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.

File: **LabTAT.csv**

**Business Problem:** Find out difference in the average TAT of reports of the laboratories.

Inputs are Laboratory 1, Laboratory 2, Laboratory 3, Laboratory 4 that is **Discrete in more than two categories,** Outputis **continuous.** as we are trying to see the difference in average TAT of reports of the laboratories and Output is a continuous variable.

We proceed with  **ANOVA one-way test**

1) **Normality test-** We will see if data is normally distributed or not

Then we can go with **Shapiro wilk** normality test.

🡪Create hypothesis for **Laboratory 1**

Ho= Data is Normally distributed, Ha=Data is not Normally distributed

**# p-value = 0.5508 >0.05 so p high null fly => It follows normal distribution**

**🡪** Create hypothesis for **Laboratory 2**

Ho= Data is Normally distributed, Ha=Data is not Normally distributed

**# p-value = 0.8637 >0.05 so p high null fly => It follows normal distribution**

🡪Create hypothesis for **Laboratory 3**

Ho= Data is Normally distributed, Ha=Data is not Normally distributed

**# p-value = 0.4205 >0.05 so p high null fly => It follows normal distribution**

**🡪** Create hypothesis for **Laboratory 4**

Ho= Data is Normally distributed, Ha=Data is not Normally distributed

**# p-value = 0.6619 >0.05 so p high null fly => It follows normal distribution**

So the data is normally distributed.

**2) Variance Test:**  We can go with var.test()

**🡪** Create hypothesis for **Laboratory 1** and **Laboratory 2**

#H0=Variance of TAT of Lab 1 is equal to variance of TAT of Lab 2

#H1=Variance of TAT of Lab 1 is not equal to variance of TAT of Lab 2

**# p-value = 0.1675 > 0.05 so p high null fly accept null hypothesis..**

**🡪** Create hypothesis for **Laboratory 2** and **Laboratory 3**

#H0=Variance of TAT of Lab 2 is equal to variance of TAT of Lab 3

#H1=Variance of TAT of Lab 2 is not equal to variance of TAT of Lab 3

**# p-value = 0.2742 > 0.05 so p high null fly accept null hypothesis.**

**🡪** Create hypothesis for **Laboratory 3** and **Laboratory 4**

#H0=Variance of TAT of Lab 3 is equal to variance of TAT of Lab 4

#H1=Variance of TAT of Lab 3 is not equal to variance of TAT of Lab 4

**# p-value = 0.3168 > 0.05 so p high null fly accept null hypothesis**

**🡪** Create hypothesis for **Laboratory 4** and **Laboratory 1**

#H0=Variance of TAT of Lab 4 is equal to variance of TAT of Lab 1

#H1=Variance of TAT of Lab 4 is not equal to variance of TAT of Lab 1

**# p-value = 0.1408 > 0.05 so p high null fly accept null hypothesis.**

As there are more than 2 discrete variables and output variable TAT is a continuous variable. Hence we will go with **Anova one way test**.

**3) 1-way ANOVA test:**

**🡪** Create hypothesis

#Ho= Average TAT for all the samples is same

#Ha= Averages TAT for atleast one sample is not same

So we proceed with **aov** test

**#Laboratory2 p-value = 0.166 > 0.05 so p high null fly accept null hypothesis.**

**#Laboratory3 p-value = 0.277 > 0.05 so p high null fly accept null hypothesis.**

**#Laboratory4 p-value = 0.215 > 0.05 so p high null fly accept null hypothesis.**

**Inferences:**

🡪there is no significant difference in the average TAT for all the laboratories.

**R-Code:**

**# Load the Dataset**

library(readxl)

**######## LabTAT.xlsx data ##########**

LabTAT <- read\_excel(file.choose())

View(LabTAT)

attach(LabTAT)

**#############Normality test###############**

#Ho= Data is normally distributed (no action take)

#Ha=Data is not normally distributed (action take)

**shapiro.test(`Laboratory 1`)**

# p-value = 0.5508 >0.05 so p high null fly => It follows normal distribution

**shapiro.test(`Laboratory 2`)**

# p-value = 0.8637 >0.05 so p high null fly => It follows normal distribution

**shapiro.test(`Laboratory 3`)**

# p-value = 0.4205 >0.05 so p high null fly => It follows normal distribution

**shapiro.test(`Laboratory 4`)**

# p-value = 0.6619 >0.05 so p high null fly => It follows normal distribution

**#############Variance test###############**

#Ho= Variance of TAT of all Labs are equal

#H1=atleast 1 Lab variance of TAT are not equal

**var.test(`Laboratory 1`,`Laboratory 2`)**

#H0=Variance of TAT of Lab 1 is equal to variance of TAT of Lab 2

#H1=Variance of TAT of Lab 1 is not equal to variance of TAT of Lab 2

# p-value = 0.1675 > 0.05 so p high null fly accept null hypothesis.

**var.test(`Laboratory 2`,`Laboratory 3`)**

# p-value = 0.2742 > 0.05 so p high null fly accept null hypothesis.

**var.test(`Laboratory 3`,`Laboratory 4`)**

# p-value = 0.3168 > 0.05 so p high null fly accept null hypothesis.

**var.test(`Laboratory 4`,`Laboratory 1`)**

# p-value = 0.1408 > 0.05 so p high null fly accept null hypothesis.

**######## 1-way ANOVA test#############**

?aov

#Ho= Average TAT for all the samples is same

#Ha= Averages TAT for atleast one sample is not same

**Anova\_results12345 <- aov(`Laboratory 1`~`Laboratory 2`+`Laboratory 3`+`Laboratory 4`,data = LabTAT)**

**summary(Anova\_results12345)**

#Laboratory2 p-value = 0.166 > 0.05 so p high null fly accept null hypothesis.

#Laboratory3 p-value = 0.277 > 0.05 so p high null fly accept null hypothesis.

#Laboratory4 p-value = 0.215 > 0.05 so p high null fly accept null hypothesis.

**Python code:**

import pandas as pd

import scipy

from scipy import stats

import statsmodels.api as sm

from statsmodels.formula.api import ols

LabTAT=pd.read\_excel("C:\RAVI\Data science\Assignments\Modue 5 Hypothesis\LabTAT.xlsx")

LabTAT

LabTAT.columns="Laboratory1","Laboratory2","Laboratory3","Laboratory4"

**##########Normality Test ############**

#Ho= Data is normally distributed (no action take)

#Ha=Data is not normally distributed (action take)

**stats.shapiro(LabTAT.Laboratory1)**

# p-value = 0.55 >0.05 so p high null fly => It follows normal distribution

**stats.shapiro(LabTAT.Laboratory2)**

# p-value = 0.86 >0.05 so p high null fly => It follows normal distribution

**stats.shapiro(LabTAT.Laboratory3)**

# p-value = 0.42 >0.05 so p high null fly => It follows normal distribution

**stats.shapiro(LabTAT.Laboratory4)**

# p-value = 0.66 >0.05 so p high null fly => It follows normal distribution

**############## Variance test #########**

#Ho= Variance of TAT of all Labs are equal

#H1=atleast 1 Lab variance of TAT are not equal

**scipy.stats.levene(LabTAT.Laboratory1,LabTAT.Laboratory2)**

#H0=Variance of TAT of Lab 1 is equal to variance of TAT of Lab 2

#H1=Variance of TAT of Lab 1 is not equal to variance of TAT of Lab 2

# p-value = 0.06 > 0.05 so p high null fly accept null hypothesis.

**scipy.stats.levene(LabTAT.Laboratory2,LabTAT.Laboratory3)**

# p-value = 0.33> 0.05 so p high null fly accept null hypothesis.

**scipy.stats.levene(LabTAT.Laboratory3,LabTAT.Laboratory4)**

# p-value = 0.15 > 0.05 so p high null fly accept null hypothesis.

**scipy.stats.levene(LabTAT.Laboratory4,LabTAT.Laboratory1)**

# p-value = 0.221 > 0.05 so p high null fly accept null hypothesis.

**############# One - Way Anova###################**

#Ho= Average TAT for all the samples is same

#Ha= Averages TAT for all the samples is not same

**mod = ols('Laboratory1~Laboratory2+Laboratory3+Laboratory4', data=LabTAT).fit()**

**aov\_table=sm.stats.anova\_lm(mod, type=2)**

**print(aov\_table)**

#Laboratory2 p-value = 0.166299 > 0.05 so p high null fly accept null hypothesis.

#Laboratory3 p-value = 0.277335 > 0.05 so p high null fly accept null hypothesis.

#Laboratory4 p-value = 0.215323 > 0.05 so p high null fly accept null hypothesis.