Between-Subjects, Two-Way Analysis of Variance

12

LEARNING OBJECTIVES

- Describe what two-way ANOVA does.
- Complete a between-subjects, two-way ANOVA.*
- Interpret a between-subjects, two-way ANOVA.

CHAPTER OVERVIEW

Chapter 11 introduced one-way ANOVA, a test that allows the examination of the impact of more than two levels of an explanatory variable at one time. Now, this chapter introduces two-way ANOVA, a technique that allows researchers to examine the impact of two explanatory variables at one time. Because each explanatory variable will have at least two levels, as shown in **Table 12.1**, a two-way ANOVA is used when at least four sample means are being compared.

TABLE 12.1 Design for the Simplest Type of Study to Be Analyzed with a Two-Way Analysis of Variance

	Level 1 of Explanatory Variable 2	Level 2 of Explanatory Variable 2
Level 1 of Explanatory Variable 1	Mean of Cell A	Mean of Cell B
Level 2 of Explanatory Variable 1	Mean of Cell C	Mean of Cell D

Two-way ANOVA is used in studies in which the effects of two explanatory variables are being studied at the same time. As each explanatory variable will have at least two levels, the simplest two-way ANOVA will analyze data from a study, like the one above, in which there are four samples (cells).

12.1 Introduction to Two-Way ANOVA

More complex tools are needed to understand, to take apart, more complex things. A basic tool, like a rock, will get you into a skull and show you that there is grey soft matter in there. But, to see finer details of the brain, a saw and a scalpel are more

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

helpful. As tools have evolved, think electron microscopes and MRI machines, so has our understanding of the brain.

Statistical tests are statisticians' tools. Tests like a two-sample t test are simple tools—all they allow a user to do is compare two groups (e.g., What is the impact of 0 vs. 1 drink of alcohol on driving?). One-way ANOVA allows more groups to be compared at once, so it allows more complex questions to be addressed (How does 0 vs. 1 vs. 2 vs. 3 vs. 4 drinks affect driving?). But, the most complex ANOVA questions involve the influence of multiple factors at once (What is the impact on driving of different doses of alcohol for men vs. women, depending on the time elapsed since the last meal?)

A "factor" is ANOVA-speak for an explanatory variable, an independent variable or a grouping variable that is thought to have some impact on the dependent variable. Factorial ANOVA can have two, three, four, even five factors or "ways." (The driving study had three: dose of alcohol, sex, and time since meal.) As the number of ways increases, a factorial ANOVA becomes harder to interpret. In this chapter, we'll limit ourselves to two factors and what is called two-way ANOVA.

To introduce two-way ANOVA, here's an example about the influence of two factors, nature *and* nurture, on personality. Suppose a researcher was studying factors that influence altruism and was interested in *both* how the children were reared (nurture) and what their nervous systems were like (nature). Using adopted children as participants, the researcher classified the children as falling into one of three levels—high, medium, or low—based on how altruistic their adoptive parents were. The adoptive parents reared them, so these three levels of altruism represented the influence of nurture. In addition, the researcher classified the children as being in one of two levels—high or low—based on how altruistic their birth parents were. Birth parents, who provide genetic material, represent the influence of nature. This two-way classification would allow the researcher to study the effect of both nature and nurture at the same time (see **Table 12.2**).

TABLE 12.2 Design for the Study of the Effect of Nature and Nurture on Altruism

	Adoptive Parents High on Altruism	Adoptive Parents Medium on Altruism	Adoptive Parents Low on Altruism
Birth Parents High on Altruism			
Birth Parents Low on Altruism			

This design is called a 3 × 2 design because there are three levels of one explanatory variable and two levels of the other. (It could also be called a 2 × 3 design.)

Each child who is a participant in the study is classified as belonging in one, and only one, of the six cells.

In Table 12.2, there are two levels of one grouping variable and three levels of the second grouping variable, so this two-way ANOVA would be called a 2 × 3 ANOVA (pronounced "two by three"). (Note that it doesn't matter which number goes first: It's fine to call it a 3 × 2 ANOVA.) Each child in the study would be classified as fitting into only one of the six cells in Table 12.2. This design would allow the researcher to examine the influence of two factors on children's altruism at once.

The type of two-way ANOVA being taught in this chapter is a *between-subjects* ANOVA. **Between-subjects** is an ANOVA term for independent samples. In a between-subjects design, different cases make up the different groups. **Within-subjects** is the ANOVA term for dependent samples. If different participants each rated the taste of a single ice cream (high-fat, low-fat, and no-fat), that would be a between-subjects design. If each taste-tester rated all three of the ice creams, that would be a within-subjects design.

A Common Question

Q Can an ANOVA have more than two ways?

A Absolutely. There can be a three-way ANOVA, a four-way ANOVA, and even more ways than that. The general term for a multiple-way ANOVA is **factorial ANOVA** (ways are also called *factors*). It is rare, however, that a researcher designs an experiment that needs more than a three-way ANOVA because the number of participants required becomes too large and the results become complicated to interpret.

There is an advantage to completing a single two-way ANOVA with two explanatory variables instead of two separate one-way tests, one for each of the explanatory variables. To explore this advantage, imagine a study that compares the final point total in a statistics class based on two explanatory variables: (1) type of instruction (whether students take the class in a classroom or online), and (2) level of math anxiety (whether students are high or low on math anxiety). Each factor has two levels (classroom vs. online, and high math anxiety vs. low math anxiety).

To conduct separate one-way tests, a researcher would need to do two separate studies:

- To examine the effect of type of instruction, the researcher would get a sample of students, assign some to take the course in a classroom, others to take it online, and then compare how much they learned.
- To examine the effect of anxiety, the researcher would put together another sample of students, classify them as high or low on math anxiety, have them all take the same course, and then compare how much they learned.

To conduct the study as a two-way ANOVA means examining both explanatory variables at once. The researcher would obtain *one* sample of students, classify them as high or low on math anxiety, and then assign half of each type to take the course in a classroom and half to take it online. That would give four groups:

- High-anxiety students taking the course in a classroom (Cell A)
- High-anxiety students taking the course online (Cell B)
- Low-anxiety students taking the course in a classroom (Cell C)
- Low-anxiety students taking the course online (Cell D)

The arrangement of the four groups in this study is diagrammed in Table 12.3. Notice how the two ways yield four cells arranged in rows (levels of math anxiety) and columns (where the course is taken). The two explanatory variables are **crossed**, which means that every level of one explanatory variable is paired with every level of the other explanatory variable.

TABLE 12.3 The Four Groups in a Two-Way ANOVA Examining the Impact of Math Anxiety and Type of Instruction on Learning

	Classroom Instruction	Online Instruction
High Math Anxiety	A	B
Low Math Anxiety	C	D

In a two-way ANOVA, there are two explanatory variables, each with at least two levels. Here, there are two levels of the row variable (degree of math anxiety) and two levels of the column variable (type of instruction). This results in four groups, one for each cell of the matrix.

There are two advantages to completing this study as a two-way ANOVA rather than as two one-way ANOVAs. First, it allows the study to be completed and analyzed in one pass, not as two different studies. Second, and more important, the person conducting the study gains more understanding because the variability in the dependent variable can now be divided into *three* effects. By doing the study as a two-way ANOVA, an effect has been gained.

The three effects being studied are:

- Does type of instruction affect learning?
- Is level of math anxiety related to learning?
- Do type of instruction and level of math anxiety *interact* to affect learning?

The first two effects are called **main effects**. Main effects examine the overall impact of an explanatory variable by itself. The third effect, the *interaction* effect, is unique to two-way ANOVA (and factorial ANOVA). An **interaction effect** occurs if the impact of one explanatory variable on the dependent variable depends on the level of the other explanatory variable.

So what exactly is an interaction effect? The easiest way to show what it is, is with a figure. Figure 12.1 gives the outcome for a 2 × 2 ANOVA of type of instruction and math anxiety on performance. When the lines are not parallel, then an interaction exists. That's what is seen in Figure 12.1, which illustrates that the low-math-anxiety people perform equally well whether in the classroom or online, while the high-math-anxiety people do more poorly when taking the class online. The two explanatory variables *interact* to determine the outcome.

When the lines are not parallel, an interaction exists.

A Common Question

Q How nonparallel do the lines have to be for an interaction to occur?

A If the lines are close to parallel, the interaction effect probably is not statistically significant. There should be a reasonable amount of deviation from parallel to speculate that an interaction effect has occurred.

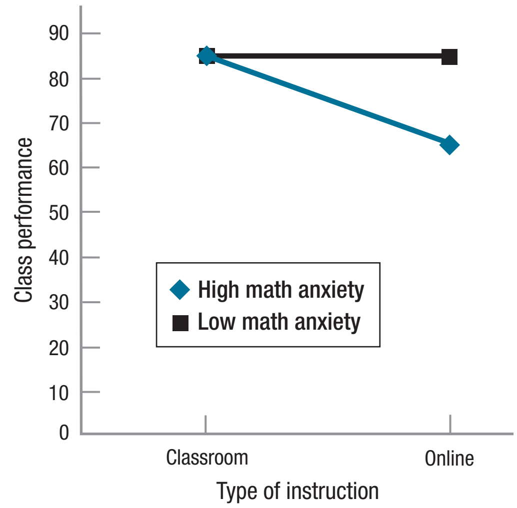


Figure 12.1 Interaction Between Level of Math Anxiety and Type of Instruction on Performance The nonparallel lines indicate that level of math anxiety and type of classroom instruction interact to determine performance. If asked whether classroom instruction or online instruction is better for learning statistics, the answer is that it depends on a student's level of math anxiety.

When an interaction takes place, the answer to the question of whether a main effect exists is, "It depends." Here are questions about the two main effects for the type of instruction/math anxiety study:

- Does type of instruction have an impact on how much one learns?
  - *It depends* on how much math anxiety one has. If a person is low on math anxiety, then he or she will learn equally well whether in the classroom or online. However, if a person is high on math anxiety, then he or she would be wise to take the course in a classroom.
- Is level of math anxiety related to how much one learns?
  - *It depends* on which course one takes. If a person is taking the course in a classroom, then math anxiety doesn't matter. However, if a person is taking the course online, then he or she will probably do better if he or she is low on math anxiety.

When there is an interaction, one needs to be careful about interpreting main effects. To understand why, look at the results for the type of instruction/level of math anxiety study, which are presented in Table 12.4.

Each of the cells has a mean for one of the four groups—a mean of 85 for high-anxiety students in the classroom, a mean of 65 for high-anxiety students online, and so on. Rather than attend to the cell means, pay attention to the means for each row and for each column. Each cell has the same sample size—and that will be true for all examples in this chapter—so the means for the cells in a row or column can be averaged to find the mean for that row or column. The formula for calculating row means and column means is shown in Equation 12.1.

TABLE 12.4 Cell Means, Row Means, and Column Means for the Effect of Level of Math Anxiety and Type of Instruction on Learning

	Classroom Instruction	Online Instruction	
High Math Anxiety	85	65	75.00
Low Math Anxiety	85	85	85.00
	85.00	75.00	

The sample size for each cell is the same, so a row mean or column mean can be calculated by averaging together the cell means in a row or a column. There is an interaction here (see Figure 12.1), so one shouldn't use the row means or column means to describe the main effects for rows or columns.

Equation 12.1 Formulas for Calculating Row Means and Column Means for Two-Way ANOVA When Each Cell Has the Same Sample Size

$$M_{\text{Row}} = \frac{\text{Add up all the cell means in a row}}{\text{The number of cells in the row}}$$

$$M_{\text{Column}} = \frac{\text{Add up all the cell means in a column}}{\text{The number of cells in the column}}$$

where M_{Row} = row mean
M_{Column} = column mean

The researcher would calculate the mean for the high-math-anxiety row as follows:

$$\begin{aligned} M_{\text{RowHigh}} &= \frac{\text{Add up all the means in the row}}{\text{The number of cells in the row}} \\ &= \frac{85 + 65}{2} \\ &= \frac{150.00}{2} \\ &= 75.00 \end{aligned}$$

The researcher would calculate the mean for the low-math-anxiety row as follows:

$$\begin{aligned} M_{\text{RowLow}} &= \frac{\text{Add up all the means in the row}}{\text{The number of cells in the row}} \\ &= \frac{85 + 85}{2} \\ &= \frac{170.00}{2} \\ &= 85.00 \end{aligned}$$

The researcher would calculate the mean for the classroom instruction column as follows:

$$
\begin{aligned}
M_{\text{ColumnClass}} &= \frac{\text{Add up all the means in the column}}{\text{The number of cells in the column}} \\
&= \frac{85 + 85}{2} \\
&= \frac{170.00}{2} \\
&= 85.00
\end{aligned}
$$

And, finally, here's the mean for the online instruction column:

$$
\begin{aligned}
M_{\text{ColumnOnline}} &= \frac{\text{Add up all the cell means in a column}}{\text{The number of cells in the column}} \\
&= \frac{65 + 85}{2} \\
&= \frac{150.00}{2} \\
&= 75.00
\end{aligned}
$$

What information do the row and column means provide?

- The row means give information about the main effect of level of anxiety on learning. Comparing the row mean of 75.00 for high-math-anxiety students to the row mean of 85.00 for the low-math-anxiety students suggests that level of anxiety predicts performance: it appears as if the higher the anxiety, the worse the performance. However, we already know, from the interaction effect, that how anxiety affects performance depends on the type of instruction. Though a main effect for type of anxiety exists, interpreting the interaction effect gives a better understanding of the results.
- The column means give information about the main effect of type of instruction on learning. Comparing the column mean of 85.00 for the classroom students to the column mean of 75.00 for the online students suggests that type of instruction affects performance: it seems as if students learn more in the classroom. However, we already know, from the interaction effect, that how type of instruction affects performance depends on a person's level of math anxiety. Again, though a main effect for type of instruction exists, interpreting the interaction seems to do a better job of explaining the results.

In a two-way ANOVA, when there is a statistically significant interaction, statistically significant main effects often play a background role.

This point is important: in a two-way ANOVA, when there is a statistically significant interaction, statistically significant main effects often play a background role.

To help make the difference between main effects and interactions clear, consider an example without an interaction. Suppose a physical therapist had equal numbers of men and women, then measured each person's hand strength. For half of each sex, the physical therapist measured hand strength in the dominant hand. For the other half, she measured hand strength in the nondominant hand.

The graph in **Figure 12.2** shows the results—men are stronger than women and the dominant hand is stronger than the nondominant hand. The two lines are parallel, so no interaction exists. In this case, it makes sense to interpret the main effects,

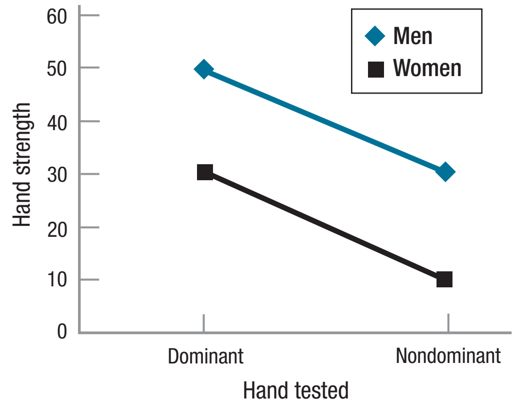


Figure 12.2 Lack of Interaction Between Sex and Tested Hand Strength The lines in this graph are parallel, indicating that sex and hand tested don't interact to determine hand strength. The graph shows a main effect for sex—men are stronger than women—and a main effect for hand tested—dominant hands are stronger than nondominant hands.

which can be seen by examining the row means and the column means found in Table 12.5. There is no "it depends" in answer to the two questions below:

- Is there a difference in hand strength by sex?
  - Yes. Men have more hand strength than women, both in the dominant hand and nondominant hand.
- Is there a difference in strength between dominant and nondominant hands?
  - Yes. Dominant hands are stronger than nondominant hands, both for men and for women.

TABLE 12.5 Effect of Sex and Hand Dominance on Hand Strength

	Dominant Hand	Nondominant Hand	
Men	50	30	40.00
Women	30	10	20.00
	40.00	20.00	

These results have no interaction (see Figure 12.2), so one can use the row means and column means to interpret the main effects for sex and hand tested. The main effects show that men are stronger than women and that dominant hands are stronger than nondominant hands.

Worked Example 12.1 For practice with main effects and interaction effects, let's look at some data on aggression in boys and girls. Suppose a developmental psychologist, Dr. O'Grady, administered an aggression scale to a random sample of 100 boys and 100 girls. Half of each sex had their level of physical aggression measured (getting into fights, throwing sticks and stones, etc.) and half had verbal aggression measured (spreading rumors, telling lies, etc.). On each scale, a higher score means more aggression.

