

QUESTION : APPLY ADVANCED STATISTICAL TECHNIQUES TO A DATASET , PRESENTING FINDINGS AND INSIGHTS

✓ IMPORTING NUMPY LIBRARY

```
import numpy as np
```

✓ CREATING A DATAFRAME CALLED SCORES

```
scores = np.array([55,60,35,70,67,40,59])
```

✓ DISPLAYING THE DATAFRAME

```
print(scores)

[55 60 35 70 67 40 59]
```

✓ CHECKING THE SHAPE OF THE ARRAY

```
scores.shape

(7,)
```

✓ SIZE OF THE ARRAY

```
scores.size

7
```

✓ TYPE OF THE ARRAY

```
scores.dtype
```

```
dtype('int64')
```

STATISTICAL TECHNIQUES

✓ MEAN OF THE ARRAY

```
mean= np.mean(scores)  
print(mean)
```

```
55.142857142857146
```

✓ MEDIAN OF THE ARRAY

```
median = np.median(scores)  
print(median)
```

```
59.0
```

✓ STANDARD DEVIATION OF THE ARRAY

```
std_dev = np.std(scores)  
print(std_dev)
```

```
12.15881304625898
```

✓ VARIANCE OF THE ARRAY

```
var = np.var(scores)  
print(var)
```

```
147.83673469387756
```

INFERENCE STATISTICS

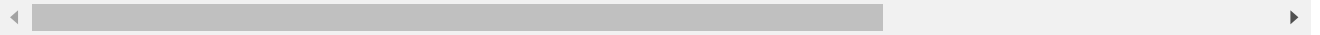
- ✓ Five student were chosen at random with scores of test 66,59,78,43,65. How is it possible for the mean score of

```
#performing t-test  
scores_1= [66,59,78,43,65]  
print(scores_1)
```

```
[66, 59, 78, 43, 65]
```

```
!pip install spacy
```

```
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages (0.16  
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from  
Requirement already satisfied: numpy<1.28.0,>=1.21.6 in /usr/local/lib/python3.10/dis
```



```
from spacy import stats as st
```

```
st.ttest_1samp(scores,65)
```

```
TtestResult(statistic=-1.9857999485524458, pvalue=0.09426313289371169, df=6)
```

since , the pvalue = 0.094 which is a strong evidence against the hypothesis, that means the hypothesis is therefore rejected.