CHI-SQUARE (χ^2) TEST

When a coin is tossed 200 times, the theoretical considerations lead us to expect 100 heads and 100 tails. But in practice, these results are rarely achieved. The quantity χ^2 (the Greek letter *chi* squared, pronounced chi-square) describes the magnitude of discrepancy between theory and observation. If $\chi = 0$, the observed and expected frequencies completely coincide. The greater the discrepancy between the observed and expected frequencies, the greater the value of χ^2. Thus χ^2 **affords a measure of the correspondence between theory and observation**.

- Chi Square statistics are formulated to determine whether an observed number differs either from chance or from what was expected.

If O_i ($i = 1, 2, \ldots, n$) is a set of observed (experimental) frequencies and E_i ($i = 1, 2, \ldots, n$) is the corresponding set of expected (theoretical or hypothetical) frequencies, then χ^2 **is defined as**

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

where $\Sigma O_i = \Sigma E_i = N$ (total frequency) and degrees of freedom ($d.f.$) = ($n - 1$).

Note. (i) If $\chi^2 = 0$, the observed and theoretical frequencies agree exactly.

(ii) If $\chi^2 > 0$ they do not agree exactly.

Procedure for testing the significance of the difference between the observed and expected frequencies.

H_0 : There is no significant difference between the observed and the expected frequencies.

H_1 : There is significant difference between the observed and the expected frequencies.

The test statistic is $\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right]$. The expected frequencies are determined on the assumption that H_0 is true.

The number of degrees of freedom $= n-1$ where n is the number of classes.

From χ^2 –table we can find for a given degrees of freedom the table value of χ^2 for a given significance level (say $\alpha = .05$ or $\alpha = 0.01$)

Example 1:

A company keeps records of accidents. During a recent safety review, a random sample of 60 accidents was selected and classified by the day of the week on which they occured.

Day	MON	TUE	WED	THU	FRI
No. of accidents	8	12	9	14	17

Test whether there is any evidence that accidents are more likely on some days than others.

Solutions:

H_0 : Accidents are equally likely to occur on any day of the week.

H_1 : Accidents are not equally likely to occur on the days of the week.

Total number of accidents = 60

On the assumption H_0, the expected number of accidents on any day

$$= \frac{60}{5} = 12$$

Let O denote observed frequency and E denote expected frequency

O	E	$O-E$	$(O-E)^2$
8	12	-4	16
12	12	0	0
9	12	-3	9
14	12	2	4
17	12	5	25
60	60		54

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = \frac{54}{12} = 4.5$$

n = number of classes = 5

$\therefore$ number of degrees of freedom $= n - 1 = 5 - 1 = 4$

For 4 of degrees of freedom the table value of $\chi^2 = 9.488$.

But the calculated value of χ^2 is 4.5.

$\therefore$ Calculated value of $\chi^2 <$ the table value of χ^2.

Hence H_0 is accepted at 5% level. This means that the accidents are equally likely to occur on any day of the week.

Example 4:

A company produces a product of four sizes – small, medium, large and extra large. In the past, the demand for these sizes has been fairly constant at 20% for small, 45% for medium, 25% for large and 10% for extra large.

A random sample of 400 recent sales included 66 small, 172 medium, 109 large and 53 extra large.

Test whether there is evidence of significant change in demand for the different sizes.

Solution:

The expected frequencies are

$$\frac{20}{100}\times 400,\ \frac{45}{100}\times 400,\ \frac{25}{100}\times 400, \text{and } \frac{10}{100}\times 400 \text{(i.e.,) } 80,\ 180,\ 100,\ 40$$

Size	O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Small	66	80	–14	196	2.450
Medium	172	180	–8	64	0.356
Large	109	100	9	81	0.810
Extra large	53	40	13	169	4.225
	400				7.841

$$\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 7.841$$

Number of degrees of Freedom = 4 – 1 = 3

The table value of χ^2 for 3 df at 5% level = 7.81

Conclusion:

Since the calculated value of χ^2 is greater than the table value χ^2, H_0 is rejected 5% level.

$\therefore$ There is evidence of significant change in demand for the different sizes.

Goodness of Fit Test

Example 4. *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 binomial law holds and the chance of male birth is equal to that of female birth, namely $p = q = 1/2$.

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

We use binomial distribution to calculate theoretical frequency given by:

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

where N is the total frequency. $N(r)$ is the number of families with r male children:

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

where p and q are the probability of male and female births, n is the number of children.

$$N(0) = \text{No. of families with 0 male children} = 800 \times {}^{4}C_{0}\left(\frac{1}{2}\right)^{4} = 800 \times 1 \times \frac{1}{2^{4}} = 50$$

$$N(1) = 800 \times {}^{4}C_{1}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{3} = 200; \quad N(2) = 800 \times {}^{4}C_{2}\left(\frac{1}{2}\right)^{2}\left(\frac{1}{2}\right)^{2} = 300$$

$$N(3) = 800 \times {}^{4}C_{3}\left(\frac{1}{2}\right)^{1}\left(\frac{1}{2}\right)^{3} = 200; \quad N(4) = 800 \times {}^{4}C_{4}\left(\frac{1}{2}\right)^{0}\left(\frac{1}{2}\right)^{4} = 50$$

Observed frequency O_i	32	178	290	236	94
Expected frequency E_i	50	200	300	200	50
$(O_i - E_i)^2$	324	484	100	1296	1936
$\dfrac{(O_i - E_i)^2}{E_i}$	6.48	2.42	0.333	6.48	38.72

$$\chi^2 = \frac{\Sigma(\mathrm{O}_i - \mathrm{E}_i)^2}{\mathrm{E}_i} = 54.433.$$

Conclusion. The table value of χ^2 at 5% level of significance for 5 – 1 = 4 d.f. is 9.49.

Since the calculated value of χ^2 is greater than the tabulated value, H_0 is rejected.

I.e., the data are not consistent with the hypothesis that the binomial law holds and that the chance of a male birth is not equal to that of a female birth.

Note. Since the fitting is binomial, the degrees of freedom $\nu = n - 1$, *i.e.*, $\nu = 5 - 1 = 4$.

Example 5. *Verify whether the Poisson distribution can be assumed from the data given below:*

No. of defects	*0*	*1*	*2*	*3*	*4*	*5*
Frequency	*6*	*13*	*13*	*8*	*4*	*3*

Example 5. *Verify whether the Poisson distribution can be assumed from the data given below:*

No. of defects	*0*	*1*	*2*	*3*	*4*	*5*
Frequency	*6*	*13*	*13*	*8*	*4*	*3*

Sol. H_0 : The Poisson fit is a good fit to the data.

Mean of the given distribution $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{94}{47} = 2$

To fit a Poisson distribution we require m. Parameter $m = \overline{x} = 2$.

By the Poisson distribution the frequency of r success is

$$N(r) = N \times e^{-m} \cdot \frac{m^r}{r!}, \text{ N is the total frequency.}$$

$$N(0)=47\times e^{-2}\cdot\frac{(2)^0}{0!}=6.36\approx 6;\qquad N(1)=47\times e^{-2}\cdot\frac{(2)^1}{1!}=12.72\approx 13$$

$$N(2)=47\times e^{-2}\cdot\frac{(2)^2}{2!}=12.72\approx 13;\qquad N(3)=47\times e^{-2}\cdot\frac{(2)^3}{3!}=8.48\approx 9$$

$$N(4)=47\times e^{-2}\cdot\frac{(2)^4}{4!}=4.24\approx 4;\qquad N(5)=47\times e^{-2}\cdot\frac{(2)^5}{5!}=1.696\approx 2.$$

X	0	1	2	3	4	5
O_i	6	13	13	8	4	3
E_i	6.36	12.72	12.72	8.48	4.24	1.696
$\frac{(O_i - E_i)^2}{E_i}$	0.2037	0.00616	0.00616	0.02716	0.0135	1.0026

$$\chi^2 = \frac{\Sigma(\mathrm{O}_i - \mathrm{E}_i)^2}{\mathrm{E}_i} = 1.2864.$$

Conclusion. The calculated value of χ^2 is 1.2864. The tabulated value of χ^2 at 5% level of significance for $\gamma = 6 - 2 = 4$ d.f. is 9.49. Since the calculated value of χ^2 is less than that of the tabulated value, H_0 is accepted, *i.e.*, the Poisson distribution provides a good fit to the data.