SHELBY J. HABERMAN
DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS

Analysis of Qualitative Data

Volume 2 *NEW DEVELOPMENTS*

1979

ACADEMIC PRESS
A Subsidiary of Harcourt Brace Jovanovich, Publishers

New York London Toronto Sydney San Francisco

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Haberman, Shelby J
Analysis of qualitative data.

Includes bibliographical references and indexes.
CONTENTS: v. 1. Introductory topics.--v. 2.
New developments.
1. Log-linear models. 2. Contingency tables.
3. Social sciences--Statistical methods.
I. Title.
HA33.H27 519.5 77-25731
ISBN 0-12-312502-2 (v. 2)

PRINTED IN THE UNITED STATES OF AMERICA

79 80 81 82 9 8 7 6 5 4 3 2 1

Preface

In Volume 2 a variety of models for qualitative data are explored; these go beyond the hierarchical log–linear models and logit models of Volume 1.

Chapter 6 discusses multinomial response models appropriate for the complete factorial tables considered in Volume 1. These models are generalizations of the hierarchical log–linear models of Chapters 2–4 and of the logit models of Chapter 5. As in the case of logit models, the models of Chapter 6 can be used to exploit ordered categories and can be used with continuous predicting variables.

Chapter 7 examines log–linear models for incomplete factorial tables. The chapter emphasizes for incomplete two-way tables quasi-independence models which have been the subject of much early work on log–linear models. Hierarchical log–linear models for incomplete multi-way tables and multinomial response models for incomplete tables are also studied. Both iterative proportional fitting and Newton–Raphson algorithms are developed.

In Chapter 8 models are considered for contingency tables in which several variables have the same categories. Symmetry models, quasi-symmetry models, marginal-homogeneity models, and distance models are introduced and related to quasi-independence models and to parametrizations developed for hierarchical log–linear models. Again, both iterative proportional fitting and the Newton–Raphson algorithm are used for numerical work.

In Chapter 9 adjustment of data is studied through numerical methods developed in earlier chapters for use with log–linear models. The methods of Deming and Stephan for adjustment of marginal totals are related to the iterative proportional fitting algorithm for hierarchical log–linear models. Alternative methods of adjustment of data are also studied. In contrast with most earlier treatments of adjustment of data, emphasis is given to estimation of standard deviations of estimates.

Chapter 10 develops a general theory of latent-class analysis in terms of log–linear models. Although latent-class analysis had been considered as early as 1950 by Guttman (1950) and Lazarsfeld (1950a,b), use of latent-class models has become much easier in recent years due to Goodman's (1974a,b) development of an iterative proportional fitting algorithm for a general collection of latent-class models. This chapter discusses Goodman's iterative proportional fitting algorithm and a scoring algorithm similar in structure to the Newton–Raphson algorithm for log–linear models.

The Appendix provides computer programs for log–linear models and latent-class models.

This volume may serve as a sequel to Volume 1 in a two-quarter or two-semester course. The two volumes together provide a thorough introduction to log–linear models and to latent-class analysis.

Work on the book has benefited from partial support from National Science Foundation Grants Nos. SOC72-05228 A04 and MCS72-04364 A04 and National Institutes of Health Grant No. G722648. Comments by Professors Leo Goodman, William Mason, and Clifford Clogg have contributed to the preparation of Volume 2, just as they aided in the preparation of Volume 1. The National Opinion Research Center of the University of Chicago has been a source of much of the data used in this volume. Thanks are also due Dr. L. Srole for permission to use Table 6.1.

Contents of Volume 1

6 Multinomial Response Models

Multinomial response models, or multinomial logit models, are generalizations of logit models in which one or more independent variables are used to predict one or more polytomous dependent variables. Multinomial response models have a much more limited literature than do logit models. The literature does not appear to go back prior to Mantel (1966). Bock (1970, 1975, pp. 520–538), Bock and Yates (1973), Nerlove and Press (1973), Press (1972, pp. 268–272), Theil (1969, 1970), and Haberman (1974a, pp. 351–373) are among those who have discussed these models.

As in the case of logit models, a special version of the Newton–Raphson algorithm can be employed with multinomial response models to produce simpler computations than those required in the Newton–Raphson algorithm of Chapters 1–3. This version of the Newton–Raphson algorithm also simplifies computation of estimated asymptotic variances of the parameters in the multinomial response model. Gains achieved are not without cost. If a parameter appears in a log-linear model formulation but not in an equivalent multinomial response formulation, then the parameter estimate and its EASD are not obtained through the version of the Newton–Raphson algorithm in this chapter.

The Newton–Raphson algorithm for multinomial response models forms the basis for the MULTIQUAL program of Bock and Yates (1973) and for the FREQ program in the appendix. The algorithm is algebraically equivalent to the algorithm for logit models when there is a simple dichotomous dependent variable. The algorithm reduces to the algorithm of Chapters 1–3 in the trivial case in which it assumed that there is a single independent variable with one category.

Since any log-linear model can be regarded as a multinomial response model, the family of multinomial response models is very large. Numerous cases deserve special attention. The possibilities can be distinguished in terms of the nature of the dependent variable or variables:

(D1) There is a single dichotomous dependent variable.

(D2) There is a single polytomous dependent variable in which scores are not assigned to categories.

(D3) There is a single polytomous dependent variable, and scores are assigned to categories.

(D4) There are several dichotomous or polytomous dependent variables, and scores are not assigned to categories.

(D5) There are several dichotomous or polytomous dependent variables, and scores are assigned to categories.

Several possibilities can be distinguished in terms of the independent variables. The first distinction involves the number of independent variables:

(M1) One independent variable is present.

(M2) More than one independent variable is present.

The second distinction involves the nature of the independent variables.

(I1) The independent variables are dichotomous or polytomous, and scores are not assigned to categories.

(I2) All independent variables are dichotomous or polytomous, at least one independent variable is polytomous, and scores are assigned to categories of the polytomous variables.

(I3) At least one independent variable is continuous.

If the dependent variable is dichotomous, as in (D1), then the multinomial response model is a logit model. Since logit models have been studied in Chapter 5, no special attention need be given to this case. However, it should be emphasized that the methods of this chapter do apply to logit models.

Even excluding logit models, 24 combinations of conditions remain. The simplest four cases involve a single polytomous dependent variable and a single dichotomous or polytomous independent variable. These four cases are examined in Section 6.1. One of these cases, the combination of (D2), (M1), and (I1), involves the type of models for two-way tables examined in Chapter 2. The remaining cases provide a basic introduction to the multinomial response version of the Newton–Raphson algorithm.

Section 6.2 considers the 12 cases in which at least three variables are involved but no independent variable is continuous. The four cases in which dependent variables satisfy (D2) or (D4) and independent variables satisfy (I1) reduce to the hierarchical models of Chapters 3 and 4. The new Newton-

Raphson algorithm introduced in this chapter is useful for computations even for the models in which iterative proportional fitting is available, especially if estimates of asymptotic variances are needed and closed-form maximum likelihood estimates do not exist. In other cases in this section, useful competitors for the Newton–Raphson algorithm are not available. This section considers a huge family of models. Just as in Chapters 3 and 4, model selection is a major problem. Methods used here will be similar to those of earlier chapters. Thus standardized values, adjusted residuals, and partitions of the likelihood-ratio chi-square receive emphasis.

Section 6.3 considers the remaining eight cases in which at least one independent variable is continuous. As in the case of logit models, problems arise in use of residual analysis and chi-square statistics, although maximum likelihood estimates retain ordinary large-sample properties. Added problems of computational cost also appear present.

6.1 Multinomial Response Models for Two-Way Tables

Multinomial response models for two-way tables may be used to illustrate many of the general principles of multinomial response models. In the case of two-way tables, interpretation in terms of simultaneous logit models is attractive. In some respects, multinomial response models are related to logit models just as multivariate regression models are related to ordinary univariate regression models.

The relationship of logit models to multinomial response models also extends to the Newton–Raphson algorithm of this section. This algorithm is a generalization of the Newton–Raphson algorithm for logit analysis, although the equations used to define the new algorithm are somewhat different from those in Chapter 5. Despite differences in appearance, it is still possible to interpret the algorithm in terms of a series of weighted regression analyses.

To define a multinomial response model for an $r \times s$ contingency table, let polytomous variables A_h and B_h, $1 \le h \le N$, be observed such that each A_h can take values from 1 to r and each B_h can take values from 1 to s. Assume that given the B_h, the A_h are independently distributed with probability $p_{i \cdot j}^{A \cdot B} > 0$ that $A_h = i$ given that $B_h = j$. Let n_{ij} be the number of observations h with $A_h = i$ and $B_h = j$. Then given that n_j^B observations have $B_h = j$, n_{ij} has expected value $m_{ij} = n_j^B p_{i \cdot j}^{A \cdot B}$, and each column n_{ij}, $1 \le i \le r$, has a multinomial distribution with sample size n_j^B and probabilities $p_{i \cdot j}^{A \cdot B}$, $1 \le i \le r$.

A multinomial response model may be defined in terms of a series of related logit models. For each value j of B_h, let q independent variables t_{jk}, $1 \le k \le q$, be given. The basic requirement in a multinomial response model is that for categories i and i' of A_h, the logit $\tau_{ii'\cdot j}^{A\cdot B} = \log(p_{i\cdot j}^{A\cdot B}/p_{i'\cdot j}^{A\cdot B}) = \log(m_{ij}/m_{i'j})$ be a linear function of the independent variables. Thus for some unknown $\eta_{ii'}$ and $\xi_{ii'k}$, $1 \le k \le q$,

$$\tau_{ii'\cdot j}^{A\cdot B} = \eta_{ii'} + \sum_{k=1}^{q} \xi_{ii'k} t_{jk}. \tag{6.1}$$

(If $q = 0$, then $\tau_{ii'\cdot j}^{A\cdot B} = \eta_{ii'}$.) Since for $1 \le j \le s$, $1 \le i \le r$, $1 \le i' \le r$, and $1 \le i'' \le r$,

$$\tau_{ii\cdot j}^{A\cdot B} = 0 \tag{6.2}$$

and

$$\tau_{ii''\cdot j}^{A\cdot B} = \tau_{ii'\cdot j}^{A\cdot B} + \tau_{i'i''\cdot j}^{A\cdot B}, \tag{6.3}$$

one may assume without loss of generality that

$$\eta_{ii} = 0, \tag{6.4}$$

$$\eta_{ii''} = \eta_{ii'} + \eta_{i'i''}, \tag{6.5}$$

$$\xi_{iik} = 0, \tag{6.6}$$

and

$$\xi_{ii''k} = \xi_{ii'k} + \xi_{i'i''k}, \qquad 1 \le k \le q. \tag{6.7}$$

Other linear constraints may also be imposed on the $\eta_{ii'}$ and $\xi_{ii'k}$. Several examples will appear in this section.

Special Cases

Several special multinomial response models have already been considered. If $r = 2$, the logit model of Chapter 5 is equivalent to the multinomial response model of (6.1), because (5.2) and (6.1) are equivalent if $\omega_j = \tau_{12\cdot j}^{A\cdot B}$, $\eta = \eta_{12}$, and $\beta_k = \xi_{12k}$, $1 \le k \le q$.

If $q = 0$, then the column homogeneity model and the multinomial response model are equivalent. Verification of the claim is straightforward. If the column homogeneity model holds, then (6.1) holds when $q = 0$ and $\eta_{ii'} = \tau_{ii'\cdot 1}^{A\cdot B}$. On the other hand, if $q = 0$, then (6.1) implies that the cross-product ratio

$$\tau_{(1i')(1j')}^{AB} = \tau_{1i'\cdot 1}^{A\cdot B} - \tau_{1i'\cdot j'}^{A\cdot B} = \eta_{1i'} - \eta_{1i'} = 0, \qquad 2 \le i' \le r, \quad 2 \le j' \le s.$$

As shown in Section 2.6, this condition on the cross-product ratio implies the additive log–linear model.

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B,$$

where

$$\sum \lambda_i^A = \sum \lambda_j^B = 0.$$

This additive log–linear model is equivalent to the column homogeneity model under the column multinomial sampling procedure of this section. The η-parameters and λ-parameters are related by the equations

$$\eta_{ii'} = \lambda_i^A - \lambda_{i'}^A \qquad \text{and} \qquad \lambda_i^A = \frac{1}{r}\sum_{i'} \eta_{ii'}.$$

The saturated model of Chapter 2 is equivalent to the multinomial response model

$$\tau_{ii'\cdot j}^{A\cdot B} = \eta_{ii'} + \eta_{ii'j}^B \tag{6.8}$$

in which

$$\sum_j \eta_{ii'j}^B = 0. \tag{6.9}$$

Equation (6.8) can be converted into the form used in (6.1) if, as in Chapter 2, $x_{jj'}^B$ is defined for $1 \leq j' \leq s - 1$ as

$$\begin{aligned} x_{jj'}^B &= 1, \qquad j = j', \\ &= 0, \qquad j \neq j', \quad j < s, \\ &= -1, \qquad j = s. \end{aligned}$$

Then (6.8) is equivalent to

$$\tau_{ii'\cdot j}^{A\cdot B} = \eta_{ii'} + \sum_{j'=1}^{s-1} \eta_{ii'j'}^B x_{jj'}^B.$$

Equation (6.8) imposes no restrictions on the $\tau_{ii'\cdot j}^{A\cdot B}$, for given any $\tau_{ii'\cdot j}^{A\cdot B}$, the η-parameters may be defined by the equations

$$\eta_{ii'} = \frac{1}{s}\sum_j \tau_{ii'\cdot j}^{A\cdot B} \tag{6.10}$$

and

$$\eta_{ii'j}^B = \tau_{ii'j}^{A\cdot B} - \eta_{ii'}. \tag{6.11}$$

In terms of the parametrization

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB},$$
$$\sum \lambda_i^A = \sum \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0,$$

one has

$$\tau_{ii'\cdot j}^{A\cdot B} = (\lambda_i^A - \lambda_{i'}^A) + (\lambda_{ij}^{AB} - \lambda_{i'j}^{AB}),$$

so that

$$\eta_{ii'} = \lambda_i^A - \lambda_{i'}^A \tag{6.12}$$

and

$$\eta_{ii'j}^B = \lambda_{ij}^{AB} - \lambda_{i'j}^{AB}. \tag{6.13}$$

Similarly,

$$\lambda_i^A = \frac{1}{r}\sum_{i'} \eta_{ii'} \tag{6.14}$$

and

$$\lambda_{ij}^{AB} = \frac{1}{r}\sum_{i'} \eta_{ii'j}^B. \tag{6.15}$$

Thus the multinomial response models include the models employed so far in this book for analysis of $r \times s$ contingency tables with column multinomial sampling.

Mental Health Status and Parental Socioeconomic Status

To illustrate use of multinomial response models, Table 6.1, which also appears in Haberman (1974b), may be considered. These data originally appear in a study by Srole, Langner, Opler, and Rennie (1962) which attempted to examine relationships between mental illness and socioeconomic status. Subjects were obtained from a probability sample of the resident midtown Manhattan population living at home. The survey was conducted in 1954. Of 1911 persons contacted, 1660 permitted themselves to be interviewed.

Details of the classification of subjects by mental health category and parental socioeconomic status are described by Srole *et al.* In this section, the task is to describe the relationship between the variables reported in the table. Questions of rating biases, sampling biases, and causation are all important, but they are not considered here.

Models for Table 6.1 will assume that mental health category A_h is the dependent variable and parental socioeconomic status B_h is the independent variable. Thus $n_{11} = 64$, $n_{21} = 94$, etc.

Table 6.1

Subjects Cross-Classified by Mental Health Status and Parental Socioeconomic Status[a]

Mental health category	Parental socioeconomic status stratum						
	A	B	C	D	E	F	Total
Well	64	57	57	72	36	21	307
Mild symptom formation	94	94	105	141	97	71	602
Moderate symptom formation	58	54	65	77	54	54	363
Impaired	46	40	60	94	78	71	389
Total	262	245	287	384	265	217	1660

[a] Srole, Langner, Michael, Opler, and Rennie (1962, p. 213).

Column homogeneity The two multinomial response models from Chapter 2 that may be applied to these data are the column homogeneity model and the saturated model. The column homogeneity model is unsatisfactory, for the Pearson chi-square is 46.0, the likelihood-ratio chi-square is 47.4, and there are (4-1) (6-1) = 15 degrees of freedom.

The saturated model Obviously, the saturated model does fit the data. Examination of estimated η^B-parameters can be used to suggest multinomial response models simpler than the saturated model but less restrictive than the column homogeneity model. Since $\tau_{ii'\cdot j}^{A\cdot B}$ has maximum likelihood estimate

$$\hat{\tau}_{ii'\cdot j}^{A\cdot B} = \log\left(\frac{n_{ij}}{n_{i'j}}\right)$$

under the saturated model, $n_{ii'}$ has maximum likelihood estimate

$$\hat{\eta}_{ii'} = \frac{1}{s}\sum_j \log\left(\frac{n_{ij}}{n_{i'j}}\right)$$

and $\eta_{ii'j}^B$ has maximum likelihood estimate

$$\hat{\eta}_{ii'j}^B = \log\left(\frac{n_{ij}}{n_{i'j}}\right) - \hat{\eta}_{ii'} = \sum_{j'=1}^{s} c_{jj'} \log\left(\frac{n_{ij'}}{n_{i'j'}}\right),$$

where

$$\begin{aligned} c_{jj'} &= 1 - 1/s, \qquad & j &= j', \\ &= -1/s, \qquad & j &\neq j'. \end{aligned}$$

The estimated asymptotic variance of $\hat{\eta}^{B}_{ii'j}$ is

$$s^2(\hat{\eta}^{B}_{ii'j}) = \sum_{j'=1}^{s} c_{jj'}^2\left(\frac{1}{n_{ij'}} + \frac{1}{n_{i'j'}}\right) = s^2(\hat{\eta}_{ii'}) + \left(1 - \frac{2}{s}\right)\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right),$$

where

$$s^2(\hat{\eta}_{ii'}) = \frac{1}{s^2}\sum_{j}\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right).$$

Results are summarized in Table 6.2. The choices of i and i' used in the table suffice to specify all $\hat{\eta}^{B}_{ii'j}$ since for $i < i'$,

$$\hat{\eta}^{B}_{ii'j} = \sum_{k=i}^{i'-1} \hat{\eta}^{B}_{k(k+1)j}.$$

The general pattern exhibited by the estimates $\hat{\eta}^{B}_{ii'j}$ is a tendency to decrease as j increases. The pattern is quite clear for $i = 1$ and $i' = 2$ and fairly clear for $i = 3$ and $i' = 4$. The pattern is less evident for $i = 2$ and $i' = 3$.

The simultaneous linear logit model A possible model suggested by Table 6.2 assumes that each logit $\tau^{A \cdot B}_{ii' \cdot j}$ is a linear function of the parental socio-

Table 6.2

Estimated η^B-Parameters for the Saturated Model for Table 6.1

i	i'	j	$\hat{\eta}^{B}_{ii'j}$	$s(\hat{\eta}^{AB}_{ii'j})$
1	2	1	0.345	0.152
		2	0.229	0.156
		3	0.119	0.154
		4	0.057	0.140
		5	−0.262	0.176
		6	−0.489	0.216
2	3	1	−0.014	0.152
		2	0.057	0.155
		3	−0.017	0.145
		4	0.108	0.134
		5	0.089	0.154
		6	−0.223	0.162
3	4	1	0.270	0.178
		2	0.338	0.186
		3	0.118	0.164
		4	−0.161	0.146
		5	−0.330	0.163
		6	−0.236	0.165

economic status score t_j, where $t_j = 2j - 7$, so that $t_1 = -5$, $t_2 = -3$, $t_3 = -1$, $t_4 = 1$, $t_5 = 3$, $t_6 = 5$, and $\sum t_j = 0$. Thus

$$\tau_{ii'\cdot j}^{A\cdot B} = \eta_{ii'} + \xi_{ii'} t_j. \tag{6.16}$$

As usual,

$$\eta_{ii} = 0, \qquad \eta_{ii''} = \eta_{ii'} + \eta_{i'i''}, \qquad \xi_{ii} = 0, \qquad \xi_{ii''} = \xi_{ii'} + \xi_{i'i''}.$$

The model appears in Simon (1974).

An equivalent log-linear representation To compute maximum likelihood estimates, the model is expressed in terms of the log m_{ij}. A series of weighted regression analyses are then performed using techniques from analysis of covariance. Since in the decomposition (6.8),

$$\eta_{ii'j}^{B} = \xi_{ii'} t_j,$$

(6.14) and (6.15) imply that

$$\begin{aligned}\lambda_i^A &= \frac{1}{r}\sum_{i'=1}^{r} \eta_{ii'} \\ &= -\frac{1}{r}\sum_{i'=1}^{i-1}\sum_{k=i'}^{i-1} \eta_{k(k+1)} + \frac{1}{r}\sum_{i'=i+1}^{r}\sum_{k=i}^{i'-1} \eta_{k(k+1)} \\ &= -\frac{1}{r}\sum_{k=1}^{i-1} k\eta_{k(k+1)} + \frac{1}{r}\sum_{k=i}^{r-1}(r-k)\eta_{k(k+1)}\end{aligned} \tag{6.17}$$

and

$$\lambda_{ij}^{AB} = \frac{1}{r}\sum_{i'=1}^{r} \eta_{ii'j}^{B} = \frac{1}{r} t_j \sum_{i'=1}^{r} \xi_{ii'} = -\frac{1}{r} t_j \sum_{k=1}^{i-1} k\xi_{k(k+1)} + \frac{1}{r} t_j \sum_{k=1}^{r-1}(r-k)\xi_{k(k+1)}. \tag{6.18}$$

Thus

$$\begin{aligned}\log m_{ij} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \\ &= \lambda + \lambda_i^B + \sum_{k=1}^{2(r-1)} \beta_k x_{ijk} = \alpha_j^B + \sum_{k=1}^{2(r-1)} \beta_k x_{ijk},\end{aligned} \tag{6.19}$$

where $\alpha_j^B = \lambda + \lambda_j^B$, $1 \le j \le s$, and for $1 \le k \le r - 1$,

$$\beta_k = \eta_{k(k+1)}, \qquad \beta_{k+r-1} = \beta_{k(k+1)},$$

$$\begin{aligned} x_{ijk} &= 1 - \frac{k}{r}, \quad i \le k, \qquad & x_{ij(k+r-1)} &= t_j\left(1 - \frac{k}{r}\right), \quad i \le k, \\ &= -\frac{k}{r}, \quad i > k, & &= -t_j\frac{k}{r}, \quad i > k.\end{aligned}$$

The x_{ijk}, $1 \leq k \leq 6$, for Table 6.1 are shown in Table 6.3. Arguments similar to those in Section 2.6 may be used to show that the maximum likelihood estimates $\hat{m}_{ij}$ of the means m_{ij} satisfy the equations

$$\hat{m}_j^B = n_j^B, \qquad 1 \leq j \leq s, \tag{6.20}$$

and

$$\sum_i \sum_j x_{ijk} \hat{m}_{ij} = \sum_i \sum_j x_{ijk} n_{ij}, \qquad 1 \leq k \leq 2(r-1). \tag{6.21}$$

As is evident from Bock (1975, p. 524) and Press (1972, p. 270), these equations are well known. For a derivation, see Exercise 6.8.

Table 6.3

Coefficients x_{ijk} for the Simultaneous Logit Model for Table 6.1

i	j	x_{ij1}	x_{ij2}	x_{ij3}	x_{ij4}	x_{ij5}	x_{ij6}
1	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{15}{4}$	$-\frac{5}{2}$	$-\frac{5}{4}$
2		$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{5}{4}$	$-\frac{5}{2}$	$-\frac{5}{4}$
3		$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{5}{4}$	$\frac{5}{2}$	$-\frac{5}{4}$
4		$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{5}{4}$	$\frac{5}{2}$	$\frac{15}{4}$
1	2	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{9}{4}$	$-\frac{3}{2}$	$-\frac{3}{4}$
2		$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}$	$-\frac{3}{2}$	$-\frac{3}{4}$
3		$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{2}$	$-\frac{3}{4}$
4		$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{3}{4}$	$\frac{3}{2}$	$\frac{9}{4}$
1	3	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$
2		$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$
3		$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$-\frac{1}{4}$
4		$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
1	4	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
2		$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
3		$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$
4		$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$
1	5	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{3}{2}$	$\frac{3}{4}$
2		$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{3}{4}$	$\frac{3}{2}$	$\frac{3}{4}$
3		$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{3}{4}$	$-\frac{3}{2}$	$\frac{3}{4}$
4		$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$-\frac{3}{4}$	$-\frac{3}{2}$	$-\frac{9}{4}$
1	6	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{15}{4}$	$\frac{5}{2}$	$\frac{5}{4}$
2		$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{5}{4}$	$\frac{5}{2}$	$\frac{5}{4}$
3		$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{5}{4}$	$-\frac{5}{2}$	$\frac{5}{4}$
4		$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$-\frac{5}{4}$	$-\frac{5}{2}$	$-\frac{15}{4}$

The Newton–Raphson algorithm for multinomial response models relies on the fact that if $m_j^B = n_j^B$ and if

$$v_{ij} = \sum_{k=1}^{2(r-1)} \beta_k x_{ijk},$$

then

$$m_j^B = \sum_i m_{ij} = \exp(\alpha_j^B) \sum_i \exp v_{ij} = n_j^B,$$

so that

$$\alpha_j^B = \log\left(n_j^B \Big/ \sum_i \exp v_{ij}\right).$$

Thus, given the constraint $m_j^B = n_j^B$, $1 \le j \le s$, the β_k, $1 \le k \le q$, determine the α_j^B, $1 \le j \le s$. Consequently, the algorithm can concentrate on computation of the $\hat{\beta}_k$. Algorithms of the type used in this section are found in the appendix and in Bock and Yates (1973).

As usual, the algorithm requires initial estimates m_{ij0} of the n_{ij}. The choice used here is $m_{ij0} = n_{ij} + \frac{1}{2}$.

From these estimates, empirical logarithms

$$y_{ij0} = \log m_{ij0}$$

are obtained. Then estimates β_{k0} are obtained as in the weighted regression problem

$$y_{ij0} = \alpha_j^B + \sum_k \beta_k x_{ijk} + \varepsilon_{ij} = \lambda + \lambda_j^B + \sum_i \beta_k x_{ijk} + \varepsilon_{ij},$$

$$\sum \lambda_j^B = 0,$$

where the ε_{ij} are independent random variables with respective means 0 and variances m_{ij0}^{-1}.

This regression problem corresponds to one-way analysis of covariance. It is well known that the β_{k0} satisfy the simultaneous equations

$$\sum_l S_{kl0} \beta_{l0} = w_{k0} = \sum_i (x_{ijk} - \theta_{jk0}) y_{ij0} m_{ij0}, \tag{6.22}$$

where

$$\theta_{jk0} = \sum_i x_{ijk} m_{ij0} \Big/ \sum_i m_{ij0} \tag{6.23}$$

and

$$S_{kl0} = \sum_i \sum_j (x_{ijk} - \theta_{jk0})(x_{ijl} - \theta_{jl0}) m_{ij0}. \tag{6.24}$$

In matrix terminology,

$$\boldsymbol{\beta}_0 = S_0^{-1}\mathbf{w}_0. \tag{6.25}$$

Given the β_{k0}, one may let

$$v_{ij0} = \sum_k \beta_{k0} x_{ijk} \tag{6.26}$$

and

$$g_{j0} = n_j^B \Big/ \sum_i \exp v_{ij0}, \tag{6.27}$$

so that

$$m_{ij1} = g_{j0} \exp v_{ij0} \tag{6.28}$$

satisfies the condition

$$m_{j1}^B = n_j^B.$$

At iteration $\nu \geq 1$, weighted regression analysis is based on the working logarithms

$$y_{ij\nu} = \log m_{ij\nu} + (n_{ij} - m_{ij\nu})/m_{ij\nu}. \tag{6.29}$$

The weighted regression model

$$y_{ij\nu} = \alpha_j^B + \sum_k \beta_k x_{ijk} + \varepsilon_{ij}$$

is used to find the $\beta_{k\nu}$, where the ε_{ij} are independent random variables with respective means 0 and variances m_{ij}. Thus the $\beta_{k\nu}$ satisfy the simultaneous equations

$$\sum_l S_{kl\nu} \beta_{l\nu} = w_{k\nu}, \tag{6.30}$$

where

$$\theta_{jk\nu} = \sum_i x_{ijk} m_{ij\nu} \Big/ \sum_i m_{ij\nu} = (n_j^B)^{-1} \sum_i x_{ijk} m_{ij\nu}, \tag{6.31}$$

$$w_{k\nu} = \sum_i \sum_j (x_{ijk} - \theta_{jk\nu}) y_{ij\nu} m_{ij\nu}, \tag{6.32}$$

and

$$S_{kl\nu} = \sum_i \sum_j (x_{ijk} - \theta_{jk\nu})(x_{ijl\nu} - \theta_{jl\nu}) m_{ij\nu}. \tag{6.33}$$

Given the $\beta_{k\nu}$, a new estimate $m_{ij(\nu+1)}$ of $\hat{m}_{ij}$ is found from the equations

$$v_{ij\nu} = \sum_k \beta_{k\nu} x_{ijk}, \tag{6.34}$$

$$g_{j\nu} = n_j^B \Big/ \sum_i \exp v_{ij\nu}, \tag{6.35}$$

and

$$m_{ij(\nu+1)} = g_{j\nu} \exp v_{ij\nu}. \tag{6.36}$$

Normally, $m_{ij\nu}$ approaches $\hat{m}_{ij}$ and $\beta_{k\nu}$ approaches $\hat{\beta}_k$ as ν becomes large. To simplify calculations, one may use the observation that if

$$a_k = \sum (x_{ijk} - \theta_{jk\nu})(n_{ij} - m_{ij\nu}) = \sum x_{ijk} n_{ij} - \sum x_{ijk} m_{ij\nu}, \tag{6.37}$$

and if the $\delta_{k\nu}$, $1 \le k \le 2(r-1)$, are defined by the simultaneous equations

$$\sum_l S_{kl\nu} \delta_{l\nu} = a_{k\nu}, \tag{6.38}$$

then

$$\beta_{k\nu} = \beta_{k(\nu-1)} + \delta_{k\nu}. \tag{6.39}$$

In matrix terms,

$$\boldsymbol{\beta}_\nu = \boldsymbol{\beta}_{\nu-1} + S_\nu^{-1} \mathbf{a}_\nu. \tag{6.40}$$

Calculations are summarized in Table 6.4. The program in the appendix performs these computations, while Bock and Yates (1973) proceed in a similar manner. Convergence is clearly very rapid in this example.

Chi-square statistics The chi-square statistics

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \tag{6.41}$$

and

$$L^2 = 2 \sum_i \sum_j n_{ij} \log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right) \tag{6.42}$$

have $rs - s - 2(r-1) = (r-1)(s-2) = (4-1)(6-2) = 12$ degrees of freedom, for there are rs counts n_{ij}, s parameters α_j^B, and $2(r-1)$ parameters β_k. In this example, $X^2 = 6.29$ and $L^2 = 6.28$, so the simultaneous linear logit model fits the data very well.

Adjusted residuals Further confirmation that the model is consistent with the data may be obtained through the adjusted residuals

$$r_{ij} = (n_{ij} - \hat{m}_{ij})/\hat{c}_{ij}^{1/2}. \tag{6.43}$$

Table 6.4

Computation of Maximum Likelihood Estimates for the Simultaneous Linear Logit Model of Table 6.1 Using the Newton–Raphson Algorithm

i	j	m_{ij0}	y_{ij0}	v_{ij0}	m_{ij1}	v_{ij1}	m_{ij2}
1	1	64.5	4.1667	0.09171	68.42	0.09511	68.56
2		94.5	4.5486	0.43737	96.67	0.44121	96.42
3		58.5	4.0690	−0.11400	55.70	−0.11264	55.70
4		46.5	3.8395	−0.41509	41.22	−0.42368	40.81
1	2	57.5	4.0518	−0.05902	55.71	−0.05871	55.68
2		94.5	4.5486	0.42814	90.68	0.43246	91.00
3		54.5	3.9982	−0.10352	53.29	−0.10304	53.27
4		40.5	3.7013	−0.26560	45.31	−0.27071	45.05
1	3	57.5	4.0518	−0.20976	56.31	−0.21253	56.11
2		105.5	4.6587	0.41891	105.58	0.42371	106.01
3		65.5	4.1821	−0.09305	63.28	−0.09344	63.20
4		60.5	4.1026	−0.11611	61.83	−0.11773	61.69
1	4	72.5	4.2836	−0.36049	64.39	−0.36636	63.95
2		141.5	4.9523	0.40968	139.10	0.41496	139.68
3		77.5	4.3503	−0.08257	85.03	−0.08385	84.82
4		94.5	4.5486	0.03338	95.48	0.03524	95.55
1	5	36.5	3.5973	−0.51123	37.63	−0.52018	37.23
2		97.5	4.5799	0.40045	93.65	0.40621	94.02
3		54.5	3.9982	−0.07210	58.38	−0.07425	58.15
4		78.5	4.3631	0.18287	75.34	0.18822	75.60
1	6	21.5	3.0681	−0.66196	25.86	−0.67400	25.47
2		71.5	4.2697	0.39122	74.13	0.39746	74.37
3		54.5	3.9982	−0.06162	47.13	−0.06466	46.85
4		71.5	4.2697	0.33236	69.89	0.34119	70.30

k	w_{k0}	S_{k10}	S_{k20}	S_{k30}	S_{k40}	S_{k50}	S_{k60}	β_{k0}
1	−95.29	248.70	135.70	68.33	−171.75	−70.98	−12.09	−0.69942
2	100.34	135.70	408.38	209.20	−70.98	0.50	73.48	0.50211
3	34.13	68.33	209.20	294.95	−12.09	2488.83	137.45	−0.04644
4	−151.17	−171.75	−70.98	−12.09	2488.83	1370.63	670.73	−0.07075
5	−239.67	−70.98	0.50	73.48	1370.63	4211.95	2124.61	−0.00985
6	−243.81	−12.09	73.48	137.45	670.73	2124.61	3084.37	−0.06951

k	a_{k1}	S_{k11}	S_{k21}	S_{k31}	S_{k41}	S_{k51}	S_{k61}
1	−1.32	247.51	135.48	68.24	−162.81	−61.40	−7.12
2	0.87	135.48	406.37	208.05	−61.40	11.31	77.60
3	0.07	68.24	208.05	293.22	−7.12	77.60	137.42
4	−4.04	−162.81	−61.40	−7.12	2610.02	1408.79	694.43
5	−3.73	−61.40	11.31	77.60	1408.79	4167.21	2119.35
6	−5.93	−7.12	77.60	137.42	694.43	2119.35	2990.72

Table 6.4 (continued)

k	δ_{k1}	β_{k1}
1	−0.00936	−0.70878
2	0.00588	0.50798
3	−0.00096	−0.04740
4	−0.00178	−0.07254
5	0.00068	−0.00917
6	−0.00218	−0.07169

The estimated asymptotic variance $\hat{c}_{ij}$ of $n_{ij} - \hat{m}_{ij}$ is given by the formula

$$\hat{c}_{ij} = \hat{m}_{ij}\left[1 - \hat{m}_{ij}/n_j^B - \sum_k \sum_l (x_{ijk} - \hat{\theta}_{jk})(x_{ijl} - \hat{\theta}_{jl})\hat{S}^{kl}\right], \tag{6.44}$$

where

$$\hat{\theta}_{jk} = \sum_i x_{ijk}\hat{m}_{ijk} \Big/ \sum_i \hat{m}_{ijk} = (n_j^B)^{-1} \sum_i x_{ijk}\hat{m}_{ij}, \tag{6.45}$$

$$\hat{S}_{kl} = \sum_i \sum_j (x_{ijk} - \hat{\theta}_{jk})(x_{ijl} - \hat{\theta}_{jl})\hat{m}_{ij}, \tag{6.46}$$

and $\hat{S}^{kl}$, $1 \le k \le 2(r-1)$, $1 \le l \le 2(r-1)$, are the elements of $\hat{S}^{-1}$, the inverse of the matrix $\hat{S}$ with elements $\hat{S}_{kl}$, $1 \le k \le 2(r-1)$, $1 \le l \le 2(r-1)$.

The right-hand side of (6.44) is the variance of $\hat{m}_{ij}R_{ij}$, where R_{ij} is the residual

$$R_{ij} = Y_{ij} - a_j^B - \sum_k b_k x_{ijk} \tag{6.47}$$

in a weighted regression model

$$Y_{ij} = \alpha_j^B + \sum_k \beta_k x_{ijk} + \varepsilon_{ij} \tag{6.48}$$

in which the Y_{ij} are hypothetical responses and each ε_{ij} is an independent random variable with mean 0 and variance $\hat{m}_{ij}^{-1}$. The weighted-least-squares estimate of α_j^B is a_j^B, while β_k has weighted-least-squares estimate b_k.

Results are summarized in Table 6.5. As should be expected given the low values of the chi-square statistics, the adjusted residuals are quite modest in size and show little pattern.

Parameter estimates The large-sample properties of the vector $\hat{\boldsymbol{\beta}}$ of maximum likelihood estimates $\hat{\beta}_k$, $1 \le k \le 2(r-1)$, are approximately the same as those of the vector $\mathbf{b}$ of weighted-least-squares estimates b_k, $1 \le k \le 2(r-1)$, in a weighted regression model in which (6.48) holds for errors ε_{ij} which are independently distributed with $N(0, m_{ij}^{-1})$ distributions. The vector

Table 6.5

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Simultaneous Linear Logit Model for Table 6.1[a]

Mental health category	Parental socioeconomic status stratum					
	A	B	C	D	E	F
Well	64	57	57	72	36	21
	68.57	55.68	56.11	63.94	37.23	25.47
	−1.00	0.24	0.15	1.30	−0.26	−1.20
Mild symptom	94	94	105	141	97	71
formation	96.92	91.00	106.01	139.68	94.02	74.37
	−0.55	0.46	−0.14	0.16	0.46	−0.65
Moderate symptom	58	54	65	77	54	54
formation	55.70	53.27	63.20	84.82	58.15	46.85
	0.50	0.13	0.28	−1.12	−0.74	1.59
Impaired	46	40	60	94	78	71
	40.81	45.05	61.68	95.55	75.60	70.30
	1.22	−0.98	−0.27	−0.21	0.39	0.14

[a] First line is observed count, second line is estimated expected count, and third line is adjusted residual.

b is equal to $S^{-1}\mathbf{w}$, where S^{-1} is the inverse of S, S is a $2(r-1)$ by $2(r-1)$ matrix with coordinates

$$S_{kl} = \sum_i \sum_j (x_{ijk} - \theta_{jk})(x_{ijl} - \theta_{jl}) m_{ij}, \tag{6.49}$$

$$\theta_{jk} = (n_j^B)^{-1} \sum_i x_{ijk} m_{ij}, \tag{6.50}$$

and **w** is a vector with coordinates

$$w_k = \sum_i \sum_j (x_{ijk} - \theta_{jk}) Y_{ij} m_{ij}.$$

Let $\boldsymbol{\beta}$ have coordinates β_k, $1 \le k \le 2(r-1)$. Then the weighted-least-squares estimate **b** has an $N(\boldsymbol{\beta}, S^{-1})$ distribution. Similarly, the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ has an approximate $N(\boldsymbol{\beta}, S^{-1})$ distribution with the approximation becoming increasingly accurate as the column totals n_j^B become large. To estimate S^{-1}, one may use $\hat{S}^{-1}$, where $\hat{S}$ is defined as in (6.46). Thus $\hat{\beta}_k$ has EASD $s(\hat{\beta}_k) = (\hat{S}^{kk})^{1/2}$, and an approximate 95 percent confidence interval for β_k has lower bound

$$\hat{\beta}_k - 1.96 s(\hat{\beta}_k)$$

and upper bound

$$\hat{\beta}_k + 1.96 s(\hat{\beta}_k).$$

Table 6.6

Parameter Estimates for the Simultaneous Linear Logit Model for Table 6.1

Parameter	Estimate	EASD	Approximate 95 percent confidence interval: Lower bound	Upper bound
$\beta_1 = \eta_{12}$	−0.709	0.072	−0.850	−0.568
$\beta_2 = \eta_{23}$	0.508	0.067	0.378	0.638
$\beta_3 = \eta_{34}$	−0.048	0.074	−0.192	0.097
$\beta_4 = \xi_{12}$	−0.073	0.022	−0.116	−0.029
$\beta_5 = \xi_{23}$	−0.009	0.021	−0.050	0.032
$\beta_6 = \xi_{34}$	−0.072	0.023	−0.117	−0.026

Parameter estimates, estimated asymptotic standard deviations, and confidence intervals are summarized in Table 6.6. The estimates $\hat{\eta}_{12}$, $\hat{\eta}_{23}$, and $\hat{\eta}_{34}$ are quite variable; however, $\hat{\xi}_{12}$, $\hat{\xi}_{23}$, and $\hat{\xi}_{34}$ are similar, especially in the case of $\hat{\xi}_{12}$ and $\hat{\xi}_{34}$. These results suggest that a log-linear model should be considered in which $\xi_{12} = \xi_{23} = \xi_{34} = \xi$, so that $\xi_{ii'} = \xi(i' - i)$.

Linear-by-Linear Interaction and Parallel Logits

Analysis for a multinomial response model of the form

$$\tau_{ii' \cdot j}^{A \cdot B} = \eta_{i(i+1)} + \xi(i' - i)t_j$$

is quite similar to analysis for the simultaneous linear logit model. Since

$$\eta_{ii'j}^{B} = \xi(i' - i)t_j,$$

the λ-parameter λ_{ij}^{AB} is equal to

$$\frac{1}{r}\sum_{i'} \xi(i' - i)t_j = -\xi\left(i - \frac{r+1}{2}\right)t_j = \xi u_i t_j.$$

In Table 6.1, $u_i = \frac{3}{2}$, $u_2 = \frac{1}{2}$, $u_3 = -\frac{1}{2}$, and $u_4 = -\frac{3}{2}$. Thus

$$\begin{aligned}\log m_{ij} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \\ &= \lambda + \lambda_j^B + \sum_{k=1}^{r} \beta_k x_{ijk} = \alpha_j^B + \sum_{k=1}^{r} \beta_k x_{ijk},\end{aligned} \tag{6.51}$$

where $\alpha_j^B = \lambda + \lambda_j^B$, λ_k, $1 \le k \le r - 1$, is defined as in (6.19); x_{ijk}, $1 \le k \le r - 1$, is defined as in Table 6.3;

$$\beta_r = \xi \qquad \text{and} \qquad x_{ijr} = u_i t_j.$$

Since the interaction λ_{ij}^{AB} is a multiple of a product of a linear score u_i for the categories i of A_h and a linear score t_j for the categories of B_h, (6.51) is called a model of linear-by-linear interaction. Haberman (1974b) and Nelder and Wedderburn (1972) have used this model.

The Newton–Raphson algorithm for this multinomial response model is very similar to the algorithm for the simultaneous linear logit model, so computation of maximum likelihood estimates and residuals is left as an exercise (Exercise 6.1). Results are summarized in Tables 6.7 and 6.8. This model also fits the data quite well, as indicated by the modest size of the

Table 6.7

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Log-Linear Model of Linear-by-Linear Interaction, as Applied to Data in Table 6.1[a]

Mental health category	Parental socioeconomic status stratum					
	A	B	C	D	E	F
Well	64	57	57	71	36	21
	65.29	54.21	55.91	65.28	38.96	27.35
	−0.23	0.49	0.18	1.05	−0.58	−1.50
Mild symptom formation	94	94	105	141	97	71
	104.42	94.94	107.20	137.04	89.56	68.84
	−1.47	−0.14	−0.30	0.48	1.06	0.35
Moderate symptom formation	58	54	65	77	54	54
	50.24	49.92	61.72	86.39	61.82	52.02
	1.36	0.70	0.52	−1.32	−1.25	0.34
Impaired	46	40	60	94	78	71
	42.14	45.93	62.18	95.29	74.66	68.80
	0.81	−1.10	−0.35	−0.18	0.53	0.40

[a] First line is observed count, second line is estimated expected count, and third line is adjust residual.

Table 6.8

Parameter Estimates for the Log Linear Model of Linear-by-Linear Interaction for Table 6.1

Parameter	Estimate	EASD	Approximate 95 percent confidence interval	
			Lower bound	Upper bound
$\beta_1 = \eta_{12}$	−0.696	0.071	−0.835	−0.558
$\beta_2 = \eta_{23}$	0.507	0.067	0.376	0.637
$\beta_3 = \eta_{34}$	−0.053	0.073	−0.197	0.091
$\beta_4 = \xi$	−0.0453	0.0075	−0.0600	−0.0306

adjusted residuals and by the Pearson chi-square of 9.73 and the likelihood-ratio chi-square of 9.90. Since there are rs cells in the table, s parameters α_j^B, and r parameters β_k, there are $rs - r - s = 24 - 4 - 6 = 14$ degrees of freedom. Thus the chi-square statistics are quite small. The difference in likelihood-ratio chi-square statistics between the simultaneous linear logit model and the linear-by-linear model is 3.61, while the difference in degrees of freedom is 2. Thus the change in likelihood-ratio chi-square is not large enough to indicate inadequacy in the linear-by-linear model.

The parameter estimate of greatest interest is $\hat{\xi}$, for $2\hat{\xi}$ estimates the change in the log odds $\tau_{i(i+1)\cdot j}^{A\cdot B}$ associated with a change of parental socioeconomic status j of 1. It is estimated that the corresponding odds $q_{i(i+1)\cdot j}^{A\cdot B} = \exp \tau_{i(i+1)\cdot j}^{A\cdot B}$ is decreased by $100(1 - e^{2(-0.0453)}) = 8.66$ percent when j is increased by 1. This change in odds between adjacent categories is modest; however, the estimate $\hat{\xi}$ corresponds to quite substantial contrasts between subjects with parental socioeconomic status stratum A and those with parental socioeconomic status stratum F. For example, given the model, the maximum likelihood estimate $\hat{q}_{14.1}^{A\cdot B} = \hat{m}_{11}/\hat{m}_{41}$ is 1.55. In contrast, $\hat{q}_{14.6}^{A\cdot B} = \hat{m}_{16}/\hat{m}_{46}$ is 0.40. The first ratio provides an estimate that given parental socioeconomic status stratum A, a subject is 1.55 times as likely to have mental health category well as the subject is likely to have mental health category impaired. The second ratio provides an estimate that given parental socioeconomic status stratum F, a subject is 0.40 times as likely to have mental health category well as the subject is likely to have mental health category impaired.

The contrast chosen is the most extreme available; nevertheless, it does serve to point out an apparent strong relationship between mental health and parental socioeconomic status. As noted earlier, the reality of the relationship depends on methodological questions that cannot be inferred from the table.

A Model for Known Row Scores and Unknown Column Scores

So far, the scores t_j for the categories j of the independent variable have been assumed known, and the interaction term λ_{ij}^{AB} has been equal to $v_i t_j$ for some unknown scores v_i, $1 \le i \le r$, for the dependent variable such that $\sum v_i = 0$. In the simultaneous linear logit model,

$$\xi_{ii'} = v_i - v_{i'} = \sum_{k=i}^{i'-1} \xi_{k(k+1)}, \qquad i < i',$$

so that $\xi_{i(i+1)}$ provides a measure of distance $v_{i+1} - v_i$ between categories i and $i + 1$. In the linear-by-linear log-linear model,

$$v_i - v_{i'} = \xi(i' - i), \qquad i < i',$$

so that the distance $v_{i+1} - v_i$ between categories i and $i + 1$ is a constant ξ.

Cases can arise in which scores u_i, $1 \leq i \leq r$ are given for the categories of the independent variable, but scores w_j for the categories of the dependent variable are unknown. It is assumed that

$$\lambda_{ij}^{AB} = u_i w_j,$$

where

$$\sum u_i = \sum w_j = 0.$$

If $\zeta_{jj'} = w_{j'} - w_j$, then

$$\zeta_{jj'} = \sum_{k=j}^{j'-1} \zeta_{k(k+1)}, \qquad j < j',$$

so that $\zeta_{j(j+1)}$ provides a measure of distance between adjacent categories j and $j+1$ of the independent variable. If the linear-by-linear model

$$\lambda_j^{AB} = \xi u_i t_j$$

of this section holds, then

$$\zeta_{jj'} = (t_{j'} - t_j) = 2\xi(j' - j).$$

To estimate the $\zeta_{j(j+1)}$, observe that

$$\begin{aligned}\lambda_{ij}^{AB} &= \frac{1}{s}\sum_{j'=1}^{s}(\lambda_{ij}^{AB} - \lambda_{ij'}^{AB}) \\ &= \frac{u_i}{s}\sum_{j'=1}^{s}\zeta_{jj'} \\ &= \frac{u_i}{s}\left[-\sum_{k=1}^{j-1} k\zeta_{k(k+1)} + \sum_{k=j}^{s}(s-k)\zeta_{k(k+1)}\right].\end{aligned}$$

Therefore,

$$\begin{aligned}\log m_{ij} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \\ &= \alpha_j^B - \frac{1}{r}\sum_{k=1}^{i-1} k\eta_{k(k+1)} + \frac{1}{r}\sum_{k=i}^{r}(r-k)\eta_{k(k+1)} \\ &\quad - \frac{u_i}{s}\sum_{k=1}^{j-1} k\zeta_{k(k+1)} + \frac{u_i}{s}\sum_{k=j}^{s}(s-k)\zeta_{k(k+1)} \\ &= \alpha_j^B + \sum_{k=1}^{r+s-2} \beta_k x_{ijk}. \end{aligned} \tag{6.52}$$

For $1 \le k \le r - 1$,

$$\begin{aligned} \beta_k &= \eta_{k(k+1)}, \\ x_{ijk} &= -k/r, \qquad & i > k, \\ &= (r - k)/r, \qquad & i \le k. \end{aligned}$$

For $1 \le k \le s - 1$,

$$\beta_{r+k-1} = \zeta_{k(k+1)}$$

and

$$\begin{aligned} x_{ij(r+k-1)} &= -u_i k/s, \qquad & j > k, \\ &= u_i(s - k)/s, \qquad & j \le k. \end{aligned}$$

The x_{ijk}, $1 \le k \le 8$, for Table 6.1 are shown in Table 6.9.

Table 6.9

Coefficients in the Model of Unknown Column Scores for Table 6.1

i	j	x_{ij1}	x_{ij2}	x_{ij3}	x_{ij4}	x_{ij5}	x_{ij6}	x_{ij7}	x_{ij8}
1	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{5}{4}$	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
2	1	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$
3	1	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$-\frac{5}{12}$	$-\frac{1}{3}$	$-\frac{1}{4}$	$-\frac{1}{6}$	$-\frac{1}{12}$
4	1	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$-\frac{5}{4}$	-1	$-\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$
1	2	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
2	2	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$
3	2	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$	$-\frac{1}{3}$	$-\frac{1}{4}$	$-\frac{1}{6}$	$-\frac{1}{12}$
4	2	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{4}$	-1	$-\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$
1	3	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
2	3	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{12}$	$-\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$
3	3	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{6}$	$-\frac{1}{4}$	$-\frac{1}{6}$	$-\frac{1}{12}$
4	3	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$-\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$
1	4	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
2	4	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{12}$	$-\frac{1}{6}$	$-\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$
3	4	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$-\frac{1}{6}$	$-\frac{1}{12}$
4	4	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	$-\frac{1}{2}$	$-\frac{1}{4}$
1	5	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	-1	$\frac{1}{4}$
2	5	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{12}$	$-\frac{1}{6}$	$-\frac{1}{4}$	$-\frac{1}{3}$	$\frac{1}{12}$
3	5	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{3}$	$-\frac{1}{12}$
4	5	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	$-\frac{1}{4}$
1	6	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	-1	$-\frac{5}{4}$
2	6	$-\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{12}$	$-\frac{1}{6}$	$-\frac{1}{4}$	$-\frac{1}{3}$	$-\frac{5}{12}$
3	6	$-\frac{1}{4}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{5}{12}$
4	6	$-\frac{1}{4}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	$\frac{5}{4}$

The Newton–Raphson algorithm for this model is very similar to the Newton–Raphson algorithm for the other models in this section. Using this algorithm, the maximum likelihood estimates and adjusted residuals in Tables 6.10 and 6.11 are found. As should be expected given the good fit provided by the linear-by-linear model, the new model fits the data quite well. Adjusted residuals are modest, and the chi-square statistics are relatively

Table 6.10

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Model of Unknown Column Scores for Table 6.1[a]

Mental health category	Parental socioeconomic status stratum					
	A	B	C	D	E	F
Well	64	57	57	72	36	21
	60.15	57.36	56.44	69.86	38.52	24.78
	0.89	−0.06	0.13	0.44	−0.63	−1.10
Mild symptom formation	94	94	105	141	97	71
	102.54	96.31	107.58	140.38	89.22	65.98
	−1.23	−0.34	−0.37	0.08	1.22	0.92
Moderate symptom formation	58	54	65	77	54	54
	52.54	48.61	61.54	84.65	62.02	52.72
	0.98	0.98	0.57	−1.11	−1.30	0.22
Impaired	46	40	60	94	78	71
	46.84	42.82	61.45	89.10	75.25	73.53
	−0.23	−0.79	−0.37	1.11	0.71	−0.70

[a] First line is observed count, second line is estimated expected count, and third line is adjusted residual.

Table 6.11

Estimated Parameters for the Model of Unknown Column Scores for Table 6.1

Coefficient	Estimate	EASD
$\beta_1 = \eta_{12}$	−0.703	0.071
$\beta_2 = \eta_{23}$	0.501	0.069
$\beta_3 = \eta_{34}$	−0.056	0.074
$\beta_4 = \zeta_{12}$	−0.013	0.087
$\beta_5 = \zeta_{23}$	0.125	0.085
$\beta_6 = \zeta_{34}$	0.053	0.075
$\beta_7 = \zeta_{45}$	0.142	0.077
$\beta_8 = \zeta_{56}$	0.139	0.088

small. The Pearson chi-square is 6.78, the likelihood ratio chi-square is 6.83, and there are

$$rs - s - (r + s - 2) = (r - 2)(s - 1) = 10$$

degrees of freedom. Since the decrease in L^2 from the linear-by-linear model is only 3.07, and the decrease in degrees of freedom is $s - 2 = 4$, the new model does not represent a clear improvement over the old one. This conclusion is supported by the limited variation observed in the estimates $\zeta_{i(i+1)}$.

If neither row nor column scores are assigned in advance, then these scores can be estimated simultaneously. The analysis is somewhat more complex than others considered in this paper since the model used is not a log-linear model. Andersen (1979) has considered this problem.

Scoring Systems

Each model considered in this section is based on assignment of one or more scores to each category of each variable. In each model, category i of variable A_h receives $r' \le r - 1$ scores $q^A_{ii'}$, $1 \le i' \le r'$, and category j of variable B_h receives $s' \le s - 1$ scores $q^B_{jj'}$, $1 \le j' \le s'$. The scores are chosen so that the sums

$$\sum_i q^A_{ii'}, \qquad 1 \le i \le r', \qquad \text{and} \qquad \sum_j q^B_{jj'}, \qquad 1 \le j \le s',$$

are all 0. Scores for each variable are independent. Thus no constants $c^A_{i'}$, $1 \le i' \le r'$, exist such that some $c^A_{i'}$ is not 0 and such that

$$\sum_{i'} c^A_{i'} q^A_{ii'} = 0, \qquad 1 \le i \le r.$$

Similarly, no $c^B_{j'}$, $1 \le j' \le s'$, exist such that some $c^B_{j'}$ is not 0 and such that

$$\sum_{j'} c^B_{j'} q^B_{jj'} = 0, \qquad 1 \le j \le s.$$

In (6.19) and (6.51) one may let the dependent variable have $r' = r - 1$ scores

$$\begin{aligned} q^A_{ii'} &= 1 - i/r, \qquad i \le i', \\ &= -i/r, \qquad i > i'. \end{aligned}$$

while the independent variable has the $s' = 1$ score

$$q^B_{j1} = t_j.$$

Then (6.19) can be written as

$$\log m_{ij} = \alpha^B_j + \sum_{i'} \gamma^A_{i'} q^A_{ii'} + \sum_{i'} \gamma^{AB}_{i'1} q^A_{ii'} q^B_{j1}, \tag{6.53}$$

where $\gamma_{i'}^{A} = \eta_{i'(i'+1)}$ and $\gamma_{i'1}^{AB} = \xi_{i'(i'+1)}$. In (6.51), each $\gamma_{i'1}^{AB}$ has a common value $\gamma_{\cdot 1}^{AB} = \xi$, so that

$$\log m_{ij} = \alpha_j^B + \sum_{i'} \gamma_{i'}^A q_{ii'}^A + \gamma_{\cdot 1}^{AB}\left(\sum_{i'} q_{ii'}^A\right) q_{jj'}^B. \tag{6.54}$$

On the other hand, in (6.52), one may set

$$\begin{aligned} q_{ii'}^A &= 1 - i/r, & i \le i', \\ &= -i/r, & i > i', \end{aligned}$$

for $1 \le i' \le r - 1$ and

$$\begin{aligned} q_{jj'}^B &= 1 - j/s, & j \le j', \\ &= -j/s, & j > j, \end{aligned}$$

for $1 \le j' \le s - 1$. If $\gamma_{i'}^A = \eta_{i'(i'+1)}$ and $\gamma_{\cdot j'}^{AB} = \zeta_{j'(j'+1)}$, then (6.51) can be written as

$$\log m_{ij} = \alpha_j^B + \sum_{i'} \gamma_i^A q_{ii'}^A + \sum_{j'} \gamma_{\cdot j'}^{AB}\left(\sum_{i'} q_{ii'}^A\right) q_{jj'}^B. \tag{6.55}$$

Orthogonal-polynomial scores Many other choices of scores can be used to obtain the same models. For example, orthogonal-polynomial scores may be applied to each variable to describe nonlinear effects associated with variables A_h and B_h. For variable A_h, there scores are chosen so that for $1 \le i' \le r - 1$, $q_{ii'}^A$ is a polynomial of degree i' in u_i, and so that for $i \le i' < i'' \le r - 1$, the orthogonality condition

$$\sum_i q_{ii'}^A q_{ii''}^A = 0$$

is satisfied. Similarly, for $1 \le j' \le s - 1$, $q_{jj'}^B$ is a polynomial of degree j in t_j, and for $1 \le j' < j'' \le s - 1$,

$$\sum_j q_{jj'}^B q_{jj''}^B = 0.$$

One choice of orthogonal-polynomial scores for Table 6.1 is listed in Table 6.12.

To illustrate the nature of these scores, consider q_{i1}^A and q_{i2}^A. The orthogonality condition holds since

$$\sum_i q_{i1}^A q_{i2}^A = (-3)(1) + (-1)(-1) + (1)(-1) + (3)(1) = 0.$$

The score q_{i1}^A is a polynomial of degree 1 in the u_i, for

$$q_{i1}^A = -2u_i.$$

Table 6.12

Orthogonal-Polynomial Scores for Table 6.1

i	$q^A_{ii'}$			j	$q^B_{jj'}$				
	i'				j'				
	1	2	3		1	2	3	4	5
1	−3	1	−1	1	−5	5	−5	1	−1
2	−1	−1	3	2	−3	1	7	−3	5
3	1	−1	−1	3	−1	−4	4	2	−10
4	3	1	3	4	1	−4	−4	2	10
				5	3	−1	−7	−3	−5
				6	5	5	5	1	1

The score q^A_{i2} is a polynomial of degree 2 in the u_i, since

$$q^A_{i2} = u_i^2 - \tfrac{5}{4}.$$

For tables of orthogonal-polynomial scores, see Fisher and Yates (1963, pp. 98–108), DeLury (1960), and Pearson and Hartley (1966, pp. 236–245). These tables apply if $u_i - u_{i+1}$ is constant for $1 \le i \le r - 1$ or if $t_j - t_{j+1}$ is constant for $1 \le j \le s - 1$. For a general procedure for calculation of orthogonal-polynomial scores, see Wishart and Metakides (1953). For some applications of these scores to statistical data, see Fisher (1921) and Bock (1975), among many others.

As noted by Haberman (1974b), among others, the λ-parameters can be described in terms of orthogonal-polynomial scores in a simple manner. The λ-parameters λ^A_i can be written as

$$\lambda^A_i = \sum_{i'} \gamma^A_{i'} q^A_{ii'},$$

where

$$\gamma^A_{i'} = \sum_i \lambda^A_i q^A_{ii'}/w^A_{i'} \qquad \text{and} \qquad w^A_{i'} = \sum_i (q^A_{ii'})^2.$$

Similarly,

$$\lambda^{AB}_{ij} = \sum_{i'} \sum_{j'} \gamma^A_{i'j'} q^A_{ii'} q^B_{jj'},$$

where

$$\gamma^A_{i'j'} = \sum_i \sum_j \lambda^{AB}_{ij} q^A_{ii'} q^B_{jj'}/(w^A_{i'} w^B_{j'}) \qquad \text{and} \qquad w^B_{j'} = \sum_j (q^B_{jj'})^2.$$

If orthogonal–polynomial scores are used, then (6.19) is equivalent to the log–linear model

$$\log m_{ij} = \alpha_j^B + \sum_{i'} \gamma_{i'}^A q_{ii'}^A + \sum_{i'} \gamma_{i'1}^{AB} q_{ii'}^A q_{j1}^B. \tag{6.56}$$

Here

$$\gamma_{i'}^A = \sum_i \lambda_i^A q_{ii'}^A / w_{i'}^A = \frac{1}{r}\left[\sum_{i=1}^{i'} (r - i)\eta_{i(i+1)} q_{ii'}^A - \sum_{i=i+1}^{r-1} i\eta_{i(i+1)} q_{ii'}^A\right] \bigg/ w_{i'}^A.$$

Thus

$$\gamma_1^A = -(1/20)(3\eta_{12} + 4\eta_{23} + 3\eta_{34}),$$

$$\gamma_2^A = \tfrac{1}{4}(\eta_{12} - \eta_{34}),$$

and

$$\gamma_3^A = -(1/20)(\eta_{12} - 2\eta_{23} + \eta_{34}).$$

Similarly,

$$\gamma_{11}^A = -(1/20)(3\xi_{12} + 4\xi_{23} + 3\xi_{34}),$$

$$\gamma_{21}^A = \tfrac{1}{4}(\xi_{12} - \xi_{34}),$$

and

$$\gamma_{31}^A = -(1/20)(\xi_{12} - 2\xi_{23} + \xi_{34}).$$

In the case of (6.51), an equivalent model is

$$\log m_{ij} = \alpha_j^B + \sum_{i'} \gamma_{i'}^A q_{ii'}^A + \gamma_{11}^{AB} q_{i1}^A q_{j1}^B. \tag{6.57}$$

Here the $\gamma_{i'}^A$ are defined as in (6.56), and

$$\gamma_{11}^{AB} = -(1/20)(3\xi_{12} + 4\xi_{23} + 3\xi_{34}) = -(1/2)\xi.$$

In the case of (6.52), the equivalent model is

$$\log m_{ij} = \alpha_j^B + \sum_{i'} \gamma_{i'}^A q_{ii'}^A + \sum_{j'} \gamma_{1j'}^{AB} q_{i1}^A q_{jj'}^B. \tag{6.58}$$

The saturated model has the form

$$\log m_{ij} = \alpha_j^B + \sum_{i'} \gamma_{i'}^A q_{ii'}^A + \sum_{i'} \sum_{j'} \gamma_{i'j'}^A q_{ii'}^A q_{jj'}^B. \tag{6.59}$$

Maximum likelihood estimates under these parametrizations are shown in Table 6.13. Computations require no special comment, except for the saturated model.

Table 6.13

Maximum Likelihood Estimates for Orthogonal-Polynomial Parameters for Models for Table 6.1

	Model							
	Saturated		Simultaneous linear logit		Linear-by-linear		Unknown column scores	
Parameter	Estimate	EASD	Estimate	EASD	Estimate	EASD	Estimate	EASD
γ_1^A	0.0158	0.0127	0.0118	0.0123	0.0110	0.0121	0.0136	0.0123
γ_2^A	−0.1728	0.0265	−0.1654	0.0258	−0.1609	0.0255	−0.1620	0.0256
γ_3^A	0.0881	0.0109	0.0886	0.0107	0.0881	0.0107	0.0880	0.0107
γ_{11}^{AB}	0.0249	0.0040	0.0235	0.0038	0.0227	0.0038	0.0232	0.0038
γ_{21}^{AB}	−0.0032	0.0083	−0.0002	0.0080	—	—	—	—
γ_{31}^{AB}	0.0063	0.0034	0.0063	0.0033	—	—	—	—
γ_{12}^{AB}	0.0059	0.0035	—	—	—	—	0.0050	0.0032
γ_{22}^{AB}	−0.0051	0.0072	—	—	—	—	—	—
γ_{32}^{AB}	−0.0014	0.0029	—	—	—	—	—	—
γ_{13}^{AB}	−0.0004	0.0024	—	—	—	—	−0.0006	0.0022
γ_{23}^{AB}	−0.0052	0.0049	—	—	—	—	—	—
γ_{33}^{AB}	−0.0018	0.0020	—	—	—	—	—	—
γ_{14}^{AB}	0.0023	0.0058	—	—	—	—	0.0021	0.0056
γ_{24}^{AB}	0.0034	0.0121	—	—	—	—	—	—
γ_{34}^{AB}	−0.0022	0.0050	—	—	—	—	—	—
γ_{15}^{AB}	−0.0013	0.0018	—	—	—	—	−0.0012	0.0017
γ_{25}^{AB}	0.0010	0.0037	—	—	—	—	—	—
γ_{35}^{AB}	0.0001	0.0016	—	—	—	—	—	—

As Haberman (1974b) notes,

$$\gamma_{i'}^A = \sum_i \sum_j q_{ii'}^A \log m_{ij}/(sw_{i'}^A) \tag{6.60}$$

and

$$\gamma_{i'j'}^{AB} = \sum_i \sum_j q_{ii'}^A q_{jj'}^B \log m_{ij}/(w_{i'}^A w_{j'}^B), \tag{6.61}$$

so that the estimates $\hat{\gamma}_{i'}^A$ and $\hat{\gamma}_{i'j'}^{AB}$ satisfy the equations

$$\hat{\gamma}_{i'}^A = \sum_i \sum_j q_{ii'}^A \log n_{ij}/(sw_{i'}^A), \tag{6.62}$$

$$s^2(\hat{\gamma}_{i'}^A) = \left[\sum_i \sum_j (q_{ii'}^A)^2/n_{ij}\right] \Big/ (sw_{i'}^A)^2, \tag{6.63}$$

$$\hat{\gamma}_{i'j'}^{AB} = \sum_i \sum_j q_{ii'}^A q_{jj'}^B \log n_{ij}/(w_{i'}^A w_{j'}^B), \tag{6.64}$$

and

$$s^2(\hat{\gamma}_{i'j'}^{AB}) = \left[\sum_i \sum_j (q_{ii'}^A)^2 (q_{jj'}^B)^2 / n_{ij}\right] \Big/ (w_{i'}^A w_{j'}^B)^2. \tag{6.65}$$

In each model, $\hat{\gamma}_2^A$, $\hat{\gamma}_3^A$, and $\hat{\gamma}_{11}^{AB}$ are the only estimates which differ from 0 by at least twice their EASD, and each of the corresponding standardized values $\hat{\gamma}_2^A/s(\hat{\gamma}_2^A)$, $\hat{\gamma}_3^A/s(\hat{\gamma}_3^A)$, and $\hat{\gamma}_{11}^{AB}/s(\hat{\gamma}_{11}^{1B})$ is at least 6 in absolute value. Thus no indication exists that a model is needed that is more complex than the model of linear-by-linear interaction, and no useful simplification of this model appears possible.

Regression Coefficients, Correlation Coefficients, and the Model of Linear-by-Linear Interaction

The model of linear-by-linear interaction leads to maximum likelihood equations that may be interpreted in terms of linear prediction. For subject h let $X_h = t_j$ if $B_h = j$, and let $Y_h = u_i$ if $A_h = i$. The sum of squares

$$\sum_h (Y_h - a - bX_h)^2$$

for linear prediction of Y_h from X_h is minimized if

$$b = \sum_h (X_h - \bar{X}) Y_h \Big/ \sum_h (X_h - \bar{X})^2$$

and

$$a = \bar{Y} - b\bar{X},$$

where

$$\bar{X} = \frac{1}{N} \sum X_h \qquad \text{and} \qquad \bar{Y} = \frac{1}{N} \sum Y_h.$$

In terms of the counts n_{ij},

$$b = \sum_i \sum_j u_i(t_j - \bar{X}) n_{ij} \Big/ \sum_j (t_j - \bar{X})^2 n_i^A,$$

$$\bar{X} = \frac{1}{N} \sum t_j n_j^B, \qquad \text{and} \qquad \bar{Y} = \frac{1}{N} \sum u_i n_i^A.$$

Given the X_h, $1 \le h \le N$, b has expected value

$$\beta = \sum_i \sum_j u_i(t_j - \bar{X}) m_{ij} \Big/ \sum_j (t_j - \bar{X})^2 n_j^B$$

and a has expected value

$$\alpha = \frac{1}{N}\sum t_j m_i^A - \beta \bar{X}.$$

Under the parametrization of the linear-by-linear model in terms of orthogonal polynomials, it follows that

$$\sum_i \sum_j u_i t_j \hat{m}_{ij} = \sum_i \sum_j u_i t_j n_{ij},$$

and

$$\sum_i u_i \hat{m}_i^A = \sum_i \sum_j u_i \hat{m}_{ij} = \sum_i \sum_j u_i n_{ij} = \sum_i u_i n_i^A.$$

Therefore, the maximum likelihood estimate

$$\hat{\beta} = \sum_i \sum_j u_i(t_j - \bar{X})\hat{m}_{ij} \Big/ \sum_j (t_j - \bar{X})^2 n_j^B$$

of the expected value β is equal to the regression coefficient b, and the maximum likelihood estimate

$$\hat{\alpha} = \frac{1}{N}\sum t_i \hat{m}_i^A - \hat{\beta}\bar{X}$$

of the expected value α is equal to the intercept a of the estimated regression line.

The estimate b also appears in residual analysis for the model of column homogeneity as

$$b = \sum_i \sum_j u_i t_j (n_{ij} - n_i^A n_j^B / N) \Big/ \sum_j (t_j - \bar{X})^2 n_j^B.$$

The corresponding adjusted residual is

$$t = \frac{b}{\hat{c}^{1/2}} = \frac{N^{1/2} \sum_i \sum_j u_i t_j (n_{ij} - n_i^A n_j^B / N)}{[\sum_i (u_i - \bar{Y})^2 n_i^A \sum_j (t_j - \bar{X})^2 n_j^B]^{1/2}},$$

where

$$\hat{c} = \frac{1}{N}\sum_i (u_i - \bar{Y})^2 n_i^A \Big/ \sum_j (t_j - \bar{X})^2 n_j^B.$$

Thus $t/N^{1/2}$ is the sample correlation coefficient for the pairs (X_h, Y_h), $1 \le h \le N$. As noted by Yates (1948) and Armitage (1955), among others, t has an approximate standard normal distribution if the column homogeneity model holds.

Other Techniques

Alternative approaches to ordered tables exist which do not involve log-linear models. Interested readers may consult Grizzle and Williams (1972), Simon (1974), and Bock (1975, pp. 541–559) for some recent treatments of the subject. Older discussions of interest include Kendall and Stuart (1967, pp. 562–578), Mantel (1963), and Williams (1952). The subject has been treated by methods involving correlation analyses since Yule (1900) and Pearson and Lee (1901). Thus the general problem of ordered classifications has a long history in the statistical literature.

6.2 Multinomial Response Models for Multi-Way Tables

Multinomial response models for three or more dichotomous or polytomous variables are direct generalizations of multinomial response models for one polytomous dependent variable and one polytomous independent variable. Since the number of models and the number of parameters to consider increases as the number of variables increases, analysis is more complex in the multi-way case than in the two-way case; however, the basic principles involved are unchanged. Possible multinomial response models will be illustrated in this section by a reanalysis of Table 6.14 (Table 4.6 of Volume 1) using models based on orthogonal scoring systems. The models considered here are of the same kind previously considered in Bock and Yates (1973), Bock (1970, 1975, pp. 528–538), and Haberman (1974a, pp. 213–214). As in Section 6.1, they include hierarchical log-linear models and logit models as special cases. The Newton–Raphson algorithm used in this section is very similar to the algorithm of Section 6.1. A special case has already been used in Section 3.4 of Volume 1. The program in the appendix can be used to implement this algorithm, although it lacks Bock and Yates' (1973) sophisticated procedures for automatic generation of models.

Scoring Systems and λ-Parameters

In hierarchical models, a set of λ-parameters with a common superscript is either set to 0 or completely unrestricted. Such a procedure does not permit any consideration of ordering of categories. Scoring systems for categories can eliminate this difficulty since they may be used to decompose all λ-parameters with a common superscript into components based on scores. Models can then be considered in which only selected components of these λ-parameters are set to 0.

Table 6.14

Attitudes toward Nontherapeutic Abortions among White, Christian Subjects in the 1972–1974 General Social Surveys[a]

Year	Religion[b]	Education in years	Attitudes[c] Positive	Mixed	Negative	Total
1972	North. Prot.	≤8	9	16	41	66
		9–12	85	52	105	242
		≥13	77	30	38	145
	South. Prot.	≤8	8	8	46	62
		9–12	35	29	54	118
		≥13	37	15	22	74
	Catholic	≤8	11	14	38	63
		9–12	47	35	115	197
		≥13	25	21	42	88
1973	North. Prot.	≤8	17	17	42	76
		9–12	102	38	84	224
		≥13	88	15	31	134
	South. Prot.	≤8	14	11	34	59
		9–12	61	30	59	150
		≥13	49	11	19	79
	Catholic	≤8	6	16	26	48
		9–12	60	29	108	197
		≥13	31	18	50	99
1974	North. Prot.	≤8	23	13	32	68
		9–12	106	50	88	244
		≥13	79	21	31	131
	South. Prot.	≤8	5	15	37	57
		9–12	38	39	54	131
		≥13	52	12	32	96
	Catholic	≤8	8	10	24	42
		9–12	65	39	89	193
		≥13	37	18	43	98

[a] Data tapes from 1972, 1973, and 1974 General Social Surveys, National Opinion Research Center, University of Chicago.

