10

Chi-Square Test

The chi-square test is an important test amongst the several tests of significance developed by statisticians. Chi-square, symbolically written as χ^2 (Pronounced as Ki-square), is a statistical measure used in the context of sampling analysis for comparing a variance to a theoretical variance. As a non-parametric* test, it "can be used to determine if categorical data shows dependency or the two classifications are independent. It can also be used to make comparisons between theoretical populations and actual data when categories are used."[1] Thus, the chi-square test is applicable in large number of problems. The test is, in fact, a technique through the use of which it is possible for all researchers to (i) test the goodness of fit; (ii) test the significance of association between two attributes, and (iii) test the homogeneity or the significance of population variance.

CHI-SQUARE AS A TEST FOR COMPARING VARIANCE

The chi-square value is often used to judge the significance of population variance i.e., we can use the test to judge if a random sample has been drawn from a normal population with mean (μ) and with a specified variance (σ_p^2). The test is based on χ^2-distribution. Such a distribution we encounter when we deal with collections of values that involve adding up squares. Variances of samples require us to add a collection of squared quantities and, thus, have distributions that are related to χ^2-distribution. If we take each one of a collection of sample variances, divided them by the known population variance and multiply these quotients by $(n - 1)$, where n means the number of items in the sample, we shall obtain a χ^2-distribution. Thus, $\frac{\sigma_s^2}{\sigma_p^2}(n-1) = \frac{\sigma_s^2}{\sigma_p^2}$ (d.f.) would have the same distribution as χ^2-distribution with $(n - 1)$ degrees of freedom.

* See Chapter 12 Testing of Hypotheses-II for more details.

[1] Neil R. Ullman, *Elementary Statistics—An Applied Approach*, p. 234.

The χ^2-distribution is not symmetrical and all the values are positive. For making use of this distribution, one is required to know the degrees of freedom since for different degrees of freedom we have different curves. The smaller the number of degrees of freedom, the more skewed is the distribution which is illustrated in Fig. 10.1:

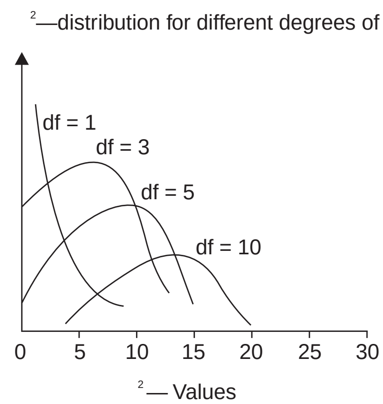

Fig. 10.1

Table given in the Appendix gives selected critical values of χ^2 for the different degrees of freedom. χ^2-values are the quantities indicated on the *x*-axis of the above diagram and in the table are areas below that value.

In brief, when we have to use chi-square as a test of population variance, we have to work out the value of χ^2 to test the null hypothesis (viz., $H_0 : \sigma_s^2 = \sigma_p^2$) as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1)$$

where σ_s^2 = variance of the sample;

σ_p^2 = variance of the population;

$(n-1)$ = degrees of freedom, n being the number of items in the sample.

Then by comparing the calculated value with the table value of χ^2 for $(n-1)$ degrees of freedom at a given level of significance, we may either accept or reject the null hypothesis. If the calculated value of χ^2 is less than the table value, the null hypothesis is accepted, but if the calculated value is equal or greater than the table value, the hypothesis is rejected. All this can be made clear by an example.

Illustration 1

Weight of 10 students is as follows:

S. No.	1	2	3	4	5	6	7	8	9	10
Weight (kg.)	38	40	45	53	47	43	55	48	52	49

Can we say that the variance of the distribution of weight of all students from which the above sample of 10 students was drawn is equal to 20 kgs? Test this at 5 per cent and 1 per cent level of significance.

Solution: First of all we should work out the variance of the sample data or σ_s^2 and the same has been worked out as under:

Table 10.1

S. No.	*X_i (Weight in kgs.)*	*$(X_i - \bar{X})$*	*$(X_i - \bar{X})^2$*
1	38	–9	81
2	40	–7	49
3	45	–2	04
4	53	+6	36
5	47	+0	00
6	43	–4	16
7	55	+8	64
8	48	+1	01
9	52	+5	25
10	49	+2	04
$n = 10$	$\sum X_i = 470$		$\sum(X_i - \bar{X})^2 = 280$

$$\bar{X} = \frac{\sum X_i}{n} = \frac{470}{10} = 47 \text{ kgs.}$$

$$\therefore \qquad \sigma_s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{280}{10-1}} = \sqrt{31.11}$$

or $$\sigma_s^2 = 31.11.$$

Let the null hypothesis be $H_0 : \sigma_p^2 = \sigma_s^2$. In order to test this hypothesis we work out the χ^2 value as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1)$$

$$= \frac{31.11}{20}(10-1) = 13.999.$$

Degrees of freedom in the given case is $(n-1) = (10-1) = 9$. At 5 per cent level of significance the table value of χ^2 = 16.92 and at 1 per cent level of significance, it is 21.67 for 9 d.f. and both these values are greater than the calculated value of χ^2 which is 13.999. Hence we accept the null hypothesis and conclude that the variance of the given distribution can be taken as 20 kgs at 5 per cent as also at 1 per cent level of significance. In other words, the sample can be said to have been taken from a population with variance 20 kgs.

Illustration 2

A sample of 10 is drawn randomly from a certain population. The sum of the squared deviations from the mean of the given sample is 50. Test the hypothesis that the variance of the population is 5 at 5 per cent level of significance.

Solution: Given information is

$$n = 10$$

$$\Sigma\left(X_i - \overline{X}\right)^2 = 50$$

$$\therefore \qquad \sigma_s^2 = \frac{\Sigma\left(X_i - \overline{X}\right)^2}{n-1} = \frac{50}{9}$$

Take the null hypothesis as $H_0: \sigma_p^2 = \sigma_s^2$. In order to test this hypothesis, we work out the χ^2 value as under:

$$\chi^2 = \frac{\sigma_s^2}{\sigma_p^2}(n-1) = \frac{\frac{50}{9}}{5}(10-1) = \frac{50}{9} \times \frac{1}{5} \times \frac{9}{1} = 10$$

Degrees of freedom = (10 – 1) = 9.

The table value of χ^2 at 5 per cent level for 9 d.f. is 16.92. The calculated value of χ^2 is less than this table value, so we accept the null hypothesis and conclude that the variance of the population is 5 as given in the question.

CHI-SQUARE AS A NON-PARAMETRIC TEST

Chi-square is an important non-parametric test and as such no rigid assumptions are necessary in respect of the type of population. We require only the degrees of freedom (implicitly of course the size of the sample) for using this test. As a non-parametric test, chi-square can be used (i) as a test of goodness of fit and (ii) as a test of independence.

As a test of goodness of fit, χ^2 test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data. When some theoretical distribution is fitted to the given data, we are always interested in knowing as to how well this distribution fits with the observed data. The chi-square test can give answer to this. If the calculated value of χ^2 is less than the table value at a certain level of significance, the fit is considered to be a good one which means that the divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the calculated value of χ^2 is greater than its table value, the fit is not considered to be a good one.

As a test of independence, χ^2 test enables us to explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, χ^2 test will helps us in deciding this issue. In such a situation, we proceed with the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent which means that new medicine is not effective in controlling fever. On this basis we first calculate the expected frequencies and then work out the value of χ^2. If the calculated value of χ^2 is less than the table value at a certain level of significance for given degrees of freedom, we conclude that null hypothesis stands which means that the two attributes are independent or not associated (i.e., the new medicine is not effective in controlling the fever). But if the calculated value of χ^2 is greater than its table value, our inference then would be that null hypothesis does not hold good which means the two attributes are associated and the association is not because of some chance factor but it exists in reality (i.e., the new medicine is effective in controlling the fever and as such may be prescribed). It may, however, be stated here that χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes.

In order that we may apply the chi-square test either as a test of goodness of fit or as a test to judge the significance of association between attributes, it is necessary that the observed as well as theoretical or expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution. χ^2 is then calculated as follows:

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where

O_{ij} = observed frequency of the cell in *i*th row and *j*th column.

E_{ij} = expected frequency of the cell in *i*th row and *j*th column.

If two distributions (observed and theoretical) are exactly alike, $\chi^2 = 0$; but generally due to sampling errors, χ^2 is not equal to zero and as such we must know the sampling distribution of χ^2 so that we may find the probability of an observed χ^2 being given by a random sample from the hypothetical universe. Instead of working out the probabilities, we can use ready table which gives probabilities for given values of χ^2. Whether or not a calculated value of χ^2 is significant can be ascertained by looking at the tabulated values of χ^2 for given degrees of freedom at a certain level of significance. If the calculated value of χ^2 is equal to or exceeds the table value, the difference between the observed and expected frequencies is taken as significant, but if the table value is more than the calculated value of χ^2, then the difference is considered as insignificant i.e., considered to have arisen as a result of chance and as such can be ignored.

As already stated, degrees of freedom* play an important part in using the chi-square distribution and the test based on it, one must correctly determine the degrees of freedom. If there are 10 frequency classes and there is one independent constraint, then there are (10 – 1) = 9 degrees of freedom. Thus, if '*n*' is the number of groups and one constraint is placed by making the totals of observed and expected frequencies equal, the d.f. would be equal to (n – 1). In the case of a contingency table (i.e., a table with 2 columns and 2 rows or a table with two columns and more than two rows or a table with two rows but more than two columns or a table with more than two rows and more than two columns), the d.f. is worked out as follows:

$$\text{d.f.} = (c - 1)\,(r - 1)$$

where '*c*' means the number of columns and '*r*' means the number of rows.

CONDITIONS FOR THE APPLICATION OF χ^2 TEST

The following conditions should be satisfied before χ^2 test can be applied:

(i) Observations recorded and used are collected on a random basis.

(ii) All the itmes in the sample must be independent.

(iii) No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.

(iv) The overall number of items must also be reasonably large. It should normally be at least 50, howsoever small the number of groups may be.

(v) The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known are know as linear constraints.

STEPS INVOLVED IN APPLYING CHI-SQUARE TEST

The various steps involved are as follows:

* For d.f. greater than 30, the distribution of $\sqrt{2\chi^2}$ approximates the normal distribution wherein the mean of $\sqrt{2\chi^2}$ distribution is $\sqrt{2\text{d.f.} - 1}$ and the standard deviation = 1. Accordingly, when d.f. exceeds 30, the quantity $\left[\sqrt{2\chi^2} - \sqrt{2\text{d.f.} - 1}\right]$ may be used as a normal variate with unit variance, i.e.,

$$z_\alpha = \sqrt{2\chi^2} - \sqrt{2\text{d.f.} - 1}$$

(i) First of all calculate the expected frequencies on the basis of given hypothesis or on the basis of null hypothesis. Usually in case of a 2 × 2 or any contingency table, the expected frequency for any given cell is worked out as under:

$$\text{Expected frequency of any cell} = \left[\frac{\text{(Row total for the row of that cell)} \times \text{(Column total for the column of that cell)}}{\text{(Grand total)}}\right]$$

(ii) Obtain the difference between observed and expected frequencies and find out the squares of such differences i.e., calculate $(O_{ij} - E_{ij})^2$.

(iii) Divide the quantity $(O_{ij} - E_{ij})^2$ obtained as stated above by the corresponding expected frequency to get $(O_{ij} - E_{ij})^2/E_{ij}$ and this should be done for all the cell frequencies or the group frequencies.

(iv) Find the summation of $(O_{ij} - E_{ij})^2/E_{ij}$ values or what we call $\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$. This is the required χ^2 value.

