ANÁLISE DE DADOS CATEGORIZADOS COM OBSERVAÇÕES OMISSAS

ANALYSIS OF MISSING CATEGORICAL DATA

Autor: Carlos Daniel Paulino

Professor Associado do Departamento de Matemática e membro do Centro de Matemática Aplicada do Instituto Superior Técnico

SUMÁRIO:

- Este artigo debruça-se sobre o problema da análise de dados categorizados quando os dados incluem unidades cuja resposta a todas as variáveis não é integralmente conhecida. Pretende-se nele rever e desenvolver a aplicação da abordagem por máxima verosimilhança baseada num modelo amostral Multinomial munido de uma estrutura não informativa para o processo de não-resposta (ou de censura). Os resultados dessa aplicação são descritos em formulação matricial apropriada para a sua implementação computacional de uma forma independente da configuração da tabela de contingência e do padrão de incompletude dos dados.

PALAVRAS-CHAVE:

- *Distribuição Multinomial, dados aleatoriamente omissos, censura ignorável, censura informativa, metodologia de máxima verosimilhança, modelos lineares, modelos log-lineares.*

ABSTRACT:

- This is an article dealing with the analysis of categorical data when some sample units are not fully classified. It is intended to review and develop the application of the maximum likelihood approach when based upon a Multinomial sampling model provided with a non-informative structure for the non-response (or censoring) process. The results of such application are described in matrix terms suitable for their computational implementation regardless of the contingency table shape and the data incompleteness pattern.

KEY-WORDS:

- *Multinomial distribution, data missing at random, ignorable censoring, informative censoring, maximum likelihood methodology, strictly linear models, log-linear models.*

1. DESCRIÇÃO DO PROBLEMA

Consideremos um processo de classificação de n unidades amostrais segundo variáveis categorizadas cuja distribuição conjunta se representa numa tabela de contingência com m celas. Razões várias fazem com que por vezes tal processo amostral seja incompletamente observado ou registado no sentido em que nem todas as unidades aparecem perfeitamente classificadas segundo cada uma das variáveis.

As unidades com alguma classificação omissa (também denominada censurada) surgem registadas em classes de pelo menos duas celas da tabela que agrupamos no conjunto Δ_c. O conjunto $\Pi_1 = \{\{i\}, i = 1,...,m\}$ agrupa as classes em que se registam as unidades completamente categorizadas, constituindo pois a partição mais fina do conjunto das m celas da tabela $\{1,...,m\}$.

Vamos admitir que o conjunto Δ_c é formado por $T-1$ partições deste conjunto, Π_t, $t = 2,\ldots,T$, sem elementos comuns entre elas, onde se toma $\Pi_t = \{C_{tj}, j = 1,\ldots,m_t\}$. Este padrão de registo ocorre quando há omissão da classificação em uma ou mais variáveis como acontece frequentemente.

Por conveniência matemática, vamos representar por $\mathbf{Z} = [\mathbf{Z}_t, t = 1,\ldots,t]$ a matriz particionada $m \times (m+l)$, $l = \sum_{t=2}^{T} m_t$, onde cada submatriz $\mathbf{Z}_t, t = 1,\ldots,t$, é representada por $\mathbf{Z}_t = [\mathbf{z}_{tj}, j = 1,\ldots,m_t]$ em que $\mathbf{z}_{tj}$ é um vector $m \times 1$ indicando as categorias que pertencem à classe C_{tj} através de uns, sendo as restantes componentes nulas. Note-se que $m_1 = m$, $C_{1j} = \{j\}$ e $\mathbf{Z}_1 = \mathbf{I}_m$ (matriz identidade de ordem m). Consistentemente com o exposto, os dados observáveis são as frequências registáveis no conjunto $\Delta = \bigcup_{t=1}^{T} \Pi_t$, denotadas por $\mathrm{N} = (\mathrm{N}_t, t = 1,\ldots,t)$, em que $\mathrm{N}_t = (n_{tj}, j = 1,\ldots,m_t)$.

A análise deste tipo de dados visa geralmente os mesmos objectivos que no caso do problema padrão de dados completos em que não existem observações omissas nem se admite a possibilidade da sua ocorrência. Ou seja, pretende-se traçar inferências sobre a distribuição conjunta das variáveis categorizadas definidoras da tabela. Para o efeito, necessita-se de estabelecer um modelo estatístico capaz de explicar os dados observados, sendo tal o objecto da Secção 2.

Na Secção 3 avançar-se-á com o ajustamento de modelos de omissão não-informativa pela metodologia de máxima verosimilhança, preparando o terreno para a análise de modelos estruturais estritamente lineares e log-lineares a desenvolver na Secção 4. A última secção ilustra a aplicação dos métodos descritos através de dois exemplos.

2. MODELAÇÃO

O primeiro pressuposto é o de que as n unidades amostrais foram seleccionadas segundo um processo de amostragem aleatória simples de uma dada população com dimensão suposta suficientemente grande. Deste modo, considera-se que o vector $\mathbf{y} = (y_i, i = 1,\ldots, m)$ de frequências hipotéticas que resultariam da classificação completa de todas as unidades possui uma distribuição amostral Multinomial com parâmetro probabilístico caracterizando a distribuição conjunta da tabela denotado por $\theta = (\theta_1,\ldots,\theta_m)$, $\sum_i \theta_i = 1$.

Admitamos para a modelação do processo de registo (ou de censura) que para toda a cela i, as y_i observações fictícias se repartem por todas as classes de Δ contendo i de acordo igualmente com um processo Multinomial de parâmetros $\lambda_{ti}, t = 1,\ldots,T$ tais que $\sum_{t=1}^{T} \lambda_{ti} = 1$. Denotando por $\{y_{ti}\}$ as frequências desdobradas, tem-se então que a distribuição amostral conjunta de $(y_{ti};\ t = 1,\ldots,T, i = 1,\ldots,m)$ condicional em $\mathbf{y}$ é um produto de distribuições Multinomiais independentes. Consequentemente, a distribuição amostral conjunta de $\{y_{ti}\}$ é uma distribuição Multinomial com parâmetros $\{\theta_i \lambda_{ti}\}$. Mas, à excepção de $\{y_{1i} = n_{1i}, i = 1,\ldots,m\}$, estas frequências não são observáveis. As frequências observáveis resultam destas frequências fictícias por agrupamento,

$$n_{tj} = \sum_{i \in C_{tj}} y_{ti},\ j = 1,\ldots,m_t,\ \ t = 2,\ldots,T\ .$$

Por conseguinte, a distribuição amostral conjunta de **N** é descrita por um modelo Multinomial com função de probabilidade

$$f\left(\mathrm{N} \mid \{\theta_i, \lambda_{ti}\}\right) = \frac{n!}{\prod_{t=1}^{T}\prod_{j=1}^{m_t} n_{tj}!} \prod_{i=1}^{m} (\theta_i \lambda_{1i})^{n_{1i}} \prod_{t=2}^{T}\prod_{j=1}^{m_t}\left(\sum_{i \in C_{tj}} \theta_i \lambda_{ti}\right)^{n_{tj}}\ .$$

