Department of Computer Engineering

Academic Year: 2024-25 **Semester: VIII**
Class / Branch: BE Computer **Subject: Applied Data Science Lab**

Experiment No. 3

1. Aim: To explore the inferential statistic t-test on the given dataset.

Dataset: In this experiment, three fictitious datasets are used. Reliance data mart dataset is used to explore One sample t-test. Crocin dataset and Prescore-Postscore datasets are used to study Two sample paired t-test.

2. Software used: Google Colaboratory / Jupyter Notebook

3. Theory :-

- Z-Test & T-Tests are Parametric Tests, where the Null Hypothesis is less than, greater than or equal to some value.
- A z-test is used if the population variance is known, or if the sample size is larger than 30.
- If the sample size is less than 30 and the population variance is unknown, we must use a t-test.

T test is a type of inferential statistic used to study if there is a statistical difference between two groups. Mathematically, it establishes the problem by assuming that the means of the two distributions are equal (H_0: $\mu_1=\mu_2$). If the t-test rejects the null hypothesis (H_0: $\mu_1=\mu_2$), it indicates that the groups are highly probably different.

The statistical test can be one-tailed or two-tailed. The **one-tailed test** is appropriate when there is a difference between groups in a specific direction. It is less common than the **two-tailed test.** When choosing a *t* test, you will need to consider two things: whether the groups being compared come from a single population or two different populations, and whether you want to test the difference in a specific direction.

There are three main types of t-test :

- **One Sample t-test :** Compares **mean of a single group against a known**/hypothesized/ population mean.
- **Two Sample: Paired Sample T Test:** Compares means from the **same group at different times.**
- **Two Sample: Independent Sample T Test:** Compares means for **two different groups.**

One Sample t-test:

$$\mathbf{t=\frac{(Sample\ Mean-\ Population\ Mean)}{Standard\ Error}}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$\bar{x}$ Sample mean
μ Population mean
s Sample standard deviation
n Sample size

Two-sample - Paired Sample t-test

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

$\bar{d}$ =Mean of the difference
s=Standard deviation of the difference
n =is the sample size (i.e., size of d)

- If the **calculated t value is less than critical t value or greater that the critical value (**obtained from a critical value table called the T-distribution table) **then reject the null hypothesis.**
- **P-value <significance level (α) => Reject your null hypothesis** in favor of your alternative hypothesis. Your result is statistically significant.
- **P-value >= significance level (α) => Fail to reject your null hypothesis.** Your result is not statistically significant.

4. Program

In [1]:
```
import numpy as np
import pandas as pd
from scipy import stats
Reliance_Mart_Data=pd.read_excel(r'C:\Users\Ramya\Desktop\ads sttp data\t val\RelianceDataMart.xlsx')
```

In [2]:
```
Reliance_Mart_Data
```

Out[2]:

	Rice_Bag_Weight
0	24.50
1	24.70
2	25.60
3	25.00
4	24.70
5	23.30
6	23.30
7	24.00
8	25.10
9	24.30
10	23.30
11	24.10
12	24.10
13	24.20
14	25.20
15	24.90
16	24.70
17	24.10
18	25.00
19	24.70
20	24.90
21	25.00
22	24.00
23	23.08
24	24.30
25	24.20
26	24.56
27	24.50
28	24.70

In [3]:
```
Reliance_Mart_Data.describe()
```

Out[3]:

	Rice_Bag_Weight
count	29.000000
mean	24.448207
std	0.569463
min	23.300000
25%	24.100000
50%	24.500000
75%	24.900000
max	25.600000

In [4]:
```
one_sample_result=stats.ttest_1samp(Reliance_Mart_Data,25)
```

In [5]:
```
one_sample_result
```

Out[5]:
```
Ttest_1sampResult(statistic=array([-5.23697685]), pvalue=array([1.45121657e-05]))
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
import pandas as pd
from scipy import stats
```

```
prepostscore=pd.read_excel('/content/drive/MyDrive/Colab Notebooks/Pre_Post_Score.xlsx')
```

```
prepostscore
```

	Pre_Score	Post_Score
0	18	22
1	21	25
2	16	17
3	22	24
4	19	16
5	24	29
6	17	20
7	21	23
8	23	19
9	18	20
10	14	15
11	16	15
12	16	18
13	19	26
14	18	18
15	20	24
16	12	18
17	22	25
18	15	19
19	17	16

```
Two_Sample_Results=stats.ttest_rel (prepostscore ["Pre_Score"], prepostscore ["Post_Score"])
```

```
Two_Sample_Results
```

```
Ttest_relResult(statistic=-3.231252665580312, pvalue=0.004394965993185664)
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount
```

```
import pandas as pd
from scipy import stats
```

```
Crocin_Data=pd.read_excel('/content/drive/MyDrive/Colab Notebooks/Crocin_Data_ST.xlsx')
```

```
Crocin_Data
```

	Before_Crocin	After_Crocin
0	101.0	99
1	99.0	98
2	101.0	97
3	99.9	99
4	99.8	98
6	98.0	97
6	97.0	99
7	101.0	98
8	102.0	96
9	103.0	98
10	99.0	94
11	99.9	96
12	99.8	97
13	99.7	99
14	101.1	98
15	102.3	97
16	101.0	99
17	99.0	98
18	101.0	97
19	99.9	99
20	99.8	98
21	98.0	96
22	97.0	97
23	101.0	99
24	102.0	97
25	103.0	99
26	99.0	98
27	99.9	97
28	99.8	99

```
Two_Sample_Results=stats.ttest_rel (Crocin_Data ["Before_Crocin"], Crocin_Data ["After_Crocin"])
```

```
Two_Sample_Results
```

```
Ttest_relResult(statistic=7.071712959273876, pvalue=1.0800112658101922e-07)
```

5.Conclusion :- One sample t-test has been done on the reliance data mart dataset and it has been found that difference exists between the rice bag population mean and rice bag sample mean.

Two sample paired t-test has been done on the prescore-post score dataset and Crocin dataset. In the prescore-post score dataset difference exists between the mean pre-score before studying the module and mean prescore after studying the module. In the crocin dataset it is found that temperature difference exists before and after having the crocin tablet.