PART III Analysis of Variance

This section covers a family of tests frequently used in psychology called analysis of variance, abbreviated as ANOVA. ANOVAs are difference tests, like *t* tests. Like *t* tests, they take a variety of forms, depending on the groups being studied. But they are unlike *t* tests in one important way. *t* tests are limited to a maximum of two samples, so they are great for classic studies that have an experimental group and control group. ANOVA can handle three or more groups. So, for example, it can be used with a study where people with headaches are randomly assigned to receive aspirin, acetaminophen, or ibuprofen to see which medication works most quickly.

ANOVAs can also be used for studies with two or more explanatory variables like the one below, in which right-handed people are tested in a driving simulator to see how which hand they have on the wheel and which eye they have on the road affect performance.

	Both Eyes Open	Right Eye Covered	Left Eye Covered
Use both hands	2 eyes, 2 hands	L eye, 2 hands	R eye, 2 hands
Use right hand	2 eyes, R hand	L eye, R hand	R eye, R hand
Use left hand	2 eyes, L hand	L eye, L hand	R eye, L hand

Analyses of variance allow behavioral scientists to ask, and to answer, complex questions, the type of questions that must be posed to understand something as complex as how humans and animals think, feel, and behave.

Between-Subjects, One-Way Analysis of Variance

10

LEARNING OBJECTIVES

- Explain when ANOVA is used and how it works.
- Complete a between-subjects, one-way ANOVA.
- Interpret the results of a between-subjects, one-way ANOVA.

CHAPTER OVERVIEW

Chapters 8 and 9 covered two-sample *t* tests, tests that compare the means of two groups to see if they have a statistically significant difference. *t* tests are great for experiments with just two groups, like classic experiments with a control group and an experimental group. A *t* test would work well, for example, to see if a medication works better than a placebo in treating an illness. However, experiments that address complex questions often require more than two groups. Finding the best way to treat an illness might involve comparing three different medications and three different doses of each medication, ending up with nine different groups. The statistical technique that compares means when there are more than two groups is called analysis of variance.

10.1 Introduction to Analysis of Variance

Analysis of Variance Terminology

Analysis of variance, called **ANOVA** for short, is a family of statistical tests used for comparing the means of two or more groups. This chapter focuses on **between-subjects, one-way ANOVA**. Between-subjects, one-way ANOVA is an extension of the independent-samples *t* test, so it is used to compare means when there are two or more *independent* samples.

- **Between-subjects** is ANOVA terminology for independent samples.
- **Way** is ANOVA terminology for an explanatory variable. Ways can either be independent variables or grouping variables. (Independent variables are controlled by the experimenter, who assigns subjects to groups. Grouping variables are naturally occurring characteristics used to classify subjects into different groups.)

- A one-way ANOVA has one explanatory variable, either an independent variable or a grouping variable. A two-way ANOVA would have two explanatory variables, a three-way ANOVA would have three, etc.
- Though the explanatory variable in ANOVA is either a grouping variable or an independent variable, it becomes tedious to use both terms. Most statisticians get a bit casual with language and simply call it an independent variable; we'll continue calling them, generically, explanatory variables. Of course, when it is an independent variable, we'll call it that.
- An explanatory variable in ANOVA is also called a **factor,** so a one-way ANOVA can also be called a one-factor ANOVA.
- **Level** is the term in ANOVA for a category of an explanatory variable. The grouping variable sex, for example, has two levels—male and female.

To clarify all this terminology, here's a question where a between-subjects, one-way ANOVA would be used: Is there a difference in artistic ability among right-handed, left-handed, and ambidextrous people? This question could be answered by gathering a sample of people and classifying them into three groups: (1) right-handed people, (2) left-handed people, and (3) ambidextrous people.

Now there are three samples. The samples are independent samples as who is in one sample does not control or determine who is in another sample.

Next, artistic ability is measured with an interval-level scale. Because the dependent variable, artistic ability, is measured at the interval level, the mean level for each of the three groups can be calculated. Means can also be calculated for ratio-level variables, so ANOVA may be used when the outcome variable is measured at the interval or ratio level.

Table 10.1 illustrates this experiment with one cell for each sample. There is one row with three columns. The row represents the explanatory variable of handedness, what in ANOVA terminology is called the *way* or the *factor.* The columns represent the three *levels* of the explanatory variable: right-handed, left-handed, and ambidextrous.

TABLE 10.1 Comparing Artistic Ability in Three Independent Samples

Level 1	Level 2	Level 3
Right-handed	Left-handed	Ambidextrous

A one-way ANOVA has one factor with multiple levels. Here, the factor (explanatory variable) is handedness, and the three levels are right-handed, left-handed, and ambidextrous. Notice how the design is diagrammed as one row with three cells. In diagrams of ANOVA designs, each cell represents a sample. This design has three samples.

Why ANOVA Is Needed

t tests compare means between groups, so why is analysis of variance even needed? ANOVA is needed in order to keep the risk of Type I error at a reasonable level when comparing means of multiple groups. (Remember: Type I error occurs when a researcher erroneously concludes that there is a statistically significant difference.)

To see why this is a problem, consider a study with five conditions—for example, four different medications and a placebo being tested in treating some disease. It would be possible to analyze the data from these five conditions using a series of *t* tests. For example, a researcher could compare the mean of Condition 1 to Condition 2, the mean of Condition 1 to Condition 3, the mean of Condition 1 to Condition 4, and so on as shown in Table 10.2. This would require completing 10 *t* tests.

TABLE 10.2 All Possible Pairs of Five Experimental Groups

1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5
2 vs. 3	2 vs. 4	2 vs. 5	
3 vs. 4	3 vs. 5		
4 vs. 5			

If a study has five experimental conditions and if each condition is compared to each other, a researcher would need to complete 10 separate *t* tests. If each individual *t* test has a 5% chance of making a Type I error, there's almost a 50% chance that 1 of the 10 would have a Type I error.

It is tedious to do 10 *t* tests, but that is not why ANOVA is preferred. The real problem is that the likelihood of making a Type I error increases as the number of *t* tests increases. Scientists usually set alpha, the probability of a Type I error, at .05, so that they have a 5% chance of making this error. But, with 10 separate tests, the *overall* alpha level rises to be close to 50%. This means that there is close to a 50% chance that 1 of the 10 *t* tests will reach the wrong conclusion, rejecting the null hypothesis when it is true. Those are not good odds. And, the experimenter won't know which, if any, of the statistically significant results is erroneous.

Analysis of variance solves the problem of having a large risk of Type I error, what statisticians call runaway alpha. With ANOVA, one test is completed with a specified chance of Type I error. As with *t* tests, the alpha level, the chance of committing a Type I error, is usually set at .05. One test—the ANOVA—compares all the means at once and determines if any two of the means have a statistically significant difference. If the ANOVA is statistically significant, then the researcher performs what is called a post-hoc test. A **post-hoc test** is a follow-up test, engineered to find out which pairs of means differ while keeping the overall alpha level at a specified level, again, usually .05.

Here's a metaphor for how analysis of variance and post-hoc tests work together. Imagine standing outside of a football stadium on the day of a big game and hearing a loud roar. From the roar, one would know that something significant has happened in the game. That's analysis of variance—it indicates, in general, that something interesting has happened, but it doesn't state specifically what happened. To find out what has happened at the game, one needs to buy a ticket and go into the stadium. Going into the stadium is like doing a post-hoc test. Post-hoc tests are only conducted in analysis of variance when one is sure there is something interesting to be found.

What ANOVA Does

With this background on ANOVA, let's learn why it is called analysis of variance and how it works. ANOVA works by analyzing a set of scores and separating out the different sources of variability in the scores.

To understand this, imagine investigating the effect of alcohol on intoxication. Suppose research participants come to a laboratory where each person consumes one beer. After 20 minutes, researchers measure the level of intoxication by observing the effects of the alcohol on each person's performance on a behavioral task, say, walking a straight line. The higher the intoxication score, the poorer is the ability to walk in a straight line.

Figure 10.1 shows the expected results—not everyone would have exactly the same intoxication score. Rather, there is variability in the scores—with some people acting quite intoxicated, some people not acting at all intoxicated, and most clustered around the average score. Even though everyone received *exactly* the same dose of alcohol, not everyone reacted in *exactly* the same way. This variability *within* a group that receives the same treatment is called **within-group variability.** Within-group variability is primarily caused by individual differences, attributes that vary from case to case. So, how much one weighs, how recently one has eaten, and how much prior experience one has had with alcohol will all affect a person's intoxication score. These are all individual difference factors.

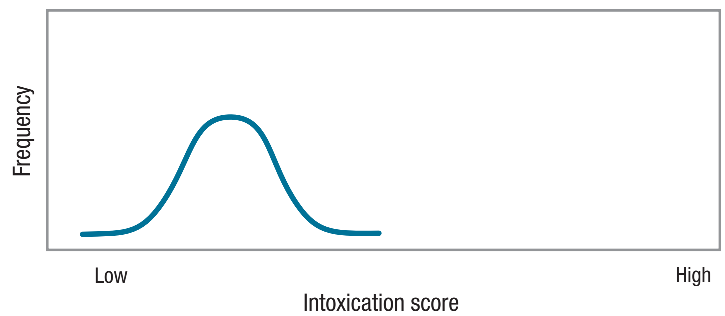

Figure 10.1 Distribution of Intoxication Scores for Participants Who Consumed One Beer Even though all participants received exactly the same dose of alcohol, there was variability in how it affected them. Some people exhibited very little intoxication and other people showed more. This variability within the group is accounted for by individual differences such as sex, weight, time since last meal, and prior experience with alcohol.

Within-group variability can be reduced by making the sample more homogeneous, but the effect of individual differences can't be eliminated entirely. For example, if all participants were men who weighed 175 pounds, had eaten dinner 30 minutes ago, and had been consuming alcohol regularly for over a year, there would still be variability within that group on the intoxication scores.

Within-group variability is one type of variability in analysis of variance. The other type of variability for a one-way ANOVA is *between-group variability.* **Between-group variability** is variability that is due to the different "treatments" that the different groups receive.

Figure 10.2 shows the distribution of intoxication scores for two groups—one group where each participant drank one beer and one group where each participant drank a six-pack. Individual differences explain the variability within a group, but the different doses of alcohol explain the differences *between* groups, why one group is more intoxicated than the other. This is called the **treatment effect** because it refers to the different ways that groups are treated. The treatment effect shows up as an impact on the outcome variable (here, the intoxication score) and is associated

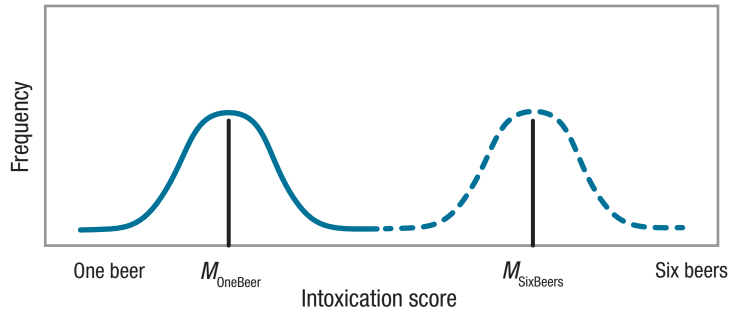

Figure 10.2 Distributions of Intoxication Scores for Participants Who Consumed Different Amounts of Alcohol The curve on the left is the distribution of intoxication scores for participants who consumed one beer. The curve on the right is the distribution of scores for participants who consumed six beers. Note that within each group there is variability in the intoxication level, variability due to individual differences. There is also variability *between* the two groups. The easiest way to see this is to observe that the two curves have different midpoints, with one group having a higher average score than the other.

with the explanatory variable (here, the dose of alcohol, which is controlled by the experimenter).

Between-group variability in one-way ANOVA is made up of two things: the treatment effect and individual differences. We've already covered how treatment plays a role in between-group variability. Now, let's see how individual differences play a role in between-group variability.

Imagine a large group of people randomly divided into two groups. Because of random assignment, the two groups should be fairly similar in terms of sex, weight, time since last meal, experience with alcohol, and so on. Now, each person in each group consumes the same dose of alcohol and is measured for intoxication level. Both groups are similar in terms of their characteristics and receive exactly the same treatment. Will the mean intoxication scores of the two groups be exactly the same? No. Because individual differences exist, the two groups will have slightly different means. Between-group variability is due *both* to individual differences and treatment effect.

How ANOVA Uses Variability

To understand how ANOVA uses within-group variability and between-group variability to see if there is a statistically significant difference, look at the two panels in Figure 10.3. Each panel represents groups that are randomly assigned to receive three different treatments for some illness. Treatment, the explanatory variable, has three levels. The top panel (A) depicts an outcome where there is little impact of treatment on outcome.

The top panel shows little effect of the independent variable because the three means (M_1, M_2, and M_3) are very close to each other. In contrast, the bottom panel (B), where the means (M_4, M_5, and M_6) are far apart, shows that treatment has an impact on outcome because the different treatments lead to dramatically different outcomes.

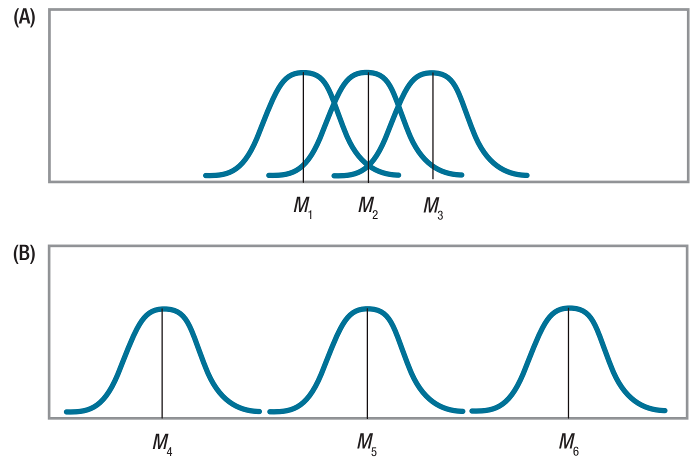

Figure 10.3 Examples of Little Impact of Treatment on Outcome and a Lot of Impact of Treatment on Outcome In the top panel (A), the three different treatments have about the same average effect on outcome. Note that the three means (M_1 to M_3) are very close to each other and that there is a lot of overlap in outcome from group to group. In contrast, the bottom panel (B) shows three treatments with very different effects on outcome. In it, the three means (M_4 to M_6) are further apart from each other and there is no overlap in outcome from group to group.

Analysis of variance could be used to analyze these results. ANOVA would show that the results in the top panel are not statistically significant, while the results in the bottom panel are statistically significant. How does analysis of variance lead to these conclusions?

Look at the means in each panel. Note that there is little variability among the means in the top panel and a lot of variability among the means in the bottom panel. Little variability exists in the top panel as all the means are close to each other. The greater distance between the means in the bottom panel indicates more variability between the means there. In the language of ANOVA, there is more *between-group variability* when the effect of treatment is large (the bottom panel) than when the effect of treatment is small (the top panel).

One-way analysis of variance calculates the ratio of between-group variability to within-group variability.

Figure 10.4 shows how the total variability in the data is partitioned into between-group variability and within-group variability. To decide if the amount of between-group variability is large or small, ANOVA compares it to within-group variability. One-way analysis of variance calculates the ratio of between-group variability to within-group variability. This is called an F ratio, in honor of Sir Ronald Fisher, who developed the procedure.

$$F = \frac{\text{Between Group Variability}}{\text{Within Group Variability}}$$

The F ratio works because within-group variability is made up of individual differences, while between-group variability includes treatment effect *and* individual differences. So, the F ratio could be rewritten as

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

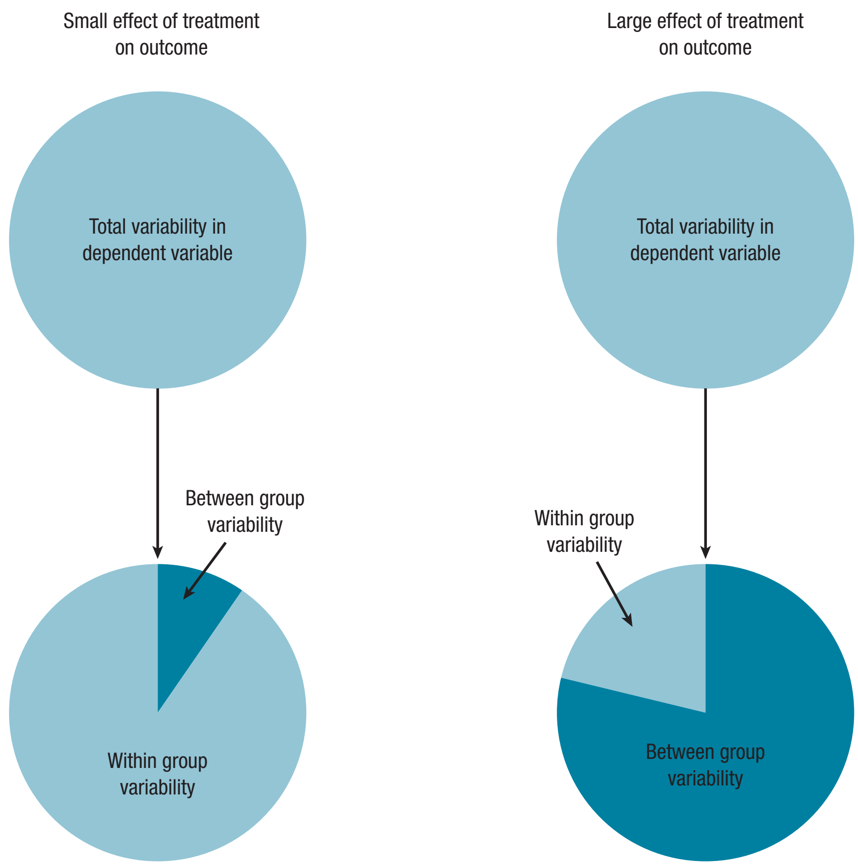

Figure 10.4 Partitioning Variability into Between-Group and Within-Group Variability The panel on the left shows little effect of treatment on outcome as only a small piece of the total variability is explained by differences between groups. In contrast, in the panel on the right, where a large piece of the total variability is accounted for by differences between groups, there is a large effect of treatment on outcome.

Here's what the *F* ratio, also known just as *F*, means:

- If there is no treatment effect, then there is no variability due to treatment and the variability indicated by the numerator of the *F* ratio is due only to individual differences.
- As a result, the *F* ratio has the same numerator (individual differences variability) and denominator (individual differences variability), so it will equal 1.
- As the effect of treatment grows, the numerator becomes larger than the denominator, and the *F* ratio climbs above 1. (Remember, treatment effect refers to the impact of the explanatory variable.)
- As the *F* ratio increases, as it climbs higher above 1, the results are more likely to be statistically significant.
- As variability is never negative, the *F* ratio can't go below 0.

Figure 10.5 gives an example of what the *F* distribution looks like. Note that it starts at 0, has a mode near 1, and tails off to the right. *F* gets bigger when there is more between-group variability than within-group variability.

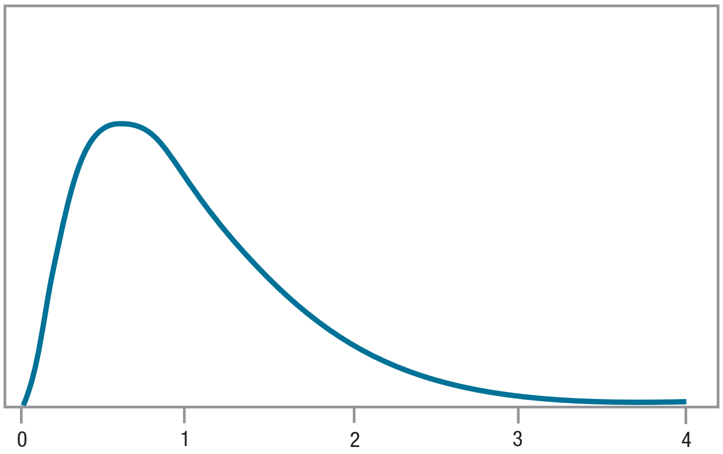

Figure 10.5 An Example of a Sampling Distribution for the *F* Ratio This is an example of a sampling distribution for the *F* ratio, the ratio of between-group variability to within-group variability. Note that *F* can't be lower than zero, that the high point of the curve is close to a value of 1 on the *X*-axis, that the distribution is positively skewed, and that the probability decreases as *F* gets larger than 1.

A Common Question

Q Looking at the ratio of between-group variability to within-group variability is a clever way to see if means differ. Could an ANOVA be used instead of a *t* test when there are just two groups?

A Yes. ANOVA can be used when comparing *two* or more means. In fact, a *t* test is just a variation on ANOVA. If a researcher has two groups, calculates a *t* value, and then squares the *t* value, it will be equal to the *F* ratio obtained by analyzing the same data with an ANOVA.

Practice Problems 10.1

Review Your Knowledge

10.1 What makes up within-group variability?

10.2 What makes up between-group variability?

10.3 An *F* ratio is a ratio of what divided by what?

10.4 When is a post-hoc test used?

10.2 Calculating Between-Subjects, One-Way ANOVA

Imagine that Dr. Chung, a psychologist who studies learning and motivation, designed a study to learn whether rats could discriminate among different types of food and if this discrimination influenced their behavior. He built a large and complex maze and trained 10 rats to run it. A rat would be placed in the start box and had to find its way to the goal box. To motivate the rats, food was placed in the goal box. Dr. Chung trained the rats with three different types of food: a low-calorie food, a normal-calorie food, and a high-calorie food. Each rat received an equal number of trials with each food and the rats were trained until they all ran the maze equally quickly.

Dr. Chung then put all the rats on a diet until they lost 10% of their normal body weight. The purpose of this was to increase their motivation to find food. Up until this point, all the rats had received the same treatment. Now, Dr. Chung implemented his experimental manipulation.

He randomly assigned the rats to three groups based on the type of food they would find in the goal box the next time they ran the maze: low-calorie, medium-calorie, or high-calorie. Each rat, individually, was placed in the maze and allowed to approach the goal box. When the rat got there, however, it found a screen that prevented it from entering. The rat could see and smell the food through the screen, but it couldn't get to the food. The purpose of this was to inform the rat of the type of food that awaited it.

Dr. Chung picked up the rat, removed the screen, and placed the rat in the start box. He then timed, in seconds, how long it took the rat to get to the goal box. This is the dependent variable in his study and the fewer seconds it took the rat to get to the goal box, the faster the rat traveled the maze. Dr. Chung figured that if the different caloric contents of the foods had been recognized by the rats and if hungry rats were more motivated by higher-calorie foods, then the time taken to run the maze should differ among the three groups. The data and the means for the three groups are shown in Table 10.3.

TABLE 10.3 Time to Run Maze (in seconds)

	Low-Calorie Food	Medium-Calorie Food	High-Calorie Food	
	30	28	24	
	31	29	25	
	32	27	26	
		28		Grand
Σ	93.00	112.00	75.00	280.00
n	3	4	3	10
M	31.00	28.00	25.00	28.00
s	1.00	0.82	1.00	2.58

These data represent the dependent variable, the number of seconds it took the 10 rats, randomly assigned to three different conditions, to travel through a maze in order to reach food in the goal box. The independent variable, the caloric density of the food, has three levels: low, medium, and high. The final column, labeled "Grand," provides information for subjects from all groups combined.

The average was 31.00 seconds to get to the low-calorie food, 28.00 seconds for the medium-calorie food, and 25.00 seconds for the high-calorie food (Figure 10.6). It appears as if mean speed increases as calorie content goes up, but the effect is not dramatic. It seems possible that the differences among the three groups may be explained by sampling error. It is also possible that the calorie content of the food in the goal box has had an effect. A statistical test is needed to decide between these two options.

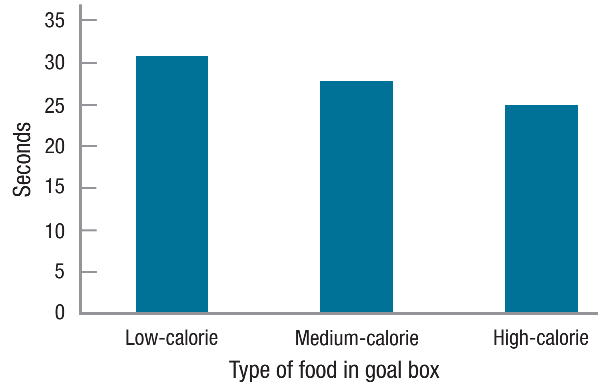

Figure 10.6 The Effect of Calorie Content on Maze Travel Time There appears to be some decrease in the mean time it takes a rat to run the maze (i.e., an increase in speed) as the calorie content of the food in the goal box increases. But, the effect looks modest at best. A between-subjects, one-way ANOVA can be used to see if the effect is a statistically significant one or if the differences are due to sampling error.

Step 1 Pick a Test

The first step of hypothesis testing is picking the right test. This study compares the means of three independent samples. There is one independent variable (calorie content), which has three levels (low, medium, and high). The dependent variable, seconds, is measured at the ratio level, so means can be calculated. This study calls for a between-subjects, one-way ANOVA.

Don't forget: **T**om **a**nd **H**arry **d**espise **c**rabby **i**nfants is a mnemonic to remember the six steps of hypothesis testing.

Step 2 Check the Assumptions

The assumptions for the between-subjects, one-way ANOVA, listed in Table 10.4, are the same as they are for an independent-samples t test: (1) each sample should be a random sample from its population, (2) the cases should be independent of each other, (3) the dependent variable should be normally distributed in each population, and (4) each population should have the same degree of variability. The robustness of the assumptions is the same as for t. The only nonrobust assumption is the second one, the assumption of independence of observations within each group.

TABLE 10.4 Assumptions for One-Way ANOVA

	Assumption	Robustness
1. Random samples	Each sample is a random sample from its population.	Robust
2. Independence of cases	Each case is not influenced by other cases in the sample.	Not robust
3. Normality	The dependent variable is normally distributed in each population.	Robust, especially if sample size (N) is large and the n's are about equal.
4. Homogeneity of variance	The degree of variability in the two populations is equivalent.	Robust, especially if sample size (N) is large and the n's are about equal

There's no hard and fast rule as to what a "large" sample size is. Some statisticians say 25 or more, some 35, and some 50.

Here is an evaluation of the assumptions for the maze data:

1. *Random samples.* Though this is an experimental study in which cases were randomly assigned to experimental conditions, the initial sample wasn't a random sample from the population of rats, so this assumption was violated. The assumption is robust, however, so Dr. Chung can proceed. However, he has to be careful about generalizing beyond the specific strain of rats he's working with.
2. *Independence of cases.* Each rat was trained and tested individually. Each rat only provided one data point for the final test. So, this assumption was not violated.
3. *Normality.* It seems reasonable to assume that, within a population of rats, running speed is normally distributed. That is, for rats, there is a mean speed around which most animals cluster and the number of rats with higher and lower speeds tails off in both directions symmetrically.
4. *Homogeneity of variance.* A look at the three standard deviations (1.00, 0.82, and 1.00) shows that they are all about the same. There is no reason to believe that this assumption has been violated.

With no nonrobust assumptions violated, Dr. Chung can proceed with the planned between-subjects, one-way ANOVA.

Step 3 List the Hypotheses

The hypotheses are statements about the populations. In Dr. Chung's study, there are three samples, with each one thought of as having come from a separate population.

His hypotheses will be about the population of rats that finds low-calorie food in the goal box, the population of rats that finds medium-calorie food, and the population of rats that finds high-calorie food.

Look at Figure 10.5, the example of a sampling distribution for an F ratio, and notice two things—there are no negative values on the x-axis and the sampling distribution is positively skewed. It looks like there is only one "tail" to the sampling distribution. This doesn't mean that all ANOVAs are one-tailed; it means just the opposite. Both "tails" are wrapped into this one side, so ANOVA always has nondirectional hypotheses. One doesn't have to decide between a one-tailed or two-tailed test. ANOVA is always two-tailed.

The nondirectional null hypothesis will state that no difference exists in the means of the populations. The generic form of this, where there are k samples, is

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$$

For the maze study, with three populations, Dr. Chung's null hypothesis is

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

The alternative hypothesis, also nondirectional, says not all population means are equal. It is not written as $\mu_1 \neq \mu_2 \neq \mu_3$ because that doesn't cover options like Populations 1 and 2 having different means but Populations 2 and 3 don't. The alternative hypothesis states that at least two of the population means are different, and maybe all three are. There is no easy way to write this mathematically, so Dr. Chung will write it as

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

Step 4 Set the Decision Rule

This step finds the critical value of F, abbreviated F_{cv}, the value that separates the rare zone from the common zone of the sampling distribution of the F ratio. Figure 10.7 is an example of a sampling distribution of F with the rare and common zones marked.

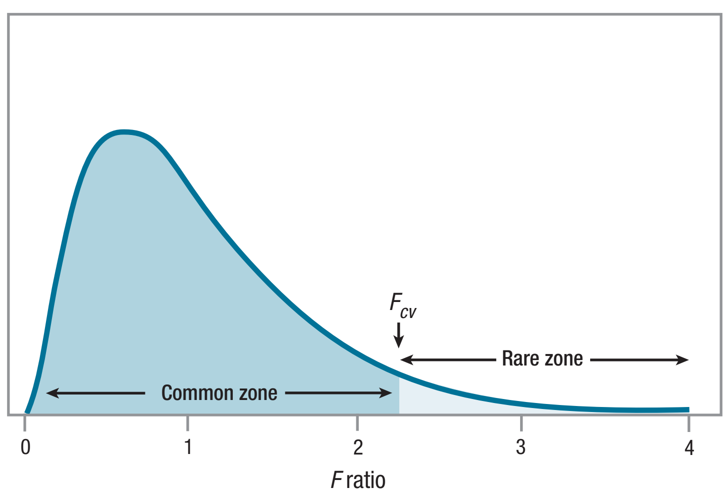

Figure 10.7 Using the Critical Value of F to Mark the Common Zone and the Rare Zone of the Sampling Distribution of F The critical value of F, F_{cv}, marks the boundary between the common zone and the rare zone of the test statistic. Here, 5% of the area under the curve falls in the rare zone that is on and to the right of F_{cv}.

The decision rule will sound familiar as it is similar to what we used for the *z* test and *t* tests: if the calculated value of the test statistic, *F*, falls on the line or in the rare zone, the null hypothesis is rejected.

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

To obtain the value of F_{cv}, use Appendix Table 4. It provides two F_{cv} tables: one for critical values of *F* with a 5% chance of making a Type I error (i.e., $\alpha = .05$) and one for a 1% chance of making a Type I error ($\alpha = .01$). A portion of the table with the most commonly used alpha level, .05, is shown in **Table 10.5**.

TABLE 10.5 Part of Appendix Table 4, Critical Values of *F* (F_{cv})

Denominator Degrees of Freedom	Numerator Degrees of Freedom					
	1	2	3	4	5	6
1	161.448	199.500	215.707	224.583	230.162	233.986
2	18.513	19.000	19.164	19.247	19.296	19.330
3	10.128	9.552	9.277	9.117	9.013	8.941
4	7.709	6.944	6.591	6.388	6.256	6.163
5	6.608	5.786	5.409	5.192	5.050	4.950
6	5.987	5.143	4.757	4.534	4.387	4.284
7	5.591	4.737	4.347	4.120	3.972	3.866
8	5.318	4.459	4.066	3.838	3.687	3.581
9	5.117	4.256	3.863	3.633	3.482	3.374
10	4.965	4.103	3.708	3.478	3.326	3.217
11	4.844	3.982	3.587	3.357	3.204	3.095
12	4.747	3.885	3.490	3.259	3.106	2.996
13	4.667	3.806	3.411	3.179	3.025	2.915
14	4.600	3.739	3.344	3.112	2.958	2.848
15	4.543	3.682	3.287	3.056	2.901	2.790

The critical value of *F*, F_{cv}, is found at the intersection of the column for the numerator degrees of freedom and the row for the denominator degrees of freedom.

To determine the critical value of *F*, find the value at the intersection of the *column* for the degrees of freedom in the numerator and the *row* for the degrees of freedom in the denominator. Equation 10.1 shows how to calculate the different degrees of freedom needed for a between-subjects, one-way ANOVA. Recall the formula for *F* ratio from the previous section:

- The numerator degrees of freedom for the *F* ratio, what defines the columns in the table of critical values of *F*, are called between-groups degrees of freedom. This is abbreviated as df_{Between}.
- The denominator degrees of freedom for the *F* ratio, what defines the rows in the table of critical values of *F*, are within-groups degrees of freedom. This is abbreviated as df_{Within}.

- The final degrees of freedom calculated in Equation 10.1 are total degrees of freedom. This is abbreviated as df_{Total} and represents the total number of degrees of freedom in the data. Note that adding together df_{Between} and df_{Within} equals df_{Total}.

Equation 10.1 Degrees of Freedom for Between-Subjects, One-Way ANOVA

$$df_{\text{Between}} = k - 1$$

$$df_{\text{Within}} = N - k$$

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

where df_{Between} = between-groups degrees of freedom (degrees of freedom for the numerator)
df_{Within} = within-groups degrees of freedom (degrees of freedom for the denominator)
df_{Total} = total degrees of freedom
k = the number of groups
N = the total number of cases

For the maze data, there are three groups (low-, medium-, and high-calorie), so $k = 3$. There are a total of 10 participants in the study, so $N = 10$. Given these values, degrees of freedom are calculated as

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

$$\begin{aligned} df_{\text{Within}} &= N - k \\ &= 10 - 3 \\ &= 7 \end{aligned}$$

$$\begin{aligned} df_{\text{Total}} &= N - 1 \\ &= 10 - 1 \\ &= 9 \end{aligned}$$

To find the critical value of F, Dr. Chung needs numerator degrees of freedom ($df_{\text{Between}} = 2$) and denominator degrees of freedom ($df_{\text{Within}} = 7$). The intersection of the column for 2 degrees of freedom and the row for 7 degrees of freedom leads to a critical value of F of 4.737. That means the decision rule for the maze data can now be written as:

- If $F \geq 4.737$, reject the null hypothesis.
- If $F < 4.737$, fail to reject the null hypothesis.

The sampling distribution of F with degrees of freedom of 2 (numerator) and 7 (denominator) is shown in Figure 10.8. The critical value of F, $F_{cv} = 4.737$, is used to separate the rare zone from the common zone. If the observed value of F, the value of the test statistic calculated in Step 5, falls on the line or in the rare zone, the null

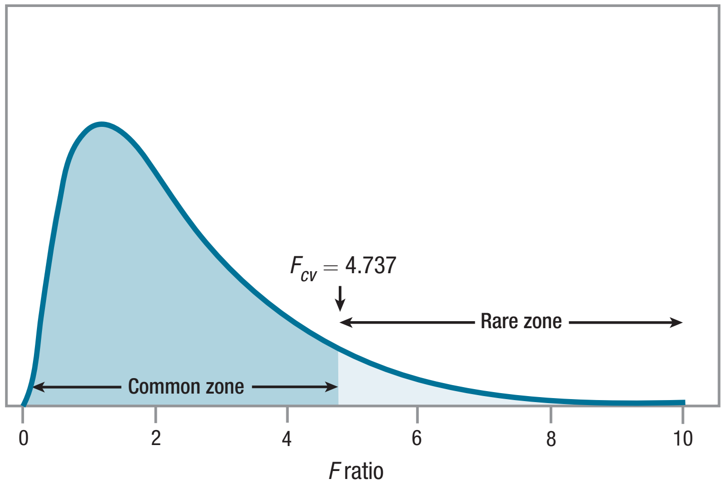

Figure 10.8 The Critical Value of F for 2 and 7 Degrees of Freedom The critical value of F, F_{cv}, with 2 degrees of freedom in the numerator and 7 degrees of freedom in the denominator, is 4.737. Note that 5% of the F distribution falls on or to the right of this point. If the observed value of F falls in the rare zone, then the null hypothesis is rejected.

hypothesis is rejected. If it falls in the common zone, Dr. Chung will fail to reject the null hypothesis.

Step 5 Calculate the Test Statistic

To complete the ANOVA, Dr. Chung's next step is to analyze the variability in the maze data. This means taking account of variability both within and between groups:

- Not each rat with the same type of food in the goal box coursed through the maze in the same amount of time, so there is variability within groups.
- The three groups all had different means, so there is variability between groups.

To understand why analysis of variance is called analysis of *variance,* it is useful to review the variance formula from Chapter 3. Equation 3.6 says:

$$s^2 = \frac{\Sigma(X - M)^2}{N - 1}$$

The *numerator* in the variance formula is calculated by following three steps: (1) calculating deviation scores by subtracting the mean from each score, (2) squaring all the deviation scores, and (3) adding up all the squared deviation scores. This numerator, a sum of squared deviation scores, is called a *sum of squares,* abbreviated as *SS*. Calculating sums of squares for the two sources of variability, between-groups and within-groups, is necessary to find the between- and within-group variances that are analyzed in an analysis of variance.

Actually, there are three sums of squares that are calculated. In addition to sum of squares between-groups (SS_{Between}) and sum of squares within groups (SS_{Within}), sum of squares total (SS_{Total}) is also calculated. Sum of squares total represents all the variability in the data. Between-subjects, one-way ANOVA divides SS_{Total} into between-groups sum of squares (SS_{Between}), which measures variability due to treatment, and within-groups sum of squares (SS_{Within}), which measures variability due to individual differences. In other words, $SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}}$.

SS_{Total} represents all the variability in the scores. Between-subjects, one-way ANOVA divides it into variability due to the different ways the groups are treated, and variability due to individual differences.

Sum of squares total is calculated by treating all the cases as if they belong in one group: the grand mean, the mean of all the scores, is subtracted from each score, these deviation scores are squared, and the squared deviation scores are added up. Voilà, a sum of squares, the numerator for a variance.

Sum of squares between represents the variability between groups. It is calculated by subtracting the grand mean from each group mean, squaring these deviation scores, multiplying each squared deviation score by the number of cases in the group, and then adding them up. The final variance numerator, the one that represents variability within groups, **sum of squares within**, is calculated by taking each score, subtracting from it its group mean, squaring each deviation score, and adding them all up.

The calculations just described are what are called definitional formulas because the calculations explain what the value being calculated is. Unfortunately, definitional formulas often are not straightforward mathematically. In contrast, computational formulas are designed to be easier to use. Equation 10.2 is a computational formula for sum of squares total.

Equation 10.2 Formula for Calculating Sum of Squares Total for Between-Subjects, One-Way ANOVA

$$SS_{\text{Total}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

where SS_{Total} = total sum of squares
X = raw score
N = the total number of cases

This formula says:

1. Square each score and add them all up.
2. Add up all the scores, square the sum, and divide by the total number of cases.
3. Subtract the result of Step 2 from the result of Step 1.

The easiest way to do this is to take Table 10.3 and add another column to it, one for squared scores. This can be seen in Table 10.6. From the table, we can see that the sum of squared scores for Step 1 is 7,900.00.

TABLE 10.6 Maze Running Time Data from Table 10.3, Squared and Summed, in Preparation for Computing Sums of Squares

	Low-Calorie		Medium-Calorie		High-Calorie		Grand	
	X	X^2	X	X^2	X	X^2	X	X^2
	30	900	28	784	24	576		
	31	961	29	841	25	625		
	32	1,024	27	729	26	676		
			28	784				
Σ	93	2,885	112	3,138	75	1,877	280	7,900
n	3		4		3		10	

Step 2 is next.

$$\begin{aligned}\frac{(\Sigma X)^2}{N} &= \frac{280^2}{10}\\ &= \frac{78,400.00}{10}\\ &= 7,840.00\end{aligned}$$

Finally,

$$\begin{aligned}SS_{\text{Total}} &= \Sigma X^2 - \frac{(\Sigma X)^2}{N}\\ &= 7,900.00 - 7,840.00\\ &= 60.00\end{aligned}$$

Next, let's use Equation 10.3 to calculate sum of squares between groups.

Equation 10.3 Formula for Calculating Between-Groups Sum of Squares for Between-Subjects, One-Way ANOVA

$$SS_{\text{Between}} = \Sigma\left(\frac{\left(\Sigma X_{\text{Group}}\right)^2}{n_{\text{Group}}}\right) - \frac{(\Sigma X)^2}{N}$$

where SS_{Between} = between-groups sum of squares
X_{Group} = raw scores for cases in a group
n_{Group} = number of cases in a group
X = raw scores
N = total number of cases

Here's how the formula works:

1. For each group, add up all the scores, square that sum, and divide that square by the number of cases in the group. Add up all these quotients.
2. Add up all the scores, square that sum, and divide that square by the total number of cases.
3. Subtract Step 2 from Step 1.

Again, most of the work has already been done in Table 10.6. Here is Step 1:

$$\begin{aligned}\Sigma\left(\frac{\left(\Sigma X_{\text{Group}}\right)^2}{n_{\text{Group}}}\right) &= \frac{93^2}{3} + \frac{112^2}{4} + \frac{75^2}{3}\\ &= \frac{8,649.00}{3} + \frac{12,544.00}{4} + \frac{5,625.00}{3}\\ &= 2,883.00 + 3,136.00 + 1,875.00\\ &= 7,894.00\end{aligned}$$

And, Step 2:

$$\begin{aligned}\frac{(\Sigma X)^2}{N} &= \frac{280^2}{10}\\ &= \frac{78{,}400.00}{10}\\ &= 7{,}840.00\end{aligned}$$

Finally,

$$\begin{aligned}SS_{\text{Between}} &= \Sigma\left(\frac{\left(\Sigma X_{\text{Group}}\right)^2}{n_{\text{Group}}}\right) - \frac{(\Sigma X)^2}{N}\\ &= 7{,}894.00 - 7{,}840.00\\ &= 54.00\end{aligned}$$

The final sum of squares to calculate is the sum of squares within. Equation 10.4 covers this.

Equation 10.4 Formula for Calculating Sum of Squares Within for Between-Subjects, One-Way ANOVA

$$SS_{\text{within}} = \Sigma\left(\Sigma X^2_{\text{Group}} - \frac{\left(\Sigma X_{\text{Group}}\right)^2}{n_{\text{Group}}}\right)$$

where SS_{Within} = sum of squares within
X_{Group} = raw scores for cases in a group
n_{Group} = number of cases in a group

1. For each group, square each score and add them all up.
2. For each group, add up all the scores, square the sum, and divide by the total number of cases.
3. For each group, subtract the result of Step 2 from the result of Step 1.
4. Add together all the remainders from Step 3.

As with the other sums of squares, Table 10.6 has the components already prepared:

$$\begin{aligned}SS_{\text{within}} &= \Sigma\left(\Sigma X^2_{\text{Group}} - \frac{\left(\Sigma X_{\text{Group}}\right)^2}{n_{\text{Group}}}\right)\\ &= \left(2{,}885 - \frac{93^2}{3}\right) + \left(3{,}138 - \frac{112^2}{4}\right) + \left(1{,}877 - \frac{75^2}{3}\right)\\ &= \left(2{,}885 - \frac{8{,}649.00}{3}\right) + \left(3{,}138 - \frac{12{,}544.00}{4}\right) + \left(1{,}877 - \frac{5{,}625.00}{3}\right)\\ &= (2{,}885 - 2{,}883.00) + (3{,}138 - 3{,}136.00) + (1{,}877 - 1{,}875.00)\\ &= 2.00 + 2.00 + 2.00\\ &= 6.00\end{aligned}$$

Earlier in the chapter, it was stated that $SS_{Total} = SS_{Between} + SS_{Within}$. Let's see if that is true. $SS_{Between}$ is 54.00 and SS_{Within} is 6.00. They sum to 60.00, which is what we calculated SS_{Total} to be. This is shown visually in Figure 10.9.

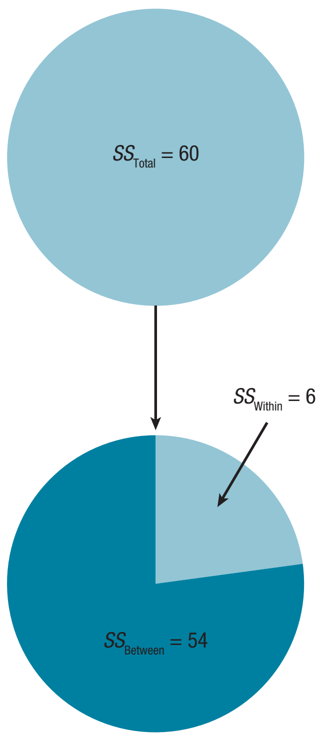

Figure 10.9 Partitioning Total Sum of Squares for Maze Running Time Data SS_{Total} can be broken down into its component parts, $SS_{Between}$ and SS_{Within}. Note that in this example the largest chunk of SS_{Total} is accounted for by $SS_{Between}$.

Now that the degrees of freedom and sums of squares have been calculated, Dr. Chung can go on to calculate the *F* ratio. By organizing the sums of squares and the degrees of freedom into an ANOVA summary table, the *F* ratio will almost calculate itself.

Table 10.7 is a template of a summary table for a between-subjects, one-way ANOVA. It is important to note the order in which it is laid out—three rows and five columns. Always arrange a summary table for a between-subjects, one-way ANOVA in exactly this way:

- There is one row for each source of variability:
  - Between-groups variability appears on the top row.
  - Within-groups variability is on the middle row.
  - Total variability goes on the bottom row.

TABLE 10.7 Template for ANOVA Summary Table for One-Way ANOVA

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups				
Within groups				
Total				

An ANOVA summary table for a one-way ANOVA is set up with these five columns and these three rows in this order.

- The five columns, in order, tell:
  - The source of variability (between, within, or total).
  - The sum of squares calculated for that source of variability.
  - The degrees of freedom for that source of variability.
  - What is called the "mean square" for that source of variability.
  - The *F* ratio.

Dr. Chung already has enough information to fill in the second and third columns. Mean square, the fourth column, sounds like something new, but it is a variance by another name. To calculate the mean square, one divides the sum of squares in a row by the degrees of freedom for that row.

To see how a mean square is a variance, let's revisit the variance formula from Chapter 3 one more time:

$$s^2 = \frac{\Sigma(X - M)^2}{N - 1}$$

The numerator in this formula is the same as the sum of squares total. And, df_{Total} is $N - 1$. Thus, if mean square total were calculated, it would be exactly the same as the variance for all the cases. Analysis of variance really does analyze variances.

The only mean squares needed to complete a between-subjects, one-way ANOVA are between-groups mean square (MS_{Between}) and within-groups mean square (MS_{Within}). As the mean square total is not needed, the convention is not to calculate it and to leave the space blank where mean square total would go.

The formula for calculating the between-groups mean square is given in Equation 10.5.

Equation 10.5 Formula for Between-Groups Mean Square, MS_{Between}, for Between-Subjects, One-Way ANOVA

$$MS_{\text{Between}} = \frac{SS_{\text{Between}}}{df_{\text{Between}}}$$

where MS_{Between} = between-groups mean square
SS_{Between} = between-groups sum of squares
df_{Between} = between-groups degrees of freedom

For the maze data, where $SS_{\text{Between}} = 54.00$ and $df_{\text{Between}} = 2$, the between-groups mean square would be calculated:

$$\begin{aligned} MS_{\text{Between}} &= \frac{SS_{\text{Between}}}{df_{\text{Between}}} \\ &= \frac{54.00}{2} \\ &= 27.00 \end{aligned}$$

Equation 10.6 offers the formula for the within-groups mean square.

