

Q1. A F&B manager wants to determine whether there is any significant difference in the diameter of the outlet between two units. A randomly selected sample of outlets 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

In [1]:

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

In [2]:

```
data=pd.read_csv('Cutlets.csv')  
data
```

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
5	7.3871	6.8110
6	6.8755	7.2212
7	7.0621	6.6606
8	6.6840	7.2402
9	6.8236	7.0503
10	7.3930	6.8810
11	7.5169	7.4059
12	6.9246	6.7652
13	6.9256	6.0380
14	6.5797	7.1581
15	6.8394	7.0240
16	6.5970	6.6672
17	7.2705	7.4314
18	7.2828	7.3070
19	7.3495	6.7478
20	6.9438	6.8889
21	7.1560	7.4220
22	6.5341	6.5217
23	7.2854	7.1688
24	6.9952	6.7594
25	6.8568	6.9399
26	7.2163	7.0133
27	6.6801	6.9182
28	6.9431	6.3346
29	7.0852	7.5459
30	6.7794	7.0992
31	7.2783	7.1180
32	7.1561	6.6965
33	7.3943	6.5780

	Unit A	Unit B
34	6.9405	7.3875

Assume Null hypothesis as $H_0: \mu_1 = \mu_2$ (There is no difference in diameters of cutlets between two units) Thus
Alternate hypothesis as $H_a: \mu_1 \neq \mu_2$ (There is significant difference in diameters of cutlets between two units) 2
Sample 2 Tail test applicable

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]:

```
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]:

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

Out[6]:

Ttest_indResult(statistic=0.7228688704678063, pvalue=0.4722394724599501)

In [7]:

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

Out[7]:

0.4722394724599501

In []:

```
# compare p_value with  $\alpha = 0.05$  (At 5% significance level)
```

Inference: As ($p_value=0.4722$) > ($\alpha = 0.05$); Accept Null Hypothesis i.e. $\mu_1 = \mu_2$ (Thus, there is no difference in diameters of cutlets between two units

In []: