Exercise.8

Chi square test – test for association and goodness of fit

χ^2 – test for goodness of fit

If 0_i, (i=1,2,…..,n) is a set of observed (experimental frequencies)and Ei (i=1,2,…..,n) is the corresponding set of expected (theoretical or hypothetical) frequencies, then,

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2 (n-1)\,\text{df}$$

Example1:

The number of yiest cells counted in a haemocytometer is compared to the theoretical value is given below. Does the experimental result support the theory?

No. of Yeast cells in the square	Obseved Frequency	Expected Frequency
0	103	106
1	143	141
2	98	93
3	42	41
4	8	14
5	6	5

Solution:

H_0: the experimental results support the theory

H_1: the esperimental results does not support the theory.

Level of significance=5%

Test Statistic:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(n-1) \text{ df}$$

O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
103	106	-3	9	0.0849
143	141	2	4	0.0284
98	93	5	25	0.2688
42	41	1	1	0.0244
8	14	-6	36	2.5714
6	5	1	1	0.2000
400	400			**3.1779**

$\therefore \chi^2 = 3.1779$

Table value:

χ^2 (6-1=5 at 5 % l.os)= 11.070

Inference

$\chi^2_{cal} < \chi^2_{tab}$

We accept the null hypothesis.

(i.e) there is a good correspondence between theory and experiment.

χ^2 test for independence of attributes

Formula:

$$\chi^2 = \sum_{i=j}^{m} \sum_{j=1}^{n} \frac{\left(o_{ij} - E_{ij}\right)}{E_{ij}} \sim \chi^2(m-1)(n-1) \text{ df}$$

O_{ij} – observed frequencies

E_{ij} – Expected frequencies

m= number of rows

n= number of columns

$$\sum Oij = \sum Eij$$

Exapmle 2:

The severity of a disease and blood group were studied in a research projest. The findings sre given in the following table, knowmn as the m xn contingency table. Can this severity of the condition and blood group are associated.

Severity of a disease classified by blood group in 1500 patients.

Condition	Blood Groups				Total
	O	A	B	AB	
Severe	51	40	10	9	110
Moderate	105	103	25	17	250
Mild	384	527	125	104	1140
Total	540	670	160	130	1500

Solution:

H_0: The two attributes severity of the condition and blood groups are not associated.

H_1: The two attributes severity of the condition and blood groups are associated.

Calculation of Expected frequencies

Condition	Blood Groups				Total
	O	A	B	AB	
Severe	39.6	49.1	11.7	9.5	110
Moderate	90.0	111.7	26.7	21.7	250
Mild	410.4	509.2	121.6	98.8	1140

Total	540	670	160	130	1500

Test statistic:

$$\chi^2 = \sum_{i=j}^{m} \sum_{j=1}^{n} \frac{\left(o_{ij} - E_{ij}\right)}{E_{ij}} \sim \chi^2 (m-1)(n-1) \text{ df}$$

Here m= 3.,n =4

Calculations:

O_i	E_i	O_i-E_i	$(O_i\text{-}E_i)^2$	$(O_i\text{-}E_i)^2/E_i$
51	39.6	11.4	129.96	3.2818
40	49.1	-9.1	82.81	1.6866
10	11.7	-1.7	2.89	0.2470
9	9.5	-0.5	0.25	0.0263
105	90.0	15	225.00	2.5000
103	111.7	-8.7	75.69	0.6776
25	26.7	-1.7	2.89	0.1082
17	21.7	-4.7	22.09	1.0180
384	410.4	-26.4	696.96	1.6982
527	509.2	17.8	316.84	0.6222
125	121.6	3.4	11.56	0.0951
104	98.8	5.2	27.04	0.2737
				12.2347

$\therefore \chi^2$ =12.2347

Table value:

χ^2 (3-1)(4-1)= χ^2 (6) at 5 % l.os)= 12.59

Inference

$\chi^2_{cal} < \chi^2_{tab}$

We accept the null hypothesis.

