χ^2 Test (CHI-SQUARE) – For Goodness of Fit

This test is used 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.

By this test we test whether differences between observed and expected frequencies are significant or not.

Application

(i) To test the goodness of fit

(ii) To test the 'Independence of attributes'.

(iii) To test if the hypothetical value of the pop. Variance is σ^2.

(iv) To test the homogenity of independent estimates of the population variances.

Conditions :

* The sample observations should be independent.
* No theoretical cell frequency should be less than 5.

Formula :

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

1. The following data gives the number of aircraft accidents that occured during the various days of a week. Find whether the accidents are uniformly distributed over the week.

Days :	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
No. of accidents :	14	16	8	12	11	9	14

Solution:

Let O → observed frequency - given

E → expected frequency - $\frac{\sum O_i}{n}$

$$\therefore \quad E = \frac{14+16+8+12+11+9+14}{7} = \frac{84}{7} = 12$$

O	E	(O - E)	$\frac{(O-E)^2}{E}$
14	12	2	0.333
16	12	4	1.333
8	12	-4	1.333
12	12	0	0
11	12	-1	0.083
9	12	-3	0.75
14	12	2	0.333

$$\sum \frac{(O-E)^2}{E} = 4.165$$

* H_0 : The accidents are uniformly distributed over the week

* H_1 : The accidents are not uniformly distributed over the week

* LOS = 5% df = n-1 = 7-1 = 6

* Table value = χ^2_α = 12.592

* Test Statistic $\chi^2 = \sum \left[\frac{(O-E)^2}{E}\right] = 4.165$

* Conclusion : Cal χ^2 < Tab χ^2_α

So we accept Ho.

2. The table below gives the number of aircraft accidents that occured during the various days of the week. Test whether the accidents are uniformly distributed over the week.

Days :	Mon	Tues	Wed	Thurs	Fri	Sat
No. of accidents :	14	18	12	11	15	14.

Soln: Let $O \rightarrow$ observed frequency - given

$E \rightarrow$ ~~Ef~~ Expected frequency $= \frac{\sum O}{n} = \frac{84}{6} = 14$

O	E	$(O-E)$	$\frac{(O-E)^2}{E}$
14	14	0	0
18	14	4	1.143
12	14	-2	0.286
11	14	-3	0.643
15	14	1	0.071
14	14	0	0

$$\sum \frac{(O-E)^2}{E} = 2.143$$

Ho : The accidents are uniformly distributed over the week

H₁ : The accidents are not uniformly distributed over the week.

Scanned by CamScanner

* LOS 5%, $df = n-1 = 6-1 = 5$

* Table value $\chi^2_\alpha = 11.070$

* Test Statistic $\chi^2 = \Sigma\left[\frac{(O-E)^2}{E}\right] = 2.143$

* Conclusion:

Cal χ^2 < Tab χ^2_α

So we accept Ho.

3. 4 coins were tossed 160 times and the following results were obtained.

No. of Heads :	0	1	2	3	4
Observed frequency :	17	52	54	31	6

under the assumption that the coins are unbaised. Find the expected frequencies of getting 0, 1, 2, 3, 4 heads and test the goodness of fit.

Soln: To find Expected frequency

$$P\{X=x\} = nCx\, p^x q^{n-x}, \quad x = 0,1,2,\ldots$$

$\boxed{p = \frac{1}{2},\ q = \frac{1}{2},\ n = 4}$

$$P\{X=x\} = 4Cx \left(\tfrac{1}{2}\right)^x \left(\tfrac{1}{2}\right)^{4-x}$$

$$= 4Cx \left(\tfrac{1}{2}\right)^4$$

$$P\{X=0\} = 4C0\left(\tfrac{1}{2}\right)^4 = 0.0625$$

$$P\{X=1\} = 4C1\left(\tfrac{1}{2}\right)^4 = 0.25$$

$$P\{X=2\} = 4C2\left(\tfrac{1}{2}\right)^4 = 0.375$$

$$P\{X=3\} = 0.25$$

$$P\{X=4\} = 0.0625$$

O	$E = 160 \times P(X=x_i)$	$(O-E)$	$\frac{(O-E)^2}{E}$
17	10		4.9
52	40		3.6
54	60		0.6
31	40		2.025
06	10		1.6

$$\sum \frac{(O-E)^2}{E} = 12.725$$

* Ho : The coins are unbiased
* H_1 : The coins are biased
* LOS = 5 DF = n-1 = 5-1 = 4
* Table value of χ^2_α = 9.488
* Test statistic $\chi^2 = \sum \frac{(O-E)^2}{E} = 12.725$

Cal χ^2 ≮ Tab. χ^2

So we reject Ho. The coins are biased.

(4) HW : Five coins are tossed 320 times. The no. of heads observed is gn below :

No. of Heads :	0	1	2	3	4	5
Frequency :	15	45	85	95	60	20

Examine whether the coin is unbiased. use 5% LOS.

5. A sample analysis of examination results of 500 students was made. It was found that 220 students have failed. 170 have secured a third class. 90 have secured a second class and the rest, a first class. So do these figures support the general belief that the above categories are in the ratio 4:3:2:1 respy ?

Soln:

Expected frequency of the 4 classes be

$\frac{4}{10} \times 500$, $\frac{3}{10} \times 500$, $\frac{2}{10} \times 500$, $\frac{1}{10} \times 500$

= 200 , 150 , = 100 , = 50

O	E	O−E	$\frac{(O-E)^2}{E}$
220	200	20	2
170	150	20	2.667
90	100	−10	1
20	50	−30	18

$$\sum \frac{(O-E)^2}{E} = 23.667$$

* Ho : The results are in the ratio 4:3:2:1

* H1 : " " are not "

* $\alpha = 5\%$ df $= n-1 = 4-1 = 3$

* Table value of $\chi^2_\alpha = 7.815$

* Test statistic $\chi^2 = \sum \left[\frac{(O-E)^2}{E}\right] = 23.667$

Conclusion

Cal $\chi^2 \nless$ Tab χ^2_α

We reject Ho.

χ^2 – test to test the Independence of attributes

Let us consider two attributes A and B. A divided into r classes $A_1, A_2, \dots A_r$ and B divided into s classes $B_1, B_2, \dots B_s$. If this is expressed as $r \times s$ matrix, the matrix is called $r \times s$ Contingency table.

Note: For a 2×2 contingency table

a	b
c	d

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

Observed Freq	Attri B_1	Attri B_2	Total
Attribute A_1	a	b	a+b
Attribute A_2	c	d	c+d
Total	a+c	b+d	a+b+c+d = N

i.e, $$\chi^2 = \frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

Calculation of Expected Frequency

	Attribute B_1	Attribute B_2
Attribute A_1	$\frac{(a+b)(a+c)}{N}$	$\frac{(a+b)(b+d)}{N}$
Attribute A_2	$\frac{(a+c)(c+d)}{N}$	$\frac{(b+d)(c+d)}{N}$

Degrees of freedom $= (r-1)(s-1) = \nu$

1. Out of 8000 graduates in a town 800 are females, out of 1600 graduate employees 120 are females. Use χ^2 to determine if any distinction is made in appointment on the basis of sex. Value of χ^2 at 5% Los for one degree of freedom is 3.84.

Solution :

	Male	Female	Total
Graduates in town	7200	800	8000
Graduate employees	1480	120	1600
Total	8680	920	9600

* Ho: There is no significance difference b/w male and female

* H_1 : There is a significance difference b/w male and female.

* LOS = 5% , DF = (r-1)(s-1) = (2-1)(2-1) = 1

* Table value $\chi^2 = 3.84$

* Test statistic $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$\text{Expected frequency} = \frac{\text{Corresponding row total} \times \text{Column total}}{\text{Grand total}}$$

For 7200 $E = \frac{8000 \times 8680}{9600} = 7233.33$

For 800 $E = \frac{8000 \times 920}{9600} = 766.67$

For 1480, $E = \frac{(1600)(8680)}{9600} = 1446.67$

120, $E = \frac{(1600)(920)}{9600} = 153.33$

O	E	(O−E)	$\frac{(O-E)^2}{E} = \chi^2$
7200	7233.33	−33.33	0.1536
800	766.67	33.33	1.4490
1480	1446.67	33.33	0.7679
120	153.33	−33.33	7.2451
		Total	$9.6156 = \chi^2$

Conclusion: Cal χ^2 $\not<$ Tab. χ^2. So we Reject Ho.

So we accept Ho.

2. 1000 Students at college level were graded according to their IQ and their economic conditions. What conclusion can you draw from the following data.

Economic Condition	IQ Level	
	High	Low
Rich	460	140
Poor	240	160