Um estudo de métodos bayesianos para dados de sobrevivência com omissão nas covariáveis.

Démerson André Polli

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO DE MESTRE
EM
CIÊNCIAS

Área de Concentração: **Estatística**
Orientador: **Prof. Dr. Antonio Carlos Pedroso de Lima**

São Paulo, abril de 2007.

Um estudo de métodos bayesianos para dados de sobrevivência com omissão nas covariáveis.

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Démerson André Polli e aprovada pela Comissão Julgadora.

São Paulo, 01 de abril de 2007.

Banca examinadora:

- Prof. Dr. Antonio Carlos Pedroso de Lima (orientador) - IME/USP
- Prof. Dr. Heleno Bolfarine - IME/USP
- Prof. Dr. Edwin Moises Marcos Ortega - ESALQ/USP

Agradecimentos

Em primeiro lugar, agradeço a Deus pela minha vida e por me dar inteligência, saúde e condições para realizar mais este trabalho. Se não fosse a Sua mão me guiando nada teria acontecido!
Gostaria de agradecer as pessoas que, de uma forma ou de outra, me ajudaram desde a minha graduação:

Ao meu orientador e amigo Antonio Carlos, pelo companheirismo, paciência e por me mostrar, com suas sugestões e críticas, como realizar um trabalho sério e competente de pesquisa.

Aos meus pais, Siléia e Júlio, e irmãos (Danilo e Gustavo) que, através da convivência familiar, fizeram parte de minha formação moral.

À minha esposa e fiel companheira Jadasa pelo amor, carinho e dedicação. Obrigado por ter suportado (e ainda estar suportando) as dificuldades e desafios decorrentes desta jornada. O caminho é difícil, mas logo veremos os resultados!

À minha sogra Rosa por sempre estar presente, fazendo o papel de mãe, inclusive durante a realização deste trabalho.

Aos meus avós (*in memorian*) Mário e Cida que, mesmo com simplicidade, sempre me apoiaram. Tenho muitas saudades de vocês e gostaria que estivessem aqui para ver um dos resultados do apoio que me deram.

Aos meus avós Hulda e Daniel; e aos meus tios Mary, Dailes e Susie pelo apoio durante minha graduação. Em especial, agradeço à tia Mary pela ajuda financeira durante toda aquela época!

Aos meus tios Jamil e Dorcas que gentilmente me deram um lar em São Paulo durante 4 anos e que, deste modo, facilitaram muito minha vida durante a graduação. Aos meus primos (e as respectivas famílias), Sônia, Sueli, Corina, Wolsey e Eliel, e à vó Corina por todo o apoio durante o período que estive por aí.

Aos muitos amigos que fiz na USP, alunos, professores e funcionários, muitos dos quais não vejo a um bom tempo, pelas dicas, críticas, sugestões e, especialmente, pela amizade e companheirismo.

Resumo

O desenvolvimento de métodos para o tratamento de omissões nos dados é recente na estatística e tem sido alvo de muitas pesquisas. A presença de omissões em covariáveis é um problema comum na análise estatística e, em particular nos modelos de análise de sobrevivência, ocorrendo com freqüência em pesquisas clínicas, epidemiológicas e ambientais.

Este trabalho apresenta propostas bayesianas para a análise de dados de sobrevivência com omissões nas covariáveis considerando modelos paramétricos da família Weibull e o modelo semi-paramétrico de Cox.

Os métodos estudados foram avaliados tanto sob o enfoque paramétrico quanto o semi-paramétrico considerando um conjunto de dados de portadores de insuficiência cardíaca. Além disso, é desenvolvido um estudo para avaliar o impacto de diferentes proporções de omissão.

Palavras-chave: omissão nas covariáveis, análise de sobrevivência, inferência bayesiana.

Abstract

The development of methods dealing with missing data is recent in Statistics and is the target of many researchers. The presence of missing values in the covariates is very common in statistical analysis and, in particular, in clinical, epidemiological and enviromental studies for survival data.

This work considers a bayesian approach to analise data with missing covariates for parametric models in the Weibull family and for the Cox semiparametric model.

The studied methods are evaluated for the parametric and semiparametric approaches considering a dataset of patients with heart insufficiency. Also, the impact of different omission proportions is assessed.

Keywords: missing covariates, survival analysis, bayesian inference.

Sumário

Notação

Neste trabalho, será adotada a seguinte notação:

- Variáveis aleatórias são escritas em maiúsculo $(X, Y, \ldots)$ e os respectivos valores observados são escritos em minúsculo $(x, y, \ldots)$.
- Matrizes e vetores aleatórios são escritos, para distinguir das variáveis, em negrito $(\boldsymbol{X}, \boldsymbol{Y}, \ldots)$ e, da mesma forma, os valores observados em minúsculo $(\boldsymbol{x}, \boldsymbol{y}, \ldots)$.
- Variáveis e vetores que representam parâmetros de distribuições de probabilidade serão escritos em letras gregas $(\lambda, \boldsymbol{\theta}, \ldots)$.
- Os elementos de matrizes e vetores aleatórios são indicados entre colchetes.
- Os elementos de conjuntos são indicados entre chaves.
- A função $\log(\cdot)$ representa o logaritmo natural.
- A função densidade de probabilidade (função de probabilidade no caso discreto) será representada por $\mathrm{f}(\cdot)$ e a correspondente função densidade acumulada (função de probabilidade acumulada) será representada por $\mathrm{F}(\cdot)$. Em algumas situações tais funções serão definidas em função de um espaço de probabilidades $(\Omega, \mathcal{F}, \mathcal{P})$ em que Ω é o espaco amostral, $\mathcal{F}$ é a σ-álgebra e $\mathcal{P}$ são as probabilidades associadas a cada elemento da σ-álgebra.
- As distribuições a *priori* e a *posteriori* serão indicadas por $\pi(\cdot)$ e $\pi(\cdot|\cdot)$.
- Os dados omissos são indicados com asterisco $(\boldsymbol{x}^*, x^*, \ldots)$
- Para simplificar a notação, foram usados índices tanto para elementos de vetores e matrizes quanto para iterações de algoritmos. Nestes casos, os tipos de índices são separados por ";" (ponto e vírgula), sendo que o índice referente à iteração vem antes dos índices de vetores e matrizes. Por exemplo, $\boldsymbol{\theta}_{k;i}$ refere-se ao valor do vetor $\boldsymbol{\theta}_i$ na k-ésima iteração do algoritmo.
- Geralmente o índice i se refere a unidades amostrais, j a covariáveis e k a iterações dos algoritmos.
- Para unificar as definições para as variáveis contínuas e discretas será usado, sempre que possível, a integral de Lebesgue-Stieltjes (Magalhães, 2006), que no caso contínuo é calculada através da integral de Riemann e no caso discreto como um somatório.

1. Introdução

O desenvolvimento de métodos para o tratamento de omissões nos dados, seja na variável resposta ou nas covariáveis, é recente na estatística (Rubin, 1976, 1980, 1996; Rubin e Schenker, 1986; Ibrahim, 1990; Efron, 1994; Rubin et al., 1994; Ibrahim et al., 1999a,b, 2002). O aumento do poder de processamento dos computadores em conjunto com o desenvolvimento de teorias que justificam o uso de simulações na análise de dados tem permitido o estudo deste problema com maior profundidade.

Em particular, em modelos de análise de sobrevivência é comum a ocorrência de omissão em covariáveis em algumas áreas de conhecimento como, por exemplo, pesquisas clínicas, epidemiológicas e ambientais. As propostas mais simples de solução deste problema consiste em ignorar os casos omissos (análise de casos completos), excluir da análise as covariáveis que apresentam omissão ou substituir as omissões pela média das observações com dados completos. Estas propostas podem causar perda de informação e vício nas estimativas (Little e Rubin, 1987; Ibrahim et al., 2001).

O uso de métodos bayesianos é uma alternativa interessante neste caso pois, definindo uma distribuição a priori para os parâmetros do modelo, é possível gerar valores que irão substituir as omissões usando o conhecimento que o pesquisador tem a respeito dos dados. No entanto, a literatura sobre modelos de análise de sobrevivência com omissões nas covariáveis é mais farta para a aplicação da inferência clássica, sendo quase inexistente para a abordagem bayesiana.

O modelo semi-paramétrico de Cox (1972) é, provavelmente, o modelo de análise de sobrevivência mais usado na prática. Na literatura existem diversas propostas para o ajuste deste modelo para conjuntos com omissões nas covariáveis.

Lin e Ying (1993) são os primeiros autores a propor um método para o ajuste de conjuntos de dados com omissões nas covariáveis para o modelo de Cox. A proposta consiste em ajustar o modelo usando apenas as informações observadas para cada unidade amostral. Deste modo, quando algum individuo na amostra apresenta omissões nas covariáveis, apenas aquelas que foram observadas serão usadas no ajuste do modelo. Uma adaptação deste método é proposta por Zhou e Pepe (1995) e consiste em maximizar a esperança da veros-

similhança parcial, sendo que para aqueles indivíduos que falharam são feitas imputações de valores nas covariáveis com omissão. É importante observar que estes modelos supõem que as omissões são geradas ao acaso (Rubin, 1976). Paik e Tsai (1997) e Paik (1997) propõem adaptações nestes modelos para os casos em que a ocorrência da omissão depende do tempo de sobrevivência e da censura ou das covariáveis observadas. O inconveniente destes modelos é a dificuldade de implementação, o que tem impedido seus usos na prática.

Pugh et al. (1993) propõe usar apenas as unidades amostrais com dados completos para a estimação dos parâmetros no modelo de Cox, dando um maior peso para aquelas que apresentam a menor probabilidade de serem completamente observadas. Tal probabilidade é estimada através de uma regressão logística. No entanto, de acordo com Therneau e Grambsch (2000), este método não tem apresentado resultados consistentes.

Por outro lado, também existem diversas propostas para os modelos paramétricos de regressão com omissões nas covariáveis. Little e Rubin (1983) propõem maximizar a função de verossimilhança tratando os valores não observados como parâmetros. Tal abordagem tem o inconveniente de que o número de parâmetros cresce em função do tamanho da amostra e, no caso em que a proporção de omissão nas covariáveis é alta, torna-se inviável. DeGroot e Goel (1980) propõe o uso desta idéia para misturas de distribuições normais bivariadas e Press e Scott (1976) apresenta uma adaptação para a inferência bayesiana em que a distribuição a posteriori conjunta dos parâmetros e das covariáveis é maximizada.

A proposta de Joseph Ibrahim e colaboradores (Ibrahim, 1990; Lipsitz e Ibrahim, 1996, 1998; Ibrahim et al., 1999a,b; Chen e Ibrahim, 2001; Herring e Ibrahim, 2002) baseia-se no uso do algoritmo EM (Dempster et al., 1977) para a maximização da verossimilhança em conjuntos de dados com omissões nas covariáveis. Uma adaptação deste método para o ajuste do modelo de Cox é proposta por Leong et al. (2001). Esta metodologia é detalhada na seção 2.1.

A proposta de Donald Rubin e co-autores (Rubin, 1980, 1987, 1996; Little e Rubin, 1983, 1987; Rubin e Schenker, 1986; Little, 1992; Liu e Rubin, 1994) consiste em gerar múltiplos conjuntos de dados imputando valores gerados empiricamente nas variáveis omissas. A proposta inicial trata apenas de omissão em uma única variável em que se deseja estimar os parâmetros da distribuição. Uma adaptação desta proposta para modelos de sobrevivência é apresentada na seção 2.2.

Este trabalho é organizado em 5 capítulos. O capítulo 1 apresenta um resumo sobre o problema de omissões em covariáveis em modelos de sobrevivência. O capítulo 2 apresenta os detalhes das propostas de analise dos dados usando o algoritmo EM (seção 2.1) ou

a imputação em múltiplos conjuntos (seção 2.2), abordando o ajuste de modelos Weibull (seção 2.3) e de Cox (seção 2.4). A implementação dos algoritmos é discutida no capítulo 3. O capítulo 4 apresenta a análise de um conjunto de dados de sobrevivência com omissão nas covariaveis e um estudo de sensibilidade dos modelos apresentados nesta dissertação. As conclusões e as propostas para trabalhos futuros são apresentadas no capítulo 5.

Os programas de computador que implementam os métodos estudados nesta dissertação podem ser obtidos em `http://www.ime.usp.br/~acarlos/missingBayes`.

2. Metodologia

Um tratamento usual para omissões em covariáveis é o método genericamente denominado *imputação* (Rubin, 1980), que consiste em completar as lacunas causadas pelas omissões nas covariáveis. Como o conjunto original de dados é "aumentado" o método também é chamado de *dados aumentados*.

Uma proposta de implementação de tal método, apresentada em Ibrahim (1990); Lipsitz e Ibrahim (1996, 1998); Ibrahim et al. (1999a,b); Chen e Ibrahim (2001) e Herring e Ibrahim (2002), consiste em maximizar a distribuição a *posteriori* dos parâmetros da regressão usando o algoritmo EM (Dempster et al., 1977). Uma alternativa, apresentada em Rubin (1980, 1987, 1996); Little e Rubin (1983, 1987); Rubin e Schenker (1986); Little (1992) e Liu e Rubin (1994), consiste em criar diversos conjuntos completos de dados e maximizar a distribuição a *posteriori* para cada um dos conjuntos. A estimativa dos parâmetros, neste caso, é a média aritmética das estimativas individuais para cada um dos conjuntos. Estas implementações são abordadas, respectivamente, nas seções 2.1 e 2.2. O modelo paramétrico Weibull é abordado na seção 2.3 e o modelo de Cox (1972) é abordado na seção 2.4.

Considere o conjunto de dados $\{n, \boldsymbol{T}, \boldsymbol{\Delta}, \boldsymbol{X}\}$, em que n é o tamanho amostral, $\boldsymbol{T}$ é o vetor de dimensão n de tempos de acompanhamento (falhas ou censuras), $\boldsymbol{\Delta}$ é o vetor de dimensão n de indicadores de falha e $\boldsymbol{X}$ é a matriz de dimensão n por p de covariáveis. Tal conjunto assume os valores $\mathcal{D} = \{n, \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x}\}$ com tempo de acompanhamento $\boldsymbol{t}' = [t_1, \ldots, t_n]$, indicadores de falha $\boldsymbol{\delta}' = [\delta_1, \ldots, \delta_n]$ e matriz $\boldsymbol{x}' = [\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_n]$ composta pelos vetores (linha) $\boldsymbol{x}_i, i = 1, \ldots, n$ de covariáveis associadas a cada unidade amostral. Os dados referentes à i-ésima unidade amostral são representados por $\mathcal{D}_i = \{t_i, \delta_i, \boldsymbol{x}_i\}$ com $\delta_i = 1$ se a unidade amostral falhou e $\delta_i = 0$ se foi censurada.

Sejam, para a i-ésima unidade amostral, C_i o tempo de censura *à direita* e T_i^0 o tempo de falha, de modo que T_i, o tempo de acompanhamento, seja dado por $T_i = \min(T_i^0, C_i)$. Assim, se $\boldsymbol{\theta}$ é o vetor de parâmetros associado à distribuição de probabilidades dos tempos de sobrevivência $T_i^0, i = 1, \ldots, n$, e $\pi(\boldsymbol{\theta})$ é a respectiva *distribuição a priori* então as

inferências sobre $\boldsymbol{\theta}$ são obtidas através da *distribuição a posteriori* definida por

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto \mathrm{L}(\boldsymbol{\theta}|\mathcal{D})\pi(\boldsymbol{\theta})$$

com $\mathrm{L}(\boldsymbol{\theta}|\mathcal{D})$, a verossimilhança de $\boldsymbol{\theta}$, dada por

$$\mathrm{L}(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{L}(\boldsymbol{\theta}|\mathcal{D}_i) = \prod_{i=1}^{n} \mathrm{f}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} \cdot \mathrm{S}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i}$$

em que $\mathrm{f}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})$ é a função densidade de probabilidade e $\mathrm{S}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})$ é a função de sobrevivência de T_i^0. Note que, a princípio, não é necessário considerar a distribuição das covariáveis $\boldsymbol{X}_i$, pois, o interesse é fazer inferências a respeito dos tempos de sobrevivência e não das covariáveis. No entanto, havendo omissões nestas covariáveis é necessário considerar suas distribuições de probabilidade. Assim, a distribuição de probabilidades do vetor $\boldsymbol{X}_i$ será denotada por $\mathrm{f}(\boldsymbol{X}_i|\boldsymbol{\alpha})$ com o vetor de parâmetros de perturbação do modelo $\boldsymbol{\alpha}$. Portanto, a distribuição a *posteriori* conjunta para os vetores $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ será

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) \propto \mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) \cdot \pi(\boldsymbol{\theta}, \boldsymbol{\alpha}) \tag{2.1}$$

com $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha})$ a priori conjunta de $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$. Se a distribuição das covariáveis for independente da distribuição do tempo de sobrevivência, os vetores $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ serão independentes e a verossimilhança conjunta $\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D})$ será dada pelo produto das verossimilhanças individuais de $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$,

$$\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{L}(\boldsymbol{\theta}|\mathcal{D}_i) \cdot \mathrm{L}(\boldsymbol{\alpha}|\mathcal{D}_i) = \prod_{i=1}^{n} \mathrm{f}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} \cdot \mathrm{S}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i} \cdot \mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha})$$

Uma situação curiosa ocorre se a distribuição a priori conjunta para $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ for obtida pelo método de Bayes-Laplace (Paulino et al., 2003), em que todos os possíveis valores de $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ são equiprováveis, isto é,

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}) \propto 1.$$

Neste caso, a posteriori será proporcional à verossimilhança conjunta para $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$,

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) \propto \prod_{i=1}^{n} \mathrm{f}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} \cdot \mathrm{S}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i} \cdot \mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha}) = \mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) \tag{2.2}$$

e, por isso, os vetores $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ de parâmetros que maximizam a distribução a posteriori será também aqueles que maximizam a verossimilhança conjunta. No entanto, esta abordagem é questionável pois, como a priori de Bayes-Laplace é *imprópria*, a posteriori também poderá ser imprópria.

Para simplificar a modelagem, é possível particionar a distribuição do vetor $\boldsymbol{X}_i = [X_{i,1}, \ldots, X_{i,p}], i = 1, \ldots, n$, em um produto de distribuições condicionais. Deste modo, sua densidade de probabilidade pode ser escrita como

$$\mathrm{f}(\boldsymbol{X}_i|\boldsymbol{\alpha}) = \mathrm{f}(X_{i,1}|\boldsymbol{\alpha}_1) \cdot \mathrm{f}(X_{i,2}|X_{i,1}, \boldsymbol{\alpha}_2) \cdots \mathrm{f}(X_{i,p}|X_{i,1}, \ldots, X_{i,p-1}, \boldsymbol{\alpha}_p) \tag{2.3}$$

em que $\boldsymbol{\alpha}_j, j = 1, \ldots, p$, são vetores *distintos* de parâmetros associados à cada uma das densidades condicionais e $\boldsymbol{\alpha}' = [\ \boldsymbol{\alpha}_1' \quad \ldots \quad \boldsymbol{\alpha}_p'\ ]$. Esta partição faz com que se evite o uso de distribuições multivariadas para as covariáveis. Além disso, as interações entre tais covariáveis são eliminadas do modelo, causando uma redução do número de parâmetros a serem estimados. Os vetores $\boldsymbol{\alpha}_j$, de parâmetros de perturbação, podem ser estimados através de modelos de regressão (linear ou logística).

Considere agora que, para a i-ésima unidade amostral, é possível particionar o vetor $\boldsymbol{X}_i$ em $\boldsymbol{X}_i = [\boldsymbol{X}_i^o, \boldsymbol{X}_i^*]$ com $\boldsymbol{X}_i^*$ um vetor de dimensão q_i de covariáveis omissas e $\boldsymbol{X}_i^o$ um vetor de dimensão $p - q_i$ de covariáveis completamente observadas. Nesta condição, se a densidade de probabilidade do vetor $\boldsymbol{X}_i^o$ é $\mathrm{f}\,(\boldsymbol{X}_i^o|\boldsymbol{\alpha}_0)$ e a densidade de probabilidade do vetor $\boldsymbol{X}_i^*$ puder ser particionada como em (2.3), então

$$\mathrm{f}(\boldsymbol{X}_i|\boldsymbol{\alpha}) = \mathrm{f}\,(\boldsymbol{X}_i^o|\boldsymbol{\alpha}_0)\, \mathrm{f}(X_{i,1}^*|\boldsymbol{X}_i^o, \boldsymbol{\alpha}_1) \cdots \mathrm{f}(X_{i,q_i}^*|X_{i,1}^*, \ldots, X_{i,q_i-1}^*, \boldsymbol{X}_i^o, \boldsymbol{\alpha}_{q_i})$$

em que a densidade condicional da covariável (omissa) $X_{i,j}^*, j = 1, \ldots, q_i$, é denotada por $\mathrm{f}(X_{i,j}^*|X_{i,1}^*, \ldots, X_{i,j-1}^*, \boldsymbol{\alpha}_j)$. As dimensões dos vetores $\boldsymbol{X}_i^o$ e $\boldsymbol{X}_i^*$ são diferentes entre as unidades amostrais pois dependem da estrutura de omissão (quais covariáveis foram ou não observadas para cada unidade amostral). Além disso, como não serão feitas imputações de dados no vetor $\boldsymbol{X}_i^o$ das covariáveis completamente observadas, é possível ignorar a distribuição associada à estas covariáveis ao menos que seja interessante fazer inferências a respeito do vetor $\boldsymbol{\alpha}_0$. Assim, a expressão acima se reduz a

$$\mathrm{f}(\boldsymbol{X}_i|\boldsymbol{\alpha}) \propto \mathrm{f}(X_{i,1}^*|\boldsymbol{X}_i^o, \boldsymbol{\alpha}_1) \cdots \mathrm{f}(X_{i,q_i}^*|X_{i,1}^*, \ldots, X_{i,q_i-1}^*, \boldsymbol{X}_i^o, \boldsymbol{\alpha}_{q_i}) \tag{2.4}$$

Note em (2.4) que as unidades amostrais completamente observadas não contribuem para a estimação dos parâmetros de perturbação.

2.1. Implementação paramétrica através do algoritmo EM

A implementação da *imputação múltipla* via algoritmo EM (Dempster et al., 1977) está presente em diversos artigos empregando modelos lineares generalizados (Ibrahim, 1990; Ibrahim et al., 1999a,b) ou modelos de análise de sobrevivência (Lipsitz e Ibrahim, 1996,

1998; Chen e Ibrahim, 2001; Herring e Ibrahim, 2002). A seção 2.1.1 apresenta a implementação para o cenário em que o processo que gera as omissões não depende das covariáveis, ou seja, o caso de *omissões não informativas*. Este é o cenário conhecido por MCAR (*missing completely at random* ou *omissões completamente ao acaso*). Uma extensão para o cenário de *omissão informativa* no qual o mecanismo gerador de omissões depende das covariáveis será apresentada na seção 2.1.2. Este cenário contempla os mecanismos MAR (*missing at random* ou *omissão ao acaso*) no qual as omissões dependem das covariáveis observadas (Rubin, 1976) e o mecanismo MNAR (*missing not at random* ou *omissão não aleatória*) no qual as omissões dependem também das covariáveis omissas.

O apêndice A apresenta detalhes sobre o algoritmo EM e as variações ECM (Meng e Rubin, 1993) e ECME (Liu e Rubin, 1994).

2.1.1. Mecanismos de omissão não informativa

O algoritmo EM (Dempster et al., 1977) foi criado para a obtenção de estimativas de máxima verossimilhança em modelos com dados incompletos (omissos ou latentes). As estimativas de máxima densidade a posteriori, no entanto, podem ser obtidas de forma análoga, maximizando-se a esperança (com relação à distribuição das variáveis latentes ou omissas) do logaritmo da posteriori. É um algoritmo iterativo, cuja k-ésima iteração consiste em calcular os vetores $(\boldsymbol{\alpha}_k, \boldsymbol{\theta}_k)$ tal que

$$(\boldsymbol{\alpha}_k, \boldsymbol{\theta}_k) = \arg\max \mathrm{Q}(\boldsymbol{\alpha}, \boldsymbol{\theta} | \boldsymbol{\alpha}_{k-1}, \boldsymbol{\theta}_{k-1}) = \arg\max \sum_{i=1}^{n} \mathrm{Q}_i(\boldsymbol{\alpha}, \boldsymbol{\theta} | \boldsymbol{\alpha}_{k-1}, \boldsymbol{\theta}_{k-1}) \tag{2.5}$$

em que,

$$\mathrm{Q}_i(\boldsymbol{\alpha}, \boldsymbol{\theta} | \boldsymbol{\alpha}_k, \boldsymbol{\theta}_k) = \int_{\boldsymbol{z} \in \mathcal{S}_i} \ell_i(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k | \mathcal{D}_{i,(\boldsymbol{z})}) \, \mathrm{dF}(\boldsymbol{z} | t_i, \delta_i, \boldsymbol{x}_i^o, \boldsymbol{\alpha}_k) \tag{2.6}$$

é a contribuição da i-ésima unidade amostral para a esperança (com relação às covariáveis omissas) do logaritmo da posteriori, $\mathcal{S}_i$ é o conjunto de todos os valores possíveis para as covariáveis omissas (espaço amostral) para a i-ésima unidade amostral, $\boldsymbol{z}$ é o vetor com os valores imputados em $\boldsymbol{X}_i^*$, $\mathcal{D}_{i,(\boldsymbol{z})}$ denota o conjunto de dados com os valores omissos substituídos pelo vetor $\boldsymbol{z}$ e

$$\begin{aligned} \ell_i(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k | \mathcal{D}_{i,(\boldsymbol{z})}) \;\; = \;\; & \delta_i \cdot \ln \mathrm{f}(t_i | \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\theta}_k) \;+\; (1 - \delta_i) \cdot \ln \mathrm{S}(t_i | \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\theta}_k) \\ & +\; \ln \mathrm{f}(\boldsymbol{z} | \boldsymbol{x}_i^o, \boldsymbol{\alpha}_k) \;+\; \ln \pi(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k). \end{aligned}$$

O algoritmo é iniciado com as estimativas iniciais $\boldsymbol{\alpha}_0$ e $\boldsymbol{\theta}_0$ para os vetores de parâmetros e, no k-ésimo passo, as estimativas dos parâmetros são denotadas pelos vetores $\boldsymbol{\alpha}_k$ e $\boldsymbol{\theta}_k$.

Note que a expressão (2.6) é a média (esperança) do logaritmo da distribuição a posteriori ponderada pela probabilidade do vetor $\boldsymbol{X}_i^*$ das covariáveis omissas assumir o valor $\boldsymbol{z}$. Tal probabilidade, no caso em que as covariáveis omissas são categorizadas (ou enumeráveis) é dada pela expressão (Lipsitz e Ibrahim, 1996)

$$w_i(\boldsymbol{z}) = \mathrm{f}(\boldsymbol{z}|t_i, \delta_i, \boldsymbol{x}_i^o, \boldsymbol{\alpha}_k) = \frac{\mathrm{f}(t_i|\delta_i, \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\theta}_k)\mathrm{f}(\boldsymbol{x}_i^o, \boldsymbol{z}|\boldsymbol{\alpha}_k)}{\sum_{\boldsymbol{z}\in\mathcal{S}_i}\mathrm{f}(t_i|\delta_i, \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\theta}_k)\mathrm{f}(\boldsymbol{x}_i^o, \boldsymbol{z}|\boldsymbol{\alpha}_k)}$$

e a expressão (2.6) é dada por

$$\mathrm{Q}_i(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{\alpha}_k, \boldsymbol{\theta}_k) = \sum_{\boldsymbol{z}\in\mathcal{S}_i} w_i(\boldsymbol{z})\ell_i(\boldsymbol{\theta}_j, \boldsymbol{\alpha}_j|\mathcal{D}_{i,(\boldsymbol{z})}).$$

Este é o algoritmo conhecido por *EM ponderado* (Ibrahim, 1990).

No caso em que as covariáveis são contínuas, a expressão (2.6) é aproximada pelo algoritmo *Monte Carlo EM* (Wei e Tanner, 1990), que consiste em gerar $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{m_i}$, uma amostra aleatória de tamanho m_i, da distribuição de probabilidades das covariáveis omissas, $\mathrm{f}(\boldsymbol{X}_i^*|\boldsymbol{X}_i^o, T_i, \Delta_i, \boldsymbol{\alpha}_k)$, e substituir a esperança definida na expressão (2.6) pela média aritmética

$$\mathrm{Q}_i(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{\alpha}_k, \boldsymbol{\theta}_k) = \frac{1}{m_i}\sum_{r=1}^{m_i}\ell_i(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k|\mathcal{D}_{i,(\boldsymbol{z}_r)})$$

em que $\mathcal{D}_{i,(\boldsymbol{z}_r)}$ é o conjunto de dados com as covariáveis omissas substituídas pelo vetor $\boldsymbol{z}_r, r = 1, \ldots, m_i$. A amostra aleatória $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{m_i}$ pode ser gerada usando os amostradores de Gibbs (Geman e Geman, 1984) com o algoritmo de rejeição adaptativa (Gilks e Wild, 1992), pois a distribuição das covariáveis é log-côncava.

O algoritmo Monte Carlo EM tende a sub-estimar a variabilidade dos parâmetros quando a proporção de omissões é alta. Além disso, de acordo com Jank (2005) e Booth et al. (2001) o tamanho amostral m_i não deve permanecer fixo, pois a geração de valores para as covariáveis causa um ruído aleatório que pode fazer com que o algoritmo não convirja (veja detalhes no apêndice A).

Ibrahim et al. (1999b) propõe, para o caso em que algumas covariáveis são categorizadas (ou enumeráveis) e outras são contínuas, combinar os procedimentos definidos acima, aplicando o algoritmo EM ponderado para as variáveis discretas e o Monte Carlo EM para as variáveis contínuas.

A variância das estimativas dos parâmetros, se não puder ser obtida diretamente da posteriori, pode ser calculada através do inverso da (matriz de) informação observada de Fisher (Ibrahim, 1990; Lipsitz e Ibrahim, 1996; Ibrahim et al., 1999a). A substituição da função de verossimilhança pela distribuição a posteriori no cálculo da informação de Fisher é adequada pois, espera-se que a informação obtida da amostra domine a informação a

priori e, assim, a variabilidade a posteriori seja muito próxima daquela devida aos dados. A informação de Fisher, por causa das omissões nas covariáveis, deve ser calculada como uma variância condicional. Assim, se $\hat{\boldsymbol{\theta}}$ e $\hat{\boldsymbol{\alpha}}$ são os valores preditos para $\boldsymbol{\theta}$ e $\boldsymbol{\alpha}$ após a convergência do algoritmo, $\nabla_{(\cdot)}$ é o operador primeira derivada, $\nabla^2_{(\cdot,\cdot)}$ é o operador segunda derivada, $\mathcal{D}_{i,(\boldsymbol{z})}$ é o conjunto de dados com as omissões substituidas por $\boldsymbol{z}$ e

$$\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z}) = \left[\begin{array}{c} \nabla_{(\boldsymbol{\theta})} \ln \pi(\boldsymbol{\theta},\boldsymbol{\alpha}|\mathcal{D}_{i,(\boldsymbol{z})}) \\ \nabla_{(\boldsymbol{\alpha})} \ln \pi(\boldsymbol{\theta},\boldsymbol{\alpha}|\mathcal{D}_{i,(\boldsymbol{z})}) \end{array}\right]$$

então a informação de Fisher é igual a

$$\mathcal{I}_{\mathrm{F}}(\boldsymbol{\theta},\boldsymbol{\alpha}) \;=\; \sum_{i=1}^{n} \left\{ \mathrm{E}\left[\mathrm{var}\left(\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right)\right] + \mathrm{var}\left(\mathrm{E}\left[\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right]\right)\right\} \tag{2.7}$$

em que

$$\mathrm{E}\left[\mathrm{var}\left(\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right)\right] \approx -\ddot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}),$$

com

$$\ddot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) = \left[\begin{array}{cc} \nabla^2_{(\boldsymbol{\theta},\boldsymbol{\theta}')}\, \mathrm{Q}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) & \nabla^2_{(\boldsymbol{\theta},\boldsymbol{\alpha}')}\mathrm{Q}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) \\ \nabla^2_{(\boldsymbol{\alpha},\boldsymbol{\theta}')}\, \mathrm{Q}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) & \nabla^2_{(\boldsymbol{\alpha},\boldsymbol{\alpha}')}\mathrm{Q}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) \end{array}\right]$$

e a função $\mathrm{Q}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})$ definida por (2.6). Além disso,

$$\mathrm{var}\left(\mathrm{E}\left[\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right]\right) = \mathrm{E}\left[\mathrm{E}\left[\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right]^2\right] - \mathrm{E}\left[\mathrm{E}\left[\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right]\right]^2$$

e,

$$\mathrm{E}\left[\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right] \approx \dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z}), \quad \mathrm{E}\left[\mathrm{E}\left[\boldsymbol{\Psi}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right]\right] \approx \dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})$$

com

$$\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z}) = \left[\begin{array}{c} \nabla_{(\boldsymbol{\theta})}\ \ell_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\mathcal{D}_{i,(\boldsymbol{z})}) \\ \nabla_{(\boldsymbol{\alpha})}\ \ell_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\mathcal{D}_{i,(\boldsymbol{z})}) \end{array}\right], \quad \dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) = \left[\begin{array}{c} \nabla_{(\boldsymbol{\theta})}\ \mathrm{Q}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) \\ \nabla_{(\boldsymbol{\alpha})}\ \mathrm{Q}_i(\boldsymbol{\theta},\boldsymbol{\alpha}|\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) \end{array}\right]$$

Substituindo estes elementos em (2.7),

$$\mathcal{I}_{\mathrm{F}}(\boldsymbol{\theta},\boldsymbol{\alpha}) \approx \mathcal{I}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) = \sum_{i=1}^{n} \left\{ -\ddot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) + \mathrm{E}\left[\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})'\right] - \dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})\dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})'\right\}$$

que é o resultado obtido por Louis (1982).

No caso em que as covariáveis omissas são discretas,

$$\mathrm{E}\left[\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})'\right] = \sum_{\boldsymbol{z}\in\mathcal{S}_i} w_i(\boldsymbol{z})\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})'$$

e a variância assintótica das estimativas é dada por (Lipsitz e Ibrahim, 1996, p. 919)

$$\mathcal{I}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) = \sum_{i=1}^{n} \left[-\ddot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) + \sum_{\boldsymbol{z}\in\mathcal{S}_i} w_i(\boldsymbol{z})\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})' - \dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})\dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})'\right]$$

em que $w_i(\boldsymbol{z})$ é o peso do algoritmo EM ponderado. No caso em que as covariáveis são contínuas

$$\mathrm{E}\left[\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z})'\right] \approx \frac{1}{m_i}\sum_{k=1}^{m_i}\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z}_k)\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z}_k)'$$

e, neste caso, a variância assintótica é dada por (Ibrahim et al., 1999a, p. 593)

$$\mathcal{I}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) = \sum_{i=1}^{n}\left[-\ddot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) + \frac{1}{m_i}\sum_{k=1}^{m_i}\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z}_k)\dot{\ell}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}|\boldsymbol{z}_k)' - \dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})\dot{\mathrm{Q}}_i(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})'\right].$$

Desta forma, a variância assintótica das estimativas dos parâmetros $(\boldsymbol{\theta},\boldsymbol{\alpha})$ é

$$\mathrm{var}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}}) = \left[\mathcal{I}(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\alpha}})\right]^{-1}$$

Lipsitz e Ibrahim (1996) e Ibrahim et al. (1999a) obtêm o intervalo de credibilidade para os parâmetros através da distribuição Normal. Entretanto, a validade acerca da suposição de normalidade não é discutida nos artigos.

2.1.2. Mecanismos de omissão informativa

Na situação em que o mecanismo de omissão depende das covariáveis, chamado de *omissão informativa*, o uso de métodos que desconsideram tal dependência gera estimativas viciadas para os parâmetros. Uma discussão detalhada deste fato pode ser vista em Little e Rubin (1987) e Rubin (1987). Esta seção apresenta o modelo definido em Ibrahim et al. (1999b).

No caso de omissão informativa é necessário considerar o mecanismo que causa as omissões. Assim, o conjunto de dados é estendido para $\{n,\boldsymbol{T},\boldsymbol{\Delta},\boldsymbol{X},\boldsymbol{R}\}$ com n, $\boldsymbol{T}$, $\boldsymbol{\Delta}$ e $\boldsymbol{X}$ definidos como antes e a matriz $\boldsymbol{R}$ representando o mecanismo de omissão dos dados. Considerando o i-ésimo elemento amostral, $\mathcal{D}_i = \{t_i,\delta_i,\boldsymbol{x}_i,\boldsymbol{r}_i\}$, $\boldsymbol{r}_i$ é um vetor composto por p variáveis indicadoras de modo que $r_{ij}=1$ se x_{ij} é um dado omisso e $r_{ij}=0$ caso contrário. Da mesma forma que na distribuição conjunta das covariáveis é possível definir

$$\mathrm{f}(\boldsymbol{r}_i|t_i,\delta_i,\boldsymbol{x}_i,\boldsymbol{\phi}) = \mathrm{f}(r_{i,p}|r_{i,1},\ldots,r_{i,p-1},t_i,\delta_i,\boldsymbol{x}_i,\phi_p)\cdots\mathrm{f}(r_{i,1}|t_i,\delta_i,\boldsymbol{x}_i,\phi_1)$$

em que cada distribuição condicional pode ser ajustada por uma regressão logística. Além disso, a expressão acima é aproximada por um modelo conjunto log-linear (Agresti, 1990).

A distribuição a posteriori será dada por

$$\pi(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\phi}|\mathcal{D}) \propto \mathrm{L}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\phi}|\mathcal{D})\cdot\pi(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\phi})$$

em que $\mathrm{L}(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\phi}|\mathcal{D})$ é a verossimilhança e $\pi(\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\phi})$ é a priori conjunta de $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ e $\boldsymbol{\phi}$. Considerando que a distribuição do tempo de sobrevivência não depende das omissões e

nem do respectivo mecanismo gerador, o vetor $\boldsymbol{\theta}$ será independente dos vetores $\boldsymbol{\alpha}$ e $\boldsymbol{\phi}$. Neste caso, a verossimilhança é dada por

$$\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\phi}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{f}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} \cdot \mathrm{S}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i} \cdot \mathrm{f}(\boldsymbol{x}_i, \boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{\alpha}, \boldsymbol{\phi})$$

e a priori é dada por

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\phi}) = \pi(\boldsymbol{\theta}) \cdot \pi(\boldsymbol{\alpha}, \boldsymbol{\phi}) = \pi(\boldsymbol{\theta}) \cdot \pi(\boldsymbol{\alpha}|\boldsymbol{\phi}) \cdot \pi(\boldsymbol{\phi}).$$

Se as covariáveis forem independentes do tempo

$$\mathrm{f}(\boldsymbol{x}_i, \boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{\alpha}, \boldsymbol{\phi}) = \mathrm{f}(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\alpha}, \boldsymbol{\phi}) \cdot \mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha}, \boldsymbol{\phi}),$$

e a verossimilhança será

$$\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\phi}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{f}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} \cdot \mathrm{S}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i} \cdot \mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha}, \boldsymbol{\phi}) \cdot \mathrm{f}(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\alpha}, \boldsymbol{\phi}).$$

No cenário MAR (*omissões ao acaso*) o mecanismo gerador das omissões depende apenas das covariáveis observadas. Neste caso, o mecanismo gerador das omissões não depende de $\boldsymbol{\alpha}$ e a verossimilhança se resume a

$$\mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\phi}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{f}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{\delta_i} \cdot \mathrm{S}(t_i|\boldsymbol{x}_i, \boldsymbol{\theta})^{1-\delta_i} \cdot \mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha}) \cdot \mathrm{f}(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\phi}).$$

com $\mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha})$ definida por (2.4).

De modo análogo ao apresentado na seção 2.1.1, no caso em que as covariáveis são discretas (ou enumeráveis) aplica-se o *algoritmo EM ponderado* (Ibrahim, 1990) com os pesos $w_i(\boldsymbol{z})$ definidos por

$$\begin{aligned} w_i(\boldsymbol{z}) &= \mathrm{f}(\boldsymbol{z}|t_i, \delta_i, \boldsymbol{x}_i^o, \boldsymbol{\alpha}_k) \\ &= \frac{\mathrm{f}(t_i|\delta_i, \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\theta}_k)\mathrm{f}(\boldsymbol{z}|\boldsymbol{x}_i^o, \boldsymbol{\alpha}_k, \boldsymbol{\phi}_k)\mathrm{f}(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\phi}_k)}{\sum_{\boldsymbol{z} \in \mathcal{S}_i} \mathrm{f}(t_i|\delta_i, \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\theta}_k)\mathrm{f}(\boldsymbol{z}|\boldsymbol{x}_i^o, \boldsymbol{\alpha}_k, \boldsymbol{\phi}_k)\mathrm{f}(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i^o, \boldsymbol{z}, \boldsymbol{\phi}_k)} \end{aligned}$$

e, no caso em que as covariáveis omissas são contínuas aplica-se o *algoritmo MCEM* (Tanner e Wong, 1987), em que a amostra $\boldsymbol{z}_1, \ldots, z_{m_i}$ é gerada da distribuição das covariáveis omissas por meio do algoritmo de rejeição adaptativa (Gilks e Wild, 1992) aplicado aos amostradores de Gibbs (Geman e Geman, 1984). Em ambos os casos,

$$\ell_i(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\phi}_k|\boldsymbol{z}) = \ln \mathrm{L}(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\phi}_k|t_i, \delta_i, \boldsymbol{x}_i^o, \mathcal{D}_{i,(\boldsymbol{z})}) + \ln \pi(\boldsymbol{\theta}_k, \boldsymbol{\alpha}_k, \boldsymbol{\phi}_k)$$

e da mesma forma que antes, m_i deve ser atualizado durante a execução do algoritmo (Booth e Hobert, 1999) e os métodos podem ser combinados caso existam covariáveis categorizadas (enumeraveis) e contínuas num mesmo conjunto de dados (Ibrahim et al., 1999b). A variância assintótica das estimativas é igual ao inverso da matriz de informação observada

$$\mathrm{var}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\phi}}) = \left[\mathcal{I}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\phi}})\right]^{-1}$$

calculada de modo análogo ao da omissão não informativa (seção 2.1.1).

2.2. Implementação paramétrica via múltiplos conjuntos

Esta seção apresenta uma implementação da imputação múltipla baseada na geração e análise de múltiplos conjuntos completos de dados. Trata-se de uma adaptação de Rubin e Schenker (1986) para o problema de omissão em covariáveis. Abordagens semelhantes ao do artigo citado são discutidas em Rubin (1980, 1987, 1996); Little e Rubin (1983, 1987); Little (1992) e Liu e Rubin (1994).

Sejam $\{n, \boldsymbol{T}, \boldsymbol{\Delta}, \boldsymbol{X}\}$ e $\{n, \boldsymbol{T}, \boldsymbol{\Delta}, \boldsymbol{X}, \boldsymbol{R}\}$, respectivamente, os conjuntos de dados para omissões não informativas e omissões informativas, em que n é o tamanho amostral, $\boldsymbol{T}$ é o vetor de dimensão n de tempos de acompanhamento (falha ou censura), $\boldsymbol{\Delta}$ é o vetor de dimensão n de indicadores de falha, $\boldsymbol{X}$ é a matriz de dimensão n por p de covariáveis e $\boldsymbol{R}$ é uma matriz de dimensão n por p que representa o mecanismo de omissão. Tal conjunto assume os valores $\mathcal{D} = \{n, \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x}\}$ caso as omissões sejam não informativas ou, caso contrário, assume $\mathcal{D} = \{n, \boldsymbol{t}, \boldsymbol{\delta}, \boldsymbol{x}, \boldsymbol{r}\}$, com tempos de acompanhamento $\boldsymbol{t}' = [t_1, \ldots, t_n]$, indicadores de falha $\boldsymbol{\delta}' = [\delta_1, \ldots, \delta_n]$, matriz de covariáveis $\boldsymbol{x}' = [\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_n]$ e matriz do mecanismo de omissão $\boldsymbol{r}' = [\boldsymbol{r}'_1, \ldots, \boldsymbol{r}'_n]$. O conjunto de dados assume, para a i-ésima unidade amostral, os valores $\mathcal{D}_i = \{t_i, \delta_i, \boldsymbol{x}_i\}$ se as omissões forem não informativas ou $\mathcal{D}_i = \{t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{r}_i\}$ se as omissões forem informativas, com $r_{ij} = 1$ se x_{ij} é um dado omisso ou $r_{ij} = 0$ caso contrário.

Suponha que o vetor de covariáveis da i-ésima unidade amostral $\boldsymbol{x}_i$, particionado em um vetor de covariáveis observadas $\boldsymbol{x}_i^o$ e um vetor de covariáveis omitidas $\boldsymbol{x}_i^*$, seja completado através da substituição das omissões em $\boldsymbol{x}_i^*$ por dados imputados, segundo os processos descritos nas subseções 2.2.1 e 2.2.2. Considere que esta imputação é realizada para todas as unidades amostrais $i = 1, \ldots, n$ e é repetida m vezes, de modo que são obtidas m matrizes de covariáveis completas, denotadas por $\boldsymbol{x}^{(k)}$ com $k = 1, \ldots, m$. Além disso, ao substituir $\boldsymbol{x}$ no conjunto $\mathcal{D}$ por $\boldsymbol{x}^{(k)}$ se obtém o conjunto $\mathcal{D}^{(k)}$. Se $\boldsymbol{\gamma}$ representa um vetor contendo os parâmetros do modelo ($\boldsymbol{\gamma}' = [\boldsymbol{\theta}', \boldsymbol{\alpha}']$ para omissão não informativa e $\boldsymbol{\gamma}' = [\boldsymbol{\theta}', \boldsymbol{\alpha}', \boldsymbol{\phi}']$ para omissão informativa), com distribuição a priori dada por $\pi(\boldsymbol{\gamma})$, então a posteriori é

$$\pi(\boldsymbol{\gamma}|\mathcal{D}) \quad \propto \quad \mathrm{L}(\boldsymbol{\gamma}|\mathcal{D})\pi(\boldsymbol{\gamma}) \tag{2.8}$$

com a verossimilhança, se as covariáveis são independentes do tempo, definida por

$$\mathrm{L}(\boldsymbol{\gamma}|\mathcal{D}) = \mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{f}\left(t_i|\boldsymbol{x}_i, \boldsymbol{\theta}\right)^{\delta_i} \cdot \mathrm{S}\left(t_i|\boldsymbol{x}_i, \boldsymbol{\theta}\right)^{1-\delta_i} \cdot \mathrm{f}\left(\boldsymbol{x}_i|\boldsymbol{\alpha}\right)$$

no caso de *omissão não informativa*, ou definida por

$$\mathrm{L}(\boldsymbol{\gamma}|\mathcal{D}) = \mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\phi}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{f}\left(t_i|\boldsymbol{x}_i, \boldsymbol{\theta}\right)^{\delta_i} \cdot \mathrm{S}\left(t_i|\boldsymbol{x}_i, \boldsymbol{\theta}\right)^{1-\delta_i} \cdot \mathrm{f}\left(\boldsymbol{x}_i|\boldsymbol{\alpha}, \boldsymbol{\phi}\right) \cdot \mathrm{f}\left(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\alpha}, \boldsymbol{\phi}\right)$$

no caso de *omissão informativa*. No caso MAR (*omissões ao acaso*), como visto anteriormente, a verossimilhança se reduz a

$$\mathrm{L}(\boldsymbol{\gamma}|\mathcal{D}) = \mathrm{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\phi}|\mathcal{D}) = \prod_{i=1}^{n} \mathrm{f}\left(t_i|\boldsymbol{x}_i, \boldsymbol{\theta}\right)^{\delta_i} \cdot \mathrm{S}\left(t_i|\boldsymbol{x}_i, \boldsymbol{\theta}\right)^{1-\delta_i} \cdot \mathrm{f}\left(\boldsymbol{x}_i|\boldsymbol{\alpha}\right) \cdot \mathrm{f}\left(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\phi}\right).$$

Logo, é possível estimar o vetor $\boldsymbol{\gamma}$ através da maximização da posteriori (2.8), utilizando o conjunto de dados $\mathcal{D}^{(k)}$ para $k = 1, \ldots, m$. Desta forma, obtém-se o vetor $\hat{\boldsymbol{\gamma}}^{(k)}$ com as estimativas para a k-ésima imputação de dados. Da repetição deste processo, para cada uma das m imputações, obtém-se a seqüência $\hat{\boldsymbol{\gamma}}^{(1)}, \ldots, \hat{\boldsymbol{\gamma}}^{(m)}$ de estimativas para o vetor $\boldsymbol{\gamma}$.

A seqüência $\hat{\boldsymbol{\gamma}}^{(1)}, \ldots, \hat{\boldsymbol{\gamma}}^{(m)}$ pode ser combinada em uma única estimativa $\hat{\boldsymbol{\gamma}}$ pela média aritmética

$$\hat{\boldsymbol{\gamma}} = \frac{1}{m} \sum_{k=1}^{m} \hat{\boldsymbol{\gamma}}^{(k)}$$

e, supondo a independência entre as estimativas $\boldsymbol{\gamma}^{(k)}, k = 1, \ldots, m$, quando m tende ao infinito, de acordo com o Teorema Central do Limite, $\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}$ tem distribuição Normal de média 0 e variância, condicional à distribuição das covariáveis completamente observadas, dada por

$$\mathrm{var}\left(\hat{\boldsymbol{\gamma}}\right) \quad = \quad \mathrm{E}\left[\mathrm{var}\left(\hat{\boldsymbol{\gamma}}|\boldsymbol{X}^o\right)\right] + \mathrm{var}\left(\mathrm{E}\left[\hat{\boldsymbol{\gamma}}|\boldsymbol{X}^o\right]\right) \tag{2.9}$$

em que

$$\mathrm{E}\left[\mathrm{var}\left(\hat{\boldsymbol{\gamma}}|\boldsymbol{X}^o\right)\right] = \mathrm{E}\left[\mathrm{var}\left(\frac{1}{m}\sum_{r=1}^{m}\hat{\boldsymbol{\gamma}}^{(r)}|\boldsymbol{X}^o\right)\right] = \frac{1}{m^2}\sum_{k=1}^{m}\mathrm{E}\left[\mathrm{var}\left(\hat{\boldsymbol{\gamma}}^{(k)}|\boldsymbol{X}^o\right)\right].$$

Note que $\hat{\boldsymbol{\gamma}}^{(1)}, \ldots, \hat{\boldsymbol{\gamma}}^{(m)}$ são identicamente distribuídos, assim, $\mathrm{E}\left[\mathrm{var}\left(\hat{\boldsymbol{\gamma}}^{(k)}|\boldsymbol{X}^o\right)\right]$ é constante e não depende do índice $k, k = 1, \ldots, m$. Desta forma,

$$\mathrm{E}\left[\mathrm{var}\left(\hat{\boldsymbol{\gamma}}|\boldsymbol{X}^o\right)\right] = \frac{1}{m}\mathrm{E}\left[\mathrm{var}\left(\hat{\boldsymbol{\gamma}}^{(k)}|\boldsymbol{X}^o\right)\right]$$

em que $\mathrm{E}\left[\mathrm{var}\left(\hat{\boldsymbol{\gamma}}^{(k)}|\boldsymbol{X}^o\right)\right]$ pode ser estimada por

$$\hat{\boldsymbol{B}} = \widehat{\mathrm{var}}\left(\hat{\boldsymbol{\gamma}}^{(k)}|\boldsymbol{x}^o\right) = \frac{1}{m-1}\sum_{r=1}^{m}\left(\hat{\boldsymbol{\gamma}}^{(r)} - \hat{\boldsymbol{\gamma}}\right)\left(\hat{\boldsymbol{\gamma}}^{(r)} - \hat{\boldsymbol{\gamma}}\right)'$$

a *variância entre imputações*.

De acordo com Rubin e Schenker (1986), o elemento $\mathrm{var}\left(\mathrm{E}\left[\hat{\boldsymbol{\gamma}}|\boldsymbol{X}^o\right]\right)$ é igual a soma da variabilidade *entre imputações* ($\hat{\boldsymbol{B}}$) e da variabilidade *dentro das imputações* estimada por

$$\hat{\boldsymbol{W}} = \frac{1}{m}\sum_{k=1}^{m}\mathcal{I}(\hat{\boldsymbol{\gamma}}^{(k)})$$

em que $\mathcal{I}(\hat{\boldsymbol{\gamma}}^{(k)}), k = 1, \ldots, m$ é a matriz de informação observada. Assim, a expressão (2.9) pode ser estimada por

$$\widehat{\text{var}}\,(\hat{\boldsymbol{\gamma}}) = \boldsymbol{T} = \hat{\boldsymbol{W}} + \frac{m+1}{m}\hat{\boldsymbol{B}}$$

que é uma estimativa da variância assintótica de $\hat{\boldsymbol{\gamma}}$ apresentada em Rubin e Schenker (1986, sec. 1.3 e 1.7).

A aproximação da distribuição do vetor de parâmetros $\hat{\boldsymbol{\gamma}}$ pela Normal é justificada pelo Teorema Central do Limite quando o número de conjuntos m tende ao infinito. Entretanto, quando este valor é pequeno a aproximação não é adequada. Neste caso, a proposta de Rubin e Schenker (1986) é aproximar a distribuição de $\hat{\boldsymbol{\gamma}}$ pela distribuição t de Student

$$\frac{\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}}{\boldsymbol{T}^{1/2}} \sim \text{t}_{\nu}$$

com o número de graus de liberdade dado por

$$\nu = (m-1) \cdot \left[1 + \left(\frac{m}{m+1}\right)\frac{\hat{\boldsymbol{W}}}{\hat{\boldsymbol{B}}}\right]^2.$$

Os detalhes para a obtenção desta expressão são encontrados em Schenker (1985) e Rubin (1986). Note que se a variabilidade entre imputações for muito maior que a variabilidade dentro das imputações ($\boldsymbol{B} \gg \boldsymbol{W}$), o número de graus de liberdade será aproximadamente igual a $m - 1$.

A geração dos conjuntos $\mathcal{D}^{(k)}, k = 1, \ldots, m$ pode ser feita a partir da distribuição a priori de $\boldsymbol{\alpha}$. Para isso, gera-se um vetor $\tilde{\boldsymbol{\alpha}}$ da distribuição $\pi(\boldsymbol{\alpha})$ e, em seguida, gera-se as imputações usando a densidade f $(X_i^*|t_i, \delta_i, \boldsymbol{x}_i^o, \tilde{\boldsymbol{\alpha}})$ das covariáveis omissas. Uma alternativa é usar os métodos empíricos descritos nas seções 2.2.1 e 2.2.2 para gerar as imputações. A geração de amostras das distribuições de probabilidade pode ser feita através do algoritmo de rejeição adaptativa (Gilks e Wild, 1992) aplicado aos amostradores de Gibbs (Geman e Geman, 1984) ou por métodos como o Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970).

2.2.1. Métodos de imputação múltipla para dados discretos

Esta seção apresenta os métodos *Imputação Aleatória Simples*, *Bootstrap Bayesiano* e *Bootstrap Aproximadamente Bayesiano*, descritos em Rubin e Schenker (1986, sec. 2), que podem ser usados para gerar as imputações quando as covariáveis são discretas e quando a distribuição a priori das covariáveis não é especificada no modelo.

Imputação Aleatória Simples: Este método consiste em substituir as omissões nas covariáveis por valores selecionados *com reposição* dentre aquelas unidades amostrais em que

tais covariáveis foram observadas (não omissão). Deste modo, a probabilidade de seleção de cada valor observado permanece constante. No entanto, este método não usa a distribuição a priori do vetor $\boldsymbol{\alpha}$ e, como os valores imputados e observados se repetem, a variabilidade é subestimada. Isto faz com que os intervalos de credibilidade sejam muito pequenos além de descartar todo o conhecimento a priori a respeito dos parâmetros.

Imputação por Bootstrap Bayesiano: O *Bootstrap Bayesiano* é proposto por Rubin (1981). Suponha que uma determinada covariável apresenta um dos z valores $v_1, \ldots, v_z$ com as respectivas probabilidades $\xi_1, \ldots, \xi_z$. Se a distribuição a priori do vetor $\boldsymbol{\xi} = [\ \xi_1 \quad \ldots \quad \xi_z\ ]$ é uma distribuição de Dirichlet imprópria, cuja densidade é

$$\pi(\boldsymbol{\xi}) \propto \prod_{r=1}^{z} \xi_r^{-1}$$

e, além disso, n_r é o número de vezes que o valor v_r é observado, então a distribuição a posteriori de $\boldsymbol{\xi}$ é a distribuição de Dirichlet

$$\pi(\boldsymbol{\xi}|\boldsymbol{x}^o) \propto \prod_{r=1}^{z} \xi_r^{n_r-1}$$

com parâmetros $(n_1 - 1, \ldots, n_z - 1)$. Note que a posteriori também tem distribuição de Dirichlet devido à conjugação (Paulino et al., 2003, p. 89) desta distribuição com relação à distribuição da covariável.