Este modelo, naturalmente sobreparametrizado (há mT parâmetros para apenas $m+l$ observações), padece previsivelmente de falta de identificabilidade, inviabilizando a realização de inferências que não recorram a informação extra. O processo de identificação do modelo mais comummente considerado consiste na imposição externa de restrições nos parâmetros do processo de registo de modo que a probabilidade condicional de registo em cada partição Π_t seja a mesma para todas as categorias reunidas em cada classe de Π_t, isto é,

$$M_1 : \lambda_{ti} = \lambda_{tk},\ i,k \in C_{tj},\ j = 1,\ldots,m_t,\ \ t = 2,\ldots,T\ .$$

Este processo M_1 é rotulado na literatura por censura não-informativa, ou ainda, segundo Rubin (1976), por censura MAR (acrónimo de *Missing At Random*). Observe-se que o modelo estatístico sob M_1 passa a ser saturado, sendo a respectiva verosimilhança expressável numa forma factorizada por

$$\begin{aligned} L\left(\theta,\left\{\lambda_{tj}\right\}\middle|\ \mathrm{N};M_1\right) &\propto \prod_{t=1}^{T}\prod_{j=1}^{m_t}\lambda_{tj}^{n_{tj}} \times \prod_{i=1}^{m}\theta_i^{\,n_{1i}} \prod_{t=2}^{T}\prod_{j=1}^{m_t}\left(\mathbf{z}_{tj}'\theta\right)^{n_{tj}} \equiv \\ &\equiv L_1\left(\left\{\lambda_{tj}\right\}\middle|\ \mathrm{N}\right) \times L_2\left(\theta\,\middle|\ \mathrm{N}\right) \end{aligned}$$

onde λ_{tj} denota a probabilidade condicional de registo em C_{tj}.

Note-se que esta factorização implica em caso de ausência de qualquer relação entre θ e $\{\lambda_{tj}\}$ (e. g., sob independência *a priori*) que é L_2 o único factor relevante para as inferências sobre θ que respeitam o Princípio da Verosimilhança. Daí que se diga que o mecanismo de registo é ignorável do ponto de vista das inferências verosimilhancistas (em particular, bayesianas) sobre θ.

Um caso especial de um mecanismo de censura MAR resulta da imposição mais restritiva de uma única probabilidade condicional de censura por cada tipo de censura, i.e., por cada Π_t, sendo pois definido por

$$M_2 : \lambda_{ti} = \alpha_t,\ i = 1,\ldots,m,\ \ t = 1,\ldots,T\ .$$

Este processo é por vezes rotulado por censura MCAR (acrónimo de *Missing Completely At Random*) e, contrariamente a M_1, implica pelo Princípio da Condicionalidade Generalizado (para detalhes vide Paulino, 1991) que as inferências frequencistas sobre θ podem basear-se apenas na distribuição de núcleo L_2, que é agora a distribuição condicional de $\{\mathrm{N}_t\}$ dados os totais $N_{t.} = \sum_{j=1}^{m_t} n_{tj}$ (Produto de T Multinomiais independentes, $M(N_{t.}, \mathbf{z}_t'\theta)$).

Grande parte das análises descritas na literatura assumem compreensivelmente um mecanismo de censura não-informativa ou ignorável do ponto de vista das inferências baseadas ou na verosimilhança (M_1) ou na teoria da amostragem (M_2). Isto é claramente patente no livro de Little and Rubin (1987) e no artigo de revisão sobre métodos frequencistas (máxima verosimilhança e mínimos quadrados generalizados) de Paulino (1991) - uma revisão suplementar mais recente (em português) do uso destes métodos encontra-se na dissertação de Rodrigues (1996).

As análises baseadas em mecanismos de censura informativa ou não ignorável do ponto de vista frequencista apoiam-se em outras estruturas sobre $\{\lambda_{ti}\}$ de modo a manter a sua dependência de classificações não observadas, ou então na fixação de valores consentâneos para $\{\lambda_{ti}\}$ com o eventual propósito de estudar a sua influência nas inferências sobre θ, no espírito de uma análise de sensibilidade. Referências recentes deste tipo de análises incluem Baker and Laird (1988), Conaway (1993), Chambers and Welsh (1993), Park and Brown (1994) e Baker (1994, 1996).

Do ponto de vista bayesiano, Dickey et al, (1987), por um lado, e Forster and Smith (1998), por outro, são artigos recentes que ilustram bem a aplicação de métodos bayesianos nos dois cenários acima indicados para a modelação do mecanismo de censura. Fazendo uso da maior flexibilidade do paradigma bayesiano em acomodar modelos inidentificáveis, Paulino and Pereira (1992, 1995), Walker (1996) e Soares and Paulino (2000) descrevem métodos apropriados para as inferências de interesse sem imporem restrições identificantes nas probabilidades condicionais de censura.

Nas secções seguintes concentrar-nos-emos em descrever e desenvolver a metodologia de máxima verosimilhança de modo a torná-la directamente aplicável a qualquer problema de dados incompletamente categorizados na forma atrás indicada, e supostamente explicáveis pelo modelo Multinomial munido da estrutura não-informativa MAR. O desenvolvimento do mesmo tipo de análise para o problema análogo de observação incompleta de processos poissonianos está em fase de ultimação (veja-se, entretanto, Paulino, 2000 e Paulino e Soares, 2000).

3. ANÁLISE SOB OMISSÃO ALEATÓRIA

Em tudo o que segue admitimos a validade do mecanismo de censura não-informativa M_1, sob o qual **N** apresenta a distribuição amostral $M(n,\{(\mathbf{z}'_{tj}\theta)\lambda_{tj}\})$ onde $\{\lambda_{tj}\}$ denotam agora as probabilidades condicionais de registo em todas as classes C_{tj}. Para a estimação por máxima verosimilhança de θ basta maximizar o factor L_2, o que conduz ao sistema de equações de verosimilhança

$$\hat{\theta} = \mathrm{N}_1 + \sum_{t=2}^{T} \mathbf{D}_{\hat{\theta}} \mathbf{Z}_t \mathbf{D}^{-1}_{\mathbf{Z}_t'\hat{\theta}} \mathrm{N}_t \, ,$$

não sendo difícil verificar que a expressão dada corresponde à forma usual de apresentação do algoritmo EM (a da igualdade das esperanças, condicional em **N** e não condicional, das frequências fictícias **y**).

