

MODELO DE REGRESSÃO LOGÍSTICO E EXPONENCIAL POR PARTES
EM ANÁLISE DE SOBREVIVÊNCIA

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sra MARIA BERNADETE SANCHES TERUEL e aprovada pela Comissão Julgadora.

Campinas, 13 de Julho de 1993.

Profa. Dra.

Cicília Yuko Wada
Cicília Yuko Wada OK

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito para obtenção do Título de MESTRE em Estatística.

Universidade Estadual de Campinas
Instituto de Matemática , Estatística e Ciência da Computação - IMECC

MODELOS DE REGRESSÃO LOGÍSTICO E EXPONENCIAL POR PARTES EM
ANÁLISE DE SOBREVIVÊNCIA

Maria Bernadete Sanches Teruel

Prof. Dra. Cícilia Yuko Wada
Orientadora

Campinas - SP
1993

Aos meus pais,
Francisco e Elenice

Agradecimentos

- À Prof. Dra Cicília Yuko Wada, não apenas pela orientação desta dissertação, como também pela amizade e incentivo constante nas horas difíceis.

- Ao Prof. Lúcio Tunes dos Santos, pela ajuda na parte computacional deste trabalho.

- À Rita Helena Antonelli Cardoso, estatística e ao Dr. Pereira Barreto médico do INCOR por cederem os dados, com os quais enriqueceram nosso trabalho.

- De modo especial à minha irmã Glória, e aos amigos Pedro e Estela, Magda e Gilberto pelo incentivo e amizade.

- Aos colegas do GAP (Grupo de Apoio a Informática e Pesquisa) do Instituto de Saúde, pelo apoio e compreensão durante o período de tese.

- E a todos os amigos, professores, parentes e colegas que de alguma forma estiveram vinculados a minha formação.

- Ao CNPQ pelo apoio financeiro para efetivação desta tarefa.

Índice

Introdução	1
Capítulo 1 - Modelos de Regressão Usuais na Análise de Sobrevida	3
1.1. Introdução	3
1.2. Notação e Conceitos Básicos de Análise de Sobrevida	4
1.3. Estimação da Função de Sobrevida	7
1.3.1. Métodos Não Paramétricos	8
1.3.2. Métodos Paramétricos	9
1.4. Estimação da Função de Sobrevida em Modelos de Regressão	10
1.4.1. Modelos Paramétricos	10
1.4.2. Modelos Semi-Paramétrico	11
1.5. Estimadores de Máxima Verossimilhança	11
1.5.1. Método de Newton-Raphson	13
1.5.2. Distribuição Assintótica dos Estimadores de Máxima Verossimilhança	13
1.6. Variância do Estimador da Função de Sobrevida	14
1.7. Testes de Hipóteses	19
1.8. Modelo de Riscos Proporcionais	20
1.8.1. Modelo Exponencial	20
1.8.2. Modelo de Riscos Proporcionais de Cox	22
1.9. Exemplo	24
1.9.1. Resultados	25
Capítulo 2 - Modelo Logísticos em Análise de Sobrevida	34
2.1. Introdução	34
2.2. Modelo Linear Logístico	36
2.2.1. Modelo Logístico	36
2.2.2. Interpretação dos Parâmetros	38
2.3. Modelo Logístico Linear para Análise de Sobrevida	40
2.4. Modelo Logístico Exponencial para Análise de Sobrevida	43
2.5. Modelo Logístico Exponencial por Partes para Análise de Sobrevida	48

2.6. Extensão do Modelo Exponencial Logístico	55
2.7. Aplicação	59
Capítulo 3 - Modelo de Regressão Exponencial por Partes	66
3.1. Introdução	66
3.2. Modelo Exponencial por Partes	68
3.3. Relação entre o Modelo Logístico por Partes e Exponencial por Partes.....	75
3.4. Unificação dos Modelos Exponencial por Partes, Poisson e Multinomial.....	77
3.4.1. Modelo de Regressão de Poisson	77
3.4.2. Modelo de Poisson em Análise de Sobrevida	80
3.4.3. Teorema de Holford	83
3.5. Exemplo	86
Capítulo 4 - Aplicação a um Caso Real.....	90
4.1. Problema	90
4.2. Objetivo	93
4.3. Metodologia de Análise	94
4.4. Resultados	95
Conclusões da Tese	116
Apêndice A - Teoremas	117
Apêndice B - Conjunto de Dados	119
Apêndice C - Resultados	135
Apêndice D - Programas Computacionais.....	146
Referências Bibliográficas	187

várias extensões cuja diferenciação ocorre na natureza das variáveis regressoras incluídas no modelo e na complexidade de sua estrutura matemática.

Um segundo modelo alternativo, é o modelo de regressão exponencial por partes (piecewise exponential, em inglês), que em particular trata da análise de dados de sobrevivência estruturados na forma de tabelas de vida, com covariáveis categóricas.

O objetivo deste trabalho é apresentar esses métodos alternativos de modelos de regressão para análise de sobrevivência, o logístico (e suas extensões) e o exponencial por partes, com o estudo das propriedades da função de sobrevivência associada a esses modelos em termos estatísticos através de estimação e construção de intervalos de confiança. A comparação desses dois modelos também foi objetivo de estudo. A formulação de programas de computação em alguns casos e a descrição do uso de pacotes computacionais já de domínio público como o SAS, foi também realizada, visando complementar os programas no que se refere a erros padrão das estimativas da função de sobrevivência (não fornecidos usualmente).

No capítulo 1 é apresentada uma introdução dos conceitos básicos, e uma revisão dos modelos de Kaplan-Meier [1958] e tabela de vida (não paramétrico) e do modelo de Riscos Proporcionais Exponencial de Glasser [1967] (paramétrico) que são os mais conhecidos nesta área, incluindo modelos de regressão com covariáveis concomitantes e o modelo de riscos proporcionais de Cox [1972]. No capítulo 2, é considerado o estudo de vários modelos de regressão logístico, usado como alternativo, aos modelos descritos no capítulo anterior, com enfoques diferentes, apresentando desde o modelo logístico simples (Elandt-Johnson [1980]), englobando a variável tempo como covariável (Mantel e Hankey [1978]), e ainda quando a distribuição dos tempos é exponencial ou exponencial por partes (Elandt-Johnson [1983]). No capítulo 3 é apresentado o modelo de regressão exponencial por partes (Laird et al [1981]) e seu uso em análise de sobrevivência, quando as covariáveis são categóricas e a sua equivalência com o modelo de regressão de Poisson e regressão multinomial (Holford [1980]). Ainda neste capítulo é apresentado um estudo comparando o modelo logístico exponencial por partes e log-linear exponencial por partes. No capítulo 4, uma aplicação é realizada com os dados cedidos pelo Dr. Pereira Barreto referente ao estudo de endomiocardiofibrose, realizado no Instituto do Coração, Hospital das Clínicas S.P. E finalmente, são apresentadas algumas considerações finais e conclusões desta tese.

INTRODUÇÃO

A **Análise de Sobrevivência** é uma área da estatística aplicada que analisa o tempo observado para a ocorrência de algum evento de interesse. Por exemplo, o tempo de sobrevivência de pacientes submetidos a um determinado tratamento. Esta situação é muito comum entre pesquisadores biomédicos; porém, nos últimos tempos tem chamado a atenção de pesquisadores de diversas áreas como por exemplo engenharia, devido ao grande aparecimento de problemas onde o tempo é a variável de interesse e nessa aplicação, é chamado Análise de Confiabilidade.

Usualmente o objetivo em problemas de Análise de Sobrevivência é prever a probabilidade de sobrevivência, ou tempo médio de vida e comparar a curva de sobrevivência dos indivíduos. A possibilidade de identificar fatores de riscos e inclusão de variáveis concomitantes tem sido alvo mais recente, com o desenvolvimento de novas técnicas e metodologias de estudo. São os modelos de regressão, que em análise de sobrevivência tem sido empregado para avaliar a dependência da variável de interesse, tempo de falhas, sob um conjunto de variáveis regressoras (variáveis concomitantes), com o intuito de prever as probabilidades de sobrevivência. Os modelos mais utilizados foram desenvolvidos nas décadas de 60 e 70 (modelos de regressão paramétrica e modelo de Cox), modelos estes que utilizam tempo como variável dependente. Estes modelos exigem, no entanto, que algumas pressuposições sejam satisfeitas, como por exemplo, algum modelo paramétrico, exponencial ou Weibull para os tempos de vida ou que os riscos sejam proporcionais, respectivamente.

É desejável estudar alguns modelos de regressão alternativos a estes dois modelos mais usuais, para que possam ser utilizados quando as condições desejáveis para o emprego dos dois métodos acima não estejam satisfeitas.

Assim, quando a variável de interesse é do tipo dicotômica, por exemplo, falha ou não falha durante um período de tempo, o modelo de regressão logístico pode ser utilizado como modelo alternativo para análise de dados de sobrevivência. Este modelo será descrito através de

CAPÍTULO 1

MODELOS DE REGRESSÃO USUAIS NA ANÁLISE DE SOBREVIVÊNCIA

1.1. Introdução.

Há situações onde o tempo de ocorrência de um evento é a variável de interesse; tempo de ocorrência é denominado "tempo de vida". Por exemplo, suponha que os eventos de interesse são mortes de indivíduos num sentido real e o "tempo de vida" é o comprimento de vida desse indivíduo; ou, talvez o tempo de sobrevivência desse indivíduo medido após uma condição inicial; por exemplo, início de tratamento com uma determinada droga até a morte. Assim métodos estatísticos para análise de tempos de sobrevivência ou tempos de resposta para um determinado tratamento ou tempos de falha, são chamados de Métodos de Análise de Sobrevivência.

Os principais objetivos da Análise de Sobrevivência são: 1. predição da probabilidade de sobrevivência; 2. estimação do tempo médio de vida e 3. comparação da distribuição de sobrevivência para experimentos com animais ou mesmo de pacientes humanos.

Os dados de sobrevivência apresentam características que criam problemas especiais na análise, por exemplo, no término do período de estudo alguns indivíduos ainda podem estar vivos. O tempo de vida exato para esses indivíduos são desconhecidos, somente é sabido ser maior do que o tempo fixado para o término do experimento. Estas observações são conhecidas

como censuradas e podem ocorrer também quando indivíduos são perdidos de observação, por exemplo pela mudança de endereço. As censuras podem ser classificadas como:

Censura tipo I: ocorre quando o tempo de estudo é fixado e neste tempo, o indivíduo ainda não falhou.

Censura tipo II: ocorre quando o número de falhas é fixo; i. e., dos n itens na amostra somente os r primeiros são consideradas falhas e o restante censuradas, $(1 \leq r \leq n)$.

Censura aleatória: ocorre quando os tempos são censurados de forma aleatórios. Por exemplo, num ensaio clínico pacientes podem entrar no estudo em uma fração mais ou menos aleatória, de acordo com os seus tempos de diagnósticos. Se o estudo termina em uma data pré-fixada, então o tempo de observação é o comprimento entre a entrada de um indivíduo no estudo até o término do estudo é aleatório, e é chamado censura aleatória. A censura do tipo I está incluída em censura aleatória. Exemplo, perdidos de observação, retirados do experimento, término do estudo.

Uma outra característica da Análise de Sobrevida é que permite a inclusão de variáveis concomitantes, ou regressoras no modelo. Por exemplo, suponha que há interesse em se estudar o tempo de sobrevida de pacientes com câncer de pulmão, fatores como idade e condição física do paciente, tipo de tumor e tempo de diagnóstico, devem ser considerados, ignorar tais fatores ao planejar e analisar os dados acima pode obter conclusões incorretas.

Os métodos de análise de sobrevida estão sendo muito utilizados, pois permite incorporar censuras e variáveis concomitantes. O objetivo desse capítulo é apresentar uma revisão sucinta dos modelos de regressão mais utilizados em análise de sobrevida, bem como realizar uma revisão sumária dos métodos de estimação de máxima verossimilhança e suas propriedades assintóticas. Deste modo, seção 1.2 introduz a notação e conceitos básicos. Estimação da função de sobrevida é abordada na seção 1.3 e nas seções 1.4 a 1.7, o método de máxima verossimilhança para estimar os parâmetros, distribuição assintótica destes estimadores e as estatísticas para testes de hipóteses considerando uma função genérica. A seção 1.8 trata do modelo de riscos proporcionais e finalmente na seção 1.9 um exemplo com intuito de ilustrar a teoria apresentada nas seções anteriores.

1.2. Notação e Conceitos Básicos de Análise de Sobrevida.

Primeiramente considera-se o caso de uma única variável T . Se T representa os tempos de falhas de um indivíduo, é uma variável aleatória não negativa usualmente caracterizada por três funções:

a - Função Densidade de Probabilidade (f.d.p.), denotado por $f(t)$ é dado por

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr \left[\text{um indivíduo falhar no intervalo } (t, t + \Delta t) \right]}{\Delta t} \quad (1.1)$$

A f.d.p. tem as seguintes propriedades:

1a. $f(t)$ é uma função não negativa

$$f(t) \geq 0 \text{ para todo } t \geq 0$$

$$f(t) = 0 \text{ para } t < 0$$

2a. A área sobre a curva da densidade e o eixo dos t é igual a 1.

b - Função de Sobrevidência denota-se por $S(t)$, logo

$$S(t) = \Pr (\text{um indivíduo sobreviver ao tempo } t) = \exp \left[- \int_0^t \lambda(u) du \right] \quad (1.2)$$

Note que a função $S(t)$ é contínua e monótona decrescente com $S(0) = 1$ e $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$.

c - Função Risco de Falha Instantâneo denota-se por $\lambda(t)$. A função risco do tempo de sobrevidência t fornece a razão de falha, i.e.

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr \{ t < T < t + \Delta t / T \geq t \}}{\Delta t} = \frac{f(t)}{S(t)} \quad (1.3)$$

A fim de exemplificar, considere:

1. $t \sim \text{Exponencial}(\theta)$

- Função Densidade de Probabilidade (f.d.p.)

$$f(t) = \theta \exp [- \theta t] \text{ para } t > 0 \text{ e } \theta > 0. \quad (1.4)$$

- Função de Sobrevidência correspondente à f.d.p. acima é

$$S(t) = \exp [- \theta t] \text{ , } t > 0 \text{ e } \theta > 0. \quad (1.5)$$

- Função Risco

$$\lambda(t) = \theta \quad t \geq 0 \text{ e } \theta > 0. \quad (1.6)$$

2. $t \sim \text{Weibull}(\alpha, \theta)$

- Função Densidade de Probabilidade (f.d.p.)

$$f(t) = \alpha \theta [\theta t]^{\alpha-1} \exp [- (\theta t)^\alpha] \text{ com } t > 0, \alpha > 0 \text{ e } \theta > 0. \quad (1.7)$$

- Função de Sobrevivência correspondente à f.d.p. acima é

$$S(t) = \exp [- (\theta t)^\alpha] \quad t > 0, \alpha > 0 \text{ e } \theta > 0. \quad (1.8)$$

- Função Risco

$$\lambda(t) = \alpha \theta [\theta t]^{\alpha-1} \quad t \geq 0, \alpha > 0 \text{ e } \theta > 0. \quad (1.9)$$

Porém, na prática tem-se situações onde a população de estudo é heterogênea, tornando importante considerar a relação dos tempos de falha com outros fatores. Seja $Z^T = (z_0, z_1, z_2, \dots, z_p)$, vetor das $p+1$ variáveis concomitantes que são consideradas para predizer falhas. Assim, pode-se reescrever as três funções acima como:

a1 - Função Densidade de Probabilidade (f.d.p.)

$$f(t | z(t)) = \lim_{\Delta t \rightarrow 0} \frac{\text{Pr} [\text{um indivíduo falhar no intervalo } (t, t + \Delta t) | z]}{\Delta t}. \quad (1.10)$$

b1 - Função de Sobrevivência

$$S(t | z(t)) = \text{Pr} [\text{um indivíduo sobreviver ao tempo } t | z] = \exp \left[- \int_0^t \lambda[u | z(u)] du \right]. \quad (1.11)$$

c1 - Função Risco de Falha Instantâneo

$$\lambda (t | z(t)) = \lim_{\Delta t \rightarrow 0} \frac{\text{Pr} \{ t < T < t + \Delta t | T \geq t \}}{\Delta t} = \frac{f(t | z(t))}{S(t | z(t))}. \quad (1.12)$$

Como exemplo, considere os dois modelos já mencionados acima

1a. $t \sim \text{Exponencial}(\theta)$

- Função Densidade de Probabilidade (f.d.p.)

$$f(t | z) = \theta_z \exp [- \theta_z t] \text{ para } t > 0 \text{ e } \theta_z > 0. \quad (1.13)$$

- Função de Sobrevida correspondente à f.d.p. acima é

$$S(t | z) = \exp [- \theta_z t] \quad t > 0 \text{ e } \theta_z > 0. \quad (1.14)$$

- Função Risco

$$\lambda(t | z) = \theta_z \quad t \geq 0 \text{ e } \theta_z > 0. \quad (1.15)$$

2a. $t \sim \text{Weibull} (\alpha, \theta)$ tem-se que:

- Função Densidade de Probabilidade (f.d.p.)

$$f(t | z) = \alpha \theta_z [\theta_z t]^{\alpha-1} \exp [- (\theta_z t)^\alpha] \text{ para } t > 0, \alpha > 0 \text{ e } \theta_z > 0. \quad (1.16)$$

- Função de Sobrevida correspondente à f.d.p. acima é

$$S(t | z) = \exp [- (\theta_z t)^\alpha] \quad t > 0, \alpha > 0 \text{ e } \theta_z > 0. \quad (1.17)$$

- Função Risco

$$\lambda(t | z) = \alpha \theta_z [\theta_z t]^{\alpha-1} \quad t \geq 0, \alpha > 0 \text{ e } \theta_z > 0. \quad (1.18)$$

1.3. Estimação da Função de Sobrevida.

A função de sobrevida $S(t)$ pode ser estimada através de métodos não-paramétricos Produto Limite de Kaplan-Meier e Tabelas de Vida, ou através de métodos paramétricos que consiste na estimação dos parâmetros da função de sobrevida, quando somente os tempos de falha e censura são considerados.

1.3.1. Métodos Não-Paramétricos.

Métodos não-paramétricos devem ser considerados quando não se conhece a distribuição dos tempos, ou a distribuição teórica não se ajusta bem aos dados ou, a procura do modelo "adequado" demanda muito tempo e dinheiro. O método não paramétrico mais conhecido é o Produto-Limite de Kaplan-Meier (PL), e o estimador da função de sobrevivência é dado por:

$$\hat{S}(t) = \prod_{j: t_j < t} \left[1 - \frac{d_j}{n_j} \right] \tag{1.19}$$

onde t_j = tempos de falhas

d_j = número de indivíduos que falharam no tempo j e

n_j = número de indivíduos em risco no tempo j (vivo e não censurado antes de t_j).

O estimador de $S(t)$ possui as seguintes propriedades:

- i. Estimador PL pode ser visto como o estimador de máxima verossimilhança.
- ii. O estimador PL, $\hat{S}(t)$ é fortemente consistente para $S(t)$, i.e.

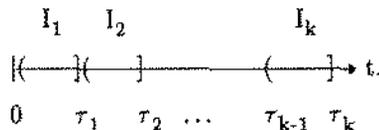
$$\hat{S}(t) \xrightarrow{q.c.} S(t) \text{ ou } \Pr \left[\lim_{n \rightarrow \infty} \hat{S}_n(t) = S(t) \right] = 1.$$

- iii. A distribuição assintótica de $Z_n(t) = \sqrt{n} \left[\hat{S}_n(t) - S(t) \right]$ é normal multivariada com média $\mathbf{0}$ e matriz de variância e covariância dada por

$$\text{Var} \left[\hat{S}(t) \right] \simeq \frac{1}{n} \left[\hat{S}(t) \right]^2 \sum_{j: t_j > t_{n_j}} \frac{d_j}{[n_j - d_j]}, \tag{1.20}$$

pois $Z_n(t)$ é um processo Gaussiano.

Lee [1980] apresentou estimativa de $S(t)$, quando os tempos de falha estão distribuídos em intervalos de classes, i.e., quando o tempo está particionado em uma seqüência de intervalos I_1, \dots, I_k , estes intervalos são quase sempre, mas não necessariamente de comprimentos iguais.



Para estimar a função de sobrevivência usou o método atuário ou tabela de vida. Assim,

$$\hat{S}(\tau_k) = \prod_{i=1}^k \hat{p}_i \tag{1.21}$$

onde

$\bar{p}_i = 1 - \bar{q}_i$ = proporção condicional de sobrevivência no intervalo i .

\bar{q}_i = proporção condicional de falha no intervalo i .

n_i = número de indivíduos vivos no início do intervalo i .

d_i = número de indivíduos que falharam no intervalo i .

l_i = número de indivíduos perdidos de observação no intervalo i .

w_i = número de indivíduos retirados vivos no intervalo i .

$$n_i' = n_i - \frac{1}{2} [l_i - w_i]$$

A variância de $\hat{S}(\tau_k)$ foi derivada por Greenwood [1926] e é

$$\text{Var}[\hat{S}(\tau_k)] \simeq \text{Var}(P_i) \simeq S^2(\tau_k) \sum_{i=1}^k \frac{d_i}{n_i' (n_i' - d_i)} \quad (1.22)$$

1.3.2. Métodos Paramétricos.

Métodos paramétricos são empregados quando a distribuição dos tempos de falha é conhecida. A procura da função distribuição adequada para os tempos de falha se baseia numa abordagem inicial não paramétrica, que fornece evidências de possíveis modelos para os dados, i.e., através das estimativas de $S(t)$ pelos métodos não paramétricos descritos na secção 1.3.1, analisa-se graficamente o ajuste de alguma distribuição paramétrica. Por exemplo; estima-se $\hat{S}(t)$ pelo método de Kaplan-Meier e traçam-se os seguintes gráficos:

1º) $\log \hat{S}(t) * \text{TEMPO}$

2º) $\log [- \log \hat{S}(t)] * \log (\text{TEMPO})$

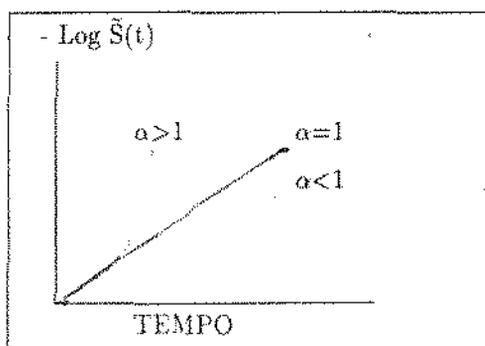


Gráfico 1.

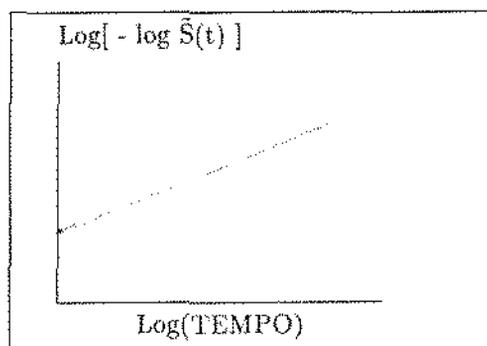


Gráfico 2.

Obs: a expressão log refere-se ao logaritmo neperiano.

1º Caso: gráfico 1, menos o logaritmo de $\hat{S}(t)$ contra tempo, se o resultado mostra uma relação linear, e passar através da origem ($\alpha = 1$) então a distribuição apropriada para t é a exponencial com $\hat{S}(t) = \exp(-\hat{\theta}t)$, onde $\hat{\theta}$ é o estimador de máxima verossimilhança de θ .

2º Caso : se o gráfico 1, menos logaritmo de $\hat{S}(t)$ contra tempo não é linear, i.e., ($\alpha \neq 1$) deve-se continuar a investigar. O próximo passo é traçar o gráfico logaritmo de menos logaritmo de $\hat{S}(t)$ contra logaritmo do tempo e verificar se a relação é linear. Em caso afirmativo a distribuição de t é Weibull com $\hat{S}(t) = \exp(-\hat{\theta}t^{\hat{\alpha}})$, onde $\hat{\theta}$ e $\hat{\alpha}$ é o estimador de máxima verossimilhança de θ e α respectivamente .

1.4. Estimação da Função de Sobrevivência em Modelos de Regressão.

Em situações, onde há interesse em avaliar a relação entre a variável resposta (tempo de falha) com outros fatores (variáveis concomitantes) utiliza-se modelos de regressão. Nesta secção, é realizada uma breve revisão dos modelos mais utilizados. Posteriormente, na secção 1.8, será apresentada uma revisão mais detalhada de modelos de riscos proporcionais.

1.4.1. Modelos Paramétricos.

Considerando t com distribuição exponencial, com f.d.p. dada em (1.13) e função de sobrevivência dada em (1.14), a dependência do tempo de falha sobre as variáveis concomitantes podem ser expressadas através dos seguintes modelos de regressão: Linear Feigl e Zelen [1965], Linear Recíproco Feigl e Zelen [1965] e Log-Linear Glasser [1967].

A tabela abaixo apresenta um resumo, contendo $\lambda(t|z)$ e $S(t|z)$ relativos aos 3 modelos acima, considerando $t \sim \text{Exponencial}[-\theta_z]$

MODELO	θ_z	$\lambda(t z)$	$S(t z)$
1. Linear Feigl e Zelen [1965]	$\beta^T z_i$	$\beta^T z_i$	$\exp\left[-t\left(\beta^T z_i\right)\right]$
2. Linear Recíproco Feigl e Zelen [1965]	$[\beta^T z_i]^{-1}$	$[\beta^T z_i]^{-1}$	$\exp\left[-t\left(\beta^T z_i\right)^{-1}\right]$
3. Log-Linear Glasser [1967]	$\exp[\beta^T z_i]$	$\exp[\beta^T z_i]$	$\exp\left[-t \exp\left(\beta^T z_i\right)\right]$

Note que as funções riscos $\lambda(t | z)$ podem ser desmembradas. Por exemplo,

$$\lambda(t | z) = \exp\left[\underline{\beta}^T z_i\right] = \exp\left[\beta_0\right] \exp\left[\underline{\beta}^{*T} z_i^*\right] = \lambda_0(t) G(z; \underline{\beta}),$$

onde $\lambda_0(t) = \exp\left[\beta_0\right]$ e $G(z; \underline{\beta}) = \exp\left[\underline{\beta}^{*T} z_i^*\right]$, com $\underline{\beta}^{*T} = (\beta_1, \beta_2, \dots, \beta_p)$ e $z_i^* = (z_1, z_2, \dots, z_p)$. Portanto

$$\lambda(t | z) = \lambda_0(t) \exp\left[\underline{\beta}^{*T} z_i^*\right] \quad (1.23)$$

que é conhecido como modelo de riscos proporcionais, pois a razão entre a função de risco relativo às variáveis concomitantes é constante com respeito ao tempo. Existem várias abordagens para analisar este modelo, que tem sido de grande utilidade em problemas envolvendo dados de sobrevivência.

1.4.2. Modelo Semiparamétrico.

Cox [1972] sugeriu que $\lambda_0(t)$ dado em (1.23) fosse considerada uma função de risco arbitrária, i.e., os tempos de falha tem distribuições arbitrárias, ele argumenta que as informações contidas em $\lambda_0(t)$ pode não contribuir sobre $\underline{\beta}$ em intervalos de tempos onde falhas não ocorrem. Desta forma convém considerar apenas a situação sobre o conjunto dos instantes em que falhas ocorrem, ou seja, é conveniente trabalhar com distribuições condicionais. Este modelo tem sido muito usado devido a flexibilidade de não necessitar assumir nenhuma distribuição paramétrica para os tempos de falha.

1.5. Estimadores de Máxima Verossimilhança.

A função de sobrevivência é estimada através da estimação de seus parâmetros. A técnica mais utilizada é o método de máxima verossimilhança. A função de verossimilhança em análise de sobrevivência incorpora a informação de censura já mencionada anteriormente; pois não considerar informações incompletas sobre os tempos de falha pode levar a inferências viciadas ou menos eficientes. Portanto, é importante considerar a natureza do mecanismo de censura atuando sobre os dados quando constrói-se a função de verossimilhança para os parâmetros.

Assumindo que os tempos de censura sejam aleatórios, suponha um estudo onde n

indivíduos são observados e associa-se ao i -ésimo indivíduo as variáveis aleatórias contínuas Y_i (tempo de falha) e W_i (tempo de censura) com função distribuição $F_\theta(\cdot)$ e $H(\cdot)$ respectivamente. As derivadas destas funções fornecem as densidades $f_\theta(\cdot)$ e $h(\cdot)$. Define-se $t_i = \min(Y_i, W_i)$ e $\delta_i = 1$ se $Y_i \leq W_i$ e $\delta_i = 0$ se $Y_i > W_i$, $i = 1, 2, \dots, n$. Lawless [1982], apresentou a f.d.p conjunta de (t_i, δ_i) por

$$\Pr\{t_i = t, \delta_i = 0\} = \Pr\{W_i = t, Y_i > W_i\} = f_\theta(t_i) [1 - H(t_i)]$$

$$\Pr\{t_i = t, \delta_i = 1\} = \Pr\{Y_i = t, Y_i \leq W_i\} = h(t_i) [1 - F_\theta(t_i)]$$

Assim, a função de verossimilhança é

$$L(\theta) = \prod_{i=1}^n [f_\theta(t_i)]^{\delta_i} [1 - F_\theta(t_i)]^{1-\delta_i} [h(t_i)]^{1-\delta_i} [1 - H(t_i)]^{\delta_i}.$$

Supondo independência entre Y_i e W_i , a distribuição de W_i não envolve nenhum parâmetro de interesse. Conseqüentemente, a função de verossimilhança é reescrita considerando apenas os termos envolvendo os parâmetros de interesse

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n [f_\theta(t_i)]^{\delta_i} [1 - F_\theta(t_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [f_\theta(t_i)]^{\delta_i} [S_\theta(t_i)]^{1-\delta_i}. \end{aligned} \tag{1.24}$$

O método de máxima verossimilhança para estimar os parâmetros do modelo implica maximização da equação (1.24). Isto é equivalente a maximizar o logaritmo da função de verossimilhança:

$$\log L = \sum_{i=1}^n \delta_i \log [f_\theta(t_i)] + \sum_{i=1}^n [1 - \delta_i] \log [S_\theta(t_i)] \tag{1.25}$$

e as equações normais

$$\frac{\partial}{\partial \theta_j} \log L(\theta) = 0, \quad j = 1, 2, \dots, p.$$

Normalmente, as equações de máxima verossimilhança não são lineares e portanto necessitam de métodos iterativos. Um método conhecido é o de Newton-Raphson.

1.5.1. Método de Newton-Raphson.

Este método é um procedimento utilizado para encontrar ponto de máximo de uma função. Para usá-lo deve-se admitir um valor inicial e aproximar a função na vizinhança desse valor por um polinômio de segundo grau. Em seguida, usa-se o valor encontrado para aproximar novamente este valor a um polinômio de segundo grau na vizinhança desse segundo valor e assim é gerada uma seqüência de valores que convergirá para o ponto de máximo. Seja $G(\theta)$ uma função com argumentos reais, por exemplo, logaritmo da função de verossimilhança e suponha que deseja-se encontrar o valor de $\hat{\theta}$ que maximiza $G(\theta)$. Nos casos onde as duas primeiras derivadas de $G(\theta)$ existem, i.e., seja $U(\theta)$ um vetor coluna de dimensão p cujos elementos são dados por $\frac{\partial}{\partial \theta_j} G(\theta)$ e $H(\theta)$ uma matriz quadrada de dimensão p cujos elementos são dados por $\frac{\partial^2}{\partial \theta_j \partial \theta_v} G(\theta)$ $j, v = 1, 2, \dots, p$. Seja $\theta^0 = (\theta_1^0, \dots, \theta_p^0)^T$ um vetor coluna de dimensão p cujos componentes são valores próximos dos componentes de θ . Expandindo $U(\theta)$ em série de Taylor em torno de θ^0 tem-se que

$$U(\theta) \simeq U(\theta^0) + (\theta - \theta^0) H(\theta^0). \tag{1.26}$$

Resolvendo a expressão acima em termos de θ , obtém-se para a $(a+1)$ -ésima iteração na seqüência de aproximações para $\hat{\theta}$.

$$\theta^{(a+1)} = \theta^{(a)} - H^{-1}(\theta^{(a)}) U(\theta^{(a)}) \tag{1.27}$$

onde $H(\theta)$ é uma matriz não singular. O processo só termina quando $\theta^{(a+1)}$ e $\theta^{(a)}$ são "bastante próximos". O processo iterativo pode ser iniciado com um vetor $\theta^{(0)}$ formado por valores arbitrários.

1.5.2. Distribuição Assintótica dos Estimadores de Máxima Verossimilhança.

Seja T_1, T_2, \dots, T_n variáveis aleatórias independentes e identicamente distribuídas (i.i.d) com função densidade $f_\theta(t)$, $\theta \in \Theta \subset \mathbb{R}^p$ satisfazendo as seguintes condições de regularidade (Cramér):

(i) - As derivadas $\frac{\partial}{\partial \theta_j} f_\theta(t)$ e $\frac{\partial^2}{\partial \theta_j \partial \theta_v} f_\theta(t)$ existem em quase toda parte e são tais que

$$\left| \frac{\partial}{\partial \theta_j} f_\theta(t) \right| \leq H_j(t) \text{ e } \left| \frac{\partial^2}{\partial \theta_j \partial \theta_v} f_\theta(t) \right| \leq G_{jv}(t) \text{ onde } \int_{\mathbb{R}} H_j(t) dt < \infty, \int_{\mathbb{R}} G_{jv}(t) dt < \infty,$$

para $j, v = 1, \dots, p$;

(ii) - Para $j, v = 1, \dots, p$, $\frac{\partial}{\partial \theta_j} \log f_{\theta}(t)$ e $\frac{\partial^2}{\partial \theta_j \partial \theta_v} \log f_{\theta}(t)$ existem em quase toda parte e são tais que:

a - a matriz de informação de Fisher,

$$I(\theta) = E \left\{ \frac{\partial}{\partial \theta_j} \log f_{\theta}(T_1) \right\}^T \left\{ \frac{\partial}{\partial \theta_j} \log f_{\theta}(T_1) \right\};$$

onde $\frac{\partial}{\partial \theta_j} \log f_{\theta}(T_1) = \left[\frac{\partial}{\partial \theta_1} \log f_{\theta}(T_1), \dots, \frac{\partial}{\partial \theta_p} \log f_{\theta}(T_1) \right]$, é finita e positiva definida.

$$b - E(\theta) \left\{ \sup_{\{h: \|h\| \leq \delta\}} \left\| \frac{\partial^2}{\partial \theta \partial \theta^T} \log f_{(\theta+h)}(T_1) - \frac{\partial^2}{\partial \theta \partial \theta^T} \log f_{\theta}(T_1) \right\| \right\} = \psi_{\delta},$$

onde $\frac{\partial^2}{\partial \theta \partial \theta^T} \log f_{\theta}(T_1) = \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f_{\theta}(T_1) \right]_v$, converge para zero com $\delta \rightarrow 0$.

Satisfeitas as condições acima, verifica-se a distribuição assintótica de $\hat{\theta}$

$$\hat{\theta} \stackrel{\approx}{\sim} N[\theta, I^{-1}(\theta)], \tag{1.28}$$

ou ainda,

$$U(\theta) \stackrel{\approx}{\sim} N[0, I(\theta)] \text{ pois}$$

$$E[U(\theta)] \approx E(\theta - \hat{\theta}) \approx E(\theta) - E(\hat{\theta}) = \theta - \theta = 0 \text{ e}$$

$$\text{var}[U(\theta)] = \text{var}[-\theta I(\theta) + \hat{\theta} I(\theta)] = \text{var}[\hat{\theta} I(\theta)] = I^2(\theta) \text{var}(\hat{\theta}) = I^2(\theta) I^{-1}(\theta) = I(\theta).$$

1.6. Variância do Estimador da Função de Sobrevida.

Se $\hat{\theta}$ é o estimador de máxima verossimilhança de θ , pela teoria assintótica apresentada na subsecção 1.5.2, $\hat{\theta} \stackrel{\approx}{\sim} N(\theta, I^{-1}(\theta))$, onde $I(\theta) = E \left[-\frac{\partial^2}{\partial \theta} \log f_{\theta}(t) \right]$, logo $\text{var}(\hat{\theta}) = I^{-1}(\theta)$.

Seja $g(\theta)$ uma função de θ e $g(\hat{\theta})$ é o estimador de máxima verossimilhança de $g(\theta)$. Se $g(\hat{\theta})$ satisfaz as condições de regularidade descritas na subsecção 1.5.2, a distribuição assintótica para o estimador $g(\hat{\theta})$ pode então ser encontrada. Desenvolvendo $g(\hat{\theta})$ em série de Taylor em torno de θ tem-se que:

$$g(\hat{\theta}) = g(\theta) + (\hat{\theta} - \theta)g'(\theta) + \frac{(\hat{\theta} - \theta)^2}{2!} g''(\theta) + \dots$$

onde $g'(\theta) = \frac{\partial}{\partial \theta} g(\theta)$ e $g''(\theta) = \frac{\partial^2}{\partial \theta^2} g(\theta)$. Então pelo Teorema A1, apêndice A, a expressão acima torna-se:

$$g(\hat{\theta}) = g(\theta) + (\hat{\theta} - \theta)g'(\theta) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

$$\left[g(\hat{\theta}) - g(\theta) \right] = (\hat{\theta} - \theta)g'(\theta) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

pelo Teorema A2, apêndice A, $\sqrt{n} [\hat{\theta} - \theta] = o_p(1)$, ou seja $[\hat{\theta} - \theta] = o_p\left(\frac{1}{\sqrt{n}}\right)$, o que implica que

$$\sqrt{n} \left[g(\hat{\theta}) - g(\theta) \right] = \sqrt{n} (\hat{\theta} - \theta) g'(\theta) + o_p(1).$$

logo,

$$\sqrt{n} (\hat{\theta} - \theta) g'(\theta) \xrightarrow{\mathcal{D}} \mathcal{N} \left[0, [g'(\theta)]^T I^{-1}(\theta) g'(\theta) \right]$$

Tem-se que pelo teorema de Slutsky, apêndice A, teorema A3, que

$$\sqrt{n} \left[g(\hat{\theta}) - g(\theta) \right] \xrightarrow{\mathcal{D}} \mathcal{N} \left[0, [g'(\theta)]^T I^{-1}(\theta) g'(\theta) \right], \tag{1.29}$$

onde

$$\text{var}[g(\hat{\theta})] = [g'(\theta)]^T I^{-1}(\theta) g'(\theta). \tag{1.30}$$

Este método é conhecido como Método Delta.

$$\text{Se } \theta \text{ é univariado tem-se que } \sqrt{n} [g(\hat{\theta}) - g(\theta)] \xrightarrow{\mathcal{D}} \mathcal{N} \left[0, [g'(\theta)]^2 I^{-1}(\theta) \right]$$

Se $g_1(\hat{\theta}_1)$ e $g_2(\hat{\theta}_2)$ são duas funções reais de $\hat{\theta}_1$ e $\hat{\theta}_2$ respectivamente, onde $g_1(\hat{\theta}_1)$ e $g_2(\hat{\theta}_2)$ são independentes, a distribuição assintótica é dada por

$$\left[g_1(\hat{\theta}_1) - g_2(\hat{\theta}_2) \right] \stackrel{\mathcal{D}}{\rightarrow} \mathcal{N} \left[0, \text{var}[g_1(\hat{\theta}_1)] + \text{var}[g_2(\hat{\theta}_2)] \right].$$

Ainda com relação a distribuição, tem-se o intervalo de 100(1- α)% de confiança para $g(\hat{\theta})$, que é dado por

$$\text{IC: } \left[g(\hat{\theta}) \pm Z_{1-\alpha/2} \sqrt{\text{var}[g(\hat{\theta})]} \right] \tag{1.31}$$

onde $Z_{1-\alpha/2}$ é o percentil 100(1 - $\alpha/2$)% da distribuição Normal Padrão.

Exemplos:

1. $t \sim \text{exponencial}(\theta)$

Utilizando $f_{\theta}(t_p)$ e $S_{\theta}(t_i)$ dados em (1.4) e (1.5), respectivamente, a função de verossimilhança em (1.24) torna-se

$$L(\theta) = \prod_{i=1}^n \theta^{\delta_i} \exp(-\theta t_i).$$

Tem-se que a Informação de Fisher para θ é

$$I^{-1}(\theta) = E \left\{ - \frac{\partial^2}{\partial \theta^2} \log L(\theta) \right\}^{-1} = \frac{\theta^2}{nF},$$

onde $nF = \sum_{i=1}^n \delta_i$.

1a. $g(\hat{\theta}) = S_{\hat{\theta}}(t) = \exp(-\hat{\theta}t)$, onde t é fixo. logo

$$g'(\theta) = \frac{\partial}{\partial \theta} S_{\theta}(t) = -t \exp(-\theta t).$$

De (1.30) tem-se

$$\text{Var}[S_{\theta}(t)] = \left[\frac{\partial}{\partial \theta} S_{\theta}(t) \right]^2 I^{-1}(\theta) = t^2 \exp(-2\theta t) \frac{\theta^2}{nF}.$$

Então,

$$S_{\hat{\theta}}(t) \sim \mathcal{N} \left[0, t^2 \exp(-2\theta t) \frac{\theta^2}{nF} \right] \text{ e } \text{var}[S_{\hat{\theta}}(t)] = t^2 \exp(-2\hat{\theta}t) \frac{\hat{\theta}^2}{nF}.$$

1b. $g(\hat{\theta}) = -\log[S_{\hat{\theta}}(t)] = \hat{\theta}t$

$$g'(\theta) = \frac{\partial}{\partial \theta}[-\log[S_{\theta}(t)]] = t.$$

Portanto

$$\text{Var}[-\log[S_{\hat{\theta}}(t)]] = \left[\frac{\partial}{\partial \theta}[-\log[S_{\theta}(t)]] \right]^2 I^{-1}(\theta) = t^2 \frac{\theta^2}{nF}.$$

$$-\log[S_{\hat{\theta}}(t)] \sim \mathcal{N}\left[0, t^2 \frac{\theta^2}{nF}\right] \text{ e } \text{vâr}[-\log[S_{\hat{\theta}}(t)]] = t^2 \frac{\theta^2}{nF}.$$

1c. $g(\hat{\theta}) = \log[-\log[S_{\hat{\theta}}(t)]] = \log \hat{\theta} + \log t.$

$$g'(\hat{\theta}) = \frac{\partial}{\partial \theta}[\log[-\log[S_{\theta}(t)]]] = \frac{1}{\theta}.$$

Portanto

$$\text{Var}[\log[-\log[S_{\hat{\theta}}(t)]]] = \left[\frac{\partial}{\partial \theta} \log[-\log[S_{\theta}(t)]] \right]^2 I^{-1}(\theta) = \frac{1}{\theta^2} \frac{\theta^2}{nF} = \frac{1}{nF}.$$

$$\log[-\log[S_{\hat{\theta}}(t)]] \sim \mathcal{N}\left[0, \frac{1}{nF}\right] \text{ e } \text{vâr}[\log[-\log[S_{\hat{\theta}}(t)]]] = \frac{1}{nF}.$$

2. $t \sim \text{Weibull}(\alpha, \xi).$

Utilizando-se da $f_{(\alpha, \xi)}(t)$ e $S_{(\alpha, \xi)}(t)$ descritas em (1.7) e (1.8), porém para simplificar faz-se $\gamma = \xi^\alpha$. A função de verossimilhança é

$$L(\alpha, \gamma) = \prod_{i=1}^n [\alpha \gamma t_i^{\alpha-1}]^{\delta_i} \exp(-\gamma t_i^\alpha).$$

A matriz $\Gamma^{-1}(\gamma, \alpha)$ é composta por:

$$A = \frac{\partial^2}{\partial \gamma^2} \log L(\alpha, \gamma) = -\frac{nF}{\gamma^2}$$

$$B = \frac{\partial^2}{\partial \gamma \partial \alpha} \log L(\alpha, \gamma) = - \sum_{i=1}^n t_i^\alpha \log t_i$$

$$C = \frac{\partial}{\partial \alpha} \log L(\alpha, \gamma) = - \frac{nF}{\alpha^2} - \gamma \sum_{i=1}^n t_i^\alpha [\log t_i]^2.$$

Portanto $I^{-1}(\theta) = \begin{bmatrix} \text{var}(\hat{\gamma}) & \text{cov}(\hat{\gamma}, \hat{\alpha}) \\ \text{cov}(\hat{\gamma}, \hat{\alpha}) & \text{var}(\hat{\alpha}) \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix}^{-1}.$

2a. $g(\hat{\theta}) = S_{\hat{\theta}}(t) = \exp(-\hat{\gamma}t^{\hat{\alpha}})$, t é fixo.

$$g'(\theta) = \begin{bmatrix} \frac{\partial}{\partial \gamma} S_{\theta}(t) \\ \frac{\partial}{\partial \alpha} S_{\theta}(t) \end{bmatrix} = \begin{bmatrix} -t^\alpha \exp(-\gamma t^\alpha) \\ -\exp(-\gamma t^\alpha) \gamma t^\alpha \log t \end{bmatrix}$$

logo

$$\text{var}[S_{\hat{\theta}}(t)] = [g'(\theta)]^T I^{-1}(\theta) g'(\theta).$$

2b. $g(\hat{\theta}) = -\log[S_{\hat{\theta}}(t)] = \hat{\gamma}t^{\hat{\alpha}}$

$$g'(\theta) = \begin{bmatrix} \frac{\partial}{\partial \gamma} [-\log S_{\theta}(t)] \\ \frac{\partial}{\partial \alpha} [-\log S_{\theta}(t)] \end{bmatrix} = \begin{bmatrix} t^\alpha \\ \gamma t^\alpha \log t \end{bmatrix}$$

2c. $g(\hat{\theta}) = \log[-\log[S_{\hat{\theta}}(t)]] = \log \hat{\gamma} + \hat{\alpha} \log t$

$$g'(\theta) = \begin{bmatrix} \frac{\partial}{\partial \gamma} \log[-\log S_{\theta}(t)] \\ \frac{\partial}{\partial \alpha} \log[-\log S_{\theta}(t)] \end{bmatrix} = \begin{bmatrix} \frac{1}{\gamma} \\ \log t \end{bmatrix}$$

1.7. Testes de Hipóteses.

Uma vez que se conhece o modelo "adequado" para explicar os dados de sobrevivência a próxima etapa é construir teste de ajuste do modelo e teste sobre os seus parâmetros.

Testes estatísticos para verificar a significância dos parâmetros no modelo. Seja $\theta = (\theta_q^T, \theta_{p-q}^T)^T$, ($q \leq p$) uma partição do vetor de parâmetros, e considere a hipótese nula definida por: $H_0: \theta_q = \theta_q^0$. Um caso particular desta hipótese é $\theta_q^0 = 0$. Três estatísticas de testes são usualmente empregadas para testar H_0 .

a. - Estatística de Wald

$$Q_w = (\hat{\theta}_q - \hat{\theta}_q^0)^T [I_{qq}(\theta)]^{-1} (\hat{\theta}_q - \hat{\theta}_q^0), \tag{1.32}$$

onde $I_{qq}(\theta)$ é a sub-matriz de $I(\theta)$ associado ao sub-vetor $\hat{\theta}_q$ e $\hat{\theta}$. Porém, quando há interesse em testar a nulidade de somente um parâmetro do modelo, i.e., $H_0: \theta_j = 0, j = 1, \dots, p$ a estatística acima toma a forma

$$Q_w = \frac{\hat{\theta}_j^2}{I^{jj}(\hat{\theta})}, \tag{1.33}$$

onde $I^{jj}(\hat{\theta})$ são os elementos da diagonal principal de $[I(\hat{\theta})]^{-1}$.

b. - Estatística do Escore de Rao

Por definição de $\hat{\theta}$, $U(\hat{\theta}) = 0$, sob a validade de H_0 , o estimador restrito $\hat{\theta}^0 = (\hat{\theta}_q^{0T}, \hat{\theta}_{p-q}^{0T})^T$. Uma forma de medir a distância entre $U(\hat{\theta}^0)$ e o vetor nulo é através da forma quadrática.

$$S = U(\hat{\theta}^0)^T [I(\hat{\theta}^0)]^{-1} U(\hat{\theta}^0). \tag{1.34}$$

Note que esta estatística não exige cálculo do estimador de máximo verossimilhança de θ .

c. - Estatística da Razão de Verossimilhança (Neyman - Pearson/Wilks)

$$R(\theta_0) = -2 [L(\hat{\theta}^0) - L(\hat{\theta})], \tag{1.35}$$

onde $L(\cdot)$ é o logaritmo da função de verossimilhança.

Sob a hipótese H_0 as três estatísticas acima tem distribuição assintótica Qui-Quadrado com q graus de liberdade.

1.8. Modelo de Riscos Proporcionais.

Os modelos de riscos proporcionais assumem que variáveis concomitantes tem um efeito multiplicativo na função risco. Esta suposição parece ser razoável em muitas situações, por isso tem sido de grande importância na análise de dados de sobrevivência. Este modelo está definido em (1.23), na secção 1.4.

Há duas situações a ser consideradas:

1. $\lambda_0(t)$ é uma função específica, i.e., envolve parâmetros desconhecidos, como um exemplo, tem-se o modelo paramétrico estudado por Glasser [1967];
2. $\lambda_0(t)$ é uma função desconhecida e não especificada, i.e., uma função de risco arbitrária, Cox [1972].

Esta secção tem como objetivo apresentar resultados específicos relativos ao modelo de riscos proporcionais, nas situações acima.

1.8.1. Modelo Exponencial (Glasser).

O Modelo de riscos proporcionais é dito ser exponencial quando a distribuição associada aos tempos de falha é exponencial, i. e. a função de risco básica é constante. Este modelo deve ser considerado quando as diferenças em θ_i estão relacionadas com outras variáveis concomitantes z , i.e

$$\theta_i = g(z_i; \beta). \quad (1.36)$$

Glasser [1967] propôs um modelo log-linear onde $g(z_i; \beta)$ em (1.35) é dado por

$$g(z_i; \beta) = \exp[\beta^T z_i], \quad (1.37)$$

e o valor esperado do tempo de sobrevivência é

$$E(t_i) = \frac{1}{\theta_i} = \exp[-\beta^T z_i],$$

com $z_{i0} \equiv 1$, $i = 1, 2, \dots, n$ onde $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$ e $z_i = (z_{i0}, z_{i1}, \dots, z_{ip})$.

A função de verossimilhança dos n tempos de sobrevivência pode ser escrito como

$$L(\underline{\beta}) = \prod_{i=1}^n \left[\exp(\underline{\beta}^T \mathbf{z}_i) \right]^{\delta_i} \exp \left[- t_i \exp(\underline{\beta}^T \mathbf{z}_i) \right]. \tag{1.38}$$

O logaritmo da função de verossimilhança é

$$\log L(\underline{\beta}) = \sum_{i=1}^n \delta_i (\underline{\beta}^T \mathbf{z}_i) - \sum_{i=1}^n t_i \exp(\underline{\beta}^T \mathbf{z}_i). \tag{1.39}$$

Os estimadores de máxima verossimilhança de $\underline{\beta}^T = (\beta_0, \beta_1, \dots, \beta_p)$ podem ser encontrados resolvendo as (p+1) equações:

$$\begin{aligned} U_j(\underline{\beta}) &= \frac{\partial}{\partial \beta_j} \log L(\underline{\beta}) = \sum_{i=1}^n \delta_i z_{ij} - \sum_{i=1}^n t_i z_{ij} \exp(\underline{\beta}^T \mathbf{z}_i) \\ &= \sum_{i=1}^n z_{ij} \left[\delta_i - t_i \exp(\underline{\beta}^T \mathbf{z}_i) \right], \quad j = 0, 1, \dots, p. \end{aligned} \tag{1.40}$$

Estas equações não tem solução explícita, logo necessita-se de métodos iterativos para resolvê-las. O procedimento de Newton-Raphson descrito na secção (1.4) é usado.

Os elementos da matriz das derivadas segundas é

$$H_{jv}(\underline{\beta}) = \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\underline{\beta}) = - \sum_{i=1}^n z_{ij} t_i \exp(\underline{\beta}^T \mathbf{z}_i) z_{iv}. \tag{1.41}$$

A matriz de covariância assintótica para os estimadores de máxima verossimilhança podem ser estimados pela inversa da Informação de Fisher.

$$V(\hat{\underline{\beta}}) = [I(\hat{\underline{\beta}})]^{-1}$$

cujos elementos de $I(\hat{\underline{\beta}}) = E \left\{ - \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\underline{\beta}) \right\}.$

Muitas vezes, as esperanças solicitadas acima são incalculáveis. Em tais casos, utiliza-se o inverso da matriz de Fisher observada, i.e., a matriz $I(\hat{\underline{\beta}})$ tem o elemento (j,v) da forma

$$- \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\underline{\beta}).$$

A distribuição de sobrevivência estimada é

$$\hat{S}(t_i) = \exp \left[- t_i \exp(\hat{\beta}^T z_i) \right]. \tag{1.42}$$

Da secção 1.6. tem-se que

$$\text{Var}(S_{\hat{\theta}}(t_i)) = [S'_{\hat{\theta}}(t_i)]^T I^{-1}(\hat{\theta}) S'_{\hat{\theta}}(t_i), \tag{1.43}$$

onde

$$S'_{\hat{\theta}}(t_i) = \frac{\partial}{\partial \beta_j} S_{\hat{\theta}}(t_i) = \exp \left[- t_i \exp(\hat{\beta}^T z_i) \right] \left[- t_i z_{ij} \exp(\hat{\beta}^T z_i) \right], \tag{1.44}$$

1.8.2. Modelo de Riscos Proporcionais de Cox.

O modelo de riscos proporcionais de Cox deve ser considerado, quando não se conhece a distribuição associada aos tempos. A teoria desenvolvida por Cox [1972] assume para os dados um modelo de riscos proporcionais dado em (1.23), onde $\lambda_0(t)$ é uma função de risco arbitrária que corresponde à função risco de um indivíduo quando $z = \theta$.

A função densidade de probabilidade de T dado z correspondente a (1.23) é dada por

$$f(t | z) = \lambda(t, z) \exp(\beta^{*T} z^*) \exp \left[- \exp(\beta^{*T} z^*) \int_0^t \lambda_0(u) du \right]. \tag{1.45}$$

Cox [1972] sugeriu a função de verossimilhança parcial para estimar os parâmetros β na equação acima. Para simplificar a notação, assume-se que as observações são ordenadas, i.e, seja $t_1 < \dots < t_k$ denota os k distintos tempos de falha ordenados e $\mathfrak{R}(i)$ o conjunto de indivíduos em risco em $t(i) - \theta$ (i.e., vivos e não censurados), representado por $\mathfrak{R} = \{ i: t_i \geq t \}$. A probabilidade condicional que um indivíduo (i) em estudo falhe até $t(i)$, dado que ocorreu exatamente uma falha em $t(i)$ e $\mathfrak{R}(i)$ é:

$$\frac{\lambda(t(i), z_i^*)}{\sum_{l \in \mathfrak{R}(i)} \lambda(t(i), z_l^*)} = \frac{\exp(\beta^{*T} z_{(i)}^*)}{\sum_{l \in \mathfrak{R}(i)} \exp(\beta^{*T} z_{(l)}^*)} \tag{1.46}$$

para $i = 1, 2, \dots, k$.

Cada instante de falha contribui com um fator para a verossimilhança, logo

$$L(\underline{\beta}^*) = \prod_{i=1}^k \frac{\exp(\underline{\beta}^{*\text{T}} z_{(i)}^*)}{\sum_{l \in \mathfrak{R}(i)} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*)}$$

O estimador de $\underline{\beta}$ é encontrado maximizando o logaritmo da função de verossimilhança parcial

$$\log L(\underline{\beta}^*) = \sum_{i=1}^k \left[\underline{\beta}^{*\text{T}} z_{(i)}^* - \log \left\{ \sum_{l \in \mathfrak{R}(i)} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*) \right\} \right] \quad (1.47)$$

A maximização é realizada igualando a zero a derivada parcial de primeira ordem da equação acima, i. e.

$$\frac{\partial}{\partial \beta_j} \log L(\underline{\beta}^*) = \sum_{i=1}^k \left[z_{ij} - \frac{\sum_{l \in \mathfrak{R}(i)} z_{lj} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*)}{\sum_{l \in \mathfrak{R}(i)} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*)} \right] \quad (1.48)$$

tem que ser resolvida iterativamente, empregando o método de Newton-Raphson (secção 1.5.1); e o elemento (jv) da matriz das segundas derivadas é da forma

$$\begin{aligned} \frac{\partial}{\partial \beta_j \partial \beta_v} \log L(\underline{\beta}^*) = & - \sum_{i=1}^k \left\{ \frac{\left\{ \sum_{l \in \mathfrak{R}(i)} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*) \right\} \left\{ \sum_{l \in \mathfrak{R}(i)} z_{lj} z_{lv} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*) \right\}}{\left\{ \sum_{l \in \mathfrak{R}(i)} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*) \right\}^2} - \right. \\ & \left. \frac{\left\{ \sum_{l \in \mathfrak{R}(i)} z_{lj} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*) \right\} \left\{ \sum_{l \in \mathfrak{R}(i)} z_{lv} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*) \right\}}{\left\{ \sum_{l \in \mathfrak{R}(i)} \exp(\underline{\beta}^{*\text{T}} z_{(l)}^*) \right\}^2} \right. \end{aligned} \quad (1.49)$$

Se $\hat{\underline{\beta}}$ representa a estimativa de $\underline{\beta}$, a estimativa da função de sobrevivência $S(t, \mathbf{z})$ é dada por

$$\hat{S}(t, \mathbf{z}) = \exp \left[- \int_0^t \lambda(u, \mathbf{z}) du \right] = \exp \left[- \int_0^t \lambda_0(u) \exp(\hat{\underline{\beta}}^{*\text{T}} \mathbf{z}^*) du \right]$$

$$\begin{aligned}
 &= \exp \left[\left\{ - \int_0^t \lambda_0(u) \, du \right\} \exp(\hat{\beta}^{*T} z^*) \right] = \left[\exp \left\{ - \int_0^t \lambda_0(u) \, du \right\} \right] \exp(\hat{\beta}^{*T} z^*) \\
 &= [\hat{S}_0(t)] \exp(\hat{\beta}^{*T} z^*). \tag{1.50}
 \end{aligned}$$

Link [1984] propôs que a variância da função de sobrevivência fosse estimada usando a fórmula da variância derivada por Tsiatis [1974], onde este usou de argumentos heurísticos para encontrar uma estimativa da variância assintótica da função de Sobrevivência $\hat{S}(t, z)$. Sua estimativa da variância assintótica para os tempos de morte observados $t = t_1, \dots, t_k$ é dada por

$$\hat{V}_T[\hat{S}(t, z)] \approx \hat{S}^2(t, z) \left\{ \exp(2\hat{\beta}^{*T} z^*) \left[\sum_{i=1}^h \frac{1}{\left(\sum_{l \in \mathfrak{R}(i)} \exp(\hat{\beta}^{*T} z_{(l)}^*) \right)^2} \right] + \exp(2\hat{\beta}^{*T} z^*) W^T \hat{V}ar(\hat{\beta}^*) W \right\} \tag{1.51}$$

onde h é tal que $t_h < t \leq t_{h+1}$, e

$$W_j = \sum_{i=1}^h \left\{ \frac{\sum_{l \in \mathfrak{R}(i)} z_{lj} \exp(\hat{\beta}^{*T} z_{(l)}^*)}{\left(\sum_{l \in \mathfrak{R}(i)} \exp(\hat{\beta}^{*T} z_{(l)}^*) \right)^2} - z_j \frac{1}{\sum_{l \in \mathfrak{R}(i)} \exp(\hat{\beta}^{*T} z_{(l)}^*)} \right\}.$$

Tsiatis [1981], usando um processo Gaussiano, mostrou que $\hat{S}(t, z)$ é assintoticamente normal.

