**CHI-SQUARE TEST**

**CONTINGENCY TABLE** is a table in which frequencies correspond to two variables. One variable is used to categorize rows, and a second to categorize columns.

A **test of independence** tests the null hypothesis that there is no association between the row variable and the column variable in a contingency table. (For the null hypothesis, we will use the statement that “the row and column variables are independent”).

**Assumptions:**
1. The sample data is a random sample.
2. The null hypothesis is that the row and column variable are independent; the alternative hypothesis H1 is the statement that the row and column variables are not independent.
3. For every cell in the contingency table, the expected frequency is at least 5.

**Test statistic** for a Test of Independence (same for homogeneity)

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

**Critical Values**
1. The critical values are found in Table A-4 by using degrees of freedom = (r-1)(c-1), r = # of rows, c = # of cols.
2. In a test of independence with a contingency table, the critical region is located at the right tail only.

**Example:** Racial Profiling

<table>
<thead>
<tr>
<th></th>
<th>Black and non-hispanic</th>
<th>White and non-hispanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stopped by police</td>
<td>a = 24</td>
<td>b = 147</td>
</tr>
<tr>
<td>Not stopped by police</td>
<td>c = 176</td>
<td>d = 1253</td>
</tr>
</tbody>
</table>

Compute expected values: use the following procedure

\[
c1 = \frac{(a+c)(a+b)}{a+b+c+d} = 21.375 \quad \frac{(b+d)(a+b)}{a+b+c+d} = 149.625
\]

\[
c2 = \frac{(a+c)(c+d)}{a+b+c+d} = 178.625 \quad \frac{(b+d)(c+d)}{a+b+c+d} = 1,250.375
\]

\[
\chi^2 = \frac{(24 - 21.375)^2}{21.375} + \frac{(147 - 149.625)^2}{149.625} + \frac{(176 - 178.625)^2}{178.625} + \frac{(1253 - 1250.375)^2}{1250.375} = 0.4125
\]

From Table A-4: Critical value for \( \alpha = .05 \), 1 d.o.f. \( \chi^2 = 3.841 \), fail to reject.