

ASSIGNMENT

PARAMETRIC AND NON PARAMETRIC TESTS

```
import pandas as pd
```

```
dataset=pd.read_csv("general_data.csv")
```

```
dataset.head()
```

Out[13]:

	Age	Attrition	...	YearsSinceLastPromotion	YearsWithCurrManager
0	51	No	...	0	0
1	31	Yes	...	1	4
2	32	No	...	0	3
3	38	No	...	7	5
4	32	No	...	0	4

[5 rows x 24 columns]

```
dataset.columns
```

Out[14]:

```
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',  
      'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',  
      'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',  
      'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',  
      'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',  
      'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],  
      dtype='object')
```

WILCOXON SIGN TEST

H0: There is no significant relationship between DistanceFromHome and StandardHours.

H1: There is a significant relationship between DistanceFromHome and StandardHours.

```
from scipy.stats import wilcoxon
```

```
stats,p=wilcoxon(dataset.DistanceFromHome,dataset.StandardHours)
```

```
print(stats,p)
```

```
4191909.0 0.043962321781016465
```

P<0.05.

So Null Hypothesis is rejected

FRIEDMAN TEST

H0: There is no significant relationship between DistanceFromHome, StandardHours and Age.

H1: There is a significant relationship between DistanceFromHome, StandardHours and Age.

```
from scipy.stats import friedmanchisquare
```

```
stats,p=friedmanchisquare(dataset.DistanceFromHome,dataset.StandardHours,dataset.Age)
```

```
print(stats,p)
```

```
6584.534715025906 0.0
```

P<0.05.

So Null Hypothesis is rejected

MANNWHITNEY TEST

H0: There is no significant difference in BusinessTravel and MaritalStatus

H1: There is a significant difference in BusinessTravel and MaritalStatus

```
from scipy.stats import mannwhitneyu
```

```
stats,p=mannwhitneyu(dataset.BusinessTravel,dataset.MaritalStatus)
```

```
print(stats,p)
```

```
634500.0 0.0
```

P<0.05.

So Null Hypothesis is rejected.

KRUSKAL WALLIS 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.

```
from scipy.stats import kruskal
```

```
stats,p=kruskal(dataset.BusinessTravel,dataset.MaritalStatus,dataset.Department)
```

```
print(stats,p)  
7330.102366193341 0.0
```

P<0.05.

So 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.

```
dataset1=dataset.dropna()
```

```
from scipy.stats import chi2_contingency
```

```
chitable=pd.crosstab(dataset1.Attrition,dataset1.Gender)
```

```
chitable
```

```
Out[40]:
```

```
Gender    Female    Male
```

```
Attrition
```

```
No         1488    2189
```

```
Yes         268     437
```

```
stats,p,dof,expeted=chi2_contingency(chitable)
```

```
print(stats,p)  
1.3825823839528295 0.23966176275638887
```

p>0.05 Null Hypothesis is Accepted.

ONE SAMPLE T TEST

H0: There is no significant difference of mean of MonthlyIncome against population mean=65029.

H1: There is a significant difference of mean of MonthlyIncome against population mean=65029.

```
from scipy.stats import ttest_1samp
```

```
stats,p=ttest_1samp(dataset.MonthlyIncome,65029)
```

```
print(stats,p)  
0.00044149505974563756 0.999647757893073
```

p>0.05 Null Hypothesis is Accepted.

TWO SAMPLE PAIRED T TEST

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

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

```
from scipy.stats import ttest_rel
```

```
stats,p=ttest_rel(dataset1.TotalWorkingYears,dataset1.MonthlyIncome)
```

```
print(stats,p)
```

```
-91.34250233404504 0.0
```

p<0.05.

So 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
```

```
stats,p=ttest_ind(data1.MonthlyIncome,data2.MonthlyIncome)
```

```
print(stats,p)
```

```
-2.0708863763619316 0.03842748490605113
```

p<0.05.

So Null Hypothesis is rejected.