12

Analysis of Variance

GOALS

When you have completed this chapter, you will be able to:

1 **List the characteristics of the *F* distribution.**

2 **Conduct a test of hypothesis to determine whether the variances of two populations are equal.**

3 **Discuss the general idea of analysis of variance.**

4 **Organize data into a one-way ANOVA table.**

5 **Conduct a test of hypothesis among three or more treatment means.**

6 **Develop confidence intervals for the difference in treatment means.**

A grocery store wants to monitor the amount of withdrawals that its customers make from automatic teller machines (ATMs) located within their stores. They sample 10 withdrawals from each location: Use a .01 level of significance to test if there is a difference in the mean amount of money withdrawn, based on the output on page 369. (See Goal 5 and Exercise 29.)

Introduction

In this chapter we continue our discussion of hypothesis testing. Recall that in Chapters 10 and 11 we examined the general theory of hypothesis testing. We described the case where a large sample was selected from the population. We used the *z* distribution (the standard normal distribution) to determine whether it was reasonable to conclude that the population mean was equal to a specified value. We tested whether two population means are the same. We also conducted both one- and two-sample tests for population proportions, again using the standard normal distribution as the distribution of the test statistic. We described methods for conducting tests of means where the populations were assumed normal but the samples were small (contained fewer than 30 observations). In that case the *t* distribution was used as the distribution of the test statistic. In this chapter we expand further our idea of hypothesis tests. We describe a test for variances and then a test that simultaneously compares several means to determine if they came from equal populations.

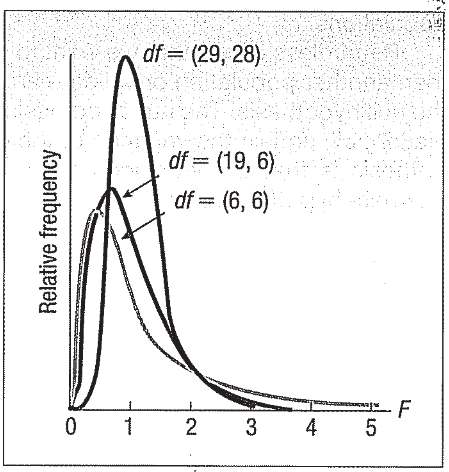
The *F* Distribution

The probability distribution used in this chapter is the *F* distribution. It was named to honor Sir Ronald Fisher, one of the founders of modern-day statistics. This probability distribution is used as the distribution of the test statistic for several situations. It is used to test whether two samples are from populations having equal variances, and it is also applied when we want to compare several population means simultaneously. The simultaneous comparison of several population means is called **analysis of variance (ANOVA).** In both of these situations, the populations must follow a normal distribution, and the data must be at least interval-scale.

What are the characteristics of the *F* distribution?

Characteristics of the *F* distribution

1. **There is a "family" of *F* distributions.** A particular member of the family is determined by two parameters: the degrees of freedom in the numerator and the degrees of freedom in the denominator. The shape of the distribution is illustrated by the following graph. There is one *F* distribution for the combination of 29 degrees of freedom in the numerator and 28 degrees of freedom in the denominator. There is another *F* distribution for 19 degrees in the numerator and 6 degrees of freedom in the denominator. Note that the shape of the curves changes as the degrees of freedom changes.



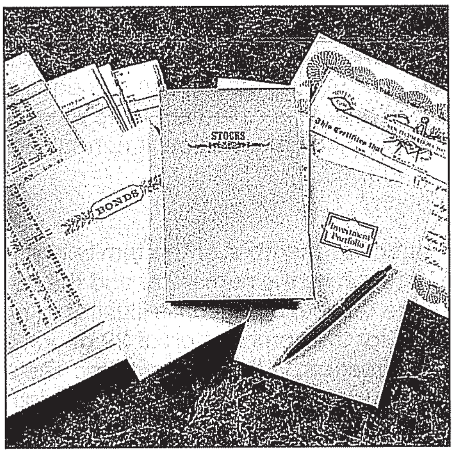
2. **The *F* distribution is continuous.** This means that it can assume an infinite number of values between zero and positive infinity.

3. **The *F* distribution cannot be negative.** The smallest value *F* can assume is 0.
4. **It is positively skewed.** The long tail of the distribution is to the right-hand side. As the number of degrees of freedom increases in both the numerator and denominator the distribution approaches a normal distribution.
5. **It is asymptotic.** As the values of *X* increase, the *F* curve approaches the *X*-axis but never touches it. This is similar to the behavior of the normal distribution, described in Chapter 7.

Comparing Two Population Variances

The *F* distribution is used to test the hypothesis that the variance of one normal population equals the variance of another normal population. The following examples will show the use of the test:

- Two Barth shearing machines are set to produce steel bars of the same length. The bars, therefore, should have the same mean length. We want to ensure that in addition to having the same mean length they also have similar variation.



- The mean rate of return on two types of common stock may be the same, but there may be more variation in the rate of return in one than the other. A sample of 10 Internet stocks and 10 utility stocks shows the same mean rate of return, but there is likely more variation in the Internet stocks.
- A study by the marketing department for a large newspaper found that men and women spent about the same amount of time per day reading the paper. However, the same report indicated there was nearly twice as much variation in time spent per day among the men than the women.

