Vol 6 Issue 2 August 2016

ISSN No :2231-5063

International Multidisciplinary Research Journal

Golden Research Thoughts

Welcome to GRT

RNI MAHMUL/2011/38595 **ISSN No.2231-5063**

Golden Research Thoughts Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial board.Readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.aygrt.isrj.org

ISSN: 2231-5063 Impact Factor : 4.6052(UIF) Volume - 6 | Issue - 2 | August - 2016

GRT Golden Research Thoughts

CHI-SQUARE TEST AND IT'S APPLICATION

Dr. Sakdeo Babita Marutirao
Associate Professor in Agricultural Development Trust's, Shardabai Pawar Mahila Mahvidyalaya, Shardanagar, Malegaon Bk., Baramati, Pune.

ABSTRACT

Various tests of significance, such as 'Z' and 't' test discussed in the previous chapters are parametric tests and applied to only quantitative data like length, weight, height, Hb%. VO_2 consumption, number of seeds per and etc. These tests were based on the assumption that the samples were drawn from the normally distributed populations.

KEYWORDS :*Chi-Square Test, quantitative data, populations. statistics.*

INTRODUCTION

There are many situations in which it is not possible to make any dependable assumption about the distribution from which samples have been drawn. In biological experiments we also get qualitative data like the colour, health, intelligence, cure response of drug etc. in which observations arc classified in a particular category, class or group. For these qualitative data a non parametric test, called chi-square test is commonly used. In many studies, especially genetical studies, it becomes necessary to test the significance of overall deviation between the observed and expected frequencies.

Chi-square test was developed by Prof. A.R. Fisher in 1870. Karl Pearson improved Fisher's Chi-square test in its modern form in the year 1900. Chi-square is derived from the Greek Letter (Chi-Z) and pronounced as ki or Ksy without S.

DEFINITION

"Chi-square test is the test of significance of overall deviation square Observed and Expected frequencies divided by Expected frequencies.

Formula for Determining Z^2

The Chi-square (x^2) statistics is solved as follows :

$$\chi^2 = \Sigma\left\{\frac{(O-E)^2}{E}\right\} \text{ or } \Sigma\left\{\frac{(fo-fe)^2}{fe}\right\}$$

Here, O or *fo* = Observed frequency in a class
E or *fe*= Expected frequency in a class
Σ = Summation

From this equation, the value of x_2 will be zero if *fo= fe* in each classes, but due to chance error this never happens and the observed test are based on the number of degree of freedom (d.f.) at the critical level probability (5% or l%). We can find the expected value of x^2 from the table. This expected value can be compared with the value calculated from the data. If the tabular value is lower than the calculated value then the results significant.

Common application of X^2 test

1. As an alternative test to find significance of difference in two or more than two proportions. Chi-square test is a very useful test which can applied to find significance in the same type of data with advantage mentioned below :

(a) To compare the values of two binomial samples even if they small. For example - While finding significance of difference in rate respiration in 5 control and 5 Thyroxine (T4) treated animal of same species significance of difference in yield of fruit in control and insecticide sprayed garden etc. :

(b) To compare the frequencies of two multinomial samples. Chi-square measures the probability of association between two discrete attributes. Two events can be studied for their association such as iron intake and Hb season and fecundity, T_4 injection and oxygen consumption, nutritional and intelligence, weight and diabetes etc. There are two possibilities, either they influence or they do not.

2. As a test of association between two events in binomial or ----- samples. Chi-square measures the difference between two or -------- events.

The X^2 test can be applied to find association between two discrete ------- when there are more than two groups as happens in multinomial samples i.e., to test the association between quantity of T_4 injection -1 mg, 1 mg, 3 mg, 4 mg, 5 mg and rate of oxygen consumption; atmospheric temperature variations and water percentage of body; incidence of filariasis will social classes; state of nutrition and intelligence quotient.

Table is prepared by enumeration of qualitative data. Since one wants to know the association between two sets of events, the table is also called association table.

3. As a test of goodness of fit. X^2 test 's also applied as a test of goodness of fit". Goodness of fit reveals the closeness of Observed frequency with those of the Expected. Thus it helps to answer whether something (Physical or Chemical factors) did or did not have an effect. If Observed and Expected frequencies are in agreement with each other then the (In square should be zero. Null hypothesis H0. But it does not happen in life science. There is always some degree of deviation.

Pre-requisites of x^2 test -

There are three basic pre-requisites of x^2 test such as (i) Sample must be random, (ii) Data should be qualitative and (iii) Preferably Observed frequency should not be less than live.

DETERMINATION OF THE VALUE OF X^2

Following steps are required to calculate the value of x^2 :

(I) Make a contingency table and note the Observed frequency (fo or O) in each class of one class, of one event, rowwise i.e., horizontally and then the numbers in each group of the other event, columnwise i.e., vertically.

(II) Calculate the Expected frequencies (fe or E).

(III) Kind the difference between the Observed and Expected frequency in each cell (fo - fe). If the deviation is large, the square deviation (fo-fe) must be large.

(IV) Calculate the X^2 value applying formula, Σ(fo - fe) / fe. This will be the value of x2 which ranges from zero to infinity.

(V) The calculated value of x^2 's compared with the table value of given degrees of freedom at either 5% or 1% level of significance the calculated value of X^2 is less than the tabulated value at a ----- level of significance; the difference between the Observed and Expected frequencies is not significant, and could have arisen ---- fluctuations of sampling. On the other hand, when the calculated is more than the tabulated value, the difference between the and the Expected values is significant.

In biological observations, apart from quantitative characters, an investigator has to record qualitative characters i.e. colour. of the flowers, seed texture etc. These qualitative characters are difficult to express in the form-Of number, yet the number of flowers having a particular colour, Of the number of seeds with smooth or rough surface falls under a category which has a numerical estimate i.e. these are expressed numerically. As mentioned earlier, in such studies, each individual having a particular qualitative character is placed into a particular category or class and frequency distribution is studied with two or more classes.

The data expressed in frequency distribution with qualitative observation are based on field observations and the frequency of each class is thus known as observed frequency. The observed frequencies may differ from the expected values, or the: frequencies may be distributed differently from expected proportion, in respect to a particular hypothesis in mind of an investigator. It is, therefore, necessary to know whether there is any deviation between the observed and expected values of frequencies, or whether both the frequencies (observed and expected) are in agreement with each other. Furthermore, it is important to test the significance of overall deviation between the observed and expected values. The measurement of such deviation and its test of significance in known as χ2 test (read as Chi-square test, pronounced as kie-square test). The value of χ^2 is given by an equation :

$$\chi^2 = \Sigma \frac{(O-E)^2}{E}$$

Where O represents the observed and E the expected values.

The χ^2 test have several applications in statistical analysis i.e. to test the deviation ob observed values from the values to test goodness of fit in breeding experiments to determine association between two or more attributes etc.

The value of χ^2 is zero when observed and expected values perfectly match with each other. With the deviation between observed (O) and expected (E) values, the value of (O-E) increases and subsequently χ^2 increases. Greater the deviation (or difference) between observed and expected values, greater will be the size of χ^2.

In χ^2 analysis the degree of freedom (n-1) is calculated from the number of classes or categories involved in the study. The degree of freedom of choice is always one less than the total number of classes.

Example :

A fieldman was requested to collect 100 rose flowers expecting in mind that he will collect 25 each with white, red yellow and pink colour. However, it was observed that, he collected 27 white, 30 red, 20 yellow and 23 pink flowers. To determine whether the flowers collected are in agreement with the expectation, calculateχ^2 as :

Colour or rose flowers

Number	White	Red	Yellow	Pink
Observed ()	27	30	20	23
Expected (E)	25	25	25	25
Deviation (O-E)	+2	+5	-5	-2

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(2)^2}{25} + \frac{(5)^2}{25} + \frac{(-5)^2}{25} + \frac{(-2)^2}{25}$$

$$= \frac{4}{25} + \frac{25}{25} + \frac{25}{25} + \frac{4}{25}$$

$$= 0.16 + 1.00 + 1.00 + 0.16$$

$$= 2.32$$

The calculated X^2 value is 2.32. To test whether the value is significant for the difference or not, Chi-square table (Appendix V) is referred. The degree of freedom is = n-1 = 4-1= 3 as the number of characteristics are 4. In the X^2 table, at 3 degree of freedom the X^2 is 7.82 for p=0.05 (or at 5 % level of significance). The calculated X^2 value is much less than the value given in X^2 distribution table, which indicates that there is no significant difference between the collected (observed) number and expected number of flowers of different colour.

The X^2 test for goodness of fit;

This test is widely applied for studies related to hybridization. The main aim is to test the closeness between observed and expected frequencies of types of plants in F_2 generation.

Example :

In a plant breeding study two varieties of plants, were crossed. All of the plants were similar in F_1 generation, while four types of plants were observed in F_2 generation. Out of 640 plants there were 366 dwarf plants with red flowers, 112 dwarf plants with yellow flowers, 129 tall plants with red flowers and 33 tall plants with yellow flowers. For testing whether these 4 types of plants follow Mendelian dihybrid ratio of 9:3:3:1 for independent segregation, the X2 test is used as:

The expected number of plants that would fall in each class are calculated according to the hypothetical ratio 9:3:3:1, as:

i) dwarf with red flowers = (9/16) (640) = 360
ii) dwarf with yellow flowers = (3/16) (640) = 120
iii) tall with red flowers = (3/16) (640) = 120
iv) tall with yellow flowers = (1/16) (640) = 40
Total = 640

Determine whether observed values of number of plants are in agreement with predicted or expected values as:

Type of the plant			
dwarf	dwarf	tall	tall
with	with	with	with
red	yellow	red	yellow
flowers	flowers	flowers	flowers

Observed number (0)	366	112	129	33
Expected number (E)	360	120	120	40
Deviation (O - E)	+6	-8	+9	-7

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = \frac{(6)^2}{360} + \frac{(-8)^2}{120} + \frac{(9)^2}{120} + \frac{(-7)^2}{25}$$

$$= 0.100 + 0.533 + 0.675 + 1.225$$

$$= 2.533$$

The calculated X^2 value is 2.533. The X^2 value at 4-l=3 degree of freedom is 7.82 for p = 0.05 in X^2 table. As the calculated of value (2.533) is much less than the tabulated value (7.82) it is concluded that there is no significant difference between the observed and expected number of different types of plants in a population of F_2 generation. This the Mendelian dihybrid ratio of 9:3:3:1 holds true.

Degrees of freedom (d.f.)

In X^2 test, while comparing the calculated value of X^2 with the tabular value, we have to calculate the degree of freedom. The degree of freedom calculated from the number of classes. Therefore, the number of degrees freedom in a X^2 test is equal to the number of classes minus one. If there two classes, three classes, and four classes; the degree of freedom would 2-1,3-1 and 4-1. respectively. In a contingency table, the degree of freedom is calculated in a different manner :

d.f. = (r-1) (c-1)

Where r = Number of rows in a table,

c= number of columns in a table.

Thus in a 2x2 contingency table, the degree of freedom is (2 - 1) < (2- 1)= 1. Similarly in a 3 x 3 contingency table, the number of degrees of freedom is (3-1) (3 - 1) = 4, Likewise in 3x 4 contingency table the deg of freedom is (3 - 1) (4 - 1) = 6 and so on.

CONCLUSION=

Based on a large number of studies of the analysis of contingency tables, the current ecommendation would be to continue to use the standard Pearson chi-square test whenever the expected cell frequencies are sufficiently large. There seems to be no problem defining large as ”at least 5.” With small expected frequencies Fisher’s Exact Test seems to perform well regardless of the sampling plan, but 3 randomization tests adapted for the actual research design, as described above, will give a somewhat more exact solution. Recently Campbell (2007) carried out a very large sampling study on 2×2 tables comparing different chi-square statistics under different sample sizes and different underlying designs. He found that across all sampling designs, a statistic suggested by Karl Pearson’s son Egon Pearson worked best in most situations. The statistic is defined as χ 2 N N - 1 . (For the justification for that adjustment see Campbell’s paper.) Campbell found that as long as the smallest expected frequency was at least one, the adjusted chi-square held the Type I error rate at very nearly a . When the smallest expected frequency fell below 1, Fisher’s Exact Test did best.

REFERENCES

1.Campbell, I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Statistics in Medicine, 26, 3661−3675, (2007).

2. Cochran, W. G. The χ 2 test of goodness of fit. Annals of Mathematical Statistics, 25, 315−345, (1952).

3. Fisher, R. A. The logic of inductive inference. Journal of the Royal Statistical Society., 98, 39−54, (1934).