Obtaining the Maximum Likelihood Estimates in Incomplete $R \times C$ Contingency Tables Using a Poisson Generalized Linear Model

Stuart R. Lipsitz, Michael Parzen, and Geert Molenberghs

This article describes estimation of the cell probabilities in an $R \times C$ contingency table with ignorable missing data. Popular methods for maximizing the incomplete data likelihood are the EM-algorithm and the Newton–Raphson algorithm. Both of these methods require some modification of existing statistical software to get the MLEs of the cell probabilities as well as the variance estimates. We make the connection between the multinomial and Poisson likelihoods to show that the MLEs can be obtained in any generalized linear models program without additional programming or iteration loops.

Key Words: EM-algorithm; Ignorable missing data; Newton–Raphson algorithm; Offset.

1. INTRODUCTION

$R \times C$ contingency tables, in which the row variable and column variable are jointly multinominal, occur often in scientific applications. The problem of estimating cell probabilities from incomplete tables—that is, when either the row or column variable is missing for some of the subjects—is very common. For example, Table 1 contains such data from the Six Cities Study, a study conducted to assess the health effects of air pollution (Ware et al. 1984). The columns of Table 1 correspond to the wheezing status (no wheeze, wheeze with cold, wheeze apart from cold) of a child at age 10. The rows represent the smoking status of the child’s mother (none, medium, heavy) during that time. For some individuals the maternal smoking variable is missing, while for others the child’s wheezing status is missing. One of the objectives is to estimate the probabilities of the joint distribution of maternal smoking and respiratory illness.

Two popular methods for maximizing the incomplete data likelihood (assuming ignorable nonresponse in the sense of Rubin [1976]) are the EM-algorithm (Dempster,
Table 1. Six Cities Data: Maternal Smoking Cross-Classified by Child’s Wheeze Status

<table>
<thead>
<tr>
<th>Maternal Smoking</th>
<th>Child’s wheeze status</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Wheeze</td>
<td>Cold from Cold</td>
<td>Wheeze apart from Cold</td>
<td>Missing</td>
</tr>
<tr>
<td>None</td>
<td>287</td>
<td>39</td>
<td>38</td>
<td>279</td>
</tr>
<tr>
<td>Moderate</td>
<td>18</td>
<td>6</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>Heavy</td>
<td>91</td>
<td>22</td>
<td>23</td>
<td>201</td>
</tr>
<tr>
<td>Missing</td>
<td>59</td>
<td>18</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

Laird, and Rubin 1977) and the Newton–Raphson algorithm (Hocking and Oxspring 1971). Both of these methods require some modification of existing statistical software to get the MLEs of the cell probabilities as well as the variance estimates. In this article, we establish a connection between the multinomial and Poisson likelihoods to show that the MLEs can be obtained in any generalized linear models program, such as SAS Proc GENMOD (SAS Institute Inc. 1993), Stata glm (Stata Corporation 1995), S function glm (Hastie and Pregibon 1993), or GLIM (Francis, Green, and Payne1993), without any additional programming or iteration loops.

In Section 2, we formulate the missing data likelihood, and the connection between the multinomial likelihood and the Poisson likelihood. Section 3 describes the Poisson generalized linear model, including the design matrix and offset. Section 4 presents the SAS Proc GENMOD code for a \((2 \times 2)\) table given in Little and Rubin (1987) and the S function glm code for the data in Table 1.

2. INCOMPLETE \(R \times C\) TABLES

Suppose that two discrete random variables, \(Y_{i1}\) and \(Y_{i2}\), are to be observed on each of \(N\) independent subjects, where \(Y_{i1}\) can take on values \(1, \ldots, R\), and \(Y_{i2}\) can take on values \(1, \ldots, C\). Let the probabilities of the joint distribution of \(Y_{i1}\) and \(Y_{i2}\) be denoted by

\[
p_{jk} = \text{pr}[Y_{i1} = j, Y_{i2} = k],
\]

for \(j = 1, \ldots, R\) and \(k = 1, \ldots, C\). Because all of the probabilities must sum to 1, there are \((RC - 1)\) nonredundant multinomial cell probabilities. We let \(p\) denote the \((RC - 1) \times 1\) probability vector of \(p_{jk}\)’s; for simplicity, we delete \(p_{RC}\). The marginal probabilities are \(p_{j+} = \text{pr}[Y_{i1} = j]\) and \(p_{+k} = \text{pr}[Y_{i2} = k]\), where ‘+’ denote summing over the subscript it replaces. The joint probability function of \(Y_{i1}\) and \(Y_{i2}\) can be written as

\[
f(y_{i1}, y_{i2}|p) = \prod_{j=1}^{R} \prod_{k=1}^{C} p_{jk}^{I[Y_{i1}=j, Y_{i2}=k]},
\]

where \(I[\cdot]\) is an indicator function.

Now, suppose, due to an ignorable missing data mechanism (Rubin 1976), some individuals have a missing value for either \(Y_{i1}\) or \(Y_{i2}\). When there is missing data, it is
convenient to introduce two indicator random variables \( R_{1i} \) and \( R_{2i} \), where \( R_{1i} \) equals 1 if \( Y_{i1} \) is observed and equals 0 if \( Y_{i1} \) is missing. Similarly, \( R_{2i} \) equals 1 if \( Y_{i2} \) is observed and equals 0 if \( Y_{i2} \) is missing. It is assumed that no individuals are missing both \( Y_{i1} \) and \( Y_{i2} \)—that is, no individuals have \( R_{1i} = R_{2i} = 0 \) (if subjects are missing on both variables, they would not be considered participants in the study). The “complete” data for subject \( i \) are \((R_{1i}, R_{2i}, Y_{i1}, Y_{i2})\), with joint distribution

\[
L(\phi, \mathbf{p}) = L_1(\phi, \mathbf{p})L_2(\phi, \mathbf{p})L_3(\phi, \mathbf{p}),
\]

where

\[
L_1(\phi, \mathbf{p}) = \prod_{i=1}^{N} \left[ f(r_{1i} = 1, r_{2i} = 1 | y_{i1}, y_{i2}, \phi) f(y_{i1}, y_{i2} | \mathbf{p}) \right]^{r_{1i}r_{2i}};
\]

\[
L_2(\phi, \mathbf{p}) = \prod_{i=1}^{N} \left[ \sum_{y_{i1}} f(r_{1i} = 0, r_{2i} = 1 | y_{i1}, y_{i2}, \phi) f(y_{i1}, y_{i2} | \mathbf{p}) \right]^{1-r_{1i}r_{2i}};
\]

\[
f(y_{i1}, y_{i2}, r_{1i}, r_{2i} | \mathbf{p}, \phi) = f(r_{1i}, r_{2i} | y_{i1}, y_{i2}, \phi)f(y_{i1}, y_{i2} | \mathbf{p}),
\]

(2.1)

where

\[
f(r_{1i}, r_{2i} | y_{i1}, y_{i2}, \phi)
\]

is the “missing data mechanism” with parameter vector \( \phi \). We assume that \( \phi \) is distinct from \( \mathbf{p} \).

Following the nomenclature of Rubin (1976) and Little and Rubin (1987), a hierarchy of missing data mechanisms can be distinguished. First, when \( f(r_{1i}, r_{2i} | y_{i1}, y_{i2}, \phi) \) is independent of both \( Y_{i1} \) and \( Y_{i2} \), then the missing data are said to be missing completely at random (MCAR). When \( f(r_{1i}, r_{2i} | y_{i1}, y_{i2}, \phi) \) depends on the observed data, but not on the missing values, the missing data are said to be missing at random (MAR). Clearly, MCAR is a special case of MAR, and often no distinction is made between these two mechanisms when they are referred to as being ignorable. We caution, however, that the use of the term ignorable does not imply that the individuals with missing data can simply be ignored. Rather, the term ignorable is used here to indicate that it is not necessary to specify a model for the missing data mechanism to estimate \( \mathbf{p} \) in a likelihood-based analysis of the data.

Our goal is to estimate \( \mathbf{p} \) using likelihood methods when the missing data are MAR. If \( Y_{i1} \) or \( Y_{i2} \) is missing, individual \( i \) will not contribute (2.1) to the likelihood of the data, but will contribute (2.1) summed over the possible values of \( Y_{i1} \) or \( Y_{i2} \). In particular, if \( Y_{i1} \) is missing \((t = 1 \text{ or } 2)\), individual \( i \) contributes

\[
\sum_{y_{i1}} f(r_{1i}, r_{2i} | y_{i1}, y_{i2}, \phi) f(y_{i1}, y_{i2} | \mathbf{p})
\]

(2.2)

to the likelihood. Then, the full likelihood can be written as

\[
L(\phi, \mathbf{p}) = L_1(\phi, \mathbf{p})L_2(\phi, \mathbf{p})L_3(\phi, \mathbf{p}),
\]

where

\[
L_1(\phi, \mathbf{p}) = \prod_{i=1}^{N} \left[ f(r_{1i} = 1, r_{2i} = 1 | y_{i1}, y_{i2}, \phi) f(y_{i1}, y_{i2} | \mathbf{p}) \right]^{r_{1i}r_{2i}};
\]

(2.3)

\[
L_2(\phi, \mathbf{p}) = \prod_{i=1}^{N} \left[ \sum_{y_{i1}} f(r_{1i} = 0, r_{2i} = 1 | y_{i1}, y_{i2}, \phi) f(y_{i1}, y_{i2} | \mathbf{p}) \right]^{1-r_{1i}r_{2i}};
\]

(2.4)
and

\[ L_3(\phi, p) = \prod_{i=1}^{N} \left( \sum_{y_{i2}} f(r_{i1} = 1, r_{i2} = 0|y_{i1}, y_{i2}, \phi) f(y_{i1}, y_{i2}|p) \right)^{r_{i1}(1-r_{i2})} \]  

(2.5)

Now, suppose the missing data are MAR. MAR implies that \( f(r_{i1} = 0, r_{i2} = 1|y_{i1}, y_{i2}, \phi) \) in (2.4) does not depend on \( y_{i1} \); that is,

\[ f(r_{i1} = 0, r_{i2} = 1|y_{i1}, y_{i2}, \phi) = f(r_{i1} = 0, r_{i2} = 1|y_{i2}, \phi). \]

Then,

\[ L_2(\phi, p) = \prod_{i=1}^{N} \left[ f(r_{i1} = 0, r_{i2} = 1|y_{i2}, \phi) \sum_{y_{i1}} f(y_{i1}, y_{i2}|p) \right]^{(1-r_{i1})r_{i2}} \]

\[ = \prod_{i=1}^{N} \left[ f(r_{i1} = 0, r_{i2} = 1|y_{i2}, \phi) f(y_{i2}|p) \right]^{(1-r_{i1})r_{i2}}, \]

where

\[ f(y_{i2}|p) = \prod_{k=1}^{C} p_{i2}^{I_{i2}=k} \]

is the marginal distribution of \( Y_{i2} \). Similarly, MAR implies that \( f(r_{i1} = 1, r_{i2} = 0|y_{i1}, y_{i2}, \phi) \) in (2.5) does not depend on \( y_{i2} \); that is,

\[ f(r_{i1} = 1, r_{i2} = 0|y_{i1}, y_{i2}, \phi) = f(r_{i1} = 1, r_{i2} = 0|y_{i1}, \phi). \]

Then,

\[ L_3(\phi, p) = \prod_{i=1}^{N} \left[ f(r_{i1} = 1, r_{i2} = 0|y_{i1}, \phi) \sum_{y_{i2}} f(y_{i1}, y_{i2}|p) \right]^{r_{i1}(1-r_{i2})} \]

\[ = \prod_{i=1}^{N} \left[ f(r_{i1} = 1, r_{i2} = 0|y_{i1}, \phi) f(y_{i1}|p) \right]^{r_{i1}(1-r_{i2})}, \]

where

\[ f(y_{i1}|p) = \prod_{j=1}^{R} p_{i1}^{I_{i1}=j} \]

is the marginal distribution of \( Y_{i1} \).