[b] Southern Protestants are Protestant respondents living in Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, Washington, D.C., or West Virginia.

[c] Responses to three questions are used to determine this variable. Subjects are asked, "Please tell me whether or not you think it should be possible for a pregnant woman to obtain a *legal* abortion if . . .," and six conditions are given. The following three are used in the table:

"B. If she is married and does not want any more children.

D. If the family has a very low income and cannot afford any more children.

F. If she is not married and does not want to marry the man."

Further details on these questions may be found in National Opinion Research Center (1974, p. 53). The attitude toward abortion is said to be positive if the subject answers "yes" to all three questions, the attitude is judged negative if the subject answers "no" to all three questions, and the attitude is said to be mixed if the subject answers "yes" to at least one question and "no" to at least one question. Subjects are excluded from tabulation if they do not answer "yes" or "no" to all three questions.

Table 6.15

Scores for Table 6.14

		First score		Second score	
Variable	Category	Name	Value	Name	Value
Year	1972	q^A_{11}	−1	q^A_{12}	1
	1973	q^A_{21}	0	q^A_{22}	−2
	1974	q^A_{31}	1	q^A_{32}	1
Religion	North. Prot.	q^B_{11}	1	q^B_{12}	1
	South. Prot.	q^B_{21}	1	q^B_{22}	−1
	Cath.	q^B_{31}	−2	q^B_{32}	0
Education in years	≤8	q^C_{11}	−1	q^C_{12}	1
	9 12	q^C_{21}	0	q^C_{22}	−2
	≥13	q^C_{31}	1	q^C_{32}	1
Attitudes	Positive	q^D_{11}	1	q^D_{12}	1
	Mixed	q^D_{21}	0	q^D_{32}	−2
	Negative	q^D_{31}	−1	q^D_{32}	1

As an illustration, consider the scores shown in Table 6.15. All four pairs of scores are orthogonal because

$$\sum_i a^A_{i1} q^A_{i2} = \sum_j q^B_{j1} q^B_{j2} = \sum_k q^C_{k1} q^C_{k2} = \sum_l q^D_{l1} q^D_{l2} = 0.$$

The scores for year, education, and attitudes are orthogonal-polynomial scores. For example, if $t^A_1 = -1$, $t^A_2 = 0$, and $t^A_3 = 1$, then

$$q^A_{i1} = t^A_i$$

and

$$q^A_{i2} = 3(t^A_i)^2 - 2.$$

In each case, the first score measures linear effects of the variable, and the second score measures quadratic effects. The scores for the remaining variable religion are termed Helmert contrasts by Bock (1975, pp. 258–260, 466). The scores q^B_{j1} compare all Protestants to Catholics, while the q^B_{j2} compare Northern and Southern Protestants without regard to Catholics. For a more complete discussion of scoring systems, see Bock and Yates (1973).

Use of these scores depends on the expression of λ-parameters for multiway tables in terms of scores. A λ^S-parameter can be expressed in terms of a sum of products of scores for variables represented by S. For example, for some unique γ^D_1 and γ^D_2,

$$\lambda^D_l = \gamma^D_1 q^D_{l1} + \gamma^D_2 q^D_{l2},$$

and for some unique γ_{11}^{CD}, γ_{21}^{CD}, γ_{12}^{CD}, and γ_{22}^{CD},

$$\lambda_{kl}^{CD} = \gamma_{11}^{CD}q_{k1}^{C}q_{l1}^{D} + \gamma_{21}^{CD}q_{k2}^{C}q_{l1}^{D} + \gamma_{12}^{CD}q_{k1}^{C}q_{l2}^{D} + \gamma_{22}^{CD}q_{k2}^{C}q_{l2}^{D}.$$

Similar results apply to other λ-parameters. The key feature is that a γ-parameter exists for each possible product of scores. The scores are helpful if λ-parameters can be expressed in terms of a limited number of γ-parameters, for fewer independent parameters need to estimated and interpretation of results may be simplified. The following examples illustrate some possible models for Table 4.6.

Model 1. The Additive Simultaneous Logit Model with Ordering Ignored

In this model, the conditional log odds $\tau_{ll'\cdot ijk}^{D\cdot ABC}$ is assumed additive in terms of the year $A_h = i$, the religion $B_h = j$ of the respondent, and the education $C_h = k$ of the respondent, where l and l' are possible values of the attitude variable D_h. Thus for some parameters $\eta_{ll'}$, $\eta_{ll'i}^{A}$, $\eta_{ll'j}^{B}$, and $\eta_{ll'k}^{C}$,

$$\tau_{ll'\cdot ijk}^{D\cdot ABC} = \eta_{ll'} + \eta_{ll'i}^{A} + \eta_{ll'j}^{B} + \eta_{ll'k}^{C}, \tag{6.66}$$

where

$$\sum_i \eta_{ll'i}^{A} = \sum_j \eta_{ll'j}^{B} = \sum_k \eta_{ll'k}^{C} = 0.$$

In terms of λ-parameters, the log odds $\tau_{ll'\cdot ijk}^{D\cdot ABC}$ satisfies the equation

$$\begin{aligned}\tau_{ll'\cdot ijk}^{D\cdot ABC} &= \log m_{ijkl} - \log m_{ijkl'} \\ &= (\lambda_l^D - \lambda_{l'}^D) + (\lambda_{il}^{AD} - \lambda_{il'}^{AD}) + (\lambda_{jl}^{BD} - \lambda_{jl'}^{BD}) \\ &\quad + (\lambda_{kl}^{CD} - \lambda_{kl'}^{CD}) + (\lambda_{ijl}^{ACD} - \lambda_{ijl'}^{ACD}) + (\lambda_{ikl}^{ABD} - \lambda_{ikl'}^{ACD}) \\ &\quad + (\lambda_{jkl}^{BCD} - \lambda_{jkl'}^{BCD}) + (\lambda_{ijkl}^{ABCD} - \lambda_{ijkl'}^{ABCD}).\end{aligned} \tag{6.67}$$

Thus for (6.66) to hold, it must be the case that

$$\eta_{ll'} = \lambda_l^D - \lambda_{l'}^D, \qquad \eta_{ll'i}^{A} = \lambda_{il}^{AD} - \lambda_{il'}^{AD},$$

$$\eta_{ll'j}^{B} = \lambda_{jl}^{BD} - \lambda_{jl'}^{BD}, \qquad \eta_{ll'k}^{C} = \lambda_{kl}^{CD} - \lambda_{kl'}^{CD},$$

and

$$\lambda_{ijl}^{ABD} = \lambda_{ikl}^{ACD} = \lambda_{jkl}^{BCD} = \lambda_{ijkl}^{ABCD} = 0.$$

The equivalent log-linear model to (6.66) is the hierarchical model

$$\begin{aligned}\log m_{ijkl} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} \\ &\quad + \lambda_{il}^{AD} + \lambda_{jk}^{BC} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD} + \lambda_{ijk}^{ABC} \\ &= \alpha_{ijk}^{ABC} + \lambda_l^D + \lambda_{il}^{AD} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD}.\end{aligned} \tag{6.68}$$

In terms of γ-parameters, (6.68) becomes

$$\log m_{ijkl} = \alpha^{ABC}_{ijk} + \sum_{l'} \gamma^{D}_{l'} q^{D}_{ll'} + \sum_{i'}\sum_{l'} \gamma^{AD}_{i'l'} q^{A}_{ii'} q^{D}_{ll'}$$

$$+ \sum_{j'}\sum_{l'} \gamma^{BD}_{j'l'} q^{B}_{jj'} q^{D}_{ll'} + \sum_{k'}\sum_{l'} \gamma^{CD}_{k'l'} q^{C}_{kk'} q^{D}_{ll'}$$

$$= \alpha^{ABC}_{ijk} + \sum_{c=1}^{14} \beta_c x_{ijklc} \tag{6.69}$$

where $\beta_1 = \gamma^D_1$, $\beta_2 = \gamma^D_2$, $\beta_3 = \gamma^{AD}_{11}$, $\beta_4 = \gamma^{AD}_{21}$, $\beta_5 = \gamma^{AD}_{12}$, $\beta_6 = \gamma^{AD}_{22}$, $\beta_7 = \gamma^{BD}_{11}$, $\beta_8 = \gamma^{BD}_{21}$, $\beta_9 = \gamma^{BD}_{12}$, $\beta_{10} = \gamma^{BD}_{22}$, $\beta_{11} = \gamma^{CD}_{11}$, $\beta_{12} = \gamma^{CD}_{21}$, $\beta_{13} = \gamma^{CD}_{12}$, $\beta_{14} = \gamma^{CD}_{22}$,

$x_{ijkl1} = q^D_{l1}$, $\quad x_{ijkl2} = q^D_{l2}$, $\quad x_{ijkl3} = q^A_{i1} q^D_{l1}$, $\quad x_{ijkl4} = q^A_{i2} q^D_{l1}$, $\quad$ etc.

To use the Newton–Raphson algorithm of Section 6.1, the table of counts n_{ijkl} may be transformed into a 3×27 table of counts $n^*_{i^*j^*}$, where

$$n^*_{i^*j^*} = n_{ijkl}, \qquad i^* = l, \quad j^* = i + 3(j-1) + 9(k+1).$$

Table 6.16

Parameter Estimates for Simultaneous Logit Models for Table 6.14

	Model 1		Model 2		Model 3	
Parameter	Estimate	EASD	Estimate	EASD	Estimate	EASD
γ^D_1	−0.184	0.025	−0.179	0.025	−0.178	0.025
γ^D_2	0.209	0.017	0.207	0.017	0.207	0.017
γ^{AD}_{11}	0.090	0.026	0.093	0.026	0.093	0.026
γ^{AD}_{21}	−0.033	0.014	−0.034	0.015	−0.034	0.014
γ^{AD}_{12}	0.008	0.018	0.007	0.017	0.007	0.018
γ^{AD}_{22}	−0.026	0.011	−0.027	0.011	−0.026	0.011
γ^{BD}_{11}	0.109	0.015	0.096	0.019	0.097	0.018
γ^{BD}_{21}	0.083	0.026	0.098	0.030	0.083	0.026
γ^{BD}_{12}	0.004	0.011	0.012	0.012	0.012	0.012
γ^{BD}_{22}	0.018	0.019	0.010	0.021	0.020	0.019
γ^{CD}_{11}	0.404	0.034	0.393	0.035	0.389	0.035
γ^{CD}_{21}	−0.025	0.015	−0.024	0.015	−0.023	0.015
γ^{CD}_{12}	0.077	0.023	0.076	0.024	0.073	0.023
γ^{CD}_{22}	−0.005	0.010	−0.004	0.010	−0.006	0.010
γ^{BCD}_{111}	—	—	0.090	0.026	0.088	0.026
γ^{BCD}_{211}	—	—	−0.041	0.042	-	—
γ^{BCD}_{121}	—	—	0.006	0.011	0.006	0.011
γ^{BCD}_{221}		—	0.012	0.018	—	—
γ^{BCD}_{112}	—	—	−0.004	0.017	−0.006	0.017
γ^{BCD}_{212}	—	—	−0.007	0.029	—	—
γ^{BCD}_{122}	—	—	0.0198	0.0074	0.0186	0.0074
γ^{BCD}_{222}	—	—	−0.0204	0.013	—	—

Thus $n_{11}^* = 9$, $n_{21}^* = 16$, $n_{31}^* = 41$, $n_{12}^* = 85$, $n_{14}^* = 8$, etc. If for $i^* = l$ and $j^* = i + 3(j-1) + 9(k-1)$,

$$\alpha_{j^*}^{B^*} = \alpha_{ijk}^{ABC}, \qquad m_{i^*j^*}^* = m_{ijkl}, \qquad \text{and} \qquad x_{i^*j^*c}^* = x_{ijklc},$$

then the mean $m_{i^*j^*}^*$ of $n_{i^*j^*}^*$ satisfies the log–linear model

$$\log m_{i^*j^*}^* = \alpha^{B^*} + \sum_{c=1}^{14} \beta_c x_{i^*j^*c}^*. \tag{6.70}$$

Using (6.70), the parameter estimates and estimated asymptotic standard deviations in Table 6.16 are found. The Pearson chi-square statistic X^2 is 57.6, while the likelihood-ratio chi-square L^2 is 57.7. There are $(3-1) \times 27 - 14 = 40$ degrees of freedom, so both chi-square statistics are significant at the 5 percent level. This lack of fit is not surprising since the λ^{BCD}-parameters have been assumed 0, even though they have been shown in Section 4.2 of Volume 1 not to be 0. Thus a model is needed which does not set the λ_{jkl}^{BCD} to 0.

Model 2. The Simultaneous Logit Model with an Interaction between Religion and Education

Here (6.66) is generalized to include an interaction of religion and education, so that

$$\tau_{ll'\cdot ijk}^{D\cdot ABC} = \eta_{ll'} + \eta_{ll'i}^{A} + \eta_{ll'j}^{B} + \eta_{ll'k}^{C} + \eta_{ll'jk}^{BC}, \tag{6.71}$$

where the constraints on the $\eta_{ll'i}^{A}$, $\eta_{ll'j}^{B}$, and $\eta_{ll'k}^{C}$ still apply and where

$$\sum_j \eta_{ll'jk}^{BC} = \sum_k \eta_{ll'jk}^{BC} = 0.$$

The equivalent log–linear model is the hierarchical model

$$\begin{aligned}\log m_{ijkl} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{il}^{AD} \\ &\quad + \lambda_{jk}^{BC} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD} + \lambda_{ijk}^{ABC} + \lambda_{jkl}^{BCD} \\ &= \alpha_{ijk}^{ABC} + \lambda_l^D + \lambda_{il}^{AD} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD} + \lambda_{jkl}^{BCD} \\ &= \alpha_{ijk}^{ABC} + \sum_{l'} \gamma_{l'}^{D} q_{ll'}^{D} + \sum_{i'}\sum_{l'} \gamma_{i'l'}^{AD} q_{ii'}^{A} q_{ll'}^{D} \\ &\quad + \sum_{j'}\sum_{l'} \gamma_{j'l'}^{BD} q_{jj'}^{B} q_{ll'}^{D} + \sum_{k'}\sum_{l'} \gamma_{k'l'}^{CD} q_{kk'}^{C} q_{ll'}^{D} + \sum_{j'}\sum_{k'}\sum_{l'} \gamma_{j'k'l'}^{BCD} q_{jj'}^{B} q_{kk'}^{C} q_{ll'}^{D}.\end{aligned} \tag{6.72}$$

Maximum likelihood estimates may be computed in the same way as in Model 1 through a transformation of the counts n_{ijkl} into a 3×27 table of counts $n_{i^*j^*}^*$. Maximum likelihood estimates are shown in Table 6.16. Since 22 γ-parameters appear in the model, there are $(3-1) \times 27 - 22 = 32$

degrees of freedom. The chi-square statistics are $X^2 = 29.2$ and $L^2 = 29.5$, so the fit is quite satisfactory. This result is quite predictable since this model is less restrictive than the model

$$\log m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{il}^{AD} + \lambda_{jk}^{BC} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD} + \lambda_{jkl}^{BCD} \tag{6.73}$$

considered in Section 4.2 of Volume 1. This latter model fits the data quite well, for X^2 here is 45.9, L^2 is 45.7, and there are 48 degrees of freedom. The difference in L^2 between the models specified by (6.72) and (6.73) is only 16.2, while the difference in degrees of freedom is 16. Thus (6.72) does not provide any obvious improvement over (6.73).

Model 3. A Simultaneous Logit Model without Interactions of Education with Type of Protestant

The interactions $\eta_{ll'jk}^{BC}$ of religion and education observed in the simultaneous logit model (6.71) appear to involve contrasts between Protestants and Catholics rather than between Northern and Southern Protestants. Under (6.72), the log cross-product ratio

$$\begin{aligned}
\tau_{(12)(ll')\cdot ik}^{BD\cdot AC} &= \tau_{ll'\cdot i1k}^{D\cdot ABC} - \tau_{ll'\cdot i2k}^{D\cdot ABC} \\
&= \log m_{i1kl} - \log m_{i1kl'} - \log m_{i2kl} + \log m_{i2kl'} \\
&= 2\gamma_{21}^{BD}(q_{l1}^D - q_{l'1}^D) + 2\gamma_{22}^{BD}(q_{l2}^D - q_{l'2}^D) \\
&= (\eta_{ll'1}^B - \eta_{ll'2}^B) + (\eta_{ll'1k}^{BC} - \eta_{ll'2k}^{BC}) \\
&= 2\gamma_{21}^{BD}(q_{l1}^D - q_{l'1}^D) + 2\gamma_{22}^{BD}(q_{l2}^D - q_{l'2}^D) \\
&\quad + 2\gamma_{211}^{BCD}q_{k1}^C(q_{l1}^D - q_{l'1}^D) + 2\gamma_{221}^{BCD}q_{k2}^C(q_{l1}^D - q_{l'1}^D) \\
&\quad + 2\gamma_{212}^{BCD}q_{k1}^C(q_{l2}^D - q_{l'2}^D) + 2\gamma_{222}^{BCD}q_{k2}^C(q_{l2}^D - q_{l'2}^D).
\end{aligned}$$

Controlling for year i and education k, this log cross-product ratio measures the difference between Northern and Southern Protestant respondents in the log odds that the attitude is l rather than l'. This cross-product ratio is independent of year i and education j if $\gamma_{211}^{BCD} = \gamma_{221}^{BCD} = \gamma_{212}^{BCD} = \gamma_{222}^{BCD} = 0$. The estimates of γ_{211}^{BCD}, γ_{221}^{BCD}, γ_{212}^{BCD}, and γ_{222}^{BCD} from Model 2 are relatively small compared to their respective estimated asymptotic standard deviations. Consequently, it is plausible that $\eta_{ll'1k}^{BC} = \eta_{ll'2k}^{BC}$ and

$$\begin{aligned}
\log m_{ijkl} = \alpha_{ijk}^{ABC} &+ \sum_{l'} \gamma_{l'}^D q_{ll'}^D + \sum_{i'}\sum_{l'} \gamma_{i'l'}^{AD} q_{ii'}^A q_{ll'}^D + \sum_{j'}\sum_{l'} \gamma_{j'l'}^{BD} q_{jj'}^B q_{ll'}^D \\
&+ \sum_{k'}\sum_{l'} \gamma_{k'l'}^{CD} q_{kk'}^C q_{ll'}^D + \sum_{k'}\sum_{l'} \gamma_{1k'l'}^{BCD} q_{j1}^B q_{kk'}^C q_{ll'}^D .
\end{aligned} \tag{6.74}$$

Under (6.74), one obtains the parameter estimates in the last two columns of Table 6.16. The new value of X^2 is 33.5, L^2 is now 34.0, and there are 36 degrees of freedom. Thus the fit remains satisfactory. This change in L^2 from Model 2 is 45.1, and the change in degrees of freedom is 4, so Model 3 appears to fit the data quite well relative to Model 2.

Model 4. Both (6.73) and (6.74) Hold

Since both (6.73) and (6.74) specify models which fit Table 6.14, it is plausible that the log-linear model in which both (6.73) and (6.74) hold may fit the table. Since (6.73) implies that

$$\begin{aligned}\log m_{ijkl} = {} & \alpha_i^A + \sum_{j'} \gamma_{j'}^B q_{jj'}^B + \sum_{k'} \gamma_{k'}^C q_{kk'}^C + \sum_{l'} \gamma_{l'}^D q_{ll'}^D \\ & + \sum_{i'}\sum_{l'} \gamma_{i'l'}^{AD} q_{ii'}^A q_{ll'}^D + \sum_{j'}\sum_{k'} \gamma_{j'k'}^{BC} q_{jj'}^B q_{kk'}^C + \sum_{j'}\sum_{l'} \gamma_{j'l'}^{BD} q_{jj'}^B q_{ll'}^D \\ & + \sum_{k'}\sum_{l'} \gamma_{k'l'}^{CD} q_{kk'}^C q_{ll'}^D + \sum_{j'}\sum_{k'}\sum_{l'} \gamma_{j'k'l'}^{BCD} q_{jj'}^B q_{kk'}^C q_{ll'}^D, \end{aligned} \tag{6.75}$$

where $\alpha_i^A = \lambda + \lambda_i^A$, and (6.74) assumes that $\gamma_{2k'l'}^{BCD} = 0$, $1 \le k' \le 2$, $1 \le l' \le 2$, the desired log-linear model has the form

$$\begin{aligned}\log m_{ijkl} = {} & \alpha_i^A + \sum_{j'} \gamma_{j'}^B q_{jj'}^B + \sum_{k'} \gamma_{k'}^C q_{kk'}^C + \sum_{l'} \gamma_{l'}^D q_{ll'}^D \\ & + \sum_{i'}\sum_{l'} \gamma_{i'l'}^{AD} q_{ii'}^A q_{ll'}^D + \sum_{j'}\sum_{k'} \gamma_{j'k'}^{BC} q_{jj'}^B q_{kk'}^C \\ & + \sum_{j'}\sum_{l'} \gamma_{j'l'}^{BD} q_{jj'}^B q_{ll'}^D + \sum_{k'}\sum_{l'} \gamma_{k'l'}^{CD} q_{kk'}^C q_{ll'}^D + \sum_{k'}\sum_{l'} \gamma_{1k'l'}^{BCD} q_{j1}^B q_{kk'}^C q_{ll'}^D. \end{aligned} \tag{6.76}$$

Equation (6.76) differs from similar equations for Models 1, 2, and 3 in the α-parameter. Earlier models have a term α_{ijk}^{ABC} which corresponds to the independent variables A_h, B_h, and C_h. The γ-parameters in these earlier models always include a superscript D corresponding to the dependent variable D_h. The term α_i^A in (6.76) corresponds to the single independent variable A_h. The variables B_h, C_h, and D_h can be regarded as dependent variables, i.e., the observed religion B_h, education C_h, and attitudes D_h of respondent depend on the year A_h of the survey. In keeping with this new division into independent and dependent variables, all γ-parameters in (6.76) include a superscript B, C, or D.

On way to use the Newton–Raphson algorithm of Section 6.1 with (6.76) involves transformation of the counts n_{ijkl} into a 27×3 table of counts $n_{i^*j^*}^*$, where

$$n_{i^*j^*}^* = n_{ijkl}, \qquad i^* = 9(j-1) + 3(k-1) + l, \qquad j^* = i.$$

Thus $n_{11}^* = 9$, $n_{21}^* = 16$, $n_{41}^* = 85$, $n_{12}^* = 17$, $n_{22}^* = 17$, etc. In terms of the mean $m_{i^*j^*}^*$ of $n_{i^*j^*}^*$, the model may be written

$$\log m_{i^*j^*}^* = \alpha_{j^*}^{B^*} + \sum_{k^*=1}^{26} \beta_{k^*} x_{i^*j^*k^*}^*, \tag{6.77}$$

where the β_{k^*} and $x_{i^*j^*k^*}^*$ are defined as in Table 6.17. Using this model, the parameter estimates in Table 6.18 are found. The Pearson chi-square is 50.4, the likelihood-ratio chi-square is 50.2, and there are $(27 - 1)(3) - 26 = 52$ degrees of freedom, so that the fit is quite good.

The reduction in L^2 from the conditional independence model of (6.73) is 4.53, while the change in degrees of freedom is 4. The reduction in L^2 from

Table 6.17

Coefficients and Parameters for Model 4 for Table 6.14[a]

k^*	γ-parameter equal to β_{k^*}	$x_{11k^*}^*$	$x_{21k^*}^*$	$x_{31k^*}^*$	$x_{41k^*}^*$	$x_{51k^*}^*$	$x_{61k^*}^*$
1	γ_1^B	1	1	1	1	1	1
2	γ_2^B	1	1	1	1	1	1
3	γ_1^C	−1	−1	−1	0	0	0
4	γ_2^C	1	1	1	−2	−2	−2
5	γ_1^D	1	0	−1	1	0	−1
6	γ_2^D	1	−2	1	1	−2	1
7	γ_{11}^{AD}	−1	0	1	−1	0	1
8	γ_{21}^{AD}	1	0	−1	1	0	−1
9	γ_{12}^{AD}	−1	2	−1	−1	2	−1
10	γ_{22}^{AD}	1	−2	1	1	−2	1
11	γ_{11}^{BC}	−1	−1	−1	0	0	0
12	γ_{21}^{BC}	−1	−1	−1	0	0	0
13	γ_{12}^{BC}	1	1	1	−2	−2	−2
14	γ_{22}^{BC}	1	1	1	−2	−2	−2
15	γ_{11}^{BD}	1	0	−1	1	0	−1
16	γ_{21}^{BD}	1	0	−1	1	0	−1
17	γ_{12}^{BD}	1	−2	1	1	−2	1
18	γ_{22}^{BD}	1	−2	1	1	−2	1
19	γ_{11}^{CD}	−1	0	1	0	0	0
20	γ_{21}^{CD}	1	0	−1	−2	0	2
21	γ_{12}^{CD}	−1	2	−1	0	0	0
22	γ_{22}^{CD}	1	−2	1	−2	4	−2
23	γ_{111}^{BCD}	−1	0	1	0	0	0
24	γ_{121}^{BCD}	1	0	−1	−2	0	2
25	γ_{112}^{BCD}	−1	2	−1	0	0	0
26	γ_{122}^{BCD}	1	−2	1	−2	4	−2

[a] Space requirements limit presentation of the $x_{i^*j^*k^*}^*$ to $x_{11k^*}^*$, $x_{21k^*}^*$, $x_{31k^*}^*$, $x_{41k^*}^*$, $x_{51k^*}^*$, $x_{61k^*}^*$.

Table 6.18

Estimated γ-Parameters for Model 4 for Table 6.14

Parameter	Maximum likelihood estimate	Estimated asymptotic standard deviation	Standardized value
γ_1^B	0.025	0.016	1.54
γ_2^B	0.203	0.025	8.02
γ_1^C	0.287	0.031	9.30
γ_2^C	−0.307	0.014	−22.69
γ_1^D	−0.179	0.025	−7.02
γ_2^D	0.207	0.017	12.13
γ_{11}^{AD}	0.091	0.025	3.70
γ_{21}^{AD}	−0.028	0.014	−1.99
γ_{12}^{AD}	0.007	0.018	0.40
γ_{22}^{AD}	−0.025	0.011	−2.32
γ_{11}^{BC}	−0.031	0.022	−1.40
γ_{21}^{BC}	0.054	0.034	1.60
γ_{12}^{BC}	0.018	0.010	1.89
γ_{22}^{BC}	−0.041	0.015	2.71
γ_{11}^{BD}	0.098	0.018	5.30
γ_{21}^{BD}	0.079	0.026	3.07
γ_{12}^{BD}	0.012	0.012	1.01
γ_{22}^{BD}	0.018	0.019	0.98
γ_{11}^{CD}	0.389	0.035	11.21
γ_{21}^{CD}	−0.024	0.015	−1.61
γ_{12}^{CD}	0.072	0.023	3.11
γ_{22}^{CD}	−0.007	0.010	−0.64
γ_{111}^{BCD}	0.084	0.026	3.26
γ_{121}^{BCD}	0.006	0.011	0.60
γ_{112}^{BCD}	−0.007	0.017	−0.42
γ_{122}^{BCD}	0.086	0.0074	2.52

the model of (6.74) is 16.3, while there is a decrease of 16 in degrees of freedom. Thus no evidence exists that (6.76) fits the data worse than (6.73) or (6.74).

Other Models

Inspection of Table 6.18 shows that γ_2^B, γ_1^C, γ_2^C, γ_1^D, γ_2^D, γ_{11}^{AD}, γ_{11}^{BD}, γ_{21}^{BD}, γ_{11}^{CD}, γ_{12}^{CD}, and γ_{111}^{BCD} all have corresponding standardized values at least 3 in magnitude. The standardized values for λ_{21}^{AD}, λ_{22}^{AD}, λ_{22}^{BC}, and λ_{122}^{BCD} have magnitudes of at least 1.99.

Thus a substantial fraction of the γ-parameters in Table 6.7 are clearly not 0. Nonetheless, modest simplifications appear possible. For example, one can consider a model in which

$$\log m_{ijkl} = \alpha_i^A + \sum_{j'=1}^{2} \gamma_{j'}^B q_{jj'}^B + \sum_{k'=1}^{2} \gamma_{k'}^C q_{kk'}^C + \sum_{l'=1}^{2} \gamma_{l'}^D q_{l'}^D$$
$$+ \sum_{i'=1}^{2} \sum_{l'=1}^{2} \gamma_{i'l'}^{AD} q_{ii'}^A q_{ll'}^D + \gamma_{11}^{BC} q_{j1}^B q_{k1}^C + \gamma_{11}^{BD} q_{j1}^B q_{k1}^D$$
$$+ \gamma_{21}^{BD} q_{j2}^B q_{k1}^D + \gamma_{11}^{CD} q_{k1}^C q_{l1}^D + \gamma_{12}^{CD} q_{k1}^C q_{l2}^D + \gamma_{111}^{BCD} q_{j1}^B q_{k1}^C q_{l1}^D.$$

This model leads to a Pearson chi-square statistic $X^2 = 73.1$ and a likelihood-ratio chi-square statistic $L^2 = 73.5$. There are $3(27 - 1) - 16 = 62$ degrees of freedom, so that these statistics have significance levels of about 15 percent. Thus this model does not provide a bad overall fit, although it is noticeably less successful than Model 4. A decrease in L^2 of 23.3 has been accompanied by a decrease in 10 in the degrees of freedom. This decrease is significant at approximately the 0.01 level. An alternate model which fares better assumes that

$$\log m_{ijkl} = \alpha_i^A + \sum_{j'=1}^{2} \gamma_{j'}^B q_{jj'}^B + \sum_{k'=1}^{2} \gamma_{k'}^C q_{kk'}^C + \sum_{l'=1}^{2} \gamma_{l'}^D q_{ll'}^D$$
$$+ \sum_{i'=1}^{2} \sum_{l'=1}^{2} \gamma_{i'l'}^{AD} q_{ii'}^A q_{ll'}^D + \gamma_{11}^{BC} q_{j1}^B q_{k1}^C + \gamma_{12}^{BC} q_{j1}^B q_{k2}^C + \gamma_{11}^{BD} q_{j1}^B q_{l1}^D$$
$$+ \gamma_{21}^{BD} q_{j2}^B q_{k1}^D + \gamma_{12}^{BD} q_{j1}^B q_{l2}^D + \sum_{k'=1}^{2} \sum_{l'=1}^{2} \gamma_{k'l'}^{CD} q_{kk'}^C q_{ll'}^D$$
$$+ \gamma_{111}^{BCD} q_{j1}^B q_{k1}^C q_{l1}^D + \gamma_{122}^{BCD} q_{j1}^B q_{k2}^C q_{l2}^D.$$

Here $X^2 = 60.2$, $L^2 = 60.1$, and the degrees of freedom are 57. The change in L^2 from Model 4 is now 9.92, while the change in degrees of freedom is 5. This change in L^2 is now only significant at the 0.08 level.

Interpretation of γ-Parameters

To illustrate interpretation of γ-parameters, consider the parameters γ_{11}^{AD} and γ_{111}^{BCD} in Model 4. Since the scores $q_{ii'}^A$, $1 \le l' \le 2$, and the scores $q_{ll'}^D$, $1 \le l' \le 2$, are orthogonal, one may write

$$\gamma_{11}^{AD} = \left(\frac{1}{2 \cdot 2 \cdot 3 \cdot 3}\right) \sum_i \sum_j \sum_k \sum_l q_{i1}^A q_{l1}^D \log m_{ijkl}.$$

Formulas of this type may be found in Good (1958), among others. Here the product $2 \cdot 2 \cdot 3 \cdot 3$ arises since variables B_h and C_h each have three categories and since

$$\sum_i (q_{i1}^A)^2 = (-1)^2 + 0^2 + 1^2 = 2$$

and

$$\sum_l (q_{l1}^D)^2 = 1^2 + 0^2 + (-1)^2 = 2.$$

Note that for fixed j and k,

$$\sum_i \sum_l q_{i1}^A q_{l1}^D \log m_{ijkl} = -\log m_{11kl} + \log m_{31kl} + \log m_{13kl} - \log m_{33kl}$$
$$= \tau_{(31)(13)\cdot jk}^{AD\cdot BC}.$$

Thus $4\gamma_{11}^{AD}$ is the average log cross-product ratio

$$\tau_{(31)(13)\cdot jk}^{AD\cdot BC}$$

for $1 \leq j \leq 3, 1 \leq k \leq 3$. Under the model under study, $\tau_{(31)(13)\cdot jk}^{AD\cdot BC}$ is constant for each j and k. This statement holds since $\tau_{(31)(13)\cdot jk}^{AD\cdot BC}$ is constant under the hierarchical model with generating class AD and BCD. The present model is more restrictive than this hierarchical model, so $\tau_{(31)(13)\cdot jk}^{AD\cdot BC}$ is surely independent of j and k. Given that orthogonal scores are present, the result will hold if $\gamma_{1j'1}^{ABD} = 0$, $1 \leq j' \leq 2$, $\gamma_{1k'1}^{ACD} = 0$, $1 \leq k' \leq 2$, and $\gamma_{1k'l'1}^{ABCD} = 0$, $1 \leq k' \leq 2$, $1 \leq l' \leq 2$, even if other γ-parameters are not 0.

The maximum likelihood estimate

$$\hat{\gamma}_{11}^{AD} = 0.091 \qquad \text{and} \qquad s(\hat{\gamma}_{11}^{AD}) = 0.025.$$

It can be shown that

$$\hat{\gamma}_{11}^{AD} = \frac{1}{4}\log\left(\frac{n_{31}^{AD}n_{13}^{AD}}{n_{11}^{AD}n_{33}^{AD}}\right) \qquad \text{and} \qquad s^2(\hat{\gamma}_{11}^{AD}) = \frac{1}{4}\left(\frac{1}{n_{31}^{AD}} + \frac{1}{n_{11}^{AD}} + \frac{1}{n_{13}^{AD}} + \frac{1}{n_{33}^{AD}}\right),$$

so that $\hat{\gamma}_{11}^{AD}$ corresponds to $-\frac{1}{4}\hat{\tau}_{(13)(13)\cdot jk}^{AD\cdot BC}$ in Table 4.11. Consequently, $4\hat{\gamma}_{11}^{AD}$ measures the change in the relative log odds of favorable and unfavorable attitudes toward nontherapeutic abortions from 1972 to 1974. The estimated percentage increase in the odds is

$$100(\exp 4\hat{\gamma}_{11}^{AD} - 1) = 44.1 \text{ percent}.$$

This increase is the same under the model for all religious and educational groups.

In the case of γ_{111}^{BCD}, one may write

$$\gamma_{111}^{BCD} = \frac{1}{6\cdot 2\cdot 2\cdot 3}\sum_i \sum_j \sum_k \sum_l q_{j1}^B q_{k1}^C q_{l1}^D \log m_{ijkl}.$$

Here, the 6 in the denominator corresponds to

$$\sum_j (q_{j1}^B)^2 = 1^2 + 1^2 + (-2)^2 = 6,$$

the next two 2's correspond to

$$\sum_k (q_{k1}^C)^2 = (-1)^2 + 0^2 + 1^2 = 2 \qquad \text{and} \qquad \sum_l (q_{l1}^D)^2 = 2,$$

and the 3 corresponds to the number of categories in variable A_h. A relatively straightforward argument shows that

$$\gamma_{111}^{BCD} = \frac{1}{3} \sum_i \frac{1}{24} (\tau_{(13)(31)(13)\cdot i}^{BCD\cdot A} + \tau_{(23)(31)(13)\cdot i}^{BCD\cdot A}).$$

Since $\tau_{(13)(31)(13)\cdot i}^{BCD\cdot A}$ and $\tau_{(23)(31)(13)\cdot i}^{BCD\cdot A}$ are independent of i, γ_{111}^{BCD} is $\frac{1}{24}$ times the constant sum

$$\tau_{(13)(31)(13)\cdot i}^{BCD\cdot A} + \tau_{(23)(31)(13)\cdot i}^{BCD\cdot A}.$$

Since γ_{211}^{BCD} is assumed 0, it also follows that

$$\begin{aligned} 0 &= \frac{1}{2\cdot 2\cdot 2\cdot 3} \sum_i \sum_j \sum_k \sum_l q_{j2}^B q_{k1}^C q_{l1}^D \log m_{ijkl} \\ &= \frac{1}{3} \sum_i \tau_{(12)(31)(13)\cdot i}^{BCD\cdot A}, \end{aligned}$$

where $\tau_{(12)(31)(13)\cdot i}^{BCD\cdot A}$ is independent of i. Since

$$\tau_{(12)(31)(13)\cdot i}^{BCD\cdot A} = \tau_{(31)(13)\cdot i1}^{CD\cdot AB} - \tau_{(31)(13)\cdot i2}^{CD\cdot AB},$$

the log cross-product ratios

$$\tau_{(31)(13)\cdot ij}^{CD\cdot AB} = \log\left(\frac{p_{1\cdot ij3}^{D\cdot ABC}\; p_{3\cdot ij1}^{D\cdot ABC}}{p_{3\cdot ij3}^{D\cdot ABC}\; p_{1\cdot ij3}^{D\cdot ABC}}\right)$$

for education and attitudes given year and religion are constant for all years i and for both Protestant groups ($j = 1$ or 2). In addition, γ_{111}^{BCD} is $\frac{1}{12}$ times the constant difference

$$\tau_{(j3)(31)(13)\cdot i}^{BCD\cdot A} = \tau_{(31)(13)\cdot ij}^{CD\cdot AB} - \tau_{(31)(13)\cdot i3}^{CD\cdot AB}$$

in log cross-product ratios of education and attitudes given year and religion between Protestant groups ($j = 1$ or 2) and Catholics. Since $\hat{\gamma}_{111}^{BCD} = 0.084$ and $s(\hat{\gamma}_{111}^{BCD}) = 0.026$, the difference in log cross-product ratios appears to be substantial, because

$$\hat{\tau}_{(j3)(31)(13)\cdot i}^{BCD\cdot A} = 12\hat{\gamma}_{111}^{BCD} = 1.01$$

and

$$s(\hat{\tau}_{(j3)(31)(13)\cdot i}^{BCD\cdot A}) = 12s(\hat{\gamma}_{111}^{BCD}) = 0.31.$$

Thus whether a respondent is Protestant or Catholic has a substantial effect on the interaction between education and attitudes when the comparison involves the lowest and highest educational groups and the least and most favorable attitudes toward abortion. The proposed model claims that this interaction is not affected by whether the respondent is a Northern or Southern Protestant.

For other examples of relationships between γ-parameters and τ-parameters, see Exercise 6.6.

This section illustrates use of orthogonal scores with contingency tables involving at least three variables. Such scores may be used in hierarchical log-linear models such as Models 1 and 2 where the model itself does not exploit the structure of the categories of variables under study, in simultaneous logit models such as Model 3 in which the structure of the categories of the independent variables is used to simplify the model, or in models such as Model 4, in which the structure of the categories of the dependent variables is used to simplify the model. There may be three independent variables, as in Models 1, 2, and 3, or one independent variable, as in Model 4. As in Models 1, 2, and 3, there may be one dependent variable, however, as in Model 4, there may be three dependent variables. Thus orthogonal scores may be employed in a very large variety of circumstances.

6.3 Multinomial Response Models for One or More Continuous Independent Variables

If one or more independent variables are continuous, then multinomial response models may still be employed; however, chi-square approximations for test statistics and normal approximations for residuals only apply in limited circumstances. The problems here are very similar to those in logit analysis. The limited literature concerning this case includes Haberman (1974a, pp. 352–373) and Nerlove and Press (1973), among other references. This section provides two methods of analysis to illustrate procedures appropriate for the sparse tables which result from the presence of at least one continuous independent variable. Both examples explore the relationship of attitudes toward women's roles to the sex and education of the respondent. One example describes attitudes by a scale with values from -2 to 2, while the other example describes attitudes in terms of four dichotomous dependent variables. Both analyses apply to Table 6.19. In this six-way table, 1330 subjects in the 1975 General Social Survey are classified by sex, education, and responses to four questions concerning women's roles. Responses are arranged so that a label of " + " corresponds to response more favorable to a public role for women.

This table is difficult to analyze due to its size and its sparseness, for there are 672 cells for only 1330 observations. Two approaches to this problem are considered in this section. In the first approach, a summary statistic is constructed for the four dichotomous responses, and a multinomial response

Table 6.19

Responses of Subjects in 1975 General Social Survey to Four Questions on Women's Roles—Classification by Education and Sex[a]

Sex	Years of education completed	Response + + + +	+ + + −	+ + − +	+ + − −	+ − + +	+ − + −	+ − − +	+ − − −	− + + +	− + + −	− + − +	− + − −	− − + +	− − + −	− − − +	− − − −	Total
Male	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	3
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
	2	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	3
	3	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	3
	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	5	0	2	0	0	2	2	1	0	1	1	0	0	1	1	1	1	13
	6	3	0	0	0	2	0	0	0	1	0	0	3	2	2	0	3	16
	7	2	0	1	0	3	0	0	0	0	0	0	1	1	3	0	5	16
	8	5	2	0	0	5	1	0	1	0	7	0	2	4	7	1	9	44
	9	7	1	0	2	1	0	1	0	0	3	0	3	0	2	0	3	23
	10	11	2	0	2	3	0	2	1	1	3	0	0	0	4	0	3	32
	11	16	3	1	0	5	6	0	0	4	3	0	3	0	1	0	5	47
	12	60	27	0	4	20	5	2	1	5	19	2	8	4	8	0	12	177
	13	17	11	2	1	3	0	0	0	2	6	0	2	0	3	0	2	49
	14	20	8	0	1	5	0	0	0	0	3	0	3	1	3	0	1	45
	15	8	2	0	2	1	1	0	0	1	2	0	1	0	0	0	0	18
	16	42	8	1	0	3	0	1	0	0	7	0	0	1	0	0	2	65
	17	4	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	7
	18	9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	11
	19	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
	20	4	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	7
	Total:	215	73	7	13	57	15	7	3	15	55	3	29	14	35	2	47	590

Female	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	4
	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	3
	4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	3
	5	4	0	0	0	0	1	0	0	0	1	0	1	1	0	0	2	10
	6	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	4	7
	7	5	1	0	0	2	1	0	0	0	0	1	2	0	4	0	7	23
	8	9	2	0	1	5	2	0	0	1	5	2	2	2	14	0	13	58
	9	11	4	1	1	2	2	0	0	0	5	0	2	0	0	0	5	33
	10	19	6	1	0	6	0	0	1	0	6	0	3	1	6	2	7	58
	11	17	9	0	1	6	3	0	0	2	6	1	6	1	5	0	3	60
	12	97	44	1	9	17	7	3	2	8	23	0	22	5	21	0	13	272
	13	30	12	1	0	2	0	0	1	0	1	0	3	0	0	0	3	53
	14	23	13	1	0	2	0	1	1	1	2	0	1	0	1	0	4	50
	15	10	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	14
	16	42	6	1	3	0	2	0	0	0	0	0	4	0	0	0	1	59
	17	11	0	0	2	0	2	0	0	0	0	0	1	0	0	0	0	16
	18	8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
	19	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
	20	2	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	4
	Total:	290	101	6	17	42	20	4	5	13	54	4	48	10	54	4	68	740
Total		505	174	13	30	99	35	11	8	28	109	7	77	24	89	6	115	1330

[a] The four questions, which are found in National Opinion Research Center (1975, pp. 551–52), are

A. "Do you agree or disagree with this statement? — Women should take care of running their homes and leaving running the country up to men.

B. Do you approve or disapprove of a married woman earning money in business or industry if she has a husband capable of supporting her?

C. If your party nominated a woman for President, would you vote for her if she were qualified for the job?

D. Tell me if you agree or disagree with this statement: Most men are better suited emotionally for politics than are most women."

The label "−" corresponds to the following responses: "agree" in A, "disapprove" in B, "no" in C, and "yes" in D. The label "+" corresponds to "disagree" in A, "approve" in B, "yes" in C, and "no" in D. The notation "+ + − −" corresponds to a subject responding "disagree" in A, "approve" in B, "no" in C, and "yes" in D. Subjects are only tabulated if responses corresponding to "+" or "−" are given to all questions and if years of education completed is reported. In all, 1330 of 1490 subjects are tabulated.

From data tape from 1975 General Social Survey, National Opinion Research Center, University of Chicago.

model is constructed for the summary statistic. This approach reduces the size of the table to be examined; however, some information may be lost concerning the relationship between the independent and dependent variables. Residual analysis may be used to detect such losses of information.

In the second approach, models based on orthogonal scores are developed as in Section 6.2. Standardized parameter estimates are used to simplify models. Results of model simplification are judged in terms of the reduction achieved in the likelihood-ratio chi-square statistic. This method of analysis is expensive in terms of computer time; however, this approach does have the virtue of great generality.

Analysis by Number of Positive Responses

A simple summary statistic for the responses in Table 6.19 is the number of positive responses. Using this statistic, Table 6.20 is obtained. Observe that

Table 6.20

Responses of Subjects in 1975 General Social Survey to Four Questions on Women's Roles—Classification by Number of Positive Responses, Education, and Sex[a]

Sex	Years of education completed	Number of positive responses				
		0	1	2	3	4
Male	0	1	0	0	2	0
	1	0	1	0	0	0
	2	0	2	1	0	0
	3	0	1	0	1	1
	4	0	0	0	0	3
	5	1	2	5	5	0
	6	3	5	2	3	3
	7	5	4	1	4	2
	8	9	11	12	7	5
	9	3	5	6	2	7
	10	3	5	7	6	11
	11	5	4	9	13	16
	12	12	17	36	52	60
	13	2	5	7	18	17
	14	1	6	5	13	20
	15	0	1	5	4	8
	16	2	0	9	12	42
	17	0	0	1	2	4
	18	0	0	0	2	9
	19	0	0	0	4	3
	20	0	0	1	2	4

Table 6.20 (continued)

Sex	Years of education completed	Number of positive responses 0	1	2	3	4
Female	0	3	1	0	0	0
	1	0	1	0	0	0
	2	0	0	0	0	0
	3	1	1	1	0	0
	4	2	0	1	0	0
	5	2	1	3	0	4
	6	4	2	1	0	0
	7	7	6	2	3	5
	8	13	16	12	8	9
	9	5	2	8	7	11
	10	7	12	7	13	19
	11	3	11	12	17	17
	12	13	45	47	70	97
	13	3	4	1	15	30
	14	4	3	3	17	23
	15	0	0	2	2	10
	16	1	4	5	7	42
	17	0	1	4	0	11
	18	0	0	0	2	8
	19	0	0	0	0	2
	20	0	1	0	1	2

[a] Computed from Table 6.19.

four positive responses in Table 6.20 correspond to the classification "+ + + +" in Table 6.19, while three positive responses correspond to the classifications "+ + + −," "+ + − +," "+ − + +," and "− + + +" in Table 6.19.

To construct a multinomial response model for Table 6.19, let subject h, $1 \le h \le N = 1330$ have Y_h years of education, sex S_h (1 for male and 2 for female), and $K_h - 1$ positive responses. Thus $1 \le S_h \le 2$, $0 \le Y_h \le 20$, and $1 \le K_h \le 5$.

Assume that the probability $P^K_{k \cdot h} > 0$ that $K_h = k$ depends only on sex S_h and the education Y_h of subject h. One simple multinomial response model for these data is a generalization of the linear-by-linear model of Section 6.1. Let $T_{h1} = 3 - 2S_h$ be a sex score for subject h, and let $T_{h2} = Y_h - 10$ be an education score. These scores are chosen so that their range is symmetric about 0. In the case of T_{h1}, $T_{h1} = 1$ for males and $T_{h1} = -1$ for females. In the case of T_{h2}, T_{h2} ranges from -10 in the case of no completed years of schooling to 10 in the case of 20 years of schooling. Assume that the log odds

$$\Omega^K_{kk' \cdot h} = \log(P^K_{k \cdot h}/P^K_{k' \cdot h})$$

for $K_h = k$ rather than $K_h = k'$ is a linear function

$$\Omega^K_{kk'\cdot h} = \eta_{kk'} + \xi_{kk'1} T_{h1} + \xi_{kk'2} T_{h2} \tag{6.78}$$

of the sex score T_{h1} and the education score T_{h2}. As in Section 6.1, assume that for some ξ_1 and ξ_2,

$$\xi_{kk'1} = (k' - k)\xi_1 \tag{6.79}$$

and

$$\xi_{kk'2} = (k' - k)\xi_2. \tag{6.80}$$

To describe this model in the terms generally considered with log-linear models, form a $5 \times N$ table of counts N^K_{kh}, $1 \le k \le 5$, $1 \le h \le N$, so that

$$\begin{aligned} N^K_{kh} &= 1, \qquad K_h = k, \\ &= 0, \qquad K_h \ne k. \end{aligned}$$

Then each column N^K_{kh}, $1 \le k \le 5$, corresponding to an individual subject h has an independent multinomial distribution with sample size 1 and probabilities $P^K_{k\cdot h}$, $1 \le k \le 5$. The mean of N^K_{kh} is $P^K_{k\cdot h}$. As in Section 6.1, an equivalent version of (6.78), (6.79), and (6.80) is the log-linear model

$$\begin{aligned} \log P^K_{k\cdot h} &= \alpha^K_h - \frac{1}{5}\sum_{l=1}^{k-1} l\eta_{l(l+1)} + \frac{1}{5}\sum_{l=k}^{4}(5 - l)\eta_{l(l+1)} \\ &\quad + \xi_1(3 - k)T_{h1} + \xi_2(3 - k)T_{h2} \\ &= \alpha^K_h + \sum_{l=1}^{6} \beta_l X^k_{khl}. \end{aligned} \tag{6.81}$$

Here $\beta_l = \eta_{l(l+1)}$, $1 \le l \le 4$, $\beta_5 = \xi_1$, $\beta_6 = \xi_2$, and

$$\begin{aligned} X^K_{khl} &= 1 - k/5, \qquad && k \le l \le 4, \\ &= -k/5, && l < k \le 5, \\ &= (3 - k)T_{h1}, && l = 5, \\ &= (3 - k)T_{h2}, && l = 6. \end{aligned}$$

Computation of Maximum Likelihood Estimates

Maximum likelihood estimates may be obtained by the same Newton–Raphson algorithm used in the rest of this chapter. The algorithm may be applied to the counts N^K_{kh}, just as for any two-way table; however, two complications do arise. The first difficulty involves cost of computations. There are 6950 probabilities $P^K_{k\cdot h}$, 6950 counts N^K_{kh}, and 41,700 variables X^K_{khl}. Consequently, computation is expensive, both in terms of computer time and in terms of computer storage requirements. The second difficulty involves starting the algorithm with an initial approximation $P^K_{k\cdot h0}$ for $\hat{P}^K_{k\cdot h}$. The usual procedure for selection of starting values leads to

$$\begin{aligned} P^K_{k\cdot h0} &= \tfrac{3}{7}, & K_h &= k, \\ &= \tfrac{1}{7}, & K_h &\neq k. \end{aligned}$$

A starting value of this sort is acceptable, however, convergence is slower than when the algorithm is applied to tables in which cell counts are large.

In this example, computations are much less expensive if based on the table of counts n_{syk}. Here there are n_{syk} subjects h with sex $S_h = s$, education $Y_h = y$, and response $K_h = k$. The probability is $p^{K\cdot SY}_{k\cdot sy}$ that $K_h = k$ if $S_h = s$ and $Y_h = y$. Thus $p^{K\cdot SY}_{k\cdot sy} = P^K_{k\cdot h}$ if $S_h = s$ and $Y_h = y$. The mean of n^{SYK}_{syk} is $n^{SYK}_{syk} = n^{SY}_{sy} p^{K\cdot SY}_{k\cdot sy}$, where n^{SY}_{sy} subjects have sex $S_h = s$ and education $Y_h = y$. If $n^{SY}_{sy} > 0$, then

$$\log m^{SYK}_{syk} = \alpha^{SY}_{sy} + \sum_{l=1}^{6} \beta_l x^{SK}_{sykl}, \tag{6.82}$$

where the β_l are defined as in (6.81) and

$$\begin{aligned} x^{SYK}_{sykl} &= 1 - k/5, & k &\leq l \leq 4, \\ &= -k/5, & l &< k \leq 5, \\ &= (3-k)(3-2s), & l &= 5, \\ &= (3-k)(y-10), & l &= 6. \end{aligned}$$

Using (6.82), we obtain the results shown in Tables 6.21 and 6.22.

Table 6.21

Parameter Estimates for Linear-by-Linear Log-Linear Model for Table 6.20.

Parameter	Estimate	EASD
$\beta_1 = \eta_{12}$	−0.472	0.121
$\beta_2 = \eta_{23}$	−0.101	0.101
$\beta_3 = \eta_{34}$	−0.203	0.090
$\beta_4 = \eta_{45}$	−0.221	0.076
$\beta_5 = \xi_1$	−0.010	0.023
$\beta_6 = \xi_2$	−0.1031	0.0082

Table 6.22

Observed and Estimated Expected Average Number of Responses for Linear-by-Linear Model for Table 6.20

	Male			Female		
Education in years	Observed average	Estimated expected average	Adjusted residual	Observed average	Estimated expected average	Adjusted residual
0–8	1.90	1.80	0.96	1.45	1.65	−2.11
9	2.48	2.36	0.41	2.55	2.31	0.99
10	2.53	2.45	0.36	2.43	2.41	0.12
11	2.66	2.63	0.16	2.57	2.60	−0.18
12	2.74	2.80	−0.77	2.71	2.77	−0.96
13	2.88	2.95	−0.46	3.23	2.92	1.92
14	3.00	3.09	−0.56	3.04	3.06	−0.15
15	3.06	3.21	−0.64	3.57	3.19	1.38
16	3.42	3.31	0.91	3.44	3.29	1.22
17–20	3.56	3.51	0.39	3.47	3.45	0.14

Parameter Estimates

If the model holds, then the parameter estimates $\hat{\beta}_l$ are approximately normally distributed with respective asymptotic variances S^{ll}, where the $S^{ll'}$, $1 \le l \le 6$, $1 \le l' \le 6$, are the elements of the inverse S^{-1} of the matrix S with coordinates

$$\begin{aligned} S_{ll'} &= \sum_k \sum_h (X^K_{kkl} - \Theta^K_{hl})(X^K_{khl} - \Theta^K_{hl'})P^K_{k \cdot h} \\ &= \sum_s \sum_y \sum_k (x^K_{sykl} - \theta^{SYK}_{syl})(x^K_{sykl} - \theta^{SYK}_{syl'})m_{ijk}, \end{aligned} \tag{6.83}$$

where

$$\Theta^K_{hl} = \sum_k X^K_{khl} P^K_{kh} \tag{6.84}$$

and

$$\theta^{SYK}_{syl} = \sum_k x^{SYK}_{sykl} m^{SYK}_{syk} / n^{SY}_{sy}. \tag{6.85}$$

(The definition of θ^{SYK}_{syl} can be arbitrary if $n^{SY}_{sy} = 0$.) The estimated asymptotic covariance matrix $\hat{S}^{-1}$ is the inverse of $\hat{S}$, where

$$\begin{aligned} \hat{S}_{ll'} &= \sum_k \sum_h (X^K_{khl} - \hat{\Theta}^K_{hl})(X^K_{khl'} - \hat{\Theta}^K_{kl'})\hat{P}^K_{k \cdot h} \\ &= \sum_s \sum_y \sum_k (x^{SYK}_{sykl} - \hat{\theta}^{SYK}_{syl})(x^{SYK}_{syk} - \hat{\theta}^{SYK}_{syl'})\hat{m}^{SYK}_{syk}, \end{aligned} \tag{6.86}$$

$$\hat{\Theta}^K_{kl} = \sum_k X^K_{khl} \hat{P}^K_{k \cdot h}, \tag{6.87}$$

and

$$\hat{\theta}_{syl}^{SYK} = \sum_k x_{sykl}^{SYK} \hat{m}_{syk}^{SYK} / n_{sy}^{SY}. \tag{6.88}$$

(Again, the definition of $\hat{\theta}_{syl}^{SYK}$ can be arbitrary if $n_{sy}^{SY} = 0$). As shown in Haberman (1974a, pp. 358–367), these large-sample approximations should become increasingly accurate as the total sample size N becomes large. The essential condition required is that the S^{ll} all become small as N becomes large.

The major observation in Table 6.21 is that a strong education effect is present, but effects of sex are not evident. Were response independent of sex and education, the model would hold with $\xi_1 = \xi_2 = 0$. Since $\hat{\xi}_2/s(\hat{\xi}_2) = 12.50$, the presence of an education effect is clear. A relationship of response to sex has not been shown since $\hat{\xi}_1/s(\hat{\xi}_1) = 0.46$.

To interpret $\hat{\xi}_2$, observe that the log odds of response $K_h = k$ rather than $k + 1$ is estimated to increase by 0.1031 if education Y_h increases by one year and sex S_h remains the same. The corresponding odds ratio is increased by

$$100(e^{0.1031} - 1) = 10.86$$

percent.

The effects of education are also quite clear if Table 6.22 is examined. This table considers the average number

$$\sum_k (k - 1) n_{syk}^{SYK} / n_{sy}^{SY}$$

of the positive responses for subjects of sex s with y years of education, $1 \le s \le 2, 9 \le y \le 16$.

Since sample sizes are modest for the lowest and highest educational levels, the table also includes the average number

$$\sum_{y=0}^{8} \sum_k (k - 1) n_{syk}^{SYK} \bigg/ \sum_{y=0}^{8} n_{sy}^{SY}$$

of positive responses of subjects of sex s and with no more than eight years of education and the average number

$$\sum_{y=17}^{20} \sum_k (k - 1) n_{syk}^{SYK} \bigg/ \sum_{y=17}^{20} n_{sy}^{SY}$$

of positive responses of subjects of sex s with at least 17 years of education. To find estimated expected averages, n_{syk}^{SYK} is replaced by $\hat{m}_{syk}$. For males, observed averages rise steadily from 1.90 to 3.56, while estimated expected averages rise from 1.80 to 3.51. For females, the observed patterns is more irregular, but a rise from 1.45 to a high of 3.57 is observed. The estimated averages rise from 1.65 to 3.45. Effects of sex are much smaller, as can be seen by comparison of observed or estimated expected averages for males and females for the same educational level.

Adjusted Residuals

As in logit analysis with ungrouped data, adjusted residuals for multinomial response models for ungrouped data must be based on selected linear combinations of the data. For example, in Table 6.22, the observed average for males with nine years of education is

$$\sum_s \sum_y \sum_k d_{syk} n_{syk}^{SYK} = \sum_k \sum_h D_{kh} N_{kh},$$

where

$$\begin{aligned} d_{syk} &= (k-1)/n_{19}^{SY}, \qquad s = 1, \quad y = 9, \\ &= 0, \qquad \text{otherwise,} \end{aligned}$$

and

$$\begin{aligned} D_{kh} &= (k-1)/n_{19}^{SY}, \qquad S_h = 1, \quad Y_h = 9, \\ &= 0, \qquad \text{otherwise.} \end{aligned}$$

The estimated expected average is

$$\sum_s \sum_y \sum_k d_{syk} \hat{m}_{syk}^{SYK} = \sum_k \sum_h D_{kh} \hat{P}_{k \cdot h}^{K}$$

and the adjusted residual is

$$\begin{aligned} t &= \left(\sum_s \sum_y \sum_k d_{syk} n_{syk}^{SYK} - \sum_s \sum_y \sum_k d_{syk} \hat{m}_{syk}^{SYK} \right) \Big/ \hat{c}^{1/2} \\ &= \left(\sum_k \sum_h D_{kh} N_{kh} - \sum_k \sum_h D_{kh} \hat{P}_{kh}^{K} \right) \Big/ \hat{c}^{1/2}, \end{aligned} \tag{6.89}$$

where

$$\begin{aligned} \hat{c} &= \sum_s \sum_y \sum_k d_{syk}^2 \hat{m}_{syk}^{SYK} - \sum_s \sum_y \left(\sum_k d_{syk} \hat{m}_{syk}^{SYK} \right)^2 \Big/ n_{sy}^{SY} - \sum_l \sum_{l'} \hat{f}_l \hat{f}_{l'} \hat{S}^{ll'} \\ &= \sum_k \sum_h D_{kh}^2 \hat{P}_{k \cdot h}^{K} - \sum_h \left(\sum_k D_{kh} \hat{P}_{k \cdot h}^{K} \right)^2 - \sum_l \sum_{l'} \hat{f}_l \hat{f}_{l'} S^{ll'} \end{aligned} \tag{6.90}$$

and

$$\begin{aligned} \hat{f}_l &= \sum_s \sum_y \sum_k d_{syk} (x_{sykl}^{SYK} - \hat{\theta}_{syl}^{SYK}) \hat{m}_{syk}^{SYK} \\ &= \sum_k \sum_h D_{kh} (X_{khl}^{K} - \hat{\Theta}_{hl}^{K}) \hat{P}_{k \cdot h}^{K}. \end{aligned} \tag{6.91}$$

Here $\hat{c}$ is the variance of the linear combination

$$\sum_k \sum_h D_{kh} \hat{P}^K_{k \cdot h} R_{kh}$$

of the residuals

$$R_{kh} = Y_{kh} - a_h - \sum_l b_l X^K_{khl}$$

from a weighted regression model in which hypothetical dependent variables Y_{kh} satisfy the equation

$$Y_{kh} = \alpha_h + \sum_l b_l X_{khl} + \varepsilon_{kh}$$

for independent errors ε_{kh} with respective means 0 and variances $(\hat{P}^K_{k \cdot h})^{-1}$. The a_h and b_l are the weighted-least-squares estimates of α_h and β_l, respectively. The adjusted residual t has an approximate standard normal distribution if the model holds and if each D_{kh} is small relative to $\hat{c}^{1/2}$.

The adjusted residuals in Table 6.22 are quite small for males, although the signs do exhibit some pattern. On the other hand, one adjusted residual for females exceeds 2 in magnitude and another is 1.92. As in Section 6.3, there is some suggestion that larger than usual changes occur between years 8 and 9 and between years 12 and 13. Nonetheless, the number of large residuals is certainly not unusual in Table 6.22.

Chi-square statistics The usual chi-square statistics

$$X^2_{SYK} = \sum_s \sum_y \sum_k (n^{SYK}_{syk} - \hat{m}^{SYK}_{syk})^2 / \hat{m}_{syk} \tag{6.92}$$

and

$$L^2_{SYK} = 2 \sum_s \sum_y \sum_k n^{SYK}_{syk} \log(n^{SYK}_{syk} / \hat{m}^{SYK}_{syk}) \tag{6.93}$$

are difficult to use in Table 6.20 since the large number of cells with small counts make the chi-square approximation too inaccurate to be useful. Problems are even more severe when the ungrouped statistics

$$X^2_{Ku} = \sum_k \sum_h (N_{kh} - \hat{P}^K_{k \cdot h})^2 / \hat{P}^K_{k \cdot h} \tag{6.94}$$

and

$$L^2_{Ku} = -2 \sum_k \sum_k N_{kh} \log \hat{P}^K_{k \cdot h} \tag{6.95}$$

are used. Nonetheless, the statistics L^2_{SYK} and L^2_{Ku} are useful for comparison of models. As in the case of logit models, differences in likelihood-ratio chi-square statistics have approximate chi-square distributions under quite general conditions. For a proof, see Haberman (1974a, pp. 372–373).

To illustrate behavior of chi-square statistics, Table 6.2 lists chi-square statistics X^2_{SYK} and L^2_{SYK} for the following five models based on (6.81):

(1) $\log P^K_{k \cdot h} = \alpha^K_h + \sum_{l=1}^{4} \beta_l X^K_{khl},$

(2) $\log P^K_{k \cdot h} = \alpha^K_h + \sum_{l=1}^{4} \beta_l X^K_{khl} + \beta_6 X^K_{kh6},$

(3) $\log P^K_{k \cdot h} = \alpha^K_h + \sum_{l=1}^{4} \beta_l X^K_{khl} + \beta_5 X^K_{kh5} + \beta_6 X^K_{kh6},$

(4) $\log \mathrm{P}^K_{k \cdot h} = \alpha^K_h + \sum_{l=1}^{4} \beta_l X^K_{khl} + \beta_5 X^K_{kh5} + \beta_6 X^K_{kh6} + \beta_7 X^K_{kh7},$

(5) $\log P^K_{k \cdot h} = \alpha^K_h + \sum_{l=1}^{4} \beta_l X^K_{khl} + \beta_6 X^K_{kh6} + \beta_8 X^K_{kh8}.$

Here

$$X^K_{kh7} = (3 - k)T_{h1} T_{h2}$$

and

$$X^K_{kh8} = [(3 - k)^2 - 2]T_{h2}.$$

Model 1 assumes that response is independent of sex and education. Model 3 is just (6.81), while Model 2 sets the sex-effect parameter β_5 to 0 in Model 3. Model 4 includes an interaction of sex and education. Under this model,

$$\begin{aligned} \Omega^K_{kk' \cdot h} &= \eta_{kk'} + \xi_{kk'1} T_{h1} + \xi_{kk'2} T_{h2} + \xi_{kk'3} T_{h3} \\ &= (\eta_{kk'} + \xi_{kk'1}) + (\xi_{kk'2} + \xi_{kk'3}) T_{h2}, \qquad S_h = 1, \\ &= (\eta_{kk'} - \xi_{kk'1}) + (\xi_{kk'2} - \xi_{kk'3}) T_{h2}, \qquad S_h = 2, \end{aligned}$$

where $\xi_{kk'1} = (k - k)\beta_5$, $\xi_{kk'2} = (k' - k)\beta_6$, and $\xi_{kk'3} = (k' - k)\beta_7$. Thus for each sex, there is a distinct linear relationship between the logits $\Omega^K_{kk' \cdot h}$ and the education scores T_{h2}. In Model 5, the logit $\Omega^K_{k(k+1) \cdot h}$ is a linear function

$$\Omega^K_{k(k+1) \cdot h} = \eta_{k(k+1)} + \xi_{k(k+1)2} T_{h2}$$

of the education score T_{h2}; however, in contrast to Model 2,

$$\xi_{k(k+1)2} = (\beta_6 + 5\beta_8) - 2\beta_8 k$$

is a linear function of k. This model provides a test of the claim in Model 2 that $\xi_{k(k+1)2}$ is constant for $1 \leq k \leq 4$.

The values of X^2_{SYK} and L^2_{SYK} are quite different in Table 6.22, so that individual Pearson or likelihood-ratio chi-square statistics are not very useful. A further problem with the Pearson statistic X^2_{SYK} is the possibility, as

Table 6.23

Chi-Square Statistics for Models for Table 6.20

Model	X^2_{SYK}	L^2_{SYK}	Degrees of freedom of X^2 and L^2	Model compared	Decrease in L^2_{SYK} compared to reference model	Decrease in degrees of freedom compared to reference model
1	392.92	395.15	160	—	—	—
2	222.94	198.32	159	1	1	196.84
3	220.72	198.10	158	2	1	0.22
4	214.90	192.92	157	3	1	5.18
5	223.28	197.72	158	2	1	0.60

in Models 2 and 5, that X^2_{SYK} may increase as the model becomes more restrictive.

To find the degrees of freedom in Table 6.23, observe that there are 41 n^{SY}_{sy} which are positive. If there are q β-parameters in the model, then there are $(5 - 1)(41) - q = 164 - q$ degrees of freedom.

Comparison of likelihood-ratio chi-square statistics for Models 1 and 2 provides overwhelming evidence that β_6 is not 0, so that response is not independent of education. Some indication of a relationship of response to sex also exists. Although the comparison between Models 2 and 3 provides no indication that β_5 is not 0, the comparison of Models 3 and 4 yields a change in L^2_{SYK} significant at the 3 percent level. Thus there is an indication that the strength of the relationship between education and response is influenced by the sex of the respondent. Comparison of Models 2 and 5 provides no evidence that β_8 is not 0, so that the assumption that all $\xi_{k(k+1)2}$ parameters are equal has not been disproven. In summary, residual paralysis, examination of parameter estimates, and chi-square statistics provide overwhelming evidence that response is related to education and provide ambiguous evidence that sex may affect the relationship of education and response.

Sufficiency of the Summary Statistic

The summary statistic K_h may be regarded as providing all information concerning the relationship of responses to sex and education if the conditional distribution of responses given in K_h is independent of the sex and education of the respondent. Otherwise, some information is lost by confining attention to the summary statistic rather than to the individual responses.

To describe the problem of information loss more formally, let E_h be defined as in Table 6.24. Observe that any possible combination of responses

Table 6.24

Definition of E_h in Table 6.19

Response	K_h	E_h	Response	K_h	E_h
+ + + +	5	1	− + + +	4	4
+ + + −	4	1	− + + −	3	4
+ + − +	4	2	− + − +	3	5
+ + − −	3	1	− + − −	2	2
+ − + +	4	3	− − + +	3	6
+ − + −	3	2	− − + −	2	3
+ − − +	3	3	− − − +	2	4
+ − − −	2	1	− − − −	1	1

to questions A, B, C, and D correspond to a unique pair (E_h, K_h). Let $P^{EK}_{ek\cdot h} > 0$ be the probability that $E_h = e$ and $K_h = k$, let $v(k)$, $1 \le k \le 5$, be the number of possible values of E_h if $K_h = k$, and let

$$P^{E\cdot K}_{e\cdot kh} = P^{EK}_{ek\cdot h}/P^{K}_{k\cdot h}, \qquad 1 \le e \le v(k), \quad 1 \le k \le 5,$$

be the conditional probability that $E_h = e$ given that $K_h = k$. If responses are conditionally independent of sex and education given the summary statistic, then for some $p^{E\cdot K}_{e\cdot k}$, $1 \le e \le v(k)$, $1 \le k \le 5$,

$$P^{E\cdot K}_{e\cdot kh} = p^{E\cdot K}_{e\cdot k},$$

so that

$$P^{EK}_{ek\cdot h} = p^{E\cdot K}_{e\cdot k} P^{K}_{k\cdot h}.$$

Thus the distribution of responses (E_h, K_h) depends on sex S_h and education Y_h only to the extent that the distribution of the summary statistic K_h depends on S_h and Y_h.

Since Table 6.19 has so many small counts, conditional independence is best tested by use of conditional log-linear models such as the simultaneous conditional logit model

$$\Omega^{E\cdot K}_{ee'\cdot kh} = \log(P^{E\cdot K}_{e\cdot kh}/P^{E\cdot K}_{e'\cdot kh}) = \eta^{K}_{ee'k} + \xi^{K}_{ee'1k}T_{h1} + \xi^{K}_{ee'2k}T_{h2}, \qquad K_h = k. \quad (6.96)$$

Equivalently,

$$\log P^{E\cdot K}_{e\cdot kh} = \alpha^{E\cdot K}_{kh} + \sum_{l=1}^{3q(k)} \beta_{lk} X_{ekhl}, \quad (6.97)$$

where $q(k) = v(k) - 1$,

$$\alpha_{kh}^{E\cdot K} = \log P_{v(k)\cdot kh}^{E\cdot K},$$

$$\begin{aligned}
\beta_{lk} &= \eta_{ev(k)}^{K}, && e = l, \quad 1 \le l \le q(k),\\
&= \xi_{ev(k)1k}^{K}, && e = l - q(k), \quad q(k) + 1 \le l \le 2q(k), \quad 1 \le l \le 2q(k),\\
&= \xi_{ev(k)2k}^{K}, && e = l - 2q(k), \quad 2q(k) + 1 \le l \le 3q(k), \quad 1 \le l \le 3q(k),\\
X_{ekhl} &= t_{ee'}^{E}, && e' = l, \quad 1 \le l \le q(k),\\
&= t_{ee'}^{E} T_{h1}, && e' = l - q(k), \quad q(k) + 1 \le l \le 2q(k),\\
&= t_{ee'}^{E} T_{h2}, && e' = l - 2q(k), \quad 2q(k) + 1 \le l \le 3q(k),
\end{aligned}$$

and

$$\begin{aligned}
t_{ee'}^{E} &= 1, && e = e',\\
&= 0, && e \ne e'.
\end{aligned}$$

If k is 1 or 5, then (6.97) is just the trivial model

$$\log P_{1\cdot kh}^{E\cdot K} = \alpha_{kh}^{E\cdot K}.$$

Under conditional independence, $\xi_{ee'1k}^{K} = \xi_{ee'2k}^{K} = 0$. Thus conditional independence can be tested by comparing the likelihood-ratio chi-square statistic for (6.96) to the likelihood-ratio chi-square statistic for the model

$$\log P_{e\cdot kh}^{E\cdot K} = \alpha_{k\cdot h}^{E\cdot K} + \sum_{l=1}^{q(k)} \beta_{lk} X_{ekhl}, \qquad K_h = k. \tag{6.98}$$

An alternate model also examined in this section ignores the sex effects $\xi_{ee'1k}^{K}$ but not the education effects $\xi_{ee'2k}^{K}$. Thus

$$\log P_{e\cdot kh}^{E\cdot K} = \alpha_{kh}^{E\cdot K} + \sum_{l=1}^{q(k)} \beta_{lk} X_{ekhl} + \sum_{l=2q(k)+1}^{3q(k)} \beta_{lk} X_{ekhl}, \qquad K_h = k. \tag{6.99}$$

Computation of maximum likelihood estimates. Maximum likelihood estimates for (6.67)–(6.99) may be computed separately for each k, $1 \le k \le 5$, from the $2 \times 21 \times v(k)$ table of counts n_{syek}^{SYEK}, $1 \le s \le 2$, $0 \le y \le 20$, $1 \le e \le v(k)$, where n_{syek}^{SYEK} subjects h have sex $S_h = s$, education $Y_h = y$, and response pair $(E_h, K_h) = (e, k)$. For example, $n_{1011}^{SYEK} = 0$ and $n_{1022}^{SYEK} = 2$. Let $p_{e\cdot syk}^{E\cdot SYK} > 0$ be the conditional probability $P_{e\cdot kh}^{E\cdot K}$ that $E_h = e$ if $S_h = e$, $Y_h = y$, and $K_h = k$. Then given n_{syk}^{SYK}, n_{SYEK}^{SYEK} has conditional expected value $n_{syek}^{SYEK} = n_{syk}^{SYK} p_{e\cdot syk}^{E\cdot SYK}$ and (6.97) is equivalent to the log–linear model

$$\log m_{syek}^{SYEK} = \alpha_{syk}^{SYK} + \sum_{l=1}^{3q(k)} \beta_{lk} x_{syek}^{SYEK}, \qquad n_{syk}^{SYK} > 0, \tag{6.100}$$

where

$$\begin{aligned} x_{syek}^{SYEK} &= t_{ee'}^{E}, \qquad e' = l, \quad 1 \le l \le q(k), \\ &= t_s^S t_{ee'}^{E}, \qquad e' = l - q(k), \quad q(k) + 1 \le l \le 2q(k), \\ &= t_y^Y t_{ee'}^{E}, \qquad e' = l - 2q(k), \quad 2q(k) + 1 \le l \le 3q(k), \end{aligned} \tag{6.101}$$

$t_1^S = -t_2^S = 1$, and $t_y^Y = y - 10$. Equation (6.98) is equivalent to

$$\log m_{syek}^{SYEK} = \alpha_{syk}^{SYK} + \sum_{l=1}^{q(k)} \beta_{lk} x_{stek}^{SYEK}, \qquad n_{syk}^{SYK} > 0, \tag{6.102}$$

while (6.99) is equivalent to

$$\log m_{stek}^{SYEK} = \alpha_{syk}^{SYK} + \sum_{l=1}^{q(k)} \beta_{lk} x_{stek}^{SYEK} + \sum_{l=2q(k)+1}^{3q(k)} \beta_{lk} x_{syek}^{SYEK}, \qquad n_{syk}^{SYK} > 0. \tag{6.103}$$

Computations are routine, except that iterative calculations are unnecessary whenever k is 1 or 5. In such cases, $\hat{m}_{sylk}^{SYEK} = n_{sylk}^{SYEK}$. Parameter estimates and estimated symptotic standard deviations are summarized in Table 6.25.

