Some Comments on Chi-Square

Thinking about Chi-square (χ^2)

There are essentially three purposes or rationales for chi-square:

- **Test of Homogeneity**. Chi-square is sometimes called a test of homogeneity (meaning "same type"), because it examines whether two groups are the same or different. You can think of this as analogous to a the t-test, but used in a situation where the dependent variable is dichotomous. In other words, chi-square is used to see if two groups are the same in their responses to a dichotomous variable (e.g., yes/no survey question).
- **Goodness-of-Fit**. Chi-square is also sometimes called a test of goodness-of-fit, because it tests the degree to which observed frequencies fit the frequencies one expects from chance. Thus, the observed frequencies from the study (f_o) are "fitted" to the frequencies that are expected from chance (f_e). This is clearly seen in the computation of the definitional formula for chi-square. Larger chi-squares indicate *less fit* (or a lack of fit). A chi-square of zero means there is a perfect fit of the observed to the expected frequencies.
- **Test of Independence**. One can also think of chi-square as a test of independence, in which the "independence" or "dependence" of two variables is tested. If two variables are uncorrelated, they are independent. If they are correlated, they are dependent. The "test of independence" of chi-square is testing whether the null hypothesis that the two variables are uncorrelated (or independent) is true or not.

So, these different interpretations or rationales of chi-square mean that this statistic is useful to test many different types of hypotheses, given a data situation that involves all categorical variables. These three interpretations of chi-square also highlight the fact that when we are testing to see if two groups are different, we are also testing the hypothesis about whether a grouping variable (i.e., the dichotomous independent variable) is correlated with the dependent variable.

Applications

Earlier, we discussed the use of chi-square to test whether or not there were more yes's or no's in a survey (e.g., when polling about elections). This same logic can be extended to test more complicated questions. Most commonly, a chi-square analysis is used to compare two groups on a yes or no survey question. So, for example, we might examine whether Republicans and Democrats differ in their opinions of a gun control bill (until very recently, this would be a pretty self-evident hypothesis). This type of test is a between-subjects test, because two separate groups are being compared.

The definitional formula for the chi-square analysis is exactly the same as before.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Where f_o is the observed frequency and f_e is the expected frequency. The complexity is figuring out what the f_e should be when we have a more complicated design. Because the expected frequencies depend on how many people overall said yes or no and how many people there are in each group, we must use the marginal frequencies to compute the f_e.

The expected frequency formula to find the f_e for each cell is below:

$$f_e = \frac{f_r(f_c)}{N_T}$$

f_r is the observed marginal frequency for that row, f_c is the observed marginal frequency for that column, and N_T is the total number of cases.

For example, the marginal frequencies of the total number of people in group 1 and the total number of people who said "no" are used to find the expected frequency for those who said no in group 1 ($f_{e1,no}$) in the table below.

	no	yes	
group 1	14	23	37
group 2	24	13	37
	38	36	74

$$f_{e1,no} = \frac{(37)(38)}{74} = 19$$

The result of the chi-square is compared to the tabled critical value based on $df = (r - 1)(c - 1)$, where r and c represent the number of rows and the number of columns, respectively.

Chi-square for within-subjects

The chi-square test for within-subjects designs is called McNemar's chi-square. As with the paired t-test or the within-subjects ANOVA, the McNemar test is used whenever the same individuals are measured (or surveyed) twice, matched on some variable (e.g., yoked by age), participants are paired in some way (e.g., twins or married couples), or responses on two measures are used (e.g., favorability to gun control compared to favorability for abolishing the second amendment).

For instance, we might examine the favorability of voters for gun control legislation in April and in June.

		June		
		No	Yes	
April	No	80	100	180
	Yes	10	110	120
		90	210	300

To compute McNemar's, the following formula is used:

$$McNemar's\ \chi^2 = \frac{(c-b)^2}{c+b}$$

c, b, and d come from labeling the cells in the table as below.

		June	
		No	Yes
April	No	a	b
	Yes	c	d

$$= \frac{(10-100)^2}{100+10}$$
$$= \frac{(90)^2}{110}$$
$$= 73.63$$

df in this test is 1.

For more than 2 related groups, one can use Cochran's Q test, which I won't detail here.

Other Comments

Planned follow-up analyses in complex chi-square contingency tables are simple chi-square analyses based on simple chi-squares for two-cell comparisons or smaller contingency tables (e.g., a 2 X 2 from a 5 X 3 design). The chi-squares for the set of all possible *orthogonal* chi-squares add up to the chi-square for the whole design (or the omnibus test).

Although 2 X 2 chi-square table looks like a 2 X 2 factorial table, they are not analogous. Because one of the columns (or rows) is for the dependent variable, it is really the three-way table that is analogous to the factorial design. So, a 2 X 2 factorial design with a dichotomous dependent variable requires an analysis of a three-way contingency table--a 2 X 2 X 2. There is a variant of the chi-square test called the Mantel-Haenszel statistic, which is not currently available in SPSS (but is available in SAS).