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?

Solution:

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 χ^2_α

with degree of freedom = (r - 1) (c - 1), and reject the null hypothesis if $\chi^2 > \chi^2_\alpha$.

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:

Here is a Table of Observed Frequency (O)

	High Scool	Bachelors	Masters	Ph.D	Total
Female	60	54	46	41	**201**
Male	40	44	53	57	**194**
Total	**100**	**98**	**99**	**98**	**395**

Here's the table of expected counts (E):

	High Scool	Bachelors	Masters	Ph.D	Total
Female	50.9	49.9	50.4	49.9	**201**
Male	49.1	48.1	48.6	48.1	**194**
Total	**100**	**98**	**99**	**98**	**395**

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

	High Scool	Bachelors	Masters	Ph.D	Total
Female	1.6	0.3	0.4	1.6	**201**
Male	1.7	0.4	0.4	1.6	**194**
Total	**100**	**98**	**99**	**98**	**395**

$\chi 2$ = 1.6+0.3+0.4+1.6+1.7+0.4+0.4+1.6 = 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 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
Mean	48.2	35.4	69.8

Intermediate steps in calculating the group variances:

Group-1

	Value	Value-Mean	Square (Value-Mean)
1	51	2.8	7.84
2	45	-3.2	10.24
3	33	-15.2	231.04
4	45	-3.2	10.24
5	67	18.8	353.44
Mean	**48.2**		612.8

Gropup-2

	Value	Value-Mean	Square (Value-Mean)
1	23	-12.4	153.76
2	43	7.6	57.76
3	23	-12.4	153.76
4	43	7.6	57.76
5	45	9.6	92.16
Mean	**35.4**		515.2

Group-3

	Value	Value-Mean	Square (Value-Mean)
1	56	-13.8	190.44
2	76	6.2	38.44
3	74	4.2	17.64
4	87	17.2	295.84
5	56	-13.8	190.44
Mean	**69.8**		732.8

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

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

$$Var_1 = \frac{612.8}{5-1} = 153.2$$

$$Var_2 = \frac{515.2}{5-1} = 128.8$$

$$Var_3 = \frac{732.8}{5-1} = 183.2$$

$$MS_{error} = \frac{153.2 + 128.8 + 183.2}{3} = 155.07$$

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:

$$df_{error} = 15 - 3 = 12$$

$$SS_{error} = (155.07)(15 - 3) = 1860.8$$

Intermediate steps in calculating the variance of the sample means:

Grand mean ($\bar{x}_{grand}$) = $48.2+35.4+69.83 = 51.13$

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 (SS_{means})$=604.58$

$Var_{means} = 604.583-1 = 302.29$

$MS_{between} = (302.29)(5) = 1511.45$

Note: This method of estimating the variance IS sensitive to group mean differences!

Calculating the remaining *between* (or *group*) terms of the ANOVA table:

$df_{groups} = 3-1 = 2$

$SS_{group} = (1511.45)(3-1) = 3022.9$

Test statistic and critical value

$F = 1511.45155.07 = 9.75$

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

Decision: reject H0

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

$\eta^2 = 3022.94883.7 = 0.62$

APA writeup

$F(2, 12) = 9.75$, $p < 0.05$, η^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:

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.