Journal of Official Statistics
Vol. 7. No. 4, 1991, pp. 393–404
© Statistics Sweden

A Framework for Analyzing Categorical Survey Data with Nonresponse

David A. Binder[1]

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

Key words: Multidimensional 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, nonresponse occurs. This nonresponse may be at the unit level (complete nonresponse) or at the item level. A concise discussion of various methods for nonresponse adjustment is given in Platek, Singh, and Tremblay (1977). Little (1988) mentions the three general strategies for handling nonresponse in survey data. These are:

- direct analysis of the incomplete data,
- imputation, and
- reweighting complete cases.

Many of the methods available are described in Little and Rubin (1987). In practice, direct analysis is usually avoided because of its complexity; reweighting is used for unit nonresponse and imputation is used to handle item nonresponse.

In some cases direct analysis will lead to reweighting procedures, especially for ignorable response patterns. One of the most popular reweighting methods in practice is poststratification or weighting class adjustment; see Oh and Scheuren (1983). However, there are some inherent problems with this method, even 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 panel surveys where the respondents are contacted on two or more occasions, and much information from the first occasion is available even for nonrespondents to later occasions; see Lepkowski (1989).
b. In most surveys, nonresponse on different items would imply different weighting classes for each item, so that the data file may have different weights for different items. Because of the

[1] Business Survey Methods Division, 11-F R.H. Coats Building, Statistics Canada, Ottawa, Canada K1A 0T6.

Acknowledgement: The author would like to thank the referees whose comments have led to a very much improved version of this paper.

difficulty in analyzing such data, reweighting is often restricted to unit nonresponse.

For item nonresponse, imputation methods are commonly used. For a discussion of imputation procedures, see Kalton and Kasprzyk (1982). Imputation yields individually clean records which are convenient for tabulation. If the nonresponse rates are low, this is probably quite suitable. The problem with higher nonresponse rates is that (i) we are adding an imputation variance to the estimate (see Kalton and Kish 1984) and (ii) the estimates of variance will usually be biased, possibly leading to misleading analytical conclusions. Rubin (1987) discusses these drawbacks and recommends multiple imputation to overcome some of the concerns. Little (1988) discusses examples of actual survey practices for imputation.

In this paper, we concentrate on surveys with qualitative responses. Fay (1986) suggests that when p variables are collected, some of which may be subject to nonresponse, the analysis proceeds with $2p$ variables, where we have p indicator variables for whether or not each variable had a response. A log-linear analysis of the $2p$-dimensional table is then possible. Such models can include both ignorable and non-ignorable response patterns. Little (1985) and Fay (1989) use this framework for the analysis of longitudinal data. Baker and Laird (1988) use this framework where only one variable is subject to nonresponse.

In these papers, emphasis is put on modelling the nonresponse mechanism, although log-linear analysis of the complete data is implicit in the framework. Here, we shall emphasize the analytic objective, where the data are to be fitted to a model. The nonresponse mechanism is included explicitly in the model. We also extend the analysis to the case of a complex survey, where we wish to incorporate the sample design weights in the analysis. We find that by modelling the data and the nonresponse mechanism we can develop a rich class of adjustment methods. The models proposed are extensions of those in Fay (1986).

In Section 3, we generalize the models to allow for arbitrary multidimensional cross-classifications. This provides a framework within which to explore a wide variety of nonresponse adjustment methods. Examples of applications can be found in Chen and Fienberg (1976), Nordheim (1984), Stasny (1986), Fay (1989) and Little and Su (1989). In Section 2, we present the simple case of a two-way table to motivate the ideas of Section 3. We also demonstrate the effect of various model assumptions on a simple numerical example. In Section 4, we discuss the implications of more complex sampling schemes.

2. Poisson Sampling with One Variable Subject to Nonresponse

2.1. Notation

First we consider the case where we have a cross-classification of categorical data, where there may be nonresponse 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 nonresponse. Therefore, for each cell (i, j) we have two possible response patterns, denoted by P_1 and P_2. Response pattern P_1 refers to complete response, so that cell (i, j) is observed, whereas response pattern P_2 refers to the incomplete response case, so that all we can observe is the row membership, i. Our data consist of $\{n_{ij} | i = 1, \ldots, I; j = 1, \ldots, J\}$ for the complete responses and $\{n_{iU} | i = 1, \ldots, I\}$ for the incomplete responses. Thus the data can be displayed as in Table 1.

