Assignment 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?

Null Hypothesis: H0 - The gender and education level are independent

Alternate Hypothesis: HA - The gender and education level are dependent

The chi-square test is calculated using the formula:

χ2 = ∑(Oi – Ei)^2/Ei

Oi – Observed Frequency

Ei – Expected Frequency

E = row total \* column total

sample size

Critical Value of χ2α with degree of freedom = (r-1)(c-1)

= (2-1)(4-1)

= 1 \* 3 = 3

Critical Value of χ2 with degree of freedom as 3 = 7.815

If χ2 > χ2α then we reject the null hypothesis

Here's the table of expected counts:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | High School | Bachelors | Masters | Ph.d. | Total |
| Female | (201\*100)/395 | (201\*98)/395 | (201\*99)/395 | (201\*98)/395 | 201 |
| Male | (194\*100)/395 | (194\*98)/395 | (194\*99)/395 | (194\*98)/395 | 194 |
| Total | 100 | 98 | 99 | 98 | 395 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 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 |

χ2 = (60−50.886) ^2 /50.886+(54-49.868)^2/49.868+(46-50.377)^2/50.377+(41-49.868)^2/49.868+(40-49.114)^2/49.114+(44-48.132)^2/48.132+(53-48.623)^2/48.623+(57−48.132)2/48.132=8.006

χ2 = 8.006

The critical value of χ2 with 3 degree of freedom is 7.815. Since 8.006 > 7.815, therefore we reject the null hypothesis i.e the gender and education level are not independent and conclude that the education level depends on gender at a 5% level of significance.

**Assignment 2**

Using the following data, perform a oneway analysis of variance using α=.05. Write up the results in APA format.

[Group1: 51, 45, 33, 45, 67]

[Group2: 23, 43, 23, 43, 45]

[Group3: 56, 76, 74, 87, 56]

Sample mean of 3 groups:

Group 1: 48.2

Group 2: 35.4

Group 3: 69.8

Variance 1 = 153.2

Variance 2 = 128.8

Variance 3 = 183.2

Mean of all variance is = (153.2 + 128.8 + 183.2)/3

= 155.0667

Dferror = 15-3 = 12

SSerror = 155.0667 \* 12 = 1860.8

Grand Mean(x\_bar grand) = (48.2 + 35.4 + 69.8)/3

Variance Mean = 302.29

MSgroup = (302.29)(5) = 1511.45

Dfgroups = 3-1 = 2

SSgroup = 1511.45 \* 2 = 3022.9

Test Statistic and critical value

F = 1511.45/155.067 = 9.75

Fcritical(2,12) = 3.89

Since F > Fcritical, hence reject Null Hypothesis.

Effect Size = η2 = SSgroup/(SSgroup +SSerror) = 3022.9/(3022.9+1860.8)

= 3022.9 / 4883.7 = 0.62

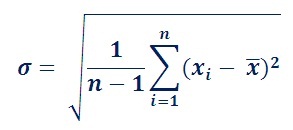
**APA format**

F(2,12) = 9.75

P < 0.05

η2=0.62.

**Assignment 3**

F Test is generally defined as ratio of the variances of the given two set of values. First [calculate standard deviation](http://ncalculators.com/statistics/mean-standard-deviation-calculator.htm) and variation of the given set of values. The formula used to calculate SD is,   
**Standard Deviation Formula**   
   
The standard deviation is represented by the symbol ᵟ and variance is square of the standard deviation.   
The formula used to calculate F Test is,   
**F Test Formula**

F = estimate of Variance from means

estimate of Variance from individuals

**For Example:**   
Calculate F Test for given 10,20,30,40,50 and 5,10,15,20,25.   
For 10, 20,30,40,50:   
  
**Calculate Variance of first set**   
  
Total Inputs (N) =(10,20,30,40,50)   
Total Inputs (N)=5   
Mean (xm)= (x1+x1+x2...xn)/N   
Mean (xm)= 150/5   
Means(xm)= 30   
SD=sqrt(1/(N-1)\*((x1-xm)2+(x2-xm)2+..+(xn-xm)2))   
=sqrt(1/(5-1)((10-30)2+(20-30)2+(30-30)2+(40-30)2+(50-30)2))   
=sqrt(1/4((-20)2+(-10)2+(0)2+(10)2+(20)2))   
=sqrt(1/4((400)+(100)+(0)+(100)+(400)))   
=sqrt(250)   
=15.8114   
Variance=SD2   
Variance=15.81142   
Variance=250   
  
**Calculate Variance of second set**   
For 5, 10,15,20,25:   
Total Inputs(N) =(5,10,15,20,25)   
Total Inputs(N)=5   
Mean (xm)= (x1+x2+x3...xN)/N   
Mean (xm)= 75/5   
Means (xm)= 15   
SD=sqrt(1/(N-1)\*((x1-xm)2+(x2-xm)2+..+(xn-xm)2))   
=sqrt(1/(5-1)((5-15)2+(10-15)2+(15-15)2+(20-15)2+(25-15)2))   
=sqrt(1/4((-10)2+(-5)2+(0)2+(5)2+(10)2))   
=sqrt(1/4((100)+(25)+(0)+(25)+(100)))   
=sqrt(62.5)   
=7.9057   
Variance=SD2   
Variance=7.90572   
Variance=62.5   
  
**To calculate F Test**   
F Test = (variance of 10, 20,30,40,50) / (variance of 5, 10, 15, 20, 25)   
= 250/62.5   
= 4.   
  
The F Test value is 4.