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]:
```
import pandas as pd
import numpy as np
from scipy import stats
import seaborn as sns
```

In [3]:
```
dff = pd.read_csv('Costomer+OrderForm.csv')
dff.head()
```

Out[3]:

	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

In [5]:
```
dff.describe()
```

Out[5]:

	Phillippines	Indonesia	Malta	India
count	300	300	300	300
unique	2	2	2	2
top	Error Free	Error Free	Error Free	Error Free
freq	271	267	269	280

In [6]:
```
Phillippines_value = dff['Phillippines'].value_counts()
Indonesia_value = dff['Indonesia'].value_counts()
Malta_value = dff['Malta'].value_counts()
India_value = dff['India'].value_counts()
print(Phillippines_value)
print(Indonesia_value)
print(Malta_value)
print(India_value)
```

```
Error Free    271
Defective      29
Name: Phillippines, dtype: int64
Error Free    267
Defective      33
Name: Indonesia, dtype: int64
Error Free    269
Defective      31
Name: Malta, dtype: int64
Error Free    280
Defective      20
Name: India, dtype: int64
```

In [7]:
```
stats.chi2_contingency([[271,267,269,280],[29,33,31,20]])
```

Out[7]:
```
(3.858960685820355,
 0.2771020991233135,
 3,
 array([[271.75, 271.75, 271.75, 271.75],
        [ 28.25,  28.25,  28.25,  28.25]]))
```

In [ ]:
```
# pvalue(0.27710)>0.05 so we accept the null hypothesis
```