

Grizzle J E, Starmer C F & Koch G G. Analysis of categorical data by linear models.
Biometrics 25:489-504, 1969.
[Depts. Biostatistics and Biomathematics, Univ. North Carolina, Chapel Hill, NC]

The paper shows how weighted least squares analysis can be used to fit models to categorized data (contingency tables). Tests of interest in contingency often correspond to tests made routinely in multiple regression, except that a preliminary transformation of the multinomial probabilities in the contingency table is necessary. [The *SCI*[®] indicates that this paper has been cited over 145 times since 1969.]

James E. Grizzle
Department of Biostatistics
School of Public Health
University of North Carolina
Chapel Hill, NC 27514

August 30, 1979

"I wrote my dissertation on analyzing binomial data by fitting logistic models.¹ From this beginning I continued to keep abreast of the research in the analysis of contingency tables, and in the mid-1960s I was asked to organize a graduate course on the subject. Data analysis was to be emphasized in the course rather than theory. Another important part of my background was courses on the analysis of contingency tables on other multivariate topics taught by S.N. Roy. Roy always emphasized general approaches to problems rather than *ad hoc* solutions.

"During a survey of the literature in preparation for the course, I became impressed that applied papers on contingency tables emphasized special case solutions to problems and that it might be possible for students to spend considerable time on problem solving without mastering any useful general techniques. I began to try

to develop an approach that could unify many of the papers I was

reviewing and at the same time be accessible to intermediate level students. It soon became clear that the connection between the modified minimum X^2 method and weighted least squares provided an easy unified approach to many analysis problems. It was not hard to develop the facility with multivariate weighted least squares that was necessary to handle the problem involving several multinomial distributions.

"At this time I was tutoring Frank Starmer, who was a student in biomathematics and an excellent computer programmer. As I developed the understanding of how to incorporate various special cases into the general methodology, Frank analyzed some typical problems. Some of them required generalization of the approach we were developing. Eventually we became satisfied with the formulation and degree of generalization presented in the paper. Gary Koch critiqued our work from time to time and made some important suggestions. Subsequently Gary and his students went much further with models of the same general kind. Their work is often cited and is partly responsible for the popularity of this paper.

"In my opinion the reason why this paper is cited is that it shows how the powerful method of multiple regression can be applied to contingency tables. This has made it possible to unify teaching. It has also had the equally important benefit of emphasizing the value of modeling for contingency tables and of testing hypotheses that are specific to the research problem being analyzed."

1. Grizzle J E. Application of the logistic model to analyzing categorical data. Unpublished thesis. North Carolina State Univ., 1960. 76 p.