

LEO A. GOODMAN¹*University of Chicago*

NEW METHODS FOR THE ANALYSIS OF
TWO-WAY CONTINGENCY TABLES:
AN ALTERNATIVE TO DIACONIS AND EFRON

I would like to congratulate Persi Diaconis and Bradley Efron on their very interesting and stimulating paper (1985). Their paper introduces into the literature on the analysis of contingency tables new ideas and methods, and a new level of mathematical sophistication and depth, pertaining to the interpretation of the chi-square statistic. I appreciate very much the merits of their approach; and I would like here to suggest an alternative approach to the analysis of contingency tables, and to indicate some of the merits of this alternative. I shall illustrate this alternative approach by applying it in an analysis of the same two contingency tables (Tables 1 and 2) that were used by Diaconis and Efron to illustrate their methods.

I begin with Table 1, a 4×4 cross-classification of 592 subjects with respect to eye color (the row classification) and hair color (the column classification). For this table, Diaconis and Efron obtain a chi-square value of 138.29, with $3 \times 3 = 9$ degrees of freedom, using the usual goodness-of-fit (i.e., Pearsonian) chi-square statistic for testing the usual model H_0 of independence between the row classification and column classification. They then proceed to provide new and interesting interpretations of this statistic. The alternative approach that I wish to suggest here, for analyzing contingency tables of this kind, leads me, when analyzing this 4×4 table, to modify the usual model H_0 of independence by introducing 2 parameters pertaining to the association in the table, and the model thus obtained (called here model H') turns out to reduce the goodness-of-fit chi-square value from 138.29 to 10.48, with $9 - 2 = 7$ degrees of freedom (d.f.). This reduction of 92 percent in the chi-square value, with the introduction of only 2 association parameters in model H' , is quite dramatic. The 2 parameters provide a parsimonious description of the association in this table; the estimate of these association parameters describes the magnitude of this association; the relatively small chi-square value indicates that the data in this table are congruent (more or less) with H' ; the 7 d.f. corresponding to H' indicate that this model makes 7 distinct assertions about the association in the table; and the relatively small chi-square value indicates that the data are congruent (more or less) with those 7 assertions.²

¹ Support for this work was provided in part by National Science Foundation Grant SES-8303838. This work was carried out while the author was a Fellow of the Center for Advanced Study in the Behavioral Sciences, with financial support provided in part by National Science Foundation Grant BNS-8011494 to the Center.

² Model H' is an example of the kind of model introduced in my work on "association models" (see, e.g., 1979, 1981), as are the other models that will be brought forward in this discussion (viz., models H'' , M' , M'' , and the other association models in Tables A and B). Due to space constraints imposed by the Editor on this discussion, the details about these particular models, and their application in the present context, will be discussed more fully elsewhere.