times each person is rearrested within 2 years of the end of the probation sentence. There will be some random fluctuations within each group; not everybody is going to have the same recidivism score. This is white noise or, more formally, within-group variance—in any sample of people, places, or objects, there will be variation. What we are attempting to discern is whether the difference *between* the groups outweighs the random

Figure 12.4 Recidivism Scores



variance among the people within each group. If the program is effective, then the treatment-completion group should have significantly lower recidivism scores than the other two groups. The impact of the program should be large relative to the random white noise. We might even expect the dropout group's recidivism to be significantly less than the no-treatment group (though probably not as low as the treatment completers). Figure 12.4 diagrams the two possible sets of results.

The *x*'s in the figure represent recidivism scores under two possible scenarios: that within-group variance trumps between-group variance, and that between-group variance is stronger than that within groups. The overlap depicted on the left side suggests that the treatment program was ineffective, since it failed to pull one or two groups away from the others. On the right side, the separation between the groups indicates that they are truly different; this implies that the treatment program did work and that those who completed or started and dropped out are significantly different from each other and from the group that did not participate. An ANOVA test for the left side would yield a small *F* statistic, because the between-group variance is minimal compared to the within-group variance. An ANOVA for the right side, though, would produce a large (statistically significant) *F* because the ratio of between-to-within is high.

The *F* distribution is bounded on the left at zero, meaning it does not have a negative side. As a result, all critical and obtained values of *F* are positive; it is impossible for a correctly calculated *F* to be negative. This is because *F* is based on variance and variance cannot be negative.

Take a moment now to read Research Example 12.1, which describes a situation in which researchers would use ANOVA to test for a difference between groups or, in other words, would attempt to determine whether there is a relationship between a multiple-class IV and a continuous DV.

Franklin and Fearn's (2010) IV (*race* coded as *white; black; Hispanic; Asian*) was a four-class, categorical variable. Their DV (*sentence length*, measured in months) was continuous. ANOVA is the correct bivariate analysis in this situation.

Let's get into an example to see the ANOVA steps and calculations in action. We will use the Juvenile Defendants in Criminal Courts (JDCC; see Data Sources 11.1). We can examine whether attorney type (measured as public defender, assigned counsel, or private attorney) affects the jail sentences received by male youth convicted of weapons offenses. Table 12.1 shows the youths' sentences in months.

Research Example 12.1

Do Asian Defendants Benefit From a "Model Minority" Stereotype?

Numerous studies have found racially based sentencing disparities that are not attributable to differences in defendants' prior records or the severity of their instant offenses. Most such studies have focused on white, black, and Hispanic/Latino defendants. One area of the race-and-sentencing research that has received very little scholarly attention is the effect of race on sentencing among Asians. Franklin and Fearn (2010) set out to determine whether Asian defendants are treated differently from those of other races. They predicted that Asians would be sentenced more leniently due to the stereotype in the United States that Asians are a "model minority," in that they are widely presumed to be an economically, academically, and socially productive group.

To test the hypothesis that Asian defendants are given lighter sentences relative to similarly situated defendants of other races, Franklin and Fearn's (2010) DV was *sentence length,* which was coded as the number of months of incarceration imposed on offenders sentenced to jail or prison. The researchers reported the statistics shown in the table with respect to the mean sentence length across race in this sample.

So, what did the researchers find? It turned out that there were no statistically significant differences between the groups. Franklin and Fearn (2010) retained the null hypothesis that there is no relationship between race and sentencing, and concluded that Asian defendants do not, in fact, receive significantly shorter jail or prison sentences relative to other racial groups once relevant legal factors (e.g., offense type) are taken into account

	Defendant Race				
	White	Black	Hispanic	Asian	Total
Mean Sentence Length (Months)	11.80	17.40	16.50	16.10	15.50

Source: Adapted from Table 1 in Franklin and Fearn (2010).

We will conduct a five-step hypothesis test to determine whether defendants' conviction histories affect their sentences. Alpha will be set at .01.

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

The null hypothesis in ANOVA is very similar to that in *t* tests. The difference is that now there are more than two means. The null is phrased as

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

The structure of the null is dependent on the number of groups—if there were four groups, there would be a μ_4 as well, and five groups would require the addition of a μ_5.

The alternative hypothesis in ANOVA is a bit different from what we have seen before because the only information offered by this test is whether at least one group is significantly different from at least one other group. The *F* statistic indicates neither the

Table 12.1 Jail Sentences (Months) of Male Juveniles Convicted of Weapons Offenses, by Attorney Type

Attorney Type		
Public Defender (x_1)	Assigned Counsel (x_2)	Private Attorney (x_3)
1	4	3
2	2	8
9	3	10
3	6	9
6		12
		11
$n_1 = 5$	$n_2 = 4$	$n_3 = 6$

number of differences nor the specific group or groups that stand out from the others. The alternative hypothesis is, accordingly, rather nondescript. It is phrased as

$$H_1\text{: some } \mu_i \neq \text{some } \mu_j$$

If the null is rejected in an ANOVA test, the only conclusion possible is that at least one group is markedly different from at least one other group—there is no way to tell which group is different or how many between-group differences there are. This is the reason for the existence of **post hoc tests**, which will be covered later in the chapter.

Post hoc tests: Analyses conducted when the null is rejected in ANOVA in order to determine the number and location of differences between groups.

Step 2. Identify the distribution, and compute the degrees of freedom.

As aforementioned, ANOVA relies on the *F* distribution. This distribution is bounded on the left at zero (meaning it has only positive values) and is a family of curves whose shapes are determined by alpha and the degrees of freedom (*df*). There are two types of degrees of freedom in ANOVA: between-group (df_B) and within-groups (df_W). They are computed as

$$df_B = k - 1 \qquad \text{Formula 12(1)}$$

$$df_W = N - k, \qquad \text{Formula 12(2)}$$

where

N = the total sample size across all groups and
k = the number of groups.

The total sample size N is derived by summing the number of cases in each group, the latter of which are called *group sample sizes* and are symbolized n_k. In the present example, there are three groups ($k = 3$) and $N = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$. The degrees of freedom are therefore

Table 12.2 Elements of ANOVA

Sample Sizes	Means	Sums of Squares	Mean Squares
n_k = the sample size of group *k*; the number of cases in each group	$\bar{x}_k$ = group mean; each group's mean on the DV	SS_B = between-groups sums of squares	MS_B = between-groups mean squares
N = the total sample size across all groups	$\bar{x}_G$ = the grand mean; the mean for the entire sample regardless of group	SS_W = within-groups sums of squares SS_T = total sums of squares; $SS_B + SS_W = SS_T$	MS_W = within-groups mean squares

$$df_B = 3 - 1 = 2$$

$$df_W = 15 - 3 = 12$$

Step 3. Identify the critical value and state the decision rule.

