Question : 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

Are gender and education level dependent at 5% level of significance? In

other words, given the data collected above, is there a relationship between the gender

of an individual and the level of education that they have obtained?

Answer : **The data displayed in tabular format for better understanding.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gender/Education** | **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, I use Chi-square test to test the independence of the categorical features.**

**Ho(Null Hypothesis) : Gender and Education are Independent.**

**Ha(Alternative Hypothesis) : Gender and Education are dependent.**

**χ2=∑(O−E)^2/E where χ2 is Chi-square, O is Observed Frequency and E is Expected Frequency**

**E = (Total Row \* Total Column)/Size of Sample where Size of Sample is given as 395**

**Hence, 1>Observation value is 60 and it's related Expected value is (201\*100)/395 = 50.886**

**2>Observation value is 40 and it's related Expected value is (194\*100)/395 = 49.113**

**3> Observation value is 54 and it's related Expected value is (201\*98)/395 = 49.868**

**4> Observation value is 44 and it's related Expected value is (194\*98)/395 = 48.131**

**5> Observation value is 46 and it's related Expected value is (201\*99)/395 = 50.377**

**6> Observation value is 53 and it's related Expected value is (194\*99)/395 = 48.622**

**7> Observation value is 41 and it's related Expected value is (201\*98)/395 = 49.868**

**8> Observation value is 57 and it's related Expected value is (194\*98)/395 = 48.131**

**So, χ2= (60-50.886)^2/50.886 + (40-49.113)^2/49.113 + (54-49.868)^2/49.868 + (44-48.131)^2/48.131 + (46-50.377)^2/50.377 + (53-48.622)^2/48.622 + (41-49.868)^2/49.868 + (57-48.131)^2/48.131 = 1.632 + 1.690 + 0.342 + 0.354 + 0.380 + 0.394 + 1.576 + 1.634 = 8.002**

**Degree of freedom = (Number of Columns -1)\*(Number of Rows -1) = (4-1)\*(2-1) = 3\*1 = 3**

**Now, referring Chi Squared table with input params (Degree of freedom = 3 and level of significance = 0.05), the critical value for χ2 is found to be 7.814.**

**It's found that the calculated χ2 value(8.002) is greater than Critical χ2 value(7.814).**

**Hence, I can infer that Alternative Hypothesis is true here which means Gender and Education are dependent each other.**