Chapter 11

Chi-Square Tests and *F*-Tests

In previous chapters you saw how to test hypotheses concerning population means and population proportions. The idea of testing hypotheses can be extended to many other situations that involve different parameters and use different test statistics. Whereas the standardized test statistics that appeared in earlier chapters followed either a normal or Student *t*-distribution, in this chapter the tests will involve two other very common and useful distributions, the chi-square and the *F*-distributions. The **chi-square distribution**[1] arises in tests of hypotheses concerning the independence of two random variables and concerning whether a discrete random variable follows a specified distribution. The ***F*-distribution**[2] arises in tests of hypotheses concerning whether or not two population variances are equal and concerning whether or not three or more population means are equal.

[1] A particular probability distribution specified by a number of degrees of freedom, df.

[2] A particular probability distribution specified by two degrees of freedom, df_1 and df_2.

11.1 Chi-Square Tests for Independence

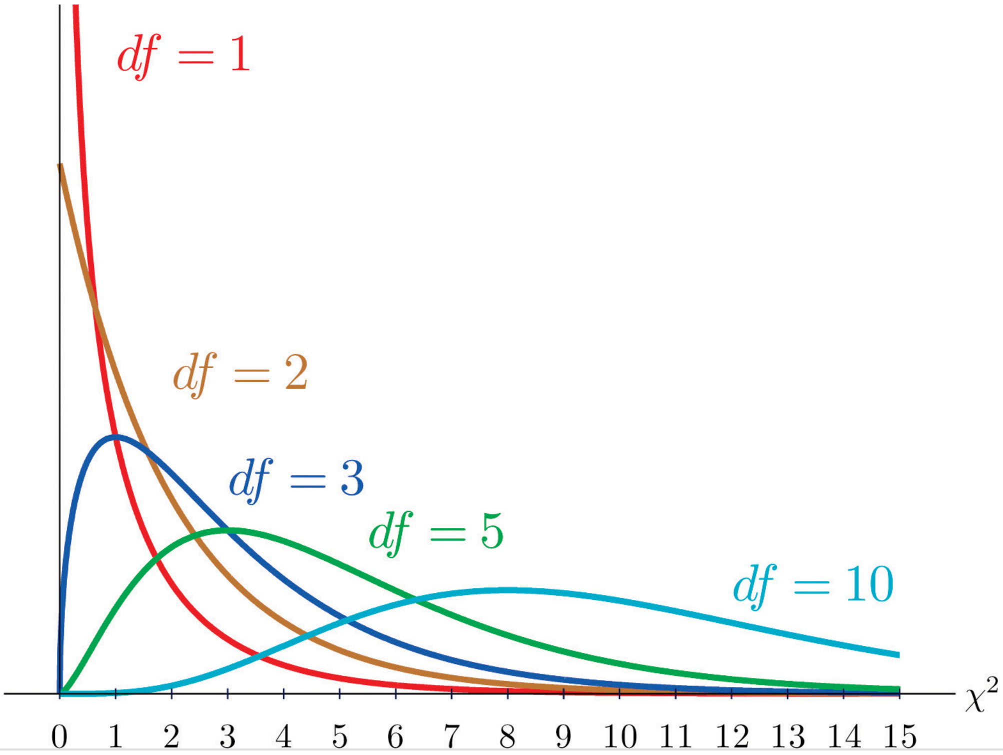
LEARNING OBJECTIVES

1. To understand what chi-square distributions are.
2. To understand how to use a chi-square test to judge whether two factors are independent.

Chi-Square Distributions

As you know, there is a whole family of *t*-distributions, each one specified by a parameter called the *degrees of freedom*, denoted df. Similarly, all the chi-square distributions form a family, and each of its members is also specified by a parameter df, the number of degrees of freedom. Chi is a Greek letter denoted by the symbol χ and chi-square is often denoted by χ^2. Figure 11.1 "Many " shows several chi-square distributions for different degrees of freedom. A **chi-square random variable**[3] is a random variable that assumes only positive values and follows a chi-square distribution.

Figure 11.1 Many χ^2 Distributions



[3] A random variable that follows a chi-square distribution.

Definition

The value of the chi-square random variable χ^2 with $df = k$ that cuts off a right tail of area c is denoted χ_c^2 and is called a **critical value**. *See Figure 11.2.*

Figure 11.2

χ_c^2 Illustrated

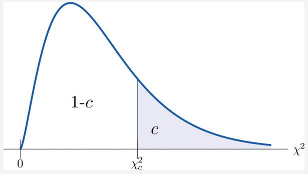


Figure 12.4 "Critical Values of Chi-Square Distributions" gives values of χ_c^2 for various values of *c* and under several chi-square distributions with various degrees of freedom.

Tests for Independence

Hypotheses tests encountered earlier in the book had to do with how the numerical values of two population parameters compared. In this subsection we will investigate hypotheses that have to do with whether or not two random variables take their values independently, or whether the value of one has a relation to the value of the other. Thus the hypotheses will be expressed in words, not mathematical symbols. We build the discussion around the following example.

There is a theory that the gender of a baby in the womb is related to the baby's heart rate: baby girls tend to have higher heart rates. Suppose we wish to test this theory. We examine the heart rate records of 40 babies taken during their mothers' last prenatal checkups before delivery, and to each of these 40 randomly selected records we compute the values of two random measures: 1) gender and 2) heart rate. In this context these two random measures are often called **factors**[4]. Since the burden of proof is that heart rate and gender are related, not that they are unrelated, the problem of testing the theory on baby gender and heart rate can be formulated as a test of the following hypotheses:

[4] A variable with several qualitative levels.

$$\begin{aligned} H_0 &: \text{Baby gender and baby heart rate are independent} \\ \text{vs. } H_a &: \text{Baby gender and baby heart rate are } \textit{not} \text{ independent} \end{aligned}$$

The factor gender has two natural categories or levels: boy and girl. We divide the second factor, heart rate, into two levels, low and high, by choosing some heart rate, say 145 beats per minute, as the cutoff between them. A heart rate below 145 beats per minute will be considered low and 145 and above considered high. The 40 records give rise to a 2 × 2 *contingency table*. By adjoining row totals, column totals, and a grand total we obtain the table shown as Table 11.1 "Baby Gender and Heart Rate". The four entries in boldface type are counts of observations from the sample of n = 40. There were 11 girls with low heart rate, 17 boys with low heart rate, and so on. They form the *core* of the expanded table.

Table 11.1 Baby Gender and Heart Rate

		Heart Rate		
		Low	High	Row Total
Gender	Girl	**11**	**7**	18
	Boy	**17**	**5**	22
Column Total		28	12	Total = 40

In analogy with the fact that the probability of independent events is the product of the probabilities of each event, if heart rate and gender were independent then we would expect the number in each core cell to be close to the product of the row total R and column total C of the row and column containing it, divided by the sample size n. Denoting such an expected number of observations E, these four expected values are:

- 1st row and 1st column: $E = (R \times C) \,/\, n = 18 \times 28 \,/\, 40 = 12.6$
- 1st row and 2nd column: $E = (R \times C) \,/\, n = 18 \times 12 \,/\, 40 = 5.4$
- 2nd row and 1st column: $E = (R \times C) \,/\, n = 22 \times 28 \,/\, 40 = 15.4$
- 2nd row and 2nd column: $E = (R \times C) \,/\, n = 22 \times 12 \,/\, 40 = 6.6$

We update Table 11.1 "Baby Gender and Heart Rate" by placing each expected value in its corresponding core cell, right under the observed value in the cell. This gives the updated table Table 11.2 "Updated Baby Gender and Heart Rate".

Table 11.2 Updated Baby Gender and Heart Rate

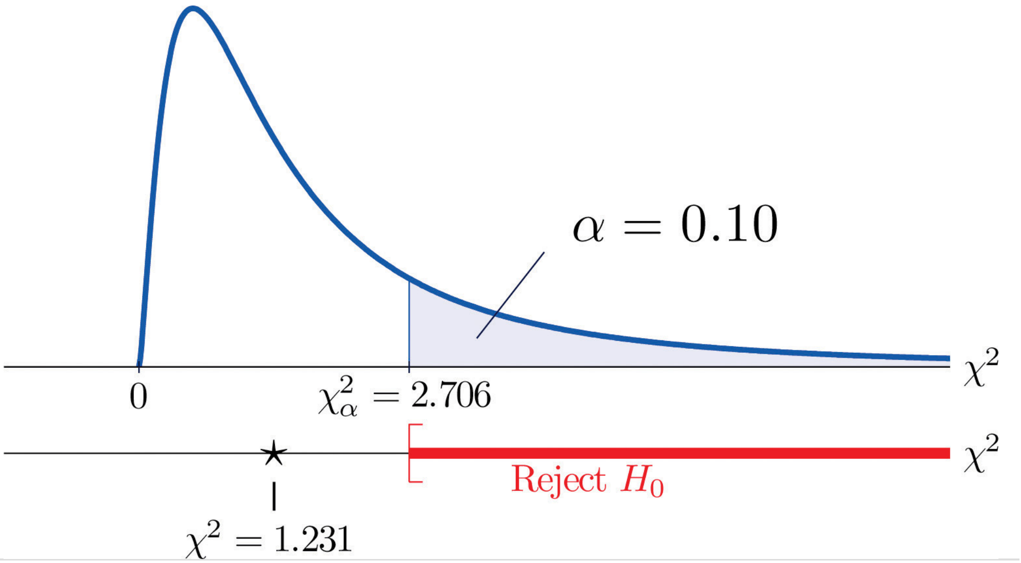
		Heart Rate		
		Low	High	Row Total
Gender	Girl	$O = 11$ $E = 12.6$	$O = 7$ $E = 5.4$	$R = 18$
	Boy	$O = 17$ $E = 15.4$	$O = 5$ $E = 6.6$	$R = 22$
Column Total		$C = 28$	$C = 12$	$n = 40$

A measure of how much the data deviate from what we would expect to see if the factors really were independent is the sum of the squares of the difference of the numbers in each core cell, or, standardizing by dividing each square by the expected number in the cell, the sum $\Sigma(O - E)^2 \,/\, E$. We would reject the null hypothesis that the factors are independent only if this number is large, so the test is right-tailed. In this example the random variable $\Sigma(O - E)^2 \,/\, E$ has the chi-square distribution with one degree of freedom. If we had decided at the outset to test at the 10% level of significance, the critical value defining the rejection region would be, reading from Figure 12.4 "Critical Values of Chi-Square Distributions", $\chi^2_\alpha = \chi^2_{0.10} = 2.706$, so that the rejection region would be the interval $\left[2.706, \infty\right)$. When we compute the value of the standardized test statistic we obtain

$$\Sigma \frac{(O - E)^2}{E} = \frac{\left(11 - 12.6\right)^2}{12.6} + \frac{\left(7 - 5.4\right)^2}{5.4} + \frac{\left(17 - 15.4\right)^2}{15.4} + \frac{(5 - 6.6}{6.6}$$

Since 1.231 < 2.706, the decision is not to reject H_0. See Figure 11.3 "Baby Gender Prediction". The data do not provide sufficient evidence, at the 10% level of significance, to conclude that heart rate and gender are related.

Figure 11.3 Baby Gender Prediction



With this specific example in mind, now turn to the general situation. In the general setting of testing the independence of two factors, call them *Factor 1* and *Factor 2*, the hypotheses to be tested are

$$\begin{aligned} H_0 &: \text{The two factors are independent} \\ \text{vs. } H_a &: \text{The two factors are } \textit{not} \text{ independent} \end{aligned}$$

As in the example each factor is divided into a number of categories or levels. These could arise naturally, as in the boy-girl division of gender, or somewhat arbitrarily, as in the high-low division of heart rate. Suppose Factor 1 has I levels and Factor 2 has J levels. Then the information from a random sample gives rise to a general $I \times J$ contingency table, which with row totals, column totals, and a grand total would appear as shown in Table 11.3 "General Contingency Table". Each cell may be labeled by a pair of indices $\left(i, j\right)$. O_{ij} stands for the observed count of observations in the cell in row i and column j, R_i for the i^{th} row total and C_j for the j^{th} column total. To simplify the notation we will drop the indices so Table 11.3 "General Contingency Table" becomes Table 11.4 "Simplified General Contingency Table". Nevertheless it is important to keep in mind that the Os, the Rs and the Cs, though denoted by the same symbols, are in fact different numbers.