The χ^2 value obtained as such should be compared with relevant table value of χ^2 and then inference be drawn as stated above.

We now give few examples to illustrate the use of χ^2 test.

Illustration 3

A die is thrown 132 times with following results:

Number turned up	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

Is the die unbiased?

Solution: Let us take the hypothesis that the die is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is 132 ×1/6 = 22. Now we can write the observed frequencies along with expected frequencies and work out the value of χ^2 as follows:

Table 10.2

No. turned up	*Observed frequency* O_i	*Expected frequency* E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$\therefore \qquad \Sigma[(O_i - E_i)^2/E_i] = 9.$

Hence, the calculated value of $\chi^2 = 9$.

$\because$ Degrees of freedom in the given problem is

$(n - 1) = (6 - 1) = 5.$

The table value* of χ^2 for 5 degrees of freedom at 5 per cent level of significance is 11.071. Comparing calculated and table values of χ^2, we find that calculated value is less than the table value and as such could have arisen due to fluctuations of sampling. The result, thus, supports the hypothesis and it can be concluded that the die is unbiased.

Illustration 4

Find the value of χ^2 for the following information:

Class	*A*	*B*	*C*	*D*	*E*
Observed frequency	8	29	44	15	4
Theoretical (or expected) frequency	7	24	38	24	7

Solution: Since some of the frequencies less than 10, we shall first re-group the given data as follows and then will work out the value of χ^2 :

Table 10.3

Class	*Observed frequency O_i*	*Expected frequency E_i*	$O_i - E_i$	$(O_i - E_i)^2/E_i$
A and *B*	(8 + 29) = 37	(7 + 24) = 31	6	36/31
C	44	38	6	36/38
D and *E*	(15 + 4) = 19	(24 + 7) = 31	–12	144/31

$$\therefore \qquad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 6.76 \text{ app.}$$

Illustration 5

Genetic theory states that children having one parent of blood type *A* and the other of blood type *B* will always be of one of three types, *A*, *AB*, *B* and that the proportion of three types will on an average be as 1 : 2 : 1. A report states that out of 300 children having one *A* parent and *B* parent, 30 per cent were found to be types *A*, 45 per cent per cent type *AB* and remainder type *B*. Test the hypothesis by χ^2 test.

Solution: The observed frequencies of type *A*, *AB* and *B* is given in the question are 90, 135 and 75 respectively.

*Table No. 3 showing some critical values of χ^2 for specified degrees of freedom has been given in Appendix at the end of the book.

The expected frequencies of type *A*, *AB* and *B* (as per the genetic theory) should have been 75, 150 and 75 respectively.

We now calculate the value of χ^2 as follows:

Table 10.4

Type	*Observed frequency* O_i	*Expected frequency* E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
A	90	75	15	225	225/75 = 3
AB	135	150	–15	225	225/150 = 1.5
B	75	75	0	0	0/75 = 0

$$\therefore \qquad \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 3 + 1.5 + 0 = 4.5$$

$$\because \qquad \text{d.f.} = (n - 1) = (3 - 1) = 2.$$

Table value of χ^2 for 2 d.f. at 5 per cent level of significance is 5.991.

The calculated value of χ^2 is 4.5 which is less than the table value and hence can be ascribed to have taken place because of chance. This supports the theoretical hypothesis of the genetic theory that on an average type *A*, *AB* and *B* stand in the proportion of 1 : 2 : 1.

Illustration 6

The table given below shows the data obtained during outbreak of smallpox:

	Attacked	*Not attacked*	*Total*
Vaccinated	31	469	500
Not vaccinated	185	1315	1500
Total	216	1784	2000

Test the effectiveness of vaccination in preventing the attack from smallpox. Test your result with the help of χ^2 at 5 per cent level of significance.

Solution: Let us take the hypothesis that vaccination is not effective in preventing the attack from smallpox i.e., vaccination and attack are independent. On the basis of this hypothesis, the expected frequency corresponding to the number of persons vaccinated and attacked would be:

$$\text{Expectation of } (AB) = \frac{(A) \times (B)}{N}$$

when *A* represents vaccination and *B* represents attack.