1.9. Exemplo.

Nesta secção será considerado um exemplo com o intuito de ilustrar os métodos descritos acima para análise de modelos de regressão com dados de sobrevivência.

Exemplo 1: Krall, Uthoff e Harley [1975] apresentaram um estudo sobre mieloma múltiplo conduzido por pesquisadores da West Virginia University Medical Center. Neste estudo 65 pacientes com "alkylating agents" são tratados ao tempo de diagnóstico. Destes pacientes 48 morreram durante o estudo e 17 estavam vivos no término do mesmo. Assim tem-se disponível os tempos de sobrevivência em meses desde o diagnóstico e a variável STATUS que assume dois valores, 0 ou 1, indicando se o paciente estava vivo ou morto, respectivamente, até o término do estudo. Assim se o valor do STATUS é 0, o correspondente valor do tempo é censurado. Além

disso, foi registrado para cada paciente os valores das seguintes variáveis concomitantes.

Tabela 1.1: Sumário das variáveis aferidas no estudo.

VARIÁVEIS	Descrição	Tipo	Notação do Apêndice B1
LOGBUN	Logaritmo da variável BUN	Contínua	Z1
HGB	Hemoglobina	Contínua	Z2
PLATELET	Plateletes	Discreta 0 = anormal 1 = normal	Z3
INFEC	Infecção	Discreta 0 = ausente 1 = presente	Z4
IDADE	Idade em anos (completos)	Contínua	Z5
SEXO	Sexo	Discreta 1 = masculino 2 = feminino	Z6
LOGWBC	Logaritmo da variável WBC	Contínua	Z7
FRAT	Fraturas	Discreta 0 = ausente 1 = presente	Z8
LOGPBM	Logaritmo da % de Células Plasmática na medula óssea	Contínua	Z9
LINF	% linfócitos no sangue periférico	Contínua	Z10
CELULAS	% células myeloid no sangue perif.	Contínua	Z11
PROTEIN	Proteínas	Contínua	Z12
PBJU	Proteína Bence Jone na urina	Discreta 1 = presente 2 = ausente	Z13
TOTSORO	Total de soro proteico	Contínua	Z14
SOROGL	Soro Globin (gm %)	Contínua	Z15
SOROCAL	Soro Cálcio (mgm %)	Contínua	Z16

Obs: Os dados estão listados na Tabela B1 (Apêndice B), página 120.

O objetivo principal desse estudo foi identificar fatores prognósticos importantes na sobrevivência de pacientes com mieloma múltiplo.

1.9.1. Resultados.

A. Descritivo.

A1. Análise descritiva da variável tempo de sobrevivência.

A análise exploratória das variáveis foi feita utilizando o procedimento UNIVARIATE do *software* estatístico SAS. Em A1 apresenta os resultados obtidos na investigação da variável resposta TEMPO e em A2 a análise descritiva das variáveis concomitantes aferidas.

0	44778	0	1122235566667779
1	1223699	1	111113456677899
2	8	2	456
3		3	257
4	1	4	11
5	37	5	1248
6		6	67
7	7	7	
8		8	89
9		9	2
	Censura		Falha

Quadro 1.1 - Ramo e Folhas dos tempos de censura e falha dos pacientes (em meses) com mieloma múltiplo.

Observando o aspecto visual da distribuição dos tempos de censura e falha no quadro acima, parece razoável supor que a distribuição associada ao tempo de sobrevivência dos pacientes com mieloma múltiplo seja exponencial.

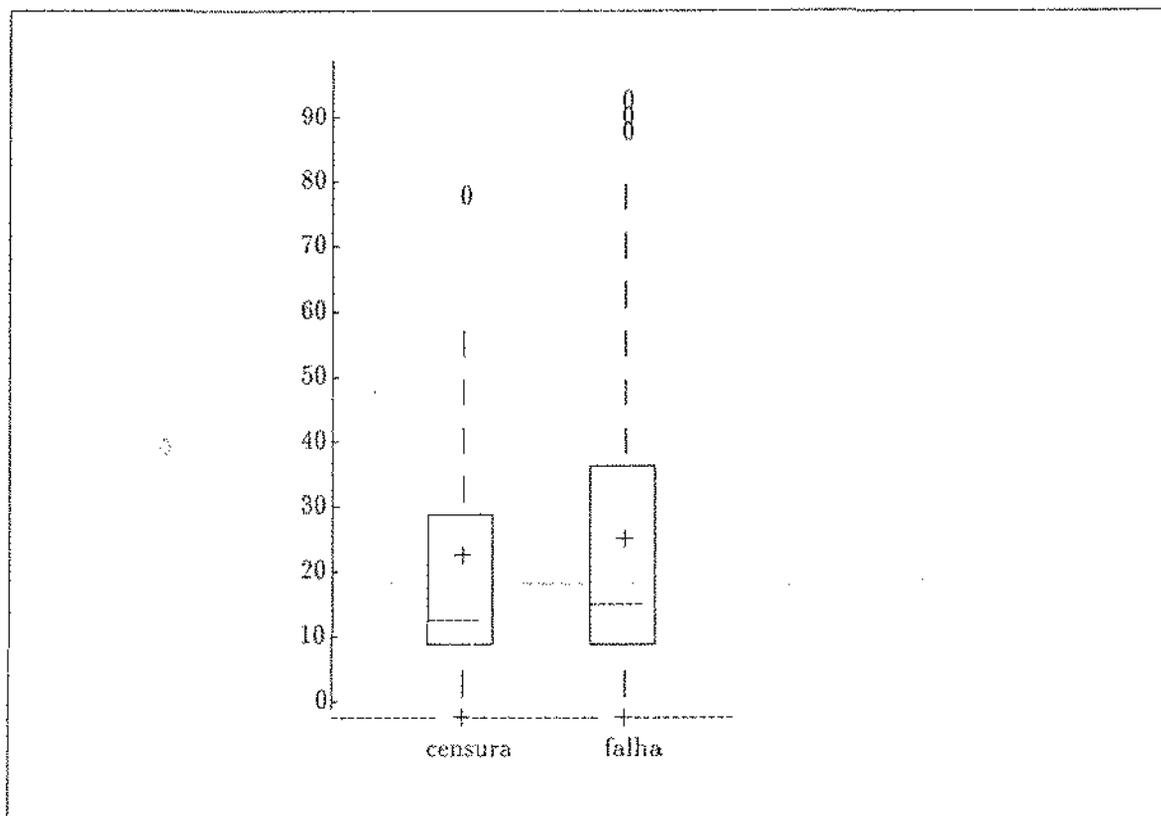


Figura 1.1 - Desenho esquemático correspondentes aos tempos de censura e falha dos pacientes com mieloma múltiplo.

A2: Análise descritiva das variáveis independentes.

Tabela 1.2 - Sumário das estatísticas para as variáveis contínuas.

VARIÁVEL	N	Média	Mediana	d.p	Máximo	Mínimo
LOGBUN	65	1.393	1.322	0.313	2.236	0.778
HGB	65	10.202	10.200	2.557	14.600	4.900
IDADE	65	60.154	60.000	10.334	82.000	38.000
LOGWBC	65	3.769	3.732	0.238	4.954	3.362
LOGPBM	65	1.549	1.623	0.364	2.000	0.477
LINF	65	6.738	6.000	6.300	24.000	0.000
CELULAS	65	30.354	35.000	20.031	68.000	0.000
PROTEIN	65	3.615	1.000	6.012	27.000	0.000
TOTSORO	65	8.600	9.000	2.249	17.000	4.000
SOROGL	65	5.215	5.000	2.190	12.000	2.000
SOROCAL	65	10.123	10.000	1.816	18.000	7.000

A2. Estimação não paramétrica da função de sobrevivência.

Estimou-se a função de sobrevivência através do método não paramétrico de Kaplan-Meier, utilizou-se do procedimento LIFETEST do SAS. Os resultados estão sumarizados na Tabela 1.3 e na Figura 1.2.

Tabela 1.3 - Função de sobrevivência estimada pelo método de Kaplan-Meier para pacientes com mieloma múltiplo.

TEMPO (t_1)	d_1	n_1	$\tilde{S}(t_1)$	d.p.[$\tilde{S}(t_1)$]
1.25	2	65	0.9692	0.0214
2.00	3	63	0.9231	0.0331
3.00	1	60	0.9077	0.0359
5.00	2	57	0.8758	0.0411
6.00	4	55	0.8121	0.0489
7.00	3	51	0.7644	0.0533
9.00	1	45	0.7474	0.0547
11.00	5	44	0.6625	0.0603
13.00	1	36	0.6441	0.0613
14.00	1	34	0.6251	0.0624
15.00	1	33	0.6062	0.0633
16.00	2	32	0.5683	0.0648
17.00	2	29	0.5291	0.0660
18.00	1	27	0.5095	0.0664
19.00	2	26	0.4703	0.0668
24.00	1	22	0.4489	0.0671
25.00	1	21	0.4275	0.0672
26.00	1	20	0.4062	0.0672

Tabela 1.3 - (Continuação).

TEMPO (t_i)	d_i	n_i	$\hat{S}(t_i)$	d.p.[$\hat{S}(t_i)$]
32.00	1	18	0.3836	0.0671
35.00	1	17	0.3610	0.0669
37.00	1	16	0.3385	0.0664
41.00	2	15	0.2933	0.0647
51.00	1	12	0.2689	0.0638
52.00	1	11	0.2445	0.0625
54.00	1	9	0.2173	0.0612
58.00	1	7	0.1863	0.0598
66.00	1	6	0.1552	0.0573
67.00	1	5	0.1242	0.0536
88.00	1	3	0.0828	0.0492
89.00	1	2	0.0414	0.0382
92.00	1	1	-	-

nota- n_i : número de pacientes em t_i
 d_i : número de pacientes mortos em t_i

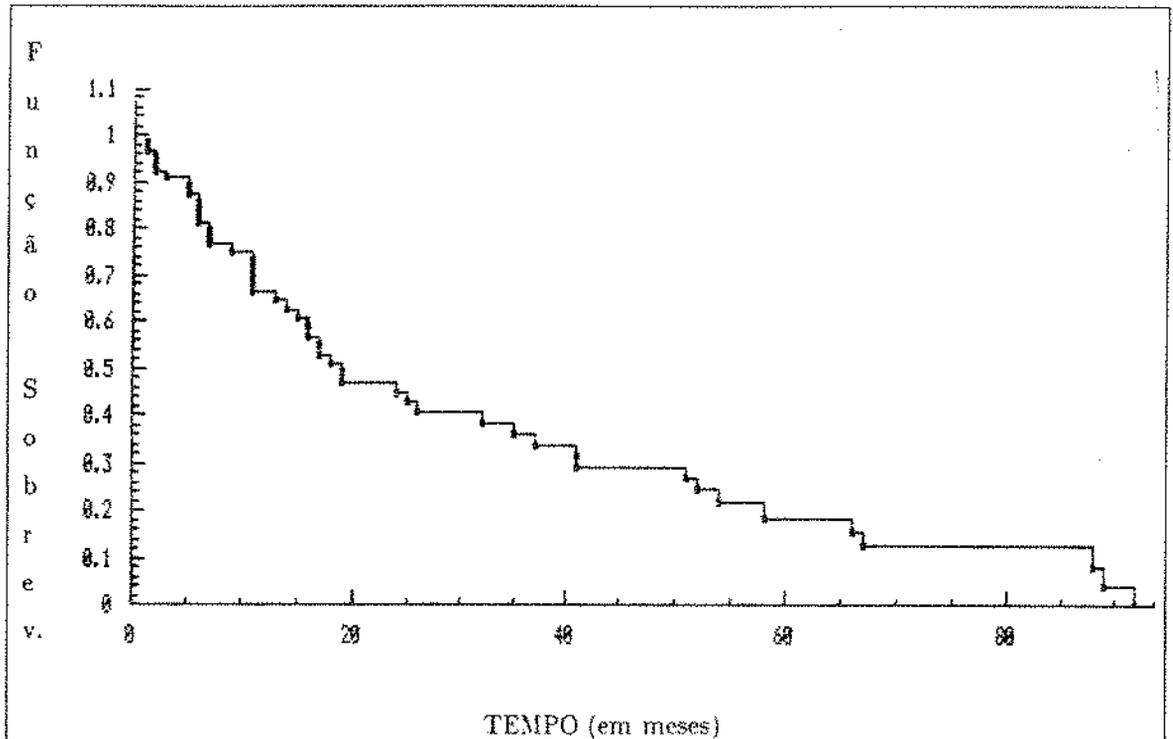


Figura 1.2 - Curva de sobrevivência estimada através de Kaplan-Meier para pacientes com mieloma múltiplo.

A3. Investigação da distribuição dos tempos através de gráficos.

Uma investigação mais detalhada dos tempos de falha dos pacientes com mieloma múltiplo foi feita através do gráfico menos logaritmo da função de sobrevivência ($-\log \hat{S}(t)$) contra tempo (ver Figura 1.3) para verificar a distribuição adequada. Considerando-se somente a variável resposta tempo e o indicador de censura, ajustou-se o modelo paramétrico exponencial através do procedimento LIFEREG do SAS. Os resultados estão mostrados na Tabela 1.4.

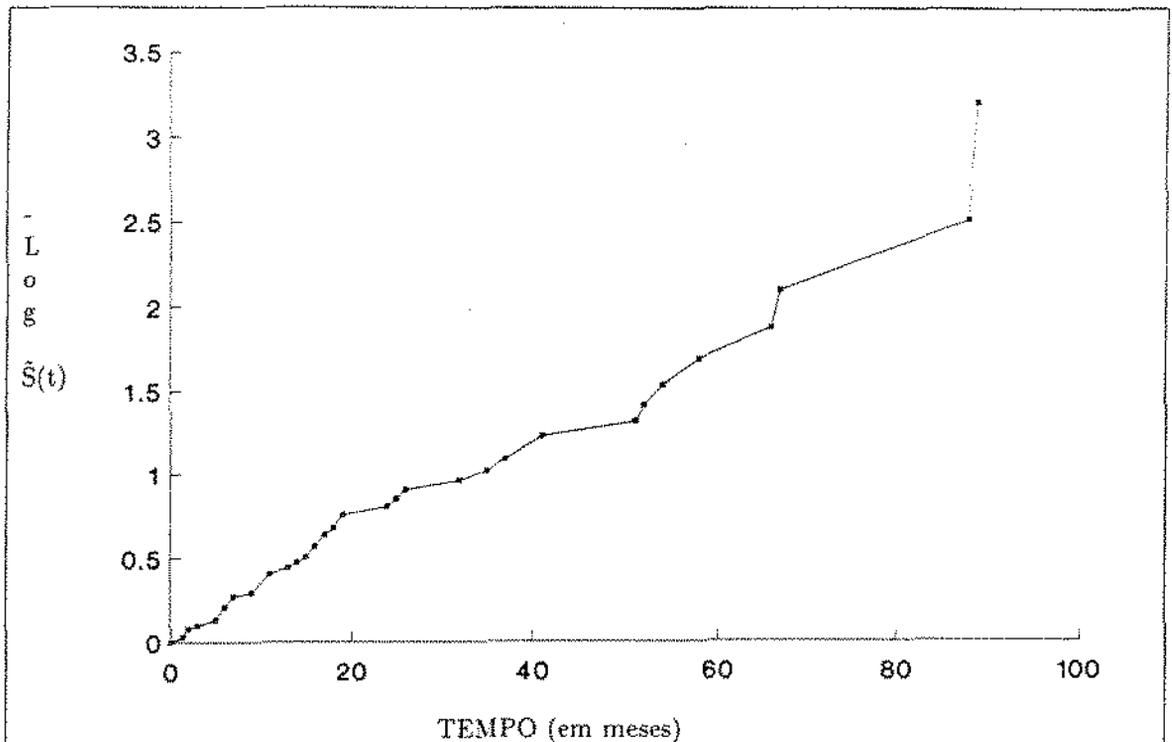


Figura 1.3 - Distribuição menos logaritmo da função de sobrevivência ao longo do tempo.

A figura acima mostra uma evidência de que a relação entre menos logaritmo da função de sobrevivência e tempo possa se considerar linear. Portanto, a distribuição exponencial dos tempos parecem se ajustar aos dados.

Tabela 1.4 - Ajuste do Modelo Exponencial.

VARIÁVEIS	$\hat{\beta}$	Erro-Padrão	p-value
INTERCEPTO	3.4816	0.1443	0.0001
ESCALAR	1.0000	0.0000	

O valor do teste multiplicador de Lagrange sob a hipótese H_0 : escalar = 1 é 0.1129 com p-value = 0.7369. Portanto, aceita-se a hipótese H_0 , i.e., os tempos possuem distribuição exponencial.

B. Ajuste do Modelo.

B1. Modelo de Cox

O modelo de Cox foi ajustado através do procedimento PHREG do *software* estatístico SAS, com o propósito de identificar fatores prognósticos importantes na sobrevivência de pacientes com mieloma múltiplo. Para selecionar as covariáveis que formarão o "melhor" modelo empregou-se o método de seleção STEPWISE. Este e outros métodos de seleção podem ser encontrados em Hosmer and Lemeshow [1984 - Cap. 3]. Duas variáveis, LOGBUN e HGB estavam relacionadas com o tempo de sobrevivência destes pacientes. Portanto, a função risco para cada indivíduo é dada por:

$$\lambda_i(t, z_i) = \lambda_0(t) \exp(\beta_1 \text{LOGBUN}_i + \beta_2 \text{HGB}_i), \quad i = 1, 2, \dots, n.$$

A tabela abaixo mostra o sumário do modelo ajustado.

Tabela 1.5 - Estimação dos coeficientes de regressão para o modelo de Cox usando o método de seleção Stepwise.

VARIÁVEIS	$\hat{\beta}$	Erro-Padrão	p-value
LOGBUN	1.6744	0.6121	0.0062
HGB	- 0.1189	0.0575	0.0385

Para testar o ajuste do modelo, i.e., $H_0: \beta_1 = \beta_2 = 0$, usou-se a estatística da razão de verossimilhança o valor encontrado foi 11.949 que é uma χ^2 com 2 graus de liberdade

(p-value = 0.0025). Há evidências de que o teste é significativo, i.e., os fatores LOGBUN e HGB são importantes na sobrevivência dos pacientes com mieloma múltiplo.

B1. Modelo Log-linear

Uma vez que os tempos de sobrevivência de pacientes com mieloma múltiplo estão distribuídos exponencialmente, ajustou-se também o modelo de regressão log-linear exponencial utilizando o procedimento LIFEREG do *software* estatístico SAS. Deste modelo também concluiu-se que somente as variáveis LOGBUN e HGB são significantes. Portanto, ajustou-se um segundo modelo onde a função risco para cada indivíduo é dada por:

$$\lambda_i(t, z_i) = \exp(\beta_0) \exp(\beta_1 \text{LOGBUN}_i + \beta_2 \text{HGB}_i), \quad i = 1, 2, \dots, n.$$

Os resultados estão sumarizados na tabela abaixo.

Tabela 1.6 - Estimação dos coeficientes de regressão para o modelo log-linear exponencial.

VARIÁVEIS	$\hat{\beta}$	Erro-Padrão	p-value
INTERCEPTO	4.4806	0.9949	0.0001
LOGBUN	- 1.5414	0.5549	0.0055
HGB	0.1041	0.0558	0.0623

Os resultados obtidos neste modelo são muito próximos aos do modelo de Cox. O critério usado para testar o ajuste do modelo, $H_0: \beta_1 = \beta_2 = 0$, foi o teste da razão de verossimilhança

$$R(\hat{\beta}) = -2 \left\{ \log L(\hat{\beta}^0) - \log L(\hat{\beta}) \right\},$$

onde $\log L(\hat{\beta}^0)$ é o logaritmo da função de verossimilhança do modelo onde $\beta_1 = \beta_2 = 0$ e $\log L(\hat{\beta})$ é o logaritmo da função de verossimilhança do modelo apresentado na Tabela 1.6. Portanto, o valor do teste encontrado é 11.2554 que é uma χ^2 com 2 gl (p-value = 0.0036). Os fatores LOGBUN e HGB mostraram-se importantes na sobrevivência dos pacientes com mieloma múltiplo. O coeficiente negativo indica que LOGBUN está relacionado negativamente com o $\log[\lambda_i(t, z_i)]$.

C. Intervalos de Confiança para a Função Sobrevida nos modelos B1 e B2.

Com o intuito de comparar as taxas de sobrevivência segundo os modelos acima, estimou-se a função de sobrevivência $\hat{S}(t; z)$ para cada paciente e construiu-se intervalos de confiança de 95% de confiança. Os resultados encontram-se na Tabela C1 (apêndice C). O procedimento PHREG do SAS, fornece estimativa da função sobrevivência $S(t; z)$, mas não oferece uma forma de estimação da variância da função de sobrevivência. Para obter tal estimativa foi desenvolvido o programa PEVSTMC.PRG utilizando o procedimento IML do SAS, (Apêndice D3.5). Por outro lado, o procedimento LIFEREG não estima a função de Sobrevida $S(t; z)$, neste caso foi desenvolvido o programa PESVSMG.PRG (apêndice D3.3) para o cálculo da estimativa de $S(t; z)$ e $\text{var}(S(t; z))$ dado em (1.41) e (1.42) respectivamente.

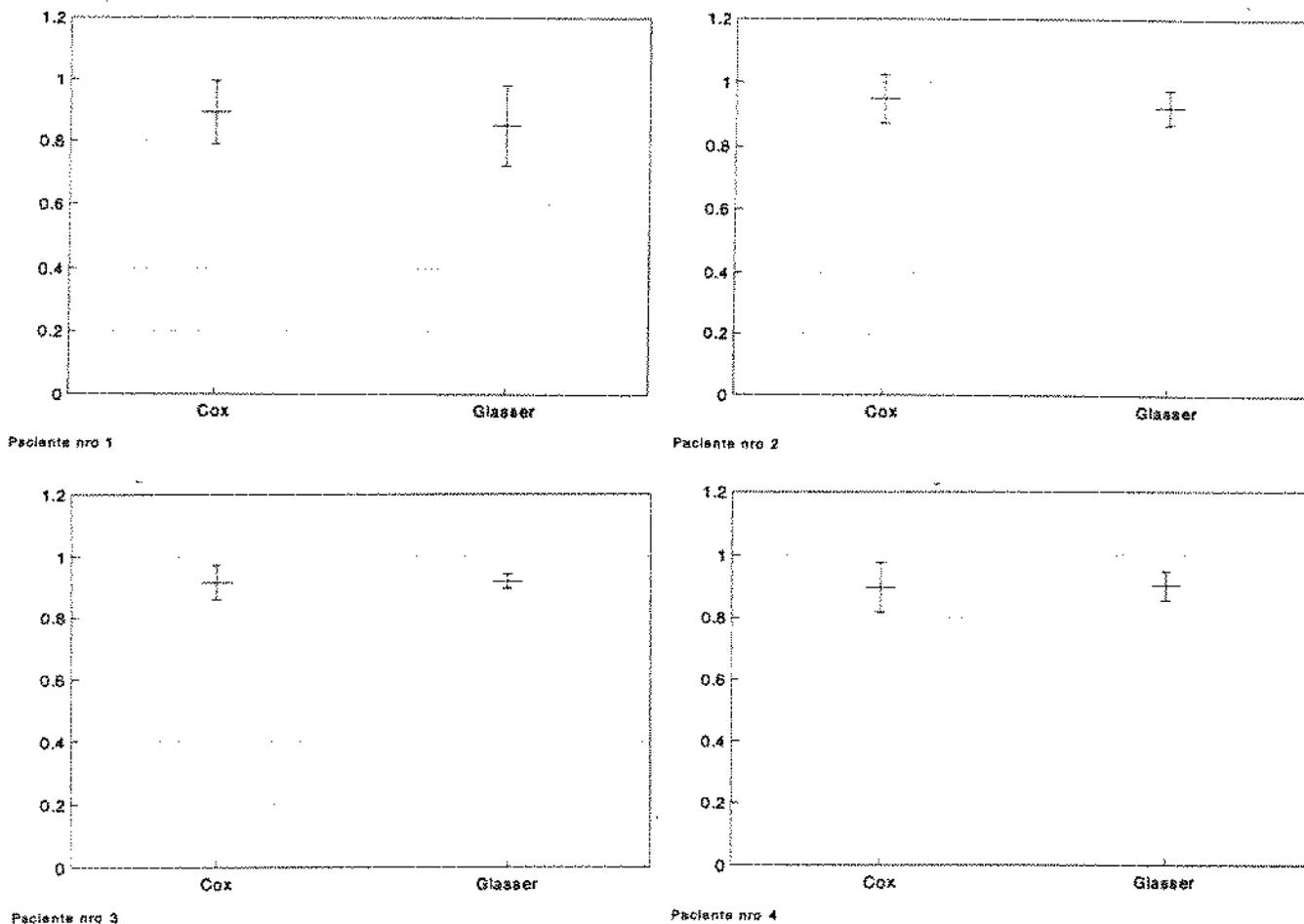


Figura 1.4 - Gráficos dos intervalos de confiança para estimativa de $\hat{S}(t; z)$ nos modelos de riscos proporcionais.

D. Conclusão.

Pelo aspecto visual dos gráficos anteriores, nota-se que em média a função de sobrevivência estimada através do modelo de Cox são inferiores à respectiva função sobrevivência estimada pelo modelo de Glasser. Entretanto, a estimativa de seu respectivo erro padrão, é inferior no modelo de Glasser que no modelo de Cox.

Ambos os modelos parecem ser adequados para a análise dos dados da Tabela B1.

CAPÍTULO 2

MODELO LOGÍSTICO EM ANÁLISE DE SOBREVIVÊNCIA

2.1. Introdução

Suponha que n indivíduos são estudados, e para o i -ésimo indivíduo observa-se a variável aleatória X_i , assumindo os valores 0 ou 1, que podem ser chamados de falha ou sucesso respectivamente, ou seja $x_i = 1$ se o i -ésimo indivíduo é um sucesso e $x_i = 0$, caso contrário. Suponha ainda que p variáveis independentes $Z_{i1}, Z_{i2}, \dots, Z_{ip}$ (características do paciente ou possíveis fatores de risco) são medidas para cada indivíduo.

Problemas desta ordem ocorrem especialmente na área médica, onde a resposta binária pode representar sucesso ou falha de um tratamento específico, morte ou sobrevivência após seguimento do tratamento durante um período de tempo, ou ainda morte por uma determinada causa em contraste com morte por alguma causa distinta. As variáveis $Z_{i1}, Z_{i2}, \dots, Z_{ip}$ podem representar quantidades ou qualidades próprias de cada indivíduo. Possíveis variáveis Z são: idade, tempo desde diagnóstico, severidade inicial dos sintomas, sexo, aspectos da histologia médica do indivíduo, entre outras.

Com tais informações, em geral, deseja-se um método que permita avaliar a dependência da probabilidade de sucesso sob valores das variáveis independentes Z , i. e, deseja-se utilizar as

covariáveis a fim de explicar ou prever a influência destas na variável resposta. A função logística é então usada por estatísticos e epidemiologistas para modelar as probabilidades de sucesso ou falha, durante um período de tempo específico, como função de várias variáveis independentes que podem influenciar a variável dicotômica tomada como resposta.

O modelo logístico também pode ser utilizado para análise de dados de sobrevivência. Ele é uma alternativa do modelo de regressão de Cox e de outros modelos de regressão paramétricos que utilizam os tempos de sobrevivência como variável dependente e incorporam observações censuradas (perdidos da observação). No modelo logístico, a variável dependente é a variável dicotômica, observada para cada indivíduo em estudo, após um período de tempo τ pré fixado. Neste modelo, os casos perdidos da observação não são incluídos na análise.

O uso do modelo logístico como modelo para a análise de sobrevivência foi questionado em diversos aspectos, em muitos estudos. Elandt-Johnson [1980] e Green, M.S e Symons, M.J [1983] estudaram as condições sob as quais o modelo logístico e de Cox se aproximam. Eles mostraram que quando o período de seguimento é curto e a doença é rara, os coeficientes do modelo de regressão logístico se aproximam daqueles do modelo de riscos proporcionais de Cox. Myers et al. [1973] questionaram o fato do modelo logístico ter a variável resposta dicotômica 0 ou 1, independente do tempo de seguimento; isto é, a resposta que ocorre perto do início do intervalo de seguimento é dado o mesmo peso que à resposta que ocorre no fim do período. Para evitar isso, eles propuseram o modelo logístico exponencial, onde incorporam a variável tempo de sobrevivência.

Elandt-Johnson [1983] estudou o uso do modelo logístico em estudos prospectivos em situações em que ocorrem censuras simples e censuras múltiplas. **Censuras simples** ocorrem, em situações onde todos os indivíduos entram no estudo ao mesmo tempo e observa-se falha ou não falha após o período τ pré-fixado. Neste tipo de estudo, existem a informação de quantos não falharam e o comprimento do período de estudo, logo não é necessário conhecer a distribuição dos tempos, argumenta Elandt-Johnson. Em contrapartida, propõe o uso do modelo logístico paramétrico (exponencial entre outros) quando ocorrem censuras múltiplas. **Censuras múltiplas** ou progressivas ocorrem em situações onde todos os indivíduos entram no estudo em diferentes tempos e o tempo de seguimento τ é fixo no calendário. Assim cada indivíduo que falhar terá um período distinto de observação uns dos outros. O mesmo ocorre para cada indivíduo que não falhar. Desse modo, argumenta Elandt-Johnson, deve-se levar em conta os tempos potenciais de falha e não falha e, portanto necessita-se caracterizar a distribuição desses tempos.

O objetivo deste capítulo é fazer uma breve revisão do modelo logístico usual e apresentar as utilizações do modelo logístico para análise de dados de sobrevivência, sob várias condições estudadas. Na secção 2.2. é apresentado o modelo logístico e a interpretação dos seus

parâmetros. A secção 2.3 trata do modelo logístico estudado por Elandt-Johnson [1980]. Na secção 2.4 é dada ênfase aos modelos paramétricos introduzidos por Elandt-Johnson [1980] e Myers et al [1973]. A secção 2.5 apresenta um enfoque ao modelo logístico exponencial por partes. Na secção 2.7 é abordada a generalização de Mantel e Hankey [1978] usando o modelo logístico.

2.2. Modelo Logístico Linear.

O modelo logístico é próprio para analisar dados binários uma vez que o modelo de regressão usual não se ajustaria a tais dados. Esta secção é dedicada à descrição do modelo logístico e à interpretação dos seus parâmetros. Referências ao modelo de regressão logístico podem ser encontrados em Cox [1970] e Hosmer and Lemeshow [1984].

2.2.1. Modelo Logístico.

Suponha um experimento com n indivíduos onde n_1 desses indivíduos falham e o restante $n_2 = n - n_1$ não falham. Denota-se por X_i a variável resposta dicotômica, $i = 1, 2, \dots, n$ onde

$$X_i = \begin{cases} 1 & \text{se o } i\text{-ésimo indivíduo falhar (sucesso)} \\ 0 & \text{se o } i\text{-ésimo indivíduo não falhar (fracasso)} \end{cases} \quad (2.1)$$

Suponha ainda que para cada indivíduo, são medidas p variáveis independentes $Z_{i1}, Z_{i2}, \dots, Z_{ip}$ (características ou fatores de risco).

Uma formulação geral para o modelo é dado por:

$$\begin{aligned} \Pr\{x_i = 1 \mid z_i\} &= \Pr\{\text{indivíduo falhar dado o vetor de covariáveis}\} = p(z_i) \\ &= \left[1 + \exp[-g(z_i, \beta)] \right]^{-1} = \frac{\exp[g(z_i, \beta)]}{\left[1 + \exp[g(z_i, \beta)] \right]}, \end{aligned} \quad (2.2)$$

que é a probabilidade de falha, onde $g(z_i, \beta)$ é alguma função real de z_i e de um vetor de parâmetros $\beta^T = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$, onde β_j é o parâmetro que determina a contribuição da variável z_j para a resposta, β_0 é o coeficiente linear e $z^T = (z_0, z_1, \dots, z_p)$, com $z_0 \equiv 1$. A probabilidade de não falha é

$\Pr[x_i = 0 | z_i] = \Pr\{ \text{indivíduo não falhar dado o vetor de covariáveis} \} = q(z_i)$

$$= 1 - p(z_i) = P[x_i = 0] = \frac{1}{1 + \exp[g(z_i, \beta)]} \quad (2.3)$$

com

$$p(z_i) + q(z_i) = 1. \quad (2.4)$$

Obter soluções para as equações (2.2) e (2.3) é complicado; porém utilizando a transformação logarítmica para a razão de $p(z_i)$ e $q(z_i)$, obtém-se a função real $g(z_i, \beta)$, de fácil solução. Então o modelo logístico linear é definido por

$$\log_e \frac{p(z_i)}{q(z_i)} = \log_e \frac{P[x_i = 1 | z_i]}{P[x_i = 0 | z_i]} = g(z_i, \beta), \quad (2.5)$$

onde $\log_e \frac{p(z_i)}{q(z_i)}$ é chamado de transformação logística da probabilidade de falha.

Cox [1970] sugeriu um modelo no qual o logaritmo da resposta esperada era linear em z , isto é

$$g(z_i, \beta) = \beta^T z_i$$

em (2.5). Substituindo em (2.2) e (2.3), obtém-se

$$p(z) = 1 - q(z) = \frac{\exp(\beta^T z)}{1 + \exp(\beta^T z)} \quad (2.6)$$

e

$$q(z) = \frac{1}{1 + \exp(\beta^T z)} = [1 + \exp(\beta^T z)]^{-1}. \quad (2.7)$$

O modelo de probabilidade para X_i , $i = 1, \dots, n$ é o Bernouille com parâmetros $p(z_i)$ e $q(z_i)$

$$P(X_i = x_i) = P(X_i = 1 | z_i)^{x_i} P(X_i = 0 | z_i)^{1 - x_i}, \quad x_i = 0, 1 \quad (2.8)$$

A função de verossimilhança é então o produto de n distribuições de Bernouille e é definida por:

$$L(\beta) = \prod_{i=1}^n [p(z_i)]^{x_i} [q(z_i)]^{1-x_i} = \prod_{i=1}^n \left[\frac{p(z_i)}{q(z_i)} \right]^{x_i} q(z_i). \quad (2.9)$$

Os parâmetros $\underline{\beta}$ podem ser estimados pelos métodos clássicos de estimação usados em modelo de análise de regressão linear. A estimação dos parâmetros do modelo logístico não se restringe ao método de máxima verossimilhança. O método dos mínimos quadrados ponderados, introduzido por Duncan e Walker [1967] também pode ser utilizado.

2.2.2. Interpretação dos Parâmetros.

A interpretação dos parâmetros envolve duas questões:

1. determinar a função de variáveis dependente e variáveis independentes,
2. e definir a unidade de mudança na variável independente.

Primeiramente, considera-se um modelo com uma única covariável z , que pode ser do tipo discreto ou contínuo. Suponha linearidade da função $g(\cdot)$ em (2.5)

$$g(z) = \log_e \frac{p(z)}{q(z)} = \beta_0 + \beta_1 z, \tag{2.10}$$

onde β_1 é o coeficiente angular, i.e., é a diferença entre os valores da variável dependente $z + 1$ até z . $(g(z) = \beta_0 + \beta_1 z \Rightarrow \beta_1 = g(z + 1) - g(z))$

1. z é uma variável discreta, onde z assume os valores $z = 0$ ou 1 . Sob este modelo existem dois valores para $p(z_i)$ e $q(z_i)$, por conveniência estes valores serão apresentados numa tabela 2x2.

Tabela 2.1. Relação entre a variável resposta e variável independente.

		variável independente z	
		$z = 1$	$z = 0$
variável resposta x	$x = 1$	$p(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$p(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
	$x = 0$	$q(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$q(0) = \frac{1}{1 + e^{\beta_0}}$
Total		1	1

A probabilidade de falha para $z = 1$ e $z = 0$ é definido por $\frac{p(1)}{q(1)}$ e $\frac{p(0)}{q(0)}$ respectivamente. O logaritmo da probabilidade de falha é

$$g(1) = \log \left\{ \frac{p(1)}{q(1)} \right\} \quad \text{e} \quad g(0) = \log \left\{ \frac{p(0)}{q(0)} \right\}.$$

Denotando por ψ a *odds ratio*, é definido pela razão dos *odds* de $z = 1$ por $z = 0$

$$\psi = \frac{p(1) / q(1)}{p(0) / q(0)}. \tag{2.11}$$

O logaritmo da *odds ratio*, denominado por logito é

$$\log(\psi) = \log \left(\frac{p(1) / q(1)}{p(0) / q(0)} \right) = g(1) - g(0) \tag{2.12}$$

que é a diferença dos logitos.

Agora, usando as expressões mostradas na Tabela 2.1 o *odds ratio* é

$$\psi = \frac{p(1) [1 - p(0)]}{p(0) [1 - p(1)]} = e^{\beta_1}. \tag{2.13}$$

Quando z assume valores genéricos, por exemplo a ou b , ao invés de 0 ou 1 $\psi = e^{\beta_1(a - b)}$, logo a interpretação dos parâmetros através de ψ não pode ser realizada sem antes conhecer a codificação adotada. Para efeito de simplificação adota-se $a = 1$ e $b = 0$. Donde, para regressão logística com uma variável independente dicotômica $\psi = e^{\beta_1}$ e a diferença do logito é $\log(\psi) = \log(e^{\beta_1}) = \beta_1$. O *odds ratio* é uma medida de associação e a interpretação desses parâmetros reflete três situações quando associa a variável x com a covariável z :

- a. $\psi = 1$ indica inexistência de associação;
- b. $0 < \psi < 1$ associação negativa;
- c. $\psi > 1$ associação positiva.

Em suma para uma variável dicotômica o parâmetro de interesse é a *odds ratio*.

2. z é uma variável contínua, o coeficiente β_1 pode ser interpretado como a variação ocorrida no logaritmo da *odds ratio* quando há um acréscimo de uma unidade em z , i.e., $\beta_1 = g(z + 1) - g(z)$ para todo valor de z , onde $g(z)$ é definida em (2.10).

Em alguns problemas o acréscimo de uma unidade não seria interessante. Por exemplo, em estudos biológicos o acréscimo de 1 ano na idade ou de 1mm hg na pressão sanguínea sistólica pode não ser importante, porém o acréscimo de 10 anos ou 10 mm hg poderia ser

considerado mais interessante. Portanto, se ao invés de impor o acréscimo de uma unidade, considerar-se um acréscimo de uma constante "c" de unidades de variação de z, o logaritmo da *odds ratio* muda para

$$g(z + c) - g(z) = c\beta_1. \tag{2.14}$$

Quando $g(z)$ não é linear em z deve-se agrupar a covariável contínua para análise, ou ajustar um outro tipo de relação que melhor se adapte aos dados, por exemplo, uma função quadrática. Para melhor compreensão desta discussão ver Hosmer and Lemeshow [1984 - Cap 4].

2.3. Modelo Logístico Linear para Análise de Sobrevida (Elandt-Johnson).

Suponha que todos os indivíduos entram no estudo ao mesmo tempo, e τ é o tempo de seguimento desde a entrada até a falha ou censura (tipo I, i. e., fixo um intervalo de tempo τ de observação do experimento) e z não depende de t.

A probabilidade de sobrevida até o tempo τ é

$$\Pr [T > \tau | z] = S(\tau | z),$$

e a probabilidade de falhar antes do tempo τ é

$$\Pr [T \leq \tau | z] = 1 - S(\tau | z).$$

O modelo logístico sob um período τ pré-fixado pode ser definido na forma

$$\log \frac{1 - S(\tau | z)}{S(\tau | z)} = G(z(\tau), \underline{\beta}) \tag{2.15}$$

onde $G(\cdot)$ é uma função real dos τ e z.

Nosso interesse é o modelo $S_T(\cdot)$ talque $g(\cdot)$ pode ser aproximado para uma função linear dos z, e o z podem ou não ser função linear de t (ou são constante).

$$\log \frac{1 - S(\tau | z)}{S(\tau | z)} \simeq \underline{\beta}^T z(t) \tag{2.16}$$

onde $\underline{\beta}^T = (\beta_0, \beta_1, \dots, \beta_p)$ vetor dos parâmetros desconhecidos. Escrevendo $z(t)$ por z, podemos considerar o seguinte modelo

$$\log \frac{1 - S(\tau | \mathbf{z})}{S(\tau | \mathbf{z})} = \beta^T \mathbf{z} \quad \text{onde } z_{j0} \equiv 1 \quad (2.17)$$

Portanto,

$$\frac{1 - S(\tau | \mathbf{z})}{S(\tau | \mathbf{z})} = \exp(\beta^T \mathbf{z}) \quad \text{e} \quad S(\tau | \mathbf{z}) = \frac{1}{1 + \exp(\beta^T \mathbf{z})} = q(\tau | \mathbf{z}) \quad (2.18)$$

é probabilidade de sobreviver e

$$1 - S(\tau | \mathbf{z}) = 1 - \frac{1}{1 + \exp(\beta^T \mathbf{z})} = \frac{\exp(\beta^T \mathbf{z})}{1 + \exp(\beta^T \mathbf{z})} = p(\tau | \mathbf{z}) \quad (2.19)$$

é probabilidade de falhar.

Suponha um estudo onde se observa n indivíduos e que destes n_1 falham até τ . Defina-se

$$\delta_i = \begin{cases} 1 & \text{se o indivíduo (i) falhar no tempo } \tau \\ 0 & \text{se o indivíduo (i) não falhar no tempo } \tau \end{cases} \quad (2.20)$$

O modelo de probabilidade associado a δ_i é exatamente igual ao (1.8), Bernouille com parâmetros $[1 - S(\tau | \mathbf{z})]$ e $S(\tau | \mathbf{z})$. Duas observações tornam-se importantes neste momento:

1. os indivíduos censurados antes de τ não são utilizados para estimar os parâmetros, o que implica que $N_T \leq N$.

2. Os tempos de falhas e censuras não são utilizados.

A função de verossimilhança é dada por

$$L = \prod_{i=1}^{N_T} [1 - S(\tau | \mathbf{z})]^{\delta_i} [S(\tau | \mathbf{z})]^{1 - \delta_i}$$

$$L(\beta) = \prod_{i=1}^{N_T} \left(\frac{\exp(\beta^T \mathbf{z}_i)}{1 + \exp(\beta^T \mathbf{z}_i)} \right)^{\delta_i} \left(\frac{1}{1 + \exp(\beta^T \mathbf{z}_i)} \right)^{1 - \delta_i} = \prod_{i=1}^{N_T} \frac{\exp(\delta_i \beta^T \mathbf{z}_i)}{1 + \exp(\beta^T \mathbf{z}_i)} \quad (2.21)$$

e

$$\log L(\beta) = \sum_{i=1}^{N_T} \delta_i \beta^T \mathbf{z}_i - \sum_{i=1}^{N_T} \log [1 + \exp(\beta^T \mathbf{z}_i)] \quad (2.22)$$

A função score $U(\underline{\beta}) = \frac{\partial}{\partial \underline{\beta}} \log L(\underline{\beta})$, tem como componente j ,

$$\frac{\partial}{\partial \beta_j} \log L(\underline{\beta}) = \sum_{i=1}^{N_T} \delta_i z_{ij} - \sum_{i=1}^{N_T} \frac{z_{ij} \exp(\underline{\beta}^T z_i)}{1 + \exp(\underline{\beta}^T z_i)} = \sum_{i=1}^{N_T} z_{ij} [\delta_i - [1 - S(\tau | z)]] \quad (2.23)$$

com $j = 0, 1, \dots, p$.

O elemento (jv) da matriz $(p+1) \times (p+1)$ das derivadas segunda é

$$H(\underline{\beta}) = \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\underline{\beta}) = - \sum_{i=1}^{N_T} \frac{z_{ij} z_{iv} \exp(\underline{\beta}^T z_i)}{\{1 + \exp(\underline{\beta}^T z_i)\}^2}.$$

$$H(\underline{\beta}) = - \sum_{i=1}^{N_T} z_{ij} \left[\frac{\exp(\underline{\beta}^T z_i)}{1 + \exp(\underline{\beta}^T z_i)} \frac{1}{1 + \exp(\underline{\beta}^T z_i)} \right] z_{iv} = - \sum_{i=1}^{N_T} z_{ij} \left[[1 - S(\tau | z)] S(\tau | z) \right] z_{iv}. \quad (2.24)$$

Esta matriz é também conhecida por matriz Hessiana da função (2.22). Portanto, a matriz de informação de Fisher é dada por

$$I(\underline{\beta}) \approx E \left[- \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\underline{\beta}) \right].$$

A inversa de $I(\underline{\beta})$ é a matriz de covariância assintótica dos estimadores de máxima verossimilhança dos parâmetros da regressão logística, cuja estimativa de máxima verossimilhança será dada por $[I(\hat{\underline{\beta}})]^{-1}$.

O estimador de $\underline{\beta}$ é obtido pelo método de máxima verossimilhança onde

$$U(\underline{\beta}) = \sum_{i=1}^{N_T} [\delta_i - [1 - S(\tau | z)]] z_{ij} = 0,$$

a solução do sistema acima requer o uso de método iterativo, já mencionado anteriormente (subsecção 1.5.1), com os valores iniciais de $\underline{\beta}^0 = 0$ o que implica $S(\tau | z) = 0.5$. Os parâmetros estimados podem ser testado utilizando as estatísticas definidas na secção 1.7.

A função de sobrevivência estimada é

$$\hat{S}(t | z) = [1 + \exp(\hat{\underline{\beta}}^T z)]^{-1}. \quad (2.25)$$

De (1.30) tem-se que $\text{Var}[\hat{S}(t|z)] = \left[\frac{\partial}{\partial \beta} \hat{S}(t|z) \right]^T I^{-1}(\hat{\beta}) \left[\frac{\partial}{\partial \beta} \hat{S}(t|z) \right]$, onde

$$\frac{\partial}{\partial \beta} \hat{S}(t|z) = -z_{ij} \frac{\exp(\hat{\beta}^T z)}{[1 + \exp(\hat{\beta}^T z)]^2} = -z_{ij} \hat{S}_i(t|z) \frac{1}{1 + \exp(-\hat{\beta}^T z)}. \quad (2.26)$$

Então $\hat{S}(t|z) \approx \mathcal{N}[0, \text{Var}(\hat{S}(t|z))]$. Do exposto pode-se construir intervalos de $100(1 - \alpha)\%$ de confiança para a função de sobrevivência, através da expressão 1.31. Foi desenvolvido um programa computacional no módulo IML do SAS, PESVSLAS.PRG (Apêndice D3.7), para efetuar os cálculos da estimativa da função de sobrevivência $S(t|z)$ expressada em (2.25) e sua variância utilizando (1.30) cuja derivada de $S(t|z)$ é dado em (2.26).

2.4. Modelo Logístico Exponencial para Análise de Sobrevida (Elandt-Johnson e Myers).

Myers et al [1973] e Elandt-Johnson [1983] utilizaram os tempos de vida no modelo de regressão logístico, em situações onde a entrada dos indivíduos no estudo se verifica em tempos diferentes e ainda com tempo de seguimento τ fixo no calendário.

Quando os tempos de vida seguem uma distribuição exponencial com parâmetro $\lambda(z)$ constante, com função sobrevivência $S(t|z)$ dada por

$$S(t|z) = \exp[-\lambda(z)t], \quad 0 < t \leq \tau \quad (2.27)$$

o logaritmo da *odds ratio* é uma função linear de z sob um período $(0, \tau)$, i. e.

$$\log \frac{p(\tau|z)}{q(\tau|z)} = \log \frac{1 - \exp(-\lambda(z)\tau)}{\exp(-\lambda(z)\tau)} = \sum_{j=0}^p \beta_j z_{ji} = \beta^T z \quad (2.28)$$

Assim

$$q(\tau|z) = \exp(-\lambda(z)\tau) = [1 + \exp(\beta^T z)]^{-1}, \quad (2.29)$$

que é igual a (2.7). Assim, para $0 < t \leq \tau$ tem-se que:

$$\lambda(z) = \frac{1}{\tau} \log [1 + \exp(\beta^T z)] \quad (2.30)$$

O período τ pode ser pré-fixado, ou como Myers et al [1973] propôs, τ pode ser um

parâmetro desconhecido que pode ser estimado.

Dois casos especiais são de interesse:

(i) - Suponha que τ é muito pequeno (i. e., $\tau \rightarrow 0$). Então (2.28) torna-se

$$\log \frac{1 - \exp(-\lambda(z)\tau)}{\exp(-\lambda(z)\tau)} = \log[1 - \exp(-\lambda(z)\tau)] \doteq \log[\lambda(z)\tau] = \underline{\beta}^T \mathbf{z}$$

ou

$$\lambda(z) = \frac{1}{\tau} \exp(\underline{\beta}^T \mathbf{z}) = \frac{1}{\tau} \exp(\beta_0 + \underline{\beta}^{*T} \mathbf{z}^*) = \lambda_0 \exp(\underline{\beta}^{*T} \mathbf{z}^*)$$

onde $\lambda_0 = \frac{1}{\tau} \exp(\beta_0)$, $\underline{\beta}^* = (\beta_1, \beta_2, \dots, \beta_p)$ e $\mathbf{z}^* = (z_1, z_2, \dots, z_p)$.

Este é um caso especial do modelo de Cox [1972]

$$\lambda(t | \mathbf{z}) = \lambda(t) \exp(\underline{\beta}^{*T} \mathbf{z}^*)$$

com a vizinhança de risco $\lambda(t) = \lambda_0$ sob o período $(0, \tau)$.

(ii) - Suponha que τ é muito grande (i. e., $\tau \rightarrow \infty$). Então (2.28) torna-se

$$\log \frac{1 - \exp(-\lambda(z)\tau)}{\exp(-\lambda(z)\tau)} = \log[\exp(\lambda(z)\tau)] = \lambda(z)\tau = \underline{\beta}^T \mathbf{z},$$

onde $\lambda(z) = \frac{1}{\tau} \underline{\beta}^T \mathbf{z} = \gamma_0 + \gamma^T \mathbf{z}^*$, $\gamma_0 = \frac{1}{\tau} \beta_0$ e $\gamma = \frac{1}{\tau} \underline{\beta}^*$ que é um modelo aditivo da função razão de risco.

De modo similar à secção anterior, considera-se n indivíduos num estudo. Seja n_i o número de mortes e t_i o tempo no qual o indivíduo i , com característica \mathbf{z}_i , foi visto pela última vez no intervalo de observação τ ($0 < t_i \leq \tau$). Assim,

$$t_i = \begin{cases} t_i & \text{se o indivíduo (i) morre antes de } \tau \\ \tau & \text{se o indivíduo (i) sobrevive a } \tau \end{cases}$$

Nota-se que os casos perdidos de observação (i.e., censurados antes de τ), não são considerados para estimar os parâmetros. A função de verossimilhança neste caso é dada por

$$L(\tau; \underline{\beta}) = \prod_{i=1}^n \lambda(z_i)^{\delta_i} \exp(-\lambda(z_i) t_i), \tag{2.31}$$

onde $\delta_i = \begin{cases} 1 & \text{se o } i^{\text{o}} \text{ indivíduo morre até } t_i \\ 0 & \text{se o } i^{\text{o}} \text{ indivíduo não morre até } t_i \end{cases}$.

logo a função a ser maximizada é

$$\log L(\tau; \beta) = \sum_{i=1}^n \left[\delta_i \log \lambda(z_i) - \lambda(z_i)t_i \right],$$

substituindo $\lambda(z_i) = \frac{1}{\tau} \log [1 + \exp(\beta^T z_i)]$ tem-se que

$$\log L(\tau; \beta) = \sum_{i=1}^n \delta_i \log \left\{ \log [1 + \exp(\beta^T z_i)] \right\} - \sum_{i=1}^n \delta_i \log \tau - \frac{1}{\tau} \sum_{i=1}^n t_i \log [1 + \exp(\beta^T z_i)] \tag{2.32}$$

Nota-se que, os tempos de falha e censura são considerados no modelo e o período de seguimento τ é usado para pondera-los. Assim tem-se duas situações a ser considerada:

1. τ é conhecido - Neste caso τ é pré-fixado, portanto a equação de verossimilhança é

$$\begin{aligned} \frac{\partial}{\partial \beta_j} \log L(\tau; \beta) &= \sum_{i=1}^n \delta_i \frac{z_{ij} \exp(\beta^T z_i)}{1 + \exp(\beta^T z_i)} \frac{1}{\log [1 + \exp(\beta^T z_i)]} - \frac{1}{\tau} \sum_{i=1}^n t_i \frac{z_{ij} \exp(\beta^T z_i)}{1 + \exp(\beta^T z_i)} \\ &= \sum_{i=1}^n \frac{z_{ij}}{1 + \exp(-\beta^T z_i)} \left\{ \frac{\delta_i}{\log [1 + \exp(\beta^T z_i)]} - \frac{t_i}{\tau} \right\} \end{aligned} \tag{2.33}$$

com $j = 0, 1, \dots, p$. A matriz das derivadas segundas é:

$$\begin{aligned} V_{\beta\beta} &= \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\tau; \beta) \\ &= \sum_{i=1}^n z_{ij} \left\{ \frac{z_{iv} \exp(-\beta^T z_i)}{[1 + \exp(-\beta^T z_i)]^2} \left[\frac{\delta_i}{\log [1 + \exp(\beta^T z_i)]} - \frac{t_i}{\tau} \right] - \frac{1}{1 + \exp(-\beta^T z_i)} \left[\frac{\delta_i}{\log [1 + \exp(\beta^T z_i)]} \frac{z_{iv} \exp(\beta^T z_i)}{1 + \exp(\beta^T z_i)} \right] \right\} = \end{aligned}$$

$$= \sum_{i=1}^n z_{ij} \frac{1}{[1 + \exp(-\beta^T z_i)]^2} z_{iv} \left[\exp(-\beta^T z_i) \left\{ \frac{\delta_i}{\log[1 + \exp(\beta^T z_i)]} - \frac{t_i}{\tau} \right\} - \frac{\delta_i}{[\log[1 + \exp(\beta^T z_i)]]^2} \right]. \quad (2.34)$$

2. τ é desconhecido - Neste segundo caso o estimador de máxima verossimilhança de τ é dado por

$$\frac{\partial}{\partial \tau} \log L(\tau; \beta) = -\frac{1}{\tau} \sum_{i=1}^n \left\{ \delta_i - \frac{1}{\tau} t_i \log[1 + \exp(\beta^T z_i)] \right\} = 0 \quad (2.35)$$

portanto

$$\hat{\tau} = \frac{1}{\sum_{i=1}^n \delta_i} \sum_{i=1}^n t_i \log[1 + \exp(\beta^T z_i)] \quad (2.36)$$

que é uma função dos elementos de β .

Substituindo (2.36) na equação (2.33) acima, obtém-se as equações de máximo verossimilhança somente como função dos parâmetros β , logo

$$\frac{\partial}{\partial \beta_j} \log L(\tau; \beta) = \sum_{i=1}^n \frac{z_{ij}}{1 + \exp(-\beta^T z_i)} \left\{ \frac{\delta_i}{\log[1 + \exp(\beta^T z_i)]} - \frac{t_i^{n_F}}{\sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)]} \right\}, \quad (2.37)$$

onde $n_F = \sum_{i=1}^n \delta_i$.

Obs:- o índice h foi introduzido com o intuito de simplificar e não confundir quando a derivada segunda da equação acima fosse calculada.

O elemento (j,v) da matriz da segundas derivadas parciais em relação aos β é:

$$\begin{aligned} V_{\beta\beta} &= \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\tau; \beta) \\ &= \sum_{i=1}^n \left\{ \frac{z_{ij} z_{iv} \exp(-\beta^T z_i)}{[1 + \exp(-\beta^T z_i)]^2} \left[\frac{\delta_i}{\log[1 + \exp(\beta^T z_i)]} - \frac{t_i^{n_F}}{\sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)]} \right] - \right. \\ &\quad \left. \frac{z_{ij}}{1 + \exp(-\beta^T z_i)} \left[\frac{\delta_i}{[\log[1 + \exp(\beta^T z_i)]]^2} \frac{z_{iv} \exp(\beta^T z_i)}{1 + \exp(\beta^T z_i)} \right] + \right. \\ &\quad \left. \frac{z_{ij}}{1 + \exp(-\beta^T z_i)} \left[\frac{t_i^{n_F}}{[\sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)]]^2} \sum_{h=1}^n t_h \frac{z_{hv} \exp(\beta^T z_h)}{1 + \exp(\beta^T z_h)} \right] \right\} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \left\{ \frac{z_{ij} z_{iv}}{[1 + \exp(-\underline{\beta}^T z_i)]^2} \left\{ \exp(-\underline{\beta}^T z_i) \left[\frac{\delta_i}{\log[1 + \exp(\underline{\beta}^T z_i)]} - \frac{t_i n_F}{\sum_{h=1}^n t_h \log[1 + \exp(\underline{\beta}^T z_h)]} \right] \right. \right. \\
 &\left. \left. - \frac{\delta_i}{[\log[1 + \exp(\underline{\beta}^T z_i)]]^2} \right\} + \frac{z_{ij}}{1 + \exp(-\underline{\beta}^T z_i)} \left[\frac{t_i n_F}{\left[\sum_{h=1}^n t_h \log[1 + \exp(\underline{\beta}^T z_h)] \right]^2} \sum_{h=1}^n t_h \frac{z_{hv}}{1 + \exp(-\underline{\beta}^T z_h)} \right] \right\}
 \end{aligned}
 \tag{2.38}$$

Na obtenção dos estimadores dos β_j , o procedimento usado foi o de Newton-Raphson, já mencionado anteriormente, com valores iniciais de $\beta_0 = \beta_1 = \dots = \beta_p = 0$. Após estimar os $\underline{\beta}$, substitui-se seus valores em (2.36) e tem-se o valor estimado de τ .

Nota-se que em algumas circunstâncias não é possível conhecer o tempo exato onde o evento ocorre, porém tem-se informações do intervalo de tempo onde eles ocorrem. A função de verossimilhança, quando os t_i são intervalos de tempos, é discutido por Myers et al [1973].

