10

Analysis of Variance

The next procedure we cover is referred to as **AN**alysis **O**f **VA**riance (commonly abbreviated as **ANOVA**). More specifically, we will take up an application known as one-way ANOVA. Many statisticians think of ANOVA as an extension of the difference of means test because it's based, in part, on a comparison of sample means. At the same time, however, the procedure involves a comparison of different estimates of population variance—hence the name analysis of variance. Because ANOVA is appropriate for research involving three or more samples, it has wide applicability.

In the field of experimental psychology, for example, researchers routinely look at results from three or more samples, often referred to as *treatment groups*. One can easily imagine an educational psychologist wanting to know if students exposed to three different treatment conditions or learning environments (positive sanction, negative sanction, and sanction neutral) exhibit different test scores. Assuming the test scores are based on an interval/ratio scale of measurement, ANOVA would be an appropriate approach to the problem.

Similarly, a geographer might be interested in the growth rates of four types of cities—manufacturing centers, government centers, retail centers, and financial centers. A study along those lines would be another research problem ideally suited for ANOVA.

What makes both of these problems appropriate for ANOVA is the fact that they involve more than two groups or samples and a single variable that has been measured at the interval/ratio level of measurement. It's true that research problems like these can be approached with a series of *t* tests, and that might be your inclination if you knew nothing about ANOVA. For example, the geographer could conduct different *t* tests—comparing the growth rates of manufacturing centers with those of government centers, followed by a comparison with financial centers, and so forth—but there are inherent problems in that approach.

A study based on just four types of cities would turn into a series of six *t* tests involving all the possible comparisons. Besides the added work of six individual tests, there's the issue of Type I errors (rejection of the null hypothesis when it is true). Without going into the mathematics of the situation, the fact is that the probability of a Type I error would be magnified. Even though the probability of a Type I error on any one of the six tests would be, let's say, .05 (if that was the designated level of significance), it would increase well beyond that for the six individual tests taken together. Given that, it's no wonder that researchers commonly turn to ANOVA. In short, ANOVA allows the comparison of multiple samples in a single application. That should be apparent once you consider the logic of ANOVA.

LEARNING CHECK

Question: ANOVA is appropriate for what types of research situations?

Answer: ANOVA is appropriate for situations involving three or more samples and a variable measured at the interval/ratio level of measurement.

Before We Begin

Up to this point, we've covered four specific hypothesis testing procedures. First there was the hypothesis test involving a single sample mean, one procedure with sigma (σ) known and then another with sigma (σ) unknown. Then we shifted to situations involving the matched and related samples, followed by situations involving two independent samples. In sum, we looked at four situations and four hypothesis testing procedures.

If you'll take the time to reflect on that—the notion that we looked at four situations and four hypothesis testing procedures—you'll likely see the repetition that occurs in the field of statistics. Just to make certain that you grasp this repetitive nature of hypothesis testing, let me urge you to think about it this way: The underlying logic remains the same; what changes is the research situation. In other words, it's the research situation or problem that dictates what procedure to use.

So, you ask, how does that relate to where we're going? The answer is pretty straight-forward. We started out with research problems involving one sample. Then we dealt with situations involving two samples. Naturally, not every research situation falls into one of those categories; it's common to encounter research situations that involve three or more samples. In a nutshell, that's where we're going in this chapter—research situations involving three or more samples.

The Logic of ANOVA

Imagine for a moment that we want to know if scores on an aptitude test actually vary for students in different types of schooling environments—home schooling, public schooling, and private schooling. This research question involves a comparison of more than two groups. Assuming that the aptitude test scores are measured at the interval/ratio level, the situation is tailor-made for an application of ANOVA. We could easily think of our study as one that asks whether or not aptitude test scores vary on the basis of school environment.

Another way to look at the question is whether or not type of school environment is a legitimate classification scheme when it comes to the matter of aptitude test scores. After all, to refer to students in terms of home, public, and private schooling is to speak in terms of a classification scheme. If aptitude test scores really do vary on the basis of school environment—if there is a significant difference between the scores in the three environments—then it's probably legitimate to speak in terms of school environments when looking at test scores. If there isn't a significant difference between the scores, however, we have to question the legitimacy of the classification scheme. In a sense, we were also dealing with the legitimacy of a classification scheme in the last chapter, particularly in reference to the test for independent sample. To suggest that two groups are different with respect to some variable is, in fact, a way of suggesting that the members of the group or cases can reasonably be classified on the basis of the variable in question. That said, let's return to the topic at hand.

To understand how all of this relates to ANOVA, consider Figure 10-1. Imagine that the three curves shown in Figure 10-1 represent the distributions of aptitude test scores for three samples—a sample of home-schooled students, a sample of public school students, and a sample of private school students.

The three distributions reflect three different means, but the means are fairly close together, and there's substantial variation in the scores within each group. Additionally, there's noticeable overlap in the distributions. The overlap exists, in part, because of those factors taken together—the fact that there's substantial variation *within* each of the distributions, coupled with minimal difference *between* the means. (Technically, the proper phrase should be *among the means* because the comparison typically involves three or more means, but in the language of ANOVA, the phrase *between the means* is used nonetheless. It's just a matter of statistical convention.)

Now consider the distributions shown in Figure 10-2. You'll note that the means of the distributions in Figure 10-2 are very different, and there's no overlap between the three curves.

A grasp of ANOVA begins with an understanding of the different patterns reflected in Figures 10-1 and 10-2. If there's more variation between groups than within groups (as suggested by the illustration in Figure 10-2), then there's support for the assertion that students in the different schooling environments are different with respect to aptitude test scores. Conversely, the illustration in Figure 10-1 would challenge the legitimacy of the classification scheme. Because the means are fairly close together in Figure 10-1, and there is a decided or noticeable overlap between the three samples (home-schooled students, public school students, and private school students), it wouldn't make much sense to speak in terms of type of schooling environment when it comes to test scores on the aptitude test.

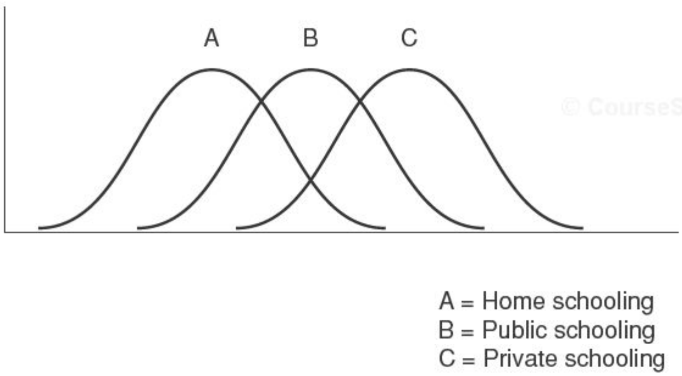


Figure 10-1 Student Performance in Three Learning Environments (Scenario #1)

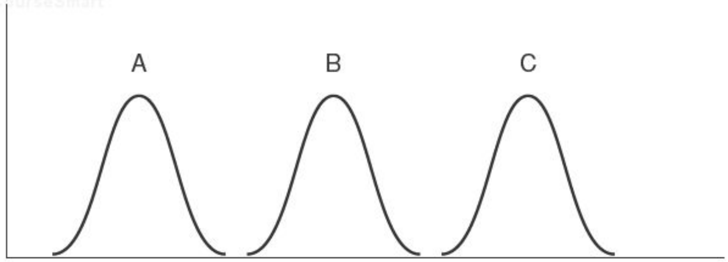


Figure 10-2 Student Performance in Three Learning Environments (Scenario #2)

LEARNING CHECK

Question: What are some ways to think about the purpose of ANOVA? What does it measure?

Answer: It measures whether there's more variation between groups than within groups. It examines the legitimacy of a classification scheme.

From Curves to Data Distributions

So far we've been speaking in rather general terms, with vague references to variation within groups and equally vague references to the variation between groups. Now it's time to take a closer look at ANOVA and how it actually measures the amount of variation we're considering. In essence, ANOVA allows us to calculate a *ratio* of the variation between groups to the variation within groups. This ratio is referred to as the ***F* ratio** (named after its developer, Sir Ronald Fisher).