Dr. O'Grady's study compares the means for groups that differ on two independent variables, sex and type of aggression, so it is an example where results are analyzed with a two-way ANOVA. Each way has two levels, so this is called a 2 × 2 ANOVA and there are four cells. The results from Dr. O'Grady's study can be found in Table 12.6, where each case fits in one, and only one, cell. For example, a boy who answered questions about physical aggressiveness would be in the top left cell.

TABLE 12.6 Measuring Aggression: Effect of Sex and Type of Aggression

	Physical Aggression	Verbal Aggression	
Boys	52	23	37.50
Girls	20	51	35.50
	36.00	37.00	

The graph of these results shows an interaction (see Figure 12.3), so one can't use the row means and column means to demonstrate main effects.

The row means for boys (37.50) and girls (35.50) don't show much of a main effect by sex. Boys are slightly more aggressive on average, but this 2-point difference could easily be due to sampling error. Even if the small difference indicated a difference between population means, it probably wouldn't be a meaningful difference.

The column means tell a similar story—the amount of physical aggression ($M = 36$) and the amount of verbal aggression ($M = 37$) are about equal. Verbal aggression is just slightly higher, but it seems possible that the levels of verbal aggression and physical aggression are the same in the populations and that the difference observed between the samples could easily be due to sampling error. Based on these column means, it seems that there is little difference in the use of the two types of aggression.

For a different view, look at the interaction effect graphed in Figure 12.3.

- Are boys more aggressive than girls?
  - *It depends* on the type of aggression. Boys are more aggressive physically, and girls are more aggressive verbally.
- Is physical aggression as likely to occur as verbal aggression?
  - *It depends* on a person's sex. Physical aggression is more likely to occur in boys and verbal aggression in girls.

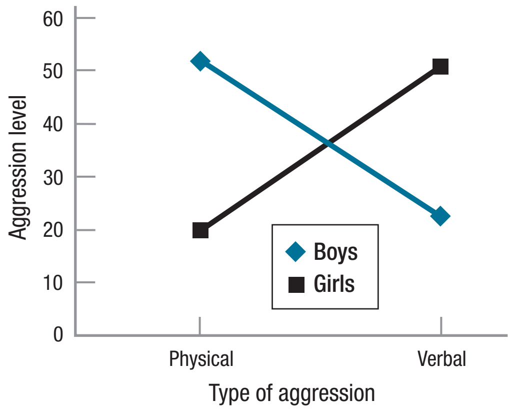


Figure 12.3 Interaction Between Sex and Type of Aggression The nonparallel lines indicate that the two variables, sex and type of aggression, interact. Are boys more aggressive than girls? It depends on the type of aggression. Is physical aggression more common than verbal aggression? It depends on whether one is talking about boys or girls.

It is clear in this example that there is an interaction effect. This is apparent both from the nonparallel lines in Figure 12.3 and from the use of "it depends" in interpreting the results. To reiterate: When an interaction effect exists, be cautious interpreting the main effects.

Practice Problems 12.1

Apply Your Knowledge

12.01 Classify each scenario in terms of the number of ways and the number of levels each way has. For example, classify each scenario as a 2 × 3 design, a 2 × 3 × 5 design, or some other variation.

a. Men and women who are right-handed or left-handed and who use razors with blades vs. electric razors are compared in terms of satisfaction with the smoothness of their shaves.

b. Male and female students studying for traditionally female careers (e.g., nursing) or for traditionally male careers (e.g., engineering) are compared in terms of how androgynous they are.

12.02 Each cell in the matrix below reports the mean of six cases. (a) Calculate the row and column means. (b) Interpret the two main effects.

	Condition 1	Condition 2
Group 1	12.00	18.00
Group 2	3.00	9.00

12.03 Thirty-six cases were randomly divided into four samples. Each person took either a low dose or a high dose of a drug and was queried about either the physical side effects or psychological side effects. Each cell in the matrix below reports the mean number of side effects reported by a sample of nine cases. (a) Graph the cell means. (b) Decide if there is an interaction. (c) Interpret the effect of the two conditions on the two groups.

	Physical Side Effects	Psychological Side Effects
Low Dose	10.00	10.00
High Dose	10.00	18.00

12.2 Calculating a Between-Subjects, Two-Way ANOVA

A two-way ANOVA allows a researcher to see, at one time, the effects of two explanatory variables by themselves (the main effects) and in combination (the interaction). A two-way ANOVA achieves this by partitioning the between-group variability differently. As shown in Figure 12.4, a two-way ANOVA takes the total variability and separates that into within-group variability and between-group variability. It then takes the between-group variability and divides that further into the two main effects, the effect of the row explanatory variable and the effect of the column explanatory variable, and the effect of the interaction of the two main effects. Each of these three effects is tested for statistical significance with its own *F* ratio.

To learn how to complete the calculations for a between-subjects, two-way ANOVA, here's an example about the effect of caffeine consumption and sleep

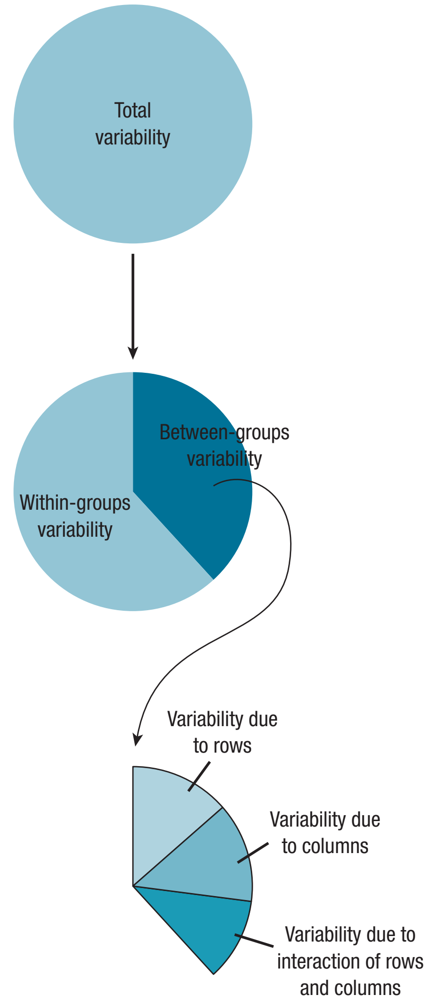


Figure 12.4 Partitioning Variability for Between-Subjects, Two-Way ANOVA This figure shows how the variability for a between-subjects, two-way ANOVA is initially partitioned into between-group variability and within-group variability, as for a between-subjects, one-way ANOVA. The between-group variability is then divided further into three parts: that due to the row variable, the column variable, and the interaction of the two.

deprivation on mental alertness. Imagine that a sleep researcher, Dr. Ballard, obtained 30 participants who were students at his college. He gave each a mental alertness task one hour after waking. (Higher scores on this test indicate more mental alertness.) The participants were randomly assigned into six groups, with five in each group. Half the participants consumed a standard cup of coffee (150 mg of caffeine) 30 minutes after waking and half did not. One third of the participants were allowed a full night's sleep (0-hours sleep deprivation), one third were awakened an hour early (1-hour sleep deprivation), and the final third were awakened two hours early (2-hours sleep deprivation).

The design of the study is shown in Table 12.7. Note that the six cells are laid out in a 3×2 format as the two explanatory variables are crossed. The two levels for dose of caffeine are the two rows and the three levels for sleep deprivation are the three columns. Each of the 30 participants is in one and only one cell.

Table 12.7 displays the cell means, row means, and column means, so Dr. Ballard can speculate about main effects. The graph in Figure 12.5 allows him to consider the presence of an interaction.

Looking at Table 12.7 and Figure 12.5, three effects are apparent: (1) a main effect for caffeine, with those receiving caffeine performing better; (2) a main effect for

TABLE 12.7 Mental Alertness Descriptive Statistics for Caffeine/Sleep Deprivation Data

	0-hours Sleep Deprivation	1-hour Sleep Deprivation	2-hours Sleep Deprivation	
Caffeine	88.00 (3.87)	90.00 (4.69)	86.00 (4.27)	88.00
No Caffeine	80.00 (4.79)	68.00 (4.43)	50.00 (5.07)	66.00
	84.00	79.00	68.00	

Means (and standard deviations) for mental alertness scores are reported for each cell. Row means and column means are also reported.

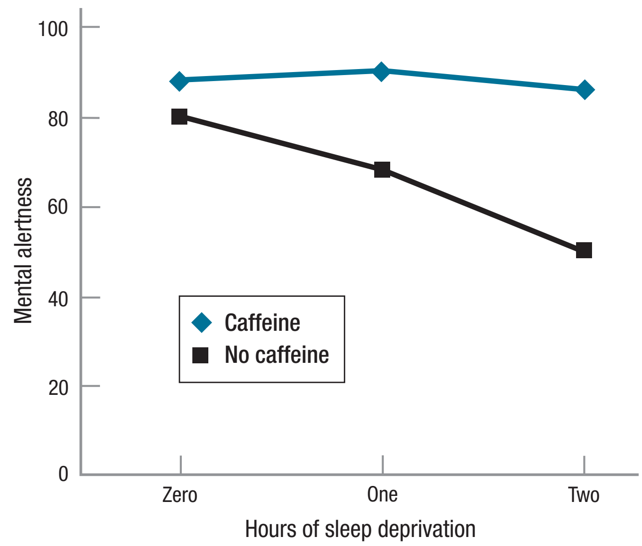


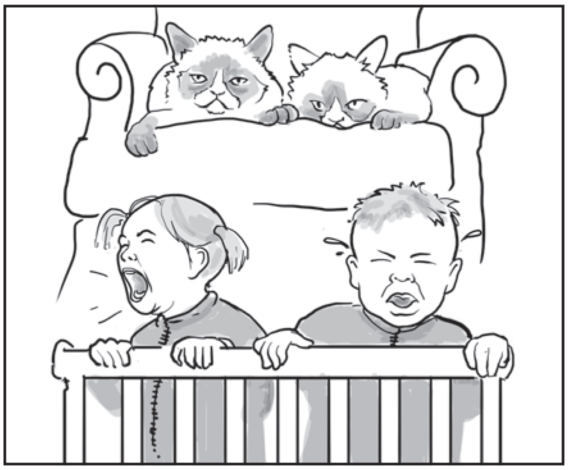
Figure 12.5 Effect of Caffeine Consumption and Sleep Deprivation on Mental Alertness This graph has nonparallel lines. That means there may be an interaction between consumption of caffeine and number of hours of sleep deprivation on mental alertness.

sleep deprivation, with those who were more sleep-deprived performing worse; and (3) an interaction between the two variables. Of course, Dr. Ballard will need to do a statistical test to see if any of these results is statistically significant. And, if the interaction effect is statistically significant, that will take precedence over the main effects.

Conducting a between-subjects, two-way ANOVA requires following the same six steps for hypothesis testing as in previous hypothesis tests.

Step 1 Pick a Test

Dr. Ballard is comparing the means of six groups, formed by the crossing of two independent variables (caffeine consumption and sleep deprivation). The groups are independent samples, so he'll use a between-subjects, two-way ANOVA. Specifically, it is a 2×3 ANOVA, as there are two levels of caffeine consumption (consume caffeine or not consume caffeine) and three levels of sleep deprivation (0, 1, or 2 hours of deprivation).



Tom **a**nd **H**arry **d**espise **c**rabby **i**nfants.

Step 2 Check the Assumptions

The assumptions for a two-way ANOVA are the same as they are for a between-subjects, one-way ANOVA: random samples, independence of observations, normality, and homogeneity of variance.

- *Random samples.* Dr. Ballard would love to be able to draw a conclusion from his study about humans in general. But, he doesn't have a random sample of participants from the human population, so the random samples assumption is violated. This is a robust assumption, however, so he can proceed with the two-way ANOVA. He'll need to be careful about generalizing the results.
- *Independence of observations.* Each participant was tested individually, so their results don't influence each other and the independence of samples assumption is not violated. This assumption is not robust, so if it had been violated, the analyses couldn't proceed.
- *Normality.* Dr. Ballard is willing to assume that the dependent variable, mental alertness, is normally distributed in the larger population, so this assumption is not violated. This assumption is robust to violation, especially if the sample size is large.
- *Homogeneity of variance.* All six of the cell standard deviations are about the same (see Table 12.7), which means that the variability in each group is about the same. This assumption is not violated in the study. (And, when the sample size is large, it is a robust assumption.)

For the caffeine/sleep deprivation study, no nonrobust assumptions were violated, so Dr. Ballard can proceed with the ANOVA.

Step 3 List the Hypotheses

For a one-way ANOVA, there is one set of hypotheses. For a two-way ANOVA, with a null hypothesis and an alternative hypothesis for both of the main effects and for the interaction effect, there are three sets of hypotheses.

- For the row main effect, the null hypothesis states that the population means for all levels of the row variable are equal to each other. The alternative hypothesis for the row main effect states that not all population means for the levels of the row variable are the same.
- For the column main effect, the null hypothesis says that the population means for all levels of the column variable are the same. The alternative hypothesis for the column main effect says that not all population means for the levels of the column variable are the same.
- For the interaction effect, the null hypothesis says that there is no interaction effect in the population. This means that the impact of one main effect on the dependent variable is independent of the impact of the other main effect on the dependent variable for all the cells. The alternative hypothesis for the interaction effect says that there is an interaction effect for at least one cell.

For the caffeine/sleep deprivation study, the row variable, caffeine consumption, has two levels, so Dr. Ballard writes the hypotheses:

$$H_{0\text{ Rows}}: \mu_{\text{Row1}} = \mu_{\text{Row2}}$$
$$H_{1\text{ Rows}}: \mu_{\text{Row1}} \neq \mu_{\text{Row2}}$$

There are three levels of the column variable, sleep deprivation, so the column null hypothesis could be written two ways: (1) $\mu_{\text{Column1}} = \mu_{\text{Column2}} = \mu_{\text{Column3}}$ or (2) all column population means are the same.

$H_{0\text{ Columns}}$: All column population means are the same.

$H_{1\text{ Columns}}$: At least one column population mean is different from at least one other column population mean.

There is no easy way to write the hypotheses for the interaction effect symbolically. So, using plain language, Dr. Ballard writes the hypotheses as

$H_{0\text{ Interaction}}$: There is no interactive effect of the two independent variables on the dependent variable in the population.

$H_{1\text{ Interaction}}$: The two independent variables in the population interact to affect the dependent variable in at least one cell.

Step 4 Set the Decision Rules

Just as three sets of hypotheses exist, there will be three decision rules for a two-way ANOVA—one for the row effect, one for the column effect, and one for the interaction effect. Because this is an ANOVA, F ratios will be calculated for each of the three effects (F_{Rows}, F_{Columns}, and $F_{\text{Interaction}}$) and compared to their critical values of F($F_{cv\text{ Rows}}$, $F_{cv\text{ Columns}}$, and $F_{cv\text{ Interaction}}$). If the observed value of F is greater than or equal to F_{cv} for an effect, the null hypothesis is rejected for that effect, the alternative hypothesis accepted, and the effect is called statistically significant.

To find a critical value of F for an F ratio, a researcher needs to decide on the alpha level and know the degrees of freedom for the F ratio. The most commonly used alpha levels for ANOVA are .05 and .01, which correspond to a 5% chance and a 1% chance of making a Type I error. (Type I error occurs when the null hypothesis is erroneously rejected.) Typically, alpha is set at .05. When the consequences of a Type I error are more severe, alpha is set at .01 instead. Dr. Ballard, of course, wants to avoid Type I error, but he can live with a 5% chance of it occurring. So, he sets alpha at .05.

Now he needs to calculate degrees of freedom both for the numerator and the denominator of the F ratios. The degrees of freedom for the numerator will change from F ratio to F ratio, moving from df_{Rows} to df_{Columns} to $df_{\text{Interaction}}$ as a researcher moves from effect to effect. The degrees of freedom for the denominator term for the between-subjects, two-way ANOVA is degrees of freedom within, df_{Within}, and it will remain so across all three of the hypothesis tests. Formulas for calculating all 4 of the degrees of freedom needed to find the critical values of F are shown in Equation 12.2. Table 12.9 (on page 437) shows how to calculate the other 2 degrees of freedom—between groups degrees of freedom and total degrees of freedom—that will be needed to complete the ANOVA summary table.

