Day 12 assignment

These are the codes used to find the short inference based on p in the given data sheet.

import pandas as pd

import matplotlib.pyplot as plt

import scipy.stats as sat

dp = pd.read\_excel("C:\letsupgrage assignment\general data.xlsx", sheet\_name=0)

dp\_yes = pd.read\_excel(r"c:\letsupgrage assignment\attrition\_yes.xlsx", sheet\_name=0)

dp\_no = pd.read\_excel(r"C:\letsupgrage assignment\attrition\_no.xlsx", sheet\_name=0)

#print(dp)

dp.dropna(subset = ["Attrition"], inplace=True)

dp.head()

dp.isnull()

dp.duplicated()

print(dp.drop\_duplicates())

from scipy.stats import wilcoxon

stats, p = wilcoxon(dp.Attrition, dp.MonthlyIncome)

print("the value of two wilcoxon test")

print(stats, p)

from scipy.stats import mannwhitneyu

stats, r = mannwhitneyu(dp\_yes.MonthlyIncome, dp\_no.MonthlyIncome)

print("the value of mann whitney test")

print(stats, r)

from scipy.stats import kruskal

stats, s = kruskal(dp.Attrition, dp.DistanceFromHome)

print("the value of kruskal wallis test")

print(stats,s)

from scipy.stats import friedmanchisquare

stats, q = friedmanchisquare(dp.Attrition, dp.Age, dp.DistanceFromHome)

print("the value of fried man's test")

print(stats, q)

from scipy.stats import chi2\_contingency

chitable = pd.crosstab(dp.Attrition, dp.MonthlyIncome)

stats, t, dof, expeted = chi2\_contingency(chitable)

print("the value of chi square test")

print(stats, t)

from scipy.stats import ttest\_1samp

stats, u = ttest\_1samp(dp.MonthlyIncome, 70000)

print("the value of one sample t-test")

print(stats, u)

from scipy.stats import ttest\_rel

stats, v = ttest\_rel(dp.Attrition, dp.MonthlyIncome)

print("the value of two sample paired t-test")

print(stats, v)

from scipy.stats import ttest\_ind

stats, w = ttest\_ind(dp.MonthlyIncome, dp.TrainingTimesLastYear)

print("the value of two sample independent t-test")

print(stats, w)

**STATEMENT:**

while we are doing this I just done all the eight test in the given data. But the only one test is suited for the given data. That is the mann whitney test. In the mann whitney’s test I just compared the both attrition yes and no data with the monthly income for the workers.

For the,

H0 = the monthly income of the both dataset attrition yes and no not be in normally distributed

Ha = the monthly income of the both dataset attrition yes and no is normally distributed.

**Output:**

the value of mann whitney test

1264900.5 0.053577283839938566

The value of the p is greater than the h0 of the value. So we can accept the statement of the H0.

That is the attrition of yes and no with the monthly income is not normally distributed.