The *F* distribution is also used to test assumptions for some statistical tests. Recall that in the previous chapter when small samples were assumed, we used the *t* test to investigate whether the means of two independent populations differed. To employ that test, we assume that the variances of two normal populations are the same. See this list of assumptions on page 323. The *F* distribution provides a means for conducting a test regarding the variances of two normal populations.

Regardless of whether we want to determine if one population has more variation than another population or validate an assumption for a statistical test, we first state the null hypothesis. The null hypothesis could be that the variance of one normal population, σ_1^2, equals the variance of the other normal population, σ_2^2. The alternate hypothesis is that the variances differ. In this instance the null hypothesis and the alternate hypothesis are:

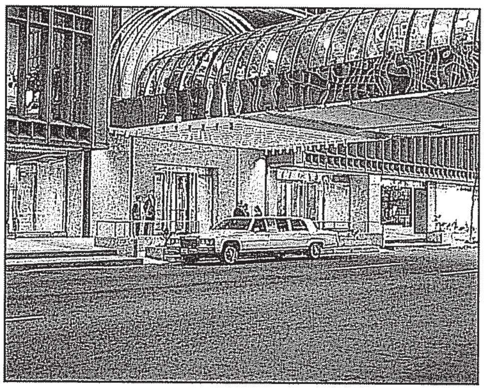
$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

To conduct the test, we select a random sample of n_1 observations from one population, and a sample of n_2 observations from the second population. The test statistic is defined as follows.

TEST STATISTIC FOR COMPARING TWO VARIANCES $$F = \frac{s_1^2}{s_2^2} \qquad [12\text{–}1]$$

The terms s_1^2 and s_2^2 are the respective sample variances. The test statistic follows the F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. In order to reduce the size of the table of critical values, the *larger* sample variance is placed in the numerator; hence, the tabled F ratio is always larger than 1.00. Thus, the right-tail critical value is the only one required. The critical value of F for a two-tailed test is found by dividing the significance level in half ($\alpha/2$) and then referring to the appropriate degrees of freedom in Appendix G. An example will illustrate.

EXAMPLE



Lammers Limos offers limousine service from the city hall in Toledo, Ohio, to Metro Airport in Detroit. Sean Lammers, president of the company, is considering two routes. One is via U.S. 25 and the other via I-75. He wants to study the time it takes to drive to the airport using each route and then compare the results. He collected the following sample data, which is reported in minutes. Using the .10 significance level, is there a difference in the variation in the driving times for the two routes?

U.S. Route 25	Interstate 75
52	59
67	60
56	61
45	51
70	56
54	63
64	57
	65

SOLUTION

The mean driving times along the two routes are nearly the same. The mean time is 58.29 minutes for the U.S. 25 route and 59.0 minutes along the I-75 route. However, in evaluating travel times, Mr. Lammers is also concerned about the variation in the travel times. The first step is to compute the two sample variances. We'll use formula (3–11) to compute the sample standard deviations. To obtain the sample variances, we square the standard deviations.

U.S. Route 25

$$\bar{X} = \frac{\Sigma X}{n} = \frac{408}{7} = 58.29 \qquad s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{485.43}{7-1}} = 8.9947$$

Interstate 75

$$\bar{X} = \frac{\Sigma X}{n} = \frac{472}{8} = 59.00 \qquad s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{134}{8-1}} = 4.3753$$

There is more variation, as measured by the standard deviation, in the U.S. 25 route than in the I-75 route. This is somewhat consistent with his knowledge of the two routes; the U.S. 25 route contains more stoplights, whereas I-75 is a limited-access interstate highway. However, the I-75 route is several miles longer. It is important that the service offered be both timely and consistent, so he decides to conduct a statistical test to determine whether there really is a difference in the variation of the two routes.

The usual five-step hypothesis-testing procedure will be employed.

Step 1: We begin by stating the null hypothesis and the alternate hypothesis. The test is two-tailed because we are looking for a difference in the variation of the two routes. We are *not* trying to show that one route has more variation than the other.

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Step 2: We selected the .10 significance level.

Step 3: The appropriate test statistic follows the F distribution.

Step 4: The critical value is obtained from Appendix G, a portion of which is reproduced as Table 12–1. Because we are conducting a two-tailed test, the tabled significance level is .05, found by $\alpha/2 = .10/2 = .05$. There are $n_1 - 1 = 7 - 1 = 6$ degrees of freedom in the numerator, and $n_2 - 1 = 8 - 1 = 7$ degrees of freedom in the denominator. To find the critical value, move horizontally across the top portion of the F table (Table 12–1 or Appendix G) for the .05 significance level to 6 degrees of freedom in the numerator. Then move down that column to the critical value opposite 7 degrees of freedom in the denominator. The critical value is 3.87. Thus, the decision rule is: Reject the null hypothesis if the ratio of the sample variances exceeds 3.87.

