NIVEA DA SILVA MATUDA

# HETEROGENEIDADE NÃO-OBSERVADA NA ANÁLISE DA HISTÓRIA DE EVENTOS

Profa. Orientadora Dra. Aida C.G. Verdugo Lazo

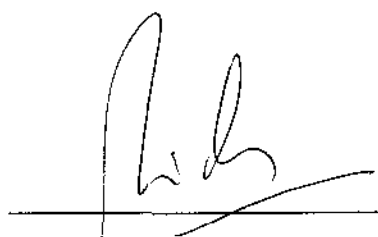
CAMPINAS
1998

UNICAMP
BIBLIOTECA CENTRAL

HETEROGENEIDADE NÃO-OBSERVADA NA ANÁLISE DA HISTÓRIA DE EVENTOS

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Nivea da Silva Matuda e aprovada pela comissão julgadora.

Campinas, 05 de julho de 1998



Profa. Dra.: Aida C.G. Verdugo Lazo
Orientadora

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

**FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP**

Matuda, Nivea da Silva

M327h Heterogeneidade não-observada na análise da história de eventos / Nivea da Silva Matuda -- Campinas, [S.P. :s.n.], 1998.

Orientador : Aida Verdugo Lazo

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Modelos log-lineares. 2. Variáveis latentes. 3. Análise da história de eventos. 4. Análise de sobrevivência. I. Lazo, Aida Verdugo. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

**Tese de Mestrado defendida e aprovada em 15 de junho de 1998**

**Pela Banca Examinadora composta pelos Profs. Drs.**

**Prof (a). Dr (a). AIDA CECÍLIA GRAZIELA VERDUGO LAZO**

**Prof (a). Dr (a). CICÍLIA YUKO WADA**

**Prof (a). Dr (a). NANCY LOPES GARCIA**

# Agradecimentos

Ao Senhor Jesus que me dá suporte para enfrentar todas as horas de crise e que participa comigo dos momentos de alegria pelas conquistas.

A professora Aida que foi uma orientadora-mãe por nunca deixar que as dificuldades que surgiram atrapalhassem o desenvolvimento deste trabalho, me incentivando sempre com sua confiança.

Aos meus familiares pelo apoio incondicional.

A todos os queridos colegas de turma pelo companheirismo e boas lembranças.

A professora Nancy e professor Cifointes que sempre me atenderam para esclarecer dúvidas cruciais.

A professora Chet que deu uma super força para que se pudesse rodar o programa Dnewton.

Ao professor Robert Mare, sua contribuição foi essencial para que se chegassem a resultados melhores.

Ao amigo Rui que não mediu esforços para me ajudar sempre que pedi socorro durante a montagem do arquivo de dados.

Ao NEPO por fornecer os dados do Censo.

A CAPES pelo suporte financeiro.

E aos caros colegas do Departamento de Estatística do UFPR por sempre estarem prontos a ajudar, possibilitando a finalização deste trabalho, em especial as professoras Suely e Silvia pela hospitalidade e apoio e o professor Caleffe por se dispor a corrigir meu Português em tempo recorde.

# Sumário

# Capítulo 1

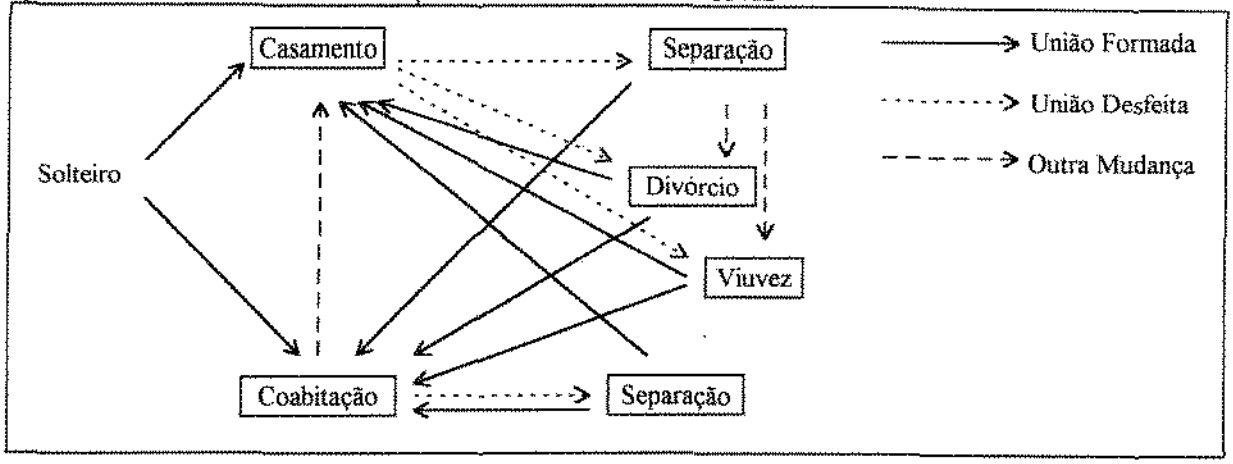
# Introdução

## 1.1 Apresentação e Motivação

Pesquisadores têm virtualmente grande interesse em eventos e suas causas. Um evento consiste de alguma mudança qualitativa que ocorre num ponto específico do tempo (Allison, 1984). Nascimento, casamento, falecimento, desemprego, reincidência de hábitos ou doenças e quebra de equipamentos são exemplos de mudanças que podem ser interpretados como eventos. Não é comum usar o termo evento para descrever uma mudança gradual de alguma variável. A mudança deve consistir de uma separação relativamente distinta entre o que precede e o que segue à transição.

Como os eventos são definidos em termos de mudanças no tempo, cada vez mais se reconhece que a melhor maneira de estudar eventos e suas causas é coletando dados da *história de eventos*. Considere, por exemplo, que um indivíduo solteiro poderia casar-se e, casado, poderia optar pela dissolução do casamento e assumir um segundo casamento ou uma primeira união consensual, e assim por diante. Um diagrama com os possíveis estágios e transições é apresentado no Quadro 1.1. Com os dados da história de eventos pode-se estudar todo o processo, admitindo interdependência entre os diferentes tipos de transições. Em sua forma mais simples, uma história de eventos é um registro longitudinal de quando acontece um tipo de transição para uma amostra de indivíduos. Se o objetivo

QUADRO 1.1 - TIPOS DE TRANSIÇÕES ENTRE ESTADOS CIVIS




for estudar as causas dos eventos, a história de eventos deveria também incluir dados sobre possíveis variáveis que, isoladamente ou em combinação, influenciam no tempo de ocorrência de um evento. Uma pesquisa sobre nupcialidade consistiria, por exemplo, da idade ao casar de solteiros e também incluiria informações sobre sexo e nível de instrução e renda na época do casamento.

Vários métodos para a análise da história de eventos foram surgindo independentemente em áreas como bioestatística, engenharia, ciências sociais e economia. Esses métodos algumas vezes competem entre si, mas na maioria das vezes complementam um ao outro. A demografia formulou o método mais antigo, mais conhecido e ainda muito utilizado para analisar dados da história de eventos – a tábua de vida. Os bioestatísticos desenvolveram mais tarde métodos para análise de dados de sobrevivência. De fato, muito da literatura sobre métodos da história de eventos aparece com o nome de análise de sobrevivência. Os engenheiros enfrentam uma situação similar na análise de dados de quebra de máquinas e componentes industriais. Os métodos desenvolvidos nesta área levam o nome de análise de confiabilidade.

Com o desenvolvimento de novas metodologias e o aumento de recursos computacionais, essas áreas tradicionais têm focalizado os *modelos de risco* que são métodos empregados para avaliar a dependência da ocorrência do evento sob um conjunto de variáveis regressoras (covariáveis), com o intuito de estimar as probabilidades de o evento ocorrer a

partir de um determinado tempo. Os modelos de risco na análise da história de eventos podem ser vistos como uma extensão da análise de regressão convencional (Allison, 1984; Trussel e Richards, 1985).

Um problema comum na análise de dados da história de eventos é a influência da heterogeneidade populacional sobre o tempo de ocorrência do evento. A idéia de que indivíduos mais frágeis [1] experimentarão o evento de interesse antes que os outros induz uma seleção de indivíduos mais fortes. Deste modo, em estudos da dinâmica de populações heterogêneas, o que se observa na história de eventos não representaria um indivíduo da população, pois seria influenciada por um processo seletivo (Vaupel e Yashin, 1985). Quase todas as análises com modelos de risco estão supostamente baseadas em que toda heterogeneidade da população de interesse é captada por meio de um conjunto de covariáveis observadas e incluídas no modelo. Levando em conta que nem toda a heterogeneidade pode ser considerada pelas covariáveis, o processo de seleção altera a composição das subpopulações formadas por determinantes não-observados, podendo distorcer a realidade e produzir estimativas viciadas dos parâmetros de interesse, caso a *heterogeneidade não-observada* for omitida (Heckman e Singer, 1982; Vaupel e Yashin 1985; Trussel e Rodriguez, 1990). Por essa razão, pesquisadores de diversas áreas têm incorporado a heterogeneidade não-observada nos modelos de risco.

Vaupel, Manton e Stallard (1979) foram os primeiros a discutir o impacto da heterogeneidade não-observada apresentando a idéia de *fragilidade*. Eles discutem este conceito no âmbito da terminologia usada na construção de tábuas de vida. Artigos posteriores apresentam modelos de fragilidade em outros contextos e áreas. Por exemplo, em análise de sobrevivência situam-se Hoougard (1984 e 1986), Aalen (1988), Oakes (1989) e Vaupel (1990). Em demografia, encontram-se entre outros, Trussel e Richards (1985), Trussel e Rodrigues (1990), Manton, Singer e Woodbury (1992) e Mare (1994). Na área de ciências sociais pode-se citar Allison (1984), Namboodiri e Suchindran (1987), Blossfeld, Hamerle

[1] Indivíduos são as unidades de análise. Por exemplo, componentes eletrônicos poderiam ser as unidades de análise em estudos de confiabilidade. É também abstrato o conceito de indivíduos mais frágeis ou mais fortes. Os indivíduos mais frágeis são os mais suscetíveis a sofrer o evento.

e Mayer (1989) e Blossfeld e Hamerle (1992) e em econometria Heckman e Singer (1982). Devido ao desenvolvimento e aplicação dos modelos de fragilidade em diversas áreas, a terminologia não é uniforme e nem todos os métodos são acessíveis a todos os usuários.

A intenção básica desta dissertação é apresentar detalhadamente os conceitos envolvidos no estudo da dinâmica de populações heterogêneas e discutir algumas tentativas que têm sido feitas para introduzir as fontes de heterogeneidade não-observada na análise da história de eventos. Inúmeros estudos da história de eventos têm buscado uma maneira eficiente de modelar a heterogeneidade não-observada com sugestões alternativas. Não se tem a intensão de expô-los integralmente. As discussões concentram-se em apresentar alguns modelos de risco com heterogeneidade não-observada e sua implicação na interpretação dos riscos populacionais. Os métodos de estimação e a sensibilidade dos modelos apresentados são, em geral, comentados com a inclusão de referências.

No presente capítulo, após a descrição dos fundamentos estatísticos da análise da história de eventos que são relevantes para descrever o caso univariado, é ressaltado, de maneira ilustrativa, o problema da heterogeneidade não-observada. No Capítulo 2 são apresentadas e discutidas estratégias paramétricas para modelar a heterogeneidade não-observada em modelos de risco univariado e algum comentário é feito sobre a estimação não-paramétrica. O Capítulo 2 termina apresentando o problema de identificabilidade com modelos de risco univariado incluindo heterogeneidade não-observada. O Capítulo 3 introduz a teoria estatística para analisar dados da história de eventos quando o tempo de ocorrência é contínuo e os indivíduos são observados aos pares. No Capítulo 3 também é apresentado o conceito de fragilidade compartilhada que viabiliza a identificabilidade dos modelos de risco bivariado para tempos contínuos com heterogeneidade não-observada. No Capítulo 4 riscos para tempos discretos são modelados por meio de uma metodologia sugerida por Mare (1994), que no caso de dados pareados também permite incluir heterogeneidade não-observada nos modelos de risco sem a perda da identificabilidade e sem a necessidade de se fazer fortes suposições paramétricas. Uma aplicação com dados de escolaridade de irmãos é apresentada no final do Capítulo 4, ilustrando a metodologia de

Mare. A análise restringe-se à quantificação das estatísticas de qualidade de ajuste e a uma descrição das estimativas dos parâmetros. O Capítulo 5 contém uma discussão geral sobre o impacto da heterogeneidade não-observada em modelos de risco e as diferentes soluções apontadas para controlá-lo.

## 1.2 Teoria Estatística na Análise da História de Eventos

### 1.2.1 Conceitos Básicos

Em muitas histórias de eventos, cada indivíduo pode experimentar eventos múltiplos e de vários tipos diferentes. Nesta seção são introduzidos modelos estatísticos em que cada indivíduo experimenta não mais do que um evento de um único tipo. Mesmo sendo simples, estes modelos podem ser aplicados em um grande número de situações como, por exemplo, em estudos de sobrevivência para análise da duração do primeiro emprego ou do tempo para imigrar.

Inicialmente, suponha que a população sob estudo seja homogênea, isto é, não há heterogeneidade entre os indivíduos. A introdução de covariáveis será considerada na Seção 1.2.2.

O tempo de espera até a ocorrência de um evento é representado no modelo estatístico por uma variável aleatória não-negativa $T$.

#### Funções do Tempo de Ocorrência do Evento

Um conceito chave da análise da história de eventos é a função taxa de risco ou simplesmente *função de risco* que é a probabilidade de um evento ocorrer num determinado tempo para um indivíduo em particular, dado que o indivíduo está exposto ao risco de sofrer o evento naquele tempo. Quando se consideram tempos contínuos, a probabilidade de um evento ocorrer exatamente no tempo $t$ é igual a zero, qualquer que seja $t$.

Portanto, a definição dada acima não é válida para tempos contínuos.

Seja $T$ uma variável aleatória contínua e não-negativa que representa o tempo de ocorrência do evento para um indivíduo. A função de distribuição e a densidade de $T$ são denotados respectivamente por $F(t)$ e $f(t)$. Como usual,

$$F(t) = P(T \leq t) \tag{1.1}$$

e para todos os pontos para os quais $F(t)$ pode ser diferenciada

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} = \frac{d}{dt} F(t) \ . \tag{1.2}$$

A função complementar de $F(t)$

$$S(t) = 1 - F(t) = P(T > t) \ , \tag{1.3}$$

denominada função de sobrevivência, representa a probabilidade de um indivíduo sobreviver a $t$, ou seja, até o tempo $t$ o evento ainda não ocorreu para o indivíduo. A função de sobrevivência é uma função monótona não-crescente e contínua, em que

$$S(0) = 1$$

e

$$\lim_{t \to +\infty} S(t) = 0$$

A função de risco para tempos contínuos é definida por

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} \ , \tag{1.4}$$

Note-se que

$$\lambda(t) \geq 0, \qquad \text{se } t \geq 0$$

e

$$\int_0^{\infty} \lambda(t)\, dt = \infty \ .$$

Não se deve pensar que a função de risco em (1.4) seja uma probabilidade, pois pode ser maior que 1. De fato não tem limite superior. A interpretação correta é dizer que $\lambda(t)$ é uma *taxa de ocorrência do evento*. Outras terminologias como taxa de intensidade e força de mortalidade ou de transição são freqüentemente encontradas em diferentes tipos de aplicações da função de risco.

A função de risco acumulada é representada pela integral

$$\Lambda(t) = \int_0^t \lambda(u)\, du \ .$$

Pela definição de probabilidade condicional, imediatamente obtém-se, de (1.4), a relação

$$\lambda(t) = \frac{f(t)}{S(t)} \ . \tag{1.5}$$

Pode-se encontrar a relação entre a função de sobrevivência e a função de risco pela integração de $\lambda(t)$. De (1.2), (1.3) e (1.5) tem-se que

$$\int_0^t \lambda(u)\, du = \int_0^t \frac{f(u)}{1 - F(u)} du = -\log\left[1 - F(t)\right] = -\log S(t)$$

$$\Leftrightarrow S(t) = \exp\left[-\int_0^t \lambda(u)\, du\right] = \exp\left\{-\Lambda(t)\right\} \tag{1.6}$$

e portanto

$$\lambda(t) = -\frac{d}{dt} \log S(t) \ . \tag{1.7}$$

A densidade $f(t)$ pode ser obtida de (1.5) e (1.6) como uma função de $\lambda(t)$, dada por

$$f(t) = \lambda(t)\, S(t) = \lambda(t) \exp\left[-\int_0^t \lambda(u)\, du\right] = \lambda(t) \exp\left[-\Lambda(t)\right] \ . \tag{1.8}$$

Considerando as relações de (1.1) a (1.3) e (1.5) a (1.8) torna-se evidente que cada uma

das três funções $f(t)$, $S(t)$ e $\lambda(t)$ pode ser usada para descrever o tempo de ocorrência de um evento. Em particular, se a função de risco for conhecida, a lei de probabilidade para o tempo de ocorrência do evento é caracterizada completamente.

Geralmente, tem-se pelo menos alguma informação qualitativa inicial de como $\lambda(t)$ pode depender do tempo. No começo do processo de vida, por exemplo, o risco de morrer é relativamente alto como atestam as elevadas taxas de mortalidade infantil; depois, o risco decresce e permanece quase constante por algum tempo em um nível baixo até crescer com o tempo por causa do envelhecimento do organismo. É possível obter alguma informação quantitativa sobre $\lambda(t)$ calculando a função de risco observada $\widehat{\lambda(t)}$, que possue pelo menos duas fórmulas alternativas. Se $\widehat{\lambda(t)}$ for calculado em intervalos de tempo de tamanho $\Delta t$ então

$$\widehat{\lambda(t)} = \frac{\text{número de indivíduos que sofreram o evento em } t}{(\text{número de sobreviventes a } t) \cdot \Delta t}$$

ou

$$\widehat{\lambda(t)} = \frac{\text{número indivíduos que sofreram o evento no intervalo } (t, t+\Delta t), \text{ por unidade de tempo}}{(\text{número de sobreviventes a } t) - \frac{1}{2}(\text{número de indivíduos que sofreram o evento no intervalo})} \quad ;$$

esta última é utilizada em ciências atuariais (Lee, 1980).

Essa informação qualitativa ou quantitativa da relação entre a função de risco e o tempo deve indicar as distribuições de probabilidades mais adequadas para $T$. Algumas distribuições especiais para o tempo de ocorrência do evento são apresentadas a seguir.

## Distribuições de Probabilidades para o Tempo de Ocorrência do Evento

### Distribuição Exponencial

Uma das aplicações mais simples de distribuições para $T$, o tempo de ocorrência do evento, é a distribuição exponencial com parâmetro $\alpha$ ($\alpha > 0$), conhecido como parâmetro de taxa. Denota-se por $T \sim \textit{exponencial}(\alpha)$. Para $t \geq 0$,

$$f(t) = \alpha \exp(-\alpha t)$$

$$S(t) = \exp(-\alpha t)$$

$$\lambda(t) = \alpha$$

Uma função de risco que não depende do tempo implica numa distribuição exponencial para o tempo de ocorrência do evento.

**Distribuição Gompertz**

Pode-se relaxar a suposição de um risco constante admitindo que o logaritmo do risco é uma função linear do tempo, como quando $T$ assume distribuição Gompertz, isto é, $T \sim Gompertz(\mu, \sigma)$, em que $\mu \in \mathbf{R}$ é o parâmetro de locação e $\sigma > 0$ é o parâmetro de escala. Para $t \in \mathbf{R}$,

$$f(t) = \sigma^{-1} \exp[(t-\mu)/\sigma] \exp\{-\exp[(t-\mu/\sigma)]\}$$

$$S(t) = \exp\{-\exp[(t-\mu)/\sigma]\}$$

$$\lambda(t) = \sigma^{-1} \exp[(t-\mu)/\sigma]$$

A distribuição de Gompertz também recebe os nomes de valor extremo e Gumbel. Note-se que a distribuição Gompertz é usada como distribuição de $T$, embora admita valores negativos com densidade positiva. É mais comum, no entanto, uma Gompertz surgir como distribuição de $\log T$. Isto equivale a assumir então que $T$ tenha uma distribuição Weibull.

**Distribuição Weibull**

Seja $T \sim Weibull(\phi, \varphi)$, em que $\phi, \varphi > 0$ são os parâmetros de escala e de forma, respectivamente. Então para $t \geq 0$,

$$f(t) = \varphi \phi^{-\varphi} t^{\varphi-1} \exp[-(t/\phi)^{\varphi}]$$

$$S(t) = \exp[-(t/\phi)^{\varphi}]$$

$$\lambda(t) = \varphi \phi^{-\varphi} t^{\varphi-1}$$

Neste caso o logaritmo da função de risco cresce ou decresce linearmente com o logaritmo do tempo. A relação entre os parâmetros das distribuições Gompertz e Weibull é a seguinte

$$\mu = \log \phi \qquad \text{e} \qquad \sigma = \frac{1}{\varphi}$$

Quando $\varphi = 1$ na distribuição Weibull, obtém-se uma distribuição exponencial com $\alpha = 1/\phi$.

Crowder et al. (1990), apresentam os gráficos destas funções para alguns valores específicos dos parâmetros e propõem outros exemplos de distribuições para o tempo de ocorrência de um evento.

### 1.2.2 Modelos de Risco

Em adição ao tempo de ocorrência, geralmente várias covariáveis são coletadas de cada indivíduo da amostra e um objetivo importante da análise estatística é verificar a influência quantitativa dessas covariáveis sobre a função de risco.

A história de eventos é ideal para o estudo de causas de eventos, mas possui duas características — censuras [2] e covariáveis variando no tempo — que criam problemas para procedimentos estatísticos padrões como a regressão múltipla. De fato, a aplicação de métodos padrões pode levar a vícios sérios ou perda de informações (Allison, 1984). Esses problemas têm sido resolvidos pelo *método de máxima verossimilhança*, apresentado mais adiante. O método de máxima verossimilhança requer especificação da forma funcional da distribuição do tempo de ocorrência do evento. Os métodos não-paramétricos não requerem tal suposição mas a presença de observações censuradas implica que métodos não-paramétricos clássicos baseados em postos não são diretamente aplicáveis. O método

[2] No término do período de observação, alguns indivíduos podem ainda não ter experimentado o evento. O tempo exato de ocorrência do evento para estes indivíduos é desconhecido e só é possível argumentar que é maior que o tempo fixado para o período de observação. Este tipo de informação, conhecido como censura aleatória, pode ocorrer também quando indivíduos são retirados do experimento ou "perdidos" durante o período de observação. Aqui são consideradas somente censuras aleatórias. Para mais detalhes ver Lawless (1982).

de Kaplan-Meier (ver Lee, 1980) para estimar a função de sobrevivência é um exemplo de método não-paramétrico para dados da história de eventos.

### Modelos para Tempos Contínuos

Suponha que se tenha uma amostra de $n$ indivíduos independentes e que cada indivíduo começa a ser observado a partir de algum ponto natural igual a zero. Por exemplo, se o evento de interesse for o divórcio, o ponto inicial é a data de casamento. Na análise da história de eventos, para cada indivíduo $i$ $(i = 1, 2, \ldots, n)$ os dados consistem de $(t_i, \delta_i, \mathbf{x}_i)$ em que $t_i$ é o tempo de ocorrência do evento ou o tempo de censura para o indivíduo $i$; $\delta_i$ é igual a 0 se $t_i$ for censurado ou é igual a 1, caso contrário; e $\mathbf{x}_i' = (x_{1i}, x_{2i}, \ldots, x_{pi})$ é um vetor dos valores de $p$ covariáveis do indivíduo $i$ que são consideradas para predizer o evento. O vetor $\mathbf{x}_i$, neste caso, não varia no tempo, mas uma generalização para covariáveis variando no tempo pode ser considerada. Assume-se que $0 \leq t_i \leq N$, sendo que $N$ é o tempo final do período de observação.

Uma formulação geral conhecida como *modelo de riscos proporcionais* é dada por

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\,\Psi(\mathbf{x}) \ ,$$

em que $\lambda_0(t)$ é uma função do tempo $t$ e $\Psi(\mathbf{x})$ é uma função positiva de $\mathbf{x}$ ([3]). Este modelo é conhecido como modelo de riscos proporcionais porque para quaisquer dois indivíduos num tempo $t$, a razão de seus riscos independe de $t$. Assume-se neste modelo que as covariáveis têm um efeito multiplicativo sobre a função de risco, suposição que parece ser razoável em muitas situações, por isso tem grande importância na análise de dados da história de eventos.

Pensar que a função de risco seja função linear das covariáveis pode ser mais simples, mas não é adequado, porque a função de risco não pode ser negativa e uma função linear pode tomar valores menores que zero. Normalmente considera-se o logaritmo da função

[3] É comum suprimir o índice $i$ e esta prática é seguida daqui por diante, sempre que possível.

de risco como função linear das covariáveis.

Uma formulação funcional muito usada é o modelo de riscos proporcionais linear dado por

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}) \ , \tag{1.9}$$

em que $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ é um vetor de parâmetros desconhecido. O vetor $\boldsymbol{\beta}$ representa os efeitos das covariáveis sobre a taxa de ocorrência do evento. Assim, se $\beta_1$ for positivo um crescimento em $x_1$ produz um crescimento na verossimilhança de que o evento ocorrerá. Assume-se que esses efeitos são constantes no tempo.

A análise estatística depende de se assumir ou não uma forma específica para $\lambda_0(t)$ em (1.9). Casos particulares deste modelo são obtidos especificando-se a função $\lambda_0(t)$. A suposição mais simples é que $\lambda_0(t) = \beta_0$; isto implica que $T$ tenha distribuição exponencial. Se for assumido que $\lambda_0(t) = \beta_0 + \beta_{01} \log t$, então $T$ tem distribuição Weibull. Alternativamente, a especificação $\lambda_0(t) = \beta_0 + \beta_{01} t$ determina uma distribuição Gompertz para $T$ (ver Seção 1.2.1). Em geral, uma especificação da função $\lambda_0(t)$ é equivalente à especificação da distribuição dos tempos contínuos de ocorrência do evento. Entretanto, note que, sob o modelo de riscos proporcionais, os parâmetros dessas distribuições são dados em função de $\mathbf{x}$.

Um problema para a análise dos dados é que $T$ não é observado para casos censurados. No entanto, o método de máxima verossimilhança (**MV**) admite fazer o uso completo da informação que se tem nos casos de censura para estimar os parâmetros do modelo de riscos proporcionais. O método **MV** é certamente uma das técnicas mais utilizadas na estimação paramétrica, quando a forma da distribuição geradora dos dados é conhecida. Em geral, produz equações bastante convenientes do ponto de vista computacional e estimadores com boas qualidades estatísticas. O método combina as observações censuradas e não-censuradas de tal modo que produz (sob certas condições) estimativas assintoticamente não-viciadas, normalmente distribuídas e eficientes.

A forma da *função de verossimilhança* depende do tipo de censura. Ao assumir censura aleatória, considera-se para um indivíduo que $T$ é o tempo de ocorrência do

evento e $L$ é o tempo de censura. Para cada indivíduo $i$ $(i = 1, 2, \ldots, n)$ observa-se que

$$t_i = \min(T_i, L_i)$$

e

$$\delta_i = \begin{cases} 1, \text{ se } T_i \leq L_i \\ 0, \text{ se } T_i > L_i \end{cases} .$$

Considerando-se os pares $(T_i, L_i)$ independentes e ainda $f_T(t_i|\mathbf{x}_i)$ e $S_T(t_i|\mathbf{x}_i)$ a função de densidade e de sobrevivência de $T$, respectivamente e $f_L(t_i|\mathbf{x}_i)$ e $S_L(t_i|\mathbf{x}_i)$ a função de densidade e de sobrevivência de $L$, respectivamente, tem-se que a probabilidade do indivíduo $i$ ser censurado em $t$ é dada por

$$\begin{aligned} P(t_i = t, \delta_i = 0|\mathbf{x}_i) &= P(L_i = t, T_i > L_i|\mathbf{x}_i) = P(L_i = t, T_i > t|\mathbf{x}_i) \\ &= P(L_i = t|\mathbf{x}_i) P(T_i > t|\mathbf{x}_i) = f_L(t|\mathbf{x}_i) S_T(t|\mathbf{x}_i) \end{aligned}$$

e a probabilidade do indivíduo $i$ sofrer o evento no tempo $t$ e não estar censurado é dada por

$$\begin{aligned} P(t_i = t, \delta_i = 1|\mathbf{x}_i) &= P(T_i = t, T_i \leq L_i|\mathbf{x}_i) = P(T_i = t, L_i \geq t|\mathbf{x}_i) \\ &= P(T_i = t|\mathbf{x}_i) P(L_i \geq t|\mathbf{x}_i) = f_T(t|\mathbf{x}_i) S_L(t|\mathbf{x}_i) . \end{aligned}$$

A *função de verossimilhança* para $(t_i, \delta_i, \mathbf{x}_i)$ $(i = 1, 2, \ldots, n)$ é definida como

$$\begin{aligned} L &= \prod_{i=1}^{n} [f_T(t_i|\mathbf{x}_i) S_L(t_i|\mathbf{x}_i)]^{\delta_i} [f_L(t_i|\mathbf{x}_i) S_T(t_i|\mathbf{x}_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} f_T(t_i|\mathbf{x}_i)^{\delta_i} S_T(t_i|\mathbf{x}_i)^{1-\delta_i} f_L(t_i|\mathbf{x}_i)^{1-\delta_i} S_L(t_i|\mathbf{x}_i)^{\delta_i} . \end{aligned} \tag{1.10}$$

Em geral, os parâmetros de interesse são aqueles associados ao tempo de ocorência do evento e as funções $f_L(t_i|\mathbf{x}_i)$ e $S_L(t_i|\mathbf{x}_i)$ não envolvem qualquer parâmetro de interesse.

Assim, a função de verossimilhança em (1.10) pode ser simplificada para

$$L = \prod_{i=1}^{n} \left[ f_T \left( t_i | \mathbf{x}_i \right) \right]^{\delta_i} \left[ S_T \left( t_i | \mathbf{x}_i \right) \right]^{1-\delta_i} \quad . \tag{1.11}$$

Assume-se em (1.11) que a censura independe da ocorrência do evento, isto é, indivíduos não são seletivamentes censurados porque eles têm mais ou menos chance de experimentar o evento (Allison,1980).

A aplicação de (1.5) em (1.11), resulta numa expressão para verossimilhança em termos da função de sobrevivência e de risco dada por

$$L = \prod_{i=1}^{n} \left[ \lambda \left( t_i | \mathbf{x}_i \right) \right]^{\delta_i} S \left( t_i | \mathbf{x}_i \right) \tag{1.12}$$

e, de (1.6) e (1.12), é possível escrever a função de verossimilhança em termos unicamente da função de risco

$$L = \prod_{i=1}^{n} \left[ \lambda \left( t_i | \mathbf{x}_i \right) \right]^{\delta_i} \exp \left[ - \int_0^{t_i} \lambda \left( u | \mathbf{x}_i \right) du \right] \quad . \tag{1.13}$$

Estimativas de **MV** são obtidas substituindo em (1.13) a expressão apropriada para $\lambda \left( t_i | \mathbf{x}_i \right)$ e então escolhendo estimativas de $\lambda_0 \left( t \right)$ e de $\boldsymbol{\beta}$ que maximizam $L$. Em geral, os estimadores de **MV** não podem ser expressos analiticamente; necessita-se de métodos iterativos para a obtenção das estimativas como, por exemplo, o conhecido método de Newton-Rapshon.

Uma desvantagem do método **MV** é a exigência de que se especifique a forma de $\lambda_0 \left( t \right)$ para poder ser usado. Geralmente, não há bases teóricas e empíricas suficientes para escolher entre especificações alternativas; além disso, os resultados podem variar dependendo da escolha.

Quando não há uma indicação sobre a forma específica de $\lambda_0 \left( t \right)$, o modelo de riscos proporcionais linear é chamado de semi-paramétrico e foi sugerido por Cox (1972). Em parte é paramétrico porque especifica um modelo de regressão com uma forma funcional exclusiva; é em parte não-paramétrico porque não especifica a forma exata da

distribuição dos tempos de ocorrência do evento. Para estimar os parâmetros do modelo semi-paramétrico é usado o método chamado de verossimilhança parcial (**VP**), que tem-se mostrado altamente eficiente e tem sido muito usado em experimentos médicos. Uma descrição detalhada dos métodos de **MV** e **VP** pode ser encontrada em Kalbfleisch e Prentice (1980).

O modelo de riscos proporcionais pode ser generalizado para admitir covariáveis que mudam no tempo. Os parâmetros podem ser estimados sem dificuldades pelo método de **VP**. As covariáveis, dependendo do tempo, também podem ser incorporadas à estimação por **MV**, mas esta estratégia, segundo Allison (1982), freqüentemente leva a procedimentos computacionais exaustivos. Somente o método de **MV** é proposto para estimar os parâmetros dos modelos discutidos nos capítulos seguintes.

Outra maneira de analisar a influência quantitativa das covariáveis sobre a função de risco é por meio do *modelo de tempo de falha acelerado*. É um modelo de regressão introduzido convenientemente como o logaritmo de $T_i$, o tempo de ocorrência do evento no indivíduo $i$, já que $T_i$ é não-negativo. O modelo especifica uma relação linear dada por

$$\log T_i = \boldsymbol{\gamma}'\mathbf{x}_i + \varepsilon U_i$$

em que $\varepsilon$ é um parâmetro de escala e $U_i$ é uma variável aleatória que representa o erro da observação $i$. Várias escolhas da distribuição do erro levam a versões diferentes de modelos de risco.

O modelo de tempo de falha acelerado pode alternativamente ser formulado via um modelo de risco, isto é

$$\lambda\left(t|\mathbf{x}\right) = \lambda_V\left[t\exp\left(-\boldsymbol{\gamma}'\mathbf{x}\right)\right]\exp\left(-\boldsymbol{\gamma}'\mathbf{x}\right)$$

em que $\lambda_V$ denota a função de risco de $V = \exp\left(\varepsilon U\right)$.

### Modelos para Tempos Discretos

Modelos de risco para tempos discretos são comumente usados em pesquisas de diversas áreas, como aproximações de modelos para tempos contínuos ou para apresentação de processos que são intrisecamente discretos (Allison, 1982).

Os modelos para tempos discretos têm várias características desejáveis como, por exemplo, maior facilidade para incorporar covariáveis dependentes do tempo e interpretação mais imediata, sem metodologia sofisticada.

A notação usada nos modelos para tempos discretos é similar àquela para tempos contínuos. Assume-se que os tempos são registrados somente sobre valores inteiros-positivos e que são observados $n$ indivíduos independentes começando de algum ponto inicial igual a um. Os dados para cada indivíduo são formados por $(t_i, \delta_i, \mathbf{x}_{ti})$ em que $t_i$ e $\delta_i$ são definidos como no caso dos modelos para tempos contínuos (por exemplo, se $t_i = a$ e $\delta_i = 0$ significa que o indivíduo $i$ é observado em $a$ mas não em $a+1$, em que $a$ é um inteiro-positivo) e o vetor de covariáveis do indivíduo $i$ no tempo $t$ é representado por $\mathbf{x}'_{ti} = (x_{1ti}, x_{2ti}, \ldots, x_{pti})$, que pode assumir valores diferentes em cada tempo discreto. Assume-se que $t_i = 1, 2, \ldots, N$ em que $N$ é o tempo final do período de observação e, como habitual, assume-se que o tempo de censura é independente do risco de ocorrência do evento (Allison, 1982).

A função de risco para tempos discretos, denotada por $h(t)$, é

$$h(t) = P(T = t | T \geq t) \ ,$$

em que $T$ é a variável aleatória discreta representando o tempo de ocorrência do evento.

O próximo passo é especificar como a função de risco depende das covariáveis $\mathbf{x}_t$ e do tempo $t$. Isto é,

$$h(t|\mathbf{x}_t) = P(T = t | T \geq t, \mathbf{x}_t) \ .$$

Uma primeira aproximação poderia ser a definição da função de risco como função linear das covariáveis. Contudo, por ser uma probabilidade, $h(t|\mathbf{x}_t)$ varia entre 0 e 1,

enquanto uma função linear pode assumir qualquer valor real. Este problema pode ser contornado aplicando-se a *transformação logito* em $h(t|\mathbf{x}_t)$:

$$\log\left[\frac{h(t|\mathbf{x}_t)}{1-h(t|\mathbf{x}_t)}\right] \tag{1.14}$$

e fazendo

$$\log\left[\frac{h(t|\mathbf{x}_t)}{1-h(t|\mathbf{x}_t)}\right] = \beta_{0t} + \boldsymbol{\beta}'\mathbf{x}_t \ ,$$

em que $\beta_{0t}$ e $\boldsymbol{\beta}$ são os parâmetros do modelo.

Como $h(t|\mathbf{x}_t)$ varia entre 0 e 1, a transformação logito varia entre menos e mais infinito. Há outras transformações que têm esta propriedade, mas a transformação logito é a mais familiar e a mais conveniente computacionalmente, apesar de ser uma escolha arbitrária. Note-se que $\beta_{0t}$ refere-se a $N$ diferentes constantes (uma constante para cada tempo $t = 1, 2, \ldots, N$ observado) e, como antes, $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ é um vetor de parâmetros desconhecido dos efeitos das covariáveis. Isto implica que o coeficiente $\beta_1$, por exemplo, representa a mudança que ocorrerá no logito se o valor em $x_1$ alterar-se.

A transformação logito em modelos de risco para tempos discretos pode ser escrita na forma de regressão logística. Isolando $h(t|\mathbf{x}_t)$ em (1.14) tem-se que

$$h(t|\mathbf{x}_t) = \frac{\exp(\beta_{0t} + \boldsymbol{\beta}'\mathbf{x}_t)}{1 + \exp(\beta_{0t} + \boldsymbol{\beta}'\mathbf{x}_t)} \ .$$

Alguns casos especiais do modelo de regressão logística são obtidos impondo-se restrições ao conjunto de constantes $\beta_{0t}$. Por exemplo:

$$\beta_{0t} = \beta_0 \ ,$$

$$\beta_{0t} = \beta_0 + \beta_{01} t \ ,$$

ou

$$\beta_{0t} = \beta_0 + \beta_{01} \log t \ .$$

É também possível generalizar o modelo assumindo que os efeitos das covariáveis variam no tempo, simplesmente substituindo $\boldsymbol{\beta}$ por $\boldsymbol{\beta}_t$.

Para estimar os parâmetros da regressão logística, pode ser usado um estimador **VP** análogo àquele para tempos contínuos. Mas este método torna-se extremamente exaustivo computacionalmente se o evento ocorrer para muitos indivíduos durante a mesma unidade de tempo (Allison, 1982). Felizmente, o método de **MV** pode ser usado sem qualquer restrição sobre $\beta_{0t}$.

