Chi-Square Test

47. The cross-tabulation data given below represent the number of males and females in a work group who feel overstressed and those who don't.

Overstressed	Women	Men
No	9	4
Yes	6	9

a. Write the hypotheses for the chi-square test for independence.

b. Find the expected frequencies.

c. Compute the chi-square statistic using a level of significance of 0.05.

Row Attribute: Feeling overstressed

Column Attribute: Gender

H_0: Feeling overstressed or not is independent of Gender

H_1: Feeling overstressed or not dependent on Gender

Overstressed	Women	Men	Total
No	9 $= O_{11}$	4 $= O_{12}$ 13 $= R_1$	
Yes	6 $= O_{13}$	9 $= O_{14}$ 15 $= R_2$	
Total	15 $= C_1$	13 $= C_2$ 28 $= n$	

$$E_{11} = \frac{R_1 C_1}{n}$$

$$E_{12} = \frac{R_1 C_2}{n}$$

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$= \frac{(9 - 6.96)^2}{6.96} + \frac{(4 - 6.03)^2}{6.03}$$

$$+ \frac{(6 - 8.03)^2}{8.03}$$

$$+ \frac{(9 - 6.96)^2}{6.96}$$

Overstressed	Women	Men
No	E11=13*15/28=6.96	E12=13*13/28=6.03
Yes	E21=15*15/28=8.03	E22=15*13/28=6.96

```
In [43]: obs = np.array([[9,4],
    ...:                  [6,9]])

In [44]: chi2_contingency(obs,correction=False)
Out[44]:
Chi2ContingencyResult(statistic=2.3924260355029583, pvalue=0.12192428411669565,
dof=1, expected_freq=array([[6.96428571, 6.03571429],
      [8.03571429, 6.96428571]]))
```

> 0.05

Conclusion: Feeling overstressed may be not be related to gender

49. The following cross-tabulation shows the number of people who rated a customer service representative as friendly and polite based on whether the representative greeted them first.

a. Write the hypotheses for the chi-square test for independence.

b. Find the expected frequencies.

c. Compute the chi-square statistic using a level of significance of 0.01.

d. Find the chi-square critical value and p-value and draw a conclusion.

Staff Greeting	Friendly/Polite	
	No	Yes
No	13	7
Yes	12	22

```
In [45]: obs = np.array([[13,7], [12,22]])

In [46]: print(chi2_contingency(obs,correction=False))
Chi2ContingencyResult(statistic=4.469403651115618, pvalue=0.034507033347004896
dof=1, expected_freq=array([[ 9.25925926, 10.74074074],        >0.01
    [15.74074074, 18.25925926]]))
```

∴ We don't reject Ho at 1 % l.o.s.
Conclusion:- Being friendly/Polite not related to staff greeting.

55. For the data in the Excel file *New Account Processing*, perform chi-square tests for independence to determine if certification is independent of gender and if certification is independent of having prior industry background.

```
In [60]:
    ...: a = pd.crosstab(index=acc['Gender'],
    ...:                  columns=acc['Certified'])
    ...: print(chi2_contingency(a,correction=False))
Chi2ContingencyResult(statistic=0.054727773462326004, pvalue=0.8150318082955932,
dof=1, expected_freq=array([[7.67741935, 6.32258065],
    [9.32258065, 7.67741935]]))
```

```
In [61]:
    ...: b = pd.crosstab(index=acc[' Prior Background'],
    ...:                  columns=acc['Certified'])
    ...: print(chi2_contingency(b, correction=False))
Chi2ContingencyResult(statistic=8.957200152788387, pvalue=0.0027637813584297775,
dof=1, expected_freq=array([[ 6.03225806,  4.96774194],
    [10.96774194,  9.03225806]]))
```

```
In [90]: ctab
Out[90]:
left              0     1
Department
IT             954   273
RandD          666   121
accounting     563   204
hr             524   215
management     539    91
marketing      655   203
product_mng    704   198
sales         3126  1010
support       1674   555
technical     2023   697
```

```
        Department  left  value
0               IT     0    954
1            RandD     0    666
2       accounting     0    563
3               hr     0    524
4       management     0    539
5        marketing     0    655
6      product_mng     0    704
7            sales     0   3126
8          support     0   1674
9        technical     0   2023
10              IT     1    273
11           RandD     1    121
12      accounting     1    204
13              hr     1    215
14      management     1     91
```