2x2 Table Analysis

The 2x2 Table Analysis command calculates the following statistics for 2-by-2 contingency tables: chi-square, Yates-corrected chi-square, the Fisher Exact Test, Phi-Square, the McNemar Change Test and also indices relevant to various special kinds of 2-by-2 tables. The command can be used to summarize the relationship between several categorical variables, it is a categorical equivalent of the scatterplot used to analyze the relationship between two continuous variables [SRB].

A 2x2 table contains four cells with frequencies:

<table>
<thead>
<tr>
<th>Test result - Yes (Positive)</th>
<th>Observed - Yes</th>
<th>Observed – No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>B</td>
<td>TP+FP</td>
</tr>
<tr>
<td>True positive (TP)</td>
<td></td>
<td>False positive (FP)</td>
<td></td>
</tr>
<tr>
<td>Test result – no (Negative)</td>
<td>C</td>
<td>D</td>
<td>FN+TN</td>
</tr>
<tr>
<td>False negative (FN)</td>
<td></td>
<td>True negative (TN)</td>
<td></td>
</tr>
<tr>
<td>Marginal total for observations</td>
<td>A+C</td>
<td>B+D</td>
<td>n = A+B+C+D</td>
</tr>
</tbody>
</table>

How To

- Run: Statistics->Nonparametric-> 2x2 Table Analysis (tabulated data).
- Enter the A, B, C, D cell values.
  - To tabulate raw data use the Cross-tabulation command.
- Run the analysis.

Results

Chi-square – is a statistics used to examine the relationship between categorical variables. The contingency chi-square is based on the same principles as the ordinary chi-square analysis where expected vs. observed frequencies are being checked.

\[
\chi^2 = \sum \frac{(O_{b s r} - E_{xp c t})^2}{E_{xp c t}}
\]

For 2x2 tables the expected value can be calculated as:

\[
f_e = \frac{(N_r)(N_c)}{N}
\]

where \( N_r \) – is the total number of cases in the particular row or TP+FP, \( N_c \) – is the total number in the particular column or A+C, \( N \) is the number of A+B+C+D in the full sample.

Yates Corrected Chi-square - is a correction made to explain the fact that both Pearson’s chi-square test and McNemar’s chi-square test are biased upwards for a 2 x 2 contingency table. It is defined as [YAC]:

\[
[YAC]
\]
\[ \chi^2_{Yates} = \frac{N(|AD - BC| - \frac{N}{2})^2}{(A + B)(C + D)(A + C)(B + D)} \]

**McNemar Test** – is applied to 2 by 2 contingency tables with a dichotomous trait, with matched pairs of subjects, to determine whether the row and column marginal frequencies are equal.

It is calculated as:

\[ \chi^2 = \frac{(B - C)^2}{B + C} \]

**Pearson’s Coefficient of Contingency** is defined as following:

\[ Pearson = \frac{\chi^2}{\sqrt{\chi^2 + n}} \]

The coefficient varies between 0 (no relationship) and 1 (strong relationship) depending on a size of the table (for a 2 × 2 table the maximum value is 0.707). That’s why it should be used only to compare tables with the same sizes.

**Cramer’s (V) Coefficient of Contingency** reflects the strength of the association in a contingency table and is calculated as:

\[ V = \frac{AD - BC}{\sqrt{(A + B)(C + D)(A + C)(B + D)}} \]

This coefficient is a modified version of the Phi-square and varies between 0 (no relationship) and 1 (strong relationship).

**Fisher Corrected** – is an alternative to the chi-square test if the total number of observations is less than 20. Also known as Fisher’s Exact Test.

**Phi-square (Mean Square Contingency Coefficient)** – is a measure of association for two binary variables and is defined as: \( \varphi^2 = \frac{\chi^2}{n} \).

**Odds Ratio (OR)** – defined as \( OR = \frac{A + B}{C + D} \). Odds Ratio is one of three main ways to quantify how strong the presence or absence of property A is associated with the presence or absence of property B in a given population. Odds ratio (OR) is related to risk ratio.

**Relative Risk (RR)**. Together with odds ratio is the main measure of association in observational studies:

\[ Relative \ Risk = \frac{A/(A + C)}{B/(B + D)} \]

**References**