A função de verossimilhança para tempos discretos e identicamente distribuídos pode ser escrita como

$$L = \prod_{i=1}^{n} P\left(T = t_i|\mathbf{x}_{ti}\right)^{\delta_i} P\left(T > t_i|\mathbf{x}_{ti}\right)^{1-\delta_i} \ , \tag{1.15}$$

que é análoga à função de verossimilhança em (1.11) para tempos contínuos. As probabilidades em (1.15) podem ser expressas como uma função de $h\left(t_i|\mathbf{x}_{ti}\right)$.

Pela propriedade de evento complementar tem-se que

$$h\left(t\right) = 1 - P\left(T > t|T \geq t\right) \ . \tag{1.16}$$

Pela definição de probabilidade condicional, (1.16) resulta em

$$h\left(t\right) = 1 - \frac{P\left(T > t\right)}{P\left(T \geq t\right)} \tag{1.17}$$

e a definição de função de risco resulta em

$$h\left(t\right) = \frac{P\left(T = t\right)}{P\left(T \geq t\right)} \ . \tag{1.18}$$

Como $t = 1, 2, \ldots, N$, tem-se que

$$P\left(T \geq t\right) = P\left(T > t - 1\right) \ . \tag{1.19}$$

Mostra-se então por (1.17) e (1.19) que

$$P(T > t) = [1 - h(t)]\, P(T \geq t) = \prod_{j=1}^{t} [1 - h(j)] \tag{1.20}$$

e por (1.18) e (1.20) que

$$P(T = t) = h(t)\, P(T \geq t) = h(t) \prod_{j=1}^{t-1} [1 - h(j)] \ . \tag{1.21}$$

A função de verossimilhança em (1.15) pode ser reescrita em termos da função de risco. Usando-se os resultados de (1.20) e (1.21) com a função de risco dependendo de $t$ e $\mathbf{x}$ e, tomando o logaritmo, tem-se a função de log-verossimilhança é dada por

$$\log L = \sum_{i=1}^{n} \left[ \delta_i \log \frac{h(t_i|\mathbf{x}_{ti})}{1 - h(t_i|\mathbf{x}_{ti})} + \log \sum_{j=1}^{t_i} [1 - h(j|\mathbf{x}_{ti})] \right] \ .$$

Ao substituir-se $h(t_i|\mathbf{x}_{ti})$ por um adequado modelo de regressão logística pode-se, neste ponto, proceder à maximização de $\log L$ com relação a $\beta_{0t}$ e $\boldsymbol{\beta}$, para obter os estimadores de **MV**.

A transformação logito também é usada quando são analisados os efeitos de covariáveis categóricas sobre uma variável dependente dicotômica. Neste caso, os dados são geralmente apresentados sob a forma de uma tabela de contingência. A transformação logito é aplicada sobre as probabilidades esperadas para a variável dependente e o modelo que estima os efeitos das covariáveis é denominado *modelo logito*. O Capítulo 4 apresenta dados categóricos da história de eventos que são analisados por meio de algumas generalizações do modelo logito. Uma descrição do modelo logito pode ser vista em Agresti (1990).

### Eventos Repetidos e de Múltiplos Tipos

Muitos dos conceitos envolvidos para o caso de um evento podem ser aplicados a situações mais complexas como *eventos repetidos*. Os eventos repetidos são processos

em que cada indivíduo pode experimentar uma sucessão de eventos. Exemplos incluem tempo de concepção dos filhos e períodos sucessivos de desemprego. Desde que os eventos ocorrem a um mesmo indivíduo, os tempos de espera, em geral, não são independentes e somente o último intervalo pode estar censurado, pois os eventos ocorrem em seqüência. Isto introduz alguma simplificação na estimação pois é possível estudar uma sucessão de eventos usando *condicionalidade seqüencial*. Considere, por exemplo, $\mathbf{T} = (T_1, T_2, T_3)$ uma variável aleatória trivariada e discreta que representa os tempos de ocorrência de 3 eventos sucessivos para um indivíduo. Pode-se sempre fatorar a probabilidade conjunta, $P(T_1 = t_1, T_2 = t_2, T_3 = t_3)$, como o produto de

$$P(T_1 = t_1) \cdot P(T_2 = t_2 | T_1 = t_1) \cdot P(T_3 = t_3 | T_1 = t_1, T_2 = t_2) \ .$$

Uma contribuição típica para a função de verossimilhança, dado que $T_3$ é o único intervalo que pode estar censurado, é

$$P(T_1 = t_1) \cdot P(T_2 = t_2 | T_1 = t_1) \cdot$$
$$P(T_3 = t_3 | T_1 = t_1, T_2 = t_2)^{\delta} \cdot P(T_3 \geq t_3 | T_1 = t_1, T_2 = t_2)^{1-\delta} \ ,$$

em que $\delta = 0$ para as observações censuradas de $T_3$ e $\delta = 1$ para observações de $T_3$ completamente observadas, como usual.

Um outro tipo de dado multivariado da história de eventos origina-se quando um indivíduo experimenta eventos de múltiplos estados, passando por vários tipos de transições. A natureza dos dados permite condicionar cada transição com a história integral das transições prévias, combinando os elementos de riscos competitivos com os modelos de eventos repetidos.

Os métodos **MV** e **VP** podem ser estendidos para manejar dados completos da história de eventos que podem envolver tipos múltiplos de eventos repetidos. Uma descrição sobre esses modelos mais complexos pode ser encontrada, por exemplo, em Crowder et al. (1990) e em Allison (1984) para modelos para tempos discretos.

Dados multivariados da hitória de eventos que se originam de indivíduos pareados ou de outros agrupamentos de maior tamanho são considerados a partir do Capítulo 3.

### 1.2.3 Misturas de Distribuições

Os modelos de risco apresentados até aqui não têm um termo de perturbação aleatória, como o erro aleatório no modelo de regressão padrão. No entanto, eles não são modelos determinísticos porque existe uma variação aleatória entre a variável dependente $\lambda(t|\mathbf{x})$ e o tempo observado ou, no caso discreto, entre $h(t|\mathbf{x}_t)$ e o tamanho do intervalo de tempo observado (Allison, 1982 e 1984). Ainda assim, alguns argumentam que deveria ser incluído um termo de perturbação aleatória nos modelos de risco, isto é, um termo que representaria o efeito da heterogeneidade não-observada, a fim de diminuir a variação aleatória. A presença de duas fontes aleatórias nos modelos de risco faz com que a distribuição de $T$ seja uma *mistura de distribuições.* Pode-se classificar o estudo da dinâmica de populações heterogêneas como uma interseção da análise da história de eventos com mistura de distribuições.

Misturas ocorrem freqüentemente quando o parâmetro $\Theta$ de uma família de distribuição estiver sujeito a uma variação. Suponha que esta família de distribuição tenha função de distribuição $F(t|\Theta)$ e que $\Theta$ tenha função de distribuição $G(\theta)$, para $t$ e $\theta$ específicos. Então a esperança de $F(t|\Theta)$, dada por

$$F(t) = E[F(t|\Theta)] = \int_{-\infty}^{\infty} F(t|\theta)\, dG(\theta) \ , \tag{1.22}$$

é também uma função de distribuição.

$F(t)$ é a função de distribuição da mistura ou a função de distribuição não-condicionada, $G(\theta)$ é chamada de função de distribuição misturadora [4] e $F(t|\theta)$ é uma função

[4] Em inglês, mixing distribution function.

de distribuição condicionada. Uma forma simbólica para representar a mistura é

$$D_1 \bigwedge_{\Theta} D_2$$

em que $D_1$ é a distribuição condicionada a $\Theta$ e $D_2$ é a distribuição de $\Theta$.

Dada a função de sobrevivência condicionada, $S(t|\Theta)$, obtém-se a função de sobrevivência não-condicionada pela linearidade da esperança

$$S(t) = E[S(t|\Theta)] = \int_{-\infty}^{\infty} S(t|\theta)\, dG(\theta) \ . \tag{1.23}$$

**Exemplo 1.1** A mistura

$$Poisson(\Theta) \bigwedge_{\Theta} gama(\eta, \nu)$$

representa uma distribuição de Poisson composta, formada quando o valor esperado $\Theta$ de uma Poisson tem distribuição gama. Esta distribuição de Poisson composta é de fato uma distribuição binomial negativa de parâmetros $\eta$ e $\nu$ (Johnson e Kotz, 1969).

**Exemplo 1.2** Suponha que a variável aleatória $T$ tenha distribuição

$$exponencial(\Theta) \bigwedge_{\Theta} gama(\eta, \nu) \ .$$

A distribuição de $T$ é uma mistura de exponenciais, formada quando o inverso do valor esperado da distribuição exponencial tem distribuição gama. De fato a distribuição de $T$ é conhecida como distribuição de Pareto do tipo II com parâmetros $\eta$ e $\nu$ (Johnson e Kotz, 1970).

A função de sobrevivência não-condicionada de $T$ é dada por

$$S(t) = \int_0^{\infty} S(t|\theta)\, g(\theta)\, d\theta \ ,$$

em que

$$S(t|\theta) = \exp(-\theta t)$$

e

$$g(\theta) = \frac{\nu^{\eta}}{\Gamma(\eta)}\theta^{\eta-1}\exp(-\nu\theta) \ , \qquad \text{para } \theta \geq 0 \ .$$

Assim,

$$\begin{aligned} S(t) &= \int_0^{\infty} \exp(-\theta t)\frac{\nu^{\eta}}{\Gamma(\eta)}\theta^{\eta-1}\exp(-\nu\theta)\,d\theta \\ &= \frac{\nu^{\eta}}{\Gamma(\eta)}\int_0^{\infty} \exp[-\theta(\nu+t)]\,\theta^{\eta-1}d\theta \ . \end{aligned}$$

Completando-se a integral para uma densidade $gama(\nu+t,\eta)$, resulta que

$$S(t) = \left(\frac{\nu}{\nu+t}\right)^{\eta}$$

e da relação em (1.7) obtém-se a função de risco não-condicionada de $T$ dada por

$$\lambda(t) = \frac{\eta}{\nu+t} \ ;$$

sendo que $\lambda(t)$ é decrescente em $t$.

Note-se que o parâmetro $\Theta$, que está sujeito a uma variação, não aparece na distribuição misturada. Mas é provável que a mistura dependa do(s) parâmetro(s) da distribuição de $\Theta$, como nos Exemplos 1.1 e 1.2. Num caso mais geral, quando a família de distribuição é multiparamétrica, a mistura pode também depender dos parâmetros que não variam na distribuição condicional, como no exemplo seguinte.

**Exemplo 1.3** Se $T$ tem uma distribuição dada por

$$Weibull(\Theta,\varphi) \bigwedge_{(\Theta)^{-\varphi}} gama(\eta,\nu)$$

então $T$ tem uma distribuição Weibull misturada, conhecida como distribuição de Burr,

a qual depende dos parâmetros $\eta, \nu, \varphi$ porém não depende de $\Theta$. Quando $\eta = 1$, a distribuição de $T$ é algumas vezes chamada de distribuição log-logística (Johnson e Kotz, 1970).

O conceito de condicionalidade é a base da definição de distribuição misturada. Seja $F(t|\Theta = \theta)$ a função de distribuição de $T$ condicionada a $\Theta = \theta$ e seja $F(t)$ a função de distribuição não-condicionada de $T$. A função de distribuição não-condicionada $F(t)$ é obtida como o valor esperado de $F(t|\Theta)$, desde que

$$F(t) = P(T \leq t) = E\left(\mathbf{I}_{\{T \leq t\}}\right) \qquad \text{e} \qquad F(t|\Theta) = P(T \leq t|\Theta) = E\left(\mathbf{I}_{\{T \leq t\}}|\Theta\right) ,$$

em que $\mathbf{I}_{\{A\}}$ é a função indicadora de $A$. Pela propriedade básica da esperança condicional tem-se que

$$F(t) = E\left(\mathbf{I}_{\{T \leq t\}}\right) = E\left[E\left(\mathbf{I}_{\{T \leq t\}}|\Theta\right)\right] = E\left[F(t|\Theta)\right] .$$

Uma situação típica em que a mistura de distribuições é aplicável é quando se investiga uma variável aleatória $T$ e a distribuição condicional de $T$, dado o valor observado da variável aleatória $\Theta$, é conhecida exatamente ou tem forma relativamente simples. Por causa da inabilidade para observar $\Theta$, a distribuição não-condicionada de $T$ é uma mistura e é usualmente muito complexa.

O maior interesse no estudo de misturas de distribuições é a estimação de seus parâmetros. No entanto, antes que o processo de estimação possa ser empreendido, é necessário considerar a *identificabilidade da mistura*, isto é, a caracterização única da mistura. Uma mistura é identificável se existe uma correspondência 1–1 entre a distribuição misturadora e a mistura resultante. Uma mistura que não é identificável não pode ser expressa unicamente como função da distribuição condicional ou da distribuição misturadora. A identificabilidade é crucial uma vez que não é compatível estimar ou testar hipóteses acerca dos parâmetros de misturas que não sejam identificáveis (Kotz e Johnson, 1985).

## 1.3 O Impacto no Modelo de Risco ao Negligenciar a Heterogeneidade Não-Observada

É necessária muita cautela ao se tentar fazer inferências acerca do efeito do tempo sobre a função de risco. Numa população homogênea, todos os indivíduos estão sujeitos à mesma função de risco. Uma população heterogênea consiste de várias subpopulações homogêneas (Vaupel e Yashin, 1985), com as correspondentes funções de risco subpopulacionais que podem ser combinadas para formar a função de risco para toda a população. Ignorar a heterogeneidade entre os indivíduos de uma população pode levar a recomendações errôneas se a intervenção depende de respostas em nível individual. Mesmo quando o risco individual é constante no tempo, diferenças entre os indivíduos que afetam a função de risco individual e que não estão incorporadas ao modelo tenderão a produzir evidências de uma função de risco populacional que decresce com o tempo. Uma explicação intuitiva é que indivíduos pertencentes às subpopulações com altos riscos experimentam o evento antes que os outros e portanto são eliminados previamente do grupo em risco. Com o passar do tempo, esta seleção produz uma população exposta ao risco, cada vez mais constituída de indivíduos com baixos riscos, resultando na observação de um risco decrescente com o tempo, que não representa os verdadeiros riscos individuais, mas um processo seletivo da população.

Se o risco de reincidência de pessoas que estão tentando parar de fumar, por exemplo, parece cair com o tempo, então, isto implica que um indivíduo que parou de fumar há pouco tempo tem maior tendência para voltar ao vício do que aquele que parou de fumar há mais tempo? Não necessariamente. Poderia haver dois grupos de indivíduos: os de alto risco de reincidência e os de baixo risco de reincidência. O risco de reincidência em cada grupo poderia ser constante no tempo, mas o risco de reincidência para toda a população, decrescendo com o tempo, seria um artifício da heterogeneidade.

Para explicar o impacto da heterogeneidade na análise do efeito do tempo sobre a função de risco, sejam $\lambda_1(t)$ e $\lambda_2(t)$ as funções de risco de duas subpopulações homogêneas

e seja $\lambda(t)$ a função de risco de toda a população no tempo $t$. Vaupel e Yashin (1985) definiram $\lambda(t)$ como uma média ponderada de $\lambda_1(t)$ e $\lambda_2(t)$. Sejam $S_1(t)$ e $S_2(t)$ as funções de sobrevivência das duas subpopulações, ou seja

$$S_i(t) = \exp\left[-\int_0^t \lambda_i(u)\,du\right], \qquad i = 1, 2\,. \tag{1.24}$$

E defina-se $\pi(t)$ como a proporção de indivíduos da primeira subpopulação que não sofreram o evento antes de $t$, isto é

$$\pi(t) = \frac{\pi(0)\,S_1(t)}{\pi(0)\,S_1(t) + [1-\pi(0)]\,S_2(t)}\,. \tag{1.25}$$

A função de risco populacional é dada por

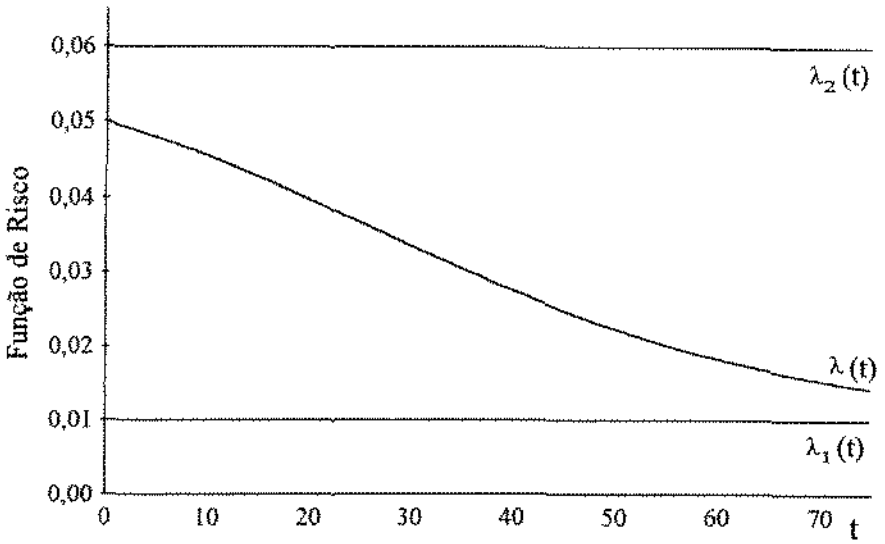
$$\lambda(t) = \pi(t)\,\lambda_1(t) + [1-\pi(t)]\,\lambda_2(t)\,. \tag{1.26}$$

Portanto, em (1.26) a relação entre a função de risco populacional e as funções de risco subpopulacionais é ponderada pela proporção (não-constante) da população exposta ao risco em cada subpopulação. Com o tempo, a função de risco de toda população se aproximará da função de risco da subpopulação mais forte, ou seja a subpopulação com menor taxa de risco. O Gráfico 1.1 ilustra um caso específico, em que $\lambda_1(t) = 0,01$; $\lambda_2(t) = 0,06$; $\pi(0) = 0,2$ e $0 \leq t \leq 75$.

Em geral, o risco estimado de modelos que negligenciam a heterogeneidade decresce mais rapidamente ou cresce mais lentamente do que os verdadeiros riscos das subpopulações homogêneas (Trussel e Richards, 1985). O Gráfico 1.2 contém um outro exemplo de dois riscos subpopulacionais que resultam em um risco populacional não-monotonicamente decrescente, em que o risco populacional $\lambda(t)$ foi obtido de (1.26) com $\lambda_1(t) = 2t$; $\lambda_2(t) = \frac{3}{2}\,t^{\frac{1}{2}}$; $\pi(0) = 0,5$ e $0 \leq t \leq 6,5$.

Não há como saber se a população é de fato homogênea com um risco $\lambda(t)$ ou se a população é heterogênea com dois riscos subpopulacionais $\lambda_1(t)$ e $\lambda_2(t)$, que resultam em um risco populacional $\lambda(t)$, porque a existência de subpopulações é uma suposição e

GRÁFICO 1.1 - RISCO POPULACIONAL DECRESCENTE NO TEMPO COM RISCOS SUBPOPULACIONAIS CONSTANTES

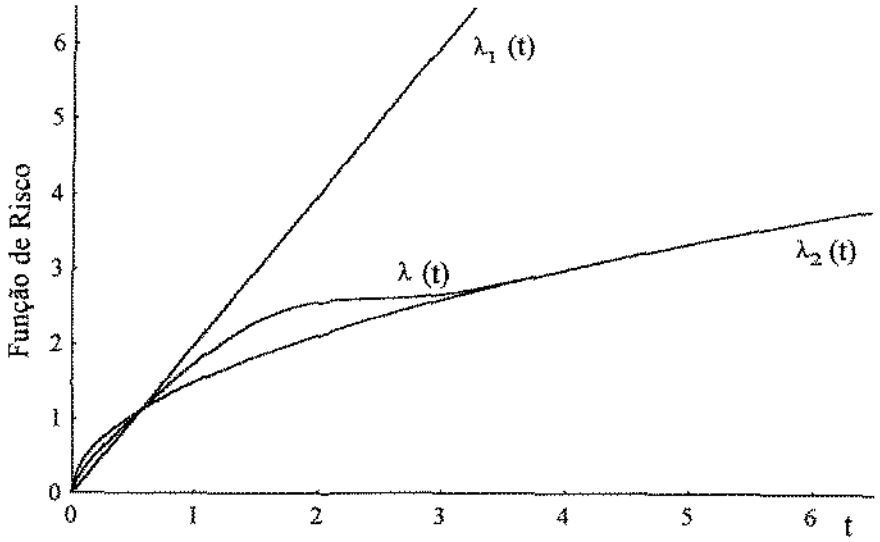



somente $\lambda(t)$ é observado.

Um modo de contornar o problema da falta de identificabilidade é incorporar, explicitamente no modelo, fontes de heterogeneidade por meio de covariáveis observadas. Mas a suposição de que todas as fontes de heterogeneidade possam ser medidas e consideradas não é realista; esta suposição é assumida em quase todas as aplicações de modelos de risco (Trussel e Richards, 1985). Por exemplo, em estudos sobre determinantes de mortalidade infantil, o analista não pode observar a fragilidade inerente a cada indivíduo (Vaupel, Manton e Stallard, 1979). Analogamente, em análise de fecundidade, freqüentemente assume-se que a capacidade de engravidar varia entre as mulheres mas não pode ser explicitamente medida (Trussel e Richards, 1985). Se a idiossincrasia não pode ser considerada num processo de estimação surge então, o problema de omissão de variáveis.

Segundo Trussel e Richards (1985), num simples modelo de riscos proporcionais com covariáveis categóricas, uma variável omitida resulta em estimativas distorcidas dos parâmetros relacionados com as covariáveis incluídas no modelo, inclusive quando a distribuição das características omitidas é inicialmente a mesma em todas as categorias das covariáveis incluídas no modelo. O problema é que as subpopulações, formadas pelas ca-

GRÁFICO 1.2 - RISCO POPULACIONAL CRESCENTE NO TEMPO COM RISCOS SUBPOPULACIONAIS CRESCENTES




tegorias das covariáveis não são homogêneas e, por isso, a distribuição das características omitidas pode, com o tempo, alterar-se dentro das categorias. Este resultado é verdadeiro porque indivíduos de alto risco com certos valores na variável omitida experimentarão o evento antes, tanto que a distribuição das características omitidas deverá mudar a uma taxa diferente em cada subpopulação. Considere como ilustração o exemplo a seguir descrito por Trussel e Rodrigues (1990).

**Exemplo 1.4** Seja uma amostra de indivíduos classificados por sexo e cor conforme a Tabela 1.1. A proporção de homens e de mulheres em cada cor é a mesma (igual a 50%), ou seja, a distribuição da variável sexo é, inicialmente, a mesma em todas as categorias da variável cor.

Suponha que fosse assumida a função de risco dada por

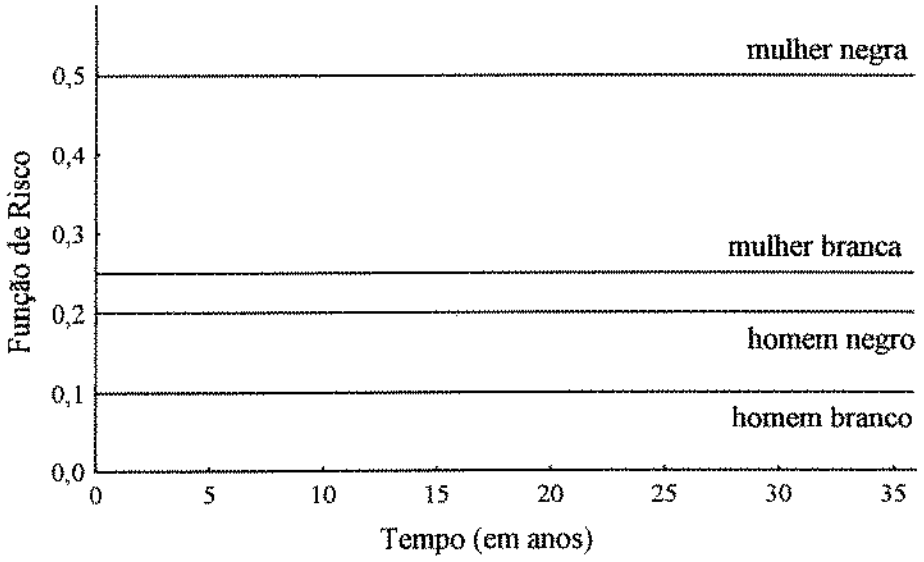
$$\lambda(t|x,z) = \exp(-2,30 + 0,69x + 0,91z) \ , \tag{1.27}$$

em que $t$ é o tempo de sobrevivência observado em anos, $x$ assume o valor 1 se o indivíduo for negro e 0 se for branco e $z$ assume o valor 1 se o indivíduo for do sexo masculino e 0 se for do sexo feminino. O modelo assume que os riscos são constantes com o tempo

TABELA 1.1 - DISTRIBUIÇÃO DA POPULAÇÃO INICIAL POR SEXO E COR EM FREQÜÊNCIAS E PORCENTAGENS

| SEXO | COR | | | |
|---|---|---|---|---|
| | FREQÜÊNCIA | | PORCENTAGEM | |
| | BRANCA | NEGRA | BRANCA | NEGRA |
| HOMEM | 50 | 100 | 50,0 | 50,0 |
| MULHER | 50 | 100 | 50,0 | 50,0 |
| TOTAL | 100 | 200 | 100,0 | 100,0 |

GRÁFICO 1.3 - RISCOS POR SEXO E COR




mas variam por sexo e cor (ver Gráfico 1.3).

Assumindo o modelo em (1.27), as proporções na Tabela 1.1 não se mantém as mesmas, nem variam igualmente.

Define-se $n_{xz}(t)$ como número de sobreviventes da cor $x$ e sexo $z$ no tempo $t$, isto é

$$n_{xz}(t) = n_{xz}(0)\, S(t|x,z) \ ,$$

em que $S(t|x,z)$ é a função de sobrevivência.

Como

$$S(t|x,z) = \exp\left[-\int_0^t \lambda(t|x,z)\right] ,$$

tem-se que

$$n_{xz}(t) = n_{xz}(0)\exp\left[-t\,\lambda(t|x,z)\right] ; \tag{1.28}$$

sendo que $n_{xz}(t)$ é decrescente em $t$.

O número de sobreviventes por cor e sexo para $t = 5$ é apresentado na Tabela 1.2. Por meio desta tabela pode-se observar que a proporção de mulheres após 5 anos é diferente em cada cor (68% na cor branca e 82% na cor negra). A composição da amostra não pode permanecer a mesma por causa dos diferentes níveis da taxa de risco em cada categoria da variável sexo, tanto em negros como em brancos.

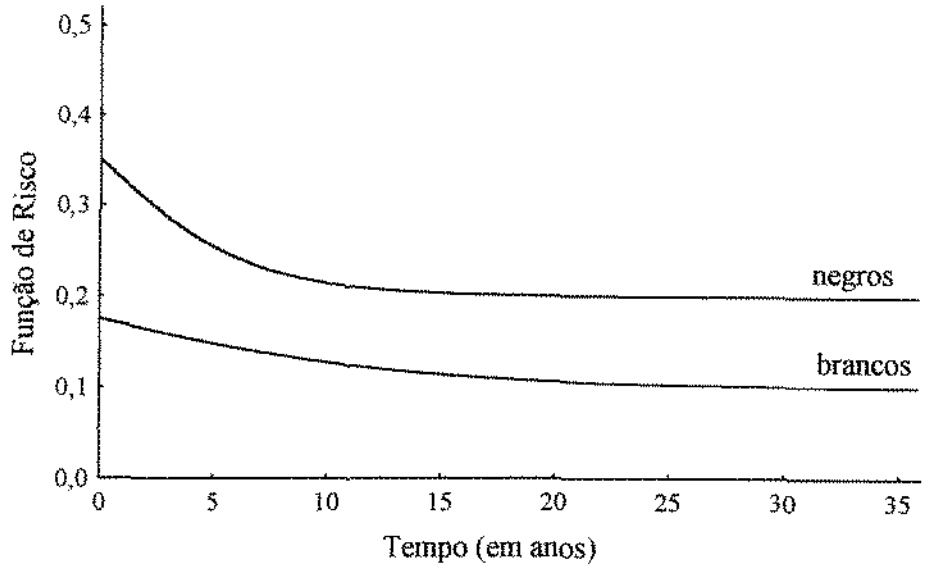
TABELA 1.2 - DISTRIBUIÇÃO DA POPULAÇÃO APÓS 5 ANOS POR SEXO E COR EM FREQÜÊNCIAS E PORCENTAGENS

| SEXO | COR | | | |
|---|---|---|---|---|
| | FREQÜÊNCIA | | PORCENTAGEM | |
| | BRANCA | NEGRA | BRANCA | NEGRA |
| HOMEM | 14 | 8 | 31,8 | 17,8 |
| MULHER | 30 | 37 | 68,2 | 82,2 |
| TOTAL | 44 | 45 | 100,0 | 100,0 |

Se for omitida a variável sexo, os riscos por cor (não-condicionados ao sexo) *declinarão* com o tempo porque os homens, que têm riscos mais altos que as mulheres independente da cor, tendem a ter menores chances de sobreviver (ver Gráfico 1.4). As curvas no Gráfico 1.4 foram traçadas com base em (1.24), (1.25) e (1.26) sendo $\pi(0) = 0,50$ para negros e brancos.

A omissão de $z$ leva a conclusões diferentes sobre os riscos de brancos e negro. Primeiro porque o Gráfico 1.4 evidencia riscos decrescentes com o tempo, para cada cor. E segundo porque observando o Gráfico 1.4 verifica-se que o risco da população negra não diminui proporcionalmente ao da branca. Isto ocorre porque a dinâmica do processo vai subtrair rapidamente a população negra masculina, que tem um risco muito alto (de

GRÁFICO 1.4 - RISCOS POR COR




0,50). Quando toda a população masculina de brancos e negros tiver sido virtualmente selecionada, a relação entre os riscos por cor volta a ser o dobro, pois neste momento o risco de brancos é aproximadamente igual ao risco de mulheres brancas e o risco de negros é aproximadamente igual ao risco de mulheres negras.

Vários pesquisadores têm reconhecido que é comum a presença de heterogeneidade não-observada e que sua omissão acarreta em diferentes conclusões com relação ao efeito do tempo e das covariáveis observadas sobre a função de risco, e incorporam-na explicitamente em seus modelos. O seguinte problema proposto por Vaupel e Yashin (1985) é uma ilustração muito simples desta área de pesquisa.

**Exemplo 1.5** Indivíduos pertencentes a uma população situam-se em duas subpopulações com funções de risco constantes $\lambda_1$ e $\lambda_2$. Como um resultado de estudos que serviram de subsídio, os valores de $\lambda_1$ e $\lambda_2$ são conhecidos, com $\lambda_2 > \lambda_1$. Observações sobre o tempo de ocorrência do evento são viáveis para cada indivíduo que experimentou o evento no período de observação. Para os indivíduos censurados, o tempo em que os indivíduos pararam de ser observados é conhecido. O que é desconhecido, e deve ser estimado, é a proporção $\pi(0)$ de indivíduos que pertencem à primeira subpopulação no tempo inicial.

Para motivar a discusão deste problema de inferência estatística são apresentadas duas situações:

a) uma população de animais usada para estudar a eficiência de uma vacina está sendo testada com relação a uma doença. Indivíduos vacinados com sucesso têm um risco $\lambda_1$ de adoecer, que pode ser zero. Os outros indivíduos vacinados podem permanecer desprotegidos com risco $\lambda_2$, se sua resposta de imunização for inadequada. O que precisa ser estimado é a proporção $\pi(0)$ de indivíduos em risco que foram imunizados. Através de dados da incidência da doença na população vacinada pode-se obter uma estimativa de $\pi(0)$;

b) uma peça de algum equipamento pode ser fabricada de dois modos — o modo comum e o modo simplificado — produzindo peças aparentemente iguais. Se a peça for feita do modo comum, o risco de falhar é alguma constante $\lambda_1$; caso contrário é $\lambda_2$. Várias peças tinham sido produzidas pelos dois modos de fabricação. O tempo de falha de algumas dessas peças é observado. A análise estatística consiste em estimar $\pi(0)$, a proporção de peças fabricadas pelo modo comum.

A proporção $\pi(0)$ pode ser estimada pelo método de máxima verossimilhança. A função de verossimilhança é $\lambda(t)S(t)$ se o evento ocorrer no tempo $t$, e a função de verossimilhança para os indivíduos censurados é simplesmente $S(t)$, em que $t$ é o tempo de observação até o término do experimento. A função de risco $\lambda(t)$ de toda a população pode ser calculada de (1.24), (1.25) e (1.26) e a função de sobrevivência populacional $S(t)$ pode ser obtida de (1.6).

Quando as observações individuais forem independentes, a função de verossimilhança para os dados é então dada por

$$L[\pi(0)] = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i) \ ,$$

em que $n$ é o tamanho da amostra, $t_i$ é o tempo de ocorrência do evento para o i-ésimo indivíduo ou o tempo de censura e $\delta_i$ é igual a 0, se o i-ésimo indivíduo for censurado e

é igual a 1, caso contrário. A estimativa de máxima verossimilhança para $\pi(0)$ pode ser encontrada por meio de algoritmos computacionais padrões.

Este simples exemplo poderia ser generalizado de diversas formas. Os valores de $\lambda_1$ e $\lambda_2$ poderiam ser desconhecidos e teriam que ser estimados; as funções de risco poderiam depender do tempo ou um vetor de covariáveis poderia ser observado para cada indivíduo, podendo ser assumido um modelo de riscos proporcionais adequado para o conjunto de dados. Alguns métodos que tratam da heterogeneidade não-observada em modelos de risco univariado são apresentados no próximo capítulo.

# Capítulo 2

# Modelos de Risco com Heterogeneidade Não-Observada

Se um modelo de risco que inclui somente covariáveis não se ajusta bem aos dados, pode-se tentar diminuir a variação aleatória introduzindo explicitamente uma fonte de variação que represente o efeito da heterogeneidade não-observada sobre a função de risco e que seja independente das covariáveis.

Considere $T$ o tempo de ocorrência do evento, contínuo, e suponha que a função de risco possa ser expressada como

$$\lambda\left(t|\mathbf{x},\mathbf{w}\right)=\lambda_0\left(t|\mathbf{x}\right)\exp\left(\boldsymbol{\gamma}'\mathbf{w}\right)\ , \tag{2.1}$$

em que $\lambda_0\left(t|\mathbf{x}\right)$ é uma função do tempo e do vetor de covariáveis, $\mathbf{w}$ é o vetor $(q\times 1)$ dos valores de $q$ características não-observadas e $\boldsymbol{\gamma}$ é o vetor $(q\times 1)$ de parâmetros desconhecidos dos efeitos de $\mathbf{w}$.

O que acontece se $\exp\left(\boldsymbol{\gamma}'\mathbf{w}\right)$, o efeito da heterogeneidade não-observada, for omitido do modelo de risco em (2.1)? As estimativas do efeito do tempo e das covariáveis sobre o risco ficariam viciadas, pois seriam contaminadas pelo processo de seleção (ver Exemplo 1.4).

Algumas estratégias têm sido adotadas para a aplicação do modelo em (2.1). Uma estratégia bastante usual supõe uma distribuição paramétrica para a heterogeneidade não-observada e uma forma funcional para $\lambda_0(t|\mathbf{x})$. Uma outra estratégia, sugerida por Heckman e Singer (1982), especifica somente um $\lambda_0(t|\mathbf{x})$ e estima uma distribuição para a heterogeneidade não-observada simultaneamente com os parâmetros do modelo, por meio de um método de estimação não-paramétrico.

Os modelos de risco para tempos discretos podem também incorporar a heterogeneidade não-observada como uma generalização dos modelos apresentados na Seção 1.2.2 (Hamerle, 1986 citado por Blossfeld, Hamerle e Mayer, 1989).

## 2.1 Um Modelo de Fragilidade Multiplicativo

Considere que o interesse inicial seja, de alguma forma, modelar a heterogeneidade não-observada. A análise de covariáveis é discutida na Seção 2.5. Então, todo tipo de heterogeneidade está sendo considerado basicamente como não-observado.

Uma versão alternativa do modelo em (2.1) supõe que os indivíduos diferem, na sua propensão para sofrer o evento de interesse, de uma maneira bastante simples. Denominada *modelo de fragilidade multiplicativo*, a função de risco para um indivíduo no tempo $t$ é dada por

$$\lambda(t|z) = \lambda_0(t)\, z\ , \tag{2.2}$$

em que $z$ é um escalar positivo, representando a fragilidade do indivíduo que se supõe atuar de forma multiplicativa sobre a função de risco e $\lambda_0(t)$ é conhecido como um risco padrão. Este modelo assume que, em qualquer tempo, um indivíduo com fragilidade 2, por exemplo, é duas vezes mais propenso a sofrer o evento que o indivíduo com fragilidade 1 (chamado de indivíduo padrão). Portanto, (2.2) pertence à classe de modelos de riscos proporcionais e o modelo em (2.2) equivale a (2.1) considerando $z = \exp(\boldsymbol{\gamma}'\mathbf{w})$.

Se $z$ fosse observado, $\lambda(t|z)$ poderia ser estimado usando o método de **MV**, assumindo uma forma funcional para $\lambda_0(t)$. Como o valor $z$, em geral, não está disponível para o

pesquisador, outra abordagem é necessária. A fragilidade individual não é observada mas assume-se que tenha uma distribuição na população.

Seja $Z$ uma variável aleatória não-negativa que represente a fragilidade individual, mensurável ou não, que pode manifestar-se de maneira indireta. Por conseguinte, a função de risco $\lambda(t|z)$ é interpretada estatisticamente como uma função de risco condicional de $T$ dado $Z = z$, ou simplesmente, *risco condicionado*.

Como Vaupel, Manton e Stallard (1979) têm mostrado, associados ao risco condicionado estão:

a) o risco condicionado acumulado

$$\Lambda(t|z) = \int_0^t \lambda(u|z)\, du \ ,$$

e por (2.2), note-se que se $\Lambda_0(t)$ é um risco padrão acumulado então

$$\Lambda(t|z) = z \int_0^t \lambda_0(u)\, du = z\Lambda_0(t) \ ; \tag{2.3}$$

b) a função de sobrevivência condicionada

$$S(t|z) = \exp\left[-\Lambda(t|z)\right] \ ,$$

e se $S_0(t)$ é uma função de sobrevivência padrão, por (2.3), tem-se que

$$S(t|z) = \exp\left[-z\Lambda_0(t)\right] = \left\{\exp\left[-\Lambda_0(t)\right]\right\}^z = \left[S_0(t)\right]^z \ . \tag{2.4}$$

A equação (2.4) implica, por exemplo, que se o indivíduo padrão tem função de sobrevivência igual a 50% em algum tempo $t$, um indivíduo com fragilidade 2 tem função de sobrevivência de somente 25% em $t$ e um indivíduo com fragilidade 3, somente 12,5%.

Ao assumir o modelo de fragilidade multiplicativo em (2.2), conseqüentemente supõe-se que, em síntese:

a) a fragilidade é definida em termos relativos, isto é

$$z = \frac{\text{risco do indivíduo com fragilidade } z}{\text{risco do indivíduo padrão}} \; ;$$

b) cada indivíduo nasce com uma certa fragilidade para sofrer o evento e permanece com esse nível de fragilidade por toda a vida;

c) acrescentando uma unidade à fragilidade, a função de sobrevivência fica $[100S_0(t)]\%$ menor.

Outras implicações bem interessantes do modelo de fragilidade multiplicativo verifica-se com as definições de função de sobrevivência e risco não-condicionados.

A função de sobrevivência não-condicionada $S(t)$ é dada pelo valor esperado da sobrevivência condicionada $S(t|Z)$ (ver Seção 1.2.3), ou seja

$$S(t) = E[S(t|Z)] \; . \tag{2.5}$$

Se $Z$ tem função de distribuição $G(z)$ então

$$S(t) = \int_0^\infty S(t|z)\, dG(z) \; . \tag{2.6}$$

Note-se que a fragilidade é não-negativa e de (2.4) e (2.6), Hoougard (1984) tem destacado que

$$S(t) = \int_0^\infty \exp[-z\Lambda_0(t)]\, dG(z) = \mathcal{L}[\Lambda_0(t)] \; , \tag{2.7}$$

em que $\mathcal{L}(s) = \int_0^\infty \exp(-zs)\, dG(z)$ é a *transformada de Laplace* para a distribuição da fragilidade no ponto $s$.

O risco não-condicionado $\lambda(t)$, entretanto, não é o valor esperado de $\lambda(t|Z)$, pois $\lambda(t)$ é o limite de uma razão cujo numerador é uma probabilidade condicionada a $T \geq t$.

Mais precisamente,

$$\lambda(t) = E[\lambda(t|Z, T \geq t)] = \int_0^\infty \lambda(t|z)\, dG(z|T \geq t) \ , \tag{2.8}$$

em que $G(z|T \geq t)$ é a função de distribuição condicional de $Z$ dado $T \geq t$. É possível mostrar que (2.8) é verdadeiro, desenvolvendo-se uma relação entre distribuição conjunta e condicional, juntamente com as equações (1.7) e (2.6) (ver Proposição 1 do Anexo A).

As funções condicionadas $S(t|z)$ e $\lambda(t|z)$, freqüentemente referidas como funções de sobrevivência e de risco individuais, respectivamente, não são estimáveis pois o valor de $Z$ é desconhecido. Diferentemente, as funções não-condicionadas (ou funções populacionais) $S(t)$ e $\lambda(t)$, são funções da distribuição misturada de $T$, cujos parâmetros são estimados por meio de dados sobre o tempo de ocorrência do evento para uma amostra de indivíduos.

Quando se substitui (2.2) em (2.8) verifica-se que

$$\lambda(t) = \lambda_0(t) \int_0^\infty z dG(z|T \geq t) = \lambda_0(t)\, E(Z|T \geq t) \ , \tag{2.9}$$

em que $E(Z|T \geq t)$ é interpretado como a fragilidade média entre os sobreviventes [1] a $t$, que decresce (ou pelo menos não cresce) com o tempo, uma vez que indivíduos mais frágeis tendem a sofrer o evento primeiro, permanecendo entre os sobreviventes somente os indivíduos de menor fragilidade. Para mostrar este resultado basta derivar $E(Z|T \geq t)$ em relação a $t$ e verificar que a derivada é não-positiva (ver Proposição 2 do Anexo A). Portanto, como $E(Z|T \geq t)$ é uma função não-crescente em $t$, tem-se que o risco não-condicionado $\lambda(t) = \lambda_0(t)\, E(Z|T \geq t)$ decresce mais rapidamente (ou cresce mais lentamente) que o risco individual $\lambda(t|z) = \lambda_0(t)\, z$ conforme o risco padrão $\lambda_0(t)$ é decrescente (ou crescente). Isto significa que se for negligenciado o modelo de fragilidade multiplicativo, quando este é realista, o risco individual é superestimado pelo risco não-condicionado.

[1] Por simplicidade, a terminologia usada é adequada a um estudo de mortalidade, porém o tempo de sobrevivência pode representar, por exemplo, o tempo desde o início do casamento até o divórcio.

## 2.2 Distribuições de Fragilidade

Quando for assumido o modelo dado por (2.2), é necessário estipular uma distribuição de probabilidades para a fragilidade $Z$. No Exemplo 1.5, assume-se o modelo de fragilidade em (2.2) com $\lambda_0(t) = 1$ e

$$Z = \begin{cases} \lambda_1, \text{ se o indivíduo pertence a população de baixo risco} \\ \lambda_2, \text{ caso contrário} \end{cases}$$

em que $\lambda_1$ e $\lambda_2$ são desconhecidos pois $Z$ não é observado e

$$P(Z = \lambda_1 | T \geq t) = \pi(t)$$

A seguir são apresentadas outras distribuições para a fragilidade e suas propriedades.

### 2.2.1 Fragilidade Gama

Uma suposição usada por vários autores (Vaupel, Manton e Stallard, 1979; Aalen, 1987) é que $Z$ tenha uma distribuição gama. A distribuição gama implica numa variação contínua para o nível de fragilidade entre os indivíduos. Esta distribuição é freqüentemente adotada para a fragilidade porque é confinada a números não-negativos, assume uma variedade de formas dependendo dos valores dos parâmetros e é bastante conveniente matematicamente. Se $Z \sim gama(\eta, \nu)$, a função densidade de $Z$ é dada por

$$g(z) = \frac{\nu^\eta}{\Gamma(\eta)} z^{\eta-1} \exp(-\nu z), \qquad \text{para } z \geq 0 \tag{2.10}$$

e

$$g(z) = 0, \qquad \text{para } z < 0,$$

em que $\Gamma(\cdot)$ é a função gama, $\eta$ é o parâmetro de forma e $\nu$ é o parâmetro de escala; $\eta, \nu > 0$.

A média e a variância de $Z$ são respectivamente

$$E(Z) = \frac{\eta}{\nu} \qquad \text{e} \qquad Var(Z) = \frac{\eta}{\nu^2} \ . \tag{2.11}$$

Assumindo (2.2) e que $Z \sim gama(\eta, \nu)$, a função de sobrevivência não-condicionada é, de (2.6), dada por

$$\begin{aligned} S(t) &= \int_0^\infty \exp\left[-z\Lambda_0(t)\right] \frac{\nu^\eta}{\Gamma(\eta)} z^{\eta-1} \exp(-\nu z)\, dz \\ &= \frac{\nu^\eta}{\Gamma(\eta)} \int_0^\infty z^{\eta-1} \exp\left\{-z\left[\nu + \Lambda_0(t)\right]\right\} dz \ . \end{aligned}$$

Completando a integral para uma densidade gama com parâmetros $\eta$ e $\nu + \Lambda_0(t)$, tem-se que

$$S(t) = \left[\frac{\nu}{\nu + \Lambda_0(t)}\right]^\eta \ .$$

Uma maneira alternativa de encontrar a função de sobrevivência não-condicionada é usando o resultado em (2.7). Como as transformadas de Laplace são bem conhecidas e muitas estão tabeladas, a tarefa torna-se mais fácil. Para a distribuição gama com parâmetros $\eta$ e $\nu$, a transformada de Laplace é dada por

$$\mathcal{L}(s) = \left[\frac{\nu}{\nu + s}\right]^\eta \ .$$

Calculando para $s = \Lambda_0(t)$, obtem-se o mesmo resultado anterior.

Para se encontrar o risco não-condicionado basta derivar $-\log[S(t)]$ em relação a $t$ e assim obter

$$\lambda(t) = \frac{\eta\lambda_0(t)}{\nu + \Lambda_0(t)} \ . \tag{2.12}$$

**Exemplo 2.1** A fragilidade gama é considerada por Aalen (1987) para analisar o risco de a mulher "expelir" o DIU (Dispositivo Intra-Uterino). Foi registrado o tempo $t$ medido em dias, desde a introdução do DIU até a sua expulsão. As mulheres que retiraram o

DIU por razões médicas ou pessoais e aquelas que tiveram remoção planejada do DIU não foram consideradas. Os dados indicam que a taxa de expulsão é decrescente com o tempo, tendo um nível muito alto no começo e depois decrescendo rapidamente. Uma interpretação para este resultado observado é que o organimo de todas as mulheres tende a adaptar-se ao DIU. Outra interpretação para o decrescimento da taxa de expulsão é que o risco de ocorrer uma expulsão varia muito entre as mulheres. As mulheres de alto risco tenderão a deixar o grupo de risco (por causa da expulsão) muito cedo, enquanto as de baixo risco tenderão a permanecer em risco. Uma das declarações ou ambas podem ser verdadeiras e não é possível por meio do conjunto de dados considerados aqui decidir os méritos relativos aos dois tipos de variação. No entanto, para ilustrar a fragilidade gama, considere um modelo em que somente a última variação mensionada está presente, isto é, assuma que cada mulher tenha ao longo do tempo um risco constante de expulsão, mas haja uma variação desse risco entre as mulheres. Suponha então que o risco individual seja dado por

$$\lambda(t|z) = z$$

ou seja, assume-se o modelo em (2.2) com $\lambda_0(t) = 1$, para que o risco individual seja constante no tempo. Suponha também que risco individual varie entre as mulheres conforme uma distribuição gama com densidade dada por (2.10). Então, de (2.12), tem-se que o risco não-condicionado é dado por

$$\lambda(t) = \frac{\eta}{\nu + t} ;$$

sendo que $\lambda(t)$ é uma função decrescente em $t$. Aalen (1987), usando os estimadores de **MV** para $\eta$ e $\nu$, verificou que o modelo estimado se ajusta bem aos dados observados, mesmo que a interpretação para os dados possa ter sido falsa. Não há como determinar se a distribuição do tempo até a expulsão do DIU é realmente uma mistura de distribuições dada por

$$exponencial(Z) \bigwedge_{z} gama(\eta, \nu) \ ,$$

porque o risco não-condicionado é o único que pode ser estimado e este é idêntico à função de risco de uma população homogênea com distribuição Pareto do Tipo II (ver Exemplo 1.2). Ou seja, a população pode ser homogênea com um risco decrescente ou a população pode ser composta de riscos individuais constantes no tempo mas, são distribuídos entre os indivíduos conforme a distribuição gama.

Uma outra propriedade da fragilidade gama, mostrada na Seção 2.3, é que a fragilidade entre os sobreviventes a um determinado tempo também tem distribuição gama com o mesmo parâmetro de forma, mas com parâmetro de escala deslocado. Outras distribuições também têm esta propriedade, como a distribuição gaussiana inversa (Hoougard, 1984).

### 2.2.2 Fragilidade Gaussiana Inversa

A distribuição gaussiana inversa tem função densidade dada por

$$g(z) = \sqrt{\frac{\eta}{\pi}} z^{-\frac{3}{2}} \exp\left(\sqrt{4\eta\nu} - \nu z - \frac{\eta}{z}\right) ,$$

dependendo de dois parâmetros $\eta$ e $\nu$; $\eta, \nu > 0$.

Esta distribuição tem, respectivamente, média e variância iguais a

$$E(Z) = \sqrt{\frac{\eta}{\nu}} \qquad \text{e} \qquad Var(Z) = \frac{\sqrt{\eta}}{2}\nu^{-\frac{3}{2}} . \tag{2.13}$$

Pela equação (2.6), se $Z \sim GI(\eta, \nu)$, isto é, se a fragilidade tem distribuição gaussiana inversa com parâmetros $\eta$ e $\nu$, a função de sobrevivência não-condicionada é

$$\begin{aligned} S(t) &= \int_0^\infty \exp\left[-z\Lambda_0(t)\right] \sqrt{\frac{\eta}{\pi}} z^{-\frac{3}{2}} \exp\left(\sqrt{4\eta\nu} - \nu z - \frac{\eta}{z}\right) dz \\ &= \int_0^\infty \sqrt{\frac{\eta}{\pi}} z^{-\frac{3}{2}} \exp\left\{\sqrt{4\eta\nu} - z\left[\nu + \Lambda_0(t)\right] - \frac{\eta}{z}\right\} dz . \end{aligned}$$

Completando a integral para uma densidade gaussiana inversa com parâmetros $\eta$ e

$\nu + \Lambda_0(t)$ obtém-se que

$$S(t) = \exp\left\{\sqrt{4\eta\nu} - \sqrt{4\eta\left[\nu + \Lambda_0(t)\right]}\right\} .$$

Da mesma forma que para a fragilidade gama, obtém-se a função risco não-condicionada para a fragilidade gaussiana inversa, que é dada por

$$\lambda(t) = \frac{\sqrt{\eta}\lambda_0(t)}{\sqrt{\nu + \Lambda_0(t)}} .$$

## 2.3 Distribuições de Fragilidade Condicionada

Se $T$ e $Z$ têm densidade conjunta, a função densidade de $Z$ condicionada entre os sobreviventes a $t$ pode ser escrita como

$$g(z|T \geq t) = \frac{P(T \geq t|z)\, g(z)}{P(T \geq t)} = \frac{S(t|z)\, g(z)}{S(t)} , \tag{2.14}$$

pela definição de probabilidade condicional e função de sobrevivência.

Obtém-se, de maneira similar, a distribuição condicional de $Z$ dado $T = t$, isto é, a distribuição da fragilidade condicionada entre os que sobreviveram até o tempo $t$. A função densidade condicional de $Z$ dado $T = t$ pode ser escrita como

$$g(z|T = t) = \frac{f(t|z)\, g(z)}{f(t)} = \frac{-\frac{\partial}{\partial t} S(t|z)\, g(z)}{-\frac{d}{dt} S(t)} . \tag{2.15}$$

Assumindo uma forma para $S(t|z)$ pode-se comparar várias escolhas para $g(z)$ a partir das equações (2.14) e (2.15). A seguir, como exemplo, são comparadas as fragilidades gama e gaussiana inversa, supondo que a função de sobrevivência condicionada é a do modelo de fragilidade multiplicativo, isto é, $S(t|z) = \exp\left[-z\Lambda_0(t)\right]$. Sob esse modelo, sabe-se que a fragilidade média entre os sobreviventes não crescerá com o tempo, qualquer que seja a distribuição atribuída a $Z$ (ver Seção 2.1). A fim de comparar a fragilidade

entre os sobreviventes para as diferentes distribuições de fragilidade, seria então conveniente considerar o coeficiente de variação de $Z$ entre os sobreviventes, já que é uma medida normalizada. O coeficiente de variação de $Z|T \geq t$ é dado por

$$CV\left(Z|T \geq t\right) = \frac{\sqrt{Var\left(Z|T \geq t\right)}}{E\left(Z|T \geq t\right)} .$$

**Distribuição Condicionada da Fragilidade Gama**

Seja $Z \sim gama\left(\eta, \nu\right)$. Substituindo em (2.14) as expressões correspondentes à distribuição gama, tem-se que a função densidade condicional de $Z$ dado $T \geq t$ é

$$g\left(z|T \geq t\right) = \frac{\left[\nu + \Lambda_0\left(t\right)\right]^{\eta}}{\Gamma\left(\eta\right)} z^{\eta-1} \exp\left\{-z\left[\nu + \Lambda_0\left(t\right)\right]\right\} ,$$

que é densidade gama com parâmetros $\eta$ e $\nu + \Lambda_0\left(t\right)$. Isto quer dizer que, se a população inicial tem fragilidade gama com parâmetros $\eta$ e $\nu$, então a fragilidade dos sobreviventes a $t$ também tem uma distribuição gama com parâmetros $\eta$ e $\nu + \Lambda_0\left(t\right)$.

Supondo $E\left(Z\right) = 1$, assume-se por (2.2) que em média $\lambda\left(t|z\right) = \lambda_0\left(t\right)$. Sob essa suposição [2], de (2.11), tem-se que

$$\eta = \nu = \frac{1}{Var\left(Z\right)}$$

e fazendo

$$\sigma^2 = Var\left(Z\right)$$

tem-se que $Z|T \geq t$ tem distribuição gama com parâmetros $\frac{1}{\sigma^2}$ e $\frac{1}{\sigma^2} + \Lambda_0\left(t\right)$. A esperança,

[2] Esta restrição não leva à perda da generalidade visto que o nível médio de fragilidade pode ser absorvido pelo risco padrão (veja Proposição 3 do Anexo A). Quando $E\left(Z\right) = 1$, as fórmulas simplificam-se e $CV\left(Z|T \geq t\right)$ representa, de certo modo, a heterogeneidade da fragilidade da população sobrevivente, pois

$$\left[CV\left(Z|T \geq t\right)\right]^2 = Var\left(Z|T \geq t\right) .$$

a variância e o coeficiente de variação de $Z|T \geq t$ são dados, respectivamente, por

$$E\left(Z|T \geq t\right) = \frac{1}{1+\sigma^2\Lambda_0\left(t\right)}\ , \tag{2.16}$$

$$Var\left(Z|T \geq t\right) = \frac{1}{\sigma^2\left[1+\sigma^2\Lambda_0\left(t\right)\right]^2} \tag{2.17}$$

e

$$CV\left(Z|T \geq t\right) = \frac{1}{\sigma}\ . \tag{2.18}$$

A função densidade condicional de $Z$ dado $T = t$, obtida de (2.15) é

$$g\left(z|T=t\right) = \frac{\left[\nu+\Lambda_0\left(t\right)\right]^{\eta+1}}{\Gamma\left(\eta+1\right)} z^{\eta} \exp\left\{-z\left[\nu+\Lambda_0\left(t\right)\right]\right\}\ .$$

Esta função é claramente uma densidade gama com parâmetros $\eta+1$ e $\nu+\Lambda_0\left(t\right)$. Portanto, se a fragilidade na população inicial tem distribuição gama com parâmetros $\eta$ e $\nu$, então a fragilidade entre os que sobreviveram até $t$ tem distribuição gama com parâmetros $\eta+1$ e $\nu+\Lambda_0\left(t\right)$. Além disso, quando $E\left(Z\right)=1$ e $Var\left(Z\right)=\sigma^2$, a fragilidade média entre os sobreviventes até $t$ é dada por

$$E\left(Z|T=t\right) = \frac{1+\sigma^2}{1+\sigma^2\Lambda_0\left(t\right)}$$

e portanto, $E\left(Z|T=t\right)$ é, coerentemente, sempre maior que a fragilidade média entre os sobreviventes a $t$.

### Distribuição Condicionada da Fragilidade Gaussiana Inversa

Seja $Z \sim GI\left(\eta,\nu\right)$. Analogamente ao que foi estabelecido para a fragilidade gama, tem-se que a função densidade condicional de $Z$ dado $T \geq t$ é

$$g\left(z|T \geq t\right) = \sqrt{\frac{\eta}{\pi}} z^{-\frac{3}{2}} \exp\left\{\sqrt{4\eta\left[\nu+\Lambda_0\left(t\right)\right]} - z\left[\nu+\Lambda_0\left(t\right)\right] - \frac{\eta}{z}\right\}\ ,$$

que é a função densidade da distribuição gaussiana inversa com parâmetros $\eta$ e $\nu+\Lambda_0(t)$.

Se $Z$ tem distribuição gaussiana inversa com média 1 e variância $\sigma^2$ então por (2.13)

$$\eta = \nu = \frac{1}{2\sigma^2}$$

e $Z|T \geq t$ tem distribuição gaussiana inversa com parâmetros $\frac{1}{2\sigma^2}$ e $\frac{1}{2\sigma^2} + \Lambda_0(t)$. A esperança, a variância e o coeficiente de variação de $Z|T \geq t$ são dados, respectivamente, por

$$E(Z|T \geq t) = \left[\sqrt{1 + 2\sigma^2\Lambda_0(t)}\right]^{-\frac{1}{2}} , \tag{2.19}$$

$$Var(Z|T \geq t) = \sigma^2 \left[1 + 2\sigma^2\Lambda_0(t)\right]^{-\frac{3}{2}} \tag{2.20}$$

e

$$CV(Z|T \geq t) = \sigma \left[1 + 2\sigma^2\Lambda_0(t)\right]^{-\frac{1}{4}} . \tag{2.21}$$

Hoougard (1984) comenta que fazendo $Z \sim GI(\eta, \nu)$, a distribuição da fragilidade entre os que sobreviveram até um determinado tempo $t$ tem distribuição gaussiana inversa generalizada com função densidade dada por

$$g(z|T = t) = \sqrt{\frac{\eta\left[\nu + \Lambda_0(t)\right]}{\pi}} z^{-\frac{1}{2}} \exp\left\{\sqrt{4\eta\left[\nu + \Lambda_0(t)\right]} - z\left[\nu + \Lambda_0(t)\right] - \frac{\eta}{z}\right\} .$$

Outras distribuições têm sido atribuídas para a fragilidade como as distribuições estáveis sugeridas por Hoougard (1986) e a distribuição de Poisson composta sugerida por Aalen (1988). Assumindo que $E(Z) = 1$, estas distribuições têm fragilidade média entre os sobreviventes expressa por

$$E(Z|T \geq t) = \frac{1}{\left[1 + \frac{\sigma^2}{c}\Lambda_0(t)\right]^c} , \tag{2.22}$$

em que $\sigma^2 = Var(Z)$ e $c$ é uma constante (Rodrigues, 1992).

A fragilidade gama e a gaussiana inversa assumem a forma de (2.22) quando $c = 1$ e

$c = \frac{1}{2}$, respectivamente.

## 2.4 Comparação entre a Fragilidade Gama e a Gaussiana Inversa

Os resultados obtidos na Seção 2.3 sugerem uma maneira de comparar a fragilidade gama com a fragilidade gaussiana inversa quando é assumido o modelo em (2.2) com $E(Z) = 1$ e $Var(Z) = \sigma^2$.

Ao assumir fragilidade gama ou gaussiana inversa verifica-se claramente que a fragilidade média entre os sobreviventes declina com o tempo, dado que os denominadores de (2.16) e (2.19) crescem quando $t$ cresce. Note-se por (2.16) e (2.19) que a fragilidade média declinará mais rapidamente, isto é, o processo de seleção operará mais rapidamente quando a população for mais heterogênea no início ($\sigma^2$ grande) ou o risco for alto ($\Lambda_0(t)$ grande).

Se a fragilidade tem distribuição gama então:

a) A variância da fragilidade entre os sobreviventes é decrescente no tempo, dado que o denominador de (2.17) cresce quando $t$ cresce. Portanto, a fragilidade da população em risco torna-se mais homogênea com o tempo;

b) o coeficiente de variação da fragilidade entre os sobreviventes em (2.18) não varia com o tempo. Em outras palavras, a fragilidade média e o desvio-padrão entre os sobreviventes diminuem com o tempo, mas diminuem proporcionalmente, uma vez que o coeficiente de variação não depende do tempo.

Se a fragilidade tem distribuição gaussiana inversa então de (2.20) e (2.21) verifica-se que tanto a variância como o coeficiente de variação da fragilidade entre os sobreviventes a $t$ decrescem com o tempo, significando que a heterogeneidade entre os sobreviventes é cada vez menor, mesmo quando comparada com a média, e portanto, a fragilidade da população sobrevivente torna-se mais homogênea com o tempo.

A fragilidade gama implica que a heterogeneidade com relação à fragilidade média da população de sobreviventes a $t$, representada pelo coeficiente de variação de $Z|T \geq t$, é constante no tempo. Se este resultado não corresponde à realidade, uma alternativa é supor fragilidade gaussiana inversa que torna a população de sobreviventes relativamente mais homogênea com o tempo.

## 2.5 Um Modelo de Fragilidade Multiplicativo com Covariáveis

Uma generalização do modelo em (2.2) é considerar um modelo de fragilidade multiplicativo, no qual o risco padrão é também função do vetor $\mathbf{x}$ dos valores de $p$ covariáveis do indivíduo, ou seja

$$\lambda(t|\mathbf{x}, z) = \lambda_0(t|\mathbf{x})\, z \ , \tag{2.23}$$

em que $\mathbf{x}$ e $z$ são independentes.

Uma forma sugestiva para a função de risco padrão é

$$\lambda_0(t|\mathbf{x}) = \lambda_0^*(t) \exp(\boldsymbol{\beta}'\mathbf{x}) \ , \tag{2.24}$$

em que $\lambda_0^*(t)$ é um risco padrão que depende somente de $t$, e resulta que

$$\lambda(t|\mathbf{x}, z) = \lambda_0^*(t) \exp(\boldsymbol{\beta}'\mathbf{x})\, z \ . \tag{2.25}$$

Associados ao risco condicionado $\lambda(t|\mathbf{x}, z)$ em (2.25) estão

a) risco condicionado acumulado

$$\Lambda(t|\mathbf{x}, z) = \int_0^t \lambda(u|\mathbf{x}, z)\, du = z\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}) \ ;$$

b) função de sobrevivência condicionada

$$S(t|\mathbf{x}, z) = \exp[-\Lambda(t|\mathbf{x}, z)] = \exp[z\Lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{x})] = [S_0(t)]^{z\exp(\boldsymbol{\beta}'\mathbf{x})}. \quad (2.26)$$

A função de sobrevivência não-condicionada é uma generalização de (2.5), dada por

$$S(t|\mathbf{x}) = E[S(t|\mathbf{x}, Z)] .$$

Se $Z$ tem função de distribuição $G(z)$ então

$$S(t|\mathbf{x}) = \int_0^\infty S(t|\mathbf{x}, z)\, dG(z) . \quad (2.27)$$

Desenvolvendo (2.27) em termos de (2.26) tem-se que

$$S(t|\mathbf{x}) = \int_0^\infty \exp[-z\Lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{x})]\, dG(z) = \mathcal{L}[\Lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{x})] .$$

E o risco não-condicionado de $T$ é

$$\lambda(t|\mathbf{x}) = E[\lambda(t|\mathbf{x}, Z, T \geq t)] = \int_0^\infty \lambda(t|\mathbf{x}, z)\, dG(z|T \geq t) . \quad (2.28)$$

Quando o modelo em (2.25) é adotado, devem ser atribuídas distribuições de probabilidades a $Z$, como a distribuição gama ou a gaussiana inversa, ficando mantidos os resultados obtidos nas Seções 2.2 e 2.3 com as devidas generalizações.

## 2.6 Sensibilidade dos Modelos de Fragilidade

Uma questão de interesse é verificar quão sensíveis são as estimativas dos parâmetros do modelo em (2.25) com relação às escolhas de $\lambda_0(t)$ e da distribuição de $Z$. Heckman e Singer (1982) demonstram que, assumindo uma representação paramétrica para $\lambda_0(t)$ em (2.25), os resultados podem ser sensíveis à escolha da distribuição para a fragilidade,

mesmo quando for escolhida uma distribuição com forma paramétrica flexível, como a distribuição gama.

Uma maneira de obter modelos de fragilidade mais robustos é estimando-se uma distribuição para a fragilidade. Como uma solução, Heckman e Singer (1982) propuseram um método de estimação de máxima verossimilhança não-paramétrico para estimar a distribuição da fragilidade, usando o algoritmo EM de Dempster, Laird e Rubin (1977).

Laird (1978) e Heckman e Singer (1982) mostram que uma abordagem de máxima verossimilhança não-paramétrica para estimação da distribuição de $Z$ leva a uma distribuição discreta, em que $Z$ assume valores $z_1, z_2, \ldots, z_k$ sendo que para $j = 1, 2, \ldots, k$, $P(Z = z_j) = \pi_j$ com $\sum_{j=1}^{k} \pi_j = 1$. Sob essas condições precedentes, a sobrevivência não-condicionada, em (2.27), é simplesmente

$$S(t|\mathbf{x}) = \sum_{j=1}^{k} S(t|\mathbf{x}, z_j)\, \pi_j$$

Trussel e Richards (1985) verificaram, por meio de estudos empíricos, que os modelos estimados pela solução de Heckman e Singer são sensíveis à escolha do risco padrão, cuja forma é supostamente paramétrica. De fato, há evidências mais recentes de que a análise pode ser mais sensível às suposições acerca do risco padrão do que às suposições para a distribuição da fragilidade. Isto é comprovado por alguns experimentos de simulação em Hamerle e Moller (1990), citados por Blossfeld e Hamerle (1992).

Um outro fato importante para ser considerado é que os modelos de fragilidade vistos até aqui, consideram que as covariáveis e a fragilidade são independentes. A suposição de independência, neste caso, é uma restrição muito forte, já que a heterogeneidade não-observada de um indivíduo deveria estar associada à parte observada da heterogeneidade. Se a independência não pode ser comprovada, a função de sobrevivência não-condicionada é então dada por

$$S(t|\mathbf{x}) = \int_0^{\infty} S(t|\mathbf{x}, z)\, dG(z|\mathbf{x}) \ , \tag{2.29}$$

em que $G(z|\mathbf{x})$ é a função de distribuição condicional de $Z$ dado $\mathbf{X} = \mathbf{x}$ e sendo que $\mathbf{X}' = (X_1, X_2, \ldots, X_p)$ é o vetor de p covariáveis observadas. Note-se que este modelo assume que a distribuição de $Z$ é determinada pelos valores observados de $\mathbf{X}$. Sensibilidades nas estimativas dos parâmetros pertencentes à classe de modelos em (2.23) podem surgir devido a um erro substancial entre a distribuição marginal $G(z)$ e a distribuição condicional $G(z|\mathbf{x})$ (Manton, Singer e Woodbury, 1992).

Se $\mathbf{X}$ e $Z$ não são independentes, o risco não-condicionado de $T$ é uma generalização de (2.28) dado por

$$\lambda(t|\mathbf{x}) = \int_0^\infty \lambda(t|\mathbf{x}, z)\, dG(z|T \geq t, \mathbf{x}) \ . \tag{2.30}$$

De (2.25) e (2.30), tem-se que

$$\begin{aligned} \lambda(t|\mathbf{x}) &= \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{x}) \int_0^\infty z dG(z|T \geq t, \mathbf{x}) \\ &= \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{x})\, E(Z|T \geq t, \mathbf{x}) \ . \end{aligned} \tag{2.31}$$

Aplicando logaritmo na equação (2.31), verifica-se que

$$\log[\lambda(t|\mathbf{x})] = a(t) + \boldsymbol{\beta}'\mathbf{x} + b(t, \mathbf{x}) \ ,$$

em que

$a(t) = \log[\lambda_0(t)]$ é o efeito do tempo,

$\boldsymbol{\beta}'\mathbf{x}$ é o efeito das covariáveis e

$b(t, \mathbf{x}) = \log[E(Z|T \geq t, \mathbf{x})]$ é o efeito da interação entre tempo e covariáveis.

Intuitivamente, $E(Z|T \geq t, \mathbf{x})$ deveria decrescer com o tempo, devido ao processo de seleção de indivíduos mais fortes e como uma generalização de (2.22), Rodrigues (1992) destaca que para algumas distribuições de fragilidade

$$E(Z|T \geq t, \mathbf{x}) = \frac{1}{\left[1 + \frac{\sigma^2}{c}\Lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{x})\right]^c}$$

com $c = 1$ e $c = 1/2$ para a fragilidade gama e gaussiana inversa, respectivamente, desde que $\sigma^2 = Var(Z)$ e $c$ é uma constante. Rodrigues (1992) sugere uma modelagem flexível para a interação $b(t, \mathbf{x})$ ou então que sejam feitas suposições para a distribuição da fragilidade, já que $b(t, \mathbf{x})$ pode depender dos parâmetros da distribuição de $Z|T \geq t, \mathbf{x}$.

Quando há interesse em decompor a heterogeneidade não-observada em um vetor $\mathbf{W} = (W_1, W_2, \ldots, W_q)$ de $q$ variáveis não-observadas, ao invés de uma única fragilidade escalar $Z$, é necessário atribuir uma distribuição multivariada para $\mathbf{W}$ e que o risco não-condicionado de $T$, é dado por

$$\lambda(t|\mathbf{x}) = \lambda_0(t|\mathbf{x}) E[\exp(\boldsymbol{\gamma}'\mathbf{W}) | T \geq t, \mathbf{x}] ,$$

com $\lambda_0(t|\mathbf{x})$ igual a (2.24), por exemplo. O processo de estimação envolveria os parâmetros da distribuição multivariada de $\mathbf{W}$ e os parâmetros do vetor $\boldsymbol{\gamma}$ dos efeitos de $\mathbf{W} = \mathbf{w}$ sobre o risco.

A principal desvantagem dos modelos de fragilidade é que, quando o tempo de ocorrência do evento é univariado, esses modelos resultam em misturas não-identificáveis, tanto para o caso discreto como para o caso contínuo (Oakes, 1989; Mare, 1994).

Estudos teóricos sobre a dinâmica de populações heterogêneas começam, primeiramente, assumindo um modelo para a história de eventos das subpopulações homogêneas e assumindo uma distribuição para essas subpopulações. Depois, é derivado o modelo para a população toda. Tais estudos têm sido muito importantes pois os modelos a nível individual podem ser mais simples e têm uma interpretação mais direta e simplificada que os modelos das populações compostas. O esforço inverso, de tentar decompor a dinâmica da população observada em dois componentes — a distribuição para as subpopulações e a dinâmica das subpopulações — tem sido mais complexo (Vaupel, 1990).

Nas discusões sobre os modelos de fragilidade apresentadas na Seção 2.1, partiu-se de um risco condicionado para um não-condicionado por meio de uma mistura, em que a transformada de Laplace poderia ser utilizada na análise. A transformada de Laplace

num ponto $s$, dada por

$$\mathcal{L}(s) = \int_0^\infty \exp(-zs)\, dG(z) \ ,$$

tem importância tanto numérica como teórica:

a) dada a função de distribuição de fragilidade $G(z)$, o cálculo da transformada de Laplace é bem conhecido com algoritmos eficientes;

b) dada a transformada de Laplace $\mathcal{L}(s)$, recuperar $G(z)$ é um problema mal condicionado, no sentido de que mudanças suaves em $\mathcal{L}(s)$ induzem a uma enorme flutuação em $G(z)$. Estatisticamente, isto significa que a estimação dos parâmetros da distribuição de fragilidade apresenta dificuldades (Rodrigues, 1992).

Assumindo uma distribuição para a fragilidade, é possível sob o modelo de fragilidade multiplicativo encontrar a forma do risco padrão (e do risco condicionado) a partir do risco não-condicionado. Na Seção 2.7 são desenvolvidas fórmulas de inversão (Rodrigues, 1992) para encontrar o risco padrão dada uma fragilidade gama, primeiramente, e dada uma fragilidade gaussiana inversa, em seguida.

## 2.7 Fórmulas de Inversão

Seja $Z$ a fragilidade e seja $T$ o tempo contínuo de ocorrência do evento para um indivíduo com risco individual dado pelo modelo de fragilidade multiplicativo

$$\lambda(t|z) = \lambda_0(t)\, z \ ,$$

como inicialmente foi proposto na Seção 2.1.

Suponha que a fragilidade tenha distribuição gama com esperança 1 e variância $\sigma^2$ (ver Seção 2.2.1), para a qual tem sido mostrado que o risco não-condicionado de $T$ é

$$\lambda(t) = \frac{\lambda_0(t)}{1 + \sigma^2 \Lambda_0(t)} \ . \tag{2.32}$$

Integrando (2.32), obtém-se uma expressão para o risco acumulado $\Lambda(t)$. Note-se que

$$\frac{d}{dt}\log\left[1+\sigma^2\Lambda_0(t)\right]=\frac{\sigma^2\lambda_0(t)}{1+\sigma^2\Lambda_0(t)}\ . \tag{2.33}$$

Como $\Lambda(0)=0$, então a integração dos dois lados de (2.32) no intervalo $[0,t]$ resulta em

$$\Lambda(t)=\frac{1}{\sigma^2}\log\left[1+\sigma^2\Lambda_0(t)\right]$$

e ao isolar $\Lambda_0(t)$ em (2.33) tem-se que

$$\Lambda_0(t)=\frac{1}{\sigma^2}\left\{\exp\left[\sigma^2\Lambda(t)\right]-1\right\}\ . \tag{2.34}$$

Para obter uma fórmula de inversão de $\lambda_0(t)$, basta derivar (2.34) em relação a $t$, resultando em

$$\lambda_0(t)=\lambda(t)\exp\left[\sigma^2\Lambda(t)\right]\ . \tag{2.35}$$

Dadas a distribuição não-condicionada de $T$ e a variância da fragilidade gama, é possível encontrar a forma de $\lambda_0(t)$ através de (2.35).

**Exemplo 2.2** Se a distribuição não-condicionada de $T$ é exponencial com parâmetro $\alpha$, isto significa que

$$\lambda(t)=\alpha \qquad \text{e} \qquad \Lambda(t)=\alpha t\ .$$

Usando a fórmula de inversão (2.35) pode-se encontrar o risco padrão

$$\lambda_0(t)=\alpha\exp\left(\sigma^2\alpha t\right)$$

e o risco condicionado é dado por

$$\lambda(t|z)=z\alpha\exp\left(\sigma^2\alpha t\right)\ .$$

Assim, sob o modelo de fragilidade multiplicativo, um risco não-condicionado cons-

tante no tempo seria conseqüência do pressuposto de que a população seja composta de indivíduos heterogêneos que variam entre si conforme uma fragilidade gama e cujo risco condicionado cresce exponencialmente com o tempo, uma vez que $z$, $\alpha$ e $\sigma^2$ são valores não-negativos.

No caso de uma fragilidade gaussiana inversa com média 1 e variância $\sigma^2$, tem-se que

$$\lambda(t) = \frac{\lambda_0(t)}{\sqrt{1 + 2\sigma^2 \Lambda_0(t)}} . \tag{2.36}$$

Note-se que o lado direito de (2.36) é de fato a derivada em relação a $t$ de

$$\frac{1}{\sigma^2}\sqrt{1 + 2\sigma^2 \Lambda_0(t)} .$$

Assim, a integração dos dois lados de (2.36) no intervalo $[0, t]$ resulta em

$$\Lambda(t) = \frac{1}{\sigma^2}\left\{\left[1 + 2\sigma^2 \Lambda_0(t)\right]^{\frac{1}{2}} - 1\right\} \tag{2.37}$$

e isolando $\Lambda_0(t)$ em (2.37), tem-se que

$$\Lambda_0(t) = \frac{\left[1 + \sigma^2 \Lambda(t)\right]^2 - 1}{2\sigma^2} . \tag{2.38}$$

Resta agora derivar (2.38) em relção a $t$ para obter o risco padrão, dado pela seguinte fórmula de inversão

$$\lambda_0(t) = \lambda(t)\left[1 + \sigma^2 \Lambda(t)\right] . \tag{2.39}$$

**Exemplo 2.3** É possível produzir um risco não-condicionado constante no tempo, como resultado de algum risco condicionado sujeito a uma fragilidade gaussiana inversa. Suponha-se que o risco não-condicionado de $T$ seja dado por $\lambda(t) = \alpha$, portanto $\Lambda(t) = \alpha t$, e se $E(Z) = 1$ e $Var(Z) = \sigma^2$, basta usar a fórmula de inversão (2.39) para

encontrar um risco padrão dado por

$$\lambda_0(t) = \alpha\left(1 + \sigma^2\alpha t\right) ,$$

que é linear em $t$ e o risco condicionado de $T$ é dado por

$$\lambda(t|z) = z\alpha\left(1 + \sigma^2\alpha t\right) .$$

Assim, sob o modelo de fragilidade multiplicativo, uma distribuição exponencial para $T$ pode vir de um risco condicionado linearmente crescente com o tempo, sujeito a uma fragilidade gaussiana inversa.

Dos Exemplos 2.2 e 2.3 surge um problema de identificabilidade. Dado um risco não-condicionado constante no tempo, há três interpretações alternativas para a composição e dinâmica populacional:

a) a população é homogênea com um risco que não varia entre os indivíduos e é constante no tempo, pois $T$, o tempo de ocorrência do evento para um indivíduo, teria uma distribução exponencial;

b) a população é heterogênea com um risco que varia de um indivíduo para outro conforme uma fragilidade gama e que cresce exponencialmente com o tempo, pois a distribuição não-condicionada de $T$ seria a mistura de uma distribuição tendo função de risco exponencialmente crescente no tempo com uma distribuição gama;

c) a população é heterogênea com um risco que varia de um indivíduo para outro conforme uma fragilidade gaussiana inversa e que cresce linearmente com o tempo, pois a distribuição não-condicionada de $T$ seria a mistura de uma distribuição tendo função de risco linearmente crescente no tempo com uma distribuição gaussiana inversa.

Nenhuma quantidade maior de dados sobre $\lambda(t)$ pode ajudar a distinguir entre estes modelos porque eles têm idênticas consequências observáveis: um risco populacional constante no tempo. A menos que se assuma uma distribuição para $Z$, não há como reconhecer os riscos em nível individual.

Pela existência das fórmulas de inversão (2.35) e (2.39), tem-se mostrado que se $Z$ tem uma distribuição gama ou uma distribuição gaussiana inversa, então para um dado $\lambda(t)$ é sempre possível encontrar uma função de risco $\lambda_0(t)$ que satisfaça

$$\lambda(t) = \lambda_0(t)\, E(Z|T \geq t) \ .$$

Isto significa que, quando a função de distribuição $G(z)$ provém da distribuição gama ou da gaussiana inversa, o modelo $(\lambda_0, G)$ para $T$ não pode ser empiricamente distinguível do modelo de população homogênea baseado somente em $\lambda(t)$, nem estes modelos de fragilidade gama e gaussiana inversa podem ser distinguidos um do outro.

O Exemplo 2.4 extraído de Rodrigues (1992) ilustra a falta de identificabilidade para interpretar riscos proporcionais.

**Exemplo 2.4** Suponha que se tenha ajustado ao tempo de ocorrência de um evento, um modelo de riscos proporcionais constantes ao longo do tempo, dado por

$$\lambda(t|\mathbf{x}) = \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x}) \ . \tag{2.40}$$

Mas se for assumido que existe uma fragilidade que se distribui entre os indivíduos conforme uma gama com média 1 e variância $\sigma^2$ (sendo $Z$ independente das covariáveis), o modelo em (2.40) é equivalente (observacionalmente) ao seguinte modelo em nível individual

$$\lambda(t|\mathbf{x}, z) = z \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x}) \exp\left[\sigma^2 t \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x})\right] \ , \tag{2.41}$$

segundo a fórmula de inversão em (2.35) adaptada para incorporar covariáveis.

O modelo em (2.41) é conhecido como modelo de tempo de falha acelerado com risco

padrão Gompertz e fragilidade gama (ver Seção 1.2.2). E o risco não-condicionado $\lambda(t|\mathbf{x})$ para este modelo é constante no tempo (ver Seção 2.5).

Outro modelo em nível individual também poderia resultar em um risco não-condicionado constante no tempo. O modelo com um risco condicionado sujeito a uma fragilidade gaussiana inversa com média 1 e variância $\sigma^2$ pode gerar um modelo com risco não-condicionado constante no tempo. Assim, aplicando-se a fórmula de inversão em (2.39), com a inclusão de covariáveis, resulta que o risco individual é dado por

$$\lambda(t|\mathbf{x}, z) = z \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x}) \left[1 + \sigma^2 t \exp(\beta_0 + \boldsymbol{\beta}'\mathbf{x})\right] , \tag{2.42}$$

quando $E(Z) = 1$, $Var(Z) = \sigma^2$ e $Z$ é independente das covariáveis.

Os riscos individuais em (2.41) e (2.42) são modelos de riscos não-proporcionais. Os modelos em (2.40), (2.41) e (2.42) têm implicações comportamentais muito diferentes, mas possuem idênticas conseqüências observacionais: um risco populacional constante no tempo. Este problema fundamental de falta de identificabilidade não pode ser resolvido sem informações externas que venham a ser assumidas na análise.

Hoem (1990) tem demonstrado que os resultados dos Exemplos 2.2 a 2.4 podem ser estendidos para qualquer distribuição de fragilidade com média finita. A prova é feita a partir de um risco $m(t)$ não-estocástico e de uma função de distribuição $G(z)$. A demonstração de Hoem é apresentada nas próximas seções deste capítulo, dividida em duas partes: na Seção 2.8 parte-se inicialmente de um modelo de risco com uma função de risco $m(t)$ que não inclui a heterogeneidade não-observada, para reescrevê-lo como um modelo de fragilidade multiplicativo com risco não-condicionado $m(t)$; e na Seção 2.9 obtém-se um resultado análogo, partindo de um modelo de fragilidade multiplicativo com risco não-condicionado $m(t)$, que é idêntico a um risco não-condicionado resultante de outro modelo de fragilidade multiplicativo.

## 2.8 Modelos de Fragilidade Equivalentes a um Modelo sem Heterogeneidade Não-Observada

Conforme proposto por Hoem (1990), seja $m(\cdot)$ uma função de risco não-estocástica e seja $Z$ uma variável aleatória não-negativa com função de distribuição $G(\cdot)$ de média finita. Define-se

$$c_G(s) = \frac{\int_0^\infty z \exp(-zs)\, dG(z)}{\int_0^\infty \exp(-zs)\, dG(z)}. \tag{2.43}$$

Note-se que $c_G(s)$ está bem definido pois $E_G(Z) = \int_0^\infty z dG(z) < \infty$.

Define-se ainda

$$C_G(s) = \int_0^s c_G(u)\, du, \qquad V_G(s) = C_G^{-1}(s),$$

$$v_G(s) = \frac{d}{ds} V_G(s), \qquad M(t) = \int_0^t m(u)\, du,$$

$$a_G(t) = m(t)\, v_G(M(t)) \tag{2.44}$$

e

$$A_G(t) = \int_0^t a_G(u)\, du.$$

Da definição de $M(t)$ tem-se que

$$m(t) = \frac{d}{dt} M(t) \tag{2.45}$$

Desenvolvendo (2.44) tem-se pela definição de $v_G(\cdot)$ e por (2.45) que

$$a_G(t) = \frac{d}{dt} M(t)\ \frac{d}{dM(t)} V_G(M(t)),$$



UNICAMP BIBLIOTECA CENTRAL

e pela aplicação da regra da cadeia de derivadas de funções compostas

$$a_G(t) = \frac{d}{dt} V_G(M(t)) \ ,$$

e portanto tem-se que

$$A_G(t) = V_G(M(t)) \ . \tag{2.46}$$

Pode ser mostrado que se $T$ é o tempo de ocorrência do evento com risco condicionado igual a $\lambda(t|z) = a_G(t)\, z$ então

$$c_G(A_G(t)) = E_G(Z|T \geq t) \ , \tag{2.47}$$

em que $E_G(Z|T \geq t) = \int_0^t z dG(z|T \geq t)$ e $Z$ representaria a fragilidade multiplicativa (ver Proposição 4(a) do Anexo A).

A demonstração de Hoem prova que $m(t)$ é o risco não-condicionado de $T$.

De (2.46) e de (2.47) tem-se que

$$E_G(Z|T \geq t) = c_G[V_G(M(t))] \ . \tag{2.48}$$

Segundo Hoem (1990), tem-se por construção de $a_G(t)$ em (2.44) e pelos resultados da Seção 2.1 que $\lambda(t)$, o risco não-condicionado de $T$, é dado por

$$\lambda(t) = a_G(t)\, E_G(Z|T \geq t) \ . \tag{2.49}$$

Como mostrado na Proposição 4(b) do Anexo A, a função $a_G(t)$ em (2.44) pode ser reescrita como

$$a_G(t) = \frac{m(t)}{c_G[V_G(M(t))]} \ . \tag{2.50}$$

Para finalizar a prova basta mostrar que $\lambda(t)$ é igual a $m(t)$, substituindo $E_G(Z|T \geq t)$ e $a_G(t)$ em (2.49) pelas expressões em (2.48) e (2.50), respectivamente.

Portanto, qualquer função de risco $m(t)$ num modelo sem heterogeneidade não-

observada pode ser gerada, alternativamente, por um modelo de fragilidade multiplicativo em que a distribuição de fragilidade, escolhida arbitrariamente, tem média finita.

Suponha-se agora que o modelo de partida inclua somente covariáveis e que a função de risco é dada por

$$m\left(t|\mathbf{x}\right) = m_0\left(t\right) U(\boldsymbol{\beta}, \mathbf{x}) \ , \tag{2.51}$$

para algum risco padrão $m_0\left(t\right)$ e uma função regressora $U(\boldsymbol{\beta}, \mathbf{x})$, em que $\mathbf{x}$ é o vetor dos valores das covariáveis e $\boldsymbol{\beta}$ é o correspondente vetor dos coeficientes de regressão. Então, como em (2.44), tem-se por construção que

$$a_G\left(t\right) = m\left(t|\mathbf{x}\right) v_G(M\left(t|\mathbf{x}\right)) = m_0\left(t\right) U(\boldsymbol{\beta}, \mathbf{x})\, v_G\left[M_0\left(t\right) U(\boldsymbol{\beta}, \mathbf{x})\right] \ ,$$

em que $M\left(t|\mathbf{x}\right) = \int_0^t m\left(u|\mathbf{x}\right) du$ e $M_0\left(t\right) = \int_0^t m_0\left(u\right) du$. Portanto, o problema de identificabilidade surge novamente com os modelos de fragilidade incluindo covariáveis, ou seja, não se pode distinguir entre riscos populacionais resultantes de modelos de fragilidade multiplicativos com covariáveis e modelos de riscos proporcionais sem heterogeneidade não-observada, porque as conseqüências observáveis são idênticas. Note-se que $a_G\left(t\right)$ não precisa ser fatorável em duas funções, uma que dependa somente de $t$ e outra que dependa somente de $\mathbf{x}$, mesmo que $m\left(t\right)$ seja decomposto desta maneira (Hoem, 1990).

## 2.9 A Falta de Identificabilidade com Modelos de Fragilidade

Seja um modelo de fragilidade multiplicativo com risco condicionado

$$\lambda_H\left(t|z\right) = z\lambda_{0H}\left(t\right) \ ,$$

sujeito a uma fragilidade $Z$ com função de distribuição misturadora $H(z)$ de média finita. O correspondente risco não-condicionado é dado por

$$\lambda_H(t) = \lambda_{0H}(t) E_H(Z|T \geq t) \ . \tag{2.52}$$

Define-se

$$\Lambda_H(t) = \int_0^t \lambda_H(u)\, du$$

e escolhe-se alguma distribuição arbitrária $G(z)$ para $Z$, com média finita e utiliza-se uma construção similar a (2.44) para definir

$$\lambda_{0G}(t) = \lambda_H(t)\, v_G(\Lambda_H(t)) \ .$$

Então, para qualquer $G$, $(\lambda_{0G}, G)$ tem o mesmo risco não-condicionado de $(\lambda_{0H}, H)$, digamos $\lambda_H(t)$. Os dois modelos são observacionalmente indistinguíveis.

Elbers e Ridder (1982), citados por Hoem (1990), demonstram que o efeito do tempo e das covariáveis sobre o risco e a distribuição de fragilidade são identificáveis a partir do risco não-condicionado de um modelo multiplicativo de riscos proporcionais com fragilidade. O ponto de partida para o resultado de identificabilidade de Elbers-Ridder é um modelo de fragilidade multiplicativo com risco não-condicionado dado por

$$\lambda_H(t|\mathbf{x}) = \lambda_{0H}(t|\mathbf{x})\, E_H(Z|T \geq t, \mathbf{x})$$

e com risco condicionado dado por

$$\lambda_H(t|\mathbf{x}, z) = z\lambda_{0H}(t|\mathbf{x})$$

em que o risco padrão $\lambda_{0H}(t|\mathbf{x})$ é decomposto em

$$\lambda_{0H}(t|\mathbf{x}) = \lambda_0(t)\, U(\boldsymbol{\beta}, \mathbf{x}) \tag{2.53}$$

uma decomposição similar a (2.51), exceto que a suposição de fatorabilidade é feita agora em $\lambda_{0H}(t|\mathbf{x})$, em vez de $m(t|\mathbf{x})$. No resultado de Elbers-Ridder assume-se que $\mathbf{x}$ toma valores num subconjunto aberto de um espaço euclidiano e que $U(\boldsymbol{\beta}, \cdot)$ é não-constante sobre este conjunto. Também é necessário que $\lambda_{0G}(t)$ seja fatorável de maneira similar. Estas condições são suficientes para excluir todas as especificações $(\lambda_{0G}, G)$ que não coincidam com $(\lambda_{0H}, H)$ tornando $\lambda_0(t)$, $U(\boldsymbol{\beta}, \mathbf{x})$ e $H(z)$ identificáveis [3].

Assim, o teorema de identificabilidade de Elbers-Ridder não contradiz o resultado geral de não-identificabilidade, pois exclue não-identificabilidade pela imposição de restrições suficientemente fortes sobre heterogeneidades permissíveis equivalentes.

Tais restrições devem vir de teoria subjacente. A sustentabilidade dessas restrições não pode ser decidida baseando-se somente em conhecimentos acerca do risco não-condicionado $\lambda_H(t)$. Mesmo se $\lambda_{0H}(t)$ for fatorável como em (2.53), $\lambda_{0G}(t)$ não precisa ter, e normalmente não satisfará, uma decomposição similar. Se $\lambda_H(t)$ é toda informação que se tem para fazer inferências, a fatorabilidade de $\lambda_{0H}(t)$ não é uma característica do modelo identificável.

Uma observação adicional com respeito à identificabilidade dos modelos de fragilidade é a necessidade de considerar a distribuição condicional de $Z$ dado $\mathbf{X} = \mathbf{x}$, em que se assume que $Z$ e o vetor de covariáveis $\mathbf{X}$ são variáveis aleatórias dependentes. A importância do problema foi apontada por Montgomery, Richards e Braun (1986) no contexto de regressão logística com covariáveis não-observadas, mas nenhuma análise rigorosa de identificabilidade tem sido conduzida para classes específicas de distribuições condicionais de $Z$.

Oakes (1989) mostra que o problema de identificabilidade é radicalmente diferente para dados bivariados, em que os tempos contínuos de ocorrência de um evento são observados em pares de indivíduos. A demonstração de Oakes é apresentada no próximo capítulo.

[3] Elbers e Ridder (1982) citados por Hoem (1990) também assumem que $E(Z) < \infty$, como tem-se estabelecido. Heckman e Singer (1984), citados por Hoem (1990), têm modificado esta suposição, mas mantêm a suposição de fatorabilidade para $\lambda_{0G}(t)$.

# Capítulo 3

# Modelos de Risco de Tempos Multivariados

Os dados multivariados na história de eventos podem consistir de tempos relacionados de ocorrências de um evento que se originam de indivíduos agrupados, tal como a história de fecundidade de irmãs, a experiência escolar de irmãos ou a sobrevivência de casais. Em geral, correspondem a grupos cujos membros têm experimentado algo em comum. Nestes casos há necessidade de modelar a dependência entre as observações de um grupo de indivíduos, pois usualmente os tempos de ocorrência do evento em diferentes indivíduos, no mesmo grupo, não serão independentes. A falta de independência pode ser devida a características observáveis. No entanto, é possível discutir se alguma dependência intra-grupos ainda permanece, mesmo depois de incorporar as covariáveis observadas. Note-se que qualquer tempo agrupado, ou até todos, podem estar censurados, tornando a estimação mais complexa. Diferentemente da estrutura de eventos repetidos esboçada na Seção 1.2.2 , não é possível adotar o princípio da condicionalidade seqüencial.

A existência de observações multivariadas por causa dos agrupamentos de indivíduos incorpora questões complexas na especificação dos modelos de risco, mas por outro lado, proporciona uma oportunidade de modelar os determinantes não-observados do risco de

uma maneira efetiva (Mare, 1994). Os modelos de risco de tempos pareados tornam possível a identificabilidade das misturas resultantes de modelos de fragilidade, quando os tempos estão associados entre si devido a uma dependência comum com a fragilidade que é compartilhada pelo par de indivíduos (Oakes, 1989).

## 3.1 Conceitos Básicos

Seja $T_{ij}$ $(i = 1, 2, \ldots, n)\,(j = 1, 2, \ldots, m)$ uma variável aleatória contínua e não-negativa que represente o tempo de ocorrência do evento no indivíduo $i$ do grupo $j$. Suponha-se primeiramente que $n = 2$, e que os modelos não admitem covariáveis. Como os resultados probabilísticos podem ser estudados para um único grupo o índice $j$ é omitido, daqui em diante, sempre que possível.

Relacionados com a distribuição conjunta de $T_1$ e $T_2$ estão:

a) a função de distribuição conjunta

$$F(t_1, t_2) = P(T_1 \leq t_1, T_2 \leq t_2) \ ;$$

b) para os pontos em que $F(t_1, t_2)$ for diferenciável, a densidade conjunta

$$f(t_1, t_2) = \lim_{\Delta t \to 0} \frac{P(t_1 \leq T_1 \leq t_1 + \Delta t, t_2 \leq T_2 \leq t_2 + \Delta t)}{\Delta t^2} = \frac{\partial^2}{\partial t_1 \partial t_2} F(t_1, t_2) \ ;$$

c) a função de sobrevivência conjunta

$$S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = \int_{t_1}^{\infty} \int_{t_2}^{\infty} f(u, v)\, du dv \ ,$$

com funções de sobrevivência marginais dadas por

$$S_1(t_1) = P(T_1 > t_1) = S(t_1, 0)$$

e

$$S_2(t_2) = P(T_2 > t_2) = S(0, t_2) \ ,$$

e se $T_1$ e $T_2$ forem independentes, tem-se que

$$S(t_1, t_2) = S_1(t_1) S_2(t_2) \ ;$$

d) a função de risco conjunta

$$\lambda(t_1, t_2) = \lim_{\Delta t \to 0} \frac{P(t_1 \leq T_1 \leq t_1 + \Delta t, t_2 \leq T_2 \leq t_2 + \Delta t \mid T_1 \geq t_1, T_2 \geq t_2)}{\Delta t^2} \ ,$$

com função de risco marginal, para $i = 1, 2$, dada por

$$\lambda_i(t_i) = \lim_{\Delta t \to 0} \frac{P(t_i \leq T_i \leq t_i + \Delta t | T_i \geq t_i)}{\Delta t} \ ,$$

e pode-se também definir funções de risco condicionais de $T_1$ dado $T_2$ como

$$\lambda(t_1 | T_2 = t_2) = \lim_{\Delta t \to 0} \frac{P(t_1 \leq T_1 \leq t_1 + \Delta t | T_1 \geq t_1, T_2 = t_2)}{\Delta t}$$

e

$$\lambda(t_1 | T_2 \geq t_2) = \lim_{\Delta t \to 0} \frac{P(t_1 \leq T_1 \leq t_1 + \Delta t | T_1 \geq t_1, T_2 \geq t_2)}{\Delta t} \ ,$$

que denotam o risco para o indivíduo 1 dado que o outro não sobreviveu a $t_2$ e dado que o outro sobreviveu a $t_2$, respectivamente.

Crowder et al. (1990) apresentam várias distribuições para tempos pareados como, por exemplo, as distribuições exponenciais bivariadas para $T_1$ e $T_2$, sugeridas por Gumbel (1960) citado por Crowder et al. (1990).

## 3.2 Um Modelo de Fragilidade Compartilhada

Uma maneira de modelar uma função de sobrevivência conjunta é assumir a existência de uma fragilidade $Z$ comum ao par, tal que dado $Z = z$, $T_1$ e $T_2$ sejam independentes. Assume-se então que

$$S(t_1, t_2|z) = S_1(t_1|z)\, S_2(t_2|z) \ , \tag{3.1}$$

em que as funções de sobrevivência em (3.1) são condicionadas a $Z = z$. Conforme o contexto, $Z$ pode representar uma quantidade não-observada e específica do par, algo que seja comum aos dois indivíduos pareados, e que possa explicar a falta de independência entre os tempos pareados de ocorrência do evento. Portanto, a *fragilidade compartilhada* $Z$ refere-se a pares de indivíduos e não a indivíduos isolados.

Nos modelos de fragilidade em (3.1), supõe-se que a fragilidade compartilhada atua multiplicativamente sobre a função de risco marginal condicionada, tanto que

$$\lambda_i(t_i|z) = \lambda_{0i}(t_i)\, z \ ,$$

em que $\lambda_{0i}(t_i)$ é o risco padrão do indivíduo $i\,(i = 1, 2)$. E, como no caso univariado, estão associados a $\lambda_i(t_i|z)$

a) o risco marginal condicionado acumulado

$$\Lambda_i(t_i|z) = \int_0^{t_i} \lambda_i(u|z)\, du = z\Lambda_{0i}(t_i) \ ,$$

com $\Lambda_{0i}(t_i) = \int_0^{t_i} \lambda_{0i}(u)\, du$;

b) a função de sobrevivência marginal condicionada

$$S_i(t_i|z) = \exp\left[-\Lambda_i(t_i|z)\right] = \exp\left[-z\Lambda_{0i}(t_i)\right] = S_{0i}(t_i)^z \ , \tag{3.2}$$

com $S_{0i}(t_i) = \exp\left[-\Lambda_{0i}(t_i)\right]$.

Sob o modelo de fragilidade compartilhada, a função de sobrevivência conjunta condicional de $T_1$ e $T_2$ dado $Z = z$, é então, de (3.1) e (3.2), dada por

$$S(t_1, t_2|z) = \exp\{-z[\Lambda_{01}(t_1) + \Lambda_{02}(t_2)]\} \ .$$

Se $Z$ tem função de distribuição $G(z)$ não-negativa, então a função de sobrevivência conjunta não-condicionada é dada, pela teoria de misturas de distribuições, por

$$\begin{aligned} S(t_1, t_2) &= \int_0^\infty S(t_1, t_2|z)\, dG(z) = \int_0^\infty \exp\{-z[\Lambda_{01}(t_1) + \Lambda_{02}(t_2)]\}\, dG(z) \\ &= \mathcal{L}(\Lambda_{01}(t_1) + \Lambda_{02}(t_2)) \ , \end{aligned} \tag{3.3}$$

em que a última expressão é a transformada de Laplace para $G(z)$ calculada no ponto $s = \Lambda_{01}(t_1) + \Lambda_{02}(t_2)$. Assim, para a estimação efetiva deste modelo é necessária alguma suposição acerca da forma de $\lambda_{0i}(t_i)$, $(i = 1, 2)$ e da distribuição de $Z$.

### 3.2.1 Identificabilidade com Modelos de Fragilidade Compartilhada

Oakes (1989) tem mostrado que as misturas para dados pareados com fragilidade compartilhada são identificáveis. A prova de Oakes faz inicialmente uma referência às distribuições arquimedianas que têm a seguinte forma geral para a função de sobrevivência bivariada

$$S(t_1, t_2) = p[q(S_1(t_1)) + q(S_2(t_2))] \ , \tag{3.4}$$

em que $p(s)$ é qualquer função decrescente não-negativa com $p(0) = 1$ e com segunda derivada não-negativa, e $q(r)$ é a função inversa de $p(s)$.

Pode ser visto facilmente que as distribuições bivariadas geradas por modelos de fragilidade compartilhada são uma subclasse das distribuições arquimedianas. Quando

$p(s) = \mathcal{L}(s)$, (3.3) é equivalente a (3.4) desde que para $i = 1, 2$

$$p(\Lambda_{0i}(t_i)) = \int_0^\infty \exp\left[-z\Lambda_{0i}(t_i)\right] dG(z) = \int_0^\infty S_i(t_i|z)\, dG(z) = S_i(t_i)$$

e por inversão, implica que

$$q(S_i(t_i)) = q[p(\Lambda_{0i}(t_i))] = \Lambda_{0i}(t_i) \ .$$

A demonstração de Oakes está baseada numa função introduzida por Clayton (1978) dada por

$$R(t_1, t_2) = \frac{\lambda(t_1|T_2 = t_2)}{\lambda(t_1|T_2 \geq t_2)} \ , \tag{3.5}$$

que é uma razão entre dois riscos condicionais de $T_1$ dado $T_2$, sendo uma razão do tipo produto cruzado.

Denote-se $\mathbf{t} = (t_1, t_2)$. Para distribuições arquimedianas, a prova de Oakes mostra primeiramente que $R(\mathbf{t})$ depende de $\mathbf{t}$ somente por meio de $S(\mathbf{t})$, por uma função $R^*(r)$ com $r = S(\mathbf{t})$, isto é

$$R(\mathbf{t}) = R^*(S(\mathbf{t})) \ . \tag{3.6}$$

A segunda parte da prova de Oakes demonstra que a recíproca desse primeiro resultado também é verdadeira, ou seja, se a razão $R(\mathbf{t})$ pode ser reescrita como $R^*(S(\mathbf{t}))$, então o vetor aleatório $(T_1, T_2)$ tem distribuição arquimediana. Note-se que nem toda distribuição arquimediana leva a um modelo de fragilidade compartilhada. Oakes (1989) mostra que a função $R^*(r)$ determina $p(s)$ unicamente, exceto por um parâmetro de escala.

Uma implicação prática dos resultados da prova de Oakes é que nos modelos de fragilidade compartilhada as misturas são identificadas e podem ser testados porque têm conseqüências observáveis, ou seja, a transformada de Laplace, denotada por $p(s)$ em Oakes (1989), pode ser determinada em termos de $R(\mathbf{t})$, que é uma razão de riscos observáveis.

**Prova de Oakes** Uma maneira de verificar (3.6) é expressar (3.5) como função somente de $q(r)$ já que $r = S(t)$. Pode-se mostrar que

$$R(\mathbf{t}) = \frac{[D_1 D_2 S(\mathbf{t})]\ S(\mathbf{t})}{[D_1 S(\mathbf{t})]\,[D_2 S(\mathbf{t})]}\ , \tag{3.7}$$

em que $D_i = -\partial/\partial t_i$ $(i = 1, 2)$ (ver Proposição 5 do Anexo A). Note-se que por (3.7), $R(\mathbf{t})$ é uma função de $\mathbf{t}$ mas não é, claramente, uma função composta de $S(\mathbf{t})$.

Para reescrever $R(\mathbf{t})$ em função de $q(r)$, primeiramente obtém-se $R(\mathbf{t})$ em função de $p(s)$. Fazendo $s = q(S_1(t_1)) + q(S_2(t_2))$, tem-se de (3.4) que

$$S(\mathbf{t}) = p(s) \tag{3.8}$$

e pela regra da cadeia de derivadas de funções compostas, tem-se que a derivada de (3.8) em relação a $t_i$ é dada por

$$\frac{\partial}{\partial t_i} S(\mathbf{t}) = \frac{\partial}{\partial t_i} p(s) = p'(s)\ q'(S_i(t_i))\ S_i'(t_i)$$

e portanto (3.7) resulta em

$$R(\mathbf{t}) = R^*(s) = \frac{p(s)\ p''(s)}{[p'(s)]^2}\ , \tag{3.9}$$

em que $s = q(S(\mathbf{t}))$, por inversão de (3.8).

Para obter a fórmula explícita de $R^*(r)$, deve-se lembrar que quando $r = S(\mathbf{t})$ e $s = q(S_1(t_1)) + q(S_2(t_2))$, tem-se de (3.8) que

$$r = p(s)\ . \tag{3.10}$$

A equação (3.10) implica, por inversão, que

$$q(r) = q(p(s)) = s\ .$$

É necessário também desenvolver a segunda derivada de $q(r)$ dada por

$$q''(r) = -\frac{p''(q(r))\, q'(r)}{[p'(q(r))]^2} = -\frac{p''(s)\, q'(r)}{[p'(s)]^2}$$

$$\Leftrightarrow \frac{q''(r)}{q'(r)} = -\frac{p''(s)}{[p'(s)]^2} \,. \tag{3.11}$$

De (3.10) e (3.11) tem-se que (3.9) resulta em

$$R(\mathbf{t}) = R^*(r) = -\frac{r q''(r)}{q'(r)} \,, \tag{3.12}$$

em que $r = S(\mathbf{t})$.

Para demonstrar a unicidade na determinação da função $p(s)$, a prova de Oakes utiliza a função $q(r)$. Isolando $q(r)$ em (3.12) tem-se, em termos de $R^*(r)$, uma fórmula implícita apresentada por Oakes (1989) como

$$q_\kappa(r) = \int_r^1 \exp\left[\int_u^{1-\kappa} \frac{R^*(v)}{v} dv\right] du \,. \tag{3.13}$$

A função $p(s)$ é determinada unicamente exceto por uma mudança de escala em $s$ e sua função inversa é determinada em termos de $R^*(v)$ por (3.13), a menos de uma constante multiplicativa especificada por $\kappa > 0$. A constante $\kappa$ apresenta-se simplesmente como um irrelevante fator de escala dentro da função $p(u)$.

Para serem feitos maiores progressos com relação às propriedades dos modelos de fragilidade compartilhada, é necessário especificar a distribuição de $Z$.

### 3.2.2 Distribuições de Fragilidade Compartilhada

#### Fragilidade Gama

Suponha que a fragilidade compartilhada $Z$ tenha uma distribuição gama com parâmetros $\eta$ e $\nu$. Então a transformada de Laplace para a distribuição de $Z$ no ponto

$s$, é dada por

$$\mathcal{L}(s) = \left[\frac{\nu}{\nu + s}\right]^{\eta} . \tag{3.14}$$

Se $\eta = \nu = 1/\sigma^2$ ([1]) então

$$\mathcal{L}(s) = \left[1 + \sigma^2 s\right]^{-\frac{1}{\sigma^2}} .$$

E por (3.3) tem-se que

$$S(\mathbf{t}) = \left[1 + \sigma^2 \left[\Lambda_{01}(t_1) + \Lambda_{02}(t_2)\right]\right]^{-\frac{1}{\sigma^2}} .$$

Seguindo a notação de Oakes tem-se que

$$\mathcal{L}(s) = p(s) = \left[1 + \sigma^2 s\right]^{-\frac{1}{\sigma^2}} ;$$

então, por inversão,

$$q(r) = \frac{1}{\sigma^2}\left(r^{-\sigma^2} - 1\right) .$$

As duas primeiras derivadas de $q(r)$ são, respectivamente

$$q'(r) = -r^{-\left(1+\sigma^2\right)}$$

e

$$q''(r) = \left(1 + \sigma^2\right) r^{-\left(2+\sigma^2\right)} .$$

Substituindo-as em (3.12) chega-se a

$$R^*(r) = 1 + \sigma^2 , \tag{3.15}$$

[1] Esta restrição sobre os parâmetros da distribuição gama implica que $E(Z) = 1$ e $Var(Z) = \sigma^2$.

resultando finalmente em

$$R(\mathbf{t}) = 1 + \sigma^2 \ .$$

O resultado em (3.15) proporciona uma nova interpretação para $\sigma^2$ ([2]). Uma variância igual a $\sigma^2$ significa que se o indivíduo 2 do par não sobreviveu a $t$, o risco para o indivíduo 1 deveria ser $(100\sigma^2)\,\%$ maior que o risco para o mesmo indivíduo 1 se o outro membro do par sobrevivesse a $t$.

No entanto, a principal conseqüência de (3.15) é que o modelo de fragilidade compartilhada gama produz uma mistura identificável porque implica que $R(\mathbf{t})$, uma razão de dois riscos estimáveis, é constante no tempo.

E finalmente, de (3.4) tem-se que a função de sobrevivência conjunta não-condicionada, em função das marginais, é dada por

$$S(\mathbf{t}) = \left[S_1(t_1)^{-\sigma^2} + S_2(t_2)^{-\sigma^2} - 1\right]^{-\frac{1}{\sigma^2}}$$

e portanto, dadas as funções de sobrevivência marginais não-condicionadas e a variância da fragilidade gama, obtém-se a distribuição arquimediana para o vetor aleatório $(T_1, T_2)$.

### Fragilidade Gaussiana Inversa

Se $Z$ tem distribuição gaussiana inversa com parâmetros $\eta$ e $\nu$, então a transformada de Laplace para $Z$ no ponto $s$ é dada por

$$\mathcal{L}(s) = \exp\left[-\sqrt{4\eta\nu} - \sqrt{4\eta(\nu + s)}\right] \ ; \tag{3.16}$$

[2] Inicialmente, $\sigma^2$ foi interpretado como o valor da variância da fragilidade compartilhada.

e para $\eta = \nu = \dfrac{1}{2\sigma^2}$ ([3]), a equação (3.16) torna-se

$$\mathcal{L}(s) = \exp\left[-\frac{1}{\sigma^2}\left(\sqrt{1+2\sigma^2 s}-1\right)\right] .$$

Então, de (3.3), a função de sobrevivência não-condicionada é

$$S(\mathbf{t}) = \exp\left\{-\frac{1}{\sigma^2}\left(\sqrt{1+2\sigma^2\left[\Lambda_{01}(t_1)+\Lambda_{02}(t_2)\right]}-1\right)\right\} .$$

Seguindo a notação de Oakes tem-se que

$$\mathcal{L}(s) = p(s) = \exp\left[-\frac{1}{\sigma^2}\left(\sqrt{1+2\sigma^2 s}-1\right)\right] ;$$

então, por inversão,

$$q(r) = \frac{\sigma^2}{2}\log r\left(\log r - \frac{2}{\sigma^2}\right) .$$

As duas primeiras derivadas de $q(r)$ são, respectivamente

$$q'(r) = \frac{\sigma^2}{r}\left[\log r - \frac{1}{\sigma^2}\right]$$

e

$$q''(r) = \left(\frac{\sigma}{r}\right)^2\left(1-\log r+\frac{1}{\sigma^2}\right) .$$

Substituindo-as em (3.12) tem-se que

$$R^*(r) = 1+\left(\frac{1}{\sigma^2}-\log r\right)^{-1} ,$$

resultando finalmente em

$$R(\mathbf{t}) = 1+\left(\frac{1}{\sigma^2}-\log S(\mathbf{t})\right)^{-1} . \tag{3.17}$$

[3]Esta restrição sobre os parâmetros da distribuição gaussina inversa implica que $E(Z) = 1$ e $Var(Z) = \sigma^2$.

Se $\sigma^2 \to 0$, em (3.17), então $R(\mathbf{t}) \to 1$, indicando que $T_1$ e $T_2$ são independentes (ver item (b) do modelo de Clayton).

### Modelo de Clayton

Clayton (1978) propôs um modelo para tempos pareados contínuos em que os dois riscos condicionais de $T_1$ dado $T_2$ são proporcionais, ou seja

$$R(\mathbf{t}) = \frac{\lambda(t_1|T_2 = t_2)}{\lambda(t_1|T_2 \geq t_2)} = c \ . \tag{3.18}$$

Em outras palavras, o risco de um indivíduo em qualquer tempo $t_1$ dado que o outro não sobreviveu a algum tempo $t_2$, é $c$ vezes o risco no mesmo tempo $t_1$ dado que o outro sobreviveu a $t_2$. Qualquer que seja $t_1, t_2 \geq 0$, note-se que (3.18) resulta na mesma razão de risco.

Usando a notação de Oakes, tem-se que

$$R^*(t) = c \ ;$$

ao ser substituído em (3.13) tem-se que

$$\begin{aligned} q_\kappa(r) &= \int_r^1 \exp\left(\int_u^{1-\kappa} \frac{c}{v} dv\right) du = \int_r^1 \exp\left(c \int_u^{1-k} \frac{dv}{v}\right) du \\ &= \int_r^1 \exp\left\{c\left[\log(1-k) - \log u\right]\right\} du \\ &= (1-k)^c \int_r^{1-k} -u^c du \ . \end{aligned}$$

Desprezando o termo constante $k$, resulta que:

a) para $c > 1$

$$q(r) = \left(\frac{1}{r}\right)^{c-1} - 1$$

e, por inversão,

$$p(s) = \left(\frac{1}{1+s}\right)^{\frac{1}{c-1}} ,$$

que é a transformada de Laplace em (3.14) quando $\eta = \frac{1}{c-1}$ e $\nu = 1$;

b) para $c = 1$

$$q(r) = -\log r$$

e, por inversão,

$$p(s) = \exp\{-s\} ,$$

que é a transformada de Laplace para distribuição degenerada no ponto 1. Por (3.4), a função de sobrevivência conjunta não-condicionada, é dada por

$$S(\mathbf{t}) = S_1(t_1) S_2(t_2)$$

e, portanto, $c = 1$ implica que $T_1$ e $T_2$ são independentes.

Os resultados de Oakes mostram que o modelo de Clayton é equivalente ao modelo de fragilidade compartilhada, em que a fragilidade tem distribuição gama com parâmetros $\eta = \frac{1}{c-1}$ e $\nu = 1$, caso $c > 1$; e também é equivalente quando a fragilidade tem distribuição degenerada no ponto 1, no caso em que $c = 1$. Segundo Oakes (1989), quando $0 < c < 1$, o modelo de Clayton resulta em uma distribuição que determina uma associação negativa entre $T_1$ e $T_2$.

Oakes (1989) propõe outras distribuições possíveis para a fragilidade compartilhada.

## 3.3 Extensões Multivariadas

Os conceitos básicos da Seção 3.1 podem ser estendidos facilmente para mais do que dois tempos pareados de ocorrência do evento (ver Crowder et al., 1990). Em geral, em estudos com tempos agrupados, é interessante considerar que a função de sobrevivência

dada uma fragilidade compartilhada é o produto das marginais condicionadas também à fragilidade compartilhada.

Seja um conjunto de dados agrupados em que, para o indivíduo $i$ em um determinado grupo $j$, $t_{ij}$ é o tempo observado, $\delta_{ij}$ é o indicador de censura e $\mathbf{x}_{ij}$ é o vetor de valores de $p$ covariáveis $(i = 1, 2, \ldots, n)\,(j = 1, 2, \ldots, m)$.

Sejam $\mathbf{t}_j = (t_{1j}, t_{2j}, \ldots, t_{nj})$ e $\mathbf{x}_j = (\mathbf{x}_{1j}, \mathbf{x}_{2j}, \ldots, \mathbf{x}_{nj})$ o vetor dos tempos de ocorrências do evento e a matriz de covariáveis do grupo $j$, respectivamente. A função de sobrevivência condicionada para um determinado grupo (neste caso, omite-se o índice $j$) é definida como

$$S(\mathbf{t}|\mathbf{x}, z) = P(T_1 \geq t_1, T_2 \geq t_2, \ldots, T_n \geq t_n|\mathbf{x}, z) \ ,$$

em que as covariáveis do grupo são independentes da fragilidade compartilhada pelo grupo. Assumindo que os $n$ tempos de ocorrência do grupo, $T_1, T_2, \ldots, T_n$ sejam independentes dado $\mathbf{x}_i$ e $z$, tem-se que

$$S(\mathbf{t}|\mathbf{x}, z) = \prod_{i=1}^{n} S_i(t_i|\mathbf{x}_i, z) \ .$$

Pelo modelo de fragilidade compartilhada, as marginais condicionadas do risco satisfazem

$$\lambda_i(t_i|\mathbf{x}_i, z) = \lambda_{0i}(t_i|\mathbf{x}_i)\, z \ , \qquad i = 1, 2, \ldots, n$$

e os efeitos das covariáveis no risco padrão $\lambda_{0i}(t_i|\mathbf{x}_i)$ podem, por vezes, seguir um modelo de riscos proporcionais

$$\lambda_{0i}(t_i|\mathbf{x}_i) = \lambda^*_{0i}(t_i) \exp(\boldsymbol{\beta}'\mathbf{x}_i) \ ,$$

em que $\lambda^*_{0i}(t_i)$ é um novo risco padrão que depende somente de $t_i$ $(i = 1, 2, \ldots, n)$. Adicionalmente, assume-se que $Z$ tenha uma função de distribuição $G(z)$, e então a função

de sobrevivência não-condicionada é, pela teoria de mistura de distribuições, dada por

$$S(\mathbf{t}|\mathbf{x}) = \int_0^\infty \prod_{i=1}^{n} S_i(t_i|\mathbf{x}_i, z)\, dG(z)$$

e pela relação entre função de sobrevivência e de risco

$$\begin{aligned} S(\mathbf{t}|\mathbf{x}) &= \int_0^\infty \prod_{i=1}^{n} \exp\left[-z\Lambda_{0i}(t_i|\mathbf{x}_i)\right] dG(z) \\ &= \int_0^\infty \exp\left[-z\sum_{i=1}^{n}\Lambda_{0i}(t_i|\mathbf{x}_i)\right] dG(z) \ , \end{aligned} \tag{3.19}$$

em que $\Lambda_{0i}(t_i|\mathbf{x}_i) = \int_0^{t_i} \lambda_{0i}(u|\mathbf{x}_i)\, du$, $(1, 2, \ldots, n)$.

A equação em (3.19) pode ser encontrada usando-se a transformada de Laplace para a distribuição de $Z$ no ponto

$$s = \sum_{i=1}^{n} \Lambda_{0i}(t_i|\mathbf{x}_i) \ .$$

Por exemplo, para a fragilidade gama com parâmetros iguais a $1/\sigma^2$, tem-se que a função de sobrevivência conjunta não-condicionada é dada por

$$S(\mathbf{t}|\mathbf{x}) = \left[1 + \sigma^2 \sum_{i=1}^{n} \Lambda_{0i}(t_i|\mathbf{x}_i)\right]^{-\frac{1}{\sigma^2}} \ .$$

Oakes (1989) comenta que suas considerações sobre identificabilidade com modelos de fragilidade compartilhada podem ser estendidas para problemas envolvendo covariáveis e distribuições multivariadas. Guo (1993) generaliza a caracterização de Clayton para o caso multivariado quando for assumida uma distribuição gama para a fragilidade compartilhada e Hoougard (1986) propõe uma distribuição estável positiva para a fragilidade compartilhada. Porém, as generalizações de Guo e de Hoougard não consideram a porção da fragilidade compartilhada que depende das covariáveis observadas. Para estabelecer resultados mais precisos, a distribuição condicionada $G(z|\mathbf{x})$ deveria substituir $G(z)$, quando a independência entre a fragilidade compartilhada e as covariáveis não pode ser satisfeita.

Quando os tempos $T_1$ e $T_2$ são discretos, a conexão exata com as distribuições arquimedianas se perde. Mesmo considerando uma versão discreta para $R(\mathbf{t})$, não é possível estabelecer uma correspondência entre os modelos de fragilidade compartilhada e riscos pareados discretos. Portanto, a prova de Oakes não pode ser generalizada para o caso discreto. No entanto, Mare (1994) propôs uma abordagem baseada num modelo logito para indivíduos pareados que é adaptado para a estrutura particular de dados categóricos da história de eventos, permitindo o controle da heterogeneidade não-observada comum ao par de observações, e que é uma generalização direta dos modelos log-lineares.

# Capítulo 4

# Um Modelo Logito para Dados Categóricos da História de Eventos

Em certos estudos da história de eventos, o tempo de ocorrência do evento é medido por uma progressão discreta, em que em cada estágio os indivíduos estão sujeitos ao risco de desistir, como a trajetória de uma pessoa em sucessivos estágios escolares, que pode ser descrita, por exemplo, em 4 etapas: nenhum grau de instrução completo, 1º grau completo, 2º grau incompleto, pelo menos o 2º grau completo. Neste caso, a escolaridade é categorizada em 4 níveis sucessivos: menos de 8, exatamente 8, de 9 a 10, e 11 ou mais anos completos de estudo. A passagem de um nível para outro define-se por *transição*. Então, se um determinado indivíduo completou 9 anos de estudo, significa que ele passou pela 1ª e 2ª transições, mas não passou pela 3ª transição (concluiu o 1º grau e ingressou no 2º grau, porém não concluiu o 2º grau). Os dados de mortalidade fetal e infantil em que a progressão dos indivíduos depende das transições feitas nos diferentes estágios sucessivos de sobrevivência desde a concepção, passando pelo período perinatal, até o período pós-neonatal, são também exemplos de tempos categóricos de ocorrência do evento. Mare (1994) inclue outros exemplos como a trajetória de mulheres em sucessivos partos, a mobilidade de pessoas em organizações hierarquicas como o serviço militar, e o progresso de casos criminais desde a prisão, passando por acusação, julgamento, condenação, até a

sentença (Mare, 1994).

Suponha que exista interesse em estimar o efeito das covariáveis sobre a probabilidade do indivíduo não fazer uma determinada transição, dado que completou todas as transições anteriores. Uma variável *dummy* $\mathbf{X}_{ti}$ é incorporada ao modelo logito para representar a covariável, ou seja, $\mathbf{X}_{ti}$ é um vetor $((p-1)\times 1)$ de variáveis dicotômicas $\mathbf{X}_{tik}$, que assumem o valor 1 se o indivíduo $i$ não passa pela transição t e pertence à categoria $k+1$ e 0, caso contrário $(k=1,2,\ldots,p-1)\,(\mathbf{t}=1,2,\ldots,N)\,(i=1,2,\ldots,n)$. Segundo o exemplo já citado das transições escolares, se um determinado indivíduo $i$ completou 9 anos de estudo então $Y_{1i}=Y_{2i}=0$ e $Y_{3i}=1$. Uma sucessão de transições, representada por $Y_{1i}$, $Y_{2i}$, ..., $Y_{Ni}$, permite a adoção do princípio de condicionalidade seqüencial para eventos repetidos (ver Seção 1.2.2). Para estimar os efeitos das covariáveis pode-se aplicar a condicionalidade seqüencial em um modelo logito como

$$l\left(Y_{ti}=1|Y_{1i}=Y_{2i}=\ldots=Y_{\mathrm{t}-1,i}=0,\mathbf{X}_{ti}=\mathbf{x}_{ti}\right)=\beta_{0\mathrm{t}}+\boldsymbol{\beta}_{\mathrm{t}}'\mathbf{x}_{ti}\ , \tag{4.1}$$

em que $l(\cdot)$ é a transformação logito (ver Seção 1.2.2), $\boldsymbol{\beta}_{\mathrm{t}}$ é o vetor $(p\times 1)$ de parâmetros desconhecidos representando o efeito das categorias da covariável sobre o logito de não fazer a transição t dado que as transições anteriores foram feitas e $\beta_{0\mathrm{t}}$ é o intercepto (escalar) do modelo que denota o logito de não fazer a transição t dado que as transições anteriores foram feitas, quando $\mathbf{X}_{ti}$ for um vetor nulo. Admite-se que $Y_{0i}=0$. Sob (4.1), os vetores $\mathbf{X}_{ti}$ e $\boldsymbol{\beta}_{\mathrm{t}}$ podem depender de t.

Mare (1994) propôs métodos para estimar (4.1) com covariáveis categóricas. Este método de estimação alternativo para analisar tempos categóricos de ocorrência do evento de um único indivíduo pode ser generalizado diretamente para o caso de indivíduos pareados, produzindo estimativas equivalentes àquelas provenientes de métodos convencionais. Para observações pareadas de tempos categóricos de ocorrência, o modelo logito pode considerar a heterogeneidade não-observada quando esta abrange características compartilhadas pelo par de indivíduos, proporcionando uma identificabilidade das misturas provenientes de modelos com heterogeneidade não-observada sem fortes suposições

restritivas acerca dos parâmetros da distribuição da heterogeneidade não-observada.

Por definição, transições podem ser observadas somente até o indivíduo desistir. Isto é, se $Y_{ti} = 1$ então $Y_{t+1,i}, \ldots, Y_{Ni}$ não podem ser observados. Mesmo assim, é útil considerar a distribuição conjunta hipotética de $Y_{1i}, Y_{2i}, \ldots, Y_{Ni}$, incluindo os resultados que não são observados. No caso de indivíduos pareados, esta abordagem proporcionará um meio eficaz para levar em conta os fatores não-observados que influenciam os tempos pareados.

O método de estimação está baseado na transformação de uma distribuição do tempo de ocorrência em uma distribuição conjunta de transições. Como uma distribuição conjunta de transições é observada somente parcialmente, neste sentido, a abordagem adotada é um caso especial da estrutura de Haberman (1988) para análise de *tabelas de contingência observadas parcialmente*. Este tipo de abordagem é discutido a seguir, para indivíduos isolados, e posteriormente generalizado para dados pareados.

## 4.1 Relação entre Tabelas de Contingência Observadas e Parcialmente Observadas

O método de estimação focaliza a distribuição conjunta de $Y_{ti}$, o indicador para transição t no indivíduo $i$ $(\mathrm{t} = 1, 2)$ $(i = 1, 2 \ldots, n)$.

A probabilidade de falha, isto é, a probabilidade de "sofrer" o evento em uma determinada categoria de tempo, pode ser expressada em termos:

a) dos parâmetros do modelo em (4.1);

b) das probabilidades de fazer uma determinada transição, que provêm da parte observada da distribuição conjunta hipotética de $Y_{1i}$ e $Y_{2i}$ $(i = 1, 2, \ldots, n)$. Dadas estas probabilidades, pode-se escrever a função de verossimilhança do modelo logito e estimar os parâmetros em (4.1) pelo método **MV** (ver Seção 1.2.2).

Para formar a conexão entre a probabilidade de falha e a probabilidade de transição usando uma estrutura mínima de dados, pode-se restringir aos casos em que somente as duas primeiras transições são de interesse, isto é, $N = 2$ e as covariáveis estão fixadas no tempo, tanto que $\mathbf{X}_{1i} = \mathbf{X}_{2i} = \mathbf{X}_i$. Então (4.1) resulta em duas equações

$$l\left(Y_{1i} = 1 | \mathbf{X}_i = \mathbf{x}_i\right) = \beta_{01} + \boldsymbol{\beta}_1' \mathbf{x}_i \tag{4.2}$$

e

$$l\left(Y_{2i} = 1 | Y_{1i} = 0, \mathbf{X}_i = \mathbf{x}_i\right) = \beta_{02} + \boldsymbol{\beta}_2' \mathbf{x}_i \ . \tag{4.3}$$

Definam-se

$$A_t = \exp\left(\beta_{0t} + \boldsymbol{\beta}_t' \mathbf{x}\right) \qquad t = 1, 2 \tag{4.4}$$

e

$$D = \left(1 + A_1\right)\left(1 + A_2\right) \ , \tag{4.5}$$

em que $\beta_{0t}$ e $\boldsymbol{\beta}_t$ são os parâmetros em (4.1). O índice $i$ foi suprimido de (4.4) e (4.5) e esta prática será seguida adiante, sempre que possível.

Com a simplificação de (4.1) em (4.2) e (4.3) e as definições em (4.4) e (4.5), pode-se mostrar (ver Proposição 6 do Anexo A) que a probabilidade de um determinado indivíduo desistir em fazer a primeira transição é dada por

$$P\left(Y_1 = 1 | \mathbf{x}\right) = \frac{A_1}{1 + A_1} = \frac{A_1}{D} + \frac{A_1 A_2}{D} \tag{4.6}$$

e a probabilidade de desistir em fazer a segunda transição, dado que passou pela primeira é

$$P\left(Y_2 = 1 | Y_1 = 0, \mathbf{x}\right) = \frac{A_2}{1 + A_2} \ .$$

Pela definição de probabilidade condicional e de evento complementar, a probabilidade

de um indivíduo fazer a 1ª transição e não fazer a 2ª é

$$P(Y_1 = 0, Y_2 = 1|\mathbf{x}) = P(Y_1 = 0|\mathbf{x})\, P(Y_2 = 1|Y_1 = 0, \mathbf{x}) = \frac{A_2}{D} \,. \tag{4.7}$$

De forma análoga, a probabilidade do indivíduo fazer as duas transições é

$$P(Y_1 = 0, Y_2 = 0|\mathbf{x}) = P(Y_1 = 0|\mathbf{x})\, P(Y_2 = 0|Y_1 = 0, \mathbf{x}) = \frac{1}{D} \,. \tag{4.8}$$

As equações (4.6) a (4.8) formam a distribuição conjunta de probabilidades dos resultados possíveis de $Y_1$ e $Y_2$ e podem ser combinadas para formar a função de verossimilhança para o modelo em (4.1) dada por

$$\begin{aligned} L(\boldsymbol{\theta}) \;=\; & \prod_{i=1}^{n} \Big[ P(Y_{1i} = 1|\mathbf{x}_i)^{Y_{1i}}\, P(Y_{1i} = 0, Y_{2i} = 1|\mathbf{x}_i)^{(1-Y_{1i})Y_{2i}} \\ & P(Y_{1i} = 0, Y_{2i} = 0|\mathbf{x}_i)^{(1-Y_{1i})(1-Y_{2i})} \Big] \,, \end{aligned} \tag{4.9}$$

em que $\boldsymbol{\theta} = (\beta_{01}, \beta_{02}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Note-se que (4.9) pode ser escrito na forma dos parâmetros $\beta_{0t}$ e $\boldsymbol{\beta}_t$, usando as definições de $A_1$, $A_2$ e $D$ e que esta expressão de verossimilhança é equivalente à função de verossimilhança do modelo logito para riscos discretos (ver Allison, 1982).

As três probabilidades em (4.6), (4.7) e (4.8) correspondem aos três resultados possíveis de ocorrência do evento (ver tabela observada no Quadro 4.1).

E as equações (4.7) e (4.8) correspondem às partes observadas da distribuição conjunta das variáveis $Y_1$ e $Y_2$ como ilustra a tabela expandida no Quadro 4.1.

O valor de $Y_1$ é observado em todos os indivíduos, mas $Y_2$ é observado somente em indivíduos com $Y_1 = 0$. O Quadro 4.1 mostra o mapeamento das caselas da tabela observada dos resultados possíveis de ocorrência do evento para as caselas de uma tabela da distribuição conjunta de $Y_1$ e $Y_2$, observada parcialmente. São as probabilidades em (4.6) até (4.8) que constroem este mapeamento. Cada um dos dois termos do lado direito da equação (4.6) correspondem a um dos resultados $[Y_1 = 1, Y_2 = 1]$ ou $[Y_1 = 1, Y_2 = 0]$

QUADRO 4.1 - MAPEAMENTO DA TABELA EXPANDIDA DE TRANSIÇÕES PARA INDIVÍDUOS ISOLADOS

Tabela Observada da Distribuição de Freqüências dos Resultados Possíveis de $Y_1$ e $Y_2$, dado **X**

| Resultados | $Y_1 = 1$ | $Y_1 = 0, Y_2 = 1$ | $Y_1 = 0, Y_2 = 0$ |
|---|---|---|---|
| Freqüências | $a$ | $b$ | $c$ |

Tabela Expandida da Distribuição Conjunta de Freqüências de $Y_1$ e $Y_2$, dado **X**

| $Y_1$ | $Y_2$ | |
|---|---|---|
| | 1 | 0 |
| 1 | $a_1$ | $a_2$ |
| 0 | $b$ | $c$ |

Nota: $a_1 + a_2 = a$

Tabela Expandida da Distribuição Conjunta de Probabilidades de $Y_1$ e $Y_2$, dado **X**

| $Y_1$ | $Y_2$ | |
|---|---|---|
| | 1 | 0 |
| 1 | $\frac{A_1 A_2}{D}$ | $\frac{A_1}{D}$ |
| 0 | $\frac{A_2}{D}$ | $\frac{1}{D}$ |

que, por definição, não podem ser observados. Assim a casela $[Y_1 = 1]$ da distribuição de freqüências dos resultados possíveis é mapeada para duas caselas da distribuição conjunta de freqüências das transições. Poder-se-ia pensar então que a probabilidade de um indivíduo não fazer a 1ª transição e também não fazer a 2ª transição, se ele estivesse em risco na 2ª transição, seja dada por

$$P(Y_1 = 1, Y_2 = 1|\mathbf{x}) = \frac{A_1 A_2}{D}$$

e que a probabilidade de um indivíduo não fazer a 1ª transição e fazer a 2ª transição, se ele estivesse em risco na 2ª transição, seja dada por

$$P(Y_1 = 1, Y_2 = 0|\mathbf{x}) = \frac{A_1}{D} .$$

## 4.2 Estimação do Modelo Logito para Indivíduos Isolados

Os modelos propostos por Mare (1994) são uma variante dos *modelos de classes latentes* para dados categóricos (ver Seção 4.3.1). Haberman (1988) apresenta os modelos de classes latentes dentro do contexto geral da análise log-linear de tabelas de contingência observadas parcialmente. O modelo log-linear para a tabela expandida do Quadro 4.1 pode ser escrito como

$$\begin{aligned}\log P\left(Y_{1i}=y_{1i}, Y_{2i}=y_{2i}, \mathbf{X}_i=\mathbf{x}_i\right) &= \mu_0+\mu_{01}y_{1i}+\mu_{02}y_{2i}+y_{1i}\sum_{k=1}^{p}\mu_{1k}x_{ik} \\ &\quad +y_{2i}\sum_{k=1}^{p}\mu_{2k}x_{ik}+\sum_{k=1}^{p}\mu_k x_{ik} \qquad (4.10)\end{aligned}$$

em que $P(\cdot)$ é a probabilidade do indivíduo $i$ pertencer a uma dada casela da tabela de contingência de $Y_{1i}$, $Y_{2i}$ e $\mathbf{X}_i$ com $y_{ti}=0,1$; $\mathrm{t}=1,2$ e $i=1,2,\ldots,i$. Os parâmetros do modelo são denotados por $\mu$'s e $x_{ki}$ indica o valor da covariável $k$ para o indivíduo $i$ $(k=1,2,\ldots,p)$.

Os parâmetros em (4.10) são de fato iguais aos parâmetros em (4.2) e (4.3) (ver Proposição 7 do Apêndice A). Portanto, os termos $\frac{A_1}{D}$, $\frac{A_1A_2}{D}$, $\frac{A_2}{D}$ e $\frac{1}{D}$ nas equações (4.6) a (4.8) definem um *modelo log-linear* para tabelas de contingência observadas parcialmente com duas dimensões observadas, $\mathbf{X}_i$ e $Y_{1i}$, e uma dimensão observada parcialmente, $Y_{2i}$ e, assim, o modelo logito simplificado em (4.2) e (4.3) para dados categóricos da história de eventos é equivalente ao modelo log-linear para uma tabela de contingência observada parcialmente com dimensões $Y_{1i}$, $Y_{2i}$ e $\mathbf{X}_i$, tanto que a equação (4.9) corresponde à função de log-verossimilhança para o modelo log-linear em (4.10) [1].

O Quadro 4.2 apresenta as linhas da *matriz de delineamento* do modelo log-linear

[1] A função de verossimilhança do modelo log-linear é equivalente ao modelo logito em (4.1), exceto que o primeiro inclue um fator para os parâmetros da distribuição marginal de $\mathbf{x}_i$. Esses parâmetros são considerados parâmetros de perturbação no modelo logito e não afetam as estimativas dos outros parâmetros (Mare, 1994).

QUADRO 4.2 - MATRIZ DE DELINEAMENTO DO MODELO LOG-LINEAR, DADO $X_i$, PARA A TABELA EXPANDIDA DO QUADRO 5.1, COM SEUS CORREPONDENTES RESULTADOS OBSERVADOS DE $Y_{1i}$ E $Y_{2i}$

| Parâmetros | $\beta_{01}$ | $\beta_{02}$ | $\beta_1$ | $\beta_2$ | | | |
|---|---|---|---|---|---|---|---|
| Variáveis | $Y_{1i}$ | $Y_{2i}$ | $Y_{1i}X_{ik}$ | $Y_{2i}X_{ik}$ | | $Y_{1i}$ | $Y_{2i}$ |
| | 1 | 1 | $x_{ik}$ | $x_{ik}$ | | 1 | |
| Matriz de | 1 | 0 | $x_{ik}$ | 0 | Resultados | 1 | |
| Delineamento | 0 | 1 | 0 | $x_{ik}$ | Observados | 0 | 1 |
| | 0 | 0 | 0 | 0 | | 0 | 0 |

em (4.10) para um determinado nível de $\mathbf{X}_i$ e os corrrespondentes resultados observados. Cada linha da matriz de delineamento no Quadro 4.2 corresponde a uma casela da tabela expandida no Quadro 4.1. As primeiras quatro colunas da matriz de delineamento denotam os parâmetros do modelo logito simplificado em (4.2) e (4.3). A quinta e sexta colunas apresentam os resultados observados de $Y_{1i}$ e $Y_{2i}$. Dada esta matriz de delineamento no Quadro 4.2 e o mapeamento representado no Quadro 4.1, as estimativas de máxima verossimilhança dos parâmetros do modelo logito podem ser obtidas por intermédio do programa DNEWTON, que usa um algorítmo de Newton-Raphson modificado (ver Anexo B).

### Função de Verossimillhança Geral

Na seção anterior discutiu-se um modelo para distribuições de tempos de ocorrência com duas transições e nenhuma censura a priori na categoria final de transição, mas a abordagem apresentada por Mare (1994) pode ser generalizada para aplicações com distribuições que incluem qualquer número de transições e que admitem censura na última categoria de transição.

A metodologia geral tem a mesma construção do caso simples, isto é:

a) expressar as probabilidades das caselas da tabela de contigência observada da distribuição dos tempos de ocorrência do evento, em termos dos parâmetros do modelo logito em (4.1);

b) mapear as probabilidades das caselas da tabela de contingência do item (a) em uma tabela de contingência observada parcialmente da distribuição conjunta de transições.

A partir do modelo logito em (4.1), define-se

$$A_{\mathtt{t}i} = \exp\left(\beta_{0\mathtt{t}} + \boldsymbol{\beta}_{\mathtt{t}}'\mathbf{x}_{\mathtt{t}i}\right)$$

e

$$D_i = \prod_{s=1}^{N} \left(1 + A_{si}\right) \ ,$$

com $\mathtt{t} = 1, 2, \ldots, N$ e $i = 1, 2, \ldots, n$.

Então, a probabilidade das caselas para indivíduos que desistem de fazer a transição t (casos não-censurados) é

$$\begin{aligned} P\left(Y_{1i} = Y_{2i} = \ldots = Y_{\mathtt{t}-1,i} = 0, Y_{\mathtt{t}i} = 1 | \mathbf{x}_{\mathtt{t}i}\right) &= \frac{A_{\mathtt{t}i}}{1 + A_{\mathtt{t}i}} \prod_{s=1}^{\mathtt{t}-1} \frac{1}{1 + A_{si}} \\ &= A_{\mathtt{t}i} \left[ \frac{\prod_{s=\mathtt{t}+1}^{N} \left(1 + A_{si}\right)}{D_i} \right] . \end{aligned} \tag{4.11}$$

E a probabilidade das caselas para censura na transição t é

$$P\left(Y_{1i} = Y_{2i} = \ldots = Y_{\mathtt{t}i} = 0 | \mathbf{x}_{\mathtt{t}i}\right) = \prod_{s=1}^{\mathtt{t}} \frac{1}{1 + A_{si}} = \frac{\prod_{s=\mathtt{t}+1}^{N} \left(1 + A_{si}\right)}{D_i} \ , \tag{4.12}$$

em que censura em t significa que se desconhece o valor de $Y_{\mathtt{t}+1,i}$.

Juntas, as equações (4.11) e (4.12) podem ser usadas para construir a função de verossimilhança a fim de estimar o modelo logito em (4.1), tanto quanto o modelo log-linear em (4.10) para uma tabela de contingência parcialmente observada.

A princípio, a análise do modelo logito pode ser estendida para incorporar hetero-

geneidade não-observada, incluindo os termos que representem a fragilidade do indivíduo em (4.1). Mas, na prática, como discutido no Capítulo 2, a identificabilidade de misturas resultantes do modelo logito para indivíduos isolados é problemática, por causa das fortes suposições requeridas. No entanto, Mare (1994) mostrou que para indivíduos pareados o modelo de logito permite a identificabilidade da mistura e estimação dos modelos com heterogeneidade não-observada, sem fazer fortes suposições paramétricas.

## 4.3 Um Modelo Logito para Indivíduos Pareados

Considere agora os casos em que a distribuição dos tempos de ocorrência do evento é observada aos pares em intervalos de tempos, como por exemplo, o progresso de pares de irmãos em sucessivos níveis de escolaridade, a trajetória de pares de irmãs em partos sucessivos, e a trajetória de casais em sucessivos estágios de sobrevivência. Mare (1994) desenvolveu um modelo logito para analisar dados pareados como nos exemplos citados acima, que é uma extensão direta do modelo logito em (4.1). Para especificar o modelo para dados pareados é preciso interpretar e parametrizar a associação entre os tempos pareados.

Igualmente como para distribuições conjuntas de quaisquer variáveis discretas, a estrutura de dependência pode ser interpretada de dois diferentes modos. Uma possível interpretação é que a distribuição conjunta surge da associação de dois tempos pareados, que podem ser descritos ou inferidos porque uma transição realizada por um membro do par pode afetar a probabilidade do outro membro fazer a transição. Uma segunda interpretação é que a distribuição conjunta dos tempos de ocorrência surge da *heterogeneidade compartilhada* pelas observações pareadas, isto é, as características observadas e não-observadas comuns ao par de indivíduos explicariam a interdependência entre os dois tempos. Na análise desenvolvida por Mare (1994), assegura-se que observações pareadas provêm um meio útil e flexível para considerar os efeitos da heterogeneidade em nível de par. Uma situação muito diferente das restrições impostas para estimar

modelos com heterogenidade em nível individual. Nesta seção, a discussão está focalizada em como a heterogeneidade compartilhada em nível de par controla distribuições pareadas de tempos categóricos de ocorrência de eventos.

Vários tipos de abordagens são viáveis para representar a heterogeneidade não-observada em modelos de risco bivariado de tempo discreto. A abordagem de Mare assume que a heterogeneidade não-observada é uma variável *dummy* tendo uma distribuição misturadora discreta com um número de categorias (classes latentes) determinado pelo analista (ver, por exemplo, Heckman e Singer, 1982; Trussel e Richards, 1985). Outra classe de modelos representa a heterogeneidade não-observada por meio de efeitos fixados para cada par, que denota a "habilidade" do par em fazer transições sucessivas. Esta última classe de modelos é análoga aos modelos de Rasch para análise de itens (Lindsay, Clogg e Grego, 1991), em que os parâmetros para "indivíduo" e "item" são representados explicitamente. No conjunto de dados de interesse, transições tomariam o papel de itens e os pares tomariam o papel de indivíduos. Mare (1994) inclue mais observações acerca da aplicação de modelos de Rasch no contexto da análise da história de eventos.

Seja $\mathbf{X}_{\mathrm{t}ij}$ um vetor $(p_{\mathrm{t}} \times 1)$ de $p_{\mathrm{t}}$ covariáveis categóricas para o membro $i$ do par $j$ na transição t e seja $Y_{\mathrm{t}ij}$ uma variável dicotômica que é igual a 1 se o $i$-ésimo indivíduo do par $j$ desiste na transição t e é 0, caso contrário $(\mathrm{t} = 1,2,\ldots,N)\,(i = 1,2)\,(j = 1,2,\ldots,m)$. Uma variável *dummy* $\mathbf{W}_j$ é incorporada ao modelo logito para representar a heterogeneidade não-observada entre os pares, isto é, $\mathbf{W}_j$ é um vetor $((q-1) \times 1)$ de variáveis dicotômicas $W_{jr}$, que assumem o valor 1 se o par $j$ pertence à classe latente $r+1$ e 0, caso contrário $(r = 1,2,\ldots,q-1)$. O valor de $q$ é determinado pelo analista, mas é limitado pela quantidade de informação na tabela de dados, para identificabilidade do modelo [2].

Um modelo logito para os efeitos das covariáveis dado por

$$l\left(Y_{\mathrm{t}ij} = 1 | Y_{1ij} = Y_{2ij} = \ldots = Y_{\mathrm{t}-1,ij} = 0, \mathbf{X}_{\mathrm{t}ij} = \mathbf{x}_{\mathrm{t}ij}, \mathbf{W}_j = \mathbf{w}_j\right) = \beta_{0\mathrm{t}i} + \boldsymbol{\beta}'_{\mathrm{t}}\mathbf{x}_{\mathrm{t}ij} + \boldsymbol{\gamma}'_{\mathrm{t}i}\mathbf{w}_j \tag{4.13}$$

[2] O número de categorias para a heterogeneidade não-observada é igual a $q$. Quando $\mathbf{W}_j = \mathbf{0}$, significa que o par $j$ pertence à classe latente 1.

é definido para o indivíduo $i$ do par $j$, em que $l(\cdot)$ e $\boldsymbol{\beta}_{\mathrm{t}}$ é definido como em (4.1), $\boldsymbol{\gamma}_{\mathrm{t}i}$ é um vetor $((q-1)\times 1)$ de parâmetros desconhecidos que representa os efeitos das classes latentes sobre o logito de não fazer a transição t, dado que as transições anteriores foram feitas e $\beta_{0\mathrm{t}i}$ é o intercepto do modelo.

Em (4.13), $\boldsymbol{\gamma}_{\mathrm{t}i}$ varia entre os membros do par e entre as transições, mas na prática esses parâmetros são tipicamente restritos, de acordo com as exigências para identificabilidade do modelo logito [3] ou suposições substantivas acerca dos efeitos da heterogeneidade não-observada. Por exemplo, se os efeitos de $\mathbf{W}_j$ não variam entre os membros do par nem entre as transições, então $\boldsymbol{\gamma}_{\mathrm{t}i} = \boldsymbol{\gamma}$ para todo t e $i$. O número de classes latentes é limitado pelo número de categorias dos tempos de ocorrência e das transições que são observadas, para garantir identificabilidade do modelo. Sob (4.13), o vetor de covariáveis $\mathbf{X}_{\mathrm{t}ij}$ pode depender da transição e variar entre observações num par e entre pares. Os efeitos das covariáveis, representados por $\boldsymbol{\beta}_{\mathrm{t}}$, dependem da transição mas, como no caso de indivíduos pareados, esses parâmetros podem ser restritos na prática.

A suposição chave nos modelos de classes latentes é a *independência condicional*, ou seja, as transições para os dois membros de um par são independentes dentro de níveis comuns de $\mathbf{X}_{\mathrm{t}ij}$ e $\mathbf{W}_j$, assim

$$P\left(Y_{\mathrm{t}1j}=y, Y_{\mathrm{t}'2j}=y' | Y_{11j}=\ldots=Y_{\mathrm{t}-1,1j}=Y_{12j}=\ldots=Y_{\mathrm{t}'-1,2j}=0, \mathbf{x}_{\mathrm{t}1j}, \mathbf{x}_{\mathrm{t}'2j}, \mathbf{w}_j\right) =$$

$$= P(Y_{\mathrm{t}1j}=y | Y_{11j}=\ldots=Y_{\mathrm{t}-1,1j}=0, \mathbf{x}_{\mathrm{t}1j}, \mathbf{w}_j)\, P\left(Y_{\mathrm{t}'2j}=y' | Y_{12j}=\ldots=Y_{\mathrm{t}'-1,2j}=0, \mathbf{x}_{\mathrm{t}'2j}, \mathbf{w}_j\right), \quad (4.14)$$

em que $i=1,2$ e $j=1,2,\ldots,m$ $(y, y'=0,1)$.

Portanto, assume-se que a heterogeneidade observada e não-observada em nível de par explica completamente a associação entre os tempos de ocorrência do evento em um par de indivíduos.

[3] Lindsay, Clogg e Grego (1991) apresentam algumas disposições gerais sobre a identificabilidade de modelos de classes latentes, mas para a sua aplicabilidade em tempos pareados, como os discutidos neste capítulo, é necessário que mais estudos teóricos sejam feitos.

### 4.3.1 Interpretação das Tabelas de Contingência Observadas Parcialmente para Indivíduos Pareados

Os parâmetros em (4.13) podem ser estimados pelo método de máxima verossimilhança. A fim de construir a função de verossimilhança para o modelo em (4.13), é necessário especificar a relação entre a distribuição conjunta observada dos tempos de ocorrência do evento para os membros do par e os parâmetros do modelo. Esta relação e a função de verossimilhança para o modelo são apresentadas a seguir, numa simplificação do modelo em (4.13), em que há apenas duas transições de interesse para cada indivíduo do par ($N = 2$), a variável latente $W_j$ é escalar, tendo duas categorias representando a heterogeneidade não-observada no par $j$ ($q = 2$) e há algumas restrições nos parâmetros do modelo. Esta simplificação em (4.13), resulta em quatro equações:

$$l\left(Y_{11j} = 1 | \mathbf{X}_j = \mathbf{x}_j, W_j = w_j\right) = \beta_{011} + \boldsymbol{\beta}_1' \mathbf{x}_j + \gamma w_j \ , \tag{4.15}$$

$$l\left(Y_{21j} = 1 | Y_{11} = 0, \mathbf{X}_j = \mathbf{x}_j, W_j = w_j\right) = \beta_{021} + \boldsymbol{\beta}_2' \mathbf{x}_j + \gamma w_j \ , \tag{4.16}$$

$$l\left(Y_{12j} = 1 | \mathbf{X}_j = \mathbf{x}_j, W_j = w_j\right) = \beta_{011} + \boldsymbol{\beta}_1' \mathbf{x}_j + \gamma w_j \tag{4.17}$$

e

$$l\left(Y_{22j} = 1 | Y_{12} = 0, \mathbf{X}_j = \mathbf{x}_j, W_j = w_j\right) = \beta_{022} + \boldsymbol{\beta}_2' \mathbf{x}_j + \gamma w_j \ , \tag{4.18}$$

em que as probabilidades das transições obedecem à condição de independência em (4.14) e cada equação corresponde a uma transição feita por um indivíduo do par $j$.

O novo modelo simplificado assume que os efeitos das covariáveis $\mathbf{X}_j$ não variam entre os membros do par, mas para cada membro do par $j$ existe um único intercepto. A heterogeneidade não-observada é incorporada ao modelo por meio de uma única variável dicotômica $W_j$, que assume o valor 1 se o par de indivíduos pertence à segunda classe latente e 0 se o par pertence à primeira classe latente.

Na ausência de associação entre os membros do par (dado $\mathbf{x}_j$), isto é, se $\gamma = 0$, é possível estimar os parâmetros de cada equação em (4.15) até (4.18) usando os métodos

apresentados para estimação de tempos categóricos de ocorrência do evento para indivíduos isolados ou os métodos padrões para estimação de modelos de risco univariado para tempos discretos. Deve-se então simplesmente considerar os $m$ pares de observações como $2m$ indivíduos independentes. Em geral, $\gamma \neq 0$ e métodos para o caso pareado são necessários. Diferentemente do caso de indivíduos isolados, não é possível particionar distribuições discretas de tempos pareados dentro de seqüências de transições condicionalmente independentes. Contudo, pode-se estimar os parâmetros em (4.15) a (4.18), usando uma generalização direta da abordagem para análise de tabelas de contingência observadas parcialmente, desenvolvida nas Seções 4.1 e 4.2 para indivíduos isolados.

Como nos modelos para indivíduos isolados, o método consiste em construir uma distribuição conjunta hipotética das transições para cada membro do par. Para indivíduos isolados, esta distribuição hipotética é uma classificação cruzada dos resultados de cada transição, admitindo-se aqueles resultados em que indivíduos poderiam experimentar cada transição, sem considerar se eles estivessem de fato em risco de transição. No caso pareado, a distribuição hipotética provê esta classificação cruzada em cada membro do par, mais a classificação cruzada das transições dos dois membros do par. Por intermédio dessa distribuição hipotética, ilustrada no Quadro 4.3, é possível considerar o comportamento de cada indivíduo em cada transição, sem considerar se o par dele estava em risco na transição em questão, como também é possível modelar a associação entre os dois membros do par.

O método requer que as probabilidades da distribuição conjunta observada dos tempos categóricos de ocorrência do evento para cada nível de $\mathbf{X}_j$ e $W_j$ sejam expressas em termos dos parâmetros em (4.15) a (4.18) e, fazendo assim, estabelecer um mapeamento das caselas dessa distribuição observada para as caselas de uma distribuição conjunta observada parcialmente de $Y_{11j}, Y_{21j}, Y_{12j}$ e $Y_{22j}$.

Define-se, para $\mathrm{t} = 1, 2$, que

$$A_{\mathrm{t}} = \exp\left(\beta_{0\mathrm{t}1} + \boldsymbol{\beta}_{\mathrm{t}}'\mathbf{x} + \gamma w\right) \ ,$$

QUADRO 4.3 - MAPEAMENTO DA TABELA EXPANDIDA DE TRANSIÇÕES DE INDIVÍDUOS PAREADOS

Tabela Observada da Distribuição de Freqüências dos Resultados Possíveis de $Y_{11}$, $Y_{21}$, $Y_{12}$, e $Y_{22}$, dado **X** e $W$

| Resultados | $Y_{11}=1$ $Y_{12}=1$ | $Y_{11}=1$ $Y_{12}=0, Y_{22}=1$ | $Y_{11}=1$ $Y_{12}=0, Y_{22}=0$ | $Y_{11}=0, Y_{21}=1$ $Y_{12}=1$ | $Y_{11}=0, Y_{21}=1$ $Y_{12}=0, Y_{22}=1$ | $Y_{11}=0, Y_{21}=1$ $Y_{12}=0, Y_{22}=0$ | $Y_{11}=0, Y_{21}=0$ $Y_{12}=1$ | $Y_{11}=0, Y_{21}=0$ $Y_{12}=0, Y_{22}=1$ | $Y_{11}=0, Y_{21}=0$ $Y_{12}=0, Y_{22}=0$ |
|---|---|---|---|---|---|---|---|---|---|
| Freqüências | $a$ | $b$ | $c$ | $d$ | $e$ | $f$ | $g$ | $h$ | $i$ |

Tabela Expandida da Distribuição Conjunta de Freqüências de $Y_{11}$, $Y_{21}$, $Y_{12}$, e $Y_{22}$, dado **X** e $W$

| $Y_{11}$ | $Y_{21}$ | $Y_{12}=1$, $Y_{22}=1$ | $Y_{12}=1$, $Y_{22}=0$ | $Y_{12}=0$, $Y_{22}=1$ | $Y_{12}=0$, $Y_{22}=0$ |
|---|---|---|---|---|---|
| 1 | 1 | $a_1$ | $a_2$ | $b_1$ | $c_1$ |
| | 0 | $a_3$ | $a_4$ | $b_2$ | $c_2$ |
| 0 | 1 | $d_1$ | $d_2$ | $e$ | $f$ |
| | 0 | $g_1$ | $g_2$ | $h$ | $i$ |

Nota: $a_1 + a_2 + a_3 + a_4 = a$, $b_1 + b_2 = b$ e assim por diante para todas as letras com índice

Tabela Expandida da Distribuição Conjunta de Probabilidades de $Y_{11}$, $Y_{21}$, $Y_{12}$, e $Y_{22}$, dado **X** e $W$

| $Y_{11}$ | $Y_{21}$ | $Y_{12}=1$, $Y_{22}=1$ | $Y_{12}=1$, $Y_{22}=0$ | $Y_{12}=0$, $Y_{22}=1$ | $Y_{12}=0$, $Y_{22}=0$ |
|---|---|---|---|---|---|
| 1 | 1 | $\frac{A_1A_2B_1B_2}{D}$ | $\frac{A_1A_2B_1}{D}$ | $\frac{A_1A_2B_2}{D}$ | $\frac{A_1A_2}{D}$ |
| | 0 | $\frac{A_1B_1B_2}{D}$ | $\frac{A_1B_1}{D}$ | $\frac{A_1B_2}{D}$ | $\frac{A_1}{D}$ |
| 0 | 1 | $\frac{A_2B_1B_2}{D}$ | $\frac{A_2B_1}{D}$ | $\frac{A_2B_2}{D}$ | $\frac{A_2}{D}$ |
| | 0 | $\frac{B_1B_2}{D}$ | $\frac{B_1}{D}$ | $\frac{B_2}{D}$ | $\frac{1}{D}$ |

$$B_t = \exp\left(\beta_{0t2} + \boldsymbol{\beta}_t'\mathbf{x} + \gamma w\right)$$

e

$$D = (1 + A_1)(1 + A_2)(1 + B_1)(1 + B_2) \ ,$$

em que o índice $j$ foi omitido e $\beta_{0ti}$ e $\boldsymbol{\beta}_t$ são os parâmetros em (4.15) a (4.18).

Dados $\mathbf{x}_j$ e $w_j$, as nove probabilidades para as correspondentes caselas da tabela observada no Quadro 4.3 podem ser obtidas usando um procedimento semelhante ao apresentado no caso de indivíduos isolados (ver Seção 4.1). A distribuição conjunta de probabilidades dos resultados possíveis de $Y_{11j}, Y_{21j}, Y_{12j}$ e $Y_{22j}$ é formada por

$$\begin{aligned} P(Y_{11} = 1, Y_{12} = 1|\mathbf{x}, w) &= \frac{A_1}{1 + A_1}\frac{B_1}{1 + B_1} \\ &= \frac{A_1 B_1}{D} + \frac{A_1 B_1 B_2}{D} + \frac{A_1 A_2 B_1}{D} + \frac{A_1 A_2 B_1 B_2}{D} \ , \end{aligned} \tag{4.19}$$

$$\begin{aligned} P(Y_{11} = 1, Y_{12} = 0, Y_{22} = 1|\mathbf{x}, w) &= \frac{A_1}{1 + A_1}\frac{1}{1 + B_1}\frac{B_2}{1 + B_2} \\ &= \frac{A_1 B_2}{D} + \frac{A_1 A_2 B_2}{D} \ , \end{aligned} \tag{4.20}$$

$$\begin{aligned} P(Y_{11} = 1, Y_{12} = 0, Y_{22} = 0|\mathbf{x}, w) &= \frac{A_1}{1 + A_1}\frac{1}{1 + B_1}\frac{1}{1 + B_2} \\ &= \frac{A_1}{D} + \frac{A_1 A_2}{D} \ , \end{aligned} \tag{4.21}$$

$$\begin{aligned} P(Y_{11} = 0, Y_{21} = 1, Y_{12} = 1|\mathbf{x}, w) &= \frac{1}{1 + A_1}\frac{A_2}{1 + A_2}\frac{B_1}{1 + B_1} \\ &= \frac{A_2 B_1}{D} + \frac{A_2 B_1 B_2}{D} \ , \end{aligned} \tag{4.22}$$

$$P(Y_{11} = 0, Y_{21} = 1, Y_{12} = 0, Y_{22} = 1|\mathbf{x}, w) = \frac{1}{1 + A_1}\frac{A_2}{1 + A_2}\frac{1}{1 + B_1}\frac{B_2}{1 + B_2}$$

$$= \frac{A_2 B_2}{D} \tag{4.23}$$

$$\begin{aligned} P(Y_{11}=0, Y_{21}=1, Y_{12}=0, Y_{22}=0|\mathbf{x}, w) &= \frac{1}{1+A_1}\frac{A_2}{1+A_2}\frac{1}{1+B_1}\frac{1}{1+B_2} \\ &= \frac{A_2}{D}\,, \end{aligned} \tag{4.24}$$

$$\begin{aligned} P(Y_{11}=0, Y_{21}=0, Y_{12}=1|\mathbf{x}, w) &= \frac{1}{1+A_1}\frac{1}{1+A_2}\frac{B_1}{1+B_1} \\ &= \frac{B_1}{D} + \frac{B_1 B_2}{D}\,, \end{aligned} \tag{4.25}$$

$$\begin{aligned} P(Y_{11}=0, Y_{21}=0, Y_{12}=0, Y_{22}=1|\mathbf{x}, w) &= \frac{1}{1+A_1}\frac{1}{1+A_2}\frac{1}{1+B_1}\frac{B_2}{1+B_2} \\ &= \frac{B_2}{D}\,, \end{aligned} \tag{4.26}$$

e

$$\begin{aligned} P(Y_{11}=0, Y_{21}=0, Y_{12}=0, Y_{22}=0|\mathbf{x}, w) &= \frac{1}{1+A_1}\frac{1}{1+A_2}\frac{1}{1+B_1}\frac{1}{1+B_2} \\ &= \frac{1}{D}\,. \end{aligned} \tag{4.27}$$

Essas equações podem ser usadas para formar a função de verossimilhança do modelo logito. Denote-se a probabilidade do par $j$ pertencer à classe latente $r$ $(r = 1, \ldots, q)$ por $\pi_{jr}$, em que $\sum_{r=1}^{q} \pi_{jr} = 1$. Então, com a simplificação de (4.13), a função de verossimilhança é dada por

$$\begin{aligned} L(\boldsymbol{\theta}) = \prod_{j=1}^{m} \Bigg\{ \sum_{r=1}^{2} \pi_{jr} \Big[ & P(Y_{11j} = Y_{12j} = 1|\mathbf{x}_j, w_j)^{Y_{11j}Y_{12j}} \\ & \cdot P(Y_{11j} = Y_{22j} = 1, Y_{12j} = 0|\mathbf{x}_j, w_j)^{Y_{11j}(1-Y_{12j})Y_{22j}} \\ & \cdot P(Y_{11j} = 1, Y_{12j} = Y_{22j} = 0|\mathbf{x}_j, w_j)^{Y_{11j}(1-Y_{12j})(1-Y_{22j})} \end{aligned}$$

$$\cdot P\left(Y_{11j}=0, Y_{21j}=Y_{12j}=1|\mathbf{x}_j, w_j\right)^{(1-Y_{11j})Y_{21j}Y_{12j}}$$
$$\cdot P\left(Y_{11j}=Y_{12j}=0, Y_{21j}=Y_{22j}=1|\mathbf{x}_j, w_j\right)^{(1-Y_{11j})Y_{21j}(1-Y_{12j})Y_{22j}}$$
$$\cdot P\left(Y_{11j}=Y_{12j}=Y_{22j}=0, Y_{21j}=1|\mathbf{x}_j, w_j\right)^{(1-Y_{11j})Y_{21j}(1-Y_{12j})(1-Y_{22j})}$$
$$\cdot P\left(Y_{11j}=Y_{21j}=0, Y_{12j}=1|\mathbf{x}_j, w_j\right)^{(1-Y_{11j})(1-Y_{21j})Y_{12j}}$$
$$\cdot P\left(Y_{11j}=Y_{21j}=Y_{12j}=0, Y_{22j}=1|\mathbf{x}_j, w_j\right)^{(1-Y_{11j})(1-Y_{21j})(1-Y_{12j})Y_{22j}}$$
$$\left.\cdot P\left(Y_{11j}=Y_{21j}=Y_{12j}=Y_{22j}=0|\mathbf{x}_j, w_j\right)^{(1-Y_{11j})(1-Y_{21j})(1-Y_{12j})(1-Y_{22j})}\right]\Bigg\},$$

em que $\boldsymbol{\theta}=(\beta_{0ti}, \boldsymbol{\beta}_{\mathrm{t}}, \gamma_{\mathrm{t}}, \pi_{jr})\,(\mathrm{t}, i, r=1,2)\,(j=1,2,\ldots,m)$.

A interpretação das probabilidades em (4.19) a (4.27) é paralela àquela feita para as equações (4.6) até (4.8) no caso de indivíduos isolados. Cada parcela do lado direito de (4.19) até (4.27) corresponde a uma probabilidade da casela da tabela observada parcialmente (para uma determinada categoria de $\mathbf{X}_j$ e $W_j$). As probabilidades das nove caselas da tabela observada dos tempos de ocorrência do evento decompõem-se em $2^4=16$ probabilidades conjuntas de $Y_{11j}$, $Y_{21j}$, $Y_{12j}$ e $Y_{22j}$, em que somente quatro são de resultados possíveis (ver Quadro 4.3). Os resultados relativos às probabilidades dadas em (4.23), (4.24), (4.26) e (4.27) são diretamente observados visto que cada um deles corresponde exatamente a um resultado possível da distribuição conjunta de $Y_{11j}$, $Y_{21j}$, $Y_{12j}$ e $Y_{22j}$. Em contraste, os resultados relativos às probabilidades dadas nas outras cinco equações não são diretamente observados, pois correspondem a resultados hipotéticos da distribuição conjunta das transições. Por exemplo, para pares nos quais $Y_{11j}=Y_{12j}=1$ os valores de $Y_{21j}$ e $Y_{22j}$ não podem ser observados. A equação (4.19) constroe o mapeamento da probabilidade desses pares em quatro probabilidades de combinações hipoteticamente possíveis de $Y_{21j}$ e $Y_{22j}$. Em geral, (4.19) até (4.27) estruturam o mapeamento das caselas da tabela observada de tempos pareados de ocorrência nas caselas de uma tabela observada parcialmente, que representa a distribuição conjunta hipotética de $Y_{11j}$, $Y_{21j}$, $Y_{12j}$ e $Y_{22j}$.

As equações de (4.19) até (4.27) definem implicitamente um modelo de classes latentes, pois esta correspondência é mais complexa que no modelo de classes latentes

ordinal [4]. No modelo logito para indivíduos isolados aparece uma estrutura parcialmente latente, mas no caso pareado é um modelo de classes latentes de duas maneiras: primeira, dentro das categorias da variável latente $W_j$, as equações (4.14) e (4.19) até (4.27) definem um modelo com classificação parcialmente latente (relativo à distribuição conjunta das quatro categorias em $Y_{21j}$ e $Y_{22j}$, que estão classificadas sob $\mathbf{X}_j$, $Y_{11j}$ e $Y_{12j}$), analogamente ao modelo de indivíduos isolados; e segunda, porque dentro das categorias da covariável $\mathbf{X}_j$, as equações (4.14) e (4.19) até (4.27) definem um modelo de classes latentes para a distribuição conjunta de $Y_{11j}$, $Y_{21j}$, $Y_{12j}$ e $Y_{22j}$, no qual a interdependência das transições pareadas é explicada por sua associação comum com a variável latente (não-observada) $W_j$.

### 4.3.2 Estimação do Modelo Logito para Indivíduos Pareados

Os 16 termos do lado direito das equações (4.19) até (4.27) definem um modelo de classes latentes no qual há uma variável não-observada $W_j$ e duas variáveis observadas parcialmente $Y_{21j}$ e $Y_{22j}$, que estão interligadas às variáveis observadas $\mathbf{X}_j$, $Y_{11j}$ e $Y_{12j}$. Analogamente, esses termos definem um modeo log-linear para uma tabela de contingência observada parcialmente com dimensões $Y_{11j}$, $Y_{21j}$, $Y_{12j}$, $Y_{22j}$, $\mathbf{X}_j$ e $W_j$. O modelo log-linear pode ser escrito como

$$\log P\left(Y_{11j}=y_{11j}, Y_{21j}=y_{21j}, Y_{12j}=y_{12j}, Y_{22j}=y_{22j}, \mathbf{X}_j=\mathbf{x}_j, W_j=w_j\right) =$$

$$= \mu_0 + \sum_{t=1}^{2}\sum_{i=1}^{2}\mu_{0ti}\,y_{tij} + \left(y_{11j}+y_{12j}\right)\sum_{k=1}^{p}\mu_{1k}x_{jk} +$$

$$+ \left(y_{21j}+y_{22j}\right)\sum_{k=1}^{p}\mu_{2k}x_{jk} + \mu\left(\sum_{t=1}^{2}\sum_{i=1}^{2}y_{tij}w_j\right) + \sum_{k=1}^{p}\mu_k x_{jk}\ , \qquad (4.28)$$

[4]Quando as variáveis de interesse são dicotômicas e um modelo com uma única variável latente com $q$ classes é postulado, este é conhecido como modelo de classes latentes. Everitt (1984) apresenta uma descrição introdutória dos modelos de classes latentes.

em que $P(\cdot)$ denota a probabilidade do par $j$ pertencer a uma dada casela da tabela de contingência de $Y_{11j}$, $Y_{21j}$, $Y_{12j}$, $Y_{22j}$, $\mathbf{X}_j$ e $W_j$ com $y_{tij} = 0, 1$; $\mathrm{t} = 1, 2$; $i = 1, 2$ e $j = 1, 2, \ldots m$. Os parâmetros do modelo são denotados por $\mu$'s e $x_{kj}$ é o valor da covariável $k$ para o par $j$ ($k = 1, 2, \ldots, p$). Como em (4.15) até (4.18), o modelo em (4.28) inclue interceptos $\mu_{0ti}$ específicos para indivíduos e transições, efeitos das covariáveis $\mu_{1k}$ e $\mu_{2k}$ específicos para 1ª e 2ª transições, respectivamente, e um único parâmetro $\mu$ para o efeito da heterogeneidade não-observada. Os parâmetros de (4.28) envolvidos com $Y_{tij}$ são de fato iguais aos de (4.15) até (4.18), isto é, $\mu_1 = \beta_{01}$, $\mu_2 = \beta_{02}$, $\mu_{1k} = \beta_{1k}$, $\mu_{2k} = \beta_{2k}$ e $\mu = \gamma$ ([5]).

O Quadro 4.4 apresenta a matriz de delineamento de um modelo de classes latentes generalizado para um dado nível de $\mathbf{X}_j$. Mare (1994) chama-o de modelo de classes latentes (não-proporcional) com duas classes. No Quadro 4.4 também são apresentados os correspondentes resultados observados. Cada linha da matriz corresponde a uma casela na tabela observada parcialmente. O modelo de classes latentes requer uma matriz de delineamento com 32 linhas (16 linhas para cada classe). As primeiras seis colunas da matriz correspondem aos quatro interceptos, um para cada membro do par e para cada uma das duas transições, mais os efeitos específicos para as transições. A sétima e oitava coluna correspondem, no modelo de classes latentes, a dois parâmetros para a heterogeneidade não-observada: um parâmetro para a distribuição da variável latente $W_j$ e outro para os efeitos de $W_j$ sobre as transições. As colunas restantes correspondem aos resultados observados de $Y_{11j}$, $Y_{21j}$, $Y_{12j}$ e $Y_{22j}$. Por meio desta matriz de delineamento no Quadro 4.4 juntamente com o mapeamento no Quadro 4.3, pode-se obter as estimativas de máxima verossimilhança dos parâmetros em (4.15) a (4.18), utilizando o programa DNEWTON (ver Anexo B).

[5] Esta equivalência de parâmetros do modelos log-linear e logito pode ser demonstrada de modo semelhante ao apresentado para o caso de indivíduos isolados (ver Proposição 7 do Apêndice A)

QUADRO 4.4 - MATRIZ DE DELINEAMENTO DO MODELO LOG-LINEAR, DADO $X_k$, E $W_j$ PARA A TABELA EXPANDIDA DO QUADRO 5.3, COM SEUS CORRESPONDENTES RESULTADOS OBSERVADOS DE $Y_{11j}$, $Y_{21j}$, $Y_{12j}$ E $Y_{22j}$

| Parâmetros | $\beta_{011}$ | $\beta_{021}$ | $\beta_{012}$ | $\beta_{022}$ | $\beta_{1k}$ | $\beta_{2k}$ | $\gamma$ | |
|---|---|---|---|---|---|---|---|---|
| Variáveis | $Y_{11j}$ | $Y_{21j}$ | $Y_{12j}$ | $Y_{22j}$ | $Y_{1j}X_{jk}$ | $Y_{2j}X_{jk}$ | $W_j$ | $Y_{..}W_j$ |
| | 1 | 1 | 1 | 1 | $2x_{jk}$ | $2x_{jk}$ | 0 | 0 |
| | 1 | 1 | 1 | 1 | $2x_{jk}$ | $2x_{jk}$ | 1 | 4 |
| | 1 | 1 | 1 | 0 | $2x_{jk}$ | $1x_{jk}$ | 0 | 0 |
| | 1 | 1 | 1 | 0 | $2x_{jk}$ | $1x_{jk}$ | 1 | 3 |
| | 1 | 0 | 1 | 1 | $2x_{jk}$ | $1x_{jk}$ | 0 | 0 |
| | 1 | 0 | 1 | 1 | $2x_{jk}$ | $1x_{jk}$ | 1 | 3 |
| | 1 | 0 | 1 | 0 | $2x_{jk}$ | 0 | 0 | 0 |
| | 1 | 0 | 1 | 0 | $2x_{jk}$ | 0 | 1 | 2 |
| | 0 | 1 | 1 | 1 | $1x_{jk}$ | $2x_{jk}$ | 0 | 0 |
| | 0 | 1 | 1 | 1 | $1x_{jk}$ | $2x_{jk}$ | 1 | 3 |
| | 0 | 1 | 1 | 0 | $1x_{jk}$ | $1x_{jk}$ | 0 | 0 |
| | 0 | 1 | 1 | 0 | $1x_{jk}$ | $1x_{jk}$ | 1 | 2 |
| | 0 | 0 | 1 | 1 | $1x_{jk}$ | $1x_{jk}$ | 0 | 0 |
| | 0 | 0 | 1 | 1 | $1x_{jk}$ | $1x_{jk}$ | 1 | 2 |
| | 0 | 0 | 1 | 0 | $1x_{jk}$ | 0 | 0 | 0 |
| Matriz de | 0 | 0 | 1 | 0 | $1x_{jk}$ | 0 | 1 | 1 |
| Delineamento | 1 | 1 | 0 | 1 | $1x_{jk}$ | $2x_{jk}$ | 0 | 0 |
| | 1 | 1 | 0 | 1 | $1x_{jk}$ | $2x_{jk}$ | 1 | 3 |
| | 1 | 0 | 0 | 1 | $1x_{jk}$ | $1x_{jk}$ | 0 | 0 |
| | 1 | 0 | 0 | 1 | $1x_{jk}$ | $1x_{jk}$ | 1 | 2 |
| | 0 | 1 | 0 | 1 | 0 | $2x_{jk}$ | 0 | 0 |
| | 0 | 1 | 0 | 1 | 0 | $2x_{jk}$ | 1 | 2 |
| | 0 | 0 | 0 | 1 | 0 | $1x_{jk}$ | 0 | 0 |
| | 0 | 0 | 0 | 1 | 0 | $1x_{jk}$ | 1 | 1 |
| | 1 | 1 | 0 | 0 | $1x_{jk}$ | $1x_{jk}$ | 0 | 0 |
| | 1 | 1 | 0 | 0 | $1x_{jk}$ | $1x_{jk}$ | 1 | 2 |
| | 1 | 0 | 0 | 0 | $1x_{jk}$ | 0 | 0 | 0 |
| | 1 | 0 | 0 | 0 | $1x_{jk}$ | 0 | 1 | 1 |
| | 0 | 1 | 0 | 0 | 0 | $1x_{jk}$ | 0 | 0 |
| | 0 | 1 | 0 | 0 | 0 | $1x_{jk}$ | 1 | 1 |
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

| | $Y_{11j}$ | $Y_{21j}$ | $Y_{12j}$ | $Y_{22j}$ |
|---|---|---|---|---|
| | 1 | | 1 | |
| | 1 | | 1 | |
| | 1 | | 1 | |
| | 1 | | 1 | |
| | 1 | | 1 | |
| | 1 | | 1 | |
| | 1 | | 1 | |
| | 1 | | 1 | |
| | 0 | 1 | 1 | |
| | 0 | 1 | 1 | |
| | 0 | 1 | 1 | |
| | 0 | 1 | 1 | |
| | 0 | 0 | 1 | |
| | 0 | 0 | 1 | |
| | 0 | 0 | 1 | |
| Resultados | 0 | 0 | 1 | |
| Observados | 1 | | 0 | 1 |
| | 1 | | 0 | 1 |
| | 1 | | 0 | 1 |
| | 1 | | 0 | 1 |
| | 0 | 1 | 0 | 1 |
| | 0 | 1 | 0 | 1 |
| | 0 | 0 | 0 | 1 |
| | 0 | 0 | 0 | 1 |
| | 1 | | 0 | 0 |
| | 1 | | 0 | 0 |
| | 1 | | 0 | 0 |
| | 1 | | 0 | 0 |
| | 0 | 1 | 0 | 0 |
| | 0 | 1 | 0 | 0 |
| | 0 | 0 | 0 | 0 |
| | 0 | 0 | 0 | 0 |

Nota: $Y_{.j} = \sum_i Y_{ij}$ e $Y_{..} = \sum_i \sum_t Y_{ij}$ e $\beta_{tk}$ é o efeito da covariável $k$ sobre o logito de não fazer a transição **t**, dado que a transição anterior foi feita (**t**, $i = 1, 2$) ($j = 1, 2, \ldots, m$) ($k = 1, 2, \ldots, p$).

### Função de Verossimilhança Geral

As equações (4.19) até (4.27) provêm os componentes da função de verossimilhança para distribuições pareadas de tempos de ocorrência com duas transições, sem censura antes da categoria final, e uma variável latente dicotômica. Nesta seção são apresentados os componentes da função de verossimilhança do modelo logito para indivíduos pareados, com distribuições que incluam um número arbitrário de transições, com censura à direita em qualquer transição depois da primeira, e com um grande número de pontos no conjunto suporte da distribuição discreta da heterogeneidade não-observada, tão grande quanto for permitido para a identificabilidade do modelo. A partir do modelo logito em (4.13) para indivíduos pareados, define-se

$$A_{\mathtt{t}ij} = \exp\left(\beta_{0\mathtt{t}i} + \boldsymbol{\beta}'_{\mathtt{t}i}\mathbf{x}_{\mathtt{t}ij} + \boldsymbol{\lambda}'_{\mathtt{t}i}\mathbf{w}_j\right)$$

e

$$D_j = \prod_{s=1}^{N}\prod_{s'=1}^{N}\left(1 + A_{s1j}\right)\left(1 + A_{s'2j}\right) \ ,$$

em que $\beta_{0\mathtt{t}i}$ e $\boldsymbol{\beta}_{\mathtt{t}i}$ são os parâmetros do modelo em (4.13) com $\mathtt{t} = 1, 2, \ldots, N$; $i = 1, 2$ e $j = 1, 2, \ldots, m$. A probabilidade das caselas para pares nos quais o indivíduo 1 desiste de fazer a transição $\mathtt{t}$ e o indivíduo 2 desiste de fazer a transição $\mathtt{t}'$ (casos não-censurados) é

$$\begin{aligned}
&P\left(Y_{11j} = Y_{21j} = \ldots = Y_{\mathtt{t}-1,1j} = Y_{12j} = Y_{22j} = \ldots = Y_{\mathtt{t}'-1,2j} = 0, Y_{\mathtt{t}1j} = Y_{\mathtt{t}'2j} = 1 | \mathbf{x}_{\mathtt{t}1j}, \mathbf{x}_{\mathtt{t}'2j}, \mathbf{w}_j\right) = \\
&= \frac{A_{\mathtt{t}1j}}{1 + A_{\mathtt{t}1j}} \frac{A_{\mathtt{t}'2j}}{1 + A_{\mathtt{t}'2j}} \prod_{s=1}^{\mathtt{t}-1}\prod_{s'=1}^{\mathtt{t}'-1}\left(\frac{1}{1 + A_{s1j}}\right)\left(\frac{1}{1 + A_{s'2j}}\right) \\
&= \frac{A_{\mathtt{t}1j}A_{\mathtt{t}'2j}}{D_j} \prod_{s=\mathtt{t}+1}^{N}\prod_{s'=\mathtt{t}'+1}^{N}\left(1 + A_{s1j}\right)\left(1 + A_{s'2j}\right) \ . \qquad (4.29)
\end{aligned}$$

A probabilidade das caselas para pares nos quais o membro $i$ desiste na transição $\mathtt{t}$ e o membro $i'$ está censurado na transição $\mathtt{t}'$ é

$$P\left(Y_{1ij} = Y_{2ij} = \ldots = Y_{\mathtt{t}-1,ij} = Y_{1i'j} = Y_{2i'j} = \ldots = Y_{\mathtt{t}'i'j} = 0, Y_{\mathtt{t}ij} = 1 | \mathbf{x}_{\mathtt{t}ij}, \mathbf{x}_{\mathtt{t}'i'j}, \mathbf{w}_j\right) =$$

$$= \frac{A_{tij}}{D_j} \prod_{s=t+1}^{N} \prod_{s'=t'+1}^{N} (1 + A_{sij})(1 + A_{s'i'j}) \ . \tag{4.30}$$

E a probabilidade das caselas para pares nos quais o indivíduo 1 está censurado na transição $\mathbf{t}$ e o indivíduo 2 está censurado na transição $\mathbf{t'}$ é

$$P\left(Y_{11j} = Y_{21j} = ... = Y_{t1j} = Y_{12j} = Y_{22j} = ... = Y_{t'2j} = 0 | \mathbf{x}_{t1j}, \mathbf{x}_{t'2j}, \mathbf{w}_j\right) =$$

$$= \frac{1}{D_j} \prod_{s=t+1}^{N} \prod_{s'=t'+1}^{N} (1 + A_{s1j})(1 + A_{s'2j}) \ . \tag{4.31}$$

Juntas, as expressões (4.29) até (4.31) podem ser usadas para formar a função de verossimilhança de (4.13).

Os modelos de Mare para análise de dados pareados de tempos categóricos são generalizações diretas dos modelos log-lineares que são os modelos mais usados quando os tempos de ocorrência são categóricos para indivíduos isolados. Quando as observações são de pares de indivíduos, o modelo para tempos pareados proporciona uma maneira útil de incorporar a heterogeneidade não-observada. Usando a informação fornecida pelas transições dentro de cada par, pode-se identificar os efeitos da heterogeneidade não-observada sobre o logito de não fazer uma determinada transição dado que passou pelas transições anteriores, sem fortes suposições arbitrárias sobre os parâmetros da distribuição da heterogeneidade não-observada. Este tipo de modelo, porém, assume que a parte essencial da heterogeneidade não-observada que vicia os efeitos das covariáveis resulta de fatores que são comuns aos membros do par. Se os fatores não-observados específicos ao membro bem como os específicos ao par são importantes, restrições adicionais devem ser aplicadas no modelo com indivíduos pareados, mantendo a identificabilidade do modelo.

A extensão dos conceitos dos modelos de dados pareados é feita diretamente para agrupamentos de maior tamanho. Segundo Mare (1994), na análise com agrupamentos maiores e mais complexos e distribuições do tempo de ocorrência do evento com um número de categorias moderadamente grande, é possível especificar os efeitos das variáveis

não-observadas. E os efeitos da heterogeneidade não-observada podem variar no tempo e entre os membros do grupo. Não há dificuldades conceituais, segundo Mare (1994), em estender estes modelos a fim de considerar riscos competitivos e transições de vários tipos. Por causa do isomorfismo com alguns modelos gerais para análise de dados categóricos ordinais, Mare (1994) sugere que os modelos discutidos aqui poderiam ser úteis para a análise de outras respostas ordinais pareadas ou de agrupamentos de maior tamanho que não sejam propriamente dados da análise da história de eventos.

Uma característica restritiva dos modelos que exploram a conexão entre os modelos log-lineares para tabelas de contingência e o modelo logito de dados agrupados é que são limitados quando o risco padrão for contínuo no modelo logito (ver Mare, 1994; Allison, 1982).

Uma desvantagem dos modelos de Mare é que eles são computacionalmente caros quando o número de categorias da distribuição do tempo de ocorrência do evento é grande. Em geral, quando a distribuição tem $N$ transições, o número de caselas na tabela expandida é $2^{2N}$ e esta tabela deve ser ainda embutida dentro da classificação conjunta das covariáveis observadas e variáveis latentes (não-observadas). Então, para grandes aplicações pode ser necessário usar modelos mais simples.

A estimação do modelo dado pela equação (4.13) pode ser ilustrada analisando-se dados de escolaridade de irmãos, como em Mare (1994). No restante deste capítulo é apresentada uma aplicação dos modelos de Mare para estimar o efeito da escolaridade do pai sobre a progressão escolar dos filhos, com os dados amostrais do censo demográfico brasileiro de 1991.

## 4.4 Transições Escolares de Irmãos

Um assunto central nos estudos sobre progressão e evasão escolar é o efeito da influência familiar sobre escolaridade dos seus membros mais novos. Embora anos de estudo seja uma variável quantitativa, a maioria das divisões do sistema escolar, como pré-escola, 1º

grau, 2º grau, 3º grau, etc, forma uma progressão discreta, e em cada estágio as pessoas estão sujeitas ao risco de desistir. Os estágios do processo escolar seriam a escala na qual os anos de estudo são mensurados e o tempo-calendário que a pessoa leva para completar uma etapa dos estágios escolares é irrelevante. Vários estudos analisam os efeitos de características observadas de antecedentes familiares sobre a probabilidade de concluir uma etapa escolar (ver Mare, 1994), mas a validade de seus estudos depende do controle adequado dos fatores observados e não-observados que afetam a escolaridade dos filhos de uma família. Uma estratégia para obter uma melhor estimativa dos efeitos dos fatores familiares é usar a informação escolar de irmãos. Dados sobre irmãos provêm um meio de estimar o efeito de fatores comuns de nível familiar, observados ou não, que influenciam na progressão escolar desses irmãos.

Segundo Mare (1994), fatores que exercem influência na escolaridade de irmãos e que não são observados podem ser classificados em dois tipos: heterogeneidade não-observada que é comum a todo descendente de uma família e heterogeneidade não-observada que varia entre os irmãos de uma família. O primeiro tipo de heterogeneidade não-observada inclui fatores que irmãos compartilham, como situação financeira estável da família, características da vizinhança e a herança genética comum. No modelo logito, essas fontes de variação não-observadas são variáveis latentes que afetam as transições escolares de todos os irmãos dentro de uma mesma família. O segundo tipo pode incluir fatores que irmãos experimentam diferentemente, como traços singulares de personalidade. Se excluídos os gêmeos, flutuações da renda familiar e experiência no mercado de trabalho também são fontes de heterogeneidade não-observadas entre irmãos. Se a heterogeneidade entre irmãos de uma mesma família é, relativamente, de menor importância quando comparada com a heterogeneidade entre famílias, então, através da informação escolar de irmãos, é possível incluir o primeiro tipo de heterogeneidade não-observada no modelo logito (ver Seção 4.3), a fim de analisar os determinantes da progressão escolar de nível familiar.

Existe ainda a possibilidade de estimar a probabilidade do indivíduo não passar por uma determinada transição, mesmo que ele tenha desistido em alguma transição ante-

rior. No contexto de dados escolares, há razões especiais para considerar as probabilidades hipotéticas. Por exemplo, a taxa de transição do 2º grau para a faculdade entre aqueles que não terminaram o 2º grau pode ou não ser igual à taxa de transição entre estudantes que de fato se graduaram no 2º grau. É razoável presumir que os que realmente concluiram o 2º grau são uma população seleta e assim, os desistentes, provavelmente não se comportariam do mesmo modo que os graduados, mesmo se a esses fosse dada a chance de entrar numa faculdade. Uma outra motivação para considerar a distribuição hipotética de desistentes deriva da avaliação sobre a dificuldade de concluir uma etapa escolar e ingressar na próxima. Como, por exemplo, verificar se concluir o 1º grau é mais fácil ou mais difícel que ingressar no 2º grau. Uma investigação para esta questão requer que se considere a distribuição hipotética do ingresso ou não de pessoas no 2º grau que falharam em concluir o 1º grau.

### 4.4.1 Descrição do Conjunto de Dados

A partir da amostra do censo demográfico brasileiro de 1991 [6], que contém os registros pessoais com informação sobre escolaridade e relação com o chefe da família, foi possível selecionar famílias que, na época em que o censo foi realizado, tinham apenas dois filhos homens (não-gêmeos) com 20 anos ou mais de idade, morando no mesmo domicílio dos pais (ver IBGE, 1994) [7]. Foram excluídas famílias em que o pai ou os filhos freqüentavam a escola na época em que o censo foi realizado e famílias com falta de informações sobre a escolaridade do pai ou dos filhos [8].

A Tabela 4.1 classifica as famílias especificadas acima por nível escolar do pai e dos

[6] Essa amostra é um levantamento que a fundação Instituto Brasileiro de Geografia e Estatística (IBGE) realizou juntamente com o recenseamento de 1991. Para maiores detalhes ver IBGE (1996).

[7] A escolha deste tipo de família mantem a mesma estrutura de análise realizada por Mare (1994) para os dados de escolaridade dos Estados Unidos, exceto a restrição para filhos morando no mesmo domicílio dos pais. Nada impede, que famílias com duas filhas com mais de 25 anos de idade e residindo no mesmo domicílio dos pais, por exemplo, possam ser selecionadas e analisadas da mesma forma. Não é possível selecionar famílias cujos filhos já não estavam morando no mesmo domicílio dos pais por causa da natureza dos dados do levantamento.

[8] Poucas exclusões foram feitas para eliminar dados censurados.

TABELA 4.1 - DISTRIBUIÇÃO DE ESCOLARIDADE DOS FILHOS E DO PAI EM UMA AMOSTRA DE 8620 FAMÍLIAS RESIDENTES NO BRASIL EM 1991

| ANOS DE ESTUDO DO PAI | ANOS DE ESTUDO DO 2º FILHO | ANOS DE ESTUDO DO 1º FILHO | | |
|---|---|---|---|---|
| | | < 8 | 8 | > 8 |
| 0 - 3 | < 8 | 2457 | 218 | 217 |
| | 8 | 240 | 179 | 128 |
| | > 8 | 322 | 132 | 488 |
| 4 | < 8 | 575 | 119 | 155 |
| | 8 | 152 | 132 | 140 |
| | > 8 | 186 | 126 | 665 |
| 5 - 7 | < 8 | 56 | 14 | 29 |
| | 8 | 10 | 18 | 18 |
| | > 8 | 16 | 13 | 132 |
| 8 | < 8 | 50 | 22 | 33 |
| | 8 | 16 | 38 | 54 |
| | > 8 | 24 | 38 | 286 |
| > 8 | < 8 | 28 | 15 | 51 |
| | 8 | 9 | 31 | 66 |
| | > 8 | 38 | 44 | 840 |

**Fonte**: Dados brutos extraídos da amostra do Censo Demográfico do Brasil, 1991 - IBGE

QUADRO 4.5 - CATEGORIAS DE ANOS DE ESTUDO DO PAI, SEGUNDO AS DIVISÕES DO SISTEMA ESCOLAR ANTIGO

| CATEGORIA | NÍVEL DE ESCOLARIDADE | ANOS DE ESTUDO |
|---|---|---|
| 1 | Sem instrução ou primário incompleto | 0 - 3 |
| 2 | Primário completo | 4 |
| 3 | Ginásio incompleto | 5 - 7 |
| 4 | Ginásio completo | 8 |
| 5 | Pelo menos colegial incompleto | > 8 |

QUADRO 4.6 - CATEGORIAS DE ANOS DE ESTUDO DOS FILHOS, SEGUNDO A DIVISÃO DO SISTEMA ESCOLAR VIGENTE

| CATEGORIA | GRAU DE ESCOLARIDADE | ANOS DE ESTUDO |
|---|---|---|
| 1 | Nenhum grau de instrução completo | 0 - 7 |
| 2 | 1º grau completo | 8 |
| 3 | Pelo menos 2º grau incompleto | > 8 |

filhos. A escolaridade do pai foi classificada em 5 categorias e a escolaridade dos filhos em 3 categorias, conforme Quadros 4.5 e 4.6, respectivamente, em que cada categoria corresponde a uma fase do período escolar. Apesar dessas classificações serem bastante gerais, são suficientes para a ilustração desejada [9].

A Tabela 4.2 resume os dados da Tabela 4.1 classificando a escolaridade do pai em dois níveis de instrução. Ao analisar a escolaridade do 1º filho (irmão mais velho), observa-se na Tabela 4.2 e no Gráfico 4.1 que, quando o pai tinha baixo nível de instrução, 59,3% dos irmãos mais velhos estudaram menos de 8 anos, 27,0% tinham mais de 8 anos de estudo e somente 13,7% tinham concluído o 1º grau mas não ingressaram no 2º grau. Em contraste, verifica-se que entre as famílias em que o pai estudou mais de 4 anos, a maioria (75, 9%) dos irmãos mais velhos estudaram mais de 8 anos, 12,4% tinham menos de 8 anos de estudo e 11,7% tinham estudado até concluirem o 1º grau. As distribuições

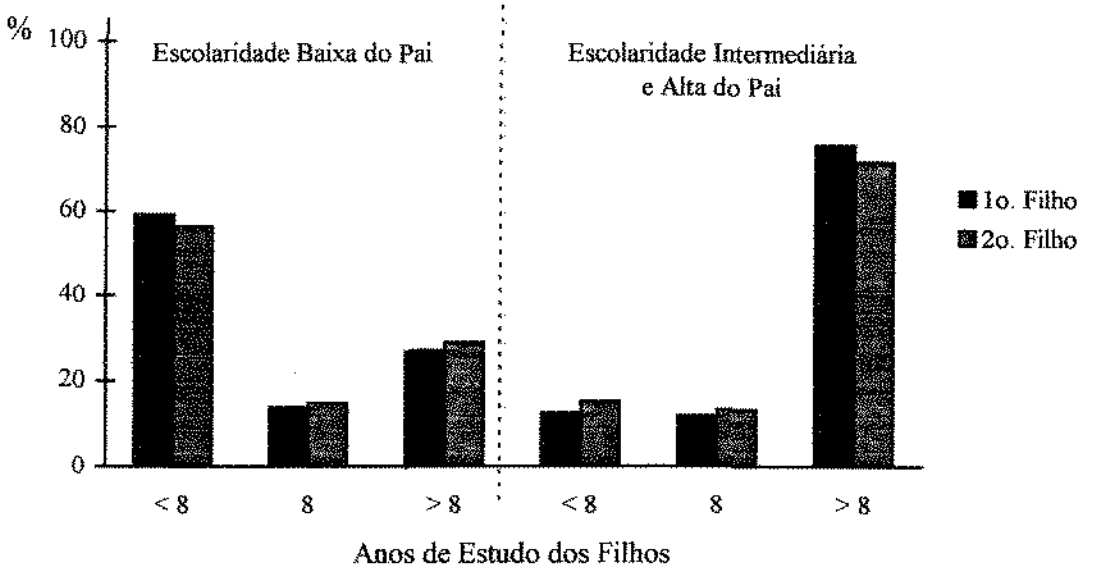
[9]Se fosse considerado um número maior de categorias, o problema de freqüências muito baixas ocorreria em certas categorias, o que prejudicaria a análise estatística.

TABELA 4.2 - DISTRIBUIÇÃO DE ESCOLARIDADE DOS FILHOS E DO PAI EM FREQÜÊNCIAS E PORCENTAGENS

| NÍVEL DE INSTRUÇÃO DO PAI | ANOS DE ESTUDO DO 2º FILHO | ANOS DE ESTUDO DO 1º FILHO | | | TOTAL |
|---|---|---|---|---|---|
| | | < 8 | 8 | > 8 | |
| Baixo (0 - 4 anos) | < 8 | 3032 | 337 | 372 | 3741 (56,4) |
| | 8 | 392 | 311 | 268 | 971 (14,6) |
| | > 8 | 508 | 258 | 1153 | 1919 (28,9) |
| | Total | 3952 (59,3) | 906 (13,7) | 1793 (27,0) | 6631 (100,0) |
| Intermediário e Alto (5 anos ou mais) | < 8 | 134 | 51 | 113 | 298 (15,0) |
| | 8 | 35 | 87 | 138 | 260 (13,3) |
| | > 8 | 78 | 95 | 1258 | 1431 (71,9) |
| | Total | 247 (12,4) | 233 (11,7) | 1509 (75,9) | 1989 (100,0) |
| TOTAL | < 8 | 3166 | 388 | 485 | 4039 (46,9) |
| | 8 | 427 | 398 | 406 | 1231 (14,3) |
| | > 8 | 586 | 353 | 2411 | 3350 (38,9) |
| | Total | 4179 (13,2) | 1139 (13,2) | 3302 (38,3) | 8620 (100,0) |

**Fonte**: Tabela 5.1
**Nota**: Números entre parênteses indicam porcentagens

GRÁFICO 4.1 - DISTRIBUIÇÃO DE ESCOLARIDADE DOS FILHOS E DO PAI EM PORCENTAGENS




de escolaridade do 2º filho (irmão mais novo) em cada um dos dois níveis de escolaridade do pai podem ser descritas de forma análoga à que foi feita para o irmão mais velho, verificando-se as mesmas tendências. Portanto, há uma indicação de que filhos com pai de baixa escolaridade tendem a estudar menos do que quando o pai tem mais de 4 anos de estudo. Além disso, nas famílias em que o pai tinha baixa escolaridade, o irmão mais novo teve uma educação mais privilegiada que o outro irmão, como mostram a Tabela 4.2 e o Gráfico 4.1. Comparando-se as porcentagens de filhos com menos de 8 anos de estudo, verifica-se que a menor porcentagem está entre os mais novos (56, 4%), contra 59,3% dos irmãos mais velhos. Por outro lado, verifica-se que entre os irmãos mais novos 28,9% tinham mais de 8 anos de estudo, um pouco maior que os 27,0% entre os irmãos mais velhos. Quando o pai tinha um maior nível de instrução verifica-se o inverso: o 1º filho aparece numa situação mais favorecida. Para explorar melhor os efeitos da escolaridade do pai sobre a escolaridade dos filhos pode-se estimar os parâmetros do modelo logito em (4.13), usando os dados de progressão escolar da Tabela 4.1.

### 4.4.2 Um Modelo Logito para Transições Escolares de Irmãos

Dentro de cada um dos 5 níveis de escolaridade do pai na Tabela 4.1, a distribuição conjunta de escolaridade dos dois irmãos é uma distribuição de escolaridade pareada discreta com três categorias em cada marginal. Os três níveis observados de escolaridade dos filhos permitem analisar duas transições escolares (ver Quadro 4.6),

- Primeira transição: concluir o 1º grau (passar da categoria 1 para a categoria 2) e
- Segunda transição: completar pelo menos 1 ano do 2º grau, dado que concluiu o 1º grau (passar da categoria 2 para a categoria 3).

O principal interesse é verificar se os efeitos da escolaridade do pai variam entre as transições escolares dos filhos e como variam, modelando os dados da Tabela 4.1. Para tanto, o modelo logito a ser considerado é dado em (4.13) para apenas duas transições ($N = 2$). O nível de escolaridade do pai na família $j$ é idêntica para cada filho $i$ e para cada transição t, portanto $\mathbf{X}_{tij} = \mathbf{X}_j$. O vetor de covariáveis foi introduzido no modelo logito como uma variável *dummy* (ver Quadro 4.7), em que $\mathbf{X}_j$ é um vetor $(4 \times 1)$ de quatro variáveis dicotômicas representadas por $X_{jk}$, que é igual a 1 se a escolaridade do pai da família $j$ pertence a categoria $k+1$ e 0 caso contrário ($k = 1, 2, 3, 4$). O vetor $\mathbf{X}_j = \mathbf{0}$ indica que o pai da família $j$ tem menos de 4 anos de estudo. Os parâmetros $\boldsymbol{\beta}_1$ e $\boldsymbol{\beta}_2$ são os vetores $(4 \times 1)$ dos efeitos categóricos da escolaridade do pai sobre o logito de evasão escolar na primeira e segunda transição, respectivamente. A variável latente $W_j$, incluída no modelo para representar a heterogeneidade não-observada compartilhada entre os irmãos, é igual a 0 se a família $j$ pertence a classe latente 1 e igual a 1 se a família $j$ pertence a classe latente 2.

Com estas especificações o modelo logito em (4.13) reduz-se nas equações (4.15) a (4.18) e o modelo log-linear correspondente é dado em (4.28) com $x_{jk} = 0, 1$ e $p = 4$.

QUADRO 4.7 - VALORES DA COVARIÁVEL ESCOLARIDADE DO PAI

| ANOS DE ESTUDO DO PAI | 0 - 3 | 4 | 5 - 7 | 8 | > 8 |
|---|---|---|---|---|---|
| $X_i$ | $\begin{pmatrix}0\\0\\0\\0\end{pmatrix}$ | $\begin{pmatrix}1\\0\\0\\0\end{pmatrix}$ | $\begin{pmatrix}0\\1\\0\\0\end{pmatrix}$ | $\begin{pmatrix}0\\0\\1\\0\end{pmatrix}$ | $\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$ |

TABELA 4.3 - ESTATÍSTICAS DE QUALIDADE DE AJUSTE E GRAUS DE LIBERDADE PARA MODELOS SELECIONADOS

| MODELO LOGITO | RAZÃO DE VEROSSIMILHANÇA | GRAUS DE LIBERDADE |
|---|---|---|
| 1- Sem Heterogeneidade Não-Observada | | |
| 1.1- Efeitos Proporcionais | 2510 | 32 |
| 1.2- Efeitos Não-Proporcionais | 2168 | 28 |
| 2- Com Heterogeneidade Não-Observada | | |
| 2.1- Efeitos Proporcionais | 412 | 30 |
| 2.2- Efeitos Não-Proporcionais | 377 | 26 |

**Fonte**: Saídas do programa Dnewton

### 4.4.3 Resultados Empíricos

A Tabela 4.3 apresenta a razão de verossimilhança para várias especificações do modelo de Mare dado por (4.15) até (4.18) ajustados para os dados da Tabela 4.1. Os modelos incorporam suposições alternativas acerca da proporcionalidade dos efeitos da escolaridade do pai sobre a escolaridade dos filhos e acerca da heterogeneidade não-observada. Todos esses 4 modelos assumem que os efeitos de escolaridade do pai são os mesmos para ambos os filhos. Os Modelos 1.1 e 1.2 não assumem heterogeneidade não-observada, isto é, a distribuição da escolaridade do 1º e 2º filhos são condicionalmente independentes, dada a mútua dependência com a escolaridade do pai. Sob estes modelos, $\lambda = 0$ em (4.15) a (4.18). O Modelo 1.1 assume efeitos iguais de escolaridade do pai na 1ª e 2ª transições, $\beta_1 = \beta_2$ (efeitos proporcionais) e o Modelo 1.2 relaxa essa suposição, $\beta_1 \neq \beta_2$ (efeitos não-proporcionais). A grande diferença da razão de verossimilhança entre os Modelos 1.1 e 1.2 sugere fortemente que os efeitos de escolaridade do pai variam entre as transições.

Os Modelos 2.1 e 2.2 introduzem o efeito dos determinantes não-observados de escolaridade que são compartilhados entre os irmãos por meio da variável latente dicotômica

$W_j$. Estes modelos, que correspondem às equações em (4.15) a (4.18) com $\lambda \neq 0$, são estimados por máxima verossimilhança e incluem dois parâmetros a mais que os correspondentes modelos sem heterogeneidade não-observada: um para a distribuição da heterogeneidade não-observada e outro para o efeito de $W_j$ sobre o logito de evasão escolar. Os Modelos 2.1 e 2.2 se diferenciam com relação à especificação de proporcionalidade dos efeitos das covariáveis observadas em (4.15) a (4.18): o Modelo 2.1 supõe efeitos proporcionais ($\beta_1 = \beta_2$) e o Modelo 2.2 não ($\beta_1 \neq \beta_2$). Por incorporar aspectos da associação entre escolaridade de irmãos, estes modelos ajustam os dados muito melhor que os modelos sem heterogeneidade não-observada [10]. Como os Modelos 1.1 e 1.2, os modelos com heterogeneidade não-observada também sugerem que os efeitos de escolaridade do pai sejam não-proporcionais, embora a diferença entre os Modelos 2.1 e 2.2 seja muito menor que a diferença entre 1.1 e 1.2.

A Tabela 4.4 registra as estimativas para os parâmetros dos Modelos 1.2 e 2.2. Cada modelo inclue separadamente interceptos para o 1º e 2º filhos tanto para a 1ª transição como para a 2ª transição. No Modelo 1.2 estes interceptos são estimativas do logito de não fazer a transição dado que o pai tinha de 0 a 3 anos de estudo. Para cada transição o modelo inclue um conjunto de parâmetros para os efeitos de escolaridade do pai, que são as estimativas do logito de não fazer a transição dado que o pai completou um certo nível de escolaridade menos a estimativa do logito de não fazer a transição dado que o pai tinha de 0 a 3 anos de estudo. As estimativas dos parâmetros para ambos os modelos sugerem que os efeitos de escolaridade do pai variam entre as transições escolares (apesar de serem estimados valores diferentes para cada intercepto). Isto é mostrado mais claramente no Gráfico 4.2, em que são plotados os riscos relativos [11] de evasão escolar do 2º filho estimados pelos Modelos 1.2 e 2.2. Nos Modelos 1.2 e 2.2, as estimativas dos riscos relativos são maiores na 2ª transição do que na 1ª transição, indicando que o

[10] Apesar da diferença muito grande na estimativa da razão de verossimilhança entre os Modelos 1.1 e 2.1 e entre os Modelos 1.2 e 2.2, estas comparações de modelos são feitas informalmente. Quando o número de classes latentes difere entre os modelos, a diferença entre as estatísticas de qualidade de ajuste dos modelos hierárquicos não segue uma distribuição assintótica qui-quadrada (Mare, 1994).

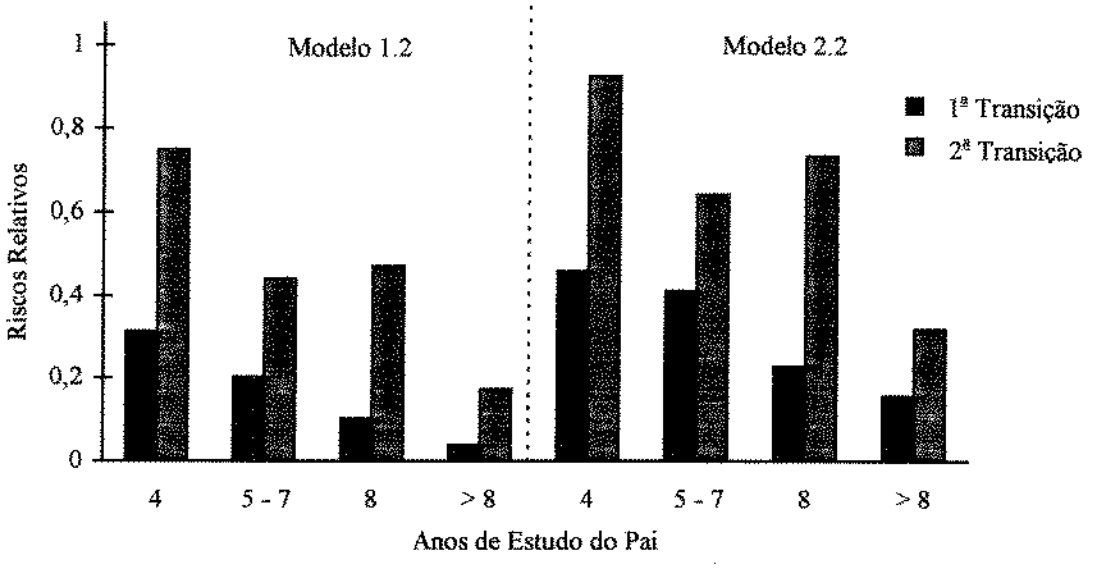
[11] Os riscos relativos plotados no Gráfico 4.2 são medidas relativas aos filhos cujo pai tinha de 0 a 3

TABELA 4.4 - PARÂMETROS ESTIMADOS PELOS MODELOS DE EFEITOS NÃO-PROPORCIONAIS

| PARÂMETROS | | MODELO 1.2 | MODELO 2.2 |
|---|---|---|---|
| Primeira Transição | | | |
| Intercepto | | | |
| | 1º Filho | 0,746 | 2,003 |
| | 2º Filho | 0,712 | 1,947 |
| Efeito de Escolaridade do Pai | | | |
| 4 | (vs. 0 - 3) | -1,154 | -0,777 |
| 5-7 | (vs. 0 - 3) | -1,597 | -0,885 |
| 8 | (vs. 0 - 3) | -2,288 | -1,474 |
| >8 | (vs. 0 - 3) | -3,237 | -1,858 |
| Segunda Transição | | | |
| Intercepto | | | |
| | 1º Filho | -0,487 | 1,776 |
| | 2º Filho | -0,513 | 1,780 |
| Efeito de Escolaridade do Pai | | | |
| 4 | (vs. 0 - 3) | -0,288 | -0,076 |
| 5-7 | (vs. 0 - 3) | -0,819 | -0,441 |
| 8 | (vs. 0 - 3) | -0,783 | -0,305 |
| >8 | (vs. 0 - 3) | -1,761 | -1,135 |
| Heterogeneidade Não-Observada | | | |
| Logito de $Y_{tij} = 0$ na classe 2 | (vs. classe 1) | .. | 6,272 |
| Efeito da classe 2 | (vs. classe 1) | .. | -3,050 |

**Fonte**: Saídas do programa Dnewton
**Nota**: Sinal convencional utilizado:
.. dado numérico não se aplica

GRÁFICO 4.2 - RISCOS RELATIVOS DE EVASÃO ESCOLAR DO 2o. FILHO POR ESCOLARIDADE DO PAI ESTIMADOS PELOS MODELOS NÃO-PROPORCIONAIS




efeito de escolaridade do pai enfraquece [12] entre a primeira e segunda transições. Mas os riscos relativos de evasão escolar estimados pelo Modelo 1.2, são bem menores que as correspondentes estimativas no Modelo 2.2. Isto sugere que os efeitos de escolaridade do pai sobre as transições escolares do 2º filho estimados pelo Modelo 1.2 são explicados em grande parte pela heterogeneidade não-observada em nível familiar.

A Tabela 4.4 também registra parâmetros para heterogeneidade não-observada do Modelo 2.2. Neste modelo de duas classes latentes misturadas, um dos parâmetros descreve o logito dos dois irmãos fazerem as duas transições ($Y_{tij} = 0$ para todo t, $i = 1, 2$), dado que a família $j$ pertence à classe latente 2 (versus classe 1) e o outro é o parâmetro para o efeito da classe latente 2 (versus classe 1) sobre o logito de evasão escolar. Sob o

---

anos de estudo, dadas por

$$\frac{\exp\{l(Y_{t2j} = 1|\mathbf{X}_j = \mathbf{x}_j, W_j = w_j)\}}{\exp\{l(Y_{t2j} = 1|\mathbf{X}_j = \mathbf{0}, W_j = w_j)\}} = \exp\{\beta_{tk}\} \ ,$$

em que $\beta_{tk}$ é o efeito da categoria $k+1$ de escolaridade do pai sobre o logito de não fazer a transição $t$, versus categoria 1 ($k = 1, 2, 3, 4$).

[12] O efeito $\beta_{tk}$ é negativo para todo $t$ e $k$ em ambos os modelos não-proporcionais (ver Tabela 4.4), e quanto mais negativo for esse efeito, mais próximo de 0 (zero) estará o risco relativo; tanto que um efeito fortemente negativo deverá resultar em um risco relativo de evasão escolar quase nulo.

Modelo 2.2, cada risco relativo condicionado [13] de evasão escolar estimado para famílias da classe latente 1 é cerca de $21 = \exp(3,05)$ vezes o correpondente para famílias da classe latente 2. Portanto os irmãos da classe latente 2 teriam uma educação mais privilegiada do que os da classe latente 1.

Os modelos de classes latentes provêm um modo direto de estimar a distribuição da heterogeneidade não-observada. Para ilustrar o quanto o modelo de duas classes latentes pode revelar sobre como a heterogeneidade não-observada afeta os parâmetros dos modelos de risco, considere a distribuição da heterogeneidade não-observada nos níveis de escolaridade do pai e do 2º filho, estimada pelo Modelo 2.2. Estas distribuições são mostradas no Gráfico 4.3 em que se plota as porcentagens estimadas de famílias na classe latente 2 (versus classe 1) por nível de escolaridade do pai e do 2º filho. Essas porcentagens derivam das freqüências esperadas do Modelo 2.2. No nível de escolaridade mais baixo do 2º filho, verifica-se que o fato da família ser membro ou não da classe latente 2 estava altamente associado à escolaridade do pai. Quando o pai tinha menos do que 4 anos de estudo somente 12,1% das famílias pertenciam à classe latente 2, enquanto que nas famílias cujo pai tinha mais de 8 anos de estudo 63,4 % estavam na classe latente 2. A medida que a escolaridade do 2º filho aumenta, verifica-se um enfraquecimento substancial na associação entre heterogeneidade não-observada e escolaridade do pai, ao ponto da distribuição da heterogeneidade não-observada ser essencialmente uniforme entre os níveis de escolaridade do pai, quando se considera o nível mais alto de escolaridade do 2º filho, em que a porcentagem mais baixa de famílias pertencentes à classe latente 2 (94,0%) está no nível inferior de escolaridade do pai e a porcentagem mais alta (98,9%) encontra-se entre os pais que tinham estudado mais de 8 anos.

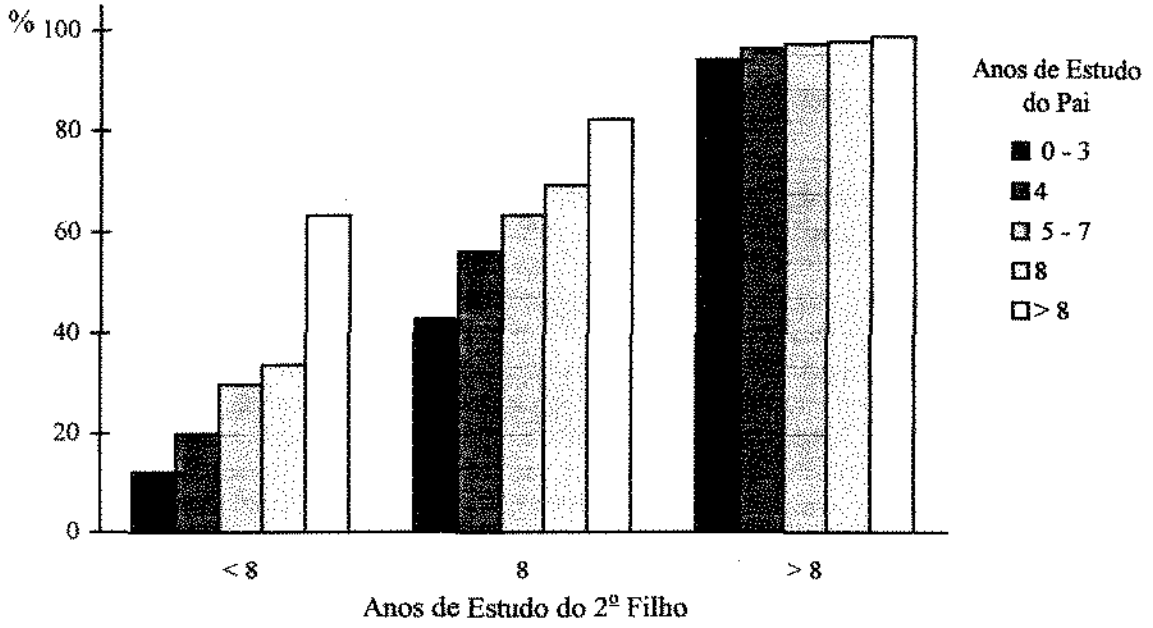
A alteração na distribuição da heterogeneidade não-observada entre os níveis de esco-

[13] Em inglês, odds. O risco relativo condicionado de uma variável dicotômica $U$ é dado por

$$\frac{P(U=u)}{1-P(U=u)}$$

e é por isso que em inglês, a transformação logito é denominada log-odds.

GRÁFICO 4.3 - PORCENTAGEM DE FAMÍLIAS NA CLASSE LATENTE 2 POR ESCOLARIDADE DO 2º FILHO E DO PAI




laridade do pai, na medida em que se progride na escolaridade do 2º filho, deve ser uma causa importante para surgir efeitos maiores de escolaridade do pai sobre as transições em modelos que não levam em consideração a heterogeneidade não-observada. Essa ilustração mostra a importância de incluir-se, nos modelos para análise de dados da história de eventos, um termo que represente a heterogeneidade não-observada.

# Capítulo 5

# Considerações Finais

O tema da dissertação foi motivado por uma discussão recente sobre o impacto da heterogeneidade não-observada na análise da história de eventos, iniciada com o trabalho de Vaupel, Maton e Stallard (1979) e ampliada nos anos 80 e 90.

A parte inicial da dissertação consistiu numa exaustiva revisão bibliográfica de duas teorias que dão base para o desenvolvimento do tema: análise da história de eventos e mistura de distribuições, incluída no Capítulo 1 e início do Capítulo 3.

Alguns modelos de fragilidade foram apresentados e analisados nos Capítulos 2 e 3. Ao avaliá-los, duas questões podem ser consideradas acerca da modelagem dos efeitos da heterogeneidade não-observada. A primeira é sobre a identificabilidade dos modelos com heterogeneidade não-observada. As funções de risco e de distribuição da heterogeneidade não-observada são claramente não-identificáveis quando os dados da história de eventos são sobre tempos univariados. Como foi mostrado no Capítulo 2, para análises de eventos não-repetidos de indivíduos isolados há um modelo sem heterogeneidade não-observada e uma infinidade de outros modelos com diferentes distribuições da heterogeneidade não-observada — e conseqüentemente, com diferentes funções de risco padrão — que se ajustam aos dados observados identicamente. Mas esses modelos têm implicações comportamentais muito diferentes. A escolha entre esses modelos é feita somente com base em conhecimento externo sobre a forma funcional da função de risco ou sobre a

distribuição da heterogeneidade não-observada. Portanto, o grau de confiança das estimativas dos parâmetros depende diretamente da própria confiança nessas informações subjacentes. As teorias existentes provêm pouca orientação na escolha entre modelos alternativos igualmente bem ajustados. Esta argumentação, no entanto, não deve ser usada como justificativa para ignorar a heterogeneidade não-observada. Quando os tempos de ocorrência do evento estão agrupados de alguma forma em pares ou em grupos de maior tamanho, a situação de identificabilidade dos modelos de risco com heterogeneidade não-observada é bem diferente. No Capítulo 3 foi discutido que se a heterogeneidade não-observada for compartilhada pelo grupo, os modelos de risco podem incluir um termo que represente a heterogeneidade não-observada a fim de buscar um "melhor" ajuste para os dados.

A segunda questão é sobre o cuidado que se deve ter quanto à qualidade de ajuste de um modelo aos dados. Não se pode simplesmente comparar a verossimilhança de um modelo com a verossimilhança de outro modelo a fim de tirar-se conclusões sobre a qualidade de ajuste dos modelos. No caso dos modelos de risco univariado as ferramentas de instabilidade de Heckman-Singer e Trussel-Richards, apresentadas no Capítulo 2, ilustram o fato de que um analista pode atingir nada mais que uma falsa sensação de segurança quando examina um primeiro modelo com somente covariáveis observadas e um segundo modelo com covariáveis observadas e uma variável não-observada, para concluir que o segundo modelo permite melhor inferência porque controla a heterogeneidade não-observada. Os resultados de instabilidade não permitem ir muito além disso, porque testes para qualidade de ajuste dos modelos poderiam revelar que nenhum dos modelos ajustam-se aos dados. Alternativamente, os testes poderiam revelar que vários modelos ajustam-se igualmente bem aos dados, embora os parâmetros estimados sejam diferentes e somente informação externa, nem sempre disponível para o analista, poderia ser usada para distinguir entre esses modelos alternativos. Quanto aos modelos de risco multivariado apresentados no Capítulo 3, estudos feitos por meio de simulações, similares àqueles referidos no Capítulo 2 para o caso univariado, devem revelar algo sobre a insta-

bilidade das estimativas dos parâmetros dos modelos multivariados com heterogeneidade não-observada.

O Capítulo 4 foi inspirado no trabalho inovador de Mare (1994) que analisa dados categóricos da história de eventos por meio de variantes do modelo logito. O estudo deste artigo permitiu que fossem aprofundados outros aspectos da análise de dados categóricos como a análise de tabelas observadas parcialmente e o modelo de classes latentes. A proposta alternativa de Mare para estimar o modelo logito pareado, vista no Capítulo 4, pode ser aplicada quando há interesse em verificar o efeito da heterogeneidade não-observada sobre as transições entre tempos de ocorrência categorizados. Se a heterogeneidade não-observada é compartilhada pelo par, a metodologia de Mare elimina a necessidade de fazer fortes suposições paramétricas e traz uma relação simples para explicar a associação entre covariáveis categóricas e a heterogeneidade não-observada, levando grande vantagem sobre os modelos de fragilidade compartilhada nestes pontos. Mare (1994) mensiona que uma outra vantagem do modelo logito pareado é que a qualidade de ajuste do modelo pode ser avaliada verificando os ajustes de outros modelos para dados categóricos que podem, conjuntamente, indicar possíveis causas de falta de ajuste.

Para concluir considera-se importante destacar os comentários feitos por Trussel (1992) em que afirma haver duas razões coagindo para se querer modelar heterogeneidade não-observada explicitamente, mesmo que os tempos de ocorrência sejam univariados. Uma razão é que em quase todas as ocasiões dificilmente alguém poderia estar confiante em que todos os fatores relacionados com um resultado de interesse estão de fato incluídos no modelo estatístico. Partindo desse suposto, é tentador adicionar um termo que capte os efeitos desses fatores omitidos. No entanto, está claro que esta alternativa não é uma solução do problema de variáveis omitidas porque esses modelos podem captar somente aquelas fontes de variação que são independentes das covariáveis incluídas. Pode-se, sim, ir mais longe, construindo um modelo no qual os determinantes não-observados e os observados são dependentes conjuntamente, mas a forma desta dependência teria de ser assumida, como sugerido nas Seções 2.4 e 3.3. A outra razão é que um modelo com

heterogeneidade não-observada poderia ser mais parcimonioso que um outro sem. Se o objetivo for descrever o comportamento e não estabelecer relações de causa, então um modelo mais simples deveria ser cogitado. Mas não se pode simplesmente modelar a heterogeneidade não-observada explicitamente e declarar que se tem uma melhor explicação para os dados. Em vez disso, um pesquisador cuidadoso deveria sempre estar ciente de que diferentes modelos de risco univariado poderiam gerar os mesmos resultados observados ou que problemas de instabilidade poderiam estar distorcendo as estimativas dos parâmetros.

# Referências Bibliográficas


[1] AALEN, O.O. (1987) Two examples of modelling heterogeneity in survival analysis. *Scandinavian Journal of Statistics* **14**:19-25.

[2] _____ (1988) Heterogeneity in survival analysis. *Statistics in Medicine* **7**:1121-37.

[3] AGRESTI, A. (1990) *Categorical Data Analysis*. New York, J. Wiley.

[4] ALLISON, P.D. (1982) Discrete-Time methods for the analysis of event histories. In: LEINHARTDT, S. (Ed.) *Sociological Methodology*. San Francisco, Jossey-Bass. pp.61-98.

[5] _____ (1984) *Event History Analysis* : Regression for Longitudinal Event Data. Bervely Hills, Sage.

[6] BLOSSFELD, H.-P.; HAMERLE, A. (1992) Unobserved heterogeneity in event history models. *Quality and Quantity* **26**:157-68.

[7] _____; _____; MAYER, K.U. (1989) *Event History Analysis* : Statistical Theory and Application in the Social Sciences. Hillsdate - New Jersey, Lawrence Erlbaum.

[8] CLAYTON, D.G. (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chromic disease incidence. *Biometrika* **65**:141-51.

[9] COX, D.R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society* B, **34**:187-202.

[10] CROWDER, M.J. et al. (1990) *Statistical Analysis of Reliability Data.* London, Chapman and Hall.

[11] DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* B **38**:1-38.

[12] EVERITT, B.S. (1984) *An Introduction to Latent Variable Models.* New York, Chapman and Hall.

[13] GUO, G. (1993) Use of sibling data to estimate family mortality effects in Guatemala. *Demography* **30**:15-32.

[14] HABERMAN, S.J. (1988) A stabilized Newton-Raphson algorithm for log-linear models for frequency tables derived by indirect observation. In: CLOGG, C.C. (Ed.) *Sociological Methodology.* Washington, American Sociological Association. pp. 193-211.

[15] HECKMAN, J.; SINGER, B. (1982) Population heterogeneity in demographic models. In: LAND, K.; ROGERS, A. (Ed.) *Multidimensional Mathematical Demography.* New York, Academic Press. pp. 567- 599.

[16] HOEM, J.M. (1990) Identifiability in hazards models with unobserved heterogeneity: The compatibility of two apparently contradictory results. *Theoretical Population Biology* **37**:124-28.

[17] HOOUGARD, P. (1984) Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* **71**:75-83.

[18] _____ (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**:387-96.

[19] IBGE (1994) *Sistema de Informática Geográfica.* 1 disco compacto (cd-rom). Rio de Janeiro.

[20] _____ (1996) *Censo Demográfico 1991* : Características da População e Instrução. v.1. Rio de Janeiro.

[21] JOHNSON, N.L.; KOTZ, S. (1969) *Distributions in Statistics* : Discrete Distributions. New York, Houghton Mifflin.

[22] _____; _____ (1970) *Distributions in Statistics* : Continuous Univariate Distributions 1. New York, Houghton Mifflin.

[23] KALBFLEISCH, J.D.; PRENTICE, R.L. (1980) *The Statistical Analysis of Failure Time Data.* New York, J. Wiley.

[24] KOTZ, S.; JOHNSON, N.L. (Ed. in Chief) (1985) *Encyclopedia of Statistical Sciences* v.5. New York, J.Wiley.

[25] LAIRD, N. (1978) Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**:805-11.

[26] LAWLESS, J.F. (1982) *Statistical Models and Methods for Lifetime Data.* New York, J. Wiley.

[27] LEE, E.T. (1980) *Statistical Methods for Survival Data Analysis.* Belmont, Wadsworth.

[28] LINDSAY, D., CLOGG, C.C. e GREGO, J. (1991) Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* **86**:96-107.

[29] MANTON, K.G.; SINGER, B.; WOODBURY, M.A. (1992) Some issues in the quantitative characterization of heterogeneous populations. In: TRUSSEL, J.; HANKINSON, R.; TILTON, J. (Ed.) *Demographic Applications of Event History Analysis.* New York, Oxford University Press. pp. 9-37.

[30] MARE, R.D. (1994) Discrete-Time bivariate hazards with unobserved heterogeneity: a partially observed contingency table approach. In: MARSDEN, P.T. (Ed.) *Sociological Methodology.* [Washington], American Sociological Association. pp. 341-383.

[31] MONTGOMERY, M.R.; RICHARDS, T.; BRAUN, H.I. (1986) Child health breastfeeding, and survival in Malaysia: a random-effects logit approach. *Journal of the American Statistical Association* **81**:297-309.

[32] NAMBOODIRI, K.; SUCHINDRAN, C.M. (1987) *Life Table Techniques and their Applications.* Orlando, Academic Press.

[33] OAKES, D. (1989) Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**:487-93.

[34] RODRIGUES, G. (1992) *Statistical Demography.* Notas de aula para o programa de Population Studies em Princeton University.

[35] TRUSSEL, J. (1992) Introduction. In: TRUSSEL, J.; HANKINSON, R.; TILTON, J. (Ed.) *Demographic Applications of Event History Analysis.* New York, Oxford University Press. pp.1-7.

[36] _____; RICHARDS, T. (1985) Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. In: TUMA, N. (Ed.) *Sociological Methodology.* San Francisco, Jossey-Bass. pp. 242-276.

[37] _____; RODRIGUES, G. (1990) Heterogeneity in demographic research. In: ADAMS, J. et al. (Ed.) *Convergent Questions in Genetics and Demography.* New York, Oxford University Press. pp. 111-132.

[38] VAUPEL, J.W. (1990) Relatives risks: frailty models of life history data. *Theoretical Population Biology* **37**:220-34.

[39] _____; MANTON, K.G.; STALLARD, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**:439-54.

[40] _____; YASHIN, I.A. (1985) Heterogeneity's ruses: some surprising effects of selection on population dynamics. *American Statistician* **39**:176-85.

# Anexo A

# Alguns Resultados de Interesse

Nas Proposições 1 a 4 considere que $T$ seja uma variável aleatória contínua e não-negativa representando tempo de sobrevivência para um determinado indivíduo e $Z$ seja uma variável aleatória não-negativa representando a fragilidade do indivíduo. E sejam $S(t|z)$ e $S(t)$ as funções de sobrevivência condicionada e não-condicionada de $T$, respectivamente.

**Proposição 1**

O risco não-condicionado de $T$ é dado por

$$\lambda(t) = \int_0^\infty \lambda(t|z)\, dG(z|T \geq t) \ ,$$

em que $\lambda(t|z)$ é o risco condicionado e $G(z|T \geq t)$ é a função de distribuição condicional de $Z$ dado $T \geq t$.

**Demonstração**

Pela relação entre função de risco e função de sobrevivência tem-se que

$$\lambda(t) = -\frac{d}{dt}\log[S(t)] = -\frac{1}{S(t)}\frac{d}{dt}S(t)$$

e como $S(t) = E[S(t|Z)]$, então

$$\lambda(t) = -\frac{1}{S(t)}\frac{d}{dt}\int_0^\infty S(t|z)\,dG(z) \ ,$$

em que $G(z)$ é a função de distribuição de $Z$.

Supondo que as operações de integração e diferenciação possam ser trocadas tem-se que

$$\lambda(t) = -\frac{1}{S(t)}\int_0^\infty \frac{\partial}{\partial t}S(t|z)\,dG(z) \ . \tag{A.1}$$

Desde que

$$S(t|z) = \exp[-\Lambda(t|z)]$$

tem-se que

$$\frac{\partial}{\partial t}S(t|z) = -\lambda(t|z)\exp[-\Lambda(t|z)] = -\lambda(t|z)\,S(t|z) \ , \tag{A.2}$$

em que $\Lambda(t|z)$ é o risco acumulado condicionado de $T$.

Substituindo (A.2) em (A.1) resulta que

$$\lambda(t) = -\frac{1}{S(t)}\int_0^\infty -\lambda(t|z)\,S(t|z)\,dG(z) = \int_0^\infty \lambda(t|z)\,\frac{S(t|z)}{S(t)}dG(z) \ .$$

Resta provar que a seguinte relação é verdadeira:

$$dG(z|T \geq t) = \frac{S(t|z)}{S(t)}dG(z) \ .$$

Pelo Teorema da Probabilidade Total, a relação entre a função de distribuição conjunta e a função de distribuição condicional pode ser escrita como

$$P(T \geq t, Z \leq z) = \int_0^z P(T \geq t|u)\,dG(u)$$

$$\Leftrightarrow \frac{P(T \geq t, Z \leq z)}{P(T \geq t)} = \int_0^z \frac{P(T \geq t|u)}{P(T \geq t)}dG(u)$$

$$\Leftrightarrow G\left(z|T \geq t\right) = \int_0^z \frac{S\left(t|u\right)}{S\left(t\right)} dG\left(u\right)$$

Logo,

$$dG\left(z|T \geq t\right) = \frac{S\left(t|z\right)}{S\left(t\right)} dG\left(z\right) \tag{A.3}$$

e de (A.3) verifica-se que

$$\lambda\left(t\right) = \int_0^\infty \lambda\left(t|z\right) \frac{S\left(t|z\right)}{S\left(t\right)} dG\left(z\right) = \int_0^\infty \lambda\left(t|z\right) dG\left(z|T \geq t\right) \ .$$

**Proposição 2**

A fragilidade média entre os sobreviventes a um determinado tempo $t$ dada por $E\left(Z|T \geq t\right)$ é uma função não-crescente em relação ao tempo.

