Non Parametric Test

Data Cleaning

import pandas as pd

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

dataset.isnull().any(axis=1).sum()

Out[3]: 28

dataset.isnull().any(axis=1).sum()

Out[5]: 28

dataset.drop\_duplicates()

dataCleaned= dataset.dropna()

dataset.isnull().any(axis=1).sum()

Out[8]: 28

dataCleaned.isnull().any(axis=1).sum()

Out[9]: 0

Wilcoxon Test

from scipy.stats import wilcoxon

df =dataCleaned[dataCleaned['YearsAtCompany'] !=0]

dataWil =df[df['TotalWorkingYears'] !=0]

H0 = There is no significance difference between YearsAtCompany and TotalWorkingYears

Ha = There is significance difference between YearsAtCompany and TotalWorkingYears

stats,p =wilcoxon(dataWil.YearsAtCompany,dataWil.TotalWorkingYears)

print(stats,p)

1. 0.0

Null hypothesis is rejected

Mann Whitney Test

H0 = There is no significance difference between Attrition and DistanceFromHome

Ha = There is significance difference between Attrition and DistanceFromHome

from scipy.stats import mannwhitneyu

stats,p = mannwhitneyu(dataCleaned.Attrition,dataCleaned.DistanceFromHome)

print(stats,p)

219255.0 0.0

Null hypothesis is rejected

Chi Square Test

H0 = There is no significance difference between Attrition and Gender

Ha = There is significance difference between Attrition and Gender

from scipy.stats import chi2\_contingency

attrGen = pd.crosstab(dataCleaned.Gender,dataCleaned.Attrition)

attrGen

Out[50]:

Attrition 0 1

Gender

Female 1488 268

Male 2189 437

stats,p,dof,expeted = chi2\_contingency(attrGen)

print(stats,p)

1.3825823839528295 0.23966176275638887

Null Hypothesis is Accepted

Parametric Tests

One Sample T test

from scipy.stats import ttest\_1samp

stats,p = ttest\_1samp(dataCleaned.MonthlyIncome,20000)

print(stats,p)

63.275111750220255 0.0

Null Hypothesis is rejected