The CHI-SQUARE TEST :
A FREQUENCY DATA BASED STATISTICAL DEVICE

Ashok kumar[1]

ABSTRACT

"Cause Effect" relationships are very important and play a most vital and significant role in medical researchers to reveal efficacy of medication as preventive and curative measures towards control and eradication of various diseases and morbidity factors. The present paper, in this direction is an attempt to highlight the main applicability of - test (chi- square test) ; a most appropriate statistical device in testing statistical significance of the above mentioned relationships in studying impact of one on the other ; supplemented by a few citations thereof.

KEWORDS: Observed and Expected Frequencies; Degrees of Freedom (df); Statistical Significance; Independence of Attributes; Null Hypothesis; Level of Significance; p-Value.

1. Associate Professor of Biostatistics, Department of Community Medicine, Universal College of Medical Sciences and Teaching Hospital, Bhairahawa, Nepal

For Correspondence
Dr. Ashok Kumar
Associate Professor of Biostatistics
Department of Community Medicine
Universal College of Medical Sciences and
Teaching Hospital Bhairahawa , Nepal
E- Mail: aku001@rediffmail.com

INTRODUCTION

The chi-square test , a frequency data based test of significance is basically used as a test of similarity/ dissimilarity between theory and experiment by testing significance of difference between the observed (experimental) and expected (theoretical /hypothetical) frequencies , to find out as to whether the postulated theory is also being testified by the experiment , or not. , towards its finally being approved/disapproved .part from above ,this test ; which is suitable for samples of large size (at least 50 units) has not only one , but a number of applications , among which one is testing association between two attributes (qualitative characters) towards establishing, "cause-effect" relationships , in terms of impact of one on the other.

ANALYTICAL METHOD

(i) test, in testing "Correspondence /Divergence" between theory and experiment:

The statistic, under the assumption of null hypothesis of no difference between observed and expected frequencies, is given by

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where, k = Number of Classes or Groups, O_i = Observed Frequency, E_i=Expected Frequency.

The insignificance of value at a given level of significance (in general 5% and 1%), results to 1:1 correspondence between theory and experiment ; while its significance indicates no correspondence between theory and experiment.

(ii) χ^2 test, in testing association between two attributes:

In testing association between two attributes, say A and B, under the assumption of null hypothesis of independence of A and B with observed frequencies in a 2x2 contingency table as under:

Number of Units	B	Not B	Total
A Not A	a c	b d	$(a+b)=R_1$ $(c+d)=R_2$
Total	$(a+c)=C_1$	$(b+d)=c_2$	(a+b+c+d)=N

χ^2 is defined as

$\chi^2 = N (ad - bc)^2 / R_1R_2C_1C_2$

The insignificance of above, at a given level of significance, refers to independence of A and B to show that "there is no impact of A on B"; while its significance refers to association between A and B, to reveal effectiveness of A on B. Here it is to be noted that as per accepted notions the significance of χ^2 is decided as

(i) if at a given level of significance, calculated chi-square is less than corresponding table value, then it is insignificant while if it is greater than or equal to table value then it is significant "or"

(ii) If corresponding p-value of chi-square is greater than 0.05 then it is not significant (insignificant), greater than 0.05 but less than 0.01 then it is significant and if it if less than 0.01 then it is highly significant.

CITATIONS

C-(1) Given following data of blood group wise distribution of patients in a hospital:

Blood Group:	O	A	B	AB	Total
Number of Patients:	100	20	40	40	200

can it be concluded that the patients are uniformly distributed over different Blood Groups?

SOLUTION:

To find out as to whether the patients are uniformly distributed over different blood groups, or not; by chi -square test, the null hypothesis will be that," there is no difference between observed and expected frequencies (i.e. patients are uniformly distributed over different blood groups).This tested by

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where, k=Number of groups=4, O_i=Observed frequency, E_i=Expected frequency

Table 1: Observed and Expected frequencies

S.No.	Blood Group	Observed Frequency	Expected Ratio	Expected Frequency
1	O	100	1	50
2	A	20	1	50
3	B	40	1	50
4	AB	40	1	50
Total	-	200	4	200

$\chi^2 = (100\text{-}50)^2 /50 + (20\text{-}50)^2/50 + (40\text{-}50)^2 /50 + (40\text{-}50)^2/50$
=50+18+2+2 = 72

Therefore, Calculated Value of Chi-Square = 72
From Table: Table value of Chi-Square at 5% level of Significance for 3 df = 7.82

Since, Calculated Chi-Square > Tabulated Chi-Square; hence it is significant at 5% level and we reject null hypothesis.

RESULT

At 5% level of significance, it cannot be concluded that patients are uniformly distributed over different blood groups.

C-(2): On the basis of following data, can you infer that Vaccination is effective in cure of disease 'D'?

Number of Patients	Disease ‘D’Cured	Disease ‘D’Not Cured	Total
Vaccinated	80	20	100
Not Vaccinated	30	70	100

SOLUTION

To find out effectiveness of vaccination on cure of disease 'D', by Chi-Square test the Null hypothesis will be that the given attributes i.e. vaccination and cure of didease 'D' is independent of each other. With following frequencies in a 2x2 Contingency table

Number of Patients	Disease'D'Cured	Disease 'D' not cured	Total
Vaccinated	80 (a)	20 (b)	100 ($a+b=R_1$)
Not Vaccinated	30 (c)	70(d)	100 ($c+d=R_2$)
Total	110 ($a+c= C_1$)	90 ($b+d= C_2$)	200($a+b+c+d = N$)

Number of Patients	Disease ‘D’ Cured	Disease ‘D’ Cured	Total
Vaccinated	80 (a)	20(b)	100($a+b=R_1$)
Not Vaccinated	-	70(d)	100($c+d=R_2$)
Total	110 ($a+c=C_1$)	90($b+d=C_2$)	200($a+b+c+d=N$)

$$\chi^2 = N(ad - bc)^2 / R_1R_2C_1C_2$$
$$= 200 (80.70 - 20.30)^2 / 100.100.110.90 = 50.50$$

Hence, Calculated Chi-Square = 50.50, df = (2-1) x (2-1) = 1x1 =1
From Table, 5% Value of Chi-Square for 1 df = 3.84

Since, Calculated Chi-Square > Table value of Chi-square, therefore it is significant at 5% level of significance and we reject the Null Hypothesis.

RESULT

At 5% Level of Significance, we can infer that vaccination is effective in cure of disease 'D'.

C-(3): Given following data, can it be said that High Status persons are more prone to HIV-Infection?

Number of Persons	HIV Infected	HIV Not Infected	Total
High Status	55	45	100
Low Status	50	50	100

To know as to whether High Status persons are more prone to HIV infection, or not, by chi-square test ; the null hypothesis will be that , “the given attributes i.e. Status and HIV infection are independent of each other .” With following frequencies in a 2x2 table

Number of Persons	**HIV Infected**	**HIV Not Infected**	**Total**
High Status	55(a)	45(b)	100($a+b=R_1$)
Low Status	50(c)	50(d)	100 ($c+d=R_2$)
Total	105($a+c=c_1$)	95($b+d=c_2$)	200($a+b+c+d=N$)

$\chi^2=N(ad-bc)^2/ R_1R_2C_1C_2=200(55.50-45.50)^2/100. 100. 105. 95 = 0.50$

Hence, Calculated Chi-Square = 0.50, df= (2-1) x (2-1) =1x1=1

From Table, Table value of Chi -Square at 5% level of significance for 1 df =3.84

Since, calculated Chi-Square < table value of Chi-Square; therefore it is not significant (insignificant) at 5% level and we can accept null hypothesis.

Result: At 5% level of significance, it cannot be said that high status persons are more prone to infection.

CONCLUSION

Among its various applications, like testing goodness of fit in terms of theoretical and experimental data, equality of proportions in different groups, homogeneity of several variances, detection of linkage in experiments of genetics; chi-square test plays a most significant role in associating two attributes to be able to know the impact of one on the other, towards cure of various diseases and knowing the corresponding morbidity factors, thereof.

REFERENCES

1. *Dixit JV. Principles and Practice of Biostatistics: 2nd Edition, Simplex Printing Press, Jabalpur. 2003.*

2. *Kapur J N, Saxena HC. Mathematical Statistics: S Chand & Co. New Delhi. 1997.*

3. *Mahajan BK. Methods in Biostatistics 7th Edition Jaypee Brothers. New Delhi. 2010.*

4. *Pant PR. Social Science Research and Thesis Writing: Buddha Academy Publishers and Distributors Pvt. Ltd. Kathmandu, Nepal. 2012.*

5. *Rao Sundar, P.S.S., Richard J. Introduction to Biostatistics and Research Methods; 4th Edition, Prentice Hall of India Pvt. Ltd. New Delhi. 2006.*

6. *Kothari CR. Research Methodology, Methods and Techniques, 2nd Revised Edition, New Age International Publications ,New Delhi, 2004.*