Speaking of Statistics

Here is a survey about college students' credit card usage. Suggest several ways that the study could have been more meaningful if confidence intervals had been used.

OTHER PEOPLE'S MONEY

Undergrads love their plastic. That means—you guessed it—students are learning to become debtors. According to the Public Interest Research Groups, only half of all students pay off card balances in full each month, 36% sometimes do and 14% never do. Meanwhile, 48% have paid a late fee. Here's how undergrads stack up, according to Nellie Mae, a provider of college loans:

Undergrads with a credit card **78%**

Average number of cards owned . . **3**

Average student card debt **$1236**

Students with 4 or more cards **32%**

Balances of $3000 to $7000 . . . . . **13%**

Balances over $7000 . . . . **9%**

4. Enter **500** in the box labeled *n*.

5. Either type in or scroll to 90% for the Confidence Level, then click [OK].

The result of the procedure is shown next.

Confidence interval—proportion

```
  90% Confidence level
 0.12 Proportion
  500 n
1.645 z
0.024 Half-width
0.144 Upper confidence limit
0.096 Lower confidence limit
```

7–4 Confidence Intervals for Variances and Standard Deviations

Objective 6

Find a confidence interval for a variance and a standard deviation.

In Sections 7–1 through 7–3 confidence intervals were calculated for means and proportions. This section will explain how to find confidence intervals for variances and standard deviations. In statistics, the variance and standard deviation of a variable are as important as the mean. For example, when products that fit together (such as pipes) are manufactured, it is important to keep the variations of the diameters of the products as small as possible; otherwise, they will not fit together properly and will have to be scrapped. In the manufacture of medicines, the variance and standard deviation of the medication in the pills play an important role in making sure patients receive the proper dosage. For these reasons, confidence intervals for variances and standard deviations are necessary.

Historical Note

The χ^2 distribution with 2 degrees of freedom was formulated by a mathematician named Hershel in 1869 while he was studying the accuracy of shooting arrows at a target. Many other mathematicians have since contributed to its development.

To calculate these confidence intervals, a new statistical distribution is needed. It is called the **chi-square distribution.**

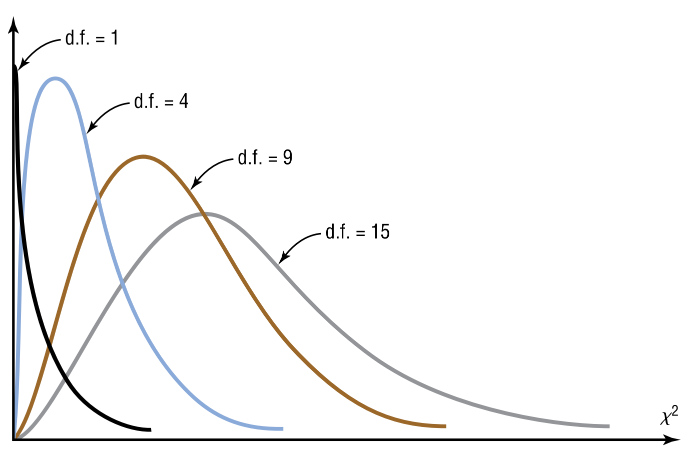
The chi-square variable is similar to the *t* variable in that its distribution is a family of curves based on the number of degrees of freedom. The symbol for chi-square is χ^2 (Greek letter chi, pronounced "ki"). Several of the distributions are shown in Figure 7–9, along with the corresponding degrees of freedom. The chi-square distribution is obtained from the values of $(n - 1)s^2/\sigma^2$ when random samples are selected from a normally distributed population whose variance is σ^2.

A chi-square variable cannot be negative, and the distributions are skewed to the right. At about 100 degrees of freedom, the chi-square distribution becomes somewhat symmetric. The area under each chi-square distribution is equal to 1.00, or 100%.

Table G in Appendix C gives the values for the chi-square distribution. These values are used in the denominators of the formulas for confidence intervals. Two different values

Figure 7–9

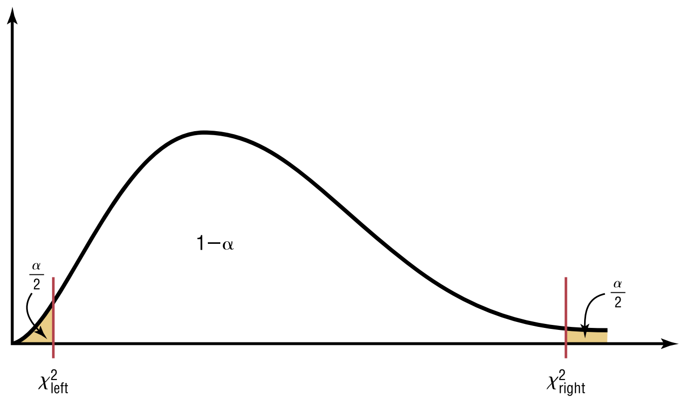
The Chi-Square Family of Curves



are used in the formula because the distribution is not symmetric. One value is found on the left side of the table, and the other is on the right. See Figure 7–10. For example, to find the table values corresponding to the 95% confidence interval, you must first change 95% to a decimal and subtract it from 1 ($1 - 0.95 = 0.05$). Then divide the answer by 2 ($\alpha/2 = 0.05/2 = 0.025$). This is the column on the right side of the table, used to get the values for χ^2_{right}. To get the value for χ^2_{left}, subtract the value of $\alpha/2$ from 1 ($1 - 0.05/2 = 0.975$). Finally, find the appropriate row corresponding to the degrees of freedom $n - 1$. A similar procedure is used to find the values for a 90 or 99% confidence interval.

Figure 7–10

Chi-Square Distribution for d.f. = $n - 1$



Example 7–13

Find the values for χ^2_{right} and χ^2_{left} for a 90% confidence interval when $n = 25$.

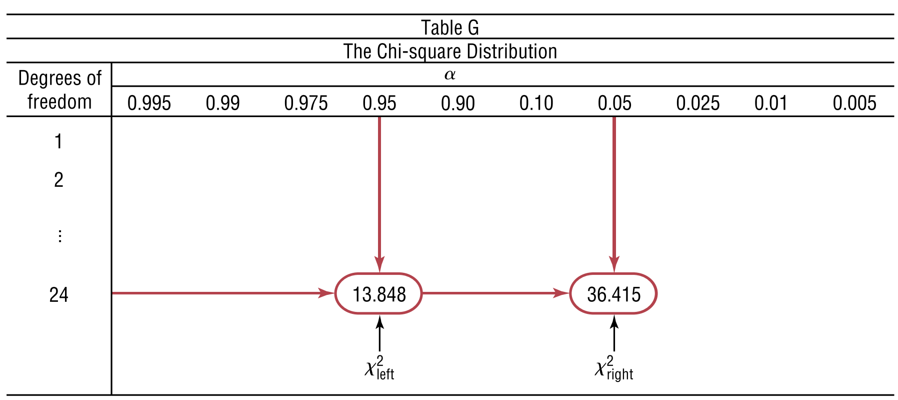
Solution

To find χ^2_{right}, subtract $1 - 0.90 = 0.10$ and divide by 2 to get 0.05.

To find χ^2_{left}, subtract $1 - 0.05$ to get 0.95. Hence, use the 0.95 and 0.05 columns and the row corresponding to 24 d.f. See Figure 7–11.

Figure 7–11

χ^2 Table for Example 7–13



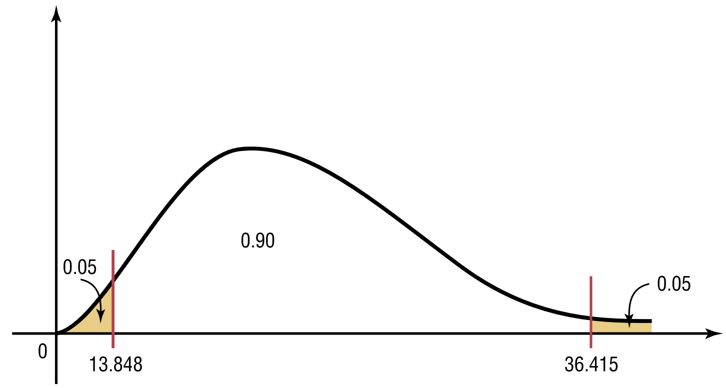
The answers are

$$\chi^2_{\text{right}} = 36.415$$
$$\chi^2_{\text{left}} = 13.848$$

See Figure 7–12.

Figure 7–12

χ^2 Distribution for Example 7–13



Useful estimates for σ^2 and σ are s^2 and s, respectively.

To find confidence intervals for variances and standard deviations, you must assume that the variable is normally distributed.

The formulas for the confidence intervals are shown here.

Formula for the Confidence Interval for a Variance

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

d.f. $= n - 1$

Formula for the Confidence Interval for a Standard Deviation

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}$$

d.f. $= n - 1$

Recall that s^2 is the symbol for the sample variance and s is the symbol for the sample standard deviation. If the problem gives the sample standard deviation s, be sure to *square* it when you are using the formula. But if the problem gives the sample variance s^2, *do not square it* when you are using the formula, since the variance is already in square units.

Assumptions for Finding a Confidence Interval for a Variance or Standard Deviation

1. The sample is a random sample.
2. The population must be normally distributed.

Rounding Rule for a Confidence Interval for a Variance or Standard Deviation When you are computing a confidence interval for a population variance or standard deviation by using raw data, round off to one more decimal place than the number of decimal places in the original data.

When you are computing a confidence interval for a population variance or standard deviation by using a sample variance or standard deviation, round off to the same number of decimal places as given for the sample variance or standard deviation.

Example 7–14 shows how to find a confidence interval for a variance and standard deviation.

Example 7–14 Nicotine Content

Find the 95% confidence interval for the variance and standard deviation of the nicotine content of cigarettes manufactured if a sample of 20 cigarettes has a standard deviation of 1.6 milligrams.

Solution

Since $\alpha = 0.05$, the two critical values, respectively, for the 0.025 and 0.975 levels for 19 degrees of freedom are 32.852 and 8.907. The 95% confidence interval for the variance is found by substituting in the formula.

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

$$\frac{(20-1)(1.6)^2}{32.852} < \sigma^2 < \frac{(20-1)(1.6)^2}{8.907}$$

$$1.5 < \sigma^2 < 5.5$$

Hence, you can be 95% confident that the true variance for the nicotine content is between 1.5 and 5.5.

For the standard deviation, the confidence interval is

$$\sqrt{1.5} < \sigma < \sqrt{5.5}$$

$$1.2 < \sigma < 2.3$$

Hence, you can be 95% confident that the true standard deviation for the nicotine content of all cigarettes manufactured is between 1.2 and 2.3 milligrams based on a sample of 20 cigarettes.

Example 7–15 Cost of Ski Lift Tickets

Find the 90% confidence interval for the variance and standard deviation for the price in dollars of an adult single-day ski lift ticket. The data represent a selected sample of nationwide ski resorts. Assume the variable is normally distributed.

59	54	53	52	51
39	49	46	49	48

Source: *USA TODAY.*

Solution

Step 1 Find the variance for the data. Use the formulas in Chapter 3 or your calculator. The variance $s^2 = 28.2$.

Step 2 Find χ^2_{right} and χ^2_{left} from Table G in Appendix C. Since $\alpha = 0.10$, the two critical values are 3.325 and 16.919, using d.f. = 9 and 0.95 and 0.05.

Step 3 Substitute in the formula and solve.

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

$$\frac{(10-1)(28.2)}{16.919} < \sigma^2 < \frac{(10-1)(28.2)}{3.325}$$

$$15.0 < \sigma^2 < 76.3$$

For the standard deviation

$$\sqrt{15} < \sigma < \sqrt{76.3}$$
$$3.87 < \sigma < 8.73$$

Hence you can be 90% confident that the standard deviation for the price of all single-day ski lift tickets of the population is between \$3.87 and \$8.73 based on a sample of 10 nationwide ski resorts. (Two decimal places are used since the data are in dollars and cents.)

Note: If you are using the standard deviation instead (as in Example 7–14) of the variance, be sure to square the standard deviation when substituting in the formula.

Applying the Concepts 7–4

Confidence Interval for Standard Deviation

Shown are the ages (in years) of the Presidents at the times of their deaths.

67	90	83	85	73	80	78	79
68	71	53	65	74	64	77	56
66	63	70	49	57	71	67	71
58	60	72	67	57	60	90	63
88	78	46	64	81	93	93	

1. Do the data represent a population or a sample?
2. Select a random sample of 12 ages and find the variance and standard deviation.
3. Find the 95% confidence interval of the standard deviation.
4. Find the standard deviation of all the data values.
5. Does the confidence interval calculated in question 3 contain the mean?
6. If it does not, give a reason why.
7. What assumption(s) must be considered for constructing the confidence interval in step 3?

See page 398 for the answers.

Exercises 7–4

1. What distribution must be used when computing confidence intervals for variances and standard deviations? Chi-square

2. What assumption must be made when computing confidence intervals for variances and standard deviations? The variable must be normally distributed.

3. Using Table G, find the values for χ^2_{left} and χ^2_{right}.

a. $\alpha = 0.05, n = 12$ 3.816; 21.920

b. $\alpha = 0.10, n = 20$ 10.117; 30.144

c. $\alpha = 0.05, n = 27$ 13.844; 41.923

d. $\alpha = 0.01, n = 6$ 0.412; 16.750

e. $\alpha = 0.10, n = 41$ 26.509; 55.758

4. Lifetimes of Wristwatches Find the 90% confidence interval for the variance and standard deviation for the lifetimes of inexpensive wristwatches if a sample of 24 watches has a standard deviation of 4.8 months. Assume the variable is normally distributed. Do you feel that the lifetimes are relatively consistent? $15.1 < \sigma^2 < 40.5$; $3.9 < \sigma < 6.4$

5. Carbohydrates in Yogurt The number of carbohydrates (in grams) per 8-ounce serving of yogurt for each of a random selection of brands is listed below. Estimate the true population variance and standard deviation for the number of carbohydrates per 8-ounce serving of yogurt with 95% confidence. $56.6 < \sigma^2 < 236.3$; $7.5 < \sigma < 15.4$

17	42	41	20	39	41	35	15	43
25	38	33	42	23	17	25	34	

6. Carbon Monoxide Deaths A study of generation-related carbon monoxide deaths showed that a sample of 6 recent years had a standard deviation of 4.1 deaths per year. Find the 99% confidence interval of the variance and standard distribution. Assume the variable is normally distributed. $5.0 < \sigma^2 < 204.0$; $2.2 < \sigma < 14.3$

Source: Based on information from Consumer Protection Safety Commission.