**Hypothesis Testing Exercise**

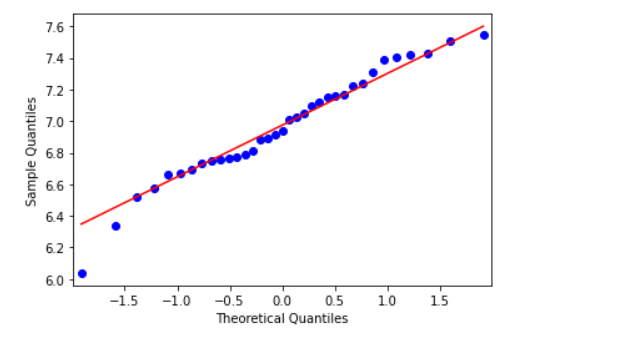
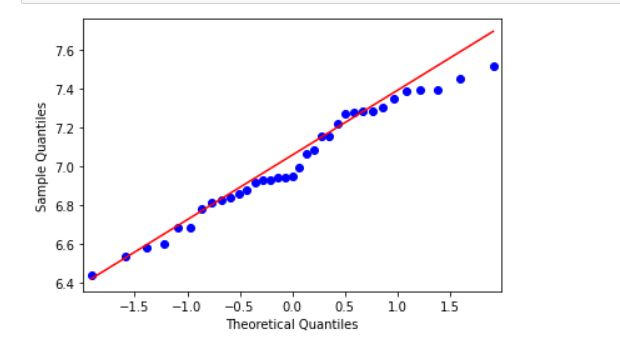
Q1) A F&B manager wants to determine whether there is any significant difference in the diameter of the cutlet between two units. A randomly selected sample of cutlets was collected from both units and measured? Analyze the data and draw inferences at 5% significance level. Please state the assumptions and tests that you carried out to check validity of the assumptions.

Ans:

Consider H0:The is no significant difference between two units

H1:The is significant difference between two units

To check if the data is following normal distribution we will need to draw a q-q plot



Unit A Unit B

As majority of the data is falling on the line we can say that the data is Normally distributed

Python Code-

import pandas as pd

import scipy

import numpy as np

from scipy import stats

import statsmodels.api as st

import matplotlib.pyplot as plt

data=pd.read\_csv("C:\\Users\\advay\\Downloads\\Assignment 3\\Cutlets.csv")

st.qqplot(data['Unit A'],line='q')

st.qqplot(data['Unit B'],line='q')

plt.show()

array1=data['Unit A'].values

array2=data['Unit B'].values

stats.ttest\_ind(array1,array2)

Result- test statistic=0.7228688704678063, pvalue=0.4722394724599501

Conclusion-As p-value is less than level of significance we accept H1, i.e. There is a significant difference between

two units

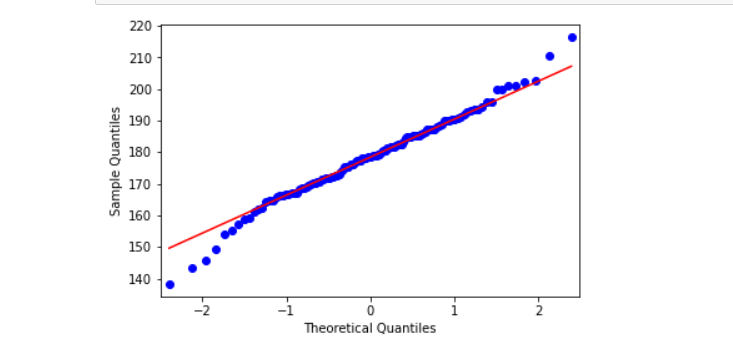
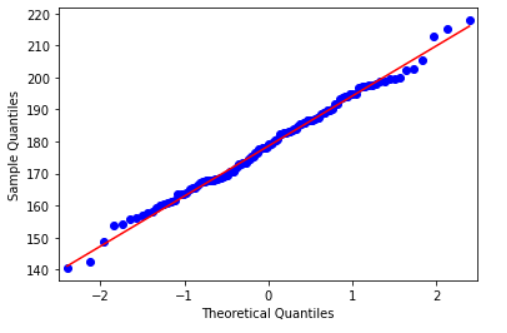
Q2)A hospital wants to determine whether there is any difference in the average Turn Around Time (TAT) of reports of the laboratories on their preferred list. They collected a random sample and recorded TAT for reports of 4 laboratories. TAT is defined as sample collected to report dispatch.Analyze the data and determine whether there is any difference in average TAT among the different laboratories at 5% significance level.

Ans-

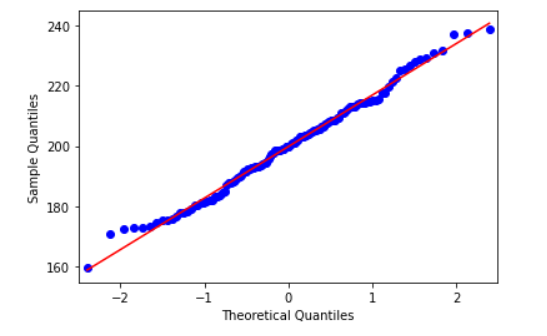
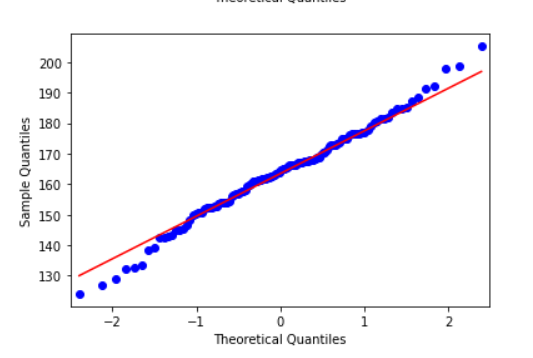
Consider H0:There is no significant difference in the average TAT among the different laboratories

  H1:There is a significant difference in the average TAT among the different laboratories

Q-Q Plot:-

Laboratory 1 Laboratory 2

Laboratory 3 Laboratory 4

As majority of the data is falling on the line we can say that the data is Normally distributed

Python Code-

*import pandas as pd*

*import scipy*

*import numpy as np*

*from scipy import stats*

*import statsmodels.api as st*

*import matplotlib.pyplot as plt*

*data=pd.read\_csv("C:\\Users\\advay\\Downloads\\Assignment 3\\LabTAT.csv")*

*st.qqplot(data['Laboratory 1'],line='q')*

*st.qqplot(data['Laboratory 2'],line='q')*

*st.qqplot(data['Laboratory 3'],line='q')*

*st.qqplot(data['Laboratory 4'],line='q')*

*plt.show()*

*scipy.stats.f\_oneway(data.iloc[:,0],data.iloc[:,1],data.iloc[:,2],data.iloc[:,3])*

Result- statistic=118.70421654401437, pvalue=2.1156708949992414e-57

Conclusion-As the p-value is less than the level of significance we accept H1 i.e. There is a significant difference in the average TAT among the different laboratories

Q3) 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 |

Consider

H0:Male-Female buyer ratios are similar across regions

H1:Male-Female buyer ratios are not similar across regions

Python Code-

*import pandas as pd*

*import scipy*

*import numpy as np*

*from scipy import stats*

*Buyer=pd.read\_csv("C:\\Users\\advay\\Downloads\\Assignment 3\\BuyerRatio.csv")*

*data=Buyer.iloc[:,1:5]*

*stats.chi2\_contingency(data)*

Result- Test Statistic-1.595945538661058 ,pvalue-0.6603094907091882

Conclusion- As p-value is more than level of significance(5%) we accept H0, i.e. Male-Female buyer ratios are similar across regions

Q4)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

H0-All the centres have same defective %

H1:All the centres have different defective %

Python Code-

*import pandas as pd*

*import scipy*

*import numpy as np*

*from scipy import stats*

*import statsmodels.api as st*

*data=pd.read\_csv("C:\\Users\\advay\\Downloads\\Assignment 3\\Costomer+OrderForm.csv")*

*data['Phillippines'].value\_counts()*

*data['Indonesia'].value\_counts()*

*data['India'].value\_counts()*

*data['Malta'].value\_counts()*

*data={'Phillippines':[279,21],'Indonesia':[267,33],'Malta':[269,31],'India':[280,20]}*

*data1=pd.DataFrame(data,index=['Error Free', 'Defective'])*

*stats.chi2\_contingency(data1)*

Result- Test Statistics= 5.6255707762557075, p value=0.1313180510950158

Conclusion- As p-value is more than level of significance(5%) we accept H0, Hence All the centres have same defective %