# Assignment - 3 - Q1 to Q4(Hypothesis Testing)

In [1]:

```
import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import norm
```

In [3]:

```
# Load the dataset
data=pd.read_csv('Downloads\\Cutlets.csv')
data.head()
```

Out[3]:

| | Unit A | Unit B |
|---|---|---|
| **0** | 6.8090 | 6.7703 |
| **1** | 6.4376 | 7.5093 |
| **2** | 6.9157 | 6.7300 |
| **3** | 7.3012 | 6.7878 |
| **4** | 7.4488 | 7.1522 |

In [4]:

```
unitA=pd.Series(data.iloc[:,0])
unitA
```

Out[4]:

```
0     6.8090
1     6.4376
2     6.9157
3     7.3012
4     7.4488
5     7.3871
6     6.8755
7     7.0621
8     6.6840
9     6.8236
10    7.3930
11    7.5169
12    6.9246
13    6.9256
14    6.5797
15    6.8394
16    6.5970
17    7.2705
18    7.2828
19    7.3495
20    6.9438
21    7.1560
22    6.5341
23    7.2854
24    6.9952
25    6.8568
26    7.2163
27    6.6801
28    6.9431
29    7.0852
30    6.7794
31    7.2783
32    7.1561
33    7.3943
34    6.9405
Name: Unit A, dtype: float64
```

In [5]:

```python
unitB=pd.Series(data.iloc[:,1])
unitB
```

Out[5]:

```
0     6.7703
1     7.5093
2     6.7300
3     6.7878
4     7.1522
5     6.8110
6     7.2212
7     6.6606
8     7.2402
9     7.0503
10    6.8810
11    7.4059
12    6.7652
13    6.0380
14    7.1581
15    7.0240
16    6.6672
17    7.4314
18    7.3070
19    6.7478
20    6.8889
21    7.4220
22    6.5217
23    7.1688
24    6.7594
25    6.9399
26    7.0133
27    6.9182
28    6.3346
29    7.5459
30    7.0992
31    7.1180
32    6.6965
33    6.5780
34    7.3875
Name: Unit B, dtype: float64
```

In [6]:

```python
# 2-sample 2-tail ttest:  stats.ttest_ind(array1,array2)   #ind -> independent sapmles
p_value=stats.ttest_ind(unitA,unitB)
p_value
```

Out[6]:

```
Ttest_indResult(statistic=0.7228688704678063, pvalue=0.4722394724599501)
```

In [7]:

```python
p_value[1]   #  2-tail probability
```

Out[7]:

```
0.4722394724599501
```

In [8]:

```python
# compare p_value with α = 0.05  (At 5% significance level)
```

# Q2 (Hypothesis Testing)

In [9]:

```python
import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import norm
```

In [10]:

```python
# Load the dataset
data=pd.read_csv('Downloads\\LabTAT.csv')
data.head()
```

Out[10]:

| | Laboratory 1 | Laboratory 2 | Laboratory 3 | Laboratory 4 |
|---|---|---|---|---|
| **0** | 185.35 | 165.53 | 176.70 | 166.13 |
| **1** | 170.49 | 185.91 | 198.45 | 160.79 |
| **2** | 192.77 | 194.92 | 201.23 | 185.18 |
| **3** | 177.33 | 183.00 | 199.61 | 176.42 |
| **4** | 193.41 | 169.57 | 204.63 | 152.60 |

In [12]:

```
# Anova ftest statistisc: stats.f_oneway(column-1,column-2,column-3,column-4)
p_value=stats.f_oneway(data.iloc[:,0],data.iloc[:,1],data.iloc[:,2],data.iloc[:,3])
p_value
```

Out[12]:

```
F_onewayResult(statistic=118.70421654401437, pvalue=2.1156708949992414e-57)
```

In [13]:

```
p_value[1]     #compare it with α = 0.05
```

Out[13]:

```
2.1156708949992414e-57
```

## Q3- (Hypothesis Testing)

In [14]:

```
import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import norm
```

In [17]:

```
df = pd.read_csv('Downloads\\BuyerRatio.csv')
df.head()
```

Out[17]:

| | Observed Values | East | West | North | South |
|---|---|---|---|---|---|
| **0** | Males | 50 | 142 | 131 | 70 |
| **1** | Females | 435 | 1523 | 1356 | 750 |

In [26]:

```
df_table=df.iloc[:,1:6]
df_table
```

Out[26]:

| | East | West | North | South |
|---|---|---|---|---|
| **0** | 50 | 142 | 131 | 70 |
| **1** | 435 | 1523 | 1356 | 750 |

In [27]:

```
df_table.values
```

Out[27]:

```
array([[  50,  142,  131,   70],
       [ 435, 1523, 1356,  750]], dtype=int64)
```

In [20]:

```
val=stats.chi2_contingency(df_tables)
```

In [21]:

```
val
```

Out[21]:

```
(1.595945538661058,
 0.6603094907091882,
 3,
 array([[  42.76531299,  146.81287862,  131.11756787,   72.30424052],
        [ 442.23468701, 1518.18712138, 1355.88243213,  747.69575948]]))
```

In [22]:

```
type(val)
```

Out[22]:

```
tuple
```

In [28]:

```
no_of_rows=len(df_table.iloc[0:2,0])
no_of_columns=len(df_table.iloc[0,0:4])
degree_of_f=(no_of_rows-1)*(no_of_columns-1)
print('Degree of Freedom=',degree_of_f)
```

```
Degree of Freedom= 3
```

In [29]:

```
Expected_value=val[3]
```

In [30]:

```
Expected_value
```

Out[30]:

```
array([[  42.76531299,  146.81287862,  131.11756787,   72.30424052],
       [ 442.23468701, 1518.18712138, 1355.88243213,  747.69575948]])
```

In [31]:

```
from scipy.stats import chi2
chi_square=sum([(o-e)**2/e for o,e in zip(df_table.values,Expected_value)])
chi_square_statestic=chi_square[0]+chi_square[1]
chi_square_statestic
```

Out[31]:

```
1.5152956451130446
```

In [32]:
```
critical_value=chi2.ppf(0.95,3)
critical_value
```

Out[32]:
```
7.814727903251179
```

In [33]:
```
if chi_square_statestic >= critical_value:
      print('Dependent (reject H0)')
else:
      print('Independent  (fail to reject H0)')
```

```
Independent  (fail to reject H0)
```

In [34]:
```
pvalue=1-chi2.cdf(chi_square_statestic,3)
pvalue
```

Out[34]:
```
0.6787446296467897
```

In [35]:
```
if pvalue  <=  0.05:
       print('Dependent  (reject H0)')
else:
    print('Independent    (fail to  reject H0)')
```

```
Independent    (fail to  reject H0)
```

# Q4 - (Hypothesis Testing)

In [36]:
```
import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import norm
from scipy.stats import chi2_contingency
```

In [37]:
```
# Load the dataset
data=pd.read_csv('Downloads\\Costomer+OrderForm.csv')
data
```

Out[37]:

| | Phillippines | Indonesia | Malta | India |
|---|---|---|---|---|
| **0** | Error Free | Error Free | Defective | Error Free |
| **1** | Error Free | Error Free | Error Free | Defective |
| **2** | Error Free | Defective | Defective | Error Free |
| **3** | Error Free | Error Free | Error Free | Error Free |
| **4** | Error Free | Error Free | Defective | Error Free |
| **...** | ... | ... | ... | ... |
| **295** | Error Free | Error Free | Error Free | Error Free |
| **296** | Error Free | Error Free | Error Free | Error Free |
| **297** | Error Free | Error Free | Defective | Error Free |
| **298** | Error Free | Error Free | Error Free | Error Free |
| **299** | Error Free | Defective | Defective | Error Free |

300 rows × 4 columns

In [38]:
```
data.Phillippines.value_counts()
```

Out[38]:
```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
```

In [39]:
```
data.Indonesia.value_counts()
```

Out[39]:
```
Error Free    267
Defective      33
Name: Indonesia, dtype: int64
```

In [40]:
```
data.Malta.value_counts()
```

Out[40]:
```
Error Free    269
Defective      31
Name: Malta, dtype: int64
```

In [41]:

```
data.India.value_counts()
```

Out[41]:

```
Error Free    280
Defective      20
Name: India, dtype: int64
```

In [42]:

```
# make a contingency table
obs=np.array([[271,267,269,280],[29,33,31,20]])
obs
```

Out[42]:

```
array([[271, 267, 269, 280],
       [ 29,  33,  31,  20]])
```

In [43]:

```
# chi2 contengency independent test
chi2_contingency(obs)   # o / p is (chi2 stats value, p_value, df, expected obsvations)

```

Out[43]:

```
(3.858960685820355,
 0.2771020991233135,
 3,
 array([[271.75, 271.75, 271.75, 271.75],
        [ 28.25,  28.25,  28.25,  28.25]]))
```

In [44]:

```
# compare p_value with α = 0.05
```

In [ ]:

```

```