χ^2 Test for Goodness of Fit:

It enables us to find if the deviation of the experiment from theory is just by chance or is it really due to the inadequacy of the theory to fit the observed data.

Let there be n samples (cells) from the Normal population with observed frequencies (experimental values) $f_1, f_2, ..., f_n$ and corresponding expected frequencies (theoretical or hypothetical values) $e_1, e_2, ..., e_n$

Then under H_0: There is no significant difference in the theoretical values and experimental data i.e. The theory is supported by experiments,

the test statistics is $\chi_t^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}$ which follow **chi square distribution** with k degree of freedom.

Note: χ^2 Test is valid if

1. The observations for all n samples is made independently.
2. The total of the frequency should be large i.e. $f_1 + f_2 + ... + f_n = N \geq 50$
3. No theoretical cells should have frequency less than 5 i.e. $f_i \geq 5 \;\&\; e_i \geq 5 \;\forall\, i$

If any cells has frequency less than 5, then it is pooled (grouped) with the preceding or succeeding frequency.

4. The degree of freedom of the test statistics
$= k =$ No. of cells (after grouping) $-$ No. of population parameter estimated from the data $- 1$

5. The **critical value** χ_α^2 at level of significance $= \alpha$ and degree of freedom $= k$ is given in the table.

6. The entire critical region lie towards the right tail. Thus there is no two tail test but only one tail test whether to accept H_0 or reject H_0.

H_0 is accepted iff $\chi_t^2 \leq \chi_\alpha^2$

① Theory predicts that the proportion of beans in 4 groups A, B, C and D should be 9:3:3:1. In an experiment among 1600 beans the numbers in the 4 groups were 882, 313, 287 and 118 respectively. Does the experiment support the theory? Test at 5% significance level.

Soln

Group	f	e	$\frac{(f-e)^2}{e}$
A	882	900	0.36
B	313	300	0.56
C	287	300	0.56
D	118	100	3.24
			4.72

H_0: The experiment support the theory.

$\Rightarrow$ proportion of beans in A, B, C, D should be 9:3:3:1

Group:	A	B	C	D
$p(x)$:	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

$\therefore$ Expected frequency is

$e = N \times p(x) = 1600 \times p(x)$

The test statistics is

$$\chi^2_{f} = \sum \frac{(f-e)^2}{e}$$

$$= 4.72$$

At L.S $= 5\%$ i.e. $\alpha = 0.05$,

d.f $= 4-0-1 = 3$

critical point is

$$\chi^2_{\alpha} = 7.815$$

$\Rightarrow \chi^2_{f} < \chi^2_{\alpha}$

$\Rightarrow$ Ho accepted

$\Rightarrow$ Experiment supports the theory that proportion of beans in the 4 groups A, B, C, D should be in ratio 9:3:3:1.

② Fit a poisson distribution to the data. Is poisson distribution is a good fit for the data?

x:	0	1	2	3	4	5	6
f:	143	90	42	12	9	3	1

Soln Ho: poisson distribution is a good fit to the given data.

Let $x \sim P(\lambda)$

$$\lambda = \text{mean}(x) = \frac{\sum xf}{\sum f} = \frac{267}{300} = 0.89$$

$$\therefore P\{x\} = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \frac{e^{-0.89}(0.89)^x}{x!}, \quad x = 0, 1, \ldots, 6$$

$\therefore$ Expected frequency is

$$e = N \times P\{x\}$$

$$= 300 \times \frac{e^{-0.89}(0.89)^x}{x!}, \quad 0, 1, \ldots, 6$$

x	f	e	$(f-e)^2/e$
0	143	123.2	3.18
1	90	109.7	3.54
2	42	48.8	0.95
3	12	14.5	
4	9	3.2	2.63
5	3 13 25	0.6 3.9 18.1	
6	1	0.1	
			10.3

$\therefore$ Test statistics is

$$\chi^2_{\text{cal}} = \sum \frac{(f-e)^2}{e} = 10.3$$

At L.S. = 5% i.e. $\alpha = 0.05$,

d.f = 4 − 1 − 1 = 2

critical point is

$\chi^2_\alpha = 5.992$

$\therefore \chi^2_t > \chi^2_\alpha$

$\Rightarrow$ H_0 is

$\Rightarrow$ Poisson distribution is not a good fit to the data.

χ^2 Test for Independence of Attribute:

When two characteristics of the population are under study, then data is represented in the form of contingency table

Y / X	y_1	y_2	....	y_n	Total
x_1	f_{11}	f_{12}	....	f_{1n}	f_1
x_2	f_{21}	f_{22}	....	f_{2n}	f_2
....	....	....	....	....	....
x_m	f_{m1}	f_{m2}	....	f_{mn}	f_m
Total	f^1	f^2	....	f^n	N

To test that two characteristics X & Y are dependent or independent we propose

H_o: X & Y are independent

Under this H_o

Expected frequency of $\left(x_i, y_j\right) = e_{ij} = \frac{f_i\, f^j}{N}$

and test statistics is

$$\chi_t^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}}$$

which follow chi square distribution with k degree of freedom.

Note:

1. No cells should have less than 5 frequency otherwise it is grouped with preceding or succeeding ones.

2. the degree of freedom for test statistics is
$k = (m-1)(n-1) -$ No. of cells grouped with other cells

3. For 2×2 contingency table,

Y / X	y_1	y_2	Total
x_1	a	b	$a+b$
x_2	c	d	$c+d$
Total	$a+c$	$b+d$	N

shortcut formula for test statistics is

$$\chi_t^2 = \frac{N\,|ad - bc|^2}{(a+c)(b+d)(a+b)(c+d)}$$

which follow chi square distribution with 1 **degree of freedom.**

4. If any of these cell frequencies a, b, c, d is less than 5 then we apply χ^2 **Test with Yate's correction** without grouping, and test statistics is given by

$$\chi_t^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{(a+c)(b+d)(a+b)(c+d)}$$

which follow chi square distribution with 1 **degree of freedom**.

(1) Table below shows the performances of students in two subjects Mathematics and physics. Test the hypothesis that performance in two subjects are independent.

		Grades in Mathematics		
		High	Medium	Low
Grades in physics	High	56	71	12
	Medium	47	163	38
	Low	4	52	85

Sol^n H_0: Performances in physics and Mathematics are independent.

				Total
	56	71	12	139
	47	163	38	248
	4	52	85	141
Total	107	286	135	528 = N

Under the Ho;
the Expected frequency is

$$e = \frac{f_i \times f^j}{N} = \frac{f_i \times f^j}{528}$$

& Test statistics is

$$\chi^2_{\#} = \sum \frac{(f-e)^2}{e}$$

f	e	$\frac{(f-e)^2}{e}$
56	28.2	27.4
71	75.3	0.25
12	35.5	15.6
47	50.3	0.22
163	134.3	6.1
$\left.\begin{matrix}38\\4\end{matrix}\right\}42$	$\left.\begin{matrix}63.49\\28.60\end{matrix}\right\}92$	27.2
42	76.4	15.5
85	36.1	66.24
		158.5

$\chi^2_f = 158.5$

At L.S. = 5 % i.e. $\alpha = 0.05$

& $d.f = (3-1)(3-1) - 1 = 2\times2-1$

$= 3$

critical point is

$\chi^2_\alpha = 7.815$

$\Rightarrow \chi^2_f > \chi^2_\alpha$

$\Rightarrow$ Ho is rejected

$\Rightarrow$ Performance in Mathematics and Physics are dependent.

(2) The manager of a chain of restaurants wants to know whether the customer satisfaction is related to the waiter. He takes a random sample of 100 customers, asking the name of the waiter and whether the service was good or poor. He then categorises the salaries of waiters as low and high. His results are shown below. Test whether the quality of service is related with waiter's salary.

		Waiter's salary	
		Low	High
Service quality	Good	24	28
	Poor	29	19

Sol^n Ho: There is no relation between the quality of service and waiter's salary.

24^a	28^b	52
29_c	19_d	48
53	47	$100 = N$

Using shortcut formula,

The test statistics is

$$\chi_f^2 = \frac{N\,|ad-bc|^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$= \frac{100\,(24\times19-29\times28)^2}{53\times47\times52\times48}$$

$$= 2.04$$

At L.S $= 5\%$ i.e. $\alpha = 0.05$,

d.f. $= (2-1)\times(2-1) = 1$

$\chi_\alpha^2 = 3.843$

$\Rightarrow \chi_f^2 < \chi_\alpha^2$

$\Rightarrow$ Ho is accepted

$\Rightarrow$ There is no relation between the quality of service and waiters salary.

(3) Investigate the association between the darkness of eyes colour in father and son from the following data.

		Father's eyes colour	
		Dark	Not dark
Son's eyes colour	Dark	48	362
	Not dark	80	3

Ho: There is no association between the darkness of eyes colour of father and son.

	a 48	b 362	410
	c 80	d 3	83
	128	365	493 = N

Using the shortcut formula with Yate's correction,
Test statistics is

$$\chi_f^2 = \frac{N\left(|ad-bc|-\frac{N}{2}\right)^2}{(a+c)(b+d)(a+b)(c+d)}$$

$$= \frac{493\left(|48\times3-80\times362|-\frac{493}{2}\right)^2}{128\times365\times410\times83}$$

$$= 0.0089$$

At L.S. = 5 % i.e. $\alpha = 0.05$,

d.f = 1

$\chi_\alpha^2 = 3.843 \Rightarrow \chi_f^2 < \chi_\alpha^2$

$\Rightarrow$ Ho is accepted

$\Rightarrow$ There is no relation.