Introduction to the Analysis of Variance

Chapter Outline

TIP FOR SUCCESS

This chapter assumes you understand the logic of hypothesis testing and the *t* test, particularly the material on the distribution of means.

In this chapter, you will learn about a procedure for comparing *more than two groups of scores,* each of which is from an *entirely separate group of people.*

We begin with an example. In a classic study, Cindy Hazan and Phillip Shaver (1987) arranged to have the *Rocky Mountain News,* a large Denver area newspaper, print a mail-in survey. The survey included the question shown in Table 1 to measure what is called attachment style. Those who selected the first choice are "secure"; those who selected the second, "avoidant"; and those who selected the third, "anxious-ambivalent." These attachment styles are thought to be different ways of behaving and thinking in close relationships that develop from a person's experience

Table 1 Question Used in Hazan and Shaver (1987) Newspaper Survey

Which of the following best describes your feelings? [Check One]
[ ] I find it relatively easy to get close to others and am comfortable depending on them and having them depend on me. I don't often worry about being abandoned or about someone getting too close to me.
[ ] I am somewhat uncomfortable being close to others; I find it difficult to trust them completely, difficult to allow myself to depend on them. I am nervous when anyone gets too close, and often, love partners want me to be more intimate than I feel comfortable being.
[ ] I find that others are reluctant to get as close as I would like. I often worry that my partner doesn't really love me or won't want to stay with me. I want to merge completely with another person, and this desire sometimes scares people away.

Source: Hazan, C. and Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology, 52,* 515.

with early caregivers (Mikulincer & Shaver, 2007). (Of course, this single item is only a very rough measure that works for a large survey but is certainly not definitive in any particular person.) Readers also answered questions about various aspects of love, including amount of jealousy in their current or most recent relationship. Hazan and Shaver then compared the amount of jealousy reported by people with the three different attachment styles.

With a *t* test for independent means, Hazan and Shaver could have compared the mean jealousy scores of any two of the attachment styles. Instead, they were interested in differences among all three attachment styles. The statistical procedure for testing variation among the means of *more than two groups* is called the **analysis of variance,** abbreviated as **ANOVA.** (You could use the analysis of variance for a study with only two groups, but the simpler *t* test gives the same result.)

In this chapter, we introduce the analysis of variance, focusing on the fundamental logic, how to carry out an analysis of variance in the most basic situation, and how to make sense of more complicated forms of the analysis of variance when reading about them in research articles.

Basic Logic of the Analysis of Variance

The null hypothesis in an analysis of variance is that the several populations being compared all have the same mean. For example, in the attachment-style example, the null hypothesis is that the populations of secure, avoidant, and anxious-ambivalent people all have the same average degree of jealousy. The research hypothesis would be that the degree of jealousy is not the same among these three populations.

Hypothesis testing in analysis of variance is about whether the means of the samples differ more than you would expect if the null hypothesis were true. This question about *means* is answered, surprisingly, by analyzing *variances* (hence the name *analysis of variance*). Among other reasons, you focus on variances because when you want to know how several means differ, you are asking about the variation among those means.

Thus, to understand the logic of analysis of variance, we consider variances. In particular, we begin by discussing *two different ways* of estimating population variances. As you will see, the analysis of variance is about a comparison of the results of these two different ways of estimating population variances.

analysis of variance (ANOVA) Hypothesis-testing procedure for studies with three or more groups.

Estimating Population Variance from Variation within Each Sample

With the analysis of variance, as with the *t* test, you do not know the true population variances. However, as with the *t* test, you can estimate the variance of each of the populations from the scores in the samples. Also, as with the *t* test, you assume in the analysis of variance that all populations have the *same* variance. This allows you to average the population variance estimates from each sample into a single pooled estimate, called the **within-groups estimate of the population variance.** It is an average of estimates figured entirely from the scores *within* each of the samples.

One of the most important things to remember about this within-groups estimate of the population variance is that it is not affected by whether the null hypothesis is true. This estimate comes out the same whether the means of the populations are all the same (the null hypothesis is true) or the means of the populations are not all the same (the null hypothesis is false). This estimate comes out the same because it focuses only on the variation *inside* each population. Thus, it doesn't matter how far apart the means of the different populations are.

If the variation in scores within each sample is not affected by whether the null hypothesis is true, what determines the level of within-group variation? The answer is that chance factors (that is, factors that are unknown to the researcher) account for why different people in a sample have different scores. These chance factors include the fact that different people respond differently to the same situation or treatment and that there may be some experimental error associated with the measurement of the variable of interest. Thus, we can think of the within-groups population variance estimate as an estimate based on chance (or unknown) factors that cause different people in a study to have different scores.

Estimating the Population Variance from Variation between the Means of the Samples

There is also a second way to estimate the population variance. Each sample's mean is a number in its own right. If there are several samples, there are several such numbers, and these numbers will have some variation among them. The variation among these means gives another way to estimate the variance in the populations that the samples come from. Just how this works is a bit tricky, so follow the next two sections closely.

When the Null Hypothesis Is True. First, consider the situation in which the null hypothesis is true. In this situation, all samples come from populations that have the *same mean.* Remember, we are always assuming that all populations have the same variance (and also that they are all normal curves). Thus, if the null hypothesis is true, all populations are identical and thus they have the same mean, variance, and shape.

However, even when the populations are identical (that is, even when the null hypothesis is true), the samples will each be a little different. And if the samples are a little different, the sample means will each be a little different. How different can the sample means be? That depends on how much variation there is in each population. If a population has very little variation in the scores in it, then the means of samples from that population (or any identical population) will tend to be very similar to each other. When the null hypothesis is true, the variability among the sample means is influenced by the same chance (or unknown) factors that influence the variability among the scores in each sample.

within-groups estimate of the population variance (S^2_{Within}) In analysis of variance, estimate of the variance of the distribution of the population of individuals based on the variation among the scores within each of the actual groups studied.

What if several identical populations (with the same population mean) have a lot of variation in the scores in each? In that situation, if you take one sample from each population, the means of those samples could easily be very different from each other. Being very different, the variance of these means will be large. The point is that the more variance in each of several identical populations, the more variance there will be among the means of samples when you take a random sample from each population.

Suppose you were studying samples of six children from each of three large playgrounds (the populations in this example). If each playground had children who were all either 7 or 8 years old, the means of your three samples would all be between 7 and 8. Thus, there would not be much variance among those means. However, if each playground had children ranging from 3 to 12 years old, the means of the three samples would probably vary quite a bit. What this shows is that the variation among the means of samples is related directly to the amount of variation in each of the populations from which the samples are taken. The more variation in each population, the more variation there is among the means of samples taken from those populations.

This principle is shown in Figure 1. The three identical populations on the left have small variances, and the three identical populations on the right have large variances. In each set of three identical populations, even though the *means of the populations* (shown by triangles) are exactly the same, the *means of the samples* from those populations (shown by Xs) are not exactly the same. Most important, the sample means from the populations that each have a small amount of variance are closer together (have less variance among them). The sample means from the populations that each have more variance are more spread out (have more variance among them).

We have now seen that the variation among the means of samples taken from identical populations is related directly to the variation of the scores in each of those populations. This has a very important implication: It should be possible to estimate the variance in each population from the variation among the means of our samples.

Such an estimate is called a **between-groups estimate of the population variance.** (It has this name because it is based on the variation between the means of the samples, the "groups." Grammatically, it ought to be *among* groups; but *between* groups is traditional.) You will learn how to figure this estimate later in the chapter.

So far, all of this logic we have considered has assumed that the null hypothesis is true, so that there is no variation among the means of the *populations.* Let's now consider what happens when the null hypothesis is not true, when instead the research hypothesis is true.

When the Null Hypothesis Is Not True. If the null hypothesis is not true (and thus the research hypothesis is true), the populations themselves do not have the same mean. In this situation, variation among the means of samples taken from these populations is still caused by the chance factors that cause variation within the populations. So the larger the variation within the populations, the larger the variation will be among the means of samples taken from the populations. However, in this situation, in which the research hypothesis is true, variation among the means of the samples is *also* caused by the variation among the population means. You can think of this variation among population means as resulting from a *treatment effect*—that is, the different treatment received by the groups (as in an experiment) causes the groups to have different means. So, when the research hypothesis is true, the means of the samples are spread out for two different reasons: (1) because of variation within each of the populations (due to unknown chance factors) and (2) because of variation among the means of the populations (that is,

TIP FOR SUCCESS

You may want to read this paragraph more than once to ensure that you fully understand the logic we are presenting.

between-groups estimate of the population variance In an analysis of variance, estimate of the variance of the population of individuals based on the variation among the means of the groups studied.

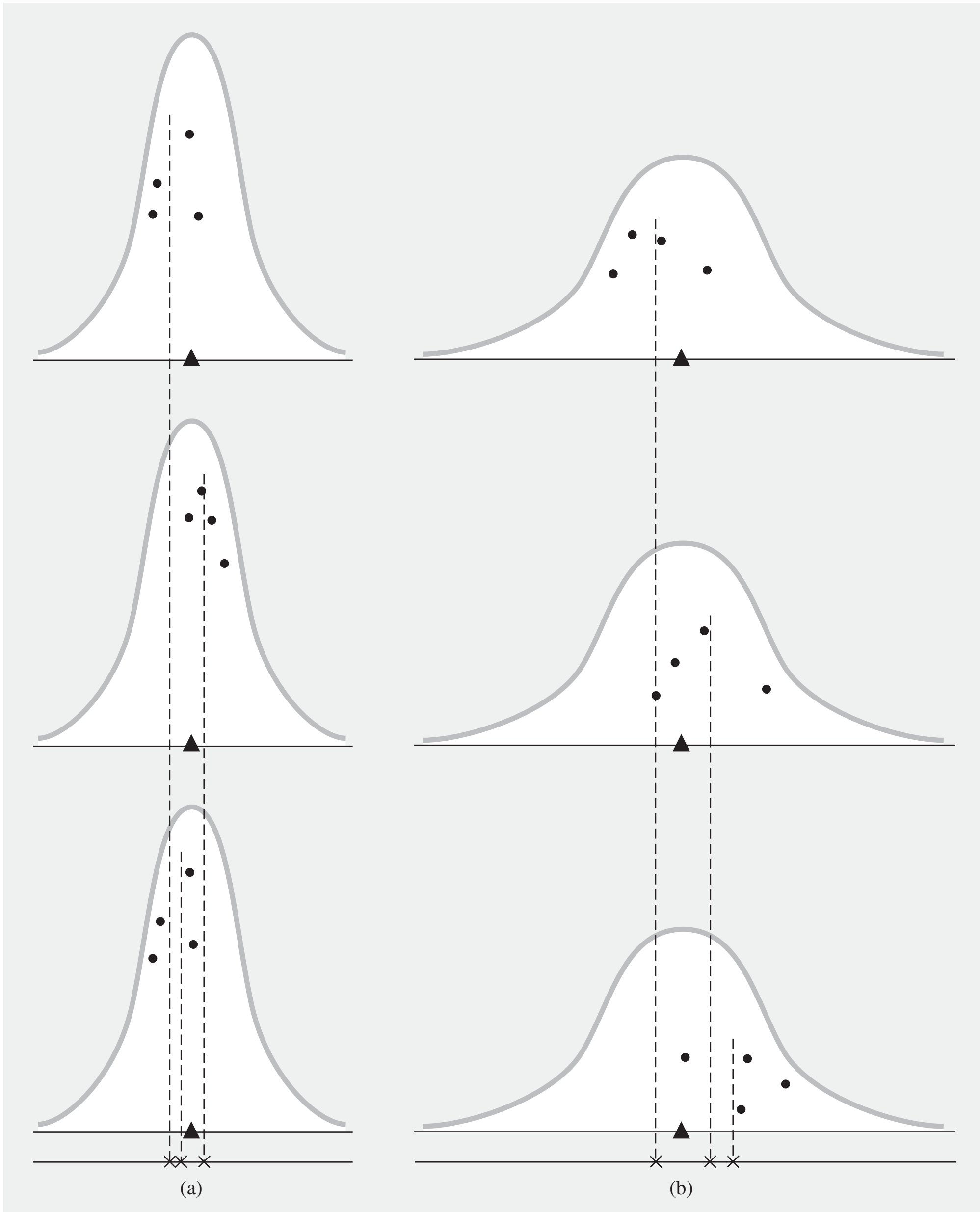


Figure 1 Means of samples from identical populations will not be identical. (a) Sample means from populations with less variation will vary less. (b) Sample means from populations with more variation will vary more. (Population means are indicated by a triangle, sample means by an X.)

a treatment effect). The left side of Figure 2 shows populations with the same means (shown by triangles) and the means of samples taken from them (shown by Xs). (This is the same situation as in both sides of Figure 1.) The right side of Figure 2 shows three populations with different means (shown by triangles) and the means of samples taken from them (shown by Xs). (This is the situation we are discussing in this section.)

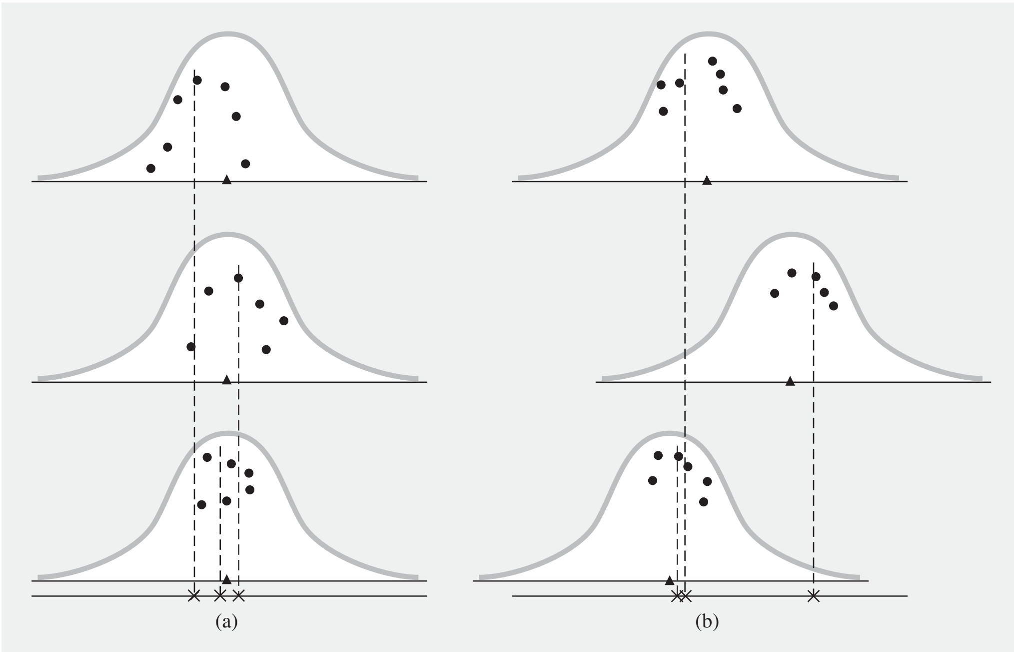


Figure 2 Means of samples from populations whose means differ (b) will vary more than sample means taken from populations whose means are the same (a). (Population means are indicated by a triangle, sample means by an X.)

Notice that the means of the samples are more spread out in the situation on the right side of Figure 2. This is true even though the variations in the populations are the same for the situation on both sides of Figure 2. This additional spread (variance) for the means on the right side of the figure is due to the populations having different means.

In summary, the between-groups estimate of the population variance is figured based on the variation among the means of the samples. If the null hypothesis is true, this estimate gives an accurate indication of the variation within the populations (that is, the variation due to chance factors). But if the null hypothesis is false, this method of estimating the population variance is influenced both by the variation within the populations (the variation due to chance factors) and the variation among the population means (the variation due to a treatment effect). It will not give an accurate estimate of the variation within the populations, because it also will be affected by the variation among the populations. This difference between the two situations has important implications. It is what makes the analysis of variance a method of testing hypotheses about whether there is a difference among means of groups.

Comparing the Within-Groups and Between-Groups Estimates of Population Variance

Table 2 summarizes what we have seen so far about the within-groups and between-groups estimates of population variance, both when the null hypothesis is true and when the research hypothesis is true. Let’s first consider the top part of the table, about when the null hypothesis is true. In this situation, the within-groups and between-groups estimates are based on the same thing (that is, the chance variation within populations). Literally, they are estimates of the same population variance. Therefore,

TIP FOR SUCCESS

Table 2 summarizes the logic of the analysis of variance. Test your understanding of this logic by trying to explain Table 2 without referring to the book. You might try writing your answer down and swapping it with someone else in your class.

Table 2 Sources of Variation in Within-Groups and Between-Groups Variance Estimates

	Variation Within Populations (Due to Chance Factors)	Variation Between Populations (Due to a Treatment Effect)
Null hypothesis is true		
Within-groups estimate reflects	✓	
Between-groups estimate reflects	✓	
Research hypothesis is true		
Within-groups estimate reflects	✓	
Between-groups estimate reflects	✓	✓

when the null hypothesis is true, both estimates should be about the same. (Only *about* the same; these are estimates.) Here is another way of describing this similarity of the between-groups estimate and the within-groups estimate when the null hypothesis is true. In this situation, the ratio of the between-groups estimate to the within-groups estimate should be approximately one to one. For example, if the within-groups estimate is 107.5, the between-groups estimate should be around 107.5, so that the ratio would be about 1. (A ratio is found by dividing one number by the other; thus $107.5/107.5 = 1$.)

Now let's turn to the bottom half of Table 2, about when the research hypothesis is true (and thus, the null hypothesis is not true). The situation is quite different here. When the research hypothesis is true, the between-groups estimate is influenced by two sources of variation: (1) the variation of the scores in each population (due to chance factors), and (2) the variation of the means of the populations from each other (due to a treatment effect). Yet even when the research hypothesis is true, the within-groups estimate still is influenced *only* by the variation in the populations. Therefore, when the research hypothesis is true, the between-groups estimate should be *larger* than the within-groups estimate.

In this situation, because the research hypothesis is true, the ratio of the between-groups estimate to the within-groups estimate should be greater than 1. For example, the between-groups estimate might be 638.9 and the within-groups estimate 107.5, making a ratio of 638.9 to 107.5 or 5.94. That is, when we divide the larger, the between-groups estimate, by the smaller, the within-groups estimate, we get not 1 but more than 1 (in the example, $638.9/107.5 = 5.94$). In this example the between-groups estimate is nearly six times bigger (5.94 times to be exact) than the within-groups estimate.

This is the central principle of the analysis of variance: *When the null hypothesis is true, the ratio of the between-groups population variance estimate to the within-groups population variance estimate should be about 1. When the research hypothesis is true, this ratio should be greater than 1.* If you figure this ratio and it comes out much bigger than 1, you can reject the null hypothesis. That is, it is unlikely that the null hypothesis could be true and the between-groups estimate be a lot bigger than the within-groups estimate.

The *F* Ratio

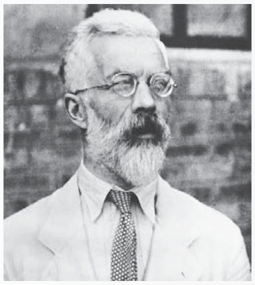
This crucial ratio of the between-groups to the within-groups population variance estimate is called an ***F* ratio.** (The *F* is for Sir Ronald Fisher, an eminent statistician who developed the analysis of variance; see Box 1.)

F ratio In analysis of variance, ratio of the between-groups population variance estimate to the within-groups population variance estimate; score on the comparison distribution (an *F* distribution) in an analysis of variance; also referred to simply as *F*.

The *F* Distribution and the *F* Table

We have said that if the crucial ratio of between-groups estimate to within-groups estimate (the *F* ratio) is a lot larger than 1, you can reject the null hypothesis. The next question is, "Just how much bigger than 1 should it be?"

BOX 1 Sir Ronald Fisher, Caustic Genius at Statistics



Courtesy of the Library of Congress

Ronald A. Fisher, a contemporary of William Gosset and Karl Pearson, was probably the brightest and certainly the most productive of this close-knit group of British statisticians. In the process of writing 300 papers and seven books, he developed many of the modern field's key concepts: variance, analysis of variance, significance levels, the null hypothesis, and almost all of our basic ideas of research design, including the fundamental importance of randomization. In fact, he has been described as producing "the scaffolding onto which the discipline [of statistics] was built" (Wright, 2009). For good measure, he was also a major contributor to the field of genetics.

A family legend is that little Ronald, born in 1890, was so fascinated by math that one day, at age 3, when put into his highchair for breakfast, he asked his nurse, "What is a half of a half?" Told it was a quarter, he asked, "What's half of a quarter?" To that answer he wanted to know what was half of an eighth. At the next answer he purportedly thought a moment and said, "Then I suppose that a half of a sixteenth must be a thirty-toof." Ah, baby stories.

As a grown man, however, Fisher seems to have been anything but darling. Some observers ascribe this to a cold and unemotional mother, but whatever the reason, throughout his life he was embroiled in bitter feuds, even with scholars who had previously been his closest allies and who certainly ought to have been comrades in research.

Fisher's thin ration of compassion extended to his readers as well; not only was his writing hopelessly obscure, but it often simply failed to supply important assumptions and proofs. Gosset said that when Fisher began a sentence with "Evidently," it meant two hours of hard work before one could hope to see why the point was evident.

Indeed, his lack of empathy extended to all of humankind. Like Galton, Fisher was fond of eugenics, favoring anything that might increase the birthrate of the upper and professional classes and skilled artisans. Not only did he see contraception as a poor idea—fearing that the least desirable persons would use it least—but he defended infanticide as serving an evolutionary function. It may be just as well that his opportunities to experiment with breeding never extended beyond the raising of his own children and some crops of potatoes and wheat.

Although Fisher eventually became the Galton Professor of Eugenics at University College, his most influential appointment probably came when he was invited to Iowa State College in Ames for the summers of 1931 and 1936 (where he was said to be so put out with the terrible heat that he stored his sheets in the refrigerator all day). At Ames, Fisher greatly impressed George Snedecor, a U.S. professor of mathematics also working on agricultural problems. Consequently, Snedecor wrote a textbook of statistics for agriculture that borrowed heavily from Fisher's work. The book so popularized Fisher's ideas about statistics and research design that its second edition sold 100,000 copies.

You can learn more about Fisher at the following Web site: http://www-history.mcs.st-and.ac.uk/~history/Biographies/Fisher.html.

Sources: Peters (1987), Salsburg (2001), Stigler (1986), Tankard (1984), Wright (2009).

Statisticians have developed the mathematics of an ***F* distribution** and have prepared tables of *F* ratios. For any given situation, you merely look up in an ***F* table** how extreme an *F* ratio is needed to reject the null hypothesis at, say, the .05 level. (You learn to use the *F* table later in the chapter.)

For an example of an *F* ratio, let's return to the attachment-style study. The results of that study, for jealousy, were as follows: The between-groups population variance estimate was 23.27 and the within-groups population variance estimate was .53. (You will learn shortly how to figure these estimates on your own.) The ratio of the between-groups to the within-groups variance estimates (23.27/.53) came out to 43.91; that is, $F = 43.91$. This *F* ratio is considerably larger than 1. The *F* ratio needed to reject the null hypothesis at the .05 level in this study is only 3.01. Thus, the researchers confidently rejected the null hypothesis and concluded that the amount of jealousy is not the same among the three attachment styles. (Mean jealousy ratings were 2.17 for secures, 2.57 for avoidants,

F distribution Mathematically defined curve that is the comparison distribution used in an analysis of variance; distribution of *F* ratios when the null hypothesis is true.

F table Table of cutoff scores on the *F* distribution for various degrees of freedom and significance levels.

and 2.88 for anxious-ambivalents. The question was "I [loved/love] _____ so much that I often [felt/feel] jealous." The blank was for the respondent's important lover's name. Responses were on a 1 to 4 scale from *strongly disagree* to *strongly agree.*)

An Analogy

Some students find an analogy helpful in understanding the analysis of variance. The analogy is to what engineers call the "signal-to-noise ratio." For example, your ability to make out the words in a staticky cell-phone conversation depends on the strength of the signal versus the amount of random noise. With the *F* ratio in the analysis of variance, the difference among the means of the samples is like the signal; it is the information of interest. The variation within the samples is like the noise. When the variation among the samples is sufficiently great in comparison to the variation within the samples, you conclude that there is a significant effect.

TIP FOR SUCCESS

These "How are you doing?" questions and answers provide a useful summary of the logic of the analysis of variance. Be sure to review them (and the relevant sections in the text) as many times as necessary to fully understand this logic.

How are you doing?

1. When do you use an analysis of variance?
2. (a) What is the within-groups population variance estimate based on? (b) How is it affected by the null hypothesis being true or not? (c) Why?
3. (a) What is the between-groups population variance estimate based on? (b) How is it affected by the null hypothesis being true or not? (c) Why?
4. What are two sources of variation that can contribute to the between-groups population variance estimate?
5. (a) What is the *F* ratio; (b) why is it usually about 1 when the null hypothesis is true; and (c) why is it usually larger than 1 when the null hypothesis is false?

Answers

1. You use the analysis of variance when you are comparing means of samples from more than two populations.
2. (a) The within-groups population variance estimate is based on the variation among the scores in each of the samples. (b) It is not affected. (c) Whether the null hypothesis is true has to do with whether the means of the populations differ. Thus, the within-groups estimate is not affected by whether the null hypothesis is true. This is because the variation *within* each population (which is the basis for the variation within each sample) is not affected by whether the population means differ.
3. (a) The between-groups population variance estimate is based on the variation among the means of the samples. (b) It is larger when the null hypothesis is false. (c) Whether the null hypothesis is true has to do with whether the means of the populations differ. When the null hypothesis is false, the means of the populations differ. Thus, the between-groups estimate is bigger when the null hypothesis is false. This is because the variation among the means of the populations (which is one basis for the variation among the means of the samples) is greater when the population means differ.
4. Two sources of variation that can contribute to the between-groups population variance estimate are: (1) variation among the scores in each of the populations (that is, variation due to chance factors) and (2) variation among the means of the populations (that is, variation due to a treatment effect).
5. (a) The *F* ratio is the ratio of the between-groups population variance estimate to the within-groups population variance estimate. (b) Both estimates are based

entirely on the same source of variation—the variation of the scores within each of the populations (that is, due to chance factors). (c) The between-groups estimate is also influenced by the variation among the means of the populations (that is, a treatment effect) whereas the within-groups estimate is not. Thus, when the null hypothesis is false (and thus the means of the populations are not the same), the between-groups estimate will be bigger than the within-groups estimate.

Carrying Out an Analysis of Variance

Now that we have considered the basic logic of the analysis of variance, we will go through an example to illustrate the details. (We will use a fictional study to keep the numbers simple.)

Suppose a researcher is interested in the influence of knowledge of previous criminal record on juries' perceptions of the guilt or innocence of defendants. The researcher recruits 15 volunteers who have been selected for jury duty (but have not yet served at a trial). The researcher shows them a DVD of a 4-hour trial in which a woman is accused of passing bad checks. Before viewing the DVD, however, all of the research participants are given a "background sheet" with age, marital status, education, and other such information about the accused woman. The sheet is the same for all 15 participants, with one difference. For five of the participants, the last section of the sheet says that the woman has been convicted several times before for passing bad checks; we will call these participants the *Criminal Record Group.* For five other participants, the last section of the sheet says the woman has a completely clean criminal record—the *Clean Record Group.* For the remaining five participants, the sheet does not mention anything about criminal record one way or the other—the *No Information Group.*

The participants are randomly assigned to groups. After viewing the DVD of the trial, all 15 participants make a rating on a 10-point scale, which runs from completely sure she is innocent (1) to completely sure she is guilty (10). The results of this fictional study are shown in Table 3. As you can see, the means of the three groups are not the same (8, 4, and 5). Yet there is also quite a bit of variation within each of the

Table 3 Results of the Criminal Record Study (Fictional Data)

	Criminal Record Group			Clean Record Group			No Information Group	
Rating	Deviation from Mean	Squared Deviation from Mean	Rating	Deviation from Mean	Squared Deviation from Mean	Rating	Deviation from Mean	Squared Deviation from Mean
10	2	4	5	1	1	4	−1	1
7	−1	1	1	−3	9	6	1	1
5	−3	9	3	−1	1	9	4	16
10	2	4	7	3	9	3	−2	4
8	0	0	4	0	0	3	−2	4
Σ: 40	0	18	20	0	20	25	0	26
$M = 40/5 = 8$			$M = 20/5 = 4$			$M = 25/5 = 5$		
$S^2 = 18/4 = 4.5$			$S^2 = 20/4 = 5.0$			$S^2 = 26/4 = 6.5$		

three groups. Population variance estimates from the scores within each of these three groups are 4.5, 5.0, and 6.5.

You need to figure the following numbers to test the hypothesis that the three populations are different: (a) a population variance estimate based on the variation of the scores within each of the samples, (b) a population variance estimate based on the differences between the group means, and (c) the ratio of the two, the *F* ratio. (In addition, we need the significance cutoff from an *F* table.)

Figuring the Within-Groups Estimate of the Population Variance

You can estimate the population variance from any one group (that is, from any one sample) using the usual method of estimating a population variance from a sample. First, you figure the sum of the squared deviation scores. That is, you take the deviation of each score from its group's mean, square that deviation score, and sum all the squared deviation scores. Second, you divide that sum of squared deviation scores by that group's degrees of freedom. (The degrees of freedom for a group are the number of scores in the group minus 1.) For the example, as shown in Table 3, this gives an estimated population variance of 4.5 based on the Criminal Record Group's scores, an estimate of 5.0 based on the Clean Record Group's scores, and an estimate of 6.5 based on the No Information Group's scores.

Once again, in the analysis of variance, as with the *t* test, we assume that the populations have the same variance. The estimates based on each sample's scores are all estimating the same true population variance. The sample sizes are equal in this example so that the estimate for each group is based on an equal amount of information. Thus, you can pool these variance estimates by straight averaging. This gives an overall estimate of the population variance based on the variation within groups of 5.33 (that is, the sum of 4.5, 5.0, and 6.5, which is 16, divided by 3, the number of groups).

To summarize, the two steps for figuring the within-groups population variance estimate are:

Ⓐ **Figure population variance estimates based on each group's scores.**
Ⓑ **Average these variance estimates.**

The estimated population variance based on the variation of the scores within each of the groups is the within-groups variance estimate. This is symbolized as S^2_{Within}. In terms of a formula (when sample sizes are all equal),

The within-groups population variance estimate is the sum of the population variance estimates based on each sample, divided by the number of groups.

$$S^2_{\text{Within}} = \frac{S^2_1 + S^2_2 + \cdots + S^2_{\text{Last}}}{N_{\text{Groups}}} \tag{1}$$

In this formula, S^2_1 is the estimated population variance based on the scores in the first group (the group from Population 1), S^2_2 is the estimated population variance based on the scores in the second group, and S^2_{Last} is the estimated population variance based on the scores in the last group. (The dots, or ellipses, in the formula show that you are to fill in a population variance estimate for as many other groups as there are in the analysis.) N_{Groups} is the number of groups. Using this formula for our figuring, we get

$$S^2_{\text{Within}} = \frac{S^2_1 + S^2_2 + \cdots + S^2_{\text{Last}}}{N_{\text{Groups}}} = \frac{4.5 + 5.0 + 6.5}{3} = \frac{16}{3} = 5.33$$

Figuring the Between-Groups Estimate of the Population Variance

Figuring the between-groups estimate of the population variance also involves two steps (though quite different ones from figuring the within-groups estimate). First, estimate the variance of a distribution of means from the means of your samples. Second, based on the variance of this distribution of means, figure the variance of the population of individuals. Here are the two steps in more detail:

Ⓐ **Estimate the variance of the distribution of means:** Add up the sample means' squared deviations from the overall mean (the mean of all the scores) and divide this by the number of means minus 1.

You can think of the means of your samples as taken from a distribution of means. You follow the standard procedure of using the scores in a sample to estimate the variance of the population from which these scores are taken. In this situation, you think of the means of your samples as the scores and the distribution of means as the population from which these scores come. What this all boils down to are the following procedures. You begin by figuring the sum of squared deviations. (You find the mean of your sample means, figure the deviation of each sample mean from this mean of means, square each of these deviations, and finally sum these squared deviations.) Then, divide this sum of squared deviations by the degrees of freedom, which is the number of means minus 1. In terms of a formula (when sample sizes are all equal),

$$S_M^2 = \frac{\Sigma(M - GM)^2}{df_{\text{Between}}} \qquad (2)$$

The estimated variance of the distribution of means is the sum of each sample mean's squared deviation from the grand mean, divided by the degrees of freedom for the between-groups population variance estimate.

In this formula, S_M^2 is the estimated variance of the distribution of means (estimated based on the means of the samples in your study). M is the mean of each of your samples. GM is the **grand mean,** the overall mean of all your scores, which is also the mean of your means. df_{Between} is the degrees of freedom in the between-groups estimate, the number of groups minus 1. Stated as a formula,

$$df_{\text{Between}} = N_{\text{Groups}} - 1 \qquad (3)$$

The degrees of freedom for the between-groups population variance estimate is the number of groups minus 1.

N_{Groups} is the number of groups.

In the criminal-record example, the three means are 8, 4, and 5. The figuring of S_M^2 is shown in Table 4.

grand mean (GM) In analysis of variance, overall mean of all the scores, regardless of what group they are in; when group sizes are equal, mean of the group means.

Table 4 Estimated Variance of the Distribution of Means Based on Means of the Three Experimental Groups in the Criminal Record Study (Fictional Data)

	Sample Means	Deviation from Grand Mean	Squared Deviation from Grand Mean
	(M)	($M - GM$)	$(M - GM)^2$
	4	−1.67	2.79
	8	2.33	5.43
	5	− .67	.45
Σ	17	−0.01	8.67

$GM = (\Sigma M)/N_{\text{Groups}} = 17/3 = 5.67$; $S_M^2 = [\Sigma(M - GM)^2]/df_{\text{Between}} = 8.67/2 = 4.34$

B **Figure the estimated variance of the population of individual scores:** Multiply the variance of the distribution of means by the number of scores in each group.

What we have just figured in Step A, from a sample of a few means, is the estimated variance of a distribution of means. From this we want to estimate the variance of the population (the distribution of individuals) on which the distribution of means is based. The variance of a distribution of means is smaller than the variance of the population (the distribution of individuals) on which it is based. This is because means are less likely to be extreme than are individual scores (because any one sample is unlikely to include several scores that are extreme in the same direction). Specifically, the variance of a distribution of means is the variance of the distribution of individual scores divided by the number of scores in each sample.

You should be familiar with figuring the variance of the distribution of means by *dividing* the variance of the distribution of individuals by the sample size. Now, however, you are going to figure the variance of the distribution of individuals by *multiplying* the variance of the distribution of means by the sample size (see Table 5). That is, to come up with the variance of the population of individuals, you multiply your estimate of the variance of the distribution of means by the sample size in each of the groups. The result of all this is the between-groups population variance estimate. Stated as a formula (for when sample sizes are equal),

The between-groups population variance estimate is the estimated variance of the distribution of means multiplied by the number of scores in each group.

$$S^2_{\text{Between}} = (S^2_M)(n) \tag{4}$$

In this formula, S^2_{Between} is the estimate of the population variance based on the variation among the means (the between-groups population variance estimate); n is the number of participants in each sample.

Let's return to our example in which there were five participants in each sample and an estimated variance of the distribution of means of 4.34. In this example, multiplying 4.34 by 5 gives a between-groups population variance estimate of 21.70. In terms of the formula,

$$S^2_{\text{Between}} = (S^2_M)(n) = (4.34)(5) = 21.70$$

To summarize, the procedure of estimating the population variance based on the differences between group means is (a) figure the estimated variance of the distribution of means and then (b) multiply that estimated variance by the number of scores in each group.

Figuring the *F* Ratio

The *F* ratio is the ratio of the between-groups estimate of the population variance to the within-groups estimate of the population variance. Stated as a formula,

The *F* ratio is the between-groups population variance estimate divided by the within-groups population variance estimate.

$$F = \frac{S^2_{\text{Between}}}{S^2_{\text{Within}}} \tag{5}$$

Table 5 Comparison of Figuring the Variance of a Distribution of Means from the Variance of a Distribution of Individuals, and the Reverse

- From distribution of individuals to distribution of means: $S^2_M = S^2/n$
- From distribution of means to distribution of individuals: $S^2 = (S^2_M)(n)$

In the example, the ratio of between and within is 21.70 to 5.33. Carrying out the division gives an F ratio of 4.07. In terms of the formula,

$$F = \frac{S^2_{\text{Between}}}{S^2_{\text{Within}}} = \frac{21.70}{5.33} = 4.07$$

TIP FOR SUCCESS

A very common mistake when figuring the F ratio is to turn the formula upside down. Just remember it is as simple as Black and White, so it is Between divided by Within.

The *F* Distribution

You are not quite done. You still need to find the cutoff for the F ratio that is large enough to reject the null hypothesis. This requires a distribution of F ratios that you can use to figure out what is an extreme F ratio.

In practice, you simply look up the needed cutoff in a table (or read the exact significance from the computer output). To understand from where that number in the table comes, you need to understand the F distribution. The easiest way to understand this distribution is to think about how you would go about making one.

Start with three identical populations. Next, randomly select five scores from each. Then, on the basis of these three samples (of five scores each), figure the F ratio. (That is, you use these scores to make a within-groups estimate and a between-groups estimate, then divide the between estimate by the within estimate.) Let's say that you do this and the F ratio you come up with is 1.36. Now you select three new random samples of five scores each and figure the F ratio using these three samples. Perhaps you get an F of 0.93. If you do this whole process many, many times, you will eventually get a lot of F ratios. The distribution of all possible F ratios figured in this way (from random samples from identical populations) is called the F distribution. Figure 3 shows an example of an F distribution. (There are many different F distributions, and each has a slightly different shape. The exact shape depends on how many samples you take each time and how many scores are in each sample. The general shape is like that shown in the figure.)

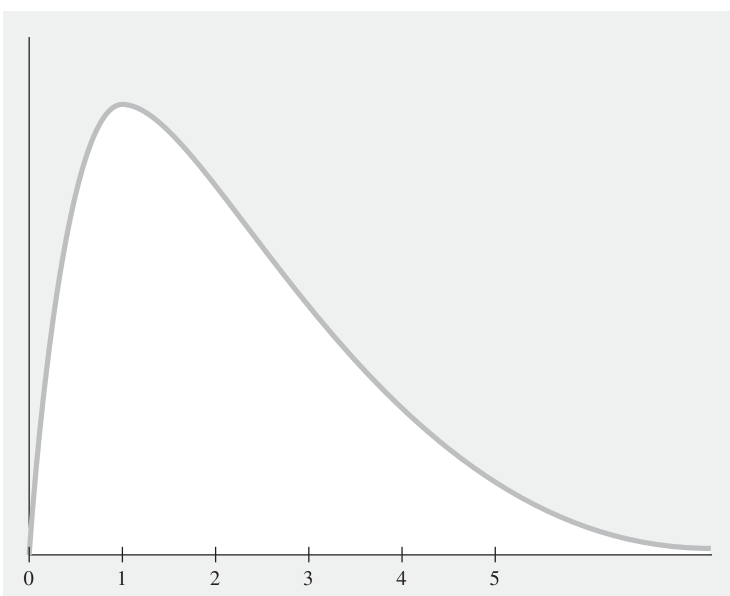


Figure 3 An *F* distribution.

No one actually goes about making F distributions in this way. It is a mathematical distribution whose exact characteristics can be found from a formula. Statisticians can also prove that the formula gives exactly the same result as if you had the patience to follow this procedure of taking random samples and figuring the F ratio of each for a very long time.

As you can see in Figure 3, the F distribution is not symmetrical, but has a long tail on the right. The reason for this positive skew is that an F distribution is a distribution of ratios of variances. Variances are always positive numbers. (A variance is an average of squared deviations, and anything squared is a positive number.) A ratio of a positive number to a positive number can never be less than 0. Yet there is nothing to stop a ratio from being a very high number. Thus, the F ratio's distribution cannot be lower than 0 and can rise quite high. (Most F ratios pile up near 1, but they spread out more on the positive side, where they have more room to spread out.)

The *F* Table

The F table is a little more complicated than the t table. This is because there is a different F distribution according to both the degrees of freedom used in the between-groups variance estimate and the degrees of freedom used in the within-groups variance estimate. That is, you have to take into account two different degrees of freedom to look up the needed cutoff. One is the **between-groups degrees of freedom.** It is also called the *numerator degrees of freedom.* This is the degrees of freedom you use in the between-groups variance estimate, the numerator of the F ratio. As shown earlier in Formula 3, the degrees of freedom for the between-groups population variance estimate is the number of groups minus 1 ($df_{\text{Between}} = N_{\text{Groups}} - 1$).

between-groups degrees of freedom (df_{Between}) Degrees of freedom used in the between-groups estimate of the population variance in an analysis of variance (the numerator of the F ratio); number of scores free to vary (number of means minus 1) in figuring the between-groups estimate of the population variance; same as *numerator degrees of freedom.*

The other type of degrees of freedom is the **within-groups degrees of freedom,** also called the *denominator degrees of freedom.* This is the sum of the degrees of freedom from each sample you use when figuring out the within-groups variance estimate, the denominator of the F ratio. Stated as a formula,

The degrees of freedom for the within-groups population variance estimate is the sum of the degrees of freedom used in making estimates of the population variance from each sample.

$$df_{\text{Within}} = df_1 + df_2 + \ldots + df_{\text{Last}} \quad (6)$$

In the criminal-record study example, the between-groups degrees of freedom is 2. (There are 3 means, minus 1.) In terms of the formula,

$$df_{\text{Between}} = N_{\text{Groups}} - 1 = 3 - 1 = 2$$

The within-groups degrees of freedom is 12. This is because each of the groups has 4 degrees of freedom on which the estimate is based (5 scores minus 1) and there are three groups overall, making a total of 12 degrees of freedom. In terms of the formula,

within-groups degrees of freedom (df_{Within}) Degrees of freedom used in the within-groups estimate of the population variance in an analysis of variance (the denominator of the F ratio); number of scores free to vary (number of scores in each group minus 1, summed over all the groups) in figuring the within-groups population variance estimate; same as *denominator degrees of freedom.*

$$\begin{aligned} df_{\text{Within}} &= df_1 + df_2 + \ldots + df_{\text{Last}} = (5 - 1) + (5 - 1) + (5 - 1) \\ &= 4 + 4 + 4 = 12 \end{aligned}$$

You would look up the cutoff for an F distribution "with 2 and 12" degrees of freedom. As shown in Table 6, for the .05 level, you need an F ratio of 3.89 to reject the null hypothesis.

Table 6 Selected Cutoffs for the *F* Distribution (with Values Highlighted for the Criminal Record Study)

Denominator Degrees of Freedom	Significance Level	Numerator Degrees of Freedom					
		1	*2*	*3*	*4*	*5*	*6*
10	.01	10.05	7.56	6.55	6.00	5.64	5.39
	.05	4.97	4.10	3.71	3.48	3.33	3.22
	.10	3.29	2.93	2.73	2.61	2.52	2.46
11	.01	9.65	7.21	6.22	5.67	5.32	5.07
	.05	4.85	3.98	3.59	3.36	3.20	3.10
	.10	3.23	2.86	2.66	2.54	2.45	2.39
12	.01	9.33	6.93	5.95	5.41	5.07	4.82
	.05	4.75	3.89	3.49	3.26	3.11	3.00
	.10	3.18	2.81	2.61	2.48	2.40	2.33
13	.01	9.07	6.70	5.74	5.21	4.86	4.62
	.05	4.67	3.81	3.41	3.18	3.03	2.92
	.10	3.14	2.76	2.56	2.43	2.35	2.28

Note: Full table is Table A-3 in the Appendix.

How are you doing?

For part (c) of each question, use the following scores involving three samples: The scores in Sample A are 5 and 7 ($M = 6$), the scores in Sample B are 6 and 9 ($M = 7.5$), and the scores in Sample C are 8 and 9 ($M = 8.5$).

1. (a) Write the formula for the within-groups population variance estimate and (b) define each of the symbols. (c) Figure the within-groups population variance estimate for these scores.
2. (a) Write the formula for the variance of the distribution of means when using it as part of an analysis of variance and (b) define each of the symbols. (c) Figure the variance of the distribution of means for these scores.
3. (a) Write the formula for the between-groups population variance estimate based on the variance of the distribution of means and (b) define each of the symbols and explain the logic behind this formula. (c) Figure the between-groups population variance estimate for these scores.
4. (a) Write the formula for the *F* ratio and (b) define each of the symbols and explain the logic behind this formula. (c) Figure the *F* ratio for these scores.
5. (a) Write the formulas for the between-groups and within-groups degrees of freedom and (b) define each of the symbols. (c) Figure the between-groups and within-groups degrees of freedom for these scores.
6. (a) What is the *F* distribution? (b) Why is it skewed to the right? (c) What is the cutoff *F* for these scores for the .05 significance level?

Answers

1. (a) Formula for the within-groups population variance estimate: $S^2_{\text{Within}} = (S^2_1 + S^2_2 + \ldots + S^2_{\text{Last}})/N_{\text{Groups}}$.

(b) S^2_{Within} is the within-groups population variance estimate; S^2_1 is the estimated population variance based on the scores in the first group (the group from Population A); S^2_2 is the estimated population variance based on the scores in the second group; S^2_{Last} is the estimated population variance based on the scores in the last group; the dots show that you are to fill in the population variance estimate for as many other groups as there are in the analysis; N_{Groups} is the number of groups.

(c) Figuring for the within-groups population variance estimate:

$$S^2_1 = [(5 - 6)^2 + (7 - 6)^2]/(2 - 1) = (1 + 1)/1 = 2;$$
$$S^2_2 = [(6 - 7.5)^2 + (9 - 7.5)^2]/(2 - 1) = (2.25 + 2.25)/1 = 4.5;$$
$$S^2_3 = [(8 - 8.5)^2 + (9 - 8.5)^2]/(2 - 1) = (.25 + .25)/1 = .5;$$
$$S^2_{\text{Within}} = (S^2_1 + S^2_2 + \ldots + S^2_{\text{Last}})/N_{\text{Groups}} = (2 + 4.5 + .5)/3 = 7/3 = 2.33.$$

2. (a) Formula for the variance of the distribution of means when using it as part of an analysis of variance: $S^2_M = \Sigma(M - GM)^2/df_{\text{Between}}$.

(b) S^2_M is the estimated variance of the distribution of means (estimated based on the means of the samples in your study). M is the mean of each of your samples. GM is the grand mean, the overall mean of all your scores, which is also the mean of your means. df_{Between} is the degrees of freedom in the between-groups estimate, the number of groups minus 1.

(c) Grand mean (GM) is $(6 + 7.5 + 8.5)/3 = 7.33$.

$$S^2_M = [\Sigma(M - GM)^2]/df_{\text{Between}}$$
$$= [(6 - 7.33)^2 + (7.5 - 7.33)^2 + (8.5 - 7.33)^2]/3 - 1$$
$$= (1.77 + .03 + 1.37)/2 = 3.17/2 = 1.585.$$

3. (a) Formula for the between-groups population variance estimate based on the variance of the distribution of means: $S^2_{\text{Between}} = (S^2_M)(n)$.

(b) S^2_{Between} is the between-groups population variance estimate; S^2_M is the estimated variance of the distribution of means (estimated based on the means of the samples in your study); n is the number of participants in each sample. The goal is to have a variance of a distribution of individuals based on the variation among the means of the groups. S^2_M is the estimate of the variance of a distribution of means from the overall population based on the means of the samples. To go from the variance of a distribution of means to the variance of a distribution of individuals, you multiply by the size of each sample. This is because the variance of the distribution of means is always smaller than the distribution of individuals (because means of samples are less likely to be extreme than are individual scores); the exact relation is that the variance of the distribution of means is the variance of the distribution of individuals divided by the sample size; thus you reverse that process here.

(c) $S^2_{\text{Between}} = (S^2_M)(n) = (1.585)(2) = 3.17$.

4. (a) Formula for the F ratio: $F = S^2_{\text{Between}}/S^2_{\text{Within}}$.

(b) F is the F ratio; S^2_{Between} is the between-groups population variance estimate; S^2_{Within} is the within-groups population variance estimate. The null hypothesis is that all the samples come from populations with the same mean. If the null hypothesis is true, the F ratio should be about 1. This is

because the two population variance estimates are based on the same thing, the variation within each of the populations (due to chance factors). If the research hypothesis is true, so that the samples come from populations with different means, the F ratio should be larger than 1. This is because the between-groups estimate is now influenced by the variation both within the populations (due to chance factors) and among them (due to a treatment effect). But the within-groups estimate is still affected only by the variation within each of the populations.
(c) $F = S^2_{Between}/S^2_{Within} = 3.17/2.33 = 1.36$.

5. (a) Formulas for the between-groups and within-groups degrees of freedom: $df_{Between} = N_{Groups} - 1$ and $df_{Within} = df_1 + df_2 + \ldots + df_{Last}$.
(b) $df_{Between}$ is the between-groups degrees of freedom; N_{Groups} is the number of groups; df_{Within} is the within-groups degrees of freedom; df_1 is the degrees of freedom for the population variance estimate based on the scores in the first sample; df_2 is the degrees of freedom for the population variance estimate based on the scores in the second sample; df_{Last} is the degrees of freedom for the population variance estimate based on the scores in the last sample; the dots show that you are to fill in the population degrees of freedom for as many other samples as there are in the analysis.
(c) $df_{Between} = N_{Groups} - 1 = 3 - 1 = 2$; $df_{Within} = df_1 + df_2 + \ldots + df_{Last} = 1 + 1 + 1 = 3$.

6. (a) The F distribution is the distribution of F ratios you would expect by chance. (b) F ratios are a ratio of variances. Variances as averages of squared numbers have to be positive, thus F ratios are ratios of two positive numbers, which always have to be positive and can't be less than 0. But there is no limit to how high an F ratio can be. The result is that the scores bunch up at the left (near 0) and spread out to the right. (c) 9.55.

Hypothesis Testing with the Analysis of Variance

Here are the five steps of hypothesis testing for the criminal record study.

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are three populations:

Population 1: Jurors told that the defendant has a criminal record.
Population 2: Jurors told that the defendant has a clean record.
Population 3: Jurors given no information of the defendant's record.

The null hypothesis is that these three populations have the same mean. The research hypothesis is that the populations' means are not the same.

❷ **Determine the characteristics of the comparison distribution.** The comparison distribution is an F distribution with 2 and 12 degrees of freedom.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** Using the F table for the .05 significance level, the cutoff F ratio is 3.89.

❹ **Determine your sample's score on the comparison distribution.** In the analysis of variance, the comparison distribution is an F distribution, and the sample's score on that distribution is thus its F ratio. In the example, the F ratio was 4.07.

❺ **Decide whether to reject the null hypothesis.** In the example, the F ratio of 4.07 is more extreme than the .05 significance level cutoff of 3.89. Thus, the researcher would reject the null hypothesis that the three groups come from populations with the same mean. This suggests that they come from populations with different means: that people exposed to different kinds of information (or no information) about the criminal record of a defendant in a situation of this kind will differ in their ratings of the defendant's guilt.

You may be interested to know that several real studies have looked at whether knowing a defendant's prior criminal record affects the likelihood of conviction (e.g., Steblay, Hosch, Culhane, & McWethy, 2006). The overall conclusion seems to be reasonably consistent with that of the fictional study described here.

Summary of Steps for Hypothesis Testing Using the Analysis of Variance

Table 7 summarizes the steps of an analysis of variance of the kind we have been considering in this chapter.[1]

Assumptions in the Analysis of Variance

The assumptions for the analysis of variance are basically the same as for the t test for independent means. That is, you get strictly accurate results only when the populations follow a normal curve and have equal variances. As with the t test, in practice you get quite acceptable results even when your populations are moderately far from normal and have moderately different variances. As a general rule, if the variance estimate of the group with the largest estimate is no more than four or five times that of the smallest and the sample sizes are equal, the conclusions using the F distribution should be reasonably accurate.

[1]There are, as usual, computational formulas that can be used for analysis of variance. They can be used in the unlikely situation in which you have to do an analysis of variance for a real study by hand (without a computer). However, they are also useful in a more common situation. The procedure you have learned to do in the chapter works (without modification) only if you have equal numbers of scores in each group, while the computational formulas below also work when there are unequal numbers of scores in each group. These formulas require that you first figure an intermediary for the two variance estimates, called "sum of squares," or SS for short. For the between-groups estimate, $S^2_{\text{Between}} = SS_{\text{Between}}/df_{\text{Between}}$. The formula for SS_{Between} is as follows:

$$SS_{\text{Between}} = \frac{(\Sigma X_1)^2}{n_1} + \frac{(\Sigma X_2)^2}{n_2} + \cdots + \frac{(\Sigma X_{\text{Last}})^2}{n_{\text{Last}}} - \frac{(\Sigma X)^2}{N} \quad (7)$$

For the within-groups estimate, $S^2_{\text{Within}} = SS_{\text{Within}}/df_{\text{Within}}$. The formula for SS_{Within} is as follows:

$$SS_{\text{Within}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N} - SS_{\text{Between}} \quad (8)$$

However, as usual, we urge you to use the definitional formulas we have presented in the chapter to work out the practice problems. The definitional formulas are closely related to the meaning of the procedures. Using the definitional formulas to work out the problems helps you learn the meaning of the analysis of variance.

Table 7 Steps for the Analysis of Variance (When Sample Sizes Are Equal)

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.**

❷ **Determine the characteristics of the comparison distribution.**

a. The comparison distribution is an F distribution.

b. The between-groups (numerator) degrees of freedom is the number of groups minus 1: $df_{\text{Between}} = N_{\text{Groups}} - 1$.

c. The within-groups (denominator) degrees of freedom is the sum of the degrees of freedom in each group (the number in the group minus 1): $df_{\text{Within}} = df_1 + df_2 + \ldots + df_{\text{Last}}$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.**

a. Decide the significance level.

b. Look up the appropriate cutoff in an F table, using the degrees of freedom from Step ❷.

❹ **Determine your sample's score on the comparison distribution.** This will be an F ratio.

a. Figure the between-groups population variance estimate (S^2_{Between}).

Figure the mean of each group.

Ⓐ **Estimate the variance of the distribution of means:** $S^2_M = [\Sigma(M - GM)^2]/df_{\text{Between}}$.

Ⓑ **Figure the estimated variance of the population of individual scores:** $S^2_{\text{Between}} = (S^2_M)(n)$.

b. Figure the within-groups population variance estimate (S^2_{Within}).

i. Figure population variance estimates based on each group's scores: For each group, $S^2 = [\Sigma(X - M)^2]/(n - 1)$.

ii. Average these variance estimates: $S^2_{\text{Within}} = (S^2_1 + S^2_2 + \cdots + S^2_{\text{Last}})/N_{\text{Groups}}$.

c. Figure the F ratio: $F = S^2_{\text{Between}}/S^2_{\text{Within}}$.

❺ **Decide whether to reject the null hypothesis:** Compare the scores from Steps ❸ and ❹.

Comparing Each Group to Each Other Group

When you reject the null hypothesis in an analysis of variance, this tells you that the population means are not all the same. What is not clear, however, is which population means differ from which. For example, in the criminal record study, the Criminal Record group jurors had the highest ratings for the defendant's guilt ($M = 8$); the No Information group jurors, the second highest ($M = 5$); and the Clean Record group jurors, the lowest ($M = 4$). From the analysis of variance results, we concluded that the true means of the three populations our samples represent are not all the same. (That is, the overall analysis of variance was significant.) However, we do not know which populations' means are *significantly different* from each other.

For this reason, researchers often do not stop after getting a significant result with an analysis of variance. Instead, they may go on to compare each population to each other population. For example, with three groups, you would compare group 1 to group 2, group 1 to group 3, and group 2 to group 3. You could do each of these comparisons using ordinary t tests for independent means. However, there is a problem with using ordinary t tests like this. The problem is that you are making three comparisons, each at the .05 level. The overall chance of at least one of them being significant just by chance is more like .15.

Some statisticians argue that it is all right to do three t tests in this situation because we have first checked that the overall analysis of variance is significant. These are called **protected *t* tests.** We are protected from making too big an error by the overall analysis of variance being significant. However, most statisticians believe that

protected *t* tests In analysis of variance, t tests among pairs of means after finding that the F for the overall difference among the means is significant.

the protected t test is not enough protection. Advanced statistics texts give procedures that provide even more protection. (Also, most standard statistics software, including SPSS, have options as part of the analysis of variance that provide various ways to compare means that provide strong protection of this kind. One widely used method is called *Tukey's HSD* test.)

How are you doing?

1. A study compares the effects of three experimental treatments, A, B, and C, by giving each treatment to 16 participants and then assessing their performance on a standard measure. The results on the standard measure are as follows: Treatment A group: $M = 20$, $S^2 = 8$; Treatment B group: $M = 22$, $S^2 = 9$; Treatment C group: $M = 18$, $S^2 = 7$. Use the steps of hypothesis testing (with the .01 significance level) to test whether the three experimental treatments create any difference among the populations these groups represent.
2. Give the two main assumptions for the analysis of variance.
3. Why do we need the equal variance assumption?
4. What is the general rule about when violations of the equal variance assumption are likely to lead to serious inaccuracies in results?
5. After getting a significant result with an analysis of variance, why do researchers usually go on to compare each population to each other population?

Answers

1. Steps of hypothesis testing:

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are three populations:

Population 1: People given experimental treatment A.
Population 2: People given experimental treatment B.
Population 3: People given experimental treatment C.

The null hypothesis is that these three populations have the same mean. The research hypothesis is that their means are not the same.

❷ **Determine the characteristics of the comparison distribution.** The comparison distribution will be an F distribution. Its degrees of freedom are figured as follows: The between-groups variance estimate is based on three groups, making 2 degrees of freedom. The within-groups estimate is based on 15 degrees of freedom (16 participants) in each of the three groups, making 45 degrees of freedom.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** Using Table 3 of appendix "Tables", the cutoff for numerator $df = 2$ and denominator $df = 45$, at the .01 significance level, is 5.11.

❹ **Determine your sample's score on the comparison distribution.**

a. Figure the between-groups population variance estimate (S^2_{Between}). First, figure the mean of each group. The group means are 20, 22, and 18.

Ⓐ **Estimate the variance of the distribution of means:** Add up the sample means' squared deviations from the grand mean and divide by the number of means minus 1:

$$S^2_M = [(20 - 20)^2 + (22 - 20)^2 + (18 - 20)^2]/(3 - 1)$$
$$= (0 + 4 + 4)/2 = 4.$$

Ⓑ **Figure the estimated variance of the population of individual scores:** Multiply the variance of the distribution of means by the number of scores in each group:

$$S^2_{\text{Between}} = (4)(16) = 64.$$

b. Figure the within-groups population variance estimate (S^2_{Within}):

Ⓐ Figure population variance estimates based on each group's scores. Treatment A group, $S^2 = 8$; Treatment B group, $S^2 = 9$; Treatment C group, $S^2 = 7$.

Ⓑ Average these variance estimates. $S^2_{\text{Within}} = (8 + 9 + 7)/3 = 8$.

The F ratio is the between-groups estimate divided by the within-groups estimate: $F = 64/8 = 8.00$.

⑤ **Decide whether to reject the null hypothesis.** The F of 8.00 is more extreme than the .01 cutoff of 5.11. Therefore, reject the null hypothesis. The research hypothesis is supported; the different experimental treatments do produce different effects on the standard performance measure.

2. The two main assumptions for the analysis of variance are that the populations are normally distributed and have equal variances.

3. We need the equal variance assumption in order to be able to justify averaging the estimates from each sample into an overall within-groups population variance estimate.

4. The analysis can lead to inaccurate results when the variance estimate from the group with the largest estimate is more than four or five times the smallest variance estimate.

5. After getting a significant result with an analysis of variance, researchers usually go on to compare each population to each other population to find out which populations' means are significantly different from each other.

Effect Size and Power for the Analysis of Variance

Effect Size

Effect size for the analysis of variance is a little more complex than for a t test. With the t test, you took the difference between the two means and divided by the pooled population standard deviation. In the analysis of variance, we have more than two means, so it is not obvious just what is the equivalent to the difference between the means—the numerator in figuring effect size. Thus, in this section we consider a quite different approach to effect size, the **proportion of variance accounted for (R^2).**

proportion of variance accounted for (R^2) Measure of effect size for analysis of variance.

To be precise, R^2 is the proportion of the total variation of scores from the grand mean that is accounted for by the variation between the means of the groups. (In other words, you consider how much of the variance in the measured variable—such as ratings of guilt—is accounted for by the variable that divides up the groups—such as what experimental condition one is in.) In terms of a formula,

$$R^2 = \frac{(S^2_{\text{Between}})(df_{\text{Between}})}{(S^2_{\text{Between}})(df_{\text{Between}}) + (S^2_{\text{Within}})(df_{\text{Within}})} \qquad (9)$$

The proportion of variance accounted for is the between-groups population variance estimate multiplied by the between-groups degrees of freedom, divided by the sum of the between-groups population variance estimate multiplied by the between-groups degrees of freedom, plus the within-groups population variance estimate multiplied by the within-groups degrees of freedom.

The between- and within-groups degrees of freedom are included in the formula to take into account the number of participants and the number of groups used in the study. But this formula really looks more complicated than it is. Basically the numerator is a function of the between-groups variance estimate and the denominator is a function of the total variance (between plus within).

Consider once again the criminal-record study. In that example, $S^2_{\text{Between}} = 21.70$, $df_{\text{Between}} = 2$, $S^2_{\text{Within}} = 5.33$, and $df_{\text{Within}} = 12$. Thus, the proportion of the total variation accounted for by the variation between groups is $(21.70)(2)/[(21.70)(2) + (5.33)(12)]$, which is .40 (or 40%). In terms of the formula,

$$R^2 = \frac{(S^2_{\text{Between}})(df_{\text{Between}})}{(S^2_{\text{Between}})(df_{\text{Between}}) + (S^2_{\text{Within}})(df_{\text{Within}})}$$

$$= \frac{(21.70)(2)}{(21.70)(2) + (5.33)(12)} = \frac{43.40}{107.36} = .40$$

The proportion of variance accounted for is a useful measure of effect size because it has the direct meaning suggested by its name. (Furthermore, researchers are familiar with R^2 and its square root R from their use in multiple regression and multiple correlation.) R^2 has a minimum of 0 and a maximum of 1. However, in practice it is rare in most behavioral and social sciences research for an analysis of variance to have an R^2 even as high as .20. Cohen's (1988) conventions for R^2 are .01, a small effect size; .06, a medium effect size; and .14, a large effect size. (You should also know that another common name for this measure of effect size often used in research articles, besides R^2, is η^2, the Greek letter eta squared.)

Power

Table 8 shows the approximate power for the .05 significance level for small, medium, and large effect sizes; sample sizes of 10, 20, 30, 40, 50, and 100 per group; and three, four, and five groups.[2]

Consider a planned study with five groups of 10 participants each and an expected large effect size (.14). Using the .05 significance level, the study would have a power of .56. Thus, even if the research hypothesis is in fact true and has a large effect size, there is only a little greater than even chance (56%) that the study will come out significant.

Determining power is especially useful when interpreting the practical implication of a nonsignificant result. For example, suppose that you have read a study using an analysis of variance with four groups of 30 participants each, and there is a nonsignificant result at the .05 level. Table 8 shows a power of only .13 for a small effect size. This suggests that even if such a small effect exists in the population, this study would be very unlikely to have come out significant. But the table shows a power of .96 for a large effect size. This suggests that if a large effect existed in the population, it almost surely would have shown up in that study.

Planning Sample Size

Table 9 gives the approximate number of participants you need in each group for 80% power at the .05 significance level for estimated small, medium, and large effect sizes for studies with three, four, and five groups.[3]

[2]More detailed tables are provided in Cohen (1988, pp. 289–354). When using these tables, note that the value of u at the top of each table refers to df_{between}, which for a one-way analysis of variance is the number of groups minus 1, not the number of groups as used in our Table 8.

[3]More detailed tables are provided in Cohen (1988, pp. 381–389). If you use these, see footnote 2 above.

Table 8 Approximate Power for Studies Using the Analysis of Variance Testing Hypotheses at the .05 Significance Level

	Effect Size		
Participants per Group (n)	**Small (R^2 = .01)**	**Medium (R^2 = .06)**	**Large (R^2 = .14)**
Three groups ($df_{\text{Between}} = 2$)			
10	.07	.20	.45
20	.09	.38	.78
30	.12	.55	.93
40	.15	.68	.98
50	.18	.79	.99
100	.32	.98	*
Four groups ($df_{\text{Between}} = 3$)			
10	.07	.21	.51
20	.10	.43	.85
30	.13	.61	.96
40	.16	.76	.99
50	.19	.85	*
100	.36	.99	*
Five groups ($df_{\text{Between}} = 4$)			
10	.07	.23	.56
20	.10	.47	.90
30	.13	.67	.98
40	.17	.81	*
50	.21	.90	*
100	.40	*	*

*Nearly 1.

For example, suppose you are planning a study involving four groups, and you expect a small effect size (and will use the .05 significance level). For 80% power, you would need 274 participants in each group, a total of 1,096 in all. However, suppose you could adjust the research plan so that it was now reasonable to predict a large effect size (perhaps by using more accurate measures and a stronger experimental procedure). Now you would need only 18 in each of the four groups, for a total of 72.

Table 9 Approximate Number of Participants Needed in Each Group (Assuming Equal Sample Sizes) for 80% Power for the One-Way Analysis of Variance Testing Hypotheses at the .05 Significance Level

	Effect Size		
	Small (R^2 = .01)	**Medium (R^2 = .06)**	**Large (R^2 = .14)**
Three groups ($df_{\text{Between}} = 2$)	322	52	21
Four groups ($df_{\text{Between}} = 3$)	274	45	18
Five groups ($df_{\text{Between}} = 4$)	240	39	16

How are you doing?

1. (a) Write the formula for effect size in analysis of variance; (b) define each of the symbols; and (c) figure the effect size for a study in which $S^2_{Between} = 12.22$, $S^2_{Within} = 7.20$, $df_{Between} = 2$, and $df_{Within} = 8$.
2. What is the power of a study with four groups of 40 participants each to be tested at the .05 significance level, in which the researchers predict a large effect size?
3. About how many participants do you need in each group for 80% power in a planned study with five groups in which you predict a medium effect size and will be using the .05 significance level?

Answers

1. (a) The formula for effect size in analysis of variance: $R^2 = (S^2_{Between})(df_{Between})/[(S^2_{Between})(df_{Between}) + (S^2_{Within})(df_{Within})]$. (b) R^2 is the proportion of variance accounted for; $S^2_{Between}$ is the between-groups population variance estimate; $df_{Between}$ is the between-groups degrees of freedom (number of groups minus 1); S^2_{Within} is the within-groups population variance estimate; df_{Within} is the within-groups degrees of freedom (the sum of the degrees of freedom for each group's population variance estimate). (c) $R^2 = (12.22)(2)/[(12.22)(2) + (7.20)(8)] = 24.44/(24.44 + 57.6) = 24.44/82.04 = .30$.
2. The power of the study is .99.
3. The number of participants needed in each group is 39.

Factorial Analysis of Variance

Factorial analysis of variance is an extension of the procedures you just learned. It is a wonderfully flexible and efficient approach that handles many types of experimental studies. The actual figuring of a factorial analysis of variance is beyond what we can cover in an introductory book. Our goal in this section is to help you understand the basic approach and the terminology so that you can make sense of research articles that use it.

We introduce factorial analysis of variance with an example. E. Aron and A. Aron (1997) proposed that people differ on a basic, inherited tendency they call "sensory-processing sensitivity" that is found in about a fourth of humans and in many animal species. People with this trait process information especially thoroughly. Table 10 gives six items from their Highly Sensitive Person Scale. If you score high on most of the items, you are probably among the 25% of people who are "highly sensitive."

factorial analysis of variance Analysis of variance for a factorial research design.

Table 10 Selected Items from the Highly Sensitive Person Scale

1. Do you find it unpleasant to have a lot going on at once?
2. Do you find yourself wanting to withdraw during busy days, into bed or into a darkened room or any place where you can have some privacy and relief from stimulation?
3. Are you easily overwhelmed by things like bright lights, strong smells, coarse fabrics, or sirens close by?
4. Do you get rattled when you have a lot to do in a short amount of time?
5. Do changes in your life shake you up?
6. Are you bothered by intense stimuli, like loud noises and chaotic scenes?

Note: Each item is answered on a scale from 1 "Not at all" to 7 "Extremely."
Source: Aron, E., & Aron, A. (1997). Sensory-processing sensitivity and its relation to introversion and emotionality. *Journal of Personality and Social Psychology, 73,* 345–368. Copyright © 1997, American Psychological Association. Reproduced with permission. The use of APA information does not imply endorsement by APA.

One implication of being a highly sensitive person, according to this model, is that such people are more affected than others by success and failure. This is because they process all experiences more completely. So one would expect highly sensitive people to feel especially good when they succeed and especially bad when they fail. To test this prediction, E. Aron, A. Aron, and Davies (2005, Study 4) conducted an experiment. In the experiment, students first completed the sensitivity questions in Table 10. (This permitted the researchers when analyzing the results to divide them into highly sensitive and not highly sensitive.) Later, as part of a series of tests on attitudes and other topics, everyone was given a timed test of their "applied reasoning ability," something important to most students. But without their knowing it, half took an extremely easy version and half took a version so difficult that some problems had no right answer. The questionnaires were handed out to people in alternate seats (it was randomly determined where this process began in each row); so if you had the hard version of the test, the people on either side had an easy version and you were probably aware that they finished quickly while you were still struggling. On the other hand, if you had the easy version, you were probably aware that you had finished easily while the others around you were still working on what you presumed to be the same test.

Right after the test, everyone was asked some items about their mood, such as how depressed, anxious, and sad they felt at the moment. (The mood items were buried in other questions.) Responses to the mood items were averaged to create a measure of overall negative mood.

In sum, the study looked at the effect of *two different factors* on negative mood: (1) whether a person was highly sensitive or not highly sensitive and (2) whether a person had taken the easy test (which caused them to feel they had succeeded) or the hard test (which caused them to feel they had failed).

Aron and colleagues could have done two studies, one comparing highly sensitive versus not highly sensitive individuals and one comparing people who took the easy test versus people who took the hard test. Instead, they studied the effects of both sensitivity and test difficulty in a single study. With this setup there were four groups of participants (see Table 11): (1) those who are not highly sensitive and took the easy test, (2) those who are highly sensitive and took the easy test, (3) those who are not highly sensitive and took the hard test, and (4) those who are highly sensitive and took the hard test.

Table 11 Factorial Research Design Employed by E. Aron and Colleagues (2005)

		Test Difficulty	
		Easy	Hard
Sensitivity	Not High	a	c
	High	b	d

Factorial Research Design Defined

The E. Aron and colleagues (2005) study is an example of a **factorial research design.** In a factorial research design the effects of *two or more variables* are examined at once by making groupings of every combination of the variables. In this example, there are two levels of sensitivity (not high and high) and two levels of test difficulty (easy and hard). This creates four possible group combinations, and the researchers used all of them in their study.

A factorial research design has a major advantage over conducting separate studies of each variable: efficiency. With a factorial design, you can study both variables at once, without needing twice as many participants. In the example, Aron and colleagues were able to use a single group of participants to study the effects of sensitivity and test difficulty on negative mood.

factorial research design Way of organizing a study in which the influence of two or more variables is studied at once by setting up the situation so that a different group of people are tested for each combination of the levels of the variables.

Interaction Effects

There is an even more important advantage of a factorial research design. A factorial design lets you study the effects of *combining two or more variables.* In this example, sensitivity and test difficulty might affect negative mood in a simple additive way.

By additive, we mean that their combined influence is the sum of their separate influences; if you are more of one and also more of the other, then the overall effect is the total of the two individual effects. For example, suppose being highly sensitive makes you more likely to experience a negative mood; similarly, suppose the test being hard makes you more likely to experience a negative mood. If these two effects are additive, then participants in the high sensitivity, hard test group will be most likely to experience a negative mood; participants who are not highly sensitive and take the easy test will be the least likely to experience a negative mood; and those in the other two conditions would have an intermediate likelihood of experiencing a negative mood.

It could also be that one variable but not the other has an effect. Or perhaps neither variable has any effect. In the additive situation, or when only one variable or neither has an effect, looking at the two variables in combination does not give any interesting additional information.

However, it is also possible that the *combination of the two variables* changes the result. In fact, as noted earlier, Aron and colleagues (2005) predicted that the effect of being highly sensitive would be especially strong in the hard test condition. A situation where the *combination* of variables has a special effect is called an **interaction effect.** An interaction effect is an effect in which the effect of one variable (that divides the groups) on the measured variable is different across the levels of the other variable that divides the groups.

In the Aron and colleagues (2005) study, there was an interaction effect. Look at Table 12. The result was that the students in the High Sensitivity/Hard Test group had the most negative mood (remember, a high number means a more negative mood), and the students in the High Sensitivity/Easy Test group had the least negative mood. The level of negative mood in the other two groups (the Not High Sensitivity groups) was similar and between that of the two High Sensitivity groups. These results show that the effect of test difficulty (on negative mood) is different according to the level of sensitivity: For students who are not highly sensitive, their level of negative mood is slightly higher with a hard test (2.56) compared to an easy test (2.43); but for students who are highly sensitive, their level of negative mood is much higher with a hard test (3.01) compared to an easy test (2.19).

Table 12 Mean Levels of Negative Mood in the E. Aron and Colleagues (2005) Study

		Test Difficulty	
		Easy	Hard
Sensitivity	Not High	2.43	2.56
	High	2.19	3.01

Suppose the researchers had studied sensitivity and test difficulty in two separate studies. In the study of sensitivity (assuming equal numbers of students in each group), they would have concluded that sensitivity had little (if any) effect on negative mood: The average level of negative mood for students who are not highly sensitive is 2.50 (that is, the average of 2.43 and 2.56) and for highly sensitive students 2.60 (the average of 2.19 and 3.01). In the study of test difficulty (again assuming equal numbers of students in each group), they would have concluded that students experienced a lower level of negative mood when taking an easy test than a hard test: The average level of negative mood for students taking the easy test was 2.31 (the average of 2.43 and 2.19), and the average for those taking the hard test was 2.79 (the average of 2.56 and 3.01). Thus, following the approach of two separate studies, the researchers would have completely missed the most important result. The most important result had to do with the *combination of the two factors.*

Some Terminology

The Aron et al. study would be analyzed with what is called a **two-way analysis of variance** (it uses a **two-way factorial research design**) because it considers the effect of two variables that separate groups. These variables are called **grouping variables,** also known as **independent variables.** By contrast, the situations we considered earlier (such as the attachment style study or the criminal record study) were examples of studies analyzed using a **one-way analysis of variance.** Such studies are called *one-way*

interaction effect Situation in a factorial analysis of variance in which the combination of variables has an effect that could not be predicted from the effects of the two variables individually.

two-way analysis of variance Analysis of variance for a two-way factorial research design.

two-way factorial research design Factorial design with two variables that each divide the groups.

grouping variable A variable that separates groups in analysis of variance.

independent variable Variable considered to be a cause, such as what group a person is in for an analysis of variance.

one-way analysis of variance Analysis of variance in which there is only one grouping variable (as distinguished from a factorial analysis of variance).

because they consider the effect of only one grouping (or independent) variable (such as a person's attachment style or information about a defendant's criminal record).

In a two-way analysis, each grouping (or independent) variable or "way" (each dimension in the diagram) is a possible **main effect.** If the result for a variable, averaging across the other grouping variable, is significant, it is a *main effect.* This is entirely different from an *interaction effect,* which is based on the *combination of grouping variables.* In the two-way Aron and colleagues (2005) study, there was a possibility of two main effects and one interaction effect. The two possible main effects are one for sensitivity and one for test difficulty. The possible interaction effect is for the combination of sensitivity and test difficulty. In a two-way analysis of variance you are always testing two possible main effects and one possible interaction effect.

Each grouping combination in a factorial design is called a **cell.** The mean of the scores in each cell is a **cell mean.** In the Aron and colleagues (2005) study, there are four cells and thus four cell means, one for each combination of the levels of sensitivity and test difficulty. That is, one cell is Not High Sensitivity and an Easy Test (as shown in Table 12, its cell mean is 2.43); one cell is High Sensitivity and an Easy Test (2.19); one cell is Not High Sensitivity and a Hard Test (2.56); and one cell is High Sensitivity and a Hard Test (3.01).

The means of one grouping variable alone are called **marginal means.** For example, in the Aron and colleagues (2005) study there are two marginal means for each grouping variable. For sensitivity, there is one marginal mean for all the students who are not highly sensitive (as we saw earlier, 2.50) and one for all the students who are highly sensitive (2.60). For the test difficulty grouping variable, there is one marginal mean for all the students who took the easy test (2.31) and one marginal mean for all the students who took the hard test (2.79). (Because we were mainly interested in the interaction, we did not show these marginal means in our tables.) To look at a main effect, you focus on the marginal means for each grouping variable. To look at the interaction effect, you focus on the pattern of individual cell means.

The individual cell means and marginal means are means of the variable that is measured in the research study. In the Aron and colleagues (2005) study, it was the participants' reports of their level of negative mood. A variable like this, which is measured and represents the effect of the experimental procedure, is called a **dependent variable.** It is *dependent* in the sense that any participant's score on this variable depends on what happens in the experiment.

main effect Difference between groups on one grouping variable in a factorial analysis of variance; result for a variable that divides the groups, averaging across the levels of the other variable that divides the groups.

cell In a factorial research design, a particular combination of levels of the variables that divide the groups.

cell mean Mean of a particular combination of levels of the variables that divide the groups in a factorial design in analysis of variance.

marginal mean In a factorial design, mean score for all the participants at a particular level of one of the variables; often shortened to *marginal.*

dependent variable Variable considered to be an effect.

How are you doing?

1. (a) What is a factorial research design? (b) and (c) Give two advantages of a factorial research design over doing two separate experiments.
2. In a factorial research design, (a) what is a main effect, and (b) what is an interaction effect?
3. In the following table are the means from a study in which participants rated the originality of paintings under various conditions. For each mean, indicate its grouping and whether it is a cell or marginal mean.

	Contemporary	Renaissance	Overall
Landscape	6.5	5.5	6
Portrait	3.5	2.5	3
Overall	5	4	

4. In each of the following studies, participants' performance on a coordination task was measured under various conditions or compared for different groups. For each study, make a diagram of the research design and indicate whether it is a one-way or two-way design: (a) a study in which people are assigned to either a high-stress condition or a low-stress condition, and within each of these conditions, half are assigned to work alone and half to work in a room with other people; (b) a study comparing students majoring in physics, chemistry, or engineering; (c) a study comparing people doing a task in a hot room versus a cold room, with half in each room doing the task with their right hand and half with their left hand. (Your diagrams for a two-way analysis should look something like the diagram in Table 12, but without the means written in; diagrams for a one-way analysis would have just one row with the categories across the top.)

Answers

1. (a) A factorial research design is a research design in which the effect of two or more variables is examined at once by making groupings of every combination of the variables.
(b) A factorial research design is more efficient. For example, you can study the effects of two grouping (or independent) variables at once with only a single group of participants.
(c) A factorial research design makes it possible to see if there are interaction effects.

2. (a) A main effect in a factorial research design is the effect of one of the grouping (or independent) variables ignoring the pattern of results on the other grouping (or independent) variable.
(b) An interaction effect is the different effect of one grouping (or independent) variable according to the level of the other grouping (or independent) variable.

3. 6.5 = cell mean for Contemporary/Landscape group; 5.5 = cell mean for Renaissance/Landscape group; 6 = marginal mean for Landscape groups; 3.5 = cell mean for Contemporary/Portrait group; 2.5 = cell mean for Renaissance/Portrait group; 3 = marginal mean for Portrait groups; 5 = marginal mean for Contemporary groups; and 4 = marginal mean for Renaissance groups.

4. (a) Two-way.

	Stress	
Task	**High**	**Low**
Alone		
With Others		

(b) One-way.

Subject		
Physics	**Chemistry**	**Engineering**

(c) Two-way.

Recognizing and Interpreting Interaction Effects

It is very important to understand interaction effects. In many experiments and other kinds of research (such as the relation between genetics and environment) the interaction effect is the main point of the research. As we have seen, an interaction effect is an effect in which the impact of one grouping variable depends on the level of another grouping variable. You can think out and describe an interaction effect in three ways: in words, in numbers, or in a graph. Note that, in discussing the examples in this section on interaction effects, we will treat all differences that have the pattern of an interaction effect or of a main effect as if they were statistically significant. (In reality, you would carry out hypothesis testing steps to test whether the particular patterns were strong enough to be statistically *significant.*) We are taking this approach here to keep the focus on the *idea* of interaction effects while you are learning this fairly abstract notion.

Identifying Interaction Effects in Words and Numbers

You can think out an interaction effect in words by saying that you have an interaction effect when the effect of one grouping variable varies according to the level of another grouping variable. In the Aron and colleagues (2005) example, the effect of test difficulty (easy vs. hard) varies according to the level of sensitivity (not high vs. high). Another way of saying this is that the effect of test difficulty depends on the level of sensitivity (the effect of test difficulty is different for highly sensitive individuals than for individuals who are not highly sensitive). (You can also talk about this interaction effect by focusing on the other grouping variable first. So you could say that the effect of sensitivity varies according to the level of test difficulty, or that the effect of sensitivity depends on the level of test difficulty. Interaction effects are completely symmetrical in that you can describe them from the point of view of either grouping variable.)

You can see an interaction effect numerically by looking at the pattern of cell means. If there is an interaction effect, the *pattern of differences in cell means* across one row will not be the same as the patterns of differences in cell means across another row. (Again, all of this is symmetrical: You can also look at the differences in cell means for one column compared to another column, but we focus here on rows.) Consider the Aron and colleagues (2005) example. In the Not High Sensitivity row, the cell mean for negative mood of the Easy Test students (2.43) was just slightly lower than the cell mean for negative mood of the Hard Test students (2.56). This is a difference of $-.13$ (that is, $2.43 - 2.56 = -.13$). However, now look at the High Sensitivity row. In this row, the cell mean for Easy Test students' negative mood (2.19) was a lot lower than the cell mean for negative mood for the Hard Test students (3.01). This difference of $-.82$ is not at all the same as the difference of $-.13$ in the Not High Sensitivity row; this *dissimilar pattern of differences in cell means* indicates an interaction effect.

Table 13 gives cell means and marginal means for six possible results of a fictional two-way factorial study on the relation of age and education (the grouping

Table 13 Possible Means for Results of a Study of the Relation of Age and Education to Income (in Thousands of Dollars)

	Result A			Result B			Result C		
Age	***High School***	***University***	***Overall***	***High School***	***University***	***Overall***	***High School***	***University***	***Overall***
Younger	50	50	50	70	50	60	30	70	50
Older	50	70	60	50	70	60	50	90	70
Overall	50	60		60	60		40	80	

	Result D			Result E			Result F		
Age	***High School***	***University***	***Overall***	***High School***	***University***	***Overall***	***High School***	***University***	***Overall***
Younger	30	30	30	50	70	60	40	55	47.5
Older	130	130	130	50	90	70	45	70	57.5
Overall	80	80		50	80		42.5	62.5	

or independent variables) to income (the dependent variable). The grouping variable age has two levels (younger, such as 25 to 34, vs. older, such as 35 to 44), and the grouping variable education has two levels (high school vs. university). These fictional results are exaggerated to make clear when there are interactions and main effects. Before you look at the six possible results, take a minute to think about what kind of results you might expect (and hope!) to see. For example, do you expect that people with a university education will earn less than or more than people with only a high school education? Would you expect younger people to earn more or less than older people? Most importantly (since we are focusing on interaction effects), what about the possibility of an interaction effect? Do you think that the effect of education (university vs. high school) will be different according to age (younger vs. older)? Let's take a look at six possible results and then we'll tell you what the results of actual research show.

Result A: Interaction. Note that in the Younger row, education makes no difference, but in the Older row, the university cell mean is much higher than the high-school cell mean. One way to say this is that for the younger group, education is unrelated to income; but for the older group, people with a university education earn much more than those with less education. There are also two main effects: Overall, older people earn more than younger people, and overall, people with a university education earn more than those with only a high school education.

Result B: Interaction. This is because in the Younger row the high-school mean income is higher than the university mean income, but in the Older row the high-school mean income is lower. Put in words, among younger people, those with only a high-school education make more money (perhaps because they entered the workplace earlier or the kinds of jobs they have start out at a higher level); but among older people, those with a university education make more money. (There are no main effects in Result B, since the marginal means for the two rows are the same and marginal means for the two columns are the same.)

Result C: No interaction. In the Younger row, the high-school mean is 40 lower than the university mean, and the same is true in the Older row. In other words, whether young or old, people with university educations earn $40,000 more. (That is, there is a main effect for education, and also a main effect for age.)

TIP FOR SUCCESS

Remember, you can tell whether there is an interaction effect by looking at the pattern of cell means. There is an interaction effect if the pattern of cell means in one row is different from the pattern of cell means across another row.

TIP FOR SUCCESS

You can tell whether there is a main effect by looking at the marginal means. The marginal means in Table 13 are the means in the "Overall" columns and rows. There is a main effect for a particular grouping variable if the marginal means for the different levels of that grouping variable are not the same.

Result D: No interaction. There is no difference in the pattern of income between the two rows. Regardless of education, older people earn $100,000 more. (That is, there is a main effect for age; but there is no main effect for education.)

Result E: Interaction. In the Younger row, the university mean is 20 higher, but in the Older row, the university mean is 40 higher than the high-school mean. So among young people, university-educated people earn a little more; but, among older people, those with a university education earn much more. (There are also main effects for both age and education.)

Result F: Interaction. There is a smaller difference between the high school and university mean in the Younger row than in the Older row. As with Result E, for people with a university education, income increases more with age than it does for those with only a high-school education.[4]

Identifying Interaction Effects Graphically

Another common way of making sense of interaction effects is by graphing the pattern of cell means. This is usually done with a bar graph, although a line graph is sometimes used. The bar graph in Figure 4 shows the results from

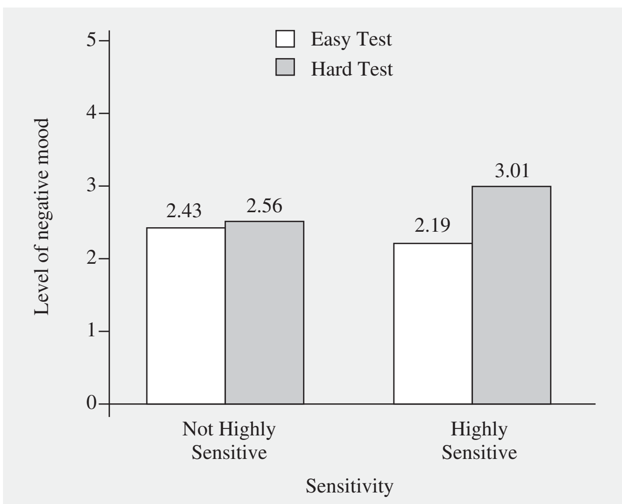


Figure 4 Bar graph for level of negative mood as a function of sensitivity (not high vs. high) and test difficulty (easy vs. hard).
Source: Based on Aron, E., Aron, A., & Davies, K. M. (2005). Adult shyness: The interaction of temperamental sensitivity and an adverse childhood environment. *Personality and Social Psychology Bulletin, 31*, No. 2, 181–197. Copyright © 2005, Society for Personality and Social Psychology, Inc. Reprinted by permission of SAGE Publications.

[4]Based on 1997–1999 statistics from the U.S. Census Bureau (the most recent time period for which statistics are available), the actual situation in the United States is closest to Result F (Day & Newberger, 2002). People with a university education earn more than those with only a high-school education in both age groups, but the difference is somewhat greater for the older group. (You may be interested to know that, based on data from 2008, individuals with a college degree in the United States earn an average of $26,000 more per year than those with only a high school education [U.S. Census Bureau, 2008]. However, it is important to keep in mind that whether people receive a college education is also related to the social class of their parents and other factors that may affect income as much or more than education does.)

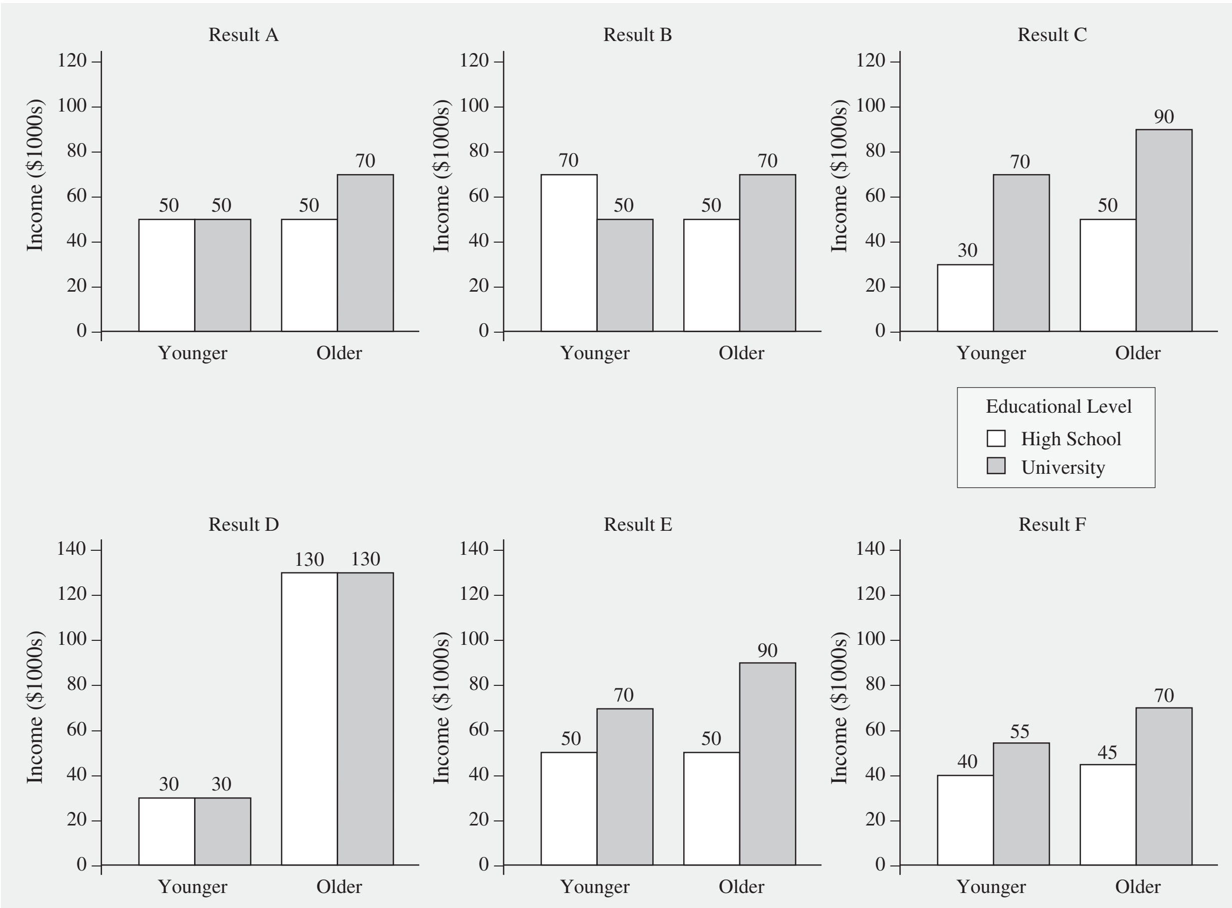


Figure 5 Bar graphs of fictional results in Table 12.

