CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES

In statistics, sometimes we have to deal with attributes or qualitative characters, which cannot be measured accurately, although items can be divided into two or more categories w.r.t. attributes.

Let A & B are two attributes of the population. A can be divided into m categories and B can be divided into n categories. The data can be shown in a form of two way table with m rows and n columns as bivariate frequency distribution.

This two way freq. table for attributes is known as (m x n) contingency table

3 x 4 contingency table

Attribute B

Attribute A

-	B_1	B_2	B_3	B_4	Total
A_1	$(A_1 B_1)$	$(A_1 B_2)$	$(A_1 B_3)$	$(A_1 B_4)$	(A_1)
A_2	$A_2 B_1$	$(A_2 B_2)$	$(A_2 B_3)$	$(A_2 B_4)$	(A_2)
A_3	$(A_3 B_1)$	(A_3, B_2)	$(A_3 B_3)$	(A_3, B_4)	(A_3)
Total	(B_1)	(B_2)	(B_3)	(B_4)	N = Total

Two attributes A and B are said to be independent they are not related to each other.

If two attributes are not independent, they are associated, on the basis of cell frequency.

It is required to test that two attributes A and B are associated or not.

We will assume the Null Hypothesis H_0 "that attributes are independent", under this assumption, the expected frequency of any cell is given by

$$f_e = \frac{\text{Row total} \times \text{column total}}{\text{Total freq.}}$$

$$\boxed{f_e = \frac{(A_i)(B_j)}{N}}$$

Test statistic:- $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

with degree of freedom :-

From contingency table :-

$\nu = (\text{no. of rows} - 1)(\text{no. of column} - 1)$

$\boxed{\nu = (r-1)(c-1)}$

If $\chi^2_{cal} < \chi^2_{tab}$, H_0 accepted

ie attributes are said to be independent

If $\chi^2_{cal} > \chi^2_{tab}$, then H_0 is rejected.

classmate

Date
Page

Q1 A sample of 400 students of undergraduate and 400 students of post graduate classes was taken to know their opinion about autonomous colleges. 290 of the undergraduate and 310 of the postgraduates students favoured the autonomous status. Present these facts in the form of a table and test at 5% level of significance, that the opinion regarding autonomous status of college is independent of level of class of students.

Ans:- N = 800

	Opinion about autonomous colleges		
	favoured	Not favoured	Total
Undergraduate	290	110	400
Postgraduate	310	90	400
Total	400	400	800

(i) **Null Hypothesis:-** there is no relation between the classes of students and opinion ie two attributes are independent.

(ii) **Alternate Hypo:-** H_1: there is relation between the classes of students and opinion.

classmate

Date
Page

Test statistic:-

Observed frequency (f_o)	Expected frequency $f_e = \frac{(A_i)(B_j)}{N}$	$\frac{(f_o - f_e)^2}{f_e}$
290	$\frac{400 \times 600}{800} = 300$	0.33
110	$\frac{400 \times 200}{800} = 100$	1.00
310	$\frac{400 \times 600}{800} = 300$	0.33
90	$\frac{400 \times 200}{800} = 100$	1.00

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\boxed{\chi^2 = 2.66}$$

Critical value:-

$$\nu = (r-1)(c-1)$$

$$= (2-1)(2-1) = 1$$

$$\chi^2_{0.5}(\nu = 1) = 3.81$$

as $\chi^2_{cal} < \chi^2_{0.5}$ ie H_0 is accepted ie there is no relation between the classes of students and their opinion.

Example-2:- In an experiment on immunisation of cattle from tuberculosis the following results are obtained.

	Affected	Not Affected	Total
Inoculated	267	27	294
Not inoculated	757	155	912
Total	1024	182.	1206

Use χ^2 test to determine the efficiency of vaccine in preventing the tuberculosis.

Solution:-

$N = 1206$

Null Hypothesis:- H_0: There is no relation b/w inoculation and effect on ~~disease~~ disease ie two attributes are independent.

Alternate Hypothesis:- H_1: There is relation b/w inoculation and effect on disease.

$\boxed{\alpha = .05}$

Observed frequency f_o	Expected freq $f_e = \frac{(A_i)(B_j)}{N}$	$\frac{(f_o - f_e)^2}{f_e}$
267	$\frac{294 \times 1024}{1206} = 250$	1.156
27	$\frac{294 \times 182}{1206} = 44$	6.568
757	$\frac{912 \times 1024}{1206} = 774$	0.37
155	$\frac{912 \times 182}{1206} = 138$	2.09

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$\alpha = .05$

Critical value : $v = (r-1)(c-1)$

$$v = (2-1)(2-1)$$

$$\boxed{v=1}$$

$$\chi^2_{tab}(v=1) = 3.84$$

as $\chi^2_{cal} > \chi^2_{tab}$ ie Null Hypo is rejected at 5% level of significance

ie vaccine is effective in preventing tuberculosis.

classmate

Date ______
Page ______

Ex-3:- The following contingency table for hair color and eye color is given. Find the value of χ^2. Is there good association b/w the two

	Hair color			
Eye color	fair	Brown	Black	Total
Blue	15	5	20	40
grey	20	10	20	50
Brown	25	15	20	60
Total	60	30	60	150

Solution:- $N = 150$

Null Hypothesis:- There is no association b/w two attributes, hair and eye color.

Alternate Hypothesis:- there is association b/w two attributes, hair and eye color.

$\boxed{\alpha = 0.05}$

Test statistic

Observed freq (f_o)	Expected (f_e)	$\frac{(f_o-f_e)^2}{f_e}$
15	$\frac{40\times60}{150} = 16$	0.0625
5	$\frac{40\times30}{150} = 8$	1.125
20	$\frac{40\times60}{150} = 16$	1
20	$\frac{50\times60}{150} = 20$	0
10	$50\times30/150 = 10$	0
20	$50\times60/150 = 20$	0
25	$60\times60/150 = 24$	.042
15	$60\times30/150 = 12$	0.75
20	$60\times60/150 = 24$	0.666

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = 3.6465$$

Critical value:-

$$v = (r-1)(c-1)$$
$$= (3-1)(3-1)$$
$$v = 4$$

$$\chi^2_{.05}(v=4) = 9.49$$

Since $\chi^2_{cal} < \chi^2_{.05}$ ie Ho is accepted

ie there is no ~~sig~~ association between two attributes hair and eye color.

Q:- The following table gives the level of education and the marriage adjustment score for a sample of married women

Level of education	Marriage Adjustment				Total
	very low	low	high	very high	
college	24	97	62	58	241
high school	22	28	30	41	121
Middle School	32	10	11	20	73
Total	78	135	103	119	435

can you conclude that from the above data the higher the level of education, the greater is the degree of adjustment in marriage.

Sol^{n}:- H_0: There is no relation between the level of education and adjustment in marriage. i.e. two attributes are independent.

H_1: there is relation b/w level of education and adjustment in marriage.

$\alpha = .05$ (assumption)

Test Statistic :-

Observed frequencies (f_o)	Expected frequencies (f_e)	$\frac{(f_o - f_e)^2}{f_e}$
24	$\frac{241 \times 78}{435} = 43$	8.3953
97	$\frac{241 \times 135}{435} = 75$	6.4533
62	$\frac{241 \times 103}{435} = 57$	0.4386
58	$\frac{241 \times 119}{435} = 66$	.9697
22	$\frac{121 \times 78}{435} = 22$	0
28	$\frac{121 \times 135}{435} = 37$	2.1892
30	$\frac{121 \times 103}{435} = 29$	0.0345
41	$\frac{121 \times 119}{435} = 33$	1.9394
32	$\frac{73 \times 78}{435} = 13$	27.7692
10	$\frac{73 \times 135}{435} = 23$	7.3478
11	$\frac{73 \times 103}{435} = 17$	2.1176
20	$\frac{73 \times 119}{435} = 20$	0.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$[\chi^2 = 57.713]$$

Critical value :- $(r-1)(c-1)$

$= (3-1)(4-1) = 6$

$$\chi^2_{.05}(\nu = 6) = 12.59$$

Decision :- Since $\chi^2_{cal} > \chi^2_{.05}(\nu = 6)$ ie

Null Hypothesis is rejected at 5% level of significance ie level of education and adjustment in marriage are related and higher the level of education, the greater the degree of adjustment in marriage.