Statistics Assigment No. 4

Q.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. That data that resulted from the survey is summarized in the following table:

	High School	Bachelors	Masters	Phd	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

Q. Are gender & education level dependent at 5% level of significance? In other words, given the data collected above, is there a relationship b/w the gender of an individual & the level of education that they have obtained?

Note: To test the independence of 2 categorical features, we use CHI-Square Test of independence.

We can summarize 2 categorical variables within a 2-way table, also called a $r \times c$ contingency table, where

r = no. of rows

c = no. of columns

H_0 : The two categorical variables are independent

H_1 : The Two categorical variables are dependent.

CHI-Square Test Statistics

$$\boxed{\chi^2 = \sum \frac{(O-E)^2}{E}}$$

where $O \rightarrow$ Observed frequency

$E \Rightarrow$ Expected frequency under the null hypothesis

$$\boxed{E = \frac{\text{row total} \times \text{column total}}{\text{sample size}}} \longrightarrow Eq^n \text{ (1)}$$

here sample size is $\rightarrow 395$

<u>Note</u>: We will compare the value of test statistics to the critical value of χ^2_α with degree of freedom $= (r-1)(c-1)$, & reject the null hypothesis if $\chi^2 > \chi^2_\alpha$.

here, degree of freedom $= (2-1)(4-1)$
$\Rightarrow 3$

Here, is the table of expected counts !
(by using Eqn ①)

	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,

$$\chi^2 = \frac{(60-50.886)^2}{50.886} + \cdots + \frac{(57-48.132)^2}{48.132}$$

$$= 8.006$$

The critical value of χ^2 with 3 degree of freedom is 7.815.
Since $8.006 > 7.815$, we reject the null hypothesis & conclude that the education level depends on gender at a 5% level of significance.

Statistics Assignment No. - 4

Q.3 Calculate F Test for given 10, 20, 30, 40, 50 and 5, 10, 15, 20, 25.

For 10, 20, 30, 40, 50 :

F Test is generally defined as ratio of the variances of the given two set of values.

1st calculate standard deviation and variation of the given set of values.

Standard Deviation Formula (σ) : $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$

Variance : σ^2

F-Test FORMULA

$$F = \frac{\text{estimate of } \sigma^2 \text{ from means}}{\text{estimate of } \sigma^2 \text{ from individuals}}$$

$$F = \frac{\text{variance between Treatments}}{\text{Variance within Treatments}}$$

$$F = \frac{\text{Variance of Treatments}}{\text{Variance of Errors}}.$$

Calculate Variance of 1st set

~~Total~~ Inputs (N) = (10, 20, 30, 40, 50)

Total Inputs (N) = 5

$$\text{Mean } (X_m) = \frac{x_1 + x_2 + x_3 + x_4 - - - x_n}{N} = \frac{150}{5} = 30$$

$$\text{Standard deviation (SD)} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\Rightarrow \sqrt{\frac{1}{(5-1)} \left[(10-30)^2 + (20-30)^2 + (30-30)^2 + (40-30)^2 + (50-30)^2 \right]}$$

$$\Rightarrow \sqrt{\frac{1}{4} (400 + 100 + 0 + 100 + 400)} \Rightarrow \sqrt{250}$$

$$\Rightarrow 15.8114$$

$$\boxed{\text{Variance} = SD^2 = 250}$$

Calculate Variance of 2nd set

Inputs = (5, 10, 15, 20, 25)

Total Inputs (N) = 5

$$(X_m) \text{ Mean} = \frac{5 + 10 + 15 + 20 + 25}{5}$$

$$= \frac{75}{5} = 15$$

$$\text{Standard deviation (SD)} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(3)

$$SD = \sqrt{\frac{1}{5-1}\left[(5-15)^2 + (10-15)^2 + (15-15)^2 + (20-15)^2 + (25-15)^2\right]}$$

$$= \sqrt{\frac{1}{4}\left[100 + 25 + 0 + 25 + 100\right]} \Rightarrow \sqrt{62.5}$$

$$SD = 7.9057$$

$$\text{Variance} = SD^2 \Rightarrow 62.5$$

Calculate F-Test

$$F \cdot \text{Test} = \frac{\text{Variance of } (10, 20, 30, 40, 50)}{\text{Variance of } (5, 10, 15, 20, 25)}$$

$$\Rightarrow \frac{250}{62.5} \Rightarrow 4$$

$$\boxed{\text{The F Test value is } 4}$$

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]

SOLUTION:

	Group 1	Group 2	Group 3
	51	23	56
	45	43	76
	33	23	74
	45	43	87
	67	45	56
means	48.2	35.4	69.8

Calculate within group variances:

Group 1

	value	mean	deviations	sq deviations
	51	48.2	2.8	7.84
	45	48.2	-3.2	10.24
	33	48.2	-15.2	231.04
	45	48.2	-3.2	10.24
	67	48.2	18.8	353.44
SS				612.8

Group 2

	value	mean	deviations	sq deviations
	23	35.4	-12.4	153.76
	43	35.4	7.6	57.76
	23	35.4	-12.4	153.76
	43	35.4	7.6	57.76
	45	35.4	9.6	92.16
SS				515.2

Group 3

	value	mean	deviations	sq deviations
	56	69.8	-13.8	190.44
	76	69.8	6.2	38.44
	74	69.8	4.2	17.64
	87	69.8	17.2	295.84
	56	69.8	-13.8	190.44
SS				732.8

Mean Square (Error or within group)	155.07
Degrees of Freedom (Error or within group)	12
Sum of Squares (Within Group)	1860.8

	Group 1	Group 2	Group 3
	51	23	56
	45	43	76
	33	23	74
	45	43	87
	67	45	56
means	48.2	35.4	69.8

group means	grand mean	deviations	sq deviations
48.2	51.13	-2.93	8.60
35.4	51.13	-15.73	247.54
69.8	51.13	18.67	348.44

Sum of squares (means)	604.59
Mean Square (Between Groups)	1511.47
Degree of Freedom (Group)	2
SS (Group)	3022.93

Test statistic and critical value

F	9.75
Fcritical(2,12)	3.89

ANOVA Table

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	3022.93	2	1511.47	9.75	0.00306	3.89
Within Groups	1860.80	12	155.07			
Total	4883.73	14				

Decision Reject Ho since F > F-critical

Effect size

η2 0.62

APA writeup

F (2, 12) = 9.75, p < 0.05, η2 = 0.62