A F&B manager wants to determine whether there is any significant difference in the diameter of the cutlet between two units. A randomly selected sample of cutlets was collected from both units and measured? Analyze the data and draw inferences at 5% significance level. Please state the assumptions and tests that you carried out to check validity of the assumptions.

#2-Sampled T-test

P = 0.05

H0 : There is no significant difference in the diameter of the cutlets between two units.

Ha : There is a significant difference in the diameter of the cutlets between two units.

In [1]:

```
import pandas as pd
from scipy import stats
```

In [2]:

```
cutlets_data = pd.read_csv('C:/Users/Ravi Kiran/Hypothesis Testing/Cutlets.csv')
cutlets_data.head()
```

Out[2]:

	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 [3]:

```
cutlets_data['Unit A'].mean()
```

Out[3]:

```
7.01909142857143
```

In [4]:

```
1 cutlets_data['Unit B'].mean()
```

Out[4]:

```
6.964297142857142
```

In [5]:

```
1 stats.ttest_ind(a = cutlets_data['Unit A'], b = cutlets_data['Unit B'])
```

Out[5]:

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

Since, pvalue > 0.05, We cannot Reject Null Hypothesis

==========

A hospital wants to determine whether there is any difference in the average Turn Around Time (TAT) of reports of the laboratories on their preferred list. They collected a random sample and recorded 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 different laboratories at 5% significance level.

ANOVA test

P = 0.05

H0 : There is no significant difference in the Average TAT among different labs

Ha : There is a significant difference in the Average TAT among different labs

In [6]:

```
Lab_TAT = pd.read_csv('C:/Users/Ravi Kiran/Hypothesis Testing/LabTAT.csv')
Lab_TAT.head()
```

Out[6]:

	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 [7]:

```
stats.f_oneway(Lab_TAT['Laboratory 1'],Lab_TAT['Laboratory 2'],Lab_TAT['Laboratory 3'],
```

Out[7]:

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

Since, P-value < 0.05, We Reject the Null Hypothesis.

==========

Sales of products in four different regions is tabulated for males and females. Find if male-female buyer rations are similar across regions.

Chi - Squared test

P = 0.05

H0 : All proportions are equal

Ha : All proportions are not equal

In [8]:

```
Buyer_data = pd.read_csv('C:/Users/Ravi Kiran/Hypothesis Testing/BuyerRatio.csv')
Buyer_data.head()
```

Out[8]:

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

In [9]:

```
del Buyer_data['Observed Values']
```

In [10]:

```
Buyer_data
```

Out[10]:

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

In [11]:

```
chi2,p_val,dof,expected = stats.chi2_contingency(observed = Buyer_data)
```

In [12]:

```
print('Chi-squared value : ',round(chi2,5))
print('P-value           : ',round(p_val,5))
print('Degree of Freedom : ',dof)
```

```
Chi-squared value :  1.59595
P-value           :  0.66031
Degree of Freedom :  3
```

Since, P-value > 0.05, We cannot Reject Null Hypothesis

==========

TeleCall uses 4 centers around the globe to process customer order forms. They audit a certain % of the customer order forms. Any error in order form renders it defective and has to be reworked 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 inferences

Chi - Squared test

P = 0.05

H0: There is no significant difference in defective % across 4 centres

Ha: There is a significant difference in defective % across 4 centres

In [13]:

```
CO_data = pd.read_csv('C:/Users/Ravi Kiran/Hypothesis Testing/costomer+orderForm.csv')
CO_data
```

Out[13]:

	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 [14]:

```
CO_data.describe()
```

Out[14]:

	Phillippines	Indonesia	Malta	India
count	300	300	300	300
unique	2	2	2	2
top	Error Free	Error Free	Error Free	Error Free
freq	271	267	269	280

In [15]:

```
CO_data['Phillippines'].value_counts()
```

Out[15]:

```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
```

In [16]:

```
CO_data['Indonesia'].value_counts()
```

Out[16]:

```
Error Free    267
Defective      33
Name: Indonesia, dtype: int64
```

In [17]:

```
CO_data['Malta'].value_counts()
```

Out[17]:

```
Error Free    269
Defective      31
Name: Malta, dtype: int64
```

In [18]:

```
CO_data['India'].value_counts()
```

Out[18]:

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

In [19]:

```
chi2,p_val,dof,expected = stats.chi2_contingency([[271,267,269,280],[29,33,31,20]])
```

In [20]:

```
print('Chi-squared value : ',round(chi2,5))
print('P-value           : ',round(p_val,5))
print('Degree of Freedom : ',dof)
```

```
Chi-squared value :  3.85896
P-value           :  0.2771
Degree of Freedom :  3
```

Since, P-value >0.05, we cannot Reject Null Hypothesis.