Week – 4

Write a program to calculate chi-square value using Python. Report your observation

```
import scipy.stats as stats

import seaborn as sns
import pandas as pd
import numpy as np
dataset=sns.load_dataset('tips')

dataset.head()
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
dataset_table=pd.crosstab(dataset['sex'],dataset['smoker'])
print(dataset_table)
```

```
smoker  Yes  No
sex
Male     60  97
Female   33  54
```

```
dataset_table.values
```

```
array([[60, 97],
       [33, 54]])
```

```
#observed values
observed_values=dataset_table.values
print("observed values:-\n",observed_values)
```

```
observed values:-
 [[60 97]
 [33 54]]
```

```
val=stats.chi2_contingency(dataset_table)

val

 (0.0, 1.0, 1, array([[59.84016393, 97.15983607],
        [33.15983607, 53.84016393]]))

expected_values=val[3]

no_of_rows=len(dataset_table.iloc[0:2,0])
no_of_coloumns=len(dataset_table.iloc[0,0:2])
ddof=(no_of_rows-1)*(no_of_coloumns-1)
print("degree_of_freedom",ddof)
alpha=0.05

 degree_of_freedom 1

from scipy.stats import chi2
chi_squre=sum([(o-
e)**2./e for o,e in zip(observed_values,expected_values)])
chi_squre_statistic=chi_squre[0]+chi_squre[1]

print("chi_squre_statistic:",chi_squre_statistic)

  chi_squre_statistic: 0.001934818536627623

critical_value=chi2.ppf(q=1-alpha,df=ddof)
print('critical_value:',critical_value)

  critical_value: 3.841458820694124
```

```
p_value=1-chi2.cdf(x=chi_squre_statistic,df=ddof)
print("p-value:",p_value)
print("significancee level:",alpha)
print("degree of freedom:",ddof)
print('p-value:',p_value)

  p-value: 0.964915107315732
  significancee level: 0.05
  degree of freedom: 1
  p-value: 0.964915107315732
```

ADITYA
ENLIGHTENS THE NESCIENCE