

# Critical factors in the chi-square test of independence: A technique for exploratory data analysis

L. W. BUCKALEW and W. H. PEARSON

*Air Force Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base  
Dayton, Ohio 45443*

The chi-square test of independence was reviewed, with concerns voiced for limitations traditionally imposed on the specificity of interpretation allowed given a significant chi square. Situations encouraging more exacting interpretation were noted, and a technique was proposed that would allow statistical consideration of factors, delineated by a row or column, that make appreciable contribution to a statistically significant chi-square value. Suggestions were offered as to the interpretation and utilitarian value of identified critical factors in a chi-square problem.

Chi square is a highly utilitarian nonparametric statistic facilitating the analysis of nominal data, typically in the form of frequencies. Traditionally, this inferential statistic compares observed frequencies with expected, with the latter generated by the null hypothesis or by theory. Conventional treatment of this statistic (Ferguson, 1980; Siegel, 1956) reflects two basic forms: tests of goodness of fit and tests of independence. In both cases, chi square reflects discrepancies between observed and expected (theoretical) frequencies. In the case of testing goodness of fit, a significant chi square allows the conclusion that the sample is biased with respect to known population values or the distribution of some theoretical population. A technique for more exacting interpretation, allowing specific cell (condition/category) consideration, was proffered by Buckalew and Pearson (1981).

The chi-square test of independence involves two variables, such as groups and conditions/categories, and seeks to determine whether one variable is independent of the other. With this test, a significant chi square indicates nonindependence, that is, that an association exists between variables. While additional information in the form of a contingency coefficient is available to reflect the strength of an association, interpretation of a significant chi square traditionally is limited to statements that groups are differentially distributed over categories or conditions. This limitation on the degree of specificity traditionally allowed with chi square can prove frustrating, particularly when a general trend appears to exist within data or when there is an a priori hypothesis as to why the chi square should be significant. This note suggests the use of computational components

The first author accomplished this work while assigned to the Summer Faculty Research Program. Requests for reprints should be addressed to L. W. Buckalew, Department of Psychology, Box 200, Alabama A & M University, Normal, Alabama 35762.

of a significant chi square to facilitate more specific interpretation as exploratory data analysis.

Sophisticated considerations and applications of chi-square analysis may be found in Goodman (1970), and Guilford (1965) and Hays (1963) have addressed the use of chi square to approximate multinomial probabilities or the log transformation of a probability into a chi square. Specifically considering the chi-square test of independence, Guilford (1965) noted that many sets of observations provide a number of frequency values, for each of which a null hypothesis can be stated and a test of departures of observed from expected frequencies is needed. This is tantamount to computing a separate chi square for each row and column of a contingency table, with the traditional chi square serving to make such combined tests. The proposed exploratory data analysis technique simply seeks to identify the source of the overall significance in chi square. From a related framework, Guilford posed that chi square constituted the sum of ratios: squared differences between observed and expected frequencies to expected frequency. The proposed technique operationally looks for disproportionate ratios for a row or column of a chi-square table, respecting the additive property of chi-square components.

In the conceptual framework proposed, the traditional two variables of a chi-square table are viewed as composed of factors, with each row and column constituting a factor. The proposed extension of chi square will be referred to as a critical factor test ( $\chi^2_f$ ). The rationale for this test is that, if a single row (i) or column (j) summation of chi-square ratios,

$$\sum_{i \text{ or } j} [(O_{ij} - E_{ij})^2 / E_{ij}],$$

meets or exceeds the chi-square value needed for significance (critical value), using the df of  $(r - 1)(c - 1)$  for

the total chi-square problem, then that row or column factor's influence may be considered significant, that is, a critical factor ( $\chi^2_f$ ). In essence, all the critical factor test requires is the summation of cell  $(O - E)^2/E$  for any given row or column and comparison of this  $\chi^2_f$  with the critical value needed for the overall chi square. If  $\chi^2_f$  is equal to or greater than the critical value for chi square, then that factor, identified by the row or column heading, is said to be significant in exploratory data analysis.

This critical factor test does not preclude computation of the overall chi square. It simply uses existing computational components and is only intended to suggest a potential cause or source of O - E differences responsible for a significant chi square. The interpretation of a  $\chi^2_f$  involves a statement relative to the origin of the overall chi-square significance. When a column (category/condition) yields a significant  $\chi^2_f$ , it may be interpreted as differences between groups (rows) in how they are represented within that particular category or condition (column). When a row (group) yields a significant  $\chi^2_f$ , it may be interpreted as differences over conditions or categories (columns) in terms of how all groups compare with a specific group. The subsequently identified factor has the potential for appreciable utilitarian value to the researcher.

The proposed test for a critical factor first requires computation of the overall chi square and may be applied only if the resulting value is significant. The research design must involve more than two groups ( $r > 2$ ) and more than two conditions/categories ( $c > 2$ ), that is, at least a 3 by 3 chi-square table. Also, if one or more

cells is a significant contributor in the sense of Buckalew and Pearson (1981), useful interpretation of a  $\chi^2_f$  is severely compromised. Also, the utilitarian value of an identified critical factor decreases appreciably with each additional such factor identified in the chi-square table. The traditional (Ferguson, 1980; Siegel, 1956) limitations or requirements for chi-square computation must, of course, be respected, with detailed considerations thereof offered by Hays (1963). Ultimately, obtaining a significant  $\chi^2_f$  would operationally identify the source of significance in the overall chi square, as alluded to by Guilford (1965), and could functionally suggest directions for further study of the identified phenomenon. Additional statistical considerations of chi-square analysis, inclusive of interactions, are offered by Goodman (1970).

#### REFERENCES

- BUCKALEW, L. W., & PEARSON, W. H. Determination of critical observed frequencies in chi square. *Bulletin of the Psychonomic Society*, 1981, **18**, 289-290.
- FERGUSON, G. A. *Statistical analysis in psychology and education* (5th ed.). New York: McGraw-Hill, 1980.
- GOODMAN, L. A. The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, 1970, **65**, 226-256.
- GUILFORD, J. P. *Fundamental statistics in psychology and education* (4th ed.). New York: McGraw-Hill, 1965.
- HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received for publication February 1, 1982.)