**STATISTICAL LEARNING**

***WILCOXON TEST***

import pandas as pd

from scipy.stats import wilcoxon

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

dataset.head()

dataset['Attrition']=dataset['Attrition'].map({'Yes':1,'No':0})

stats, p=wilcoxon( dataset.Attrition,dataset.PercentSalaryHike)

print(stats,p)

**Outpu**t:

Stats=0.0

P= 0.0

***FriedManTest***

import pandas as pd

from scipy.stats import friedmanchisquare

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

dataset.head()

dataset['Attrition']=dataset['Attrition'].map({'Yes':1,'No':0})

stats, p=friedmanchisquare( dataset.Attrition,dataset.PercentSalaryHike,dataset.MonthlyIncome)

print(stats,p)

**Output:**

Stats : 8820.0

P:0.0

***MannWhiteyTest***

import pandas as pd

from scipy.stats import mannwhitneyu

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

dataset.head()

dataset['Attrition']=dataset['Attrition'].map({'Yes':1,'No':0})

stats, p=mannwhitneyu( dataset.Attrition,dataset.PercentSalaryHike)

print(stats,p)

**Output:**

1. 0.0

***Kruskal Test***

import pandas as pd

from scipy.stats import kruskal

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

dataset.head()

dataset['Attrition']=dataset['Attrition'].map({'Yes':1,'No':0})

stats, p=kruskal( dataset.Attrition,dataset.MonthlyIncome,dataset.PercentSalaryHike)

print(stats,p)

**Output:**

Stats: 7145.04395599951

P= 0.0

**Analysis**

P value is less than 0.005, So we have to reject null hypothesis. Hence Attrition is happening due to Monthly Income & Percentage in salary hike

***Chisquare***

import pandas as pd

from scipy.stats import chi2\_contingency

dataset=pd.read\_excel('5 Chi square Test.xlsx',sheet\_name=0)

dataset.head()

dataset=pd.crosstab(dataset.Gender,dataset.Smoking)

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

print(stats,p)

**Output:**

Stats=3.1712567666931584

P= 2

**Parametric Test**

***One Sample Test***

import pandas as pd

from scipy.stats import ttest\_1samp

dataset=pd.read\_excel('1. One Sample.xlsx',sheet\_name=0)

dataset.head()

stats, p=ttest\_1samp(dataset.Height,65)

print(stats,p)

**Output:**

Stats=11.498800238580099

P=1.087893570160242e-26

***Two Sample Paired test***

import pandas as pd

from scipy.stats import ttest\_rel

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

dataset.head()

dataset['Attrition']=dataset['Attrition'].map({'Yes':1,'No':0})

stats,p=ttest\_rel(dataset.Attrition,dataset.PercentSalaryHike)

print(stats,p)

**Output**:

Stats= -272.62099503261675

P= 0.0

***Two Sample Independent Test***

import pandas as pd

from scipy.stats import ttest\_ind

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

dataset.head()

dataset['Attrition']=dataset['Attrition'].map({'Yes':1,'No':0})

stats,p=ttest\_ind(dataset.Attrition,dataset.PercentSalaryHike)

print(stats,p)

**Output:**

Stats=-271.7370409532765

P= 0.0

***OneWay ANOVA***

import pandas as pd

import statsmodels.api as sm

from statsmodels.formula.api import ols

dataset=pd.read\_excel('ANCOVA1.xlsx',sheet\_name=0)

print(dataset.columns)

model=ols('Sales~C(Promotion)',dataset).fit()

oneway=sm.stats.anova\_lm(model,typ=2)

print(oneway)

**Output:**

C(Promotion) 106.066667 2.0 17.943609 0.000011

Residual 79.800000 27.0 NaN NaN

***Twoway Anova***

import pandas as pd

import statsmodels.api as sm

from statsmodels.formula.api import ols

dataset=pd.read\_excel('ANCOVA1.xlsx',sheet\_name=0)

print(dataset.columns)

model=ols('Sales~C(Promotion)+C(Coupon)',dataset).fit()

twoway=sm.stats.anova\_lm(model,typ=2)

print(twoway)

**Output:**

sum\_sq df F PR(>F)

C(Promotion) 106.066667 2.0 52.098237 8.032739e-10

C(Coupon) 53.333333 1.0 52.392947 1.095036e-07

Residual 26.466667 26.0 NaN NaN

***ANCOVA***

import pandas as pd

import statsmodels.api as sm

from statsmodels.formula.api import ols

dataset=pd.read\_excel('ANCOVA1.xlsx',sheet\_name=0)

print(dataset.columns)

model=ols('Sales~C(Promotion)+C(Coupon)+ClietelRatings',dataset).fit()

ANCOVA=sm.stats.anova\_lm(model,typ=2)

print(ANCOVA)

**Output:**

sum\_sq df F PR(>F)

C(Promotion) 106.066667 2.0 51.732296 1.301539e-09

C(Coupon) 53.333333 1.0 52.024936 1.471502e-07

ClietelRatings 0.837931 1.0 0.817375 3.745789e-01

Residual 25.628736 25.0 NaN NaN

Finding ETA square

import pandas as pd

import statsmodels.api as sm

from statsmodels.formula.api import ols

dataset=pd.read\_excel('ANCOVA1.xlsx',sheet\_name=0)

print(dataset.columns)

model=ols('Sales~C(Promotion)',dataset).fit()

oneway=sm.stats.anova\_lm(model,typ=2)

print(oneway)

def anova\_table(oneway):

oneway['eta\_sq']=oneway[:-1]['sum\_sq']/sum(oneway['sum\_sq'])

cols=['sum\_sq','df','F','PR(>F)','eta\_sq']

oneway=oneway[cols]

return oneway;

print(anova\_table(oneway))