In [8]:

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-pytho
n
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
from scipy import stats
from scipy.stats import chi2
# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files
under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserve
d as output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of
the current session
```

```
/kaggle/input/data-antibody/data_antibody.csv
/kaggle/input/covid-data-gender/data_covid_fe.csv
/kaggle/input/covid-data-gender/both_covid_data_gender.csv
/kaggle/input/covid-data-gender/data_covid_ma.csv
/kaggle/input/pcr-data/pcr_data.csv
/kaggle/input/both-covid-data/both_covid_data.csv
```

In [9]:

```
both_data = pd.read_csv("/kaggle/input/both-covid-data/both_covid_data.csv")
pcr = pd.read_csv("/kaggle/input/pcr-data/pcr_data.csv")
antibody = pd.read_csv("/kaggle/input/data-antibody/data_antibody.csv")
both_data.head()
```

Out[9]:

	Symptom-Throat Pain	Symptom-Dyspnea	Symptom-Fever	Symptom-Cough	Symptom-Headache	Symptom-Taste Disorders	Symptom-Olfactory Disorders	Symptom-Coryza	Gender	Are you a health professional?	Class
0	0	1	0	0	1	1	1	1	1	1.0	0
1	1	1	1	1	1	0	0	1	0	1.0	0
2	0	1	0	0	0	0	0	1	1	1.0	0
3	0	1	0	0	1	0	0	0	0	1.0	0
4	1	0	0	0	1	0	0	1	0	1.0	0

dictionary:

- **Class: 0 (positive) and 1(negative)**

1. **sum: sum of total values with (0 and 1)**
2. **pt: percentage of total values**
3. **p_p: percentage of positive test result(pt)**
4. **p_n: percentage of negative test result(pt)**

Setting cross tab with frequency distribution!

In [10]:

```
def set_crosstab(index_value,dt):
    data=pd.crosstab(index=dt[index_value],columns=dt["Class"])
    data['sum']=None
    data['p_t']=None
    data['p_p']=None
    data['p_n']=None
    data.iloc[:1,2:3]=data[0][0]+data[1][0]
    data.iloc[1:2,2:3]=data[0][1]+data[1][1]
    data.iloc[:1,3:4]=data.iloc[:1,2:3].values[0][0]*100/len(both_data)
    data.iloc[1:2,3:4]=data.iloc[1:2,2:3].values[0][0]*100/len(both_data)
    data.iloc[:1,4:5]=data.iloc[:1,:1].values[0][0]*data.iloc[:1,3:4].values[0][0]/data.
iloc[:1,2:3].values[0][0]
    data.iloc[1:2,4:5]=data.iloc[1:2,:1].values[0][0]*data.iloc[1:2,3:4].values[0][0]/da
ta.iloc[1:2,2:3].values[0][0]
    data.iloc[:1,5:6]=data.iloc[:1,1:2].values[0][0]*data.iloc[:1,3:4].values[0][0]/data
.iloc[:1,2:3].values[0][0]
    data.iloc[1:2,5:6]=data.iloc[1:2,1:2].values[0][0]*data.iloc[1:2,3:4].values[0][0]/d
ata.iloc[1:2,2:3].values[0][0]
    return data
```

Chi2 test analyze!

In [11]:

```
def test_statistic_chi2(value,data):
    test_result=pd.crosstab(index=data[value],columns=data["Class"])

    test_statistic, p_value, dof, expected=stats.chi2_contingency([test_result.iloc[0].v
alues,
                                                   test_result.iloc[1].values],correction=True)
    alpha = 0.01
    if p_value <= alpha:
        print('Dependent (reject H0)')
    else:
        print('Independent (fail to reject H0)')
```

BOTH DATA

Symptom- Throat Pain

In [12]:

```
set_crosstab('Symptom- Throat Pain',both_data)
```

Out[12]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Throat Pain						
0	322	147	469	17.5393	12.0419	5.49738
1	1015	1190	2205	82.4607	37.9581	44.5026

In [13]:

