Philippine Journal of Psychology
1868, Vol. 1, No. 1, 16–17

A NOTE ON THE COMPUTATION OF THE CHI-SQUARE FROM A FOURFOLD TABLE

LEONEL CAMPOS
Ateneo de Manila University

The most general expression of the chi-square is,

$$\text{Chi-square} = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad (1)$$

where,

O_{ij} represents an empirical, observed frequency, and

E_{ij} represents a theoretical, expected frequency.

Formula (1) is well-known and needs no further elaboration.

For a four-fold table of contingencies with one degree of freedom (*df*), the difference $O_{ij} - E_{ij}$ is identical for all cells in the table. For this reason, formula (1) is often given (Guilford, 1956, p. 236; Snedecor, 1956, p. 224) as follows:

$$\text{Chi-square} = (O_{ij} - E_{ij})^2 \sum \left(\frac{1}{E}\right) \qquad (2)$$

Formulas (1) and (2) require the laborious and error-producing technique of finding and adding the reciprocals of the E_{ij}'s for all the cells in the table. That procedure may involve a considerable degree of inaccuracy introduced by round-off errors.

A short-cut formula is also available, which requires that the cells of the fourfold table be labelled, A, B, C, and D, letting A and D be diagonally opposed regardless of their location to represent the O_{ij}'s of the cells so labelled. The expression for chi-square then becomes:

$$\text{Chi-square} = \frac{N(AD - BC)^2}{(A + B)(A + C)(C + D)(B + D)} \qquad (3)$$

where $N = A + B + C + D$, and the terms in the parentheses in the denominator represent each one of the marginal totals of the table in question.

Formula (3) is still laborious, but is the most accurate approximation of the chi-square in a 2 x 2 table, and the most convenient if a desk calculator is at hand. Its main drawback is that it involves the use of very large numbers. It is quite inconvenient when there is no access to mechanical computational aids. It is possible, however, to strike a balance between the degree of inaccuracy involved in formulas (1) and (2), and the amount of labor involved in formula (3) by the following operations: Since we have let A, B, C, and D stand for the O_{ij}'s of the cells so labelled, for the purpose of using formula (3), we may let a, b, c, and d stand for their corresponding E_{ij}'s. In this case, formula (2) may be re-written,

$$\text{Chi-square} = (A - a)^2 \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right) \qquad (4)$$

$$\text{Chi-square} = (A - a)^2 \left(\frac{a + d}{ad} + \frac{b + c}{bc}\right) = \frac{N(A - a)^2}{ad} \qquad (4a)$$

The simplication expressed in formula (4a) follows from the fact that $a + b + c + d = N$, and from the fact that, ad = bc, which can easily be shown. Formula (4a) above is then used by means of the following logical steps:

Step 1: Set up the table of frequencies as usual;

Step 2. Compute,

$$a = \frac{(A + B)\ (A + C)}{N};$$

Step 3: Compute,

$$d = D - (A - a);$$

Step 4: Substitute numerical values for the symbols in formula (4a) and compute for chi-square.

The correction for lack of continuity necessary in this case is incorporated as follows.

$$\text{Chi-square} = \frac{N(\ |A - a| - 0.5)^2}{ad} \qquad (4b)$$

A Numerical Example: Consider the following arrangement:

32	0	32
18	14	32
50	14	64

A	B	A + B
C	D	C + D
A + C	B + D	N

In this case,

$$a = \frac{32 \times 50}{64} = 25$$

$A - a = 32 - 25 = 7$, and
$d = 14 - 7 = 7$

then

$$\text{Chi-square} = \frac{64(|7| - .5)^2}{25 \times 7} = 15.45$$

When a desk calculator is available, formulas (4a, b) have no particular advantage over formula (3), but definitely reduce the amount of labor and inaccuracy inherent in formulas (1) and (2). In addition, formulas (4a, b) have the conceptual advantage that, with the help of simple algebra, they show the relationship between the normal deviate *z*, and the one-*df* chi-square, when the normal question is used as an approximation to the binomial.

Let, $$p = \frac{a}{N} \qquad (5)$$

p, the probability that a random outcome will possess the attributes predicated in the row and the column converging in the cell labelled A.

Similarly, let, $$q = \frac{d}{N} \qquad (6)$$

q, the probability that a random outcome will possess the attributes predicated in the row and the column converging in the cell labelled D. Then, a = Np and Npq = ad/N. From these two relationships, it follows that formula (4a) can be written as:

$$\text{Chi-square} = \frac{(A - a)^2}{\frac{ad}{N}} = \frac{(A - Np)^2}{Npq} \qquad (7)$$

Finally, we get the square root of both sides of the equation (7) we obtain,

$$\text{Chi-square} = \frac{A - Np}{\sqrt{Npq}} = z,$$

the normal deviate

a relationship which is not obvious in the more common expressions of the chi-square, but which is given to the beginner to accept as an act of faith.

REFERENCES

GUILFORD, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill; 1956.

SNEDECOR, G. W. *Statistical methods.* (5th Ed.) Ames, Iowa: Iowa State Press, 1956.