The main goal of this section is to analyze two-way frequency tables. A contingency table (or two-way frequency table) is a table in which frequencies correspond to two categorical variables. A relationship between the two variables can be deduced using the Chi-Square \((\chi^2)\) Test. This test can show if there is a relationship between the row variable and the column variable. It also can be thought of as testing the claim that different populations have the same proportions of some characteristics. Each is essentially the same. The assumptions of this test are:

1. The data are randomly selected
2. The data consist of frequency counts for several cells spread across a row variable and a column variable.
3. No more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater.
4. In particular, for a \(2 \times 2\) table, the expected frequency for each cell should be at least 5.

The following notation is used:

\[
\begin{align*}
O &= \text{represents the observed frequency of an outcome} \\
E &= \text{represents the expected frequency of an outcome} \\
r &= \text{represents the number of different row categories or outcomes} \\
c &= \text{represents the number of different column categories or outcomes}
\end{align*}
\]

The chi-square test tests the hypotheses

\[
\begin{align*}
H_0 : & \text{ there is no relationship between the two categorical variables} \\
H_a : & \text{ there is a relationship between the two categorical variables}
\end{align*}
\]

The test statistic is

\[
\chi^2 = \sum \frac{(O - E)^2}{E},
\]

with \((r - 1)(c - 1)\) degrees of freedom. The \(P\)-value for this test statistic is always the area to the right of the \(\chi^2\) test statistic.

**Example:**

The accompanying table summarizes successes and failures when subjects used different methods in trying to stop smoking. The determination of smoking or not smoking was made five months after the treatment was begun, and the data are based on results from the Centers for Disease Control and Prevention. Use a 0.05 significance test to test the claim that success is independent of the method used. If someone wants to stop smoking, does the choice of the method make a difference?

<table>
<thead>
<tr>
<th></th>
<th>Nicotine Gum</th>
<th>Nicotine Patch</th>
<th>Nicotine Inhaler</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>191</td>
<td>263</td>
<td>95</td>
<td>549</td>
</tr>
<tr>
<td>Not Smoking</td>
<td>59</td>
<td>57</td>
<td>27</td>
<td>143</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>320</td>
<td>122</td>
<td>692</td>
</tr>
</tbody>
</table>

To compute the expected frequency for each cell, we use the formula

\[
E = \frac{\text{(row total)(column total)}}{\text{(grand total)}}.
\]

For example, for the cell corresponding to Nicotine Gum and Smoking, the expected count would be

\[
E = \frac{250(549)}{692}.
\]

An updated table that includes the expected counts is constructed on the next page.
<table>
<thead>
<tr>
<th></th>
<th>Nicotine Gum</th>
<th>Nicotine Patch</th>
<th>Nicotine Inhaler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>191</td>
<td>263</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>198.33815</td>
<td>253.87283</td>
<td>96.78902</td>
</tr>
<tr>
<td>Not Smoking</td>
<td>59</td>
<td>57</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>51.66185</td>
<td>66.12717</td>
<td>25.21098</td>
</tr>
</tbody>
</table>

Table 1: Observed and Expected Counts for Method versus Smoking Status.

\[ H_0: \]

\[ H_a: \]

\[ \alpha = \]

**Test Statistic:**

\[ \chi^2 = \sum \frac{(O - E)^2}{E} = \]

\[ P\text{-value}: \]

**Decision:** (Circle one:) Reject \( H_0 \) or Fail to Reject \( H_0 \)

**Conclusion:**
**Solution:**

| **H₀:** There is no relationship between the method used and being able to stop smoking |
| **H₁:** There is a relationship between the method used and being able to stop smoking |

| **α = 0.05** |
| **Test Statistic:** |

\[
\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(191 - 198.33815)^2}{198.33815} + \frac{(263 - 253.87283)^2}{253.87283} + \frac{(95 - 96.78902)^2}{96.78902} + \frac{(59 - 51.66185)^2}{51.66185} + \frac{(27 - 25.21098)^2}{25.21098} + \frac{(57 - 66.12717)^2}{66.12717} = 3.0618
\]

| **P-value:** P-value = 0.2163 |

| **Decision:** Since P-value > 0.05, then Fail to Reject H₀ |

| **Conclusion:** There is not sufficient evidence that there is a relationship between the method used and being able to stop smoking. It doesn’t seem to matter which method we choose to stop smoking. |

**Note:** When using a TI-83 or TI-84 calculator, select **STAT** → **TESTS** → **χ²-Test**, and enter the appropriate data or statistics. You must enter the two way table as a matrix. To enter the matrix, do the following:

1. Select **2ND** → **MATRIX** → **EDIT** → **[A]**
2. Select the size of the two way table. For this example, it is 2 × 3.
3. Enter the data in the matrix.
4. Select **STAT** → **TESTS** → **χ²-Test**
5. Select **Calculate**
6. Note that the expected values of each of the entries in **[A]** is found in the matrix **[B]**. To display **[B]**, press **2ND** → **MATRIX** → **NAMES** → **[B]**