To exemplify the effect of various model

Table 1. Observed counts in a two-way table with one variable subject to nonresponse

	Complete Responses			Incomplete Responses	Totals
	n_{11}	. . .	n_{1J}	n_{1U}	n_{1+}
	.		.	.	.
	.		.	.	.
	.		.	.	.
	n_{I1}	. . .	n_{IJ}	n_{IU}	n_{I+}
Totals	n_{+1}	. . .	n_{+J}	n_{+U}	n_{++}

assumptions for the data, we use the numerical example which appeared in Little (1985). Certain special cases of our overall model framework are exactly those as discussed by Little. Table 2 shows the frequencies.

Now, it is well known (see, for example Bishop, Fienberg, and Holland 1985, p. 477) that maximum likelihood estimates for proportions from a multinomial distribution are identical to those obtained 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 nonresponse mechanism, we assume that given the complete response was in cell (i, j), the probability of observing a complete response (response pattern P_1) is π_{ij1}. Alternatively, the probability of a nonresponse to the column variable (response pattern P_2) is π_{ij2}, where $\pi_{ij1} + \pi_{ij2} = 1$. These assumptions imply that the observed data $\{n_{ij}, n_{iU}\}$ are independent Poisson with means according to Table 3.

Table 2. Example of a 2 by 2 table with one variable subject to nonresponse

	Complete Responses		Incomplete Responses	Totals
	100	20	40	160
	30	50	60	140
Totals	130	70	100	300

The log-likelihood function for the observations is

$$\ell = -\lambda_{++} + \sum_i \sum_j n_{ij}[\log\{\lambda_{ij}\} + \log\{\pi_{ij1}\}] + \sum_i n_{iU} \log\left\{\sum_j \lambda_{ij}\pi_{ij2}\right\} \qquad (2.1)$$

subject to $\pi_{ij1} + \pi_{ij2} = 1$.

Now, in general we have $3IJ$ unknown parameters with only $I(J + 1)$ observations and IJ restrictions, so that the model parameters cannot be uniquely estimated unless there are at least $I(J - 1)$ further restrictions. Fay (1986) and Baker and Laird (1988) point out that having $I(J - 1)$ restrictions is a necessary but not sufficient condition for estimability. Problems with estimability can occur if some of the π's (the response propensities) are estimated to be 0 or 1, in which case a number of boundary solutions may exist.

The requirement for restrictions on the parameters leads to some arbitrariness in the selected models. This is where it becomes important to use external knowledge or belief about the response mechanism. Although producers of official statistics tend to avoid methods which depend on model assumptions, for problems such as nonresponse such models are inevitable, even if they are used implicitly rather than explicitly. It is

Table 3. Means of Poisson random variable in a two-way table with one variable subject to nonresponse

	Complete Responses			Incomplete Responses	Totals
	$\lambda_{11}\pi_{111}$	. . .	$\lambda_{1J}\pi_{1J1}$	$\sum_j \lambda_{1j}\pi_{1j2}$	λ_{1+}
	.		.	.	.
	.		.	.	.
	.		.	.	.
	$\lambda_{I1}\pi_{I11}$	. . .	$\lambda_{IJ}\pi_{IJ1}$	$\sum_j \lambda_{Ij}\pi_{Ij2}$	λ_{I+}
Totals	$\sum_i \lambda_{i1}\pi_{i11}$	. . .	$\sum_i \lambda_{iJ}\pi_{iJ1}$	$\sum_i\sum_j \lambda_{ij}\pi_{ij2}$	λ_{++}

desirable, though, to use methods which are not too sensitive to the model assumptions. The framework of this paper allows sufficient flexibility to include a wide variety of models which might be considered.

In the following, we assume that $\lambda_{ij} = \lambda_{ij}(\boldsymbol{\theta})$ and $\pi_{ijk} = \pi_{ijk}(\boldsymbol{\beta})$, where the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are distinct; that is, the parameter space for $\{\boldsymbol{\theta}, \boldsymbol{\beta}\}$ can be represented as a Cartesian product, $\Theta \times B$. One important consequence of this is that if $\pi_{ij2}(\boldsymbol{\beta})$ is independent of j for all i, then the maximum likelihood estimates for $\{\lambda_{ij}\}$ will not depend on the estimated π_{ijk}'s, so that the model for the π_{ijk}'s is inconsequential for estimating $\{\lambda_{ij}\}$. This is a special case of Rubin's (1976) notion of "missing at random." In particular, the maximum likelihood estimator for $\{\lambda_{ij}\}$ is a solution to

$$\sum_i\sum_j \frac{1}{\hat{\lambda}_{ij}} \frac{\partial\hat{\lambda}_{ij}}{\partial\hat{\boldsymbol{\theta}}} \left(n_{ij} + \frac{\hat{\lambda}_{ij}}{\hat{\lambda}_{i+}} n_{iU}\right) = \sum_i\sum_j \frac{\partial\hat{\lambda}_{ij}}{\partial\hat{\boldsymbol{\theta}}}. \tag{2.2}$$

