

Question:

The following table shows the Myers-Briggs personality preferences for a random sample of 389 past computer science graduates (from a University) in the listed professions

Occupation	Personality preference Type	
	E	I
Faculty	62	45
Data Scientist	56	81
Entrepreneur	94	51

Determine if there is any association between the listed occupations and the personality preferences at 5% level of significance.

Report:

Based on the question I will choose "Chi-square test of independence" to find is there any relationship between the listed occupations and the personality preferences.

H0: the listed occupations and the personality preferences are independent.

H1: the listed occupations and the personality preferences are not independent.

For calculate X^2 , first I will calculate expected value of each observation value. They are

58.31362	48.68638
74.66324	62.33676
79.02314	65.97686

And then I use the function to calculate X^2_{obs}

Chi-square statistic

$$\chi^2 = \sum_{k=1}^N \frac{\{Obs(k) - Exp(k)\}^2}{Exp(k)}.$$

$((58.31362-62)^2)/62+((48.68638-45)^2)/45+((74.66324-56)^2)/56+((62.33676-81)^2)/81+((79.02314-94)^2)/94+((65.97686-51)^2)/51$

The answer is [1] 17.82571

So that the **test statistics** is 17.82571.

The **null distribution** is Chi-Square test of independence with 2 degrees of freedom.

The **expression of p-value:** $P\{X^2 > 17.82571\}$

So that first I will build the matrix in R language by making the operations below:

```
x<-c(62, 45, 56, 81, 94, 51)
xmat<-matrix(x, byrow=T, ncol=2)
xmat
[,1] [,2]
[1,]  62  45
[2,]  56  81
[3,]  94  51
```

And then I use the function of “Chi-square test of independence” to get the answer. The degree of freedom is $(3-1)*(2-1)=2$.

```
chisq.test(xmat)
```

The answer shows that

Pearson's Chi-squared test

```
data: xmat
```

```
X-squared = 17.003, df = 2, p-value = 0.0002031
```

Because P-value is smaller than 0.05, so that reject H_0 at 5% level of significance.

So based on the data, the listed occupations and the personality preferences are **NOT** independent at 5% level of significance.