Statistics4 –session 18----assignment

Problem 1

Null Hypothesis

The two categorical variables are independent

Alternative Hypothesis

The two categorical variables are dependent

Chi-Square Test Statistic

$\chi 2=\sum(O-E)2/E$

where *O* represents the observed frequency. *E* is the expected frequency under the null hypothesis and computed by:

$$E=\text{row total}\times\text{column total}/\text{sample size}$$

We will compare the value of the test statistic to the critical value of $\chi\alpha 2$ with degree of freedom = (*r* - 1) (*c* - 1), and reject the null hypothesis.

Here's 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, $\chi 2=(60-50.886)250.886+\cdots+(57-48.132)248.132=8.006$

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 and conclude that the education level depends on gender at a 5% level of significance.

Problem 2

Sample means ($\bar{x}\bar{x}$) 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
```

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

```
[1] 612.8 515.2 732.8
```

Var1=612.85−1=153.2Var1=612.85−1=153.2
Var2=515.25−1=128.8Var2=515.25−1=128.8
Var3=732.85−1=183.2Var3=732.85−1=183.2
MSerror=153.2+128.8+183.23=155.07MSerror=153.2+128.8+183.23=155.07 ***Note: this is just the average within-group variance; it is not sensitive to group mean differences!***
Calculating the remaining *error* (or *within*) terms for the ANOVA table:

dferror=15−3=12dferror=15−3=12
SSerror=(155.07)(15−3)=1860.8SSerror=(155.07)(15−3)=1860.8
Intermediate steps in calculating the variance of the sample means:

Grand mean ($\bar{x}_{grand}$$\bar{x}_{grand}$) = 48.2+35.4+69.83=51.1348.2+35.4+69.83=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**

Sum of squares (SSmeans)=604.58(SSmeans)=604.58
Varmeans=604.583−1=302.29Varmeans=604.583−1=302.29
MSbetween=(302.29)(5)=1511.45MSbetween=(302.29)(5)=1511.45 ***Note: This method of estimating the variance IS sensitive to group mean differences!***
Calculating the remaining *between* (or *group*) terms of the ANOVA table:

dfgroups=3−1=2dfgroups=3−1=2
SSgroup=(1511.45)(3−1)=3022.9SSgroup=(1511.45)(3−1)=3022.9
Test statistic and critical value

F=1511.45155.07=9.75F=1511.45155.07=9.75
Fcritical(2,12)=3.89Fcritical(2,12)=3.89
Decision: reject H0 Decision: reject H0
ANOVA table

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

Effect size

η2=3022.94883.7=0.62η2=3022.94883.7=0.62
APA writeup

***F*(2, 12)=9.75, *p* <0.05, η2η2=0.62.**

Problem 3

Calculate Variance of first set

Total Inputs (N) =(10,20,30,40,50)
Total Inputs (N)=5
Mean (xm)= $(x_1+x_1+x_2...x_n)/N$
Mean (xm)= 150/5
Means(xm)= 30
SD=sqrt(1/(N-1)*($(x_1-xm)^2+(x_2-x_m)^2+..+(x_n-x_m)^2$))
=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=SD^2
Variance=15.8114^2
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)*((x_1-x_m)^2+(x_2-x_m)^2+..+(x_n-x_m)^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=SD^2
Variance=7.9057^2
Variance=62.5

<u>To calculate F Test</u>
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.