Statistical Tests

import numpy as n

import pandas as pd

import matplotlib.pyplot as m

from scipy.stats import pearsonr

from sklearn.preprocessing import LabelEncoder

count=0

ds=pd.read\_csv("general\_data.csv")

ds.dropna()

ds.drop\_duplicates()

le=LabelEncoder()

ds["Attrition"]=le.fit\_transform(ds["Attrition"])

data\_yes=ds[ds["Attrition"]==1]

data\_no=ds[ds["Attrition"]==0]

1. Mann whitneyu Test:

**A.**Attrition vs Distance from home

H0: Distance from home does not effect attrition

Ha: Distance from home does effect attrition

#=============================================================================

from scipy.stats import mannwhitneyu as man

s1=data\_yes.DistanceFromHome

s2=data\_no.DistanceFromHome

stat,p=man(s1,s2)

print("The value of p is",p)

#=============================================================================

**The value of p is 0.4629185205822659**

Since the value of p is >0.05 we accept the null hypothesis

**B.**Attrition vs job level:

H0:job level does not effect attrition

Ha:job level effects attrition

s3=data\_yes.JobLevel

s4=data\_no.JobLevel

stat,p=man(s3,s4)

print(""The value of p is" ,p)

**The value of p is 0.4211326530832555**

Since the value of p is >0.05 we accept the null hypothesis

**C.**Attrition vs Years With Current Manager:

H0: Years with Current Manager doesn’t effect attrition

HA: Years with Current Manager Effect’s attrition

s5=data\_yes.YearsWithCurrManager

s6=data\_no.YearsWithCurrManager

stat,p=man(s6,s5)

print("The value of p is",p)

**The value of p is 1.2365483142169503e-31**

The value of p <0.05 hence we reject the null hypothesis

**2.INDEPENDENT T TEST**

**A.**Attrition vs Distance from home

H0: Distance from home does not effect attrition

Ha: Distance from home does effect attrition s1=data\_yes.DistanceFromHome

s2=data\_no.DistanceFromHome

stat,p=ttest(s1,s2)

print("The value of p is",p)

**The value of p is 0.518286042805572**

Since the value of p is >0.05 we accept the null hypothesis

**B.** Attrition vs job level:

H0:job level does not effect attrition

Ha:job level effects attrition

s3=data\_yes.JobLevel

s4=data\_no.JobLevel

stat,p=ttest(s3,s4)

print("The value of p is",p)

The value of p is 0.4945171727187496

Since the value of p is >0.05 we accept the null hypothesis