Equation 12.2 Formulas for Calculating Degrees of Freedom for Between-Subjects, Two-Way ANOVA

$$df_{\text{Rows}} = R - 1$$
$$df_{\text{Columns}} = C - 1$$
$$df_{\text{Interaction}} = df_{\text{Rows}} \times df_{\text{Columns}}$$
$$df_{\text{Within}} = N - (R \times C)$$

where df_{Rows} = degrees of freedom for the row main effect
df_{Columns} = degrees of freedom for the column main effect
$df_{\text{Interaction}}$ = degrees of freedom for the interaction effect
df_{Within} = degrees of freedom for the within-group effect
R = number of rows
C = number of columns
N = total number of cases

For the caffeine/sleep deprivation study, there are two rows. Here are Dr. Ballard's calculations for the degrees of freedom for the row main effect:

$$\begin{aligned} df_{\text{Rows}} &= R - 1 \\ &= 2 - 1 \\ &= 1 \end{aligned}$$

With three columns, his calculations for the degrees of freedom for the column main effect look like this:

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

Once df_{Rows} and df_{Columns} are known, it is easy to calculate the numerator degrees of freedom for the interaction effect:

$$\begin{aligned} df_{\text{Interaction}} &= df_{\text{Rows}} \times df_{\text{Columns}} \\ &= 1 \times 2 \\ &= 2 \end{aligned}$$

Finally, checking back and seeing that there were 30 cases, Dr. Ballard calculates the degrees of freedom for the within-group effect:

$$\begin{aligned} df_{\text{Within}} &= N - (R \times C) \\ &= 30 - (2 \times 3) \\ &= 30 - 6 \\ &= 24 \end{aligned}$$

Once all the degrees of freedom have been calculated, they can be used to find the three critical values of F (see Table 12.8). The row main effect has 1 degree of freedom in the numerator, that's df_{Rows}, and 24 in the denominator, df_{Within}. Dr. Ballard looks in Appendix Table 4, at the intersection of the column with 1 degree of freedom and the row with 24 degrees of freedom in the F critical values table for $\alpha = .05$.

TABLE 12.8 Guide to Finding the Critical Value of *F* for Each Effect in a Between-Subjects, Two-Way ANOVA

Effect	Numerator *df*	Denominator *df*
Rows	df_{Rows}	df_{Within}
Columns	df_{Columns}	df_{Within}
Interaction	$df_{\text{Interaction}}$	df_{Within}

The critical values of F, F_{cv}, are found in Appendix Table 4 at the intersection of the column for the correct numerator degrees of freedom with the row for the correct denominator degrees of freedom. For a between-subjects, two-way ANOVA, this table reveals what the numerator and denominator degrees of freedom are for each effect.

There, he finds that the critical value of F, with $\alpha = .05$, is 4.260 for the row main effect: $F_{cv\ \text{Rows}} = 4.260$.

Next, he finds the critical value of F for the column main effect. In Appendix Table 4, he uses the column with 2 degrees of freedom and the row with 24 degrees of freedom to find that F_{cv} for the column effect is 3.403: $F_{cv\ \text{Columns}} = 3.403$. The interaction effect has the same numerator and denominator degrees of freedom as the column effect, 2 and 24. So, the critical value of F for the interaction effect is the same as the column effect: $F_{cv\ \text{Interaction}} = 3.403$.

Dr. Ballard can now write the decision rules for the three effects for the caffeine/sleep deprivation study as shown:

- Row main effect (caffeine)
  - If $F_{\text{Rows}} \geq 4.260$, reject $H_{0\ \text{Rows}}$.
  - If $F_{\text{Rows}} < 4.260$, fail to reject $H_{0\ \text{Rows}}$.
- Column main effect (sleep deprivation)
  - If $F_{\text{Columns}} \geq 3.403$, reject $H_{0\ \text{Columns}}$.
  - If $F_{\text{Columns}} < 3.403$, fail to reject $H_{0\ \text{Columns}}$.
- Interaction effect
  - If $F_{\text{Interaction}} \geq 3.403$, reject $H_{0\ \text{Interaction}}$.
  - If $F_{\text{Interaction}} < 3.403$, fail to reject $H_{0\ \text{Interaction}}$.

Step 5 Calculate the Test Statistics

Between-subjects, two-way ANOVA starts the same way between-subjects, one-way ANOVA does. It separates the total variability in a set of scores into between-group variability and within-group variability. Then it goes a step further and separates between-group variability into three subcomponents—variability due to the row effect, variability due to the column effect, and variability due to the interaction effect. Finally, each of these three effects is tested individually with its own F ratio.

The sources of variability in the data are separated out by calculating sums of squares, as was done for the one-way ANOVA. The sums of squares that need to be calculated are:

- Sum of squares between groups (SS_{Between})
- Sum of squares rows (SS_{Rows})
- Sum of squares columns (SS_{Columns})
- Sum of squares interaction ($SS_{\text{Interaction}}$)
- Sum of squares within groups (SS_{Within})
- Sum of squares total (SS_{Total})

Remember that a sum of squares in an ANOVA is a sum of squared deviation scores (see Chapter 11). What scores are used and what mean is subtracted from them vary depending on the source of variability being calculated. The sums of squares are arranged in an ANOVA summary table, **Table 12.9**, along with the formulas to complete the other necessary values. Note that the same columns that were used for a one-way and repeated-measures ANOVA summary table—source of variability, sum of squares, degrees of freedom, mean square, and *F* ratio—are present in the same order in the summary table for the two-way ANOVA. But, the sources of variability are different. Making sure that the correct order is followed for the rows and columns in an ANOVA summary table is key to completing an ANOVA. The three effects that are being tested fall under the between-groups effect. Note how they are indented in our table to show that they are derived from the between-groups effect.

TABLE 12.9 Template for ANOVA Summary Table for Between-Subjects, Two-Way ANOVA

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	$SS_{\text{Rows}} + SS_{\text{Columns}} + SS_{\text{Interaction}}$	$df_{\text{Rows}} + df_{\text{Columns}} + df_{\text{Interaction}}$		
Rows	Given	$R - 1$	$\frac{SS_{\text{Rows}}}{df_{\text{Rows}}}$	$\frac{MS_{\text{Rows}}}{MS_{\text{Within}}}$
Columns	Given	$C - 1$	$\frac{SS_{\text{Columns}}}{df_{\text{Columns}}}$	$\frac{MS_{\text{Columns}}}{MS_{\text{Within}}}$
Interaction	Given	$df_{\text{Rows}} \times df_{\text{Columns}}$	$\frac{SS_{\text{Interaction}}}{df_{\text{Interaction}}}$	$\frac{MS_{\text{Interaction}}}{MS_{\text{Within}}}$
Within groups	Given	$N - (R \times C)$	$\frac{SS_{\text{Within}}}{df_{\text{Within}}}$	
Total	$SS_{\text{Between}} + SS_{\text{Within}}$	$N - 1$		

Be sure to label and order the rows and columns in a between-subjects, two-way ANOVA summary table as they are here. Note that some cells are left blank.

Here are the sums of squares for the row effect, the column effect, the interaction effect, and for within-group variability for the caffeine/sleep deprivation study (formulas for calculating sums of squares for a two-way ANOVA are given in an appendix to this chapter):

- $SS_{\text{Rows}} = 3{,}630.00$
- $SS_{\text{Between}} = 5{,}950.00$
- $SS_{\text{Columns}} = 1{,}340.00$
- $SS_{\text{Interaction}} = 980.00$
- $SS_{\text{Within}} = 748.00$
- $SS_{\text{Total}} = 6{,}698.00$

Equation 12.2 has already been used to calculate four degrees of freedom—rows, columns, interaction, and within. Table 12.9 shows how to calculate the other two degrees of freedom—between groups and total. Between-groups variability is broken down into variability for rows, columns, and interaction, so degrees of freedom between groups, df_{Between}, is found by adding up df_{Rows}, df_{Columns}, and $df_{\text{Interaction}}$. For the caffeine/sleep deprivation study, this is

$$\begin{aligned} df_{\text{Between}} &= df_{\text{Rows}} + df_{\text{Columns}} + df_{\text{Interaction}} \\ &= 1 + 2 + 2 \\ &= 5 \end{aligned}$$

As shown in Table 12.9, the degrees of freedom total, df_{Total}, is calculated by subtracting 1 from the total number of subjects:

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

The next step is to calculate mean squares for rows, columns, interaction, and within groups. Following the instructions in Table 12.9:

$$\begin{aligned} MS_{\text{Rows}} &= \frac{SS_{\text{Rows}}}{df_{\text{Rows}}} \\ &= \frac{3{,}630.00}{1} \\ &= 3{,}630.00 \end{aligned}$$

$$\begin{aligned} MS_{\text{Columns}} &= \frac{SS_{\text{Columns}}}{df_{\text{Columns}}} \\ &= \frac{1{,}340.00}{2} \\ &= 670.00 \end{aligned}$$

$$\begin{aligned} MS_{\text{Interaction}} &= \frac{SS_{\text{Interaction}}}{df_{\text{Interaction}}} \\ &= \frac{980.00}{2} \\ &= 490.00 \end{aligned}$$

$$\begin{aligned} MS_{\text{Within}} &= \frac{SS_{\text{Within}}}{df_{\text{Within}}} \\ &= \frac{748.00}{24} \\ &= 31.1667 \\ &= 31.17 \end{aligned}$$

As directed in Table 12.9, Dr. Ballard finds the three F ratios by dividing the mean squares for the row main effect, the column main effect, and the interaction effect by the mean square for within-group variability:

$$\begin{aligned} F_{\text{Rows}} &= \frac{MS_{\text{Rows}}}{MS_{\text{Within}}} \\ &= \frac{3{,}630.00}{31.17} \\ &= 116.4581 \\ &= 116.46 \end{aligned}$$

$$\begin{aligned} F_{\text{Columns}} &= \frac{MS_{\text{Columns}}}{MS_{\text{Within}}} \\ &= \frac{670.0}{31.17} \\ &= 21.4950 \\ &= 21.50 \end{aligned}$$

$$\begin{aligned} F_{\text{Interaction}} &= \frac{MS_{\text{Interaction}}}{MS_{\text{Within}}} \\ &= \frac{490.00}{31.17} \\ &= 15.7202 \\ &= 15.72 \end{aligned}$$

With all the F ratios calculated, the ANOVA summary table is complete (see Table 12.10). We'll come back to see how the results are interpreted after getting more practice with calculations.

TABLE 12.10 Completed ANOVA Summary Table Showing the Effects of Caffeine and Sleep Deprivation on Mental Alertness

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Between groups	5,950.00	5		
Rows	3,630.00	1	3,630.00	116.46
Columns	1,340.00	2	670.00	21.50
Interaction	980.00	2	490.00	15.72
Within groups	748.00	24	31.17	
Total	6,698.00	29		

The effect of caffeine consumption is the row variable, and the effect of sleep deprivation is the column variable.

Worked Example 12.2 The question "What makes relationships work?" fills the covers of supermarket magazines, but it also interests research psychologists. Dr. Larue, a social psychologist, conducted a study to investigate two variables that she believed were associated with relationship satisfaction. The variables were (1) arguing style and (2) type of parental relationship model.

Dr. Larue found college seniors who were in serious relationships and assessed their ability to argue in a positive or negative way. People who argue positively don't become threatened or defensive, don't attack their partners, and help arguments reach a successful conclusion that is satisfactory to both sides. Dr. Larue classified students as positive arguers, mixed arguers, or negative arguers.

Dr. Larue also asked these students to rate the quality of their parents' marriage as being good, average, or bad. With three levels of parental marital quality and three levels of arguing style, there were nine cells in Dr. Larue's design. Dr. Larue randomly selected eight students from each cell. That is, there were eight who perceived their parents' marriages as good and who were positive arguers, eight who rated their parents' marriages as good and who had a mixed arguing style, and so on. With nine cells and eight participants per cell, Dr. Larue had a total of 72 participants.

Dr. Larue then measured the second grouping variable by asking each student to rate his or her satisfaction with his or her current relationship. (Dr. Larue made sure that no one in the study was in a relationship with someone else in the study.) The interval-level satisfaction scores could range from a low of 5 to a high of 35.

Table 12.11 shows the mean (and standard deviation) for each cell, as well as the row means and the column means. Figure 12.6 displays the data graphically. Looking at Figure 12.6 and Table 12.11 together, there are three things to note:

1. It appears that there is a row main effect for arguing style. Mean satisfaction scores decrease as the amount of negative arguing increases, from 27.92 for positive arguers, to 21.25 for mixed arguers, and 16.00 for negative arguers.
2. It appears that there is a column main effect for perceived quality of parental marriage. Mean satisfaction scores decrease as the perception of parental marriage becomes more negative, from 24.42 for those with a positive perception, to 22.08 for those with an average perception, and down to 18.67 for those with a negative perception.

TABLE 12.11 Descriptive Statistics for Relationship Satisfaction Study

	Parents' Marriage Perceived as "Good"	Parents' Marriage Perceived as "Average"	Parents' Marriage Perceived as "Bad"	
Positive Arguing Style	30.50 (3.07)	28.25 (2.43)	25.00 (2.93)	27.92
Mixed Arguing Style	23.75 (3.33)	21.50 (2.45)	18.50 (3.21)	21.25
Negative Arguing Style	19.00 (2.67)	16.50 (2.07)	12.50 (3.21)	16.00
	24.42	22.08	18.67	

Means (and standard deviations) for satisfaction with the current relationship are reported for each cell. Row means and column means are also reported. Satisfaction scores can range from a low of 5 to a high of 35.

3. The three lines are parallel, so there appears to be no interaction between arguing style and quality of parental marriage on relationship satisfaction. Of course, Dr. Larue needs to do a statistical test to see if her observations pan out statistically.

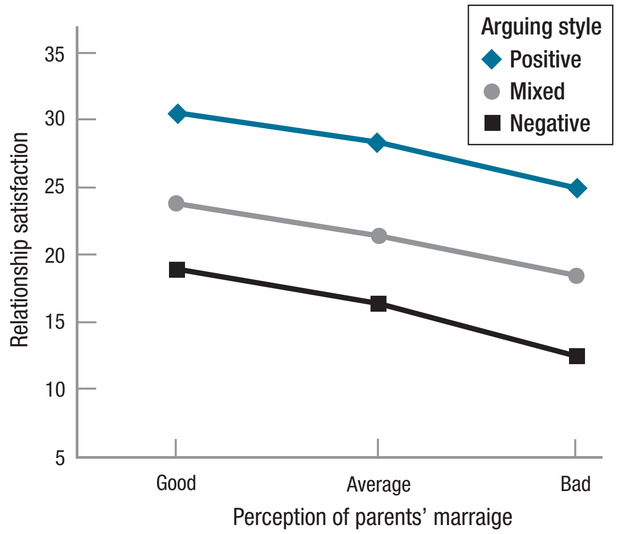


Figure 12.6 Impact of Arguing Style and Perception of Parental Marriage on Relationship Satisfaction This graph shows a main effect of arguing style and a main effect of perception of parents' marriage on relationship satisfaction. The lines are parallel, so there is no interaction between the two grouping variables.

Step 1 Pick a Test. There are two grouping variables (arguing style and parental marriage). Each grouping variable has three levels (positive, mixed, and negative arguing style; good, average, or bad parental marriage). When three levels of arguing style are crossed with three levels of parental marriage, this forms nine groups. Each case is in just one group and each case is not paired with another case, so the groups are all independent. That means this is a between-subjects design. Comparing the means of groups defined by two grouping variables and of groups that are independent samples calls for a between-subjects, two-way ANOVA. Specifically, this is a 3×3 ANOVA.

Step 2 Check the Assumptions.

- *Random samples.* The cases used in this study are random samples from the populations of seniors at Dr. Larue's university who are in relationships and fit in one of the nine cells. This assumption is not violated as long as that is the population to which she wishes to generalize her results.
- *Independence of observations.* No person in the study was in a relationship with anyone else in the study, so this assumption is not violated.
- *Normality.* Dr. Larue is willing to assume that relationship satisfaction is normally distributed within each of the nine populations represented here. Therefore, this assumption is not violated.
- *Homogeneity of variance.* The standard deviations for all nine groups are shown in Table 12.11 and they are all about the same. No standard deviation is twice another, so this assumption is not violated.

Dr. Larue can proceed with the planned statistical test.

Step 3 List the Hypotheses. There are three hypotheses in a two-way ANOVA: one for the row main effect (arguing style), one for the column main effect (parental marriage quality), and one for the interaction effect (arguing style × parental marriage quality).

- Row main effect:

$$H_{0\text{ Rows}}: \mu_{\text{Row1}} = \mu_{\text{Row2}} = \mu_{\text{Row3}}$$
$$H_{1\text{ Rows}}: \text{At least one row population mean is different from at least one other.}$$

- Column main effect:

$$H_{0\text{ Columns}}: \mu_{\text{Column1}} = \mu_{\text{Column2}} = \mu_{\text{Column3}}$$
$$H_{1\text{ Columns}}: \text{At least one column population mean is different from at least one other.}$$

- Interaction effect:

$H_{0\text{ Interaction}}$: There is no interactive effect of the two grouping variables on the dependent variable in the population.

$H_{1\text{ Interaction}}$: The two grouping variables interact to affect the dependent variable in at least one cell in the population.

Step 4 Set the Decision Rules. Setting the decision rules requires deciding on an alpha level and knowing the degrees of freedom for the numerator and the degrees of freedom for the denominator for each of the three effects being tested. Dr. Larue is comfortable having a 5% chance of Type I error, so she sets alpha at .05.

Next, to calculate degrees of freedom, she needs to know R (the number of rows), C (the number of columns), and N (the total number of cases):

- $R = 3$ (There are three levels of arguing style.)
- $C = 3$ (There are three levels of parental marriage quality.)
- $N = 72$ (Each cell has eight cases and there are nine cells: $8 \times 9 = 72$.)

Using Equation 12.2, Dr. Larue calculates the various degrees of freedom:

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

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

$$\begin{aligned} df_{\text{Interaction}} &= df_{\text{Rows}} \times df_{\text{Columns}} \\ &= 2 \times 2 \\ &= 4 \end{aligned}$$

$$\begin{aligned} df_{\text{Within}} &= N - (R \times C) \\ &= 72 - (3 \times 3) \\ &= 72 - 9 \\ &= 63 \end{aligned}$$

Table 12.8 offers guidance as to which source provides the numerator and denominator degrees of freedom for each effect. The critical value of F for the main effect of rows has 2 degrees of freedom in the numerator (df_{Rows}) and 63 degrees of freedom in the denominator (df_{Within}). Looking in Appendix Table 4 at the intersection of the column with 2 degrees of freedom and the row with 63 degrees of freedom, Dr. Larue discovers no such row exists. In these situations, apply *The Price Is Right* rule and use the degrees of freedom value that is closest without going over. Here, that is 60, which makes $F_{cv\text{ Rows}} = 3.150$.

This study has the same degrees of freedom for the numerator and the denominator for the column main effect, 2 and 63. This means $F_{cv\text{ Columns}} = 3.150$.

The degrees of freedom for the numerator for the interaction effect ($df_{\text{Interaction}}$) are 4 and the denominator degrees of freedom (df_{Within}) are 63. Using Table 4 in the Appendix and *The Price Is Right* rule, $F_{cv\text{ Interaction}} = 2.525$.

Here are Dr. Larue's three decision rules:

- Main effect of rows:
  - If $F_{\text{Rows}} \geq 3.150$, reject $H_{0\text{ Rows}}$.
  - If $F_{\text{Rows}} < 3.150$, fail to reject $H_{0\text{ Rows}}$.
- Main effect of columns:
  - If $F_{\text{Columns}} \geq 3.150$, reject $H_{0\text{ Columns}}$.
  - If $F_{\text{Columns}} < 3.150$, fail to reject $H_{0\text{ Columns}}$.
- Interaction effect:
  - If $F_{\text{Interaction}} \geq 2.525$, reject $H_{0\text{ Interaction}}$.
  - If $F_{\text{Interaction}} < 2.525$, fail to reject $H_{0\text{ Interaction}}$.

Step 5 **Calculate the Test Statistics.** Here are the sums of squares for Dr. Larue's study, derived from the formulas provided in the chapter appendix:

- $SS_{\text{Between}} = 2{,}117.00$
- $SS_{\text{Rows}} = 1{,}712.00$
- $SS_{\text{Columns}} = 401.00$
- $SS_{\text{Interaction}} = 4.00$
- $SS_{\text{Within}} = 469.00$
- $SS_{\text{Total}} = 2{,}586.00$

Dr. Larue has already calculated degrees of freedom for rows, columns, interaction, and within groups (2, 2, 4, and 63, respectively), so she can now calculate degrees of freedom for between groups and total degrees of freedom as directed by Table 12.9:

$$\begin{aligned} df_{\text{Between}} &= df_{\text{Rows}} + df_{\text{Columns}} + df_{\text{Interaction}} \\ &= 2 + 2 + 4 \\ &= 8 \end{aligned}$$

$$\begin{aligned} df_{\text{Total}} &= N - 1 \\ &= 72 - 1 \\ &= 71 \end{aligned}$$

The next step is the calculation of the four mean squares. To do this, Dr. Larue divides each of the four sums of squares by its degrees of freedom:

$$MS_{\text{Rows}} = \frac{SS_{\text{Rows}}}{df_{\text{Rows}}} = \frac{1{,}712.00}{2} = 856.00$$

$$MS_{\text{Columns}} = \frac{SS_{\text{Columns}}}{df_{\text{Columns}}} = \frac{401.00}{2} = 200.50$$

$$MS_{\text{Interaction}} = \frac{SS_{\text{Interaction}}}{df_{\text{Interaction}}} = \frac{4.00}{4} = 1.00$$

$$MS_{\text{Within}} = \frac{SS_{\text{Within}}}{df_{\text{Within}}} = \frac{469.00}{63} = 7.4444 = 7.44$$

Once the four mean squares have been calculated, all that is left is to find the three F ratios. To do this, Dr. Larue divides each of the mean squares—rows, columns, and interaction—by the mean square within groups:

$$F_{\text{Rows}} = \frac{MS_{\text{Rows}}}{MS_{\text{Within}}} = \frac{856.00}{7.44} = 115.0538 = 115.05$$

$$F_{\text{Columns}} = \frac{MS_{\text{Columns}}}{MS_{\text{Within}}} = \frac{200.50}{7.44} = 26.9489 = 26.95$$

$$F_{\text{Interaction}} = \frac{MS_{\text{Interaction}}}{MS_{\text{Within}}} = \frac{1.00}{7.44} = 0.1344 = 0.13$$

The completed ANOVA summary table is shown in **Table 12.12**. How to interpret the results found in a summary table is the next order of business.

TABLE 12.12 Completed ANOVA Summary Table for Dr. Larue's Relationship Satisfaction Study

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	2,117.00	8		
Rows	1,712.00	2	856.00	115.05
Columns	401.00	2	200.50	26.95
Interaction	4.00	4	1.00	0.13
Within groups	469.00	63	7.44	
Total	2,586.00	71		

Arguing style is the row variable, and perception of parental marriage is the column variable.

Practice Problems 12.2

Apply Your Knowledge

12.04 Read each scenario and decide what statistical test should be used. Select from a single-sample *z* test; single-sample *t* test; independent-samples *t* test; paired-samples *t* test; between-subjects, one-way ANOVA; one-way, repeated-measures ANOVA; and between-subjects, two-way ANOVA.

a. People who are classified as (1) overweight, (2) normal weight, or (3) underweight are randomly assigned to drink either (1) regular soda, (2) diet soda, or (3) water. Thirty minutes later they indicate, on an interval scale, how thirsty they are.

b. Backpacks of elementary school students, middle school students, and high school students are weighed to see if there are differences in how heavy they are.

12.05 List the hypotheses for a between-subjects, two-way ANOVA in which there are four rows and three columns.

12.06 Given $df_{\text{Rows}} = 2$, $df_{\text{Columns}} = 4$, $df_{\text{Interaction}} = 8$, and $df_{\text{Within}} = 165$, list the critical values of *F* for the three *F* ratios for a between-subjects, two-way ANOVA for $\alpha = .05$.

12.07 Given $n = 7$, $R = 2$, $C = 2$, $SS_{\text{Between}} = 650.00$, $SS_{\text{Rows}} = 250.00$, $SS_{\text{Columns}} = 300.00$, $SS_{\text{Interaction}} = 100.00$, $SS_{\text{Within}} = 800.00$, and $SS_{\text{Total}} = 1{,}450.00$, complete an ANOVA summary table for a between-subjects, two-way ANOVA.

12.3 Interpreting a Between-Subjects, Two-Way ANOVA

Interpretation of a statistical test involves stating in plain language what the results mean. The interpretation plan for a two-way ANOVA addresses the same three questions as were addressed for a one-way ANOVA but for more effects: (1) Were the null hypotheses rejected? (2) How large are the effects? (3) Where are the effects and what is their direction?

Let's start with Dr. Ballard's study about the impact of caffeine and sleep deprivation on mental alertness. In this study, 30 college students were randomly assigned to six groups. Before completing a mental alertness task one hour after waking up, half the participants consumed a cup of caffeinated coffee and half didn't. Further, one third of the participants had a full night's sleep, one third were sleep-deprived by one hour, and one third by two hours. The results of the study are shown in Figure 12.5 (see page 432) and **Figure 12.7**, and in **Table 12.13**.

- Figure 12.7 graphs the row means for the main effect of caffeine and the column means for the main effect of sleep deprivation.
- Figure 12.5 uses the cell means to show what appears to be an interaction effect.
- Table 12.13 shows the ANOVA summary table, from which Dr. Ballard will need a number of values as he interprets the results.

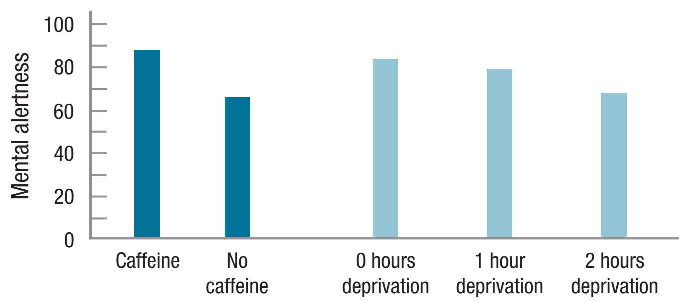


Figure 12.7 Mean Mental Alertness for Row Main Effect (Caffeine Consumption) and Column Main Effect (Sleep Deprivation) The row means and column means appear to show that caffeine improves performance and that sleep deprivation harms it.

TABLE 12.13 ANOVA Summary Table Showing the Effects of Caffeine and Sleep Deprivation on Mental Alertness

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	5,950.00	5		
Rows	3,630.00	1	3,630.00	116.46
Columns	1,340.00	2	670.00	21.50
Interaction	980.00	2	490.00	15.72
Within groups	748.00	24	31.17	
Total	6,698.00	29		

Caffeine consumption is the row variable, and sleep deprivation is the column variable.

Were the Null Hypotheses Rejected?

To determine if any of the three null hypotheses was rejected—one for the row main effect, one for the column main effect, and one for the interaction effect—Dr. Ballard needs the decision rules generated in Step 4 and the F ratios calculated in Step 5. The critical values of F were $F_{cv\text{ Rows}} = 4.260$, $F_{cv\text{ Columns}} = 3.403$, and $F_{cv\text{ Interaction}} = 3.403$. The values of F calculated were $F_{\text{Rows}} = 116.46$, $F_{\text{Columns}} = 21.50$, and $F_{\text{Interaction}} = 15.72$ (see Table 12.13). Applying the three decision rules:

- $116.46 \geq 4.260$, so reject $H_{0\text{ Rows}}$, accept $H_{1\text{ Rows}}$, and call the row effect statistically significant.
- $21.50 \geq 3.403$, so reject $H_{0\text{ Columns}}$, accept $H_{1\text{ Columns}}$, and call the column effect statistically significant.
- $15.72 \geq 3.403$, so reject $H_{0\text{ Interaction}}$, accept $H_{1\text{ Interaction}}$, and call the interaction effect statistically significant.

The next step is to write the results in APA format. APA format for the results of an ANOVA means reporting five pieces of information: (1) stating what test was done (an F test), (2) indicating the numerator and denominator degrees of freedom for the F ratio, (3) reporting the observed value of the test statistic, (4) naming the selected alpha level, and (5) telling whether the observed F fell in the rare zone ($p < .05$, i.e., null hypothesis was rejected) or in the common zone ($p > .05$, the null hypothesis was not rejected).

APA format for the main effect of rows is

$$F(1, 24) = 116.46, p < .05$$

For the main effect of columns, APA format is

$$F(2, 24) = 21.50, p < .05$$

For the interaction effect, APA format is

$$F(2, 24) = 15.72, p < .05$$

After determining the status—statistically significant or not—of each F ratio, a researcher can begin to interpret the results. It is tempting to start the interpretation with one of the main effects displayed in Figure 12.7, but remember—with a two-way ANOVA and a statistically significant interaction effect, the interaction takes precedence.

When the interaction effect is statistically significant, the null hypothesis that each main effect has an independent effect on the dependent variable is rejected. The alternative hypothesis that the two independent variables interact to affect the dependent variable in at least one cell is accepted. As a result, the main effects are less relevant. To explore further what the interaction means, Dr. Ballard will need to use a post-hoc test to compare individual cell means.

Dr. Ballard can't be sure which mean differences are statistically significant until the post-hoc tests are completed, but inspecting Figure 12.5 gives some sense of the interaction. Figure 12.5 suggests that the amount of sleep deprivation has little effect on mental alertness when people are dosed with caffeine. However, for those who don't receive caffeine, alertness declines as sleep deprivation increases.

What about the statistically significant main effects? Does sleep deprivation affect mental alertness? Does caffeine? Both answers are of the "it depends" variety. Look at Figure 12.5. Whether sleep deprivation affects performance depends on whether one has consumed caffeine. And, whether caffeine affects performance depends on whether a person is sleep-deprived. It doesn't look like the main effects will add much to our interpretation.

How Big Are the Effects?

The same measure of effect, eta squared, is used for two-way ANOVA as was used for repeated-measures ANOVA. For a two-way ANOVA, eta squared can be calculated for each main effect and for the interaction. Eta squared, like r^2, calculates the percentage of variability in the dependent variable that is explained by an explanatory variable or by the interaction of the explanatory variables.

The formulas for calculating eta squared for the row main effect, the column main effect, and the interaction effect are given in Equation 12.3.

Equation 12.3 Formulas for Calculating Eta Squared (η^2) for Row Main Effect, Column Main Effect, and Interaction Effect

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

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

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

where η^2_{Rows} = eta squared for the row main effect, the percentage of variability in the dependent variable that is explained by the row explanatory variable

η^2_{Columns} = eta squared for the column main effect, the percentage of variability in the dependent variable that is explained by the column explanatory variable

$\eta^2_{\text{Interaction}}$ = eta squared for the interaction effect, the percentage of variability in the dependent variable that is explained by the interaction between the row explanatory variable and the column explanatory variable

SS_{Rows} = sum of squares rows

SS_{Columns} = sum of squares columns

$SS_{\text{Interaction}}$ = sum of squares interaction

SS_{Total} = sum of squares total

For the caffeine/sleep deprivation study, let's start with η^2 for the statistically significant interaction.

Dr. Ballard calculates eta squared for the interaction effect as follows:

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

$$= \frac{980.00}{6{,}698.00} \times 100$$

$$= 0.1463 \times 100$$

$$= 14.63\%$$

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

$$= \frac{1{,}340.00}{6{,}698.00} \times 100$$

$$= 20.00\%$$

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

$$= \frac{3{,}630.00}{6{,}698.00} \times 100$$

$$= 54.20\%$$

The same standards are used for interpreting eta squared as were used for r^2 for one-way ANOVA:

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

Even though the rows (caffeine) effect is very large at 54% and the column (sleep deprivation) effect at 20% is stronger than the interaction effect, the focus of the interpretation will be on the interaction. The interaction of the two variables, which explains about 15% of the variability in mental alertness, is a medium effect. Most of the variability that sleep deprivation explains is due to the effect on the no caffeine participants. The line for the caffeine-receiving participants in Figure 12.5 is mostly flat, indicating that degree of sleep deprivation explains little of the variability in mental alertness for these subjects. In contrast, the line for the no caffeine participants is on a downward trajectory as sleep deprivation increases, suggesting that amount of sleep deprivation explains a lot of the variability in mental alertness for these subjects. Does the amount of sleep deprivation explain a lot of the variability in mental alertness? It depends. The main effects are trumped by the interaction effect.

Where Are the Effects, and What Is Their Direction?

Just as with the other ANOVA tests, finding where the effects lie for two-way ANOVA involves the use of post-hoc tests. And, just as with the other ANOVAs, post-hoc tests for two-way ANOVA should be used only when the effect is statistically significant.

- If the row main effect is statistically significant, and if there are three or more levels of the row explanatory variable, then a post-hoc test for the row effect can be used to find which pairs of *row* means differ statistically. (If there are only two row means and the row main effect is statistically significant, then the two existing row means must differ statistically.)
- If the column main effect is statistically significant, and if there are three or more levels of the column explanatory variable, then a post-hoc test for the column effect can be used to find which pairs of *column* means differ statistically. (If there are only two column means and the column main effect is statistically significant, then the two existing column means must differ statistically.)
- If the interaction effect was statistically significant, then a post-hoc test for the interaction effect can be used to find which pair(s) of *cell* means differ statistically.

The post-hoc test for the between-subjects, two-way ANOVA is the same one used for other ANOVAs, the Tukey *HSD*. *HSD,* remember, stands for "honestly significant difference." If a pair of means differs by the *HSD* value or more than the *HSD* value, then the difference is a statistically significant one. The formulas for the calculation of *HSD* values are found in Equation 12.4.

Equation 12.4 Formulas for Calculating Tukey *HSD* Values for Post-Hoc Tests for Between-Subjects, Two-Way ANOVA

$$HSD_{\text{Rows}} = q_{\text{Rows}}\sqrt{\frac{MS_{\text{Within}}}{n_{\text{Rows}}}}$$

$$HSD_{\text{Columns}} = q_{\text{Columns}}\sqrt{\frac{MS_{\text{Within}}}{n_{\text{Columns}}}}$$

$$HSD_{\text{Cells}} = q_{\text{Cells}}\sqrt{\frac{MS_{\text{Within}}}{n_{\text{Cells}}}}$$

where HSD_{Rows} = HSD value for the row main effect
q_{Rows} = q value for the row main effect, from Appendix Table 5, where k = number of rows and $df = df_{\text{Within}}$
MS_{Within} = within-groups mean square
n_{Rows} = number of cases in a row
HSD_{Columns} = HSD value for the column main effect
q_{Columns} = q value for the column main effect, from Appendix Table 5, where k = the number of columns and $df = df_{\text{Within}}$
n_{Columns} = number of cases in a column
HSD_{Cells} = HSD value for the interaction effect
q_{Cells} = q value for the interaction effect, from Appendix Table 5, where k = the number of cells and $df = df_{\text{Within}}$
n_{Cells} = number of cases in a cell

For the caffeine/sleep deprivation data, there is little need to do post-hoc tests for the statistically significant main effects as our focus will be on the interaction. Instead, the *HSD* test will be used to interpret the interaction effect only.

The *HSD* to be calculated will be used to compare cell means—any two cell means that differ by the HSD_{Cells} value have a difference that is large enough to represent a statistically significant difference. And statistically significant sample differences provide evidence for population differences.

Here's what one needs to calculate HSD_{Cells}:

- Determine the alpha level, .05 or .01. Typically, the same alpha level as used in the decision rule for the *F* ratio is utilized. For Dr. Ballard's study, this means $\alpha = .05$.
- To find the q_{Cells} value in Appendix Table 5, know that $k = 6$, because there are six cells, and that $df = 24$, because $df_{\text{Within}} = 24$. The intersection of the column for $k = 6$ and the row for $df = 24$ gives $q_{\text{Cells}} = 4.37$.
- From the ANOVA summary table, note that $MS_{\text{Within}} = 31.17$.
- Each cell has five cases, so $n_{\text{Cells}} = 5$.

Here is Equation 12.4, with those values substituted:

$$\begin{aligned} HSD_{\text{Cells}} &= q_{\text{Cells}}\sqrt{\frac{MS_{\text{Within}}}{n_{\text{Cells}}}} \\ &= 4.37\sqrt{\frac{31.17}{5}} \\ &= 4.37\sqrt{6.2340} \\ &= 4.37 \times 2.4968 \\ &= 10.9110 \\ &= 10.91 \end{aligned}$$

The HSD_{Cells} value is 10.91 and any two cell sample means that differ by that much or more have a statistically significant difference. Table 12.14 contains all six cell means and there are 15 possible cell-by-cell comparisons (see Table 12.15). Note, in the points below, how Dr. Ballard approaches the comparisons in an organized fashion and indicates the directions of the difference.

- For participants who consumed no caffeine, each increase in sleep deprivation—from 0 hours ($M = 80.00$) to 1 hour ($M = 68.00$), and from 1 hour to 2 hours ($M = 50.00$)—caused a statistically significant decline in mental alertness.
- Consuming caffeine seems to protect against the negative effects of sleep deprivation as there was no statistically significant change in cell means for the caffeine group. (The means for 0, 1, and 2 hours of sleep deprivation were, respectively, 88.00, 90.00, and 86.00.)
- Looking at differences between the caffeine and no-caffeine groups, there is no evidence that caffeine consumption helped performance if no sleep deprivation occurred (means of 88.00 vs. 80.00). But, it did help performance with 1 hour of sleep deprivation (90.00 vs. 68.00) and 2 hours of sleep deprivation (86.00 vs. 50.00).

TABLE 12.14 Mental Alertness Cell Means for Caffeine/Sleep Deprivation Data

	0-hours Sleep Deprivation	1-hour Sleep Deprivation	2-hours Sleep Deprivation
Caffeine	88.00	90.00	86.00
No Caffeine	80.00	68.00	50.00

Any cell means that differ by at least the HSD_{Cells} value of 10.91 are statistically significantly different.

TABLE 12.15 Possible Comparisons of Cell Means for the Caffeine Consumption/Sleep Deprivation Study

	0-hours Sleep Deprivation	1-hour Sleep Deprivation	2-hours Sleep Deprivation
Caffeine	A	B	C
No Caffeine	D	E	F

Possible Cell-to-Cell Comparisons	
A vs.	B, C, D, E, and F
B vs.	C, D, E, and F
C vs.	D, E, and F
D vs.	E and F
E vs.	F

Putting It All Together

Before writing an interpretation of a two-way ANOVA, it is helpful to review the interaction graph, Figure 12.5. The graph shows two things:

- For those who don't consume caffeine, mental alertness decreases as sleep deprivation moves from 0 hours of deprivation to 1 hour and to 2 hours.
- Consuming caffeine keeps mental alertness from deteriorating, at least with 1 or 2 hours of sleep deprivation.

Here's Dr. Ballard's interpretation in which he addresses the following four points: What was done? What was found? What does it mean? What suggestions exist for future research?

> This study explored the effects of caffeine consumption and sleep deprivation on mental alertness. Using a between-subjects design, 30 college students were assigned to six groups and then had their mental alertness tested. Half received caffeine before testing and half didn't; one third had a full night's sleep, one third were awakened 1 hour early, and one third 2 hours early. There was a statistically significant interaction effect of the two variables on mental alertness $F(2, 24) = 15.72$, $p < .05$ as well as statistically significant main effects for caffeine $F(1, 24) = 116.46$, $p < .05$ and sleep deprivation $F(2, 24) = 21.50$, $p < .05$. In general, caffeine

consumption kept mental alertness elevated and, as sleep deprivation increased, performance deteriorated. However, these two variables did not independently affect mental alertness.

The interaction effect was moderately strong and showed that the impact of sleep deprivation on mental alertness depended on whether one consumed caffeine before testing or not. Further, how caffeine affected mental alertness depended on how sleep-deprived one was. For people who did not consume caffeine, increasing sleep deprivation caused a worsening of mental alertness. In contrast, consuming caffeine kept sleep deprivation from affecting mental alertness. This study suggests that a person can compensate for the mental alertness deficit caused by an hour or two of sleep deprivation by drinking a cup of coffee. It would be wise to replicate this study, to see if the effect is found in a different population. If it is, future research should investigate the effect of different doses of caffeine on different amounts of sleep deprivation.

(By the way, the data in this example were made up. Don't put too much faith in caffeine being an effective antidote to sleep deprivation. Sorry.)

Worked Example 12.3 For practice in interpreting two-way ANOVA, a return to Dr. Larue's study of factors affecting relationship satisfaction is in order. In that study, there were three levels of arguing style—positive, mixed, and negative—crossed with three different perceptions of the quality of one's parents' marriage—good, average, and bad. The mean level of relationship satisfaction for each of the nine conditions, measured on a scale ranging from 5 (very low) to 35 (very high), is shown in Table 12.16. The apparent lack of interaction is shown graphically in Figure 12.6. And, Figure 12.8 displays the main effects for arguing style and parental marital quality.

TABLE 12.16 Cell, Row, and Column Means for Relationship Satisfaction for Different Arguing Styles and Different Perceptions of Parental Marriages

	Parents' Marriage Perceived as "Good"	Parents' Marriage Perceived as "Average"	Parents' Marriage Perceived as "Bad"	
Positive Arguing Style	30.50 (3.07)	28.25 (2.43)	25.00 (2.93)	27.92
Mixed Arguing Style	23.75 (3.33)	21.50 (2.45)	18.50 (3.21)	21.25
Negative Arguing Style	19.00 (2.67)	16.50 (2.07)	12.50 (3.21)	16.00
	24.42	22.08	18.67	

The row and column means are suggestive of main effects for the two grouping variables. However, until the absence of an interaction effect is established and until it is determined that the main effects are statistically significant, they cannot be interpreted.

The three critical values of F were $F_{cv\ \text{Rows}} = 3.150$, $F_{cv\ \text{Columns}} = 3.150$, and $F_{cv\ \text{Interaction}} = 2.525$. The ANOVA summary table, which makes a return appearance in Table 12.17, shows that the observed values of F for the three effects were $F_{\text{Rows}} = 115.05$, $F_{\text{Columns}} = 26.95$, and $F_{\text{Interaction}} = 0.13$.

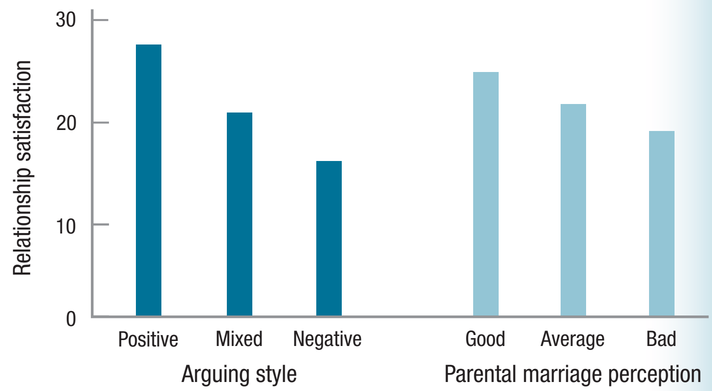


Figure 12.8 Main Effects of Arguing Style and Perception of Parental Marriage on Relationship Satisfaction According to this graph, it looks like there are two main effects: (1) relationship satisfaction decreases as arguing style becomes more negative; (2) relationship satisfaction decreases as perception of quality of parental marriage worsens.

TABLE 12.17 ANOVA Summary Table: Impact of Arguing Style and Perception of Parents' Marriage on Relationship Satisfaction

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	2,117.00	8		
Rows	1,712.00	2	856.00	115.05
Columns	401.00	2	200.50	26.95
Interaction	4.00	4	1.00	0.13
Within groups	469.00	63	7.44	
Total	2,586.00	71		

This is the ANOVA summary table for the results of the relationship satisfaction study. Arguing style is the row variable and perception of parental marriage is the column variable.

Were the null hypotheses rejected?

- Row main effect: $115.05 \geq 3.150$, so reject $H_{0\ \text{Rows}}$ and accept $H_{1\ \text{Rows}}$.
  - In APA format: $F(2, 63) = 115.05$, $p < .05$.
  - The row effect, arguing style, is statistically significant.
  - It is reasonable to conclude, in the larger population, that at least one arguing style differs from at least one other in mean relationship satisfaction.
- Column main effect: $26.95 \geq 3.150$, so reject $H_{0\ \text{Columns}}$ and accept $H_{1\ \text{Columns}}$.
  - In APA format: $F(2, 63) = 26.95$, $p < .05$.
  - The column effect, perception of parental marital quality, is statistically significant.
  - It is reasonable to conclude, in the larger population, that at least one level of perceived marital quality differs from at least one other in mean relationship satisfaction.
- Interaction effect: $0.13 < 2.525$, so fail to reject $H_{0\ \text{Interaction}}$.
  - In APA format: $F(4, 63) = 0.13$, $p > .05$.
  - The interaction effect is not statistically significant.
  - There is not enough evidence to conclude, in the larger population, that the effect of arguing style interacts with the effect of perceived parental marital quality to affect relationship satisfaction.

In the caffeine/sleep deprivation study, the interaction effect was statistically significant and, as a result, the statistically significant main effects were ignored. Now, in the relationship satisfaction study, the two main effects are statistically significant and the interaction is not. How does interpretation work with this set of results?

How big are the effects? Effect size is measured by calculating eta squared (Equation 12.3):

$$\eta^2_{\text{Rows}} = \frac{SS_{\text{Rows}}}{SS_{\text{Total}}} \times 100$$
$$= \frac{1{,}712.00}{2{,}586.00} \times 100$$
$$= 0.6620 \times 100$$
$$= 66.20\%$$

$$\eta^2_{\text{Columns}} = \frac{SS_{\text{Columns}}}{SS_{\text{Total}}} \times 100$$
$$= \frac{401.00}{2586.00} \times 100$$
$$= 0.1551 \times 100$$
$$= 15.51\%$$

$$\eta^2_{\text{Interaction}} = \frac{SS_{\text{Interaction}}}{SS_{\text{Total}}} \times 100$$
$$= \frac{4.00}{2586.00} \times 100$$
$$= 0.0015 \times 100$$
$$= 0.15\%$$

Note that eta squared for the not-statistically-significant interaction effect was calculated. Even though the interaction effect was not significant, it is possible for eta squared to be sizable. If that happened, it would serve to alert a researcher to the possibility of Type II error. In the current situation, with the percentage of variability near zero, there is nothing to make Dr. Larue think she missed finding an interaction effect that really exists. From here on out, it is safe to ignore the interaction effect.

Eta squared for the rows effect was about 66% and for the columns effect it was about 16%. Both main effects have an impact on relationship satisfaction, but one more than the other. The impact of arguing style on relationship satisfaction is quite strong. The impact of perceived parental marital quality is smaller but still meaningful. These results suggest that higher levels of relationship satisfaction are associated more with being a positive arguer than with perceiving one's parents' marriage as good, though having a good model of a marriage is associated meaningfully with relationship satisfaction.

Where are the effects, and what is their direction? Now it is time to use Equation 12.4 to conduct some post-hoc tests and find out what is causing the statistically significant effects. Remember, only conduct a post-hoc test when the effect is statistically significant. With the current example, Dr. Larue has no need to find out what caused the interaction effect because there is no evidence that an interaction effect exists.

To apply Equation 12.4, first find the q value from Appendix Table 5. This depends on α (.05), how many means are being compared, and what the degrees of freedom are. For both the row effect and column effect, there are three means, so $k = 3$ in both instances. Both instances have the same degrees of freedom as well: $df_{\text{Within}} = 63$. Turning to Appendix Table 5, there is a column with $k = 3$, but no row for $df = 63$. In these situations, apply *The Price Is Right* rule and use the df value that is closest to 63 without going over. Here, that is $df = 60$. The q value, at the intersection of $k = 3$ and $df = 60$, is 3.40 for $\alpha = .05$.

To apply Equation 12.4, one also needs to know MS_{Within}, which is 7.44, and how many cases are in a row and a column. Each row contains 24 cases, as does each column. All the values are the same for our calculations for HSD_{Rows} and HSD_{Columns}: $q = 3.40$, $MS_{\text{Within}} = 7.44$, and $n = 24$, so both HSD values can be calculated in one pass:

$$\begin{aligned}
HSD_{\text{Rows}} = HSD_{\text{Columns}} &= q\sqrt{\frac{MS_{\text{Within}}}{n}} \\
&= 3.40\sqrt{\frac{7.44}{24}} \\
&= 3.40\sqrt{.3100} \\
&= 3.40 \times 0.5568 \\
&= 1.8931 \\
&= 1.89
\end{aligned}$$

With an HSD value of 1.89, any two row means or any two column means that differ by that amount or more are statistically significantly different.

For the row main effect, relationship satisfaction grows statistically significantly worse as there is less positive arguing and more negative arguing:

- Positive arguers ($M = 27.92$) have statistically significantly more relationship satisfaction than do mixed arguers ($M = 21.25$). The difference, 6.67, is greater than the HSD value of 1.89.
- Mixed arguers ($M = 21.25$) have statistically significantly more relationship satisfaction than do negative arguers ($M = 16.00$). The difference, 5.25. is greater than the HSD value.