Then, under MAR, the likelihood \( L(\phi, p) \) factors into two components, \( L(\phi, p) = L(\phi) L(p) \), where \( L(\phi) \) is a function only of \( \phi \), given by

\[ L(\phi) = \prod_{i=1}^{N} \left[ f(r_{i1} = 1, r_{i2} = 1|y_{i1}, y_{i2}, \phi) \right]^{r_{i1}r_{i2}} \]

\[ \times \left[ f(r_{i1} = 0, r_{i2} = 1|y_{i2}, \phi) \right]^{(1-r_{i1})r_{i2}} \left[ f(r_{i1} = 1, r_{i2} = 0|y_{i1}, \phi) \right]^{r_{i1}(1-r_{i2})}, \]

(2.6)
Table 2. Notation for a \((3 \times 3)\) Table With Missing Data

<table>
<thead>
<tr>
<th>Variable (Y_{i1})</th>
<th>Variable (Y_{i2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (u_{11}) (u_{12}) (u_{13}) (w_{1+})</td>
<td></td>
</tr>
<tr>
<td>2 (u_{21}) (u_{22}) (u_{23}) (w_{2+})</td>
<td></td>
</tr>
<tr>
<td>3 (u_{31}) (u_{32}) (u_{33}) (w_{3+})</td>
<td></td>
</tr>
<tr>
<td>Missing (z_{+1}) (z_{+2}) (z_{+3})</td>
<td></td>
</tr>
</tbody>
</table>

and \(L(p)\) is a function only of \(p\), given by

\[
L(p) = \prod_{i=1}^{N} \left[ f(y_{i1}, y_{i2} | p)^r_{i1} r_{i2} f(y_{i1} | p)^{r_{i1}(1-r_{i2})} f(y_{i2} | p)^{(1-r_{i1})r_{i2}} \right]. \tag{2.7}
\]

Because the likelihood factors, the maximum likelihood estimate (MLE) of \(\phi\) can be obtained from \(L(\phi)\) and the MLE of \(p\) can be obtained from \(L(p)\). We note here that an appropriate MAR missing data mechanism for \(f(r_{i1}, r_{i2} | y_{i1}, y_{i2}, \phi)\) is described in detail in Chen and Fienberg (1974) and Little (1985).

We can simplify the likelihood in (2.7). We let

\[
u_{jk} = \sum_{i=1}^{N} R_{i1} R_{i2} I[Y_{i1} = j, Y_{i2} = k]
\]

denote the number of subjects who are observed on both \(Y_{i1}\) and \(Y_{i2}\), with response level \(j\) on \(Y_{i1}\) and level \(k\) on \(Y_{i2}\). Also, we let

\[
w_{j+} = \sum_{i=1}^{N} R_{i1} (1 - R_{i2}) I[Y_{i1} = j]
\]

denote the number of subjects with response level \(j\) on \(Y_{i1}\) who are missing \(Y_{i2}\), and let

\[
z_{+k} = \sum_{i=1}^{N} (1 - R_{i1}) R_{i2} I[Y_{i2} = k]
\]

denote the number of subjects with response level \(k\) on \(Y_{i2}\) who are missing \(Y_{i1}\). The cell counts for a \((3 \times 3)\) table using this notation are shown in Table 2. Using this notation, the likelihood for the cell probabilities \(p\) in (2.7) reduces to

\[
L(p) = \left[ \prod_{j=1}^{R} \prod_{k=1}^{C} p_{jk}^{u_{jk}} \right] \left[ \prod_{j=1}^{R} p_{j+}^{w_{j+}} \right] \left[ \prod_{k=1}^{C} p_{+k}^{z_{+k}} \right]; \tag{2.8}
\]

that is, the complete cases (subjects with \(R_{i1} = R_{i2} = 1\)) contribute

\[
L_1(p) = \left[ \prod_{j=1}^{R} \prod_{k=1}^{C} p_{jk}^{u_{jk}} \right]. \tag{2.9}
\]
a multinomial likelihood with sample size ‘$u_+$’ and $(RC - 1)$ probabilities $p$; subjects observed only on $Y_{i1}$ contribute

$$L_2(p) = \prod_{j=1}^{R} p_{j+}^{w_{j+}},$$

(2.10)

a multinomial likelihood with sample size ‘$w_+$’ and $(R - 1)$ nonredundant probabilities $\{p_{1+}, \ldots, p_{R-1,+}\}$; and subjects observed only on $Y_{i2}$ contribute

$$L_3(p) = \prod_{k=1}^{C} p_{+k}^{z_{+k}},$$

(2.11)

a multinomial likelihood with sample size ‘$z_+$’ and $(C - 1)$ nonredundant probabilities $\{p_{1+}, \ldots, p_{C-1+}\}$.

First, we consider the contribution to the likelihood from the complete cases ($u_{jk}$’s). It is a well-known fact that a multinomial likelihood can be written as a Poisson likelihood (Agresti 1990). The multinomial likelihood is written in terms of the probabilities $p_{jk}$, whereas the Poisson is written in terms of the expected cell counts $E(u_{jk}) = u_{++} p_{jk}$.

Letting the $u_{jk}$’s be independent Poisson random variables, the Poisson likelihood contributed by the $u_{jk}$’s is

$$L_1(p) = e^{E(u_{++})} \prod_{j=1}^{R} \prod_{k=1}^{C} E(u_{jk})^{u_{jk}}$$

$$= e^{(u_{++} p_{++})} \prod_{j=1}^{R} \prod_{k=1}^{C} [u_{++} p_{jk}]^{u_{jk}},$$

(2.12)

where $p_{++} = \sum_{j=1}^{R} \sum_{k=1}^{C} p_{jk} = 1$. To constrain the Poisson likelihood to be equivalent to the multinomial likelihood with sample size $u_{++}$, we must restrict $E(u_{++}) = u_{++} p_{++}$ to equal $u_{++}$—that is, we must make sure that $p_{++} = 1$. This can be accomplished by rewriting $p_{RC} = 1 - \sum_{j \neq RC} p_{jk}$ in the likelihood, which is most easily handled using an offset in the Poisson generalized linear model, as described in the next section.

Formally, $E(u_{jk}) = u_{++} p_{jk}$ is only true if the data are missing completely at random. However, the MLEs under missing completely at random and the weaker missing at random are identical, and it is easier to explain the connection between Poisson and multinomial likelihoods when the data are missing completely at random. Thus, we discuss obtaining the MLEs under missing completely at random, keeping in mind that the MLEs are the same as under missing at random.

Next, we consider the contribution to the likelihood from the subjects only observed on $Y_{i1}$—that is, the $w_{j+}$’s. Again, we use the connection between the multinomial likelihood and the Poisson likelihood. We again write the Poisson likelihood in terms of the expected cell counts $E(w_{j+}) = w_{++} p_{j+}$. Letting the $w_{j+}$’s be independent Poisson random variables, the Poisson likelihood contributed by the $w_{j+}$’s is
The Poisson likelihood will be equivalent to the multinomial if we force $E(w_{++}) = w_{+++}$ to equal $w_{++}$, which is again accomplished by rewriting $p_{RC} = 1 - \sum_{jk \neq RC} P_{jk}$ in the likelihood, resulting again in an offset in the part of the Poisson generalized linear model corresponding to (2.13).

Finally, we consider the contribution to the likelihood from the subjects only observed on $Y_i$—that is, the $z_{++}$’s. Letting the $z_{++}$’s be independent Poisson random variables with $E(z_{++}) = z_{++}p_{++}$, the Poisson likelihood contributed by the $z_{++}$’s is

$$L_3(p) = e^{E(z_{++})} \prod_{k=1}^{C} [E(z_{++})]^{z_{++}}$$

$$= e^{(z_{++}p_{++})} \prod_{k=1}^{C} [w_{++}p_{++}]^{z_{++}}.$$  \hspace{1cm} (2.14)

Here, (2.14) will be equivalent to a multinomial likelihood when rewriting $p_{RC} = 1 - \sum_{jk \neq RC} P_{jk}$ so that $E(z_{++}) = z_{++}p_{++} = z_{++}$. As before, we need an offset in the part of the Poisson generalized linear model corresponding to (2.14).

Then, letting the $u_{jj}$’s, $w_{++}$’s, and $z_{++}$’s be independent Poisson random variables, we can maximize $L(p) = L_1(p)L_2(p)L_3(p)$ using a Poisson generalized linear model; in particular, a Poisson linear model, as we show in the following section. To constrain $E(u_{++}), E(w_{++}),$ and $E(z_{++})$ to equal $u_{++}, w_{++}$ and $z_{++}$, respectively, we must use an offset in the generalized linear model, which will be described next.

3. THE DESIGN MATRIX FOR THE POISSON LINEAR MODEL

Suppose we stack the $(RC \times 1)$ vector $u = [u_{11}, \ldots, u_{RC}]'$; the $(R \times 1)$ vector $w = [w_{1}, \ldots, w_{R+}]'$; and the $(C \times 1)$ vector $z = [z_{+1}, \ldots, z_{+C}]'$ to form the outcome vector

$$F = [u', w', z']',$$

whose elements are independent Poisson random variables. In this section, we show that the MLE for the $p_{jk}$’s can be obtained from a Poisson linear model for $F$ of the form

$$E(F) = Xp + \gamma,$$

where $\gamma$ is an offset, and the elements of the design matrix $X$ are easy to obtain.
Obtaining Maximum Likelihood Estimate

The model for the elements of \( \mathbf{u} \) is

\[
E(u_{jk}) = u_{++}p_{jk} \quad (jk \neq \text{RC});
\]

and

\[
E(u_{\text{RC}}) = u_{++} \left( 1 - \sum_{jk \neq \text{RC}} p_{jk} \right) = u_{++} - u_{++} \sum_{jk \neq \text{RC}} p_{jk},
\]

(3.2)

For example, if \( R = C = 3 \), then

\[
\begin{bmatrix}
  u_{11} \\
  u_{12} \\
  u_{13} \\
  u_{21} \\
  u_{22} \\
  u_{23} \\
  u_{31} \\
  u_{32} \\
  u_{33}
\end{bmatrix}
= u_{++}
\begin{bmatrix}
  1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1
\end{bmatrix}
\begin{bmatrix}
  p_{11} \\
  p_{12} \\
  p_{13} \\
  p_{21} \\
  p_{22} \\
  p_{23} \\
  p_{31} \\
  p_{32} \\
  u_{++}
\end{bmatrix}
\]

(3.3)

or

\[
E(\mathbf{u}) = (u_{++}A_1)p + \gamma_1,
\]

where, in general, the \( RC \times (RC - 1) \) matrix \( A_1 \) has its first \( RC - 1 \) rows equal to the \( RC - 1 \) identity matrix and the last row containing -1’s. Also, \( \gamma_1 \) is an \( (RC \times 1) \) “offset” vector with all elements 0 except the last, which equals \( u_{++} \). Finally, we write

\[
E(\mathbf{u}) = X_1p + \gamma_1,
\]

where \( X_1 = u_{++}A_1 \). The general partitioned matrix form of \( E(\mathbf{u}) \) is given in the appendix.

Next, consider the linear model for the subjects who have \( Y_{ij} \) observed, but are missing \( Y_{i2} \), the \( w_{j+} \)’s. In terms of the vector \( p \), we have

\[
E(w_{j+}) = w_{++}p_{j+} = w_{++} \sum_{k=1}^{C} p_{jk} \quad (j = 1, \ldots, R - 1);
\]

and

\[
E(w_{R+}) = w_{++} \left[ 1 - \sum_{j=1}^{R-1} p_{j+} \right] = w_{++} \left[ 1 - \sum_{j=1}^{R-1} \sum_{k=1}^{C} p_{jk} \right] = w_{++} - w_{++} \left[ \sum_{j=1}^{R-1} \sum_{k=1}^{C} p_{jk} \right],
\]
where \(w_{++}\) becomes an offset. When \(R = C = 3\) as in Table 1, we have

\[
\begin{bmatrix}
E(w_{1+}) \\
E(w_{2+}) \\
E(w_{3+})
\end{bmatrix} = w_{++} \begin{bmatrix}
p_{1+} \\
p_{2+} \\
p_{3+}
\end{bmatrix}
\]

\[
= w_{++} \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 & 0
\end{bmatrix}
\begin{bmatrix}
p_{11} \\
p_{12} \\
p_{13} \\
p_{21} \\
p_{22} \\
p_{23} \\
p_{31} \\
p_{32}
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
0 \\
0 \\
w_{++}
\end{bmatrix}
\]

or \(E(w) = (w_{++} + A_2) p + \gamma_2\). In general, the \(R \times (RC - 1)\) matrix \(A_2\) has first \((R - 1)\) rows and \((R - 1)C\) columns equal to a block diagonal matrix, with the blocks being a \((1 \times C)\) vector of 1’s. The last \((C - 1)\) columns of \(A_2\) are all zeroes. The last row of \(A_2\) is the negative of the sum of the first \(R - 1\) rows, and thus has first \((R - 1)C\) elements equal to -1, and last \((C - 1)\) elements equal to 0. Also, \(\gamma_2\) is an \((R \times 1)\) “offset” vector with all elements 0 except the last, which equals \(w_{++}\). Finally, we write

\[
E(w) = X_2 p + \gamma_2,
\]

where \(X_2 = w_{++} + A_2\). The general partitioned matrix form of \(E(w)\) is given in the appendix.

Finally, consider the linear model for the subjects who have \(Y_{i2}\) observed, but are missing \(Y_{i1}\), the \(z_{+k}\)’s. In terms of the vector \(p\), we have

\[
E(z_{+k}) = z_{++} p_{+k} = z_{++} \sum_{j=1}^{R} p_{jk} \quad (k = 1, \ldots, C - 1);
\]

and, similar to the above derivation for \(w_{R+}\), we have

\[
E(z_{+C}) = z_{++} \left[1 - \sum_{k=1}^{C-1} p_{+k}\right]
\]

\[
= z_{++} \left[1 - \sum_{j=1}^{R} \sum_{k=1}^{C-1} p_{jk}\right]
\]

\[
= z_{++} \left[1 - \sum_{k=1}^{C-1} p_{+k}\right] - z_{++} \left[\sum_{j=1}^{R} \sum_{k=1}^{C-1} p_{jk}\right],
\]

where \(z_{++}\) becomes an offset. When \(R = C = 3\) as in Table 1, we have
or $E(z) = (z_{++} A_3) p + \gamma_3$. In general, the first $C - 1$ rows of the $C \times (RC - 1)$ matrix $A_3$ horizontally concatenates $R$ identical matrices of the form

$$[I_{C-1} \ 0_{C-1}],$$

where $I_{C-1}$ is a $C - 1$ identity matrix, and $0_{C-1}$ is a $(C - 1) \times 1$ vector of 0’s. After concatenating these $R$ matrices, the last $(RC - 1)$th column is deleted. The last row of $A_3$ is the negative of the sum of the first $(C - 1)$ rows. Also, $\gamma_3$ is a $(C - 1)$ “offset” vector with all elements 0 except the last, which equals $z_{++}$. Finally, we write

$$E(z) = X_3 p + \gamma_3,$$

where $X_3 = z_{++} A_3$. The general partitioned matrix form of $E(z)$ is given in the appendix.

Then, in any generalized linear models program, we specify a Poisson error structure, a linear link, an outcome vector of the form

$$F = \begin{bmatrix} u \\ w \\ z \end{bmatrix},$$

a design matrix

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix},$$

and, finally, an offset

$$\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}.$$

Note that the design matrix $X$ does not contain an intercept.
Table 3. Data from Little and Rubin (1987, p. 183)

<table>
<thead>
<tr>
<th>Variable Y_{i1}</th>
<th>Variable Y_{i2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>Missing</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Missing</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>60</td>
</tr>
</tbody>
</table>

4. EXAMPLE

In order to illustrate our method, we use the (2 × 2) table with missing data found in Little and Rubin (1987), which is given here in Table 3. Little and Rubin (1987, p. 183), show the MLEs are $\hat{p}_{11} = .28$, $\hat{p}_{12} = .17$, $\hat{p}_{21} = .24$, and $\hat{p}_{22} = 0.31$. Using the results of the previous section, the Poisson linear model is

$$E = \begin{bmatrix} u_{11} \\ u_{12} \\ u_{21} \\ u_{22} \\ w_{1+} \\ w_{2+} \\ z_{+1} \\ z_{+2} \end{bmatrix} = \begin{bmatrix} u_{++} & 0 & 0 \\ 0 & u_{++} & 0 \\ 0 & 0 & u_{++} \\ -u_{++} & -u_{++} & -u_{++} \\ w_{++} & w_{++} & 0 \\ -w_{++} & -w_{++} & 0 \\ z_{++} & 0 & z_{++} \\ -z_{++} & 0 & -z_{++} \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{21} \\ p_{22} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ u_{++} \\ 0 \\ w_{++} \\ 0 \\ z_{++} \end{bmatrix},$$

where $u_{++} = 300$, $w_{++} = 90$, and $z_{++} = 88$. The SAS Proc Genmod commands and selected output are given in Table 4. The estimates of the $p_{jk}$'s in Table 4 agree with those given in Little and Rubin (1987).

Next, we use the S function glm to obtain the MLEs for the Six Cities data in Table 1. The models for $E(u)$, $E(w)$ and $E(z)$ are given in formulas (3.3), (3.4), and (3.5), respectively, with $u_{++} = 578$, $w_{++} = 507$, and $z_{++} = 103$. The commands and output are given in Table 5.

Recall that the MLEs are the same under missing at random or missing completely at random. However, when a generalized linear model program is used to estimate the $p_{jk}$'s, the inverse of the observed information matrix (i.e., the inverse of the negative of the Hessian matrix) should be used to estimate the variance. The inverse of the expected information is consistent only if the data are missing completely at random, whereas the inverse of the observed information is consistent under the weaker missing at random (Efron and Hinkley 1978).

5. EXTENSIONS

5.1 Multiway Tables

In this article, we have shown how any generalized linear models program can be used to obtain the MLEs of the cell probabilities in incomplete (R × C) tables. Other methods for obtaining the MLEs, such as the EM algorithm, require additional
programming steps. The method described here also easily extends to estimating the cell probabilities of multiway contingency tables. Although the notation gets more involved, the MLE can be calculated in any generalized linear models program with an offset and a Poisson error structure. For example, consider data from the Muscatine Coronary Risk Factor Study (Woolson and Clarke 1984), a longitudinal study to assess coronary risk factors in 4,856 school children. Children were examined in the years 1977, 1979, and 1981. The response of interest at each time is the binary variable obesity (obese, not obese). We let \( Y_{it} \) be the binary random variable (1=obese, 0=not obese) at time \( t \), \( t = 1 \) for 1977, \( t = 2 \) for 1979, and \( t = 3 \) for 1981. We are interested in estimating the joint probabilities

\[
p_{jkt} = \Pr[Y_{i1} = j, Y_{i2} = k, Y_{i3} = \ell],
\]

\( j, k, \ell = 0, 1 \). If there were no missing data, we would have a \((2 \times 2 \times 2)\) contingency table. Unfortunately, 3,086 (63.6%) of the subjects are observed at some subset of the three occasions. The data are given in Table 6.
Table 5. S function glm commands for Six Cities data in Table 1

<table>
<thead>
<tr>
<th>count</th>
<th>p11</th>
<th>p12</th>
<th>p13</th>
<th>p21</th>
<th>p22</th>
<th>p23</th>
<th>p31</th>
<th>p32</th>
<th>offs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>287</td>
<td>528</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>0</td>
<td>528</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>528</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>528</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>528</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>528</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>91</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>528</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>528</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>-528</td>
<td>-528</td>
<td>-528</td>
<td>-528</td>
<td>-528</td>
<td>-528</td>
<td>-528</td>
<td>528</td>
</tr>
<tr>
<td>10</td>
<td>279</td>
<td>507</td>
<td>507</td>
<td>507</td>
<td>507</td>
<td>507</td>
<td>507</td>
<td>507</td>
<td>507</td>
</tr>
<tr>
<td>11</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>59</td>
<td>103</td>
<td>0</td>
<td>0</td>
<td>103</td>
<td>0</td>
<td>0</td>
<td>103</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>18</td>
<td>0</td>
<td>103</td>
<td>0</td>
<td>0</td>
<td>103</td>
<td>0</td>
<td>0</td>
<td>103</td>
</tr>
<tr>
<td>15</td>
<td>26</td>
<td>-103</td>
<td>-103</td>
<td>0</td>
<td>-103</td>
<td>-103</td>
<td>0</td>
<td>-103</td>
<td>103</td>
</tr>
</tbody>
</table>

Call:

```r
glm(formula = counts -1 + p11 + p12 + p13 + p21 + p22 + p23 + p31 + p32 +
    offset(offs), family = poisson(link = identity))
```

Coefficients:

<table>
<thead>
<tr>
<th>p11</th>
<th>p12</th>
<th>p13</th>
<th>p21</th>
<th>p22</th>
<th>p23</th>
<th>p31</th>
<th>p32</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4747281</td>
<td>0.07005741</td>
<td>0.0741947</td>
<td>0.03273146</td>
<td>0.01195075</td>
<td>0.008743729</td>
<td>0.206066</td>
<td>0.05585518</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 15 Total; 7 Residual
Residual Deviance: 36.00067
Again, we assume the missing data are MAR. We let $a_{jk\ell}$ denote the number of subjects who are observed at all three times, with response level $j$ on $Y_{i1}$, level $k$ on $Y_{i2}$, and level $\ell$ on $Y_{i3}$; we let $b_{j+k+}$ denote the number of subjects with response level $j$ on $Y_{i1}$, level $k$ on $Y_{i2}$, and missing $Y_{i3}$; we let $c_{j+\ell+}$ denote the number of subjects with response level $j$ on $Y_{i1}$, level $\ell$ on $Y_{i3}$ and missing $Y_{i2}$; we let $d_{+k\ell+}$ denote the number of subjects with response level $k$ on $Y_{i2}$, level $\ell$ on $Y_{i3}$ and missing $Y_{i1}$; we let $e_{j++k}$ denote the number of subjects with response level $j$ on $Y_{i1}$, and missing $Y_{i2}$ and $Y_{i3}$; we let $f_{++k\ell}$ denote the number of subjects with response level $k$ on $Y_{i2}$, and missing $Y_{i1}$ and $Y_{i3}$; and we let $g_{++\ell+}$ denote the number of subjects with response level $\ell$ on $Y_{i3}$, and missing $Y_{i1}$ and $Y_{i2}$.

Suppose we stack the vectors
\[
a = [a_{000}, a_{001}, a_{010}, a_{011}, a_{100}, a_{101}, a_{110}, a_{111}]';
\]
\[
b = [b_{00+}, b_{01+}, b_{10+}, b_{11+}]';
\]
\[
c = [c_{0+0}, c_{0+1}, c_{1+0}, c_{1+1}]';
\]
\[
d = [d_{+00}, d_{+01}, d_{+10}, d_{+11}]';
\]
\[ e = [e_{0++}, e_{1++}]', \]
\[ f = [f_{0++}, f_{1++}]', \]
and
\[ g = [g_{0++}, g_{1++}]', \]
to form the outcome vector
\[ \mathbf{F} = [a_0, b_0, c_0, d_0, e_0, f_0, g_0]', \]
whose elements are independent Poisson random variables. Using similar ideas as in the previous sections, the MLE for the \( p_{jk}'s \) can be obtained from a Poisson linear model for \( \mathbf{F} \) of the form
\[ E(\mathbf{F}) = Xp + \gamma, \]
where \( \gamma \) is an offset, \( X \) is the design matrix, and \( p = [p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}] \). We delete \( p_{111} = (1 - p_{000} - p_{001} - p_{010} - p_{011} - p_{100} - p_{101} - p_{110}) \) from \( p \) since it is redundant. In particular, the model for \( a \) is
\[
E \begin{bmatrix}
  a_{000} \\
  a_{001} \\
  a_{010} \\
  a_{011} \\
  a_{100} \\
  a_{101} \\
  a_{110} \\
  a_{111}
\end{bmatrix} = a_{+++} \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1
\end{bmatrix} + \begin{bmatrix}
  p_{000} \\
  p_{001} \\
  p_{010} \\
  p_{011} \\
  p_{100} \\
  p_{101} \\
  p_{110} \\
  p_{111}
\end{bmatrix} ; \]
(5.1)
the model for \( b \) is
\[
E(\mathbf{b}_{00+}) = b_{+++} \begin{bmatrix}
  p_{00+} \\
  p_{01+} \\
  p_{10+} \\
  p_{11+}
\end{bmatrix} \]
\[
= b_{+++} \begin{bmatrix}
  1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
  -1 & -1 & -1 & -1 & -1 & -1 & -1 & 0
\end{bmatrix} \]
\[
+ \begin{bmatrix}
  p_{000} \\
  p_{001} \\
  p_{010} \\
  p_{011} \\
  p_{100} \\
  p_{101} \\
  p_{110} \\
  p_{111}
\end{bmatrix} ; \]
(5.2)
the model for $c$ is

\[
\begin{bmatrix}
E(c_{00}) \\
E(c_{01}) \\
E(c_{10}) \\
E(c_{11}) \\
\end{bmatrix}
= c_{+++} \begin{bmatrix}
p_{00} \\
p_{01} \\
p_{10} \\
p_{11} \\
\end{bmatrix}
\]

\[
= c_{+++} \begin{bmatrix}
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
-1 & -1 & -1 & -1 & 0 & 1 & -1 \\
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
0 \\
0 \\
0 \\
c_{+++} \\
\end{bmatrix} ; \quad (5.3)
\]

the model for $d$ is

\[
\begin{bmatrix}
E(d_{00}) \\
E(d_{01}) \\
E(d_{10}) \\
E(d_{11}) \\
\end{bmatrix}
= d_{+++} \begin{bmatrix}
p_{00} \\
p_{01} \\
p_{10} \\
p_{11} \\
\end{bmatrix}
\]

\[
= d_{+++} \begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 \\
-1 & -1 & -1 & -1 & 0 & -1 & -1 \\
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
0 \\
0 \\
0 \\
d_{+++} \\
\end{bmatrix} ; \quad (5.4)
\]

the model for $e$ is
\[
\begin{bmatrix}
E(e_{0++}) \\
E(e_{1++})
\end{bmatrix}
= e_{+++} \begin{bmatrix}
p_{00+} \\
p_{1++}
\end{bmatrix}
\]

\[
= e_{+++} \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 \\
-1 & -1 & -1 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
p_{000} \\
p_{001} \\
p_{010} \\
p_{011} \\
p_{100} \\
p_{101} \\
p_{110}
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
0 \\
e_{+++}
\end{bmatrix}.
\]

(5.5)

The model for \( f \) is
\[
\begin{bmatrix}
E(f_{00+}) \\
E(f_{11+})
\end{bmatrix}
= f_{+++} \begin{bmatrix}
p_{00+} \\
p_{11+}
\end{bmatrix}
\]

\[
= f_{+++} \begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 1 & 0 \\
-1 & -1 & 0 & 0 & -1 & -1 & 0
\end{bmatrix}
\begin{bmatrix}
p_{000} \\
p_{001} \\
p_{010} \\
p_{011} \\
p_{100} \\
p_{101} \\
p_{110}
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
0 \\
f_{+++}
\end{bmatrix}.
\]

(5.6)

and the model for \( g \) is
\[
\begin{bmatrix}
E(g_{000}) \\
E(g_{111})
\end{bmatrix}
= g_{+++} \begin{bmatrix}
p_{+++} \\
p_{++1}
\end{bmatrix}
\]

\[
= g_{+++} \begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
-1 & 0 & -1 & 0 & -1 & 0 & -1
\end{bmatrix}
\begin{bmatrix}
p_{000} \\
p_{001} \\
p_{010} \\
p_{011} \\
p_{100} \\
p_{101} \\
p_{110}
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
0 \\
g_{+++}
\end{bmatrix}.
\]

(5.7)

The SAS Proc Genmod commands for obtaining \( p \) and selected output are given in Table 7.

