**Hypothesis Testing Exercise**

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.

We will use 2 sample 2 tail t test

H0 = Diameter of the cutlets between 2 units is same

HA = Diameter of the cutlets between 2 units is different

α=0.05

df=pd.read\_csv('cutlets.csv')

dfA = df["Unit A"]

dfB = df["Unit B"]

tStat,pValue =stats.ttest\_ind(dfA,dfB)

**P-Value:0.4722394724599501 T-Statistic:0.7228688704678061**

if pValue <0.05:

print('we reject null hypothesis')

else:

print('we accept null hypothesis')

S**o result came as 'we accept null hypothesis i.e. Diameter of the cutlets between 2 units is same, no significance difference**

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.

4 samples are given.

α=0.05

H0 = TAT is same for all labs

HA = TAT is different for 4 labs.

We will use Anova test

-----------------

dfMed = pd.read\_csv('LabTAT.csv')

dfLab1 = dfMed["Laboratory 1"]

dfLab2= dfMed["Laboratory 2"]

dfLab3 = dfMed["Laboratory 3"]

dfLab4 = dfMed["Laboratory 4"]

tStat, pValue = stats.f\_oneway(dfLab1,dfLab2,dfLab3,dfLab4)

---------------

p value is 2.1156

as p value > 0.05, we accept null hypothesis, **so TAT is same for all labs, no significance difference**

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 | 435 | 1523 | 1356 | 750 |

4 samples are given

H0 = Ratios are same

HA = Ratios are different

import scipy as sp

Male = [50,142,131,70]

Female=[435,1523,1356,750]

Sales=[Male,Female]

chiStats = sp.stats.chi2\_contingency(Sales)

print('Test t=%f p-value=%f' % (chiStats[0], chiStats[1]))

Test t=1.595946 p-value=0.660309

P value > 0.05

**So accept H0 -> Male female ratios are same all regions**

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

Here both centers and output are discrete.

So we want to check if the % varies by centers.

We will sue Chi square test

H0 = Defective % are same for all centers

HA = Defective % varies

We will take value counts for error free and defective

Phillippines\_value=Customer['Phillippines'].value\_counts()

Indonesia\_value=Customer['Indonesia'].value\_counts()

Malta\_value=Customer['Malta'].value\_counts()

India\_value=Customer['India'].value\_counts()

import scipy as sp

Error\_free = [Phillippines\_value['Error Free'],Indonesia\_value['Error Free'],Malta\_value['Error Free'],India\_value['Error Free']]

Defective = [Phillippines\_value['Defective'],Indonesia\_value['Defective'],Malta\_value['Defective'],India\_value['Defective']]

chiStats = sp.stats.chi2\_contingency([Error\_free,Defective])

print('Test t=%f p-value=%f' % (chiStats[0], chiStats[1]))

Test t=3.858961 p-value=0.277102

P value > 0.05

**So accept H0 -> Defective % are same for all centers**