Table 11.3 General Contingency Table

		Factor 2 Levels					
		1	· · ·	***j***	· · ·	***J***	**Row Total**
Factor 1 Levels	1	O_{11}	· · ·	O_{1j}	· · ·	O_{1J}	R_1
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	i	O_{i1}	· · ·	O_{ij}	· · ·	O_{iJ}	R_i
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	I	O_{I1}	· · ·	O_{Ij}	· · ·	O_{IJ}	R_I
Column Total		C_1	· · ·	C_j	· · ·	C_J	n

Table 11.4 Simplified General Contingency Table

		Factor 2 Levels					
		1	· · ·	***j***	· · ·	***J***	**Row Total**
Factor 1 Levels	1	O	· · ·	O	· · ·	O	R
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	i	O	· · ·	O	· · ·	O	R
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	I	O	· · ·	O	· · ·	O	R
Column Total		C	· · ·	C	· · ·	C	n

As in the example, for each core cell in the table we compute what would be the *expected number* E of observations if the two factors were independent. E is computed for each core cell (each cell with an O in it) of Table 11.4 "Simplified General Contingency Table" by the rule applied in the example:

$$E = \frac{R \times C}{n}$$

where R is the row total and C is the column total corresponding to the cell, and n is the sample size.

After the expected number is computed for every cell, Table 11.4 "Simplified General Contingency Table" is updated to form Table 11.5 "Updated General Contingency Table" by inserting the computed value of E into each core cell.

Table 11.5 Updated General Contingency Table

		Factor 2 Levels					
		1	**· · ·**	***j***	**· · ·**	***J***	**Row Total**
Factor 1 Levels	1	O E	· · ·	O E	· · ·	O E	R
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	i	O E	· · ·	O E	· · ·	O E	R
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	I	O E	· · ·	O E	· · ·	O E	R
Column Total		C	· · ·	C	· · ·	C	n

Here is the test statistic for the general hypothesis based on Table 11.5 "Updated General Contingency Table", together with the conditions that it follow a chi-square distribution.

Test Statistic for Testing the Independence of Two Factors

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

where the sum is over all core cells of the table.

If

1. the two study factors are independent, and
2. the observed count O of each cell in Table 11.5 "Updated General Contingency Table" is at least 5,

then χ^2 approximately follows a chi-square distribution with $df = (I-1) \times (J-1)$ degrees of freedom.

The same five-step procedures, either the critical value approach or the *p*-value approach, that were introduced in Section 8.1 "The Elements of Hypothesis Testing" and Section 8.3 "The Observed Significance of a Test" of Chapter 8 "Testing Hypotheses" are used to perform the test, which is always right-tailed.

EXAMPLE 1

A researcher wishes to investigate whether students' scores on a college entrance examination (CEE) have any indicative power for future college performance as measured by GPA. In other words, he wishes to investigate whether the factors CEE and GPA are independent or not. He randomly selects $n = 100$ students in a college and notes each student's score on the entrance examination and his grade point average at the end of the sophomore year. He divides entrance exam scores into two levels and grade point averages into three levels. Sorting the data according to these divisions, he forms the contingency table shown as Table 11.6 "CEE versus GPA Contingency Table", in which the row and column totals have already been computed.

TABLE 11.6 CEE VERSUS GPA CONTINGENCY TABLE

		GPA			
		<2.7	**2.7 to 3.2**	**>3.2**	**Row Total**
CEE	< 1800	**35**	**12**	**5**	52
	≥ 1800	**6**	**24**	**18**	48
Column Total		41	36	23	$\text{Total} = 100$

Test, at the 1% level of significance, whether these data provide sufficient evidence to conclude that CEE scores indicate future performance levels of incoming college freshmen as measured by GPA.

Solution:

We perform the test using the critical value approach, following the usual five-step method outlined at the end of Section 8.1 "The Elements of Hypothesis Testing" in Chapter 8 "Testing Hypotheses".

- Step 1. The hypotheses are

$$\begin{aligned} H_0 &: \text{CEE and GPA are independent factors} \\ \text{vs. } H_a &: \text{CEE and GPA are not independent factors} \end{aligned}$$

- Step 2. The distribution is chi-square.
  - Step 3. To compute the value of the test statistic we must first computed the expected number for each of the six core cells (the ones whose entries are boldface):
    - 1st row and 1st column: $E = (R \times C) / n = 41 \times 52 / 100 = 21.32$
    - 1st row and 2nd column: $E = (R \times C) / n = 36 \times 52 / 100 = 18.72$
    - 1st row and 3rd column: $E = (R \times C) / n = 23 \times 52 / 100 = 11.96$
    - 2nd row and 1st column: $E = (R \times C) / n = 41 \times 48 / 100 = 19.68$
    - 2nd row and 2nd column: $E = (R \times C) / n = 36 \times 48 / 100 = 17.28$
    - 2nd row and 3rd column: $E = (R \times C) / n = 23 \times 48 / 100 = 11.04$

    Table 11.6 "CEE versus GPA Contingency Table" is updated to Table 11.7 "Updated CEE versus GPA Contingency Table".

TABLE 11.7 UPDATED CEE VERSUS GPA CONTINGENCY TABLE

		GPA			
		<2.7	**2.7 to 3.2**	**>3.2**	**Row Total**
CEE	< 1800	$O = 35$ $E = 21.32$	$O = 12$ $E = 18.72$	$O = 5$ $E = 11.96$	*R* = 52
	≥ 1800	$O = 6$ $E = 19.68$	$O = 24$ $E = 17.28$	$O = 18$ $E = 11.04$	*R* = 48
Column Total		*C* = 41	*C* = 36	*C* = 23	*n* = 100

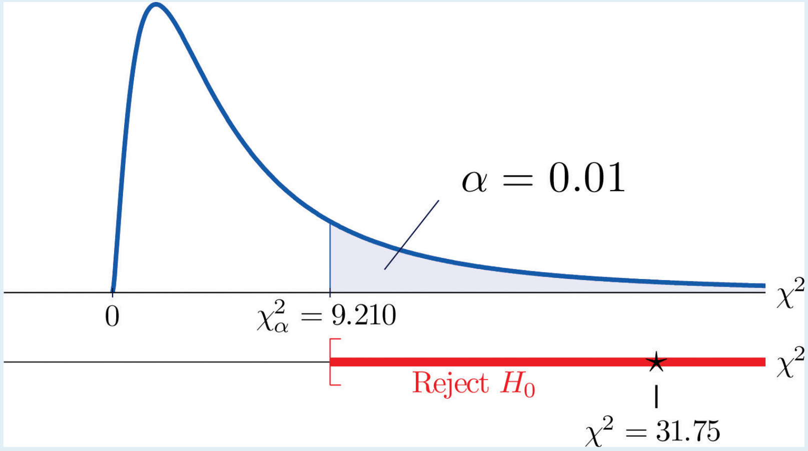
The test statistic is

$$\begin{aligned}\chi^2 &= \Sigma \frac{(O-E)^2}{E} \\ &= \frac{(35-21.32)^2}{21.32} + \frac{(12-18.72)^2}{18.72} + \frac{(5-11.96)^2}{11.96} \\ &\quad + \frac{(6-19.68)^2}{19.68} + \frac{(24-17.28)^2}{17.28} + \frac{(18-11.04)^2}{11.04} \\ &= 31.75\end{aligned}$$

- Step 4. Since the CEE factor has two levels and the GPA factor has three, $I = 2$ and $J = 3$. Thus the test statistic follows the chi-square distribution with $df = (2-1) \times (3-1) = 2$degrees of freedom.

  Since the test is right-tailed, the critical value is $\chi^2_{0.01}$. Reading from Figure 12.4 "Critical Values of Chi-Square Distributions", $\chi^2_{0.01} = 9.210$, so the rejection region is $[9.210, \infty)$.
- Step 5. Since 31.75 > 9.21 the decision is to reject the null hypothesis. See Figure 11.4. The data provide sufficient evidence, at the 1% level of significance, to conclude that CEE score and GPA are not independent: the entrance exam score has predictive power.

Figure 11.4
Note 11.9 "Example 1"



KEY TAKEAWAYS

- Critical values of a chi-square distribution with degrees of freedom df are found in Figure 12.4 "Critical Values of Chi-Square Distributions".
- A **chi-square test**[5] can be used to evaluate the hypothesis that two random variables or factors are independent.

[5] A test based on a chi-square statistic to check whether two factors are independent.

EXERCISES

BASIC

1. Find $\chi^2_{0.01}$ for each of the following number of degrees of freedom.

   a. $df = 5$
   b. $df = 11$
   c. $df = 25$

2. Find $\chi^2_{0.05}$ for each of the following number of degrees of freedom.

   a. $df = 6$
   b. $df = 12$
   c. $df = 30$

3. Find $\chi^2_{0.10}$ for each of the following number of degrees of freedom.

   a. $df = 6$
   b. $df = 12$
   c. $df = 30$

4. Find $\chi^2_{0.01}$ for each of the following number of degrees of freedom.

   a. $df = 7$
   b. $df = 10$
   c. $df = 20$

5. For $df = 7$ and $\alpha = 0.05$, find

   a. χ^2_{α}
   b. $\chi^2_{\frac{\alpha}{2}}$

6. For $df = 17$ and $\alpha = 0.01$, find

   a. χ^2_{α}
   b. $\chi^2_{\frac{\alpha}{2}}$

7. A data sample is sorted into a 2 × 2 contingency table based on two factors, each of which has two levels.

		Factor 1		
		Level 1	Level 2	Row Total
Factor 2	Level 1	20	10	*R*
	Level 2	15	5	*R*
Column Total		*C*	*C*	*n*

a. Find the column totals, the row totals, and the grand total, *n*, of the table.
b. Find the expected number *E* of observations for each cell based on the assumption that the two factors are independent (that is, just use the formula $E = (R \times C) / n$).
c. Find the value of the chi-square test statistic χ^2.
d. Find the number of degrees of freedom of the chi-square test statistic.

8. A data sample is sorted into a 3 × 2 contingency table based on two factors, one of which has three levels and the other of which has two levels.

		Factor 1		
		Level 1	Level 2	Row Total
Factor 2	Level 1	20	10	*R*
	Level 2	15	5	*R*
	Level 3	10	20	*R*
Column Total		*C*	*C*	*n*

a. Find the column totals, the row totals, and the grand total, *n*, of the table.
b. Find the expected number *E* of observations for each cell based on the assumption that the two factors are independent (that is, just use the formula $E = (R \times C) / n$).
c. Find the value of the chi-square test statistic χ^2.
d. Find the number of degrees of freedom of the chi-square test statistic.

APPLICATIONS

9. A child psychologist believes that children perform better on tests when they are given perceived freedom of choice. To test this belief, the psychologist carried out an experiment in which 200 third graders were randomly assigned to two groups, *A* and *B*. Each child was given the same simple logic test. However in group *B*, each child was given the freedom to choose a text booklet

from many with various drawings on the covers. The performance of each child was rated as Very Good, Good, and Fair. The results are summarized in the table provided. Test, at the 5% level of significance, whether there is sufficient evidence in the data to support the psychologist's belief.

		Group	
		A	***B***
Performance	Very Good	32	29
	Good	55	61
	Fair	10	13

10. In regard to wine tasting competitions, many experts claim that the first glass of wine served sets a reference taste and that a different reference wine may alter the relative ranking of the other wines in competition. To test this claim, three wines, *A*, *B* and *C*, were served at a wine tasting event. Each person was served a single glass of each wine, but in different orders for different guests. At the close, each person was asked to name the best of the three. One hundred seventy-two people were at the event and their top picks are given in the table provided. Test, at the 1% level of significance, whether there is sufficient evidence in the data to support the claim that wine experts' preference is dependent on the first served wine.

		Top Pick		
		A	***B***	***C***
First Glass	*A*	12	31	27
	B	15	40	21
	C	10	9	7

11. Is being left-handed hereditary? To answer this question, 250 adults are randomly selected and their handedness and their parents' handedness are noted. The results are summarized in the table provided. Test, at the 1% level of significance, whether there is sufficient evidence in the data to conclude that there is a hereditary element in handedness.

		Number of Parents Left-Handed		
		0	**1**	**2**
Handedness	Left	8	10	12
	Right	178	21	21

12. Some geneticists claim that the genes that determine left-handedness also govern development of the language centers of the brain. If this claim is true, then it would be reasonable to expect that left-handed people tend to have stronger language abilities. A study designed to text this claim randomly selected 807 students who took the Graduate Record Examination (GRE). Their scores on the language portion of the examination were classified into three categories: *low*, *average*, and *high*, and their handedness was also noted. The results are given in the table provided. Test, at the 5% level of significance, whether there is sufficient evidence in the data to conclude that left-handed people tend to have stronger language abilities.

		GRE English Scores		
		Low	**Average**	**High**
Handedness	Left	18	40	22
	Right	201	360	166

13. It is generally believed that children brought up in stable families tend to do well in school. To verify such a belief, a social scientist examined 290 randomly selected students' records in a public high school and noted each student's family structure and academic status four years after entering high school. The data were then sorted into a 2 × 3 contingency table with two factors. Factor 1 has two levels: *graduated* and *did not graduate*. Factor 2 has three levels: *no parent*, *one parent*, and *two parents*. The results are given in the table provided. Test, at the 1% level of significance, whether there is sufficient evidence in the data to conclude that family structure matters in school performance of the students.

		Academic Status	
		Graduated	**Did Not Graduate**
Family	No parent	18	31
	One parent	101	44
	Two parents	70	26

14. A large middle school administrator wishes to use celebrity influence to encourage students to make healthier choices in the school cafeteria. The cafeteria is situated at the center of an open space. Everyday at lunch time students get their lunch and a drink in three separate lines leading to three separate serving stations. As an experiment, the school administrator displayed a poster of a popular teen pop star drinking milk at each of the three areas where drinks are provided, except the milk in the poster is different at each location: one shows white milk, one shows strawberry-flavored pink milk,

and one shows chocolate milk. After the first day of the experiment the administrator noted the students' milk choices separately for the three lines. The data are given in the table provided. Test, at the 1% level of significance, whether there is sufficient evidence in the data to conclude that the posters had some impact on the students' drink choices.

	Student Choice		
	Regular	Strawberry	Chocolate
Poster Choice			
Regular	38	28	40
Strawberry	18	51	24
Chocolate	32	32	53

LARGE DATA SET EXERCISE

15. Large Data Set 8 records the result of a survey of 300 randomly selected adults who go to movie theaters regularly. For each person the gender and preferred type of movie were recorded. Test, at the 5% level of significance, whether there is sufficient evidence in the data to conclude that the factors "gender" and "preferred type of movie" are dependent.

    http://www.gone.2012books.lardbucket.org/sites/all/files/data8.xls

ANSWERS

1. a. 15.09,
   b. 24.72,
   c. 44.31

3. a. 10.64,
   b. 18.55,
   c. 40.26

5. a. 14.07,
   b. 16.01

7. a. $C_1 = 35, C_2 = 15, R_1 = 30, R_2 = 20, n = 50$,
   b. $E_{11} = 21, E_{12} = 9, E_{21} = 14, E_{22} = 6$,
   c. $\chi^2 = 0.3968$,
   d. $df = 1$

9. $\chi^2 = 0.6698, \chi^2_{0.05} = 5.99$, do not reject H_0

11. $\chi^2 = 72.35, \chi^2_{0.01} = 9.21$, reject H_0

13. $\chi^2 = 21.2784, \chi^2_{0.01} = 9.21$, reject H_0

15. $\chi^2 = 28.4539$. $df = 3$. Rejection Region: $[7.815, \infty)$. Decision: Reject H_0 of independence.

11.2 Chi-Square One-Sample Goodness-of-Fit Tests

LEARNING OBJECTIVE

1. To understand how to use a chi-square test to judge whether a sample fits a particular population well.

Suppose we wish to determine if an ordinary-looking six-sided die is fair, or balanced, meaning that every face has probability 1/6 of landing on top when the die is tossed. We could toss the die dozens, maybe hundreds, of times and compare the actual number of times each face landed on top to the expected number, which would be 1/6 of the total number of tosses. We wouldn't expect each number to be exactly 1/6 of the total, but it should be close. To be specific, suppose the die is tossed n = 60 times with the results summarized in Table 11.8 "Die Contingency Table". For ease of reference we add a column of expected frequencies, which in this simple example is simply a column of 10s. The result is shown as Table 11.9 "Updated Die Contingency Table". In analogy with the previous section we call this an "updated" table. A measure of how much the data deviate from what we would expect to see if the die really were fair is the sum of the squares of the differences between the observed frequency O and the expected frequency E in each row, or, standardizing by dividing each square by the expected number, the sum $\Sigma(O - E)^2 \Big/ E$. If we formulate the investigation as a test of hypotheses, the test is

$$H_0 : \text{The die is fair}$$
$$\text{vs. } H_a : \text{The die is } \textit{not} \text{ fair}$$

Table 11.8 Die Contingency Table

Die Value	Assumed Distribution	Observed Frequency
1	1/6	9
2	1/6	15
3	1/6	9
4	1/6	8
5	1/6	6
6	1/6	13

Table 11.9 Updated Die Contingency Table

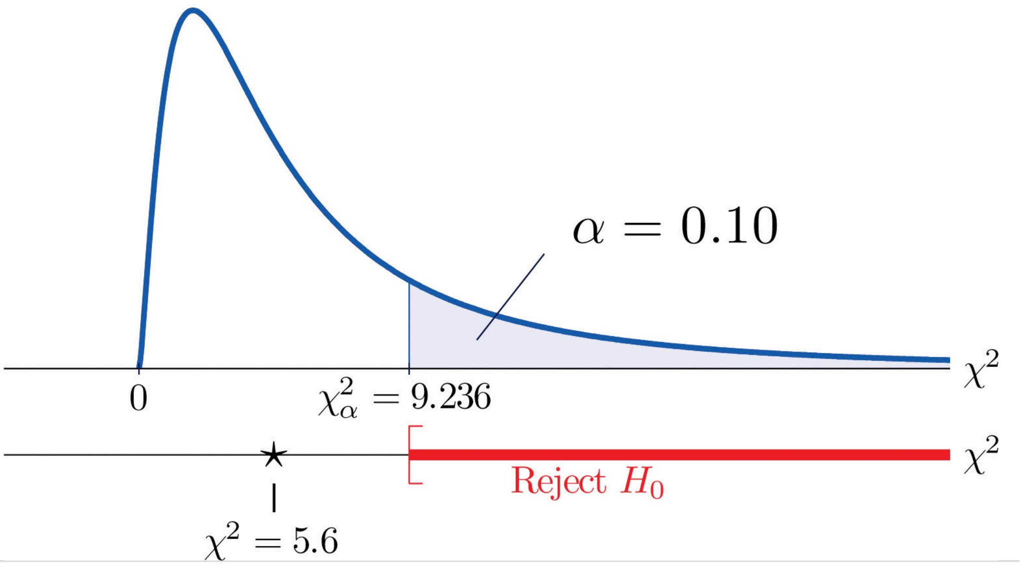
Die Value	Assumed Distribution	Observed Freq.	Expected Freq.
1	1/6	9	10
2	1/6	15	10
3	1/6	9	10
4	1/6	8	10
5	1/6	6	10
6	1/6	13	10

We would reject the null hypothesis that the die is fair only if the number $\Sigma(O-E)^2 \big/ E$ is large, so the test is right-tailed. In this example the random variable $\Sigma(O-E)^2 \big/ E$ has the chi-square distribution with five degrees of freedom. If we had decided at the outset to test at the 10% level of significance, the critical value defining the rejection region would be, reading from Figure 12.4 "Critical Values of Chi-Square Distributions", $\chi^2_\alpha = \chi^2_{0.10} = 9.236$, so that the rejection region would be the interval $\left[9.236, \infty\right)$. When we compute the value of the standardized test statistic using the numbers in the last two columns of Table 11.9 "Updated Die Contingency Table", we obtain

$$\begin{aligned}
&\Sigma \frac{(O-E)^2}{E} \\
&= \frac{(-1)^2}{10} + \frac{5^2}{10} + \frac{(-1)^2}{10} + \frac{(-2)^2}{10} + \frac{(-4)^2}{10} + \frac{3^2}{10} \\
&= 0.1 + 2.5 + 0.1 + 0.4 + 1.6 + 0.9 \\
&= 5.6
\end{aligned}$$

Since 5.6 < 9.236 the decision is not to reject H_0. See Figure 11.5 "Balanced Die". The data do not provide sufficient evidence, at the 10% level of significance, to conclude that the die is loaded.

Figure 11.5 Balanced Die



In the general situation we consider a discrete random variable that can take I different values, $x_1, x_2, \ldots, x_I$, for which the default assumption is that the probability distribution is

x	x_1	x_2	$\ldots$	x_I
$P(x)$	p_1	p_2	$\ldots$	p_I

We wish to test the hypotheses

$$\begin{aligned} H_0 &: \text{The assumed probability distribution for } X \text{ is valid} \\ \text{vs. } H_a &: \text{The assumed probability distribution for } X \text{ is } \textit{not} \text{ valid} \end{aligned}$$

We take a sample of size n and obtain a list of observed frequencies. This is shown in Table 11.10 "General Contingency Table". Based on the assumed probability distribution we also have a list of assumed frequencies, each of which is defined and computed by the formula

$$E_i = n \times p_i$$

Table 11.10 General Contingency Table

Factor Levels	Assumed Distribution	Observed Frequency
1	p_1	O_1
2	p_2	O_2
$\vdots$	$\vdots$	$\vdots$
I	p_I	O_I

Table 11.10 "General Contingency Table" is updated to Table 11.11 "Updated General Contingency Table" by adding the expected frequency for each value of X. To simplify the notation we drop indices for the observed and expected frequencies and represent Table 11.11 "Updated General Contingency Table" by Table 11.12 "Simplified Updated General Contingency Table".

Table 11.11 Updated General Contingency Table

Factor Levels	Assumed Distribution	Observed Freq.	Expected Freq.
1	p_1	O_1	E_1
2	p_2	O_2	E_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	p_I	O_I	E_I

Table 11.12 Simplified Updated General Contingency Table

Factor Levels	Assumed Distribution	Observed Freq.	Expected Freq.
1	p_1	O	E
2	p_2	O	E
$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	p_I	O	E

Here is the test statistic for the general hypothesis based on Table 11.12 "Simplified Updated General Contingency Table", together with the conditions that it follow a chi-square distribution.

Test Statistic for Testing Goodness of Fit to a Discrete Probability Distribution

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

where the sum is over all the rows of the table (one for each value of X).

If

1. the true probability distribution of X is as assumed, and
2. the observed count O of each cell in Table 11.12 "Simplified Updated General Contingency Table" is at least 5,

then χ^2 approximately follows a chi-square distribution with $df = I-1$ degrees of freedom.

The test is known as a *goodness-of-fit* χ^2 test since it tests the null hypothesis that the sample fits the assumed probability distribution well. It is always right-tailed, since deviation from the assumed probability distribution corresponds to large values of χ^2.

Testing is done using either of the usual five-step procedures.

EXAMPLE 2

Table 11.13 "Ethnic Groups in the Census Year" shows the distribution of various ethnic groups in the population of a particular state based on a decennial U.S. census. Five years later a random sample of 2,500 residents of the state was taken, with the results given in Table 11.14 "Sample Data Five Years After the Census Year" (along with the probability distribution from the census year). Test, at the 1% level of significance, whether there is sufficient evidence in the sample to conclude that the distribution of ethnic groups in this state five years after the census had changed from that in the census year.

TABLE 11.13 ETHNIC GROUPS IN THE CENSUS YEAR

Ethnicity	White	Black	Amer.-Indian	Hispanic	Asian	Others
Proportion	0.743	0.216	0.012	0.012	0.008	0.009

TABLE 11.14 SAMPLE DATA FIVE YEARS AFTER THE CENSUS YEAR

Ethnicity	Assumed Distribution	Observed Frequency
White	0.743	1732
Black	0.216	538
American-Indian	0.012	32
Hispanic	0.012	42
Asian	0.008	133
Others	0.009	23

Solution:

We test using the critical value approach.

- Step 1. The hypotheses of interest in this case can be expressed as

$$H_0 : \text{The distribution of ethnic groups has not changed}$$
$$\text{vs. } H_a : \text{The distribution of ethnic groups } \textit{has} \text{ changed}$$

- Step 2. The distribution is chi-square.
  - Step 3. To compute the value of the test statistic we must first compute the expected number for each row of Table 11.14 "Sample Data Five Years After the Census Year". Since $n = 2500$, using the formula $E_i = n \times p_i$ and the values of p_i from either Table 11.13 "Ethnic Groups in the Census Year" or Table 11.14 "Sample Data Five Years After the Census Year",

$$\begin{aligned} E_1 &= 2500 \times 0.743 = 1857.5 \\ E_2 &= 2500 \times 0.216 = 540 \\ E_3 &= 2500 \times 0.012 = 30 \\ E_4 &= 2500 \times 0.012 = 30 \\ E_5 &= 2500 \times 0.008 = 20 \\ E_6 &= 2500 \times 0.009 = 22.5 \end{aligned}$$

Table 11.14 "Sample Data Five Years After the Census Year" is updated to Table 11.15 "Observed and Expected Frequencies Five Years After the Census Year".

