# ***ASSIGNMENT***

# **PARAMETRIC AND NON PARAMETRIC TESTS**

Python 3.6.4 |Anaconda, Inc.| (default, Jan 16 2018, 10:22:32) [MSC v.1900 64 bit (AMD64)]

Type "copyright", "credits" or "license" for more information.

IPython 6.2.1 -- An enhanced Interactive Python.

import pandas as pd

from scipy.stats import wilcoxon

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

dataset.head()

Out[5]:

Age Attrition BusinessTravel Department DistanceFromHome \

0 51 No Travel\_Rarely Sales 6

1 31 Yes Travel\_Frequently Research & Development 10

2 32 No Travel\_Frequently Research & Development 17

3 38 No Non-Travel Research & Development 2

4 32 No Travel\_Rarely Research & Development 10

Education EducationField EmployeeCount EmployeeID Gender \

0 2 Life Sciences 1 1 Female

1 1 Life Sciences 1 2 Female

2 4 Other 1 3 Male

3 5 Life Sciences 1 4 Male

4 1 Medical 1 5 Male

NumCompaniesWorked Over18 PercentSalaryHike \

0 ... 1.0 Y 11

1 ... 0.0 Y 23

2 ... 1.0 Y 15

3 ... 3.0 Y 11

4 ... 4.0 Y 12

StandardHours StockOptionLevel TotalWorkingYears TrainingTimesLastYear \

0 8 0 1.0 6

1 8 1 6.0 3

2 8 3 5.0 2

3 8 3 13.0 5

4 8 2 9.0 2

YearsAtCompany YearsSinceLastPromotion YearsWithCurrManager

0 1 0 0

1 5 1 4

2 5 0 3

3 8 7 5

4 6 0 4

[5 rows x 24 columns]

# **WILCOXON SIGN TEST**

**H0: There is no significant relationship between JobLevel and MonthlyIncome.**

**H1: There is a significant relationship between JobLevel and MonthlyIncome.**

stats,p=wilcoxon(dataset.JobLevel,dataset.MonthlyIncome)

print(stats,p)

0.0 0.0

**SINCE,**

**p<0.05, null hypothesis is rejected.**

# **FRIEDMAN TEST**

**H0: There is no significant relationship between PercentSalaryHike, TotalWorkingYears and MonnthlyIncome.**

**H1: There is a significant relationship between PercentSalaryHike, TotalWorkingYears and MonnthlyIncome.**

from scipy.stats import friedmanchisquare

stats2,p2=friedmanchisquare(dataset.MonthlyIncome,dataset.TotalWorkingYears,dataset.PercentSalaryHike)

print(stats2,p2)

7113.22493873597 0.0

**SINCE,**

**p<0.05, null hypothesis is rejected.**

# **MANNWHITNEY TEST**

**H0: There is no significant difference in DistanceFromHome for employees leaving or not leaving the company.**

**H1: There is a significant difference in DistanceFromHome for employees leaving or not leaving the company.**

from scipy.stats import mannwhitneyu

data1=dataset[dataset['Attrition']==0]

data1=dataset[dataset['Attrition']=='Yes']

data2=dataset[dataset['Attrition']=='No']

stats,p=mannwhitneyu(data1.DistanceFromHome,data2.DistanceFromHome)

print(stats,p)

1312110.0 0.4629185205822659

**SINCE,**

**p>=0.05, null hypothesis is accepted.**

# **KRUSKAL WALLLIS TEST**

**H0: There is no significant difference in DistanceFromHome for employees with different business history.**

**H1: There is a significant difference in DistanceFromHome for employees with different business history.**

data3=dataset[dataset['BusinessTravel']=='Travel\_Rarely']

data4=dataset[dataset['BusinessTravel']=='Non-Travel']

data5=dataset[dataset['BusinessTravel']=='Travel\_Frequently']

from scipy.stats import kruskal

stats3,p3=kruskal(data3.DistanceFromHome,data4.DistanceFromHome,data5.DistanceFromHome)

print(stats3,p3)

7.437949655278909 0.02425882457875373

**SINCE,**

**p<0.05, null hypothesis is rejected.**

# **CHI SQUARE TEST**

**H0: There is no dependency of Attrition on Gender.**

**H1: There is a dependency of Attrition on Gender.**

from scipy.stats import chi2\_contingency

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

chitable=pd.crosstab(dataset.Gender,dataset.Attrition)

chitable

Out[6]:

Attrition No Yes

Gender

Female 1494 270

Male 2205 441

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

print(stats,p)

1.349904410246582 0.24529482862926827

**SINCE,**

**p>=0.05, null hypothesis is accepted.**

# **ONE SAMPLE T TEST**

**H0: There is no significant difference of mean of age of employees against population mean=30.**

**H1: There is a significant difference of mean of age of employees against population mean=30.**

from scipy.stats import ttest\_1samp

stats4,p4=ttest\_1samp(dataset.Age,30)

print(stats4,p4)

50.342713783882296 0.0

**SINCE,**

**p<0.05, null hypothesis is rejected.**

# **TWO SAMPLE PAIRED T TEST**

**H0: There is no significant difference in mean of PercentSalaryHike and MonthlyIncome.**

**H1: There is a significant difference in of PercentSalaryHike and MonthlyIncome.**

from scipy.stats import ttest\_rel

stats5,p5=ttest\_rel(dataset.MonthlyIncome,dataset.PercentSalaryHike)

print(stats5,p5)

91.7261306360495 0.0

**SINCE,**

**p<0.05, null hypothesis is rejected.**

# **TWO SAMPLE SEPARATE T TEST**

**H0: There is no significant difference in mean of MonthlyIncome of employees leaving or not leaving the company.**

**H1: There is a significant difference in mean of MonthlyIncome of employees leaving or not leaving the company.**

from scipy.stats import ttest\_ind

data1=dataset[dataset['Attrition']=='Yes']

data2=dataset[dataset['Attrition']=='No']

stats,p=ttest\_ind(data1.MonthlyIncome,data2.MonthlyIncome)

print(stats,p)

-2.0708863763619316 0.03842748490605113

**SINCE,**

**p<0.05, null hypothesis is rejected.**