STA 4033
Mathematical Statistics
with Computer Applications
Lecture 13*

Brett Presnell
Department of Statistics
University of Florida

6 June 2001
5. Chi-Square Tests and Analysis of Contingency Tables

5.1. Comparing Multinomial Distributions

Example. (From Mooore, The Basic Practice of Statistics, 2000) A three-year study compared the antidepressant desipramine with lithium and a placebo as treatments for cocaine addiction. The subjects were 72 chronic users who wanted to break their cocaine habit. Twenty-four of the subjects were randomly assigned to each treatment, with the following results:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Y</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Desipramine</td>
<td>14</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>2. Lithium</td>
<td>6</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>3. Placebo</td>
<td>4</td>
<td>20</td>
<td>24</td>
</tr>
</tbody>
</table>

The sample proportions of patients not relapsing are

\[ \hat{p}_1 = \frac{14}{24} = 0.583 \quad \hat{p}_2 = \frac{6}{24} = 0.250 \quad \hat{p}_3 = \frac{4}{24} = 0.167 \]

Question: Could the differences in observed success rates be due to chance alone?
In the last example, let $p_1$, $p_2$, and $p_3$ be the proportion of all possible patients not relapsing under each of the three treatments. We wish to compare these proportions in a statistical test.

- Could carry out three separate tests (D vs L, D vs P, and L vs P) for equality of proportions.

\[
\begin{align*}
H_0 : p_1 &= p_2 & \text{vs} & & H_a : p_1 \neq p_2 \\
H_0 : p_1 &= p_3 & \text{vs} & & H_a : p_1 \neq p_3 \\
H_0 : p_2 &= p_3 & \text{vs} & & H_a : p_2 \neq p_3
\end{align*}
\]

- Problem is that overall error rate is greater than the nominal $\alpha$ (say .05) because we are making *multiple comparisons*.

- Should do an overall test first:

\[
H_0 : p_1 = p_2 = p_3 \quad \text{vs} \quad H_a : \text{at least two of the } p_j \text{ are different}
\]
Suppose there is no difference between the three treatments. Then an estimate of the overall success rate is

$$\hat{p}_{\text{overall}} = \frac{14 + 6 + 4}{72} = \frac{24}{72} = \frac{1}{3} = 0.333$$

and the estimated overall failure rate is

$$\hat{q}_{\text{overall}} = \frac{10 + 18 + 20}{72} = \frac{48}{72} = \frac{2}{3} = 0.667$$

Thus under the hypothesis of no difference in the treatments we would estimate the expected counts in the table to be

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Desipramine</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Lithium</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Placebo</td>
<td>4</td>
<td>20</td>
</tr>
</tbody>
</table>

In general, the “expected” count for each table cell is calculated as

$$\text{Expected} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$
The chi-square test statistics is based on the difference in observed and expected frequencies. The general form is

\[ X^2 = \sum_{\text{all cells}} \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}} \]

Large values of \( X^2 \) favor \( H_a \) over \( H_0 \).

**Example.** In the example, we compute

\[
\begin{pmatrix}
\frac{(14 - 8)^2}{8} & \frac{(10 - 16)^2}{16} \\
\frac{(6 - 8)^2}{8} & \frac{(18 - 16)^2}{16} \\
\frac{(4 - 8)^2}{8} & \frac{(20 - 16)^2}{16}
\end{pmatrix}
= \begin{pmatrix}
4.5 & 2.25 \\
0.5 & 0.25 \\
2.0 & 1.00
\end{pmatrix}
\]

so

\[ X^2 = 4.5 + 2.25 + 0.5 + 0.25 + 2.0 + 1.00 = 10.5 \]

Under \( H_0 \), the sampling distribution of \( X^2 \) is approximately chi-square with

\[ \text{df} = (\text{number of rows} - 1) \times (\text{number of columns} - 1) \]

So

- reject \( H_0 \) at significance level \( \alpha \) if \( X^2 \geq \chi^2_{\alpha} \)
- \( P\text{-value} = P(\chi^2 \geq X_{\text{obs}}^2) \)

**Example.** In the example,

\[ P\text{-value} = .0052 \]

There is strong evidence of a difference in the effectiveness of the three treatments.
5.2. Other Uses of the Chi-Square Test

In the previous example, we compared three populations. The variable measured was a two-category response (relapse: no or yes). We can compare the distribution of a multicategory response across several populations in the same way.

The chi-square test is also appropriate in the following situation. Suppose we draw a random sample from some population and measure two categorical variables, e.g., political party affiliation and religious affiliation, for each sampled unit.

- Let

  \( p_{ij} \) represent the probability that an observation has level \( i \) of the first variable and level \( j \) of the second.

  \( p_{i+} \) represent the probability that an observation has level \( i \) of the first variable.

  \( p_{+j} \) represent the probability that an observation has level \( j \) of the second variable.

- The two variables are independent if and only if

  \[ p_{ij} = p_{i+} \times p_{+j}. \]
The data can be summarized in a contingency table, with $n_{ij}$ giving the number of observations having level $i$ of the first variable and level $j$ of the second.

- If the total sample size is $n$, then the expected value of $n_{ij}$ is $n \times p_{ij}$.
- Under the null hypothesis of independence, estimate the expected value of $n_{ij}$ by

$$
\hat{\mu}_{ij} = n \times \hat{p}_{i+} \times \hat{p}_{+j} = n \times \frac{n_{i+}}{n} \times \frac{n_{+j}}{n} = \frac{n_{i+}n_{+j}}{n}
$$

$$
= \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}
$$

This leads to the same test statistic as before as a test of independence of the two variables:

$$
X^2 = \sum_{ij} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} = \sum_{\text{all cells}} \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}
$$

As before, the sampling distribution of $X^2$ under the null hypothesis is approximately chi-square with

$$
df = (\text{number of rows} - 1) \times (\text{number of columns} - 1)
$$
**Example.** (1998 General Social Survey) Consider the following table:

<table>
<thead>
<tr>
<th>Religious Affiliation</th>
<th>Party Preference</th>
<th>Democrat</th>
<th>Independent</th>
<th>Republican</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td></td>
<td>525</td>
<td>502</td>
<td>468</td>
<td>1495</td>
</tr>
<tr>
<td>Catholic</td>
<td></td>
<td>252</td>
<td>281</td>
<td>158</td>
<td>691</td>
</tr>
<tr>
<td>Jewish</td>
<td></td>
<td>21</td>
<td>18</td>
<td>11</td>
<td>50</td>
</tr>
<tr>
<td>None</td>
<td></td>
<td>121</td>
<td>201</td>
<td>55</td>
<td>377</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>919</td>
<td>1002</td>
<td>692</td>
<td>2613</td>
</tr>
</tbody>
</table>

Wish to test

\[ H_0 : \text{religious and party affiliations are independent} \]

against the alternative that religious and party affiliations are dependent.

Under the null hypothesis of independence, the expected number of, e.g., democratic protestants is

\[ \hat{\mu}_{11} = \frac{1495 \times 919}{2613} = 525.8 \]

For the entire table, the expected counts are

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>Independent</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td>525.8</td>
<td>573.3</td>
<td>395.9</td>
</tr>
<tr>
<td>Catholic</td>
<td>243.0</td>
<td>265.0</td>
<td>183.0</td>
</tr>
<tr>
<td>Jewish</td>
<td>17.6</td>
<td>19.2</td>
<td>13.2</td>
</tr>
<tr>
<td>None</td>
<td>132.6</td>
<td>144.6</td>
<td>99.8</td>
</tr>
</tbody>
</table>
Thus the components of the chi-square statistic are

\[
\begin{pmatrix}
\frac{(525 - 525.8)^2}{525.8} & \frac{(502 - 573.3)^2}{573.3} & \frac{(468 - 395.9)^2}{395.9} \\
\frac{(252 - 243)^2}{243.0} & \frac{(281 - 265)^2}{265.0} & \frac{(158 - 183)^2}{183.0} \\
\frac{(21 - 17.6)^2}{17.6} & \frac{(18 - 19.2)^2}{19.2} & \frac{(11 - 13.2)^2}{13.2} \\
\frac{(121 - 132.6)^2}{132.6} & \frac{(201 - 144.6)^2}{144.6} & \frac{(55 - 99.8)^2}{99.8}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
0.0 & 8.9 & 13.1 \\
0.3 & 1.0 & 3.4 \\
0.7 & 0.1 & 0.4 \\
1.0 & 22.0 & 20.1
\end{pmatrix}
\]

and the chi-square statistics is

\[X^2 = 0.0 + 8.9 + \cdots + 20.1 = 71.0\]

Under the null hypothesis of independence, the distribution of \(X^2\) is approximately chi-square with

\[df = (4 - 1) \times (3 - 1) = 6.\]

The \(P\)-value for this test is tiny (not surprising because of the large sample size):

\[P\text{-value} \approx 2.55 \times 10^{-13}\]

Thus there is very strong evidence against the null hypothesis of independence (though we don’t take the precise value of the \(P\)-value too seriously).

### 5.3. Follow Up

A significant chi-square test alone is not very informative. One simple yet effective way to describe the dependence in the data is to compare either row or column percentages, perhaps graphically. Note that
• if the row variable is best thought of as a predictor and the column variable as response, then we should look at row percentages.

• if the column variable is best thought of as a predictor and the row variable as response, then we should look at column percentages.

Another way to explore the results of a chi-square test, is to look for the table cells that contribute a large amount to the value of the chi-square statistic. By comparing observed and expected values, one then sees how the count in each of these cells differs from what would be expected under independence. However, this leaves open the question of what constitutes a large contribution to the chi-square statistic.

A better approach is to use the adjusted residuals,

$$r_{ij} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1 - p_i)(1 - p_j)}}$$

When the sample size is large, each adjusted residuals has approximately a standard normal distribution under the null hypothesis. An adjusted residual that exceeds 2 or 3 in absolute value indicates a lack of fit of $H_0$ (the independence hypothesis) in that cell. The sign of the adjusted residual indicates whether the cell count was greater or less than the expected value under independence.
Example. In the last example, the row proportions are

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>Independent</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td>0.351</td>
<td>0.336</td>
<td>0.313</td>
</tr>
<tr>
<td>Catholic</td>
<td>0.365</td>
<td>0.407</td>
<td>0.229</td>
</tr>
<tr>
<td>Jewish</td>
<td>0.420</td>
<td>0.360</td>
<td>0.220</td>
</tr>
<tr>
<td>None</td>
<td>0.321</td>
<td>0.533</td>
<td>0.146</td>
</tr>
</tbody>
</table>

These are plotted in Figure 1.

The adjusted residuals are

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>Independent</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td>-0.1</td>
<td>-5.8</td>
<td>6.5</td>
</tr>
<tr>
<td>Catholic</td>
<td>0.8</td>
<td>1.5</td>
<td>-2.5</td>
</tr>
<tr>
<td>Jewish</td>
<td>1.0</td>
<td>-0.3</td>
<td>-0.7</td>
</tr>
<tr>
<td>None</td>
<td>-1.4</td>
<td>6.5</td>
<td>-5.7</td>
</tr>
</tbody>
</table>

From this we see that the

- Protestant-Independent count is lower than expected
- Protestant-Republican count is higher than expected
- Catholic-Republican count is lower than expected
- None-Independent count is higher than expected
- None-Republican count is lower than expected
Figure 1: Political and religious affiliation.