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

In [3]:
```
CO = pd.read_csv('/Users/SAURABH/Saurabh patil/DATA SCIENCE/Hypothesis/Costomer+OrderForm
```

In [4]:
```
CO
```

Out[4]:

	Phillippines	Indonesia	Malta	India
0	Error Free	Error Free	Defective	Error Free
1	Error Free	Error Free	Error Free	Defective
2	Error Free	Defective	Defective	Error Free
3	Error Free	Error Free	Error Free	Error Free
4	Error Free	Error Free	Defective	Error Free
...	...	...	...	...
295	Error Free	Error Free	Error Free	Error Free
296	Error Free	Error Free	Error Free	Error Free
297	Error Free	Error Free	Defective	Error Free
298	Error Free	Error Free	Error Free	Error Free
299	Error Free	Defective	Defective	Error Free

300 rows × 4 columns

In [5]:
```
CO.Phillippines.value_counts()
```

Out[5]:
```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
```

In [6]:
```
CO.Indonesia.value_counts()
```

Out[6]:
```
Error Free    267
Defective      33
Name: Indonesia, dtype: int64
```

In [7]:
```
CO.Malta.value_counts()
```

Out[7]:
```
Error Free    269
Defective      31
Name: Malta, dtype: int64
```

In [8]:
```
CO.India.value_counts()
```

Out[8]:
```
Error Free    280
Defective      20
Name: India, dtype: int64
```

In [9]:
```
co = pd.DataFrame(index=['Error Free','Defective'],
                  data={'Phillippines':[271,29],'Indonesia':[267,33],'Malta':[269,31],'In
```

Out[10]:

	Phillippines	Indonesia	Malta	India
Error Free	271	267	269	280
Defective	29	33	31	20

In [11]:
```
#Inputs are 4 discrete variables(east,west,north,south).
#Output is also discrete.
#We are trying to find out if proportions of male and female are similar or not across th
#Hence, we'll proceed with chi-square test
```

In [12]:
```
from scipy.stats import chi2_contingency
```

In [13]:
```
#Create hypothesis
#Ho= Percentages of Defective across all the centres is same
#Ha= Percentages of Defective across all the centres is not same
```

In [14]:
```
chi2_stat, p_val, dof, ex =stats.chi2_contingency(co)

print("===Chi2 Stat===")
print(chi2_stat)
print("\n")
print("===Degrees of Freedom===")
print(dof)
print("\n")
print("===P-Value===")
print(p_val)
print("\n")
print("===Contingency Table===")
print(ex)
```

```
===Chi2 Stat===
3.858960685820355


===Degrees of Freedom===
3


===P-Value===
0.2771020991233135


===Contingency Table===
[[271.75 271.75 271.75 271.75]
 [ 28.25  28.25  28.25  28.25]]
```

In [15]:
```
#Since p-value (0.277)> alpha (0.05), hence we can't reject the null hypothesis
#Conclusion: Percentages of Defective across all the centres is same
```

In [ ]:
```

```