An example of comparing frequencies: 2 x 2 chi-square

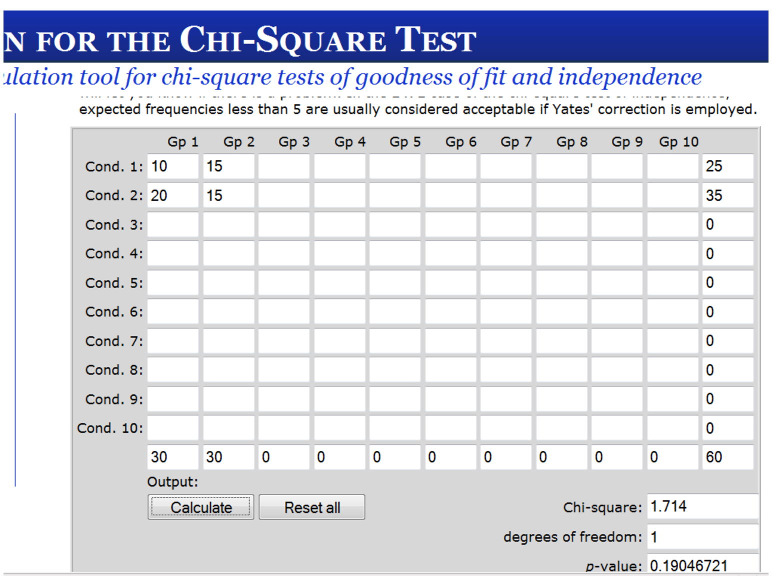
Suppose you have two groups (one who took Vitamin C and another who took placebo), and you would like to see if the frequency of getting the flu is different or not. 33% of the vitamin C group had the flu. 50% of the placebo group got the flue. Is it statistically different?

The 2 x 2 contingency table:

	Had flu	Did not have flu
Group (A) Vitamin C	10	20
Group (B) Placebo	15	15

There are many online applets you can use. Here is one:

http://www.quantpsy.org/chisq/chisq.htm



The difference was not found to be significant (χ^2 (1) = 1.714, *p* = .19).

See Chi.pdf below for more details, which is an excerpt (358 – 360) from: Sprinthall (2003). Basic Statistical Analysis. Boston: Allyn and Bacon.

Chi.pdf

The r × k Chi Square (r by k)

The chi squares shown so far are called $1 \times k$ chi squares, meaning that one sample group of subjects has been assigned to any number, k, of categories. Sometimes, however, a researcher wishes to select more than one group—for example, an experimental and a control group—and then to compare these groups with respect to some observed frequency. For this, the $r \times k$ chi square is used.

Although the basic equation remains the same, there are two differences in the steps required to complete the $r \times k$ chi square. First, the frequency expected on the basis of chance cannot be computed in the same way. Second, the degrees of freedom are assigned in a slightly different manner.

A researcher is interested in discovering whether vitamin C aids in the prevention of influenza. Two groups are randomly selected, with 30 subjects in each group. Group A is given 250 mg of vitamin C daily for a period of three months, while Group B is given a placebo. In Group A, 10 subjects report having caught influenza during that time, while in Group B, 15 subjects report having the flu.

Contingency Table. The data are set up in what is called a contingency table, shown in Table 13.2. The cells in the contingency table are lettered a, b, c, and d and represent specific and unique categories. For example, cell a contains only those subjects who took vitamin C and also caught influenza, cell b contains only those who took vitamin

TABLE 13.2 Contingency table for vitamin C/influenza data.

	Had influenza	*Did Not Have influenza*	
Group A: vitamin C	a 10	b 20	30 = a + b
Group B: placebo	c 15	d 15	30 = c + d
	a + c = 25	b + d = 35	60 = a + b + c = N

C and did not catch influenza, and so on. Note that this particular contingency table has two rows and two columns; it is called a 2×2 contingency table. Also note that the two groups define the rows, whereas the categories on which the subjects are being measured (flu versus no flu) are heading the columns. Another way to remember this is that the independent variable (whether or not the subjects took the vitamin) is set in the rows and the dependent variable (whether or not they caught the flu) in the columns.

To the right of each row, and at the bottom of each column, we place the marginal totals: 30 and 30 for the rows and 25 and 35 for the columns. This is an important step; these marginal totals are used for calculating the values of the frequency expected. Also, as a check, we verify that the row total (30 + 30) equals the column total (25 + 35) and that *each adds up to the total* N, in this case 60.

Calculating the Chi Square Value. As we did for the $1 \times k$, we set up a table. Here, each column is headed by a particular cell. As we fill in this table, we must keep the contingency table (Table 13.2) clearly in view because some of the values are taken from it.

The value of the r by k chi square is obtained by the following steps (see Table 13.3):

1. We take the values for f_o directly from the contingency table and put them in the first row: 10 for cell a, 20 for cell b, and so on.

TABLE 13.3 An $r \times k$ chi square with data from Table 13.2.

	a	*b*	*c*	*d*	
f_o	10	20	15	15	
f_e	12.500	17.500	12.500	17.500	
$f_o - f_e$	−2.500	2.500	2.500	−2.500	
$(f_o - f_e)^2$	6.250	6.250	6.250	6.250	
$(f_o - f_e)^2/f_e$	.500	.357	.500	.357	$\chi^2 = 1.714$

2. To fill the second row, we calculate the values for f_e for all cells of the contingency table. For a given cell, f_e equals the product of its column total and its row total divided by N, the total number of cases. Cell a is in the column whose total is 25 and in the row whose total is 30. Therefore, f_e for cell a equals $(25 \times 30) \div 60$, or 12.500. By this same formula, we obtain values of f_e for b, c, and d of 17.500, 12.500, and 17.500, respectively.

3. In the third row, we place the difference between f_o and f_e. For cell a, this value is −2.500, from 10 − 12.500, and so on.

4. We square the values in the third row and place the squared differences in the fourth row, that is, 6.250 for all cells.

5. We divide each squared difference by the f_e for that cell and put the resulting values in the fifth row: .500, .357, .500, and .357.

6. We add across all the values in the fifth row to obtain the value for chi square, that is, $\chi^2 = 1.714$.

Interpreting the Chi Square Value. The degrees of freedom for the $r \times k$ chi square are found by the following equation: $\mathrm{df} = (r - 1)(k - 1)$. The numbers of rows and columns are taken *from the contingency table*. Thus, in the vitamin C study, we had two rows (vitamin C versus no vitamin C) and two columns (influenza versus no influenza). The degrees of freedom, then, are equal to $(2 - 1)(2 - 1)$, or 1. Then

$$\chi^2_{.05(1)} = 3.84$$
$$\chi^2 = 1.714 \qquad \text{Accept } H_0\text{; not significant.}$$

Checking our obtained chi square value of 1.714 against the critical value of 3.84 for an alpha level of .05 shows that the statistical decision must be an acceptance of the null hypothesis. There are no significant differences between the two groups with respect to their frequency of catching flu. Since this was experimental methodology between-subjects, we conclude that the independent variable (vitamin C) did not affect the dependent variable (influenza). We would write the conclusions as follows: A 2×2 chi square was computed comparing the frequency of influenza between groups who had either received vitamin C or a placebo. The difference was found not to be significant (chi square(1) = 1.714, $p > .05$, ns.).