Chi Square Test

The Chi-Square test of independence is a statistical test to determine if there is a significant relationship between 2 categorical variables.  In simple words, the Chi-Square statistic will test whether there is a significant difference in the observed vs the expected frequencies of both variables.

The Chi-Square statistic is calculated as follows:

http://latex.codecogs.com/gif.latex?\chi&space;=&space;\sum&space;\frac%7b(Observed-Expected)%5e%7b2%7d%7d%7bExpected%7d

http://latex.codecogs.com/gif.latex?Expected&space;=&space;\frac%7bRowTotal&space;\times&space;Column&space;Total%7d%7bOverall&space;Total%7d

http://latex.codecogs.com/gif.latex?DF&space;=&space;\left&space;(&space;&hash;&space;Rows&space;-&space;1&space;\right&space;)&space;\times&space;\left&space;(&space;&hash;&space;Columns&space;-&space;1&space;\right&space;)

The **Null** hypothesis is that there is NO association between both variables.

The **Alternate** hypothesis says there is evidence to suggest there is an association between the two variables.

To reject the null hypothesis, the calculated P-Value needs to be below a defined threshold. Say, if we use an alpha of .05, if the p-value < 0.05 we reject the null hypothesis.

**Rules to use the Chi-Square Test:**

1. Variables are Categorical

2. Frequency is at least 5

3. Variables are sampled independently