AD-A121 969   LOGLINEAR MODELS AND CATEGORICAL DATA ANALYSIS WITH   1/1
              PSYCHOMETRIC AND ECON. (U) CARNEGIE-MELLON UNIV
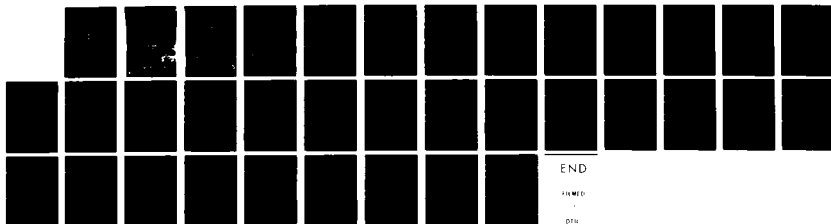   PITTSBURGH PA DEPT OF STATISTICS   S E FIENBERG ET AL.
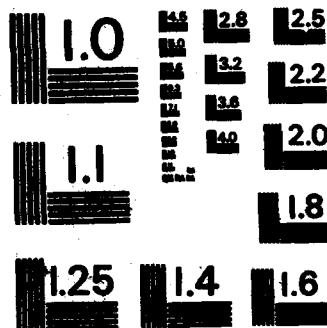UNCLASSIFIED   NOV 82 TR-259 N00014-80-C-0637   F/G 12/1   NL

END
FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD A121969

LOGLINEAR MODELS AND
CATEGORICAL DATA ANALYSIS
WITH PSYCHOMETRIC AND
ECONOMETRIC APPLICATIONS

by

Stephen E. Fienberg

and

Michael M. Meyer*

DEPARTMENT

OF

STATISTICS

DTIC

NOV 3 0

E

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

82 11 29 074

LOGLINEAR MODELS AND
CATEGORICAL DATA ANALYSIS
WITH PSYCHOMETRIC AND
ECONOMETRIC APPLICATIONS

by

Stephen E. Fienberg

and

Michael M. Meyer*

Technical Report No. 259

Department of Statistics

Carnegie-Mellon University

Pittsburgh, PA 15213

August, 1982

Revised November, 1982

*Department of Statistics, University of Wisconsin, Madison, WI 53706

DTIC
ELECTE
NOV 30 1982
E

ABSTRACT

The past decade has seen the publication of a large number of books and papers on the analysis of multi-way contingency tables using loglinear and logit models. The present article presents a summary of the statistical theory that underlies much of this work, and provides some linkage to models and methods of special interest to psychometricians and econometricians. The discussion includes a review of recent and current research on the computation of maximum likelihood estimates for loglinear and logit models, especially for large multi-way contingency tables.

Key Words: Bradley-Terry paired comparisons model; Contingency tables; Logit models; Loglinear models; Rasch model.

Accession For		
NTIS GRA&I	☒	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution/		
Availability Codes		
	Avail and/or	
Dist	Special	
A		

1. INTRODUCTION

When elementary statistics texts write about contingency table analysis they usually discuss only two-way tables, and the Pearson chi-square test for independence. Yet, almost 50 years ago, Bartlett (1935) began statisticians on a path of research on the analysis of three-way and higher-way contingency tables that has branched into many different directions. This paper attempts to present a concise summary of one of these branches, that dealing with the use of loglinear models and their analysis using the method of maximum likelihood. Some basic references for this literature include Andersen (1980), Bishop, Fienberg, and Holland (1975), Birch (1963), Bock (1975, Chapter 8), Fienberg (1980), Goodman (1978), and Haberman (1974, 1978, 1979). Chapter 1 of Fienberg (1980) provides further details on the historical development of this literature.

In Section 2, we summarize the basic theory associated with the loglinear model and its constrained counterpart, the logit model. We consider various sampling structures for contingency tables, and make use of more general results on maximum likelihood estimation for exponential family distributions. This material on loglinear models is linked to the recent econometric literature on retrospective choice-based sampling, and simultaneous logit and loglinear models.

In Section 3, we illustrate the use of the general loglinear model results for two statistical problems, whose literatures have evolved separately. More specifically, we show how these non-contingency-table problems can be given contingency-table-like representations. Then in Section 4 we briefly summarize the recent literature on correspondence analysis, and its links to the loglinear model literature.

Finally, in Section 5 we focus on computational aspects of maximum likelihood estimation for loglinear models, and we outline some current research efforts on computation for very large contingency tables.

2. LOGLINEAR MODELS AND METHODS

For many examples of discrete–response data it is natural to assume a Poisson or multinomial distribution for the observed counts. When such an assumption seems reasonable the methods associated with loglinear models may be appropriate for the data. Subject to the distributional requirements, the full power of the methodology is available for any problem where the response variables are discrete and the explanatory variables are either discrete, continuous, or some combination thereof. Loglinear models are most commonly applied to contingency tables. These arise when all of the explanatory variables are discrete, and the data can be represented as a p–way cross–classification, where p is the total number of response and explanatory variables. When some of the explanatory variables are continuous it is not possible to display the complete data as a cross–classification but loglinear model methods are still appropriate.

The loglinear model is a categorical data tool which closely resembles the regression and analysis of variance (ANOVA) models for continuous data. Regression models define a decomposition of the expected values of the data (actually conditional expectations for responses given the values of certain explanatory variables). Most applications take a *linear* decomposition of the space of all possible expected values. Just as linear regression is the natural and most easily manipulated model (contrasted with, say, nonlinear regression) for continuous data, the natural model·for categorical data is linear in the logarithms of the expected values. This formulation is mathematically simple, and more significantly, the important concepts of independence and conditional independence can be easily expressed in the loglinear formulation (for details, see Bishop, Fienberg, and Holland, 1975).

In order to describe the loglinear model and its justification we need to develop some notation. Let I be a finite index set and consider a vector of observed counts,

$$\mathbf{x} = \{x_i ; i \in I , x_i \in Z \} \tag{2.1}$$

which are considered to be realizations of a set of random variables,

$$\mathbf{X} = \{X_i ; i \in I \}. \tag{2.2}$$

The set I indexes the possible outcomes. The random variable X has expectation

$$m = E(X),$$
(2.3)

and log-expectation

$$\mu = \ln(m) ,$$
(2.4)

where the expectation and logarithm operators are applied element-wise to their arguments. Although the observations must be counts, the only constraint on the means, m, is that they be positive and thus the log-means, μ, can take on any real values. We view μ as an element of the real vector space R^I and we use $\langle \cdot, \cdot \rangle$ to denote the usual inner product on this space.

Thus a 2 × 3 contingency table,

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{21} & X_{23} \end{pmatrix}$$

could be described in terms of the index set

$$I = \{ (i,j) ; i \in \{1,2\}, j \in \{1,2,3\} \} ,$$

or by

$$I^* = \{1,2,3,4,5,6\}$$

if some ordering convention is adopted. Suppose that the cells of the 2×3 table contained continuous or measurement data. Then the standard ANOVA model for the (i,j) cell entry would be:

$$E(X_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$$
(2.5)

with constraints

$$\sum_i u_{1(i)} = \sum_j u_{2(j)} = \sum_i u_{12(ij)} = \sum_j u_{12(ij)} = 0 ,$$
(2.6)

i.e. a row-plus-column-plus-interaction effects model. When the entries $\{X_{ij}\}$ are counts, the corresponding model for the expectations of the counts is

$$\mu_{ij} = \ln(E(X_{ij})) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} .$$
(2.7)

with the same constraints as before, i.e. those in expression (2.6). Setting $u_{12(ij)} = 0$ for all i and j in (2.7) implies that the variables corresponding to rows and columns are independent. In general independence and, in higher dimensions, conditional independence models have a natural formulation as loglinear models. Just as nonlinear models in ANOVA and regression

situations are a possibility, non-loglinear models for counts are possible but the mathematics of estimation is more difficult than for the loglinear models described in this paper.

The statistical literature on frequency data has focussed on two basic sampling schemes and associated probability distributions for the random variables X. The sampling models, the Poisson and product-multinomial paradigms, are intimately related (as product multinomial sampling can be obtained as a conditional version of Poisson sampling), and both result in theories where linear restrictions on μ produce natural models of interest. Extensive accounts of the theory of loglinear models are available in Bishop, Fienberg and Holland (1975), Plackett (1981), and especially Haberman (1974). Extensions to more complex sampling structures can be found in Brier (1980), Fellegi (1980), and Rao and Scott (1981).

2.1. THE POISSON MODEL

Under the Poisson model the elements of X are thought of as independent Poisson random variables X with expectation $m = E(x)$. The probability density for this situation is

$$\Pi_{i \in I} \frac{m_i^{x_i} e^{-m_i}}{x_i!} \ .$$

(2.8)

which results in the following log-likelihood function for $\mu = \ln(m)$ given x:

$$\ell_p(\mu;x) \propto \langle x, \mu \rangle - \langle m, 1 \rangle \ .$$

(2.9)

The most natural realization of this model is when the counts represent the results of simultaneous independent Poisson processes, with means m_i, observed over a fixed period of time.

2.2. PRODUCT-MULTINOMIAL MODEL

Let us partion the index set into r disjoint parts such that

$$I = \bigcup_{k=1}^{r} I_k \ .$$

(2.10)

For each k, $\{x_i ; i \in I_k\}$ has a multinomial distribution with mean $\{m_i ; i \in I_k\}$. In more standard parlance, if n_k is the sample size then $\{x_i ; i \in I_k\}$ has a multinomial distribution with parameters n_k and probability vector

$$p_k = \{ m_i / n_k ; i \in I_k \}. \qquad (2.11)$$

The important distinction is that here the sample size is known a-priori. The full product-multinomial model assumes that the r multinomial distributions so determined are independent with possibly some relationship between the means of the different distributions. Let us denote by v_k the indicator vector of I_k, i.e.

$$v_k = \begin{cases} 1 & i \in I_k \\ 0 & \text{otherwise.} \end{cases} \qquad (2.12)$$

With this notation $n_k = \langle v_k, x \rangle$. The probability density under product multinomial sampling is

$$\Pi^r_{k=1} \binom{n_k}{x} \Pi_{i \in I_k} \left(\frac{m_i}{n_k} \right)^{x_i} , \qquad (2.13)$$

where $\binom{n_k}{x}$ is the usual multinomial coefficient. The log-likelihood for this model is

$$\ell_M(\mu; x, \{n_k\}) = \Sigma^r_{k=1} \{ \ln(n_k!) + \Sigma_{i \in I_k}(x_i \mu_i - x_i \ln(n_k) - \ln(x_i!)) \}$$

$$\propto \langle \mu, x \rangle , \qquad (2.14)$$

subject to the constraints that $\langle m, v_k \rangle = \langle x, v_k \rangle$ for $k = 1, ..., r$. When $r = 1$ we have observations from a single multinomial; otherwise we have a set of r independent multinomials. The crucial point to note is that the kernel of both the Poisson and product-multinomial likelihoods are equivalent.

Before proceeding with the theory consider two examples of frequency date, one which can be represented as a classical contingency table and one which cannot.

2.3. TWO EXAMPLES

We are interested in understanding the factors which affect academic performance of a group of school students. The performance will be assigned to one of three positions; above average, average, and below average.

(i) We select 40 students from each of 4 schools and ascertain the performance level of each student. The resulting data can be presented as a 4 × 3 cross classification with rows representing schools and columns representing perfomance. The row margins are fixed by the

product multinomial sampling design to be 40. If we had chosen students at random until we ran out of money and cross-classified them then, the row margins would be random. In this circumstance the row margins would give some information about the relative sizes of the schools but would not otherwise give any information about the dependence of performance on school. Either way a tabular presentation of the data is possible.

(ii) Consider a similar experiment except that now we take 160 students and record their performance and family income. As income is an essentially continuous variable we cannot present these data as a cross-classification without discarding some information, but we might produce a table listing each person, their performance (the response) and the family income (the explanatory variable).

Of course for this type of study there are many potentially useful explanatory variables and any well-designed survey would record some combination of discrete and continuous predictors. The study could also be extended by recording an extra response variable, e.g. whether the students' performance had improved, remained constant, or declined over the past year. One might then be additionally be interested in the relationship between the two response variables after the effects of the explanatory variables have been taken into account. Many of the questions on relationships among variables can be explored in the context of loglinear models of the sort described here.

Lest we be led to believe that *standard* loglinear model theory is all powerful, let us suppose that we still regard performance as the response variable but that we had first selected 50 people from each performance category, i.e. we stratified on the value of the response variable. A loglinear model may still be appropriate for the resulting probabilities, but, as the sampling scheme fixes the totals for the categories of the response variable, it is neither Poisson nor product-multinomial. Thus the standard methods of frequency data analysis are not necessarily appropriate. For a discussion of this choice-based paradigm see the articles by Manski and McFadden and by Coslett in Manski and McFadden (1981), and further references therein. We will restrict ourselves here to the two sampling schemes we have outlined above.

2.4. THE BASIC THEORETICAL RESULTS

As we have seen, both the Poisson and the product multinomial sampling families have log-likelihoods which can be written in the form

$$\ell(\mu;x) = K(m) + H(x) + \langle \mu, x \rangle . \qquad (2.15)$$

This is the general form for the exponential family of densities (e.g. see Barndorff-Nielsen, 1978). Exponential family theory suggests that the log-expectations, μ, should be the parameter of interest. Even more important, if we choose models which are linear in these natural parameters, then there is a body of theory which is directly applicable to the problem. Fortunately models which consider linear restrictions on μ are not only convenient but are often readily interpreted. For example if we have a purely response contingency table (one in which all the categorizing variables are responses) then the family of loglinear models includes all of the independence and conditional independence relationships between the responses (e.g. see the discussion in Chapters 3 and 4 of Bishop, Fienberg, and Holland, 1975).

Under both Poisson and product-multinomial sampling the likelihood is

$$\ell(\mu;x) \propto \langle x; \mu \rangle , \qquad (2.16)$$

but under product-multinomial sampling there is the additional sampling constraint that

$$P_N(m) = P_N(x) , \qquad (2.17)$$

where

$$N = \text{linear space spanned by } \{v_k\} , \qquad (2.18)$$

and P_N is the orthogonal projection onto N. Under Poisson sampling we can consider any model of the form $\mu \in M$, where M is an arbitrary linear subspace. Under product-multinomial sampling, the sampling constraints dictate that we consider only those models for which $N \subset M$.

Under such a loglinear model, elementary properties of inner products can be invoked to show that

$$\langle x, \mu \rangle = \langle x, P_M \mu \rangle = \langle P_M x, \mu \rangle , \qquad (2.19)$$

and hence that $P_M x$ is a sufficient statistic. In .act the exponential family theory assures us that $P_M x$ is in fact a complete minimal sufficient statistic. In the contingency table context, M is often the space spanned by the row or column indices of a table, in which case $P_M x$ can be obtained from the row or column margins of the table.

Even more can be gained from viewing the loglinear model in the exponential family framework. A standard result states that if the maximum likelihood estimate (MLE) exists it is unique and satisfies (i) the model, $\mu \in M$, and (ii) the likelihood equations

$$P_M \hat{m} = P_M x \quad . \tag{2.20}$$

i.e. the MLE is obtained by equating the minimal sufficient statistics with their expected values. When $N \ M$, expression (2.17) forces expression (2.15) to be satisfied.

The likelihood equations for the two sampling schemes are identical but the interpretation of the resulting parameter estimates differs. In product–multinomial sampling some of the parameter comparisons are meaningless for the very reason that the differences are induced by the sampling design.

In order to utilize the likelihood equations we need to be sure that the MLE exists. A theorem of Haberman (1974) asserts that the MLE exists if and only if there is a $d \in M^\perp$, the orthogonal complement of M, such that $x + d > 0$. Obviously if x has no zero components then $d = 0$ will suffice. When there are observed zeros in the data then the potential for non–existence of the MLE is there. Even when the MLE does not exist the log–likelihood function is still convex. The only problem is that the maximum lies ouside the allowable parameter space, i.e. some components of μ are $-\infty$ at the maximum. This corresponds to an estimated mean value, $\hat{m}_p$ of zero or equivalently to a zero estimated probability. If one is willing to consider this extended interpretation of the MLE, then the MLE always exists and is unique. Unfortunately, it is important to know about such singularities for computational reasons. We return to this problem in Section 5 on computation.

In principle estimation for loglinear models is quite simple – see section 5 on computing for

further details. Without some indication of the variances of the fitted values (or, equivalently, the parameter estimates), however, estimation would be of little value. Again the general theory of exponential families comes to our aid. MLE's in exponential families have an asymptotic variance-covariance matrix equal to the inverse of the Fisher information matrix. It is easy to show that for the Poisson likelihood the first and second derivatives of the loglikelihood are $x - m$, and $-\text{diag}(m_i,...,m_I) = -D_m$, respectively. Using this notation, Haberman (1974, pp.75-78) has shown that as the total sample size, N, tends to infinity

$$L[(N^{\frac{1}{2}}(\hat{\mu}-\mu)] \rightarrow N(0,P_M^i D_m^{-1}) ,\qquad(2.21)$$

where m is the true mean parameter and P_M^i is the orthogonal projection onto M with respect to the inner product $\langle .,D_m.\rangle$. From this it follows that

$$L[(N^{\frac{1}{2}}(\hat{m} - m)] \rightarrow N(0,D_m.P_M^i) .\qquad(2.22)$$

By using these asymptotic normal distributions it is possible to directly show that the usual goodness-of-fit measures,

$$X^2 = \Sigma(x_i - \hat{m}_i)^2 / \hat{m}_i \qquad(2.23)$$

and

$$G^2 = 2 \Sigma x_i \ln(x_i/\hat{m}_i) ,\qquad(2.24)$$

both have asymptoic X^2 distributions with degrees of freedom equal to the number of elements in the index set I minus the dimension of M. These statistics are of course quite well known; the first one, (2.23), is just Pearson's chi-square while the second, (2.24), is the log-likelihood-ratio statistic. Thus we see that under Poisson or product-multinomial sampling the loglinear model arises naturally from consideration of the likelihood and the properties of exponential families.

Once we have decided on a loglinear model there is nothing that dictates that we must use maximum likelihood estimation. An alternative estimation strategy would be to set up the problem in the Kullback-Leibler information theory framework. Using the result of Csiszár (1975) and Kullback (1959) it is possible to show that the natural Kullback-Leibler

representation is in fact the convex dual of the maximum likelihood representation and that the estimated means from either approach are equivalent. For further details and references see Meyer (1982). Alternative estimation strategies which have appeared in the statistical literature include weighted least squares and minimum X^2 methods (e.g. see Grizzle, Starmer, and Koch, 1969).

2.5. LOGLINEAR VERSUS LOGIT MODELS

An important topic that we have overlooked thus far is the relationship of loglinear and logit models to each other. Consider a vector of binomial responses $x = \{x_i; i \in I\}$ and sample sizes $n = \{n_i; i \in I\}$ with probability of success p. The logit model postulates the relationship

$$\text{logit}(p) = \ln(p/(1 - p)) \in M^* , \qquad (2.25)$$

where M^* is a linear space spanned by appropriate explanatory variables. It is possible to show that every logit model can be represented by an equivalent loglinear model for the $I \times 2$ contingency table formed from the number of success and number of failures. Thus all of the results for loglinear models carry over directly to logit models. However it is often much easier to interpret models when they are presented in terms of the log odds or logit scale. Even more important, it is often quite inefficient to use numerical algorithms which are suitable for the general loglinear model in place of the more specific methods appropriate for logit models. The numerical efficiency of logit model methods is largely due to the presence of sampling constraints which are implicit in the logit formulation but which must be considered explicitly in the loglinear model.

Similar comments are in general true for the simultaneous or multinomial logit models which are sometimes used when a multinomial rather than binomial response is observed. Consider a trinomial response model. The data we observe consists of three vectors, x_1, x_2, and x_3 which represent say success, partial success, and failure with corresponding probabilities p_1, p_2, and p_3, and n the number of trials, together with whatever explanatory variables are necessary. The simultaneous logit model (e.g. see Fienberg, 1980) postulates that

$$\ln(p_1/p_2) \in M_1 \qquad (2.26)$$

and

$$\ln(p_2/p_3) \in M_2 \qquad (2.27)$$

are simultaneously true, and that

$$p_1 + p_2 + p_3 = 1. \qquad (2.28)$$

The multinomial logit model (e.g. see Nerlove and Press, 1979, Chamberlain, 1980, or Manski, 1981) is equivalent to the simultaneous logit model. For these models there is always, again, an equivalent loglinear model. As before, it is the ease of interpretation and computational advantages which make the simultaneous logit model useful, but the theoretical results necessary to manipulate these models are direct consequences of the more general loglinear model properties. König, Nerlove, and Oudiz (1981) provide a detailed economic application of the multinomial logit model.

Overall the loglinear model is a very powerful and practical tool for discrete data analysis. However it must be viewed in much the same light as the general linear model. It is a very general tool and one must be careful only to use it when appropriate. Specialized tools and models for certain discrete data problems are often more appropriate just as it is easier to consider the analysis of designed experiments independently of the linear model interpretation.

3. USING LOGLINEAR MODELS FOR SOME "NON-CONTINGENCY" TABLE PROBLEMS

The application of the loglinear model results from Section 2 to multidimensional contingency tables focusses on models where each set of the parameters in the logarithmic scale is associated with one or more dimensions of the table. For example, for a three-dimensional table and the loglinear model representing no second-order interaction (see Fienberg, 1980),

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} , \tag{3.1}$$

the term $u_{23(jk)}$ is associated with dimensions 2 and 3, and so on. Other applications of the results will not necessarily have this feature, as will be apparent in the examples below.