Como meio de tornear a lentidão do simples algoritmo EM pode-se optar pelos algoritmos alternativos de Newton-Raphson e "scoring" de Fisher, para o que se necessita de começar por determinar o vector gradiente do logaritmo da verosimilhança L_2 que exprimimos em termos de $\bar{\theta} = (\mathbf{I}_{m-1},\ \mathbf{0}_{m-1})\theta$, i. e., de todas as componentes de θ à excepção da última. Denotando por $\overline{\mathbf{Z}}_t$ a matriz $(m-1)\times(m_t-1)$ obtida de $\mathbf{Z}_t$ por remoção da última linha e coluna e definindo $\bar{\theta}_t = \overline{\mathbf{Z}}_t'\bar{\theta}$ e $\overline{\mathbf{Z}} = \left[\overline{\mathbf{Z}}_t, t=1,\ldots,T\right]$, resulta que a função *score* é

$$\mathbf{U}_2(\bar{\theta}) = \sum_{t=1}^{T} \overline{\mathbf{Z}}_t \left[\mathbf{\Sigma}_t(\bar{\theta})\right]^{-1} \bar{\mathbf{p}}_t \; - \left(\overline{\mathbf{Z}}\left[\mathbf{\Sigma}(\bar{\theta})\right]^{-1} \overline{\mathbf{Z}}'\right)\bar{\theta}$$

onde $\bar{\mathbf{p}}_t = (\mathbf{I}_{m_t-1},\ \mathbf{0}_{m_t-1})\mathbf{N}_t / N_{t.}$, $\mathbf{\Sigma}_t(\bar{\theta}) = \left(\mathbf{D}_{\bar{\theta}_t} - \bar{\theta}_t\, \bar{\theta}_t'\right)\big/ N_{t.}$, $t = 1,\ldots,T$ e $\mathbf{\Sigma}(\bar{\theta})$ é a matriz diagonal em blocos com blocos diagonais $\mathbf{\Sigma}_t(\bar{\theta})$. Segue-se então que a matriz hessiana de $\ln L_2(\bar{\theta} \,|\, \mathbf{N})$ pode ser compactamente representada por:

$$\mathbf{H}_2(\bar{\theta}) = -\sum_{t=1}^{T} \overline{\mathbf{Z}}_t \left[\mathbf{D}_{\mathbf{v}_t} + \frac{n_{tm_t}}{(1-\mathbf{J}_t'\bar{\theta}_t)^2}\mathbf{J}_t\mathbf{J}_t'\right] \overline{\mathbf{Z}}_t'$$

onde $\mathbf{v}_t = \left(n_{tj} / (\mathbf{z}_{tj}'\theta)^2,\ j = 1,\ldots,m_t-1\right)'$ e $\mathbf{J}_t$ é o vector $(m_t-1)\times 1$ de uns.

O algoritmo *scoring* de Fisher exige a estimação adicional dos parâmetros $\lambda_t = \left(\lambda_{t1},\ldots,\lambda_{tm_t}\right), t = 1,\ldots,T$ pois $E\left(n_{tj} \middle| \theta, \{\lambda_{tj}\}, M_1\right) = n\lambda_{tj}\mathbf{z}_{tj}'\theta,\ \forall t, j$. Por maximização restringida de $\ln L_1(\lambda(M_1) \,|\, \mathbf{N})$, onde $\lambda(M_1) = \left(\lambda_t, t = 1,\ldots,T\right)$, ou mais simplesmente, atendendo ao facto de o modelo estatístico sob M_1 ser saturado, as estimativas de máxima verosimilhança das probabilidades condicionais de registo são $\hat{\lambda}(M_1) = \left(\hat{\lambda}_t, t = 1,\ldots,T\right)$, $\hat{\lambda}_t = \left(\hat{\lambda}_{t1},\ldots,\hat{\lambda}_{tm_t}\right) = n^{-1}\mathbf{D}^{-1}_{\mathbf{Z}_t'\hat{\theta}}\mathbf{N}_t$. Deste modo, obtém-se como matriz de covariância assintótica estimada de $\hat{\bar{\theta}}$ a matriz $\left(-\mathbf{H}_2(\hat{\bar{\theta}})\right)^{-1}$.

A forma do factor L_1 da verosimilhança sob M_2

$$L_1\left(\{\lambda(M_2)\}\middle|\mathbf{N};M_2\right)=\prod_{t=1}^{T}\alpha_t^{N_{t.}}$$

conduz a que a estimativa de máxima verosimilhança dos parâmetros desse tipo de processo de censura seja $\hat{\alpha}_t = N_{t.}/N_{..}$, $t=1,\ldots,T$ com $N_{..} = \sum_{t=1}^{T} N_{t.}$. Deste modo, o teste da razão de verosimilhanças de Wilks de ajustamento do modelo especial de censura não-informativa M_2 é definido pela estatística

$$Q_V\left(M_2|M_1\right)=-2\ln\frac{L_1\left(\hat{\alpha}_t\mid\mathbf{N};M_2\right)}{L_1\left(\hat{\lambda}_{tj}\mid\mathbf{N};M_1\right)}$$
$$=-2\sum_{t=1}^{T}\sum_{j=1}^{m_t} n_{tj}\left[\ln(\mathbf{z}'_{tj}\hat{\theta})-\ln(n_{tj}/N_{t.})\right]$$

com distribuição nula aproximada $\chi^2_{(s)}$, $s=l-T+1$. Em alternativa, o teste de Pearson para o ajustamento de M_2 apoia-se na estatística, assintoticamente equivalente a Q_V ,

$$Q_P\left(M_2|M_1\right)=\sum_{t=1}^{T}\sum_{j=1}^{m_t}\frac{\left(n_{tj}-N_{t.}\mathbf{z}'_{tj}\hat{\theta}\right)^2}{N_{t.}\mathbf{z}'_{tj}\hat{\theta}}.$$

Querendo testar o ajustamento de qualquer outro modelo particularizando a censura MAR basta aplicar *mutatis mutandi* o procedimento acabado de descrever.

