11-3
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>722</td>
</tr>
<tr>
<td>Acquaintance or Relative</td>
<td>39</td>
<td>106</td>
<td>642</td>
</tr>
</tbody>
</table>

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 column 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

H₀: The row variable is independent of the column variable

H₁: 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.

Test of Independence

Test Statistic

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

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.

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

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>722</td>
</tr>
<tr>
<td>Acquaintance</td>
<td>39</td>
<td>106</td>
<td>642</td>
</tr>
</tbody>
</table>

### Expected Frequency for Contingency Tables

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

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

\[ E = n \times p \]

(\text{probability of a cell})

\[ E = n \times p \]
Expected Frequency for Contingency Tables

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

(expected frequency for 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>
</tr>
</thead>
<tbody>
<tr>
<td>Stranger</td>
<td>12</td>
<td>379</td>
<td>727</td>
</tr>
<tr>
<td>Acquaintance or Relative</td>
<td>33</td>
<td>106</td>
<td>642</td>
</tr>
</tbody>
</table>
Is the type of crime independent of whether the criminal is a stranger?

<table>
<thead>
<tr>
<th>Type of Crime</th>
<th>Stranger</th>
<th>Acquaintance or Relative</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homicide</td>
<td>12</td>
<td>39</td>
<td>1118</td>
</tr>
<tr>
<td>Robbery</td>
<td>379</td>
<td>106</td>
<td>705</td>
</tr>
<tr>
<td>Assault</td>
<td>727</td>
<td>642</td>
<td>1369</td>
</tr>
</tbody>
</table>

H₀: Type of crime is independent of knowing the criminal
H₁: Type of crime is dependent with knowing the criminal
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 or Relative</td>
<td>39</td>
<td>106</td>
<td>642</td>
<td>797</td>
</tr>
<tr>
<td>Column Total</td>
<td>51</td>
<td>485</td>
<td>1309</td>
<td>1905</td>
</tr>
</tbody>
</table>

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

\[ E = \frac{1118 \times 51}{1905} = 29.93 \]

\[ E = \frac{1118 \times 485}{1905} = 284.64 \]

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

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

<table>
<thead>
<tr>
<th></th>
<th>Homicide</th>
<th>Robbery</th>
<th>Forgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stranger</td>
<td><img src="29.93" alt="Homicide" />, <img src="21.07" alt="Robbery" />, <img src="284.64" alt="Forgery" /></td>
<td><img src="284.64" alt="Homicide" />, <img src="106" alt="Robbery" />, <img src="200.36" alt="Forgery" /></td>
<td><img src="803.43" alt="Homicide" />, <img src="642" alt="Robbery" />, <img src="565.57" alt="Forgery" /></td>
</tr>
<tr>
<td>Acquaintance or Relative</td>
<td><img src="38" alt="Homicide" />, <img src="21.07" alt="Robbery" />, <img src="284.64" alt="Forgery" /></td>
<td><img src="106" alt="Homicide" />, <img src="206.36" alt="Robbery" />, <img src="642" alt="Forgery" /></td>
<td><img src="803.43" alt="Homicide" />, <img src="565.57" alt="Robbery" />, <img src="10.329" alt="Forgery" /></td>
</tr>
</tbody>
</table>

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

Test Statistic \[ X^2 = 10.741 + 31.281 + \ldots + 10.329 = 119.319 \]
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)

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)
Test Statistic: $X^2 = 119.319$
with $\alpha = 0.05$ and $(r - 1)(c - 1) = (2 - 1)(3 - 1) = 2$ degrees of freedom
Critical Value: $X^2 = 5.991$ (from Table A-4)

$H_0$: The type of crime and knowing the criminal are independent
$H_1$: The type of crime and knowing the criminal are dependent

Sample data: $X^2 = 119.319$

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

Relationships Among Components in $X^2$ Test of Independence

Figure 11-8

- Compare the observed $O$ values to the corresponding expected $E$ values.
- Small $O$ value, large $P$-value, $X^2$ here
- Large $O$ value, small $P$-value, Reject 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>Seat Belt Use in Taxi Cabs</th>
<th>New York</th>
<th>Chicago</th>
<th>Pittsburgh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test has usable seat belt?</td>
<td>Yes</td>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>74</td>
<td>87</td>
</tr>
</tbody>
</table>

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

H₀: The 3 cities have the same proportion of taxis with usable seat belts.

H₁: The proportion of taxis with usable seat belts is not the same in all 3 cities.

Sample data:

\[ X^2 = 42.004 \]

\[ a = 0.05 \]

\[ X^2 = 5.991 \]

Fail to Reject homogeneity

Reject homogeneity
Example - Test of Homogeneity

<table>
<thead>
<tr>
<th></th>
<th>New York</th>
<th>Chicago</th>
<th>Pittsburgh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test has usable seat belt?</td>
<td>Yes</td>
<td>74</td>
<td>87</td>
</tr>
<tr>
<td>Yes</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Claim: The 3 cities have the same proportion of taxis with usable seat belts
H₀: The 3 cities have the same proportion of taxis with usable seat belts
H₁: The proportion of taxis with usable seat belts is not the same in all 3 cities

Sample data:
\[ \chi^2 = 42.004 \]
\[ a = 0.05 \]
\[ \chi^2 = 5.991 \]

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.