

EXAMPLE (a):

Major roasters and distributors of coffee in the United States have long felt the great uncertainty in the price of coffee beans. Over the course of one year, for example, coffee futures prices went from a low of \$1.40 per pound up to \$2.50 and then down to \$2.03. The main reason for such wild fluctuations in price, which strongly affect the performance of coffee distributors, is the constant danger of drought in Brazil. Since Brazil produces 30% of the world's coffee, the market for coffee beans is very sensitive to the annual rumors of impending drought.³

Recently a domestic coffee distributor decided to avert the problem altogether by eliminating Brazilian coffee from all blends the company distributes. Before taking such action, the distributor wanted to minimize the chances of suffering losses in sales volume. Therefore, the distributor hired a marketing research firm to conduct a statistical test of consumers' taste preferences. The research firm made arrangements with several large restaurants to serve randomly chosen groups of their customers different kinds of after-dinner coffee. Three kinds of coffee were served: a group of 21 randomly chosen customers were served pure Brazilian coffee; another group of 20 randomly chosen customers were served pure Columbian coffee; and a third group of 22 randomly chosen customers were served pure African-grown coffee.

This is the *completely randomized design* part of the name of the ANOVA technique we mentioned at the end of the last section. In completely randomized design, the experimental units (in this case, the people involved in the experiment) are randomly assigned to the three treatments, the treatment being the kind of coffee they are served. Later in this chapter, we will encounter other designs useful in many situations. To prevent a response bias, the people in this experiment were not told the kind of coffee they were being served. The coffee was listed as a "house blend."

Suppose that data for the three groups were consumers' ratings of the coffee on a scale of 0 to 100 and that certain computations were carried out with these data (computations will be discussed in the next section), leading to the following value of the ANOVA test statistic: $F = 2.02$. Is there evidence to conclude that any of the three kinds of coffee leads to an average consumer rating different from the other two kinds?

SOLUTION:

The null and alternative hypotheses here are, by Equation 8-1,

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

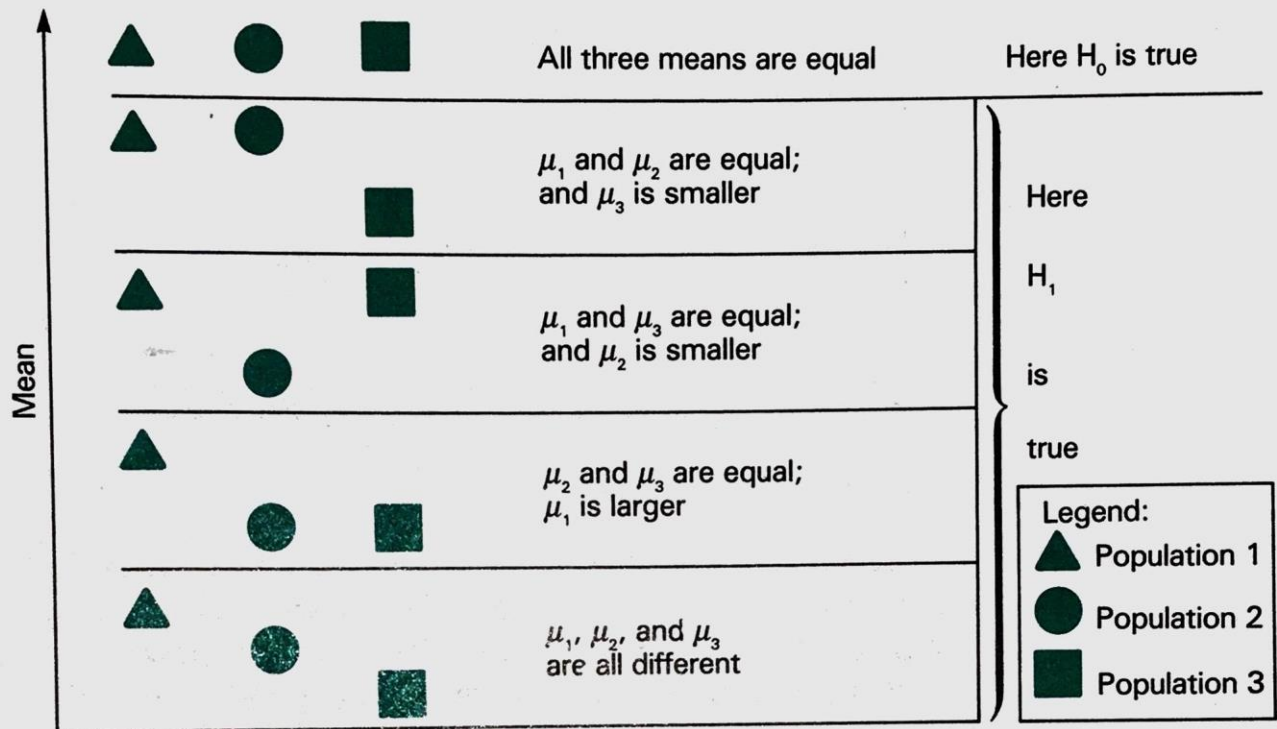
$$H_1: \text{Not all three } \mu_i \text{ are equal}$$

Let us examine the meaning of the null and alternative hypotheses in this example. The null hypothesis states that average consumer responses to each of the three kinds of coffee are equal. The alternative hypothesis says that not all three population means are equal. What are the possibilities covered under the alternative hypothesis? The possible relationships among the relative magnitudes of any three real numbers μ_1 , μ_2 , and μ_3 are shown in Figure 8-3.

As you can see from Figure 8-3, the alternative hypothesis is composed of several different possibilities—it includes all the cases where *not all* three means are equal. Thus, if we reject the null hypothesis, all we know is that there is statistical evidence to conclude that not all three population means are equal. However, we do not know in what way the means are

³ See "Coffee Nerves that May Not Last Long," *Business Week* (September 8, 1986).

FIGURE 8-3 The Possible Relationships among the Relative Magnitudes of the Three Population Means μ_1 , μ_2 , and μ_3



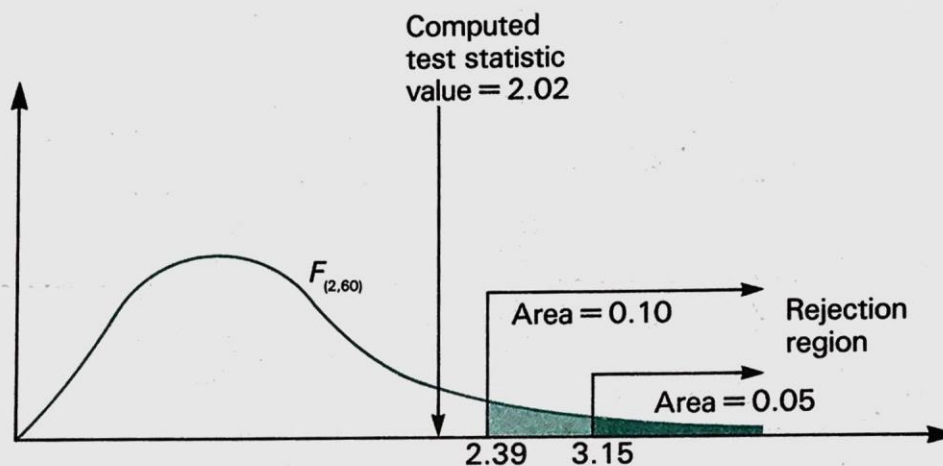
different. Therefore, once we reject the null hypothesis, we need to conduct further analysis to determine which population means are different from one another. The further analysis following ANOVA will be discussed in a later section.