(i.e) the two attributes severity of the condition and blood group are independent.

2x2 – contingency table

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2 \text{ (2-1) (2-1) df} \quad = \chi^2 \text{(1) df}$$

Example 3:

In order to determine the possible effectof a chemical treatment on the rate of germination of cotton seeds a pot culture experiment was conducted. The results are given below

Chemical treatment and germination of cotton seeds

	Germinated	Not germinated	Total
Chemically Treated	118	22	140
Untreated	120	40	160
Total	238	62	300

Does the chemical treatrment improve the germination rate of cotton seeds at 1 % level?

Solution:

H_0:The chemical treatment does not improve the germination rate of cotton seeds.

H_1: The chemical treatment improves the germination rate of cotton seeds.

L.O.S = 1 %

Test statistic

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2 \text{ (2-1) (2-1) df} \quad = \chi^2 \text{(1) df}$$

$$\chi^2 = \frac{300(118 \times 40 - 22 \times 120)^2}{140 \times 160 \times 62 \times 238} = 3.927$$

Table value:

χ^2 (1) df at 1 % L.O.S = 6.635

Inference

$\chi^2_{cal} < \chi^2_{tab}$

We accept the null hypothesis.

(i.e)The chemical treatmentwill not improve the germination rate of cotton seeds significantly.

Yates correction for continuity

Then use, $\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2$ (1) df

(or)

Directly use the χ^2 - statistic as

$$\chi^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2 \text{ (1) df}$$

Example 4:

In an experiment on the effect of a growth regulator on fruit setting in muskmelon the following results were obtained. Test whether the fruit setting in muskmelon and the application of growth regulator are independent at 1% level.

	Fruit set	Fruit not set	Total
Treated	16	9	25

Control	4	21	25
Total	20	30	50

Solution:

H_0:Fruit setting in muskmelon does not depend on the application of growth regulator.

H_1: Fruit setting in muskmelon depend on the application of growth regulator.

L.O.S = 1 %

Tet statistic

$$\chi^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2 \text{ (1) df}$$

$$\chi^2 = \frac{50\left[|16 \times 21 - 9 \times 4| - \frac{50}{2}\right]^2}{25 \times 25 \times 20 \times 30} = 10.08$$

Table value:

χ^2 (1) df at 1 % L.O.S = 6.635

Inference

$\chi^2_{\text{cal}} > \chi^2_{\text{tab}}$

We reject the null hypothesis.

(i.e) Fruit setting in muskmelon is influenced by the growth regulator.

Learning Exercise

1. The theory predicts the proportion of beans in the 4 groups A, B, C, D should be 9:3:3:1. In an experiment among 1600 beans, the number in the four groups were 882, 313, 287 and 118. Does the experimental result support the theory.

2. A study was conducted, among 100 professors from 3 different divisions for the preference on beverages of 3 categories test if there is any relationship between the field of teaching and preference of beverage.

Field of teaching				
Beverage	Business	Social Sciences	Agri	Total
Tea	20	10	10	40
Coffee	10	10	15	35
Cold drinks	10	8	7	25
Total	40	28	32	100

3. A random sample of 600 students from Delhi University are selected and asked their opinion ab out autonomous Status of Colleges. The results were given below. Test the hypothesis at 5% level that opinions are independent of class groupings.

Class grouping	Favour of	Against	
Commerce	120	80	200
Science	130	70	200
Arts	70	30	100
Total	400	200	600

4. In a survey of preference of new coverage 100 pearsons are collected and taste preference of average was surveyed according to sex of the person. We conclude that the taste preference and sex of the person are associated.

	Male	Female	

Favour	35	25	60
Against	25	15	40
Total	60	40	100

5. Two new food stuffs were introduced and public opinion was sought based on the taste of the food stuff. The results are given below. Examine whether there is an association between the category of the food stuffs and the taste of the food stuffs.

Category			
	A	B	
Tasty	620	380	1000
Not tasty	550	450	1000
Total	1170	830	2000