A função de sobrevivência estimada para $0 < t_i \leq \tau$ é

$$\hat{S}(t | z) = \exp\left(- \int_0^{t_i} \lambda(z) du\right) = \exp\left(- \lambda(z) \int_0^{t_i} du\right) = \exp\left(- \lambda(z) t_i\right),$$

substituindo o valor de $\lambda(z)$ dado por (2.30) na expressão acima tem-se

$$\begin{aligned}
 \hat{S}(t | z) &= \exp\left(- \frac{t_i}{\tau} \log[1 + \exp(\underline{\beta}^T z)]\right) = \exp\left(\log[1 + \exp(\underline{\beta}^T z)]^{-\frac{t_i}{\tau}}\right) = \\
 &= [1 + \exp(\underline{\beta}^T z)]^{-\frac{t_i}{\tau}}.
 \end{aligned}
 \tag{2.39}$$

Assim a variância assintótica de $\hat{S}(t | z)$ é dado por $[G(\theta)]^T \text{var}(\hat{\underline{\beta}}) G(\theta)$, onde $G(\theta)$ é um vetor linha sendo que os seus elementos são as derivadas de $\hat{S}(t | z)$ em relação a $\underline{\beta}$, porém deve-se lembrar que as derivadas serão diferentes considerando as duas situações acima, logo tem-se

1. τ conhecido;

$$\frac{\partial}{\partial \beta_j} S(t | z) = - z_{ij} \frac{t_i}{\tau} \exp(\underline{\beta}^T z_i) \{1 + \exp(\underline{\beta}^T z_i)\}^{-\frac{t_i}{\tau} - 1}
 \tag{2.40}$$

2. τ desconhecido; substituiu-se o valor de $\hat{\tau}$ dado em (2.36) na equação (2.39), obter-se-a uma equação somente em função dos β . A derivada é

$$\begin{aligned} \frac{\partial}{\partial \beta_j} S(t | z) = & - z_{ij} \exp(\beta^T z_i) \frac{t_i n_F}{\sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)]} [1 + \exp(\beta^T z_i)]^{-\frac{t_i n_F}{\sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)]} - 1} + \\ & [1 + \exp(\beta^T z_i)]^{-\frac{t_i n_F}{\sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)]}} \left\{ \log[1 + \exp(\beta^T z_i)] \right\} \left\{ \frac{t_i n_F}{\left\{ \sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)] \right\}^2} \right\}^* \\ & \sum_{h=1}^n t_h \frac{z_{ij} \exp(\beta^T z_h)}{1 + \exp(\beta^T z_h)}. \end{aligned}$$

A expressão acima pode ser simplificada substituindo alguns valores já conhecido, logo

$$\begin{aligned} \frac{\partial}{\partial \beta_j} S(t | z) = & - z_{ij} \exp(\beta^T z_i) \frac{t_i}{\tau} \left\{ 1 + \exp(\beta^T z_i) \right\}^{-\frac{t_i}{\tau} - 1} + \\ S(t | z) & \left\{ \log[1 + \exp(\beta^T z_i)] \right\} \frac{t_i}{\tau \sum_{h=1}^n t_h \log[1 + \exp(\beta^T z_h)]} \sum_{h=1}^n t_h \frac{z_{ij} \exp(\beta^T z_h)}{1 + \exp(\beta^T z_h)}. \end{aligned} \quad (2.41)$$

Com base nas informação acima, o intervalo de $100(1 - \alpha)\%$ de confiança para $\hat{S}(t | z)$ é dado por (1.31).

Utilizando o módulo IML do SAS, foi desenvolvido o programa PAMLEAS.PRG (Apêndice D3.8) para ajustar o modelo logístico exponencial quando τ é conhecido (2.28) e testar seus parâmetros, tal como calcular as estimativas da função de sobrevivência (2.39) e a variância de $S(t | z)$ dado por (1.30) com a derivada da função de sobrevivência expressada em (2.40).

2.5. Modelo Logístico Exponencial por Partes para Análise de Sobrevidência.

Na análise de sobrevivência há situações onde a função de risco não é constante durante todo o período do estudo; porém, pode ser constante em fragmentos deste período. Por exemplo, suponha uma situação onde o risco é uma constante λ_1 até um determinado tempo t_1 , e uma

constante λ_2 até um tempo t_2 , e assim por diante.

Então,

$$\lambda(t) = \begin{cases} \lambda_1 & \text{se } 0 \leq t < t_1 \\ \lambda_2 & \text{se } t_1 \leq t < \infty \end{cases} \quad (2.42)$$

a função densidade é dada por

$$f(t) = \begin{cases} \lambda_1 \exp(-\lambda_1 t) & 0 < t < t_1 \\ \lambda_2 \exp(-\lambda_1 t_1) \exp(-\lambda_2(t - t_1)) & t_1 \leq t < \infty \end{cases} \quad (2.43)$$

e a função de sobrevivência é dado por

$$S(t) = \begin{cases} \exp(-\lambda_1 t) & 0 < t < t_1 \\ \exp(-\lambda_1 t_1) \exp(-\lambda_2(t - t_1)) & t_1 \leq t < \infty \end{cases} \quad (2.44)$$

O estudo descrito acima pode ser estendido a uma situação onde as covariáveis usadas no modelo tem características contínuas, sujeitas a mudanças no tempo, e o tipo de mudança na maioria das vezes é desconhecida. Poderia ser por exemplo, variações aleatórias dependendo de muitos fatores não controlados, ou alguma dependência de covariáveis no tempo, cuja relação é difícil de ser estabelecida. Uma solução é transformar uma variável contínua em discreta, entretanto, pode haver perdas de informação e uma pergunta que surge é se esta é a forma correta de resolver tais problemas.

Suponha-se que pacientes com câncer indicam uma alta taxa de mortalidade nos primeiros anos após diagnósticos, mas pacientes que sobrevivem à este período crítico reagem melhor a doença. Neste caso não é possível ajustar uma função de distribuição simples, i.e., de forma matemática simples, de modo que cubra a variação dos dados.

Seja τ o período de investigação e $M = 1, 2, \dots, k$ são pontos fixos ($0 < t_1 < \dots < t_k < \infty$) até que uma mudança na função de sobrevivência ocorra. Estes pontos podem ser sugeridos por disposição gráfica. Cada intervalo apresentará um novo conjunto de medidas, i.e., para o i -ésimo indivíduo no intervalo $I_k = [t_k, t_{k+1})$, de comprimento $\tau_k = t_{k+1} - t_k$, $k = 0, 1, 2, \dots, M-1$; o conjunto de covariáveis é $\mathbf{z}_{ki}^T = (z_{1ki}, z_{2ki}, \dots, z_{pki})$, que podem ser tomadas até o último contato com o i -ésimo indivíduo observado durante o intervalo I_k . Geralmente as funções de sobrevivência, podem pertencer a diferentes famílias, entretanto na prática, observa-se que são membros diferentes de uma mesma família. Frequentemente aproxima-se esta distribuição desconhecida por uma exponencial por partes, com taxa de risco para o k -ésimo período da forma:

$$\lambda_k(t | \mathbf{z}) = \lambda_k \exp(\beta^{*T} \mathbf{z}_i^*) \quad t \in I_k \quad (2.45)$$

A taxa de risco do i -ésimo indivíduo no k -ésimo intervalo é denotado por λ_{ik} , assumindo que $\lambda_{ik} > 0$ para cada (i, k) .

A relação logístico-exponencial sob o período I_k é escrito por

$$\log \frac{p_k(\tau_k | z)}{q_k(\tau_k | z)} = \log \frac{1 - \exp[-\lambda_k(z)\tau_k]}{\exp[-\lambda_k(z)\tau_k]} = \gamma_k + \beta^{*T}z^* \quad (2.46)$$

para $k = 0, 1, \dots, M-1$.

Para $t_k \leq t < t_{k+1}$ as probabilidades de falhas e sobrevivências são

$$q_k(z) = [1 + \exp(\gamma_k + \beta^{*T}z_{ik}^*)]^{-1}, \quad (2.47)$$

$$p_k(z) = 1 - q_k(z) = \frac{\exp(\gamma_k + \beta^{*T}z_{ik}^*)}{1 + \exp(\gamma_k + \beta^{*T}z_{ik}^*)}, \quad e \quad (2.48)$$

$$\lambda_k(z) = \frac{1}{\tau_k} \log [1 + \exp(\gamma_k + \beta^{*T}z_{ik}^*)] \quad (2.49)$$

para $t_k \leq t < t_{k+1}$. Onde γ_k (β_{0k}) assume valores diferentes em cada intervalo.

Supondo que os dados são agrupados em intervalos fixos, mas também se conhecem os tempos individuais de mortes, então t_{ki} representa o tempo no qual o i -ésimo indivíduo foi visto pela última vez no intervalo $[t_k, t_{k+1})$, i.e

$$t_{ki} = \begin{cases} t_{ki} = t_i & \text{se } t_i < t_{k+1} \\ t_{ki} = t_{k+1} & \text{se } t_i \geq t_{k+1} \end{cases}$$

Seja $\tau_{ki} = t_{ki} - t_k$ ($0 < \tau_{ki} \leq \tau_k$) o tempo que o indivíduo i permaneceu no intervalo k . Para melhor entendimento considera-se, por exemplo, um estudo onde observam-se alguns indivíduos por um período de 12 meses. Supondo que este dados foram agrupados em intervalos de 6 meses, logo $k=0, 1$. Suponha-se que um indivíduo tem tempo de falha $t_i = 10$ meses. Este indivíduo sobrevive ao primeiro intervalo de 6 meses e vem a falhar no segundo intervalo; então $t_{0i} = 6$ e $t_{1i} = 10$, com $\tau_{0i} = 6 - 0 = 6$ e $\tau_{1i} = 10 - 6 = 4$.

Seja z_{ki} o conjunto de características medidas para o i -ésimo indivíduo no intervalo $[t_k, t_{k+1})$. Nota-se que o vetor das covariáveis pode ser observado a cada intervalo, porém, quando z são medidos uma única vez, os z_{ki} podem ser substituídos por z_i .

Define-se a variável dicotômica δ_{ki} ,

$$\delta_{ki} = \begin{cases} 1 & \text{se o } i\text{-ésimo indivíduo morre até } t_{ki} \\ 0 & \text{se o } i\text{-ésimo indivíduo vive até } t_{ki} \end{cases}$$

Substituindo δ_i , z_i e t_i em (2.32) por δ_{ki} , z_{ki} e τ_{ki} , respectivamente, a função de verossimilhança para o intervalo $[t_k, t_{k+1})$ é

$$L_k(\tau_k, \gamma_k, \beta) = \prod_{i=1}^n [\lambda_k(z_{ki})]^{\delta_{ki}} \exp[-\lambda_k(z_{ki}) \tau_{ki}]. \quad (2.50)$$

Nota-se que, somente os indivíduos mortos no intervalo e ainda vivos após τ_k são excluídos a cada intervalo, e a verossimilhança total é

$$L(\tau, \gamma, \beta) = \prod_{k=0}^{M-1} L_k(\beta). \quad (2.51)$$

Certamente

$$\log L(\tau, \gamma, \beta) = \sum_{k=0}^{M-1} \sum_{i=1}^n \left\{ \delta_{ki} \log \left[\log \left[1 + \exp(\gamma_k + \beta^{*T} z_{ki}^*) \right] \right] - \delta_{ki} \log \tau_k - \frac{\tau_{ki}}{\tau_k} \log \left[1 + \exp(\gamma_k + \beta^{*T} z_{ki}^*) \right] \right\} \quad (2.52)$$

As equações normais são:

$$U(\gamma_k) = \frac{\partial}{\partial \gamma_k} \log L_k(\tau_k, \gamma_k, \beta) = \sum_{i=1}^n \left\{ \frac{\delta_{ki}}{\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right]} \frac{\exp(\gamma_k + \beta^{*T} z_i^*)}{1 + \exp(\gamma_k + \beta^{*T} z_i^*)} - \frac{1}{\tau_k} \frac{\tau_{ki}}{\tau_k} \frac{\exp(\gamma_k + \beta^{*T} z_i^*)}{1 + \exp(\gamma_k + \beta^{*T} z_i^*)} \right\} = \sum_{i=1}^n \frac{1}{1 + \exp(-\gamma_k - \beta^{*T} z_i^*)} \left\{ \frac{\delta_{ki}}{\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right]} - \frac{\tau_{ki}}{\tau_k} \right\}, \quad (2.53)$$

com $k = 0, 1, \dots, M-1$;

$$\frac{\partial}{\partial \beta_j} \log L_k(\tau_k, \gamma_k, \beta) = \sum_{i=1}^n \left\{ \frac{z_{kij} \delta_{ki}}{\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right]} \frac{\exp(\gamma_k + \beta^{*T} z_i^*)}{1 + \exp(\gamma_k + \beta^{*T} z_i^*)} - \frac{1}{\tau_k} \frac{\tau_{ki}}{\tau_k} \frac{z_{kij} \exp(\gamma_k + \beta^{*T} z_i^*)}{1 + \exp(\gamma_k + \beta^{*T} z_i^*)} \right\} =$$

$$\sum_{i=1}^n \frac{z_{kij}}{1 + \exp(-\gamma_k - \beta^{*T} z_i^*)} \left\{ \frac{\delta_{ki}}{\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]} - \frac{\tau_{ki}}{\tau_k} \right\},$$

com $j = 1, 2, \dots, p$; portanto

$$U(\beta_j) = \frac{\partial}{\partial \beta_j} \log L(\tau_k, \gamma_k, \beta) = \sum_{k=1}^{M-1} \left[\sum_{i=1}^n \frac{z_{kij}}{1 + \exp(-\gamma_k - \beta^{*T} z_i^*)} \left\{ \frac{\delta_{ki}}{\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]} - \frac{\tau_{ki}}{\tau_k} \right\} \right] \quad (2.54)$$

A matriz das derivadas segundas é:

$$\begin{aligned} V_{\gamma\gamma} &= \frac{\partial^2}{\partial \gamma_k^2} \log L(\tau_k, \gamma_k, \beta) = \sum_{i=1}^n \left\{ \frac{\exp(-\gamma_k - \beta^{*T} z_i^*)}{[1 + \exp(-\gamma_k - \beta^{*T} z_i^*)]^2} \left[\frac{\delta_{ki}}{\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]} - \frac{\tau_{ki}}{\tau_k} \right] - \right. \\ &\quad \left. \frac{1}{1 + \exp(-\gamma_k - \beta^{*T} z_i^*)} \left[\frac{\delta_{ki}}{[\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]^2} \frac{\exp(\gamma_k + \beta^{*T} z_{ki}^*)}{1 + \exp(\gamma_k + \beta^{*T} z_{ki}^*)} \right] \right\} \\ &= \sum_{i=1}^n \frac{1}{[1 + \exp(-\gamma_k - \beta^{*T} z_i^*)]^2} \left\{ \exp(-\gamma_k - \beta^{*T} z_i^*) \left[\frac{\delta_{ki}}{\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]} - \frac{\tau_{ki}}{\tau_k} \right] - \right. \\ &\quad \left. \frac{\delta_{ki}}{[\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]^2} \right\}, \quad (2.55) \end{aligned}$$

$$\begin{aligned} V_{\beta\beta} &= \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\tau_k, \gamma_k, \beta) = \sum_{k=0}^{M-1} \sum_{i=1}^n \left\{ \frac{z_{kij} \exp(-\gamma_k - \beta^{*T} z_i^*) z_{kiv}}{[1 + \exp(-\gamma_k - \beta^{*T} z_i^*)]^2} \left[\frac{\delta_{ki}}{\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]} - \frac{\tau_{ki}}{\tau_k} \right] - \right. \\ &\quad \left. \frac{z_{kij}}{1 + \exp(-\gamma_k - \beta^{*T} z_i^*)} \left[\frac{\delta_{ki}}{[\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]^2} \frac{z_{kiv} \exp(\gamma_k + \beta^{*T} z_{ki}^*)}{1 + \exp(\gamma_k + \beta^{*T} z_{ki}^*)} \right] \right\} \end{aligned}$$

$$= \sum_{k=0}^{M-1} \sum_{i=1}^n z_{kij} \frac{1}{[1 + \exp(-\gamma_k - \beta^{*T} z_i^*)]^2} z_{kiv} \left\{ \exp(-\gamma_k - \beta^{*T} z_i^*) \left[\frac{\delta_{ki}}{\log[1 + \exp(\gamma_k + \beta^{*T} z_i^*)]} - \frac{\tau_{ki}}{\tau_k} \right] - \right.$$

$$\left. \frac{\delta_{ki}}{\left[\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right] \right]^2} \right\} \quad (2.56)$$

$$\begin{aligned} V_{\gamma\beta} &= \frac{\partial^2}{\partial \gamma_k \partial \beta_j} \log L(\tau_k, \gamma_k, \beta) = \sum_{i=1}^n \left\{ \frac{z_{kij} \exp(-\gamma_k - \beta^{*T} z_i^*)}{\left[1 + \exp(-\gamma_k - \beta^{*T} z_i^*) \right]^2} \left[\frac{\delta_{ki}}{\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right]} - \frac{\tau_{ki}}{\tau_k} \right] \right. \\ &\quad \left. \frac{1}{1 + \exp(-\gamma_k - \beta^{*T} z_i^*)} \left[\frac{\delta_{ki}}{\left[\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right] \right]^2} \frac{z_{kij} \exp(\gamma_k + \beta^{*T} z_{ki}^*)}{1 + \exp(\gamma_k + \beta^{*T} z_{ki}^*)} \right] \right\} \\ &= \sum_{i=1}^n \frac{z_{kij}}{\left[1 + \exp(-\gamma_k - \beta^{*T} z_i^*) \right]^2} \left\{ \exp(-\gamma_k - \beta^{*T} z_i^*) \left[\frac{\delta_{ki}}{\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right]} - \frac{\tau_{ki}}{\tau_k} \right] \right. \\ &\quad \left. \frac{\delta_{ki}}{\left[\log \left[1 + \exp(\gamma_k + \beta^{*T} z_i^*) \right] \right]^2} \right\}. \quad (2.57) \end{aligned}$$

Na obtenção dos estimadores dos parâmetros desconhecidos, o procedimento usado foi o de Newton -Raphson, já mencionado anteriormente. Como valores iniciais podem-se considerar $\beta_1 = \dots = \beta_p = 0$ e $\gamma_k = 0, k = 1, 2, \dots, M-1$.

Usando (2.54) - (2.56) segue que

$$\hat{V}_{(p+k) \times (p+k)} = \begin{bmatrix} \hat{V}_{\gamma\gamma} & \hat{V}_{\gamma\beta} \\ \hat{V}_{\beta\gamma} & \hat{V}_{\beta\beta} \end{bmatrix}$$

é a matriz assintótica de variância e covariância para (γ, β) .

A função de sobrevivência estimada para $t_k < t_{ki} \leq t_{k+1}$, i.e para $t_{ki} \in I_k$ é

$$\begin{aligned} p_k &= \exp\left(- \int_{t_k}^{t_{ki}} \lambda_k(z) du\right) = \exp\left(- \lambda_k(z) \int_{t_k}^{t_{ki}} du\right) = \exp\left(- \lambda_k(z) [t_{ki} - t_k]\right) \\ &= \exp\left(- \lambda_k(z) \tau_{ki}\right), \end{aligned}$$

substituindo o valor de $\lambda_k(z)$ dado por (2.49) na expressão acima tem-se

$$\begin{aligned}
 p_k &= \exp\left(-\frac{\tau_{ki}}{\tau_k} \log\left[1 + \exp(\hat{\gamma}_k + \hat{\beta}^{*T} z_{ki}^*)\right]\right) = \exp\left(\log\left[1 + \exp(\hat{\gamma}_k + \hat{\beta}^{*T} z_{ki}^*)\right]^{-\frac{\tau_{ki}}{\tau_k}}\right) \\
 &= \left[1 + \exp(\hat{\gamma}_k + \hat{\beta}^{*T} z_{ki}^*)\right]^{-\frac{\tau_{ki}}{\tau_k}}.
 \end{aligned} \tag{2.58}$$

Portanto,

$$\hat{S}_k(t | z) = \begin{cases} \left[1 + \exp(\hat{\gamma}_0 + \hat{\beta}^{*T} z_{0i}^*)\right]^{-\frac{\tau_{0i}}{\tau_0}} & t_{0i} \in I_0 \\ \left[1 + \exp(\hat{\gamma}_0 + \hat{\beta}^{*T} z_{0i}^*)\right]^{-\frac{\tau_{0i}}{\tau_0}} \left[1 + \exp(\hat{\gamma}_1 + \hat{\beta}^{*T} z_{1i}^*)\right]^{-\frac{\tau_{1i}}{\tau_1}} & t_{1i} \in I_1 \\ \vdots & \end{cases} \tag{2.59}$$

Assim, a variância assintótica de $\hat{S}_k(t | z)$ é dado por $[G(\theta)]^T \text{var}(\hat{\beta}) G(\theta)$, onde $G(\theta)$ é um vetor linha sendo que os seus elementos são as derivadas de $S(t | z)$ em relação a γ_k e β_j . Por exemplo, as derivadas de $S(t | z)$ quando os dados estão agrupados em três intervalos são:

$$\begin{aligned}
 \frac{\partial}{\partial \gamma_0} S(t | z) &= -\frac{\tau_{0i}}{\tau_0} \exp(\gamma_0 + \beta^{*T} z_{0i}^*) \left\{1 + \exp(\gamma_0 + \beta^{*T} z_{0i}^*)\right\}^{-\frac{\tau_{0i}}{\tau_0} - 1} * \\
 &\quad \left[1 + \exp(\gamma_1 + \beta^{*T} z_{1i}^*)\right]^{-\frac{\tau_{1i}}{\tau_1}} \left[1 + \exp(\gamma_2 + \beta^{*T} z_{2i}^*)\right]^{-\frac{\tau_{2i}}{\tau_2}}
 \end{aligned} \tag{2.60}$$

$$\begin{aligned}
 \frac{\partial}{\partial \gamma_1} S(t | z) &= -\frac{\tau_{1i}}{\tau_1} \exp(\gamma_1 + \beta^{*T} z_{1i}^*) \left\{1 + \exp(\gamma_1 + \beta^{*T} z_{1i}^*)\right\}^{-\frac{\tau_{1i}}{\tau_1} - 1} * \\
 &\quad \left[1 + \exp(\gamma_0 + \beta^{*T} z_{0i}^*)\right]^{-\frac{\tau_{0i}}{\tau_0}} \left[1 + \exp(\gamma_2 + \beta^{*T} z_{2i}^*)\right]^{-\frac{\tau_{2i}}{\tau_2}}
 \end{aligned} \tag{2.61}$$

$$\begin{aligned}
 \frac{\partial}{\partial \gamma_2} S(t | z) &= -\frac{\tau_{2i}}{\tau_2} \exp(\gamma_2 + \beta^{*T} z_{2i}^*) \left\{1 + \exp(\gamma_2 + \beta^{*T} z_{2i}^*)\right\}^{-\frac{\tau_{2i}}{\tau_2} - 1} * \\
 &\quad \left[1 + \exp(\gamma_0 + \beta^{*T} z_{0i}^*)\right]^{-\frac{\tau_{0i}}{\tau_0}} \left[1 + \exp(\gamma_1 + \beta^{*T} z_{1i}^*)\right]^{-\frac{\tau_{1i}}{\tau_1}}
 \end{aligned} \tag{2.62}$$

$$\begin{aligned}
 \frac{\partial}{\partial \beta_j} S(t | z) = & -z_{ij} \left\{ \frac{r_{0i}}{r_0} \exp(\gamma_0 + \beta^{*T} z_{0i}^*) \left[1 + \exp(\gamma_0 + \beta^{*T} z_{0i}^*) \right]^{-\frac{r_{0i}}{r_0} - 1} \right. \\
 & \left[1 + \exp(\gamma_1 + \beta^{*T} z_{1i}^*) \right]^{-\frac{r_{1i}}{r_1}} \left[1 + \exp(\gamma_2 + \beta^{*T} z_{2i}^*) \right]^{-\frac{r_{2i}}{r_2}} + \\
 & \frac{r_{1i}}{r_1} \exp(\gamma_1 + \beta^{*T} z_{1i}^*) \left[1 + \exp(\gamma_1 + \beta^{*T} z_{1i}^*) \right]^{-\frac{r_{1i}}{r_1} - 1} \\
 & \left. \left[1 + \exp(\gamma_0 + \beta^{*T} z_{0i}^*) \right]^{-\frac{r_{0i}}{r_0}} \left[1 + \exp(\gamma_2 + \beta^{*T} z_{2i}^*) \right]^{-\frac{r_{2i}}{r_2}} + \right. \\
 & \left. \frac{r_{2i}}{r_2} \exp(\gamma_2 + \beta^{*T} z_{2i}^*) \left[1 + \exp(\gamma_2 + \beta^{*T} z_{2i}^*) \right]^{-\frac{r_{2i}}{r_2} - 1} \right. \\
 & \left. \left[1 + \exp(\gamma_0 + \beta^{*T} z_{0i}^*) \right]^{-\frac{r_{0i}}{r_0}} \left[1 + \exp(\gamma_1 + \beta^{*T} z_{1i}^*) \right]^{-\frac{r_{1i}}{r_1}} \right\}. \quad (2.63)
 \end{aligned}$$

Usando (2.60) - (2.63) segue que

$$\begin{bmatrix} \frac{\partial}{\partial \gamma} S(t | z) \\ \frac{\partial}{\partial \beta} S(t | z) \end{bmatrix},$$

é o vetor das derivadas de $S(t | z)$. Utilizando as informações acima contruiu-se intervalos de $100(1 - \alpha)\%$ de confiança para $\hat{S}(t | z)$.

O programa PAMLEPAS.PRG (Apêndice D3.9), foi desenvolvido com o intuito de ajustar o modelo logístico exponencial por partes (2.46), tal como estimar a função de sobrevivência $S(t | z)$ através de (2.59) e a variância da função de sobrevivência dada em (1.30).

2.6. Extensão do Modelo Logístico Exponencial.

Mantel e Hankey [1978] assumiram que o logaritmo da probabilidade condicional de

falha é uma função das variáveis regressoras e do tempo, e a equação (2.17) pode ser modificada para

$$\log \frac{p_{ki}}{q_{ki}} = \underline{\beta}^T z_i + h(k) \quad (2.63)$$

$i = 1, 2, \dots, n$ e $k = 0, 1, \dots, T_i$, onde $q_{ki} = 1 - p_{ki}$ é a probabilidade que o i -ésimo indivíduo falhe no intervalo de tempo k condicionado a não ter respondido em nenhum intervalo de tempo anterior. Os valores de q_{ki} dependem das variáveis regressoras z_i e do intervalo de tempo k . Quando $h(k)$ é conhecido ou especificado, este modelo pode ser empregado a fim de analisar os dados, tal como avaliar o papel das variáveis regressoras, sem necessariamente assumir função de risco constante. Na ausência de uma forma conhecida ou específica para $h(k)$ pode ser aproximado de uma maneira grosseira por um polinômio de grau s em k

$$h(k) \simeq \sum_{l=1}^s \theta_l k^l \quad (2.64)$$

onde θ_0 é omitido, pois será embutido em β_0 na equação (2.63).

Então, substituindo $h(k)$ em (2.63) obtém-se

$$\log \frac{p_{ki}}{q_{ki}} = \underline{\beta}^T z_i + \sum_{l=1}^s \theta_l k^l \quad (2.65)$$

portanto

$$p_{ki} = \frac{\exp \left[\underline{\beta}^T z_i + \sum_{l=1}^s \theta_l k^l \right]}{1 + \exp \left[\underline{\beta}^T z_i + \sum_{l=1}^s \theta_l k^l \right]} \quad (2.66)$$

e

$$q_{ki} = \frac{1}{1 + \exp \left[\underline{\beta}^T z_i + \sum_{l=1}^s \theta_l k^l \right]} \quad (2.67)$$

Seja t_i o tempo de resposta para o i -ésimo indivíduo. Este tempo pode ser contínuo ou discreto.

1. **Tempo contínuo** quando o tempo exato de resposta é observado, neste caso o tempo entra no modelo como uma covariável e a extensão do modelo logístico fica semelhante ao modelo logístico da secção 2.3, onde (2.64) é da forma

$$h(t) \simeq \sum_{l=1}^s \theta_l t^l,$$

sendo que o s é o grau do polinômio a ser considerado no modelo.

2. **Tempo discreto** quando o tempo está representado por intervalos, T_i representa o intervalo de tempo onde as falhas ocorrem ($\delta_i = 1$) ou T_i representa o número de intervalo no fim do qual não houve falha ($\delta_i = 0$). Este modelo é discutido por Mantel e Hankey [1978] e serão apresentados alguns de seus resultados.

A função de verossimilhança é dado por

$$L = \prod_{i=1}^n \left[\frac{1 - q_{iT_i}}{q_{iT_i}} \right]^{\delta_i} \prod_{k=1}^{T_i} q_{ki} \quad (2.68)$$

e

$$\begin{aligned} \log L &= \sum_{i=1}^n \left[\delta_i \log \left[\frac{p_{iT_i}}{q_{iT_i}} \right] + \sum_{k=1}^{T_i} \log q_{ki} \right] \\ &= \sum_{i=1}^n \left[\delta_i \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right) - \sum_{k=1}^{T_i} \log \left[1 + \exp \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right) \right] \right] \end{aligned} \quad (2.69)$$

que é linear nos parâmetros.

A primeira e segunda derivadas parciais do logaritmo da verossimilhança são

$$\frac{\partial}{\partial \beta_j} \log L = \sum_{i=1}^n \left[\delta_i z_{ij} - \sum_{k=1}^{T_i} z_{ij} \frac{\exp \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right)}{1 + \exp \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right)} \right] = \sum_{i=1}^n z_{ij} \left[\delta_i - \sum_{k=1}^{T_i} p_{ki} \right] \quad (2.70)$$

e

$$\frac{\partial}{\partial \theta_1} \log L = \sum_{i=1}^n \left[\delta_i T_i - \sum_{k=1}^{T_i} k^1 \frac{\exp \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right)}{1 + \exp \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right)} \right] = \sum_{i=1}^n \left[\delta_i T_i - \sum_{k=1}^{T_i} k^1 p_{ki} \right] \quad (2.71)$$

e as derivadas segundas são

$$\frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L = - \sum_{i=1}^n \left[\sum_{k=1}^{T_i} z_{ij} z_{iv} \frac{\exp \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right)}{\left[1 + \exp \left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l \right) \right]^2} \right] = - \sum_{i=1}^n \sum_{k=1}^{T_i} z_{ij} z_{iv} p_{ki} q_{ki} \quad (2.72)$$

e

$$\frac{\partial^2}{\partial \beta_j \partial \theta_1} \log L = - \sum_{i=1}^n \left[\sum_{k=1}^{T_i} z_{ij} k^l \frac{\exp(\beta^T z_i + \sum_{l=1}^s \theta_l k^l)}{\left[1 + \exp(\beta^T z_i + \sum_{l=1}^s \theta_l k^l) \right]^2} \right] = - \sum_{i=1}^n \sum_{k=1}^{T_i} z_{ij} k^l p_{ki} q_{ki} \tag{2.73}$$

$$\frac{\partial^2}{\partial \theta_1 \partial \theta_u} \log L = - \sum_{i=1}^n \sum_{k=1}^{T_i} k^l k^u \frac{\exp(\beta^T z_i + \sum_{l=1}^s \theta_l k^l)}{\left[1 + \exp(\beta^T z_i + \sum_{l=1}^s \theta_l k^l) \right]^2} = - \sum_{i=1}^n \sum_{k=1}^{T_i} k^l k^u p_{ki} q_{ki} \tag{2.74}$$

Para estimar os parâmetros, considera-se que a função risco é constante, θ_j são todos iguais a zero, e entra-se com valores iniciais para os β_j . Um possível conjunto de valores iniciais para os β_j são $\beta_0 = \beta_1 = \dots = \beta_p = 0$. Estima-se os parâmetros pelo método de Newton-Raphson da mesma forma que foi estimado na secção 2.2. Em seguida usa-se estes parâmetros estimados como valores iniciais dos β_j e para os $\theta_l = 0$, para estimar os parâmetros quando todos estão incluídos.

A função de sobrevivência estimada é

$$\hat{S}(t | z) = \prod_{k=1}^{T_i} \frac{1}{1 + \exp(\hat{\beta}^T z_i + \sum_{l=1}^s \hat{\theta}_l k^l)} \tag{2.75}$$

Portanto a variância assintótica de $S(t | z)$ é dado por $[G(\theta)]^T \text{var}(\hat{\beta}) G(\theta)$, onde $G(\theta)$ é um vetor linha sendo que os seus elementos são as derivadas de $S(t | z)$ em relação a β_j e a θ_j .

A derivada acima fica complicada e extensa quando T_i é muito grande. Uma solução é considerar $g(\theta) = \log S(t | z)$, assim

$$G(\theta) = - \sum_{k=1}^{T_i} \log \left[1 + \exp(\beta^T z_i + \sum_{l=1}^s \theta_l k^l) \right], \text{ e}$$

$$\frac{\partial}{\partial \beta_j} G(\theta) = - \sum_{k=1}^{T_i} z_{ij} \frac{\exp(\beta^T z_i + \sum_{l=1}^s \theta_l k^l)}{\left[1 + \exp(\beta^T z_i + \sum_{l=1}^s \theta_l k^l) \right]}, \tag{2.76}$$

$$\frac{\partial}{\partial \theta_1} G(\theta) = - \sum_{k=1}^{T_i} k^l \frac{\exp\left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l\right)}{\left[1 + \exp\left(\beta^T z_i + \sum_{l=1}^s \theta_l k^l\right)\right]}. \quad (2.77)$$

De (2.79) e (2.80) tem-se que

$$\begin{bmatrix} \frac{\partial}{\partial \beta} G(\theta) \\ \frac{\partial}{\partial \theta} G(\theta) \end{bmatrix},$$

é a vetor das derivadas de logaritmo da função de sobrevivência. Utilizando as informações acima, contruiu-se intervalos de $100(1 - \alpha)\%$ de confiança para logaritmo de $\hat{S}(t | z)$.

2.7. Aplicação.

Nesta secção, estuda-se novamente os dados do exemplo 1 introduzido no Capítulo 1 , sobre sobrevivência de pacientes com myeloma múltiplo, com o intuito de ilustrar os métodos expostos acima.

Algumas suposições a respeito do planejamento se fez necessário, antes de ajustar qualquer modelo. Supõe-se que todos pacientes entraram no estudo no mesmo instante e foram seguidos por um tempo τ ou entrando em tempos diferentes porém seguidos por um mesmo tempo τ pré-fixado.

Foi considerado o tempo $\tau = 20$ meses como o tempo fixado para observação, portanto, os dados censurados antes deste período foram descartados, assim como os pacientes mortos após este período foram considerados no modelo como vivos.

As suposições acima foram consideradas para todos os modelos aqui analisados.

Para o modelo logístico e sua extensão, o procedimento LOGISTIC do SAS estima e testa os parâmetros, porém para calcular a estimativa da função de sobrevivência foi desenvolvido um programa através do procedimento IML do SAS. PESVSLAS.PRG (Apêndice D3.7).

Para os modelos logístico exponencial e exponencial por partes foram desenvolvidos os programas de computador PAMLEAS.PRG (Apêndice D3.8) e PAMLEPAS.PRG (Apêndice D3.9), respectivamente, para estimar e testar os parâmetros do modelo, bem como para calcular as estimativas da função de sobrevivência e os respectivos intervalos de confiança.

A. Ajuste dos Modelos.

A.1. Modelo Logístico.

No caso desta aplicação, há interesse em determinar quais fatores prognósticos importantes no status de sobrevivência de pacientes com mieloma múltiplo. No ajuste do modelo foram considerado 53 pacientes dos 65, pois 12 destes foram perdidos de observação, i.e., censurados antes de 20 meses. Define-se a variável resposta por

$$\delta_i = \begin{cases} 1 & \text{se indivíduo (i) morreu antes dos 20 meses} \\ 0 & \text{se indivíduo (i) sobreviveu aos 20 meses} \end{cases}$$

Este modelo, pode ser ajustado usando o procedimento LOGISTIC do *software* estatístico SAS, que tem como opção métodos de seleção de covariáveis. Neste caso empregou-se o método STEPWISE. (ver Hosmer and Lemeshow [1984] - Cap. 3). Três variáveis, LOGBUN, PBJU e TOTSORO estavam relacionadas com o status do paciente após o período de observação. Portanto, o modelo ajustado para cada paciente é dado por:

$$\log \frac{1 - S(\tau | \mathbf{z})}{S(\tau | \mathbf{z})} = \beta_0 + \beta_1 \text{LOGBUN}_i + \beta_2 \text{PBJU}_i + \beta_3 \text{TOTSORO}_i.$$

Os resultados estão sumarizados na tabela abaixo.

Tabela 2.2 - Estimação dos coeficientes de regressão para o modelo logístico.

VARIÁVEIS	$\hat{\beta}$	Erro-padrão	p-value
INTERCEPTO	10.6866	3.4457	0.0019
LOGBUN	-3.4352	1.3716	0.0123
PBJU	-2.0750	0.7645	0.0066
TOTSORO	-0.3746	0.1800	0.0374

Para testar, a hipótese $\beta_1 = \beta_2 = \beta_3 = 0$, usou-se o teste da razão de verossimilhança (1.35). O valor do teste estatístico encontrado foi 17.532, que é uma χ^2 com 3 graus de liberdade (p-value = 0.0005). Portanto há evidências de que as covariáveis LOGBUN, PBJU e TOTSORO estão relacionados com o status do paciente com mieloma múltiplo.

A2. Modelo Logístico Exponencial.

Como foi verificado no Capítulo 1, os tempos de sobrevivência relacionados aos pacientes com mieloma múltiplo possuem distribuição exponencial; uma condição necessária para o ajuste do modelo logístico exponencial. A variável resposta é a mesma usada no modelo anterior. Seja

$$t_i = \begin{cases} t_i & \text{se o indivíduo (i) morreu antes dos 20 meses} \\ \tau & \text{se o indivíduo (i) sobreviveu aos 20 meses} \end{cases}$$

Este modelo foi ajustado usando o programa PAMLEAS.PRG (Apêndice D3.8). Procedeu-se da seguinte forma:

- (1) ajustou-se o modelo onde somente o intercepto foi considerado;
- (2) ajustou-se para cada covariável um modelo logístico exponencial simples, i.e., modelo incluindo o intercepto e a respectiva covariável e comparou-se este último com o modelo inicial (1) através do teste da razão de verossimilhança dado em (1.35). O nível descritivo (p-value) do teste determina a inclusão da respectiva covariável (p-value < 25%);
- (3) ajustou-se o modelo considerando todas as covariáveis selecionadas em (2) e testou-se a nulidade de cada um de seus coeficientes, através da estatística de Wald expressada em (1.33). A exclusão de cada covariável é determinada pelo nível descritivo do respectivo teste de Wald (p-value > 5%);
- (4) ajustou-se o modelo sem as covariáveis excluídas em (3) e comparou-se este com o modelo (3) através da estatística da razão de verossimilhança. Assim as covariáveis LOGBUN e PBJU mostraram evidências de risco para os pacientes observados nos 20 primeiros meses com mieloma múltiplo. Portanto, o modelo ajustado para cada paciente é dado por

$$\log \frac{1 - \exp(\lambda(z_i) \tau)}{\exp(\lambda(z_i) \tau)} = \beta_0 + \beta_1 \text{LOGBUN}_i + \beta_2 \text{PBJU}_i$$

Os resultados estão sumarizados na tabela abaixo.

Tabela 2.3 - Estimação dos coeficientes de regressão para o modelo logístico exponencial.

VARIÁVEIS	$\hat{\beta}$	Erro-padrão	p-value
INTERCEPTO	- 5.9964	1.8947	0.0015
LOGBUN	3.1087	1.1013	0.0048
PBJU	1.2726	0.6058	0.0357

Para testar a hipótese $\beta_1 = \beta_2 = 0$ usou-se o teste da razão de verossimilhança, que resultou em $\chi^2 = 33.9162$ com 2 gl (p-value = 0.0000). Logo rejeitou-se a hipótese dos β serem iguais a zero, i.e, as covariáveis LOGBUN e PBJU estão relacionados com status do paciente com mieloma múltiplo.

A3. Modelo Logístico Exponencial por Partes.

Neste modelo, o objetivo é verificar se a sobrevivência dos pacientes com mieloma múltiplo é diferente ao longo do tempo, isto é, paciente que sobrevive aos primeiros meses reagem melhor ao mieloma múltiplo, e se os fatores prognósticos estão sujeitos a mudança no tempo.

Supondo que o tempo de seguimento ($\tau = 20$), seja dividido em dois pontos fixos, ou seja $[0, 10)$ e $[10, 20)$. O status do paciente é observado ao término de cada intervalo, logo define-se

$$\delta_{ki} = \begin{cases} 1 & \text{se o indivíduo (i) morreu antes de } \tau_k \\ 0 & \text{se o indivíduo (i) sobreviveu a } \tau_k \end{cases}$$

e o tempo de sobrevivência relacionado ao i-ésimo paciente no intervalo (k) é

$$t_{ki} = \begin{cases} t_{ki} = t_i & \text{se } t_i \in \tau_k \\ t_{ki} = \tau_k & \text{se } t_i > \tau_k \end{cases}$$

Neste estudo as covariáveis foram aferidas uma única vez para cada paciente.

Na busca do "melhor" modelo procedeu-se da mesma forma descrita no modelo anterior. Para ajustar e testar o modelo usou-se o programa PAMLEPAS.PRG desenvolvido no módulo IML do SAS. O modelo logístico exponencial para cada paciente sob o período τ_k é da forma

$$\log \frac{1 - \exp(\lambda_k(z_i)\tau_k)}{\exp(\lambda_k(z_i)\tau_k)} = \gamma_k + \beta_1 \text{LOGBUN}_{ik} + \beta_2 \text{INFEC}_{ik} + \beta_3 \text{PBJU}_{ik} + \beta_4 \text{TOTSORO}_{ik}$$

Os resultados estão sumarizados na Tabela 2.4.

Tabela 2.4 - Estimação dos coeficientes de regressão para o modelo logístico exponencial por partes.

VARIÁVEIS	$\hat{\beta}$	Erro-padrão	p-value
INTERCEPTO1	- 7.1661	0.7556	0.0000
INTERCEPTO2	- 6.2664	0.7479	0.0000
LOGBUN	2.3321	0.4344	0.0000
INFEC	0.8112	0.3928	0.0389
PBJU	0.8042	0.1510	0.0000
TOTSORO	0.1770	0.0392	0.0000

Nota-se que quando se divide o tempo de seguimento em intervalos, os resultados se alteram se comparado com os modelos anteriores; ou seja, neste modelo além dos fatores de risco já detectados nos modelos anteriores, aparece também como fator de risco para os pacientes com myeloma múltiplo a covariável infecção (INFEC).

A4. Extensão do Modelo Logístico.

Não foi possível ajustar o modelo logístico, extensão de Mantel e Hankey [1978], para os dados referentes ao estudo de pacientes com myeloma múltiplo, porque ao incluir o polinômio do tempo de grau 1 ou grau 2 como covariáveis, o método de Newton-Raphson empregado para estimar os parâmetros β não convergiu. (número de observações insuficiente para o modelo)

B. Intervalos de Confiança para a Função Sobrevida nos Modelos A1, A2 e A3.

Com o intuito de comparar os modelos logísticos estimou-se a função de sobrevivência $\hat{S}(t | z)$ para cada paciente e construiu-se intervalos de confiança de 95% de confiança. Os resultados encontram-se na Tabela C2 (Apêndice C). Os cálculos foram feitos utilizando os programas PESVSLAS.PRG no modelo logístico linear, PAMLEAS.PRG no modelo logístico exponencial e PAMLEPAS.PRG no modelo logístico exponencial por partes.

D. Gráfico.

Para ilustrar os resultados, procedeu-se da mesma forma do capítulo anterior. Construiu-

se gráficos das estimativas da função de sobrevivência e seus respectivos erros padrão, considerando-se os três modelos ajustados acima, para os quatro primeiros pacientes, com o intuito de se ter uma idéia das possíveis vantagens de um sobre os outros.

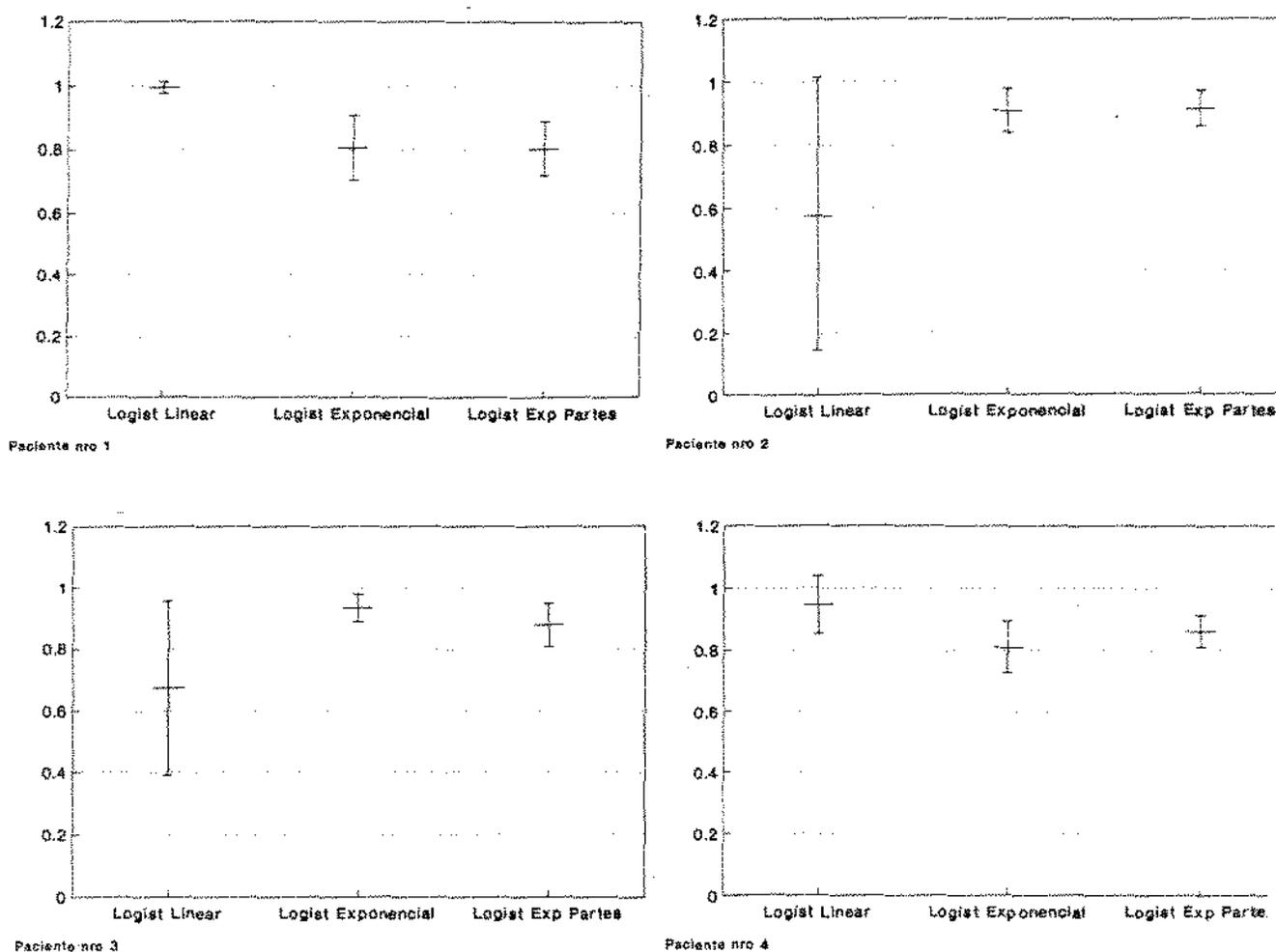


Figura 2.1. - Intervalos de confiança para estimativa da função de sobrevivência no modelo Logístico Linear, Logístico Exponencial e Logístico Exponencial por Partes.

D. Conclusão.

Os gráficos anteriores, mostram que em média a função de sobrevivência estimada através do modelo A1 se distancia dos valores estimados nos outros dois modelos quando o tempo é considerado, e seus erros padrão apresentaram estimativas superiores aos demais, com exceção para alguns paciente.

Considerando os modelos A2 e A3, as estimativas da função de sobrevivência individual são próximas nos dois modelos, e os intervalos de confiança aparentam ser consistentes em ambos os modelos. Portanto, uma vez que se conhece o tempo de falha individual destes pacientes e este por sua vez possui distribuição exponencial, ambos os modelos parecem, explicar bem os riscos relacionados a estes pacientes.

CAPÍTULO 3

MODELO DE REGRESSÃO EXPONENCIAL POR PARTES

3.1 - Introdução

Em situações onde o número de indivíduos observados é muito grande, tal que os tempos de sobrevivência podem ser agrupados em intervalos e, as variáveis concomitantes aferidas para cada indivíduo são categóricas, o método de tabela de vida, usualmente empregado para estimar a distribuição de sobrevivência, pode ser de grande utilidade. Porém, para analisar o efeito das variáveis concomitantes na sobrevivência dos indivíduos é necessário construir para cada combinação dos níveis das covariáveis uma tabela de vida distinta, embora a análise possa ser realizada globalmente incluindo essas distintas tabelas. Para melhor compreensão considere o exemplo abaixo.

Exemplo 3.1. Laird e Olivier[1981], apresentaram um conjunto de dados, onde estudaram a sobrevivência de 1975 pacientes submetidos a transplante de rim. Os fatores de interesse são:

1. Tipo de Doador, com 2 categorias: CAD (doador morto) e LRD (doador vivo);
2. Grau de Compatibilidade, com 5 categorias crescentes: 0, 1, 2, 3, 4;

O período de seguimento foi dividido em 16 intervalos de tempos. A partir de tais informações,

construiu-se 10 tabelas de vidas referentes a combinações das categorias das 2 variáveis de interesse ($2 \times 5 = 10$), como pode ser visualizado na Tabela 3.1 abaixo.

Tabela 3.1 - Tabelas de vida para pacientes submetidos a transplante de rim.

Tempo	CAD	Grau de Compatibilidade														
		0			1			2			3			4		
		E ^a	W ^b	D ^c	E	W	D	E	W	D	E	W	D	E	W	D
0 - 7d	242	0	13	386	0	16	393	0	20	130	0	8	18	0	0	
7 - 15d	229	0	18	370	0	25	373	0	20	122	0	4	18	0	1	
15 - 21d	211	0	14	345	0	12	353	0	18	118	0	8	17	0	0	
21 - 30d	197	0	13	333	0	17	335	0	18	110	0	10	17	0	0	
1 - 2m	184	0	21	316	0	43	317	0	36	100	0	14	17	0	1	
2 - 3m	163	0	14	273	0	30	281	0	26	86	0	2	16	0	3	
3 - 6m	149	0	23	243	0	35	255	0	36	84	0	10	13	0	2	
6 - 9m	126	0	8	208	0	11	219	0	6	74	0	5	11	0	0	
9 - 12m	118	0	3	197	0	16	213	0	10	69	0	1	11	0	0	
1 - 1.5a	115	0	4	181	0	9	203	0	11	68	0	3	11	0	0	
1.5 - 2a	111	2	9	172	6	6	192	8	7	65	2	4	11	0	1	
2 - 2.5a	100	17	2	160	34	8	177	43	8	59	14	0	10	5	1	
2.5 - 3a	81	21	3	118	26	4	126	31	6	45	11	0	4	1	1	
3 - 3.5a	57	19	2	88	27	3	89	22	5	34	11	1	2	0	0	
3.5 - 4a	36	19	1	58	33	1	62	27	1	22	11	0	2	0	0	
> 4a	16	16	0	24	23	1	34	34	0	11	9	2	2	2	0	

Tempo	LRD	Grau de Compatibilidade														
		0			1			2			3			4		
		E	W	D	E	W	D	E	W	D	E	W	D	E	W	D
0 - 7d	20	0	0	48	0	2	410	0	14	164	0	4	164	0	1	
7 - 15d	20	0	1	46	0	0	396	0	7	160	0	3	163	0	2	
15 - 21d	19	0	1	46	0	1	386	0	3	157	0	5	161	0	1	
21 - 30d	18	0	0	45	0	2	386	0	4	152	0	7	160	0	1	
1 - 2m	18	0	3	43	0	5	382	0	31	145	0	8	159	0	5	
2 - 3m	15	0	1	38	0	3	351	0	21	137	0	1	154	0	1	
3 - 6m	14	0	2	35	0	2	330	0	27	136	0	8	153	0	4	
6 - 9m	12	0	0	33	0	0	303	0	14	128	0	3	149	0	1	
9 - 12m	12	0	0	33	0	1	289	0	11	125	0	1	148	0	2	
1 - 1.5a	12	0	0	32	0	2	278	0	9	124	0	0	146	0	0	
1.5 - 2a	12	0	0	30	0	0	269	8	10	124	2	1	146	4	2	
2 - 2.5a	12	2	0	30	4	0	251	53	7	121	17	2	140	23	1	
2.5 - 3a	10	0	0	26	8	1	191	45	3	102	17	1	116	33	1	
3 - 3.5a	10	5	0	17	3	0	143	45	2	84	26	1	82	27	1	
3.5 - 4a	5	1	0	14	7	0	96	39	0	57	24	0	54	23	1	
> 4a	4	4	0	7	7	0	57	57	0	33	33	0	30	30	0	

E^a = entrada

W^b = censura

D^c = morte

Holford[1980] e Laird et al [1981], estudaram o uso do modelo log-linear, empregado em análise de dados categóricos, para caracterizar dados de sobrevivência dispostos como no exemplo acima. Holford havia demonstrado que os estimadores de máxima verossimilhança do modelo de risco log-linear, quando as covariáveis são categóricas e os modelos estruturais são Poisson, exponencial por partes e multinomial, são equivalentes. Laird et al demonstraram a equivalência do modelo log-linear para médias das caselas da tabela de contingência com estrutura de Poisson, com o modelo de risco log-linear quando a distribuição dos tempos de sobrevivência é exponencial por partes e as covariáveis são categóricas.

O objetivo deste capítulo é estudar o uso do modelo exponencial por partes em situações de análise de sobrevivência, como o segundo modelo alternativo aos usuais modelos de regressão paramétrico e de Cox e comparar este modelo com o logístico exponencial por partes. A correspondência entre os modelos log-lineares com estrutura de Poisson, exponencial por partes e multinomial nas abordagens de Holford e Laird et al são revistas com o objetivo de fornecer fundamentos para o uso de pacotes computacionais.

Na seção 3.2 é apresentado o modelo exponencial por partes para dados de sobrevivência provenientes de tabelas de vida. A relação entre a regressão logística exponencial por partes e exponencial por partes é dada na seção 3.3. A seção 3.4 trata da correspondência entre os modelos de Poisson, exponencial por partes e multinomial. E finalmente na seção 3.5 a reprodução dos resultados utilizando os dados do exemplo 3.1 apresentados por Laird et al [1981] e uma aplicação aos dados do exemplo 1 do Capítulo 1.

3.2. Modelo Exponencial por Partes.

O modelo exponencial por partes pode ser utilizado quando o intervalo de tempo é particionado em k intervalos de tempo mutuamente exclusivos I_k , $k = 1, \dots, K$, tal que a função risco é constante em cada intervalo. As presunções básicas neste modelo são:

1. censuras ocorrem uniformemente;
2. as probabilidades de falha são pequenas dentro de cada intervalo.

Quando as covariáveis são consideradas, os modelos exponenciais por partes irão depender do tipo de covariáveis consideradas. Entre outros, Holford [1980] considerou no seu modelo covariáveis do tipo contínuo e/ou discretas e Laird et al (1981) utilizou o modelo exponencial por partes quando as covariáveis eram categóricas.

Basicamente, quando as covariáveis são contínuas e/ou discretas, a função risco é apresentada por

$$\lambda(t, \mathbf{z}) = \lambda_k \exp(\underline{\beta}^{*T} \mathbf{z}^*), \quad t \in I_k, k=1,2, \dots, K, \quad (3.1)$$

onde:

$\mathbf{z}^* = (z_1, z_2, \dots, z_p)$ vetor de covariáveis,

$\underline{\beta}^*$: é o vetor coluna dos parâmetros desconhecidos que especifica os efeitos das covariáveis.

I_k : representa o intervalo.

λ_k : é a função risco da exponencial com função densidade de probabilidade $f(t) = \lambda \exp(-\lambda t)$.

No modelo com variáveis categóricas, denota-se por λ_k a função risco constante em cada intervalo I_k e a função risco do modelo será:

$$\lambda_k(t, \mathbf{x}) = \lambda_k \exp(\underline{\beta}^{*T} \mathbf{x}^*) \quad \lambda \in I_k, \text{ com } k=1,2, \dots, K; \quad (3.2)$$

onde:

\mathbf{x}^* : é a matriz de planejamento que especifica os níveis das p covariáveis categóricas; isto é,

$$\mathbf{x}^* = (x_1, x_2, \dots, x_p).$$

De (3.2), obtém-se que

$$\theta_k = \lambda_k(t, \mathbf{x}) = \lambda_k \exp(\underline{\beta}^{*T} \mathbf{x}^*) = \exp(\log \lambda_k + \underline{\beta}^{*T} \mathbf{x}^*), \quad (3.3)$$

com $\underline{\beta}^* = (\underline{\beta}_1, \dots, \underline{\beta}_p)$, $\underline{\beta}_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{il_i})$, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, l_i$, onde j é o número de níveis da covariável i .

Considera-se o exemplo 3.1 para as covariáveis tipo de doador, com 2 níveis ($l_1 = 2$) e grau de compatibilidade, com 5 níveis ($l_2 = 5$).

Como $k = 1, \dots, K$, o modelo completo incorporando o efeito do tempo é o seguinte:

$$\log \theta = \underline{\beta}^T \mathbf{x} \quad (3.4)$$

com $\underline{\beta}^T = (\underline{\beta}_1, \dots, \underline{\beta}_p, \underline{\beta}_T)$ e $\mathbf{x} = (x_1, \dots, x_p, x_T)$.

Laird et al empregaram a notação usual de modelo log-linear em tabelas de contingência para caracterizar o modelo de risco exponencial por partes (3.4). A equação acima pode ser reescrita como

$$\log \theta = \mu + \mu_0(I_k) + \sum_{j=1}^{l_i} \mu_{i(j)} + \sum_{j,k}^{l_i, K} \mu_{0i(I_k, j)} + \sum_{j,j'}^{l_i, l_{i'}} \mu_{ii'(j, j')} + \dots \quad (3.5)$$

onde:

μ : média geral

$\mu_{i(j)}$: efeito principal do j -ésimo nível da covariável i na sobrevivência,

$\mu_{ii'(j, j')}$: efeito de interação nos níveis j e j' das covariáveis i e i' na sobrevivência.

O modelo da matriz de planejamento é o de diferença das médias, de tal modo que

$$\sum_{j=1}^{l_i} \mu_i(j) = \sum_{j=1}^{l_i} \sum_{j'=1}^{l_{i'}} \mu_{ij'}(j, j') = \dots = 0.$$

Observa-se que a função risco do modelo (3.3) considera na matriz de planejamento x^* unicamente o efeito das covariáveis e o efeito do tempo é dado por λ_k . Incorporando este último efeito na matriz x ,

$$\log \lambda_k = \beta_0 + \beta_T$$

sendo β_0 a média geral. Portanto,

$$\lambda_k = \exp(\beta_0 + \beta_T).$$

Observa-se que a matriz x reparametrizada como explicado acima, tem $(1+(l_1-1)+\dots+(l_p-1)+(k-1))$ colunas de efeitos principais e $(l_1 \times l_2 \times \dots \times l_p \times k)$ linhas.

Para ilustração, suponha um estudo onde foram aferidas para cada indivíduo 2 covariáveis categóricas, $i = 1, 2$, a covariável 1 com 2 níveis de resposta e a covariável 2 com 3 níveis, i.e. $l_1 = 2$ e $l_2 = 3$. Suponha ainda que estes indivíduos foram seguidos por um período de tempo dividido em três intervalos mutuamente exclusivo, isto é l_k , $k = 1, 2, 3$. Assim o modelo exponencial por partes com efeito principal é dado por

$$\theta_k = \lambda_k(t, x) = \lambda_k \exp(\beta^{*T} x^*), \quad k = 1, 2, 3.$$

Portanto,

$$x^* = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \quad \text{e } \beta^* = \begin{bmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{22} \end{bmatrix}.$$

Logo,

$$\log \theta = x^T \beta = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 0 \\ 1 & 1 & -1 & -1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 1 \\ 1 & -1 & 0 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \beta_{21} \\ \beta_{22} \\ \beta_{31} \\ \beta_{32} \end{bmatrix}$$

onde:

β_0 = média geral;

β_{11} = efeito principal do nível 1 da covariável 1 na sobrevivência;

β_{21} = efeito principal do nível 1 da covariável 2 na sobrevivência;

β_{22} = efeito principal do nível 2 da covariável 2 na sobrevivência;

β_{31} = efeito principal do intervalo de tempo 1 na sobrevivência;

β_{32} = efeito principal do intervalo de tempo 2 na sobrevivência;

sob as restrições do modelo:

$\beta_{12} = -\beta_{11}$ é o efeito principal do nível 2 da covariável 1 na sobrevivência;

$\beta_{23} = -(\beta_{21} + \beta_{22})$ é o efeito principal do nível 3 da covariável 2 na sobrevivência;

$\beta_{33} = -(\beta_{31} + \beta_{32})$ é o efeito principal do intervalo de tempo 3 na sobrevivência; portanto,

$$\lambda_1 = \exp(\beta_0 + \beta_{31}), \lambda_2 = \exp(\beta_0 + \beta_{32}) \text{ e } \lambda_3 = \exp(\beta_0 + \beta_{33}).$$

Supondo que n indivíduos são observados e seja θ_k a função de risco constante no intervalo I_k e δ_{ki} a variável dicotômica tal que

$$\delta_{ki} = \begin{cases} 1 & \text{se o } i\text{-ésimo indivíduo falha no intervalo } I_k \\ 0 & \text{se o } i\text{-ésimo indivíduo não falha no intervalo } I_k \end{cases}, \text{ para } k = 1, \dots, K.$$

A função de verossimilhança para os indivíduos no intervalo I_k é dada por

$$L_k(\theta_k) = \prod_{i=1}^n [\theta_k]^{\delta_{ki}} \exp(-\theta_k t_{ki}), \quad (3.7)$$

e a verossimilhança total é

$$L(\theta) = \prod_{k=1}^K L_k(\theta_k). \quad (3.8)$$

Portanto, a função a ser maximizada é

$$\log L(\theta) = \sum_{k=1}^K \sum_{i=1}^n [\delta_{ki} \log \theta_k - \theta_k t_{ki}] = \sum_{k=1}^K \sum_{i=1}^n [d_k \log \theta_k - \theta_k T_k] \quad (3.9)$$

onde:

$d_k = \sum_{i=1}^n \delta_{ki}$ - é o número de falhas no intervalo I_k ;

$T_k = \sum_{i=1}^n t_{ki}$ - é o tempo total de seguimento dos indivíduos exposto ao risco no intervalo I_k .

Quando os dados estão representados em tabelas de vida, como mostra a Tabela 3.1, T_k pode ser calculado diretamente, i.e.,

$$T_k = [E_k - (D_k + W_k)0.5] \tau_k, \tag{3.10}$$

onde E_k , D_k e W_k são dados da Tabela 3.1 e τ_k é o comprimento do intervalo I_k .

Substituindo-se o valor de θ_k dado em (3.3) na equação (3.9), tem-se que

$$\log L(\underline{\beta}) = \sum_{k=1}^K [d_k \underline{\beta}^T x_k - T_k \exp(\underline{\beta}^T x_k)],$$

diferenciando a equação acima com respeito a cada um dos parâmetros, obtém-se

$$U(\underline{\beta}) = \frac{\partial}{\partial \beta_j} \log L(\underline{\beta}) = \sum_{k=1}^K [d_k x_{kj} - T_k x_{kj} \exp(\underline{\beta}^T x_k)] = \sum_{k=1}^K x_{kj} [d_k - T_k \exp(\underline{\beta}^T x_k)], \tag{3.11}$$

e os elementos (jv) da matriz das derivadas segunda é dado por

$$H_{jv}(\underline{\beta}) = \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\underline{\beta}) = - \sum_{k=1}^K x_{kj} T_k \exp(\underline{\beta}^T x_k) x_{kv}. \tag{3.12}$$

O estimador de máxima verossimilhança $\hat{\underline{\beta}}$ para os parâmetros do modelo é obtido igualando (3.11) a zero, obtendo um sistema de equações que usualmente não são lineares. Assim são necessários procedimentos iterativos para encontrar $\hat{\underline{\beta}}$. O procedimento empregado foi o Newton-Raphson, subsecção 1.5.1, com valores iniciais de $\underline{\beta} = 0$.

Neste caso é de interesse estimar a probabilidade de um indivíduo sobreviver ao intervalo $I_{j+1} = [t_j, t_{j+1}]$ dado que sobreviveu a t_j , $j = 0, 1, \dots, k-1$. Assim, para cada combinação dos níveis das covariáveis, a taxa de risco é constante em cada intervalo. Logo

$$p_{j+1} = \exp \left[- \int_{t_j}^{t_{j+1}} \theta_{j+1} du \right] = \exp [- \theta_{j+1} (t_{j+1} - t_j)] \tag{3.13}$$

A probabilidade de um indivíduo sobreviver a t , onde $t_k \leq t < t_{k+1}$ é

$$S_k(t | x) = \left[\prod_{j=0}^{k-1} p_{j+1} \right] p_k,$$

onde $p_k = \exp \left[- \int_{t_k}^t \theta_k du \right] = \exp [- \theta_k (t - t_k)]$. Logo,

$$S_k(t | \mathbf{x}) = \left[\prod_{j=0}^{k-1} \exp\{-\theta_{j+1} (t_{j+1} - t_j)\} \right] \exp\{-\theta_k(t - t_k)\} \quad (3.14)$$

Portanto, o logaritmo de menos $S_k(t | \mathbf{x})$ é

$$-\log S_k(t | \mathbf{x}) = \sum_{j=0}^{k-1} \theta_{j+1} (t_{j+1} - t_j) + \theta_k(t - t_k), \quad (3.15)$$

onde θ_k expressado em (3.3) é o risco referente ao intervalo I_k . A estimativa de (3.15) é

$$-\log \hat{S}_k(t | \mathbf{x}) = \left[\sum_{j=0}^{k-1} \hat{\theta}_{j+1} (t_{j+1} - t_j) + \hat{\theta}_k(t - t_k) \right] \quad (3.16)$$

Da secção 1.6 tem-se que

$$\text{Var}\{-\log S_k(t | \mathbf{x})\} = \left[\frac{\partial}{\partial \beta_j} (-\log S(t | \mathbf{x})) \right]^T \mathbf{I}^{-1}(\beta) \left[\frac{\partial}{\partial \beta_j} (-\log S(t | \mathbf{x})) \right],$$

onde

$$\frac{\partial}{\partial \beta_j} (-\log S(t | \mathbf{x})) = \sum_{j=0}^{k-1} x_{kj} \exp(\beta^T \mathbf{x}_k) (t_{j+1} - t_j) + x_{kj} \exp(\beta^T \mathbf{x}_k)(t - t_k). \quad (3.17)$$

Com base nas informações acima, o intervalo de $100(1-\alpha)\%$ de confiança para $(-\log \hat{S}(t))$ é dado por (1.31).

Foi desenvolvido o programa de computador no módulo IML do SAS (PAMEPTV.PRG, Apêndice D3.10), para estimar os parâmetros do modelo exponencial por partes com covariáveis categóricas, tal como o menos logaritmo da função de sobrevivência em cada intervalo e para construir intervalos de confiança.

Por outro lado, quando as covariáveis são contínuas e/ou discretas, a função risco (3.1) para o modelo exponencial por partes, pode ser expressada por

$$\lambda_k(t, \mathbf{z}) = \exp(\beta_{0k} + \beta^{*T} \mathbf{z}_k^*), \quad t \in I_k, k=1,2, \dots, K \quad (3.18)$$

onde:

$\beta_{0k} = \log(\lambda_k)$ que é o risco constante no intervalo I_k , que assume valores diferentes a cada intervalo;

β_j^* : é o efeito da j -ésima covariável no modelo;

\mathbf{z}_k^* : é o conjunto de fatores de risco (covariáveis) medidas para o i -ésimo indivíduo no intervalo I_k .

Nota-se que o vetor das covariáveis pode ser observado uma única vez ou a cada intervalo.

Para ilustrar, suponha que n indivíduos foram estudados durante um certo período de tempo e observou-se que o risco associado não era constante durante todo o período, porém constante em intervalos de tempo. Assim particionou-se o tempo em k intervalos mutuamente exclusivo, onde foram aferidas p covariáveis contínuas e/ou discretas medidas a cada intervalo I_k . O modelo exponencial por partes incluindo tais covariáveis é dado por (3.18) onde

$$z_k = \begin{bmatrix} 1 & z_{k11} & z_{k12} & \dots & z_{k1p} \\ 1 & z_{k21} & z_{k22} & \dots & z_{k2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & z_{kn1} & z_{kn2} & \dots & z_{knp} \end{bmatrix} \text{ e } \underline{\beta} = \begin{bmatrix} \beta_{0k} \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

z_k : é a matriz de covariáveis medidas no intervalo I_k ;

β_{0k} : é o risco constante no intervalo I_k , assumindo valores diferentes a cada intervalo;

β_j : é o efeito associado a j -ésima covariável.

Nota-se que os indivíduos que falharam no k -ésimo intervalo, não são mais observados no $(k+1)$ -ésimo intervalo. Assim, a função (3.9) pode ser expressada por

$$\log L(\beta_{0k}, \underline{\beta}^*) = \sum_{k=1}^K \sum_{i=1}^n [\delta_{ki} (\beta_{0k} + \underline{\beta}^{*T} z_{ki}^*) - t_{ki} \exp(\beta_{0k} + \underline{\beta}^{*T} z_{ki}^*)]. \quad (3.19)$$

Logo,

$$U(\beta_{0k}) = \frac{\partial}{\partial \beta_{0k}} \log L(\beta_{0k}, \underline{\beta}^*) = \sum_{i=1}^n [\delta_{ki} - t_{ki} \exp(\beta_{0k} + \underline{\beta}^{*T} z_{ki}^*)], \quad k = 1, \dots, K \quad (3.20)$$

$$U(\underline{\beta}) = \frac{\partial}{\partial \beta_j} \log L(\beta_{0k}, \underline{\beta}^*) = \sum_{k=1}^K \sum_{i=1}^n z_{kij} [\delta_{ki} - t_{ki} \exp(\beta_{0k} + \underline{\beta}^{*T} z_{ki}^*)]. \quad (3.21)$$

E a matriz das derivadas segundas é:

$$V_{\beta_{0k}^2} = \frac{\partial^2}{\partial \beta_{0k}^2} \log L(\beta_{0k}, \underline{\beta}^*) = - \sum_{i=1}^n t_{ki} \exp(\beta_{0k} + \underline{\beta}^{*T} z_{ki}^*), \quad k = 1, \dots, K \quad (3.22)$$

$$V_{\underline{\beta}\underline{\beta}} = \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\beta_{0k}, \underline{\beta}^*) = - \sum_{k=1}^K \sum_{i=1}^n z_{kij} t_{ki} \exp(\beta_{0k} + \underline{\beta}^{*T} z_{ki}^*) z_{kiv}, \quad (3.23)$$

$$V_{\beta_{0k}\underline{\beta}} = \frac{\partial^2}{\partial \beta_{0k} \partial \beta_j} \log L(\beta_{0k}, \underline{\beta}^*) = - \sum_{i=1}^n z_{kij} t_{ki} \exp(\beta_{0k} + \underline{\beta}^{*T} z_{ki}^*), \quad k = 1, \dots, K \quad (3.24)$$

Usando (3.22) - (3.24) a matriz assintótica de variância e covariância é dada por

$$\hat{V}_{(k+p) \times (k+p)} = \begin{bmatrix} \hat{V}_{\beta_{0k}^2} & \hat{V}_{\beta_{0k}\beta} \\ \hat{V}_{\beta\beta_{0k}} & \hat{V}_{\beta\beta} \end{bmatrix}$$

Para estimar os parâmetros do modelo exponencial por partes quando as covariáveis são contínuas e/ou discretas e as informações são individuais, foi desenvolvido o programa PAMEP.PRG no módulo IML do SAS, Apêndice D3.11. Nota-se que neste novo contexto T_k é a soma dos tempos de sobrevivência exatos de cada indivíduo no intervalo I_k . Este programa permite estimar a função de sobrevivência e construir intervalos de $100(1-\alpha)\%$ de confiança para a mesma.

3.3. Relação entre o Modelo Logístico Exponencial por Partes e Exponencial por Partes.

Uma vez que o modelo logístico exponencial por partes e log-linear exponencial por partes já foram definidos nas secções (2.5) e (3.2), respectivamente, neste momento será introduzido somente o necessário para atingir o objetivo de verificar a relação dos dois modelos considerando covariáveis contínuas e/ou discretas.

Recordando, seja τ_k o comprimento do intervalo entre os tempos t_k e t_{k+1} e a função logística exponencial por partes dada em (2.46), sob o período I_k , é

$$\log \frac{p_k(\tau_k | z)}{q_k(\tau_k | z)} = \log \frac{1 + \exp[-\lambda_k(z) \tau_k]}{\exp[-\lambda_k(z) \tau_k]} = \beta_{0k} + \beta^T z_k, \quad k=0, 1, \dots, K \quad (3.25)$$

onde $q_k(\tau_k | z)$ e $p_k(\tau_k | z)$ estão definidos em (2.47) e (2.48), respectivamente. A probabilidade de sobrevivência correspondente ao tempo $t \in I_k$, i.e., $t_k \leq t < t_{k+1}$ foi verificado em (2.59) no Capítulo 2, secção 2.5 que é dada por

$$\begin{aligned} \hat{S}_k(t | z; \beta) &= \left\{ \prod_{j=0}^{k-1} \left[\frac{1}{1 + \exp(\beta_{0j} + \beta^{*T} z_j^*)} \right]^{\frac{(t_{j+1} - t_j)}{\tau_j}} \right\} \left\{ \left[\frac{1}{1 + \exp(\beta_{0k} + \beta^{*T} z_k^*)} \right]^{\frac{(t - t_k)}{\tau_k}} \right\} \\ &= \left\{ \prod_{j=0}^{k-1} \left[\frac{\exp(-\beta_{0j} - \beta^{*T} z_j^*)}{1 + \exp(-\beta_{0j} - \beta^{*T} z_j^*)} \right]^{\frac{(t_{j+1} - t_j)}{\tau_j}} \right\} \left\{ \left[\frac{\exp(-\beta_{0k} - \beta^{*T} z_k^*)}{1 + \exp(-\beta_{0k} - \beta^{*T} z_k^*)} \right]^{\frac{(t - t_k)}{\tau_k}} \right\} \end{aligned} \quad (3.26)$$

Por outro lado, o modelo baseado no risco da exponencial por partes é dado por

$$\lambda(t, z) = \lambda_k(t) \exp(\alpha^T z_k^*), \quad (3.27)$$

onde o vetor dos coeficientes α relacionados aos fatores de risco z não são necessariamente os mesmos que os coeficientes β^* nas equações (3.25) e (3.26). A função sobrevivência correspondente a (3.27) para $t_k \leq t < t_{k+1}$ é

$$S_k(t | z; \alpha) = \left\{ \prod_{j=0}^{k-1} \exp[-\lambda_j(t) \exp(\alpha^T z_j^*) [t_{j+1} - t_j]] \right\} \left\{ \exp[-\lambda_k(t) \exp(\alpha^T z_k^*) [t - t_k]] \right\}, \quad (3.28)$$

e

$$-\log \hat{S}_k(t | z; \alpha) = \sum_{j=0}^{k-1} \lambda_j(t) \exp(\alpha^T z_j^*) [t_{j+1} - t_j] + \lambda_k(t) \exp(\alpha^T z_k^*) [t - t_k]. \quad (3.29)$$

Uma relação entre uma regressão logística exponencial por partes e o caso do modelo de risco log-linear exponencial por partes com $\lambda_k(t) = \lambda_k$, onde λ_k é constante dentro do intervalo I_k é agora estudado. Para o modelo logístico exponencial por partes em (3.26) o menos logaritmo de $\hat{S}_k(t | z; \beta)$ é dado por

$$\begin{aligned} -\log \hat{S}_k(t | z; \beta) &= \sum_{j=0}^{k-1} \frac{(t_{j+1} - t_j)}{\tau_j} \log \left[\frac{1 + \exp[-\beta_{0j} - \beta^{*T} z_j^*]}{\exp[-\beta_{0j} - \beta^{*T} z_j^*]} \right] + \frac{(t - t_k)}{\tau_k} \log \left[\frac{1 + \exp[-\beta_{0k} - \beta^{*T} z_k^*]}{\exp[-\beta_{0k} - \beta^{*T} z_k^*]} \right] \\ &= \sum_{j=0}^{k-1} \frac{(t_{j+1} - t_j)}{\tau_j} \log \left[1 + \frac{1}{u_j} \right] + \frac{(t - t_k)}{\tau_k} \log \left[1 + \frac{1}{u_k} \right], \end{aligned} \quad (3.30)$$

onde $u_k = \exp[-\beta_{0k} - \beta^{*T} z_k^*]$. A expressão pode ser aproximada através da expansão de série de Taylor, como

$$\log \left[1 + \frac{1}{u_k} \right] \approx \frac{1}{u_k}$$

se $\frac{1}{u_k}$ é pequeno. A aproximação

$$-\log \hat{S}_k(t | z; \beta) \approx \sum_{j=0}^{k-1} \frac{(t_{j+1} - t_j)}{\tau_j} \exp[\beta_{0j} + \beta^{*T} z_j^*] + \frac{(t - t_k)}{\tau_k} \exp[\beta_{0k} + \beta^{*T} z_k^*]. \quad (3.31)$$

Assim, a equação (3.30) pode ser aproximada a expressão (3.29)

$$\sum_{j=0}^{k-1} \frac{(t_{j+1} - t_j)}{\tau_j} \exp[\beta_{0j} + \beta^{*T} z_j^*] + \frac{(t - t_k)}{\tau_k} \exp[\beta_{0k} + \beta^{*T} z_k^*] \approx \sum_{j=0}^{k-1} \lambda_j(t) \exp(\alpha^T z_j^*) [t_{j+1} - t_j] + \lambda_k(t) \exp(\alpha^T z_k^*) [t - t_k]. \quad (3.32)$$

a equação (3.32) sugere a correspondência natural $\beta_{0k} = \log [\tau_k \lambda_k]$, $k = 1, \dots, K$ e $\beta_j \simeq \alpha_j$, $j = 1, \dots, p$. Nota-se que o comprimento do intervalo τ_k afeta diretamente o coeficiente β_{0k} na regressão logística exponencial por partes.

3.4. Unificação dos Modelos Exponencial por Partes, Poisson e Multinomial.

O objetivo desta secção é apresentar a equivalência estudada por Holford [1980], entre os parâmetros da regressão de Poisson, regressão exponencial por partes e regressão multinomial, sob certas condições particulares, com uma finalidade primordial: permitir o uso de pacotes computacionais conhecidos a fim de estimar os parâmetros de regressão de um modelo, através de seu equivalente correspondente.

Para tanto, é apresentado o modelo de regressão de Poisson para análise de dados de sobrevivência provenientes de tabelas de vida (Tabela 3.1) e equivalência deste com o modelo exponencial por partes secção 3.2 e o multinomial.

3.4.1 Modelo de Regressão de Poisson.

A regressão de Poisson foi definida por Koch [1983] como método estatísticos para analisar a relação entre uma contagem observada que segue distribuição de Poisson e um conjunto de variáveis exploratórias independentes.

O modelo de regressão pode representar, por exemplo:

1. Contagens de bactérias ou vírus para um conjunto de diferentes diluição e/ou condições ambientais.
2. Número de falhas de um determinado equipamento durante operações diversas.
3. Estatísticas vitais pertencentes à mortalidade ou morbidade infantil.
4. Incidência de câncer de uma amostra classificada segundo característica demográficas e outras características.

Suponha-se que d_x representa contagens independentes, com distribuição de Poisson, com valor esperado $\theta = \theta_x$, onde x é a matriz de planejamento que especifica os níveis das p covariáveis categóricas. O modelo de regressão de Poisson apresentado por Koch [1983] é

$$m_x = T_x G(x; \beta), \tag{3.33}$$

onde:

m_x é o valor esperado do número de eventos d_x ; i.e., $m_x = E[d_x]$

T_x é a população exposta ao risco

$G(x, \beta)$ é uma função real conhecida dos x e de β que especifica a relação entre os riscos.

Especificando, $G(x, \beta) = \exp(\beta^T x)$, onde $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$ é o vetor de parâmetros desconhecidos, com β_j representando o parâmetro que determina a contribuição de x_j , $j = 0, 1, \dots, p$ ($x_0 \equiv 1$). O modelo (3.33) é o modelo log-linear com estrutura de Poisson.

Considere agora um estudo envolvendo uma amostra de n indivíduos, onde para cada indivíduo está associado um conjunto de características x , alguns desses indivíduos apresentam as mesmas características, podendo assim ser agrupados em subconjuntos da amostra. Seja $k = 1, \dots, K$ indexador dos subconjuntos da amostra, para os quais $x_k = (x_{k1}, x_{k2}, \dots, x_{kp})^T$ representa o vetor de p variáveis concomitantes linearmente independente, onde $p \leq K$, seja $d_k = d_{x_k}$ o número de eventos para o k -ésimo subconjunto da amostra, e $T_k = T_{x_k}$ a população exposta ao risco correspondente. Sob o suposto que d_k tem distribuição de Poisson independente, com valor esperado $m_k = m_{x_k}$, a função de verossimilhança para os dados é

$$L(d | m) = \prod_{k=1}^K \frac{m_k^{d_k} \exp(-m_k)}{d_k!} \tag{3.34}$$

onde $d = (d_1, d_2, \dots, d_K)^T$ e $m = (m_1, m_2, \dots, m_K)^T$. As equações de máxima verossimilhança são obtidas substituindo-se o valor de m_k dado em (3.33) na equação acima, e diferenciando o logaritmo da máxima verossimilhança com respeito a cada um dos parâmetros, i.e.

$$\log L(d | m) = \sum_{k=1}^K \left\{ d_k \beta^T x_k - T_k \exp(\beta^T x_k) \right\} + A \tag{3.35}$$

onde $A = \sum_{k=1}^K \left\{ d_k \log T_k - \log d_k! \right\}$ é constante. A derivada primeira é

$$\begin{aligned} U(\beta) &= \frac{\partial}{\partial \beta_j} \log L(d | m) = - \sum_{k=1}^K T_k x_{kj} \exp(\beta^T x_k) + \sum_{k=1}^K d_k x_{kj} \\ &= \sum_{k=1}^K x_{kj} \left[d_k - T_k \exp(\beta^T x_k) \right], \end{aligned} \tag{3.36}$$

e as derivadas segundas são

$$H_{jv}(\beta) = \frac{\partial^2}{\partial \beta_j \partial \beta_v} \log L(\beta) = - \sum_{k=1}^K x_{kj} T_k \exp(\beta^T x_k) x_{kv}. \quad (3.37)$$

O estimador de máxima verossimilhança $\hat{\beta}$ para os parâmetros do modelo log-linear (3.33) é obtido igualando as equações (3.36) a zero, obtendo-se um sistema de equações que usualmente não tem solução explícita, i.e., não são lineares. Assim, são necessários procedimentos iterativos para calcular $\hat{\beta}$. O procedimento empregado é o método de Newton-Raphson dado na subsecção 1.5.1, pois este converge rapidamente se as seguintes condições são satisfeitas:

- i - x tem posto completo p ;
- ii - O modelo apresenta os resíduos $(d - \hat{m})$ pequenos e não correlacionados com outras variáveis exploratórias;
- iii - As contagens d são suficientemente grandes afim de que $d^T x \stackrel{a}{\approx} N_n$ em virtude do Teorema Central do Limite.

As condições (ii) e (iii) garantem que $\hat{\beta} \stackrel{a}{\approx} N[\beta; V(\hat{\beta})]$ onde $V(\hat{\beta})$ é a inversa de $H(\hat{\beta})$. Logo, pode-se testar a hipótese usando a estatística de Wald descrita em (1.32).

Exemplo 3.2: Aplicação da Regressão de Poisson em modelos log-lineares. Os dados apresentados na Tabela 3.2 são contagens de células de bactérias em divisão, considerando 4 grupos experimentais. Eles foram estudados por Vieira [1988] e analisados por Wada.

Tabela 3.2 - Contagem de bactérias em divisão para os 4 grupos experimentais.

Grupos Experimentais	Índice Mitótico Células em divisão
CONTROLE	6, 7, 7, 8, 8, 10, 8, 9, 6, 5
CoAc	4, 3, 7, 5, 4, 5, 4, 7, 5, 7
1 $\mu\text{g/ml}$	6, 5, 8, 4, 6, 11, 3, 11, 9, 6
0.02 $\mu\text{g/ml}$	5, 7, 1, 7, 6, 3, 9, 7, 3, 7

Para cada grupo experimental $h = 1, 2, 3, 4$, a compatibilidade dos dados com a distribuição de Poisson com valor esperado $\exp(\beta_h)$ pode ser investigado ajustando o modelo

log-linear (3.32) com $T_h = 1_{s_h}$ e $x_h = 1_{s_h}$, onde $s_h = 10, 10, 10, 10$, denota o número de réplicas para cada grupo experimental. Neste caso, as equações de máxima verossimilhança (3.36) tem solução explícita

$$\hat{\beta}_h = \text{Log}_e \left[\sum_{i=1}^{s_h} \frac{d_{ih}}{s_h} \right]$$

onde d_{ih} denota a contagem para a i -ésima réplica no h -ésimo grupo experimental. A variância estimada correspondente é

$$V(\hat{\beta}_h) = [s_h \exp(\hat{\beta}_h)]^{-1}.$$

Assim, os resultados $\hat{\beta}_h$ e seus respectivos erros padrão para os dados da Tabela 3.2 estão representados na tabela abaixo.

Grupo	$\hat{\beta}$	Desvio-Padrão
Experimental		
CONTROLE	2.0015	0.1162
CoAc	1.6292	0.1400
1 $\mu\text{g/ml}$	1.9315	0.1204
0.02 $\mu\text{g/ml}$	1.7047	0.1349

Para testar se a divisão de células no grupo controle é igual ao do grupo CoAc, a hipótese de interesse é $H_0: \beta_1 = \beta_2 \Rightarrow H_0: \beta_1 - \beta_2 = 0$ que é a mesma coisa que testar $H_0: C\beta = 0$, onde $C = [1 \ -1 \ 0 \ 0]$ e $\beta = [\beta_1, \beta_2, \beta_3, \beta_4]$. Neste caso a estatística de Wald é

$$Q_W = [\hat{\beta}C]^T [CV(\hat{\beta})C^T]^{-1} C\hat{\beta},$$

e seu resultado $Q_W = 4.1875$ indica que o teste é significativo ao nível de significância de 5%, i.e, existe evidências de que as contagens das células em divisão do grupo controle diferem da contagem do grupo CoAc.

3.4.2. Modelo de Poisson em Análise de Sobrevivência.

Holford [1980] seguindo a mesma linha do modelo discutido na subsecção anterior, discutiu a aplicação do modelo log-linear para taxas de Poisson, porém, usando um novo contexto, i. e., T deixa de ser a população exposta ao risco e passa a representar o tempo total de seguimento dos indivíduos expostos.

Em seguida, Laird et al [1981], compartilhando os pensamentos de Holford,

apresentaram o mesmo modelo quando o tempo total é particionado em k intervalos, $k = 1, \dots, K$ e o número de falhas entre $(t_k, t_{k+1}]$, condicionado a T_k (tempo total de seguimento dos indivíduos expostos no intervalo I_k) tem distribuição de Poisson com $E(d_k | T_k) = m_k = \theta_k T_k$, i.e., com parâmetros que são o produto do número esperado de falha com o tempo total de seguimento dos indivíduos expostos ao intervalo I_k .

Assim, as informações apresentadas na Tabela 3.1, podem ser resumidas em tabelas de contingência contendo as duas informações necessárias para a análise: o número de mortes (d_k) e o tempo total de exposição dos indivíduos (T_k calculado por 3.10), em cada combinação dos níveis das covariáveis e dos intervalos do tempo como mostram as tabelas abaixo.

Tabela 3.3 - Número de mortes no intervalo pelo tipo de doador e grau de compatibilidade.

TIPO	CAD(morto)					LRD(vivo)					
	Grau	0	1	2	3	4	0	1	2	3	4
Tempo Comp.											
7d		13	16	20	8	0	0	2	14	4	1
15d		18	25	20	4	1	1	0	7	3	2
21d		14	12	18	8	0	1	1	3	5	1
1m		13	17	18	10	0	0	2	4	7	1
2m		21	43	36	14	1	3	5	31	8	5
3m		14	30	26	2	3	1	3	21	1	1
6m		23	35	36	10	2	2	2	27	8	4
9m		8	11	6	5	0	0	0	14	3	1
1a		3	16	10	1	0	0	1	11	1	2
1.5a		4	9	11	3	0	0	2	9	0	0
2a		9	6	7	4	1	0	0	10	1	2
2.5a		2	8	8	0	1	0	0	7	2	1
3a		3	4	6	0	1	0	1	3	1	1
3.5a		2	3	5	1	0	0	0	2	1	1
4a		1	1	1	0	0	0	0	0	0	1
> 4a		0	1	0	2	0	0	0	0	0	0

Tabela 3.4 - Total de tempo seguimento de indivíduos expostos ao risco no intervalo pelo tipo de doador e grau de compatibilidade.

TIPO	CAD(morto)					LRD(vivo)					
	Grau										
Tempo	Comp.	0	1	2	3	4	0	1	2	3	4
7d	1648.5	2646.0	2681.0	882.0	126.0	140.0	329.0	2821.0	1134.0	1144.5	
15d	1760.0	2860.0	2904.0	960.0	140.0	156.0	368.0	3140.0	1268.0	1296.0	
21d	1224.0	2034.0	2064.0	684.0	102.0	111.0	273.0	2325.0	927.0	963.0	
1m	1714.5	2920.5	2934.0	945.0	153.0	162.0	396.0	3456.0	1336.5	1435.5	
2m	5205.0	8835.0	8970.0	2790.0	495.0	495.0	1215.0	10995.0	4230.0	4695.0	
3m	4680.0	7740.0	8040.0	2550.0	435.0	435.0	1095.0	10215.0	4095.0	4605.0	
6m	12375.0	20295.0	21330.0	7110.0	1080.0	1170.0	3060.0	28485.0	11880.0	13590.0	
9m	10980.0	18225.0	19440.0	6435.0	990.0	1080.0	2970.0	26640.0	11385.0	13365.0	
1a	10485.0	17010.0	18720.0	6165.0	990.0	1080.0	2925.0	25515.0	11205.0	13230.0	
1.5a	20340.0	31770.0	35550.0	11970.0	1980.0	2160.0	5580.0	49230.0	22320.0	26280.0	
2a	18990.0	30420.0	33210.0	11160.0	1890.0	2160.0	5400.0	46800.0	22050.0	25740.0	
2.5a	16290.0	25020.0	27270.0	9360.0	1260.0	1980.0	5040.0	39780.0	20070.0	23040.0	
3a	12420.0	18540.0	19350.0	7110.0	540.0	1800.0	3870.0	30060.0	16740.0	17820.0	
3.5a	8370.0	13140.0	13590.0	5040.0	360.0	1350.0	2790.0	21510.0	12690.0	12240.0	
4a	4680.0	7380.0	8640.0	2970.0	360.0	810.0	1890.0	13770.0	8100.0	7560.0	
> 4a	1440.0	2160.0	3060.0	990.0	180.0	360.0	630.0	5130.0	2970.0	2700.0	

Nota: $T_k = (E_k - (D_k + W_k)/2)\tau_k$, onde E, D, W são dados na Tabela 3.1 e τ é o comprimento do intervalo em dias.

A definição ou escolha da função de ligação torna-se assim instrumento de grande importância para a investigação da adequação do modelo. Suponha que foram aferidas para cada indivíduo, p covariáveis categóricas de interesse. Assumindo-se que a taxa de falha $\theta_k = m_k / T_k$, está relacionado com o vetor das p covariáveis categóricas através de uma função log-linear, tem-se

$$\theta_x = \exp(\beta^T x) \tag{3.38}$$

onde:

x: é a matriz de planejamento que especifica os níveis das p covariáveis concomitantes e incorpora o efeito do tempo, como visto na secção 3.2.

β : é o vetor coluna dos parâmetros desconhecidos que especifica os efeitos das covariáveis e do tempo.

Lembrando ainda que os dados podem, ser representados por tabelas de contingência, sendo que o valor esperado em cada casela é $m_k = \theta_k T_k$ que é a média da distribuição de Poisson, tem-se

$$\log \hat{m}_k = \log \hat{\theta}_k + \log T_k$$

onde $\hat{\theta}_k$ é dado por (3.38). Logo

$$\log \hat{\theta}_k = \log \left[\frac{\hat{m}_k}{T_k} \right] = \hat{\beta}^T x_k \quad (3.39)$$

que é a taxa de risco a ser modelada em cada casela.

Assim, as equações de máxima verossimilhança usadas para estimar os β são equivalentes a (3.36), pois a única diferença entre estas e o modelo da subsecção 3.4.1 é a definição de T.

Para ajustar o modelo de Poisson pode-se usar o programa PAMEPTV.PRG, desenvolvido para o ajuste do modelo exponencial por partes, uma vez que as equações máxima verossimilhança expressadas em (3.36) e (3.11), respectivamente são as mesmas. Levando em conta estes resultados, Holford [1980] enunciou um teorema unificando o estimador de máxima verossimilhança para o modelos exponencial por partes, Poisson e Multinomial.

3.4.3. Teorema de Holford.

Suponha que $\theta_k = \exp(\alpha + \beta^{*T} x_k^*)$, onde $x_k^{*T} = (x_{k1}, x_{k2}, \dots, x_{kp})$ é o conjunto de covariáveis categóricas e $\beta^T = (\beta_1, \beta_2, \dots, \beta_p)$ os efeitos associados as covariáveis. O estimador de máxima verossimilhança para β é equivalente para

$$(i) \quad t_{ki} \sim \text{Exponencial}(\theta_k) \quad k = 1, 2, \dots, K \quad \text{e} \quad i = 1, 2, \dots, n_k,$$

onde i indexa o indivíduo e t_{ki} é o tempo de falha do i-ésimo indivíduo;

$$(ii) \quad d_k \sim \text{Poisson}(\theta_k T_k) \quad k = 1, 2, \dots, K,$$

onde d_k é o número de falhas no intervalo I_k e T_k é o tempo total de seguimento dos indivíduos expostos ao risco no intervalo I_k .

$$(iii) \quad d_{K \times 1} \sim \text{Multinomial}(P_{K \times 1}, d) \quad k = 1, 2, \dots, K.$$

onde

$$P_k = \frac{T_k}{\sum_{k=1}^K T_k} \exp(\alpha + \beta^{*T} x_k^*), \quad d = \sum_{k=1}^K d_k \quad \text{e} \quad \sum_{k=1}^K P_k = 1.$$

Prova:

A função de verossimilhança exponencial por partes da secção 3.2 é agora expressa por

$$L(\theta) = \prod_{k=1}^K \prod_{i=1}^{n_k} [\theta_k]^{\delta_{ki}} \exp(-\theta_k t_{ki}),$$

o logaritmo da função de verossimilhança é

$$\phi_{EP} = \sum_{k=1}^K \sum_{i=1}^{n_k} [\delta_{ki} \log \theta_k - \theta_k T_{ki}] = \sum_{k=1}^K [d_k \log \theta_k - \theta_k T_k] \quad (3.40)$$

Substituindo $\theta_k = \exp(\alpha + \underline{\beta}^{*T} \mathbf{z}_k^*)$ na equação anterior e derivando em relação aos parâmetros α e $\underline{\beta}$ e igualando as equações à zero. A j -ésima equação que deve ser resolvida em ordem para encontrar o estimador de máxima verossimilhança para $\underline{\beta}$ é

$$U(\alpha) = \frac{\partial}{\partial \alpha} \phi_{EP} = \sum_{k=1}^K [d_k - T_k \exp(\alpha + \underline{\beta}^{*T} \mathbf{z}_k^*)] = 0, \quad (3.41)$$

$$U(\underline{\beta}) = \frac{\partial}{\partial \beta_j} \phi_{EP} = \sum_{k=1}^K x_{kj} [d_k - T_k \exp(\alpha + \underline{\beta}^{*T} \mathbf{z}_k^*)] = 0. \quad (3.42)$$

De (3.41) tem-se que

$$\exp(\alpha) = \frac{\sum_{k=1}^K d_k}{\sum_{k=1}^K T_k \exp(\underline{\beta}^{*T} \mathbf{z}_k^*)} = \frac{d}{\sum_{k=1}^K T_k \exp(\underline{\beta}^{*T} \mathbf{z}_k^*)}, \quad (3.43)$$

substituindo (3.43) em (3.42) obtém-se a equação

$$\sum_{k=1}^K x_{kj} d_k - \frac{d \sum_{k=1}^K x_{kj} T_k \exp(\underline{\beta}^{*T} \mathbf{z}_k^*)}{\sum_{k=1}^K T_k \exp(\underline{\beta}^{*T} \mathbf{z}_k^*)} = 0. \quad (3.44)$$

que é a equação a ser resolvida para estimar $\underline{\beta}$.

Em contrapartida, a função de verossimilhança da Poisson é dada por

$$L(\theta) = \prod_{k=1}^K \frac{[T_k \theta_k]^{d_k} \exp(-\theta_k T_k)}{d_k!} = \prod_{k=1}^K \left\{ [\theta_k]^{d_k} \exp(-\theta_k T_k) \right\} \left\{ \frac{[T_k]^{d_k}}{d_k!} \right\}$$

o logaritmo da função acima é

$$\phi_{PO} = \sum_{k=1}^K [d_k \log \theta_k - \theta_k T_k] + \sum_{k=1}^K [d_k \log T_k - \log d_k!], \quad (3.45)$$

o primeiro termo de (3.45) é igual a (3.40) e o segundo é uma constante $A = \sum_{k=1}^K [d_k \log T_k - \log d_k!]$ que não depende dos parâmetros. Portanto, (3.45) pode ser

representada por

$$\phi_{PO} = \phi_{EP} + A.$$

Logo o estimador de máxima verossimilhança de $\underline{\beta}$ em (3.45) e (3.40) são equivalentes e expressados em (3.44).

Finalmente a função de verossimilhança da distribuição multinomial é

$$L(P_k) \approx d! \prod_{k=1}^K \frac{[P_k]^{d_k}}{d_k!},$$

onde $\log L(P_k) = \log d! + \sum_{k=1}^K d_k \log P_k - \sum_{k=1}^K \log d_k!$, sujeito a restrição $\sum_{k=1}^K P_k = 1$, tem-se que

$$\phi_M = \text{Log } L(P_k) = \log d! + \sum_{k=1}^K d_k \log P_k - \lambda \left(\sum_{k=1}^K P_k - 1 \right) - \sum_{k=1}^K \log d_k!$$

Substituindo-se $P_k = \frac{T_k}{\sum_{k=1}^K T_k} \exp(\alpha + \underline{\beta}^{*T} \mathbf{x}_k^*)$ obtém-se

$$\log L(P_k) \approx \sum_{k=1}^K d_k \log \frac{T_k}{\sum_{k=1}^K T_k} \exp(\alpha + \underline{\beta}^{*T} \mathbf{x}_k^*) - \lambda \left(\sum_{k=1}^K \frac{T_k}{\sum_{k=1}^K T_k} \exp(\alpha + \underline{\beta}^{*T} \mathbf{x}_k^*) - 1 \right). \quad (3.46)$$

Derivando-se a equação acima em relação aos parâmetros α , $\underline{\beta}$ e λ obtém-se

$$U(\alpha) = \frac{\partial}{\partial \alpha} \phi_M = \sum_{k=1}^K d_k - \lambda \left[\sum_{k=1}^K \frac{T_k}{\sum_{k=1}^K T_k} \exp(\alpha + \underline{\beta}^{*T} \mathbf{x}_k^*) \right] = 0, \quad (3.47)$$

$$U(\underline{\beta}) = \frac{\partial}{\partial \beta_j} \phi_M = \sum_{k=1}^K d_k x_{kj} - \lambda \left[\sum_{k=1}^K \frac{T_k x_{kj}}{\sum_{k=1}^K T_k} \exp(\alpha + \underline{\beta}^{*T} \mathbf{x}_k^*) \right] = 0, \quad (3.48)$$

$$U(\lambda) = \frac{\partial}{\partial \lambda} \phi_M = - \sum_{k=1}^K \frac{T_k}{\sum_{k=1}^K T_k} \exp(\alpha + \underline{\beta}^{*T} \mathbf{x}_k^*) + 1 = 0, \quad (3.49)$$

representando-se $T = \sum_{k=1}^K T_k$, de (3.47) tem-se que

$$\exp(\alpha) = \frac{T}{\sum_{k=1}^K T_k \exp(\underline{\beta}^{*T} \mathbf{x}_k^*)}. \quad (3.50)$$

De (3.48)

$$\lambda = \frac{Td}{\exp(\alpha) \sum_{k=1}^K T_k \exp(\underline{\beta}^{*T} \underline{x}_k^*)} \tag{3.51}$$

Por fim, substituindo-se (3.50) e (3.51) em (3.49) obtém-se

$$\sum_{k=1}^K x_{kj} d_k - \frac{d \sum_{k=1}^K x_{kj} T_k \exp(\underline{\beta}^{*T} \underline{x}_k^*)}{\sum_{k=1}^K T_k \exp(\underline{\beta}^{*T} \underline{x}_k^*)} = 0. \tag{3.52}$$

Logo verifica-se que a equação acima é idêntica a equação obtida em (3.44). Assim o estimador de máxima verossimilhança de $\hat{\underline{\beta}}$ em (i) - (iii) são equivalentes, o que implica que os estimadores de máxima verossimilhança de $\hat{\underline{\beta}}$ podem ser encontrados usando qualquer um dos modelos acima. Entretanto, se há interesse em estimar a probabilidade de sobrevivência associada a cada indivíduo, o modelo exponencial por partes ou Poisson é o adequado, sendo inviável a utilização do modelo multinomial uma vez que o estimador de máxima verossimilhança de $\hat{\alpha}$ difere neste com os demais.

O programa PAMLOGHM (Apêndice D3.12), foi desenvolvido com intuito de ajustar o modelo log-linear multinomial definido em (iii), uma vez que os procedimentos conhecidos para o ajuste de tal modelo, por exemplo, CATMOD (procedimento próprio para ajuste de modelo log-linear para dados categóricos) do SAS não fornecem o estimador de máxima verossimilhança para o modelo com estrutura da multinomial apresentada por Holford.

3.5. Exemplo.

Com o intuito de ilustrar a teoria acima desenvolvida, considerou-se num primeiro instante os dados da Tabela 3.1 para apresentar alguns resultados discutidos por Laird et al [1981]. Estes resultados foram obtidos pelo ajuste do modelo de regressão exponencial por partes sob estimadores de máxima verossimilhança, através do programa computacional PAMEPTV.PRG (ver Apêndice D3.10).

Como uma primeira análise usou-se os dados como mostram as Tabelas 3.3 e 3.4, i.e., considerou-se os 16 intervalos de tempo e as 5 categorias de grau de compatibilidade, ajustou-se um modelo com efeitos principais do tipo de doador, grau de compatibilidade e tempo. O modelo é dado por:

$$\log \lambda(t; \underline{x}) = \mu + \mu_0 (I_k) + \sum_{j=1}^{I_1} \mu_1 (j) + \sum_{j=1}^{I_2} \mu_2 (j),$$

onde:

μ : efeito da média total da tabela;

$\mu_1(l_1)$: efeito principal associado ao tipo de doador na sobrevivência, $l_1 = 2$;

$\mu_2(l_2)$: efeito principal associado ao grau de compatibilidade na sobrevivência, $l_2 = 5$;

$\mu_0(l_k)$: efeito principal associado ao intervalo de tempo, $k = 1, 2, \dots, 16$.

As restrições do modelo são $\sum_j^{l_1} \mu_1(j) = \sum_j^{l_2} \mu_2(j) = \sum_k^K \mu_0(k) = 0$. Os efeitos estimados para este

modelo estão sumarizados na tabela abaixo.

Tabela 3.5 - Estimativas de máxima verossimilhança dos parâmetros e seus respectivos erros padrão.

PARÂMETRO	Estimativa	Erro-Padrão
μ	- 7.4686	0.0677
$\mu_1(1)$	0.3403	0.0409
* $\mu_1(2)$	- 0.3403	-
$\mu_2(0)$	0.2579	0.0809
$\mu_2(1)$	0.2337	0.0703
$\mu_2(2)$	0.1982	0.0598
$\mu_2(3)$	0.0031	0.0849
* $\mu_2(4)$	- 0.6929	-
$\mu_0(1)$	2.0434	0.1216
$\mu_0(2)$	1.9967	0.1198
$\mu_0(3)$	2.0809	0.1321
$\mu_0(4)$	1.8550	0.1253
$\mu_0(5)$	1.5749	0.0937
$\mu_0(6)$	1.1833	0.1101
$\mu_0(7)$	0.5695	0.0970
$\mu_0(8)$	- 0.4737	0.1475
$\mu_0(9)$	- 0.4938	0.1516
$\mu_0(10)$	- 1.3132	0.1630
$\mu_0(11)$	- 1.2092	0.1594
$\mu_0(12)$	- 1.3613	0.1836
$\mu_0(13)$	- 1.4444	0.2175
$\mu_0(14)$	- 1.3852	0.2488
$\mu_0(15)$	- 2.2073	0.4715
* $\mu_0(16)$	- 1.4156	-
Ajuste do modelo $Q_L = 154.8109$		(p-value = 0.1699)

Nota: (*) - Valores obtidos sob as restrições do modelo.

Estas estimativas indicam que o risco é menor para os pacientes com maior grau de compatibilidade e o doador vivo (LRD), e que o período imediatamente seguido ao transplante

apresenta um risco relativamente alto, que diminui com o tempo. As estimativas sugerem que os efeitos do grau de compatibilidade é não linear, sendo que as diferenças entre os três primeiros graus são pequenas; assim, decidiu-se juntar os três reduzindo para 3 níveis o grau de compatibilidade (0-2,3,4). Os efeitos relacionados aos intervalos de tempo sugerem que podem ser redivididos, reduzindo de 16 para 5 intervalos. A divisão correspondente aos novos intervalos é 0-1 mês, 1-3 meses, 3-6 meses, 6 meses - 1 ano, e > 1 ano. Assim, os dados da Tabela 3.3 e 3.4 foram reduzidos a uma tabela 3x2x5 ($l_1 = 2$ e $l_2 = 3$).

Por fim, apresenta-se o modelo reduzido contendo o efeito principal do tempo e os efeitos combinados dos níveis de grau de compatibilidade e tipo de doador. O modelo é

$$\text{Log } \lambda(t; \mathbf{x}) = \mu + \mu_0(I_k) + \sum_{jj'=1}^{l_1 l_2} \mu_{12}(jj')$$

onde

μ : efeito da média total da tabela;

$\mu_0(I_k)$: efeito principal associado ao intervalo de tempo, $k = 1, 2, \dots, 5$;

$\mu_{12}(jj')$: efeito combinado ao tipo de doador com grau de compatibilidade na sobrevivência com $j = 1, 2$ e $j' = 1, 2, 3$; por exemplo $\mu_{12}(11)$ é o efeito combinado do tipo de doador (CAD) com grau de compatibilidade (0-2).

As restrições do modelo são $\sum_k \mu_0(I_k) = \sum_{jj'} \mu_{12}(jj') = 0$. Os resultados estão sumarizados na tabela abaixo.

Tabela 3.6 - Estimativas de máxima verossimilhança dos parâmetros e seus respectivos erros padrão.

PARÂMETRO	Estimativa	Erro-Padrão
μ	- 7.1190	0.0728
$\mu_0(1)$	1.5639	0.0575
$\mu_0(2)$	0.9826	0.0590
$\mu_0(3)$	0.1450	0.0727
$\mu_0(4)$	- 0.9077	0.0878
* $\mu_0(5)$	- 1.7838	-
$\mu_{12}(11)$	0.6214	0.0789
$\mu_{12}(12)$	0.5307	0.1200
$\mu_{12}(13)$	0.4190	0.2680
$\mu_{12}(21)$	0.0054	0.0929
$\mu_{12}(22)$	- 0.4248	0.1413
* $\mu_{12}(23)$	- 1.1517	-
Ajuste do modelo $Q_L = 28.0099$		(p-value = 0.109)

Nota: (*) - Valores obtidos sob as restrições do modelo.

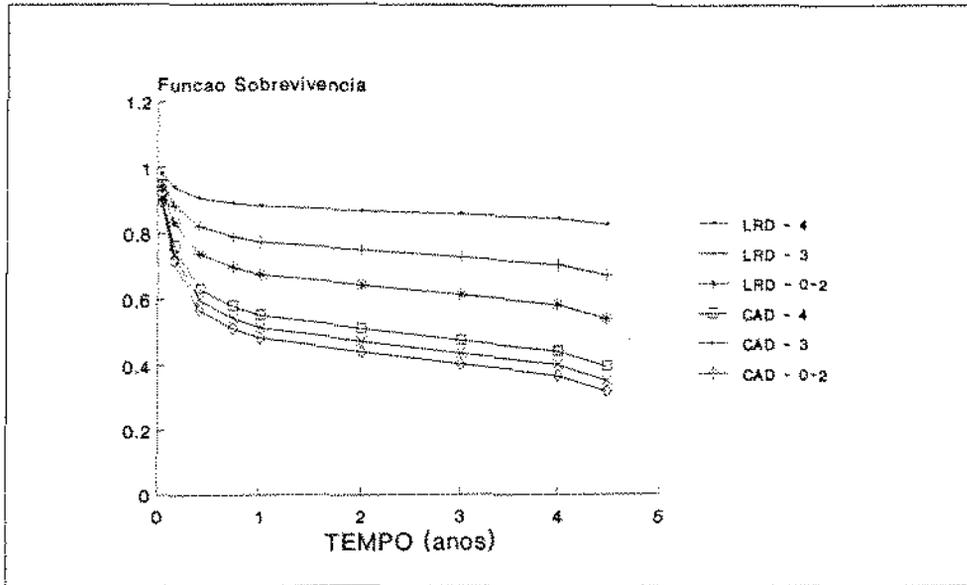


Figura 3.1 - Ajuste das curvas de sobrevivência.

A Figura 3.1 mostra a curva da sobrevivência estimada, considerando os seis grupos, i.e., para as combinações dos níveis de grau de compatibilidade e tipo de doador ao longo do tempo. Nota-se que os indivíduos cujo tipo de doador é vivo e com alto grau de compatibilidade possui maior probabilidade de sobrevivência que os outros grupos. Conclui-se que o tipo de doador mais adequado é vivo e quanto maior o grau de compatibilidade, maior a sobrevivência. Nota-se ainda pelo gráfico, que dois ou três intervalos de tempo devem ser suficientes para caracterizar a função de sobrevivência dos pacientes transplantados.

CAPÍTULO 4

APLICAÇÃO A UM CASO REAL.

4.1. Problema.

Será considerado um estudo de ensaios clínico sobre endomiocardiofibrose¹, conduzido pelo INCOR (Instituto do Coração - Hospital das Clínicas S.P). Este estudo envolvem 151 pacientes que poderiam ser indicados para cirurgia ou simplesmente receber tratamento clínico dependendo da sua classe funcional. A classe funcional é constituída das categorias:

- 1 - quando o paciente não apresenta sintomas que é cardíaco
- 2 - quando o paciente é cardíaco e não se altera quando exerce um grande esforço
- 3 - quando o paciente é cardíaco necessitando de pouco esforço para se cansar e
- 4 - quando o paciente é cardíaco e se cansa ao menor esforço.

Os pacientes que forem classificados tendo as classes funcionais 3 ou 4 são designados para cirurgia exceto aqueles que apresentam em seu quadro clínico algum fator que venha interferir

¹: é uma doença caracterizada por fibrose nos ventrículos esquerdo, direito ou em ambos; i.e. , os ventrículos são tomados por fibras que impedem a circulação do sangue, fazendo com que o coração tenha um mal funcionamento.

no sucesso da operação; por exemplo, pacientes hipertensos. Assim, os pacientes foram divididos em 2 tratamentos: cirúrgico (com 81 pacientes) e clínicos (com 70 pacientes). Dos 81 pacientes do tratamento cirúrgico 21 faleceram, sendo que dos 70 do grupo clínico faleceram 29. Foram aferidas para cada paciente dados relativos à identificação do paciente, estado clínico, cirúrgico, eletrocardiograma (ECG), exames laboratoriais, ecocardiograma (ECO) e cateterismo (CAT). Os dados estão listados, na Tabela B2, do Apêndice B, página 122.

Tabela 4.1. - Sumário das variáveis aferidas no estudo .

VARIÁVEIS	Descrição	Tipo	Notação do Tabela B2
DADOS PESSOAIS			
NOME	Iniciais do Nome do Paciente	Caracter	
IDADE	Idade do Paciente	Continua (em anos)	Z6
SEXO	Sexo do Paciente	Discreta F = Feminino M = Masculino	Z7
COR	Cor do Paciente	Discreta BCA = Branca MUL = Mulata PTA = Preta	Z8
PRIMEIRA	Data da Primeira Consulta	Data	
CIRURGIA	Data da Cirurgia	Data	
ULTIMA	Data da última Consulta	Data	
OBITO	Data do Óbito	Data	
ESTADO CLÍNICO			
CF1	Classe Funcional no Início	Discreta 1 = Moderada 2 = Leve 3 = Alta 4 = Intensa	Z2
CF2	Classe Funcional Final	Discreta 1 = Moderada 2 = Leve 3 = Alta 4 = Intensa	Z1
THIST	Tempo de História em anos	Continua	Z9
INTFIBVD	Intensidade de Fibrose de Ventri- culo Direito	Discreta 1 = Discreta 2 = Moderada 3 = Intensa	Z3
INTFIBVE	Intensidade de Fibrose do Ventri- culo Esquerdo	Discreta 1 = Discreta 2 = Moderada 3 = Intensa	Z4
DISPNEIA	Dispneia	Discreta S = Sim N = Não	Z10
EDEMA	Edema	Discreta S = Sim N = Não	Z11
ASCITE	Ascite	Discreta S = Sim N = Não	Z12
DORPREC	Dor Precordial	Discreta S = Sim N = Não	Z13
SMITRAL	Sopro Mitral	Discreta S = Sim N = Não	Z14

Tabela 4.1 - (Continuação) .

VARIÁVEIS	Descrição	Tipo	Notação da Tabela B2
STRICUP	Sopro Tricuspide	Discreta S = Sim N = Não	Z15
FIGADO	Fígado (tamanho em cm)	Contínua	Z16
ESTALIDO	Estalido	Discreta S = Sim N = Não	Z17
CIRURGIA			
RESSECVE	Ressecção da Fibrose VE	Discreta S = Sim N = Não	Z18
RESSECVD	Ressecção da Fibrose VD	Discreta S = Sim N = Não	Z19
PROTESEM	Prótese Mitral	Discreta S = Sim N = Não	Z20
PROTESET	Prótese Tricuspide	Discreta S = Sim N = Não	Z21
PLASTICA	Plástica da Valvula	Discreta S = Sim N = Não	Z22
RESSECPO	Ressecção pelo Ápice	Discreta S = Sim N = Não	Z23
ECG			
FA	Fibrilação Atrial	Discreta S = Sim N = Não	Z24
SAD	Sobrecarga Atrial Direita	Discreta S = Sim N = Não	Z25
SVD	Sobrecarga Ventricular Direita	Discreta S = Sim N = Não	Z26
SAE	Sobrecarga Atrial Esquerda	Discreta S = Sim N = Não	Z27
SVE	Sobrecarga Ventricular Esquerda	Discreta S = Sim N = Não	Z28
BRD	Bloqueio do Ramo Direito	Discreta S = Sim N = Não	Z29
BDAS	Broqueio	Discreta S = Sim N = Não	Z30
BRE	Bloqueio do Ramo Esquerdo	Discreta S = Sim N = Não	Z31
AINATIVA	Área Inativa	Discreta S = Sim N = Não	Z32
ALTT	Alteração da Repolarização Ventricular	Discreta S = Sim N = Não	Z33
LABORATORIO			
TP	Tempo de Protrombina	Contínua (NI \pm 100 %)	Z34
EOSINOFI	Eosinofilia	Contínua (\pm - 5 %)	Z35
ECO			
ECOVE	Diâmetro Ventrículo Esquerdo	Contínua	Z36
ECOAE	Diâmetro Aorta Esquerda	Contínua	Z37
ECOVD	Diâmetro Ventrículo Direito	Contínua	Z38
ECOAD	Diâmetro Aorta Direita	Contínua	Z39
ECOFE	Fração de Ejeção	Contínua	Z5

Tabela 4.1 - (Continuação) .

VARIÁVEIS	Descrição	Tipo	Notação da Tabela B2
ECOIM	Insuficiência Mitral	Discreta S = Sim N = Não	Z40
ECOIT	Insuficiência Tricuspidé	Discreta S = Sim N = Não	Z41
CAT			
CATAD	Pressão Átrio Direito	Contínua	Z45
CATVDP	Pressão Sistólica do VD	Contínua	Z46
CATVDPDI	Pressão Diastólica Inicial VD	Contínua	Z47
CATVDPDF	Pressão Diastólica Final VD	Contínua	Z48
CATTPPS	Pressão Sistólica Tronco Pulmonar	Contínua	Z49
CATTPPD	Pressão Diastólica Tronco Pulmonar	Contínua	Z50
CATCP	Capilar Pulmonar	Contínua	Z51
CATVEPS	Pressão Sistólica VE	Contínua	Z52
CATVEPDI	Pressão Diastólica Inicial VE	Contínua	Z53
CATVEPDF	Pressão Diastólica Final VE	Contínua	Z54
CATAOPS	Pressão Sistólica Aórtica	Contínua	Z55
CATAOPD	Pressão Diastólica Aórtica	Contínua	Z56
CATIM	Insuficiência Mitral	Discreta S = Sim N = Não	Z57
CATIT	Insuficiência Tricuspidé	Discreta S = Sim N = Não	Z58
CALCIF	Calcificação	Discreta S = Sim N = Não	Z59
TEMPO			
T1	Data do óbito-Data cirurgia	Contínua (em dias)	
T2	Data do óbito-Data 1ª consulta	Contínua (em dias)	
T3	Data da ult cons-Data cirurgia	Contínua (em dias)	
T4	Data da ult - Data 1ª consulta	Contínua (em dias)	
T5	Data cirurgia-Data 1ª consulta	Contínua (em dias)	
TEMPO	Tempo: T2 se o paciente morreu T4 se o paciente esta vivo	Contínua (em dias)	
δ	Status do Paciente	Discreta 1 = Falha 0 = Censura	
TRATAM	Tipo de Tratamento	Discreta 1 = Cirúrgico 2 = Clínico	

4.2. Objetivo.

O objetivo é estabelecer os fatores prognósticos de sobrevivência de pacientes com Endomiocardiofibrose e verificar se a cirurgia prolonga a vida dos pacientes em comparação com tratamento clínico.

4.3. Metodologia de Análise.

Como a variável de interesse é tempo de sobrevivência dos pacientes, que pode ser tempo de falha ou tempo censurado. Deve-se utilizar métodos de análise de sobrevivência, pois estes como já vistos, permitem incorporar informações a respeito do tempo de censura. O tempo de sobrevivência em dias para cada paciente é:

$$\text{Tempo} = \text{Data do óbito} - \text{Data da primeira consulta} = T2,$$

se o paciente falhar, ou

$$\text{Tempo} = \text{Data da última consulta} - \text{Data da primeira consulta} = T4,$$

se o paciente é censurado.

Por exemplo, o primeiro paciente tem $\text{Tempo} = \text{Data óbito} - \text{Data primeira consulta} = 12/09/81 - 10/09/80 = 367$ dias. (ver Tabela B2, Apêndice B, página 122)

A. Descritiva.

Sabe-se que na maioria das vezes, a distribuição associada à variável aleatória Tempo é a exponencial. Foi feita uma análise exploratória utilizando as técnicas de ramo e folhas (forma de sumarizar os dados) e desenho esquemático (traduz em um desenho as informações mais importantes, tais como mediana, média, mínimo, máximo e quartis) na busca de indicações de possíveis distribuições associadas aos tempos.

Numa segunda etapa da investigação utilizou-se os tempos, levando em conta tempos censurado, para estimar a função de sobrevivência usando a tabela de vida. Traçou-se os gráficos de interesse (secção 1.3.2) para verificar a distribuição adequada aos dados. Considerando somente a variável tempo e o indicador de censura, ajustou-se o modelo paramétrico Weibull e exponencial.

Uma exploração inicial se faz necessário, afim de selecionar somente as variáveis que contribuem na sobrevivência de pacientes que sofrem de endomiocardiofibrose. Portanto, foi feita uma análise descritiva para investigar tais variáveis, verificar o status de sobrevivência (0 = vivos, 1 = falecidos) versus outras variáveis categóricas (no caso de contínuas a variável foi categorizada).

Primeiramente foi feita uma análise descritiva dos dados composta de cálculos de médias, medianas, desvios padrão, valores máximos e mínimos. Para cada uma das variáveis discretas construiu-se uma tabela de contingência, a partir da qual testou-se a hipótese de igualdade de proporções (falha ou não falha) para 2 amostras independentes, utilizando-se da estatística Qui - Quadrado de Pearson. Para cada uma das variáveis contínuas os dados foram

dicotomizados, isto é, dividiu-se em duas categorias (\geq mediana e $<$ mediana) e procedeu-se a análise empregando-se a mesma técnica usada para as variáveis discretas. Nos casos onde havia evidências de se rejeitar a hipótese, prosseguiu-se a análise para detectar quais realmente diferiam através do teste log-rank, para comparação dos tempos de sobrevivência dos grupos construídos como descrito acima (pela mediana).

Após verificar-se detalhadamente quais as variáveis independentes que mostraram ser importantes na sobrevivência dos pacientes, foi investigada também a existência de interação destas com os tratamentos cirúrgico e clínico. Procedeu-se da seguinte maneira: contruiu-se tabelas de contingências para verificar a relação entre tratamentos (1 = cirúrgico, 2 = clínico) versus outras variáveis concomitantes utilizando-se a estatística χ^2 .

B. Modelos de Regressão.

Finalmente, com o intuito de identificar fatores prognósticos importantes na sobrevivência de pacientes com endomiocardiopfibrose, ajustou-se modelos de regressão.

Todos os cálculos foram feitos utilizando-se o *software* estatístico SAS (Statistical Analysis System).

4.4. Resultados.

A. Descritiva.

A1. Análise descritiva da variável dependente tempo de sobrevivência.

0	00011111222222333344444	0	00000000000111111111222233344
0	55555566667777889999999	0	5567789
1	0011122222334444	1	0113334
1	5566678888	1	6677
2	01222333444	2	3
2	67789	2	
3	044	3	1
3	5889	3	
4	002	4	
4	59	4	6
	Censura		Falha

Quadro 4.1 - Ramo e Folhas dos tempos de censura e falha dos pacientes com endomiocardiopfibrose. (em milhares de dias)

Observando o quadro anterior pode-se fazer as seguintes observações a respeito dos conjuntos de valores em questão:

i) No Ramo e Folhas dos tempos de falha há um destaque ao valor 46 - correspondente ao tempo de falha do paciente 18. Além deste está bem destacado o valor 31. Os demais valores estão concentrados entre 0 e 23. Há uma acentuada assimetria em direção a valores grandes.

(ii) Olhando-se para o aspecto visual da distribuição dos tempos de falha e censura, faz sentido supor que tal conjunto de valores pudesse ser, de alguma forma, considerado como provenientes de um conjunto de dados com distribuição exponencial.

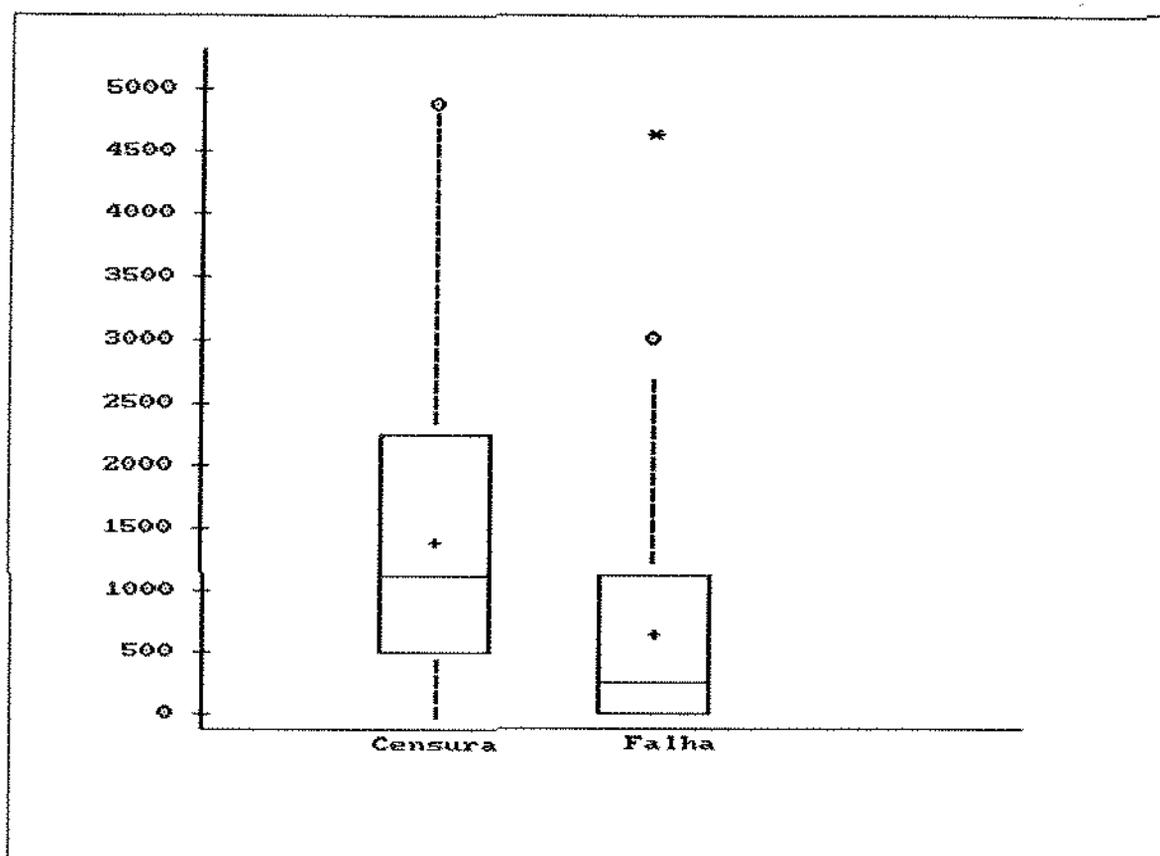


Figura 4.1 - Desenho esquemático correspondentes aos tempos de censura e falha dos pacientes com endomiocardiopatia.

Na Figura 4.1, mais do que no Quadro 4.1 aparece o real destaque do valor correspondente ao paciente 18. Pode-se observar que para os tempos de censura a assimetria é menos acentuada do que nos tempos de falha. O tempo mediano de sobrevivência é maior nos pacientes censurados.

Tabela 4.2 - Sumário das estatísticas para as variáveis contínuas.

VARIÁVEL	N	Média	Mediana	d.p	Máximo	Mínimo
IDADE	151	36.92	38	13.06	64	5
THIST	136	4.29	3	4.17	20	1
FIGADO	135	3.71	4	3.19	13	0
TP	99	67.79	70	18.36	99	10
EOSINOFI	108	5.64	2	11.48	78	0
ECOVE	79	52.73	53	9.26	76	27
ECOAE	77	42.54	41	9.62	71	26
ECOVD	66	23.80	20	10.14	58	10
ECOFE	0.7	0.73	0.75	0.09	0.89	0.37
CATAD	111	14.62	16	6.81	29	2
CATVDPS	112	44.87	40	21.18	99	12
CATVDPDI	100	6.15	6	5.65	24	0
CATVDPDF	112	15.95	16	6.56	32	2
CATTPPS	112	44.17	39.5	21.27	99	12
CATTPPD	111	23.43	20	11.14	58	6
CATCP	116	20.94	20	9.85	52	4
CATVEPS	115	119.52	119	20.72	219	82
CATVEPDI	106	4.42	0	6.37	26	0
CATVEPDF	115	23.29	23	10.17	50	6
CATAOPS	115	116.60	114	20.27	218	82
CATAOPD	114	76.94	78	10.94	99	47

A2. Estimação da função de sobrevivência.

Tabela 4.3 - Tabela de vida para pacientes com Endomiocardiofibrose.

INTERVALO					
TEMPO (t_i)	n_i'	d_i	$w_i + l_i$	n_i	$\hat{S}(t_i)$
0 + 500	151	30	29	136.5	1.0000
500 + 1000	92	7	18	83.0	0.7802
1000 + 1500	67	6	18	58.0	0.7144
1500 + 2000	43	4	8	39.0	0.6405
2000 + 2500	31	1	11	25.5	0.5748
2500 + 3000	19	0	6	16.0	0.5523
3000 + 3500	13	1	3	11.5	0.5523
3500 + 4000	9	0	4	7.0	0.5043
4000 + 4500	5	0	3	3.5	0.5043
4500 +	2	1	1	1.5	0.5043

Notam n_i' : número de pacientes vivos no início do intervalo t_i
 d_i : número de pacientes mortos durante o intervalo t_i
 $w_i + l_i$: número de pacientes retirados vivos ou perdidos no intervalo t_i
 n_i : $n_i' - \frac{1}{2} [w_i + l_i]$
 $\hat{S}(t_i)$: proporção acumulada de sobrevivência do início do estudo até o término do intervalo.

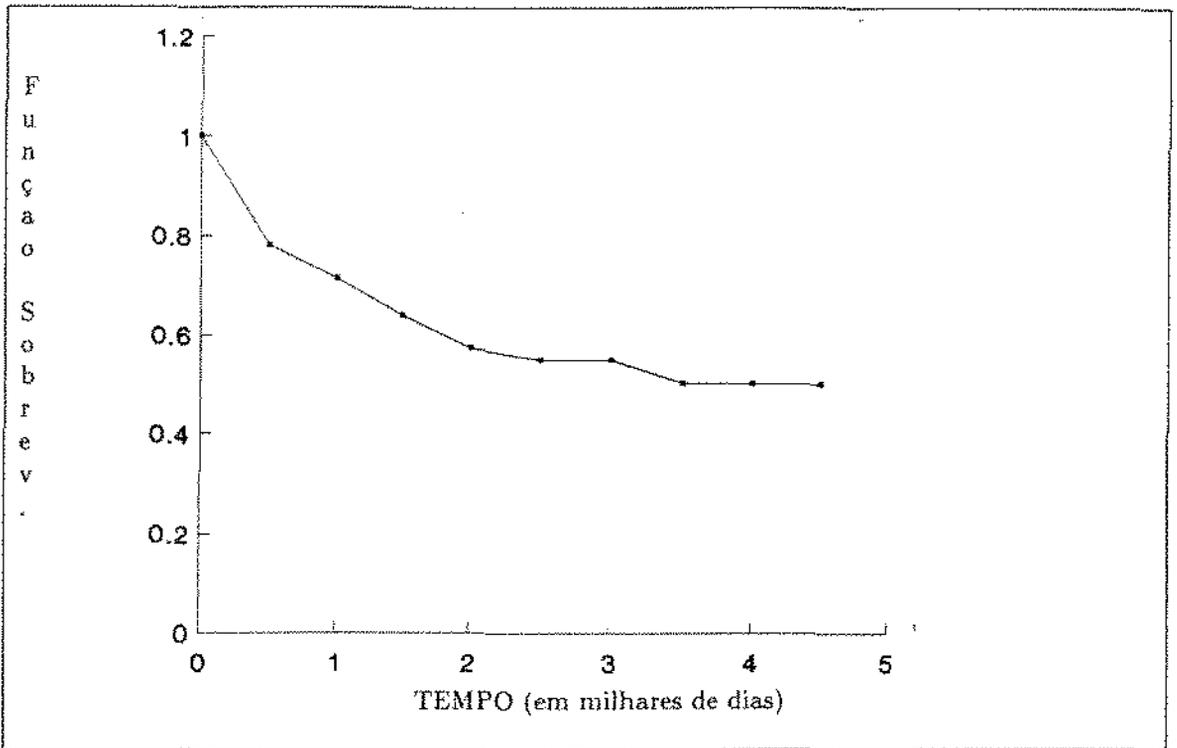


Figura 4.2 - Curva de sobrevivência para pacientes com endomiocardiopatia.

A3. Investigação da distribuição dos tempos através de gráficos.

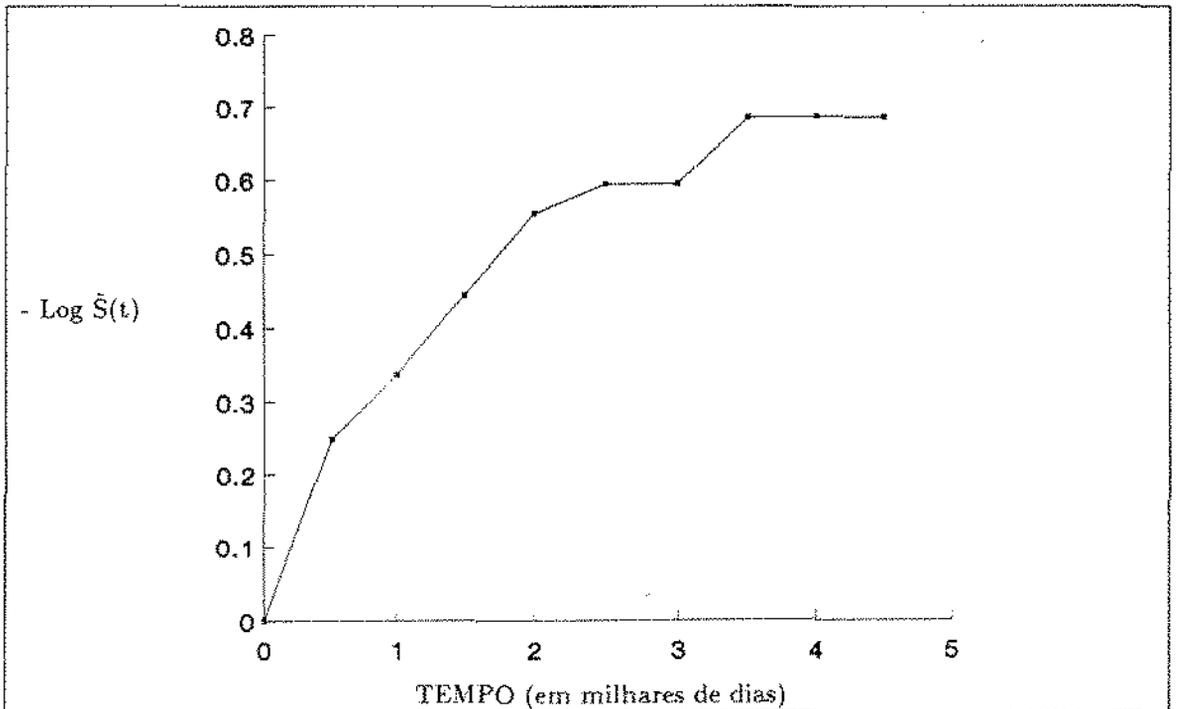


Figura 4.3 - Menos logaritmo da taxa de sobrevivência dos pacientes com endomiocardiopatia ao longo do tempo.

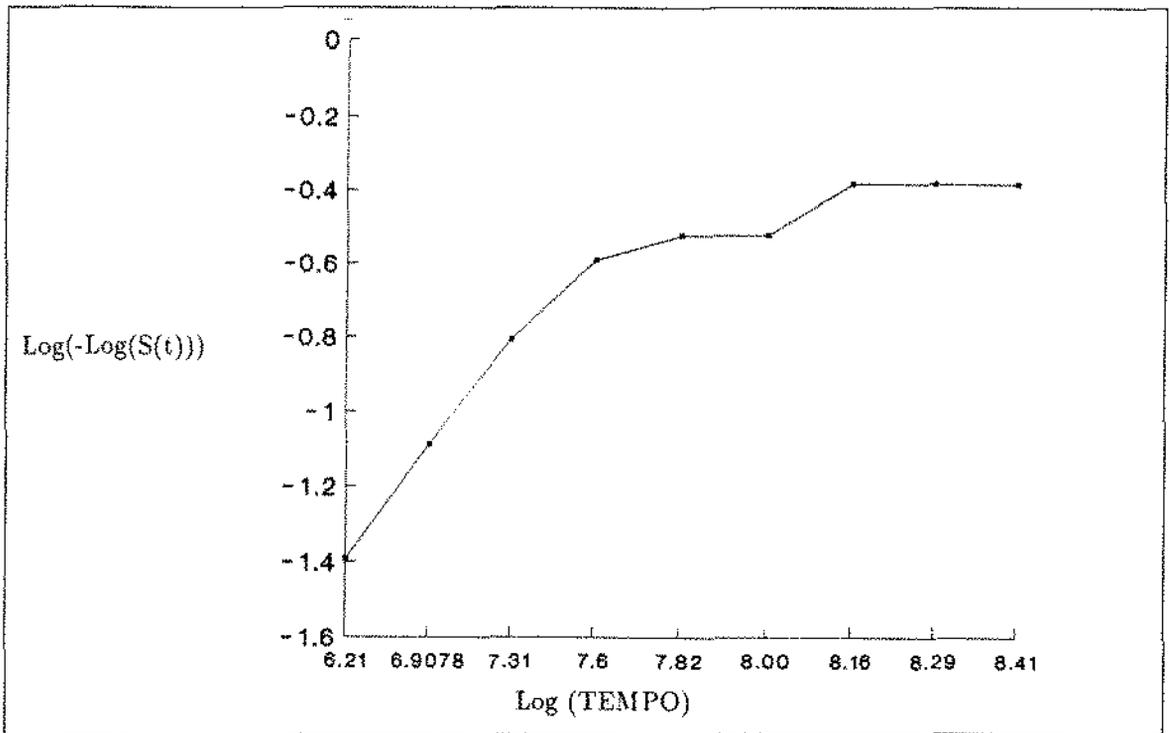


Figura 4.4 - Logaritmo do menos logaritmo da taxa de sobrevivência dos pacientes com o logaritmo do tempo.

A Figura 4.3 mostra que a relação entre menos o logaritmo da taxa de sobrevivência e o tempo de sobrevivência não é linear, portanto, a distribuição dos tempos não parece ser exponencial. Com relação a Figura 4.4, também se nota que a relação não é linear excluindo assim a possibilidade de associar a distribuição Weibull para os tempos de sobrevivência.

Tabela 4.4 - Ajuste dos modelos de Weibull e exponencial.

VARIÁVEIS	WEIBULL			EXPONENCIAL		
	$\hat{\beta}$	Erro Padrão	P-value	$\hat{\beta}$	Erro Padrão	P-value
INTERCEPTO	8.8323	0.3317	0.0001	8.1613	0.1414	0.0000
ESCALAR	1.8371	0.2284		1	0	

O valor do teste multiplicador de Lagrange sob a hipótese H_0 : escalar = 1 é 17.6685 (p-value = 0.0001). Portanto, rejeita-se a hipótese, i.e., os tempos não possuem distribuição exponencial.

A4. Relação entre variáveis independentes e status de sobrevivência.

Tabela 4.5 - Sumário das estatísticas.

DISCRETAS			CONTINUAS		
VARIÁVEL	χ^2	p-value	VARIÁVEL	χ^2	p-value
SEXO	2.077	0.150	IDADE	0.571	0.450
COR	1.993	0.574	THIST	3.152	0.076 (*)
CF1	29.069	0.000 (*)	FIGADO	12.452	0.000 (*)
CF2	87.672	0.000 (*)	TP	0.370	0.543
INTFIBVD	9.250	0.010 (*)	EOSINOFI	3.138	0.076 (*)
INTFIBVE	4.078	0.130	ECOVE	0.017	0.898
DISPNEIA	0.910	0.340	ECOAE	0.002	0.969
EDEMA	7.745	0.005 (*)	ECOVD	0.055	0.815
ASCITE	16.267	0.000 (*)	ECOFE	1.932	0.165
DORPREC	15.625	0.000 (*)	CATAD	3.310	0.069
SMITRAL	0.129	0.720	CATVDPS	0.162	0.687
STRICUP	0.347	0.556	CATVDPDI	8.261	0.004 (*)
ESTALIDO	0.844	0.358	CATVDPDF	2.581	0.108
RESSECVE	1.271	0.260	CATTPPS	3.366	0.067
RESSECVD	0.315	0.574	CATTPPD	4.382	0.036 (*)
PROTESEM	0.752	0.386	CATCP	2.048	0.152
PROTESET	0.864	0.353	CATVEPS	0.179	0.672
PLASTICA	6.049	0.014 (*)	CATVEPDI	1.878	0.171
RESSECPO	9.164	0.002 (*)	CATVEPDF	2.191	0.139
FA	4.511	0.034 (*)	CATAOPS	0.298	0.585
SAD	0.003	0.953	CATAOPD	1.718	0.190
SVD	0.263	0.608			
SAE	0.027	0.870			
SVE	0.536	0.464			
BRD	0.666	0.415			
BDAS	0.014	0.907			
BRE	1.891	0.169			
AINATIVA	3.921	0.048 (*)			
ALTT	0.213	0.645			
ECOIM	2.451	0.117			
ECOIT	0.096	0.756			
CATIM	3.039	0.081			
CATIT	3.647	0.056 (*)			
CALCIF	0.538	0.463			

Nota: (*)- variáveis cujo teste de igualdade de proporção ou teste log-rank foram rejeitadas ao nível de significância < 0.05 .

As variáveis tempo de história em anos (THIST), eosinofia (EOSINOFI), pressão do átrio direito (CATAD) e pressão sistólica do tronco pulmonar (CATTPPS) foram investigadas utilizando-se também o teste log-rank. Este, por sua vez, mostrou-se significativo para as

variáveis THIST ($\chi^2 = 3.5934$, p-value = 0.0580) e EOSINOFI ($\chi^2=3.7741$, p-value = 0.0521).

A5. Relação entre tratamento e variáveis.

Tabela 4.6 - Sumário das estatísticas para a variável TEMPO nos tratamentos.

		N	MÉDIA	MEDIANA	d.p.	MÁX.	MÍN.
CIRÚRGICO	Falha	60	1562.83	1262	1131.62	4869	116
	Censura	21	699.38	348	1019.51	4563	13
CLÍNICO	Falha	41	1177.44	774	1222.16	4464	8
	Censura	29	634.86	229	798.35	3054	3

Para os pacientes cirúrgicos também se fez uma análise descritiva para o tempo de espera, i.e.

$$\text{Tempo Espera} = \text{Data da Cirúrgia} - \text{Data da primeira consulta} = T5,$$

com intuito de verificar o tempo mediano de espera. A tabela abaixo mostra estes resultados.

Tabela 4.7 - Sumário das estatísticas para a variável TEMPO DE ESPERA.

VARIÁVEL	N	Média	Mediana	d.p	Máximo	Mínimo
TEMPO DE ESPERA	81	349.90	61	794.01	4557	0

Tabela 4.8 - Sumário das estatísticas para as variáveis contínuas considerando tratamentos.

VARIÁVEL	CIRÚRGICO						CLÍNICO					
	N	\bar{x}	MED	d.p	MAX	MIN	N	\bar{x}	MED	d.p	MAX	MIN
THIST	71	3.8	2	3.62	15	1	65	4.8	3	4.68	20	1
FIGADO	69	3.9	4	2.75	13	0	66	3.4	3	3.59	12	0
EOSINOFI	66	5.3	2	8.70	57	0	42	6.2	2	14.94	78	0
CATVDPDI	49	5.6	5	5.09	20	0	52	6.6	6	6.15	24	0
CATTPPD	57	22.1	19	10.12	54	8	54	24.9	23	12.03	58	6

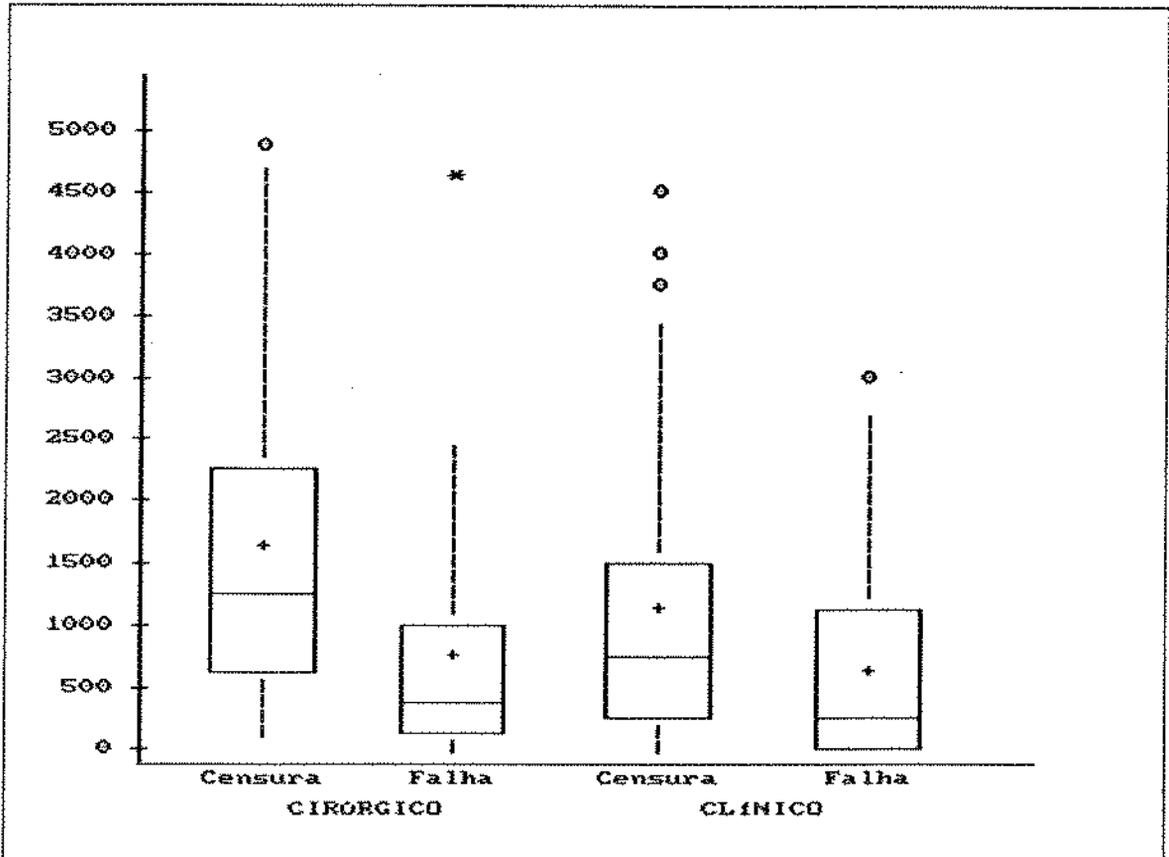


Figura 4.5 - Desenho esquemático correspondente aos tempos de censura e falha dos pacientes com endomiocardiopatia, nos tratamentos cirúrgico e clínico .

O tempo médio de sobrevivência e o tempo médio de falha no grupo cirúrgico é maior do que no grupo clínico, nota-se também que a média se distancia da mediana, confirmando a assimetria da distribuição dos tempos.

A6. Estimação da função de sobrevivência nos tratamentos.

Tabela 4.9 - Tabela de vida para pacientes com endomiocardiopatia no tratamento cirúrgico e clínico.

INTERVALO	CIRÚRGICO					CLÍNICO					
	TEMPO(t_i)	n'_i	d_i	$w_i + l_i$	n_i	$\hat{S}(t_i)$	n'_i	d_i	$w_i + l_i$	n_i	$\hat{S}(t_i)$
0 - 500	500	81	12	14	74.0	1.0000	70	18	15	62.5	1.0000
500 - 1000	1000	55	4	7	51.5	0.8378	37	3	11	31.5	0.7120
1000 - 1500	1500	44	3	13	37.5	0.7728	23	3	5	20.5	0.6442
1500 - 2000	2000	28	1	7	24.5	0.7109	15	3	1	14.5	0.5499
2000 - 2500	2500	20	0	8	16.0	0.6819	11	1	3	9.5	0.4361
2500 - 3000	3000	12	0	4	10.0	0.6819	7	0	2	6.0	0.3902
3000 - 3500	3500	8	0	3	6.5	0.6819	5	1	0	5.0	0.3902
3500 - 4000	4000	5	0	2	4.0	0.6819	4	0	2	3.0	0.3122
4000 - 4500	4500	3	0	1	2.5	0.6819	2	0	2	1.0	0.3122
4500 +		2	1	1	1.5	0.6819	-	-	-	-	-

Notas:
 n'_i : número de pacientes vivos no início do intervalo t_i
 d_i : número de pacientes mortos durante o intervalo t_i
 $w_i + l_i$: número de pacientes retirados vivos ou perdidos no intervalo t_i
 n_i : $n'_i - \frac{1}{2} [w_i + l_i]$
 $\hat{S}(t_i)$: proporção acumulada de sobrevivência do início do estudo até o término do intervalo.

O valor da estatística log-rank referente ao teste de igualdade de tratamento é 7.1513 (p-value = 0.0075). Portanto rejeita-se hipótese, i.e., existe efeito de tratamento cirúrgico e clínico na sobrevivência dos pacientes com endomiocardiopatia.

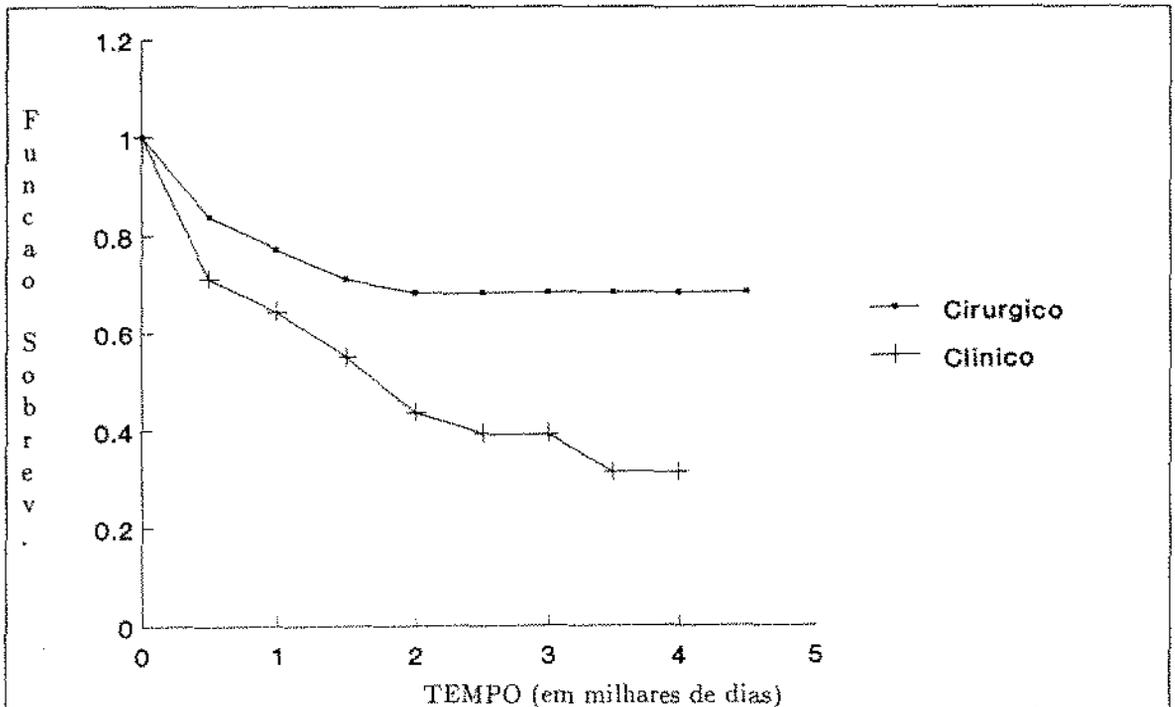


Figura 4.6 - Curva de sobrevivência para os pacientes com endomiocardiopatia no tratamento cirúrgico e clínico.

Observando-se a Figura 4.6, nota-se que os pacientes submetidos a cirurgia apresentam tempo de sobrevivência sempre maior do que aqueles que receberam tratamento clínicos.

A7. Relação entre as variáveis independentes significantes na sobrevivência dos pacientes com tipo de tratamento.

Tabela 4.10 - Sumário das Estatísticas.

VARIÁVEL	χ^2	p-value
CF1	36.687	0.000 (*)
CF2	11.160	0.011 (*)
THIST	2.740	0.098
INTFIBVD	2.027	0.363
EDEMA	5.531	0.019 (*)
ASCITE	0.987	0.321
DORPREC	2.282	0.131
FIGADO	0.025	0.874
PLASTICA	31.019	0.000 (*)
RESSECPO	27.143	0.000 (*)
FA	0.294	0.587
AINATIVA	0.286	0.593
EOSINOFI	0.259	0.611
CATVDPDI	1.093	0.296
CATTPPD	1.517	0.218
CATIT	16.295	0.000 (*)

Nota: (*)- variáveis cujo teste de igualdade de proporção foram rejeitadas ao nível de significância < 0.05 .

B. Ajuste do Modelo.

Uma vez que se verifica na secção anterior (secção A) que a distribuição associada aos tempos de sobrevivência para os pacientes com endomiocardiofibrose não é exponencial, optou-se por ajustar modelos onde esta suposição não fosse necessária. São o modelo de riscos proporcionais de Cox, o modelo logístico ou ainda modelos considerando o tempo particionado em intervalos mutuamente exclusivo com risco constante.

B1. Modelo de Cox

Dos 151 pacientes estudados somente 72 foram consideradas para o ajuste do modelo,

sendo que 79 foram eliminados por não possuir informações completas a respeito das 14 covariáveis selecionadas na secção A.

O modelo de Cox foi ajustado usando o procedimento PHREG do SAS, e a seleção das covariáveis no modelo foi através do método de seleção STEPWISE. As covariáveis que foram incluídas no modelo final foram: classe funcional final (CF2), fibrilação atrial (FA), tipo de tratamento (TRATAM). Assim a função de risco para cada indivíduo é

$$\lambda(t; z) = \lambda_0(t) \exp[\beta_1 CF2_i + \beta_2 FA_i + \beta_3 TRATAM_i], \quad \text{com } i=1, \dots, 72.$$

Os resultados estão sumarizados na tabela abaixo.

Tabela 4.11 - Estimação dos coeficientes de regressão.

VARIÁVEIS	$\hat{\beta}$	Erro Padrão	p-value
CF2	1.5518	0.3082	0.0001
FA	- 0.9909	0.4769	0.0377
TRATAMENTO	- 1.1241	0.5287	0.0335

Para testar a hipótese $H_0: \beta_1 = \beta_2 = \beta_3 = 0$, usou-se a estatística da razão de verossimilhança. O valor encontrado é 43.289 que é uma χ^2 com 3 graus de liberdade (p-value = 0.0001). Portanto, existem evidências de que os fatores de risco classe funcional final (CF2), fibrilação atrial (FA) e tipo de Tratamento (TRATAM) são importantes na sobrevivência dos pacientes com endomiocardiopatia.

Com o objetivo de comparar as taxas de sobrevivências associadas a cada paciente segundo o modelo acima, construiu-se intervalos de 95% de confiança, onde a estimativa de $\hat{S}(t|z)$ é encontrada através do procedimento PHREG do SAS e o erro associado através do programa PEVSTMC.PRG (Apêndice D3.5). Os resultados encontram-se na Tabela C4, Apêndice C. O gráfico a seguir, mostra estes resultados para os 10 primeiros pacientes divididos igualmente para tipos de tratamentos, i.e, 5 pacientes para cada grupo.

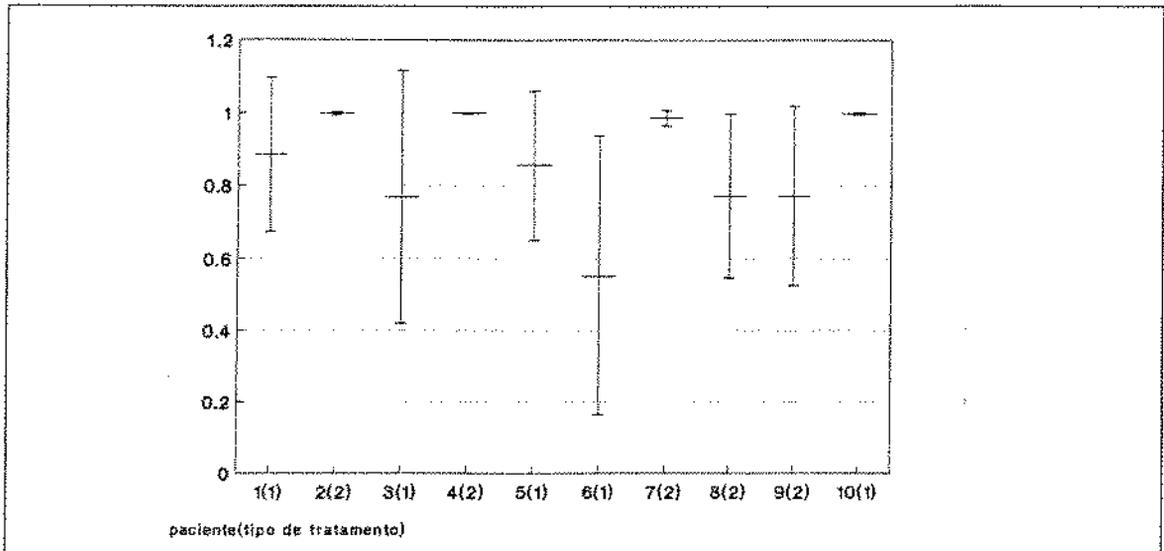


Figura 4.7 - Intervalo de confiança para taxa de sobrevivência estimada no modelo de Cox, para os 10 primeiros pacientes.

Pelo aspecto visual do gráfico acima, nota-se que em média a taxa de sobrevivência para os pacientes que receberam tratamento clínico e/ou possuíam fibrilação atrial são inferiores aos que receberam tratamento cirúrgico e/ou não possuíam fibrilação atrial.

B2. Modelo Logístico.

Neste problema, como se conhece o tempo calendário de entrada e saída (data primeira consulta e data última consulta ou óbito) associado a cada indivíduo no estudo, é possível fixar um tempo τ para o término do mesmo. A figura a seguir, ilustra o possível ensaio.

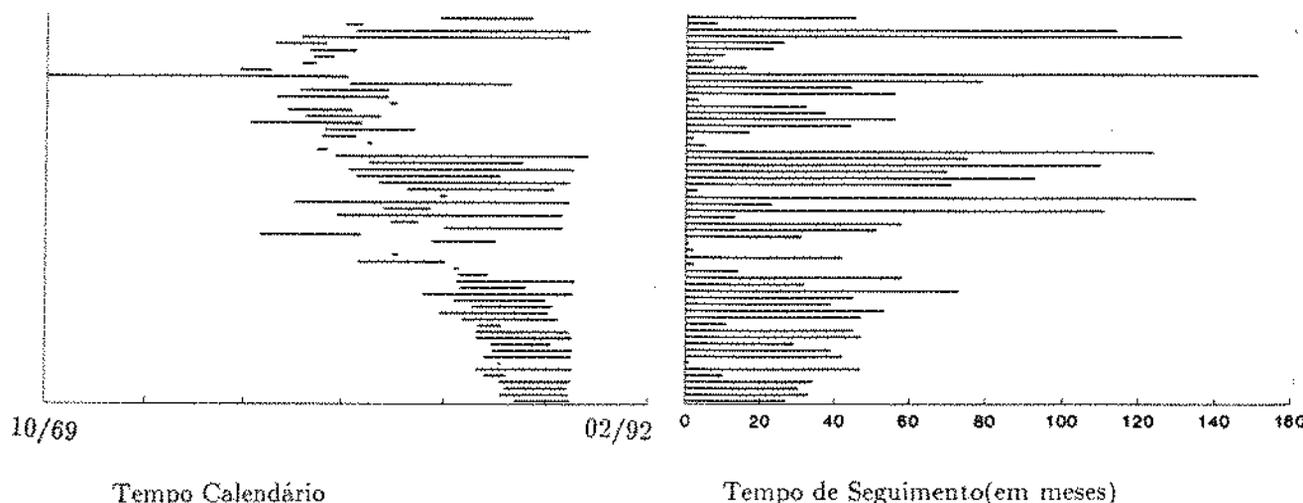


Figura 4.8 - Entrada e saída dos pacientes com endomiocardiopatia no estudo.

Com base na figura acima fixou-se $\tau = 30$ meses, pois como pode ser visto a maioria dos pacientes foram seguidos pelo menos por 30 meses.

Com o intuito de verificar quais fatores prognósticos são importantes no *status* de sobrevivência de pacientes com endomiocardiopatia, ajustou-se o modelo logístico sem levar em consideração os tempos associados a cada paciente. Porém, quando planejou-se o estudo a fim de ajustar o modelo logístico, somente 54 pacientes foram considerados dos 72 que possuíam informações completas com respeito as covariáveis; sendo que 18 destes foram considerados perdidos de observação, i.e, censurados antes de 30 meses. Assim, a variável resposta pode ser definida por

$$\delta_i = \begin{cases} 1 & \text{se o paciente (i) morre antes de 30 meses} \\ 0 & \text{se o paciente (i) sobrevive aos 30 meses} \end{cases}$$

Este modelo, foi ajustado usando o procedimento LOGISTIC do SAS, e o método empregado para selecionar covariáveis foi STEPWISE. Detectou-se apenas a covariável classe funcional final (CF2) como fator de risco importante no *status* da sobrevivência dos pacientes com endomiocardiopatia. Portanto, o modelo ajustado é

$$\log \frac{1 - S(\tau | z)}{S(\tau | z)} = \beta_0 + \beta_1 \text{CF2}_i$$

Os resultados estão sumarizados na tabela seguinte.

Tabela 4.12 - Estimação dos coeficientes de regressão.

VARIÁVEIS	$\hat{\beta}$	Erro Padrão	p-value
INTERCEPTO	6.7233	1.8184	0.0002
CF2	- 1.8391	0.5150	0.0004

Este modelo mostrou-se pouco recomendado para explicar a sobrevivência de pacientes com endomiocardiopatia, talvez pelo fato do estudo não ter sido planejado como deveria, i.e., pré-fixar o tempo τ no início do estudo. Para poder ajustar o modelo foi necessário estipular este período correndo o risco de perder pacientes (18 pacientes vivos antes de $\tau = 30$ meses) além de atribuir uma resposta para o paciente que poderia não ser condizente com a realidade, i.e., pacientes que morreram após os 30 meses foram considerados como vivos, estes fatores afetam diretamente no ajuste do modelo fazendo com que este somente detectasse uma covariável como fator de risco para tais pacientes.

B3. Modelo Logístico Exponencial por Partes.

Numa segunda etapa investigou-se vários intervalos de tempos através do procedimento LIFEREG do SAS, na busca de intervalos de riscos constantes, encontrando como possíveis partições 0 - 6 meses e 6 - 30 meses. Porém quando planejou-se o estudo para que pudesse ser ajustado o modelo logístico exponencial por partes, o número de pacientes reduziu-se muito tornando impossível o ajuste do modelo, pois o método iterativo empregado para encontrar as estimativas para o vetor dos parâmetros não atingiu convergência.

B4. Modelo Exponencial por Partes.

Para o ajuste do modelo exponencial por partes foram considerados 151 pacientes com endomiocardiopatia e os fatores de interesse fornecidos pelo ajuste do modelo de Cox

1. Tipo de tratamento com 2 categorias: cirúrgico e clínico ($I_1 = 2$);
2. Classe funcional final (CF2) com 4 categorias: Moderada, Leve, Alta e Intensa ($I_2 = 4$);
3. Fibrilação atrial (FA) com 2 categorias: Sim e Não ($I_3 = 2$).

O período de seguimento foi dividido em 10 intervalos. Para o ajuste do modelo exponencial particionado levando em conta os três fatores seria necessário um número maior de pacientes, pois só assim seria possível construir tabelas de vida referentes as combinações das categorias, porém como dispõem-se de apenas 151 pacientes decidiu-se fazer análise separada, i.e., uma vez que há interesse em verificar se o tipo de tratamento prolonga a vida destes pacientes,

considerou-se num primeiro instante tipo de tratamento (TRATAM) e classe funcional final (CF2) e em seguida tipo de tratamento e fibrilação atrial (FA).

B4.1. Modelo Exponencial por Partes (tipo de tratamento e classe funcional final).

Considerando os fatores tipo de tratamento (TRATAM) e classe funcional final (CF2) construiu-se 8 tabelas de vida(Tabela C5, Apêndice C) referentes à combinação das categorias destas covariáveis, e tabelas de contingências contendo as duas informações necessária para a análise: o número de mortes e o tempo total de exposição dos pacientes em cada combinação dos níveis das covariáveis e dos intervalos de tempo. Nota-se que a informação do tempo individual não é necessário, i.e, T_k é dado por (3.10).

A tabela de contingência considerando as quatro categorias de classe funcional final (CF2) apresentaram muitas caselas iguais a zero, assim optou-se por juntar os níveis 1-2 e 3-4 reduzindo-se de quatro para duas classes funcionais, $l_2 = 2$ (Tabela C6, Apêndice C). Numa primeira análise usou-se os dados como mostra a Tabela C6, i.e, considerando os 10 intervalos de tempo, e ajustou-se um modelo com efeitos principais do tipo de tratamento (TRATAM), classe funcional final (CF2) e tempo de sobrevivência. Deste ajuste verificou-se que os efeitos relacionados aos intervalos de tempos sugerem uma redução de dez para seis intervalos. A divisão correspondente aos novos intervalos é 0-1 mês, 1-3 meses, 3-6 meses, 6 meses- 1 ano, 1-2 anos e ≥ 2 anos, assim os dados da Tabela C6 são reduzidos a uma tabela (2x2x6). Portanto, o modelo reduzido, contendo o efeito principal do tempo e os efeitos principais do tipo de tratamento e classe funcional final é dado por:

$$\log \lambda(t; \mathbf{x}) = \mu + \mu_{0(I_k)} + \sum_{j=1}^{l_1} \mu_{1(j)} + \sum_{j=1}^{l_2} \mu_{2(j)}$$

com as restrições $\sum_{k=1}^K \mu_{0(I_k)} = \sum_{j=1}^{l_1} \mu_{1(j)} = \sum_{j=1}^{l_2} \mu_{2(j)} = 0$, $k= 1, \dots, 6$, $l_1=2$ e $l_2 = 2$ onde:

μ : média geral da tabela;

$\mu_{0(I_k)}$: efeito principal associado ao intervalo de tempo I_k na sobrevivência dos pacientes com endomiocardiofibrose;

$\mu_{1(j)}$: efeito principal associado ao tipo de tratamento na sobrevivência dos pacientes com endomiocardiofibrose.

$\mu_{2(j)}$: efeito principal associado a classe funcional final na sobrevivência dos pacientes com endomiocardiofibrose.

Para o ajuste do modelo usou-se o programa PAMEPTV.PRG (Apêndice D3.10). Os resultados estão apresentados na tabela abaixo.

Tabela 4.13 - Estimativas de máximo verossimilhança dos parâmetros e seus respectivos erros padrão.

PARÂMETRO	Estimativa	Erro Padrão	P-value
μ	- 7.7036	0.2025	0.0000
$\mu_{0(1)}$	1.0357	0.3276	0.0016
$\mu_{0(2)}$	0.5168	0.3274	0.1144
$\mu_{0(3)}$	0.1574	0.3449	0.6480
$\mu_{0(4)}$	- 0.9388	0.4363	0.0314
$\mu_{0(5)}$	- 0.7916	0.3452	0.0218
* $\mu_{0(6)}$	0.0205	-	-
$\mu_{1(1)}$	- 0.1208	0.1462	0.4086
* $\mu_{1(2)}$	0.1208	0.1462	0.4086
$\mu_{2(1)}$	- 1.3409	0.1958	0.0000
* $\mu_{2(2)}$	1.3409	0.1958	0.0000

Ajuste do modelo $Q_L = 22.9721$ (p-value = 0.1145)

Nota: (*) - são valores obtidos sob as restrições do modelo.

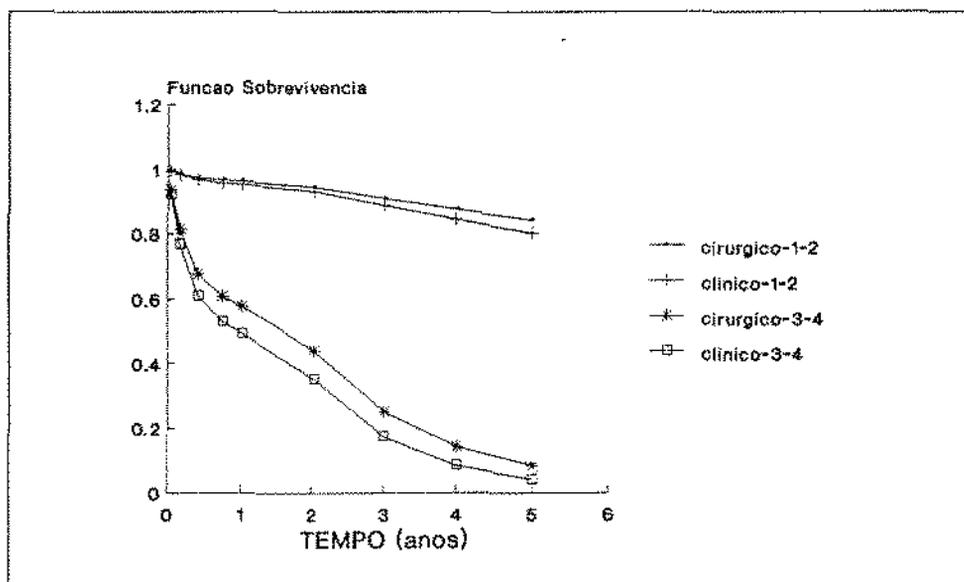


Figura 4.9 - Ajuste das curvas de sobrevivência para pacientes com endomiocardiopatia considerando os fatores de risco tipo de tratamento e classe funcional final.

A Figura 4.9 mostra a curva da sobrevivência estimada, considerando os 4 grupos, i.e., para as combinações dos níveis do tipo de tratamento (TRATAM) e classe funcional final (CF2)

ao longo do tempo. Nota-se apenas que os pacientes que possuem classe funcional final 1-2 parecem ter taxas de sobrevivência superior aos pacientes que possuem classe funcional final 3-4 (p-value = 0.0000).

A fim de verificar o efeito do tipo de tratamento dentro dos níveis de classe funcional ajustou-se um modelo hierárquico dado por:

$$\log \lambda(t; \mathbf{x}) = \mu + \mu_{0(I_k)} + \sum_{j=1}^{l_2} \mu_{2(j)} + \sum_{l=1}^{l_2} \sum_{j=1}^{l_1} \mu_{1(j)}^{(2(l))}$$

com as restrições $\sum_{k=1}^K \mu_{0(I_k)} = \sum_{j=1}^{l_2} \mu_{2(j)} = \sum_{l=1}^{l_2} \sum_{j=1}^{l_1} \mu_{1(j)}^{(2(l))} = 0, k= 1, \dots, 6, l_1=2$ e $l_2 = 2$ onde:

μ : média geral da tabela;

$\mu_{0(I_k)}$: efeito principal associado ao intervalo de tempo I_k na sobrevivência dos pacientes com endomiocardiofibrose;

$\mu_{2(j)}$: efeito principal associado a classe funcional final na sobrevivência dos pacientes com endomiocardiofibrose.

$\mu_{1(j)}^{(2(l))}$: efeito do tipo de tratamento dentro do l -ésimo nível de classe funcional final na sobrevivência dos pacientes com endomiocardiofibrose.

Os resultados estão sumarizados na tabela abaixo.

Tabela 4.14 - Estimativas de máximo verossimilhança dos parâmetros e seus respectivos erros padrão.

PARÂMETRO	Estimativa	Erro Padrão	P-value
μ	- 7.6915	0.2025	0.0000
$\mu_{0(1)}$	1.0355	0.3276	0.0016
$\mu_{0(2)}$	0.5175	0.3274	0.1144
$\mu_{0(3)}$	0.1597	0.3449	0.6480
$\mu_{0(4)}$	- 0.9396	0.4363	0.0314
$\mu_{0(5)}$	- 0.7942	0.3452	0.0218
* $\mu_{0(6)}$	0.0211	-	-
$\mu_{2(1)}$	- 1.3357	0.1940	0.0000
* $\mu_{2(2)}$	1.3357	0.1940	0.0000
$\mu_{1(1)}^{(2(1))}$	- 0.3081	0.3536	0.3836
* $\mu_{1(2)}^{(2(1))}$	0.3081	0.3536	0.3836
$\mu_{1(1)}^{(2(2))}$	- 0.0833	0.1588	0.5999
* $\mu_{1(2)}^{(2(2))}$	0.0833	0.1588	0.5999

Ajuste do modelo $Q_L = 22.6374$

(p-value = 0.0921)

Nota: (*) - são valores obtidos sob as restrições do modelo.

As taxas de sobrevivência não diferem quanto ao tipo de tratamento que o paciente recebe dentre os que possuem classe funcional final 1-2 (p-value=0.3836) e nem dentre os que possuem classe funcional 3-4 (p-value = 0.5999)

B4.2. Modelo Exponencial por Partes (tipo de tratamento e fibrilação atrial).

Nas Tabelas C7 e C8 (Apêndice C) encontram-se as 4 tabelas de vidas, referentes a combinações das categorias dos fatores tipo de tratamento (TRATAM) e fibrilação atrial (FA) e as tabelas de contingência contendo o número de mortes e o tempo total de exposição, respectivamente. O procedimento empregado foi o mesmo descrito em B4.1. Assim, o modelo reduzido, contendo o efeito principal do tempo, considerando os mesmos seis intervalos do modelo anterior, e os efeitos principais do tipo de tratamento e fibrilação atrial é

$$\log \lambda(t; x) = \mu + \mu_{0(I_k)} + \sum_{j=1}^{l_1} \mu_{1(j)} + \sum_{j=1}^{l_3} \mu_{3(j)}$$

com as restrições $\sum_{k=1}^K \mu_{0(I_k)} = \sum_{j=1}^{l_1} \mu_{1(j)} = \sum_{j=1}^{l_3} \mu_{3(j)} = 0$, $k = 1, \dots, 6$, $l_1 = 2$ e $l_3 = 2$ onde:

μ : média geral da tabela;

$\mu_{0(I_k)}$: efeito principal associado ao intervalo de tempo I_k na sobrevivência dos pacientes com endomiocardiofibrose;

$\mu_{1(j)}$: efeito principal associado ao tipo de tratamento na sobrevivência dos pacientes com endomiocardiofibrose.

$\mu_{3(j)}$: efeito principal associado a fibrilação atrial na sobrevivência dos pacientes com endomiocardiofibrose.

Para o ajuste do modelo usou-se o programa PAMEPTV.PRG (Apêndice D3.10). Os resultados estão apresentados na tabela seguinte.

Tabela 4.15 - Estimativas de máximo verossimilhança dos parâmetros e seus respectivos erros padrão.

PARÂMETRO	Estimativa	Erro Padrão	P-value
μ	- 7.3692	0.1559	0.0000
$\mu_{0(1)}$	1.1433	0.3280	0.0004
$\mu_{0(2)}$	0.5522	0.3276	0.0917
$\mu_{0(3)}$	0.1525	0.3448	0.6581
$\mu_{0(4)}$	- 0.9506	0.4363	0.0293
$\mu_{0(5)}$	- 0.8866	0.3452	0.0101
* $\mu_{0(6)}$	- 0.0108	-	-
$\mu_{1(1)}$	- 0.3240	0.1438	0.0242
* $\mu_{1(2)}$	0.3240	0.1438	0.0242
$\mu_{3(1)}$	0.3857	0.1424	0.0067
* $\mu_{3(2)}$	- 0.3857	0.1424	0.0067

Ajuste do modelo $Q_L = 32.4275$ (p-value = 0.0087)

Nota: (*) - são valores obtidos sob as restrições do modelo.

