10

John Bryson/Time Life Pictures/Getty Images

"So what!"

—ANONYMOUS

Norman Rockwell

We have all heard the exclamation, "So what!" Philologists (people who study cultural linguistics) tell us that this expression is a shortened version of "So what is the difference!" They also tell us that there are similar popular or slang expressions about differences in all languages and cultures. Norman Rockwell (1894–1978) painted everyday people and situations. In this cover, "Girl with Black Eye," for the *Saturday Evening Post* (May 23, 1953), a young lady is about to have a conference with her school principal. So what! It is human nature to challenge the claim that something is better, worse, or just simply different. In this chapter, we will focus on this very human theme by studying a variety of topics regarding questions of whether or not differences exist between two population variances or among several population means.

For online student resources, visit the Brase/Brase, *Understandable Statistics,* 10th edition web site at **http://www.cengage.com/statistics/brase**.

CHI-SQUARE AND *F* DISTRIBUTIONS

How do you decide if random variables are dependent or independent? (SECTION 10.1)

How do you decide if different populations share the same proportions of specified characteristics? (SECTION 10.1)

How do you decide if two distributions are not only dependent, but actually the same distribution? (SECTION 10.2)

How do you compute confidence intervals and tests for σ? (SECTION 10.3)

How do you test two variances σ_1^2 and σ_2^2? (SECTION 10.4)

What is one-way ANOVA? Where is it used? (SECTION 10.5)

What about two-way ANOVA? Where is it used? (SECTION 10.6)

The Image Works

Mesa Verde National Park

FOCUS PROBLEM

Archaeology in Bandelier National Monument

Archaeologists at Washington State University did an extensive summer excavation at Burnt Mesa Pueblo in Bandelier National Monument. Their work is published in the book *Bandelier Archaeological Excavation Project: Summer 1990 Excavations at Burnt Mesa Pueblo and Casa del Rito,* edited by T. A. Kohler.

One question the archaeologists asked was: Is raw material used by prehistoric Indians for stone tool manufacture independent of the archaeological excavation site? Two different excavation sites at Burnt Mesa Pueblo gave the information in the table below.

Use a chi-square test with 5% level of significance to test the claim that raw material used for construction of stone tools and excavation site are independent. (See Problem 17 of Section 10.1.)

Stone Tool Construction Material, Burnt Mesa Pueblo

Material	Site A	Site B	Row Total
Basalt	731	584	1315
Obsidian	102	93	195
Pedernal chert	510	525	1035
Other	85	94	179
Column Total	**1428**	**1296**	**2724**

Robert Brenner/PhotoEdit

Archaeological excavation site

591

PART I: INFERENCES USING THE CHI-SQUARE DISTRIBUTION

Overview of the Chi-Square Distribution

So far, we have used several probability distributions for hypothesis testing and confidence intervals, with the most frequently used being the normal distribution and the Student's *t* distribution. In this chapter, we will use two other probability distributions, namely, the chi-square distribution (where *chi* is pronounced like the first two letters in the word *kite*) and the *F* distribution. In Part I, we will see applications of the chi-square distribution, whereas in Part II, we will see some important applications of the *F* distribution.

Chi is a Greek letter denoted by the symbol χ, so chi-square is denoted by the symbol χ^2. Because the distribution is of chi-*square* values, the χ^2 values begin at 0 and then are all positive. The graph of the χ^2 distribution is not symmetrical, and like the Student's *t* distribution, it depends on the number of degrees of freedom. Figure 10-1 shows χ^2 distributions for several degrees of freedom (*d.f.*).

As the degrees of freedom increase, the graph of the chi-square distribution becomes more bell-like and begins to look more and more symmetric.

> The **mode (high point)** of a chi-square distribution with *n* degrees of freedom occurs over $n - 2$ (for $n \geq 3$).

Table 7 of Appendix II shows critical values of chi-square distributions for which a designated area falls to the *right* of the critical value. Table 10-1 gives an excerpt from Table 7. Notice that the row headers are degrees of freedom, and the column headers are areas in the *right* tail of the distribution. For instance, according to the table, for a χ^2 distribution with 3 degrees of freedom, the area occurring to the *right* of $\chi^2 = 0.072$ is 0.995. For a χ^2 distribution with 4 degrees of freedom, the area falling to the *right* of $\chi^2 = 13.28$ is 0.010.

In the next three sections, we will see how to apply the chi-square distribution to different applications.

FIGURE 10-1

The χ^2 Distribution
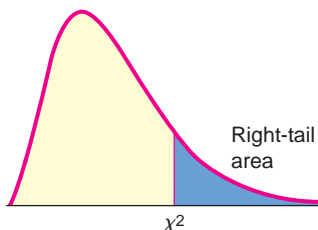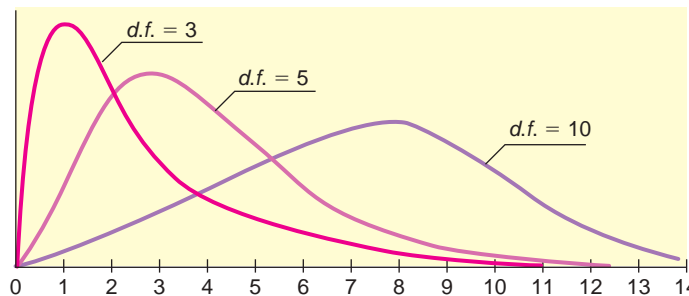
TABLE 10-1   Excerpt from Table 7 (Appendix II): The χ^2 Distribution

	Area of the Right Tail					
d.f.	0.995	0.990	0.975	. . .	0.010	0.005
⋮	⋮	⋮	⋮		⋮	⋮
3	0.072	0.115	0.216		11.34	12.84
4	0.207	0.297	0.484		13.28	14.86

SECTION 10.1

Chi-Square: Tests of Independence and of Homogeneity

FOCUS POINTS

- Set up a test to investigate independence of random variables.
- Use contingency tables to compute the sample χ^2 statistic.
- Find or estimate the *P*-value of the sample χ^2 statistic and complete the test.
- Conduct a test of homogeneity of populations.

Innovative Machines Incorporated has developed two new letter arrangements for computer keyboards. The company wishes to see if there is any relationship between the arrangement of letters on the keyboard and the number of hours it takes a new typing student to learn to type at 20 words per minute. Or, from another point of view, is the time it takes a student to learn to type *independent* of the arrangement of the letters on a keyboard?

To answer questions of this type, we test the hypotheses

Hypotheses

H_0: Keyboard arrangement and learning times *are independent*.
H_1: Keyboard arrangement and learning times *are not independent*.

Test of independence

Chi-square distribution

In problems of this sort, we are testing the *independence* of two factors. The probability distribution we use to make the decision is the *chi-square distribution*. Recall from the overview of the chi-square distribution that *chi* is pronounced like the first two letters of the word *kite* and is a Greek letter denoted by the symbol χ. Thus, chi-square is denoted by χ^2.

Innovative Machines' first task is to gather data. Suppose the company took a random sample of 300 beginning typing students and randomly assigned them to learn to type on one of three keyboards. The learning times for this sample are shown in Table 10-2. These learning times are the observed frequencies O.

Observed frequency O
Contingency table

Table 10-2 is called a *contingency table*. The *shaded boxes* that contain observed frequencies are called *cells*. The row and column totals are not considered to be cells. This contingency table is of size 3×3 (read "three-by-three") because there are three rows of cells and three columns. When giving the size of a contingency table, we always list the number of *rows first*.

> To determine the **size** of a contingency table, count the number of rows containing data and the number of columns containing data. The size is
>
> Number of rows $\times$ Number of columns
>
> where the symbol "$\times$" is read "by." The number of rows is always given first.

TABLE 10-2   **Keyboard versus Time to Learn to Type at 20 wpm**

Keyboard	21–40 h	41–60 h	61–80 h	Row Total
A	#1 25	#2 30	#3 25	80
B	#4 30	#5 71	#6 19	120
Standard	#7 35	#8 49	#9 16	100
Column Total	90	150	60	300
				Sample size

GUIDED EXERCISE 1 | **Size of contingency table**

Give the sizes of the contingency tables in Figures 10-2(a) and (b). Also, count the number of cells in each table. (Remember, each pink shaded box is a cell.)

(a) FIGURE 10-2(a) Contingency Table

				Row total
Column total				

(b) FIGURE 10-2(b) Contingency Table

			Row total
Column total			

(a) ⟹ There are two rows and four columns, so this is a 2 × 4 table. There are eight cells.

(b) ⟹ Here we have three rows and two columns, so this is a 3 × 2 table with six cells.

We are testing the null hypothesis that the keyboard arrangement and the time it takes a student to learn to type are *independent*. We use this hypothesis to determine the *expected frequency* of each cell.

Expected frequency E

For instance, to compute the expected frequency of cell 1 in Table 10-2, we observe that cell 1 consists of all the students in the sample who learned to type on keyboard A and who mastered the skill at the 20-words-per-minute level in 21 to 40 hours. By the assumption (null hypothesis) that the two events are independent, we use the multiplication law to obtain the probability that a student is in cell 1.

$$P(\text{cell 1}) = P(\text{keyboard A } and \text{ skill in } 21-40 \text{ h})$$
$$= P(\text{keyboard A}) \cdot P(\text{skill in } 21-40 \text{ h})$$

Because there are 300 students in the sample and 80 used keyboard A,

$$P(\text{keyboard A}) = \frac{80}{300}$$

Also, 90 of the 300 students learned to type in 21–40 hours, so

$$P(\text{skill in } 21-40 \text{ h}) = \frac{90}{300}$$

Using these two probabilities and the assumption of independence,

$$P(\text{keyboard A } and \text{ skill in } 21-40 \text{ h}) = \frac{80}{300} \cdot \frac{90}{300}$$

Finally, because there are 300 students in the sample, we have the *expected frequency E* for cell 1.

$$E = P(\text{student in cell 1}) \cdot (\text{no. of students in sample})$$
$$= \frac{80}{300} \cdot \frac{90}{300} \cdot 300 = \frac{80 \cdot 90}{300} = 24$$

We can repeat this process for each cell. However, the last step yields an easier formula for the expected frequency *E*.

<div style="background: #fdf3cf; padding: 1em;">

Formula for expected frequency E

$$E = \frac{(\text{Row total})(\text{Column total})}{\text{Sample size}}$$

</div>

Row total
Column total

Note: If the expected value is not a whole number, do *not* round it to the nearest whole number.

Let's use this formula in Example 1 to find the expected frequency for cell 2.

EXAMPLE 1   EXPECTED FREQUENCY

Find the expected frequency for cell 2 of contingency Table 10-2.

SOLUTION: Cell 2 is in row 1 and column 2. The *row total* is 80, and the *column total* is 150. The size of the sample is still 300.

$$E = \frac{(\text{Row total})(\text{Column total})}{\text{Sample size}}$$

$$= \frac{(80)(150)}{300} = 40$$

GUIDED EXERCISE 2   **Expected frequency**

Table 10-3 contains the *observed frequencies O* and *expected frequencies E* for the contingency table giving keyboard arrangement and number of hours it takes a student to learn to type at 20 words per minute. Fill in the missing expected frequencies.

TABLE 10-3   Complete Contingency Table of Keyboard Arrangement and Time to Learn to Type

Keyboard	21–40 h	41–60 h	61–80 h	Row Total
A	#1 $O = 25$ $E = 24$	#2 $O = 30$ $E = 40$	#3 $O = 25$ $E = \_\_$	80
B	#4 $O = 30$ $E = 36$	#5 $O = 71$ $E = \_\_$	#6 $O = 19$ $E = \_\_$	120
Standard	#7 $O = 35$ $E = \_\_$	#8 $O = 49$ $E = 50$	#9 $O = 16$ $E = 20$	100
Column Total	90	150	60	300 Sample Size

For cell 3, we have

$$E = \frac{(80)(60)}{300} = 16$$

For cell 5, we have

$$E = \frac{(120)(150)}{300} = 60$$

For cell 6, we have

$$E = \frac{(120)(60)}{300} = 24$$

For cell 7, we have

$$E = \frac{(100)(90)}{300} = 30$$

Computing the sample test statistic χ^2

Now we are ready to compute the sample statistic χ^2 for the typing students. The χ^2 value is a measure of the sum of the differences between *observed frequency O* and *expected frequency E* in each cell. These differences are listed in Table 10-4.

As you can see, if we sum the differences between the observed frequencies and the expected frequencies of the cells, we get the value zero. This total certainly does not reflect the fact that there were differences between the observed and expected frequencies. To obtain a measure whose sum does reflect the magnitude of the differences, we square the differences and work with the quantities $(O - E)^2$. But instead of using the terms $(O - E)^2$, we use the values $(O - E)^2/E$.

TABLE 10-4	Differences Between Observed and Expected Frequencies		
Cell	Observed O	Expected E	Difference $(O - E)$
1	25	24	1
2	30	40	−10
3	25	16	9
4	30	36	−6
5	71	60	11
6	19	24	−5
7	35	30	5
8	49	50	−1
9	16	20	−4
			$\Sigma(O - E) - 0$

We use this expression because a small difference between the observed and expected frequencies is not nearly as important when the expected frequency is large as it is when the expected frequency is small. For instance, for both cells 1 and 8, the squared difference $(O - E)^2$ is 1. However, this difference is more meaningful in cell 1, where the expected frequency is 24, than it is in cell 8, where the expected frequency is 50. When we divide the quantity $(O - E)^2$ by E, we take the size of the difference with respect to the size of the expected value. We

Sample test statistic χ^2   use the sum of these values to form the sample statistic χ^2:

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

where the sum is over all cells in the contingency table.

COMMENT   If you look up the word *irony* in a dictionary, you will find one of its meanings to be "the difference between actual (or observed) results and expected results." Because irony is so prevalent in much of our human experience, it is not surprising that statisticians have incorporated a related chi-square distribution into their work.

GUIDED EXERCISE 3 | *Sample χ^2*

(a) Complete Table 10-5.

➡   The last two rows of Table 10-5 are

TABLE 10-5   Data of Table 10-4

Cell	O	E	$O - E$	$(O - E)^2$	$(O - E)^2/E$
1	25	24	1	1	0.04
2	30	40	−10	100	2.50
3	25	16	9	81	5.06
4	30	36	−6	36	1.00
5	71	60	11	121	2.02
6	19	24	−5	25	1.04
7	35	30	5	25	0.83
8	49	50	____	____	____
9	16	20	____	____	____

$$\Sigma \frac{(O - E)^2}{E} = \text{____}$$

Cell	O	E	$O - E$	$(O - E)^2$	$(O - E)^2/E$
8	49	50	−1	1	0.02
9	16	20	−4	16	0.80

$$\Sigma \frac{(O - E)^2}{E} = \text{total of last column} = 13.31$$

(b) Compute the statistic χ^2 for this sample.

➡   Since $\chi^2 = \Sigma \dfrac{(O - E)^2}{E}$, then $\chi^2 = 13.31$.

Notice that when the observed frequency and the expected frequency are very close, the quantity $(O - E)^2$ is close to zero, and so the statistic χ^2 is near zero. As the difference increases, the statistic χ^2 also increases. To determine how large the sample statistic can be before we must reject the null hypothesis of independence, we find the *P*-value of the statistic in the chi-square distribution, Table 7 of Appendix II, and compare it to the specified level of significance α. The *P*-value depends on the number of degrees of freedom. To test independence, the degrees of freedom *d.f.* are determined by the following formula.

Degrees of freedom for test of independence

Degrees of freedom for test of independence

Degrees of freedom = (Number of rows − 1) · (Number of columns − 1)

or $d.f. = (R - 1)(C - 1)$

where R = number of cell rows

C = number of cell columns

GUIDED EXERCISE 4   **Degrees of freedom**

Determine the number of degrees of freedom in the example of keyboard arrangements (see Table 10-2). Recall that the contingency table had three rows and three columns.

$\Rightarrow$ $d.f. = (R - 1)(C - 1)$
$= (3 - 1)(3 - 1) = (2)(2) = 4$

**Finding the *P*-value for tests of independence**

To test the hypothesis that the letter arrangement on a keyboard and the time it takes to learn to type at 20 words per minute are independent at the $\alpha = 0.05$ level of significance, we estimate the *P*-value shown in Figure 10-3 below for the sample test statistic $\chi^2 = 13.31$ (calculated in Guided Exercise 3). We then compare the *P*-value to the specified level of significance α.

For tests of independence, we always use a *right-tailed* test on the chi-square distribution. This is because we are testing to see if the χ^2 measure of the difference between the observed and expected frequencies is too large to be due to chance alone.

In Guided Exercise 4, we found that the degrees of freedom for the example of keyboard arrangements is 4. From Table 7 of Appendix II, in the row headed by $d.f. = 4$, we see that the sample $\chi^2 = 13.31$ falls between the entries 13.28 and 14.86.
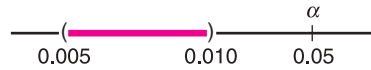
FIGURE 10-3

P-value

P-value
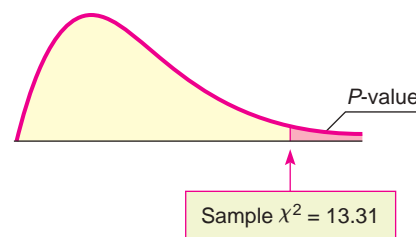
Sample $\chi^2 = 13.31$

Right-tail Area	0.010	0.005
$d.f. = 4$	13.28	14.86
	↑ Sample $\chi^2 = 13.31$	

The corresponding *P*-value falls between 0.005 and 0.010. From technology, we get *P*-value ≈ 0.0098.

Since the *P*-value is less than the level of significance $\alpha = 0.05$, we reject the null hypothesis of independence and conclude that keyboard arrangement and learning time are *not* independent.

Tests of independence for two statistical variables involve a number of steps. A summary of the procedure follows.

PROCEDURE

HOW TO TEST FOR INDEPENDENCE OF TWO STATISTICAL VARIABLES

Setup

Construct a contingency table in which the rows represent one statistical variable and the columns represent the other. Obtain a random sample of observations, which are assigned to the cells described by the rows and columns. These assignments are called the **observed values O** from the sample.

Procedure

1. Set the level of significance α and use the hypotheses

    H_0: The variables are independent.

    H_1: The variables are not independent.

2. For each cell, compute the **expected frequency E** (do not round but give as a decimal number).

$$E = \frac{(\text{Row total})(\text{Column total})}{\text{Sample size}}$$

Requirement

You need a sample size large enough so that, for each cell, $E \geq 5$.

Now each cell has two numbers, the observed frequency O from the sample and the expected frequency E.

Next, compute the sample *chi-square test statistic*

$$\chi^2 = \Sigma \frac{(O - E)^2}{E} \quad \text{with degrees of freedom } d.f. = (R - 1)(C - 1)$$

where the sum is over all cells in the contingency table and

    R = number of rows in contingency table

    C = number of columns in contingency table

3. Use the chi-square distribution (Table 7 of Appendix II) and a *right-tailed test* to find (or estimate) the *P-value* corresponding to the test statistic.

4. *Conclude* the test. If *P*-value $\leq \alpha$, then reject H_0. If *P*-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

GUIDED EXERCISE 5 | **Testing independence**

Super Vending Machines Company is to install soda pop machines in elementary schools and high schools. The market analysts wish to know if flavor preference and school level are independent. A random sample of 200 students was taken. Their school level and soda pop preferences are given in Table 10-6. Is independence indicated at the $\alpha = 0.01$ level of significance?

STEP 1: State the null and alternate hypotheses.

STEP 2:

(a) Complete the contingency Table 10-6 by filling in the required expected frequencies.

H_0: School level and soda pop preference are independent.

H_1: School level and soda pop preference are not independent.

The expected frequency

for cell 5 is $\dfrac{(40)(80)}{200} = 16$

for cell 6 is $\dfrac{(40)(120)}{200} = 24$

for cell 7 is $\dfrac{(20)(80)}{200} = 8$

for cell 8 is $\dfrac{(20)(120)}{200} = 12$

Note: In this example, the expected frequencies are all whole numbers. If the expected frequency has a decimal part, such as 8.45, do *not* round the value to the nearest whole number; rather, give the expected frequency as the decimal number.

Table 10-6 **School Level and Soda Pop Preference**

Soda Pop	High School	Elementary School	Row Total
Kula Kola	O = 33 #1 E = 36	O = 57 #2 E = 54	90
Mountain Mist	O = 30 #3 E = 20	O = 20 #4 E = 30	50
Jungle Grape	O = 5 #5 E = ____	O = 35 #6 E = ____	40
Diet Pop	O = 12 #7 E = ____	O = 8 #8 E = ____	20
Column Total	80	120	200 Sample Size

(b) Fill in Table 10-7 and use the table to find the sample statistic χ^2.

The last three rows of Table 10-7 should read as follows:

Table 10-7 **Computational Table for χ^2**

Cell	O	E	O − E	(O − E)²	(O − E)²/E
1	33	36	−3	9	0.25
2	57	54	3	9	0.17
3	30	20	10	100	5.00
4	20	30	−10	100	3.33
5	5	16	−11	121	7.56
6	35	24	11	____	____
7	12	8	____	____	____
8	8	12	____	____	____

Cell	O	E	O − E	(O − E)²	(O − E)²/E
6	35	24	11	121	5.04
7	12	8	4	16	2.00
8	8	12	−4	16	1.33

χ^2 = total of last column

$= \Sigma \dfrac{(O - E)^2}{E} = 24.68$

(c) What is the size of the contingency table? Use the number of rows and the number of columns to determine the degrees of freedom.

The contingency table is of size 4×2. Since there are four rows and two columns,

$d.f. = (4 - 1)(2 - 1) = 3$

Continued

GUIDED EXERCISE 5   *continued*

STEP 3: Use Table 7 of Appendix II to estimate the *P*-value of the sample statistic $\chi^2 = 24.68$ with *d.f.* $= 3$.
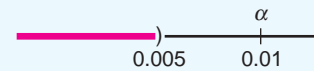
Right-tail Area	0.005
d.f. = 3	12.84
	↑
	Sample $\chi^2 = 24.68$

As the χ^2 values increase, the area to the right decreases, so

P-value < 0.005

STEP 4: Conclude the test by comparing the *P*-value of the sample statistic to the level of significance $\alpha = 0.01$.

Since the *P*-value is less than α, we reject the null hypothesis of independence. Technology gives *P*-value ≈ 0.00002.

STEP 5: *Interpret* the test result in the context of the application.

At the 1% level of significance, we conclude that school level and soda pop preference are dependent.

TECH NOTES   The TI-84Plus/TI-83Plus/TI-*n*spire calculators, Excel 2007, and Minitab all support chi-square tests of independence. In each case, the observed data are entered in the format of the contingency table.

TI-84Plus/TI-83Plus/TI-*n*spire (with TI-84Plus keypad) Enter the observed data into a matrix. Set the dimension of matrix [**B**] to match that of the matrix of observed values. Expected values will be placed in matrix [**B**]. Press **STAT, TESTS,** and select option **C:χ^2-Test.** The output gives the sample χ^2 with the *P*-value.

Excel 2007 Enter the table of observed values. Use the formulas of this section to compute the expected values. Enter the corresponding table of expected values. Finally, click **insert function** (*f*ₓ). Select **Statistical** for the category and **CHITEST** for the function. Excel returns the *P*-value of the sample χ^2 value.

Minitab Enter the contingency table of observed values. Use the menu selection **Stat ➤ Tables ➤ Chi-Square Test.** The output shows the contingency table with expected values and the sample χ^2 with *P*-value.

Tests of Homogeneity

We've seen how to use contingency tables and the chi-square distribution to test for independence of two random variables. The same process enables us to determine whether several populations share the same proportions of distinct categories. Such a test is called a *test of homogeneity*.

According to the dictionary, among the definitions of the word *homogeneous* are "of the same structure" and "composed of similar parts." In statistical jargon, this translates as a test of homogeneity to see if two or more populations share specified characteristics in the same proportions.

Test of homogeneity

> A **test of homogeneity** tests the claim that *different populations* share the *same proportions* of specified characteristics.

The computational processes for conducting tests of independence and tests of homogeneity are the same. However, there are two main differences in the initial setup of the two types of tests, namely, the sampling method and the hypotheses.

> Tests of independence compared to tests of homogeneity
>
> 1. **Sampling method**
>   For tests of independence, we use one random sample and observe how the sample members are distributed among distinct categories.
>   For tests of homogeneity, we take random samples from each different population and see how members of each population are distributed over distinct categories.
>
> 2. **Hypotheses**
>   For tests of independence,
>
> H_0: The variables are independent.
> H_1: The variables are not independent.
>
>   For tests of homogeneity,
>
> H_0: Each population shares respective characteristics in the same proportion.
> H_1: Some populations have different proportions of respective characteristics.

EXAMPLE 2 TEST OF HOMOGENEITY

Pets—who can resist a cute kitten or puppy? Tim is doing a research project involving pet preferences among students at his college. He took random samples of 300 female and 250 male students. Each sample member responded to the survey question "If you could own only one pet, what kind would you choose?" The possible responses were: "dog," "cat," "other pet," "no pet." The results of the study follow.

Pet Preference

Gender	Dog	Cat	Other Pet	No Pet
Female	120	132	18	30
Male	135	70	20	25

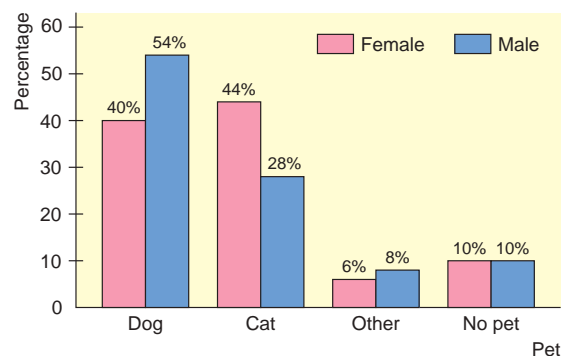Does the same proportion of males as females prefer each type of pet? Use a 1% level of significance.

We'll answer this question in several steps.

(a) First make a cluster bar graph showing the percentages of females and the percentages of males favoring each category of pet. From the graph, does it appear that the proportions are the same for males and females?

SOLUTION: The cluster graph shown in Figure 10-4 was created using Minitab. Looking at the graph, it appears that there are differences in the proportions of females and males preferring each type of pet. However, let's conduct a statistical test to verify our visual impression.

FIGURE 10-4

Pet Preference by Gender

Copyright 2010 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(b) Is it appropriate to use a test of homogeneity?

SOLUTION: Yes, since there are separate random samples for each designated population, male and female. We also are interested in whether each population shares the same proportion of members favoring each category of pet.

(c) State the hypotheses and conclude the test by using the Minitab printout.

H_0: The proportions of females and males naming each pet preference are the same.

H_1: The proportions of females and males naming each pet preference are not the same.

```
Chi-Square Test: Dog, Cat, Other, No Pet
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts
              Dog          Cat        Other       No Pet        Total
1             120          132           18           30          300
           139.09       110.18        20.73        30.00
            2.620        4.320        0.359        0.000

2             135           70           20           25          250
           115.91        91.82        17.27        25.00
            3.144        5.185        0.431        0.000

Total         255          202           38           55          550
Chi-Sq = 16.059, DF = 3, P-Value = 0.001
```

Since the P-value is less than α, we reject H_0 at the 1% level of significance.

(d) *Interpret* the results.

SOLUTION: It appears from the sample data that male and female students at Tim's college have different preferences when it comes to selecting a pet.

PROCEDURE

HOW TO TEST FOR HOMOGENEITY OF POPULATIONS

Setup

Obtain random samples from each of the populations. For each population, determine the number of members that share a distinct specified characteristic. Make a contingency table with the different populations as the rows (or columns) and the characteristics as the columns (or rows). The values recorded in the cells of the table are the **observed values O** taken from the samples.

Procedure

1. Set the level of significance and use the hypotheses

   H_0: The proportion of each population sharing specified characteristics is the same for all populations.

   H_1: The proportion of each population sharing specified characteristics is not the same for all populations.

2. Follow steps 2–5 of the procedure used to test for independence.

It is important to observe that when we reject the null hypothesis in a test of homogeneity, we don't know which proportions differ among the populations. We know only that the populations differ in some of the proportions sharing a characteristic.

Multinomial Experiments (Optional Reading)

Here are some observations that may be considered "brain teasers." In Chapters 6, 7, and 8, you studied normal approximations to binomial experiments. This concept resulted in some important statistical applications. Is it possible to extend this idea and obtain even more applications? Well, read on!

Consider a *binomial experiment* with n trials. The probability of success on each trial is p, and the probability of failure is $q = 1 - p$. If r is the number of successes out of n trials, then, from Chapter 5, you know that

$$P(r) = \frac{n!}{r!(n-r)!} \, p^r q^{n-r}$$

The binomial setting has just two outcomes: success or failure. What if you want to consider more than just two outcomes on each trial (for instance, the outcomes shown in a contingency table)? Well, you need a new statistical tool.

Multinomial experiments

Consider a *multinomial experiment*. This means that

1. The trials are independent and repeated under identical conditions.

2. The outcome on each trial falls into exactly one of $k \geq 2$ categories or cells.

3. The probability that the outcome of a single trial will fall into the ith category or cell is p_i (where $i = 1, 2, \ldots, k$) and remains the same for each trial. Furthermore, $p_1 + p_2 + \cdots + p_k = 1$.

4. Let r_i be a random variable that represents the number of trials in which the outcome falls into category or cell i. If you have n trials, then $r_1 + r_2 + \cdots + r_k = n$. The multinomial probability distribution is then

$$P(r_1, r_2, \ldots r_k) = \frac{n!}{r_1! r_2! \cdots r_k!} \, p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k}$$

How are the multinomial distribution and the binomial distribution related? For the special case $k = 2$, we use the notation $r_1 = r$, $r_2 = n - r$, $p_1 = p$, and $p_2 = q$. In this special case, the multinomial distribution becomes the binomial distribution.

There are two important tests regarding the cell probabilities of a multinomial distribution.

I. **Test of Independence** (Section 10.1)
   In this test, the null hypothesis of independence claims that each cell probability p_i will equal the product of its respective row and column probabilities. The alternate hypothesis claims that this is not so.

II. **Goodness-of-Fit Test** (Section 10.2)
    In this test, the null hypothesis claims that each category or cell probability p_i will equal a prespecified value. The alternate hypothesis claims that this is not so.

So why don't we use the multinomial probability distribution in Sections 10.1 and 10.2? The reason is that the exact calculation of probabilities associated with type I errors using the multinomial distribution is very tedious and cumbersome. Fortunately, the British statistician Karl Pearson discovered that the chi-square distribution can be used for this purpose, provided the expected value of each cell or category is at least 5.

It is Pearson's chi-square methods that are presented in Sections 10.1 and 10.2. In a sense, you have seen a similar application to statistical tests in Section 8.3, where you used the normal approximation to the binomial when np, the expected number of successes, and nq, the expected number of failures, were both at least 5.

Was there a time in the past when people worked for the same company all their lives, regularly purchased the same brand names, always voted for candidates from the same political party, and loyally cheered for the same sports team? One way to look at this question is to consider tests of statistical independence. Is customer loyalty independent of company profits? Can a company maintain its productivity independent of loyal workers? Can politicians do whatever they please independent of the voters back home? Americans may be ready to act on a pent-up desire to restore a sense of loyalty in their lives. For more information, see American Demographics*, Vol. 19, No. 9.*

SECTION 10.1 PROBLEMS

1. *Statistical Literacy* In general, are chi-square distributions symmetric or skewed? If skewed, are they skewed right or left?

2. *Statistical Literacy* For chi-square distributions, as the number of degrees of freedom increases, does any skewness increase or decrease? Do chi-square distributions become more symmetric (and normal) as the number of degrees of freedom becomes larger and larger?

3. *Statistical Literacy* For chi-square tests of independence and of homogeneity, do we use a right-tailed, left-tailed, or two-tailed test?

4. *Critical Thinking* In general, how do the hypotheses for chi-square tests of independence differ from those for chi-square tests of homogeneity? Explain.

5. *Critical Thinking* Zane is interested in the proportion of people who recycle each of three distinct products: paper, plastic, electronics. He wants to test the hypothesis that the proportion of people recycling each type of product differs by age group: 12–18 years old, 19–30 years old, 31–40 years old, over 40 years old. Describe the sampling method appropriate for a test of homogeneity regarding recycled products and age.

6. *Critical Thinking* Charlotte is doing a study on fraud and identity theft based both on source (checks, credit cards, debit cards, online banking/finance sites, other) and on gender of the victim. Describe the sampling method appropriate for a test of independence regarding source of fraud and gender.

7. *Interpretation: Test of Homogeneity* Consider Zane's study regarding products recycled and age group (see Problem 5). Suppose he found a sample $\chi^2 = 16.83$.
   (a) How many degrees of freedom are used? Recall that there were 4 age groups and 3 products specified. Approximate the *P*-value and conclude the test at the 1% level of significance. Does it appear that the proportion of people who recycle each of the specified products differ by age group? Explain.
   (b) From this study, can Zane identify how the different age groups differ regarding the proportion of those recycling the specified product? Explain.

8. *Interpretation: Test of Independence* Consider Charlotte's study of source of fraud/identity theft and gender (see Problem 6). She computed sample $\chi^2 = 10.2$.
   (a) How many degrees of freedom are used? Recall that there were 5 sources of fraud/identity theft and, of course, 2 genders. Approximate the *P*-value and conclude the test at the 5% level of significance. Would it seem that gender and source of fraud/identity theft are independent?
   (b) From this study, can Charlotte identify which source of fraud/identity theft is dependent with respect to gender? Explain.

For Problems 9–19, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Find the value of the chi-square statistic for the sample. Are all the expected frequencies greater than 5? What sampling distribution will you use? What are the degrees of freedom?

(c) Find or estimate the *P*-value of the sample test statistic.
(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis of independence?
(e) *Interpret* your conclusion in the context of the application.
Use the expected values *E* to the hundredths place.

9. *Psychology: Myers–Briggs* The following table shows the Myers–Briggs personality preferences for a random sample of 406 people in the listed professions (*Atlas of Type Tables* by Macdaid, McCaulley, and Kainz). E refers to extroverted and I refers to introverted.

Occupation	Personality Preference Type		Row Total
	E	I	
Clergy (all denominations)	62	45	107
M.D.	68	94	162
Lawyer	56	81	137
Column Total	186	220	406

Use the chi-square test to determine if the listed occupations and personality preferences are independent at the 0.05 level of significance.

10. *Psychology: Myers–Briggs* The following table shows the Myers–Briggs personality preferences for a random sample of 519 people in the listed professions (*Atlas of Type Tables* by Macdaid, McCaulley, and Kainz). T refers to thinking and F refers to feeling.

Occupation	Personality Preference Type		Row Total
	T	F	
Clergy (all denominations)	57	91	148
M.D.	77	82	159
Lawyer	118	94	212
Column Total	252	267	519

Use the chi-square test to determine if the listed occupations and personality preferences are independent at the 0.01 level of significance.

11. *Archaeology: Pottery* The following table shows site type and type of pottery for a random sample of 628 sherds at a location in Sand Canyon Archaeological Project, Colorado (*The Sand Canyon Archaeological Project,* edited by Lipe).

Site Type	Pottery Type			Row Total
	Mesa Verde Black-on-White	McElmo Black-on-White	Mancos Black-on-White	Black-on-White
Mesa Top	75	61	53	189
Cliff-Talus	81	70	62	213
Canyon Bench	92	68	66	226
Column Total	248	199	181	628

Use a chi-square test to determine if site type and pottery type are independent at the 0.01 level of significance.

12. *Archaeology: Pottery* The following table shows ceremonial ranking and type of pottery sherd for a random sample of 434 sherds at a location in the Sand Canyon Archaeological Project, Colorado (*The Architecture of Social Integration in Prehistoric Pueblos,* edited by Lipe and Hegmon).

Use a chi-square test to determine if ceremonial ranking and pottery type are independent at the 0.05 level of significance.

Ceremonial Ranking	Cooking Jar Sherds	Decorated Jar Sherds (Noncooking)	Row Total
A	86	49	135
B	92	53	145
C	79	75	154
Column Total	257	177	434

13. *Ecology: Buffalo* The following table shows age distribution and location of a random sample of 166 buffalo in Yellowstone National Park (based on information from *The Bison of Yellowstone National Park,* National Park Service Scientific Monograph Series).

Age	Lamar District	Nez Perce District	Firehole District	Row Total
Calf	13	13	15	41
Yearling	10	11	12	33
Adult	34	28	30	92
Column Total	57	52	57	166

Use a chi-square test to determine if age distribution and location are independent at the 0.05 level of significance.

14. *Psychology: Myers–Briggs* The following table shows the Myers–Briggs personality preference and area of study for a random sample of 519 college students (*Applications of the Myers–Briggs Type Indicator in Higher Education,* edited by Provost and Anchors). In the table, IN refers to introvert, intuitive; EN refers to extrovert, intuitive; IS refers to introvert, sensing; and ES refers to extrovert, sensing.

Myers–Briggs Preference	Arts & Science	Business	Allied Health	Row Total
IN	64	15	17	96
EN	82	42	30	154
IS	68	35	12	115
ES	75	42	37	154
Column Total	289	134	96	519

Use a chi-square test to determine if Myers–Briggs preference type is independent of area of study at the 0.05 level of significance.

15. *Sociology: Movie Preference* Mr. Acosta, a sociologist, is doing a study to see if there is a relationship between the age of a young adult (18 to 35 years old) and the type of movie preferred. A random sample of 93 young adults revealed the following data. Test whether age and type of movie preferred are independent at the 0.05 level.

Movie	Person's Age			Row Total
	18–23 yr	24–29 yr	30–35 yr	
Drama	8	15	11	34
Science fiction	12	10	8	30
Comedy	9	8	12	29
Column Total	29	33	31	93

16. *Sociology: Ethnic Groups* After a large fund drive to help the Boston City Library, the following information was obtained from a random sample of contributors to the library fund. Using a 1% level of significance, test the claim that the amount contributed to the library fund is independent of ethnic group.

Ethnic Group	Number of People Making Contribution					Row Total
	$1–50	$51–100	$101–150	$151–200	Over $200	
A	83	62	53	35	18	251
B	94	77	48	25	20	264
C	78	65	51	40	32	266
D	105	89	63	54	29	340
Column Total	360	293	215	154	99	1121

17. *Focus Problem: Archaeology* The Focus Problem at the beginning of the chapter refers to excavations at Burnt Mesa Pueblo in Bandelier National Monument. One question the archaeologists asked was: Is raw material used by prehistoric Indians for stone tool manufacture independent of the archaeological excavation site? Two different excavation sites at Burnt Mesa Pueblo gave the information in the following table. Use a chi-square test with 5% level of significance to test the claim that raw material used for construction of stone tools and excavation site are independent.

Material	Stone Tool Construction Material, Burnt Mesa Pueblo		
	Site A	Site B	Row Total
Basalt	731	584	1315
Obsidian	102	93	195
Pedernal chert	510	525	1035
Other	85	94	179
Column Total	1428	1296	2724

18. *Political Affiliation: Spending* Two random samples were drawn from members of the U.S. Congress. One sample was taken from members who are Democrats and the other from members who are Republicans. For each sample, the number of dollars spent on federal projects in each congressperson's home district was recorded.
    (i) Make a cluster bar graph showing the percentages of Congress members from each party who spent each designated amount in their respective home districts.
    (ii) Use a 1% level of significance to test whether congressional members of each political party spent designated amounts in the same proportions.

Party	Dollars Spent on Federal Projects in Home Districts			Row Total
	Less than 5 Billion	5 to 10 Billion	More than 10 Billion	
Democratic	8	15	22	45
Republican	12	19	16	47
Column Total	20	34	38	92

19. *Sociology: Methods of Communication* Random samples of people ages 15–24 and 25–34 were asked about their preferred method of (remote) communication with friends. The respondents were asked to select one of the methods from the following list: cell phone, instant message, e-mail, other.

(i) Make a cluster bar graph showing the percentages in each age group who selected each method.

(ii) Test whether the two populations share the same proportions of preferences for each type of communication method. Use $\alpha = 0.05$.

| | Preferred Communication Method | | | | |
Age	Cell Phone	Instant Message	E-mail	Other	Row Total
15–24	48	40	5	7	100
25–34	41	30	15	14	100
Column Total	89	70	20	21	200

SECTION 10.2

Chi-Square: Goodness of Fit

FOCUS POINTS

- Set up a test to investigate how well a sample distribution fits a given distribution.
- Use observed and expected frequencies to compute the sample χ^2 statistic.
- Find or estimate the *P*-value and complete the test.

Last year, the labor union bargaining agents listed five categories and asked each employee to mark the *one* most important to her or him. The categories and corresponding percentages of favorable responses are shown in Table 10-8. The bargaining agents need to determine if the *current* distribution of responses "fits" last year's distribution or if it is different.

In questions of this type, we are asking whether a population follows a specified distribution. In other words, we are testing the hypotheses

Hypotheses

H_0: The population fits the given distribution.

H_1: The population has a different distribution.

Computing sample χ^2

We use the chi-square distribution to test "goodness-of-fit" hypotheses. Just as with tests of independence, we compute the sample statistic:

$$\chi^2 = \Sigma \frac{(O - E)^2}{E} \text{ with degrees of freedom} = k - 1$$

where E = expected frequency

O = observed frequency

$\dfrac{(O - E)^2}{E}$ is summed for each category in the distribution

k = number of categories in the distribution

Next we use the chi-square distribution table (Table 7, Appendix II) to estimate the *P*-value of the sample χ^2 statistic. Finally, we compare the *P*-value to the level of significance α and conclude the test.

TABLE 10-8	Bargaining Categories (last year)

Category	Percentage of Favorable Responses
Vacation time	4%
Salary	65%
Safety regulations	13%
Health and retirement benefits	12%
Overtime policy and pay	6%

Goodness-of-fit test

In the case of a *goodness-of-fit test,* we use the null hypothesis to compute the expected values for the categories. Let's look at the bargaining category problem to see how this is done.

In the bargaining category problem, the two hypotheses are

H_0: The present distribution of responses is the same as last year's.
H_1: The present distribution of responses is different.

The null hypothesis tells us that the *expected frequencies* of the present response distribution should follow the percentages indicated in last year's survey. To test this hypothesis, a random sample of 500 employees was taken. If the null hypothesis is true, then there should be 4%, or 20 responses, out of the 500 rating vacation time as the most important bargaining issue. Table 10-9 gives the other expected values and all the information necessary to compute the sample statistic χ^2. We see that the sample statistic is

$$\chi^2 = \Sigma \frac{(O - E)^2}{E} = 14.15$$

Larger values of the sample statistic χ^2 indicate greater differences between the proposed distribution and the distribution followed by the sample. The larger the χ^2 statistic, the stronger the evidence to reject the null hypothesis that the population distribution fits the given distribution. Consequently, goodness-of-fit tests are always *right-tailed* tests.

Type of test

For *goodness-of-fit tests,* we use a *right-tailed* test on the chi-square distribution. This is because we are testing to see if the χ^2 measure of the difference between the observed and expected frequencies is too large to be due to chance alone.

To test the hypothesis that the present distribution of responses to bargaining categories is the same as last year's, we use the chi-square distribution (Table 7

TABLE 10-9	Observed and Expected Frequencies for Bargaining Categories

Category	O	E	$(O - E)^2$	$(O - E)^2/E$
Vacation time	30	4% of 500 = 20	100	5.00
Salary	290	65% of 500 = 325	1225	3.77
Safety	70	13% of 500 = 65	25	0.38
Health and retirement	70	12% of 500 = 60	100	1.67
Overtime	40	6% of 500 = 30	100	3.33
	$\Sigma O = 500$	$\Sigma E = 500$		$\Sigma \dfrac{(O - E)^2}{E} = 14.15$

Degrees of freedom

of Appendix II) to estimate the *P*-value of the sample statistic $\chi^2 = 14.15$. To estimate the *P*-value, we need to know the number of *degrees of freedom*. In the case of a goodness-of-fit test, the degrees of freedom are found by the following formula.

> **Degrees of freedom for goodness-of-fit test**
>
> $$d.f. = k - 1$$
>
> where k = number of categories

Notice that when we compute the expected values *E*, we must use the null hypothesis to compute all but the last one. To compute the last one, we can subtract the previous expected values from the sample size. For instance, for the bargaining issues, we could have found the number of responses for overtime policy by adding the other expected values and subtracting that sum from the sample size 500. We would again get an expected value of 30 responses. The degrees of freedom, then, is the number of *E* values that *must* be computed by using the null hypothesis.

For the bargaining issues, we have

$$d.f. = 5 - 1 = 4$$

where $k = 5$ is the number of categories.

P-value

We now have the tools necessary to use Table 7 of Appendix II to estimate the *P*-value of $\chi^2 = 14.15$. Figure 10-5 shows the *P*-value. In Table 7, we use the row headed by *d.f.* = 4. We see that $\chi^2 = 14.15$ falls between the entries 13.28 and 14.86. Therefore, the *P*-value falls between the corresponding right-tail areas 0.005 and 0.010. Technology gives the *P*-value ≈ 0.0068.

To test the hypothesis that the distribution of responses to bargaining issues is the same as last year's at the 1% level of significance, we compare the *P*-value of the statistic to $\alpha = 0.01$.


We see that the *P*-value is less than α, so we reject the null hypothesis that the distribution of responses to bargaining issues is the same as last year's.

Interpretation At the 1% level of significance, we can say that the evidence supports the conclusion that this year's responses to the issues are different from last year's.

Goodness-of-fit tests involve several steps that can be summarized as follows.

FIGURE 10-5

P-value
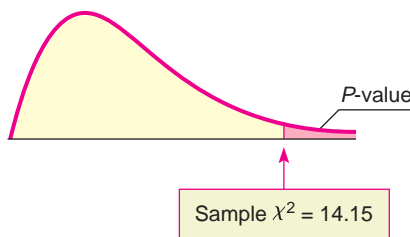
Right-tail area	0.010	0.005
d.f. = 4	13.28	14.86
		↑ Sample $\chi^2 = 14.15$

PROCEDURE	HOW TO TEST FOR GOODNESS OF FIT

Setup

First, each member of a population needs to be classified into exactly one of several different categories. Next, you need a specific (theoretical) distribution that assigns a fixed probability (or percentage) that a member of the population will fall into one of the categories. You then need a random sample size n from the population. Let O represent the *observed number* of data from the sample that fall into each category. Let E represent the *expected number* of data from the sample that, in theory, would fall into each category.

$O = $ observed frequency count of a category using sample data

$E = $ expected frequency of a category

$= $ (sample size n)(probability assigned to category)

Requirement

The sample size n should be large enough that $E \geq 5$ in each category.

Procedure

1. Set the *level of significance* α and use the *hypotheses*

   H_0: The population fits the specified distribution of categories.

   H_1: The population has a different distribution.

2. For each category, compute $(O - E)^2/E$, and then compute the *sample test statistic*

$$\chi^2 = \Sigma \frac{(O - E)^2}{E} \text{ with } d.f. = k - 1$$

   where the sum is taken over all categories and $k = $ number of categories.

3. Use the chi-square distribution (Table 7 of Appendix II) and a *right-tailed test* to find (or estimate) the *P-value* corresponding to the sample test statistic.

4. *Conclude* the test. If *P*-value $\leq \alpha$, then reject H_0. If *P*-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

One important application of goodness-of-fit tests is to genetics theory. Such an application is shown in Guided Exercise 6.

GUIDED EXERCISE 6 | **Goodness-of-fit test**

According to genetics theory, red-green colorblindness in humans is a recessive sex-linked characteristic. In this case, the gene is carried on the X chromosome only. We will denote an X chromosome with the colorblindness gene by X_c and one without the gene by X_n. Women have two X chromosomes, and they will be red-green colorblind only if both chromosomes have the gene, designated X_cX_c. A woman can have normal vision but still carry the colorblind gene if only one of the chromosomes has the gene, designated X_cX_n. A man carries an X and a Y chromosome; if the X chromosome carries the colorblind gene (X_cY), the man is colorblind.

According to genetics theory, if a man with normal vision (X_nY) and a woman carrier (X_cX_n) have a child, the probabilities that the child will have red-green colorblindness, will have normal vision and not carry the gene, or will have normal vision and carry the gene are given by the *equally likely* events in Table 10-10.

Continued

GUIDED EXERCISE 6   *continued*

$P(\text{child has normal vision and is not a carrier}) = P(X_nY) + P(X_nX_n) = \dfrac{1}{2}$

$P(\text{child has normal vision and is a carrier}) = P(X_cX_n) = \dfrac{1}{4}$

$P(\text{child is red-green colorblind}) = P(X_cY) = \dfrac{1}{4}$

Table 10-10 Red-Green Colorblindness

Mother	Father X_n	Y
X_c	X_cX_n	X_cY
X_n	X_nX_n	X_nY

To test this genetics theory, Genetics Labs took a random sample of 200 children whose mothers were carriers of the colorblind gene and whose fathers had normal vision. The results are shown in Table 10-11. We wish to test the hypothesis that the population follows the distribution predicted by the genetics theory (see Table 10-10). Use a 1% level of significance.

(a) State the null and alternate hypotheses. What is α?

⟹ H_0: The population fits the distribution predicted by genetics theory.

H_1: The population does not fit the distribution predicted by genetics theory.

$\alpha = 0.01$

(b) Fill in the rest of Table 10-11 and use the table to compute the sample statistic χ^2.

TABLE 10-11 Colorblindness Sample

Event	O	E	$(O - E)^2$	$(O - E)^2/E$
Red-green colorblind	35	50	225	4.50
Normal vision, noncarrier	105	___	___	___
Normal vision, carrier	60	___	___	___

⟹ **TABLE 10-12 Completion of Table 10-11**

Event	O	E	$(O - E)^2$	$(O - E)^2/E$
Red-green colorblind	35	50	225	4.50
Normal vision, noncarrier	105	100	25	0.25
Normal vision, carrier	60	50	100	2.00

The sample statistic is $\chi^2 = \Sigma \dfrac{(O - E)^2}{E} = 6.75$

(c) There are $k = 3$ categories listed in Table 10-11. Use this information to compute the degrees of freedom.

⟹ $d.f. = k - 1$
$= 3 - 1 = 2$

(d) Find the P-value for $\chi^2 = 6.75$.

⟹ Using Table 7 of Appendix II and the fact that goodness-of-fit tests are right-tailed tests, we see that
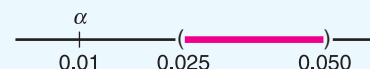
Right-tail area	0.050	0.025
$d.f. = 2$	5.99	7.38

Sample $\chi^2 = 6.75$

$0.025 < P\text{-value} < 0.050$
Technology gives P-value ≈ 0.0342.

(e) Conclude the test for $\alpha = 0.01$.

⟹ For $\alpha = 0.01$, we have

α
0.01   0.025   0.050

Since P-value $> \alpha$, do not reject H_0.

(f) *Interpret* the conclusion in the context of the application.

At the 1% level of significance, there is insufficient evidence to conclude that the population follows a distribution different from that predicted by genetics theory.

| Run! Run! Run!

What description would you use for marathon runners? How about age distribution? Body weight? Length of stride? Heart rate? Blood pressure? What countries do these runners come from? What are their best running times? Make your own estimated distribution for these variables, and then consider a goodness-of-fit test for your distribution compared with available data. For more information on marathon runners, visit the Brase/Brase statistics site at **http://www.cengage.com/statistics/brase** *and find links to the Honolulu marathon site and to the Runners World site.*

**SECTION 10.2
PROBLEMS**

1. | *Statistical Literacy* For a chi-square goodness-of-fit test, how are the degrees of freedom computed?

2. | *Statistical Literacy* How are expected frequencies computed for goodness-of-fit tests?

3. | *Statistical Literacy* Explain why goodness-of-fit tests are always right-tailed tests.

4. | *Critical Thinking* When the sample evidence is sufficient to justify rejecting the null hypothesis in a goodness-of-fit test, can we tell exactly how the distribution of observed values over the specified categories differs from the expected distribution? Explain.

For Problems 5–16, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Find the value of the chi-square statistic for the sample. Are all the expected frequencies greater than 5? What sampling distribution will you use? What are the degrees of freedom?
(c) Find or estimate the *P*-value of the sample test statistic.
(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis that the population fits the specified distribution of categories?
(e) *Interpret* your conclusion in the context of the application.

5. | *Census: Age* The age distribution of the Canadian population and the age distribution of a random sample of 455 residents in the Indian community of Red Lake Village (Northwest Territories) are shown below (based on *U.S. Bureau of the Census, International Data Base*).

Age (years)	Percent of Canadian Population	Observed Number in Red Lake Village
Under 5	7.2%	47
5 to 14	13.6%	75
15 to 64	67.1%	288
65 and older	12.1%	45

Use a 5% level of significance to test the claim that the age distribution of the general Canadian population fits the age distribution of the residents of Red Lake Village.

6. | *Census: Type of Household* The type of household for the U.S. population and for a random sample of 411 households from the community of Dove Creek, Montana, are shown (based on *Statistical Abstract of the United States*).

Type of Household	Percent of U.S. Households	Observed Number of Households in Dove Creek
Married with children	26%	102
Married, no children	29%	112
Single parent	9%	33
One person	25%	96
Other (e.g., roommates, siblings)	11%	68

Use a 5% level of significance to test the claim that the distribution of U.S. households fits the Dove Creek distribution.

7. *Archaeology: Stone Tools* The types of raw materials used to construct stone tools found at the archaeological site Casa del Rito are shown below (*Bandelier Archaeological Excavation Project,* edited by Kohler and Root). A random sample of 1486 stone tools was obtained from a current excavation site.

Raw Material	Regional Percent of Stone Tools	Observed Number of Tools at Current Excavation Site
Basalt	61.3%	906
Obsidian	10.6%	162
Welded tuff	11.4%	168
Pedernal chert	13.1%	197
Other	3.6%	53

Use a 1% level of significance to test the claim that the regional distribution of raw materials fits the distribution at the current excavation site.

8. *Ecology: Deer* The types of browse favored by deer are shown in the following table (*The Mule Deer of Mesa Verde National Park,* edited by Mierau and Schmidt). Using binoculars, volunteers observed the feeding habits of a random sample of 320 deer.

Type of Browse	Plant Composition in Study Area	Observed Number of Deer Feeding on This Plant
Sage brush	32%	102
Rabbit brush	38.7%	125
Salt brush	12%	43
Service berry	9.3%	27
Other	8%	23

Use a 5% level of significance to test the claim that the natural distribution of browse fits the deer feeding pattern.

9. *Meteorology: Normal Distribution* The following problem is based on information from the *National Oceanic and Atmospheric Administration (NOAA) Environmental Data Service.* Let *x* be a random variable that represents the average daily temperature (in degrees Fahrenheit) in July in the town of Kit Carson, Colorado. The *x* distribution has a mean μ of approximately 75°F and standard deviation σ of approximately 8°F. A 20-year study (620 July days) gave the entries in the rightmost column of the following table.

I	II	III	IV
Region under Normal Curve	$x°$F	Expected % from Normal Curve	Observed Number of Days in 20 Years
$\mu - 3\sigma \leq x < \mu - 2\sigma$	$51 \leq x < 59$	2.35%	16
$\mu - 2\sigma \leq x < \mu - \sigma$	$59 \leq x < 67$	13.5%	78
$\mu - \sigma \leq x < \mu$	$67 \leq x < 75$	34%	212
$\mu \leq x < \mu + \sigma$	$75 \leq x < 83$	34%	221
$\mu + \sigma \leq x < \mu + 2\sigma$	$83 \leq x < 91$	13.5%	81
$\mu + 2\sigma \leq x < \mu + 3\sigma$	$91 \leq x < 99$	2.35%	12

(i) Remember that $\mu = 75$ and $\sigma = 8$. Examine Figure 6-5 in Chapter 6. Write a brief explanation for Columns I, II, and III in the context of this problem.

(ii) Use a 1% level of significance to test the claim that the average daily July temperature follows a normal distribution with $\mu = 75$ and $\sigma = 8$.

10. *Meteorology: Normal Distribution* Let x be a random variable that represents the average daily temperature (in degrees Fahrenheit) in January for the town of Hana, Maui. The x variable has a mean μ of approximately 68°F and standard deviation σ of approximately 4°F (see reference in Problem 9). A 20-year study (620 January days) gave the entries in the rightmost column of the following table.

I	II	III	IV
Region Under Normal Curve	$x°$F	Expected % from Normal Curve	Observed Number of Days in 20 Years
$\mu - 3\sigma \leq x < \mu - 2\sigma$	$56 \leq x < 60$	2.35%	14
$\mu - 2\sigma \leq x < \mu - \sigma$	$60 \leq x < 64$	13.5%	86
$\mu - \sigma \leq x < \mu$	$64 \leq x < 68$	34%	207
$\mu \leq x < \mu + \sigma$	$68 \leq x < 76$	34%	215
$\mu + \sigma \leq x < \mu + 2\sigma$	$72 \leq x < 76$	13.5%	83
$\mu + 2\sigma \leq x < \mu + 3\sigma$	$76 \leq x < 80$	2.35%	15

(i) Remember that $\mu = 68$ and $\sigma = 4$. Examine Figure 6-5 in Chapter 6. Write a brief explanation for Columns I, II, and III in the context of this problem.

(ii) Use a 1% level of significance to test the claim that the average daily January temperature follows a normal distribution with $\mu = 68$ and $\sigma = 4$.

11. *Ecology: Fish* The Fish and Game Department stocked Lake Lulu with fish in the following proportions: 30% catfish, 15% bass, 40% bluegill, and 15% pike. Five years later it sampled the lake to see if the distribution of fish had changed. It found that the 500 fish in the sample were distributed as follows.

Catfish	Bass	Bluegill	Pike
120	85	220	75

In the 5-year interval, did the distribution of fish change at the 0.05 level?

12. *Library: Book Circulation* The director of library services at Fairmont College did a survey of types of books (by subject) in the circulation library. Then she used library records to take a random sample of 888 books checked out last

term and classified the books in the sample by subject. The results are shown below.

Subject Area	Percent of Books in Circulation Library on This Subject	Number of Books in Sample on This Subject
Business	32%	268
Humanities	25%	214
Natural science	20%	215
Social science	15%	115
All other subjects	8%	76

Using a 5% level of significance, test the claim that the subject distribution of books in the library fits the distribution of books checked out by students.

13. *Census: California* The accuracy of a census report on a city in southern California was questioned by some government officials. A random sample of 1215 people living in the city was used to check the report, and the results are shown here:

Ethnic Origin	Census Percent	Sample Result
Black	10%	127
Asian	3%	40
Anglo	38%	480
Latino/Latina	41%	502
Native American	6%	56
All others	2%	10

Using a 1% level of significance, test the claim that the census distribution and the sample distribution agree.

14. *Marketing: Compact Discs* Snoop Incorporated is a firm that does market surveys. The Rollum Sound Company hired Snoop to study the age distribution of people who buy compact discs. To check the Snoop report, Rollum used a random sample of 519 customers and obtained the following data:

Customer Age (years)	Percent of Customers from Snoop Report	Number of Customers from Sample
Younger than 14	12%	88
14–18	29%	135
19–23	11%	52
24–28	10%	40
29–33	14%	76
Older than 33	24%	128

Using a 1% level of significance, test the claim that the distribution of customer ages in the Snoop report agrees with that of the sample report.

15. *Accounting Records: Benford's Law* Benford's law states that the first nonzero digits of numbers drawn at random from a large complex data file have the following probability distribution (Reference: American Statistical Association, *Chance*, Vol. 12, No. 3, pp. 27–31; see also the Focus Problem of Chapter 9).

First nonzero digit	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Suppose that $n = 275$ numerical entries were drawn at random from a large accounting file of a major corporation. The first nonzero digits were recorded for the sample.

First nonzero digit	1	2	3	4	5	6	7	8	9
Sample frequency	83	49	32	22	25	18	13	17	16

Use a 1% level of significance to test the claim that the distribution of first nonzero digits in this accounting file follows Benford's law.

16. *Fair Dice: Uniform Distribution* A gambler complained about the dice. They seemed to be loaded! The dice were taken off the table and tested one at a time. One die was rolled 300 times and the following frequencies were recorded.

Outcome	1	2	3	4	5	6
Observed frequency O	62	45	63	32	47	51

Do these data indicate that the die is unbalanced? Use a 1% level of significance. *Hint:* If the die is balanced, all outcomes should have the same expected frequency.

17. *Highway Accidents: Poisson Distribution* A civil engineer has been studying the frequency of vehicle accidents on a certain stretch of interstate highway. Long-term history indicates that there has been an average of 1.72 accidents per day on this section of the interstate. Let r be a random variable that represents number of accidents per day. Let O represent the number of observed accidents per day based on local highway patrol reports. A random sample of 90 days gave the following information.

r	0	1	2	3	4 or more
O	22	21	15	17	15

(a) The civil engineer wants to use a Poisson distribution to represent the probability of r, the number of accidents per day. The Poisson distribution is

$$P(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

where $\lambda = 1.72$ is the average number of accidents per day. Compute $P(r)$ for $r = 0, 1, 2, 3$, and 4 or more.

(b) Compute the expected number of accidents $E = 90P(r)$ for $r = 0, 1, 2, 3$, and 4 or more.

(c) Compute the sample statistic $\chi^2 = \Sigma \dfrac{(O - E)^2}{E}$ and the degrees of freedom.

(d) Test the statement that the Poisson distribution fits the sample data. Use a 1% level of significance.

18. *Bacteria Colonies: Poisson Distribution* A pathologist has been studying the frequency of bacterial colonies within the field of a microscope using samples of throat cultures from healthy adults. Long-term history indicates that there is an average of 2.80 bacteria colonies per field. Let r be a random variable that represents the number of bacteria colonies per field. Let O represent the number of observed bacteria colonies per field for throat cultures from healthy adults. A random sample of 100 healthy adults gave the following information.

r	0	1	2	3	4	5 or more
O	12	15	29	18	19	7

(a) The pathologist wants to use a Poisson distribution to represent the probability of r, the number of bacteria colonies per field. The Poisson distribution is

$$P(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

where $\lambda = 2.80$ is the average number of bacteria colonies per field. Compute $P(r)$ for $r = 0, 1, 2, 3, 4,$ and 5 or more.

(b) Compute the expected number of colonies $E = 100P(r)$ for $r = 0, 1, 2, 3, 4,$ and 5 or more.

(c) Compute the sample statistic $\chi^2 = \Sigma\dfrac{(O - E)^2}{E}$ and the degrees of freedom.

(d) Test the statement that the Poisson distribution fits the sample data. Use a 5% level of significance.

SECTION 10.3

Testing and Estimating a Single Variance or Standard Deviation

FOCUS POINTS

- Set up a test for a single variance σ^2.
- Compute the sample χ^2 statistic.
- Use the χ^2 distribution to estimate a *P*-value and conclude the test.
- Compute confidence intervals for σ^2 or σ.

Testing σ^2

Many problems arise that require us to make decisions about variability. In this section, we will study two kinds of problems: (1) we will test hypotheses about the variance (or standard deviation) of a population, and (2) we will find confidence intervals for the variance (or standard deviation) of a population. It is customary to talk about variance instead of standard deviation because our techniques employ the sample variance rather than the standard deviation. Of course, the standard deviation is just the square root of the variance, so any discussion about variance is easily converted to a similar discussion about standard deviation.

Test of variance

Let us consider a specific example in which we might wish to test a hypothesis about the variance. Almost everyone has had to wait in line. In a grocery store, bank, post office, or registration center, there are usually several checkout or service areas. Frequently, each service area has its own independent line. However, many businesses and government offices are adopting a "single-line" procedure.

In a single-line procedure, there is only one waiting line for everyone. As any service area becomes available, the next person in line gets served. The old, independent-lines procedure has a line at each service center. An incoming customer simply picks the shortest line and hopes it will move quickly. In either procedure, the number of clerks and the rate at which they work is the same, so the average waiting time is the *same*. What is the advantage of the single-line procedure? The difference is in the *attitudes* of people who wait in the lines. A lengthy waiting line will be more acceptable, even though the average waiting time is the same, if the variability of waiting times is smaller. When the variability is small, the inconvenience of waiting (although it might not be reduced) does become more predictable. This means impatience is reduced and people are happier.

To test the hypothesis that variability is less in a single-line process, we use the chi-square distribution. The next theorem tells us how to use the sample and population variance to compute values of χ^2.

THEOREM 10.1 If we have a *normal* population with variance σ^2 and a random sample of n measurements is taken from this population with sample variance s^2, then

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

has a chi-square distribution with degrees of freedom $d.f. = n - 1$.

Recall that the chi-square distribution is *not* symmetrical and that there are different chi-square distributions for different degrees of freedom. Table 7 of Appendix II gives chi-square values for which the area α is to the *right* of the given chi-square value.

EXAMPLE 3 χ^2 DISTRIBUTION

(a) Find the χ^2 value such that the area to the right of χ^2 is 0.05 when $d.f. = 10$.

SOLUTION: Since the area to the right of χ^2 is to be 0.05, we look in the right-tail area $= 0.050$ column and the row with $d.f. = 10$. $\chi^2 = 18.31$ (see Figure 10-6a).

(b) Find the χ^2 value such that the area to the *left* of χ^2 is 0.05 when $d.f. = 10$.

SOLUTION: When the area to the left of χ^2 is 0.05, the corresponding area to the *right* is $1 - 0.05 = 0.95$, so we look in the right-tail area $= 0.950$ column and the row with $d.f. = 10$. We find $\chi^2 = 3.94$ (see Figure 10-6b).

FIGURE 10-6

χ^2 Distribution with $d.f. = 10$.

(a)

Area to the right $= 0.05$

18.31 χ^2

(b)

Area is 0.05

Area to the right $= 0.95$
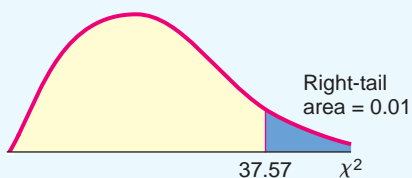
3.94 χ^2
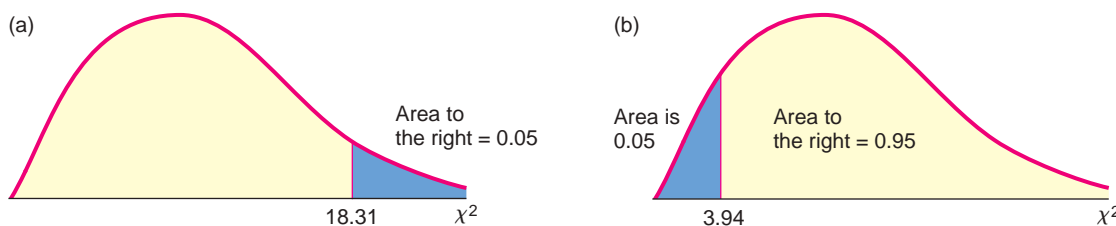
GUIDED EXERCISE 7 χ^2 **distribution**

(a) Find the area to the *right* of $\chi^2 = 37.57$ when $d.f. = 20$.

Use Table 7 of Appendix II. In the row headed by $d.f. = 20$, find the column with $\chi^2 = 37.57$. The column header 0.010 is the area to the right (see Figure 10-7).

FIGURE 10-7 χ^2 Distribution with $d.f. = 20$

Right-tail area $= 0.01$

37.57 χ^2

Continued

GUIDED EXERCISE 7   *continued*

(b) Find the area to the *left* of $\chi^2 = 8.26$ when $d.f. = 20$.

   (i) First use Table 7 of Appendix II to find the area to the right of $\chi^2 = 8.26$.

   (ii) To get the area to the *left* of $\chi^2 = 8.26$, subtract the area to the right from 1.

⇨ Find the column with $\chi^2 = 8.26$ in the row headed by $d.f. = 20$. The column header 0.990 gives the area to the *right* of $\chi^2 = 8.26$.

Next, subtract the area to the right from 1. Area to *left* = 1 − Area to *right* = 1 − 0.990 = 0.010 (see Figure 10-8).

FIGURE 10-8 χ^2 Distribution with $d.f. = 20$

Area is 0.01

Right-tail area = 0.99

8.26

χ^2
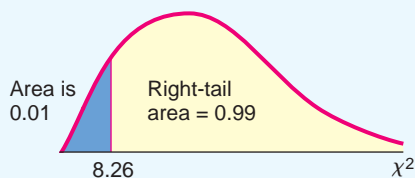
Table 10-13 summarizes the techniques for using the chi-square distribution (Table 7 of Appendix II) to find *P*-values for a right-tailed test, a left-tailed test, and a two-tailed test. Example 4 demonstrates the technique of finding *P*-values for a left-tailed test. Example 5 demonstrates the technique for a two-tailed test, and Guided Exercise 8 uses a right-tailed test.

Now let's use Theorem 10.1 and our knowledge of the chi-square distribution to determine if a single-line procedure has less variance of waiting times than independent lines.

TABLE 10-13   *P*-values for Chi-Square Distribution Table (Table 7, Appendix II)

(a) Two-tailed test

   Remember that the *P*-value is the probability of getting a test statistic as extreme as, or more extreme than, the test statistic computed from the sample. For a two-tailed test, we need to account for corresponding equal areas in both the upper and lower tails. This means that in each tail, we have an area of *P*-value/2. The total *P*-value is then

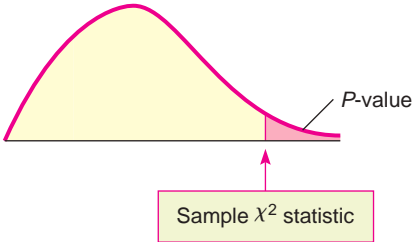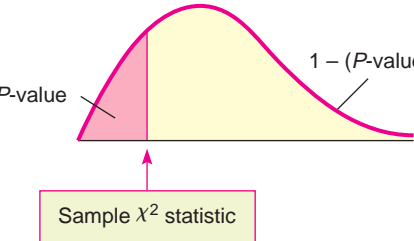$$P\text{-value} = 2\left(\frac{P\text{-value}}{2}\right)$$

$\dfrac{P\text{-value}}{2}$

Sample χ^2 statistic

$\dfrac{P\text{-value}}{2}$

Sample χ^2 statistic
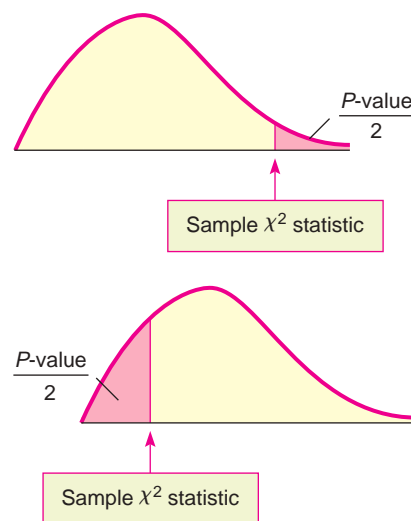
Be sure to choose the area in the appropriate tail (left or right) so that

$$\frac{P\text{-value}}{2} \leq 0.5.$$

TABLE 10-13	***P*-values for Chi-Square Distribution Table (Table 7, Appendix II)** (*continued*)

(b) Right-tailed test Since the chi-square table gives right-tail probabilities, you can use the table directly to find or estimate the *P*-value.	*P*-value Sample χ^2 statistic
(c) Left-tailed test Since the chi-square table gives right-tail probabilities, you first find or estimate the quantity $1 - (P\text{-value})$ right tail. Then subtract from 1 to get the *P*-value of the left-tail.	*P*-value $1 - (P\text{-value})$ Sample χ^2 statistic

EXAMPLE 4   TESTING THE VARIANCE (LEFT-TAILED TEST)

For years, a large discount store has used independent lines to check out customers. Historically, the standard deviation of waiting times is 7 minutes. The manager tried a new, single-line procedure. A random sample of 25 customers using the single-line procedure was monitored, and it was found that the standard deviation for waiting times was only $s = 5$ minutes. Use $\alpha = 0.05$ to test the claim that the variance in waiting times is reduced for the single-line method. Assume the waiting times are normally distributed.

Establish H_0 and H_1

SOLUTION: As a null hypothesis, we assume that the variance of waiting times is the same as that of the former independent-lines procedure. The alternate hypothesis is that the variance for the single-line procedure is less than that for the independent-lines procedure. If we let σ be the standard deviation of waiting times for the single-line procedure, then σ^2 is the variance, and we have

Sample χ^2 value and degrees of freedom

$$H_0: \sigma^2 = 49 \qquad\qquad H_1: \sigma^2 < 49 \qquad\qquad (\text{use } 7^2 = 49)$$

We use the chi-square distribution to test the hypotheses. Assuming that the waiting times are normally distributed, we compute our observed value of χ^2 by using Theorem 10.1, with $n = 25$.

$$s = 5 \quad \text{so} \quad s^2 = 25 \quad \text{(observed from sample)}$$
$$\sigma = 7 \quad \text{so} \quad \sigma^2 = 49 \quad \text{(\textit{from } } H_0: \sigma^2 = 49)$$
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)25}{49} \approx 12.24$$
$$d.f. = n - 1 = 25 - 1 = 24$$

Checkout lines

Next we estimate the *P*-value for $\chi^2 = 12.24$. Since we have a left-tailed test, the *P*-value is the area of the chi-square distribution that lies to the *left* of $\chi^2 = 12.24$, as shown in Figure 10-9 on the next page.

P-value

To estimate the *P*-value on the left, we consider the fact that the area of the right tail is between 0.975 and 0.990. To find an estimate for the area of the left tail, we *subtract* each right-tail endpoint from 1. The *P*-value (area of the left tail) is in the interval
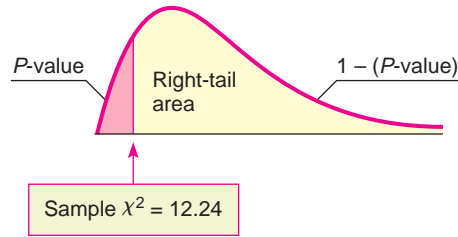
P-value

Right-tail Area	0.990	0.975
d.f. = 24	10.86	12.40

$$1 - 0.990 < P\text{-value of left tail} < 1 - 0.975$$
$$0.010 < P\text{-value} < 0.025$$

Test conclusion

To conclude the test, we compare the *P*-value to the level of significance $\alpha = 0.05$.

Since the *P*-value is less than α, we reject H_0.

Interpretation

Interpretation At the 5% level of significance, we conclude that the variance of waiting times for a single line is less than the variance of waiting times for multiple lines.

The steps used in Example 4 for testing the variance σ^2 are summarized as follows.

PROCEDURE

HOW TO TEST σ^2

Requirements

You first need to know that a random variable *x* has a normal distribution. In testing σ^2, the normal assumption must be strictly observed (whereas in testing means, we can say "normal" or "approximately normal"). Next you need a random sample (size $n \geq 2$) of values from the *x* distribution for which you compute the sample variance s^2.

Procedure

1. In the context of the problem, state the *null hypothesis H_0* and the *alternate hypothesis H_1*, and set the *level of significance α*.

2. Use the value of σ^2 given in the null hypothesis H_0, the sample variance s^2, and the sample size *n* to compute the *sample test statistic*

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \text{ with degrees of freedom } d.f. = n - 1$$

3. Use a chi-square distribution and the type of test to find or estimate the *P-value*. Use the procedures shown in Table 10-13 and Table 7 of Appendix II.

4. *Conclude* the test. If *P*-value $\leq \alpha$, then reject H_0. If *P*-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

EXAMPLE 5   TESTING THE VARIANCE (TWO-TAILED TEST)

Let *x* be a random variable that represents weight loss (in pounds) after following a certain diet for 6 months. After extensive study, it is found that *x* has a normal distribution with $\sigma = 5.7$ pounds. A new modification of the diet has been implemented. A random sample of $n = 21$ people use the modified diet for 6 months. For these people, the sample standard deviation of weight loss is $s = 4.1$ pounds. Does this result indicate that the variance of weight loss for the modified diet is

different (either way) from the variance of weight loss for the original diet? Use $\alpha = 0.01$. Assume weight loss for each diet follows a normal distribution.

(a) What is the level of significance? State the null and alternate hypotheses.

SOLUTION: We are using $\alpha = 0.01$. The standard deviation of weight loss for the original diet is $\sigma = 5.7$ pounds, so the variance is $\sigma^2 = 32.49$. The null hypothesis is that the weight loss variance for the modified diet is the same as that for the original diet. The alternate hypothesis is that the variance is different.

$$H_0: \sigma^2 = 32.49 \quad H_1: \sigma^2 \neq 32.49$$

(b) Compute the sample test statistic χ^2 and the degrees of freedom.

SOLUTION: Using sample size $n = 21$, sample standard deviation $s = 4.1$ pounds, and $\sigma^2 = 32.49$ from the null hypothesis, we have
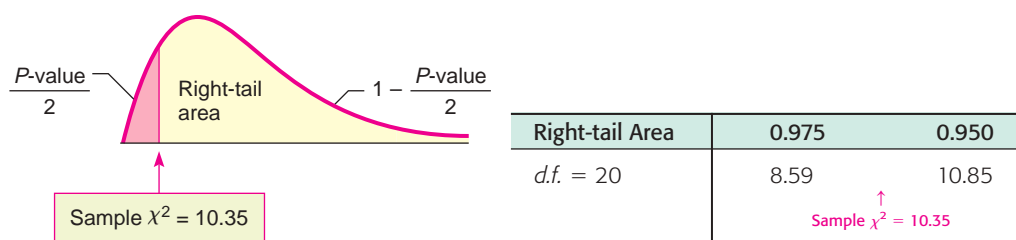
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(21-1)4.1^2}{32.49} \approx 10.35$$

with degrees of freedom $d.f. = n - 1 = 21 - 1 = 20$.

(c) Use the chi-square distribution (Table 7 of Appendix II) to estimate the *P*-value.

FIGURE 10-10

P-value

Right-tail Area	0.975	0.950
d.f. = 20	8.59	10.85
		Sample χ^2 = 10.35

SOLUTION: For a *two-tailed* test, the area beyond the sample χ^2 represents *half* the total *P*-value or (*P*-value)/2. Figure 10-10 shows this region, which is to the left of $\chi^2 \approx 10.35$. However, Table 7 of Appendix II gives the areas in the *right tail*. We use Table 7 to find the area in the right tail and then subtract from 1 to find the corresponding area in the left tail.

From the table, we see that the right-tail area falls in the interval between 0.950 and 0.975. Subtracting each endpoint of the interval from 1 gives us an interval containing (*P*-value)/2. Multiplying by 2 gives an interval for the *P*-value.
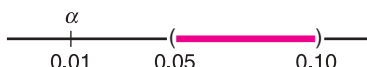
$$1 - 0.975 < \frac{P\text{-value}}{2} < 1 - 0.950 \quad \text{Subtract right-tail-area endpoints from 1.}$$

$$0.025 < \frac{P\text{-value}}{2} < 0.050$$

$$0.05 < P\text{-value} < 0.10 \quad \text{Multiply each part by 2.}$$

(d) *Conclude* the test.

SOLUTION: The *P*-value is greater than $\alpha = 0.01$, so we do not reject H_0.

(e) *Interpret* the conclusion in the context of the application.

SOLUTION: At the 1% level of significance, there is insufficient evidence to conclude that the variance of weight loss using the modified diet is different from the variance of weight loss using the original diet.

GUIDED EXERCISE 8 | **Testing the variance (right-tailed test)**

Certain industrial machines require overhaul when wear on their parts introduces too much variability to pass inspection. A government official is visiting a dentist's office to inspect the operation of an x-ray machine. If the machine emits too little radiation, clear photographs cannot be obtained. However, too much radiation can be harmful to the patient. Government regulations specify an average emission of 60 millirads with standard deviation σ of 12 millirads, and the machine has been set for these readings. After examining the machine, the inspector is satisfied that the average emission is still 60 millirads. However, there is wear on certain mechanical parts. To test variability, the inspector takes a random sample of 30 x-ray emissions and finds the sample standard deviation to be $s = 15$ millirads. Does this support the claim that the variance is too high (i.e., the machine should be overhauled)? Use a 1% level of significance. Assume the emissions follow a normal distribution.

