CHI-SQUARE TEST (χ^2 Test)

χ^2 describes the magnitude of discrepency betn theory and observation.

eg: In tossing of a coin 200 times, the theoritical considerations leads to a result giving head 100 times and tail 100 times but these results are rarely achievable.

If O_i ($i = 1, 2, 3, \ldots$) is a set of observed (experimental) frequencies and E_i ($i = 1, 2, \ldots, n$) is the corresponding set of expected (theoritical) frequencies.

Then

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

where $\sum O_i = \sum E_i = N$

degree of freedom (d.f.) $= n-1$

Note: ① If $\chi^2 = 0$ then observed and expected freq. agree exactly.

② If $\chi^2 > 0$ then they don't agree exactly.

Degree of Freedom

Degree of Freedom (d.f.), $\nu = n - k$

where n = total no. of obs.

k = no. of independent constraints

Q. The following table gives the no. of accidents that took place in an industry various days of the week. Test if accidents are uniformly distributed over the week.

Day	M	T	W	Th	F	S
No. of accidents	14	18	12	11	15	14

Soln:

Null Hypothesis H_0 : The accidents are uniformly distributed over the week.

Under H_0, the expected freq. of accidents on each day $= \frac{84}{6} = 14$

Observed Freq. O_i	14	18	12	11	15	14
Expected Freq. E_i	14	14	14	14	14	14
$(O_i - E_i)^2$	0	16	4	9	1	0

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i}$$

$$= \frac{0+16+4+9+1+0}{14}$$

$$= \frac{30}{14}$$

$$= 2.1428$$

Significance: 5%
$n = 6$ $d.f = n-1$ $= 5$

Conclusion: Table value of χ^2 at 5% level of significance for $(6-1=)$ 5 d.f is 11.070

$5\% = 0.05$
$\uparrow$
level of significance

Row – 5 (d.f)
Col^m – 0.05
Ans – 11.070

$\therefore$ calculated value of χ^2 is less than tabulated value

$\therefore$ H_0 is accepted

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

Q. A die is thrown 270 times and the results of these throws are given below

No. on the die	1	2	3	4	5	6
Frequency	40	32	29	59	57	59

Test whether the die is biased or not.

Q. Records taken of the number of male and female births in 800 families having four children are as follows

No. of male births	0	1	2	3	4
No. of female births	4	3	2	1	0
No. of families	32	178	290	236	94

Test whether the data are consistent with the hypothesis that the binomal law holds and the chance of male birth is equal to that of female namely $p = q = 1/2$

Soln: H_0: The data are consistent with the hypothesis of equal probability for male and female birth. i.e. $p = q = 1/2$.

The theoretical frequency is given by

$$N(r) = N \times P(X=r)$$

Here, N = total freq.

$N(r)$ = no. of families with r male children

$$P(X=r) = {}^nC_r\, p^r q^{n-r}$$

where p = prob. of a male birth

q = " " " female "

n = no. of children

$N(0)$ = No. families with 0 male children

$= N \times P(X=0)$

$= 800 \left[{}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 \right]$

$= 800 \times 1/2^4 \quad = 50$

$$N(1) = 800 \times P(x=1) = 800\left[{}^4C_1\left(\tfrac{1}{2}\right)^1\left(\tfrac{1}{2}\right)^3\right]$$

$$= 200$$

$$N(2) = 800 \times P(x=2) = 800\left[{}^4C_2\left(\tfrac{1}{2}\right)^2\left(\tfrac{1}{2}\right)^2\right]$$

$$= 300$$

$$N(3) = 800 \times P(x=3) = 800\left[{}^4C_3\left(\tfrac{1}{2}\right)^3\left(\tfrac{1}{2}\right)\right]$$

$$= 200$$

$$N(4) = 800 \times P(x=4) = 800\left[{}^4C_4\left(\tfrac{1}{2}\right)^4\right]$$

$$= 50$$

Observed Frequency O_i	32	178	290	236	94
Expected Frequency E_i	50	200	300	200	50
$O_i - E_i$	-18	-22	-10	36	44
$(O_i - E_i)^2$	324	484	100	1296	1936
$\frac{(O_i - E_i)^2}{E_i}$	6.48	2.42	0.333	6.48	38.72

$$\chi^2 = \frac{\Sigma (O_i - E_i)^2}{E_i} = 54.433$$

Conclusion: Table value of χ^2 at 5% level of significance for $5-1=4$ d.f. is 9.488

$\because$ calculated value of χ^2 is greater than the tabulated value

$\therefore$ H_0 is rejected

Hence, the data are not consistent with the hypothesis that the binomial law holds and that the probability of male and femal birth is same.

Q. Fit a Poisson distribution distribution to the following data and test the goodness of fit

x	0	1	2	3	4
f	109	65	22	3	1

Note :

① If the data is given in a series of 'n' numbers then d.f. $= n-1$

② In case of Binomial dist. , d.f $= n-1$

③ " " " Poisson dist. , d.f $= n-2$

④ " " " Normal dist. , d.f $= n-3$.