Seja agora $M\subset M_1$ um modelo arbitrário mais restritivo que M_1 para o vector λ das probabilidades condicionais de censura MAR e *H* um modelo reduzido para o vector θ das probabilidades das celas. A estatística de teste do modelo conjunto $M\cap H$ sob M_1 pelo critério da razão das verosimilhanças de Wilks é desdobrável na soma das correspondentes estatísticas de teste de *M* e de *H*, i.e.,

$$Q_V\left(M\cap H\mid M_1\right)=-2\ln\frac{L_1\left(\hat{\lambda}(M)|\mathbf{N}\right)L_2\left(\hat{\theta}(H)\mid\mathbf{N}\right)}{L_1\left(\hat{\lambda}\mid\mathbf{N}\right)L_2\left(\hat{\theta}\mid\mathbf{N}\right)}$$
$$=Q_V\left(M\mid M_1\right)+Q_V\left(H\mid M_1\right)$$

onde $\hat{\lambda}(M)$ e $\hat{\theta}(H)$ denotam as estimativas de máxima verosimilhança de λ e θ restringidas a *M* e *H*, respectivamente. Como notaram Williamson and Haber (1994), este particionamento de Q_V mostra que a comparação por esse critério de qualquer par de modelos para as probabilidades das categorias (respectivamente, probabilidades condicionais de registo) não depende da estrutura tão ou mais reduzida que se imponha ao processo de registo (respectivamente, processo de amostragem Multinomial).

Se o parâmetro de interesse for apenas θ e se se pretende testar pelo critério Q_V o ajustamento de um modelo estrutural *H* para ele, a estrutura apropriada é

$$Q_V\left(H \mid M_1\right) = -2\ln\frac{L_2\left(\hat{\theta}(H) \mid \mathbf{N}\right)}{L_2\left(\hat{\theta} \mid \mathbf{N}\right)}$$

sendo independente do modelo M mais restrito que M_1 (e.g., o modelo de censura MCAR M_2) que se queira aceitar como válido. A este respeito observe-se que o recurso à estatística de Pearson,

$$Q_P\left(H \mid M_1\right) = \sum_{t=1}^{T}\sum_{j=1}^{m_t}\frac{\left[n_{tj} - n\hat{\lambda}_{tj}\mathbf{z}'_{tj}\hat{\theta}(H)\right]^2}{n\hat{\lambda}_{tj}\mathbf{z}'_{tj}\hat{\theta}(H)}$$

$$= \sum_{t=1}^{T}\sum_{j=1}^{m_t}\frac{n_{tj}}{\mathbf{z}'_{tj}\hat{\theta}}\frac{\left[\mathbf{z}'_{tj}\left(\hat{\theta} - \hat{\theta}(H)\right)\right]^2}{\mathbf{z}'_{tj}\hat{\theta}(H)}$$

já não proporciona o mesmo vantajoso resultado pois a sua expressão depende claramente de M através das respectivas estimativas de $\boldsymbol{\lambda}$. Querendo testar simultaneamente o ajuste de M_2 e H pelo critério Q_P a estatística a usar é

$$Q_P\left(M_2 \cap H \mid M_1\right) = \sum_{t=1}^{T}\sum_{j=1}^{m_t}\frac{\left[n_{tj} - N_{t.}\mathbf{z}'_{tj}\hat{\theta}(H)\right]^2}{N_{t.}\mathbf{z}'_{tj}\hat{\theta}(H)} .$$

4. CONCRETIZAÇÃO A MODELOS LINEARES E LOG-LINEARES

Para uma concretização adicional de resultados anteriores vamos considerar para H algumas das estruturas habituais em dados categorizados, a saber, os modelos estritamente lineares e log-lineares no vector das probabilidades θ.

O modelo estrutural estritamente linear é formulado em termos de equações livres por

$$H : \overline{\theta} = \mathbf{X}\beta$$

onde $\mathbf{X}$ é uma matriz especificada $(m-1)\times p$ de característica completa $r(\mathbf{X}) = p < m$. A incorporação desta estrutura no logaritmo da verosimilhança $L_2\left(\overline{\theta} \middle| \mathbf{N}\right)$ e a sua diferenciação em ordem a β conduzem à função *score* $\mathbf{U}_{2L}(\beta) = \mathbf{X}'\,\mathbf{U}_2(\overline{\theta}(\beta))$ e às seguintes equações de verosimilhança

$$\mathbf{X}'\sum_{t=1}^{T}\overline{\mathbf{Z}}_t\mathbf{D}^{-1}_{\overline{\mathbf{Z}}'_t\overline{\theta}(\hat{\beta})}\overline{\mathbf{N}}_t = \mathbf{X}'\sum_{t=1}^{\mathrm{T}}\overline{\mathbf{Z}}_t\{1 - \mathbf{J}'_1\overline{\theta}(\hat{\beta})\}^{-1}n_{tm_t}\mathbf{J}'_t ,$$

onde $\overline{\mathbf{N}}_t = \overline{\mathbf{p}}_t N_{t.}$.

A matriz hessiana associada está relacionada com a do correspondente modelo saturado pela expressão

$$\mathbf{H}_{2L}(\beta) = \mathbf{X}'\,\mathbf{H}_2(\overline{\theta}(\beta))\,\mathbf{X}$$

e, por conseguinte, a correspondente matriz de informação de Fisher é

$$\mathbf{I}_{2L}(\beta,\lambda) = \mathbf{X}'\sum_{t=1}^{T}\overline{\mathbf{Z}}_t\{\mathbf{D}_{\mathbf{u}} + [1-\mathbf{J}_t'\,\overline{\mathbf{Z}}_t'\overline{\theta}(\beta)]^{-1}\lambda_{tm_t}\mathbf{J}_t\mathbf{J}_t'\}\,\overline{\mathbf{Z}}_t'\,\mathbf{X}$$

onde $\mathbf{u}_t = \mathbf{u}_t(\beta,\lambda) = \left(n\lambda_{tj}/\mathbf{z}_{tj}'\theta(\beta),\ j=1,\ldots,m_t\right)$.

