**Ex.No: 6 Implementation of T-Test – Two sample t-test and Paired T-Test**

**AIM:**

To Write a program implantation of T-Test – Two sample t-test and Paired T-Test.

**ALGORITHM:**

Step 1: Create the data

Step 2: Conduct a two sample t-test.

Step 3: Interpret the results

**PROGRAM:**

**6 A) Two-Sample T-Test:**

import numpy as np

import pandas as pd

np.random.seed(12)

wisconsin\_ages1 = stats.poisson.rvs(loc=18, mu=33, size=30)

wisconsin\_ages2 = stats.poisson.rvs(loc=18, mu=13, size=20)

wisconsin\_ages = np.concatenate((wisconsin\_ages1, wisconsin\_ages2))

print(wisconsin\_ages.mean() )

stats.ttest\_ind(a= minnesota\_ages, b= wisconsin\_ages, equal\_var=False)

Ttest\_indResult(statistic=-1.7083870793286842, pvalue=0.09073104343957748)

**6 B) Paired T-Test:**

np.random.seed(11)

before= stats.norm.rvs(scale=30, loc=250, size=100)

after = before + stats.norm.rvs(scale=5, loc=-1.25, size=100)

weight\_df = pd.DataFrame({"weight\_before":before,"weight\_after":after,

"weight\_change":after-before})

weight\_df.describe()

stats.ttest\_rel(a = before, b = after)

Ttest\_relResult(statistic=2.5720175998568284, pvalue=0.011596444318439857)

**6 C) Type I and Type II Error:**

plt.figure(figsize=(12,10))

plt.fill\_between(x=np.arange(-4,-2,0.01),

y1= stats.norm.pdf(np.arange(-4,-2,0.01)) ,

facecolor='red',

alpha=0.35)

plt.fill\_between(x=np.arange(-2,2,0.01),

y1= stats.norm.pdf(np.arange(-2,2,0.01)) ,

facecolor='grey',

alpha=0.35)

plt.fill\_between(x=np.arange(2,4,0.01),

y1= stats.norm.pdf(np.arange(2,4,0.01)) ,

facecolor='red',

alpha=0.5)

plt.fill\_between(x=np.arange(-4,-2,0.01),

y1= stats.norm.pdf(np.arange(-4,-2,0.01),loc=3, scale=2) ,

facecolor='grey',

alpha=0.35)

plt.fill\_between(x=np.arange(-2,2,0.01),

y1= stats.norm.pdf(np.arange(-2,2,0.01),loc=3, scale=2) ,

facecolor='blue',

alpha=0.35)

plt.fill\_between(x=np.arange(2,10,0.01),

y1= stats.norm.pdf(np.arange(2,10,0.01),loc=3, scale=2),

facecolor='grey',

alpha=0.35)

plt.text(x=-0.8, y=0.15, s= "Null Hypothesis")

plt.text(x=2.5, y=0.13, s= "Alternative")

plt.text(x=2.1, y=0.01, s= "Type 1 Error")

plt.text(x=-3.2, y=0.01, s= "Type 1 Error")

plt.text(x=0, y=0.02, s= "Type 2 Error");

**OUTPUT:**

**RESULT:**

**Ex.No: 7 Implementation of** Variance Analysis ( ANOVA)

**AIM:**

To write a program implementation of Variance analysis (ANOVA).

**ALGORITHM:**

A. Input:

A bunch of students from different colleges taking the same exam. You want to see if one college outperforms the other, hence your null hypothesis is that the meansof GPAs in each group are equivalent to those of the other groups. To keep it simple,we will consider 3 groups (college ‘A’, ‘B’, ‘C’) with 6 students each.

A=[25,25,27,30,23,20]

B=[30,30,21,24,26,28]

C=[18,30,29,29,24,26]

Null Hypothesis: GPAs in each group are equivalent to those of the other groups.

Alternate Hypothesis – There is a significant difference among the groups

B. Output:

To find the null hypothesis or alternate hypothesis is acceptable or not.

1. Rows are grouped according to their value in the category column.

2. The total mean value of the value column is computed.



3. The mean within each group is computed.

4. The difference between each value and the mean value for the group is calculated and squared.

5. The squared difference values are added. The result is a value that relates tothe total deviation of rows from the mean of their respective groups. This valueis referred to as the sum of squares within groups, or S2Wthn.

6. For each group, the difference between the total mean and the group mean issquared and multiplied by the number of values in the group. The results areadded. The result is referred to as the sum of squares between groups or S2Btwn.



7. The two sums of squares are used to obtain a statistic for testing the nullhypothesis, the so called F-statistic. The F-statistic is calculated as:



wheredfBtwn (degree of freedom between groups) equals the number of groupsminus 1, and dfWthn (degree of freedom within groups) equals the totalnumber of values minus the number of groups

8. The F-statistic is distributed according to the F-distribution (commonlypresented in mathematical tables/handbooks). The F-statistic, in combinationwith the degrees of freedom and an F-distribution table, yields the p-value.

The p-value is the probability of the actual or a more extreme outcome underthe null-hypothesis. The lower the p-value, the larger the difference.

**PROGRAM:**

import pandas as pd

importnumpy as np

importscipy.stats as stats

a=[25,25,27,30,23,20]

b=[30,30,21,24,26,28]

c=[18,30,29,29,24,26]

list\_of\_tuples = list(zip(a, b,c))

df = pd.DataFrame(list\_of\_tuples, columns = ['A', 'B', 'C'])

df

m1=np.mean(a)

m2=np.mean(b)

m3=np.mean(c)

print('Average mark for college A: {}'.format(m1))

print('Average mark for college B: {}'.format(m2))

print('Average mark for college C: {}'.format(m3))

m=(m1+m2+m3)/3

print('Overall mean: {}'.format(m))

SSb=6\*((m1-m)\*\*2+(m2-m)\*\*2+(m3-m)\*\*2)

print('Between-groups Sum of Squared Differences: {}'.format(SSb))

MSb=SSb/2

print('Between-groups Mean Square value: {}'.format(MSb))

err\_a=list(a-m1)

err\_b=list(b-m2)

err\_c=list(c-m3)

err=err\_a+err\_b+err\_c

ssw=[]

fori in err:

ssw.append(i\*\*2)

SSw=np.sum(ssw)

print('Within-group Sum of Squared Differences: {}'.format(SSw))

MSw=SSw/15

print('Within-group Mean Square value: {}'.format(MSw))

F=MSb/MSw

print('F-score: {}'.format(F))

print(stats.f\_oneway(a,b,c))

**OUTPUT:**

**RESULT:**