Chi-Square Test of Independence

Null Hypothesis: The two categorical variables (Gender, Education Level) are independent.

Alternative Hypothesis: The two categorical variables(Gender, Education Level) are dependent.

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

*χ*2=∑(*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 total /sample size

We will compare the value of the test statistic to the critical value of *χ*2*α* with degree of freedom = (*r* - 1) (*c* - 1), and reject the null hypothesis if *χ*2>*χ*2*α*.

Here's the table of Observed counts ( data that resulted from the survey):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 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, *χ*2=(60−50.886)^2/50.886+⋯+(57−48.132)^2/48.132=8.006

The critical value of *χ*2 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.