Statitics_4_Assignment

March 18, 2019

In [2]:
```
"""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 gende
of an individual and the level of education that they have obtained?"""

# We can perform a chi square test to find the dependance between Gender and Education

import numpy as np
from scipy.stats import chi2_contingency

Female = [60,54,46,41]
Male = [40,44,53,57]
table=np.array([Male,Female])
chi2_stat,p_val,dof,ex = chi2_contingency(table)

print("===Chi2 Stat===")
print(chi2_stat)
print("\n")
print("===Degrees of Freedom===")
print(dof)
print("\n")
print("===P-Value===")
print(p_val)
print("\n")
print("===Contingency Table===")
print(ex)

if p_val<=0.05:
    print("There is dependency between Gender and Education")  #Reject Null Hypothesis
else:
```

```
            print("Education is Gender Independent")  #Fail to reject Null Hypothesis
```

```
===Chi2 Stat===
8.006066246262538


===Degrees of Freedom===
3


===P-Value===
0.045886500891747214


===Contingency Table===
[[49.11392405 48.13164557 48.62278481 48.13164557]
 [50.88607595 49.86835443 50.37721519 49.86835443]]
There is dependency between Gender and Education
```

In [3]:
```
"""Problem Statement 2:
Using the following data, perform a oneway analysis of variance using =.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]"""

import numpy as np
import pandas as pd
import scipy.stats as stats

groups = pd.DataFrame({"Group1":[51, 45, 33, 45, 67],"Group2":[23, 43, 23, 43, 45],"Gr

statistic,p_val = stats.f_oneway(groups["Group1"],groups["Group2"],groups["Group3"])

print("------")
print("Statistic=", statistic)
print("-------")
print("pvalue=", p_val)

if p_val< 0.05:
    print("There is some difference across the means of groups")
else:
    print("There is no big difference across the means of groups")
```

```
------
Statistic= 9.747205503009463
```

```
-------
pvalue= 0.0030597541434430556
There is some difference across the means of groups
```

In [4]:
```
"""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:"""

a = [10,20,30,40,50]
b = [5,10,15,20,25]
F = np.var(a)/np.var(b)

print("The F test result is :",F)
```

```
The F test result is : 4.0
```