Assignment-03-Q1-Hypothesis-Testing

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.

In [1]:

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

In [2]:

```
# Load dataset
```

In [3]:

```
cutlet=pd.read_csv("C:/Users/LENOVO/Documents/Custom Office Templates/Cutlets.csv")
cutlet.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

In [4]:

```
unitA=pd.Series(cutlet.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(cutlet.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
tStat,p_value=stats.ttest_ind(unitA,unitB)
print("P-Value:{0} T-Statistic{1}".format(p_value,tStat))
```

```
P-Value:0.4722394724599501 T-Statistic0.7228688704678063
```

In [7]:

```
print(p_value)   # 2-tail probability
```

```
0.4722394724599501
```

In [8]:

```
# compared p_value with aplpha = 0.05 (At 5% significance level)
if p_value < 0.05:
    print('we reject null hypothesis')
else:
    print('we accept null hypoythesis')
```

```
we accept null hypoythesis
```

In [ ]:

```

```

Assignment-3-Q2 (Hypothesis Testing)

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.

In [11]:

```
# load the dataset
tat=pd.read_csv("C:/Users/LENOVO/Documents/Custom Office Templates/LabTAT.csv")
tat.head()
```

Out[11]:

	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 [12]:

```
# Anova ftest statistic: stats.f_oneway(column-1,column-2,column3,column4)
f_Stat,pvalue=stats.f_oneway(tat.iloc[:,0],tat.iloc[:,1],tat.iloc[:,2],tat.iloc[:,3])
print("P-value:{0} f-Statistic{1}".format(pvalue,f_Stat))
```

```
P-value:2.1156708949992414e-57 f-Statistic118.70421654401437
```

In [13]:

```
print(pvalue) # compare with alpha= 0.05
```

```
2.1156708949992414e-57
```

In [14]:

```
if pvalue < 0.05:
    print('we reject null hypothesis')
else:
    print('we accept null hypoythesis')
```

```
we reject null hypothesis
```

In [ ]:

```

```

Assignment-3-Q3 (Hypothesis Testing)

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

East West North South

Male 50 142 131 70

Female 550 351 480 350

H0 All proportion are equal 1. Check p-value

Ha Not all proportion are equal 2. if p-value< alpha we reject Null Hypothesis

In [15]:

```
from scipy.stats import chi2_contingency
```

In [16]:

```
buyer=pd.read_csv("C:/Users/LENOVO/Documents/Custom Office Templates/BuyerRatio.csv")
buyer
```

Out[16]:

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

In [17]:

```
# Make dimensional array
obs=np.array([[50,142,131,70],[435,1523,1356,750]])
obs
```

Out[17]:

```
array([[  50,  142,  131,   70],
       [ 435, 1523, 1356,  750]])
```

In [18]:

```
# Chi2 contengency independent test
chi2_contingency(obs)  # o/p is (chi2 stats value, p_value, df, expected observation)
```

Out[18]:

```
(1.595945538661058,
 0.6603094907091882,
 3,
 array([[  42.76531299,  146.81287862,  131.11756787,   72.30424052],
        [ 442.23468701, 1518.18712138, 1355.88243213,  747.69575948]]))
```

p_value > 0.05 we accept the Null Hypothesis

In [ ]:

```

```

Assignment-3-Q4 (Hypothesis Testing)

In [19]:

```
# Load the dataset
data=pd.read_csv("C:/Users/LENOVO/Documents/Custom Office Templates/Costomer+OrderForm.csv"
data
```

Out[19]:

	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 [20]:

```
data.Phillippines.value_counts()
```

Out[20]:

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

In [21]:

```
data.Indonesia.value_counts()
```

Out[21]:

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

In [22]:

```
data.Malta.value_counts()
```

Out[22]:

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

In [23]:

```
data.India.value_counts()
```

Out[23]:

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

In [24]:

```
# make a contingency table
obs=np.array([[271,267,269,280],[29,33,31,20]])
obs
```

Out[24]:

```
array([[271, 267, 269, 280],
       [ 29,  33,  31,  20]])
```

In [25]:

```
# Chi2 contigency independence test
chi2_contingency(obs)  # o/p is (Chi2 stats value, p_value, df, expected observation)
```

Out[25]:

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

the p_value > 0.05 we aacept Null hypothesis

In [ ]:

```

```