TABLE 11.15 OBSERVED AND EXPECTED FREQUENCIES FIVE YEARS AFTER THE CENSUS YEAR

Ethnicity	Assumed Dist.	Observed Freq.	Expected Freq.
White	0.743	1732	1857.5
Black	0.216	538	540
American-Indian	0.012	32	30
Hispanic	0.012	42	30
Asian	0.008	133	20
Others	0.009	23	22.5

The value of the test statistic is

$$\begin{aligned}\chi^2 &= \Sigma \frac{(O-E)^2}{E} \\ &= \frac{\left(1732 - 1857.5\right)^2}{1857.5} + \frac{\left(538 - 540\right)^2}{540} + \frac{(32-30)^2}{30} + \frac{(42}{} \\ &\qquad + \frac{(133-20)^2}{20} + \frac{\left(23 - 2\right.}{22.} \\ &= 651.881\end{aligned}$$

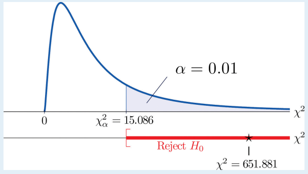
- Since the random variable takes six values, $I = 6$. Thus the test statistic follows the chi-square distribution with $df = 6 - 1 = 5$ degrees of freedom.

  Since the test is right-tailed, the critical value is $\chi^2_{0.01}$. Reading from Figure 12.4 "Critical Values of Chi-Square Distributions", $\chi^2_{0.01} = 15.086$, so the rejection region is $\left[15.086, \infty\right)$.

- Since 651.881 > 15.086 the decision is to reject the null hypothesis. See Figure 11.6. The data provide sufficient evidence, at the 1% level of significance, to conclude that the ethnic distribution in this state has changed in the five years since the U.S. census.

Figure 11.6

Note 11.15 "Example 2"



KEY TAKEAWAY

- The **chi-square goodness-of-fit test**[6] can be used to evaluate the hypothesis that a sample is taken from a population with an assumed specific probability distribution.

[6] A test based on a chi-square statistic to check whether a sample is taken from a population with a hypothesized probability distribution.

EXERCISES

BASIC

1. A data sample is sorted into five categories with an assumed probability distribution.

Factor Levels	Assumed Distribution	Observed Frequency
1	$p_1 = 0.1$	10
2	$p_2 = 0.4$	35
3	$p_3 = 0.4$	45
4	$p_4 = 0.1$	10

   a. Find the size *n* of the sample.
   b. Find the expected number *E* of observations for each level, if the sampled population has a probability distribution as assumed (that is, just use the formula $E_i = n \times p_i$).
   c. Find the chi-square test statistic χ^2.
   d. Find the number of degrees of freedom of the chi-square test statistic.

2. A data sample is sorted into five categories with an assumed probability distribution.

Factor Levels	Assumed Distribution	Observed Frequency
1	$p_1 = 0.3$	23
2	$p_2 = 0.3$	30
3	$p_3 = 0.2$	19
4	$p_4 = 0.1$	8
5	$p_5 = 0.1$	10

   a. Find the size *n* of the sample.
   b. Find the expected number *E* of observations for each level, if the sampled population has a probability distribution as assumed (that is, just use the formula $E_i = n \times p_i$).
   c. Find the chi-square test statistic χ^2.
   d. Find the number of degrees of freedom of the chi-square test statistic.

APPLICATIONS

3. Retailers of collectible postage stamps often buy their stamps in large quantities by weight at auctions. The prices the retailers are willing to pay depend on how old the postage stamps are. Many collectible postage stamps at auctions are described by the proportions of stamps issued at various periods in the past. Generally the older the stamps the higher the value. At one particular auction, a lot of collectible stamps is advertised to have the age distribution given in the table provided. A retail buyer took a sample of 73 stamps from the lot and sorted them by age. The results are given in the table provided. Test, at the 5% level of significance, whether there is sufficient evidence in the data to conclude that the age distribution of the lot is different from what was claimed by the seller.

Year	Claimed Distribution	Observed Frequency
Before 1940	0.10	6
1940 to 1959	0.25	15
1960 to 1979	0.45	30
After 1979	0.20	22

4. The litter size of Bengal tigers is typically two or three cubs, but it can vary between one and four. Based on long-term observations, the litter size of Bengal tigers in the wild has the distribution given in the table provided. A zoologist believes that Bengal tigers in captivity tend to have different (possibly smaller) litter sizes from those in the wild. To verify this belief, the zoologist searched all data sources and found 316 litter size records of Bengal tigers in captivity. The results are given in the table provided. Test, at the 5% level of significance, whether there is sufficient evidence in the data to conclude that the distribution of litter sizes in captivity differs from that in the wild.

Litter Size	Wild Litter Distribution	Observed Frequency
1	0.11	41
2	0.69	243
3	0.18	27
4	0.02	5

5. An online shoe retailer sells men's shoes in sizes 8 to 13. In the past orders for the different shoe sizes have followed the distribution given in the table

provided. The management believes that recent marketing efforts may have expanded their customer base and, as a result, there may be a shift in the size distribution for future orders. To have a better understanding of its future sales, the shoe seller examined 1,040 sales records of recent orders and noted the sizes of the shoes ordered. The results are given in the table provided. Test, at the 1% level of significance, whether there is sufficient evidence in the data to conclude that the shoe size distribution of future sales will differ from the historic one.

Shoe Size	Past Size Distribution	Recent Size Frequency
8.0	0.03	25
8.5	0.06	43
9.0	0.09	88
9.5	0.19	221
10.0	0.23	272
10.5	0.14	150
11.0	0.10	107
11.5	0.06	51
12.0	0.05	37
12.5	0.03	35
13.0	0.02	11

6. An online shoe retailer sells women’s shoes in sizes 5 to 10. In the past orders for the different shoe sizes have followed the distribution given in the table provided. The management believes that recent marketing efforts may have expanded their customer base and, as a result, there may be a shift in the size distribution for future orders. To have a better understanding of its future sales, the shoe seller examined 1,174 sales records of recent orders and noted the sizes of the shoes ordered. The results are given in the table provided. Test, at the 1% level of significance, whether there is sufficient evidence in the data to conclude that the shoe size distribution of future sales will differ from the historic one.

Shoe Size	Past Size Distribution	Recent Size Frequency
5.0	0.02	20
5.5	0.03	23

Shoe Size	Past Size Distribution	Recent Size Frequency
6.0	0.07	88
6.5	0.08	90
7.0	0.20	222
7.5	0.20	258
8.0	0.15	177
8.5	0.11	121
9.0	0.08	91
9.5	0.04	53
10.0	0.02	31

7. A chess opening is a sequence of moves at the beginning of a chess game. There are many well-studied named openings in chess literature. French Defense is one of the most popular openings for black, although it is considered a relatively weak opening since it gives black probability 0.344 of winning, probability 0.405 of losing, and probability 0.251 of drawing. A chess master believes that he has discovered a new variation of French Defense that may alter the probability distribution of the outcome of the game. In his many Internet chess games in the last two years, he was able to apply the new variation in 77 games. The wins, losses, and draws in the 77 games are given in the table provided. Test, at the 5% level of significance, whether there is sufficient evidence in the data to conclude that the newly discovered variation of French Defense alters the probability distribution of the result of the game.

Result for Black	Probability Distribution	New Variation Wins
Win	0.344	31
Loss	0.405	25
Draw	0.251	21

8. The Department of Parks and Wildlife stocks a large lake with fish every six years. It is determined that a healthy diversity of fish in the lake should consist of 10% largemouth bass, 15% smallmouth bass, 10% striped bass, 10% trout, and 20% catfish. Therefore each time the lake is stocked, the fish population in the lake is restored to maintain that particular distribution. Every three years, the department conducts a study to see whether the distribution of the fish in the lake has shifted away from the target proportions. In one particular year, a

research group from the department observed a sample of 292 fish from the lake with the results given in the table provided. Test, at the 5% level of significance, whether there is sufficient evidence in the data to conclude that the fish population distribution has shifted since the last stocking.

Fish	Target Distribution	Fish in Sample
Largemouth Bass	0.10	14
Smallmouth Bass	0.15	49
Striped Bass	0.10	21
Trout	0.10	22
Catfish	0.20	75
Other	0.35	111

LARGE DATA SET EXERCISE

9. Large Data Set 4 records the result of 500 tosses of six-sided die. Test, at the 10% level of significance, whether there is sufficient evidence in the data to conclude that the die is not “fair” (or “balanced”), that is, that the probability distribution differs from probability 1/6 for each of the six faces on the die.

   http://www.gone.2012books.lardbucket.org/sites/all/files/data4.xls

ANSWERS

1.
   a. $n = 100$,
   b. $E = 10, E = 40, E = 40, E = 10$;
   c. $\chi^2 = 1.25$,
   d. $df = 3$

3. $\chi^2 = 4.8082, \chi^2_{0.05} = 7.81$, do not reject H_0

5. $\chi^2 = 26.5765, \chi^2_{0.01} = 23.21$, reject H_0

7. $\chi^2 = 2.1401, \chi^2_{0.05} = 5.99$, do not reject H_0

9. $\chi^2 = 2.944.\ df = 5.$ Rejection Region: $[9.236, \infty)$. Decision: Fail to reject H_0 of balance.

11.3 *F*-tests for Equality of Two Variances

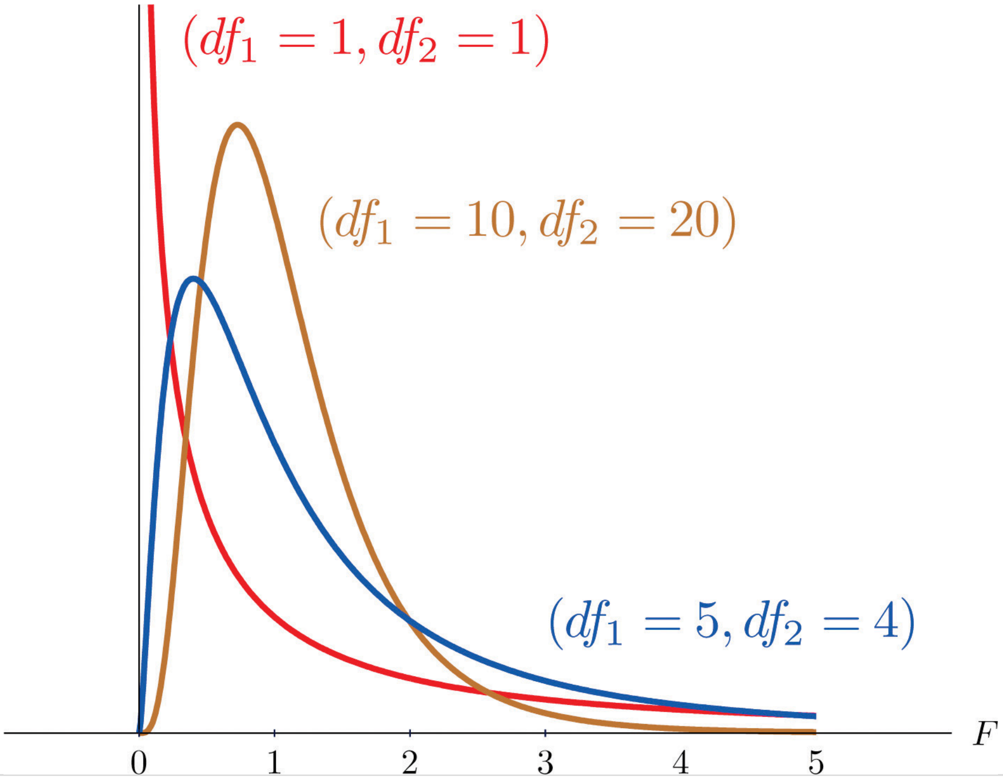
LEARNING OBJECTIVES

1. To understand what *F*-distributions are.
2. To understand how to use an *F*-test to judge whether two population variances are equal.

F-Distributions

Another important and useful family of distributions in statistics is the family of *F*-distributions. Each member of the *F*-distribution family is specified by a pair of parameters called *degrees of freedom* and denoted df_1 and df_2. Figure 11.7 "Many " shows several *F*-distributions for different pairs of degrees of freedom. An **F random variable**[7] is a random variable that assumes only positive values and follows an *F*-distribution.

Figure 11.7 Many F-Distributions



[7] A random variable following an *F*-distribution.

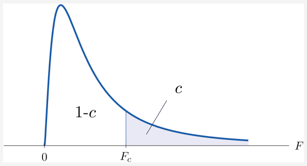
The parameter df_1 is often referred to as the *numerator* degrees of freedom and the parameter df_2 as the *denominator* degrees of freedom. It is important to keep in mind that they are not interchangeable. For example, the *F*-distribution with degrees of freedom $df_1 = 3$ and $df_2 = 8$ is a different distribution from the *F*-distribution with degrees of freedom $df_1 = 8$ and $df_2 = 3$.