For the column main effect, relationship satisfaction gets statistically significantly worse as the perception of the parents' marriage worsens:

- Students who rated their parents' marriage as good ($M = 24.42$) rated their own relationship as statistically significantly more satisfying than those who rated their parents' marriage as average ($M = 22.08$). The difference, 2.34, is greater than the HSD value of 1.89.
- Rating one's parents' marriage as average ($M = 22.08$) is associated with a statistically significantly higher score on relationship satisfaction than rating it as bad ($M = 18.67$). The difference, 3.41, is greater than the HSD value.

- *Putting it all together.* Before writing her four-point interpretation, Dr. Larue reviewed the graph displaying the main effects (Figure 12.8), so she had a clear picture of the results in her mind. Here's what she wrote (notice, this is a quasi-experimental study in which nothing is manipulated, so she avoids cause-and-effect language to describe the results):

In this social psychology study, the abilities of two variables, arguing style and perceived quality of parental marriage, to predict relationship satisfaction in college students were examined. Students were classified into three categories of arguing style (positive, negative, or mixed) and with three different perceptions of their parents' marriages (good, average, or bad). Eight students from each of these possible combinations who were in current relationships were randomly selected and completed a survey measuring degree of satisfaction with their current relationship.

Using a between-subjects, two-way ANOVA, both arguing style and parental marital perception had statistically significant effects on relationship satisfaction [respectively, $F(2, 63) = 115.05$, $p < .05$, and $F(2, 63) = 26.95$, $p < .05$]. There was no interaction effect for the two variables $F(4, 63) = 0.13$, $p > .05$.

Arguing style was the stronger predictor of relationship satisfaction. As students' arguing style moved from positive to negative, there was a decrease in relationship satisfaction. The role played by perceived parental marriage, though less powerful, was still meaningful. A more negative view of parents' marriages was associated with lower relationship satisfaction.

This study suggests that the relationships one sees as a child influence one's future relationship satisfaction. The good news is that a more powerful influence on relationship satisfaction is arguing style and a positive arguing style is a skill that can be learned. Future research should examine whether teaching positive arguing skills improves relationship satisfaction.

Practice Problems 12.3

12.08 Given $\alpha = .05$, $df_{\text{Rows}} = 3$, $df_{\text{Columns}} = 2$, $df_{\text{Interaction}} = 6$, $df_{\text{Within}} = 60$, $F_{\text{Rows}} = 3.25$, $F_{\text{Columns}} = 1.22$, and $F_{\text{Interaction}} = 0.83$, (a) write each result for this between-subjects, two-way ANOVA in APA format and (b) for each result report whether the effect is statistically significant.

12.09 Given this ANOVA summary table, (a) calculate η^2 for each effect and (b) classify each effect as small, medium, or large. Use $\alpha = .05$.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	5,357.00	15		
Rows	3,725.00	3	1,241.67	37.63
Columns	1,312.00	3	437.33	13.25
Interaction	320.00	9	35.56	1.08
Within groups	7,392.00	224	33.00	
Total	12,749.00	239		

12.10 Here is a table of cell means for a 3 × 2 between-subjects, two-way ANOVA with seven cases in each cell. Note that the row means and column means have been calculated.

	Column 1	Column 2	Column 3	
Row 1	26.00	30.00	34.00	30.00
Row 2	18.00	22.00	26.00	22.00
	22.00	26.00	30.00	

Here is the between-subjects, two-way ANOVA summary table for these data:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	1,120.00	5	224.00	
Rows	672.00	1	672.00	35.99
Columns	448.00	2	224.00	12.00
Interaction	0.00	2	0.00	0.00
Within groups	672.00	36	18.67	
Total	1,792.00	41		

As appropriate, calculate *HSD* values and comment on the direction of the differences for the effects.

12.11 A kinesiologist wanted to investigate the effect of temperature and humidity on human performance. He found 28 college students and randomly assigned them to four different conditions, during which they were to walk at their normal pace on a treadmill for 60 minutes. He measured how far, in miles, they walked. The conditions varied in temperature and humidity: (1) normal temperature and normal humidity; (2) normal temperature and high humidity; (3) high temperature and normal humidity; (4) high temperature and high humidity. The results looked like this:

	Normal Humidity	High Humidity	
Normal Temperature	3.00 miles	2.80 miles	2.90
High Temperature	2.80 miles	2.00 miles	2.40
	2.90	2.40	

Here is the ANOVA summary table:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	4.13	3		
Rows	1.75	1	1.75	25.00
Columns	1.75	1	1.75	25.00
Interaction	0.63	1	0.63	9.00
Within groups	1.58	24	0.07	
Total	5.71	27		

The critical value of *F* for each effect is 4.260. Eta squared for the row, column, and interaction effects, respectively, are 30.65%, 30.65%, and 11.03%. The *HSD* value for comparing cells is 0.39. Given all this information, write a four-point interpretation.

Application Demonstration

The two examples offered in this chapter so far—Dr. Ballard's caffeine/sleep deprivation study and Dr. Larue's relationship satisfaction study—have both involved statistically significant results. Unfortunately, research doesn't always turn out that way. Here's an example of a two-way ANOVA study in which not one of the three null hypotheses is rejected. How are results interpreted in this situation?

Imagine a sensory psychologist, Dr. Porter, who wanted to explore the threshold for perceiving low-frequency sounds. She tested six men and six women to see how low a sound they could perceive. Half the participants were tested in their left ears and half in their right ears. Sounds were measured in hertz (Hz), and the lower the hertz the better a person's hearing. (Want to see how well you do? Search for "Ultimate Sound Test [10000 Hz–1 Hz]" on YouTube.)

Table 12.18 and Figure 12.9 show the results. There appears to be no interaction and slightly better performance exists (a) for women and (b) for the left ear. Whether the effects are statistically significant or can be explained by sampling error remains to be seen.

TABLE 12.18 Cell, Row, and Column Means for Lowest Hertz Tone Perceived

	Left Ear	Right Ear	
Men	140.00	150.00	145.00
Women	120.00	130.00	125.00
	130.00	140.00	

The lower the mean, the better the hearing. Each cell contains three cases. The row means suggest a small advantage in hearing for women over men. The column means suggest a small advantage for the left ear over the right ear. The lack of interaction between the two variables can be seen in Figure 12.9.

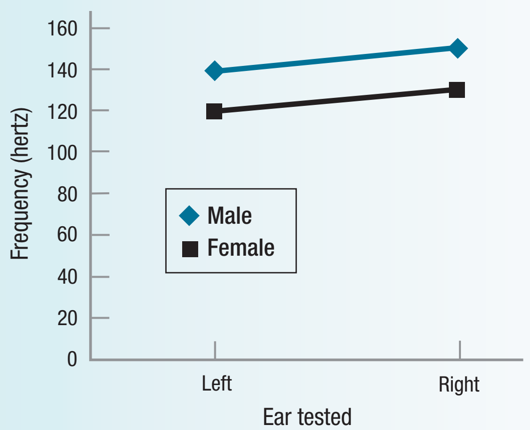


Figure 12.9 Impact of Sex and Ear Tested on Lowest Frequency Perceived
With the lines being parallel, there is no interaction between sex and ear tested. This graph also shows a slightly better performance by women and for the left ear.

The appropriate statistical test to compare the four means from these two independent variables is a two-way ANOVA. The four groups are all independent samples, so the test is a between-subjects, two-way ANOVA, specifically a 2×2 ANOVA.

No assumptions were violated and the ANOVA summary table is presented in Table 12.19. The critical value of *F* for all three effects was 5.318 and no effect was statistically significant.

- There is not enough evidence to conclude that men and women differ in their ability to hear low-frequency sounds.
- There is not enough evidence to conclude that the right ear differs from the left ear in its ability to hear low-frequency sounds.
- There is not enough evidence to conclude that a person's sex and which ear is tested interact to influence the perception of low-frequency sounds.

TABLE 12.19 Completed ANOVA Summary Table for Low-Frequency Perception Data

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	1,500.00	3		
Rows	1,200.00	1	1,200.00	1.30
Columns	300.00	1	300.00	0.32
Interaction	0.00	1	0.00	0.00
Within groups	7,400.00	8	925.00	
Total	8,900.00	11		

Sex is the row variable, and ear (left vs. right) is the column variable.

With no statistically significant effect, no need exists to do any post-hoc testing. A researcher would be wise, however, to calculate eta squared in order to consider the possibility of a Type II error, the possibility of having failed to find an effect that does occur.

Applying Equation 12.3, Dr. Porter finds

$$\begin{aligned}\eta^2_{\text{Rows}} &= \frac{SS_{\text{Rows}}}{SS_{\text{Total}}} \times 100 \\ &= \frac{1{,}200.00}{8{,}900.00} \times 100 \\ &= 0.1348 \times 100 \\ &= 13.48\%\end{aligned}$$

$$\begin{aligned}\eta^2_{\text{Columns}} &= \frac{SS_{\text{Columns}}}{SS_{\text{Total}}} \times 100 \\ &= \frac{300.00}{8{,}900.00} \times 100 \\ &= 0.0337 \times 100 \\ &= 3.37\%\end{aligned}$$

$$\begin{aligned}\eta^2_{\text{Interaction}} &= \frac{SS_{\text{Interaction}}}{SS_{\text{Total}}} \times 100 \\ &= \frac{0.00}{8{,}900.00} \times 100 \\ &= 0.0000 \times 100 \\ &= 0.00\%\end{aligned}$$

These results show that the row main effect, $\eta^2 = 13.48\%$, is medium. The column main effect, $\eta^2 = 3.37\%$, is small. There is no interaction effect. Not enough evidence exists to say that the row effect of sex—male vs. female—affects the perception of low frequencies, but there is enough of a hint here that Dr. Porter might want to draw attention to it. As always, looking at a picture of the effects, like that shown in Figure 12.9, helps to clarify what the effects were. Here's her four-point interpretation:

A sensory psychologist conducted a study testing the ears (left vs. right) of both men and women to see if one ear or one sex had a lower threshold for low-frequency sounds. According to this study, there was not enough evidence to conclude that either one sex or one ear was better than the other at perceiving low-frequency sounds. However, the results suggested that women ($M = 125.00$ Hz) may have an edge over men ($M = 145.00$ Hz) in perceiving low-frequency sounds. To investigate this, it would be advisable to replicate the study with a larger sample size.

SUMMARY

Describe what two-way ANOVA does.

- Two-way ANOVA measures the impact of two explanatory variables on a dependent variable. It divides the variability in the dependent variable into that due to each explanatory variable separately (the main effects) *and* that due to them together (the interaction effect).
- The interaction effect is an advantage of two-way ANOVA over one-way ANOVA. An interaction effect means that the impact of one explanatory variable depends on the level of the other explanatory variable. When there is a statistically significant interaction effect, the main effects are not interpreted.

Complete a between-subjects, two-way ANOVA.

- Use between-subjects, two-way ANOVA when there are two explanatory variables, each with at least two levels, the samples are independent, and the dependent variable is measured at the interval or ratio level.
- Conducting a two-way ANOVA follows the same procedure as is used for other hypothesis tests. That is, check the assumptions (random samples, independence of observations, normality, and homogeneity of variance); form null and alternative hypotheses for each main effect and for the interaction; calculate degrees of freedom; find the critical value of F; set the decision rule; and, finally, find the value of the test statistic by completing an ANOVA summary table.

Interpret a between-subjects, two-way ANOVA.

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

DIY

Two-way ANOVA allows us to examine the impact of two explanatory variables at once. For example, we can look at two sporting events, like swimming, that are structured the same for men and women. Table 12.20 displays the mean times, in seconds, for the top five finishers in the 100-meter freestyle and 100-meter backstroke at the 2012 Summer Olympics.

TABLE 12.20 Mean Times for Top 5 Finishers in Men's and Women's 100-meter Freestyle and Backstroke at the 2012 Summer Olympics

	Freestyle	Backstroke	
Men	48	53	50.5
Women	53	58	55.5
	50.5	55.5	

The graph of this table, Figure 12.10, shows no interaction but the main effects of sex and stroke.

Figure 12.10 is a graph of the results. What does the graph suggest about the "main effects" of sex and type of stroke as well as the interaction of these two variables? From the graph, it is clear that there is no interaction between sex and stroke and that there are probably main effects of sex and stroke. The main effect of sex, if it is statistically significant, says that no matter the stroke, men swim about five seconds faster over a 100-meter distance. If the main effect of stroke is statistically significant, it would be interpreted as saying that both men and women are about five seconds faster when they cover 100 meters by freestyle rather than backstroke.

Note that the lack of an interaction means that these effects are independent and cumulative. Men are five seconds faster than women and freestyle is five seconds faster than backstroke. Thus, men swimming the freestyle are ten seconds faster than women doing the backstroke.

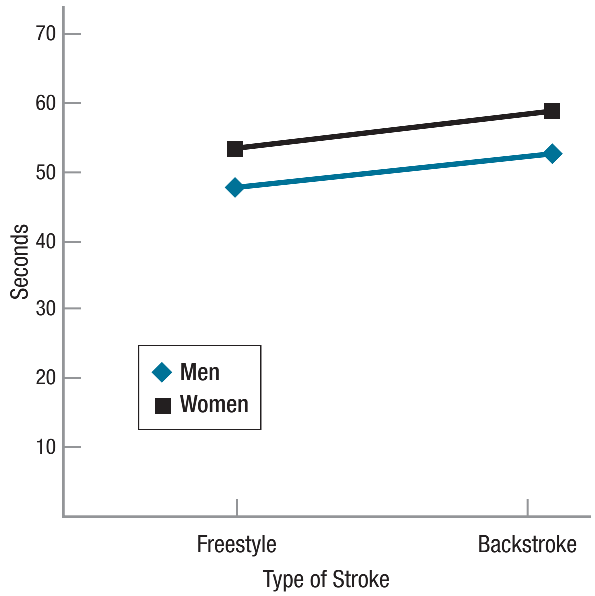


Figure 12.10 2012 Summer Olympics: Average Times of Top 5 Finishers This figure shows the effects of sex and stroke on how long it takes to swim 100 yards.

Now it's your turn. Find some data that vary on two dimensions. Need ideas? See the bulleted examples below. Put the data in a table like the one above. Be sure to label the rows and columns and to calculate marginal values. Graph the results and interpret the graph.

- Sporting events that are structured the same for men and women, at the Olympic, national, collegiate, or high school level
- Sporting events compared across different levels (e.g., high school vs. college)
- Average SAT scores, math vs. reading/writing, for different colleges
- Average salaries, men vs. women, for different professions

KEY TERMS

between-subjects – ANOVA terminology for independent samples.

crossed – a factorial ANOVA in which each level of each explanatory variable occurs with each level of the other explanatory variable.

factorial ANOVA – an analysis of variance in which there is more than one explanatory variable.

interaction effect – situation, in factorial ANOVA, in which the impact of one explanatory variable on the dependent variable depends on the level of another explanatory variable.

main effect – the impact of an explanatory variable, by itself, on the dependent variable.

within-subjects – ANOVA terminology for dependent samples.

CHAPTER EXERCISES

Review Your Knowledge

12.01 ____-way ANOVA examines the impact of two explanatory variables at the same time.

12.02 A 2 × 3 ANOVA could also be called a ____ × ____ ANOVA.

12.03 "Between subjects" means ____ samples and "____ subjects" means dependent samples.

12.04 Two-way ANOVA, three-way ANOVA, and four-way ANOVA are all examples of ____ ANOVA.

12.05 There is an advantage in performing one ____-way ANOVA over two ____-way ANOVAs.

12.06 If every level of one explanatory variable occurs with every level of the other explanatory variable, then the two explanatory variables are said to be ____.

12.07 A two-way ANOVA has two ____ effects and one ____ effect.

12.08 An interaction occurs when the effect of one explanatory variable on the dependent variable ____ on the level of the other explanatory variable.

12.09 If the lines in a graph are ____, then an interaction exists.

12.10 If each cell has the same sample size, the cell means in a row can be averaged to find the ____.

12.11 Row means and column means give information about ____ effects.

12.12 If the interaction effect is statistically significant, ____ interpret the main effects.

12.13 In a between-subjects, two-way ANOVA, the groups are ____ samples.

12.14 The random samples assumption says that the sample is a ____ from the ____ to which one wishes to generalize the results.

12.15 If cases in a cell influence each other's scores on the dependent variable, then the ____ assumption is violated.

12.16 The ____ assumption is tested by looking at cell standard deviations.

12.17 In a two-way ANOVA, there are ____ sets of hypotheses.

12.18 The row's null hypothesis states that the ____ for all levels of the row variable are equal.

12.19 The alternative hypothesis for the columns says that ____ one of the population column means differs from ____ one of the other population column means.

12.20 The alternative hypothesis for the interaction effect says that there is ____ for at least one cell.

12.21 Decision rules exist for the columns main effect, the rows main effect, and the ____.

12.22 The decision rule for the interaction effect involves comparing $F_{\text{Interaction}}$ to ____.

12.23 The number of degrees of freedom for the row effect depends on ____.

12.24 The denominator degrees of freedom for all the *F* ratios calculated for a between-subjects, two-way ANOVA is degrees of freedom ____.

12.25 Finding F_{cv} in Appendix Table 4 depends on knowing the numerator ____ and denominator ____ for the *F* ratio.

12.26 If $F_{\text{Interaction}} < F_{cv\ \text{Interaction}}$, then the null hypothesis is ____.

12.27 Total variance in between-subjects, two-way ANOVA is divided into two factors: ____ and ____.

12.28 In between-subjects, two-way ANOVA, between-group variability is divided into variability due to ____, ____, and ____.

12.29 *SS* is the abbreviation for a sum of ____.

12.30 To calculate a mean square, one divides a ____ by its ____.

12.31 $MS_{\text{Interaction}}$ divided by MS_{Within} calculates ____.

12.32 When an ANOVA summary table for a between-subjects, two-way ANOVA is completed, there are ____ *F* ratios.

12.33 If a null hypothesis is rejected, the result is called ____.

12.34 If results, in APA format, are written as $p > .05$, the null hypothesis was ____.

12.35 $p < .05$ means that alpha was set at ____.

12.36 Only interpret a main effect if ____.

12.37 ____ is the effect size used for between-subjects, two-way ANOVA.

12.38 Eta squared is calculated for both ____ effects and for the ____.

12.39 An effect of ≈ ____% is considered a medium effect.

12.40 Eta squared can alert one to the risk of ____ error if the effect was not statistically significant.

12.41 Post-hoc tests should only be used when an effect is ____.

12.42 If the row main effect is statistically significant and there are only two rows, it *is/is not* necessary to do a post-hoc test for the row effect.

12.43 If two cells differ by exactly the *HSD* amount, then the difference is ____.

12.44 The *HSD* value for the interaction effect is used to compare ____ means.

Apply Your Knowledge

Select the right test for the scenario from among a single-sample z test; single-sample t test; independent-samples t test; paired-samples t test; between-subjects, one-way ANOVA; one-way, repeated-measures ANOVA; and between-subjects, two-way ANOVA.

12.45 A dentist randomly assigns people to use one of three different forms of dental hygiene: (1) dental floss, (2) wooden toothpicks, or (3) anti-plaque rinse. After six months he measures, in grams, how much plaque he scrapes off the teeth. What statistical test should he use to see if form of dental hygiene has an impact on the amount of plaque?

12.46 The American Psychological Association wanted to determine if there was a difference in post-graduation success between BA and BS degrees in psychology. It put together a random sample of students with each degree and found out what the annual salary was for their first job after graduation. What statistical test should be used to analyze these data for the effect of type of degree?

12.47 Male and female athletes exercised for an hour on a treadmill. During this time, half of each sex was assigned to drink water and half was assigned to drink a sport beverage. At the end of the hour, lactic acid levels were measured and mean levels were calculated. What statistical test should be used to see how sex and type of beverage affect lactic acid production?

12.48 To study accommodation to the loss of an eye, a sensory psychologist obtained 10 volunteers who agreed to wear an eye patch over one eye for 4 weeks. To test visual ability, the psychologist used a batting test: a pitching machine lobbed 50 pitches to each participant and the psychologist counted how many were hit. This test was conducted (a) before the eye patch was worn, (b) immediately after the eye patch was first put on, (c) 1 week into the study, (d) 2 weeks into the study, (e) 3 weeks into the study, (f) on the last day wearing the eye patch, and (g) 1 week later. What test should be used to see if batting ability changed over the course of the study?

Calculating row and column means

12.49 There are 10 participants in each cell. Each cell is an independent sample. The design is fully crossed. The value in each cell is the cell mean. Calculate (a) row means and (b) column means for this matrix.

	Column 1	Column 2	Column 3
Row 1	25	35	45
Row 2	25	15	5

12.50 There are 17 participants in each cell. Each cell is an independent sample. The design is fully crossed. The value in each cell is the cell mean. Calculate (a) row means and (b) column means for this matrix.

	Column 1	Column 2	Column 3
Row 1	100	110	140
Row 2	140	200	180

Speculating about main effects and interactions

12.51 Given the cell means, row means, and column means below, (a) graph the results to examine if there is an interaction; (b) speculate about which effects would be statistically significant; (c) indicate which effects should be interpreted.

	Column 1	Column 2	
Row 1	17	13	15.00
Row 2	16	12	14.00
	16.50	12.50	

12.52 Given the cell means, row means, and column means below, (a) graph the results to determine if you think there is an interaction; (b) speculate about which effects would be statistically significant; (c) indicate which effects should be interpreted.

	Column 1	Column 2	
Row 1	80	60	70.00
Row 2	60	40	50.00
	70.00	50.00	

Checking the assumptions

12.53 Eighty volunteers who read about a study in a newspaper were randomly assigned to be in the experimental or control group. Within each group, participants were randomly assigned to one of four different conditions. Each participant then participated in the study individually. After this, each participant's status on the interval-level psychological variable was measured. Below are the means (and standard deviations). (a) Evaluate all four assumptions and (b) decide if it is OK to proceed with the test.

	I	II	III	IV
Control	12.34 (2.89)	13.76 (3.89)	14.17 (4.95)	16.88 (6.91)
Experimental	14.88 (3.12)	15.66 (4.18)	16.99 (5.54)	17.33 (6.48)

12.54 A researcher was interested in the impact of different types of day care on children's development. She obtained samples of children who (a) had been watched by a babysitter at home; (b) had attended small, home-based day care; or (c) had gone to a large day-care center. She also classified the children as having received paid care for (1) less than 40 hours a week or (2) 40 or more hours per week. There were eight children in each cell and no siblings were included. She measured each child's adjustment level, on an interval scale, in the first grade. Below are the means (and standard deviations). (a) Evaluate all four assumptions and (b) decide if it is OK to proceed with the test.

	Babysitter	Small Day Care	Large Day Care
<40 Hours	112.66 (14.83)	108.28 (16.76)	106.80 (17.65)
≥40 Hours	114.88 (13.12)	108.71 (15.66)	109.28 (16.77)

Listing hypotheses

12.55 List all the hypotheses for the ANOVA described in Exercise 12.53 for (a) rows, (b) columns, and (c) interaction.

12.56 List all the hypotheses for the ANOVA described in Exercise 12.54 for (a) rows, (b) columns, and (c) interaction.

Calculating degrees of freedom

12.57 If $R = 3$, $C = 4$, and $n = 8$, calculate the degrees of freedom for the following effects: (a) rows, (b) columns, (c) interaction, (d) within groups, (e) between groups, and (f) total.

12.58 If $R = 2$, $C = 3$, and $n = 11$, calculate the degrees of freedom for the following effects: (a) rows, (b) columns, (c) interaction, (d) within groups, (e) between groups, and (f) total.

Finding F_{cv}

12.59 If $\alpha = .05$, $df_{\text{Rows}} = 2$, $df_{\text{Columns}} = 2$, $df_{\text{Interaction}} = 4$, and $df_{\text{Within}} = 36$, find (a) $F_{cv\ \text{Rows}}$, (b) $F_{cv\ \text{Columns}}$, and (c) $F_{cv\ \text{Interaction}}$.

12.60 If $\alpha = .05$, $df_{\text{Rows}} = 3$, $df_{\text{Columns}} = 1$, $df_{\text{Interaction}} = 3$, and $df_{\text{Within}} = 40$, find (a) $F_{cv\ \text{Rows}}$, (b) $F_{cv\ \text{Columns}}$, and (c) $F_{cv\ \text{Interaction}}$.

Writing the decision rule

12.61 If $F_{cv\ \text{Interaction}}$ is 2.668, what is the decision rule for the interaction effect?

12.62 If $F_{cv\ \text{Rows}}$ is 3.295, what is the decision rule for the rows effect?

Calculating mean squares

12.63 If $SS_{\text{Within}} = 168.48$ and $df_{\text{Within}} = 40$, calculate MS_{Within}.

12.64 If $SS_{\text{Interaction}} = 0.60$ and $df_{\text{Interaction}} = 4$, calculate $MS_{\text{Interaction}}$.

Calculating F ratios

12.65 If $MS_{\text{Rows}} = 17.30$ and $MS_{\text{Within}} = 3.33$, what is F_{Rows}?

12.66 If $MS_{\text{Columns}} = 66.54$ and $MS_{\text{Within}} = 78.88$, what is F_{Columns}?

Completing a summary table

12.67 Given $R = 2$, $C = 2$, $n = 5$, $SS_{\text{Between}} = 61.00$, $SS_{\text{Rows}} = 15.00$, $SS_{\text{Columns}} = 12.00$, $SS_{\text{Interaction}} = 34.00$, $SS_{\text{Within}} = 120.00$, and $SS_{\text{Total}} = 181.00$, complete the ANOVA summary table for a between-subjects, two-way ANOVA.

12.68 Given $R = 2$, $C = 4$, $n = 12$, $SS_{\text{Between}} = 225.00$, $SS_{\text{Rows}} = 78.00$, $SS_{\text{Columns}} = 23.00$, $SS_{\text{Interaction}} = 124.00$, $SS_{\text{Within}} = 350.00$, and $SS_{\text{Total}} = 575.00$, complete the ANOVA summary table for a between-subjects, two-way ANOVA.

Applying the decision rule

12.69 If $F_{cv} = 3.443$ and $F = 17.55$, was the null hypothesis rejected?

12.70 If $F_{cv} = 3.191$ and $F = 2.86$, was the null hypothesis rejected?

Writing results in APA format

12.71 If $F_{\text{Interaction}} = 3.70$, $df_{\text{Interaction}} = 2$, $df_{\text{Within}} = 66$, and $F_{cv\ \text{Interaction}} = 3.138$, write the results in APA format. Use $\alpha = .05$.

12.72 If $F_{\text{Rows}} = 2.87$, $df_{\text{Rows}} = 2$, $df_{\text{Within}} = 171$, and $F_{cv\ \text{Rows}} = 3.053$, write the results in APA format. Use $\alpha = .05$.

12.73 If $F_{\text{Columns}} = 2.45$, $df_{\text{Columns}} = 3$, and $df_{\text{Within}} = 168$, write the results in APA format. Use $\alpha = .05$.

12.74 If $F_{\text{Columns}} = 4.00$, $df_{\text{Columns}} = 1$, and $df_{\text{Within}} = 36$, write the results in APA format. Use $\alpha = .05$.

12.75 Given the ANOVA summary table below, write the results in APA format for (a) the rows effect, (b) the columns effect, and (c) the interaction effect. Use $\alpha = .05$.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	108.56	5		
Rows	45.40	2	22.70	3.21
Columns	7.62	1	7.62	1.08
Interaction	55.54	2	27.77	3.93
Within groups	212.22	30	7.07	
Total	320.78	35		

12.76 Given the ANOVA summary table below, write the results in APA format for (a) the rows effect, (b) the columns effect, and (c) the interaction effect. Use $\alpha = .05$.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	70.40	7		
Rows	32.34	3	10.78	2.77
Columns	7.62	1	7.62	1.96
Interaction	30.44	3	10.15	2.61
Within groups	342.66	88	3.89	
Total	413.06	95		

Deciding which effects to interpret

12.77 If the null hypothesis is rejected for all three effects (rows, columns, and interaction), which effects should be interpreted?

12.78 If the null hypothesis is rejected for the rows effect and the columns effect but not for the interaction effect, which effects should be interpreted?

12.79 If the null hypothesis is rejected for the interaction effect but not for the rows effect or the columns effect which effects should be interpreted?

12.80 If the null hypothesis is rejected for the rows effect and the interaction effect but not for the columns effect, which effects should be interpreted?

Calculating eta squared

12.81 If $SS_{Rows} = 3.78$ and $SS_{Total} = 47.83$, (a) calculate η^2_{Rows} and (b) classify the effect as small, medium, or large.

12.82 If $SS_{Columns} = 17.91$ and $SS_{Total} = 783.54$, (a) calculate $\eta^2_{Columns}$ and (b) classify the effect as small, medium, or large.

12.83 Given the ANOVA summary table below, calculate η^2 for (a) the rows effect, (b) the columns effect, and (c) the interaction effect.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	132.00	7		
Rows	33.00	3	11.00	4.56
Columns	54.00	1	54.00	22.41
Interaction	45.00	3	15.00	6.22
Within groups	212.00	88	2.41	
Total	344.00	95		

12.84 Given the ANOVA summary table below, calculate η^2 for (a) the rows effect, (b) the columns effect, and (c) the interaction effect.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	125.00	5		
Rows	45.00	2	22.50	18.29
Columns	68.00	1	68.00	55.28
Interaction	12.00	2	6.00	4.88
Within groups	140.00	114	1.23	
Total	265.00	119		

Finding q

12.85 Given $R = 3$, $C = 2$, and $n = 8$, find (a) q_{Rows}, (b) $q_{Columns}$, and (c) q_{Cells}.

12.86 Given $R = 3$, $C = 3$, and $n = 11$, find (a) q_{Rows}, (b) $q_{Columns}$, and (c) q_{Cells}.

Calculating HSD

12.87 If $q_{Rows} = 3.49$, $MS_{Within} = 4.56$, and $n_{Rows} = 12$, what is HSD_{Rows}?

12.88 If $q_{Cells} = 2.96$, $MS_{Within} = 12.75$, and $n_{Cells} = 4$, what is HSD_{Cells}?

12.89 If $\alpha = .05$, for which effects should an *HSD* value be calculated?

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	324.00	8		
Rows	70.00	2	35.00	28.46
Columns	88.00	2	44.00	35.77
Interaction	166.00	4	41.50	33.74
Within groups	100.00	81	1.23	
Total	424.00	89		

12.90 If $\alpha = .05$, for which effects should an *HSD* value be calculated?

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	118.00	8		
Rows	88.00	2	44.00	9.91
Columns	23.00	2	11.50	2.59
Interaction	7.00	4	1.75	0.39
Within groups	600.00	135	4.44	
Total	718.00	143		

Interpreting HSD

12.91 If $M_{\text{Row 1}} = 117.66$, $M_{\text{Row 2}} = 113.63$, $M_{\text{Row 3}} = 128.91$, and $HSD_{\text{Rows}} = 5.89$, determine (a) which rows have a statistically significant difference and (b) the direction of the differences.

12.92 If $M_{\text{Cell 1}} = 55.54$, $M_{\text{Cell 2}} = 48.34$, $M_{\text{Cell 3}} = 36.44$, $M_{\text{Cell 4}} = 59.40$, and $HSD_{\text{Cells}} = 7.83$, determine (a) which cells have a statistically significant difference and (b) the direction of the differences.

Interpreting results

12.93 A consumer psychologist classified shoppers at a grocery store as (a) being males or females, and (b) shopping with or without children. These two variables were crossed and he took a random sample of five shoppers from each of the four cells. Then he gave them the Enjoyment of Shopping Experience Scale (ESES) to be completed for the current shopping experience. The ESES is an interval-level measure, with a mean of 50 indicating neutral feelings about shopping. Scores can range from 20 to 80, with scores above 50 an indication of enjoying shopping; scores below 50 indicate that shopping is not enjoyable. Here are the means:

	Shopping Without Children	Shopping With Children	Row Means
Male Shopper	43.20	56.80	50.00
Female Shopper	57.40	44.80	51.10
Column Means	50.30	50.80	

No nonrobust assumptions were violated and a between-subjects, two-way ANOVA was completed. Here is the ANOVA summary table:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	865.35	3		
Rows	6.05	1	6.05	0.29
Columns	1.25	1	1.25	0.06
Interaction	858.05	1	858.05	41.15
Within groups	333.60	16	20.85	
Total	1,198.95	19		

The researcher calculated $\eta^2_{\text{Rows}} = 0.50\%$, $\eta^2_{\text{Columns}} = 0.10\%$, $\eta^2_{\text{Interaction}} = 71.57\%$, and $HSD_{\text{Cells}} = 8.27$. Complete a four-point interpretation for the results. (*Hint:* Graphing the results will help.)

