Statistical Test

Non Parametric Test :

1. **Wilcoxon :**
   Its is used to compare two paired samples

   H0: There is no significant difference between employees Years Since Last Promotion and Years With Current Manager.
   H1: There is significant difference between employees Years Since Last Promotion and Years With Current Manager.

   ```
   import pandas as pd
   dataset = pd.read_csv("general_data.csv")
   dataset.head()
   Out[3]:
     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]

   from scipy.stats import Wilcoxon
   stats,p = wilcoxon(dataset.YearsSinceLastPromotion,dataset.YearsWithCurrManager)
   print(stats,p)
   699082.5 7.578916343856332e-279
   ```

   Since $p < 0.05$, Ho is rejected.
   Hence proved There is significant difference between employees Years Since Last Promotion and Years With Current Manager.

2. **Friedman :**
   Its is used to compare more than two paired samples

   H0: There is no significant difference between employees number of companies work, total working years and years at company.
   H1: There is significant difference between employees number of companies work, total working years and years at company.

Statistical Test

```
from scipy.stats import friedmanchisquare
stats,p = friedmanchisquare(dataset.NumCompaniesWorked,dataset.YearsAtCompany,
dataset.TotalWorkingYears)
print(stats,p)
5100.159963317181 0.0
```

Since $p < 0.05$, Ho is rejected.
Hence proved There is significant difference between employees number of companies work, total working years and years at company.

3. **Mann-whitney :**
   Its is used to compare two independent samples

   H0: There is no significant difference between Education Field and Job Role.
   H1: There is significant difference between Education Field and Job Role.

   ```
   from scipy.stats import mannwhitneyu
   stats,p = mannwhitneyu(dataset.EducationField,dataset.JobRole)
   print(stats,p)
   8566884.0 8.531789873710365e-23
   ```

   Since $p < 0.05$, Ho is rejected.
   Hence proved There is significant difference between employees Education Field and Job Role.

4. **Kruskal-walls :**
   Its is used to compare more than two independent samples

   H0: There is no significant difference between Department, Education Field and Job Role.
   H1: There is significant difference between Department, Education Field and Job Role.

   ```
   from scipy.stats import Kruskal
   stats,p = kruskal(dataset.EducationField,dataset.JobRole, dataset.Department)
   print(stats,p)
   2079.6462052037637 0.0
   ```

   Since $p < 0.05$, Ho is rejected.
   Hence proved There is significant difference between Department, Education Field and Job Role.

Statistical Test

5. **Chi-square :**
   It is used to check the dependency between categorical variables.

   H0: There is no significant difference between Attrition and Gender.
   H1: There is significant difference between Attrition and Gender.

   ```
   from scipy.stats import chi2_contingency
   chitable = pd.crosstab(dataset.Attrition,dataset.Gender)
   stats,p,dof,expected = chi2_contingency(chitable)
   print(stats,p)
   1.349904410246582 0.24529482862926827
   ```

   Since $p > 0.05$, Ho is accepted.
   Hence proved There is no significant difference between Attrition and Gender.

Parametric Test :

6. **One sample t-test :**
   It is used to compare sample mean with population mean

   H0: There is no significant difference between training time of employee and standard training time as 3.
   H1: There is significant difference between training time of employee and standard training time as 3.

   ```
   from scipy.stats import ttest_1samp
   stats,p = ttest_1samp(dataset.TrainingTimesLastYear,3)
   print(stats,p)
   -10.338997107228291 8.987949368189617e-25
   ```

   Since $p < 0.05$, Ho is rejected.
   Hence proved There is significant difference between training time of employee and standard training time as 3.

Statistical Test

7. **Two sample paired t-test :**
   It is used to compare mean of two paired samples

   H0: There is no significant difference between Monthly Income and years at company.
   H1: There is significant difference between Monthly Income and years at company.

   ```
   from scipy.stats import ttest_rel
   stats,p = ttest_rel(dataset.MonthlyIncome,dataset.YearsAtCompany)
   print(stats,p)
   91.73768218543985 0.0
   ```

   Since p < 0.05 , Ho is rejected.
   Hence proved There is significant difference between There is significant difference between Monthly Income and years at company.

8. **Two sample Independent :**
   It is used to compare mean of two independent samples

   H0: There is no significant difference between Stock Option Level and Job Level.
   H1: There is significant difference between Stock Option Level and Job Level.

   ```
   from scipy.stats import ttest_ind
   stats,p = ttest_ind(dataset.StockOptionLevel,dataset.JobLevel)
   print(stats,p)
   -60.391685443496 0.0
   ```

   Since p < 0.05 , Ho is rejected.
   Hence proved There is significant difference between Stock Option Level and Job Level.