Methodology Branch
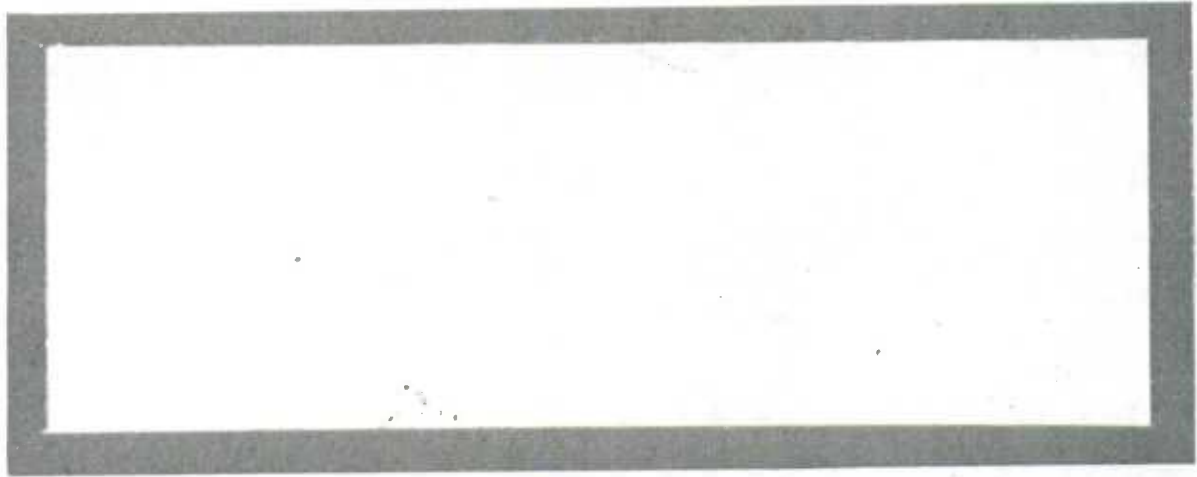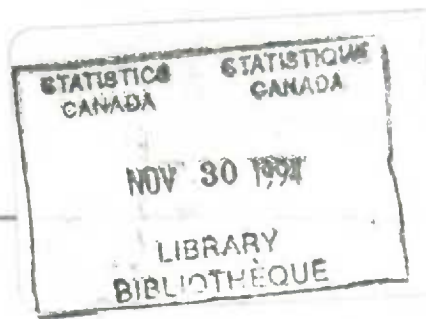
Institutions & Agriculture
Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquête
institutions et agriculture

Canada

NON-RESPONSE & LOGLINEAR MODELS FOR CATEGORICAL DATA

by

D.A. Binder

Non-response and loglinear models for categorical data

D.A. Binder
Statistics Canada, Ottawa, Canada

Summary

A general framework for analyzing multidimensional contingency tables with non-response is discussed. Emphasis is placed on modelling the data cells and the non-response mechanism. The implications of general log-linear models are discussed. Extensions to complex survey designs are given.

Some key words: Multi-dimensional contingency tables; Poisson models; Response mechanisms; Complex survey designs.

1. INTRODUCTION

In most surveys, in spite of all reasonable follow-up efforts and careful control of the survey process, non-response occurs. This non-response may be at the unit level (complete non-response) or at the item level. An excellent discussion of various methods for non-response adjustment is given in Platek, Singh and Tremblay(1977).

One of the most popular methods for non-response adjustment is post-stratification or weighting class adjustment; see Oh and Scheuren (1983). However, there are some inherent problems with this method, especially for large scale surveys. Some of these are:

(a) There may be so many potential weighting classes that the number of respondents in some classes is too small. This is especially true with surveys where the respondents are contacted on two or more occasions, and much information from the first occasion is available even for non-respondents to later occasions.

(b) In most surveys, non-response on different items would imply

Résumé

La présente étude donne un cadre général pour l'analyse des tableaux de
contingence multidimensionnels pour les non-réponses. Elle met l'accent sur
la modélisation des cellules de données et sur le mécanisme de non-réponse.
Elle passe par ailleurs en revue l'incidence des modèles généraux log-
linéaires, et fournit des exemples d'applications des plans de sondage
complexes.

Mots-clés: tableaux de contingence multidimensionnels; modèles de Poisson;
mécanismes de réponse; plans de sondage complexes.

different weighting classes for each item. This makes the
analysis difficult.

Because of these difficulties, a popular alternative is hot-deck
imputation. This yields individually clean records which are convenient
for tabulation. If the non-response rates are low, this is probably
quite suitable. The problem with higher non-response rates is that
(i) we are adding an imputation variance to the estimates (see Kalton
and Kish; 1984), and (ii) the estimates of variance will usually be
biased, possibly leading to misleading analytical conclusions.

In this paper, we concentrate on surveys with qualitative responses.
We find that by modelling the data and the non-response mechanism we
can develop a rich class of adjustment methods.

2. POISSON SAMPLING WITH ONE VARIABLE SUBJECT TO NON-RESPONSE

2.1 Notation

Suppose we have a cross-classification of categorical data, where
there may be non-response in only one of the variables. We let
subscript i index the variables which always have complete response
and subscript j indexes the variable which may be subject to non-
response. Our data consists of $\{n_{ij}\}$ (i=1, ..., I and j=1, ..., J) for
the complete responses and $\{n_{iU}\}$ (i=1, ..., I) for the incomplete
responses. Thus the data can be displayed in Tableau 1.

Now, it is well-known (see, for example Bishop, Fienberg and
Holland; 1975, p. 447) that maximum likelihood estimates for propor-
tions from a multinomial distribution are identical to those obtained

TABLEAU 1

	Complete Responses		Incomplete Responses	Totals
	n_{11} $\cdots$ n_{1J}		n_{1U}	n_{1+}
	$\vdots$ $\vdots$		$\vdots$	$\vdots$
	n_{I1} $\cdots$ n_{IJ}		n_{IU}	n_{I+}
Totals	n_{+1} $\cdots$ n_{+J}		n_{+U}	n_{++}

from independent Poisson samples for each cell. We therefore derive our results for the independent Poisson model. In particular, we assume that, had we observed all the complete responses, the distribution of cell (i,j) would be Poisson with mean λ_{ij}. To model the non-response mechanism, we assume that given the complete response is in cell (i,j), the probability of response is π_{ij}. These assumptions imply that the observed data $\{n_{ij}, n_{iU}\}$ are independent Poisson with means according to Tableau 2.

TABLEAU 2

Complete Responses	Incomplete Responses
$\lambda_{11}\,\pi_{11}$ $\cdots$ $\lambda_{1J}\,\pi_{1J}$	$\sum_j \lambda_{1j}(1 - \pi_{ij})$
$\vdots$ $\vdots$	$\vdots$
$\lambda_{I1}\,\pi_{I1}$ $\cdots$ $\lambda_{IJ}\,\pi_{IJ}$	$\sum_j \lambda_{Ij}(1 - \pi_{Ij})$

The log-likelihood function for the observations is thus:

$$\ell = - \lambda_{++} + \sum_{ij} n_{ij}\left[\log \lambda_{ij} + \log \pi_{ij}\right] + \sum_i n_{iU} \log\{\sum_j \lambda_{ij}(1-\pi_{ij})\},$$

$$(2.1)$$

where $\lambda_{++} = \Sigma\Sigma_{ij} \lambda_{ij}$. We also define $\lambda_{i+} = \Sigma_j \lambda_{ij}$.

