Problem Statement:

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.

[redacted]$_0$[redacted]$_A$[redacted]$_0$[redacted]$_A$

Exploratory Data Analysis:

	A	B
0	0.44	0.14
1	0.61	0.15
2	0.47	0.31
3	0.30	0.16
4	0.15	0.37

Dataset has 2 variables A and B, which has the measurement of moisture present per 100 sq. ft.

Both variables are float in data types.

Descriptive Statistics for both variable:

A		***B***	
Mean	0.316666667	**Mean**	0.273548
Standard Error	0.022621804	**Standard Error**	0.024659
Median	0.29	**Median**	0.23
Mode	0.2	**Mode**	0.11
Standard Deviation	0.135730826	**Standard Deviation**	0.137296
Sample Variance	0.018422857	**Sample Variance**	0.01885
Kurtosis	0.979774347	**Kurtosis**	-0.90955
Skewness	0.950618572	**Skewness**	0.513424
Range	0.59	**Range**	0.48
Minimum	0.13	**Minimum**	0.1
Maximum	0.72	**Maximum**	0.58
Sum	11.4	**Sum**	8.48
Count	36	**Count**	31

Standard Deviation for both variables are almost identical.

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

For the A shingles, the null and alternative hypothesis to test whether the population mean moisture content is less than 0.35 pound per 100 square feet is given:

Variable A has the mean of 0.32 that means the Variable A has moisture 0.32 per 100 sq. ft. on an average. Variable B has moisture of 0.27 per 100 sq. ft. on an average.

To check whether the mean moisture control for A shingles is within permissible limits, the following null and alternative hypotheses are formulated

H_0 <= 0.35

H_A > 0.35

```
t_statistic,p_value=stats.ttest_1samp(data.A,0.35) #by default t-test is a two sided ttest in python
#t-distribution moves left to right (-ve to +ve side)
print('The test statistic is {}'.format(t_statistic))
p_value_greater= 1-(p_value/2) #p_value given will be two sided p-value
#hence dividing pvalue by 2 and subtracting it by 1 will give pvalue of right side of t-distribution, since alternate lies on right side
if p_value_greater>0.05:
    print('The p-value is {} which is greater than the level of significance , hence, we fail to reject the Null Hypothesis'.format(p_value_greater))
else:     print('The p-value is {} which is less than the level of significance , hence, we reject the Null Hypothesis'.format(p_value_greater))
```

```
The test statistic is -1.4735046253382782
The p-value is 0.9252236685509249 which is greater than the level of significance , hence, we fail to reject the Null Hypothesis
```

Since p-value of the test is less than α = 0.05, H_0 is not rejected. We may conclude that for A shingles the mean moisture content is within the permissible limits

For the B shingles, the null and alternative hypothesis to test whether the population mean moisture content is less than 0.35 pound per 100 square feet is given:

H_0<=0.35

H_A>0.35

```
[ ] t_statistic,p_value=stats.ttest_1samp(data.B.dropna(),0.35) #by default t-test is a two sided ttest in python
    print('The test statistic is {}'.format(t_statistic))
    p_value_greater= 1-(p_value/2) #hence dividing pvalue by 2 and subtracting it by 1 will give pvalue of right side of t-distribution
    if p_value_greater>0.05:
        print('The p-value is {} which is greater than the level of significance , hence, we fail to reject the Null Hypothesis'.format(p_value_greater))
    else:     print('The p-value is {} which is less than the level of significance , hence, we reject the Null Hypothesis'.format(p_value_greater))
```

```
The test statistic is -3.1003313069986995
The p-value is 0.9979095225996808 which is greater than the level of significance , hence, we fail to reject the Null Hypothesis
```

Since p-value of the test is less than α = 0.05, H_0 is not rejected. We may conclude that for A shingles the mean moisture content is within the permissible limits

Commented [1]: Added code snippet for A Shingle, p_value is greater than alpha

Commented [2]: Snippet for B Shingles added

b. **Do you think that the population means for shingles A and B are equal? Form a hypothesis and conduct a test.**

Denote , the population mean for shingles A by μ_A and the population mean for shingles B by μ_B

Null hypothesis states that the population means of Shingles A and B are equal,
X
Alternative hypothesis states that the population means of Shingles A and B are not equal,

H_0: μA = μB
H_A: $\mu_A \neq \mu_B$

To perform the hypothesis testing, the following assumptions must hold

The variables must follow continuous distribution The sample must be randomly collected from the populationThe underlying distribution must be normal Alternatively, if the data is continuous, but may not be assumed to follow a normal distribution, a reasonably large sample size is required. Central Limit Theorem (CLT) asserts that sample mean follws a normal distribution, even if the population distribution is not normal, when sample size is at least 30.For 2-sample t-test the population variances of the two distributions must be equal.

```
[ ] t_statistic, p_value =  ttest_ind(data['A'], data['B'], axis=0, equal_var=False, nan_policy='omit')

    print('Our t test \nt statistic: {0} p value: {1} '.format(t_statistic, p_value))

    # p_value > 0.05 => Null hypothesis:
    # p_value < 0.05 => Alternate hypothesis:
    # the population means for shingles A and B are equal

    alpha_value = 0.05 # Level of significance
    print('Level of significance: %.2f' %alpha_value)
    print ("Our t-test p-value=", p_value)

    if p_value < alpha_value:
        print('We have evidence to reject the null hypothesis since p value < Level of significance')
    else:
        print('We have no evidence to reject the null hypothesis since p value > Level of significance')
```

```
One sample t test
t statistic: 1.2885080295255027 p value: 0.2022582205021782
Level of significance: 0.05
Our one-sample t-test p-value= 0.2022582205021782
We have no evidence to reject the null hypothesis since p value > Level of significance
```

Commented [3]: Code snippet for both A and B added

t statistic: 1.2885080295255027 p value: 0.2022582205021782

Level of significance: 0.05

Our t-test p-value= 0.2022582205021782

Conclusion: We have no evidence to reject the null hypothesis, since p value > Level of significance