In [1]:
```
import numpy as np
import scipy.stats as stats
```

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?

In [2]:
```
df = [[60,54,46,41],[40,44,53,57]]
pvalue = stats.chi2_contingency(df)
α = 0.05
print(f"Decision Rule at α = 0.05 is: 3.841")
```

```
Decision Rule at α = 0.05 is: 3.841
And pvalue is :8.006066246262538 which is greater than α :3.841.
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]

In [3]:
```
gp1=[51, 45, 33, 45, 67]
gp2=[23, 43, 23, 43, 45]
gp3=[56, 76, 74, 87, 56]
df =pd.DataFrame([[51, 45, 33, 45, 67],[23, 43, 23, 43, 45],[56, 76, 74
def degreeOfFreedom(data_f):
    try:
        global df_between,df_within, df_total
        df_between =len(data_f.index)-1
        df_within = (len(data_f.iloc[0]) + len(data_f.iloc[1]))-len(dat
        df_total = df_between + df_within
    except Exception as e:
```

```
        print('Error in Function degreeOfFreedom',e)
degreeOfFreedom(df)
st , pv = stats.f_oneway(gp1,gp2,gp3)
print(f"Decision Rule {df_within,df_between} at α = 0.05 is: 3.25744")
print(f"And pvalue is :{pv} which is less than α :{3.25744} \nWe reject
```

```
Decision Rule (7, 2) at α = 0.05 is: 3.25744
And pvalue is :0.0030597541434430556 which is less than α :3.25744.
We reject the 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:

In [4]:

```
F_testresult = stats.f_oneway([10, 20, 30, 40, 50],[5,10,15, 20, 25])
gp1_var = np.var([10, 20, 30, 40, 50])
gp2_var = np.var([5,10,15, 20, 25])
F_value = gp1_var/gp2_var
print(f"F test Result:\nStatistics:{F_testresult[0]}\nP value:{F_testre
```

```
F test Result:
Statistics:3.6
P value:0.0943497728424377

F value for given 10, 20, 30, 40, 50 and 5,10,15, 20, 25 =>
4.0
```

In [ ]: