Sri SAI RAM ENGINEERING COLLEGE
INSTITUTE OF TECHNOLOGY
West Tambaram, Chennai - 44

SAIRAM DIGITAL RESOURCES

YEAR II

SEM IV

MA8391

PROBABILITY AND STATISTICS

DEPARTMENT OF INFORMATION TECHNOLOGY

UNIT III

TESTING OF HYPOTHESIS

3.7 CHI-SQUARE PROBLEMS BASED ON INDEPENDECE OF ATTRIBUTES

SCIENCE & HUMANITIES

χ^2 -test is used to test the independence of attributes:

An attributes means a equality or characteristic. χ^2 - test is used to test whether the two attributes are associated or independent. Let us consider two attributes A and B. A is divided into three classes and B is divided into three classes.

		Attribute B			
		B1	B2	B3	Total
Attribute A	A1	a11	a12	a13	R1
	A2	a21	a22	a23	R2
	A3	a31	a32	a33	R3
	Total	C1	C2	C3	N

Now, under the null hypothesis H_0: The attributes A and B are independent and we calculate the expected frequency E_{ij} for varies cells using the following formula

$$E_{ij} = \frac{R_i \times C_j}{N}, \; i = 1, 2, \ldots r, \; j = 1, 2, \ldots s$$

$E(a_{11}) = \frac{R_1 \times C_1}{N}$	$E(a_{12}) = \frac{R_1 \times C_2}{N}$	$E(a_{13}) = \frac{R_1 \times C_3}{N}$	R_1
$E(a_{21}) = \frac{R_2 \times C_1}{N}$	$E(a_{22}) = \frac{R_2 \times C_2}{N}$	$E(a_{23}) = \frac{R_2 \times C_3}{N}$	R_2
$E(a_{31}) = \frac{R_3 \times C_1}{N}$	$E(a_{32}) = \frac{R_3 \times C_2}{N}$	$E(a_{33}) = \frac{R_3 \times C_3}{N}$	R_3
C_1	C_2	C_3	**N**

and we compute

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(0_{ij}-E_{iJ}.\right)^2}{E_{ij}}$$

Which follows χ^2 distribution with $\mathrm{n} = $ (r-1)(s-1) degrees of freedom at 5% or 1% level of significance.

1. Calculate the expected frequencies for the following data presuming two attributes viz., conditions of home and condition of child as independent.

	Condition of home		
		Clean	Dirty
Condition of Child	Clean	70	50
	Fair	80	20
	Dirty	35	45

Use Chi-Square test at 5% level of significance to state whether the two attributes are independent.

Solution:

Null hypothesis $H\,0$: Conditions of home and conditions of child are independent.

Alternate hypothesis $H1$: Conditions of home and conditions of child are not independent.

Level of significance: $a = 0.05$

		Condition of home		Total
Condition of Child		Clean	Dirty	
	Clean	70	50	120
	Fair	80	20	100
	Dirty	35	45	80
Total		185	115	300

The test statistics: $\chi^2 = \sum_{i=1}^{r} \sum_{i=1}^{s} \frac{\left(O_{jj}-E_{ij}\right)^2}{E_{ij}}$ Analysis:

$$\text{Expected Frequency} = \frac{\text{Corresponding row tota1}\times column\ total}{\text{Grand Tota1}}$$

$$\text{Expected Frequency for } 70 = \frac{120\times185}{300} = 74\ ,$$

$$\text{Expected Frequency for } 80 = \frac{100\times185}{300} = 61.67,$$

$$\text{Expected Frequency for } 35 = \frac{80\times185}{300} = 49.33\ ,$$

$$\text{Expected Frequency for } 50 = \frac{120\times115}{300} = 46,$$

$$\text{Expected Frequency for } 20 = \frac{100\times115}{300} = 38.33\ ,$$

$$\text{Expected Frequency for } 45 = \frac{80\times115}{300} = 30.67$$

O_{ij}	E_{ij}	O_{ij} - E_{ij}	$(O_{ij} - E_{ij})^2$	$\dfrac{(O_{ij} - E_{ij})^2}{E_{ij}}$
70	74	-4	16	$\dfrac{16}{74} = 0.216$
50	46	4	16	0.348
80	61.67	18.33	335.99	5.448
20	38.33	-18.33	335.99	8.766
35	49.33	-14.33	205.35	4.163
45	30.67	14.33	205.35	6.695
Total				25.636

$\alpha = 0.05$

Degrees of freedom

$= (r-1)(c-1) = (3-1)(2-1) = 2 \; \chi_{\alpha}^{2} = 5.991$

$\therefore \chi^{2} = 25.636$

Conclusion:

Since $\chi^{2} > \chi_{\alpha}^{2}$, we Reject our Null Hypothesis H_0.

Hence, Conditions ofhome and conditions of child are not independent.

2. The following contingency table presents the reactions of legislators to a tax plan according to party affiliation. Test whether party affiliation influences the reaction to the tax plan at 0.01 level of signification.

	Reaction			
Party	Infavour	Neutral	Opposed	Total
Party A	120	20	20	160
Party B	50	30	60	140
Party C	50	10	40	100
Total	220	60	120	400

Soln:

Given

Null hypothesis H_0 : Party affiliation and tax plan are independent.

Alternate hypothesis H_1 : Party affiliation and tax plan are not independent.

Level of significance: $\alpha = 0.05$

The test statistics: $\chi^2 = \sum_{i=1}^{r} \quad \sum_{i=1}^{s} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

Reaction				
Party	Infavour	Neutral	Opposed	Total
Party A	120	20	20	160
Party B	50	30	60	140
Party C	50	10	40	100
Total	220	60	120	400

E(120)= $\frac{160x220}{400} = 88$

E(20)= $\frac{160x60}{400} = 24$

E(20)= $\frac{160x120}{400} = 48$

E(50)= $\frac{140x220}{400} = 77$

E(30)= $\frac{140x60}{400} = 21$

E(60)= $\frac{140x120}{400} = 42$

E(50)= $\frac{100x220}{400} = 55$

E(10)= $\frac{100x60}{400} = 15$

E(40)= $\frac{120x100}{400} = 30$

120	88	32	1024	11.64
20	24	-4	16	0.67
20	48	-28	784	16.33
50	77	-27	729	9.47
30	21	9	81	3.86
60	42	18	324	7.71
50	55	-5	25	0.45
10	15	-5	25	1.67
40	30	10	100	3.33

$$\therefore \chi^2 = 55.13$$

$\alpha = 0.05$

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

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

$$\therefore \chi_\alpha^2 = 13.28$$

Conclusion:

Since $\chi^2 > \chi_\alpha^2$, we Reject our Null Hypothesis H_0

Hence, the Party Affiliation and tax plan are dependent.

3. From a poll of 800 television viewers, the following data have been accumulated as to, their levels of education and their preference of television stations. We are interested in determining if the selection of a TV station is independent of the level of education

	Educational Level			
Public	High School	Bachelor	Graduate	**Total**
Broadcasting	50	150	80	**280**
Commercial Stations	150	250	120	**520**
Total	**200**	**400**	**200**	**800**

	Educational Level			
Public	High School	Bachelor	Graduate	Total
Broadcasting	50	150	80	280
Commercial Stations	150	250	120	520
Total	200	400	200	800

a. State the null and alternative hypotheses.

b. Show the contingency table of the expected frequencies.

c. (iii) Compute the test statistic.

(iv) The null hypothesis is to be tested at 95% confidence.

Determine the critical value for this test.

Solution:

Null Hypothesis: Selection of TV station is independent of level of education

Alternative Hypothesis: Selection of TV station is not independent of level of education

Level of significance: $\alpha = 0.05$

(i)

	Educational Level			
Public	High School	Bachelor	Graduate	Total
Broadcasting	50	150	80	280
Commercial Stations	150	250	120	520
Total	200	400	200	800

To Find Expected frequency:

$$\text{Expected Frequency} = \frac{\text{Corresponding row total} \times \text{Column total}}{\text{GrandTota1}}$$

$$\text{Expected Frequency for } 50 = \frac{280 \times 200}{800} = 70\,,$$

$$\text{Expected Frequency for } 150 = \frac{280 \times 400}{800} = 140$$

$$\text{Expected Frequency for } 80 = \frac{280 \times 200}{800} = 70\,,$$

$$\text{Expected Frequency for } 150 = \frac{520\times 200}{800} = 130$$

$$\text{Expected Frequency for } 250 = \frac{520\times 400}{800} = 260\ ,$$

$$\text{Expected Frequency for } 120 = \frac{520\times 200}{800} = 130$$

The test statistic: $\chi^2 \ = \sum_{i=1}^{r}\sum_{i=1}^{s}\frac{\left(O_{lj}-E_{ij}\right)^2}{E_{ij}}$ Analysis:

O_{ij}	E_{ij}	$O_{ij} - E_{ij}$	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
50	70	-20	400	5.714
150	140	10	100	0.174
80	70	10	100	1.428
150	130	20	400	3.076
250	260	-10	100	0.385
120	130	-10	100	0.769
TOTAL				11.546

O_{ij}	E_{ij}	O_{ij} - E_{ij}	$(O_{ij} - E_{ij})^2$	$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
50	70	-20	400	5.714
150	140	10	100	0.174
80	70	10	100	1.428
150	130	20	400	3.076
250	260	-10	100	0.385
120	130	-10	100	0.769
TOTAL				11.546

(iii)Test statistic $= 11.546$

(iv) Critical Chi-Square $= 5.991$,

Conclusion: Calculated value $>$ table value Hence, we reject Null Hypothesis.

4. From the data given below about the treatment of 250 patients suffering from a disease, state whether new treatment is superior to the conventional test.

Data	Number of patients		
	Favourable	Not favorable	Total
New one	140	30	170
Conventional	60	20	80
Total	200	50	280

Solution: We set up null hypothesis as there is no significance in results due to the two procedures adopted. The alternate hypothesis may be assumed as there could be some difference in the results.

Set up level of significance as $+\left(\frac{(112-100)^2}{100}\right)+\left(\frac{(71-50)^2}{50}\right)+\left(\frac{(32-10)^2}{10}\right)$ $\alpha = 5\%$ then tabulated value is $\chi^2|_{\alpha=0.05,\gamma=1} = 3.841.$

Consider finding expected values given by the formula, $\text{Expectation}(AB) = \frac{RT \cdot CT}{N}$

where RT means that the row total for the row containing the cell, CT means that the total for the column containing the cell and N, total number of frequencies. Keeping these in view, we find that expected frequencies are

		B	
A	136	34	170
	64	16	80
	200	50	250

Note: $\frac{170\cdot 200}{250} = 136;\quad \frac{170\cdot 50}{250} = 34,\quad \frac{80\cdot 200}{250} = 64$ and $\frac{80\cdot 50}{250} = 16.$

140	136	4	16	0.118
60	64	-4	16	0.250
30	34	-4	16	0.471
20	16	4	16	1.000
			Total	1.839

Note: $\frac{170\cdot 200}{250} = 136;$ $\frac{170\cdot 50}{250} = 34,$ $\frac{80\cdot 200}{250} = 64$ and $\frac{80\cdot 50}{250} = 16.$

140	136	4	16	0.118
60	64	-4	16	0.250
30	34	-4	16	0.471
20	16	4	16	1.000
			Total	1.839

As the calculated value 1.839 is lower than the tabulated value $\chi^2|_{\alpha=0.05,\gamma=1} = 3.841$, we accept the null hypothesis, namely, that there is not much significant difference between the two procedures.