

Importing the necessary libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.stats as stats
```

### ✓ 1. One-sample t-test

The mass of a sample of N 20 acorns from a forest subjected to acid rain from a coal power plant are m [8.8, 6.6, 9.5, 11.2, 10.2, 7.4, 8.0, 9.6, 9.9, 9.0, 7.6, 7.4, 10.4, 11.1, 8.5, 10.0, 11.6, 10.7, 10.3, and 7.0 g] Does this sample provide enough evidence ( $\alpha = 0.05$ ) to say that the average mass of all acorns is different from 10 g?

```
m = np.array([8.8, 6.6, 9.5, 11.2, 10.2, 7.4, 8.0, 9.6, 9.9, 9.0, 7.6, 7.4,
              10.4, 11.1, 8.5, 10.0, 11.6, 10.7, 10.3, 7.0])
avg = 10

t_stat_onesample, p_value_onesample = stats.ttest_1samp(m, avg)
t_stat_onesample, p_value_onesample
```

```
↩ (-2.2491611580763973, 0.03655562279112415)
```

The p-value (0.0366) is less than the significance level of 0.05, we reject the null hypothesis. There is enough evidence to say that the average mass of all acorns is different from 10 g.

### ✓ 2. Independent (unpaired) two-sample t-test

The mass of  $N_1 = 20$  acorns from oak trees up wind from a coal power plant and  $N_2 = 30$  acorns from oak trees down wind from the same coal power plant are measured. Is the mass of acorns from trees down wind different from the ones from up wind at a significance level of 0.05?

The sample sizes are not equal but we will assume that the population variance for sample 1 and sample 2 are equal. sample up wind:  $x_1 = [10.8, 10.0, 8.2, 9.9, 11.6, 10.1, 11.3, 10.3, 10.7, 9.7, 7.8, 9.6, 9.7, 11.6, 10.3, 9.8, 12.3, 11.0, 10.4, 10.4]$  sample down wind:  $x_2 = [7.8, 7.5, 9.5, 11.7, 8.1, 8.8, 8.8, 7.7, 9.7, 7.0, 9.0, 9.7, 11.3, 8.7, 8.8, 10.9, 10.3, 9.6, 8.4, 6.6, 7.2, 7.6, 11.5, 6.6, 8.6, 10.5, 8.4, 8.5, 10.2, 9.2]$

```
x1 = np.array([10.8, 10.0, 8.2, 9.9, 11.6, 10.1, 11.3, 10.3, 10.7, 9.7, 7.8, 9.6, 9.7, 11.6, 10.3, 9.8, 12.3, 11.0, 10.4, 10.4])
x2 = np.array([7.8, 7.5, 9.5, 11.7, 8.1, 8.8, 8.8, 7.7, 9.7, 7.0, 9.0, 9.7, 11.3, 8.7, 8.8, 10.9, 10.3, 9.6, 8.4, 6.6, 7.2, 7.6, 11.5, 6.6,
              8.6, 10.5, 8.4, 8.5, 10.2, 9.2])

t_stat_twosample, p_value_twosample = stats.ttest_ind(x1, x2, equal_var=True)
t_stat_twosample, p_value_twosample
```

```
↩ (3.5981947686898033, 0.0007560337478801464)
```

As the p-value (0.00076) is less than 0.05, we reject the null hypothesis. This means there is significant evidence that the mass of acorns from trees downwind is different from those upwind.

### ✓ 3. ANOVA test


The marks obtained by 5 randomly picked students in Mathematics exam from three sections A, B, and C are as follows: Marks of 5 randomly picked students from Section A A[51, 45, 33, 45, 67] Marks of 5 randomly picked students from Section B B[23, 43, 23, 43, 45] Marks of 5 randomly picked students from Section C C[56, 76, 74, 87, 56] Does the sample provide enough evidence to say that the mean marks of students in the three sections are different?

```
import scipy.stats as stats

A = np.array([51, 45, 33, 45, 67])
B = np.array([23, 43, 23, 43, 45])
C = np.array([56, 76, 74, 87, 56])

f_stat, p_value_anova = stats.f_oneway(A, B, C)

f_stat, p_value_anova
```

 (9.747205503009463, 0.0030597541434430556)

The p-value (0.0030) is less than the significance level of 0.05, we reject the null hypothesis. This means there is sufficient evidence to conclude that the mean marks of students in the three sections are significantly different.