O uso de uma destas matrizes no correspondente algoritmo de Newton-Raphson ou "scoring" de Fisher conduz-nos à determinação da estimativa de máxima verosimilhança $\hat{\beta}$ de β, da qual se obtêm os valores observados das estatísticas de teste atrás citadas através das estimativas de máxima verosimilhança $\hat{\overline{\theta}}(H) = \mathbf{X}\hat{\beta}$. A distribuição nula assintótica de $Q_V(M_2 \cap H \mid M_1)$ e de $Q_P(M_2 \cap H \mid M_1)$ é $\chi^2_{(s+m-1-p)}$ enquanto a de $Q_V(H \mid M_1)$ e $Q_P(H \mid M_1)$ é $\chi^2_{(m-1-p)}$. Pelo método delta, conclui-se que a matriz de covariância aproximada de $\hat{\overline{\theta}}(H)$ é

$$V\hat{a}r\left[\hat{\overline{\theta}}(H)\right] = \mathbf{X}\left[\mathbf{I}_{2L}(\hat{\beta},\hat{\lambda})\right]^{-1}\mathbf{X}'.$$

Saliente-se ainda que, não havendo interesse em β nem no estudo de eventuais reduções do modelo *H*, o ajustamento de *H* pode processar-se alternativamente através de equivalente formulação em restrições de *H*, i.e.,

$$H: \mathbf{C}\overline{\theta} = \mathbf{0}$$

onde $\mathbf{C}$ é uma matriz $a \times (m-1)$ de característica completa $a = m-1-p$ cujas linhas são ortogonais às colunas de $\mathbf{X}$. Com efeito, a maximização restringida de $\ln L_2(\overline{\theta} \mid \mathbf{N})$ pelo método dos multiplicadores lagrangianos conduz-nos ao sistema de equações

$$\hat{\overline{\theta}}(H) = \left\{\mathbf{I}_{m-1} - \hat{\mathbf{V}}\mathbf{C}'\left[\mathbf{C}\hat{\mathbf{V}}\mathbf{C}'\right]^{-1}\mathbf{C}\right\}\hat{\mathbf{V}}\sum_{t=1}^{T}\overline{\mathbf{Z}}_t\,\hat{\mathbf{\Sigma}}_t\,\overline{\mathbf{p}}_t,$$

onde $\hat{\mathbf{V}} = \left[\overline{\mathbf{Z}}\left[\mathbf{\Sigma}(\hat{\overline{\theta}}(H))\right]^{-1}\overline{\mathbf{Z}}'\right]^{-1}$ e $\hat{\mathbf{\Sigma}}_t = \mathbf{\Sigma}_t(\hat{\overline{\theta}}(H))$, passível de resolução iterativa directa em ordem a $\hat{\overline{\theta}}(H)$.

O modelo estrutural log-linear para as probabilidades das celas é compactamente definido por

$$H: \ln\theta = \mathbf{1}_m\eta + \mathbf{X}\beta$$

onde $\mathbf{X}$ é uma matriz $m \times p$ tal que a característica de $[\mathbf{1}_m, \mathbf{X}]$ é $1 + p$. Com a inclusão da restrição natural $\mathbf{1}'\theta = 1$, esta estrutura pode ser expressa por

$$\theta(\beta) = \exp(\mathbf{X}\beta) / \mathbf{1}'_m \exp(\mathbf{X}\beta) .$$

Diferenciando $\ln L_2(\theta(\beta) \mid \mathbf{n})$ em ordem a β obtemos a função *score*

$$\mathbf{U}_{2LL}(\beta) = \mathbf{X}' \left\{ \mathbf{N}_1 + \sum_{t=2}^{T} \mathbf{D}_{\theta(\beta)} \mathbf{Z}_t \mathbf{D}^{-1}_{\mathbf{Z}'_t \theta(\beta)} \mathbf{N}_t \right\} - \mathbf{X}'\mu(\beta)$$

onde $\mu(\beta) = N\theta(\beta)$, pelo que o sistema de equações de verosimilhança é

$$\mathbf{X}' \left\{ \mathbf{N}_1 + \sum_{t=2}^{T} \mathbf{D}_{\theta(\beta)} \mathbf{Z}_t \mathbf{D}^{-1}_{\mathbf{Z}'_t \theta(\beta)} \mathbf{N}_t \right\} = \mathbf{X}'\mu(\beta) .$$

A forma destas equações mostra que a sua resolução iterativa em ordem a β ou a θ pode ser feita via algoritmo EM onde o passo M pode exigir, dependendo de **X**, o recurso a outro algoritmo à semelhança do que se passa com dados completos (vide Paulino, 1991).

Procedendo à diferenciação adicional em ordem a β' do vector gradiente de $\ln L_2$ obtém-se a matriz hessiana

$$\mathbf{H}_{2LL}(\beta) = \mathbf{X}' \left\{ -n\, \mathbf{I}_m + \sum_{t=2}^{T} \left[\mathbf{D}_{\mathbf{a}_t(\beta)} - \mathbf{D}_{\mathbf{b}_t(\beta)} \mathbf{Z}_t \mathbf{Z}'_t \right] \right\} \left\{ \mathbf{D}_{\theta(\beta)} - \theta(\beta)[\theta(\beta)]' \right\} \mathbf{X}$$

onde $\mathbf{a}_t(\beta) = \mathbf{Z}_t \mathbf{D}^{-1}_{\mathbf{Z}'_t \theta(\beta)} \mathbf{N}_t$, e $\mathbf{b}_t(\beta) = \mathbf{D}_{\theta(\beta)} \mathbf{Z}_t \mathbf{D}^{-2}_{\mathbf{Z}'_t \theta(\beta)} \mathbf{N}_t$ que, juntamente com a função *score*, permitem obter iterativamente a estimativa de máxima verosimilhança $\hat{\beta}$ pelo método de Newton-Raphson. Optando pelo algoritmo "scoring" de Fisher, necessitamos da matriz de informação de Fisher que é expressável por

$$\mathbf{I}_{2LL}(\beta, \lambda) = \mathbf{X}' \left\{ n\, \mathbf{I}_m - \sum_{t=2}^{T} \left[\mathbf{D}_{\mathbf{A}_t(\beta,\lambda)} - \mathbf{D}_{\mathbf{B}_t(\beta,\lambda)} \mathbf{Z}_t \mathbf{Z}'_t \right] \right\} \left\{ \mathbf{D}_{\theta(\beta)} - \theta(\beta)[\theta(\beta)]' \right\} \mathbf{X}$$

onde

$$\begin{aligned} \mathbf{A}_t(\beta, \lambda) &\equiv E\left(\mathbf{a}_t(\beta) \middle| \beta, \{\lambda_{tj}\}; M_1 \right) = n \mathbf{Z}_t \lambda_t \\ \mathbf{B}_t(\beta, \lambda) &\equiv E\left(\mathbf{b}_t(\beta) \middle| \beta, \{\lambda_{tj}\}; M_1 \right) = \mathbf{D}_{\theta(\beta)} \mathbf{Z}_t \mathbf{D}^{-1}_{\mathbf{Z}'_t \theta(\beta)} n \lambda_t \end{aligned} .$$

Uma vez obtido $\hat{\beta}$ e a sua matriz de covariância aproximada estimada, $\hat{\mathbf{V}}_{\hat{\beta}} = \left[\mathbf{I}_{2LL}(\hat{\beta}, \hat{\lambda}) \right]^{-1} \equiv \left[-\mathbf{H}_{2LL}(\hat{\beta}) \right]^{-1}$, com $\hat{\lambda}_t = n^{-1} \mathbf{D}^{-1}_{\mathbf{Z}'_t \hat{\theta}(\beta)} \mathbf{N}_t$, rapidamente se determinam as correspondentes estimativas para as probabilidades $\theta(\beta)$ e para a matriz de covariância aproximada associada

$$\mathbf{V}_{\hat{\theta}}(\beta,\lambda)=\left\{\mathbf{D}_{\theta(\beta)}-\theta(\beta)[\theta(\beta)]'\right\}\mathbf{X}\,[\mathbf{I}_2(\beta,\lambda)]^{-1}\mathbf{X}'\left\{\mathbf{D}_{\theta(\beta)}-\theta(\beta)[\theta(\beta)]'\right\},$$

bem como os valores observados das estatísticas de teste Q_V e Q_P para o ajustamento de H mencionadas anteriormente e cuja distribuição nula assintótica comum é

$$Q_V(M_2\cap H\mid M_1)\equiv^a Q_P(M_2\cap H\mid M_1)\xrightarrow[sob\ M_2\cap H]{a}\chi^2_{(s+m-1-p)}$$

.

$$Q_V(H\mid M_1)\equiv^a Q_P(H\mid M_1)\xrightarrow[sob\ M_2\cap H]{a}\chi^2_{(m-1-p)}\ .$$

5. ILUSTRAÇÃO

A metodologia descrita nas secções anteriores é aqui aplicada a dois exemplos. Os cálculos foram executados recorrendo a partes do programa desenvolvido por Rodrigues (1996) e a programas adicionais, todos elaborados na linguagem matricial do sistema NTIA (1996), semelhante à IML/SAS - é nossa intenção desenvolver a curto/médio prazo um novo programa integrado que inclua os desenvolvimentos aqui expostos e que seja executável por um *software* de maior difusão.

Exemplo 1: Este problema retracta um estudo de comparação de dois métodos de avaliação da susceptibilidade à cárie dentária numa amostra de crianças em idade escolar e já com dentição permanente. O método convencional, de difícil aplicação em grande escala e de custos elevados, baseou-se na contagem de bactérias *Lactobacillus* em amostras salivares. De acordo com o menor ou maior número destas bactérias, assim se classificou cada unidade em baixo, médio ou alto risco. O método de fácil aplicabilidade e de baixos custos operou esta classificação segundo a coloração obtida com a reacção de uma amostra de saliva com a resarzurina (azul, violeta e rosa, indicando baixo, médio e alto risco, respectivamente).

As frequências observadas, exibidas na tabela 1.1, mostram que um número significativo de crianças não conseguiu ser classificado de acordo com esta escala gradativa devido à ocorrência de padrões de cor intermédios. O conjunto de frequências pode ser encaixado no figurino estabelecido de partições do seguinte modo:

$$\Pi_1 = \textit{partição da classificação completa}\ ,\ \mathbf{Z}_1 = \mathbf{I}_9, \mathrm{N}_{1.} = 51;$$

$$\Pi_2 = \{\{(1,j),(2,j)\}, j=1,2,3\} \cup \{\{(3,1),(3,2),(3,3)\}\},\ \mathbf{Z}_2' = \begin{bmatrix} \mathbf{1}_2' \otimes \mathbf{I}_3 & \mathbf{0}_{3,3} \\ \mathbf{0}_6' & \mathbf{1}_3' \end{bmatrix}, \mathrm{N}_{2.} = 18;$$

$$\Pi_3 = \{\{(2,j),(3,j)\}, j=1,2,3\} \cup \{\{(1,1),(1,2),(1,3)\}\},\ \mathbf{Z}_3' = \begin{bmatrix} \mathbf{1}_3' & \mathbf{0}_6' \\ \mathbf{0}_{3,3} & \mathbf{1}_2' \otimes \mathbf{I}_3 \end{bmatrix}, \mathrm{N}_{3.} = 28.$$

Tabela 1.1: Frequências observadas dos graus de risco à cárie dentária

	Teste padrão		
Teste simplificado	alta	média	baixa
alta	7	11	2
média	3	9	5
baixa	0	10	4
alta/média	8	7	3
média/baixa	7	14	7

As grandes questões de interesse neste problema respeitam à comparação dos dois métodos em termos quer das probabilidades marginais dos graus de risco quer de alguma medida de concordância entre as classificações operadas.

Tabela 1.2: Estimativas MV de $\{\theta_{ij}\}$ e respectivos erros padrões na subamostra completa e na amostra global.

Celas ij	p_{ij}	$\hat{\sigma}(p_{ij})$ $(\times 10)$	$\hat{\theta}_{ij}$	$\hat{\sigma}(\hat{\theta}_{ij})$ $(\times 10)$	$\hat{\theta}_{ij}(H)$	$\hat{\sigma}(\hat{\theta}_{ij}(H))$ $(\times 10)$
11	0.137	0.482	0.106	0.359	0.105	0.354
12	0.216	0.576	0.142	0.390	0.135	0.275
13	0.039	0.272	0.026	0.179	0.026	0.171
21	0.059	0.330	0.152	0.656	0.160	0.488
22	0.177	0.534	0.219	0.529	0.223	0.510
23	0.098	0.416	0.124	0.387	0.131	0.337
31	0.000	0.000	$1.9\text{x}10^{-7}$	0.797	$1.9\text{x } 10^{-7}$	0.480
32	0.196	0.556	0.165	0.455	0.156	0.393
33	0.078	0.377	0.066	0.302	0.065	0.299

Para a resposta a estas questões, já previamente abordadas segundo outros métodos ou análises preliminares por Soares e Paulino (1998, 2000) e Rodrigues e Paulino (1998), expõem-se na tabela 1.2 estimativas MV das probabilidades de classificação conjunta e respectivos desvios padrões assintóticos estimados. As colunas 2 e 3 referem-se às proporções observadas na subamostra completa e exibem-se aqui apenas para mostrar que uma análise condicional à amostra completamente categorizada conduz a estimativas bem diferentes das que se obtêm com todas as unidades amostradas sob uma censura MAR e que se apresentam nas colunas 4 e 5.

