```
import scipy.stats as stats
from scipy.stats import chi2
import seaborn as sns
import pandas as pd
import numpy as np


dataset = sns.load_dataset("tips")
dataset.head(20)
```

	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
5	25.29	4.71	Male	No	Sun	Dinner	4
6	8.77	2.00	Male	No	Sun	Dinner	2
7	26.88	3.12	Male	No	Sun	Dinner	4
8	15.04	1.96	Male	No	Sun	Dinner	2
9	14.78	3.23	Male	No	Sun	Dinner	2
10	10.27	1.71	Male	No	Sun	Dinner	2
11	35.26	5.00	Female	No	Sun	Dinner	4
12	15.42	1.57	Male	No	Sun	Dinner	2
13	18.43	3.00	Male	No	Sun	Dinner	4
14	14.83	3.02	Female	No	Sun	Dinner	2
15	21.58	3.92	Male	No	Sun	Dinner	2
16	10.33	1.67	Female	No	Sun	Dinner	3
17	16.29	3.71	Male	No	Sun	Dinner	3
18	16.97	3.50	Female	No	Sun	Dinner	3
19	20.65	3.35	Male	No	Sat	Dinner	3

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

```
array([[60, 97],
       [33, 54]], dtype=int64)
```

```
val = stats.chi2_contingency(dataset_table)
val
```

```
Chi2ContingencyResult(statistic=0.0, pvalue=1.0, dof=1, expected_freq=array([[59.84016393, 97.15983607],
       [33.15983607, 53.84016393]]))
```

```
expected_values = val[3]
expected_values
```

```
array([[59.84016393, 97.15983607],
       [33.15983607, 53.84016393]])
```

```
rows = len(dataset_table.iloc[0:2,0])
col = len(dataset_table.iloc[0,0:2])
dof = val[2] # (rows-1)*(col-1)
alpha = 0.05
rows, col, dof, alpha
```

```
(2, 2, 1, 0.05)
```

```
chai_square = sum([(o - e)**2 / e for o, e in zip(observed_values, expected_values)])
chai_square_statistic = chai_square[0] + chai_square[1]
chai_square_statistic
```

```
0.001934818536627623
```

```
p_value = 1 - chi2.cdf(chai_square_statistic, df=dof)
print('P Value : ', p_value, '\nSignifinance Level :', alpha, '\nDegree of Freedom :', dof)
```

```
P Value :  0.964915107315732
Signifinance Level : 0.05
Degree of Freedom : 1
```

```
if p_value >= alpha:
    print('As', p_value, 'is greater than', alpha, ', there is \033[1mno relation\033[0m between two categorical variables.')
else:
    print('As', p_value, 'is less than', alpha, ', there is a \033[1mrelation\033[0m between two categorical variables.')
```

```
As 0.964915107315732 is greater than 0.05 , there is no relation between two categorical variables.
```