## Problem Statement

A hospital wants to determine whether there is any difference in the average Turn Around Time (TAT) of reports of the laboratories on their preferred list. They collected a random sample and recorded TAT for reports of 4 laboratories. TAT is defined as sample collected to report dispatch. Analyze the data and determine whether there is any difference in average TAT among the different laboratories at 5% significance level.

In [1]:
```python
import pandas as pd
import numpy as np
import scipy as sc
from scipy import stats
```

In [2]:
```python
lab = pd.read_csv("D:/DataScience/Class/assignment working/Hypothesis/lab_tat_updated.csv")
```

In [3]:
```python
lab.head()
```
Out[3]:

|   | Laboratory_1 | Laboratory_2 | Laboratory_3 | Laboratory_4 |
|---|---|---|---|---|
| 0 | 185.35 | 165.53 | 176.70 | 166.13 |
| 1 | 170.49 | 185.91 | 198.45 | 160.79 |
| 2 | 192.77 | 194.92 | 201.23 | 185.18 |
| 3 | 177.33 | 183.00 | 199.61 | 176.42 |
| 4 | 193.41 | 169.57 | 204.63 | 152.60 |

In [4]:
```python
lab.shape
```
Out[4]:

(120, 4)

In [5]:
```python
lab.describe()
```
Out[5]:

|  | Laboratory_1 | Laboratory_2 | Laboratory_3 | Laboratory_4 |
|---|---|---|---|---|
| count | 120.000000 | 120.000000 | 120.000000 | 120.00000 |
| mean | 178.257333 | 178.902917 | 200.210167 | 163.68275 |
| std | 13.919668 | 14.957114 | 15.794801 | 15.08508 |
| min | 140.250000 | 140.550000 | 170.580000 | 124.06000 |
| 25% | 170.267500 | 168.025000 | 190.182500 | 154.05000 |
| 50% | 179.055000 | 178.870000 | 198.610000 | 164.42500 |
| 75% | 187.222500 | 189.112500 | 211.197500 | 172.88250 |
| max | 216.390000 | 217.860000 | 238.700000 | 205.18000 |

## Normality test

In [8]:
```python
stats.shapiro(lab.Laboratory_1)
```
Out[8]:

ShapiroResult(statistic=0.9886691570281982, pvalue=0.42317795753479004)

In [9]:
```python
stats.shapiro(lab.Laboratory_2)
```
Out[9]:

ShapiroResult(statistic=0.9936322569847107, pvalue=0.8637524843215942)

In [10]:
```python
stats.shapiro(lab.Laboratory_3)
```
Out[10]:

ShapiroResult(statistic=0.9796067476272583, pvalue=0.06547004729509354)

In [11]:
```python
stats.shapiro(lab.Laboratory_4)
```
Out[11]:

ShapiroResult(statistic=0.9913753271102905, pvalue=0.6618951559066772)

**from above re sults we failed to reject null hypothesis, so we can say data is normal.**

## One way ANOVA

*Variance Test*

In [19]:

```
# levene test to check variance of population with more than two groups
```

In [13]:

```
stats.levene(lab.Laboratory_1,lab.Laboratory_2,lab.Laboratory_3,lab.Laboratory_4)
```

Out[13]:

```
LeveneResult(statistic=1.025294593220823, pvalue=0.38107781677304564)
```

**failed to reject null hypothesis as p value is greater than 0.05, so we can coclude that the deta have equal variance**

**We have following Hypothesis for the test:**
***Null Hypothesis:* TAT1= TAT2= TAT3 = TAT4**
***Alternate Hypothese:* atleast TAT for one lab is different**

In [20]:

```
F, p = stats.f_oneway(lab.Laboratory_1,lab.Laboratory_2,lab.Laboratory_3,lab.Laboratory_4)
```

In [22]:

```
print(F)
print(p)
```

```
121.39264646442368
2.143740909435053e-58
```

## Summary and inference

**from above results we can say that all above group have different popuation mean.**

**we can conclude that all the above laboratories have different mean time**