Para gerar os dados para as imputações a partir da posteriori acima, considere que Y_r é uma variável aleatória com distribuição Gama de parâmetros $n_r - 1$ e 1 e que

$$T = \sum_{r=1}^{z} Y_r$$

é uma variável aleatória com distribuição Gama de parâmetros $\sum(n_r - 1)$ e 1. Então, o vetor $\hat{\boldsymbol{\xi}}' = [\ Y_1/T \quad \ldots \quad Y_z/T\ ]$ tem a distribuição de Dirichlet dada pela posteriori acima. Deste modo, gera-se a seqüência $Y_1, \ldots, Y_z$ a partir das respectivas distribuições Gama e calcula-se o vetor $\hat{\boldsymbol{\xi}}$. Os valores para as imputações são selecionados dentre $v_1, \ldots, v_z$ com as probabilidades dadas por $\hat{\boldsymbol{\xi}}$.

Imputação por Bootstrap Aproximadamente Bayesiano: O *Bootstrap Aproximadamente Bayesiano* é uma aproximação para o método acima, que consiste em selecionar, *com reposição*, um número fixo n de valores observados na covariável que se deseja "completar" e, depois, selecionar com reposição dentre tais valores aqueles que serão imputados. Neste caso, quando n se torna muito grande (tende ao infinito) a distribuição das imputações

se aproxima daquela descrita no método anterior. Se dentre os n valores selecionados da covariáveis existem z valores distintos $(v_1, \ldots, v_z)$ cujas proporções (probabilidades) de ocorrência são $p_1, \ldots, p_z$ então a seleção feita pelo *Bootstrap Aproximadamente Bayesiano* é o equivalente à seleção de observações de uma distribuição multinomial com parâmetros $(n, p_1, \ldots, p_z)$.

2.2.2. Métodos de imputação múltipla para dados contínuos

Esta seção apresenta os métodos *Imputação Completamente Normal* e *Imputação Ajustada pela Incerteza na Média e na Variância*, que podem ser usados para gerar as imputações quando as covariáveis são contínuas e quando a distribuição a priori das covariáveis não é especificada no modelo.

Imputação Completamente Normal: Suponha que as observações da covariável em que se deseja fazer imputações são independentes e identicamente distribuidas, com distribuição normal de média μ e variância σ^2. Se a priori de (μ, σ^2) é proporcional a σ^{-2}, então a posteriori de σ^2 é dada por (Box e Tiao, 1973, cap. 2)

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n^o-1}$$

em que s^2 é a variância amostral e a posteriori de μ dado σ^2 será normal de média $\bar{x}$, a média dos valores observados na covariável, e variância σ^2/n^o com n^o representando o número de observações completas na covariável.

Para as imputações, $\tilde{\sigma}^2$ é gerado da distribuição a posteriori de σ^2, $\tilde{\mu}$ é gerado da distribuição a posteriori de μ dado $\sigma^2 = \tilde{\sigma}^2$ e, posteriormente, os valores que serão imputados na covariável são gerados da distribuição $\mathrm{N}(\tilde{\mu}, \tilde{\sigma}^2)$.

Imputação Ajustada pela Incerteza na Média e na Variância: Este método faz com que a forma da distribuição dos valores observados na covariável influencie nas imputações. É uma forma de diminuir o efeito da suposição de normalidade para as covariáveis quando esta não é valida.

Suponha que na covariável existem n^* omissões. Da mesma forma que no método anterior, gera-se os parâmetros $\tilde{\sigma}^2$ e $\tilde{\mu}$. Dentre os valores observados na covariável, seleciona-se com reposição n^* valores e normaliza-se tais valores para que tenham média 0 e variância 1. Desta forma, denotando os valores selecionados por $x^o_1, \ldots, x^o_{n^*}$, obtém-se a seqüência $z_1, \ldots, z_{n^*}$ na qual

$$z_i = \frac{x^o_i - \bar{x}^o}{\sqrt{(n^*-1)s^2_{x^o}/n^*}}, \quad i = 1, \ldots, n^*$$

em que $\bar{x}^o$ é a média e $s^2_{x^o}$ é a variância amostral dos valores selecionados. Para cada uma das n^* omissões imputa-se o valor

$$\tilde{\mu} + \tilde{\sigma} z_i, \quad i = 1, \ldots, n^*.$$

2.3. Imputação para o modelo paramétrico da família Weibull

Denotando $\boldsymbol{\theta} = \{\rho, \boldsymbol{\beta}\}$, o vetor de parâmetros do modelo Weibull, em que ρ é o parâmetro de escala e $\boldsymbol{\beta}$ é o vetor de parâmetros de regressão do tempo de sobrevivência, define-se

$$\lambda_i = \exp(-\boldsymbol{\beta}'\boldsymbol{x}_i), \quad i = 1, \ldots, n$$

o parâmetro de locação associado à i-ésima unidade amostral. Desta forma, a função de sobrevivência para a família Weibull é dada por

$$\mathrm{S}\left(t_i|\boldsymbol{x}_i, \rho, \lambda_i\right) = \exp\left(-(\lambda_i t_i)^\rho\right) = \quad i = 1, \ldots, n$$

e a respectiva densidade de probabilidade é dada por

$$\mathrm{f}\left(t_i|\boldsymbol{x}_i, \rho, \lambda_i\right) = \rho(\lambda_i t_i)^{\rho-1} \exp\left(-(\lambda_i t_i)^\rho\right), \quad i = 1, \ldots, n.$$

Reescrevendo em função dos parâmetros de regressão, tem-se que

$$\begin{aligned} \mathrm{S}\left(t_i|\boldsymbol{x}_i, \rho, \boldsymbol{\beta}\right) &= \exp\left(-\left(\exp(-\boldsymbol{\beta}'\boldsymbol{x}_i)t_i\right)^\rho\right), \quad i = 1, \ldots, n. \\ \mathrm{f}\left(t_i|\boldsymbol{x}_i, \rho, \boldsymbol{\beta}\right) &= \rho\left(\exp(-\boldsymbol{\beta}'\boldsymbol{x}_i)t_i\right)^{\rho-1} \exp\left(-\left(\exp(-\boldsymbol{\beta}'\boldsymbol{x}_i)t_i\right)^\rho\right), \quad i = 1, \ldots, n. \end{aligned}$$

O modelo de Valor-Extremo, usado devido a maior estabilidade dos algoritmos, é obtido com a transformação

$$Y_i = \ln T_i, i = 1, \ldots, n.$$

em que a função de sobrevivência se torna

$$\mathrm{S}\left(y_i|\boldsymbol{x}_i, \rho, \lambda_i\right) = \exp\left(-(\lambda_i \exp(y_i))^\rho\right) = \exp\left(-\exp\left(\rho(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right)\right) \tag{2.10}$$

e a função densidade do tempo de sobrevivência se torna

$$\mathrm{f}\left(y_i|\boldsymbol{x}_i, \rho, \boldsymbol{\beta}\right) = \rho \exp\left(\rho(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right) \exp\left(-\exp\left(\rho(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right)\right). \tag{2.11}$$

Assim, para a implementação dos métodos descritos nas seções 2.1 e 2.2, para a família Weibull de distribuições, foi adotada a função de sobrevivência dada por (2.10) e a função densidade dada por (2.11).

Observe que o modelo Exponencial é um caso particular do modelo Weibull em que a escala é fixada em $\rho = 1$.

2.4. Imputação para o modelo de Cox

Uma abordagem utilizando o modelo de Cox (1972) para lidar com omissões nas covariáveis é apresentado em Leong et al. (2001). Trata-se do modelo de Cox ajustado através do algoritmo EM com o uso de múltiplos conjuntos de dados "completados".

Se os tempos de sobrevivência $\boldsymbol{T}^0$ seguem o modelo de Cox (1972) então sua função de risco é dada por

$$\lambda(t^0|\boldsymbol{x}_i, \boldsymbol{\beta}) = \lambda_0(t^0)\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)$$

em que $\lambda_0(t)$ é a função de risco basal, arbitrária, que depende apenas do tempo de sobrevivência (t^0) e $\boldsymbol{\beta}$ é o vetor de parâmetros de regressão, desconhecidos (neste caso $\boldsymbol{\theta} = \boldsymbol{\beta}$).

A função de risco acumulado é dada por

$$\Lambda(t^0|\boldsymbol{x}_i, \boldsymbol{\beta}) = \int_0^{t^0} \lambda(s|\boldsymbol{x}_i, \boldsymbol{\beta})\,\mathrm{d}s = \exp(\boldsymbol{\beta}'\boldsymbol{x}_i)\Lambda_0(t^0)$$

com

$$\Lambda_0(t^0) = \int_0^{t^0} \lambda_0(s)\,\mathrm{d}s.$$

Segue que a função de sobrevivência é dada por

$$\mathrm{S}(t^0|\boldsymbol{x}_i, \boldsymbol{\theta}) = \mathrm{S}(t^0|\boldsymbol{x}_i, \boldsymbol{\beta}) = \exp\left(-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)\Lambda_0(t^0)\right)$$

e a função densidade de probabilidade do tempo de sobrevivência é dada por

$$\mathrm{f}\left(t^0|\boldsymbol{x}_i, \boldsymbol{\theta}\right) = \mathrm{f}\left(t^0|\boldsymbol{x}_i, \boldsymbol{\beta}\right) = \lambda_0(t^0)\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)\exp\left(-\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)\Lambda_0(t^0)\right).$$

Para descrever as equações de estimação para o modelo de Cox será considerada a notação de processos de contagem (Gill, 1984). Assim, define-se o processo de contagem multivariado $\{\mathrm{N}_i(t) : 0 \leq t < \infty, i = 1, \ldots, n\}$ que é um processo estocástico com n componentes em que $N_i(t)$ é função aleatória de t que vale 1 se a i-ésima unidade amostral falhou até o instante t ou 0 caso contrário, ou seja,

$$\mathrm{N}_i(t) = \mathrm{I}(T_i \leq t, T_i^0 \leq C_i)$$

em que C_i e T_i^0 são, respectivamente, os tempos de censura e de falha da i-ésima unidade amostral, e o processo de intensidade associado a $\mathrm{N}_i(t)$ é

$$\{\Lambda_i(t) : 0 \leq t < \infty; i = 1, \ldots, n\}.$$

Definindo o processo

$$\mathrm{Y}_i(t) = \mathrm{I}(T_i^0 \geq t, C_i \geq t)$$

tal que $\mathrm{Y}_i(t) = 1$ se a i-ésima unidade amostral está em risco no instante imediatamente anterior a t e 0 caso contrário, a verossimilhança parcial (Cox, 1975) pode ser escrita, em termos de processos de contagem, por

$$\mathrm{L}_p(\boldsymbol{\beta}) \;=\; \prod_{t\geq 0}^{*}\prod_{i=1}^{n}\left(\frac{\mathrm{Y}_i(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_i)}{\sum_j \mathrm{Y}_j(t)\exp(\boldsymbol{\beta}'\boldsymbol{x}_j)}\right)^{\mathrm{dN}_i(t)} \tag{2.12}$$

em que $\mathrm{dN}_i(t) = 1$ se a i-ésima unidade amostral falhou em $[t, t+\mathrm{d}t)$ ou 0 caso contrário. Na expressão acima, o primeiro produtório (um *produto-integral*) é um produto finito calculado apenas nos instantes de falha.

De acordo com Cox (1975, sec. 4), se as covariáveis são independentes do tempo, a verossimilhança parcial de Cox é equivalente ao produto de uma função (de verossimilhança) associada ao risco basal e a verossimilhança marginal associada ao vetor de parâmetros $\boldsymbol{\beta}$. Assim, é possível escrever a distribuição a posteriori de $\boldsymbol{\beta}$ como

$$\pi(\boldsymbol{\beta}|\mathcal{D}) \propto \mathrm{L}_p(\boldsymbol{\beta})\pi(\boldsymbol{\beta}).$$

Considere que o vetor de covariáveis associado à i-ésima unidade amostral é definido por $\boldsymbol{X}_i = [X_{i,1}, \ldots, X_{i,p}], i = 1, \ldots, n$, um vetor de dimensão p que se distribui de acordo com a densidade denotada por $\mathrm{f}(\boldsymbol{X}_i|\boldsymbol{\alpha})$, cujo parâmetro $\boldsymbol{\alpha}$ tem distribuição a priori $\pi(\boldsymbol{\alpha})$. Na presença de omissões nas covariáveis tal vetor pode ser particionado como $\boldsymbol{X}_i = [\boldsymbol{X}_i^o, \boldsymbol{X}_i^*]$ com $\boldsymbol{X}_i^*$ um vetor de dimensão q_i de covariáveis omissas e $\boldsymbol{X}_i^o$ um vetor de dimensão $p - q_i$ de variáveis completamente observadas. O mecanismo gerador das omissões é representado pelo vetor $\boldsymbol{R}_i$ cujos elementos assumem os valores $r_{i,j}, i = 1, \ldots, n, j = 1, \ldots, p$ com $r_{i,j} = 1$ se $X_{i,j}$ é um dado omisso ou $r_{i,j} = 0$ se $X_{i,j} = x_{i,j}$. A densidade de probabilidade associada ao processo gerador de omissões é denotada por $\mathrm{f}(\boldsymbol{R}_i|\boldsymbol{\phi}, \ldots)$ e o parâmetro $\boldsymbol{\phi}$ tem distribuição a priori $\pi(\boldsymbol{\phi})$. Assim, se as omissões forem não-informativas a distribuição a posteriori é dada por

$$\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}|\mathcal{D}) \;\propto\; \prod_{i=1}^{n}\mathrm{L}_p(\boldsymbol{\beta})\pi(\boldsymbol{\beta})\mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha})\,\pi(\boldsymbol{\alpha}) \tag{2.13}$$

em que $\boldsymbol{\alpha}$ e $\boldsymbol{\beta}$ são independentes. Por outro lado, se as omissões são informativas, a posteriori será dada por

$$\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\phi}|\mathcal{D}) \;\propto\; \prod_{i=1}^{n}\mathrm{L}_p(\boldsymbol{\beta})\pi(\boldsymbol{\beta})\mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha}, \boldsymbol{\phi})\,\mathrm{f}(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\alpha}, \boldsymbol{\phi})\,\pi(\boldsymbol{\alpha}|\boldsymbol{\phi})\pi(\boldsymbol{\phi}) \tag{2.14}$$

e se as omissões ocorrem ao acaso (*missing at random* – MAR)

$$\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\phi}|\mathcal{D}) \;\propto\; \prod_{i=1}^{n}\mathrm{L}_p(\boldsymbol{\beta})\pi(\boldsymbol{\beta})\mathrm{f}(\boldsymbol{x}_i|\boldsymbol{\alpha})\,\mathrm{f}(\boldsymbol{r}_i|t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{\phi})\,\pi(\boldsymbol{\alpha}|\boldsymbol{\phi})\pi(\boldsymbol{\phi}).$$

No modelo de Cox (1972), a estimação do vetor $\boldsymbol{\beta}$ de parâmetros é feita a partir do logaritmo da verossimilhança parcial (expressão 2.12),

$$\ell(\boldsymbol{\beta}) = \int_0^\infty \sum_{i=1}^n \left[\boldsymbol{\beta}'\boldsymbol{x}_i - \ln \left(\sum_{j=1}^n \mathrm{Y}_j(t) \exp(\boldsymbol{\beta}'\boldsymbol{x}_j) \right) \right] \mathrm{dN}_i(t),$$

cuja derivada com relação a $\boldsymbol{\beta}$ é o vetor escore dado por

$$\mathrm{u}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty [\boldsymbol{x}_i - \mathrm{E}_{\boldsymbol{\beta}}(t;\boldsymbol{\beta})] \ \mathrm{dN}_i(t)$$

com

$$\mathrm{E}_{\boldsymbol{\beta}}(t;\hat{\boldsymbol{\beta}}) = \frac{\sum_j \mathrm{Y}_j(t)\boldsymbol{x}_j \exp(\boldsymbol{\beta}'\boldsymbol{x}_j)}{\sum_j \mathrm{Y}_j(t) \exp(\boldsymbol{\beta}'\boldsymbol{x}_j)}$$

A estimativa de máxima verossimilhança ($\hat{\boldsymbol{\beta}}$) para $\boldsymbol{\beta}$ é o vetor que torna o escore igual a zero ($\mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) = 0$).

De modo análogo, derivando o logaritmo da posteriori (2.13) ou (2.14) com relação a $\boldsymbol{\beta}$, se obtém o vetor escore

$$\mathrm{u}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) \ = \ \sum_{i=1}^n \left\{ \int_0^t [\boldsymbol{x}_i - \mathrm{E}_{\boldsymbol{\beta}}(t;\boldsymbol{\beta})] \ \mathrm{dN}_i(t) \right\} + \frac{\partial \ln \pi(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$$

do qual, fazendo $\mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}) = 0$, se obtém o vetor $\hat{\boldsymbol{\beta}}$ que maximiza a posteriori.

O vetor $\boldsymbol{\alpha}$ é estimado do escore $\mathrm{u}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$, a derivada com relação a $\boldsymbol{\alpha}$ do logaritmo da posteriori (2.13) ou (2.14), tal que o vetor $\hat{\boldsymbol{\alpha}}$ que maximiza a posteriori é obtido igualando-se o escore a zero ($\mathrm{u}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}) = 0$). A estimativa para $\boldsymbol{\phi}$, o vetor que modela o mecanismo de omissão na posteriori (2.14), é obtida analogamente a partir do escore

$$\mathrm{u}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}) \ = \ \frac{\partial \ln \pi(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\phi} | \mathcal{D})}{\partial \boldsymbol{\phi}}.$$

Fazendo $\mathrm{u}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}) = 0$ se obtém as estimativas desejadas.

A estimação da função de risco basal no instante t é obtida do estimador de Breslow (1974)

$$\hat{\lambda}_0(t) \ = \ \frac{\sum_j \mathrm{dN}_i(t)}{\sum_j \mathrm{Y}_i(t) \exp(\hat{\boldsymbol{\beta}}\boldsymbol{x}_j)} \tag{2.15}$$

O modelo de Cox pode ser ajustado pelo método descrito na seção 2.2 usando as prioris (2.13) ou (2.14). O ajuste através do algoritmo EM é feito da seguinte forma:

1. A cada passo do algoritmo, a matriz $\boldsymbol{x}$ de covariáveis é completada substituindo cada omissão por um valor correspondente, gerado da distribuição $\mathrm{f}(\boldsymbol{x}_i^* | \boldsymbol{x}_i^o, \boldsymbol{\alpha})$ das covariáveis omissas para $i = 1, \ldots, n$. Este processo é repetido m vezes de modo que se obtém m matrizes completas $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ de covariáveis.

