⊛ Chi-square test (χ^2 test):

If 'x' is normally distributed with mean μ and standard deviation σ then $z = \frac{x-\mu}{\sigma}$ is a standard normal variate with 0 & 1

i.e $Z = \frac{x-\mu}{\sigma} \sim N(\mu, \sigma^2) = N(0,1)$.

then $z^2 = \left(\frac{x-\mu}{\sigma}\right)^2$ is a chi-square variate with 1 degree of freedom. Chi-square variate denoted by the symbol χ^2.

If $x_1, x_2, \ldots x_n$ are n independent normal variates with mean $\mu_1, \mu_2, \ldots \mu_n$ and standard deviations $\sigma_1, \sigma_2, \ldots \sigma_n$ respectively then

$$\chi^2 = \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 + \cdots + \left(\frac{x_n-\mu_n}{\sigma_n}\right)^2$$

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{x_i-\mu_i}{\sigma_i}\right)^2 ;$$ is a chi-square variate with n degree of freedom.

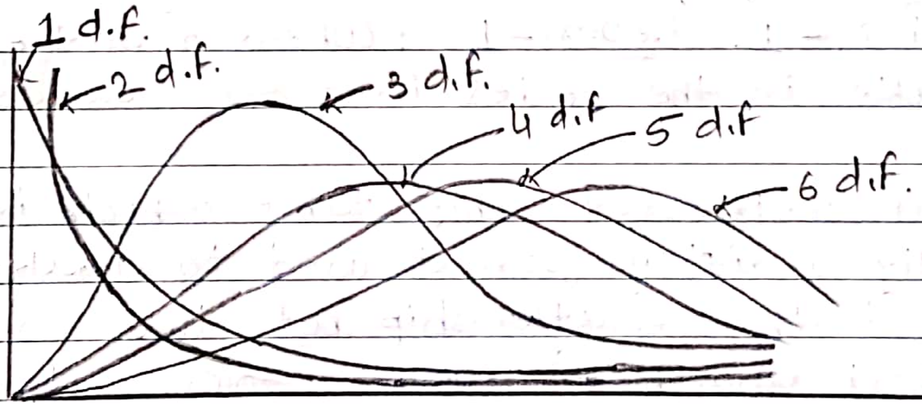
→ It is computed on the basis of frequencies in a sample and is applied only for qualitative data such as intelligence, colour, immunity, health, response to drug, etc.

→ If a random variable x has a chi-square distribution with n d.f., we write $x \sim \chi^2_{(n)}$ and P.D.F is

$$f(x) = \frac{1}{2^{n/2}\,\Gamma\frac{n}{2}} \cdot e^{-x/2} \cdot x^{\frac{n}{2}-1} \quad ; \quad 0 \le x < \infty .$$

Charecteristics of chi-square (χ^2) distribution

1) chi-square is always positively skewed.
   i.e χ^2 value is always positive.
2) Chi-square values increase with the increase in d.f.
3) The S.D. of χ^2 distribution is equal to $\sqrt{2v}$; where v is the degree of freedom (d.f.) & Mode $= v-2$.
4) The mean of χ^2 distribution is the number of d.f. (v)
5) The value of χ^2 lies bet^n 0 and ∞ i.e $0 \le \chi^2 < \infty$.
6) The sum of two χ^2 distribution is again a χ^2 distribution
   i.e If χ_1^2 and χ_2^2 are two independent distributions with d.f. n_1 and n_2 respectively then $\chi_1^2 + \chi_2^2$ is also a χ^2 distribution with d.f. (n_1+n_2).
7) For different d.f., the shape of curve will be differed
8) Chi-square (χ^2) is a statistic hypothesis and not a parameter.



Conditions for applying Chi-square (χ^2) test.

1) Every observation of the sample for this test should be independent of all other observations.
2) The expected frequency of any item should not be less than 5.

3) The total number of observations used for the test should be large. i.e $n \geq 50$.

4) Chi-square is wholly dependent on degree of freedom.

5) This test is used only for drawing inferences by testing hypothesis. It cannot be used for estimation of parameter or any other value.

6) The frequencies used in X^2 should be absolute and relative in terms.

7) The observations collected for x^2 test should be on random sampling.

⊛ Degree of freedom:

case-I. If the data is given in the form of a series of variables in a row or column, then the degree of freedom = number of items in the series − 1. i.e $v = n - 1$; where n is the number of variables in the series in a row or column.

case-II. For a contingency table (table which is essentially a display format used to analyze and record the relationship betn two or more categorical variables) with r rows and c columns, the degree of freedom $(v) = (r-1)(c-1)$

⊛ Chi-square test for goodness of fit:

This test is used to determine how well a theoretical distribution fits an empirical distribution. To investigate the agreement between observed and expected frequencies, we compute the value of the statistic.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$\left(\begin{array}{l} \sum O_i = \sum E_i = N \text{ (total frequency)} \\ E_i = \dfrac{\text{total frequency}}{\text{No of frequency}} \end{array} \right)$

Here χ^2 follows chi-square distribution with $(n-1)$ d.f.

We find $\chi^2_{n-1, 0.05}$ from the chi-square table.

If $\chi^2 \le \chi^2_{n-1, 0.05}$, we accept H_0: the given theoretical distribution fits the empirical distribution, otherwise we reject H_0.

Note: If we have estimated k parameters in fitting the given theoretical distribution the d.f. for χ^2 are $n-k-1$.

Ex-1 The following table gives the number of aircraft accidents that occurred during the various days of the week. Find whether the accidents are uniformly distributed over the week. (Given that value of χ^2 at 5% level of significance for 6 d.f. is 12.59)

Days	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Total.
No. of accidents.	14	16	8	12	11	9	14	84.

Sol^n: Taking the hypothesis that the accidents are uniformly distributed over the week, the Expected frequency for each day is

$$= \frac{\text{Total Frequency}}{\text{No. of Days}} = \frac{84}{7} = 12. = E_i$$

$$\therefore \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} ;$$

where O_i = observed frequency
E_i = expected frequency.

$$\therefore \chi^2 = \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$\chi^2 = \frac{(4+16+16+0+1+9+4)}{12} = 4.12.$$

As the calculated value of $\chi^2 = 4.12 < \chi^2_{6,0.05}$ which is 12.59.

$\therefore$ H_0 is accepted.

i.e the accidents are uniformly distributed over the week.

Ex-2 In an ecological study, water samples were collected in alternate months to study the phytoplantons population. The following table shows the number of organisms per cubic centimeter in the six samples.

sample No.	1	2	3	4	5	6
No. of organisms	79	84	102	64	90	85

Test the hypothesis that the number of organisms present in each sample does not depend on the particular sample.

Soln

Let H_0: The number of organisms does not depend on the water sample.

Now, the total number of organisms in the six samples is 504. Hence the mean number of organisms per sample is $504/6 = 84$.

∴ The expected frequency for each sample will be 84.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad ; \quad O_i = \text{observed frequency}, \quad E_i = \text{expected frequency}.$$

$$\therefore \chi^2 = \frac{(79-84)^2}{84} + \frac{(84-84)^2}{84} + \frac{(102-84)^2}{84} + \frac{(64-84)^2}{84} + \frac{(90-84)^2}{84} + \frac{(85-84)^2}{84}$$

$$\chi^2 = \frac{25 + 0 + 324 + 400 + 36 + 1}{84} = \frac{786}{84} = 9.36.$$

$d.f. = n-1 = 6-1 = 5.$

From the chi-square table, we find $\chi^2_{5,0.05} = 11.1$

Here $\chi^2 < \chi^2_{5,0.05}$

∴ We accept H_0 and conclude that the number of organisms present in each sample does not depend on the particular sample.

Ex-3 The demand for a particular spare part in a factory was found to very from day to day. In a sample study the following information was obtained :

Day	Mon	Tue.	Wed.	Thur.	Fri.	Sat.
No. of parts demanded	1124	1125	1110	1120	1126	1115

Test the hypothesis that the number of parts demanded has no association with the days of the week.

Soln: Let Ho: The number of spare parts demanded does not depend on the day of the week

Now, the total number of spare parts demanded during the six days = 6,720.

∴ We should except 6720/6 = 1,120 parts to be demanded on each of the six days.

$$\therefore \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(1124-1120)^2}{1120} + \frac{(1125-1120)^2}{1120} + \frac{(1110-1120)^2}{1120}$$

$$+ \frac{(1120-1120)^2}{1120} + \frac{(1126-1120)^2}{1120} + \frac{(1115-1120)^2}{1120}$$

$$= \frac{16+25+100+0+36+25}{1120} = \frac{202}{1120} = 0.1804$$

d.f. = n-1 = 6 - 1 = 5

From chi-square table, we find $\chi^2_{5,0.05} = 11.1$

Hence $\chi^2 < \chi^2_{5,0.005}$ ∴ We accept Ho.

Scanned with CamScanner