

Problem 3

An important quality characteristic used by the manufacturers of ABC asphalt shingles is the amount of moisture the shingles contain when they are packaged. Customers may feel that they have purchased a product lacking in quality if they find moisture and wet shingles inside the packaging. In some cases, excessive moisture can cause the granules attached to the shingles for texture and colouring purposes to fall off the shingles resulting in appearance problems. To monitor the amount of moisture present, the company conducts moisture tests. A shingle is weighed and then dried. The shingle is then reweighed, and based on the amount of moisture taken out of the product, the pounds of moisture per 100 square feet are calculated. The company would like to show that the mean moisture content is less than 0.35 pound per 100 square feet.

The file (A & B shingles.csv) includes 36 measurements (in pounds per 100 square feet) for A shingles and 31 for B shingles.

Here we have to first import all the required libraries

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import scipy.stats as stats
from scipy.stats import ttest_1samp, ttest_ind
from statsmodels.stats.power import ttest_power
from scipy.stats import f
```

Note that two libraries imported

ttest_1samp: To perform 1 sample t test

ttest_ind : To perform 2 sample t test

in Shingles B, there are Some rows are having no values, we will discard it while doing calculations.

3.1 Do you think there is evidence that mean moisture contents in both types of shingles are within the permissible limits? State your conclusions clearly showing all steps.

Step 1:

State H_0 and H_a :

H_0 : mean moisture content ≤ 0.35 H_a : mean moisture content > 0.35

Step 2:

Value of Alpha: Confidence level:

Alpha: 0.05

Step 3:

Decide Test

1 Sample T test

Step 4:

Calculate t stat value and P value in Python

```
t1,p1_value = stats.ttest_1samp(df['A'],0.35)
print(t1,p1_value)
```

The T statistic is: -3.1003313069986995 The corresponding p value is : 0.0020904774003191826

For Shingles A value of p is 0.07477633144907513 is greater than alpha, 0.07477633144907513 > 0.05, Null Hypothesis is accepted. mean moisture contents in type A of shingles are within the permissible limits.

For Shingles B value of p is 0.0020904774003191826 is less than alpha, 0.0020904774003191826 < 0.05, Null Hypothesis will be rejected. mean moisture contents in type B of shingles are not within the permissible limits. to get more correct results we need more no of samples.

3.2 Do you think that the population means for shingles A and B are equal? Form the hypothesis and conduct the test of the hypothesis. What assumption do you need to check before the test for equality of means is performed?

Step 1:

State H_0 and H_a :

H_0 : Population mean of Shingles A = Population mean of Shingles B

H_a : Population mean of Shingles A greater than or less than Population mean of Shingles B

Step2:

Confidence level:

Alpha = 0.05

Step 3:

We are using 2 sample t Test.

```
# We are using 2 sample t test

alpha = 0.05

t3, p3_value = ttest_ind(df['A'], df['B'], equal_var=True, nan_policy='omit')
print(t3, p3_value)

1.2896282719661123 0.2017496571835306
```

Step 4:

The T statistic is: 1.2896282719661123 The corresponding pvalue is : 0.2017496571835306

Here, P value is 0.2017496571835306 is greater than 0.05 hence Null hypothesis will be accepted

Two-Sample T-Test Assumptions we made:

The assumptions of the two-sample t-test are: The data are continuous (not discrete).

The data follow the normal probability distribution.

The two samples are independent. Both samples are simple random samples from their respective populations.