guage that was not understandable. Going back again to them LM appeal, we may say that what was most meaningful to them would be those that were correctly expressed with implications that they could see personally applicable.

(b) The distortion that the message might have undergone as it trickled down from the government to these people;

(c) The possible blocks in the line of communication as well as the differences in background between that of the sender (in this case the government) and the receiver (in this case the people); and lastly

(d) The fact that communication, if it existed at all, seemed to be one-way in that these people were always receiving and were never able to send their own messages so that then the government could respond to it.

If the aforementioned speculations were true, then there is implied a responsibility to the effect that government officials should see to it that the projects and activities of the government be made understandable to these people, who, after all, may very well represent the majority of our population. The same could be said regarding the other established institutions of our society such as the Roman Catholic Church and civic organizations.

## *Comments:*

# On the Computation of the Chi Square from Derived Tables of Contingencies*

LEONEL CAMPOS**
*Ateneo de Manila University*
*Loyola Heights, Quezon City*

Occasionally it is desirable to modify a complex table of contingencies with multiple degrees of freedom (df), for the purpose of obtaining more meaningful information. A complex table of contingencies can be modified by: a) combining frequencies from adjacent categories; b) eliminating some category or categories altogether; or c) by a combination of both (a) and (b). A table that results after carrying out these procedures we will call, a *derived table of contingencies* (DTC for short).

Being able to modify a complex table of contingencies by any of the procedures described above becomes a need when:

1. some of the cells contain expected frequencies smaller than 3, since very small expected frequencies will tend to give spuriously large chi-square; and

2. the researcher wishes to test specific hypotheses about the data, which require the combination or/and the elimination of some categories.

In any case, the investigator is anxious to know that in modifying his original information he is following procedures

* This comment illustrates a statistical tool derived by others in the study of society.

** The author is connected with the Department of Behavioral Sciences of the Ateneo de Manila University.

which are both qualitatively meaningful and methodologically sound.

Let us remember, first, that the intact table of contingencies presumes a self-contained population defined exactly by the marginal totals. The relationships between the original categories imply specific relationships between the attributes existing in the population at hand.

In a DTC these relationships are disrupted, and we are left with the uncertainty that the conclusions drawn from it may not generalize to the whole intact table, which renders our operations invalid, and, at best misleading.

What we are trying to convey is the fact that DTC's introduce the serious problem of lack of goodness-of-fit: the possibility that our new table describes a population completely different from the population we intended to study.

Bresnahan and Shapiro (1966) have suggested a formula which allows the researcher to interpret his chi-square solely in terms of interactions, or the lack of them—independence—without having to worry about the lack of goodness-of-fit. The use of their formula obeys the observation of one simple rule:

*The expected frequencies are to be estimated from the marginal totals of the whole table, and not from the marginal totals of the derived table.*

Their equation is:

$$X^2d = \sum_j \sum_i \frac{(Oij - Eij)^2}{Eij} - \sum_i \frac{(Oi. - Ei.)^2}{Ei.}$$

$$\sum_j \frac{(O.j - E.j)^2}{E.j} + \frac{(Od - Ed)^2}{Ed}$$

where, $X^2d$, a chi-square for a derived table;

Oij and Eij, the observed and expected frequencies of the cell located in row *i* and column j, respectively.

Oi. and Ei., the observed and expected marginal totals of the *rows* of the derived table;

O.j and E.j, the observed and expected marginal totals of the *columns* of the derived table; and,

Od and Ed, the observed and expected total values of the derived table.

The observation of the rule given above assures the researcher that he is testing his derived table against a fair, representative sample of the original, intact population. It ought to be evident that combining or eliminating categories is analogous to sampling the population at hand; hence, the possibility of lack of goodness-of-fit.

The last three terms of equation (1) introduce a correction for lack of goodness-of-fit. The resulting chi-square for the DTC admits the same interpretation as an ordinary chi-square.

*An Example:*

The usefulness of formula (1), as well as its use, are now illustrated. In a recent study Pal, 1966)—

"A two-page schedule was administered to 162 informants in the barrios, and to 156 college students. The barrio informants supplied information themselves while the student informants gave information about their parents. The informants were asked to estimate the socio-economic status of their families in their respective communities, whether in the 1st 25, 2nd 25, 3rd 25, or 4th 25.

In Table 1 [reproduced here as table 1, also] the socio-economic status distribution of the informants' families is shown. (pp. 31-32)

TABLE 1. SOCIO-ECONOMIC STATUS DISTRIBUTIONS OF INFORMANTS' FAMILIES. (EXPECTED FREQUENCIES IN PARENTHESIS).

| Residence of Informants | 1st 25 | 2nd 25 | 3rd 25 | 4th 25 | Total |
|---|---|---|---|---|---|
| Rural | 27 | 71 | 52 | 12 | 162 |
| | (46.87) | (61.13) | (46.36) | (7.64) | (162) |
| College | 65 | 49 | 39 | 3 | 156 |
| | (46.13) | (58.87) | (44.64) | (7.36) | (156) |
| Total | 92 | 120 | 91 | 15 | 318 |
| | (92) | (120) | (91) | (15) | (318) |

$X^2 = 26.8884$
df = P < .001

TABLE 2. SAME DATA AS IN TABLE 1. THE 2ND-TO-4TH-GROUP CATEGORIES HAVE BEEN COMBINED (EIJ'S IN PARENTHESIS).

| Residence of Informants | 1st 25 | 2nd-4th 25 | Total |
|---|---|---|---|
| Rural | 27 | 135 | 162 |
| | (46.87) | (115.13) | (162) |
| College | 65 | 91 | 156 |
| | (45.13) | (110.87) | (162) |
| Total | 92 | 226 | 318 |
| | (92) | (226) | (318) |

In table 1 all Eij's are satisfactory, but it is possible to obtain derived tables for the purpose of testing specific hypotheses.

In reference to the data of table 1, (and with the aid of additional information, not reproduced here) Pal (1966) has commented that:

> "The 1st 25 group is a class apart from the rest while the 2nd 25, the 3rd 25, and the 4th 25 belong to the same class. On the basis of the economic criterion, the informants' families reflect communities of two-class social structure (p. 32)."

The raw information does not make this conclusion a compelling one; yet, it can be easily tested for accuracy.

First of all, notice that the chi-square for the whole table is 26.8894, which df = 3, has P < .001. We may conclude that a significant interaction exists, and Pal's viewpoint of homogeneity for the lower three fourths of the population yields the following hypotheses:

a) If we combine the 2nd-to-4th-group categories, a significant interaction will remain. This procedure will yield a partial test of the "two-class structure" viewpoint. We obtain the arrangement of table 2.

The combined observed frequency for the upper, right-hand cell is, 135 = 71 + 52 + 12. And its corresponding combined expected frequency is, 115.13 ≐ 61.13 + 46.36 + 7.64.

TABLE 3. SAME INFORMATION AS IN TABLE 1. THE 1ST-GROUP CATEGORY HAS BEEN ELIMINATED. (EIJ'S IN PARENTHESIS.)

| Residence of Informants | 2nd 25 | 3rd 25 | 4th 25 | Total |
|---|---|---|---|---|
| Rural | 71 | 52 | 12 | 135 |
| | (61.13) | (46.36) | (7.64) | (115.13) |
| College | 49 | 39 | 3 | 91 |
| | (58.87) | (44.64) | (7.36) | (110.87) |
| Total | 120 | 91 | 15 | 226 |
| | (120) | (91) | (15) | (226) |

91, and 110.87, in the lower right hand cell, are to be interpreted in an analogous manner.

It turns out, that in this instance, O.j = E.j; Oi = Ei.; and Od = Ed. As a consequence, we may conclude that table 2 is a fair sample of the original population (table 1), and the last three terms become zero, thus leaving the customary formula for the chi-square. This occurrence is not always obtained, of course.

The chi-square for table 2 is by formula (1).

$X^2d = 24.1609$, df $= 1$ $P < .001$. partially substantiating the "two-class structure" viewpoint.

