

In [1]:

```
import pandas as pd
import numpy as np
from scipy import stats
```

In [2]:

```
customer=pd.read_csv('Customer+OrderForm.csv')
customer
```

Out[2]:

	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]:

```
customer['Phillippines'].value_counts()
```

Out[5]:

```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
```

In [7]:

```
customer['Indonesia'].value_counts()
```

Out[7]:

```
Error Free    267
Defective     33
Name: Indonesia, dtype: int64
```

In [8]:

```
customer['Malta'].value_counts()
```

Out[8]:

```
Error Free    269
Defective     31
Name: Malta, dtype: int64
```

In [9]:

```
customer['India'].value_counts()
```

Out[9]:

```
Error Free    280
Defective     20
Name: India, dtype: int64
```

In [10]:

```
new=np.array([[271,267,269,280],[29,33,31,20]])
new
```

Out[10]:

```
array([[271, 267, 269, 280],
       [ 29,  33,  31,  20]])
```

## Formulation of Hypothesis

**Ho : There is no difference of % of defective from centre**

**Ha : There is difference of % of defective from centre**

## Test

**We use Chi -Squared test because we need proportion and we have qualitative data**

In [13]:

```
p=stats.chi2_contingency(new)
```

In [14]:

```
p[1]
```

Out[14]:

```
0.2771020991233135
```

**Here significance value is 0.05% and we got p-value as 0.277 which is greater**

## Takeaway

**As we got 27% as p-value we don't reject  $H_0$  that is we can say that there is no difference of % of defective from centre**

In [ ]: