*Session 5: Assignment 1*

**Table of Contents**

1. Introduction
2. Problem Statement
3. Output

1. Introduction

This assignment will help you to consolidate the concepts learnt in the session.

1. Problem Statement

**Problem Statement 1:**

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

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 the gender

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

**Note: Solution submitted via github must contain all the detailed steps.**

**3. Output:**

Step 1: Use formulae χ2=∑(O−E)2/E: O is Observed Frequency, to find the chi-square test statistic

Step 2: Calculate Expected Frequency (E)= row total × column total / sample size

Step 3: Compare the value of the test statistic to the critical value of χ2α with degree of freedom, (r - 1) (c - 1) and reject the null hypothesis if χ2>χ2α.

Step 4: **Find** Df = (r-1)(c-1) = (4-1)(2-1) = (3)(1) = 3

Step 5: **Given Sample Size=395**, then:

Ef (High School) = 201 \* 100/395 = **50.886 &** Em (High School) = 194\*100/395 = **49.114**

Ef (Bachelors)= 201\*98/395 = **49.868** & Em(Bachelors)=194\*98/395 = **48.132**

Ef(Masters) = 201\*99/395 = **50.377 &** Em(Masters)= 194\*99/395 = **48.623**

Ef(Phd.) = 201\*98/395 = **49.868 &** Em(Phd.) = 194\*98/395 = **48.132**

Step 6: Prepare the table of expected frequency counts:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Expected Frequency** | **High School** | **Bachelors** | **Masters** | **Ph.d.** | **Total** |
| **Female (Ef)** | **50.886** | **49.868** | **50.377** | **49.868** | **201** |
| **Male (Em)** | **49.114** | **48.132** | **48.623** | **48.132** | **194** |
| **Total** | **100** | **98** | **99** | **98** | **395** |

So, **χ2=** (60−50.886)2/50.886+⋯+ (57−48.132)2/48.132

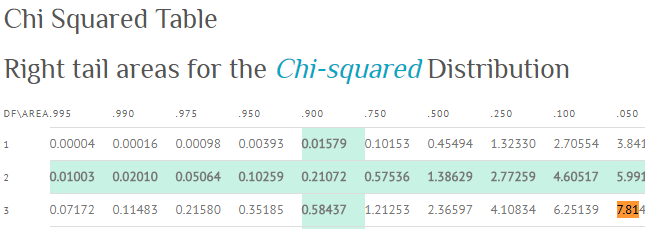
=> (60-50.886)(60-50.886)/50.886 + (54- 49.868)(54-49.868)/49.868 + (46-50.377)(46-50.377)/50.377 +

(41-49.868)(41-49.868)/49.868 + (40-49.114)(40-49.114)/49.114 + (44-48.132)(44-48.132)/48.132 +

(53-48.623)(53-48.623)/48.623 + (57-48.132)(57-48.132)/48.132

=> 1.632 + 0.342 + 0.380 + 1.577 + 1.691 + 0.355 + 0.394 + 1.634 = **8.006**

Step 7: Critical value of **χ2**with **3 df** for alpha value of 0.05 is **7.814** as per the table below:



***As, 8.006 > 7.815, therefore we reject the null hypothesis.***

***Yes, gender and education level are dependent at 5% level of significance.***