

Problem 3 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 is calculated. The company claims that the mean moisture content cannot be greater 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.

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:

$$H_0 \leq 0.35$$

$$H_A > 0.35$$

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:

$$H_0 \leq 0.35$$

$$H_A > 0.35$$

Exploratory Data Analysis:

A	B
0.44	0.14
0.61	0.15
0.47	0.31
0.3	0.16
0.15	0.37
0.24	0.18
0.16	0.42
0.2	0.58
0.2	0.25
0.2	0.41
0.26	0.17
0.14	0.13
0.33	0.23
0.13	0.11
0.72	0.1

The given data set contains 2 columns of float type data and there are 5 null values in column B

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  ---
0    A         36 non-null      float64
1    B         31 non-null      float64
dtypes: float64(2)
memory usage: 704.0 bytes
```

Descriptive statistics for the dataset:

	A	B
count	36.000000	31.000000
mean	0.316667	0.273548
std	0.135731	0.137296
min	0.130000	0.100000
25%	0.207500	0.160000
50%	0.290000	0.230000
75%	0.392500	0.400000
max	0.720000	0.580000

3.1 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?

In testing whether the population means for shingles A and B are equal , the null hypothesis states that the mean are equal and the alternate hypothesis states that the mean are different

Ho : mean A = mean B

Ha : mean A not = mean B

We have two samples and we don't have population standard deviation, we cannot proceed with Z test , we have to perform T test.

We select 5% level of significance

```
ttest_ind(sv['A'],sv['B'],nan_policy = 'omit')
```

```
Ttest_indResult(statistic=1.2896282719661123, pvalue=0.2017496571835306)
```

pvalue =0.2017496571835306 is greater than 5% level of significance i.e, 0.05, Therefore we accept the null hypothesis , which means at 95% confidence we can say that population means for shingles A and B are equal.

3.2 What assumption about the population distribution is needed in order to conduct the hypothesis tests above?

We assume that the samples are randomly selected, independent and come from a normally distributed population and also equal variance exists when the standard deviations of samples are approximately equal.