10-3 Contingency Tables

Definition

- Contingency Table (or two-way frequency table)
  a table in which frequencies correspond to two variables.
  (One variable is used to categorize rows, and a second variable is used to categorize columns.)

Contingency tables have at least two rows and at least two columns.

Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>Homicide</th>
<th>Robbery</th>
<th>Assault</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stranger</td>
<td>12</td>
<td>379</td>
<td>727</td>
</tr>
<tr>
<td>Acquaintance or Relative</td>
<td>39</td>
<td>106</td>
<td>642</td>
</tr>
</tbody>
</table>
Definition

Test of Independence

tests the null hypothesis that there is no association between the row variable and the column variable.

(The null hypothesis is the statement that the row and column variables are independent.)

Assumptions

1. The sample data are randomly selected.
2. The null hypothesis \( H_0 \) is the statement that the row and column variables are independent; the alternative hypothesis \( H_1 \) is the statement that the row and variables are dependent.
3. For every cell in the contingency table, the expected frequency \( E \) is at least 5. (There is no requirement that every observed frequency must be at least 5.)

Tests of Independence

\( H_0: \) The row variable is independent of the column variable

\( H_1: \) The row variable is dependent (related to) the column variable

This procedure cannot be used to establish a direct cause-and-effect link between variables in question.

Dependence means only there is a relationship between the two variables.
Chapter 10. Section 10

Test of Independence

Test Statistic

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

Critical Values

1. Found in Table A-4 using
   degrees of freedom = (r - 1)(c - 1)
   r is the number of rows and c is the number of columns
2. Tests of Independence are always right-tailed.

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

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

Total number of all observed frequencies in the table

Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>Homicide</th>
<th>Robbery</th>
<th>Assault</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stranger</td>
<td>12</td>
<td>379</td>
<td>727</td>
</tr>
<tr>
<td>Acquaintance or Relative</td>
<td>39</td>
<td>106</td>
<td>642</td>
</tr>
</tbody>
</table>
Expected Frequency for Contingency Tables

\[ E = \frac{\text{row total} \times \text{column total}}{\text{grand total}} \]

(probability of a cell)

\[ E = \frac{(\text{row total})(\text{column total})}{\text{grand total}} \]
Is the type of crime independent of whether the criminal is a stranger?

<table>
<thead>
<tr>
<th></th>
<th>Homicide</th>
<th>Robbery</th>
<th>Assault</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stranger</td>
<td>12</td>
<td>379</td>
<td>727</td>
<td>1118</td>
</tr>
<tr>
<td>Acquaintance</td>
<td>39</td>
<td>106</td>
<td>642</td>
<td>787</td>
</tr>
<tr>
<td>or Relative</td>
<td></td>
<td></td>
<td></td>
<td>1905</td>
</tr>
</tbody>
</table>

$H_0$: Type of crime is independent of knowing the criminal

$H_1$: Type of crime is dependent with knowing the criminal

\[ E = \frac{(1118)(51)}{1905} = 29.93 \quad E = \frac{(1118)(485)}{1905} = 284.64 \]

etc.
Is the type of crime independent of whether the criminal is a stranger?

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

<table>
<thead>
<tr>
<th></th>
<th>Homicide</th>
<th>Robbery</th>
<th>Assault</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stranger</td>
<td>12 (29.93)</td>
<td>379 (803.43)</td>
<td>127 (284.64)</td>
</tr>
<tr>
<td>Acquaintance or Relative</td>
<td>59 (21.07)</td>
<td>106 (31.281)</td>
<td>642 (7.271)</td>
</tr>
</tbody>
</table>

Upper left cell: \[ \frac{(O - E)^2}{E} = \frac{(12 - 29.93)^2}{29.93} = 10.741 \]

Test Statistic \( \chi^2 = 10.741 + 31.281 + \ldots + 10.329 = 119.319 \)

Test Statistic: \( \chi^2 = 119.319 \)

with \( \alpha = 0.05 \) and \( (r - 1)(c - 1) = (2 - 1)(3 - 1) = 2 \) degrees of freedom

Critical Value: \( \chi^2 = 5.991 \) (from Table A-4)

Reject independence

\( H_0: \) The type of crime and knowing the criminal are independent

\( H_1: \) The type of crime and knowing the criminal are dependent
It appears that the type of crime and knowing the criminal are related.

\[ \chi^2 = 119.319 \]

with \( \alpha = 0.05 \) and \( (r - 1)(c - 1) = (2 - 1)(3 - 1) = 2 \) degrees of freedom

Critical Value: \( \chi^2 = 5.991 \) (from Table A-4)

Test Statistic: \( \chi^2 = 119.319 \)

Reject independence

\[ \chi^2 = 5.991 \] (from Table A-4)

It appears that the type of crime and knowing the criminal are related.

**Relationships Among Components in \( \chi^2 \) Test of Independence**

**Definition**

Test of Homogeneity tests the claim that different populations have the same proportions of some characteristics
Example - Test of Homogeneity

<table>
<thead>
<tr>
<th></th>
<th>New York</th>
<th>Chicago</th>
<th>Pittsburgh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>3</td>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td>No</td>
<td>74</td>
<td>87</td>
<td>70</td>
</tr>
</tbody>
</table>

Claim: The 3 cities have the same proportion of taxis with usable seat belts

$H_0$: The 3 cities have the same proportion of taxis with usable seat belts

$H_1$: The proportion of taxis with usable seat belts is not the same in all 3 cities

Sample data:

$X^2 = 5.991$

$\alpha = 0.05$

Fail to Reject homogeneity

There is sufficient evidence to warrant rejection of the claim that the 3 cities have the same proportion of usable seat belts in taxis; appears from Table Chicago has a much higher proportion.