Equation 10.6 Formula for Within-Groups Mean Square, MS_{Within}, for a Between-Subjects, One-Way ANOVA

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{df_{\text{Within}}}$$

where MS_{Within} = within-groups mean square
SS_{Within} = within-groups sum of squares
df_{Within} = within-groups degrees of freedom

For the maze data, $SS_{\text{Within}} = 6.00$ and $df_{\text{Within}} = 7$, so MS_{Within} would be calculated:

$$\begin{aligned} MS_{\text{Within}} &= \frac{SS_{\text{Within}}}{df_{\text{Within}}} \\ &= \frac{6.00}{7} \\ &= 0.8571 \\ &= 0.86 \end{aligned}$$

Once the two mean squares are calculated, it is time to calculate the *F* ratio. *F* is the ratio of variability due to between-group factors and individual differences divided by the variability due to individual differences. *F* is calculated as shown in Equation 10.7.

Equation 10.7 Formula for Calculating *F* for a Between-Subjects, One-Way ANOVA

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}$$

where F = the F ratio
MS_{Between} = the between-groups mean square
MS_{Within} = the within-groups mean square

For the maze data, where $MS_{\text{Between}} = 27.00$ and $MS_{\text{Within}} = 0.86$, F is calculated as

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{27.00}{0.86} = 31.3953 = 31.40$$

The complete ANOVA summary table for the maze data is shown in Table 10.8. Table 10.9 is a summary table that provides the organization and formulas for completing a between-subjects, one-way ANOVA.

TABLE 10.8 Completed ANOVA Summary Table for Maze Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Between groups	54.00	2	27.00	31.40
Within groups	6.00	7	0.86	
Total	60.00	9		

This shows what a complete ANOVA summary table looks like.

TABLE 10.9 How to Complete an ANOVA Summary Table for a Between-Subjects, One-Way ANOVA

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Between groups	Equation 10.3	$k - 1$	$\frac{SS_{\text{Between}}}{df_{\text{Between}}}$	$\frac{MS_{\text{Between}}}{MS_{\text{Within}}}$
Within groups	Equation 10.4	$N - k$	$\frac{SS_{\text{Within}}}{df_{\text{Within}}}$	
Total	Equation 10.2	$N - 1$		

This table summarizes all the steps necessary to complete an ANOVA summary table for a between-subjects, one-way ANOVA.

Worked Example 10.1

For practice with a between-subjects, one-way ANOVA, let's explore how effective pain relievers like acetaminophen (Tylenol) and ibuprofen (Advil) are. Dr. Douglas, a sensory psychologist, wanted to see if medications such as these affected the pain threshold (how long it takes to perceive a stimulus as painful). Twelve male undergraduates volunteered for the study, and she randomly assigned them to three groups, one control group and two experimental groups, but tested each person individually. Each participant was given a pill—either a placebo, ibuprofen, or acetaminophen—waited an hour for the medication to take effect, then placed his hand in a bucket of ice water. As this was a test of pain threshold, he was told to remove his hand from the ice water as soon as it became painful. The researcher recorded the elapsed time, in seconds, as shown in Table 10.10. Note that this table has squared and summed scores so that the ANOVA computational formulas can be easily completed.

TABLE 10.10 Results of Pain Threshold Test for Different Pain Relievers (in seconds)

	Placebo		Ibuprofen		Acetaminophen		Grand	
	X	X^2	X	X^2	X	X^2	X	X^2
	15	225	17	289	24	576		
	18	324	24	576	16	256		
	23	529	30	900	26	676		
	14	196	21	441	18	324		
Σ	70	1,274	92	2,206	84	1,832	246	5,312
n	4		4		4		12	
M	17.50		23.00		21.00		20.50	
s	4.04		5.48		4.76		4.95	

These values in the columns labeled X represent how many seconds it took, in the ice water, before a participant removed his hand because it had become painful. This table has been set up with X^2 values and sums so that it is prepared for use in calculating sums of squares.

Figure 10.10 shows what appears to be an effect of the pain relievers on the pain threshold. Compared to the placebo group, those using acetaminophen took longer on average to feel pain, and those given ibuprofen took even longer on average. It is necessary, of course, to conduct a hypothesis test to see if the effect is statistically significant and isn't plausibly accounted for by sampling error.

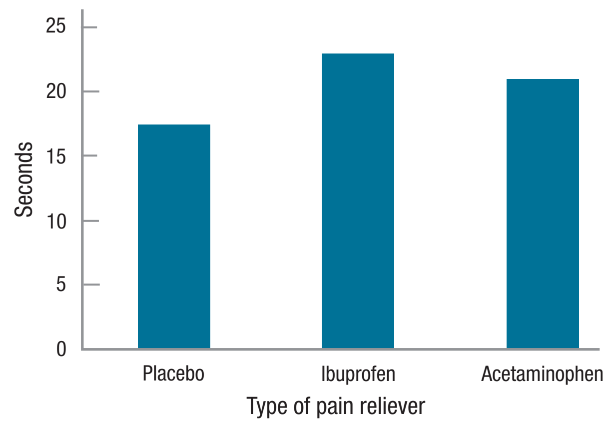

Figure 10.10 Mean Seconds to Mean Pain Threshold for Different Pain Relievers
It appears that ibuprofen, and perhaps acetaminophen, raise the pain threshold compared to placebo. But, a between-subjects, one-way ANOVA has to be conducted to make sure that the effect is statistically significant.

Step 1 Pick a Test

There are a number of conditions that lead to Dr. Douglas's selection of between-subjects, one-way ANOVA as the appropriate test to analyze these data.

- There is one independent variable (factor), the type of drug.
- This factor has three levels (placebo, ibuprofen, and acetaminophen), so there are three samples or groups.

- The cases were randomly assigned to the groups, so the samples are independent samples.
- The dependent variable, number of seconds to pain threshold, is measured at the ratio level so means can be calculated for each group.

Step 2 Check the Assumptions

1. Random samples: The samples are made up of male, college-student volunteers, so they are not random samples from the human population. The random samples assumption is violated, but it is a robust assumption so the researcher can proceed. She'll just have to be careful about the population to which she generalizes the results.
2. Independence of cases: Each participant was in the study only once. Plus, each participant was tested individually, uninfluenced by the other participants. This assumption was not violated.
3. Normality: Researchers are commonly willing to assume that psychological variables, like pain threshold, are normally distributed. So, this assumption is not violated.
4. Homogeneity of variance: All the standard deviations (4.04, 5.48, and 4.76, from Table 10.10) are very similar, so the amount of variability in each population seems about equal.

With no nonrobust assumptions violated, Dr. Douglas can proceed with the planned between-subjects, one-way ANOVA.

Step 3 List the Hypotheses

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

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

Step 4 Set the Decision Rule

In order to find the critical value of F, Dr. Douglas needs to know the degrees of freedom for the numerator (df_{Between}) and for the denominator (df_{Within}). To do so, she uses Equation 10.1:

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

$$\begin{aligned} df_{\text{Within}} &= N - k \\ &= 12 - 3 \\ &= 9 \end{aligned}$$

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

Looking at the $\alpha = .05$ version of Appendix Table 4—at the intersection of the column for 2 degrees of freedom in the numerator and the row for 9 degrees of freedom in the denominator—she finds F_{cv} is 4.256. Here is the decision rule:

- If $F \geq 4.256$, reject H_0.
- If $F < 4.256$, fail to reject H_0.

Step 5 Calculate the Test Statistic

Dr. Douglas's first step on the path to F is to calculate SS_{Total}, using Equation 10.2 and values from Table 10.10:

$$\begin{aligned}
SS_{\text{Total}} &= \Sigma X^2 - \frac{(\Sigma X)^2}{N} \\
&= 5{,}312 - \frac{246^2}{12} \\
&= 5{,}312 - \frac{60{,}516.00}{12} \\
&= 5{,}312 - 5043.00 \\
&= 269.00
\end{aligned}$$

Next, she calculates SS_{Between} following Equation 10.3:

$$\begin{aligned}
SS_{\text{Between}} &= \Sigma\left(\frac{(\Sigma X_{\text{Group}})^2}{n_{\text{Group}}}\right) - \frac{(\Sigma X)^2}{N} \\
&= \left(\frac{70^2}{4} + \frac{92^2}{4} + \frac{84^2}{4}\right) - \frac{246^2}{12} \\
&= \left(\frac{4{,}900.00}{4} + \frac{8{,}464.00}{4} + \frac{7{,}056.00}{4}\right) - \frac{60{,}516.00}{12} \\
&= (1{,}225.00 + 2{,}116.00 + 1{,}764.00) - 5{,}043.00 \\
&= 5{,}105.00 - 5{,}043.00 \\
&= 62.00
\end{aligned}$$

Finally, using Equation 10.4, she calculates SS_{Within}:

$$\begin{aligned}
SS_{\text{Within}} &= \Sigma\left(\Sigma X^2_{\text{Group}} - \frac{(\Sigma X_{\text{Group}})^2}{n_{\text{Group}}}\right) \\
&= \left(1{,}274 - \frac{70^2}{4}\right) + \left(2{,}206 - \frac{92^2}{4}\right) + \left(1{,}832 - \frac{84^2}{4}\right) \\
&= \left(1{,}274 - \frac{4{,}900.00}{4}\right) + \left(2{,}206 - \frac{8{,}464.00}{4}\right) + \left(1{,}832 - \frac{7{,}056.00}{4}\right) \\
&= (1{,}274 - 1{,}225.00) + (2{,}206 - 2{,}116.00) + (1{,}832 - 1{,}764.00) \\
&= 49.00 + 90.00 + 68.00 \\
&= 207.00
\end{aligned}$$

As a check that she has done the math correctly, she adds together SS_{Between} (62.00) and SS_{Within} (207.00) to make sure that they sum to sum of squares total (269.00). They do. Figure 10.11 shows this visually.

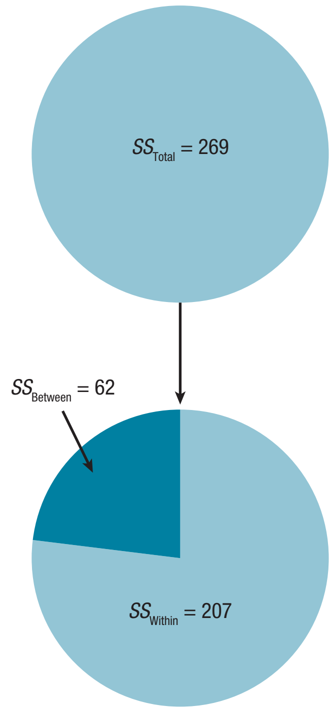

Figure 10.11 Partitioning Total Sum of Squares for Mean Pain Threshold Data
SS_{Total} can be broken down into its component parts, SS_{Between} and SS_{Within}. Note that the largest chunk of SS_{Total} is accounted for by SS_{Within}.

The final step is to complete an ANOVA summary table for the pain threshold data. The degrees of freedom and sums of squares already calculated can be used to start to fill in the summary table. Then, following the guidelines in Table 10.9, Dr. Douglas completes the ANOVA summary table (see Table 10.11), finding $F = 1.35$. In the next section, we'll learn how to interpret the results.

TABLE 10.11 ANOVA Summary Table for One-Way ANOVA for Mean Pain Threshold Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	62.00	2	31.00	1.35
Within groups	207.00	9	23.00	
Total	269.00	11		

This is the completed ANOVA summary table for the study examining the impact of different pain relievers on mean pain threshold.

Practice Problems 10.2

Apply Your Knowledge

10.5 Select the correct test for this example: A gerontologist wants to determine which exercise program, yoga or stretching, leads to greater limberness in elderly people. He uses a ratio-level range-of-motion test to measure limberness and randomly assigns elderly people to receive either eight weeks of yoga or eight weeks of stretching.

10.6 Select the correct test for this example: A human factors psychologist is comparing three different adhesives used in sealing cereal boxes to see which one is easiest to open. He obtained 90 consumers, randomly assigned each participant to open one cereal box, and measured how long it took to open the box. Thirty of the boxes were sealed with Adhesive A, 30 with Adhesive B, and 30 with Adhesive C.

10.7 If the numerator *df* for a between-subjects, one-way ANOVA are 4 and the denominator *df* are 20, what is the decision rule if $\alpha = .05$?

10.8 If a between-subjects, one-way ANOVA has 32 cases randomly assigned to four equally sized groups, what are df_{Between}, df_{Within}, and df_{Total}?

10.9 Here are data on an interval-level variable for three independent samples. Prepare this table so that it would be ready for use in computing sums of squares. (Do not compute the sums of squares.)

Group 1	Group 2	Group 3
16	12	13
17	14	15
20		18

10.10 Given the data in the table below, calculate (a) SS_{Total}, (b) SS_{Between}, and (c) SS_{Within}.

10.11 Given $df_{\text{Between}} = 3$, $df_{\text{Within}} = 12$, $df_{\text{Total}} = 15$, $SS_{\text{Between}} = 716.00$, $SS_{\text{Within}} = 228.00$, and $SS_{\text{Total}} = 944.00$, complete an ANOVA summary table.

	Group 1		Group 2		Group 3		Grand	
	X	X^2	X	X^2	X	X^2	X	X^2
	16	256	19	361	26	676		
	18	324	20	400	22	484		
	14	196	22	484	25	625		
	18	324	24	576	27	729		
Σ	66.00	1,100.00	85.00	1,821.00	100.00	2,514.00	251.00	5,435.00
n	4		4		4		12	

10.3 Interpreting Between-Subjects, One-Way ANOVA

The interpretation of a between-subjects, one-way ANOVA starts the same way it did for *t* tests, addressing questions of whether the null hypothesis was rejected and how big the effect is. After that, the interpretation answers a different question, where the effect is found.

Why is there a need to worry about where the effect is? This question didn't have to be addressed for a *t* test because in a *t* test just two groups exist. If the results of a *t* test were statistically significant, it was clear that the mean of Group 1 was different from the mean of Group 2. However, there can be more than two groups in an ANOVA. If the results of the ANOVA are statistically significant, then all that is known is that the mean of at least one group is different from at least one other group. If there are three groups, the mean of Group 1 could differ from the mean of Group 2, the mean of Group 1 could differ from the mean of Group 3, or the mean of Group 2 could differ from the mean of Group 3. It is also possible that two of the mean of the three pairs could differ or that all three pairs could differ. This is why it's important to address the question of where the effect is when interpreting a statistically significant ANOVA.

Was the Null Hypothesis Rejected?

Dr. Chung's maze study compared rats that expected to find different foods in the goal box of a maze. Ten rats were assigned to three groups—one that expected to find low-calorie food in the goal box, one that expected medium-calorie food, and one that expected high-calorie food. The rats expecting low-calorie food took 31 seconds from start box to goal box, the rats expecting medium-calorie food took 28 seconds, and the rats expecting high-calorie food took 25 seconds. Dr. Chung had determined $F_{cv} = 4.737$ and formulated the decision rule:

- If $F \geq 4.737$, reject the null hypothesis.
- If $F < 4.737$, fail to reject the null hypothesis.

As shown in the ANOVA summary table, Table 10.8, $F = 31.40$ for the maze data. Applying the decision rule, it is clear that 31.40 is greater than or equal to 4.737, and Figure 10.12 shows how the observed value of F falls in the rare zone. This means that the null hypothesis is rejected and the alternative hypothesis accepted. This leads to the conclusion that at least one of the three groups—rats expecting low-calorie food, rats expecting medium-calorie food, or rats expecting high-calorie food—differs from at least one of the others in terms of the mean time it takes to get to the food. It is not yet clear where the difference lies, so the best interpretation Dr. Chung can make at present is: "The results of the between-subjects, one-way ANOVA were statistically significant [$F(2, 7) = 31.40$, $p < .05$], indicating that there is a statistically significant difference in the time it takes to get from the start box to the goal box depending on the calorie content of the food a rat expects to find."

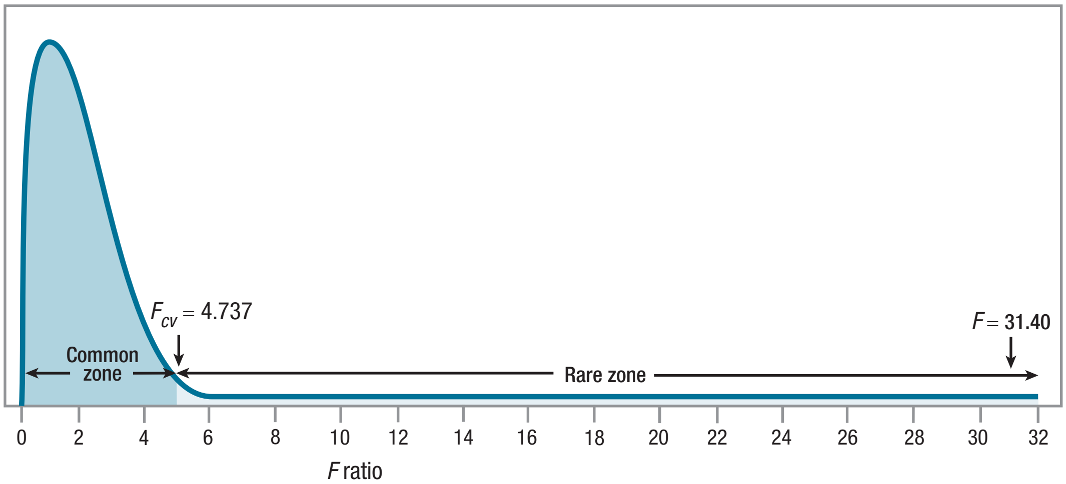

Figure 10.12 Determining Whether the Observed Value of the Test Statistic Falls in the Rare Zone or the Common Zone for the Maze Study The observed value of the test statistic, $F = 31.40$, falls in the rare zone, so the null hypothesis is rejected.

Dr. Chung used APA format in reporting the results of the analysis of variance. APA format provides five pieces of information:

1. F indicates that an F test, an analysis of variance, was used.
2. The numbers in parentheses, 2 and 7, are the degrees of freedom for SS_{Between}, the numerator of the F ratio, and for SS_{Within}, the denominator of the F ratio. By adding these two numbers together and adding 1, it is possible to find the sample size:

$$N = 2 + 7 + 1 = 10$$

3. 31.40 is the observed (calculated) value of the test statistic.
4. .05 reveals that alpha was set at .05 and there is a 5% chance of making a Type I error.
5. $p < .05$ indicates that the null hypothesis was rejected. It means that an F value of 31.40 is a rare occurrence (it happens less than 5% of the time) when the null hypothesis is true.

How Big Is the Effect?

The second question to be addressed in interpreting an ANOVA is how big the *overall* effect is. That is, how much impact does the independent variable have on the dependent variable? For the maze data, this involves asking how much impact the three different types of food (low-, medium-, and high-calorie) have on the time it takes to run the maze.

Once again, r^2 will be used. Equation 10.8 shows how to calculate r^2, the percentage of variability in the outcome variable that is accounted for by group status.

Equation 10.8 Formula for r^2, the Percentage of Variability in the Dependent Variable Accounted for by the Explanatory Variable

$$r^2 = \frac{SS_{\text{Between}}}{SS_{\text{Total}}} \times 100$$

where r^2 = the percentage of variability in the dependent variable that is accounted for by the explanatory variable

SS_{Between} = between-groups sum of squares (Equation 10.3)

SS_{Total} = total sum of squares (Equation 10.2)

This formula says that r^2 is calculated as the between-groups sum of squares divided by the total sum of squares. Then, to turn it into a percentage, the ratio is multiplied by 100. These calculations reveal the percentage of total variability in the scores that is accounted for by the treatment effect. For the maze data, these calculations would lead to the conclusion that $r^2 = 90\%$:

$$\begin{aligned} r^2 &= \frac{SS_{\text{Between}}}{SS_{\text{Total}}} \times 100 \\ &= \frac{54.00}{60.00} \times 100 \\ &= .9000 \times 100 \\ &= 90.00\% \end{aligned}$$

By Cohen's standards, an r^2 of 90% is a huge effect. Figure 10.13 is a visual demonstration of how big an effect this is. Here is what Dr. Chung's interpretation would look like for the maze data now that r^2 is known:

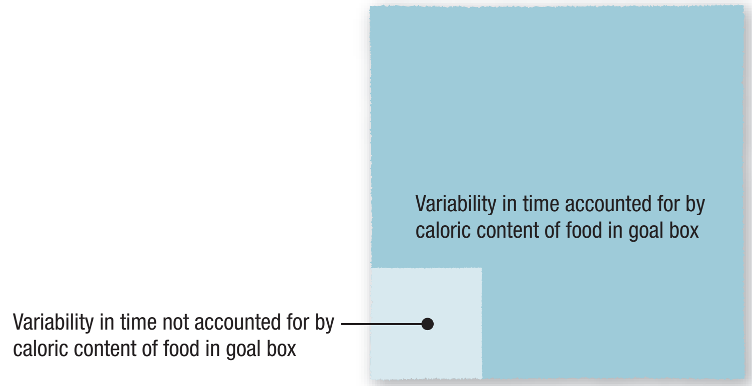

Figure 10.13 Percentage of Variability in Time to Run the Maze Scores That Is Accounted for by Type of Food in Goal Box 90% of the variability in time it takes for rats to run the maze is accounted for by the type of food they expect to find in the goal box. Only 10% of the variability remains, to be accounted for by other factors such as individual differences.

> The results of the one-way ANOVA were statistically significant [$F(2, 7) = 31.40$, $p < .05$], indicating that the time it takes a hungry rat to run a maze is affected by the type of food it expects to find in the goal box. When all the rats are of the same age and from the same strain, as they were here, then the calorie content of the food in the goal box has a very large effect on the speed with which the animal travels the maze. In fact, the type of food in the goal box explains 90% of the variability in running speed.

Here's a heads up for future chapters and for reading results sections in psychology articles—there's another measure of effect size for a between-subjects, one-way ANOVA that is calculated exactly the same way as r^2, but is called something different. It is called "eta squared" and abbreviated η^2. It is calculated the same way as r^2 and provides the same information, how much of the variability in the dependent variable is explained by the independent variable.

A Common Question

Q How is η pronounced?

A η is the lowercase version of the Greek letter eta and there is no consensus on how it should be pronounced. I have heard "eat-uh," "etta," and "ey-tuh" as in "hey." Whichever one you choose, just say it with confidence and you'll be fine.

Where Is the Effect?

So far, Dr. Chung knows that the goal affects running time and how much variability is explained, but he doesn't know where the effect occurs. Is the 3-second difference between the mean of low-calorie time and the mean of medium-calorie time a statistically significant one? Or, does it take the 6-second difference between low-calorie and high-calorie goals to be statistically significant?

Post-hoc tests are meant to be used only after *a statistically significant* F *ratio has been found.*

The *F* ratio, which was statistically significant, reveals that at least one pair of means has a statistically significant difference, but it doesn't tell which one or which ones. For that, a *post-hoc* test is needed. "Post-hoc" is Latin for "after this" and post-hoc tests are meant to be used only *after* a statistically significant *F* ratio has been found. Post-hoc tests are mathematically designed to allow multiple comparisons to be made while keeping alpha at the desired level.

There are a wide variety of post-hoc tests with cool names like the Scheffé, the Newman-Keuls, and the Bonferroni-Dunn. Post-hoc tests vary in terms of what type of error they are more likely to make, a Type I error or a Type II error. No post-hoc test can guarantee avoiding a mistake. Compared to each other, some post-hoc tests are more likely to say that there is a statistically significant difference between means that don't differ (Type I error). Other post-hoc tests are more likely to find no statistically significant difference when a difference exists (Type II error).

The test taught here, the Tukey *HSD*, is the latter type, what is called a conservative test. If it finds a statistically significant difference between a pair of means, then the population means probably really are different. Which is why *HSD* stands for "honestly significant difference." However, there is a cost to this conservatism: the *HSD* has less statistical power and so may overlook a pair of means that are different.

The Tukey *HSD* works by calculating an *HSD* value, which is the minimum difference needed between two means in order for the difference to be considered statistically significant. If the observed difference between a pair of means is greater than or equal to the *HSD* value, then the researcher can conclude that there is a statistically significant difference between the two groups. For example, the mean for the low-calorie group was 31 seconds and for the medium-calorie group it was 28 seconds. There is a 3.00-second difference between the mean of these two groups. If the *HSD* value were, say, 2.50, then the difference between the mean of these two groups would be a statistically significant one. This is the case because the observed difference, 3.00, met or exceeded the *HSD* value of 2.50.

The formula to calculate *HSD* is shown in Equation 10.9. In order to apply that formula, one needs a value called *q*. Values of *q* can be found in Appendix Table 5, a part of which is shown in Table 10.12.

TABLE 10.12 Part of Appendix Table 5, Table of Values of *q* for Use in the Tukey *HSD* Post-Hoc Test

	k		
df	**2**	**3**	**4**
2	6.09	8.33	9.80
3	4.50	5.91	6.83
4	3.93	5.04	5.76
5	3.64	4.60	5.22
6	3.46	4.34	4.90
7	3.34	4.17	4.68
8	3.26	4.04	4.53
9	3.20	3.95	4.42
10	3.15	3.88	4.33

A *q* value is needed to calculate a Tukey *HSD* post-hoc test. The *q* value for a between-subjects, one-way ANOVA is found at the intersection of the column for *k*, the number of groups in the ANOVA, and the row where $df = df_{\text{Within}}$.

There are q tables for $\alpha = .05$ and $\alpha = .01$, for a 5% and a 1% chance of making a Type I error. The q values are found at the intersections of rows and columns.

- The columns represent different values of k, the number of groups being compared in the ANOVA.
- The columns represent within-groups degrees of freedom, df_{Within}.

For the maze data, there are three groups and within-groups degrees of freedom are 7. The q value of 4.17 is found at the intersection of the column where $k = 3$ with the row where $df = 7$.

Equation 10.9 Formula to Calculate Tukey *HSD* Value for a Post-Hoc Test for Between Subjects, One-Way ANOVA

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

where HSD = value by which, if two means differ, the difference is statistically significant

q = value of q, from Appendix Table 5

$MS_{\text{Within}} = MS_{\text{Within}}$ (from ANOVA summary table)

n = sample size for the smallest group

To calculate HSD, one needs to know q, MS_{Within}, and n:

- From Appendix Table 5, it is known that $q = 4.17$.
- Referring back to the ANOVA summary table for the maze data (Table 10.7), $MS_{\text{Within}} = 0.86$.
- The only missing piece is n, the sample size for the smallest group. In Table 10.3, where the maze data were first presented, it can be seen that the smallest group has three cases, so $n = 3$. (Tukey's HSD was designed to be used when all sample sizes are equal. When sample sizes are unequal, using the smallest sample size makes the HSD value larger, which keeps the test conservative.)

With all the parts in place, the HSD value is calculated as

$$\begin{aligned} HSD &= q\sqrt{\frac{MS_{\text{Within}}}{n}} \\ &= 4.17\sqrt{\frac{0.86}{3}} \\ &= 4.17\sqrt{0.2867} \\ &= 4.17 \times 0.5354 \\ &= 2.2326 \\ &= 2.23 \end{aligned}$$

The *HSD* value is 2.23. Any two means that differ by at least 2.23 seconds are honestly significantly different. Here are the three sample means:

- Group 1: Expecting low-calorie food in the goal box = 31.00 seconds
- Group 2: Expecting medium-calorie food in the goal box = 28.00 seconds
- Group 3: Expecting high-calorie food in the goal box = 25.00 seconds

Here are the comparisons:

- Group 1 vs. Group 2: There is a 3.00-second difference between the mean of Group 1 and the mean of Group 2. $3.00 \geq 2.23$, so this is a statistically significant difference.
- Group 1 vs. Group 3: There is a 6.00-second difference between the mean of Group 1 and the mean of Group 3. $6.00 \geq 2.23$, so this is a statistically significant difference.
- Group 2 vs. Group 3: There is a 3.00-second difference between the mean of Group 2 and the mean of Group 3. $3.00 \geq 2.23$, so this is a statistically significant difference.

Once a researcher knows whether a difference is statistically significant, he or she needs to think about the direction of the difference. As with *t* tests, the actual sample means can be used to determine the direction of the difference. Here are Dr. Chung's conclusions:

- Group 1 vs. Group 2: Expecting medium-calorie food compared to low-calorie food leads to rats taking significantly less time to travel through the maze. That is, the medium-calorie rats ran through the maze statistically more quickly on average.
- Group 1 vs. Group 3: There's a significant difference between expecting low-calorie food and expecting medium-calorie food, so it's not surprising that a similar difference exists between expecting low-calorie food and expecting high-calorie food. Rats expecting high-calorie food were statistically faster on average than rats expecting low-calorie food.
- Group 2 vs. Group 3: Rats expecting high-calorie food were statistically faster on average than rats expecting medium-calorie food.

A statistically significant ANOVA reveals that at least one pair of population means differs. Now after the post-hoc test, Dr. Chung knows which pairs differ and what the differences mean. His challenge is to summarize the findings. He could report that the mean of Group 1 differs from the mean of Group 2 and that the mean of Group 1 differs from the mean of Group 3, and so forth, but that would be a focus on the trees, not the forest. Interpretations are more useful if they find a pattern in the results. Dr. Chung's findings could be summarized by saying that rats, this strain of rats at least, appear able to figure out that different foods have different caloric contents and, when hungry, are more motivated to seek nutritionally richer foods.

With all three interpretation questions addressed—Was the null hypothesis rejected? How big is the effect? Where is the effect?—Dr. Chung is ready to write a four-point interpretation that (1) tells what the study was about, (2) presents the main results, (3) explains what the results mean, and (4) makes suggestions for future research:

> In this study, rats were exposed to three different types of food (low-calorie, medium-calorie, and high-calorie) in the goal box while learning a maze. After learning the maze, they were made hungry, randomly assigned to three groups,

and timed while they ran the maze. The three groups expected to find different foods in the goal box: low-calorie, medium-calorie, or high-calorie. There was a statistically significant decrease in mean time, that is, an increase in mean speed, as the caloric content of the food increased [$F(2, 7) = 31.40$, $p < .05$]. Time decreased from a mean of 31.00 seconds for the rats expecting low-calorie food to 28.00 seconds for those expecting medium-calorie food and to 25.00 seconds for the ones expecting high-calorie food. The effect of caloric content on speed in this study was quite dramatic, accounting for 90% of the variability in speed. This is probably an overestimate of the effect of caloric content as all rats were from the same strain and were the same age.

These results suggest that this strain of rats is able to discriminate the caloric content of food and that, when they are hungry, they are more motivated to seek nutritionally richer food. This motivation makes sense for increasing the likelihood of survival in lean times. An alternative explanation for the results is that the richer-calorie foods tasted better and that rats hurry to food if it tastes good. Future research should be sure to control for food taste. Future research should also employ different strains of rats.

Worked Example 10.2 For practice in interpretation, let's return to Dr. Douglas's data from the study of over-the-counter pain relievers (ibuprofen and acetaminophen) on pain threshold. In that study, 12 male undergraduates were randomly assigned by a sensory psychologist to take a placebo, ibuprofen, or acetaminophen. They then placed their hands in ice water until they felt pain. The participants who received placebos kept their hands in ice water for an average of 17.50 seconds, the ibuprofen participants for 23.00 seconds, and the acetaminophen participants for 21.00 seconds. F_{cv} was 4.256 and the ANOVA summary table is reprinted here (see Table 10.13).

TABLE 10.13 ANOVA Summary Table for One-Way ANOVA for Pain Threshold Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	62.00	2	31.00	1.35
Within groups	207.00	9	23.00	
Total	269.00	11		

This is the same ANOVA summary table that was printed earlier as Table 10.11.

Was the null hypothesis rejected? From the ANOVA summary table (Table 10.13), it is clear that the observed value of F is 1.35. 1.35 is less than 4.256, F_{cv}, so the null hypothesis was not rejected. Figure 10.14 shows how the observed value of F falls in the common zone. With these data, Dr. Douglas does not have enough evidence to state that any one of these three groups differs from another in terms of the mean pain threshold. In APA format, she would write the results as

$$F(2, 9) = 1.35, p > .05$$

- F tells the reader that an analysis of variance, also known as an F test, was done.

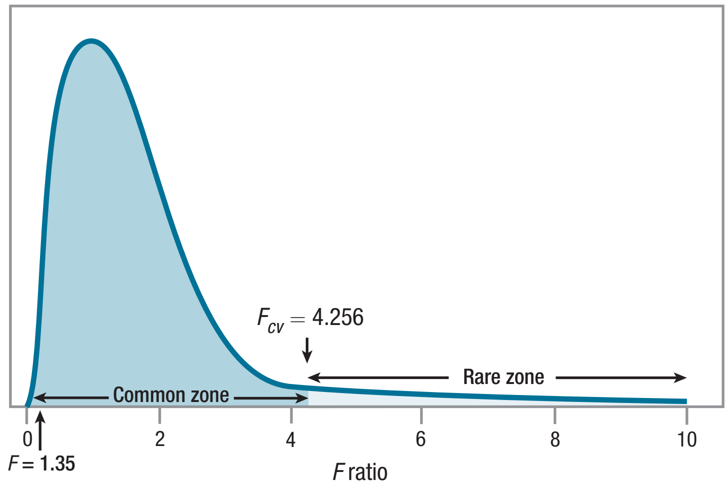

Figure 10.14 Determining Whether the Observed Value of the Test Statistic Falls in the Rare Zone or the Common Zone for the Pain Threshold Study The observed value of the test statistic, $F = 1.35$, falls in the common zone, so we fail to reject the null hypothesis.

- The 2 and 9 in the parentheses are the numerator and denominator degrees of freedom. These are df_{Between} and df_{Within}. If they are added together and 1 is added, that gives N.
- 1.35 is the value of F that Dr. Douglas calculated.
- The .05 is the alpha level that had been selected.
- The $p > .05$ means that a result like $F = 1.35$ is a common one when the null hypothesis is true, as under those conditions it occurs more than 5% of the time. The sensory psychologist has failed to reject the null hypothesis.

How big is the effect? Equation 10.5 is used to calculate the effect size, r^2, the percentage of variability in the dependent variable (seconds to pain threshold) that is accounted for by the independent variable (type of painkiller). To use the equation, two values, SS_{Between} (62.00) and SS_{Total} (269.00), are obtained from the ANOVA summary table:

$$\begin{aligned} r^2 &= \frac{SS_{\text{Between}}}{SS_{\text{Total}}} \times 100 \\ &= \frac{62.00}{269.00} \times 100 \\ &= .2305 \times 100 \\ &= 23.05\% \end{aligned}$$

An r^2 of 23% is close to a large effect, according to Cohen. This gives Dr. Douglas two conflicting pieces of information:

- Not enough evidence exists to say there is an effect.
- However, the effect seems to be large.

One way to reconcile these two conflicting pieces of information is to look at the possibility of a Type II error. Type II errors occur when a researcher fails to find statistically significant evidence of an effect that really does exist. That possibility can't be ruled out here. When interpreting such a situation, the researcher should point out the possibility and suggest that the study be replicated with a larger sample size. Larger sample sizes give tests more power and so increase the likelihood of being able to find an effect when it does exist.

Where is the effect? Post-hoc tests are designed to be conducted only when the null hypothesis is rejected. In this study, the null hypothesis was not rejected. This means there is not enough evidence that an effect exists, so there is no reason to try to find where the effect is.

Putting it all together. Here's Dr. Douglas's four-point interpretation for the pain threshold study. She reveals what the study was about, identifies its main results, interprets them, and makes suggestions for future research.

> Analysis of data for a study comparing the impact of placebo, ibuprofen, and acetaminophen on pain threshold found no evidence that either of the over-the-counter pain relievers had any impact on the threshold for perceiving pain. There was no statistically significant difference in the mean time it took participants to perceive pain when their hands were in ice water [$F(2, 9) = 1.35$, $p > .05$]. There was some evidence that medication may have an impact, but the small sample size ($N = 12$) meant the study did not have enough power to detect it. In order to determine if these over-the-counter pain relievers have an impact on pain threshold, the study should be replicated with a larger sample size.

Practice Problems 10.3

Apply Your Knowledge

10.12 Given $\alpha = .05$, $F_{cv} = 2.690$, $df_{\text{Between}} = 4$, $df_{\text{Within}} = 30$, and $F = 7.37$, write the results in APA format.

10.13 Given $SS_{\text{Between}} = 1{,}827.50$ and $SS_{\text{Total}} = 4{,}631.50$, (a) calculate r^2 and (b) comment on the size of the effect.

10.14 Four groups are being compared in a between-subjects, one-way ANOVA. Each group has 10 cases. $MS_{\text{Within}} = 77.89$ and $df_{\text{Within}} = 36$. The F ratio was statistically significant with $\alpha = .05$.

a. Use this information to complete a post-hoc test comparing $M_1 = 29.00$ and $M_2 = 18.00$; interpret the difference between the two population means.

b. Use this information to complete a post-hoc test comparing $M_1 = 29.00$ and $M_3 = 22.00$; interpret the difference between the two population means.

10.15 An industrial/organizational psychologist wanted to investigate the effect of different management styles on employee performance. She obtained 27 employees at a large company and randomly assigned them, in equal-sized groups, to be supervised by managers who believed in motivating employees by using (a) rewards for good performance, (b) punishment for poor performance, or (c) a mixture of rewards and punishments. No nonrobust assumptions of a between-subjects, one-way ANOVA were violated. One year later, she recorded how much of a raise (as a percentage) each employee received. She believed that employees who did better work would receive bigger raises. The employees supervised by managers who believed in using rewards saw a mean raise of 8%, those supervised by managers who believed in using punishment had a mean raise of 5%, and those supervised by managers who used a mixture of rewards and punishment had a mean raise of 7%.

A between-subjects, one-way ANOVA was used to analyze the results. The results were statistically significant. Given the ANOVA summary table, $r^2 = 30.43\%$ and $HSD = 2.35$, write a four-point interpretation.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	42.00	2	21.00	5.25
Within groups	96.00	24	4.00	
Total	138.00	26		

Application Demonstration

America is a land of movers. If we don't like where we're living, if we think the grass is greener somewhere else, we just pull up our stakes and go. Are salaries for psychologists consistent throughout the United States or do some areas have higher or lower salaries? If so, this information could guide a recent graduate's decision about where to seek employment.

The Census Bureau divides the United States into four regions: the Northeast, Midwest, South, and West. Four states from each region were randomly selected and the mean salary for "clinical, counseling, and school psychologists" was obtained from a Bureau of Labor Statistics website for each state. Salaries were recorded in thousands of dollars; \$79.52, for example, means \$79,520. Recording the values this way keeps the sums of squares from becoming unwieldy numbers in the hundreds of millions. The mean salaries for the four regions (Northeast = \$77.73, Midwest = \$63.55, West = \$68.83, and South = \$61.37) are seen in Figure 10.15.

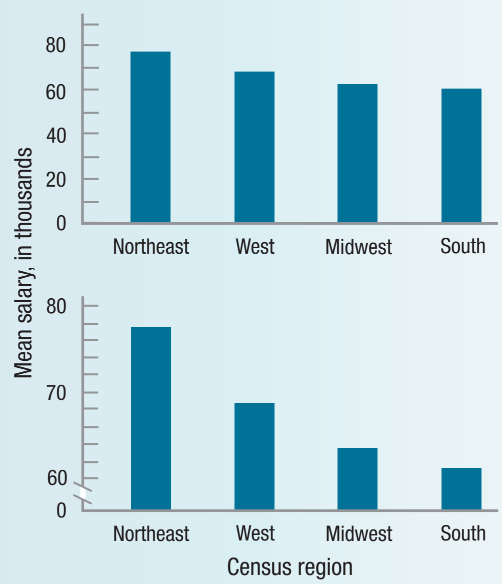

Figure 10.15 Mean Salary for Psychologists by Census Region: Two Different Pictures The mean salary for clinical, counseling, and school psychologists is highest in the Northeast and lowest in the South. Whether the differences among regions look significant or not depends on how the graph is drawn. The two graphs here differ only in the scale on the *Y*-axis.

The top panel in Figure 10.15 shows some slight differences in salary by region. Based on this graph, it seems unlikely that there are any statistically significant differences. The bottom panel in Figure 10.15, where the *y*-axis doesn't start at zero, does make it look like there are significant differences by region. As mentioned in Chapter 2, a graph is a subjective interpretation of the data.

The question of whether there are differences in mean salary is best resolved in an objective manner with a significance test.

Step 1 Pick a Test. There are four samples, they are independent, and the dependent variable is measured at the ratio level, so the test should be a between-subjects, one-way ANOVA.

Step 2 Check the Assumptions.

- The random samples assumption is not violated. Cases were randomly selected from their populations.
- The independence of observations assumption is not violated. Each state (case) is measured only once.
- The normality assumption might be violated. The populations are not large. For example, there are only nine states in the Northeast. With such a small population, it is unlikely to be normally distributed. But, this assumption is robust.
- The homogeneity of variance assumption is not violated. All standard deviations are comparable.

Step 3 List the Hypotheses.

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

H_1: At least one population mean differs from the others.

Step 4 Set the Decision Rule. First, find the numerator, denominator, and total degrees of freedom:

$$\text{Numerator degrees of freedom} = df_{\text{Between}} = k - 1 = 4 - 1 = 3$$

$$\text{Denominator degrees of freedom} = df_{\text{Within}} = N - k = 16 - 4 = 12$$

$$\text{Total degrees of freedom} = df_{\text{Total}} = N - 1 = 16 - 1 = 15$$

Next, set the alpha level:

$\alpha = .05$ because willing to have a 5% chance of making a Type I error

Then, use the degrees of freedom and alpha level to find the critical value of F in Appendix Table 4:

$$F_{cv} = 3.490$$

Then, set the decision rule:

- If $F \geq 3.490$, reject H_0.
- If $F < 3.490$, fail to reject H_0.

Step 5 Calculate the Test Statistic. Calculate the between-groups sum of squares using Equation 10.3, the within-groups sum of squares using Equation 10.4, and the total sum of squares using Equation 10.2. To save time, these values have already been calculated:

$$SS_{\text{Between}} = 636.41$$
$$SS_{\text{Within}} = 723.85$$
$$SS_{\text{Total}} = 1{,}360.26$$

Use the degrees of freedom calculated in Step 4 and the sums of squares calculated in Step 4 to fill in the appropriate columns of the ANOVA summary table. Then follow the guidelines in Table 10.7 to complete the ANOVA summary table (Table 10.14).

TABLE 10.14 ANOVA Summary Table for Salaries of Psychologists by Census Regions

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	636.41	3	212.14	3.52
Within groups	723.85	12	60.32	
Total	1,360.26	15		

Step 6 Interpret the Results. *Was the null hypothesis rejected?* To determine if the null hypothesis should be rejected, apply the decision rule. Which of these statements is true?

- Is $3.52 \geq 3.490$?
- Is $3.52 < 3.490$?

The first statement is true, so the null hypothesis should be rejected. The results are called statistically significant, the alternative hypothesis is accepted, and it is concluded that at least one population mean differs from at least one other population mean. A post-hoc test will be needed to determine where the differences lie. For now, though, the results can be written in APA format:

$$F(3, 12) = 3.52, p < .05$$

How big is the effect? This question is answered by calculating r^2:

$$r^2 = \frac{SS_{\text{Between}}}{SS_{\text{Total}}} \times 100 = \frac{636.41}{1,360.26} \times 100 = .4679 \times 100 = 46.79\%$$

On the basis of Cohen's criteria, the fact that region explains 46.79% of the variability in salary means this is a very large effect.

Where is the effect? Tukey's *HSD* will be used as a post-hoc test to find out which pair, or pairs, of means are statistically significantly different. To calculate the *HSD* value, a researcher must know q, MS_{Within}, and n. q is found in Appendix Table 5 for $\alpha = .05$, in the column for $k = 4$ and the row for $df_{\text{Within}} = 12$. With $q = 4.20$, $MS_{\text{Within}} = 60.32$, and $n = 4$, Equation 10.9 is used to calculate *HSD:*

$$HSD = q\sqrt{\frac{MS_{\text{Within}}}{n}} = 4.20\sqrt{\frac{60.32}{4}} = 4.20\sqrt{15.0800} = 4.20 \times 3.8833$$
$$= 16.3099 = 16.31$$

Any groups that have a mean difference greater than or equal to 16.31 points are statistically significantly different. The largest mean is the Northeast at \$77.73 thousand and the smallest is the South at \$61.37 thousand. Those two are 16.36 points apart and this difference is a statistically significant difference. All other differences are smaller than 16.31 and thus not statistically significant.

The effect seems to be due to the $16.36 thousand difference between the mean salaries for psychologists in the Northeast and mean salaries in the South. Salaries are statistically significantly higher in the Northeast than they are in the South.

Here is a four-point interpretation:

In this study, a random sample of states from each of the four census regions was taken and the average salary in each state for clinical, counseling, and school psychologists was recorded. The mean salary in the Northeast region was $77,730, $63,550 in the Midwest, $61,370 in the South, and $68,830 in the West. The overall ANOVA showed a statistically significant effect for the region [$F(3, 12) = 3.52$, $p < .05$] and region accounted for almost 50% of the variability in salary. However, post-hoc testing showed that the only statistically significant difference that existed was between the Northeast and the South—the mean salaries for clinical, counseling, and school psychologists are statistically higher in the Northeast than in the South. The results of this study suggest that it might be worthwhile for a psychologist living in the South to move to the North, but there would be no advantage to moving to the West or Midwest. If this study is replicated, it would be advisable to take into account the cost of living as higher salaries in a region may be mitigated by a higher cost of living.

SUMMARY

Explain when ANOVA is used and how it works.

- Analysis of variance (ANOVA) is a family of tests for comparing the means of three or more groups. Between-subjects, one-way ANOVA is used to compare the means of two or more independent samples when there is just one explanatory variable.
- Between-subjects, one-way ANOVA separates variability in scores into that due to the different ways the groups are treated and that due to variability between cases. If there is more variability between groups due to treatment than within groups due to individual differences, then there is a statistically significant difference among the groups.

Complete a between-subjects, one-way ANOVA.

- To complete an ANOVA, the assumptions must be met and the hypotheses set. The null hypothesis says that all population means are the same and the alternative that at least one population mean differs from the others. The decision rule compares the observed F ratio (variability due to treatment divided by variability due to individual differences) to the critical value of F. Calculating an F ratio involves finding sums of squares and degrees of freedom for three sources of variability (between group, within group, and total) and organizing that information in an ANOVA summary table.

Interpret the results of a between-subjects, one-way ANOVA.

- If the results are statistically significant, at least one population mean differs from at least one other. Then calculate the size of the effect using r^2 and use post-hoc testing with the Tukey *HSD* to which pair(s) of means differ.
- If results are not statistically significant, then there is not enough evidence to conclude any population means are different. Nonetheless, calculate r^2 in order to consider the possibility of Type II error.

DIY

ANOVA is an extension of the *t* test, so let's extend the DIY from the independent-samples *t* test to the between-subjects, one-way ANOVA. In that DIY, you were asked to compare two groups of states, say, northern states vs. southern states, on some outcome variable, say, murder rate. Now, divide states into three groups, say, northern, middle, and southern states, and compare them on your outcome variable.

KEY TERMS

analysis of variance (ANOVA) – a family of statistical tests for comparing the means of two or more groups.

between-group variability – variability in scores that is primarily due to the different treatments that different groups receive.

between subjects – ANOVA terminology for independent samples.

between-subjects, one-way ANOVA – a statistical test used to compare the means of two or more independent samples when there is just one explanatory variable.

factor – term for an explanatory variable in ANOVA.

level – ANOVA terminology for a category of an explanatory variable.

post-hoc test – a follow-up test to a statistically significant ANOVA, engineered to find out which pairs of means differ while keeping the overall alpha level at the chosen level.

sum of squares between (**$SS_{Between}$**) – A sum of the squared deviations scores representing the variability between groups.

sum of squares total (**SS_{Total}**) – A sum of the squared deviation scores representing the all variability in the scores.

sum of squares within (**SS_{Within}**) – A sum of the squared deviation scores representing the variability within groups.

treatment effect – the impact of the explanatory variable on the dependent variable.

way – term for an explanatory variable in ANOVA.

within-group variability – variability within a sample of cases, all of which have received the same treatment.

CHAPTER EXERCISES

Review Your Knowledge

10.01 A ____-sample test compares the mean of one group to the mean of another group.

10.02 Analysis of variance is used when comparing the means of ____ or more groups.

10.03 ____ is short for analysis of variance.

10.04 ____, one-way ANOVA compares the means of two or more independent samples.

10.05 Other terms used for explanatory variables in ANOVA are ____ and ____.

10.06 The categories of an explanatory variable in ANOVA are called ____.

10.07 ANOVA keeps the risk of ____ error at a reasonable level.

10.08 The risk of making a Type I error ____ as the number of statistical tests being completed increases.

10.09 A follow-up test in ANOVA is called a ____ test.

10.10 A ____ is used to find out which pairs of means in an ANOVA are statistically significantly different.

10.11 ANOVA works by separating out the different sources of ____ in the scores.

10.12 Variability within a set of scores is called ____ variability.

10.13 Within-group variability is caused by ____.

10.14 Making a sample more homogeneous ____ within-group variability.

10.15 ____ variability is due to the different treatments different groups receive.

10.16 ____ effect is the term used to label the effect of the explanatory variable.

10.17 Between-group variability is caused by ____ as well as treatment effect.

10.18 If treatment has ____ impact on outcome, the sample means are close together.

10.19 When sample means are far apart, the treatment effect is ____.

10.20 When treatment has a large impact, there is a lot of variability between group ____.

10.21 The ratio of between-group variability to within-group variability is called an ____.

10.22 Between-group variability, the numerator in an *F* ratio, is made up of ____ and ____.

10.23 If treatment has no impact on outcome, the *F* ratio should be near ____.

10.24 As treatment has an impact on outcome, the value of the *F* ratio climbs above ____.

10.25 If there are just two groups, $F =$ ____ squared.

10.26 Between-subjects, one-way ANOVA is used when there is one ____ and ____ groups.

10.27 The assumptions for between-subjects, one-way ANOVA are the same as they are for ____.

10.28 For a between-subjects, one-way ANOVA, the cases in each group should be ____ from the population.

10.29 If the cases in the groups for a between-subjects, one-way ANOVA are paired together, a between-subjects, one-way ANOVA *can* / *cannot* be used.

10.30 Between-subjects, one-way ANOVA assumes that the dependent variable in each population is ____ distributed.

10.31 The assumption that the variability in all groups is about the same is called the ____ assumption.

10.32 The null hypothesis states that there is ____ mean difference between any of the populations.

10.33 The null hypothesis for a between-subjects, one-way ANOVA is always ____ directional.

10.34 The alternative hypothesis says that at least ____ population mean is different from at least ____ other population mean.

10.35 The decision rule for a between-subjects, one-way ANOVA says that if *F* falls in the rare zone of the sampling distribution, the null hypothesis is ____.

10.36 The critical value of *F* depends on the degrees of freedom for the ____ and the degrees of freedom for the ____ of the *F* value.

10.37 The numerator degrees of freedom for a between-subjects, one-way ANOVA *F* ratio is df____.

10.38 The denominator degrees of freedom for a between-subjects, one-way ANOVA *F* ratio is df____.

10.39 In order to calculate total degrees of freedom for a between-subjects, one-way ANOVA, one could add together ____ and ____.

10.40 If $F = F_{cv}$, the null hypothesis is ____.

10.41 ΣX^2 is called a ____.

10.42 Numerators in variance formulas are ____.

10.43 To calculate ____, the grand mean is subtracted from each score, the difference scores are squared, and they are all added up.

10.44 SS_{Total} can be broken down into two components: ____ and ____.

10.45 SS_{Between} isolates the variability in scores that is primarily due to ____.

10.46 $SS_{\text{Total}} - SS_{\text{Between}} =$ ____

10.47 The table used to organize ANOVA results is called an ____.

10.48 The sources of variability in an ANOVA are listed in the ____ column of an ANOVA summary table.

10.49 A ____ is calculated by dividing a sum of squares by its degrees of freedom.

10.50 An *F* ratio in between-subjects, one-way ANOVA is calculated as the ratio of *MS*____ to *MS*____.

10.51 Interpretation of results for a statistically significant ANOVA differs from *t* in that it has to address where the ____ is located.

10.52 One determines if an *F* ratio is statistically significant by comparing *F* to ____.

10.53 If the null hypothesis is rejected, then ____ is accepted.

10.54 The numbers in parentheses when reporting the results of a between-subjects, one-way ANOVA in APA format are ____ and ____.

10.55 If the null hypothesis is not rejected, ANOVA results are reported in APA format as *p* ____ .05.

10.56 ____ is used as the effect size in between-subjects, one-way ANOVA.

10.57 r^2 is calculated as the percentage of total variability in scores that is accounted for by ____ variability.

10.58 The higher the percentage of variability in the dependent variable explained by the explanatory variable, the ____ is the size of the effect.

10.59 The closer r^2 is to ____%, the stronger the effect; the closer r^2 is to ____%, the weaker the effect.

10.60 Cohen considers an r^2 of 9% to be a ____ effect.

10.61 Post-hoc tests are only used if the ANOVA results are ____.

10.62 Some post-hoc tests have a greater likelihood of making a ____ error and others a greater chance of a ____ error.

10.63 The Tukey *HSD* is a conservative test. It is more likely to make a ____ error than a ____ error.

10.64 *HSD* stands for ____.

10.65 If the difference between two means meets or exceeds the *HSD* value, the difference is ____.

10.66 To find a *q* value, one needs to know df_{Within} and ____.

10.67 The *n* in the *HSD* equation represents the sample size for the ____ group.

10.68 The *HSD* test is used to determine which pairs of means have a ____ difference.

Apply Your Knowledge

For the first set of questions, select the appropriate statistical test from (a) single-sample z test; (b) single-sample t test; (c) independent-samples t test; (d) paired-samples t test; (e) between-subjects, one-way ANOVA; or (f) none of the above.

10.69 An adhesives researcher tests the holding power of a wood glue under different humidities. He glues together 100 small pieces of wood and randomly assigns each one to sit for an hour in a room with either 10%, 30%, 50%, 70%, or 90% relative humidity. Each glued piece of wood is then tested to see how many pounds of weight it takes to break the glue bond. The means of the five groups are compared.

10.70 Some parents demand that their children friend them on Facebook and other parents don't. A developmental psychologist wondered if such parental supervision had any impact on behavior. She obtained a sample of first-year college students and classified them as having parents who were or were not Facebook friends. From each student she also learned the number of days, during the first month of college, that he or she had consumed any alcohol. The two means are compared.

10.71 A nutritionist investigates the impact of type of breakfast on mid-morning concentration. He puts together a sample of 50 adults. He feeds them a breakfast and 4 hours later measures their concentration on an interval-level concentration test. On one day he feeds them bacon and eggs, on a second day he feeds them oatmeal, and on a third day he has them skip breakfast. He then calculates the mean concentration score for each condition.

10.72 A social psychologist is interested in how people perceive age-discrepant couples. She gets a random sample of shoppers at a mall, has them read a wedding announcement, and then asks them to predict—on an interval-level scale—the newlyweds' degree of marital happiness after 10 years of marriage. Participants are randomly assigned to groups: one-third read an announcement in which the bride is 25 and the groom 35, one-third read an announcement in which the bride is 35 and the groom 25, and the final third read an announcement in which both bride and groom are 30.

For the next set of questions, determine if assumptions have been violated and if a between-subjects, one-way ANOVA can be completed.

10.73 A sensory psychologist wants to determine if a sensory threshold differs depending on which nostril is used. He obtains 30 introductory psychology student volunteers and measures the absolute threshold, in parts per million, for detecting the scent of peppermint. Participants are randomly assigned to three equal-size groups: (1) scent administered to the left nostril, (2) scent administered to the right nostril, and (3) scent administered to both nostrils. Each participant is tested individually by an experimenter who does not know what the hypothesis is and does not know which nostril is being used.

10.74 A biomedical engineer compared the abrasion resistance—as measured by number of rotations until failure—of three different artificial hips. Manufacturer X makes hip joints out of metal, manufacturer Y out of ceramic, and manufacturer Z out of a metal/ceramic composite. The biomedical engineer gets a random sample of 6 hips from each manufacturer's production line and tests each one individually until failure occurs.

Writing the hypotheses

10.75 Write the null and alternative hypotheses for the between-subjects, one-way ANOVA in Exercise 10.73.

10.76 Write the null and alternative hypotheses for the between-subjects, one-way ANOVA in Exercise 10.74.

Calculating degrees of freedom

10.77 If $N = 48$ and $k = 4$, calculate (a) df_{Total}, (b) $df_{Between}$, and (c) df_{Within}.

10.78 If $N = 120$ and $k = 5$, calculate (a) df_{Total}, (b $df_{Between}$, and (c) df_{Within}.

10.79 If there are four independent samples, each with 15 participants, what are (a) df_{Total}, (b) $df_{Between}$, and (c) df_{Within}?

10.80 If $n_1 = 15$, $n_2 = 12$, and $n_3 = 18$, what are (a) df_{Total}, (b) $df_{Between}$, and (c) df_{Within}?

Finding F_{cv}

10.81 If $df_{Within} = 44$ and $df_{Between} = 2$, what is F_{cv} if $\alpha = .01$?

10.82 If $df_{Between} = 3$ and $df_{Within} = 36$, what is F_{cv} if $\alpha = .05$?

10.83 If $\alpha = .05$, $N = 50$, and $k = 4$, what is F_{cv}?

10.84 If $\alpha = .05$, $N = 80$, and $k = 5$, what is F_{cv}?

Stating the decision rule

10.85 If $df_{Within} = 40$ and $df_{Between} = 2$, what is the decision rule if $\alpha = .05$? Draw a sampling distribution of F and mark the rare and common zones.

10.86 If $df_{Within} = 10$ and $df_{Between} = 3$, what is the decision rule if $\alpha = .05$? Draw a sampling distribution of F and mark the rare and common zones.

Calculating sums of squares

10.87 Prepare the data table for use by computational formulas for sums of squares.

Group 1	Group 2	Group 3
108	100	99
102	105	95
		91

10.88 Prepare the data table for use by computational formulas for sums of squares.

Group 1	Group 2	Group 3
46	54	74
48	58	80

10.89 Given the data in this table, calculate SS_{Total}, $SS_{Between}$, and SS_{Within}.

	Group 1		Group 2		Group 3			
	X	X^2	X	X^2	X	X^2		
	112	12,544	98	9,604	88	7,744		
	104	10,816	90	8,100	85	7,225		
			88	7,744	76	5,776	Grand	
					80		X	X^2
$\sum$	216.00	23,360.00	276.00	25,448.00	329.00	27,145.00	821.00	75,953.00
n	2		3		4		9	

10.90 Given the data in this table, calculate SS_{Total}, $SS_{Between}$, and SS_{Within}.

	Group 1		Group 2		Group 3			
	X	X^2	X	X^2	X	X^2		
	55	3,025	47	2,209	68	4,624		
	63	3,969	58	3,364	66	4,356		
	72	5,184	63	3,969	62	3,844	Grand	
	59	3,481	48	2,304	73	5,329	X	X^2
$\sum$	249	15,659	216	11,846	269	18,153	734	45,658
n	4		4		4		12	

10.91 If $SS_{Total} = 98.75$ and $SS_{Between} = 40.33$, what is SS_{Within}?

10.92 If $SS_{Within} = 168.43$ and $SS_{Between} = 764.13$, what is SS_{Total}?

Calculating mean squares

10.93 If $SS_{Between} = 2{,}378.99$ and $df_{Between} = 3$, what is $MS_{Between}$?

10.94 If $SS_{Between} = 138.76$ and $df_{Between} = 4$, what is $MS_{Between}$?

10.95 If $SS_{Within} = 78.95$ and $df_{Within} = 32$, what is MS_{Within}?

10.96 If $SS_{Within} = 452.86$ and $df_{Within} = 102$, what is MS_{Within}?

Calculating F

10.97 If $MS_{Between} = 38.88$ and $MS_{Within} = 17.44$, what is F?

10.98 If $MS_{Within} = 764.55$ and $MS_{Between} = 898.00$, what is F?

Completing an ANOVA summary table

10.99 Complete this ANOVA summary table:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Between groups	172.80	2		
Within groups	6,410.80	12		
Total	6,583.60	14		

10.100 Complete this ANOVA summary table:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Between groups	47,843.15	3		
Within groups	3,053.88	30		
Total	50,897.03	33		

Deciding if the null hypothesis is rejected

10.101 If $F_{cv} = 3.238$, draw a sampling distribution of F, label the rare and common zones, locate $F = 1.96$, and determine, for this F value, if the null hypothesis should be rejected.

10.102 If $F_{cv} = 2.337$, draw a sampling distribution of F, label the rare and common zones, locate $F = 5.66$, and determine, for this F value, if the null hypothesis should be rejected.

10.103 If $F_{cv} = 3.467$ and $F = 7.64$, is the null hypothesis rejected?

10.104 If $F_{cv} = 2.486$ and $F = 1.98$, is the null hypothesis rejected?

Using APA format

10.105 If $df_{\text{Between}} = 3$, $df_{\text{Within}} = 17$, $F = 5.34$, and $\alpha = .05$, write the results in APA format. (Use df_{Between} and df_{Within} to find F_{cv} in order to determine if the null hypothesis was rejected.)

10.106 If $df_{\text{Between}} = 6$, $df_{\text{Within}} = 30$, $F = 2.81$, and $\alpha = .05$, write the results in APA format. (Use df_{Between} and df_{Within} to find F_{cv} in order to determine if the null hypothesis was rejected.)

10.107 Given this ANOVA summary table, write the results in APA format using $\alpha = .05$. (Use df_{Between} and df_{Within} to find F_{cv} in order to determine if the null hypothesis was rejected.)

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	59.98	3	19.99	3.28
Within groups	268.60	44	6.10	
Total	328.58	47		

10.108 Given this ANOVA summary table, write the results in APA format using $\alpha = .05$. (Use df_{Between} and df_{Within} to find F_{cv} in order to determine if the null hypothesis was rejected.)

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	42.34	2	21.17	2.23
Within groups	1,896.22	200	9.48	
Total	1,938.56	202		

Interpreting APA format

10.109 The results of a between-subjects, one-way ANOVA, in APA format, are $F(3, 26) = 4.53$, $p < .05$. What interpretative statement can one make about the differences among the four population means?

10.110 The results of a between-subjects, one-way ANOVA, in APA format, are $F(3, 30) = 2.66$, $p > .05$. What interpretative statement can one make about the differences among the four population means?

Calculating r^2

10.111 If $SS_{\text{Between}} = 128.86$ and $SS_{\text{Total}} = 413.67$, what is r^2?

10.112 If $SS_{\text{Between}} = 17.48$ and $SS_{\text{Total}} = 342.88$, what is r^2?

10.113 Using the information in the ANOVA summary table in Exercise 10.107, calculate r^2.

10.114 Using the information in the ANOVA summary table in Exercise 10.108, calculate r^2.

Deciding whether to do a post-hoc test

10.115 Should a post-hoc test be completed for the results in the ANOVA summary table in Exercise 10.107?

10.116 Should a post-hoc test be completed for the results in the ANOVA summary table in Exercise 10.108?

Finding q

10.117 If $k = 4$ and $df_{\text{Within}} = 16$, what is q if $\alpha = .01$?

10.118 If $k = 3$ and $df_{\text{Within}} = 26$, what is q if $\alpha = .05$?

Calculating HSD

10.119 If $q = 3.55$, $MS_{\text{Within}} = 10.44$, and $n = 8$, what is HSD?

10.120 If $q = 4.05$, $MS_{\text{Within}} = 6.87$, and $n = 11$, what is HSD?

Interpreting HSD

10.121 If $M_1 = 13.09$, $M_2 = 8.89$, and $HSD = 4.37$, is the difference a statistically significant one? What conclusion would one draw about the direction of the difference between the two population means?

10.122 If $M_1 = 123.65$, $M_2 = 144.56$, and $HSD = 10.64$, is the difference a statistically significant one? What conclusion would one draw about the direction of the difference between the two population means?

10.123 If $M_1 = 67.86$, $M_2 = 53.56$, $M_3 = 61.55$, and $HSD = 8.30$, which pairs of means have statistically significant differences? What is the direction of the differences?

10.124 If $M_1 = 12.55$, $M_2 = 13.74$, $M_3 = 5.49$, and $HSD = 4.44$, which pairs of means have statistically significant differences? What is the direction of the differences?

Completing an interpretation (HSD values are given whether needed or not).

10.125 An addictions researcher was curious about which drug was hardest to quit: alcohol, cigarettes, or heroin. She obtained samples of alcoholics, smokers, and heroin addicts who were in treatment for the second time and found out how long they had remained abstinent, in months, after their first treatment. The mean time to relapse for the 8 alcoholics was 4.63 months; for the 11 smokers, it was 4.91 months; for the 6 heroin addicts, it was 5.17 months. Given the ANOVA summary table below, $r^2 = 0.66\%$, and $HSD = 3.83$, complete a four-point interpretation of the results using $\alpha = .05$.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	1.02	2	0.51	0.07
Within groups	153.62	22	6.98	
Total	154.64	24		

10.126 A sleep specialist investigated the impact of watching TV and using computers, before bedtime, on sleep onset. He obtained 30 college student volunteers and randomly assigned them to three equally sized groups: (1) work on a computer for 30 minutes before going to bed, (2) watch TV for 30 minutes before going to bed, and (3) don't work on a computer or watch TV before bedtime. He then measured time to sleep onset (in minutes), finding means for the three groups, respectively, of 19.60, 17.40, and 5.30. Given the ANOVA summary table, $r^2 = 65.63\%$, and $HSD = 5.32$, complete a four-point interpretation of the results using $\alpha = .05$.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	1,185.80	2	592.90	25.78
Within groups	620.90	27	23.00	
Total	1,806.70	29		

For Exercises 10.127–10.128, complete all six steps of hypothesis testing for these data sets. Remember to keep the six steps in the right order.

10.127 A consumer researcher gave consumers a sample shampoo. After using the shampoo, each consumer used an interval-level scale to rate his or her satisfaction with it. Scores could range from 0 to 100, with higher scores indicating greater satisfaction. Consumers didn't know each other and were randomly assigned to three groups: (1) receive a store brand of shampoo in a bottle clearly labeled as such, (2) receive a premium brand of shampoo in the premium brand's bottle, or (3) receive a store brand of shampoo in the premium brand's bottle. Here are the collected data:

Store Brand, Store Bottle	Premium Brand, Premium Bottle	Store Brand, Premium Bottle	
70	85	85	
65	90	80	
65	95	90	
60	90	85	
$M = 65.00$ $s = 4.08$	$M = 90.00$ $s = 4.08$	$M = 85.00$ $s = 4.08$	$M_{Grand} = 80.00$

10.128 An environmental psychologist investigated ways to reduce waste. He randomly assigned office workers in small businesses to three groups: (1) to be in a control group, (2) to receive daily e-mail reminders about the importance of recycling, or (3) to have their current wastebaskets replaced with much smaller wastebaskets. To make sure they didn't influence each other, each worker was selected from a different small business. At the end of the week, the psychologist measured how many pounds of office waste each person had generated. Here are the data:

Control Group	Daily e-mail Reminders	Smaller Wastebaskets	
12	14	6	
14	12	4	
18	10	8	
12			
$M = 14.00$ $s = 2.83$	$M = 12.00$ $s = 2.00$	$M = 6.00$ $s = 2.00$	$M_{\text{Grand}} = 11.00$

Expand Your Knowledge

10.129 Which of the following is in descending order of size?

- **a.** $SS_{\text{Total}}, SS_{\text{Within}}, SS_{\text{Between}}$
- **b.** $SS_{\text{Total}}, SS_{\text{Between}}, SS_{\text{Within}}$
- **c.** $SS_{\text{Total}}, SS_{\text{Within}}, MS_{\text{Within}}$
- **d.** $SS_{\text{Within}}, SS_{\text{Total}}, MS_{\text{Within}}$
- **e.** $MS_{\text{Within}}, SS_{\text{Within}}, SS_{\text{Total}}$
- **f.** None of the above.

10.130 If treatment has an effect, then there is:

- **a.** more variability between groups than in total.
- **b.** more variability within groups than in total.
- **c.** more variability between groups than within groups.
- **d.** more variability within groups than between groups.
- **e.** None of the above.

For Exercises 10.131–10.138, indicate whether what is written could be true or is false.

10.131 $F(3, 18) = 10.98, p < .05$

10.132 $F(3, 25) = 1.98, p < .05$

10.133 $F(4, 20) = -17.89, p < .05$

10.134 $N = 14, k = 3$

10.135 $N = 6, k = 8$

10.136 If $\alpha = .05, q = 1.96$.

10.137 $SS_{\text{Between}} = 25, SS_{\text{Within}} = 10, SS_{\text{Total}} = 35$

10.138 $SS_{\text{Between}} = 12.50, SS_{\text{Within}} = 12.50, SS_{\text{Total}} = 25.00$

SPSS

The data for a between-subjects, one-way ANOVA have to be entered in the SPSS data editor in a specific way. Figure 10.16 contains the data for psychologists' salaries for states in the four census regions. The first column in the data editor tells which state the data come from. The second variable, the column labeled "region," contains information about which group a case is in. Here, a 1 indicates that a case is in the Northeast, a 2 indicates the Midwest, a 3 is for cases in the South, and a 4 is for the West. Note that SPSS does not need cases in a region to be grouped together. The final variable, the column labeled "psych," contains the cases' scores on the dependent variable, the mean salary for the state.

The command to start a one-way ANOVA can be found under "Analyze" and "Compare Means," as shown in Figure 10.17.

Clicking on "One-Way ANOVA. . ." opens up the commands shown in Figure 10.18. "psych," the dependent variable, has already been moved over to the box labeled "Dependent List." "region," the explanatory variable, is moved over to the box labeled "Factor."

	state	region	psych
1	Alaska	4	72.13
2	Connecticut	1	79.52
3	Delaware	3	66.89
4	Florida	3	68.60
5	Georgia	3	62.76
6	Idaho	4	61.49
7	Indiana	2	64.48
8	Kansas	2	55.96
9	Michigan	2	71.22
10	Mississippi	3	47.21
11	Nevada	4	67.12
12	New York	1	84.09
13	North Dakota	2	62.55
14	Rhode Island	1	82.10
15	Vermont	1	65.21
16	Washington	4	74.59

Figure 10.16 Data Entry in SPSS for a One-Way ANOVA At least two columns are needed to enter data for a one-way ANOVA in SPSS (just as for the independent-samples *t* test). One column, "region," contains information about which group a case belongs to and another column, "psych," has the case's score on the dependent variable. (Source: SPSS)

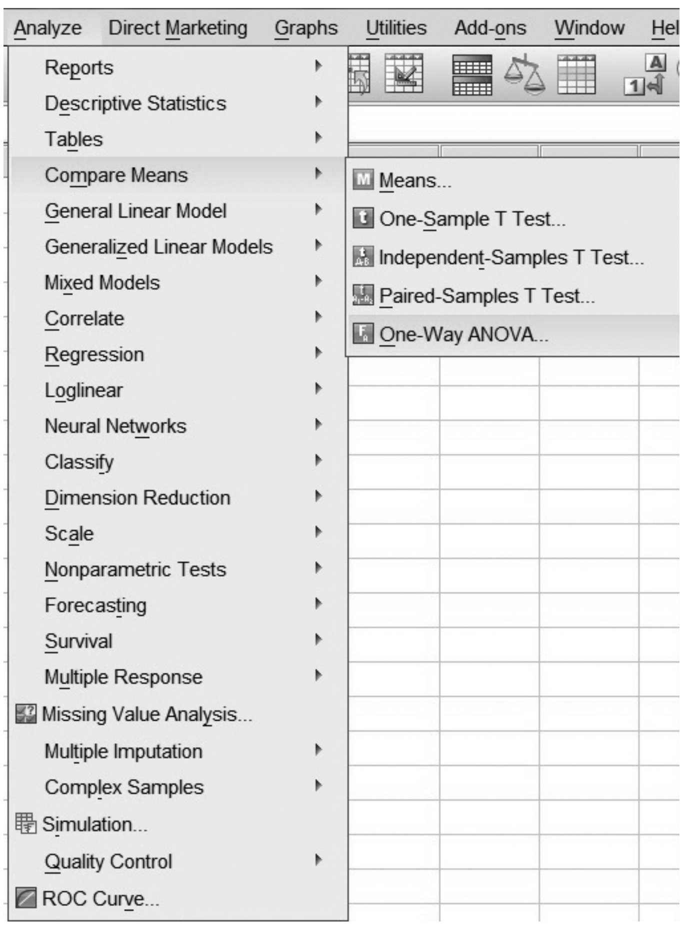

Figure 10.17 Starting a One-Way ANOVA in SPSS The command "One-Way ANOVA. . ." is found under the "Compare Means" heading under "Analyze." (Source: SPSS)

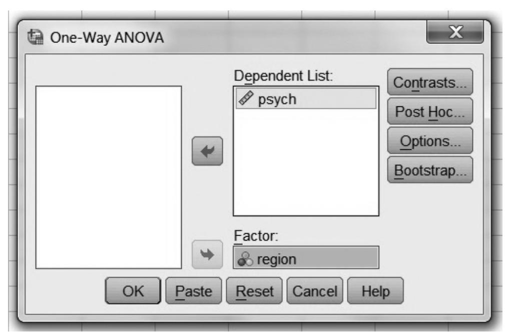

Figure 10.18 Defining the Dependent Variable and Explanatory Variable for a One-Way ANOVA in SPSS Once the dependent variable (psych) has been moved to "Dependent List:" and the explanatory variable (region) to "Factor:", clicking on the "OK" button will start the analysis of variance. (Source: SPSS)

Clicking on “OK” produces the output shown in Figure 10.19. Note that SPSS produces an ANOVA summary table like the one in Table 10.14. The SPSS summary table has an extra column on the right, labeled “Sig.” This column reports the exact significance level for the *F* ratio. If the value reported in the “Sig.” column is less than or equal to .05, then the results are statistically significant and the results are written in APA format as $p < .05$. If the “Sig.” level is greater than .05, then one has failed to reject the null hypothesis and the results are written in APA format as $p > .05$.

ANOVA

psych

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	636.407	3	212.136	3.517	.049
Within Groups	723.851	12	60.321		
Total	1360.258	15			

Figure 10.19 Example of One-Way ANOVA Output from SPSS SPSS produces an ANOVA summary table. The column labeled “Sig.” provides the exact significance level for the *F* ratio. If the “Sig.” value is $\leq .05$, the results are written as $p < .05$ in APA format. If the “Sig.” value is $> .05$, write the results as $p > .05$. (Source: SPSS)

Here, the exact significance level is .049, which is less than .05, so the null hypothesis is rejected. If the exact significance level is known, as it is here, then it should be reported in APA format. These results would be reported as $F(3, 12) = 3.52$, $p = .049$.

Once a researcher knows that the *F* ratio was statistically significant, he or she can proceed to select a post-hoc test by clicking the “Post Hoc” button seen in Figure 10.17. This opens up the selection of post-hoc tests seen in Figure 10.20. “Tukey,” which is the Tukey *HSD* test covered in this chapter, has been selected. If a researcher wants to change the significance level from .05, now is the time to do so.

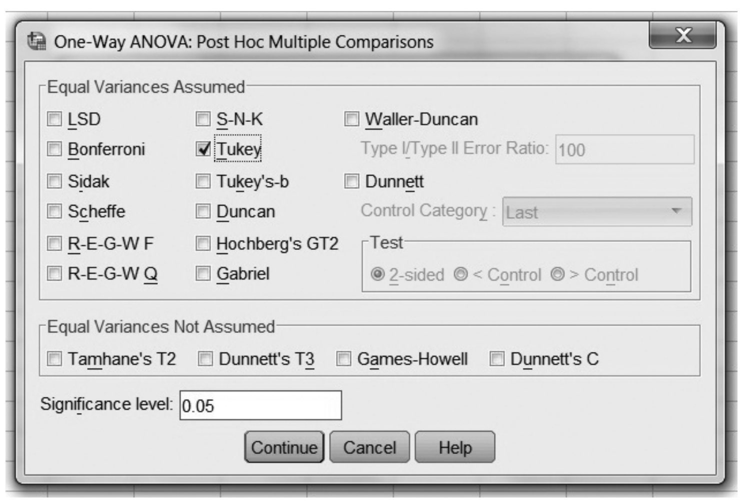

Figure 10.20 Choosing a Post-Hoc Test in SPSS This is the dialog box that opens up when one clicks on “Post Hoc” in Figure 10.18. Note the wide variety of post-hoc tests that are available. “Tukey,” which is Tukey’s *HSD*, has been selected. (Source: SPSS)

Multiple Comparisons

Dependent Variable: psych

Tukey HSD

(I) region	(J) region	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 North East	2 Midwest	14.17750	5.49185	.097	-2.1273	30.4823
	3 South	16.36500*	5.49185	.049	.0602	32.6698
	4 West	8.89750	5.49185	.404	-7.4073	25.2023
2 Midwest	1 North East	-14.17750	5.49185	.097	-30.4823	2.1273
	3 South	2.18750	5.49185	.978	-14.1173	18.4923
	4 West	-5.28000	5.49185	.773	-21.5848	11.0248
3 South	1 North East	-16.36500*	5.49185	.049	-32.6698	-.0602
	2 Midwest	-2.18750	5.49185	.978	-18.4923	14.1173
	4 West	-7.46750	5.49185	.546	-23.7723	8.8373
4 West	1 North East	-8.89750	5.49185	.404	-25.2023	7.4073
	2 Midwest	5.28000	5.49185	.773	-11.0248	21.5848
	3 South	7.46750	5.49185	.546	-8.8373	23.7723

* The mean difference is significant at the 0.05 level.

Figure 10.21 Output for Post-Hoc Tests in SPSS Each row in this table compares a pair of means. SPSS is redundant and compares each pair of means twice. For example, the first row (Northeast vs. Midwest) is the same as the fourth row (Midwest vs. Northeast). All that changes is the direction of the mean difference. Also, note the footnote on the table—asterisks that follow the mean difference indicate the difference is a statistically significant one. The exact significance level is found in the column labeled “Sig.” (Source: SPSS)

The output from the *HSD* test is shown in Figure 10.21. SPSS is redundant and has 12 rows in the table when it only needs 6. The first row, for example, compares the Northeast region to the Midwest region, and the fourth row turns this around to compare the Midwest region to the Northeast region. It is exactly the same comparison, though one finds a mean difference of 14.17750 and the other –14.17750. SPSS doesn’t report the *HSD* values, but it does report each mean difference that is statistically significant at the .05 level. It does this, as mentioned in the table footnote, by placing an asterisk after each mean difference that is statistically significant.