

12.2 CONTINGENCY TABLE *p*-VALUE

Example 12.2 (adapted from Keller, p. 503) A cola company sells four types of cola in North America. To see if the same marketing approach used in the United States can be used in Canada and Mexico, the firm first analyzes the association between cola preference and consumer nationality. The analyst classifies the population of cola drinkers by cola preference: regular, both caffeine- and sugar-free, caffeine-free only, and sugar-free only. A second classification consists of the three nationalities: American, Canadian, and Mexican. The marketing analyst then interviews a random sample of 250 cola drinkers from the three countries, classifies each by the two criteria, and records the observed frequency shown here.

Figure 12.4 Observed Frequency

	Regular	Both-free	Caffeine-free	Sugar-free
American	72	8	12	23
Canadian	26	10	16	33
Mexican	7	10	14	19

Because the marketing analyst seeks the association between cola preference and nationality, the null and alternative hypotheses to be tested are

H_0 : The two classifications are statistically independent.

H_A : The two classifications are dependent.

The following steps describe how to construct the Actual Frequency and Expected Frequency tables. Excel's CHITEST formula then obtains the chi-square *p*-value for reporting results. Although the cell references shown are appropriate for a table with three rows and four columns, this approach can be used for tables with any number of rows and columns.

Constructing the Actual Frequency Table

1. Enter the title (**Actual Frequency**), column category labels (**Regular**, and so on), Row Total label, row category labels (**American**, and so on), and **Column Total** label as shown in Figure 12.5. Make the title bold, add borders to the body of the table, and modify column widths. Enter the observed frequencies in the body of the table (B5:E7).
2. To obtain row totals, enter the SUM formula in the Row Total column of the first data row, cell F5. The formula is =SUM(B5:E5). Copy the formula to the other Row Total cells (F6:F7).

- To obtain column totals, enter the SUM formula in the Column Total row of the first data column, cell B8. The formula is =SUM(B5:B7). Copy the formula to the other Column Total cells (C8:F8).

Figure 12.5 Worksheet for Contingency Table p-Value

	A	B	C	D	E	F
1	Contingency Table					
2						
3	Actual Frequency					
4		Regular	Both-free	Caffeine-free	Sugar-free	Row Total
5	American	72	8	12	23	115
6	Canadian	26	10	16	33	85
7	Mexican	7	10	14	19	50
8	Column Total	105	28	42	75	250
9						
10	Expected Frequency					
11		Regular	Both-free	Caffeine-free	Sugar-free	Row Total
12	American	48.30	12.88	19.32	34.50	115
13	Canadian	35.70	9.52	14.28	25.50	85
14	Mexican	21.00	5.60	8.40	15.00	50
15	Column Total	105	28	42	75	250
16						
17	p-value:	1.3E-07				

Constructing the Expected Frequency Table

- Select the entire Actual Frequency table (A3:F8)—including title, labels, and marginal totals—and copy to the clipboard (click the Copy tool).
- Select a cell (A10) directly below the upper-left corner of the Actual Frequency table. Click the Paste tool or choose Paste from the Edit menu.
- Enter **Expected Frequency** in cell A10 as the title for this new table, replacing the original title.
- The formula for expected frequency in each cell is

$$\text{Expected Frequency} = \text{Column Total} * \text{Row Total} / \text{Overall Total}.$$

By using mixed references (absolute and relative) for this formula, you can enter it once in a single cell and copy it to the other cells. To do this, select the upper-left corner of the data (B12) and enter =B\$8*\$F5/\$F\$8. The mixed references to column and row totals are explained later.

Or, enter the formula in the upper-left corner of the data (B12) by pointing. Begin by typing the equals sign (=); then click on the Column Total cell (B8) at the bottom of the Actual Frequency table. With the insertion bar on the Column Total cell reference in the formula bar, press F4 twice to change the relative

reference to a mixed reference (B\$8, relative column reference and absolute row reference). Type the asterisk multiplication sign (*) and click on the Row Total cell (F5) on the right side of the Actual Frequency table. With the insertion bar on the Row Total cell reference in the formula bar, press F4 three times to change the relative reference to a mixed reference (\$F5, absolute column reference and relative row reference). Type the slash division sign (/) and click on the overall total cell (F8) in the lower-right corner of the Actual Frequency table. With the insertion bar on the overall total cell reference in the formula bar, press F4 once to change the relative reference to an absolute reference (\$F\$8, both absolute row reference and absolute column reference). Finish by clicking the formula checkmark icon, or press Enter.

8. In the new table, copy the formula in the upper-left data cell (B12) to the other cells in the body of the table (B12:E14), excluding the Column Total row and the Row Total column.
9. To format, select all numerical values in the body of the table and click the Decrease Decimal button repeatedly until the desired number of decimal places are displayed. Also, change the decimals displayed for the values in the margin of the table.

Calculating the Chi-Square *p*-Value

10. Enter the label **p-value:** in an empty cell (A17).
11. Select the next cell to the right of the label (B17). Click the Insert Function tool button (icon f_x) or from the Insert menu choose Function.
12. In the Insert Function dialog box, select Statistical for the Function Category and select CHITEST as the Function Name. Click OK.
The syntax for the CHITEST() function is
CHITEST(actual_range,expected_range).
13. In the CHITEST dialog box, select the Actual_range edit box and type **B5:E7**. Alternatively, select the edit box and point to the appropriate cells on the worksheet (click B5 and drag to E7).
14. Select the Expected_range edit box and type **B12:E14**. Alternatively, select and point.
15. After entering Actual_range and Expected_range, click the OK button.

Reporting-Approach Interpretation: The *p*-value for Example 12.2 is 1.3E-07—that is, 0.00000013. If the null hypothesis is true and the two classifications are statistically independent, the probability of obtaining the observed frequencies in a random sample of

250 is 0.00000013. Because the observed results are so unlikely under the null hypothesis, most people would reject the null hypothesis and conclude that the two classifications are dependent. Different nationalities prefer different types of cola.

12.3 CHI-SQUARE STATISTIC

To calculate the chi-square statistic, start by constructing a table containing its individual terms. Then specify a significance level (alpha) for the hypothesis test, use Excel's CHINV function to determine the critical chi-square value for the test, compare the calculated statistic with the critical chi-square value, and, finally, decide to accept or reject the hypothesis. The eventual results are shown in Figure 12.6.

Figure 12.6 Worksheet for the Chi-Square Statistic

	A	B	C	D	E
19	Chi-Square Terms				
20		Regular	Both-free	Caffeine-free	Sugar-free
21	American	11.63	1.85	2.77	3.83
22	Canadian	2.64	0.02	0.21	2.21
23	Mexican	9.33	3.46	3.73	1.07
24					
25	Chi-Square:	42.75			
26					
27	Alpha:	0.01			
28	Critical Value:	16.81			
29					
30	Decision:	Reject Ho			

1. Select part of the Actual Frequency table (A3:E7), excluding marginal totals, and copy to the clipboard. (Select and click the Copy tool button.)
2. Select a cell (A19) directly below the upper-left corner of the Actual Frequency and Expected Frequency tables and click the Paste tool button.
3. Enter **Chi-Square Terms** as the title for this new table, replacing the original title.
4. The formula for individual terms of the chi-square statistic is

$$\text{Chi-Square Term} = (\text{Actual} - \text{Expected})^2 / \text{Expected}.$$

In the Chi-Square Terms table, select the upper-left corner of the data (B21) and enter the formula $=\text{(B5-B12)}^2/\text{B12}$ by typing or by pointing to the upper-left corner cells in the Actual Frequency and Expected Frequency tables. See Figure 12.7.

Figure 12.7 Formulas for the Chi-Square Statistic (after Step 6)

	A	B
19	Chi-Square Terms	
20		Regular
21	American	= $(B5-B12)^2/B12$
22	Canadian	
23	Mexican	
24		
25	Chi-Square:	=SUM(B21:E23)
26		
27	Alpha:	0.01
28	Critical Value:	=CHIINV(B27,6)
29		
30	Decision:	=IF(B25>B28, "Reject Ho", "Accept Ho")

- In the new table, copy the formula in the upper-left data cell (B21) to the other cells in the body of the table (B21:E23).
- To complete the worksheet, enter the labels and formulas shown in Figure 12.7. The CHIINV function has the following syntax:

CHIINV(probability,deg_freedom).

The probability argument is a right-tail probability, in our case the user-specified level of significance (alpha). The deg_freedom argument for a contingency table is

Degrees of Freedom = (Number of Rows – 1) * (Number of Columns – 1),

which is $(3 - 1) * (4 - 1) = 2 * 3 = 6$ in Example 12.2.

Decision-Approach Interpretation: The calculated chi-square statistic (42.75) measures the differences between the actual frequencies and the frequencies expected under the null hypothesis. It is greater than the critical chi-square value (16.81), which corresponds to a 1% rejection region in the right tail. Therefore, there is sufficient evidence to reject the null hypothesis at the 1% level (right-tail test) and conclude that different nationalities prefer different types of cola.