Day -10 Assignments

Statistical Test – Mann-whitney

import pandas as pd

data=pd.read\_excel("general\_data.xls",sheet\_name=1)

data.head()

Out[3]:

DistanceFromHome\_YES ... YearsWithCurrManager\_NO

0 10.0 ... 0

1 11.0 ... 3

2 1.0 ... 5

3 1.0 ... 4

4 4.0 ... 7

[5 rows x 14 columns]

**Attrition vs DistanceFromHome**

from scipy.stats import mannwhitneyu

a1=data.DistanceFromHome\_YES

a2=data.DistanceFromHome\_NO

stats,p=mannwhitneyu(a1,a2)

print(stats,p)

1317879.0 0.0

Value of p is less than 0.05 which means null hypothesis is Rejected and alternate hypothesis is accepted

H0- There is NO significant difference between DistanceFromHome and Attrition.

Ha- There is significanct difference between DistanceFromHome and Attrition.

**Attrition vs PercentSalaryHike**

b1=data.PercentSalaryHike\_YES

b2=data.PercentSalaryHike\_NO

stats,p=mannwhitneyu(b1,b2)

print(stats,p)

1250640.0 0.0

Value of p is less than 0.05 which means null hypothesis is Rejected and alternate hypothesis is accepted

H0- There is NO significant difference between PercentSalaryHike and Attrition.

Ha- There is significanct difference between PercentSalaryHike and Attrition.

**Attrition vs MonthlyIncome**

c1=data.MonthlyIncome\_YES

c2=data.MonthlyIncome\_NO

stats,p=mannwhitneyu(c1,c2)

print(stats,p)

1365088.5 0.0

Value of p is less than 0.05 which means null hypothesis is Rejected and alternate hypothesis is accepted

H0- There is NO significant difference between MonthlyIncome and Attrition.

Ha- There is significanct difference between MonthlyIncome and Attrition.

**Attrition vs NumCompaniesWorked**

d1=data.NumCompaniesWorked\_YES

d2=data.NumCompaniesWorked\_NO

stats,p=mannwhitneyu(d1,d2)

print(stats,p)

1279115.5 0.0

Value of p is less than 0.05 which means null hypothesis is Rejected and alternate hypothesis is accepted

H0- There is NO significant difference between NumCompaniesWorked and Attrition.

Ha- There is significanct difference between NumCompaniesWorked and Attrition.

**Attrition vs TotalWorkingYears**

e1=data.TotalWorkingYears\_YES

e2=data.TotalWorkingYears\_NO

stats,p=mannwhitneyu(e1,e2)

print(stats,p)

1726497.0 0.0

Value of p is less than 0.05 which means null hypothesis is Rejected and alternate hypothesis is accepted

H0- There is NO significant difference between TotalWorkingYears and Attrition.

Ha- There is significanct difference between TotalWorkingYears and Attrition.

**Attrition vs YearsAtCompany**

f1=data.YearsAtCompany\_YES

f2=data.YearsAtCompany\_NO

stats,p=mannwhitneyu(f1,f2)

print(stats,p)

1706751.0 0.0

Value of p is less than 0.05 which means null hypothesis is Rejected and alternate hypothesis is accepted

H0- There is NO significant difference between YearsAtCompany and Attrition.

Ha- There is significanct difference between YearsAtCompany and Attrition.

**Attrition vs YearsWithCurrManager**

g1=data.YearsWithCurrManager\_YES

g2=data.YearsWithCurrManager\_NO

stats,p=mannwhitneyu(g1,g2)

print(stats,p)

1672735.5 0.0

Value of p is less than 0.05 which means null hypothesis is Rejected and alternate hypothesis is accepted

H0- There is NO significant difference between YearsWithCurrManager and Attrition.

Ha- There is significanct difference between YearsWithCurrManager and Attrition.