12.94 A cognitive psychologist decided to investigate whether two beliefs about healthy living had any real impact on performance in college. She took 40, first-semester college student volunteers and randomly assigned half to eat breakfast every day and the other half to skip breakfast every day. This was crossed with another variable. Half the students were randomly assigned to sleep at least 8 hours a night and half were randomly assigned to sleep less than 8 hours a night. At the end of the semester, she recorded each participant's GPA. The results are shown in this table:

	Sleep ≥8 Hours	Sleep <8 Hours	Row Means
Eat Breakfast	3.20	2.72	2.96
Skip Breakfast	3.08	2.77	2.93
Column Means	3.14	2.75	

No nonrobust assumptions for a between-subjects, two-way ANOVA were violated and the ANOVA was completed. Here is the summary table:

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	1.64	3		
Rows	0.01	1	0.01	0.06
Columns	1.56	1	1.56	9.18
Interaction	0.07	1	0.07	0.41
Within groups	6.03	36	0.17	
Total	7.67	39		

The eta squared values for rows, columns, and interaction are, respectively, 0.13%, 20.34%, and 0.91%. HSD_{Cells} is 0.50. Interpret the results. (*Hint:* Graphing the results will help.)

12.95 A developmental psychologist investigated the influence of two crossed variables—exposure to televised violence and type of parental discipline—on teens' acceptance of violence. He obtained a random sample of seniors at the local high school and classified them on two dimensions: (1) whether or not their parents had restricted their television viewing and (2) whether their parents used positive reinforcement or punishment. He took a random sample of six teens from each of the four samples (no siblings were included) and his dependent variable was each teen's score on the interval-level Approval of Violence Scale (AVS). Scores on the AVS range from 0 to 20, with higher scores indicating a greater acceptance of violence as a way to solve disagreements. The cell means (and standard deviations) are shown below. SS_{Rows} was calculated to be 210.04, SS_{Columns} as 222.04, $SS_{\text{Interaction}}$ as 15.04, and SS_{Within} as 204.83.

	Restricted TV	Unrestricted TV
Positive Reinforcement	6.00 (2.10)	10.50 (3.62)
Punishment	10.33 (4.08)	18.00 (2.61)

12.96 An industrial organizational psychologist investigated the crossed effects of two variables—being a member of a team sport in high school and extroversion level—on how well professors are liked by their colleagues. The dependent variable was the interval-level Colleague Collegiality Scale (CCS). Scores on the CCS range from 0 to 24; higher scores indicate greater liking by colleagues. From a national and representative sample of college professors, the psychologist randomly selected professors until there were eight in each cell. The table below shows the cell means (and standard deviations). SS_{Rows} was 36.13, $SS_{\text{Columns}} = 1.13$, $SS_{\text{Interaction}} = 36.13$, and $SS_{\text{Within}} = 217.50$.

	Extroverted	Introverted
Played a Team Sport in High School	16.50 (2.51)	14.00 (2.39)
Did Not Play a Team Sport in High School	12.25 (3.62)	14.00 (2.45)

Expand Your Knowledge

12.97 Each cell in the table contains seven cases. (a) Given the cell, row, and column means below, calculate the missing cell means. If it can't be done, say so. (b) Calculate the mean for all 28 cases. If it can't be done, say so.

	Condition 1	Condition 2	Row Means
Control Group	A 4.00	B	8.00
Experimental Group	C	D	7.00
Column Means	6.00	9.00	

12.98 Each cell in this table has the same number of cases, but that number is unknown. (a) Given the cell, row, and column means below, calculate the missing cell means. If it can't be done, say so. (b) Calculate the mean for all the cases. If it can't be done, say so.

	Condition 1	Condition 2	Row Means
Control Group	A	B	6.00
Experimental Group	C 5.00	D	5.50
Column Means	7.00	4.50	

12.99 Each cell in the table has five cases. (a) Given the cell, row, and column means below, calculate the missing cell means. If it can't be done, say so. (b) Calculate the mean for all 30 cases. If it can't be done, say so.

	Condition 1	Condition 2	Condition 3	Row Means
Control Group	A 1.00	B	C	2.00
Experimental Group	D	E	F	6.00
Column Means	2.50	5.50	4.00	

12.100 The means of five independent samples are being compared. Which ANOVA is being used?
a. between-subjects, one-way ANOVA
b. one-way, repeated-measures ANOVA
c. between-subjects, two-way ANOVA
d. (a) or (b)
e. (a) or (c)
f. not enough information provided to decide

12.101 The means of six independent samples are being compared. Which ANOVA is being used?
a. between-subjects, one-way ANOVA
b. one-way, repeated-measures ANOVA
c. between-subjects, two-way ANOVA
d. (a) or (b)
e. (a) or (c)
f. not enough information provided to decide

12.102 Given $R = 3$, $C = 3$, $n = 12$, and the ANOVA summary table below, calculate *HSD* values for the effects as necessary and appropriate.

Source of Variability	Sum of Squares	Degrees of Freedom	Mean Square	*F* ratio
Between groups	184.00	8		
Rows	115.00	2	57.50	9.49
Columns	12.00	2	6.00	0.99
Interaction	57.00	4	14.25	2.35
Within groups	600.00	99	6.06	
Total	784.00	107		

12.103 If $SS_{Rows} = 17.50$, $SS_{Columns} = 13.75$, $SS_{Interaction} = 22.25$, and $SS_{Within} = 44.75$, calculate (a) $SS_{Between}$ and (b) SS_{Total}.

12.104 If $SS_{Rows} = 123.60$, $SS_{Columns} = 80.30$, $SS_{Interaction} = 20.80$, and $SS_{Within} = 168.20$, calculate (a) $SS_{Between}$ and (b) SS_{Total}.

SPSS

- Data to be analyzed for a between-subjects, two-way ANOVA with SPSS have to be entered in the data editor in a certain manner. Each case goes on a separate row and each of the three variables—the two independent variables and the one dependent variable—has a column. Figure 12.11 shows how the caffeine/sleep deprivation data are entered.
- The first column contains the data for the independent variable "Caffeine," which has two levels. Cases who consumed caffeine have a value of 1 and those who did not consume caffeine are given a value of 0.

- The second column, “Sleep,” contains information about which amount of sleep deprivation a case has experienced. Those with no sleep deprivation have the value of 0, one hour of sleep deprivation gets a 1, and two hours of sleep deprivation gets a 2.
- The third column, “Alertness,” is the case’s score on the mental alertness task.

There are two menus that must be accessed to run a between-subjects, two-way ANOVA in SPSS. To start, as shown in Figure 12.12, click on “Analyze,” then “General Linear Model,” and then “Univariate. . . .”

	Caffeine	Sleep	Alertness
1	1	0	82
2	1	0	85
3	1	0	88
4	1	0	91
5	1	0	94
6	1	1	82
7	1	1	87
8	1	1	90
9	1	1	93
10	1	1	98
11	1	2	79
12	1	2	83
13	1	2	86
14	1	2	89
15	1	2	93
16	0	0	72
17	0	0	76
18	0	0	80

Figure 12.11 Data Entry for Between-Subjects, One-Way ANOVA for SPSS Each case is on a separate row. The values for the two independent variables and for the dependent variable fall in the columns. (Source: SPSS)

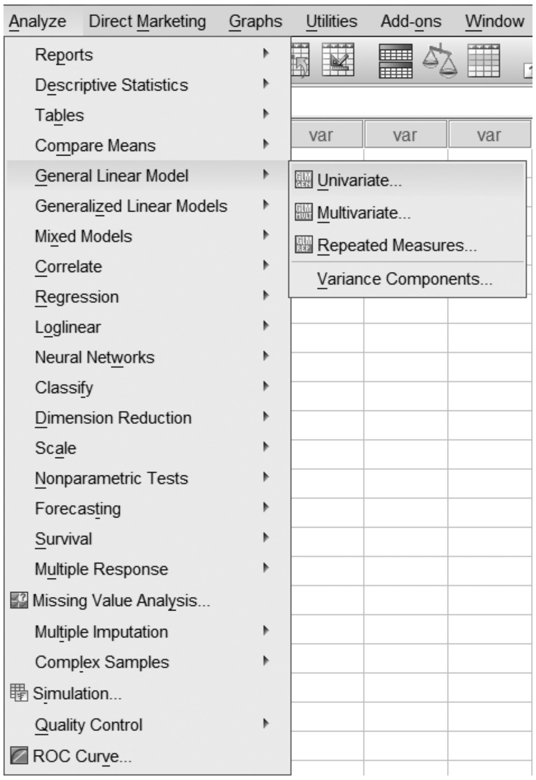


Figure 12.12 Initiating a Between-Subjects, One-Way ANOVA in SPSS Between-subjects, one-way ANOVA is found under “General Linear Model” in SPSS. (Source: SPSS)

Clicking on “Univariate. . .” opens up the menu seen in Figure 12.13. Note that the arrow buttons have already been used to send “Alertness” to the dependent variable box and “Caffeine” to the fixed factors box. Once the arrow button is used to send “Sleep” to the fixed factors box, press the “Go” button at the bottom to initiate the calculations.

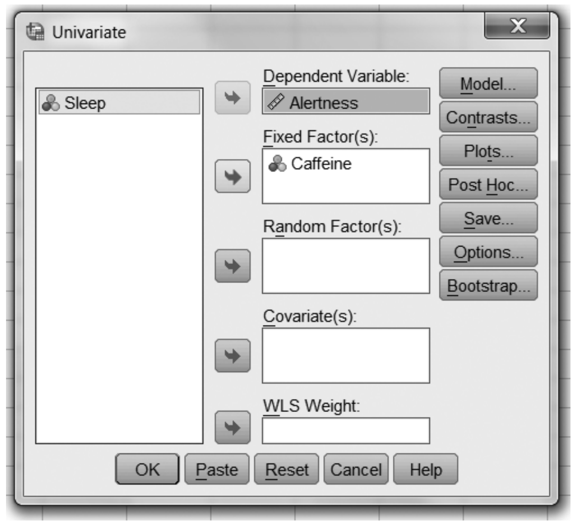


Figure 12.13 Commands to Run a Between-Subjects, One-Way ANOVA in SPSS To run a between-subjects, one-way ANOVA in SPSS, the independent variables are listed as "Fixed Factors" and the dependent variable as the "Dependent Variable." (Source: SPSS)

The results are seen in Figure 12.14. The SPSS ANOVA summary table is arranged a little differently from the one in the text. The columns in the summary table are the same, but there is one additional column labeled "Sig." This gives the exact significance level for an *F* ratio.

Tests of Between-Subjects Effects

Dependent Variable: Alertness

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5950.000[a]	5	1190.000	38.182	.000
Intercept	177870.000	1	177870.000	5707.059	.000
Caffeine	3630.000	1	3630.000	116.471	.000
Sleep	1340.000	2	670.000	21.497	.000
Caffeine * Sleep	980.000	2	490.000	15.722	.000
Error	748.000	24	31.167		
Total	184568.000	30			
Corrected Total	6698.000	29			

a. R Squared = .888 (Adjusted R Squared = .865)

Figure 12.14 ANOVA Summary Table Generated by SPSS for a Between-Subjects, One-Way ANOVA What the text calls row, column, and interaction effects are labeled by SPSS with their variable names: "Caffeine," "Sleep," and "Caffeine * Sleep." The asterisk in the interaction effect is an abbreviation for "by," so the interaction is pronounced "Caffeine by sleep." (Source: SPSS)

The rows in the summary table have some different sources of variability. Focus on the three that are labeled "Caffeine," "Sleep," and "Caffeine * Sleep."

- The row labeled "Caffeine" gives the results for the main effect of the caffeine variable.
- The row labeled "Sleep" gives the results for the main effect of the sleep deprivation variable.

- The interaction effect is the one in which the two main effects are connected with an asterisk, "Caffeine * Sleep."
- The row labeled "Error" is what the text calls "Within." This is the denominator term for all three of the *F* ratios.

The *F* ratios calculated by SPSS have one more decimal place than those calculated in this book, but otherwise they are the same. The final column in the summary table gives the exact significance level for an effect. If the alpha level is set at .05, then a result is statistically significant as long as the significance level reported in this column is $\leq$.05.

Appendix

Calculating Sums of Squares for Between-Subjects, Two-Way ANOVA

We'll take a few shortcuts in calculating sums of squares for between-subjects, two-way ANOVA. Because $SS_{Total} = SS_{Between} + SS_{Within}$, if one calculates any two, one can figure out the missing sum of squares. And as $SS_{Between} = SS_{Rows} + SS_{Columns} + SS_{Interaction}$, once $SS_{Between}$ is known, if we calculate SS_{Rows} and $SS_{Columns}$, we can figure out $SS_{Interaction}$ by subtraction.

As shown in Table 12.21, the first step is to sum and square the data. These data are the raw data for the caffeine and sleep deprivation study. Of SS_{Total}, SS_{Within}, and $SS_{Between}$, the easiest to calculate are SS_{Total} and $SS_{Between}$. Following Equation 10.2, SS_{Total} is calculated as follows:

$$\begin{aligned} SS_{Total} &= \Sigma X^2 - \frac{(\Sigma X)^2}{N} \\ &= 184{,}568.00 - \frac{2{,}310^2}{30} \\ &= 6{,}698.00 \end{aligned}$$

These data have been summed, squared, and arranged to expedite calculations of sums of squares for a between-subjects, two-way ANOVA.

Using Equation 10.3, $SS_{Between}$ is

$$\begin{aligned} SS_{Between} &= \Sigma\left(\frac{(\Sigma X_{Group})^2}{n}\right) - \frac{(\Sigma X)^2}{N} \\ &= \frac{440.00^2}{5} + \frac{450.00^2}{5} + \frac{430.00^2}{5} + \frac{400.00^2}{5} + \frac{340.00^2}{5} \\ &\quad + \frac{250.00^2}{5} - \frac{2{,}310^2}{30} = 5{,}950.00 \end{aligned}$$

Then, SS_{Within} is calculated as

$$\begin{aligned} SS_{Within} &= SS_{Total} - SS_{Between} \\ &= 6{,}698.00 - 5{,}950.00 \\ &= 748.00 \end{aligned}$$

TABLE 12.21 Raw Data Used in Calculating Sums of Squares for a Between-Subjects, Two-Way ANOVA

	Column 1		Column 2		Column 3			
	X	X^2	X	X^2	X	X^2		
	82	6,724.00	82	6,724.00	79	6,241.00		
	85	7,225.00	87	7,569.00	83	6,889.00		
	88	7,744.00	90	8,100.00	86	7,396.00		
	91	8,281.00	93	8,649.00	89	7,921.00		
Row 1	94	8,836.00	98	9,604.00	93	8,649.00	Row total	
n	5		5		5		15	
Σ	440.00	38,810.00	450.00	40,646.00	430.00	37,096.00	1,320.00	116,552.00
	72	5,184.00	62	3,844.00	44	1,936.00		
	76	5,776.00	64	4,096.00	45	2,025.00		
	80	6,400.00	68	4,624.00	50	2,500.00		
	83	6,889.00	72	5,184.00	55	3,025.00		
Row 2	89	7,921.00	74	5,476.00	56	3,136.00	Row total	
n	5		5		5		15	
Σ	400.00	32,170.00	340.00	23,224.00	250.00	12,622.00	990.00	68,016.00
	Column totals						Grand total	
n	10		10		10		30	
Σ	840.00	70,980.00	790.00	63,870.00	680.00	49,718.00	2,310.00	184,568.00

These data have been summed, squared, and arranged to expedite calculations of sums of squares for a between-subjects, two-way ANOVA.

The next step is to compute SS_{Rows} and SS_{Columns}:

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

$$= \frac{1{,}320.00^2}{15} + \frac{990.00^2}{15} + \frac{2{,}310^2}{30}$$

$$= 3{,}630.00$$

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

$$= \frac{840.00^2}{10} + \frac{790.00^2}{10} + \frac{680.00^2}{10} - \frac{2{,}310^2}{30}$$

$$= 1{,}340.00$$

The final sum of squares, $SS_{\text{Interaction}}$, is calculated by subtraction:

$$\begin{aligned} SS_{\text{Interaction}} &= SS_{\text{Between}} - SS_{\text{Rows}} - SS_{\text{Columns}} \\ &= 5{,}950.00 - 3{,}630.00 - 1{,}340.00 \\ &= 980.00 \end{aligned}$$