

Problem Statement

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.

1. Understanding Business Problem

To find out if there is any significant difference in the diameter of the cutlet between two units (Unit-A & Unit-B)

2. Given Data

Level of Significance $\rightarrow \alpha = 0.05$

3. Import Necessary Libraries

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import statsmodels.api as sm
import statsmodels.formula.api as smf
from matplotlib import pyplot as plt
from scipy import stats

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

4. Import Data

```
In [21]: cutlets_df = pd.read_csv('Cutlets.csv')
```

```
In [3]: cutlets_df.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

5. Perform Initial Analysis

```
In [4]: cutlets_df.shape
```

```
Out[4]: (35, 2)
```

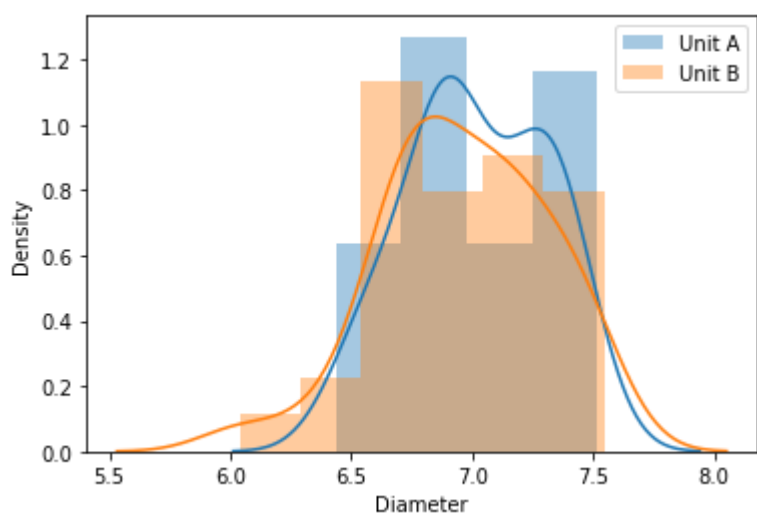
```
In [6]: cutlets_df.dtypes # to know the data type
```

```
Out[6]: Unit A      float64
Unit B      float64
dtype: object
```

```
In [5]: cutlets_df.isna().sum() # finding null/NaN values in data
```

```
Out[5]: Unit A      0
Unit B      0
dtype: int64
```

```
In [20]: plt.title('of UnitA ')
sns.distplot(cutlets_df['Unit A'], label='Unit A')
sns.distplot(cutlets_df['Unit B'], label='Unit B')
plt.xlabel('Diameter')
plt.legend()
plt.show()
```



6. Hypothesis Formulation

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

H1 : There is a significant difference in the diameter of the cutlet between two units.

α : 5%**

7. Perform Paired Test (Dependent)

Unit A and Unit B are related to each other so, we can perform paired test to get the p-values.

Since Unit A and Unit B are related to each other so, we will perform Paired test

```
In [9]: paired_test_cut_dia,p_val_cut_dia = stats.ttest_rel(a= cutlets_df['Unit A'] ,b= cutlets_df['Unit B'])
```

```
In [10]: print('*****')
print('Paired T Test Value      :',paired_test_cut_dia)
print('P-Value for cutlets diameter :',p_val_cut_dia)
print('*****')
```

```
*****
Paired T Test Value      : 0.7536787225614314
P-Value for cutlets diameter : 0.4562300768038412
*****
```

8. Conclusion

```
In [22]: # checking if ' $\alpha$  > P-value' or ' $\alpha$  < P-value'
```

```
In [11]: if p_val_cut_dia < 0.05:
print('At 5% level of significance we can reject the Null Hypothesis and we can state that there is a significant difference in the diameter of the cutlet between two units.')
else:
print('At 5% level of significance we cannot reject the Null Hypothesis and we can state that there is no significant difference in the diameter of the cutlet between two units.')
```

9. Verifying the above conclusion manually

Verifying the result by calculating the mean of 'Unit A' and 'Unit B'

```
In [12]: unit_a_mean = cutlets_df['Unit A'].mean() # calculation mean of 'Column A'
unit_a_mean
```

```
Out[12]: 7.01909142857143
```

```
In [13]: unit_b_mean = cutlets_df['Unit B'].mean() # calculation mean of 'Column B'
unit_b_mean
```

```
Out[13]: 6.964297142857142
```