HYPOTHESIS TESTING FOR CUTLETS DATA

In [4]:
```
!pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\91998\anaconda3\lib\site-packages (1.2.4)
Requirement already satisfied: pytz>=2017.3 in c:\users\91998\anaconda3\lib\site-packages (from pandas) (2021.1)
Requirement already satisfied: numpy>=1.16.5 in c:\users\91998\anaconda3\lib\site-packages (from pandas) (1.20.1)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\91998\anaconda3\lib\site-packages (from pandas) (2.8.1)
Requirement already satisfied: six>=1.5 in c:\users\91998\anaconda3\lib\site-packages (from python-dateutil>=2.7.3->pandas) (1.15.0)
```

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

In [18]:
```
data=cutlets_data=pd.read_csv('Cutlets.csv')
data.head()
```

Out[18]:

	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

Initial Analysis

In [20]:
```
cutlets_data.shape
```

Out[20]:
```
(35, 2)
```

In [21]:
```
cutlets_data.dtypes
```

Out[21]:
```
Unit A    float64
Unit B    float64
dtype: object
```

In [22]:
```
cutlets_data.isna().sum()
```

Out[22]:
```
Unit A    0
Unit B    0
dtype: int64
```

In [10]:
```
cutlets_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Unit A  35 non-null     float64
 1   Unit B  35 non-null     float64
dtypes: float64(2)
memory usage: 688.0 bytes
```

In [13]:
```
unitA=pd.Series(data.iloc[:,0])
unitA
```

Out[13]:
```
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 [14]:
```
unitB=pd.Series(data.iloc[:,1])
unitB
```

Out[14]:
```
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
```

H0 == (There is no significance difference between diameter of the Culets) H1 == (There is a significance difference between diameter of the Culets)

In [23]:
```
p_value=stats.ttest_ind(unitA,unitB)
p_value
```

Out[23]:
```
Ttest_indResult(statistic=0.7228688704678063, pvalue=0.4722394724599501)
```

p>0.05--where p is greater than significance level. So, here we dont reject null hypothesis.

HYPOTHESIS FOR

In [ ]:

In [ ]:

In [ ]: