DOCUMENT RESUME

ED 397 113 TM 025 246

AUTHOR Anderson, Carolyn J.; Hsieh, Ju-Shan
TITLE Logit Multiplicative Models: Alternatives to Saturated Logit/Loglinear Models for 3-Way Tables.
PUB DATE 3 Apr 96
NOTE 25p.; Sections of this paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Interaction; *Mathematical Models; *Predictor Variables; Scaling; *Statistical Analysis
IDENTIFIERS *Logit Multiplicative Models; Log Multiplicative Models

ABSTRACT

When the highest-way association is present in a 3-way cross-classification of frequencies, standard logit and loglinear models have an many parameters as there are cells in the table; that is, the models are "saturated." Extensions of logit and loglinear models are described here that provide more parsimonious alternatives to saturated models. The new models, logit multiplicative models, and their equivalent log multiplicative models, are introduced here for the case where there is one dichotomous response or criterion variable and two (polytomous) explanatory or predictor variables. In logit multiplicative models, the interaction between the explanatory variables is represented by the product of scale values for the categories of the explanatory variables and a measure of the strength of the association. Plots of the scale values provide graphical representations and descriptions of the interaction. The new models are illustrated by modeling a 3-way interaction between whether an elementary school student attends an extracurricular tutoring program, the highest educational level attained by the student's father, and the student's grade level. (Contains 1 figure, 3 tables, and 33 references.) (Author)

Logit Multiplicative Models: Alternatives to Saturated Logit/Loglinear Models for 3–Way Tables

Carolyn J. Anderson
and
Ju-Shan Hsieh

Department of Educational Psychology

University of Illinois

Last Revised April 3, 1996

Sections of this paper were presented at the 1996 Annual Meeting of the American Educational Research Association, NY City, NY. Computer programs used to fit models presented are available from the author. Correspondence should be addressed to Carolyn Anderson, University of Illinois, MC-708, 1310 South Sixth Street, Champaign, IL, 61820–6990, (217) 244-3537, FAX (217) 244-7620, e-mail: cja@uiuc.edu

Abstract

Logit Multiplicative Models: Alternatives to Saturated Logit/Loglinear Models for 3–Way Tables

When the highest–way association is present in a 3–way cross-classification of frequencies, standard logit and loglinear models have as many parameters as there are cells in the table; that is, the models are "saturated". Extensions of logit and loglinear models are described here that provide more parsimonious alternatives to saturated models. The new models, logit multiplicative models and their equivalent log multiplicative models, are introduced here for the case where there is one dichotomous response or criterion variable and two (polytomous) explanatory or predictor variables. In logit multiplicative models, the interaction between the explanatory variables is represented by the product of scale values for the categories of the explanatory variables and a measure of the strength of the association. Plots of the scales values provide graphical representations and descriptions of the interaction. The new models are illustrated by modeling a 3–way interaction between whether an elementary school student attends an extra-curricular tutoring program, the highest educational level attained by the student's father, and the student's grade level.

Keywords: loglinear models, logit models, 3-way interactions, latent variables, scaling, categorical data analysis.

1 Introduction

Variables in educational and social science research are often discrete, such as gender, race, grade level, types of programs of study, whether a course is passed or failed, the highest degree earned by a parent, plans after graduation, desired occupation, and actual occupation. Other variables are continuous, but are measured descretely, such as socio-economic status, ability, and achievement. With respect to ability and/or achievement, the observed variables are typically whether a student selects a correct or incorrect answer on an objective test item, or the actual response option selected on a multiple choice item.

The standard approach to analyzing multivariate categorical is to use loglinear or logit models. These models are extremely useful for identifying interactions that are present in multivariate categorical data; however, they are not as useful for helping to describe the nature of the relationships that do exist. Unlike continuous variables where a single number such as a correlation coefficient is sufficient to summarize the association between two variables. the number of statistics (i.e., odds ratios) needed to characterize the association between categorical variables is an increasing function the number of categories of the variables. For example, the minimum number of odds ratios that are needed to completely describe the relationship between just two variables each of which has 4 levels is $(4-1)(4-1)=9$. When associations exist among three or more variables, the number of statistics needed to describe the relationship is larger making the problem of interpreting and describing interactions even more difficult.

In the case of two categorical variables, the multidimensional row–column or "$RC(M)$" association model developed by Goodman (1979, 1985, 1986, 1991) is extremely useful for summarizing and describing the relationship between two variables (also see Agresti, 1990; Clogg & Shihadeh, 1994; Wickens, 1989). The $RC(M)$ association model is an extension of the loglinear model for two–way tables. Interactions are represented in $RC(M)$ models by the product of scale values assigned to categories of the variables and a measure of the strength of the relationship. Plots of scale values provide graphical representations of the association between variables, which greatly aid interpretation. Numerous generalizations of the $RC(M)$ association model for three or more variables have been proposed (Anderson, 1996; Becker, 1989a; Becker & Clogg, 1989; Choulakian, 1988; Clogg, 1982a, 1982b; Gilula

& Haberman, 1988; Goodman, 1986; Mooijaart, 1991); however, the existing model generalizations do not address the situation where one variable is a dichotomous response or criterion variable and the other variables are explanatory variables. It is this situation that we are concerned with here.

Loglinear and $RC(M)$ association models represent the relationships between categorical variables without making distinctions regarding the role that particular variables play in an analysis. When one variable is a response or criterion variable and the rest of the variables are explanatory or predictor variables, logit models are often preferable to loglinear models, even though logit models are equivalent to loglinear models. Logit models are simpler than their equivalent log linear models. Logit models only include terms that represent the associations between the criterion and the explanatory variables, and they do not contain terms that represent the relationship between the explanatory variables. Extensions of logit models, logit multiplicative models, are presented here that are analogous to the $RC(M)$ association model. In the logit multiplicative models, the interactions between the explanatory variables are represented by products of scale values multiplied times a measure of the strength of the association. The models are equivalent to one of the models in the family of association model generalizations proposed by Anderson (1996).

In Section 2, the logit multiplicative model for the specific case of a dichotomous response variable and two explanatory variables is described. Since fitting the models is a non-trivial problem and cannot be done using standard procedures in readily available statistical software packages, two methods currently available for estimating the new models are discussed in Section 3. As an example, the models are used to analyze data from a study by Hsieh (1996) on the effects of extra-curricular tutoring programs on mathematics achievement in elementary school children in Taiwan.

2 The Logit Multiplicative Model

In Section 2.1, the logit multiplicative model is presented as an extension of a standard logit model, followed in Section 2.2 by a discussion of the identification constraints needed to estimate model parameters. Lastly, in Section 2.3, the interpretation of the model and its parameters is discussed.

2.1 The Basic Model

Let F_{ijk} equal the number of individuals (students, parents, subjects, objects, etcetera) who fall into categories i, j, and k of variables A, B, and C, respectively, where variables A and B are explanatory variables and variable C is a dichotomous response variable (i.e., $k = 1, 2$). The number of categories of variable A equals I (i.e., $i = 1, \ldots I$) and the number of categories of variable B equals J (i.e., $j = 1, \ldots, J$).

We start with the standard logit model with two explanatory variables where the "dependent" variable is the odds of one response versus the other (i.e., F_{ij1}/F_{ij2}). The most complex or saturated model is

$$\frac{F_{ij1}}{F_{ij2}} = \beta\beta_{A(i)}\beta_{B(j)}\beta_{AB(ij)} \tag{1}$$

where β is a constant, $\beta_{A(i)}$ and $\beta_{B(j)}$ are "main" or marginal effect terms for variables A and B, respectively, and $\beta_{AB(ij)}$ is the interaction term. Since the odds of response 1 versus response 2 is a multiplicative function of the model parameters, the logarithm of F_{ij1}/F_{ij2} is a linear function of model parameters; that is,

$$\ln\left(\frac{F_{ij1}}{F_{ij2}}\right) = \tau + \tau_{A(i)} + \tau_{B(j)} + \tau_{AB(ij)} \tag{2}$$

where $\tau = \ln(\beta)$ is a constant, $\tau_{A(i)} = \ln(\beta_{A(i)})$ and $\tau_{B(j)} = \ln(\beta_{B(j)})$ are "main" effect terms, and $\tau_{AB(ij)} = \ln(\beta_{AB(ij)})$ is the interaction term. This model has as many unique parameters as there are data points, which in this case equals the number of odds that can be formed for the various combinations of variables A and B (i.e., $I \times J$). The saturated logit model (equations 1 and 2) will always fit the data perfectly.

To obtain a more parsimonious and simpler representation of the data, we note that the interaction terms $\tau_{AB(ij)}$ in equation 2 (or $\beta_{AB(ij)}$ in equation 1) are unstructured in the sense that they equal whatever they need to equal so that the data are fit perfectly. The new models presented here impose a multiplicative structure on these terms and break the interaction down into component pieces as follows:

$$\ln\left(\frac{F_{ij1}}{F_{ij2}}\right) = \tau + \tau_{A(i)} + \tau_{B(j)} + \sum_{m=1}^{M} \phi_m\mu_{im}\nu_{jm} \tag{3}$$

where M equals the number of components or dimensions used to represent the interaction, μ_{im} and ν_{jm} are scale values for categories i and j of variables A and B, respectively, on dimension m, and ϕ_m is a measure of the strength of the interaction between A and B on dimension m. When $M = \min(I, J) - 1$, equation 3 is equivalent to the saturated logit model; however, when $M < \min(I, J) - 1$, the model is not saturated and provides a summary of the interaction. In practice, models such as equation 3, the $RC(M)$ association model and it's various generalizations, typically only need a small number of dimensions (i.e., 1 or 2) to adequately fit data.

Assigning numbers to the categories of "ordinal" variables seems natural, but what about scaling the categories of "nominal" variables? While no *a priori* ordering of categories may exist (i.e., a variable is "nominal"), when considering the relationship between observed variables, there may be an ordering of the categories on some underlying or latent (continuous) dimension. Even when categories have an *a priori* ordering (i.e., an "ordinal" variable), this ordering may not be the appropriate one for describing an interaction between variables. Since the parameters of model 3, including the scale values and association parameters, are estimated from the data, we can discover the appropriate ordering and relative spacing between categories that is needed to summarize the interactions by fitting the model to data.

2.2 Identification Constraints

To estimate the parameters of equation 3, identification constraints on the model parameters are required. These constraints do not effect the predicted or "fitted" values, and thus do not effect how well the model fits the data. The identification constraints do effect the actual numerical values of estimated parameters. The constraints imposed on the $\tau_{A(i)}$'s and $\tau_{B(j)}$'s are the same as those typically imposed on the analogous terms in logit and loglinear models; that is, either a particular term is set equal to a constant (e.g., $\tau_{A(1)} = \tau_{B(1)} = 0$), or the sum of the terms for a variable is set equal to a constant (e.g., $\sum_{i=1}^{I} \tau_{A(i)} = \sum_{j=1}^{J} \tau_{B(j)} = 0$).

The sets of scale values for each variable need to be centered and scaled. The centering constraints set the location of the scale. The centering constraints used here are

$$\sum_{i=1}^{I} \mu_{im} h_{A(i)} = 0 \tag{4}$$

$$\sum_{j=1}^{J} \nu_{jm} h_{B(j)} = 0 \tag{5}$$

where $h_{A(i)}$ and $h_{B(j)}$ are fixed and known weights for categories i and j of variables A and B, respectively. Possible choices for $h_{A(i)}$ and $h_{B(j)}$ include unit weights, uniform weights (i.e., $h_{A(i)} = 1/I$ and $h_{B(j)} = 1/J$), or marginal probabilities. Becker and Clogg (1989) discuss the choice of weights for the $RC(M)$ association model, and their results apply here as well.

The scaling constraints used here are

$$\sum_{i=1}^{I} \mu_{im}^2 h_{A(i)} = \delta_{mm'} \tag{6}$$

$$\sum_{j=1}^{J} \nu_{jm}^2 h_{B(j)} = \delta_{mm'} \tag{7}$$

where $\delta_{mm'} = 1$ for $m = m'$ and 0 for $m \neq m'$. These constraints set the unit of measurement and constrain the scale to be orthogonal across dimensions.

In sum, the constraints on the scale values can be thought of as setting the means of the sets of scale values equal to zero, the variances equal to one, and the covariances (or correlations) between scales equal to zero. Thus, the scale values are an interval level measure on underlying continuous variables. Linear transformations of the scale values will not effect the fitted (predicted) values.

Given the identification constraints, the degrees of freedom for the logit multiplicative model in equation 3 can now be computed. The degrees of freedom equals the number of data points minus the number of unique parameters (i.e., the number of parameters in equation 3 minus the number of constraints needed to identify them). Thus, the degrees of freedom equal

$$df = (I - M - 1)(J - M - 1) \tag{8}$$

2.3 Interpretation of Model Parameters

In logit multiplicative models, as well as in standard logit, loglinear, and $RC(M)$ association models, interactions between variables are defined in terms of odds ratios. A direct relationship exists between odds ratios and the model parameters. Since interactions between categorical variables are defined in terms of odds ratios (and for three variables, ratios of odds ratios), this relationship has implications regarding the proper interpretation of the logit multiplicative model.

Let $\theta_{ii'(j)}$ equal the ratio of the odds F_{ij1}/F_{ij2} to $F_{i'j1}/F_{i'j2}$, which is an odds ratio for variables A and C conditional on category j of variable B; that is,

$$\theta_{ii'(j)} = \frac{F_{ij1}/F_{ij2}}{F_{i'j1}/F_{i'j2}} = \frac{F_{ij1}F_{i'j2}}{F_{ij2}F_{i'j1}}$$

If there is no interaction between variables A and B in their relationship to variable C (i.e., no 3–way association among variables A, B and C), then the conditional odds ratio given category j equals the conditional odds ratio for any other category of variable B (i.e., $\theta_{ii'(j)} = \theta_{ii'(j')}$ for all $i, i' = 1, \ldots, I$, and $j, j' = 1, \ldots, J$). Alternatively, we can consider odds ratios conditioning on the categories of variable A; that is,

$$\theta_{jj'(i)} = \frac{F_{ij1}F_{ij'2}}{F_{ij'1}F_{ij2}}$$

If there is no interaction between variables A, B and C, then the conditional odds ratios for all categories of variable A will all be equal (i.e., $\theta_{jj'(i)} = \theta_{jj'(i')}$ for all $i, i' = 1, \ldots, I$, and $j, j' = 1, \ldots, J$). Thus, when no 3–way association exists,

$$\Theta_{ii',jj'} = \frac{\theta_{jj'(i)}}{\theta_{jj'(i')}} = \frac{\theta_{ii'(j)}}{\theta_{ii'(j')}} = 1$$

If a 3–way interaction does exist, then $\Theta_{ii',jj'} \neq 1$, or equivalently $\ln(\Theta_{ii',jj'}) \neq 0$ for at least some $i, i' = 1, \ldots, I$, and $j, j' = 1, \ldots, J$.

In terms of the parameters of the logit multiplicative model, the logarithm of the ratio of conditional odds ratios equals

$$\ln(\Theta_{ii',jj'}) = \sum_m \phi_m(\mu_{im} - \mu_{i'm})(\text{[illegible]}) \tag{9}$$

The associations in the data that are attributable to the 3–way relationship are represented by the scale values and the association parameter, and not by any of the other terms in the model. The scale values provide information about the structure of the interaction between variables A and B in their relation with variable C. Since the scale values provide an interval level measure on latent continuous variables, only the relative differences between scale values for the categories of variable A are meaningful, as well as the relative differences between the scale values for categories of variable B. Categories with the relatively larger differences between their scale values have (conditional) odds ratios that are more dissimilar and thus their ratio $\Theta_{ii',jj'}$ is further from 1 than do categories with relatively smaller differences. Alternatively, categories with nearly equivalent scale values, have nearly equivalent odds ratios, and thus $\Theta_{ii',jj'}$ for these categories will be close to 1, the point of no association.

Plots of scale values provide "pictures" of the relationship among the variables. Such "pictures" are graphical representations of all possible ratios of odds ratios that can be formed. These plots greatly facilitate the substantive interpretation and description of interactions in data. The geometry and interpretation of such plots is similar to that of plots of scale values from the $RC(M)$ association model (see Goodman, 1986; Clogg, 1986), except that interactions are defined in terms of ratios of odds ratios rather than just odds ratios. In these plots, the relative distances between points provides information about the relationship between the variables. An example of such a plot and its interpretation is given in Section 4.

To gain insight into the meaning and interpretation of ϕ, we first examine the simple case of the one dimensional model where

$$\ln(\Theta_{ii',jj'}) = \phi(\mu_i - \mu_{i'})(\nu_j - \nu_{j'})$$

For a one unit change in the scale for variable A (i.e., $(\mu_i - \mu_{i'}) = 1$) and a one unit change in the scale for variable B (i.e., $(\nu_j - \nu_{j'}) = 1$), ϕ is the logarithm of the ratio of conditional odds ratios. In other words, the association parameter ϕ is a measure of the strength of the relationship between variables A, B and C. In the case of two or more dimensions, ϕ_m measures the strength of the relationship on the mth latent dimension.

At times, we may wish to consider the "effect" of one variable holding the other variable constant; that is, we may want to consider what happens to the odds F_{ij1}/F_{ij2} when we change to category i' of variable A. To examine such effects, we look at the appropriate conditional odds ratios, which in this case is $\theta_{ii'(j)}$. In terms of the logit model parameters, this equals

$$\ln(\theta_{ii'(j)}) = (\tau_{A(i)} - \tau_{A(i')}) + \sum_m \phi_m(\mu_{im} - \mu_{i'm})\nu_{jm}$$

Thus, the difference between the odds ratio for category i and that for category i' depends on both the main effect of variable A and the interaction effect between variables A and B. Alternatively, if we hold the level of variable A constant and examine the change in odds ratios between categories j and j' of variable B, we find that the change in odds ratios depends on both the marginal effect of B and the interaction between A and B; that is,

$$\ln(\theta_{jj'(i)}) = (\tau_{B(j)} - \tau_{B(j')}) + \sum_m \phi_m\mu_{im}(\nu_{jm} - \nu_{j'm})$$

Rather than odds and odds ratios, for some purposes it is more convienent to use predicted probabilities or relative frequencies. The fitted odds can be transformed to probabilities as follows:

$$\begin{aligned} \pi_{ij1} &= \frac{1}{1 + (F_{ij1}/F_{ij2})^{-1}} \\ &= \frac{1}{1 + e^{-(\tau + \tau_{A(i)} + \tau_{B(j)} + \sum_m \phi_m\mu_{im}\nu_{jm})}} \end{aligned} \tag{10}$$

where π_{ij1} is the predicted probability of response 1 given categories i and j of variables A and B, respectively. However, with respect to interpreting the model parameters, the interpretation using equation 10 is not as simple, straight forward, or direct as it is when we discuss odds and odds ratios.

3 Estimation

While the model can be estimated by least squares, only maximum likelihood estimation under the standard sampling assumptions of independent, homogeneous observations from either a Binomial or Poisson distribution is discussed here. The same inherent difficulties encountered when estimating the $RC(M)$ association model apply to estimating logit multiplicative models (see Haberman (1995) for a discussion of the difficulties involved in estimating the $RC(M)$ association model). Two currently available methods of estimating the models are briefly outlined here. One method uses common statistical packages, but this method requires specially written modules and can only be used to estimate a one dimensional model (i.e., $M = 1$). This method is described in Section 3.1. The other method makes use of the equivalence between the logit multiplicative model and a special case of one model from the family of models proposed by Anderson (1996). This family of models, 3–mode association models, are generalizations of the $RC(M)$ association model to 3–way tables. This latter method makes use of a program written to estimate the entire family of 3–mode associations models developed by Anderson (1996). In Section 3.2, the equivalence between the logit multiplicative model and the 3–mode association model is given, as well as some general comments about the algorithm used in the program.

3.1 Uni-Dimensional Model

Uni-dimensional model can be estimated using software that estimates *generalized linear* models (Dobson, 1990; McCullagh & Nelder, 1990), such as *GLIM* (Francis, Green & Payne, 1993) or *SAS/GENMOD* (SAS Inc., 1994). Generalized linear models are extensions of traditional linear models that have two basic parts: a structural component and a random component. The structural component is a linear function of the predictor or explanatory variables. The random component is a probability distribution for the response variable, which can be any distribution from an exponential family of distributions. The "link function" describes how the mean of the response variable is related to the linear predictor. The procedure used to estimate generalized linear models can be used to fit the one dimensional logit multiplicative model.

Both the one dimensional RC model and the logit model are generalized bilinear models (as opposed to generalized linear models). The RC model has a *log* link function and its random component is the *Poisson* distribution, while the logit bilinear model has a *logit* link function and its random component is the *Binomial* distribution. By changing the link function and the distribution, the algorithm given by Becker (1989b) for estimating the one dimensional RC association model using generalized linear models can be modified to fit the one dimensional logit model (or any generalized bilinear model).

The procedure to estimate logit bilinear models is iterative and requires starting values for the scales values. To describe the basic steps required for the iterative portion of the procedure, we write equation 3 as

$$\ln(F_{ij1}/F_{ij2}) = \tau + \tau_{A(i)} + \tau_{B(j)} + x_i y_j \tag{11}$$

where $\phi\mu_i\nu_j = x_i y_j$. Variables A and B are declared to be classification variables and x and y are alternately treated as numerical variables and parameters that are to be estimated. In one step, equation . is fit to data with x_i's treated as the values of a numerical variable and the y_j's are estimated parameters. On the next step, the new estimates of y_j are treated as the values of a numerical variable and the x_j's are estimated parameters. This process is repeated until the change between fitted values on successive cycles is less than some specified criterion (i.e., a very small number). After the solution has converged, the identification constraints are imposed on the scales values (i.e., $\hat{\mu}_{im} = a\hat{x}_i + b$, and $\hat{\nu}_{jm} = c\hat{y}_j + d$, where a, b, c and d are constants such that $\sum_i \hat{\mu}_i h_{A(i)} = \sum_j \hat{\nu}_j h_{B(j)} = 0$ and $\sum_i \hat{\mu}_i^2 h_{A(i)} = \sum_j \hat{\nu}_j^2 h_{B(j)} = 1$). The model is estimated a final time using the product $\hat{\mu}_i\hat{\nu}_j$ as a numerical variable to obtain an estimate of ϕ and the final estimates of the other parameters in the model. The fit statistics for the final model are correct, but the degrees of freedom and estimated standard errors of the model parameters given by the program are not correct. The correct degrees of freedom are given in equation 8.

Since the procedure requires iteratively fitting models, in practice, modules written to perform the cycles and steps within each cycle are used. Becker (1989b) describes one such module for the program *GLIM*. A module using *SAS/GENMOD* is available from the author. One advantage of this procedure is that it uses existing and generally available software. Another advantage is that the model statement can be readily modified to fit more complex bilinear models. For example, more variables can be included and additional bilinear terms for other 2–way interactions can be estimated (e.g., Anderson & Wasserman, 1995). The major disadvantage of this method is that it cannot be used to estimate multiple dimensions, which is why we turn to a second method of estimation.

3.2 Multi-Dimensional Model

The estimation procedure described here can be used to estimate both uni- and multi-dimensional models. Equation 3 is a special case of one class of 3-mode association models proposed by Anderson (1996). Three-mode association models, which are log multiplicative models, are extensions of the saturated loglinear model for 3-way tables, as well as generalizations of the $RC(M)$ association model to 3-way tables. The 3-mode association model that is equivalent to the logit multiplicative model is

$$\begin{aligned} \ln(F_{ijk}) &= \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)} + \lambda_{AB(ij)} + \lambda_{AC(ik)} + \\ &\quad \lambda_{BC(jk)} + \sum_{r=1}^{R}\sum_{s=1}^{S}\sum_{t=1}^{T} \phi_{rst}\mu_{ir}\nu_{js}\eta_{kt} \end{aligned} \tag{12}$$

where the terms $\lambda, \lambda_{A(i)}, \lambda_{B(j)}, \lambda_{C(k)}, \lambda_{AB(ij)}, \lambda_{AC(ik)}$, and $\lambda_{BC(jk)}$ are marginal effect terms for the various margins of the table; μ_{ir}, ν_{js} and η_{kt} are scale values for variables A, B and C, respectively, on dimensions r, s, and t, respectively; and ϕ_{rst} is the association parameter measuring the strength of the relationship among dimensions r, s and t. Given the identification constraints for the 3-mode association model (Anderson, 1996), if variable C only has 2 levels (i.e., $k = 1, 2$), then $T = 1$, $R = S$, $\eta_{11} = -\eta_{21} = 1/\sqrt{2}$, and $\phi_{rs1} = 0$ for $r \neq s$, such that equation 12 reduces to

$$\begin{aligned} \ln(F_{ijk}) &= \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{C(k)} + \lambda_{AB(ij)} + \lambda_{AC(ik)} + \\ &\quad \lambda_{BC(jk)} + \sum_{r=1}^{R} \phi_{rr1}\mu_{ir}\nu_{jr} \left(\frac{-1^{(k+1)}}{\sqrt{2}} \right) \end{aligned} \tag{13}$$

This log multiplicative model is equivalent to the logit multiplicative model given in equation 3. This equvialence becomes readily apparent when we express $\ln(F_{ij1}/F_{ij2})$ in terms of the parameters of equation 13:

$$\begin{aligned} \ln(F_{ij1}/F_{ij2}) &= (\lambda_{C(1)} - \lambda_{C(2)}) + (\lambda_{AC(i1)} - \lambda_{AC(i2)}) \\ &\quad (\lambda_{BC(j1)} - \lambda_{BC(j2)}) + \sum_{r=1}^{R} \sqrt{2}\phi_{rr1}\mu_{ir}\nu_{jr} \end{aligned}$$

The correspondence between parameters in equations 3 and 13 is

$$
\begin{aligned}
\tau &= (\lambda_{C(1)} - \lambda_{C(2)}) \\
\tau_{A(i)} &= (\lambda_{AC(i1)} - \lambda_{AC(i2)}) \\
\tau_{B(j)} &= (\lambda_{BC(j1)} - \lambda_{BC(j2)}) \\
\phi_m &= \sqrt{2}\phi_{rr1} \\
\mu_{im} &= \mu_{ir} \\
\nu_{jm} &= \nu_{jr}
\end{aligned}
$$

Since the logit and log multiplicative models are equivalent, a program that fits equation 12 (and thus equation 13) can be used to fit the multidimensional logit multiplicative model.

Anderson (1993, 1996) gives maximum likelihood equations and an algorithm using univariate Netwon-Raphson procedure to fit 3–mode association models, including model 12. The FORTRAN program, *3mode*, implementing this algorithm is available from the author. The major disadvantage of this method is that a global solution is not guaranteed (Anderson, 1996; Haberman, 1995); however, in practice, the method works well. Furthermore, to ensure convergence, the program can be run iteratively with different starting values.

4 Example

As an example, logit multiplicative models are fit to data from a study by Hsieh (1996) on the effects of extra-curricular tutoring programs on mathematics achievement test scores of elementary school children in Taiwan. High achievement test scores are critical for children to gain access to higher educational and thus occupational opportunities. One question in this study are what are potential factors or determinants of whether a student attends an extra-curricular tutoring program. These tutoring programing, known as "cramming schools", purport to increase students' achievement test scores. The data used here, given in Table 1, consist of frequencies of children cross-classified by their grade level (3^{rd}, 4^{th}, 5^{th}, or 6^{th}), the highest education level attained by their father (less than a sixth grade education, graduated from elementary school, graduated from junior high, graduated from high school, graduated from college, or attended graduate or professional school), and whether the student attends a cramming school in mathematics. The goal of this analysis is to determine whether grade level and/or father's education are related to whether a student attends a cramming school. Whether a student attends a cramming school is the criterion variable, and grade level and father's education are the explanatory or predictor variables.

Logit models for various numbers of dimensions were fit to the data in Table 1. The fit statistics for these models are reported in Table 2, where G^2 is the likelihood ratio statistic and X^2 is Pearson's chi-square statistic. The first column of Table 2 indicates how many dimensions were fit. The first model with $M = 0$, which is the additive effects logit model (i.e., the no 3–factor interaction loglinear model), does not fit the data. A 3–way interaction among grade level, father's education, and whether a student attends cramming school exists Rather than having to settle for the saturated logit model (the last model in the table, $M = 3$), we have two intermediate models ($M = 1$ and $M = 2$), both of which appear to fit based on the global fit statistics.

The (standardized) residuals and estimated parameters for both of the one and two dimensional models were examined. The fitted values from the one dimensional model are reported in Table 1. Comparing these to the observed frequencies, there is unusually large residual for fourth graders whose father graduated from junior high. The estimated scale values and association parameters for the two dimensional model reveal that the second dimension essentially accounts for the cell for fourth graders whose father's completed junior high. We selected the one dimensional model as the better of the two models partially on the basis of parsimony and partially on the basis of its interpretation, which is given below.

To describe the interaction between grade level and father's education with respect to whether a student attends cramming school, we examine the estimated scale values of the one dimensional model. The estimates of the model parameters are reported in Table 3 and the estimated scale values are plotted in Figure 1. In estimating these parameters, zero sum constraints were imposed on the main (marginal) effect terms and unit weights (i.e., $h_{G(i)} = h_{F(j)} = 1$) were used for the scale values.

In Figure 1 (or Table 3), we see that the scale values for student's whose father have had some education are nearly equivalent and that the scale values

Table 1: Frequencies (first row) and fitted values (second row) from the logit(1) multiplicative model cross-classified by whether a student attends cramming school, student's grade level, and father's education level.

		Father's Education Level					
Cramming School	Grade Level	None 0	Primary 1	Junior High 2	Senior High 3	College 4	Post-Grad. 5
Yes	3rd	3	1	3	13	35	4
		3.00	0.54	3.94	14.65	33.79	3.08
	4th	1	2	9	19	27	4
		0.99	1.88	4.83	17.75	29.77	6.78
	5th	0	4	12	35	44	10
		0.01	4.59	15.30	34.59	42.41	8.10
	6th	0	19	25	64	51	5
		0.00	18.99	24.93	64.01	51.03	5.04
No	3rd	2	2	8	29	76	4
		2.00	2.46	7.06	27.35	77.21	4.92
	4th	1	8	4	30	67	13
		1.01	8.12	8.17	31.25	64.23	10.22
	5th	3	16	24	48	72	8
		2.99	15.41	20.70	48.41	73.59	9.90
	6th	4	13	22	47	78	7
		4.00	13.01	22.07	46.99	77.97	6.96

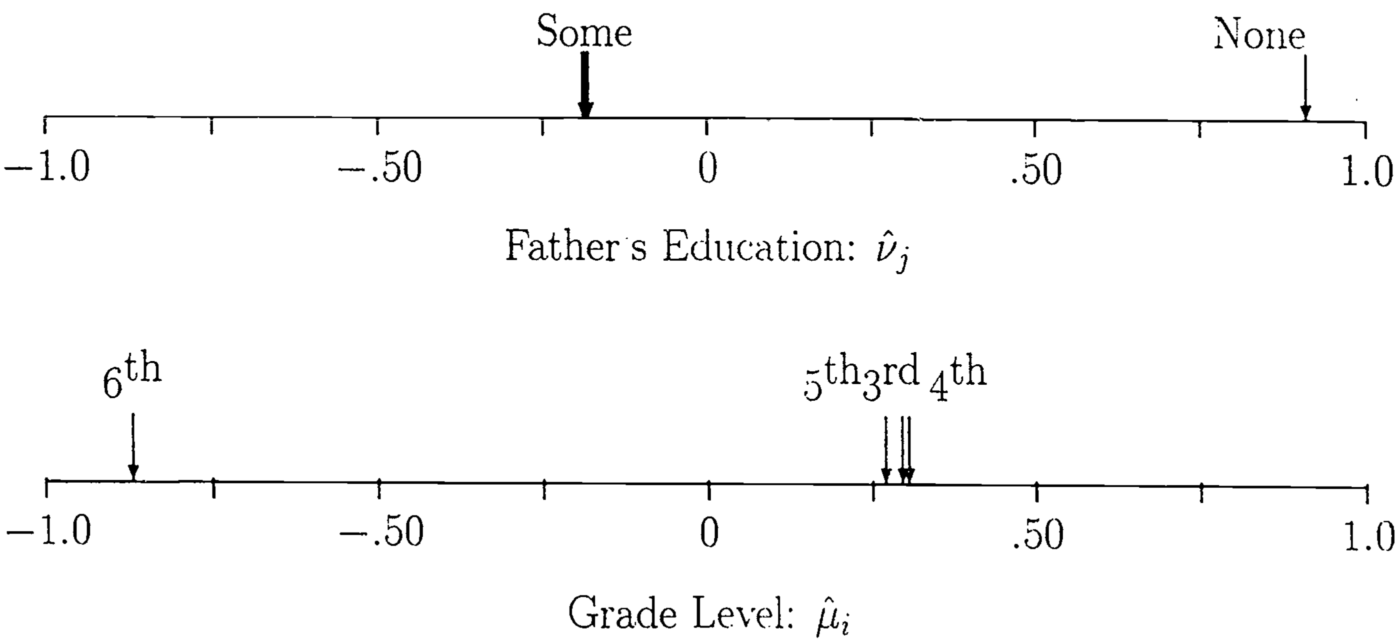

Figure 1: Estimated scales values from the logit(1) multiplicative model fit to the odds of attending cramming school cross-classified by grade level and father's education.

Table 2: Fit Statistics for logit multiplicative models for various numbers of dimensions.

Model (M)	df	G^2	p-value	X^2	p-value
0	15	29.33	.01	27.71	.02
1	8	11.91	.16	12.03	.15
2	3	.75	.86	.87	.93
3	0	.00	1.00	.00	1.00

for student's in the third, fourth and fifth grades are also nearly equivalent. This implies that odds ratios for students in grades 3 through 5 and whose father have had some education are nearly equivalent (i.e., $\Theta_{ii',jj'} \sim 1$). There is a relatively large distance from the scale values for "None" and "Some" and between the scale values for grades 3 through 5 and grade 6. The observed pattern of scale values indicates that the odds that children in the 3rd, 4th and 5th grades attend a cramming school are greater than the odds for children in the 6th grade when their fathers have no education versus when their father has had some education. The odds that children in the 6th grade attend a cramming school are greater than those for children in the younger grades given that their parents had more than an elementary school education. Overall (except for children whose fathers have the lowest level of education, the odds (and in this case, probability) that children attend tutoring program are larger when they're in the sixth grade versus one of the other grades.

We should note that the estimated association parameter $\hat{\phi} = 220.62$ is extremely large. Given that the difference between scale values for "Some" and "None" is approximately equal to 1 and that the difference between the scale values for sixth graders and children in the other grades is a little larger than 1, the odds ratio for sixth graders whose father has had some education versus no education is more than exp(220.62) (a very large numer) times larger than the corresponding odds ratios for childern in the other grades. Alternatively, the odds ratio for children whose father has had some education and who are in the sixth versus one of the other grades is more than exp(220.62) times larger than the corresponding odds ratios for children

Table 3: Parameter Estimates from the logit bilinear model (i.e. $M = 1$).

Variable	Category	Marginal Effect		Scale Value	
—	—	$\hat{\tau} =$	-12.3351	$\hat{\phi} =$	220.6313
Grade Level	3rd	$\hat{\tau}_{G(1)} =$	11.7323	$\hat{\mu}_1 =$	.2984
	4th	$\hat{\tau}_{G(2)} =$	11.7096	$\hat{\mu}_2 =$	.2964
	5th	$\hat{\tau}_{G(3)} =$	10.9070	$\hat{\mu}_3 =$	.2709
	6th	$\hat{\tau}_{G(4)} =$	-34.3490	$\hat{\mu}_4 =$	$-.8658$
Father's Education	None	$\hat{\tau}_{F(1)} =$	-59.1020	$\hat{\nu}_1 =$	.9128
	1	$\hat{\tau}_{F(2)} =$	11.3814	$\hat{\nu}_2 =$	$-.1867$
	2	$\hat{\tau}_{F(3)} =$	12.0125	$\hat{\nu}_3 =$	$-.1821$
	3	$\hat{\tau}_{F(4)} =$	12.0309	$\hat{\nu}_4 =$	$-.1830$
	4	$\hat{\tau}_{F(5)} =$	11.6925	$\hat{\nu}_5 =$	$-.1809$
	5	$\hat{\tau}_{F(6)} =$	11.9845	$\hat{\nu}_6 =$	$-.1799$

whose fathers have had no education.

In this data set, the interaction between father's education and grade level is due to the sixth graders *and* the student's whose father had no formal education. The additive logit model fit to all the data except the cell for the sixth graders whose father had no education does not adequately fit the data (i.e., $G^2 = 24.53$, $df = 14$, $p-$value$= .03$)[1]. However, if we delete the sixth graders and those students whose fathers had no education, we find that the additive logit model with just marginal effects for grade and father's education (i.e., the model without an interaction between grade level and father's education) fits the data quite well ($G^2 = 11.80$, $df = 8$, p-value$= .16$)[2]. Furthermore, the additive logit model fit to data with just children in the sixth grade deleted provides an adequate fit ($G^2 = 16.38$, $df = 10$, p-value$= .09$), and the model fit to the data with just children

[1] Fitting this model makes use of methodology for incomplete tables. The odds for sixth graders whose father had no education was fit perfectly, which uses up one degree of freedom relative to the logit model with $M = 0$

[2] This result was an additional reason for selecting the model with one versus two dimensions

whose father had [illegible] education also fits adequately (G^2 = 20.08, df = 12, p–value= .07). While the nature of the interaction in this example is a rather simple one, it illustrates the power of logit multiplicative models with respect to identifying where the interaction is and its ability to give a parsimonious representation of the data.

5 Discussion

The logit model extension proposed here provides a means not just of testing whether there is a relationship between discretely measured variables, but it provides a metric description and representation of the interactions. While the specific logit model described here was designed for the case of one dichotomous response or criterion variable and two explanatory variables, extensions of this model to polytomous responses and/or more predictor variables in a similar fashion that the $RC(M)$ association model has been extended to higher–way tables is straight forward (see especially Becker & Clogg, 1989; Clogg & Shihadeh, 1994).

The new logit model and related models such as the $RC(M)$ association model and its generalizations are very powerful tools for representing and describing associations in cross–classifications. Such models have been primarily (and successfully) used in sociology (e.g., Clogg, 1982b; Clogg, Eliason & Wahl, 1990; Faust & Wasserman, 1993; Yamaguchi, 1987; Xie, 1992). Clogg (1982b) gives just a sample of the potential applications of these models. Examples of their use in educational research are surprisingly rare, especially given that variables in educational research are often measured discretely (e.g., see first paragraph of this paper). These models can be used in observational studies such as the one described in this paper or in qualitative studies where behaviors are observed and coded according to some defined scheme (e.g., Anderson & Kramer, 1996).

Due to the latent (continuous) variable interpretations of the models (e.g., Bartholomew, 1980, 1987; Goodman, 1981, 1985; Lauritzen and Wermuth, 1989; Whittaker, 1989), the models have potential applications in the area of educational measurement where concern is focused on the measurement of underlying abilities. The resemblance of equation 10 to an item response model is not by coincidence. There are very close relationships between models for categorical data and more commonly (and some not so commonly) known latent variable models (e.g., see Agresti, 1995; Anderson, 1986; Clogg, 1982b).

Logit multiplicative models, as well as the $RC(M)$ association model and its various generalizations, are relatively recent developments in the methodology for categorical data analysis. Given the wide range of potential applications in educational research for such models, we anticipate that researchers will find these models valuable tools.

6 References

Agresti, A. (1990.) *Categorical Data Analysis.* New York: Wiley.

Agresti, A. (1995). Logit models and related quasi-symmetric log-linear models for comparing responses to similar items in a survey. *Sociological Methods & Research, 24*, 68–95.

Anderson, C.J. (1993). *The Analysis of Multivariate, Longitudinal Categorical Data by Log–Multilinear Models.* Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Anderson, C.J. (1996, in press). The analysis of three–way contingency tables by three–mode association models. *Psychometrika.*

Anderson, C.J., and Kramer, L. (1996). Children's friendships during the transition to becoming a sibling: The use of association models to analyze longitudinal categorical data. (manuscript under review).

Anderson, C.J., and Wasserman, S. (1995). Log-multiplicative models for valued social network relations. *Sociological Methods & Research, 24*, 96–127.

Anderson, E.B. (1986). Discussion of paper by L.A. Goodman. *International Statistical Review, 54*, 271–272.

Bartholomew, D.J. (1980). Factor analysis for categorical data. *Journal of the Royal S tistical Society, B, 42*, 292-321.

Bartholomew, D.J. (1987). *Latent Variable Models and Factor Analysis.* New York, NY: Oxford University Press.

Becker, M.P. (1989a). Models for the analysis of association in multivariate contingency tables. *Journal of the American Statistical Association, 84*, 1014–1019.

Becker, M.P. (1989b). Square contingency tables having ordered categories and GLIM. *GLIM Newsletter, 19*, 22–31.

Becker, M.P. and Clogg, C.C. (1989). Analysis of sets of two-way contingency tables using association models. *Journal of the American Statisticial Association, 84*, 142–151.

Choulakian, V. (1988). Exploratory analysis of contingency tables by loglinear formulations and generalizations of correspondence analysis. *Psychometrika, 53*, 235–250.

Clogg, C.C. (1982a). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association, 77*, 803–815.

Clogg, C.C. (1982b). Using association models in sociological research: Some examples. *American Journal of Sociology, 88*, 114–134.

Clogg, C.C. (1986). Statistical modelling versus singular value decomposition. *International Statistical Review, 54*, 284–288.

Clogg, C.C., Eliason, S.R., and Wahl, R.J. (1990). Labor-market experiences and labor-force outcomes. *American Journal of Sociology, 95*, 1537–1576.

Clogg, C.C. and Shihadeh, E.S. (1994). *Statistical Models for Ordinal Variables.* Thousand Oaks, CA: Sage.

Dobson, A.J. (1990). *An Introduction to Generalized Linear Models.* New York: Chapman & Hall.

Francis, B., Green, M, and Payne, C. (1993) *The GLIM System Release 4 Manual.* Oxford: Clarendon.

Gilula, Z., and Haberman, S.J. (1988). The analysis of multiway contingency tables by restricted canonical and restricted association models. *Journal of the American Statistical Association, Theory and Methods, 83*, 760–771.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classificiations having ordered categories. *Journal of the American Statistical Association, 74*, 537–552.

Goodman, L.A. (1981). Association models and the bivariate normal distribution for contingency tables with ordered categories. *Biometrika, 68*, 347–355.

Goodman, L.A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics, 13*, 10–69.

Goodman, L.A. (1986). Some useful extenstions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review, 54*, 243–309.

Goodman, L.A. (1991). Measures, models, and graphical displays in the analysis of cross-classified data. *Journal of the American Statistical Association, 86*, 1085–1111.

Haberman, Z. (1995). Computation of maximum likelihood estimates in association models. *Journal of the American Statistical Association, 90*, 1438–1446.

Lauritzen, S.L., and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics, 17*, 31–57.

McCullagh, P., and Nelder, J.A. (1990). *Generalized Linear Models*, (2nd Edition). London: Chapman and Hall.

SAS Inc. (1994). *SAS Technical Report P-243.* Cary, NC: SAS Inc.

Wickens, T. (1989). *Mulitway Contingency Tables Anlaysis of the Social Sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Whittaker. J. (1989). Discussion of paper by van der Heijden, de Falugerolles & de Leeuw. *Applied Statistics, 38*, 278–279.

Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobibility tables. *American Sociological Review, 57*, 380–395.