import pandas as pd

dataset=pd.read\_excel("Correlation.xlsx",sheet\_name=0)

dataset.head()

**Out[3]:**

**Respondent Number Attitude Duration**

**0 1 6 10**

**1 2 9 12**

**2 3 8 12**

**3 4 3 4**

**4 5 10 12**

* **CORRELATION**

from scipy.stats import pearsonr

stats,p=pearsonr(dataset.Attitude,dataset.Duration)

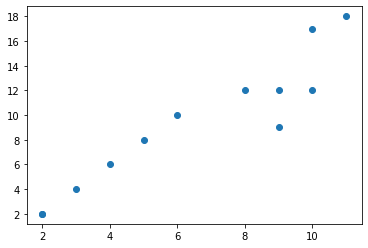
print(stats,p)

**0.9360778239640097 7.545161167077795e-06**

**Ho – Null hypothesis is rejected.There is significant difference between Attitude & Duration**

**import matplotlib.pyplot as plt**

**plt.scatter(dataset.Attitude,dataset.Duration)**

****

**+vely correlated because of the upward direction of the line.**

**dataset.corr()**

**Out[9]:**

**Respondent Number Attitude Duration**

**Respondent Number 1.000000 -0.041827 -0.052700**

**Attitude -0.041827 1.000000 0.936078**

**Duration -0.052700 0.936078 1.000000**

* **NON PARAMETRIC TESTS**
* **WILCOXON TESTING**

from scipy.stats import wilcoxon

dataset2=pd.read\_excel("1 Wilcoxon.xlsx",sheet\_name=0)

stats,p=wilcoxon(dataset2.TOTALCIN,dataset2.TOTALCW2)

print(stats,p)

**29.5 0.00259741456482452**

**H0 is rejected.There is significant change in Calcium in second week than the first week**

* **FRIEDMAN TEST**

dataset3=pd.read\_excel("1 Wilcoxon.xlsx",sheet\_name=0)

from scipy.stats import friedmanchisquare

stats,p=friedmanchisquare(dataset3.TOTALCIN,dataset3.TOTALCW2,dataset3.TOTALCW4)

print(stats,p)

**27.927710843373504 8.62133745016363e-07**

**H0 is rejected.there is significant change in calcium in second week and fourth week than the first week**

* **MANNWHITNEY TEST**

dataset4=pd.read\_excel("3 Mann Whitney.xlsx",sheet\_name=1)

from scipy.stats import mannwhitneyu

stats,p=mannwhitneyu(dataset4.Design1,dataset4.Design2)

print(stats,p)

**9.0 0.2641796636354743**

**H0 is accepted.No significant sales diffrence between design 1 and design 2**

* **KRUSKAL WALLEY TEST**

dataset5=pd.read\_excel("4 Kruskal Wallis.xlsx",sheet\_name=0)

from scipy.stats import kruskal

stats,p=kruskal(dataset5.Design1,dataset5.Design2,dataset5.Design3)

print(stats,p)

**2.7345323741007226 0.25480259087913626**

**H0 is accepted.No significant sales difference between design1 ,design2 & design3**

* **CHISQUARE TEST**

**dataset6=pd.read\_excel("5 Chi square Test.xlsx",sheet\_name=0)**

**from scipy.stats import chi2\_contingency**

**chitable=pd.crosstab(dataset6.Gender,dataset6.Smoking)**

**chitable**

**Out[34]:**

**Smoking CurrentSmoker NonSmoker PastSmoker**

**Gender**

**Female 31 149 13**

**Male 37 148 24**

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

**print(stats,p)**

**3.1712567666931584 0.20481904779163013**

**H0 is accepted.Smoking is not depends upon gender.both men and women are smoking**

* **PARAMETRIC TEST**
* **ONE SAMPLE T-TEST**

dataset6=pd.read\_excel("1. One Sample.xlsx",sheet\_name=0)

from scipy.stats import ttest\_1samp

stats,p=ttest\_1samp(dataset6.Height,65)

print(stats,p)

**11.498800238580099 1.087893570160242e-26**

**H0 is rejected.**

* **2 SAMPLE PAIRED T-TEST**

dataset7=pd.read\_excel("2. Paired Sample.xlsx",sheet\_name=0)

from scipy.stats import ttest\_rel

stats,p=ttest\_rel(dataset7.English,dataset7.Math)

print(stats,p)

**36.312568981719856 3.0710987192210606e-128**

**H0 is rejected**

* **2 SAMPLE INDEPENDENT T-TEST**

from scipy.stats import ttest\_ind

dataset8=pd.read\_excel("3. Independent Sample.xlsx",sheet\_name=3)

stats,p=ttest\_ind(dataset8.Nonathelete,dataset8.Athelete)

print(stats,p)

**13.368790432137319 7.116633157230895e-33**

**There is significant difference between Athelete and Non athlete about time**

* **ONE WAY ANOVA**

dataset9=pd.read\_excel("ANCOVA1.xlsx",sheet\_name=0)

dataset.head()

**Out[58]:**

**Respondent Number Attitude Duration**

**0 1 6 10**

**1 2 9 12**

**2 3 8 12**

**3 4 3 4**

**4 5 10 12**

import statsmodels.api as sm

model=ols('Sales~(Promotion)',dataset9).fit()

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

print(oneway)

**df sum\_sq mean\_sq F PR(>F)**

**C(Promotion) 2.0 106.066667 53.033333 17.943609 0.000011**

**Residual 27.0 79.800000 2.955556 NaN NaN**

* **TWOWAY ANOVA**

model=ols('Sales~C(Promotion)+C(Coupon)',dataset9).fit()

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

print(twoway)

**df sum\_sq mean\_sq F PR(>F)**

**C(Promotion) 2.0 106.066667 53.033333 52.098237 8.032739e-10**

**C(Coupon) 1.0 53.333333 53.333333 52.392947 1.095036e-07**

**Residual 26.0 26.466667 1.017949 NaN NaN**

* **ANCOVA**

model=ols('Sales~C(Promotion)+C(Coupon)+C(ClietelRatings)',dataset9).fit()

ancova=sm.stats.anova\_lm(model,type=2)

print(ancova)

**df sum\_sq mean\_sq F PR(>F)**

**C(Promotion) 2.0 106.066667 53.033333 50.875000 5.605989e-09**

**C(Coupon) 1.0 53.333333 53.333333 51.162791 3.592805e-07**

**C(ClietelRatings) 4.0 3.533333 0.883333 0.847384 5.103673e-01**

**Residual 22.0 22.933333 1.042424 NaN NaN**

Significant difference in sales using promotion and coupon..no difference with clietel rating