4) 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
from scipy import stats as stats
from scipy.stats import chi2_contingency
from scipy.stats import chi2
```

In [2]:

```
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]:

```
data.head(2)
```

Out[3]:

	Phillippines	Indonesia	Malta	India
0	Error Free	Error Free	Defective	Error Free
1	Error Free	Error Free	Error Free	Defective

In [4]:

```
print(data['Phillippines'].value_counts(),
      data['Indonesia'].value_counts(),
      data['Malta'].value_counts(),
      data['India'].value_counts())
```

```
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 [5]:

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

Out[5]:

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

In [6]:

```
stat, p, dof, expected = chi2_contingency([[271,267,269,280],[29,33,31,20]])
```

In [7]:

```
stat
```

Out[7]:

```
3.858960685820355
```

In [8]:

```
p
```

Out[8]:

```
0.2771020991233135
```

In [9]:

```
print('dof=%d' % dof)
print(expected)
```

```
dof=3
[[271.75 271.75 271.75 271.75]
 [ 28.25  28.25  28.25  28.25]]
```

In [14]:

```
alpha = 0.05
prob=1-alpha
critical = chi2.ppf(prob, dof)
print('probability=%.3f, critical=%.3f, stat=%.3f' % (prob, critical, stat))
if abs(stat) >= critical:
    print('Dependent (reject H0),variables are related')
else:
    print('Independent (fail to reject H0), variables are not related')
```

```
probability=0.950, critical=7.815, stat=3.859
Independent (fail to reject H0), variables are not related
```

In [11]:

```
print('significance=%.3f, p=%.3f' % (alpha, p))
if p <= alpha:
    print('Dependent (reject H0)')
else:
    print('Independent (fail to reject H0)')
```

```
significance=0.050, p=0.277
Independent (fail to reject H0)
```