One of the values of general theoretical results is that they are often applicable to specific settings beyond those which led to the formulation of the general structure. This is certainly true for results on the analysis of categorical data problems. Fortunately many of the "non-contingency table" applications of the loglinear model results have contingency-table-like representations so that we can *interpret* the results of our analyses using whatever intuition we have gleaned from the analysis of contingency table data using loglinear models. This section is based on a related discussion in Fienberg (1981), and contains two examples of the use of such contingency-table representations.

3.1. THE BRADLEY-TERRY PAIRED COMPARISONS MODEL

Early in the psychometric literature, Thurstone (1927a,b) proposed a model for binary paired comparisons using linear model structure, and this model was amplified by Mosteller (1951) and Bock and Jones (1969, Chapters 6 and 7). Bradley and Terry (1952) suggested a simple variation on the Thurstone approach which is, in effect, a loglinear or logit model. Extensions of their approach have been developed over the intervening three decades (for an excellent review of this literature see Bradley, 1976).

Suppose t items (e.g., different types of chocolate pudding) or treatments, labeled T_1, T_2,...,T_t, are compared in pairs by sets of judges. (Or suppose that t football teams compete

in pairs in a series of matches.) The Bradley-Terry model postulates that the probability of T_i being preferred to T_j is

$$Pr(T_i \rightarrow T_j) = \frac{\pi_i}{\pi_i + \pi_j} , \qquad i,j = 1,2,...,t,$$

$$i \neq j. \qquad (3.2)$$

where each $\pi_i \geq 0$ and we add the constraint that $\sum_{i=1}^{t} \pi_i = 1$. The model assumes independence of the same pair by different judges and different pairs by the same judge. In the example of the football matches we assume the independence of outcomes of the matches.

TABLE 3-1

Layout for Data in Paired-Comparisons Study with t = 4

		Against			
		T_1	T_2	T_3	T_4
	T_1	—	X_{12}	X_{13}	X_{14}
For	T_2	X_{21}	—	X_{23}	24
	T_3	X_{31}	X_{32}	—	X_{34}
	T_4	X_{41}	X_{42}	X_{43}	—

In the typical paired comparison experiment, T_i is compared with T_j $n_{ij} \geq 0$ times, and we let x_{ij} be the observed number of times T_i is preferred to T_j in these n_{ij} comparisons. Table 3-1 shows the typical layout for the observed data when $t = 4$, with preference (for, against) defining rows and columns. Clearly the binomial constraint,

$$x_{ij} + x_{ji} = n_{ij} ,$$

is of the form of the multinomial constraints described in Section 2, and using the basic loglinear result on the equivalence of MLE's for product-multinomial and Poisson sampling schemes, we can convert expression (3.2) into a model for expected values for a Poisson sampling setting, i.e.

$$\log m_{ij} = \alpha_i + \beta_j + \gamma_{ij} \qquad (3.3)$$

where

$$\gamma_{ij} = \gamma_{ji} , \qquad (3.4)$$

with suitable side constraints. But this, as was noted in Fienberg and Larntz (1976), is simply the model of quasi-symmetry in a square contingency table (see Bishop, Fienberg, and Holland, 1975, Chapter 8). The minimal sufficient statistics are

$$\{x_{i+}\}, \quad \{x_{+j}\}, \quad \{x_{ij} + x_{ji}\} \qquad (3.5)$$

(actually either the row or column totals are redundant), and we can use a trick, suggested in Bishop, Fienberg, and Holland, to transform the problem to one for a three-way table of expected counts. We generate duplicate tables and set

$$m_{ijk} = \begin{cases} m_{ij} & k = 1 , \\ m_{ji} & k = 2 , \end{cases} \qquad (3.6)$$

and, for the observed counts,

$$x_{ijk} = \begin{cases} x_{ij} & k = 1 , \\ x_{ji} & k = 2 . \end{cases} \qquad (3.7)$$

Then the loglinear version of the Bradley-Terry model given by (3.3) and (3.4) becomes the model of no-second-order interaction in the new 3-dimensional table, whose minimal sufficient statistics are ($\{x_{ij+}\}$, $\{x_{i+k}\}$, $\{x_{+jk}\}$). Thus we can analyze the fit of the model and variations on it in a familiar contingency table setting of the sort described in Section 2. The model of expressions (3.3) and (3.4) could be approached directly, but the duplication involved in expressions (3.6) and (3.7) turns what is otherwise a "new loglinear" structure into a recognizable one for multi-dimensional tables, and simplifies the computations of MLE's no matter what iterative method is used for their calculation.

As an example, we reconsider a t = 4 illustration given by Dykstra (1960), in which $n_{34} = 0$ (this does not affect the identification and estimability of the Bradley-Terry parameters). The observed data are given in part (i) of the table and the estimated expected frequencies are given in part (ii) to one decimal place. The goodness-of-fit of the model is summarized by the statistics,

16

TABLE 3-2

Analysis of Dykstra (1960) Example

(a) Observed frequencies

	T_1	T_2	T_3	T_4
T_1	—	28	15	23
T_2	112	—	46	47
T_3	39	17	—	—
T_4	34	11	—	—

(b) Expected frequencies under Bradley-Terry model

	T_1	T_2	T_3	T_4
T_1	-	24.1	17.3	24.5
T_2	115.7	-	43.7	45.5
T_3	36.7	19.3	-	-
T_4	32.5	12.5	-	-

$$X^2 = 2.0 \quad \text{and} \quad G^2 = 2.4 \quad ,$$

and there are 2 degrees of freedom. The actual maximum likelihood estimates of the Bradley-Terry parameters are

$$\hat{\pi}_1 = .1082 \quad \hat{\pi}_2 = .5793 , \quad \hat{\pi}_3 = .2294, \text{ and } \hat{\pi}_4 = .1431$$

as given by Dykstra.

These results on the loglinear representation for the Bradley-Terry model are by now reasonably well-known, and they can be extended to more complex settings involving ties, multiple comparisons, order effects, and rankings (e.g. see Fienberg and Larntz, 1976, Fienberg, 1979, and Duncan and Brody, 1982). Recent results by Meyer (1981) are of special use in giving contingency table representations to some of these generalizations.

3.2. THE RASCH MODEL

We now turn to a problem which begins with a representation as a two-way table of 0's and 1's, and ends up as a relatively standard multi-dimensional contingency table problem. The results of *ability tests* are often structured in the form of sequences of 1's for correct answers and 0's for incorrect answers. For a test with k problems or items administered to n individuals, we let

$$Y_{ij} = \begin{cases} 1 & \text{if individual } i \text{ answers item } j \text{ correctly} \\ 0 & \text{otherwise.} \end{cases} \quad (3.8)$$

Thus we have a two-way table of random variables $\{Y_{ij}\}$ with realizations $\{y_{ij}\}$. An alternative representation of the data is in the form of a $n \times 2^k$ table $\{W_{ij_1 j_2 \cdots j_k}\}$ where the subscript i still indexes individuals and now $j_1, j_2, \ldots, j_k$ refer to the correctness of the responses on items $1, 2, \ldots, k$, respectively, i.e.

$$W_{ij_1 j_2 \cdots j_k} = \begin{cases} 1 & \text{if } i \text{ responds } (j_1, j_2, \ldots, j_k) \\ 0 & \text{otherwise.} \end{cases} \quad (3.9)$$

The simple Rasch model (see Rasch, 1960 as reprinted in 1980; and Andersen, 1980, 1983) for the $\{Y_{ij}\}$ is

$$\log \frac{P(Y_{ij}=1)}{P(Y_{ij}=0)} = \gamma + \mu_i + \nu_j . \tag{3.10}$$

where

$$\Sigma \mu_i = \Sigma \nu_j = 0 . \tag{3.11}$$

Differences of the form $\mu_i - \mu_r$ are typically described as measuring the relative abilities of individuals i and r, while those of the form $\nu_j - \nu_s$ are described as measuring the relative difficulties of items j and s. Expression (3.10) is a *logit* model in the usual contingency table sense for a 3-dimensional array whose first layer is $\{y_{ij}\}$ and whose marginal totals adding across layers is an n×k table of 1's. Because the Rasch model, when viewed as a model for $P(Y_{ij}=1)$, depends on the item parameters in a non-linear way, it is not at all clear whether we can collapse the array $\{w_{i_1 i_2 \cdots i_k}\}$ by adding over subjects for estimation purposes. We return to this matter below.

Maximum likelihood estimation for the parameters of the Rasch model of expression (3.10) has been the focus of several authors, including Rasch and Andersen. Unconditional maximum likelihood (UML) estimates can be derived but they have rather problematic asymptotic properties, e.g. the estimates are inconsistent as $n \to \infty$ and k remains moderate, although they are consistent when both n and $k \to \infty$ (Haberman, 1977).

Before turning to an alternative to the UML approach, we point out a recently-derived result for UML estimates for the Rasch model which links up in yet another way with loglinear structures for contingency tables. In order to derive necessary and sufficient conditions for the existence of UML estimates (a problem not really discussed for any of the data structures in this paper), Fischer (1981) embeds the matrix $y = \{y_{ij}\}$ into a larger (n+k)× (n+k) matrix of the form:

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & e^T - y^T \\ y & 0 \end{bmatrix} \tag{3.12}$$

where e is an nxk matrix of 1's, so that, for all (i,j),

$$a_{ij} + a_{ji} = 1 . \tag{3.13}$$

Then he notes that the Rasch model of expression (3.10) is transformed into an incomplete version of the Bradley-Terry model of expression (3.2) discussed at the beginning of this section, i.e.

$$P(a_{ij}=1) = \frac{\pi_i}{\pi_i + \pi_j} \qquad i = k+1,...,k+n,$$

$$j = 1,2,...,k, \qquad (3.14)$$

and similarly for the other non-zero block of entries in A, where

$$\log \frac{\pi_i}{\pi_r} = \mu_i - \mu_r, \qquad i,r = 1,2,...,n,$$

$$i \neq r, \qquad (3.15)$$

and

$$\log \frac{\pi_j}{\pi_s} = \nu_j - \nu_s, \qquad j,s = 1,2,...k,$$

$$j \neq s . \qquad (3.16)$$

Thus, using a three-dimensional representation for A alluded to at the beginning of this section, we can show that estimation results for the UML approach to the Rasch model correspond to those for the no-second-order interaction model (see expression (3.1)) applied to an incomplete three-dimensional contingency consisting of two zero blocks of dimension $k \times k \times 2$ and $n \times n \times 2$, and a duplicated version of the $n \times k \times 2$ table with layers y, and e $-$ y.

Now, we turn to a conditional approach to likelihood estimation (CML) advocated initially by Rasch, who noted that the conditional distribution of Y given the individual marginal totals $\{Y_{i+} = y_{i+}\}$ depends only on the item parameters, $\{\nu_j\}$. Then each of the row sums $\{y_{i+}\}$ can take only $k+1$ distinct values corresponds to the number of correct responses. Next, we recall the alternate representation of the data in the form of an $n \times 2^k$ array, $\{W_{i_1 i_2 \cdots i_k}\}$, as given by expression (3.9). Adding across individuals we create a 2^k contingency table, X, with entries

$$X_{i_1 i_2 \cdots i_k} = W_{+i_1 i_2 \cdots i_k}. \qquad (3.17)$$

Earlier, we asked the question of whether we could work with this collapsed array. The answer is yes, since all of the information we need to preserve is the response pattern, i.e.

$\{j_1, j_2, \dots, j_k\}$, and the number of "correct" responses that correspond to that pattern. Such information allows us to completely reconstruct the original matrix of responses, Y, except for the labelling of individuals, and thus we can use the 2^k array x to represent the conditional distribution of X given $\{Y_{i+} = y_{i+}\}$.

Duncan (1982) and Tjur (1982) independently noted that we can estimate the item parameters for the Rasch model of expression (3.10) using the 2^k array x, and the loglinear model

$$\log m_{j_1 j_2 \dots j_k} = \omega + \sum_{s=1}^{k} \delta_{j_s} \nu_s + \gamma_{j_+} , \qquad (3.18)$$

where the subscript $j_+ = \sum_{s=1}^{k} j_s, \delta_{j_s} = 1$ if $j_s = 1$ and is 0 otherwise, and

$$\sum_{p=0}^{k} \gamma_p = 0 . \qquad (3.19)$$

More specifically Tjur (1982) shows that maximum likelihood estimation of the 2^k contingency table of expected values, $m = \{m_{j_1 j_2 \dots j_k}\}$, using a Poisson sampling scheme and the loglinear model of expression (3.18) produces the conditional maximum xlikelihood estimates of $\{\nu_j\}$ for the original Rasch model. Tjur proves this equivalence by (1) assuming that the individual parameters are independent identically distributed random variables from some completely unknown distribution, π; (2) integrating the conditional distribution of Y given $\{Y_{i+} = y_{i+}\}$ over the mixing distribution, π; (3) embedding this "random effects" model in an "extended random model"; and (4) noting that the likelihood for the extended model is equivalent to that for expression (3.18) applied to x.

TABLE 3-3

Multiplicative Representation of Expected Values of Model (3.18) for the Case k = 3

Item C

| | | Yes
Item A | | No
Item A | |
		Yes	No	Yes	No
Item B	Yes	$abcS_3$	bcS_2	abS_2	bS_1
	No	acS_2	cS_1	aS_1	S_0

For k=3, the loglinear version of the Rasch model for the 2^3 table, i.e. (3.18), can be represented in multiplicative form for the expected values m as in Table 3-3. The multiplicative parameters a, b, and c in Table 3-3 correspond to the $\{v_a\}$ in (3.18), and the multiplicative parameters $\{S_i\}$ correspond to $\{\gamma_i\}$. The minimal sufficient statistics are

$$\{x_{i++}\}, \quad \{x_{+j+}\}, \quad \{x_{++k}\} \quad , \tag{3.20}$$

and

$$\{x_{111}, \quad x_{110} + x_{101} + x_{011}, \quad x_{100} + x_{010} + x_{001}, \quad x_{000}\} \quad . \tag{3.21}$$

But these are the minimal sufficient statistics of the model of quasi-symmetry preserving one-dimensional marginal totals which was first proposed by Bishop, Fienberg, and Holland (1975, Chapter 8). Indeed, the quasi-symmetry model is equivalent to that of expression (3.18). Thus following the prescription of Bishop, Fienberg, and Holland (1975, p.305), we can re-represent the data in a 4-dimensional redundant form (as a 2×2×2×6 table) and estimate the Rasch model item parameters using a standard loglinear model fitted to a 4-way table (although not the 4-way table w of expression (3.9)). Additional simplifications ensue here because

$$\hat{m}_{111} = x_{111} \quad .$$

$$\hat{m}_{000} = x_{000} \quad . \tag{3.22}$$

Duncan (1982) gives several examples of the application of the Rasch model to survey research problems, and he presents several extensions of the model, indicating how they can be represented in a multi-dimensional table format such as that of Table 3-3.

Plackett (1981), in a very brief section of the 2nd edition of his monograph on categorical data analysis, notes that the Q-statistic of Cochran (1950) can be viewed as a means of testing that the item parameters in the Rasch model are all equal and thus zero, i.e. $v_j = 0$ for all j. This observation is intimately related to the results just described, and our original data representation in the form of an n×k (individual by item) array y is exactly the same representation used by Cochran. By carrying out a conditional test for the equality of marginal proportions given model (3.18) i.e. quasi-symmetry preserving one-dimensional

marginals, we get a test that is essentially equivalent to Cochran's test. But this is also the test for $\{\nu_j = 0\}$ within model (3.18).

4. CORRESPONDENCE ANALYSIS FOR CONTINGENCY TABLE DATA

At the same time as the methods for loglinear and logit models were being developed in the United States and the United Kingdom, an alternative method known as *correspondence analysis* was being developed and practiced by a group of French statisticians (e.g. see Benzécri, 1973). Correspondence analysis can be thought of as a special case of canonical correlation analysis, which is used in general to study linear relationships between two sets of variables. Keller and Wansbeek (1983) set correspondence analysis in the context of an errors-in-variables model. In this section, we briefly outline two different ways that correspondence analysis can be used to analyze contingency table data, and we contrast these approaches with the loglinear model approach which is the main focus of this paper.

The first approach to the use of correspondence analysis is as a technique for displaying the rows and columns of a two-way contingency table as points in corresponding low-dimensional vector spaces (e.g. see Greenacre, 1981, or Heiser and Meulman, 1983). When the dimensionality is one or two, these spaces are often superimposed and used for a pictorial joint display. Let x be an r×c matrix of counts, and A and B corresponding matrices of row and column proportions, respectively. Then correspondence analysis finds a linear mapping between the co-ordinates f of the rows and g of the columns, with respect to the principal axes defined by the following eigen-equations:

$$A^T B^T f = \lambda f$$

$$B^T A^T g = \lambda f . \tag{4.1}$$

The eigenvalues here are usually arranged in descending order

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0 , \tag{4.2}$$

and the choice of m = 2 allows for a two-dimensional graphical display. The method as outlined here is limited to two-way tables, and as such has little to offer for the analysis of multi-way tables unless several variables are merged.

O'Neill (1978) has developed a formal test for independence in a two-way table using a canonical correlation approach and Haberman (1981) has demonstrated its relationship to a

simple extension of the loglinear model suggested in Fienberg (1968) and developed by Andersen (1980) and Goodman (1979). The canonical correlation test extracts the largest eigenvalue, R_1, of the $r \times r$ matrix B with elements

$$b_{ii'} = \Sigma^c_{j=1} \, d_{ij} \, d_{i'j} \qquad (4.3)$$

where

$$d_{ij} = \frac{x_{ij} - x_{i+}x_{+j}/n}{[x_{i+}x_{+j}]^{\frac{1}{2}}} \; . \qquad (4.4)$$

Then NR_1^2 is then referred to as the distribution of the maximum eigenvalue of an $(r-1) \times (r-1)$ central Wishart matrix with $c-1$ degrees of freedom. Thus when $r > 2$ and $c > 2$, the asymptotic distribution of NR_1^2 is *not* the chi-squared distribution. Haberman (1981) has shown that this test is equivalent under the null hypothesis to the test for $\lambda = 0$ in the model

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} + \lambda \, u'_{1(i)} \, u'_{2(j)} \; . \qquad (4.5)$$

where

$$\Sigma u_{1(i)} = \Sigma u_{2(j)} = \Sigma u'_{(i)} = \Sigma u'_{2(j)} = 0 \; . \qquad (4.6)$$

De Leeuw (1983) describes the interaction structure in expression (4.5) as bilinear. For a discussion of extensions of the model of expressions (4.5) and (4.6) to multi-way tables and the use of such models for ordered categorical data, see Fienberg (1982).

A second approach to the use of correspondence analysis (under its psychometric name, *dual scaling*) for multi-way tables is presented by Nishisato (1980) who describes the loglinear model approach to multi-way tables as a form of analysis of variance on the log-expected values. (We note parenthetically that such an ANOVA description ignores the interpretation of many loglinear models in terms of conditional independence of variables and related ideas, as well as the interpretation of individual parameters in terms of odds ratios.) He then proposes a pair of dual scaling or reciprocal approaches in which the unit of analysis is not a cell frequency but a single response from a single subject, and the data are arrayed in a two-dimensional response pattern matrix. Method I first derives an optimal response-score vector y_r and then subjects it to a standard ANOVA. Method II essentially attempts to carry out

the two parts of Method I simultaneously (see the related discussion in de Leeuw et al., 1976, and Young et al., 1976. Deville and Saporta (1983) present an alternative approach to Nishisato's which they label as *multiple correspondence analysis.*

Most of the examples of these optimal scaling or correspondence analysis methods applied to multiway contingency table data that appear in the literature are not especially revealing. For example, Nishisato presents an example with four binary explanatory variables and 3 observations for each of the $2^4 = 16$ experimental conditions, and where there are 8 response variables each with 5 categories. Thus, the 3 observations are spread over a 5^8 response structure and make the estimation of a standard loglinear model virtually impossible. The heroic and untestable assumptions implicit in Nishisato's dual scaling analysis need to be made explicit before a direct comparison with loglinear models is possible. On the other hand, Kester and Schriever (1982) analyze a standard $2 \times 3 \times 3 \times 4$ contingency table example given in Fienberg (1980, p.91). A reasonable loglinear model for the data in that table has as its minimal sufficient statistics only two-way marginal totals. Kester and Schriever's analysis of this example relies only on these bivariate two-way totals, and introduces scalings for the three non-binary dimensions. In a sense their analysis can be thought of as supplementing the standard loglinear analysis.

5. COMPUTATIONAL METHODS FOR LOGLINEAR MODELS

There are numerous compuational methods which can be used to fit general loglinear models. For any particular application the investigator must be careful to select a method which is not only feasible but efficient. The loglikelihood function for the loglinear Poisson model is a convex function and when the MLE exist (i.e. there are no fitted zeros) it is a strictly convex function. As the first and second derivatives of the loglikelihood are analytically available, a natural contender for the maximization procedure is Newton's method.

Newton's method is indeed useful for loglinear models where the dimension of the parameter space is small. Using Newton's method has the advantage that the calculations necessary to obtain asymptotic covariances for the fitted values or parameter estimates (i.e. the Fisher information matrix or some decomposition of it) are a by product of the algorithm. A further advantage is that in general Newton's method can be expected to have a quadratic rate of convergence. A widely available implementation of Newton's method for discrete data is in the statistical package GLIM (Baker and Nelder, 1978), while a version that is useful only for logit models can be found in the BMDP statistical system. There are however several disadvantages to using this algorithm. First, fitted zeros can cause Newton's method to converge much more slowly than one would otherwise expect. Thus it is imperative that potential singularities in the data be detected before formal fitting is attempted. Fienberg, Meyer and Stewart (1982) use a result of Stewart (1980) to first screen the data for singularity problems before fitting loglinear or logit models using a version of Newton's algorithm which is efficient in terms of both time and storage. Second, a more limiting problem with Newton's method is that for large models (greater than about 200 parameters) the amount of computer memory required to store the second derivative matrix exceeds the capacity of most computers. Under these circumstances a possible alternative is the family of conjugate gradient algorithms. McIntosh (1982) has investigated the possibility of using conjugate-gradient methods for many statistical applications. Conjugate gradient algorithms have the property that they do not use any second derivative information for the function being maximized. This means that asymptotic covariances are not available (other than in some special closed-form situations), but

the method can consider very large problems. Whereas Newton's method spends a lot of effort finding a good search direction, conjugate-gradient methods merely find a reasonable direction. Thus each iteration of the algorithm requires relatively little computation but the algorithm is likely to require more iterations than Newton's method.

Both Newton's method and the conjugate-gradient methods are important general optimization tools applicable to a wide variety of problems. A third optimization method is the cyclic-coordinate ascent algorithm. This algorithm is not widely used for general problems but is of special importance in loglinear model applications. In this context it is known as the *Iterative Proportional Fitting Procedure* (IPFP). The distinguishing feature of IPFP is that it attempts to maximize the likelihood by searching along a series of fixed directions, and thus no computational effort is expended in searching for good directions. What makes IPFP attractive for loglinear model applications is that for many models there is a closed-form answer for maximization along the fixed directions. Specifically, let $\hat{m}$ be the current vector of fitted values. If we are attempting to maximize the likelihood along a vector, β, which consists of only zeros and ones, then

$$\hat{m}_{new} = \hat{m} \cdot \langle x, \beta \rangle / \langle \hat{m}, \beta \rangle , \qquad (5.1)$$

i.e. the new vector of fitted values is just a simple proportional adjustment of the old vector. Thus we can use this method for any model M, which can be spanned by a set of vectors $\{\beta_k; k=1,...,s\}$ where each β_k is a vector of only zeros and ones. For each direction, β_k, there is a step in the iteration of the form (5.1), and thus a complete cycle of the iteration consists of a set of s steps along the s fixed directions $\{\beta_k\}$ corresponding to the model M. The class of models which satisfy these requirements is surprisingly large and includes all the ANOVA-like loglinear models for contingency tables. This class itself includes all the independence and conditional independence models.

The IPFP generally takes many iterations to converge, but as each iteration is so simple that this is of little concern. A special feature of the IPFP is that when a closed-form MLE exists the algorithm will converge in at most two iterations and one can assure convergence in one

iteration by choosing an appropriate order for selection of the spanning vectors (e.g., see Haberman, 1974). The main advantage of the IPFP is that it needs to store only the table of fitted values and the sufficient statistics in computer memory and thus can be used on very large problems. The BMDP package uses this method in its contingency table program, BMDP3F.

When one has a large contingency table problem which is not amenable to the IPFP, there are several alternatives. An algorithm which is based on the IPFP but which is useful for any model is the generalized iterative scaling method of Darroch and Ratcliff (1972). Unfortunately this algorithm has the slow convergence disadvantage of the IPFP, without the important advantage of simple computations at each iteration. For special applications there are several alternatives. Under some circumstances it is possible to transform a contingency table and associated models into a form where the simple IPFP can be used. An example of this is given in Meyer (1982). For the simultaneous logit problem it is possible to combine the IPFP with some other algorithm, e.g. Newton's method to gain a useful hybrid. Rather than consider the problem as a large contingency table one can fit each of the individual logit models which comprise the system and combine the estimate using ideas based on the IPFP. It is then necessary to repeat this process until the estimates converge. Some details on this approach are available in Meyer (1981).

For general applications of the loglinear model to small problems, Newton's method is probably the most efficient and useful algorithm. As the size of the problem increases the IPFP, when it can be applied, is the most useful algorithm.

REFERENCES

Andersen, E.B. (1980). *Discrete Statistical Models with Social Science Applications.* Amsterdam: North Holland.

Andersen, E.B. (1983). "Latent trait models." *Annals of Applied Econometrics* (this issue).

Baker, R.J. and Nelder, J.A. (1978). *The GLIM System, Release 3, Manual.* Oxford, England: Numerical Algorithms Group.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* New York: Wiley.

Bartlett, M.S. (1935). "Contingency table interactions." *J. Roy. Statist. Soc. Suppl.* 2, 248-252.

Benzécri, J. P. (1973). *L'Analyse des Correspondances* (Volume 2 of *L'Analyse des Donnees.* Paris: Dunod.

Birch, M.W. (1963). "Maximum likelihood in three-way contingency tables." *J. Roy. Statist. Soc. B,* 25, 220-233.

Bishop, Y.M.M. (1959). "Full contingency tables, logits, and split contingency tables." *Biometrics,* 25, 383-400.

Bishop, Yvonne, M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass.: MIT Press.

Bock, R.D. (1975). *Multivariate Statistical Methods in Behavioral Research.* New York: McGraw-Hill.

Bock, R.D. and Jones, L.V. (1968). *The Measurement and Prediction of Judgement and Choice.* San Francisco: Holden-Day.

Bradley, R.A. (1976). "Science, statistics, and paired comparisons." *Biometrics* 32, 213-232.

Bradley, R.A. and Terry, M.E. (1952). "The rank analysis of incomplete block designs. I. The method of paired comparisons." *Biometrika* 39, 324-345.

Brier, S.S. (1980). "Analysis of contingency tables under cluster sampling." *Biometrika* 67, 591-596.

Chamberlain, G. (1980). "Analysis of covariance with qualitative data." *Rev. Econ. Studies* 47, 225-238.

Cochran, W.G. (1950). "The comparison of percentages in matched samples." *Biometrika* 37, 256-66.

Csiszár, I. (1975). "I-divergence geometry of probability distributions and minimization problems." *The Annals of Probability,* 3, 146-158.

Darroch, J.N. and Ratcliff, D. (1972). "Generalized iterative scaling for loglinear models." *Ann. Math. Statist.* 43, 1470–1480.

de Leeuw, J., Young, F.W., and Takane, Y. (1976). "Additive structure in qualitative data: an alternating least squares method with optimal scaling features." *Psychometrika* 41, 471–503.

de Leeuw, J. (1983). "Models and methods for the analysis of correlation coefficients." *Annals of Applied Econometrics* (this issue).

Deville, J.C. and Saporta, G. (1983). "Correspondence analysis, application to nominal time series." *Annals of Applied Econometrics* (this issue).

Duncan, O.D. (1982). "Rasch measurement in survey research: further examples and discussion." To appear, in C.F. Turner and E. Martin (eds.) *Survey Measurement of Subjective Phenomena*, Vol. 2, (forthcoming).

Duncan, O.D. and Brody, C. (1982). "Analysing n rankings of three items." In *Social Structure and Behavior: Essays in Honor of William Hamilton Sewell*. New York: Academic Press, 269–310.

Dykstra, O. (1960). "Rank analysis of incomplete block designs: A method of paired comparisons employing unequal repetitions on pairs." *Biometrics* 16, 176–188.

Fellegi, I.P. (1980). "Approximate tests of independence and goodness-of-fit based on stratified multistage samples." *J. Amer. Statist. Assoc.* 75, 261–268.

Fienberg, S.E. (1968). *The Estimation of Cell Probabilities in Two-way Contingency Tables*. Ph.D. Thesis, Department of Statistics, Harvard University.

Fienberg, S.E. (1979). "Loglinear representation for paired comparison models with ties and within-pair order effects." *Biometrics* 35, 479–481.

Fienberg, S.E. (1980). *The Analysis of Cross-Classified Categorical Data* (2nd ed.). Cambridge, Mass.: The MIT Press.

Fienberg, S.E. (1981). "Recent advances in theory and methods for the analysis of categorical data: making the link to statistical practice." *Bull. of Int. Stat. Inst.*, 43rd Session, in press.

Fienberg, S.E. (1982). "Using information on ordering for loglinear model analysis of multidimensional contingency tables." *Proceedings of the XIth International Biometrics Conference*, Toulouse, France, 133–139.

Fienberg, S. E. and Larntz, K. (1976). "Loglinear representation for paired and multiple comparisons models." *Biometrika*, 63, 245–254.

Fienberg, S.E., Meyer, M.M., and Stewart, G.W. (1982). "The numerical analysis of contingency table data." Unpublished manuscript.

Fischer, G.H. (1981). "On the existence and uniqueness of maximum-likelihood estimates in the

Rasch model." *Psychometrika*, 46, 59-78.

Goodman, L.A. (1978). *Analyzing Qualitative/Categorical Data: Log Linear Models and Latent Structure Analysis.* Cambridge, Mass.: Abt.

Goodman, L.A. (1979). "Simple models for the analysis of association in cross-classifications having ordered categories." *J. Amer. Statist. Assoc.* 74, 537-552.

Greenacre, M. (1981). "Practical correspondence analysis." In *Interpreting Multivariate Data* (V. Barnett, ed.), Chichester: Wiley, 119-146.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). "Analysis of categorical data by linear models." *Biometrics*, 25, 489-504.

Haberman, S.J. (1974). *The Analysis of Frequency Data.* Chicago: The University of Chicago Press.

Haberman, S.J. (1977). "Maximum likelihood estimates in exponential response models." *Ann. Statist.* 5, 815-841.

Haberman, S.J. (1978). *Analysis of Qualitative Data, Volume 1 (Introductory Topics).* New York: Academic Press.

Haberman, S.J. (1979). *Analysis of Qualitative Data, Volume 2 (New Developments).* New York: Academic Press.

Haberman, S.J. (1981). "Tests for independence in two-way contingency tables based on canonical correlation and on linear by linear interaction." *Ann. Statistics* 9, 1178-1186.

Heiser, W.J. and Meulman, J. (1983). "Analyzing rectangular tables by joint and constrained multidimensional scaling." *Annals of Applied Econometrics* (this issue).

Keller, W.J. and Wansbeek, J. (1983). "Multivariate methods for quantitative and qualitative data." *Annals of Applied Econometrics* (this issue).

Kester, N.K. and Schriever, B.F. (1982). "Analysis of association of categorical variables by numerical scores and graphical representation." Preprint. Stiching Mathematisch Centrum, Amsterdam.

König, H., Nerlove, M., and Oudiz, G. (1981). "Micro-analysis of realizations, plans, and expectations in the IFO and INSEE business texts by multivariate log-linear probability models." In E.G. Charatsis (ed.) *Proceedings of the Econometric Society European Meeting, 1979.* Amsterdam: North-Holland, 393-420.

Kullback, S. (1959). *Information Theory and Statistics.* New York: Wiley.

Manski, C.F. (1981). "Structural models for discrete data: the analysis of discrete choice." *Sociological Methodology 1981*, 58-109.

Manski, C.F. and McFadden, D. (eds.) (1981). *Structural Analysis of Discrete Data: With Econometric Applications.* Cambridge, Mass.: The MIT Press.

McIntosh, A. (1982). *Fitting Linear Models: An Application of Conjugate Gradient Algorithms. Lecture Notes in Statistics*, Vol. 10, New York: Springer-Verlag.

Meyer, M.M. (1981). *Applications and Generalizations of the Iterative Proportional Fitting Procedure.* Ph.D. Thesis, School of Statistics, University of Minnesota.

Meyer, M.M. (1982). "Transforming contingency tables." *Ann. Statist.* 10, (in press).

Mosteller, F. (1951). "Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations." *Psychometrika* 16, 3-9.

Nerlove, M. and Press, S.J. (1979). "Multivariate log-linear probability models for the analysis of qualitative data." *Bull. of Int. Statist. Inst.*, 42nd Session.

Nishisato, S. (1980). *Analysis of Categorical Data: Dual Scaling and Its Applications.* Toronto: University of Toronto Press.

O'Neill, M.E. (1978). "Distribution expansions for canonical correlations from contingency tables." *J. Roy. Statist. Soc. B* 40, 303-312.

Plackett, R.L. (1981). *The Analysis of Categorical Data.* 2nd edition, London: Griffin.

Rao, J.N.K., and Scott, A.J. (1981). "The analysis of categorical data from complex surveys: chi-squared tests for goodness of fit and independence in two-way tables." *J. Amer. Statist. Assoc.*, 76, 221-230.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests.* (Reprinting of the 1960 book.) Chicago: Univ. of Chicago Press.

Stewart, G.W. (1980). "A new method for solving linear inequalities." Tech. Report 970, Dept. of Computer Science, University of Maryland.

Thurstone, L.L. (1927a). "A law of comparative judgement." *Psych. Rev.* 34, 278-286.

Thurstone, L.L. (1927b). "Psychophysical analysis." *Amer. J. Psych.* 38, 368-389.

Tjur, T. (1982). "A connection between Rasch's item analysis model and a multiplicative Poisson model." *Scan. J. Statistics* 9, 23-30.

Young, F.W., de Leeuw, J., and Takane, Y. (1976). "Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features." *Psychometrika* 40, 505-529.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 259	2. GOVT ACCESSION NO. AO-A121962	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Loglinear Models and Categorical Data Analysis with Psychometric and Econometric Applications		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stephen E. Fienberg Michael M. Meyer		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0637
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Carnegie-Mellon University Pittsburgh, PA 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Contracts Office Carnegie-Mellon University Pittsburgh, PA 15213		12. REPORT DATE November, 1982
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bradley-Terry paired comparisons model; Contingency tables; Logit models; Loglinear models; Rasch model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The past decade has seen the publication of a large number of books and papers on the analysis of multi-way contingency tables using loglinear and logit models. The present article presents a summary of the statistical theory that underlies much of this work, and provides some linkage to models and methods of special interest to psychometricians and econometricians. The discussion includes a review of recent and current research on the computation of maximum likelihood estimates for loglinear and logit models, especially for large multi-way contingency tables.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-5601