We have a null hypothesis and an alternative hypothesis. We also assume that the conditions required for ANOVA are met; that is, we assume that the three populations of consumer responses are (approximately) normally distributed with equal population variance. Now we need to conduct the test.

Since there are three populations, or treatments, under study, the degrees of freedom for the numerator are $r - 1 = 3 - 1 = 2$. Since the total sample size is: $n = n_1 + n_2 + n_3 = 21 + 20 + 22 = 63$, we find that the degrees of freedom for the denominator are $n - r = 63 - 3 = 60$. Thus, when the null hypothesis is true, our test statistic has an F distribution with 2 and 60 degrees of freedom: $F_{(2,60)}$. From Appendix C, Table 5, we find that the right-hand-tail critical point at $\alpha = 0.05$ for an F distribution with 2 and 60 degrees of freedom is $C = 3.15$. Since the computed value of the test statistic is equal to 2.02, we may conclude that at the 0.05 level of significance there is insufficient evidence to conclude that the three means are different. The null hypothesis that all three population means are equal must be accepted. Since the critical point for $\alpha = 0.10$ is 2.39, we find that the p -value is greater than 0.10.

Based on our data, there is no evidence that consumers tend to prefer the Brazilian coffee to the other two brands. The distributor may substitute one of the other brands for the price-unstable Brazilian coffee. Note that we usually prefer to make conclusions based on the *rejection* of a null hypothesis because acceptance is often considered a weak conclusion. The results of our test are shown in Figure 8-4.

FIGURE 8-4 Carrying out the Test of Example (a)



In this section, we have seen the basic elements of the hypothesis test underlying analysis of variance: the null and alternative hypotheses, the required assumptions, the test statistic, and the decision rule. We have not, however, seen how the test statistic is computed from the data or the reasoning behind its computation. The theory and the computations of ANOVA are explained in the following sections.

PROBLEMS

- 8-1. Four populations are compared by analysis of variance. What are the possible relations among the four population means covered under the null and alternative hypotheses?
- 8-2. What are the assumptions of ANOVA?
- 8-3. Three methods of training managers are to be tested for relative effectiveness. The management training institution proposes to test the effectiveness of the three methods by comparing two methods at a time, using a paired t test. Explain why this is a poor procedure.
- 8-4. In an analysis of variance comparing the output of five plants, data sets of 21 observations per plant are analyzed. The computed F statistic value is 3.6. Do you believe that there are differences in average output among the five plants? What is the approximate p -value? Explain.
- 8-5. A real estate development firm wants to test whether there are differences in the average price of a lot of a given size in the center of each of four cities: Philadelphia, New York, Washington, and Baltimore. Random samples of 52 lots in Philadelphia, 38 lots in New York, 43 lots in Washington, and 47 lots in Baltimore lead to a computed test statistic value of 12.53. Do you believe that average lot prices among the four cities are equal? How confident are you of your conclusion? Explain.

THE THEORY AND THE COMPUTATIONS OF ANOVA:

Recall that the purpose of analysis of variance is to detect differences among several population means based on evidence provided by random samples from these populations. How can this be done? We want to compare r population means. We use r

random samples, one from each population. Each random sample has its own mean. The mean of the sample from population i will be denoted by $\bar{x}_i$. We may also compute the mean of all data points in the study, regardless of which population they come from. The mean of all the data points (when all data points are considered a single set) is called the *grand mean* and is denoted by $\bar{\bar{x}}$. These means are given by the following equations.

The mean of sample i ($i = 1, \dots, r$):

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{i,j}}{n_i} \quad (8-3)$$

The **grand mean**, the mean of all the data points:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^r \sum_{j=1}^{n_i} x_{i,j}}{n} \quad (8-4)$$

where $x_{i,j}$ is the particular data point in position j within the sample from population i . The subscript i denotes the population, or treatment, and runs from 1 to r . The subscript j denotes the data point within the sample from population i ; thus, j runs from 1 to n_i .

In Example (a), $r = 3$, $n_1 = 21$, $n_2 = 20$, $n_3 = 22$, and $n = n_1 + n_2 + n_3 = 63$. The third data point (person) in the group of 21 people who consumed Brazilian coffee is denoted by $x_{1,3}$ (that is, $i = 1$ denotes treatment 1 and $j = 3$ denotes the third point in that sample).

We will now define the main principle behind the analysis of variance.

If the r population means are different (that is, at least two of the population means are *not* equal), then it is likely that the variation of the data points about their respective sample means, $\bar{x}_i$, will be *small* when compared with the variation of the r sample means about the grand mean, $\bar{\bar{x}}$.

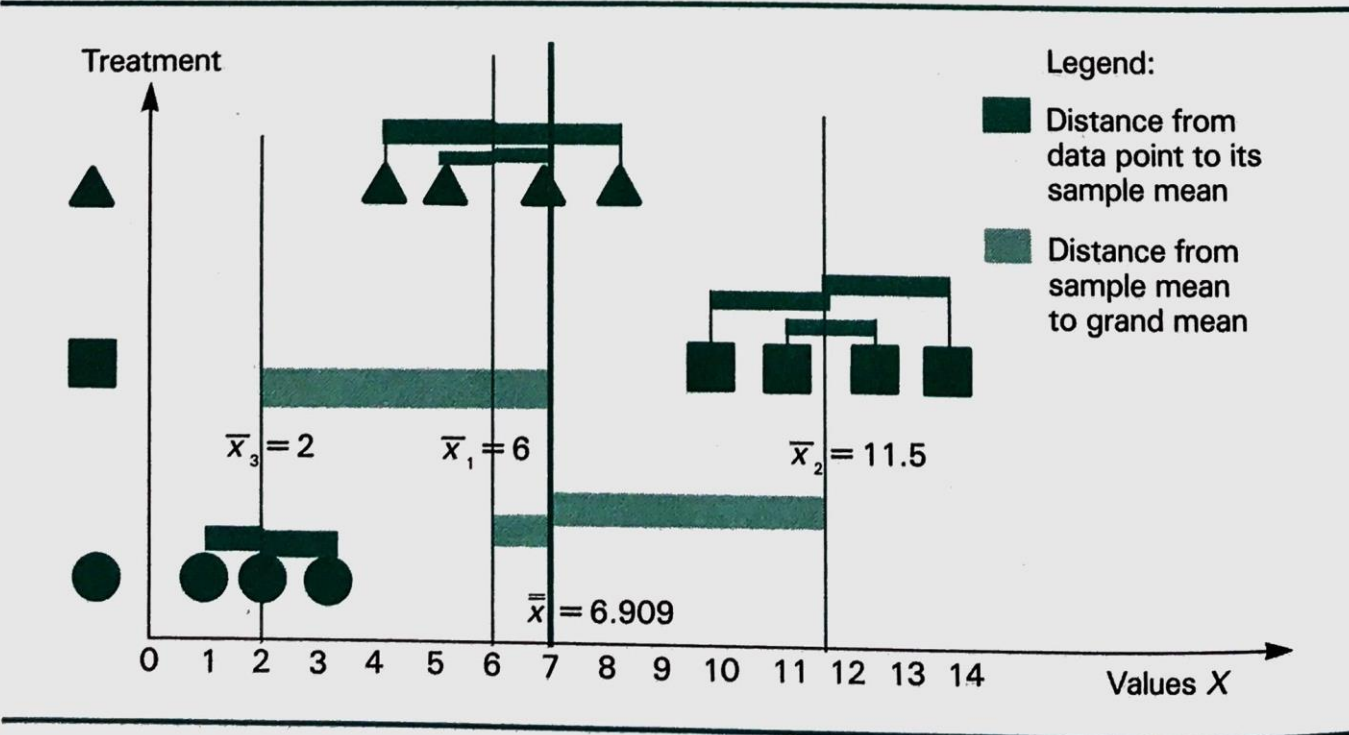
We will demonstrate the ANOVA principle using three hypothetical populations, which we will call the triangles, the squares, and the circles. Table 8-1 gives the values of the sample points from the three populations. For demonstration purposes, we use very small samples. In real situations, the sample sizes should be much larger. The data given in Table 8-1 are shown in Figure 8-5. The figure also shows the deviations of the data points from their sample means and the deviations of the sample means from the grand mean.

TABLE 8-1 Data and the Various Sample Means for Triangles, Squares, and Circles

Treatment (<i>i</i>)	Sample Point (<i>j</i>)	Value ($x_{i,j}$)
<i>i</i> = 1 Triangle	1	4
Triangle	2	5
Triangle	3	7
Triangle	4	8
Mean of triangles		6
<i>i</i> = 2 Square	1	10
Square	2	11
Square	3	12
Square	4	13
Mean of squares		11.5
<i>i</i> = 3 Circle	1	1
Circle	2	2
Circle	3	3
Mean of circles		2
Grand mean of all data points		6.909

Look carefully at Figure 8-5. Note that the *average* distance (in absolute value) of data points from their respective group means (that is, the average distance, in absolute value, of a triangle from the mean of the triangles, $\bar{x}_1$; and similarly for the squares and the circles) is *relatively small* compared with the average distance (in absolute value) of the three sample means from the grand mean. If you are not convinced of this, note that there are only three distances of sample means to the grand mean (in the computation, each distance is weighted by the actual number of points in the group), and that only one of them, the smallest distance—that of $\bar{x}_1$ to $\bar{\bar{x}}$ —is

FIGURE 8-5 The Deviations of the Triangles, Squares, and Circles from Their Sample Means and the Deviations of the Sample Means from the Grand Mean



of the relative magnitude of the distances between the data points and their respective sample means. The two other distances are much greater; hence, the average distance of the sample means from the grand mean is greater than the average distance of all data points to their respective sample means.

The *average* deviation from a mean is zero. We talk about the average absolute deviation—actually, we will use the average *squared* deviation—to prevent the deviations from canceling out. This should remind you of the definition of the sample variance in Chapter 0. Now let us define some terms that will make our discussion simpler.

We define an **error deviation** as the difference between a data point and its sample mean. Errors are denoted by e , and we have:

$$e_{i,j} = x_{i,j} - \bar{x}_i \quad (8-5)$$

Thus, all the distances from the data points to their sample means in Figure 8–5 are errors (some are positive, and some are negative). The reason these distances are called errors is that they are unexplained by the fact that the corresponding data points belong to population i . The errors are assumed to be due to natural variation, or pure randomness, within the sample from treatment i .

On the other hand,

We define a **treatment deviation** as the deviation of a sample mean from the grand mean. Treatment deviations, t_i , are given by:

$$t_i = \bar{x}_i - \bar{\bar{x}} \quad (8-6)$$

The ANOVA principle thus says:

When the population means are not equal, the “average” error is relatively small compared with the “average” treatment deviation.

Again, if we actually averaged all the deviations, we would get zero. Therefore, when we apply the principle computationally, we will square the error and treatment deviations before averaging them. This way, we will maintain the relative (squared) magnitudes of these quantities. The averaging process is further complicated because we have to average based on degrees of freedom (recall that degrees of freedom were used in the definition of a sample variance). For now, let the term *average* be used in a simplified, intuitive sense.

Since we noted that the average error in our triangle-square-circle example looks

small relative to the average treatment deviation, let us see what the populations that brought about our three samples look like. Figure 8–6 shows the three populations, assumed normally distributed with equal variance. (This can be seen from the equal width of the three normal curves. Note also that the three samples seem to have equal dispersion about their sample means.) The figure also shows that the three population means are not equal.

Figure 8–7, in contrast, shows three samples of triangles, squares, and circles in which the average error deviation is of about the same magnitude (*not* smaller than) as the average treatment deviation. As can be seen from the superimposed normal populations from which the samples have arisen in this case, the three population means, μ_1 , μ_2 , and μ_3 , are all equal. Compare the two figures to convince yourself of the ANOVA principle.

The Sum of Squares Principle

We have seen how, when the population means are different, the error deviations in the data are small when compared with the treatment deviations. We made general statements about the average error being small when compared with the average treatment deviation. The error deviations measure how close the data *within* each group are to their respective group means. The treatment deviations measure the distances *between* the various groups. It therefore seems intuitively plausible (as seen in Figures 8–5 to 8–7) that when these two kinds of deviations are of about equal magnitude, the population means are about equal. Why? Because when the average

FIGURE 8–6 The Samples of Triangles, Squares, and Circles and Their Respective Populations (the Three Populations Are Normal with Equal Variance but with Different Means)