Aron and colleagues' (2005) study. The graphs in Figure 5 show the graphs for the fictional age and education results we just considered (the ones shown in Table 13).

One thing to notice about such graphs is this: Whenever there is an interaction, the pattern of bars on one section of the graph is different from the pattern on the other section of the graph. Thus, in Figure 4, the pattern for not highly sensitive is a small step up, but the pattern for highly sensitive is a much larger step up. The bars having a different pattern is just a graphic way of saying that the pattern of differences between the cell means from row to row is not the same.

Consider Figure 5, based on the age and education example. First, look at Results C and D. In Result C the younger and older sets of bars have the same pattern; both step up by 40. In Result D, both are flat. Within both Results C and D, the younger bars and the older bars have the same pattern. These were the examples that *did not have interactions.* All the other results, which did have interactions, have patterns of bars that are not the same for younger and older age groups. For example, in Result A, the two younger bars are flat but the older bars show a step up. In Result B, the younger bars show a step down from high school to university, but the older bars show a step up from high school to university. In

Results E and F, both younger and older bars show a step up, but the younger bars show a smaller step up than the older bars.

You can also see *main effects* from these graphs. In Figure 5, a main effect for age would be shown by the bars for younger being overall higher or lower than the bars for older. For example, in Result C, the bars for older are clearly higher than the bars for younger. What about the main effect for the bars that are not grouped together—university versus high school in this example? With these bars, you have to see whether the overall step pattern goes up or down. For example, in Result C, there is also a main effect for education, because the general pattern of the bars for high school to university goes up, and it does this for both the younger and older bars. Result D shows a main effect for age (the older bars are higher than the younger bars). But Result D does not show a main effect for education; the pattern is flat for both the older and younger bars.

Relation of Interaction and Main Effects

A study can have any combination of main and interaction effects can be significant. For example, they may all be significant, as in the pattern in Result F of Table 13 (or Figure 5). In this result, as you saw, older students earn more (a main effect for age), university students earn more (a main effect for level of education), and how much more university students earn depends on age (the interaction effect).

There can also be an interaction effect with no main effects. Result B of Table 13 is an example. The average level of income is the same for younger and older (no main effect for age), and it is the same for university and high school (no main effect for level of education).

There can also be one main effect significant along with an interaction, one main effect significant by itself, or for there to be no significant main or interaction effects. See how many of these possibilities you can find in the fictional results in Table 13 (or Figure 5).

When there is no interaction, a main effect has a straightforward meaning. However, when there is an interaction along with a main effect, things are more complicated. This is because the main effect may be created by just one of the levels of the variable. For example, in Result A the main effect for education (50 vs. 60) is entirely due to the difference for older people. It would be misleading to say that education makes a difference without noting that it really only matters when you are older.

On the other hand, even when there is an interaction, sometimes the main effect clearly holds up over and above the interaction. For example, in Result F in the age and education example, it seems clear that the main effect for age holds up over and above the interaction. That is, it is true for both people with and without a university education that older people earn more. (There is still an interaction, of course, because how much more older people earn depends on education.)

Extensions and Special Cases of the Factorial Analysis of Variance

Analysis of variance is very versatile. Factorial designs can be extended into three-way and higher designs (in which the effects of three or more grouping variables are examined). There are also procedures to handle situations in which the same participants are tested more than once. This is like the t test for dependent means but works with more than two testings. It is called a *repeated measures analysis of variance.*

How are you doing?

Questions 1–3 are based on the following results for a fictional study of the effects of vividness and length of examples (the grouping variables) on number of examples recalled (the dependent variable).

Example Length	Vividness		
	Low	***High***	***Overall***
Short	5	7	6
Long	3	1	2
Overall	4	4	

1. Describe the pattern of results in words.
2. Explain the pattern in terms of numbers.
3. Make two bar graphs of these results: One graph should show vividness (low and high) on the horizontal axis, with bars for short and long example length; another graph should show example length (short and long) on the horizontal axis, with bars for low and high vividness.
4. For a two-way factorial design, what are the possible combinations of main and interaction effects?
5. When there is both a main and an interaction effect, (a) under what conditions must you be careful in interpreting the main effect, and (b) under what conditions can you still be confident in the overall main effect?

Answers

1. There is a main effect in which short examples are recalled better, there is no main effect for vividness, and there is an interaction effect in which there is a bigger advantage of short over long examples when they are highly vivid.

2. The main effect is that on the average people recall six short examples but only two long examples. There is no main effect for vividness because people on the average recall four examples, regardless of how vivid the examples are. The interaction effect is that for short examples, people recall two more highly vivid than low vivid; but for long examples, they recall two fewer highly vivid than low vivid.

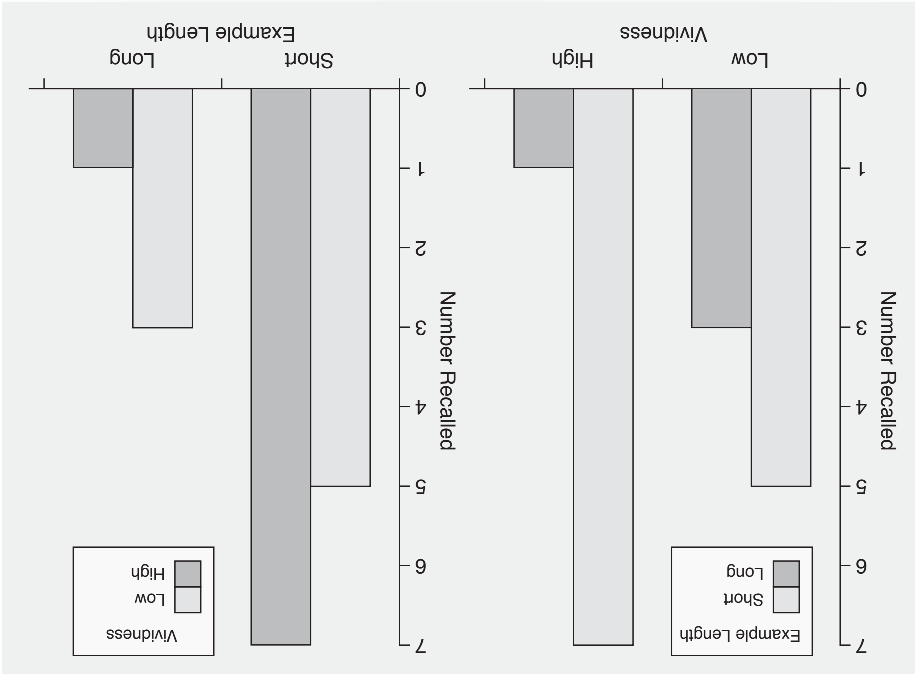
3. See Figure 6.

4. All possible combinations: no main or interaction effects, either main effect only, the interaction only, both main effects but no interaction effect, an interaction effect with either main effect, or an interaction effect with both main effects.

5. (a) You should be careful in interpreting the main effect when it is found for only one level of the other grouping variable.

(b) You can still be confident in the overall main effect when it holds at each level of the other grouping variable.

Figure 6 Answer to "How are you doing?" question 3.



Analyses of Variance in Research Articles

A one-way analysis of variance is usually described in a research article by giving the F, the degrees of freedom, and the significance level. For example, $F(3, 68) = 5.21$, $p < .01$. The means for the groups usually are given in a table. However, if there are only a few groups and only one or a few measures, the means may be given in the regular text of the article. Returning to the criminal-record study example, we could describe the analysis of variance results this way: "The means for the Criminal Record, Clean Record, and No Information conditions were 7.0, 4.0, and 5.0, respectively. These were significantly different, $F(2, 12) = 4.07$, $p < .05$."

In a factorial analysis of variance, researchers usually give a description in the text plus a table or figure. The text gives the F ratio and the information that goes with it for each main and interaction effect. The table gives the cell means and sometimes also the marginal means. For example, Aron and colleagues (2005) described the result we used for our example as follows: "As hypothesized . . . the interaction of sensitivity and condition [easy vs. hard test difficulty] predicted state negative affect [mood] . . . , $t(155) = 1.72$, $p < .05$." We showed a graph based on this result in Figure 4.

Here is another example. Shao and Skarlicki (2009) conducted a study to examine whether there was an interaction effect of mindfulness and gender (the two grouping or independent variables) on academic performance (the measured or dependent variable) among students in an MBA (Master's of Business Administration) program. Mindfulness can be defined as a person's tendency to be "attentive to and aware of what is taking place in the present" (Brown & Ryan, 2003, p. 822) and the researchers measured it using a 15-item survey. Based on previous neuroscience research on mindfulness and gender differences in activation of the left and right brain hemispheres, the

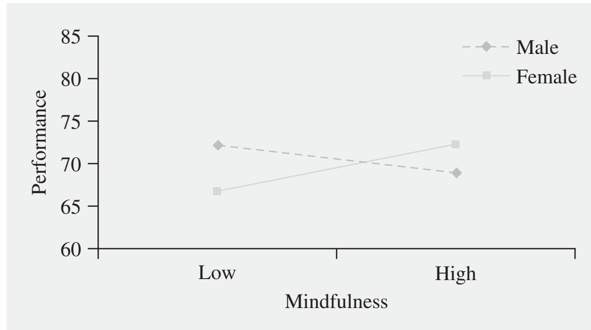


Figure 7 Two-way interaction of mindfulness and gender on performance.
Source: Shao, R., & Skarlicki, D. P. (2009). The role of mindfulness in predicting individual performance. *Canadian Journal of Behavioral Science, 41,* 195–201. Copyright © 2009 Canadian Psychological Association. Reprinted by permission.

researchers hypothesized that women would show a stronger positive association between mindfulness and academic performance than men. The results of the study are shown in the line graph in Figure 7. As you can see in the figure, the results supported the researchers' hypothesis: Women with high mindfulness had better academic performance than women with low mindfulness, but that was not the case among men (and in fact, the researchers noted that mindfulness was not significantly associated with performance among men). In a line graph, such as Figure 7, you can tell there is an interaction by the lines not being parallel. As with a bar graph, this shows that there is a different pattern of cell means from row to row.

Learning Aids

1. The analysis of variance (ANOVA) is used to test hypotheses based on differences among means of more than two samples. The procedure compares two estimates of population variance. One, the within-groups estimate, is figured by averaging the population variance estimates from each of the samples. The other, the between-groups estimate, is based on the variation among the means of the samples.
2. The F ratio is the between-groups estimate divided by the within-groups estimate. The null hypothesis is that all the samples come from populations with the same mean. If the null hypothesis is true, the F ratio should be about 1. This is because the two population variance estimates are based on the same thing, the variation within each of the populations (due to chance factors). If the research hypothesis is true, so that the samples come from populations with different means, the F ratio should be larger than 1. This is because the between-groups estimate is now influenced by the variation both within the populations (due to chance factors) and among them (due to a treatment effect). But the within-groups estimate is still affected only by the variation within each of the populations.
3. When the samples are of equal size, the within-groups population variance estimate is the ordinary average of the estimates of the population variance figured from each sample. The between-groups population variance estimate is done in two steps. First, you estimate the variance of the distribution of means based on

the means of your samples. (This is figured with the usual formula for estimating population variance from sample scores.) Second, you multiply this estimate by the number of participants in each group. This step takes you from the variance of the distribution of means to the variance of the distribution of individual scores.

4. The distribution of F ratios when the null hypothesis is true is a mathematically defined distribution that is skewed to the right. Significance cutoffs are given on an F table according to the degrees of freedom for each population variance estimate: the between-groups (numerator) degrees of freedom is the number of groups minus 1 and the within-groups (denominator) degrees of freedom is the sum of the degrees of freedom within all samples.
5. The assumptions for the analysis of variance are the same as for the t test. The populations must be normally distributed, with equal variances. Like the t test, the analysis of variance is adequately accurate even when there are moderate violations of these assumptions.
6. The proportion of variance accounted for (R^2) (also called eta squared, η^2), is a measure of analysis of variance effect size. R^2 represents how much of the variance in the measured variable is accounted for by the variable that divides up the groups. Power depends on effect size, number of people in the study, significance level, and number of groups.
7. In a factorial research design, participants are put into groupings according to the combinations of the variables whose effects are being studied. In such designs, you can study the effects of two grouping (or independent) variables on a dependent (or measured) variable without needing twice as many participants. Also, such designs allow you to study the effects of combinations of the two grouping variables.
8. An interaction effect is when the impact of one grouping variable varies according to the level of the other grouping variable. A main effect is the impact of one grouping variable, ignoring the effect of the other grouping variable. Interaction effects and main effects can be described verbally, numerically, and graphically (usually on a graph with bars for each combination of the grouping variables, with the height of the bar being the score on the measured or dependent variable).
9. Analysis of variance results are reported in a standard fashion in research articles, such as $F(3, 68) = 5.21$, $p < .01$. In a factorial analysis of variance, there is usually a description in the text plus a table. A graph may also be used to show any significant interactions.

analysis of variance (ANOVA)
within-groups estimate of the population variance (S^2_{Within})
between-groups estimate of the population variance (S^2_{Between})
F ratio
F distribution
F table
grand mean (GM)
between-groups degrees of freedom (df_{Between})
within-groups degrees of freedom (df_{Within})
protected t tests
proportion of variance accounted for (R^2)
factorial analysis of variance
factorial research design
interaction effect
two-way analysis of variance
two-way factorial research design
grouping variable
independent variable
one-way analysis of variance
main effect
cell
cell mean
marginal means
dependent variable

Example Worked-Out Problems

Overall Analysis of Variance

An experiment compares the effects of four treatments, giving each treatment to 20 participants and then assessing their performance on a standard measure. The results on the standard measure are as follows. Treatment 1: $M = 15$, $S^2 = 20$; Treatment 2: $M = 12$, $S^2 = 25$; Treatment 3: $M = 18$, $S^2 = 14$; Treatment 4: $M = 15$, $S^2 = 27$. Use the five steps of hypothesis testing (and the .05 significance level) to determine whether treatment matters.

Answer

Steps of hypothesis testing:

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are four populations:

Population 1: People given experimental treatment 1.
Population 2: People given experimental treatment 2.
Population 3: People given experimental treatment 3.
Population 4: People given experimental treatment 4.

The null hypothesis is that these four populations have the same mean. The research hypothesis is that the four population means are not the same.

❷ **Determine the characteristics of the comparison distribution.** The comparison distribution will be an F distribution.

$$df_{\text{Between}} = N_{\text{Groups}} - 1 = 4 - 1 = 3;$$

$$df_{\text{Within}} = df_1 + df_2 + \ldots + df_{\text{Last}} = 19 + 19 + 19 + 19 = 76.$$

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** Using Table 3 of the appendix "Tables" for $df = 3, 75$ (the closest below 3, 76) at the .05 level, the cutoff F is 2.73.

❹ **Determine your sample's score on the comparison distribution.**

a. Figure the between-groups population variance estimate (S^2_{Between}): Figure the mean of each group. The group means are 15, 12, 18, and 15.

Ⓐ **Estimate the variance of the distribution of means:** Sum the sample means' squared deviations from the grand mean and divide by the number of means minus 1:

$$GM = (15 + 12 + 18 + 15)/4 = 15.$$

$$\begin{aligned} S^2_M &= \Sigma(M - GM)^2/df_{\text{Between}} \\ &= [(15 - 15)^2 + (12 - 15)^2 + (18 - 15)^2 + (15 - 15)^2]/(4 - 1) \\ &= (0 + 9 + 9 + 0)/3 = 18/3 = 6. \end{aligned}$$

Ⓑ **Figure the estimated variance of the population of individual scores:** Multiply the variance of the distribution of means by the number of scores in each group.

$$S^2_{\text{Between}} = (S^2_M)(n) = (6)(20) = 120.$$

b. Figure the within-groups population variance estimate (S^2_{Within}).

i. Figure population variance estimates based on each group's scores. Treatment 1 group, $S^2 = 20$; Treatment 2 group, $S^2 = 25$; Treatment 3 group, $S^2 = 14$; Treatment 4 group, $S^2 = 27$.
ii. Average these variance estimates.

$$S^2_{\text{Within}} = (20 + 25 + 14 + 27)/4 = 86/4 = 21.5.$$

c. Figure the F ratio:

$$F = S^2_{\text{Between}}/S^2_{\text{Within}} = 120/21.5 = 5.58.$$

❺ **Decide whether to reject the null hypothesis.** The F of 5.58 is more extreme than the .05 cutoff F of 2.73. Therefore, reject the null hypothesis. The research hypothesis is supported; the different experimental treatments do produce different effects on the standard performance measure.

Figuring Effect Size for an Analysis of Variance

Figure the effect size for the analysis of variance question above.

Answer

$$R^2 = (S^2_{\text{Between}})(df_{\text{Between}})/[(S^2_{\text{Between}})(df_{\text{Between}}) + (S^2_{\text{Within}})(df_{\text{Within}})]$$
$$= (120)(3)/[(120)(3) + (21.5)(76)] = (360)/[(360) + (1634)] = .18.$$

Outline for Writing Essays for a One-Way Analysis of Variance

1. Explain that the one-way analysis of variance is used for hypothesis testing when you have scores from three or more entirely separate groups of people. Be sure to explain the meaning of the research hypothesis and the null hypothesis in this situation.
2. Describe the core logic of hypothesis testing in this situation. Be sure to mention that the analysis of variance involves comparing the results of two ways of estimating the population variance. One population variance estimate (the within-groups estimate) is based on the variation within each sample and the other estimate (the between-groups estimate) is based on the variation among the means of the samples. Be sure to describe these estimates in detail (including how they are figured, why they are figured that way, and how each is affected by whether or not the null hypothesis is true); explain how and why they are used to figure an F ratio.
3. Explain the logic of the comparison distribution that is used with a one-way analysis of variance (the F distribution).
4. Describe the logic and process for determining the cutoff sample F score on the comparison distribution at which the null hypothesis should be rejected.
5. Explain how and why the scores from Steps ❸ and ❹ of the hypothesis-testing process are compared. Explain the meaning of the result of this comparison with regard to the specific research and null hypotheses being tested.

Practice Problems

These problems involve figuring. Most real-life statistics problems are done on a computer with special statistical software. Even if you have such software, do these problems by hand to ingrain the method in your mind. To learn how to use a computer to solve statistics problems like those in this chapter, refer to the "Using SPSS" section at the end of this chapter.

All data are fictional unless an actual citation is given.

Set I (for answers, see the end of this chapter)

1. For each of the following studies, decide whether you can reject the null hypothesis that the groups come from identical populations. Use the .05 level. In addition, figure the effect size and approximate power for each. Note that studies (b) and (c) provide S, not S^2.

(a)	Group 1	Group 2	Group 3
n	10	10	10
M	7.4	6.8	6.8
S^2	.82	.90	.80

(b)	Group 1	Group 2	Group 3	Group 4
n	25	25	25	25
M	94	101	124	105
S	24	28	31	25

(c)	Group 1	Group 2	Group 3	Group 4	Group 5
n	25	25	25	25	25
M	94	101	124	105	106
S	24	28	31	25	27

2. For each of the following studies, decide whether you can reject the null hypothesis that the groups come from identical populations. Use the .01 level. In addition, figure the effect size for each. (Be sure to show your calculations throughout.)

(a)	Group 1	Group 2	Group 3
	8	6	4
	8	6	4
	7	5	3
	9	7	5

(b)	Group 1	Group 2	Group 3
	12	10	8
	4	2	0
	12	10	8
	4	2	0

3. A social worker at a small psychiatric hospital was asked to determine whether there was any clear difference in the length of stay of patients with different categories of diagnosis. Looking at the last four clients in each of the three major categories, the results (in terms of weeks of stay) were as follows:

Diagnosis Category		
Affective Disorders	**Personality Disorders**	**Drug-Related Conditions**
8	12	8
6	8	10
4	9	12
6	11	10

(a) Using the .05 level and the five steps of hypothesis testing, determine if there is a significant difference in length of stay among diagnosis categories; (b) figure the effect size for the study; and (c) explain your answer to (a) to someone who understands everything involved in conducting a *t* test for independent means but who is unfamiliar with the analysis of variance.

4. A study compared the felt intensity of unrequited love (loving someone who doesn't love you) among three groups: 50 individuals who were currently experiencing unrequited love, who had a mean experienced intensity, $M = 3.5$, $S^2 = 5.2$; 50 who had previously experienced unrequited love and described their experiences retrospectively, $M = 3.2$, $S^2 = 5.8$; and 50 who had never experienced unrequited love but described how they thought they would feel if they were to experience it, $M = 3.8$, $S^2 = 4.8$. (a) Using the .05 level and the five steps of hypothesis testing, determine the significance of the difference among groups; (b) figure the effect size and approximate power; and (c) explain your answer to (a) to someone who has never had a course in statistics.
5. A researcher studying genetic influences on learning compares the maze performance of four genetically different strains of mice using eight mice per strain. Performance for the four strains were as follows:

Strain	Mean	*S*
J	41	3.5
M	38	4.6
Q	14	3.8
W	37	4.9

(a) Using the .01 significance level and the five steps of hypothesis testing, is there an overall difference in maze performance among the four strains?; (b) figure the effect size for the study; and (c) explain your answer to (a) to someone who is familiar with hypothesis testing with known populations, but is unfamiliar with the *t* test or the analysis of variance.

6. What is the power of each of the following planned studies, using the analysis of variance with $p < .05$?

	Predicted Effect Size	Number of Groups	Participants in Each Group
(a)	Small	3	20
(b)	Small	3	30
(c)	Small	4	20
(d)	Medium	3	20

7. About how many participants do you need in each group for 80% power in each of the following planned studies, using the analysis of variance with $p < .05$?

	Predicted Effect Size	Number of Groups
(a)	Small	3
(b)	Large	3
(c)	Small	4
(d)	Medium	3

8. Each of the following is a table of means showing results of a study using a factorial design. Assuming that any differences are statistically significant, for each table, (a) and (b) make two bar graphs showing the results (in one graph grouping the bars according to one variable and in the other graph grouping the bars according to the other variable); (c) indicate which effects (main and interaction), if any, are found; and (d) describe the meaning of the pattern of means (that is, any main or interaction effects or the lack thereof) in words.

   i. Measured variable: Income (thousands of U.S. dollars)

		Age	
		Young	*Old*
Class	*Lower*	25	40
	Upper	30	110

   ii. Measured variable: Grade point average

		Major	
		Science	*Arts*
College	*Community*	2.1	2.8
	Liberal Arts	2.8	2.1

   iii. Measured variable: Days sick per month

		Gender	
		Females	*Males*
Group	*Exercisers*	2.0	2.5
	Controls	3.1	3.6

9. Each of the following is a table of means showing results of a study using a factorial design. Assuming that any differences are statistically significant, for each table, (a) and (b) make two bar graphs showing the results (in one graph grouping

the bars according to one variable and in the other graph grouping the bars according to the other variable); (c) indicate which effects (main and interaction), if any, are found; and (d) describe the meaning of the pattern of means (that is, any main or interaction effects or the lack thereof) in words.

i. Measured variable: Conversation length

Relationship	Topic: *Nonpersonal*	Topic: *Personal*
Friend	16	20
Parent	10	6

ii. Measured variable: Rated restaurant quality

Cost	City: *New York*	City: *Chicago*	City: *Vancouver*
Expensive	9	5	7
Moderate	6	4	6
Inexpensive	4	3	5

iii. Measured variable: Ratings of flavor

Type	Coffee Brand: *X*	Coffee Brand: *Y*	Coffee Brand: *Z*
Regular	7	4	6
Decaf	5	2	6

10. Grilo, Walker, Becker, Edell, and McGlashan (1997) conducted a study of the relationship of depression and substance use to personality disorders. Personality disorders are persistent, problematic traits and behaviors that exceed the usual range of individual differences. The researchers conducted interviews assessing personality disorders with adolescents who were psychiatric inpatients and had one of three diagnoses: those with major depression, those with substance abuse, and those with both major depression and substance abuse. The mean number of disorders was as follows: major depression, $M = 1.0$; substance abuse, $M = .7$; those with both conditions, $M = 1.9$. The researchers reported, "The three study groups differed in the average number of diagnosed personality disorders, $F(2, 112) = 10.18$, $p < .0001$." Explain this result to someone who is familiar with hypothesis testing with known populations, but is unfamiliar with the t test or the analysis of variance.
11. A researcher wants to know whether the need for health care among prisoners varies according to the different types of prison facilities. The researcher randomly selects 40 prisoners from each of the three main types of prisons in a particular Canadian province and conducts exams to determine their need for health care. In the article describing the results, the researcher reported the means for each group and then added: "The need for health care among prisoners in the three types of prison systems appeared to be clearly different, $F(2, 117) = 5.62$, $p < .01$." Explain this result to a person who has never had a course in statistics.
12. Which type of English words are longer: nouns, verbs, or adjectives? Go to a book of at least 400 pages (not this book or any other textbook) and turn to random

pages (using the random numbers listed at the end of this paragraph). Go down the page until you come to a noun. Note its length (in number of letters). Do this for 10 different nouns. Do the same for 10 verbs and then for 10 adjectives. Using the .05 significance level, carry out an analysis of variance comparing the three types of words. (Be sure also to give the full bibliographic information on the book you used: authors, title, year published, publisher, city.)
73, 320, 179, 379, 323, 219, 176, 167, 102, 228, 352, 4, 335, 118, 12, 333, 123, 38, 49, 399, 17, 188, 264, 342, 89, 13, 77, 378, 223, 92, 77, 152, 34, 214, 75, 83, 198, 210.

Set II

13. For each of the following studies, decide whether you can reject the null hypothesis that the groups come from identical populations. Use the .05 level.

(a)	Group 1	Group 2	Group 3
n	5	5	5
M	10	12	14
S^2	4	6	5

(b)	Group 1	Group 2	Group 3
n	10	10	10
M	10	12	14
S^2	4	6	5

(c)	Group 1	Group 2	Group 3
n	5	5	5
M	10	14	18
S^2	4	6	5

(d)	Group 1	Group 2	Group 3
n	5	5	5
M	10	12	14
S^2	2	3	2.5

14. For each of the following studies, decide whether you can reject the null hypothesis that the groups come from identical populations. Use the .01 level. In addition, figure the effect size for each.

(a)	Group 1	Group 2	Group 3
	1	1	8
	2	2	7
	1	1	8
	2	2	7

(b)	Group 1	Group 2	Group 3
	1	4	8
	2	5	7
	1	4	8
	2	5	7

15. An organizational researcher was interested in whether individuals working in different sectors of a company differed in their attitudes toward the company. The results for the three people surveyed in engineering were 10, 12, and 11; for the three in the marketing department, 6, 6, and 8; for the three in accounting, 7, 4, and 4; and for the three in production, 14, 16, and 13 (higher numbers mean more positive attitudes). Was there a significant difference in attitude toward the company among employees working in different sectors of the company at the .05 level? (a) Carry out the five steps of hypothesis testing; (b) figure the effect size; and (c) explain your answer to (a) to someone who understands everything involved in conducting a *t* test for independent means, but is unfamiliar with the analysis of variance.
16. Do students at various colleges differ in how sociable they are? Twenty-five students were randomly selected from each of three colleges in a particular region and were asked to report on the amount of time they spent socializing each day with other students. The results for College X was a mean of 5 hours and an estimated population variance of 2 hours; for College Y, $M = 4$, $S^2 = 1.5$; and for College Z, $M = 6$, $S^2 = 2.5$. What should you conclude? Use the .05 level. (a) Use the five steps of hypothesis testing; (b) figure the effect size; and (c) explain your answers to (a) and (b) to someone who understands everything involved in conducting a *t* test for independent means but who has never heard of the analysis of variance.
17. A researcher studying artistic preference randomly assigns a group of 45 participants to one of three conditions in which they view a series of unfamiliar abstract paintings. The 15 participants in the Famous condition are led to believe that these are each famous paintings; their mean rating for liking the paintings is 6.5 $(S = 3.5)$. The 15 in the Critically Acclaimed condition are led to believe that these are paintings that are not famous but are very highly thought of by a group of professional art critics; their mean rating is 8.5 $(S = 4.2)$. The 15 in the Control condition are given no special information about the paintings; their mean rating is 3.1 $(S = 2.9)$. Does what people are told about paintings make a difference in how well they are liked? Use the .05 level. (a) Use the five steps of hypothesis testing; (b) figure the effect size; and (c) explain your answer to (a) to someone who is familiar with the *t* test for independent means, but is unfamiliar with analysis of variance.
18. What is the power of each of the following planned studies, using the analysis of variance with $p < .05$?

	Predicted Effect Size	Number of Groups	Participants in Each Group
(a)	Small	4	50
(b)	Medium	4	50
(c)	Large	4	50
(d)	Medium	5	50

19. About how many participants do you need in each group for 80% power in each of the following planned studies, using the analysis of variance with $p < .05$?

	Predicted Effect Size	Number of Groups
(a)	Small	5
(b)	Medium	5
(c)	Large	5
(d)	Medium	3

20. Each of the following is a table of means showing results of a study using a factorial design. Assuming that any differences are statistically significant, for each table (a) and (b), make two bar graphs showing the results (in one graph grouping the bars according to one variable and in the other graph grouping the bars according to the other variable); (c) indicate which effects (main and interaction), if any, are found; and (d) describe the meaning of the pattern of means (that is, any main or interaction effects or the lack thereof) in words.
   i. Measured variable: Degree of envy of another person's success

Status of Other	Degree of Success: *Great*	Degree of Success: *Small*
Friend	8	5
Stranger	1	4

   ii. Measured variable: Observed engagement in the activity

Situation	Play Activity: *Blocks*	Play Activity: *Dress Up*
Alone	4.5	2.5
With playmate	2.5	4.5

   iii. Measured variable: Intensity of attention

Type of Balletgoer	Program: *Swan Lake*	Program: *Modern*
Regular	20	15
Sometime	15	15
Novice	10	5

21. Each of the following is a table of means showing results of a study using a factorial design. Assuming that any differences are statistically significant, for each table (a) and (b), make two bar graphs showing the results (in one graph grouping the bars according to one variable and in the other graph grouping the bars according to the other variable); (c) indicate which effects (main and interaction), if any, are found; and (d) describe the meaning of the pattern of means (that is, any main or interaction effects or the lack thereof) in words.

i. Measured variable: Right frontal neural activity in brain during memory task

Times Presented	Items Remembered: *Words*	*Pictures*
Once	45	68
Twice	30	30

ii. Measured variable: Approval rating of the U.S. president

Class	Region: *West*	*East*	*Midwest*	*South*
Middle	70	45	55	50
Lower	50	25	35	30

iii. Measured variable: Satisfaction with education

Time After Obtaining BA	Gender: *Females*	*Males*
1 month	3	3
1 year	4	4
5 years	9	9

22. An experiment is conducted in which 60 participants each fill out a personality test, but not according to the way the participants see themselves. Instead, 15 are randomly assigned to fill it out according to the way they think their mothers see them (that is, the way they think their mothers would fill it out to describe the participants); 15 as their fathers would fill it out for them; 15 as their best friends would fill it out for them; and 15 as the professors they know best would fill it out for them. The main results appear in Table 14. Explain these results to a person who has never had a course in statistics.
23. Sinclair and Kunda (2000) tested the idea that, if you want to think well of someone (for example, because they have said positive things about you), you are less influenced by the normal stereotypes when evaluating them. Participants filled out

Table 14 Means for Main Personality Scales for Each Experimental Condition (Fictional Data)

Scale	Mother	Father	Friend	Professor	$F(3, 56)$
Conformity	24	21	12	16	4.21**
Extroversion	14	13	15	13	2.05
Maturity	15	15	22	19	3.11*
Self-confidence	38	42	27	32	3.58*

$^*p < .05$, $^{**}p < .01$.

a questionnaire on their social skills and then received feedback from either a male or female "manager in training." The study was rigged so that the managers gave half the participants positive feedback and half negative feedback. The participants then rated the managers for their skill at evaluating them. The question was whether the usual tendency to stereotype women as less skillful managers would be undermined when people got positive ratings. Sinclair and Kunda described their results as follows:

> Participants' ratings of the manager's skill at evaluating them were analyzed with a 2 (feedback) $\times$ 2 (manager gender) [two-way] ANOVA. Managers who had provided positive feedback ($M = 9.08$) were rated more highly than were managers who had provided negative feedback ($M = 7.46$), $F(1,46) = 19.44$, $p < .0001$. However, as may be seen in [Figure 8], the effect was qualified by a significant interaction, $F(1, 46) = 4.71$, $p < .05$. . . . (pp. 1335–1336)

Describe the meaning of these results to a person who has never had a course in statistics. (Do not go into the details of the figuring, just the basic logic of the pattern of means, and the meaning of any significant results.)

24. Cut up 100 little pieces of paper of about the same size and write "1" on 16, "2" on 34, "3" on 34, and "4" on 16 of them. (You are creating an approximately normal distribution.) Put the slips into a bowl or hat, mix them up, draw out two, write the numbers on them down, and put them back. Then draw out another two, write down their numbers, and put them back, and finally another two, write down their numbers, and put them back. (Strictly speaking, you should sample "with replacement." That means putting each one, not two, back after writing its number down. But we want to save you a little time, and it should not make much difference in this case.) Figure an analysis of variance for these three randomly selected groups of two each. Write down the F ratio, and repeat the entire drawing process and analysis of variance again. Do this entire process at least 20 times,

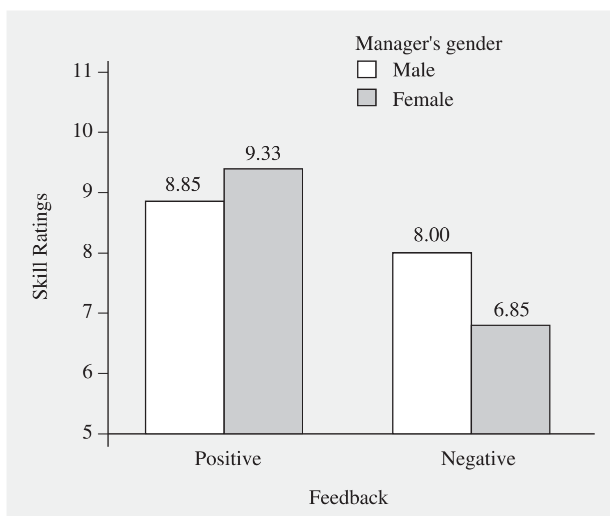


Figure 8 Participants' ratings of the manager's skill at evaluating them as a function of feedback favorability and the manager's gender (Study 2). *Source:* Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin, 26, 11,* 1329–1342. Copyright © 2000, Society for Personality and Social Psychology, Inc. Reprinted by permission SAGE Publications.

and make a frequency polygon of your results. You are creating an F distribution for 2 (3 groups − 1) and 3 (2 − 1 in each of three groups) degrees of freedom. At what point do the top 5% of your F scores begin? Compare that to the 5% cutoff given in Table 3 in the appendix "Tables" for 2 and 3 degrees of freedom.

Using SPSS

The in the following steps indicates a mouse click. (We used SPSS version 17.0 for Windows to carry out these analyses. The steps and output may be slightly different for other versions of SPSS.)

One-Way Analysis of Variance

It is easier to learn these steps using actual numbers, so we will use the criminal record example from earlier in the chapter. The scores for that example are shown in Table 3.

