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:

Gender	High School	Bachelors	Masters	Ph.d.	Tot
Female	60	54	46	41	2
Male	40	44	53	57	1
Total	100	98	99	98	3

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?

Create table to expected count following formulaes to be used

E= (Row Total×Column Total) / Sample Size

New Table would look like this

Gender	High School	Bachelors	Masters	Ph.d.	Tot
Female	50.886	49.868	50.377	49.868	2
Male	49.114	48.132	48.623	48.132	1
Total	100	98	99	98	3

So $\chi 2 = (60-50.886)2 / 50.886 + ... + (57-48.132)2 /48.132 = 8.006$

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 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]

Let's there are 3 Groups G1, G2 and G3

Mean of the all three groups are 48.2, 35.4, 69.8 respectively

In [1]:
```
''' 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
'''
```

Out[1]:
```
' Intermediate steps in calculating the group variances:\n\n [[1]]\n   value mean deviatio
ns sq deviations\n 1    51 48.2        2.8          7.84\n 2    45 48.2       -3.2
10.24\n 3    33 48.2      -15.2        231.04\n 4    45 48.2       -3.2         10.24\n 5
   67 48.2       18.8        353.44\n\n[[2]]\n  value mean deviations sq deviations\n1
23 35.4      -12.4        153.76\n2    43 35.4        7.6         57.76\n3    23 35.4
-12.4        153.76\n4    43 35.4        7.6         57.76\n5    45 35.4        9.6
  92.16\n\n[[3]]\n  value mean deviations sq deviations\n1    56 69.8      -13.8        19
0.44\n2    76 69.8        6.2         38.44\n3    74 69.8        4.2         17.64\n4    8
7 69.8       17.2        295.84\n5    56 69.8      -13.8        190.44\n'
```

Sum of squared deviations from the mean (SS) for the groups G1, G2 and G3:

612.8

515.2

732.8

In [2]:
```
Var1 = 612.8/(5-1)

Var2 = 515.2/(5-1)

Var3 = 732.8/(5-1)
print(Var1)
print(Var2)
print(Var3)
```

```
153.2
128.8
183.2
```

In [3]:
```
MS_error = (Var1 + Var2 + Var3) /3
MS_error
```

Out[3]:
```
155.06666666666666
```

Calculating the remaining error (or within) terms for the ANOVA table:

In [4]:
```
df_error = 15-3
df_error
```

Out[4]:
```
12
```

In [5]:
```
SS_error1=(MS_error)*(15-3)
SS_error1
```

Out[5]:
```
1860.8
```

In [6]:
```
## Intermediate steps in calculating the variance of the sample means:

x_mean = (48.2 + 35.4 + 69.8) /3
x_mean
```

Out[6]:
```
51.133333333333326
```

In [7]:
```
'''
   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

'''
```

Out[7]:
```
'\n   group mean       grand mean     deviations      sq deviations\n       48.2          5
1.13          -2.93           8.58\n       35.4          51.13          -15.73          24
7.43\n       69.8          51.13           18.67          348.57\n\n'
```

In [8]:
```
#Sum of squares
SS_means=604.58
Var_means = SS_means/(3-1)
Var_means
```

Out[8]:
```
302.29
```

In [9]:
```
MS_between=Var_means*5
MS_between
```

Out[9]:
```
1511.45
```

In [10]:
```
df_groups=3-1
SS_group=MS_between*df_groups
SS_group
```

Out[10]:
```
3022.9
```

In [11]:
```
## Test statistic and critical value
F = MS_between /MS_error
F
```

Out[11]:
```
9.747098022355976
```

In [12]:
```
## F_critical(2,12)=3.89
#  Decision: reject H0
```

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:

$$F = \frac{\text{estimate of } \sigma^2 \text{ from means}}{\text{estimate of } \sigma^2 \text{ from individuals}}$$

$$F = \frac{\text{Variance between Treatments}}{\text{Variance within Treatments}}$$

$$F = \frac{\text{Variance of Treatments}}{\text{Variance of Error}}$$

In [13]:
```
# Calculate Variance of first set
list1 = [10, 20, 30, 40, 50]
#list1
import numpy as np
np.mean(list1)
import statistics as stats

x1 = stats.mean(list1)
print("Mean ",x1)
sd1 = stats.stdev(list1)
print("Standard Deviation -->", sd1)
varaiance1 = sd1**2
print("Varaince -->",varaiance1)
```

```
Mean  30
Standard Deviation --> 15.811388300841896
Varaince --> 250.0
```

In [14]:
```
## Calculate Variance of second set
list2 = [5,10,15, 20, 25]
#list1
import numpy as np
np.mean(list2)
import statistics as stats

x2 = stats.mean(list2)
print("Mean ",x2)
sd2 = stats.stdev(list2)
print("Standard Deviation -->", sd2)
varaiance2 = sd2**2
print("Varaince -->",varaiance2)
```

```
Mean  15
Standard Deviation --> 7.905694150420948
Varaince --> 62.5
```

In [15]:
```
# To calculate F Test
# F Test = (variance of 10, 20,30,40,50) / (variance of 5, 10, 15, 20, 25)
F_Test = varaiance1/varaiance2
F_Test
```

Out[15]:
```
4.0
```