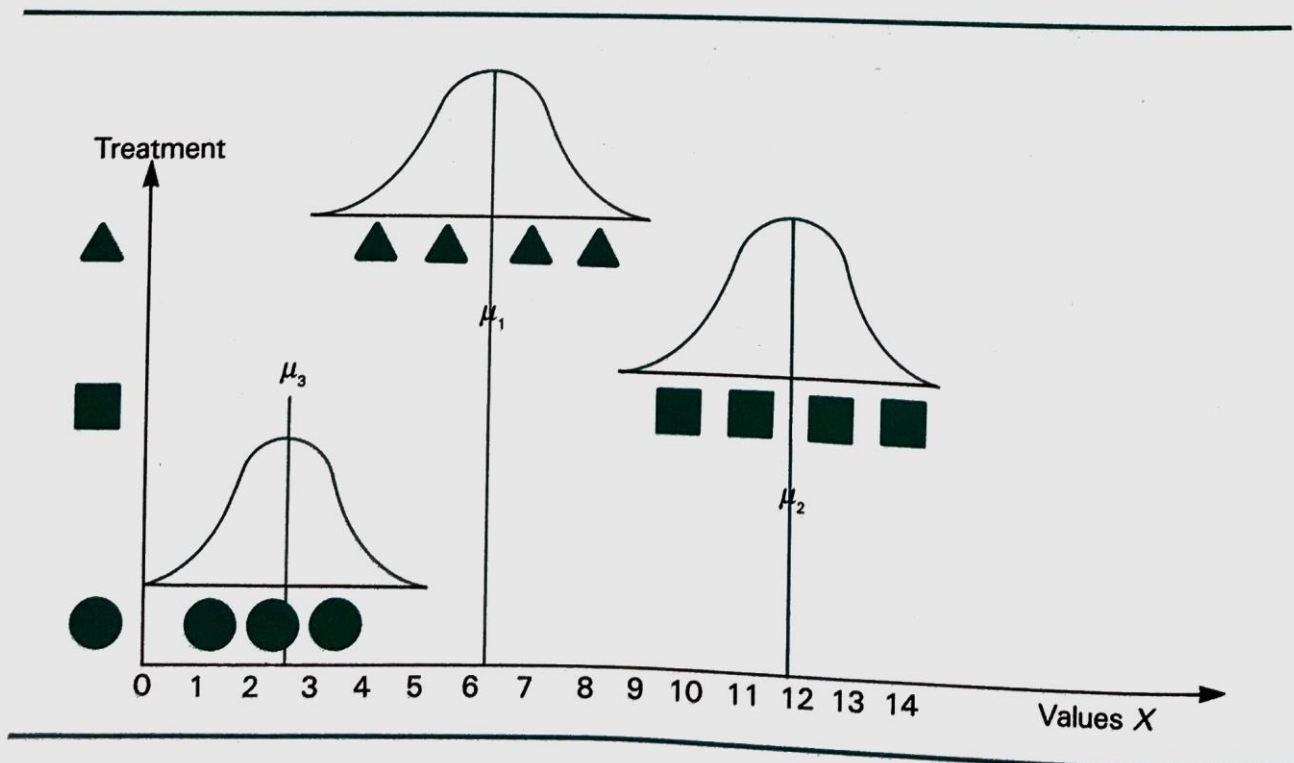
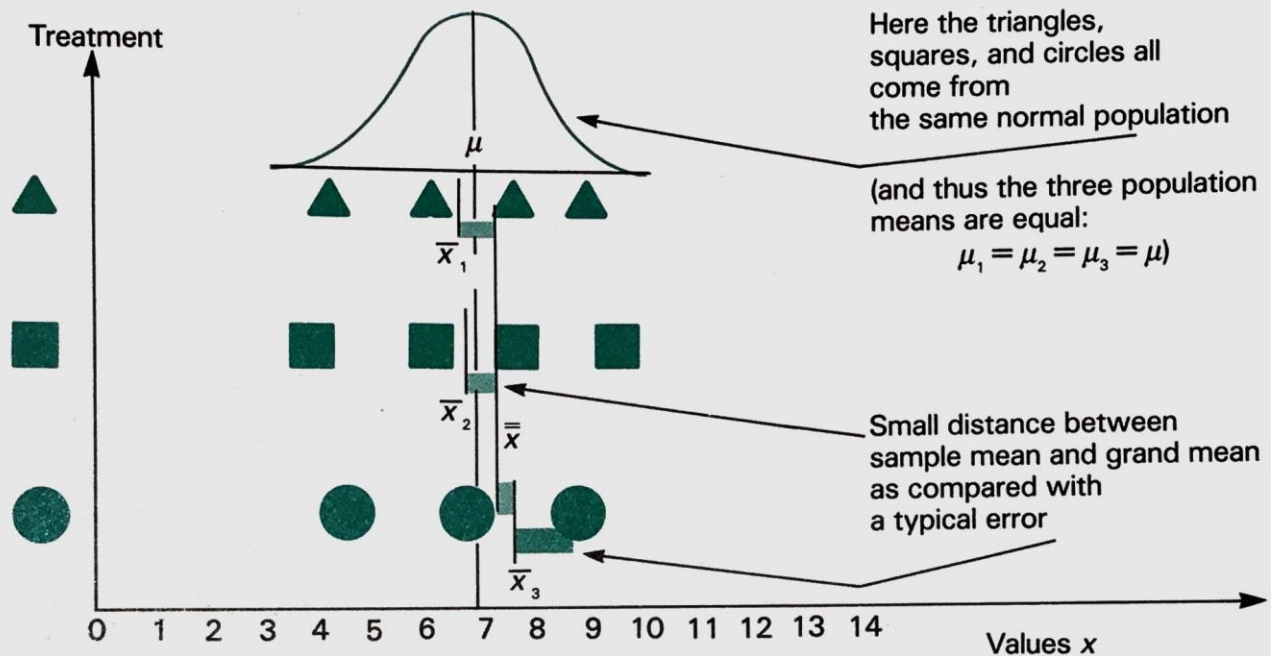


FIGURE 8-7 Samples of Triangles, Squares, and Circles Where the Average Error Deviation is Not Smaller than the Average Treatment Deviation



error is about equal to the average treatment deviation, then the treatment deviation may itself be viewed as just another error. That is, the treatment deviation in this case is due to pure chance rather than to any real difference among the population means. In other words, when the average t is of the same magnitude as the average e , then both are estimates of the internal variation within the data and carry no information about a difference between any two groups—about a difference in population means.

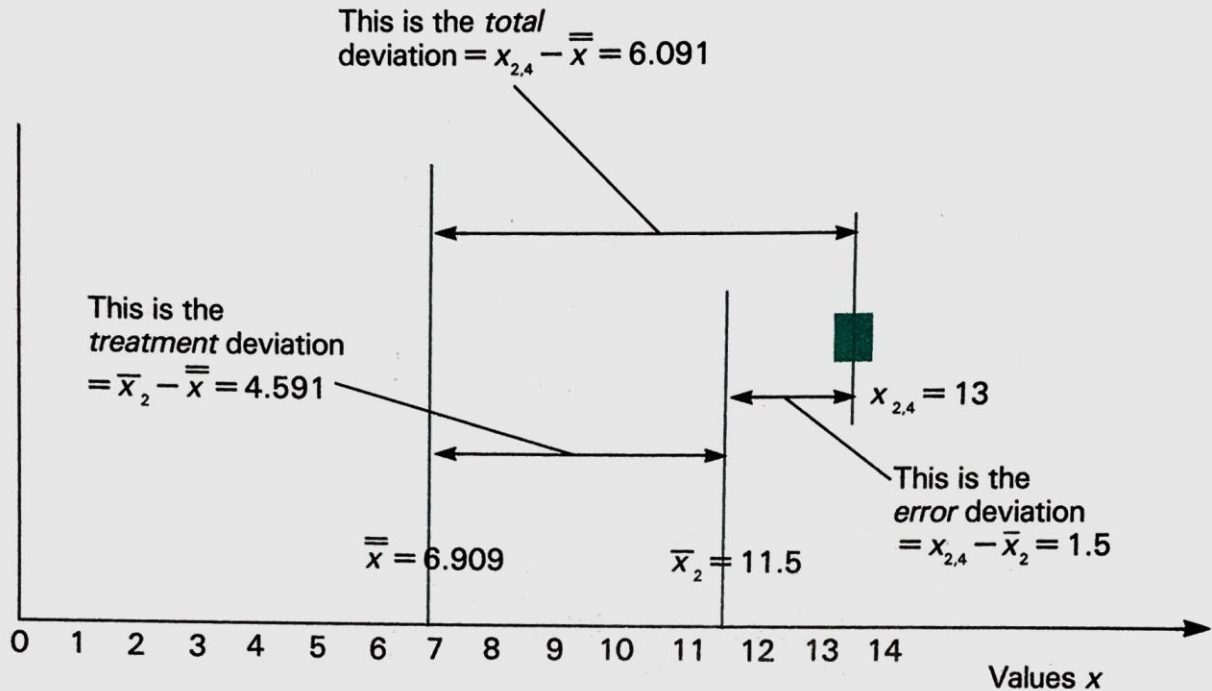
We will now make everything quantifiable, using the *sum of squares principle*. We start by returning to Figure 8-5, looking at a particular data point, and analyzing distances associated with the data point. We choose the fourth data point from the sample of squares (population 2). This data point is $x_{2,4} = 13$ (verify this from Table 8-1). We now magnify a section of Figure 8-5, the section surrounding this particular data point. This is shown in Figure 8-8.

