Problem Statement 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 Phd. 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:

Gender/Education	**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

Let use the **Chi-Square test of Independence** to test the independence of two categorical variables.
By summarizing two categorical variables within a two-way table (contingency table) where our main interests are could be two variables independent (either Null Hypothesis, Alternate Hypothesis)

$\chi 2=\sum(O-E)^2/E$ ⇒ **Where $\chi 2$ is Chi-square,**
O is Observed Frequency
and E is Expected Frequency.

E = (Total Row * Total Column)/Sample Size ⇒ **Where Size of Sample is given as 395**

<u>Each observation value:</u>

1. Female/High School value 60 ⇒ **(201*100)/395 = 50.886**
2. Male/High School value 40 ⇒ **(194*100)/395 = 49.113**
3. Female/Bachelors value 54 ⇒ **(201*98)/395 = 49.868**
4. Male/Bachelors value 44 ⇒ **(194*98)/395 = 48.131**
5. Female/Masters value 46 ⇒ **(201*99)/395 = 50.377**
6. Male/Masters value **53** ⇒ **(194*99)/395 = 48.622**
7. Female/Phd. value 41 ⇒ **(201*98)/395 = 49.868**
8. Male/Phd value 57 ⇒ **(194*98)/395 = 48.131**

Here's the table of expected counts:

Gender/Education	High School	Bachelors	Masters	Phd.	Total
Female	50.886	49.868	50.377	49.868	201
Male	49.113	48.131	48.622	48.131	194
Total	100	98	99	98	395

So if we do this,

$\chi 2$= (60-50.886)^2/50.886 + (40-49.113)^2/49.113 + (54-49.868)^2/49.868 + (44-48.131)^2/48.131 + (46-50.377)^2/50.377 + (53-48.622)^2/48.622 + (41-49.868)^2/49.868 + (57-48.131)^2/48.131

= 1.632 + 1.690 + 0.342 + 0.354 + 0.380 + 0.394 + 1.576 + 1.634 = 8.002

Degree of freedom = (Number of Columns -1)*(Number of Rows -1)

= (4-1)*(2-1) = 3*1 = 3

The critical value of $\chi 2$ with 3 Degree of Freedom is 7.815. Since 8.006 > 7.815. Therefore we reject the Null Hypothesis(Ho) and conclude that the education level depends on gender at a 5% level of significance.

Problem Statement 2:

Using the following data, perform a one way 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]

Group1 Mean : (51+45+33+45+67)/5 = 241/5 = 48.2

Group2 Mean: (23+43+23+43+45)/5 = 177/5 = 35.4

Group3 Mean: (56+76+74+87+56)/5 = 349/5 = 69.8

Sample means for the groups: = 48.2, 35.4, 69.8

Intermediate steps in calculating the group variances:

[1]

	value	mean	deviations	sq deviations
1	51	48.2	2.8	7.84
2	45	48.2	-3.2	10.24
3	33	48.2	-15.2	231.04
4	45	48.2	-3.2	10.24
5	67	48.2	18.8	353.44

[2]

	value	mean	deviations	sq deviations
1	23	35.4	-12.4	153.76
2	43	35.4	7.6	57.76
3	23	35.4	-12.4	153.76
4	43	35.4	7.6	57.76
5	45	35.4	9.6	92.16

[3]

	value	mean	deviations	sq deviations
1	56	69.8	-13.8	190.44
2	76	69.8	6.2	38.44

3	**74**	**69.8**	**4.2**	**17.64**
4	**87**	**69.8**	**17.2**	**295.84**
5	**56**	**69.8**	**-13.8**	**190.44**

The Sum of squared deviations from the mean (SS) for the groups are:

⇒ 612.8 515.2 732.8

Now we can calculate the Groups Mean respectively:

[[1]] Group1 Mean: 612.8 / 5-1 = 153.2

Group2 Mean: 515.2 / 5-1 = 128.8

Group3 Mean: 732.8 / 5-1 = 183.2

Mean Square-within group variance = (153.2+128.8+183.2)/3 = 155.07

Let calculate the remaining error terms for the ANOVA table:
df error = 15 - 3 = 12

SumofSquares-within error = Mean Square-within * df-within = **(155.07)*(15-3) = 1860.8**

Mean of Group Means = **(48.2+35.4+69.8)/3 = 153.4/3 = 51.13**

group mean	grand mean	deviations	sq deviations
48.2	**51.13**	**-2.93**	**8.58**
35.4	**51.13**	**-15.73**	**247.43**
69.8	**51.13**	**18.67**	**348.57**

Variance of Groups Means = (48.2-51.13)^2+(35.4-51.13)^2+(69.8-51.13)^2/3-1
= 604.58/2
= 302.29

Mean Square-between = 302.29*5 = 1511.45

SumofSquares-between = Mean Square-between* df-between = (1511.45)*(3-1) = 3022.9

Test statistic and critical value F**=1511.45 / 155.07 = 9.75**

$F_{critical}(2,12) = 3.89$

***We can reject H0 by decision.**

ANOVA table

source	SS	df	MS	F
group	3022.9	2	1511.45	9.75
error	1860.8	12	155.07	
total	4883.7			

Effect size

η^2**=3022.9/4883.7=0.62**

APA writeup ⇒ $F(2, 12)=9.75$, $p < 0.05$, $\eta^2=0.62$.

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 known as ratio of variance of set of values.

Set 1 : 10, 20, 30, 40, 50 And **Mean of Set 1** = (10+20+30+40+50)/5= 30

Variance of Set 1 =((10-30)^2+(20-30)^2+(30-30)^2+(40-30)^2+(50-30)^2)/5-1 = 250.25

Set 2 : 5,10,15, 20, 25 And **Mean of Set 2** = (5+10+15+20+25)/5= 15

Variance of Set 2 =((5-15)^2+(10-15)^2+(15-15)^2+(20-15)^2+(25-15)^2)/5-1 = 62.75

F Test for 10, 20, 30, 40, 50

⇒ **Variance of Set 1/ Variance of Set 2** = 250.25/62.75 = 3.988