Hypothesis Assignment

211091055 Gaytri Thakre

```
In [4]: '''1) 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.'''
```

Out[4]: '1) A F&B manager wants to determine whether there is any\nsignificant difference in the diameter of the cut let between two\nunits. A randomly selected sample of cutlets was collected from\nboth units and measured? A nalyze the data and draw inferences\nat 5% significance level. Please state the assumptions and tests\nthat you carried out to check validity of the assumptions.'

```
In [5]: import pandas as pd
        import numpy as np
        from scipy.stats import ttest_ind
        from scipy.stats import f_oneway
        from scipy.stats import stats
```

```
In [6]: #Reading data
        data1 = pd.read_csv(r'C:\Users\Hypothesis Assignment\Cutlets.csv')
        data1.head()
```

Out[6]:

	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 [7]: unit1_diameters = data1['Unit A']
        unit2_diameters = data1['Unit B']
```

```
In [8]: #Defining significance level
        alpha = 0.05
        #Performing Ttest as 'N' is less
        t_stat, p_value = ttest_ind(unit1_diameters, unit2_diameters)
```

```
In [9]: p_value
```

Out[9]: 0.4722394724599501

```
In [10]: t_stat
```

Out[10]: 0.7228688704678063

```
In [11]: # Deciding Correct hypothesis

         if p_value < alpha:
             print("Rejecting null hypothesis as there is a significant difference in the diameter of cutlets between
         else:
             print("Fail to reject null hypothesis there is no significant difference in the diameter of cutlets betwe
```

Fail to reject null hypothesis there is no significant difference in the diameter of cutlets between the two
units.

```
In [12]: ''' 2) 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. '''
```

Out[12]: ' 2) A hospital wants to determine whether there is any\ndifference in the average Turn Around Time (TAT) of reports\nof the laboratories on their preferred list. They collected a\nrandom sample and recorded TAT for reports of 4\nlaboratories. TAT is defined as sample collected to report\ndispatch.\n\nAnalyze the data and determine whether there is any\ndifference in average TAT among the different laboratories\nat 5% significance level. '

```
In [13]: data2 = pd.read_csv(r'C:\Users\Hypothesis Assignment\LabTAT.csv')
         data2
```

Out[13]:

	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
...	...	...	...	...
115	178.49	170.66	193.80	172.68
116	176.08	183.98	215.25	177.64
117	202.48	174.54	203.99	170.27
118	182.40	197.18	194.52	150.87
119	182.09	215.17	221.49	162.21

120 rows × 4 columns

```
In [14]: tat_laboratory1 = data2['Laboratory 1']
         tat_laboratory2 = data2['Laboratory 2']
         tat_laboratory3 = data2['Laboratory 3']
         tat_laboratory4 = data2['Laboratory 4']
```

```
In [15]: #Defining Significance level
         alpha = 0.05
         #Performing anova test
         f_stat, p_value = stats.f_oneway(tat_laboratory1, tat_laboratory2, tat_laboratory3, tat_laboratory4)
```

```
C:\Users\HOME\AppData\Local\Temp\ipykernel_8148\1099572350.py:4: DeprecationWarning: Please use `f_oneway` f
rom the `scipy.stats` namespace, the `scipy.stats.stats` namespace is deprecated.
  f_stat, p_value = stats.f_oneway(tat_laboratory1, tat_laboratory2, tat_laboratory3, tat_laboratory4)
```

```
In [16]: f_stat
```

Out[16]: 118.70421654401437

```
In [17]: p_value
```

Out[17]: 2.1156708949992414e-57

```
In [18]: # Deciding Correct hypothesis

         if p_value < alpha:
             print("Rejecting null hypothesis as there is a significant difference in the diameter of cutlets between
         else:
             print("Fail to reject null hypothesis there is no significant difference in the diameter of cutlets betwe
```

```
Rejecting null hypothesis as there is a significant difference in the diameter of cutlets between the two un
its.
```

```
In [19]: '''3) Sales of products in four different regions is tabulated for males and females.
         Find if male-female buyer rations are similar across regions.'''
```

Out[19]: '3) Sales of products in four different regions is tabulated for males and females.\nFind if male-female buy
er rations are similar across regions.'

```python
In [20]: from scipy.stats import chi2_contingency
```

```python
In [21]: # Creating table of given data
         data3 = np.array([[50, 142, 131, 70],
                          [435, 1523, 1356, 750]])
```

```python
In [22]: #Defining alpha
         alpha =0.05
         #performing chi-square test
         chi, p_value, dof, expected = chi2_contingency(data3)
         chi
```

Out[22]: 1.595945538661058

```python
In [23]: dof
```

Out[23]: 3

```python
In [24]: p_value
```

Out[24]: 0.6603094907091882

```python
In [25]: # Deciding Correct hypothesis

         if p_value < alpha:
             print("Rejecting null hypothesis as there is a significant difference in the diameter of cutlets between
         else:
             print("Fail to reject null hypothesis there is no significant difference in the diameter of cutlets betwe
```

Fail to reject null hypothesis there is no significant difference in the diameter of cutlets between the two
units.

In [26]:
```
'''4) 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'''
```

Out[26]: '4) TeleCall uses 4 centers around the globe to process\ncustomer order forms. They audit a certain % of the \ncustomer order forms. Any error in order form\nrenders it defective and has to be reworked before\nprocess ing. The manager wants to check whether\nthe defective % varies by centre. Please analyze the\ndata at 5% si gnificance level and help the manager\ndraw appropriate inferences'

In [27]:
```
#Reading data
data4 = pd.read_csv(r'C:\Users\Hypothesis Assignment\Costomer+OrderForm.csv')
data4.head()
```

Out[27]:

	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

In [28]:
```
country1 = data4['Phillippines']
country2 = data4['Indonesia']
country3 = data4['Malta']
country4 = data4['India']
```

```
In [29]: data4.replace({'Error Free': 1, 'Defective': 0}, inplace=True)
         data4
```

Out[29]:

	Phillippines	Indonesia	Malta	India
0	1	1	0	1
1	1	1	1	0
2	1	0	0	1
3	1	1	1	1
4	1	1	0	1
...	...	...	...	...
295	1	1	1	1
296	1	1	1	1
297	1	1	0	1
298	1	1	1	1
299	1	0	0	1

300 rows × 4 columns

```
In [30]: #Defining Significance level
         alpha = 0.05
         #Performing anova test
         f_stat, p_value = stats.f_oneway(country1, country2, country3, country4)
```

```
C:\Users\HOME\AppData\Local\Temp\ipykernel_8148\3786107032.py:4: DeprecationWarning: Please use `f_oneway` f
rom the `scipy.stats` namespace, the `scipy.stats.stats` namespace is deprecated.
  f_stat, p_value = stats.f_oneway(country1, country2, country3, country4)
```

```
In [31]: f_stat
```

Out[31]: 1.286168556089167

```
In [32]:  p_value
```

Out[32]: 0.2776780955705948

```
In [33]:  # Deciding Correct hypothesis

          if p_value < alpha:
              print("Rejecting null hypothesis as there is a significant difference in the diameter of cutlets between
          else:
              print("Fail to reject null hypothesis there is no significant difference in the diameter of cutlets betwe
```

Fail to reject null hypothesis there is no significant difference in the diameter of cutlets between the two
units.