**Session 18 Assignment 1**

**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?

**Answer:**

Here is 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, χ2 = (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.**

chi - sq = 8.006, DF = 3, P-Value = 0.046

The Chi-square test of independence value that Minitab calculated is 8.006, which is the same as we calculated above. The Chi-square test for independence is an important method for determining. If there is a relationship between variables, where the chance that something falls into a particular category depends on whether the variable falls into another category comes into play. This relationship of independence / dependence is important to be able to understand and use.

**Problem Statement 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]

Answer:

**Step 1: Calculate all the means**

**= (51** + 45 + 33 + 45 + 67 + 23 +43+23+43+45+56+76+74+87+56)/15

=51.13

Mean = 51.13

**Step 2: Set up the Null and Alternate Hypothesis**

Calculate using H0: μ1 = μ2 = μ3 Ha: μ1 ≠ μ2 ≠ μ3

We also specify the as well as the rejection criteria. α = 0.05 Rejection criteria: K0.05 < F This means that if the critical value of F from tables is less than the calculated value of F, we reject the null hypothesis

**Step 3: Calculate the Sum of Squares**

The formula for sum of squares are given below

SS(total) = SS(between) + SS(within)

**Step 4: Calculate the Degrees of Freedom**

The degrees of freedom are calculated using the formula below.

***dftotal = n - 1 = 15 - 1 = 14***

***dfwithin = n - k = 15 - 3 = 12***

***dfbetween = k - 1 = 3 - 1 = 2***

Verify that dftotal = dfbetween + dfwithin

**Step 5: Calculate the Mean Squares**

The formula is given below for the mean squares between

***MS(Between) = SS(between) / df(between)***

***MS(Between) = 3022/2 = 1511.47***

**The formula for the mean squares within**

***MS(Within) = SS(Within) / Df(within)***

***MS(Within) = 1860/12 = 155.07***

**Step 6: Calculate the F Statistic**

Then we go ahead to calculate the F statistic using MSB and MSW

***F = MS(Between)/MS(Within)***

***F = 1511.47/155.07 = 9.75***

**Step 7: Look up F from table and state your Conclusion**

Access table from here From the table of F distribution, the critical value of F for 0.05 significance and degrees of freedom of(df1 = 12 and df2 = 2) we have:

F = 3.89

**Problem Statement 3:**

Calculate F Test for given 10, 20, 30, 40, 50 and 5,10,15, 20, 25.

For 10, 20, 30, 40, 50:

Answer:

#### F Test is generally defined as ratio of the variances of the given two set of values. First calculate standard deviation and variation of the given set of values. The formula used to calculate SD is,[¶](https://render.githubusercontent.com/view/ipynb?commit=0d13f208d788ac60cd8a81293373eef46c7b16c5&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f626b61727468696b6579616e313938382f53746174697374696373342d41737369676e6d656e74332f306431336632303864373838616336306364386138313239333337336565663436633762313663352f53746174697374696373342d41737369676e6d656e74332e6970796e62&nwo=bkarthikeyan1988%2FStatistics4-Assignment3&path=Statistics4-Assignment3.ipynb&repository_id=141235797&repository_type=Repository#F-Test-is-generally-defined-as-ratio-of-the-variances-of-the-given-two-set-of-values.-First-calculate-standard-deviation-and-variation-of-the-given-set-of-values.-The-formula-used-to-calculate-SD-is,)

Standard Deviation Formula 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 Test Example

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.