$\therefore$

$$(A) = 500$$
$$(B) = 216$$
$$N = 2000$$

$$\text{Expectation of } (AB) = \frac{500 \times 216}{2000} = 54$$

Now using the expectation of (*AB*), we can write the table of expected values as follows:

	Attacked: B	*Not attacked: b*	*Total*
Vaccinated: *A*	(*AB*) = 54	(*Ab*) = 446	500
Not vaccinated: *a*	(*aB*) = 162	(*ab*) = 1338	1500
Total	216	1784	2000

Table 10.5: Calculation of Chi-Square

Group	*Observed frequency* O_{ij}	*Expected frequency* E_{ij}	$(O_{ij} - E_{ij})$	$(O_{ij} - E_{ij})^2$	$(O_{ij} - E_{ij})^2/E_{ij}$
AB	31	54	–23	529	529/54 = 9.796
Ab	469	446	+23	529	529/44 = 1.186
aB	158	162	+23	529	529/162 = 3.265
ab	1315	1338	–23	529	529/1338 = 0.395

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 14.642$$

$\because$ Degrees of freedom in this case = $(r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$.

The table value of χ^2 for 1 degree of freedom at 5 per cent level of significance is 3.841. The calculated value of χ^2 is much higher than this table value and hence the result of the experiment does not support the hypothesis. We can, thus, conclude that vaccination is effective in preventing the attack from smallpox.

Illustration 7

Two research workers classified some people in income groups on the basis of sampling studies. Their results are as follows:

Investigators	*Income groups*			*Total*
	Poor	*Middle*	*Rich*	
A	160	30	10	200
B	140	120	40	300
Total	300	150	50	500

Show that the sampling technique of at least one research worker is defective.

Solution: Let us take the hypothesis that the sampling techniques adopted by research workers are similar (i.e., there is no difference between the techniques adopted by research workers). This being so, the expectation of *A* investigator classifying the people in

(i) Poor income group $= \dfrac{200 \times 300}{500} = 120$

(ii) Middle income group $= \dfrac{200 \times 150}{500} = 60$

(iii) Rich income group $= \dfrac{200 \times 50}{500} = 20$

Similarly the expectation of *B* investigator classifying the people in

(i) Poor income group $= \dfrac{300 \times 300}{500} = 180$

(ii) Middle income group $= \dfrac{300 \times 150}{500} = 90$

(iii) Rich income group $= \dfrac{300 \times 50}{500} = 30$

We can now calculate value of χ^2 as follows:

Table 10.6

Groups	*Observed frequency* O_{ij}	*Expected frequency* E_{ij}	$O_{ij} - E_{ij}$	$(O_{ij} - E_{ij})^2\ E_{ij}$
Investigator A				
classifies people as poor	160	120	40	1600/120 = 13.33
classifies people as middle class people	30	60	–30	900/60 = 15.00
classifies people as rich	10	20	–10	100/20 = 5.00
Investigator B				
classifies people as poor	140	180	–40	1600/180 = 8.88
classifies people as middle class people	120	90	30	900/90 = 10.00
classifies people as rich	40	30	10	100/30 = 3.33

Hence, $$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = 55.54$$

$\because$ Degrees of freedom $= (c - 1)\,(r - 1)$

$= (3 - 1)\,(2 - 1) = 2.$

The table value of χ^2 for two degrees of freedom at 5 per cent level of significance is 5.991.

The calculated value of χ^2 is much higher than this table value which means that the calculated value cannot be said to have arisen just because of chance. It is significant. Hence, the hypothesis does not hold good. This means that the sampling techniques adopted by two investigators differ and are not similar. Naturally, then the technique of one must be superior than that of the other.

Illustration 8

Eight coins were tossed 256 times and the following results were obtained:

Numbers of heads	0	1	2	3	4	5	6	7	8
Frequency	2	6	30	52	67	56	32	10	1

Are the coins biased? Use χ^2 test.