We define the **total deviation** of a data point $x_{i,j}$ (denoted by $\text{Tot}_{i,j}$) as the deviation of the data point from the grand mean:

$$\text{Tot}_{i,j} = x_{i,j} - \bar{\bar{x}} \quad (8-7)$$

Figure 8-8 shows that the total deviation is equal to the treatment deviation plus the error deviation. This is true for *any* point in our data set (even when some of the numbers are negative).

FIGURE 8-8 The Total Deviation as the Sum of the Treatment Deviation and the Error Deviation for a Particular Data Point



For any data point, $x_{i,j}$,

$$\text{Tot} = t + e$$

(8-8)

In words:

Total deviation = Treatment deviation + Error deviation

In the case of our chosen data point, $x_{2,4}$, we have:

$$t_2 + e_{2,4} = 4.591 + 1.5 = 6.091 = \text{Tot}_{2,4}$$

Equation 8-8 works for every data point in our data set. Here is how it is algebraically derived:

$$t_i + e_{i,j} = (\bar{x}_i - \bar{x}) + (x_{i,j} - \bar{x}_i) = x_{i,j} - \bar{x} = \text{Tot}_{i,j} \quad (8-9)$$

As seen in Equation 8-9, the term $\bar{x}_i$ cancels out when the two terms in parentheses are added. This shows that for every data point, the total deviation is equal to the treatment part of the deviation plus the error part. This is also seen in Figure 8-8. The *total* deviation of a data point from the grand mean is thus partitioned into a deviation due to *treatment* and a deviation due to *error*. The deviation due to treatment differences is the *between-treatments* deviation, while the deviation due to error is the *within-treatment* deviation.

We have only considered one point, $x_{2,4}$. In order to determine whether or not the error deviations are small when compared with the treatment deviations, we need to

aggregate the partition over all data points. This is done, as we noted earlier, by averaging the deviations. We take the partition of the deviations in Equation 8–9 and we square each of the three terms (otherwise our averaging process would lead to zero).⁴ The squaring of the terms in Equation 8–9 gives, on one side:

$$t_i^2 + e_{i,j}^2 = (\bar{x}_i - \bar{\bar{x}})^2 + (x_{i,j} - \bar{x}_i)^2 \quad (8-10)$$

And on the other side:

$$\text{Tot}_{i,j}^2 = (x_{i,j} - \bar{\bar{x}})^2 \quad (8-11)$$

Note an interesting thing: The two sides of Equation 8–9 are equal, but when all three terms are squared, the two sides (now Equations 8–10 and 8–11) are *not* equal. Try this with any of the data points. The surprising thing happens next.

We take the squared deviations of Equations 8–10 and 8–11, and we *sum them over all our data points*. Interestingly, the sum of the squared error deviations and the sum of the squared treatment deviations do add up to the sum of the squared total deviations. Mathematically, cross-terms in the equation drop out, allowing this to happen. The result is the sum of squares principle.

We have the following:

$$\sum_{i=1}^r \sum_{j=1}^{n_i} \text{Tot}_{i,j}^2 = \sum_{i=1}^r n_i t_i^2 + \sum_{i=1}^r \sum_{j=1}^{n_i} e_{i,j}^2$$

This can be written in longer form as:

$$\sum_{i=1}^r \sum_{j=1}^{n_i} (x_{i,j} - \bar{\bar{x}})^2 = \sum_{i=1}^r n_i (\bar{x}_i - \bar{\bar{x}})^2 + \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2$$

The Sum of Squares Principle

The total sum of squares (SSTO) is the sum of the two terms: the sum of squares for treatment (SSTR) and the sum of squares for error (SSE).

$$\text{SSTO} = \text{SSTR} + \text{SSE} \quad (8-12)$$

The sum of squares principle partitions the total sum of squares within the data, SSTO, into a part due to treatment effect, SSTR, and a part due to errors, SSE. The squared deviations of the treatment means from the grand mean are *counted for every data point*—hence the term n_i in the first summation on the right side (SSTR) of Equation 8–12. The second term on the right-hand side is the sum of the squared

⁴ This can be seen from the data in Table 8–1. Note that the sum of the deviations of the triangles from their mean of 6 is $(4 - 6) + (5 - 6) + (7 - 6) + (8 - 6) = 0$; hence, an average of these deviations, or those of the squares or circles, leads to zero.

errors, i.e., the sum of the squared deviations of the data points from their respective sample means.

See Figure 8–8 for the different deviations associated with a single point. Imagine a similar relation among the three kinds of deviations for every one of the data points, as shown in Figure 8–5. Then imagine all these deviations squared and added together—errors to errors, treatments to treatments, and totals to totals. The result is Equation 8–12, the sum of squares principle.

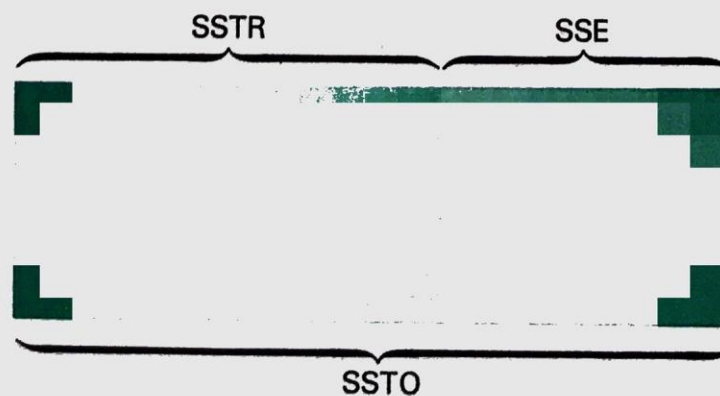
Sums of squares measure variation within the data. SSTO is the total amount of variation within the data set. SSTR is that part of the variation within the data that is due to differences among the groups, and SSE is that part of the variation within the data that is due to error—the part that cannot be explained by differences among the groups. Therefore, SSTR is sometimes called the sum of squares *between* (variation among the groups), and SSE is called the sum of squares *within* (within-group variation). SSTR is also called the *explained variation* (because it is the part of the total variation that can be explained by the fact that the data points belong to several different groups). SSE is, then, called the *unexplained variation*. The partition of the sum of squares in analysis of variance is shown in Figure 8–9.

Breaking down the sum of squares is not enough, however. If we want to determine whether or not the errors are small as compared with the treatment part, we need to find the *average* (squared) error and the *average* (squared) treatment deviation. Averaging, in the context of variances, is achieved by dividing by the appropriate number of degrees of freedom associated with each sum of squares.

Degrees of Freedom

Recall our definition of degrees of freedom in Chapter 4. The degrees of freedom are the number of data points that are “free to move,” that is, the number of elements in the data set, minus the number of restrictions. A restriction on a data set is a quantity already computed from the entire data set under consideration; thus, knowledge of this quantity makes one data point fixed and reduces by one the effective number of data points that are free to move. This is why, as was shown in Chapter 4, knowledge of the sample mean reduces the degrees of freedom of the

FIGURE 8–9 Partition of the Total Sum of Squares into Treatment and Error Parts



sample variance to $n - 1$. What are the degrees of freedom in the context of analysis of variance?

Consider the total sum of squares, SSTO. In computing this sum of squares, we use the entire data set and information about *one* quantity computed from the data: the grand mean (because, by definition, SSTO is the sum of the squared deviations of all data points from the grand mean). Since we have a total of n data points and one restriction,

The number of degrees of freedom associated with SSTO is $n - 1$.

The sum of squares for treatment, SSTR, is computed from the deviations of r sample means from the grand mean. The r sample means are considered r independent data points, and the grand mean (which can be considered as having been computed from the r sample means) thus reduces the degrees of freedom by one.

The number of degrees of freedom associated with SSTR is $r - 1$.

The sum of squares for error, SSE, is computed from the deviations of a total of n data points ($n = n_1 + n_2 + \cdots + n_r$) from r different sample means. Since each of the sample means acts as a restriction on the data set, the degrees of freedom for error are $n - r$. This can be seen another way: There are r groups with n_i data points in group i . Thus, each group, with its own sample mean acting as a restriction, has degrees of freedom equal to $n_i - 1$. The total number of degrees of freedom for error is the sum of the degrees of freedom in the r groups: $df = (n_1 - 1) + (n_2 - 1) + \cdots + (n_r - 1) = n - r$.

The number of degrees of freedom associated with SSE is $n - r$.

An important principle in analysis of variance is that the degrees of freedom of the three components are *additive* in the same way that the sums of squares are additive.

$$df(\text{total}) = df(\text{treatment}) + df(\text{error}) \quad (8-13)$$

This can easily be verified by noting the following: $n - 1 = (r - 1) + (n - r)$ —the r drops out. We are now ready to compute the average squared deviation due to treatment and the average squared deviation due to error.

The Mean Squares

In finding the average squared deviations due to treatment and to error, we divide each sum of squares by its degrees of freedom. We call the two resulting averages **mean square treatment (MSTR)** and **mean square error (MSE)**, respectively.

$$\text{MSTR} = \frac{\text{SSTR}}{r - 1} \quad (8-14)$$

$$\text{MSE} = \frac{\text{SSE}}{n - r} \quad (8-15)$$

The Expected Values of the Statistics MSTR and MSE under the Null Hypothesis

When the null hypothesis of ANOVA is true, all r population means are equal, and in this case there are *no treatment effects*. In such a case, the average squared deviation due to “treatment” is just another realization of an average squared error. In terms of the expected values of the two mean squares, we have:

$$E(\text{MSE}) = \sigma^2 \quad (8-16)$$

and

$$E(\text{MSTR}) = \sigma^2 + \frac{\sum n_i(\mu_i - \mu)^2}{r - 1} \quad (8-17)$$

where μ_i is the mean of population i and μ is the combined mean of all r populations.

Equation 8-16 says that *MSE is an unbiased estimator of σ^2 , the assumed common variance of the r populations*. The mean square error in ANOVA is therefore just like the sample variance in the one-population case of earlier chapters.

The mean square treatment, however, comprises two components, as seen from Equation 8-17. The first component is σ^2 , as in the case of MSE. The second component is a measure of the differences among the r population means, μ_i . If the null hypothesis is true, all r population means are equal—they are all equal to μ . In such a case, the second term in Equation 8-17 is equal to *zero*. When this happens, the expected value of MSTR and the expected value of MSE are each equal to σ^2 .

When the null hypothesis of ANOVA is true and all r population means are equal, MSTR and MSE are two independent, unbiased estimators of the common population variance, σ^2 .

507

If, on the other hand, the null hypothesis is not true and differences do exist among the r population means, then *MSTR will tend to be larger than MSE*. This happens because when not all population means are equal, the second term in Equation 8–17 is a positive number.

The F Statistic

The preceding discussion suggests that the ratio of MSTR to MSE is a good indicator of whether or not the r population means are equal. If the r population means are equal, then MSTR/MSE would tend to be close to 1.00. Remember that both MSTR and MSE are sample statistics derived from our data. As such, MSTR and MSE will have some randomness associated with them, and they are not likely to exactly equal their expected values. Thus, when the null hypothesis is true, MSTR/MSE will vary around the value 1.00. When the r population means are not all equal, the ratio MSTR/MSE will tend to be greater than 1.00 because the expected value of MSTR, from Equation 8–17, will be larger than the expected value of MSE. How large is “large enough” for us to reject the null hypothesis?

This is where statistical inference comes in. We want to determine whether the difference between our observed value of MSTR/MSE and the number 1.00 is due just to chance variation, or whether MSTR/MSE is *significantly* greater than 1.00—implying that the population means are not all equal. We will make the determination with the aid of the F distribution.

Under the assumptions of ANOVA, the ratio MSTR/MSE possesses an F distribution with $r - 1$ degrees of freedom for the numerator and $n - r$ degrees of freedom for the denominator when the null hypothesis is true.

In Chapter 7, we saw how the F distribution is used in determining differences between two population variances—noting that if the two variances are equal, then the ratio of the two independent, unbiased estimators of the assumed common variance follows an F distribution. There, too, the appropriate degrees of freedom for the numerator and the denominator of F came from the degrees of freedom of the sample variance in the numerator and the sample variance in the denominator of the ratio. In ANOVA, the numerator is MSTR and has $r - 1$ degrees of freedom; the denominator is MSE and has $n - r$ degrees of freedom. We thus have the following.

The test statistic in analysis of variance:

$$F_{(r-1, n-r)} = \frac{\text{MSTR}}{\text{MSE}} \quad (8-18)$$

In this section, we have seen the theoretical rationale for the F statistic we used in

Section 8-2. We also saw the computations required for arriving at the value of the test statistic. In the next section, we will encounter a convenient tool for keeping track of computations and reporting our results: the ANOVA table.