Likelihood-ratio chi-squares The likelihood-ratio chi-square statistics for (6.100)–(6.103) may be found from the equations

$$L^2 = \sum_k L_k^2 \tag{6.104}$$

and

$$L_k^2 = 2 \sum_s \sum_y \sum_e n_{syek}^{SYEK} \log(n_{syek}^{SYEK}/\hat{m}_{syek}^{SYEK}). \tag{6.105}$$

Since for k equal 1 or 5, $\hat{m}_{syek}^{SYEK} = n_{syek}^{SYEK}$, $L_1^2 = L_5^2 = 0$, so that

$$L^2 = L_2^2 + L_3^2 + L_4^2.$$

Given the n_{syk}^{SYK}, the L_k^2 statistics are independently distributed. There are $[d(k) - f]q(k)$ degrees of freedom associated with each L_k^2, where $d(k)$ is the number of positive n_{syk}^{SYK}, $1 \le s \le 2$, $0 \le y \le 20$, and f is 1 in (6.103), 2 in (6.102), and 3 in (6.100). In this expression for degrees of freedom, $fq(k)$ is the number of β_{lk} parameters that are not assumed 0. From the formula for degrees of freedom of the L_k^2, it follows that L^2 has

$$\sum_k [d(k) - f]q(k) = 3[d(2) - f] + 5[d(3) - f] + 3[d(4) - f]$$

associated degrees of freedom. The difference in degrees of freedom between the L^2 statistics for (6.103) and (6.100) is then 22, while the difference for (6.103) and (6.102) is 11. Under conditional independence, both differences have approximate chi-square distributions. As is evident from Table 6.26, the

Table 6.25

Maximum Likelihood Estimates for the Conditional Log Linear Models of (6.100), (6.102), and (6.103)

	Defining equation					
	(6.100)		(6.102)		(6.103)	
Parameter	Estimate	EASD	Estimate	EASD	Estimate	EASD
$\beta_{12} = \eta^K_{142}$	0.736	0.716	0.620	0.681	0.228	0.540
$\beta_{22} = \eta^K_{242}$	3.076	0.604	2.964	0.566	2.552	0.424
$\beta_{32} = \eta^K_{342}$	3.262	0.601	3.153	0.564	2.697	0.422
$\beta_{42} = \xi^K_{1412}$	0.367	0.595	—	—	—	—
$\beta_{52} = \xi^K_{2412}$	0.341	0.479	—	—	—	—
$\beta_{62} = \xi^K_{3412}$	0.328	0.472	—	—	—	—
$\beta_{72} = \xi^K_{1422}$	0.477	0.180	0.459	0.177	—	—
$\beta_{82} = \xi^K_{2422}$	0.403	0.126	0.387	0.123	—	—
$\beta_{92} = \xi^K_{3422}$	0.302	0.120	0.287	0.117	—	—
$\beta_{13} = \eta^K_{163}$	−0.255	0.337	−0.242	0.331	0.223	0.274
$\beta_{23} = \eta^K_{263}$	0.305	0.277	0.304	0.274	0.337	0.265
$\beta_{33} = \eta^K_{363}$	−0.900	0.399	−0.883	0.391	−0.780	0.364
$\beta_{43} = \eta^K_{463}$	1.411	0.234	1.397	0.231	1.513	0.225
$\beta_{53} = \eta^K_{563}$	−1.337	0.469	−1.338	0.463	−1.232	0.430
$\beta_{63} = \xi^K_{1613}$	−0.366	0.287	—	—	—	—
$\beta_{73} = \xi^K_{2613}$	−0.326	0.271	—	—	—	—
$\beta_{83} = \xi^K_{3613}$	0.096	0.378	—	—	—	—
$\beta_{93} = \xi^K_{4613}$	−0.178	0.232	—	—	—	—
$\beta_{(10)3} = \xi^K_{5613}$	−0.318	0.437	—	—	—	—
$\beta_{(11)3} = \xi^K_{1623}$	0.373	0.105	0.366	0.105	—	—
$\beta_{(12)3} = \xi^K_{2623}$	0.157	0.094	0.154	0.094	—	—
$\beta_{(13)3} = \xi^K_{3623}$	0.173	0.130	0.179	0.132	—	—
$\beta_{(14)3} = \xi^K_{4623}$	0.190	0.080	0.190	0.080	—	—
$\beta_{(15)3} = \xi^K_{5623}$	−0.136	0.111	−0.135	0.140	—	—
$\beta_{14} = \eta^K_{144}$	1.710	0.262	1.924	0.258	2.281	0.198
$\beta_{24} = \eta^K_{244}$	−1.494	0.523	−1.570	0.522	−0.767	0.336
$\beta_{34} = \eta^K_{344}$	1.413	0.254	1.447	0.260	1.263	0.214
$\beta_{44} = \eta^K_{1414}$	−0.639	0.204	—	—	—	—
$\beta_{54} = \eta^K_{2414}$	−0.120	0.351	—	—	—	—
$\beta_{64} = \eta^K_{3414}$	0.054	0.216	—	—	—	—
$\beta_{74} = \eta^K_{1424}$	0.207	0.090	0.176	0.094	—	—
$\beta_{84} = \eta^K_{2424}$	0.302	0.145	0.335	0.147	—	—
$\beta_{94} = \eta^K_{3424}$	−0.126	0.088	−0.149	0.096	—	—

Table 6.26

Likelihood-Ratio Chi-Square Statistics for the Conditional Log-Linear Models of (6.100), (6.102), and (6.103)

Defining equation	L^2	Degrees of freedom	Decrease in L^2 from (6.103)	Decrease in degrees of freedom from (6.103)
(6.103)	415.93	319		
(6.102)	344.19	308	71.74	11
(6.100)	305.92	297	110.01	22

large differences between the L^2 statistics for the models specified by (6.103) and the other two L^2 statistics provides clear evidence that conditional independence does not hold. Therefore, the relationship between the response variables and the predictors sex and education cannot be entirely described if the number of positive responses is used as a summary statistic.

Models for Multiple Responses

Instead of seeking a summary statistic to describe the relationship of responses to sex and education, the relationship of all responses to the predictors may be simultaneously considered. Analyses of this type for ungrouped data have been explored by Nerlove and Press (1973). They are based on a decomposition of individual response probabilities analogous to the decomposition of log means into λ-parameters.

For example, in Table 6.19, let $P_{abcd \cdot h} > 0$ be the probability that $A_h = a$, $B_h = b$, $C_h = c$, and $D_h = d$, where A_h, B_h, C_h, and D_h are the respective responses to items A, B, C, and D. The indexes a, b, c, and d can each be 1 or 2, with 1 corresponding to the responses scored as positive in Table 6.19. One can write

$$\begin{aligned}\log P_{abcd \cdot h} = \Lambda_h &+ \Lambda_{ah}^{A} + \Lambda_{bh}^{B} + \Lambda_{ch}^{C} + \Lambda_{dh}^{D} + \Lambda_{abh}^{AB} \\ &+ \Lambda_{ach}^{AC} + \Lambda_{adh}^{AD} + \Lambda_{bch}^{BC} + \Lambda_{bdh}^{BD} + \Lambda_{cdh}^{CD} + \Lambda_{abch}^{ABC} \\ &+ \Lambda_{abdh}^{ABD} + \Lambda_{acdh}^{ACD} + \Lambda_{bcdh}^{BCD} + \Lambda_{abcdh}^{ABCD},\end{aligned} \tag{6.106}$$

where

$$\begin{aligned}\sum_a \Lambda_{bh}^{A} &= \sum_b \Lambda_{bh}^{B} = \sum_c \Lambda_{ch}^{C} = \sum_d \Lambda_{dh}^{D} \\ &= \sum_a \Lambda_{abh}^{AB} = \sum_b \Lambda_{abh}^{AB} = \cdots = \sum_a \Lambda_{abcdh}^{ABCD} = \sum_b \Lambda_{abcdh}^{ABCD} \\ &= \sum_c \Lambda_{abcdh}^{ABCD} = \sum_d \Lambda_{abcdh}^{ABCD} = 0.\end{aligned}$$

Models based on (6.106) assume that some of the Λ-parameters are linear functions of the predicting variables, some of the Λ-parameters are independent of the predictors, and some of the Λ-parameters are 0.

For example, consider a model in which all two-factor and three-factor interactions are constant, all four-factor interactions are 0, and all main effects are linear functions of the sex score T_{h1}, the education score T_{h2}, and the sex-by-education interaction score $T_{h3} = T_{h1}T_{h2}$. Thus for all h,

$$\Lambda_{abh}^{AB} = \lambda_{ab}^{AB}, \quad \Lambda_{ach}^{AC} = \lambda_{ac}^{AC}, \quad \Lambda_{adh}^{AD} = \lambda_{ad}^{AD},$$

$$\Lambda_{bch}^{BC} = \lambda_{bc}^{BC}, \quad \Lambda_{bdh}^{BD} = \lambda_{bd}^{BD}, \quad \Lambda_{cdh}^{CD} = \lambda_{cd}^{CD}, \quad \Lambda_{abch}^{ABC} = \lambda_{abc}^{ABC},$$

$$\Lambda_{ah}^{A} = \lambda_{a}^{A} + \rho_{a1}^{A}T_{h1} + \rho_{a2}^{A}T_{h2} + \rho_{a3}^{A}T_{h3},$$

$$\Lambda_{bh}^{B} = \lambda_{b}^{B} + \rho_{b1}^{B}T_{h1} + \rho_{b2}^{B}T_{h2} + \rho_{b3}^{B}T_{h3},$$

$$\Lambda_{ch}^{C} = \lambda_{c}^{C} + \rho_{c1}^{C}T_{h1} + \rho_{c2}^{C}T_{h2} + \rho_{c3}^{C}T_{h3},$$

$$\Lambda_{dh}^{D} = \lambda_{d}^{D} + \rho_{d1}^{D}T_{h1} + \rho_{d2}^{D}T_{h2} + \rho_{d3}^{D}T_{h3},$$

where the λ-parameters satisfy the usual constraints and the ρ-parameters satisfy constraints corresponding to the λ-parameters with the same superscripts. For example,

$$\sum_a \lambda_a^A = \sum_a \rho_{a1}^A = \sum_a \rho_{a2}^A = \sum_a \rho_{a3}^A = 0$$

and

$$\sum_d \lambda_d^D = \sum_d \rho_{d1}^D = \sum_d \rho_{d2}^D = \sum_d \rho_{d3}^D = 0.$$

Thus

$$\begin{aligned}\log P_{abcd\cdot h} &= \Lambda_h + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_d^D + \lambda_{ab}^{AB} + \lambda_{ac}^{AC} + \lambda_{ad}^{AD} + \lambda_{bc}^{BC} + \lambda_{bd}^{BD} \\ &\quad + \lambda_{cd}^{CD} + \lambda_{abc}^{ABC} + \lambda_{abd}^{ABD} + \lambda_{acd}^{ACD} + \lambda_{bcd}^{BCD} \\ &\quad + \sum_{g=1}^{3} (\rho_{ag}^A + \rho_{bg}^B + \rho_{cg}^C + \rho_{dg}^D)T_{hg} \\ &= \alpha_h + \sum_{f=1}^{26} \beta_f X_{abcdhf},\end{aligned} \tag{6.107}$$

where the β_f and X_{abcdhf} are defined as in Table 6.27.

This model can be interpreted in terms of a collection of four simultaneous logit models. Let $P_{bcd\cdot h}^{BCD}$ denote the marginal probability that $B_h = b$, $C_h = c$, and $D_h = d$, and let

Table 6.27

Coefficients and Parameters in the Model Defined by (6.107)[a]

f	Corresponding λ-parameter or ρ-parameter	X_{abcdhf}	$x_{syabcdf}$
1	λ_1^A	q_a	q_a
2	λ_1^B	q_b	q_b
3	λ_1^C	q_c	q_c
4	λ_1^D	q_d	q_d
5	λ_{11}^{AB}	$q_a q_b$	$q_a q_b$
6	λ_{11}^{AC}	$q_a q_c$	$q_a q_c$
7	λ_{11}^{AD}	$q_a q_d$	$q_a q_d$
8	λ_{11}^{BC}	$q_b q_c$	$q_b q_c$
9	λ_{11}^{BD}	$q_b q_d$	$q_b q_d$
10	λ_{11}^{CD}	$q_c q_d$	$q_c q_d$
11	λ_{111}^{ABC}	$q_a q_b q_c$	$q_a q_b q_c$
12	λ_{111}^{ABD}	$q_a q_b q_d$	$q_a q_b q_d$
13	λ_{111}^{ACD}	$q_a q_c q_d$	$q_a q_c q_d$
14	λ_{111}^{BCD}	$q_b q_c q_d$	$q_b q_c q_d$
15	ρ_{11}^A	$q_a T_{h1}$	$q_a(3 - 2s)$
16	ρ_{12}^A	$q_b T_{h2}$	$q_a(y - 10)$
17	ρ_{13}^A	$q_a T_{h1} T_{h2}$	$q_a(3 - 2s)(y - 10)$
18	ρ_{11}^B	$q_b T_{h1}$	$q_b(3 - 2s)$
19	ρ_{12}^B	$q_b T_{h2}$	$q_b(y - 10)$
20	ρ_{13}^B	$q_b T_{h1} T_{h2}$	$q_b(3 - 2s)(y - 10)$
21	ρ_{11}^C	$q_c T_{h1}$	$q_c(3 - 2s)$
22	ρ_{12}^C	$q_c T_{h2}$	$q_c(y - 10)$
23	ρ_{13}^C	$q_c T_{h1} T_{h2}$	$q_c(3 - 2s)(y - 10)$
24	ρ_{11}^D	$q_d T_{h1}$	$q_d(3 - 2s)$
25	ρ_{12}^D	$q_d T_{h2}$	$q_d(y - 10)$
26	ρ_{13}^D	$q_d T_{h1} T_{h2}$	$q_d(3 - 2s)(y - 10)$

[a] Note that $q_1 = 1$ and $q_2 = -1$.

denote the conditional probability that $A_h = a$, given that $B_h = b$, $C_h = c$, and $D_h = d$. Let

$$\Omega_{12\cdot bcdh}^{A\cdot BCD} = \log\left(\frac{P_{1\cdot bcdh}^{A\cdot BCD}}{P_{2\cdot bcdh}^{A\cdot BCD}}\right) = \log\left(\frac{P_{1bcd\cdot h}}{P_{2bcd\cdot h}}\right)$$

denote the log odds that A_h is 1 rather than 2 given that $B_h = b$, $C_h = c$, and $D_h = d$. Let

$$\Omega_{12\cdot acdh}^{B\cdot ACD} = \log\left(\frac{P_{a1cd\cdot h}}{P_{a2cd\cdot h}}\right),$$

$$\Omega_{12\cdot abdh}^{C\cdot ABD} = \log\left(\frac{P_{ab1d\cdot h}}{P_{ab2d\cdot h}}\right),$$

and

$$\Omega_{12\cdot abch}^{D\cdot ABC} = \log\left(\frac{P_{abc1\cdot h}}{P_{abc2\cdot h}}\right),$$

be defined in an analogous manner. Then (6.107) holds if and only if the following logit models are all satisfied:

$$\Omega_{12\cdot bcdh}^{A\cdot BCD} = 2(\lambda_1^A + \lambda_{1b}^{AB} + \lambda_{1c}^{AC} + \lambda_{1d}^{AD} + \lambda_{1bc}^{ABC} + \lambda_{1bd}^{ABD} + \lambda_{1cd}^{ACD}) + 2\sum_g \rho_{1g}^A T_{hg},$$

$$\Omega_{12\cdot acdh}^{A\cdot BCD} = 2(\lambda_1^B + \lambda_{a1}^{AB} + \lambda_{1c}^{BC} + \lambda_{1d}^{BD} + \lambda_{a1c}^{ABC} + \lambda_{a1d}^{ABD} + \lambda_{1cd}^{BCD}) + 2\sum_g \rho_{1g}^B T_{hg},$$

$$\Omega_{12\cdot abd}^{C\cdot ABD} = 2(\lambda_1^C + \lambda_{1c}^{AC} + \lambda_{b1}^{BC} + \lambda_{1d}^{CD} + \lambda_{ab1}^{ABC} + \lambda_{a1d}^{ACD} + \lambda_{b1d}^{BCD}) + 2\sum_g \rho_{1g}^C T_{hg},$$

and

$$\Omega_{12\cdot abc}^{D\cdot ABC} = 2(\lambda_1^D + \lambda_{a1}^{AD} + \lambda_{b1}^{BD} + \lambda_{c1}^{CD} + \lambda_{ab1}^{ABD} + \lambda_{ac1}^{ACD} + \lambda_{bc1}^{BCD}) + 2\sum_g \rho_{1g}^D T_{hg}.$$

Computation of maximum likelihood estimates Computation of maximum likelihood estimates for (6.107) is unusually tedious. If grouping is completely ignored, computations may be based on the $2 \times 2 \times 2 \times 2 \times N$ tables of counts N_{abcdh} in which

$$N_{abcdh} = 1, \quad A_h = a, \quad B_h = b, \quad C_h = c, \quad D_h = d,$$
$$= 0, \quad \text{otherwise.}$$

Each table N_{abcdh}, $1 \le a \le 2$, $1 \le b \le 2$, $1 \le c \le 2$, $1 \le d \le 2$, has an independent multinomial distribution with means $P_{abcd\cdot h}$, $1 \le a \le 2$, $1 \le b \le 2$, $1 \le c \le 2$, $1 \le d \le 2$. The usual Newton–Raphson algorithm can be applied to (6.107).

In practice, it is more efficient to use the $2 \times 21 \times 2 \times 2 \times 2 \times 2$ table of counts n_{syabcd}, $1 \le s \le 2$, $0 \le y \le 20$, $1 \le a \le 2$, $1 \le b \le 2$, $1 \le c \le 2$, $1 \le d \le 2$, in which n_{syabcd} is the number of subjects h with $S_h = s$, $Y_h = y$, $A_h = a$, $B_h = b$, $C_h = c$, and $D_h = d$. Let $p_{abcd\cdot sy}^{ABCD\cdot SY}$ be the conditional probability that $A_h = a$, $B_h = b$, $C_h = c$, and $D_h = d$ given that $S_h = s$ and $Y_h = y$. Then given the n_{sy}^{SY}, n_{syabcd} has conditional mean $m_{syabcd} = n_{sy}^{SY} p_{abcd\cdot sy}^{ABCD\cdot SY}$ and the log-linear model

$$\log m_{syabcd} = \alpha_{sy}^{SY} + \sum_{f=1}^{26} \beta_f x_{syabcdf}, \qquad n_{sy}^{SY} > 0, \tag{6.108}$$

holds, where the $x_{syabcdf}$ are defined as in Table 6.25. Results are summarized in Table 6.27. For some further details on derivation of Table 6.27, see

Exercise 6.7. The computational technique is essentially the same as in Section 6.2.

Large-Sample Properties of Estimates

As is the case with the marginal multinomial response models and the conditional logit-linear models of this section, maximum likelihood estimates of β-parameters are approximately normally distributed. If the model holds, then $\hat{\beta}_f$ has asymptotic mean β_f and asymptotic variance $\sigma^2(\hat{\beta}_f) = S^{ff}$, where the $S^{ff'}$, $1 \le f \le 26$, $1 \le f' \le 26$, are the elements of the inverse of the matrix S with elements

$$S_{ff'} = \sum_s \sum_y \sum_a \sum_b \sum_c \sum_d (x_{syabcdf} - \theta^{SY}_{syf})(x_{syabcdf} - \theta^{SY}_{syf'})m_{syabcd}$$

$$= \sum_h \sum_a \sum_b \sum_c \sum_d (X_{abcdhf} - \theta_{hf})(X_{abcdhf'} - \theta_{hf'})P_{abcd\cdot h}. \quad (6.109)$$

$$\theta^{SY}_{syf} = \sum_a \sum_b \sum_c \sum_d x_{syabcdf}\, m_{syabcd}/n^{SY}_{sy} \quad (6.110)$$

and

$$\theta_{hf} = \sum_a \sum_b \sum_c \sum_d X_{abcdhf} P_{abcd\cdot h}. \quad (6.111)$$

In the analogous weighted regression model,

$$Y_{abcdh} = \alpha_h + \sum_f \beta_f X_{abcdhf} + \varepsilon_{abcdh},$$

where the ε_{abcdh} are independent $N(0, P^{-1}_{abcd\cdot h})$ variables and the T_{abcdh} are hypothetical dependent variables, the weighted-least-squares estimate $\hat{\beta}_f$ of β_f has an $N(\beta_f, \sigma^2(\hat{\beta}_f))$ distribution. If the model holds and $\beta_f = 0$, then the standardized value $\hat{\beta}_f/s(\hat{\beta}_f)$ has an approximate $N(0, 1)$ distribution. Such standardized values are found in Table 6.28.

Table 6.28 suggests that quite a few parameters can be eliminated from the model without much effect. The only standardized values greater than 2 in magnitude correspond to λ^B_1, λ^C_1, λ^D_1, λ^{AB}_{11}, λ^{AC}_{11}, λ^{AD}_{11}, λ^{BC}_{11}, λ^{CD}_{11}, ρ^A_{12}, and ρ^B_{12}. Thus one might consider the model

$$\log P_{abcd\cdot h} = \lambda_h + \lambda^A_a + \lambda^B_b + \lambda^C_c + \lambda^D_d + \lambda^{AB}_{ab} + \lambda^{AC}_{ac}$$
$$+ \lambda^{AD}_{ad} + \lambda^{BC}_{bc} + \lambda^{CD}_{cd} + \rho^A_{a2} T_{h2} + \rho^B_{b2} T_{h2}. \quad (6.112)$$

Here λ^A_a has been added due to the hierarchy principle for multi-way tables. This model can also be written as

$$\log m_{syabcd} = \alpha^{SY}_{sy} + \sum_{f=1}^{8} \beta_f x_{syabcdf} + \beta_{10} x_{syabcd(10)}$$
$$+ \beta_{16} x_{syabcd(16)} + \beta_{19} x_{syabcd(19)}. \quad (6.113)$$

Table 6.28

Estimated Parameters in the Model Defined by (6.107)

Parameter	Estimate	EASD	Standardized value
λ_1^A	−0.004	0.059	−0.06
λ_1^B	0.155	0.057	2.70
λ_1^C	0.717	0.058	12.35
λ_1^D	−0.414	0.057	−7.32
λ_{11}^{AB}	0.250	0.053	4.68
λ_{11}^{AC}	0.360	0.057	6.33
λ_{11}^{AD}	0.597	0.057	10.43
λ_{11}^{BC}	0.174	0.052	3.32
λ_{11}^{BD}	−0.031	0.057	−0.55
λ_{11}^{CD}	0.353	0.054	6.49
λ_{111}^{ABC}	0.025	0.049	0.50
λ_{111}^{ABD}	−0.032	0.043	−0.73
λ_{111}^{ACD}	0.012	0.056	0.22
λ_{111}^{BCD}	0.045	0.056	0.81
ρ_{11}^A	−0.004	0.044	−0.09
ρ_{12}^A	0.091	0.014	6.55
ρ_{13}^A	−0.019	0.014	−1.39
ρ_{11}^B	−0.040	0.036	−1.09
ρ_{12}^B	0.104	0.012	8.49
ρ_{13}^B	−0.007	0.012	−0.58
ρ_{11}^C	0.027	0.041	0.65
ρ_{12}^C	0.026	0.013	1.93
ρ_{13}^C	0.006	0.012	0.46
ρ_{11}^D	0.075	0.041	1.84
ρ_{12}^D	0.000	0.012	0.02
ρ_{13}^D	−0.015	0.011	−1.32

Under this model, the maximum likelihood estimates in Table 6.29 are obtained. The reduction in the number of parameters estimated both simplifies computations and reduces estimated asymptotic standard deviations. The latter effect is evident in a comparison of Tables 6.28 and 6.29.

Comparison of (6.107) and (6.112) The likelihood-ratio chi-square statistic for grouped data for Table 6.19 is

$$L^2 = 2\sum_s\sum_y\sum_a\sum_b\sum_c\sum_d n_{syabcd}\log(n_{syabcd}/\hat{m}_{syabcd}), \qquad (6.114)$$

while the Pearson chi-square is

$$X^2 = \sum_s\sum_y\sum_a\sum_b\sum_c\sum_d (n_{syabcd} - \hat{m}_{syabcd})^2/\hat{m}_{syabcd}. \qquad (6.115)$$

As usual with sparse tables, these statistics by themselves are of little use. A clear indication of the problem is provided by the discrepancy between L^2 and X^2 in (6.107) and (6.112). If (6.107) is used with Table 6.19, then $L^2 = 514.5$ and $X^2 = 799.6$. The nominal degrees of freedom are $41 \times 15 - 26 = 589$. (Since no woman surveyed has 2 years of education, only 41 combinations of sex and years of education are present.) The corresponding values of L^2 and X^2 for (6.112) are 532.8 and 847.3, respectively. There are now 604 degrees of freedom. Thus the two chi-square statistics are dissimilar in value, and X^2 appears to be quite unstable.

On the other hand, the difference in L^2 statistics does have an approximate chi-square distribution on $26 - 11 = 604 - 589 = 15$ degrees of freedom if (6.112) holds. Since the observed difference in L^2 is 18.32, little evidence exists that the more complete model of (6.107) provides a better fit than does (6.112).

Comparison of (6.112) to simultaneous conditional logit models If (6.112) holds, then the simultaneous conditional logit model (6.96) holds with

$$\xi_{ev(k)1k} = 0, \qquad 1 \le e \le q(k), \quad 2 \le k \le 4,$$

$$\xi^{K}_{1422} = \xi^{K}_{2623} = \xi^{K}_{3623} = \xi^{K}_{1424} = \xi^{K}_{2434} = 2\rho^{A}_{12},$$

$$\xi^{K}_{3424} = 2(\rho^{A}_{12} - \rho^{B}_{12}),$$

$$\xi^{K}_{1623} = 2(\rho^{A}_{12} + \rho^{B}_{12}),$$

$$\xi^{K}_{2422} = \xi^{K}_{5623} = 2\rho^{B}_{12},$$

and

$$\xi^{K}_{3422} = 0.$$

For example,

$$\Omega^{E\cdot K}_{14\cdot 2h} = \log\left(\frac{P_{1222\cdot h}}{P_{2221\cdot h}}\right) = 2\lambda^{A}_{1} - 2\lambda^{D}_{1} - 2\lambda^{AB}_{12} - 2\lambda^{AC}_{12} + 2\lambda^{CD}_{11} + 2\rho^{A}_{12}T_{h2},$$

so that $\xi^{K}_{1422} = 2\rho^{A}_{12}$. This comparison of models suggests that (6.112) is not fully satisfactory as (6.112) implies (6.102). As can be seen from Table 6.26, the difference in L^2 statistics between (6.102) and (6.100) is 38.27 while the difference in degrees of freedom is 11. Thus the difference in L^2 statistics is highly significant. Consequently, a fully satisfactory model for the data must take some account of the effects of sex of respondent. Unfortunately, it is not very clear from Table 6.28 how best to accomplish this task. Consequently, attention will be given now to interpretation of (6.112).

Interpretation of parameter estimates for (6.112) Use of (6.112) to obtain a relatively satisfactory model for Table 6.19 suggests that the influence of sex on response has not been demonstrated in these data but that a relationship between education and responses to A and B is clearly present. This latter relationship is suggested by the significant deviations from 0 of $\hat{\rho}_{12}^{A}$ and $\hat{\rho}_{12}^{B}$.

To assess the importance of the relationship of education Y_h to responses A_h and B_h, consider the logit equations

$$\Omega_{12 \cdot bcdh}^{A \cdot BCD} = 2(\lambda_1^A + \lambda_{1b}^{AB} + \lambda_{1c}^{AC} + \lambda_{1d}^{AD} + \rho_{12}^{A} T_{h2})$$

and

$$\Omega_{12 \cdot acdh}^{B \cdot ACD} = 2(\lambda_1^B + \lambda_{a1}^{AB} + \lambda_{1c}^{BC} + \rho_{12}^{B} T_{h2}).$$

Given the same responses to questions B, C, and D, a change of one year in education corresponds to a change in $2\rho_{12}^{A}$ in the log odds of disagreement rather than agreement with a statement that women should take care of running their homes rather than the country. The estimate of $2\rho_{12}^{A}$ is $2\hat{\rho}_{12}^{A} = 0.196$. As can be seen in Table 6.29, the estimate is moderately well determined. Its value is about eight times as large as its EASD. The estimate 0.196 corresponds to multiplication of the odds ratio for disagreement rather than agreement by $e^{0.196} = 1.22$ if education is increased by a year and multiplication of the ratio by $e^{1.96} = 7.10$ if education is increased by 10 years. Results for the second question, which deals with married women working, are similar. Here ρ_{12}^{B} has an estimate of 0.212, $e^{0.212} = 1.24$, and $e^{2.12} = 8.33$. A link between education and the other responses has not been demonstrated.

The responses themselves are related, even given education. This result is shown by the large ratios between estimates for λ_{11}^{AB}, λ_{11}^{AC}, λ_{11}^{AD}, λ_{11}^{BC}, and λ_{11}^{CD}

Table 6.29

Estimated Parameters in the Model Defined by (6.112)

Parameter	Estimate	EASD
λ_1^A	−0.016	0.053
λ_1^B	0.172	0.041
λ_1^C	0.761	0.049
λ_1^D	−0.428	0.054
λ_{11}^{AB}	0.260	0.037
λ_{11}^{AC}	0.380	0.047
λ_{11}^{AD}	0.593	0.040
λ_{11}^{BC}	0.158	0.040
λ_{11}^{CD}	0.364	0.053
ρ_{12}^{A}	0.098	0.012
ρ_{12}^{B}	0.106	0.012

and the corresponding estimated asymptotic standard deviations. The log odds $\Omega^{A\cdot BCD}_{12\cdot bcdh}$ is dependent on the responses to B, C, and D, with the log odds increasing whenever a response to B, C, or D is scored " + " rather than " −." The strongest relationship appears to be between items A and D. The log cross-product ratio

$$\Omega^{AD\cdot BC}_{(12)(12)\cdot bch} = \Omega^{A\cdot BCD}_{12\cdot 1cdh} - \Omega^{A\cdot BCD}_{12\cdot 2cdh} = 4\lambda^{AD}_{11}$$

has an estimate of 2.37, which corresponds to a cross-product ratio of 10.7. Note that A refers to women's role in the home and D refers to women's relative emotion suitability for politics. The log odds $\Omega^{B\cdot ACD}_{12\cdot acdh}$ appears to have links with responses A and C, as can be seen from the estimates $\hat{\lambda}^{AB}_{11}$ and $\hat{\lambda}^{BC}_{11}$. Other relationships can be observed between C and D.

In summary, the model suggests that the four responses to questions concerning women are related, even taking into account education and sex. The model implies that given education, sex and the four responses are unrelated and that education is related to responses to A and B. These detailed conclusions can be reached with models for multiple responses. They are not available with marginal models.

The various models of this section illustrate some of the available methods for the analysis of one or more polytomous responses when at least one predictor is continuous or has many categories. The methods used are similar to those used with multi-way contingency tables in which cell counts are not large, except that special caution is required in residual analysis and in use of chi-square tests. Some difficulty also arises since computations are relatively expensive due to the large tables involved and due to the lack of accurate initial estimates of maximum likelihood estimates.

EXERCISES

6.1 Derive Tables 6.7 and 6.8.

Solution

Details are omitted due to the length of the computations. Calculations of estimates are similar to those in Table 6.4. The program in the appendix may be used to obtain Tables 6.7 and 6.8.

6.2 Derive Tables 6.10 and 6.11.

Solution

The approach here is similar to that in Exercise 6.1. Table 6.9 is helpful in using the program in the appendix.

6.3 Compare the linear-by-linear model, the model of unknown column scores, and the model of unknown row scores using Table 6.30 (Table 2.7 of Volume 1). Let $u_1 = 1$, $u_2 = -1$, $u_3 = 0$, $u_4 = 0$, $u_5 = 0$, $t_1 = 3$, $t_2 = 1$, $t_3 = -1$, and $t_4 = -3$ in the linear-by-linear model.

Table 6.30

Cross-Classification of Respondents in 1972–1975 General Social Surveys by Attitude to Treatment of Criminals by Courts and by Year of Survey[a,b]

	Year of survey				
Response	1972	1973	1974	1975	Total
Too harshly	105	68	42	61	276
Not harshly enough	1066	1092	580	1174	3912
About right	265	196	72	144	677
Don't know	173	138	51	104	466
No answer	4	10	8	7	29
Total	1613	1504	753	1490	5360

[a] National Opinion Research Center (1972, p. 29; 1973, p. 58; 1974, p. 55; 1975, p. 58).

[b] The question asked is, "In general, do you think the courts in this area deal too harshly or not harshly enough with criminals?" The reduced 1974 sample results from an experimental change in the question used with half the sample.

Solution

Results are summarized in Tables 6.31–6.33. The chi-square statistics for the linear-by-linear model are $X^2 = 31.0$ and $L_2 = 30.1$. The degrees of freedom are 11, so that the model does not appear to fit the data. The significance level is about 0.001. The model of unknown column scores yields an X^2 of 28.6 and an L^2 of 28.3 on 9 degrees of freedom, so this model is not satisfactory. The model of unknown row scores yields $X^2 = 14.0$ and $L^2 = 13.9$. The degrees of freedom are 8, so these chi-square statistics have significance levels of about 8 percent. Thus there is reason to question this model, although it is far more successful than the other two. Examination of residuals suggest some difficulty with the response "no answer."

6.4 In the memory example of Chapter 1 of Volume 1, subjects may be divided into those that remember exactly one event and those that remember

Table 6.31

Coefficients for Models for Table 6.30[a]

						Linear-by-linear	Unknown column scores			Unknown row scores			
i	j	x_{ij1}^a	x_{ij2}^a	x_{ij3}^a	x_{ij4}^a	x_{ij5}	x_{ij5}	x_{ij6}	x_{ij7}	x_{ij5}	x_{ij6}	x_{ij7}	x_{ij8}
1	1	1	0	0	0	3	1	0	0	3	0	0	0
2	1	0	1	0	0	−3	−1	0	0	0	3	0	0
3	1	0	0	1	0	0	0	0	0	0	0	3	0
4	1	0	0	0	1	0	0	0	0	0	0	0	3
5	1	−1	−1	−1	−1	0	0	0	0	−3	−3	−3	−3
1	2	1	0	0	0	1	0	1	0	1	0	0	0
2	2	0	1	0	0	−1	0	−1	0	0	1	0	0
3	2	0	0	1	0	0	0	0	0	0	0	1	0
4	2	0	0	0	1	0	0	0	0	0	0	0	1
5	2	−1	−1	−1	−1	0	0	0	0	−1	−1	−1	−1
1	3	1	0	0	0	−1	0	0	1	−1	0	0	0
2	3	0	1	0	0	1	0	0	−1	0	−1	0	0
3	3	0	0	1	0	0	0	0	0	0	0	−1	0
4	3	0	0	0	1	0	0	0	0	0	0	0	−1
5	3	−1	−1	−1	−1	0	0	0	0	1	1	1	1
1	4	1	0	0	0	−3	−1	−1	−1	−3	0	0	0
2	4	0	1	0	0	3	1	1	1	0	−3	0	0
3	4	0	0	1	0	0	0	0	0	0	0	−3	0
4	4	0	0	0	1	0	0	0	0	0	0	0	−3
5	4	−1	−1	−1	−1	0	0	0	0	3	3	3	3

[a] Note that x_{ij1}, x_{ij2}, and x_{ij3} are the same for all three models.

more than one event. The data used in Table 1.14 may then be reclassified to produce Table 6.34. Test the hypothesis that

$$\log m_{ij} = \alpha_j^B + \beta i.$$

In your analysis, consider the alternate model

$$\log m_{ij} = \alpha_j^B + \beta_j i.$$

Solution

In the first model, one has

$$\log m_{ij} = \alpha_j^B + \beta_1 x_{ij1},$$

where

$$x_{ij1} = i,$$

Table 6.32

Estimated Cell Mean and Adjusted Residuals for Models for Table 6.30

	1972			1973			1974			1975		
Response	A[a]	B[a]	C[a]	A	B	C	A	B	C	A	B	C
Too harshly	114.2	119.7	98.4	82.4	75.1	81.1	31.7	31.1	35.5	47.8	50.1	60.9
	−1.36	−2.58	1.38	−1.95	−1.36	−1.80	2.04	2.84	1.23	2.57	2.31	0.02
Not harshly	1093.7	1080.7	1079.2	1079.6	1099.1	1076.8	567.6	569.1	570.9	1171.1	1163.1	1185.1
enough	−2.70	−2.58	−1.38	0.85	−1.36	1.05	1.16	2.84	0.86	0.34	2.31	−1.46
About right	234.0	238.4	257.9	197.5	190.5	201.0	88.8	88.2	83.2	156.6	159.9	134.9
	2.88	2.56	0.98	−0.14	0.51	−0.45	−2.05	−2.25	−1.45	−1.34	−1.76	1.63
Don't know	161.1	164.1	170.6	136.0	131.1	137.6	61.1	60.1	58.9	107.8	110.1	98.9
	1.27	0.97	0.39	0.22	0.80	0.05	−1.46	−1.52	−1.20	−0.46	−0.76	1.07
No answer	10.0	10.2	6.8	8.5	8.2	7.5	3.8	3.8	4.4	6.7	6.9	10.2
	−2.6	−2.43	−1.90	0.63	0.76	1.10	2.32	2.34	1.86	0.13	0.01	−2.41

[a] Model A is the linear-by-linear model, Model B is the model of unknown column scores, and Model C is the model of unknown row scores. The first line is the estimated cell mean for the model and the second line is the corresponding adjusted residual.

Table 6.33

Estimated Parameters for Models for Table 6.30

	Model					
	Linear-by-linear		Unknown column scores		Unknown row scores	
Parameter	Estimate	EASD	Estimate	EASD	Estimate	EASD
β_1	−0.413	0.063	−0.415	0.064	−0.370	0.063
β_2	2.317	0.043	2.318	0.044	2.311	0.043
β_3	0.540	0.051	0.540	0.051	0.509	0.052
β_4	0.166	0.054	0.167	0.055	0.147	0.056
β_5	0.078	0.011	0.267	0.059	0.041	0.027
β_6			0.025	0.042	−0.055	0.018
β_7			−0.086	0.055	0.069	0.022
β_8					0.052	0.023

Table 6.34

Distribution by Number of Events Recalled and Months Prior to Interview of Stressful Events Reported by Subjects[a]

Months i before interview	Number of subjects reporting one event	Number of subjects reporting more than one event
1	15	34
2	11	44
3	14	28
4	17	26
5	5	30
6	11	24
7	10	32
8	4	27
9	8	29
10	10	11
11	7	28
12	9	31
13	11	18
14	3	19
15	6	23
16	1	11
17	1	14
18	4	11
Total	147	440

[a] One event randomly selected for each subject reporting events from one of 18 months prior to interview.

while in the second model

$$\log m_{ij} = \alpha_j^B + \beta_1 x_{ij1} + \beta_2 x_{ij2},$$

where $x_{ijk} = i, j = k, x_{ijk} = 0, j \neq k$. Neither model is very successful. For the first model, $X^2 = 49.3$, $L^2 = 53.0$, and there are 33 degrees of freedom, while in the second model, $X^2 = 48.0$, $L^2 = 50.5$, and there are 32 degrees of freedom.

It is difficult to find in Table 6.35 much pattern to departures from the two models by examining adjusted residuals, although much of the problem involves the interval from 10 to 13 months ago. This result is not inconsistent with difficulties noted in Table 1.14.

6.5 Derive Table 6.29 by means of the computer program in the Appendix.

Solution

The work is straightforward given Table 6.17, although the tedium involved in construction of the $x_{i'j'k'}$ can be greatly reduced by some relatively simple programming.

Table 6.35

Estimated Cell Means and Adjusted Residuals for Table 6.34[a]

	Subjects reporting one event		Subjects reporting more than one event	
Months i before interview	First model	Second model	First model	Second model
1	13.1(0.57)	15.2(−0.05)	39.1(−0.92)	37.2(−0.60)
2	12.3(−0.39)	14.0(−0.89)	36.8(1.30)	35.2(1.65)
3	11.6(0.75)	12.8(0.36)	34.6(−1.21)	33.4(−1.01)
4	10.9(1.94)	11.8(1.61)	32.6(−1.22)	31.6(−1.06)
5	10.2(−1.70)	10.9(−1.86)	30.6(−0.12)	30.0(0.01)
6	9.6(0.46)	10.0(0.34)	28.8(−0.93)	28.4(−0.86)
7	9.1(0.32)	9.2(0.28)	27.1(0.97)	26.9(1.02)
8	8.5(−1.59)	8.4(−1.58)	25.5(0.31)	25.5(0.31)
9	8.0(−0.00)	7.8(0.09)	24.0(1.06)	24.2(1.01)
10	7.5(0.92)	7.1(1.11)	22.6(−2.51)	22.9(−2.56)
11	7.1(−0.03)	6.6(0.18)	21.2(1.52)	21.7(1.40)
12	6.7(0.93)	6.0(1.26)	20.0(2.56)	20.6(2.40)
13	6.3(1.94)	5.6(2.42)	18.8(−0.19)	19.5(−0.35)
14	5.9(−1.22)	5.1(−0.98)	17.7(0.33)	18.5(0.13)
15	5.6(0.20)	4.7(0.64)	16.6(1.65)	17.5(1.40)
16	5.2(−1.90)	4.3(−1.70)	15.6(−1.24)	16.6(−1.47)
17	4.9(−1.82)	4.0(−1.60)	14.7(−0.19)	15.7(−0.46)
18	4.6(−0.30)	3.7(0.20)	13.8(−0.81)	14.9(−1.10)

[a] Adjusted residuals are in parentheses.

A virtue of the Bock and Yates (1973) program is its ability to perform the tedious construction of the $x_{i'j'k'}$ automatically.

6.6 Interpret the parameters γ_{21}^{AD} and γ_{21}^{CD} of Table 6.18 in terms of log odds or cross-product ratios.

Solution

In the case of γ_{21}^{AD},

$$\gamma_{21}^{AD} = \frac{1}{6 \cdot 3 \cdot 3 \cdot 2} \sum_i \sum_j \sum_k \sum q_{i2}^{A} q_{l1}^{D} \log m_{ijkl}$$

$$= \frac{1}{9} \sum_j \sum_k \frac{1}{12} (\tau_{(12)(13)\cdot jk}^{AD\cdot BC} + \tau_{(32)(13)\cdot jk}^{AD\cdot BC})$$

$$= \tfrac{1}{12} (\tau_{(12)(13)}^{AD} + \tau_{(32)(13)}^{AD})$$

for A_h and (B_h, C_h) are conditionally independent given D_h.

In the case of γ_{21}^{CD},

$$\gamma_{21}^{CD} = \frac{1}{3 \cdot 3 \cdot 6 \cdot 2} \sum_i \sum_j \sum_k \sum_l q_{k2}^{C} q_{l1}^{D} \log m_{ijkl}$$

$$= \frac{1}{9} \sum_i \sum_j \frac{1}{12} (\tau_{(12)(13)\cdot ij}^{CD\cdot AB} + \tau_{(32)(13)\cdot ij}^{CD\cdot AB})$$

$$= \frac{1}{3} \sum_j \frac{1}{12} (\tau_{(12)(13)\cdot j}^{CB\cdot B} + \tau_{(31)(13)\cdot j}^{CD\cdot B}).$$

6.7 Derive Table 6.28 and Table 6.29.

Solution

The MULTIQUAL program of Bock and Yates (1973) or the program in the Appendix can be used. Both programs require expanded storage capacity for this problem. Given Table 6.27, implementation is straightforward. The key observation is that the counts n_{syabcd} can be transformed to a 16 by 42 array of counts $n_{i^*j^*}^*$, with the 16 corresponding to all values of a, b, c, and d and the 42 corresponding to all values of s and y. For example, one can let $n_{i^*j^*}^* = n_{syabcd}$ if $j^* = 21(s-1) + y$ and $i^* = 8(a-1) + 4(b-1) + 2(c-1) + d$. In Table 6.29, only 11 of the $x_{syabcdf}$ in Table 6.27 are required, while in Table 6.28, all the $x_{syabcdf}$ are needed.

6.8 Show that (6.20) and (6.21) hold.

Solution

Observe that in (6.19),

$$\log m_{ij} = \lambda + \sum_{j'=1}^{s-1} \lambda_{j'}^{B} x_{jj'}^{B} + \sum_{k=1}^{2(r-1)} \beta_k x_{ijk},$$

where $x_{jj'}^{B}$ is 1 for $j = j'$, 0 for $j \neq j', j < s$, and -1 for $j = s$. Thus

$$\hat{m}_j^B - \hat{m}_s^B = \sum_i \sum_j x_{jj'}^B \hat{m}_{ij} = \sum_i \sum_j x_{jj'}^B n_{ij} = n_j^B - n_s^B$$

and (6.21) holds. Since

$$\sum_j \hat{m}_j^B = N = \sum_j n_j^B,$$

it follows that (6.20) must hold.

7 Incomplete Tables

Incomplete contingency tables are those in which some cells are ignored. The cells may be ignored because they are unobserved, because they pertain to events that cannot occur, or because they are unusual in other respects.

A large fraction of current literature on incomplete tables has been devoted to two-way tables with excluded cells. In this literature, an additive log-linear model has been used which is based on the one used in Chapter 2 of Volume 1 with complete tables. The resulting models, called quasi-independence models, have been discussed by Caussinus (1965), Goodman (1968), Bishop and Fienberg (1969), and Bishop, Fienberg, and Holland (1975, pp. 177–210), among many others. Two examples of use of these models are provided in Sections 7.1 and 7.2.

A limited literature exists concerning incomplete three-way or higher-way tables. In this literature, models are based on the hierarchical log-linear models of Chapters 3 and 4. Fienberg (1972a, b), Haberman (1974a, pp. 228–302), and Bishop, Fienberg, and Holland (1975, pp. 210–228) have examined such models. Section 7.3 illustrates use of these models.

Incomplete tables can also arise in multinomial response problems. Bock (1975, pp. 538–541) and Bock and Yates (1973) consider this possibility. Section 7.4 provides an example involving an ordered classification.

7.1 Incomplete Two-Way Tables and Migration

Table 7.1 provides a simple example of an $r \times s$ contingency table in which log-linear models for incomplete tables can aid in analysis. The observations compare current residence and residence at age 16 for subjects in the 1974 General Social Survey who were residing in the United States at age 16. Variables such as age or sex which may affect responses in this table will be investigated in Section 7.4.

The row and column variables are obviously dependent, but much of this dependence follows from the fact that 82 percent of the subjects interviewed still live in the region in which they were living at age 16. A more subtle issue to investigate is the relationship between residence at age 16 and current residence, given that the two are different. For assessment of this relationship, a table of migrants may be formed, as in Table 7.2.

To construct a model for Table 7.2 to express lack of relationship between former and current residence, given that they differ, let A_h denote the region of residence of subject h, $1 \le h \le N = 1430$, at age 16 and let B_h denote the current region of residence. Let $A_h = 1$ if the former residence is the Northeast, let $A_h = 2$ if the former residence is the South, etc. Let B_h be defined in a similar manner, so that $B_h = 1$ if the current residence is the Northeast, $B_h = 2$ if the current residence is the South, etc. For $i \neq j$, let n_{ij} be the

Table 7.1

Subjects in 1974 General Social Survey Cross-Classified by Residence at Age 16 and Current Residence, Excluding Subjects Resident in Foreign Countries at Age 16[a,b]

Residence at age 16	Current residence: Northeast	South	North Central	West	Total
Northeast	263	22	14	13	312
South	26	399	36	30	491
North Central	10	41	368	46	465
West	1	8	5	148	162
Total	300	470	423	237	1430

[a] Data tape from the 1974 General Social Survey of the National Opinion Research Center, University of Chicago.

[b] Regions are defined so that the Northeast includes New England and the Middle Atlantic States, New York, New Jersey, and Pennsylvania. The South includes the District of Columbia, the South Atlantic states Delaware, Maryland, West Virginia, Virginia, North and South Carolina, Georgia, and Florida, and the South Central states Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Oklahoma, Louisiana, and Texas. The North Central region includes Wisconsin, Illinois, Indiana, Michigan, Ohio, Minnesota, Iowa, Missouri, North and South Dakota, Nebraska, and Kansas. The West includes the Mountain states Montana, Idaho, Wyoming, Nevada, Utah, Colorado, Arizona, and New Mexico and the Pacific states Washington, Oregon, California, Alaska, and Hawaii.

Table 7.2

Subjects in 1974 General Social Survey Resident in the United States at Age 16 in a Different Region Than Their Current Residence

Residence at age 16	Current Residence				
	Northeast	South	North Central	West	Total
Northeast	--	22	14	13	49
South	26	·	36	30	92
North Central	10	41	—	46	97
West	1	8	5	--	14
Total	37	71	55	89	252

number of subjects h with $A_h = i$ and $B_h = j$. Let $n_{ii} = 0$. Assume that given the B_h, $1 \le h \le N = 1430$, the A_h are independently distributed with probability $p_{i\cdot j}^{A\cdot B}$ that $A_h = i$ if $B_h = j$. For $i \ne j$, assume $p_{i\cdot j}^{A\cdot B} > 0$.

Independence of residence at age 16 and current residence implies that $p_{i\cdot j}^{A\cdot B}$ is independent of j. Although independence does not hold, one consequence of independence should be noted. Let m_{ij} be the expected value of n_{ij}, given that n_j^B subjects have current residence $B_h = j$ different from former residence A_h. Then

$$\begin{aligned} m_{ij} &= n_j^B p_{i\cdot j}^{A\cdot B}/(1 - p_{j\cdot j}^{A\cdot B}), \qquad && i \ne j, \\ &= 0, && i = j. \end{aligned}$$

Note in this formula that given that $B_h = j$, $1 - p_{j\cdot j}^{A\cdot B}$ is the probability that $A_h \ne j$ and for $i \ne j$, $p_{i\cdot j}^{A\cdot B}/(1 - p_{j\cdot j}^{A\cdot B})$ is the conditional probability that $A_h = i$ given that $A_h \ne j$. Under independence,

$$p_{i\cdot j}^{A\cdot B}/(1 - p_{j\cdot j}^{A\cdot B}) = p_{i\cdot 1}^{A\cdot B}/(1 - p_{j\cdot 1}^{A\cdot B}),$$

so that

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B, \qquad i \ne j, \tag{7.1}$$

for some λ, λ_i^A, and λ_j^B such that

$$\sum \lambda_i^A = \sum \lambda_j^B = 0. \tag{7.2}$$

Indeed, one may let

$$\begin{aligned} \lambda &= \tfrac{1}{4} \sum_{i'} \log p_{i'\cdot 1}^{A\cdot B} + \tfrac{1}{4} \sum_{j'} \log[n_{j'}^B/(1 - p_{j'\cdot 1}^{A\cdot B})], \\ \lambda_i^A &= p_{i\cdot 1}^{A\cdot B} - \tfrac{1}{4} \sum_{i'} \log p_{i'\cdot 1}^{A\cdot B}, \\ \lambda_j^B &= \log[n_j^B/(1 - p_{j\cdot 1}^{A\cdot B})] - \tfrac{1}{4} \sum_{j'} \log[n_{j'}^B/(1 - p_{j'\cdot 1}^{A\cdot B})] \end{aligned}$$

Equations (7.1) and (7.2) may hold without independence holding. All that is implied is that

$$p_{i \cdot j}^{A \cdot B} = c_i/(1 - c_j), \qquad i \neq j,$$

where the c_i are positive numbers such that

$$\sum c_i = 1.$$

Under independence, $c_i = p_{i \cdot 1}^{A \cdot B}$. More generally, it is easily shown that

$$C_i = \exp(\lambda_i^A) \Big/ \sum_{i'} \exp(\lambda_{i'}^A)$$

(see Exercise 7.1). The model defined by (7.1) and (7.2) has been used by Savage and Deutsch (1960) and Goodman (1963, 1964) in analysis of transaction flows and by Goodman (1965, 1968, 1969) in analysis of social-mobility tables.

The model is an example of a quasi-independence model for an $r \times s$ contingency table. In such a model, counts n_{ij}, $1 \leq i \leq r$, $1 \leq j \leq s$, are observed. The counts n_{ij} are 0 for (i, j) not in I and have respective means m_{ij} for (i, j) in I. It is assumed that

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B, \qquad (i, j) \text{ in } I, \tag{7.3}$$

for some λ, λ_i^A, and λ_j^B such that

$$\sum \lambda_i^A = \sum \lambda_j^B = 0. \tag{7.4}$$

The counts n_{ij} may be independent Poisson variables, they may form a table with a multinomial distribution, each row of counts n_{ij}, $1 \leq j \leq s$, may have an independent multinomial distribution, or each column of counts n_{ij}, $1 \leq i \leq r$, may have an independent multinomial distribution. If I includes all pairs (i, j), $1 \leq i \leq r$, $1 \leq j \leq s$, then the quasi-independence model reduces to the independence model of Chapter 2. If $r = s = 4$ and I contains the pairs (i, j), $1 \leq i \leq 4$, $1 \leq j \leq 4$, $i \neq j$, then one obtains the model of (7.1) and (7.2).

Maximum Likelihood Equations

In a quasi-independence model, the row marginal totals $\hat{m}_i^A$ for the table of estimated means $\hat{m}_{ij}$ are equal to the observed row marginal totals n_i^A of the table and the column marginal totals $\hat{m}_j^B$ for the table of estimated means are equal to the observed column marginal totals n_j^B. More formally,

$$\hat{m}_i^A = n_i^A, \qquad 1 \le i \le r, \tag{7.5}$$

$$\hat{m}_j^B = n_j^B, \qquad 1 \le j \le s, \tag{7.6}$$

$$\hat{m}_{ij} = 0, \qquad (i, j) \text{ not in } I, \tag{7.7}$$

$$\log \hat{m}_{ij} = \hat{\lambda} + \hat{\lambda}_i^A + \hat{\lambda}_j^B, \qquad (i, j) \text{ in } I, \tag{7.8}$$

$$\sum \hat{\lambda}_i^A = \sum \hat{\lambda}_j^B = 0. \tag{7.9}$$

The arguments used in Section 2.6 of Volume 1 to derive maximum likelihood equations for the independence model require only minor changes to lead to these maximum likelihood equations. In the case of (7.1) and (7.2), one finds that

$$\hat{m}_i^A = n_i^A, \qquad 1 \le i \le 4,$$

$$\hat{m}_j^B = n_j^B, \qquad 1 \le j \le 4,$$

$$\hat{m}_{ii} = 0, \qquad 1 \le i \le 4,$$

$$\log \hat{m}_{ij} = \hat{\lambda} + \hat{\lambda}_i^A + \hat{\lambda}_j^B, \quad i \ne j; \qquad \sum_i \hat{\lambda}_i^A = \sum_j \hat{\lambda}_j^B = 0.$$

These formulas are equivalent to formulas derived by Savage and Deutsch (1960) and Goodman (1963, 1964) in connection with transaction flow analysis. Equivalent formulas are obtained by Watson (1956) in the case in which I consists of all (i, j) such that $i > 1$ or $j > 1$.

Kastenbaum (1958) obtains an equivalent formula in the case in which $r \ne 2$, c is 4, and I contains all pairs (i, j), $1 \le i \le 2$, $1 \le j \le 4$, except (1, 1) and (2, 2). Detailed discussion of these formulas for general quasi-independence models appear in Caussinus (1965) and Goodman (1968), both of whom provide further references to earlier literature.

Iterative Proportional Fitting

Unlike the independence model, the quasi-independence model generally does not have maximum likelihood estimates expressed in closed form. In particular, maximum likelihood estimates for the proposed model for Table 7.2 cannot be found without iterative computations. Two general approaches are commonly used. The simplest is a version of iterative proportional fitting first applied to quasi-independence models by Goodman (1964). As in Chapters 3 and 4 of Volume 1, iterative proportional fitting does not help in computation of asymptotic variances, but it is easily implemented with common computer packages such as those described by Haberman (1972, 1973b) and Fay and Goodman (1975). Brown (1977) is also of interest. Convergence proofs may be found in Darroch and Ratcliff (1972) and Haberman (1974a, pp. 65–66).

Table 7.3

Use of Iterative Proportional Fitting to Obtain Maximum Likelihood Estimates for the Quasi-Independence Model for Table 7.2

i	m_{i10}	m_{i20}	m_{i30}	m_{i40}	m_{i0}^A	n_i^A
1	0	1	1	1	3	49
2	1	0	1	1	3	92
3	1	1	0	1	3	97
4	1	1	1	0	3	14
Total:	3	3	3	3	12	252
Corresponding n_j^B	37	71	55	89		

i	m_{i11}	m_{i21}	m_{i32}	m_{i41}	m_{i1}^A	n_i^A
1	0	16.333	16.333	16.333	49.000	49
2	30.667	0	30.667	30.667	92.000	92
3	32.333	32.333	0	32.333	97.000	97
4	4.667	4.667	4.667	0	14.000	14
Total:	67.667	53.333	51.667	79.333	252.000	252
Corresponding n_j^B	37	71	55	89		

i	m_{i22}	m_{i22}	m_{i32}	m_{i42}	m_{i2}^A	n_i^A
1	0	21.744	17.387	18.324	57.454	49
2	16.768	0	32.645	34.403	83.817	92
3	17.680	43.044	0	36.273	96.997	97
4	2.552	6.212	4.968	0	13.732	14
Total:	37.000	71.000	55.000	89.000	252.000	252
Corresponding n_j^B	37	71	55	89		

i	$m_{i1(10)}$	$m_{i2(10)}$	$m_{i3(10)}$	$m_{i4(10)}$	$m_{i(10)}^A$	n_i^A
1	0	19.397	14.310	15.295	49.000	49
2	17.929	0	35.789	38.253	92.000	92
3	16.616	44.959	0	35.451	97.000	97
4	2.456	6.644	4.902	0	14.000	14
Total:	37.004	70.952	55.037	89.007	0	252
Corresponding n_j^B	37	71	55	89		

For a review of earlier proofs, which are either less general or less rigorous, see Fienberg (1970). Various algebraically equivalent approaches are found in the literature. A reasonable approach found in Caussinus (1965) lets

$$\begin{aligned} m_{ij0} &= 1, \quad (i,j) \text{ in } I, & m_{ij(t+1)} &= m_{ijt} n_i^A / m_{it}^A, \quad t \text{ even}, \\ &= 0, \quad (i,j) \text{ not in } I, & &= m_{ijt} n_j^B / m_{jt}^B, \quad t \text{ odd}. \end{aligned}$$

A summary of computations for Table 7.2 is provided in Table 7.3.

The Newton–Raphson Algorithm

The Newton–Raphson algorithm is easily adapted for use with quasi-independence models. For $1 \le k \le r$, let

$$\begin{aligned} x_{ijk} &= 1, \quad i = k, \\ &= 0, \quad i \ne k, \quad i < r, \\ &= -1, \quad i = r. \end{aligned}$$

Note that

$$\log m_{ij} = \alpha_j + \sum_{h=1}^{r-1} \beta_h x_{ijk},$$

where $\alpha_j = \lambda + \lambda_j^B$ and $\beta_k = \lambda_k^A$, $1 \le k \le r - 1$. Let

$$\begin{aligned} y_{ij0} &= \log(n_{ij} + \tfrac{1}{2}), \quad (i,j) \text{ in } I, \\ &= 0, \quad (i,j) \text{ not in } I, \\ m_{ij0} &= n_{ij} + \tfrac{1}{2}, \quad (i,j) \text{ in } I, \\ &= 0, \quad (i,j) \text{ not in } I. \\ z_{ij} &= 1, \quad (i,j) \text{ in } I \\ &= 0, \quad (i,j) \text{ not in } I. \end{aligned}$$

Consider the regression model

$$\begin{aligned} y_{ij0} &= \alpha_j + \sum_{k=1}^{r-1} \beta_k x_{ijk} + \varepsilon_{ij}, \quad (i,j) \text{ in } I, \\ &= 0, \quad (i,j) \text{ not in } I, \end{aligned}$$

where the ε_{ij} are independent and have respective means 0 and variances m_{ij0}^{-1}. The estimates β_{k0} of β_k, $1 \le k \le r - 1$, obtained from the model are given by the equations

$$\sum_l S_{kl0} \beta_{l0} = w_{k0}, \qquad 1 \le k \le r - 1,$$

where

$$w_{k0} = \sum_i \sum_j (x_{ijk} - \theta_{ik0}) y_{ij0} m_{ij0},$$

$$S_{kl0} = \sum_i \sum_j (x_{ijk} - \theta_{jk0})(x_{ijl} - \theta_{jl0}) m_{ij0},$$

and

$$\theta_{jk0} = \sum_i x_{ijk} m_{ij0} \Big/ \sum_i m_{ij0}.$$

Given β_{k0}, $1 \le k \le r - 1$, one may let

$$\begin{aligned} v_{ij0} &= \sum_k \beta_{k0} x_{ijk} = \beta_{i0}, & i < r, \\ &= -\sum_{k=1}^{r-1} \beta_{k0}, & i = r, \end{aligned}$$

$g_{j0} = n_j^B / \sum_i z_{ij} \exp v_{ij0}$, and $m_{ij1} = z_{ij} g_{j0} \exp v_{ij0}$. Similarly to the Newton–Raphson algorithm of Chapter 6, the next iteration produces new approximations β_{k1} of $\hat{\beta}_k$, $1 \le k \le r - 1$, where $\beta_{k1} = \beta_{k0} + \delta_{k1}$,

$$\sum_l S_{kl} \delta_{l1} = a_k, \qquad 1 \le k \le r - 1,$$

$$a_{k1} = \sum_i \sum_j x_{ijk}(n_{ij} - m_{ij1}),$$

$$S_{kl1} = \sum_i \sum_j (x_{ijk} - \theta_{jk1})(x_{ijl} - \theta_{jl1}) m_{ij1},$$

and

$$\theta_{jkl} = \sum_i x_{ijk} m_{ij1} \Big/ \sum_i m_{ij1} = (n_j^B)^{-1} \sum_i x_{ijk} m_{ij1}.$$

Given β_{k1}, $1 \le k \le r - 1$,

$$\begin{aligned} v_{ij1} &= \sum_k \beta_{k1} x_{ijk} = \beta_{i1}, & i < r, \\ &= -\sum_{k=1}^{r-1} \beta_{k1}, & i = r, \end{aligned}$$

$$g_{j1} = n_j^B \Big/ \sum_i z_{ij} \exp v_{ij1}, \qquad \text{and} \qquad m_{ij2} = z_{ij} g_{j1} \exp v_{ij1}.$$

As is evident from Table 7.4, further iterations are unnecessary. Thus convergence is much more rapid than in the case of iterative proportional fitting.

The algorithm described here may be implemented by use of the program in the appendix or by use of Bock and Yates (1973). Thus a large number of programs are available for computation of maximum likelihood estimates under quasi-independence.

Table 7.4

Use of the Newton–Raphson Algorithm to Obtain Maximum Likelihood Estimates for the Quasi-Independence Model for Table 7.2

i	j	n_{ij}	x_{ij1}	x_{ij2}	x_{ij3}	y_{ij0}	m_{ij0}	z_{ij}	v_{ij0}	$z_{ij} \exp v_{ij0}$ [a]	m_{ij2}	v_{ij1}	$z_{ij} \exp v_{ij1}$ [b]	m_{ij2}	m_{ij3}
1	1	0	1	0	0	0.0000	0.0	0	−0.18440	0.0000	0.000	−0.17255	0.0000	0.000	0.000
2	1	26	0	1	0	3.2771	26.5	1	0.70235	2.0185	17.597	0.74467	2.1058	17.931	17.936
3	1	10	0	0	1	2.3514	10.5	1	0.64940	1.9144	16.690	0.66778	1.9499	16.604	16.609
4	1	1	−1	−1	−1	0.4055	1.5	1	−1.16735	0.3112	2.713	−1.23990	0.2894	2.464	2.455
1	2	22	1	0	0	3.1135	22.5	1	−0.18440	0.8316	19.313	−0.17255	0.8415	19.393	19.401
2	2	0	0	1	0	0.0000	0.0	0	0.70235	0.0000	0.000	0.74467	0.0000	0.000	0.000
3	2	41	0	0	1	3.7257	41.5	1	0.64940	1.9144	44.460	0.66778	1.9499	44.937	44.954
4	2	8	−1	−1	−1	2.1401	8.5	1	−1.16735	0.3112	7.227	−1.23990	0.2894	6.670	6.645
1	3	14	1	0	0	2.6742	14.5	1	−0.18440	0.8316	14.468	−0.17255	0.8415	14.300	14.305
2	3	36	0	1	0	3.5973	36.5	1	0.70235	2.0185	35.118	0.74467	2.1058	35.782	35.795
3	3	0	0	0	1	0.0000	0.0	0	0.64940	0.0000	0.000	0.66778	0.0000	0.000	0.000
4	3	5	−1	−1	−1	1.7048	5.5	1	−1.16735	0.3112	5.414	−1.23990	0.2894	4.918	4.900
1	4	13	1	0	0	2.6027	13.5	1	−0.18440	0.8316	15.534	−0.17255	0.8415	15.294	15.294
2	4	30	0	1	0	3.4177	30.5	1	0.70235	2.0185	37.705	0.74467	2.1058	38.269	38.269
3	4	46	0	0	1	3.8394	46.5	1	0.64940	1.9144	35.761	0.66778	1.9499	35.437	35.437
4	4	0	−1	−1	−1	0.0000	0.0	0	−1.16735	0.0000	0.000	−1.23990	0.0000	0.000	0.000

[a]
$$S_0 = \begin{bmatrix} 59.791 & 8.6287 & 3.418 \\ 8.628 & 64.482 & 0.871 \\ 3.418 & 0.871 & 72.448 \end{bmatrix}; \quad \mathbf{w}_0 = \begin{bmatrix} -2.746 \\ 44.263 \\ 47.029 \end{bmatrix}; \quad \boldsymbol{\beta}_0 = \begin{bmatrix} -0.18440 \\ 0.70235 \\ 0.64940 \end{bmatrix}$$

[b]
$$S_1 = \begin{bmatrix} 58.212 & 6.205 & 4.691 \\ 6.205 & 67.035 & 1.296 \\ 4.691 & 1.296 & 72.558 \end{bmatrix}; \quad \mathbf{a}_1 = \begin{bmatrix} 1.0385 \\ 2.9342 \\ 1.4444 \end{bmatrix}; \quad \boldsymbol{\delta}_1 = \begin{bmatrix} 0.01185 \\ 0.04232 \\ 0.01838 \end{bmatrix}; \quad \boldsymbol{\beta}_1 = \begin{bmatrix} -0.17255 \\ 0.74467 \\ 0.66778 \end{bmatrix}$$

Tests of Fit

The Pearson and likelihood-ratio chi-square statistics are readily applied to quasi-independence models. The only complication that may sometimes arise involves degrees of freedom. If

$$\sum_i \sum_j{}'$$

is used to denote summation over pairs (i, j) in I, then

$$X^2 = \sum_i \sum_j{}' \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \qquad \text{and} \qquad L^2 = 2 \sum_i \sum_j{}' n_{ij} \log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right).$$

Normally, if I has q elements, then the degrees of freedom are $q - r - s + 1$. For example, in Table 7.2, I has 12 elements and $r = s = 4$, so that the degrees of freedom are $12 - 4 - 4 + 1 = 5$. In Table 7.2, $X^2 = 13.4$ and $L^2 = 13.5$, so that the observed significance level is about 2 percent under either chi-square statistic. Thus the quasi-independence model does not appear adequate.

Unfortunately, this normal formula for degrees of freedom is not always correct. The formula only holds if the following conditions hold:

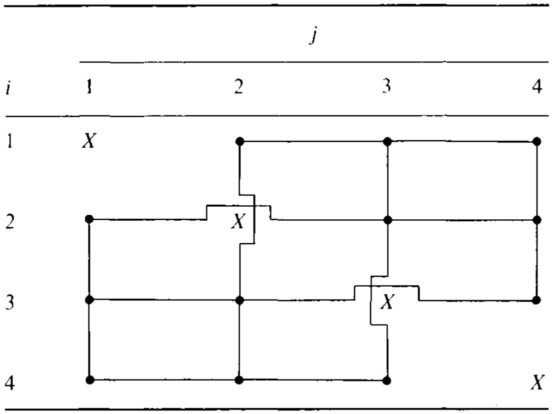
(a) Each row i contains some cell (i, j) in I.
(b) Each column j contains some cell (i, j) in I.
(c) The table is inseparable [see Goodman (1968)]. That is, for each distinct (i, j) and (k, l) in I, a series of elements (i_t, j_t), $1 \leq t \leq u$, of I exist such that $i_1 = i, j_1 = j, i_u = k, j_u = l$, and for $1 \leq t \leq u - 1$, $i_t = i_{t+1}$ or $j_t = j_{t+1}$.

If these conditions are satisfied, one may then say that (i, j) and (k, l) are connected [See Fienberg (1972b) or Bishop, Fienberg, and Holland (1975, p. 182) for discussion of origins of this term].

All these conditions are satisfied in Table 7.2. Conditions (a) and (b) are obvious. It is easily shown that condition (c) may be verified by choosing a fixed (i, j) in I and showing that for any other (k, l) in I, (i, j) and (k, l) are connected. For instance, in Table 7.2, one may let $(i, j) = (2, 1)$. If $(k, l) = (2, 3)$, then $(i_1, j_1) = (2, 1)$, $(i_2, j_2) = (2, 3)$ is a suitable sequence connecting (2, 1) and (2, 3), while if $(k, l) = (4, 3)$, then $(i_1, j_1) = (2, 1)$, $(i_2, j_2) = (2, 3)$, $(i_3, j_3) = (4, 3)$ is suitable. The diagram in Table 7.5 may be helpful. Note that if the table is viewed as part of a chessboard, then inseparability holds if a rook can move from (i, j) to (k, l) by always stopping at spaces in I. For further examples and discussion of degrees of freedom in incomplete two-way tables, see Goodman (1968), Haberman (1974a, pp. 245–264), and Bishop, Fienberg, and Holland (1975, pp. 182–185, 187–188) and Exercises 7.2 to 7.4.

Table 7.5

Verification That Table 7.2 Is Inseparable[a]



[a] If one can travel from one dot at (i, j) to another at (k, l) along a connected sequence of lines, then (i, j) and (k, l) are connected. Note that a move from one line to another can only occur at a dot. Thus one may travel from (2, 1) to (3, 3) but not from (2, 1) to (3, 2) by way of (2, 2). The table is inseparable if one can go from any dot to any other dot along a sequence of lines.

Use of Adjusted Residuals and Linear Contrasts

Adjusted residuals for Table 7.2 are shown in Table 7.6. They are obtained from the formula

$$r_{ij} = (n_{ij} - \hat{m}_{ij})/\hat{c}_{ij}^{1/2}, \qquad i \neq j,$$

where

$$\hat{c}_{ij} = \hat{m}_{ij}\left[1 - \hat{m}_{ij}/n_j^B - \sum_k \sum_l (x_{ijk} - \hat{\theta}_{jk})(x_{ijl} - \hat{\theta}_{jl})\hat{S}^{kl}\right],$$

$$\hat{\theta}_{jk} = (n_j^B)^{-1} \sum_i x_{ijk}\hat{m}_{ij},$$

$$\hat{S}_{kl} = \sum_i \sum_j (x_{ijk} - \hat{\theta}_{jk})(x_{ijl} - \hat{\theta}_{jl})\hat{m}_{ij},$$

and $\hat{S}^{kl}$, $1 \leq k \leq r - 1$, $1 \leq l \leq r - 1$, are the elements of the inverse of the matrix $\hat{S}$ with elements $\hat{S}_{kl}$, $1 \leq k \leq r - 1$, $1 \leq l \leq r - 1$.

