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* and a *two-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.

7 Conduct a test of hypothesis among treatment means using a blocking variable.

8 Conduct a two-way ANOVA with interaction.

A computer manufacturer is about to unveil a new, faster personal computer. The new machine clearly is faster, but initial tests indicate there is more variation in the processing time, which depends on the program being run, and the amount of input and output data. A sample of 16 computer runs, covering a range of production jobs, showed that the standard deviation of the processing time was 22 (hundredths of a second) for the new machine and 12 (hundredths of a second) for the current machine. At the .05 significance level, can we conclude that there is more variation in the processing time of the new machine? (Exercise 24, Goal 2.)

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 sample was selected from the population. We used the *z* distribution (the standard normal distribution) or the *t* 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, using the standard normal distribution as the distribution of the test statistic. In this chapter we expand 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.

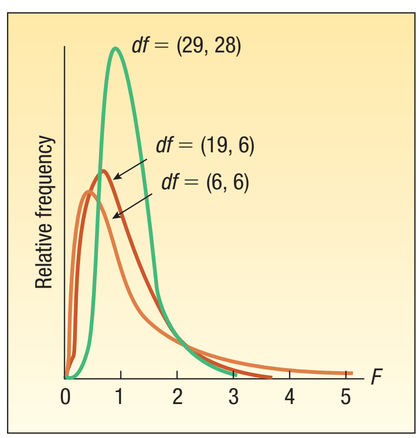
The *F* Distribution

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 (*df*) 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. The final distribution shown has 6 degrees of freedom in the numerator and 6 degrees of freedom in the denominator. We will describe the concept of degrees of freedom later in the chapter. Note that the shape of the curves change as the degrees of freedom change.



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 probability 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 technology and 10 utility stocks shows the same mean rate of return, but there is likely more variation in the technology 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 surfing the Net. 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 we used the t test to investigate whether the means of two independent populations differed. To employ that test, we sometimes assume that the variances of two normal populations are the same. See this list of assumptions on page 381. The F distribution provides a means for conducting a test regarding the variances of two normal populations.

Regardless of whether we want to determine whether 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 is that the variance of one normal population, σ_1^2, equals the variance of the other normal population, σ_2^2. The alternate hypothesis could be 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 random 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. If the null hypothesis is true, 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 B.4. 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 B.4, 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 B.4) 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.

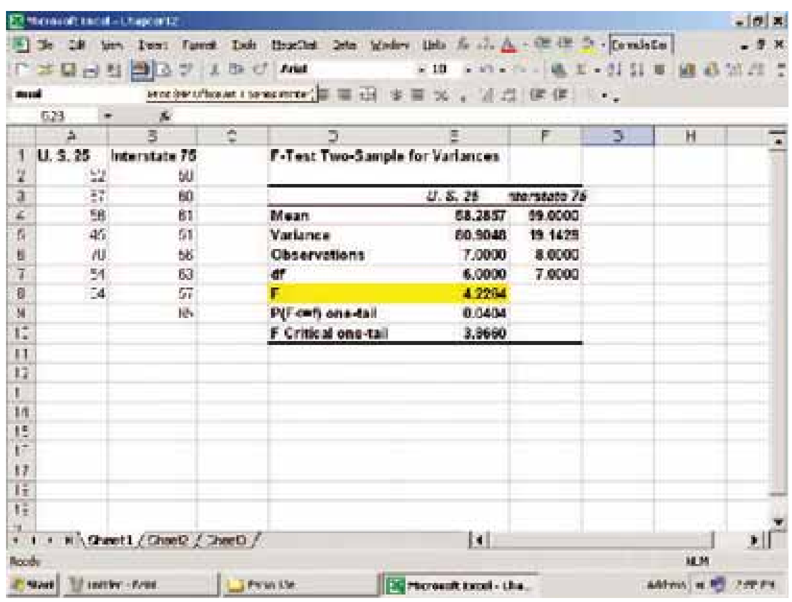
A logical question arises regarding one-tailed tests. For example, suppose in the previous example we suspected that the variance of the times using the U.S. 25 route is *larger* than the variance of the times along the I-75 route. We would state the null and the alternate hypothesis as

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

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 B.4 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).

Excel



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 iPod 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 population 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 error in any of the six independent tests is:

$$P(\text{All correct}) = (.95)(.95)(.95)(.95)(.95)(.95) = .735$$

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 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. For example, treatment refers to how a plot of ground was treated with a particular type of fertilizer. 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 were 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 means are *not* the same.

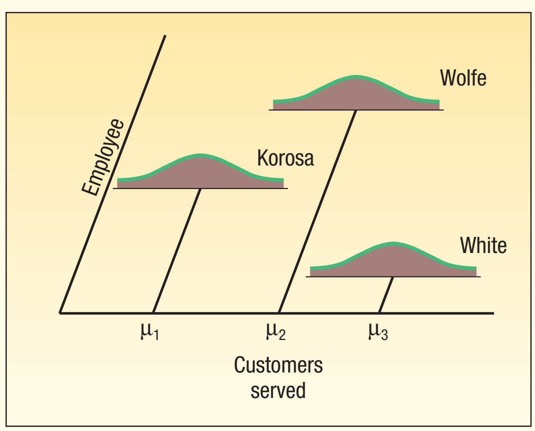


CHART 12–1 Case Where Treatment Means Are Different

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.

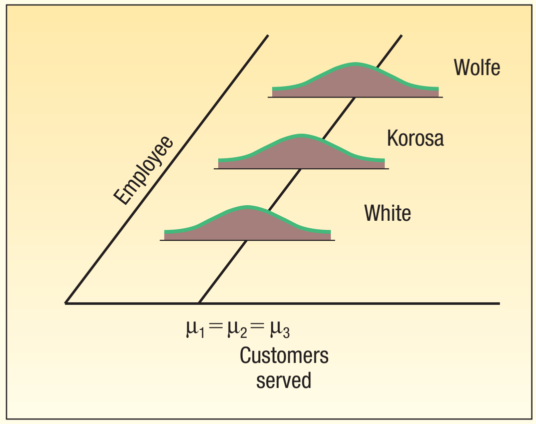


CHART 12–2 Case Where Treatment Means Are the Same

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 412 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.

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.

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

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 treatment means are not the same. There is a difference in the mean number of customers served by the three employees.

Here's another example, which deals with samples of different sizes.

Example

Recently airlines have cut services, such as meals and snacks during flights, and started charging extra for some services such as accommodating overweight luggage, last-minute flight changes, and pets traveling in the cabin. However, they are still very concerned about service. Recently a group of four carriers (we have used historical names for confidentiality) hired Brunner Marketing Research, Inc., to survey recent passengers regarding their level of satisfaction with a recent flight. The survey included questions on ticketing, boarding, in-flight service, baggage handling, pilot communication, and so forth. Twenty-five questions offered a range of possible answers: excellent, good, fair, or poor. A response of excellent was given a score of 4, good a 3, fair a 2, and poor a 1. These responses were then totaled, so the total score was an indication of the satisfaction with the flight. The greater the score, the higher the level of satisfaction with the service. The highest possible score was 100.

Brunner randomly selected and surveyed passengers from the four airlines. Below is the sample information. Is there a difference in the mean satisfaction level among the four airlines? Use the .01 significance level.

Eastern	TWA	Allegheny	Ozark
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 use the 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 airlines.

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

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

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 scores for the four airlines. If H_0 is rejected, we conclude that there is a difference in at least one pair of mean scores, 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 B.4. 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.

$$\text{Degrees of freedom in the numerator} = k - 1 = 4 - 1 = 3$$

$$\text{Degrees of freedom in the denominator} = n - k = 22 - 4 = 18$$

Refer to Appendix B.4 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 or the random error.

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:

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

where:

X is each sample observation.
$\overline{X}_G$ is the overall or 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:

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

where:

$\overline{X}_c$ is the sample mean for treatment c.

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.

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

	Eastern	TWA	Allegheny	Ozark	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	1,664
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 sampled passenger had a score of 94 and the overall or grand mean is 75.64. So $(X - \overline{X}_G) = 94 - 75.64 = 18.36$. For the last passenger $(X - \overline{X}_G) = 65 - 75.64 = -10.64$. The calculations for all other passengers follow.

Eastern	TWA	Allegheny	Ozark
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 passenger:

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

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

	Eastern	TWA	Allegheny	Ozark	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	
	19.01	54.17	5.57	113.21	
		152.77	19.01	2.69	
			58.37	113.21	
			113.21		
Total	649.92	267.57	235.07	332.54	1,485.10

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 passengers on Eastern Airlines) is 87.25, found by $\overline{X}_E = 349/4$. The subscript E refers to Eastern Airlines.

The first passenger rated Eastern a 94, so $(X - \overline{X}_E) = (94 - 87.25) = 6.75$. The first passenger in the TWA group responded with a total score of 75, so $(X - \overline{X}_{TWA}) = (75 - 78.20) = -3.2$. The detail for all the passengers follows.

Eastern	TWA	Allegheny	Ozark
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	

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.

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

	Eastern	TWA	Allegheny	Ozark	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 $\Sigma(X - \overline{X}_c)^2 = 594.41$.

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

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

For this example:

$$\text{SST} = \text{SS total} - \text{SSE} = 1{,}485.10 - 594.41 = 890.69.$$

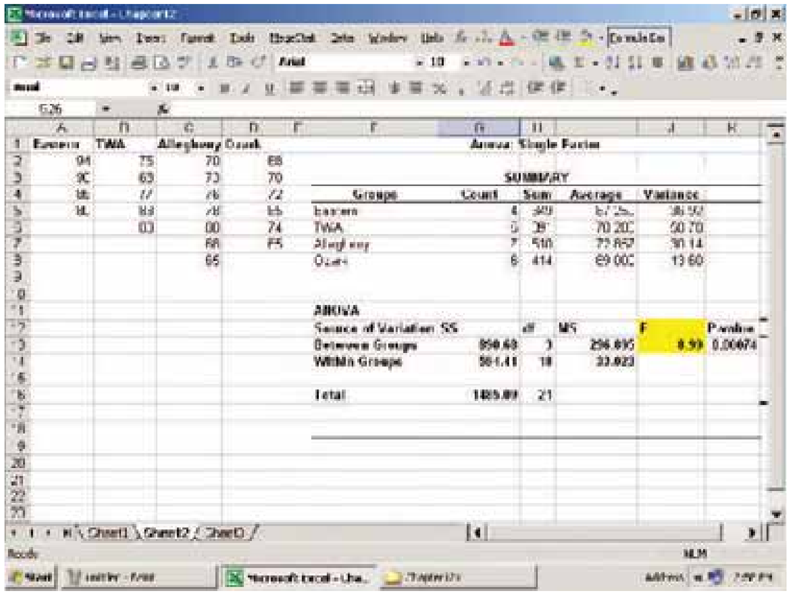
To find the computed value of *F*, work your way across the ANOVA table. The degrees of freedom for the numerator and the denominator are the same as in step 4 on page 416 when we were finding the critical value of *F*. The term **mean square** is another expression for an estimate of the variance. The 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	890.69	3	296.90	8.99
Error	594.41	18	33.02	
Total	1,485.10	21		

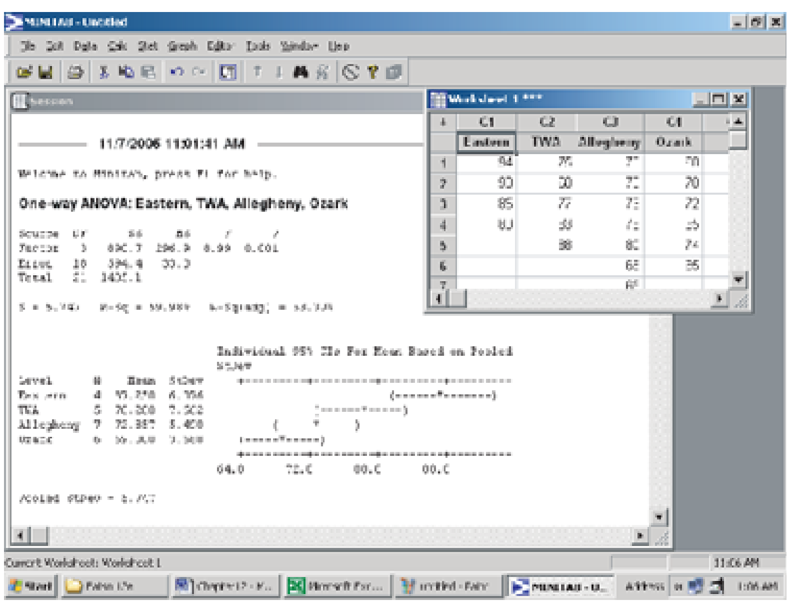
The computed value of *F* is 8.99, which is greater than the critical value of 5.09, so the null hypothesis is rejected. We conclude the population means are not all equal. The mean scores are not the same for the four airlines. It is likely that the passenger scores are related to the particular airline. 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 airlines and passenger ratings. 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 airline passenger 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 hypotheses.
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 hypotheses.
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 by type of industry spend at their desktop computers. 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 that in the Brunner Research example regarding airline passenger ratings there was a difference in the treatment means. That is, the null hypothesis was rejected and the alternate hypothesis accepted. If the passenger ratings 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 computer output of the previous example (see page 420), note that the sample mean score for those passengers rating Eastern's service is 87.25, and for those rating Ozark's service the sample mean score is 69.00. Is there enough disparity to justify the conclusion that there is a significant difference in the mean satisfaction scores of the two airlines?

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 populations is found by:

CONFIDENCE INTERVAL FOR THE DIFFERENCE IN TREATMENT MEANS

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

where

$\bar{X}_1$ is the mean of the first sample.
$\bar{X}_2$ is the mean of the second sample.
t is obtained from Appendix B.2. The degrees of freedom is equal to $n - k$.
MSE is the mean square error term obtained from the ANOVA table [SSE/$(n - k)$].
n_1 is the number of observations in the first sample.
n_2 is the number of observations in the second sample.

How do we decide whether there is a difference in the treatment means? If the confidence interval includes zero, there is *not* a difference between the treatment means. For example, if the left endpoint of the confidence interval has a negative sign and the right endpoint has a positive sign, the interval includes zero and the two means do not differ. So if we develop a confidence interval from formula (12–5) and find the difference in the sample means was 5.00, that is, if $\bar{X}_1 - \bar{X}_2 = 5$ and $t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 12$, the confidence interval would range from -7.00 up to 17.00. To put it in symbols:

$$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 5.00 \pm 12.00 = -7.00 \text{ up to } 17.00$$

Note that zero is included in this interval. Therefore, we conclude that there is no significant difference in the selected treatment means.

On the other hand, if the endpoints of the confidence interval have the same sign, this indicates that the treatment means differ. For example, if $\bar{X}_1 - \bar{X}_2 = -0.35$ and $t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.25$, the confidence interval would range from -0.60 up to -0.10. Because -0.60 and -0.10 have the same sign, both negative, zero is not in the interval and we conclude that these treatment means differ.

Using the previous airline example let us compute the confidence interval for the difference between the mean scores of passengers on Eastern and Ozark Airlines. With a 95 percent level of confidence, the endpoints of the confidence interval are 10.46 and 26.04.

$$\begin{aligned}(\bar{X}_E - \bar{X}_O) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_E} + \frac{1}{n_O}\right)} &= (87.25 - 69.00) \pm 2.101\sqrt{33.0\left(\frac{1}{4} + \frac{1}{6}\right)} \\ &= 18.25 \pm 7.79\end{aligned}$$

where

$\bar{X}_E$ is 87.25.
$\bar{X}_O$ is 69.00.
t is 2.101: from Appendix B.2 with $(n - k) = 22 - 4 = 18$ degrees of freedom.
MSE is 33.0: from the ANOVA table with SSE/$(n - k)$ = 594.4/18.
n_E is 4.
n_O is 6.

The 95 percent confidence interval ranges from 10.46 up to 26.04. Both endpoints are positive; hence, we can conclude these treatment means differ significantly.

That is, passengers on Eastern rated service significantly different from those on Ozark.

Approximate results can also be obtained directly from the MINITAB output. Following is the lower portion of the output from page 420. On the left side is the number of observations, the mean, and the standard deviation for each treatment. Seven passengers on Allegheny rated the service as 72.857 with a standard deviation of 5.490.

```
                                Individual 95% CIs For Mean
                                Based on Pooled StDev
Level       N    Mean   StDev   ---------+---------+---------+---------
Eastern     4  87.250   6.076                          (------*-----)
TWA         5  78.200   7.662               (-----*-----)
Allegheny   7  72.857   5.490         (-----*-----)
Ozark       6  69.000   3.688   (-----*-----)
                                ---------+---------+---------+---------
                              64.0      72.0      80.0      88.0
```

On the right side of the printout is a confidence interval for each treatment mean. The asterisk (*) indicates the location of the treatment mean and the open parenthesis and close parenthesis, the endpoints of the confidence interval. In those instances where the intervals overlap, the treatment means may not differ. If there is no common area in the confidence intervals, that pair of means differ.

The endpoints of a 95 percent confidence interval for mean passenger scores for Allegheny are about 69 and 77. For Ozark the endpoints of the 95 percent confidence interval for the mean passenger score are about 64 and 73. There is common area between these points, so we conclude this pair of means does not differ. In other words, there is no significant difference between the mean passenger ratings for Allegheny and Ozark. The difference in the mean scores is due to chance.

There are two pairs of means that differ. The mean scores of passengers on Eastern Airlines are significantly different from the mean ratings of passengers on Allegheny and on Ozark. There is no common area between these two pairs of confidence intervals.

We should emphasize that this investigation is a step-by-step process. The initial step is to conduct the ANOVA test. Only if the null hypothesis that the treatment means are equal is rejected should any analysis of the individual treatment means be attempted.

Self-Review 12–3

The following data are the semester tuition charges ($000) for a sample of private colleges in various regions of the United States. At the .05 significance level, can we conclude there is a difference in the mean tuition rates for the various regions?

Northeast ($000)	Southeast ($000)	West ($000)
10	8	7
11	9	8
12	10	6
10	8	7
12		6

(a) State the null and the alternate hypotheses.
(b) What is the decision rule?
(c) Develop an ANOVA table. What is the value of the test statistic?
(d) What is your decision regarding the null hypothesis?
(e) Could there be a significant difference between the mean tuition in the Northeast and that of the West? If so, develop a 95 percent confidence interval for that difference.

Exercises

11. Given the following sample information, test the hypothesis that the treatment means are equal at the .05 significance level.

Treatment 1	Treatment 2	Treatment 3
8	3	3
11	2	4
10	1	5
	3	4
	2	

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.
f. If H_0 is rejected, can we conclude that treatment 1 and treatment 2 differ? Use the 95 percent level of confidence.

12. Given the following sample information, test the hypothesis that the treatment means are equal at the .05 significance level.

Treatment 1	Treatment 2	Treatment 3
3	9	6
2	6	3
5	5	5
1	6	5
3	8	5
1	5	4
	4	1
	7	5
	6	
	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.
f. If H_0 is rejected, can we conclude that treatment 2 and treatment 3 differ? Use the 95 percent level of confidence.

13. A senior accounting major at Midsouth State University has job offers from four CPA firms. To explore the offers further, she asked a sample of recent trainees how many months each worked for the firm before receiving a raise in salary. The sample information is submitted to MINITAB with the following results:

```
Analysis of Variance
Source    DF      SS       MS       F       P
Factor     3    32.33    10.78    2.36    0.133
Error     10    45.67     4.57
Total     13    78.00
```

At the .05 level of significance, is there a difference in the mean number of months before a raise was granted among the four CPA firms?

14. A stock analyst wants to determine whether there is a difference in the mean rate of return for three types of stock: utility, retail, and banking stocks. The following output is obtained:

```
Analysis of Variance
Source      DF        SS        MS        F        P
Factor       2     86.49     43.25    13.09    0.001
Error       13     42.95      3.30
Total       15    129.44
                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level        N      Mean     StDev    ----------+----------+----------+------
Utility      5    17.400     1.916                      (------*------)
Retail       5    11.620     0.356     (------*------)
Banking      6    15.400     2.356                  (------*------)
                                      ----------+----------+----------+------
Pooled StDev = 1.818                          12.0       15.0       18.0
```

a. Using the .05 level of significance, is there a difference in the mean rate of return among the three types of stock?

b. Suppose the null hypothesis is rejected. Can the analyst conclude there is a difference between the mean rates of return for the utility and the retail stocks? Explain.

Two-Way Analysis of Variance

In the airline passenger ratings example, we divided the total variation into two categories: the variation between the treatments and the variation within the treatments. We also called the variation within the treatments the error or the random variation. To put it another way, we considered only two sources of variation, that due to the treatments and the random differences. In the airline passenger ratings example there may be other causes of variation. These factors might include, for example, the season of the year, the particular airport, or the number of passengers on the flight.

The benefit of considering other factors is that we can reduce the error variance. That is, if we can reduce the denominator of the *F* statistic (reducing the error variance or, more directly, the SSE term), the value of *F* will be larger, causing us to reject the hypothesis of equal treatment means. In other words, if we can explain more of the variation, then there is less "error." An example will clarify the reduction in the error variance.

Example

WARTA, the Warren Area Regional Transit Authority, is expanding bus service from the suburb of Starbrick into the central business district of Warren. There are four routes being considered from Starbrick to downtown Warren: (1) via U.S. 6, (2) via the West End, (3) via the Hickory Street Bridge, and (4) via Route 59. WARTA conducted several tests to determine whether there was a difference in the mean travel times along the four routes. Because there will be many different drivers, the test was set up so each driver drove along each of the four routes. Below is the travel time, in minutes, for each driver–route combination.

	Travel Time From Starbrick to Warren (minutes)			
Driver	**U.S. 6**	**West End**	**Hickory St.**	**Rte. 59**
Deans	18	17	21	22
Snaverly	16	23	23	22
Ormson	21	21	26	22
Zollaco	23	22	29	25
Filbeck	25	24	28	28

At the .05 significance level, is there a difference in the mean travel time along the four routes? If we remove the effect of the drivers, is there a difference in the mean travel time?

Solution

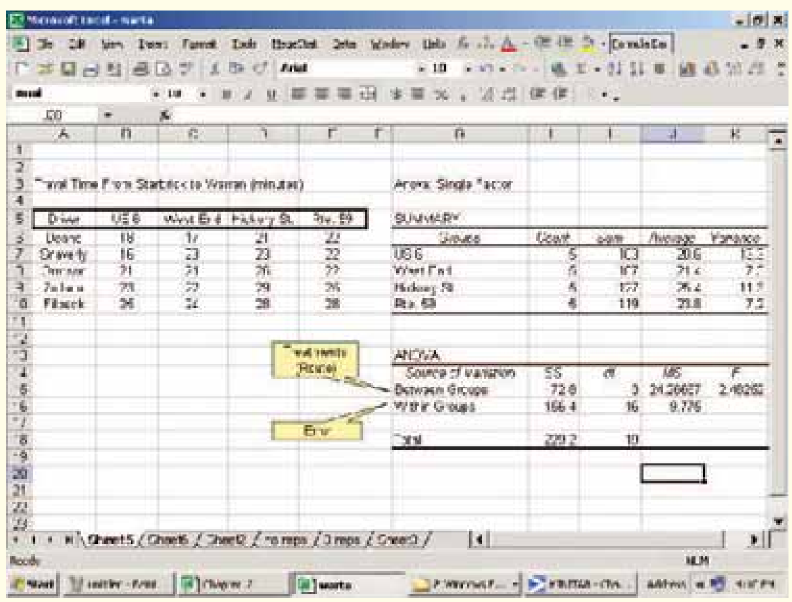
To begin, we conduct a test of hypothesis using a one-way ANOVA. That is, we consider only the four routes. Under this condition the variation in travel times is either due to the treatments or it is random. The null hypothesis and the alternate hypothesis for comparing the mean travel time along the four routes are:

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1: Not all treatment means are the same.

There are four routes, so for the numerator the degrees of freedom is $k - 1 = 4 - 1 = 3$. There are 20 observations, so the degrees of freedom in the denominator is $n - k = 20 - 4 = 16$. From Appendix B.4, with the .05 significance level, the critical value of F is 3.24. The decision rule is to reject the null hypothesis if the computed value of F is greater than 3.24.

We use Excel to perform the calculations. The computed value of F is 2.482, so our decision is to not reject the null hypothesis. We conclude there is no difference in the mean travel time along the four routes. There is no reason to select one of the routes as faster than the other.

Excel



From the above Excel output the mean travel times along the routes were: 20.6 minutes along U.S. 6, 21.4 minutes along the West End route, 25.4 minutes using Hickory Street, and 23.8 minutes using Route 59. We conclude these differences could reasonably be attributed to chance. From the ANOVA table we note: SST is 72.8, SSE is 156.4, and SS total is 229.2.

In the above example we considered the variation due to the treatments (routes) and took all the remaining variation to be random. If we could consider the effect of the several drivers, this would allow us to reduce the SSE term, which would lead to a larger value of F. The second treatment variable, the drivers in this case, is referred to as a **blocking variable.**

BLOCKING VARIABLE A second treatment variable that when included in the ANOVA analysis will have the effect of reducing the SSE term.

In this case we let the drivers be the blocking variable, and removing the effect of the drivers from the SSE term will change the *F* ratio for the treatment variable. First, we need to determine the sum of squares due to the blocks.

In a two-way ANOVA, the sum of squares due to blocks is found by the following formula.

$$\text{SSB} = k\Sigma(\overline{X}_b - \overline{X}_G)^2 \qquad \textbf{[12–6]}$$

where

k is the number of treatments.
b is the number of blocks.
$\overline{X}_b$ is the sample mean of block b.
$\overline{X}_G$ is the overall or grand mean.

From the calculations below, the means for the respective drivers are 19.5 minutes, 21 minutes, 22.5 minutes, 24.75 minutes, and 26.25 minutes. The overall mean is 22.8 minutes, found by adding the travel time for all 20 drives (456 minutes) and dividing by 20.

	Travel Time From Starbrick to Warren (minutes)					
Driver	**U.S. 6**	**West End**	**Hickory St.**	**Rte. 59**	**Driver Sums**	**Driver Means**
Deans	18	17	21	22	78	19.5
Snaverly	16	23	23	22	84	21
Ormson	21	21	26	22	90	22.5
Zollaco	23	22	29	25	99	24.75
Filbeck	25	24	28	28	105	26.25

Substituting this information into formula (12–6) we determine SSB, the sum of squares due to the drivers (the blocking variable), is 119.7.

$$\begin{aligned}\text{SSB} &= k\Sigma(\overline{X}_b - \overline{X}_G)^2\\ &= 4(19.5 - 22.8)^2 + 4(21.0 - 22.8)^2 + 4(22.5 - 22.8)^2\\ &\quad + 4(24.75 - 22.8)^2 + 4(26.25 - 22.8)^2\\ &= 119.7\end{aligned}$$

The same format is used in the two-way ANOVA table as in the one-way case, except there is an additional row for the blocking variable. SS total and SST are calculated as before, and SSB is found from formula (12–6). The SSE term is found by subtraction.

SUM OF SQUARES ERROR, TWO-WAY $\quad \text{SSE} = \text{SS total} - \text{SST} - \text{SSB} \quad$ **[12–7]**

The values for the various components of the ANOVA table are computed as follows.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Treatments	SST	$k - 1$	$\text{SST}/(k - 1) = \text{MST}$	MST/MSE
Blocks	SSB	$b - 1$	$\text{SSB}/(b - 1) = \text{MSB}$	MSB/MSE
Error	SSE	$(k - 1)(b - 1)$	$\text{SSE}/(k - 1)(b - 1) = \text{MSE}$	
Total	SS total	$n - 1$		

SSE is found by formula (12–7).

$$SSE = SS\ total - SST - SSB = 229.2 - 72.8 - 119.7 = 36.7$$

Source of Variation	(1) Sum of Squares	(2) Degrees of Freedom	(3) Mean Square (1)/(2)
Treatments	72.8	3	24.27
Blocks	119.7	4	29.93
Error	36.7	12	3.06
Total	229.2	19	

There is disagreement at this point. If the purpose of the blocking variable (the drivers in this example) was only to reduce the error variation, we should not conduct a test of hypothesis for the difference in block means. That is, if our goal was to reduce the MSE term, then we should not test a hypothesis regarding the blocking variable. On the other hand, we may wish to give the blocks the same status as the treatments and conduct a test of hypothesis. In the latter case, when the blocks are important enough to be considered as a second factor, we refer to this as a **two-factor experiment.** In many cases the decision is not clear. In our example we are concerned about the difference in the travel time for the different drivers, so we will conduct the test of hypothesis. The two sets of hypotheses are:

1. H_0: The treatment means are the same ($\mu_1 = \mu_2 = \mu_3 = \mu_4$).
   H_1: The treatment means are not the same.
2. H_0: The block means are the same ($\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$).
   H_1: The block means are not the same.

First, we will test the hypothesis concerning the treatment means. There are $k - 1 = 4 - 1 = 3$ degrees of freedom in the numerator and $(b - 1)(k - 1) = (5 - 1)(4 - 1) = 12$ degrees of freedom in the denominator. Using the .05 significance level, the critical value of F is 3.49. The null hypothesis that the mean times for the four routes are the same is rejected if the F ratio exceeds 3.49.

$$F = \frac{MST}{MSE} = \frac{24.27}{3.06} = 7.93$$

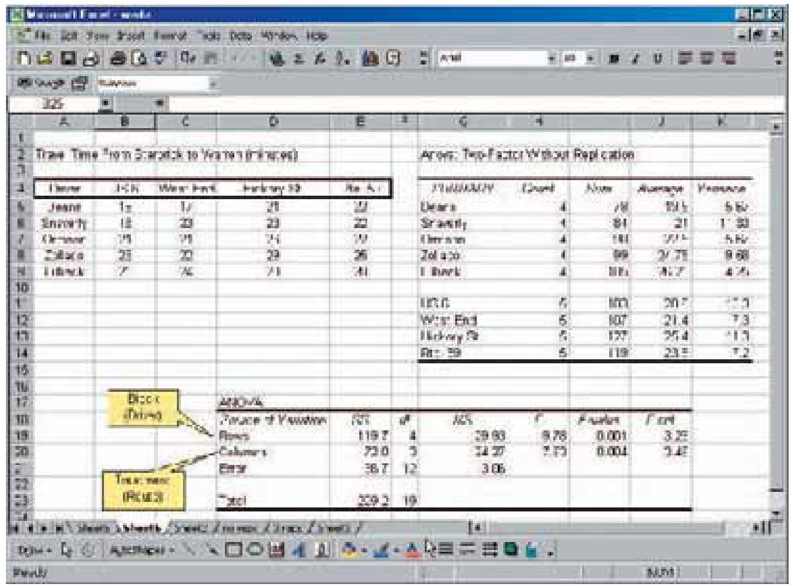
The null hypothesis is rejected and the alternate accepted. We conclude that the mean travel time is not the same for all routes. WARTA will want to conduct some tests to determine which treatment means differ.

Next, we test to find whether the travel time is the same for the various drivers. The degrees of freedom in the numerator for blocks is $b - 1 = 5 - 1 = 4$. The degrees of freedom for the denominator are the same as before: $(b - 1)(k - 1) = (5 - 1)(4 - 1) = 12$. The null hypothesis that the block means are the same is rejected if the F ratio exceeds 3.26.

$$F = \frac{MSB}{MSE} = \frac{29.93}{3.06} = 9.78$$

The null hypothesis is rejected, and the alternate is accepted. The mean time is not the same for the various drivers. Thus, WARTA management can conclude, based on the sample results, that there is a difference in the routes and in the drivers.

The Excel spreadsheet has a two-factor ANOVA procedure. The output for the WARTA example just completed follows. The results are the same as reported earlier. In addition the Excel output reports the p-values. The p-value for the null hypothesis regarding the drivers is .001 and .004 for the routes. These p-values confirm that the null hypotheses for treatments and blocks should both be rejected because the p-value is less than the significance level.



Self-Review 12–4

Rudduck Shampoo sells three shampoos, one each for dry, normal, and oily hair. Sales, in millions of dollars, for the past five months are given in the following table. Using the .05 significance level, test whether the mean sales differ for the three types of shampoo or by month.

	Sales ($ million)		
Month	**Dry**	**Normal**	**Oily**
June	7	9	12
July	11	12	14
August	13	11	8
September	8	9	7
October	9	10	13

Exercises

For exercises 15 and 16, conduct a test of hypothesis to determine whether the block or the treatment means differ. Using the .05 significance level: (a) state the null and alternate hypotheses for treatments; (b) state the decision rule for treatments; and (c) state the null and alternate hypotheses for blocks. Also, state the decision rule for blocks, then: (d) compute SST, SSB, SS total, and SSE; (e) complete an ANOVA table; and (f) give your decision regarding the two sets of hypotheses.

15. The following data are given for a two-factor ANOVA.

	Treatment	
Block	**1**	**2**
A	46	31
B	37	26
C	44	35

16. The following data are given for a two-factor ANOVA.

	Treatment		
Block	**1**	**2**	**3**
A	12	14	8
B	9	11	9
C	7	8	8

17. Chapin Manufacturing Company operates 24 hours a day, five days a week. The workers rotate shifts each week. Management is interested in whether there is a difference in the number of units produced when the employees work on various shifts. A sample of five workers is selected and their output recorded on each shift. At the .05 significance level, can we conclude there is a difference in the mean production rate by shift or by employee?

	Units Produced		
Employee	**Day**	**Afternoon**	**Night**
Skaff	31	25	35
Lum	33	26	33
Clark	28	24	30
Treece	30	29	28
Morgan	28	26	27

18. There are three hospitals in the Tulsa, Oklahoma, area. The following data show the number of outpatient surgeries performed at each hospital last week. At the .05 significance level, can we conclude there is a difference in the mean number of surgeries performed by hospital or by day of the week?

	Number of Surgeries Performed		
Day	**St. Luke's**	**St. Vincent**	**Mercy**
Monday	14	18	24
Tuesday	20	24	14
Wednesday	16	22	14
Thursday	18	20	22
Friday	20	28	24

Two-Way ANOVA with Interaction

In the previous section, we studied the separate or independent effects of two variables, routes into the city and drivers, on mean travel time. The sample results indicated differences in mean time among the routes. Perhaps this is simply related to differences in the distance among the routes. The results also indicated differences in the mean drive time among the several drivers. Perhaps this difference is explained by differing average speeds by the drivers regardless of the route. There is another effect that may influence travel time. This is called an **interaction effect** between route and driver on travel time. For example, is it possible that one of the drivers is especially good driving one or more of the routes? Perhaps one driver knows how to effectively time the traffic lights or how to avoid heavily congested intersections for one or more of the routes. In this case, the combined effect of driver and route may also explain differences in mean travel time. To measure interaction effects it is necessary to have at least two observations in each cell.

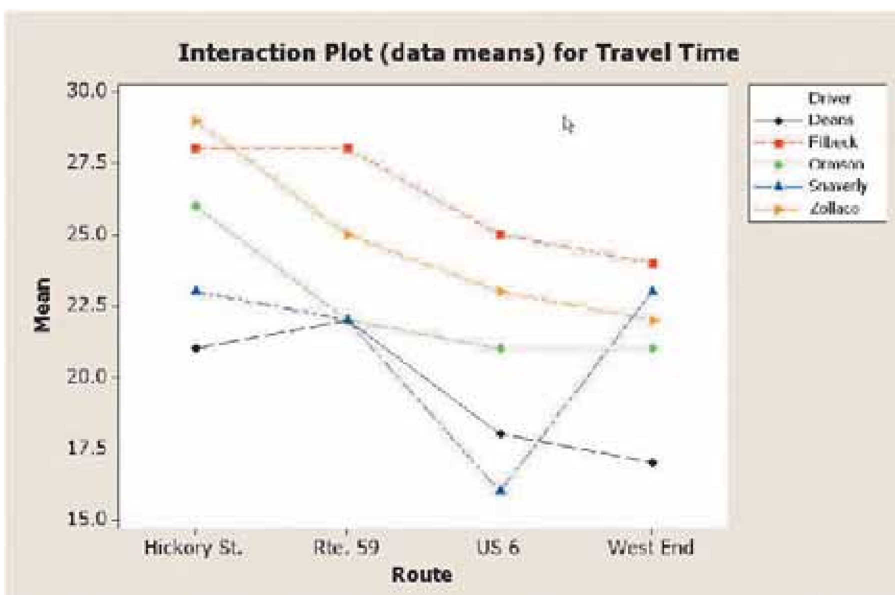
When we use a two-way ANOVA to study interaction, instead of using the terms treatments and blocks, we now call the two variables **factors.** So in this method there is a route factor, a driver factor, and an interaction of the two factors. That is, there is an *effect* for the routes, for the driver, and for the interaction of drivers and routes.

Interaction occurs if the combination of two factors has some effect on the variable under study, in addition to each factor alone. We refer to the variable being studied as the **response** variable. An everyday illustration of interaction is the effect of diet and exercise on weight. It is generally agreed that a person's weight (the response variable) can be controlled with two factors, diet and exercise. Research shows that weight is affected by diet alone and that weight is affected by exercise alone. However, the general recommended method to control weight is based on the combined or *interaction* effect of diet and exercise.

INTERACTION The effect of one factor on a response variable differs depending on the value of another factor.

Interaction Plots

One way to study interaction is by plotting factor means in a graph called an interaction plot. Consider the bus driver example in the previous section. WARTA, the Warren Area Regional Transit Authority, wants to study the mean travel time for different routes and drivers. To complete the study, we should also explore the prospect of interaction between driver and route. We begin by graphing the mean travel times on each route for each driver and connect the points. We compute Deans' mean travel times for each route and plot them in a graph of mean travel times versus route. We repeat this process for each of the drivers. The interaction plot follows.



This plot helps us understand the interaction between the effects of drivers and routes on travel time. If the line segments for the drivers appear essentially parallel, then there is probably no interaction. On the other hand, if the line segments **do not**

seem to be parallel or cross, this suggests an interaction between the factors. The above plot suggests interaction because:

- The line segments for Zollaco and Filbeck cross over each other.
- Snaverly's line segment from U.S. 6 to West End crosses over three other line segments.

These observations suggest an interaction between driver and route.

Hypothesis Test for Interaction

The next step is to conduct statistical tests to further investigate the possible interaction effects. In summary, our study of travel times has several questions:

- Is there an interaction between routes and drivers?
- Are the travel times for drivers the same?
- Are the travel times for the routes the same?

Of the three questions, we are most interested in the test for interactions.

We can investigate these questions statistically by extending the two-way ANOVA procedure presented in the previous section. We add another source of variation, namely, the interaction. However, in order to estimate the "error" sum of squares, we need at least two measurements for each driver/route combination. So suppose the experiment reported on page 426 is repeated by measuring two more travel times for each driver and route combination. So, we **replicate** the experiment. Now we have three observations for each driver/route combination. Using the mean of three travel times for each driver/route combination we get a more reliable measure of the mean travel time. The results of replicating the experiment are in the following Excel table. Note that the data should be entered in this exact format so we can use statistical software.

Microsoft Excel - two-way anova with interaction

	A	B	C	D	E	F	G	H
1			US 6	West End	Hickory St	Route 59		
2		Deans	18	14	20	19		
3		Deans	15	17	21	22		
4		Deans	21	20	22	25		
5		Snaverly	19	20	24	24		
6		Snaverly	15	24	23	22		
7		Snaverly	14	25	22	20		
8		Ormson	19	23	25	23		
9		Ormson	21	21	29	23		
10		Ormson	23	19	24	20		
11		Zollaco	24	20	30	26		
12		Zollaco	20	24	28	25		
13		Zollaco	25	22	29	24		
14		Filbeck	27	24	28	28		
15		Filbeck	25	24	28	30		
16		Filbeck	23	24	28	26		
17								

Sheet4 | Sheet1 | Sheet2 | Sheet3

Ready NUM

To explain the spreadsheet, consider the "20, 21, 22" for the rows labeled "Deans" and column labeled "Hickory St." These are the three travel time measurements for Deans to drive the Hickory Street route. Specifically, Deans drove the Hickory Street route the first time in 20 minutes, 21 minutes on the second trip, and 22 minutes on the third trip.

The ANOVA now has three sets of hypotheses to test:

1. H_0: There is no interaction between drivers and routes.
   H_1: There is interaction between drivers and routes.
2. H_0: The driver means are the same.
   H_1: The driver means are *not* the same.
3. H_0: The route means are the same.
   H_1: The route means are *not* the same.

Note that we will label the driver effect as **Factor A** and the route effect as **Factor B.**

Each of these hypotheses is tested using the familiar F statistic. We can use a decision rule for each of the above tests or we can use the p-values for each test. In this case we will use the .05 significance level and compare the p-value generated by the statistical software with the significance level. So the various null hypotheses are rejected if the computed p-value is less than .05. Instead of computing treatment and block sum of squares, we compute factor and interaction sum of squares. The computations for the factor sum of squares are very similar to the SST and SSB as computed before. See formulas (12–4) and (12–6). The sum of squares due to possible interaction is:

$$\text{SSI} = (k-1)(b-1)\Sigma\Sigma(\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_G)^2 \qquad \textbf{[12–8]}$$

where

- i is a subscript or label representing a route.
- j is a subscript or label representing a driver.
- k is the number of Factor A (route effect) levels.
- b is the number of Factor B (driver effect) levels.
- n is the number of observations.
- $\bar{X}_{ij}$ is the mean travel time on a route, *i,* for driver, *j*. Note these are the means that we plotted in the graph on page 432.
- $\bar{X}_{i.}$ is the mean travel time for route *i*. Note the dot shows that the mean is calculated over all drivers. These are the route means that we compared on page 429.
- $\bar{X}_{.j}$ is the mean travel time for driver *j*. Note the dot shows that the mean is calculated over all routes. These are the driver means that we compared on page 429.
- $\bar{X}_G$ is the grand mean.

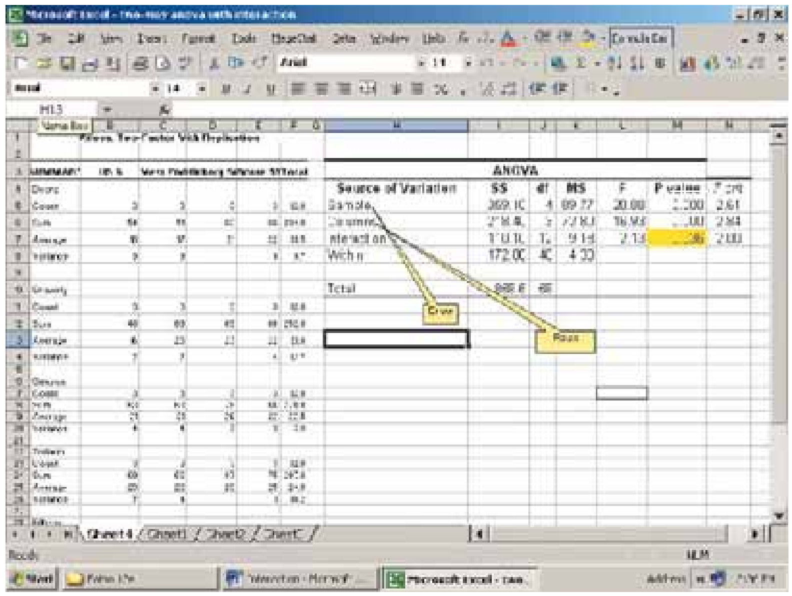
Once you have SSI, then SSE is found as:

$$\text{SSE} = \text{SS total} - \text{SS Factor A} - \text{SS Factor B} - \text{SSI} \qquad \textbf{[12–9]}$$

The complete ANOVA table including interactions is:

Source	Sum of Squares	*df*	Mean Square	*F*
Route	Factor A	$k - 1$	$\text{SSA}/(k-1) = \text{MSA}$	MSA/MSE
Driver	Factor B	$b - 1$	$\text{SSB}/(b-1) = \text{MSB}$	MSB/MSE
Interaction	SSI	$(k-1)(b-1)$	$\text{SSI}/(k-1)(b-1) = \text{MSI}$	MSI/MSE
Error	SSE	$n - kb$	$\text{SSE}/(n-kb) = \text{MSE}$	
Total	SS total	$n - 1$		

The resulting Excel output shows the summary descriptive statistics for each driver and an ANOVA table.



The *p*-value for interactions of 0.036 (noted in yellow), is less than our significance level of 0.05. So our decision is to reject the null hypothesis of no interaction and conclude that the combination of route and driver has a significant effect on the response variable travel time.

Interaction effects provide information about the combined effects of variables. If interaction is present, then you should conduct a one-way ANOVA to test differences in the factor means for each level of the other factor. This analysis requires some time and work to complete but the results can show some very interesting results.

We will continue the analysis by conducting a one-way ANOVA for each driver testing the hypothesis: H_0: Route travel times are equal. The results follow.

Deans: H_0: Route travel times are equal.

Source	DF	SS	MS	F	P
Dean RTE	3	51.00	17.00	2.43	0.140
Error	8	56.00	7.00		
Total	11	107.00			

Snaverly: H_0: Route travel times are equal.

Source	DF	SS	MS	F	P
SN RTE	3	102.00	34.00	7.16	0.012
Error	8	38.00	4.75		
Total	11	140.00			

Ormson: H_0: Route travel times are equal.

Source	DF	SS	MS	F	P
Ormson RTE	3	51.00	17.00	3.78	0.059
Error	8	36.00	4.50		
Total	11	87.00			

Zollaco: H_0: Route travel times are equal.

Source	DF	SS	MS	F	P
Z-RTE	3	86.25	28.75	8.85	0.006
Error	8	26.00	3.25		
Total	11	112.25			

Filbeck: H_0: Route travel times are equal.

Source	DF	SS	MS	F	P
Filbeck RTE	3	38.25	12.75	6.38	0.016
Error	8	16.00	2.00		
Total	11	54.25			

Recall the results from the two-way ANOVA without interaction on page 429. In that analysis, the results clearly showed that the factor "route" had a significant effect on travel time. However, now that we include the interaction effect, the results show that conclusion is not generally true. As we review the *p*-values for the five one-way ANOVAs above (reject the null if the *p*-value is less than 0.05), we know that route mean travel times are different for three drivers: Filbeck, Snaverly, and Zollaco. However, for Deans and Ormson, their mean route travel times are not significantly different.

Now that we know this new and interesting information, we would want to know why these differences exist. Further investigation of the driving habits of the five drivers would be required.

In review, our presentation of two-way ANOVA with interaction shows the power of statistical analysis. In this analysis, we were able to test the combined effect of driver and route on travel time, and to show that different drivers evidently behave differently as they travel their routes. Gaining understanding of interaction effects is extremely important in many applications, from scientific areas such as agriculture and quality control to managerial fields like human resource management and gender equality in salary and performance ratings.

Self-Review 12–5 See the following ANOVA table.

ANOVA

Source of Variation	SS	*df*	MS	*F*	*p*-value
Factor A	6.41	3	2.137	3.46	0.0322
Factor B	5.01	2	2.507	1.06	0.0304
Interaction	33.15	6	5.525	8.94	0.0000
Error	14.83	24	0.618		
Total	59.41	59			

Use the .05 significance level to answer the following questions.

(a) How many levels does Factor A have? Is there a significant difference among the Factor A means? How do you know?

(b) How many levels does Factor B have? Is there a significant difference among the Factor B means? How do you know?

(c) How many observations are there in each cell? Is there a significant interaction between Factor A and Factor B on the response variable? How do you know?

Exercises

19. Consider the following sample data for a two-factor ANOVA experiment:

		Factor A		
		Level 1	Level 2	Level 3
Factor B	Level 1	23	20	11
		21	32	20
		25	26	20
	Level 2	13	20	11
		32	17	23
		17	15	8

Use the .05 significance level to answer the following questions.

a. Is there a difference in the Factor A means?

b. Is there a difference in the Factor B means?

c. Do Factors A and B have a significant interaction?

20. Consider the following partially completed two-way ANOVA table. Suppose there are four levels of Factor A and three levels of Factor B. The number of replications per cell is 5. Complete the table and test to determine if there is a significant difference in Factor A means, Factor B means, or the interaction means. Use the .05 significance level. (Hint: estimate the values from the *F* table.)

ANOVA

Source	SS	*df*	MS	*F*
Factor A	75			
Factor B	25			
Interaction	300			
Error	600			
Total	1000			

21. The distributor of the *Wapakoneta Daily News,* a regional newspaper serving southwestern Ohio, is considering three types of dispensing machines or "racks." Management wants to know if the different machines affect sales. These racks are designated as J-1000, D-320, and UV-57. Management also wants to know if the placement of the racks either inside or outside supermarkets affects sales. Each of six similar stores was randomly assigned a machine and location combination. The data below is the number of papers sold over four days.

Position/Machine	J-1000	D-320	UV-57
Inside	33, 40, 30, 31	29, 28, 33, 33	47, 39, 39, 45
Outside	43, 36, 41, 40	48, 45, 40, 44	37, 32, 36, 35

a. Draw the interaction graph. Based on your observations, is there an interaction effect? Based on the graph, describe the interaction effect of machine and position.
b. Use the 0.05 level to test for position, machine, and interaction effects on sales. Report the statistical results.
c. Compare the inside and outside mean sales for each machine using statistical techniques. What do you conclude?

22. A large company is organized into three functional areas: manufacturing, marketing, and research and development. The employees claim that the company pays women less than men for similar jobs. The company randomly selected four males and four females in each area and recorded their weekly salaries in dollars.

Area/Gender	Female	Male
Manufacturing	1016, 1007, 875, 968	978, 1056, 982, 748
Marketing	1045, 895, 848, 904	1154, 1091, 878, 876
Research and Development	770, 733, 844, 771	926, 1055, 1066, 1088

a. Draw the interaction graph. Based on your observations, is there an interaction effect? Based on the graph, describe the interaction effect of gender and area on salary.
b. Use the 0.05 level to test for gender, area, and interaction effects on salary. Report the statistical results.
c. Compare the male and female mean sales for each area using statistical techniques. What do you recommend to the distributor?

Chapter Summary

I. The characteristics of the F distribution are:
- **A.** It is continuous.
- **B.** Its values cannot be negative.
- **C.** It is positively skewed.
- **D.** There is a family of F distributions. Each time the degrees of freedom in either the numerator or the denominator changes, a new distribution is created.

II. The F distribution is used to test whether two population variances are the same.
- **A.** The sampled populations must follow the normal distribution.
- **B.** The larger of the two sample variances is placed in the numerator, forcing the ratio to be at least 1.00.
- **C.** The value of F is computed using the following equation:

$$F = \frac{s_1^2}{s_2^2} \quad \textbf{[12–1]}$$

III. A one-way ANOVA is used to compare several treatment means.
- **A.** A treatment is a source of variation.
- **B.** The assumptions underlying ANOVA are:
  - **1.** The samples are from populations that follow the normal distribution.
  - **2.** The populations have equal standard deviations.
  - **3.** The samples are independent.
- **C.** The information for finding the value of F is summarized in an ANOVA table.
  - **1.** The formula for SS total, the sum of squares total, is:

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

  - **2.** The formula for SSE, the sum of squares error, is:

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

  - **3.** The formula for the SST, the sum of squares treatment, is found by subtraction.

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

  - **4.** This information is summarized in the following table and the value of F determined.

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	$n - 1$		

IV. If a null hypothesis of equal treatment means is rejected, we can identify the pairs of means that differ from the following confidence interval.

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

V. In a two-way ANOVA we consider a second treatment variable.
- **A.** The second treatment variable is called the blocking variable.
- **B.** It is determined using the following equation:

$$\text{SSB} = k\Sigma(\overline{X}_b - \overline{X}_G)^2 \quad \textbf{[12–6]}$$

- **C.** The SSE term, or sum of squares error, is found from the following equation.

$$\text{SSE} = \text{SS total} - \text{SST} - \text{SSB} \quad \textbf{[12–7]}$$

D. The *F* statistics for the treatment variable and the blocking variable are determined in the following table.

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

VI. In a two-way ANOVA with repeated observations we consider two treatment variables and the possible interaction between the variables.

A. The sum of squares due to possible interactions is found by:

$$SSI = (k - 1)(b - 1)\Sigma\Sigma(\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_G)^2 \quad \textbf{[12–8]}$$

B. The SSE term is found by subtraction.

$$SSE = SS\ total - SSA - SSB - SSI \quad \textbf{[12–9]}$$

C. The complete ANOVA table including interactions is:

Source	Sum of Squares	df	Mean Square	F
Factor A	SSA	$k - 1$	$SSA/(k - 1) = MSA$	MSA/MSE
Factor B	SSB	$b - 1$	$SSB/(b - 1) = MSB$	MSB/MSE
Interaction	SSI	$(k - 1)(b - 1)$	$SSI/(k - 1)(b - 1) = MSI$	MSI/MSE
Error	SSE	$n - kb$	$SSE/(n - kb) = MSE$	
Total	SS total	$n - 1$		

Pronunciation Key

SYMBOL	MEANING	PRONUNCIATION
SS total	Sum of squares total	*S S total*
SST	Sum of squares treatment	*S S T*
SSE	Sum of squares error	*S S E*
MSE	Mean square error	*M S E*
SSB	Block sum of squares	*S S B*
SSI	Sum of squares interaction	*S S I*

Chapter Exercises

23. A real estate agent in the coastal area of Georgia wants to compare the variation in the selling price of homes on the oceanfront with those one to three blocks from the ocean. A sample of 21 oceanfront homes sold within the last year revealed the standard deviation of the selling prices was $45,600. A sample of 18 homes, also sold within the last year, that were one to three blocks from the ocean revealed that the standard deviation was $21,330. At the .01 significance level, can we conclude that there is more variation in the selling prices of the oceanfront homes?

24. A computer manufacturer is about to unveil a new, faster personal computer. The new machine clearly is faster, but initial tests indicate there is more variation in the processing time. The processing time depends on the particular program being run, the amount of input data, and the amount of output. A sample of 16 computer runs, covering a range of production jobs, showed that the standard deviation of the processing time was 22 (hundredths of a second) for the new machine and 12 (hundredths of a second) for the current machine. At the .05 significance level can we conclude that there is more variation in the processing time of the new machine?

25. There are two Chevrolet dealers in Jamestown, New York. The mean monthly sales at Sharkey Chevy and Dave White Chevrolet are about the same. However, Tom Sharkey, the owner of Sharkey Chevy, believes his sales are more consistent. Below is the number of new cars sold at Sharkey in the last seven months and for the last eight months at Dave White. Do you agree with Mr. Sharkey? Use the .01 significance level.

Sharkey	98	78	54	57	68	64	70	
Dave White	75	81	81	30	82	46	58	101

26. Random samples of five were selected from each of three populations. The sum of squares total was 100. The sum of squares due to the treatments was 40.
a. Set up the null hypothesis and the alternate hypothesis.
b. What is the decision rule? Use the .05 significance level.
c. Complete the ANOVA table. What is the value of *F*?
d. What is your decision regarding the null hypothesis?

27. In an ANOVA table MSE was equal to 10. Random samples of six were selected from each of four populations, where the sum of squares total was 250.
a. Set up the null hypothesis and the alternate hypothesis.
b. What is the decision rule? Use the .05 significance level.
c. Complete the ANOVA table. What is the value of *F*?
d. What is your decision regarding the null hypothesis?

28. The following is a partial ANOVA table.

Source	Sum of Squares	*df*	Mean Square	*F*
Treatment		2		
Error			20	
Total	500	11		

Complete the table and answer the following questions. Use the .05 significance level.
a. How many treatments are there?
b. What is the total sample size?
c. What is the critical value of *F*?
d. Write out the null and alternate hypotheses.
e. What is your conclusion regarding the null hypothesis?

29. A consumer organization wants to know whether there is a difference in the price of a particular toy at three different types of stores. The price of the toy was checked in a sample of five discount stores, five variety stores, and five department stores. The results are shown below. Use the .05 significance level.

Discount	Variety	Department
$12	$15	$19
13	17	17
14	14	16
12	18	20
15	17	19

30. A physician who specializes in weight control has three different diets she recommends. As an experiment, she randomly selected 15 patients and then assigned 5 to each diet.

After three weeks the following weight losses, in pounds, were noted. At the .05 significance level, can she conclude that there is a difference in the mean amount of weight loss among the three diets?

Plan A	Plan B	Plan C
5	6	7
7	7	8
4	7	9
5	5	8
4	6	9

31. The City of Maumee comprises four districts. Chief of Police Andy North wants to determine whether there is a difference in the mean number of crimes committed among the four districts. He recorded the number of crimes reported in each district for a sample of six days. At the .05 significance level, can the chief of police conclude there is a difference in the mean number of crimes?

Number of Crimes			
Rec Center	Key Street	Monclova	Whitehouse
13	21	12	16
15	13	14	17
14	18	15	18
15	19	13	15
14	18	12	20
15	19	15	18

32. A study of the effect of television commercials on 12-year-old children measured their attention span, in seconds. The commercials were for clothes, food, and toys. At the .05 significance level is there a difference in the mean attention span of the children for the various commercials? Are there significant differences between pairs of means? Would you recommend dropping one of the three commercial types?

Clothes	Food	Toys
26	45	60
21	48	51
43	43	43
35	53	54
28	47	63
31	42	53
17	34	48
31	43	58
20	57	47
	47	51
	44	51
	54	

33. When only two treatments are involved, ANOVA and the Student *t* test (Chapter 10) result in the same conclusions. Also, $t^2 = F$. As an example, suppose that 14 randomly selected students were divided into two groups, one consisting of 6 students and the other of 8. One group was taught using a combination of lecture and programmed instruction, the other using a combination of lecture and television. At the end of the course, each group was given a 50-item test. The following is a list of the number correct for each of the two groups.

Lecture and Programmed Instruction	Lecture and Television
19	32
17	28
23	31
22	26
17	23
16	24
	27
	25

a. Using analysis of variance techniques, test H_0 that the two mean test scores are equal; $\alpha = .05$.
b. Using the t test from Chapter 10, compute t.
c. Interpret the results.

34. There are four auto body shops in Bangor, Maine, and all claim to promptly serve customers. To check if there is any difference in service, customers are randomly selected from each repair shop and their waiting times in days are recorded. The output from a statistical software package is:

Summary

Groups	Count	Sum	Average	Variance
Body Shop A	3	15.4	5.133333	0.323333
Body Shop B	4	32	8	1.433333
Body Shop C	5	25.2	5.04	0.748
Body Shop D	4	25.9	6.475	0.595833

ANOVA

Source of Variation	SS	*df*	MS	*F*	*p*-value
Between Groups	23.37321	3	7.791069	9.612506	0.001632
Within Groups	9.726167	12	0.810514		
Total	33.09938	15			

Is there evidence to suggest a difference in the mean waiting times at the four body shops? Use the .05 significance level.

35. The fuel efficiencies for a sample of 27 compact, midsize, and large cars are entered into a statistical software package. Analysis of variance is used to investigate if there is a difference in the mean mileage of the three cars. What do you conclude? Use the .01 significance level.

Summary

Groups	Count	Sum	Average	Variance
Compact	12	268.3	22.35833	9.388106
Midsize	9	172.4	19.15556	7.315278
Large	6	100.5	16.75	7.303

Additional results are shown below.

ANOVA					
Source of Variation	**SS**	***df***	**MS**	***F***	***p*-value**
Between Groups	136.4803	2	68.24014	8.258752	0.001866
Within Groups	198.3064	24	8.262766		
Total	334.7867	26			

36. Three assembly lines are used to produce a certain component for an airliner. To examine the production rate, a random sample of six hourly periods is chosen for each assembly line and the number of components produced during these periods for each line is recorded. The output from a statistical software package is:

Summary				
Groups	**Count**	**Sum**	**Average**	**Variance**
Line A	6	250	41.66667	0.266667
Line B	6	260	43.33333	0.666667
Line C	6	249	41.5	0.7

ANOVA					
Source of Variation	**SS**	***df***	**MS**	***F***	***p*-value**
Between Groups	12.33333	2	6.166667	11.32653	0.001005
Within Groups	8.166667	15	0.544444		
Total	20.5	17			

a. Use a .01 level of significance to test if there is a difference in the mean production of the three assembly lines.
b. Develop a 99 percent confidence interval for the difference in the means between Line B and Line C.

37. A grocery store wants to monitor the amount of withdrawals that its customers make from automatic teller machines (ATMs) located within its stores. It samples 10 withdrawals from each location and the output from a statistical software package is:

Summary				
Groups	**Count**	**Sum**	**Average**	**Variance**
Location X	10	825	82.5	1,808.056
Location Y	10	540	54	921.1111
Location Z	10	382	38.2	1,703.733

ANOVA					
Source of Variation	**SS**	***df***	**MS**	***F***	***p*-value**
Between Groups	1,0081.27	2	5,040.633	3.411288	0.047766
Within Groups	3,9896.1	27	1,477.633		
Total	4,9977.37	29			

a. Use a .01 level of significance to test if there is a difference in the mean amount of money withdrawn.
b. Develop a 90 percent confidence interval for the difference in the means between Location X and Location Z.

38. One reads that a business school graduate with an undergraduate degree earns more than a high school graduate with no additional education, and a person with a master's degree or a doctorate earns even more. To investigate we select a sample of

25 mid-level managers of companies in southeast rural communities. Their incomes, classified by highest level of education, follow.

Income ($ thousands)		
High School or Less	Undergraduate Degree	Master's Degree or More
75	79	81
77	87	103
83	115	112
92	103	89
69	111	124
73	114	119
84	119	119
	122	125
	92	103

Test at the .05 level of significance that there is no difference in the arithmetic mean salaries of the three groups. If the null hypothesis is rejected, conduct further tests to determine which groups differ.