Definition

The value of the F random variable F with degrees of freedom df_1 and df_2 that cuts off a right tail of area c is denoted F_c and is called a **critical value**. *See Figure 11.8.*

Figure 11.8

F_c Illustrated



Tables containing the values of F_c are given in Chapter 11 "Chi-Square Tests and ". Each of the tables is for a fixed collection of values of *c*, either 0.900, 0.950, 0.975, 0.990, and 0.995 (yielding what are called "lower" critical values), or 0.005, 0.010, 0.025, 0.050, and 0.100 (yielding what are called "upper" critical values). In each table critical values are given for various pairs (df_1, df_2) .We illustrate the use of the tables with several examples.

EXAMPLE 3

Suppose F is an F random variable with degrees of freedom $df_1 = 5$ and $df_2 = 4$. Use the tables to find

a. $F_{0.10}$
b. $F_{0.95}$

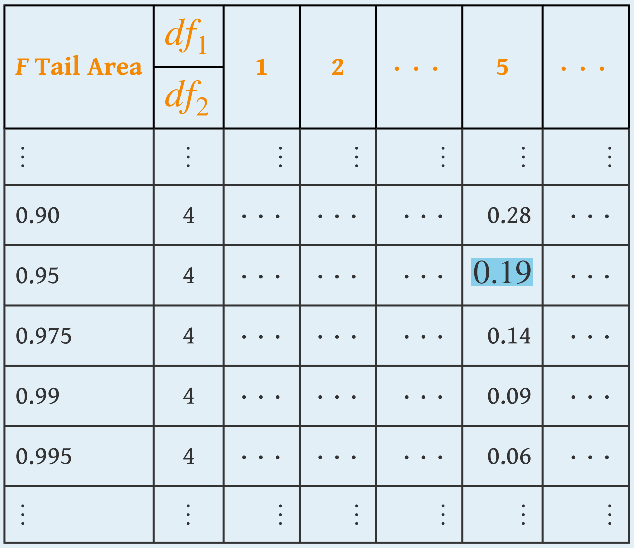
Solution:

a. The column headings of all the tables contain $df_1 = 5$. Look for the table for which 0.10 is one of the entries on the extreme left (a table of upper critical values) and that has a row heading $df_2 = 4$ in the left margin of the table. A portion of the relevant table is provided. The entry in the intersection of the column with heading $df_1 = 5$ and the row with the headings 0.10 and $df_2 = 4$, which is shaded in the table provided, is the answer, $F_{0.10} = 4.05$.

F Tail Area	df_1 df_2	1	2	. . .	5	. . .
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.005	4	. . .	. . .	. . .	22.5	. . .
0.01	4	. . .	. . .	. . .	15.5	. . .
0.025	4	. . .	. . .	. . .	9.36	. . .
0.05	4	. . .	. . .	. . .	6.26	. . .
0.10	4	. . .	. . .	. . .	4.05	. . .
⋮	⋮	⋮	⋮	⋮	⋮	⋮

b. Look for the table for which 0.95 is one of the entries on the extreme left (a table of lower critical values) and that has a row heading $df_2 = 4$ in the left margin of the table. A portion of the relevant table is provided. The entry in the intersection of the column with heading $df_1 = 5$ and the row with the headings

0.95 and $df_2 = 4$, which is shaded in the table provided, is the answer, $F_{0.95} = 0.19$.



F Tail Area	df_1 df_2	1	2	. . .	5	. . .
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.90	4	. . .	. . .	. . .	0.28	. . .
0.95	4	. . .	. . .	. . .	0.19	. . .
0.975	4	. . .	. . .	. . .	0.14	. . .
0.99	4	. . .	. . .	. . .	0.09	. . .
0.995	4	. . .	. . .	. . .	0.06	. . .
⋮	⋮	⋮	⋮	⋮	⋮	⋮

EXAMPLE 4

Suppose $\mathbf{F}$ is an *F* random variable with degrees of freedom $df_1 = 2$ and $df_2 = 20.$ Let $\alpha = 0.05.$ Use the tables to find

a. F_{α}
b. $F_{\alpha/2}$
c. $F_{1-\alpha}$
d. $F_{1-\alpha/2}$

Solution:

a. The column headings of all the tables contain $df_1 = 2.$ Look for the table for which $\alpha = 0.05$ is one of the entries on the extreme left (a table of upper critical values) and that has a row heading $df_2 = 20$ in the left margin of the table. A portion of the relevant table is provided. The shaded entry, in the intersection of the column with heading $df_1 = 2$ and the row with the headings 0.05 and $df_2 = 20$ is the answer, $F_{0.05} = 3.49.$

F Tail Area	df_1	1	2	. . .
	df_2			
⋮	⋮	⋮	⋮	⋮
0.005	20	. . .	6.99	. . .
0.01	20	. . .	5.85	. . .
0.025	20	. . .	4.46	. . .
0.05	20	. . .	3.49	. . .
0.10	20	. . .	2.59	. . .
⋮	⋮	⋮	⋮	⋮

b. Look for the table for which $\alpha / 2 = 0.025$ is one of the entries on the extreme left (a table of upper critical values) and that has a row heading $df_2 = 20$ in the left margin of the table. A portion of the relevant table is provided. The shaded entry, in the intersection of the column with heading $df_1 = 2$ and the row with the headings 0.025 and $df_2 = 20$ is the answer, $F_{0.025} = 4.46.$

F Tail Area	df_1	1	2	· · ·
	df_2			
⋮	⋮	⋮	⋮	⋮
0.005	20	· · ·	6.99	· · ·
0.01	20	· · ·	5.85	· · ·
0.025	20	· · ·	4.46	· · ·
0.05	20	· · ·	3.49	· · ·
0.10	20	· · ·	2.59	· · ·
⋮	⋮	⋮	⋮	⋮

c. Look for the table for which $1 - \alpha = 0.95$ is one of the entries on the extreme left (a table of lower critical values) and that has a row heading $df_2 = 20$ in the left margin of the table. A portion of the relevant table is provided. The shaded entry, in the intersection of the column with heading $df_1 = 2$ and the row with the headings 0.95 and $df_2 = 20$ is the answer, $F_{0.95} = 0.05.$

F Tail Area	df_1	1	2	· · ·
	df_2			
⋮	⋮	⋮	⋮	⋮
0.90	20	· · ·	0.11	· · ·
0.95	20	· · ·	0.05	· · ·
0.975	20	· · ·	0.03	· · ·
0.99	20	· · ·	0.01	· · ·
0.995	20	· · ·	0.01	· · ·
⋮	⋮	⋮	⋮	⋮

d. Look for the table for which $1 - \alpha / 2 = 0.975$ is one of the entries on the extreme left (a table of lower critical values) and that has a row heading $df_2 = 20$ in the left margin of the table. A portion of the relevant table is provided. The shaded entry, in the intersection of the column with heading $df_1 = 2$ and the row with the headings 0.975 and $df_2 = 20$ is the answer, $F_{0.975} = 0.03$.

F Tail Area	df_1	**1**	**2**	**. . .**
	df_2			
⋮	⋮	⋮	⋮	⋮
0.90	20	. . .	0.11	. . .
0.95	20	. . .	0.05	. . .
0.975	20	. . .	0.03	. . .
0.99	20	. . .	0.01	. . .
0.995	20	. . .	0.01	. . .
⋮	⋮	⋮	⋮	⋮

A fact that sometimes allows us to find a critical value from a table that we could not read otherwise is:

If F_u (r, s)denotes the value of the F-distribution with degrees of freedom $df_1 = r$and $df_2 = s$that cuts off a right tail of area u, then

$$F_c\left(k, \ell\right) = \frac{1}{F_{1-c}\left(\ell, k\right)}$$

EXAMPLE 5

Use the tables to find

a. $F_{0.01}$ for an F random variable with $df_1 = 13$ and $df_2 = 8$
b. $F_{0.975}$ for an F random variable with $df_1 = 40$ and $df_2 = 10$

Solution:

a. There is no table with $df_1 = 13$, but there is one with $df_1 = 8$. Thus we use the fact that

$$F_{0.01}\ (13,8) = \frac{1}{F_{0.99}\ (8,13)}$$

Using the relevant table we find that $F_{0.99}\ (8,13) = 0.18$, hence $F_{0.01}\ (13,8) = 0.18^{-1} = 5.556$.

b. There is no table with $df_1 = 40$, but there is one with $df_1 = 10$. Thus we use the fact that

$$F_{0.975}\ (40,10) = \frac{1}{F_{0.025}\ (10,40)}$$

Using the relevant table we find that $F_{0.025}\ (10,40) = 3.31$, hence $F_{0.975}\ (40,10) = 3.31^{-1} = 0.302$.

F-Tests for Equality of Two Variances

[8]In Chapter 9 "Two-Sample Problems" we saw how to test hypotheses about the difference between two population means μ_1 and μ_2. In some practical situations the difference between the population standard deviations σ_1 and σ_2 is also of interest. Standard deviation measures the variability of a random variable. For example, if the random variable measures the size of a machined part in a manufacturing process, the size of standard deviation is one indicator of product quality. A smaller standard deviation among items produced in the manufacturing process is desirable since it indicates consistency in product quality.

[8] A test based on an *F* statistic to check whether two population variances are equal.

For theoretical reasons it is easier to compare the squares of the population standard deviations, the population variances σ_1^2 and σ_2^2. This is not a problem, since $\sigma_1 = \sigma_2$ precisely when $\sigma_1^2 = \sigma_2^2$, $\sigma_1 < \sigma_2$ precisely when $\sigma_1^2 < \sigma_2^2$, and $\sigma_1 > \sigma_2$ precisely when $\sigma_1^2 > \sigma_2^2$.

The null hypothesis always has the form $H_0 : \sigma_1^2 = \sigma_2^2$. The three forms of the alternative hypothesis, with the terminology for each case, are:

Form of H_a	Terminology
$H_a : \sigma_1^2 > \sigma_2^2$	Right-tailed
$H_a : \sigma_1^2 < \sigma_2^2$	Left-tailed
$H_a : \sigma_1^2 \neq \sigma_2^2$	Two-tailed

Just as when we test hypotheses concerning two population means, we take a random sample from each population, of sizes n_1 and n_2, and compute the sample standard deviations s_1 and s_2. In this context the samples are always independent. The populations themselves must be normally distributed.

Test Statistic for Hypothesis Tests Concerning the Difference Between Two Population Variances

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

If the two populations are normally distributed and if $H_0 : \sigma_1^2 = \sigma_2^2$ is true then under independent sampling F approximately follows an F-distribution with degrees of freedom $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$.

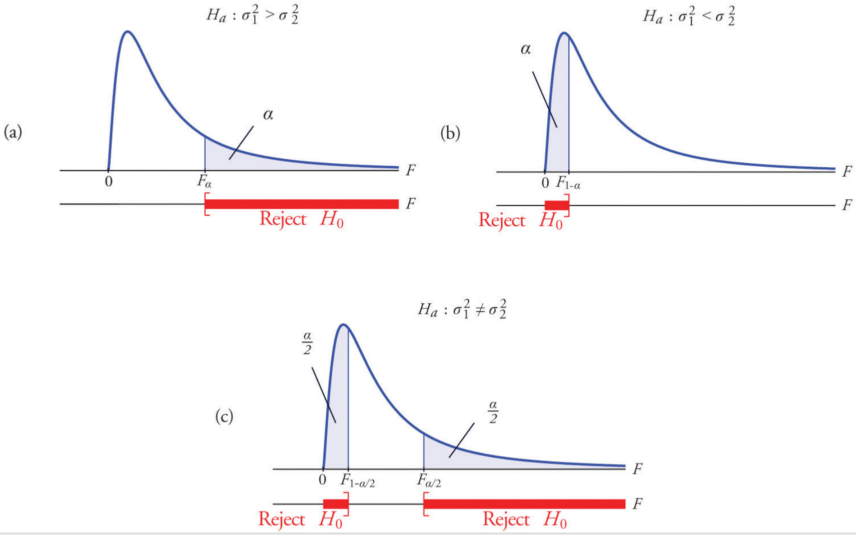
A test based on the test statistic $\boldsymbol{F}$ is called an F-test.