2. Na k-ésima iteração do algoritmo EM, calcula-se $\hat{\boldsymbol{\beta}}_k$, como solução da equação

$$\frac{1}{m}\sum_{l=1}^{m} \mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_k|\boldsymbol{x}^{(l)}) = 0,$$

$\hat{\boldsymbol{\alpha}}_k$ como solução de

$$\frac{1}{m}\sum_{l=1}^{m} \mathrm{u}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}_k|\boldsymbol{x}^{(l)}) = 0$$

e $\hat{\boldsymbol{\phi}}_k$ como solução de

$$\frac{1}{m}\sum_{l=1}^{m} \mathrm{u}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}_k|\boldsymbol{x}^{(l)}) = 0,$$

em que $\mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_k|\boldsymbol{x}^{(l)})$, $\mathrm{u}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}_k|\boldsymbol{x}^{(l)})$ e $\mathrm{u}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}_k|\boldsymbol{x}^{(l)})$ são, respectivamente, os escores $\mathrm{u}_{\boldsymbol{\beta}}(\boldsymbol{\beta})$, $\mathrm{u}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ e $\mathrm{u}_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ com a matriz de covariáveis dadas por $\boldsymbol{x}^{(l)}, l = 1, \ldots, m$.

3. Estima-se a função de risco basal usando (2.15).

De acordo com Leong et al. (2001, p. 475),

$$\begin{aligned}
\mathrm{var}(\hat{\boldsymbol{\beta}}) &= \left[\frac{1}{m}\sum_{l=1}^{m}\frac{\partial \mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}|\boldsymbol{x}^{(l)})}{\partial\boldsymbol{\beta}} - \frac{1}{m}\sum_{l=1}^{m}\mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}|\boldsymbol{x}^{(l)})\mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}|\boldsymbol{x}^{(l)})'\right]^{-1} \\
\mathrm{var}(\hat{\boldsymbol{\alpha}}) &= \left[\frac{1}{m}\sum_{l=1}^{m}\frac{\partial \mathrm{u}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}|\boldsymbol{x}^{(l)})}{\partial\boldsymbol{\alpha}} - \frac{1}{m}\sum_{l=1}^{m}\mathrm{u}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}|\boldsymbol{x}^{(l)})\mathrm{u}_{\boldsymbol{\alpha}}(\hat{\boldsymbol{\alpha}}|\boldsymbol{x}^{(l)})'\right]^{-1} \\
\mathrm{var}(\hat{\boldsymbol{\phi}}) &= \left[\frac{1}{m}\sum_{l=1}^{m}\frac{\partial \mathrm{u}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}|\boldsymbol{x}^{(l)})}{\partial\boldsymbol{\phi}} - \frac{1}{m}\sum_{l=1}^{m}\mathrm{u}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}|\boldsymbol{x}^{(l)})\mathrm{u}_{\boldsymbol{\phi}}(\hat{\boldsymbol{\phi}}|\boldsymbol{x}^{(l)})'\right]^{-1}
\end{aligned}$$

são estimadores consistentes para as variâncias das estimativas $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$ e $\hat{\boldsymbol{\phi}}$. A obtenção destas expressões é análoga ao feito na seção 2.1.1. A expressão para o vetor $\boldsymbol{\beta}$ é derivada da equação 2.7 em que

$$\sum_{i=1}^{n}\mathrm{E}\left[\mathrm{var}\left(\boldsymbol{\Psi}_i(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right)\right] \approx \frac{1}{m}\sum_{l=1}^{m}\frac{\partial \mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}|\boldsymbol{x}^{(l)})}{\partial\boldsymbol{\beta}}$$

e

$$\sum_{i=1}^{n}\mathrm{var}\left(\mathrm{E}\left[\boldsymbol{\Psi}_i(\boldsymbol{\beta},\boldsymbol{\alpha}|\boldsymbol{z})|\boldsymbol{X}_i^*\right]\right) \approx -\frac{1}{m}\sum_{l=1}^{m}\mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}|\boldsymbol{x}^{(l)})\mathrm{u}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}|\boldsymbol{x}^{(l)})'.$$

As expressões para os vetores $\boldsymbol{\alpha}$ e $\boldsymbol{\phi}$ são obtidas da mesma forma. Detalhes podem ser vistos na página 10.

3. Implementação dos algoritmos

Este capítulo apresenta alguns detalhes a respeito da implementação dos algoritmos descritos no capítulo 2. As aplicações destes modelos podem ser vistas no capítulo 4.

Para ajustar o modelo paramétrico da família Weibull considere que, se $T_i, i = 1, \ldots, n$, são variáveis aleatórias com distribuição Weibull de parâmetros ρ e λ_i,

$$Y_i = \ln T_i, \quad i = 1, \ldots, n$$

tem distribuição Valor-Extremo (VE) com parâmetros ρ e λ_i. Esta transformação se justifica pela estabilidade computacional.

Para construir o modelo de regressão, as covariáveis são associadas ao modelo por

$$\lambda_i = \exp(-\boldsymbol{\beta}'\boldsymbol{x}_i), \quad i = 1, \ldots, n$$

em que $\boldsymbol{\beta}$ é o vetor dos coeficientes de regressão. Além disso, se adotou a transformação $\rho = \exp(-\tau)$ para simplificar a implementação do modelo, pois, deste modo, estimando-se τ a restrição $\rho > 0$ é automaticamente satisfeita. Neste caso, a função de sobrevivência é dada por

$$\mathrm{S}(y_i|\boldsymbol{x}_i, \tau, \boldsymbol{\beta}) = \exp\left(-\exp\left(\mathrm{e}^{\tau}(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right)\right),$$

a respectiva função densidade de probabilidade por

$$\mathrm{f}\left(y_i|\boldsymbol{x}_i, \tau, \boldsymbol{\beta}\right) = \mathrm{e}^{\tau} \exp\left(\mathrm{e}^{\tau}(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right) \exp\left(-\exp\left(\mathrm{e}^{\tau}(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right)\right)$$

e a verossimilhança de $\boldsymbol{\theta} = \{\tau, \boldsymbol{\beta}\}$ é dada por

$$\mathrm{L}(\tau, \boldsymbol{\beta}|\mathcal{D}) \propto \prod_{i=1}^{n} \left[\mathrm{e}^{\tau} \exp\left(\mathrm{e}^{\tau}(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right)\right]^{\delta_i} \exp\left(-\exp\left(\mathrm{e}^{\tau}(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\right)\right).$$

Um fato importante é que, se os vetores $\boldsymbol{\alpha}$ e $\boldsymbol{\theta}$ forem independentes, a distribuição a posteriori é dada por

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) = \mathrm{L}(\boldsymbol{\theta}|\mathcal{D})\mathrm{L}(\boldsymbol{\alpha}|\mathcal{D})\pi(\boldsymbol{\theta})\pi(\boldsymbol{\alpha}),$$

e como é possível fatorar esta expressão como

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathcal{D}) = \pi(\boldsymbol{\theta}|\mathcal{D})\pi(\boldsymbol{\alpha}|\mathcal{D})$$

então os vetores $\boldsymbol{\theta}$ e $\boldsymbol{\alpha}$ são ortogonais. Isto permite estimar tais vetores separadamente.

3.1. Distribuições a priori usadas para a família Weibull

A distribuição a priori, no modelo Valor-Extremo, usada para o parâmetro ρ foi a Gama com hiperparâmetros κ e ν (Ibrahim et al., 2001, p. 38). Assim, τ tem distribuição dada pela densidade

$$\pi(\tau|\kappa,\nu) = \frac{1}{\Gamma(\kappa)}\mathrm{e}^{\tau\kappa}\nu^{\kappa}\exp(-\nu\mathrm{e}^{\tau}).$$

Foram usados na implementação do algoritmo os hiperparâmetros $\kappa = 0,001$ e $\nu = 0,001$, de modo que a priori de ρ tem média 1 e variância 1000. Esta é uma priori que traz pouca informação sobre o parâmetro devido à grande variabilidade. Para o vetor de parâmetros $\boldsymbol{\beta}$ foi adotada a priori Normal p-variada com hiperparâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ (Ibrahim et al., 2001, p. 38),

$$\pi(\boldsymbol{\beta}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left(-\frac{(\boldsymbol{\beta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})}{2}\right).$$

A distribuição conjunta das covariáveis omissas $\boldsymbol{X}_i^*$ é escrita como

$$\mathrm{f}\left(X_{i,q_i}^*|X_{i,q_i-1}^*,\ldots,X_{i,1}^*,\boldsymbol{X}_i^o|\boldsymbol{\alpha}_{q_i},\sigma_{q_i}^2\right)\cdots\mathrm{f}\left(X_{i,1}^*|\boldsymbol{X}_i^o,\boldsymbol{\alpha}_1,\sigma_1^2\right)$$

em que q_i é o número de omissões para o indivíduo i, $(\boldsymbol{\alpha}_j,\sigma_j^2)$ é o vetor de parâmetros associado à j-ésima covariável $(j = 1,\ldots,q_i)$ e cada distribuição condicional acima é uma Normal com média $\boldsymbol{\alpha}_j'\boldsymbol{x}_{i,(j)}$ e variância σ_j^2 com

$$\boldsymbol{x}_{i,(j)} = \left[\begin{array}{cccc} z_{i,j-1} & \cdots & z_{i,1} & \boldsymbol{x}_i^o \end{array}\right].$$

Tal distribuição foi definida desta forma para que as covariáveis omissas sejam modeladas como uma regressão linear simples. Detalhes podem ser vistos em (Paulino et al., 2003). Considerando que a priori para $(\boldsymbol{\alpha}_j,\sigma_j^2)$ é dada por

$$\pi(\boldsymbol{\alpha}_j,\sigma_j^2) \propto \sigma_j^{-2}, \quad j = 1,\ldots,q_i$$

então as distribuições a posteriori são dadas por (Paulino et al., 2003, p. 211)

$$\sigma_j|\boldsymbol{x}_i \sim \mathrm{GI}\left(\frac{n-j}{2}, \sum_{i\in\mathcal{O}(j)}\frac{(x_{i,j}^*-\boldsymbol{x}_{i,(j)}\hat{\boldsymbol{\alpha}}_j)'(x_{i,j}^*-\boldsymbol{x}_{i,(j)}\hat{\boldsymbol{\alpha}}_j)}{2}\right), \quad j = 1,\ldots,q_i$$

uma Gama-Invertida e

$$\boldsymbol{\alpha}_j|\boldsymbol{x}_i,\sigma_j^2 \sim \mathrm{N}_{q_i}\left(\hat{\boldsymbol{\alpha}}_j,\sigma_j^2(\boldsymbol{x}_{(j)}'\boldsymbol{x}_{(j)})^{-1}\right), \quad j = 1,\ldots,q_i$$

uma Normal multivariada. Na notação acima, $z_{i,j}$ representa o valor imputado na j-ésima covariável da i-ésima unidade amostral, $\mathcal{O}(j)$ é o conjunto de todas as unidades amostrais em que a covariável j foi imputada e

$$\hat{\boldsymbol{\alpha}}_j = \sum_{i\in\mathcal{O}(j)}(\boldsymbol{x}_{i,(j)}'\boldsymbol{x}_{i,(j)})^{-1}\boldsymbol{x}_{i,(j)}'z_{i,j}, \quad j = 1,\ldots,q_i$$

Note que, na distribuição Normal o valor que maximiza a densidade de probabilidade é a média. Por sua vez, na distribuição Gama-Invertida com parâmetros a e b o valor de máxima densidade de probabilidade é

$$\frac{b}{a+1}.$$

Desta forma, não será necessário estimar os parâmetros de perturbação através de métodos numéricos. Basta tomar,

$$\boldsymbol{\alpha}_j = \hat{\boldsymbol{\alpha}}_j, j = 1, \ldots, q_i$$

e

$$\sigma_j^2 = \sum_{i \in \mathcal{O}(j)} \frac{(z_{i,j} - \boldsymbol{x}_{i,(j)}\hat{\boldsymbol{\alpha}}_j)'(z_{i,j} - \boldsymbol{x}_{i,(j)}\hat{\boldsymbol{\alpha}}_j)}{n - j + 2}, \quad j = 1, \ldots, q_i.$$

Para a implementação do algoritmo descrito na seção 2.1.1, foi feito o mesmo número de imputações para cada unidade amostral ($m_i = m, i = 1, \ldots, n$). Desta forma, para se construir o conjunto com os dados imputados,

1. Cada indivíduo que *não possui* omissões nas covariáveis é replicado m vezes.

2. Cada indivíduo que *possui* omissões nas covariáveis é replicado m vezes e as omissões são substituídas pelos vetores $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$ gerados como no algoritmo *Monte Carlo EM*.

A matriz de dados resultante tem mn linhas.

A escala do modelo foi estimada separadamente dos parâmetros de regressão. Assim, o algoritmo ECME foi usado ao invés do algoritmo EM. Esta estratégia melhora a convergência do algoritmo, pois estima em separado a escala e os parâmetros de regressão. Tal algoritmo consiste em atualizar o parâmetro de escala e, fixado tal parâmetro, atualiza-se os parâmetros da regressão. Detalhes podem ser consultados no apêndice A.

Para a implementação do método apresentado na seção 2.2, ao invés de se replicar as unidades amostrais para gerar a matriz aumentada, gera-se m matrizes completas de dados e aplica-se o algoritmo de Newton-Raphson para maximizar a posteriori para cada uma destas matrizes. A estimativa final é a média das m estimativas individuais (veja seção 2.1).

3.2. Distribuições a priori usadas para o modelo de Cox

Para o ajuste do modelo de Cox (1972) foi adotada a priori de Bayes-Laplace para o vetor $\boldsymbol{\beta}$ de parâmetros principais. Os parâmetros de perturbação $\boldsymbol{\alpha}$ foram modelados da mesma

forma que no modelo da família Weibull, ou seja, a posteriori é dada por

$$\sigma_j|\boldsymbol{x}_i \sim \text{GI}\left(\frac{n-j}{2}, \sum_{i\in\mathcal{O}(j)} \frac{(x^*_{i,j} - x_{i,(j)}\hat{\boldsymbol{\alpha}}_j)'(x^*_{i,j} - x_{i,(j)}\hat{\boldsymbol{\alpha}}_j)}{2}\right), \quad j = 1, \ldots, q_i$$

e

$$\boldsymbol{\alpha}_j|\boldsymbol{x}_i, \sigma_j^2 \sim \text{N}_{q_i}\left(\hat{\boldsymbol{\alpha}}_j, \sigma_j^2(\boldsymbol{x}'_{(j)}\boldsymbol{x}_{(j)})^{-1}\right)$$

em que

$$\boldsymbol{x}_{i,(j)} = \left[\begin{array}{cccc} z_{i,j-1} & \ldots & z_{i,1} & \boldsymbol{x}_i^o \end{array}\right], \quad \hat{\boldsymbol{\alpha}}_j = \sum_{i\in\mathcal{O}(j)} \left(\boldsymbol{x}'_{i,(j)}\boldsymbol{x}_{i,(j)}\right)^{-1} \boldsymbol{x}'_{i,(j)} z_{i,j},$$

$z_{i,j}$ representa o valor imputado na j-ésima covariável da i-ésima unidade amostral e $\mathcal{O}(j)$ é o conjunto de todas as unidades amostrais em que a covariável j foi imputada.

4. Aplicação e resultados

Os métodos descritos nos capítulos anteriores serão aplicados aos dados estudados por Lima et al. (2005), proveniente de um estudo prospectivo feito pelo Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da USP, com 1343 pacientes, portadores de insuficiência cardíaca causada por diferentes etiologias. Os pacientes foram admitidos no período entre abril de 1991 e junho de 2003 através de um protocolo de avaliação clínica. Foram submetidos a avaliações médicas para verificar as funções cardíacas e definição de tratamento clínico, sendo acompanhados até o óbito (falha), término do estudo ou perda de contato (censuras). Este conjunto de dados apresenta uma elevada taxa de omissão nas covariáveis, pois, dos 1343 pacientes apenas 23 apresentam os dados completos. O conjunto de dados é composto por 31 covariáveis, ou seja, existem mais covariáveis de que unidades amostrais completas. Além disto, 510 pacientes foram censurados (62% de censura).

Para ilustrar os métodos será usado na análise apenas um subconjunto destas covariáveis originais. A descrição do conjunto de dados encontra-se na seção 4.1. A seção 4.2 apresenta o ajuste dos dados de insuficiência cardíaca pelos diferentes métodos, a seção 4.3 apresenta um estudo do efeito da omissão nas estimativas dos modelos paramétricos ajustados e a seção 4.4 apresenta um estudo semelhante para o modelo de Cox.

4.1. Descrição das covariáveis

Neste trabalho serão consideradas as seguintes covariáveis:

1. **Idade** – idade do paciente em anos.
2. **Sexo** – sexo do paciente. O valor 0 indica sexo feminino e 1 sexo masculino.
3. **IMC** – índice de massa corporal em quilogramas por metro quadrado.
4. **Etiologia** – causa da insuficiência cardíaca: 0 - doença de Chagas, 1 - hipertensão, 2 - idiopatia, 3 - isquêmia e 9 - outras etiologias.
5. **HDL** – nível de colesterol HDL (alta densidade) no sangue, em miligramas por decilitro.

6. **LDL** – nível de colesterol LDL (baixa densidade) no sangue, em miligramas por decilitro.

7. **Creatinina** – nível de creatinina, uma proteína excretada pelo rim e que indica o funcionamento do orgão, em miligramas por decilitro.

8. **Fração de ejeção do ventrículo esquerdo** (FE) – porcentagem do sangue recebido pelo coração na diástole (fase de relaxamento do órgão) que é ejetado para a circulação pulmonar durante a sístole (fase de contração).

9. **Fração de ejeção do ventrículo esquerdo por ventriculografia radioisotópica** (FER) – porcentagem do sangue recebido pelo coração na diástole (fase de relaxamento do órgão) que é ejetado para a circulação pulmonar durante a sístole (fase de contração), medido através do exame de ventriculografia radioisotópica.

4.2. Análise dos dados - comparação entre os métodos

As variáveis idade, sexo, etiologia, níveis de colesterol HDL e LDL, nível de creatinina e as duas medidas da fração de ejeção do ventrículo esquerdo foram usadas para a análise dos dados de insuficiência cardíaca. A Tabela 4.1 apresenta a proporção de omissões em cada uma destas variáveis.

Foram considerados, para cada método, diferentes distribuições a priori para o vetor de parâmetros do modelo de sobrevivência. Na análise inicial foi considerada a priori não informativa de Bayes-Laplace,

$$\pi(\boldsymbol{\theta}) \propto 1$$

e, posteriormente, uma priori Normal multivariada para os coeficientes $\boldsymbol{\beta}$ de regressão com média

$$\boldsymbol{\mu}' = \left[\begin{array}{cccccccccccc} 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

e matriz de variância-covariância dada por

$$\boldsymbol{\Sigma} = 100\boldsymbol{I}$$

com $\boldsymbol{I}$ a matriz identidade de dimensão 12 e para a escala do modelo ρ uma priori com média 1 e variância 1000. Estas prioris representam um modelo cuja a média do logaritmo do tempo de sobrevivência é 7, que é uma aproximação ao valor observado na amostra.

Assume-se para todos os modelos que o tempo de sobrevivência tem distribuição Weibull. Tais modelos foram estimados através do algoritmo EM (seção 2.1) e através da

Covariável	Omissões	Porcentagem
Idade	0	0%
Sexo	0	0%
Etiologia	0	0%
Nível de colesterol HDL	631	47%
Nível de colesterol LDL	632	47%
Nível de creatinina	110	8%
Fração de ejeção	387	29%
Fração de ejeção por ventriculografia	323	24%
TOTAL	898	67%

Tabela 4.1.: Omissões nas covariáveis.

imputação em múltiplos conjuntos completos de dados (seção 2.2). Neste último método foram consideradas para a geração das imputações nas covariáveis omissas três diferentes metodologias: imputação geradas diretamente da distribuição preditiva das covariáveis, imputação completamente normal e imputação normal ajustada pela incerteza.

As Tabelas 4.2 e 4.3 apresentam, respectivamente, os ajustes com priori de Bayes-Laplace e Normal-Gama Invertida (ver descrição no capítulo 3) para os parâmetros de regressão. Nestas tabelas, **EM** se refere ao método de estimação através do algoritmo EM, **MC-P**, **MC-N** e **MC-A** se referem ao método de estimação através de múltiplos conjuntos de dados em que as imputações são geradas pela distribuição **P**reditiva das covariáveis, por imputação completamente **N**ormal ou por imputação **A**justada pela incerteza.

Parâmetro	EM	MC-P	MC-N	MC-A
Intercepto	6,1929 (0,3842)	5,8704 (0,1352)	6,2740 (0,1898)	6,1748 (0,2479)
idade	-0,0031 (0,0043)	-0,0099 (0,0017)	-0,0031 (0,0016)	-0,0033 (0,0016)
sexo	-0,1313 (0,1222)	-0,3074 (0,0469)	-0,2216 (0,0544)	-0,2251 (0,0490)
hipertensos	1,2164 (0,1535)	1,2464 (0,0721)	1,3652 (0,0784)	1,3673 (0,0649)
idiopáticos	0,6857 (0,1128)	0,7817 (0,0499)	0,7313 (0,0514)	0,7495 (0,0479)
isquêmicos	0,7393 (0,1433)	0,7985 (0,0689)	0,8258 (0,0659)	0,8349 (0,0565)
outras etiol.	0,8708 (0,1614)	0,9226 (0,0692)	0,9612 (0,0736)	0,9690 (0,0616)
Nível HDL	0,0136 (0,0035)	0,0102 (0,0023)	0,0126 (0,0032)	0,0141 (0,0022)
Nível LDL	0,0022 (0,0012)	0,0023 (0,0008)	0,0027 (0,0014)	0,0028 (0,0015)
creatinina	-0,4268 (0,0956)	-0,0059 (0,0072)	-0,4904 (0,0487)	-0,4732 (0,0468)
FE	0,0333 (0,0059)	0,0457 (0,0028)	0,0380 (0,0044)	0,0390 (0,0050)
FER	0,0024 (0,0064)	0,0039 (0,0011)	0,0021 (0,0053)	0,0013 (0,0040)
Escala	1,13071	1,13608	1,15908	1,15714

Tabela 4.2.: Estimativas (erro padrão) dos parâmetros para o modelo Weibull - parâmetros com priori de Bayes-Laplace.

Na Tabela 4.2 observa-se que as estimativas pontuais dos parâmetros da regressão do tempo de sobrevivência são bastante parecidas entre os diferentes métodos. Note que a estimativa por múltiplos conjuntos completos com imputações geradas da distribuição preditiva (MC-P) é aquela que mais difere das demais e a que apresentou o menor erro padrão. A estimativa com maior erro padrão foi aquela obtiva através do algoritmo EM.

Parâmetro	**EM**	**MC-P**	**MC-N**	**MC-A**
Intercepto	6,3383 (0,0844)	5,8719 (0,1056)	6,2756 (0,1491)	6,1767 (0,2172)
idade	-0,0045 (0,0043)	-0,0099 (0,0015)	-0,0031 (0,0015)	-0,0033 (0,0015)
sexo	-0,1551 (0,0721)	-0,3075 (0,0420)	-0,2217 (0,0500)	-0,2252 (0,0442)
hipertensos	1,4523 (0,0769)	1,2457 (0,0643)	1,3644 (0,0703)	1,3665 (0,0552)
idiopáticos	0,8147 (0,0671)	0,7811 (0,0455)	0,7308 (0,0468)	0,7490 (0,0429)
isquêmicos	0,8998 (0,0744)	0,7980 (0,0632)	0,8252 (0,0593)	0,8344 (0,0489)
outras etiol.	0,9909 (0,0769)	0,9218 (0,0607)	0,9604 (0,0651)	0,9683 (0,0509)
Nível HDL	0,0115 (0,0036)	0,0102 (0,0022)	0,0126 (0,0031)	0,0140 (0,0022)
Nível LDL	0,0016 (0,0013)	0,0023 (0,0008)	0,0027 (0,0014)	0,0028 (0,0015)
creatinina	-0,4784 (0,0644)	-0,0059 (0,0072)	-0,4904 (0,0463)	-0,4732 (0,0449)
FE	0,0391 (0,0064)	0,0457 (0,0028)	0,0380 (0,0043)	0,0390 (0,0050)
FER	0,0027 (0,0060)	0,0039 (0,0011)	0,0020 (0,0052)	0,0013 (0,0039)
Escala	1,16369	1,13604	1,15904	1,15710

Tabela 4.3.: Estimativas (erro padrão) dos parâmetros para o modelo Weibull - parâmetros com priori informativa.

Em comparação com a Tabela 4.3 é possível observar que as estimativas foram pouco afetadas quando se considerou a distribuição a priori informativa para os parâmetros de regressão. No entanto, o método aparentemente mais afetado pelo uso da distribuição a priori informativa foi aquele do algoritmo EM. Porém, o erro padrão de todos os métodos foram sensivelmente menores do que aquele observado na estimativa com priori de Bayes-Laplace. Além disso, as diferenças entre os métodos com relação ao erro padrão é menos perceptível neste caso.

4.3. Efeito das omissões nos diferentes métodos

Esta seção apresenta um estudo de sensibilidade dos diferentes métodos com relação a proporção de omissão em uma dada covariável. Para isso, foram selecionados 500 unidades amostrais dentre os indivíduos que possuiam dados sobre o IMC (Índice de Massa Corporal). Dentre as unidades amostrais selecionadas existem 335 indivíduos censurados (67% de censura). Com tal sub-amostra foi ajustado um modelo Weibull no software `R` (R Development Core Team, 2006) através da rotina `survreg` e verificou-se que a variável idade tinha um efeito significante no tempo de sobrevivência ($p = 0,041$) quando se controlava o sexo, o IMC e a etiologia.

Para o conjunto de dados sem omissões e a priori de Bayes-Laplace para os parâmetros de regressão se obtém a posteriori

$$\pi(\theta|\mathcal{D}) \propto \mathrm{L}(\theta|\mathcal{D})$$

e, por isso, a estimativa que maximiza a distribuição a posteriori coincide com a estimativa de máxima verossimilhança. Por este motivo, o ajuste do modelo Weibull feito pela rotina

`survreg` do software `R` (numa abordagem freqüêntista) foi comparado com o ajuste obtido pelos métodos de estimação através do algoritmo EM e por imputação em múltiplos conjuntos de dados. A priori considerada é a de Bayes-Laplace. O resultado destes ajustes estão na Tabela 4.4.

	Método		
Parâmetro	**R**	**EM**	**MC**
Intercepto	8,1391 (0,6630)	8,0371 (0,5835)	8,1209 (0,2074)
idade	-0,0181 (0,0089)	-0,0162 (0,0078)	-0,0177 (0,0028)
sexo	-0,3559 (0,2335)	-0,3242 (0,2056)	-0,3519 (0,0736)
IMC	0,0248 (0,0232)	0,0221 (0,0206)	0,0244 (0,0074)
hipertensos	0,7915 (0,3281)	0,7173 (0,2871)	0,7806 (0,1019)
idiopáticos	0,6814 (0,2502)	0,6310 (0,2196)	0,6742 (0,0769)
isquêmicos	0,4502 (0,2820)	0,4074 (0,2478)	0,4434 (0,0865)
outras etiol.	0,7595 (0,3426)	0,7044 (0,3014)	0,7520 (0,1007)
escala	1,13315	1,09315	1,12726

Tabela 4.4.: Estimativas com priori de Bayes-Laplace para os parâmetros principais e dados completos (comparação com ajuste do `survreg`).

O ajuste do modelo através da imputação em múltiplos conjuntos completos (MC) foi o que mais se aproximou do ajuste clássico e a diferença entre estes dois ajustes se deve à precisão usada nas respectivas rotinas. Os ajustes através do algoritmo EM apresentam um erro padrão maior que o ajuste por múltiplos conjuntos completos, no entanto, as estimativas são próximas àquelas do modelo clássico (freqüêntista).

Com os 500 indivíduos selecionados para a sub-amostra foram gerados 4 conjuntos de dados, com proporção de omissão crescente para a variável idade. Foram selecionados ao acaso 75 indivíduos que tiveram a idade omitida artificialmente, gerando-se assim um conjunto com 15% de omissão nesta covariável. A partir do conjunto com 15% de omissão foram selecionados outros 75 indivíduos que tiveram a idade propositalmente omitida, gerando-se um conjunto com 30% de omissão e, novamente, deste último conjunto, foram selecionados 150 indivíduos cujas idades também foram removidas para gerar um conjunto com 60% de omissão. Estes 4 conjuntos com proporções 0%, 15%, 30% e 60% foram analisados para comparar os métodos, resultando nas quantidades apresentadas nas Tabelas 4.5 a 4.12.

As Tabelas 4.5 e 4.6 apresentam os ajustes através do algoritmo EM, as Tabelas 4.7 a 4.9 apresentam os ajustes por imputação em múltiplos conjuntos completos com priori de Bayes-Laplace para os parâmetros da regressão e as Tabelas 4.10 a 4.12 apresentam os ajustes por imputação em múltiplos conjuntos completos, com priori informativa para os parâmetros da regressão. A priori usada para os parâmetros da regressão foi a Normal

multivariada com média

$$\boldsymbol{\mu}' = \left[\begin{array}{cccccccc} 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

O número de réplicas foi fixado em $m = 10$.

É importante observar que quando há omissões, as estimativas podem variar entre diferentes execuções do algoritmo pois, como as imputações são aleatórias, se o gerador de números aleatórios for eficiente, em cada execução serão construídos conjuntos de dados diferentes. Porém, é esperado que esta variação seja pequena.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,0371 (0,5835)	8,1150 (0,5820)	8,1246 (0,5792)	7,9587 (0,5930)
idade	-0,0162 (0,0078)	-0,0188 (0,0079)	-0,0196 (0,0079)	-0,0115 (0,0076)
sexo	-0,3242 (0,2056)	-0,3172 (0,2054)	-0,3344 (0,2051)	-0,3408 (0,2048)
IMC	0,0221 (0,0206)	0,0231 (0,0206)	0,0227 (0,0206)	-0,0157 (0,0195)
hipertensos	0,7173 (0,2871)	0,7416 (0,2851)	0,7883 (0,2855)	0,7331 (0,2851)
idiopáticos	0,6310 (0,2196)	0,6312 (0,2179)	0,7051 (0,2164)	0,6861 (0,2161)
isquêmicos	0,4074 (0,2478)	0,4246 (0,2470)	0,4731 (0,2500)	0,3945 (0,2481)
outras etiol.	0,7044 (0,3014)	0,7054 (0,2995)	0,7520 (0,2974)	0,7351 (0,2998)
escala	1,09315	1,09270	1,09399	1,09549

Tabela 4.5.: Estimativas através do algoritmo EM – priori de Bayes-Laplace para os parâmetros principais.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,1006 (0,3680)	7,8826 (0,3607)	8,0504 (0,3625)	7,9110 (0,3674)
idade	-0,0177 (0,0080)	-0,0129 (0,0077)	-0,0178 (0,0071)	-0,0100 (0,0071)
sexo	-0,3562 (0,2043)	-0,3634 (0,2039)	-0,3660 (0,2034)	-0,3946 (0,2027)
IMC	0,0253 (0,0184)	0,0240 (0,0180)	0,0249 (0,0176)	-0,0185 (0,0168)
hipertensos	0,8068 (0,2510)	0,8462 (0,2504)	0,8578 (0,2502)	0,8485 (0,2517)
idiopáticos	0,6950 (0,1951)	0,7333 (0,1941)	0,7544 (0,1935)	0,7557 (0,1943)
isquêmicos	0,4610 (0,2196)	0,4598 (0,2195)	0,5225 (0,2209)	0,4344 (0,2195)
outras etiol.	0,7783 (0,2589)	0,8459 (0,2573)	0,8029 (0,2578)	0,8473 (0,2585)
escala	1,13356	1,13362	1,13304	1,13865

Tabela 4.6.: Estimativas através do algoritmo EM – priori informativa para os parâmetros principais.

Analisando as Tabelas 4.5 e 4.6 observa-se que no método de estimação através do algoritmo EM, mesmo com 60% de omissão, as estimativas para os conjuntos são razoavelmente parecidas com aquelas do conjunto sem omissão. Isto indica que, embora exista um efeito devido às omissões, tal efeito não inviabiliza o uso do método mesmo quando a proporção de omissões é alta. Uma outra observação importante sobre este método é que o erro padrão tende a diminuir conforme a proporção de omissões aumenta mas, como há um efeito aleatório devido às imputações, esta tendência não é tão sensível.

As Tabelas 4.7, 4.8 e 4.9 apresentam o ajuste do modelo através da imputação em múltiplos conjuntos de dados, com priori de Bayes-Laplace para os parâmetros de regressão. É interessante observar que, ao contrário da estimação através do algoritmo EM, o erro padrão tende a aumentar em função da proporção de omissão. Isto é esperado pois os diferentes conjuntos gerados pela imputação de dados devem se tornar cada vez mais distintos em função do aumento da proporção de omissão. Este método apresentou uma menor precisão na análise do conjunto de dados com 60% de omissão, fato este mais perceptível para a Imputação Ajustada pelas Incertezas (Tabela 4.9), indicando uma possível vulnerabilidade ao aumento da proporção de omissão.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,1209 (0,2074)	8,1041 (0,2454)	8,0685 (0,2634)	7,6066 (0,3782)
idade	-0,0177 (0,0028)	-0,0177 (0,0047)	-0,0164 (0,0052)	-0,0026 (0,0078)
sexo	-0,3519 (0,0736)	-0,3482 (0,0762)	-0,3602 (0,0753)	-0,3855 (0,0755)
IMC	0,0244 (0,0074)	0,0244 (0,0076)	0,0231 (0,0077)	0,0179 (0,0071)
hipertensos	0,7806 (0,1019)	0,8123 (0,1067)	0,8235 (0,1059)	0,7873 (0,1062)
idiopáticos	0,6742 (0,0769)	0,6874 (0,0842)	0,7149 (0,0777)	0,7340 (0,0799)
isquêmicos	0,4434 (0,0865)	0,4542 (0,0892)	0,4624 (0,0918)	0,3816 (0,0896)
outras etiol	0,7520 (0,1067)	0,7731 (0,1139)	0,7834 (0,1106)	0,8353 (0,1171)
escala	1,12726	1,12812	1,12848	1,13036

Tabela 4.7.: Estimativas por imputação em múltiplos conjuntos geradas pela distribuição preditiva das covariáveis – priori de Bayes-Laplace para os parâmetros principais.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,1209 (0,2074)	8,1245 (0,2370)	7,9695 (0,2495)	7,6903 (0,3097)
idade	-0,0177 (0,0028)	-0,0182 (0,0042)	-0,0135 (0,0052)	-0,0044 (0,0061)
sexo	-0,3519 (0,0736)	-0,3518 (0,0748)	-0,3638 (0,0756)	-0,3827 (0,0757)
IMC	0,0244 (0,0074)	0,0246 (0,0076)	0,0221 (0,0077)	0,0178 (0,0071)
hipertensos	0,7806 (0,1019)	0,8132 (0,1092)	0,8164 (0,1066)	0,7815 (0,1053)
idiopáticos	0,6742 (0,0769)	0,6822 (0,0810)	0,7159 (0,0809)	0,7319 (0,0794)
isquêmicos	0,4434 (0,0865)	0,4479 (0,0918)	0,4364 (0,0910)	0,3790 (0,0879)
outras etiol.	0,7520 (0,1067)	0,7761 (0,1097)	0,7997 (0,1105)	0,8330 (0,1135)
escala	1,12726	1,12798	1,12896	1,13144

Tabela 4.8.: Estimativas por imputação em múltiplos conjuntos geradas por *Imputação Completamente Normal* – priori de Bayes-Laplace para os parâmetros principais.

As Tabelas 4.10, 4.11 e 4.12 apresentam o ajuste do modelo através da imputação em múltiplos conjuntos de dados, com priori informativa para os parâmetros de regressão. Assim como no ajuste com a priori de Bayes-Laplace, a tendência foi o erro padrão aumentar em função da proporção de omissão e perda de eficiência quando se tem 60% de omissão.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,1209 (0,2074)	8,0989 (0,2342)	8,0286 (0,2771)	7,5035 (0,3694)
idade	-0,0177 (0,0028)	-0,0173 (0,0043)	-0,0149 (0,0056)	0,0000 (0,0081)
sexo	-0,3517 (0,0736)	-0,3449 (0,0749)	-0,3626 (0,0751)	-0,3894 (0,0751)
IMC	0,0244 (0,0074)	0,0238 (0,0077)	0,0225 (0,0074)	0,0175 (0,0071)
hipertensos	0,7806 (0,1019)	0,8148 (0,1059)	0,8117 (0,1084)	0,7814 (0,1069)
idiopáticos	0,6742 (0,0769)	0,6917 (0,0813)	0,7112 (0,0802)	0,7344 (0,0820)
isquêmicos	0,4434 (0,0865)	0,4518 (0,0914)	0,4447 (0,0903)	0,3760 (0,0902)
outras etiol.	0,7520 (0,1067)	0,7751 (0,1101)	0,7898 (0,1116)	0,8526 (0,1145)
escala	1,12726	1,12878	1,12848	1,12995

Tabela 4.9.: Estimativas por imputação em múltiplos conjuntos geradas pela Imputação Ajustada pela Incerteza – com priori de Bayes-Laplace para os parâmetros principais.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,0873 (0,1164)	8,0719 (0,1821)	8,0371 (0,2002)	7,5760 (0,3321)
idade	-0,0175 (0,0025)	-0,0175 (0,0046)	-0,0162 (0,0051)	-0,0024 (0,0077)
sexo	-0,3516 (0,0646)	-0,3480 (0,0675)	-0,3599 (0,0664)	-0,3851 (0,0664)
IMC	0,0249 (0,0058)	0,0249 (0,0061)	0,0236 (0,0061)	0,0184 (0,0055)
hipertensos	0,7957 (0,0794)	0,8271 (0,0856)	0,8380 (0,0845)	0,8020 (0,0843)
idiopáticos	0,6873 (0,0617)	0,7003 (0,0710)	0,7275 (0,0633)	0,7466 (0,0659)
isquêmicos	0,4546 (0,0695)	0,4654 (0,0724)	0,7134 (0,0752)	0,3931 (0,0724)
outras etiol.	0,7697 (0,0819)	0,7905 (0,0919)	0,8005 (0,0874)	0,8525 (0,0956)
escala	1,12759	1,12844	1,12882	1,13081

Tabela 4.10.: Estimativas por imputação em múltiplos conjuntos de dados geradas pela distribuição preditiva das covariáveis – priori informativa para os parâmetros principais.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,0873 (0,1164)	8,0920 (0,1646)	7,9386 (0,1879)	7,6582 (0,2489)
idade	-0,0175 (0,0025)	-0,0180 (0,0041)	-0,0133 (0,0051)	-0,0042 (0,0060)
sexo	-0,3516 (0,0646)	-0,3514 (0,0659)	-0,3635 (0,0667)	-0,3825 (0,0668)
IMC	0,0249 (0,0058)	0,0251 (0,0061)	0,0226 (0,0063)	0,0183 (0,0055)
hipertensos	0,7957 (0,0794)	0,8278 (0,0888)	0,8309 (0,0849)	0,7965 (0,0835)
idiopáticos	0,6873 (0,0617)	0,6951 (0,0670)	0,7285 (0,0672)	0,7446 (0,0653)
isquêmicos	0,4546 (0,0695)	0,4592 (0,0762)	0,4476 (0,0744)	0,3908 (0,0707)
outras etiol.	0,7697 (0,0819)	0,7934 (0,0862)	0,8168 (0,0872)	0,8504 (0,0907)
escala	1,12759	1,12830	1,12933	1,13187

Tabela 4.11.: Estimativas por imputação em múltiplos conjuntos de dados geradas por *Imputação Completamente Normal* – priori informativa para os parâmetros principais.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
Intercepto	8,0873 (0,1164)	8,0664 (0,1597)	7,9971 (0,2189)	7,4722 (0,3254)
idade	-0,0175 (0,0025)	-0,0171 (0,0041)	-0,0147 (0,0055)	0,0002 (0,0080)
sexo	-0,3516 (0,0646)	-0,3448 (0,0661)	-0,3624 (0,0661)	-0,3891 (0,0660)
IMC	0,0249 (0,0058)	0,0243 (0,0062)	0,0229 (0,0059)	0,0179 (0,0054)
hipertensos	0,7957 (0,0794)	0,8295 (0,0846)	0,8263 (0,0875)	0,7965 (0,0850)
idiopáticos	0,6873 (0,0617)	0,7046 (0,0675)	0,7238 (0,0662)	0,7473 (0,0687)
isquêmicos	0,4546 (0,0695)	0,4629 (0,0754)	0,4559 (0,0737)	0,3877 (0,0726)
outras etiol.	0,7697 (0,0819)	0,7926 (0,0870)	0,8070 (0,0887)	0,8698 (0,0926)
escala	1,12759	1,12910	1,12884	1,13042

Tabela 4.12.: Estimativas por imputação em múltiplos conjuntos geradas pela Imputação Ajustada pela Incerteza – com priori informativa para os parâmetros da regressão.

