**StatisticalTests Assignment:-**

**Data Loading and Treatment:**

In [**58**]: import pandas as pd

In [**59**]: dataset = pd.read\_excel('general\_data.xlsx',sheet\_name = 'general\_data')

In [**60**]: dataset.isnull()

Out[**60**]:

Age Attrition ... YearsSinceLastPromotion YearsWithCurrManager

0 False False ... False False

1 False False ... False False

2 False False ... False False

3 False False ... False False

4 False False ... False False

... ... ... ... ... ...

4405 False False ... False False

4406 False False ... False False

4407 False False ... False False

4408 False False ... False False

4409 False False ... False False

[4410 rows x 24 columns]

In [**61**]: dataset.duplicated()

Out[**61**]:

0 False

1 False

2 False

3 False

4 False

...

4405 False

4406 False

4407 False

4408 False

4409 False

Length: 4410, dtype: bool

In [**62**]: dataset1 = dataset.drop\_duplicates()

In [**63**]: dataset2 = dataset1.dropna()

In [**64**]: dataset2.columns

Out[**64**]:

Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',

'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',

'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',

'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',

'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],

dtype='object')

**WILCOXON TEST:**

H0 - There is no significant difference the in the years the employee has worked with manager and for the company.

H1 - There is a significant difference the in the years the employee has worked with manager and for the company.

In [**65**]: from scipy.stats import wilcoxon

In [**66**]: stats, p = wilcoxon(dataset2.YearsWithCurrManager,dataset2.YearsAtCompany)

In [**67**]: print(stats,p)

0.0 0.0

Inference:

P value is less than 0.05, hence Alernate hypothesis is accepted.

There is a significant difference the in the years the employee has worked with manager and for the company.

**FRIEDMAN TEST:**

H0 - There is no significant difference the in the years the employee has worked with manager and for the company and the total years worked.

H1 - There is a significant difference the in the years the employee has worked with manager and for the company and the total years worked.

In [**69**]: stats,p= friedmanchisquare(dataset2.YearsWithCurrManager,dataset2.YearsAtCompany,dataset2.TotalWorkingYears)

In [**70**]: print(stats,p)

7600.426186058835 0.0

Inference:

P value is less than 0.05, hence Alernate hypothesis is accepted.

There is a significant difference the in the years the employee has worked with manager and for the company and the total years worked.

**MANN -WHITNEY TEST:**

H0 - There is no significant difference the in the monthlyIncome for males and females.

H1 - There is a significant difference the in the monthlyIncome for males and females.

In [**75**]: dataset3 = pd.read\_excel('general\_data.xlsx',sheet\_name = 1)

In [**76**]: from scipy.stats import mannwhitneyu

In [**77**]: stats,p =mannwhitneyu(dataset3.Male,dataset3.Female)

In [**78**]: print(stats,p)

55.0 0.1703893067814773

Inference:

P value is more than 0.05, hence Null hypothesis is accepted.

There is no significant difference the in the monthlyIncome for males and females.

**KRUSHKAL WALLIS TEST:**

H0 - There is no significant difference the in the monthlyIncome for managers, HR and Sales Executive.

H1 - There is a significant difference the in the monthlyIncome for managers, HR and Sales Executive.

In [**79**]: dataset4 = pd.read\_excel('general\_data.xlsx',sheet\_name = 2)

In [**80**]: from scipy.stats import kruskal

In [**81**]: stats,p =kruskal(dataset4.Manager,dataset4.HR,dataset4.SalesExec)

In [**82**]: print(stats,p)

1.9723833757421623 0.3729944689982891

In [**83**]: dataset4

Out[**83**]:

Manager HR SalesExec

0 68700 21050 193280

1 162910 51260 23420

2 44480 30720 58130

3 97130 35640 31430

4 65000 68040 55380

5 81030 66960 41270

6 28630 30720 54050

7 55610 190450 46840

8 28970 14200 23350

Inference:

P value is more than 0.05, hence Null hypothesis is accepted.

There is no significant difference the in the monthlyIncome for managers, HR and Sales Executive.

**CHI-SQUARE TEST:**

H0 - There is no dependency between gender and businessTravel

H1 - There is dependency between gender and businessTravel

In [**84**]: from scipy.stats import chi2\_contingency

chitable1 = pd.crosstab(dataset2.Gender,dataset2.BusinessTravel)

In [89]: chitable1

Out[89]:

BusinessTravel Non-Travel Travel\_Frequently Travel\_Rarely

Gender

Female 153 328 1275

Male 295 497 1834

stats,p,dof,expected = chi2\_contingency(chitable1)

print(stats,p)

7.7114689255919355 0.021158057795728463

Inference:

P value is less than 0.05, hence Alernate hypothesis is accepted.

There is dependency between gender and businessTravel