Hypothesis Testing Exercise

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**

**Ans) Given, significance level = 5%**

**Alternate Hypothesis [HA] =µ1 ≠ µ2**

**Null Hypothesis [Ho] = µ1= µ2**

**Alpha = 0.05 , because it is two tailed alpha/2 = 0.05/2**

**=0.025**

**Conditions: If the P value is< alpha reject the Null hypothesis**

**If the P value > alpha accept the Null Hypothesis. i.e no difference in the both units measured.**

**If the P value< alpha reject the Null Hypothesis. I.e both units are different**

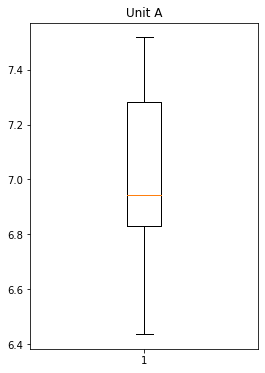
plt.subplots (figsize =(9,6))                          # box plot for Unit A

plt.subplot (121)

plt.boxplot(df['Unit A'])

plt.title ('Unit A')

plt.show()



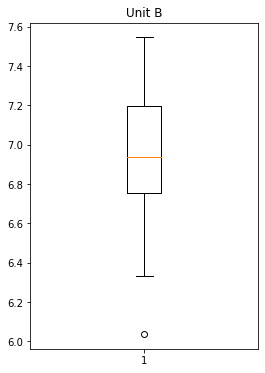
plt.subplots (figsize = (9,6))

plt.subplot (122)

plt. boxplot (df['Unit B'])                          # Boxplot for Unit B

plt.title ('Unit B')

plt.show()



import statsmodels.api as sm

sm.stats.ttest\_ind(df['Unit A'],df['Unit B'])                   # tvalue, Pvalue and degree of freedom

(0.7228688704678063, 0.4722394724599501, 68.0)

p\_value= 0.4722394724599501

alpha =0.025

alpha = 0.025

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

if p\_value <= alpha:

    print('We reject Null Hypothesis there is a significance difference between two Units A and B')

else:

    print('We fail to reject Null hypothesis')

Significnace=0.025, p=0.472

We fail to reject Null hypothesis

CodeText

So, we accept the null hypothesis that the untis are same and there is no difference between them.

1. 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**

Ans) Null Hypothesis = µ1 = µ2 = µ3 = µ4

Alternative hypothesis = there is some difference between the average of the four test report.

Significance = 5%

Alpha = 0.05%

import pandas as pd

import numpy as np

import seaborn as sns

from scipy import stats

import matplotlib.pyplot as plt

from google.colab import files

uploaded=files.upload()

df = pd.read\_csv('LabTAT.csv')

df

df.isnull().sum()

plt.subplots(figsize = (17,7))

plt.subplot(221)

plt.boxplot(df['Laboratory 1'])

plt.title('Laboratory 1')

plt.subplot(222)

plt.boxplot(df['Laboratory 2'])

plt.title('Laboratory 2')

plt.subplot(223)

plt.boxplot(df['Laboratory 3'])

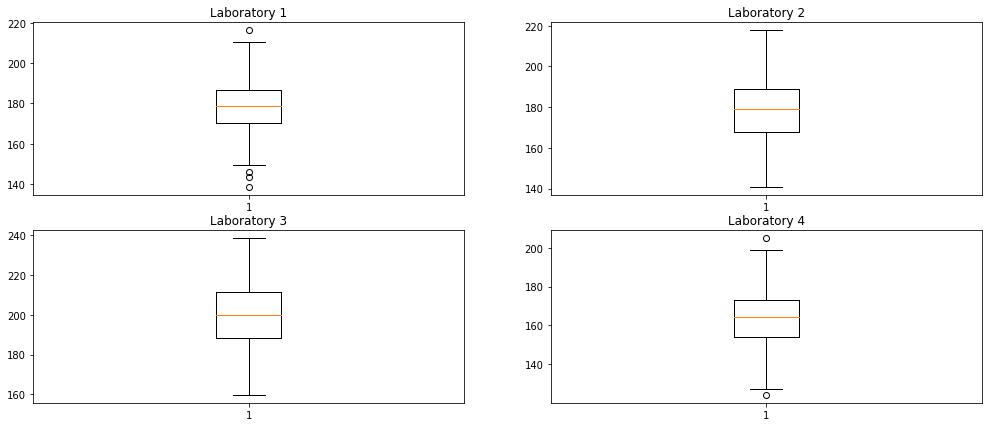
plt.title('Laboratory 3')

plt.subplot(224)

plt.boxplot(df['Laboratory 4'])

plt.title('Laboratory 4')

plt.show()



labtat = stats.f\_oneway(df.iloc[:,0],df.iloc[:,1],df.iloc[:,2],df.iloc[:,3])

labtat

F\_onewayResult(statistic=118.70421654401437, pvalue=2.1156708949992414e-57)

alpha = 0.05

if p\_value <= 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')

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

S0,their a difference among the average of the reports.

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

Ans)Given the buyers ration of male and female across regions.

Null Hypothesis = There is no dependency between the male and female ration.

Alternate Hypothesis = There is a dependency between the male and female ration.

import pandas as pd

import numpy as np

from scipy import stats

from google.colab import files

uploaded = files.upload()

Ratio = pd.read\_csv('BuyerRatio.csv')

table = [[50,142,131,70],                                                       # contigency table

        [435,1523,1356,750]]

stat, p, dof, exp = stats.chi2\_contingency(table)                         # applying chi square converting ‘O’ into ‘E’

print(stat,"\n", p,"\n", dof,"\n", exp)

1.595945538661058

0.6603094907091882

3

[[ 42.76531299 146.81287862 131.11756787 72.30424052]

[ 442.23468701 1518.18712138 1355.88243213 747.69575948]]

stats.chi2\_contingency(table

(1.595945538661058,

0.6603094907091882,

3,

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

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

Expected = np.array ([  42.76531299,  146.81287862,  131.11756787,   72.30424052, 442.23468701, 1518.18712138, 1355.88243213,  747.69575948])

statistics, p\_value = stats.chisquare(observed, Expected, ddof = 3)             #Compare the value

print("Statistics = ",statistics,"\n",'P\_Value = ', p\_value)

Statistics = 1.5959455390914483

P\_Value = 0.8095206646905712

alpha = 0.05

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

if p\_value <= alpha:

    print('We reject Null Hypothesis that all portions are equal')

else:

    print('We fail to reject Null hypothesis')

Significnace=0.050, p=0.810

We fail to reject Null hypothesis

CodeText

So, we fail to reject the null hypothesis that means there is no dependency between male and female ration they are similar across the regions. The sample are independent.

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**

**Ans)** Given the data of the 4 Tele call centers around the globe.

Null Hypothesis = All the customer orders form are same from the centre.

Alternate Hypothesis = Atleast one is different.

import pandas as pd

import numpy as np

from scipy import stats

from google.colab import files

uploaded = files.upload()

df = pd.read\_csv('Costomer+OrderForm.csv')

df.isnull().any()                                                               #checking the null values

Phillippines False

Indonesia False

Malta False

India False

dtype: bool

df.info()                                                                        # information of the data type

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 300 entries, 0 to 299

Data columns (total 4 columns):

# Column Non-Null Count Dtype

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

0 Phillippines 300 non-null object

1 Indonesia 300 non-null object

2 Malta 300 non-null object

3 India 300 non-null object

dtypes: object(4)

memory usage: 9.5+ KB

print(df['Phillippines'].value\_counts(),'\n',df['Indonesia'].value\_counts(),'\n',df['Malta'].value\_counts(),'\n',df['India'].value\_counts)       #value counts

Error Free 271

Defective 29

Name: Phillippines, dtype: int64

Error Free 267

Defective 33

Name: Indonesia, dtype: int64

Error Free 269

Defective 31

Name: Malta, dtype: int64

<bound method IndexOpsMixin.value\_counts of 0 Error Free

1 Defective

2 Error Free

3 Error Free

4 Error Free

...

295 Error Free

296 Error Free

297 Error Free

298 Error Free

299 Error Free

Name: India, Length: 300, dtype: object>

contingency\_table = [[271,267,269,280], # contingency table

                    [29,33,31,20]]

print(contingency\_table)

[[271, 267, 269, 280], [29, 33, 31, 20]]

stat, p, df, exp = stats.chi2\_contingency(contingency\_table)                         # observed and expected value

print("Statistics = ",stat,"\n",'P\_Value = ', p,'\n', 'degree of freedom =', df,'\n', 'Expected Values = ', exp)

Statistics = 3.858960685820355

P\_Value = 0.2771020991233135

degree of freedom = 3

Expected Values = [[271.75 271.75 271.75 271.75]

[ 28.25 28.25 28.25 28.25]]

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

expected = np.array([271.75, 271.75, 271.75, 271.75, 28.25, 28.25, 28.25, 28.25])

test\_statistic , p\_value = stats.chisquare(observed, expected, ddof = df)

print("Test Statistic = ",test\_statistic,'\n', 'p\_value =',p\_value)                            #t value and P value

Test Statistic = 3.858960685820355

p\_value = 0.4254298144535761

alpha = 0.05

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

if p\_value <= alpha:

    print('We reject Null Hypothesis that all are equal')

else:

    print('We fail to reject Null hypothesis')

Significnace=0.050, p=0.425

We fail to reject Null hypothesis

so, we fail to reject the null hypothesis it means all are the same.