

Assignment 19.1

Is gender independent of education level? A random sample of 395 people was 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	Bachelor	Master	Phd	Total
Female	60	54	46 46	41	201
Male	40	44	53 53	57	194
Total	100	98	99 99	98	395

Question:- Are gender and education level dependent at 5% level of significance? In other words, given the data collected above, is there a relationship between gender of an individual and the level of education that they have obtained?

Solution:- I will be using Chi-Square Test of independence to test independence of two categorical variables Gender and Education level

Null Hypothesis (H_0):- Gender and Education level are independent

Alternative Hypothesis (H_A):- Gender and Education level are dependent

First, Calculate E, the Expected Frequency under the Null hypothesis using formula:- $E = \frac{\text{row total} \times \text{column total}}{\text{sample size}}$ for each entry in table as below

	High School	Bachelor	Master	Phd	Total
Female	$\frac{201 \times 100}{395} = 50.886$	$\frac{201 \times 98}{395} = 49.868$	$\frac{201 \times 99}{395} = 50.377$	$\frac{201 \times 98}{395} = 49.868$	201
Male	$\frac{194 \times 100}{395} = 49.114$	$\frac{194 \times 98}{395} = 48.132$	$\frac{194 \times 99}{395} = 48.623$	$\frac{194 \times 98}{395} = 48.132$	194
Total	100	98	99	98	395

The chi-square test statistic (χ^2) is calculated using the formula

$$\chi^2 = \sum (O - E)^2 / E$$

Where O is Observed frequency

E is Expected frequency

$$\begin{aligned}\chi^2 &= \frac{(60 - 50.886)^2}{50.886} + \frac{(51 - 49.868)^2}{49.868} + \frac{(46 - 50.377)^2}{50.377} \\ &+ \frac{(41 - 49.868)^2}{49.868} + \frac{(40 - 49.114)^2}{49.114} + \frac{(44 - 48.132)^2}{48.132} \\ &+ \frac{(53 - 48.623)^2}{48.623} + \frac{(57 - 48.132)^2}{48.132} \\ &= 1.632 + 0.342 + 0.380 + 1.577 + 1.691 + 0.355 + 0.394 + 1.634 \\ &= 8.005\end{aligned}$$

~~Degree of freedom = (No. of rows - 1) * (No. of columns - 1)~~
Degree of freedom = (No. of rows - 1) * (No. of columns - 1)
 $= (2 - 1) * (4 - 1) = 3$

Critical value of χ^2 with 3 degree of freedom
at 5% significance level = 7.81

Since $8.005 > 7.81$, therefore, rejecting the Null hypothesis
Hence Education level depends on gender at 5% level of
significance.