Now, in general we have 2IJ unknown parameters with only I(J+1)
observations, so that the model is not identifiable unless there are
at least I(J-1) restrictions. In the following, we assume that these
restrictions can be parametrized so that $\lambda_{ij} = \lambda_{ij}(\theta)$ and $\pi_{ij} = \pi_{ij}(\beta)$,
where the unknown parameters θ and β are distinct (that is, the
parameter space for $\{\theta,\beta\}$ can be represented as a Cartesian product,
$\Theta \times B$). One simple consequence of this is that if $\pi_{i1}(\beta) = \ldots = \pi_{iJ}(\beta)$
for all i=1, ..., I, then the maximum likelihood estimates for $\{\lambda_{ij}\}$
will not depend on the estimated π_{ij}'s, so that the model for the π_{ij}'s
is inconsequential for estimating $\{\lambda_{ij}\}$. This is equivalent to Rubin's
(1976) notion of "missing at random". In particular, the maximum
likelihood estimator for $\{\lambda_{ij}\}$ is a solution to

$$\Sigma\Sigma_{ij} \frac{1}{\hat{\lambda}_{ij}} \frac{\partial \hat{\lambda}_{ij}}{\partial \hat{\theta}} \left[n_{ij} + \frac{\hat{\lambda}_{ij}}{\hat{\lambda}_{i+}} n_{iU}\right] = \Sigma\Sigma \frac{\partial \hat{\lambda}_{ij}}{\partial \hat{\theta}} \tag{2.2}$$

We see that this may be solved via a straight-form application of the
EM algorithm, where the complete data is estimated by $\{n_{ij} + \frac{\hat{\lambda}_{ij}}{\hat{\lambda}_{i+}} n_{iU}\}$
on each iteration, using the current estimates for $\{\hat{\lambda}_{ij}\}$. A more
efficient algorithm such as Newton-Raphson iteration may be preferable
in practice; see Section 2.3.

We call the model for the $\{\lambda_{ij}\}$ the data-model, whereas the model
for $\{\pi_{ij}\}$ will be referred to as the response-model.

2.2 Saturated Data Model

We now demonstrate that a saturated model for $\{\lambda_{ij}\}$ and a missing-at-random model for $\{\pi_{ij}\}$ leads to weighting class adjustment or post-stratification adjustment methods. In the saturated data model, $\lambda_{ij}(\theta) = \theta_{ij}$, so that (2.2) yields:

$$\frac{1}{\hat{\lambda}_{ij}} \left[n_{ij} + \frac{\hat{\lambda}_{ij}}{\hat{\lambda}_{i+}} n_{iU} \right] = 1$$

$$\Rightarrow \hat{\lambda}_{ij} = n_{ij} + \frac{\hat{\lambda}_{ij}}{\hat{\lambda}_{i+}} n_{iU}$$

$$\Rightarrow \hat{\lambda}_{i+} = n_{i+},$$

so that

$$\hat{\lambda}_{ij} = \alpha_i \, n_{ij} \, ,$$

where

$$\alpha_i = n_{i+}/(n_{i+} - n_{iU}) \, .$$

Thus each row i of $\{n_{ij}\}$ is reweighted by the factor α_i.

An alternative response model to that of missing at random is one where $\pi_{ij}(\beta) = \beta_j$. This model is identifiable only when $J \leq I$. If $J = I$, the observations will fit the estimated cell means exactly. In this case the data cannot be used to assess the relative merits of this model against the model where $\pi_{ij}(\beta) = \beta_i$. This decision must be based on subjective considerations. It is possible, though, to test whether β_j is constant. Now, for $\pi_{ij}(\beta) = \beta_j$, the maximum likelihood equations may be simplified to:

$$\hat{\lambda}_{ij} = n_{ij} + n_{iU} \frac{\hat{\lambda}_{ij} (1-\hat{\beta}_j)}{\sum_{\ell} \hat{\lambda}_{i\ell} (1-\hat{\beta}_\ell)}$$

$$\hat{\beta}_j = n_{+j}/\hat{\lambda}_{+j} ,$$

where $\qquad \hat{\lambda}_{+j} = \underset{i}{\Sigma} \, \hat{\lambda}_{ij} .$

2.3 Log-Linear Data Models

We now consider the implications of various log-linear model assumptions on the $\{\lambda_{ij}\}$, to accommodate a rich class of non-saturated models. We first consider the case where the response mechanism is assumed to be a missing-at-random model so that attention is focussed on the data model.

Suppose $\log \lambda_{ij}(\theta) = x_{ij}^T \, \theta$, where x_{ij} and θ are q-dimensional column vectors. Now, the estimating equations (2.2) may be written as:

$$X^T(\hat{\Lambda} - N - \hat{M}) = 0, \qquad\qquad (2.3)$$

where X is an $IJ \times q$ matrix with (i,j)-th row being x_{ij}^T, $\hat{\Lambda}$ is an $IJ \times 1$ vector of $\{\hat{\lambda}_{ij}\}$, N is an $IJ \times 1$ vector of $\{n_{ij}\}$ and $\hat{M}$ is an $IJ \times 1$ vector of $\{n_{iu} \, \hat{\lambda}_{ij}/\hat{\lambda}_{i+}\}$. To solve this system of equations, we suggest an iterative Newton-Raphson approach, so that the $(t+1)$-th iteration, $\hat{\theta}^{(t+1)}$ satisfies:

$$\hat{\theta}^{(t+1)} = \hat{\theta}^{(t)} - [\hat{J}^{(t)}]^{-1} X^T(\hat{\Lambda}^{(t)} - N - \hat{M}^{(t)}), \qquad (2.4)$$

where $\qquad \hat{J}^{(t)} = X^T[\hat{D}_1^{(t)} - \hat{D}_2^{(t)} + \hat{C}^{(t)} D_3 (\hat{C}^{(t)})^T] \, X,$

for $\qquad \hat{D}_1^{(t)} = \text{diag} \, \{\hat{\lambda}_{ij}^{(t)}\}$

$$\hat{D}_2^{(t)} = \text{diag} \, \{n_{iU} \, \hat{\lambda}_{ij}^{(t)}/\hat{\lambda}_{i+}^{(t)}\}$$

$$D_3 = \text{diag} \, \{n_{iU}\}$$

$$\hat{C}^{(t)} = \begin{vmatrix} \hat{c}_1^{(t)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \hat{c}_I^{(t)} \end{vmatrix}$$

$$c_i^{(t)} = \frac{1}{\hat{\lambda}_{i+}^{(t)}} (\hat{\lambda}_{i1}^{(t)}, \ldots, \hat{\lambda}_{iJ}^{(t)})^T$$

An initial starting point $\hat{\theta}^{(o)}$ must be identified. Normally the first component of θ represents an intercept term; that is, $x_{ij1} = 1$. If so, a convenient value for $\hat{\theta}^{(o)}$ is $[\log n_{++}/IJ, 0, \ldots, 0]^T$.

We now turn to the situation where the response mechanism is not assumed to be missing at random. In this case, we propose models of the form

$$\log \left[\frac{\pi_{ij}(\beta)}{1-\pi_{ij}(\beta)} \right] = z_{ij}^T \beta \ ,$$

where z_{ij} and β are r-dimensional column vectors. To ensure identifiability, we require $q + r \leq I(J + 1)$. We also require the matrices X and Z to be full rank. Now, the log-likelihood function (2.1) may be written as:

$$\ell = - \sum_i \sum_j \exp \{x_{ij}^T \theta\} + \sum\sum_{ij} n_{ij} \left[x_{ij}^T \theta + z_{ij}^T \beta - \log \{1 + \exp(z_{ij}^T \beta)\} \right]$$

$$+ \sum_i n_{iU} \left[\log \sum_j \{\exp(x_{ij}^T \theta)/(1 + \exp(z_{ij}^T \beta))\} \right] \ . \tag{2.5}$$

Differentiating with respect to θ and β, we obtain the following likelihood equation:

$$X^T(\hat{\Lambda} - N - \hat{M}) = 0 \tag{2.6a}$$

$$Z^T \left[(I - \hat{D}_\pi)N - \hat{D}_\pi \hat{M} \right] = 0 \tag{2.6b}$$

where $\quad$ X, $\hat{\Lambda}$ and N are defined in (2.3), $\hat{M}$ is modified to the IJ × 1 vector of $\{n_{iU}(1 - \hat{\pi}_{ij}) \hat{\lambda}_{ij} / \underset{k}{\Sigma}(1 - \hat{\pi}_{ik}) \hat{\lambda}_{ik}\}$ and $\hat{D}_\pi = \text{diag} \{\hat{\pi}_{ij}\}$. These equations could be solved iteratively using either Newton-Raphson iterations or applying the EM algorithm to the augmented table with 2IJ cells where IJ of the cells have missing values. To apply Newton-Raphson iterations, we need the following derivatives:

$$\frac{\partial \Lambda}{\partial \theta} = \left[\text{diag} \{\lambda_{ij}\}\right] X$$

$$\frac{\partial M}{\partial \theta} = \left[\text{diag} \{\frac{n_{iU}(1 - \pi_{ij}) \lambda_{ij}}{\underset{k}{\Sigma}(1 - \pi_{ik}) \lambda_{ik}}\}\right] X - C\left[\text{diag}\{n_{iU}\}\right] C^T X \,,$$

where

$$C = \begin{bmatrix} c_1 & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & \cdots & c_I \end{bmatrix}$$

for $\quad c_i = \dfrac{1}{\underset{k}{\Sigma}(1 - \pi_{ik}) \lambda_{ik}} \left[(1 - \pi_{i1}) \lambda_{i1}, \ldots, (1 - \pi_{iJ}) \lambda_{iJ}\right]^T$

$$\frac{\partial M}{\partial \beta} = - \left[\text{diag} \{\frac{n_{iu} \pi_{ij}(1 - \pi_{ij})\lambda_{ij}}{\underset{k}{\Sigma}(1 - \pi_{ik}) \lambda_{ik}}\}\right] Z + C\left[\text{diag}\{n_{iU}\}\right] C^T D_\pi Z$$

$$\frac{\partial D_\pi N}{\partial \beta} = \left[\text{diag}\{\pi_{ij}(1 - \pi_{ij}) n_{ij}\}\right] Z$$

$$\frac{\partial D_\pi M}{\partial \theta} = D_\pi \frac{\partial M}{\partial \theta} \,,$$

$$\frac{\partial D_\pi M}{\partial \beta} = D_\pi \frac{\partial M}{\partial \beta} + \left[\text{diag} \{\frac{n_{iU} \pi_{ij}(1 - \pi_{ij})^2 \lambda_{ij}}{\underset{k}{\Sigma}(1 - \pi_{ik}) \lambda_{ik}}\}\right] Z \,.$$

3. SUBSET SELECTION - THE CASE OF ARBITARY RESPONSE PATTERNS

3.1 Structure

In the previous section, we considered only the case where one of the variables is subject to non-response. In order to extend this to more general situations, we need to introduce the concepts of response patterns and subset selection.

A response pattern, P_k, defines which variables are observed and which variables have non-response. If we are studying an N-way table, the number of possible response patterns is $K \leq 2^N$. The set of all response patterns is denoted by $\{P_1, \ldots, P_K\}$. For example in the case of only one variable subject to non-response, $K = 2$, P_1 refers to complete response and P_2 refers to non-response on one variable.

Now, for response pattern P_k, there are associated subsets of cells which can be observed, denoted by $\{A_{k\ell}\}$, where for fixed k, the elements are mutually exclusive and exhaustive. For example, with complete response, this set contains all individual cells; whereas in the case of one variable having non-response, this set contains all subsets of cells where the variables with no non-response are specified. We see, therefore, that for a particular response pattern, P_k, the observations consist of counts corresponding to the number of times each of the subsets $\{A_{k\ell}\}$ is selected. Thus we have the concept of subset selection. To reflect the relationship between subsets and response patterns we require that $A_{k\ell} \neq A_{jm}$ whenever $k \neq j$ or $\ell \neq m$.