39. Shank's, Inc., a nationwide advertising firm, wants to know whether the size of an advertisement and the color of the advertisement make a difference in the response of magazine readers. A random sample of readers is shown ads of four different colors and three different sizes. Each reader is asked to give the particular combination of size and color a rating between 1 and 10. Assume that the ratings follow the normal distribution. The rating for each combination is shown in the following table (for example, the rating for a small red ad is 2).

	Color of Ad			
Size of Ad	Red	Blue	Orange	Green
Small	2	3	3	8
Medium	3	5	6	7
Large	6	7	8	8

Is there a difference in the effectiveness of an advertisement by color and by size? Use the .05 level of significance.

40. There are four McBurger restaurants in the Columbus, Georgia, area. The numbers of burgers sold at the respective restaurants for each of the last six weeks are shown below. At the .05 significance level, is there a difference in the mean number sold among the four restaurants, when the factor of week is considered?

	Restaurant			
Week	Metro	Interstate	University	River
1	124	160	320	190
2	234	220	340	230
3	430	290	290	240
4	105	245	310	170
5	240	205	280	180
6	310	260	270	205

a. Is there a difference in the treatment means?
b. Is there a difference in the block means?

41. The city of Tucson, Arizona, employs people to assess the value of homes for the purpose of establishing real estate tax. The city manager sends each assessor to the same five homes and then compares the results. The information is given below, in thousands of dollars. Can we conclude that there is a difference in the assessors, at $\alpha = .05$?

	Assessor			
Home	Zawodny	Norman	Cingle	Holiday
A	$53.0	$55.0	$49.0	$45.0
B	50.0	51.0	52.0	53.0
C	48.0	52.0	47.0	53.0
D	70.0	68.0	65.0	64.0
E	84.0	89.0	92.0	86.0

a. Is there a difference in the treatment means?
b. Is there a difference in the block means?

42. Martin Motors has in stock three cars of the same make and model. The president would like to compare the gas consumption of the three cars (labeled car A, car B, and car C) using four different types of gasoline. For each trial, a gallon of gasoline was added to an empty tank, and the car was driven until it ran out of gas. The following table shows the number of miles driven in each trial.

	Distance (miles)		
Types of Gasoline	Car A	Car B	Car C
Regular	22.4	20.8	21.5
Super regular	17.0	19.4	20.7
Unleaded	19.2	20.2	21.2
Premium unleaded	20.3	18.6	20.4

Using the .05 level of significance:
a. Is there a difference among types of gasoline?
b. Is there a difference in the cars?

43. A research firm wants to compare the miles per gallon of unleaded regular, mid-grade, and super premium gasolines. Because of differences in the performance of different automobiles, seven different automobiles were selected and treated as blocks. Therefore, each brand of gasoline was tested with each type of automobile. The results of the trials, in miles per gallon, are shown in the following table. At the .05 significance level, is there a difference in the gasolines or automobiles?

Automobile	Regular	Mid-grade	Super Premium
1	21	23	26
2	23	22	25
3	24	25	27
4	24	24	26
5	26	26	30
6	26	24	27
7	28	27	32

44. Three supermarket chains in the Denver area each claim to have the lowest overall prices. As part of an investigative study on supermarket advertising, the *Denver Daily News* conducted a study. First, a random sample of nine grocery items was selected. Next, the price of each selected item was checked at each of the three chains on the same day.

At the .05 significance level, is there a difference in the mean prices at the supermarkets or for the items?

Item	Super$	Ralph's	Lowblaws
1	$1.12	$1.02	$1.07
2	1.14	1.10	1.21
3	1.72	1.97	2.08
4	2.22	2.09	2.32
5	2.40	2.10	2.30
6	4.04	4.32	4.15
7	5.05	4.95	5.05
8	4.68	4.13	4.67
9	5.52	5.46	5.86

45. Listed below are the weights (in grams) of a sample of M&M's Plain candies, classified according to color. Use a statistical software system to determine whether there is a difference in the mean weights of candies of different colors. Use the .05 significance level.

Red	Orange	Yellow	Brown	Tan	Green
0.946	0.902	0.929	0.896	0.845	0.935
1.107	0.943	0.960	0.888	0.909	0.903
0.913	0.916	0.938	0.906	0.873	0.865
0.904	0.910	0.933	0.941	0.902	0.822
0.926	0.903	0.932	0.838	0.956	0.871
0.926	0.901	0.899	0.892	0.959	0.905
1.006	0.919	0.907	0.905	0.916	0.905
0.914	0.901	0.906	0.824	0.822	0.852
0.922	0.930	0.930	0.908		0.965
1.052	0.883	0.952	0.833		0.898
0.903		0.939			
0.895		0.940			
		0.882			
		0.906			

46. There are four radio stations in Midland. The stations have different formats (hard rock, classical, country/western, and easy listening), but each is concerned with the number of minutes of music played per hour. From a sample of 10 hours from each station, the following sample means were offered.

$$\bar{X}_1 = 51.32 \qquad \bar{X}_2 = 44.64 \qquad \bar{X}_3 = 47.2 \qquad \bar{X}_4 = 50.85$$

$$\text{SS total} = 650.75$$

a. Determine SST.
b. Determine SSE.
c. Complete an ANOVA table.
d. At the .05 significance level, is there a difference in the treatment means?
e. Is there a difference in the mean amount of music time between station 1 and station 4? Use the .05 significance level.

We recommend you complete the following exercises using a statistical software package such as Excel, MegaStat, or MINITAB.

47. The American Accounting Association recently conducted a study to compare the weekly wages of men and women employed in either the public or private sector of accounting.

	Sector	
Gender	**Public**	**Private**
Men	$ 978	$1,335
	1,035	1,167
	964	1,236
	996	1,317
	1,117	1,192
Women	$ 863	$1,079
	975	1,160
	999	1,063
	1,019	1,110
	1,037	1,093

At the .05 significance level:

a. Draw an interaction plot of men and women means by sector.
b. Test the interaction effect of gender and sector on wages.
c. Based on your results in part (b), conduct the appropriate tests of hypotheses for differences in factor means.
d. Interpret the results in a brief report.

48. Robert Altoff is vice president for engineering for a manufacturer of household washing machines. As part of new product development, he wishes to determine the optimal length of time for the washing cycle. A part of the development is to study the relationship between the detergent used (four brands) and the length of the washing cycle (18, 20, 22, or 24 minutes). In order to run the experiment 32 standard household laundry loads (having equal amounts of dirt and the same total weights) are randomly assigned to the 16 detergent–washing cycle combinations. The results (in pounds of dirt removed) are shown below.

	Cycle Time (min)			
Detergent Brand	**18**	**20**	**22**	**24**
A	0.13	0.12	0.19	0.15
	0.11	0.11	0.17	0.18
B	0.14	0.15	0.18	0.20
	0.10	0.14	0.17	0.18
C	0.16	0.15	0.18	0.19
	0.17	0.14	0.19	0.21
D	0.09	0.12	0.16	0.15
	0.13	0.13	0.16	0.17

At the .05 significance level:

a. Draw an interaction plot of the detergent means by cycle time.
b. Test the interaction effect of brand and cycle time on "dirt removed."
c. Based on your results in part (b), conduct the appropriate tests of hypotheses for differences in factor means.
d. Interpret the results in a brief report.

exercises.com

49. Many real estate companies and rental agencies now publish their listings on the Web. One example is Dunes Realty Company, located in Garden City Beach, South Carolina. Go to its website, http://www.dunes.com, select **Vacation Rentals,** then **Beach Home w/Pool Search,** then indicate *5* bedrooms, accommodations for *14* people, *second row* (this means it is across the street from the beach), no amenities. Select a period in July and August, indicate that you are willing to spend *$8,000* per week, and then click on

Search the Beach Homes w/Pools. The output should include details on the beach houses that met your criteria. At the .05 significance level, is there a difference in the mean rental prices for the different number of bedrooms? (You may want to combine some of the larger homes, such as 8 or more bedrooms.) Which pairs of means differ?

50. The percentages of quarterly changes in the Gross Domestic Product for 20 countries are available at the following site: http://www.oecd.org. Select **Statistics, National Accounts, Quarterly National Accounts** and select **Quarterly Growth Rates of GDP at Constant Price** for OECD Countries. Copy the data for Germany, Japan, and the United States into three columns in MINITAB or Excel. Perform an ANOVA test to see whether there is a difference in the means. What can you conclude?

Data Set Exercises

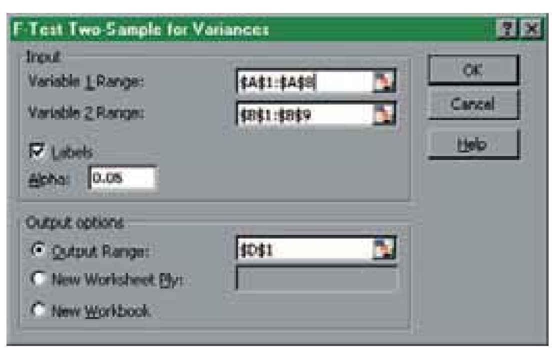
51. Refer to the Real Estate data, which report information on the homes sold in the Denver, Colorado, area last year.
- **a.** At the .02 significance level, is there a difference in the variability of the selling prices of the homes that have a pool versus those that do not have a pool?
- **b.** At the .02 significance level, is there a difference in the variability of the selling prices of the homes with an attached garage versus those that do not have an attached garage?
- **c.** At the .05 significance level, is there a difference in the mean selling price of the homes among the five townships?

52. Refer to the Baseball 2005 data, which report information on the 30 Major League Baseball teams for the 2005 season.
- **a.** At the .10 significance level, is there a difference in the variation in team salary among the American and National league teams?
- **b.** Create a variable that classifies a team's total attendance into three groups: less than 2.0 (million), 2.0 up to 3.0, and 3.0 or more. At the .05 significance level, is there a difference in the mean number of games won among the three groups? Use the .01 significance level.
- **c.** Using the same attendance variable developed in part (b), is there a difference in the mean team batting average? Use the .01 significance level.
- **d.** Using the same attendance variable developed in part (b), is there a difference in the mean salary of the three groups? Use the .01 significance level.

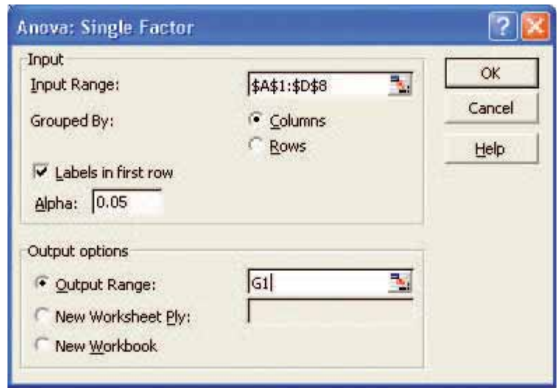
53. Refer to the Wage data, which report information on annual wages for a sample of 100 workers. Also included are variables relating to industry, years of education, and gender for each worker.
- **a.** Conduct a test of hypothesis to determine if there is a difference in the mean annual wages for workers in the three industries. If there is a difference in the means, which pair or pairs of means differ? Use the .05 significance level.
- **b.** Conduct a test of hypothesis to determine if there is a difference in the mean annual wages for workers in the six different occupations. If there is a difference in the means, which pair or pairs of means differ? Use the .05 significance level.

Software Commands

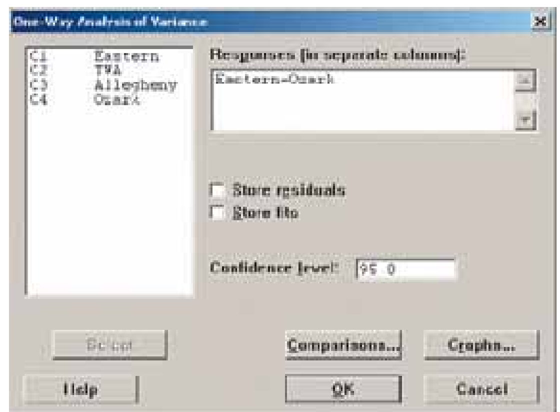
1. The Excel commands for the test of variances on page 411 are:
- **a.** Enter the data for U.S. 25 in column A and for I-75 in column B. Label the two columns.
- **b.** Click on **Tools, Data Analysis,** select **F-Test Two-Sample for Variances,** and click **OK.**
- **c.** The range of the first variable is *A1:A8* and *B1:B9* for the second. Click on **Labels,** enter *0.05* for **Alpha,** select *D1* for the **Output Range,** and click **OK.**



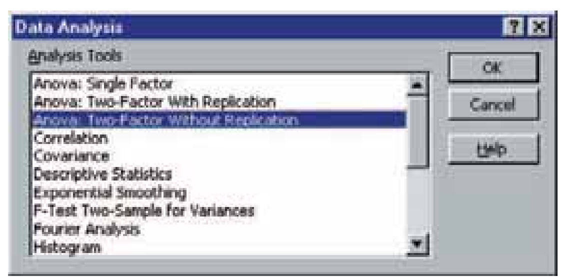
2. The Excel commands for the one-way ANOVA on page 420 are:
   a. Key in data into four columns labeled: *Eastern, TWA, Allegheny,* and *Ozark*.
   b. Click on **Tools** on the Excel Toolbar and select **Data Analysis.** In the dialog box select **ANOVA: Single Factor,** and click **OK.**
   c. In the subsequent dialog box make the input range *A1:D8,* click on **Grouped by Columns,** click on **Labels in first row,** the **Alpha** text box is *.05,* and finally select **Output Range** as *G1* and click **OK.**



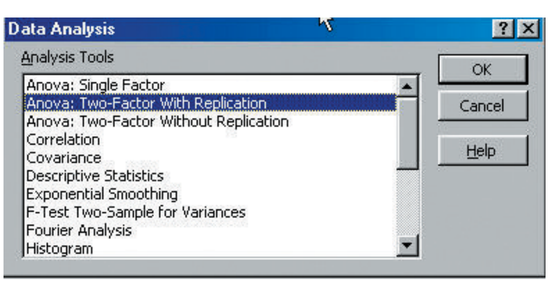
3. The MINITAB commands for the one-way ANOVA on page 420 are:
   a. Input the data into four columns and identify the columns as *Eastern, TWA, Allegheny,* and *Ozark*.
   b. Select **Stat, ANOVA,** and **One-way (Unstacked),** select the data in columns C1 to C4, check on **Select** in the lower left, and then click **OK.**



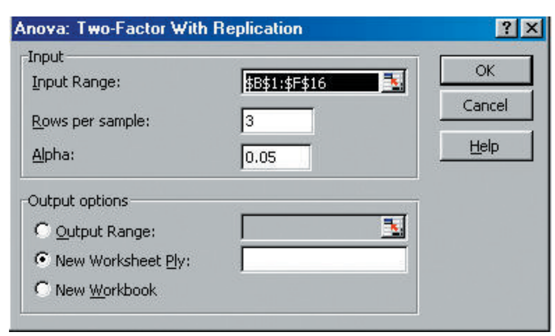
4. The Excel commands for the two-way ANOVA on page 430 are:
   a. In the first row of the first column write the word *Driver,* then list the five drivers in the first column. In the first row of the next four columns enter the name of the routes. Enter the data under each route name.
   b. Select **Tools, Data Analysis,** and **ANOVA: Two-Factor Without Replication,** and then click **OK.**
   c. In the dialog box the **Input Range** is *A4:E9,* click on **Labels,** select *G2* for **Output Range,** and then click **OK.**



5. The Excel commands for the two-way ANOVA with interaction on page 435 are:
   a. Enter the data into Excel as shown on page 433.
   b. Select **Tools, Data Analysis,** and **ANOVA: Two-Factor with Replication,** and then click **OK.**



   c. In the dialog box, enter the **Input Range** as *B2:F16,* enter **Rows per sample** as *3,* select **New Worksheet Ply,** and then click **OK.**



Chapter 12 Answers to Self-Review

12–1 Let Mark's assemblies be population 1, then H_0: $\sigma_1^2 \leq \sigma_2^2$; H_1: $\sigma_1^2 > \sigma_2^2$; $df_1 = 10 - 1 = 9$; and df_2 also equals 9. H_0 is rejected if $F > 3.18$.

$$F = \frac{(2.0)^2}{(1.5)^2} = 1.78$$

H_0 is not rejected. The variation is the same for both employees.

12–2 **a.** H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: At least one treatment mean is different.

b. Reject H_0 if $F > 4.26$

c. $$\bar{X} = \frac{240}{12} = 20$$

$$\text{SS total} = (18 - 20)^2 + \cdots + (32 - 20)^2 = 578$$

$$\text{SSE} = (18 - 17)^2 + (14 - 17)^2 + \cdots + (32 - 29)^2 = 74$$

$$\text{SST} = 578 - 74 = 504$$

d.

Source	Sum of Squares	Degrees of Freedom	Mean Square	*F*
Treatment	504	2	252	30.65
Error	74	9	8.22	
Total	578	11		

e. H_0 is rejected. There is difference in the mean number of bottles sold at the various locations.

12–3 **a.** H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: Not all means are equal.

b. H_0 is rejected if $F > 3.98$.

c. $\bar{X}_G = 8.86$, $\bar{X}_1 = 11$, $\bar{X}_2 = 8.75$, $\bar{X}_3 = 6.8$
SS total = 53.71
SST = 44.16
SSE = 9.55

Source	Sum of Squares	*df*	Mean Square	*F*
Treatment	44.16	2	22.08	25.43
Error	9.55	11	0.8682	
Total	53.71	13		

d. H_0 is rejected. The treatment means differ.

e. $(11.0 - 6.8) \pm 2.201\sqrt{0.8682\left(\frac{1}{5} + \frac{1}{5}\right)} = 4.2 \pm 1.30 = 2.90$ and 5.50
These treatment means differ because both endpoints of the confidence interval are of the same sign—positive in this problem.

12–4 For types:

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: The treatment means are not equal.

Reject H_0 if $F > 4.46$.
For months:

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
H_1: The block means are not equal.

Reject H_0 if $F > 3.84$.
The analysis of variance table is as follows:

Source:	*df*	SS	MS	*F*
Types	2	3.60	1.80	0.39
Months	4	31.73	7.93	1.71
Error	8	37.07	4.63	
Total	14	72.40		

The null hypotheses cannot be rejected for either types or months. There is no difference in the mean sales among types or months.

12–5 **a.** There are four levels of Factor A. The *p*-value is less than .05, so Factor A means differ.

b. There are three levels of Factor B. The *p*-value is less than .05, so the Factor B means differ.

c. There are three observations in each cell. There is an interaction between Factor A and Factor B means, because the *p*-value is less than .05.

A Review of Chapters 10–12

This section is a review of the major concepts and terms introduced in Chapters 10, 11, and 12. Chapter 10 began our study of hypothesis testing. A hypothesis is a statement about the value of a population parameter. In statistical hypothesis testing, we begin by making a statement about the value of the population parameter in the null hypothesis. We establish the null hypothesis for the purpose of testing. When we complete the testing, our decision is either to reject or to fail to reject the null hypothesis. If we reject the null hypothesis, we conclude that the alternate hypothesis is true. The alternate hypothesis is "accepted" only if we show that the null hypothesis is false. We also refer to the alternate hypothesis as the research hypothesis. Most of the time we want to prove the alternate hypothesis.

In Chapter 10 we selected random samples from a single population and tested whether it was reasonable that the population parameter under study equaled a particular value. For example, we wish to investigate whether the mean tenure time of those holding the position of CEO in large firms is 12 years. We select a sample of CEOs, compute the sample mean, and compare the mean of the sample to the population. The single population under consideration is the length of tenure of CEOs of large firms. We described methods for conducting the test when the population standard deviation was available and when it was not available. Also, in Chapter 10 we conducted tests of hypothesis about a population proportion. A proportion is the fraction of individuals or objects possessing a certain characteristic. For example, industry records indicate that 70 percent of gasoline sales for automobiles are for the regular grade of gasoline. A sample of the 100 sales from last month at the Pantry in Conway revealed 76 were for the regular grade. Can the owners conclude that more than 70 percent of their customers purchase the regular grade?

In Chapter 11 we extended the idea of hypothesis testing to whether two independent random samples came from populations having the same or equal population means. For example, St. Mathews Hospital operates an urgent care facility on both the north and south sides of Knoxville, Tennessee. The research question is: Is the mean waiting time for patients visiting the two facilities the same? To investigate, we select a random sample from each of the facilities and compute the sample means. We test the null hypothesis that the mean waiting time is the same at the two facilities. The alternate hypothesis is that the mean waiting time is not the same for the two facilities. If the population standard deviations are known, we use the z distribution as the distribution of the test statistic. If the population standard deviations are not known, the test statistic follows the t distribution.

Our discussion in Chapter 11 also concerned dependent samples. For *dependent* samples, we applied the *paired difference test*. The test statistic is the t distribution. One typical paired sample problem calls for recording an individual's blood pressure before administering medication and then again afterward in order to evaluate the effectiveness of the medication. We also considered the case of testing two population proportions. For example, the production manager wished to compare the proportion of defects on the day shift with that of the second shift.

Chapter 11 dealt with the difference between two population means. Chapter 12 presented tests for variances and a procedure called the *analysis of variance,* or *ANOVA*. ANOVA is used to simultaneously determine whether several independent normal populations have the same mean. This is accomplished by comparing the variances of the random samples selected from these populations. We apply the usual hypothesis-testing procedure, but we use the F distribution as the test statistic. Often the calculations are tedious, so a software package is recommended.

As an example of analysis of variance, a test could be conducted to resolve whether there is any difference in effectiveness among five fertilizers on the weight of popcorn ears. This type of analysis is referred to as *one-factor ANOVA* because we are able to draw conclusions about only one factor, called a *treatment*. If we want to draw conclusions about the simultaneous effects of more than one factor or variable, we use the *two-factor ANOVA* technique. Both the one-factor and two-factor tests use the *F distribution* as the distribution of the test statistic. The F distribution is also the distribution of the test statistic used to find whether one normal population has more variation than another.

Two-factor analysis of variance is further complicated by the possibility that interactions may exist between the factors. There is an *interaction* if the response to one of the factors depends on the level of the other factor. Fortunately, the ANOVA technique is easily extended to include a test for interactions.

Glossary

Chapter 10

Alpha The probability of a Type I error or the level of significance. Its symbol is the Greek letter α.

Alternate hypothesis The conclusion we accept when we demonstrate that the null hypothesis is false. It is also called the research hypothesis.

Critical value A value that is the dividing point between the region where the null hypothesis is not rejected and the region where it is rejected.

Degrees of freedom The number of items in a sample that are free to vary. Suppose there are two items in a sample, and we know the mean. We are free to specify only one of the two values, because the other value is automatically determined (since the two values total twice the mean). Example: If the mean is \$6, we are free to choose only one value. Choosing \$4 makes the other value \$8 because \$4 + \$8 = 2(\$6). So there is 1 degree of freedom in this illustration. We can determine the degrees of freedom by $n - 1 = 2 - 1 = 1$. If n is 4, then there are 3 degrees of freedom, found by $n - 1 = 4 - 1 = 3$.

Hypothesis A statement or claim about the value of a population parameter. Examples: 40.7 percent of all persons 65 years old or older live alone. The mean number of people in a car is 1.33.

Hypothesis testing A statistical procedure, based on sample evidence and probability theory, used to determine whether the statement about the population parameter is a reasonable statement.

Null hypothesis A statement about the value of a population parameter, H_0 that is put up for testing in the face of numerical evidence.

One-tailed test Used when the alternate hypothesis states a direction, such as H_1: $\mu > 40$, read "the population mean is greater than 40." Here the rejection region is only in one tail (the right tail).

Proportion A fraction or percentage of a sample or a population having a particular trait. If 5 out of 50 in a sample liked a new cereal, the proportion is 5/50, or .10.

***p*-value** The probability of computing a value of the test statistic at least as extreme as the one found in the sample data when the null hypothesis is true.

Significance level The probability of rejecting the null hypothesis when it is true.

Two-tailed test Used when the alternate hypothesis does not state a direction, such as H_1: $\mu \neq 75$, read "the population mean is not equal to 75." There is a region of rejection in each tail.

Type I error Occurs when a true H_0 is rejected.

Type II error Occurs when a false H_0 is accepted.

Chapter 11

Dependent samples Dependent samples are characterized by a measurement, then some type of intervention, followed by another measurement. Paired samples are also dependent because the same individual or item is a member of both samples. Example: Ten participants in a marathon were weighed prior to and after competing in the race. We wish to study the mean amount of weight loss.

Independent samples The samples chosen at random are not related to each other. We wish to study the mean age of the inmates at the Auburn and Allegheny prisons. We select a random sample of 28 inmates from the Auburn prison and a sample of 19 inmates at the Allegheny prison. A person cannot be an inmate in both prisons. The samples are independent, that is, unrelated.

Pooled estimate of the population variance A weighted average of s_1^2 and s_2^2 used to estimate the common variance, σ^2, when using small samples to test the difference between two population means.

