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:

The chi squared test can also be used to test independence.

Hypothesis
Null Hypothesis: Both gender and level of education is independent
Alternate Hypothesis: Both gender and level of education is dependent

Observed:

	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

Expected:

	High School	Bachelors	Masters	Ph.d	Total
Female	50.88	49.868	50.377	49.868	201
Male	49.114	48.132	48.623	48.132	194
Total	100	98	99	98	395

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Col1 = (60 – 50.88)^2/50.88= 1.634
Col2 = (54-49.868)^2/49.868= 0.342
Col3 := (46-50.377)^2/50.377= 0.380
col4 := (41-49.868)^2/49.868= 1.576
C21 : = (60 – 49.114)^2/49.114= 2.412
C22 : = (54-48.132)^2/48.132= 0.715
C23 := (46-48.623)^2/48.623= 0.141
C24: = (41-48.132)^2/48.132 = 1.056
χ2 = 8.256

Degrees of Freedom: (4-1)*(2-1) = 3

At 5% significant level = 7.815

Conclusion:

Our calculated Chi squared value greater than the critical value and hence we reject the Null Hypothesis.

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]

	GROUP 1	GROUP 2	GROUP 3	(GROUP1-GROUP1 MEAN)^2	(GROUP2 -GROUP2 MEAN)^2	(GROUP3 -GROUP3 MEAN)^2
	51	23	56	7.84	153.76	190.44
	45	43	76	10.24	57.76	38.44
	33	23	74	231.04	153.76	17.64
	45	43	87	10.24	57.76	295.84
	67	45	56	353.44	92.16	190.44
TOTAL	241	177	349	612.8	515.2	732.8
MEAN	48.2	35.4	69.8			
GRAND MEAN	51.33	51.33	51.33			
NO OF SAMPLES	5	5	5			
k	3					

N	15					
Sum of Squares between	48.9845	1268.8245	1705.705	3023.5135		
Mean between =(sum of squares between)/(k-1)	1511.75675					
Sum of squares within	1860.8					
Mean within = (sum of squares between)/(N-K)	155.0666667					
F data = MSS between / MSS within	9.749076204					
F Test =						
Degree of Freedom	Sample -1 = 2					
$F_{critical}(2,12)=3.89$						

Conclusion: Reject Null hypothesis

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:

Solution:

Standard Deviation:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Variance for 1st set : (10,20,30,40,50)

Mean = (10+20+30+40+50)/5 = 30

Standard Deviation : 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)
=15.811
Variance = SD^2
Variance = 249.98

Variance for 2nd set : (5,10,15,20,25)

Mean = (5+10+15+20+25)/5 = 15

Standard Deviation : 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)
=7.90
Variance = SD^2
Variance = 62.5

F Test : (Variance for 1st set) / (Variance for 2nd set)

F Test : 250/62.5
F Test = 4