**Assignment 3**

1. 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.

Minitab File : Cutlets.mtw

🡪Step1= Assume Null hyposthesis as Ho: μ1 = μ2 (There is no difference in diameters of cutlets between two units).

Thus Alternate hypothesis as Ha: μ1 ≠ μ2 (There is significant difference in diameters of cutlets between two units) 2 Sample 2 Tail test applicable

### Step2=Decide a cut-off value

* Significance 5%
* alpha = 0.05

As it is a two-tailed test

* alpha/2 = 0.025

Step3=Import Files

data=pd.read\_csv('/content/Cutlets.csv')

data.head()

unitA=pd.Series(data.iloc[:,0])

unitA

unitB=pd.Series(data.iloc[:,1])

unitB

### Step4= Compare Evidences with Hypothesis using t-statistics

p\_value=stats.ttest\_ind(unitA,unitB)

p\_value

Ttest\_indResult(statistic=0.7228688704678063, pvalue=0.4722394724599501)

p\_value[1]

0.4722

**Step 6=**

Compare p\_value with 'Alpha '(Significane Level)

### If p\_value is ≠ 'α' we failed to reject Null Hypothesis because of lack of evidence

### If p\_value is = 'α' we reject Null Hypothesis

Hence, We fail to reject Null Hypothesis because of lack of evidence, there is no significant difference between the two samples.

2. 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.

Minitab File: **LabTAT.mtw**

## **🡪**We are going to conduct a ANOVA Test on 4 Independent samples with Numerical Data

### We need to check whether the mean of any of these samples are different or the same?

## Step 1

#### Make two Hypothesis one contradicting to other

#### Null Hypothesis is want we want to prove

* **Null Hypothesis:** μ1=μ2=μ3=μ4
* **Alternative Hypthosis:** Atleast One of them is Differente

Step 2

Decide a cut-off value

* Significance 5%
* α=0.05

Step 3

Import data set

data=pd.read\_csv('/content/LabTAT.csv')

data.head()

plt.subplots(figsize = (16,9))

plt.subplot(221)

plt.hist(data['Laboratory 1'])

plt.title('Laboratory 1')

plt.subplot(222)

plt.hist(data['Laboratory 2'])

plt.title('Laboratory 2')

plt.subplot(223)

plt.hist(data['Laboratory 3'])

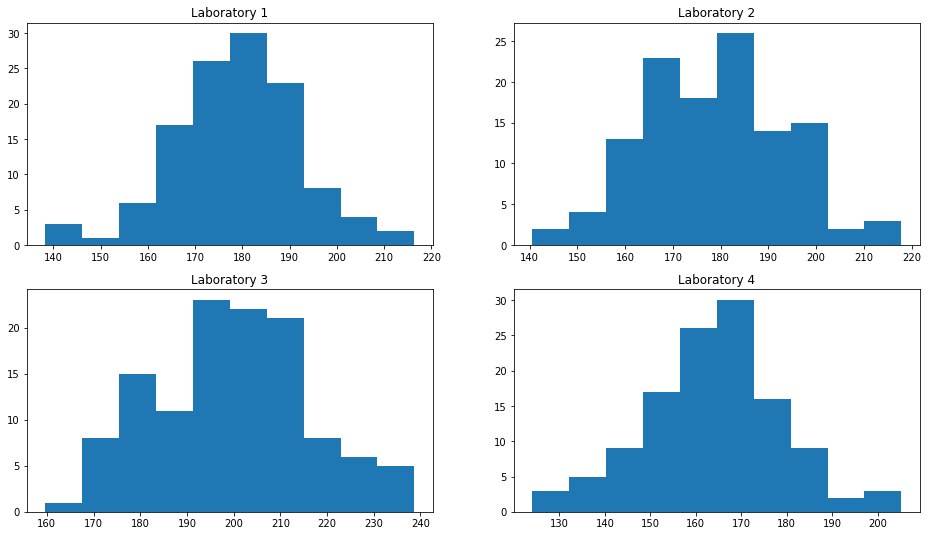
plt.title('Laboratory 3')

plt.subplot(224)

plt.hist(data['Laboratory 4'])

plt.title('Laboratory 4')

plt.show()



# Anova ftest statistics: stats.f\_oneway(column-1,column-2,column-3,column-4)

p\_value=stats.f\_oneway(data.iloc[:,0],data.iloc[:,1],data.iloc[:,2],data.iloc[:,3])

p\_value=2.1156e-57

### Compare p\_value with 'α'(Significane Level)

### If p\_value is ≠ 'α ' we failed to reject Null Hypothesis because of lack of evidence

### If p\_value is = 'α ' we reject Null Hypothesis

alpha = 0.05

print('Significnace=%.3f, p=%.3f' % (alpha, p\_value[1]))

if p\_value[1] <= alpha:

    print('We reject Null Hypothesis there is a significance difference between TAT of reports of the laboratories')

else:

    print('We fail to reject Null hypothesis')

Significnace=0.050, p=0.000

We reject Null Hypothesis there is a significance difference between TAT of reports of the laboratories

**Hence, We fail to reject Null Hypothesis because of lack evidence, there is no significant difference between the samples**

3. Sales of products in four different regions is tabulated for males and females. Find if male-female buyer rations are similar across regions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **East** | **West** | **North** | **South** |
| Males | 50 | 142 | 131 | 70 |
| Females | 550 | 351 | 480 | 350 |

1. Check p-value
2. If p-Value < alpha, we reject Null Hypothesis

* All proportions are equal

H0

* Not all Proportions are equal

Ha

r4

Buyer Ratio.mtw

🡪

Assume Null Hypothesis as Ho: Independence of categorical variables (male-female buyer rations are similar across regions (does not vary and are not related) Thus Alternate Hypothesis as Ha: Dependence of categorical variables (male-female buyer rations are NOT similar across regions (does vary and somewhat/significantly related)

data=pd.read\_csv('/content/BuyerRatio.csv')

data

# Make dimensional array

obs=np.array([[50,142,131,70],[435,1523,1356,750]])

obs

array([[ 50, 142, 131, 70], [ 435, 1523, 1356, 750]])

# Chi2 contengency independence test

chi2\_contingency(obs) # o/p is (Chi2 stats value, p\_value, df, expected obsvations)

(1.595945538661058, 0.6603094907091882, 3, array([[ 42.76531299, 146.81287862, 131.11756787, 72.30424052], [ 442.23468701, 1518.18712138, 1355.88243213, 747.69575948]]))

p value=0.66030

**Inference: As (p-value = 0.6603) > (α = 0.05); Accept the Null Hypothesis i.e. Independence of categorical variables Thus, male-female buyer rations are similar across regions and are not related**

4. TeleCall uses 4 centers around the globe to process customer order forms. They audit a certain % of the customer order forms. Any error in order form renders it defective and has to be reworked before processing. The manager wants to check whether the defective % varies by centre. Please analyze the data at *5%* significance level and help the manager draw appropriate inferences

Minitab File: **CustomerOrderForm.mtw**

🡪Assume Null Hypothesis as Ho: Independence of categorical variables (customer order forms defective % does not varies by centre) Thus, Alternative hypothesis as Ha Dependence of categorical variables (customer order forms defective % varies by centre)

data=pd.read\_csv('/content/Costomer+OrderForm.csv')

data.head()

data.describe()

data.info()

data.isnull().sum()

data.Phillippines.value\_counts()

data.Indonesia.value\_counts()

data.Malta.value\_counts()

data.India.value\_counts()

# Make a contingency table

obs=np.array([[271,267,269,280],[29,33,31,20]])

obs

# Chi2 contengency independence test

chi2\_contingency(obs) # o/p is (Chi2 stats value, p\_value, df, expected obsvations)

(3.858960685820355, 0.2771020991233135, 3, array([[271.75, 271.75, 271.75, 271.75], [ 28.25, 28.25, 28.25, 28.25]]))

**Inference: As (p\_value = 0.2771) > (α = 0.05); Accept Null Hypthesis i.e. Independence of categorical variables Thus, customer order forms defective % does not varies by centre.e=0.2771**