However, if Pal's viewpoint of a "two-class structure" is an accurate description of table 1, it ought to occur that, if we eliminate the 1st 25 group, the remaining data ought to yield a small chi-square value. This is the crucial test of Pal's interpretation. Table 3 shows the results of eliminating the 1st 25 group.

In this case, the marginal totals for the columns conform to expectation, (i.e. O.j = E.j in every instance); in the same manner, Od = Ed. However, the marginal totals of the rows, do not conform to expectation, that is, the goodness-of-fit is deficient in this part of the table. In computing $X^2d$, we obtain,

$$\sum_j \sum_i \frac{(Oij - Eij)^2}{Eij} = \frac{(71 - 61.13)^2}{61.13} + \frac{(52 - 46.36)^2}{46.36}$$

$$+ \frac{(12 - 7.64)^2}{7.64} + \frac{(49 - 58.87)^2}{58.87}$$

$$+ \frac{(39 - 44.64)^2}{44.64} + \frac{(3 - 7.36)^2}{7.36}$$

$$= 9.7178$$

Then,

$$\sum_i \frac{(Oi. - Ei.)^2}{Ei.} = 0$$

$$\frac{(O.i - E.j)^2}{E.j} = \frac{(135 - 115.13)^2}{115.13} + \frac{(91 - 110.87)^2}{110.87} = 6.9903.$$

Finally,

$$(Od - Ed)^2 = O.$$

Therefore,

$$X^2d = [9.7178 - O - 6.9903 + 0 = 2.7275.$$

By eliminating the first column we lost df = 1, and we are left with df = 2. With df = 2, $X^2d$ = 2.7175 had $P > .05$, clearly, not significant, an indication that this part of the table is homogeneous. The soundness of Dr. Pal's conclusion is upheld.

Notice that if we had tested the homogeneity of table 3 by means of ordinary chi-square formulas, we would have obtained, $X^2 = 6.7366$ with $P < 0.5$, and would have arrived at false conclusions. The reasons for this discrepancy are:

a. the ordinary chi-square formula gives estimates of Eij from the DTC. Since the information from the original table is not fully used the procedure violates the requirement of exhaustiveness in the utilization of data, necessary for correct interpretation of chi-square.

b. the ordinary chi-square formula does not correct for lack of goodness-of-fit; as a matter of fact, it does not even consider the possibility that this may occur.

The Bresnahan-Shapiro formula (Bresnahan and Shapiro, 1966), by requiring that Eij's be estimated from the intact table satisfies the criterion of exhaustiveness and by correcting for lack of goodness-of-fit, assures a fair sample of the population at hand. In this manner, conclusions arrived at on the basis of DTC's can be generalized with confidence to the whole intact table.

Finally, it is possible to use the Bresnahan-Shapiro formula to systematically explore a complex table of contingencies by successive combinations and/or elimination of categories of data. This procedure, requires additional rules to assure independence of the chi-square derived from derived table.

For the time being we will simply point that we obtained two DTC's with 1 and 2 df, respectively, which add up to df = 3, the df of the intact table. Also the sum of the $X^2d$'s is 24.1609 + 2.7275 = 26.8884 ($X^2 = 26.8884$) within negligible computational error replicate the total chi-square obtained for the intact table.

We intend to discuss this matter in a subsequent communication.

Bresnahan, J., and Shapiro, M. "A General Equation for the Exact Partitioning of Chi-Square Contingency Tables," 66, 1966 *Psychological Bulletin,* 252-262.

Pal, Agaton, "Aspects of Lowland Philippine Social Structure,"XIV: 1 *Philippine Sociological Review,* 1966, 31-39.

# Book Reviews

Burgess, E., Locke and Thomes, *The Family,* third edition. New York: The American Book Company, 1963, 582 pp.

There are four outstanding features of *The Family* by Burgess, Locke and Thomes: (1) the use of personal documents which are effectively presented at the beginning of each important topic to illustrate the problems and to present illumination of concepts; (2) the employment of the ideal type method developed by Max Weber of identification, isolation and accentuation of logical extremes; (3) the presentation of findings from various studies on the family; and (4) the suggestion of other areas to be studied.