PROBLEMS

- 8-6. Define *treatment* and *error*.
- 8-7. Explain why trying to compute a simple average of all error deviations and of all treatment deviations will not lead to any results.
- 8-8. Explain how the total deviation is partitioned into the treatment deviation and the error deviation.
- 8-9. Explain the sum of squares principle.
- 8-10. Where do errors come from, and what do you think are their sources?
- 8-11. If, in an analysis of variance, you find that $MSTR$ is greater than MSE , why can you not immediately reject the null hypothesis without determining the F ratio and its distribution? Explain.
- 8-12. What is the main principle behind analysis of variance?
- 8-13. Explain how information about the variance components in a data set can lead to conclusions about population means.
- 8-14. Explain the meaning of the terms *within*, *between*, *unexplained*, *explained*, and the context in which these terms arise.
- 8-15. By the sum of squares principle, SSE and $SSTR$ are additive, and their sum is $SSTO$. Does such a relation exist between MSE and $MSTR$? Explain.
- 8-16. Does the quantity $MSTR/MSE$ follow an F distribution when the null hypothesis of ANOVA is false? Explain.
- 8-17. (A mathematically demanding problem) Prove the sum of squares principle, Equation 8-12.

8-4 THE ANOVA TABLE AND EXAMPLES:

Table 8-2 shows the data for our triangles, squares, and circles. In addition, the table shows the deviations from the group means, and their squares. From these quantities, we find the sum of squares and mean squares.

As we see in the last row of the table, the sum of all the deviations of the data points from their group means is zero, as expected. The sum of the *squared* deviations from the sample means (which, from Equation 8-12, is SSE) is equal to 17.00.

$$SSE = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2 = 17.00$$

Now we want to compute the sum of squares for treatment. Recall from Table 8-1 that $\bar{\bar{x}} = 6.909$. Again using the definitions in Equation 8-12, we have:

$$\begin{aligned} SSTR &= \sum_{i=1}^r n_i (\bar{x}_i - \bar{\bar{x}})^2 = 4(6 - 6.909)^2 + 4(11.5 - 6.909)^2 + 3(2 - 6.909)^2 \\ &= 159.9 \end{aligned}$$

TABLE 8-2 Computations for Triangles, Squares, and Circles

Treatment (<i>i</i>)	<i>j</i>	Value ($x_{i,j}$)	$(x_{i,j} - \bar{x}_j)$	$(x_{i,j} - \bar{x}_j)^2$
Triangle	1	4	$4 - 6 = -2$	$(-2)^2 = 4$
Triangle	2	5	$5 - 6 = -1$	$(-1)^2 = 1$
Triangle	3	7	$7 - 6 = 1$	$(1)^2 = 1$
Triangle	4	8	$8 - 6 = 2$	$(2)^2 = 4$
Square	1	10	$10 - 11.5 = -1.5$	$(-1.5)^2 = 2.25$
Square	2	11	$11 - 11.5 = -0.5$	$(-0.5)^2 = 0.25$
Square	3	12	$12 - 11.5 = 0.5$	$(0.5)^2 = 0.25$
Square	4	13	$13 - 11.5 = 1.5$	$(1.5)^2 = 2.25$
Circle	1	1	$1 - 2 = -1$	$(-1)^2 = 1$
Circle	2	2	$2 - 2 = 0$	$(0)^2 = 0$
Circle	3	3	$3 - 2 = 1$	$(1)^2 = 1$
			Sum = 0	Sum = 17

We now compute the mean squares. From Equations 8-14 and 8-15, respectively, we get:

$$\text{MSTR} = \frac{\text{SSTR}}{r - 1} = \frac{159.9}{2} = 79.95$$

$$\text{MSE} = \frac{\text{SSE}}{n - r} = \frac{17}{8} = 2.125$$

Using Equation 8-18, we get the computed value of the F statistic:

$$F_{(2,8)} = \frac{\text{MSTR}}{\text{MSE}} = \frac{79.95}{2.125} = 37.62$$

We are finally in a position to conduct the ANOVA hypothesis test to determine whether or not the means of the three populations are equal. From Appendix C, Table 5, we find that the critical point at $\alpha = 0.01$ (for a right-hand-tailed test) for the F distribution with 2 degrees of freedom for the numerator and 8 degrees of freedom for the denominator is 8.65. We can therefore reject the null hypothesis. Since 37.62 is much greater than 8.65, the p -value is much smaller than 0.01. This is shown in Figure 8-10.

Note that, as usual, we must exercise caution in the interpretation of results based on such small samples. As we noted earlier, in real situations we use large data sets, and the computations are usually done by computer. In the rest of our examples, we will assume that sums of squares and other quantities are produced by a computer.⁵

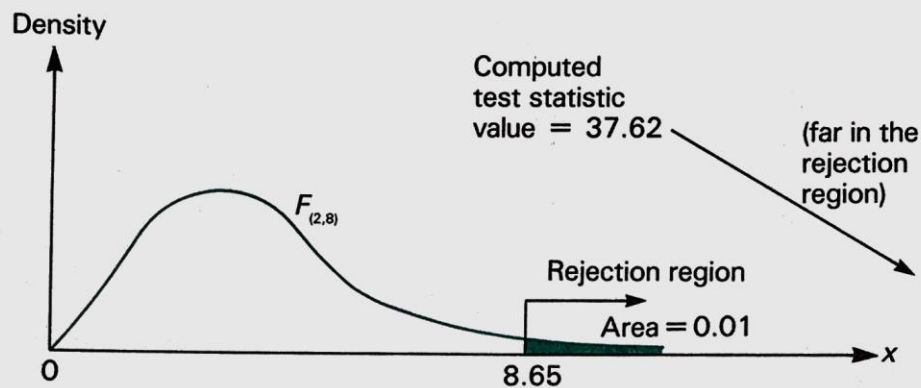
⁵ If you must carry out ANOVA computations by hand, there are equivalent computational formulas for the sums of squares that may be easier to apply than Equation 8-12. These are:

$$\text{SSTO} = \sum_i \sum_j (x_{i,j})^2 - (\sum_i \sum_j x_{i,j})^2 / n$$

$$\text{SSTR} = \sum_i ((\sum_j x_{i,j})^2 / n_i) - (\sum_i \sum_j x_{i,j})^2 / n$$

and we obtain SSE by subtraction: $\text{SSE} = \text{SSTO} - \text{SSTR}$.

FIGURE 8-10 Rejecting the Null Hypothesis in the Triangles, Squares, and Circles Example.



All statistical computer packages contain ANOVA routines, and in Section 8-9, we will present examples of input and output.

An essential tool for reporting the results of an analysis of variance is the ANOVA table. An ANOVA table lists the sources of variation: treatment, error, and total. (In the two-factor ANOVA, which we will see in later sections, there will be more sources of variation.) The ANOVA table lists the sums of squares, the degrees of freedom, the mean squares, and the F ratio. The table format simplifies the analysis and the interpretation of the results. The structure of the ANOVA table is based on the fact that both the sums of squares and the degrees of freedom are additive. We will now present an ANOVA table for the triangles, squares, and circles example. Table 8-3 shows the results computed above.

Note that the entries in the second and third columns, sum of squares and degrees of freedom, are both additive. The entries in the fourth column, mean square, are obtained by dividing the appropriate sums of squares by their degrees of freedom. We do not define a mean square total, which is why there is no entry in that particular position in the table. The last entry in the table is the main objective of our analysis: the F ratio, which is computed as the ratio of the two entries in the previous column. There are no other entries in the last column. Example (b), which follows, demonstrates the use of the ANOVA table.

TABLE 8-3 ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	$F = \text{Ratio}$
Treatment	$SSTR = 159.9$	$r - 1 = 2$	$MSTR = \frac{SSTR}{r - 1}$ $= 79.95$	$F = \frac{MSTR}{MSE}$ $= 37.62$
Error	$SSE = 17.0$	$n - r = 8$	$MSE = \frac{SSE}{n - r}$	
Total	$SSTO = 176.9$	$n - 1 = 10$	$= 2.125$	

EXAMPLE (b):

Club Med has over 30 major resorts worldwide, from Tahiti to Switzerland. Many of the beach resorts are in the Caribbean, and at one point the club wanted to test whether or not the resorts on Guadeloupe, Martinique, Haiti, Paradise Island, and St. Lucia were all equally well liked by vacationing club members. The analysis was to be based on a survey questionnaire filled out by a random sample of 40 respondents in each of the resorts. From every returned questionnaire, a general satisfaction score, on a scale of 0 to 100, was computed. Analysis of the survey results yielded the statistics given in Table 8–4.

The results were computed from the responses using a computer program that calculated the sums of squared deviations from the sample means and from the grand mean. Given the values of SSTO and SSE, construct an ANOVA table and conduct the hypothesis test. (Note: the reported sample means in Table 8–4 will be used in the next section.)

SOLUTION:

Let us first construct an ANOVA table and fill in the information we have: SSTO = 112,564; SSE = 98,356; $n = 200$; and $r = 5$. This has been done in Table 8–5. We now compute SSTR as the difference between SSTO and SSE and enter it in the appropriate place in the table. We then divide SSTR and SSE by their respective degrees of freedom to give us MSTR and MSE. Finally, we divide MSTR by MSE to give us the F ratio. All these quantities are entered in the ANOVA table. The result is the complete ANOVA table for the study, Table 8–6.

Table 8–6 contains all the pertinent information for this study. We are now ready to conduct the hypothesis test.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (average vacationer satisfaction for each of the five resorts is equal)

TABLE 8–4 Club Med Survey Results

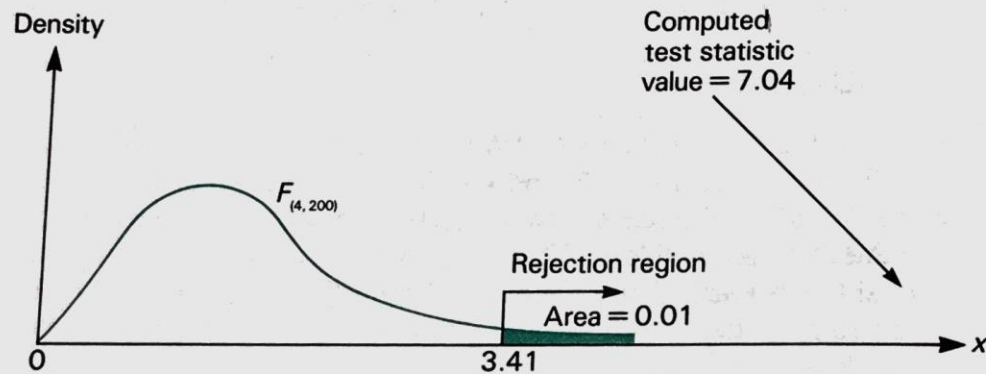
Resort (i)	Mean Response ($\bar{x}_i$)
1. Guadeloupe	89
2. Martinique	75
3. Haiti	73
4. Paradise Island	91
5. St. Lucia	85
SSTO = 112,564 SSE = 98,356	

TABLE 8–5 Preliminary ANOVA Table for Club Med Example

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Treatment	SSTR =	$r - 1 = 4$	MSTR =	$F =$
Error	SSE = 98,356	$n - r = 195$	MSE =	
Total	SSTO = 112,564	$n - 1 = 199$		

TABLE 8-6 ANOVA Table for Club Med Example

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Treatment	SSTR = 14,208	$r - 1 = 4$	MSTR = 3,552	$F = 7.04$
Error	SSE = 98,356	$n - r = 195$	MSE = 504.4	
Total	SSTO = 112,564	$n - 1 = 199$		

FIGURE 8-11 Club Med Test

H_1 : Not all μ_i ($i = 1, \dots, 5$) are equal (on the average, vacationer satisfaction is not equal among the five resorts)

As shown in Table 8-6, the test statistic value is $F_{(4, 195)} = 7.04$. As often happens, the exact number of degrees of freedom we need does not appear in Appendix C, Table 5. We use the nearest entry, which is the critical point for F with 4 degrees of freedom for the numerator and 200 degrees of freedom for the denominator. The critical point for $\alpha = 0.01$ is $C = 3.41$. The test is illustrated in Figure 8-11.

Since the computed test statistic value falls in the rejection region for $\alpha = 0.01$, we reject the null hypothesis and note that the p -value is smaller than 0.01. We may conclude that, based on the survey results and our assumptions, it is likely that the five resorts studied are not equal in terms of average vacationer satisfaction. Which resorts are more satisfying than others? This question will be answered when we return to this example in the next section.

EXAMPLE (c):

An article in the *Journal of Marketing Research* reports the results of a study of job involvement of salespeople in the four major career stages: exploration, establishment, maintenance, and disengagement.⁶ As part of the study, the authors report the results of an analysis of vari-

⁶ W. L. Cron and J. W. Slocum, "The Influence of Career Stages on Salespeople's Performance," *Journal of Marketing Research* (May 1986).

TABLE 8-7 ANOVA Table for Job Involvement

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Treatment	SSTR = 879.3	$r - 1 = 3$	MSTR = 293.1	$F = 8.52$
Error	SSE = 18,541.6	$n - r = 539$	MSE = 34.4	
Total	SSTO = 19,420.9	$n - 1 = 542$		

ance aimed at determining whether or not salespeople in each of the four career stages are, on the average, equally involved with their jobs. Involvement is measured on a special scale developed by psychologists. The analysis is based on questionnaires returned by a total of 543 respondents, and the reported F value is 8.52. The authors note the result is “significant at $p < .01$.” Assuming that $MSE = 34.5$, construct an ANOVA table for this example. Also verify the authors’ claim about the significance of their results.

SOLUTION:

In this problem, another exercise in the construction of ANOVA tables, we are doing the opposite of what is usually done: we are going from the final result of an F ratio to the earlier stages of an analysis of variance. First, multiplying the F ratio by MSE gives us MSTR. Then, from the sample size $n = 543$ and from $r = 4$, we get the number of degrees of freedom for treatment, error, and total. Using our information, we construct the ANOVA table (Table 8-7).

From Appendix C, Table 5, we find that the critical point for a right-hand-tailed test at $\alpha = 0.01$ for an F distribution with 3 and 400 degrees of freedom (the entry for degrees of freedom closest to the needed 3 and 539) is 3.83. Thus, we may conclude that differences do exist among the four career stages with respect to average job involvement. The authors’ statement about the p -value is also true: the p -value is much smaller than 0.01.