Four adjusted residuals are large. They are associated with a residence at age 16 in the South or North Central region and a current residence in the Northeast or West. The suggestion from the pattern of signs of r_{21}, r_{31}, r_{24}, and r_{34} is that residents of the Northeast are relatively more likely to

Table 7.6

Adjusted Residuals for the Quasi-Independence Model for Table 7.2

Residence at age 16	Current residence Northeast	South	North Central	West
Northeast	—	0.93	−0.12	−0.80
South	3.07	—	0.07	−2.57
North Central	−2.51	−1.35	—	3.22
West	−1.06	0.77	0.06	—

come from the South rather than the North Central region than are residents of the West. Consider the log cross-product ratio

$$\log(n_{21} + \tfrac{1}{2}) - \log(n_{31} + \tfrac{1}{2}) - \log(n_{24} + \tfrac{1}{2}) + \log(n_{34} + \tfrac{1}{2}),$$

which estimates

$$\log\left(\frac{m_{21}}{m_{31}}\bigg/\frac{m_{24}}{m_{34}}\right) = \tau^{A \cdot B}_{23 \cdot 1} - \tau^{A \cdot B}_{23 \cdot 4}.$$

This estimate is 1.35, and the corresponding EASD

$$\left(\frac{1}{n_{21} + \frac{1}{2}} + \frac{1}{n_{31} + \frac{1}{2}} + \frac{1}{n_{24} + \frac{1}{2}} + \frac{1}{n_{34} + \frac{1}{2}}\right)^{1/2}$$

is 0.43. Thus the standardized value of the cross-product ratio is 1.35/0.43 = 3.11. Even given the problem of *post hoc* examination of data, this contrast appears important. Thus failure of the quasi-independence model appears to largely involve this interaction between origin in the South or North Central region and current residence in the Northeast or West. Exercise 7.7 may be instructive in this regard.

7.2 Quasi-Independence, Age at First Marriage and Current Age

The 1974 General Social Survey asked ever-married respondents for age at first marriage and current age. Obviously, these variables are not independent, for age at first marriage cannot exceed current age. Nonetheless, the question still remains whether age at first marriage has any further relationship to current age. To help investigate this situation, consider Table 7.7, in which responses of ever-married women are tabulated. A similar analysis for men is developed in Exercise 7.8.

Table 7.7

Ever-Married Women in 1974 General Social Survey, Cross-Classified by Age at First Marriage and Current Age

Age at first marriage	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71	Total
≤20	9	43	51	103	68	65	39	22	400
21–25	—	20	40	53	45	43	24	26	251
26–30	—	—	3	4	5	7	12	7	38
≥31	—	—	—	1	3	9	4	4	21
Total:	9	63	94	161	121	124	79	59	710

Let A_h denote the age group at marriage of ever-married woman h, $1 \le h \le 710$, and let B_h denote the current age group of woman h, so that if her age at marriage is between 21 and 25 and she is currently between 41 and 50, then $A_h = 2$ and $B_h = 5$. Let $p^{A \cdot B}_{i \cdot j}$ denote the conditional probability that $A_h = i$ given that $B_h = j$. Let n_{ij} be the number of women with age group at marriage $A_h = i$ and current age group $B_h = j$, so that $n_{24} = 45$. Assume that the pairs (A_h, B_h) are independent and identically distributed, so that the table of counts n_{ij} has a multinomial distribution. Let m_{ij} be the expected value of n_{ij}.

A simple quasi-independence model for these data assumes that

$$\log m_{ij} = \lambda + \lambda^A_i + \lambda^B_j, \qquad i < j,$$

so that for i and j less than or equal to k, the log odds

$$\tau^{A \cdot B}_{ij \cdot k} = \log(p^{A \cdot B}_{i \cdot k}/p^{A \cdot B}_{j \cdot k}) = \log(m_{ik}/m_{jk}) = \lambda^A_i - \lambda^A_j$$

does not depend on k.

For example, it is assumed that for each group k of women who are at least 26, the relative log odds $\tau^{A \cdot B}_{12 \cdot k}$ of marriage by 20 rather than marriage between 21 and 25 is constant. This assumption might in principle lead to problems for women who married when more than 30 since the older the respondent the more years exist after age 30 in which a first marriage is conceivable. Nonetheless, the practical problem is actually negligible; only two women in the sample married for the first time after they were 40.

Closed-Form Estimation

As in Section 7.1, one can use the Newton–Raphson or iterative proportional fitting algorithm with Table 7.7; but, it is possible to obtain exact maximum likelihood estimates for the quasi-independence model of Table

7.7 in a finite number of steps. This table is one of many incomplete tables such that maximum likelihood estimates under quasi-independence can be expressed in closed form. Among the many authors who have developed such closed-form expressions are Waite (1915), Watson (1956), Kastenbaum (1958), Caussinus (1965), Asano (1965), Goodman (1968), Bishop and Fienberg (1969), Haberman (1974a, pp. 266–282), and Bishop, Fienberg, and Holland (1975, pp. 192–206). Exercises 7.9, 7.10, and 7.11 provide some examples of such closed-form expressions of maximum-likelihood estimates.

To obtain maximum likelihood estimates for Table 7.7, it is only necessary to consider the n_{ij} for $i < j$. For such i and j, let

$$
\begin{aligned}
m_{ij1} &= n_{ij}, \\
m_{ij2} &= \left(\sum_{j'=3}^{8} m_{ij'1}\right)\left(\sum_{i'=1}^{2} m_{i'j1}\right)\Big/ \sum_{i'=1}^{2}\sum_{j'=3}^{8} m_{i'j'1}, \qquad i \geq 3, \quad j \leq 2, \\
&= m_{ij1}, \qquad \text{otherwise}, \\
m_{ij3} &= \left(\sum_{j'=4}^{8} m_{ij'2}\right)\left(\sum_{i'=1}^{3} m_{i'j2}\right)\Big/ \sum_{i'=1}^{3}\sum_{j'=4}^{8} m_{i'j'2}, \qquad i \geq 4, \quad j \leq 3, \\
&= m_{ij2}, \qquad \text{otherwise}, \\
m_{ij4} &= \left(\sum_{j'=5}^{8} m_{ij'3}\right)\left(\sum_{i'=1}^{4} m_{i'j3}\right)\Big/ \sum_{i'=1}^{4}\sum_{j'=5}^{8} m_{i'j'3}, \qquad i \geq 5, \\
&= m_{ij3}, \qquad \text{otherwise}.
\end{aligned}
$$

Then the maximum likelihood estimate $\hat{m}_{ij}$ of m_{ij} is m_{ij4}. Results are summarized in Table 7.8. The chi-square statistics for testing goodness of fit are

$$X^2 = \sum_{j=2}^{8} \sum_{i=1}^{\min(j-1,4)} (n_{ij} - \hat{m}_{ij})^2/\hat{m}_{ij} = 29.3$$

and

$$L^2 = 2\sum_{j=2}^{8} \sum_{i=1}^{\min(j-1,4)} n_{ij} \log(n_{ij}/\hat{m}_{ij}) = 27.9.$$

As noted by Caussinus (1965), among others, the general formula for degrees of freedom is $q - a - b + s$, where q is the number of elements in the index set I, a is the number of rows i such that some (i, j) is in I, b is the number of columns j such that some (i, j) is in I, and s is the largest number of cells (i_t, j_t), $1 \leq t \leq s$, such that (i_t, j_t) and $(i_{t'}, j_{t'})$ are not connected for $t \neq t'$.

The set I in this example includes the pairs (i, j), $1 \leq i < j \leq 8$ and $1 \leq i \leq 4$, so that q is 22, a is 4, b is 7, and s is 1. Thus there are 12 degrees of freedom. Note that $s = 1$ if the table is inseparable. The chi-square

Table 7.8

Computation of Maximum Likelihood Estimates for Table 7.7 under Quasi-Independence[a]

Age at first marriage	Iteration	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71
≤20	1	—	43.00	51.00	103.00	68.00	65.00	39.00	22.00
	2	—	43.00	54.69	93.76	67.92	64.91	37.87	28.85
	3	—	43.00	54.69	89.73	66.18	64.49	42.06	30.84
	4	—	43.00	54.69	89.73	64.31	65.91	41.99	31.36
21–25	1	—	—	40.00	53.00	45.00	43.00	24.00	26.00
	2	—	—	36.31	62.24	45.08	43.09	25.13	19.15
	3	—	—	36.31	59.56	43.93	42.81	27.92	20.47
	4	—	—	36.31	59.56	42.69	43.75	27.87	20.82
26–30	1	—	—	—	4.00	5.00	7.00	12.00	7.00
	2	—	—	—	4.00	5.00	7.00	12.00	7.00
	3	—	—	—	10.71	7.90	7.70	5.02	3.68
	4	—	—	—	10.71	7.67	7.86	5.01	3.74
≥31	1	—	—		—	3.00	9.00	4.00	4.00
	2	—	—		—	3.00	9.00	4.00	4.00
	3	—			—	3.00	9.00	4.00	4.00
	4	—	—	—	—	6.32	6.48	4.13	3.08

[a] Entries are values m_{ijt}.

statistics are significant at about the 0.005 level, so there appears to be substantial evidence against the quasi-independence model.

To examine the model failure more closely, consider the adjusted residuals in Table 7.9. These adjusted residuals can be obtained through an adaptation of a formula in Haberman (1973a) for triangular tables or by use of general formulas in Section 7.1. The only new point if the formulas in Section 7.1 are used is that only indices j between 2 and 8 need ever be considered in computations, for there is no index $(i, 1)$ in I. Note that the adjusted residual

Table 7.9

Adjusted Residuals for the Quasi-Independence Model for Table 7.7

Age at first marriage	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71
≤20	—		−0.86	2.46	0.75	−0.18	−0.72	−2.57
21–25	—		0.86	−1.25	0.49	−0.16	−0.97	1.48
26–30		—	—	−2.55	−1.13	−0.36	3.49	1.84
≥31			—	—	−1.64	1.24	−0.07	0.58

r_{12} is undefined since $\hat{m}_{12}$ is always equal to n_{12}. Several of the adjusted residuals are rather large in magnitude. The overall pattern provides some suggestion that age at marriage tends to decrease as respondents become younger. This possibility is considered in more detail in Section 7.4.

The two examples of quasi-independence models presented in this section and in Section 7.1 should be regarded as only a limited introduction. In these sections, a plausible quasi-independence model is given in advance. Other problems can arise in which quasi-independence is thought to be helpful but the choice of the set I is unclear. Such problems have been treated in various ways by Goodman (1965, 1969, 1971) and Fienberg (1969), among many others.

7.3 Hierarchical Models for Incomplete Multi-Way Tables

Analysis of incomplete multi-way tables by hierarchical models is quite similar to analysis of complete multi-way tables by hierarchical models. Occasional new complications involves difficulties in determination of degrees of freedom for chi-square statistics and in estimation of λ-parameters.

The literature to date is somewhat less extensive than is the literature for quasi-independence models for incomplete two-way tables. Available references include Fienberg (1972a,b), Bishop, Fienberg, and Holland (1975, pp. 210–256), Goodman (1975), Haberman (1974a, pp. 228–302), and Duncan (1975). All these references consider the iterative proportional fitting algorithm used in this section. Very little literature appears to deal with the Newton–Raphson algorithm for incomplete tables explored in this section, although the algorithm of Bock and Yates (1973) can be used with incomplete tables.

As an example of a hierarchical model for an incomplete table, consider the log-linear model

$$\log m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^S + \lambda_{ij}^{AB} + \lambda_{ik}^{AS} + \lambda_{jk}^{BS}, \qquad i \le j,$$

$$\sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_k \lambda_k^S = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB}$$

$$= \sum_i \lambda_{ik}^{AS} = \sum_k \lambda_{ik}^{AS} = \sum_j \lambda_{jk}^{BS} = \sum_k \lambda_{jk}^{BS} = 0,$$

for Table 7.10.

Here m_{ijk} is the expected value of n_{ijk}, and n_{ijk} is the observed number of ever-married subjects h with age group $A_h = 1$ at first marriage, current age group $B_h = j$, and sex $S_h = k$. Note that $n_{122} = 24$ and that $n_{ijk} = 0$ for $i > j$.

Table 7.10

Ever-Married Subjects in 1974 General Social Survey, Cross-Classified by Age at First Marriage, Current Age, and by Sex

Sex	Age at first marriage	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71	Total
Female	≤20	9	43	51	103	68	65	39	22	400
	21–25	—	20	40	53	45	43	24	26	251
	26–30	—	—	3	4	5	7	12	7	38
	≥31	—	—	—	1	3	9	4	4	21
	Total:	9	63	94	161	121	124	79	59	710
Male	≤20	2	24	21	30	22	19	16	11	145
	21–25	—	23	34	62	49	50	38	19	275
	26–30	—	—	3	10	20	27	23	19	102
	≥31	—	—	—	4	10	15	17	11	57
	Total:	2	47	58	106	101	111	94	60	579
Total	≤20	11	67	72	133	90	84	55	33	545
	21–25	—	43	74	115	94	93	62	45	526
	26–30	—	—	6	14	25	34	35	26	140
	≥31	—	—		5	13	24	21	15	78
	Total:	11	110	152	267	222	235	173	119	1289

To help interpret this model, consider age groups at first marriage i and i'. Let j be a current age group such that $i \le j$ and $i' \le j$. Let $p_{i\cdot jk}^{A\cdot BS}$ be the conditional probability that subject h was first married at age $A_h = i$ given that the current age is $B_h = j$ and that the subject is of sex $S_h = k$. The log odds of first marriage at age i rather than i' is

$$\tau_{ii'\cdot jk}^{A\cdot BS} = \log(p_{i\cdot jk}^{A\cdot BS}/p_{i'\cdot jk}^{A\cdot BS}) = \log(m_{ijk}/m_{i'jk}) = \log m_{ijk} - \log m_{i'jk}.$$

The variation in this log odds resulting from a change from sex $S_h = 1$ to sex $S_h = 2$ is

$$\begin{aligned}\tau_{(ii')(12)\cdot j}^{AS\cdot B} &= \tau_{ii'\cdot j1}^{A\cdot BS} - \tau_{ii'\cdot j2}^{A\cdot BS} \\ &= \log m_{ij1} - \log m_{i'j1} - \log m_{ij2} + \log m_{i'j2},\end{aligned}$$

which is equal to $\lambda_{i1}^{AS} - \lambda_{i'1}^{AS} - \lambda_{i2}^{AS} + \lambda_{i'2}^{AS}$. Thus for $i \le j$ and $i' \le j$, $\tau_{(ii')(12)\cdot j}^{AS\cdot B}$ is independent of current age j. Thus the interaction between sex and age at first marriage is independent of current age.

Maximum Likelihood Estimation

As usual, the first step in examination of the proposed log-linear model is computation of maximum likelihood estimates $\hat{m}_{ijk}$. Two possible methods, iterative proportional fitting and the Newton-Raphson algorithm, are available.

Iterative Proportional Fitting

Except for the choice of starting value, iterative proportional fitting proceeds just as in a model for a complete three-way table in which there is no three-factor interaction. One may let

$$\begin{aligned} m_{ijk0} &= 1, \qquad i \le j, \\ &= 0, \qquad i > j, \\ m_{ijk1} &= m_{ijk0} n_{ij}^{AB}/m_{ij0}^{AB}, \\ m_{ijk2} &= m_{ijk1} n_{ik}^{AS}/m_{ik1}^{AS}, \\ m_{ijk3} &= m_{ijk2} n_{jk}^{BS}/m_{jk2}^{BS}, \\ m_{ijk4} &= m_{ijk3} n_{ij}^{AB}/m_{ij3}^{AB}, \\ &\text{etc.} \end{aligned}$$

Results are summarized in Table 7.11. Standard computer algorithms for iterative proportional fitting can perform these calculations with little difficulty. The general rule is that iterative proportional fitting for a hierarchical model for an incomplete multi-way table is the same as iterative proportional fitting for a complete table, except that the initial values for cell means are 1 if the cell mean is not assumed 0 and 0 if the cell mean is assumed 0.

Unfortunately, currently available algorithms using iterative proportional fitting do not provide parameter estimates and do not necessarily yield the correct degrees of freedom for chi-square tests. Thus these algorithms are less useful than in the case of complete factorial tables.

The Newton-Raphson Algorithm

Due to problems of identifiability of parameters, use of the Newton-Raphson algorithm requires care. The simplest approach involves use of sex S_h as a response variable and ages A_h and B_h as predictor variables. Thus

$$\log m_{ijk} = \alpha_{ij}^{AB} + \sum_{l=1}^{11} \beta_l x_{ijkl}, \qquad i \le j.$$

Table 7.11

Results of Iterative Proportional Fitting with the Quasi-Independence Model of Table 7.10[a]

Sex	Age at first marriage	Iteration	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71
Female	≤20	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1	5.50	33.50	36.00	66.50	45.00	42.00	27.50	16.50
		2	8.07	49.17	52.84	97.61	66.06	61.65	40.37	24.22
		3	9.00	44.45	55.33	99.70	65.95	62.80	37.47	25.16
		9	9.00	45.41	54.62	99.18	66.03	62.42	38.54	24.78
	21–25	0	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1	0.00	21.50	37.00	57.50	47.00	46.50	31.00	22.50
		2	0.00	20.52	35.31	54.88	44.86	44.38	29.59	21.47
		3	0.00	18.55	36.97	56.05	44.78	45.21	27.46	22.31
		9	0.00	17.59	37.59	56.44	44.68	45.32	27.00	22.39
	26–30	0	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
		1	0.00	0.00	3.00	7.00	12.50	17.00	17.50	13.00
		2	0.00	0.00	1.63	3.80	6.79	9.23	9.50	7.06
		3	0.00	0.00	1.71	3.88	6.77	9.40	8.82	7.33
		9	0.00	0.00	1.79	3.97	6.77	9.54	8.42	7.51
	≥31	0	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
		1	0.00	0.00	0.00	2.50	6.50	12.00	10.50	7.50
		2	0.00	0.00	0.00	1.35	3.50	6.46	5.65	4.04
		3	0.00	0.00	0.00	1.37	3.49	6.58	5.25	4.20
		9	0.00	0.00	0.00	1.41	3.51	6.71	5.04	4.32
Male	≤20	0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1	5.50	33.50	36.00	66.50	45.00	42.00	27.50	16.50
		2	2.93	17.83	19.16	35.39	23.94	22.35	14.63	8.78
		3	2.00	20.79	17.86	34.30	23.99	21.90	15.65	8.47
		9	2.00	21.58	17.39	33.83	23.97	21.59	16.42	8.23
	21–25	0	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		1	0.00	21.50	37.00	57.50	47.00	46.50	31.00	22.50
		2	0.00	22.48	38.69	60.12	49.14	48.62	32.41	23.53
		3	0.00	26.21	36.07	58.27	49.24	47.64	34.67	22.69
		9	0.00	25.42	36.40	58.55	49.32	47.68	35.00	22.61
	26–30	0	0.00	0.00	1.00	1.00	1.00	1.00	1.00	1.00
		1	0.00	0.00	3.00	7.00	12.50	17.00	17.50	13.00
		2	0.00	0.00	4.37	10.20	18.21	24.77	25.50	18.94
		3	0.00	0.00	4.08	9.89	18.25	24.27	27.27	18.27
		9	0.00	0.00	4.21	10.03	18.23	24.46	26.60	18.48
	≥31	0	0.00	0.00	0.00	1.00	1.00	1.00	1.00	1.00
		1	0.00	0.00	0.00	2.50	6.50	12.00	10.50	7.50
		2	0.00	0.00	0.00	3.65	9.50	17.54	15.35	10.96
		3	0.00	0.00	0.00	3.54	9.52	17.19	16.41	10.57
		9	0.00	0.00	0.00	3.59	9.49	17.28	15.97	10.68

[a] Entries are values of m_{ijkl}.

Here

$$\begin{aligned} \alpha_{ij}^{AB} &= \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, & \\ \beta_l &= \lambda_1^S, & l = 1, \\ &= \lambda_{(l-1)1}^{AS}, & 2 \le l \le 4, \\ &= \lambda_{(l-4)1}^{BS}, & 5 \le l \le 11. \end{aligned}$$

For $l = 1$,

$$\begin{aligned} x_{ijkl} &= 1, & k = 1, \\ &= -1, & k = 2. \end{aligned}$$

For $2 \le l \le 4$,

$$\begin{aligned} x_{ijkl} &= 1, & i = l - 1 \text{ and } k = 1 \text{ or } i = 4 \text{ and } k = 2, \\ &= 0, & i \ne l - 1, \ i \le 3, \\ &= -1, & i = 4 \text{ and } k = 1 \text{ or } i = l - 1 \text{ and } k = 2. \end{aligned}$$

For $5 \le l \le 11$,

$$\begin{aligned} x_{ijkl} &= 1, & j = l - 4 \text{ and } k = 1 \text{ or } j = 8 \text{ and } k = 2, \\ &= 0, & j \ne l - 4, \ j \le 7, \\ &= -1, & j = 8 \text{ and } k = 1 \text{ or } j = l - 4 \text{ and } k = 2. \end{aligned}$$

Results are summarized in Table 7.12 and in Table 7.13. Computations may be accomplished by use of the program in the Appendix or by use of Bock and Yates (1973). As described in Exercise 7.12, some rearrangement of subscripts is required for these programs. Since sex has only two values, algorithms for logit analysis may also be employed.

Goodness of Fit

The Pearson chi-square statistic X^2 and the likelihood-ratio chi-square statistic L^2 are readily calculated given approximations for $\hat{m}_{ijk}$. One finds that $X^2 = 13.6$ and $L^2 = 13.3$.

There are 52 cells of the table in which m_{ijk} is not necessarily 0. These are 26 coefficients α_{ij}^{AB} and 11 coefficients β_l, so there are $52 - 26 - 11 = 15$ degrees of freedom. Thus the observed fit is quite satisfactory. This appearance is also supported by the adjusted residuals in Table 7.12. These residuals are all modest in size, and no obvious patterns are present other than the algebraic identities $r_{ij1} = -r_{ij2}$ and $r_{221} = -r_{121} = -r_{222} = r_{122}$.

Inspection of the standardized values in Table 7.13 suggests that it may be possible to obtain a model consistent with the data in which the λ^{BS}-parameters are assumed 0. On the other hand, the standardized values for λ_{11}^{AS} and λ_{31}^{AS} are so large that removal of all λ^{AS}-parameters is not feasible.

Table 7.12

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Model of No Three-Factor Interaction for Table 7.10[a]

Sex	Age at first marriage	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71
Male	≤20	9.00	43.00	51.00	103.00	68.00	65.00	39.00	22.00
		9.00	45.42	54.62	99.18	66.04	62.42	38.56	24.78
		—	−1.04	−1.38	1.16	0.65	0.86	0.17	−1.34
	21–25	—	20.00	40.00	53.00	45.00	43.00	24.00	26.00
		—	17.58	37.60	56.44	44.68	45.32	26.99	22.39
		—	1.04	0.90	−1.03	0.10	−0.73	−1.08	1.52
	26–30	—	—	3.00	4.00	5.00	7.00	12.00	7.00
		—	—	1.79	3.97	6.78	9.54	8.41	7.51
		—	—	1.14	0.02	−0.94	−1.20	1.77	−0.28
	≥31	—	—	—	1.00	3.00	9.00	4.00	4.00
		—	—	—	1.41	3.51	6.72	5.03	4.32
		—	—	—	−0.43	−0.36	1.32	−0.64	−0.22
Female	≤20	2.00	24.00	21.00	30.00	22.00	19.00	16.00	11.00
		2.00	21.58	17.38	33.82	23.97	21.58	16.44	8.22
		—	1.04	1.38	−1.16	−0.65	−0.86	−0.17	1.34
	21–25	—	23.00	34.00	62.00	49.00	50.00	38.00	19.00
		—	25.42	36.40	58.56	49.32	47.68	35.01	22.61
		—	−1.04	−0.90	1.03	−0.10	0.73	1.08	−1.52
	26–30	—	—	3.00	10.00	20.00	27.00	23.00	19.00
		—	—	4.21	10.03	18.22	24.46	26.59	18.49
		—	—	−1.14	−0.02	0.94	1.20	−1.77	0.28
	≥31	—	—	—	4.00	10.00	15.00	17.00	11.00
		—	—	—	3.59	9.49	17.28	15.97	10.68
		—	—	—	0.43	0.36	−1.32	0.64	0.22

[a] First line is observed count, second line is estimated expected counts, and third line is adjusted residual.

Lack of estimates of λ^A-, λ^B-, and λ^{AB}-parameters may appear to result from the particular implementation of the Newton–Raphson algorithm that was employed; however, a much deeper problem is actually involved. These parameters are not uniquely defined by the log means $\log m_{ijk}$, $i \leq j$. The basic problem is that the 26 parameters $\alpha_{ij}^{AB} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$ cannot simultaneously determine the 32 independent parameters λ, λ_i^A, $1 \leq i \leq 3$, λ_j^B, $1 \leq j \leq 7$, and λ_{ij}^{AB}, $1 \leq i \leq 3$, $1 \leq j \leq 7$.

Table 7.13

Estimated Parameters for the Model of No Three-Factor Interaction for Table 7.10

Parameter	Estimate	EASD	Standardized value
λ_1^S	−0.109	0.065	−1.67
λ_{11}^{AS}	0.640	0.057	11.14
λ_{21}^{AS}	0.084	0.054	1.55
λ_{31}^{AS}	−0.361	0.081	−4.48
λ_{11}^{BS}	0.221	.0.345	0.64
λ_{21}^{BS}	−0.159	0.104	−1.53
λ_{31}^{BS}	0.041	0.095	0.44
λ_{41}^{BS}	0.007	0.080	0.08
λ_{51}^{BS}	−0.024	0.084	−0.29
λ_{61}^{BS}	−0.000	0.083	0.00
λ_{71}^{BS}	−0.105	0.092	−1.15

Despite the difficulty in estimating λ^{AB}-parameters, it is not difficult to use results of Section 7.2 and Exercise 7.8 to show that the λ_{ij}^{AB} parameters and the λ_{ijk}^{ABS} parameters cannot all be 0. If all such parameters were 0, then one would have

$$\log m_{ijk} = (\lambda + \lambda_k^S) + (\lambda_i^A + \lambda_{ik}^{AS}) + (\lambda_j^B + \lambda_{jk}^{BS}), \qquad i \le j.$$

The results of Section 7.2 imply that

$$\log m_{ij1} = (\lambda + \lambda_1^S) + (\lambda_i^A + \lambda_{i1}^{AS}) + (\lambda_j^B + \lambda_{j1}^{BS}), \qquad i < j,$$

is not a viable model for the observations n_{ij1}, $i < j$. Similarly, Exercise 7.8 implies that

$$\log m_{ij2} = (\lambda + \lambda_2^S) + (\lambda_i^A + \lambda_{i2}^{AS}) + (\lambda_j^B + \lambda_{j2}^{BS}), \qquad i < j,$$

is not a viable model for the observations n_{ij2}, $i < j$. Therefore, the more general model proposed for the n_{ijk}, $i \le j$, $1 \le k \le 2$, cannot be used. Thus no hierarchical log-linear model can be expected to hold in which the λ^{AB}-parameters are set to 0.

A Simpler Hierarchical Model

The only possibly viable hierarchical model simpler than the model of no three-factor interaction assumes that

$$\log m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^S + \lambda_{ij}^{AB} + \lambda_{ik}^{AS}, \qquad i \le j.$$

This model implies that given age A_h at marriage, current age B_h and sex S_h are independent. It is easily verified that the maximum likelihood estimates $\hat{m}_{ijk}$ for this model satisfy the equation

$$\hat{m}_{ijk} = n_{ij}^{AB} n_{ik}^{AS} / n_i^A.$$

(Note that $n_{ij}^{AB} = 0$ for $i > j$, so that $\hat{m}_{ijk} = 0$ for $i > j$.)

As with hierarchical models for complete tables, it suffices to show that for some $\hat{\lambda}$, $\hat{\lambda}_i^A$, $\hat{\lambda}_j^B$, $\hat{\lambda}_k^S$, $\hat{\lambda}_{ij}^{AB}$, and $\hat{\lambda}_{ik}^{AS}$,

$$\log \hat{m}_{ijk} = \hat{\lambda} + \hat{\lambda}_i^A + \hat{\lambda}_j^B + \hat{\lambda}_k^S + \hat{\lambda}_{ij}^{AB} + \hat{\lambda}_{ik}^{AS},$$

$$\hat{m}_{ij}^{AB} = n_{ij}^{AB}, \qquad \hat{m}_{ik}^{AS} = n_{ik}^{AS},$$

$$\sum_i \hat{\lambda}_i^A = \sum_j \hat{\lambda}_j^B = \sum_k \hat{\lambda}_k^S = \sum_i \hat{\lambda}_{ij}^{AB} = \sum_j \hat{\lambda}_{ij}^{AB} = \sum_i \hat{\lambda}_{ik}^{AS} = \sum_k \hat{\lambda}_{ik}^{AS} = 0.$$

To show that $\hat{m}_{ijk}$ has the required form, note that

$$\sum_k n_{ij}^{AB} n_{ik}^{AS} / n_i^A = n_{ij}^{AB}, \qquad \sum_j n_{ij}^{AB} n_{ik}^{AS} / n_i^A = n_{ik}^{AS},$$

and

$$\log\left(\frac{m_{ijk}^* m_{ij'k'}^*}{m_{ij'k}^* m_{ijk'}^*}\right) = 0,$$

where

$$\begin{aligned} m_{ijk}^* &= n_{ij}^{AB} n_{ik}^{AS} / n_i^A, & i \le j, \\ &= n_{ik}^{AS} / n_i^A, & i > j. \end{aligned}$$

As in Chapter 3 of Volume 1, it follows that for all i, j, and k, unique $\hat{\lambda}$, $\hat{\lambda}_i^A$, $\hat{\lambda}_j^B$, $\hat{\lambda}_k^S$, $\hat{\lambda}_{ij}^{AB}$, and $\hat{\lambda}_{ik}^{AS}$ exist such that the required constraints on $\hat{\lambda}$-parameters hold and

$$\log m_{ijk}^* = \hat{\lambda} + \hat{\lambda}_i^A + \hat{\lambda}_j^B + \hat{\lambda}_k^S + \hat{\lambda}_{ij}^{AB} + \hat{\lambda}_{ik}^{AS}.$$

Restriction to the case $i \le j$ leads to the conclusion that $\hat{m}_{ijk}$ satisfies all maximum likelihood equations. Note, however, that the choice of parameters $\hat{\lambda}$, $\hat{\lambda}_i^A$, $\hat{\lambda}_j^B$, and $\hat{\lambda}_{ij}^{AB}$ remains somewhat arbitrary. Many choices of m_{ijk}^* are possible.

Results for this model are summarized in Table 7.14 and Table 7.15. In Table 7.14, it is helpful to note that the adjusted residual

$$r_{ijk} = \frac{n_{ijk} - n_{ij}^{AB} n_{ik}^{AS} / n_i^A}{[(n_{ij}^{AB} n_{ik}^{AS} / n_i^A)(1 - n_{ij}^{AB} / n_i^A)(1 - n_{ik}^{AS} / n_i^A)]^{1/2}}, \qquad i \le j,$$

just as in the corresponding model for a complete three-way table. In Table 7.15, note that parameter estimates may be computed by use of the equation

$$\log(\hat{m}_{i81} / \hat{m}_{i82}) = 2(\hat{\lambda}_1^S + \hat{\lambda}_{i1}^{AS}) = \log(n_{i1}^{AS} / n_{i2}^{AS}).$$

Table 7.14

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Model of No Interaction of Sex and Current Age for Table 7.10

Sex	Age at first marriage	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71
Female	≤20	9.00	43.00	51.00	103.00	68.00	65.00	39.00	22.00
		8.07	49.17	52.84	97.62	66.06	61.65	40.37	24.22
		0.64	−1.82	−0.53	1.22	0.51	0.90	−0.44	−0.90
	21–25		20.00	40.00	53.00	45.00	43.00	24.00	26.00
			20.52	35.31	54.88	44.86	44.38	29.59	21.47
			−0.16	1.18	−0.40	0.03	−0.32	−1.51	1.41
	26–30	—	—	3.00	4.00	5.00	7.00	12.00	7.00
		—	—	1.63	3.80	6.79	9.23	9.50	7.06
				1.29	0.13	−0.89	−0.99	1.10	−0.03
	≥31	—	—	—	1.00	3.00	9.00	4.00	4.00
		—	—	—	1.35	3.50	6.46	5.65	4.04
		—	—	—	−0.36	−0.34	1.40	−0.95	−0.02
Male	≤20	2.00	24.00	21.00	30.00	22.00	19.00	16.00	11.00
		2.93	17.83	19.16	35.38	23.94	22.35	14.63	8.78
		−0.64	1.82	0.53	−1.22	−0.51	−0.90	0.44	0.90
	21–25	—	23.00	34.00	62.00	49.00	50.00	38.00	19.00
		—	22.48	38.69	60.12	49.14	48.62	32.41	23.53
			0.16	−1.18	0.40	−0.03	0.32	1.51	−1.41
	26–30	—	—	3.00	10.00	20.00	27.00	23.00	19.00
			—	4.37	10.20	18.21	24.77	25.50	18.94
			—	−1.29	−0.13	0.89	0.99	−1.10	0.03
	≥31	—	—	—	4.00	10.00	15.00	17.00	11.00
		—	—		3.65	9.50	17.54	15.35	10.96
			—	—	0.36	0.34	−1.40	0.95	0.02

Table 7.15

Estimated Parameters for the Model of No Interaction of Sex and Current Age for Table 7.10

Parameter	Estimate	EASD
λ_1^S	−0.133	0.043
λ_{11}^{AS}	0.640	0.055
λ_{21}^{AS}	0.087	0.053
λ_{31}^{AS}	−0.361	0.080
λ_{41}^{AS}	−0.366	0.100

Since the sum $\sum_i \hat{\lambda}_{i1}^{AS} = 0$, it follows that

$$\hat{\lambda}_1^S = \frac{1}{8}\sum_{i=1}^{4}\log\left(\frac{n_{i1}^{AS}}{n_{i2}^{AS}}\right).$$

Since $\log(n_{i1}^{AS}/n_{i2}^{AS})$ has estimated asymptotic variance $(1/n_{i1}^{AS} + 1/n_{i2}^{AS})$, $\hat{\lambda}_1^S$ has estimated asymptotic variance

$$s^2(\hat{\lambda}_1^S) = \frac{1}{64}\sum_{i=1}^{4}\left(\frac{1}{n_{i1}^{AS}} + \frac{1}{n_{i2}^{AS}}\right).$$

Given the formula for $\hat{\lambda}_1^S$, it follows that

$$\hat{\lambda}_{i1}^{AS} = \frac{1}{2}\log\left(\frac{n_{i1}^{AS}}{n_{i2}^{AS}}\right) - \hat{\lambda}_1^S = \frac{1}{8}\sum_{i'=1}^{4} c_{ii'}\log\left(\frac{n_{i1}^{AS}}{n_{i2}^{AS}}\right),$$

where $c_{ii'}$ is 3 if $i = i'$ and $c_{ii'}$ is -1 otherwise. Thus

$$\begin{aligned} s^2(\hat{\lambda}_{i1}^{AS}) &= \frac{1}{64}\sum_{i'=1}^{4} c_{ii'}^2\left(\frac{1}{n_{i'1}^{AS}} + \frac{1}{n_{i'2}^{AS}}\right) \\ &= \frac{1}{64}\sum_{i'=1}^{4}\left(\frac{1}{n_{i'1}^{AS}} + \frac{1}{n_{i'2}^{AS}}\right) + \frac{1}{8}\left(\frac{1}{n_{i1}^{AS}} + \frac{1}{n_{i2}^{AS}}\right). \end{aligned}$$

Since $X^2 = 17.9$, $L^2 = 17.6$, and there are 22 degrees of freedom, the fit remains quite satisfactory. Note that removal of the 7 independent parameters λ_{j1}^{BS}, $1 \le j \le 7$, has resulted in an increase of L^2 of only 4.30. It also should be noted that the $\hat{\lambda}_{i1}^{AS}$ decrease in a fairly regular fashion as i increases. Possible exploitation of this fact is considered in the next section.

To interpret the estimates $\hat{\lambda}_{11}^{AS}$, note that the log cross-product ratio

$$\begin{aligned} \tau_{(ii')(12)\cdot j}^{AS\cdot B} &= \tau_{ii'\cdot j1}^{A\cdot BS} - \tau_{ii'\cdot j2} \\ &= \log m_{ij1} - \log m_{i'j1} - \log m_{ij2} + \log m_{i'j2} \\ &= \lambda_{i1}^{AS} - \lambda_{i'1}^{AS} - \lambda_{i2}^{AS} + \lambda_{i'2}^{AS} \\ &= 2(\lambda_{i1}^{AS} - \lambda_{i'1}^{AS}), \qquad i \le j, \quad i' \le j, \end{aligned}$$

measures the difference between the relative log odds of first marriage of females at age i rather i' given current age j and the corresponding log odds for males of current age j. This difference is independent of j and is estimated to be rather substantial. Note that in the most extreme case, for $j \ge 4$,

$$\hat{\tau}_{(14)(12)\cdot j}^{AS\cdot B} = 2(\hat{\lambda}_{11}^{AS} - \hat{\lambda}_{41}^{AS}) = 2.013.$$

This estimate corresponds to multiplication of the odds ratio by $e^{2.013} = 7.49$. The overall pattern suggested by the decreasing values of $\hat{\lambda}_{i1}^{AS}$ is that for all current ages, men tend to have their first marriage later than women. Some check on the regularity of this phenomenon is made in the next section where the hypothesis is tested that the λ_{i1}^{AS} are linear in i.

Degrees of Freedom for Hierarchical Models for Incomplete Multi-Way Tables

The reader should note that the correct degrees of freedom for chi-square statistics are often far from obvious when hierarchical models for multi-way tables are employed. Haberman (1974a, pp. 235–244) considers this problem in detail. If all λ-parameters in a hierarchical model are uniquely defined by those cell means which are positive, then the degrees of freedom are the degrees of freedom for the corresponding hierarchical model for a complete table minus the number of cells with means of 0. For example, the complete 4×4 table corresponding to Table 7.2 has $(4 - 1) \times (4 - 1) = 9$ degrees of freedom under the additive model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B,$$
$$\sum \lambda_i^A = \sum \lambda_j^B = 0.$$

There are 4 means set to 0, so there are $9 - 4 = 5$ degrees of freedom. The rule used here is essentially the only rule used in Haberman (1973b) to compute degrees of freedom. Unfortunately, it is not uncommon for this rule to be inadequate. Indeed, there are failures encountered in Sections 7.2 and 7.3. For example, in Section 7.3, the model of no three-factor interaction has $(4 - 1)(8 - 1)(2 - 1) = 21$ degrees of freedom when applied to a complete table. There are 12 cells with means of 0, so the degrees of freedom should be $21 - 12 = 9$. However, the correct figure is 15.

Simple improvements on the procedure of subtracting the number of empty cells can be made. For example, one can infer the 15 degrees of freedom in the model of no three-factor interaction by adding 6, the number of m_{ij}^{AB} equal 0, to the difference $21 - 12 = 9$. Goodman (1968) and Bishop, Fienberg, and Holland (1975, pp. 115–116, 218–219) provide some rules for computation of degrees of freedom. Unfortunately, no simple rules for degrees of freedom appear to be entirely adequate when the generating class of the hierarchical model has at least three members. The only procedures that will always work are based on algorithms for efficient solution of simultaneous linear equations. They are far beyond the scope of this book. Haberman (1974a) does describe a numerical technique that will always work. The program in the Appendix does have provisions to provide correct degrees of freedom; however, it is not completely immune from rounding errors.

It should also be noted that if the λ-parameters are not uniquely determined by the cell means which are positive, then interpretation of parameters may become quite complex. Treatments such as in Haberman (1974a, pp. 296–302) are far beyond the level of this book. Thus more caution should be exercised in treatment of incomplete tables by hierarchical models than is needed with complete tables.

7.4 Models for Incomplete Tables with Ordered Categories

Models for complete tables with ordered categories are readily adapted for use with incomplete tables with ordered categories. As an example, consider a model for Table 7.10 of the form

$$\log m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^S + \lambda_{ij}^{AB} + \gamma_{11}^{AS} q_i^A q_k^S, \qquad i \le j,$$

where $q_i^A = 2i - 5$ and $q_k^S = 2k - 3$ are scores for the categories of A_h and S_h. Thus the categories corresponding to age at first marriage have respective scores -3, -1, 1, and 3. Thus the scoring is linear. In the case of sex, males receive a score of 1 and females a score of -1. The model proposed simplifies the λ^{AS}-parameters in the hierarchical model

$$\log m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^S + \lambda_{ij}^{AB} + \lambda_{ik}^{AS},$$

for

$$\lambda_{ik}^{AS} = \gamma^{AS} q_i^A q_k^S,$$

so that $2\lambda^{AS}(q_i^A - q_{i'}^A)$ is the difference

$$\begin{aligned} \tau_{(ii')(12)\cdot j}^{AS\cdot B} &= \tau_{ii'\cdot j1}^{A\cdot BS} - \tau_{ii'\cdot j2}^{A\cdot BS} \\ &= \log m_{ij1} - \log m_{i'j1} - \log m_{ij2} + \log m_{i'j2} \end{aligned}$$

in the log odds of first marriage at age i rather than i' for females of current age j, $j \ge i$, $j \ge i'$, and the corresponding log odds for males of age j. Thus this difference is a linear function of the difference $i - i'$ between ranks of age groups. The Newton–Raphson algorithm is readily applied, especially if one writes

$$\log m_{ijk} = \alpha_{ij}^{AB} + \lambda_1^S q_k^S + \gamma_{11}^{AS} q_i^A q_k^S, \qquad i \le j,$$

so that $\alpha_{ij}^{AB} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$. The MULTIQUAL program of Bock and Yates (1973) has provisions for ordered classifications in incomplete multi-way tables. The program in the Appendix can also be used, although more effort is perhaps involved.

Results are summarized in Tables 7.16 and 7.17. Considered without respect to previous models, this model fits fairly well, for the Pearson chi-square statistic is 30.3, and the likelihood-ratio chi-square is 28.5, and there are 24 degrees of freedom. However, the model is not fully satisfactory. Note the large adjusted residuals for the combination of age of at least 31 at first marriage and current age of 51 to 60. Also note that a substantial increase in the likelihood ratio chi-square has resulted from the restriction made on the λ^{AS}-parameters. Without this restriction, the likelihood-ratio chi-square was 17.6 and there were 22 degrees of freedom. The change of

Table 7.16

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for Table 7.10 under the Model of Section 7.4

Sex	Age of first marriage	Current age ≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71
Female	≤20	9.00	43.00	51.00	103.00	68.00	65.00	39.00	22.00
		7.81	47.60	57.15	94.48	63.94	59.67	39.07	23.44
		0.80	−1.30	−0.04	1.82	1.01	1.37	−0.02	−0.57
	21–25	—	20.00	40.00	53.00	45.00	43.00	24.00	26.00
		—	21.90	37.69	58.57	47.87	47.36	31.58	22.92
		—	−0.59	0.56	−1.10	−0.62	−0.95	−1.98	0.94
	26–30	—	—	3.00	4.00	5.00	7.00	12.00	7.00
		—	—	1.83	4.27	7.63	10.38	10.68	7.93
		—	—	1.04	−0.16	−1.18	−1.32	0.51	−0.41
	≥31	—	—		1.00	3.00	9.00	4.00	4.00
		—	—	—	0.78	2.04	3.76	3.29	2.35
		—	—	—	0.27	0.76	3.11	0.45	1.21
Male	≤20	2.00	24.00	21.00	30.00	22.00	19.00	16.00	11.00
		3.19	19.40	20.85	38.52	26.06	24.33	15.93	9.56
		−0.80	1.30	0.04	−1.82	−1.01	−1.37	0.02	0.57
	21–25	—	23.00	34.00	62.00	49.00	50.00	38.00	19.00
		—	21.10	36.31	56.43	46.13	45.64	30.42	22.08
		—	0.59	−0.56	1.10	0.62	0.95	1.98	−0.94
	26–30	—	—	3.00	10.00	20.00	27.00	23.00	19.00
		—	—	4.17	9.73	17.37	23.62	24.32	18.07
		—	—	−1.04	0.16	1.18	1.32	−0.51	0.41
	≥31	—		—	4.00	10.00	15.00	17.00	11.00
		—	—	—	4.22	10.96	20.24	17.71	12.65
			—	—	−0.27	−0.76	−3.11	−0.45	−1.21

Table 7.17

Estimated Parameters for the Model of Section 7.4

Parameter	Estimate	EASD
λ_1^S	−0.196	0.041
γ_{11}^{AS}	0.215	0.019

28.5 − 17.6 = 10.9 in the likelihood-ratio chi-square involves 24 − 22 = 2 degrees of freedom, so it is significant at about the 0.004 level. Thus $\tau^{AS\cdot B}_{(ii')(1\,2)\cdot j}$ does not appear to be a linear function of $i - i'$. Nonetheless, results of Section 7.3 yield no indication that $\tau^{AS\cdot B}_{(ii')(1\,2)\cdot j}$ depends on j, provided j is as great as the maximum of i and i'.

EXERCISES

7.1. In Section 7.1, show that (7.1) and (7.2) imply that

$$p^{A\cdot B}_{i\cdot j}/(1 - p^{A\cdot B}_{j\cdot j}) = c_i/(1 - c_j), \qquad i \neq j,$$

where

$$c_i = \exp(\lambda^A_i) \Big/ \sum_{i'} \exp(\lambda^A_{i'}).$$

Solution

Let

$$b_i = \exp(\lambda^A_i)$$

and

$$d_j = \exp(\lambda + \lambda^B_j),$$

so that

$$c_i = b_i \Big/ \sum_{i'} b_{i'}$$

and

$$\begin{aligned} m_{ij} &= c_i d_j, \qquad & i \neq j, \\ &= 0, & i = j. \end{aligned}$$

Since

$$n^B_j = \sum_{i'} m_{i'j} = d_j\left(\sum_{i' \neq j} b_{i'}\right),$$

it follows that

$$m_{ij} = n^B_j b_i \Big/ \sum_{i' \neq j} b_{i'}, \qquad i \neq j,$$

and

$$p^{AB}_{i\cdot j}/(1 - p^{A\cdot B}_{j\cdot j}) = m_{ij}/n^B_j = b_i \Big/ \sum_{i' \neq j} b_{i'}, \qquad i \neq j.$$

Since

$$\sum_{i' \neq j} b_{i'} = \sum_{i'} b_{i'} - b_j,$$

it follows that

$$p_{i \cdot j}^{A \cdot B}/(1 - p_{j \cdot j}^{A \cdot B}) = c_i/(1 - c_j)$$

7.2. Verify that an $r \times s$ table is connected if for some (i, j) in I, (i, j) is connected to every other (k, l) in I.

Solution

Consider (k_1, l_1) in I and (k_2, l_2) in I. There exist (i_t, j_t), $1 \leq t \leq u$, in I such that $(i_1, j_1) = (i, j)$, $(i_u, j_u) = (k_1, l_1)$, and for $1 \leq t \leq u - 1$, either $i_t = i_{t+1}$ or $j_t = j_{t+1}$. There also exist (i'_t, j'_t), $1 \leq t \leq u'$, in I such that $(i'_1, j'_1) = (i, j)$, $(i'_{u'}, j'_{u'}) = (k_2, l_2)$, and for $1 \leq t \leq u' - 1$, $i'_t = i'_{t+1}$ or $j'_t = j'_{t+1}$. Let

$$\begin{aligned} (i_t^*, j_t^*) &= (i_{u-t+1}, j_{u-t+1}), \qquad & 1 \leq t \leq u, \\ &= (i'_{t-u}, j'_{t-u}), & u < t \leq u + u'. \end{aligned}$$

Then (i_t^*, j_t^*), $1 \leq t \leq u + u^*$, is a suitable sequence for connection of (k_1, l_1) and (k_2, l_2). Note that each (i_t^*, j_t^*) is in I, $(i_1^*, j_1^*) = (k_1, l_1)$, $(i_{u+u'}^*, j_{u+u'}^*) = (k_2, l_2)$, and either $i_t^* = i_{t+1}^*$ or $j_t^* = j_{t+1}^*$ for $1 \leq t < u + u'$.

7.3. Show that Table 7.18 is separable (not inseparable). Here an X indicates that (i, j) is not in I.

Table 7.18

A Separable Table

	j			
i	1	2	3	4
1	X	n_{12}	X	n_{14}
2	n_{21}	X	n_{23}	X
3	X	n_{32}	X	n_{34}
4	n_{41}	X	n_{43}	X

Solution

Consider the following diagram of cells connected to (2, 1) (Table 7.19). Note that (1, 2), (1, 4), (3, 2), and (3, 4) are not connected to (2, 1).

Table 7.19

Cells in Table 7.18 Connected to (2, 1)[a]

	j			
i	1	1	3	4
1	*X*		*X*	
2	•	*X*	•	*X*
3	*X*		*X*	
4	•	*X*	•	*X*

[a] Note: a similar separable table is described in Harris (1910) and later in Goodman (1968).

7.4. Show that Table 7.20 is inseparable. As in Exercise 7.3, an X indicates that (i, j) is not in I.

Table 7.20

An Inseparable Table

	j			
i	1	2	3	4
1	X	n_{12}	X	n_{14}
2	n_{21}	n_{22}	n_{23}	X
3	X	n_{32}	X	n_{34}
4	n_{41}	X	n_{43}	X

Solution

Consider Table 7.21. Note that all cells are connected to (2, 1).

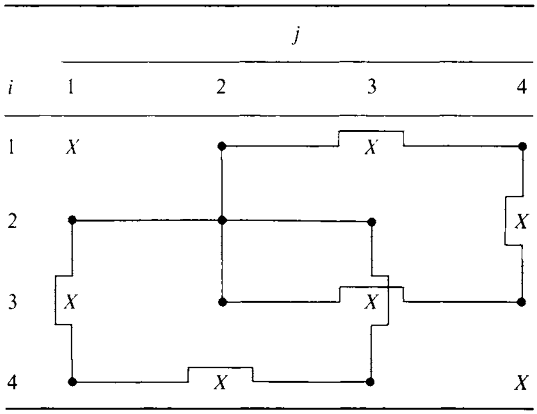
7.5. Find the degrees of freedom in a quasi-independence model for Table 7.21.

Solution

Here $q = 8$, $a = 4$, $b = 4$, and $s = 2$. Note that the table has two parts. The cells (2, 1), (2, 3), (4, 1), and (4, 3) are mutually connected, as are the cells (1, 2), (1, 4), (3, 2), and (3, 4). No cell in the first part is connected to a cell in the second part. Thus there are $8 - 4 - 4 + 2 = 2$ degrees of freedom.

Table 7.21

Cells in Table 7.20 Connected to (2, 1)



7.6. Show that under a quasi-independence model, Table 7.22 has one degree of freedom.

Table 7.22

A Table with Empty Rows and Columns

i	j = 1	2	3	4
1	X			X
2	X			X
3	X	X	X	X
4		X	X	X
5			X	X

Solution

Here $q = 7$, $a = 4$, $b = 3$, and $s = 1$. Note that (5, 1) is connected to (5, 2), (2, 2), (1, 2), (4, 1), (2, 3), and (3, 3). Thus there is $7 - 4 - 3 + 1 = 1$ degree of freedom.

7.7. Consider a model for Table 7.2 in which

$$\log m_{ij} = \alpha_j + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4},$$

where the x_{ijk}, $1 \le k \le 3$, are defined as in Table 7.4 and

$$\begin{aligned} x_{ij4} &= 1, && i = 2 \text{ and } j = 1 \quad \text{or} \quad i = 3 \text{ and } j = 4, \\ &= -1, && i = 2 \text{ and } j = 4 \quad \text{or} \quad i = 3 \text{ and } j = 1, \\ &= 0, && \text{otherwise.} \end{aligned}$$

In this model

$$\tau^{AB}_{(23)(14)} = \tau^{A\cdot B}_{23\cdot 1} - \tau^{A\cdot B}_{23\cdot 4} = 4\beta_4,$$

so that the apparently nonzero interaction noted in Section 7.1 can be accounted for. Find the Pearson chi-square, likelihood-ratio chi-square, degrees of freedom, maximum-likelihood estimates $\hat{m}_{ij}$, and adjusted residuals r_{ij} for this model.

Solution.

Results are summarized in Table 7.23. One has $X^2 = 1.44$ and $L^2 = 1.61$, and there are 4 degrees of freedom. Thus the fit is rather satisfactory.

Table 7.23

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Model of Exercise 7.7[a]

Residence at age 16	Current residence			
	Northeast	South	North Central	West
Northeast	—	22	14	13
		20.75	13.25	15.00
		0.45	0.30	−0.70
South	26	—	36	30
	25.62		37.15	29.23
	0.28		−0.43	0.44
North Central	10	41	—	46
	9.17	43.06		44.77
	0.64	−0.72		0.64
West	1	8	5	—
	2.20	7.20	4.60	
	−0.92	0.46	0.24	

[a] First line is observed value, second line is fitted value, and third line is adjusted residual.

7.8. Test the quasi-independence model

$$\log m_{ij} = \lambda + \lambda^A_i + \lambda^B_j, \qquad i < j,$$

using the data in Table 7.24.

Solution

Results are summarized in Table 7.25. Since $X^2 = 24.9$, $L^2 = 27.0$, and there are 12 degrees of freedom, substantial evidence exists that quasi-

Table 7.24

Ever-Married Men in 1974 General Social Survey, Cross-Classified by Age of Marriage and Current Age

Age at first marriage	Current age								
	≤20	21 25	26–30	31–40	41–50	51–60	61–70	≥71	Total
≤20	2	24	21	30	22	19	16	11	145
21–25	—	23	34	62	49	50	38	19	275
26–30	—	—	3	10	20	27	23	19	102
≥31	—	—	—	4	10	15	17	11	57
Total:	2	47	58	106	101	111	94	60	579

Table 7.25

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Quasi-Independence Model of Table 7.24[a]

Age at first marriage	Current age							
	≤20	21–25	26–30	31–40	41–50	51–60	61–70	≥71
≤20	—	24	21	30	32	19	16	11
	—	24.00	17.64	24.91	21.10	23.18	19.63	12.53
	—	—	1.05	1.32	0.24	−1.09	−1.01	−0.52
21–25	—	—	34	62	49	50	31	19
	—	—	37.36	52.76	44.67	49.10	41.58	26.54
	—	—	−1.05	2.08	0.97	0.20	−0.82	−2.09
26–30	—	—	—	10	20	27	23	19
	—	—	—	24.33	20.60	22.64	19.18	12.24
	—	—	—	−3.83	−0.17	1.17	1.09	2.32
≥31	—	—	—	—	10	15	17	11
	—	—	—	—	14.63	16.07	13.61	8.69
	—	—		—	−1.54	−0.35	1.15	0.93

[a] First line is observed count, second line is estimated expected counts, and third line is adjusted residual.

independence does not hold. Much of the lack of fit appears to result from the unexpected small number of men currently 31–40 who were married between ages 26 and 30. If the quasi-independence model is

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B, \qquad i < j, \quad (i, j) \neq (3, 4),$$

then $X^2 = 10.0$, $L^2 = 10.1$, and there are 11 degrees of freedom.

7.9. Use the iterative proportional fitting algorithm to find the maximum likelihood estimates listed in Table 7.8. Note that the m_{ij2} need not equal the

Table 7.26

Results of Iterative Proportional Fitting in the Quasi-Independence Model for Table 7.7[a]

Age at marriage	Iteration	Current age ≤ 20	21–25	26–30	31–40	41–50	51–60	61–70	≥ 71	Total m_{it}^A	Total n_i^A
≤ 20	1	0	1	1	1	1	1	1	1	7	391
	2	0	55.86	55.86	55.86	55.86	55.86	55.86	55.86	391.00	
	3	0	43.00	53.87	58.17	63.55	65.12	41.49	30.99	386.19	
21–25	1	0	0	1	1	1	1	1	1	6	231
	2	0	0	38.50	38.50	38.50	38.50	38.50	38.50	231.00	
	3	0	0	37.13	60.78	43.80	44.89	28.60	21.36	236.55	
26–30	1	0	0	0	1	1	1	1	1	5	35
	2	0	0	0	7	7	7	7	7	35.00	
	3	0	0	0	11.05	7.96	8.16	5.20	3.88	36.20	
≥ 31	1	0	0	0	0	1	1	1	1	4	20
	2	0	0	0	0	5.00	5.00	5.00	5.00	20.00	
	3	0	0	0	0	5.69	5.83	3.71	2.77	18.01	
Total m_j^B	1	0	1	2	3	4	4	4	4	22.00	
	2	0	55.86	94.36	101.36	106.36	106.34	106.36	106.36	677.0	
	3	0	43.00	91.00	160.00	121.00	124.00	79.00	59.00	677.00	
Total n_j^B		0	43	91	160	121	124	79	59		677

maximum likelihood estimates $\hat{m}_{ij}$ (see Haberman (1974a, pp. 295–296) for a similar result).

Solution

The three results m_{ij0}, m_{ij1}, and m_{ij2} are listed in Table 7.26. Note that one may let

$$m_{ij0} = 1, \qquad i < j,$$
$$= 0, \qquad i \geq j,$$

$$m_{ij1} = m_{ij0} n_i^A / m_{i0}^A, \qquad m_{ij2} = m_{ij1} n_j^B / m_{j1}^B, \qquad \text{etc.}$$

Here n_{ij} has been set to 0 if $i < j$.

7.10. Let a contingency table be defined as in Table 7.27. Consider the quasi-independence model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B, \qquad i \leq 2 \quad \text{or} \quad j \leq 2.$$

Show that if $n_i^A > 0$, $1 \leq i \leq 4$, $n_j^B > 0$, $1 \leq j \leq 4$, and $E > 0$, then

$$\begin{aligned} \hat{m}_{ij} &= E n_i^A n_j^B / (CD), && 1 \leq i \leq 2, \quad 1 \leq j \leq 2, \\ &= n_i^A n_j^B / C, && 1 \leq i \leq 2, \quad 3 \leq j \leq 4, \\ &= n_i^A n_j^B / D, && 3 \leq i \leq 4, \quad 1 \leq j \leq 2, \\ &= 0, && 3 \leq i \leq 4, \quad 3 \leq j \leq 4, \end{aligned}$$

Table 7.27

The Contingency Table of Exercise 7.10

	j			
i	1	2	3	4
1	n_{11}	n_{12}	n_{13}	n_{14}
2	n_{21}	n_{22}	n_{23}	n_{24}
3	n_{31}	n_{32}	—	—
4	n_{41}	n_{42}	—	—

where

$$C = n_1^A + n_2^A, \qquad D = n_1^B + n_2^B,$$

and

$$E = n_{11} + n_{12} + n_{21} + n_{22}.$$

Solution

Note that

$$\begin{aligned}
\hat{m}_i^A &= (E/C)n_i^A + [(C - E)/C]n_i^A = n_i^A, && 1 \le i \le 2, \\
&= n_i^A(D/D) = n_i^A, && 3 \le i \le 4, \\
\hat{m}_j^B &= (E/D)n_j^B + [(D - E)/D]n_j^B = n_j^B, && 1 \le j \le 2, \\
&= (C/C)n_j^B = n_j^B, && 3 \le j \le 4.
\end{aligned}$$

Let

$$\hat{\lambda} = \tfrac{1}{4} \sum_{i'=1}^{4} \log n_{i'}^A + \tfrac{1}{4} \sum_{j'=1}^{4} \log n_{j'}^B - \tfrac{1}{2} \log C - \tfrac{1}{2} \log D,$$

$$\hat{\lambda}_i^A = \log n_i^A - \tfrac{1}{4} \sum_{i'=1}^{4} \log n_{i'}^A + a(i)(\tfrac{1}{2} \log E - \tfrac{1}{2} \log C),$$

$$\hat{\lambda}_j^B = \log n_j^B - \tfrac{1}{4} \sum_{j'=1}^{4} \log n_{j'}^B + a(j)(\tfrac{1}{2} \log E - \tfrac{1}{2} \log D),$$

where $a(i) = 1$ for $1 \le i \le 2$ and $a(i) = -1$ for $3 \le i \le 4$. Note that

$$\sum_i \hat{\lambda}_i^A = \sum_j \hat{\lambda}_j^B = 0$$

and

$$\log \hat{m}_{ij} = \hat{\lambda} + \hat{\lambda}_i^A + \hat{\lambda}_j^B, \qquad i \le 2 \quad \text{or} \quad j \le 2.$$

Thus $\hat{m}_{ij}$ satisfies all requirements of a maximum likelihood estimate. Note that similar examples are fairly common in the literature. For instance, see Goodman (1968).

7.11. Consider the contingency table defined in Table 7.28. Show that under the quasi-independence model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B, \quad i \le 1 \text{ and } j \le 2 \text{ or } i \ge 3 \text{ and } j \ge 2,$$

$$\begin{aligned}
\hat{m}_{ij} &= n_i^A n_j^B / C, && i \le 2, j \le 1, \\
&= n_1^A (n_{12} + n_{22})/(n_1^A + n_2^A), && i \le 2, \quad j = 2, \\
&= n_i^A (n_{32} + n_{42})/(n_1^A + n_2^A), && i \ge 3, \quad j = 2, \\
&= n_i^A n_j^B / D, && i \ge 3, \quad j \ge 3, \\
&= 0, && \text{otherwise,}
\end{aligned}$$

where

$$C = n_1^A + n_2^A, \qquad D = n_3^A + n_4^A.$$

Table 7.28

The Contingency Table of Exercise 7.11

i	j = 1	2	3	4
1	n_{11}	n_{12}	—	—
2	n_{21}	n_{22}	—	—
3	—	n_{32}	n_{33}	n_{34}
4	—	n_{42}	n_{43}	n_{44}

Solution

The argument is very similar to that in Exercise 7.11.

7.12. Derive Table 7.12 and Table 7.13 by use of the Newton–Raphson algorithm.

Solution

Details require too much space to be described here. One may use the logit model

$$\omega_{ij} = \tau_{12\cdot ij}^{S\cdot AB} = 2\beta_1 + \sum_{l=2}^{11} 2\beta_l x_{ij1l}, \qquad i \le j,$$

as in Chapter 5 or the multinomial response form

$$\log m'_{i'j'} = \alpha'_{j'} + \sum_{l=1}^{11} \beta_l x'_{i'j'l}, \qquad 1 \le i' \le 2, \quad 1 \le j' \le 26.$$

Here the observed table $n'_{i'j'}$ is defined so that

$$\begin{aligned} n'_{i'j'} &= n_{ijk}, & i' &= k, \quad j' = a(i,j), \\ a(1,1) &= i+1, & 1 &\le i \le 2, \\ a(i,3) &= i+3, & 1 &\le i \le 3, \\ a(i,j) &= i+6+4(j-1), & 1 &\le i \le 4, \quad 4 \le j \le 8. \end{aligned}$$

The expected value $m'_{i'j'}$ is m_{ijk} and $x'_{i'j'}$ is x_{ijkl} if $i' = k$ and $j' = a(i,j)$.

7.13. White Christian subjects in the 1972, 1973, and 1974 General Social Surveys are classified in Table 7.29 by their responses to three questions on abortion. Let n_{ijkl} be the number of subjects h who in year $Y_h = l$ give responses $B_h = j$ to question B, $D_h = j$ to question D, and $F_h = k$ to question F, so that $n_{1111} = 334$, $n_{1121} = 34$, etc. Consider the following quasi-independence model for subjects with mixed responses:

$$\log m_{ijkl} = \lambda + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y, \qquad i \ne j \quad \text{or} \quad i \ne k,$$
$$\sum_i \lambda_i^B = \sum_j \lambda_j^D = \sum_k \lambda_k^F = \sum_l \lambda_l^Y = 0.$$

What does this model imply? Does it fit the data? Find approximate 95 percent confidence limits for λ_1^B, λ_1^D, and λ_1^F. Interpret these limits.

Solution

Either the Newton–Raphson algorithm or iterative proportional fitting may be used to obtain maximum likelihood estimates m_{ijkl}. In the case of iterative proportional fitting, set n_{ijkl} to 0 if $i = j = k$, let $m_{ijkl0} = 0$ if $i = j = k$, and let $m_{ijkl0} = 1$, otherwise. Let

$$m_{ijkl1} = n_i^B m_{ijkl0}/m_{i0}^B, \qquad m_{ijkl2} = n_j^D m_{ijkl1}/m_{j1}^D,$$
$$m_{ijkl3} = n_k^F m_{ijkl2}/m_{k2}^F, \qquad m_{ijkl4} = n_l^Y m_{ijkl3}/m_{l3}^Y, \qquad \text{etc.}$$

In the case of the Newton–Raphson algorithm, consider the model

$$\log m_{ijkl} = \alpha_l + \lambda_1^B q_i + \lambda_1^D q_j + \lambda_1^F q_h, \qquad i \ne j \quad \text{or} \quad i \ne k,$$

where $q_1 = 1$ and $q_2 = -1$. The Newton–Raphson algorithm has the advantage that it can be used to obtain adjusted residuals, estimates $\hat{\lambda}_1^B$, $\hat{\lambda}_1^D$, and $\hat{\lambda}_1^F$, and estimated asymptotic standard deviations.

Results are summarized in Table 7.30 and Table 7.31. Since $X^2 = 6.91$, $L^2 = 6.84$, and there are 12 degrees of freedom, the model fits quite well.

Table 7.29

White Christian Subjects in the 1972 to 1974 General Social Surveys, Cross-Classified by Year of Survey and Responses to Three Questions on Abortion Attitudes[a, b]

Year	Response to B	Response to D	Response to F	Observed count
1972	Yes	Yes	Yes	334
	Yes	Yes	No	34
	Yes	No	Yes	12
	Yes	No	No	15
	No	Yes	Yes	53
	No	Yes	No	63
	No	No	Yes	43
	No	No	No	501
1973	Yes	Yes	Yes	428
	Yes	Yes	No	29
	Yes	No	Yes	13
	Yes	No	No	17
	No	Yes	Yes	42
	No	Yes	No	53
	No	No	Yes	31
	No	No	No	453
1974	Yes	Yes	Yes	413
	Yes	Yes	No	29
	Yes	No	Yes	16
	Yes	No	No	18
	No	Yes	Yes	60
	No	Yes	No	57
	No	No	Yes	37
	No	No	No	430

[a] Data tapes for 1972, 1973, and 1974 General Social Surveys, National Opinion Research Center, University of Chicago.

[b] For questions used, see Table 6.14. Subjects are only included here if included in that table.

The model implies first that, given that not all responses are the same, responses are independent of year of survey. The model also implies that the log odds

$$\tau_{12\cdot jkl}^{B\cdot DFY} = \log(p_{1\cdot jkl}^{B\cdot DFY}/p_{2\cdot jkl}^{B\cdot DFY}) = 2\lambda_1^B$$

of response $B_h = 1$ rather than $B_h = 2$ given $D_h = j$, $F_h = k$, and $Y_h = l$ is independent of year l and is the same for $j = 1$ and $k = 2$ as for $j = 2$ and $k = 1$. Similar conclusions apply to $\tau_{12\cdot ikl}^{D\cdot BFY} = 2\lambda_1^D$ and $\tau_{12\cdot ijl}^{F\cdot BDY}$.

Table 7.30

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for Table 7.29 under Quasi-Independence

Year	Response to *B*	Response to *D*	Response to *F*	Observed count	Estimated expected count	Adjusted residual
1972	Yes	Yes	Yes	—	—	—
	Yes	Yes	No	34	29.27	1.06
	Yes	No	Yes	12	16.64	−1.28
	Yes	No	No	15	18.82	−1.00
	No	Yes	Yes	53	55.96	−0.54
	No	Yes	No	63	63.32	−0.06
	No	No	Yes	43	35.99	1.46
	No	No	No	—		—
1973	Yes	Yes	Yes	—	—	—
	Yes	Yes	No	29	24.61	1.05
	Yes	No	Yes	13	13.99	−0.29
	Yes	No	No	17	15.83	0.33
	No	Yes	Yes	42	47.06	−0.97
	No	Yes	No	53	53.25	−0.05
	No	No	Yes	31	30.26	0.16
	No	No	No	—	-	—
1974	Yes	Yes	Yes	—	—	—
	Yes	Yes	No	29	28.77	0.03
	Yes	No	Yes	16	16.41	−0.11
	Yes	No	No	18	18.57	−0.15
	No	Yes	Yes	60	55.20	0.88
	No	Yes	No	57	62.46	−0.96
	No	No	Yes	37	35.49	0.31
	No	No	No	—	—	—

Table 7.31

Estimated Parameters under the Quasi-Independence Model for Table 7.29

Parameter	Estimate	EASD	Lower limit of 95 percent confidence interval	Upper limit of 95 percent confidence interval
λ_1^B	−0.386	0.052	−0.488	−0.284
λ_1^D	0.221	0.051	0.184	0.321
λ_1^F	−0.062	0.048	−0.156	0.032

To aid in interpretation of confidence limits for λ_1^B, λ_1^D, and λ_1^F, note that corresponding confidence limits for the odds ratios

$$q_{12\cdot jkl}^{B\cdot DFY} = p_{1\cdot jkl}^{B\cdot DFY}/p_{2\cdot jkl}^{B\cdot DFY}, \qquad j \neq k,$$

are

$$\exp[2(-0.488)] = 0.377 \qquad \text{and} \qquad \exp[2(-0.284)] = 0.567.$$

Similar confidence limits for $q_{12\cdot ikl}^{D\cdot BFY}$, $i \neq k$, are 1.445 and 1.900. Limits for $q_{12\cdot ijl}^{F\cdot BDY}$, $i \neq j$, are 0.732 and 1.04. Thus the odds ratio for B, which involves a married woman who does not want more children, is somewhat less than 1, so that respondents are likely to oppose abortion for this grounds, given that they are divided on the other two questions. On the other hand, respondents tend to favor abortion due to economic hardships (D) given that responses to B and F differ. No preponderance in one direction or the other has been established in the case of F, which deals with an unmarried woman who does not want to marry the man.

7.14. An alternate analysis of Table 7.29 may be based on redefinition of variables. For related examples, see Bishop, Fienberg, and Holland (1975, pp. 225–228, 281–309), Bloomfield (1974), and Duncan (1975).

Let W_h be $7 - B_h - D_h - F_h$, and let X_h be $B_h - D_h + 2$. One then obtains the incomplete table shown in Table 7.32. Note that W_h provides an overall

Table 7.32

Table 7.29 Rearranged

Year	Number of positive responses to items B, D, and F ($W_h - 1$)	Relationship of responses to items B and D: Response to B positive and response to D negative ($X_h = 1$)	Responses the same ($X_h = 2$)	Response to B negative and response to D positive ($X_h = 3$)
1972	0	—	501	—
	1	15	43	63
	2	12	34	53
	3	—	334	—
1973	0	—	453	—
	1	17	31	53
	2	13	29	42
	3		482	—
1974	0	—	430	—
	1	18	37	57
	2	16	29	60
	3	—	413	—

measure of favorable attitudes toward legalized abortion and X_h reflects relative preference for abortions due to a desire of a married woman not to have more children rather than abortions due to economic hardship. Let n_{ijk} be the number of subjects with $W_h = i$, $X_h = j$, and $Y_h = k$.

Test the following quasi-independence models:

$$\text{Model 1:} \quad \log m_{ijk} = \lambda + \lambda_i^W + \lambda_i^X + \lambda_k^Y, \qquad 2 \le i \le 3 \quad \text{or} \quad j = 2,$$

$$\text{Model 2:} \quad \log m_{ijk} = \lambda + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_{ik}^{WY}, \qquad 2 \le i \le 3 \quad \text{or} \quad j = 2,$$

$$\text{Model 3:} \quad \log m_{ijk} = \lambda + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_{jk}^{XY}, \qquad 2 \le i \le 3 \quad \text{or} \quad j = 2,$$

$$\text{Model 4:} \quad \log m_{ijk} = \lambda + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \lambda_{ij}^{WX}, \qquad 2 \le i \le 3 \quad \text{or} \quad j = 2.$$

Based on results from these models, construct models with fewer independent parameters using scores

$$q_{i1}^Y = q_{i1}^X = i - 2, \qquad 1 \le i \le 3,$$
$$q_{i2}^Y = q_{i2}^X = 3(i-2)^2 - 2, \qquad 1 \le i \le 3,$$
$$q_{i1}^W = 2i - 5, \qquad 1 \le i \le 4,$$

$$\begin{aligned} q_{i2}^W &= 1, & i &= 1 \text{ or } 4, \\ &= -1, & 2 &\le i \le 3, \\ q_{i3}^W &= -1, & i &= 1, \\ &= 3, & i &= 2, \\ &= -3, & i &= 3, \\ &= 1, & i &= 4. \end{aligned}$$

Table 7.33

Chi-Square Statistics for Models for Table 7.32

Model number	x^2	L^2	Degrees of freedom d	L^2/d
1	26.48	26.71	16	1.67
2	3.42	3.48	10	0.35
3	22.21	22.39	12	1.87
4	26.16	26.45	12	2.20
5	12.81	12.80	15	0.85
6	17.33	17.18	17	1.01

Table 7.34

Parameter Estimates for Model 6 for Table 7.32

Parameter	Estimate	EASD	Standardized value
γ_1^W	−0.0287	0.0065	−4.39
γ_2^W	1.313	0.025	53.38
γ_1^X	0.606	0.054	11.30
γ_{11}^{WY}	0.0299	0.0080	3.72

Solution

Some results are summarized in Tables 7.33 and 7.34. Note that iterative proportional fitting or the Newton–Raphson algorithm may be used in computations for the four hierarchical models presented. As an example of models which use ordering of categories, consider

$$\text{Model 5:}\quad \log m_{ijk} = \lambda + \lambda_i^W + \lambda_j^X + \lambda_k^Y + \gamma_{11}^{WY} q_{11}^W q_{k1}^Y$$

$$= \alpha_k^Y + \sum_{i'=1}^{3} \gamma_{i'}^W q_{ii'}^W + \sum_{j'=1}^{2} \gamma_{j'}^X q_{jj'}^X + \gamma_{11}^{WY} q_i^W q_{k1}^Y,$$

$$2 \le i \le 3 \quad \text{or} \quad j = 2,$$

and

$$\text{Model 6:}\quad \log m_{ijk} = \alpha_k^Y + \sum_{i'=1}^{2} \gamma_{i'}^W q_{ii'}^W + \gamma_1^X q_{j1}^X + \gamma_{11}^{WY} q_{11}^W q_{k1}^W,$$

$$2 \le i \le 3 \quad \text{or} \quad j = 2.$$

Note that the only relationship between variables which is clearly present is between year of survey Y_h and number of positive responses $W_h - 1$, as is evident by comparison of likelihood-ratio chi-square statistics for Model 1 and Model 2. Although Model 5 and Model 6 both fit the data rather well, they do not yield the remarkably low chi-square statistic of Model 2.

Examination of parameter estimates from Model 6 shows that further simplification is not practical.

8 Symmetrical Tables

Multi-way contingency tables often arise in which two or more of the cross-classified variables have the same categories. In such cases, it may be helpful to consider models that assume various kinds of symmetrical relationships between variables. An assortment of such models are examined in this chapter. Section 8.1 considers symmetry, quasi-symmetry and marginal-homogeneity models for two-way tables, and Section 8.2 considers distance models for two-way tables. In Section 8.3, applications to higher-way tables are explored.

8.1 Symmetry, Quasi-Symmetry, and Marginal Homogeneity for Two-Way Tables

To illustrate use of symmetry models in two-way tables, consider Table 8.1. For subject h, $1 \leq h \leq N = 1055$, let A_h denote husband's highest degree, let B_h denote wife's highest degree, and let n_{ij} denote the number of subjects h with $A_h = i$ and $B_h = j$, so that $n_{11} = 259$, $n_{21} = 82$, etc. Assume that the $N = 1055$ pairs (A_h, B_h) are independently distributed with probability $p_{ij} > 0$ that $A_h = i$ and $B_h = j$. This assumption is approximate given the complex sampling method used, but it is probably adequate for most analysis.

Even a casual inspection of Table 8.1 indicates that degrees of husbands and wives are related. Formal tests of independence yield $X^2 = 505.3$ and $L^2 = 473.4$. Since there are 9 degrees of freedom, the evidence of dependence is overwhelmingly strong.

Table 8.1

Married Respondents in 1974 General Social Survey, Cross-Classified by Highest Degrees Attained[a,b]

Husband's highest degree	Wife's highest degree: Less than high school diploma	High school diploma or junior college degree	Bachelor's degree	Graduate degree	Total
Less than high school diploma	259	123	2	0	384
High school diploma or junior college degree	82	370	30	7	489
Bachelor's degree	5	59	34	4	102
Graduate degree	2	41	29	8	80
Total:	348	593	95	19	1055

[a] Data tape from 1974 General Social Survey of the National Opinion Research Center, University of Chicago.
[b] No cross-classification available on 10 respondents, and 419 respondents in survey not married.

Further casual inspection also shows that the table is not symmetric, in other words, it is not true that the cell means $m_{ij} = Np_{ij}$ satisfy the equations

$$m_{ij} = m_{ji}$$

for all i and j. The hypothesis of symmetry corresponds to the general log-linear model for $r \times r$ tables in which it is assumed that the cell means m_{ij} satisfy the equation

$$\log m_{ij} = \log m_{ji} = \alpha_{ij}, \qquad i \leq j.$$

In terms of the parametrization

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB},$$

where

$$\sum \lambda_i^A = \sum \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB},$$

the symmetry model assumes that

$$\lambda_i^A = \lambda_i^B, \qquad i \le i \le r,$$

and

$$\lambda_{ij}^{AB} = \lambda_{ji}^{AB}, \qquad i \le j.$$

Such a model can be used if the table of counts n_{ij}, $1 \le i \le r$, $1 \le j \le r$ has a multinomial distribution or if the counts have independent Poisson distributions.

Under this symmetry model, the maximum likelihood estimates $\hat{m}_{ij}$ of m_{ij} satisfy the equations

$$\log \hat{m}_{ij} = \log \hat{m}_{ji} = \hat{\alpha}_{ij}, \qquad i \le j,$$

$$\hat{m}_{ij} + \hat{m}_{ji} = n_{ij} + n_{ji}, \qquad i \le j,$$

as is readily verified given results of Section 2.6. Thus

$$\hat{m}_{ij} = \tfrac{1}{2}(n_{ij} + n_{ji}).$$

After some manipulation, the statistics X^2 and L^2 can be written

$$X^2 = \sum_{i=2}^{r} \sum_{j=1}^{i-1} \frac{(n_{ij} - n_{ji})^2}{(n_{ij} + n_{ji})}$$

and

$$L^2 = 2\sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij} \log\left(\frac{2n_{ij}}{n_{ij} + n_{ji}}\right)$$

$$= 2\sum_{i=2}^{r} \sum_{j=1}^{i-1} \{n_{ij} \log n_{ij} + n_{ji} \log n_{ji} - (n_{ij} + n_{ji}) \log[\tfrac{1}{2}(n_{ij} + n_{ji})]\}.$$

Since there are r^2 cells in the table and $r(r + 1)/2$ parameters α_{ij}, the number of degrees of freedom for these statistics is $r^2 - r(r + 1)/2 = r(r - 1)/2$. Since r in this example is 4, this number is 6. In the example under study, $X^2 = 64.0$ and $L^2 = 70.0$. Thus the symmetry hypothesis is not tenable.

The material presented here concerning symmetry models is rather old. McNemar (1947) obtained the formula for X^2 in the case in which r is 2. For general r, the formulas for $\hat{m}_{ij}$ and X^2 are due to Bowker (1948). Keats (1957) obtained the formula for L^2 in the case in which r is 2. The general formula for L^2 is found in Kullback (1968[1959], pp. 177–180). For some further details concerning the symmetry model, see Exercises 8.1 and 8.2.

The most obvious area in which the symmetry hypothesis fails is that of graduate degrees. Far more husbands than wives have graduate degrees. The adjusted residual r_{ij} for $i \ne j$ is readily shown to satisfy the equation

$$r_{ij} = (n_{ij} - n_{ji})/(n_{ij} + n_{ji})^{1/2}.$$

In this example, $r_{42} = 4.91$ and $r_{43} = 4.35$. The symmetry model also has failings when neither husband nor wife has a graduate degree. Note that $r_{21} = -2.86$ and $r_{32} = 3.7$.

Quasi-Symmetry

A more limited hypothesis can still be tried. In this model, it is only assumed that in the decomposition

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB},$$
$$\sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0,$$

one has

$$\lambda_{ij}^{AB} = \lambda_{ji}^{AB}$$

for all i and j. This model is termed a quasi-symmetry model by Caussinus (1965). It contrasts with the symmetry model in that the symmetry model assumes both that

$$\lambda_{ij}^{AB} = \lambda_{ji}^{AB}$$

and that

$$\lambda_i^A = \lambda_i^B.$$

The quasi-symmetry model has been used by Berger and Snell (1957) in a study of the theory of social mobility, but they do not consider formal estimation and testing procedures. The treatment here is derived from Caussinus (1965).

Under the quasi-symmetry model, maximum likelihood estimates $\hat{m}_{ij}$ are determined by the equations

$$\log \hat{m}_{ij} = \hat{\lambda} + \hat{\lambda}_i^A + \hat{\lambda}_j^B + \hat{\lambda}_{ij}^{AB},$$

$$\sum_i \hat{\lambda}_i^A = \sum_j \hat{\lambda}_j^B = \sum_i \hat{\lambda}_{ij}^{AB} = \sum_j \hat{\lambda}_{ij}^{AB} = 0,$$

$$\hat{\lambda}_{ij}^{AB} = \hat{\lambda}_{ji}^{AB}, \qquad \hat{m}_i^A = n_i^A, \qquad \hat{m}_j^B = n_j^B,$$

and

$$\hat{m}_{ij} + \hat{m}_{ji} = n_{ij} + n_{ji}.$$

The last equation implies that

$$\hat{m}_{ii} = n_{ii}.$$

As usual, n_i^A is the observed number of subjects with $A_h = i$, and n_j^B is the observed number of subjects with $B_h = j$. Corresponding definitions apply to $\hat{m}_i^A$ and $\hat{m}_j^B$. Verification of these equations is a bit tricky. For some details, see Exercise 8.3 or Caussinus (1965).

Iterative Proportional Fitting

A simple iterative algorithm due to Caussinus (1965) may be used to find the $\hat{m}_{ij}$. In this algorithm, which corresponds to the iterative proportional fitting algorithm of Chapters 3 and 4 of Volume 1 and Chapter 7 of this volume,

$$m_{iit} = n_{ii}$$

for each cycle $t \geq 0$, at the start of cycle 0,

$$m_{ij0} = 1, \qquad i \neq j,$$

and for $v \geq 0$ and $i \neq j$, cycles $3v$ to $3v + 2$ are defined by

$$m_{ij(3v+2)} = m_{ij(3v+1)}(n_j^B - n_{jj})/(m_{j(3v+1)}^B - n_{jj}),$$

$$m_{ij(3v+2)} = m_{ij(3v+1)}(n_j^B - n_{jj})/(m_{j(3v+1)}^B - n_{jj}),$$

$$m_{ij(3v+3)} = m_{ij(3v+2)}(n_{ij} + n_{ji})/(m_{ij(3v+2)} + m_{ji(3v+2)}).$$

It follows from general results of Darroch and Ratcliff (1972) or Haberman (1974a, pp. 65–68) that m_{ijt} converges to $\hat{m}_{ij}$ whenever $\hat{m}_{ij}$ exists. If the algorithm is used with Table 8.1, the results of Table 8.2 are obtained. The estimates $m_{ij(12)}$ are quite adequate approximation to the $\hat{m}_{ij}$.

As Bishop, Fienberg, and Holland (1975, p. 290) note, standard computer packages for iterative proportional fitting can be applied to the Caussinus algorithm by rewriting the table under study as an incomplete table. In one approach, a table of counts n'_{ijkl}, $1 \leq i \leq r$, $1 \leq j \leq r$, $1 \leq k \leq r$, $1 \leq l \leq r$, is defined so that

$$\begin{aligned} n'_{ijkl} &= n_{ij}, \qquad k = \min(i,j), \quad l = \max(i,j), \quad i \neq j, \\ &= 0, \qquad \text{otherwise.} \end{aligned}$$

Thus $n'_{1111} = 259$, $n'_{2112} = 82$, $n'_{1212} = 123$, etc. The quasi-symmetry model holds if and only if the mean m'_{ijkl} of n'_{ijkl} satisfies the model

$$\log m'_{ijkl} = \lambda' + \lambda_i'^A + \lambda_j'^B + \lambda_k'^C + \lambda_l'^D + \lambda_{kl}'^{CD}, \qquad k = \min(i,j), \quad l = \max(i,j), \quad i \neq j,$$

where

$$\sum \lambda_i'^A = \sum \lambda_j'^B = \sum \lambda_k'^C = \sum \lambda_l'^D = \sum_k \lambda_{kl}'^{CD} = \sum_l \lambda_{kl}'^{CD} = 0.$$

Table 8.2

Computation of Maximum Likelihood Estimates of Expected Cell Counts under the Quasi-Symmetry Model for Table 8.1

t	i	m_{i1t}	m_{i2t}	m_{i3t}	m_{i4t}
0	1	259	1	1	1
	2	1	370	1	1
	3	1	1	34	1
	4	1	1	1	8
1	1	259	41.67	41.67	41.67
	2	39.67	370	39.67	39.67
	3	22.67	22.67	34	22.67
	4	24.00	24.00	24.00	8
2	1	259	105.19	24.13	4.41
	2	40.98	370	22.97	4.20
	3	23.37	57.22	34	2.40
	4	24.74	60.59	13.90	8
3	1	259	147.61	3.56	0.30
	2	57.38	370	25.49	3.11
	3	3.44	63.51	34	4.85
	4	1.70	44.89	28.15	8
6	1	259	120.94	3.10	0.30
	2	84.06	370	31.70	5.17
	3	3.90	57.30	34	5.91
	4	1.70	42.83	27.09	8
12	1	259	121.67	3.05	0.28
	2	83.33	370	30.80	4.89
	3	3.95	58.20	34	5.83
	4	1.72	43.11	27.17	8

The iterative proportional fitting algorithm for this incomplete table corresponds to the Caussinus algorithm if

$$\begin{aligned} m'_{ijkl0} &= 1, \qquad k = \min(i,j), \quad l = \max(i,j), \quad i \neq j, \\ &= 0, \qquad \text{otherwise}, \\ m'_{ijkl1} &= m'_{ijkl0}\, n_i'^{A}/m_{i0}'^{A}, \\ m'_{ijkl2} &= m'_{ijkl1}\, n_j'^{B}/m_{j1}'^{B} \\ m'_{ijkl3} &= m'_{ijkl2}\, n_{kl}'^{CD}/m_{kl2}'^{CD}, \qquad \text{etc.} \end{aligned}$$

As usual, 0/0 is taken to be 0. Note that $n_i'^A = n_i^A$, $n_j'^B = n_j^B$, and $n_{kl}'^{CD} = n_{ij} + n_{ji}$ if $k < l$. One has

$$m'_{ijklt} = m_{ijt}$$

if $k = \min(i, j)$, $l = \max(i, j)$, and $i \neq j$.

The chi-square statistics X^2 and L^2 have respective values 2.75 and 3.02. To find degrees of freedom, note as in Exercise 8.3 that in an $r \times r$ table, there are r^2 cells, a parameter λ, $r - 1$ independent parameters λ_i^A, $1 \leq i \leq r - 1$, $r - 1$ independent parameters λ_j^B, $1 \leq j \leq r - 1$, and $r(r - 1)/2$ independent parameters λ_{ij}, $1 \leq i \leq j \leq r - 1$. Thus there are

$$r^2 - 1 - (r - 1) - (r - 1) - r(r - 1)/2 = (r - 1)(r - 2)/2$$

degrees of freedom. If $r = 4$, there are 3 degrees of freedom. Thus the quasi-symmetry model does fit Table 8.1 quite well.

The Newton–Raphson Algorithm

A number of possible implementations are available for the Newton–Raphson algorithm. The choice depends somewhat on the parameters of interest. If all λ-parameters except λ are to be considered, then one may let

$$\log m_{ij} = \lambda + \sum_{i'=1}^{r-1} \lambda_{i'}^A q_{ii'} + \sum_{j'=1}^{r-1} \lambda_{j'}^B q_{jj'} + \sum_{i'=1}^{r-1} \sum_{j'=1}^{i'} \lambda_{i'j'}^{AB}(q_{ii'} q_{jj'} + q_{ij'} q_{ji'}),$$

where for $1 \leq i \leq r$, $1 \leq i' = r - 1$,

$$\begin{aligned} q_{ii'} &= 1, \quad i = i', \\ &= 0, \quad i \neq i', \quad i < r, \\ &= -1, \quad i = r. \end{aligned}$$

This parametrization also appears in Exercise 8.3. Implementation of the Newton–Raphson algorithm for Table 8.1 is straightforward.

In the usual fashion, adjusted residuals and estimated asymptotic standard deviations are obtained, just as in Tables 8.3 and 8.4. Note that since $\hat{m}_{ii} = n_{ii}$, the adjusted residual r_{ii} is undefined. Also note that $r_{ij} = -r_{ji}$ for $i \neq j$. Details are left as an exercise for the reader (Exercise 8.4). The reader should note that the method presented here for derivation of Table 8.3 is not the most efficient one for this purpose since 12 simultaneous equations must be solved. An alternate parametrization presented in Exercise 8.5 involves only 3 simultaneous equations but does not lead immediately to the parameter estimates of Table 8.4. As is evident from the relatively small values of X^2 and L^2, and from the lack of large adjusted residuals, the quasi-symmetry model appears quite adequate. The sample size for women with graduate

Table 8.3

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Quasi-Symmetry Model for Table 8.1[a]

Husband's highest degree	Wife's highest degree: Less than high school diploma	High school diploma or junior college degree	Bachelor's degree	Graduate degree
Less than high school diploma	259 259.00 —	123 121.67 1.01	2 3.05 −0.84	0 0.28 −0.59
High school diploma or junior college degree	82 83.33 −1.01	370 370.00 –	30 30.78 −0.43	7 4.89 1.44
Bachelor's degree	5 3.95 −0.84	59 58.22 0.43	34 34.00 —	4 5.83 −1.25
Graduate degree	2 1.72 0.58	41 43.11 −1.44	29 27.17 1.25	8 8.00 –

[a] First line is observed count, second line is estimated expected count, and third line is adjusted residual.

Table 8.4

Parameter Estimates for the Quasi-Symmetry Model for Table 8.1

Parameter	Estimate	EASD
λ_1^A	−0.320	0.131
λ_2^A	0.964	0.125
λ_3^A	−0.192	0.136
λ_1^B	0.100	0.135
λ_2^B	1.762	0.125
λ_3^B	−0.310	0.139
λ_{11}^{AB}	1.452	0.154
λ_{21}^{AB}	0.486	0.121
λ_{22}^{AB}	0.157	0.068
λ_{31}^{AB}	−1.408	0.258
λ_{32}^{AB}	−0.380	0.133
λ_{33}^{AB}	0.438	0.105

degrees is rather small, so that some questions can be raised about accuracy of asymptotic approximations. Nonetheless, there seems no reason to believe the model does not fit the data.

The Meaning of Quasi-Symmetry

Interpretation of the quasi-symmetry model requires some care. The key observation to keep in mind is that quasi-symmetry holds if and only if for all integers i, j, k, and l between 1 and r,

$$q^{AB}_{(ij)(kl)} = \frac{m_{ik} m_{jl}}{m_{il} m_{jk}} = \frac{m_{ki} m_{lj}}{m_{li} m_{kj}} = q^{AB}_{(kl)(ij)}.$$

The necessity of the condition follows since under quasi-symmetry,

$$\begin{aligned}\log\left(\frac{m_{ik} m_{jl}}{m_{il} m_{jk}}\right) &= \lambda^{AB}_{ik} - \lambda^{AB}_{il} - \lambda^{AB}_{jk} + \lambda^{AB}_{jl} \\ &= \lambda^{AB}_{ki} - \lambda^{AB}_{li} - \lambda^{AB}_{kj} + \lambda^{AB}_{lj} \\ &= \log\left(\frac{m_{ki} m_{lj}}{m_{li} m_{kj}}\right).\end{aligned}$$

Sufficiency follows since if

$$\lambda^{AB}_{ik} - \lambda^{AB}_{il} - \lambda^{AB}_{jk} + \lambda^{AB}_{jl} = \lambda^{AB}_{ki} - \lambda^{AB}_{li} - \lambda^{AB}_{kj} + \lambda^{AB}_{lj},$$

then

$$r^{-2} \sum_{l=1}^{r} \sum_{j=1}^{r} (\lambda^{AB}_{ik} - \lambda^{AB}_{il} - \lambda^{AB}_{jk} + \lambda^{AB}_{jl}) = \lambda^{AB}_{ik} = \lambda^{AB}_{ki}.$$

Observe that $q^{B\cdot A}_{kl\cdot i} = m_{ik}/m_{il}$ is the odds that a husband with degree i has a wife with degree k, given that the wife has degree k or l. The corresponding odds ratio for a husband with degree j is $q^{B\cdot A}_{k\cdot j} = m_{jk}/m_{jl}$. The cross-product ratio

$$q^{AB}_{(ij)(kl)} = \frac{m_{ik} m_{jl}}{m_{il} m_{jk}}$$

is the ratio of these two sets of odds. The ratio $q^{A\cdot B}_{kl\cdot i} = m_{ki}/m_{li}$ is the odds that a wife with degree i has a husband with degree k, given that the husband has degree k or l. The corresponding odds ratio for a wife with degree j is $q^{A\cdot B}_{kl\cdot j} = m_{kj}/m_{lj}$. The cross-product ratio

$$q^{AB}_{(kl)(ij)} = \frac{m_{ki} m_{lj}}{m_{li} m_{kj}}$$

is the ratio of these two sets of odds.

For example, the maximum likelihood estimate of

$$q^{AB}_{(23)(12)} = \frac{m_{21} m_{32}}{m_{22} m_{31}}$$

is 3.32. Thus given that a wife has less than a bachelor's degree, it is estimated that the odds that she has less than a high school diploma are 3.32 times as great given that her husband has a high school diploma or junior college degree as they are given that her husband has a bachelor's degree.

Given that a husband has less than a bachelor's degree, it is also estimated that the odds that he has less than a high school diploma are 3.32 times as great given that his wife has a high school diploma or junior college degree as they are given that his wife has a bachelor's degree. Thus the model implies a type of symmetry in the relationship between degrees of husbands and wives.

To use the parameter estimates $\hat{\lambda}_{ij}^{AB}$, note that $\lambda_{ij}^{AB} = \lambda_{ji}^{AB}$ is both the average

$$\frac{1}{r^2}\sum_k \sum_l \tau_{(ik)(jl)}^{AB}$$

of all log cross-product ratios

$$\tau_{(ik)(jl)}^{AB} = \log q_{(ik)(jl)}^{AB},$$

and the average

$$\frac{1}{r^2}\sum_k \sum_l \tau_{(jl)(ik)}^{AB}$$

of all log cross-product ratios $\tau_{(jl)(ik)}^{AB}$.

Since $\tau_{(ii)(jl)}^{AB} = \tau_{(ik)(jj)}^{AB} = 0$, $[r/(r-1)]^2\lambda_{ij}^{AB}$ is the average

$$\left(\frac{1}{r-1}\right)^2 \sum_{k \neq i} \sum_{l \neq j} \tau_{(ik)(jl)}^{AB}$$

of the nontrivial log cross-product ratios $\tau_{(ik)(jl)}^{AB}$, $k \neq i$, $l \neq j$. Thus

$$\exp\left[\left(\frac{r}{r-1}\right)^2 \lambda_{ij}^{AB}\right]$$

is the geometric mean of the cross-product ratios

$$q_{(jl)(ik)}^{AB} = q_{(ik)(jl)}^{AB}, \qquad k \neq i, \quad l \neq j.$$

For example, consider λ_{11}^{AB}. The estimated value of λ_{11}^{AB} is 1.452, and an approximate 95 percent confidence interval for λ_{11}^{AB} has bounds

$$1.452 - 1.96(0.154) = 1.150$$

and

$$1.452 + 1.96(0.154) = 1.763.$$

The corresponding estimate

$$\exp[(\tfrac{4}{3})^2 \hat{\lambda}_{11}^{AB}]$$

is 13.21, and 95 percent confidence bounds for $\exp[(\frac{4}{3})^2\lambda_{11}^{AB}]$ are approximately

$$\exp[(\tfrac{4}{3})^2 1.150] = 7.72$$

and

$$\exp[(\tfrac{4}{3})^2 1.753] = 22.57.$$

Thus there is a very clear tendency for coefficients $q_{(1k)(1l)}^{AB} = q_{1k\cdot 1}^{B\cdot A}/q_{1k\cdot l}^{B\cdot A} = q_{1k\cdot 1}^{A\cdot B}/q_{1k\cdot l}^{A\cdot B}$, $k > 1$, $l > 1$, to exceed 1. This coefficient corresponds to a strong tendency for individuals with less than a high school diploma to have spouses with less than a high school diploma. The positive coefficients estimates $\hat{\lambda}_{22}^{AB}$ and $\hat{\lambda}_{33}^{AB}$ have similar meanings.

Quasi-Symmetry and Quasi-Independence

The quasi-independence model in which diagonal cells are ignored implies the quasi-symmetry model. If

$$\log m_{ij} = \kappa + \kappa_i^A + \kappa_j^B, \qquad i \neq j,$$
$$\sum \kappa_i^A = \sum \kappa_j^B = 0,$$

then

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB},$$
$$\sum_i \lambda_i^A = \sum_j \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0,$$
$$\lambda_{ij}^{AB} = \lambda_{ji}^{AB},$$

where

$$\lambda = \left(\frac{r-1}{r}\right)\kappa + \frac{1}{r^2}\sum_{i'} \log m_{i'j},$$

$$\lambda_i^A = \left(\frac{r-1}{r}\right)\kappa_i^A + \frac{1}{r}\log m_{ii} - \frac{1}{r^2}\sum_{i'}\log m_{i'i'},$$

$$\lambda_j^B = \left(\frac{r-1}{r}\right)\kappa_j^B + \frac{1}{r}\log m_{jj} - \frac{1}{r^2}\sum_{i'}\log m_{i'i'},$$

$$\lambda_{ij}^{AB} = \frac{1}{r}(\kappa + \kappa_i^A + \kappa_j^B) - \frac{1}{r}\log m_{ii} - \frac{1}{r}\log m_{jj} + \frac{1}{r^2}\sum_{i'}\log m_{i'i'}, \quad i \neq j,$$

$$= -\left(\frac{r-1}{r}\right)(\kappa + \kappa_i^A + \kappa_j^B) + \left(1 - \frac{2}{r}\right)\log m_{ii} + \frac{1}{r^2}\sum_{i'}\log m_{i'i'}, \quad i = j.$$

In Table 8.1, the quasi-independence model does not apply, for the Pearson chi-square $X^2 = 122.5$, the likelihood-ratio chi-square $L^2 = 91.2$, and there are 5 degrees of freedom. Thus quasi-symmetry provides a far better model than quasi-independence.

Quasi-Symmetry and Ordered Categories

In Section 6.1, models for tables such as Table 8.1 are considered in which

$$\lambda_{ij}^{AB} = \gamma^{AB} u_i t_j, \qquad \text{where} \qquad \sum u_i = \sum t_j = 0.$$

If $u_i = t_i$, then this model implies the quasi-symmetry model. Unfortunately, this model for ordered categories is not promising in Table 8.1. With $u_1 = t_1 = 3$, $u_2 = t_2 = 1$, $u_3 = t_3 = -1$, and $u_4 = t_4 = -3$, one finds a Pearson chi-square of 104.0, a likelihood-ratio chi-square of 70.3, and 8 degrees of freedom. Thus the previously introduced method to exploit ordered categories is not very helpful here. Other methods for ordered categories with symmetric tables are considered in Section 8.2.

Quasi-Independence, Symmetry, and Marginal Homogeneity

This section has emphasized symmetry and quasi-symmetry models. Other treatments of square contingency tables also deal with the hypothesis of marginal homogeneity. Under this hypothesis,

$$m_i^A = m_i^B, \qquad \text{or equivalently,} \qquad p_i^A = p_i^B.$$

In Table 8.1, this model says that the marginal distribution of degrees attained by husbands is the same as the marginal distribution of degrees attained by wives. If r is 2, marginal homogeneity and symmetry are equivalent, for

$$m_1^A = m_{12} + m_{12} = m_{11} + m_{21} = m_1^B$$

if and only if $m_{12} = m_{21}$. McNemar's (1947) chi-square test uses this observation for 2×2 tables. The case in which r is 2 is also unusual in that an exact test of marginal homogeneity or symmetry is available. The essential observation is that conditional on $n_{12} + n_{21}$, n_{12} has a binomial distribution with sample size $n_{12} + n_{21}$ and probability $m_{12}/(m_{12} + m_{21})$. This observation can be found in Mosteller (1947). The conditional significance level of $X^2 = (n_{12} - n_{21})^2/(n_{12} + n_{21})$ is then the probability that $Y \leq \min(n_{12}, n_{21})$, where Y has a binomial distribution with probability $\frac{1}{2}$ and sample size $n_{12} + n_{21}$.

Cochran (1950) notes that a continuity correction for X^2 is available in the form

$$X^2 = (|n_{12} - n_{21}| - 1)^2/(n_{12} + n_{21}).$$

This continuity correction generally improves the normal approximation to the significance level of the exact test. The approximation is satisfactory for $n_{12} + n_{21}$ greater than 30.

If r exceeds 2, then quasi-symmetry and marginal homogeneity are distinct hypotheses. As Caussinus (1965) notes, symmetry holds if and only if both quasi-symmetry and marginal homogeneity hold. To test for marginal homogeneity, one may then compute the likelihood ratio chi-square statistic L_1^2 for the hypothesis that quasi-symmetry holds and compute the likelihood-ratio chi-square statistic L_2^2 for the hypothesis that symmetry holds. The difference $L_1^2 - L_2^2$ has an approximate chi-square distribution with $r - 1$ degrees of freedom if the marginal homogeneity hypothesis holds and if the quasi-independence model is approximately true. In the specific example under study, $r - 1$ is 3 and $L_1^2 - L_2^2 = 70.0 - 3.0 = 67.0$. Thus the evidence against marginal homogeneity is very strong.

The preceding test for marginal homogeneity is most attractive if the quasi-symmetry model is true or is nearly true. If the quasi-symmetry model provides a poor fit to the observed data, then other testing procedures are more attractive. The simplest alternative test procedure is discussed by Stuart (1955). To implement the Stuart test, let

$$f_i = n_i^A - n_i^B, \qquad 1 \le i \le r - 1,$$

and for $1 \le i \le r - 1$, $1 \le j \le r - 1$, let

$$\begin{aligned} v_{ij} &= n_i^A + n_i^B - 2n_{ii}, \qquad & i = j, \\ &= -(n_{ij} + n_{ji}), \qquad & i \ne j. \end{aligned}$$

Let the inverse of $V = \{v_{ij}\}$ be denoted by $V^{-1} = \{v^{ij}\}$. If the m_{ij} become large and if marginal homogeneity holds, the distribution of

$$Q = \sum_{i=1}^{r-1} \sum_{j=1}^{r-1} f_i f_j v^{ij}$$

is approximated by a chi-square distribution with $r - 1$ degrees of freedom. The statistic Q may appear to depend on the way in which categories are numbered, but such is not the case. In the example under study, Q is 60.6 (see Exercise 8.8), so that the same conclusions may be drawn from this test as from the use of $L_1^2 - L_2^2$. A slight variation on the statistic Q is proposed by Bhapkar (1966). Ireland, Ku, and Kullback (1969) note that Bhapkar's statistic is equal to $Q/(1 - Q/n)$. Whether the chi-square approximation for Q is better than the chi-square approximation for $Q/(1 - Q/n)$ appears to be unknown. Madansky (1963) considers likelihood ratio tests for marginal homogeneity. Kullback (1968[1959], pp. 177–180) and Ireland, Ku, and Kullback (1969) treat square contingency tables from the point of view of information theory. The approach in Ireland, Ku, and Kullback (1969) is closely related to the methods for survey adjustment of Chapter 9.

8.2 Distance Models

Distance models are a class of models which are appropriate in square contingency tables in which the categories of the row and column variables are ordered. They have been used with social-mobility tables by Goodman (1972) and Haberman (1974a, pp. 215–227). In this section, a few examples of such models are applied to Table 8.1.

A very simple distance model assumes that

$$\log m_{ij} = \lambda' + \lambda_i'^A + \lambda_j'^B - \eta|i - j|,$$

where

$$\sum \lambda_i'^A = \sum \lambda_j'^B = 0.$$

The primes are used to distinguish the λ'-parameters from the λ-parameters in the usual parametrization

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB},$$

$$\sum \lambda_i^A = \sum \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0.$$

The λ-parameters are related to the λ'-parameters and η by the equations

$$\lambda = \lambda' - \frac{\eta}{r^2}\sum_i \sum_j |i - j| = \lambda' - \frac{\eta}{3r}(r-1)(r+1)$$

$$\lambda_i^A = \lambda_i'^A - \frac{\eta}{r}\sum_j |i-j| = \lambda_i'^A - \frac{\eta}{2r}[i(i-1) + (r-i+1)(r-i)]$$

$$\lambda_j^B = \lambda_j'^B - \frac{\eta}{r}\sum_j |i-j| = \lambda_j'^B - \frac{\eta}{2r}[j(j-1) + (r-j+1)(r-j)]$$

$$\lambda_{ij}^{AB} = -\eta\left\{|i-j| + \frac{1}{2r}[i(i-1) + j(j-1) + (r-i+1)(r-i)\right.$$

$$\left. + (r-j+1)(r-j)] + \frac{1}{3r}(r-1)(r+1)\right\}.$$

Thus as in the quasi-symmetry model, $\lambda_{ij}^{AB} = \lambda_{ji}^{AB}$.

To use the Newton–Raphson algorithm, let $\alpha_j^B = \lambda' + \lambda_j'^B$, let

$$w_{ij} = -|i - j|,$$

and for $1 \le k \le r - 1$, let

$$\begin{aligned} x_{ijk} &= 1, & i &= k, \\ &= 0, & i &\neq k, \quad i < r, \\ &= -1, & i &= r. \end{aligned}$$

Thus the models under study assume that

$$\log m_{ij} = \alpha_j^B + \sum_{k=1}^{r-1} \lambda_k'^A x_{ijk} + \eta w_{ij}.$$

In the fixed-distance model for off-diagonals, the equation is assumed for $i \neq j$. Otherwise, the equation is assumed for all i and j.

Results are summarized in Tables 8.5, 8.6, and 8.7. If diagonal cells are ignored, then $X^2 = 3.19$, $L^2 = 3.48$, and there are 4 degrees of freedom. Thus the fit is quite satisfactory. If all cells are considered, $X^2 = 38.26$, $L^2 = 35.97$, and there are 8 degrees of freedom, so that the fit is quite unsatisfactory.

Numerous variations on these two fixed-distance models are explored by Goodman (1972) and Haberman (1974a). For example, in a variable-distance model, for some η_k, $1 \leq k \leq r - 1$,

$$\begin{aligned}\log m_{ij} &= \lambda' + \lambda_i'^A + \lambda_j'^B - \sum_{k=g}^{h-1} \eta_k, \qquad g = \min(i,j) < h = \max(i,j)\\ &= \lambda' + \lambda_i'^A + \lambda_j'^B, \qquad i = j.\end{aligned}$$

As usual, this model may be applied only to off-diagonal cells or it may be applied to the entire table. Note that in the fixed-distance case, each $\eta_k = \eta$.

Table 8.5

Definitions of Variables in Fixed-Distance Models for Table 8.1

i	j	x_{ij1}	x_{ij2}	x_{ij3}	w_{ij}
1	1	1	0	0	0
2	1	0	1	0	−1
3	1	0	0	1	−2
4	1	−1	−1	−1	−3
1	2	1	0	0	−1
2	2	0	1	0	0
3	2	0	0	1	−1
4	2	−1	−1	−1	−2
1	3	1	0	0	−2
2	3	0	1	0	−1
3	3	0	0	1	0
4	3	−1	−1	−1	−1
1	4	1	0	0	−2
2	4	0	1	0	−2
3	4	0	0	1	−1
4	4	−1	−1	−1	0

Table 8.6

Estimated Expected Cell Counts and Adjusted Residuals for Fixed-Distance Models for Table 8.1[a]

Husband's highest degree	Wife's highest degree: Less than high school diploma	High school diploma or junior college degree	Bachelor's degree	Graduate degree
Less than	—	121.92	2.67	0.41
high school		0.80	−0.52	−0.67
diploma	250.96	123.86	8.64	0.54
	2.10	−0.17	−2.58	−0.76
High school	83.08	—	31.13	4.78
diploma or	−0.80		−0.60	1.50
junior college	78.96	381.73	26.63	1.67
degree	0.60	−2.62	0.89	4.48
Bachelor's	3.44	58.76	—	5.80
degree	1.10	0.12	—	−1.24
	10.92	52.77	36.08	2.26
	−2.04	1.41	−0.56	1.27
Graduate	2.48	42.33	27.20	—
degree	−0.37	−0.70	1.24	—
	7.16	34.64	23.67	14.53
	−2.15	1.68	1.56	−3.82

[a] First line is estimated expected count for fixed-distance model ignoring diagonal cells, second line is corresponding adjusted residual, third line is estimated expected count for fixed-distance model for all cells, and fourth line is adjusted residual for the latter model.

Table 8.7

Parameter Estimates for Fixed-Distance Models for Table 8.1

Parameter	Ignoring off-diagonals: Estimate	Ignoring off-diagonals: EASD	Complete table: Estimate	Complete table: EASD
$\lambda_1'^A$	−0.167	0.124	0.250	0.069
$\lambda_2'^A$	0.160	0.130	0.235	0.065
$\lambda_3'^A$	−0.897	0.130	−0.603	0.084
η	1.689	0.204	1.141	0.059

A special complication which arises when only off-diagonal cells are considered involves degrees of freedom. If each η_k and each λ, λ_i^A, and λ_j^B had a uniquely defined maximum likelihood estimate, then the degrees of freedom would be

$$r(r-1) - 1 - (r-1) - (r-1) - (r-1) = (r-1)(r-3) - 1$$
$$= r^2 - 4r + 2,$$

for there are $r(r-1)$ off-diagonal cells, 1 parameter λ, $r-1$ independent parameters λ_i^A, $1 \le i \le r-1$, $r-1$ independent parameters λ_j^B, $1 \le j \le r-1$, and $r-1$ parameters η_k, $1 \le k \le r-1$. However, the actual degrees of freedom are

$$(r-2)^2 = r^2 - 4r + 4,$$

as noted in Haberman (1974a, p. 217). If $r = 4$, then there are 4 degrees of freedom, just as in the fixed-distance model. In other words, if $r = 4$ and if off-diagonal cells are ignored, then the fixed-distance and variable-distance models coincide. If $r > 4$, then the models do differ. In use of the variable-distance model, it is helpful to resolve the lack of identifiability of parameters by the arbitrary choice $\eta_1 = \eta_2$ and $\eta_{r-2} = \eta_{r-1}$. Note that in the case of $r = 4$, $\eta_1 = \eta_2 = \eta_3$, just as for a fixed-distance model.

If all cells are considered, then the variable-distance model does differ from the fixed-distance model and the parameters are all uniquely defined. Use of this model is an exercise for the reader (Exercise 8.9).

8.3 Symmetry Models for Multi-Way Tables

Generalization of symmetry models from two-way to higher-way tables is straightforward, although the choice of models does increase rapidly. For some discussion and some helpful references, see Bishop, Fienberg, and Holland (1975, pp. 303–309). Here, attention will be restricted to some models for Table 7.29.

To begin consideration of symmetry models for this table, it is helpful to examine a saturated model and to estimate all interaction terms for this model. This task is accomplished in the usual fashion discussed in Chapter 4 of Volume 1. Results are summarized in Table 8.8. Note that the saturated model has the form

$$\log m_{ijkl} = \lambda + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y + \lambda_{ij}^{BD} + \lambda_{ik}^{BF} + \lambda_{il}^{BY}$$
$$+ \lambda_{jk}^{DF} + \lambda_{jl}^{DY} + \lambda_{kl}^{FY} + \lambda_{ijk}^{BDF} + \lambda_{ijl}^{BDY} + \lambda_{ikl}^{BFY} + \lambda_{jkl}^{DFY} + \lambda_{ijkl}^{BDFY}$$

Table 8.8

Parameter Estimates for the Saturated Model for Table 7.29

Parameter	Estimate	EASD	Standardized value
λ_1^B	−0.365	0.035	−10.39
λ_1^D	0.277	0.035	7.90
λ_1^F	−0.038	0.035	1.09
λ_{11}^{BD}	0.714	0.035	20.35
λ_{11}^{BF}	0.624	0.035	17.77
λ_{11}^{BY}	−0.080	0.050	−1.61
λ_{12}^{BY}	0.060	0.050	1.20
λ_{13}^{BY}	0.020	0.049	0.41
λ_{11}^{DF}	0.645	0.035	18.37
λ_{11}^{DY}	0.008	0.050	0.15
λ_{12}^{DY}	−0.004	0.050	−0.08
λ_{13}^{DY}	−0.004	0.049	0.07
λ_{11}^{FY}	−0.032	0.050	−0.65
λ_{12}^{FY}	−0.023	0.050	−0.46
λ_{13}^{FY}	0.056	0.049	1.14
λ_{111}^{BDF}	0.042	0.035	1.19
λ_{111}^{BDY}	0.037	0.050	0.74
λ_{112}^{BDY}	0.019	0.050	0.39
λ_{113}^{BDY}	−0.056	0.049	−1.15
λ_{111}^{BFY}	−0.037	0.050	−0.75
λ_{112}^{BFY}	0.044	0.050	0.87
λ_{113}^{BFY}	−0.006	0.049	−0.13
λ_{111}^{DFY}	−0.046	0.050	−0.92
λ_{112}^{DFY}	0.031	0.050	0.62
λ_{113}^{DFY}	0.015	0.049	0.30
λ_{1111}^{BDFY}	−0.014	0.050	−0.27
λ_{1112}^{BDFY}	0.022	0.050	0.44
λ_{1113}^{BDFY}	−0.008	0.049	−0.17

and that $\lambda_2^B = -\lambda_1^B$, $\lambda_2^D = -\lambda_1^D$; $\lambda_2^F = -\lambda_1^F$, $\lambda_{22}^{BD} = -\lambda_{12}^{BD} = -\lambda_{21}^{BD} = \lambda_{11}^{BD}$, etc. Note that estimates of λ and λ_l^Y are excluded since the number of subjects in a sample in a given year is regarded as fixed.

The only large standardized values observed in Table 8.8 correspond to λ_1^B, λ_1^D, λ_{11}^{BD}, λ_{11}^{BF}, and λ_{11}^{DF}. Thus one may consider the hierarchical model

$$\log m_{ijkl} = \lambda + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y + \lambda_{ij}^{BD} + \lambda_{jk}^{BF} + \lambda_{jk}^{DF}.$$

Using this model, $X^2 = 27.6$, $L^2 = 27.9$, and there are 15 degrees of freedom. Since the significance level for both chi-square statistics is about 0.02, the model is somewhat questionable. On the other hand, the model

$$\log m_{ijkl} = \lambda + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y + \lambda_{ij}^{BD} + \lambda_{ik}^{BF} + \lambda_{il}^{BY} + \lambda_{jk}^{DF} + \lambda_{jl}^{DY} + \lambda_{kl}^{FY}$$

including all two-factor interactions has $X^2 = 6.83$, $L^2 = 6.83$, and 9 degrees of freedom. Thus the fit is quite satisfactory. The improvement over the previous model is quite substantial. The decrease of L^2 of 21.0 corresponds to a decrease of freedom of 6. The significance level is only 0.002.

To proceed further, consider Table 8.9. The estimated interactions $\hat{\lambda}_{11}^{BD}$, $\hat{\lambda}_{11}^{BF}$, and $\hat{\lambda}_{11}^{DF}$ are rather similar in value, especially in the case of $\hat{\lambda}_{11}^{BD}$ and $\hat{\lambda}_{11}^{DF}$. Thus one may consider models in which $\lambda_{11}^{BD} = \lambda_{11}^{BF} = \lambda_{11}^{DF}$. To start, consider the model (Model 1)

$$\begin{aligned}\log m_{ijkl} &= \lambda + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y + \lambda_{ij}^{BD} + \lambda_{ik}^{BD} + \lambda_{il}^{BY} + \lambda_{jk}^{BD} + \lambda_{jl}^{DY} + \lambda_{kl}^{FY} \\ &= \alpha_l^Y + \lambda_1^B q_i + \lambda_1^D q_j + \lambda_1^F q_k + \lambda_{11}^{BD}(q_i q_j + q_i q_k + q_j q_k) \\ &\quad + \lambda_{11}^{BY} q_i q_{l1}^Y + \lambda_{12}^{BY} q_i q_{l2}^Y + \lambda_{11}^{DY} q_j q_{l1}^Y + \lambda_{12}^{DY} q_j q_{l2}^Y \\ &\quad + \lambda_{11}^{FY} q_k q_{l1}^Y + \lambda_{12}^{FY} q_k q_{l2}^Y,\end{aligned}$$

where $\alpha_l^Y = \lambda + \lambda_{l2}^Y$, $q_1 = 1$, $q_2 = -1$, $q_{11}^Y = 1$, $q_{22}^Y = 0$, $q_{31}^Y = -1$, $q_{12}^Y = 0$, $q_{22}^Y = 1$, and $q_{32}^Y = -1$. This model is implemented by the Newton–Raphson algorithm. The resulting value of X^2 and L^2 are 10.12 and 9.98, respectively. The degrees of freedom are 11, so that the fit appears to be rather satisfactory. Thus the data are consistent with the symmetry hypothesis that the interactions between questions are equal, so that $\lambda_{11}^{BD} = \lambda_{22}^{BD} = \lambda_{11}^{BF} = \lambda_{22}^{BF} = \lambda_{11}^{DF} = \lambda_{22}^{DF} = -\lambda_{12}^{BD} = -\lambda_{21}^{BD} = -\lambda_{12}^{BF} = -\lambda_{21}^{BF} = -\lambda_{12}^{DF} = -\lambda_{21}^{DF}$.

Table 8.9

Parameter Estimates for the Model of Two-Factor Interactions for Table 7.29

Parameter	Estimate	EASD	Standardized value
λ_1^B	−0.356	0.034	−10.50
λ_1^D	0.286	0.034	8.47
λ_1^F	−0.017	0.031	−0.55
λ_{11}^{BD}	0.714	0.035	20.45
λ_{11}^{BF}	0.639	0.032	19.90
λ_{11}^{BY}	−0.084	0.042	−1.98
λ_{12}^{BY}	0.092	0.042	2.18
λ_{13}^{BY}	−0.008	0.042	−0.20
λ_{11}^{DF}	0.626	0.032	19.89
λ_{11}^{DY}	0.000	0.041	0.01
λ_{12}^{DY}	−0.020	0.042	−0.48
λ_{13}^{DY}	0.020	0.041	0.47
λ_{11}^{FY}	−0.033	0.041	−0.80
λ_{12}^{FY}	−0.020	0.042	−0.49
λ_{13}^{FY}	0.053	0.041	1.29

Given the pattern of λ-parameters involving year of survey which is encountered in Table 8.9, one may consider the still simpler model (Model 2)

$$\log m_{ijkl} = \lambda + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_{ij}^{BD} + \lambda_{ik}^{BD} + \lambda_{il}^{BY} + \lambda_{jk}^{BD}$$
$$= \alpha_l^Y + \lambda_1^B q_i + \lambda_1^D q_j + \lambda_1^F q_k + \lambda_{11}^{BD}(q_i q_j + q_i q_k + q_j q_k)$$
$$+ \lambda_{11}^{BY} q_i q_{l1}^Y + \lambda_{12}^{BY} q_i q_{l2}^Y.$$

One finds that $X^2 = 13.1$ and $L^2 = 12.9$, and there are 15 degrees of freedom. Thus this last model also provides a satisfactory fit. For detailed results, see Table 8.10 and Table 8.11. At this point, attractive simplifications are not available.

The last model suggests that there is some change over time in the distribution of responses to question B, which concerns a married woman who does

Table 8.10

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for Model 2 for Table 7.29

Year of survey	Response to B	Response to D	Response to F	Observed count	Estimated expected count	Adjusted residual
1972	Yes	Yes	Yes	334	340.36	−1.08
	Yes	Yes	No	34	25.44	1.88
	Yes	No	Yes	12	14.34	−0.67
	Yes	No	No	15	14.85	0.04
	No	Yes	Yes	53	60.00	−1.06
	No	Yes	No	63	62.14	0.13
	No	No	Yes	43	35.04	1.52
	No	No	No	501	502.82	−0.20
1973	Yes	Yes	Yes	428	419.64	1.33
	Yes	Yes	No	29	31.36	−0.48
	Yes	No	Yes	13	17.68	−1.22
	Yes	No	No	17	18.31	−0.34
	No	Yes	Yes	42	52.64	−1.70
	No	Yes	No	53	54.52	−0.24
	No	No	Yes	31	30.74	0.05
	No	No	No	453	441.11	1.39
1974	Yes	Yes	Yes	413	410.16	0.46
	Yes	Yes	No	29	30.65	−0.34
	Yes	No	Yes	16	17.28	−0.34
	Yes	No	No	18	17.90	0.03
	No	Yes	Yes	60	53.10	1.10
	No	Yes	No	57	54.98	0.32
	No	No	Yes	37	31.00	1.21
	No	No	No	430	444.92	−1.74

Table 8.11

Parameter Estimates for Model 2 for Table 7.29

Parameter	Estimate	EASD
λ_1^B	−0.338	0.031
λ_1^D	0.269	0.031
λ_1^F	−0.017	0.030
λ_{11}^{BD}	0.657	0.012
λ_{11}^{BY}	−0.108	0.026
λ_{12}^{BY}	0.062	0.025
λ_{13}^{BU}	0.046	0.025

not want more children. This effect is evident from the marginal totals for this question. In 1972, 395 of 1055 subjects, or 37.4 percent answered "yes." In 1973, the fraction increased to 487 of 1066, or 45.7 percent, while in 1974 the fraction was 476 of 1060 or 44.9 percent. Since the model implies that given the response to B, year of survey is independent of responses to the other two questions (no λ-parameters include Y and D or F), the estimated log cross-product ratio $\tau^{BY\cdot DF}_{(12)(ll')\cdot jk}$ for interaction of year l and l' of survey and response to "yes" or "no" to B given responses j to D and k to F satisfies the relationships

$$\begin{aligned}\hat{\lambda}^{BY\cdot DF}_{(12)(ll')\cdot jk} &= \hat{\lambda}^{BY}_{(12)(ll')}\\ &= \log[(n^{BY}_{1l}n^{BY}_{2l'})/(n^{BY}_{2l}n^{BY}_{1l'})]\\ &= \hat{\lambda}^{BY}_{1l} - \hat{\lambda}^{BY}_{2l} - \hat{\lambda}^{BY}_{1l'} + \hat{\lambda}^{BY}_{2l'}\\ &= 2(\hat{\lambda}^{BY}_{1l} - \hat{\lambda}^{BY}_{1l'}).\end{aligned}$$

For example, consider subjects who answer "yes" to questions D and F. Note that for years 1972 and 1974,

$$\hat{\tau}^{BY\cdot DF}_{(12)(13)\cdot 11} = \log\left(\frac{m_{1111}m_{2113}}{m_{2111}m_{1113}}\right) = \log\left(\frac{340.36 \times 53.10}{60.00 \times 410.16}\right) = -0.309,$$

so that for such subjects, the odds of a positive response to B are $0.734 = e^{-0.309}$ times as great in 1972 as they are in 1974. Note that

$$\log[(n^{BY}_{11}n^{BY}_{23})/(n^{BY}_{21}n^{BY}_{13})] = \log[(395 \times 584)/(660 \times 476)] = -0.309$$

and

$$2(\hat{\lambda}^{BY}_{11} - \hat{\lambda}^{BY}_{13}) = 2(-0.108 - 0.046) = -0.309.$$

(Rounding of numbers for purposes of presentation accounts for the slight discrepancy in the last equation.)

According to the model under study, the responses to different questions are strongly related, and the relationship is symmetrical. This relationship does not depend on the year of survey. Note that the log odds ratio $\tau_{12\cdot ijl}^{B\cdot DFY}$ of a response "yes" rather than "no" to B given response i to D, response j to F, and year of survey l is

$$2(\lambda_1^B + \lambda_{1i}^{BD} + \lambda_{1j}^{BD}),$$

which is independent of year of survey. Similarly,

$$\tau_{12\cdot ijl}^{D\cdot BFY} = 2(\lambda_1^D + \lambda_{1i}^{BD} + \lambda_{1j}^{BD})$$

and

$$\tau_{12\cdot ijl}^{F\cdot BDY} = 2(\lambda_1^F + \lambda_{1i}^{BD} + \lambda_{1j}^{BD}).$$

The estimates $\hat{\tau}_{12\cdot ijl}^{B\cdot DFY}$, $\hat{\tau}_{12\cdot ijl}^{D\cdot BFY}$, and $\hat{\tau}_{12\cdot ijl}^{F\cdot BDY}$ are strongly dependent on i and j. For the case, $i = j = 1$ which corresponds to two positive responses,

$$\hat{\tau}_{12\cdot 11l}^{B\cdot DFY} = 2(\hat{\lambda}_1^B + 2\hat{\lambda}_{11}^{BD}) = 1.95,$$

$$\hat{\tau}_{12\cdot 11l}^{D\cdot BFY} = 2(\hat{\lambda}_1^D + 2\hat{\lambda}_{11}^{BD}) = 3.17,$$

$$\hat{\tau}_{12\cdot 11l}^{F\cdot BDY} = 2(\hat{\lambda}_1^F + 2\hat{\lambda}_{11}^{BD}) = 2.59.$$

With mixed responses ($i \neq j$) on other questions, one has

$$\hat{\tau}_{12\cdot 12l}^{B\cdot DFY} = \hat{\tau}_{12\cdot 21l}^{B\cdot DFY} = 2\hat{\lambda}_1^B = -0.677,$$

$$\hat{\tau}_{12\cdot 12l}^{D\cdot BFY} = \hat{\tau}_{12\cdot 21l}^{D\cdot BFY} = 2\hat{\lambda}_1^D = 0.538,$$

$$\hat{\tau}_{12\cdot 12l}^{F\cdot BDY} = \hat{\tau}_{21\cdot 12l}^{F\cdot BDY} = 2\hat{\lambda}_1^F = -0.035.$$

With negative responses ($i = j = 2$),

$$\hat{\tau}_{12\cdot 22l}^{B\cdot DFY} = 2(\hat{\lambda}_1^B - 2\hat{\lambda}_{11}^{BD}) = -3.31,$$

$$\hat{\tau}_{12\cdot 22l}^{D\cdot BFY} = 2(\hat{\lambda}_1^D - 2\hat{\lambda}_{11}^{BD}) = -2.09,$$

$$\hat{\tau}_{12\cdot 22l}^{F\cdot BDY} = 2(\hat{\lambda}_1^F - 2\hat{\lambda}_{11}^{BD}) = -2.66.$$

The odds of a positive response to B are estimated to be

$$\exp(8\hat{\lambda}_{11}^{BD}) = 192$$

times as large if the other responses are positive than if the other responses are negative. This same ratio applies to questions D and F.

The size of the relationship is fairly well determined. Note that a 95 percent confidence interval for λ_{11}^{BD} has bounds

$$0.657 - 1.96(0.012) = 0.633$$

and

$$0.657 + 1.96(0.012) = 0.681.$$

Corresponding bounds for $\exp(8\lambda_{11}^{BD})$ are

$$e^{8(0.633)} = 158 \qquad \text{and} \qquad e^{8(0.681)} = 233.$$

Thus the model of symmetrical interaction has resulted in a simple description of the relationship of the parameters and in a relatively accurate assessment of the magnitude of the relationship. This model is no more parsimonious than Model 6 of Exercise 7.29; however, its symmetry and greater ease of interpretation are attractive.

EXERCISES

8.1 Show that in the symmetry model of Section 8.1,

$$X^2 = \sum_{i=2}^{r} \sum_{j=1}^{i-1} (n_{ij} - n_{ji})^2/(n_{ij} + n_{ji})$$

and

$$L^2 = 2 \sum_{i=2}^{r} \sum_{j=1}^{i-1} \{n_{ij} \log n_{ij} + n_{ji} \log n_{ji} - (n_{ij} + n_{ji}) \log[\tfrac{1}{2}(n_{ij} + n_{ji})]\}.$$

Solution

Since $\hat{m}_{ij} = \frac{1}{2}(n_{ij} + n_{ji})$,

$$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{r} [n_{ij} - \tfrac{1}{2}(n_{ij} + n_{ji})]^2/[\tfrac{1}{2}(n_{ij} + n_{ji})]$$

$$= \sum_{i=1}^{r} \sum_{j=1}^{r} (n_{ij} - n_{ji})^2/(n_{ij} + n_{ji}).$$

Note that

$$(n_{ij} - n_{ji})^2/(n_{ij} + n_{ji}) = (n_{ji} - n_{ij})^2/(n_{ji} + n_{ij})$$

and

$$(n_{ii} - n_{ii})^2/(n_{ii} + n_{ii}) = 0.$$

Thus X^2 has the desired form.

In the case of L^2,

$$L^2 = 2 \sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij} \log[2n_{ij}/(n_{ij} + n_{ji})]$$

$$= 2 \sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij} \log n_{ij} - 2 \sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij} \log[\tfrac{1}{2}(n_{ij} + n_{ji})].$$

Since

$$n_{ii} \log[\tfrac{1}{2}(n_{ii} + n_{ii})] = n_{ii} \log n_{ii},$$

it follows that L^2 has the desired form.

8.2 Show that in the symmetry model of Section 8.1,

$$r_{ij} = (n_{ij} - n_{ji})/(n_{ij} + n_{ji})^{1/2}, \qquad i \neq j.$$

Solution

Note that the estimated variance of $n_{ij} - \hat{m}_{ij} = \frac{1}{2}(n_{ij} - n_{ji})$ is

$$\hat{c}_{ij} = \hat{m}_{ij}[1 - \hat{m}_{ij}/(\hat{m}_{ij} + \hat{m}_{ji})] = \tfrac{1}{4}(n_{ij} + n_{ji}), \qquad i \neq j.$$

Thus

$$r_{ij} = \tfrac{1}{2}(n_{ij} - n_{ji})/[\tfrac{1}{4}(n_{ij} + n_{ji})]^{1/2} = (n_{ij} - n_{ji})/(n_{ij} + n_{ji})^{1/2}.$$

8.3 Verify that under the quasi-symmetry model of Section 8.1,

$$\hat{m}_i^A = n_i^A, \qquad \hat{m}_j^B = n_j^B,$$

and

$$\hat{m}_{ij} + \hat{m}_{ji} = n_{ij} + n_{ji}.$$

Solution

For $1 \leq i' \leq r - 1$, let

$$\begin{aligned} q_{ii'} &= 1, && i = i', \\ &= 0, && i \neq i', \quad i < r, \\ &= -1, && i = r. \end{aligned}$$

One now proceeds much as in Section 2.6. Note that it is assumed that

$$\begin{aligned} \log m_{ij} &= \lambda + \sum_{i'=1}^{r-1} \lambda_{i'}^A q_{ii'} + \sum_{j'=1}^{r-1} \lambda_{j'}^B q_{jj'} + \sum_{i'=1}^{r-1} \sum_{j'=1}^{r-1} \lambda_{i'j'}^{AB} q_{ii'} q_{jj'} \\ &= \lambda + \sum_{i'=1}^{r-1} \lambda_{i'}^A q_{ii'} + \sum_{j'=1}^{r-1} \lambda_{j'}^B q_{jj'} + \sum_{i'=1}^{r-1} \sum_{j'=1}^{i'} \lambda_{i'j'}^{AB} (q_{ii'} q_{jj'} + q_{ij'} q_{ji'}). \end{aligned}$$

Thus

$$\sum_i \sum_j \hat{m}_{ij} = N = \sum_i \sum_j n_{ij},$$

$$\sum_i \sum_j q_{ii'} \hat{m}_{ij} = \sum_i \sum_j q_{ii'} n_{ij}, \qquad 1 \leq i' \leq r - 1,$$

$$\sum_i \sum_j q_{jj'} \hat{m}_{ij} = \sum_i \sum_j q_{jj'} n_{ij}, \qquad 1 \leq j' \leq r - 1,$$

$$\sum_i \sum_j (q_{ii'} q_{jj'} + q_{ij'} q_{ji'}) \hat{m}_{ij} = \sum_i \sum_j (q_{ii'} q_{jj'} + q_{ij'} q_{ji'}) n_{ij}, \quad 1 \leq j' \leq i' \leq r - 1.$$

As in Section 2.6, the first three equations are equivalent to the equations

$$\hat{m}_i^A = n_i^A, \qquad 1 \le i \le r,$$
$$\hat{m}_j^B = n_j^B, \qquad 1 \le j \le r.$$

The last equation reduces to

$$\hat{m}_{i'j'} - \hat{m}_{rj'} - \hat{m}_{i'r} + \hat{m}_{rr} + \hat{m}_{j'i'} - \hat{m}_{j'r} - \hat{m}_{rj'} + \hat{m}_{rr}$$
$$= n_{i'j'} - n_{rj'} - n_{i'r} + n_{rr} + n_{j'i'} - n_{j'r} - n_{rj'} + n_{rr},$$
$$1 \le i' \le r - 1, \quad 1 \le j' \le r - 1.$$

This new equation is implied by the equation

$$\hat{m}_{ij} + \hat{m}_{ji} = n_{ij} + n_{ji}.$$

On the other hand, summation over i' and j' shows that

$$\hat{m}_{rr} = n_{rr}.$$

Summation over j' for a fixed i' shows that

$$\hat{m}_{i'r} + \hat{m}_{ri'} = n_{i'r} + n_{ri'}, \qquad 1 \le i' \le r - 1.$$

It then follows that

$$\hat{m}_{i'j'} + \hat{m}_{j'i'} = n_{i'j'} + n_{j'i'}, \qquad 1 \le i' \le r - 1, \quad 1 \le j' \le r - 1.$$

Thus the general equation

$$\hat{m}_{ij} + \hat{m}_{ji} = n_{ij} + n_{ji}$$

must hold.

8.4 Derive Tables 8.3 and 8.4.

Solution

The programm in the Appendix can be used. Note that

$$\log m_{ij} = \alpha + \sum_{k=1}^{12} \beta_k x_{ijk},$$

where $\alpha = \lambda$ and the β_k and x_{ijk} are defined as in Table 8.12.

8.5 Derive Table 8.3 with the Newton–Raphson algorithm, using the parametrization

$$\log m_{ij} = \alpha_{ij} + \tfrac{1}{2}(\lambda_i^A - \lambda_i^B) - \tfrac{1}{2}(\lambda_j^A - \lambda_j^B), \qquad i < j,$$
$$= \alpha_{ji} + \tfrac{1}{2}(\lambda_i^A - \lambda_i^B) - \tfrac{1}{2}(\lambda_j^A - \lambda_j^B), \qquad i > j,$$

where

$$\alpha_{ij} = \lambda + \tfrac{1}{2}(\lambda_i^A + \lambda_j^A) + \tfrac{1}{2}(\lambda_i^B + \lambda_j^B) + \lambda_{ij}^{AB}, \qquad i < j.$$

Table 8.12

Coefficients for the Quasi-Symmetry Model for Table 8.1

k	1	2	3	4	5	6	7	8	9	10	11	12
β_k	λ_1^A	λ_2^A	λ_3^A	λ_1^B	λ_2^B	λ_3^B	λ_{11}^{AB}	λ_{12}^{AB}	λ_{22}^{AB}	λ_{13}^{AB}	λ_{23}^{AB}	λ_{33}^{AB}
x_{11k}	1	0	0	1	0	0	2	0	0	0	0	0
x_{21k}	0	1	0	1	0	0	0	1	0	0	0	0
x_{31k}	0	0	1	1	0	0	0	0	0	1	0	0
x_{41k}	−1	−1	−1	1	0	0	−2	−1	0	−1	0	0
x_{12k}	1	0	0	0	1	0	0	1	2	0	0	0
x_{22k}	0	1	0	0	1	0	0	0	0	0	0	0
x_{32k}	0	0	1	0	1	0	0	0	−2	0	1	0
x_{42k}	−1	−1	−1	0	1	0	0	−1	0	0	−1	0
x_{13k}	1	0	0	0	0	1	0	0	0	1	0	0
x_{23k}	0	1	0	0	0	1	0	0	0	0	1	0
x_{33k}	0	0	1	0	0	1	0	0	0	0	0	2
x_{43k}	−1	−1	−1	0	0	1	0	0	0	−1	−1	−2
x_{14k}	1	0	0	−1	−1	−1	−2	−1	0	−1	0	0
x_{24k}	0	1	0	−1	−1	−1	0	−1	−2	0	−1	0
x_{34k}	0	0	1	−1	−1	−1	0	0	0	−1	−1	−2
x_{44k}	−1	−1	−1	−1	−1	−1	2	2	2	2	2	2

Solution

Convert Table 8.1 into a two-way table of counts $n'_{i'j'}$, $1 \le i' \le 2$, $1 \le j' \le 6$, as shown in Table 8.13. Let $m'_{i'j'}$ be the expected value of $n'_{i'j'}$, and note that the proposed parametrization is equivalent to the model

$$\log m'_{i'j'} = \alpha'_{j'} + \sum_{k=1}^{3} \beta_k x_{i'j'k},$$

where $\alpha'_{j'} = \alpha_{ij}$ for j' corresponding to i and j,

$$\beta_k = (\lambda_k^A - \lambda_k^B), \qquad 1 \le k \le 3,$$

and the $x_{i'j'k}$ are defined as in Table 8.13. Given this parametrization, the program in the Appendix or other related programs are readily used. The coefficient estimates $\hat{\beta}_k$ and the corresponding EASD's $s(\hat{\beta}_k)$ listed in Table 8.14 also provide a measure of the degree of asymmetry found in the table. Note the very large standardized value $(\hat{\lambda}_2^A - \hat{\lambda}_2^B)/s(\hat{\lambda}_2^A - \hat{\lambda}_2^B)$.

8.6 Verify that the quasi-symmetry model of Section 8.1 holds if and only if

$$m_{ij}m_{jk}m_{ki}/(m_{ji}m_{kj}m_{ik}) = 1.$$

Table 8.13

Reparametrization of Table 8.1

i'	j'	i	j	$n'_{i'j'} = n_{ij}$	$x_{i'j'1}$	$x_{i'j'2}$	$x_{i'j'3}$
1	1	1	2	123	1	−1	0
2	1	2	1	82	−1	1	0
1	2	1	3	2	1	0	−1
2	2	3	1	5	−1	0	1
1	3	1	4	0	2	1	1
2	3	4	1	2	−2	−1	−1
1	4	2	3	30	0	1	−1
2	4	3	2	59	0	−1	1
1	5	2	4	7	1	2	1
2	5	4	2	41	−1	−2	−1
1	6	3	4	4	1	1	2
2	6	4	3	29	−1	−1	−2

Table 8.14

Estimates of $\frac{1}{2}(\lambda_i^A - \lambda_i^B)$ in Table 8.1

Parameter	Estimate	EASD	Standardized value
$\frac{1}{2}(\lambda_1^A - \lambda_1^B)$	−0.210	0.075	−2.80
$\frac{1}{2}(\lambda_2^A - \lambda_2^B)$	−0.399	0.059	−6.72
$\frac{1}{2}(\lambda_3^A - \lambda_3^B)$	−0.080	0.077	−1.05

Solution

The observation here is verified in Caussinus (1965). See also Bishop, Fienberg, and Holland (1975, p. 290). The necessity of the condition follows since

$$q^{AB}_{(ij)(ki)} = q^{AB}_{(ki)(ij)}$$

under quasi-symmetry. Thus

$$\left(\frac{m_{ki}m_{ij}}{m_{ii}m_{kj}}\right)\bigg/\left(\frac{m_{ik}m_{ji}}{m_{jk}m_{ii}}\right) = \frac{m_{ij}m_{jk}m_{ki}}{m_{ji}m_{kj}m_{ik}} = 1.$$

On the other hand,

$$q^{AB}_{(ij)(kl)} = q^{AB}_{(ij)(ki)}/q^{AB}_{(ij)(li)}$$

and

$$q^{AB}_{(kl)(ij)} = q^{AB}_{(ki)(ij)}/q^{AB}_{(li)(ij)}.$$

If

$$m_{ij}m_{jk}m_{ki}/(m_{ji}m_{kj}m_{ik}) = 1$$

for all i, j, and k, then

$$q^{AB}_{(ij)(ki)} = q^{AB}_{(ki)(ij)}.$$

Thus

$$q^{AB}_{(ij)(kl)} = q^{AB}_{(kl)(ij)}$$

and quasi-symmetry holds.

8.7 Determine whether a quasi-symmetry model provides a satisfactory fit for Table 7.1.

Solution

Results are summarized in Table 8.15 and Table 8.16. Since $X^2 = 3.86$, $L^2 = 3.84$, and there are 3 degrees of freedom, the fit is quite satisfactory.

8.8 Verify that the Stuart $Q = 60.6$ in Table 8.1.

Table 8.15

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for Table 7.1 under the Quasi-Symmetry Model[a]

Residence at age 16	Current residence			
	Northeast	South	North Central	West
Northeast	263	22	14	13
	263	25.00	11.86	12.13
	—	−1.52	1.13	0.76
South	26	399	36	30
	23.00	399	36.46	32.55
	1.52		−0.21	−1.58
North Central	10	41	368	46
	12.14	40.54	368	44.32
	−1.13	0.21	··	1.02
West	1	8	5	148
	1.87	5.45	6.68	148
	−0.76	1.58	−1.02	

[a] The first entry is the observed frequency, the second entry is the maximum likelihood estimate of the expected frequency, and the third entry is the adjusted residual.

Table 8.16

Definitions of Variables in Variable-Distance Model for Table 8.1

i	j	x_{ij1}	x_{ij2}	x_{ij3}	w_{ij1}	w_{ij2}	w_{ij3}
1	1	1	0	0	0	0	0
2	1	0	1	0	−1	0	0
3	1	0	0	1	−1	−1	0
4	1	−1	−1	−1	−1	−1	−1
1	2	1	0	0	−1	0	0
2	2	0	1	0	0	0	0
3	2	0	0	1	0	−1	0
4	2	−1	−1	−1	0	−1	−1
1	3	1	0	0	−1	−1	0
2	3	0	1	0	0	−1	0
3	3	0	0	1	0	0	0
4	3	−1	−1	−1	0	0	−1
1	4	1	0	0	−1	−1	−1
2	4	0	1	0	0	−1	−1
3	4	0	0	1	0	0	−1
4	4	−1	−1	−1	0	0	0

Solution

Note that $f_1 = 36$, $f_2 = -104$, $f_3 = 7$, $v_{11} = 214$, $v_{21} = v_{12} = -205$, $v_{22} = 342$, $v_{31} = v_{13} = -7$, $v_{32} = v_{23} = -89$, $v_{33} = 129$. Thus $v^{11} = 0.0176$, $v^{21} = v^{12} = 0.0132$, $v^{32} = 0.0134$, $v^{31} = v^{13} = 0.0101$, $v^{32} = v^{23} = 0.0100$, $v^{33} = 0.0152$, and $Q = 60.60$.

8.9 Using the variable-distance model for a complete table with Table 8.1, obtain parameter estimates and estimated asymptotic standard deviations for λ_i^A, $1 \le i \le 3$, and η_k, $1 \le k \le 3$, find estimated expected counts and adjusted residuals, and compute X^2, L^2 and the corresponding degrees of freedom. Does the model fit the data?

Solution

One may write

$$\log m_{ij} = \alpha_j + \sum_{k=1}^{3} \lambda_k'^A x_{ijk} + \sum_{k=1}^{3} \eta_k w_{ijk},$$

where $\alpha_j^B = \lambda' + \lambda_j'^B$, x_{ijk} is defined as in Section 8.2, and

$$\begin{aligned} w_{ijk} &= -1, \qquad i \le k < j \quad \text{or} \quad j \le k < i, \\ &= 0, \qquad \text{otherwise.} \end{aligned}$$

Table 8.17

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for Variable-Distance Model for Table 8.1

Husband's highest degree	Wife's highest degree: Less than high school diploma	High school diploma or junior college degree	Bachelor's degree	Graduate degree
Less than	259	123	2	0
high school	259.00	116.00	7.76	1.24
diploma	—	2.76	−2.40	−1.20
High school	82	370	30	7
diploma	72.18	386.82	25.86	4.14
junior college	2.90	−4.02	1.56	1.86
degree				
Bachelor's	5	59	34	4
degree	9.63	51.62	35.13	5.62
	−1.09	1.96	−0.34	−1.04
Graduate	2	41	29	8
degree	7.20	38.56	26.25	8.00
	−2.15	0.69	0.88	—

Thus x_{ijk} and w_{ijk} are defined as in Table 8.16. Using the Newton–Raphson algorithm, one obtains results shown in Table 8.17 and Table 8.18. Since $X^2 = 18.6$, $L^2 = 23.2$, and there are 6 degrees of freedom, the fit is not unsatisfactory.

8.10 Derive Table 8.10 and Table 8.11.

Solution

The Newton–Raphson algorithm is easily used. Note that

$$\log m_{ijkl} = \alpha_l^Y + \sum_{h=1}^{6} \beta_h x_{ijklh},$$

where the definitions in Table 8.19 are used and where $\beta_1 = \lambda_1^B$, $\beta_2 = \lambda_1^D$, $\beta_3 = \lambda_1^F$, $\beta_4 = \lambda_{11}^{BD}$, $\beta_5 = \lambda_{11}^{BY}$, and $\beta_6 = \lambda_{12}^{BY}$. The coefficient $\hat{\lambda}_{13}^{BY} = -\hat{\lambda}_{11}^{BY} - \hat{\lambda}_{12}^{BY}$. The value of $s(\hat{\lambda}_{13}^{BY})$ may be obtained by noting that

$$\hat{\lambda}_{13}^{BY} = \frac{1}{2}\log(n_{13}^{BY}/n_{23}^{BY}) - \frac{1}{6}\sum_{l=1}^{3}\log(n_{1l}^{BY}/n_{2l}^{BY})$$

$$= \frac{1}{3}\log(n_{13}^{BY}/n_{23}^{BY}) - \frac{1}{6}\log(n_{11}^{BY}/n_{21}^{BY}) - \frac{1}{6}\log(n_{12}^{BY}/n_{22}^{BY}).$$

Table 8.18

Parameter Estimates for Variable-Distance Model for Table 8.1

Parameter	Estimate	EASD
λ_1^A	0.447	0.091
λ_2^A	0.410	0.081
λ_3^A	−0.444	0.105
η_1	1.241	0.079
η_2	1.160	0.114
η_3	0.322	0.246

Table 8.19

Coefficients for Model 2 for Table 7.29

i	j	k	l	x_{ijkl1}	x_{ijkl2}	x_{ijkl3}	x_{ijkl4}	x_{ijkl5}	x_{ijkl6}
1	1	1	1	1	1	1	3	1	0
1	1	2	1	1	1	−1	−1	1	0
1	2	1	1	1	−1	1	−1	1	0
1	2	2	1	1	−1	−1	−1	1	0
2	1	1	1	−1	1	1	−1	−1	0
2	1	2	1	−1	1	−1	−1	−1	0
2	2	1	1	−1	−1	1	−1	−1	0
2	2	2	1	−1	−1	−1	3	−1	0
1	1	1	2	1	1	1	3	0	1
1	1	2	2	1	1	−1	−1	0	1
1	2	1	2	1	−1	1	−1	0	1
1	2	2	2	1	−1	−1	−1	0	1
2	1	1	2	−1	1	1	−1	0	−1
2	1	2	2	−1	1	−1	−1	0	−1
2	2	1	2	−1	−1	1	−1	0	−1
2	2	2	2	−1	−1	−1	3	0	−1
1	1	1	3	1	1	1	3	−1	−1
1	1	2	3	1	1	−1	−1	−1	−1
1	2	1	3	1	−1	1	−1	−1	−1
1	2	2	3	1	−1	−1	−1	−1	−1
2	1	1	3	−1	1	1	−1	1	1
2	1	2	3	−1	1	−1	−1	1	1
2	2	1	3	−1	−1	1	−1	1	1
2	2	2	3	−1	−1	−1	3	1	1

Thus

$$s(\hat{\lambda}_{13}^{BY}) = \left[\frac{1}{9}\left(\frac{1}{n_{13}^{BY}} + \frac{1}{n_{23}^{BY}}\right) + \frac{1}{36}\left(\frac{1}{n_{11}^{BY}} + \frac{1}{n_{21}^{BY}} + \frac{1}{n_{12}^{BY}} + \frac{1}{n_{22}^{BY}}\right)\right]^{1/2}$$

$$= \left[\frac{1}{9}\left(\frac{1}{476} + \frac{1}{584}\right) + \frac{1}{36}\left(\frac{1}{395} + \frac{1}{660} + \frac{1}{487} + \frac{1}{579}\right)\right]^{1/2}$$

$$= 0.025.$$

8.11 Consider an $r \times r \times r$ contingency table with counts n_{ijk}, $1 \le i \le r$, $1 \le j \le r$, $1 \le k \le r$. Find maximum likelihood estimates $\hat{m}_{ijk}$ for the symmetry model $m_{ijk} = m_{ikj} = m_{kij}$. Find degrees of freedom for chi-square tests. Note that by Feller (1968, p. 38), there are $(r+2)(r+1)r/6$ integers i, j, and k such that $1 \le i \le j \le k \le r$.

Solution

As noted by Bishop, Fienberg, and Holland (1975, p. 302),

$$\hat{m}_{ijk} = \tfrac{1}{6}(n_{ijk} + n_{jik} + n_{ikj} + n_{kji} + n_{kij} + n_{jki}).$$

Note that the corresponding log-linear model is

$$\log m_{ijk} = \alpha_{i'j'k'},$$

where i' is the minimum of i, j, and k, k' is the maximum of i, j, and k, and j' is the second smallest (or second largest) of i, j, and k. Thus

$$\log m_{123} = \log m_{213} = \log m_{132} = \log m_{321} = \log m_{312} = \log m_{231} = \alpha_{123}$$

and

$$\log m_{112} = \log m_{121} = \log m_{211} = \alpha_{112}.$$

It follows that

$$\hat{m}_{ijk} = \hat{m}_{jik} = \hat{m}_{ikj} = \hat{m}_{kji} = \hat{m}_{kij} = \hat{m}_{jki}$$

and

$$\hat{m}_{ijk} + \hat{m}_{jik} + \hat{m}_{ikj} + \hat{m}_{kji} + \hat{m}_{kij} + \hat{m}_{jki}$$
$$= n_{ijk} + n_{jik} + n_{ikj} + n_{kji} + n_{kij} + n_{jki}.$$

Thus $\hat{m}_{ijk}$ has the desired form.

To find degrees of freedom, note that there are $(r+2)r(r-1)/6$ parameters α_{ijk}, $1 \le i \le j \le k \le r$ and r^3 cells. Thus there are $r^3 - (r+1)r(r-1)/6$ degrees of freedom. Note that the formula in Bishop, Fienberg, and Holland (1975, p. 302) appears to be in error.

9 Adjustment of Data

Survey data can be used to estimate the joint distribution of several polytomous variables in the population under study. If the survey data are supplemented by some information from population censuses concerning the distribution of these variables, then the joint distribution can be estimated with increased precision. Section 9.1 examines the approach to this problem developed by Deming and Stephan (1940) and discussed by Deming (1964 [1943]) and Ireland and Kullback (1968). Section 9.2 considers extensions of the method of Deming and Stephan (1940) discussed in Haberman (1974a, pp. 376–386). Section 9.3 provides a brief survey of related applications that appear in the literature.

Methods developed in this chapter will be illustrated through use of data on educations of husbands and wives that come from two sources, the 1972 General Social Survey and the 1970 United States Census. The object of methods in this chapter is simultaneous use of information from both sources of data to obtain a better description of the joint distribution of educations of husbands and wives than can be obtained from either source separately.

9.1 Adjustment of Marginal Tables

The method of adjustment developed by Deming and Stephan (1940) applies if a table providing the joint distribution of two or more polytomous variables has been obtained from a population sample and some tables of marginal distributions of these variables have been obtained from a population census. The joint population distribution of the variables is obtained through an iterative proportional fitting algorithm. The Deming–Stephan procedure produces consistent estimates of joint population probabilities if the sample is a simple random sample. In this case, relatively simple formulas are also available for estimation of asymptotic standard deviations of the probability estimates. The Deming–Stephan algorithm also yields consistent probability estimates if the sample is not random, provided the interaction structure associated with the sample distribution is the same as the interaction structure associated with the population distribution. Unfortunately, formulas for asymptotic variances are much more complicated in this case.

The basic operation of the Deming–Stephan method may be seen by an examination of the two-variable case. Two sources of data are present in this case. In one case, N independently distributed pairs (A_h, B_h), $1 \leq h \leq N$, are observed, where each A_h can assume integral values i from 1 to r and each B_h can assume integral values j from 1 to s. The frequency count n_{ij} is the number of pairs in which $A_h = i$ and $B_h = j$, and m_{ij} is the expected value of n_{ij}. As in Chapter 2, the $r \times s$ table of counts n_{ij} may be distributed as in a Poisson sampling model, a simple multinomial sampling model, a row-multinomial sampling model, or a column-multinomial sampling model.

In the second source of data, N' other pairs (A'_h, B'_h), $1 \leq h \leq N'$, are present, each A'_h can assume integral values i from 1 to r, and each B'_h can assume integral values j from 1 to s. However, the pairs (A'_h, B'_h) are not directly observed, and the number n'_{ij} of pairs such that $A'_h = i$ and $B'_h = j$ is not known. Instead, marginal totals are available. Thus the number $n_i'^A$ of variables A'_h equal to i may be known, the number $n_j'^B$ of variables B'_h equal to j may be known, or both $n_i'^A$, $1 \leq i \leq r$, and $n_j'^B$, $1 \leq j \leq s$, may be known. The table of n'_{ij} may represent a complete population or it may be derived from Poisson sampling, simple multinomial sampling, row-multinomial sampling, or column-multinomial sampling. The expected value of n'_{ij} is m'_{ij}. (Here $m'_{ij} = n'_{ij}$ if a complete population is observed.) The available data are used to estimate the m'_{ij} or to estimate functions of the m'_{ij} such as the expected proportion $p'_{ij} = m'_{ij}/N'$ of observations with $A'_h = i$ and $B'_h = j$.

Information from the two separate sources of data can be combined if the interaction structures of the table of n_{ij} and the table of n'_{ij} are sufficiently similar. As in Chapter 2, the m_{ij} and m'_{ij} may be parametrized so that

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, \tag{9.1}$$

$$\sum \lambda_i^A = \sum \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0, \tag{9.2}$$

$$\log m'_{ij} = \lambda' + \lambda_i'^A + \lambda_j'^B + \lambda_{ij}'^{AB}, \tag{9.3}$$

and

$$\sum \lambda_i'^A = \sum \lambda_j'^B = \sum_i \lambda_{ij}'^{AB} = \sum_j \lambda_{ij}'^{AB} = 0. \tag{9.4}$$

Deming and Stephan's (1940) approach applies under the following conditions:

(A) The row totals $n_i'^A$, $1 \le i \le r$, are known, $\lambda_j^B = \lambda_j'^B$, $1 \le j \le s$, and $\lambda_{ij}^{AB} = \lambda_{ij}'^{AB}$, $1 \le i \le r$, $1 \le j \le s$.

(B) The column totals $n_j'^B$, $1 \le j \le s$, are known, $\lambda_i^A = \lambda_i'^A$, $1 \le i \le r$, and $\lambda_{ij}^{AB} = \lambda_{ij}'^{AB}$, $1 \le i \le r$, $1 \le j \le s$.

(C) The row totals $n_i'^A$, $1 \le i \le r$, and the column totals $n_j'^B$, $1 \le j \le s$, are both known and $\lambda_{ij}^{AB} = \lambda_{ij}'^{AB}$, $1 \le i \le r$, $1 \le j \le s$.

In other words, λ^S and λ'^S parameters must be equal if S-marginal totals for the n_{ij}' are unknown.

To illustrate use of the Deming–Stephan algorithm in Cases (A), (B), and (C), consider Tables 9.1 and 9.2. Both tables have $r = 4$ rows and $s = 4$ columns, and both tables measure the joint distribution of years of education of married couples in the United States. In Table 9.1, which is derived from the 1972 General Social Survey, A_h denotes the educational group of the husband and B_h denotes the educational group of the wife, so that $A_h = 2$ and $B_h = 3$ if the husband has 12 years of education and the wife has 13 to 15 years of education. The observed count n_{ij}, $1 \le i \le 4$, $1 \le j \le 4$, is the number of subjects h with $A_h = i$ and $B_h = j$. It is assumed that the pairs (A_h, B_h) are independent and identically distributed with probability $p_{ij} > 0$ that $A_h = i$ and $B_h = j$. In Table 9.2, which is obtained from the 1970 United States Census, A_h' denotes the educational group of the husband and B_h'

Table 9.1

Distribution of Years of Education of Married Couples in 1972 General Social Survey[a]

Years of education of husband	Years of education of wife: 0–11	12	13–15	16+	Total
0–11	283	141	25	4	453
12	82	180	43	14	319
13–15	20	104	43	20	187
16+	4	52	41	69	166
Total:	389	477	152	107	1125

[a] Data tape from the 1972 General Social Survey, National Opinion Research Center, University of Chicago.

Table 9.2

Distribution of Years of Education of Married Couples in the 1970 United States Census[a,b]

Years of education of husband	Years of education of wife 0–11	12	13–15	16+	Total
0–11	13,501,277	5,289,985	838,513	304,007	19,933,782
12	3,399,824	8,157,086	1,258,299	460,704	13,275,913
13–15	802,064	2,547,540	1,319,579	517,813	5,186,966
16+	348,900	1,865,294	1,685,198	2,301,491	6,200,883
Total:	18,052,065	17,859,905	5,101,589	3,584,015	44,597,774

[a] U.S. Bureau of the Census (1972, p. 269).

[b] Respondents with 1 year of high school have 9 years of education, respondents with 1 year of college have 13 years of education, etc.

denotes the educational group of the wife. The number of husbands with $A'_h = i$ and $B'_h = j$ is n'_{ij}. Table 9.2 actually provides an estimate of n'_{ij} based on a very large sample; however, exposition will be simplified and the analysis will hardly be affected if the numbers in the table are regarded as the observed counts n'_{ij}.

The Census does provide a cross-classification of educations of husbands and wives. If such a cross-classification were not available but marginal totals such as n'^{A}_{i} were available, the cross-classification for the 1970 Census could be estimated by use of the available marginal totals, together with the cross-classification from the 1972 General Social Survey. Estimation will be considered in this section under Cases (A), (B), and (C).

Case (A)

In this case, it is assumed that the only data available from the 1970 Census are the marginal totals n'^{A}_{i} which specify the distribution of education among husbands, and it is assumed that $\lambda^{B}_{j} = \lambda'^{B}_{j}$ and $\lambda^{AB}_{ij} = \lambda'^{AB}_{ij}$. Thus the conditional logit

$$\tau^{B\cdot A}_{jj'\cdot i} = \log(p_{ij}/p_{ij'}) = \log(m_{ij}/m_{ij'}) = (\lambda^{B}_{j} - \lambda^{B}_{j'}) + (\lambda^{AB}_{ij} - \lambda^{AB}_{ij'})$$

is equal to the logit

$$\tau'^{B\cdot A}_{jj'\cdot i} = \log(n'_{ij}/n'_{ij'}) = (\lambda'^{B}_{j} - \lambda'^{B}_{j'}) + (\lambda'^{AB}_{ij} - \lambda'^{AB}_{ij'}).$$

Equivalently, the conditional probability $p^{B\cdot A}_{j\cdot i} = p_{ij}/p^{A}_{i}$ that $B_h = j$ given that $A_h = i$ is equal to the fraction $p'^{B\cdot A}_{j\cdot i} = n'_{ij}/n'^{A}_{i}$ of husbands in the Census in educational group i with wives in educational group j.

The assumption that $p_{j \cdot i}^{B \cdot A} = p_{j \cdot i}'^{B \cdot A}$ is actually a questionable one given the complete version of Table 9.2. The chi-square statistics for this model are

$$X^2 = \sum_i \sum_j (n_{ij} - \hat{m}_{ij})^2 / \hat{m}_{ij} = 25.3$$

and

$$L^2 = 2 \sum_i \sum_j n_{ij} \log(n_{ij}/\hat{m}_{ij}) = 25.6,$$

where $m_{ij} = n_i^A p_{j \cdot i}'^{B \cdot A}$. Since the corresponding multinomial response model has the form

$$\log(m_{ij}/p_{j \cdot 1}'^{B}) = \alpha_i^A,$$

where $\alpha_i^A = \log n_i^A$, there are $(4 - 1)4 = 12$ degrees of freedom. Thus X^2 and L^2 have significance levels between 1 and 2 percent. In practice, no way would normally exist to test the assumption that $p_{j \cdot i}^{B \cdot A} = p_{j \cdot i}'^{B \cdot A}$, for the adjustment procedure would normally be used when $p_{j \cdot i}'^{B \cdot A}$ is unknown. This test of the hypothesis $p_{j \cdot i}^{B \cdot A} = p_{j \cdot i}'^{B \cdot A}$ serves to indicate that the assumptions commonly made in adjustment procedures may well fail to hold. Despite the questionable nature of the assumption that $p_{j \cdot i}^{B \cdot A} = p_{j \cdot i}'^{B \cdot A}$, this case can still be studied to illustrate procedures appropriate when the assumption is valid. This case will also be useful as an indication of how the distribution of education of wives in the 1970 Census would change if the $p_{j \cdot i}'^{B \cdot A}$ were the $p_{j \cdot i}^{B \cdot A}$ in the 1972 General Social Survey.

Estimation Equations

The estimate $\hat{m}'_{ij}$ of $m'_{ij} = n'_{ij}$ is defined in the Deming and Stephan method by the following equations:

$$m_i'^A = n_i'^A, \tag{9.5}$$

$$\hat{m}_{ij} = n_{ij}, \tag{9.6}$$

$$\hat{m}'_{ij} = \hat{m}_{ij} \exp(\hat{\kappa} + \hat{\kappa}_i^A), \tag{9.7}$$

and

$$\sum \hat{\kappa}_i^A = 0. \tag{9.8}$$

Equation (9.5) restrains $m_i'^A$ so that it is equal to the known marginal total $n_i'^A$. Equation (9.6) provides the maximum likelihood estimate of m_{ij} under the saturated model used in this section for the $\log m_{ij}$. Equations (9.7) and (9.8) reflect the relationship

$$\begin{aligned} \log m'_{ij} - \log m_{ij} &= (\lambda' - \lambda) + (\lambda_i'^A - \lambda_i^A) + (\lambda_j'^B - \lambda_j^B) + (\lambda_{ij}'^{AB} - \lambda_{ij}^{AB}) \\ &= \kappa + \kappa_i^A \end{aligned} \tag{9.9}$$

which results if $\lambda_j'^B = \lambda_j^B$, $\lambda_{ij}'^{AB} = \lambda_{ij}^{AB}$, $\kappa = \lambda' - \lambda$, and $\kappa_i^A = \lambda_i'^A - \lambda_i^A$.

Equations (9.5) to (9.8) define maximum likelihood estimates of the m'_{ij} if n'^B_j and n^B_j are not fixed by the sampling procedure. In the case under study, the n'_{ij} are all fixed, so that these equations provide an approximation to the maximum likelihood estimates.

These equations are satisfied if

$$\hat{\kappa} = \frac{1}{s} \sum_i \log(n'^A_i/n^A_i), \tag{9.10}$$

$$\hat{\kappa}^A_i = \log(n'^A_i/n^A_i) - \hat{\kappa}, \tag{9.11}$$

and

$$\hat{m}'_{ij} = (n'^A_i/n^A_i)\hat{m}_{ij} = (n'^A_i/n^A_i)n_{ij}. \tag{9.12}$$

Thus $\hat{m}'_{ij}$ is found by a proportionate adjustment of each row of Table 9.1.

Confidence Intervals

Since n'^A_i is fixed, $\hat{m}'_{ij}$ has an EASD of

$$s(\hat{m}'_{ij}) = n'^A_i s(n_{ij}/n^A_i) = n'^A_i[n_{ij}(n^A_i - n_{ij})/(n^A_i)^3]^{1/2}. \tag{9.13}$$

The formula for $s(n_{ij}/n^A_i)$ may be found in Section 2.1 of Volume 1. An approximate 95 percent confidence interval for m'_{ij} has lower bound

$$\hat{m}'_{ij} - 1.96s(\hat{m}'_{ij})$$

and upper bound

$$\hat{m}'_{ij} + 1.96s(\hat{m}'_{ij}).$$

The approximation improves as the m_{ij} become large. Results are summarized in Table 9.3. The most notable feature of the table is a tendency for the fitted educational levels of wives to be higher than the levels found in Table 9.2.

Case (B)

This case is very similar to Case (A), except that now the only data available from the 1970 Census are the marginal totals n'^B_j which specify the distribution of education among wives. It is now assumed that $\lambda^A_i = \lambda'^A_i$ and $\lambda^{AB}_{ij} = \lambda'^{AB}_{ij}$, so that the conditional probability $p^{A \cdot B}_{i \cdot j} = p_{ij}/p^B_j$ that a wife in the 1972 General Social Survey with education $B_h = j$ has a husband with education $A_h = i$ is equal to the corresponding fraction $p'^{A \cdot B}_{i \cdot j} = n'_{ij}/n'^B_j$ of wives in the Census in education group j with husbands in educational group i.

The assumption that $p^{A \cdot B}_{i \cdot j} = p'^{A \cdot B}_{i \cdot j}$ is even more doubtful than the assumption $p^{B \cdot A}_{j \cdot i} = p'^{B \cdot A}_{j \cdot i}$, for

$$X^2 = \sum_i \sum_j (n_{ij} - \hat{m}_{ij})^2/\hat{m}_{ij} = 36.0$$

Table 9.3

Estimated Distribution of Education of Husbands and Wives in 1970 Based on Table 9.1 and on the Marginal Distribution of Education of Husbands in Table 9.2[a]

Education in years of husband	Education in years of wife			
	0–11	12	13–15	16+
0–11	12,453,000	6,205,000	1,100,000	176,000
	453,000	434,000	214,000	88,000
	11,564,000	5,355,000	681,000	4,000
	13,342,000	7,054,000	1,519,000	348,000
12	3,413,000	7,491,000	1,790,000	583,000
	325,000	369,000	254,000	152,000
	2,776,000	6,769,000	1,292,000	284,000
	4,049,000	8,214,000	2,287,000	881,000
13–15	555,000	2,885,000	1,193,000	535,000
	117,000	188,000	160,000	117,000
	325,000	2,515,000	879,000	325,000
	785,000	3,254,000	1,506,000	785,000
16+	149,000	1,942,000	1,532,000	2,577,000
	74,000	223,000	208,000	237,000
	5,000	1,505,000	1,125,000	2,113,000
	294,000	2,380,000	1,938,000	3,042,000

[a] First line is estimated cell counts, second line is EASD, and third and fourth lines are respective approximate lower and upper 95 percent confidence bounds for cell count.

and

$$L^2 = 2 \sum_i \sum_j n_{ij} \log(n_{ij}/\hat{m}_{ij}) = 34.8,$$

where $\hat{m}_{ij} = n_j^B p_{i \cdot j}^{\prime A \cdot B}$. As in Case (A), there are 12 degrees of freedom, so the significance levels are less than 0.001. Nonetheless, this case will also be examined for illustrative purposes.

Estimation Equations

In this case, the Deming–Stephan approach yields the equations

$$m_j'^B = n_j'^B, \tag{9.14}$$

$$\hat{m}_{ij} = n_{ij}, \tag{9.15}$$

$$\hat{m}'_{ij} = \hat{m}_{ij} \exp(\hat{\kappa} + \hat{\kappa}_j^B), \tag{9.16}$$

$$\sum \hat{\kappa}_j^B = 0. \tag{9.17}$$

Their motivation is similar to motivation of (9.5) to (9.8). They are satisfied if

$$\hat{m}'_{ij} = n_j'^B(n_{ij}/n_j^B).$$

The EASD is

$$s(\hat{m}'_{ij}) = n_j'^B[n_{ij}(n_j^B - n_{ij})/(n_j^B)^3]^{1/2} = n_j'^B s(n_{ij}/n_j^B),$$

and an approximate 95 percent confidence interval has lower bound

$$\hat{m}'_{ij} - 1.96s(\hat{m}'_{ij})$$

and upper bound

$$\hat{m}'_{ij} + 1.96s(\hat{m}'_{ij}).$$

Results for this case are summarized in Table 9.4. Note a tendency for the estimated educational levels of husbands to be lower than the observed levels in Table 9.2.

Table 9.4

Estimated Distribution of Educations of Husbands and Wives in 1970 Based on Table 9.1 and on the Marginal Distributions of Educations of Wives in Table 9.2[a]

Education in years of husband	Education in years of wife			
	0–11	12	13–15	16+
0–11	13,133,000	5,279,000	839,000	134,000
	408,000	373,000	153,000	65,700
	12,334,000	4,548,000	538,000	5,000
	13,932,000	6,011,000	1,140,000	253,000
12	3,805,000	6,740,000	1,443,000	469,000
	373,000	396,000	186,000	117,000
	3,074,000	5,963,000	1,078,000	240,000
	4,537,000	7,516,000	1,809,000	698,000
13–15	928,000	3,894,000	1,443,000	670,000
	202,000	338,000	186,000	135,000
	532,000	3,232,000	1,078,000	405,000
	1,324,000	4,556,000	1,809,000	935,000
16+	186,000	1,947,000	1,376,000	2,311,000
	92,000	255,000	184,000	166,000
	5,000	1,447,000	1,016,000	1,986,000
	367,000	2,447,000	1,736,000	2,636,000

[a] The format is the same as in Table 9.4.

Case (C)

Here the marginal distributions of educations are known from the Census for both husbands and wives but the joint distribution is unknown. The assumption is also made that $\lambda_{ij}^{AB} = \lambda_{ij}^{\prime AB}$, so that the log cross-product ratios

$$\tau_{(ii')(jj')}^{AB} = \log\left(\frac{m_{ij}m_{i'j'}}{m_{ij'}m_{i'j}}\right) \quad \text{and} \quad \tau_{(ii')(jj')}^{\prime AB} = \log\left(\frac{n'_{ij}n'_{i'j'}}{n'_{ij'}n'_{i'j}}\right)$$

are equal for all i, i', j, and j'. As noted in Exercise 9.1, this assumption is questionable, just as were the assumptions in Cases (A) and (B). As in these cases, analysis will still proceed to illustrate procedures and to indicate the effects on the Census of changing the log cross-product ratios $\tau_{(ii')(jj')}^{\prime AB}$ observed in the Census to the corresponding cross-product ratios encountered in the 1972 General Social Survey. As in the other cases, the assumption that $\lambda_{ij}^{AB} = \lambda_{ij}^{\prime AB}$ cannot be checked in most real examples since the individual n'_{ij} are not known. This example is exceptional in the respect.

Estimation Equations

Here the estimation equations are

$$m_i^{\prime A} = n_i^{\prime A}, \tag{9.18}$$

$$m_j^{\prime B} = n_j^{\prime B}, \tag{9.19}$$

$$\hat{m}_{ij} = n_{ij}, \tag{9.20}$$

$$\hat{m}'_{ij} = \hat{m}_{ij} \exp(\hat{\kappa} + \hat{\kappa}_{ij}^{AB} + \hat{\kappa}_j^B) \tag{9.21}$$

$$\sum \hat{\kappa}_i^A = \sum \hat{\kappa}_j^B = 0. \tag{9.22}$$

Equations (9.18) and (9.19) reflect the fact that $n_i^{\prime A}$ and $n_j^{\prime B}$ are known, (9.20) reflects the lack of any restriction on m_{ij}, and (9.21) and (9.22) are based on the fact that

$$\begin{aligned}\log m'_{ij} - \log m_{ij} &= (\lambda' - \lambda) + (\lambda_i^{\prime A} - \lambda_i^A) + (\lambda_j^{\prime B} - \lambda_j^B) \\ &= \kappa + \kappa_i^A + \kappa_j^B \end{aligned} \tag{9.23}$$

whenever $\lambda_{ij}^{AB} = \lambda_{ij}^{\prime AB}$.

Iterative Proportional Fitting

These equations cannot generally be solved in closed form; however, the iterative proportional fitting algorithm of Deming and Stephan (1940) can be used to solve these equations. As in Section 2.5 of Volume 1, one may let the starting values m'_{ij0} be equal to the observed counts n_{ij} from the General

Table 9.5

Estimated Distribution of Educations of Husbands and Wives in the 1970 Census Based on Table 9.1 and on Marginal Totals from Table 9.2[a]

Education in years of husband	Education in years of wife			
	0–11	12	13–15	16+
0–11	13,285,000	5,598,000	909,000	141,000
	288,000	290,000	156,000	68,000
	12,720,000	5,031,000	603,000	7,000
	13,850,000	6,166,000	1,215,000	275,000
12	3,915,000	7,269,000	1,590,000	501,000
	281,000	299,000	183,000	118,000
	3,364,000	6,687,000	1,231,000	271,000
	4,466,000	7,856,000	1,950,000	733,000
13–15	664,000	2,919,000	1,105,000	498,000
	156,000	133,000	140,000	101,000
	358,000	2,658,000	830,000	301,000
	970,000	3,181,000	1,380,000	695,000
16+	188,000	2,073,000	1,497,000	2,442,000
	91,000	290,000	167,000	144,000
	11,000	1,701,000	1,170,000	2,160,000
	366,000	2,445,000	1,824,000	2,725,000

[a] Format is the same as in Table 9.3 and 9.4.

Social Survey. Proportional adjustments are then made so that for odd v, the row total m'^{A}_{iv} is equal to the observed row total n'^{A}_{i} and for even v, the column total m'^{B}_{jv} is equal to the observed column total n'^{B}_{j}. Thus

$$\begin{aligned} m'_{ij0} &= \hat{m}_{ij} = n_{ij}, \\ m'_{ij1} &= (n'^{A}_{i}/m'^{A}_{i0})m'_{ij0}, \\ m'_{ij2} &= (n'^{B}_{j}/m'^{B}_{j1})m'_{ij1}, \\ m'_{ij3} &= (n'^{A}_{i}/m'^{A}_{i2})m'_{ij2}, \end{aligned}$$

etc. As v becomes large, m'_{ijv} approaches $\hat{m}'_{ij}$. Results are shown in Table 9.5.

Standard computer programs for iterative proportional fitting can perform the necessary calculations with little difficulty. The only complication is that most programs do not use marginal totals such as n'^{A}_{i} and n'^{B}_{j} as input. Instead, a complete table with these marginal totals must be read. For example, one may set $n'_{ij} = n'^{A}_{i}n'^{B}_{j}/N'$ for use in programs such as Fay and Goodman (1975) or Haberman (1972, 1973b).

Estimated Asymptotic Variances

In Case (C), estimation of asymptotic variances is far more complicated than in Cases (A) and (B). The problem has been studied by Haberman

(1974a, pp. 381–382). Unlike other formulas for asymptotic variances which have previously been considered, the formula for $s^2(\hat{m}'_{ij})$ does not have any simple analog in weighted regression analysis. Given known weights c_{ij}, $1 \le i \le r$, $1 \le j \le s$, the asymptotic variance $s^2(\hat{d})$ of the estimate

$$\hat{d} = \sum_i \sum_j c_{ij}\hat{m}'_{ij} \qquad \text{of} \qquad d = \sum_i \sum_j c_{ij}m'_{ij}$$

may be found in a two-stage process.

In the first stage, $\hat{d}$ is decomposed into a constant term and the variable term in the following manner. Let γ, γ_i^A, γ_j^B, e_{ij}, and f_{ij} be defined so that

$$c_{ij} = e_{ij} + f_{ij}, \tag{9.24}$$

$$\sum_i f_{ij}m'_{ij} = \sum_j f_{ij}m'_{ij} = 0, \tag{9.25}$$

$$e_{ij} = \gamma + \gamma_i^A + \gamma_j^B, \tag{9.26}$$

and

$$\sum \gamma_i^A = \sum \gamma_j^B = 0. \tag{9.27}$$

Let

$$\hat{a} = \sum_i \sum_j e_{ij}\hat{m}'_{ij}$$

and

$$\hat{b} = \sum_i \sum_j f_{ij}\hat{m}_{ij},$$

so that

$$\hat{d} = \hat{a} + \hat{b}.$$

Then $\hat{a}$ has a constant value a equal to

$$\sum_i \sum_j (\gamma + \gamma_i^A + \gamma_j^B)\hat{m}'_{ij} = \gamma \sum \sum \hat{m}'_{ij} + \sum_i \gamma_i^A \hat{m}_i'^A + \sum_j \gamma_j^B m_j'^B$$
$$= \gamma N' + \sum_i \gamma_i^A n_i'^A + \sum_j \gamma_j^B n_j'^B.$$

If $\hat{b}$ has asymptotic mean

$$b = \sum_i \sum_j f_{ij}m'_{ij}$$

and asymptotic variance $\sigma^2(\hat{b})$, then $\hat{d}$ has asymptotic mean

$$d = b + a$$

and asymptotic variance $\sigma^2(\hat{d}) = \sigma^2(\hat{b})$. Thus computation of $\sigma^2(\hat{d})$ is equivalent to computation of $\sigma^2(\hat{b})$.

The second stage involves computation of $\sigma^2(\hat{b})$. This step is simple since Haberman's (1974a) results show that $\hat{b}$ has the same asymptotic variance as

$$\sum_i \sum_j m'_{ij} f_{ij} \log n_{ij}.$$

Therefore

$$\sigma^2(\hat{d}) = \sigma^2(\hat{b}) = \sum_i \sum_j f_{ij}^2 (m'_{ij})^2 / m_{ij}. \tag{9.28}$$

If $\sigma^2(b) > 0$, then the distribution of $(\hat{b} - b)/\sigma(\hat{b}) = (\hat{d} - d)/\sigma(\hat{d})$ is approximately $N(0, 1)$, with the approximation increasingly accurate as the m_{ij} become large.

To estimate $\sigma^2(\hat{d})$, the quantities $\hat{e}_{ij}$, $\hat{f}_{ij}$, $\hat{\gamma}$, $\hat{\gamma}_i^A$, and $\hat{f}_{ij}$ are defined in an analogous manner to (9.24) through (9.27). Thus

$$\hat{c}_{ij} = \hat{e}_{ij} + \hat{f}_{ij}, \tag{9.29}$$

$$\sum_i \hat{f}_{ij} \hat{m}'_{ij} = \sum_j \hat{f}_{ij} \hat{m}'_{ij} = 0, \tag{9.30}$$

$$\hat{e}_{ij} = \hat{\gamma} + \hat{\gamma}_i^A + \hat{\gamma}_j^B, \tag{9.31}$$

and

$$\sum \hat{\gamma}_i^A = \sum \hat{\gamma}_j^B = 0. \tag{9.32}$$

The estimated asymptotic variance is then

$$s^2(\hat{d}) = \sum_i \sum_j \hat{f}_{ij}^2 (\hat{m}'_{ij})^2 / \hat{m}_{ij} = \sum_i \sum_j \hat{f}_{ij}^2 (\hat{m}'_{ij})^2 / n_{ij}. \tag{9.33}$$

Solution of (9.25) to (9.28) corresponds to minimization of the weighted sum of squares

$$\sum_i \sum_j \hat{m}'_{ij} (c_{ij} - e_{ij})^2$$

subject to the constraint that for some γ, γ_i^A, and γ_j^B,

$$e_{ij} = \gamma + \gamma_i^A + \gamma_j^B \qquad \text{and} \qquad \sum \gamma_i^A = \sum \gamma_j^B = 0$$

A solution can be found by letting $\alpha_j^B = \gamma + \gamma_j^B$, $\beta_k = \gamma_k^A$, $1 \le k \le r - 1$, and

$$\begin{aligned} x_{ijk} &= 1, \qquad i = k \\ &= -1, \qquad i = r, \\ &= 0, \qquad i \ne k, \qquad i \ne r, \end{aligned}$$

so that

$$e_{ij} = \alpha_j^B + \sum_k \beta_k x_{ijk}.$$

As in Section 6.1, the weighted sum of squares is minimized if $e_{ij} = \hat{e}_{ij}$, where

$$\hat{e}_{ij} = \hat{\alpha}_j^B = \sum_k \hat{\beta}_k x_{ijk}, \tag{9.34}$$

$$\hat{\alpha}_j^B = \sum_i x_{ijk} \hat{m}'_{ij}/n_j'^B, \tag{9.35}$$

and the $\hat{\beta}_k$, $1 \leq k \leq r - 1$, satisfy the simultaneous equations

$$\sum_{l=1}^{r-1} \hat{S}_{kl}\hat{\beta}_l = \sum_i \sum_j x_{ijk} c_{ij} \hat{m}'_{ij}, \qquad 1 \leq k \leq r - 1. \tag{9.36}$$

Here

$$\hat{S}_{kl} = \sum_i \sum_j (x_{ijk} - \hat{\theta}_{jk})(x_{ijl} - \hat{\theta}_{jl})\hat{m}'_{ij} \tag{9.37}$$

and

$$\hat{\theta}_{jk} = \sum_i x_{ijk} \hat{m}'_{ij}/n_j'^B. \tag{9.38}$$

To find $s^2(\hat{m}'_{ij})$, observe that $\hat{d} = \hat{m}'_{i'j'}$ if

$$\begin{aligned} c_{i'j'} &= 1, \qquad i = i', \quad j = j', \\ &= 0, \qquad i \neq i' \quad \text{or} \quad j \neq j'. \end{aligned} \tag{9.39}$$

Thus (9.34) to (9.39) may be used to obtain the values of $s(\hat{m}'_{ij})$ shown in Table 9.5. The approximate 95 percent confidence intervals have bounds

$$\hat{m}'_{ij} - 1.96s(\hat{m}'_{ij})$$

and

$$\hat{m}'_{ij} + 1.96s(\hat{m}'_{ij}).$$

One indication in Table 9.5 of problems with the assumption that $\lambda_{ij}^{AB} = \lambda_{ij}'^{AB}$ can be seen by comparing the actual value of n'_{22} in Table 9.2 of 8,157,086 to the upper limit 7,856,000 of the corresponding confidence interval in Table 9.5. The standardized deviate $(n'_{22} - \hat{m}'_{22})/s(\hat{m}'_{22})$ has an approximate $N(0, 1)$ distribution if $\lambda_{ij}^{AB} = \lambda_{ij}'^{AB}$. Since this deviate is

$$(8{,}157{,}086 - 7{,}269{,}000)/299{,}000 = 3.30,$$

it has an approximate significance level of $2[1 - \Phi(3.30)] = 0.00096$. Even adjusting for the presence of 16 estimates $\hat{m}'_{ij}$, this deviate is still larger than can be expected under the assumption that Tables 9.1 and 9.2 have the same two-factor interactions.

In this section, the basic procedures for adjustment of data using saturated models have been explored. The Deming–Stephan procedure for estimation of unobserved counts has been provided, and a procedure has been given for estimation of EASD's. Generalizations to higher-way tables are straightforward. For examples, see Exercises 9.4 and 9.5. In this next section, a generalization to unsaturated models is provided.

9.2 Adjustment of Marginal Totals Using Unsaturated Models

The adjustment procedures of the preceding section may also be applied if an unsaturated model is used with the table of n_{ij}. The only requirement is that the only λ-parameters to be restricted must be those assumed equal in Tables 9.1 and 9.2 to the corresponding λ'-parameters. For example, in Cases (A), (B), and (C) of Section 9.1, one may assume quasi-symmetry, in which case, $\lambda_{ij}^{AB} = \lambda_{ji}^{AB}$. This restriction is permitted since in Cases (A), (B), and (C), $\lambda_{ij}^{AB} = \lambda_{ij}^{\prime AB}$.

The estimation procedure for the unsaturated case is very similar to the estimation procedure in the saturated case, except that $\hat{m}_{ij}$ is now the maximum likelihood estimate of m_{ij} under the unsaturated model for the n_{ij}. Computation of estimated asymptotic variances is more complex than in the cases explored in Section 9.1 since the estimated asymptotic variance of a linear combination of the log $\hat{m}_{ij}$ in an unsaturated model has a relatively complicated expression compared to the estimated asymptotic variance of a linear combination of the log n_{ij}.

Computation of $\hat{m}_{ij}$

Procedures will be illustrated in the section by an application of quasi-symmetry to Table 9.1 when the assumptions in Case (C) are made. To begin analysis, estimates $\hat{m}_{ij}$ are computed as in Section 8.1 under the quasi-symmetry model

$$\log m_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB},$$

$$\sum \lambda_i^A = \sum \lambda_j^B = \sum_i \lambda_{ij}^{AB} = \sum_j \lambda_{ij}^{AB} = 0,$$

$$\lambda_{ij}^{AB} = \lambda_{ji}^{AB}.$$

Results are shown in Table 9.6. Since there are 3 degrees of freedom, $X^2 = 3.50$, and $L^2 = 3.44$, the quasi-symmetry model is consistent with Table 9.1. Thus the $\hat{m}_{ij}$ do provide acceptable estimates of the means m_{ij}.

Table 9.6

Estimated Cell Means in Table 9.1 under the Quasi-Symmetry Model

Education in years of husband	Education in years of wife			
	0–11	12	13–15	16+
0–11	283.00	145.82	21.52	2.67
12	77.18	180.00	48.01	13.81
13–15	23.48	98.99	43.00	21.53
16+	5.33	52.19	39.47	69.00

Computation of $\hat{m}'_{ij}$

Given the $\hat{m}_{ij}$, the estimates $\hat{m}'_{ij}$ for Table 9.2 satisfy

$$\hat{m}_i'^A = n_i'^A, \tag{9.40}$$

$$\hat{m}_j'^B = n_j'^B, \tag{9.41}$$

$$\hat{m}'_{ij} = \hat{m}_{ij} \exp(\hat{\kappa} + \hat{\kappa}_i^A + \hat{\kappa}_j^B), \tag{9.42}$$

$$\sum \hat{\kappa}_i^A = \sum \hat{\kappa}_j^B = 0. \tag{9.43}$$

These equations may be solved by iterative proportional fitting as in Section 9.1. The only change is in the starting value $\hat{m}'_{ij0}$, which is now $\hat{m}_{ij}$. At subsequent steps,

$$m'_{ij1} = (n_i'^A / m_{i0}'^A) m'_{ij0}, \tag{9.44}$$

$$m'_{ij2} = (n_j'^B / m_{j1}'^B) m'_{ij1}, \tag{9.45}$$

etc. Results are summarized in Table 9.7.

Estimation of Asymptotic Variances

As in Case (C) of Section 9.1, estimation of the asymptotic variance of a linear combination of the estimated means $\hat{m}'_{ij}$ is a two-stage process. The first stage is the same as in Section 9.1. The second stage is more complicated than in Section 9.1 due to the use of an unsaturated model.

As in Section 9.1, it is helpful to begin by considering the asymptotic variance $\sigma^2(\hat{d})$ of a linear combination

$$\hat{d} = \sum_i \sum_j c_{ij} \hat{m}'_{ij}$$

Table 9.7

Estimated Distribution of Educations of Husbands and Wives in the 1970 Census Based on the Quasi-Symmetry Model for Table 9.1 and on the Marginal Totals in Table 9.2[a]

Years of education of husband	Years of education of wife			
	0–11	12	13–15	16+
0–11	13,304,000	5,760,000	780,000	94,000
	288,000	274,000	107,000	35,000
	12,740,000	5,218,000	570,000	26,000
	13,869,000	6,292,000	990,000	163,000
12	3,716,000	7,276,000	1,783,000	501,000
	300,000	141,000	79,000	85,000
	3,218,000	6,689,000	1,508,000	334,000
	4,214,000	7,863,000	2,059,000	667,000
13–15	781,000	2,764,000	1,103,000	539,000
	107,000	108,000	140,000	79,000
	571,000	2,552,000	828,000	384,000
	991,000	2,975,000	1,378,000	695,000
16+	251,000	2,065,000	1,435,000	2,449,000
	85,000	170,000	136,000	143,000
	85,000	1,731,000	1,168,000	2,169,000
	418,000	2,399,000	1,701,000	2,730,000

[a] The format is the same as in Tables 9.3, 9.4, and 9.5

of the estimated means m'_{ij}. As in Section 9.1, if $\sigma^2(\hat{d}) > 0$ and

$$d = \sum_i \sum_j c_{ij} m'_{ij},$$

then $(\hat{d} - d)/\sigma(\hat{d})$ is approximately a standard normal deviate and $\sigma^2(\hat{d})$ is the asymptotic variance of

$$\hat{g} = \sum_i \sum_j (f_{ij} m'_{ij}) \log \hat{m}_{ij},$$

where f_{ij} is defined as in (9.24) to (9.27). The only change is in the formula for the asymptotic variance of $\hat{g}$.

Since

$$\log \hat{m}_{ij} = \hat{\lambda} + \hat{\lambda}_i^A + \hat{\lambda}_j^B + \hat{\lambda}_{ij}^{AB},$$

where

$$\sum \hat{\lambda}_i^A = \sum \hat{\lambda}_j^B = \sum_i \hat{\lambda}_{ij}^{AB} = \sum_j \hat{\lambda}_{ij}^{AB} = 0$$

and $\hat{\lambda}_{ij}^{AB} = \hat{\lambda}_{ji}^{AB}$, (9.25) implies that

$$\hat{g} = \hat{\lambda} \sum_i \sum_j f_{ij} m'_{ij} + \sum_i \hat{\lambda}_i^A \sum_j f'_{ij} m'_{ij} + \sum_j \hat{\lambda}_j^B \sum_i f_{ij} m'_{ij} + \sum_i \sum_j \hat{\lambda}_{ij}^{AB} m'_{ij} f_{ij}$$
$$= \sum_i \sum_j m'_{ij} f_{ij} \hat{\lambda}_{ij}^{AB}$$
$$= \sum_{i=2}^{r} \sum_{j=1}^{i-1} u_{ij} \hat{\lambda}_{ij}^{AB},$$

where

$$u_{ij} = m'_{ij} f_{ij} + m'_{ji} f_{ji} - m'_{ii} f_{ii} - m'_{jj} f_{jj}. \tag{9.46}$$

Thus $\hat{g}$ has asymptotic variance

$$\sigma^2(\hat{d}) = \sum_{i=2}^{r} \sum_{j=1}^{i-1} \sum_{i'=2}^{r} \sum_{j'=1}^{i-1} u_{ij} u_{i'j'} v_{iji'j'}, \tag{9.47}$$

where $v_{iji'j'}$ is the asymptotic covariance of $\hat{\lambda}_{ij}^{AB}$ and $\hat{\lambda}_{i'j'}^{AB}$.

To find the $v_{iji'j'}$ for Table 9.1, observe that

$$\log m_{ij} = \alpha_j^B + \sum_{k=1}^{9} \beta_k x_{ijk},$$

where the $\alpha_j^B = \lambda + \lambda_j^B$ and the β_k and x_{ijk} are defined as in Table 9.8. The asymptotic covariance matrix of the $\hat{\beta}_k$, $1 \le k \le 9$, has elements S^{kl}, $1 \le k \le 9$, $1 \le l \le 9$, where the S_{kl} are elements of the inverse S^{-1} of the matrix S with elements

$$S_{kl} = \sum_i \sum_j (x_{ijk} - \theta_{jk})(x_{ijl} - \theta_{jl}) m_{ij} \tag{9.48}$$

and where

$$\theta_{jk} = \sum_i x_{ijk} m_{ij} / m_j^B. \tag{9.49}$$

Thus $v_{2121} = S^{44}$, $v_{3121} = S^{54}$, etc.

To compute estimated asymptotic variances $s^2(\hat{d})$, let $\hat{f}_{ij}$ be defined as in Section 9.1 and let

$$\hat{u}_{ij} = \hat{m}'_{ij} \hat{f}_{ij} + \hat{m}'_{ji} \hat{f}_{ji} - \hat{m}'_{ii} \hat{f}_{ii} - \hat{m}'_{jj} \hat{f}_{jj}. \tag{9.50}$$

Then

$$s^2(\hat{d}) = \sum_{i=2}^{r} \sum_{j=1}^{i-1} \sum_{i'=2}^{r} \sum_{j'=1}^{i'-1} \hat{u}_{ij} \hat{u}_{i'j'} \hat{v}_{iji'j'}. \tag{9.51}$$

Here $\hat{v}_{2121} = \hat{S}^{44}$, $\hat{v}_{3121} = \hat{S}^{54}$, $\hat{v}_{4121} = \hat{S}^{64}$, etc, where $\hat{S}^{-1}$ is the inverse of the matrix S with elements

$$\hat{S}_{kl} = \sum_i \sum_j (x_{ijk} - \hat{\theta}_{jk})(x_{ijl} - \hat{\theta}_{jl}) \hat{m}_{ij} \tag{9.52}$$

Table 9.8

Coefficients for the Quasi-Symmetry Model of Table 9.1

k	1	2	3	4	5	6	7	8	9
β_k	λ_1^A	λ_2^A	λ_3^A	λ_{21}^A	λ_{31}^A	λ_{41}^A	λ_{32}^A	λ_{42}^A	λ_{43}^A
x_{11k}	1	0	0	−1	−1	−1	0	0	0
x_{21k}	0	1	0	1	0	0	0	0	0
x_{31k}	0	0	1	0	1	0	0	0	0
x_{41k}	−1	−1	−1	0	0	1	0	0	0
x_{12k}	1	0	0	1	0	0	0	0	0
x_{22k}	0	1	0	−1	0	0	−1	−1	0
x_{32k}	0	0	1	0	0	0	1	0	0
x_{42k}	−1	−1	−1	0	0	0	0	1	0
x_{13k}	1	0	0	0	1	0	0	0	0
x_{23k}	0	1	0	0	0	0	1	0	0
x_{33k}	0	0	1	0	−1	0	−1	0	−1
x_{43k}	−1	−1	−1	0	0	0	0	0	1
x_{14k}	1	0	0	0	0	1	0	0	0
x_{24k}	0	1	0	0	0	0	0	1	0
x_{34k}	0	0	1	0	0	0	0	0	1
x_{44k}	−1	−1	−1	0	0	−1	0	−1	−1

and

$$\hat{\theta}_{jk} = \sum_i x_{ijk}\hat{m}_{ij}/n_j^B. \tag{9.53}$$

Estimates $\hat{v}_{iji'j'}$ are shown in Table 9.9. Estimates $s^2(\hat{m}'_{ij})$ are shown in Table 9.7, together with corresponding approximate 95 percent confidence intervals. Some modest gain is generally achieved compared to Table 9.5 in terms of width of the confidence intervals. This gain is the basic motivation for use of unsaturated models in the adjustment process.

Table 9.9

Estimates $v_{iji'j'}$ for Table 9.1

i	j	$1000\hat{v}_{ij21}$	$1000\hat{v}_{ij31}$	$1000\hat{v}_{ij41}$	$1000\hat{v}_{ij32}$	$1000\hat{v}_{ij42}$	$1000\hat{v}_{ij43}$
2	1	5.426	−0.701	−10.024	−2.677	−0.259	3.834
3	1	−0.701	12.327	−11.792	−3.339	4.750	−0.661
4	1	−10.024	−11.792	50.935	10.753	−10.269	−10.517
3	2	−2.677	−3.339	10.753	6.875	−3.953	−2.912
4	2	−0.259	4.750	−10.269	−3.395	10.344	−0.586
4	3	3.834	−0.661	−10.517	−2.912	−0.586	10.725

9.3 Related Applications

The adjustment methods described in this chapter have been used extensively to describe relationships between discrete variables. An early example appears in Yule (1912). He considers a 2×2 table with counts n_{ij}, $1 \le i \le 2$, $1 \le j \le 2$. The relationship between the two cross-classified dichotomous variables can be described by the cross-product ratio

$$q = n_{11}n_{22}/(n_{12}n_{21}).$$

Yule notes that a 2×2 table of probabilities P_{11}, P_{12}, P_{21}, and P_{22} exists with marginal probabilities

$$P_i^A = P_{i1} + P_{i2} = P_j^B = P_{ij} + P_{2j} = \tfrac{1}{2}$$

and

$$P_{11}P_{22}/(P_{12}P_{21}) = q.$$

In this table,

$$P_{22} = P_{11} = 1/[2(1 + q^{1/2})]$$

and

$$P_{12} = P_{21} = q/[2(1 + q^{1/2})].$$

Yule's coefficient of colligation, which was mentioned in Chapter 2, is

$$\begin{aligned} P_{1\cdot 1}^{A\cdot B} - P_{1\cdot 2}^{A\cdot B} &= P_{11}/P_1^B - P_{12}/P_2^B \\ &= (1 - q^{1/2})/(1 + q^{1/2}). \end{aligned}$$

Mosteller (1968) considers a more general standardization in which a 5×5 table n_{ij}, $1 \le i \le 5$, $1 \le j \le 5$, is adjusted to yield a new table of estimates M_{ij}, $1 \le i \le 5$, $1 \le j \le 5$, where for some κ, κ_i^A, and κ_j^B,

$$M_i^A = \sum_j M_{ij} = 100,\ 1 \le i \le 5,$$

$$M_j^B = \sum_i M_{ij} = 100,\ 1 \le j \le 5,$$

$$\log(n_{ij}/M_{ij}) = \kappa + \kappa_i^A + \kappa_j^B,$$

and

$$\sum_i \kappa_i^A = \sum_j \kappa_j^B = 0.$$

Computations use the iterative proportional fitting algorithm described in Section 9.1. The only change is replacement of the $n_i'^A$ and $n_j'^B$ of that section by the M_i^A and M_j^B. Similar analyses of a 3×3 and an 8×3 table appear in Fienberg (1971).

The material of this chapter is also closely related to a literature on standardization of rates that dates back at least to Registrar General of Birth, Deaths, and Marriages in England (1883). Let subjects h, $1 \le h \le N$, be observed on polytomous variables $A_h = i$ and $B_h = j$. For example, A_h might be an age group and B_h might be 1 if the subject dies in a given year and 2 otherwise. Let $p_{ij} > 0$ be the probability that $A_h = i$ and $B_h = j$, let p_i^A be the marginal probability that $A_h = i$, and let $p_{j \cdot i}^{B \cdot A}$ be the conditional probability that $B_h = j$ given that $A_h = i$. Thus the p_i^A might determine the age distribution in the population, and the $p_{1 \cdot i}^{B \cdot A}$ would be age-specific death rates. Let $\hat{p}_i^A$ be an estimate of p_i^A, and let $\hat{p}_{j \cdot i}^{B \cdot A}$ be an estimate of $p_{j \cdot i}^{B \cdot A}$. Consider a reference population with probabilities P_{ij} corresponding to p_{ij}. Let the marginal totals P_i^A be known, and assume that $P_{j \cdot i}^{B \cdot A} = p_{j \cdot i}^{B \cdot A}$. Then the marginal total P_j^B has the estimate

$$\hat{P}_j^B = \sum_i \hat{p}_{j \cdot i}^{B \cdot A}(P_i^A/\hat{p}_i^A).$$

For example, the estimates $\hat{P}_1^B$ might be an age-adjusted death rate. The standard population would have an age distribution defined by the P_i^A.

EXERCISES

9.1 Use Tables 9.1 and 9.2 to test the hypothesis

$$\log(m_{ij}/n'_{ij}) = \kappa + \kappa_i^A + \kappa_j^B,$$

where

$$\sum \kappa_i^A = \sum \kappa_j^B = 0.$$

Solution

The Newton–Raphson algorithm or Deming–Stephan algorithm may be used. In the latter case,

$$\begin{aligned} m_{ij0} &= n'_{ij}, \\ m_{ij1} &= m_{ij0} n_i^A/m_{i0}^A, \\ m_{ij2} &= m_{ij1} n_j^B/m_{j1}^B, \end{aligned}$$

etc. The chi-square statistics $X^2 = 16.6$ and $L^2 = 17.3$ have $(4 - 1)(4 - 1) = 9$ degrees of freedom, so they have approximate significance levels of 5 percent. Thus the model is doubtful. Much of the difficulty appears to involve diagonal cells, as can be seen in Table 9.10.

Table 9.10

Observed Counts and Estimated Expected Counts for the Model of Exercise 9.1[a]

Years of education of husband	Years of education of wife			
	0–11	12	13–15	16+
0–11	283	141	25	4
	286.95	134.75	22.65	8.56
12	82	180	43	14
	70.49	202.69	33.16	12.66
13–15	20	104	43	20
	24.12	91.82	50.44	20.64
16+	4	52	41	69
	7.45	47.73	45.74	65.13

[a] First line is observed count and second line is estimated expected count.

The model

$$\log(m_{ij}/n'_{ij}) = \kappa + \kappa_i^A + \kappa_j^B, \qquad i \neq j,$$

leads to an X^2 of 2.96 and an L^2 of 3.14. The degrees of freedom are 6, so the fit is quite satisfactory.

9.2 Use the Newton–Raphson algorithm to compute the estimate m'_{ij} in Table 9.5.

Solution

The estimates m'_{ij} are found from use of the Newton–Raphson algorithm for the log-linear model

$$\log(m'_{ij}/z_{ij}) = \alpha_j^B + \sum_{k=1}^{3} \beta_k x_{ijk},$$

where

$$\begin{aligned} x_{ijk} &= 1, \quad && i = k, \\ &= 0, && i \neq k, \quad i < 4, \\ &= -1, && i = 4. \end{aligned}$$

The z_{ij} are set equal to n_{ij}. The counts n'_{ij} corresponding to the m'_{ij} are chosen so that they have the observed marginal totals. For example, one may let $m'_{ij} = n_i'^A n_j'^B / N'$. One may cheat a bit since the n'_{ij} are actually known and use the actual observed values.

9.3 Verify Table 9.5.

9.4 Let polytomous variables A_h, B_h, and C_h be observed, $1 \leq h \leq N$. Assume that the triples (A_h, B_h, C_h) are independently distributed with probability $p_{ijk} > 0$ that $A_h = i$, $B_h = j$, and $C_h = k$, $1 \le i \le r$, $1 \le j \le s$, and $1 \le k \le t$. Let n_{ijk} be the observed number of h such that $A_h = i$, $B_h = j$, and $C_h = k$, so that n_{ijk} has mean $m_{ijk} = Np_{ijk}$. Assume that in a population under study, variables A'_h, B'_h, and C'_h, $1 \le h \le N$, are distributed so that $p'_{ijk} > 0$ is the probability that $A'_h = i$, $B'_h = j$, and $C'_h = k$, $1 \le i \le r$, $1 \le j \le s$, and $1 \le k \le t$. Let n_{ijk} be the number of h such that $A'_h = i$, $B'_h = j$, and $C'_h = k$, and let $m'_{ijk} = N'p'_{ijk}$ denote the mean of n'_{ijk}. Let

$$\log m_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

and

$$\log m'_{ijk} = \lambda + \lambda_i'^A + \lambda_j'^B + \lambda_k'^C + \lambda_{ij}'^{AB} + \lambda_{ik}'^{AC} + \lambda_{jk}'^{BC} + \lambda_{ijk}'^{ABC}.$$

Assume that $\lambda_{ijk}^{ABC} = \lambda'^{ABC}_{ijk}$, where the usual constraints on λ- and λ'-parameters apply. Let the marginal totals $n_{ij}'^{AB}$, $n_{ik}'^{AC}$, and $n_{jk}'^{BC}$ be observed. Describe an iterative proportional fitting procedure for use in estimation of the mean $m_{ijk} = N'p'_{ijk}$ of n'_{ijk}.

Solution

One may let

$$\begin{aligned} m'_{ijk0} &= n_{ijk} \\ m'_{ijk1} &= m'_{ijk0} n_{ij}'^{AB} / m_{ij0}'^{AB}, \\ m'_{ijk2} &= m'_{ijk1} n_{ik}'^{AC} / m_{ik0}'^{AC}, \\ m'_{ijk3} &= m'_{ijk2} n_{jk}'^{BC} / m_{jk0}'^{BC}, \end{aligned}$$

etc. The limit $\hat{m}'_{ijk}$ of the $\hat{m}'_{ijk\nu}$ is the maximum likelihood estimate of m'_{ijk}.

9.5 Let A_h, B_h, C_h, n_{ijk}, N, p_{ijk}, m_{ijk}, A'_h, B'_h, C'_h, n'_{ijk}, N', p'_{ijk}, and m'_{ijk} be defined as in Exercise 9.4. Assume that the λ- and λ'-parameters in Exercise 9.4 satisfy $\lambda_{ij}^{AB} = \lambda_{ij}'^{AB}$, $\lambda_{ik}^{AC} = \lambda_{ik}'^{AC}$, and $\lambda_{ijk}^{ABC} = \lambda_{ijk}'^{ABC}$, and assume that the observed marginal totals are the $n_i'^A$ and $n_{jk}'^{BC}$. Find an iterative proportional fitting algorithm for estimation of m'_{ijk}.

Solution

Here

$$\begin{aligned} m'_{ijk0} &= n_{ijk}, \\ m'_{ijk1} &= n_i'^A m'_{ijk0} / m_{ijk0}'^A, \\ m'_{ijk2} &= n_{jk}'^{BC} m'_{ijk1} / n_{ijk1}'^{BC}, \\ m'_{ijk3} &= n_i'^A m'_{ijk2} / m_{i2}'^A, \end{aligned}$$

etc. The $m'_{ijk\nu}$ approach a limit $\hat{m}'_{ijk}$ which is the maximum likelihood estimate of m'_{ijk}.

9.6 Derive Table 9.7.

10 *Latent-Class Models*

The latent-class models of this chapter assume that the relationships between several observed (manifest) discrete variables can be explained by use of a log–linear model involving both these variables and one or more unobserved (latent) discrete variables. The usual assumption made in these models is that the manifest variables are conditionally independent given the latent variable or variables, so that for any given manifest variable, the other observed variables provide no information on that given variable beyond information provided by the latent variable or variables. In this sense, the latent variables or variables fully account for the observed relationships among the manifest variables. Readers familiar with factor analysis may find parallels between factor analysis models for continuous data and latent-class models for discrete data.

The treatment of latent-class analysis in this chapter is most closely related to work of Goodman (1974a,b) and Haberman (1974c, 1976b, 1977). This work builds on a substantial earlier literature. The most extensive treatment of latent-class analysis is in Lazarsfeld and Henry (1968). Important earlier papers include Lazarsfeld (1950a,b), Green (1951, 1952), Anderson (1954, 1959), Gibson (1955), McHugh (1956), and Madansky (1960). Except for McHugh (1956), this earlier literature emphasizes estimation by determinantal equations. This method is generally cumbersome and inefficient. The approach of Goodman and Haberman has roots in the gene-counting method for estimation of gene frequencies. For early discussion of this algorithm, see Ceppellini, Siniscalco, and Smith (1955) and Smith (1956). The models of this chapter are related to an extensive literature on incomplete data. For a very helpful review, see Dempster, Laird, and Rubin (1977).

Sections 10.1–10.3 of this chapter consider computation of maximum likelihood estimates, chi-square statistics, and adjusted residuals for the basic latent-class model in which a single latent variable is present and the only assumption made is that the manifest variables are conditionally independent given the latent variable. Techniques used in these sections parallel techniques previously used with log-linear models. Maximum likelihood equations are similar in form, an iterative proportional fitting algorithm is available, and a variant of the Newton–Raphson algorithm called the scoring algorithm reduces to a series of weighted regression problems. As with log-linear models, asymptotic variances can be computed through analogies to weighted regression problems. Despite these similarities, analysis is more difficult than in the case of log–linear models. Starting values for iterative computations are not easily chosen, multiple solutions to the maximum likelihood equations exist, and iterative computations proceed more slowly than in the case of comparable log-linear models.

In Sections 10.4 and 10.5, a variety of less traditional latent-structure models are considered. The general estimation and testing procedures for these models are quite similar to those for the more traditional model. Section 10.6 summarizes the general theory of latent-class models used in this chapter.

10.1 Maximum Likelihood Equations for the Traditional Latent-Class Model

In the traditional latent-class model, one dichotomous or polytomous latent variable and more than one dichotomous or polytomous manifest variables are present. The only assumption made is the local independence assumption that the manifest variables are conditionally independent given the latent variable.

To illustrate the traditional model, it will be applied to Table 7.29 under the condition that the latent variable is dichotomous. For subject h, let X_h denote the latent variable. As in Chapter 7, let B_h, D_h, and F_h denote the responses to questions B, D, and F, and let Y_h denote the year of survey. Thus B_h, D_h, F_h, and Y_h are the manifest variables. In the traditional latent-class model, B_h, D_h, F_h, and Y_h are conditionally independent given X_h. Thus if X_h were known, knowledge of the attitude B_h toward legal abortions for married women not desiring more children provides no added information concerning the attitude D_h toward legal abortions for women in families with low incomes or the attitude F_h toward legal abortions for unmarried women not wishing to marry the father. To describe the appropriate log–linear model, assume that given the Y_h, $1 \leq h \leq N$, the responses (X_h, B_h, D_h, F_h), $1 \leq h \leq N$, are independently distributed, with $p_{xijk \cdot l}^{XBDF \cdot Y} > 0$ the probability that $X_h = x$, $B_h = i$, $D_h = j$, and $F_h = k$ given that $Y_h = l$. Let n_{xijkl} be the number of subjects with $X_h = x$, $B_h = i$, $D_h = j$, $F_h = k$, and $Y_h = l$, and let $m_{xijkl} = n_l^Y p_{xijk \cdot l}^{XBDF \cdot Y}$ denote the conditional expected value of n_{xijkl} given the number n_l^Y of subjects responding in year l. Then it is assumed that

$$\begin{aligned}\log m_{xijkl} = \lambda + \lambda_x^X + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y \\ + \lambda_{xi}^{XB} + \lambda_{xj}^{XD} + \lambda_{xk}^{XF} + \lambda_{xl}^{XY},\end{aligned} \tag{10.1}$$

where

$$\begin{aligned}\sum_x \lambda_x^X = \sum_i \lambda_i^B = \cdots = \sum_l \lambda_l^Y = \sum_x \lambda_{xi}^{XB} = \sum_i \lambda_{xi}^{XB} \\ = \cdots = \sum_x \lambda_{xl}^{XY} = \sum_l \lambda_{xl}^{XY} = 0.\end{aligned} \tag{10.2}$$

Since the X_h are not observed, the frequencies n_{xijkl} are not known. Therefore, ordinary methods for analysis of log–linear models cannot be employed. However, it is still possible to estimate the means m_{xijkl} and the λ-parameters from the observed marginal totals n_{ijkl}^{BDFY}. These totals are known since n_{ijkl}^{BDFY} is the number of subjects h in year $Y_h = l$ with the observed responses $B_h = i$, $D_h = j$, and $F_h = k$. Given the observed marginal totals n_{ijkl}^{BDFY}, the following general rule for derivation of maximum likelihood estimates in latent-structure models may be used. The same maximum likelihood equations apply as in the ordinary case in which all frequency counts are directly observed, except that the unobserved counts are replaced by their estimated conditional expected values given the observed marginal totals. This rule can be derived from Haberman (1974c).

To illustrate the rule, consider the model defined by (10.1) and (10.2). Since the model is of hierarchical form and the generating class consists of XB, XD, XF, and XY, the maximum likelihood equations under direct observation would be

$$\hat{m}_{xi}^{XB} = n_{xi}^{XB}, \qquad \hat{m}_{xj}^{XD} = n_{xj}^{XD},$$
$$\hat{m}_{xk}^{XF} = n_{xk}^{XF}, \qquad \hat{m}_{xl}^{XY} = n_{xl}^{XY},$$
$$\log \hat{m}_{xijkl} = \hat{\lambda} + \hat{\lambda}_i^B + \hat{\lambda}_j^D + \hat{\lambda}_k^F + \cdots + \hat{\lambda}_{xl}^{XY},$$
$$\sum_x \hat{\lambda}_x^X = \sum_i \hat{\lambda}_i^B = \cdots = \sum_x \hat{\lambda}_{xl}^{XY} = \sum_i \hat{\lambda}_{xl}^{XY} = 0.$$

Since n_{xi}^{XB}, n_{xj}^{XD}, n_{xk}^{XF}, and n_{xl}^{XY} are not observed, these equations cannot be used in practice. However, given the n_{ijkl}^{BDFY}, the conditional expected value of n_{xijkl} has an estimate

$$\hat{n}_{xijkl} = (\hat{m}_{xijkl}/\hat{m}_{ijkl}^{BDFY})n_{ijkl}^{BDFY}.$$

The maximum likelihood equations can then be written as

$$\hat{m}_{xi}^{XB} = \hat{n}_{xi}^{XB}, \qquad \hat{m}_{xj}^{XD} = \hat{n}_{xj}^{XB},$$
$$\hat{m}_{xk}^{XF} = \hat{n}_{xk}^{XF}, \qquad \hat{m}_{xl}^{XY} = \hat{n}_{xl}^{XY},$$
$$\log \hat{m}_{xijkl} = \hat{\lambda} + \hat{\lambda}_i^B + \hat{\lambda}_j^D + \hat{\lambda}_k^F + \cdots + \hat{\lambda}_{xl}^{XY},$$
$$\sum_x \hat{\lambda}_x = \sum_i \hat{\lambda}_i^B = \cdots = \sum_x \hat{\lambda}_{xl}^{XY} = \sum_l \hat{\lambda}_{xl}^{XY} = 0.$$

Despite the general resemblance between maximum likelihood equations for latent-class models and maximum likelihood equations for ordinary log–linear models, many difficulties arise in the latent-class case that are not encountered in the ordinary case. These problems primarily reflect the lack of a unique solution to the maximum likelihood equations for latent-class models. An added problem arises since solutions exist that are not maximum likelihood estimates. For example the maximum likelihood equations are satisfied if $a_1 + a_2 = 1$, $a_1 > 0$, $a_2 > 0$, and

$$\hat{m}_{xijkl} = a_x n_i^B n_j^D n_k^F n_l^Y / N^3;$$

however, in general, this estimate for m_{xijk} is a maximum likelihood estimate not for the latent-class model of interest but for the independence model in which $a_x = n_x^X/N$, the proportion of X_h equal x, and

$$\log m_{xijkl} = \lambda + \lambda_x^X + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y.$$

In addition, if $\hat{m}_{xijkl}$ satisfies the maximum likelihood equations, then so does the estimate

$$\hat{m}'_{xijkl} = \hat{m}_{(3-x)ijkl}$$

in which the indexes $x = 1$ and $x = 2$ are interchanged. In other words, the data cannot be used to determine which latent class is which. Consequently, maximum likelihood estimates are not uniquely defined. Given these problems, some care must be taken to ensure that a solution of the maximum likelihood equations is indeed a maximum likelihood estimate. The next two sections consider this issue in terms of two computational procedures for derivation of the estimates $\hat{m}_{xijkl}$.

10.2 Iterative Proportional Fitting and Latent-Class Models

Iterative proportional fitting has been adapted by Goodman (1974a,b) and Haberman (1976b) for use with latent-class models. The basic concept used goes back at least to Ceppellini, Siniscalco, and Smith (1955). The iterative proportional fitting algorithm discussed in this section can be applied to any latent-class model based on a hierarchical log-linear model. The algorithm normally converges to a solution of the maximum likelihood equations; however, convergence is much slower generally than with directly observed frequency counts. A precise statement concerning convergence conditions may be found in Haberman (1976b).

Ordinary routines for iterative proportional fitting of hierarchical log-linear models can readily be adapted for use with latent-class models. Computations in this section are performed through the iterative proportional fitting algorithm available in the SNAP system at the University of Chicago. Several variations of the same basic algorithm are possible. The version presented here corresponds to that of Goodman (1974a,b), although the equivalence may not be immediately apparent to many readers.

To begin the algorithm, an initial approximation m_{xijkl0} for $\hat{m}_{xijkl}$ is required. This approximation must satisfy the log-linear model under study. Thus if (10.1) and (10.2) specify the model, some λ_0, λ^X_{x0}, λ^B_{i0}, λ^D_{j0}, λ^F_{k0}, λ^Y_{l0}, λ^{XB}_{xi0}, λ^{XD}_{xj0}, λ^{XF}_{xk0}, and λ^{XY}_{xl0} must exist such that

$$\log m_{xijkl0} = \lambda_0 + \lambda^X_{x0} + \lambda^B_{i0} + \lambda^D_{j0} + \lambda^F_{k0} + \lambda^Y_{l0} + \lambda^{XB}_{xi0} + \lambda^{XD}_{xj0} + \lambda^{XF}_{xk0} + \lambda^{XY}_{xl0}$$

and the usual constraints on λ-parameters hold. The choice of m_{xijkl0} is much more important than in ordinary iterative proportional fitting. It is especially important to observe that it is *not* appropriate to set all m_{xijkl0} to 1. Determinantal estimates of m_{xijkl} may be used for m_{xijkl0}, although they are rather tedious to compute. For details, see Lazarsfeld and Henry (1968). In this section, the rather crude and arbitrary approach is adopted in which

$$\lambda^{XB}_{110} = \lambda^{XD}_{110} = \lambda^{XF}_{110} = \lambda^{XY}_{130} = -\lambda^{XY}_{110} = 1$$

and

$$\lambda_0 = \lambda^X_{x0} = \lambda^B_{i0} = \lambda^D_{j0} = \lambda^F_{k0} = \lambda^Y_{l0} = \lambda^{XY}_{120} = 0.$$

The remaining λ-parameters are determined from these equations and the given constraints. Thus

$$\log m_{1ijkl} = -\log m_{2ijkl} = l - 2i - 2j - 2k + 7.$$

This condition links $X_h = 1$ with positive responses to questions concerning abortion.

Given m_{xijkl0}, one defines

$$n_{xijkl0} = m_{xijkl0} n^{BDFY}_{ijkl} / m^{BDFY}_{ijkl0}, \qquad m_{xijkl1} = m_{xijkl0} n^{XB}_{xi0} / m^{XB}_{xi0},$$
$$m_{xijkl2} = m_{xijkl1} n^{XD}_{xj0} / m^{XD}_{xj1}, \qquad m_{xijkl3} = m_{xijkl2} n^{XF}_{xk0} / m^{XF}_{xk2},$$
$$m_{xijkl4} = m_{xijkl3} n^{XY}_{xl0} / m^{XY}_{xl3}, \qquad n_{xijkl4} = m_{xijkl4} n^{BDFY}_{ijkl} / m^{BDFY}_{ijkl4},$$
$$m_{xijkl5} = m_{xijkl4} n^{XB}_{xi4} / m^{XB}_{xi4}, \qquad \text{etc.}$$

Table 10.1

Results of Iterative Proportional Fitting with the Simple Latent-Class Model for Table 7.29

l	i	j	k	n^{BDFY}_{ijkl}	m_{1ijkl0}	m_{2ijkl0}	m_{1ijkl4}	m_{2ijkl4}	$m_{1ijkl(24)}$	$m_{2ijkl(24)}$	$m^{BDFY}_{ijkl(24)}$
1	1	1	1	334	7.399	0.135	286.52	0.27	345.58	0.16	345.74
			2	34	1.000	1.000	27.09	3.15	25.44	2.02	27.46
		2	1	12	1.000	1.000	15.68	2.08	11.03	1.32	12.35
			2	15	0.135	7.389	1.48	24.05	0.81	17.15	17.96
	2	1	1	53	1.000	1.000	42.50	5.94	41.91	4.55	46.46
			2	63	0.135	7.389	4.02	68.86	3.09	59.10	62.19
		2	1	43	0.135	7.389	2.33	45.35	1.34	38.69	40.03
			2	501	0.018	54.598	0.22	525.46	0.10	502.72	502.81
2	1	1	1	428	20.086	0.050	387.79	0.22	412.26	0.14	412.39
			2	29	2.718	0.368	36.66	2.58	30.35	1.78	32.14
		2	1	13	2.718	0.368	21.22	1.70	13.16	1.17	14.33
			2	17	0.368	2.718	2.01	19.66	0.97	15.18	16.15
	2	1	1	42	2.718	0.050	57.52	4.86	50.00	4.03	54.02
			2	53	0.368	2.718	5.44	56.29	3.68	52.32	56.00
		2	1	31	0.368	2.718	3.15	37.07	1.60	34.25	35.84
			2	453	0.050	20.086	0.30	429.54	0.12	445.01	445.01
3	1	1	1	413	54.598	0.018	437.29	0.19	416.63	0.13	416.76
			2	29	7.389	0.135	41.34	2.24	30.67	1.75	32.42
		2	1	16	7.389	0.135	23.93	1.48	13.30	1.14	14.45
			2	18	1.000	1.000	2.26	17.11	0.98	14.87	15.85
	2	1	1	60	7.389	0.135	64.87	4.23	50.53	3.94	54.47
			2	57	1.000	1.000	6.13	48.98	3.72	51.24	54.96
		2	1	37	1.000	1.000	3.55	32.26	1.61	33.54	35.16
			2	430	0.135	7.389	0.34	373.80	0.12	435.82	435.94

Convergence is leisurely; one finds that after 24 iterations, results in Table 10.1 are obtained.

To test the model, the chi-square tests statistics

$$X^2 = \sum_i \sum_j \sum_k \sum_l (n^{BDFY}_{ijkl} - \hat{m}^{BDFY}_{ijkl})^2 / \hat{m}^{BDFY}_{ijkl}$$

or

$$L^2 = 2 \sum_i \sum_j \sum_k \sum_l n^{BDFY}_{ijkl} \log(n^{BDFY}_{ijkl} / \hat{m}^{BDFY}_{ijkl})$$

may be used. Since there are 24 counts n^{BDFY}_{ijkl} and

$$1 + (2-1) + (2-1) + (2-1) + (2-1) + (3-1) + (2-1)(2-1) \\ + (2-1)(2-1) + (2-1)(2-1) + (2-1)(3-1) = 12$$

independent λ-parameters, there are 12 degrees of freedom. The observed test statistics $X^2 = 9.97$ and $L^2 = 10.07$ suggest that the fit is quite satisfactory. A more detailed examination of the model is considered in the next section, in which the scoring algorithm is used to estimate asymptotic variances of parameter estimates.

10.3 The Scoring Algorithm

A variant of the Newton–Raphson algorithm called the scoring algorithm provides an attractive method for computation of m_{xijkl0} in the model defined by (10.1) and (10.2). Versions of this procedure can be found in Smith (1956), McHugh (1956), and Haberman (1977). The algorithm applies to any latent-class model based on a log–linear model.

Just as in the case of the Newton–Raphson algorithm for directly observed data, computations resemble those in a weighted regression analysis. The model can be expressed as in Section 6.2 in terms of the independent variable Y_h and the dependent variables X_h, B_h, D_h, a d F_h. Thus

$$\begin{aligned}\log m_{xijkl} = {} & \alpha_l^Y + \lambda_1^X q_x^X + \lambda_1^B q_i^B + \lambda_1^D q_j^D + \lambda_1^F q_k^F \\ & + \lambda_{11}^{XB} q_x^X q_i^B + \lambda_{11}^{XD} q_x^X q_j^D + \lambda_{11}^{XF} q_x^X q_k^F \\ & + \lambda_{11}^{XY} q_x^X q_{l1}^Y + \lambda_{12}^{XY} q_x^X q_{l2}^Y,\end{aligned}$$

where

$$\alpha_l^Y = \lambda + \lambda_l^Y,$$

$$q_1^X = -q_2^X = q_1^B = -q_2^B = q_1^D = -q_2^D = q_1^F = -q_2^F = 1,$$

and for $1 \le l' \le 2$,

$$\begin{aligned} q_{ll'}^Y &= 1, && l = l', \\ &= 0, && l \ne l', \quad l \ne 3, \\ &= -1, && l = 3.\end{aligned}$$

Thus one may write

$$\log m_{xijkl} = \alpha_l^Y + \sum_{h=1}^{q} \beta_h u_{xijklh},$$

where

$$\beta_1 = \lambda_1^X, \qquad \beta_2 = \lambda_1^B, \qquad u_{xijkl1} = q_x^X, \qquad u_{xijkl2} = q_i^B, \qquad \text{etc.}$$

The corresponding weighted regression model has the form

$$Z_{ijkl} = \alpha_l^Y + \sum_{h=1}^{9} \beta_h \bar{u}_{ijklh} + \varepsilon_{ijkl},$$

where

$$\bar{u}_{ijklh} = \sum_x p_{x \cdot ijkl}^{X \cdot BDFY} u_{xijklh} = \sum_x m_{xijkl} u_{xijklh} / m_{ijkl}^{BDFY}$$

and the ε_{ijkl} are independently distributed errors with means of 0 and variances of $1/m_{ijkl}^{BDFY}$. Unlike the case of ordinary log-linear models, this regression model has independent variables dependent on the unknown means m_{xijkl}. This dependence will cause special complications in selection of starting values. Nonetheless, the regression model is otherwise plausible, for

$$\sum_x (\log n_{xijkl}) m_{xijkl} / m_{ijkl}^{BDFY}$$

has asymptotic mean

$$\alpha_l^Y + \sum_h \beta_h \bar{u}_{ijklh}$$

and asymptotic variance $1/m_{ijkl}^{BDFY}$.

To begin the algorithm, an initial approximation m_{xijkl0} for m_{xijkl} must be given. Determination of good initial approximations appears tedious. One possible choice is the value of m_{xijkl4} in Table 10.1. This approximation has the useful feature that m_{ijkl0}^{BDFY} is equal to n_{ijkl}^{BDFY}, which in turn is an estimate of $\hat{m}_{ijkl}^{BDFY}$.

At iteration $t \geq 0$, a working value

$$\begin{aligned} Z_{ijklt} &= \sum_x (\log m_{xijklt}) m_{xijklt} / m_{ijklt}^{BDFY} + (n_{ijkl}^{BDFY} - m_{ijklt}^{BDFY}) / m_{ijklt}^{BDFY}, \qquad t > 0, \\ &= \sum_x (\log m_{xijklt}) m_{xijklt} / m_{ijklt}^{BDFY}, \qquad t = 0, \end{aligned}$$

is computed, and estimates β_{ht} of β_h are found as in the regression model

$$Z_{ijklt} = \alpha_l^Y + \sum \beta_h \bar{u}_{ijklht} + \varepsilon_{ijklt},$$

where the ε_{ijklt} have respective means 0 and variances $1/m_{ijklt}^{BDFY}$. Thus the β_{ht} satisfy the simultaneous equations

$$\sum_h S_{ght} \beta_{ht} = w_{gt},$$

where

$$w_{gt} = \sum_i \sum_j \sum_k \sum_l (\bar{u}_{ijklgt} - \theta_{lgt})(Z_{ijklt} - \bar{Z}_t) m_{ijklt}^{BDFY},$$

$$S_{ght} = \sum_i \sum_j \sum_k \sum_l (\bar{u}_{ijklgt} - \theta_{lgt})(\bar{u}_{ijklht} - \theta_{lht}) m_{ijklt}^{BDFY},$$

and

$$\theta_{lht} = \sum_i \sum_j \sum_k \bar{u}_{ijklht} m_{ijklt}^{BDFY} / m_{lt}^Y.$$

For $t > 0$,

$$\beta_{ht} = \beta_{h(t-1)} + \delta_{ht},$$

where the δ_{ht} satisfy the simultaneous equations

$$\sum_h S_{ght}\delta_{ht} = g_{gt}$$

and

$$a_{gt} = \sum_i \sum_j \sum_k \sum_l \bar{u}_{ijklgt}(n_{ijkl}^{BDFY} - m_{ijklt}^{BDFY}).$$

Given the β_{ht}, a new estimate $m_{xijkl(t+1)}$ of $\hat{m}_{xijkl}$ is found from the equations

$$v_{xijklt} = \sum_h \beta_{ht} u_{xijklh},$$

$$g_{lt} = n_l^Y \Big/ \sum_x \sum_i \sum_j \sum_k \exp v_{xijklt},$$

and

$$m_{xijkl(t+1)} = g_{lt} \exp v_{xijklt}.$$

Normally, the m_{xijklt} approach $\hat{m}_{xijkl}$ and the β_{ht} approach $\hat{\beta}_h$ as t becomes large.

Calculations are summarized in Tables 10.2 and 10.3. Note that convergence is much more rapid than with iterative proportional fitting. As an added dividend, the asymptotic covariance matrix of the vector $\hat{\boldsymbol{\beta}}$ of $\hat{\beta}_h$ has an estimate $\hat{S}^{-1}$, where $\hat{S}$ has elements

$$\hat{S}_{gh} = \sum_i \sum_j \sum_k \sum_i (\hat{u}_{ijklg} - \hat{\theta}_{lg})(\hat{u}_{ijklh} - \hat{\theta}_{lh})\hat{m}_{ijkl}^{BDFY},$$

$$\hat{u}_{ijklh} = \sum_x u_{xijkl}\hat{m}_{xijkl}/\hat{m}_{ijkl}^{BDFY},$$

Table 10.2

Results of the Scoring Algorithms as Applied to the Simple Latent-Class Model of Table 7.29

Parameter	h	β_{h0}	β_{h1}	β_{h2}	β_{h3}	$s(\hat{\beta}_h)$
λ_1^X	1	−0.042	−0.095	−0.105	−0.106	0.086
λ_1^B	2	−0.312	−0.314	−0.316	−0.316	0.045
λ_1^D	3	0.274	0.316	0.327	0.327	0.049
λ_1^F	4	−0.006	0.010	0.012	0.012	0.039
λ_{11}^{XB}	5	1.249	1.362	1.372	1.372	0.045
λ_{11}^{XD}	6	1.239	1.379	1.396	1.397	0.049
λ_{11}^{XF}	7	1.178	1.285	1.293	1.293	0.039
λ_{11}^{XY}	8	−0.068	−0.105	−0.105	−0.105	0.026
λ_{12}^{XY}	9	0.048	0.044	0.045	0.045	0.026

Table 10.3

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Simple Latent-Class Model of Table 7.29

Year	Response to B	Response to D	Response to F	Observed count	Estimated expected count	Adjusted residual
1972	Yes	Yes	Yes	334	345.76	−1.82
			No	34	27.42	1.54
		No	Yes	12	12.32	−0.11
			No	15	18.00	−0.90
	No	Yes	Yes	53	46.41	1.20
			No	63	62.25	0.12
		No	Yes	43	40.08	0.60
			No	501	502.75	−0.23
1973	Yes	Yes	Yes	428	412.46	2.28
			No	29	32.09	−0.70
		No	Yes	13	14.29	−0.43
			No	17	16.19	0.25
	No	Yes	Yes	42	53.96	−2.13
			No	53	56.05	−0.52
		No	Yes	31	35.89	−1.02
			No	453	445.07	1.08
1974	Yes	Yes	Yes	413	416.81	−0.56
			No	29	32.37	−0.76
		No	Yes	16	14.41	0.53
			No	18	15.88	0.65
	No	Yes	Yes	60	54.40	0.99
			No	57	55.01	0.34
		No	Yes	37	35.20	0.38
			No	430	435.91	−0.81

and

$$\hat{\theta}_{lh} = \sum_i \sum_j \sum_k \hat{u}_{ijklh} \hat{m}_{ijkl}^{BDFY} / \hat{m}_l^Y.$$

Adjusted residuals shown in Table 10.3 are based on the formula $r_{ijkl} = (n_{ijkl}^{BDFY} - \hat{m}_{ijkl}^{BDFY})/\hat{c}_{ijkl}^{1/2}$, where

$$\hat{c}_{ijkl} = \hat{m}_{ijkl}^{BDFY}\left[1 - \hat{m}_{ijkl}^{BDFY}/\hat{m}_l^Y - \hat{m}_{ijkl}^{BDFY} \sum_g \sum_h (\hat{u}_{ijklg} - \hat{\theta}_{lg})(\hat{u}_{ijklh} - \hat{\theta}_{lh})\hat{S}^{gh}\right].$$

This formula corresponds to those used in Sections 6.1 and 6.2. The modified program presented at the end of the Appendix may be used to perform the necessary computations.

Given that 24 adjusted residuals are present, none of the observed adjusted residuals is unusually large. Thus Table 10.3 confirms the satisfactory nature of the fit.

Table 10.2 suggests a very strong relationship between the latent variable and each response to an abortion question. Note that an estimate $\hat{\lambda}_{11}^{XB}$ of 1.372 corresponds to an estimated log cross-product ratio

$$\hat{\tau}_{(12)(12)}^{XB} = 4\hat{\lambda}_{11}^{XB} = 5.489$$

and an estimated cross-product ratio

$$\hat{q}_{(12)(12)}^{XB} = 3^{5.49} = 242.$$

Thus the odds are estimated to be 242 times as great that subject h with a latent attitude $X_h = 1$ has a favorable attitude toward legal abortions for married women not wishing more children than are the corresponding odds if $X_h = 2$. The coefficient λ_{11}^{XB} is also fairly well determined, for an approximate 95 percent confidence interval for λ_{11}^{XB} has bounds

$$1.372 - 1.96(0.045) = 1.284 \qquad \text{and} \qquad 1.372 + 1.96(0.945) = 1.460.$$

Similar results apply with the other two questions.

By contrast, the relationship of the latent variable X_h to the year Y_h of the survey is much weaker, although X_h and Y_h are not independent. Note that $\hat{\lambda}_{11}^{XY}$ differs from 0 by about 4 times $s(\hat{\lambda}_{11}^{XY})$. Nonetheless, the coefficients $\hat{\lambda}_{11}^{XY}$, $\hat{\lambda}_{12}^{XY}$, and $\hat{\lambda}_{13}^{XY} = -\hat{\lambda}_{11}^{XY} - \hat{\lambda}_{12}^{XY}$ are all of relatively modest size. This observation is quite consistent with the earlier findings that changes in abortion attitudes from 1972 to 1974 appear quite modest in size, although not nonexistant.

As can be seen from the variations in the coefficient estimates $\hat{\lambda}_1^B$, $\hat{\lambda}_1^D$, and $\hat{\lambda}_1^F$, substantial variations exist within each latent class in probabilities of approval of legal abortions for differing reasons. Given that $X_h = 1$, the conditional probability of a favorable response to item B is estimated to be

$$\frac{\exp(\hat{\lambda}_1^B + \hat{\lambda}_{11}^{XB})}{\exp(\hat{\lambda}_1^B + \hat{\lambda}_{11}^{XB}) + \exp(-\hat{\lambda}_1^B - \hat{\lambda}_{11}^{XB})} = 0.892.$$

Corresponding estimated conditional probabilities for $X_h = 1$ and items D and F are 0.969 and 0.932, respectively. In the case of the latent class of subjects with $X_h = 2$, the estimated conditional probabilities of favorable responses to items B, D, and F are 0.033, 0.105, and 0.072, respectively. For example,

$$\frac{\exp(\hat{\lambda}_1^B - \hat{\lambda}_{11}^{XB})}{\exp(\hat{\lambda}_1^B - \hat{\lambda}_{11}^{XB}) + \exp(-\hat{\lambda}_1^B + \hat{\lambda}_{11}^{XB})} = 0.033.$$

Similar formulas apply for the other estimates. The variations in estimated conditional probabilities depend much more strongly on the latent class than on the item answered. In addition, it should be noted that for both latent classes, the estimates suggest that favorable attitudes are most likely for item D, next most likely for item F, and least likely for item B.

Given year $Y_h = l$, the estimated conditional probability that the latent variable X_h is 1 is $\hat{m}_{1l}^{XY}/\hat{m}_1^Y$. From Table 10.1, one finds estimates for 1972, 1973, and 1974 are 0.407, 0.480, and 0.488. Thus a fairly even division into latent classes exists, although in each year the latent class corresponding to unfavorable attitudes toward abortion does predominate. It should be emphasized, however, that the specific abortion situations considered do affect responses. For an example, see Exercise 10.3.

The last three sections illustrate procedures appropriate for the most conventional conventional latent-class model. A number of alternate models can also be used with few changes in technique. Some of these models based on a single latent variable are discussed in Section 10.4. Section 10.5 considers models based on several latent variables.

10.4 Alternative Latent-Class Models for Table 7.29

Numerous latent-class models based on one latent variable can be applied to Table 7.29. A few such examples are presented in this section.

The Model of Constant Interaction

A symmetry model with

$$\lambda_{xi}^{XB} = \lambda_{xj}^{XD} = \lambda_{xk}^{XF}$$

is suggested by Table 10.2. To examine this model, note that it is assumed that

$$\log m_{xijkl} = \alpha_l^Y + \lambda_1^X q_x^X + \lambda_1^B q_i^B + \lambda_1^D q_j^D + \lambda_1^F q_k^F \\ + \lambda_{11}^{XB} q_x^X (q_i^B + q_j^D + q_k^F) + \lambda_{11}^{XY} q_x^X q_{l1}^Y + \lambda_{l12}^{XY} q_x^X q_{l2}^Y.$$

The term $\lambda_{11}^{XB} q_x^X (q_i^B + q_j^D + q_k^F)$ arises since the condition $\lambda_{11}^{XB} = \lambda_{11}^{XD} = \lambda_{11}^{XF}$ implies that

$$\lambda_{11}^{XB} q_x^X q_i^B + \lambda_{11}^{XD} q_x^X q_j^D + \lambda_{11}^{XF} q_x^X q_k^F = \lambda_{11}^{XB} q_x^X (q_i^B + q_j^D + q_k^F).$$

The scoring algorithm is readily applied using for m_{xijkl0} the estimates $\hat{m}_{xijkl}$ in Table 10.1.

Results for this model are summarized in Table 10.4 and 10.5. Since $X^2 = 13.22$, $L^2 = 13.17$, and there are 14 degrees of freedom, the fit remains

Table 10.4

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Latent-Class Model of Constant Interaction, as Applied to Table 7.29

Year	Response to B	Response to D	Response to F	Observed count	Estimated expected count	Adjusted residual
1972	Yes	Yes	Yes	334	345.74	−1.81
			No	34	24.74	2.12
		No	Yes	12	14.07	−0.60
			No	15	19.18	−1.07
	No	Yes	Yes	53	47.31	1.00
			No	63	64.50	−0.24
		No	Yes	43	36.68	1.24
			No	501	502.77	−0.23
1973	Yes	Yes	Yes	428	412.40	2.29
			No	29	28.82	0.04
		No	Yes	13	16.39	−0.92
			No	17	17.24	−0.06
	No	Yes	Yes	42	55.11	−2.20
			No	53	57.96	−0.80
		No	Yes	31	32.96	−0.40
			No	453	445.12	1.07
1974	Yes	Yes	Yes	413	416.90	−0.57
			No	29	29.07	−0.02
		No	Yes	16	16.53	−0.14
			No	18	16.91	0.29
	No	Yes	Yes	60	55.59	0.74
			No	57	56.86	0.02
		No	Yes	37	32.33	0.96
			No	430	435.81	−0.80

Table 10.5

Parameter Estimates Corresponding to Table 10.4

Parameter	Estimate	EASD
λ_1^X	−0.093	0.076
λ_1^B	−0.037	0.040
λ_1^D	0.300	0.040
λ_1^F	0.018	0.041
$\lambda_{11}^{XB} = \lambda_{11}^{XD} = \lambda_{11}^{XF}$	1.349	0.023
λ_{11}^{XY}	−0.105	0.026
λ_{12}^{XY}	0.044	0.026

quite satisfactory. The increase in L^2 from the traditional model is 3.07, and the corresponding increase in degrees of freedom is 2. Thus the chi-square statistics do not indicate any problems with the model. The adjusted residuals exhibit a similar pattern to that seen in Table 10.3. As should be expected, Table 10.5 and Table 10.2 are quite similar, except that $s(\hat{\lambda}_{11}^{XB})$ is somewhat smaller in Table 10.5. This latter result is quite expected given the simplifying assumption that $\lambda_{11}^{XB} = \lambda_{11}^{XD} = \lambda_{11}^{XF}$.

The model of constant interaction has the important property that the log odds $\tau_{12\cdot x}^{G\cdot X}$ of a positive response to question $G(=B, D$, or $F)$ given latent class x has the form

$$2(\lambda_1^G + \lambda_{x1}^{XB}),$$

so that each log cross-product ratio $\tau_{(12)(12)}^{GX}$ is $4\lambda_{11}^{XB}$. Thus the conditional odds of a positive response reflect λ_1^G, a measure of the general favorability in the population of abortions for reason G, and λ_{x1}^{XB}, a measure of the general attitude of a subject in latent class x toward legal abortions.

Incomplete Latent-Class Models

So far, the models considered permit all possible combinations of the latent variable X_h and the manifest variables B_h, D_h, F_h, and Y_h. Since Guttman's (1950) early work on scaling models, many latent-class models have been proposed in which given certain values of X_h, only a limited number of combinations (B_h, D_h, F_h, Y_h) are possible. Goodman (1975) provides a thorough discussion of many such latent-class models. One such model appropriate for Table 7.29 assumes that the latent variable X_h has three possible classes. Given $X_h = 1$, the subject favors a legal abortion for all of the reasons B, D, or F. Thus $X_h = 1$ implies $B_h = D_h = F_h = 1$. Similarly, given $X_h = 3$, the subject always opposes a legal abortion for any of the reasons B, D, or F. Thus $X_h = 3$ implies $B_h = D_h = F_h = 2$. The intermediate latent class X_h corresponds to subjects who do not necessarily favor or oppose legal abortions for reasons B, D, or F. Given $X_h = 2$, it is assumed that B_h, D_h, F_h, and Y_h are conditionally independent. Thus one obtains the log-linear model

$$\log m_{xijkl} = \lambda + \lambda_x^X + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y + \lambda_{xl}^{XY}, \qquad z_{xijkl} > 0,$$
$$m_{xijkl} = 0, \quad z_{xijkl} = 0,$$

where z_{xijkl} is 1 if $x = 2$, if $i = j = k = l = x = 1$, or if $i = j = k = l = x - 1 = 2$ and z_{xijkl} is 0 otherwise. The terms λ_{xi}^{XB}, λ_{xj}^{XD}, and λ_{xk}^{XF} disappear since only one value of X_h exists under which B_h, D_h, and F_h are not fixed.

The maximum likelihood equations correspond to those for a hierarchical model with generating class consisting of XY, B, D, and F. Were the counts n_{xijkl} observed, the maximum likelihood equations would be

$$\log \hat{m}_{xijkl} = \hat{\lambda} + \hat{\lambda}_x^X + \hat{\lambda}_i^B + \hat{\lambda}_j^D + \hat{\lambda}_k^F + \hat{\lambda}_l^Y + \hat{\lambda}_{xl}^{XY}, \qquad z_{xijkl} > 0,$$
$$\hat{m}_{xijkl} = 0, \quad z_{xijkl} = 0,$$
$$\hat{m}_{xl}^{XY} = n_{xl}^{XY}, \qquad \hat{m}_i^B = n_i^B,$$
$$\hat{m}_j^D = n_j^D, \qquad \hat{m}_k^F = n_k^F.$$

Since not all the n_{xijkl} are known, the maximum likelihood equations become

$$\log \hat{m}_{xijkl} = \hat{\lambda} + \hat{\lambda}_x^X + \hat{\lambda}_i^B + \hat{\lambda}_j^D + \hat{\lambda}_k^F + \hat{\lambda}_l^Y + \hat{\lambda}_{xl}^{XY}, \qquad z_{xijkl} > 0,$$
$$\hat{m}_{xijkl} = 0, \quad z_{xijkl} = 0,$$
$$\hat{m}_{xl}^{XY} = \hat{n}_{xl}^{XY}, \qquad \hat{m}_i^B = \hat{n}_i^B = n_i^B,$$
$$\hat{m}_j^D = \hat{n}_j^D = n_j^D, \qquad \hat{m}_k^F = \hat{n}_k^F = n_k^F,$$

where

$$\hat{n}_{xijkl} = 0, \qquad z_{xijkl} = 0,$$
$$= (\hat{m}_{xijkl}/\hat{m}_{ijkl}^{BDFY})n_{ijkl}^{BDFY}, \qquad z_{xijkl} > 0.$$

The expression for $\hat{n}_{xijkl}$ simplifies if x is 2 and i, j, and k are not all equal, for then $n_{ijkl}^{BDFY} = n_{xijkl}$, $\hat{m}_{xijkl} = \hat{m}_{ijkl}^{BDFY}$, and $\hat{n}_{xijkl} = n_{ijkl}^{BDFY}$.

For an alternate form of these equations which is helpful in computations, define the counts

$$n_{ijkl}^* = n_{ijkl}, \qquad i \neq j, \quad i \neq k, \quad \text{or} \quad j \neq k,$$
$$= 0, \qquad i = j = k,$$

and let m_{ijkl}^* denote the expected value of n_{ijkl}^*. Then

$$n_{ijkl} = n_{2ijkl} = n_{ijkl}^{BDFY}, \qquad i \neq j, i \neq k, \quad \text{or} \quad j \neq k.$$

Observe that

$$\hat{m}_{1l}^{XY} = \hat{m}_{1111l} = \hat{n}_{1l}^{XY} = \hat{n}_{1111l} = n_{111l}^{BDFY}\hat{m}_{1111l}/\hat{m}_{111l}^{BDFY},$$

so that

$$\hat{m}_{111l}^{BDFY} = n_{111l}^{BDFY}.$$

Similarly,

$$\hat{m}_{222l}^{BDFY} = n_{222l}^{BDFY}.$$

Subtraction of the equation $\hat{m}_{111l}^{BDFY} = n_{111l}^{BDFY}$ from the equation $\hat{m}_1^{*B} = n_1^B$ yields the equation $\hat{m}_1^{*B} = n_1^{*B}$. Similar arguments show that

$$\hat{m}_i^{*B} = n_i^{*B}, \qquad \hat{m}_j^{*D} = n_j^{*D},$$
$$\hat{m}_k^{*F} = n_k^{*F}, \qquad \hat{m}_l^{*Y} = n_l^{*Y}.$$

Furthermore, the estimates $\hat{m}^*_{ijkl}$ satisfy the log–linear model

$$\log \hat{m}^*_{ijkl} = \log \hat{m}_{2ijkl} = \hat{\lambda}^* + \hat{\lambda}^B_i + \hat{\lambda}^D_j + \hat{\lambda}^F_k + \hat{\lambda}^{*Y}_l, \quad i \neq j, \quad i \neq k, \quad \text{or} \quad j \neq k,$$
$$\hat{m}^*_{ijkl} = 0, \quad i = j = k,$$

where $\hat{\lambda}^* = \hat{\lambda} + \hat{\lambda}^X_2$ and $\hat{\lambda}^{*Y}_l = \hat{\lambda}^Y_l + \hat{\lambda}^{XY}_{2l}$. Thus the $\hat{m}^*_{ijkl}$ are the maximum likelihood estimates from the quasi-independence model considered in Exercise 7.13 and $\hat{\lambda}^B_i$, $\hat{\lambda}^D_j$, and $\hat{\lambda}^F_k$ have the same values as in Exercise 7.13. Chi-square statistics, estimated asymptotic standard diviations of the $\hat{\lambda}^B_i$, $\hat{\lambda}^D_j$, and $\hat{\lambda}^F_k$, and adjusted residuals are also the same as in Exercise 7.13.

The only new features involve estimation of the parameters λ^{XY}_{xl} and of the means m_{1111l}, m_{2111l}, m_{2222l}, and m_{3222l}. One approach to this problem is considered in Exercise 10.5. An alternate approach can be based on the scoring algorithm for latent-class models for incomplete tables.

The Scoring Algorithm for Incomplete Tables

To apply the scoring algorithm to the three-latent-class model under study, let

$$\log m_{xijkl} = \alpha^Y_l + \lambda^X_1 q^X_{x1} + \lambda^X_2 q^X_{x2} + \lambda^B_1 q^B_i + \lambda^D_1 q^D_j + \lambda^F_1 q^F_k + \lambda^{XY}_{11} q^X_{11} q^Y_{l1} + \lambda^{XY}_{21} q^X_{x2} q^Y_{l1} + \lambda^{XY}_{12} q^X_{x1} q^Y_{l2} + \lambda^{XY}_{22} q^X_{x2} q^Y_{l2},$$

where q^B_i, q^D_j, q^F_k, and $q^Y_{ll'}$ are defined as in Section 10.3 and for $1 \leq x' \leq 2$,

$$q^X_{xx'} = \begin{cases} 1, & x = x', \\ 0, & x \neq x', \quad x \neq 3, \\ -1, & x = 3. \end{cases}$$

For starting values, let

$$\begin{aligned} m_{xijkl0} &= n^{BDFY}_{ijkl}, && i \neq j, \quad i \neq k, \quad \text{or} \quad j \neq k, \quad x = 2, \\ &= 0, && i \neq j, \quad i \neq k, \quad \text{or} \quad j \neq k, \quad x = 1 \\ & && \text{or} \quad x = 3, \\ &= n^{BDFY}_{211l} n^{BDFY}_{121l} / n^{BDFY}_{221l}, && i = j = k = 1, \quad x = 2, \\ &= n^{BDFY}_{212l} n^{BDFY}_{222l} / n^{BDFY}_{112l}, && i = j = k = 2, \quad x = 2, \\ &= n^{BDFY}_{111l} - m_{2111l0}, && x = i = j = k = 1, \\ &= n^{BDFY}_{222l} - m_{2222l0}, && x = 3, \quad i = j = k = 2, \\ &= 0, && i = j = k = 2, \quad x = 3, \\ &= 0, && i = j = k = 2, \quad x = 1, \end{aligned}$$

These estimates are based on the following considerations. If $i \neq j$, $i \neq k$, and $j \neq k$, then n_{ijkl}^{BDFY} may be used to estimate $m_{2ijkl} = m_{ijkl}^{BDFY}$. Since the log cross-product ratio

$$\begin{aligned}\tau_{(12)(12)\cdot 21l}^{BD\cdot XFY} &= \log m_{2111l} - \log m_{2211l} - \log m_{2121l} + \log m_{2221l} \\ &= \log m_{2111l} - \log m_{211l}^{BDFY} - \log m_{221l}^{BDFY} + \log m_{221l}^{BDFY} \\ &= 0,\end{aligned}$$

it follows that

$$m_{2111l} = m_{211l}^{BDFY} m_{121l}^{BDFY} / m_{221l}^{BDFY}$$

has an estimate $n_{211l}^{BDFY} n_{121l}^{BDFY} / n_{221l}^{BDFY}$. A similar argument applies to m_{2222l}. Finally, since

$$m_{111l}^{BDFY} = m_{1111l} + m_{2111l} \qquad \text{and} \qquad m_{222l}^{BDFY} = m_{2222l} + m_{3222l},$$

the remaining nonzero estimates follow. Given these initial values, the scoring algorithm presented in Section 10.3 applies with only two changes. If z_{xijkl} is 0, then the term $\log m_{xijklt}$ in the formula for Z_{ijklt} is set to 0. In addition, the formulas for g_{lt} and m_{xijklt} are now

$$g_{lt} = n_l^Y \Big/ \sum_x \sum_i \sum_j \sum_k z_{xijkl} \exp v_{xijklt}$$

and

$$m_{xijklt} = g_{lt} z_{xijkl} \exp v_{xijklt}.$$

Results of computations are summarized in Table 10.6. The principal gain from the analysis is the information provided on the interactions λ_{xl}^{XY} between latent response X_h and year Y_h of survey. The estimates $\hat{\lambda}_{11}^{XY}$, $\hat{\lambda}_{12}^{XY}$,

Table 10.6

Parameter Estimates for the Three-Latent-Class Model for Table 7.29

Parameter	Estimate	EASD
λ_1^X	1.005	0.134
λ_2^X	−1.702	0.036
λ_1^B	−0.386	0.052
λ_1^D	0.221	0.051
λ_1^F	−0.062	0.048
λ_{11}^{XY}	−0.162	0.039
λ_{21}^{XY}	0.069	0.045
λ_{12}^{XY}	0.111	0.038
λ_{22}^{XY}	−0.107	0.046

and $\hat{\lambda}_{22}^{XY}$ are also large relative to their estimated asymptotic standard deviations. Thus a relationship between response and year has been demonstrated.

10.5 Models with Several Latent Variables

Goodman (1974a,b) has studied latent-class models with several latent variables. Such models are also discussed in Haberman (1977). The same general procedures discussed earlier in this chapter apply to these models; however, computational procedures for such models often converge quite slowly and care must be taken to ensure that parameters are identifiable. The number of possible models employing several latent variables is very large. This section will consider only a few cases involving two latent variables.

One model considered by Goodman for tables such as Table 7.29 assumes that two dichotomous latent variables W_h and X_h are associated with each subject. Given the pair (W_h, X_h), B_h, D_h, F_h, and Y_h are independent. It is assumed that given W_h, B_h and Y_h are independent of X_h, and given X_h, D_h and F_h are independent of W_h. It is also assumed that given the Y_h, the vectors $(W_h, X_h, B_h, D_h, F_h)$ are independently distributed with probability $p_{wxijk\cdot l}^{WXBDF\cdot Y} > 0$ that $W_h = w$, $X_h = x$, $B_h = i$, $D_h = j$, and $F_h = k$ given that $Y_h = l$. If n_{wxijkl} is the number of subjects h such that $W_h = w$, $X_h = x$, $B_h = i$, $D_h = j$, $F_h = k$, and $Y_h = l$, then n_{wxijkl} has a mean m_{wxijkl} that satisfies the log-linear model

$$\log m_{wxijkl} = \lambda + \lambda_w^W + \lambda_x^X + \lambda_i^B + \lambda_j^D + \lambda_k^F + \lambda_l^Y + \lambda_{wx}^{WX} + \lambda_{wi}^{WB} + \lambda_{wl}^{WY} + \lambda_{xj}^{WD} + \lambda_{xk}^{XF}.$$

This model is a hierarchical log-linear model with generating class consisting of WX, WB, WY, XD, and XF. Thus the maximum likelihood equations are

$$\hat{m}_{wx}^{WX} = \hat{n}_{wx}^{WX}, \quad \hat{m}_{wi}^{WB} = \hat{n}_{wi}^{WB}, \quad \hat{m}_{wl}^{WY} = \hat{n}_{wl}^{WY},$$
$$\hat{m}_{xj}^{XD} = \hat{n}_{xj}^{XD}, \quad \hat{m}_{xk}^{XF} = \hat{n}_{xk}^{XF},$$

and

$$\log \hat{m}_{wxijkl} = \hat{\lambda} + \hat{\lambda}_w^W + \hat{\lambda}_x^X + \hat{\lambda}_i^B + \hat{\lambda}_j^D + \hat{\lambda}_k^F + \hat{\lambda}_l^Y + \hat{\lambda}_{wx}^{WX} + \hat{\lambda}_{wi}^{WB} + \hat{\lambda}_{wl}^{WY} + \hat{\lambda}_{xj}^{XD} + \hat{\lambda}_{xk}^{XF},$$

where

$$\hat{n}_{wxijkl} = (n_{ijkl}^{BDFY}/\hat{m}_{ijkl}^{BDFY})\hat{m}_{wxijkl}.$$

Maximum likelihood estimates can be computed using the scoring algorithm or using the iterative proportional fitting algorithm. In the iterative proportional fitting algorithm, an initial estimate $m_{wxijkl0}$ is used such that $\log m_{wxijkl0}$ satisfies the log-linear model. At subsequent iterations,

$$n_{wxijkl0} = (n_{ijkl}^{BDFY}/m_{ijkl0}^{BDFY})m_{wxijkl0}, \qquad m_{wxijkl1} = (n_{wx0}^{WX}/m_{wx0}^{WX})m_{wxijkl0},$$
$$m_{wxijkl2} = (n_{wi0}^{WB}/m_{wi1}^{WB})m_{wxijkl1}, \qquad m_{wxijkl3} = (n_{wl0}^{WY}/m_{wl2}^{WY})m_{wxijkl2},$$
$$m_{wxijkl4} = (n_{xj0}^{XD}/m_{xj3}^{XD})m_{wxijkl3}, \qquad m_{wxijkl5} = (n_{xk0}^{XF}/m_{xk4}^{XF})m_{wxijkl4},$$
$$n_{wxijkl5} = (n_{ijkl}^{BDFY}/m_{ijkl5}^{BDFY})m_{wxijkl5}, \qquad m_{wxijkl6} = (n_{wx5}^{WX}/m_{wx5}^{WX})m_{wxijkl5}, \qquad \text{etc.}$$

This algorithm converges unusually slowly in the case of Table 7.29 and the scoring algorithm is also difficult to apply, apparently due to very poorly determined parameters. Goodman's applications of this model also lead

Table 10.7

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Asymmetric Model for Table 7.29

Year	Response to B	Response to D	Response to F	Observed count	Estimated expected count	Adjusted residual
1972	Yes	Yes	Yes	334	341.77	−1.36
			No	34	26.76	1.72
		No	Yes	12	11.93	0.03
			No	15	14.54	0.14
	No	Yes	Yes	53	56.12	−0.54
			No	63	62.63	0.06
		No	Yes	43	40.19	0.57
			No	501	501.06	−0.01
1973	Yes	Yes	Yes	428	421.37	1.10
			No	29	32.99	−0.90
		No	Yes	13	14.70	−0.56
			No	17	17.93	−0.28
	No	Yes	Yes	42	49.23	−1.30
			No	53	54.95	−0.33
		No	Yes	31	35.25	−0.90
			No	453	439.57	1.58
1974	Yes	Yes	Yes	413	411.86	0.19
			No	29	32.25	−0.74
		No	Yes	16	14.37	0.54
			No	18	17.53	0.14
	No	Yes	Yes	60	49.65	1.86
			No	57	55.42	0.27
		No	Yes	37	35.56	0.30
			No	430	443.37	−1.57

to slow convergence, but the scoring algorithm is relatively successful in his examples.

A model which is more successful assumes that $W_h = B_h$, so that $p^{WXBDF \cdot Y}_{wxijk \cdot l}$ is only positive if $w = i$. If n_{xijkl} is again used to denote the number of observations such that $X_h = x$, $B_h = i$, $D_h = j$, $F_h = k$, and $Y_h = l$, then the mean m_{xijkl} of n_{xijkl} satisfies the log-linear model

$$\begin{aligned}\log m_{xijkl} &= \lambda + \lambda^X_x + \lambda^B_i + \lambda^D_j + \lambda^F_k + \lambda^{XB}_{xi} + \lambda^{XD}_{xj} + \lambda^{XF}_{xk} + \lambda^{BY}_{il} \\ &= \alpha^Y_l + \lambda^X_1 q^X_x + \lambda^B_1 q^B_i + \lambda^D_1 q^D_j + \lambda^F_1 q^F_k + \lambda^{XB}_{11} q^X_x q^B_i \\ &\quad + \lambda^{XD}_{11} q^X_x q^D_j + \lambda^{XF}_{11} q^X_x q^F_k + \lambda^{BY}_{11} q^B_i q^Y_{l1} + \lambda^{BY}_{12} q^b_i q^Y_{l2}.\end{aligned}$$

This model differs from the traditional model only in substitution of λ^{BY}_{il} for λ^{XY}_{xl}. Thus it will be called an asymmetric model. One obtains results shown in Tables 10.7 and 10.8. The values of $X^2 = 8.60$ and $L^2 = 8.42$ are a little smaller than those for the simple latent-class model, and both models have 12 degrees of freedom. Coefficient estimates are quite similar, provided λ^{BY}_{il} is regarded as comparable to λ^{XY}_{il}. Furthermore, the adjusted residuals in Table 10.8 are never greater in magnitude than 1.86, an improvement over Table 10.2.

Table 10.8

Parameter Estimates for the Asymmetric Model for Table 7.29

Parameter	Estimate	EASD
λ^X_1	−0.103	0.086
λ^B_1	−0.318	0.045
λ^D_1	0.327	0.049
λ^F_1	0.011	0.039
λ^{XB}_{11}	1.372	0.045
λ^{XD}_{11}	1.397	0.049
λ^{XF}_{11}	1.293	0.039
λ^{BY}_{11}	−0.108	0.026
λ^{BY}_{12}	0.062	0.025

Model Selection

As is apparent from results of this chapter and Chapter 7, a variety of log-linear and latent-class models fit Table 7.29 about equally well. As is often the case in the analysis of observational data, many models, not all of which are mutually consistent, can be used to approximate the observed joint distribution of variables. One should not assume when a plausible model fits the data that it is the only model that fits.

10.6 General Latent-Class Models

In general, any latent-class model can be described in terms of a three-dimensional array representing a cross-classification of N polytomous variables (X_h, A_h, B_h), $1 \le h \le N$. For example, in most cases in this chapter X_h is dichotomous, A_h in this representation corresponds to the three responses (B_h, D_h, F_h), and B_h corresponds to year Y_h of survey. Given the B_h, the pairs (X_h, A_h), $1 \le h \le N$, are assumed independently distributed with probability $p_{xi \cdot j}^{XA \cdot B} > 0$ that $X_h = x$ and $A_h = i$ given that $B_h = j$, where $1 \le x \le q$, $1 \le i \le r$, and $1 \le j \le s$. If n_{xij} is the number of observations h with $X_h = x$, $A_h = i$, and $B_h = j$, then the conditional mean m_{xij} of n_{xij} given the B_h is $n_j^B p_{xi \cdot j}^{XA \cdot B}$. It is assumed that the marginal table n_{ij}^{AB} is observed, and that

$$\log(m_{xij}/z_{xij}) = \alpha_j + \sum_{k=1}^{p} \beta_k u_{xijk}, \qquad z_{xij} > 0$$

for some unknown α_j and β_k, some known $z_{xij} > 0$, and some known u_{xijk}. The maximum likelihood estimate m_{xij} satisfies the equations

$$\hat{m}_j^B = n_j^B,$$

$$\log \hat{m}_{xij} = \hat{\alpha}_j + \sum_{k=1}^{p} \hat{\beta}_k u_{xijk},$$

and

$$\sum_x \sum_i \sum_j u_{xijk} \hat{m}_{xij} = \sum_x \sum_i \sum_j u_{xijk} \hat{n}_{xij},$$

where

$$\hat{n}_{xij} = n_{ij}^{AB} \hat{m}_{xij}/\hat{m}_{ij}^{AB}.$$

Unlike regular log-linear models, these equations need not uniquely specify the $\hat{m}_{xij}$ when they exist and may even lead to solutions $\hat{m}_{xij}$ that are not maximum likelihood estimates. In practice, it appears unusual to obtain a solution $\hat{m}_{xij}$ such that $\hat{m}_{xij}$ is not a maximum likelihood estimate but $\hat{S}$ has an inverse, where

$$\hat{S}_{kl} = \sum_i \sum_j (\hat{u}_{ijk} - \hat{\theta}_{jk})(\hat{u}_{ijl} - \hat{\theta}_{jl}) \hat{m}_{ij}^{AB},$$

$$\hat{\theta}_{jk} = \sum_i \hat{u}_{ijk} \hat{m}_{ij}^{AB}/n_j^B,$$

and

$$\hat{u}_{ijk} = \sum_x u_{xijk} \hat{m}_{xij}/m_{ij}^{AB}.$$

Chi-Square Tests

If S has an inverse, where

$$S_{kl} = \sum_i \sum_j (\bar{u}_{ijk} - \theta_{jk})(\bar{u}_{ijl} - \theta_{jl}) m_{ij}^{AB},$$

$$\theta_{jk} = \sum_i \bar{u}_{ijk} m_{ij}^{AB} / n_j^B,$$

and

$$\bar{u}_{ijk} = \sum_x u_{xijk} m_{xij} / m_{ij}^{AB},$$

then one may test the latent-class model by means of the chi-square statistics

$$X^2 = \sum_i \sum_j (n_{ij}^{AB} - \hat{m}_{ij}^{AB})^2 / \hat{m}_{ij}^{AB}$$

and

$$L^2 = 2 \sum_i \sum_j n_{ij}^{AB} \log(n_{ij}^{AB} / \hat{m}_{ij}^{AB}).$$

These statistics have approximate chi-square distributions with $(r - 1)s - p - k$ degrees of freedom if the model holds, if the n_j^B are all large, and if k of the m_{ij}^{AB} are 0.

Large-sample properties of the β_k are complicated often by lack of uniqueness of the estimates. For example, in the model of Section 10.1, if $\hat{\lambda}_1^X$ is a maximum likelihood estimate of λ_1^X, then $-\hat{\lambda}_1^X$ is also a possible maximum likelihood estimate of λ_1^X. Thus some convention must generally be adopted to ensure uniqueness. For example, in Section 10.1, all maximum likelihood estimates are uniquely determined subject to the condition that $\hat{\lambda}_{11}^{XB} > 0$. Given that S is invertible, it is possible to define the $\hat{\beta}_k$, $1 \leq k \leq p$, so that $\hat{\boldsymbol{\beta}}$ has an approximate multivariate normal distribution with asymptotic mean $\boldsymbol{\beta}$ and asymptotic covariance matrix S^{-1}. The asymptotic covariance matrix can be estimated by $\hat{S}^{-1}$.

The adjusted residual $(n_{ij}^{AB} - \hat{m}_{ij}^{AB})/\hat{c}_{ij}^{1/2}$, where

$$\hat{c}_{ij} = \hat{m}_{ij}^{AB}[1 - \hat{m}_{ij}^{AB}/n_j^B - \hat{m}_{ij}^{AB} \sum_k \sum_l (\hat{u}_{ijk} - \hat{\theta}_{jk})(\hat{u}_{ijl} - \hat{\theta}_{jl})\hat{S}^{kl}],$$

has an approximate standard normal distribution if the model holds and the n_j^B are large. More generally,

$$\sum_i \sum_j d_{ij}(n_{ij}^{AB} - \hat{m}_{ij}^{AB})/\hat{c}^{1/2},$$

has an approximate standard normal distribution, where

$$\hat{c} = \sum_i \sum_j (d_{ij} - \hat{d}_j)^2 \hat{m}_{ij}^{AB} - \sum_k \sum_l \hat{f}_k \hat{f}_l \hat{S}^{kl},$$

$$\hat{d}_j = \sum_i d_{ij} \hat{m}_{ij}^{AB} / n_j^B,$$

and

$$\hat{f}_k = \sum_i \sum_j d_{ij} (\hat{u}_{ijk} - \hat{\theta}_{jk}) \hat{m}_{ij}^{AB}.$$

The scoring algorithm for computation of maximum likelihood estimates can always be implemented by beginning with an initial approximation $m_{xij0} > 0$ to $\hat{m}_{xij}$. Given m_{xij0}, an approximation β_{k0} is defined so that

$$\sum_l S_{kl0} \beta_{l0} = w_{k0},$$

where

$$S_{kl0} = \sum_i \sum_j (\bar{u}_{ijk0} - \theta_{jk0})(\bar{u}_{ijl0} - \theta_{jl0}) m_{ij0}^{AB},$$

$$w_{k0} = \sum_i \sum_j (\bar{u}_{ijk0} - \theta_{jk0}) \bar{Z}_{ij0} m_{ij0}^{AB},$$

$$\theta_{jk0} = \sum_i \bar{u}_{ijk0} m_{ij0}^{AB} / m_{j0}^B,$$

$$\bar{u}_{ijk0} = \sum_x u_{xijk0} m_{xij0} / m_{ij0}^{AB},$$

and

$$\bar{Z}_{ij0} = \sum_x (\log m_{xij0} - \log z_{xij}) m_{xij0} / m_{ij0}^{AB}.$$

The usual conventions $0 \log 0 = 0$ and $0/0 = 0$ are used in these formulas. Then one continues iterations by letting

$$v_{xij0} = \sum_k \beta_{k0} u_{xijk}, \qquad g_{j0} = n_j^B \Big/ \sum_x \sum_i z_{xij} \exp v_{xij0},$$

$$m_{xij1} = z_{xij} g_{j0} \exp v_{xij0}.$$

For $t \geq 0$,

$$\sum_l S_{klt} \delta_{lt} = a_{kt} = \sum_i \sum_j (n_{ij}^{AB} - m_{ijt}^{AB}) \bar{u}_{ijkt},$$

where

$$\bar{u}_{ijkt} = \sum_x u_{xijk} m_{xijt} / m_{ijt}^{AB}, \qquad \theta_{jkt} = \sum_i \bar{u}_{ijkt} m_{ijt}^{AB} / m_{jt}^B,$$

and

$$S_{klt} = \sum_i \sum_j (\hat{u}_{ijkt} - \theta_{ijkt})(\bar{u}_{ijlt} - \theta_{ijlt}) m_{ijt}^{AB}.$$

Then

$$\beta_{kt} = \beta_{k(t-1)} + \delta_{kt}, \qquad v_{xijt} = \sum_k \beta_{kt} u_{xijk},$$

$$g_{jt} = n_j^B \Big/ \sum_x \sum_i z_{xij} \exp v_{xijt},$$

and

$$m_{xij(t+1)} = z_{xij} g_{jt} \exp v_{xijt}.$$

The scoring algorithm generally leads to relatively rapid convergence whenever $\hat{S}$ has an inverse. Complications can arise if m_{xij0} is sufficiently badly chosen or if some estimated asymptotic variances $s^2(\hat{\beta}_k)$ are very large.

For further details concerning results in this section, see Haberman (1974c, 1977) and Sundberg (1972, 1974).

EXERCISES

10.1 Derive Table 10.1 using the algorithm presented. Show that

$$m_{xijkl4} = n_{xi0}^{XB} n_{xj0}^{XD} n_{xk0}^{XF} n_{xl0}^{XT} / (n_{x0}^{X})^3.$$

Solution

Note that m_{xijkl0} can be written as a product $a_{xi} b_{xj} c_{xk} d_{xl}$. Thus

$$m_{xi0}^{XB} = a_{xi}\left(\sum_j b_{xj}\right)\left(\sum_k c_{xk}\right)\left(\sum_l d_{xl}\right)$$

and

$$m_{xijkl1} = b_{xj1} c_{xk1} d_{xl1} n_{xi0}^{XB},$$

for some b_{xj1}, c_{xk1}, and d_{xl2}. Similarly,

$$m_{xijl2} = c_{xk1} d_{xl1} n_{xi0}^{XB} n_{xj0}^{XD} / n_{x0}^{X},$$
$$m_{xijkl3} = d_{xl1} n_{xi0}^{XB} n_{xj0}^{XD} n_{xk0}^{XF} / (n_{x0}^{X})^2,$$

and

$$m_{xijkl4} = n_{xi0}^{XB} n_{xj0}^{XD} n_{xk0}^{XF} n_{xl0}^{XY} / (n_{x0}^{X})^3.$$

This last formula is more similar to those of Goodman (1974a).

Table 10.9

Responses of White Christian Subjects in the 1972 to 1974 General Social Surveys to Three Questions Concerning Therapeutic Abortions[a,b]

Year	Response to A	Response to C	Response to E	Count
1972	Yes	Yes	Yes	806
			No	55
		No	Yes	7
			No	11
	No	Yes	Yes	45
			No	43
		No	Yes	17
			No	82
1973	Yes	Yes	Yes	859
			No	72
		No	Yes	5
			No	4
	No	Yes	Yes	49
			No	38
		No	Yes	11
			No	50
1974	Yes	Yes	Yes	869
			No	52
		No	Yes	13
			No	3
	No	Yes	Yes	47
			No	34
		No	Yes	11
			No	58

[a] Data tapes from the 1972 to 1974 General Social Surveys, National Opinion Research Center, University of Chicago.

[b] For a list of questions, see Table 4.8 in Volume 1. Only subjects cross-classified in that Table are included here.

10.2 Verify Tables 10.2 and 10.3.

10.3 Apply the model of Section 10.1 to Table 10.9

Solution

Results are summarized in Table 10.10 and Table 10.11. Note that the fit is quite satisfactory, with $X^2 = 12.5$, $L^2 = 12.3$, and 12 degrees of freedom. In addition, there is only one adjusted residual greater in magnitude than 2. The interactions between X_h and the responses A_h, C_h, and E_h are comparable to the corresponding interactions in Table 7.29 between X_h and B_h, D_h and F_h and the parameter $\hat{\lambda}_1^X$ is not much changed. However, $\hat{\lambda}_1^C$, which involves

Table 10.10

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Simple Latent-Class Model for Table 10.9

Year	Response to *A*	Response to *C*	Response to *E*	Observed count	Estimated expected count	Adjusted residual
1972	Yes	Yes	Yes	806	800.93	0.69
			No	55	57.62	−0.43
		No	Yes	7	8.23	−0.52
			No	11	7.47	1.76
	No	Yes	Yes	45	46.96	−0.35
			No	43	47.73	−1.08
		No	Yes	17	16.34	0.23
			No	82	80.72	0.28
1973	Yes	Yes	Yes	859	868.57	−1.27
			No	72	60.86	1.82
		No	Yes	5	8.35	−1.45
			No	4	5.18	−0.62
	No	Yes	Yes	49	47.20	0.33
			No	38	32.92	1.23
		No	Yes	11	11.04	−0.02
			No	50	53.88	−0.93
1974	Yes	Yes	Yes	869	864.43	0.61
			No	52	60.67	−1.42
		No	Yes	13	8.35	2.00
			No	3	5.33	−1.22
	No	Yes	Yes	47	47.20	−0.04
			No	34	33.92	0.02
		No	Yes	11	11.40	−0.15
			No	58	55.69	0.55

Table 10.11

Parameter Estimates for the Simple Latent-Class Model for Table 10.9

Parameter	Estimate	EASD
λ_1^X	−0.165	0.130
λ_1^A	0.151	0.074
λ_1^C	1.045	0.069
λ_1^E	0.271	0.050
λ_{11}^{XA}	1.373	0.072
λ_{11}^{XC}	1.335	0.067
λ_{11}^{XE}	1.080	0.049
λ_{11}^{XY}	−0.156	0.039
λ_{12}^{XY}	0.087	0.043

abortions for reasons of serious danger to the health of the mother, is much larger than the corresponding estimated parameters for questions A and E in Table 10.9 or questions B, D, and F in Table 7.29.

10.4 Apply the model of constant interaction to Table 10.9.

Solution

Results are summarized in Tables 10.12 and 10.13. The chi-square statistics are $X^2 = 27.0$ and $L^2 = 27.1$, and there are 14 degrees of freedom, so the fit is not very satisfactory. Note that the decrease in L^2 from the model of the preceding exercise is 14.8 on 2 degrees of freedom, which is significant at the 0.001 level. The problem appears to involve question E, which deals

Table 10.12

Observed Counts, Estimated Expected Counts, and Adjusted Residuals for the Model of Constant Interaction for Table 10.9

Year	Response to A	Response to C	Response to E	Observed count	Estimated expected count	Adjusted residual
1972	Yes	Yes	Yes	806	801.10	0.66
			No	55	53.43	0.26
		No	Yes	7	10.42	−1.16
			No	11	10.88	0.04
	No	Yes	Yes	45	48.65	−0.63
			No	43	50.80	−1.63
		No	Yes	17	9.90	2.52
			No	82	80.81	0.26
1973	Yes	Yes	Yes	859	868.04	−1.19
			No	72	55.35	2.77
		No	Yes	5	10.79	−1.94
			No	4	7.50	−1.40
	No	Yes	Yes	49	50.40	−0.24
			No	38	35.03	0.69
		No	Yes	11	6.83	1.74
			No	50	54.06	−0.98
1974	Yes	Yes	Yes	869	864.79	0.56
			No	52	55.26	−0.54
		No	Yes	13	10.77	0.75
			No	3	7.67	−1.85
		No	Yes	47	50.32	−0.58
			No	34	35.82	−0.42
		No	Yes	11	6.98	1.66
			No	58	55.39	0.62

Table 10.13

Parameter Estimates for the Model of Constant Interaction for Table 10.9

Parameter	Estimate	EASD
λ_1^X	−0.071	0.096
λ_1^A	0.217	0.053
λ_1^C	0.987	0.051
λ_1^E	0.170	0.052
$\lambda_{11}^{XA} = \lambda_{11}^{XC} = \lambda_{11}^{XE}$	1.245	0.033
λ_{11}^{XY}	−0.156	0.039
λ_{12}^{XY}	0.085	0.043

with rape. Note that the estimated interaction $\hat{\lambda}_{11}^{XE}$ in Table 10.11 is somewhat smaller than $\hat{\lambda}_{11}^{XA}$ or $\hat{\lambda}_{11}^{XC}$.

10.5 Determine the estimated probabilities $p_{x \cdot l}^{X;Y}$ for the three-class latent-variable model of Table 7.29.

Solution

Note that

$$\hat{m}_{2l}^{XY} - \hat{m}_{2111l} - \hat{m}_{2222l} = n_l^Y - n_{111l}^{BDFY} - n_{222l}^{BDFY},$$
$$\hat{m}_{2111l} = \hat{m}_{211l}^{BDFY}\hat{m}_{121l}^{BDFY}/\hat{m}_{221l}^{BDFY},$$

and

$$\hat{m}_{2222l} = \hat{m}_{221l}^{BDFY}\hat{m}_{212l}^{BDFY}/\hat{m}_{211l}^{BDFY}.$$

Table 10.14

Estimated Latent-Class Probabilities for Table 7.29

Year	Class	Probability
1972	1	0.292
	2	0.272
	3	0.436
1973	1	0.381
	2	0.226
	3	0.393
1974	1	0.366
	2	0.267
	3	0.368

Thus

$$\hat{m}_{2l}^{XY} = n_l^Y + \hat{m}_{2111l} + \hat{m}_{2222l} - n_{111l}^{BDFY} - n_{222l}^{BDFY},$$

$$\hat{m}_{1l}^{XY} = n_{111l}^{BDFY} - \hat{m}_{2111l}, \qquad \hat{m}_{3l}^{XY} = n_{222l}^{BDFY} - \hat{m}_{3222l}, \qquad \hat{p}_{x\cdot l}^{X\cdot Y} = \hat{m}_{xl}^{XY}/n_l^Y.$$

Results are summarized in Table 10.14.

10.6 Use an iterative proportional fitting algorithm to derive estimates $\hat{m}_{xijkl}$ for Table 7.29 under the asymmetric model.

Solution

Results are summarized in Table 10.15. One possible algorithm has the form

$$n_{xijkl0} = m_{xijkl0} n_{ijkl}^{BDFY}/m_{ijkl0}^{BDFY}, \qquad m_{xijkl1} = m_{xijkl0} n_{xi0}^{XB}/m_{xi0}^{XB},$$

$$m_{xijkl2} = m_{xijkl1} n_{xj0}^{XD}/m_{xj1}^{XD}, \qquad m_{xijkl3} = m_{xijkl2} n_{xk0}^{XF}/m_{xk2}^{XF},$$

$$m_{xijkl4} = m_{xijkl3} n_{jl0}^{BY}/m_{jl3}^{BY}, \qquad n_{xijkl4} = m_{xijkl4} n_{ijkl}^{BDFY}/m_{ijkl4}^{BDFY}, \qquad \text{etc.}$$

Table 10.15

Estimated Means for Table 7.29 under the Asymmetric Latent-Variable Model

l	i	j	k	$\hat{m}_{1ijkl}$	$\hat{m}_{2ijkl}$
1	1	1	1	341.65	0.12
			2	25.15	1.61
		2	1	10.87	1.06
			2	0.80	13.74
	2	1	1	51.59	4.53
			2	3.80	58.84
		2	1	1.64	38.55
			2	0.12	500.94
2	1	1	1	421.22	0.15
			2	31.00	1.99
		2	1	13.40	1.30
			2	0.99	16.94
	2	1	1	45.26	3.97
			2	3.33	51.62
		2	1	1.44	33.81
			2	0.11	439.46
3	1	1	1	411.71	0.15
			2	30.30	1.95
		2	1	13.10	1.27
			2	0.96	16.56
	2	1	1	45.65	4.01
			2	3.36	52.06
		2	1	1.45	34.11
			2	0.11	443.26

Appendix Computer Programs for Computation of Maximum Likelihood Estimates

The program FREQ presented in this appendix can be used to compute maximum likelihood estimates for any log–linear model. A modest modification of FREQ called LAT may be used for the latent-class models of Chapter 10. This Appendix provides program descriptions, listings, and a few examples of input and output. Both programs are written in ASA Standard Fortran. Output is arranged to facilitate terminal use.

A.1 FREQ

FREQ is a program for analysis of log–linear models. Observations consist of counts n_{ij}, $1 \le i \le r$, $1 \le j \le s$. For known constants $z_{ij} \ge 0$, $1 \le i \le r$, $1 \le j \le s$ and x_{ijk}, $1 \le i \le r$, $1 \le j \le s$, $1 \le k \le p$, and unknown parameters α_j, $1 \le j \le s$, and β_k, $1 \le k \le p$, it is assumed that the expected values m_{ij} of n_{ij} satisfy the log–linear model

$$\log(m_{ij}/z_{ij}) = \alpha_j + \sum_{k=1}^{p} \beta_k x_{ijk}, \quad z_{ij} > 0.$$

It is also assumed that conditional on the observed counts

$$n_j^B = \sum_{i=1}^{r} n_{ij},$$

the subtables $\{n_{ij}: 1 \le i \le r\}$ have independent multinomial distributions for $1 \le j \le s$.

The program uses the method of maximum likelihood to obtain estimates $\hat{\beta}_k$ of the parameters β_k and estimated asymptotic standard deviations $s(\hat{\beta}_k)$

of the $\hat{\beta}_k$. Likelihood-ratio and Pearson chi-square statistics are computed, and residual analyses are performed.

Control Cards

Cards must appear in the following order.

Card 1. *Run description* This card contains 9 entries in the following format:

Column	Format	Entry
1-4	I	The number r of rows. If the entry is not positive, $r = 1$.
5-8	I	The number s of columns. If the entry is not positive, $s = 1$.
9-12	I	The number p of independent variables x_{ijk}, $1 \leq k \leq p$, for each row i and column j. If the entry is not positive, $p = 0$.
13-16	I	If positive, the z_{ij} are supplied by the user. If not positive, each $z_{ij} = 1$.
17-20	I	If positive, initial approximations for the $\hat{\beta}_k$, $1 \leq k \leq p$, are supplied by the user. Otherwise, the program generates its own initial value.
21-24	I	If positive, observed values n_{ij}, estimated expected values $\hat{m}_{ij}$, and adjusted residuals $(n_{ij} - \hat{m}_{ij})/s(n_{ij} - \hat{m}_{ij})$ are printed, where $s(n_{ij} - \hat{m}_{ij})$ is the estimated asymptotic standard deviation of $n_{ij} - \hat{m}_{ij}$.
25-28	I	The number of generalized contrasts to be computed. For each contrast array d_{ij}, $1 \leq i \leq r$, $1 \leq j \leq s$, $$\text{O} = \sum\sum d_{ij}n_{ij}, \quad E = \sum\sum d_{ij}\hat{m}_{ij},$$ and $(\text{O} - E)/s(\text{O} - E)$ is computed, where $s(\text{O} - E)$ is the estimated asymptotic standard deviation of $\text{O} - E$.
29-32	I	The maximum number of iterations used in computing the $\hat{\beta}_k$. If blank, 10 is the maximum.
33-36	I	The stopping criterion g. The tolerance is 10^{-g} if $g \neq 0$ and 10^{-3} if $g = 0$. If no successive approximations for $\sum \hat{\beta}_k x_{ijk}$, $1 \leq i \leq r$, $1 \leq j \leq s$, differ in magnitude by more than the tolerance, then iterations stop.

Card 2. *Format of data* This card is a variable format card describing the format in which the data are to be read. The card is a FORMAT statement in which the work FORMAT has been removed and in which the initial parenthesis is in the first column. The data should be read in F format.

Card 3. *Title* Up to 80 columns of alphanumeric title information.

Label Cards

Each 8 columns of each card contains the name of a variable. Thus columns 1–8 of the first label card contains the same of the first variable x_{ij1} columns

9–16 contain the name of the second variable x_{ij2}, etc. Note that if $p \leq 10$, one label card is needed, if $11 \leq p \leq 20$, two label cards are needed, etc.

Data Cards

The data are read in standard FORTRAN order according to the format of card 2. For example, in a 2×3 table with $n_{11} = 10$, $n_{21} = 11$, $n_{12} = 20$, $n_{22} = 25$, $n_{13} = 37$, $n_{23} = 31$, one might have a single data card

101120253731

The corresponding Card 2 is

(6F2.0)

Z Cards

If the entry in columns 13–16 of Card 1 is positive, then the coefficients z_{ij} are read in the same format used for the data. Observations are read in standard FORTRAN order so that if $r = 2$, $c = 3$, $z_{ij} = j$, and if Card 2 is

(6F2.0)

then one Z card is required. The card is

010102020303

X Cards

For each k, $1 \leq k \leq p$, a card or series of cards lists the x_{ijk}, $1 \leq i \leq r$, $1 \leq j \leq s$, in the same format used for the data. If $p = 2$,

$$\begin{aligned} x_{ij1} &= 1, & i &= 1, \\ &= 0, & i &= 2, \\ x_{ij2} &= j, & i &= 1, \\ &= 0, & i &= 2, \end{aligned}$$

and if Card 2 is

(6F2.0)

then the two X cards are

010001000100
010002000300

Initial Valve Cards

If the entry in columns 17–20 of card 1 is positive, then initial approximations of the $\hat{\beta}_k$ are read in F10.5 format. The approximation for $\hat{\beta}_1$ is in columns 1–10 of the first initial value card, the approximation for $\hat{\beta}_2$ is in columns 11–20 of this card, etc. If $p \leq 8$, one initial value card is needed. If $9 \leq p \leq 16$, two such cards are needed, etc.

Contrast Label Cards

For each contrast requested in columns 25–28 of card 1, a contrast label card and contrast specification cards are required. The contrast label card contains a contrast label in columns 1–8.

The contrast specification cards specify a contrast array

$$d_{ij}, \qquad 1 \leq i \leq r, \qquad 1 \leq j \leq s,$$

using the same format used for the data cards. If card 2 is

(6F2.0)

and if

$$\begin{aligned} d_{ij} &= 1, & i &= 1, \; j = 1 \quad \text{or} \quad 3, \\ &= 0, & i &= 2, \\ &= -2, & i &= 1, \; j = 2, \end{aligned}$$

then a contrast label name and contrast name card might be

QUADRAT
0100-2000100

Example 1

Consider the log-linear time trend model of Chapter 1. The expected number m_{i1} of events n_{i1} reported in month i satisfies a model of the form

$$\log m_{i1} = \alpha + \beta i.$$

The following control cards may be used with this model. Note that r is 18, while s and p may be taken to be 1. Note that x_{i11} is i and that the three contrasts requested involve the sums

$$\sum_{i=1}^{6} n_{i1}, \qquad \sum_{i=7}^{12} n_{i1}, \qquad \text{and} \qquad \sum_{i=13}^{18} n_{i1},$$

where n_{i1} is the number of events reported in month i.

Following the control cards, some output from a terminal is included.

Control Cards

```
 18  1   1   0      1   3
(18F2.0)
STRESS EVENTS RECALLED
DECAY
495542433535423137213540292229121515
 1 2 3 4 5 6 7 8 9101112131415161718
EARLY
 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1
MIDDLE
 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0
LATE
 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0
```

Output

```
STRESS EVENTS RECALLED

COEFFICIENTS AND STANDARD ERRORS

  DECAY
  -0.06117
   0.00820

 LIKELIHOOD RATIO CHI SQUARE =   23.68111
          PEARSON CHI SQUARE =   23.79869
NUMBER OF DEGREES OF FREEDOM =   16

    I     J  OBSERVED COUNT  EXPECTED COUNT  ADJUSTED RESIDUAL
    1     1          49.000          52.183             -0.511
    2     1          55.000          49.087              0.942
    3     1          42.000          46.174             -0.667
    4     1          43.000          43.434             -0.070
    5     1          35.000          40.857             -0.962
    6     1          35.000          38.432             -0.576
    7     1          42.000          36.152              1.005
    8     1          31.000          34.006             -0.531
    9     1          37.000          31.989              0.913
   10     1          21.000          30.090             -1.710
   11     1          35.000          28.305              1.303
   12     1          40.000          26.625              2.696
   13     1          29.000          25.045              0.827
   14     1          22.000          23.559             -0.338
   15     1          29.000          22.161              1.543
   16     1          12.000          20.846             -2.074
   17     1          15.000          19.609             -1.124
   18     1          15.000          18.445             -0.874

  CONTRAST   OBSERVED VALUE  EXPECTED VALUE  ADJUSTED RESIDUAL
  EARLY             122.000         129.666             -1.235
  MIDDLE            206.000         187.167              1.701
  LATE              259.000         270.167             -1.690
```

Example 2

In the fixed-distance model for Table 8.1, the following control cards and output may be used if diagonal cells are ignored.

Control Cards

```
  4  4  4  1  0  1
(16F3.0)
FIXED DISTANCE MODEL FOR TABLE 8.1, NO DIAGONALS
A1       A2       A3       ETA
    82  5 2123    59 41  2 30    29  0  7  4
  0  1  1  1  1  0  1  1  1  1  0  1  1  1  1  0
  1  0  0 -1  1  0  0 -1  1  0  0 -1  1  0  0 -1
  0  1  0 -1  0  1  0 -1  0  1  0 -1  0  1  0 -1
  0  0  1 -1  0  0  1 -1  0  0  1 -1  0  0  1 -1
  6 -1 -2 -3 -1  4 -1 -2 -2 -1  4 -1 -3 -2 -1  6
```

Output

```
FIXED DISTANCE MODEL FOR TABLE 8.1, NO DIAGONALS

COEFFICIENTS AND STANDARD ERRORS

   A1         A2         A3          ETA
 -0.16676   0.59926   -0.89667   1.68884
  0.12358   0.16048    0.12991   0.20413

 LIKELIHOOD RATIO CHI SQUARE =    3.47776
          PEARSON CHI SQUARE =    3.19373
NUMBER OF DEGREES OF FREEDOM =    4

   I    J  OBSERVED COUNT  EXPECTED COUNT  ADJUSTED RESIDUAL
   1    1           0.0             0.0                 0.0
   2    1          82.000          83.084              -0.805
   3    1           5.000           3.439               1.110
   4    1           2.000           2.477              -0.368
   1    2         123.000         121.916               0.805
   2    2           0.0             0.0                 0.0
   3    2          59.000          58.758               0.123
   4    2          41.000          42.327              -0.701
   1    3           2.000           2.673              -0.516
   2    3          30.000          31.130              -0.604
   3    3           0.0             0.0                 0.0
   4    3          29.000          27.196               1.236
   1    4           0.0             0.411              -0.672
   2    4           7.000           4.785               1.505
   3    4           4.000           5.804              -1.236
   4    4           0.0             0.0                 0.0
```

Error Conditions

An error is reported and execution terminates if some z_{ij} is negative or if space requirements are too large. Space may be adjusted by changing the region size of the array work in the main program and giving the constant NW the same value. The space in words needed in WORK is $(3 + p)rs + sp + 4p + p^2$ if all z_{ij} are assumed 1. Otherwise, rs more words is needed. The appropriate changes are made in lines 31 and 33 of the listing. The current value of NW is 10,000.

If the independent variables x_{ijk}, $1 \leq k \leq p$, are linearly dependent, the program ignores the redundant variables. The output will contain estimates and standard errors of 0 for the corresponding coefficients.

Program Listing

The following listing includes a main program, together with subroutines CONTIN, CHOL, SOLVE, INVERT, CHISQ, RESID, and GENRES.

```
C
C          THIS PROGRAM PERFORMS MAXIMUM-LIKELIHOOD ESTIMATION
C          FOR THE LOG-LINEAR MODEL
C          LOG(M(I,J)/Z(I,J)) = A(J)+U(I,J), WHERE
C          U(I,J) = B(1)*X(I,J,1)+...+B(NV)*X(I,J,NV).
C          HERE A TABLE N(I,J) IS GIVEN WITH NR ROWS AND
C          NC COLUMNS.  THE EXPECTED VALUE OF N(I,J) IS M(I,J).
C          NORMALLY Z(I,J) IS POSITIVE.  IF Z(I,J) IS NOT POSITIVE,
C          THEN CELL N(I,J) IS IGNORED DURING ANALYSIS.  THIS
C          PROVISION IS USEFUL IN ANALYSIS OF INCOMPLETE
C          TABLES.
```

```
C
C        FOR A DESCRIPTION OF CONTROL CARDS, SEE THE PROGRAM
C        DESCRIPTION.
C
C        IN THIS PROGRAM, I1 IS 0 IF Z(I,J) IS 1 AND POSITIVE
C        IF ANY OTHER CHOICE OF Z(I,J) IS TO BE MADE.
C        THE PARAMETER IS EQUALS 0 IF INITIAL APPROXIMATIONS
C        B(K,0), 1<=K<=NV, ARE TO BE FOUND BY THE PROGRAM.
C        IF THE USER WISHES TO SUPPLY INITIAL
C        ESTIMATES, THEN IS SHOULD BE POSITIVE.
C        ADJUSTED RESIDUALS ARE SUPPLIED IF IR IS POSITIVE.
C        GENERALIZED RESIDUALS ARE SUPPLIED IF IC IS POSITIVE.
C        MAX ITERATIONS MAY BE PERFORMED IF NEEDED.  IF
C        THE USER DEFINES MX TO BE POSITIVE, THEN MAX IS MX.
C        IF MX IS NOT POSITIVE, MAX IS 10.
C        ITERATIONS STOP IF NO CHANGE U(I,J,T)-U(I,J,T-1)
C        EXCEEDS 10**IT IN MAGNITUDE.  NORMALLY IT IS -3,
C        BUT THE USER MAY CHANGE THIS CHOICE.
C
      REAL WORK(10000),FMT(20),TITLE(20)
      LOGICAL ONE,INIT
      DATA MXS,TOLS,NW/10,0.001,10000/
C
C        READ PARAMETERS.
C
      READ(5,1) NR,NC,NV,I1,IS,IR,IC,MX,IT
    1 FORMAT(9I4)
      IF(NR.LE.0) NR = 1
      IF(NC.LE.0) NC = 1
      ONE = .TRUE.
      INIT = .TRUE.
      IF(I1.GT.0) ONE = .FALSE.
      IF(IS.GT.0) INIT = .FALSE.
      IF(MX.LE.0) MX = MXS
      IF(IT.NE.0) TOL = 10.0**IT
      IF(IT.EQ.0) TOL = TOLS
C
C        READ A VARIABLE FORMAT FOR THE DATA.
C
      READ(5,2) FMT
C
C        READ A TITLE FOR THE RUN.  NOTE THAT UP TO 80 CHARACTERS
C        ARE AVAILABLE.
C
      READ(5,2) TITLE
    2 FORMAT(20A4)
      WRITE(6,11) TITLE
   11 FORMAT(1H1,20A4)
C
C        TO SET UP WORK AREAS, FIRST FIND THE SIZE OF THE TABLE.
C
      NSIZE = NR*NC
C
C        NF IS THE LOCATION IN THE WORK AREA OF THE FITTED
C        TABLE M OF EXPECTED VALUES.
C        NZ IS THE LOCATION OF THE TABLE Z IF I1>0.
C
      NF = NSIZE+1
      IF(ONE) NZ = NF
      IF(.NOT.ONE) NZ = NF+NSIZE
C
C        NX IS THE LOCATION OF THE TABLE X.
C        NT IS THE LOCATION OF THE TABLE THETA OF AVERAGES.
C
      NX = NZ+NSIZE
      NX1 = NX-1
      NT = NX+NSIZE*NV
C
C        PROVIDE SPACE FOR COVARIANCE MATRIX AT NS.
C
      NS = NT+NC*NV
C
C        PROVIDE SPACE FOR VARIABLE NAMES AT NL.
C
      NL = NS+NV*NV
```

```
C
C          PLACE RESIDUALS AT NR.
C
      NRS = NL+NV+NV
      NR2 = NRS-1
      NEN = NR2+NSIZE
C
C          PLACE PARAMETER ESTIMATES AT NB.
C
      NB = NRS+NSIZE
      ND = NB+NV
      ND1 = ND-1
      NE = ND+NV-1
C
C          CHECK FOR EXCESSIVE SPACE DEMANDS.
C
      IF(NE.GT.NW) WRITE(6,3)
    3 FORMAT(38HOTOO MUCH SPACE REQUESTED FOR ANALYSIS)
      IF(NE.GT.NW) RETURN
C
C          READ IN VARIABLE NAMES.
C
      IF(NV.LE.0) GO TO 10
      KK = NL
    4 LL = KK+19
      IF(LL.GT.NR2) LL = NR2
      READ (5,2) (WORK(I),I=KK,LL)
      KK = LL+1
      IF(LL.LT.NR2) GO TO 4
C
C     READ IN TABLE
C
  10  READ(5,FMT) (WORK(I),I=1,NSIZE)
C
C          SEE IF Z IS TO BE READ.  READ X.
C
      IF(.NOT.ONE) READ(5,FMT) (WORK(I),I=NZ,NX1)
      IF(NV.LE.0) GO TO 8
      KK = NX
      LL = KK+NSIZE-1
      DO 5 J=1,NV
      READ(5,FMT) (WORK(I),I=KK,LL)
      KK = KK+NSIZE
      LL = LL+NSIZE
    5 CONTINUE
C
C          READ INITIAL VALUES OF B IF DESIRED.
C
      IF(INIT) GO TO 8
      KK = NB
    6 LL = KK+7
      IF(LL.GE.ND) LL = ND1
      READ(5,7) (WORK(I),I=KK,LL)
    7 FORMAT(8F10.5)
      KK = LL+1
      IF(LL.LT.ND1) GO TO 6
C
C          COMPUTE ESTIMATES.
C
    8 CALL CONTIN(NR,NC,NV,MX,TOL,WORK(1),WORK(NF),WORK(NZ),
     *   WORK(NX),WORK(NT),WORK(NB),WORK(ND),WORK(NRS),WORK(NS),
     *   NDF,INIT,ONE,IFAULT)
C
C          PRINT RESULTS.
C
      IF(IFAULT.EQ.1) WRITE(6,9)
    9 FORMAT(16HOINPUT IS FAULTY)
      IF(IFAULT.GT.0) RETURN
      IF(NV.LE.0) GO TO 27
      DO 12 J=1,NV
      KK = NS+(J-1)*(NV+1)
      LL = ND+J-1
      WORK(LL) = SQRT(WORK(KK))
   12 CONTINUE
      WRITE(6,13)
   13 FORMAT(33HOCOEFFICIENTS AND STANDARD ERRORS)
```

```
      KK = 1
   14 LL = KK+6
      IF(LL.GT.NV) LL = NV
      KK1 = NL+KK+KK-2
      LL1 = NL+LL+LL-1
      KK2 = NB+KK-1
      LL2 = NB+LL-1
      KK3 = ND+KK-1
      LL3 = ND+LL-1
      WRITE(6,15) (WORK(I),I=KK1,LL1)
   15 FORMAT(/1X,7(2X,2A4))
      WRITE(6,16) (WORK(I),I=KK2,LL2)
   16 FORMAT(1X,7F10.5)
      WRITE(6,16) (WORK(I),I=KK3,LL3)
      KK = KK+7
      IF(LL.LT.NV) GO TO 14
C
C         PRINT INFORMATION ON GOODNESS OF FIT.
C
   27 CALL CHISQ(RATIO,CHI,NSIZE,WORK(1),WORK(NF))
      WRITE(6,17) RATIO,CHI,NDF
   17 FORMAT(/32H  LIKELIHOOD RATIO CHI SQUARE = ,F10.5/
     *        32H            PEARSON CHI SQUARE = ,F10.5/
     *        32H NUMBER OF DEGREES OF FREEDOM = ,I4)
C
C         OBTAIN ADJUSTED RESIDUALS IF DESIRED.
C
      IF(IR.LE.0) GO TO 22
      CALL RESID(NR,NC,NV,WORK(1),WORK(NF),WORK(NS),WORK(NX),
     *  WORK(NT),WORK(NRS))
      WRITE(6,18)
   18 FORMAT(/11H     I    J,16H  OBSERVED COUNT,
     *  16H  EXPECTED COUNT,19H  ADJUSTED RESIDUAL)
      KK1 = 1
      KK2 = NF
      KK3 = NRS
      DO 20 J=1,NC
      DO 19 I=1,NR
      WRITE(6,21) I,J,WORK(KK1),WORK(KK2),WORK(KK3)
   21 FORMAT(1X,2I5,2F16.3,F19.3)
      KK1 = KK1+1
      KK2 = KK2+1
      KK3 = KK3+1
   19 CONTINUE
   20 CONTINUE
C
C         FIND GENERALIZED RESIDUALS IF DESIRED.
C
   22 IF(IC.LE.0) RETURN
      WRITE(6,23)
   23 FORMAT(/11H   CONTRAST,16H  OBSERVED VALUE,
     *  16H  EXPECTED VALUE,19H  ADJUSTED RESIDUAL)
      DO 30 K=1,IC
      READ(5,2) NAME1,NAME2
      READ(5,FMT) (WORK(I),I=NRS,NEN)
      CALL GENRES(NR,NC,NV,WORK(1),WORK(NF),WORK(NX),
     *  WORK(NT),WORK(NS),WORK(NRS),WORK(ND),OBS,EXP,RES)
      WRITE(6,24) NAME1,NAME2,OBS,EXP,RES
   24 FORMAT(3X,2A4,2F16.3,F19.3)
   30 CONTINUE
      RETURN
      END
C
C
      SUBROUTINE CONTIN(NR,NC,NV,MX,TOL,TABLE,FIT,Z,X,THETA,B,D,U,S,
     *  NDF,INIT,ONE,IFAULT)
C
C         THIS SUBROUTINE COMPUTES MAXIMUM-LIKELIHOOD ESTIMATES
C         FOR THE PARAMETERS IN A LOG-LINEAR MODEL.  THE
C         MULTINOMIAL VERSION OF THE NEWTON-RAPHSON ALGORITHM IS
C         USED.  IT IS ASSUMED THAT THE FREQUENCY TABLE UNDER STUDY
C         IS AN NR BY NC ARRAY.  TO EACH CELL OF TABLE(I,J) CORRESPOND
C         NV OBSERVATIONS X(I,J,K) AND AN OBSERVATION Z(I,J).
C         THE ESTIMATED EXPECTED VALUE FIT(I,J) SATISFIES THE MODEL
C         FIT(I,J) = Z(I,J)*EXP(A(J)+U(I,J))
C         AND U(I,J) = B(1)*X(I,J,1)+...+B(NV)*X(I,J,NV).
```

```
C           THE ASYMPTOTIC COVARIANCE MATRIX OF B IS S.
C           ITERATIONS CONTINUE UNTIL NO CHANGE IN COORDINATES OF
C           U IS GREATER THAN TOL OR UNTIL MAX ITERATIONS HAVE BEEN
C           COMPLETED.
C           THE CHANGE IN B IN THE LAST ITERATION IS D.
C           IF INIT IS .TRUE., THE INITIAL VALUES OF B ARE FOUND BY
C           A WEIGHTED REGRESSION BASED ON THE LOGS OF 0.5+TABLE(I,J)
C           FOR I BETWEEN 1 AND NR AND J BETWEEN 1 AND NC.
C           IF INIT IS .FALSE., THEN THE USER SETS INITIAL VALUES.
C           IFAULT IS RETURNED 0 UNDER NORMAL OPERATION.
C           IF NEGATIVE ELEMENTS OF TABLE OR Z ARE FOUND, IFAULT IS 1.
C           IF ONE IS .TRUE., ALL Z(I,J) ARE ASSUMED 1.
C           THETA(J,K) IS THE WEIGHTED AVERAGE
C           (X(1,J,K)*FIT(1,J)+...+X(NR,J,K)*FIT(NR,J))/(FIT(1,J)+...
C           +FIT(NR,J)).
C           NDF IS THE NUMBER OF DEGREES OF FREEDOM FOR CHI-SQUARE
C           TESTS.
C
      REAL TABLE(NR,NC),FIT(NR,NC),Z(NR,NC),U(NR,NC),X(NR,NC,NV),
     *  THETA(NC,NV),B(NV),D(NV),S(NV,NV)
      LOGICAL INIT,INIT1,ONE
      DOUBLE PRECISION SUM,SUM1
C
C           INITIALIZE ALGORITHM.
C
      F = 1.0E30
      IFAULT = 0
      NR1 = NR
      NC1 = NC
      NV1 = NV
      NDF = (NR1-1)*NC1
      INIT1 = INIT
      IF(NV1.LE.0) INIT1 = .FALSE.
      IF(.NOT.INIT1.AND.ONE) GO TO 5
      DO 2 J=1,NC1
      SUM = 0.0D0
      SUM1 = 0.0D0
      K = 0
      IF(ONE) K = 1
      DO 1 I=1,NR1
      IF(TABLE(I,J).LT.0.0) IFAULT = 1
      IF(IFAULT.GT.0) RETURN
      IF(INIT1) FIT(I,J) = TABLE(I,J)+0.5
      IF(INIT1) U(I,J) = ALOG(FIT(I,J))
      IF(ONE) GO TO 1
      IF(Z(I,J).LT.0.0) IFAULT = 1
      IF(IFAULT.GT.0) RETURN
      IF(Z(I,J).EQ.0.0) FIT(I,J) = 0.0
      IF(Z(I,J).EQ.0.0) NDF = NDF-1
      IF(INIT1.AND.Z(I,J).GT.0.0) U(I,J) = U(I,J)-ALOG(Z(I,J))
      IF(Z(I,J).GT.0.0) K = 1
    1 CONTINUE
      NDF = NDF+1-K
      DO 101 I=1,NR1
      SUM = SUM+FIT(I,J)
      SUM1 = SUM1+FIT(I,J)*U(I,J)
  101 CONTINUE
      IF(SUM.LE.0.0D0) GO TO 2
      W = SUM1/SUM
      DO 102 I=1,NR1
  102 U(I,J) = U(I,J)-W
    2 CONTINUE
      IF(.NOT.INIT1) GO TO 5
      DO 4 K=1,NV1
      B(K) = 0.0
      SUM = 0.0D0
      DO 3 J=1,NC1
      DO 3 I=1,NR1
    3 SUM = SUM+FIT(I,J)*U(I,J)*X(I,J,K)
      D(K) = SUM
    4 CONTINUE
C
C           PERFORM AN ITERATION.
C
    5 DO 36 IT=1,MX
```

```
      IF(IT.EQ.1.AND..NOT.INIT1) GO TO 20
      DO 10 J=1,NC1
      SUM = 0.0D0
      DO 7 I=1,NR1
    7 SUM = SUM+FIT(I,J)
      W = SUM
      DO 9 K=1,NV1
      SUM = 0.0D0
      DO 8 I=1,NR1
    8 SUM = SUM+FIT(I,J)*X(I,J,K)
      IF(W.GT.0.0) THETA(J,K) = SUM/W
      IF(W.LE.0.0) THETA(J,K) = 0.0
    9 CONTINUE
   10 CONTINUE
C
C         OBTAIN WEIGHTED SUMS OF CROSS-PRODUCTS.
C
      DO 15 K=1,NV1
      DO 14 L=1,K
      SUM = 0.0D0
      DO 13 J=1,NC1
      DO 12 I=1,NR1
   12 SUM = SUM+(X(I,J,K)-THETA(J,K))*(X(I,J,L)-THETA(J,L))*FIT(I,J)
   13 CONTINUE
      S(K,L) = SUM
   14 CONTINUE
   15 CONTINUE
C
C         OBTAIN CHOLESKY DECOMPOSITION OF S.
C
      CALL CHOL(NV1,S,S,IRANK)
C
C         SEE IF FURTHER STEPS ARE NEEDED.
C
      IF(F.LT.TOL) GO TO 37
C
C         OBTAIN NEW VALUE OF B.
C
      CALL SOLVE(NV1,S,D,D)
      DO 16 K=1,NV1
   16 B(K) = B(K)+D(K)
C
C         UPDATE U AND FIT AND CHECK FOR CONVERGENCE.
C
      IF(IT.GT.1) F = 0.0
   20 DO 33 J=1,NC1
      DO 30 I=1,NR1
      SUM = 0.0D0
      IF(NV1.LE.0) GO TO 22
      DO 21 K=1,NV1
   21 SUM = SUM+B(K)*X(I,J,K)
      IF(IT.EQ.1) GO TO 22
      E = SUM
      E = ABS(E-U(I,J))
      IF(E.GT.F) F=E
   22 U(I,J) = SUM
      FIT(I,J) = EXP(U(I,J))
      IF(ONE) GO TO 30
      IF(Z(I,J).LT.0.0) IFAULT = 1
      IF(IFAULT.GT.0) RETURN
      FIT(I,J) = Z(I,J)*FIT(I,J)
   30 CONTINUE
      SUM = 0.0D0
      W = 0.0
      DO 31 I=1,NR1
      IF(.NOT.ONE.AND.Z(I,J).LE.0.0) GO TO 31
      W = W+TABLE(I,J)
      SUM = SUM+FIT(I,J)
   31 CONTINUE
      IF(SUM.GT.0.0D0) W = W/SUM
      IF(SUM.LE.0.0D0) W = 0.0
      DO 32 I=1,NR1
   32 FIT(I,J) = W*FIT(I,J)
   33 CONTINUE
      IF(NV1.LE.0) RETURN
```

```
C
C          PREPARE DIFFERENCES BETWEEN FITTED AND OBSERVED LINEAR
C          COMBINATIONS.
C
      DO 35 K=1,NV1
      SUM = 0.0D0
      DO 34 J=1,NC1
      DO 34 I=1,NR1
      IF(FIT(I,J).GT.0.0) SUM = SUM+(TABLE(I,J)-FIT(I,J))*X(I,J,K)
   34 CONTINUE
      D(K) = SUM
   35 CONTINUE
   36 CONTINUE
   37 CALL INVERT(NV1,S,S)
      NDF = NDF-IRANK
      RETURN
      END
C
C
      SUBROUTINE SOLVE(N,LU,B,X)
C
C          SOLVE LUX = B GIVEN U(I,I) NONZERO FOR I = 1,.....,N.
C
      REAL X(N),LU(N,N),B(N)
      DOUBLE PRECISION SUM
      NN = N
C
C          SOLVE LOWER TRIANGLE.
C
      X(1) = B(1)
      IF(LU(1,1).LE.0.0) X(1) = 0.0
      IF(NN.EQ.1) GO TO 3
      DO 2 I=2,NN
      SUM = 0.0D0
      I1 = I-1
      DO 1 J=1,I1
    1 SUM = SUM+LU(I,J)*X(J)
      X(I) = B(I)-SUM
      IF(LU(1,1).LE.0.0) X(I) = 0.0
    2 CONTINUE
C
C          SOLVE UPPER TRIANGLE.
C
    3 I = NN
      IF(LU(I,I).GT.0.0) X(I) = X(I)/LU(I,I)
      IF(LU(I,I).LE.0.0) X(I) = 0.0
      IF(NN.EQ.1) RETURN
      DO 5 K=2,NN
      SUM = 0.0D0
      I1 = I
      I = I-1
      IF(LU(I,I).LE.0.0) X(I) = 0.0
      IF(LU(I,I).LE.0.0) GO TO 5
      DO 4 J=I1,NN
    4 SUM = SUM+LU(I,J)*X(J)
      X(I) = (X(I)-SUM)/LU(I,I)
    5 CONTINUE
      RETURN
      END
C
C
      SUBROUTINE CHISQ(RATIO,CHI,NSIZE,TABLE,FIT)
C
C          COMPUTE LIKELIHOOD RATIO(RATIO) AND PEARSON(CHI)
C          CHI SQUARE STATISTICS.  NOTE THAT THE TABLE OF
C          OBSERVATIONS HAS NSIZE ELEMENTS, AS DOES THE TABLE OF
C          FITTED CELL MEANS.
C
      INTEGER NSIZE
      REAL TABLE(NSIZE),FIT(NSIZE)
      DOUBLE PRECISION SUM,SUM1
      SUM = 0.0D0
      SUM1 = 0.0D0
      DO 1 I=1,NSIZE
C
C        NOTE THAT FIT(I) = 0 IMPLIES TABLE(I) = 0.
```

```
C
      IF(FIT(I).LE.0.0) GO TO 1
      SUM1 = SUM1+(TABLE(I)-FIT(I))**2/FIT(I)
      IF(TABLE(I).LE.0.0) GO TO 1
      SUM = SUM+TABLE(I)*ALOG(TABLE(I)/FIT(I))
    1 CONTINUE
      CHI = SUM1
      RATIO = 2.0*SUM
      RETURN
      END
C
C
      SUBROUTINE RESID(NR,NC,NV,TABLE,FIT,S,X,THETA,R)
C
C         RESID COMPUTES ADJUSTED RESIDUALS R FOR THE FITTED
C         TABLE FIT, THE ARRAYS X AND THETA, AND THE ASYMPTOTIC
C         COVARIANCE MATRIX S OF THE SUBROUTINE CONTIN.
C         PARAMETERS ARE DEFINED AS IN CONTIN.
C
      REAL TABLE(NR,NC),FIT(NR,NC),S(NV,NV),X(NR,NC,NV),R(NR,NC),
     *  THETA(NC,NV)
      DO 8 J=1,NC
      SUM1 = 0.0
      DO 10 I=1,NR
   10 SUM1 = SUM1+FIT(I,J)
   11 DO 7 I=1,NR
      SUM = 0.0
      IF(NV.LE.1) GO TO 3
      DO 2 K=2,NV
      LL = K-1
      DO 1 L=1,LL
      SUM = SUM+S(K,L)*(X(I,J,K)-THETA(J,K))*(X(I,J,L)-THETA(J,L))
    1 CONTINUE
    2 CONTINUE
      SUM = 2.0*SUM
    3 IF(NV.LE.0) GO TO 5
      DO 4 K=1,NV
      SUM = SUM+S(K,K)*(X(I,J,K)-THETA(J,K))**2
    4 CONTINUE
    5 SUM = FIT(I,J)*(1.0-FIT(I,J)*SUM)
      IF(SUM1.GT.0.0) SUM = SUM-FIT(I,J)*FIT(I,J)/SUM1
      R(I,J) = 0.0
      IF(SUM.LE.0.0) GO TO 7
      R(I,J) = (TABLE(I,J)-FIT(I,J))/SQRT(SUM)
    7 CONTINUE
    8 CONTINUE
      RETURN
      END
C
C
      SUBROUTINE INVERT(N,TRI,XINV)
C
C         COMPUTE THE INVERSE OF A MATRIX WITH MODIFIED CHOLESKY
C         DECOMPOSITION TRI.  OUTPUT IS XINV.
C
      REAL TRI(N,N),XINV(N,N)
      DOUBLE PRECISION SUM
      NN = N
C
C         INVERT DIAGONAL ELEMENTS.
C
      DO 1 I=1,NN
      IF(TRI(I,I).GT.0.0) XINV(I,I) = 1.0/TRI(I,I)
      IF(TRI(I,I).LE.0.0) XINV(I,I) = 0.0
    1 CONTINUE
C
C         IF N IS 1, ALL IS FINISHED.
C
      IF(NN.LE.1) RETURN
C
C         NOW FOR OFF-DIAGONALS.
C
      DO 5 I=2,NN
      I1 = I-1
      DO 4 K=1,I1
      SUM = -TRI(I,K)
```

```
      IF(K.GE.I1) GO TO 3
      K1 = K+1
      DO 2 J=K1,I1
    2 SUM = SUM-TRI(I,J)*XINV(J,K)
    3 XINV(I,K) = SUM
      XINV(K,I) = XINV(I,K)*XINV(I,I)
    4 CONTINUE
    5 CONTINUE
C
C         GET FULL INVERSE.
C
      DO 9 I=1,NN
      I1 = I+1
      DO 8 J=1,I
      SUM = XINV(J,I)
      IF(I.EQ.NN) GO TO 7
      DO 6 K=I1,NN
    6 SUM = SUM+XINV(K,I)*XINV(J,K)
    7 XINV(I,J) = SUM
    8 CONTINUE
    9 CONTINUE
      DO 10 I=2,NN
      I1 = I-1
      DO 10 J=1,I1
   10 XINV(J,I) = XINV(I,J)
      RETURN
      END
C
C
      SUBROUTINE CHOL(N,SYM,TRI,IRANK)
C
C         COMPUTE THE MODIFIED CHOLESKY DECOMPOSITION OF THE
C         N X N MATRIX SYM, WHERE SYM IS NONNEGATIVE DEFINITE.
C         OUTPUT IS TRI.  IRANK IS THE RANK OF SYM.
C
      DATA TOL/1.0E-4/
      REAL SYM(N,N),TRI(N,N)
      DOUBLE PRECISION SUM
      NN = N
      IRANK = 0
      DO 6 I=1,NN
      X = TOL*SYM(I,I)
      TRI(1,I) = SYM(I,1)
      IF(I.EQ.1) GO TO 4
      IF(TRI(1,1).GT.0.0) TRI(I,1) = TRI(1,I)/TRI(1,1)
      DO 3 J=2,I
      IF(TRI(J,J).LE.0.0) GO TO 3
      SUM = 0.0D0
      K = J-1
      DO 2 L=1,K
    2 SUM = SUM+TRI(I,L)*TRI(L,J)
      TRI(J,I) = SYM(I,J)-SUM
      IF(J.LT.I) TRI(I,J) = TRI(J,I)/TRI(J,J)
    3 CONTINUE
    4 IF(TRI(I,I).GT.X) IRANK = IRANK+1
      IF(TRI(I,I).GT.X) GO TO 6
      DO 5 J=1,NN
      TRI(I,J) = 0.0
      TRI(J,I) = 0.0
    5 CONTINUE
    6 CONTINUE
      RETURN
      END
C
C
      SUBROUTINE GENRES(NR,NC,NV,TABLE,FIT,X,THETA,S,C,D,OBS,EXP,RES)
C
C         GENRES COMPUTES GENERALIZED RESIDUALS.  NR,NC,NV,TABLE,
C         FIT, X, THETA, AND S ARE DEFINED AS IN CONTIN.
C         C IS THE ARRAY OF WEIGHTS USED AND D IS A WORK ARRAY
C         OF LENGTH NV.  OUTPUT CONSISTS OF THE OBSERVED LINEAR
C         COMBINATION OBS, THE ESTIMATED EXPECTED VALUE EXP, AND
C         THE ADJUSTED RESIDUAL RES.
```

```
C
      REAL TABLE(NR,NC),FIT(NR,NC),X(NR,NC,NV),THETA(NC,NV),S(NV,NV),
     *  C(NR,NC),D(NV)
      DOUBLE PRECISION SUM,SUM1,SUM2
C
C         INITIALIZE.
C
      NV1 = NV
      NR1 = NR
      NC1 = NC
C
C         OBTAIN ASYMPTOTIC VARIANCE ESTIMATE.
C
      SUM2 = 0.0D0
      DO 2 J=1,NC1
      SUM = 0.0D0
      SUM1 = 0.0D0
      DO 1 I=1,NR1
      SUM1 = SUM1+FIT(I,J)*C(I,J)
    1 SUM = SUM+FIT(I,J)
      SUM2 = SUM2+SUM1*SUM1/SUM
    2 CONTINUE
      IF(NV1.LE.0) GO TO 14
      DO 9 K=1,NV1
      SUM = 0.0D0
      DO 8 J=1,NC1
      DO 7 I=1,NR1
      SUM = SUM+(X(I,J,K)-THETA(J,K))*FIT(I,J)*C(I,J)
    7 CONTINUE
    8 CONTINUE
      D(K) = SUM
    9 CONTINUE
      SUM = 0.0D0
      IF(NV1.LE.1) GO TO 12
      DO 11 K=2,NV1
      LL = K-1
      DO 10 L=1,LL
   10 SUM = SUM+S(K,L)*D(K)*D(L)
   11 CONTINUE
      SUM = SUM+SUM
   12 DO 13 K=1,NV1
   13 SUM = SUM+S(K,K)*D(K)*D(K)
   14 SUM1 = 0.0D0
      DO 16 J=1,NC1
      DO 15 I=1,NR1
      SUM1 = SUM1+FIT(I,J)*C(I,J)**2
   15 CONTINUE
   16 CONTINUE
      ST = SUM1-SUM-SUM2
C
C         OBTAIN OBSERVED AND ESTIMATED VALUES.
C
      RES = 0.0
      SUM = 0.0D0
      SUM1 = 0.0D0
      DO 18 J=1,NC1
      DO 17 I=1,NR1
      SUM = SUM+TABLE(I,J)*C(I,J)
      SUM1 = SUM1+FIT(I,J)*C(I,J)
   17 CONTINUE
   18 CONTINUE
      OBS = SUM
      EXP = SUM1
C
C         OBTAIN ADJUSTED RESIDUAL.
C
      IF(ST.LE.0.0) RETURN
      ST = SQRT(ST)
      RES = (OBS-EXP)/ST
      RETURN
      END
```

A.2 LAT

The program LAT is written in much the same style as FREQ, except that LAT is used with latent-class models. The program can be used with a table of counts n_{ijk}, $1 \le i \le r$, $1 \le j \le s$, $1 \le k \le t$, with corresponding expected values m_{ijk}. The observations are the marginal totals

$$n_{jk}^{BC} = \sum_{i=1}^{r} n_{ijk}.$$

For known constants $z_{ijk} > 0$, $1 \le i \le r$, $1 \le j \le s$, $1 \le k \le t$, and x_{ijkl}, $1 \le i \le r$, $1 \le j \le s$, $1 \le k \le t$, $1 \le l \le p$, and unknown parameters α_k, $1 \le k \le t$, and β_l, $1 \le l \le p$, it is assumed that the expected values m_{ijk} satisfy the log-linear model

$$\log(m_{ijk}/z_{ijk}) = \alpha_k + \sum_{l=1}^{p} \beta_l x_{ijk}, \qquad z_{ijk} > 0.$$

It is also assumed that conditional on the observed counts

$$n_k^C = \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ijk},$$

the subtables $\{n_{ijk}: 1 \le i \le r,\ 1 \le j \le s\}$ have independent multinomial distributions for $1 \le k \le t$.

The program uses the method of maximum likelihood to obtain estimates $\hat{\beta}_l$ of the parameters β_l and estimated asymptotic standard deviations $s(\hat{\beta}_l)$ of the $\hat{\beta}_l$. Likelihood-ratio and Pearson chi-square statistics are computed, and residual analyses are performed.

Control Cards

Cards must appear in the following order.

Card 1. *Run description* This card contains 10 entries in the following format at the top of page 587.

Card 2. *Format of data* This card is a variable format card describing the format in which the data are to be read. The card is a FORMAT statement in which the work FORMAT has been removed and in which the initial parenthesis is in the first column. The data should be read in *F* format.

Card 3. *Format of independent variables* This card is a variable format card describing the format in which the independent variables x_{ijkl} are to be read. The card has the same structure as card 2.

Card 4. *Title* Up to 80 columns of alphanumeric title information.

Column	Format	Entry
1–4	I	The number r of rows. If the entry is not positive, $r = 1$.
5–8	I	The number s of columns. If the entry is not positive, $s = 1$
9–12	I	The number t of blocks. If the entry is not positive, $t = 1$.
13–16	I	The number p of independent variables x_{ijkl}, $1 \le l \le p$, for each row i, column j, and block t. If the entry is not positive, $p = 0$.
14–20	I	If positive, the z_{ijk} are supplied by the user. If not positive, each $z_{ijk} = 1$.
21–24	I	If positive, initial approximations for the β, $1 \le l \le p$, are supplied by the user. Otherwise, the program generates its own initial values from estimates of $\hat{m}_{ijk}$ supplied by the user.
25–28	I	If positive, observed values n_{jk}^{BC}, estimated expected values $\hat{m}_{jk}^{BC}$, and adjusted residuals $(n_{jk}^{BC} - \hat{m}_{jk}^{BC})/s(n_{jk}^{BC} - \hat{m}_{jk}^{BC})$ are printed, where $s(n_{jk}^{BC} - \hat{m}_{jk}^{BC})$ is the estimated asymptotic standard deviation of $n_{jk}^{BC} - \hat{m}_{jk}^{BC}$.
29–32	I	The number of generalized contrasts to be computed. For each contrast array d_{jk}, $1 \le j \le s$, $1 \le k \le t$, $$\mathrm{O} = \sum\sum d_{jk} n_{jk} \quad E = \sum\sum d_{jk} m_{jk}^{BC},$$ and $(\mathrm{O} - E)/s(\mathrm{O} - E)$ is computed, where $s(\mathrm{O} - E)$ is the estimated asymptotic standard deviation of $\mathrm{O} - E$.
33–36	I	The maximum number of iterations used in computing the $\hat{\beta}_l$. If blank, 10 is the maximum.
37–40	I	The stopping criterion g. The tolerance is 10^{-g} is $g \ne 0$ and 10^{-3} if $g = 0$. If no successive approximations for $\sum \hat{\beta}_l x_{ijkl}$, $1 \le i \le r$, $1 \le j \le s$, $1 \le k \le t$, differ in magnitude by more than the tolerance, then iterations stop.

Label Cards

Each 8 columns of each card contains the name of a variable. Thus columns 1–8 of the first label card contains the name of the first variable, x_{ijk1}, columns 9–16 contain the name of the second variable x_{ijk2}, etc. Note that if $p \le 10$, one label card is needed, if $11 \le p \le 20$, two label cards are needed, etc.

Data Cards

The data are read in standard FORTRAN order according to the format of card 2. For example, in a $2 \times 3 \times 2$ table with $n_{11}^{BC} = 10$, $n_{21}^{BC} = 11$, $n_{31}^{BC} = 20$, $n_{12}^{BC} = 25$, $n_{22}^{BC} = 37$, $n_{32}^{BC} = 31$, one might have a single data card

```
101120253731
```

The corresponding card 2 is

```
(6F2.0)
```

Z Cards

If the entry in columns 17–20 of card 1 is positive, then the coefficients z_{ijk} are read in the same format used for the independent variables. Observations are read in standard FORTRAN order so that if $r = 2$, $s = 3$, $t = 2$, $z_{ijk} = j$, and if card 3 is

(12F2.0)

then one Z card is required. The card is

010102020303010102020303

X Cards

For each l, $1 \leq l \leq p$, a card or series of cards lists the x_{ijkl}, $1 \leq i \leq r$, $1 \leq j \leq s$, $1 \leq k \leq t$. If $p = 2$,

$$x_{ijk1} = 1, \quad i = 1, \qquad x_{ijk2} = j, \quad i = 1,$$
$$= 0, \quad i = 2, \qquad = 0, \quad i = 2,$$

and if card 3 is

(12F2.0)

then the two X cards are

010001000100010001000100
010002000300010002000300

Initial Value Cards

If the entry in columns 21–24 of card 1 is positive, then initial approximations of the $\hat{\beta}_l$ are read in F10.5 format. The approximation for $\hat{\beta}_1$ is in columns 1–10 of the first initial value card, the approximation for $\hat{\beta}_2$ is in columns 11–20 of this card, etc. If $p \leq 8$, one initial value card is needed. If $9 \leq p \leq 16$, two such cards are needed, etc.

Initial Approximation Means Cards

If the entry in columns 21–24 of card 1 is not positive, initial values of m_{ijk0} are read in the same format used for the independent variables.

Contrast Label Cards

For each contrast requested in columns 29–32 of card 1, a contrast label card and contrast specification cards are required. The contrast label card contains a contrast label in columns 1–8.

The following listing contains a main program and a subroutine LATENT. $1 \leq k \leq t$ using the same format used for the data cards. If card 2 is

(6F2.0)

and if

$$\begin{aligned} d_{jk} &= 1, \qquad j = 1 \text{ or } 3, \quad k = 1, \\ &= 0, \qquad k = 2, \\ &= -2, \qquad j = 2, \quad k = 1, \end{aligned}$$

then a contrast label name and contrast card might be

QUADRAT.

01-201000000

Example 3. The simple latent-class model for Table 7.29

Control cards and output for a run with this model are shown below. The contrasts used are $n_{11}^{BY} - \hat{m}_{11}^{BY}$, $n_{11}^{DY} - \hat{m}_{11}^{DY}$, and $n_{11}^{FY} - \hat{m}_{11}^{FY}$.

Control Cards

```
  2  8  3  9  0        1  3
(24F3.0)
(3(12F6.2/),12F6.2)
SIMPLE LATENT-CLASS MODEL FOR TABLE 7.29
X1      B1      D1      F1      XB11    XD11    XF11    XY11    XY12
334 34 12 15 53 63 43501428 29 13 17 42 53 31453413 29 16 18 60 57 37430
    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.
    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.
    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.
    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.
    1.    1.    1.    1.    1.    1.    1.    1.   -1.   -1.   -1.   -1.
   -1.   -1.    1.   -1.    1.    1.    1.    1.    1.    1.    1.    1.
   -1.   -1.   -1.   -1.   -1.   -1.   -1.   -1.    1.    1.    1.    1.
    1.    1.    1.    1.   -1.   -1.   -1.   -1.   -1.   -1.   -1.   -1.
    1.    1.    1.    1.   -1.   -1.   -1.   -1.    1.    1.    1.    1.
   -1.   -1.   -1.   -1.   -1.    1.    1.    1.   -1.   -1.   -1.   -1.
    1.    1.    1.    1.   -1.   -1.   -1.   -1.    1.    1.    1.    1.
   -1.   -1.   -1.   -1.    1.    1.    1.    1.   -1.   -1.   -1.   -1.
    1.    1.   -1.   -1.    1.    1.   -1.   -1.    1.    1.   -1.   -1.
    1.    1.   -1.   -1.    1.    1.   -1.   -1.    1.    1.   -1.   -1.
    1.    1.   -1.   -1.    1.    1.   -1.   -1.    1.    1.   -1.   -1.
    1.    1.   -1.   -1.    1.    1.   -1.   -1.    1.    1.   -1.   -1.
    1.   -1.    1.   -1.    1.   -1.    1.   -1.   -1.    1.   -1.    1.
   -1.    1.   -1.    1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.
   -1.    1.   -1.    1.   -1.    1.   -1.    1.    1.   -1.    1.   -1.
    1.   -1.    1.   -1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.
    1.   -1.    1.   -1.   -1.    1.   -1.    1.    1.   -1.    1.   -1.
   -1.    1.   -1.    1.    1.   -1.    1.   -1.   -1.    1.   -1.    1.
    1.   -1.    1.   -1.   -1.    1.   -1.    1.    1.   -1.    1.   -1.
   -1.    1.   -1.    1.    1.   -1.    1.   -1.   -1.    1.   -1.    1.
    1.   -1.   -1.    1.    1.   -1.   -1.    1.    1.   -1.   -1.    1.
    1.   -1.   -1.    1..   1.   -1.   -1.    1.    1.   -1.   -1.    1.
    1.   -1.   -1.    1.    1.   -1.   -1.    1.    1.   -1.   -1.    1.
    1.   -1.   -1.    1.    1.   -1.   -1.    1.    1.   -1.   -1.    1.
    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.
    1.   -1.    1.   -1.    0.    0.    0.    0.    0.    0.    0.    0.
    0.    0.    0.    0.    0.    0.    0.    0.   -1.    1.   -1.    1.
   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.
    0.    0.    0.    0.    0.    0.    0.    0.    0.    0.    0.    0.
    0.    0.    0.    0.    1.   -1.    1.   -1.    1.   -1.    1.   -1.
    1.   -1.    1.   -1.    1.   -1.    1.   -1.   -1.    1.   -1.    1.
   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.   -1.    1.
327.99  6.01 17.00 17.00  6.00  6.00  0.27 14.73 26.50 26.50  1.13 61.87
  0.77 42.23  0.17500.83426.94  1.06 25.54  3.46 11.45  1.55  2.03 14.97
 36.99  5.01  6.32 46.68  3.70 27.30  1.12451.88412.86  0.14 28.48  0.52
 15.71  0.29  9.00  9.00 58.92  1.08 28.50 28.50 18.50 18.50  7.73422.27
BY11
  1  1  1  1
DY11
  1  1  0  0  1  1
FY11
  1  0  1  0  1  0  1
```

Output

```
SIMPLE LATENT-CLASS MODEL FOR TABLE 7.29

COEFFICIENTS AND STANDARD ERRORS

   X1          B1          D1          F1          XB11        XD11        XF11
  -0.10575    -0.31598     0.32711     0.01201     1.37164     1.39679     1.29334
   0.08644     0.04523     0.04933     0.03909     0.04484     0.04893     0.03876

   XY11        XY12
  -0.10471     0.04453
   0.02605     0.02575

 LIKELIHOOD RATIO CHI SQUARE =    10.09669
          PEARSON CHI SQUARE =     9.98288
NUMBER OF DEGREES OF FREEDOM =    12

   J     K  OBSERVED COUNT  EXPECTED COUNT  ADJUSTED RESIDUAL
   1     1         334.000         345.757             -1.815
   2     1          34.000          27.420              1.535
   3     1          12.000          12.321             -0.110
   4     1          15.000          18.004             -0.903
   5     1          53.000          46.407              1.203
   6     1          63.000          62.252              0.125
   7     1          43.000          40.085              0.597
   8     1         501.000         502.753             -0.232
   1     2         428.000         412.459              2.281
   2     2          29.000          32.091             -0.699
   3     2          13.000          14.292             -0.430
   4     2          17.000          16.186              0.249
   5     2          42.000          53.961             -2.126
   6     2          53.000          56.052             -0.518
   7     2          31.000          35.893             -1.022
   8     2         453.000         445.066              1.081
   1     3         413.000         416.809             -0.558
   2     3          29.000          32.374             -0.762
   3     3          16.000          14.406              0.531
   4     3          18.000          15.882              0.649
   5     3          60.000          54.405              0.994
   6     3          57.000          55.013              0.339
   7     3          37.000          35.204              0.377
   8     3         430.000         435.907             -0.810

  CONTRAST  OBSERVED VALUE  EXPECTED VALUE  ADJUSTED RESIDUAL
  BY11             395.000         403.501             -1.428
  DY11             484.000         481.836              0.342
  FY11             442.000         444.570             -0.410
```

Error Conditions

An error is reported and execution terminates if some z_{ijk} is negative or if space requirements are too large. Space may be adjusted by changing the region size of the array work in the main program and giving the constant NW the same value. The space in words needed in WORK is $pt(1 + s + rs) + st(2 + 3r) + 4p + p^2$ if all z_{ijk} are assumed 1. Otherwise, rst more words is needed. The appropriate changes are made in lines 35 and 37 of the program. Currently NW is 10,000.

Program Listing

The following listing contains a main program and a subroutine LATENT. The subroutines CHOL, SOLVE, INVERT, CHISQ, RESID, and GENRES listed earlier are also required.

```
C
C         THIS PROGRAM PERFORMS MAXIMUM-LIKELIHOOD ESTIMATION
C         FOR THE LATENT-CLASS MODEL
C         LOG(M(I,J,K)/Z(I,J,K)) = A(K)+U(I,J,K), WHERE
C         U(I,J,K) = B(1)*X(I,J,K,1)+...+B(NV)*X(I,J,K,NV).
C         HERE A TABLE N(J,K) IS GIVEN WITH NR ROWS AND
C         NC COLUMNS.  THIS TABLE IS A MARGINAL TABLE FOR AN
C         UNOBSERVED NL BY NR BY NC TABLE WITH EXPECTED VALUE
C         M(I,J,K) IN CELL (I,J,K).
C         NORMALLY Z(I,J,K) IS POSITIVE.  IF Z(I,J,K) IS NOT POSITIVE,
C         THEN CELL (I,J,K) IS IGNORED DURING ANALYSIS.  THIS
C         PROVISION IS USEFUL IN ANALYSIS OF INCOMPLETE
C         TABLES.
C
C         FOR A DESCRIPTION OF CONTROL CARDS, SEE THE PROGRAM
C         DESCRIPTION.
C
C         IN THIS PROGRAM, I1 IS 0 IF Z(I,J,K) IS 1 AND POSITIVE
C         IF ANY OTHER CHOICE OF Z(I,J,K) IS TO BE MADE.
C         THE PARAMETER IS EQUALS 0 IF INITIAL APPROXIMATIONS
C         B(L,0), 1<=L<=NV, ARE TO BE FOUND BY THE PROGRAM
C         FROM AN INITIAL APPROXIMATION OF M(I,J,K) SUPPLIED
C         BY THE USER.
C         IF THE USER WISHES TO SUPPLY INITIAL
C         ESTIMATES OF THE B(L), THEN IS SHOULD BE POSITIVE.
C         ADJUSTED RESIDUALS ARE SUPPLIED IF IR IS POSITIVE.
C         GENERALIZED RESIDUALS ARE SUPPLIED IF IC IS POSITIVE.
C         MAX ITERATIONS MAY BE PERFORMED IF NEEDED.  IF
C         THE USER DEFINES MX TO BE POSITIVE, THEN MAX IS MX.
C         IF MX IS NOT POSITIVE, MAX IS 10.
C         ITERATIONS STOP IF NO CHANGE U(I,J,K,T)-U(I,J,K,T-1)
C         EXCEEDS 10**IT IN MAGNITUDE.  NORMALLY IT IS -3,
C         BUT THE USER MAY CHANGE THIS CHOICE.
C
      REAL WORK(10000),FMT(20),FMT1(20),TITLE(20)
      LOGICAL ONE,INIT
      DATA MXS,TOLS,NW/10,0.001,10000/
C
C         READ PARAMETERS.
C
      READ(5,1) NL,NR,NC,NV,I1,IS,IR,IC,MX,IT
    1 FORMAT(10I4)
      IF(NL.LE.0) NL = 1
      IF(NR.LE.0) NR = 1
      IF(NC.LE.0) NC = 1
      ONE = .TRUE.
      INIT = .TRUE.
      IF(I1.GT.0) ONE = .FALSE.
      IF(IS.GT.0) INIT = .FALSE.
      IF(MX.LE.0) MX = MXS
      IF(IT.NE.0) TOL = 10.0**IT
      IF(IT.EQ.0) TOL = TOLS
C
C         READ A VARIABLE FORMAT FOR THE DATA.
C
      READ(5,2) FMT
C
C         READ A VARIABLE FORMAT FOR THE MEANS AND INDEPENDENT
C         VARIABLES.
C
      READ(5,2) FMT1
C         READ A TITLE FOR THE RUN.  NOTE THAT UP TO 80 CHARACTERS
C         ARE AVAILABLE.
C
      READ(5,2) TITLE
    2 FORMAT(20A4)
      WRITE(6,11) TITLE
   11 FORMAT(1H1,20A4)
C
C         TO SET UP WORK AREAS, FIRST FIND THE SIZE OF THE TABLE.
C
      NSIZEM = NR*NC
      NSIZE = NL*NSIZEM
C
C         NF IS THE LOCATION IN THE WORK AREA OF THE FITTED
C         TABLE M OF EXPECTED VALUES.
C         NZ IS THE LOCATION OF THE TABLE Z IF I1>0.
```

```
C
      NF = NSIZEM+1
      NFE = NF+NSIZE-1
      IF(ONE) NZ = NF
      IF(.NOT.ONE) NZ = NF+NSIZE
C
C         NMAR IS THE LOCATION OF THE TABLE MB OF EXPECTED VALUES
C         OF THE MARGINAL TABLE N.
C         NX IS THE LOCATION OF THE TABLE X.
C         NT IS THE LOCATION OF THE TABLE THETA OF AVERAGES.
C         NXB IS THE LOCATION OF THE TABLE XB.
C
      NMAR = NZ+NSIZE
      NX = NMAR+NSIZEM
      NX1 = NMAR-1
      NXB = NX+NSIZE*NV
      NT = NXB+NSIZEM*NV
C
C         PROVIDE SPACE FOR COVARIANCE MATRIX AT NS.
C
      NS = NT+NC*NV
C
C         PROVIDE SPACE FOR VARIABLE NAMES AT NLB.
C
      NLB = NS+NV*NV
C
C         PLACE RESIDUALS AT NRS.
C
      NRS = NLB+NV+NV
      NR2 = NRS-1
      NEN = NR2+NSIZEM
C
C         PLACE PARAMETER ESTIMATES AT NB.
C
      NB = NRS+NSIZE
      ND = NB+NV
      ND1 = ND-1
      NE = ND+NV-1
C
C         CHECK FOR EXCESSIVE SPACE DEMANDS.
C
      IF(NE.GT.NW) WRITE(6,3)
    3 FORMAT(38HOTOO MUCH SPACE REQUESTED FOR ANALYSIS)
      IF(NE.GT.NW) RETURN
C
C         READ IN VARIABLE NAMES.
C
      IF(NV.LE.0) GO TO 10
      KK = NLB
    4 LL = KK+19
      IF(LL.GT.NR2) LL = NR2
      READ (5,2) (WORK(I),I=KK,LL)
      KK = LL+1
      IF(LL.LT.NR2) GO TO 4
C
C         READ DATA.
C
   10 READ(5,FMT) (WORK(I),I=1,NSIZEM)
C
C         SEE IF Z IS TO BE READ.  READ X.
C
      IF(.NOT.ONE) READ(5,FMT1) (WORK(I),I=NZ,NX1)
      IF(NV.LE.0) GO TO 8.
      KK = NX
      LL = KK+NSIZE-1
      DO 5 J=1,NV
      READ(5,FMT1) (WORK(I),I=KK,LL)
      KK = KK+NSIZE
      LL = LL+NSIZE
    5 CONTINUE
C
C         READ INITIAL VALUES OF B IF DESIRED.
C
      IF(INIT) READ(5,FMT1) (WORK(I),I=NF,NFE)
      IF(INIT) GO TO 8
      KK = NB
```

```
    6 LL = KK+7
      IF(LL.GE.ND) LL = ND1
      READ(5,7) (WORK(I),I=KK,LL)
    7 FORMAT(8F10.5)
      KK = LL+1
      IF(LL.LT.ND1) GO TO 6
C
C         COMPUTE ESTIMATES.
C
    8 CALL LATENT(NL,NR,NC,NV,MX,TOL,WORK(1),WORK(NF),WORK(NZ),
     *  WORK(NMAR),WORK(NX),WORK(NXB),WORK(NT),WORK(NB),WORK(ND),
     *  WORK(NRS),WORK(NS),NDF,INIT,ONE,IFAULT)
C
C         PRINT RESULTS.
C
      IF(IFAULT.EQ.1) WRITE(6,9)
    9 FORMAT(16HOINPUT IS FAULTY)
      IF(IFAULT.GT.0) RETURN
      IF(NV.LE.0) GO TO 27
      DO 12 J=1,NV
      KK = NS+(J-1)*(NV+1)
      LL = ND+J-1
      WORK(LL) = SQRT(WORK(KK))
   12 CONTINUE
      WRITE(6,13)
   13 FORMAT(33HOCOEFFICIENTS AND STANDARD ERRORS)
      KK = 1
   14 LL = KK+6
      IF(LL.GT.NV) LL = NV
      KK1 = NLB+KK+KK-2
      LL1 = NLB+LL+LL-1
      KK2 = NB+KK-1
      LL2 = NB+LL-1
      KK3 = ND+KK-1
      LL3 = ND+LL-1
      WRITE(6,15) (WORK(I),I=KK1,LL1)
   15 FORMAT(/1X,7(2X,2A4))
      WRITE(6,16) (WORK(I),I=KK2,LL2)
   16 FORMAT(1X,7F10.5)
      WRITE(6,16) (WORK(I),I=KK3,LL3)
      KK = KK+7
      IF(LL.LT.NV) GO TO 14
C
C         PRINT INFORMATION ON GOODNESS OF FIT.
C
   27 CALL CHISQ(RATIO,CHI,NSIZEM,WORK(1),WORK(NMAR))
      WRITE(6,17) RATIO,CHI,NDF
   17 FORMAT(/32H  LIKELIHOOD RATIO CHI SQUARE = ,F10.5/
     *        32H            PEARSON CHI SQUARE = ,F10.5/
     *        32H NUMBER OF DEGREES OF FREEDOM = ,I4)
C
C         OBTAIN ADJUSTED RESIDUALS IF DESIRED.
C
      IF(IR.LE.0) GO TO 22
      CALL RESID(NR,NC,NV,WORK(1),WORK(NMAR),WORK(NS),WORK(NXB),
     *  WORK(NT),WORK(NRS))
      WRITE(6,18)
   18 FORMAT(/11H    J    K,16H  OBSERVED COUNT,
     *  16H  EXPECTED COUNT,19H  ADJUSTED RESIDUAL)
      KK1 = 1
      KK2 = NMAR
      KK3 = NRS
      DO 20 K=1,NC
      DO 19 J=1,NR
      WRITE(6,21) J,K,WORK(KK1),WORK(KK2),WORK(KK3)
   21 FORMAT(1X,2I5,2F16.3,F19.3)
      KK1 = KK1+1
      KK2 = KK2+1
      KK3 = KK3+1
   19 CONTINUE
   20 CONTINUE
C
C         FIND GENERALIZED RESIDUALS IF DESIRED.
C
   22 IF(IC.LE.0) RETURN
      WRITE(6,23)
```

```
   23 FORMAT(/11H   CONTRAST,16H  OBSERVED VALUE,
     *  16H  EXPECTED VALUE,19H  ADJUSTED RESIDUAL)
      DO 30 K=1,IC
      READ(5,2) NAME1,NAME2
      READ(5,FMT) (WORK(I),I=NRS,NEN)
      CALL GENRES(NR,NC,NV,WORK(1),WORK(NMAR),WORK(NXB),
     *  WORK(NT),WORK(NS),WORK(NRS),WORK(ND),OBS,EXP,RES)
      WRITE(6,24) NAME1,NAME2,OBS,EXP,RES
   24 FORMAT(3X,2A4,2F16.3,F19.3)
   30 CONTINUE
      RETURN
      END
C
C
      SUBROUTINE LATENT(NL,NR,NC,NV,MX,TOL,TABLE,FIT,Z,FITM,X,XB,
     *  THETA,B,D,U,S,NDF,INIT,ONE,IFAULT)
C
C         THIS SUBROUTINE COMPUTES MAXIMUM-LIKELIHOOD ESTIMATES
C         FOR THE PARAMETERS IN A LATENT-CLASS MODEL.  THE
C         MULTINOMIAL VERSION OF THE SCORING ALGORITHM IS
C         USED.  IT IS ASSUMED THAT THE OBSERVED FREQUENCY TABLE
C         IS AN NR BY NC ARRAY AND THE LATENT TABLE IS AN NL BY NR
C         BY NC ARRAY.  TO CELL (I,J,K) OF THE LATENT TABLE
C         CORRESPOND NV OBSERVATIONS X(I,J,K,L), AN OBSERVATION
C         Z(I,J,K), AND AN ESTIMATED MEAN FIT(I,J,K).
C         THE ESTIMATED EXPECTED VALUE FIT(I,J,K) SATISFIES THE MODEL
C         FIT(I,J,K) = Z(I,J,K)*EXP(A(K)+U(I,J,K))
C         AND U(I,J,K) = B(1)*X(I,J,K,1)+...+B(NV)*X(I,J,K,NV).
C         THE ASYMPTOTIC COVARIANCE MATRIX OF B IS S.
C         ITERATIONS CONTINUE UNTIL NO CHANGE IN COORDINATES OF
C         U IS GREATER THAN TOL OR UNTIL MAX ITERATIONS HAVE BEEN
C         COMPLETED.
C         THE CHANGE IN B IN THE LAST ITERATION IS D.
C         IF INIT IS .TRUE., THE INITIAL VALUES OF B ARE FOUND BY
C         A WEIGHTED REGRESSION BASED ON A WEIGHTED AVERAGE OF THE
C         LOGS OF THE INITIAL M(I,J,K) FOR I FROM 1 TO NL.
C         IF INIT IS .FALSE., THEN THE USER SETS INITIAL VALUES.
C         IFAULT IS RETURNED 0 UNDER NORMAL OPERATION.
C         IF NEGATIVE ELEMENTS OF TABLE OR Z ARE FOUND, IFAULT IS 1.
C         IF ONE IS .TRUE., ALL Z(I,J,K) ARE ASSUMED 1.
C         THETA(K,L) IS THE WEIGHTED AVERAGE
C         (X(1,1,K,L)*FIT(1,1,K)+...+X(NL,NR,K,L)*FIT(NL,NR,K))/
C         (FIT(1,1,K)+...+FIT(NL,NR,K)).
C         NDF IS THE NUMBER OF DEGREES OF FREEDOM FOR CHI-SQUARE
C         TESTS.
C
      REAL TABLE(NR,NC),FIT(NL,NR,NC),Z(NL,NR,NC),FITM(NR,NC),
     *  U(NL,NR,NC),X(NL,NR,NC,NV),XB(NR,NC,NV),THETA(NC,NV),
     *  B(NV),D(NV),S(NV,NV)
      LOGICAL INIT,INIT1,ONE
      DOUBLE PRECISION SUM,SUM1,SUM2,SUM3
C
C         INITIALIZE ALGORITHM.
C
      F = 1.0E30
      IFAULT = 0
      NL1 = NL
      NR1 = NR
      NC1 = NC
      NV1 = NV
      NDF = (NR1-1)*NC1
      INIT1 = INIT
      IF(NV1.LE.0) INIT1 = .FALSE.
      IF(.NOT.INIT1.AND.ONE) GO TO 5
      DO 2 K=1,NC1
      L = 0
      DO 1 J=1,NR1
      IF(TABLE(J,K).LT.0.0) IFAULT = 1
      IF(IFAULT.GT.0) RETURN
      M = 0
      IF(ONE) M = 1
      DO 103 I=1,NL1
      IF(INIT1.AND.FIT(I,J,K).GT.0.0) U(I,J,K) = ALOG(FIT(I,J,K))
      IF(ONE) GO TO 103
      IF(Z(I,J,K).LT.0.0) IFAULT = 1
      IF(IFAULT.GT.0) RETURN
```

```
      IF(Z(I,J,K).EQ.0.0) FIT(I,J,K) = 0.0
      IF(INIT2.AND.FIT(I,J,K).LT.0.0) IFAULT = 1
      IF(IFAULT.GT.0) RETURN
      IF(Z(I,J,K).GT.0.0) M = 1
      IF(INIT1.AND.Z(I,J,K).GT.0.0) U(I,J,K) = U(I,J,K)-ALOG(Z(I,J,K))
  103 CONTINUE
      IF(M.EQ.1) L = 1
      IF(M.EQ.0) NDF = NDF-1
    1 CONTINUE
      NDF = NDF+1-L
      SUM2 = 0.0D0
      SUM3 = 0.0D0
      DO 104 J=1,NR1
      SUM = 0.0D0
      SUM1 = 0.0D0
      DO 101 I=1,NL1
      SUM = SUM+FIT(I,J,K)
      SUM1 = SUM1+FIT(I,J,K)*U(I,J,K)
  101 CONTINUE
      SUM2 = SUM2+SUM
      SUM3 = SUM3+SUM1
      IF(SUM.GT.0.0D0) U(1,J,K) = SUM1/SUM
      FITM(J,K) = SUM
  104 CONTINUE
      IF(SUM2.LE.0.0D0) GO TO 2
      W = SUM3/SUM2
      DO 102 J=1,NR1
  102 U(1,J,K) = U(1,J,K)-W
    2 CONTINUE
      IF(.NOT.INIT1) GO TO 5
      DO 4 L=1,NV1
      B(L) = 0.0
      SUM = 0.0D0
      DO 3 K=1,NC1
      DO 3 J=1,NR1
      SUM1 = 0.0D0
      DO 113 I=1,NL1
  113 SUM1 = SUM1+FIT(I,J,K)*X(I,J,K,L)
      XB(J,K,L) = SUM1/FITM(J,K)
    3 SUM = SUM+U(1,J,K)*SUM1
      B(L) = SUM
    4 CONTINUE
C
C          PERFORM AN ITERATION.
C
    5 DO 36 IT=1,MX
      IF(IT.EQ.1.AND..NOT.INIT1) GO TO 20
      DO 10 K=1,NC1
      SUM = 0.0D0
      DO 7 J=1,NR1
    7 SUM = SUM+FITM(J,K)
      W = SUM
      DO 9 L=1,NV1
      SUM = 0.0D0
      DO 8 J=1,NR1
    8 SUM = SUM+FITM(J,K)*XB(J,K,L)
      IF(W.GT.0.0) THETA(K,L) = SUM/W
      IF(W.LE.0.0) THETA(K,L) = 0.0
    9 CONTINUE
   10 CONTINUE
C
C          OBTAIN WEIGHTED SUMS OF CROSS-PRODUCTS.
C
      DO 15 L=1,NV1
      DO 14 M=1,L
      SUM = 0.0D0
      DO 13 K=1,NC1
      DO 12 J=1,NR1
   12 SUM = SUM+(XB(J,K,L)-THETA(K,L))*(XB(J,K,M)-THETA(K,M))*FITM(J,K)
   13 CONTINUE
      S(L,M) = SUM
   14 CONTINUE
   15 CONTINUE
C
C          OBTAIN CHOLESKY DECOMPOSITION OF S.
```

```
C
      CALL CHOL(NV1,S,S,IRANK)
C
C         SEE IF FURTHER STEPS ARE NEEDED.
C
      IF(F.LT.TOL) GO TO 37
C
C         OBTAIN NEW VALUE OF B.
C
      CALL SOLVE(NV1,S,D,D)
      DO 16 L=1,NV1
   16 B(L) = B(L)+D(L)
C
C         UPDATE U AND FIT AND CHECK FOR CONVERGENCE.
C
      IF(IT.GT.1) F = 0.0
   20 DO 33 K=1,NC1
      DO 30 J=1,NR1
      DO 30 I=1,NL1
      SUM = 0.0D0
      IF(NV1.LE.0) GO TO 22
      DO 21 L=1,NV1
   21 SUM = SUM+B(L)*X(I,J,K,L)
      IF(IT.EQ.1) GO TO 22
      E = SUM
      E = ABS(E-U(I,J,K))
      IF(E.GT.F) F=E
   22 U(I,J,K) = SUM
      FIT(I,J,K) = EXP(U(I,J,K))
      IF(ONE) GO TO 30
      IF(Z(I,J,K).LT.0.0) IFAULT = 1
      IF(IFAULT.GT.0) RETURN
      FIT(I,J,K) = Z(I,J,K)*FIT(I,J,K)
   30 CONTINUE
      SUM = 0.0D0
      W = 0.0
      DO 31 J=1,NR1
      W = W+TABLE(J,K)
      SUM1 = 0.0D0
      DO 131 I=1,NL1
      IF(.NOT.ONE.AND.Z(I,J,K).LE.0.0) GO TO 131
      SUM1 = SUM1+FIT(I,J,K)
  131 CONTINUE
      FITM(J,K) = SUM1
      SUM = SUM+SUM1
   31 CONTINUE
      IF(SUM.GT.0.0D0) W = W/SUM
      IF(SUM.LE.0.0D0) W = 0.0
      DO 32 J=1,NR1
      FITM(J,K) = W*FITM(J,K)
      DO 32 I=1,NL1
   32 FIT(I,J,K) = W*FIT(I,J,K)
   33 CONTINUE
      IF(NV1.LE.0) RETURN
C
C         PREPARE DIFFERENCES BETWEEN FITTED AND OBSERVED LINEAR
C         COMBINATIONS.
C
      DO 35 L=1,NV1
      SUM = 0.0D0
      DO 34 K=1,NC1
      DO 34 J=1,NR1
      XB(J,K,L) = 0.0
      SUM1 = 0.0D0
      DO 134 I=1,NL1
  134 SUM1 = SUM1+FIT(I,J,K)*X(I,J,K,L)
      IF(FITM(J,K).GT.0.0) XB(J,K,L) = SUM1/FITM(J,K)
      IF(FITM(J,K).GT.0.0) SUM = SUM+(TABLE(J,K)-FITM(J,K))*XB(J,K,L)
   34 CONTINUE
      D(L) = SUM
   35 CONTINUE
   36 CONTINUE
   37 CALL INVERT(NV1,S,S)
      NDF = NDF-IRANK
      RETURN
      END
```

References

Andersen, E. B. (1979). *Discrete statistical models with social science applications.* North-Holland, Amsterdam.

Anderson, T. W. (1954). On estimation of parameters in latent structure analysis. *Psychometrika* **19**, 1–10.

Anderson, T. W. (1959). Some scaling models and estimation procedures in the latent class model. U. Gernander (ed.), *Probability and statistics: The Harald Cramér volume*, pp. 9–38. Wiley, New York.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* **11**, 375–386.

Asano, C. (1965). On estimating multinomial probabilities by pooling incomplete samples. *Ann. Inst. Statist. Math.* **17**, 1–13.

Berger, J., and Snell, J. L. (1957). On the concept of equal exchange. *Behavioral Sci.* **2**, 111–118.

Bhapkar, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *J. Amer. Statist. Assoc.* **61**, 228–235.

Bishop, Y. M. M., and Fienberg, S. E. (1969). Incomplete two-dimensional contingency tables. *Biometrics* **22**, 119–128.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* MIT Press, Cambridge, Massachusetts.

Bloomfield, P. (1974). Linear transformations for multivariate binary data. *Biometrics* **30**, 609–617.

Bock, R. D. (1970). Estimating multinomial response relations. R. C. Bose *et al.* (eds.), *Essays in Probability and statistics*, pp. 111–132. Univ. of North Carolina Press, Chapel Hill, North Carolina.

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research.* McGraw-Hill, New York.

Bock, R. D., and Yates, G. (1973). MULTIQUAL: Log-linear analysis of nominal or ordinal qualitative data by the method of maximum likelihood. International Education Services, Chicago, Illinois.

Bowker, A. H. (1948). A test for symmetry in contingency tables. *J. Am. Statist. Assoc.* **43**, 572–574.

Brown, M. (1977). P2F: two-way frequency tables—empty cells and departures from independence. W. J. Dixon and M. B. Brown (eds.), *BMDP-77: Biomedical computer programs P-series*, pp. 278–296. Univ. of California Press, Berkeley, California, 1–38.

Caussinus, H. (1965). Contribution à l'analyse statistique des tableaux des corrélation. *Ann. Fac. Sci. Univ. Toulouse* **29**, 77–182.

Ceppellini, R., Siniscalco, M., and Smith, C. A. B. (1955). The estimation of gene frequencies in a random mating population. *Ann. Hum. Genet.* **20**, 97–115.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika* **37**, 256–266.

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

Delury, D. B. (1960). *Values and integrals of the orthogonal polynomials up to n = 26*. Univ. of Toronto Press, Toronto.

Deming, W. E. (1964[1943]). *Statistical adjustment of data*. Dover, New York.

Deming, W. E., and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* **11**, 427–444.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1–38.

Duncan, O. D. (1975). Partitioning polytomous variables in multiway contingency analysis. *Soc. Sci. Res.* **4**, 167–182.

Fay, R. E., and Goodman, L. A. (1975). ECTA Program: description for users. Dept. of Statistics, Univ. of Chicago, Chicago, Illinois.

Feller, W. (1968). *An introduction to probability theory and its applications*. Vol. I, 3rd ed. Wiley, New York.

Fienberg, S. E. (1969). Preliminary graphical analysis and quasi-independence for two-way contingency tables. *Appl. Statist.* **18**, 153–168.

Fienberg, S. E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables. *J. Amer. Statist. Assoc.* **65**, 1610–1616.

Fienberg, S. E. (1971). A statistical technique for historians: standardizing tables of counts. *J. Interdisciplinary Hist.* **1**, 305–315.

Fienberg, S. E. (1972a). The multiple-recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* **59**, 591–603.

Fienberg, S. E. (1972b). The analysis of incomplete multiway contingency tables. *Biometrics* **28**, 177–202.

Fisher, R. A. (1921). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *J. Agric. Sci.* **11**, 107–135.

Fisher, R. A., and Yates, F. (1963[1938]). *Statistical tables for biological, agricultural and medical research*, 6th ed. Hafner, New York.

Gibson, W. A. (1955). An extension of Anderson's solution for the latent structure equations. *Psychometrika* **20**, 69–73.

Goodman, L. A. (1963). Statistical methods for the preliminary analysis of transaction flows. *Econometrica* **31**, 197–208.

Goodman, L. A. (1964). A short computer program for the analysis of transaction flows. *Behavioral Sci.* **9**, 176–186.

Goodman, L. A. (1965). On the statistical analysis of mobility tables. *Am. J. Sociol.* **70**, 564–585.

Goodman, L. A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. *J. Am. Statist. Assoc.* **63**, 1091–1131.

Goodman, L. A. (1969). How to ransack social mobility tables and other kinds of cross-classified tabies. *Am. J. Sociol.* **75**, 1–40.

Goodman, L. A. (1971). A simple simultaneous test procedure for quasi-independence in contingency tables. *Appl. Statist.* **20**, 165–177.

Goodman, L. A. (1972). Some multiplicative models for the analysis of cross-classified data. *Proc. 6th Berkeley Symp. Math. Statist. Probability* **1**, 649–696.

Goodman, L. A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I—a modified latent structure approach. *Am. J. Sociol.* **79**, 1179–1259.

Goodman, L. A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.

Goodman, L. A. (1975). A new model for scaling response patterns: an application of the quasi-independence concept. *J. Am. Statist. Assoc.* **70**, 755–768.

Green, B. F. (1951). A general solution for the latent class model of latent structure analysis. *Psychometrika* **16**, 151–166.

Green, B. F. (1952). Latent structure analysis and its relation to factor analysis. *J. Am. Statist. Assoc.* **47**, 71–76.

Grizzle, J. E., and Williams, O. D. (1972). Analysis of contingency tables having ordered response categories. *J. Am. Statist. Assoc.* **67**, 55–63.

Guttman, L. (1950). The basis for scalogram analysis. S. A. Stouffer *et al.* (eds.), *Measurement and Prediction*, pp. 60–90. Princeton Univ. Press, Princeton, New Jersey.

Haberman, S. J. (1972). Log-linear fit for contingency table. *Appl. Statist.* **21**, 218–225.

Haberman, S. J. (1973b). The analysis of residuals in cross-classified tables. *Biometrics* **29**, 205–220.

Haberman, S. J. (1973c). CTAB: Analysis of multidimensional contingency tables by log-linear models: User's guide. International Educational Services, Chicago, Illinois.

Haberman, S. J. (1974a). *The analysis of frequency data.* Univ. of Chicago Press, Chicago, Illinois.

Haberman, S. J. (1974b). Log-linear models for frequency tables with ordered classifications. *Biometrics* **30**, 589–600.

Haberman, S. J. (1974c). Log-linear models for frequency tables derived by indirect observation. Maximum-likelihood equations. *Ann. Statist.* **2**, 911–924.

Haberman, S. J. (1976a). Generalized residuals for log-linear models. *Proc. Internat. Biometrics Conf., 9th* **1**, 104–122.

Haberman, S. J. (1976b). Iterative scaling procedures for log-linear models for frequency data derived by indirect observation. *Statist. Comput. Sect. Proc. Am. Statist. Assoc., 1975* 45–50.

Haberman, S. J. (1977). Product models for frequency tables involving indirect observation. *Ann. Statist.* **5**, 1124–1147.

Harris, S. J. (1910). On the selective elimination during the development of the fruits of *Staphylea*. *Biometrika* **7**, 452–504.

Ireland, C. T., and Kullback, S. (1968). Contingency tables with given marginals. *Biometrika* **55**, 179–188.

Ireland, C. T., Ku, H. H., and Kullback, S. (1969). Symmetry and marginal homogeneity of an r × r contingency table. *J. Am. Statist. Assoc.* **64**, 1323–1341.

Kastenbaum, M. (1958). Estimation of relative frequencies of four sperm types in *Drosophila Melanogaster*. *Biometrics* **14**, 223–228.

Keats, J. A. (1957). A test of significance of difference between correlated proportions. *Aust. J. Psychol.* **9**, 149–153.

Kendall, M. G., and Stuart, A. (1967). *The advanced theory of statistics*, Vol. 2, 2nd ed. Hafner, New York.

Kullback, S. (1968[1959]). *Information theory and statistics.* Dover, New York.

Lazarsfeld, P. F. (1950a). The logical and mathematical foundation of latent structure analysis. S. A. Stouffer *et al.* (eds.), *Measurement and prediction*, pp. 362–412. Princeton Univ. Press, Princeton, New Jersey.

Lazarsfeld, P. F. (1950b). The interpretation and computation of some latent structures. S. A. Stouffer *et al.* (eds.), *Measurement and prediction*, pp. 413–472. Princeton Univ. Press, Princeton, New Jersey.

Lazarsfeld, P. F., and Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin, Boston, Massachusetts.

Madansky, A. (1960). Determinantal methods in latent class analysis. *Psychometrika* **25**, 183–198.

Madansky, A. (1963). Tests of homogeneity for correlated samples. *J. Am. Statist. Assoc.* **58**, 97–119.

Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *J. Am. Statist. Assoc.* **58**, 690–700.

Mantel, N. (1966). Models for complex contingency tables and polytomous dosage response curves. *Biometrics* **22**, 83–95.

McHugh, R. B. (1956). Efficient estimation and local identification in latent class analysis. *Psychometrika* **21**, 331–347.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **12**, 153–157.

Mosteller, F. (1947). Equality of margins. *Am. Statist.* **1**, 2, 12.

Mosteller, F. (1968). Association and estimation in contingency tables. *J. Am. Statist. Assoc.* **63**, 1–28.

National Opinion Research Center (1972). Codebook for the Spring 1972 General Social Survey. National Opinion Research Center, Univ. of Chicago, Chicago, Illinois.

National Opinion Research Center (1973). Codebook for the Spring 1973 General Social Survey. National Opinion Research Center, Univ. of Chicago, Chicago, Illinois.

National Opinion Research Center (1974). Codebook for the Spring 1974 General Social Survey. National Opinion Research Center, Univ. of Chicago, Chicago, Illinois.

National Opinion Research Center (1975). Codebook for the Spring 1975 General Social Survey. National Opinion Research Center, Univ. of Chicago, Chicago, Illinois.

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135**, 370–384.

Nerlove, M., and Press, S. J. (1973). Univariate and multivariate loglinear and logistic models. Rand Corp., Santa Monica, California.

Pearson, E. S., and Hartley, H. O. (1966). *Biometrika tables for statisticians*, Vol. 1. Cambridge Univ. Press, London and New York.

Pearson, K., and Lee, A. (1901). Mathematical contributions to the theory of evolution. VIII. On the inheritance of characters not capable of exact quantitative measurement. *Phil. Trans. Ser A.* **195**, 79–150.

Press, S. J. (1972). *Applied multivariate analysis*. Holt, New York.

Registrar General of Births, Deaths, and Marriages in England (1883). *Annual summary of births, deaths, and causes of death in London and other great towns, 1883*. Eyre and Spottiswoode, London.

Savage, I. R., and Deutsch, K. W. (1960). A statistical model of the gross analysis of transaction flows. *Econometrics* **28**, 551–572.

Simon, G. (1974). Alternative analyses for the singly-ordered contingency table. *J. Am. Statist. Assoc.* **69**, 971–976.

Smith, C. A. B. (1956). Counting methods in genetical statistics. *Ann. Human. Genet.* **21**, 254–273.

Srole, L., Langner, T. S., Michael, S. T., Opler, M. K., and Rennie, T. A. C. (1962). *Mental health in the metropolis: The midtown Manhattan study*. McGraw-Hill, New York.

Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* **42**, 412–416.

Sundberg, R. (1972). Maximum likelihood theory and applications for distributions generated when observing a function of an exponential family variable. Dissertation, Institute of Mathematical Statistics, Stockholm Univ.

Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scand. J. Statist.* **1**, 49–58.

Theil, H. (1969). A multinomial extension of the linear logit model. *Internat. Econ. Rev.* **10**, 251–259.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *Am. J. Sociol.* **76**, 103–154.

Uhlenhuth, E. H., Balter, M. B., Lipman, R. S., and Haberman, S. J. (1975). Remembering life events. J. S. Strauss, H. M. Babigian, and M. Roff (eds.), *The origins and course of psychopathology: Methods of logitudinal research*, pp. 117–134. Plenum, New York.

Uhlenhuth, E. H., Lipman, R. S., Balter, M. B., and Stern, M. (1974). Symptom intensity and life stress in the city. *Arch. Gen. Psychiatry* **31**, 759–764.

U. S. Bureau of the Census (1972). *1970 census of population, subject reports*, Series PC(2)-4C.

U. S. Gov. Printing Office, Washington, D. C.

Waite, H. (1915). Association of finger-prints. *Biometrika* **10**, 421–478.

Watson, G. S. (1956). Missing and "mixed-up" frequencies in contingency tables. *Biometrics* **12**, 47–50.

Williams, E. J. (1952). Use of scores for the analysis of association in contingency tables. *Biometrika* **39**, 274–289.

Wishart, J., and Metakides, T. (1953). Orthogonal polynomial fitting *Biometrika* **40**, 361–369.

Yates, F. (1948). The analysis of contingency tables with groupings based on quantitative characteristics. *Biometrika* **35**, 176–181.

Yule, G. U. (1900). On the association of attributes in statistics. *Phil. Trans. Ser. A* **194**, 257–319.

Yule, G. U. (1912). On the methods of measuring association between two attributes. *J. Roy. Statist. Soc.* **75**, 579–642.

*Example Index**

*Index for Volumes 1 and 2.

Subject Index*

*Index for Volumes 1 and 2.

N

O