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:

Here's the table of expected counts as

$$\mathrm{E} = \frac{\text{Row total} \times \text{Column total}}{\text{Sample Size}}$$

	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, $\chi^2 = \frac{(60-50.886)^2}{50.886} + \dots + \frac{(57-48.132)^2}{48.132} = 8.006$

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

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:

Sample means ($\bar{x}$) 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

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

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

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

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

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:

$\boldsymbol{df_{\text{error}}} = \mathbf{15-3 = 12}$

$\boldsymbol{SS_{\text{error}}} = \mathbf{(155.07)(15-3) = 1860.8}$

Intermediate steps in calculating the variance of the sample means:

Grand mean $(\bar{x}_{\text{grand}}) = \frac{48.2+35.4+69.8}{3} = \mathbf{51.13}$

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 $(\boldsymbol{SS}_{\text{means}})$ = **604.58**

$\boldsymbol{Var}_{\text{means}} = \frac{\mathbf{604.58}}{\mathbf{3-1}} = \mathbf{302.29}$

$\boldsymbol{MS}_{\text{between}} = \mathbf{(302.29)(5)=1511.45}$

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

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

$\boldsymbol{Df}_{\text{groups}} = \mathbf{3-1 = 2}$

$\boldsymbol{SS}_{\text{groups}} = \mathbf{(1511.45)(3-1)=3022.9}$

Test statistic and critical value

$\boldsymbol{F} = \frac{\mathbf{1511.4}}{\mathbf{5155.07}} = \mathbf{9.75}$

$\boldsymbol{F}_{\text{Critical}}(2,12) = 3.89$

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

$\eta^2 = \frac{\mathbf{3022.9}}{\mathbf{4883.7}} = 0.62$

APA writeup

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

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 = $sqrt(1/(N-1)*((x_1-x_m)^2+(x_2-x_m)^2+..+(x_n- x_m)^2))$
   $= sqrt(1/(5-1)((10-30)^2+(20-30)^2+(30-30)^2+(40-30)^2+(50-30)^2))$
   $= sqrt(1/4((-20)^2+(-10)^2+(0)^2+(10)^2+(20)^2))$
   = 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

   Total Inputs(N) = (5,10,15,20,25)
   Total Inputs(N) = 5
   Mean (xm) = $(x_1+x_1+x_2... x_n)/N$
   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

   To calculate F Test:
   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.