TABLE 12–1 Critical Values of the F Distribution, $\alpha = .05$

Degrees of Freedom for Denominator	Degrees of Freedom for Numerator			
	5	6	7	8
1	230	234	237	239
2	19.3	19.3	19.4	19.4
3	9.01	8.94	8.89	8.85
4	6.26	6.16	6.09	6.04
5	5.05	4.95	4.88	4.82
6	4.39	4.28	4.21	4.15
7	3.97	3.87	3.79	3.73
8	3.69	3.58	3.50	3.44
9	3.48	3.37	3.29	3.23
10	3.33	3.22	3.14	3.07

Step 5: The final step is to take the ratio of the two sample variances, determine the value of the test statistic, and make a decision regarding the null hypothesis. Note that formula (12–1) refers to the sample *variances* but we calculated the sample *standard deviations*. We need to square the standard deviations to determine the variances.

$$F = \frac{s_1^2}{s_2^2} = \frac{(8.9947)^2}{(4.3753)^2} = 4.23$$

The decision is to reject the null hypothesis, because the computed F value (4.23) is larger than the critical value (3.87). We conclude that there is a difference in the variation of the travel times along the two routes.

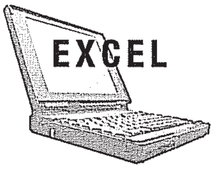
As noted, the usual practice is to determine the *F* ratio by putting the larger of the two sample variances in the numerator. This will force the *F* ratio to be at least 1.00. This allows us to always use the right tail of the *F* distribution, thus avoiding the need for more extensive *F* tables.

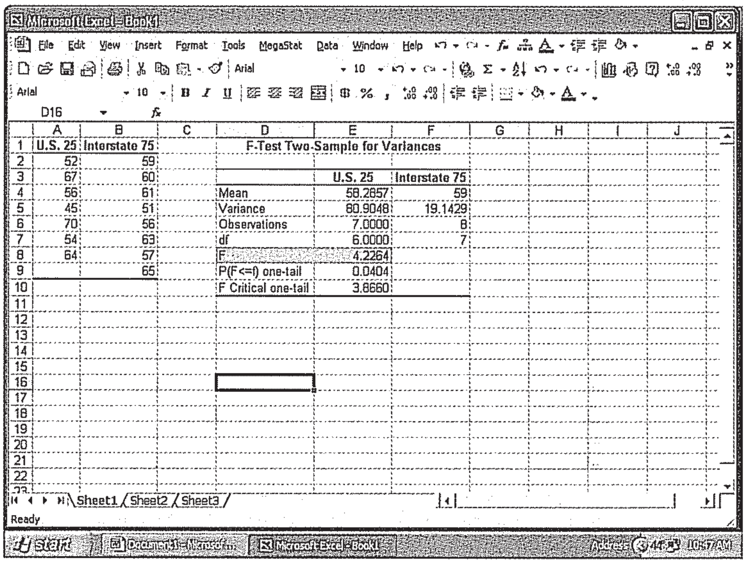
A logical question arises regarding one-tailed tests. For example, suppose in the previous example we suspected that the variance of the times using the U.S. 25 route is *larger* than the variance of the times along the I-75 route. We would state the null and the alternate hypothesis as

$$H_0\text{: } \sigma_1^2 \leq \sigma_2^2$$
$$H_1\text{: } \sigma_1^2 > \sigma_2^2$$

The test statistic is computed as s_1^2/s_2^2. Notice that we labeled the population with the suspected largest variance as population 1. So s_1^2 appears in the numerator. The *F* ratio will be larger than 1.00, so we can use the upper tail of the *F* distribution. Under these conditions, it is not necessary to divide the significance level in half. Because Appendix G gives us only the .05 and .01 significance levels, we are restricted to these levels for one-tailed tests and .10 and .02 for two-tailed tests unless we consult a more complete table or use statistical software to compute the *F* statistic.

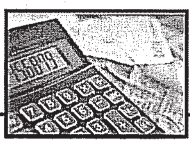
The Excel software system has a procedure to perform a test of variances. Below is the output. The computed value of *F* is the same as determined by using formula (12–1). The computed value of *F* is highlighted by shading.





	A	B	C	D	E	F
1	U.S. 25	Interstate 75		F-Test Two-Sample for Variances		
2	52	59				
3	67	60			U.S. 25	Interstate 75
4	56	61		Mean	58.2857	59
5	45	51		Variance	80.9048	19.1429
6	70	56		Observations	7.0000	8
7	54	63		df	6.0000	7
8	64	57		F	4.2264	
9		65		P(F<=f) one-tail	0.0404	
10				F Critical one-tail	3.8660	

Self-Review 12–1



Steele Electric Products, Inc. assembles electrical components for cell phones. For the last 10 days Mark Nagy has averaged 9 rejects, with a standard deviation of 2 rejects per day. Debbie Richmond averaged 8.5 rejects, with a standard deviation of 1.5 rejects, over the same period. At the .05 significance level, can we conclude that there is more variation in the number of rejects per day attributed to Mark?

Exercises

1. What is the critical *F* value for a sample of six observations in the numerator and four in the denominator? Use a two-tailed test and the .10 significance level.

2. What is the critical F value for a sample of four observations in the numerator and seven in the denominator? Use a one-tailed test and the .01 significance level.
3. The following hypotheses are given.

H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$

A random sample of eight observations from the first population resulted in a standard deviation of 10. A random sample of six observations from the second population resulted in a standard deviation of 7. At the .02 significance level, is there a difference in the variation of the two populations?
4. The following hypotheses are given.