No modelo paramétrico, as estimativas através de múltiplos conjuntos completos de dados se mostraram bastante eficientes, principalmente para proporção de omissão baixa ou moderada, apresentando uma leve sensibilidade às altas proporções de omissão. Este método é o mais indicado para uso prático pois, além de ser o de implementação mais simples, permite que as imputações sejam feitas através de métodos empíricos (ver seções 2.2.1 e 2.2.2). Por este motivo, a análise dos dados é facilitada pois a atribuição de uma distribuição para as covariáveis torna-se desnecessária.

4.4. Ajuste e estudo do efeito de omissões para o modelo de Cox

O modelo de Cox (1972) foi implementado usando o algoritmo EM (seção 2.4) e uma adaptação do modelo de imputação em múltiplos conjuntos (seção 2.2), com *priori* não informativa de Bayes-Laplace para os parâmetros da regressão ($\boldsymbol{\beta}$) e a mesma *priori* usada nos modelos paramétricos para os parâmetros de perturbação, ou seja, normal multivariada com hiperparâmetros estimados a partir das observações completas.

Para verificar a validade das rotinas implementadas foi feita a análise do conjunto completo de dados usados na seção 4.3. Os ajustes do modelo de Cox feitos pela rotina `coxph` do `R` (R Development Core Team, 2006) e através do algoritmo EM (**EM**) e imputação em múltiplos conjuntos (**MC**) são apresentados na Tabela 4.13. As estimativas neste caso são praticamente idênticas.

As Tabelas 4.14 a 4.17 apresentam o estudo de sensibilidade relacionada a alterações na proporção de omissões para o modelo de Cox. Os conjuntos de dados utilizados foram os mesmos do estudo para as distribuições da família Weibull. Para o ajuste do modelo de Cox (1972) foi adotada a priori de Bayes-Laplace para os parâmetros principais e priori informativa (Normal-Gama Invertida, ver cap. 3) para os parâmetros de perturbação.

	Método		
Parâmetro	**R**	**EM**	**MC**
sexo	0,2983 (0,2059)	0,2989 (0,0651)	0,2989 (0,0651)
IMC	-0,0218 (0,0205)	-0,0217 (0,0065)	-0,0217 (0,0065)
hipertensos	-0,7020 (0,2878)	-0,7014 (0,0910)	-0,7014 (0,0910)
idiopáticos	-0,5987 (0,2204)	-0,5984 (0,0697)	-0,5984 (0,0697)
isquêmicos	-0,3996 (0,2485)	-0,3987 (0,0786)	-0,3987 (0,0786)
outras etiol.	-0,6721 (0,3031)	-0,6719 (0,0959)	-0,6719 (0,0959)
idade	0,0159 (0,0078)	0,0158 (0,0025)	0,0158 (0,0025)

Tabela 4.13.: Estimativas para o modelo de Cox e priori de Bayes-Laplace usando dados completos em comparação com ajuste do `coxph`.

Note que tanto a estimativa obtida pelo algoritmo EM (Tabela 4.14) quanto aquelas obtidas através de múltiplos conjuntos completos (Tabelas 4.15 a 4.17) não foram muito afetadas com o aumento de proporção de omissões. Além disso, ao contrário do que ocorreu com os modelos da família Weibull, o ajuste através de múltiplos conjuntos completos se mostrou tão eficiente quanto o ajuste pelo algoritmo EM. Isto deve ter ocorrido pois, neste caso, o algoritmo EM é implementado através de múltiplos conjuntos de dados (Seção 2.4).

É importante observar as diferenças entre os métodos pois, apesar da similaridade, eles não são idênticos. Cada iteração do método de ajuste pelo algoritmo EM consiste em obter uma estimativa dos parâmetros usando todos os múltiplos conjuntos completos como descrito na seção 2.4. Além disso, a estimativa obtida é usada como um "chute" inicial para a iteração seguinte. Por outro lado, o método de ajuste através de múltiplos conjuntos consiste em obter estimativas independentes dos parâmetros do modelo em cada iteração e, no final, combinar todas elas através de uma média aritmética (seção 2.2).

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
sexo	0,2988 (0,0651)	0,2524 (0,0660)	0,2537 (0,0666)	0,3106 (0,0761)
IMC	-0,0217 (0,0065)	-0,0196 (0,0063)	-0,0189 (0,0063)	-0,0159 (0,0062)
hipertensos	-0,7014 (0,0910)	-0,7166 (0,0910)	-0,7243 (0,0913)	-0,6974 (0,0913)
idiopáticos	-0,5984 (0,0697)	-0,6283 (0,0693)	-0,6548 (0,0692)	-0,6509 (0,0693)
isquêmicos	-0,3987 (0,0786)	-0,4046 (0,0789)	-0,4110 (0,0799)	-0,3351 (0,0787)
outras etiol.	-0,6719 (0,0959)	-0,6936 (0,0956)	-0,7212 (0,0952)	-0,7433 (0,0953)
idade	0,0158 (0,0025)	0,0123 (0,0022)	0,0095 (0,0021)	0,0009 (0,0018)

Tabela 4.14.: Estimativas para o modelo de Cox através do método EM – imputação geradas da distribuição preditiva.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
sexo	0,2988 (0,0651)	0,2500 (0,0692)	0,2557 (0,0715)	0,3043 (0,1211)
IMC	-0,0217 (0,0065)	-0,0197 (0,0064)	-0,0182 (0,0064)	-0,0162 (0,0063)
hipertensos	-0,7014 (0,0910)	-0,7100 (0,0919)	-0,7104 (0,0918)	-0,6961 (0,0918)
idiopáticos	-0,5984 (0,0697)	-0,6270 (0,0715)	-0,6439 (0,0697)	-0,6481 (0,0702)
isquêmicos	-0,3987 (0,0786)	-0,3915 (0,0800)	-0,3699 (0,0807)	-0,3349 (0,0794)
outras etiol.	-0,6719 (0,0959)	-0,6230 (0,0981)	-0,7142 (0,0965)	-0,7418 (0,0966)
idade	0,0158 (0,0025)	0,0120 (0,0036)	0,0074 (0,0034)	0,0012 (0,0044)

Tabela 4.15.: Estimativas para o modelo de Cox através do múltiplos conjuntos – imputação geradas da distribuição preditiva.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
sexo	0,2988 (0,0651)	0,2979 (0,0657)	0,3080 (0,0657)	0,3239 (0,0663)
IMC	-0,0217 (0,0065)	-0,0219 (0,0066)	-0,0198 (0,0067)	-0,0161 (0,0063)
hipertensos	-0,7014 (0,0910)	-0,7287 (0,0941)	-0,7304 (0,0934)	-0,6967 (0,0913)
idiopáticos	-0,5984 (0,0697)	-0,6045 (0,0718)	-0,6324 (0,0710)	-0,6441 (0,0700)
isquêmicos	-0,3987 (0,0786)	-0,4010 (0,0816)	-0,3900 (0,0815)	-0,3369 (0,0790)
outras etiol.	-0,6719 (0,0959)	-0,6919 (0,0972)	-0,7113 (0,0975)	-0,7341 (0,0971)
idade	0,0158 (0,0025)	0,0162 (0,0036)	0,0121 (0,0045)	0,0042 (0,0054)

Tabela 4.16.: Estimativas para o modelo de Cox através do múltiplos conjuntos – Imputação Completamente Normal.

	Proporção de omissão			
Parâmetro	0%	15%	30%	60%
sexo	0,2988 (0,0651)	0,2917 (0,0658)	0,3072 (0,0660)	0,3309 (0,0655)
IMC	-0,0217 (0,0065)	-0,0212 (0,0067)	-0,0201 (0,0065)	-0,0159 (0,0062)
hipertensos	-0,7014 (0,0911)	-0,7298 (0,0930)	-0,7257 (0,0927)	-0,6973 (0,0928)
idiopáticos	-0,5984 (0,0697)	-0,6120 (0,0718)	-0,6284 (0,0708)	-0,6470 (0,0711)
isquêmicos	-0,3987 (0,0786)	-0,4047 (0,0820)	-0,3968 (0,0820)	-0,3349 (0,0816)
outras etiol.	-0,6719 (0,0959)	-0,6903 (0,0973)	-0,7025 (0,0984)	-0,7507 (0,0984)
idade	0,0158 (0,0025)	0,0155 (0,0037)	0,0134 (0,0051)	0,0006 (0,0071)

Tabela 4.17.: Estimativas para o modelo de Cox através do múltiplos conjuntos – Imputação Normal Ajustada pela Incerteza.

Parâmetro	**COX-EM**	**COX-P**	**COX-N**	**COX-A**
idade	-0,0069 (0,0015)	-0,0069 (0,0015)	0,0023 (0,0014)	0,0025 (0,0015)
sexo	0,4656 (0,0370)	0,4650 (0,0370)	0,1898 (0,0442)	0,1928 (0,0424)
hipertensos	-1,0151 (0,0504)	-1,0145 (0,0505)	-1,1748 (0,0533)	-1,1800 (0,0628)
idiopáticos	-0,6039 (0,0364)	-0,6036 (0,0364)	-0,6288 (0,0417)	-0,6460 (0,0455)
isquêmicos	-0,5380 (0,0460)	-0,5375 (0,0461)	-0,7115 (0,0512)	-0,7216 (0,0563)
outras etiol.	-0,7531 (0,0518)	-0,7529 (0,0519)	-0,8256 (0,0594)	-0,8358 (0,0596)
Nível HDL	0,0001 (0,0001)	0,0001 (0,0002)	-0,0107 (0,0026)	-0,0120 (0,0020)
Nível LDL	0,0001 (0,0000)	0,0001 (0,0001)	-0,0023 (0,0012)	-0,0024 (0,0013)
creatinina	-0,0106 (0,0010)	-0,0106 (0,0010)	0,4171 (0,0427)	0,4030 (0,0397)
FE	0,0002 (0,0001)	0,0002 (0,0001)	-0,0326 (0,0043)	-0,0333 (0,0045)
FER	0,0001 (0,0000)	0,0001 (0,0000)	-0,0015 (0,0046)	0,0010 (0,0034)

Tabela 4.18.: Estimativas (erro padrão) dos parâmetros do modelo de Cox - parâmetros com priori não informativa para os parâmetros de regressão.

A Tabela 4.18 apresenta o ajuste do modelo de Cox (1972) para os dados de insuficiência cardíaca. Os modelos são implementados por meio do algoritmo EM (**COX-EM**) e as implementações por imputações em múltiplos conjuntos geradas pela distribuição preditiva das covariáveis (**COX-P**), pelas Imputações Completamente Normais (**COX-N**) e pelas Imputações Ajustadas pela Incerteza (**COX-A**).

Observa-se que nos métodos em que as imputações foram geradas pela distribuição preditiva das covariáveis as estimativas são concordantes. Tais métodos são a estimação através do algoritmo EM (**COX-EM**) e através de imputações em múltiplos conjuntos geradas da distribuição preditiva (**COX-P**). Os métodos com imputações geradas empiricamente também concordam entre si (**COX-N** e **COX-A**); entretanto, divergem dos dois outros métodos. É importante observar que as omissões ocorreram completamente ao acaso, pois o esquema adotado para gerá-las independe das covariáveis.

5. Conclusão

A ocorrência de omissão em covariáveis em modelos de análise de sobrevivência é comum em algumas áreas de conhecimento como, por exemplo, em pesquisas clínicas, epidemiológicas e ambientais. O interesse por este problema tem sido crescente na literatura acadêmica, tanto entre freqüêntistas como entre os bayesianos.

Este trabalho apresenta dois métodos encontrados na literatura para o tratamento de omissões em covariáveis para dados de sobrevivência. Ambos os métodos fornecem estimativas pontuais para os parâmetros e uma aproximação assintótica para a distribuição a posteriori. Embora não seja a situação ideal, é uma possibilidade de análise de conjuntos de dados que apresentam omissões.

O primeiro método apresentado (seção 2.1) faz múltiplas imputações de dados, substituindo as omissões por valores gerados das respectivas distribuições de probabilidade, em um único conjunto de dados, que é estendido em um conjunto tantas vezes maior quanto o número de imputações. Este método, no caso paramétrico, se mostrou menos sensível ao aumento da proporção de omissão se comparado com o segundo método abordado. Entretanto, na situação em que as estimativas são comparáveis com as de máxima verossimilhança (freqüêntista) foi o método que apresentou a pior aproximação. Este fato pode ter ocorrido pois, como o conjunto de dados é maior que o original, o algoritmo torna-se mais sensível aos arredondamentos feitos durante o cálculo.

O segundo método apresentado (seção 2.2) também faz múltiplas imputações de dados mas, neste caso, gerando vários conjuntos completos de dados. Estes conjuntos são analisados separadamente e as diversas estimativas são combinadas ao final em uma única estimativa. No ajuste do modelo paramétrico Weibull, este método foi o que mais se aproximou da estimativa de máxima verossimilhança e, no estudo de sensibilidade, foi o mais sofreu a influência da alta proporção de omissões (60% de omissão).

O modelo de Cox (1972) foi implementado usando o algoritmo EM (seção 2.4) e uma adaptação do método de imputação em múltiplos conjuntos de dados (seção 2.2). Foi observado que os métodos que usaram imputações geradas a partir da distribuição preditiva para as covariáveis apresentam estimativas semelhantes e, da mesma forma, isto também

ocorre entre os métodos com imputações geradas empiricamente. No entanto, os métodos com imputações geradas de modo diferente (empírica comparado com distribuição preditiva) são divergentes. Na análise de sensibilidade ao aumento de proporção de omissão todos os métodos se mostraram equivalentes.

A estimação através dos múltiplos conjuntos completos de dados é mais simples de implementar que o algoritmo EM. Além disto, apesar deste método ter mostrado uma leve sensibilidade ao aumento da proporção de omissão, tal método não torna obrigatória a definição da distribuição de probabilidade das covariáveis, pois as imputações podem ser feitas por métodos empíricos. Isto faz com que este método seja mais indicado para aplicações práticas que o método de estimação através do algoritmo EM.

Neste trabalho foi usado um conjunto de dados de insuficiência cardíaca do Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da USP (Lima et al., 2005). Em tal conjunto de dados, as omissões estão relacionadas com um pior prognóstico dos pacientes. Entretanto, as análises foram feitas supondo omissões não informativas (*Missing Completely at Random* - MCAR). Note que o objetivo das análises apresentadas para este conjunto de dados foi apenas a demonstração dos métodos propostos. Uma proposta para análises futuras é a comparação entre os métodos com relação aos diferentes mecanismos de omissão.

É necessário uma metodologia para a definição das prioris, principalmente aquelas do mecanismo gerador das omissões ($\pi(\boldsymbol{\phi})$) e do mecanismo gerador das imputações ($\pi(\boldsymbol{\alpha})$). A definição da priori para os parâmetros da regressão neste trabalho foi baseada no conjunto de dados sem as unidades amostrais que apresentaram omissões (*casos completos*). Trata-se de uma abordagem do tipo Bayes Empírica (ver Paulino et al. (2003)) que pode ser evitada em análises futuras.

Alguns aspectos teóricos precisam ser explorados futuramente. A obtenção dos intervalos de credibilidade à partir da posteriori precisam ser estudada, e as aproximações usadas neste trabalho devem ser evitadas sempre que possível. Porém, em algumas situações pode ser muito difícil gerar tais intervalos diretamente. Neste caso, existe a necessidade de se verificar a suposição de normalidade usada em tais aproximações. Há também a necessidade de se definir métodos para o teste de hipóteses a respeito dos parâmetros de regressão e dos mecanismos gerador de imputações e de omissões.

O comportamento dos métodos com relação à interação entre as proporções de censura e de omissão nas covariáveis e, também, a influência do tamanho amostral na estimativa da variabilidade dos parâmetros precisam ser estudados.

A. Aspectos computacionais

A.1. Descrição dos algoritmos

Algoritmo EM O algoritmo EM (Dempster et al., 1977) é fundamentado apenas em noções simples de teoria de probabilidade e é muito útil para a obtenção de estimativas de máxima verossimilhança para dados incompletos. Se $\boldsymbol{X}^o$ é um conjunto observado, $\boldsymbol{X}^*$ é um conjunto desconhecido (não observado) de dados e θ é um parâmetro de interesse então, iniciando com θ_0 uma estimativa inicial do parâmetro θ, o algoritmo consiste nos seguintes passos

Passo E: (Expectation) Calcula-se a esperança da log-verossimilhança com relação ao vetor $\boldsymbol{X}^*$ de variáveis latentes.

$$Q(\theta|\theta_k) = \mathrm{E}_{\boldsymbol{X}^*|\boldsymbol{X}^o}\left[\ln \mathrm{L}(\theta; \boldsymbol{X}^o, \boldsymbol{X}^*)\right]$$

Passo M: (Maximization) Encontra-se θ_{k+1} tal que

$$\theta_{k+1} = \arg\max Q(\theta|\theta_k)$$

é o argumento que maximiza a esperança calculada no passo **E**.