5.2 Other Models for \( p \)

The method can be used to obtain the MLE for any linear model for \( p \), in any generalized linear model program, without any additional programming. For example,
Table 7. SAS Proc GENMOD Commands and Selected Output for Data in Table 6

```plaintext
data one;
  input count p000 p001 p010 p011 p100 p101 p110 off margtot;
  p000 = p000*margtot; p001 = p001*margtot;
p010 = p010*margtot; p011 = p011*margtot;
p100 = p100*margtot; p101 = p101*margtot;
p110 = p110*margtot; p111 = p111*margtot;
  off = off*margtot;
cards;
1209 1 0 0 0 0 0 0 0 1770
  91 0 1 0 0 0 0 0 0 1770
  66 0 0 1 0 0 0 0 0 1770
  78 0 0 0 1 0 0 0 0 1770
  64 0 0 0 0 1 0 0 0 1770
  31 0 0 0 0 0 1 0 0 1770
  62 0 0 0 0 0 0 1 0 1770
  169 -1 -1 -1 -1 -1 -1 -1 1 1770
  426 1 1 0 0 0 0 0 0 631
  54 0 0 1 1 0 0 0 0 631
  33 0 0 0 0 1 1 0 0 631
  118 -1 -1 -1 -1 -1 -1 0 1 631
  125 1 0 1 0 0 0 0 0 184
  27 0 1 0 1 0 0 0 0 184
  5 0 0 0 0 1 0 1 0 184
  27 -1 -1 -1 -1 -1 0 -1 1 184
  463 1 0 0 0 1 0 0 0 645
  63 0 1 0 0 0 1 0 0 645
  37 0 0 1 0 0 0 1 0 645
  82 -1 -1 -1 0 -1 -1 -1 1 645
  583 1 1 1 1 1 0 0 0 756
  173 -1 -1 -1 -1 0 0 0 1 756
  293 1 1 0 0 1 1 0 0 370
  77 -1 -1 0 0 -1 -1 0 1 370
  381 1 0 1 0 1 0 1 0 500
  119 -1 0 -1 0 -1 0 -1 1 500
;

proc genmod data=one;
  model count = p000 p001 p010 p011 p100 p101 p110 /
    dist=poi link=id offset=off noint;
run;
```
Table 7. continued

/* SELECTED OUTPUT */

Analysis Of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Std Err</th>
<th>ChiSquare</th>
<th>Pr¿Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>P000</td>
<td>1</td>
<td>0.6633</td>
<td>0.0078</td>
<td>7190.7713</td>
<td>0.0001</td>
</tr>
<tr>
<td>P001</td>
<td>1</td>
<td>0.0578</td>
<td>0.0048</td>
<td>143.2562</td>
<td>0.0001</td>
</tr>
<tr>
<td>P010</td>
<td>1</td>
<td>0.0348</td>
<td>0.0037</td>
<td>86.5835</td>
<td>0.0001</td>
</tr>
<tr>
<td>P011</td>
<td>1</td>
<td>0.0439</td>
<td>0.0042</td>
<td>110.6773</td>
<td>0.0001</td>
</tr>
<tr>
<td>P100</td>
<td>1</td>
<td>0.0356</td>
<td>0.0039</td>
<td>83.4538</td>
<td>0.0001</td>
</tr>
<tr>
<td>P101</td>
<td>1</td>
<td>0.0207</td>
<td>0.0033</td>
<td>40.2450</td>
<td>0.0001</td>
</tr>
<tr>
<td>P110</td>
<td>1</td>
<td>0.0357</td>
<td>0.0039</td>
<td>82.6427</td>
<td>0.0001</td>
</tr>
<tr>
<td>SCALE</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

the marginal homogeneity model (Firth 1989) is a linear model for \( p \), say \( p = D \beta \). Then \( \beta \) can be calculated using any generalized linear model program without additional programming since the model for \( F \) is still linear

\[
E(F) = Xp + \gamma \\
= X(D\beta) + \gamma \\
= X^*\beta + \gamma,
\]

where \( X^* = XD \).

With additional programming in a generalized linear model program, our method can be used to fit a nonlinear model for \( p \). For example, log-linear models of the form \( p = \exp(D\beta) \), give rise to

\[
E(F) = Xp + \gamma \\
= X[\exp(D\beta)] + \gamma. \tag{5.8}
\]

Unfortunately, without additional programming, the generalized linear models programs discussed in the introduction only allow Poisson regression models of the form

\[
E(F) = g(A\beta + \gamma),
\]

for a design matrix \( A \), and where \( g(\cdot) \) is the linear, exponential, or inverse functions. Unfortunately, (5.8) does not fall in this class, and will require additional programming. To use a generalized linear models program, one must write a macro that calculates (5.8) as well as the derivative of (5.8) with respect to \( \beta \). Alternatively, a maximization procedure based on ideas similar to this article is discussed in Molenberghs and Goetghebeur (1997).
APPENDIX

In this appendix, we give the partitioned matrix form for the models for $E(u)$, $E(w)$, and $E(z)$. We define the following notation: $I_a$ is an $a \times a$ identity matrix, $1_a$ is an $a \times 1$ vector of 1's, $0_{a,b}$ is an $a \times b$ matrix of 0's, and $\otimes$ is the “direct product.”

In partitioned matrix form, the models are

$$E(u) = u_{++} \begin{bmatrix} I_{RC-1} \\ -I_{RC-1} \end{bmatrix} p + \begin{bmatrix} 0_{RC-1,1} \\ u_{++} \end{bmatrix};$$

$$E(w) = w_{++} \begin{bmatrix} I_{R-1} \otimes 1_C \\ -I_{(R-1)C} \end{bmatrix} \begin{bmatrix} 0_{R-1,C-1} \\ 0_{1,C-1} \end{bmatrix} p + \begin{bmatrix} 0_{R-1,1} \\ w_{++} \end{bmatrix};$$

and

$$E(z) = z_{++} \begin{bmatrix} 1_R \otimes \begin{bmatrix} I_{C-1} \\ -I_{C-1} \end{bmatrix} \begin{bmatrix} 0_{C-1,1} \\ 0 \end{bmatrix} \end{bmatrix} \begin{bmatrix} I_{RC-1} \\ 0_{1,RC-1} \end{bmatrix} p + \begin{bmatrix} 0_{C-1,1} \\ z_{++} \end{bmatrix}.$$

ACKNOWLEDGMENTS

We are very grateful for the support provided by grants CA 57253 and CA 55576 from the NIH (U.S.A.). The first author would like to thank the faculty at L.U.C. who were extremely helpful during his stay as visiting professor.

[Received March 1996. Revised November 1997.]

REFERENCES