A most important point is that while the rejection region for a right-tailed test is exactly as in every other situation that we have encountered, because of the asymmetry in the F-distribution the critical value for a left-tailed test and the lower critical value for a two-tailed test have the special forms shown in the following table:

Terminology	Alternative Hypothesis	Rejection Region
Right-tailed	$H_a : \sigma_1^2 > \sigma_2^2$	$\boldsymbol{F} \geq F_\alpha$
Left-tailed	$H_a : \sigma_1^2 < \sigma_2^2$	$\boldsymbol{F} \leq F_{1-\alpha}$
Two-tailed	$H_a : \sigma_1^2 \neq \sigma_2^2$	$\boldsymbol{F} \leq F_{1-\alpha/2}$ or $\boldsymbol{F} \geq F_{\alpha/2}$

Figure 11.9 "Rejection Regions: (a) Right-Tailed; (b) Left-Tailed; (c) Two-Tailed" illustrates these rejection regions.

Figure 11.9 *Rejection Regions: (a) Right-Tailed; (b) Left-Tailed; (c) Two-Tailed*



The test is performed using the usual five-step procedure described at the end of Section 8.1 "The Elements of Hypothesis Testing" in Chapter 8 "Testing Hypotheses".

EXAMPLE 6

One of the quality measures of blood glucose meter strips is the consistency of the test results on the same sample of blood. The consistency is measured by the variance of the readings in repeated testing. Suppose two types of strips, *A* and *B*, are compared for their respective consistencies. We arbitrarily label the population of Type *A* strips Population 1 and the population of Type *B* strips Population 2. Suppose 15 Type *A* strips were tested with blood drops from a well-shaken vial and 20 Type *B* strips were tested with the blood from the same vial. The results are summarized in Table 11.16 "Two Types of Test Strips". Assume the glucose readings using Type *A* strips follow a normal distribution with variance σ_1^2 and those using Type *B* strips follow a normal distribution with variance with σ_2^2. Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude that the consistencies of the two types of strips are different.

TABLE 11.16 TWO TYPES OF TEST STRIPS

Strip Type	Sample Size	Sample Variance
A	$n_1 = 16$	$s_1^2 = 2.09$
B	$n_2 = 21$	$s_2^2 = 1.10$

Solution:

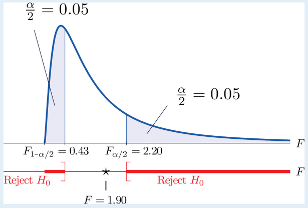
- Step 1. The test of hypotheses is

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 \\ \text{vs. } H_a &: \sigma_1^2 \neq \sigma_2^2 \quad @\ \alpha = 0.10 \end{aligned}$$

- Step 2. The distribution is the *F*-distribution with degrees of freedom $df_1 = 16 - 1 = 15$ and $df_2 = 21 - 1 = 20$.
- Step 3. The test is two-tailed. The left or lower critical value is $F_{1-\alpha/2} = F_{0.95} = 0.43$. The right or upper critical value is $F_{\alpha/2} = F_{0.05} = 2.20$. Thus the rejection region is $[0, -0.43] \cup [2.20, \infty)$, as illustrated in Figure 11.10 "Rejection Region and Test Statistic for ".

Figure 11.10

Rejection Region and Test Statistic for Note 11.27 "Example 6"



- Step 4. The value of the test statistic is

$$F = \frac{s_1^2}{s_2^2} = \frac{2.09}{1.10} = 1.90$$

- Step 5. As shown in Figure 11.10 "Rejection Region and Test Statistic for ", the test statistic 1.90 does not lie in the rejection region, so the decision is not to reject H_0. The data do not provide sufficient evidence, at the 10% level of significance, to conclude that there is a difference in the consistency, as measured by the variance, of the two types of test strips.

EXAMPLE 7

In the context of Note 11.27 "Example 6", suppose Type *A* test strips are the current market leader and Type *B* test strips are a newly improved version of Type *A*. Test, at the 10% level of significance, whether the data given in Table 11.16 "Two Types of Test Strips" provide sufficient evidence to conclude that Type *B* test strips have better consistency (lower variance) than Type *A* test strips.

Solution:

- Step 1. The test of hypotheses is now

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 \\ \text{vs. } H_a &: \sigma_1^2 > \sigma_2^2 \quad @\ \alpha = 0.10 \end{aligned}$$

- Step 2. The distribution is the *F*-distribution with degrees of freedom $df_1 = 16 - 1 = 15$ and $df_2 = 21 - 1 = 20$.

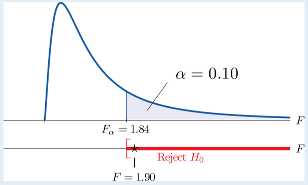
- Step 3. The value of the test statistic is

$$\boldsymbol{F} = \frac{s_1^2}{s_2^2} = \frac{2.09}{1.10} = 1.90$$

- Step 4. The test is right-tailed. The single critical value is $F_\alpha = F_{0.10} = 1.84$. Thus the rejection region is $[1.84, \infty)$, as illustrated in Figure 11.11 "Rejection Region and Test Statistic for ".

Figure 11.11

Rejection Region and Test Statistic for Note 11.28 "Example 7"



- Step 5. As shown in Figure 11.11 "Rejection Region and Test Statistic for ", the test statistic 1.90 lies in the rejection region, so the decision is to reject H_0. The data provide sufficient evidence, at the 10% level of significance, to conclude that Type *B* test strips have better consistency (lower variance) than Type *A* test strips do.

KEY TAKEAWAYS

- Critical values of an *F*-distribution with degrees of freedom df_1 and df_2 are found in tables in Chapter 12 "Appendix".
- An *F*-test can be used to evaluate the hypothesis of two identical normal population variances.

EXERCISES

BASIC

1. Find $F_{0.01}$ for each of the following degrees of freedom.

   a. $df_1 = 5$ and $df_2 = 5$
   b. $df_1 = 5$ and $df_2 = 12$
   c. $df_1 = 12$ and $df_2 = 20$

2. Find $F_{0.05}$ for each of the following degrees of freedom.

   a. $df_1 = 6$ and $df_2 = 6$
   b. $df_1 = 6$ and $df_2 = 12$
   c. $df_1 = 12$ and $df_2 = 30$

3. Find $F_{0.95}$ for each of the following degrees of freedom.

   a. $df_1 = 6$ and $df_2 = 6$
   b. $df_1 = 6$ and $df_2 = 12$
   c. $df_1 = 12$ and $df_2 = 30$

4. Find $F_{0.90}$ for each of the following degrees of freedom.

   a. $df_1 = 5$ and $df_2 = 5$
   b. $df_1 = 5$ and $df_2 = 12$
   c. $df_1 = 12$ and $df_2 = 20$

5. For $df_1 = 7, df_2 = 10$ and $\alpha = 0.05$, find

   a. F_α
   b. $F_{1-\alpha}$
   c. $F_{\alpha/2}$
   d. $F_{1-\alpha/2}$

6. For $df_1 = 15, df_2 = 8,$ and $\alpha = 0.01$, find

   a. F_α
   b. $F_{1-\alpha}$
   c. $F_{\alpha/2}$
   d. $F_{1-\alpha/2}$

7. For each of the two samples

$$\text{Sample 1 : } \{8,2,11,0,-2,\}$$
$$\text{Sample 2 : } \{-2,0,0,0,2,4,-1\}$$

find

a. the sample size,
b. the sample mean,
c. the sample variance.

8. For each of the two samples

$$\text{Sample 1 : } \{0.\ 8,1.\ 2,1.\ 1,0.\ 8,-2.0\}$$
$$\text{Sample 2 : } \{-2.\ 0,0.\ 0,0.\ 7,0.\ 8,2.\ 2,4.\ 1,-1.9\}$$

find

a. the sample size,
b. the sample mean,
c. the sample variance.

9. Two random samples taken from two normal populations yielded the following information:

Sample	Sample Size	Sample Variance
1	$n_1 = 16$	$s_1^2 = 53$
2	$n_2 = 21$	$s_2^2 = 32$

a. Find the statistic $F = s_1^2 \,/\, s_2^2$.
b. Find the degrees of freedom df_1 and df_2.
c. Find $F_{0.05}$ using df_1 and df_2 computed above.
d. Perform the test the hypotheses $H_0 : \sigma_1^2 = \sigma_2^2$vs. $H_a : \sigma_1^2 > \sigma_2^2$at the 5% level of significance.

10. Two random samples taken from two normal populations yielded the following information:

Sample	Sample Size	Sample Variance
1	$n_1 = 11$	$s_1^2 = 61$
2	$n_2 = 8$	$s_2^2 = 44$

a. Find the statistic $F = s_1^2 \,/\, s_2^2$.
b. Find the degrees of freedom df_1 and df_2.
c. Find $F_{0.05}$ using df_1 and df_2 computed above.

d. Perform the test the hypotheses $H_0 : \sigma_1^2 = \sigma_2^2$vs. $H_a : \sigma_1^2 > \sigma_2^2$at the 5% level of significance.

11. Two random samples taken from two normal populations yielded the following information:

Sample	Sample Size	Sample Variance
1	$n_1 = 10$	$s_1^2 = 12$
2	$n_2 = 13$	$s_2^2 = 23$

a. Find the statistic $F = s_1^2 / s_2^2$.
b. Find the degrees of freedom df_1 and df_2.
c. For $\alpha = 0.05$ find $F_{1-\alpha}$ using df_1 and df_2 computed above.
d. Perform the test the hypotheses $H_0 : \sigma_1^2 = \sigma_2^2$vs. $H_a : \sigma_1^2 < \sigma_2^2$at the 5% level of significance.

12. Two random samples taken from two normal populations yielded the following information:

Sample	Sample Size	Sample Variance
1	$n_1 = 8$	$s_1^2 = 102$
2	$n_2 = 8$	$s_2^2 = 603$

a. Find the statistic $F = s_1^2 / s_2^2$.
b. Find the degrees of freedom df_1 and df_2.
c. For $\alpha = 0.05$ find $F_{1-\alpha}$ using df_1 and df_2 computed above.
d. Perform the test the hypotheses $H_0 : \sigma_1^2 = \sigma_2^2$vs. $H_a : \sigma_1^2 < \sigma_2^2$at the 5% level of significance.

13. Two random samples taken from two normal populations yielded the following information:

Sample	Sample Size	Sample Variance
1	$n_1 = 9$	$s_1^2 = 123$
2	$n_2 = 31$	$s_2^2 = 543$

a. Find the statistic $F = s_1^2 / s_2^2$.
b. Find the degrees of freedom df_1 and df_2.
c. For $\alpha = 0.05$ find $F_{1-\alpha/2}$ and $F_{\alpha/2}$ using df_1 and df_2 computed above.

d. Perform the test the hypotheses $H_0 : \sigma_1^2 = \sigma_2^2$vs. $H_a : \sigma_1^2 \neq \sigma_2^2$at the 5% level of significance.

14. Two random samples taken from two normal populations yielded the following information:

Sample	Sample Size	Sample Variance
1	$n_1 = 21$	$s_1^2 = 199$
2	$n_2 = 21$	$s_2^2 = 66$

a. Find the statistic $F = s_1^2 / s_2^2$.
b. Find the degrees of freedom df_1 and df_2.
c. For $\alpha = 0.05$ find $F_{1-\alpha/2}$ and $F_{\alpha/2}$ using df_1 and df_2 computed above.
d. Perform the test the hypotheses $H_0 : \sigma_1^2 = \sigma_2^2$vs. $H_a : \sigma_1^2 \neq \sigma_2^2$at the 5% level of significance.

APPLICATIONS

15. Japanese sturgeon is a subspecies of the sturgeon family indigenous to Japan and the Northwest Pacific. In a particular fish hatchery newly hatched baby Japanese sturgeon are kept in tanks for several weeks before being transferred to larger ponds. Dissolved oxygen in tank water is very tightly monitored by an electronic system and rigorously maintained at a target level of 6.5 milligrams per liter (mg/l). The fish hatchery looks to upgrade their water monitoring systems for tighter control of dissolved oxygen. A new system is evaluated against the old one currently being used in terms of the variance in measured dissolved oxygen. Thirty-one water samples from a tank operated with the new system were collected and 16 water samples from a tank operated with the old system were collected, all during the course of a day. The samples yield the following information:

$$\begin{array}{llll} \text{New} & \text{Sample 1:} & n_1 = 31 & s_1^2 = 0.0121 \\ \text{Old} & \text{Sample 2:} & n_2 = 16 & s_2^2 = 0.0319 \end{array}$$

Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude that the new system will provide a tighter control of dissolved oxygen in the tanks.

16. The risk of investing in a stock is measured by the volatility, or the variance, in changes in the price of that stock. Mutual funds are baskets of stocks and offer generally lower risk to investors. Different mutual funds have different focuses and offer different levels of risk. Hippolyta is deciding between two mutual funds, *A* and *B*, with similar expected returns. To make a final decision, she examined the annual returns of the two funds during the last ten years and obtained the following information:

$$\begin{array}{lll} \text{Mutual Fund } A & & \\ \text{Sample 1 :} & n_1 = 10 & s_1^2 = 0.012 \\ \text{Mutual Fund } B & & \\ \text{Sample 2 :} & n_2 = 10 & s_2^2 = 0.005 \end{array}$$

Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that the two mutual funds offer different levels of risk.

17. It is commonly acknowledged that grading of the writing part of a college entrance examination is subject to inconsistency. Every year a large number of potential graders are put through a rigorous training program before being given grading assignments. In order to gauge whether such a training program really enhances consistency in grading, a statistician conducted an experiment in which a reference essay was given to 61 trained graders and 31 untrained graders. Information on the scores given by these graders is summarized below:

$$\begin{array}{llll} \text{Trained} & \text{Sample 1:} & n_1 = 61 & s_1^2 = 2.15 \\ \text{Untrained} & \text{Sample 2:} & n_2 = 31 & s_2^2 = 3.91 \end{array}$$

Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that the training program enhances the consistency in essay grading.

18. A common problem encountered by many classical music radio stations is that their listeners belong to an increasingly narrow band of ages in the population. The new general manager of a classical music radio station believed that a new playlist offered by a professional programming agency would attract listeners from a wider range of ages. The new list was used for a year. Two random samples were taken before and after the new playlist was adopted. Information on the ages of the listeners in the sample are summarized below:

$$\begin{array}{lll} \text{Before} & \text{Sample 1:} & n_1 = 21 \quad s_1^2 = 56.25 \\ \text{After} & \text{Sample 2:} & n_2 = 16 \quad s_2^2 = 76.56 \end{array}$$

Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude that the new playlist has expanded the range of listener ages.

19. A laptop computer maker uses battery packs supplied by two companies, *A* and *B*. While both brands have the same average battery life between charges (LBC), the computer maker seems to receive more complaints about shorter LBC than expected for battery packs supplied by company *B*. The computer maker suspects that this could be caused by higher variance in LBC for Brand *B*. To check that, ten new battery packs from each brand are selected, installed on the same models of laptops, and the laptops are allowed to run until the battery packs are completely discharged. The following are the observed LBCs in hours.

Brand A	Brand B
3.2	3.0
3.4	3.5
2.8	2.9
3.0	3.1
3.0	2.3
3.0	2.0
2.8	3.0
2.9	2.9
3.0	3.0
3.0	4.1

Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that the LBCs of Brand *B* have a larger variance that those of Brand *A*.

20. A manufacturer of a blood-pressure measuring device for home use claims that its device is more consistent than that produced by a leading competitor. During a visit to a medical store a potential buyer tried both devices on himself repeatedly during a short period of time. The following are readings of systolic pressure.

Manufacturer	Competitor
132	129
134	132
129	129
129	138
130	
132	

a. Test, at the 5% level of significance, whether the data provide sufficient evidence to conclude that the manufacturer's claim is true.
b. Repeat the test at the 10% level of significance. Quote as many computations from part (a) as possible.

LARGE DATA SET EXERCISES

21. Large Data Sets 1A and 1B record SAT scores for 419 male and 581 female students. Test, at the 1% level of significance, whether the data provide sufficient evidence to conclude that the variances of scores of male and female students differ.

    http://www.gone.2012books.lardbucket.org/sites/all/files/data1A.xls

    http://www.gone.2012books.lardbucket.org/sites/all/files/data1B.xls

22. Large Data Sets 7, 7A, and 7B record the survival times of 140 laboratory mice with thymic leukemia. Test, at the 10% level of significance, whether the data provide sufficient evidence to conclude that the variances of survival times of male mice and female mice differ.

    http://www.gone.2012books.lardbucket.org/sites/all/files/data7.xls

    http://www.gone.2012books.lardbucket.org/sites/all/files/data7A.xls

    http://www.gone.2012books.lardbucket.org/sites/all/files/data7B.xls

ANSWERS

1.
   a. 11.0,
   b. 5.06,
   c. 3.23

3.
   a. 0.23,
   b. 0.25,
   c. 0.40

5.
   a. 3.14,
   b. 0.27,
   c. 3.95,
   d. 0.21

7. Sample 1:

   a. $n_1 = 5,$
   b. $\overline{x}_1 = 3.8,$
   c. $s_1^2 = 30.2.$

   Sample 2:

   a. $n_2 = 7,$
   b. $\overline{x}_2 = 0.4286,$
   c. $s_2^2 = 3.95$

9.
   a. 1.6563,
   b. $df_1 = 15, df_2 = 20,$
   c. $F_{0.05} = 2.2$
   d. do not reject H_0

11.
   a. 0.5217
   b. $df_1 = 9, df_2 = 12,$
   c. $F_{0.95} = 0.3254,$
   d. do not reject H_0

13.
   a. 0.1692
   b. $df_1 = 8, df_2 = 30$
   c. $F_{0.975} = 0.26, F_{0.025} = 2.65,$
   d. reject H_0

15. $F = 0.3793, F_{0.90} = 0.58,$ reject H_0

17. $F = 0.5499, F_{0.95} = 0.61,$ reject H_0

19. $F = 0.0971$, $F_{0.95} = 0.31$, reject H_0

21. $F = 0.893131$. $df_1 = 418$ and $df_2 = 580$. Rejection Region: $(0,0.7897] \cup [1.2614,\infty)$. Decision: Fail to reject H_0 of equal variances.

11.4 *F*-Tests in One-Way ANOVA

LEARNING OBJECTIVE

1. To understand how to use an *F*-test to judge whether several population means are all equal.

In Chapter 9 "Two-Sample Problems" we saw how to compare two population means μ_1 and μ_2. In this section we will learn to compare three or more population means at the same time, which is often of interest in practical applications. For example, an administrator at a university may be interested in knowing whether student grade point averages are the same for different majors. In another example, an oncologist may be interested in knowing whether patients with the same type of cancer have the same average survival times under several different competing cancer treatments.

In general, suppose there are *K* normal populations with possibly different means, $\mu_1, \mu_2, \ldots, \mu_K$, but all with the same variance σ^2. The study question is whether all the *K* population means are the same. We formulate this question as the test of hypotheses

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \cdots = \mu_K \\ \text{vs. } H_a &: \text{ not all } K \text{ population means are equal} \end{aligned}$$

To perform the test *K* independent random samples are taken from the *K* normal populations. The *K* sample means, the *K* sample variances, and the *K* sample sizes are summarized in the table:

Population	Sample Size	Sample Mean	Sample Variance
1	n_1	$\overline{x}_1$	s_1^2
2	n_2	$\overline{x}_2$	s_2^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
K	n_K	$\overline{x}_K$	s_K^2

Define the following quantities:

The **combined sample size**:

$$n = n_1 + n_2 + \cdots + n_K$$

The **mean of the combined sample** of all *n* observations:

$$\overline{x} = \frac{\Sigma x}{n} = \frac{n_1 \overline{x}_1 + n_2 \overline{x}_2 + \cdots + n_K \overline{x}_K}{n}$$

The **mean square for treatment**:

$$MST = \frac{n_1 (\overline{x}_1 - \overline{x})^2 + n_2 (\overline{x}_2 - \overline{x})^2 + \cdots + n_K (\overline{x}_K - \overline{x})^2}{K-1}$$

The **mean square for error**:

$$MSE = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2 + \cdots + (n_K - 1)\, s_K^2}{n - K}$$

MST[9] can be thought of as the variance between the *K* individual independent random samples and **MSE**[10] as the variance within the samples. This is the reason for the name "analysis of variance," universally abbreviated **ANOVA**[11]. The adjective "one-way" has to do with the fact that the sampling scheme is the simplest possible, that of taking one random sample from each population under consideration. If the means of the *K* populations are all the same then the two quantities MST and MSE should be close to the same, so the null hypothesis will be rejected if the ratio of these two quantities is significantly greater than 1. This yields the following test statistic and methods and conditions for its use.

[9] Mean square for treatment.

[10] Mean square for error.

[11] Analysis of variance.

Test Statistic for Testing the Null Hypothesis that K Population Means Are Equal

$$F = \frac{MST}{MSE}$$

If the K populations are normally distributed with a common variance and if $H_0 : \mu_1 = \cdots = \mu_K$ is true then under independent random sampling F approximately follows an F-distribution with degrees of freedom $df_1 = K-1$ and $df_2 = n - K$.

The test is right-tailed: H_0 is rejected at level of significance α if $\boldsymbol{F} \geq F_\alpha$.

As always the test is performed using the usual five-step procedure.

EXAMPLE 8

The average of grade point averages (GPAs) of college courses in a specific major is a measure of difficulty of the major. An educator wishes to conduct a study to find out whether the difficulty levels of different majors are the same. For such a study, a random sample of major grade point averages (GPA) of 11 graduating seniors at a large university is selected for each of the four majors mathematics, English, education, and biology. The data are given in Table 11.17 "Difficulty Levels of College Majors". Test, at the 5% level of significance, whether the data contain sufficient evidence to conclude that there are differences among the average major GPAs of these four majors.

TABLE 11.17 DIFFICULTY LEVELS OF COLLEGE MAJORS

Mathematics	English	Education	Biology
2.59	3.64	4.00	2.78
3.13	3.19	3.59	3.51
2.97	3.15	2.80	2.65
2.50	3.78	2.39	3.16
2.53	3.03	3.47	2.94
3.29	2.61	3.59	2.32
2.53	3.20	3.74	2.58
3.17	3.30	3.77	3.21
2.70	3.54	3.13	3.23
3.88	3.25	3.00	3.57
2.64	4.00	3.47	3.22

Solution:

- Step 1. The test of hypotheses is

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ \text{vs. } H_a &: \text{not all four population means are equal} \quad @\ \alpha = 0.05 \end{aligned}$$

- Step 2. The test statistic is $\boldsymbol{F} = MST \mathbin{/} MSE$ with (since n = 44 and K = 4) degrees of freedom $df_1 = K{-}1 = 4 - 1 = 3$ and $df_2 = n - K = 44 - 4 = 40.$

- Step 3. If we index the population of mathematics majors by 1, English majors by 2, education majors by 3, and biology majors by 4, then the sample sizes, sample means, and sample variances of the four samples in Table 11.17 "Difficulty Levels of College Majors" are summarized (after rounding for simplicity) by:

Major	Sample Size	Sample Mean	Sample Variance
Mathematics	$n_1 = 11$	$\overline{x}_1 = 2.90$	$s_1^2 = 0.188$
English	$n_2 = 11$	$\overline{x}_2 = 3.34$	$s_2^2 = 0.148$
Education	$n_3 = 11$	$\overline{x}_3 = 3.36$	$s_3^2 = 0.229$
Biology	$n_4 = 11$	$\overline{x}_4 = 3.02$	$s_4^2 = 0.157$

The average of all 44 observations is (after rounding for simplicity) $\overline{x} = 3.15.$ We compute (rounding for simplicity)

$$\begin{aligned} MST &= \frac{11\left(2.90 - 3.15\right)^2 + 11\left(3.34 - 3.15\right)^2 + 11\left(3.36 - 3.15\right.}{4 - 1} \\ &= \frac{1.7556}{3} \\ &= 0.585 \end{aligned}$$

and

$$\begin{aligned} MSE &= \frac{(11-1)\ (0.188) + \ (11-1)\ (0.148) + \ (11-1)\ (0.229)}{44-4} \\ &= \frac{7.22}{40} \\ &= 0.181 \end{aligned}$$

so that

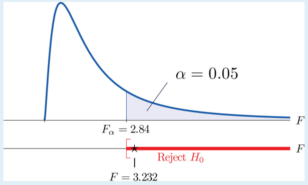
$$\boldsymbol{F} = \frac{MST}{MSE} = \frac{0.585}{0.181} = 3.232$$

- Step 4. The test is right-tailed. The single critical value is (since $df_1 = 3$ and $df_2 = 40$) $F_\alpha = F_{0.05} = 2.84$. Thus the rejection region is $[2.84, \infty)$, as illustrated in Figure 11.12.

Figure 11.12

Note 11.36 "Example 8"

Rejection Region



- Step 5. Since $F = 3.232 > 2.84$, we reject H_0. The data provide sufficient evidence, at the 5% level of significance, to conclude that the averages of major GPAs for the four majors considered are not all equal.

EXAMPLE 9

A research laboratory developed two treatments which are believed to have the potential of prolonging the survival times of patients with an acute form of thymic leukemia. To evaluate the potential treatment effects 33 laboratory mice with thymic leukemia were randomly divided into three groups. One group received Treatment 1, one received Treatment 2, and the third was observed as a control group. The survival times of these mice are given in Table 11.18 "Mice Survival Times in Days". Test, at the 1% level of significance, whether these data provide sufficient evidence to confirm the belief that at least one of the two treatments affects the average survival time of mice with thymic leukemia.

TABLE 11.18 MICE SURVIVAL TIMES IN DAYS

Treatment 1		Treatment 2	Control
71	75	77	81
72	73	67	79
75	72	79	73
80	65	78	71
60	63	81	75
65	69	72	84
63	64	71	77
78	71	84	67
		91	

Solution:

- Step 1. The test of hypotheses is

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \mu_3 \\ \text{vs. } H_a &: \text{not all three population means are equal} \quad @ \alpha = 0.01 \end{aligned}$$

- Step 2. The test statistic is $\boldsymbol{F} = MST \ / \ MSE$ with (since n = 33 and K = 3) degrees of freedom $df_1 \ = K{-}1 = 3 - 1 = 2$ and $df_2 = n - K = 33 - 3 = 30.$

  - Step 3. If we index the population of mice receiving Treatment 1 by 1, Treatment 2 by 2, and no treatment by 3, then the sample sizes, sample means, and sample variances of the three samples in Table 11.18 "Mice Survival Times in Days" are summarized (after rounding for simplicity) by:

Group	Sample Size	Sample Mean	Sample Variance
Treatment 1	$n_1 \ = 16$	$\overline{x}_1 \ = 69.75$	$s_1^2 \ = 34.47$
Treatment 2	$n_2 \ = 9$	$\overline{x}_2 \ = 77.78$	$s_2^2 \ = 52.69$
Control	$n_3 \ = 8$	$\overline{x}_3 \ = 75.88$	$s_3^2 \ = 30.69$

The average of all 33 observations is (after rounding for simplicity) $\overline{x} \ = 73.42.$ We compute (rounding for simplicity)

$$MST = \frac{16\left(69.75 - 73.42\right)^{\ 2} + 9(77.78 - 73.42)^{\ 2} + 8\left(75.88 - 73\right.}{3 - 1}$$

and

$$MSE = \frac{\left(16 - 1\right) \ (34.47) + \ (9 - 1) \ \left(52.69\right) + \ (8 - 1) \ \left(30.69\right)}{33 - 3} =$$

so that

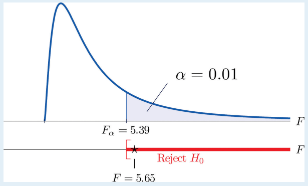
$$\boldsymbol{F} = \frac{MST}{MSE} = \frac{217.50}{38.45} = 5.65$$

- Step 4. The test is right-tailed. The single critical value is $F_\alpha \ = F_{0.01} \ \ = 5.39.$ Thus the rejection region is $\left[5.39, \infty\right)$, as illustrated in Figure 11.13.

Figure 11.13

Note 11.37 "Example 9"

Rejection Region



- Step 5. Since $F = 5.65 > 5.39$, we reject H_0. The data provide sufficient evidence, at the 1% level of significance, to conclude that a treatment effect exists at least for one of the two treatments in increasing the mean survival time of mice with thymic leukemia.

It is important to to note that, if the null hypothesis of equal population means is rejected, the statistical implication is that not all population means are equal. It does not however tell which population mean is different from which. The inference about where the suggested difference lies is most frequently made by a follow-up study.

KEY TAKEAWAY

- An *F*-test can be used to evaluate the hypothesis that the means of several normal populations, all with the same standard deviation, are identical.

EXERCISES

BASIC

1. The following three random samples are taken from three normal populations with respective means μ_1, μ_2, and μ_3, and the same variance σ^2.

Sample 1	Sample 2	Sample 3
2	3	0
2	5	1
3	7	2
5		1
3		

   a. Find the combined sample size *n*.
   b. Find the combined sample mean $\overline{x}$.
   c. Find the sample mean for each of the three samples.
   d. Find the sample variance for each of the three samples.
   e. Find MST.
   f. Find MSE.
   g. Find $F = MST\ /\ MSE$.

2. The following three random samples are taken from three normal populations with respective means μ_1, μ_2, and μ_3, and a same variance σ^2.

Sample 1	Sample 2	Sample 3
0.0	1.3	0.2
0.1	1.5	0.2
0.2	1.7	0.3
0.1		0.5
		0.0

   a. Find the combined sample size *n*.
   b. Find the combined sample mean $\overline{x}$.
   c. Find the sample mean for each of the three samples.
   d. Find the sample variance for each of the three samples.
   e. Find MST.

f. Find MSE.
g. Find $F = MST \,/\, MSE$.

3. Refer to Exercise 1.

a. Find the number of populations under consideration K.
b. Find the degrees of freedom $df_1 = K-1$ and $df_2 = n - K$.
c. For $\alpha = 0.05$, find F_α with the degrees of freedom computed above.
d. At $\alpha = 0.05$, test hypotheses

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \mu_3 \\ \text{vs. } H_a &: \text{at least one pair of the} \\ &\quad \text{population means are not equal} \end{aligned}$$

4. Refer to Exercise 2.

a. Find the number of populations under consideration K.
b. Find the degrees of freedoms $df_1 = K-1$ and $df_2 = n - K$.
c. For $\alpha = 0.01$, find F_α with the degrees of freedom computed above.
d. At $\alpha = 0.01$, test hypotheses

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \mu_3 \\ \text{vs. } H_a &: \text{at least one pair of the} \\ &\quad \text{population means are not equal} \end{aligned}$$

APPLICATIONS

5. The Mozart effect refers to a boost of average performance on tests for elementary school students if the students listen to Mozart's chamber music for a period of time immediately before the test. In order to attempt to test whether the Mozart effect actually exists, an elementary school teacher conducted an experiment by dividing her third-grade class of 15 students into three groups of 5. The first group was given an end-of-grade test without music; the second group listened to Mozart's chamber music for 10 minutes; and the third groups listened to Mozart's chamber music for 20 minutes before the test. The scores of the 15 students are given below:

Group 1	Group 2	Group 3
80	79	73
63	73	82

Group 1	Group 2	Group 3
74	74	79
71	77	82
70	81	84

Using the **ANOVA *F*-test**[12] at $\alpha = 0.10$, is there sufficient evidence in the data to suggest that the Mozart effect exists?

6. The Mozart effect refers to a boost of average performance on tests for elementary school students if the students listen to Mozart's chamber music for a period of time immediately before the test. Many educators believe that such an effect is not necessarily due to Mozart's music per se but rather a relaxation period before the test. To support this belief, an elementary school teacher conducted an experiment by dividing her third-grade class of 15 students into three groups of 5. Students in the first group were asked to give themselves a self-administered facial massage; students in the second group listened to Mozart's chamber music for 15 minutes; students in the third group listened to Schubert's chamber music for 15 minutes before the test. The scores of the 15 students are given below:

Group 1	Group 2	Group 3
79	82	80
81	84	81
80	86	71
89	91	90
86	82	86

Test, using the ANOVA *F*-test at the 10% level of significance, whether the data provide sufficient evidence to conclude that any of the three relaxation method does better than the others.

7. Precision weighing devices are sensitive to environmental conditions. Temperature and humidity in a laboratory room where such a device is installed are tightly controlled to ensure high precision in weighing. A newly designed weighing device is claimed to be more robust against small variations of temperature and humidity. To verify such a claim, a laboratory tests the new device under four settings of temperature-humidity conditions. First, two levels of *high* and *low* temperature and two levels of *high* and *low* humidity are identified. Let *T* stand for temperature and *H* for humidity. The four experimental settings are defined and noted as (*T*, *H*): (high, high), (high, low),

[12] a test based on an *F* statistic to check whether several population means are equal.

(low, high), and (low, low). A pre-calibrated standard weight of 1 kg was weighed by the new device four times in each setting. The results in terms of error (in micrograms mcg) are given below:

(high, high)	(high, low)	(low, high)	(low, low)
−1.50	11.47	−14.29	5.54
−6.73	9.28	−18.11	10.34
11.69	5.58	−11.16	15.23
−5.72	10.80	−10.41	−5.69

Test, using the ANOVA *F*-test at the 1% level of significance, whether the data provide sufficient evidence to conclude that the mean weight readings by the newly designed device vary among the four settings.

8. To investigate the real cost of owning different makes and models of new automobiles, a consumer protection agency followed 16 owners of new vehicles of four popular makes and models, call them TC, HA, NA, and FT, and kept a record of each of the owner's real cost in dollars for the first five years. The five-year costs of the 16 car owners are given below:

TC	HA	NA	FT
8423	7776	8907	10333
7889	7211	9077	9217
8665	6870	8732	10540
	7129	9747	
	7359	8677	

Test, using the ANOVA *F*-test at the 5% level of significance, whether the data provide sufficient evidence to conclude that there are differences among the mean real costs of ownership for these four models.

9. Helping people to lose weight has become a huge industry in the United States, with annual revenue in the hundreds of billion dollars. Recently each of the three market-leading weight reducing programs claimed to be the most effective. A consumer research company recruited 33 people who wished to lose weight and sent them to the three leading programs. After six months their weight losses were recorded. The results are summarized below:

Statistic	Prog. 1	Prog. 2	Prog. 3
Sample Mean	$\overline{x}_1 = 10.65$	$\overline{x}_2 = 8.90$	$\overline{x}_3 = 9.33$

Statistic	Prog. 1	Prog. 2	Prog. 3
Sample Variance	$s_1^2 = 27.20$	$s_2^2 = 16.86$	$s_3^2 = 32.40$
Sample Size	$n_1 = 11$	$n_2 = 11$	$n_3 = 11$

The mean weight loss of the combined sample of all 33 people was $\overline{x} = 9.63$. Test, using the ANOVA *F*-test at the 5% level of significance, whether the data provide sufficient evidence to conclude that some program is more effective than the others.

10. A leading pharmaceutical company in the disposable contact lenses market has always taken for granted that the sales of certain peripheral products such as contact lens solutions would automatically go with the established brands. The long-standing culture in the company has been that lens solutions would not make a significant difference in user experience. Recent market research surveys, however, suggest otherwise. To gain a better understanding of the effects of contact lens solutions on user experience, the company conducted a comparative study in which 63 contact lens users were randomly divided into three groups, each of which received one of three top selling lens solutions on the market, including one of the company's own. After using the assigned solution for two weeks, each participant was asked to rate the solution on the scale of 1 to 5 for satisfaction, with 5 being the highest level of satisfaction. The results of the study are summarized below:

Statistics	Sol. 1	Sol. 2	Sol. 3
Sample Mean	$\overline{x}_1 = 3.28$	$\overline{x}_2 = 3.96$	$\overline{x}_3 = 4.10$
Sample Variance	$s_1^2 = 0.15$	$s_2^2 = 0.32$	$s_3^2 = 0.36$
Sample Size	$n_1 = 18$	$n_2 = 23$	$n_3 = 22$

The mean satisfaction level of the combined sample of all 63 participants was $\overline{x} = 3.81$. Test, using the ANOVA *F*-test at the 5% level of significance, whether the data provide sufficient evidence to conclude that not all three average satisfaction levels are the same.

LARGE DATA SET EXERCISE

11. Large Data Set 9 records the costs of materials (textbook, solution manual, laboratory fees, and so on) in each of ten different courses in each of three different subjects, chemistry, computer science, and mathematics. Test, at the

1% level of significance, whether the data provide sufficient evidence to conclude that the mean costs in the three disciplines are not all the same.

http://www.gone.2012books.lardbucket.org/sites/all/files/data9.xls

ANSWERS

1.
   a. $n = 12$,
   b. $\overline{x} = 2.8333$,
   c. $\overline{x}_1 = 3, \overline{x}_2 = 5, \overline{x}_3 = 1$,
   d. $s_1^2 = 1.5, s_2^2 = 4, s_3^2 = 0.6667$,
   e. $MST = 13.83$,
   f. $MSE = 1.78$,
   g. $F = 7.7812$

3.
   a. $K = 3$;
   b. $df_1 = 2, df_2 = 9$;
   c. $F_{0.05} = 4.26$;
   d. $F = 5.53$, reject H_0

5. $F = 3.9647$, $F_{0.10} = 2.81$, reject H_0

7. $F = 9.6018$, $F_{0.01} = 5.95$, reject H_0

9. $F = 0.3589$, $F_{0.05} = 3.32$, do not reject H_0

11. $F = 1.418$. $df_1 = 2$ and $df_2 = 27$. Rejection Region: $\left[5.4881, \infty\right)$. Decision: Fail to reject H_0 of equal means.