Repetindo-se o processo acima até a convergência se obtém a estimativa $\hat{\theta}$ de máxima verossimilhança. Neste trabalho a mesma idéia foi usada para maximizar a distribuição a posteriori. Desta forma, a estimativa é a *moda* da distribuição a posteriori.

Um exemplo bastante simples que ilustra o algoritmo é apresentado em Flury e Zoppé (2000, sec. 2). Neste exemplo se obtém uma forma fechada para o passo **M**, ou seja,

$$\theta_{k+1} = \mathrm{f}\left(\theta_k\right)$$

e, assim, a atualização de θ consiste na iteração desta expressão.

Note que nem sempre chega-se a expressões fechadas como acima e, neste caso, a maximização deve ser feita por métodos iterativos como o Newton-Raphson.

Algoritmo ECM O algoritmo ECM (Meng e Rubin, 1993) é uma aceleração do algoritmo EM que consiste em uma modificação da etapa de maximização (Passo M) quando se deseja estimar vetores de parâmetros.

Se $\boldsymbol{\theta}$ é um vetor de parâmetro que se deseja obter a estimativa de máxima verossimilhança então a etapa de maximização do algoritmo EM pode ser feita elemento a elemento do vetor $\boldsymbol{\theta}$ da seguinte forma: atualiza-se o primeiro elemento, atualiza-se o segundo elemento mantendo o primeiro fixo, atualiza-se o terceiro elemento mantendo os dois primeiros fixos, e assim por diante.

O algoritmo EM assim implementado é chamado de ECM (Expectation-Conditional Maximization). A vantagem é que o algoritmo tende a ter uma convergência mais rápida do ponto de vista de tempo de computação se comparado com o EM. O número de iterações pode até mesmo ser maior que o EM mas, como os cálculos são todos feitos com escalares ao invés de matrizes, o custo computacional é bem menor. Este é o algoritmo usado pela maioria dos softwares estatísticos.

Algoritmo ECME O algoritmo ECME (Liu e Rubin, 1994) é uma aceleração proposta ao EM que combina os dois algoritmos anteriories (EM e ECM). Suponha que o vetor $\boldsymbol{\theta}$ de parâmetros pode ser separado em dois vetores $[\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}]$. A k-ésima iteração do algoritmo é definida por

- Encontra-se $\boldsymbol{\theta}_{k+1}^{(1)}$ tal que

$$\boldsymbol{\theta}_{k+1}^{(1)} = \arg\max \mathrm{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}_k^{(1)}, \boldsymbol{\theta}_k^{(2)})$$

  usando o algoritmo EM.

- Encontra-se $\boldsymbol{\theta}_{k+1}^{(2)}$ tal que

$$\boldsymbol{\theta}_{k+1}^{(2)} = \arg\max \mathrm{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}_{k+1}^{(1)}, \boldsymbol{\theta}_k^{(2)})$$

  usando o algoritmo ECM.

Note que a ordem do uso dos algoritmos EM e ECM podem ser permutadas e que os elementos de $\boldsymbol{\theta}$ que compõem os vetores $\boldsymbol{\theta}^{(1)}$ e $\boldsymbol{\theta}^{(2)}$ pode ser definido como se queira.

Este foi o algoritmo usado neste trabalho quando, para o modelo paramétrico de Weibull se fez

$$\boldsymbol{\theta}^{(1)} = \rho, \quad \boldsymbol{\theta}^{(2)} = \boldsymbol{\beta}$$

Algoritmo de Newton-Raphson O algoritmo de Newton-Raphson é um algoritmo iterativo usado para encontrar raízes (zeros) de funções reais. Ele é sempre usado quando existe dificuldade para se obter formas fechadas para as soluções de tais funções reais.

Assim, se $\mathrm{f}(x)$ é uma função da variável x a sua raiz pode ser encontrada iterativamente com os passos definidos por

$$x_{k+1} = x_k - \frac{\mathrm{f}(x)}{\mathrm{f}'(x)}$$

em que x_k é a estimativa de x na k-ésima iteração e $\mathrm{f}'(x)$ é a derivada de $\mathrm{f}(x)$.

Para maximizar a verossimilhança pode-se aplicar o algoritmo como

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - [\mathcal{I}(\boldsymbol{\theta}_k, \boldsymbol{X})]^{-1}\, \mathrm{u}(\boldsymbol{\theta}_k, \boldsymbol{X})$$

em que $\mathcal{I}(\boldsymbol{\theta}_k, \boldsymbol{X})$ é a matriz Hessiana do logaritmo da verossimilhança cujos elementos são os valores das derivadas de segunda ordem calculadas em $\boldsymbol{\theta}_k$ e $\mathrm{u}(\boldsymbol{\theta}_k, \boldsymbol{X})$ é o vetor escore cujos elementos são os valores das derivadas de primeira ordem do logaritmo da verossimilhança calculados também em $\boldsymbol{\theta}_k$.

As iterações são feitas até que $\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k$ seja menor que uma constante fixa ou, se não houver convergência, até que se atinja um limite fixado de iterações.

A.2. Atualização do tamanho do amostrador de Gibbs no algoritmo MCEM

O algoritmo conhecido por MCEM (Monte Carlo EM) é uma aproximação ao algoritmo EM cuja etapa do cálculo da esperança (Passo E) é substituido por uma média entre m elementos gerados ao acaso da distribuição de probabilidade das variáveis desconhecidas $\boldsymbol{X}^*$ dado as observadas $\boldsymbol{X}^o$. Assim, obtendo-se uma seqüência $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$ de *imputações* para $\boldsymbol{x}^*$ se maximiza

$$Q(\theta|\theta_k) = \frac{1}{m}\sum_{l=1}^{m} \mathrm{E}_{\boldsymbol{X}^*|\boldsymbol{X}^o}\left[\ln \mathrm{L}(\theta; \boldsymbol{X}^o, \boldsymbol{z}_l)\right]$$

Nos métodos descritos neste trabalho a expressão acima foi ainda separada por unidade amostral, de modo que

$$Q(\theta|\theta_k) = \sum_{i=1}^{n} Q_i(\theta|\theta_k) = \sum_{i=1}^{n} \frac{1}{m_i}\sum_{l=1}^{m_i} \mathrm{E}_{\boldsymbol{X}^*|\boldsymbol{X}^o}\left[\ln \mathrm{L}(\theta; \boldsymbol{X}^o, \boldsymbol{z}_l)\right]$$

permitindo que o número de imputações $m_i, i = 1, \ldots, n$ varie entre diferentes indivíduos.

Neste cenário, Jank (2005); Booth et al. (2001) afirmam que se o valor m_i permanecer fixo durante toda a simulação não haverá convergência no algoritmo MCEM devido a um componente aleatório imbutido no resíduo do modelo. No entanto, este efeito diminui

assintoticamente em função de m_i, assim, um valor grande o suficiente em m_i deve tornar tal efeito aleatório insignificante. Entretanto, esta solução é ineficiente, pois no início da simulação não existe a necessidade de se controlar a magnitude dos resíduos do modelo e, por isso, serão realizados muitos cálculos desnecessários. Booth e Hobert (1999) propõem como solução para tal problema que o tamanho da amostra m_i do amostrador de Gibbs cresça à medida que a simulação evolui. Este tamanho amostral deve ser atualizado sempre que $\boldsymbol{\theta}_{k+1}$ estiver dentro de um elipsoide com aproximadamente $100(1-\gamma)\%$ de credibilidade calculado com base na seguinte distribuição Normal multivariada atribuída a $\boldsymbol{\theta}_{k+1}$:

$$\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k \sim \mathrm{N}(\hat{\boldsymbol{\theta}}_{k+1}, \mathrm{var}(\hat{\boldsymbol{\theta}}_{k+1}|\hat{\boldsymbol{\theta}}_k))$$

em que

$$\mathrm{var}(\hat{\boldsymbol{\theta}}_{k+1}|\hat{\boldsymbol{\theta}}_k) \approx \ddot{Q}\left(\hat{\boldsymbol{\theta}}_{k+1}\right)^{-1} \cdot \mathrm{var}\left(\dot{Q}\left(\hat{\boldsymbol{\theta}}_{k+1}\right)\right) \cdot \ddot{Q}_i\left(\hat{\boldsymbol{\theta}}_{k+1}\right)^{-1}$$

com

$$\begin{aligned}
\dot{Q}\left(\hat{\boldsymbol{\theta}}_{k+1}\right) &= \frac{\partial}{\partial \boldsymbol{\theta}} Q(\boldsymbol{\theta}_{k+1}|\hat{\boldsymbol{\theta}}_k)\Big|_{\boldsymbol{\theta}_{k+1}=\hat{\boldsymbol{\theta}}_{k+1}} \\
\ddot{Q}\left(\hat{\boldsymbol{\theta}}_{k+1}\right) &= \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} Q(\boldsymbol{\theta}_{k+1}|\hat{\boldsymbol{\theta}}_k)\Big|_{\boldsymbol{\theta}_{k+1}=\hat{\boldsymbol{\theta}}_{k+1}}
\end{aligned}$$

e $\mathrm{var}\left(\dot{Q}\left(\hat{\boldsymbol{\theta}}_{k+1}\right)\right)$ é estimada por

$$\widehat{\mathrm{var}}\left(\dot{Q}(\hat{\boldsymbol{\theta}}_{k+1}|\hat{\boldsymbol{\theta}}_k)\right) = \frac{1}{m_i}\sum_{l=1}^{m_i}\left(\omega_{k,l}\frac{\partial}{\partial \boldsymbol{\theta}}\ell_i(\hat{\boldsymbol{\theta}}_k|\boldsymbol{z}_l)\right)\cdot\left(\omega_{k,l}\frac{\partial}{\partial \boldsymbol{\theta}}\ell_i(\hat{\boldsymbol{\theta}}_k|\boldsymbol{z}_l)\right)'$$

e $\omega_{k,l}$ são os pesos usados no algoritmo *importance sampling* se a geração da amostra aleatória usou tal algoritmo ou, se a amostra aleatória $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{m_i}$ for obtida por outros métodos, serão todos iguais a 1. Este algoritmo foi testado pelos autores com os valores $\gamma = 0,25$ e $\nu \in \{3, 4, 5\}$ aplicados à estimação de modelos lineares generalizados.

Referências Bibliográficas

AGRESTI, Alan. **Categorical data analysis**. New York: Wiley, 1990.

BOOTH, J. G.; HOBERT, J. P. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. **Journal of the Royal Statistical Society**, *B*, v. 61, p. 265 – 285, 1999.

BOOTH, J. G.; HOBERT, J. P.; JANK, W. S. A survey of Monte Carlo algorithms for maximizing the likelihood of a two-stage hierarchical model. **Statistical Modelling**, v. 1, p. 333 – 349, 2001.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. MA: Addison-Wesley, 1973.

BRESLOW, N. Covariance analysis of censored survival data. **Biometrics**, v. 30, p. 89 – 99, 1974.

CHEN, Ming-Hui; IBRAHIM, Joseph G. Maximum likelihood methods for cure rate models with missing covariates. **Biometrics**, v. 57, p. 43 – 52, mar. 2001.

COX, D. R. Regression models and life-tables (with discussion). **Journal of the Royal Statistical Society**, *B*, v. 34, n. 2, p. 187 – 220, 1972.

COX, D. R. Partial likelihood. **Biometrika**, v. 62, n. 2, p. 269 – 276, 1975.

DEGROOT, M. H.; GOEL, K. Estimation of the correlation coefficient from a broken random sample. **Annals of Statistics**, v. 8, p. 264 – 278, 1980.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society**, *B*, v. 39, p. 1 – 38, 1977.

EFRON, Bradley. Missing data, imputation, and the bootstrap. **Journal of the American Statistical Association**, v. 89, n. 426, p. 463 – 475, jun. 1994.

FLURY, Bernard; ZOPPÉ, Alice. Exercices in EM. **The American Statistician**, v. 54, n. 3, p. 207 – 209, ago. 2000.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. **IEEE Transactions on Patterns Analysis and Machine Intelligence**, v. 6, p. 721–741, 1984.

GILKS, W. R.; WILD, P. Adaptative rejection sampling for Gibbs sampling. **Applied Statistics**, v. 88, p. 948–993, 1992.

GILL, Richard. Understanding Cox's regression model: a martingale approach. **Journal of the American Statistical Association**, v. 79, n. 386, p. 441 – 447, jun. 1984.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, p. 97–109, 1970.

HERRING, Amy H.; IBRAHIM, Joseph G. Maximum likelihood estimation in random effects cure rate models with nonignorable missing covariates. **Biostatistics**, v. 3, n. 3, p. 387 – 405, 2002.

IBRAHIM, J.; CHEN, M.-H.; LIPSITZ, S.R. Bayesian methods for generalized linear models with covariates missing at random. **The Canadian Journal of Statistics**, v. 30, p. 55–78, 2002.

IBRAHIM, J.G.; CHEN, M.-H.; SINHA, D. **Bayesian survival analysis**. New York: Springer-Verlag, 2001.

IBRAHIM, Joseph G. Incomplete data in generalized linear models. **Journal of the American Statistical Association**, v. 85, p. 765 – 769, 1990.

IBRAHIM, Joseph G.; CHEN, Ming-Hui; LIPSITZ, Stuart R. Monte Carlo EM for missing covariates in parametric regression models. **Biometrics**, v. 55, p. 591 – 596, jun. 1999a.

IBRAHIM, Joseph G.; LIPSITZ, Stuart R.; CHEN, Ming-Hui. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. **Journal of the Royal Statistical Society**, *B*, v. 61, p. 173 – 190, 1999b.

JANK, Wolfgang. **Stochastic variants of EM: Monte Carlo, Quasi-Monte Carlo and more**. 2005. Available in internet at:
http://www.smith.umd.edu/faculty/wjank/StochasticVariantsOfEMProceedingsJSM2005.pdf.

LEONG, Traci; LIPSITZ, Stuart R.; IBRAHIM, Joseph G. Incomplete covariates in Cox model with applications to biological mark data. **Applied Statistics**, v. 50, n. 4, p. 467 – 484, 2001.

LIMA, Antonio Carlos Pedroso de; MELLO, Guilherme Machado de; FAUSTINO, Wanessa da Silva. **Relatório de análise estatística sobre o projeto: perfil lipídico em pacientes com insuficiência cardíaca**. São Paulo: IME - USP, 2005. Universidade de São Paulo. (Relatório Técnico) RAE-CEA-05P20.

LIN, D. Y.; YING, Z. Cox regression with incomplete covariate measurements. **Journal of the American Statistical Association**, v. 88, n. 424, p. 1341 – 1349, dez. 1993.

LIPSITZ, Stuart R.; IBRAHIM, Joseph G. A conditional model for incomplete covariates in parametric regression models. **Biometrika**, v. 83, n. 4, p. 916 – 922, dez. 1996.

LIPSITZ, Stuart R.; IBRAHIM, Joseph G. Estimating equations with incomplete categorical covariates in the Cox model. **Biometrics**, v. 54, n. 3, p. 1002 – 1013, set. 1998.

LITTLE, R. J. A. Regression with missing X's: a review. **Journal of the American Statistical Association**, v. 81, p. 1227 – 1237, 1992.

LITTLE, R. J. A.; RUBIN, D. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. **Journal of the American Statistical Association**, v. 37, n. 3, p. 218–220, 1983.

LITTLE, R. J. A.; RUBIN, D. **Statistical analysis with missing data**. New York: Wiley, 1987.

LIU, C.; RUBIN, D. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. **Biometrika**, v. 81, n. 4, p. 633–648, 1994.

LOUIS, Thomas A. Finding the observed information matrix when using the EM algorithm. **Journal of the Royal Statistical Society**, *B: Methodological*, v. 44, n. 2, p. 226–233, 1982.

MAGALHÃES, Marcos Nascimento. **Probabilidade e variáveis aleatórias**. 2. ed. São Paulo: EDUSP, 2006.

MENG; RUBIN. Maximum likelihood estimation via the ECM algorithm: a general framework. **Biometrika**, v. 80, n. 2, p. 267 – 278, 1993.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, v. 21, p. 1087–1092, 1953.

PAIK, Myunghee Cho. Multiple imputation for the Cox proportional hazards model with missing covariates. **Lifetime Data Analysis**, v. 3, p. 289 – 298, 1997.

PAIK, Myunghee Cho; TSAI, Wei-Yann. On using the Cox proportional hazards model with missing covariates. **Biometrika**, v. 84, n. 3, p. 579 – 593, set. 1997.

PAULINO, Carlos Daniel; TURKMAN, Maria Antónia Amaral; MURTEIRA, Bento. **Estatística bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003.

PRESS, S. J.; SCOTT, A. J. Missing variables in bayesian regression. **Journal of the American Statistical Association**, v. 71, p. 366 – 369, 1976.

PUGH, M.; ROBBINS, J.; LIPSITZ, S.; HARRINGTON, D. **Inference in the Cox proportional hazards model with missing covariate data**. Boston: Harvard School of Public Health, 1993. Department of Biostatists. (Relatório Técnico).

R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2006.

RUBIN, D. **Multiple imputation for survey nonresponse**. New York: John Wiley, 1986.

RUBIN, D. B.; STERN, H.; VEHOVAR, V. Handling "don't know"survey responses: The case of the slovenian plebiscite. **Journal of the American Statistical Association**, v. 90, p. 822–828, 1994.

RUBIN, Donald B. Inference and missing data. **Biometrika**, v. 63, n. 3, p. 581 – 592, dez. 1976.

RUBIN, Donald B. **Handling non-response in sample surveys by multiple imputation**. [S.l]: US Department of Commerce Bureu of Census Monograph, 1980.

RUBIN, Donald B. The bayesian bootstrap. **The Annals of Statistics**, v. 9, n. 1, p. 130 – 134, jan. 1981.

RUBIN, Donald B. **Multiple imputation for nonresponse in surveys**. New York: Wiley, 1987.

RUBIN, Donald B. Multiple imputation after 18+ years. **Journal of the American Statistical Association**, v. 91, n. 494, p. 473 – 489, jun. 1996.

RUBIN, Donald B.; SCHENKER, Nathaniel. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. **Journal of the American Statistical Association**, v. 81, n. 394, p. 366 – 374, jun. 1986.

SCHENKER, N. **Multiple imputation for interval estimation from surveys with ignorable nonresponse**. 1985. Tese (Doutorado em Estatística) - University of Chicago, 1985.

TANNER, Martin A.; WONG, Wing Hung. The calculation of posterior distributions by data augmentation. **Journal of the American Statistical Association**, v. 82, n. 398, p. 528 – 540, jun. 1987.

THERNEAU, Terry; GRAMBSCH, Patricia. **Modeling Survival Data: Extending the Cox Model**. 1. ed. New York: Springer Verlag, Statistics for Biology and Health, 2000.

WEI, Greg C. G.; TANNER, Martin A. A Monte Carlo implementation of the EM Algorigthm and the Poor Man's data augmentation algotithms. **Journal of the American Statistical Association**, v. 85, n. 411, p. 699 – 704, set. 1990.

ZHOU, H.; PEPE, M. S. Auxiliary covariate data in failure time regression. **Biometrika**, v. 82, p. 139 – 149, 1995.