HYPOTHESIS TESTING (T-TEST)

Example1: A manufacturer claims that the average weight of a bag of potato chips is 150 grams. A sample of 25 bags is taken, and the average weight is found to be 148 grams, with a standard deviation of 5 grams. Test the manufacturer's claim using a one-tailed t-test with a significance level of 0.05.

Null Hypothesis : the average weight is 150 grams

Alternative hypothesis : The avg weight is less than 150 grams.

given that:

```
mu = 150
n = 25
x_bar = 148
std = 5
alpha = 0.05
```

Importing libraries

```
import numpy as np
import scipy.stats as st
import math as m
```

calculate t-statistics

```
# Get critical value for 95% confidence (one-tailed test)
t = (x_bar-mu)/(std/m.sqrt(n))
print(f"calculated T score is:{t}")
```

```
calculated T score is:-2.0
```

Degree of freedom

```
df = n-1  #it's needed to determine the correct t-distribution.
```

find critical value

```
T_table = st.t.ppf(alpha,df)
print(f'critical T-value is :{T_table:.2}')
```

```
critical T-value is :-1.7
```

Take decision

```
if t<T_table:
    print("reject the null hypothesis")
else:
    print("fail to reject the null hypothesis")
```

```
reject the null hypothesis
```

Example 2: A company wants to test whether there is a difference in productivity between two teams. They randomly select 20 employees from each team and record their productivity scores. The mean productivity score for Team A is 80 with a standard deviation of 5, while the mean productivity score for Team B is 75 with a standard deviation of 6. Test at a 5% level of significance whether there is a difference in productivity between the two teams.

Null Hypothesis: There is no difference in productivity between two teams.

Alternative Hypothesis: There is a difference in productivity between two teams.

```
#given data:
a = 20
b = 20
a_mean = 80
a_std = 5
b_mean = 75
b_std = 6
aplha = 0.05
```

```python
#Degree of freedom
df = a+b-2
print(df)
```

```
38
```

```python
# calculate T score
T_score = (80-75)/m.sqrt((25/20)+m.sqrt(36/20))
print(f"T score is :{T_score:.2}")
```

```
T score is :3.1
```

```python
#finding critical value:
T_critical= st.t.ppf(aplha,df)
print(f"T_critical value is:{T_critical}")
```

```
T_critical value is:-1.6859544601667373
```

```python
# Decision
if T_score>T_critical:
    print("reject null hypothesis")
else:
    print("fail to reject null hypothesis")
```

```
reject null hypothesis
```

```python
#Given data
std = 0.5
n = 30
x_bar = 4.4
alpha = 0.05
mu = 4.2

#compute z score
z = (x_bar - mu)/(std/math.sqrt(n))
print(f'z score is : {z:.2f}')
```

```
z score is : 2.19
```

```python
# finding critical value
z_table = st.norm.ppf(1-alpha)
print(f'z critical value is: {z_table:.2f}')
```

```
z critical value is: 1.64
```

```python
#decision based on critical value
if z>z_table:
    print('null hypothesis is rejected')
else:
    print('fail to reject null hypothesis')
```

```
null hypothesis is rejected
```

```
#re check by p-value
p_value = 1-st.norm.cdf(z)
print(p_value)
```

```
0.014229868458155215
```

```
if p_value<alpha:
    print('reject h0')
else:
    print('accept h0')
```

```
reject h0
```

A company wants to test whether a new training program improves the typing speed of its employees. The typing speed of 20 employees was recorded before and after the training program. The data is given below. Test at a 5% level of significance whether the training program has an effect on the typing speed of the employees.

Null Hypothesis:The training program has no effect, i.e., the mean difference is zero.

Alternative Hypothesi: The training program has an effect, i.e., the mean difference is not zero.

```
#given:
Before=np.array([50, 60, 45, 65, 55, 70, 40, 75, 80, 65, 70, 60, 50,
55, 45, 75, 60, 50, 65, 70])

After=np.array([60, 70, 55, 75, 65, 80, 50, 85, 90, 70, 75, 65, 55,
60, 50, 80, 65, 55, 70, 75])

n = 20
alpha = 0.05
difference = After - Before
mean_difference = np.mean(difference)
std_diff = np.std(difference, ddof=1)
#degree of freedom:
df = n-1
```

```python
#compute t score
T = mean_difference / (std_diff / m.sqrt(n))
print(f"T score is :{T}")
```

```
T score is :12.704488986340237
```

```python
# finding the critical value
T_critical = st.t.ppf(aplha,df)
print(f"T critical value is :{T_critical}")
```

```
T critical value is :-1.7291328115213678
```

```python
#deciiosn
if T > T_critical:
    print("Reject the null hypothesis: There is a significant
difference.")
else:
    print("Fail to reject the null hypothesis: No significant
difference.")
```

```
Reject the null hypothesis: There is a significant difference.
```