At the risk of jumping ahead, let me point you in the right direction here. Assuming that we're in search of significant results in a hypothesis-testing situation, what we'll be looking for is more variation between the means of several groups, relative to the variation within the groups. In short, we'll be looking for more variation between than within. Because the *F* ratio is an expression of the between-to-within ratio, we'll be looking for a large *F* value. All factors being equal, the larger our *F* ratio, the greater the probability that we'll reject the null hypothesis.

LEARNING CHECK

Question: What is the *F* ratio? What does it reflect?
Answer: The *F* ratio is the test statistic calculated for ANOVA. It is the ratio of the variation between the samples to the variation within the samples.

The details of how we calculate the *F* ratio is something we'll cover later. Right now, the issue is the underlying logic. So let me give you some more examples, just to get you thinking on the right track.

> A market researcher wants to determine if there's a significant difference between the response rates to five different marketing campaigns. In other words, she wants to know if there's more response rate variation *between* than *within* the different campaigns. If there's more variation in the response rates between than within the campaigns, then it's likely that response rates really do vary by type of campaign.
>
> A sociologist wants to determine if different types of school personnel (teachers, counselors, and coaches) vary in their abilities to recognize risk factors for youth suicide. Assuming he has some sort of interval/ratio level scale to measure risk factor awareness, the question has to do with how the scores on the scale vary by personnel classification. The researcher would have to find more variation *between* different samples (teachers, coaches, and counselors) than *within* the samples to suggest that risk factor recognition actually varies by personnel classification.

By now you should be getting the message: We'll be looking for more variation between the samples than within the samples, at least if we're going to achieve significant results. That, of course, brings us to the matter of how we measure the variation. As you might have guessed, the concept of variation relates to deviations from the mean. And that, in turn, brings us to the various means we might consider.

The Different Means

We can begin with a look at Figure 10-3, but let me warn you in advance. Figure 10-3 is rather abstract. There aren't any values or scores or data of any sort. There isn't any information about a specific research question. It's all very abstract, but it's that way for a reason. One of the best ways to sharpen your thinking about the logic of ANOVA is to think about it in purely abstract terms.

Take a few moments to look at Figure 10-3, and even replicate the illustration on a sheet of paper if you want to (just so you can add some of your

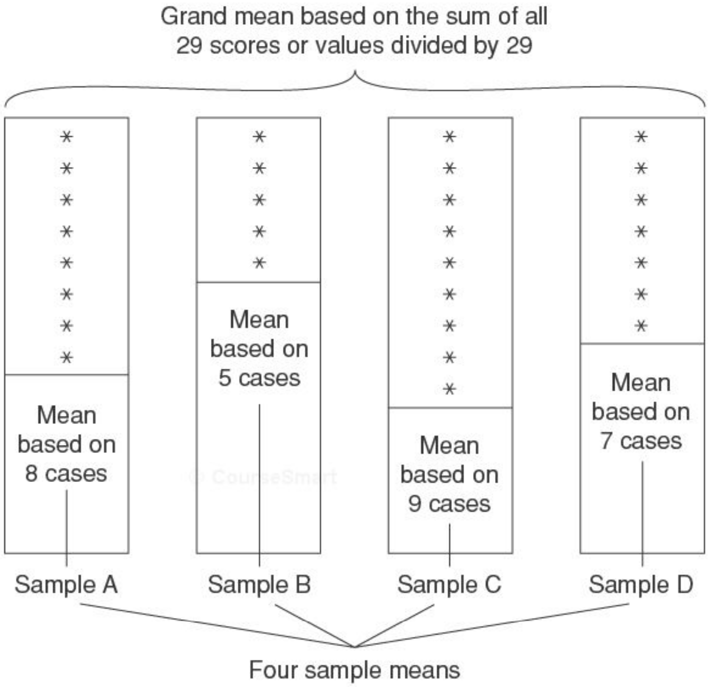


Figure 10-3 The Various Means Involved in ANOVA

own notes or doodles). Figure 10-3 depicts four samples—Sample A, Sample B, Sample C, and Sample D. It doesn't make any difference at this point what those samples relate to. Each sample has its own distribution of scores or values (represented by individual asterisks).

Note that there are 8 cases in Sample A, 5 cases in Sample B, 9 cases in Sample C, and 7 cases in Sample D. Taken together, there are 4 samples and a total of 29 cases. Remember: Each case could be a person (a total of 29 persons), an organization (a total of 29 organizations), a city (a total of 29 cities), or anything else. Each asterisk represents one case—an individual score or value.

Now think about the various means we could calculate. First, there's a mean for Sample A (based on 8 cases), a mean for Sample B (based on 5 cases), a mean for Sample C (based on 9 cases), and a mean for Sample D (based on 7 cases). There are four samples, so there are four sample means.

So far we have four sample means, but there's still another mean to consider. We could, if we wanted to, calculate a **grand mean**—an overall mean based on the 29 cases. We could add all of the individual scores or values (all 29 of them) and then divide by 29. The result would be an overall or grand mean.

Note that we wouldn't calculate the grand mean by adding the 4 sample means and dividing by 4. We could do that if all the samples had the same number of cases, but that's not what we have in this example. Instead, we have 4 samples, and each sample has a different number of cases. Each sample mean, or **group**

mean, is a function, in part, of the number of cases in the sample. Therefore, we can't treat them equally (which is what we would be doing if we simply added the 4 means and divided by 4).

Take another look at Figure 10-3. Even though it's very abstract, think about what the illustration reveals—the notion of a grand mean, as well as a mean for each sample.

> **LEARNING CHECK**
>
> *Question:* What two types of means come into play in ANOVA?
> *Answer:* The grand mean and the individual sample means.

Let me suggest that you spend some time reviewing Figure 10-3 to grasp the notion of a grand mean, along with the individual sample means. The different means are highlighted in the illustration. Once you've done that, we can move to the question of variation and how we measure it.

From Different Means to Different Types of Variation

To understand the matter of variation, think back to the idea of the deviation of a score or value from a mean (a concept introduced in Chapter 3). The concept of variation typically involves the extent to which various scores in a distribution deviate from the mean of the distribution. We can easily apply the same idea to the problem we're considering here.

Let's start with the sample or group means. We'll begin with Sample A. We already know that Sample A has a mean based on the scores from eight cases, so it's easy to think in terms of how far each of the eight scores or values deviates from the mean of Sample A. For an illustration of that point, take a look at the first column in Figure 10-4.

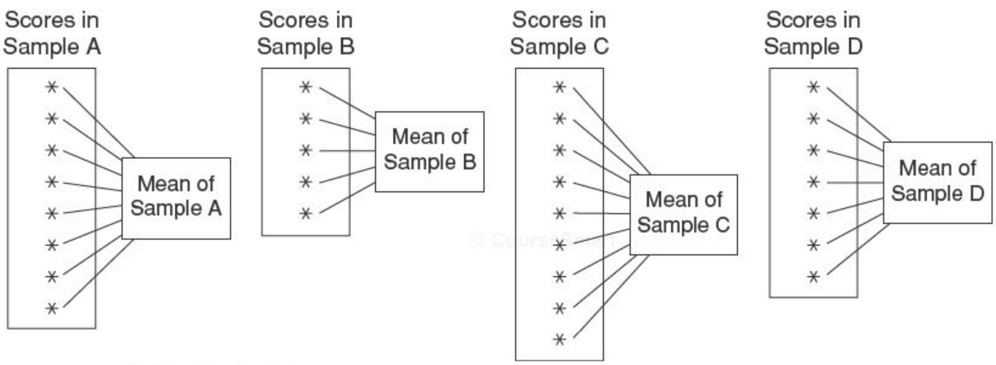


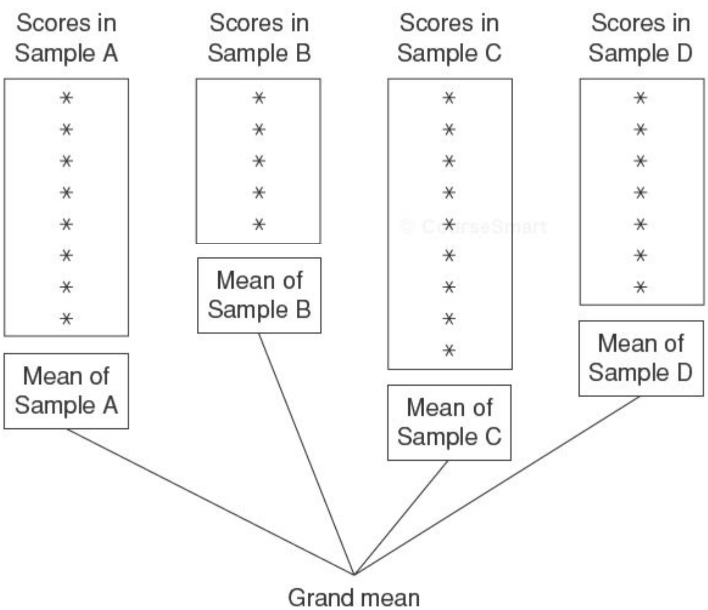
Figure 10-4 Illustration of Within-Groups Variation (Deviation of Individual Sample Scores from the Mean of the Sample)

Figure 10-4 is much the same as Figure 10-3, but with some added information. It illustrates the notion that each score or value in Sample A deviates by some amount from the mean of Sample A. Moving across to Samples B, C, and D, we encounter the same idea again and again. Each sample has a mean and individual scores or values within each sample deviate or vary from the sample mean. In other words, there is a certain amount of variation associated with each sample. This sort of deviation is what we mean by *within-groups variation*.

Now let's turn our attention to another form of variation. You'll recall from our previous discussion that we could obtain a grand mean by adding all the scores and dividing by 29 (since there are 29 cases or scores in our example). Assuming we did that, we could then calculate the difference between the mean of each sample and the grand mean—another form of variation. To get a picture of this sort of variation, take a look at Figure 10-5.

As shown in Figure 10-5, the mean of Sample A deviates a certain number of points from the grand mean, the mean of Sample B deviates a certain number of points from the grand mean, and so on. This sort of deviation is what we mean by *between-groups variation*.

Your success in understanding the ANOVA procedure will largely depend on your ability to fully comprehend these two forms of variation, so let me urge you to take a dark room moment at this point. Allow yourself to think in totally abstract terms—three samples, or seven samples, or whatever number suits you.



Also allow yourself to think in terms of however many cases you want to have in each sample. Imagine that you've calculated a mean for each sample or group, and you've calculated a grand or overall mean. The specifics aren't important at this point. What's important is the notion of two forms of variation. First, there's the variation of scores or values from the individual sample means. Then, there's the variation of each sample mean from the grand mean.

Whenever you think about the variation of individual scores from a sample mean, remind yourself that you're thinking about *within-groups variation* (simply the variation within each sample). Whenever you think about the variation of a sample mean from the grand mean, remind yourself that you're thinking in terms of *between-groups variation* (or the variation of each sample mean from the grand mean). Repeat the process over and over with different mental images. Repeat the process until you're totally comfortable with the concepts of within-groups and between-groups variation. Assuming you've spent sufficient time thinking about those concepts, we can move on to a statement of the null hypothesis.

LEARNING CHECK

Question: What is between-groups variation, and what is within-groups variation?

Answer: Between-groups variation is an expression of the amount of deviation of sample means from the grand mean. Within-groups variation is an expression of the amount of deviation of sample scores from sample means.

The Null Hypothesis

To understand the null hypothesis that's appropriate in the case of ANOVA, let's take up a less abstract example. Let's say, for example, that we're interested in urban unemployment and whether or not the unemployment levels in cities vary by region of the country. Let's also assume that we've used a random sampling technique to select cites in four different regions, and we've recorded the unemployment levels (measured as the percentage of the labor force currently unemployed in each city). Further, let's assume that we've calculated a group mean for each region (four means, one for each of the regions), and an overall mean (based on the unemployment levels in all the cities in our study). The null hypothesis for our study simply states that the means of the regions are equal. It can be stated symbolically as follows:

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

In terms of the *F* ratio, recall that there has to be more variation *between* the regions than *within* the regions for the *F* ratio to be significant. It all goes back to the notion that the *F* ratio is an expression of the ratio of the variation between groups to the variation within groups; the larger the *F* ratio, the more likely it is to be significant. If all the sample means were equal, there wouldn't be any between-groups variation. That, of course, is the situation described by the null hypothesis.

We'll eventually calculate the *F* ratio (our test statistic) as a test of the null hypothesis, and we'll arrive at a conclusion. If our calculated test statistic (the *F* ratio) meets or exceeds the critical value, we'll reject the null hypothesis (with a known probability of having committed a Type I error). All of that will eventually unfold as we work through an application of ANOVA, so that's where we'll turn next.

The Application

We'll begin our application by looking at the data presented in Table 10-1. The table presents the unemployment data for cities in four regions, described in the previous scenario. The same assumptions we encountered in the difference of means test apply in this case—namely, that the unemployment levels (expressed as a percentage of the labor force) represent interval/ratio level data and that the cities were randomly selected. Following the normal convention, we want to select a level of significance in advance, so we'll set that at .05.

Take a few moments to examine the data presented in Table 10-1. First, take note that the sample sizes are different. ANOVA doesn't require the different

Table 10-1 Levels of Unemployment by Region

North	South	East	West
3.8	4.2	8.8	4.8
7.1	6.5	5.1	1.2
9.6	4.4	12.7	8.0
8.4	8.1	6.4	9.4
5.1	7.6	9.8	3.6
11.6	5.8	6.3	8.7
6.2	4.0	10.2	6.5
7.9	7.3	8.5	
9.0	5.2	11.9	
10.3	4.8	8.6	
$\overline{X} = 7.90$	$\overline{X} = 5.79$	$\overline{X} = 8.83$	$\overline{X} = 6.03$
$n = 10$	$n = 10$	$n = 10$	$n = 7$
Grand (Overall) Mean = 7.23			

Sample mean

Number of cases in a given sample

samples to be based on the same number of cases. Second, give some thought to what an informal inspection of the data suggests. The levels of unemployment appear to be relatively high in the northern region, but that's also the case in the eastern region. In contrast, the levels of unemployment in the southern and western regions appear to be somewhat lower.

Apparent differences here or there might suggest that it's reasonable to speak in terms of regional variation (at least when it comes to levels of unemployment), but the mere appearance of variation isn't enough in the world of statistical analysis. What's required is a measure of variation that is precise—and that's what the ANOVA procedure is all about. ANOVA allows us to go beyond mere visual inspection of the data and to accurately measure the ratio (F ratio) of between-groups variation to within-groups variation. With ANOVA applied to the problem, we'll be in a position to arrive at a conclusion grounded in measurement.

With all of that as a background, we can begin the calculation of the F ratio. Up to this point, I've been using the term *variation* in a very general sense. As it turns out, what we're actually going to calculate are two estimates of variance. More specifically, we're going to develop an estimate of the between-groups variance and an estimate of the within-groups variance. In other words, the F ratio will be an expression as follows:

$$F \text{ ratio} = \frac{\text{Estimate of between-groups variance}}{\text{Estimate of within-groups variance}}$$

The process used to develop the estimates isn't difficult, but it is a little tedious (particularly if you calculate them by hand, as opposed to relying on a computer and some statistical software). Much of the complexity can be reduced, however, if the process is broken down into its component parts:

1. Calculate what's known as the sums of squares.
2. Convert the sums of squares to estimates of variance.

The process sounds more complicated than it really is, so don't be discouraged. First, we'll approach everything in a step-by-step fashion. Second, the process is remarkably similar to one we encountered earlier, in Chapter 3, when we first encountered the concept of variance. Just as we did in Chapter 3, we'll start with a calculation of the squared deviations—what we refer to in ANOVA as the *sum of squares*.

Calculating the Within-Groups Sum of Squares (SS_W)

My preference is always to begin with the calculation of the **within-groups sum of squares** (SS_W), simply because it is a bit more straightforward than the calculation of the between-groups sum of squares. We begin the process by focusing on the mean of each sample in our study. The mean unemployment level for each region is shown at the bottom of each column in Table 10-1, along with the number of cases.

First, we'll focus on the extent to which the level of unemployment for each city in a particular region deviates, or varies, from the regional mean. For example, we'll look at the extent to which the unemployment level for each city in the southern region deviates from the mean for that region, the extent to which the unemployment level for each city in the northern region deviates from the mean of that region, and so on. We will get a measure of the deviation by subtracting the regional mean from the unemployment level of each city within that region. In other words, we'll get a mathematical expression of the deviation through a simple process of subtraction.

As you learned in Chapter 3, however, the sum of the deviations from the mean always equals 0, so we'll have to square the deviations. Then we'll sum the squared deviations in each region to obtain the sum of squares for each region. In other words, each region will eventually have its own sum of squared deviations. Finally, we'll add up all the sums of squares for all the regions. This total will be the within-groups sum of squares (SS_W).

This portion of the ANOVA calculation is illustrated in Table 10-2. As you can see, the result of the within-groups sum of squares calculation is 179.29 ($SS_W = 179.29$).

I suspect you'll agree that the expression *within-groups sum of squares* is an apt phrase. After all, the process consists of calculating deviations, squaring the deviations, and summing the squared deviations across the different samples. The individual steps in the computation of the SS_W are shown below. Note how these steps correspond to the computations reflected in Table 10-2.

$$
\begin{aligned}
SS_W = {} & (3.8 - 7.90)^2 + (7.1 - 7.90)^2 + (9.6 - 7.90)^2 + (8.4 - 7.90)^2 + (5.1 - 7.90)^2 \\
& + (11.6 - 7.90)^2 + (6.2 - 7.90)^2 + (7.9 - 7.90)^2 + (9.0 - 7.90)^2 + (10.3 - 7.90)^2 \\
& + (4.2 - 5.79)^2 + (6.5 - 5.79)^2 + (4.4 - 5.79)^2 + (8.1 - 5.79)^2 + (7.6 - 5.79)^2 \\
& + (5.8 - 5.79)^2 + (4.0 - 5.79)^2 + (7.3 - 5.79)^2 + (5.2 - 5.79)^2 + (4.8 - 5.79)^2 \\
& + (8.8 - 8.83)^2 + (5.1 - 8.83)^2 + (12.7 - 8.83)^2 + (6.4 - 8.83)^2 + (9.8 - 8.83)^2 \\
& + (6.3 - 8.83)^2 + (10.2 - 8.83)^2 + (8.5 - 8.83)^2 + (11.9 - 8.83)^2 + (8.6 - 8.83)^2 \\
& + (4.8 - 6.03)^2 + (1.2 - 6.03)^2 + (8.0 - 6.03)^2 + (9.4 - 6.03)^2 + (3.6 - 6.03)^2 \\
& + (8.7 - 6.03)^2 + (6.5 - 6.03)^2 \\
= {} & 51.98 + 20.39 + 53.59 + 53.33 \\
= {} & 179.29
\end{aligned}
$$

LEARNING CHECK

Question: What is the symbol for the within-groups sum of squares, and how is it calculated?

Answer: The symbol is SS_W. It is calculated by finding the deviation of each score in a sample from the sample mean, squaring the deviations, adding the squared deviations for each sample, and summing across all the samples.

Table 10-2 Calculating the Within-Groups Sum of Squares

North

X	$X - \bar{X}$	$(X - \bar{X})^2$
3.8	–4.10	16.81
7.1	–0.80	0.64
9.6	1.70	2.89
8.4	0.50	0.25
5.1	–2.80	7.84
11.6	3.70	13.69
6.2	–1.70	2.89
7.9	0.00	0.00
9.0	1.10	1.21
10.3	2.40	5.76
$\Sigma X = 79.0$		$\Sigma(X - \bar{X})^2 = 51.98$
$\bar{X} = \frac{79.0}{10} = 7.90$		$SS_{North} = 51.98$

South

X	$X - \bar{X}$	$(X - \bar{X})^2$
4.2	–1.59	2.53
6.5	0.71	0.50
4.4	–1.39	1.93
8.1	2.31	5.34
7.6	1.81	3.28
5.8	0.01	0.00
4.0	–1.79	3.20
7.3	1.51	2.28
5.2	–0.59	0.35
4.8	–0.99	0.98
$\Sigma X = 57.9$		$\Sigma(X - \bar{X})^2 = 20.39$
$\bar{X} = \frac{57.9}{10} = 5.79$		$SS_{South} = 20.39$

East

X	$X - \bar{X}$	$(X - \bar{X})^2$
8.8	–0.03	0.00
5.1	–3.73	13.91
12.7	3.87	14.98
6.4	–2.43	5.90
9.8	0.97	0.94
6.3	–2.53	6.40
10.2	1.37	1.88
8.5	–0.33	0.11
11.9	3.07	9.42
8.6	–0.23	0.05
$\Sigma X = 88.3$		$\Sigma(X - \bar{X})^2 = 53.59$
$\bar{X} = \frac{88.3}{10} = 8.83$		$SS_{East} = 53.59$

West

X	$X - \bar{X}$	$(X - \bar{X})^2$
4.8	–1.23	1.51
1.2	–4.83	23.33
8.0	1.97	3.88
9.4	3.37	11.36
3.6	–2.43	5.90
8.7	2.67	7.13
6.5	0.47	0.22
$\Sigma X = 42.2$		$\Sigma(X - \bar{X})^2 = 53.33$
$\bar{X} = \frac{42.2}{7} = 6.03$		$SS_{West} = 53.33$

$SS_W = 51.98 + 20.39 + 53.59 + 53.33 = 179.29$

Calculating the Between-Groups Sum of Squares (SS_B)

Now we turn to the between-groups element. The grand mean (7.23) was reported in Table 10-1, along with the mean for each region. To calculate the **between-groups sum of squares** (SS_B), we'll follow a procedure similar to the previous one, but with a slight hitch in the process. Let me explain.

As noted previously, this part of the ANOVA procedure requires that we calculate the deviation (or, more correctly, the squared deviation) of each regional mean from the grand mean and sum those squared deviations across the regions. This will give us our between-groups sum of squares (SS_B). Unfortunately, however, it's not as straightforward as it might appear at first glance. As it turns out, we have to take into account the number of cases that went into the production of each regional mean. In other words, a regional mean based on 10 cases is one thing, but a regional mean based on, let's say, 7 cases is a different matter. Here's why.

We're going to focus on how far each regional mean departs from the grand mean, but we have to start by recognizing that the grand mean was, in part, a function of the total number of cases spread over several regions. Different regions, however, made different contributions to the grand mean. Three regions contributed 10 values or cases each, but another region (the western region) contributed only 7 values or cases. It's only appropriate, therefore, that we take into account the different contribution of each region as we move forward with our calculations. We'll do that by *weighting* our results by the number of cases in each region.

Yes, we're going to subtract the grand mean from the mean of each region to obtain a deviation. Then we're going to square that deviation. But then we're going to weight the result. We do that by multiplying the squared deviation of each region by the number of cases in the region. To better understand this weighting procedure, take a close look at Table 10-3.

As shown in Table 10-3, we subtract the grand mean from each regional mean, square the deviation, and then multiply it by the number of cases in that region. Finally, we sum across the regions to obtain the between-groups sum of squares (SS_B). Remember: We need to take into account the number of cases that were involved in the production of each sample or group mean. Therefore, we weight each group's squared deviation by the number of cases in the group. This important step is one you have to take, even if all the groups or samples have an equal number of cases.

The computations underlying the SS_B are summarized below. My suggestion is that you make a thorough study of those computations, as well as the details of Table 10-3. Once you do that, you'll be in a better position to see how we arrived at a between-groups sum of squares 60.88 ($SS_B = 60.88$).

$$\begin{aligned} SS_B &= 10(7.90 - 7.23)^2 + 10(5.79 - 7.23)^2 + 10(8.83 - 7.23)^2 \\ &\quad + 7(6.03 - 7.23)^2 \\ &= 4.50 + 20.70 + 25.60 + 10.08 \\ &= 60.88 \end{aligned}$$

Table 10-3 Calculating the Between-Groups Sum of Squares

Grand Mean = 7.23

North	South
$SS_B = n(\overline{X} - \overline{X}_{\text{Grand}})^2$	$SS_B = n(\overline{X} - \overline{X}_{\text{Grand}})^2$
$= 10(7.90 - 7.23)^2$	$= 10(5.79 - 7.23)^2$
$= 10(0.67)^2$	$= 10(-1.44)^2$
$= 10(0.45)$	$= 10(2.07)$
$= \mathbf{4.50}$	$= \mathbf{20.70}$

Mean of North = 7.90
$n = 10$

Mean of South = 5.79
$n = 10$

East	West
$SS_B = n(\overline{X} - \overline{X}_{\text{Grand}})^2$	$SS_B = n(\overline{X} - \overline{X}_{\text{Grand}})^2$
$= 10(8.83 - 7.23)^2$	$= 7(6.03 - 7.23)^2$
$= 10(1.60)^2$	$= 7(-1.20)^2$
$= 10(2.56)$	$= 7(1.44)$
$= \mathbf{25.60}$	$= \mathbf{10.08}$

Mean of East = 8.83
$n = 10$

Mean of West = 6.03
$n = 7$

$$SS_B = 4.50 + 20.70 + 25.60 + 10.08 = 60.88$$

LEARNING CHECK

Question: What is the symbol for the between-groups sum of squares, and how is it calculated?

Answer: The symbol is SS_B. It is calculated by finding the deviation of each sample mean from the grand mean, squaring the deviation, weighting the squared deviation for each sample, and summing across all the samples.

Even if you feel totally comfortable with the notion of the between-groups sum of squares concept, let me suggest that you take a short break at this point. We've covered quite a bit. Spend a little time thinking about the sum of squares within (SS_W) and the sum of squares between (SS_B). Concentrate on how you calculated each, and think of these as the first important steps toward the computation of the *F* ratio. Take whatever amount of time is necessary—there's still another important step ahead.

From Sums of Squares to Estimates of Variance

Assuming you took the suggested break, our next task is to transform the two sum of squares elements (SS_B and SS_W) into estimates of variance. It's actually a simple process. All we have to do is divide each sum of squares element (SS_B and SS_W) by an appropriate number of degrees of freedom. The procedure is essentially the same as the calculation of the variance for a sample (as presented in Chapter 3). Let me urge you to review that chapter if you sense you're unsure about any of this. The estimates of variance are referred to as the **mean square between** (MS_B) and the **mean square within** (MS_W). Just to solidify the two in your thinking, they are summarized as follows:

> MS_B = mean square between (an estimate of the between-groups variance)
> MS_W = mean square within (an estimate of the within-groups variance)

LEARNING CHECK

Question: What do the mean square between and mean square within represent?

Answer: The mean square between is the estimate of the between-groups variance. The mean square within is the estimate of the within-groups variance.

Question: What are the symbols for the mean square between and mean square within?

Answer: The symbols are MS_B and MS_W, respectively.

Since the fundamental nature of the ANOVA procedure can sometimes get lost in the midst of different symbols and notations, let's take a moment to review where we've been and where we're going:

1. The goal is to calculate an F ratio.
2. The F ratio is the ratio of an estimate of the between-groups variance to an estimate of the within-groups variance.
3. These estimates are derived through a two-step process.
   a. First, we compute sums of squares (between and within).
   b. Then we transform the sums of squares to estimates of variance (known as mean squares).
4. The F ratio is derived by dividing the estimate of the between-groups variance (MS_B) by the estimate of the within-groups variance (MS_W).

Note that we haven't executed Steps 3b and 4 just yet; those will be our final steps.

Since we already have our within-groups and between-groups sums of squares, our next task is to convert the sums of squares into the mean squares, or estimates of variance. As a prelude to that, a little review of the variance is in order.

The Concept of Variance. Think back to what you learned in Chapter 3 about the variance of a distribution. Recall that the variance allowed us to get around the problem that the sum of the deviations from the mean always equals 0. You'll probably also recall how the variance was computed, both for a population and a sample. You learned that the variance for a population was computed as follows:

$$\text{Variance of a population} = \frac{\sum (X - \mu)^2}{N}$$

Looking carefully at the formula for the population variance, you'll note that the numerator actually amounts to the sum of squared deviations (not unlike the sum of squares we've been discussing so far), and the denominator is simply the number of cases in the population (N).

When it came to the variance of a sample, however, we introduced a slight correction factor. Instead of using N in the denominator, we used $n - 1$ (or the degrees of freedom). The denominator $n - 1$ (degrees of freedom) was used in an effort to arrive at a sample variance that would be a more accurate estimate of the population variance. If your memory is a little faulty on this point, let me suggest you take the time to review the material in Chapter 3. My guess is that it will be important to your understanding of what we encounter next.

Assuming you've taken that time, or you feel secure without the review, let's focus now on the estimates of variance that we're going to develop. First we'll develop an estimate of the between-groups variance. Then we'll develop an estimate of the within-groups variance. Both estimates are developed in much the same way.

First, the **between-groups estimate of variance** (known as the mean square between or MS_B) is derived by dividing the between-groups sum of squares (SS_B) by the appropriate number of degrees of freedom (df_B). Then the **within-groups estimate of variance** (known as the mean square within, or MS_W) is derived by dividing the within-groups sum of squares (SS_W) by the appropriate number of degrees of freedom (df_W). The process can be summarized as follows:

Between-Groups Estimate of Variance

$$\text{Mean Square Between } (MS_B) = \frac{\text{Between-Groups Sum of Squares } (SS_B)}{\text{Between-Groups Degrees of Freedom } (df_B)}$$

Within-Groups Estimate of Variance

$$\text{Mean Square Within } (MS_W) = \frac{\text{Within-Groups Sum of Squares } (SS_W)}{\text{Within-Groups Degrees of Freedom } (df_W)}$$

Obviously, we have to determine the appropriate number of degrees of freedom for each element, so that's where we'll turn now.

Degrees of Freedom. For this portion of the discussion, let's start with the between-groups sum of squares. Think back for a moment to how we computed the between-groups sum of squares (SS_B). If necessary, review the computations outlined in Table 10-3 and the associated discussion. First, we calculated the deviation of each sample mean from the grand or overall mean. Then, we squared the deviations. Next, we multiplied the squared deviation for each sample by the number of cases in each sample. Finally, we summed the squared deviations (multiplied by the number of cases in the sample) across all the samples. The result (the between-groups sum of squares) was 60.88.

The problem we're considering here involves four samples (four regions). In the language of ANOVA, the four samples represent four categories (symbolized by k = 4). The number of degrees of freedom associated with the between-groups estimate of variance (df_B) is $k - 1$. Since we have four categories, the **between-groups degrees of freedom** can be calculated as follows:

$$df_B = k - 1$$
$$df_B = 4 - 1$$
$$df_B = 3$$

LEARNING CHECK

Question: How many degrees of freedom are associated with the between-groups estimate of variance (MS_B)?

Answer: The number of degrees of freedom for MS_B is $k - 1$, where k = the number of categories or samples in the study.

To obtain our between-groups estimate of variance (MS_B), we'll simply divide our between-groups sum of squares (SS_B = 60.88) by the between-groups degrees of freedom (df_B = 3).

$$MS_B = \frac{SS_B}{df_B}$$
$$MS_B = \frac{60.88}{3}$$
$$MS_B = 20.29$$

At this point, you should note how closely this relates to the notion of using $n - 1$ in the computation of the sample variance to obtain an unbiased estimate of the population variance.

> **LEARNING CHECK**
>
> *Question:* How is the between-groups estimate of variance (MS_B) obtained?
>
> *Answer:* The between-groups estimate of variance (MS_B) is obtained by dividing the between-groups sum of squares (SS_B) by the between-groups degrees of freedom (df_B).

Our next step is to develop our within-groups estimate of variance, and we'll use a similar procedure—we'll divide the within-groups sum of squares by an appropriate number of degrees of freedom. Now, of course, the question is how to determine the appropriate number of degrees of freedom for the within-groups element.

The degrees of freedom in the case of the within-groups sum of squares is a function of the total number of cases, as well as the number of samples or categories. In the present instance, we have a total of 37 cases spread over four categories. The appropriate number of degrees of freedom for the estimate of within-groups variance is equal to $n_{total} - k$, or the total number of cases minus the number of categories. With 37 cases and four categories, the **within-groups degrees of freedom** (df_W) can be calculated as follows:

$$df_W = n_{total} - k$$
$$df_W = 37 - 4$$
$$df_W = 33$$

If you take a close look at the formula for the within-groups degrees of freedom ($n_{total} - k$), you'll see that it's actually equal to the sum of the number of cases in each sample minus 1:

$$(n_1 - 1) + (n_2 - 1) + (n_3 - 1) + (n_4 - 1) = 33$$
$$(10 - 1) + (10 - 1) + (10 - 1) + (7 - 1) = 9 + 9 + 9 + 6 = 33$$

> **LEARNING CHECK**
>
> *Question:* How many degrees of freedom are associated with the within-groups estimate of variance (MS_W)?
>
> *Answer:* The number of degrees of freedom for MS_W is $n_{total} - k$ where n_{total} = the number of cases in the study and k = the number of categories or samples in the study.

Having determined that the appropriate number of degrees of freedom for the within-groups sum of squares is equal to 33, we can calculate the within-groups estimate of variance, or MS_W, as follows:

$$MS_W = \frac{SS_W}{df_W}$$

$$MS_W = \frac{179.29}{33}$$

$$MS_W = 5.43$$

LEARNING CHECK

Question: How is the within-groups estimate of variance (MS_W) obtained?

Answer: The within-groups estimate of variance (MS_W) is obtained by dividing the within-groups sum of squares (SS_W) by the within-groups degrees of freedom (df_W).

We've already been through several steps, so let me suggest that you take a look at Table 10-4. This summary table outlines the important elements we've encountered along the way and gives you a look ahead toward the final step.

Calculating the F Ratio

Having developed the estimates of the between-groups variance ($MS_B = 20.29$) and within-groups variance ($MS_W = 5.43$), we're now in a position to calculate the F ratio. This ratio is obtained by dividing the between-groups estimate of variance (MS_B) by the within-groups estimate of variance (MS_W). The calculation is as follows:

$$F = \frac{MS_B}{MS_W}$$

$$F = \frac{20.29}{5.43}$$

$$F = 3.74$$

Table 10-4 Components of ANOVA

Group or Sample Means

Mean of Northern Cities	= 7.90	n = 10
Mean of Southern Cities	= 5.79	n = 10
Mean of Eastern Cities	= 8.83	n = 10
Mean of Western Cities	= 6.03	n = 7

Grand Mean = 7.23

Calculate the Sums of Squares

Between-groups sum of squares (SS_B) = 60.88
Within-groups sum of squares (SS_W) = 179.29

Degrees of Freedom

Between-groups degrees of freedom (df_B) = 3
Within-groups degrees of freedom (df_W) = 33

Divide Sums of Squares by Appropriate Degrees of Freedom to Obtain the Estimates of Variance (the Mean Square Component)

Between-groups estimate of variance, or mean square between (MS_B) = 20.29
Within-groups estimate of variance, or mean square within (MS_W) = 5.43

→ Calculate the F Ratio $\frac{MS_B}{MS_W}$

LEARNING CHECK

Question: How is the F ratio calculated?
Answer: The F ratio is calculated as follows:

$$\frac{MS_B}{MS_W}$$

So, the calculated F ratio (our test statistic) equals 3.74. We have a final answer—but what does it really mean? By now, you should find yourself in very familiar territory. After all, it's really just another hypothesis-testing situation.

The Interpretation

As we've already done in what probably seems like countless situations before, we find ourselves looking at a calculated test statistic—in this case, an F ratio. But now the question is whether or not the F ratio is significant. As before, the answer turns on the critical value. If our calculated test statistic (the calculated value of *F*) meets or exceeds the critical value, we have significant results, and we can reject the null hypothesis. If, on the other hand, our calculated test statistic falls below the critical value, we'll fail to reject the null hypothesis.

Interpretation of the F Ratio

Our next task, then, is to locate the critical value. For that information, we turn to Appendix D: Distribution of *F* at the .05 Level of Significance (the level of significance that we selected at the outset). Once again, to use the table we have to take into account our degrees of freedom. We know that the degrees of freedom associated with the between-groups estimate of variance is 3, and the degrees of freedom associated with the within-groups estimate of variance is 33.

If you take a close look at Appendix D, you'll note that the degrees of freedom for the between-groups variance element (the numerator in the *F* ratio) are listed across the top row of the table. The degrees of freedom for the within-groups variance element (the denominator in the *F* ratio) are listed in the first column. Once we've identified the appropriate degrees of freedom in the top row and first column, we locate the point at which the two intersect in the table. Note, however, that there is no listing for 33 degrees of freedom. At this point, you should recall our earlier rule of thumb (noted in Appendix B)—namely, find the next lower number of degrees of freedom. Therefore, you should use the value associated with 30 degrees of freedom. That value—2.92—is our appropriate critical value.

All that remains is to compare our calculated *F* ratio to the critical value. As it turns out, our calculated *F* value of 3.74 exceeds the critical value. Therefore, we reject the null hypothesis. As before, we're rejecting the null hypothesis with a known probability of having committed a Type I or alpha error (.05).

In rejecting the null hypothesis, we move a step toward suggesting that levels of unemployment in cities do vary by region. That, of course, is another way of saying it's probably legitimate to think in terms of a regional classification scheme when speaking about levels of unemployment.

Had we failed to achieve significant results, however, we would have failed to reject the null hypothesis. Since the null was a statement that the means would be equal, failing to reject the null would be tantamount to saying that there is no significant variation across the regions. In that case, it wouldn't make much sense to speak in terms of regional variation.

Whatever the final outcome of an ANOVA application, it's always important to keep in mind what the bigger picture is all about. As we've done before, we return to the central notion that what we're really interested in are populations—not samples. In this instance, our interest was in the population of all cities in the northern region, all cities in the southern region, all cities in the eastern region, and all cities in the western region. That we had sample data to work with was important in reaching our final goal, but we were ultimately interested in the larger picture.

Working at the .05 level of significance, we determined that we could reject the null hypothesis. Is it possible that the results of the sample data gave us a false picture? Yes, of course that's possible. There's always a chance that the ratio of our estimates was the result of sampling error and that the calculated *F* ratio isn't a reflection of what's really going on in the population. If that's what happened, then we would have made a Type I error.

As we know all too well, however, we'll never know if that was the case. That's just the way it is, and there's no getting around it (short of collecting data on all cities). We always have to live with the chance of a Type I error. On the positive side, however, we always know the probability that we've made such an error. In the case of our example, it was only 5 times out of 100.

Had we wanted to, we could have set our level of significance at .01, and that would have reduced the probability of a Type I error. In fact, that's what Appendix E is all about; it shows the Distribution of *F* at the .01 Level of Significance. A quick check of the appropriate critical value in Appendix E would tell us that our results were not significant at the .01 level. In other words, had we been working at the .01 level of significance, we would have failed to reject the null hypothesis. In this case, however, we were working at the .05 level of significance, and our results were significant. As a result, we were in position to reject the null hypothesis.

If you think about the ANOVA procedure for any length of time, you're apt to conclude that it only gives us a general picture regarding the null hypothesis. ANOVA allows us to determine whether or not there's a significant difference across groups or samples, but it doesn't tell us much about the specific nature of any difference. As Gravetter and Wallnau (1999, p. 338) note:

> When you reject the null hypothesis, you conclude that the means are not all the same. Although this appears to be a simple conclusion, in most cases it actually creates more questions than it answers.

To better understand that observation, think back to our interpretation of the *F* ratio in the problem we just considered. We found a significant *F* ratio, but the conclusion left the door open to further questioning. Recall how the conclusion was phrased: It is probably legitimate to think in terms of a regional classification scheme when speaking about levels of unemployment. But questions still remain as to what produced the significant *F* ratio in the first place. To get the answers to those questions, a statistician typically turns to post hoc testing procedures.

Post Hoc Testing

As the expression implies, post hoc testing allows us to go beyond the determination that we have a significant *F* ratio. As noted previously, a significant *F* ratio does not necessarily mean that there was a significant difference between all means when examined in terms of all possible combinations. Maybe the difference between the means of the first and second samples was so large that it had a major impact on the calculation of the between-groups variation. On the other hand, maybe it was an unusually large difference between the means of the third and fourth samples. Maybe the significant results derived from noticeable differences between *all* the means. In short, having significant results is one thing; understanding the origin of the significance is another.

Fortunately, procedures are available that allow us to peel back the findings, so to speak, and gain a better understanding of which differences of means were really responsible for the final F value. Tukey's Honestly Significant Difference (HSD) is such a procedure. It is considered a post hoc test, in that it's employed after significant results are found. In short, Tukey's HSD allows us to determine where the significant differences between individual means are to be found.

The HSD procedure involves the calculation of what's known as the Q statistic. It rests on a pair-by-pair comparison of sample means. In the example used throughout this chapter, we have four samples and, therefore, four sample means. The HSD procedure applied to our problem would involve the following six comparisons:

Mean of Sample 1	and	Mean of Sample 2
Mean of Sample 1	and	Mean of Sample 3
Mean of Sample 1	and	Mean of Sample 4
Mean of Sample 2	and	Mean of Sample 3
Mean of Sample 2	and	Mean of Sample 4
Mean of Sample 3	and	Mean of Sample 4

The calculation of Q is fairly straightforward. It is calculated once for each comparison, so in this case, Q will be calculated six times.

For each comparison, we calculate the absolute difference (the difference without regard to positive or negative sign) between the two sample means. This absolute difference becomes the numerator in the test statistic (Q). The denominator is partly a function of the MS_W that was calculated in the ANOVA procedure. There are actually two different ways to calculate the denominator of the Q statistic. One version is for situations in which the sample sizes are equal; the other version is appropriate for ANOVA applications with unequal sample sizes. The example we considered in this chapter was based on unequal sample sizes, but here are both formulas.

When All Sample Sizes Are Equal

$$Q = \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{\dfrac{MS_W}{n}}}$$

Where $\overline{X}_1$ and $\overline{X}_2$ are any two means and n represents the number of cases in each sample.

When Any Two Samples Sizes Are Unequal

$$Q = \frac{|\overline{X}_1 - \overline{X}_2|}{\sqrt{\dfrac{MS_W}{\tilde{n}}}}$$

Where $\overline{X}_1$ and $\overline{X}_2$ are any two means and $\tilde{n}$ represents the harmonic mean sample size. The harmonic mean is calculated as follows:

$$\tilde{n} = \frac{k}{\dfrac{1}{n_1} + \dfrac{1}{n_2} + \dfrac{1}{n_3} + \dfrac{1}{n_4}}$$

Since our example involves unequal sample sizes, we will use the second formula for Q. We'll start by calculating the harmonic mean. Recall that the formula for the harmonic mean is as follows:

$$\tilde{n} = \frac{k}{\dfrac{1}{n_1} + \dfrac{1}{n_2} + \dfrac{1}{n_3} + \dfrac{1}{n_4}}$$

Since we have four samples (groups or categories), the harmonic mean is calculated as follows:

$$\begin{aligned}
\tilde{n} &= \frac{k}{\dfrac{1}{n_1} + \dfrac{1}{n_2} + \dfrac{1}{n_3} + \dfrac{1}{n_4}} \\
&= \frac{4}{\dfrac{1}{10} + \dfrac{1}{10} + \dfrac{1}{10} + \dfrac{1}{7}} \\
&= \frac{4}{.10 + .10 + .10 + .14} \\
&= \frac{4}{0.44} \\
&= 9.09
\end{aligned}$$

Armed with the value of the harmonic mean ($\tilde{n} = 9.09$), our next step is to bring in the mean square within (MS_W). Our previous ANOVA computations tell us that $MS_W = 5.43$. We now divide the MS_W (5.43) by the harmonic mean (9.09) and take the square root of the result. This gives us the denominator for our calculation of Q.

$$\begin{aligned}
\text{Denominator in } Q \text{ calculation} &= \sqrt{\frac{MS_W}{\tilde{n}}} \\
&= \sqrt{\frac{5.43}{9.09}} \\
&= \sqrt{0.60} \\
&= 0.77
\end{aligned}$$

Having calculated the denominator for our Q statistic, we can move through the remainder of the computations with relative ease. For each comparison, it is simply a matter of finding the absolute difference between two means, treating that value as the numerator, and dividing by the denominator that we just calculated. The steps in the process are summarized in Table 10-5.

As the summary indicates, we now have six Q values, or six different calculated Q test statistics. Each calculated Q test statistic relates to a particular comparison of means. Now all that remains is to examine whether or not the Q test statistic in question is significant for each individual comparison. That brings us to the matter of the critical value for Q—the value against which we will evaluate the individual Qs that we've calculated.

Appendix F provides the critical values of Q at the .05 level of significance. (If we were working at the .01 level of significance, we would use Appendix G.) The numbers across the top of the table refer to the number of groups or samples in the ANOVA that preceded application of the HSD measure. Since our problem is based on four samples or groups (northern, southern, eastern, and western cities), our focus will be on the column labeled 4. The within-groups degrees of freedom (df_W) is something we dealt with earlier. You will recall that the appropriate number of degrees of freedom for the within-groups element was $n - k$, or the total number of cases (37) minus the number of categories or groups (4). Therefore, the number of degrees of freedom within is 37 – 4, or 33. As before, the table has no entry for 33 degrees of freedom, but we are safe in using the row for 30 degrees of freedom. The entry associated with 30 degrees of freedom and four samples is 3.85—and that becomes our critical value.

Now all that remains is to compare the various Q values that we calculated to the critical value of Q found in Appendix F. The results are shown in Table 10-6.

Having calculated Q for each comparison and having checked each against the critical value (3.85), we determine that the only significant difference is found between the southern region and the eastern region. It is not the case

Table 10-5 Calculation of Q for Tukey's HSD

Possible Comparisons	$\lvert \bar{X}_1 - \bar{X}_2 \rvert$	$Q = \dfrac{\lvert \bar{X}_1 - \bar{X}_2 \rvert}{\sqrt{\dfrac{MS_W}{\tilde{n}}}}$
North and South	|7.90 – 5.79| = 2.11	2.11/0.77 = 2.74
North and East	|7.90 – 8.83| = 0.93	0.93/0.77 = 1.21
North and West	|7.90 – 6.03| = 1.87	1.87/0.77 = 2.43
South and East	|5.79 – 8.83| = 3.04	3.04/0.77 = 3.95
South and West	|5.79 – 6.03| = 0.24	0.24/0.77 = 0.31
East and West	|8.83 – 6.03| = 2.80	2.80/0.77 = 3.64

Table 10-6 Interpreting Tukey's HSD

Possible Comparisons	Q	Results
North and South	2.74	Not Significant
North and East	1.21	Not Significant
North and West	2.43	Not Significant
South and East	3.95	***Significant***
South and West	0.31	Not Significant
East and West	3.64	Not Significant

that there is a significant difference across all regions. Rather, the significant difference is found only between two regions.

A finding like that would, no doubt, send us back to the drawing board, at least when it comes to the matter of a regional classification scheme. In a real-life situation, now would be the time to consider other types of regional classifications—maybe, for example, one that rests on only three designated regions of the country.

Questions like that are for another time and place. It's time to bring our discussion of one-way ANOVA to a close. Before leaving the topic, though, it might be useful to review several points and to underscore a few things you may want to think about.

- Think about ANOVA as being appropriate in situations involving three or more samples, provided you have interval/ratio level data to work with.
- Think about the fact that ANOVA can be appropriate even if the samples have an unequal number of cases.
- Think about the F ratio as a ratio of two estimates of variance—the estimate of variance between groups and the estimate of variance within groups.
- Think about how the computation of the F ratio is essentially a two-step process—first the calculation of between and within sums of squares, and then a transformation of the sums of squares into estimates of variance.
- Think about how degrees of freedom come into play in the transformation of the sums of squares into the estimates of variance, with $k - 1$ degrees of freedom for the estimate between, and $n - k$ for the estimate within.

At the conclusion of this chapter you'll find several problems to consider. Some problems direct you to calculate the F ratio from beginning to end. Many of the problems, though, just pose questions about the component parts of the ANOVA procedure. Others give you the component parts of the ANOVA procedure; your job is to finish the calculations and provide an appropriate interpretation. My guess is that you'll find the questions sufficient to shore up your understanding of the topic.

Chapter Summary

With your introduction to ANOVA, you have been exposed to a widely used statistical procedure. Ideally, you have gained an understanding of why many statisticians think of it as an extension of the two-sample difference of means tests and why it can be thought of as a procedure that tests the legitimacy of a classification scheme. By the same token, you should have developed an understanding of why the procedure carries the name of *analysis of variance,* inasmuch as the F ratio is based on two estimates of variance.

As to the specific components of the ANOVA procedure, you encountered the concepts of the between- and within-groups sums of squares, as well as the between- and within-groups estimates of variance. You also developed an appreciation for the F ratio as an expression of the ratio of the two estimates of variance.

Beyond all of that, however, was an unstated lesson I hope you discovered along the way—namely, that the process of learning statistical applications gets easier and easier. It's true that different research situations call for different procedures. It's true that different procedures rest on different logical foundations and different calculations. But beyond that, the process of testing a null hypothesis remains fundamentally the same from application to application. State the null; set the level of significance; calculate the test statistic; compare the test statistic to a critical value; state a conclusion. As I mentioned before, you keep encountering the same process, over and over and over again.

Some Other Things You Should Know

There are still a few more things you should be aware of in connection with ANOVA. In a sense, we've just scratched the surface of ANOVA, so let me mention a few related matters.

As noted at the outset, the ANOVA procedure we considered in this chapter is technically known as one-way analysis of variance. It is referred to as one-way ANOVA because it is used in problems that deal with the relationship between one variable and another variable. For example, we dealt with an application that examined the variation in levels of unemployment (one variable) by region (the other variable).

A more complex application of ANOVA is available, however. For example, let's say we wanted to look at how levels of unemployment vary by region *and* type of city (manufacturing, retail, service, or other). In that case, we could opt for a two-way ANOVA application. The procedure is referred to as two-way ANOVA because it looks at how one variable varies on the basis of two other variables. To take another example, we might be interested in how student test scores vary by teaching method (lecture only, lecture plus discussion) and gender composition of the class (all-male classes, all-female classes, and combined male/female classes). This research question would also be suited for a two-way

ANOVA application. For an excellent discussion of the two-way ANOVA procedure, consult Pagano (2001).

Second, the ANOVA application that we just considered was based on the assumption that the samples were independent random samples. As was the case with the difference of means tests, a modified ANOVA procedure is available when the samples under consideration are matched or related. For a discussion of that application, see Dunn (2001).

Finally, the Tukey's HSD test that we considered is only one of a variety of post hoc test procedures that are available for use following an ANOVA application. The selection of one post hoc test over another is usually a function of several considerations. Discussions of various post hoc options are typically found in more advanced texts.

Key Terms

ANOVA (one-way)
between-groups degrees of freedom
between-groups estimate of variance
between-groups sum of squares (SS_B)
F ratio
grand (overall) mean
group (sample) mean
mean square between (MS_B)
mean square within (MS_W)
within-groups degrees of freedom
within-groups estimate of variance
within-groups sum of squares (SS_W)

Chapter Problems

Fill in the blanks, calculate the requested values, or otherwise supply the correct answer.

General Thought Questions

1. The calculated test statistic for ANOVA is known as the _______ ratio.
2. The *F* ratio is the ratio of the amount of variation _________ the groups to the amount of variation _________ the groups.
3. Explain how to calculate the within-groups sum of squares.
4. Explain how to calculate the between-groups sum of squares.
5. The between-groups sum of squares is transformed into an estimate of the between-groups variance by dividing the between-groups sum of squares by an appropriate number of ______________.
6. The within-groups sum of squares is transformed into an estimate of the within-groups variance by dividing the within-groups sum of squares by an appropriate number of ______________.
7. The formula for the number of degrees of freedom for the within-groups estimate of variance is ______, where *n* equals the total number of cases under consideration.

8. The formula for the number of degrees of freedom for the between-groups estimate of variance is ________, where *k* equals the number of groups or samples under consideration.
9. Another name for the between-groups estimate of variance is ________.
10. Another name for the within-groups estimate of variance is ________.
11. If you had a research problem appropriate for ANOVA and it was based on the results from three samples, what would be the null hypothesis?

Application Questions/Problems

1. Assume you had a research problem appropriate for ANOVA that was based on six samples and a total of 36 cases.
   a. How many degrees of freedom would be associated with the between-groups estimate of variance?
   b. How many degrees of freedom would be associated with the within-groups estimate of variance?
2. Assume the following:
   .05 level of significance; five samples; 21 cases; $SS_B = 26$; $SS_W = 29$
   a. Calculate the *F* ratio.
   b. What is the critical value?
   c. What would you conclude?
3. Assume the following:
   .05 level of significance; four samples; 30 cases; $SS_B = 80$; $SS_W = 258$
   a. Calculate the *F* ratio.
   b. What is the critical value?
   c. What would you conclude?
4. Assume the following:
   .05 level of significance; three samples; 27 cases $SS_B = 13$; $SS_W = 23$
   a. Calculate the *F* ratio.
   b. What is the critical value?
   c. What would you conclude?
5. Consider the following research data:

Sample 1	Sample 2	Sample 3
10	6	5
10	7	10
9	2	8
11	8	8
6	9	8
11	5	9
9	3	6
7	8	10
4		12
5		14
6		

a. State an appropriate null hypothesis.
b. What are the values of each category mean?
c. What is the value of the grand mean?
d. What is the value of the SS_B?
e. What is the value of the SS_W?
f. What is the value of the df_B?
g. What is the value of the df_W?
h. What is the value of the MS_W?
i. What is the value of the MS_B?
j. What is the value of F?
k. Assuming that you were working at the .05 level of significance, what would you conclude?

6. An evaluation survey, designed to measure perceived program effectiveness, was administered to a sample of 39 citizens who attended a community crime-prevention meeting. Using a scale of 0 to 10, the respondents were asked to rate the meeting in terms of effectiveness in presenting useful information. The responses were analyzed, based upon the place of residence of the respondent—northern sector, southern sector, eastern, or western sector—and the following results were found.

Northern	Southern	Eastern	Western
2	3	4	2
4	5	2	5
6	7	8	6
3	1	7	7
5	4	7	2
1	5	6	4
7	3	8	5
1	4	6	6
4		6	6
5			6
6			6

a. State an appropriate null hypothesis.
b. What are the values of each category mean?
c. What is the value of the grand mean?
d. What is the value of the SS_B?
e. What is the value of the SS_W?
f. What is the value of the df_B?
g. What is the value of the df_W?
h. What is the value of the MS_W?
i. What is the value of the MS_B?
j. What is the value of F?
k. Assuming you were working at the .05 level of significance, what would you conclude?

7. An industrial psychologist has examined the levels of absenteeism (measured in terms of days absent per year) of workers in three different work environments (morning shift, afternoon shift, and night shift). The results of the study are summarized as follows:

Day Shift	Afternoon Shift	Night Shift
3	6	5
4	4	6
3	5	4
5	4	3
7		5
$n = 5$	$n = 4$	$n = 5$

a. State an appropriate null hypothesis.
b. What are the values of each category mean?
c. What is the value of the grand mean?
d. What is the value of the SS_B?
e. What is the value of the SS_W?
f. What is the value of the df_B?
g. What is the value of the df_W?
h. What is the value of the MS_W?
i. What is the value of the MS_B?
j. What is the value of F?
k. Assuming you were working at the .05 level of significance, what would you conclude?

8. A social psychologist has been studying the relationship between group composition and level of cooperation on the part of preschool children in a task-completion exercise. Each group is observed, and the number of cooperative acts exhibited by each member of the group is recorded. Three types of groups are under study: all male, all female, and mixed (both male and female members). Results of the investigation are as follows:

All Male	All Female	Mixed Gender
4	6	3
4	9	5
3	8	6
1	4	4
3	8	7
4	8	6
$n = 6$	$n = 6$	$n = 6$

a. State an appropriate null hypothesis.
b. What are the values of each category mean?
c. What is the value of the grand mean?
d. What is the value of the SS_B?
e. What is the value of the SS_W?
f. What is the value of the df_B?
g. What is the value of the df_W?
h. What is the value of the MS_W?
i. What is the value of the MS_B?
j. What is the value of F?
k. Assuming you were working at the .05 level of significance, what would you conclude?