Now, the observations consist of $\{n_{k\ell}\}$ where $n_{k\ell}$ is the number of times subset $A_{k\ell}$ is selected. To model these observations, we start

with all indices i=1, ..., I corresponding to complete response. We assume that before imposing the non-response mechanism, the cell counts are independent Poisson samples with mean λ_i for the i-th cell. To model the non-response mechanism, we denote by π_{ik} the probability that we obtain response pattern P_k given that the complete response is in cell i.

Note that

$$\sum_k \pi_{ik} = 1.$$

Given this structure, we see that $\{n_{k\ell}\}$ are independent Poisson with means

$$\mu_{k\ell} = \sum_{i \varepsilon A_{k\ell}} \lambda_i \, \pi_{ik}$$

Therefore the log-likelihood function for this data is

$$- \sum_i \lambda_i + \sum\sum_{k\ell} n_{k\ell} \, \log(\sum_{i \varepsilon A_{k\ell}} \lambda_i \, \pi_{ik}) \tag{3.1}$$

subject to $\quad \sum_k \pi_{ik} = 1.$

We see that if π_{ik} is constant for all $i \varepsilon A_{k\ell}$, then we have a missing-at-random response mechanism.

3.2 Log-Linear Model

Analogously to Section 2, we impose a log-linear model structure on the $\{\lambda_i\}$, so we have $\log \lambda_i = x_i^T \theta$. To model the response mechanism $\{\pi_{ik}\}$ we use the log-linear analogue of the logit model used in Section 2; in particular

$$\log \pi_{ik} = z_{ik}^T \beta$$

subject to $\sum_k \pi_{ik} = 1$, $(i=1, \ldots, I)$.

The vectors x_i and θ are r-dimensional; the vectors z_{ik} and β are s-dimensional. To ensure identifiability, we require $r+s-I \leq \sum_k L_k$, where L_k is the number of subsets in $\{A_{k\ell}\}$ for given k. To estimate θ and β, we set to zero the derivatives with respect to θ, β and α of:

$$- \sum_i \exp\{x_i^T\theta\} + \sum\sum_{k\ell} n_{k\ell} \left[\log \sum_{i \in A_{k\ell}} \exp\{x_i^T\theta + z_{ik}^T \beta\} \right]$$

$$- \sum_i \alpha_i \sum_k \left[\exp\{z_{ik}^T\beta\} - 1 \right] \tag{3.2}$$

This results in the following likelihood equations:

$$\sum_i \hat{\lambda}_i x_{ia} - \sum\sum_{k\ell} n_{k\ell} \left[\frac{\sum\limits_{i \in A_{k\ell}} \hat{\lambda}_i \hat{\pi}_{ik} x_{ia}}{\sum\limits_{i \in A_{k\ell}} \hat{\lambda}_i \hat{\pi}_{ik}} \right] = 0; \text{ for } a=1, \ldots, r \tag{3.3a}$$

$$\sum\sum_{k\ell} \left[n_{k\ell} \left\{ \frac{\sum\limits_{i \in A_{k\ell}} \hat{\lambda}_i \hat{\pi}_{ik} z_{ikb}}{\sum\limits_{i \in A_{k\ell}} \hat{\lambda}_i \hat{\pi}_{ik}} \right\} - \sum_{i \in A_{k\ell}} \hat{\alpha}_i \hat{\pi}_{ik} z_{ikb} \right] = 0; \text{ for } b=b, \ldots, s \tag{3.3b}$$

$$\sum_k \hat{\pi}_{ik} = 1 \quad i \text{ for } i=1, \ldots, I \tag{3.3c}$$

We see that if the response mechanism is a missing-at-random model, equations (3.3a) reduces to

$$\sum_i \hat{\lambda}_i x_{ia} - \sum\sum_{k\ell} n_{k\ell} \frac{\sum\limits_{i \in A_{k\ell}} \hat{\lambda}_i x_{ia}}{\sum\limits_{i \in A_{k\ell}} \hat{\lambda}_i} = 0; \text{ for } i=1, \ldots, r. \tag{3.4}$$

This is simply the solution to the EM algorithm as described in Fuchs (1982). It is also the same result as found in Haberman (1974).

In general, expressions (3.3) define a system of I+r+s equations in I+r+s unknowns, which must be solved iteratively. However, expression (3.4) contains only r equations in r unknowns, so it is easier to solve.

Example: Consider a saturated two-way cross-classification as a data model, with a missing-at-random data mechanism. The data consist of $\{n_{ij}\}$ for complete response, $\{n_{iU}\}$ for non-response in the column variable, $\{n_{Uj}\}$ for non-response in the row variable and n_{UU} for complete unit non-response. The parameters of interest are denoted $\{\lambda_{ij}\}$. Equation (3.4) yields:

$$\hat{\lambda}_{ij} - n_{ij} - \sum_i n_{iU} \frac{\hat{\lambda}_{ij}}{\lambda_{i+}} - \sum_j n_{Uj} \frac{\hat{\lambda}_{ij}}{\lambda_{+j}} - n_{UU} \frac{\hat{\lambda}_{ij}}{\lambda_{++}} = 0. \tag{3.5}$$

We see that this leads to raking ratio estimators. Consider now another response mechanism where

$$E[n_{ij}] = \lambda_{ij} \pi_1, \quad E[n_{iU}] = \sum_j \lambda_{ij} \pi_{2j} \quad E[n_{Uj}] = \sum_i \lambda_{ij} \pi_{3i},$$

$$E[n_{UU}] = \sum_{ij} \lambda_{ij} \pi_{4ij}, \text{ where}$$

$$\pi_1 + \pi_{2j} + \pi_{3i} + \pi_{4ij} = 1 .$$

The estimating equations for this model are:

$$\lambda_{ij} - n_{ij} - n_{iU} \frac{\lambda_{ij} \pi_{2j}}{\sum_j \lambda_{ij} \pi_{2j}} - n_{Uj} \frac{\lambda_{ij} \pi_{3i}}{\sum_i \lambda_{ij} \pi_{3i}}$$

$$+ n_{UU} \frac{\lambda_{ij} \pi_{4ij}}{\sum_{ij} \lambda_{ij} \pi_{4ij}} = 0 \tag{3.6a}$$

$$\sum_{ij} \sum n_{ij} - \pi_1 \sum_{ij} \sum \alpha_{ij} = 0 \qquad (3.6b)$$

$$\sum_i n_{iU} \frac{\lambda_{ij} \pi_{2j}}{\sum_j \lambda_{ij} \pi_{2j}} - \pi_{2j} \sum_i \alpha_{ij} = 0 \qquad (3.6c)$$

$$\sum_j n_{Uj} \frac{\lambda_{ij} \pi_{3i}}{\sum_i \lambda_{ij} \pi_{3i}} - \pi_{3i} \sum_j \alpha_{ij} = 0 \qquad (3.6d)$$

$$n_{UU} \frac{\lambda_{ij} \pi_{4ij}}{\sum_{ij} \sum \lambda_{ij} \pi_{4ij}} - \pi_{4ij} \alpha_{ij} = 0 \qquad (3.6e)$$

$$\pi_1 + \pi_{2j} + \pi_{3i} + \pi_{4ij} = 1 \ . \qquad (3.6f)$$

4. OTHER SAMPLING SCHEMES

In Sections 2 and 3 we derived the maximum likelihood estimates for θ and β under Poisson sampling models. By standard treatments, the covariance matrix for these estimated parameters could be estimated, thus making available methods for constructing confidence intervals and performing tests of hypotheseses. However, suppose now that the $\{n_{k\ell}\}$ defined in Section 3 are not actual cell counts, but instead are population estimates of cell totals denoted by $\{N_{k\ell}\}$, based on a complex sample design. The estimation techniques described in Sections 2 and 3 could still be applied, yielding "pseudo-maximum likelihood estimates" of the parameters θ and β. Under the assumed response mechanism this would yield consistent estimates of these parameter values, provided the $\{N_{k\ell}\}$ are design-consistent estimates. If a consistent estimate of the covariance matrix of $\{N_{k\ell}\}$ is available, the covariance for $\hat{\theta}$ and

$\hat{\beta}$ could be obtained by, for example, using Taylor linearization. The derivation of these would be analogous to that given in Binder (1983).

For example, for $\hat{\theta}$ defined by (3.4), it can be shown that

$$V_{\hat{\theta}} = B \, V_{\hat{N}} \, B^T$$

where $\quad B = C^{-1} D$, for

$$C = \{c_{ab}\} \qquad (a=1, \ldots, r; \; b=1, \ldots, r)$$

where

$$c_{ab} = \sum_i \lambda_i \, x_{ia} \, x_{ib} - \sum_k \sum_\ell N_{k\ell} \, \frac{\sum\limits_{i \in A_{k\ell}} \lambda_i \, x_{ia} \, x_{ib}}{\sum\limits_{i \in A_{k\ell}} \lambda_i}$$

$$+ \sum_k \sum_\ell N_{k\ell} \, \frac{\sum\limits_{i \in A_{k\ell}} \sum\limits_{j \in A_{k\ell}} \lambda_i \, \lambda_j \, x_{ia} \, x_{jb}}{\left(\sum\limits_{i \in A_{k\ell}} \lambda_i \right)^2}$$

and $\quad D = \{d_{ab}\} \qquad (a=1, \ldots, r; \; b=1, \ldots, \sum\limits_k L_k)$

where $\quad d_{ab} = \dfrac{\sum\limits_{i \in A_{k\ell}} \lambda_i \, x_{ia}}{\sum\limits_{i \in A_{k\ell}} \lambda_i} \quad .$

5. SUMMARY

The procedures outlined in this paper have developed a rich class of models which could be used to adjust for non-response in multi-way contingency tables. However, because of problems of identifiability, alternative models may fit the data equally well. Little (1985) and Fay (1985) discuss a number of special cases within the framework described above. It is still very much a judgement decision as to which of

several competing models are appropriate. It is important here to try to model the causes of the non-response mechanism, and not just do the data analysis blindly. Once a data model is developed, if it is fairly parsimonious compared to the saturated model, more scope is available to model this mechanism and to test competing models.

REFERENCES

BINDER, D.A. (1983. On the variances of asymptotically normal estimators from complex surveys. Intl. Statist. Rev., 51, 279-292.

BISHOP, Y.M.M., FIENBERG, S.E. and HOLLAND, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Mass.

FAY, R.E. (1985). Causal models for patterns of non-response. Unpublished manuscript.

FUCHS, C. (1982). Maximum likelihood estimation and model selection in contingency tables with missing data. J. Amer. Statisti. Assoc., 77, 270-278.

HABERMAN, S.J. (1974). Log-linear models for frequency tables derived by indirect observation: Maximum likelihood equations. Ann. Statist., 2, 911-924.

KALTON, G. and KISH, L. (1984). Some efficient random imputation methods. Commun. Statist. Theor. Meth., 13(16), 1919-1939.

LITTLE, R.J.A. (1985). Nonresponse adjustments in longitudinal surveys: Models for categorical data. Invited paper at 45th Session of the Int'l. Statist. Inst., Amsterdam.

OH, H.L. and SCHEUREN, F.J. (1983). Weighting adjustment for unit non-response. Incomplete Data in Sample Surveys. Vol. 2. (Madow, Olkin and Rubin, eds.) 143-184. Academic Press, New York.

PLATEK, R., SINGH, M.P. and TREMBLAY, V. (1977). Adjustment for non-response in surveys. Survey Methodology, 3, 1-24.

RUBIN, D.B. (1976). Inference and missing data. Biometrika, 63, 581-592.