The *F* distribution is located in Appendix E. There are different distributions for different alpha levels, so take care to ensure that you are looking at the correct one! You will find the between-group *df* across the top of the table and the within-group *df* down the side. The critical value is located at the intersection of the proper column and row. With $\alpha = .01$, $df_B = 2$, and $df_W = 12$, $F_{crit} = 6.93$. The decision rule is that *if* $F_{obt} > 6.93$*, the null will be rejected*. The decision rule in ANOVA is always phrased using a *greater than* inequality because the *F* distribution contains only positive values, so the critical region is always in the right-hand tail.

Step 4. Compute the obtained value of the test statistic.

Step 4 entails a variety of symbols and abbreviations, all of which are listed and defined in Table 12.2. Stop for a moment and study this chart. You will need to know these symbols and what they mean in order to understand the concepts and formulas about to come.

You already know that each group has a sample size (n_k) and that the entire sample has a total sample size (N). Each group also has its own mean ($\bar{x}_k$), and the entire sample has a grand mean ($\bar{x}_G$). These sample sizes and means, along with other numbers that will be discussed shortly, are used to calculate the three types of sums of squares. The sums of squares are then used to compute mean squares, which, in turn, are used to derive the obtained value of *F*. We will first take a look at the formulas for the three types of sums of squares: total (SS_T), between-group (SS_B), and within-group (SS_W).

$$SS_T = \sum_i \sum_k x^2 - \frac{(\sum_i \sum_k x)^2}{N} \qquad \textit{Formula 12(3)}$$

where

$\sum_i$ = the sum of all scores *i* in group *k*,

$\sum_k$ = the sum of each group total across all groups in the sample,

x = the raw scores, and

N = the total sample size across all groups.

$$SS_B = \sum_k n_k \left(\bar{x}_k - \bar{x}_G\right)^2 \qquad \textit{Formula 12(4)}$$

where

n_k = the number of cases in group *k*,

$\bar{x}_k$ = the mean of group *k*, and

$\bar{x}_G$ = the grand mean across all groups.

$$SS_w = SS_T - SS_B \qquad \textit{Formula 12(5)}$$

The double summation signs in the SS_T formula are instructions to sum sums. The *i* subscript denotes individual scores and *k* signifies groups, so the double sigmas direct you to first sum the scores within each group and to then add up all the group sums to form a single sum representing the entire sample.

Sums of squares are measures of variation. They calculate the amount of variation that exists within and between the groups' raw scores, squared scores, and means. The SS_B formula should look somewhat familiar—in Chapters 4 and 5, we calculated deviation scores by subtracting the sample mean from each raw score. Here, we are going to subtract the grand mean from each group mean. See the connection? This strategy produces a measure of variation. The sums of squares provide information about the level of variability within each group and between the groups.

The easiest way to compute the sums of squares is to use a table. What we ultimately want from the table are (a) the sums of the raw scores for each group, (b) the sums of each group's *squared* raw scores, and (c) each group's mean. All of these numbers are displayed in Table 12.3.

We also need the grand mean, which is computed by summing all of the raw scores across groups and dividing by the total sample size *N*, as such:

$$\bar{x}_G = \frac{\sum_i \sum_k x}{N} \qquad \textit{Formula 12(6)}$$

Here,

$$\bar{x}_G = \frac{21+15+53}{5+4+6} = \frac{89}{15} = 5.93$$

Table 12.3 ANOVA Computation Table for *Attorney Type* and *Jail Sentence*

Attorney Type					
Public Defender (x_1)	x_1^2	Assigned Counsel (x_2)	x_2^2	Private Attorney (x_3)	x_3^2
1	1	4	16	3	9
2	4	2	4	8	64
9	81	3	9	10	100
3	9	6	36	9	81
6	36			12	144
				11	121
$n_1 = 5$		$n_2 = 4$		$n_3 = 6$	
$\Sigma x_1 = 21$	$\Sigma x_1^2 = 131$	$\Sigma x_2 = 15$	$\Sigma x_2^2 = 65$	$\Sigma x_3 = 53$	$\Sigma x_3^2 = 519$
$\bar{x}_1 = \frac{21}{5} = 4.20$		$\bar{x}_2 = \frac{15}{4} = 3.75$		$\bar{x}_3 = \frac{53}{6} = 8.83$	

With all of this information, we are ready to compute the three types of sums of squares, as follows. The process begins with SS_T:

$$SS_T = \sum_i \sum_k x^2 - \frac{(\Sigma_i \Sigma_k x)^2}{N}$$

$$= (131 + 65 + 519) - \frac{(21 + 15 + 53)^2}{15}$$

$$= 715 - \frac{89^2}{15}$$

$$= 715 - \frac{7,921}{15}$$

$$= 715 - 528.07$$

$$= 186.93$$

Then it is time for the between-groups sums of squares:

$$SS_B = \sum_k n_k \left(\bar{x}_k - \bar{x}_G\right)^2$$

$$= 5(4.20 - 5.93)^2 = 4(3.75 - 5.93)^2 + 6(8.83 - 5.93)^2$$

$$= 5(-1.73)^2 + 4(-2.18)^2 + 6(2.90)^2$$

$$= 5(2.99) + 4(4.75) + 6(8.41)$$

$$= 14.95 + 19.00 + 50.46$$

$$= 84.41$$

Next, we calculate the within-groups sums of squares:

$$SS_W = SS_T - SS_B = 186.93 - 84.41 = 102.52$$

Learning Check 12.1

A great way to help you check your math as you go through Step 4 of ANOVA is to remember that the final answers for any of the sums of squares, mean squares, or F_{obt} will never be negative. If you get a negative number for any of your final answers in Step 4, you will know immediately that you made a calculation error, and you should go back and locate the mistake. Can you identify the reason why all final answers are positive? Hint: The answer is in the formulas.

We now have what we need to compute the mean squares (symbolized *MS*). Mean squares transform sums of squares (measures of variation) into variances by dividing SS_B and SS_w by their respective degrees of freedom, df_B and df_w. This is a method of standardization. The mean squares formulas are

$$MS_B = \frac{SS_B}{k-1} \qquad \textit{Formula 12(7)}$$

$$MS_W = \frac{SS_W}{N-k} \qquad \textit{Formula 12(8)}$$

Plugging in our numbers,

$$MS_B = \frac{84.41}{3-1} = 42.21$$

$$MS_W = \frac{102.52}{15-3} = 8.54$$

We now have what we need to calculate F_{obt}. The *F* statistic is the ratio of between-group variance to within-group variance and is computed as

$$F_{obt} = \frac{MS_B}{MS_W} \qquad \textit{Formula 12(9)}$$

Inserting the numbers from the present example,

$$F_{obt} = \frac{42.21}{8.54} = 4.94$$

Step 4 is done! $F_{obt} = 4.94$.

Step 5. Make a decision about the null hypothesis and state the substantive conclusion.

