

Experiment - 03

Aim: To explore the inferential statistics t-test on the given dataset

Theory: Inferential statistics allows us to make conclusions about a population based on a sample of data. One of the key methods used in inferential statistics is hypothesis testing, which helps us determine if observed differences between groups are statistically significant.

A t-test is a statistical test used to compare the means of two independent groups to determine whether the observed difference is due to chance or a significant factor.

The t-test is used to determine whether the mean age of passengers who survived is significantly different from those who did not.

Null Hypothesis (H_0): There is no significant difference in the average age between passengers who survived and those who did not.

Alternative Hypothesis (H_1): There is a significant difference in the average age between the two groups.

There are different types of t-tests used in statistical analysis:

1. **Independent (Unpaired) T-Test** – Compares means between two unrelated groups.
1. **Paired T-Test** – Compares means within the same group before and after a condition.
2. **One-Sample T-Test** – Compares the mean of a single group against a known population mean.

For this experiment, we use an **Independent T-Test** since we are comparing two separate groups:

- **Passengers who survived (Survived = 1)**
- **Passengers who did not survive (Survived = 0)**

Code:

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
import pandas as pd

df = pd.read_csv("../Titanic-Dataset.csv")
df.shape
display(df.head())

# Drop missing values in the 'Age' column
df = df.dropna(subset=['Age'])
```

```
# Split data into Survived and Not Survived groups
survived_age = df[df["Survived"] == 1]["Age"]
not_survived_age = df[df["Survived"] == 0]["Age"]

# Perform independent t-test
t_stat, p_value = stats.ttest_ind(survived_age, not_survived_age,
equal_var=False)

# Print the results
print("\nT-Test Results:")
print(f"T-Statistic: {t_stat:.4f}")
print(f"P-Value: {p_value:.4f}")

# Interpret the results
alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis: There is a significant
difference in average age between survivors and non-survivors.")
else:
    print("Fail to reject the null hypothesis: No significant
difference in average age between survivors and non-survivors.")
```

Output:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

T-Test Results:

T-Statistic: -2.0460

P-Value: 0.0412

Reject the null hypothesis: There is a significant difference in average age between survivors and non-survivors.

Conclusion: Hence, we performed inferential statistics t-test on the given dataset.