**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.

Ans:-

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:-

There is no significant difference in the diameter of cutlets between unit A and unit B

Test:- Two Sample T-test

Code:-

# Two sample Z test

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

df=pd.read\_csv("Cutlets.csv")

list(df)

sns.histplot(data=df, x='Unit A', kde=True, bins=20, color='skyblue', label='Unit A')

sns.histplot(data=df, x='Unit B', kde=True, bins=20, color='orange', label='Unit B')

plt.title('Distribution of Unit A and Unit B')

plt.legend()

plt.show()

plt.figure(figsize=(10, 6))

sns.scatterplot(x='Unit A', y='Unit B', data=df, color='green')

plt.title('Scatter Plot between Unit A and Unit B')

plt.xlabel('Unit A')

plt.ylabel('Unit B')

plt.show()

correlation\_matrix = df.corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')

plt.title('Correlation Matrix')

plt.show()

sns.boxplot(data=df[['Unit A', 'Unit B']])

plt.title('Box Plot for Unit A and Unit B')

plt.show()

df["Unit A"]

df["Unit B"]

from scipy import stats

zcalc,pval = stats.ttest\_ind(df["Unit A"],df["Unit B"])

zcalc

pval

alpha=0.05

if pval < alpha:

print("Ho is rejected and H1 is accepted")

if pval > alpha:

print("H1 is rejected and Ho is accepted")

Inferences:

H1 is rejected and Ho is accepted,means the null hypothesis is accepted,So there is not

enough evidence to conclude a significant difference in the diameter of cutlets between the two units.

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.

Ans:-

import pandas as pd

from scipy import stats

data=pd.read\_csv("D:\\assignments\\LabTAT.csv")

f\_statistic, p\_value = stats.f\_oneway(data['Laboratory 1'], data['Laboratory 2'], data['Laboratory 3'], data['Laboratory 4'])

print(f"Anova ftest statistics: F = {f\_statistic}, p-value = {p\_value}")

alpha = 0.05

if p\_value < alpha:

print("Inference: Reject Null Hypothesis")

print("Conclusion: At least one sample TAT population mean is different.")

print("There is variance or difference in average Turn Around Time (TAT) of reports of the laboratories on their preferred list.")

else:

print("Inference: Fail to reject Null Hypothesis")

print("Conclusion: There is no sufficient evidence to conclude a difference in average Turn Around Time (TAT) among the laboratories.")

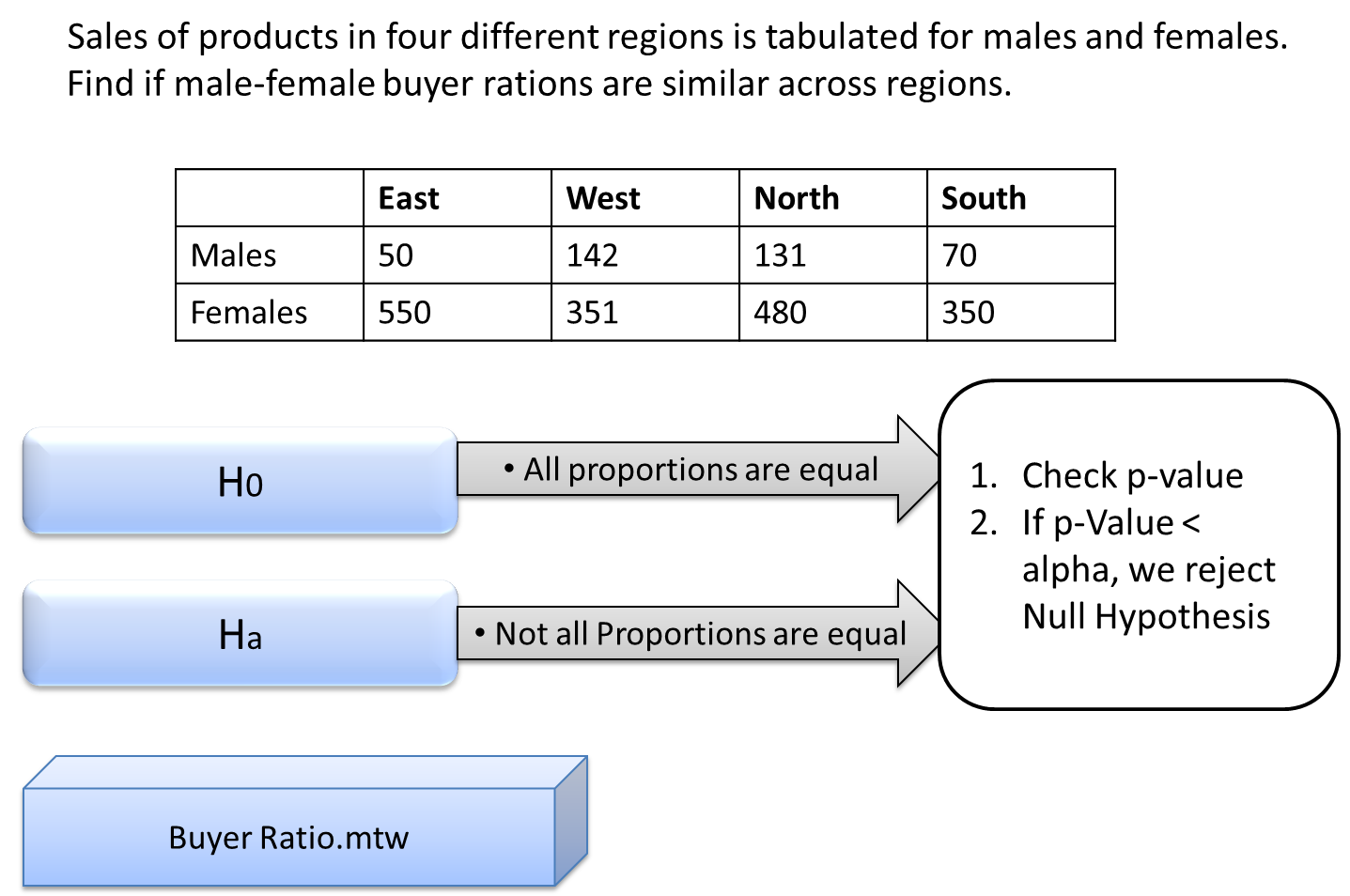
OUTPUT:-

Anova ftest statistics: F = 118.70421654401437, p-value = 2.1156708949992414e-57

Inference: Reject Null Hypothesis

Conclusion: At least one sample TAT population mean is different.

There is variance or difference in average Turn Around Time (TAT) of reports of the laboratories on their preferred list.



Ans:-

import pandas as pd

from scipy import stats

from scipy.stats import chi2

df = pd.read\_csv("BuyerRatio.csv")

df\_table = df.iloc[:, 1:6]

chi2\_stat, p\_value, dof, expected\_value = stats.chi2\_contingency(df\_table.values)

print(f"P-value: {p\_value}")

critical\_value = chi2.ppf(0.95, dof)

print(f"Critical value: {critical\_value}")

p\_value\_chi2 = 1 - chi2.cdf(chi2\_stat, dof)

if p\_value\_chi2 <= 0.05:

print('All proportions are equal')

else:

print('Not all proportions are equal')

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

Ans:-

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from scipy.stats import chi2\_contingency

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

list(df)

plt.figure(figsize=(10, 6))

sns.countplot(data=df.melt(), x='variable', hue='value', palette='Set2')

plt.title('Order Status in Each Country')

plt.xlabel('Country')

plt.ylabel('Count')

plt.show()

columns = ['Phillippines', 'Indonesia', 'Malta', 'India']

for i in range(len(columns)):

for j in range(i+1, len(columns)):

contingency\_table = pd.crosstab(df[columns[i]], df[columns[j]])

print(f"Contingency Table for {columns[i]} and {columns[j]}:")

print(contingency\_table)

chi2, p, dof, expected = chi2\_contingency(contingency\_table)

print(f"Chi-Square: {chi2}")

print(f"P-value: {p}")

print(f"Degrees of Freedom: {dof}")

alpha = 0.05

if p < alpha:

print("Reject the null hypothesis(Ho),means there is a evidence of significant difference in defective percentages.")

else:

print("Fail to reject the null hypothesis(Ho),means there is no significant difference in defective percentages.")