Ques- TeleCall uses 4 centers around the globe to process customer order forms. They audit a certain % of the customer order forms. Any error in order form renders it defective and has to be reworked before processing. The manager wants to check whether the defective % varies by centre. Please analyze the data at 5% significance level and help the manager draw appropriate inferences

In [1]:

```python
import pandas as pd
import numpy as np
from scipy import stats
from scipy.stats import norm
from scipy.stats import chi2_contingency
```

In [2]:

```python
data=pd.read_csv('Costomer+OrderForm.csv')
data
```

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 [3]:

```python
data.Phillippines.value_counts()
```

Out[3]:

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

In [4]:

```
data.Indonesia.value_counts()
```

Out[4]:

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

In [5]:

```
data.Malta.value_counts()
```

Out[5]:

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

In [6]:

```
data.India.value_counts()
```

Out[6]:

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

In [7]:

```
# Make a contingency table
array=np.array([[271,267,269,280],[29,33,31,20]])
array
```

Out[7]:

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

Assume Null Hypothesis as Ho: Independence of categorical variables (customer order forms defective % does not varies by centre) Thus, Alternative hypothesis as Ha Dependence of categorical variables (customer order forms defective % varies by centre)

In [8]:

```
# Chi2 contengency independence test
chi2_contingency(array) #o/p is (chi2 stats value,p_value,df,expected observations)
```

Out[8]:

```
(3.858960685820355,
 0.2771020991233135,
 3,
 array([[271.75, 271.75, 271.75, 271.75],
        [ 28.25,  28.25,  28.25,  28.25]]))
```

In [ ]:

```python
# Compare p_value with α = 0.05
```

Inference: As (p_value = 0.2771) > (α = 0.05); Accept Null Hypthesis i.e. Independence of categorical variables
Thus, customer order forms defective % does not varies by centre

In [ ]: