**Null Hypothesis**: The two categorical variables are independent.

**Alternative Hypothesis**: The two categorical variables are dependent.

The chi-square test statistic is calculated by using the formula:

 χ2=∑(O−E)2/Eχ2=∑(O−E)2/E

where O represents the observed frequency. E is the expected frequency under the null hypothesis and computed by:

 E=row total×column totalsample size

We will compare the value of the test statistic to the critical value of χ2αχα2 with degree of freedom = (*r* - 1) (*c* - 1), and reject the null hypothesis if χ2>χ2αχ2>χα2.

Example

Is gender independent of education level? A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | High School | Bachelors | Masters | Ph.d. | Total |
| Female | 60 | 54 | 46 | 41 | 201 |
| Male | 40 | 44 | 53 | 57 | 194 |
| Total | 100 | 98 | 99 | 98 | 395 |

Here's the table of expected counts:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | High School | Bachelors | Masters | Ph.d. | Total |
| Female | 50.886 | 49.868 | 50.377 | 49.868 | 201 |
| Male | 49.114 | 48.132 | 48.623 | 48.132 | 194 |
| Total | 100 | 98 | 99 | 98 | 395 |

So, working this out, χ2=(60−50.886)2/50.886+⋯+(57−48.132)2/48.132=8.006χ2=(60−50.886)2/50.886+⋯+(57−48.132)2/48.132=8.006

The critical value of χ2χ2 with 3 degree of freedom is 7.815. Since 8.006 > 7.815, therefore we reject the null hypothesis and conclude that the education level depends on gender at a 5% level of significance.

Session5\_Assignment1