Let σ be the (population) standard deviation of emissions (in millirads) of the machine in its present condition.

(a) What is α? State H_0 and H_1.

➡ $\alpha = 0.01$. Government regulations specify that $\sigma = 12$. This means that the variance $\sigma^2 = 144$. We are to test the claim that the variance is higher than government specifications allow.

$$H_0: \sigma^2 = 144 \quad \text{and} \quad H_1: \sigma^2 > 144$$

(b) Compute the sample test statistic χ^2 and corresponding degrees of freedom.

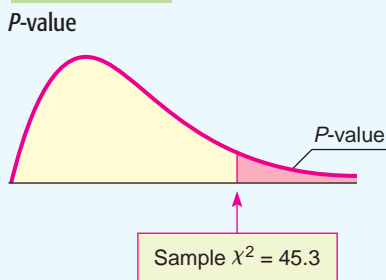➡ Using $n = 30$, $s = 15$, and $\sigma^2 = 144$ from H_0,

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(30-1)15^2}{144} \approx 45.3$$

Degrees of freedom $d.f. = n - 1 = 30 - 1 = 29$

(c) Estimate the *P*-value for the sample $\chi^2 = 45.3$ with $d.f. = 29$.

➡ Since this is a *right-tailed* test, we look up *P*-values directly in the chi-square table (Table 7 of Appendix II).
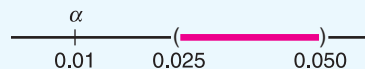
FIGURE 10-11

P-value

Sample $\chi^2 = 45.3$

Right-tail Area	0.050	0.025
$d.f. = 29$	42.56	45.72

↑
Sample $\chi^2 = 45.3$

$0.025 < P\text{-value} < 0.050$

➡ The *P*-value for $\chi^2 = 45.3$ is greater than $\alpha = 0.01$.

(d) *Conclude* the test.

Fail to reject H_0.

(e) *Interpret* the conclusion in the context of the application.

➡ At the 1% level of significance, there is insufficient evidence to conclude that the variance of the radiation emitted by the machine is greater than that specified by government regulations. The evidence does not indicate that an adjustment is necessary at this time.

Confidence Interval for σ^2

Confidence Interval

Sometimes it is important to have a *confidence interval* for the variance or standard deviation. Let us look at another example.

Mr. Wilson is a truck farmer in California who makes his living on a large single-vegetable crop of green beans. Because modern machinery is being used, the entire crop must be harvested at the same time. Therefore, it is important to plant a variety of green beans that mature all at once. This means that Mr. Wilson wants a small standard deviation between maturing times of individual plants. A seed company is trying to develop a new variety of green bean with a small standard deviation of maturing times. To test the new variety, Mr. Wilson planted 30 of the new seeds and carefully observed the number of days required for each plant to arrive at its peak of maturity. The maturing times for these plants had a sample standard deviation of $s = 3.4$ days. How can we find a 95% confidence interval for the population standard deviation of maturing times for this variety of green bean? The answer to this question is based on the following procedure.

PROCEDURE

> ### HOW TO FIND A CONFIDENCE INTERVAL FOR σ^2 AND σ
>
> *Requirements*
>
> Let x be a random variable with a normal distribution and unknown population standard deviation σ. Take a random sample of size n from the x distribution and compute the sample standard deviation s.
>
> *Procedure*
>
> A **confidence interval for the population variance** σ^2 is
>
> $$\frac{(n-1)s^2}{\chi_U^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2} \tag{1}$$
>
> and a **confidence interval for the population standard deviation** σ is
>
> $$\sqrt{\frac{(n-1)s^2}{\chi_U^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_L^2}}$$
>
> where
>
> $\quad c = $ confidence level $(0 < c < 1)$
>
> $\quad n = $ sample size $(n \geq 2)$
>
> $\quad \chi_U^2 = $ chi-square value from Table 7 of Appendix II using $d.f. = n - 1$ and right-tail area $= (1 - c)/2$
>
> $\quad \chi_L^2 = $ chi-square value from Table 7 of Appendix II using $d.f. = n - 1$ and right-tail area $= (1 + c)/2$
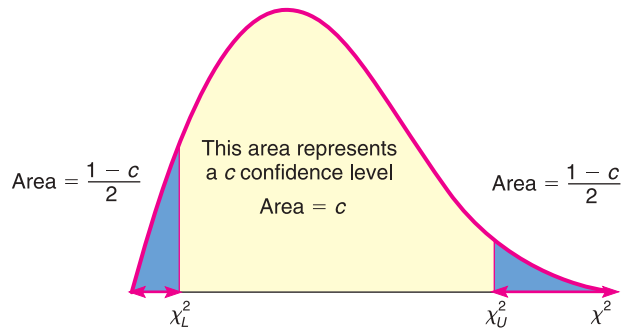
χ_U^2

χ_L^2

From Figure 10-12, on the next page, we see that a c confidence level on a chi-square distribution with equal probability in each tail does not center the middle of the corresponding interval under the peak of the curve. This is to be expected because a chi-square curve is skewed to the right.

COMMENT   Note that the method of computing confidence intervals for variances is different from the method of computing confidence intervals for means or proportions as studied in Chapter 7. Confidence intervals for σ^2 do not involve a maximal error of estimate E. Rather, the endpoints of the confidence interval are computed directly using the sample statistic s^2, the sample size, and the critical values.

$$\text{Area} = \frac{1 - c}{2}$$

This area represents
a *c* confidence level
Area = *c*

$$\text{Area} = \frac{1 - c}{2}$$

χ_L^2 χ_U^2 χ^2

Now let us finish our example regarding the variance of maturing times for green beans.

EXAMPLE 6   CONFIDENCE INTERVALS FOR σ^2 AND σ

A random sample of $n = 30$ green bean plants has a sample standard deviation of $s = 3.4$ days for maturity. Find a 95% confidence interval for the population variance σ^2. Assume the distribution of maturity times is normal.

SOLUTION: To find the confidence interval, we use the following values:

  $c = 0.95$ confidence level
  $n = 30$ sample size
  $d.f. = n - 1 = 30 - 1 = 29$ degrees of freedom
  $s = 3.4$ sample standard deviation

To find the value of χ_U^2, we use Table 7 of Appendix II with $d.f. = 29$ and right-tail area $= (1 - c)/2 = (1 - 0.95)/2 = 0.025$. From Table 7, we get

  $\chi_U^2 = 45.72$

To find the value of χ_L^2, we use Table 7 of Appendix II with $d.f. = 29$ and right-tail area $= (1 + c)/2 = (1 + 0.95)/2 = 0.975$. From Table 7, we get

  $\chi_L^2 = 16.05$

Formula (1) tells us that our desired 95% confidence interval for σ^2 is

$$\frac{(n - 1)s^2}{\chi_U^2} < \sigma^2 < \frac{(n - 1)s^2}{\chi_L^2}$$

$$\frac{(30 - 1)(3.4)^2}{45.72} < \sigma^2 < \frac{(30 - 1)(3.4)^2}{16.05}$$

$$7.33 < \sigma^2 < 20.89$$

To find a 95% confidence interval for σ, we simply take square roots; therefore, a 95% confidence interval for σ is

$$\sqrt{7.33} < \sigma < \sqrt{20.89}$$
$$2.71 < \sigma < 4.57$$

GUIDED EXERCISE 9 | *Confidence intervals for σ^2 and σ*

A few miles off the Kona coast of the island of Hawaii, a research vessel lies anchored. This ship makes electrical energy from the solar temperature differential of (warm) surface water versus (cool) deep water. The basic idea is that the warm water is flushed over coils to vaporize a special fluid. The vapor is under pressure and drives electrical turbines. Then some electricity is used to pump up cold water to cool the vapor back to a liquid, and the process is repeated. Even though some electricity is used to pump up the cold water, there is plenty left to supply a moderate-sized Hawaiian town.

Continued

GUIDED EXERCISE 9   *continued*

The subtropic sun always warms up surface water to a reliable temperature, but ocean currents can change the temperature of the deep, cooler water. If the deep-water temperature is too variable, the power plant cannot operate efficiently or possibly cannot operate at all. To estimate the variability of deep ocean water temperatures, a random sample of 25 near-bottom readings gave a sample standard deviation of 7.3°C.

Find a 99% confidence interval for the variance σ^2 and standard deviation σ of the deep-water temperatures. Assume deep-water temperatures are normally distributed.

(a) Determine the following values: $c =$ _____;
    $n =$ _____; $d.f =$ _____; $s =$ _____.

→ $c = 0.99$; $n = 25$; $d.f. = 24$; $s = 7.3$

(b) What is the value of χ_U^2? of χ_L^2?

→ We use Table 7 of Appendix II with $d.f. = 24$.

For χ_U^2, right-tail area $= (1 - 0.99)/2 = 0.005$

$$\chi_U^2 = 45.56$$

For χ_L^2, right-tail area $= (1 + 0.99)/2 = 0.995$

$$\chi_L^2 = 9.89$$

(c) Find a 99% confidence interval for σ^2.

→
$$\frac{(n-1)s^2}{\chi_U^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

$$\frac{(24)(7.3)^2}{45.56} < \sigma^2 < \frac{24(7.3)^2}{9.89}$$

$$28.07 < \sigma^2 < 129.32$$

(d) Find a 99% confidence interval for σ.

→
$$\sqrt{28.07} < \sqrt{\sigma^2} < \sqrt{129.32}$$

$$5.30 < \sigma < 11.37$$

VIEWPOINT | Adoption—A Good Choice!

Cuckoos are birds that are known to lay their eggs in the nests of other (host) birds. The host birds then hatch the eggs and adopt the cuckoo chicks as their own. Birds such as the meadow pipit, tree pipit, hedge sparrow, robin, and wren have all played host to cuckoo eggs and adopted their chicks. L. H. C. Tippett (1902–1985) was a pioneer in the field of statistical quality control who collected data on cuckoo eggs found in the nests of other birds. For more information and data from Tippett's study, visit the Brase/Brase statistics site at **http://www.cengage.com/statistics/brase** *and find a link to DASL, the Carnegie Mellon University Data and Story Library. Find Biology under Data Subjects, and then select the Cuckoo Egg Length Data file.*

**SECTION 10.3
PROBLEMS**

1. *Statistical Literacy* Does the x distribution need to be normal in order to use the chi-square distribution to test the variance? Is it acceptable to use the chi-square distribution to test the variance if the x distribution is simply mound-shaped and more or less symmetric?

2. *Critical Thinking* The x distribution must be normal in order to use a chi-square distribution to test the variance. What are some methods you can use to assess whether the x distribution is normal? *Hint:* See Chapter 6 and goodness-of-fit tests.

For Problems 3–11, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Find the value of the chi-square statistic for the sample. What are the degrees of freedom? What assumptions are you making about the original distribution?
(c) Find or estimate the P-value of the sample test statistic.

(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis of independence?

(e) *Interpret* your conclusion in the context of the application.

(f) Find the requested confidence interval for the population variance or population standard deviation. Interpret the results in the context of the application.

In each of the following problems, assume a normal population distribution.

3. *Archaeology: Chaco Canyon* The following problem is based on information from *Archaeological Surveys of Chaco Canyon, New Mexico,* by A. Hayes, D. Brugge, and W. Judge, University of New Mexico Press. A *transect* is an archaeological study area that is 1/5 mile wide and 1 mile long. A *site* in a transect is the location of a significant archaeological find. Let x represent the number of sites per transect. In a section of Chaco Canyon, a large number of transects showed that x has a population variance $\sigma^2 = 42.3$. In a different section of Chaco Canyon, a random sample of 23 transects gave a sample variance $s^2 = 46.1$ for the number of sites per transect. Use a 5% level of significance to test the claim that the variance in the new section is greater than 42.3. Find a 95% confidence interval for the population variance.

4. *Sociology: Marriage* The following problem is based on information from an article by N. Keyfitz in the *American Journal of Sociology* (Vol. 53, pp. 470–480). Let x = age in years of a rural Quebec woman at the time of her first marriage. In the year 1941, the population variance of x was approximately $\sigma^2 = 5.1$. Suppose a recent study of age at first marriage for a random sample of 41 women in rural Quebec gave a sample variance $s^2 = 3.3$. Use a 5% level of significance to test the claim that the current variance is less than 5.1. Find a 90% confidence interval for the population variance.

5. *Mountain Climbing: Accidents* The following problem is based on information taken from *Accidents in North American Mountaineering* (jointly published by The American Alpine Club and The Alpine Club of Canada). Let x represent the number of mountain climbers killed each year. The long-term variance of x is approximately $\sigma^2 = 136.2$. Suppose that for the past 8 years, the variance has been $s^2 = 115.1$. Use a 1% level of significance to test the claim that the recent variance for number of mountain climber deaths is less than 136.2. Find a 90% confidence interval for the population variance.

6. *Professors: Salaries* The following problem is based on information taken from *Academe, Bulletin of the American Association of University Professors*. Let x represent the average annual salary of college and university professors (in thousands of dollars) in the United States. For all colleges and universities in the United States, the population variance of x is approximately $\sigma^2 = 47.1$. However, a random sample of 15 colleges and universities in Kansas showed that x has a sample variance $s^2 = 83.2$. Use a 5% level of significance to test the claim that the variance for colleges and universities in Kansas is greater than 47.1. Find a 95% confidence interval for the population variance.

7. *Medical: Clinical Test* A new kind of typhoid shot is being developed by a medical research team. The old typhoid shot was known to protect the population for a mean time of 36 months, with a standard deviation of 3 months. To test the time variability of the new shot, a random sample of 23 people were given the new shot. Regular blood tests showed that the sample standard deviation of protection times was 1.9 months. Using a 0.05 level of significance, test the claim that the new typhoid shot has a smaller variance of protection times. Find a 90% confidence interval for the population standard deviation.

8. *Veterinary Science: Tranquilizer* Jim Mead is a veterinarian who visits a Vermont farm to examine prize bulls. In order to examine a bull, Jim first gives the animal a tranquilizer shot. The effect of the shot is supposed to last an average of 65 minutes, and it usually does. However, Jim sometimes gets chased out of the pasture by a bull that recovers too soon, and other times he becomes worried about prize bulls that take too long to recover. By reading journals, Jim has found that the tranquilizer should have a mean duration time of 65 minutes, with a standard deviation of 15 minutes. A random sample of 10 of Jim's bulls had a mean tranquilized duration time of close to 65 minutes but a standard deviation of 24 minutes. At the 1% level of significance, is Jim justified in the claim that the variance is larger than that stated in his journal? Find a 95% confidence interval for the population standard deviation.

9. *Engineering: Jet Engines* The fan blades on commercial jet engines must be replaced when wear on these parts indicates too much variability to pass inspection. If a single fan blade broke during operation, it could severely endanger a flight. A large engine contains thousands of fan blades, and safety regulations require that variability measurements on the population of all blades not exceed $\sigma^2 = 0.18$ mm^2. An engine inspector took a random sample of 61 fan blades from an engine. She measured each blade and found a sample variance of 0.27 mm^2. Using a 0.01 level of significance, is the inspector justified in claiming that all the engine fan blades must be replaced? Find a 90% confidence interval for the population standard deviation.

10. *Law: Bar Exam* A factor in determining the usefulness of an examination as a measure of demonstrated ability is the amount of spread that occurs in the grades. If the spread or variation of examination scores is very small, it usually means that the examination was either too hard or too easy. However, if the variance of scores is moderately large, then there is a definite difference in scores between "better," "average," and "poorer" students. A group of attorneys in a Midwest state has been given the task of making up this year's bar examination for the state. The examination has 500 total possible points, and from the history of past examinations, it is known that a standard deviation of around 60 points is desirable. Of course, too large or too small a standard deviation is not good. The attorneys want to test their examination to see how good it is. A preliminary version of the examination (with slight modifications to protect the integrity of the real examination) is given to a random sample of 24 newly graduated law students. Their scores give a sample standard deviation of 72 points.
    (i) Using a 0.01 level of significance, test the claim that the population standard deviation for the new examination is 60 against the claim that the population standard deviation is different from 60.
    (ii) Find a 99% confidence interval for the population variance.
    (iii) Find a 99% confidence interval for the population standard deviation.

11. *Engineering: Solar Batteries* A set of solar batteries is used in a research satellite. The satellite can run on only one battery, but it runs best if more than one battery is used. The variance σ^2 of lifetimes of these batteries affects the useful lifetime of the satellite before it goes dead. If the variance is too small, all the batteries will tend to die at once. Why? If the variance is too large, the batteries are simply not dependable. Why? Engineers have determined that a variance of $\sigma^2 = 23$ months (squared) is most desirable for these batteries. A random sample of 22 batteries gave a sample variance of 14.3 months (squared).
    (i) Using a 0.05 level of significance, test the claim that $\sigma^2 = 23$ against the claim that σ^2 is different from 23.
    (ii) Find a 90% confidence interval for the population variance σ^2.
    (iii) Find a 90% confidence interval for the population standard deviation σ.

PART II: INFERENCES USING THE *F* DISTRIBUTION

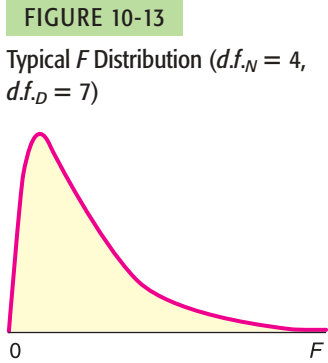### Overview of the *F* Distribution

The *F* probability distribution was first developed by the English statistician Sir Ronald Fisher (1890–1962). Fisher had a long and distinguished career in statistics, including research work at the agricultural station at Rothamsted (see Problems 5 and 6 at the end of Section 10.4). During his time there he developed the subjects of experimental design and ANOVA (see Sections 10.5 and 10.6).

The *F* distribution is a ratio of two independent chi-square random variables, each with its own degrees of freedom,

$d.f._N$ = degrees of freedom in the numerator
$d.f._D$ = degrees of freedom in the denominator

The *F* distribution depends on these *two* degrees of freedom, $d.f._N$ and $d.f._D$. Figure 10-13 shows a typical *F* distribution.

Typical *F* Distribution ($d.f._N = 4$, $d.f._D = 7$)

Properties of the *F* distribution

- The *F* distribution is not symmetrical. It is skewed to the right.
- Values of *F* are always greater than or equal to zero.
- A specific *F* distribution (see Table 8 of Appendix II) is determined from *two* degrees of freedom. These are called *degrees of freedom for the numerator $d.f._N$* and *degrees of freedom for the denominator $d.f._D$*.
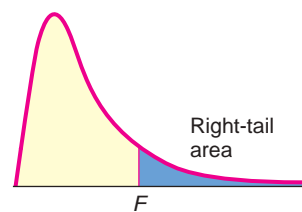- Area under the entire *F* distribution is one.

The degrees of freedom used in the *F* distribution depend on the particular application. Table 8 of Appendix II shows areas in the right-tail of different *F* distributions according to the degrees of freedom in both the numerator, $d.f._N$, and the denominator, $d.f._D$.

Table 10-14 shows an excerpt from Table 8. Notice that $d.f._D$ are row headers. For each $d.f._D$, right-tail areas from 0.100 down to 0.001 are provided in the next column. Then, under column headers for $d.f._N$ values of *F* are given corresponding to $d.f._D$, the right-tail area, and $d.f._N$. For example, for $d.f._D = 2$, Right-tail area = 0.010, and $d.f._N = 3$, the corresponding value of *F* is 99.17

In this text we present three applications of the *F* distribution: testing two variances (Section 10.4), one-way ANOVA (Section 10.5), and two-way ANOVA (Section 10.6). Sections 10.4 and 10.5 are self-contained and can be studied independently. Section 10.5 should be studied before Section 10.6.

TABLE 10-14   Excerpt from Table 8 (Appendix II): The *F* Distribution

Right-tail
area

$d.f._D$	Right-tail Area	Degrees of Freedom for Numerator $d.f._N$			
		1	2	3	4 ...
⋮	⋮	⋮	⋮	⋮	⋮
	0.100	8.53	9.00	9.16	9.24
	0.050	18.51	19.00	19.16	19.25
✓ 2	0.025	38.51	39.00	39.17	39.25
	0.010	98.50	99.00	99.17	99.25
	0.001	998.50	999.00	999.17	9999.25

Testing Two Variances

FOCUS POINTS

- Set up a test for two variances σ_1^2 and σ_2^2.
- Use sample variances to compute the sample F statistic.
- Use the F distribution to estimate a P-value and conclude the test.

In this section, we present a method for testing two variances (or, equivalently, two standard deviations). We use *independent random samples* from two populations to test the claim that the population variances are equal. The concept of variation among data is very important, and there are many possible applications in science, industry, business administration, social science, and so on.

In Section 10.3, we tested a *single* variance. The main mathematical tool we used was the chi-square probability distribution. In this section, the main tool is the F probability distribution.

Let us begin by stating what we need to assume for a test of two population variances.

Basic requirements

- The two populations are independent of each other. Recall from Section 8.5 that two sampling distributions are *independent* if there is no relation whatsoever between specific values of the two distributions.

- The two populations each have a *normal* probability distribution. This is important because the test we will use is sensitive to changes away from normality.

Setup for test

Now that we know the basic requirements, let's consider the setup.

How to Set Up the Test

STEP 1: Get Two Independent Random Samples, One from Each Population

We use the following notation:

Population I (larger s^2)	Population II (smaller s^2)
n_1 = sample size	n_2 = sample size
s_1^2 = sample variance	s_2^2 = sample variance
σ_1^2 = population variance	σ_2^2 = population variance

To simplify later discussion, we make the notational choice that

$$s_1^2 \geq s_2^2$$

This means that we *define* population I as the population with the *larger* (or equal, as the case may be) sample variance. This is only a notational convention and does not affect the general nature of the test.

STEP 2: Set Up the Hypotheses

Hypothesis tests about two variances

The null hypothesis will be that we have equal population variances.

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

Reflecting on our notation setup, it makes sense to use an alternate hypothesis, either

$$H_1: \sigma_1^2 \neq \sigma_2^2 \quad \text{or} \quad H_1: \sigma_1^2 > \sigma_2^2$$

Notice that the test makes claims about variances. However, we can also use it for corresponding claims about standard deviations.

Hypotheses about Variances	Equivalent Hypotheses about Standard Deviations
$H_0: \sigma_1^2 = \sigma_2^2$	$H_0: \sigma_1 = \sigma_2$
$H_1: \sigma_1^2 \neq \sigma_2^2$	$H_1: \sigma_1 \neq \sigma_2$
$H_1: \sigma_1^2 > \sigma_2^2$	$H_1: \sigma_1 > \sigma_2$

STEP 3: Compute the Sample Test Statistic

Sample test statistic F

$$F = \frac{s_1^2}{s_2^2}$$

For two normally distributed populations with equal variances ($H_0: \sigma_1^2 = \sigma_2^2$), the sampling distribution we will use is the F *distribution* (see Table 8 of Appendix II).
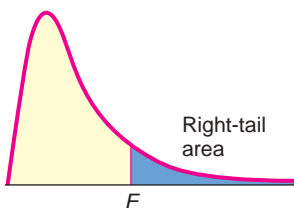
The F distribution depends on *two* degrees of freedom.

For **tests of two variances,** it can be shown that

$$d.f._N = n_1 - 1 \quad \text{and} \quad d.f._D = n_2 - 1$$

STEP 4: Find (or Estimate) the *P*-value of the Sample Test Statistic

Use the F distribution (Table 8 of Appendix II) to find the *P*-value of the sample test statistic. You need to know the degrees of freedom for the numerator, $d.f._N = n_1 - 1$, and the degrees of freedom for the denominator, $d.f._D = n_2 - 1$. Find the block of entries with your $d.f._D$ as row header and your $d.f._N$ as column header. Within that block of values, find the position of the sample test statistic F. Then find the corresponding right-tail area. For instance, using Table 10-15 (Excerpt from Table 8), we see that for $d.f._D = 2$ and $d.f._N = 3$, sample $F = 55.2$ lies between 39.17 and 99.17, with corresponding right-tail areas of 0.025 and 0.010. The interval containing the *P*-value for $F = 55.2$ is $0.010 < P\text{-value} < 0.025$.

TABLE 10-15   **Excerpt from Table 8 (Appendix II): The F Distribution**

		Degrees of Freedom for Numerator $d.f._N$			
$d.f._D$	Right-tail Area	1	2	3	4 ...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	0.100	8.53	9.00	9.16	9.24
	0.050	18.51	19.00	19.16	19.25
✓ 2	0.025	38.51	39.00	39.17	39.25
	0.010	98.50	99.00	99.17	99.25
	0.001	998.50	999.00	999.17	9999.25

Right-tail area

F

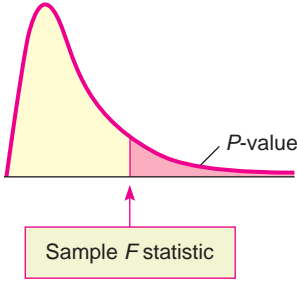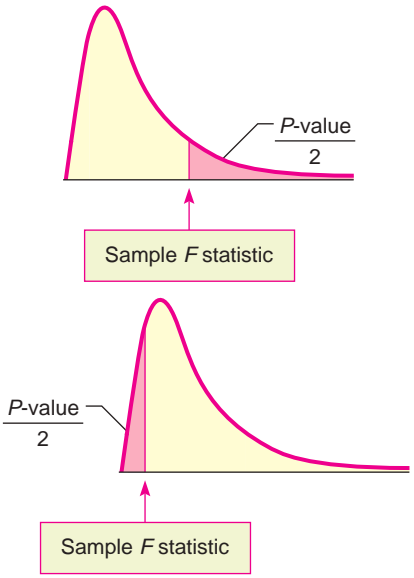| TABLE 10-16	*P*-values for Testing Two Variances (Table 8, Appendix II)

(a) Right-tailed test

Since the *F*-distribution table gives right-tail probabilities, you can use the table directly to find or estimate the *P*-value.

P-value

Sample *F* statistic

(b) Two-tailed test

Remember that the *P*-value is the probability of getting a test statistic as extreme as or more extreme than the test statistic computed from the sample. For a two-tailed test, we need to account for corresponding equal areas in *both* the upper and lower tails. This means that in each tail, we have an area of (*P*-value)/2. The total *P*-value is then

$$P\text{-value} = 2\left(\frac{P\text{-value}}{2}\right)$$

$\dfrac{P\text{-value}}{2}$

Sample *F* statistic

$\dfrac{P\text{-value}}{2}$

Sample *F* statistic

Be sure to choose the area in the appropriate tail (left or right) so that $\dfrac{P\text{-value}}{2} \leq 0.5$.

Table 10-16 gives a summary for computing the *P*-value for both right-tailed and two-tailed tests for two variances.

Now that we have steps 1 to 4 as an outline, let's look at a specific example.

EXAMPLE 7   TESTING TWO VARIANCES

Prehistoric Native Americans smoked pipes for ceremonial purposes. Most pipes were either carved-stone pipes or ceramic pipes made from clay. Clay pipes were easier to make, whereas stone pipes required careful drilling using hollow-core-bone drills and special stone reamers. An anthropologist claims that because clay pipes were easier to make, they show a greater variance in their construction. We want to test this claim using a 5% level of significance. Data for this example are taken from the Wind Mountain Archaeological Region (Source: *Mimbres Mogollon Archaeology* by A. I. Woosley and A. J. McIntyre, University of New Mexico Press). Assume the diameters of each type of pipe are normally distributed.

Ceramic Pipe Bowl Diameters (cm)

1.7	5.1	1.4	0.7	2.5	4.0
3.8	2.0	3.1	5.0	1.5	

Stone Pipe Bowl Diameters (cm)

1.6	2.1	3.1	1.4	2.2	2.1
2.6	3.2	3.4			

SOLUTION:

(a) *Check requirements* Assume that the pipe bowl diameters follow normal distributions and that the given data make up independent random samples of pipe measurements taken from archaeological excavations at Wind Mountain.

Use a calculator to verify the following:

Population I: Ceramic Pipes	Population II: Stone Pipes
$n_1 = 11$	$n_2 = 9$
$s_1^2 \approx 2.266$	$s_2^2 \approx 0.504$
$\sigma_1^2 =$ population variance	$\sigma_2^2 =$ population variance

Note: Because the sample variance for ceramic pipes (2.266) is larger than the sample variance for stone pipes (0.504), we designate population I as ceramic pipes.

(b) Set up the null and alternate hypotheses.

$$H_0: \sigma_1^2 = \sigma_2^2 \quad \text{(or the equivalent, } \sigma_1 = \sigma_2)$$
$$H_1: \sigma_1^2 > \sigma_2^2 \quad \text{(or the equivalent, } \sigma_1 > \sigma_2)$$

The null hypothesis states that there is no difference. The alternate hypothesis supports the anthropologist's claim that clay pipes have a larger variance.

(c) The sample test statistic is

$$F = \frac{s_1^2}{s_2^2} \approx \frac{2.266}{0.504} \approx 4.496$$

Now, if $\sigma_1^2 = \sigma_2^2$, then s_1^2 and s_2^2 also should be close in value. If this were the case, $F = s_1^2/s_2^2 \approx 1$. However, if $\sigma_1^2 > \sigma_2^2$, then we see that the sample statistic $F = s_1^2/s_2^2$ should be larger than 1.

(d) Find an interval containing the *P*-value for $F = 4.496$.

This is a right-tailed test (see Figure 10-14) with degrees of freedom

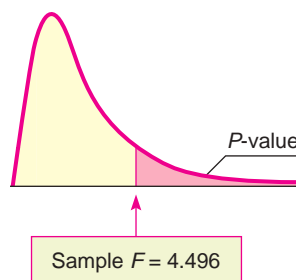$$d.f._N = n_1 - 1 = 11 - 1 = 10 \quad \text{and} \quad d.f._D = n_2 - 1 = 9 - 1 = 8$$

The interval containing the *P*-value is

$$0.010 < P\text{-value} < 0.025$$

FIGURE 10-14
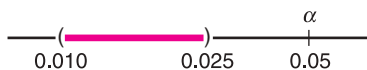
P-value

Right-tail Area		d.f.$_N$10
	0.100	2.54
d.f.$_D$	0.050	3.35
✓ 8	0.025	4.30
	0.010	5.81
	0.001	11.54

Sample $F = 4.496$

(e) Conclude the test and *interpret* the results.

Since the *P*-value is less than $\alpha = 0.05$, we reject H_0.

Technology gives *P*-value ≈ 0.0218. At the 5% level of significance, the evidence is sufficient to conclude that the variance for the ceramic pipes is larger.

We summarize the steps involved in testing two variances with the following procedure.

PROCEDURE

HOW TO TEST TWO VARIANCES σ_1^2 AND σ_2^2

Requirements

Assume that x_1 and x_2 are random variables that have *independent normal distributions* with unknown variances σ_1^2 and σ_2^2. Next, you need independent random samples of x_1 values and x_2 values, from which you compute sample variances s_1^2 and s_2^2. Use samples of sizes n_1 and n_2, respectively, with both samples of size at least 2. Without loss of generality, we may assume the notational setup is such that $s_1^2 \geq s_2^2$.

Procedure

1. Set the *level of significance* α. Use the *null hypothesis* H_0: $\sigma_1^2 = \sigma_2^2$. In the context of the problem, choose the *alternate hypothesis* to be H_1: $\sigma_1^2 > \sigma_2^2$ or H_1: $\sigma_1^2 \neq \sigma_2^2$.

2. Compute the *sample test statistic*

$$F = \frac{s_1^2}{s_2^2}$$

   where $d.f._N = n_1 - 1$ (degrees of freedom numerator)

   $d.f._D = n_2 - 1$ (degrees of freedom denominator)

3. Use the *F* distribution and the type of test to find or estimate the *P-value*. Use Table 8 of Appendix II and the procedure shown in Table 10-16.

4. *Conclude* the test. If *P*-value $\leq \alpha$, then reject H_0. If *P*-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

GUIDED EXERCISE 10 | **Testing two variances**

A large variance in blood chemistry components can result in health problems as the body attempts to return to equilibrium. J. B. O'Sullivan and C. M. Mahan conducted a study reported in the *American Journal of Clinical Nutrition* (Vol. 19, pp. 345–351) that concerned the glucose (blood sugar) levels of pregnant and nonpregnant women at Boston City Hospital. For both groups, a fasting (12-hour fast) blood glucose test was done. The following data are in units of milligrams of glucose per 100 milliliters of blood. Assume blood glucose levels for each group are normally distributed.

Continued

GUIDED EXERCISE 10   *continued*

Glucose Test: Nonpregnant Women

| 73 | 61 | 104 | 75 | 85 | 65 | 62 | 98 | 92 | 106 |

Glucose Test: Pregnant Women

| 72 | 84 | 90 | 95 | 66 | 70 | 79 | 85 |

Medical researchers question if the variance of the glucose test results for nonpregnant women is *different* (either way) from the variance for pregnant women. Let's conduct a test using a 5% level of significance.

(a) *Check requirements* What assumptions must be made about the two populations and the samples?

→ The population measurements must follow independent normal distributions. The samples must be random samples from each population.

(b) Use a calculator to compute the sample variance for each data group. Then complete the following:

→ Recall that we choose our notation so that population I has the *larger* sample variance.

Population I	Population II
$n_1 =$ _____	$n^2 =$ _____
$s_1^2 =$ _____	$s_2^2 =$ _____

Population I	Population II
$n_1 = 10$	$n_2 = 8$
$s_1^2 \approx 298.25$	$s_2^2 \approx 103.84$

(c) What is α? State the null and alternate hypotheses.

→ $\alpha = 0.05$; H_0: $\sigma_1^2 = \sigma_2^2$; H_1: $\sigma_1^2 \neq \sigma_2^2$

(d) Compute the sample test statistic F, $d.f._N$, and $d.f._D$.

→ $F = \dfrac{s_1^2}{s_2^2} \approx \dfrac{298.25}{103.84} \approx 2.87$

$d.f._N = n_1 - 1 = 10 - 1 = 9$

$d.f._D = n_2 - 1 = 8 - 1 = 7$

(e) Estimate the *P*-value.

→ Because this is a two-tailed test, we look up the area to the right of $F = 2.87$ and double it.

FIGURE 10-15 *P-value*

Sample *F* = 2.87
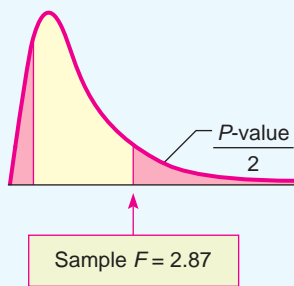
	Right-tail Area	$d.f._N$ 9
	0.100	2.72
$d.f._D$	0.050	3.68
✓ 7	0.025	4.82
	0.010	6.72
	0.001	14.33

$0.050 < \dfrac{P\text{-value}}{2} < 0.100$

$0.100 < P\text{-value} < 0.200$

(f) Conclude the test.

→ Since the *P*-value is greater than $\alpha = 0.05$, we do not reject H_0.
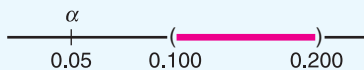
Technology gives *P*-value ≈ 0.1780.

(g) *Interpret* the results.

→ At the 5% level of significance, the evidence is insufficient to reject the claim of equal variances.

TECH NOTES   The TI-84Plus/TI-83Plus/TI-*n*spire calculators support tests of two variances. Use **STAT**, **TESTS**, and option **D:2-SampFTest**. For results consistent with the notational convention that the larger variance goes in the numerator of the sample F statistic, put the data with the larger variance in **List1**.

Minitab Release 15 of Minitab supports tests of two variances. Use menu choices **Stat ➤ *Basic Statistics* ➤ 2 Variances.**

Variety is said to be the spice of life. However, in statistics, when we want to compare two populations, we will often need the assumption that the population variances are the same. As long as the two populations follow normal distributions, we can use the methods of this section and random samples from the populations to determine if the assumption of equal variances is reasonable at a given level of significance.

VIEWPOINT | Mercury in Bass?

Largemouth bass were studied in 53 different lakes to examine factors that influence the level of mercury contamination. In many cases, the contamination was fairly low, except for older (trophy) fish, in which the mercury levels were much higher. Using information you have learned in this section, you can test variances of mercury contamination for different lakes and/or different regions. For more information, see the article by Lange, Royals, and Connor, "Transactions of the American Fisheries Society." To find this article, visit the Brase/Brase statistics site at **http://www.cengage.com/statistics/brase** *and find a link to DASL, the Carnegie Mellon University Data and Story Library. From the DASL site, link to Nature under Data Subjects, and select Mercury in the Bass Data file.*

SECTION 10.4 PROBLEMS

1. *Statistical Literacy* When using the F distribution to test variances from two populations, should the random variables from each population be independent or dependent?

2. *Statistical Literacy* When using the F distribution to test two variances, is it essential that each of the two populations be normally distributed? Would it be all right if the populations had distributions that were mound-shaped and more or less symmetrical?

3. *Statistical Literacy* In general, is the F distribution symmetrical? Can values of F be negative?

4. *Statistical Literacy* To use the F distribution, what degrees of freedom need to be calculated?

For Problems 5–12, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Find the value of the sample F statistic. What are the degrees of freedom? What assumptions are you making about the original distribution?
(c) Find or estimate the P-value of the sample test statistic.
(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis?
(e) *Interpret* your conclusion in the context of the application.

Assume that the data values in each problem come from independent populations and that each population follows a normal distribution.

5. *Agriculture: Wheat* Rothamsted Experimental Station (England) has studied wheat production since 1852. Each year, many small plots of equal size but different soil/fertilizer conditions are planted with wheat. At the end of the growing season, the yield (in pounds) of the wheat on the plot is measured. The following data are based on information taken from an article by G. A. Wiebe in the *Journal of Agricultural Research* (Vol. 50, pp. 331–357). For a random sample of years, one plot gave the following annual wheat production (in pounds):

4.15	4.21	4.27	3.55	3.50	3.79	4.09	4.42
3.89	3.87	4.12	3.09	4.86	2.90	5.01	3.39

Use a calculator to verify that, for this plot, the sample variance is $s^2 \approx 0.332$.

Another random sample of years for a second plot gave the following annual wheat production (in pounds):

4.03	3.77	3.49	3.76	3.61	3.72	4.13	4.01
3.59	4.29	3.78	3.19	3.84	3.91	3.66	4.35

Use a calculator to verify that the sample variance for this plot is $s^2 \approx 0.089$.

Test the claim that the population variance of annual wheat production for the first plot is larger than that for the second plot. Use a 1% level of significance.

6. *Agriculture: Wheat* Two plots at Rothamsted Experimental Station (see reference in Problem 5) were studied for production of wheat straw. For a random sample of years, the annual wheat straw production (in pounds) from one plot was as follows:

6.17	6.05	5.89	5.94	7.31	7.18
7.06	5.79	6.24	5.91	6.14	

Use a calculator to verify that, for the preceding data, $s^2 \approx 0.318$.

Another random sample of years for a second plot gave the following annual wheat straw production (in pounds):

6.85	7.71	8.23	6.01	7.22	5.58	5.47	5.86

Use a calculator to verify that, for these data, $s^2 \approx 1.078$.

Test the claim that there is a difference (either way) in the population variance of wheat straw production for these two plots. Use a 5% level of significance.

7. *Economics: Productivity* An economist wonders if corporate productivity in some countries is more *volatile* than that in other countries. One measure of a company's productivity is annual percentage yield based on total company assets. Data for this problem are based on information taken from *Forbes Top Companies*, edited by J. T. Davis. A random sample of leading companies in France gave the following percentage yields based on assets:

4.4	5.2	3.7	3.1	2.5	3.5	2.8	4.4	5.7	3.4	4.1
6.8	2.9	3.2	7.2	6.5	5.0	3.3	2.8	2.5	4.5	

Use a calculator to verify that $s^2 \approx 2.044$ for this sample of French companies.

Another random sample of leading companies in Germany gave the following percentage yields based on assets:

3.0	3.6	3.7	4.5	5.1	5.5	5.0	5.4	3.2
3.5	3.7	2.6	2.8	3.0	3.0	2.2	4.7	3.2

Use a calculator to verify that $s^2 \approx 1.038$ for this sample of German companies.

Test the claim that there is a difference (either way) in the population variance of percentage yields for leading companies in France and Germany. Use a 5% level of significance. How could your test conclusion relate to the economist's question regarding *volatility* (data spread) of corporate productivity of large companies in France compared with large companies in Germany?

8. *Economics: Productivity* A random sample of leading companies in South Korea gave the following percentage yields based on assets (see reference in Problem 7):

2.5	2.0	4.5	1.8	0.5	3.6	2.4
0.2	1.7	1.8	1.4	5.4	1.1	

Use a calculator to verify that $s^2 = 2.247$ for these South Korean companies.

Another random sample of leading companies in Sweden gave the following percentage yields based on assets:

2.3	3.2	3.6	1.2	3.6	2.8	2.3	3.5	2.8

Use a calculator to verify that $s^2 = 0.624$ for these Swedish companies.

Test the claim that the population variance of percentage yields on assets for South Korean companies is higher than that for companies in Sweden. Use a 5% level of significance. How could your test conclusion relate to an economist's question regarding *volatility* of corporate productivity of large companies in South Korea compared with that in Sweden?

9. *Investing: Mutual Funds* You don't need to be rich to buy a few shares in a mutual fund. The question is, "How *reliable* are mutual funds as investments?" That depends on the type of fund you buy. The following data are based on information taken from *Morningstar*, a mutual fund guide available in most libraries. A random sample of percentage annual returns for mutual funds holding stocks in aggressive-growth small companies is shown below.

−1.8	14.3	41.5	17.2	−16.8	4.4	32.6	−7.3	16.2	2.8	34.3
−10.6	8.4	−7.0	−2.3	−18.5	25.0	−9.8	−7.8	−24.6	22.8	

Use a calculator to verify that $s^2 \approx 348.43$ for the sample of aggressive-growth small company funds.

Another random sample of percentage annual returns for mutual funds holding value (i.e., market underpriced) stocks in large companies is shown below.

16.2	0.3	7.8	−1.6	−3.8	19.4	−2.5	15.9	32.6	22.1	3.4
−0.5	−8.3	25.8	−4.1	14.6	6.5	18.0	21.0	0.2	−1.6	

Use a calculator to verify that $s^2 \approx 137.31$ for value stocks in large companies.

Test the claim that the population variance for mutual funds holding aggressive-growth small stocks is larger than the population variance for mutual funds holding value stocks in large companies. Use a 5% level of significance. How could your test conclusion relate to the question of *reliability* of returns for each type of mutual fund?

10. *Investing: Mutual Funds* How *reliable* are mutual funds that invest in bonds? Again, this depends on the bond fund you buy (see reference in Problem 9). A random sample of annual percentage returns for mutual funds holding short-term U.S. government bonds is shown below.

4.6	4.7	1.9	9.3	−0.8	4.1	10.5
4.2	3.5	3.9	9.8	−1.2	7.3	

Use a calculator to verify that $s^2 \approx 13.59$ for the preceding data.

A random sample of annual percentage returns for mutual funds holding intermediate-term corporate bonds is shown below.

−0.8	3.6	20.2	7.8	−0.4	18.8	−3.4	10.5
8.0	−0.9	2.6	−6.5	14.9	8.2	18.8	14.2

Use a calculator to verify that $s^2 \approx 72.06$ for returns from mutual funds holding intermediate-term corporate bonds.

Use $\alpha = 0.05$ to test the claim that the population variance for annual percentage returns of mutual funds holding short-term government bonds is different from the population variance for mutual funds holding intermediate-term corporate bonds. How could your test conclusion relate to the question of *reliability* of returns for each type of mutual fund?

11. *Engineering: Fuel Injection* A new fuel injection system has been engineered for pickup trucks. The new system and the old system both produce about the same average miles per gallon. However, engineers question which system (old or new) will give better *consistency* in fuel consumption (miles per gallon) under a variety of driving conditions. A random sample of 31 trucks were fitted with the new fuel injection system and driven under different conditions. For these trucks, the sample variance of gasoline consumption was 58.4. Another random sample of 25 trucks were fitted with the old fuel injection system and driven under a variety of different conditions. For these trucks, the sample variance of gasoline consumption was 31.6. Test the claim that there is a difference in population variance of gasoline consumption for the two injection systems. Use a 5% level of significance. How could your test conclusion relate to the question regarding the *consistency* of fuel consumption for the two fuel injection systems?

12. *Engineering: Thermostats* A new thermostat has been engineered for the frozen food cases in large supermarkets. Both the old and the new thermostats hold temperatures at an average of 25°F. However, it is hoped that the new thermostat might be more *dependable* in the sense that it will hold temperatures closer to 25°F. One frozen food case was equipped with the new thermostat, and a random sample of 21 temperature readings gave a sample variance of 5.1. Another, similar frozen food case was equipped with the old thermostat, and a random sample of 16 temperature readings gave a sample variance of 12.8. Test the claim that the population variance of the old thermostat temperature readings is larger than that for the new thermostat. Use a 5% level of significance. How could your test conclusion relate to the question regarding the *dependability* of the temperature readings?

SECTION 10.5

One-Way ANOVA: Comparing Several Sample Means

FOCUS POINTS

- Learn about the risk α of a type I error when we test several means at once.
- Learn about the notation and setup for a one-way ANOVA test.
- Compute mean squares between groups and within groups.
- Compute the sample *F* statistic.
- Use the *F* distribution to estimate a *P*-value and conclude the test.

In our past work, to determine the existence (or nonexistence) of a significant difference between population means, we restricted our attention to only two data groups representing the means in question. Many statistical applications in psychology, social science, business administration, and natural science involve many means and many data groups. Questions commonly asked are: Which of *several* alternative methods yields the best results in a particular setting? Which of *several* treatments leads to the highest incidence of patient recovery? Which of *several* teaching methods leads to greatest student retention? Which of *several* investment schemes leads to greatest economic gain?

Using our previous methods (Sections 8.4 and 8.5) of comparing only *two* means would require many tests of significance to answer the preceding questions. For example, even if we had only 5 variables, we would be required to

perform 10 tests of significance in order to compare each variable to each of the other variables. If we had the time and patience, we could perform all 10 tests, but what about the risk of accepting a difference where there really is no difference (a type I error)? If the risk of a type I error on each test is $\alpha = 0.05$, then on 10 tests we expect the number of tests with a type I error to be 10(0.05), or 0.5 (see expected value, Section 5.3). This situation may not seem too serious to you, but remember that in a "real-world" problem and with the aid of a high-speed computer, a researcher may want to study the effect of 50 variables on the outcome of an experiment. Using a little mathematics, we can show that the study would require 1225 separate tests to check *each pair* of variables for a significant difference of means. At the $\alpha = 0.05$ level of significance for each test, we could expect (1225)(0.05), or 61.25, of the tests to have a type I error. In other words, these 61.25 tests would say that there are differences between means when there really are no differences.

Introduction

To avoid such problems, statisticians have developed a method called *analysis of variance* (abbreviated *ANOVA*). We will study single-factor analysis of variance (also called *one-way ANOVA*) in this section and two-way ANOVA in Section 10.6. With appropriate modification, methods of single-factor ANOVA generalize to *n*-dimensional ANOVA, but we leave that topic to more advanced studies.

ANOVA

EXAMPLE 8   ONE-WAY ANOVA TEST

A psychologist is studying the effects of dream deprivation on a person's anxiety level during waking hours. Brain waves, heart rate, and eye movements can be used to determine if a sleeping person is about to enter into a dream period. Three groups of subjects were randomly chosen from a large group of college students who volunteered to participate in the study. Group I subjects had their sleep interrupted four times each night, but never during or immediately before a dream. Group II subjects had their sleep interrupted four times also, but on two occasions they were wakened at the onset of a dream. Group III subjects were wakened four times, each time at the onset of a dream. This procedure was repeated for 10 nights, and each day all subjects were given a test to determine their levels of anxiety. The data in Table 10-17 record the total of the test scores for each person over the entire project. Higher totals mean higher anxiety levels.

Explanation of notation and formulas

TABLE 10-17   Dream Deprivation Study

Group I $n_1 = 6$ Subjects		Group II $n_2 = 7$ Subjects		Group III $n_3 = 5$ Subjects	
x_1	x_1^2	x_2	x_2^2	x_3	x_3^2
9	81	10	100	15	225
7	49	9	81	11	121
3	9	11	121	12	144
6	36	10	100	9	81
5	25	7	49	10	100
8	64	6	36		
		8	64		
$\Sigma x_1 = 38$	$\Sigma x_1^2 = 264$	$\Sigma x_2 = 61$	$\Sigma x_2^2 = 551$	$\Sigma x_3 = 57$	$\Sigma x_3^2 = 671$

$N = n_1 + n_2 + n_3 = 18$

$\Sigma x_{TOT} = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 = 156$

$\Sigma x_{TOT}^2 = \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 = 1486$

From Table 10-17, we see that group I had $n_1 = 6$ subjects, group II had $n_2 = 7$ subjects, and group III had $n_3 = 5$ subjects. For each subject, the anxiety score (x value) and the square of the test score (x^2 value) are also shown. In addition, special sums are shown.

We will outline the procedure for single-factor ANOVA in six steps. Each step will contain general methods and rationale appropriate to all single-factor ANOVA tests. As we proceed, we will use the data of Table 10-17 for a specific reference example.

Basic requirements of ANOVA

Our application of ANOVA has three basic requirements. In a general problem with k groups:

1. We require that each of our k groups of measurements is obtained from a population with a *normal* distribution.

2. Each group is randomly selected and is *independent* of all other groups. In particular, this means that we will not use the same subjects in more than one group and that the scores of one subject will not have an effect on the scores of another subject.

3. We assume that the variables from each group come from distributions with approximately the *same standard deviation.*

STEP 1: Determine the Null and Alternate Hypotheses

The purpose of an ANOVA test is to determine the existence (or nonexistence) of a statistically significant difference *among* the group means. In a general problem with k groups, we call the (population) mean of the first group μ_1, the population mean of the second group μ_2, and so forth. The null hypothesis is simply that *all* the group population means are the same. Since our basic requirements state that each of the k groups of measurements comes from normal, independent distributions with common standard deviation, the null hypothesis states that all the sample groups come from *one and the same* population. The alternate hypothesis is that *not all* the group population means are equal. Therefore, in a problem with k groups, we have

Hypotheses

Hypotheses for one-way ANOVA

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$

$H_1:$ At least two of the means $\mu_1, \mu_2, \ldots, \mu_k$ are not equal.

Notice that the alternate hypothesis claims that *at least* two of the means are not equal. If more than two of the means are unequal, the alternate hypothesis is, of course, satisfied.

In our dream problem, we have $k = 3$; μ_1 is the population mean of group I, μ_2 is the population mean of group II, and μ_3 is the population mean of group III. Therefore,

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

$H_1:$ At least two of the means μ_1, μ_2, μ_3 are not equal.

We will test the null hypothesis using an $\alpha = 0.05$ level of significance. Notice that only one test is being performed even though we have $k = 3$ groups and three corresponding means. Using ANOVA avoids the problem mentioned earlier of using multiple tests.

Sum of squares, SS

The concept of *sum of squares* is very important in statistics. We used a sum of squares in Chapter 3 to compute the sample standard deviation and sample variance.

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \quad \text{sample standard deviation}$$

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \quad \text{sample variance}$$

The numerator of the sample variance is a special sum of squares that plays a central role in ANOVA. Since this numerator is so important, we give it the special name SS (for "sum of squares").

$$SS = \Sigma(x - \bar{x})^2 \tag{2}$$

Using some college algebra, it can be shown that the following, simpler formula is equivalent to Equation (2) and involves fewer calculations:

$$SS = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \tag{3}$$

where n is the sample size.

In future references to SS, we will use Equation (3) because it is easier to use than Equation (2).

The **total sum of squares SS_{TOT}** can be found by using the entire collection of all data values in all groups:

$$SS_{TOT} = \Sigma x_{TOT}^2 - \frac{(\Sigma x_{TOT})^2}{N} \tag{4}$$

where $N = n_1 + n_2 + \cdots + n_k$ is the total sample size from all groups.

$$\Sigma x_{TOT} = \text{sum of all data} = \Sigma x_1 + \Sigma x_2 + \cdots + \Sigma x_k$$
$$\Sigma x_{TOT}^2 = \text{sum of all data squares} = \Sigma x_1^2 + \cdots + \Sigma x_k^2$$

Using the specific data given in Table 10-17 for the dream example, we have

$$k = 3 \quad \text{total number of groups}$$
$$N = n_1 + n_2 + n_3 = 6 + 7 + 5 = 18 \quad \text{total number of subjects}$$
$$\Sigma x_{TOT} = \text{total sum of } x \text{ values} = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 = 38 + 61 + 57 = 156$$
$$\Sigma x_{TOT}^2 = \text{total sum of } x^2 \text{ values} = \Sigma x_1^2 + \Sigma x_2^2 + \Sigma x_3^2 = 264 + 551 + 671 = 1486$$

Therefore, using Equation (4), we have

$$SS_{TOT} = \Sigma x_{TOT}^2 - \frac{(\Sigma x_{TOT})^2}{N} = 1486 - \frac{(156)^2}{18} = 134$$

The numerator for the total variation for all groups in our dream example is $SS_{TOT} = 134$. What interpretation can we give to SS_{TOT}? If we let $\bar{x}_{TOT}$ be the mean of all x values for all groups, then

$$\text{Mean of all } x \text{ values} = \bar{x}_{TOT} = \frac{\Sigma x_{TOT}}{N}$$

Under the null hypothesis (that all groups come from the same normal distribution), $SS_{TOT} = \Sigma(x_{TOT} - \bar{x}_{TOT})^2$ represents the numerator of the sample variance

for all groups. Therefore, SS_{TOT} represents total variability of the data. Total variability can occur in two ways:

SS_{BET}

1. Scores may differ from one another because they belong to *different groups* with different means (recall that the alternate hypothesis states that the means are not all equal). This difference is called **between-group variability** and is denoted SS_{BET}.

SS_W

2. Inherent differences unique to each subject and differences due to chance may cause a particular score to be different from the mean of its *own group*. This difference is called **within-group variability** and is denoted SS_W.

Because total variability SS_{TOT} is the sum of between-group variability SS_{BET} and within-group variability SS_W, we may write

$$SS_{TOT} = SS_{BET} + SS_W$$

As we will see, SS_{BET} and SS_W are going to help us decide whether or not to reject the null hypothesis. Therefore, our next two steps are to compute these two quantities.

STEP 3: Find SS_{BET}

Sum of squares between groups

Recall that $\bar{x}_{TOT}$ is the mean of all x values from all groups. Between-group variability (SS_{BET}) measures the variability of group means. Because different groups may have different numbers of subjects, we must "weight" the variability contribution from each group by the group size n_i.

$$SS_{BET} = \sum_{\text{all groups}} n_i(\bar{x}_i - \bar{x}_{TOT})^2$$

where $n_i =$ sample size of group i
$\quad\quad \bar{x}_i =$ sample mean of group i
$\quad\quad \bar{x}_{TOT} =$ mean for values from all group

If we use algebraic manipulations, we can write the formula for SS_{BET} in the following computationally easier form:

Sum of squares between groups

$$SS_{BET} = \sum_{\text{all groups}} \left(\frac{(\Sigma x_i)^2}{n_i} \right) - \frac{(\Sigma x_{TOT})^2}{N} \tag{5}$$

where, as before, $N = n_1 + n_2 + \cdots + n_k$
$\quad\quad\quad\quad\quad \Sigma x_i =$ sum of data in group i
$\quad\quad\quad\quad\quad \Sigma x_{TOT} =$ sum of data from all groups

Using data from Table 10-17 for the dream example, we have

$$
\begin{aligned}
SS_{BET} &= \sum_{\text{all groups}} \left(\frac{(\Sigma x_i)^2}{n_i} \right) - \frac{(\Sigma x_{TOT})^2}{N} \\
&= \frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} - \frac{(\Sigma x_{TOT})^2}{N} \\
&= \frac{(38)^2}{6} + \frac{(61)^2}{7} + \frac{(57)^2}{5} - \frac{(156)^2}{18} \\
&= 70.038
\end{aligned}
$$

Therefore, the numerator of the between-group variation is

$$SS_{BET} = 70.038$$

STEP 4: Find SS_W

Sum of squares within groups

We could find the value of SS_W by using the formula relating SS_{TOT} to SS_{BET} and SS_W and solving for SS_W:

$$SS_W = SS_{TOT} - SS_{BET}$$

However, we prefer to compute SS_W in a different way and to use the preceding formula as a check on our calculations.

SS_W is the numerator of the variation within groups. Inherent differences unique to each subject and differences due to chance create the variability assigned to SS_W. In a general problem with k groups, the variability within the ith group can be represented by

$$SS_i = \Sigma(x_i - \bar{x}_i)^2$$

or by the mathematically equivalent formula

$$SS_i = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n_i} \tag{6}$$

Because SS_i represents the variation within the ith group and we are seeking SS_W, the variability within *all* groups, we simply add SS_i for all groups:

Sum of squares within groups

$$SS_W = SS_1 + SS_2 + \cdots + SS_k \tag{7}$$

Using Equations (6) and (7) and the data of Table 10-17 with $k = 3$, we have

$$SS_1 = \Sigma x_1^2 - \frac{(\Sigma x_1)^2}{n_1} = 264 - \frac{(38)^2}{6} = 23.333$$

$$SS_2 = \Sigma x_2^2 - \frac{(\Sigma x_2)^2}{n_2} = 551 - \frac{(61)^2}{7} = 19.429$$

$$SS_3 = \Sigma x_3^2 - \frac{(\Sigma x_3)^2}{n_3} = 671 - \frac{(57)^2}{5} = 21.200$$

$$SS_W = SS_1 + SS_2 + SS_3 = 23.333 + 19.429 + 21.200 = 63.962$$

Let us check our calculation by using SS_{TOT} and SS_{BET}.

$$SS_{TOT} = SS_{BET} + SS_W$$
$$134 = 70.038 + 63.962 \quad \text{(from steps 2 and 3)}$$

We see that our calculation checks.

STEP 5: Find Variance Estimates (Mean Squares)

Mean squares

In steps 3 and 4, we found SS_{BET} and SS_W. Although these quantities represent variability between groups and within groups, they are not yet the variance estimates we need for our ANOVA test. You may recall our study of the Student's t distribution, in which we introduced the concept of degrees of freedom. Degrees of freedom represent the number of values that are free to vary once we have placed certain restrictions on our data. In ANOVA, there are two types of degrees

of freedom: $d.f._{BET}$, representing the degrees of freedom between groups, and $d.f._W$, representing degrees of freedom within groups. A theoretical discussion beyond the scope of this text would show

---

Degrees of freedom between and within groups

$$d.f._{BET} = k - 1 \qquad \text{where } k \text{ is the number of groups}$$
$$d.f._W = N - k \qquad \text{where } N \text{ is the total sample size}$$

(Note: $d.f._{BET} + d.f._W = N - 1$.)

The variance estimates we are looking for are designated as follows:

MS_{BET}

MS_W

MS_{BET}, the variance between groups (read "mean square between")

MS_W, the variance within groups (read "mean square within")

In the literature of ANOVA, the variances between and within groups are usually referred to as *mean squares between* and *within* groups, respectively. We will use the mean-square notation because it is used so commonly. However, remember that the notations MS_{BET} and MS_W both refer to *variances,* and you might occasionally see the variance notations S^2_{BET} and S^2_W used for these quantities. The formulas for the variances between and within samples follow the pattern of the basic formula for sample variance.

$$\text{Sample variance} = s^2 = \frac{\Sigma(x - \overline{x})^2}{n - 1} = \frac{SS}{n - 1}$$

However, instead of using $n - 1$ in the denominator for MS_{BET} and MS_W variances, we use their respective degrees of freedom.

$$\text{Mean square between} = MS_{BET} = \frac{SS_{BET}}{d.f._{BET}} = \frac{SS_{BET}}{k - 1}$$

$$\text{Mean square within} = MS_W = \frac{SS_W}{d.f._W} = \frac{SS_W}{N - k}$$

Using these two formulas and the data of Table 10-17, we find the mean squares within and between variances for the dream deprivation example:

$$MS_{BET} = \frac{SS_{BET}}{k - 1} = \frac{70.038}{3 - 1} = 35.019$$

$$MS_W = \frac{SS_W}{N - k} = \frac{63.962}{18 - 3} = 4.264$$

STEP 6: Find the *F* Ratio and Complete the ANOVA Test

The logic of our ANOVA test rests on the fact that one of the variances, MS_{BET}, *can* be influenced by population differences among means of the several groups, whereas the other variance, MS_W, *cannot* be so influenced. For instance, in the dream deprivation and anxiety study, the variance between groups MS_{BET} will be affected if any of the treatment groups has a population mean anxiety score that is *different* from that of any other group. On the other hand, the variance within groups MS_W compares anxiety scores of each treatment group to its own group anxiety mean, and the fact that group means might differ *does not* affect the MS_W value.

Recall that the null hypothesis claims that all the groups are samples from populations having the *same* (normal) distributions. The alternate hypothesis states that at least two of the sample groups come from populations with *different* (normal) distributions.

If the *null* hypothesis is *true*, MS_{BET} and MS_W should both estimate the *same* quantity. Therefore, if H_0 is true, the *F ratio*

F ratio

Sample *F* statistic

> **Sample *F* ratio**
>
> $$F = \frac{MS_{BET}}{MS_W}$$
>
> The *F* ratio is the *sample test statistic F* for ANOVA tests.

should be approximately 1, and variations away from 1 should occur only because of sampling errors. The variance within groups MS_W is a good estimate of the overall population variance, but the variance between groups MS_{BET} consists of the population variance *plus* an additional variance stemming from the differences between samples. Therefore, if the *null* hypothesis is *false*, MS_{BET} will be larger than MS_W, and the *F* ratio will tend to be *larger* than 1.

The decision of whether or not to reject the null hypothesis is determined by the relative size of the *F* ratio. Table 8 of Appendix II gives *F* values.

For our example about dreams, the computed *F* ratio is

$$F = \frac{MS_{BET}}{MS_W} = \frac{35.019}{4.264} = 8.213$$

ANOVA test is right-tailed

Because large *F* values tend to discredit the null hypothesis, we use a *right-tailed test* with the *F distribution.* To find (or estimate) the *P*-value for the sample *F* statistic, we use the *F*-distribution table, Table 8 of Appendix II. The table requires us to know *degrees of freedom for the numerator* and *degrees of freedom for the denominator.*

Degrees of freedom for sample test statistic *F*

> **Degrees of freedom for sample test statistic *F* in one-way ANOVA**
>
> Degrees of freedom numerator $= d.f._N = d.f._{BET} = k - 1$
>
> Degrees of freedom denominator $= d.f._D = d.f._W = N - k$
>
> where $k =$ number of groups
>
> $N =$ total sample size across all groups

For our example about dreams,

$$d.f._N = k - 1 = 3 - 1 = 2 \qquad d.f._D = N - k = 18 - 3 = 15$$

Finding the *P*-value

Let's use the *F*-distribution table (Table 8, Appendix II) to find the *P*-value of the sample statistic $F = 8.213$. The *P*-value is a *right-tail area,* as shown in Figure 10-16. In Table 8, look in the block headed by column $d.f._N = 2$ and row $d.f._D = 15$. For convenience, the entries are shown in Table 10-18 (Excerpt from Table 8). We see that the sample $F = 8.213$ falls between the entries 6.36 and 11.34, with corresponding right-tail areas 0.010 and 0.001.

Test conclusion

The *P*-value is in the interval $0.001 < P\text{-value} < 0.010$. Since $\alpha = 0.05$, we see that the *P*-value is less than α and we reject H_0.
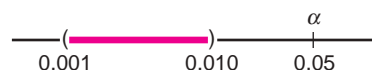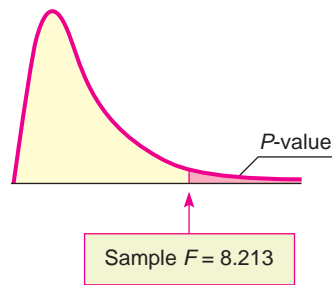
P-value

	Right-tail Area	$d.f._N$ 2
	0.100	2.70
$d.f._D$	0.050	3.68
✓ 15	0.025	4.77
	0.010	6.36
	0.001	11.34

TABLE 10-18   **Excerpt from Table 8, Appendix II**

Sample $F = 8.213$

At the 5% level of significance, we reject H_0 and conclude that not all the means are equal. The amount of dream deprivation *does* make a difference in mean anxiety level. *Note:* Technology gives *P*-value ≈ 0.0039.

This completes our single-factor ANOVA test. Before we consider another example, let's summarize the main points.

PROCEDURE

HOW TO CONSTRUCT A ONE-WAY ANOVA TEST

Requirements

You need *k* independent data groups, with each group belonging to a normal distribution and all groups having (approximately) the same standard deviation. *N* is the total number of data values across all groups.

Procedure

1. Set the *level of significance* α and the *hypotheses*

   $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$
   $H_1:$ not all of $\mu_1, \mu_2, \ldots, \mu_k$, are equal

   where μ_i is the population mean of group *i*.

2. Compute the *sample test statistic F* using the following steps or appropriate technology.

   (a) $SS_{TOT} = \Sigma x_{TOT}^2 - \dfrac{(\Sigma x_{TOT})^2}{N}$

   where Σx_{TOT} is the sum of all data elements from all groups
   Σx_{TOT}^2 is the sum of all data elements squared from all groups
   N is the total sample size

   (b) $SS_{TOT} = SS_{BET} + SS_W$

   where $SS_{BET} = \displaystyle\sum_{\text{all groups}} \left(\dfrac{(\Sigma x_i)^2}{n_i}\right) - \dfrac{(\Sigma x_{TOT})^2}{N}$

   n_i is the number of data elements in group *i*
   Σx_i is the sum of the data elements in group *i*

   $SS_W = \displaystyle\sum_{\text{all groups}} \left(\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n_i}\right)$

   Continued

(c) $MS_{BET} = \dfrac{SS_{BET}}{d.f._{BET}}$ where $d.f._{BET} = k - 1$

$MS_W = \dfrac{SS_W}{d.f._W}$ where $d.f._W = N - k$

(d) $F = \dfrac{MS_{BET}}{MS_W}$ with $d.f._N = k - 1$ and $d.f._D = N - k$

Because an ANOVA test requires a number of calculations, we recommend that you summarize your results in a table such as Table 10-19. This is the type of table that is often generated by computer software.

3. Find (or estimate) the *P-value* using the *F* distribution (Table 8, Appendix II). The test is a *right-tailed* test.

4. *Conclude* the test. If P-value $\leq \alpha$, then reject H_0. If P-value $> \alpha$, then do not reject H_0.

5. *Interpret your conclusion* in the context of the application.

Summary table for ANOVA

TABLE 10-19   Summary of ANOVA Results

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square (Variance)	F Ratio	P-value	Test Decision
Basic Model						
Between groups	SS_{BET}	$d.f._{BET}$	MS_{BET}	$\dfrac{MS_{BET}}{MS_W}$	From table	Reject H_0 or fail to reject H_0
Within groups	SS_W	$d.f._W$	MS_W			
Total	SS_{TOT}	$N - 1$				
Summary of ANOVA Results from Dream Experiment (Example 8)						
Between groups	70.038	2	35.019	8.213	< 0.010	Reject H_0
Within groups	63.962	15	4.264			
Total	134	17				

GUIDED EXERCISE 11   **One-way ANOVA test**

A psychologist is studying pattern-recognition skills under four laboratory settings. In each setting, a fourth-grade child is given a pattern-recognition test with 10 patterns to identify. In setting I, the child is given *praise* for each correct answer and no comment about wrong answers. In setting II, the child is given *criticism* for each wrong answer and no comment about correct answers. In setting III, the child is given no praise or criticism, but the observer expresses *interest* in what the child is doing. In setting IV, the observer remains *silent* in an adjacent room watching the child through a one-way mirror. A random sample of fourth-grade children was used, and each child participated in the test only once. The test scores (number correct) for each group follow. (See Table 10-20.)

(a) Fill in the missing entries of Table 10-20.

$\Sigma x_{TOT} = $ _____

$\Sigma x_{TOT} = \Sigma x_1 + \Sigma x_2 + \Sigma x_3 + \Sigma x_4$

$= 41 + 14 + 38 + 28 = 121$

Continued

GUIDED EXERCISE 11 *continued*

$\Sigma x^2_{TOT} =$ _____

$N =$ _____

$k =$ _____

$\Sigma x^2_{TOT} = \Sigma x^2_1 + \Sigma x^2_2 + \Sigma x^2_3 + \Sigma x^2_4$

$= 339 + 54 + 264 + 168 = 825$

$N = n_1 + n_2 + n_3 + n_4 = 5 + 4 + 6 + 5 = 20$

$k = 4$ groups

(b) What assumptions are we making about the data to apply a single-factor ANOVA test?

⟹ Because each of the groups comes from independent random samples (no child was tested twice), we need assume only that each group of data came from a normal distribution, and that all the groups came from distributions with about the same standard deviation.

Table 10-20 Pattern-Recognition Experiment

Group I (Praise) $n_1 = 5$		Group II (Criticism) $n_2 = 4$		Group III (Interest) $n_3 = 6$		Group IV (Silence) $n_4 = 5$	
x_1	x^2_1	x_2	x^2_2	x_3	x^2_3	x_4	x^2_4
9	81	2	4	9	81	5	25
8	64	5	25	3	9	7	49
8	64	4	16	7	49	3	9
9	81	3	9	8	64	6	36
7	49			5	25	7	49
				6	36		
$\Sigma x_1 = 41$		$\Sigma x_2 = 14$		$\Sigma x_3 = 38$		$\Sigma x_4 = 28$	
$\Sigma x^2_1 = 339$		$\Sigma x^2_2 = 54$		$\Sigma x^2_3 = 264$		$\Sigma x^2_4 = 168$	
$\Sigma x_{TOT} =$ _____		$\Sigma x^2_{TOT} =$ _____		$N =$ _____		$k =$ _____	

(c) What are the null and alternate hypotheses?

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

In words, all the groups have the same population mean, and this hypothesis, together with the basic assumptions of part (b), states that all the groups come from the same population.

H_1: not all the means μ_1, μ_2, μ_3, μ_4 are equal.

In words, not all the groups have the same population mean, so at least one group did not come from the same population as the others.

(d) Find the value of SS_{TOT}.

⟹ $SS_{TOT} = \Sigma x^2_{TOT} - \dfrac{(\Sigma x_{TOT})^2}{N} = 825 - \dfrac{(121)^2}{20}$

$= 92.950$

(e) Find SS_{BET}.

⟹ $SS_{BET} = \displaystyle\sum_{\text{all groups}} \left(\dfrac{(\Sigma x_i)^2}{n_i} \right) - \dfrac{(\Sigma x_{TOT})^2}{N}$

$= \dfrac{(41)^2}{5} + \dfrac{(14)^2}{4} + \dfrac{(38)^2}{6}$

$+ \dfrac{(28)^2}{5} - \dfrac{(121)^2}{20} = 50.617$

(f) Find SS_W and check your calculations using the formula

⟹ $SS_W = \displaystyle\sum_{\text{all groups}} \left(\Sigma x^2_i - \dfrac{(\Sigma x_i)^2}{n_i} \right)$

Continued

$$SS_{TOT} = SS_{BET} + SS_W$$

$$SS_W = SS_1 + SS_2 + SS_3 + SS_4$$

$$SS_1 = \Sigma x_1^2 - \frac{(\Sigma x_1)^2}{n_1} = 339 - \frac{(41)^2}{5} = 2.800$$

$$SS_2 = \Sigma x_2^2 - \frac{(\Sigma x_2)^2}{n_2} = 54 - \frac{(14)^2}{4} = 5.000$$

$$SS_3 = \Sigma x_3^2 - \frac{(\Sigma x_3)^2}{n_3} = 264 - \frac{(38)^2}{6} \approx 23.333$$

$$SS_4 = \Sigma x_4^2 - \frac{(\Sigma x_4)^2}{n_4} = 168 - \frac{(28)^2}{5} = 11.200$$

$$SS_W = 42.333$$

Check: $SS_{TOT} = SS_{BET} + SS_W$

$$92.950 = 50.617 + 42.333 \text{ checks}$$

(g) Find $d.f._{BET}$ and $d.f._W$.

$d.f._{BET} = k - 1 = 4 - 1 = 3$

$d.f._W = N - k = 20 - 4 = 16$

Check: $N - 1 = d.f._{BET} + d.f._W$

$20 - 1 = 3 + 16$ checks

(h) Find the mean squares MS_{BET} and MS_W.

$$MS_{BET} = \frac{SS_{BET}}{d.f._{BET}} = \frac{50.617}{3} \approx 16.872$$

$$MS_W = \frac{SS_W}{d.f._W} = \frac{42.333}{16} \approx 2.646$$

(i) Find the F ratio (sample test statistic F).

$$F = \frac{MS_{BET}}{MS_W} = \frac{16.872}{2.646} \approx 6.376$$

(j) Estimate the P-value for the sample $F = 6.376$. (Use Table 8 of Appendix II.)

$d.f._N = d.f._{BET} = k - 1 = 3$

$d.f._D = d.f._W = N - k = 16$

FIGURE 10-17 *P-value*

Sample $F = 6.376$
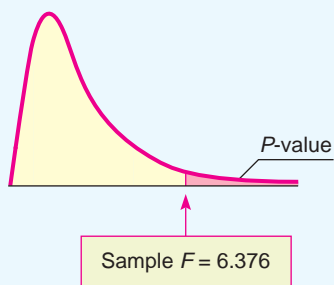
	Right-tail Area	$d.f._N$ 3
	0.100	2.46
$d.f._D$	0.050	3.24
✓ 16	0.025	4.08
	0.010	5.29
	0.001	9.01

$0.001 < P\text{-value} < 0.010$

(k) Conclude the test using a 1% level of significance. Does the test indicate that we should reject or fail to reject the null hypothesis? Explain.

$\alpha = 0.01$. Technology gives P-value ≈ 0.0048.

Since the P-value is less than 0.01, we reject H_0 and conclude that there is a significant difference in population means among the four groups. The laboratory setting *does* affect the mean scores.

(l) Make a summary table of this ANOVA test.

See Table 10-21.

Continued

TABLE 10-21   Summary of ANOVA Results for Pattern-Recognition Experiment

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square (Variance)	F Ratio	P-value	Test Decision
Between groups	50.617	3	16.872	6.376	< 0.01	Reject H_0
Within groups	42.333	16	2.646			
Total	92.950	19				

TECH NOTES   After you understand the process of ANOVA, technology tools offer valuable assistance in performing one-way ANOVA. The TI-84Plus/TI-83Plus/TI-*n*spire calculators, Excel 2007, and Minitab all support one-way ANOVA. Both the TI-84Plus/TI-83Plus/TI-*n*spire calculators and Minitab use the terminology

Factor for Between Groups

Error for Within Groups

In all the technologies, enter the data for each group in separate columns. The displays show results for Guided Exercise 11.

TI-84Plus/TI-83Plus/TI-*n*spire (with TI-84Plus keypad) Use **STAT, TESTS,** and option **F:ANOVA** and enter the lists containing the data.

```
One-way ANOVA
 F=6.376902887
 p=.0047646422
 Factor
  df=3
  SS=50.6166667
↓ MS=16.8722222
```

```
One-way ANOVA
↑  MS=16.8722222
 Error
  df=16
  SS=42.3333333
  MS=2.64583333
 Sxp=1.62660177
```

Excel 2007 Enter the data for each group in separate columns. On the home screen, click the **Data** tab. Then in the Analysis group, click **Data Analysis.** In the dialogue box, select **Anova: Single Factor.**

Anova: Single Factor							
SUMMARY							
Groups	*Count*	*Sum*	*Average*	*Variance*			
x1	5	41	8.2	0.7			
x2	4	14	3.5	1.666666667			
x3	6	38	6.333333333	4.666666667			
x4	5	28	5.6	2.8			
ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
Between Groups	50.61666667	3	16.87222222	6.376902887	0.004764642	5.292235983	
Within Groups	42.33333333	16	2.645833333				
Total	92.95	19					

Minitab Enter the data for each group in separate columns. Use **Stat ➤ ANOVA ➤ Oneway(unstacked).**

```
One-way Analysis of Variance

Analysis of Variance
Source   DF     SS      MS      F       P
Factor    3    50.62   16.87   6.38    0.005
Error    16    42.33    2.65
Total    19    92.95
                             Individual 95% CIs For Mean
                          Based on Pooled StDev
Level    N    Mean    StDev   ---+---------+---------+---------+---
x1       5    8.200   0.837                     (-----*-----)
x2       4    3.500   1.291   (------*------)
x3       6    6.333   2.160               (-----*-----)
x4       5    5.600   1.673           (------*------)
                             ---+---------+---------+---------+---
Pooled StDev = 1.627            2.5       5.0       7.5      10.0
```
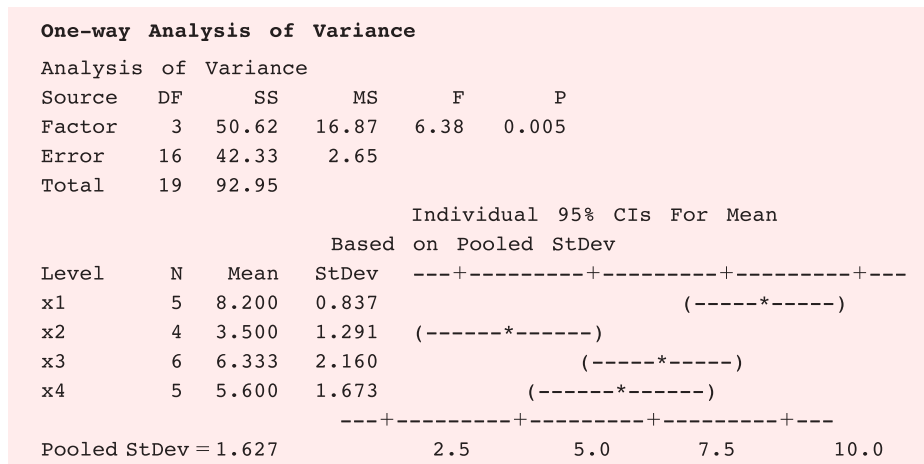
VIEWPOINT | Gear Guide!

How much does that backpack cost? Do brand names really make any difference in cost? One way to answer such a question is to construct a one-way ANOVA test to study list prices for various backpacks with similar features but different brand names. For more information, see Backpacker Magazine, *Vol. 25, issue 157.*

SECTION 10.5 PROBLEMS

In each problem, assume that the distributions are normal and have approximately the same population standard deviation.

In each problem, please provide the following information.
(a) What is the level of significance? State the null and alternate hypotheses.
(b) Find SS_{TOT}, SS_{BET}, and SS_W and check that $SS_{TOT} = SS_{BET} + SS_W$. Find $d.f._{BET}$, $d.f._W$, MS_{BET}, and MS_W. Find the value of the sample test statistic F (F ratio). What are the degrees of freedom?
(c) Find (or estimate) the P-value of the sample test statistic.
(d) Based on your answers in parts (a) to (c), will you reject or fail to reject the null hypothesis?
(e) *Interpret* your conclusion in the context of the application.
(f) Make a summary table for your ANOVA test.

1. *Archaeology: Ceramics* Wind Mountain is an archaeological study area located in southwestern New Mexico. Pot sherds are broken pieces of prehistoric Native American clay vessels. One type of painted ceramic vessel is called *Mimbres classic black-on-white*. At three different sites, the number of such sherds was counted in local dwelling excavations (Source: Based on information from *Mimbres Mogollon Archaeology* by A. I. Woosley and A. J. McIntyre, University of New Mexico Press).

Site I	Site II	Site III
61	25	12
34	18	36
25	54	69
12	67	27

data are continued on next page

Site I	Site II	Site III
79		18
55		14
20		

Shall we reject or not reject the claim that there is no difference in population mean Mimbres classic black-on-white sherd counts for the three sites? Use a 1% level of significance.

2. *Archaeology: Ceramics* Another type of painted ceramic vessel is called *three-circle red-on-white* (see reference in Problem 1). At four different sites in the Wind Mountain archaeological region, the number of such sherds was counted in local dwelling excavations.

Site I	Site II	Site III	Site IV
17	18	32	13
23	4	19	19
6	33	18	14
19	8	43	34
11	25		12
	16		15

Shall we reject or not reject the claim that there is no difference in the population mean three-circle red-on-white sherd counts for the four sites? Use a 5% level of significance.

3. *Economics: Profits per Employee* How productive are U.S. workers? One way to answer this question is to study annual profits per employee. A random sample of companies in computers (I), aerospace (II), heavy equipment (III), and broadcasting (IV) gave the following data regarding annual profits per employee (units in thousands of dollars) (Source: *Forbes Top Companies*, edited by J. T. Davis, John Wiley and Sons).

I	II	III	IV
27.8	13.3	22.3	17.1
23.8	9.9	20.9	16.9
14.1	11.7	7.2	14.3
8.8	8.6	12.8	15.2
11.9	6.6	7.0	10.1
	19.3		9.0

Shall we reject or not reject the claim that there is no difference in population mean annual profits per employee in each of the four types of companies? Use a 5% level of significance.

4. *Economics: Profits per Employee* A random sample of companies in electric utilities (I), financial services (II), and food processing (III) gave the following information regarding annual profits per employee (units in thousands of dollars). (See reference in Problem 3.)

I	II	III
49.1	55.6	39.0
43.4	25.0	37.3
32.9	41.3	10.8
27.8	29.9	32.5
38.3	39.5	15.8
36.1		42.6
20.2		

Shall we reject or not reject the claim that there is no difference in population mean annual profits per employee in each of the three types of companies? Use a 1% level of significance.

5. | *Ecology: Deer* Where are the deer? Random samples of square-kilometer plots were taken in different ecological locations of Mesa Verde National Park. The deer counts per square kilometer were recorded and are shown in the following table (Source: *The Mule Deer of Mesa Verde National Park,* edited by G. W. Mierau and J. L. Schmidt, Mesa Verde Museum Association).

Mountain Brush	Sagebrush Grassland	Pinon Juniper
30	20	5
29	58	7
20	18	4
29	22	9

Shall we reject or accept the claim that there is no difference in the mean number of deer per square kilometer in these different ecological locations? Use a 5% level of significance.

6. | *Ecology: Vegetation* Wild irises are beautiful flowers found throughout the United States, Canada, and northern Europe. This problem concerns the length of the sepal (leaf-like part covering the flower) of different species of wild iris. Data are based on information taken from an article by R. A. Fisher in *Annals of Eugenics* (Vol. 7, part 2, pp. 179–188). Measurements of sepal length in centimeters from random samples of *Iris setosa* (I), *Iris versicolor* (II), and *Iris virginica* (III) are as follows:

I	II	III
5.4	5.5	6.3
4.9	6.5	5.8
5.0	6.3	4.9
5.4	4.9	7.2
4.4	5.2	5.7
5.8	6.7	6.4
5.7	5.5	
	6.1	

Shall we reject or not reject the claim that there are no differences among the population means of sepal length for the different species of iris? Use a 5% level of significance.

7. | *Insurance: Sales* An executive at the home office of Big Rock Life Insurance is considering three branch managers as candidates for promotion to vice president. The branch reports include records showing sales volume for each salesperson in the branch (in hundreds of thousands of dollars). A random sample of these records was selected for salespersons in each branch. All three branches are located in cities in which per capita income is the same. The executive wishes to compare these samples to see if there is a significant difference in performance of salespersons in the three different branches. If so, the information will be used to determine which of the managers to promote.

Branch Managed by Adams	Branch Managed by McDale	Branch Managed by Vasquez
7.2	8.8	6.9
6.4	10.7	8.7
10.1	11.1	10.5
11.0	9.8	11.4
9.9		
10.6		

Use an $\alpha = 0.01$ level of significance. Shall we reject or not reject the claim that there are no differences among the performances of the salespersons in the different branches?

8. *Ecology: Pollution* The quantity of dissolved oxygen is a measure of water pollution in lakes, rivers, and streams. Water samples were taken at four different locations in a river in an effort to determine if water pollution varied from location to location. Location I was 500 meters above an industrial plant water discharge point and near the shore. Location II was 200 meters above the discharge point and in midstream. Location III was 50 meters downstream from the discharge point and near the shore. Location IV was 200 meters downstream from the discharge point and in midstream. The following table shows the results. Lower dissolved oxygen readings mean more pollution. Because of the difficulty in getting midstream samples, ecology students collecting the data had fewer of these samples. Use an $\alpha = 0.05$ level of significance. Do we reject or not reject the claim that the quantity of dissolved oxygen does not vary from one location to another?

Location I	Location II	Location III	Location IV
7.3	6.6	4.2	4.4
6.9	7.1	5.9	5.1
7.5	7.7	4.9	6.2
6.8	8.0	5.1	
6.2		4.5	

9. *Sociology: Ethnic Groups* A sociologist studying New York City ethnic groups wants to determine if there is a difference in income for immigrants from four different countries during their first year in the city. She obtained the data in the following table from a random sample of immigrants from these countries (incomes in thousands of dollars). Use a 0.05 level of significance to test the claim that there is no difference in the earnings of immigrants from the four different countries.

Country I	Country II	Country III	Country IV
12.7	8.3	20.3	17.2
9.2	17.2	16.6	8.8
10.9	19.1	22.7	14.7
8.9	10.3	25.2	21.3
16.4		19.9	19.8

SECTION 10.6

Introduction to Two-Way ANOVA

FOCUS POINTS

- Learn the notation and setup for two-way ANOVA tests.
- Learn about the three main types of deviations and how they break into additional effects.
- Use mean-square values to compute different sample *F* statistics.
- Use the *F* distribution to estimate *P*-values and conclude the test.
- Summarize experimental design features using a completely randomized design flow chart.

Suppose that Friendly Bank is interested in average customer satisfaction regarding the issue of obtaining bank balances and a summary of recent account transactions. Friendly Bank uses two systems, the first being a completely automated voice mail information system requiring customers to enter account numbers and passwords using the telephone keypad, and the second being the use of bank

tellers or bank representatives to give the account information personally to customers. In addition, Friendly Bank wants to learn if average customer satisfaction is the same regardless of the time of day of contact. Three times of day are under study: morning, afternoon, and evening.

Friendly Bank could do two studies: one regarding average customer satisfaction with regard to type of contact (automated or bank representative) and one regarding average customer satisfaction with regard to time of day. The first study could be done using a difference-of-means test because there are only two types of contact being studied. The second study could be accomplished using one-way ANOVA.

Two-way ANOVA

Interaction

However, Friendly Bank could use just *one* study and the technique of *two-way analysis of variance* (known as *two-way ANOVA*) to simultaneously study average customer satisfaction with regard to the variable type of contact and the variable time of day, and *also* with regard to the *interaction* between the two variables. An interaction is present if, for instance, the difference in average customer satisfaction regarding type of contact is much more pronounced during the evening hours than, say, during the afternoon hours or the morning hours.

Factor
Levels

Let's begin our study of two-way ANOVA by considering the organization of data appropriate to two-way ANOVA. Two-way ANOVA involves *two* variables. These variables are called *factors*. The *levels* of a factor are the different values the factor can assume. Example 9 demonstrates the use of this terminology for the information Friendly Bank is seeking.

EXAMPLE 9   FACTORS AND LEVELS

For the Friendly Bank study discussed earlier, identify the factors and their levels, and create a table displaying the information.

SOLUTION: There are two factors. Call factor 1 *time of day*. This factor has three levels: morning, afternoon, and evening. Factor 2 is *type of contact*. This factor has two levels: automated contact and personal contact through a bank representative. Table 10-22 shows how the information regarding customer satisfaction can be organized with respect to the two factors.

TABLE 10-22   **Table for Recording Average Customer Response**

Factor 1: Time of Day	Factor 2: Type of Contact	
	Automated	**Bank Representative**
Morning	Morning Automated	Morning Bank Representative
Afternoon	Afternoon Automated	Afternoon Bank Representative
Evening	Evening Automated	Evening Bank Representative

Cell

When we look at Table 10-22, we see six contact–time-of-day combinations. Each such combination is called a *cell* in the table. The number of cells in any two-way ANOVA data table equals the product of the number of levels of the row factor times the number of levels of the column factor. In the case illustrated by Table 10-22, we see that the number of cells is 3 × 2, or 6.

Just as for one-way ANOVA, our application of two-way ANOVA has some basic requirements:

1. The measurements in each cell of a two-way ANOVA model are assumed to be drawn from a population with a normal distribution.

2. The measurements in each cell of a two-way ANOVA model are assumed to come from distributions with approximately the same variance.

3. The measurements in each cell come from *independent* random samples.

4. There are the *same number of measurements* in each cell.

Procedure to Conduct a Two-Way ANOVA Test (More Than One Measurement per Cell)

We will outline the procedure for two-way ANOVA in five steps. Each step will contain general methods and rationale appropriate to all two-way ANOVA tests with more than one data value in each cell. As we proceed, we will see how the method is applied to the Friendly Bank study.

Let's assume that Friendly Bank has taken random samples of customers fitting the criteria of each of the six cells described in Table 10-22. This means that a random sample of four customers fitting the morning-automated cell were surveyed. Another random sample of four customers fitting the afternoon-automated cell were surveyed, and so on. The bank measured customer satisfaction on a scale of 0 to 10 (10 representing highest customer satisfaction). The data appear in Table 10-23. Table 10-23 also shows cell means, row means, column means, and the total mean $\overline{\overline{x}}$ computed for all 24 data values. We will use these means as we conduct the two-way ANOVA test.

As in any statistical test, the first task is to establish the hypotheses for the test. Then, as in one-way ANOVA, the *F* distribution is used to determine the test conclusion. To compute the sample *F* value for a given null hypothesis, many of the same kinds of computations are done as are done in one-way ANOVA. In particular, we will use degrees of freedom $d.f. = N - 1$ (where N is the total sample size) allocated among the row factor, the column factor, the interaction, and the error (corresponding to "within groups" of one-way ANOVA). We look at the sum of squares *SS* (which measures variation) for the row factor, the column factor, the interaction, and the error. Then we compute the mean square *MS* for each category by taking the *SS* value and dividing by the corresponding degrees of freedom. Finally, we compute the sample *F* statistic for each factor and for the interaction by dividing the appropriate *MS* value by the *MS* value of the error.

TABLE 10-23   **Customer Satisfaction at Friendly Bank**

Factor 1: Time of Day	Factor 2: Type of Contact		Row Means
	Automated	Bank Representative	
Morning	6, 5, 8, 4 $\overline{x} = 5.75$	8, 7, 9, 9 $\overline{x} = 8.25$	Row 1 $\overline{x} = 7.00$
Afternoon	3, 5, 6, 5 $\overline{x} = 4.75$	9, 10, 6, 8 $\overline{x} = 8.25$	Row 2 $\overline{x} = 6.50$
Evening	5, 5, 7, 5 $\overline{x} = 5.50$	9, 10, 10, 9 $\overline{x} = 9.50$	Row 3 $\overline{x} = 7.50$
Column means	Column 1 $\overline{x} = 5.33$	Column 2 $\overline{x} = 8.67$	Total $\overline{x} = 7.00$

STEP 1: Establish the Hypotheses

Because we have two factors, we have hypotheses regarding each of the factors separately (called *main effects*) and then hypotheses regarding the interaction between the factors.

These three sets of hypotheses are

Hypotheses

1. H_0: There is no difference in population means among the levels of the row factor.
    H_1: At least two population means are different among the levels of the row factor.
2. H_0: There is no difference in population means among the levels of the column factor.
    H_1: At least two population means are different among the levels of the column factor.
3. H_0: There is no interaction between the factors.
    H_1: There is an interaction between the factors.

In the case of Friendly Bank, the hypotheses regarding the main effects are

H_0: There is no difference in population mean satisfaction depending on time of contact.
H_1: At least two population mean satisfaction measures are different depending on time of contact.
H_0: There is no difference in population mean satisfaction between the two types of customer contact.
H_1: There is a difference in population mean satisfaction between the two types of customer contact.

The hypotheses regarding interaction between factors are

H_0: There is no interaction between type of contact and time of contact.
H_1: There is an interaction between type of contact and time of contact.

STEP 2: Compute Sum of Squares (*SS*) Values

The calculations for the SS values are usually done on a computer. The main questions are whether population means differ according to the factors or the interaction of the factors. As we look at the Friendly Bank data in Table 10-23, we see that sample averages for customer satisfaction differ not only in each cell but also across the rows and across the columns. In addition, the total sample mean (designated $\bar{\bar{x}}$) differs from almost all the means. We know that different samples of the same size from the same population certainly can have different sample means. We need to decide if the differences are simply due to chance (sampling error) or are occurring because the samples have been taken from different populations with means that are not the same.

The tools we use to analyze the differences among the data values, the cell means, the row means, the column means, and the total mean are similar to those we used in Section 10.5 for one-way ANOVA. In particular, we first examine deviations of various measurements from the total mean $\bar{\bar{x}}$, and then we compute the sum of the squares *SS*.

There are basically three types of deviations:

Total deviation	**= Treatment deviation**	**+ Error deviation**
Compare each data value with the total mean $\overline{\overline{x}}$, $(x - \overline{\overline{x}})$.	For each data value, compare the mean of each cell with the total mean $\overline{\overline{x}}$, (cell $\overline{x} - \overline{\overline{x}}$).	Compare each data value with the mean of its cell, $(x - \text{cell } \overline{x})$.

The treatment deviation breaks down further as

Treatment deviation	**= Deviation for main effect of factor 1**	**+ Deviation for main effect of factor 2**	**+ Deviation for interaction**
For each data value, (cell $\overline{x} - \overline{\overline{x}}$)	For each data value, compare the row mean with the total mean $\overline{\overline{x}}$, (row $\overline{x} - \overline{\overline{x}}$).	For each data value, compare the column mean with the total mean $\overline{\overline{x}}$, (column $\overline{x} - \overline{\overline{x}}$).	For each data value, (cell $\overline{x} -$ corresponding row $\overline{x} -$ corresponding column $\overline{x} + \overline{\overline{x}}$).

The deviations for each data value, row mean, column mean, or cell mean are then *squared and totaled over all the data*. This results in sums of squares, or variations. The *treatment variations* correspond to *between-group variations* of one-way ANOVA. The *error variation* corresponds to the *within-group variation* of one-way ANOVA.

Total varition = Treatment variation + Error variation

$$\underset{\text{all data}}{\sum} (x - \overline{\overline{x}})^2 = \underset{\text{all data}}{\sum} (\text{cell } \overline{x} - \overline{\overline{x}})^2 + \underset{\text{all data}}{\sum} (x - \text{cell } \overline{x})^2$$

$$SS_{TOT} = SS_{TR} + SS_E$$

where

Treatment variation = Factor 1 variation + Factor 2 variation + Interaction variation

$$SS_{TR} = SS_{F1} + SS_{F2} + SS_{F1 \times F2}$$

$$\underset{\text{all data}}{\sum} (\text{cell } \overline{x} - \overline{\overline{x}})^2 = \underset{\text{all data}}{\sum} (\text{row } \overline{x} - \overline{\overline{x}})^2 + \underset{\text{all data}}{\sum} (\text{col } \overline{x} - \overline{\overline{x}})^2 + \underset{\text{all data}}{\sum} (\text{cell } \overline{x} - \text{row } \overline{x} - \text{col } \overline{x} + \overline{\overline{x}})^2$$

The actual calculation of all the required SS values is quite time-consuming. In most cases, computer packages are used to obtain the results. For the Friendly Bank data, the following table is a Minitab printout giving the sum of squares *SS* for the type-of-contact factor, the time-of-day factor, the interaction between time and type of contact, and the error.

```
Minitab Printout for Customer Satisfaction at Friendly Bank

Analysis of Variance for Response
Source          DF        SS        MS        F         P
Time             2       4.00      2.00      1.24      0.313
Type             1      66.67     66.67     41.38      0.000
Interaction      2       2.33      1.17      0.72      0.498
Error           18      29.00      1.61
Total           23     102.00
```

We see that $SS_{type} = 66.77$, $SS_{time} = 4.00$, $SS_{interaction} = 2.33$, $SS_{error} = 29.00$, and $SS_{TOT} = 102$ (the total of the other four sums of squares).

STEP 3: Compute the Mean Square (*MS*) Values

The calculations for the MS values are usually done on a computer. Although the sum of squares computed in step 2 represents variation, we need to compute mean-square (*MS*) values for two-way ANOVA. As in one-way ANOVA, we compute *MS* values by dividing the *SS* values by respective degrees of freedom:

$$\text{Mean square } MS = \frac{\text{Corresponding sum of squares } SS}{\text{Respective degrees of freedom}}$$

For two-way ANOVA with more than one data value per cell, the degrees of freedom are

Degrees of freedom

d.f. of row factor = $r - 1$ d.f. of interaction = $(r - 1)(c - 1)$

d.f. of column factor = $c - 1$ d.f. of error = $rc(n - 1)$

d.f. of total = $nrc - 1$

where r = number of rows, c = number of columns, and n = number of data values in one cell.

The Minitab table shows the degrees of freedom and the *MS* values for the main effect factors, the interaction, and the error for the Friendly Bank study.

STEP 4: Compute the Sample *F* Statistic for Each Factor and for the Interaction

Under the assumption of the respective null hypothesis, we have

$$\text{Sample } F \text{ for row factor} = \frac{MS \text{ for row factor}}{MS \text{ for error}}$$

with degrees of freedom numerator, $d.f._N = d.f.$ of row factor

degrees of freedom denominator, $d.f._D = d.f.$ of error

$$\text{Sample } F \text{ for column factor} = \frac{MS \text{ for column factor}}{MS \text{ for error}}$$

with degrees of freedom numerator, $d.f._N = d.f.$ of column factor

degrees of freedom denominator, $d.f._D = d.f.$ of error

$$\text{Sample } F \text{ for interaction} = \frac{MS \text{ for interaction}}{MS \text{ for error}}$$

with degrees of freedom numerator, $d.f._N = d.f.$ of interaction

degrees of freedom denominator, $d.f._D = d.f.$ of error

For the Friendly Bank study, the sample *F* values are

$$\text{Sample } F \text{ for time:} \quad F = \frac{MS_{time}}{MS_{error}} = \frac{2.00}{1.61} = 1.24$$

$$d.f._N = 2 \quad \text{and} \quad d.f._D = 18$$

$$\text{Sample } F \text{ for type of contact: } \quad F = \frac{MS_{\text{type}}}{MS_{\text{error}}} = \frac{66.67}{1.61} = 41.41$$

$$d.f._N = 1 \quad \text{and} \quad d.f._D = 18$$

$$\text{Sample } F \text{ for interaction: } \quad F = \frac{MS_{\text{interaction}}}{MS_{\text{error}}} = \frac{1.17}{1.61} = 0.73$$

$$d.f._N = 2 \quad \text{and} \quad d.f._D = 18$$

Due to rounding, the sample F values we just computed differ slightly from those shown in the Minitab printout.

STEP 5: Conclude the Test

As with one-way ANOVA, larger values of the sample F statistic discredit the null hypothesis that there is no difference in population means across a given factor. The smaller the area to the right of the sample F statistic, the more likely there is an actual difference in some population means across the different factors. Smaller areas to the right of the sample F for interaction indicate greater likelihood of interaction between factors. Consequently, the P-value of a sample F statistic is the area of the F distribution to the *right* of the sample F statistic. Figure 10-18 shows the P-value associated with a sample F statistic.

Finding the P-value

Most statistical computer software packages provide P-values for the sample test statistic. You can also use the F distribution (Table 8 of Appendix II) to estimate the P-value. Once you have the P-value, compare it to the preset level of significance α. If the P-value is less than or equal to α, then reject H_0. Otherwise, do not reject H_0.

Be sure to test for interaction between the factors *first*. If you *reject* the null hypothesis of no interaction, then you should *not* test for a difference of means in the levels of the row factors or for a difference of means in the levels of the column factors because the interaction of the factors makes interpretation of the results of the main effects more complicated. A more extensive study of two-way ANOVA beyond the scope of this book shows how to interpret the results of the test of the main factors when there is interaction. For our purposes, we will simply stop the analysis rather than draw misleading conclusions.

FIGURE 10-18
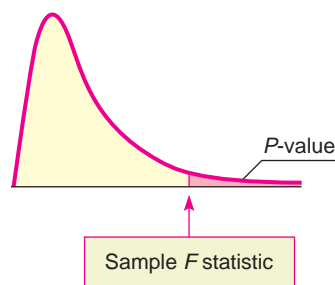
For two-way ANOVA, the P-value of the sample statistic is the area of the F distribution to the *right* of the sample F statistic.

In two-way ANOVA, test for *interaction* first. If you reject the null hypothesis of no interaction, then do not continue with any tests of differences of means among other factors (unless you know techniques more advanced than those presented in this section).

If the test for interaction between the factors indicates that there is no evidence of interaction, then proceed to test the hypotheses regarding the

levels of the row factor and the hypotheses regarding the levels of the column factor.

For the Friendly Bank study, we proceed as follows:

1. First, we determine if there is any evidence of interaction between the factors. The sample test statistic for interaction is $F = 0.73$, with P-value ≈ 0.498. Since the P-value is greater than $\alpha = 0.05$, we do not reject H_0. There is no evidence of interaction. Because there is no evidence of interaction between the main effects of type of contact and time of day, we proceed to test each factor for a difference in population mean satisfaction among the respective levels of the factors.

2. Next, we determine if there is a difference in mean satisfaction according to type of contact. The sample test statistic for type of contact is $F = 41.41$, with P-value ≈ 0.000 (to three places after the decimal). Since the P-value is less than $\alpha = 0.05$, we reject H_0. At the 5% level of significance, we conclude that there is a difference in average customer satisfaction between contact with an automated system and contact with a bank representative.

3. Finally, we determine if there is a difference in mean satisfaction according to time of day. The sample test statistic for time of day is $F = 1.24$, with P-value ≈ 0.313. Because the P-value is greater than $\alpha = 0.05$, we do not reject H_0. We conclude that at the 5% level of significance, there is no evidence that population mean customer satisfaction is different according to time of day.

Special Case: One Observation in Each Cell with No Interaction

In the case where our data consist of only one value in each cell, there are no measures for sum of squares *SS* interaction or mean-square *MS* interaction, and we cannot test for interaction of factors using two-way ANOVA. If it *seems reasonable* (based on other information) to assume that there is *no* interaction between the factors, then we can use two-way ANOVA techniques to test for average response differences due to the main effects. In Guided Exercise 12, we look at two-way ANOVA applied to the special case of only one measurement per cell and no interactions.

GUIDED EXERCISE 12 | **Special-case two-way ANOVA**

Let's use two-way ANOVA to test if the average fat content (grams of fat per 3-oz serving) of potato chips is different according to the brand or according to which laboratory made the measurement. Use $\alpha = 0.05$. (See the following Minitab tables.)

Average Grams of Fat in a 3-oz Serving of Potato Chips

Brand	Lab I	Lab II	Lab III
	Laboratory		
Texas Chips	32.4	33.1	32.9
Great Chips	37.9	37.7	37.8
Chip Ooh	29.1	29.4	29.5

Minitab Printout for Potato Chip Data

Analysis of Variance for Fat

Source	DF	SS	MS	F	P
Brand	2	108.7022	54.3511	968.63	0.000
Lab	2	0.1422	0.0711	1.27	0.375
Error	4	0.2244	0.0561		
Total	8	109.0689			

Continued

GUIDED EXERCISE 12 *continued*

(a) List the factors and the number of levels for each.

→ The factors are brand and laboratory. Each factor has three levels.

(b) Assuming there is no interaction, list the hypotheses for each factor.

→ For brand,
H_0: There is no difference in population mean fat by brands.
H_1: At least two brands have different population mean fat contents.
For laboratory,
H_0: There is no difference in mean fat content as measured by the labs.
H_1: At least two of the labs give different mean fat measurements.

(c) Calculate the sample *F* statistic for brands and compare it to the value given in the Minitab printout. Look at the *P*-value in the printout. What is your conclusion regarding average fat content among brands?

→ For brand,

$$\text{Sample } F = \frac{MS_{\text{brand}}}{MS_{\text{error}}} \approx \frac{54.35}{0.056} \approx 970$$

Using the Minitab printout, we see *P*-value ≈ 0.000 (to three places after the decimal). Using Table 8 of Appendix II, we see *P*-value < 0.001. Since the *P*-value is less than $\alpha = 0.05$, we reject H_0 and conclude that at the 5% level of significance, at least two of the brands have different mean fat contents.

(d) Calculate the sample *F* statistic for laboratories and compare it to the value given in the Minitab printout. What is your conclusion regarding average fat content as measured by the different laboratories?

→ For laboratories,

$$\text{Sample } F = \frac{MS_{\text{lab}}}{MS_{\text{error}}} \approx \frac{0.0711}{0.056} \approx 1.27$$

Using the Minitab printout, we see *P*-value ≈ 0.375. Using Table 8 of Appendix II, we see *P*-value > 0.100. Because the *P*-value is greater than $\alpha = 0.05$, we conclude that at the 5% level of significance, there is no evidence of differences in average measurements of fat content as determined by the different laboratories.

TECH NOTES   The calculations involved in two-way ANOVA are usually done using statistical or spreadsheet software. Basic printouts from various software packages are similar. Specific instructions for using Excel 2007, Minitab, and SPSS are given in the Using Technology section at the end of the chapter.

Experimental Design

In the preceding section and in this section, we have seen aspects of one-way and two-way ANOVA, respectively. Now let's take a brief look at some experimental design features that are appropriate for the use of these techniques.

For one-way ANOVA, we have one factor. Different levels for the factor form the treatment groups under study. In a *completely randomized design*, independent random samples of experimental subjects or objects are selected for each treatment group. For example, suppose a researcher wants to study the effects of different treatments for the condition of slightly high blood pressure. Three treatments are under study: diet, exercise, and medication. In a completely randomized design, the people participating in the experiment are *randomly* assigned to each treatment group. Table 10-24 shows the process.

Completely randomized design

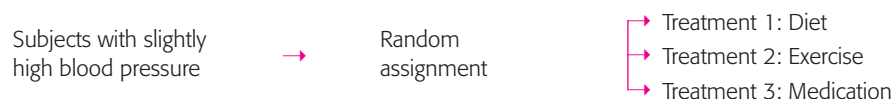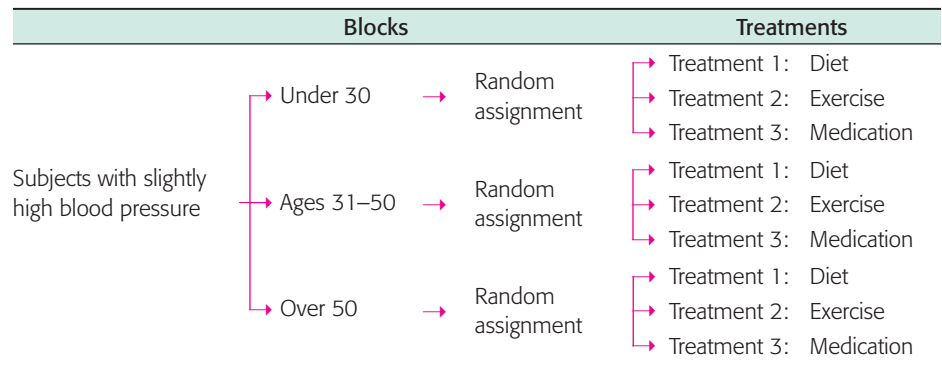**TABLE 10-24   Completely Randomized Design Flow Chart**

Subjects with slightly high blood pressure → Random assignment →
Treatment 1: Diet
Treatment 2: Exercise
Treatment 3: Medication

TABLE 10-25   Randomized Block Design Flow Chart

Blocks		Treatments
Subjects with slightly high blood pressure	Under 30 → Random assignment	Treatment 1: Diet Treatment 2: Exercise Treatment 3: Medication
	Ages 31–50 → Random assignment	Treatment 1: Diet Treatment 2: Exercise Treatment 3: Medication
	Over 50 → Random assignment	Treatment 1: Diet Treatment 2: Exercise Treatment 3: Medication

For two-way ANOVA, there are *two* factors. When we *block* experimental subjects or objects together based on a similar characteristic that might affect responses to treatments, we have a *block design*. For example, suppose the researcher studying treatments for slightly high blood pressure believes that the age of subjects might affect the response to the three treatments. In such a case, blocks of subjects in specified age groups are used. The factor "age" is used to form blocks. Suppose age has three levels: under age 30, ages 31–50, and over age 50. The same number of subjects is assigned to each block. Then the subjects in each block are randomly assigned to the different treatments of diet, exercise, or medication. Table 10-25 shows the *randomized block design*.

Experimental design is an essential component of good statistical research. The design of experiments can be quite complicated, and if the experiment is complex, the services of a professional statistician may be required. The use of blocks helps the researcher account for some of the most important sources of variability among the experimental subjects or objects. Then, randomized assignments to different treatment groups help average out the effects of other remaining sources of variability. In this way, differences among the treatment groups are more likely to be caused by the treatments themselves rather than by other sources of variability.

Randomized block design

VIEWPOINT | Who Watches Cable TV?

Consider the following claim: Average cable TV viewers are, generally speaking, as affluent as newspaper readers and better off than radio or magazine audiences. How do we know whether this claim is true? One way to answer such a question is to construct a two-way ANOVA test in which the rows represent household income levels and the columns represent cable TV viewers, news paper readers, magazine audiences, and radio listeners. The response variable is an index that represents the ratio of the specific medium compared with U.S. averages. For more information and data, see American Demographics *(Vol. 17, No. 6).*

SECTION 10.6 PROBLEMS

1. *Statistical Literacy: Physical Therapy* Does talking while walking slow you down? A study reported in the journal *Physical Therapy* (Vol. 72, No. 4) considered mean cadence (steps per minute) for subjects using no walking device, a standard walker, and a rolling walker. In addition, the cadence was measured when the subjects had to perform dual tasks. The second task was to respond vocally to a signal while walking. Cadence was measured for subjects who were

just walking (using no device, a standard walker, or a rolling walker) and for subjects required to respond to a signal while walking. List the factors and the number of levels of each factor. How many cells are there in the data table?

2. *Statistical Literacy: Salary Survey* Academe, Bulletin of the American Association of University Professors (Vol. 83, No. 2) presents results of salary surveys (average salary) by rank of the faculty member (professor, associate, assistant, instructor) and by type of institution (public, private). List the factors and the number of levels of each factor. How many cells are there in the data table?

3. *Critical Thinking: Physical Therapy* For the study regarding mean cadence (see Problem 1), two-way ANOVA was used. Recall that the two factors were walking device (none, standard walker, rolling walker) and dual task (being required to respond vocally to a signal or no dual task required). Results of two-way ANOVA showed that there was no evidence of interaction between the factors. However, according to the article, "The ANOVA conducted on the cadence data revealed a main effect of walking device." When the hypothesis regarding no difference in mean cadence according to which, if any, walking device was used, the sample *F* was 30.94, with $d.f._N = 2$ and $d.f._D = 18$. Further, the *P*-value for the result was reported to be less than 0.01. From this information, what is the conclusion regarding any difference in mean cadence according to the factor "walking device used"?

4. *Education: Media Usage* In a study of media usage versus education level (*American Demographics,* Vol. 17, No. 6), an index was used to measure media usage, where a measurement of 100 represents the U.S. average. Values above 100 represent above-average media usage.

	Media				
Education Level	Cable Network	Prime-Time TV	Radio	Newspaper	Magazine
Less than high school	80	112	87	76	85
High school graduate	103	105	100	99	101
Some college	107	94	106	105	107
College graduate	108	90	106	116	108

Source: American Demographics by Staff. Copyright 1995 by CRAIN COMMUNICATIONS INC. Reproduced with permission of CRAIN COMMUNICATIONS INC in the formats Textbook and Other book via Copyright Clearance Center.

(a) List the factors and the number of levels of each factor.
(b) Assume there is no interaction between the factors. Use two-way ANOVA and the following Minitab printout to determine if there is a difference in population mean index based on education. Use $\alpha = 0.05$.
(c) Determine if there is a difference in population mean index based on media. Use $\alpha = 0.05$.

```
Minitab Printout for Media/Education Data

Analysis of Variance for Index
Source    DF      SS      MS      F          P
Edu        3     961     320    2.96      0.075
Media      4       5       1    0.01      1.000
Error     12    1299     108
Total     19    2264
```

5. *Income: Media Usage* In the same study described in Problem 4, media usage versus household income also was considered. The media usage indices for the various media and income levels follow.

	Media				
Income Level	Cable Network	Prime-Time TV	Radio	Newspaper	Magazine
Less than $20,000	78	112	89	80	91
$20,000–$39,999	97	105	100	97	100
$40,000–$74,999	113	92	106	111	105
$75,000 or more	121	94	107	121	105

Source: American Demographics by Staff. Copyright 1995 by CRAIN COMMUNICATIONS INC. Reproduced with permission of CRAIN COMMUNICATIONS INC in the formats Textbook and Other book via Copyright Clearance Center.

(a) List the factors and the number of levels of each factor.
(b) Assume there is no interaction between the factors. Use two-way ANOVA and the following Minitab printout to determine if there is a difference in population mean index based on income. Use $\alpha = 0.05$.
(c) Determine if there is a difference in population mean index based on media. Use $\alpha = 0.05$.

```
Minitab Printout for Media/Income Data

Analysis of Variance for Index
Source     DF      SS       MS       F        P
Income     3      1078     359      2.77     0.088
Media      4        15       4      0.03     0.998
Error     12      1558     130
Total     19      2651
```

6. *Gender: Grade Point Average* Does college grade point average (GPA) depend on gender? Does it depend on class (freshman, sophomore, junior, senior)? In a study, the following GPA data were obtained for random samples of college students in each of the cells.

	Class							
Gender	Freshman		Sophomore		Junior		Senior	
Male	2.8	2.1	2.5	2.3	3.1	2.9	3.8	3.6
	2.7	3.0	2.9	3.5	3.2	3.8	3.5	3.1
Female	2.3	2.9	2.6	2.4	2.6	3.6	3.2	3.5
	3.5	3.9	3.3	3.6	3.3	3.7	3.8	3.6

(a) List the factors and the number of levels of each factor.
(b) Use two-way ANOVA and the following Minitab printout to determine if there is any evidence of interaction between the two factors at a level of significance of 0.05.

```
Minitab Printout of GPA Based on Gender and Class

Analysis of Variance for GPA
Source        DF      SS       MS       F       P
Gender         1     0.281    0.281    1.26    0.273
Class          3     2.226    0.742    3.32    0.037
Interaction    3     0.286    0.095    0.43    0.736
Error         24     5.365    0.224
Total         31     8.159
```

(c) If there is no evidence of interaction, use two-way ANOVA and the Minitab printout to determine if there is a difference in mean GPA based on class. Use $\alpha = 0.05$.

(d) If there is no evidence of interaction, use two-way ANOVA and the Minitab printout to determine if there is a difference in mean GPA based on gender. Use $\alpha = 0.05$.

7. *Experimental Design: Teaching Style* A researcher forms three blocks of students interested in taking a history course. The groups are based on grade point average (GPA). The first group consists of students with a GPA less than 2.5, the second group consists of students with a GPA between 2.5 and 3.1, and the last group consists of students with a GPA greater than 3.1. History courses are taught in three ways: traditional lecture, small-group collaborative method, and independent study. The researcher randomly assigns 10 students from each block to sections of history taught each of the three ways. Sections for each teaching style then have 10 students from each block. The researcher records the scores on a common course final examination administered to each student. Draw a flow chart showing the design of this experiment. Does the design fit the model for randomized block design?

Chapter Review

SUMMARY

In this chapter, we introduced applications of two probability distributions: the chi-square distribution and the *F* distribution.

- The chi-square distribution is used for tests of independence or homogeneity, tests of goodness of fit, tests of variance σ^2, and tests to estimate a variance σ^2.

- The *F* distribution is used for tests of two variances, one-way ANOVA, and two-way ANOVA.

ANOVA tests are used to determine whether there are differences among means for several groups.

- If groups are based on the value of only one variable, we have one-way ANOVA.

- If groups are formed using two variables, we use two-way ANOVA to test for differences of means based on either variable or on an interaction between the variables.

IMPORTANT WORDS & SYMBOLS

Section 10.1
Independence test 593
Chi-square distribution overview 592
Observed frequency of a cell, *O* 593
Contingency table with cells 593
Expected frequency of a cell, *E* 594
Row total 595
Column total 595
Degrees of freedom, $d.f. = (R - 1)(C - 1)$ for χ^2 distribution and tests of independence 597
Test of homogeneity 600
Multinomial experiments 603

Section 10.2
Goodness-of-fit test 609

Degrees of freedom $d.f.$ for χ^2 distribution and goodness-of-fit tests 611

Section 10.3
Hypotheses tests about σ^2 624
Confidence interval for σ^2 625
χ_U^2, χ_L^2 625

Section 10.4
Hypothesis tests about two variances 631
F distribution overview 630

Section 10.5
ANOVA 641
Sum of squares, *SS* 643
SS_{TOT}, SS_{BET}, SS_W 645

VIEWPOINT | Movies and Money!

Young adults are the movie industry's best customers. However, going to the movies is expensive, which may explain why attendance rates increase with household income. Using what you have learned in this chapter, you can create appropriate chi-square tests to determine how good a fit exists between national percentage rates of attendance by household income and attendance rates in your demographic area. For more information and national data, see American Demographics *(Vol. 18, No. 12).*

CHAPTER REVIEW PROBLEMS

1. *Statistical Literacy* Of the following random variables, which have only non-negative values: z, t, chi-square, F?

2. *Statistical Literacy* Of the following probability distributions, which are always symmetric: normal, Student's t, chi-square, F?

3. *Critical Thinking* Suppose you took random samples from three distinct age groups. Through a survey, you determined how many respondents from each age group preferred to get news from TV, newspapers, the Internet, or another source (respondents could select only one mode). What type of test would be appropriate to determine if there is sufficient statistical evidence to claim that the proportions of each age group preferring the different modes of obtaining news are not the same? Select from tests of independence, homogeneity, goodness of fit, and ANOVA.

4. *Critical Thinking* Suppose you take a random sample from a normal population and you want to determine whether there is sufficient statistical evidence to claim that the population variance differs from a corresponding variance specified in a government contract. Which type of test is appropriate, a test of one variance or a test of two variances?

Before you solve Problems 5–14, first classify the problem as one of the following:

   Chi-square test of independence or homogeneity
   Chi-square goodness of fit
   Chi-square for testing or estimating σ^2 or σ
   F test for two variances
   One-way ANOVA
   Two-way ANOVA

Then, in each of the problems when a test is to be performed, do the following:

   (i) Give the value of the level of significance. State the null and alternate hypotheses.
   (ii) Find the sample test statistic.
   (iii) Find or estimate the *P*-value of the sample test statistic.
   (iv) Conclude the test.

(v) *Interpret* the conclusion in the context of the application.

(vi) In the case of one-way ANOVA, make a summary table.

5. *Sales: Packaging* The makers of Country Boy Corn Flakes are thinking about changing the packaging of the cereal in the hope of improving sales. In an experiment, five stores of similar size in the same region sold Country Boy Corn Flakes in different-shaped containers for 2 weeks. Total packages sold are given in the following table. Using a 0.05 level of significance, shall we reject or fail to reject the hypothesis that the mean sales are the same, no matter which box shape is used?

Cube	Cylinder	Pyramid	Rectangle
120	110	74	165
88	115	62	98
65	180	110	125
95	96	66	87
71	85	83	118

6. *Education: Exams* Professor Fair believes that extra time does not improve grades on exams. He randomly divided a group of 300 students into two groups and gave them all the same test. One group had exactly 1 hour in which to finish the test, and the other group could stay as long as desired. The results are shown in the following table. Test at the 0.01 level of significance that time to complete a test and test results are independent.

Time	A	B	C	F	Row Total
1 h	23	42	65	12	142
Unlimited	17	48	85	8	158
Column Total	40	90	150	20	300

7. *Tires: Blowouts* A consumer agency is investigating the blowout pressures of Soap Stone tires. A Soap Stone tire is said to blow out when it separates from the wheel rim due to impact forces usually caused by hitting a rock or a pothole in the road. A random sample of 30 Soap Stone tires were inflated to the recommended pressure, and then forces measured in foot-pounds were applied to each tire (1 foot-pound is the force of 1 pound dropped from a height of 1 foot). The customer complaint is that some Soap Stone tires blow out under small-impact forces, while other tires seem to be well made and don't have this fault. For the 30 test tires, the sample standard deviation of blowout forces was 1353 foot-pounds.

(a) Soap Stone claims its tires will blow out at an average pressure of 20,000 foot-pounds, with a standard deviation of 1020 foot-pounds. The average blowout force is not in question, but the variability of blowout forces is in question. Using a 0.01 level of significance, test the claim that the variance of blowout pressures is more than Soap Stone claims it is.

(b) Find a 95% confidence interval for the variance of blowout pressures, using the information from the random sample.

8. *Computer Science: Data Processing* Anela is a computer scientist who is formulating a large and complicated program for a type of data processing. She has three ways of storing and retrieving data: CD, tape, and disk. As an experiment, she sets up her program in three different ways: one using CDs, one using tapes, and the other using disks. Then she makes four test runs of this type of data processing on each program. The time required to execute each program is shown in the following table (in minutes). Use a 0.01 level of significance to test the hypothesis that the mean processing time is the same for each method.

CD	Tape	Disks
8.7	7.2	7.0
9.3	9.1	6.4
7.9	7.5	9.8
8.0	7.7	8.2

9. | *Teacher Ratings: Grades* Professor Stone complains that students' teacher ratings depend on the grade students receive. In other words, according to Professor Stone, a teacher who gives good grades gets good ratings, and a teacher who gives bad grades gets bad ratings. To test this claim, the Student Assembly took a random sample of 300 teacher ratings on which the students' grades for the course also were indicated. The results are given in the following table. Test the hypothesis that teacher ratings and student grades are independent at the 0.01 level of significance.

Rating	A	B	C	F (or withdrawal)	Row Total
Excellent	14	18	15	3	50
Average	25	35	75	15	150
Poor	21	27	40	12	100
Column Total	60	80	130	30	300

10. | *Packaging: Corn Flakes* A machine that puts corn flakes into boxes is adjusted to put an average of 15 ounces into each box, with standard deviation of 0.25 ounce. If a random sample of 12 boxes gave a sample standard deviation of 0.38 ounce, do these data support the claim that the variance has increased and the machine needs to be brought back into adjustment? (Use a 0.01 level of significance.)

11. | *Sociology: Age Distribution* A sociologist is studying the age of the population in Blue Valley. Ten years ago, the population was such that 20% were under 20 years old, 15% were in the 20- to 35-year-old bracket, 30% were between 36 and 50, 25% were between 51 and 65, and 10% were over 65. A study done this year used a random sample of 210 residents. This sample showed

Under 20	20–35	36–50	51–65	Over 65
26	27	69	68	20

At the 0.01 level of significance, has the age distribution of the population of Blue Valley changed?

12. | *Engineering: Roller Bearings* Two processes for manufacturing large roller bearings are under study. In both cases, the diameters (in centimeters) are being examined. A random sample of 21 roller bearings from the old manufacturing process showed the sample variance of diameters to be $s^2 = 0.235$. Another random sample of 26 roller bearings from the new manufacturing process showed the sample variance of their diameters to be $s^2 = 0.128$. Use a 5% level of significance to test the claim that there is a difference (either way) in the population variances between the old and new manufacturing processes.

13. | *Engineering: Light Bulbs* Two processes for manufacturing 60-watt light bulbs are under study. In both cases, the life (in hours) of the bulb before it burns out is being examined. A random sample of 18 light bulbs manufactured using the old process showed the sample variance of lifetimes to be $s^2 = 51.87$. Another random sample of 16 light bulbs manufactured using the new process showed the sample variance of the lifetimes to be $s^2 = 135.24$. Use a 5% level of significance to test the claim that the population variance of lifetimes for the new manufacturing process is larger than that of the old process.

14. *Advertising: Newspapers* Does the newspaper section in which an ad is placed make a difference in the average daily number of people responding to the ad? Is there a difference in average daily number of people responding to the ad if it is placed in the Sunday newspaper as compared with the Wednesday newspaper? Video Entertainment is a video club that sells videotapes featuring movies of all kinds—instructional videos, videos of TV specials, etc. To attract new customers, Video Entertainment ran ads in the Wednesday newspaper and in the Sunday newspaper offering six free videotapes or DVDs to new members. In addition, the ads were placed in the sports section, the entertainment section, and the business section of the local newspaper. Ads running in the different sections carried different promotion codes. Different codes also were used for Wednesday ads compared with Sunday ads. The numbers of people responding to the ads for days selected at random were recorded according to the promotion codes. The results follow:

Day	Section of Newspaper								
	Sports			Entertainment			Business		
Wed.	12	15	11	22	14	17	2	4	3
	12	15	20	22	18	12	5	0	1
Sun.	20	23	25	32	26	28	13	16	13
	33	15	17	31	25	41	15	14	10

(a) List the factors and the number of levels for each factor.
(b) Use the following Minitab printout and the *F* distribution table (Table 8 of Appendix II) to test for interaction between the variables. Use a 1% level of significance.

```
Minitab Printout for Mean Number of Responses per Day

Analysis of Variance for Response
Source          DF          SS          MS          F          P
Day             1        1024.0      1024.0      55.28      0.000
Section         2        1573.6       786.8      42.48      0.000
Interaction     2          38.0        19.0       1.03      0.371
Error          30         555.7        18.5
Total          35        3191.2
```

(c) If there is no evidence of interaction between the factors, test for a difference in mean number of daily responses for the levels of the day factor. Use $\alpha = 0.01$.
(d) If there is no evidence of interaction between the factors, test for a difference in mean number of daily responses for the levels of the section factor. Use $\alpha = 0.01$.

DATA HIGHLIGHTS: GROUP PROJECTS

Break into small groups and discuss the following topics. Organize a brief outline in which you summarize the main points of your group discussion.

The *Statistical Abstract of the United States* reported information about the percentage of arrests of all drunk drivers according to age group. In the following table, the entry 3.7 in the first row means that in the entire United States, about 3.7% of all people arrested for drunk driving were in the age group 16–17 years. The Freemont County Sheriff's Office obtained data about the number of drunk drivers arrested in each age group over the past several years. In the following table, the entry 8 in the first row means that eight people in the age group 16–17 years were arrested for drunk driving in Freemont County.

Use a chi-square test with 5% level of significance to test the claim that the age distribution of drunk drivers arrested in Freemont County is the same as the national age distribution of drunk drivers arrested.

(a) State the null and alternate hypotheses.

(b) Find the value of the chi-square test statistic from the sample.

(c) Find the degrees of freedom and the *P*-value of the test statistic.

(d) Decide whether you should reject or not reject the null hypothesis.

(e) State your conclusion in the context of the problem.

Distribution of Drunk Driver Arrests by Age

Age	National Percentage	Number in Freemont County
16–17	3.7	8
18–24	18.9	35
25–29	12.9	23
30–34	10.3	19
35–39	8.5	12
40–44	7.9	14
45–49	8.0	16
50–54	7.9	13
55–59	6.8	10
60–64	5.7	9
65 and over	9.4	15
	100%	174

(f) How could you gather data and conduct a similar test for the city or county in which you live? Explain.

LINKING CONCEPTS: WRITING PROJECTS

Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you summarize the main points. Please include formulas and graphs as appropriate.

1. In this chapter, you studied the chi-square distribution and three principal applications of the distribution.

   (a) Outline the basic ideas behind the chi-square test of independence. What is a contingency table? What are the null and alternate hypotheses? How is the test statistic constructed? What basic assumptions underlie this application of the chi-square distribution?

   (b) Outline the basic ideas behind the chi-square test of goodness of fit. What are the null and alternate hypotheses? How is the test statistic constructed? There are a number of direct similarities between tests of independence and tests of goodness of fit. Discuss and summarize these similarities.

   (c) Outline the basic ideas behind the chi-square method of testing and estimating a standard deviation. What basic assumptions underlie this process?

   (d) Outline the basic ideas behind the chi-square test of homogeneity. What are the null and alternate hypotheses? How is the test statistic contructed? What basic assumptions underlie the application of the chi-square distribution?

2. The *F* distribution is used to construct a one-way ANOVA test for comparing several sample means.

   (a) Outline the basic purpose of ANOVA. How does ANOVA avoid high risk due to multiple type I errors?

   (b) Outline the basic assumption for ANOVA.

   (c) What are the null and alternate hypotheses in an ANOVA test? If the test conclusion is to reject the null hypothesis, do we know which of the population means are different from each other?

   (d) What is the *F* distribution? How are the degrees of freedom for numerator and denominator determined?

   (e) What do we mean by a summary table of ANOVA results? What are the main components of such a table? How is the final decision made?

USING TECHNOLOGY

Application
Analysis of Variance (One-Way ANOVA)

The following data comprise a winter mildness/severity index for three European locations near 50° north latitude. For each decade, the number of unmistakably mild months minus the number of unmistakably severe months for December, January, and February is given.

Decade	Britain	Germany	Russia
1800	−2	−1	+1
1810	−2	−3	−1
1820	0	0	0
1830	−3	−2	−1
1840	−3	−2	+1
1850	−1	−2	+3
1860	+8	+6	+1
1870	0	0	−3
1880	−2	0	+1
1890	−3	−1	+1
1900	+2	0	+2
1910	+5	+6	+1
1920	+8	+6	+2
1930	+4	+4	+5
1940	+1	−1	−1
1950	0	+1	+2

Table is based on data from Exchanging Climate by H. H. Lamb; copyright © 1966. Reprinted by permission of Routledge, UK.

1. We wish to test the null hypothesis that the mean winter indices for Britain, Germany, and Russia are all equal against the alternate hypothesis that they are not all equal. Use a 5% level of significance.

2. What is the sum of squares between groups? Within groups? What is the sample F ratio? What is the P-value? Shall we reject or fail to reject the statement that the mean winter indices for these locations in Britain, Germany, and Russia are the same?

3. What is the smallest level of significance at which we could conclude that the mean winter indices for these locations are not all equal?

Technology Hints (One-Way ANOVA)

SPSS

Enter all the data in one column. In another column, use integers to designate the group to which each data value belongs. Use the menu selections **Analyze ➤ Compare Means ➤ One-Way ANOVA.** Move the column containing the data to the dependent list. Move the column containing group designation to the factor list.

Technology Hints (Two-Way ANOVA)

Excel 2007

Excel has two commands for two-way ANOVA, depending on how many data values are in each cell. On the home screen, click the **Data** tab. Then in the Analysis group, click **Data Analysis.** Then use **ANOVA: Two-Factor with Replication** if there are two or more sample measurements for each factor combination cell. Again, there must be the same number of data in each cell.

ANOVA: Two-Factor without Replication if there is only one data value in each factor combination cell.

Data entry is fairly straightforward. For example, look at the Excel spreadsheet for the data of Guided Exercise 12 regarding the fat content of different brands of potato chips as measured by different labs.

	A	B	C	D
1		**Lab 1**	**Lab 2**	**Lab 3**
2	Texas	32.4	33.1	32.9
3	Great	37.9	37.7	37.8
4	Chip Ooh	29.1	29.4	29.5

Minitab

For Minitab, all the data for the response variable (in this case, fat content) are entered into a single column. Create two more columns, one for the row number of the cell containing the data value and one for the column number of the cell containing the data value. For the potato chip example, the rows correspond to the brand and the columns to the lab doing the analysis. Use the menu choices **Stat ➤ ANOVA ➤ Two-Way.**

	C1	C2	C3
	Brand	Lab	Fat
1	1	1	32.4
2	1	2	33.1
3	1	3	32.9
4	2	1	37.9
5	2	2	37.7
6	2	3	37.8
7	3	1	29.1
8	3	2	29.4
9	3	3	29.5

SPSS

Data entry for SPSS is similar to that for Minitab. Enter all the data in one column. Use a separate column for each factor and use a label or integer to designate the group in the particular factor. Under Variable View, type appropriate labels for the columns of data. Use the menu selections **Analyze ➤ General Linear Model ➤ Univariate. . . .** In the dialogue box, the dependent variable is the quantity represented by the data. The factors are those found in each factor column. For the special case of only one datum per cell, click the Model button. Select Custom and move the desired factors into Model.

SPSS Data Editor

File   Edit   View   Data   Transform   Analyze   Graphs

	fat	lab	brand
1	32.40	1	Texas
2	37.90	1	Great
3	29.10	1	Chip
4	33.10	2	Texas
5	37.70	2	Great
6	29.40	2	Chip
7	32.90	3	Texas
8	37.80	3	Great
9	29.50	3	Chip