Degrees of freedom

- The number of independent pieces of information which are free to vary, that go into the estimate of a parameter is called the degrees of freedom.
- In general, the degrees of freedom of an estimate of a parameter is equal to
 the number of independent scores that go into the estimate minus the
 number of parameters used as intermediate steps in the estimation of the
 parameter itself (i.e. the sample variance has N-1 degrees of freedom, since
 it is computed from N random scores minus the only 1 parameter estimated
 as intermediate step, which is the sample mean).
- The number of degrees of freedom for 'n' observations is 'n-k' and is usually denoted by 'v', where 'k' is the number of independent linear constraints imposed upon them. It is the only parameter of the chi-square distribution.
- The degrees of freedom for a chi squared contingency table can be calculated as:

$$v = (Number of rows - 1)*(Number of columns - 1)$$

Chi Square formula

- The chi-squared test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.
- The value of χ 2 is calculated as:

$$\chi^2 = \sum \frac{(O_1 - E_1)^2}{E_1} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \dots + \frac{(O_n - E_n)^2}{E_n}$$

Where, O_1 , O_2 , O_3On are the observed frequencies and E_1 , E_2 , E_3 ... E_n are the corresponding expected or theoretical frequencies.

The observed frequencies are the frequencies obtained from the observation, which are sample frequencies.

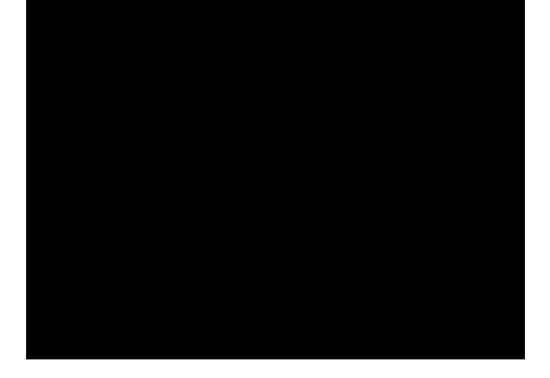
The expected frequencies are the calculated frequencies.

Alternate χ 2 Formula

Disease			
Exposure	Yes	No	Total
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	n

$$\chi_1^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

The alternate χ 2 formula applies only to 2x2 tables



Characteristics of Chi-Square test

- It is often regarded as a non-parametric test where no parameters regarding the rigidity of populations are required, such as mean and SD.
- 2. It is based on frequencies.
- 3. It encompasses the *additive property* of differences between observed and expected frequencies.
- 4. It tests the hypothesis about the independence of attributes.
- 5. It is preferred in analyzing complex contingency tables.



Steps in solving problems related to Chi-Square test

STEP 1

Calculate the expected frequencies

 $E = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$

STEP 2

 Take the difference between the observed and expected frequencies and obtain the squares of these differences (O-E)²

STEP 3

 Divide the values obtained in Step 2 by the respective expected frequency, E and add all the values to get the value according to the formula given by:

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$