The decision rule stated that if the obtained value exceeded 6.93, the null would be rejected. With an F_{obt} of 4.94, the null is retained. The substantive conclusion is that there is no significant difference between the groups in terms of sentence length received. In other words, male juvenile weapons offenders' jail sentences do not vary as a function of the type of attorney they had. That is, attorney type does not influence jail sentences. This finding makes sense. Research is mixed with regard to whether privately retained attorneys (who cost defendants a lot of money) really are better than publicly funded defense attorneys (who are provided to indigent defendants for free). While there is a popular assumption that privately retained attorneys are better, the reality is that publicly funded attorneys are frequently as or even more skilled than private ones are.

We will go through another ANOVA example. If you are not already using your calculator to work through the steps as you read and make sure you can replicate the results obtained here in the book, start doing so. This is an excellent way to learn the material.

For the second example, we will study handguns and murder rates. Handguns are a prevalent murder weapon and, in some locations, they account for more deaths than all other modalities combined. In criminal justice and criminology researchers' ongoing efforts to learn about violent crime, the question arises as to whether there are geographical differences in handgun-involved murders. Uniform Crime Report (UCR) data can be used to find out whether there are significant regional differences in handgun murder rates (calculated as the number of murders by handgun per 100,000 residents in each state). A random sample of states was drawn, and the selected states were divided by region. Table 12.4 contains the data in the format that will be used for computations. Alpha will be set at .05.

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

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

$$H_1: \text{some } \mu_i \neq \text{some } \mu_j$$

Step 2. Identify the distribution and compute the degrees of freedom.

This being an ANOVA, the F distribution will be employed. There are four groups, so $k = 4$. The total sample size is $N = 5 + 5 + 7 + 6 = 23$. Using Formulas 12(1) and 12(2), the degrees of freedom are

$$df_B = 4 - 1 = 3$$

$$df_W = 23 - 4 = 19$$

Step 3. Identify the critical value and state the decision rule.

With $\alpha = .05$ and the earlier derived *df* values, $F_{crit} = 3.13$. The decision rule states that *if* $F_{obt} > 3.13$, H_0 *will be rejected.*

Table 12.4 Handgun Murder Rate by Region

Region							
Northeast		Midwest		South		West	
(x_1)	$(x_1)^2$	(x_2)	$(x_2)^2$	(x_3)	$(x_3)^2$	(x_4)	$(x_4)^2$
1.45	2.10	2.12	4.49	4.16	17.31	.14	.02
.30	.09	.10	.01	1.87	3.50	1.09	1.19
.71	.50	1.35	1.82	3.29	10.82	.92	.85
2.17	4.71	.66	.44	2.82	7.95	2.50	6.25
.00	.00	.15	.02	2.52	6.35	1.13	1.28
				2.67	7.13	.19	.04
				1.37	1.88		
$n_1 = 5$		$n_2 = 5$		$n_3 = 7$		$n_4 = 6$	
$\sum x_1 = 4.63$	$\sum x_1^2 = 7.40$	$\sum x_2 = 4.38$	$\sum x_2^2 = 6.78$	$\sum x_3 = 18.70$	$\sum x_3^2 = 54.94$	$\sum x_4 = 5.97$	$\sum x_4^2 = 9.63$
$\bar{x}_1 = .93$		$\bar{x}_2 = .88$		$\bar{x}_3 = 2.67$		$\bar{x}_4 = 1.00$	

Step 4. Calculate the obtained value of the test statistic.

We begin by calculating the total sums of squares:

$$SS_T = (7.40 + 6.78 + 54.94 + 9.63) - \frac{(4.63 + 4.38 + 18.70 + 5.97)^2}{23}$$

$$= 78.75 - \frac{33.68^2}{23}$$

$$= 78.75 - \frac{1{,}134.34}{23}$$

$= 78.75 - 49.32$

$= 29.43$

Before computing the between-groups sums of squares, we need the grand mean:

$$\bar{x}_G = \frac{4.63 + 4.38 + 18.70 + 5.97}{23} = \frac{33.68}{23} = 1.46$$

Now SS_B can be calculated:

$$SS_B = 5(.93 - 1.46)^2 + 5(.88 - 1.46)^2 + 7(2.67 - 1.46)^2 + 6(1.00 - 1.46)^2$$

$= 5(-.53)^2 + 5(-.58)^2 + 7(1.21)^2 + 6(-.46)^2$

$= 5(.28) + 5(.34) + 7(1.46) + 6(.21)$

$= 1.40 + 1.70 + 10.22 + 1.26$

$= 14.58$

Next, we calculate the within-groups sums of squares:

$$SS_w = 29.43 - 14.58 = 14.85$$

Plugging our numbers into Formulas 12(7) and 12(8) for mean squares gives

$$MS_B = \frac{14.58}{4 - 1} = 4.86$$

$$MS_W = \frac{14.85}{23 - 4} = .78$$

Finally, using Formula 12(9) to derive F_{obt},

$$F_{obt} = \frac{4.86}{.78} = 6.23$$

This is the obtained value of the test statistic. $F_{obt} = 6.23$, and Step 4 is complete.

Step 5. Make a decision about the null and state the substantive conclusion.

In Step 3, the decision rule stated that if F_{obt} turned out to be greater than 3.13, the null would be rejected. Since F_{obt} ended up being 6.23, the null is indeed rejected. The substantive interpretation is that there is a significant difference across regions in the handgun-murder rate.

Research Example 12.2

Are Juveniles Who Are Transferred to Adult Courts Seen as More Threatening?

Recent decades have seen a shift in juvenile-delinquency policy. There has been an increasing zero tolerance sentiment with respect to juveniles who commit serious offenses. The reaction by most states has been to make it easier for juveniles to be tried as adults, which allows their sentences to be more severe than they would be in juvenile court. The potential problem with this strategy is that there is a prevalent stereotype about juveniles who get transferred or waived to adult court: They are often viewed as vicious, cold-hearted predators. Judges, prosecutors, and jurors might be biased against transferred juveniles, simply because they got transferred. This means that a juvenile and an adult could commit the same offense and yet be treated very differently by the court, potentially even ending up with different sentences.

Tang, Nuñez, and Bourgeois (2009) tested mock jurors' perceptions about the dangerousness of 16-year-olds who were transferred to adult court, 16-year-olds who were kept in the juvenile justice system, and 19-year-olds in adult court. They found that mock jurors rated transferred 16-year-olds as committing more serious crimes, being more dangerous, and having a greater likelihood of chronic offending relative to non-transferred juveniles and to 19-year-olds. The following table shows the means, standard deviations, and *F* tests.

	16-Year-Olds Transferred	*16-Year-Olds Not Transferred*	*19-Year-Olds*	
	Mean (sd)	*Mean (sd)*	*Mean (sd)*	F
Serious	7.44 (.24)	5.16 (.23)	5.66 (.24)	26.30
Dangerous	6.92 (.25)	4.76 (.25)	5.47 (.25)	19.45
Chronic	7.18 (.25)	5.74 (.24)	5.87 (.25)	10.22

Source: Adapted from Table 1 in Tang et al. (2009).

As you can see, all the *F* statistics were large; the null was rejected for each test. The transferred juveniles' means are higher than the other two groups' means for all measures. These results suggest that transferring juveniles to adult court could have serious implications for fairness. In some cases, prosecutors have discretion in deciding whether to waive a juvenile over to adult court, which means that two juveniles guilty of similar crimes could end up being treated very differently. Even more concerning is the disparity between transferred youths and 19-year-olds—it appears that juveniles who are tried in adult court could face harsher penalties than adults, even when their crimes are the same.

As another example, we will analyze data from the Firearm Injury Surveillance Study (FISS; Data Sources 8.2) to find out whether victim age varies significantly across the different victim–offender relationships. There are four relationship categories, and a total sample size of 22. Table 12.5 shows the data and calculations of the numbers needed to complete the hypothesis test. We will proceed using the five steps. Alpha will be set at .05.

Step 1. State the null (H_0) and alternative (H_1) hypotheses.

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

$$H_1: \textit{some } \mu_i \neq \textit{some } \mu_j$$

Step 2. Identify the distribution and compute the degrees of freedom.

This being an ANOVA, the F distribution will be employed. There are four groups, so $k = 4$. The total sample size is $N = 7 + 6 + 4 + 5 = 22$. Using Formulas 12(1) and 12(2), the degrees of freedom are

$$df_B = 4 - 1 = 3$$

$$df_W = 22 - 4 = 18$$

Step 3. Identify the critical value and state the decision rule.

With α =.05 and the earlier derived df values, $F_{crit} = 3.16$. The decision rule states that *if $F_{obt} > 3.16$, H_0 will be rejected.*

Table 12.5 Victim Age and Victim–Offender Relationship

Victim–Offender Relationship							
Stranger		Friend/Acquaintance		Intimate Partner		Relative	
(x_1)	$(x_1)^2$	(x_2)	$(x_2)^2$	(x_3)	$(x_3)^2$	(x_4)	$(x_4)^2$
29	841	19	361	45	2,025	23	529
31	961	25	625	36	1,296	22	484
24	576	22	484	39	1,521	19	361
27	729	24	576	40	1,600	18	324
33	1,089	30	900			25	625
28	784	25	625				
34	1,156						
$n_1 = 7$		$n_2 = 6$		$n_3 = 4$		$n_4 = 5$	
$\Sigma x_1 = 206$	$\Sigma x_1^2 = 6{,}136$	$\Sigma x_1 = 145$	$\Sigma x_2^2 = 3{,}571$	$\Sigma x_1 = 160$	$\Sigma x_3^2 = 6{,}442$	$\Sigma x_4^2 = 107$	$\Sigma x_4^2 = 2{,}323$
$\bar{x}_1 = 29.43$		$\bar{x}_2 = 24.17$		$\bar{x}_3 = 40.00$		$\bar{x}_4 = 21.40$	

Step 4. Calculate the obtained value of the test statistic.

The total sums of squares for the data in Table 12.5 is

$$SS_T = \sum_i \sum_k x^2 - \frac{(\sum_i \sum_k x)^2}{N}$$

$$= (6{,}136 + 3{,}571 + 6{,}442 + 2{,}323) - \frac{(206 + 145 + 160 + 107)^2}{22}$$

$$= 18{,}472 - \frac{618.00^2}{22}$$

$$= 18{,}472 - \frac{381{,}924}{22}$$

$$= 18{,}472 - 17360.18$$

$$= 1{,}111.82$$

Next, we need the grand mean:

$$\bar{x}_G = \frac{\sum_i \sum_k x}{N} = \frac{206 + 145 + 160 + 107}{22} = \frac{618.00}{22} = 28.09$$

Now SS_B can be calculated:

$$SS_B = \sum_k n_k \left(\bar{x}_k - \bar{x}_G\right)^2$$

$$= 7\left(29.43 - 28.09\right)^2 + 6\left(24.17 - 28.09\right)^2 + 4\left(40.00 - 28.09\right)^2 + 5\left(21.40 - 28.09\right)^2$$

$$= 7(1.34)^2 + 6(-3.92)^2 + 4(11.91)^2 + 5(-6.69)^2$$

$$= 7(1.80) + 6(15.37) + 4(141.85) + 5(44.76)$$

$$= 12.60 + 92.22 + 567.40 + 223.80$$

$$= 896.02$$

Next, we calculate the within-groups sums of squares:

$$SS_w = 1{,}111.82 - 896.02 = 215.80$$

And the mean squares are

$$MS_B = \frac{SS_B}{k-1} = \frac{896.02}{4-1} = 298.67$$

$$MS_W = \frac{SS_W}{N-k} = \frac{215.80}{22-4} = 11.99$$

Finally, F_{obt} is calculated as

$$F_{obt} = \frac{MS_B}{MS_W} = \frac{298.67}{11.99} = 24.91$$

And Step 4 is done. $F_{obt} = 24.91$.

Step 5. Make a decision about the null and state the substantive conclusion.

In Step 3, the decision rule stated that if F_{obt} turned out to be greater than 3.16, the null would be rejected. Since F_{obt} is 24.91, we reject the null. It appears that victim age does vary across the different victim–offender relationship categories.

After finding a significant F indicating that at least one group stands out from at least one other one, the obvious question is, "Which group or groups are different?" We might want to know which region or regions have a significantly higher or lower rate than the others or which victim–offender relationship or relationships contain significantly younger or older victims. The F statistic is silent with respect to the location and number of differences, so post hoc tests are used to get this information. The next section covers post hoc tests and measures of association that can be used to gauge relationship strength.

When the Null Is Rejected: A Measure of Association and Post Hoc Tests

If the null is not rejected in ANOVA, then the analysis stops because the conclusion is that the IVs and DV are not related. If the null is rejected, however, it is customary to explore the statistically significant results in more detail using measures of association (MAs) and post hoc tests. Measures of association permit an assessment of the strength of the relationship between the IV and the DV, and post hoc tests allow researchers to determine which groups are significantly different from which other ones. The MA that will be discussed here is fairly easy to calculate by hand, but the post hoc tests will be discussed and then demonstrated in the SPSS section, because they are computationally intensive.

Omega squared: A measure of association used in ANOVA when the null has been rejected in order to assess the magnitude of the relationship between the independent and dependent variables. This measure shows the proportion of the total variability in the sample that is attributable to between-group differences.

Omega squared (ω^2) is an MA for ANOVA that is expressed as the proportion of the total variability in the sample that is due to between-group differences. Omega squared can be left as a proportion or multiplied by 100 to form a percentage. Larger values of ω^2 indicate stronger IV–DV relationships, whereas smaller values signal weaker associations. Omega squared is computed as

$$\omega^2 = \frac{SS_B - (k-1)MS_W}{MS_W + SS_T} \qquad \textit{Formula 12(10)}$$

Earlier, we found a statistically significant relationship between region and handgun murder rates. Now we can calculate how strong the relationship is. Using ω^2,

$$\omega^2 = \frac{14.58-(4-1).78}{.78+29.43} = \frac{14.58-2.34}{30.21} = \frac{12.24}{30.21} = .41$$

Omega squared shows that 41% of the total variability in the states' handgun-murder rates is a function of regional characteristics. Region appears to be a very important determinate of the prevalence of handgun murders.

We can do the same for the test showing significant differences in victims' ages across four different types of victim–offender relationships. Plugging the relevant numbers into Formula 12(10) yields

$$\omega^2 = \frac{896.02-(4-1)11.99}{11.99+1{,}111.82} = \frac{896.02-35.97}{1{,}123.81} = \frac{860.05}{1{,}123.81} = .77$$

This means that 77% of the variability in victims' ages is attributable to the relationship between the victim and the shooter. This points to age being a function of situational characteristics. Younger people are more at risk of firearm injuries in certain types of situations, while older people face greater risk in other circumstances. Of course, we still do not know which group or groups are significantly different from which other group or groups. For this, post hoc tests are needed.

Tukey's honest significant difference: A widely used post hoc test that identifies the number and location(s) of differences between groups.

Bonferroni: A widely used and relatively conservative post hoc test that identifies the number and location(s) of differences between groups.

There are many different types of post hoc tests, so two of the most popular ones are presented here. The first is **Tukey's honest significant difference (HSD)**. Tukey's test compares each group to all the others in a series of two-variable hypothesis tests. The null hypothesis in each comparison is that both group means are equal; rejection of the null means that there is a significant difference between them. In this way, Tukey's is conceptually similar to a series of *t* tests, though the HSD method sidesteps the problem of familywise error.

Bonferroni is another commonly used test and owes its popularity primarily to the fact that it is fairly conservative. This means that it minimizes Type I error (erroneously rejecting a true null) at the cost of increasing the likelihood of a Type II error (erroneously retaining a false null). The Bonferroni, though, has been criticized for being too conservative. In the end, the best method is to select both Tukey's and Bonferroni in order to garner a holistic picture of your data and make an informed judgment.

The computations of both post hoc tests are complex, so we will not attempt them by hand and will instead demonstrate their use in SPSS.

Learning Check 12.2

Would it be appropriate to compute omega squared and post hoc tests for the ANOVA in the example pertaining to juvenile defendants' attorneys and sentences? Why or why not?

Research Example 12.3

Does Crime Vary Spatially and Temporally in Accordance With Routine Activities Theory?

Crime varies across space and time; in other words, there are places and times it is more (or less) likely to occur. Routine activities theory has emerged as one of the most prominent explanations for this variation. Numerous studies have shown that the characteristics of places can attract or prevent crime and that large-scale patterns of human behavior shape the way crime occurs. For instance, a tavern in which negligent bartenders frequently sell patrons too much alcohol might generate alcohol-related fights, car crashes, and so on. Likewise, when schools let out for summer break, cities experience a rise in the number of unsupervised juveniles, many of whom get into mischief. Most of this research, however, has been conducted in Western nations. De Melo, Pereira, Andresen, and Matias (2017) extended the study of spatial and temporal variation in crime rates to Campinas, Brazil, to find out if crime appears to vary along these two dimensions. They broke crime down by type and ran ANOVAs to test for temporal variation across different units of time (season, month, day of week, hour of day). The table displays the results for the ANOVAs that were statistically significant. (Nonsignificant findings have been omitted.)

Unit of Time	Crime	F value	p Value
Season	Homicide	3.252	.059
Day of week	Homicide	3.821	.009
	Robbery	16.77	.000
Hour of day	Burglary	3.88	.009
	Theft	39.98	.000
	Homicide	3.05	.000
	Rape	3.14	.000
	Robbery	98.02	.000
	Burglary	18.09	.000
	Theft	101.00	.000

Source: Adapted from Table 1 in De Melo et al. (2017).

As the table shows, homicide rates vary somewhat across month. Post hoc tests showed that the summer months experienced spikes in homicide, likely because people are outdoors more often when the weather is nice, which increases the risk for violent victimization and interpersonal conflicts. None of the variation across month was statistically significant (which is why there are no rows for this unit of time in the table). There was significant temporal variation across weeks and hours. Post hoc tests revealed interesting findings across crime type. For example, homicides are more likely to occur on weekends (since people are out and about more during weekends than during weekdays), while burglaries are more likely to happen on weekdays (since people are at work). The variation across hours of the day was also significant for all crime types, but the pattern was different within each one. For instance, crimes of violence were more common in late evenings and into the night, while burglary was most likely to occur during the daytime hours.

SPSS

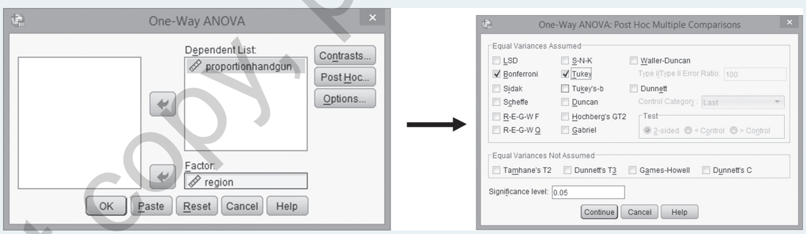
Let us revisit the question asked in Example 2 regarding whether handgun murder rates vary by region. To run an ANOVA in SPSS, follow the steps depicted in Figure 12.5. Use the *Analyze* → *Compare Means* → *One-Way ANOVA* sequence to bring up the dialog box on the left side in Figure 12.5 and then select the variables you want to use.

Move the IV to the *Factor* space and the DV to the *Dependent List*. Then click *Post Hoc* and select the *Bonferroni* and *Tukey* tests. Click *Continue* and *OK* to produce the output shown in Figure 12.6.

The first box of the output shows the results of the hypothesis test. You can see the sums of squares, *df*, and mean squares for within groups and between groups. There are also total sums of squares and total degrees of freedom. The number in the *F* column is F_{obt}. Here, you can see that F_{obt} = 6.329. When we did the calculations by hand, we got 6.23. Our hand calculations had some rounding error, but this did not affect the final decision regarding the null because you can also see that the significance value (the *p* value) is .004, which is less than .05, the value at which α was set. The null hypothesis is rejected in the SPSS context just like it was in the hand calculations.

The next box in the output shows the Tukey and Bonferroni post hoc tests. The difference between these tests is in the *p* values in the *Sig.* column. In the present case, those differences are immaterial because the results are the same across both types of tests. Based on the asterisks that flag significant results and the fact that the *p* values associated with the flagged numbers are less than .05, it is apparent that the South is the region that stands out from the others. Its mean is significantly greater than all three of the other regions' means. The Northeast, West, and Midwest do not differ significantly from one another, as evidenced by the fact that all of their *p* values are greater than .05.

Figure 12.5 Running an ANOVA in SPSS



In Figure 12.7, you can see that the F_{obt} SPSS produces (24.719) is nearly identical to the 12.91 we arrived at by hand. Looking at Tukey's and Bonferroni, it appears that the categories "relative" and "friend/acquaintance" are the only ones that do not differ significantly from one another. In the full data set, the mean age of victims shot by relatives is 21.73 and that for the ones shot by friends and acquaintances is 24.05. These means are not significantly different from each other, but they are both distinct from the means for stranger-perpetrated shootings (mean age of 29.58) and intimate-partner shootings (39.12).

Figure 12.6 ANOVA Output

Handgun murders per 100,000

	Sum of Squares	df	Mean Square	F	Sig
Between Groups	14.674	3	4.891	6.329	.004
Within Groups	14.684	19	.773		
Total	29.358	22			

Post hoc Tests

Multiple Comparisons

Dependent Variable: Handgun Murders per 100,000

	(I) Region	(J) Region	Mean Difference (I- J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	Northeast	Midwest	.05037	.55600	1.000	−1.5130	1.6138
		South	−1.74371*	.51475	.015	−3.1911	−.2963
		West	-0.6898	.53233	.999	−1.5658	1.4278
	Midwest	Northeast	−.05037	.55600	1.000	−1.6138	1.5130
		South	−1.79408*	.51475	.012	−3.2415	−.3467
		West	−.11935	.53233	.996	−1.6162	1.3775
	South	Northeast	1.74371*	.51475	.015	.2963	3.1911
		Midwest	1.79408*	.51475	.012	.3467	3.2415
		West	1.67473*	.48909	.014	.2995	3.0500
	west	Northeast	.06898	.53233	.999	−1.4278	1.5658
		Midwest	.11935	.53233	.996	−1.3775	1.6162
		South	−1.67473*	.48909	.014	−3.0500	−.2995
Bonferroni	Northeast	Midwest	.05037	.55600	1.000	−1.5864	1.6872
		South	−1.74371*	.51475	.019	−3.2591	−.2283
		West	−.06899	.53233	1.000	−1.6361	1.4991
	Midwest	Northeast	−05037	.55600	1.000	−1.6872	1.5864
		South	−1.79408*	.51475	.015	−3.3095	−.2787
		West	−.11935	.53233	1.000	−1.6865	1.4478
	South	Northeast	1.74371*	.51475	.019	.2283	3.2591
		Midwest	1.79408*	.51475	.015	.2787	3.3095
		West	1.67473*	.48909	.017	.2349	3.1146
	West	Northeast	.06898	.53233	1.000	−1.4981	1.6361
		Midwest	.11935	.53233	1.000	−1.4478	1.6865
		South	−1.67473*	.48909	.017	−3.1146	−.2349

*The mean difference is significant at the 0.05 level.

We can also use SPSS and the full FISS to reproduce the analysis we did by hand using a sample of cases. Figure 12.7 shows the ANOVA and post hoc tests.

Figure 12.7 ANOVA Output

ANOVA

Victim's age

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11795.008	3	3931.669	24.719	.000
Within Groups	163827.511	1030	159.056		
Total	175622.518	1033			

Post HOC Tests

Multiple Comparisons

Dependent Variable: Victim's age

	(I) Offender's relationship to victim	(J) Offender's relationship to victim	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	Stranger	Friend/acq	5.524*	.970	.000	3.03	8.02
		Intimate partner	–9.542*	2.567	.001	–16.15	–2.94
		Relative	7.848*	1.420	.000	4.19	11.50
	Friend/acq	Stranger	–5.524*	.970	.000	–8.02	–3.03
		Intimate partner	–15.066*	2.660	.000	–21.91	–8.22
		Relative	2.323	1.581	.456	–1.75	6.39
	Intimate partner	Stranger	9.542*	2.567	.001	2.94	16.15
		Friend/acq	15.066*	2.660	.000	8.22	21.91
		Relative	17.390*	2.855	.000	10.04	24.74
	Relative	Stranger	–7.848*	1.420	.000	–11.50	–4.19
		Friend/acq	–2.323	1.581	.456	–6.39	1.75
		Intimate partner	–17.390*	2.855	.000	–24.74	–10.04
Bonferroni	Stranger	Friend/acq	5.524*	.970	.000	2.96	8.09
		Intimate partner	–9.542*	2.660	.001	–16.33	–2.76
		Relative	7.848*	1.581	.000	4.10	11.60
	Friend/acq	Stranger	–5.524*	.970	.000	–8.09	–2.96
		Intimate partner	–15.066*	2.660	.001	–22.10	–8.03
		Relative	2.323	1.581	.000	–1.86	6.50
	Intimate partner	Stranger	9.542*	2.567	.000	2.76	16.33
		Friend/acq	15.066*	2.660	.000	8.03	22.10
		Relative	17.390*	2.855	.852	9.84	24.94
	Relative	Stranger	–7.848*	1.420	.001	–11.60	–4.10
		Friend/acq	-2.323	1.581	.000	–6.50	1.86
		Intimate partner	–17.390*	2.855	.000	–24.94	–9.84

*The mean difference is significant at the 0.05 level.

CHAPTER SUMMARY

This chapter taught you what to do when you have a categorical IV with three or more classes and a continuous DV. A series of *t* tests in such a situation is not viable because of the familywise error rate. In an analysis of variance, the researcher conducts multiple between-group comparisons in a single analysis. The *F* statistic compares between-group variance to within-group variance to determine whether between-group variance (a measure of true effect) substantially outweighs within-group variance (a measure of error). If it does, the null is rejected; if it does not, the null is retained.

The ANOVA *F*, though, does not indicate the size of the effect, so this chapter introduced you to an MA that allows for a determination of the strength of a relationship. This measure is omega squared (w^2), and it is used only when the null has been rejected—there is no sense in examining the strength of an IV–DV relationship that you just said does not exist! Omega squared is interpreted as the proportion of the variability in the DV that is attributable to the IV. It can be multiplied by 100 to be interpreted as a percentage.

The *F* statistic also does not offer information about the location or number of differences between groups. When the null is retained, this is not a problem because a retained null means that there are no differences between groups; however, when the null is rejected, it is desirable to gather more information about which group or groups differ from which others. This is the reason for the existence of post hoc tests. This chapter covered Tukey's HSD and Bonferroni, which are two of the most commonly used post hoc tests in criminal justice and criminology research. Bonferroni is a conservative test, meaning that it is more difficult to reject the null hypothesis of no difference between groups. It is a good idea to run both tests and, if they produce discrepant information, make a reasoned judgment based on your knowledge of the subject matter and data. Together, MAs and post hoc tests can help you glean a comprehensive and informative picture of the relationship between the independent and dependent variables.

THINKING CRITICALLY

1. What implications does the relationship between shooting victims' ages and these victims' relationships with their shooters have for efforts to prevent firearm violence? For each of the four categories of victim–offender relationship, consider the mean age of victims and devise a strategy that could be used to reach people of this age group and help them lower their risks of firearm victimization.
2. A researcher is evaluating the effectiveness of a substance abuse treatment program for jail inmates. The researcher categorizes inmates into three groups: those who completed the program, those who started it and dropped out, and those who never participated at all. He follows up with all people in the sample six months after their release from jail and asks them whether or not they have used drugs since being out. He codes drug use as 0 = no and 1 = yes. He plans to analyze the data using an ANOVA. Is this the correct analytic approach? Explain your answer.

REVIEW PROBLEMS

1. A researcher wants to know whether judges' gender (measured as *male; female*) affects the severity of sentences they impose on convicted defendants (measured as *months of incarceration*). Answer the following questions:
   a. What is the independent variable?
   b. What is the level of measurement of the independent variable?
   c. What is the dependent variable?

d. What is the level of measurement of the dependent variable?
e. What type of hypothesis test should the researcher use?

2. A researcher wants to know whether judges' gender (measured as *male; female*) affects the types of sentences they impose on convicted criminal defendants (measured as *jail; prison; probation; fine; other*). Answer the following questions:
   a. What is the independent variable?
   b. What is the level of measurement of the independent variable?
   c. What is the dependent variable?
   d. What is the level of measurement of the dependent variable?
   e. What type of hypothesis test should the researcher use?

3. A researcher wishes to find out whether arrest deters domestic violence offenders from committing future acts of violence against intimate partners. The researcher measures arrest as *arrest; mediation; separation; no action* and recidivism as *number of arrests for domestic violence within the next 3 years*. Answer the following questions:
   a. What is the independent variable?
   b. What is the level of measurement of the independent variable?
   c. What is the dependent variable?
   d. What is the level of measurement of the dependent variable?
   e. What type of hypothesis test should the researcher use?

4. A researcher wishes to find out whether arrest deters domestic violence offenders from committing future acts of violence against intimate partners. The researcher measures arrest as *arrest; mediation; separation; no action* and recidivism as whether these offenders were arrested for domestic violence within the next 2 years (measured as *arrested; not arrested*). Answer the following questions:
   a. What is the independent variable?
   b. What is the level of measurement of the independent variable?
   c. What is the dependent variable?
   d. What is the level of measurement of the dependent variable?
   e. What type of hypothesis test should the researcher use?

5. A researcher wants to know whether poverty affects crime. The researcher codes neighborhoods as being *lower-class*, *middle-class*, or *upper-class* and obtains the crime rate for each area (measured as the number of index offenses per 10,000 residents). Answer the following questions:
   a. What is the independent variable?
   b. What is the level of measurement of the independent variable?
   c. What is the dependent variable?
   d. What is the level of measurement of the dependent variable?
   e. What type of hypothesis test should the researcher use?

6. A researcher wants to know whether the prevalence of liquor-selling establishments (such as bars and convenience stores) in neighborhoods affects crime in those areas. The researcher codes neighborhoods as having *0–1*, *2–3*, *4–5*, or *6+* liquor-selling establishments. The researcher also obtains the crime rate for each area (measured as the number of index offenses per 10,000 residents). Answer the following questions:
   a. What is the independent variable?
   b. What is the level of measurement of the independent variable?
   c. What is the dependent variable?
   d. What is the level of measurement of the dependent variable?
   e. What type of hypothesis test should the researcher use?

7. Explain within-groups variance and between-groups variance. What does each of these concepts represent or measure?

8. Explain the *F* statistic in conceptual terms. What does it measure? Under what circumstances will *F* be small? Large?

9. Explain why the *F* statistic can never be negative.

10. When the null hypothesis in an ANOVA test is rejected, why are MA and post hoc tests necessary?

11. The Omnibus Crime Control and Safe Streets Act of 1968 requires state and federal courts to report information on all wiretaps sought by and authorized for law enforcement agencies (Duff, 2010). One question of interest to someone studying wiretaps is whether wiretap use varies by crime type; that is, we might want to know whether law enforcement agents use wiretaps with greater frequency in certain types of investigations than in other types. The following table contains data from the U.S. courts website (www.uscourts.gov/Statistics.aspx) on the number of wiretaps sought by law enforcement agencies in a sample of states. The wiretaps are broken down by offense type, meaning that each number in the table represents the number of wiretap authorizations received by a particular state for a particular offense. Using an alpha level of .05, test the null hypothesis of no difference between the group means against the alternative hypothesis that at least one group mean is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Offense Type		
Homicide and Assault (x_1)	Narcotics (x_2)	Racketeering (x_3)
2	25	1
0	1	1
0	4	0
1	2	3
14	21	0
1	3	0
2	12	0
$n_1 = 7$	$n_2 = 7$	$n_3 = 7$

12. Some studies have found that people become more punitive as they age, such that older people, as a group, hold harsher attitudes toward people who commit crimes. The General Social Survey (GSS) asks people for their opinions about courts' handling of criminal defendants. This survey also records respondents' ages. Use the data below and an alpha level of .05 to test the null hypothesis of no difference between the group means against the alternative hypothesis that at least one group mean is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Courts' Handling of Criminal Defendants		
Too Harsh (x_1)	Not Harsh Enough (x_2)	About Right (x_3)
47	50	49
46	56	48
44	52	50
41	46	47
39	49	52
50	47	49
$n_1 = 6$	$n_2 = 6$	$n_3 = 6$

13. In the ongoing effort to reduce police injuries and fatalities resulting from assaults, one issue is the technology of violence against officers or, in other words, the type of implements

offenders use when attacking police. Like other social events, weapon use might vary across regions. The UCRs collect information on weapons used in officer assaults. These data can be used to find out whether the percentage of officer assaults committed with firearms varies by region. The following table contains the data. Using an alpha level of .01, test the null of no difference between means against the alternative that at least one region is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Region			
Northeast (x_1)	Midwest (x_2)	South (x_3)	West (x_4)
.76	1.05	2.86	3.55
1.53	5.28	2.41	4.52
2.65	4.92	3.49	3.64
.00	.96	2.12	2.29
.23	1.41	3.39	3.88
$n_1 = 5$	1.50	$n_3 = 5$	4.90
	$n_2 = 6$		.68
			$n_4 = 7$

14. An ongoing source of question and controversy in the criminal court system are the possible advantages that wealthier defendants might have over poorer ones, largely as a result of the fact that the former can pay to hire their own attorneys, whereas the latter must accept the services of court-appointed counsel. There is a common perception that privately retained attorneys are more skilled and dedicated than their publicly appointed counterparts. Let us examine this issue using a sample of property defendants from the JDCC data set. The IV is *attorney type* and the DV is *days to pretrial release*, which measures the number of days between arrest and pretrial release for those rape defendants who were released pending trial. (Those who did not make bail or were denied bail are not included.) Using an alpha level of .05, test the null of no difference between means against the alternative that at least one region is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Attorney Type		
Public Defender (x_1)	Assigned Counsel (x_2)	Private Attorney (x_3)
2	0	0
42	0	0
5	6	0
4	51	1
8	5	3
1	5	24
0	$n_2 = 6$	5
$n_1 = 7$		34
		$n_3 = 8$

15. In Research Example 12.1, we read about a study that examined whether Asian defendants were sentenced more leniently than offenders of other races. Let us run a similar test using data from the JDCC. The following table contains a sample of juveniles convicted of property offenses and sentenced to probation. The IV is *race*, and the DV is each person's *probation sentence in months*. Using an alpha level of .01, test the null of no difference between means against the alternative that at least one region is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Race			
Asian (x_1)	Black (x_2)	White (x_3)	Other (x_4)
3	12	6	2
9	10	18	6

Race			
Asian (x_1)	Black (x_2)	White (x_3)	Other (x_4)
14	2	3	18
8	6	2	12
24	72	24	$n_3 = 4$
12	$n_2 = 5$	$n_3 = 5$	
$n_1 = 6$			

16. Across police agencies of different types, is there significant variation in the prevalence of bachelor's degrees among sworn personnel? The table contains Law Enforcement Management and Administrative Statistics (LEMAS) data showing a sample of agencies broken down by type. The numbers represent the percentage of sworn personnel that has a bachelor's degree or higher. Using an alpha level of .01, test the null of no difference between means against the alternative that at least one facility type is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Agency Type			
Municipal (x_1)	County (x_2)	State (x_3)	Tribal (x_4)
6.27	1.87	4.00	4.91
6.45	1.90	3.88	4.96
5.89	2.10	4.19	4.80
6.35	1.45	3.94	5.21
6.30	1.78		
6.28			
$n_1 = 6$	$n_2 = 5$	$n_3 = 4$	$n_4 = 4$

17. Let's continue using the LEMAS survey and exploring differences across agencies of varying types. Problem-oriented policing has been an important innovation in the police approach to reducing disorder and crime. This approach encourages officers to investigate ongoing problems, identify their source, and craft creative solutions. The LEMAS survey asks agency top managers whether they encourage patrol officers to engage in problem solving and, if they do, what percentage of their patrol officers are encouraged to do this type of activity. Using an alpha level of .05, test the null of no difference between means against the alternative that at least one agency type is significantly different from at least one other. Use all five steps. If appropriate, compute and interpret omega squared.

Agency Type			
Municipal (x_1)	County (x_2)	State (x_3)	Tribal (x_4)
45	36	29	57
44	37	28	58
45	36	31	59
48	34	28	56
39	38	29	57
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$n_4 = 5$

18. Do the number of contacts people have with police officers vary by race? The Police–Public Contact Survey (PPCS) asks respondents to report their race and the total number of face-to-face contacts they have had with officers in the past year. The following table shows the data. Using an alpha level of .05, test the null of no difference between means against the alternative that at least one facility type is significantly different from at least one other.

Use all five steps. If appropriate, compute and interpret omega squared.

Race		
White (x_1)	Black (x_2)	Other (x_3)
5.13	6.03	1.74
5.00	5.90	1.70
4.98	6.10	2.00
5.20	6.18	1.61
5.10	6.00	$n_3 = 4$
5.15	$n_2 = 5$	
$n_1 = 6$		

19. Are there race differences among juvenile defendants with respect to the length of time it takes them to acquire pretrial release? The data set *JDCC for Chapter 12.sav* (**edge.sagepub.com/gau3e**) can be used to test for whether time-to-release varies by race for juveniles accused of property crimes. The variables are *race* and *days*. Using SPSS, run an ANOVA with *race* as the IV and *days* as the DV. Select the appropriate post hoc tests.
    a. Identify the obtained value of *F*.
    b. Would you reject the null at an alpha of .01? Why or why not?
    c. State your substantive conclusion about whether there is a relationship between race and days to release for juvenile property defendants.
    d. If appropriate, interpret the post hoc tests to identify the location and total number of significant differences.
    e. If appropriate, compute and interpret omega squared.

20. Are juvenile property offenders sentenced differently depending on the file mechanism used to waive them to adult court? The data set *JDCC for Chapter 12.sav* (**edge.sagepub.com/gau3e**) contains the variables *file* and *jail*, which measure the mechanism used to transfer each juvenile to adult court (discretionary, direct file, or statutory) and the number of months in the sentences of those sent to jail on conviction. Using SPSS, run an ANOVA with *file* as the IV and *jail* as the DV. Select the appropriate post hoc tests.
    a. Identify the obtained value of *F*.
    b. Would you reject the null at an alpha of .05? Why or why not?
    c. State your substantive conclusion about whether there is a relationship between attorney type and days to release for juvenile defendants.
    d. If appropriate, interpret the post hoc tests to identify the location and total number of significant differences.
    e. If appropriate, compute and interpret omega squared.

21. The data set *FISS for Chapter 12.sav* (**edge.sagepub.com/gau3e**) contains the FISS variables capturing shooters' intentions (accident, assault, and police involved) and victims' ages. Using SPSS, run an ANOVA with *intent* as the IV and *age* as the DV. Select the appropriate post hoc tests.
    a. Identify the obtained value of *F*.
    b. Would you reject the null at an alpha of .05? Why or why not?
    c. State your substantive conclusion about whether victim age appears to be related to shooters' intentions.
    d. If appropriate, interpret the post hoc tests to identify the location and total number of significant differences.
    e. If appropriate, compute and interpret omega squared.

KEY TERMS

Analysis of variance (ANOVA) 281
Familywise error 281
Between-group variance 282
Within-group variance 282
F statistic 282
F distribution 282
Post hoc tests 286
Omega squared 297
Tukey's honest significant difference (HSD) 298
Bonferroni 298

GLOSSARY OF SYMBOLS AND ABBREVIATIONS INTRODUCED IN THIS CHAPTER

F	The statistic and sampling distribution for ANOVA
n_k	The sample size of group *k*
N	The total sample size across all groups
$\bar{x}_k$	The mean of group *k*
$\bar{x}_G$	The grand mean across all cases in all groups
SS_B	Between-groups sums of squares; a measure of true group effect
SS_W	Within-groups sums of squares; a measure of error
SS_T	Total sums of squares; equal to $SS_B + SS_w$
MS_B	Between-group mean square; the variance between groups and a measure of true group effect
MS_W	Within-group mean square; the variance within groups and a measure of error
ω^2	A measure of association that indicates the proportion of the total variability that is due to between-group differences