NON PARAMETRIC TESTS

1>Wilcoxon sign test

import pandas as pd

data=pd.read_excel("1 Wilcoxon.xlsx")

data.head()

```
In [5]: data.head()
Out[5]:
   ID  TRT  AGE  WEIGHIN  STAGE  TOTALCIN  TOTALCW2  TOTALCW4 TOTALCW6
0   1    0   52    124.0      2         6         6         6        7
1   5    0   77    160.0      1         9         6        10        9
2   6    0   60    136.5      4         7         9        17       19
3   9    0   61    179.6      1         6         7         9        3
4  11    0   59    175.8      2         6         7        16       13
```

H_0 = There is no significant difference between the calcium levels of patients from initial to first 2 weeks

H_A = There is no significant difference between the calcium levels of patients from initial to first 2 weeks

from scipy.stats import Wilcoxon

stats,p=wilcoxon(data.TOTALCIN,data.TOTALCW2)

```
In [8]: print(stats,p)
29.5 0.00259741456482452
```

As the p value is less than .05 the null hypothesis is rejected .

So there is significant difference in calcium levels of patients in 2 weeks.

2>FRIDMAN TEST

H_0 = There is no significant difference between the calcium levels of patients from initial to first 2 weeks and to 4 weeks

H_A = There is no significant difference between the calcium levels of patients from initial to first 2 weeks and to 4 weeks

from scipy.stats import friedmanchisquare

d1=data.TOTALCIN

d2=data.TOTALCW2

d3=dataset.TOTALCW4

stat,p=friedmanchisquare(d1,d2,d3)

```
In [18]: print(stat,p)
27.927710843373504 8.62133745016363e-07
```

As the p value is less than .05 the null hypothesis is rejected .

So there is significant difference in calcium levels of patients in 2 weeks and 4 weeks.

3>MANN WHITNEY TEST

H_0 = There is no significant difference in sales due to design 1 and design 2

H_A = There is significant difference in sales due to design1 and design2.

from scipy.stats import mannwhitneyu

datat1=pd.read_excel("3 Mann Whitney.xlsx",shee_tname=1

a1=dataset1.Design1

a2=dataset1.Design2

stat,p=mannwhitneyu(a1,a2)

print(stat,p)

```
In [28]: print(stat,p)
9.0 0.2641796636354743
```

As the p value is more than .05 the null hypothesis is accepted .

So there is no difference in design1 and design 2 on sales.

4>KRUSHAL WALLIS TEST

```
from scipy.stats import Kruskal
data2=pd.read_excel("4 Kruskal Wallis.xlsx",sheet_name=0)
data2.head()
```

```
In [6]: data2.head()
Out[6]:
   Design1  Design2  Design3
0       11       12       23
1       17       10       20
2       16       15       18
3       14       19       17
4       15       11       10
```

H_0 = There is no significant difference in sales due to design 1 and design 2 and design 3.

H_A = There is significant difference in sales due to design1 and design2 and design 3.

```
b1=data2.Design1
b2=data2.Design2
b3=data2.Design3
stat,p=kruskal(b1,b2,b3)
```

```
In [12]: print(stat,p)
2.7345323741007226 0.25480259087913626
```

As the p value is more than .05 the null hypothesis is accepted .

So there is no difference in design1 and design 2 and design 3 on sales.

5>CHI SQUARE TEST

data3=pd.read_excel("5 Chi square Test.xlsx")

data4=data3.dropna()

data4.head()

```
In [16]: data4.head()
Out[16]:
     ids  Gender        Smoking
0  43783  Female      NonSmoker
1  20278  Female      NonSmoker
2  20389  Female      NonSmoker
4  24559    Male  CurrentSmoker
5  28980  Female      NonSmoker
```

H_0 = There is no significant dependency between in gender and smoking.

H_A = There is significant dependency between in gender and smoking.

from scipy.stats import chi2_contingency

chitable=pd.crosstab(data4.Gender,data4.Smoking)

stats,p,dof,expected=chi2_contingency(chitable)

```
In [20]: print(stats,p)
3.1712567666931584 0.20481904779163013
```

As the p value is more than .05 the null hypothesis is accepted .

So there is no dependency in gender for smoking.

PARAMETRIC TEST

1>ONE SAMPLE T-TEST

from scipy.stats import ttest_1samp

dataset6=pd.read_excel("1. One Sample.xlsx",sheet_name=0)

```
In [6]: dataset6.head()
Out[6]:
     ids  Height
0  43783   72.35
1  20278   70.66
2  20389   70.68
3  24559   67.43
4  28980   68.45
```

H_0 = There is no significant difference between sample mean and population mean.

H_A = There is significant difference between sample mean and population mean.

h1=dataset6.Height

stat,p=ttest_1samp(h1,65)

```
In [9]: print(stat,p)
11.498800238580099 1.087893570160242e-26
```

As the p value is less than .05 the null hypothesis is rejected .

So there is difference in sample and population mean .

2> TWO SAMPLE PAIRED T-TEST

from scipy.stats import ttest_rel

dataset3=pd.read_excel("2. Paired Sample.xlsx",sheet_name=0)

```
In [12]: dataset3.head()
Out[12]:
     ids  English   Math
0  43783    88.24  60.02
1  20278    89.45  70.19
2  20389    96.73  71.20
3  22820    74.06  55.89
4  24559    82.61  65.52
```

H_0 = There is no significant difference between mean of Maths and English marks.

H_A = There is significant difference between mean of Maths and English marks.

p1=dataset3.English

p2=dataset3.Math

stat,p=ttest_rel(p1,p2)

```
In [16]: print(stat,p)
36.312568981719856 3.0710987192210606e-128
```

As the p value is less than .05 the null hypothesis is rejected .

So there is difference in mean of English and maths marks.

3>TWO SEPARATE/INDEPENDENT T-TEST

from scipy.stats import ttest_ind

dataset4=pd.read_excel("3. Independent Sample.xlsx",sheetname=3)

dataset4.head()

```
In [20]: dataset4.head()
Out[20]:
   Nonathelete  Athelete
0     0.004413  0.004462
1     0.004872  0.005146
2     0.008851  0.004023
3     0.006508  0.003941
4     0.006314  0.004764
```

H_0 = There is no significant difference between mean of duration by atheletes and non atheletes.

H_A = There is significant difference between mean of duration by atheletes and non atheletes.

z1=dataset4.Nonathelete

z2=dataset4.Athelete

stat,p=ttest_ind(z1,z2)

print(stat,p)

```
In [24]: print(stat,p)
13.368790432137319 7.116633157230895e-33
```

As the p value is less than .05 the null hypothesis is rejected .

So there is difference in mean of duration of atheletes and non atheletes.

4> ONE SAMPLE F-TEST(ONE WAY ANOVA)

data6=pd.read_excel("ANCOVA1.xlsx")

data6.head()

Out[26]:

	Store Number	Sales	Promotion	Coupon	ClietelRatings
0	1	10	1	1	9
1	2	9	1	1	10
2	3	10	1	1	8
3	4	8	1	1	4
4	5	9	1	1	6

H_0 = There is no significant difference in promotion of low, medium, high on sales

H_A = There is significant difference in promotion of low, medium, high on sales

import statsmodels.api as sm

from statsmodels.formula.api import ols

mode=ols('Sales~C(Promotion)',data6).fit()

oneway=sm.stats.anova_lm(mode,typ=2)

print(oneway)

```
In [36]: print(oneway)
                 sum_sq    df          F    PR(>F)
C(Promotion)  106.066667   2.0  17.943609  0.000011
Residual       79.800000  27.0        NaN       NaN
```

As the p value is less than .05 the null hypothesis is rejected .

So there is difference in promotion of low, medium, high on sales.

5> TWO SAMPLE F-TEST(TWO WAY ANOVA)

H_0 = There is no significant difference in coupon and promotion of low, medium, high on sales

H_A = There is significant difference in coupon and promotion of low, medium, high on sales

```
mode=ols('Sales~C(Promotion)+C(Coupon)',data6).fit()

twoway=sm.stats.anova_lm(mode,typ=2)

print(twoway)
```

```
In [45]: print(twoway)
                 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
```

As the p value is less than .05 the null hypothesis is rejected .

So there is difference in coupon and promotion on sales.