***t* distribution** Investigated and reported by William S. Gossett in 1908 and published under the pseudonym *Student*. It is similar to the standard normal distribution presented in Chapter 7. The major characteristics of t are:

1. It is a continuous distribution.
2. It can assume values between minus infinity and plus infinity.
3. It is symmetrical about its mean of zero. However, it is more spread out and flatter at the apex than the standard normal distribution.
4. It approaches the standard normal distribution as n gets larger.
5. There is a family of t distributions. One t distribution exists for a sample of 15 observations, another for 25, and so on.

Chapter 12

Analysis of variance (ANOVA) A technique used to test simultaneously whether the means of several populations are equal. It uses the F distribution as the distribution of the test statistic.

Block A second source of variation, in addition to treatments.

***F* distribution** It is used as the test statistic for ANOVA problems, as well as others. The major characteristics of the F distribution are:

1. It is never negative.
2. It is a continuous distribution approaching the X-axis but never touching it.
3. It is positively skewed.
4. It is based on two sets of degrees of freedom.
5. Like the t distribution, there is a family of F distributions. There is one distribution for 17 degrees of freedom in the numerator and 9 degrees of freedom in the denominator, there is another F distribution for 7 degrees of freedom in the numerator and 12 degrees of freedom in the denominator, and so on.

Interaction Two variables interact if the effect that one factor has on the variable being studied is different for different levels of the other factor.

Exercises

Part I—Multiple Choice

1. In a one-tailed test using the z distribution as the test statistic and the .01 significance level, the critical value is either

a. −1.96 or +1.96.
b. −1.65 or +1.65.
c. −2.58 or +2.58.

   d. 0 or 1.
   e. None of these is correct.
2. A Type II error is committed if we:
   a. Reject a true null hypothesis.
   b. Accept a true alternate hypothesis.
   c. Reject a true alternate hypothesis.
   d. Accept both the null and alternate hypotheses at the same time.
   e. None of these is correct.
3. The hypotheses are H_0: $\mu = 240$ pounds of pressure and H_1: $\mu \neq 240$ pounds of pressure.
   a. A one-tailed test is being applied.
   b. A two-tailed test is being applied.
   c. A three-tailed test is being applied.
   d. The wrong test is being applied.
   e. None of these is correct.
4. The .01 significance level is used in a one-tailed test of hypothesis with the rejection region in the lower tail. The computed value of z is -1.8. This indicates:
   a. H_0 should not be rejected.
   b. We should reject H_0 and accept H_1.
   c. We should take a larger sample.
   d. We should have used the .05 level of significance.
   e. None of these is correct.
5. The test statistic for testing a hypothesis for sample means when the population standard deviation is not known is:
   a. z.
   b. t.
   c. F.
   d. χ^2.
6. We want to test a hypothesis for the difference between two population means. The null and alternate hypotheses are stated as

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

   a. A left-tailed test should be applied.
   b. A two-tailed test should be applied.
   c. A right-tailed test should be applied.
   d. We cannot determine whether a left-, right-, or two-tailed test should be applied based on the information given.
   e. None of these is correct.
7. The F distribution:
   a. Cannot be negative.
   b. Is negatively skewed.
   c. Is the same as the t distribution.
   d. Is the same as the z distribution.
   e. None of these is correct.
8. As the sample size increases, the t distribution approaches:
   a. ANOVA.
   b. The standard normal or z distribution.
   c. The Poisson distribution.
   d. Zero.
   e. None of these is correct.
9. To conduct a paired difference test, the samples must be:
   a. Infinitely large.
   b. Equal to ANOVA.
   c. Independent.
   d. Dependent.
   e. None of these is correct.
10. An ANOVA test was conducted with respect to the population mean. The null hypothesis was rejected. This indicates:
   a. There were too many degrees of freedom.
   b. There is no difference between the population means.
   c. There is a difference between at least two population means.
   d. A larger sample should be selected.
   e. None of these is correct.

Part II—Problems

For problems 11–16, state: (a) the null and the alternate hypothesis, (b) the decision rule, and (c) the decision regarding the null hypothesis, (d) then interpret the result.

11. A machine is set to produce tennis balls so the mean bounce is 36 inches when the ball is dropped from a platform of a certain height. The supervisor suspects that the mean bounce has changed and is less than 36 inches. As an experiment 42 balls were dropped from the platform and the mean height of the bounce was 35.5 inches, with a standard deviation of 0.9 inches. At the .05 significance level, can the supervisor conclude that the mean bounce height is less than 36 inches?

12. Research by First Bank of Illinois revealed that 8 percent of its customers wait more than five minutes to do their banking when not using the drive-through facility. Management considers this reasonable and will not add more tellers unless the proportion becomes larger than 8 percent. The branch manager at the Litchfield Branch believes that the wait is longer than the standard at her branch and requested additional part-time tellers. To support her request she found that, in a sample of 100 customers, 10 waited more than five minutes. At the .01 significance level, is it reasonable to conclude that more than 8 percent of the customers wait more than five minutes?

13. It was hypothesized that road construction workers do not engage in productive work 20 minutes on the average out of every hour. Some claimed the nonproductive time is greater than 20 minutes. An actual study was conducted at a construction site, using a stopwatch and other ways of checking the work habits. A random check of workers revealed the following unproductive times, in minutes, during a one-hour period (exclusive of regularly scheduled breaks):

10	25	17	20	28	30	18	23	18

Using the .05 significance level, is it reasonable to conclude the mean unproductive time is greater than 20 minutes?

14. A test is to be conducted involving the mean holding power of two glues designed for plastic. First, a small plastic hook was coated at one end with Epox glue and fastened to a sheet of plastic. After it dried, weight was added to the hook until it separated from the sheet of plastic. The weight was then recorded. This was repeated until 12 hooks were tested. The same procedure was followed for Holdtite glue, but only 10 hooks were used. The sample results, in pounds, were:

	Epox	Holdtite
Sample mean	250	252
Sample standard deviation	5	8
Sample size	12	10

At the .01 significance level, is there a difference between the mean holding power of Epox and that of Holdtite?

15. Pittsburgh Paints wishes to test an additive formulated to increase the life of paints used in the hot and arid conditions of the Southwest. The top half of a piece of wood was painted using the regular paint. The bottom half was painted with the paint including the additive. The same procedure was followed for a total of 10 pieces. Then each piece was subjected to brilliant light. The data, the number of hours each piece lasted before it faded beyond a certain point, follow:

	Number of Hours by Sample									
	A	B	C	D	E	F	G	H	I	J
Without additive	325	313	320	340	318	312	319	330	333	319
With additive	323	313	326	343	310	320	313	340	330	315

Using the .05 significance level, determine whether the additive is effective in prolonging the life of the paint.

16. A Buffalo, New York, cola distributor is featuring a super-special sale on 12-packs. She wonders where in the grocery store to place the cola for maximum attention. Should it be near the front door of the grocery stores, in the cola section, at the checkout registers, or near the milk and other dairy products? Four stores with similar total sales cooperated in an experiment.

In one store the 12-packs were stacked near the front door, in another they were placed near the checkout registers, and so on. Sales were checked at specified times in each store for exactly four minutes. The results were:

Cola at the Door	In Soft Drink Section	Near Registers	Dairy Section
$6	$ 5	$ 7	$10
8	10	10	9
3	12	9	6
7	4	4	11
	9	5	
		7	

The Buffalo distributor wants to find out whether there is a difference in the mean sales for cola stacked at the four locations in the store. Use the .05 significance level.

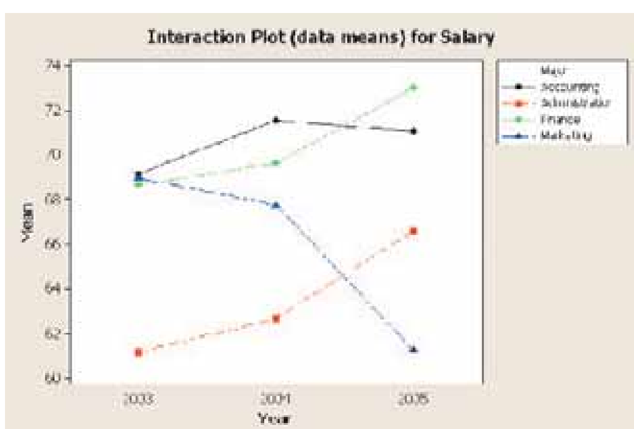
17. Williams Corporation is investigating the effects of educational background on employee performance. A potential relevant variable in this case is the self-rated social status of the employee. The company has recorded the annual sales volumes (in $000) achieved by sales employees in each of the categories below. Perform a complete two-way analysis of variance (including the possibility of interactions) on the data and describe what your results suggest.

	School Type		
Self-rated Social Status	**Ivy League**	**State-supported**	**Small Private**
Low	62, 61	68, 64	70, 70
Medium	68, 64	74, 68	62, 65
High	70, 71	57, 60	57, 56

18. A school supervisor is reviewing initial wages of former students (in $000). Samples were taken over three years for four different majors (accounting, administration, finance, and marketing).

Major/Year	2003	2004	2005
Accounting	75.4, 69.8, 62.3	73.9, 78.8, 62.0	64.2, 80.8, 68.2
Administration	61.5, 59.9, 62.1	63.9, 57.6, 66.5	74.2, 67.5, 58.1
Finance	63.6, 70.2, 72.2	69.2, 72.5, 67.2	74.7, 66.4, 77.9
Marketing	71.3, 69.2, 66.4	74.0, 67.6, 61.7	60.0, 61.3, 62.5

a. Here is an interaction plot of the information. What does it reveal?



b. Write out all of the pairs of null and alternative hypotheses you would apply for a two-way ANOVA.
c. Here is the statistical software output. Use the 0.05 level to check for interactions.

```
Source          DF       SS        MS       F        P
Major            3    329.20   109.732    3.39    0.034
Year             2      7.32     3.659    0.11    0.894
Interaction      6    183.57    30.595    0.94    0.482
Error           24    777.29    32.387
Total           35   1297.37
```

d. If proper, test the other hypotheses at the 0.05 significance level. If it is not appropriate, describe why you should not do the tests.

Cases

A. Century National Bank

Refer to the description of Century National Bank at the end of the Review of Chapters 1–4 on page 136.

With many other options available, customers no longer let their money sit in a checking account. For many years the mean checking balance has been $1,600. Does the sample data indicate that the mean account balance has declined from this value?

Recent years have also seen an increase in the use of ATM machines. When Mr. Selig took over the bank, the mean number of transactions per month per customer was 8; now he believes it has increased to more than 10. In fact, the advertising agency that prepares TV commercials for Century would like to use this on the new commercial being designed. Is there sufficient evidence to conclude that the mean number of transactions per customer is more than 10 per month? Could the advertising agency say the mean is more than 9 per month?

The bank has branch offices in four different cities: Cincinnati, Ohio; Atlanta, Georgia; Louisville, Kentucky; and Erie, Pennsylvania. Mr. Selig would like to know whether there is a difference in the mean checking account balances among the four branches. If there are differences, between which branches do these differences occur?

Mr. Selig is also interested in the bank's ATMs. Is there a difference in ATM use among the branches? Also, do customers who have debit cards tend to use ATMs differently from those who do not have debit cards? Is there a difference in ATM use by those with checking accounts that pay interest versus those that do not? Prepare a report for Mr. Selig answering these questions.

B. Bell Grove Medical Center

Ms. Gene Dempsey manages the emergency care center at Bell Grove Medical Center. One of her responsibilities is to have enough nurses so that incoming patients needing service can be handled promptly. It is stressful for patients to wait a long time for emergency care even when their care needs are not life threatening. Ms. Dempsey gathered the following information regarding the number of patients over the last several weeks. The center is not open on weekends. Does it appear that there are any differences in the number of patients served by the day of the week? If there are differences, which days seem to be the busiest?

Date	Day	Patients
9-29-06	Monday	38
9-30-06	Tuesday	28
10-1-06	Wednesday	28
10-2-06	Thursday	30
10-3-06	Friday	35
10-6-06	Monday	35
10-7-06	Tuesday	25
10-8-06	Wednesday	22
10-9-06	Thursday	21
10-10-06	Friday	32
10-13-06	Monday	37
10-14-06	Tuesday	29
10-15-06	Wednesday	27
10-16-06	Thursday	28
10-17-06	Friday	35
10-20-06	Monday	37
10-21-06	Tuesday	26
10-22-06	Wednesday	28
10-23-06	Thursday	23
10-24-06	Friday	33