1:- Chi_Square_test

Categorical Data: The data must be in categorical form, meaning it is divided into distinct categories. These categories can be nominal (e.g., gender, color, type) or ordinal (e.g., rankings, scales).

Count Data: The test is applied to frequency counts of occurrences within the categories. For instance, how many times different categories appear in a dataset.

import librares

```
In [1]:  import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
         from scipy import stats
```

load the data

```
In [2]:  df=sns.load_dataset("titanic")
         df
```

Out[2]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	NaN	Southampton	no
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Southampton	yes
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	NaN	Southampton	no
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	Cherbourg	yes
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	NaN	Queenstown	no

891 rows × 15 columns

Null Hypothesis (H0): There is no difference in the survival rates between male and female passengers on the Titanic.

Alternative Hypothesis (H1): There is a difference in the survival rates between male and female passengers on the Titanic.

In [3]:
```python
# Create a contingency Table
contingency_table=pd.crosstab(df['sex'], df['survived'])
contingency_table
```

Out[3]:

survived	0	1
sex		
female	81	233
male	468	109

In [4]:
```python
# preform chi_square test
chi2, p, dof, expected= stats.chi2_contingency(contingency_table)

print('chi_square: ', chi2)
print('P_value: ', p)
print('degree of freedom: ', dof)
print(f'Expected:\n {expected}')
```

```
chi_square:  260.71702016732104
P_value:  1.1973570627755645e-58
degree of freedom:  1
Expected:
 [[193.47474747 120.52525253]
 [355.52525253 221.47474747]]
```

Rejecting the null hypothesis at the 0.05 significance level means that we have very strong evidence to suggest that survival rates were indeed different between male and female passengers on the Titanic. This supports the alternative hypothesis that female passengers had a higher likelihood of survival compared to male passengers.

In [ ]: