In [15]:
```
#import libraries
import pandas as pd
import numpy as np
import scipy
from scipy import stats
```

In [ ]:
```
## Q1. A F&B manager wants to determine whether there is any significant difference in th
##the cutlet between two units. A randomly selected sample of cutlets was collected from
##and measured? Analyze the data and draw inferences at 5% significance level. Please sta
##assumptions and tests that you carried out to check validity of the assumptions.
## Minitab File : Cutlets.mtw
```

In [16]:
```
df = pd.read_csv(r"cutlets.csv")
df.head()
```

Out[16]:

	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 [ ]:
```
##as we have two check for diametre of tw units have to perform 2 sample test.
Assumption of null and alternate Hypothesis
#Ho= Mean of diameters of Unit A is equal to Mean of diameters of unit B
#Ha= Mean of diameters of Unit A is not equal to Mean of diameters of unit B
```

In [4]:
```
df.mean()
```

Out[4]:
```
Unit A    7.019091
Unit B    6.964297
dtype: float64
```

In [5]:
```
df.var()
```

Out[5]:
```
Unit A    0.083179
Unit B    0.117924
dtype: float64
```

In [15]:
```
p_value = stats.ttest_ind(df['Unit A'],df['Unit B'])
```

In [14]:
```
p_value
```

Out[14]:
```
0.4722394724599501
```

In [ ]:
```
#The significance level is 5% so, α = 0.05
#and p value is greater than 0.05
#we accept Alternate hypothesis
#i.e we can't see a significant difference in diametre of the two units A and B.
```

In [ ]:
```

```

In [ ]:

In [ ]:

In [ ]:
```
#Q2. A hospital wants to determine whether there is any difference in the average Turn Ar
#of reports of the laboratories on their preferred list. They collected a random sample a
#TAT for reports of 4 laboratories. TAT is defined as sample collected to report dispatch
# Analyze the data and determine whether there is any difference in average TAT among the
#laboratories at 5% significance level.


    #Minitab File: LabTAT.mtw
```

In [17]:
```
#import the libraries
import pandas as pd
import scipy
import numpy as np
from scipy import stats
```

In [12]:
```
df1 = pd.read_csv(r"LabTAT.csv")
df1.head()
```

Out[12]:

	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 [6]:
```
# As we are having more than 2 samples we have to perform ANNOVA one way test as we have
```

In [20]:
```
#Ho= Mean TAT for all the samples is same.
#Ha= mean TAT for all the samples is not same.
import scipy
```

In [4]:
```
df1.mean()
```

Out[4]:
```
Laboratory 1    178.361583
Laboratory 2    178.902917
Laboratory 3    199.913250
Laboratory 4    163.682750
dtype: float64
```

In [22]:
```
import scipy.stats as stats
p_value=stats.f_oneway(df1.iloc[:,0],df1.iloc[:,1],df1.iloc[:,2],df1.iloc[:,3])
```

In [23]:
```
p_value
```

Out[23]:
```
F_onewayResult(statistic=118.70421654401437, pvalue=2.1156708949992414e-57)
```

In [ ]:
```
# here as p_value is less than 0.05 we accept NULL hypothesis
#i.e there is NO difference in average TAT among the different laboratories at 5% signifi
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:
```
#Q3.Sales of products in four different regions is tabulated for males and females.
#Find if male-female buyer rations are similar across regions.
```

In [18]:
```
import pandas as pd
import scipy
import seaborn as sns
import scipy.stats as stats

import warnings
warnings.filterwarnings("ignore")
```

In [10]:
```
BuyerRatio=pd.read_csv(r"BuyerRatio.csv")
BuyerRatio.head()
```

Out[10]:

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

In [11]:
```
BuyerRatio.columns
```

Out[11]:
```
Index(['Observed Values', 'East', 'West', 'North', 'South'], dtype='object')
```

In [13]:
```
df3 = BuyerRatio.drop(['Observed Values'], axis=1)
```

In [14]:
```
df3.values
```

Out[14]:
```
array([[  50,  142,  131,   70],
       [ 435, 1523, 1356,  750]], dtype=int64)
```

In [21]:
```
chisq_result= scipy.stats.chi2_contingency(df3)
chisq_result
```

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]:
```
print(["T statistic","P_value"],[chisq_result[0],chisq_result[1]])
```

```
['T statistic', 'P_value'] [1.595945538661058, 0.6603094907091882]
```

In [ ]:
```
# Since we got P > 0.05 so we can say that the Null Hypothesis is True (Ho).
#In all the region Praportion of male and female is same.
```

In [ ]:

In [ ]:

In [ ]:

In [ ]:
```
#     TeleCall uses 4 centers around the globe to process customer order forms. They audi
#%  of the customer order forms. Any error in order form renders it defective and has to
#before processing.  The manager wants to check whether the defective %  varies by centre
#Please analyze the data at 5% significance level and help the manager draw appropriate i

#Minitab File: CustomerOrderForm.mtw
```

In [2]:
```
import pandas as pd
import scipy
import seaborn as sns
import scipy.stats as stats

import warnings
warnings.filterwarnings("ignore")
```

In [3]:
```
customer=pd.read_csv("Costomer+OrderForm.csv")
customer.head()
```

Out[3]:

	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

In [6]:
```
customer.Phillippines.value_counts()
```

Out[6]:
```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
```

In [8]:
```
customer.Indonesia.value_counts()
```

Out[8]:
```
Error Free    267
Defective      33
Name: Indonesia, dtype: int64
```

In [9]:
```
customer.Malta.value_counts()
```

Out[9]:
```
Error Free    269
Defective      31
Name: Malta, dtype: int64
```

In [10]:
```
customer.India.value_counts()
```

Out[10]:
```
Error Free    280
Defective      20
Name: India, dtype: int64
```

In [19]:
```
custmer=pd.DataFrame(index=['Error free','Defective'],
                     data={'Phillippines':[271,29],'Indonesia':[267,33],'Malta':[269,31],'Ind
```

In [20]:
```
custmer
```

Out[20]:

	Phillippines	Indonesia	Malta	India
Error free	271	267	269	280
Defective	29	33	31	20

In [ ]:
```
# As we got data in descrete form we have to perform chi square test.
```

In [ ]:
```
#Ho= % of all defective centre is same
#HA= % of all defective centre is not same
```

In [21]:
```
chisq_result= scipy.stats.chi2_contingency(custmer)
chisq_result
```

Out[21]:
```
(3.858960685820355,
 0.2771020991233135,
 3,
 array([[271.75, 271.75, 271.75, 271.75],
        [ 28.25,  28.25,  28.25,  28.25]]))
```

In [22]:
```
print(["chisq_value","P_value"],[chisq_result[0],chisq_result[1]])
```

```
['chisq_value', 'P_value'] [3.858960685820355, 0.2771020991233135]
```

In [ ]:
```
#as we can see p value is greater than 0.05 so u consider Alternate hypothesis
#i.e Percentage defective Across Centres is same..
```