Sections 11.2

Diana Pell

Section 11.2: Tests Using Contingency Tables

The test of independence of variables is used to determine whether two variables are independent of or related to each other when a single sample is selected.

The test of homogeneity of proportions is used to determine whether the proportions for a variable are equal when several samples are selected from different populations.

The chi-square independence test is used to test whether two variables are independent of each other.

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

with degrees of freedom equal to the number of categories minus 1, and where

- \( O \) = observed frequency
- \( E \) = expected frequency

Assumptions for Chi-Square Independence Test

1. The data are obtained from a random sample.
2. The expected value in each cell must be 5 or more. If the expected values are not 5 or more, combine categories.

The null hypotheses for the chi-square independence test are generally, with some variations, stated as follows:

\( H_0 \): The variables are independent of each other.
\( H_1 \): The variables are dependent upon each other.
The data for the two variables are placed in a contingency table (data arranged in table form for the chi-square independence test, with \( R \) rows and \( C \) columns).

Example of a \( 2 \times 3 \) contingency table:

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>( C_{1,1} )</td>
<td>( C_{1,2} )</td>
<td>( C_{1,3} )</td>
</tr>
<tr>
<td>Row 2</td>
<td>( C_{2,1} )</td>
<td>( C_{2,2} )</td>
<td>( C_{2,3} )</td>
</tr>
</tbody>
</table>

The formula for computing the expected values for each cell is

\[
\text{Expected value} = \frac{(\text{row sum})(\text{column sum})}{\text{grand total}}
\]

---

**Procedure Table**

**The Chi-Square Independence Test**

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value for the right tail. Use Table \( G \).

**Step 3** Compute the test value. To compute the test value, first find the expected values. For each cell of the contingency table, use the formula

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

To get the expected value. To find the test value, use the formula

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

**Step 4** Make the decision.

**Step 5** Summarize the results.
Exercise 1. Suppose a new postoperative procedure is administered to a number of patients in a large hospital. The researcher can ask the question, Do the doctors feel differently about this procedure from the nurses, or do they feel basically the same way? Note that the question is not whether they prefer the procedure but whether there is a difference of opinion between the two groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Prefer new procedure</th>
<th>Prefer old procedure</th>
<th>No preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurses</td>
<td>(75) 100</td>
<td>(100) 80</td>
<td>(25) 20</td>
</tr>
<tr>
<td>Doctors</td>
<td>(75) 50</td>
<td>(100) 120</td>
<td>(25) 30</td>
</tr>
</tbody>
</table>

1) $H_0$: The opinion about the procedure is independent of the profession.  
$H_1$: The opinion about the procedure is dependent on the profession.  

($claim$)

Note: If the null hypothesis is not rejected, the test means that both professions feel basically the same way about the procedure and that the differences are due to chance. If the null hypothesis is rejected, the test means that one group feels differently about the procedure from the other.

$3a. d.f. = (R-1)(C-1) = (2-1)(3-1) = 1 \cdot 2 = 2$

$x = 0.05$  
$\chi^2 \text{ table } G^+ \quad CV = 5.991$

$3b. \ E_{11} = \frac{200 \cdot 150}{400} = 75 \quad E_{12} = \frac{200 \cdot 200}{400} = 100 \quad E_{13} = \frac{200 \cdot 50}{400} = 25$

$E_{21} = \frac{200 \cdot 150}{400} = 75 \quad E_{22} = \frac{200 \cdot 200}{400} = 100 \quad E_{23} = \frac{200 \cdot 50}{400} = 25$

$E_{ij}$: Expected values for each cell.

$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(100-75)^2}{75} + \frac{(80-100)^2}{100} + \frac{(20-25)^2}{15} + \cdots + \frac{(30-25)^2}{25}$

$= 26.667$

$4) \ 26.667 > 5.991 \quad \text{Reject } H_0.$  

This is enough evidence to support the claim that opinion is related to profession.
Exercise 2. A researcher wishes to see if there is a relationship between the hospital and the number of patient infections. A random sample of 3 hospitals was selected, and the number of infections for a specific year has been reported. The data are shown next.

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Sternal Infections</th>
<th>Pneumonia Infections</th>
<th>Bloodstream Infections</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50,3041</td>
<td>27,81</td>
<td>40.89</td>
<td>119</td>
</tr>
<tr>
<td>B</td>
<td>32,29</td>
<td>18,76</td>
<td>27.15</td>
<td>79</td>
</tr>
<tr>
<td>C</td>
<td>162,31</td>
<td>84,72</td>
<td>131.96</td>
<td>384</td>
</tr>
<tr>
<td>Total</td>
<td>246</td>
<td>136</td>
<td>200</td>
<td>582</td>
</tr>
</tbody>
</table>

1) \( H_0: \) The \# of infections is independent of the hospital

2) \( H_1: \) The \# of infections is dependent on the hospital (claim)

2) Table 6, \( \alpha = 0.05 \)

\[ \text{d.f. } = (2-1)(2-1) = 1 \]

3) \( \chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(41 - 50.30)^2}{50.30} + \frac{(27 - 27.81)^2}{27.81} + \ldots + \frac{(109 - 131.96)^2}{131.96} = 30.698 \]

4) Reject \( H_0 \)

5) There is enough evidence to support the claim that the \# of infections is related to the hospital where they occurred.
Exercise 3. A researcher wished to see if there is a difference in the favorite sport of males and the favorite sport of females. She selected a sample of 32 males and 48 females and asked them which of three sports was their favorite. The results are shown.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Football</th>
<th>Baseball</th>
<th>Hockey</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>18 (6.2)</td>
<td>10 (6.7)</td>
<td>4 (6.4)</td>
<td>32</td>
</tr>
<tr>
<td>Female</td>
<td>20 (11.3)</td>
<td>16 (18.2)</td>
<td>12 (24.8)</td>
<td>48</td>
</tr>
<tr>
<td>Total</td>
<td>38</td>
<td>26</td>
<td>16</td>
<td>80</td>
</tr>
</tbody>
</table>

1) \( H_0 \): Sports preference is independent of the gender of the person.
   \( H_1 \): Sports preference is related to the gender of the person (claim).

2) \( C.V. = 4.605 \)

   \( d.f. = (2-1)(3-1) = 2 \)
   
   Table \( 6 \), \( \alpha = 0.10 \)

3) \( E_{ij} = \frac{(22)(38)}{30} = 15.2 \)

   \[ X^2 = \sum \frac{(O-E)^2}{E} = \frac{(18-15.2)^2}{15.2} + \ldots + \frac{(12-9.6)^2}{9.6} = 2.385 \]

4) \( 2.385 < 4.605 \)

   Do not reject the null hypothesis.

5) There is not enough evidence to support the claim that sports preference is related to the gender of the person.