1. Enter the scores into SPSS. SPSS assumes that all scores in a row are from the same person. In this example, each person is in only one of the three groups (the Criminal Record group, the Clean Record group, or the No Information group). Thus, to tell SPSS which person is in each group, enter the numbers as shown in Figure 9. In the first column (labeled "group"), we used the number "1" to indicate that a person is in the Criminal Record group, the number "2" to indicate that a person is in the Clean Record group, and the number "3" to indicate a person is in the No Information group.
2. *Analyze.*
3. *Compare means.*

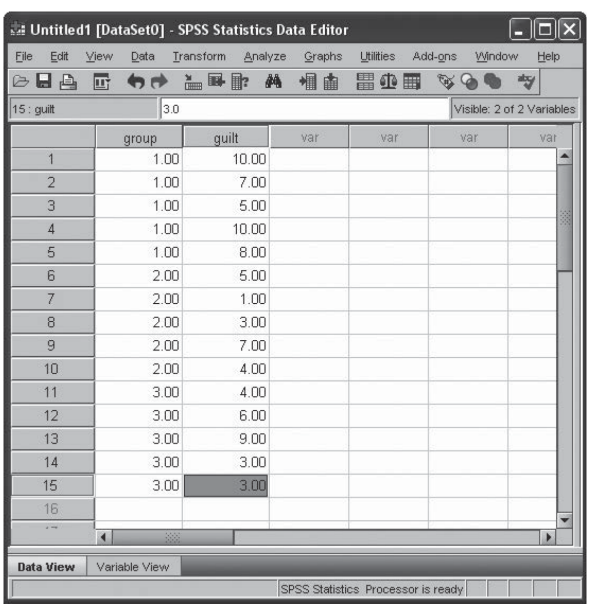
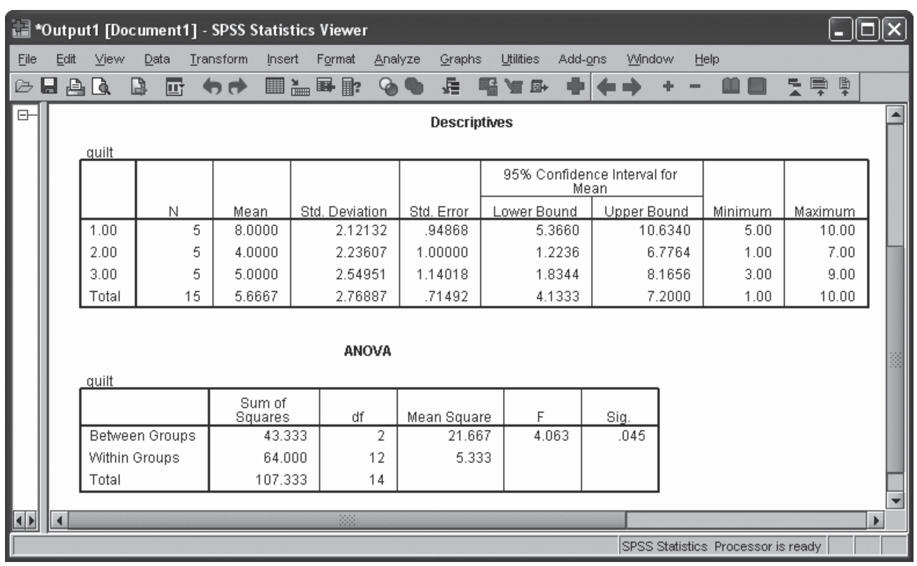


Figure 9 SPSS data editor window for the criminal record study example (in which 15 individuals rated the guilt of a defendant after being randomly assigned to one of three groups that were given different information about the defendant's previous criminal record).

4. *One-Way ANOVA.*
5. on the variable called "guilt" and then the arrow next to the box labeled "Dependent List" (this is the name used by SPSS to refer to the measured variable). This tells SPSS that the analysis of variance should be carried out on the scores for the "guilt" variable.
6. the variable called "group" and then the arrow next to the box labeled "Factor." This tells SPSS that the variable called "group" shows which person is in which group.
7. *Options.* the box labeled *Descriptive* (this checks the box). This tells SPSS to provide descriptive statistics (such as the mean and standard deviation) for each group. *Continue.* (Step 7 is optional, but we strongly recommend requesting descriptive statistics for any hypothesis testing situation.)
8. *OK.* Your SPSS output window should look like Figure 10.

The first table in the SPSS output provides descriptive statistics (number of individuals, mean, estimated population standard deviation, and other statistics) for the "guilt" scores for each of the three groups.

The second table in the SPSS output shows the actual results of the one-way analysis of variance. The first column lists the types of population variance estimates (between groups and within groups). For our purposes, you can ignore the second column (labeled "Sum of Squares"). The third column, "df," gives the degrees of freedom. In the between groups row, this corresponds to df_{Between}; in the within groups row, this corresponds to df_{Within}. The fourth column, labeled "Mean Square," gives the population variance estimates (S^2_{Between} and S^2_{Within}), with the between-groups estimate first and then the within-groups estimate. The next column gives the F ratio for the analysis of variance. Allowing for rounding error, the values for "df," "Mean Square," and "F" are the same as those given earlier in the chapter. The final column, "Sig.," shows the exact significance level of the F ratio. The significance level of .045 is less than our .05 cutoff for this example. Thus, you can reject the null hypothesis and the research hypothesis is supported (that is, the result is statistically significant).



Descriptives

guilt

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean: Lower Bound	95% Confidence Interval for Mean: Upper Bound	Minimum	Maximum
1.00	5	8.0000	2.12132	.94868	5.3660	10.6340	5.00	10.00
2.00	5	4.0000	2.23607	1.00000	1.2236	6.7764	1.00	7.00
3.00	5	5.0000	2.54951	1.14018	1.8344	8.1656	3.00	9.00
Total	15	5.6667	2.76887	.71492	4.1333	7.2000	1.00	10.00

ANOVA

guilt

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	43.333	2	21.667	4.063	.045
Within Groups	64.000	12	5.333		
Total	107.333	14			

Figure 10 SPSS output window for a one-way analysis of variance for the criminal record example.

Answers to Set I Practice Problems

1. (a) F needed $(df = 2, 27; p < .05) = 3.36$; $S_M^2 = [\Sigma(M - GM)^2]/df_{\text{Between}} = [(7.4 - 7)^2 + (6.8 - 7)^2 + (6.8 - 7)^2]/2 = .12$; $S_{\text{Between}}^2 = (S_M^2)(n) = (.12)(10) = 1.2$; $S_{\text{Within}}^2 = (S_1^2 + S_2^2 + \ldots + S_{\text{Last}}^2)/N_{\text{Groups}} = (.82 + .90 + .80)/3 = .84$; $F = 1.2/.84 = 1.43$; do not reject the null hypothesis; effect size, $R^2 = (S_{\text{Between}}^2)(df_{\text{Between}})/[(S_{\text{Between}}^2)(df_{\text{Between}}) + (S_{\text{Within}}^2)(df_{\text{Within}})] = (1.2)(2)/[(1.2)(2) + (.84)(27)] = .10$; from Table 8, approximate power = between .20 and .45.
   (b) F needed $(df = 3, 96; p < .05) = 2.70$ (actually using $df = 3, 95$); $S_M^2 = 164.67$; $S_{\text{Between}}^2 = (164.67)(25) = 4116.75$; $S_{\text{Within}}^2 = 736.5$; $F = 5.59$; reject the null hypothesis; effect size, $R^2 = .15$; approximate power = between .85 and .96.
   (c) F needed $(df = 4, 120; p < .05) = 2.46$ (actually using 4, 100); $S_M^2 = 123.5$; $S_{\text{Between}}^2 = (123.5)(25) = 3087.5$; $S_{\text{Within}}^2 = 735$; $F = 4.20$; reject the null hypothesis; effect size, $R^2 = .12$; approximate power = between .90 and .98 (using the values for $R^2 = .14$ from Table 8).
2. (a) F needed $(df = 2, 9; p < .01) = 8.02$; Group 1: $M = 8, S^2 = .67$; Group 2: $M = 6, S^2 = .67$; Group 3: $M = 4, S^2 = .67$; $S_{\text{Between}}^2 = (4)(4) = 16$; $S_{\text{Within}}^2 = .67$; $F = 16/.67 = 23.88$; reject the null hypothesis; effect size, $R^2 = .84$.
   (b) F needed $(df = 2, 9; p < .01) = 8.02$; Group 1: $M = 8, S^2 = 21.33$; Group 2: $M = 6, S^2 = 21.33$; Group 3: $M = 4, S^2 = 21.33$; $S_{\text{Between}}^2 = (4)(4) = 16$; $S_{\text{Within}}^2 = 2.133$; $F = 16/21.33 = .75$; do not reject the null hypothesis; effect size, $R^2 = .14$.
3. (a)
   ❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.** There are three populations of interest:

   Population 1: Patients with affective disorders.
   Population 2: Patients with personality disorders.
   Population 3: Patients with drug-related conditions.

   The null hypothesis is that the three populations have the same mean. The research hypothesis is that the three populations do not have the same mean.
   ❷ **Determine the characteristics of the comparison distribution.** F distribution with 2 and 9 degrees of freedom.
   ❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** 5% level, F (2, 9) needed = 4.26.
   ❹ **Determine your sample's score on the comparison distribution.** $S_{\text{Between}}^2 = (5.33)(4) = 21.32$; $S_{\text{Within}}^2 = (2.67 + 3.33 + 2.67)/3 = 2.89$; $F = 21.32/2.89 = 7.38$.
   ❺ **Decide whether to reject the null hypothesis.** F from Step ❹ (7.38) is more extreme than cutoff F from Step ❸ (4.26); reject the null hypothesis.
   (b) Effect size, $R^2 = (S_{\text{Between}}^2)(df_{\text{Between}})/[(S_{\text{Between}}^2)(df_{\text{Between}}) + (S_{\text{Within}}^2)(df_{\text{Within}})] = .62$.
   (c) The scores in this study are from three entirely separate groups of people. An analysis of variance is the appropriate analysis in this situation. The null hypothesis is that the three groups are from populations of length-of-stay scores with equal means. Also, as with a *t* test, we must be able to assume that they have equal variances. If this null hypothesis is true, then you can estimate the variance of these equal populations in two ways:
   (i) You can estimate from the variation within each of the three groups and then average them. (This is just what you would do in a *t* test for independent means, except now you are averaging three instead of just two. Also in a *t* test you would weight these variances according to the degrees of freedom they contribute to the overall estimate. However, because all three groups have equal numbers, you can simply average them—in effect weighting them equally.) In this example, the three variance estimates were 2.67, 3.33, and 2.67, which gave a pooled estimate of 2.89. This is called the within-groups estimate of the population variance.
   (ii) You can estimate the variance using the three means. If we assume the null hypothesis is true, then the means of the three groups are based on samples taken from identical populations. Each of these identical populations will have an identical distribution of means of samples taken from that population. The means of our three samples are all from identical populations, which is the same as if they were all from the same population. Thus, the amount of variation among our three means should reflect the variation in the distribution of means that they can be thought of as coming from. As a result, I can use these three means (6, 10, and 10) to estimate the variance in this distribution of means. Using the usual formula for estimating a population variance, I get 5.33.

   However, what we want is the variance of a distribution of individuals. So the question is this: What would be the distribution of individuals that would produce a distribution of means (of four scores each) with a variance of 5.33? To find the distribution of means from a distribution of individuals, you divide the variance of the distribution of individuals by the size of the samples. However, in our present situation, you want to do the reverse. Thus, you multiply the variance of the distribution of means by the size of the samples to get the variance of the distribution of individuals. This comes out to 5.33 times 4, or 21.32. This is called the between-groups estimate of the population variance.

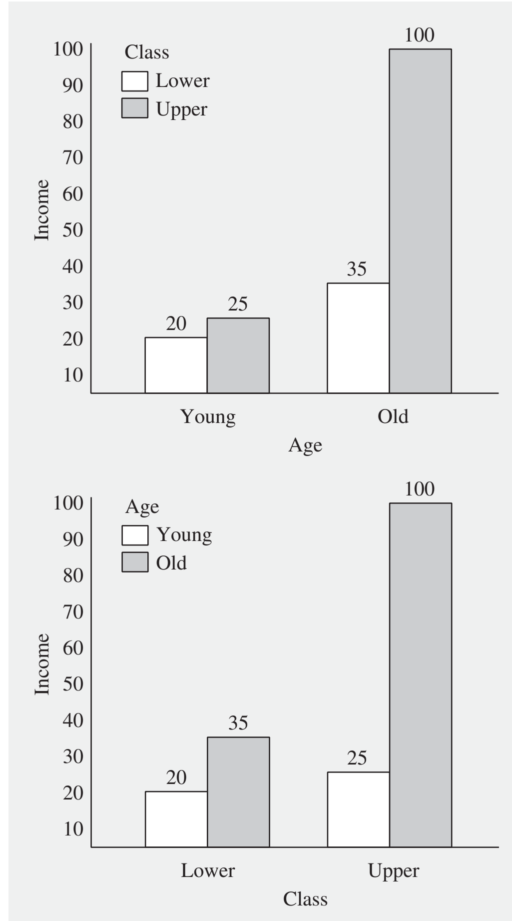
   If the null hypothesis is true, the within-groups and between-groups estimates of the population variance should be about the same. This is because they are estimates of the same thing, the variance of any of the populations (which are all assumed to be the same). This means that the ratio of the between-groups estimate divided by the within-groups estimate should be about 1.

However, suppose the null hypothesis is false so that the three populations from which these groups come in fact have different means. In that situation, the estimate based on the variation among the group means will be bigger than the estimate based on the variation within the groups. The reason it will be bigger is as follows: If the null hypothesis is true, the only reason that the means of our groups vary is because of the variance inside of each of the three identical distributions of means. But if the null hypothesis is false, each of those distributions of means also has a different mean. Thus, the variation in our means is due to *both* the variation inside of each of these now *not* identical distributions of means and also to the differences in the means of these distributions of means. Thus, there is an additional source of variation in the means of our groups. If you estimate the variance of the population using these three means, it will be larger than if the null hypothesis were true. At the same time, however, the within-groups variance is not affected by whether the three groups have different means. This is because the within-groups variance considers variation only within each of the groups. The within-groups variance thus does not get any bigger if the null hypothesis is false. The result of all this is that when the null hypothesis is false, the ratio of the between-groups variance to the within-groups variance will be more than 1.

The ratio of the between-groups estimate to the within-groups estimate is called an *F* ratio. In this example, our *F* ratio is 21.32 to 2.89: $21.32/2.89 = 7.38$.

Statisticians have made tables of what happens when you figure *F* ratios based on the situation in which you randomly take a group of four scores from each of three identical populations. This is the situation in which the null hypothesis is true. Looking at these tables, it turns out that there is less than a 5% chance of getting an *F* ratio larger than 4.26. Because our actual *F* ratio is bigger than this, we can reject the null hypothesis.

4. (a), (b), and (c). Hypothesis-testing steps, estimated effect size, and explanation similar to 3 above (except for part c you need to include extra material). *F* needed $(df = 2, 147; p < .05) = 3.09$ (actually using $df = 2, 100$); $S^2_{\text{Between}} = (.09)(50) = 4.5$; $S^2_{\text{Within}} = (5.2 + 5.8 + 4.8)/3 = 5.27$; $F = .85$; do not reject the null hypothesis. Effect size, $R^2 = .01$.
5. (a), (b), and (c). Hypothesis-testing steps and explanations similar to 3 above (except for part c you need to include extra material). *F* needed $(df = 3, 28; p < .01) = 4.57$; $S^2_{\text{Between}} = (155)(8) = 1240$; $S^2_{\text{Within}} = (3.5^2 + 4.6^2 + 3.8^2 + 3.8^2 + 4.9^2)/4 = 17.97$; $F = 69.0$; reject the null hypothesis. Effect size, $R^2 = .88$.
6. From Table 8: (a) .09; (b) .12; (c) .10; (d) .38.
7. From Table 9: (a) 322; (b) 21; (c) 274; (d) 52.
8. (i) (a) and (b)



(c) Main effects for class and age, interaction effect; (d) income is greater in general for upper-class and for older individuals, but the combination of older and upper class has a higher income than would be expected just from the effects of either variable alone.

(ii) (a) and (b) Graphs of the same kind as in (i) (a) and (b) above; (c) no main effects, interaction effect; (d) neither type of college nor type of major, by itself, predicts grades, but there is a clear pattern if one considers the combinations: Grades are highest for community college arts majors and for liberal arts college science majors.

(iii) (a) and (b) Graphs of the same kind as in (i) (a) and (b) above; (c) both main effects significant; no interaction; (d) females miss fewer days per month than males; those who exercise miss fewer days per month than controls. Each combination misses the number of days you would expect knowing their level of each variable separately.

9. (i) (a) and (b) Graphs of the same kind as in 8 (i) (a) and (b) above; (c) main effect for relationship and an interaction; (d) conversations are longer with friends; but the difference is much greater for personal than for nonpersonal topics.

(ii) (a) and (b) Graphs of the same kind as in 8 (i) (a) and (b) above; (c) main effect for city and cost, plus an interaction; (d) restaurant quality is different in different cities, with New York highest and Chicago lowest. Restaurant quality is different in different price ranges, with expensive the best and inexpensive the least. The two factors do not simply combine, however, as price makes more difference in New York than in other cities.

(iii) (a) and (b) Graphs of the same kind as in 8 (i) (a) and (b) above; (c) main effects for brand and type and an interaction; (d) flavor is rated on the average more positively for regular than decaf and brands Z and X are rated more favorably than Y. However, there is an interaction in which there is no difference between regular and decaf for brand Z, but for brands Z and Y, regular is rated 2 points higher.

10. Similar to 3c but focusing on this study's results.
11. Similar to 3c but focusing on this study's results.

Steps of Hypothesis Testing for Major Procedures

One-way analysis of variance

❶ **Restate the question as a research hypothesis and a null hypothesis about the populations.**

❷ **Determine the characteristics of the comparison distribution.** **(a)** F distribution.
(b) $df_{\text{Between}} = N_{\text{Groups}} - 1$.
(c) $df_{\text{Within}} = df_1 + df_2 + \ldots + df_{\text{Last}}$.

❸ **Determine the cutoff sample score on the comparison distribution at which the null hypothesis should be rejected.** Use F table.

❹ **Determine your sample's score on the comparison distribution.** **(a)** Figure S^2_{Between}: Figure the mean of each group.
Ⓐ $S^2_M = [\Sigma(M - GM)^2]/df_{\text{Between}}$
Ⓑ $S^2_{\text{Between}} = (S^2_M)(n)$
(b) Figure S^2_{Within}: **(i)** For each group, $S^2 = [\Sigma(X - M)^2]/(n - 1)$; **(ii)** $S^2_{\text{Within}} = (S^2_1 + S^2_2 + \ldots + S^2_{\text{Last}})/N_{\text{Groups}}$
(c) $F = S^2_{\text{Between}}/S^2_{\text{Within}}$

❺ **Decide whether to reject the null hypothesis.** Compare scores from Steps ❸ and ❹.