```
test_statistic_chi2('Symptom- Throat Pain',both_data)
```

```
Dependent (reject H0)
```

Symptom- Dyspnea

In [14]:

```
set_crosstab('Symptom- Dyspnea',both_data)
```

Out[14]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Dyspnea						
0	487	293	780	29.1698	18.2124	10.9574
1	850	1044	1894	70.8302	31.7876	39.0426

In [15]:

```
test_statistic_chi2('Symptom- Dyspnea',both_data)
```

```
Dependent (reject H0)
```

Symptom- Fever

In [16]:

```
set_crosstab('Symptom- Fever',both_data)
```

Out[16]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Fever						
0	1062	767	1829	68.3994	39.7158	28.6836
1	275	570	845	31.6006	10.2842	21.3164

In [17]:

```
test_statistic_chi2('Symptom- Fever',both_data)
```

```
Dependent (reject H0)
```

Symptom- Cough

In [18]:

```
set_crosstab('Symptom- Cough',both_data)
```

Out[18]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Cough						
0	816	572	1388	51.9073	30.5161	21.3912
1	521	765	1286	48.0927	19.4839	28.6088

In [19]:

```
test_statistic_chi2('Symptom- Cough',both_data)
```

```
Dependent (reject H0)
```

Symptom- Headache

In [20]:

```
set_crosstab('Symptom- Headache',both_data)
```

Out[20]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Headache						
0	215	49	264	9.87285	8.04039	1.83246
1	1122	1288	2410	90.1272	41.9596	48.1675

In [21]:

```
test_statistic_chi2('Symptom- Headache',both_data)
```

```
Dependent (reject H0)
```

Symptom- Taste Disorders

In [22]:

```
set_crosstab('Symptom- Taste Disorders',both_data)
```

Out[22]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Taste Disorders						
0	217	40	257	9.61107	8.11518	1.49589
1	1120	1297	2417	90.3889	41.8848	48.5041

In [23]:

```
test_statistic_chi2('Symptom- Taste Disorders',both_data)
```

```
Dependent (reject H0)
```

Symptom- Olfactory Disorders

In [24]:

```
set_crosstab('Symptom- Olfactory Disorders',both_data)
```

Out[24]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Olfactory Disorders						
0	302	39	341	12.7524	11.2939	1.45849
1	1035	1298	2333	87.2476	38.7061	48.5415

In [25]:

```
test_statistic_chi2('Symptom- Olfactory Disorders',both_data)
```

```
Dependent (reject H0)
```

Symptom- Coryza

In [26]:

```
set_crosstab('Symptom- Coryza',both_data)
```

Out[26]:

Class	0	1	sum	p_t	p_p	p_n
Symptom- Coryza						
0	129	20	149	5.57218	4.82423	0.747943
1	1208	1317	2525	94.4278	45.1758	49.2521

In [27]:

```
test_statistic_chi2('Symptom- Coryza',both_data)
```

```
Dependent (reject H0)
```

Gender

In [28]:

```
set_crosstab('Gender',both_data)
```

Out[28]:

Class	0	1	sum	p_t	p_p	p_n
Gender						
0	399	1049	1448	54.1511	14.9215	39.2296
1	938	288	1226	45.8489	35.0785	10.7704

In [29]:

```
test_statistic_chi2('Gender',both_data)
```

```
Dependent (reject H0)
```

Are you a health professional?

In [30]:

```
set_crosstab('Are you a health professional?',both_data)
```

Out[30]:

Class	0	1	sum	p_t	p_p	p_n
Are you a health professional?						
0.0	93	108	201	7.51683	3.47794	4.03889
1.0	1244	1229	2473	92.4832	46.5221	45.9611

In [31]:

```
test_statistic_chi2('Are you a health professional?',both_data)
```

```
Independent (fail to reject H0)
```