Lab Practical -01

Aim - To perform hypothesis testing on given studies cases

Case - 1 A company, let's call it XYZ Electronics, is planning to launch a new marketing campaign to promote their latest product. They believe that this campaign will boost sales compared to their historical average. To validate this belief, they have collected historical sales data for the past two years.

1) Relevant Data Set - Superstore from kaggle

```
import pandas  as pd
import numpy as np
from scipy.stats import ttest_1samp
import warnings
warnings.filterwarnings("ignore")
```

2) Data Preparation

```
data = pd.read_csv("/content/Sample - Superstore.csv",usecols = ['Order Date','Category','Sales'], encoding = 'windows-1252')
data.head()
```

	Order Date	Category	Sales
0	11/8/2016	Furniture	261.9600
1	11/8/2016	Furniture	731.9400
2	6/12/2016	Office Supplies	14.6200
3	10/11/2015	Furniture	957.5775
4	10/11/2015	Office Supplies	22.3680

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Order Date  9994 non-null   object
 1   Category    9994 non-null   object
 2   Sales       9994 non-null   float64
dtypes: float64(1), object(2)
memory usage: 234.4+ KB
```

```
tech_data = data[data['Category'] == 'Technology']
office_supply_data = data[data['Category'] == 'Office Supplies']
```

3) Define Hypothesis

i. Null Hypothesis - There is no signififcant difference in sales between the Technology and Office Supplier categories.

ii. Alternative Hypothesis - There is significant difference between two categories.

4) Select the Appropriate Test - Two sample T-Test

5) Perform the Hypothesis Testing

```
from scipy.stats import ttest_ind
t_statistic, p_value = ttest_ind(tech_data['Sales'], office_supply_data['Sales'])
print(f'T Statistic value is {t_statistic}')
print(f'P value is {p_value}')
```

```
T Statistic value is 19.817734170475198
P value is 2.459502339530819e-85
```

6) Interpret the result

```
alpha = 0.05
if p_value < alpha:
  print("Reject the null hypothesis and there is difference between technology and office supply in sales")
else:
  print("Fail to reject the null hypothesis, There is no significant  difference")
```

```
Reject the null hypothesis and there is difference between technology and office supply in sales
```

Case - 2) - ABC Manufacturing, a company that produces electronic components,has been experiencing quality issues with one of its product lines. They suspect that a recent change in the manufacturing process may have had an adverse effect on product quality. To address this issue, they decide to investigate whether reverting to the previous manufacturing process will lead to a significant improvement in product quality.

1) Relevant data set - micro chip quality dataset

```
df = pd.read_csv("/content/microchip_quality.csv", names = ['Test_score','Rough_score','Quality'])
df.head()
```

	Test_score	Rough_score	Quality
0	34.623660	78.024693	0
1	30.286711	43.894998	0
2	35.847409	72.902198	0
3	60.182599	86.308552	1
4	79.032736	75.344376	1

2) Data Preparation

```
roughness_accepted = list(df[df['Quality'] == 1]["Rough_score"])
roughness_rejected = list(df[df['Quality'] == 0]["Rough_score"])
```

3) Define Hypothesis

i. Null Hypothesis - There is no signififcant difference in between accepted and rejected microchips

ii. Alternative Hypothesis - There is significant difference between two accepted and rejected microchips

4) Select the Appropriate Test - Two sample T-Test

5) Perform the Hypothesis Test

```
from scipy.stats import ttest_ind
t_statistic, p_value = ttest_ind(roughness_accepted, roughness_rejected)
```

```
print(f'T Statistic value is {t_statistic}')
print(f'P value is {p_value}')
```

```
T Statistic value is 5.905665563839061
P value is 5.0730596140718295e-08
```

6) Interpret the result

```
alpha = 0.05
if p_value < alpha:
  print("Reject the null hypothesis and there is difference between accepted and rejected microchips")
else:
  print("Fail to reject the null hypothesis, There is no significant difference between accepted and rejected microchips")
```

```
Reject the null hypothesis and there is difference between accepted and rejected microchips
```

Conclusion - We understood the concept of Hypothesis Testing and perform two sample T-Test on both case studies