H_0: $\sigma_1^2 \leq \sigma_2^2$
H_1: $\sigma_1^2 > \sigma_2^2$

A random sample of five observations from the first population resulted in a standard deviation of 12. A random sample of seven observations from the second population showed a standard deviation of 7. At the .01 significance level, is there more variation in the first population?
5. Arbitron Media Research, Inc. conducted a study of the radio listening habits of men and women. One facet of the study involved the mean listening time. It was discovered that the mean listening time for men was 35 minutes per day. The standard deviation of the sample of the 10 men studied was 10 minutes per day. The mean listening time for the 12 women studied was also 35 minutes, but the standard deviation of the sample was 12 minutes. At the .10 significance level, can we conclude that there is a difference in the variation in the listening times for men and women?
6. A stockbroker at Critical Securities reported that the mean rate of return on a sample of 10 oil stocks was 12.6 percent with a standard deviation of 3.9 percent. The mean rate of return on a sample of 8 utility stocks was 10.9 percent with a standard deviation of 3.5 percent. At the .05 significance level, can we conclude that there is more variation in the oil stocks?

ANOVA Assumptions

Another use of the F distribution is the analysis of variance (ANOVA) technique in which we compare three or more population means to determine whether they could be equal. To use ANOVA, we assume the following:

1. The populations follow the normal distribution.
2. The populations have equal standard deviations (σ).
3. The populations are independent.

When these conditions are met, F is used as the distribution of the test statistic.

Why do we need to study ANOVA? Why can't we just use the test of differences in population means discussed in the previous chapter? We could compare the treatment means two at a time. The major reason is the unsatisfactory buildup of Type I error. To explain further, suppose we have four different methods (A, B, C, and D) of training new recruits to be firefighters. We randomly assign each of the 40 recruits in this year's class to one of the four methods. At the end of the training program, we administer to the four groups a common test to measure understanding of firefighting techniques. The question is: Is there a difference in the mean test scores among the four groups? An answer to this question will allow us to compare the four training methods.

Using the *t* distribution leads to a buildup of Type I error.

Using the t distribution to compare the four population means, we would have to conduct six different t tests. That is, we would need to compare the mean scores for the four methods as follows: A versus B, A versus C, A versus D, B versus C, B versus D, and C versus D. If we set the significance level at .05, the probability of a correct statistical decision is .95, found by $1 - .05$. Because we conduct six separate (independent) tests the probability that we do *not* make an incorrect decision due to sampling in any of the six independent tests is:

$$P(All\ correct) = (.95)(.95)(.95)(.95)(.95)(.95) = .735$$

To find the probability of at least one error due to sampling, we subtract this result from 1. Thus, the probability of at least one incorrect decision due to sampling is 1 − .735 = .265. To summarize, if we conduct six independent tests using the *t* distribution, the likelihood of rejecting a true null hypothesis because of sampling error is increased from .05 to an unsatisfactory level of .265. It is obvious that we need a better method than conducting six *t* tests. ANOVA will allow us to compare the treatment means simultaneously and avoid the buildup of the Type I error.

ANOVA was developed for applications in agriculture, and many of the terms related to that context remain. In particular the term *treatment* is used to identify the different populations being examined. The following illustration will clarify the term *treatment* and demonstrate an application of ANOVA.

EXAMPLE

Joyce Kuhlman manages a regional financial center. She wishes to compare the productivity, as measured by the number of customers served, among three employees. Four days are randomly selected and the number of customers served by each employee is recorded. The results are:

Wolfe	White	Korosa
55	66	47
54	76	51
59	67	46
56	71	48

SOLUTION

Is there a difference in the mean number of customers served? Chart 12–1 illustrates how the populations would appear if there was a difference in the treatment means. Note that the populations follow the normal distribution and the variation in each population is the same. However, the population means are *not* the same.

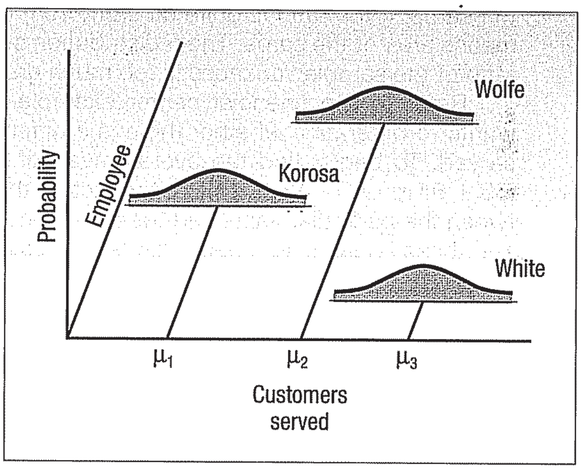


CHART 12–1 Case Where Treatment Means Are Different

Suppose the populations are the same. That is, there is no difference in the (treatment) means. This is shown in Chart 12–2. This would indicate that the population means are the same. Note again that the populations follow the normal distribution and the variation in each of the populations is the same.

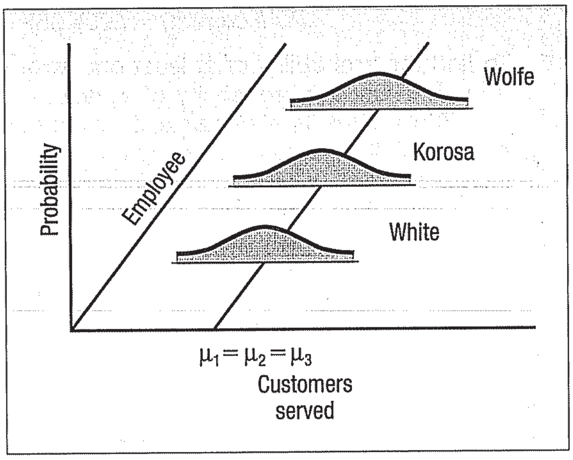


CHART 12–2 Case Where Treatment Means Are the Same

The ANOVA Test

How does the ANOVA test work? Recall that we want to determine whether the various sample means came from a single population or populations with different means. We actually compare these sample means through their variances. To explain, recall that on page 350 we listed the assumptions required for ANOVA. One of those assumptions was that the standard deviations of the various normal populations had to be the same. We take advantage of this requirement in the ANOVA test. The underlying strategy is to estimate the population variance (standard deviation squared) two ways and then find the ratio of these two estimates. If this ratio is about 1, then logically the two estimates are the same, and we conclude that the population means are the same. If the ratio is quite different from 1, then we conclude that the population means are not the same. The *F* distribution serves as a referee by indicating when the ratio of the sample variances is too much greater than 1 to have occurred by chance.

Refer to the financial center example in the previous section. The manager wants to determine whether there is a difference in the mean number of customers served. To begin, find the overall mean of the 12 observations. It is 58, found by $(55 + 54 + \cdots + 48)/12$. Next, for each of the 12 observations find the difference between the particular value and the overall mean. Each of these differences is squared and these squares summed. This term is called the **total variation.**

> **TOTAL VARIATION** The sum of the squared differences between each observation and the overall mean.

In our example the total variation is 1,082, found by $(55 - 58)^2 + (54 - 58)^2 + \cdots + (48 - 58)^2$.

Next, break this total variation into two components: that which is due to the **treatments** and that which is **random.** To find these two components, determine the mean of each of the treatments. The first source of variation is due to the treatments.

> **TREATMENT VARIATION** The sum of the squared differences between each treatment mean and the grand or overall mean.

In the example the variation due to the treatments is the sum of the squared differences between the mean of each employee and the overall mean. This term is 992. To calculate it we first find the mean of each of the three treatments. The mean for Wolfe is 56, found by (55 + 54 + 59 + 56)/4. The other means are 70 and 48, respectively. The sum of the squares due to the treatments is:

$$(56 - 58)^2 + (56 - 58)^2 + \cdots + (48 - 58)^2 = 4(56 - 58)^2 + 4(70 - 58)^2 + 4(48 - 58)^2$$
$$= 992$$

If there is considerable variation among the treatment means, it is logical that this term will be large. If the treatment means are similar, this term will be a small value. The smallest possible value would be zero. This would occur when all the treatment means are the same and equal to the overall mean.

The other source of variation is referred to as the **random** component, or the error component.

> **RANDOM VARIATION** The sum of the squared differences between each observation and its treatment mean.

In the example this term is the sum of the squared differences between each value and the mean for that particular employee. The error variation is 90.

$$(55 - 56)^2 + (54 - 56)^2 + \cdots + (48 - 48)^2 = 90$$

We determine the test statistic, which is the ratio of the two estimates of the population variance, from the following equation.

$$F = \frac{\text{Estimate of the population variance based on the differences among the sample means}}{\text{Estimate of the population variance based on the variation within the samples}}$$

Our first estimate of the population variance is based on the treatments, that is, the difference *between* the means. It is 992/2. Why did we divide by 2? Recall from Chapter 3, to find a sample variance [see formula (3–11)], we divide by the number of observations minus one. In this case there are three treatments, so we divide by 2. Our first estimate of the population variance is 992/2.

The variance estimate *within* the treatments is the random variation divided by the total number of observations less the number of treatments. That is 90/(12 − 3). Hence, our second estimate of the population variance is 90/9. This is actually a generalization of formula (11–5), where we pooled the sample variances from two populations.

The last step is to take the ratio of these two estimates.

$$F = \frac{992/2}{90/9} = 49.6$$

Because this ratio is quite different from 1, we can conclude that the population means are not the same. There is a difference in the mean number of customers served by the three employees.

Here's another example of the ANOVA technique which deals with samples of different sizes. It will provide additional insight into the technique.

EXAMPLE

Professor James Brunner had the 22 students in his 10 A.M. Introduction to Marketing rate his performance as Excellent, Good, Fair, or Poor. A graduate student collected the ratings and assured the students that Professor Brunner would not receive them until after course grades had been sent to the Registrar's office. The rating (i.e., the treatment) a student gave the professor was matched with his or her course grade, which could range from 0 to 100. The sample information is reported below. Is there a difference in the mean score of the students in each of the four rating categories? Use the .01 significance level.

Course Grades			
Excellent	**Good**	**Fair**	**Poor**
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

SOLUTION

We will follow the usual five-step hypothesis-testing procedure.

Step 1: State the null hypothesis and the alternate hypothesis. The null hypothesis is that the mean scores are the same for the four ratings.

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

The alternate hypothesis is that the mean scores are not all the same for the four ratings.

H_1: The mean scores are not all equal.

We can also think of the alternate hypothesis as "at least two mean scores are not equal."

If the null hypothesis is not rejected, we conclude that there is no difference in the mean course grades based on the instructor ratings. If H_0 is rejected, we conclude that there is a difference in at least one pair of mean ratings, but at this point we do not know which pair or how many pairs differ.

Step 2: Select the level of significance. We selected the .01 significance level.

Step 3: Determine the test statistic. The test statistic follows the *F* distribution.

Step 4: Formulate the decision rule. To determine the decision rule, we need the critical value. The critical value for the *F* statistic is found in Appendix G. The critical values for the .05 significance level are found on the first page and the .01 significance level on the second page. To use this table we need to know the degrees of freedom in the numerator and the denominator. The degrees of freedom in the numerator equals the number of treatments, designated as *k,* minus 1. The degrees of freedom in the denominator is the total number of observations, *n,* minus the number of treatments. For this problem there are four treatments and a total of 22 observations.

Degrees of freedom in the numerator $= k - 1 = 4 - 1 = 3$
Degrees of freedom in the denominator $= n - k = 22 - 4 = 18$

Refer to Appendix G and the .01 significance level. Move horizontally across the top of the page to 3 degrees of freedom in the numerator. Then move down that column to the row with 18 degrees of freedom. The value at this intersection is 5.09. So the decision rule is to reject H_0 if the computed value of *F* exceeds 5.09.

Step 5: Select the sample, perform the calculations, and make a decision. It is convenient to summarize the calculations of the *F* statistic in an

ANOVA table. The format for an ANOVA table is as follows. Statistical software packages also use this format.

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Treatments	SST	$k - 1$	$SST/(k - 1) = MST$	MST/MSE
Error	SSE	$n - k$	$SSE/(n - k) = MSE$	
Total	SS total	$n - 1$		

There are three values, or sum of squares, used to compute the test statistic F. You can determine these values by obtaining SS total and SSE, then finding SST by subtraction. The SS total term is the total variation, SST is the variation due to the treatments, and SSE is the variation within the treatments.

We usually start the process by finding SS total. This is the sum of the squared differences between each observation and the overall mean. The formula for finding SS total is:

$$\text{SS total} = \Sigma(X - \bar{X}_G)^2 \quad \textbf{[12–2]}$$

where:

X is each sample observation.
$\bar{X}_G$ is the overall or grand mean. The subscript G refers to the "grand" mean.

Next determine SSE or the sum of the squared errors. This is the sum of the squared differences between each observation and its respective treatment mean. The formula for finding SSE is:

$$\text{SSE} = \Sigma(X - \bar{X}_c)^2 \quad \textbf{[12–3]}$$

where:

$\bar{X}_c$ is the sample mean for treatment c. The subscript c refers to the particular column.

The detailed calculations of SS total and SSE for this example follow. To determine the values of SS total and SSE we start by calculating the overall or grand mean. There are 22 observations and the total is 1,664, so the grand mean is 75.64.

$$\bar{X}_G = \frac{1{,}664}{22} = 75.64$$

	Excellent	Good	Fair	Poor	Total
	94	75	70	68	
	90	68	73	70	
	85	77	76	72	
	80	83	78	65	
		88	80	74	
			68	65	
			65		
Column total	349	391	510	414	1664
n	4	5	7	6	22
Mean	87.25	78.20	72.86	69.00	75.64

Next we find the deviation of each observation from the grand mean, square those deviations, and sum this result for all 22 observations. For example, the first student who rated Dr. Brunner Excellent had a score of 94 and the overall or grand mean is 75.64. So $(X - \bar{X}_G) = 94 - 75.64 = 18.36$. To find the value for the first student who rated Dr. Brunner Good: $75 - 75.64 = -0.64$. The calculations for all students follow.

Excellent	Good	Fair	Poor
18.36	−0.64	−5.64	−7.64
14.36	−7.64	−2.64	−5.64
9.36	1.36	0.36	−3.64
4.36	7.36	2.36	−10.64
	12.36	4.36	−1.64
		−7.64	−10.64
		−10.64	

Then square each of these differences and sum all the values. Thus for the first student:

$$(X - \bar{X}_G)^2 = (94 - 75.64)^2 = (18.36)^2 = 337.09.$$

Finally sum all the squared differences as formula (12–2) directs. Our SS total value is 1,485.09.

	Excellent	Good	Fair	Poor	Total
	337.09	0.41	31.81	58.37	
	206.21	58.37	6.97	31.81	
	87.61	1.85	0.13	13.25	
	19.0	54.17	5.57	113.21	
		152.77	19.01	2.69	
			58.37	113.21	
			113.21		
Total	649.91	267.57	235.07	332.54	1,485.09

To compute the term SSE find the deviation between each observation and its treatment mean. In the example the mean of the first treatment (that is the students who rated Professor Brunner "Excellent") is 87.25. The first student earned a score of 94, so $(X - \bar{X}_c) = (94 - 87.25) = 6.75$. For the first student in the "Good" group $(X - \bar{X}_c) = (75 - 78.20) = -3.20$. The details of each of these calculations follow.

Excellent	Good	Fair	Poor
6.75	−3.2	−2.86	−1
2.75	−10.2	0.14	1
−2.25	−1.2	3.14	3
−7.25	4.8	5.14	−4
	9.8	7.14	5
		−4.86	−4
		−7.86	

Each of these values is squared and then summed for all 22 observations. The values are shown in the following table.

Statistics in Action

Have you ever waited in line for a telephone and it seemed like the person using the phone talked on and on? There is evidence that people actually talk longer on public telephones when someone is waiting. In a recent survey, researchers measured the length of time that 56 shoppers in a mall spent on the phone (1) when they were alone, (2) when a person was using the adjacent phone, and (3) when a person was using an adjacent phone and someone was waiting to use the phone. The study, using the one-way ANOVA technique, showed that the mean time using the telephone was significantly less when the person was alone.

	Excellent	Good	Fair	Poor	Total
	45.5625	10.24	8.18	1	
	7.5625	104.04	0.02	1	
	5.0625	1.44	9.86	9	
	52.5625	23.04	26.42	16	
		96.04	50.98	25	
			23.62	16	
			61.78		
Total	110.7500	234.80	180.86	68	594.41

So the SSE value is 594.41. That is $\Sigma(X - \bar{X}_c)^2 = 594.41$.

Finally we determine SST, the sum of the squares due to the treatments, by subtraction.

$$SST = \text{SS total} - SSE \qquad [12\text{–}4]$$

For this example:

$$SST = \text{SS total} - SSE = 1{,}485.09 - 594.41 = 890.68.$$

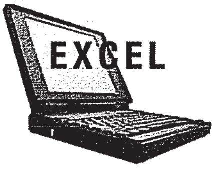
To find the computed value of F, work your way across the ANOVA table. The degrees of freedom for the numerator and the denominator are the same as in **step 4** above when we were finding the critical value of F. The term **mean square** is another expression for an estimate of the variance. The mean square for treatments is SST divided by its degrees of freedom. The result is the **mean square for treatments** and is written MST. Compute the **mean square error** in a similar fashion. To be precise, divide SSE by its degrees of freedom. To complete the process and find F, divide MST by MSE.

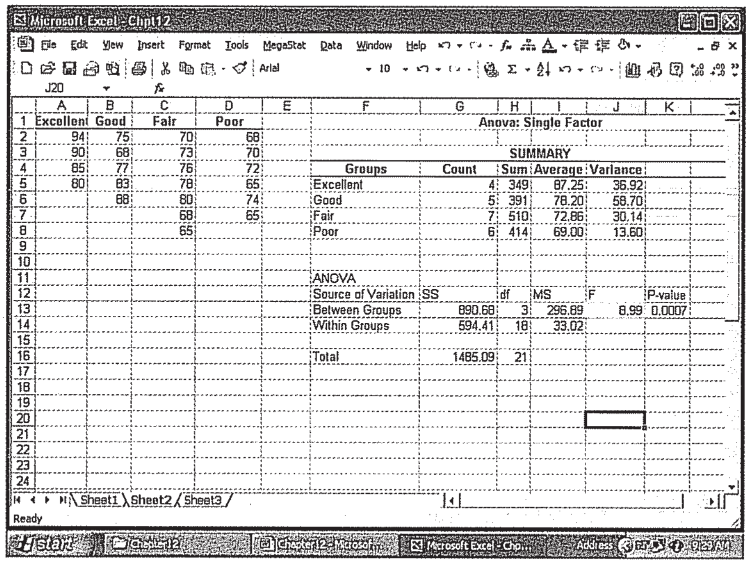
Insert the particular values of F into an ANOVA table and compute the value of F as follows.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Treatments	890.68	3	296.89	8.99
Error	594.41	18	33.02	
Total	1,485.09	21		

The computed value of F is 8.99, which is greater than the critical value of 5.09, so the null hypothesis is rejected. We conclude the population means are not all equal. The mean scores are not the same in each of the four ratings groups. It is likely that the grades students earned in the course are related to the opinion they have of the overall competency and classroom performance of Dr. Brunner the instructor. At this point we can only conclude there is a difference in the treatment means. We cannot determine which treatment groups differ or how many treatment groups differ.

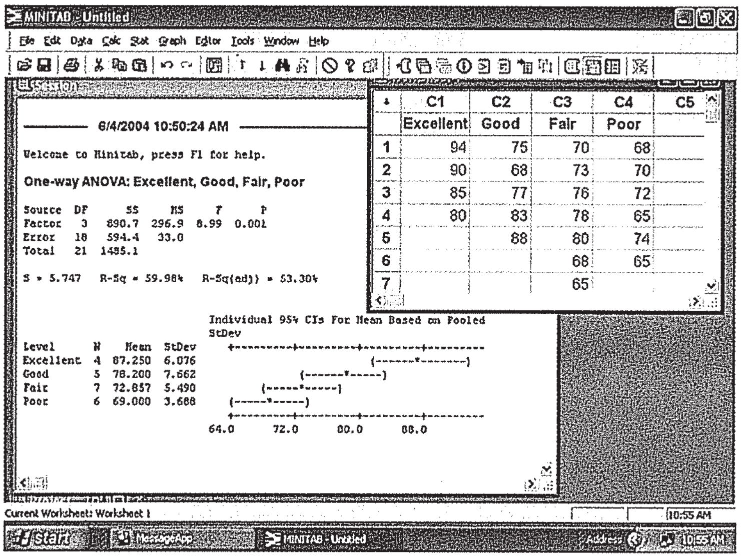
As noted in the previous example, the calculations are tedious if the number of observations in each treatment is large. There are many software packages that will output the results. Following is the Excel output in the form of an ANOVA table for the previous example involving student ratings in Dr. Brunner's Introduction to Marketing. There are some slight differences between the software output and the previous calculations. These differences are due to rounding.





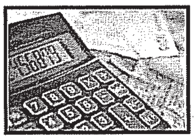
Notice Excel uses the term "Between Groups" for "Treatments" and "Within Groups" for "Error." However, they have the same meanings. The p-value is .0007. This is the probability of finding a value of the test statistic this large or larger when the null hypothesis is true. To put it another way, it is the likelihood of calculating an F value larger than 8.99 with 3 degrees of freedom in the numerator and 18 degrees of freedom in the denominator. So when we reject the null hypothesis in this instance there is a very small likelihood of committing a Type I error!

Following is the MINITAB output from the student ratings example, which is similar to the Excel output. The output is also in the form of an ANOVA table. In addition, MINITAB provides information about the differences between means. This is discussed in the next section.



The MINITAB system uses the term *factor* instead of *treatment,* with the same intended meaning.

Self-Review 12–2



Citrus Clean is a new all-purpose cleaner being test marketed by placing displays in three different locations within various supermarkets. The number of 12-ounce bottles sold from each location within the supermarket is reported below.

Near bread	18	14	19	17
Near beer	12	18	10	16
Other cleaners	26	28	30	32

At the .05 significance level, is there a difference in the mean number of bottles sold at the three locations?

(a) State the null hypothesis and the alternate hypothesis.
(b) What is the decision rule?
(c) Compute the values of SS total, SST, and SSE.
(d) Develop an ANOVA table.
(e) What is your decision regarding the null hypothesis?

Exercises

7. The following is sample information. Test the hypothesis that the treatment means are equal. Use the .05 significance level.

Treatment 1	Treatment 2	Treatment 3
8	3	3
6	2	4
10	4	5
9	3	4

a. State the null hypothesis and the alternate hypothesis.
b. What is the decision rule?
c. Compute SST, SSE, and SS total.
d. Complete an ANOVA table.
e. State your decision regarding the null hypothesis.

8. The following is sample information. Test the hypothesis at the .05 significance level that the treatment means are equal.

Treatment 1	Treatment 2	Treatment 3
9	13	10
7	20	9
11	14	15
9	13	14
12		15
10		

a. State the null hypothesis and the alternate hypothesis.
b. What is the decision rule?
c. Compute SST, SSE, and SS total.
d. Complete an ANOVA table.
e. State your decision regarding the null hypothesis.

9. A real estate developer is considering investing in a shopping mall on the outskirts of Atlanta, Georgia. Three parcels of land are being evaluated. Of particular importance is the income in the area surrounding the proposed mall. A random sample of four families is selected near each proposed mall. Following are the sample results. At the .05 significance level, can the developer conclude there is a difference in the mean income? Use the usual five-step hypothesis testing procedure.

Southwyck Area ($000)	Franklin Park ($000)	Old Orchard ($000)
64	74	75
68	71	80
70	69	76
60	70	78

10. The manager of a computer software company wishes to study the number of hours senior executives spend at their desktop computers by type of industry. The manager selected a sample of five executives from each of three industries. At the .05 significance level, can she conclude there is a difference in the mean number of hours spent per week by industry?

Banking	Retail	Insurance
12	8	10
10	8	8
10	6	6
12	8	8
10	10	10

Inferences about Pairs of Treatment Means

Suppose we carry out the ANOVA procedure and make the decision to reject the null hypothesis. This allows us to conclude that all the treatment means are not the same. Sometimes we may be satisfied with this conclusion, but in other instances we may want to know which treatment means differ. This section provides the details for such a test.

Recall the example regarding student opinions and final scores in Dr. Brunner's Introduction to Marketing. We concluded that there was a difference in the treatment means. That is, the null hypothesis was rejected and the alternate hypothesis accepted. If the student opinions do differ, the question is: Between which groups do the treatment means differ?

Several procedures are available to answer this question. The simplest is through the use of confidence intervals, that is, formula (9–2). From the MINITAB output of the previous example (see page 358), note that the sample mean score for those students rating the instruction Excellent is 87.250, and for those rating the instruction Poor it is 69.000. Thus, those students who rated the instruction Excellent seemingly earned higher grades than those who rated the instruction Poor. Is there enough disparity to justify the conclusion that there is a significant difference in the mean scores of the two groups?

The t distribution, described in Chapters 10 and 11, is used as the basis for this test. Recall that one of the assumptions of ANOVA is that the population variances are the same for all treatments. This common population value is the **mean square error,** or MSE, and is determined by SSE/($n - k$). A confidence interval for the difference between two population means is found by:

CONFIDENCE INTERVAL FOR THE DIFFERENCE IN TREATMENT MEANS

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \textbf{[12–5]}$$