A hipótese de um mecanismo de censura classicamente ignorável, M_2, é claramente desmentida pelos dados,

$$Q_V(M_2 \mid M_1) = 35.93 \underset{\nu=6}{\Rightarrow} P \approx 0$$

$$Q_P(M_2 \mid M_1) = 24.40 \underset{\nu=6}{\Rightarrow} P \approx 0,$$

como seria antecipável da análise das estimativas MV das probabilidades condicionais de censura MAR registadas na tabela 1.3.

Tabela 1.3: Estimativa MV das probabilidades condicionais de registo sob censura MAR.

Método simplificado	*Método convencional*		
	alta (1)	média (2)	baixa (3)
alta (1)	0.680	0.800	0.794
média (2)	0.204	0.424	0.415
baixa (3)	0	0.624	0.621
alta/média	0.320	0.200	0.206
média/baixa	0.476	0.376	0.379

O modelo de homogeneidade marginal $H : \mathbf{C}\overline{\theta} = \mathbf{0}$, com

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 \end{bmatrix},$$

apresenta um bom ajustamento,

$$Q_V(H \mid M_1) = 0.13 \approx Q_P(H \mid M_1) \underset{\nu=2}{\Rightarrow} P \approx 0.94 .$$

As estimativas MV dos θ_{ij} e erros padrões associados sob este modelo estão registados nas colunas 6 e 7 da tabela 1.2.

Utilizando como medida de concordância entre os dois métodos o coeficiente kapa de Cohen,

$$k=\frac{\sum_{i=1}^{3}\theta_{ii}-\sum_{i=1}^{3}\theta_{i.}\theta_{.i}}{1-\sum_{i=1}^{3}\theta_{i.}\theta_{.i}}\ ,$$

a sua estimativa MV e o erro padrão associado (calculado pelo método delta) são dados por $\hat{k}=0.016\,,\hat{\sigma}(\hat{\mathrm{k}})=0.090$.

Em suma, a análise efectuada revela que existem fortes evidências de que os dois métodos conduzem a idênticas proporções dos graus de risco atribuídos mas com uma grande discordância, pelo que o método simplificado não parece constituir um substituto válido do método padrão mais caro.

Exemplo 2: Este problema diz respeito a um estudo de caso-controlo, descrito em Williamson and Haber (1994), envolvendo mulheres com e sem cancro do colo do útero (variável A), classificadas adicionalmente pelo número de parceiros sexuais (variável B categorizada em *poucos* se inferior ou igual a 3 e *muitos* no caso contrário) e pelo rendimento anual do agregado familiar (variável C categorizada em *baixo* se inferior ou igual a 25000 USD e *alto* no caso contrário).

A tabela 2.1 de frequências observadas mostra que é apreciável o número de casos e controlos com classificação omissa nas variáveis número de parceiros sexuais e rendimento. A totalidade das frequências é agrupável nas seguintes partições:

$\Pi_1=$ *partição da classificação completa* , $\mathrm{N}_{1.}=652\,,\mathbf{Z}_1=\mathbf{I}_8$;

$\Pi_2=$ *partição da classificação em* $(\mathrm{X}_1,\mathrm{X}_2)\,,\mathrm{N}_{2.}=34\,,\mathbf{Z}_2=\mathbf{I}_4\otimes\mathbf{1}_2$;

$\Pi_3=$ *partição da classificação em* $(\mathrm{X}_1,\mathrm{X}_3)\,,\mathrm{N}_{3.}=102\,,\mathbf{Z}_3=\mathbf{I}_2\otimes(\mathbf{1}_2\otimes\mathbf{I}_2)$;

$\Pi_4=$ *partição da classificação em* $\mathrm{X}_1\,,\mathrm{N}_{4.}=26\,,\mathbf{Z}_4=\mathbf{I}_2\otimes\mathbf{1}_4$.

Tabela 2.1: Frequências observadas em mulheres

Statu do cancro cervical	Número de parceiros sexuais	Rendimento baixo	Rendimento alto	Rendimento omisso
controlo	poucos	77	87	14
controlo	muitos	94	70	8
controlo	omisso	25	18	14
caso	poucos	67	36	3
caso	muitos	143	78	9
caso	omisso	43	16	12

A questão central de interesse reside em averiguar o tipo de associação eventualmente existente com vista a determinar o possível efeito na doença das restantes variáveis. Para tal vamos considerar que tais objectivos são alcançáveis de um delineamento cruzado compatível com o modelo Multinomial em que se baseia a análise descrita neste artigo.

À semelhança do exemplo anterior, a tabela 2.2 contém várias estimativas MV das probabilidades das oito categorias das variáveis (tidas como) respostas. O teste de ajustamento do modelo MCAR apresenta resultados não significativos,

$$Q_V(M_2 \mid M_1) = 9.56 \underset{\nu=7}{\Rightarrow} P \cong 0.21$$

$$Q_P(M_2 \mid M_1) = 9.90 \underset{\nu=7}{\Rightarrow} P \cong 0.19\,,$$

pelo que a análise subsequente de estruturas log-lineares para $\{\theta_{ijk}\}$ poderá fazer uso de tal resultado, admitindo a validade do mecanismo de censura MCAR.

Tabela 2.2: Estimativas MV de $\{\theta_{ijk}\}$ e respectivos erros padrões na subamostra completa e na amostra global

Celas ijk	p_{ijk}	$\hat{\sigma}(p_{ijk})$ $(\times 10)$	$\hat{\theta}_{ijk}$	$\hat{\sigma}(\hat{\theta}_{ijk})$ $(\times 10)$	$\hat{\theta}_{ijk}(H)$	$\hat{\sigma}(\hat{\theta}_{ijk}(H))$ $(\times 10)$
111	0.118	0.126	0.121	0.123	0.127	0.117
112	0.133	0.133	0.133	0.127	0.127	0.116
121	0.144	0.138	0.143	0.131	0.137	0.120
122	0.107	0.121	0.103	0.115	0.109	0.107
211	0.103	0.119	0.105	0.116	0.099	0.104
212	0.055	0.089	0.053	0.086	0.059	0.076
221	0.219	0.162	0.226	0.154	0.232	0.149
222	0.120	0.127	0.117	0.119	0.111	0.107