Solution: Let us take the hypothesis that the coins are not biased. If that is so, the probability of any one coin falling with head upward is 1/2 and with tail upward is 1/2 and it remains the same whatever be the number of throws. In such a case the expected values of getting 0, 1, 2, … heads in a single throw in 256 throws of eight coins will be worked out as follows*.

Table 10.7

Events or No. of heads	*Expected frequencies*
0	${}^8C_0\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^8 \times 256 = 1$
1	${}^8C_1\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7 \times 256 = 8$
2	${}^8C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 \times 256 = 28$

contd.

*The probabilities of random variable i.e., various possible events have been worked out on the binomial principle viz., through the expansion of $(p + q)^n$ where $p = 1/2$ and $q = 1/2$ and $n = 8$ in the given case. The expansion of the term ${}^nC_r\, p^r\, q^{n-r}$ has given the required probabilities which have been multiplied by 256 to obtain the expected frequencies.

Events or No. of heads	*Expected frequencies*
3	${}^8C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^5 \times 256 = 56$
4	${}^8C_4\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^4 \times 256 = 70$
5	${}^8C_5\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3 \times 256 = 56$
6	${}^8C_6\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2 \times 256 = 28$
7	${}^8C_7\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^1 \times 256 = 8$
8	${}^8C_8\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^0 \times 256 = 1$

The value of χ^2 can be worked out as follows:

Table 10.8

No. of heads	*Observed frequency* O_i	*Expected frequency* E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
0	2	1	1	1/1 = 1.00
1	6	8	–2	4/8 = 0.50
2	30	28	2	4/28 = 0.14
3	52	56	–4	16/56 = 0.29
4	67	70	–3	9/70 = 0.13
5	56	56	0	0/56 = 0.00
6	32	28	4	16/28 = 0.57
7	10	8	2	4/8 = 0.50
8	1	1	0	0/1 = 0.00

$$\therefore \qquad \chi^2 = \Sigma\frac{(O_i - E_i)^2}{E_i} = 3.13$$

$\therefore$ Degrees of freedom = $(n - 1) = (9 - 1) = 8$

The table value of χ^2 for eight degrees of freedom at 5 per cent level of significance is 15.507. The calculated value of χ^2 is much less than this table and hence it is insignificant and can be ascribed due to fluctuations of sampling. The result, thus, supports the hypothesis and we may say that the coins are not biased.

ALTERNATIVE FORMULA

There is an alternative method of calculating the value of χ^2 in the case of a (2×2) table. If we write the cell frequencies and marginal totals in case of a (2×2) table thus,

a	b	$(a + b)$
c	d	$(c + d)$
$(a + c)$	$(b + d)$	N

then the formula for calculating the value of χ^2 will be stated as follows:

$$\chi^2 = \frac{(ad - bc)^2 \cdot N}{(a + c)(b + d)(a + b)(c + d)}$$

where N means the total frequency, ad means the larger cross product, bc means the smaller cross product and $(a + c)$, $(b + d)$, $(a + b)$, and $(c + d)$ are the marginal totals. The alternative formula is rarely used in finding out the value of chi-square as it is not applicable uniformly in all cases but can be used only in a (2×2) contingency table.

YATES' CORRECTION

F. Yates has suggested a correction for continuity in χ^2 value calculated in connection with a (2×2) table, particularly when cell frequencies are small (since no cell frequency should be less than 5 in any case, through 10 is better as stated earlier) and χ^2 is just on the significance level. The correction suggested by Yates is popularly known as Yates' correction. It involves the reduction of the deviation of observed from expected frequencies which of course reduces the value of χ^2. The rule for correction is to adjust the observed frequency in each cell of a (2×2) table in such a way as to reduce the deviation of the observed from the expected frequency for that cell by 0.5, but this adjustment is made in all the cells without disturbing the marginal totals. The formula for finding the value of χ^2 after applying Yates' correction can be stated thus: