A F&B manager wants to determine whether there is any significant difference in the diameter of the cutlet between two units. A randomly selected sample of cutlets was collected from both units and measured? Analyze the data and draw inferences at 5% significance level. Please state the assumptions and tests that you carried out to check validity of the assumptions.

File : **Cutlets.csv**

**Business Problem:** Find out difference in the diameter of the cutlet between two units.

Inputs are Unit A and Unit B that is **Discrete in two categories,** Outputis **continuous.** as we are trying to see the difference in diameter of the cutlet between two units and diameter is a continuous variable.

We proceed with **2-sample t test**

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

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

🡪Create hypothesis for **Unit A**

Ho= Data is Normally distributed

Ha=Data is not Normally distributed

**#p-value = 0.32 >0.05 so p high null fly => It follows normal distribution**

**🡪** Create hypothesis for **Unit B**

Ho= Data is Normally distributed

Ha=Data is not Normally distributed

**# p-value = 0.5225 >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 **Unit A** and **Unit B**

Ho= Variance of diameters of Unit A is equal to the variance of diameters of Unit B

Ha= Variance of diameters of Unit A is not equal to the variance of diameters of Unit B

**# p-value = 0.3136 > 0.05 so p high null fly => Equal variances accept null hypothesis.**

As **Unit A** and **Unit B** are 2 Discrete variables and output variable diameter is a continuous, we will go with **2-sample T test.**

**3) 2-sample T test.**

**🡪** Create hypothesis for **Unit A** and **Unit B**

Ho= Averages of diameters of Unit A is equal to Averages of diameters of unit B

Ha= Averages of diameters of Unit A is not equal to Averages of diameters of unit B

So we proceed with **t.test**

**# p-value = 0.4723 > 0.05 p high null fly accept null Hypothesis.**

**Inferences:**

🡪Inference is that there is no significant difference in the diameters of Unit A and Unit B.

**R-Code:**

**# Load the Dataset**

install.packages("readxl")

library(readxl)

**######## Cutlets.xlsx data ##########**

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

View(Cutlets)

attach(Cutlets)

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

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

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

shapiro.test(`Unit A`)

**# p-value = 0.32 >0.05 so p high null fly => It follows normal distribution**

shapiro.test(`Unit B`)

**# p-value = 0.5225 >0.05 so p high null fly => It follows normal distribution**

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

var.test(`Unit A`,`Unit B`)

**# p-value = 0.3136 > 0.05 so p high null fly => Equal variances accept null hypothesis.**

**#Ho= Variance of diameters of Unit A is equal to the variance of diameters of Unit B**

**#Ha= Variance of diameters of Unit A is not equal to the variance of diameters of Unit B**

**############2 sample t Test ###############**

?t.test()

t.test(`Unit A`,`Unit B`, alternative = "two.sided",conf.level = 0.95)

**# p-value = 0.4723 > 0.05 p high null fly accept null Hypothesis**

**# alternative = "two.sided" means we are checking for equal and unequal**

**#Ho= Averages of diameters of Unit A is equal to Averages of diameters of unit B**

**#Ha= Averages of diameters of Unit A is not equal to Averages of diameters of unit B.**

**Python code:**

import pandas as pd

import scipy

from scipy import stats

import statsmodels.api as sm

**############2 sample T Test(Cutlets) ##################**

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

Cutlets

Cutlets.columns="UnitA","UnitB"

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

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

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

Cutlets=Cutlets.iloc[0:35,]

stats.shapiro(Cutlets.UnitA)

**# p-value = 0.319 >0.05 so p high null fly => It follows normal distribution**

print(stats.shapiro(Cutlets.UnitB))

**# p-value = 0.522 >0.05 so p high null fly => It follows normal distribution**

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

**#Ho= Variance of diameters of Unit A is equal to the variance of diameters of Unit B**

**#Ha= Variance of diameters of Unit A is not equal to the variance of diameters of Unit B**

scipy.stats.levene(Cutlets.UnitA,Cutlets.UnitB)

**# p-value = 0.41 > 0.05 so p high null fly => Equal variances accept null hypothesis.**

**######## 2 Sample T test ################**

**#Ho= Averages of diameters of Unit A is equal to Averages of diameters of unit B**

**#Ha= Averages of diameters of Unit A is not equal to Averages of diameters of unit B**

scipy.stats.ttest\_ind(Cutlets.UnitA,Cutlets.UnitB)

**# p-value = 0.472 > 0.05 p high null fly accept null Hypothesis**