Faculty Name: Dr. Kumari Sushma Saroj

Department: Zoology

College: Dr. L. K. V. D College, Tajpur, Samastipur

Class: B.Sc Part –II (Hons.)

Paper: IV, Group – B

Topic: Chi -Square Test

Chi- Square Test

- A **chi-squared test**, also written as χ^2 **test**, is a statistical hypothesis test that is valid to perform when the test statistic is chi-squared distributed under the null hypothesis, specifically Pearson's chi-squared test and variants thereof. Pearson's chi-squared test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table.
- In the standard applications of this test, the observations are classified into mutually exclusive classes. If the null hypothesis is true, the test statistic computed from the observations follows a χ^2 frequency distribution. The purpose of the test is to evaluate how likely the observed frequencies would be assuming the null hypothesis is true.
- Test statistics that follow a χ^2 distribution occur when the observations are independent and normally distributed, which assumptions are often justified under the central limit theorem. There are also χ^2 tests for testing

the null hypothesis of independence of a pair of random variables based on observations of the pairs.

- Chi-squared tests often refers to tests for which the distribution of the test statistic approaches the χ^2 distribution asymptotically, meaning that the sampling distribution (if the null hypothesis is true) of the test statistic approximates a chi-squared distribution more and more closely as sample sizes increase.

Chi-Square Statistic:

The formula for the chi-square statistic used in the chi square test is:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

The chi-square formula.

Where, χ^2 = chi-square

O= observed number

E= expected number

d^2 =square of deviation for individual observation

$\sum$ = sum of

- Chi Square Test in SPSS.

- Chi Square P-Value in Excel.

A chi-square statistic is one way to show a relationship between two categorical variables.

In statistics, there are two types of variables: numerical (countable) variables and non-numerical (categorical) variables.

The chi-squared statistic is a single number that difference exists between observed counts and the counts expect if there were no relationship at all in the population.

There are a **few variations** on the chi-square statistic. Which one use depends upon how collected the data and which hypothesis is being tested.

However, all of the variations use the same idea, which are comparing expected values and with the values of actually collect.

One of the most common forms can be used for contingency tables:

$$c^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Where O is the observed value,

E is the expected value and "i" is the "ith" position in the contingency table.

A **low value** for chi-square means there is a high correlation between two sets of data.

In theory, if observed and expected values were equal ("no difference") then chi-square would be zero — an event that is unlikely to happen in real life.

Deciding whether a chi-square test statistic is large enough to indicate a statistically significant difference isn't as easy it seems.

It would be nice if we could say a chi-square test statistic >10 means a difference, but unfortunately that isn't the case.

A calculated chi-square value and compare it to a critical value from a chi-square table.

If the chi-square value is more than the critical value, then there is a significant difference.

Also use a p-value. First state the null hypothesis and the alternate hypothesis.

Then generate a chi-square curve results along with a p-value (Calculate a chi-square p-value Excel).

Small p-values (under 5%) usually indicate that a difference is significant (or "small enough").

Example:- In a cross between Tall (T) and dwarf (t) 1270 tall and 502 dwarf were obtained. Suggest if a ratio of 3 : 1 is suitable or not .

Solution:

Total number = 1270 + 502 = 1772

Hence expected 3: 1 ratio will be = 1329 : 443

Observed ratio = 1270 : 502

the formula:

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(1270 - 1329)^2}{1329} + \frac{(502 - 443)^2}{443}$$

$$= \frac{(-59)^2}{1329} + \frac{(59)^2}{443} = \frac{3481}{1329} + \frac{3481}{443}$$

$$= 2.619 + 7.767 = 10.386$$

The calculated value for χ^2 is compared to those given in the χ^2 – table after knowing the degree of freedom which is always one less than the actual number of classes. Here there are two classes (Tall and dwarf). Hence the degree of freedom 'n' is 2-1 =1

Here the calculated value of χ^2 is greater than show in the table, hence the ratio of 3:1 not suitable for it.

Chi Square P-Values.

A chi-square test will give a p-value. The p-value will tell, if the test results are significant or not.

In order to perform a chi square test and get the p-value, need two pieces of information:

- **Degrees of freedom**. That's just the number of categories minus 1.

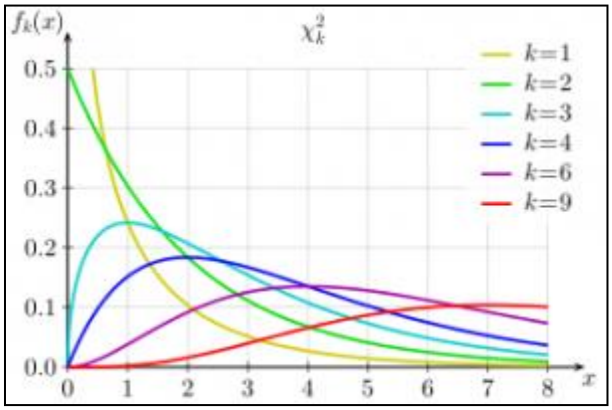
The alpha level (α). The usual alpha levels are 0.05 (5%), but also have other levels like 0.01 or 0.10.

In elementary statistics or AP statistics, both the degrees of freedom (df) and the alpha level are usually given in a question.

Degrees of freedom are placed as a subscript after the chi-square (X^2) symbol. For example, the following chi square shows 6 df:X^2_6.

And this chi square shows 4 df:X^2_4.

The Chi-Square Distribution



The chi-square distribution (also called the chi-squared distribution) is a special case of the gamma distribution.

A chi square distribution with n degrees of freedom is equal to a gamma distribution with a = n / 2 and b = 0.5 (or β = 2).

The chi-square test (a goodness of fit test).

- **Chi Distribution**

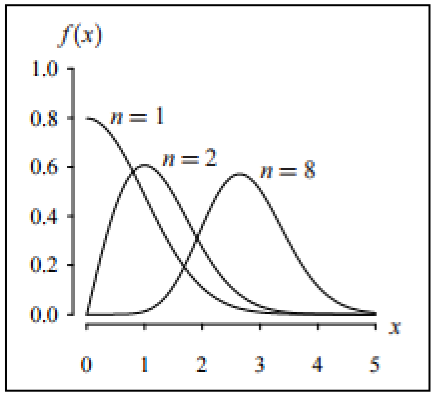
A similar distribution is the **chi distribution**.

This distribution describes the **square root** of a variable distributed according to a chi-square distribution,

with df = n > 0 degrees of freedom has a probability density function of:

$f(x) = 2^{(1-n/2)}\ x^{(n-1)}\ e^{(-(x2)}/2)\ /\ \Gamma(n/2)$

For values where X is positive.



The cdf for this function does not have a closed form, but it can be approximated with a series of integrals, using calculus.