A tabela 2.3 indica as estimativas MV e correspondentes erros padrões dos parâmetros log-lineares do modelo sem interacção de segunda ordem (definido segundo a parametrização desvios de médias). As decorrentes estimativas associadas às probabilidades θ_{ijk} estão nas colunas 6 e 7 da tabela 2.2. O resultado não significativo dos testes de ajustamento desse modelo,

$$Q_V(H \mid M_1) = 1.66 \underset{\nu=1}{\Rightarrow} P \cong 0.20 \ ,$$

$$Q_P(H \mid M_2) = 1.64 \underset{\nu=1}{\Rightarrow} P \cong 0.20 \, ,$$

mostra que não há indícios significativos de interacção de segunda ordem entre as três variáveis, ou equivalentemente, de interacção entre B e C nos logitos de ocorrência da doença. Contudo, os modelos log-lineares que se lhe sucedem na escala hierárquica apresentam todos resultados altamente significativos. Por exemplo, o teste Q_V condicional de anulamento da interacção entre a doença e o número de parceiros sexuais, $H_0 : u_{11}^{AB} = 0$, dado o modelo anterior apresenta o resultado $Q_V(H_0 \mid H; M_1) = 24.39$ ($P \cong 0$ para 1 g. l.).

Tabela 2.3: Estimativas MV dos parâmetros log-lineares do modelo sem interacção de 2ª ordem e respectivos erros padrões.

β	u_1^A	u_1^B	u_1^C	u_1^{AB}	u_1^{AC}	u_1^{BC}
$\hat{\beta}$	0.060	- 0.176	0.185	0.196	- 0.128	- 0.056
$\hat{\sigma}(\hat{\beta})$ $(\times 10)$	0.377	0.403	0.381	0.400	0.382	0.413

Assim, a associação presente entre as três variáveis parece ser razoavelmente descrita por um modelo linear aditivo nos logitos da ocorrência de cancro, reflectindo que os casos estão positivamente associados com baixo rendimento do agregado familiar e com muitos parceiros sexuais, conforme transparece das estimativas, reproduzidas na tabela 2.4, das razões de produtos cruzados parciais logaritmizadas para tal modelo enquadrado na estrutura MCAR.

Tabela 2.4: Estimativas MV do logaritmo das RPC parciais (e respectivos erros padrões) sob o modelo sem interacção de 2ª ordem com censura MCAR.

$\ln \hat{\Delta}^A$	$\hat{\sigma}(\ln \hat{\Delta}^A)$	$\ln \hat{\Delta}^B$	$\hat{\sigma}(\ln \hat{\Delta}^B)$	$\ln \hat{\Delta}^C$	$\hat{\sigma}(\ln \hat{\Delta}^C)$
- 0.225	0.166	- 0.513	0.153	0.782	0.160

REFERÊNCIAS BIBLIOGRÁFICAS

BAKER, S. G. and LAIRD, N. M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *J. Amer. Statist. Assoc.* 83, 62-69.

BAKER, S. G. (1994). Regression analysis of grouped survival data with incomplete covariates: nonignorable missing data and censoring mechanisms. *Biometrics* 50, 821-826.

BAKER, S. G. (1996). The analysis of categorical case-control data subject to nonignorable nonresponse. *Biometrics* 52, 362-369.

CHAMBERS, R. L. and WELSH, A. H. (1993). Log-linear models for survey data with non-ignorable non-response. *J. Royal Statist. Soc. B* 55, 157-170.

CONAWAY, M. R. (1993). Non-ignorable non-response models for time-ordered categorical variables. Appl. Statistics 42, 105-115.

DICKEY, J. M., JYANG, T. J. and KADANE, J. B. (1987). Bayesian methods for censored categorical data. *J. Amer. Statist. Assoc.* 87, 773-781.

FORSTER, J. J. and SMITH; W. F. (1998). Model-based inference for categorical survey data subject to non-ignorable non-response. *J. Royal Statist. Soc. B* 60, 57-70.

LITTLE, R. J. A. and RUBIN, D. B. (1987). *Statistical analysis with missing data*. Wiley, New York.

PARK, T. and BROWN, M. B. (1994). Models for categorical data with nonignorable nonresponse. *J. Amer. Statist. Assoc.* 89, 44-52.

PAULINO, C. D. (1991). Analysis of incomplete categorical data: a survey of the conditional maximum likelihood and weighted least squares approaches. *The Brazilian Journal of Probability and Statistics* 5, 1- 42.

PAULINO, C. D. (2000). Modelação e Análise de Dados Poisson sob Censura não Informativa: Parte I – Modelação e Ajustamento de Modelos de Censura. *Submetido às Actas do VII Congresso da SPE.*

PAULINO, C. D. and PEREIRA, C. A. B. (1992). Bayesian analysis of categorical data informatively censored. *Comm. in Statistics - Theory Meth.*, 21(9), 2689-2705.

PAULINO, C. D. and PEREIRA, C. A. B. (1995). Bayesian methods for categorical data under informative general censoring. *Biometrika*, 82, 2, 439-446.

PAULINO, C. D. e SOARES, P. (2000). Modelação e Análise de Dados Poisson sob Censura não Informativa: Parte II – Ajustamento de Modelos Estruturais para as Taxas e Aplicações. *Submetido às Actas do VII Congresso da SPE*.

RODRIGUES, I.M.A. (1996). Implementação computacional de análises clássicas de dados categorizados incompletos. *Tese de mestrado*, Instituto Superior Técnico, Lisboa.

RODRIGUES, I.M.A. e PAULINO, C.D.M. (1998). Avaliação comparada da susceptibilidade à cárie dentária: uma análise de dados categorizados incompletos. *In Estatística: a diversidade na unidade,* pp. 109-116, Col. Novas Tecnologias/Estatística, Salamandra, Lisboa.

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

SOARES, P. e PAULINO, C. D. (1998). Análise bayesiana de dados categorizados informativamente omissos: uma abordagem por simulação. *In Estatística: a diversidade na unidade,* pp. 405-410, Col. Novas Tecnologias/Estatística, Salamandra, Lisboa.

SOARES, P. and PAULINO, C.D. (2000). Incomplete categorical data analysis: A Bayesian perspective. *Submitted.*

SOFTWARE NTIA, versão 4.2.1 (1996). Centro Nacional de Pesquisa Tecnológica em Informática para a Agricultura, EMBRAPA, Campinas; SP, Brasil.

WALKER, S. (1996). A Bayesian maximum *a posteriori* algorithm for categorical data under informative general censoring. *The Statistician* 45, 293-298.

WILLIAMSON, G. D. and HABER, M. (1994). Models for three-dimensional contingency tables with completely and partially cross-classified data. *Biometrics* 49, 194-203.