We see that this may be solved via a straightforward application of the EM algorithm (Dempster, Laird, and Rubin 1977), where the complete data are estimated by $\{n_{ij} + (\hat{\lambda}_{ij}/\hat{\lambda}_{i+})n_{iU}\}$ 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 $\{\pi_{ijk}\}$ 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 $\{\pi_{ijk}\}$ lead to weighting class adjustment or poststratification adjustment methods. In the saturated data model, $\lambda_{ij}(\boldsymbol{\theta}) = \theta_{ij}$, so that (2.2) yields

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

This implies that

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

where

$$\alpha_i = \frac{n_{i+} + n_{iU}}{n_{i+}},$$

$$n_{i+} = \sum_{j=1}^{J} n_{ij}.$$

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

Little (1985) showed that the missing-at-random response model with the saturated data model, for the data given in Table 2, yields the estimated means in Table 4.

Table 4. Estimated means under the missing-at-random response model

	Estimated Means		Totals
	133.33	26.67	160
	52.5	87.5	140
Totals	185.83	114.17	300

An alternative response model to that of missing-at-random is one where $\pi_{ijk}(\boldsymbol{\beta}) = \beta_{jk}$, $(\beta_{j2} = 1 - \beta_{j1})$. This model is identifiable only when $J \leqslant I$. 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 $\pi_{ijk}(\boldsymbol{\beta}) = \beta_{ik}$. This decision must be based on external considerations. Now, for $\pi_{ijk}(\boldsymbol{\beta}) = \beta_{jk}$, the maximum likelihood equations may be simplified to

$$\hat{\lambda}_{ij} = n_{ij} + n_{iU} \frac{\hat{\lambda}_{ij}\hat{\beta}_{j2}}{\sum_{\ell} \hat{\lambda}_{i\ell}\hat{\beta}_{\ell 2}}$$

$$\hat{\beta}_{j1} = \frac{n_{+j}}{\hat{\lambda}_{+j}},$$

$$\hat{\beta}_{j2} = 1 - \hat{\beta}_{j1}$$

where

$$\hat{\lambda}_{+j} = \sum_{i} \hat{\lambda}_{ij}.$$

Little (1985) showed that, for the data given in Table 2, this model specification yields the estimates given in Table 5.

Other models for nonresponse could also be plausible for these data. For example, we might suppose that

$$\pi_{ijk} = \beta_{1k} \quad \text{if } i = j$$
$$= \beta_{2k} \quad \text{if } i \neq j.$$

This model might be used in a longitudinal survey where the probability of nonresponse to the second occasion depends only on whether there is a change of classification from the previous occasion. In this case we obtain estimates given by Table 6.

Table 5. Estimated means under a response model with column effects

	Estimated Means		Totals
	118.18	41.82	160
	35.45	104.55	140
Totals	153.63	146.37	300

Table 6. Estimated means under a response model with diagonal/off-diagonal effects

	Estimated Means		Totals
	100	60	160
	90	50	140
Totals	190	110	300

As can be seen from these examples, the final estimates can be sensitive to the assumed response model and the data alone cannot be used to judge which model is most appropriate. Each application must be assessed on its own merits to decide on the appropriate adjustment method.

2.3. *Log-linear data models*

In the previous section we considered only the case where the data model was saturated. This is the implicit model when the complete cross-classification is produced (at least when there is no nonresponse). As we have seen, this can severely limit the range of nonresponse models which are estimable. In data analysis, though, nonsaturated models are often considered. We now consider the implications of various log-linear model assumptions on the $\{\lambda_{ij}\}$ to accommodate a rich class of nonsaturated models. We first consider the case where the response mechanism is assumed to be a missing-at-random model, so that attention is focused on the data model.

Suppose $\log\{\lambda_{ij}(\boldsymbol{\theta})\} = \boldsymbol{x}_{ij}'\boldsymbol{\theta}$, where $\boldsymbol{x}_{ij}$ and

$\boldsymbol{\theta}$ are q-dimensional column vectors. This general formulation includes the analysis of main effects and interaction effects for contingency tables; see Bishop, Fienberg, and Holland (1975). It also includes more general models. Now, the estimating equations (2.2) may be written as

$$X'(\hat{\boldsymbol{\Lambda}} - N - \hat{M}) = \mathbf{0} \qquad (2.3)$$

where X is an $IJ \times q$ matrix with (i, j)th row being x'_{ij}, $\hat{\boldsymbol{\Lambda}}$ 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 $\{n_{iU}\hat{\lambda}_{ij}/\hat{\lambda}_{i+}\}$. Fuchs (1982) suggested using the EM algorithm to solve this system of equations, when direct estimates are not available.

This would result in the following scheme:

1. For current estimate $\hat{\boldsymbol{\theta}}^{(t)}$, compute $\hat{M}^{(t)}$ as described above;
2. Obtain new estimates $\hat{\boldsymbol{\Lambda}}^{(t+1)}$ by solving

$$X'\hat{\boldsymbol{\Lambda}}^{(t+1)} = X'[N + \hat{M}^{(t)}].$$

Since the second step is itself often iterative for many models (for example, iterative proportional fitting), this could result in many calculations to obtain convergence. However, for data models where direct estimators are available (e.g., independence of rows and columns, or the saturated data model), the EM algorithm may be preferable to alternatives such as Newton-Raphson iterations.

For the iterative Newton-Raphson approach, the $(t + 1)$th iteration, $\hat{\boldsymbol{\theta}}^{(t+1)}$, satisfies

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \hat{\boldsymbol{\theta}}^{(t)} - [\boldsymbol{J}^{(t)}]^{-1} X'[\hat{\boldsymbol{\Lambda}}^{(t)} - N - \hat{M}^{(t)}] \qquad (2.4)$$

where

$$\boldsymbol{J}^{(t)} = X'[\hat{D}_1^{(t)} - \hat{D}_2^{(t)} + \hat{C}^{(t)}D_3\hat{C}^{(t)\prime}]X$$

for

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

$$\hat{D}_2^{(t)} = \operatorname{diag}\left\{n_{iU}\frac{\hat{\lambda}_{ij}^{(t)}}{\hat{\lambda}_{i+}^{(t)}}\right\},$$

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

$$\hat{c}^{(t)} = \begin{bmatrix} \hat{c}_1^{(t)} & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \hat{c}_I^{(t)} \end{bmatrix},$$

$$\hat{c}_1^{(t)} = \frac{1}{\hat{\lambda}_{i+}^{(t)}}[\hat{\lambda}_{i1}^{(t)}, \ldots, \hat{\lambda}_{iJ}^{(t)}]'.$$

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

Returning to the data given in Table 2, suppose the data model contains no two-factor interactions, so that $\lambda_{ij} = \lambda_{i+}\lambda_{+j}/\lambda_{++}$. A missing-at-random response model leads to estimating equations given by

$$\lambda_{i+} = n_{i+} + n_{iU}$$

$$\lambda_{+j} = \left(\frac{\lambda_{++}}{n_{++}}\right)n_{+j}$$

so that the estimates are given by the values in Table 7.

Table 7. Estimated means under no two-factor interactions and a missing-at-random response model

	Estimated Means		Totals
	104	56	160
	91	49	140
Totals	195	105	300

Now, if the missing-at-random response model is assumed, the hypothesis of no two-factor interaction is testable. The expected cell counts under the model are given in Table 8.

A Pearson χ^2 test yields $X^2 = 54.85$ on one degree of freedom, which is highly significant, so we would reject the hypothesis of no two-factor interactions under a missing-at-random response mechanism.

Table 8. Estimated means under no two-factor interactions and a saturated response model

	Complete Responses		Incomplete Responses	Totals
	78	42	40	160
	52	28	60	140
Totals	130	70	100	300

We now turn to the situation where the response mechanism is not assumed to be missing-at-random. There are many models which may be used to explain the nonresponse mechanism. Here, we suggest considering those models which fit within the general framework of log-linear models for categorical data. This provides a rich class of alternatives from which to choose. In particular, the models suggested by Fay (1986) all fit into this class. However, the general log-linear framework allows for more arbitrary explanatory variables than are available in Fay's framework. For example, a logistic response function on continuous or discrete variables can be accommodated within this class.

The models we propose have the form

$$\log \pi_{ijk}(\boldsymbol{\beta}) = z'_{ijk}\boldsymbol{\beta}$$

where z_{ijk} and $\boldsymbol{\beta}$ are r-dimensional column vectors. We define $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ to be the matrices with elements $\{z_{ij1}\}$ and $\{z_{ij2}\}$, respectively, each having dimension $IJ \times r$. Now, the log-likelihood function (2.1) may be written as

$$\begin{aligned}\ell = & -\sum_i \sum_j \exp\{\boldsymbol{x}'_{ij}\boldsymbol{\theta}\} \\ & + \sum_i \sum_j n_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\theta} + \boldsymbol{z}'_{ij1}\boldsymbol{\beta}) \\ & + \sum_i n_{iU} \log\left\{\sum_j \exp(\boldsymbol{x}'_{ij}\boldsymbol{\theta} + \boldsymbol{z}'_{ij2}\boldsymbol{\beta})\right\}\end{aligned}$$

subject to

$$\exp\{z'_{ij1}\boldsymbol{\beta}\} + \exp\{z'_{ij2}\boldsymbol{\beta}\} = 1. \qquad (2.5)$$

Differentiating with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, we obtain the following likelihood equations

$$\boldsymbol{X}'(\hat{\boldsymbol{\Lambda}} - \boldsymbol{N} - \hat{\boldsymbol{M}}) = \boldsymbol{0}, \qquad (2.6a)$$

$$\boldsymbol{Z}'_1(\boldsymbol{N} - \hat{\boldsymbol{D}}_1\hat{\boldsymbol{A}}) + \boldsymbol{Z}'_2(\hat{\boldsymbol{M}} - \hat{\boldsymbol{D}}_2\hat{\boldsymbol{A}}) = \boldsymbol{0}, \qquad (2.6b)$$

$$\hat{\boldsymbol{D}}_1 + \hat{\boldsymbol{D}}_2 = \text{Identity Matrix} \qquad (2.6c)$$

where $\boldsymbol{X}$, $\hat{\boldsymbol{\Lambda}}$, and $\boldsymbol{N}$ are defined in (2.3), $\hat{\boldsymbol{M}}$ is modified to the $IJ \times 1$ vector of $\{n_{iU}\hat{\pi}_{ij2}\hat{\lambda}_{ij} / \Sigma_\ell \hat{\pi}_{i\ell 2}\hat{\lambda}_{i\ell}\}$, $\hat{\boldsymbol{D}}_1 = \operatorname{diag}\{\hat{\pi}_{ij1}\}$, $\hat{\boldsymbol{D}}_2 = \operatorname{diag}\{\hat{\pi}_{ij2}\}$ and $\hat{\boldsymbol{A}}$ is an $IJ \times 1$ vector of Lagrange multipliers.

3. Selecting Response Patterns – The Multidimensional Case

3.1. Structure

In the previous section, we considered only the case where one of the variables is subject to nonresponse. We now extend this to more general situations.

For a p-way table, Fay (1986) considered a response pattern as defining which variables are observed and which variables are missing. However, in general, a response pattern could be defined as simply the cell or combination of cells to which the observation is known to belong. Complete response gives information on the specific cell. Unit nonresponse indicates that the observation can be in any cell. Item nonresponse identifies a "slice" of the table to which the observation belongs. In general, though, we need not restrict ourselves to only these structures. For example, we may know that some observation belongs to one of the cells which have been collapsed. This may be used for variables such as income where the respondent is unwilling to fully disclose the exact category, but is willing to give the information when the available ranges are broader.

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 nonresponse, discussed in Section 2, P_1 refers to complete response and P_2 refers to nonresponse on the one variable.

In general, for response pattern P_k, there are associated subsets of cells within which the observations are known to belong, denoted by $\{A_{k\ell}\}$. For a given response pattern P_k, $\{A_{k\ell}\}$ consists of all observable cell combinations. For example, with complete response, this set contains all individual cells; whereas in the case of one variable having nonresponse, this set contains all subsets of cells where the variables with 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.

These observations are denoted by $\{n_{k\ell}\}$ where $n_{k\ell}$ is the number of times subset $A_{k\ell}$ is selected. To model these observations, we start with all indices $i = 1, \ldots, I$ corresponding to complete response. We assume that before imposing the nonresponse mechanism, the cell counts are independent Poisson samples with mean λ_i for the ith cell. To model the nonresponse mechanism, we denote by π_{ik} the probability that we obtain response pattern P_k, given that the complete observation belongs to cell i. Note that

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

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

$$\mu_{k\ell} = \sum_{i \in A_{k\ell}} \lambda_i \pi_{ik}.$$

Therefore the log-likelihood function for these data is

$$-\sum_i \lambda_i + \sum_k \sum_\ell n_{k\ell} \log\left(\sum_{i \in A_{k\ell}} \lambda_i \pi_{ik}\right) \quad (3.1)$$

subject to

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

We see that for a given k and ℓ, if π_{ik} is constant for all $i \in A_{k\ell}$, then we have an ignorable response mechanism.

3.2. Log-linear model

The categorical data models used to analyze the data will again be assumed to have a log-linear structure, so we have $\log \lambda_i = \boldsymbol{x}_i'\boldsymbol{\theta}$. In general, we have K response patterns. We consider here the implications of assuming a log-linear model for the response structure

$$\log \pi_{ik} = z_{ik}'\boldsymbol{\beta}$$

subject to

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

This log-linear model for the response patterns admits a rich class of models. For example, in a longitudinal survey, we might assume a logistic regression model for nonresponse to certain sets of variables, using variables from previous waves. We can include a number of variables, without requiring that the response probabilities depend on all the interaction terms, which is the implicit assumption in reweighting using poststratification.

The vectors $\boldsymbol{x}_i$ and $\boldsymbol{\theta}$ are q-dimensional; the vectors z_{ik} and $\boldsymbol{\beta}$ are r-dimensional. Assuming the parameters are estimable, we estimate $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ by setting to zero the derivatives with respect to $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ of

$$\begin{aligned} &-\sum_i \exp\{\boldsymbol{x}_i'\boldsymbol{\theta}\} \\ &\quad + \sum_k \sum_\ell n_{k\ell}\left(\log \sum_{i \in A_{k\ell}} \exp\{\boldsymbol{x}_i'\boldsymbol{\theta} + z_{ik}'\boldsymbol{\beta}\}\right) \\ &\quad - \sum_i \alpha_i \left(\sum_k \exp\{z_{ik}'\boldsymbol{\beta}\} - 1\right). \end{aligned} \quad (3.2)$$

This results in the following likelihood equations

$$\boldsymbol{X}'\left(\hat{\boldsymbol{\Lambda}} - \sum_k \hat{\boldsymbol{M}}_k\right) = \boldsymbol{0}, \quad (3.3a)$$

$$\sum_k Z_k'(\hat{M}_k - \hat{D}_k\hat{A}) = \mathbf{0}, \tag{3.3b}$$

$$\sum_k \hat{D}_k = \text{Identity Matrix}, \tag{3.3c}$$

where X is the $I \times q$ matrix with ith row being x_i', $\hat{\Lambda}$ is an $I \times 1$ vector of $\{\hat{\lambda}_i\}$, $\hat{M}_k$ is an $I \times 1$ vector with ith component

$$\sum_\ell n_{k\ell}\left\{\frac{\hat{\lambda}_i\hat{\pi}_{ik}}{\sum_{a\in A_{k\ell}}\hat{\lambda}_{ak}\hat{\pi}_{ak}}\right\},$$

Z_k is the $I \times r$ matrix with ith row being z_{ik}', $\hat{D}_k = \operatorname{diag}\{\hat{\pi}_{ik}\}$ for fixed k, and $\hat{A}$ is an $I \times 1$ vector of Lagrange multipliers $\{\hat{\alpha}_i\}$.

We see that if the response mechanism is ignorable, the ith component of $\hat{M}_k$ becomes

$$\sum_\ell n_{k\ell}\left\{\frac{\hat{\lambda}_i}{\sum_{a\in A_{k\ell}}\hat{\lambda}_{ak}}\right\}. \tag{3.4}$$

Thus, (3.3a) gives the estimator for $\{\lambda_i\}$ which is identical to that described in Fuchs (1982) and in Haberman (1974).

In general, expressions (3.3) define a system of $I + q + r$ equations in $I + q + r$ unknowns, which must be solved iteratively. However, for the case of ignorable response probabilities, expression (3.3a) contains only q equations in q unknowns, so it is easier to solve. For example, consider a saturated two-way cross-classification as a data model where either variable is subject to nonresponse and an ignorable response mechanism. The data consist of $\{n_{ij}\}$ for complete response, $\{n_{Uj}\}$ for nonresponse in the row variable, $\{n_{iU}\}$ for nonresponse in the column variable, and n_{UU} for complete unit nonresponse. The parameters of interest are denoted by $\{\lambda_{ij}\}$. Equation (3.3a) yields

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

Consider now a nonignorable response mechanism where

$$E(n_{ij}) = \lambda_{ij}\pi_1,$$
$$E(n_{iU}) = \sum_j \lambda_{ij}\pi_{j2},$$
$$E(n_{Uj}) = \sum_i \lambda_{ij}\pi_{i3},$$
$$E(n_{UU}) = \sum_i\sum_j \lambda_{ij}\pi_{ij4},$$
$$\pi_1 + \pi_{j2} + \pi_{i3} + \pi_{ij4} = 1.$$

The estimating equations for this model are

$$\hat{\lambda}_{ij} - n_{ij} - n_{iU}\frac{\hat{\lambda}_{ij}\hat{\pi}_{j2}}{\sum_j \hat{\lambda}_{ij}\hat{\pi}_{j2}} - n_{Uj}\frac{\hat{\lambda}_{ij}\hat{\pi}_{i3}}{\sum_i \hat{\lambda}_{ij}\hat{\pi}_{i3}} + n_{UU}\frac{\hat{\lambda}_{ij}\hat{\pi}_{ij4}}{\sum_i\sum_j \hat{\lambda}_{ij}\hat{\pi}_{ij4}} = 0, \tag{3.6a}$$

$$\sum_i\sum_j n_{ij} - \hat{\pi}_1\sum_i\sum_j \hat{\alpha}_{ij} = 0, \tag{3.6b}$$

$$\sum_i n_{iU}\frac{\hat{\lambda}_{ij}\hat{\pi}_{j2}}{\sum_j \hat{\lambda}_{ij}\hat{\pi}_{j2}} - \hat{\pi}_{j2}\sum_i \hat{\alpha}_{ij} = 0, \tag{3.6c}$$

$$\sum_j n_{Uj}\frac{\hat{\lambda}_{ij}\hat{\pi}_{i3}}{\sum_i \hat{\lambda}_{ij}\hat{\pi}_{i3}} - \hat{\pi}_{i3}\sum_j \hat{\alpha}_{ij} = 0, \tag{3.6d}$$

$$n_{UU}\frac{\hat{\lambda}_{ij}\hat{\pi}_{ij4}}{\sum_i\sum_j \hat{\lambda}_{ij}\hat{\pi}_{ij4}} - \hat{\pi}_{ij4}\hat{\alpha}_{ij} = 0, \tag{3.6e}$$

$$\hat{\pi}_1 + \hat{\pi}_{j2} + \hat{\pi}_{i3} + \hat{\pi}_{ij4} = 1. \tag{3.6f}$$

We see that although these equations can be solved, the assumed nonresponse mechanism is unusual in that the probability of complete response is constant, whereas the probability of complete nonresponse depends on the cell (i, j). The analyst must decide whether this is indeed reasonable. The choice of model within this framework will be somewhat specific to the context of the data, as a number of models will yield similar fits. It is important, therefore, for the analyst to choose the model carefully.

4. Other Sampling Schemes

In Sections 2 and 3 we derived the maximum likelihood estimates for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ 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 hypotheses. However, suppose that the $\{n_{k\ell}\}$ defined in Section 3 are not actual cell counts, but instead are population estimates of cell totals denoted by $\{\hat{N}_{k\ell}\}$, based on a complex sample design. The estimation techniques de.cribed in Sections 2 and 3 could still be applied, yielding "pseudo maximum likelihood estimates" of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. One rationalization for this is that we would obtain similar parameter estimates as we would have obtained under simple random sampling for the assumed model, provided the $\{\hat{N}_{k\ell}\}$ are design consistent. The nonresponse probabilities are estimated using a design consistent estimator. This may offer some protection against model failure. In particular, the population parameters which are being estimated are defined to be the maximum likelihood estimate if the whole population is surveyed under the same nonresponse mechanism as the mechanism for the actual data at hand. If a consistent estimate of the covariance matrix of $\{\hat{N}_{k\ell}\}$ is available, the covariance for $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ could be obtained by, for example, using Taylor linearization. The derivation would be analogous to that given in Binder (1983).

For example, for the ignorable case given by (3.3a) where $\hat{\boldsymbol{M}}_k$ is defined in (3.4), we have

$$\boldsymbol{V}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{B}\boldsymbol{V}_{\hat{N}}\boldsymbol{B}'$$

where

$$\boldsymbol{B} = \boldsymbol{C}^{-1}\boldsymbol{D},$$

$$\boldsymbol{C} = \frac{\partial}{\partial\boldsymbol{\theta}}\left[\boldsymbol{X}'\left(\boldsymbol{\Lambda} - \sum_k \boldsymbol{M}_k\right)\right],$$

$\boldsymbol{D}$ is a matrix with columns

$$\frac{\sum_{i\in A_{k\ell}} \lambda_i \boldsymbol{x}_i}{\sum_{i\in A_{k\ell}} \lambda_i}.$$

5. Discussion

In this paper we have proposed a framework which allows for a rich class of models to adjust for nonresponse when analyzing categorical survey data. However, because of problems of estimability, alternative models may fit the data equally well. Many of the models used for nonresponse adjustment which are discussed in the literature are special cases within this framework. The choice of model must be made based on external considerations. It is important here to try to model the causes of the nonresponse mechanism, and not just do the data analysis blindly.

Without some external information, it is virtually impossible to determine whether the nonresponse is ignorable. For example, suppose the analyst fits the data under the assumption of an ignorable response mechanism. Assuming that this results in some reduction of parameters compared to the saturated model, extra parameters which result in a nonignorable response pattern could be added. However, since the model under ignorable nonresponse already fits the data, any test of significance for ignorability would generally have fewer degrees of freedom than that used to fit the ignorable model. Therefore, the test would not reject ignorability. The exception to this would occur if $\{z_{ik}\}$ in the ignorable model includes variables which are additional to $\{\boldsymbol{x}_i\}$ used to fit the ignorable model.

6. References

Baker, S.G. and Laird, N.M. (1988). Regression Analysis for Categorical Variables with Outcome Subject to Nonignorable Nonresponse. Journal of the American Statistical Association, 83, 62–89.

Binder, D.A. (1983). On the Variances of Asymptotically Normal Estimators from Complex Surveys. International Statistical Review, 51, 279–292.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: MIT Press.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society, Ser. B, 39, 1–38.

Chen, T. and Fienberg, S.E. (1976). The Analysis of Contingency Tables with Incompletely Classified Data. Biometrics, 32, 133–144.

Fay, R.E. (1986). Causal Models for Patterns of Nonresponse. Journal of the American Statistical Association, 81, 354–365.

Fay, R.E. (1989). Estimating Nonignorable Nonresponse in Longitudinal Surveys Causal Modelling. In D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh (eds.), Panel Surveys, 375–399, New York: John Wiley.

Fuchs, C. (1982). Maximum Likelihood Estimation and Model Selection in Contingency Tables with Missing Data. Journal of the American Statistical Association, 77, 270–278.

Haberman, S.J. (1974). Log-Linear Models for Frequency Tables Derived by Indirect Observations: Maximum Likelihood Equations. Annals of Statistics, 2, 911–924.

Kalton, G. and Kasprzyk, D. (1982). Imputing for Missing Survey Responses. Proceedings of the Section on Survey Research Methods, American Statistical Assocation, 22–31.

Kalton, G. and Kish, L. (1984). Some Efficient Random Imputation Methods. Communications in Statistics – Theory and Methods, 13, 1919–1939.

Lepkowski, J.M. (1989). Treatment of Wave Nonresponse in Panel Surveys. In D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh (eds.), Panel Surveys, 348–374, New York: John Wiley.

Little, R.J.A. (1985). Nonresponse Adjustments in Longitudinal Surveys: Models for Categorical Data. Bulletin of the International Statistical Institute, 15, 1–15.

Little, R.J.A. (1988). Missing-Data Adjustments in Large Surveys. Journal of Business and Economic Statistics, 6, 287–296.

Little, R.J.A. and Rubin, D.B. (1987). The Statistical Analysis of Data with Missing Values. New York: John Wiley.

Little, R.J.A. and Su, H.-L. (1989). Item Nonresponse in Panel Surveys. In D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh (eds.), Panel Surveys, 400–425, New York: John Wiley.

Nordheim, E.V. (1984). Inference from Nonrandomly Missing Categorical Data: An Example from a Genetic Study on Turner's Syndrome. Journal of the American Statistical Association, 79, 772–780.

Oh, H.L. and Scheuren, F.J. (1983). Weighting Adjustment for Unit Nonresponse. In W. Madow, I. Olkin and D. Rubin (eds.), Incomplete Data in Sample Surveys, Vol. 2, 143–184, New York: Academic Press.

Platek, R., Singh, M.P., and Tremblay, V. (1977). Adjustment for Nonresponse in Surveys. Survey Methodology, 3, 1–24.

Rubin, D.B. (1976). Inference and Missing Data. Biometrika, 63, 581–592.

Rubin, D.B. (1987). Multiple Imputation

for Nonresponse in Surveys. New York: John Wiley.

Stasny, E.A. (1986). Estimating Gross Flows Using Panel Data with Nonresponse: An Example from the Canadian Labour Force Survey. Journal of the American Statistical Association, 81, 42–47.

Received January 1989
Revised July 1991