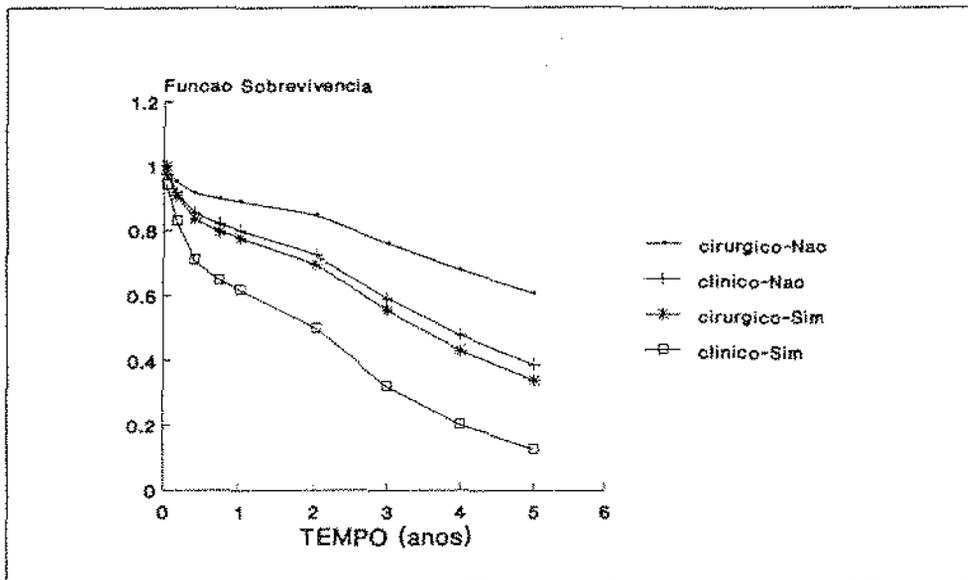


Figura 4.10 - Ajuste das curvas de sobrevivência para pacientes com endomiocardiopatia considerando os fatores de risco tipo de tratamento e fibrilação atrial.

Olhando a figura acima, observa-se que as taxas de sobrevivência para os pacientes no tratamento cirúrgico parecem ser superiores ao tratamento clínico, tanto na presença ou ausência de fibrilação atrial. Então afim de verificar o efeito do tipo de tratamento dentro dos níveis de

fibrilação atrial, ajustou-se o seguinte modelo hierárquico

$$\log \lambda(t; \mathbf{x}) = \mu + \mu_{0(I_k)} + \sum_{j=1}^{l_3} \mu_{3(j)} + \sum_{l=1}^{l_3} \sum_{j=1}^{l_1} \mu_{1(j)}(3(l))$$

com as restrições $\sum_{k=1}^K \mu_{0(I_k)} = \sum_{j=1}^{l_3} \mu_{3(j)} = \sum_{l=1}^{l_3} \sum_{j=1}^{l_1} \mu_{1(j)}(3(l)) = 0$, $k= 1, \dots, 6$, $l_1=2$ e $l_3 = 2$ onde:

μ : média geral da tabela;

$\mu_{0(I_k)}$: efeito principal associado ao intervalo de tempo I_k na sobrevivência dos pacientes com endomiocardiofibrose;

$\mu_{3(j)}$: efeito principal associado a fibrilação atrial na sobrevivência dos pacientes com endomiocardiofibrose.

$\mu_{1(j)}(3(l))$: efeito do tipo de tratamento dentro do l -ésimo nível de fibrilação atrial na sobrevivência dos pacientes com endomiocardiofibrose.

Os resultados estão sumarizados na tabela abaixo.

Tabela 4.16 - Estimativas de máximo verossimilhança dos parâmetros e seus respectivos erros padrão.

PARÂMETRO	Estimativa	Erro Padrão	P-value
μ	- 7.4003	0.1608	0.0000
$\mu_{0(1)}$	1.1317	0.3280	0.0005
$\mu_{0(2)}$	0.5384	0.3276	0.1003
$\mu_{0(3)}$	0.1433	0.3447	0.6776
$\mu_{0(4)}$	- 0.9454	0.4363	0.0303
$\mu_{0(5)}$	- 0.8730	0.3452	0.0114
* $\mu_{0(6)}$	0.0050	-	-
$\mu_{3(1)}$	0.3452	0.1491	0.0206
* $\mu_{3(2)}$	- 0.3452	0.1491	0.0206
$\mu_{1(1)}(3(1))$	- 0.5797	0.2276	0.0108
* $\mu_{1(2)}(3(1))$	0.5797	0.2276	0.0108
$\mu_{1(1)}(3(2))$	- 0.1245	0.1926	0.5181
* $\mu_{1(2)}(3(2))$	0.1245	0.1926	0.5181
Ajuste do modelo $Q_L = 30.0148$		(p-value = 0.0118)	

Nota: (*) - são valores obtidos sob as restrições do modelo.

Com base nestes resultados e na Figura 4.10 conclui-se que pacientes que não tem fibrilação atrial possuem sobrevivência superior aos pacientes que tem fibrilação atrial (p-value =0.0067).

Os pacientes que receberam tratamento cirúrgico possuem taxa de sobrevivência melhores que os pacientes que receberam tratamento clínico, *significante* dentre os que tem fibrilação atrial ($p\text{-value} = 0.0108$) mas não *significantes* dentre os que não tem fibrilação atrial ($p\text{-value} = 0.5181$).

CONCLUSÃO DA TESE.

Neste trabalho, procurou-se discutir dois métodos de regressão alternativos aos métodos comumente utilizados na análise de sobrevivência que são a regressão de riscos proporcionais de Cox [1972] e a regressão exponencial de Glasser [1967]. Assim, foram estudados sob o aspecto teórico e aplicado os métodos de regressão logístico, considerando diversas situações quando a variável resposta é dicotômica (falha ou não falha) e o método de regressão exponencial por partes, quando os dados de sobrevivência estão estruturados na forma de tabelas de vida com covariáveis categóricas e contínuas.

Foram desenvolvidos programas computacionais no módulo IML do SAS, com o intuito de aplicar os métodos de regressão alternativos aqui sugeridos. Também foram desenvolvidos programas complementares para encontrar estimativa da função de sobrevivência e seus respectivos erros quando necessário, uma vez que nos programas existentes não são fornecidos.

As aplicações numéricas destes métodos foram realizadas usando-se um conjunto de dados da literatura e outro de dados reais. A adequação do uso do modelo logístico e exponencial por partes foi prejudicado pela adaptação a posteriori destes modelos, no que se refere ao tempo de seguimento no modelo logístico e tamanho reduzido da amostra no modelo exponencial por partes. Esse prejuízo deve ser evitado com um planejamento adequado do experimento, tanto na sua condução, no tamanho da amostra requerido e, na definição a priori dos métodos estatísticos para a análise.

APÊNDICE A

TEOREMAS

Neste apêndice são apresentados alguns teoremas necessários para o desenvolvimento da teoria da secção 1.6 relativo à normalidade assintótica dos estimadores da variância da função de sobrevivência.

Teorema A1: Sejam $\{X_n\}_{n \geq 1}$ uma seqüência de variáveis aleatórias, $\{a_n\}_{n \geq 1}$ uma seqüência de números reais positivos com $a_n \rightarrow 0$ quando $n \rightarrow \infty$, e f uma função real de variável real derivável até a ordem k em um intervalo que contém um ponto a . Se $X_n - a = o_p(a_n)$, então

$$f(X_n) = \sum_{j=0}^k \frac{f^{(j)}(a)}{j!} (X_n - a)^j + o_p(a_n^k).$$

(veja prova em Leite et al [1990], Cap. 2 página 44).

Teorema A2: Se uma seqüência de variáveis aleatórias $\{X_n\}_{n \geq 1}$ converge em distribuição para uma variável aleatória X , então $X_n = o_p(1)$. (veja prova em Leite et al [1990], Cap. 2 página 58)

Teorema A3: (Teorema de Slutsky) Seja X_1, X_2, \dots e Y_1, Y_2, \dots variáveis aleatórias tais que $X_n \xrightarrow{\mathcal{D}} X$ e $Y_n \xrightarrow{\mathcal{D}} c$, onde c é uma constante. Então

i. $X_n + Y_n \xrightarrow{\mathcal{D}} X + c;$

ii. $X_n Y_n \xrightarrow{\mathcal{D}} Xc;$

iii. se $c \neq 0$, $\frac{X_n}{Y_n} \xrightarrow{\mathcal{D}} \frac{X}{c}.$

(veja prova em Leite et al [1990], Cap. 2 página 63)

APÊNDICE B

CONJUNTO DE DADOS.

Neste apêndice encontram-se os conjuntos de dados analisados nesta tese. O primeiro deles, exibido na Tabela B1, é proveniente de um estudo sobre myeloma múltiplo analisado por Krall, Uthoff e Harley [1975]. O segundo conjunto de dados, exibido na Tabela B2, são dados de um estudo real sobre endomiocardiofibrose realizado no Instituto do Coração de São Paulo pelo Dr. Pereira Barreto e analisado por Rita Helena Antonelli Cardoso [1991].

Tabela B1: Dados referentes aos 65 pacientes com mieloma múltiplo.

Pac	Tempo	δ	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	Z15	Z16
01	1.25	1	2.2175	9.4	1	0	67	1	3.6628	1	1.9542	0	0	12	2	11	7	10
02	1.25	1	1.9395	12.0	1	1	38	1	3.9868	1	1.9542	0	0	20	1	6	3	18
03	2.00	1	1.5185	9.8	1	1	81	1	3.8751	1	2.0000	0	0	2	1	11	8	15
04	2.00	1	1.7482	11.3	0	0	75	1	3.8062	1	1.2553	13	52	0	2	9	6	12
05	2.00	1	1.3010	5.1	0	0	57	1	3.7243	1	2.0000	0	0	3	1	10	6	9
06	3.00	1	1.5441	6.7	1	1	46	2	4.4757	0	1.9345	3	5	12	2	9	3	10
07	4.00	0	1.9542	10.2	1	0	59	1	4.0453	0	0.7782	21	51	12	1	7	3	10
08	4.00	0	1.9243	10.0	1	1	49	2	3.9590	0	1.6232	21	24	0	2	17	12	13
09	5.00	1	2.2355	10.1	1	0	50	2	4.9542	1	1.6628	8	27	4	1	6	4	9
10	5.00	1	1.6812	6.5	1	0	74	1	3.7324	0	1.7324	1	25	5	2	10	7	9
11	6.00	1	1.3617	9.0	1	1	77	1	3.5441	0	1.4624	7	51	1	2	7	5	8
11	6.00	1	2.1139	10.2	0	0	70	2	3.5441	1	1.3617	6	68	1	2	9	7	8
13	6.00	1	1.1139	9.7	1	0	60	1	3.5185	1	1.3979	11	30	0	2	11	7	10
14	6.00	1	1.4150	10.4	1	0	67	2	3.9294	1	1.6902	10	33	0	2	10	6	8
15	7.00	1	1.9777	9.5	1	0	48	1	3.3617	1	1.5682	9	41	5	2	9	3	10
16	7.00	1	1.0414	5.1	0	0	61	2	3.7324	1	2.0000	0	0	1	2	8	4	10
17	7.00	1	1.1761	11.4	1	0	53	2	3.7243	1	1.5185	13	42	1	1	9	6	13
18	7.00	0	1.1139	12.4	1	0	48	2	3.7993	1	1.8573	1	11	0	2	6	3	10
19	7.00	0	1.5315	10.2	1	1	81	1	3.5911	0	1.8808	1	16	0	2	12	10	11
20	8.00	0	1.0792	9.9	1	0	57	2	3.8325	1	1.6532	1	40	0	2	8	4	8
21	9.00	1	1.7243	8.2	1	0	55	1	3.7993	1	1.7404	4	31	0	2	9	6	12
22	11.00	1	1.1139	14.0	1	0	61	1	3.8808	1	1.2788	8	49	0	2	8	5	10
23	11.00	1	1.2304	12.0	1	0	43	1	3.7709	1	1.1761	7	48	1	1	12	8	9
24	11.00	1	1.3010	13.2	1	0	65	1	3.7993	1	1.8195	4	21	1	2	8	4	10
25	11.00	1	1.5682	7.5	1	0	70	1	3.8865	0	1.6721	5	35	0	2	7	5	12
26	11.00	1	1.0792	9.6	1	0	51	2	3.5051	1	1.9031	5	10	0	2	7	3	9
27	11.00	0	1.6128	14.0	1	0	60	1	3.7324	1	1.8451	0	0	3	1	6	2	9
28	12.00	0	1.1461	11.6	1	0	46	2	3.6435	0	1.1461	7	47	0	2	7	3	7
29	12.00	0	1.3979	8.8	1	0	66	2	3.8388	1	1.3617	8	50	0	2	9	5	9
30	13.00	1	0.7762	5.5	0	1	60	2	3.5798	1	1.3979	12	55	2	2	10	8	10
31	13.00	0	1.6628	4.9	0	0	71	2	3.6435	0	1.7924	2	21	0	2	10	11	9
32	14.00	1	1.3979	14.6	1	1	66	1	3.7243	1	1.2553	7	51	2	1	9	4	10
33	15.00	1	1.6021	10.6	1	0	70	1	3.6902	1	1.4314	16	46	0	2	9	5	11
34	16.00	1	1.3424	9.0	1	0	48	1	3.9345	1	2.0000	0	0	0	2	8	4	10
35	16.00	1	1.3222	8.8	1	0	62	2	3.6990	1	0.6990	5	41	17	2	8	4	10
36	16.00	0	1.1461	13.0	1	0	55	1	3.8573	0	0.9031	13	53	0	2	9	5	9
37	17.00	1	1.2304	10.0	1	0	53	1	3.8808	1	1.4472	9	50	4	2	10	7	9
38	17.00	1	1.5911	11.2	1	1	68	1	3.4314	0	1.6128	1	46	1	2	9	6	10
39	18.00	1	1.4472	7.5	1	0	65	2	3.5682	0	0.9031	17	46	7	2	6	4	8
40	19.00	1	1.0792	14.4	1	0	51	1	3.9191	1	2.0000	0	0	6	2	13	7	15
41	19.00	1	1.2553	7.5	0	0	60	2	3.7924	1	1.9294	2	10	5	1	6	3	9
42	19.00	0	1.3222	13.0	1	0	59	2	3.7709	1	2.0000	0	0	1	2	8	4	10
43	19.00	0	1.3222	10.8	1	0	69	2	3.8808	1	1.5185	0	0	0	2	10	7	10
44	24.00	1	1.3010	14.6	1	0	56	2	4.0899	1	0.4771	12	52	0	2	4	3	9
45	25.00	1	1.0000	12.4	1	0	67	1	3.8195	1	1.6435	3	23	0	2	7	3	10
46	26.00	1	1.2304	11.2	1	0	49	2	3.6021	1	2.0000	0	0	27	1	6	3	11
47	28.00	0	1.2304	7.3	1	1	82	2	3.7482	1	1.6721	7	40	0	2	7	4	9
48	32.00	1	1.3222	10.6	1	0	46	1	3.6990	1	1.6335	6	32	1	2	12	8	9
49	35.00	1	1.1139	7.0	0	0	48	1	3.6532	1	1.1761	16	42	4	1	12	9	10

Tabela B1: (Continuação).

Pac	Tempo δ	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	Z15	Z16	
50	37.00	1	1.6021	11.0	1	0	63	1	3.9542	0	1.2041	10	53	7	1	7	4	9
51	41.00	1	1.0000	10.2	1	0	69	1	3.4771	1	1.4771	5	30	6	1	10	6	10
52	41.00	1	1.1461	5.0	1	0	70	2	3.5185	1	1.3424	18	53	0	2	7	4	9
53	41.00	0	1.7559	12.8	1	0	72	1	3.7243	1	1.4472	2	56	1	1	7	3	9
54	51.00	1	1.5682	7.7	0	0	74	1	3.4150	1	1.0414	10	50	4	1	12	8	13
55	52.00	1	1.0000	10.1	1	0	60	2	3.8573	1	1.6532	1	27	4	1	10	7	10
56	53.00	0	1.1139	12.0	1	0	66	1	3.6128	1	2.0000	0	0	1	2	6	3	11
57	54.00	1	1.2553	9.0	1	0	49	1	3.7243	1	1.6990	8	37	2	1	8	3	10
58	57.00	0	1.2553	12.5	1	0	66	1	3.9685	0	1.9542	0	0	0	2	8	4	11
59	58.00	1	1.2041	12.1	1	0	42	2	3.6990	1	1.5798	9	35	22	1	7	4	10
60	66.00	1	1.4472	6.6	1	0	59	1	3.7853	1	1.8195	4	22	0	2	5	3	9
61	67.00	1	1.3222	12.8	1	0	52	1	3.6435	1	1.0414	24	54	1	1	11	7	10
62	77.00	0	1.0792	14.0	1	0	60	1	3.6812	0	0.9542	4	50	0	2	6	3	12
63	88.00	1	1.1761	10.6	1	0	47	2	3.5563	0	1.7559	22	17	21	1	6	3	9
64	89.00	1	1.3222	14.0	1	1	63	1	3.6532	1	1.6232	7	38	1	1	9	7	9
65	92.00	1	1.4314	11.0	1	0	58	2	4.0755	1	1.4150	12	36	4	1	9	6	11

Fonte: Krall, Uthoff e Harley [1975].

OBS: As siglas referentes as variáveis encontram-se definidas na Tabela 1.1, página 25.

Tabela B2: Dados para os 151 pacientes com Endomiocardiofibrose.

Pac	Z1	Z2	Z3	Z4	Z5	NOME	Z6	Z7	Z8	PRIMEIRA	CIRURGIA	ULTIMA	OBITO	Z9
001	04	04	3	2	-	NJC	49	F	BCA	10/09/80	-	-	12/09/81	16
002	02	03	1	2	0.84	NA	23	F	BCA	09/01/86	05/09/86	31/10/89	-	1
003	02	04	2	1	-	ERM	34	F	BCA	13/04/78	15/06/78	12/08/91	-	-
004	03	04	1	3	0.82	AGM	36	F	BCA	18/03/82	-	09/11/82	-	7
005	04	04	3	3	0.72	MACLB38		F	BCA	07/02/80	-	-	20/01/83	4
006	01	03	1	3	0.77	HAVA	46	F	BCA	30/08/82	28/01/92	14/02/92	-	2
007	02	03	2	2	0.72	MLCJ	50	F	BCA	15/05/80	-	09/05/91	-	10
008	04	04	2	3	0.75	TF	43	F	BCA	13/03/79	-	-	22/05/81	12
009	04	04	2	2	-	IG	42	M	BCA	16/11/78	-	-	27/05/80	5
010	04	04	2	2	-	ALP	35	M	BCA	18/12/80	-	-	22/12/80	6
011	04	04	2	2	-	VGR	22	M	BCA	12/09/80	-	-	16/08/82	3
012	03	03	1	2	-	DFA	37	F	-	27/08/82	-	-	28/12/82	15
013	04	04	3	2	0.78	AMM	38	F	BCA	26/11/80	-	-	15/09/81	7
014	04	04	2	3	-	JCB	48	F	BCA	02/05/80	-	-	17/12/80	16
015	04	04	3	1	-	JML	46	F	BCA	13/09/79	19/11/79	-	06/12/79	10
016	03	04	3	2	-	ZGA	44	F	BCA	21/09/77	08/12/78	-	14/01/79	1
017	04	04	3	2	-	MJF	29	F	MUL	17/03/80	-	-	12/09/80	10
018	03	03	2	2	-	NC	53	M	BCA	15/10/69	07/04/82	-	13/04/82	1
019	02	04	1	3	0.77	MCP	35	F	BCA	25/05/82	15/10/84	08/12/88	-	1
020	04	04	3	2	0.77	LST	51	F	BCA	24/04/80	-	-	19/12/83	20
021	04	04	3	1	-	MAL	21	M	BCA	15/04/79	-	-	12/12/83	3
022	04	04	2	2	-	JSR	37	M	BCA	09/01/84	-	-	26/04/84	1
023	02	04	2	3	0.78	PCV	27	M	BCA	30/10/79	24/11/79	-	20/06/82	2
024	03	03	2	2	-	EDRV	29	F	BCA	16/07/80	25/08/83	-	27/08/83	3
025	04	04	3	2	-	AMS	47	M	BCA	10/03/78	-	-	16/11/82	7
026	04	04	3	2	-	JMA	18	M	BCA	12/12/83	07/02/84	-	08/02/84	1
027	04	04	3	3	0.74	ASSM	32	F	-	06/05/81	23/01/85	-	23/01/85	1
028	04	04	1	2	0.72	HAS	30	F	BCA	12/03/81	16/08/82	-	18/08/82	1
029	04	04	2	2	0.60	CSF	47	F	BCA	28/02/83	04/04/83	-	10/04/83	15
030	04	04	3	2	-	GLS	24	F	BCA	16/01/81	-	26/06/81	-	5
031	02	02	2	1	-	MCS	32	F	BCA	25/06/80	-	03/06/91	-	10
032	02	03	1	3	0.78	RS	60	M	BCA	22/10/81	-	14/02/92	-	3
033	02	03	3	2	0.78	MJAC	49	F	BCA	08/12/80	20/10/89	30/06/91	-	2
034	02	03	1	1	0.75	MCLB	19	F	MUL	20/06/84	-	06/06/86	-	1
035	04	03	2	2	0.75	MLS	34	F	BCA	09/03/83	-	-	10/06/89	1
036	02	04	1	3	0.73	AAA	43	M	BCA	05/05/82	24/02/89	12/07/91	-	12
037	04	04	1	2	0.78	TJMS	29	F	BCA	11/05/84	-	13/06/84	-	14
038	02	04	2	2	0.69	TMF	47	F	BCA	23/03/85	15/09/85	-	06/02/86	2
039	02	04	2	2	0.77	EC	32	M	MUL	14/09/82	10/11/82	16/07/88	-	1
040	02	04	2	1	0.78	AJP	27	F	BCA	02/08/83	13/09/83	02/06/91	-	5
041	01	03	1	2	0.77	VFNL	23	F	-	25/10/84	08/02/85	23/10/90	-	3
042	04	04	3	1	0.70	OFS	22	M	-	25/02/86	29/04/86	-	01/05/86	4
043	02	04	3	1	-	LICS	15	F	MUL	11/10/82	18/02/83	13/09/85	-	3
044	02	02	2	2	0.76	ASC	26	M	BCA	13/06/85	-	-	15/11/85	8
045	01	03	1	1	0.75	VMO	36	F	PTA	09/02/80	31/05/85	30/06/91	-	4
046	01	03	1	3	0.67	JRL	57	M	BCA	30/01/84	24/04/84	15/06/91	-	2
047	02	03	2	2	0.78	AFM	41	F	BCA	26/10/83	09/12/83	09/09/85	-	11
048	01	03	2	3	-	JAGI	42	F	BCA	15/02/84	08/03/84	08/09/88	-	3
049	01	01	3	3	0.71	ABA	57	F	BCA	24/01/84	-	04/07/91	-	3

Tabela B2: (Continuação).

Pac	Z1	Z2	Z3	Z4	Z5	NOME	Z6	Z7	Z8	PRIMEIRA	CIRURGIA	ULTIMA	OBITO	Z9
050	01	04	3	1	0.70	SMS	9	M	MUL	26/09/83	18/11/83	22/03/89	-	1
051	01	04	2	2	0.69	ECFB	36	F	MUL	23/11/81	12/07/83	14/03/91	-	4
052	04	04	3	2	0.62	MJAS	35	F	PTA	03/02/84	-	-	08/03/85	10
053	02	04	2	2	-	ELS	22	M	BCA	07/04/86	09/05/86	06/03/91	-	1
054	04	04	3	3	-	DBS	24	F	MUL	17/08/78	-	-	20/11/82	6
055	02	04	3	2	-	NSL	26	F	MUL	14/12/79	11/01/80	-	20/04/80	1
056	02	04	3	2	-	RRFC	28	F	-	15/06/83	26/07/83	11/08/91	-	-
057	04	04	3	3	-	MNPL	35	F	-	10/04/80	-	-	25/05/80	-
058	04	04	3	2	-	AIA	40	F	-	05/06/81	-	-	09/06/81	2
059	02	04	2	2	-	GMG	38	F	-	14/04/81	28/05/81	11/08/91	-	-
060	02	02	2	1	0.86	OBS	55	M	BCA	21/07/85	-	-	17/05/86	-
061	02	03	1	1	-	VJF	13	M	-	21/10/85	-	05/05/88	-	1
062	02	04	3	3	-	MJS	29	F	BCA	09/05/83	20/07/83	11/05/89	-	2
063	01	03	3	2	0.78	MGFD	22	F	BCA	19/09/84	26/10/84	13/08/89	-	1
064	04	04	1	3	0.69	MLM	39	F	BCA	17/10/86	30/10/86	-	30/10/86	1
065	04	04	2	2	-	NFS	29	F	BCA	29/03/84	22/05/84	-	28/05/84	2
066	03	04	1	3	0.55	CS	56	M	BCA	08/10/82	15/09/83	-	11/04/86	1
067	03	04	1	2	0.37	JBB	40	M	BCA	04/09/79	-	09/05/83	-	1
068	01	01	1	2	-	ARB	64	F	BCA	20/11/84	-	02/04/91	-	1
069	04	04	2	3	-	MCSP	18	M	BCA	09/05/86	-	-	12/05/86	4
070	04	04	3	2	-	MLFP	44	F	BCA	12/06/84	-	-	22/06/84	4
071	04	04	1	2	-	ACV	53	F	-	11/11/85	-	-	01/12/85	1
072	01	01	2	1	0.74	LBS	27	M	-	23/09/83	-	29/11/89	-	5
073	02	03	3	1	0.89	JS	7	M	PTA	30/09/86	-	26/11/86	-	2
074	01	01	2	3	-	RASXF	30	F	BCA	15/10/86	-	31/03/87	-	0
075	02	04	2	2	-	JZAC	22	F	BCA	22/11/86	16/12/86	01/11/88	-	3
076	01	01	1	3	0.70	WMP	22	M	-	03/12/86	-	11/12/86	-	-
077	02	03	3	2	-	LCS	19	F	BCA	14/10/86	20/01/87	01/08/91	-	2
078	02	04	1	1	0.76	OMP	48	F	-	06/08/86	16/09/86	25/06/91	-	8
079	02	02	2	2	0.77	MVS	18	F	BCA	10/12/86	-	25/08/89	-	6
080	02	02	1	2	0.76	LAC	13	F	BCA	25/11/82	-	10/08/90	-	1
081	02	03	3	2	0.69	JND	34	F	-	04/06/85	18/03/87	14/07/91	-	1
082	01	03	2	2	0.71	MPR	48	F	BCA	17/02/87	20/03/87	21/05/91	-	12
083	01	01	1	2	0.83	VJQ	45	F	BCA	25/08/86	-	14/04/88	-	1
084	01	02	1	2	0.71	MCS	22	F	BCA	07/07/81	-	-	16/11/89	5
085	03	03	1	2	0.69	ECT	28	M	-	23/09/86	-	07/06/90	-	1
086	01	02	1	3	0.82	EMSM	35	F	-	03/07/86	-	14/12/88	-	2
087	02	02	3	2	0.70	MAS	40	F	-	03/12/86	-	08/04/87	-	5
088	01	03	1	3	0.72	LMF	38	F	MUL	10/06/87	15/07/87	12/10/90	-	9
089	02	04	1	2	0.71	MJGS	36	F	-	28/02/86	14/07/86	04/07/90	-	4
090	02	03	1	2	0.78	RBS	48	M	BCA	23/01/87	-	15/01/91	-	1
091	03	04	3	2	-	ELS	5	M	MUL	27/07/87	20/08/87	-	01/07/89	2
092	03	04	3	2	0.81	MSP	27	F	BCA	02/09/87	19/10/87	-	15/08/88	2
093	04	04	1	3	-	SMG	38	F	BCA	14/10/87	-	-	23/10/87	1
094	02	03	2	2	0.76	EMS	29	M	-	21/08/87	24/11/87	19/06/91	-	2
095	03	04	1	3	-	ICS	11	F	BCA	12/01/84	01/02/84	-	04/05/88	1
096	02	02	1	2	0.89	FCRM	62	M	BCA	23/11/87	-	-	23/12/87	2
097	01	02	2	2	0.76	IAFF	51	F	-	11/08/87	-	19/08/91	-	6
098	01	02	1	3	0.80	MRR	32	F	BCA	23/01/87	-	16/06/91	-	4
099	02	04	3	2	0.77	NBA	50	F	BCA	07/03/88	04/04/88	-	10/08/90	6
100	01	03	3	1	-	TSS	25	F	BCA	23/02/88	29/03/88	06/06/91	-	5

Tabela B2: (Continuação).

Pac	Z1	Z2	Z3	Z4	Z5	NOME	Z6	Z7	Z8	PRIMEIRA	CIRURGIA	ULTIMA	OBITO	Z9
101	02	04	2	2	0.50	NES	48	M	PTA	10/02/88	-	02/06/88	-	1
102	02	02	1	3	0.75	ABS	48	F	-	14/03/88	-	10/02/90	-	1
103	02	04	1	3	0.77	NRS	41	F	-	11/07/85	-	18/09/91	-	1
104	02	04	2	2	0.72	QD	39	M	BCA	23/03/88	05/05/88	23/05/91	-	4
105	03	03	2	2	0.76	JPC	40	M	BCA	07/04/88	12/07/88	11/08/91	-	3
106	02	03	1	3	0.58	JAO	45	M	-	17/12/87	22/07/88	12/07/91	-	1
107	02	03	2	3	0.52	MCP	52	F	BCA	03/07/88	-	04/08/88	-	1
108	02	03	1	3	0.79	ONS	40	F	-	31/08/87	26/03/90	18/08/91	-	10
109	02	02	1	3	0.75	VPC	54	F	-	04/03/88	-	18/06/91	-	1
110	02	03	3	1	0.76	FJF	60	F	-	16/12/87	-	24/10/88	-	1
111	03	04	3	1	0.68	FSS	39	M	-	09/08/88	13/10/88	21/12/90	-	1
112	04	04	2	2	-	JRMP	10	M	BCA	12/08/87	25/08/87	-	27/08/87	5
113	03	04	3	1	0.73	JBS	43	M	BCA	03/08/88	03/10/88	01/07/91	-	2
114	02	02	2	2	0.40	LMC	22	M	-	26/10/88	-	15/05/91	-	3
115	01	04	2	2	0.72	HJF	52	F	-	26/04/88	27/04/89	12/04/91	-	10
116	01	03	3	3	0.65	LSB	30	F	BCA	13/04/88	11/01/89	24/06/91	-	1
117	03	04	3	3	0.72	IRF	38	F	BCA	03/02/89	05/04/89	10/09/91	-	10
118	02	04	3	3	0.74	MCO	29	F	BCA	22/08/88	19/04/89	07/06/91	-	9
119	02	04	2	2	0.76	ZVJ	49	F	-	16/03/89	01/08/89	22/07/91	-	2
120	01	01	1	3	0.73	MS	35	F	-	29/03/89	-	12/05/91	-	0
121	02	03	3	3	-	MCGXIA1	F	BCA	01/08/89	-	08/11/90	-	-	5
122	02	03	2	3	0.75	AD	16	M	BCA	09/11/87	-	-	05/07/91	5
123	02	03	3	2	0.88	RAS	44	F	-	23/10/89	20/12/89	10/07/91	-	5
124	04	04	2	3	0.52	CSS	47	F	BCA	14/10/89	26/12/89	-	05/01/90	14
125	02	02	1	2	0.77	LLL	26	M	BCA	06/10/89	-	16/11/90	-	1
126	01	03	1	3	0.71	GRB	41	F	-	31/05/79	-	20/08/91	-	1
127	01	04	2	3	0.76	NTMR	57	F	-	23/01/90	12/02/90	16/05/91	-	2
128	01	01	1	2	-	MFSBP	51	F	-	30/04/89	-	27/07/91	-	5
129	02	03	3	3	0.75	MS	11	F	BCA	20/02/90	13/03/90	30/05/91	-	1
130	01	03	1	3	-	TV	58	F	BCA	10/01/90	15/09/91	23/10/91	-	1
131	02	04	3	3	0.59	QMCS	32	F	BCA	07/03/90	24/07/90	24/06/91	-	2
132	02	02	1	3	0.78	HCLF	61	F	-	09/03/90	-	19/08/91	-	1
133	03	03	3	3	0.69	DZG	52	F	BCA	26/01/90	-	-	14/06/90	16
134	02	03	3	2	0.80	MJAC	49	F	-	05/06/90	-	13/08/91	-	9
135	01	01	2	2	0.70	GAB	43	F	-	14/12/89	-	03/07/91	-	1
136	02	04	3	2	0.76	ISS	41	F	-	13/08/90	01/04/91	12/08/91	-	6
137	01	03	1	3	0.84	ATC	38	F	BCA	10/05/90	30/04/91	19/08/91	-	1
138	01	03	2	2	0.79	APD	25	F	MUL	20/08/90	09/01/91	27/05/91	-	3
139	01	03	1	1	0.82	MNNM	35	F	BCA	09/01/91	08/02/91	13/07/91	-	8
140	01	03	3	2	0.78	UBG	48	M	BCA	03/12/90	14/04/91	19/06/91	-	1
141	01	03	3	1	-	GAS	39	F	-	04/07/85	04/07/85	31/12/91	-	-
142	01	03	2	2	-	MESM	61	F	-	31/10/84	29/12/84	31/12/91	-	-
143	01	03	1	3	-	MIC	37	F	-	25/04/85	28/05/85	31/12/91	-	-
144	02	04	1	3	-	AMS	37	M	-	10/10/90	11/11/90	31/12/91	-	-
145	01	04	2	2	-	MLA	59	F	-	20/10/90	27/11/90	30/12/91	-	-
146	01	03	1	2	-	NMA	42	F	-	08/09/90	09/10/90	17/08/91	-	-
147	02	04	1	2	-	MF	34	F	-	05/01/91	12/01/91	17/11/91	-	-
148	01	03	1	3	0.52	PJM	50	M	BCA	17/06/91	03/07/91	10/02/92	-	5
149	01	03	3	1	-	ZRB	22	F	MUL	05/09/91	26/09/91	30/12/91	-	1
150	02	02	1	3	-	AFPF	40	F	BCA	26/07/91	-	20/01/92	-	1
151	03	04	2	2	0.82	MSS	57	F	BCA	06/12/91	-	14/02/92	-	1

Tabela B2: (Continuação).

Pac	Z10	Z11	Z12	Z13	Z14	Z15	Z16	Z17	Z18	Z19	Z20	Z21	Z22	Z23	Z24	Z25	Z26	Z27
001	N	S	S	N	N	N	0	N	N	N	N	N	N	N	N	S	S	S
002	N	S	S	N	S	S	0	N	S	N	S	N	N	N	N	N	S	N
003	N	S	S	N	N	S	3	N	S	S	S	N	N	N	S	N	N	N
004	S	N	N	N	S	N	7	S	N	N	N	N	N	N	N	N	S	S
005	S	S	S	N	N	N	-	S	N	N	N	N	N	N	N	S	S	N
006	S	N	N	S	S	N	0	S	S	N	N	N	S	S	N	N	N	N
007	S	S	S	N	S	N	0	N	N	N	N	N	N	N	N	S	N	S
008	S	S	S	N	N	N	10	N	N	N	N	N	N	N	S	N	S	N
009	S	S	N	N	N	N	7	N	N	N	N	N	N	N	N	N	N	N
010	S	N	N	N	S	N	6	S	N	N	N	N	N	N	S	N	S	N
011	S	S	S	N	N	N	3	N	N	N	N	N	N	N	N	N	S	S
012	S	S	N	N	N	N	5	N	N	N	N	N	N	N	S	N	N	N
013	S	S	S	N	S	S	7	S	N	N	N	N	N	N	S	N	N	N
014	S	S	S	N	S	S	5	S	N	N	N	N	N	N	S	N	N	N
015	S	S	S	N	S	S	6	S	N	S	N	S	N	N	S	N	N	N
016	S	S	S	N	S	S	5	N	S	S	S	S	N	N	N	S	S	S
017	S	S	S	N	S	N	4	S	N	N	N	N	N	N	S	N	S	N
018	N	S	N	S	N	N	0	N	S	N	N	N	N	N	N	N	N	S
019	S	S	S	N	S	S	5	S	S	N	S	N	N	N	N	S	N	S
020	S	S	N	N	S	N	5	S	N	N	N	N	N	N	S	N	N	N
021	S	S	S	N	S	S	10	S	N	N	N	N	N	N	S	N	S	N
022	S	S	S	N	N	N	7	S	N	N	N	N	N	N	S	N	S	N
023	S	S	S	N	S	S	4	S	S	S	S	S	N	N	N	N	S	S
024	S	N	N	N	S	N	1	N	S	S	S	S	N	N	N	S	N	S
025	S	S	S	N	N	N	10	N	N	N	N	N	N	N	N	N	N	S
026	N	N	S	N	N	S	2	N	S	S	S	S	N	N	N	S	S	S
027	N	S	S	N	N	N	4	N	S	S	S	S	N	N	N	S	N	S
028	S	S	S	N	S	N	10	N	S	N	S	N	N	N	N	S	N	S
029	S	S	S	N	N	N	6	S	S	S	S	S	N	N	S	N	N	N
030	S	S	S	N	S	N	4	S	N	N	N	N	N	N	S	N	S	N
031	S	S	N	N	N	N	-	N	N	N	N	N	N	N	N	S	S	N
032	S	N	N	S	N	N	0	N	N	N	N	N	N	N	N	N	N	N
033	S	S	N	N	N	N	5	N	S	S	N	N	S	S	S	N	S	N
034	S	S	S	N	S	N	3	S	N	N	N	N	N	N	N	S	N	S
035	S	S	S	N	S	S	12	S	N	N	N	N	N	N	N	S	S	S
036	S	N	N	S	S	N	3	N	S	N	N	N	S	S	N	N	N	S
037	S	S	N	N	S	N	0	S	N	N	N	N	N	N	N	N	N	N
038	S	S	N	N	N	N	-	N	S	S	S	S	N	N	N	S	N	S
039	S	S	N	N	S	N	3	S	S	S	S	N	N	N	N	S	N	S
040	S	S	S	N	S	N	6	S	S	S	S	S	N	N	N	S	S	S
041	S	S	S	N	S	S	6	S	S	N	S	N	N	N	N	N	S	S
042	S	S	S	N	N	N	8	S	N	S	N	S	N	N	S	N	N	N
043	S	S	S	N	N	N	-	S	N	S	N	S	N	N	N	N	N	S
044	N	N	N	N	S	N	0	N	N	N	N	N	N	N	N	N	N	N
045	N	S	S	N	N	N	3	N	N	S	N	S	N	N	N	S	N	S
046	S	S	N	N	S	N	3	S	S	N	S	N	N	N	S	N	N	N
047	S	S	N	N	S	N	4	S	S	S	S	N	N	N	S	N	N	N
048	S	N	N	N	N	N	0	S	S	N	S	N	N	N	N	N	N	S
049	N	N	N	S	N	N	0	N	N	N	N	N	N	N	N	N	N	S
050	S	S	S	N	N	N	4	N	N	S	N	S	N	N	N	S	S	N
051	S	S	N	N	S	N	3	N	S	S	S	N	N	N	N	S	S	S

Tabela B2: (Continuação).

Pac	Z10	Z11	Z12	Z13	Z14	Z15	Z16	Z17	Z18	Z19	Z20	Z21	Z22	Z23	Z24	Z25	Z26	Z27
052	S	S	S	N	S	N	8	N	N	N	N	N	N	N	S	N	N	N
053	S	S	S	S	S	S	7	S	S	S	S	S	N	N	N	S	S	N
054	S	S	S	N	S	N	6	N	N	N	N	N	N	N	N	S	N	S
055	S	S	S	N	N	N	4	N	S	S	S	S	N	N	S	N	S	N
056	-	-	-	-	-	-	-	-	S	S	S	S	N	N	N	N	S	N
057	-	-	-	-	-	-	-	-	N	N	N	N	N	N	N	N	N	S
058	S	S	S	S	N	S	12	S	N	N	N	N	N	N	S	N	N	N
059	-	-	-	-	-	-	-	-	S	S	S	S	N	N	N	S	S	N
060	N	N	N	N	N	N	0	N	N	N	N	N	N	N	N	N	N	S
061	S	N	N	N	N	N	5	N	N	N	N	N	N	N	N	S	N	S
062	S	S	S	N	N	N	4	S	S	S	S	S	N	N	S	N	S	N
063	S	S	N	N	N	N	3	N	N	S	N	S	N	N	S	N	N	N
064	S	S	N	N	S	N	8	N	S	N	S	N	N	N	S	N	N	N
065	S	S	S	N	S	N	8	N	S	S	S	N	N	N	N	S	S	S
066	S	S	N	N	N	N	6	S	S	N	S	N	N	N	N	S	N	S
067	S	N	N	S	S	N	4	S	N	N	N	N	N	N	N	N	N	N
068	N	N	N	S	N	N	0	N	N	N	N	N	N	N	N	N	N	N
069	S	S	S	N	N	N	-	N	N	N	N	N	N	N	S	N	S	N
070	S	S	N	N	N	N	3	N	N	N	N	N	N	N	N	N	N	N
071	S	S	N	N	S	S	4	N	N	N	N	N	N	N	S	N	N	N
072	N	N	N	S	N	N	0	N	N	N	N	N	N	N	N	S	N	N
073	N	S	N	N	N	N	5	N	N	N	N	N	N	N	N	S	S	S
074	N	N	N	N	N	N	0	N	N	N	N	N	N	N	-	-	-	-
075	S	N	N	N	S	N	0	N	S	N	S	N	N	N	N	S	N	S
076	N	N	N	S	S	N	0	N	N	N	N	N	N	N	N	N	N	S
077	S	N	N	N	N	N	5	S	S	S	S	S	N	N	N	S	N	S
078	S	N	N	N	S	S	6	N	N	N	N	N	N	N	S	N	N	N
079	S	N	N	S	S	N	0	S	N	N	N	N	N	N	N	S	N	N
080	S	N	N	S	S	N	0	N	N	N	N	N	N	N	N	S	S	S
081	S	S	N	N	N	N	10	S	N	S	N	S	N	N	S	N	S	N
082	N	S	N	S	N	N	0	N	N	S	N	S	N	N	N	S	N	S
083	N	N	N	S	N	N	0	N	N	N	N	N	N	N	N	N	N	N
084	S	N	N	N	N	N	2	S	N	N	N	N	N	N	S	N	S	N
085	N	S	S	S	N	N	6	N	N	N	N	N	N	N	S	N	S	N
086	S	N	N	N	N	N	0	N	N	N	N	N	N	N	N	N	N	S
087	S	S	N	S	S	N	9	N	N	N	N	N	N	N	S	N	N	N
088	S	S	N	N	S	N	4	S	S	N	S	N	N	N	S	N	N	N
089	S	S	N	N	S	N	3	N	S	N	S	N	N	N	N	S	S	S
090	S	S	N	N	S	S	0	N	N	N	N	N	N	N	N	N	N	S
091	S	S	S	N	S	N	9	S	S	S	N	N	S	N	N	S	N	S
092	S	S	S	N	N	N	3	N	N	S	N	S	N	N	N	N	N	S
093	S	S	S	N	S	N	8	S	N	N	N	N	N	N	N	S	N	S
094	S	S	N	N	N	N	13	S	N	S	N	N	N	N	S	N	S	N
095	S	S	N	N	S	N	4	N	S	N	S	N	N	N	N	N	S	S
096	S	N	N	N	S	N	0	N	N	N	N	N	N	N	S	N	N	N
097	S	N	N	S	S	N	2	S	N	N	N	N	N	N	N	S	S	S
098	S	N	N	S	N	N	0	N	N	N	N	N	N	N	N	N	N	S
099	S	N	N	S	N	N	2	N	S	S	N	N	S	S	N	S	N	S
100	S	N	N	S	N	N	2	N	N	S	N	S	N	N	S	N	S	N
101	S	S	N	N	N	N	2	S	N	N	N	N	N	N	S	N	N	N
102	S	S	N	S	N	N	0	S	N	N	N	N	N	N	S	N	N	N

Tabela B2: (Continuação).

Pac	Z28	Z29	Z30	Z31	Z32	Z33	Z34	Z35	Z36	Z37	Z38	Z40	Z41	Z42	Z43	Z44	Z45	Z46
052	N	N	N	N	S	S	70	6	-	-	-	-	S	S	N	S	19	20
053	N	N	S	N	S	S	70	9	-	-	-	-	-	-	-	-	18	50
054	S	N	N	N	N	S	51	2	-	-	-	-	-	-	-	-	19	41
055	N	N	N	N	N	S	-	-	-	-	-	-	-	-	-	-	-	-
056	N	N	N	N	S	S	-	-	-	-	-	-	-	-	-	-	-	-
057	N	N	N	N	S	S	-	-	-	-	-	-	-	-	-	-	22	46
058	N	N	N	N	S	S	-	-	-	-	-	-	-	-	-	-	-	-
059	N	N	N	N	N	N	-	-	-	-	-	-	-	-	-	-	-	-
060	N	N	N	N	S	S	52	1	-	-	-	-	-	-	-	-	-	-
061	S	N	N	N	N	N	55	1	-	-	-	-	-	-	S	S	8	71
062	N	N	N	N	N	S	82	4	-	-	-	-	-	-	-	-	-	-
063	S	N	N	N	S	S	54	10	42	30	28	N	S	N	N	S	-	-
064	S	N	N	N	N	S	-	0	51	58	20	-	-	S	-	-	4	50
065	N	N	N	N	S	N	37	0	-	-	-	S	S	-	-	-	7	47
066	S	N	N	N	S	S	99	7	-	-	-	-	-	-	-	-	5	56
067	N	N	S	N	S	S	-	-	76	46	-	-	-	-	-	-	18	71
068	N	N	N	N	N	S	-	-	-	-	-	-	-	-	-	-	-	-
069	N	S	N	N	N	N	92	2	-	-	-	-	-	-	-	-	16	99
070	N	N	N	N	S	S	-	-	-	-	-	-	-	S	-	S	19	28
071	N	N	S	N	S	S	-	-	-	-	-	-	-	-	-	-	-	-
072	N	N	N	N	N	N	-	-	40	26	20	-	S	-	N	S	10	64
073	N	N	N	N	N	N	42	11	27	32	29	S	S	N	S	S	16	22
074	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	30
075	N	N	N	N	N	S	66	1	-	-	-	S	-	-	S	-	5	38
076	N	N	N	N	N	S	-	-	55	39	10	S	S	N	-	-	14	81
077	S	N	N	N	N	S	46	8	-	-	-	-	-	-	-	-	15	22
078	S	N	N	N	N	S	48	0	34	50	22	S	S	N	N	N	-	-
079	N	S	N	N	N	S	86	2	52	43	10	S	S	N	-	-	10	44
080	S	N	N	N	N	N	-	-	40	42	10	S	N	N	N	N	17	64
081	N	N	N	N	N	N	33	20	43	30	33	-	-	-	-	-	16	22
082	N	N	N	N	N	S	70	0	50	35	20	S	S	N	N	S	8	12
083	N	N	N	N	N	S	-	-	47	29	10	S	N	N	N	N	4	35
084	N	N	S	N	N	N	86	1	42	38	10	-	-	N	-	-	8	52
085	S	N	N	N	N	N	49	1	43	35	24	S	S	N	-	-	24	48
086	S	N	N	N	N	S	-	-	55	37	20	S	N	N	S	N	-	26
087	N	N	N	N	N	S	-	-	45	37	12	S	S	N	S	S	18	21
088	S	N	N	N	N	S	74	0	55	49	39	S	S	N	S	S	11	65
089	N	N	N	N	N	N	85	0	55	46	18	S	N	N	-	-	10	80
090	S	N	N	N	N	N	-	78	61	43	28	S	S	-	-	-	12	40
091	N	N	N	N	N	N	86	0	-	-	-	-	-	-	-	-	26	-
092	N	N	N	N	N	S	55	7	49	-	-	-	-	S	S	S	26	27
093	S	N	N	N	N	S	35	2	-	-	-	-	-	-	-	-	-	-
094	N	N	N	N	N	S	99	15	48	49	58	N	S	S	S	S	19	24
095	S	N	N	N	N	S	44	1	-	-	-	-	-	-	-	-	-	-
096	S	N	N	N	N	S	-	62	-	-	-	S	N	-	S	N	5	45
097	N	N	N	N	S	S	-	4	53	27	20	-	-	-	-	-	9	54
098	S	N	N	N	N	S	-	-	50	30	19	N	N	N	N	N	4	40
099	S	N	N	N	N	S	74	0	47	41	32	-	-	N	S	S	15	40
100	N	N	N	N	N	S	74	4	-	-	-	-	-	-	S	S	15	17
101	N	S	N	N	S	S	-	0	73	47	43	-	-	-	-	-	-	-
102	S	N	N	N	N	S	-	-	59	47	27	S	N	-	-	-	7	36

Tabela B2: (Continuação).

Pac	Z28	Z29	Z30	Z31	Z32	Z33	Z34	Z35	Z36	Z37	Z38	Z40	Z41	Z42	Z43	Z44	Z45	Z46
001	-	-	-	-	-	S	74	2	-	-	-	-	-	-	-	-	-	-
002	N	N	N	N	S	S	56	6	59	54	-	S	S	S	N	S	22	83
003	N	N	N	N	N	N	98	2	-	-	-	-	-	-	-	-	17	49
004	N	N	N	N	N	S	56	5	-	-	-	-	-	-	S	S	4	99
005	S	N	N	N	N	S	58	2	-	-	-	-	-	-	S	S	28	36
006	S	N	N	N	S	S	99	1	58	44	-	-	-	N	S	S	13	75
007	S	N	S	N	S	S	82	6	60	48	20	S	S	N	S	S	25	90
008	N	N	N	N	S	S	72	4	-	-	-	-	-	-	-	-	23	76
009	N	S	S	N	N	N	-	-	-	-	-	-	-	-	-	-	17	80
010	S	N	S	N	N	S	64	0	-	-	-	-	-	-	-	-	-	-
011	N	N	N	N	N	S	64	1	-	-	-	-	-	-	-	-	22	43
012	N	N	N	N	N	S	-	-	-	-	-	-	-	-	-	-	-	-
013	N	N	N	N	S	S	99	0	-	-	-	-	-	S	-	-	16	36
014	S	N	N	N	S	S	80	3	-	-	-	-	-	-	-	-	24	43
015	S	N	N	N	N	S	86	1	-	-	-	-	-	S	-	-	-	-
016	N	N	N	N	S	N	-	12	-	-	-	-	-	-	-	-	29	42
017	N	N	N	N	N	S	61	7	-	-	-	-	-	-	-	-	16	22
018	N	N	N	N	S	S	-	1	-	-	-	-	-	-	-	-	5	61
019	N	S	N	N	S	N	82	8	47	38	-	-	-	N	S	N	16	65
020	S	N	N	N	S	S	72	0	-	-	-	-	-	S	-	-	27	55
021	S	N	N	N	N	S	10	0	-	-	-	-	-	S	-	-	27	31
022	N	N	N	N	N	N	80	5	-	-	-	-	-	-	-	-	7	26
023	S	N	N	N	N	S	76	57	-	-	-	S	-	-	-	-	24	50
024	N	N	N	N	S	N	76	4	-	-	-	-	-	-	-	-	7	45
025	S	N	N	N	N	S	61	1	-	-	-	-	-	-	-	-	15	19
026	S	N	N	N	N	N	35	-	-	-	-	-	-	-	-	-	-	-
027	N	N	N	N	N	S	96	0	-	-	-	-	-	-	-	-	20	25
028	S	N	N	N	N	S	48	11	65	51	-	S	N	-	-	-	21	66
029	N	N	N	N	S	S	70	0	-	-	-	-	-	S	-	-	21	25
030	N	N	N	N	S	S	72	1	-	-	-	-	-	S	-	-	22	36
031	N	N	N	N	N	N	70	1	-	-	-	-	-	-	-	S	10	29
032	N	N	N	N	S	S	62	4	55	40	16	S	N	S	S	N	6	30
033	S	N	N	N	N	S	88	4	38	35	53	S	S	N	N	S	16	25
034	N	N	N	N	N	S	36	14	43	42	-	-	-	N	S	S	-	-
035	N	N	N	N	N	S	44	3	54	-	-	S	S	N	-	-	23	48
036	S	N	N	N	N	S	-	3	65	45	-	S	N	N	N	N	11	28
037	N	N	S	N	N	S	-	-	-	-	-	-	-	-	-	-	-	-
038	N	N	N	N	N	S	47	1	-	-	-	-	-	-	-	-	-	-
039	N	N	N	N	N	S	59	1	60	60	25	S	S	S	-	-	12	70
040	N	N	N	N	N	S	76	2	-	-	-	N	S	S	N	S	16	33
041	N	N	N	N	N	S	87	1	61	45	12	-	-	N	-	-	17	76
042	N	N	N	N	S	S	67	1	-	-	-	-	S	S	N	S	16	23
043	N	S	N	N	N	S	58	4	-	-	-	-	-	S	-	-	18	23
044	N	N	N	N	N	N	-	-	58	39	-	S	-	-	-	-	9	51
045	N	N	N	N	S	S	64	2	43	32	32	N	S	N	-	-	17	29
046	N	N	N	N	S	S	99	8	61	45	20	S	N	N	-	-	-	-
047	S	N	N	N	N	N	51	0	-	-	-	-	-	-	S	-	23	99
048	S	N	N	N	N	S	28	5	-	-	-	-	-	-	-	-	-	-
049	S	N	N	N	N	S	-	-	47	36	19	N	N	N	S	S	-	-
050	N	N	N	N	N	N	53	-	-	-	-	-	-	-	N	S	-	-
051	S	N	N	N	N	S	52	2	-	-	-	-	-	-	-	S	9	55

Tabela B2: (Continuação).

Pac	Z28	Z29	Z30	Z31	Z32	Z33	Z34	Z35	Z36	Z37	Z38	Z40	Z41	Z42	Z43	Z44	Z45	Z46
103	N	N	N	N	N	S	70	3	62	41	17	S	N	N	S	S	4	56
104	N	S	S	N	S	S	-	-	69	71	38	S	S	N	-	-	10	30
105	N	N	N	N	N	S	84	31	58	50	37	S	S	N	-	-	6	37
106	S	N	N	N	N	S	97	6	67	37	17	S	N	N	S	N	5	65
107	S	N	N	N	N	S	74	0	64	64	25	S	S	N	S	S	11	40
108	S	N	S	N	N	N	80	1	61	41	19	S	N	N	S	N	12	66
109	N	N	N	N	N	S	-	-	54	33	15	-	-	-	S	-	-	39
110	N	N	N	N	N	S	-	1	53	38	30	S	S	N	N	S	17	21
111	N	N	N	N	N	S	54	1	44	36	15	N	-	S	N	N	18	20
112	S	N	N	N	N	N	78	-	-	-	-	S	S	-	S	S	-	-
113	N	S	N	N	N	S	74	5	-	-	-	-	-	-	-	-	10	17
114	S	N	N	N	N	S	57	1	60	60	20	S	S	N	-	-	27	75
115	N	N	N	N	S	S	76	-	55	38	34	S	S	N	-	-	8	35
116	N	N	N	N	N	S	75	0	44	30	18	N	S	S	S	S	18	43
117	N	N	N	N	N	S	-	4	55	60	26	S	S	S	-	-	25	51
118	N	N	N	N	N	S	46	8	53	44	44	S	S	N	S	S	16	37
119	S	N	N	N	N	S	72	2	48	44	12	N	S	S	-	-	17	23
120	S	N	N	N	N	S	-	-	57	44	20	-	-	N	S	N	2	34
121	N	N	N	N	N	N	-	-	-	-	-	-	-	-	-	-	16	32
122	N	N	N	N	N	N	-	1	67	52	20	-	-	-	S	S	2	45
123	N	N	N	N	N	S	58	9	46	52	35	S	S	N	S	S	19	62
124	S	N	N	N	N	S	66	2	60	51	22	-	-	-	-	-	-	-
125	N	N	N	N	N	N	84	14	63	43	19	S	S	N	S	-	-	-
126	N	N	N	N	N	S	-	-	59	45	21	S	S	S	-	-	14	92
127	S	N	N	N	N	S	82	11	50	43	22	S	S	N	S	S	7	70
128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
129	S	N	N	N	N	S	83	20	38	28	16	S	S	S	S	S	20	25
130	S	N	N	N	N	S	-	-	-	-	-	-	-	-	-	-	-	-
131	N	N	N	N	N	S	72	15	58	41	15	S	S	N	S	S	18	25
132	N	N	N	N	S	S	-	-	58	33	20	N	N	N	N	N	-	-
133	N	N	N	S	N	S	-	-	46	41	40	S	S	N	S	S	16	53
134	N	N	N	N	N	N	70	2	48	46	-	-	-	S	S	S	18	23
135	S	N	N	N	N	S	-	-	51	37	20	-	-	N	S	S	12	85
136	N	N	N	N	N	N	78	1	42	38	30	N	S	S	N	S	15	25
137	N	N	N	N	N	S	89	1	53	36	20	S	N	N	S	N	4	24
138	N	N	N	N	N	S	94	6	49	40	27	N	S	N	-	S	21	24
139	N	N	N	N	N	S	83	0	41	27	20	N	S	N	-	-	4	35
140	S	N	N	N	N	S	50	0	52	37	19	N	N	N	-	-	13	37
141	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
142	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
143	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
144	-	-	-	-	-	-	-	-	52	62	37	S	S	-	-	-	-	-
145	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
146	-	-	-	-	-	-	-	-	55	45	-	S	-	-	-	-	-	-
147	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
148	N	N	S	N	N	S	45	1	69	68	22	S	N	N	S	N	8	34
149	N	S	N	N	N	S	66	2	-	-	-	-	-	-	-	-	22	26
150	N	N	N	N	S	S	-	-	-	-	-	-	-	-	-	-	16	24
151	N	N	N	N	N	S	73	7	39	40	-	S	S	N	S	S	16	24

Tabela B2: (Continuação).

Pac	Z47	Z48	Z49	Z50	Z51	Z52	Z53	Z54	Z55	Z56	Z57	Z58	Z59	TEMPO	δ	GRUPO
001	-	-	-	-	-	-	-	-	-	-	-	-	-	367	1	2
002	6	28	80	48	35	127	0	35	114	88	S	S	N	1391	0	1
003	4	17	48	24	25	125	3	25	130	82	S	S	N	4869	0	1
004	0	12	99	40	38	123	0	37	121	80	N	N	N	236	0	2
005	19	32	35	30	23	123	0	23	104	75	S	S	N	1078	1	2
006	0	17	75	35	26	142	0	35	142	79	S	N	S	3455	0	1
007	0	25	90	50	35	110	0	35	110	70	S	N	N	4011	0	2
008	10	23	76	41	40	121	14	38	121	90	S	S	N	801	1	2
009	0	17	80	45	35	140	0	25	140	99	S	N	N	558	1	2
010	-	-	-	-	-	-	-	-	-	-	-	-	-	4	1	2
011	15	25	37	23	26	95	8	26	85	64	S	S	N	703	1	2
012	-	-	-	-	-	0	-	-	-	-	-	-	-	123	1	2
013	7	16	36	21	20	117	4	23	114	84	S	N	N	293	1	2
014	8	25	41	27	28	155	0	27	137	91	S	S	N	229	1	2
015	-	-	-	-	-	-	-	-	-	-	-	-	-	84	1	1
016	20	30	41	29	30	119	21	35	120	94	S	S	N	480	1	1
017	13	16	22	16	13	104	7	24	97	67	-	-	-	179	1	2
018	7	9	60	29	38	156	7	32	150	91	N	N	S	4563	1	1
019	6	15	61	24	21	103	7	25	101	70	S	S	N	2389	0	1
020	10	27	55	29	34	123	8	24	120	83	S	S	N	1334	1	2
021	19	29	31	25	20	138	6	30	122	92	N	S	N	1702	1	2
022	0	7	25	10	9	103	9	13	103	81	S	S	N	108	1	2
023	2	24	50	30	30	110	0	30	110	90	S	S	N	964	1	1
024	2	13	46	20	21	109	4	16	104	72	S	S	N	1137	1	1
025	11	15	21	17	11	120	0	19	119	70	N	S	N	1712	1	2
026	-	-	-	-	-	-	-	-	-	-	-	-	-	58	1	1
027	12	22	25	16	17	128	0	37	128	81	S	S	N	1358	1	1
028	9	19	64	41	40	123	7	34	120	97	S	N	N	524	1	1
029	14	20	26	21	16	101	10	13	92	72	S	S	N	41	1	1
030	24	30	37	29	29	134	10	35	120	80	S	N	N	161	0	2
031	7	11	25	10	10	124	5	15	124	78	N	N	N	3995	0	2
032	0	7	30	13	15	122	1	16	122	72	N	N	N	3767	0	2
033	-	13	25	13	11	114	-	16	114	80	S	S	N	3856	0	1
034	-	-	-	-	-	-	-	-	-	-	-	-	-	716	0	2
035	3	20	47	23	21	95	0	19	83	73	N	S	N	2285	1	2
036	6	15	28	16	14	116	0	13	116	74	S	N	S	3355	0	1
037	-	-	-	-	-	-	-	-	-	-	-	-	-	33	0	2
038	-	-	-	-	-	-	-	-	-	-	-	-	-	320	1	1
039	5	15	70	39	37	113	11	26	109	67	N	N	N	2132	0	1
040	11	20	31	19	16	126	8	25	127	89	N	S	N	2861	0	1
041	3	19	78	35	26	119	0	26	111	77	S	S	N	2189	0	1
042	10	14	21	14	5	123	0	6	122	78	N	S	N	65	1	1
043	10	18	22	17	9	98	0	9	92	59	-	-	N	1068	0	1
044	0	16	40	14	24	131	0	39	112	80	S	N	N	155	1	2
045	0	20	29	16	16	155	0	16	150	99	N	S	N	4159	0	1
046	-	-	-	-	-	-	-	-	-	-	S	-	N	2693	0	1
047	0	26	99	54	40	122	21	37	117	86	S	N	N	684	0	1
048	-	-	-	-	-	-	-	-	-	-	N	N	N	1667	0	1
049	-	-	-	-	-	-	-	-	-	-	N	N	S	2718	0	2
050	-	-	-	-	-	-	-	-	-	-	N	S	N	2004	0	1
051	3	13	57	28	24	120	7	28	124	86	S	S	N	3398	0	1

Tabela B2: (continuação)

Pac	Z47	Z48	Z49	Z50	Z51	Z52	Z53	Z54	Z55	Z56	Z57	Z58	Z59	TEMPO	δ	GRUPO
052	14	17	20	15	15	105	0	8	104	63	S	S	N	399	1	2
053	4	18	50	30	20	110	6	20	110	60	S	S	N	1794	0	1
054	13	26	40	23	24	116	8	30	116	83	S	S	N	1556	1	2
055	-	-	-	-	-	-	-	-	-	-	S	S	N	128	1	1
056	-	-	-	-	-	-	-	-	-	-	S	S	N	2979	0	1
057	11	24	46	27	26	95	6	28	90	70	S	N	N	45	1	2
058	-	-	-	-	-	-	-	-	-	-	-	-	-	4	1	2
059	-	-	-	-	-	-	-	-	-	-	-	-	-	3771	0	1
060	-	-	-	-	-	-	-	-	-	-	-	-	-	177	1	2
061	0	21	70	29	25	113	14	43	108	67	S	N	N	927	0	2
062	-	-	-	-	-	-	-	-	-	-	S	S	N	2194	0	1
063	-	-	-	-	-	-	-	-	-	-	N	S	N	1789	0	1
064	3	5	50	15	12	120	0	20	130	82	S	N	N	13	1	1
065	1	12	47	20	18	111	0	17	109	80	S	S	N	60	1	1
066	0	9	57	27	23	112	2	26	106	85	N	N	N	1281	1	1
067	5	21	67	35	38	120	13	43	115	79	N	N	N	1343	0	2
068	-	-	-	-	12	150	0	10	150	82	N	N	N	2324	0	2
069	10	16	99	55	30	108	17	25	108	74	S	S	N	3	1	2
070	12	21	23	19	16	-	-	-	-	-	-	-	N	10	1	2
071	-	-	-	-	-	-	-	-	-	-	-	-	-	20	1	2
072	0	16	54	20	14	120	0	11	120	70	N	N	N	2259	0	2
073	10	18	20	17	11	89	0	12	89	47	S	N	N	57	0	2
074	0	8	30	10	8	130	0	12	130	90	N	N	N	167	0	2
075	0	12	35	13	15	104	0	24	104	68	S	S	N	710	0	1
076	4	15	80	45	52	219	26	50	218	99	N	N	N	8	0	2
077	7	17	26	16	11	105	2	15	100	75	N	S	N	1752	0	1
078	-	-	-	-	-	-	-	-	-	-	-	-	-	1784	0	1
079	9	13	44	24	24	101	21	36	101	72	N	N	N	989	0	2
080	2	22	63	35	32	90	20	45	90	63	N	N	N	2815	0	2
081	8	17	21	16	16	113	4	12	112	77	N	S	N	2231	0	1
082	-	8	12	8	10	120	-	12	120	80	N	S	N	1554	0	1
083	0	7	27	10	12	147	-	13	133	78	N	N	N	598	0	2
084	0	8	52	30	30	108	14	30	108	60	N	N	N	3054	1	2
085	12	24	48	28	36	89	21	32	87	53	N	S	N	1353	0	2
086	0	2	26	6	4	165	0	11	146	84	N	N	N	895	0	2
087	15	17	22	16	21	125	0	12	126	75	N	S	S	126	0	2
088	3	10	63	31	25	145	0	27	143	99	S	S	N	1220	0	1
089	0	10	80	38	30	104	10	30	104	76	S	N	N	1587	0	1
090	8	12	40	20	20	120	0	20	120	80	S	S	N	1453	0	2
091	-	24	-	24	21	106	-	21	109	-	N	S	N	705	1	1
092	12	21	25	12	19	102	0	25	102	70	S	S	N	348	1	1
093	-	-	-	-	-	-	-	-	-	-	-	-	-	9	1	2
094	11	18	24	17	9	122	0	11	109	75	S	S	N	1398	0	1
095	-	-	-	-	-	-	-	-	-	-	S	S	S	1574	1	1
096	-	8	45	19	19	123	-	29	123	74	N	N	N	30	1	2
097	6	14	54	26	22	163	0	26	162	95	S	N	N	1469	0	2
098	0	7	40	14	16	136	0	33	136	84	N	N	N	1605	0	2
099	5	22	40	25	20	150	5	20	150	99	N	S	N	886	1	1
100	-	13	17	13	13	107	-	12	104	68	S	S	N	1199	0	1
101	-	-	-	-	-	-	-	-	-	-	-	-	-	113	0	2
102	6	9	34	16	16	139	0	20	139	85	N	N	N	698	0	2

Tabela B2: (continuação)

Pac	Z47	Z48	Z49	Z50	Z51	Z52	Z53	Z54	Z55	Z56	Z57	Z58	Z59	TEMPO	δ	GRUPO
103	-	6	56	22	37	103	0	38	103	67	N	N	N	2260	0	2
104	-	10	30	17	17	110	-	20	110	70	S	N	N	1156	0	1
105	1	6	39	15	14	85	6	18	85	59	N	S	N	1221	0	1
106	0	10	78	28	28	122	0	31	96	57	S	N	N	1303	0	1
107	5	12	37	20	16	130	8	20	130	80	S	N	N	32	0	2
108	0	12	66	30	30	114	0	30	114	74	S	S	N	1448	0	1
109	0	0	39	16	15	145	0	26	145	80	N	N	N	1201	0	2
110	13	17	21	17	12	110	2	12	110	84	N	S	N	313	0	2
111	-	17	20	17	12	138	-	13	138	78	S	N	N	864	0	1
112	-	-	-	-	-	-	-	-	-	-	S	N	N	15	1	1
113	4	9	17	10	10	97	0	10	97	64	N	S	S	1062	0	1
114	6	27	73	58	26	90	16	26	88	72	S	N	N	931	0	2
115	0	8	35	11	13	98	0	22	98	64	S	S	N	1081	0	1
116	-	22	43	20	23	118	-	36	117	91	S	S	N	1167	0	1
117	-	22	51	30	30	91	-	28	90	69	S	S	N	949	0	1
118	3	16	35	16	14	91	11	16	96	64	S	S	N	1019	0	1
119	8	18	20	16	20	92	0	20	92	47	S	S	N	858	0	1
120	0	6	34	10	11	133	0	11	133	74	S	N	N	774	0	2
121	-	16	23	-	16	110	-	24	110	80	S	S	N	464	0	2
122	0	3	45	26	9	82	0	13	82	53	S	N	N	1334	1	2
123	13	22	59	29	24	140	0	25	133	88	S	S	N	625	0	1
124	-	-	-	-	-	-	-	-	-	-	S	S	N	83	1	1
125	-	-	-	-	-	-	-	-	-	-	S	N	N	406	0	2
126	6	18	91	37	41	128	11	43	115	80	S	S	N	4464	0	2
127	6	14	70	33	43	117	15	46	117	78	S	N	N	478	0	1
128	-	-	-	-	-	160	0	9	152	86	N	N	S	818	0	2
129	14	19	25	18	22	101	3	22	95	72	S	S	N	464	0	1
130	-	-	-	-	20	170	0	20	170	90	S	N	S	651	0	1
131	-	12	25	12	15	100	1	15	100	75	S	S	N	474	0	1
132	-	-	-	-	11	135	0	11	125	81	N	N	S	528	0	2
133	6	17	48	35	23	122	11	26	116	80	S	S	N	139	1	2
134	-	18	23	18	15	110	-	14	110	70	S	S	N	434	0	2
135	8	15	85	45	40	118	16	44	112	79	S	N	N	566	0	2
136	8	13	25	10	6	107	2	10	116	74	S	S	N	364	0	1
137	0	4	24	8	12	150	0	8	150	90	N	N	N	466	0	1
138	9	22	27	19	14	111	0	23	108	79	S	S	N	280	0	1
139	0	5	35	8	6	-	-	-	-	-	N	S	N	185	0	1
140	3	13	37	14	15	100	0	14	100	68	N	N	N	198	0	1
141	-	-	-	-	-	-	-	-	-	-	N	N	N	2371	0	1
142	-	-	-	-	-	-	-	-	-	-	-	-	-	2617	0	1
143	-	-	-	-	-	-	-	-	-	-	-	-	-	2441	0	1
144	-	-	-	-	-	-	-	-	-	-	-	-	-	447	0	1
145	-	-	-	-	-	-	-	-	-	-	-	-	-	436	0	1
146	-	-	-	-	-	-	-	-	-	-	-	-	-	343	0	1
147	-	-	-	-	-	-	-	-	-	-	-	-	-	316	0	1
148	6	8	37	21	21	132	0	21	121	77	S	S	N	238	0	1
149	18	20	26	-	10	110	0	10	110	76	S	S	N	116	0	1
150	0	16	70	25	12	120	0	35	120	76	N	N	N	178	0	2
151	11	12	24	15	13	126	0	10	126	64	S	S	N	70	0	2

Fonte: Instituto do Coração - Hospital das Clínicas São Paulo - [1991] - Dados usado com permissão do pesquisador.

OBS:1. A variável Z39 não consta da tabela acima, pois seus valores são todos *missing*.

2. As siglas referentes as variáveis encontram-se definidas na Tabela 4.1, página 91.

APÊNDICE C

RESULTADOS

Neste apêndice encontram-se algumas tabelas referentes a resultados numéricos obtidos ao longo desta tese.

Na Tabela C1, estimativas da função de sobrevivência e os respectivos intervalos de confiança a 95 % para o modelo de Cox, usou-se o programa PEVSTMC.PRG e para o modelo exponencial de Glasser utilizou-se o programa PESVSMG.PRG.

As estimativas da função de sobrevivência e os respectivos intervalos de confiança a 95 %, para os modelos logístico e logístico exponencial estão apresentados na Tabela C2 e para o modelo logístico por partes na Tabela C3. Os programas utilizados nos cálculos foram PESVSLAS.PRG, PAMLEAS.PRG e PAMLEPAS.PRG, respectivamente.

Na Tabela C4, estimativas da função de sobrevivência e os respectivos intervalos de confiança no modelo de Cox, para os pacientes com endomiocardiopatia utilizados no Capítulo 4.

Nas Tabelas C5 e C7 encontram-se as tabelas de vidas referentes aos pacientes com endomiocardiopatia para as combinações dos níveis dos fatores tipo de tratamento (TRATAMENTO) e classe funcional final (CF2) e tipo de tratamento e fibrilação do atrio (FA), respectivamente. Nas Tabelas C6 e C8 apresentam-se as tabelas de contingência para o número de mortes e total de tempo exposto ao risco para pacientes com endomiocardiopatia e as respectivas combinações dos níveis dos fatores indicados.

Na Tabela C9 encontram-se as matrizes de planejamento utilizadas no ajuste do modelo exponencial por partes do Capítulo 4.

Tabela CI. Intervalos de confiança para a função de sobrevivência estimada para cada paciente com mieloma múltiplo nos modelos de Cox [1972] e de Glasser [1967].

IND	Modelo de Cox			Modelo de Glasser		
	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]
1	0.8919	0.0524	(0.7890; 0.9947)	0.8501	0.0660	(0.7207; 0.9795)
2	0.9486	0.0383	(0.8735; 1.0238)	0.9225	0.0270	(0.8696; 0.9754)
3	0.9153	0.0287	(0.8591; 0.9715)	0.9187	0.0125	(0.8942; 0.9431)
4	0.8970	0.0408	(0.8169; 0.9770)	0.9018	0.0243	(0.8541; 0.9494)
5	0.8980	0.0440	(0.8118; 0.9842)	0.9058	0.0280	(0.8509; 0.9607)
6	0.8501	0.0571	(0.7381; 0.9621)	0.8329	0.0366	(0.7613; 0.9046)
7	0.8084	0.0795	(0.6526; 0.9641)	0.7272	0.0798	(0.5708; 0.8836)
8	0.8128	0.0863	(0.6436; 0.9821)	0.7330	0.0748	(0.5864; 0.8796)
9	0.6146	0.1234	(0.3727; 0.8565)	0.5375	0.1635	(0.2171; 0.8580)
10	0.7443	0.0984	(0.5515; 0.9371)	0.6810	0.0731	(0.5377; 0.8242)
11	0.8085	0.0488	(0.7127; 0.9042)	0.8048	0.0270	(0.7518; 0.8577)
12	0.5223	0.1075	(0.3115; 0.7331)	0.5427	0.1415	(0.2654; 0.8200)
13	0.8788	0.0368	(0.8066; 0.9510)	0.8712	0.0256	(0.8211; 0.9214)
14	0.8214	0.0550	(0.7135; 0.9292)	0.8156	0.0245	(0.7675; 0.8637)
15	0.4735	0.0990	(0.2794; 0.6676)	0.5370	0.1182	(0.3053; 0.7687)
16	0.7686	0.0578	(0.6554; 0.8819)	0.7928	0.0686	(0.6583; 0.9274)
17	0.8557	0.0391	(0.7790; 0.9324)	0.8622	0.0258	(0.8116; 0.9128)
18	0.8828	0.0331	(0.8179; 0.9477)	0.8857	0.0266	(0.8336; 0.9377)
19	0.7219	0.0743	(0.5762; 0.8675)	0.7478	0.0358	(0.6776; 0.8181)
20	0.8535	0.0438	(0.7677; 0.9393)	0.8431	0.0326	(0.7793; 0.9070)
21	0.5354	0.1045	(0.3306; 0.7401)	0.5383	0.0824	(0.3769; 0.6998)
22	0.8476	0.0307	(0.7874; 0.9077)	0.8509	0.0416	(0.7693; 0.9325)
23	0.7749	0.0448	(0.6871; 0.8627)	0.7882	0.0380	(0.7137; 0.8627)
24	0.7797	0.0458	(0.6900; 0.8694)	0.7912	0.0438	(0.7054; 0.8770)
25	0.4644	0.0872	(0.2935; 0.6353)	0.5273	0.0727	(0.3848; 0.6698)
26	0.7684	0.0509	(0.6687; 0.8681)	0.7850	0.0434	(0.7001; 0.8700)
27	0.6829	0.0688	(0.5481; 0.8178)	0.7058	0.0746	(0.5595; 0.8520)
28	0.7927	0.0502	(0.6943; 0.8912)	0.7884	0.0405	(0.7091; 0.8678)
29	0.6101	0.0848	(0.4439; 0.7764)	0.6256	0.0461	(0.5353; 0.7160)
30	0.7571	0.0560	(0.6473; 0.8670)	0.7591	0.0961	(0.5708; 0.9474)
31	0.2687	0.0969	(0.0787; 0.4587)	0.3175	0.1236	(0.0753; 0.5598)
32	0.7510	0.0572	(0.6389; 0.8632)	0.7414	0.0664	(0.6113; 0.8716)
33	0.4984	0.0897	(0.3227; 0.6741)	0.5138	0.0653	(0.3858; 0.6418)
34	0.5352	0.0780	(0.3823; 0.6882)	0.5699	0.0501	(0.4717; 0.6682)
35	0.5386	0.0804	(0.3809; 0.6963)	0.5732	0.0521	(0.4712; 0.6753)
36	0.7561	0.0530	(0.6523; 0.8600)	0.7604	0.0535	(0.6555; 0.8652)
37	0.5917	0.0706	(0.4534; 0.7299)	0.6357	0.0491	(0.5396; 0.7319)
38	0.4350	0.0857	(0.2670; 0.6030)	0.4980	0.0688	(0.3632; 0.6329)
39	0.3384	0.0842	(0.1734; 0.5034)	0.4193	0.0722	(0.2778; 0.5609)
40	0.7468	0.0461	(0.6565; 0.8371)	0.7760	0.0646	(0.6494; 0.9025)
41	0.4103	0.0796	(0.2542; 0.5663)	0.5054	0.0743	(0.3597; 0.6510)
42	0.5957	0.0702	(0.4581; 0.7334)	0.6526	0.0630	(0.5291; 0.7761)
43	0.5102	0.0813	(0.3509; 0.6695)	0.5847	0.0492	(0.4882; 0.6812)
44	0.6435	0.0662	(0.5137; 0.7733)	0.6429	0.0852	(0.4759; 0.8100)
45	0.6920	0.0584	(0.5775; 0.8064)	0.6951	0.0730	(0.5520; 0.8381)
46	0.5150	0.0768	(0.3646; 0.6655)	0.5426	0.0605	(0.4240; 0.6612)

Tabela C1. (Continuação).

IND	Modelo de Cox			Modelo de Glasser		
	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]
47	0.3481	0.0847	(0.1821; 0.5141)	0.3722	0.0841	(0.2074; 0.5371)
48	0.4125	0.0822	(0.2513; 0.5737)	0.3974	0.0564	(0.2868; 0.5080)
49	0.3601	0.0802	(0.2028; 0.5174)	0.3447	0.1007	(0.1475; 0.5420)
50	0.2161	0.0706	(0.0777; 0.3545)	0.2069	0.0644	(0.0807; 0.3331)
51	0.4961	0.0684	(0.3621; 0.6301)	0.4723	0.0926	(0.2908; 0.6538)
52	0.1898	0.0644	(0.0636; 0.3159)	0.1990	0.1116	(-0.0197; 0.4177)
53	0.1614	0.0630	(0.0378; 0.2850)	0.1596	0.0890	(-0.0147; 0.3340)
54	0.0709	0.0391	(-0.0057; 0.1475)	0.0547	0.0332	(-0.0104; 0.1198)
55	0.4336	0.0751	(0.2865; 0.5807)	0.3824	0.0961	(0.1940; 0.5707)
56	0.4464	0.0809	(0.2878; 0.6050)	0.3836	0.0869	(0.2133; 0.5538)
57	0.2018	0.0712	(0.0622; 0.3414)	0.1903	0.0546	(0.0832; 0.2973)
58	0.3481	0.0885	(0.1746; 0.5215)	0.2962	0.0773	(0.1447; 0.4478)
59	0.3237	0.0870	(0.1532; 0.4943)	0.3034	0.0767	(0.1531; 0.4538)
60	0.0252	0.0212	(-0.0164; 0.0669)	0.0302	0.0248	(-0.0184; 0.0788)
61	0.1915	0.0735	(0.0474; 0.3356)	0.2151	0.0721	(0.0739; 0.3564)
62	0.3852	0.1034	(0.1825; 0.5879)	0.3425	0.1146	(0.1179; 0.5670)
63	0.1299	0.0733	(-0.0134; 0.2736)	0.1319	0.0508	(0.0322; 0.2315)
64	0.1110	0.0677	(-0.0218; 0.2438)	0.1651	0.0805	(0.0073; 0.3228)
65	0.0000	-	-	0.0492	0.0235	(0.0031; 0.0954)

Tabela C2. Intervalos de confiança para a função de sobrevivência estimada para cada paciente com mieloma múltiplo nos modelos logístico linear e logístico exponencial.

IND	Modelo Logístico			Modelo Logístico Exponencial		
	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]
1	0.9945	0.0092	(0.9766; 1.0125)	0.8048	0.0515	(0.7038; 0.9059)
2	0.5741	0.2211	(0.1407; 1.0075)	0.9079	0.0348	(0.8396; 0.9762)
3	0.6738	0.1447	(0.3902; 0.9574)	0.9332	0.0235	(0.8872; 0.9792)
4	0.9448	0.0475	(0.8518; 1.0378)	0.8096	0.0426	(0.7261; 0.8931)
5	0.4022	0.1283	(0.1506; 0.6537)	0.9598	0.0158	(0.9289; 0.9907)
6	0.8946	0.0653	(0.7667; 1.0226)	0.7890	0.0415	(0.7078; 0.8703)
7	0.6734	0.1923	(0.2965; 1.0503)	0.7289	0.0919	(0.5488; 0.9089)
8	0.9984	0.0035	(0.9915; 1.0053)	0.5937	0.0838	(0.4294; 0.7580)
9	0.7884	0.1996	(0.3972; 1.1797)	0.5588	0.1332	(0.2976; 0.8199)
10	0.9519	0.0430	(0.8675; 1.0362)	0.6170	0.0715	(0.4769; 0.7571)
11	0.6821	0.1025	(0.4812; 0.8830)	0.7064	0.0532	(0.6022; 0.8106)
12	0.9836	0.0219	(0.9408; 1.0265)	0.3871	0.1058	(0.1798; 0.5945)
13	0.8038	0.1132	(0.5821; 1.0256)	0.8109	0.0520	(0.7090; 0.9127)
14	0.8880	0.0692	(0.7524; 1.0236)	0.6821	0.0558	(0.5727; 0.7915)
15	0.9741	0.0299	(0.9155; 1.0327)	0.3803	0.1016	(0.1813; 0.5794)
16	0.5094	0.1296	(0.2555; 0.7633)	0.8129	0.0599	(0.6955; 0.9303)
17	0.2315	0.1054	(0.0249; 0.4380)	0.9018	0.0428	(0.8178; 0.9857)
18	0.3864	0.1418	(0.1086; 0.6642)	0.7830	0.0585	(0.6683; 0.8978)
19	0.9616	0.0396	(0.8840; 1.0391)	0.5816	0.0695	(0.4454; 0.7178)

Tabela C2. (Continuação).

IND	Modelo Logístico			Modelo Logístico Exponencial		
	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]
20	0.5418	0.1226	(0.3014; 0.7821)	0.7723	0.0656	(0.6438; 0.9008)
21	0.9404	0.0495	(0.8434; 1.0374)	0.3980	0.0902	(0.2213; 0.5747)
22	0.5712	0.1164	(0.3429; 0.7994)	0.6809	0.0800	(0.5241; 0.8377)
23	0.5275	0.1785	(0.1777; 0.8774)	0.8288	0.0670	(0.6975; 0.9601)
24	0.7170	0.0924	(0.5359; 0.8980)	0.5666	0.0728	(0.4240; 0.7092)
25	0.8135	0.0923	(0.6326; 0.9944)	0.4059	0.0822	(0.2448; 0.5670)
26	0.4484	0.1315	(0.1906; 0.7062)	0.7009	0.0818	(0.5406; 0.8613)
27	0.3050	0.1533	(0.0045; 0.6054)	0.6271	0.1073	(0.4168; 0.8374)
28	0.5057	0.1229	(0.2648; 0.7465)	0.6375	0.0822	(0.4763; 0.7987)
29	0.8371	0.0778	(0.6845; 0.9897)	0.4760	0.0757	(0.3277; 0.6244)
30	0.4707	0.1948	(0.0890; 0.8524)	0.8203	0.0972	(0.6299; 1.0107)
31	0.9489	0.0446	(0.8614; 1.0364)	0.2941	0.0854	(0.1267; 0.4615)
32	0.3922	0.1188	(0.1593; 0.6251)	0.6940	0.0949	(0.5080; 0.8801)
33	0.9120	0.0602	(0.7940; 1.0300)	0.2742	0.0812	(0.1150; 0.4334)
34	0.7449	0.0893	(0.5699; 0.9199)	0.4089	0.0800	(0.2521; 0.5658)
35	0.7315	0.0907	(0.5537; 0.9093)	0.4228	0.0808	(0.2645; 0.5811)
36	0.6839	0.1081	(0.4721; 0.8958)	0.5486	0.0944	(0.3636; 0.7336)
37	0.8078	0.0936	(0.6244; 0.9913)	0.4664	0.0879	(0.2942; 0.6386)
38	0.9089	0.0612	(0.7891; 1.0288)	0.2363	0.0775	(0.0843; 0.3883)
39	0.6643	0.1241	(0.4210; 0.9076)	0.2972	0.0773	(0.1457; 0.4486)
40	0.8850	0.1029	(0.6833; 1.0867)	0.5413	0.1091	(0.3274; 0.7552)
41	0.1139	0.0820	(-0.0468; 0.2746)	0.7073	0.1026	(0.5063; 0.9083)
42	0.7315	0.0907	(0.5537; 0.9093)	0.3598	0.0816	(0.1999; 0.5198)
43	0.8521	0.0804	(0.6945; 1.0098)	0.3598	0.0816	(0.1999; 0.5198)
44	0.3615	0.1803	(0.0080; 0.7149)	0.3559	0.0831	(0.1930; 0.5188)
45	0.3824	0.1404	(0.1073; 0.6576)	0.5848	0.1218	(0.3460; 0.8237)
46	0.1055	0.0783	(-0.0479; 0.2589)	0.7107	0.1044	(0.5061; 0.9154)
47	0.5774	0.1128	(0.3564; 0.7985)	0.4077	0.0903	(0.2306; 0.5847)
48	0.9242	0.0650	(0.7968; 1.0516)	0.3410	0.0814	(0.1814; 0.5005)
49	0.4280	0.1834	(0.0685; 0.7876)	0.7792	0.0969	(0.5894; 0.9691)
50	0.3809	0.1469	(0.0930; 0.6687)	0.4362	0.1324	(0.1766; 0.6958)
51	0.1930	0.1135	(-0.0294; 0.4154)	0.8341	0.0882	(0.6613; 1.0070)
52	0.5057	0.1229	(0.2648; 0.7465)	0.4722	0.1015	(0.2732; 0.6711)
53	0.5106	0.1758	(0.1660; 0.8552)	0.3242	0.1388	(0.0522; 0.5962)
54	0.7809	0.1391	(0.5082; 1.0535)	0.4623	0.1299	(0.2076; 0.7169)
55	0.1930	0.1135	(-0.0294; 0.4154)	0.8341	0.0882	(0.6613; 1.0070)
56	0.3864	0.1418	(0.1086; 0.6642)	0.4971	0.1062	(0.2890; 0.7053)
57	0.2137	0.0995	(0.0187; 0.4088)	0.6946	0.1060	(0.4868; 0.9023)
58	0.6840	0.0966	(0.4948; 0.8733)	0.3891	0.0875	(0.2176; 0.5607)
59	0.1355	0.0848	(-0.0307; 0.3017)	0.7273	0.1028	(0.5258; 0.9287)
60	0.5764	0.1582	(0.2663; 0.8864)	0.2597	0.0750	(0.1127; 0.4067)
61	0.5128	0.1505	(0.2177; 0.8078)	0.6488	0.1104	(0.4324; 0.8652)
62	0.3585	0.1430	(0.0783; 0.6388)	0.5241	0.1112	(0.3061; 0.7421)
63	0.0892	0.0706	(-0.0492; 0.2276)	0.7442	0.1010	(0.5462; 0.9421)
64	0.3322	0.1137	(0.1094; 0.5551)	0.6488	0.1104	(0.4324; 0.8652)
65	0.4199	0.1216	(0.1816; 0.6583)	0.5681	0.1186	(0.3356; 0.8006)

Tabela C3. Intervalos de confiança para a função de sobrevivência estimada para cada paciente com mieloma múltiplo no modelo logístico exponencial por partes.

IND	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]	IND	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]
1	0.8034	0.0431	(0.7190; 0.8878)	34	0.4998	0.0437	(0.4142; 0.5853)
2	0.9150	0.0292	(0.8579; 0.9722)	35	0.5127	0.0423	(0.4297; 0.5957)
3	0.8759	0.0361	(0.8052; 0.9466)	36	0.5756	0.0398	(0.4976; 0.6536)
4	0.8606	0.0267	(0.8082; 0.9130)	37	0.4429	0.0511	(0.3427; 0.5430)
5	0.9625	0.0045	(0.9536; 0.9714)	38	0.1135	0.0600	(-0.0042; 0.2312)
6	0.7539	0.0584	(0.6394; 0.8684)	39	0.4709	0.0514	(0.3702; 0.5716)
7	0.8353	0.0391	(0.7587; 0.9120)	40	0.3367	0.0725	(0.1946; 0.4788)
8	0.3247	0.0794	(0.1691; 0.4804)	41	0.7596	0.0273	(0.7062; 0.8130)
9	0.7222	0.0753	(0.5747; 0.8698)	42	0.4261	0.0470	(0.3339; 0.5182)
10	0.6833	0.0525	(0.5803; 0.7862)	43	0.3270	0.0546	(0.2200; 0.4340)
11	0.7229	0.0722	(0.5813; 0.8644)	44	0.6143	0.0494	(0.5175; 0.7111)
12	0.4618	0.0886	(0.2882; 0.6354)	45	0.6576	0.0405	(0.5783; 0.7370)
13	0.8304	0.0243	(0.7827; 0.8781)	46	0.7557	0.0283	(0.7003; 0.8111)
14	0.7504	0.0344	(0.6829; 0.8178)	47	0.2846	0.1041	(0.0805; 0.4888)
15	0.4734	0.0839	(0.3090; 0.6378)	48	0.2148	0.0587	(0.0997; 0.3298)
16	0.8904	0.0126	(0.8658; 0.9151)	49	0.5672	0.0561	(0.4573; 0.6770)
17	0.9170	0.0081	(0.9010; 0.9329)	50	0.4956	0.0590	(0.3800; 0.6111)
18	0.9069	0.0111	(0.8852; 0.9286)	51	0.7206	0.0363	(0.6494; 0.7918)
19	0.4093	0.0934	(0.2263; 0.5923)	52	0.5687	0.0404	(0.4895; 0.6479)
20	0.8661	0.0143	(0.8379; 0.8942)	53	0.3919	0.0744	(0.2461; 0.5377)
21	0.5223	0.0688	(0.3874; 0.6572)	54	0.2699	0.0706	(0.1316; 0.4083)
22	0.7900	0.0198	(0.7513; 0.8287)	55	0.7206	0.0363	(0.6494; 0.7918)
23	0.7593	0.0338	(0.6931; 0.8256)	56	0.6355	0.0397	(0.5577; 0.7133)
24	0.7083	0.0273	(0.6549; 0.7618)	57	0.6654	0.0311	(0.6045; 0.7263)
25	0.6077	0.0484	(0.5128; 0.7027)	58	0.4457	0.0454	(0.3567; 0.5347)
26	0.8297	0.0173	(0.7957; 0.8636)	59	0.7323	0.0274	(0.6785; 0.7860)
27	0.7891	0.0318	(0.7268; 0.8515)	60	0.4699	0.0559	(0.3604; 0.5794)
28	0.7757	0.0205	(0.7356; 0.8159)	61	0.4797	0.0552	(0.3715; 0.5878)
29	0.5708	0.0440	(0.4846; 0.6571)	62	0.6557	0.0395	(0.5784; 0.7331)
30	0.6418	0.0995	(0.4468; 0.8368)	63	0.7795	0.0268	(0.7269; 0.8321)
31	0.3228	0.0685	(0.1885; 0.4570)	64	0.3483	0.1126	(0.1276; 0.5691)
32	0.4850	0.1078	(0.2738; 0.6962)	65	0.5082	0.0486	(0.4129; 0.6035)
33	0.3239	0.0625	(0.2014; 0.4464)				

Tabela C4. Intervalos de confiança para a função de sobrevivência estimada para cada paciente com endomiocardiopatia no modelo de riscos proporcionais de Cox.

IND	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]	IND	$\hat{S}(t_i)$	DP[$\hat{S}(t_i)$]	IC[$\hat{S}(t_i)$; 95%]
64	0.8837	0.1081	(0.6718; 1.0956)	118	0.9346	0.0577	(0.8214; 1.0477)
107	0.9982	0.0028	(0.9928; 1.0036)	113	0.4229	0.3336	(-0.2309; 1.0768)
29	0.7672	0.1776	(0.4191; 1.1153)	24	0.6990	0.2084	(0.2904; 1.1075)
73	0.9986	0.0016	(0.9953; 1.0018)	88	0.9576	0.0373	(0.8846; 1.0307)
65	0.8552	0.1046	(0.6501; 1.0602)	105	0.6990	0.2292	(0.2498; 1.1481)
42	0.5504	0.1969	(0.1645; 0.9363)	66	0.6701	0.2242	(0.2307; 1.1095)
151	0.9849	0.0106	(0.9640; 1.0057)	106	0.9187	0.0688	(0.7839; 1.0535)
22	0.7720	0.1155	(0.5457; 0.9983)	20	0.1285	0.1868	(-0.2376; 0.4947)
30	0.7720	0.1262	(0.5246; 1.0193)	122	0.9664	0.0254	(0.9166; 1.0162)
139	0.9972	0.0021	(0.9931; 1.0013)	85	0.6475	0.2234	(0.2096; 1.0854)
140	0.9972	0.0025	(0.9922; 1.0022)	27	0.0707	0.1296	(-0.1832; 0.3247)
14	0.7166	0.1732	(0.3771; 1.0560)	2	0.8879	0.0785	(0.7340; 1.0418)
4	0.9741	0.0207	(0.9336; 1.0147)	94	0.7259	0.1836	(0.3660; 1.0858)
148	0.9903	0.0090	(0.9727; 1.0079)	108	0.8879	0.0896	(0.7124; 1.0634)
138	0.9964	0.0039	(0.9888; 1.0040)	90	0.9621	0.0343	(0.8948; 1.0294)
13	0.6606	0.2390	(0.1922; 1.1290)	97	0.9919	0.0085	(0.9753; 1.0084)
110	0.9816	0.0179	(0.9464; 1.0167)	54	0.3694	0.3318	(-0.2808; 1.0197)
92	0.8859	0.0898	(0.7098; 1.0620)	89	0.8715	0.1150	(0.6461; 1.0968)
136	0.9332	0.0579	(0.8197; 1.0467)	19	0.0412	0.1049	(-0.1644; 0.2468)
52	0.5530	0.2515	(0.0599; 1.0460)	25	0.2312	0.2174	(-0.1950; 0.6574)
121	0.9701	0.0244	(0.9222; 1.0180)	77	0.8168	0.1179	(0.5859; 1.0478)
137	0.9936	0.0060	(0.9819; 1.0053)	53	0.8168	0.1301	(0.5619; 1.0718)
127	0.9936	0.0067	(0.9805; 1.0066)	39	0.8168	0.1430	(0.5366; 1.0971)
16	0.8461	0.1291	(0.5930; 1.0992)	41	0.9580	0.0414	(0.8769; 1.0392)
28	0.4028	0.2860	(-0.1578; 0.9633)	81	0.5798	0.3415	(-0.0895; 1.2492)
123	0.9600	0.0328	(0.8957; 1.0243)	35	0.1293	0.2035	(-0.2695; 0.5282)
47	0.8959	0.0873	(0.7247; 1.0670)	19	0.7539	0.1898	(0.3819; 1.1258)
11	0.7125	0.1877	(0.3446; 1.0805)	40	0.7539	0.2186	(0.3253; 1.1824)
75	0.9543	0.0371	(0.8815; 1.0270)	84	0.9245	0.0653	(0.7966; 1.0524)
8	0.3530	0.2668	(-0.1699; 0.8760)	36	0.6548	0.2767	(0.1125; 1.1971)
119	0.9480	0.0394	(0.8708; 1.0252)	51	0.9142	0.0994	(0.7193; 1.1091)
99	0.9413	0.0418	(0.8595; 1.0232)	6	0.9142	0.1179	(0.6831; 1.1452)
61	0.9805	0.0154	(0.9504; 1.0107)	32	0.8714	0.1984	(0.4825; 1.2604)
114	0.9484	0.0429	(0.8643; 1.0326)	7	0.8714	0.2297	(0.4212; 1.3217)
23	0.9346	0.0510	(0.8346; 1.0345)	45	0.9142	0.1812	(0.5590; 1.2694)
79	0.9782	0.0187	(0.9416; 1.0149)	18	0.0000	-	-

Tabela C5 - Tabelas de vida para pacientes com endomiocardiofibrose referente ao tipo de tratamento (TRATAMENTO) e classe funcional final (CF2).

CIRÚRGICO	CF2											
	1			2			3			4		
	E	W	b	E	W	d	E	W	d	E	W	d
Tempo												
0 - 1 m	26	0	0	34	0	0	11	0	0	10	0	2
1 - 3 m	26	0	0	34	0	0	11	0	0	8	0	6
3 - 6 m	26	1	0	34	0	1	11	0	0	2	0	0
6 - 9 m	25	3	0	33	0	0	11	0	0	2	0	0
9 -12 m	22	2	0	33	2	1	11	0	1	2	0	0
1 -1.5 a	20	3	0	30	3	0	10	0	1	2	0	1
1.5 - 2 a	17	1	0	27	3	0	9	0	1	1	0	0
2 -2.5 a	16	0	0	24	1	1	8	3	0	1	0	0
2.5 - 3 a	16	1	0	22	2	1	7	2	0	1	0	0
> 3 a	15	15	0	19	19	0	5	1	4	1	0	1

CLÍNICO	CF2											
	1			2			3			4		
	E	W	b	E	W	d	E	W	d	E	W	d
Tempo												
0 - 1 m	14	1	0	26	0	0	6	0	0	24	0	6
1 - 3 m	13	0	0	26	2	1	6	1	0	18	1	1
3 - 6 m	13	1	0	23	3	2	5	0	2	16	1	2
6 - 9 m	12	0	0	18	0	0	3	1	0	13	0	1
9 -12 m	12	2	0	18	1	0	2	0	0	12	0	1
1 -1.5 a	10	0	0	17	4	0	2	0	0	11	0	2
1.5 - 2 a	10	0	0	13	2	0	2	0	0	9	0	2
2 -2.5 a	10	3	0	11	0	0	2	0	0	7	0	1
2.5 - 3 a	7	0	0	11	3	0	2	0	0	6	0	1
> 3 a	7	7	0	8	7	1	2	2	0	5	0	5

Tabela C6 - Número de mortes e total de tempo de seguimento dos pacientes expostos ao risco no intervalo, pelo tipo de tratamento (TRATAMENTO) e classe funcional final (CF2).

tempo	(a) número de mortes				(b) total de tempo expostos ao risco			
	Cirúrgico		Clínico		Cirúrgico		Clínico	
	1-2	3-4	1-2	3-4	1-2	3-4	1-2	3-4
0 - 1 m	0	2	0	6	1800	600	1185	810
1 - 3 m	0	6	1	1	3600	960	2250	1350
3 - 6 m	1	0	2	4	5310	1170	2970	1665
6 - 9 m	0	0	0	1	5085	1170	2700	1350
9 -12 m	1	1	0	1	4725	1125	2565	1215
1 -1.5 a	0	2	0	2	8460	1980	4500	1980
1.5 - 2 a	0	1	0	2	7560	1710	3960	1620
2 -2.5 a	1	0	0	1	7020	1530	3510	1530
2.5 - 3 a	1	0	0	1	6480	1260	2970	1350
> 3 a	0	5	1	5	3060	540	1350	630

Tabela C7 - Tabelas de vida para pacientes com endomiocardiopatia referente ao tipo de tratamento (TRATAMENTO) e fibrilação do atrio (FA).

Tempo	FA	CIRÚRGICO						CLÍNICO					
		1			2			1			2		
		E	W	b	E	W	d	E	W	d	E	W	d
0 - 1 m		24	0	2	50	0	0	25	0	4	43	1	2
1 - 3 m		22	0	4	50	0	2	21	1	1	40	3	1
3 - 6 m		18	0	1	48	1	0	19	3	4	36	1	2
6 - 9 m		17	1	0	47	2	0	12	0	1	33	1	0
9 -12 m		16	1	0	45	1	2	11	1	1	32	0	0
1 -1.5 a		15	1	0	42	3	2	9	0	1	32	4	1
1.5 - 2 a		14	1	0	37	3	1	8	1	0	27	3	2
2 -2.5 a		13	1	0	33	2	1	7	0	1	22	2	0
2.5 - 3 a		12	2	0	30	2	1	6	1	0	20	2	0
> 3 a		10	10	0	27	22	5	5	2	3	18	13	5

Tabela C8 - Número de mortes e total de tempo de seguimento dos pacientes expostos ao risco no intervalo, pelo tipo de tratamento (TRATAMENTO) e fibrilação do atrio (FA).

tempo	(a) número de mortes				(b) total de tempo expostos ao risco			
	Cirúrgico		Clínico		Cirúrgico		Clínico	
	1	2	1	2	1	2	1	2
0 - 1 m	2	0	4	2	690	1500	690	1245
1 - 3 m	4	2	1	1	1200	2940	1200	2280
3 - 6 m	1	0	4	2	1575	4275	1395	3105
6 - 9 m	0	0	1	0	1485	4140	1035	2925
9 -12 m	0	2	1	0	1395	3915	900	2880
1 -1.5 a	0	2	1	1	2610	7110	1530	5310
1.5 - 2 a	0	1	0	2	2430	6300	1350	4410
2 -2.5 a	0	1	1	0	2250	5670	1170	3780
2.5 - 3 a	0	1	0	0	2070	5130	990	3420
> 3 a	0	5	3	5	900	2430	450	1620

Tabela C9 - Matriz de planejamento para ajuste dos modelos no Capítulo 4.

(a) Matriz de planejamento para efeitos principais do tempo, tipo de tratamento (TRATAMENTO) e classe funcional final (CF2) ou fibrilação do atrio (FA).

$$X = \{1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0, \\
1 \ 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0, \\
1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0, \\
1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0, \\
1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1, \\
1 \ 1 \ 1 \ -1 \ -1 \ -1 \ -1 \ -1, \\
1 \ 1 \ -1 \ 1 \ 0 \ 0 \ 0 \ 0, \\
1 \ 1 \ -1 \ 0 \ 1 \ 0 \ 0 \ 0, \\
1 \ 1 \ -1 \ 0 \ 0 \ 1 \ 0 \ 0, \\
1 \ 1 \ -1 \ 0 \ 0 \ 0 \ 1 \ 0, \\
1 \ 1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 1, \\
1 \ 1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1, \\
1 \ -1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0, \\
1 \ -1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0, \\
1 \ -1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0, \\
1 \ -1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0, \\
1 \ -1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1, \\
1 \ -1 \ 1 \ -1 \ -1 \ -1 \ -1 \ -1, \\
1 \ -1 \ -1 \ 1 \ 0 \ 0 \ 0 \ 0, \\
1 \ -1 \ -1 \ 0 \ 1 \ 0 \ 0 \ 0, \\
1 \ -1 \ -1 \ 0 \ 0 \ 1 \ 0 \ 0, \\
1 \ -1 \ -1 \ 0 \ 0 \ 0 \ 1 \ 0, \\
1 \ -1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 1, \\
1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1\};$$

Tabela C9 - (Continuação).

(b) Matriz de planejamento para o modelo hierárquico.

$$\begin{aligned}
 X = \{ & 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0, \\
 & 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ -1 \ 0, \\
 & 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ -1 \ 0 \ 1, \\
 & 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ -1 \ 0 \ -1, \\
 & 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0, \\
 & 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ -1 \ 0, \\
 & 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ -1 \ 0 \ 1, \\
 & 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ -1 \ 0 \ -1, \\
 & 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0, \\
 & 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ -1 \ 0, \\
 & 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ -1 \ 0 \ 1, \\
 & 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ -1 \ 0 \ -1, \\
 & 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0, \\
 & 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 1 \ -1 \ 0, \\
 & 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ -1 \ 0 \ 1, \\
 & 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ -1 \ 0 \ -1, \\
 & 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0, \\
 & 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ -1 \ 0, \\
 & 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ -1 \ 0 \ 1, \\
 & 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ -1 \ 0 \ -1, \\
 & 1 \ -1 \ -1 \ -1 \ -1 \ 1 \ 1 \ 0, \\
 & 1 \ -1 \ -1 \ -1 \ -1 \ 1 \ -1 \ 0, \\
 & 1 \ -1 \ -1 \ -1 \ -1 \ -1 \ 0 \ 1 \\
 & 1 \ -1 \ -1 \ -1 \ -1 \ -1 \ 0 \ -1 \};
 \end{aligned}$$

APÊNDICE D

PROGRAMAS COMPUTACIONAIS

Neste apêndice encontram-se os programas computacionais utilizados para a obtenção dos resultados desta tese. Utilizou-se o *software* estatístico SAS (Statistical Analysis System). Na secção D1 apresenta-se uma seqüência de procedimentos SAS utilizados no ajuste de alguns dos modelos. Na secção D2, uma introdução aos programas complementares desenvolvidos no módulo IML do SAS. Finalmente na secção D3 apresentam-se os programas. Quando necessário serão feitos breves comentários.

D1. SAS (Statistical Analysis System)

SAS (SAS Institute 1987), é um pacote desenvolvido por meio de procedimentos. Os procedimentos tratados nesta tese são:

Procedimento LIFETEST: é utilizado quando os dados apresentam censuras à direita para o cálculo de estimativas não paramétricas da função distribuição de sobrevivência, tal como produto-limite de Kaplan-Meier e tabela de vida introduzidos no Capítulo 1. Também fornece teste de associação entre a variável resposta e outras covariáveis.

Procedimento **LIFEREG**: é usado para ajuste de modelos paramétricos onde a variável resposta é o tempo, que podem ser censurados à direita. A classe de modelo inclui exponencial, Weibull, log-normal e modelos log logístico. Os parâmetros são estimados através da máxima verossimilhança usando o algoritmo de Newton-Raphson. As estimativas do erro padrão dos parâmetros estimados são calculados da inversa da matriz de informação observada.

Procedimento **LOGISTIC**: permite ajustar modelos de regressão logística para dados binários e dados de resposta ordinal pelo método de máxima verossimilhança usando o método iterativo de mínimos quadrados ponderados para cálculo das estimativas dos parâmetros. O procedimento incorpora procedimentos para selecionar covariáveis para obter o "melhor modelo". As opções de tais procedimentos são **FORWARD** (parte do modelo mais simples em direção aos modelos mais complexos), **BACKWARD** (parte do modelo mais complexo em direção aos modelos mais simples) e **STEPWISE** (é a composição dos métodos **FORWARD** e **BACKWARD**, que inclui ou exclui as covariáveis conforme sua importância).

Procedimento **PHREG**: é empregado em análise de dados de sobrevivência, quando a distribuição dos tempos é desconhecida. Ajusta o modelo semi-paramétrico de Cox. O modelo de riscos proporcionais de Cox é muito usado para explicar os efeitos das variáveis exploratórias no tempo de sobrevivência. Usa o método da verossimilhança parcial de Cox para estimar os parâmetros. O procedimento **PHREG** fornece quatro métodos de seleção de variáveis: Seleção **FORWARD**, eliminação **BACKWARD**, seleção **STEPWISE** e seleção "BEST". Também fornece teste de linearidade nos parâmetros e cria um novo conjunto de dados contendo a função de sobrevivência estimada.

D2. Programas Desenvolvidos.

Os programas abaixo relacionados são todos desenvolvidos no módulo **IML** do **SAS**.

Programa **PEVSTMC.PRG**: Programa desenvolvido para estimar a variância da função de sobrevivência através da fórmula de Tsiatis expressada em (1.51), apresentada por Link [1984] e construir intervalos $100(1 - \alpha)\%$ de confiança para a função de verossimilhança quando os parâmetros são estimados através do modelo de riscos proporcionais de Cox.

Programa **PESVSMG.PRG**: Programa desenvolvido para estimar a função de Sobrevivência dada em (1.42) e a respectiva variância, empregando o método Delta apresentado na secção 1.6 no modelo de riscos proporcionais de Glasser, e construir intervalos de $100(1 - \alpha)\%$ confiança.

Programa **PESVSLAS.PRG**: Programa desenvolvido para estimar a função de

sobrevivência e a variância da mesma no modelo logístico em análise de sobrevivência, apresentadas na secção 2.3. Calcula também intervalos de $100(1 - \alpha)\%$ de confiança.

Programa **PAMLEAS.PRG**: Programa desenvolvido para ajuste do modelo logístico exponencial em análise de sobrevivência. Desenvolvido para estimar os parâmetros através do método de máxima verossimilhança empregando o algoritmo de Newton-Raphson, e calcular as estimativas da função de sobrevivência e as respectivas variâncias, para construir intervalos de $100(1 - \alpha)\%$ de confiança apresentados na secção 2.4.

Programa **PAMLEPAS.PRG**: Programa desenvolvido para ajuste do modelo logístico exponencial por partes em análise de sobrevivência. Estima os parâmetros através do método de máxima verossimilhança usando o algoritmo de Newton-Raphson, assim como fornece estimativas para a função de sobrevivência e sua variância, apresentadas na secção 2.5. Constrói também, intervalos de $100(1 - \alpha)\%$ de confiança.

Programa **PAMEPTV.PRG**: Programa desenvolvido para ajuste do modelo exponencial por partes para análise de sobrevivência quando os dados estão apresentados em tabelas de vida como mostra a Tabela 3.1 e as covariáveis são categóricas. Este programa estima os parâmetros através do método iterativo de Newton-Raphson para encontrar estimador de máxima verossimilhança. Estima o logaritmo da função de sobrevivência dada em (3.16) e controla intervalos de $100(1 - \alpha)\%$ de confiança para logaritmo de $S(t | \mathbf{x})$.

Programa **PAMEP.PRG**: Programa desenvolvido para ajuste do modelo exponencial por partes para análise de sobrevivência quando as covariáveis são contínuas e/ou discretas. Este programa estima os parâmetros através do método de Newton-Raphson. Estima o logaritmo da função de sobrevivência dada em (3.16) e controla intervalos de $100(1 - \alpha)\%$ de confiança para logaritmo de $S(t | \mathbf{x})$.

Programa **PAMLOGHM.PRG**: Programa desenvolvido para ajuste do modelo multinomial sugerido por Holford [1980]. Este programa estima os parâmetros através do método de Newton-Raphson.

D3. Listagens dos Programas.

D3.1. Programa em SAS (LIFETEST) para calcular a estimativa da função de sobrevivência através de métodos não paramétricos e desenhar os gráficos de interesse.

```
LIBNAME A ' ';
PROC LIFETEST DATA = a.dados PLOTS = (S,LS,LLS);
    TIME TEMPO*STATUS(0);
    STRATA var;                               /* valido somente quando os dados estao divididos
                                                em estratos - Capitulo 4 */
RUN;
```

Obs: a.dados é um SAS data set com os dados.

D3.2. Programa em SAS (LIFEREG) para ajustar o modelo paramétrico log-linear de Glasser.

```
LIBNAME A ' ';
PROC LIFEREG DATA = a.dados;
    MODEL TEMPO*STATUS(0) = covariaveis/ DIST = EXPONENTIAL
                                                COVB;
                                                /* especifica o modelo a ser ajustado */
    OUTPUT OUT = a.dados1 XBETA = zb STD = dpzb;
                                                /* cria um novo conjunto de dados com os valores
                                                preditos zb e seus respectivos erros padrao */
RUN;
```

Obs: - covariáveis são os nomes das covariáveis que serão incluídas no modelo, para o ajuste do modelo sem covariáveis deixar em branco.

- as opções: DIST = EXPONENTIAL -mostra que o modelo a ser ajustado é o modelo log-linear exponencial.

COVB - escreve a matriz de covariância dos parâmetros estimados.

D3.3. Programa PESVSMG.PRG (IML-SAS), desenvolvido para calcular a estimativa da função de sobrevivência e de sua variância quando o modelo ajustado é o modelo de riscos proporcionais exponencial de Glasser [1967].

```

/*----- PESVSMG.PRG----- versao 1.0 - Junho 93 ---- */
/* PROGRAMA PARA ESTIMAR: FUNCAO DE SOBREVIVENCIA          */
/*              VARIANCIA DA FUNCAO DE SOBREVIVENCIA*/
/*----- */
/*----- MODELO LOG-LINEAR DE GLASSER----- */
/*----- */
/* ENTRADA DOS DADOS: A = matriz com os dados            */
/*              B = vetor dos parametros estimados (atraves de D3.2) */
/*              VB = matriz de covariancia estimada (atraves de D3.2) */
/*----- */

PROC IML WORKSIZE = 50;
RESET NOPRINT;
/*----- - leitura dos dados ----- */
A = { conjunto de dados };

B = {4.4806, -1.5414, 0.1041};          /* parametros estimados atraves de D3.2 */
VB = { 0.989862 -0.452383 -0.034167,
      - 0.452383 0.307992 0.002339,
      - 0.034167 0.002339 0.003119};   /* matriz de covariancia dada por D3.2 */

/*----- - matrizes auxiliares utilizadas nos calculos ----- */
n = NROW(A);          /* numero de observacoes (por exemplo pacientes) */
UM = J(n,1,1);       /* vetor nx1 de elementos 1 */
Z = UM || A[,{4 5}]; /* matriz para ajuste do modelo */
p = NCOL(Z);        /* numero de parametros a ser estimado */
DELTA1 = A[,{3}];   /* vetor nx1 com indicador de censura */
T = A[,{2}];        /* vetor nx1 com os Tempos */
ZB = A[,{7}];       /* vetor nx1 dos preditos zb (atraves de D3.2) */
IND = A[,{1}];      /* vetor nx1 identificacao do paciente */
DPB = VECDIAG(VB)##0.5; /* erro padrao para os parametros */

```

```

/*-----*/
/*          CALCULO DA ESTIMATIVA DA FUNCAO DE SOBREVIVENCIA          */
/*-----*/
AUX1 = EXP(- ZB);
S = EXP(- T#AUX1);          /* funcao sobrevivencia estimada */
/*-----*/
/*          CALCULO DA ESTIMATIVA DA VARIANCIA          */
/*          METODO DELTA          */
/*-----*/
GB = J(p,n,0);
VS = J(n,1,0);
  DO I = 1 TO N;
    DO J = 1 TO P;
      Gb[J,I] = GB[J,I] + T[I,]*S[I,]*Z[I,J]*AUX1[I,];          /* derivada S(t | z) em b */
    END;
    VS[I,] = VS[I,] + ((GB[I,])' * VB * (GB[I,]));          /* variancia de S(t | z) */
  END;
DPS = VS##0.5;          /* erro padrao de S(t | z) */
/*-----*/
/*          Intervalo de 95% de confianca para a funcao sobrevivencia          */
/*          no modelo log-linear de Glasser [1967]          */
/*-----*/
AUX2 = 1.96;          /* valor tabelado da normal padrao com probabilidade 95 % */
ICLI = J(N,1,0);
ICLS = J(N,1,0);
  DO I = 1 TO N;
    ICLI[I,] = ICLI[I,] + S[I,] - AUX2*DPS[I,];          /* limite inferior */
    ICLS[I,] = ICLS[I,] + S[I,] + AUX2*DPS[I,];          /* limite superior */
  END;
/*-----*/
/*          IMPRESSAO DOS RESULTADOS          */
/*-----*/
VARIABEL = { "INTERCEPTO","nomes das covariaveis includas no modelo" };
PRINT , " MODELO DE REGRESSAO LOG-LINEAR DE GLASSER ";
PRINT, " ";

```

```

PRINT, " ";
PRINT, "PARAMETROS ESTIMADOS E ERRO PADRAO";
PRINT, VARIABEL B [FORMAT = 8.4] DPB [FORMAT = 8.4];
PRINT, " ";
PRINT / "FUNCAO DE SOBREVIVENCIA ESTIMADA, ERRO PADRAO DE S(t | z) E";
PRINT "INTERVALO DE CONFIANCA DE 95% PARA S(t | z)";
PRINT, " ";
PRINT, OBS T S [FORMAT=8.4] DPS [FORMAT=8.4] ICLI [FORMAT=8.4]
      ICLS [FORMAT=8.4];
QUIT;

```

Obs: Este programa só pode ser utilizado após a execução do programa D3.2.

D3.4. Programa em SAS (PHREG) para ajuste do modelo de risco proporcional de Cox.

```

LIBNAME A ' ';
PROC PHREG DATA = a.dados;
      MODEL TEMPO*STATUS(0) = covariáveis/COVB
                                SELECTION = STEPWISE
                                SLENTRY = 0.25
                                SLSTAY = 0.15 DETAILS;
      OUTPUT OUT = a.dados1 XBETA = zb STDXBETA = dpzb SURVIVAL = s;
      /* cria um novo conjunto de dados com os valores preditos zb
      e seus erros padrao e a estimativa da funcao de sobrevivencia */
RUN;

```

D3.5. Programa PEVSTMC.PRG (IML-SAS), desenvolvido para calcular a estimativa da variância da função de sobrevivência quando o modelo ajustado é o modelo de riscos proporcionais de Cox.

```

/*----- -PEVSTMC.PRG----- versao 1.0 - Junho 93----- */
/* PROGRAMA PARA ESTIMAR: VARIANCA DA FUNCAO DE SOBREVIVENCIA*/
/*----- - */
/*-----MODELO DE RISCOS PROPORCIONAIS DE COX -----*/

```

```

/*----- - */
/* ENTRADA DOS DADOS: A = matriz com os dados */
/*          B = vetor dos parametros estimados (atraves de D3.4) */
/*          VB = matriz de covariancia estimada (atraves de D3.4) */
/*----- - */

PROC IML WORKSIZE = 50;
RESET NOPRINT;

/*----- - leitura dos dados ----- - */
A = { conjunto de dados };
B = {1.674399, -0.118987};          /* parametros estimados atraves de D3.4 */
VB = {0.374649 0.001273,
      0.001273 0.003307};          /* matriz de covariancia dada por D3.4 */

/*----- - matrizes auxiliares utilizadas nos calculos ----- - */
N = NROW(A);          /* numero de observacoes (por exemplo pacientes) */
Z = A[,{4 5}];          /* matriz para ajuste do modelo */
T = A[,{2}];          /* vetor nx1 com os Tempos ordenados t1 < t2 < ... < tn */
DELTAI = A[,{3}];          /* vetor nx1 com indicador de censura */
ZB = A[,{6}];          /* vetor nx1 dos preditos zb (atraves de D3.4) */
S = A[,{8}];          /* vetor nx1 das sobrevivencias estimadas atraves de D3.4 */
IND= A[,{1}];          /* vetor nx1 a identificacao do paciente */
P = NCOL(Z);          /* numero de parametros a ser estimado */
DPB = VECDIAG(VB)##0.5;          /* erro padrao para os parametros */

/*----- - */
/*          CALCULO DA VARIANCIA SEGUNDO TSIATIS [1978] */
/*----- - */

VS = J(n,1,0);
W = J(p,1,0);
AUX3 = 0;
DO I = 1 TO N;
  AUX1 = 0;
  AUX2 = J(p,1,0);
  DO L = I TO N;
    IF T[L,] >= T[I,] THEN DO;

```

```

AUX1 = AUX1 + EXP(Z[L,]*B);
AUX2 = AUX2 + Z[L,]' *EXP(Z[L,]*B);
END;
END;
DO J = 1 TO P;
  W[J,] = W[J,] + AUX2[J,]/(AUX1##2) - Z[L,J]/AUX1;
END;
AUX3 = AUX3 + 1/(AUX1##2);

VS[L,] = VS[L,] + ((S[L,]##2)*(EXP(2*Z[L,]*B)*AUX3 + EXP(2*Z[L,]*B)*W' * VB*W));
/* variancia de S(t | z) */

END;
DPS = VS##0.5; /* erro padrao de S(t | z) */

/*----- */
/* Intervalo de 95% de confianca para a funcao sobrevivencia */
/* no modelo regressao de Cox [1972] */
/*----- */
aux4 = 1.96; /* valor tabelado da normal padrao com probabilidade 95% */
ICLI = J(N,1,0);
ICLS = J(N,1,0);
DO I=1 TO N;
  ICLI[I,] = ICLI[I,] + S[I,] - AUX4*DPS[I,]; /* limite inferior */
  ICLS[I,] = ICLS[I,] + S[I,] + AUX4*DPS[I,]; /* limite superior */
END;

/*----- */
/* IMPRESSAO DOS RESULTADOS */
/*----- */
VARIABEL = { "nomes das covariaveis incluidas no modelo"};
PRINT , " MODELO DE REGRESSAO DE COX ";
PRINT , " ";
PRINT , " ";
PRINT , "PARAMETROS ESTIMADOS E ERRO PADRAO";
PRINT , VARIABEL B [FORMAT=8.4] DPB [FORMAT=8.4];
PRINT , " ";

```

```

PRINT /"FUNCAO DE SOBREVIVENCIA ESTIMADA , ERRO PADRAO DE S(t | z) E";
PRINT "INTERVALO DE CONFIANCA DE 95% PARA S(t | z)";
PRINT , " ";
PRINT ,IND T S [FORMAT=8.4] DPS [FORMAT=8.4] ICLI [FORMAT=8.4]
      ICLS [FORMAT=8.4];
QUIT;

```

Obs: Este programa só pode ser utilizado após a execução do programa D3.4.

D3.6. Programa em SAS (LOGISTIC) para ajuste do modelo logístico.

```

LIBNAME A ' ';
PROC LOGISTIC DATA = a.dados;
      MODEL RESPOSTA = covariáveis / COVB
                        SELECTION = STEPWISE
                        SLENTRY = 0.25
                        SLSTAY = 0.15 DETAILS;
      OUTPUT OUT = a.dados1 XBETA = zb STDXBETA = dpzb;
                        /* cria um novo conjunto de dados com os valores preditos zb
                        e seus erros padrao */
RUN;

```

Obs: RESPOSTA é a variável dicotômica descrita em (2.20).

D3.7. Programa PESVSLAS.PRG (IML-SAS), desenvolvido para calcular a estimativa da função de sobrevivência e de sua variância quando o modelo ajustado é o modelo logístico.

```

/*----- PESVSLAS.PRG ----- versão 1.0 - Junho 93 ---- */
/* PROGRAMA PARA ESTIMAR: FUNCAO DE SOBREVIVENCIA */
/* VARIANCIA DA FUNCAO DE SOBREVIVENCIA */
/*----- - */
/*----- MODELO LOGISTICO ----- - */
/*----- - */
/* ENTRADA DOS DADOS: A = matriz com os dados */
/* B = vetor dos parametros estimados (atraves de D3.6) */
/* VB = matriz de covariância estimada (atraves de D3.6) */
/*----- - */

```

```

PROC IML WORKSIZE = 50;
RESET NOPRINT;

/*----- leitura dos dados ----- */
A = { conjunto de dados };
B = {10.6866,-3.4352,-2.0750,-0.3746}; /* parametros estimados atraves de D3.6 */
VB = {11.872648049 -3.776101392 -1.912350299 -0.45970572,
      -3.776101392 1.8814219429 0.4094242542 0.0794092465,
      -1.912350299 0.4094242542 0.5844034202 0.055785167,
      -0.45970572 0.0794092465 0.055785167 0.0324035679};
/* matriz de covariancia dada por D3.6 */

/*----- matrizes auxiliares utilizadas nos calculos ----- */
n = NROW(A); /* numero de observacoes (por exemplo pacientes) */
UM = J(n,1,1); /* vetor nxl de elementos 1 */
Z = UM || a[,{3 15 16}]; /* matriz para ajuste do modelo */
p = NCOL(Z); /* numero de parametros a ser estimado */
DELTAI = A[,2]; /* vetor nxl com indicador de censura */
T = A[,1]; /* vetor nxl com os Tempos */
DPB = VECDIAG(VB)##0.5; /* erro padrao para os parametros */

/*----- VALORES PREDITOS ----- */
P0 = UM/(UM + exp(-Z*B)); /* Probabilidade de Falha */
ZB = Z*B; /* preditor linear */
VZB = Z*HINV*Z' ; /* matriz de covariancia de ZB */
DPZB = VECDIAG(VZB)##0.5; /* desvio padrao de ZB */

/*-----
/* CALCULO DA ESTIMATIVA DA FUNCAO DE SOBREVIVENCIA
/*-----
S = UM /(UM + EXP(ZB)); /* funcao de sobrevivencia estimada */

/*-----
/* CALCULO DA ESTIMATIVA DA VARIANCIA
/*
/* METODO DELTA
/*-----
GB = J(p,n,0);
VS = J(N,1,0);
DO I = 1 TO N;

```

```

DO J = 1 TO P;
  GB[J,I] = GB[J,I] - (Z[I,J]*S[I,1]/(1 + EXP(Z[I,J]*B))); /* derivada S(t | z) em b */
END;
VS[I,] = VS[I,] + ((GB[I,])'*VB*(GB[I,])); /* variancia de S(t | z) */
END;
DPS = VS##0.5; /* erro padrao de S(t | z) */
/*----- - */
/* Intervalo de 95% de confianca para a funcao sobrevivencia */
/* no modelo logistico */
/*----- - */
aux3 = 1.96; /* valor tabelado da normal padrao com probabilidade 95% */
ICLI = J(N,1,0);
ICLS = J(N,1,0);
DO I = 1 TO N;
  ICLI[I,] = ICLI[I,] + S[I,] - AUX3*DPS[I,]; /* limite inferior */
  ICLS[I,] = ICLS[I,] + S[I,] + AUX3*DPS[I,]; /* limite superior */
END;
IND = J(N,1,0);
DO I = 1 TO N;
  IND[I,] = I;
END;
/*----- - */
/* IMPRESSAO DOS RESULTADOS */
/*----- - */
VARIAVEL = { "INTERCEPTO", "nomes das covariaveis includas no modelo" };
PRINT / "MODELO DE REGRESSAO LOGISTICO ";
PRINT, " ";
PRINT, " ";
PRINT, "PARAMETROS ESTIMADOS E ERRO PADRAO ";
PRINT, VARIAVEL B [FORMAT=8.4] VB [FORMAT=8.4];
PRINT, " ";
PRINT / "VALORES PREDITOS";
PRINT, ZB [FORMAT=8.4] DPZB [FORMAT=8.4];
PRINT / "FUNCAO DE SOBREVIVENCIA ESTIMADA, ERRO PADRAO DE S(t | z) E";
PRINT "INTERVALO DE CONFIANCA DE 95% PARA S(t | z)";

```

```
PRINT ,IND T [FORMAT = 5.2] P0 [FORMAT=8.4] S [FORMAT=8.4]
      DPS [FORMAT = 8.4] ICLI [FORMAT=8.4] ICLS [FORMAT=8.4];
QUIT;
```

Obs: Este programa só pode ser utilizado após a execução do programa D3.6.

D3.8. Programa PAMLEAS.PRG (IML-SAS), desenvolvido para ajustar o modelo logístico exponencial.

```
/*----- - PAMLEAS.PRG----- versao 1.0 - Junho 93 --- */
/* PROGRAMA PARA ESTIMAR: PARAMETROS */
/* FUNCAO DE SOBREVIVENCIA */
/* VARIANCIA DA FUNCAO DE SOBREVIVENCIA */
/*----- - */
/*----- MODELO LOGISTICO EXPONENCIAL -----*/
/*----- - */
/* ENTRADA DOS DADOS: A = matriz com os dados */
/* TAU = tempo de observacao */
/*----- - */

PROC IML WORKSIZE = 50;
RESET NOPRINT;
/*----- - leitura dos dados ----- - */
A = { conjunto de dados };
TAU = 20;
/*----- matrizes auxiliares utilizadas nos calculos ----- - */
n = NROW(A); /* numero de observacoes (por exemplo pacientes) */
UM = J(n,1,1); /* vetor nx1 de elementos 1 */
Z = UM || a[3 15]; /* matriz para ajuste do modelo */
p = NCOL(Z); /* numero de parametros a ser estimado */
STATUS = A[2]; /* vetor nx1 com indicador de censura */
T1 = A[1]; /* vetor nx1 com os Tempos */
/*----- - definindo variavel resposta deltai e a variavel tempo ----- - */
DELTAI = J(N,1,0);
T = J(N,1,0);
```

```

DO I=1 TO N;
  IF T1[I.] < TAU THEN DO;
    T[I.] = T1[I.];
    IF STATUS[I.] = 1 THEN DELTAI[I.] = 1;
    IF STATUS[I.] = 0 THEN DELTAI[I.] = 2;
  END;
  ELSE DO;
    T[I.] = TAU;
  END;
END;

/* ----- - */
/*           ESTIMACAO DOS PARAMETROS           */
/* ----- - */
/* ----- - valores iniciais ----- */
INA0 = J(p,1,0);
PARADA = J(p,1,1);
B = INA0; /* valores iniciais para vetor dos parametros */
it = 0;
LAMBDA = 1;
/* ----- - funcao para calcular as derivadas de MV ----- - */
START DERIVADA;
P0 = J(N,1,0);
AUX1 = J(N,1,0);
UB = J(P,1,0);
I0B = J(P,P,0);
DO I = 1 TO N;
  IF DELTAI[I.] = 2 THEN DO;
    P0[I.] = 1/(1+EXP(-Z[I.]*B)); /* probabilidade de falha */
    AUX1[I.] = 1/(LOG(1+EXP(Z[I.]*B)));
    UB = UB + (Z[I.]*P0[I.]+(DELTAI[I.]*AUX1[I.] - T[I.]/TAU));
    /* derivada em relacao a B */
    AUX2 = EXP(-Z[I.]*B)*(DELTAI[I.]*AUX1[I.] - T[I.]/TAU) - DELTAI[I.]*AUX1[I.]**2;
    I0B = I0B + Z[I.]*P0[I.]**2*AUX2*Z[I.];
    /* derivada parcial de segunda ordem em B */
  END;
END;

```

```

END;
FINISH;
/*----- -funcao para calcular funcao maxima verossimilhanca ----- */
START MV;
  LV = 0;
  DO I = 1 TO N;
    IF DELTAI[I] = 2 THEN DO;
      LV = LV + DELTAI[I]*(LOG(LOG(1 + EXP(Z[I]*B))) - DELTAI[I]*LOG(TAU)

          (T[I]/TAU)*LOG(1 + EXP(Z[I]*B)));      /* funcao de verossimilhanca */
    END;
  END;
FINISH;
/*----- - processo iterativo de Newton-Raphson ----- */
DO WHILE ((MAX(ABS(PARADA)) > 0.0001));          /* criterio de parada */
  RUN DERIVADA;                                /* executa a funcao derivada */
  RUN MV;                                       /* executa a funcao maxima verossimilhanca */
  LVA = LV;
  U0 = UB ;
  I0 = I0B;
  HINV = INV((-1)*I0);                         /* inversa da matriz de Fisher */
  DELTA0 = HINV*U0;
  B1 = B;
  PARADA = LAMBDA*DELTA0;
  B = B + PARADA;                              /* vetor dos parametros */
  RUN MV;                                       /* executa a funcao maxima verossimilhanca */

  IF LV <= LVA THEN DO;
    CONT = 0;
    DO WHILE ((LV <= LVA) & (CONT <= 10));
      LAMBDA = LAMBDA/2;
      PARADA = LAMBDA*DELTA0;
      B = B1 + PARADA;
      RUN MV;                                  /* executa a funcao maxima verossimilhanca */
      CONT = CONT + 1;
    END;
  END;

```



```

/*-----*/ - */
/*  CALCULO DA ESTIMATIVA DA FUNCAO DE SOBREVIVENCIA  */
/*-----*/ - */
S = J(N,1,0);
DO I = 1 TO N;
    S[I.] = (1 + EXP(ZB[I.]))##(-T[I.]/TAU);      /* funcao de sobrevivencia estimada*/
END;

/*-----*/ - */
/*          CALCULO DA ESTIMATIVA DA VARIANCIA          */
/*          METODO DELTA                                */
/*-----*/ - */
GB = J(P,N,0);
VS = J(N,1,0);
DO I = 1 TO N;
    DO J=1 TO P;
        GB[J,I] = GB[J,I] - (Z[I,J]*(T[I.]/TAU)*EXP(ZB[I.] + ((1 + EXP(ZB[I.]))##(-T[I.]/TAU - 1)));
                                                    /* derivada S(t | z) em b */
    END;
END;
DO I= 1 TO N ;
    VS[I.] = VS[I.] + (GB[I.]' * VB * GB[I.]);      /* variancia de S(t | z) */
END;
DPS = VS##0.5;      /* erro padrao de S(t | z) */
/*-----*/ - */
/*          Intervalo de 95% de confianca para a funcao sobrevivencia          */
/*          no modelo logistico                                                */
/*-----*/ - */
AUX3 = 1.96;      /* valor tabelado da normal padrao com probabilidade 95% */
ICLI = J(N,1,0);
ICLS = J(N,1,0);
DO I = 1 TO N;
    ICLI[I.] = ICLI[I.] + S[I.] - AUX3*DPS[I.];      /* limite inferior */
    ICLS[I.] = ICLS[I.] + S[I.] + AUX3*DPS[I.];      /* limite superior */
END;
IND = J(N,1,0);

```

```

DO I = 1 TO N;
  IND[1,] = 1;
END;

/*-----*/
/*          IMPRESSAO DOS RESULTADOS          */
/*-----*/

VARIAVEL = { "INTERCEPTO", "nomes das covariaveis incluidas no modelo" };
PRINT / "MODELO DE REGRESSAO LOGISTICO EXPONENCIAL ";
PRINT, " ";
PRINT, " ";
PRINT, "PARAMETROS ESTIMADOS E ERRO PADRAO";
PRINT, VARIAVEL B [FORMAT=8.4] DPB [FORMAT=8.4];
PRINT, " ";
PRINT, "VEROSSIMILHANCA";
PRINT, "ITF VL";
PRINT / " VALORES PREDITOS ";
PRINT, ZB [FORMAT=8.4] DPZB [FORMAT=8.4];
PRINT / "FUNCAO DE SOBREVIVENCIA ESTIMADA, ERRO PADRAO DE S(t | z) E";
PRINT "INTERVALO DE CONFIANCA DE 95% PARA S(t | z)";
PRINT IND T [FORMAT=5.2] S [FORMAT=8.4] DPS [FORMAT=8.4]
      ICLI [FORMAT=8.4] ICLS [FORMAT=8.4];
QUIT;

```

D3.9. Programa PAMLEPAS.PRG (IML-SAS), desenvolvido para ajustar o modelo logístico exponencial por partes.

```

/*----- PAMLEPAS.PRG ----- versao 1.0-Junho 93 ---- */
/* PROGRAMA PARA ESTIMAR: PARAMETROS          */
/*          FUNCAO DE SOBREVIVENCIA          */
/*          VARIANCA DA FUNCAO DE SOBREVIVENCIA */
/*-----*/
/*----- MODELO LOGISTICO EXPONENCIAL POR PARTES -----*/
/*-----*/

```

```

/* ENTRADA DOS DADOS: A = matriz com os dados */
/*
/*          T1 = pontos fixos de tempos de observacao */
/*----- - */

PROC IML WORKSIZE = 100;
RESET NOPRINT;

/*----- - leitura dos dados ----- - */
A = { conjunto de dados };
T1 = {0,10,20};

/*----- - matrizes auxiliares utilizadas nos calculos ----- - */
n = NROW(A); /* numero de observacoes (por exemplo pacientes) */
Z = a[,{3 6 15 16}]; /* matriz das covariaveis */
p = NCOL(Z); /* numero de covariaveis no modelo */
STATUS = A[,2]; /* vetor nx1 com indicador de censura */
TEMPO = A[,1]; /* vetor nx1 com os Tempos */
y = NROW(T1); /* numero de pontos fixos no tempo */

/*----- - definindo algumas variaveis ----- - */
Z1 = J(n,(y-1),0); /* interceptos para cada intervalo */
TIK = J(n,(y-1),0); /* tempos de falhas dentro de cada intervalo */
DELTAI = J(n,(y-1),0); /* variavel indicadora de censura em cada intervalo */
/*----- - definindo matriz de interceptos ----- */
DO K = 1 TO (y-1);
  DO I = 1 TO n;
    IF TEMPO[I,] >= T1[K,] THEN DO;
      Z1[I,K] = 1;
    END;
    ELSE DO;
      Z1[I,K] = 0;
    END;
  /*----- - definindo matriz de tempos ----- - */
  IF (TEMPO[I,] >= T1[K,] & TEMPO[I,] < T1[(k+1),]) THEN DO;
    TIK[I,K] = TEMPO[I,];
  END;
  IF (TEMPO[I,] >= T1[(k+1),] ) THEN DO ;
    TIK[I,K] = T1[(k+1),];

```

```

END;
IF (TEMPO[L] < T1[K,]) THEN DO;
    TIK[I,K] = 0;
END;

/*----- - definindo variavel resposta ----- */
IF TEMPO[I] >= T1[K,] THEN DO;
    IF TEMPO[I] < T1[(k+1),] THEN DO;
        IF STATUS[I] = 1 THEN DELTAI[I,K] = 1;
        IF STATUS[I] = 0 THEN DELTAI[I,K] = 2;
    END;
END;
IF TEMPO[I] < T1[K,] THEN DO;
    DELTAI[I,K] = 2;
END;
END;

END;

/*----- -definindo matriz dos TAUIK----- */
TAUIK = J(n,(y-1),0);
DO K = 1 TO (y-1);
    DO I = 1 TO N;
        TAUIK[I,K] = TIK[I,K] - T1[K,];
        IF TAUIK[I,K] < 0 THEN TAUIK[I,K] = 0;
    END;
END;

/*----- - definindo comprimento do intervalo----- */
TAU = J((y-1),1,0);
DO K = 1 TO (y-1);
    TAU[K,] = T1[(k+1),] - T1[k,];
END;

/*----- - */
/*          ESTIMACAO DOS PARAMETROS          */
/*----- - */
/*----- - valores iniciais ----- */
INA0 = J((p+y-1),1,0);
w = NROW(INA0);                                     /* numero de parametros a ser estimado */

```

```

PARADA = J((p+y-1),1,1);
GAMA = INA0[1:(y-1),];          /* coeficiente em relacao aos interceptos */
BETA = INA0[y:w,];             /* coeficiente em relacao as covariaveis */
B = GAMA // BETA;              /* valores iniciais para vetor dos parametros */
IT = 0;
LAMBDA = 1;
/*----- - funcao para calcular as derivadas de MV----- */
START DERIVADA;
  P0 = J(n,(y-1),0);
  AUX1 = J(n,(y-1),0);
  UGAMA = J((y-1),1,0);
  UBETA = J(p,1,0);
  I0GAMA = J((y-1),1,0);
  I0BETA = J(p,p,0);
  I0BGAMA = J(p,(y-1),0);
  DO K = 1 TO (y-1);
    DO I= 1 TO N;
      IF DELTAI[I,K] = 2 THEN DO;
        P0[I,K] = 1/(1 + EXP(Z1[I,K]*GAMA[K,] + Z[I,]*BETA));
                                /* probabilidade de falha para cada intervalo */
        AUX1[I,K] = 1/(LOG(1 + EXP(Z1[I,K]*GAMA[K,] + Z[I,]*BETA)));
        UGAMA[K,] = UGAMA[K,] + P0[I,K]*(DELTAI[I,K]*AUX1[I,K] -
          TAU[K,]/TAU[K,]);          /* derivada em relacao gama k */
        UBETA = UBETA + Z[I,]* P0[I,K]*(DELTAI[I,K]*AUX1[I,K] -
          TAU[K,]/TAU[K,]);          /* derivada em relacao beta j */
        AUX2 = EXP(-(Z1[I,K]*GAMA[K,] + Z[I,]*BETA))*(DELTAI[I,K]*AUX1[I,K] -
          TAU[K,]/TAU[K,]) - DELTAI[I,K]*AUX1[I,K]**2;
        I0GAMA[K,] = I0GAMA[K,] + Z1[I,K]*P0[I,K]**2*AUX2*Z1[I,K];
                                /* derivada parcial de segunda ordem em relacao gama k */
        I0BETA = I0BETA + Z[I,]* P0[I,K]**2*AUX2*Z[I,];
                                /* derivada parcial de segunda ordem em relacao beta j */
        I0BGAMA[K,] = I0BGAMA[K,] + Z[I,]* P0[I,K]**2*AUX2*Z1[I,K];
                                /* derivada parcial de segunda ordem em relacao gama k e beta j */
      END;
    END;
  END;
END;

```

```

END;
FINISH;
/*----- - funcao para calcular funcao maxima verossimilhanca----- - */
START MV;
  LV = 0;
  DO I = 1 TO N;
    DO I= 1 TO N;
      IF DELTAI[I,K] = 2 THEN DO;
        LV = LV + DELTAI[I,K]*LOG(LOG(1 + EXP(ZI[I,K]*GAMA[K,] + Z[I,]*BETA)))
          - DELTAI[I,K]*LOG(TAU[K,]) - (TAUIK[I,K]/TAU[K,])*LOG(1 +
            EXP(ZI[I,K]*GAMA[K,] + Z[I,]*BETA));
      END;
    END;
  END;
FINISH;
/*----- - processo iterativo de Newton-Raphson ----- - */
DO WHILE ((MAX(ABS(PARADA)) > 0.0001)); /* criterio de parada */
  RUN DERIVADA; /* executa a funcao derivada */
  RUN MV; /* executa a funcao maxima verossimilhanca */
  LVA = LV;
  U0 = UGAMA // UBETA;
  I0GAM = DIAG(I0GAMA);
  I01 = I0GAM || (I0BGAMA)';
  I02 = I0BGAMA || I0BETA;
  I0 = I01 // I02;
  HINV = INV((-1)*I0); /* inversa da matriz de Fisher */
  DELTA0 = HINV*U0;
  B1 = B;
  PARADA = LAMBDA*DELTA0;
  B = B + PARADA; /* vetor dos parametros */
  GAMA = B[1:(y-1),];
  BETA = B[y:w,];
  RUN MV; /* executa a funcao maxima verossimilhanca */

  IF LV <= LVA THEN DO;

```



```

PRINT,"TESTE INTERCEPTO2 = 0";
C = {0 1 0 0 0 0};
RUN TESTE; /* executa a funcao teste */
PRINT,"TESTE LOGBUN = 0";
C = {0 0 1 0 0 0};
RUN TESTE; /* executa a funcao teste */
PRINT,"TESTE INFEC = 0";
C = {0 0 0 1 0 0};
RUN TESTE; /* executa a funcao teste */
PRINT,"TESTE PBJU = 0";
C = {0 0 0 0 1 0};
RUN TESTE; /* executa a funcao teste */
PRINT,"TESTE TOTOSORO = 0";
C = {0 0 0 0 0 1};
RUN TESTE; /* executa a funcao teste */
/* ----- */
/* CALCULO DA ESTIMATIVA DA FUNCAO DE SOBREVIVENCIA */
/* ----- */
S = J(N,1,1);
DO I = 1 TO N;
    DO K = 1 TO (y-1);
        S[I,] = S[I,]*(1 + EXP(Z1[I,K]*GAMA[k,] + ZB[I,]))##(-TAUIK[I,K]/TAU[K,]);
        /* funcao sobrevivencia estimada */
    END;
END;
/* ----- */
/* CALCULO DA ESTIMATIVA DA VARIANCIA */
/* METODO DELTA */
/* ----- */
SAUX = J(n,(y-1),0);
DO I=1 TO N;
    DO K = 1 TO (y-1);
        SAUX[I,K] = (1 + EXP(GAMA[K,] + ZB[I,]))##(-TAUIK[I,K]/TAU[K,]);
    END;
END;

```

```

GGAMA = J((y-1),N,0);
GBETA = J(p,n,0);
VS = J(N,1,0);
DO I = 1 TO N;
  GGAMA[1,I] = GGAMA[1,I] - (TAUIK[I,1]/TAU[1,1])*((1+EXP(GAMA[1,] +
    ZB[I,]))##(-TAUIK[I,1]/TAU[1,] - 1))*EXP(GAMA[1,] +
    ZB[I,])*SAUX[I,2];
    /* derivada S(t | z) em relacao a gama 1 */
  GGAMA[2,I] = GGAMA[2,I] - (TAUIK[I,2]/TAU[2,2])*((1+EXP(GAMA[2,] +
    ZB[I,]))##(-TAUIK[I,2]/TAU[2,] - 1))*EXP(GAMA[2,] +
    ZB[I,])*SAUX[I,1];
    /* derivada S(t | z) em relacao a gama 2 */
  DO J = 1 TO p;
    GBETA[J,I] = GBETA[J,I] - Z[I,J] *(TAUIK[I,1]/TAU[1,1])*((1+EXP(GAMA[1,] +
      ZB[I,]))##(-TAUIK[I,1]/TAU[1,] - 1))*EXP(GAMA[1,] +
      ZB[I,])*SAUX[I,2] - Z[I,J] *(TAUIK[I,2]/TAU[2,2])*((1+EXP(GAMA[2,] +
      ZB[I,]))##(-TAUIK[I,2]/TAU[2,] - 1))*EXP(GAMA[2,] +
      ZB[I,])*SAUX[I,1] ;
    /* derivada S(t | z) em relacao a beta */
  END;
END;
G = GGAMA // GBETA;
DO I= 1 TO N ;
  VS[I,] = VS[I,] + (G[I,])^2 * VB * G[I,];
  /* variancia de S(t | z) */
END;
DPS = VS##0.5;
/* erro padrao de S(t | z) */
/* ----- */
/* Intervalo de 95% de confianca para a funcao sobrevivencia */
/* no modelo logistico */
/* ----- */
AUX3 = 1.96;
/* valor tabelado da normal padrao com probabilidade 95% */
ICLI = J(N,1,0);
ICLS = J(N,1,0);
DO I = 1 TO N;
  ICLI[I,] = ICLI[I,] + S[I,] - AUX3*DPS[I,];
  /* limite inferior */

```

```

      ICLS[I,] = ICLS[I,] + S[I,] + AUX3*DPS[I,];          /* limite superior */
END;

IND = J(N,1,0);
DO I = 1 TO N;
  IND[I,] = I;
END;

/* ----- */
/*          IMPRESSAO DOS RESULTADOS          */
/* ----- */
VARIAVEL = { "INTERCEPTOS", "nomes das covariaveis incluidas no modelo" };
PRINT / "MODELO DE REGRESSAO LOGISTICO EXPONENCIAL POR PARTES";
PRINT, " ";
PRINT, " ";
PRINT, "PARAMETROS ESTIMADOS E ERRO PADRAO";
PRINT, VARIAVEL B [FORMAT=8.4] DPB [FORMAT=8.4];
PRINT, " ";
PRINT / "FUNCAO DE SOBREVIVENCIA ESTIMADA, ERRO PADRAO DE S(t | z) E";
PRINT "INTERVALO DE CONFIANCA DE 95% PARA S(t | z)";
PRINT IND TEMPO [FORMAT=5.2] S [FORMAT=8.4] DPS [FORMAT=8.4]
      ICLI [FORMAT=8.4] ICLS [FORMAT=8.4];
QUIT;

```

D3.10. Programa PAMEPTV.PRG (IML-SAS), desenvolvido para ajustar o modelo exponencial por partes. (tabelas de vida)

```

/* ----- PAMEPAS.PRG ----- versão 1.0 - Junho 93 ----- */
/* PROGRAMA PARA ESTIMAR : PARAMETROS */
/*          - LOG DA FUNCAO DE SOBREVIVENCIA */
/*          VARIANCA DO LOG FUNCAO DE SOBREVIVENCIA */
/* ----- */
/* ----- MODELO EXPONENCIAL POR PARTES ----- */
/*          (tabelas de vida) */
/* ----- */

```

```

/* ENTRADA DOS DADOS: D = vetor com numero de falhas          */
/*                    T = vetor com tempo total de seguimento  */
/*                    X = matriz de planejamento              */
/* -----                                                    */
PROC IML WORKSIZE = 100;
RESET NOPRINT;
/* ----- entrada dos dados -----                          */
D = {204,170,94,54,91,30,16,10,6,10,1,4,2,0,3,35,64,31,26,34,19,9,8,4,5,5,6,4,3,6};

T = {27390.5,43470,54000,94860,351630,3471,5340,7110,12600,48600,
     521,930,1080,1980,6570,13677,24450,32715,60210,242100,4665.5,
     8.325,11880,22590,104940,4839,9300,13590,26595,115380};

X = { 1  1  0  0  0  0  1  0  0  0,
      1  1  0  0  0  0  0  1  0  0,
      1  1  0  0  0  0  0  0  1  0,
      1  1  0  0  0  0  0  0  0  1,
      1  1  0  0  0  0 -1 -1 -1 -1,
      1  0  1  0  0  0  1  0  0  0,
      1  0  1  0  0  0  0  1  0  0,
      1  0  1  0  0  0  0  0  1  0,
      1  0  1  0  0  0  0  0  0  1,
      1  0  1  0  0  0 -1 -1 -1 -1,
      1  0  0  1  0  0  1  0  0  0,
      1  0  0  1  0  0  0  1  0  0,
      1  0  0  1  0  0  0  0  1  0,
      1  0  0  1  0  0  0  0  0  1,
      1  0  0  1  0  0 -1 -1 -1 -1,
      1  0  0  0  1  0  1  0  0  0,
      1  0  0  0  1  0  0  1  0  0,
      1  0  0  0  1  0  0  0  1  0,
      1  0  0  0  1  0  0  0  0  1,
      1  0  0  0  1  0 -1 -1 -1 -1,
      1  0  0  0  0  1  1  0  0  0,
      1  0  0  0  0  1  0  1  0  0,

```

```

1 0 0 0 0 1 0 0 1 0,
1 0 0 0 0 1 0 0 0 1,
1 0 0 0 0 1 -1 -1 -1 -1,
1 -1 -1 -1 -1 -1 1 0 0 0,
1 -1 -1 -1 -1 -1 0 1 0 0,
1 -1 -1 -1 -1 -1 0 0 1 0,
1 -1 -1 -1 -1 -1 0 0 0 1,
1 -1 -1 -1 -1 -1 -1 -1 -1 -1};

/*----- - matrizes auxiliares utilizadas nos calculos ----- */
N = NROW(X);          /* numero de linhas da matriz de planejamento */
P = NCOL(X);          /* numero de parametros a ser estimado */
UM = J(N,1,1);

/* ----- */
/*          ESTIMACAO DOS PARAMETROS          */
/* ----- */
/* ----- valores iniciais ----- */
INA0 = J(P,1,0);
PARADA = J(P,1,1);
B = INA0;
IT = 0;
LAMBDA = 1;
/* ----- funcao para calcular as derivadas de MV ----- */
START DERIVADA;
  UB = J(P,1,0);
  I0B = J(P,P,0);
  DO I= 1 TO N;
    UB = UB + X[I,]'*(D[I,] - T[I,]*EXP(X[I,]*B));  /* derivada em relacao a B */
    I0B = I0B + X[I,]'*T[I,]*EXP(X[I,]*B)*X[I,];  /* derivada segunda ordem em B */
  END;
FINISH;
/* ----- funcao para calcular funcao maxima verossimilhanca ----- */
START MV;
  LV = 0;
  DO I = 1 TO N;

```

```

      LV = LV + D[I,]*X[I,]*B - T[I,]*EXP(X[I,]*B) + D[I,]*LOG(T[I,]/DK[I,]);
    END;
  FINISH;
/* ----- processo iterativo de Newton-Raphson ----- */
DO WHILE ((MAX(ABS(PARADA)) > 0.0001));
  RUN DERIVADA;                                /* executa a funcao derivada */
  RUN MV;                                       /* executa a funcao maxima verossimilhanca */
  LVA = LV;
  HINV = INV(I0B);                             /* inversa da matriz de Fisher */
  DELTA0 = HINV*UB;
  B1 = B;
  PARADA = LAMBDA*DELTA0;
  B = B + PARADA;                               /* vetor dos parametros */
  RUN MV;                                       /* executa a funcao maxima verossimilhanca */
/* ----- VALORES PREDITOS ----- */
XB = X*B;                                       /* preditor linear */
VXB = X*HINV*X';                               /* matriz de covariancia de XB */
DPZB = VECDIAG(VXB)#0.5;                       /* erro padrao de XB */
/* ----- funcao para testar os parametros ----- */
START TESTE;
  CB = C*B;
  CVBC = C*VB*C';
  QW1 = CB' *INV(CVBC)*CB;                     /* estatistica de Wald */
  q = NROW(C);                                 /* numero de graus de liberdade */
  X = PROBCHI(QW1,q);
  PVALUE = 1 - X;
  QW = QW1 || PVALUE;
  PRINT QW;
FINISH;
/* ----- funcao para estatistica do ajuste ----- */
DE = T#EXP(XB);
AUX = 0;
DO I=1 TO N;
  IF D[I,] ^= 0 THEN DO;
    AUX1 = D[I,]/DE[I,];

```

```

      AUX = AUX + D[I,]*LOG(AUX1);
    END;
  END;
  QL = 2*AUX;
  GL = N-P;
  Y = PROBCHI(QL,GL);
  PVALUE = 1 - Y;
  PRINT QL GL PVALUE;

/* ----- */
/*          IMPRESSAO DOS RESULTADOS          */
/* ----- */

VARIABEL = { "efeitos a ser incluidos no modelo" };
PRINT / "MODELO EXPONENCIAL PARTICIONADO";
PRINT, " ";
PRINT, " ";
PRINT, "PARAMETROS ESTIMADOS E ERRO PADRAO";
PRINT, VARIABEL B [FORMAT=8.4] DPB [FORMAT=8.4];
PRINT, " ";
PRINT / " VALORES PREDITOS ";
PRINT, XB [FORMAT=8.4] DPXB [FORMAT=8.4];
QUIT;

```

D3.11. Programa PAMEP.PRG (IML-SAS), desenvolvido para ajustar o modelo exponencial por partes.

```

/* ----- PAMEP.PRG ----- versão 1.0 - Junho 93 ----- */
/* PROGRAMA PARA ESTIMAR: PARAMETROS */
/*          - LOG DA FUNCAO DE SOBREVIVENCIA          */
/*          VARIANCIA DO LOG FUNCAO DESOBREVIVENCIA */
/* ----- */
/* ----- MODELO EXPONENCIAL POR PARTES ----- */
/* ----- */
/* ENTRADA DOS DADOS: A = matriz com os dados */

```

```

/*                                T1 = pontos fixos dos tempos de observacao                                */
/* -----                                                                    ----- */

PROC IML WORKSIZE = 100;
RESET NOPRINT;

/* ----- entrada dos dados ----- */
A = { conjunto de dados};
T1 = {0,10,20};

/* ----- matrizes auxiliares utilizadas nos calculos ----- */
n = NROW(A);                                /* numero de observacoes (por exemplo pacientes) */
Z = a[, {3 15 16}];                          /* matriz das covariaveis */
p = NCOL(Z);                                /* numero de covariaveis no modelo */
STATUS = A[,2];                              /* vetor nx1 com indicador de censura */
TEMPO = A[,1];                                /* vetor nx1 com os Tempos */
y = NROW(T1);                                /* numero de pontos fixos no tempo */

/* ----- definindo algumas variaveis ----- */
Z1 = J(n,(y-1),0);                          /* interceptos para cada intervalo */
TIK = J(n,(y-1),0);                          /* tempos de falhas dentro de cada intervalo */
DELTAI = J(n,(y-1),0);                      /* variavel indicadora de censura em cada intervalo */

/* ----- definindo matriz de interceptos ----- */
DO K = 1 TO (y-1);
  DO I = 1 TO n;
    IF TEMPO[I,] >= T1[K,] THEN DO;
      Z1[I,K] = 1;
    END;
    ELSE DO;
      Z1[I,K] = 0;
    END;
  /* ----- definindo matriz de tempos ----- */
  IF (TEMPO[I,] >= T1[K,] & TEMPO[I,] < T1[(k+1),]) THEN DO;
    TIK[I,K] = TEMPO[I,] - T1[k,];
  END;
  IF (TEMPO[I,] >= T1[(k+1),]) THEN DO ;
    TIK[I,K] = T1[(k+1),] - T1[k,];
  END;

```

```

    IF (TEMPO[I,] < T1[K,]) THEN DO;
        TIK[I,K] = 0;
    END;

/*----- - definindo variavel resposta ----- */
    IF TEMPO[I,] >= T1[K,] THEN DO;
        IF TEMPO[I,] < T1[(k+1),] THEN DO;
            IF STATUS[I,] = 1 THEN DELTAI[I,K] = 1;
        END;
    END;
    IF TEMPO[I,] < T1[K,] THEN DO;
        DELTAI[I,K] = 2;
    END;
END;
END;

/*----- - valores iniciais ----- */
INA0 = J((p+y-1),1,0);
w = NROW(INA0); /* numero de parametros a ser estimado */
PARADA = J((p+y-1),1,1);
GAMA = INA0[1:(y-1),]; /* coeficiente em relacao aos interceptos */
BETA = INA0[y:w,]; /* coeficiente em relacao as covariáveis */
B = GAMA // BETA; /* valores iniciais para vetor dos parametros */
IT = 0;
LAMBDA = 1;

/*----- - funcao para calcular as derivadas de MV ----- */
START DERIVADA;
    UGAMA = J((y-1),1,0);
    UBETA = J(p,1,0);
    I0GAMA = J((y-1),1,0);
    I0BETA = J(p,p,0);
    I0BGAMA = J(p,(y-1),0);
    DO K = 1 TO (y-1);
        DO I= 1 TO N;

```

```

IF DELTAI[I,K] ^= 2 THEN DO;
  UGAMA[K,] = UGAMA[K,] + (DELTAI[I,K] - TIK[I,K]*EXP(Z1[I,K]*GAMA[K,]
    + Z[I,]*BETA));          /* derivada em relacao gama k */
  UBETA = UBETA + Z[I,]*(DELTAI[I,K] - TIK[I,K]*EXP(Z1[I,K]*GAMA[K,]
    TAU[K,]/TAU[I,]));      /* derivada em relacao beta j */
  I0GAMA[K,] = I0GAMA[K,] + Z1[I,K]*TIK[I,K]*EXP(Z1[I,K]*GAMA[K,] +
    Z[I,]*BETA);
    /* derivada parcial de segunda ordem em relacao gama k */
  I0BETA = I0BETA + Z[I,]*TIK[I,K]*EXP(Z1[I,K]*GAMA[K,] +
    Z[I,]*BETA)*Z[I,];
    /* derivada parcial de segunda ordem em relacao beta j */
  I0BGAMA[,K] = I0BGAMA[,K] + I0BGAMA[,K] + Z[I,]*(TIK[I,K]*EXP(Z1[I,K]
    *GAMA[K,] + Z[I,]*BETA)*Z1[I,K]);
    /* derivada parcial de segunda ordem em relacao gama k e beta j */
  END;
END;
END;
END;
FINISH;
/*----- - funcao para calcular funcao maxima verossimilhanca----- */
START MV;
  LV = 0;
  DO I = 1 TO N;
    DO I= 1 TO N;
      IF DELTAI[I,K] ^= 2 THEN DO;
        LV = LV + (DELTAI[I,K]*(Z1[I,K]*GAMA[K,] + Z[I,]*BETA) -
          TIK[I,K]*EXP(Z1[I,K]*GAMA[K,] + Z[I,]*BETA));
      END;
    END;
  END;
FINISH;
/*----- - processo iterativo de Newton-Raphson ----- */
DO WHILE ((MAX(ABS(PARADA)) > 0.0001));          /* criterio de parada */
  RUN DERIVADA;          /* executa a funcao derivada */
  RUN MV;                /* executa a funcao maxima verossimilhanca */
  LVA = LV;

```

```

U0 = UGAMA // UBETA;
I0GAM = DIAG(I0GAMA);
I01 = I0GAM || (I0BGAMA)' ;
I02 = I0BGAMA || I0BETA;
I0 = I01 // I02;
HINV = INV(I0);                               /* inversa da matriz de Fisher */
DELTA0 = HINV*U0;
B1 = B;
PARADA = LAMBDA*DELTA0;
B = B + PARADA;                               /* vetor dos parametros */
GAMA = B[1:(y-1),];
BETA = B[y:w,];
RUN MV;                                       /* executa a funcao maxima verossimilhanca */

IF LV <= LVA THEN DO;
  CONT = 0;
  DO WHILE ((LV <= LVA) & (CONT <= 10));
    LAMBDA = LAMBDA/2;
    PARADA = LAMBDA*DELTA0;
    B = B1 + PARADA;
    GAMA = B[1:(y-1),];
    BETA = B[y:w,];
    RUN MV;                                   /* executa a funcao maxima verossimilhanca */
    CONT = CONT + 1;
  END;
END;
IT = IT + 1;
END;
ITF = IT;                                     /* numero de iteracoes necessarias para estimar os parametros */
VB = HINV;                                    /* matriz de covariancia dos parametros */
DPB = VECDIAG(VB)##0.5;                      /* desvio padrao de B */
/*----- VALORES PREDITOS----- */
ZB = Z*B;                                    /* preditor linear */
/*----- funcao para testar os parametros----- */
START TESTE;

```

```

CB = C*B;
CVBC = C*VB*C' ;
QW1 = CB' *INV(CVBC)*CB;                                /* estatística de Wald */
q = NROW(C);                                            /* numero de graus de liberdade */
X = PROBCHI(QW1,q);
PVALUE = 1 - X;
QW = QW1 || PVALUE;
PRINT QW;
FINISH;
/*----- TESTAR OS PARAMETROS----- */
/*          entrar com a matriz C          */
/*----- */
PRINT,"TESTE INTERCEPTO1 = 0";
C = {1 0 0 0 0 0};
RUN TESTE;                                             /* executa a funcao teste */
/*----- */
/*  CALCULO DA ESTIMATIVA DA FUNCAO DE SOBREVIVENCIA  */
/*----- */
S = J(N,1,1);
LNS = J(N,1,0);
DO I = 1 TO N;
  DO K = 1 TO (y-1);
    S[I,] = S[I,]*EXP(-TIK[I,K]*EXP(Z1[I,K]*GAMA[K,] + Z[I,]*BETA));
                                                    /* funcao sobrevivencia estimada */
  END;
END;
LNS = (-1)*LOG(S);                                     /* funcao -log da funcao sobrevivencia estimada*/
/*----- */
/*          CALCULO DA ESTIMATIVA DA VARIANCIA - LOG S          */
/*          METODO DELTA          */
/*----- */
GGAMA = J((y-1),N,0);
GBETA = J(p,n,0);
VLNS = J(N,1,0);
DO I = 1 TO N;

```

```

DO K = 1 TO (y-1);
  GGAMA[k,I] = GGAMA[K,I] + TIK[I,K]*EXP(Z1[I,]*GAMA[K,] + Z[I,]*BETA);
                                     /* derivada - log S(t | z) em relacao a gama i */
DO J = 1 TO p;
  GBETA[J,I] = GBETA[J,I] + TIK[I,K]*Z[I,J]*EXP(Z1[I,]*GAMA[K,] +
                                     Z[I,]*BETA);      /* derivada - log S(t | z) em relacao a beta */
END;
END;
END;
G = GGAMA // GBETA;
DO I= 1 TO N ;
  VLNS[I,] = VLNS[I,] + (G[I,]* * VB * G[I,]);      /* variancia de - log S(t | z) */
END;
DPLNS = VLNS##0.5;      /* erro padrao de - log S(t | z) */
/* ----- */
/* Intervalo de 95% de confianca para a funcao sobrevivencia */
/* no modelo logistico */
/* ----- */
AUX = 1.96;      /* valor tabelado da normal padrao com probabilidade 95% */
ICLI = J(N,1,0);
ICLS = J(N,1,0);
DO I = 1 TO N;
  ICLI[I,] = ICLI[I,] + LNS[I,] - AUX*DPLNS[I,];      /* limite inferior */
  ICLS[I,] = ICLS[I,] + LNS[I,] + AUX*DPLNS[I,];      /* limite superior */
END;

IND = J(N,1,0);
DO I = 1 TO N;
  IND[I,] = I;
END;
/* ----- */
/* IMPRESSAO DOS RESULTADOS */
/* ----- */
VARIABEL = { "INTERCEPTOS", "nomes das covariaveis includas no modelo" };
PRINT / "MODELO DE REGRESSAO LOGISTICO EXPONENCIAL POR PARTES";

```

```

PRINT, " ";
PRINT, " ";
PRINT, "PARAMETROS ESTIMADOS E ERRO PADRAO";
PRINT, VARIABEL B [FORMAT=8.4] DPB [FORMAT=8.4];
PRINT, " ";
PRINT /"- LOG FUNCAO SOBREVIVENCIA ESTIMADA , ERRO PADRAO - LOGS(t | z) E";
PRINT "INTERVALO DE CONFIANCA DE 95% PARA - LOG S(t | z)";
PRINT IND TEMPO [FORMAT=5.2] LNS [FORMAT=8.4] DPLNS [FORMAT=8.4]
      ICLI [FORMAT=8.4] ICLS [FORMAT=8.4];
QUIT;

```

D3.12. Programa PAMLOGHM.PRG (IML-SAS), desenvolvido para ajustar o modelo log-linear multinomial.(Holford [1980])

```

/* ----- PAMLOGHM.PRG ----- versão 1.0 - Junho 93 ----*/
/* PROGRAMA PARA ESTIMAR: PARAMETROS */
/* ----- */
/* ----- MODELO LOG LINEAR MULTINOMIAL -----*/
/* (Holford [1980]) */
/* ----- */
/* ENTRADA DOS DADOS: D = vetor com numero de falhas */
/* T = vetor com tempo total de seguimento */
/* X = matriz de planejamento */
/* ----- */
PROC IML WORKSIZE = 100;
RESET NOPRINT;
/* ----- entrada dos dados ----- */
D = {204,170,94,54,91,30,16,10,6,10,1,4,2,0,3,35,64,31,26,34,19,9,8,4,5,5,6,4,3,6};

T = {27390.5,43470,54000,94860,351630,3471,5340,7110,12600,48600,
      521,930,1080,1980,6570,13677,24450,32715,60210,242100,4665.5,
      8.325,11880,22590,104940,4839,9300,13590,26595,115380};

X = { 1 0 0 0 0 1 0 0 0,
      1 0 0 0 0 0 1 0 0,

```

```

1 0 0 0 0 0 0 1 0,
1 0 0 0 0 0 0 0 1,
1 0 0 0 0 -1 -1 -1 -1,
0 1 0 0 0 1 0 0 0,
0 1 0 0 0 0 1 0 0,
0 1 0 0 0 0 0 1 0,
0 1 0 0 0 0 0 0 1,
0 1 0 0 0 -1 -1 -1 -1,
0 0 1 0 0 1 0 0 0,
0 0 1 0 0 0 1 0 0,
0 0 1 0 0 0 0 1 0,
0 0 1 0 0 0 0 0 1,
0 0 1 0 0 -1 -1 -1 -1,
0 0 0 1 0 1 0 0 0,
0 0 0 1 0 0 1 0 0,
0 0 0 1 0 0 0 1 0,
0 0 0 1 0 0 0 0 1,
0 0 0 1 0 -1 -1 -1 -1,
0 0 0 0 1 1 0 0 0,
0 0 0 0 1 0 1 0 0,
0 0 0 0 1 0 0 1 0,
0 0 0 0 1 0 0 0 1,
0 0 0 0 1 -1 -1 -1 -1,
-1 -1 -1 -1 -1 1 0 0 0,
-1 -1 -1 -1 -1 0 1 0 0,
-1 -1 -1 -1 -1 0 0 1 0,
-1 -1 -1 -1 -1 0 0 0 1,
-1 -1 -1 -1 -1 -1 -1 -1 -1});

```

```

/*----- matrizes auxiliares utilizadas nos calculos ----- */
N = NROW(X);          /* numero de linhas da matriz de planejamento */
P = NCOL(X);          /* numero de parametros a ser estimado */
TD = SUM(D);          /* total de falhas */
TT = SUM (T);         /* total de tempo de seguimento */
/*----- */

```

```

/*          ESTIMACAO DOS PARAMETROS          */
/* ----- */
/* ----- valores iniciais ----- */
INA0 = J(P,1,0);
PARADA = J(P,1,1);
B = INA0;
IT = 0;
LAMBDA = 1;
/* ----- funcao para calcular as derivadas de MV ----- */
START DERIVADA;
  UB = J(P,1,0);
  I0B = J(P,P,0);
  AUX1 = J(P,1,0);
  AUX2 = J(P,1,0);
  AUX3 = 0;
  AUX4 = J(P,P,0);
  DO I= 1 TO N;
    AUX1 = AUX1 + X[I,]' * D[I,];
    AUX2 = AUX2 + X[I,]' * T[I,]*EXP(X[I,]*B);
    AUX3 = AUX3 + T[I,]*EXP(X[I,]*B);
    AUX4 = AUX4 + X[I,]' * T[I,]*EXP(X[I,]*B)*X[I,];
  END;
  UB = UB + AUX1 - TD*AUX2/AUX3;;          /* derivada em relacao a B */
  I0B = I0B + TD*((AUX3*AUX4 - AUX2*AUX2')/(AUX3*#2));
                                          /* derivada segunda ordem em B */
FINISH;
/* ----- funcao para calcular funcao maxima verossimilhanca ----- */
START MV;
  LV = 0;
  AUX5 = 0;
  DO L = 1 TO N;
    AUX5 = AUX5 + T[L,]*EXP(X[L,]*B);
  END;
  DO I = 1 TO N;
    LV = LV + D[I,]*LOG(((T[I,]*EXP(X[I,]*B))/AUX5));
  END;

```

```

                                                    /* funcao de verossimilhanca */
      END;
FINISH;
/* ----- processo iterativo de Newton-Raphson ----- */
DO WHILE ((MAX(ABS(PARADA)) > 0.0001));
  RUN DERIVADA;                                /* executa a funcao derivada */
  RUN MV;                                       /* executa a funcao maxima verossimilhanca */
  LVA = LV;
  HINV = INV(I0B);                             /* inversa da matriz de Fisher */
  DELTA0 = HINV*UB;
  B1 = B;
  PARADA = LAMBDA*DELTA0;
  B = B + PARADA;                               /* vetor dos parametros */
  RUN MV;                                       /* executa a funcao maxima verossimilhanca */

  IF (LV <= LVA) THEN DO;
    CONT = 0;
    DO WHILE ((LV <= LVA) & (CONT <= 10));
      LAMBDA = LAMBDA/2;
      PARADA = LAMBDA*DELTA0;
      B = B1 + PARADA;
      RUN MV;                                   /* executa a funcao maxima verossimilhanca */
      CONT = CONT + 1;
    END;
  END;
  END;
  IT = IT + 1;
END;
ITF = IT;                                       /* numero de iteracoes necessarias para estimar os parametros */
VB = HINV;                                     /* matriz de covariância dos parametros */
DPB = VECDIAG(VB)##0.5;                       /* erro padrao de B */
/* ----- calculo do intercepto ----- */
START ALPHA;
  AUX6 = 0;
  DO I = 1 TO N;
    AUX6 = AUX6 + T[I,]*EXP(X[I,]*B);

```

```
END;
ALPHA = LOG((TT/AUX6));          /* coeficiente do intercepto */
FINISH;
RUN ALPHA;                       /* executa a funcao alpha */
B1 = ALPHA // B;
/* ----- */
/*          IMPRESSAO DOS RESULTADOS          */
/* ----- */

VARIABEL = { "efeitos a ser incluidos no modelo"};
PRINT / "MODELO MULTINOMIAL - HOLFORD [1980]";
PRINT, " ";
PRINT, " ";
PRINT, "PARAMETROS ESTIMADOS E ERRO PADRAO";
PRINT, VARIABEL B1 [FORMAT=8.4]
PRINT, DPB [FORMAT=8.4];
QUIT;
```

REFERÊNCIAS BIBLIOGRÁFICAS.

- AGRESTI, A. (1990). *Categorical Data Analysis*. New York, John Wiley.
- BIRCH, M.W. (1963). Maximum Likelihood in Three-Way Contingency Tables. *Journal Royal Statistic Society, Série B*, 25: 220-223.
- BISHOP, Y.M.M.; FIENBERG, S.E. e HOLLAND, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MIT Press.
- COX, D.R. (1972). Regression Models and Life Tables. *Journal Royal Statistic Society, Série B*, 34: 187-220.
- CRAMÉR, H. (1946). *Mathematical Methods of Statistics* Princeton. New Jersey.
- ELANDT-JONHSON, R.C. (1980a). Time Dependent Logistic Models in Follow-up Clinical Trials: I Binary Data. Chapel Hill, N.C.. *Institute of Statistic Mimeo Series Nro 1309*: 1-19.
- ELANDT-JONHSON, R.C. (1983). Logistic Linear Models in Prospective Studies: Their Relation to Exponential Hazard Rate Models. *Contributions to Statistics: Essays in honour of N.L. Johnson*, North-Holl and P.Co:135-145.
- ELANDT-JONHSON, R.C. e JOHNSON, N.L. (1980). *Survival Models and Data Analysis*. New York, John Wiley.
- FEIGL, P. e ZELEN, M. (1965). Estimation of Exponential Survival Probabilities with Concomitant Information. *Biometrics*. 21: 826-838.
- FROME, E.L.; KUTNER, M.H. e BEAUCHAMP, J.J. (1973). Regression Analysis of Poisson Distributed Data. *Journal of the American Statistical Association*. 68: 935-940.

- GLASSER, M. (1967). Exponential Survival with Covariates. *Journal of the American Statistical Association*. **62**: 561-568.
- GREEN, M.S. e SYMONS, M.J. (1983). A Comparison of the Logistic Risk Function and the Prospective Epidemiologic Studies. *Journal Chron. Dis.* . **36**; **10**: 715-724.
- HOLFORD, T.R. (1976). Life Tables With Concomitant Information. *Biometrics*. **32**: 587-597.
- HOLFORD, T.R. (1980). The Analysis os Rates and of Survivorship Using Log-Linear Models. *Biometrics*. **36**: 299-305.
- HOSMER, D.W.,Jr. e LEMESHOW, S. (1984). *Applied Logistic Regression*. New York, John Wiley & Sons. Inc..
- JORGENSEN, D.W. (1961). Multiple Regression Analysis of a Poisson Process. *Journal of the American Statistical Association*. **56**: 235-245.
- KALBFLEISH, J.D. e PRENTICE, R.L. (1980). *The Statistical Analysis of Failure Time Data*. New York, Jonh Wiley.
- KOCH, G.G.; ATKINSON, S.S e STOKES, M.E. (1986). Regression Poisson. *Encyclopedia of Statistical Sciences*. **7**: 32-41. New York, John Wiley.
- KRALL, J.M.; UTHOFF, V.A. e HARLEY, J.B. (1975). A Step-up Procedure for Selecting Variables Associated with Survival. *Biometrics*, **31**: 49-57.
- LAIRD, N. e OLIVIER, D. (1981). Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. *Journal of the Americam Statistical Association*. **76**, **374**: 231- 240.
- LAWLESS, J.F. (1982). *Statistical Models and Methods for Lifetime Data*. New York, Jonh Wiley.
- LEE, E.T. (1981). *Survival Analysis*. New York, John Wiley.

- LEITE, J.G. e SINGER, J.M. (1990). **Métodos Assintóticos em Estatística: Fundamentos e Aplicações**. São Paulo: Associação Brasileira de Estatística - ABE. 9º SINAPE.
- LIMA, A.C.P. (1988). **Estimação em Modelos de Sobrevivência**. São Paulo: Dissertação de Mestrado, Universidade de São Paulo.
- MILLER, R.G. (1981). **Survival Analysis**. New York, John Wiley.
- MYERS, M.H.; HANKEY, B.F. e MANTEL, N. (1973). A Logistic Exponential Model for Use with Response Time Data Involving Regressor Variables. *Biometrics*. 29: 257-269.
- REIS, I.M. (1989). **Modelos Log-Lineares para Análise de Dados Categorizados**. São Paulo: Dissertação de Mestrado, Universidade Estadual de São Paulo.
- SAS/IML: **Users' Guide** . Release 6.03 Edition. Cary, N.C.: SAS Institute Inc 1988.
- WALKER, S.H. e DUNCAN, D.R. (1967). Estimation of the Probability of an Event as a Function of Several Independent Variables. *Biometrika*. 54: 167-179.
- VIEIRA, E.G (1988). **Investigação de Feitos Citotoxicos e Glastogenicos do Carrapaticida Bovino Amitraz em Sistemas de Testes in Vivo in Vitro**. Campinas: Dissertação de Mestrado, Biologia, Universidade Estadual de Campinas.