**Demonstração**

Por definição

$$E\left(Z|T \geq t\right) = \int_0^\infty z\, dG\left(z|T \geq t\right) \ ,$$

em que $G\left(z|T \geq t\right)$ é a função de distribuição de Z dado $T \geq t$.

Usando o resultado em (A.3) e supondo que as operações de integração e diferenciação possam ser trocadas, tem-se que a derivada de $E\left(Z|T \geq t\right)$ em relação a $t$ é dada por

$$\begin{aligned}
\frac{\partial}{\partial t} E\left(Z|T \geq t\right) &= \int_0^\infty \left( \frac{\partial}{\partial t} z \frac{S\left(t|z\right)}{S\left(t\right)} \right) dG(z) \\
&= \int_0^\infty z \left[ \frac{S\left(t\right) \frac{\partial}{\partial t} S\left(t|z\right) - S\left(t|z\right) \frac{\partial}{\partial t} S\left(t\right)}{\left[S\left(t\right)\right]^2} \right] dG\left(z\right) \\
&= \int_0^\infty z \frac{\frac{\partial}{\partial t} S\left(t|z\right)}{S\left(t\right)} dG\left(z\right) - \frac{\frac{\partial}{\partial t} S\left(t\right)}{\left[S\left(t\right)\right]^2} \int_0^\infty z\, S\left(t|z\right)\, dG\left(z\right) \ . \qquad \text{(A.4)}
\end{aligned}$$

Como $S(t) = E[S(t|Z)]$, então, supondo que as operações de integração e derivação possam ser tracadas, tem-se que

$$\frac{\partial}{\partial t}S(t) = \frac{\partial}{\partial t}\int_0^\infty S(t|z)\,dG(z) = \int_0^\infty \frac{\partial}{\partial t}S(t|z)\,dG(z)\ . \tag{A.5}$$

Substituindo (A.2) e (A.5) em (A.4) tem-se que

$$\frac{\partial}{\partial t}E(Z|T \geq t) = \int_0^\infty -z\lambda(t|z)\frac{S(t|z)}{S(t)}dG(z) +$$

$$+\frac{1}{[S(t)]^2}\int_0^\infty \lambda(t|z)\,S(t|z)\,dG(z)\int_0^\infty zS(t|z)\,dG(z)\ .$$

Como $\lambda(t|z) = \lambda_0(t)\,z$, note-se que

$$\frac{\partial}{\partial t}E(Z|T \geq t) = \lambda_0(t)\left\{-\int_0^\infty z^2\frac{S(t|z)}{S(t)}dG(z) + \left[\int_0^\infty z\frac{S(t|z)}{S(t)}dG(z)\right]^2\right\}\ .$$

De (A.3) verifica-se que

$$\begin{aligned}\frac{\partial}{\partial t}E(Z|T \geq t) &= \lambda_0(t)\left\{-\int_0^\infty z^2 dG(z|T \geq t) + \left[\int_0^\infty z dG(z|T \geq t)\right]^2\right\}\\ &= -\lambda_0(t)\left\{E\left(Z^2|T \geq t\right) - [E(Z|T \geq t)]^2\right\}\\ &= -\lambda_0(t)\,Var(Z|T \geq t)\ .\end{aligned}$$

Como $Var(Z|T \geq t) \geq 0$ e $\lambda_0(t) \geq 0$, então

$$\frac{\partial}{\partial t}E(Z|T \geq t) \leq 0\ .$$

**Proposição 3**

No modelo multiplicativo de fragilidade pode-se assumir que $E(Z) = 1$, sendo que o verdadeiro valor esperado da fragilidade é "absorvido" pelo risco padrão.

**Demonstração**

Seja $E(Z) < \infty$ e considere $Z^* = Z/E(Z)$. Note-se que $E(Z^*) = 1$ e de (2.2) tem-se que

$$\lambda(t|z) = \lambda_0(t) \frac{z}{E(Z)} E(Z) = \lambda(t|z^*) = \lambda_0^*(t) z^* ,$$

em que $\lambda_0^*(t) = E(Z) \lambda_0(t)$ é um novo risco padrão e $z^* = z/E(Z)$.

**Proposição 4**

a) Seja $\lambda(t|z)$ o risco condicional de $T$ dado $Z = z$, dado por

$$\lambda(t|z) = a_G(t) z .$$

Definindo

$$A_G(t) = \int_0^t a_G(u) du$$

tem-se que

$$c_G(A_G(t)) = E_G(Z|T \geq t) ;$$

b) Se $a_G(t) = m(t) v_G(M(t))$, então

$$a_G(t) = \frac{m(t)}{c_G[V_G(M(t))]} ,$$

em que $c_G(s)$, $v_G(s)$, $V_G(s)$ e $M(t)$ são funções definidas na Seção 2.8.

**Demonstração de (a)**

Seja $G(z|T \geq t)$ a função de distribuição de $Z$ dado $T \geq t$. Pela definição de esperança condicional e por (A.3),

$$E_G(Z|T \geq t) = \int_0^\infty z dG(z|T \geq t) = \int_0^\infty z \frac{S(t|z)}{S(t)} dG(z) = \frac{1}{S(t)} \int_0^\infty z S(t|z) dG(z) \ .$$

Como

$$S(t|z) = \exp[-z A_G(t)]$$

e

$$S(t) = E_G[S(t|Z)] = \int_0^\infty S(t|z) dG(z) \ ,$$

então,

$$E_G(Z|T \geq t) = \frac{\int_0^\infty z \exp[-z A_G(t)] dG(z)}{\int_0^\infty \exp[-z A_G(t)] dG(z)} = c_G(A_G(t)) \ .$$

**Demonstração de (b)**

Pela aplicação da regra da cadeia de derivadas de funções compostas, tem-se que

$$\frac{d}{dt} C_G(V_G(s)) = c_G(V_G(s)) v_G(s)$$

e como

$$C_G(V_G(s)) = s \Leftrightarrow \frac{d}{ds} C_G(V_G(s)) = 1 \ ,$$

tem-se que

$$c_G(V_G(s)) v_G(s) = 1 \Leftrightarrow v_G(s) = \frac{1}{c_G(V_G(s))} \ .$$

Conclui-se então que

$$a_G(t) = m(t) v_G(M(t)) = \frac{m(t)}{c_G[V_G(M(t))]} \ .$$

**Proposição 5**

A razão $R(\mathbf{t})$, definida em (3.4), pode ser reescrita em termos de $S(\mathbf{t})$, como

$$R(\mathbf{t}) = \frac{[D_1 D_2 S(\mathbf{t})]\, S(\mathbf{t})}{[D_1 S(\mathbf{t})]\, [D_2 S(\mathbf{t})]} .$$

**Demostração**

Desenvolvendo (3.4) tem-se, pela relação entre a função densidade, a função de sobrevivência e a função de risco, que

$$\begin{aligned} R(t_1, t_2) &= \frac{\lambda(t_1|T_2 = t_2)}{\lambda(t_1|T_2 \geq t_2)} = \frac{f(t_1|T_2 = t_2)/S(t_1|T_2 = t_2)}{f(t_1|T_2 \geq t_2)/S(t_1|T_2 \geq t_2)} \\ &= \frac{-\frac{\partial}{\partial t_1} S(t_1|T_2 = t_2)/S(t_1|T_2 = t_2)}{-\frac{\partial}{\partial t_1} S(t_1|T_2 \geq t_2)/S(t_1|T_2 \geq t_2)} . \end{aligned} \tag{A.6}$$

Tem-se por definição que

$$\begin{aligned} S(t_1|T_2 = t_2) &= P(T_1 \geq t_1|T_2 = t_2) = \frac{P(T_1 \geq t_1, T_2 = t_2)}{P(T_2 = t_2)} \\ &= \frac{-\frac{\partial}{\partial t_2}\int_{t_1}^{\infty}\int_{t_2}^{\infty} f(u,v)\,dudv}{-\frac{\partial}{\partial t_2}\int_{0}^{\infty}\int_{t_2}^{\infty} f(u,v)\,dudv} = \frac{-\frac{\partial}{\partial t_2} S(\mathbf{t})}{-\frac{\partial}{\partial t_2} S_2(t_2)} \end{aligned} \tag{A.7}$$

e

$$S(t_1|T_2 \geq t_2) = P(T_1 \geq t_1|T_2 \geq t_2) = \frac{P(T_1 \geq t_1, T_2 \geq t_2)}{P(T_2 \geq t_2)} = \frac{S(\mathbf{t})}{S_2(t_2)} . \tag{A.8}$$

Derivando (A.7) e (A.8) em relação a $t_1$ tem-se, respectivamente, que

$$-\frac{\partial}{\partial t_1} S(t_1|T_2 = t_2) = \frac{-\frac{\partial}{\partial t_1}\left[-\frac{\partial}{\partial t_2} S(\mathbf{t})\right]}{-\frac{\partial}{\partial t_2} S_2(t_2)}$$

e

$$-\frac{\partial}{\partial t_1} S(t_1|T_2 \geq t_2) = \frac{-\frac{\partial}{\partial t_1} S(\mathbf{t})}{S_2(t_2)} .$$

Substituindo (A.7), (A.8) e suas derivadas em (A.6) resulta que

$$R(\mathbf{t}) = \frac{\left[-\frac{\partial}{\partial t_1}\left[-\frac{\partial}{\partial t_2} S(\mathbf{t})\right]\right] S(\mathbf{t})}{\left[-\frac{\partial}{\partial t_1} S(\mathbf{t})\right]\left[-\frac{\partial}{\partial t_2} S(\mathbf{t})\right]} = \frac{[D_1 D_2 S(\mathbf{t})]\, S(\mathbf{t})}{[D_1 S(\mathbf{t})]\,[D_2 S(\mathbf{t})]} .$$

**Proposição 6**

Seja o modelo logito em (4.1) com $\mathbf{t} = 1, 2$ e $\mathbf{X_t} = \mathbf{X}$ [1] representado por

$$l(Y_1 = 1|\mathbf{X} = \mathbf{x}) = \beta_{01} + \boldsymbol{\beta}_1' \mathbf{x}$$

e

$$l(Y_2 = 1|Y_1 = 0, \mathbf{X} = \mathbf{x}) = \beta_{02} + \boldsymbol{\beta}_2' \mathbf{x} .$$

Dado que

$$A_{\mathbf{t}} = \exp(\beta_{0\mathbf{t}} + \boldsymbol{\beta}_{\mathbf{t}}' \mathbf{x}) \qquad \mathbf{t} = 1, 2$$

tem-se que as probabilidades condicionadas das transições são dadas por

$$P(Y_1 = 1|\mathbf{x}) = \frac{A_1}{1 + A_1}$$

e

$$P(Y_2 = 1|Y_1 = 0, \mathbf{x}) = \frac{A_2}{1 + A_2} .$$

[1] Foi omitido o índice $i$, conforme procedimento adotado na Seção 4.1.

**Demonstração**

Por definição da transformação logito tem-se que

$$l\left(Y_1 = 1|\mathbf{X} = \mathbf{x}\right) = \log \frac{P\left(Y_1 = 1|\mathbf{x}\right)}{1 - P\left(Y_1 = 1|\mathbf{x}\right)} = \beta_{01} + \boldsymbol{\beta}_1'\mathbf{x} \tag{A.9}$$

e

$$l\left(Y_2 = 1|Y_1 = 0, \mathbf{X} = \mathbf{x}\right) = \log \frac{P\left(Y_2 = 1Y_1 = 0, |\mathbf{x}\right)}{1 - P\left(Y_2 = 1Y_1 = 0, |\mathbf{x}\right)} = \beta_{02} + \boldsymbol{\beta}_2'\mathbf{x}\ . \tag{A.10}$$

Aplicando a função exponencial em ambos os lados de (A.9) e (A.10), e isolando as probabilidades, tem-se, respectivamente, que

$$P\left(Y_1 = 1|\mathbf{x}\right) = \frac{\exp\left(\beta_{01} + \boldsymbol{\beta}_1'\mathbf{x}\right)}{1 + \exp\left(\beta_{01} + \boldsymbol{\beta}_1'\mathbf{x}\right)} = \frac{A_1}{1 + A_1}$$

e

$$P\left(Y_2 = 1|Y_1 = 0, \mathbf{x}\right) = \frac{\exp\left(\beta_{02} + \boldsymbol{\beta}_2'\mathbf{x}\right)}{1 + \exp\left(\beta_{02} + \boldsymbol{\beta}_2'\mathbf{x}\right)} = \frac{A_2}{1 + A_2}\ .$$

**Proposição 7**

Os parâmetros de (4.10) que envolvem $Y_{1i}$ e $Y_{2i}$ são de fato iguais aos de (4.1), isto é, para $\mathbf{t} = 1, 2$ e $k = 1, 2, \ldots, p$, tem-se

$$\mu_{0\mathbf{t}} = \beta_{0\mathbf{t}}$$

e

$$\mu_{\mathbf{t}k} = \beta_{\mathbf{t}k}\ ,$$

quando $\mathbf{X}_i$ é um vetor de variáveis dummy e em que $\beta_{\mathbf{t}k}$ é uma componente do vetor $\boldsymbol{\beta}_\mathbf{t}$.

**Demonstração**

Usando a definição de probabilidade condicional e assumindo que $Y_{1i}$ e $Y_{2i}$ são condicionalmente independentes dado $\mathbf{X}_i$, tem-se que

$$P\left(Y_{1i}=y_{1i},Y_{2i}=y_{2i},\mathbf{X}_i=\mathbf{x}_i\right)=P\left(Y_{1i}=y_{1i},Y_{2i}=y_{2i}|\mathbf{x}_i\right)P\left(\mathbf{X}_i=\mathbf{x}_i\right)=$$

$$=P\left(Y_{1i}=y_{1i}|\mathbf{x}_i\right)P\left(Y_{2i}=y_{2i}|\mathbf{x}_i\right)P\left(\mathbf{X}_i=\mathbf{x}_i\right)\ . \tag{A.11}$$

Aplicando logaritmo em ambos os lados de (A.11), resulta que

$$\log P\left(Y_{1i}=y_{1i},Y_{2i}=y_{2i},\mathbf{X}_i=\mathbf{x}_i\right)=$$

$$=\log P\left(Y_{1i}=y_{1i}|\mathbf{x}_i\right)+\log P\left(Y_{2i}=y_{2i}|\mathbf{x}_i\right)+\log P\left(\mathbf{X}_i=\mathbf{x}_i\right)\ . \tag{A.12}$$

Então, de (A.12) e (4.10) tem-se que

$$\log\frac{P\left(Y_{1i}=1|\mathbf{x}_i\right)}{P\left(Y_{1i}=0|\mathbf{x}_i\right)}=\log P\left(Y_{1i}=1,Y_{2i}=y_{2i},\mathbf{X}_i=\mathbf{x}_i\right)-$$

$$-\log P\left(Y_{1i}=0,Y_{2i}=y_{2i},\mathbf{X}_i=\mathbf{x}_i\right)=\mu_{01}+\sum_{k=1}^{p}\mu_{1k}x_{ik}\ .$$

Por outro lado, pela definição de distribuição marginal e pelas equações (4.6) a (4.8), tem-se que

$$\log\frac{P\left(Y_{1i}=1|\mathbf{x}_i\right)}{P\left(Y_{1i}=0|\mathbf{x}_i\right)}=\log P\left(Y_{1i}=1|\mathbf{x}_i\right)-$$

$$-\log\left[P\left(Y_{1i}=0,Y_{2i}=1|\mathbf{x}_i\right)+P\left(Y_{1i}=0,Y_{2i}=0|\mathbf{x}_i\right)\right]=$$

$$=\log\left(\frac{A_{1i}}{D_i}+\frac{A_{1i}A_{2i}}{D_i}\right)-\log\left(\frac{A_{2i}}{D_i}+\frac{1}{D_i}\right)=\log A_{1i}=\beta_{01}+\boldsymbol{\beta}_1'\mathbf{x}_i\ ,$$

e portanto,

$$\mu_{01} + \sum_{k=1}^{p} \mu_{1k} x_{ik} = \beta_{01} + \boldsymbol{\beta}_1' \mathbf{x}_i \ .$$

Como $\mathbf{X}$ é uma variável dummy, tem-se que

$$\mathbf{x}_i = \mathbf{0} \Rightarrow \mu_{01} = \beta_{01}$$

e

$$x_{ik} = 1 \text{ e } x_{is} = 0 \Rightarrow \mu_{1k} = \beta_{1k} \ , \quad \text{para } k, s = 1, 2, \ldots, p; (s \neq k)$$

Similarmente, de (A.12) e (4.10) tem-se que

$$\log \frac{P\left(Y_{2i} = 1 | \mathbf{x}_i\right)}{P\left(Y_{2i} = 0 | \mathbf{x}_i\right)} = \log P\left(Y_{1i} = y_{1i}, Y_{2i} = 1, \mathbf{X}_i = \mathbf{x}_i\right) -$$

$$- \log P\left(Y_{1i} = y_{1i}, Y_{2i} = 0, \mathbf{X}_i = \mathbf{x}_i\right) = \mu_{02} + \sum_{k=1}^{p} \mu_{2k} x_{ik} \ ,$$

mas pela definição de distribuição marginal e usando as probabilidades apresentadas na tabela expandida do Quadro 4.1, tem-se que

$$\log \frac{P\left(Y_{2i} = 1 | \mathbf{x}_i\right)}{P\left(Y_{2i} = 0 | \mathbf{x}_i\right)} = \log \left[P\left(Y_{1i} = 1, Y_{2i} = 1 | \mathbf{x}_i\right) + P\left(Y_{1i} = 0, Y_{2i} = 1 | \mathbf{x}_i\right)\right] -$$

$$- \log \left[P\left(Y_{1i} = 1, Y_{2i} = 0 | \mathbf{x}_i\right) + P\left(Y_{1i} = 0, Y_{2i} = 0 | \mathbf{x}_i\right)\right] =$$

$$= \log \left(\frac{A_{1i} A_{2i}}{D_i} + \frac{A_{2i}}{D_i}\right) - \log \left(\frac{A_{1i}}{D_i} + \frac{1}{D_i}\right) = \log A_{2i} = \beta_{02} + \boldsymbol{\beta}_2' \mathbf{x}_i \ ,$$

e portanto,

$$\mu_{02} + \sum_{k=1}^{p} \mu_{2k} x_{ik} = \beta_{02} + \boldsymbol{\beta}_2' \mathbf{x}_i \ .$$

Como $\mathbf{X}$ é uma variável dummy, conclui-se que $\mu_{02} = \beta_{02}$ e $\mu_{2k} = \beta_{2k}$, para $k = 1, 2, \ldots, p$.

# Anexo B

# Listagem do Programa Dnewton

A estimação dos modelos propostos por Mare (1994) requer o uso de um programa específico desenvolvido por Haberman (1988) em Fortran77 e não publicado. Neste anexo é apresentada uma descrição do programa com esquemas de entrada de dados.

O Programa Dnewton considera um tabela-base com freqüência $n(i,j)$ na casela da linha $i$ e coluna $j(i=1,2,...,r,)(j=1,2,...,s(i))$. Cada linha é uma amostra multinomial independente. Um modelo log-linear é usado para os valores esperados $m(i,j)$. Tem-se que

$$\log\left[\frac{m(i,j)}{z(i,j)}\right] = a(j) + b(1)\,x(i,j,1) + b(2)\,x(i,j,2) + \ldots + b(p)\,x(i,j,p)$$

para $z(i,j)$ e $x(i,j,k)$ conhecidos e $a(j)$ e $b(k)$ desconhecidos $(k=1,2,\ldots,p)$.

Como nem todas as freqüências $n(i,j)$ são observadas, utiliza-se $N(h)$ para representar o número de indivíduos nas colunas $j(h,s)$ da linha $g(h)$ com $h=1,2,\ldots q$ e $s=1,2,\ldots,m(h)$. Se $m(h)>1$ o número de indivíduos na coluna $j(h,s)$ não é conhecido.

As entradas do programa são descritas a seguir.

1. A primeira linha de entrada é do título dado ao problema. A leitura é em formato 20a4.

2. A segunda linha de entrada contém 9 parâmetros:

   O primeiro parâmetro **nr** é o número de caselas na tabela expandida, igual a soma dos $s(i)$ $(i = 1, 2, \ldots r)$.

   O segundo parâmetro **np** é o número de parâmetros do modelo log-linear, igual a $p$.

   O terceiro parâmetros **nc** é o número de amostras independentes (número de combinações das categorias das covariáveis), igual a $r$.

   O quarto parâmetro **n** é o número de caselas na tabela observada, igual a $q$.

   O quinto parâmetro **iz** é zero se $z(i,j) = 1$ para todo $(i,j)$ e é positivo, caso contrário.

   O sexto parâmetros **it** é o número de iterações permitido. Se it $= 0$, um máximo de 20 iterações é usado.

   O sétimo parâmetro **ms** é positivo se os valores esperados da tabela-base forem desejados.

   O oitavo parâmetro **to** é um critério de convergência. O padrão é igual a 0,001.

   O nono parâmetro **icov** é positivo se uma estimativa da covariância assitótica dos parâmetros for desejada.

   A leitura dos parâmetros, via arquivo in1.dat, é em formato livre.

3. A próxima linha de entrada contém as variáveis independentes do modelo para tabela-base. Para $x(i,j,k)$, $k$ varia mais rapidamente e $i$ varia mais devagar. A leitura é por meio do arquivo in2.dat.

4. Se forem requeridos $z(i,j)$ não-triviais, então eles seguem com $j$ variando mais rapidamente.

5. Os nomes das variáveis do modelo seguem em formato a8, contidos no arquivo in3.dat.

6. A próxima linha de entrada lista a posição ordenada dos $n(i,j)$, em formato livre, tanto que as entradas são $1, 1+s(1), 1+s(1)+s(2)$, etc., contidas no arquivo in4.dat.

7. A tabela observada é lida em duas linhas por casela. A primeira linha está em formato livre, com a primeira entrada para a casela $s$ igual a $N(h)$, a segunda entrada é igual a $g(h)$ e a terceira entrada é igual a $m(h)$ ,contidas nos arquivos in5.dat. A segunda linha se refere aos valores de $i$ por casela, contidos no arquivo in6.dat., em formato livre.

8. Após as entradas para a tabela observada estão os valores iniciais de $b(k)$ em formato livre, contidos no arquivo in7.dat.

**Exemplo B.1** - Modelo Logito para Indivíduos Independentes

QUADRO B.1 - Valor de $(i, j, h, s)$ para cada casela da tabela expandida

| | $X'$ | | | | |
|---|---|---|---|---|---|
| $(Y_1, Y_2)$ | (0,0,0,0) | (1,0,0,0) | (0,1,0,0) | (0,0,1,0) | (0,0,0,1) |
| (1,1) | (1,1,1,1) | (1,2,4,1) | (1,3,7,1) | (1,4,10,1) | (1,4,13,1) |
| (1,0) | (2,1,1,2) | (2,2,4,2) | (2,3,7,2) | (2,4,10,2) | (2,4,13,2) |
| (0,1) | (3,1,2,1) | (3,2,5,1) | (3,3,8,1) | (3,4,11,1) | (3,4,14,1) |
| (0,0) | (4,1,3,1) | (4,2,6,1) | (4,3,9,1) | (4,4,12,1) | (4,4,15,1) |

Tem-se que

$$r = 5 \quad \rightarrow \quad \boxed{\text{nc} = 5}$$

$$s(i) = 4\,; \quad i = 1, 2, \ldots, 5 \quad \rightarrow \quad \boxed{\text{nr} = \sum_{i=1}^{5} s(i) = \sum_{i=1}^{5} 4 = 20}$$

$$p = 6 \quad \rightarrow \quad \boxed{\text{np} = 6}$$

$$q = 15 \quad \rightarrow \quad \boxed{\text{n} = 15}$$

**Exemplo B.2** - Modelo Logito para Indivíduos Pareados com heterogeneidade não-observada.

QUADRO B.2 - Valor de $(i, j, h, s)$ para cada casela da tabela expandida, dado $w$

| | $X$ | | | | |
|---|---|---|---|---|---|
| $(Y_{11}, Y_{21}, Y_{12}, Y_{22})\|w$ | (0,0,0,0) | (1,0,0,0) | (0,1,0,0) | (0,0,1,0) | (0,0,0,1) |
| (1,1,1,1)\|$w=0$ | (1,1,1,1) | (1,2,10,1) | (1,3,19,1) | (1,4,28,1) | (1,5,37,1) |
| (1,1,1,1)\|$w=1$ | (2,1,1,2) | (2,2,10,2) | (2,3,19,2) | (2,4,28,2) | (2,5,37,2) |
| (1,1,1,0)\|$w=0$ | (3,1,1,3) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,1,1,0)\|$w=1$ | (4,1,1,4) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,1,1)\|$w=0$ | (5,1,1,5) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,1,1)\|$w=1$ | (6,1,1,6) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,1,0)\|$w=0$ | (7,1,1,7) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,1,0)\|$w=1$ | (8,1,1,8) | (8,2,10,8) | (8,3,19,8) | (8,4,28,8) | (8,5,37,8) |
| (1,1,0,1)\|$w=0$ | (9,1,2,1) | (9,2,11,1) | (9,3,20,1) | (9,4,29,1) | (9,5,38,1) |
| (1,1,0,1)\|$w=1$ | (10,1,2,2) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,0,1)\|$w=0$ | (11,1,2,3) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,0,1)\|$w=1$ | (12,1,2,4) | (12,2,11,4) | (12,3,20,4) | (12,4,29,4) | (12,5,38,4) |
| (1,1,0,0)\|$w=0$ | (13,1,3,1) | (13,2,12,1) | (13,3,21,1) | (13,4,30,1) | (13,5,39,1) |
| (1,1,0,0)\|$w=1$ | (14,1,3,2) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,0,0)\|$w=0$ | (15,1,3,3) | ⋮ | ⋮ | ⋮ | ⋮ |
| (1,0,0,0)\|$w=1$ | (16,1,3,4) | (16,2,12,4) | (16,3,21,4) | (16,4,30,4) | (16,5,39,4) |
| (0,1,1,1)\|$w=0$ | (17,1,4,1) | (17,2,13,1) | (17,3,22,1) | (17,4,31,1) | (17,5,40,1) |
| (0,1,1,1)\|$w=1$ | (18,1,4,2) | ⋮ | ⋮ | ⋮ | ⋮ |
| (0,1,1,0)\|$w=0$ | (19,1,4,3) | ⋮ | ⋮ | ⋮ | ⋮ |
| (0,1,1,0)\|$w=1$ | (20,1,4,4) | (20,2,13,4) | (20,3,22,4) | (20,4,31,4) | (20,5,40,4) |
| (0,1,0,1)\|$w=0$ | (21,1,5,1) | (21,2,14,1) | (21,3,23,1) | (21,4,32,1) | (21,5,41,1) |
| (0,1,0,1)\|$w=1$ | (22,1,5,2) | (22,2,14,2) | (22,3,23,2) | (22,4,32,2) | (22,5,41,2) |
| (0,1,0,0)\|$w=0$ | (23,1,6,1) | (23,2,15,1) | (23,3,24,1) | (23,4,33,1) | (23,5,42,1) |
| (0,1,0,0)\|$w=1$ | (24,1,6,2) | (24,2,15,2) | (24,3,24,2) | (24,4,33,2) | (24,5,42,2) |
| (0,0,1,1)\|$w=0$ | (25,1,7,1) | (25,2,16,1) | (25,3,25,1) | (25,4,34,1) | (25,5,43,1) |
| (0,0,1,1)\|$w=1$ | (26,1,7,2) | ⋮ | ⋮ | ⋮ | ⋮ |
| (0,0,1,0)\|$w=0$ | (27,1,7,3) | ⋮ | ⋮ | ⋮ | ⋮ |
| (0,0,1,0)\|$w=1$ | (28,1,7,4) | (28,2,16,4) | (28,3,25,4) | (28,4,34,4) | (28,5,43,4) |
| (0,0,0,1)\|$w=0$ | (29,1,8,1) | (29,2,17,1) | (29,3,26,1) | (29,4,35,1) | (29,5,44,1) |
| (0,0,0,1)\|$w=1$ | (30,1,8,2) | (30,2,17,2) | (30,3,26,2) | (30,4,35,2) | (30,5,44,2) |
| (0,0,0,0)\|$w=0$ | (31,1,9,1) | (31,2,18,1) | (31,3,27,1) | (31,4,36,1) | (31,5,45,1) |
| (0,0,0,0)\|$w=1$ | (32,1,9,2) | (32,2,18,2) | (32,3,27,2) | (32,4,36,2) | (32,5,45,2) |

Tem-se que

$$r = 5 \quad \rightarrow \quad \boxed{\text{nc} = 5}$$

$$s(i) = 32\ ; \quad i = 1, 2, \ldots, 5 \quad \rightarrow \quad \boxed{\mathrm{nr} = \sum_{i=1}^{5} s(i) = \sum_{i=1}^{5} 32 = 160}$$

$$p = 14 \quad \rightarrow \quad \boxed{\mathrm{np} = 14}$$

$$q = 45 \quad \rightarrow \quad \boxed{\mathrm{n} = 45}$$

Segue uma versão do programa fonte, em que as linhas com letra maiúscula destacam alterações feitas no original.

```
c PROGRAM DNEWTON.FOR
c This routine obtains maximum-likelihood estimates for
c indirectly observed log-linear models by use of the
c Newton-Raphson algorithm.
double precision work(30000),x,y,c,sum,sumlik,sumwas,ratio,
* d,der
real title(20)
integer ip(1000),DAT(1000)
nw = 30000
nw1 = 1000
tol = 0.001
xkappa = 10
tau = 0.1
c INPUT AND OUTPUT FILES
OPEN(1,FILE='IN1.DAT')
OPEN(2,FILE='IN2.DAT')
OPEN(3,FILE='IN3.DAT')
OPEN(4,FILE='IN4.DAT')
OPEN(5,FILE='IN5.DAT',FORM='FORMATTED',ACCESS='DIRECT',RECL=9)
OPEN(6,FILE='IN6.DAT',FORM='FORMATTED',ACCESS='DIRECT',RECL=3)
OPEN(7,FILE='IN7.DAT')
OPEN(8,FILE='OUT.DAT')
WRITE(*,*) ' TITLE'
read(*,'(20a4)') title
c Read the x matrices.
READ (1,*) nr,np,nc,n,iz,it,ms,to,icov
if(it.le.0) it = 20
if(to.gt.0) tol=to
nt = nr*np
c Check size limits.
```

```
np2 = np*np
nsize = 2*(nt+n+np+np+np2+nc)+np+np
if(iz.gt.0) nsize = nsize+nt
if(nsize.gt.nw) go to 900
READ(2,*) (work(i), i=1,nt)
nz = nt+1
nlab = nz
nzz = nt+nr
if(iz.gt.0) nlab = nz+nr
c Read z if needed.
if(iz.gt.0) read(*,*) (work(i),i=nz,nzz)
if(iz.le.0) go to 3
do 2 i=nz,nzz
if(work(i).le.0) go to 800
2 continue
3 ns = nlab+np
nlabb = ns-1
ntab = ns+nc
ndat = nc+1
c Read labels.
READ(3,'(10(10a8/))') (work(i), i=nlab,nlabb)
c Read the pointer list.
if(nc.gt.nw1) go to 900
READ(4,*) (ip(i),i=1,nc)
ip(ndat) = nr+1
ndat = ndat+1
c Initialize sample size list.
nss = ntab-1
do 5 i=ns,nss
5 work(i) = 0
c Read the cases.
idf = 0
jj = ntab
ii = ndat
do 10 j=1,n
READ(5,99,REC=j) DAT(j),l,m
99 FORMAT(I4,I2,I1)
WORK(JJ)=DAT(J)
if(l.lt.1.or.m.lt.1) go to 800
if(l.gt.nc) go to 800
if(work(jj).gt.0) idf = idf+1
c Get sample sizes.
work(ns+l-1) = work(ns+l-1)+work(jj)
```

```
im = ii+m+1
if(im.gt.nw1) go to 900
ip(ii) = l
ip(ii+1) = m
ii = ii+2
READ(6,999,REC=j) (ip(i), i=ii,im)
999 FORMAT(I2)
do 7 i=ii,im
ip(i) = ip(l)+ip(i)-1
if(ip(i).ge.ip(l+1)) go to 800
7 continue
ii = im+1
10 jj = jj+1
c Count nonempty samples.
do 11 i=ns,nss
if(work(i).gt.0) idf = idf-1
11 continue
c Set up array locations.
nb = jj
nb0 = nb+np
nd = nb0+np
nbb0 = nd-1
ne = nd+np
ncov = ne+np
nccv = ncov+np2
nexp = nccv+np2
ncexp = nexp+nr
nsexp = ncexp+n
nlast = nsexp+nc-1
nbb = nb0-1
nee = ncov-1
ncc = nexp-1
c Get starting values.
READ(7,*) (work(i),i=nb,nbb)
c Report title.
WRITE(8,'(1x,20a4)') title
c Commence iterations.
do 100 l=1,it
ic = 0
c = 1
c Compute the table of exponents.
12 ii = 1
```

```
ic = ic+1
if(ic.gt.100) go to 700
jj = nexp
y = 0
do 20 i=1,nr
sum = 0.0
kk = nb
do 15 j=1,np
sum = sum+work(ii)*work(kk)
ii = ii+1
15 kk = kk+1
if(l.gt.1) then
x = sum-dlog(work(jj))
if(iz.gt.0) x = x+dlog(work(nz+i-1))
if(x.lt.0) x = -x
if(x.gt.y) y = x
endif
if(y.le.xkappa) work(jj) = dexp(sum)
if(iz.gt.0) work(jj) = work(jj)*work(nz+i-1)
20 jj = jj+1
if(l.gt.1.and.ic.eq.1.and.y.gt.xkappa) then
if(ir.ge.0) then
call chol(np,work(ncov),work(nccv),irank)
call solve(np,work(nccv),work(ne),work(nd))
ic = 0
ir = -1
else
c = xkappa/y
do 200 i=1,np
200 work(nb+i-1) = work(ne+i-1)+c*work(nd+i-1)
endif
go to 12
endif
c Compute sums of exponents for samples.
jj = nsexp
ii = nexp
ll = 0
do 30 j=1,nc
sum = 0
kk = ll+1
ll = ip(j+1)-1
do 25 i=kk,ll
```

```
sum = sum+work(ii)
25 ii = ii+1
work(jj) = sum
30 jj = jj+1
c Compute sums of exponents for observations.
jj = ncexp
ii = ndat
sumlik = 0
do 40 j=1,n
sum = 0
ll = ip(ii)
m = ip(ii+1)
ii = ii+1
do 35 i=1,m
ii = ii+1
kk = ip(ii)
35 sum = sum+work(nexp+kk-1)
work(jj) = sum
x = work(ntab+j-1)
y = work(ns+ll-1)
if(y.gt.0.0) ratio = work(nsexp+ll-1)*x/(sum*y)
if(x.gt.0.0) sumlik = sumlik+x*dlog(ratio)
ii = ii+1
40 jj = jj+1
c Test for instability.
if(l.gt.1) then
if(sumwas-sumlik.lt.(c*der/16.0)) then
d = c*tau
c = c*der/(2.0*(der-(sumwas-sumlik)/c))
if(c.lt.d) c = d
do 400 i=1,np
400 work(nb+i-1) = work(nb0+i-1)+c*work(nd+i-1)
go to 12
endif
endif
c Compute gradient and Hessian.
do 45 i=ne,ncc
45 work(i) = 0
ii = ndat
jj = ntab
mm = ncexp
do 55 i=1,n
```

```
ll = ip(ii)
m = ip(ii+1)
ii = ii+2
im = ii+m-1
do 50 j=1,np
sum = 0
do 48 k=ii,im
kk = ip(k)-1
ll = kk*np+j
48 sum = sum+work(nexp+kk)*work(ll)
work(nd+j-1) = sum/work(mm)
50 work(ne+j-1) = work(ne+j-1)+work(jj)*work(nd+j-1)
do 53 j=1,np
do 53 k=1,j
sum = 0
do 52 kk=ii,im
kkk=ip(kk)-1
ll = kkk*np+j
lll = ll+k-j
x = work(ll)-work(nd+j-1)
y = work(lll)-work(nd+k-1)
52 sum = sum+work(nexp+kkk)*x*y
jjj = nccv+j-1+(k-1)*np
53 work(jjj) = work(jjj)+work(jj)*sum/work(mm)
mm = mm+1
jj = jj+1
55 ii = im+1
do 70 i=1,nc
ii = ip(i)
jj = ip(i+1)-1
do 65 j=1,np
sum = 0
do 60 k=ii,jj
ll = (k-1)*np+j
60 sum = sum+work(nexp+k-1)*work(ll)
work(nd+j-1) = sum/work(nsexp+i-1)
65 work(ne+j-1) = work(ne+j-1)-work(ns+i-1)*work(nd+j-1)
do 68 j=1,np
do 68 k=1,j
sum = 0
do 67 kk=ii,jj
ll = (kk-1)*np+j
```

```
lll = ll+k-j
x = work(ll)-work(nd+j-1)
y = work(lll)-work(nd+k-1)
67 sum = sum+work(nexp+kk-1)*x*y
jjj = ncov+j-1+(k-1)*np
68 work(jjj) = work(jjj)+work(ns+i-1)*sum/work(nsexp+i-1)
70 continue
c Find Newton-Raphson step.
do 75 i=1,np
do 75 j=1,i
jj = nccv+(j-1)*np+i-1
ii = jj-np2
75 work(jj) = work(ii)-work(jj)
call chol(np,work(nccv),work(nccv),irank)
ir = irank
if(irank.lt.0) call chol(np,work(ncov),work(nccv),irank)
call solve(np,work(nccv),work(ne),work(nd))
c Check on directional derivative.
der = 0
do 77 i=1,np
77 der = der+work(ne+i-1)*work(nd+i-1)
ii = nb
jj = nd
kk = nb0
do 80 i=1,np
work(kk) = work(ii)
work(ii) = work(ii)+work(jj)
ii = ii+1
kk = kk+1
80 jj = jj+1
sumwas = sumlik
sumlik = sumlik+sumlik
WRITE(8,90) l-1,irank,sumlik,c
90 format(' Iteration = ',i2,', rank = ',i2,', chi-square = ',
* g14.5,', step size is',f8.5,/' estimates are')
WRITE(8,'(8f10.5/)') (work(i),i=nb0,nbb0)
if(ir.lt.0) WRITE(8,*) ' Hessian was not used at this step'
if(der.le.tol) go to 110
100 continue
110 WRITE(8,'(/a)') ' Coefficients and Standard Errors'
call invert(np,work(nccv),work(nccv))
do 114 i=1,np
```

```
114 work(nd+i-1) = sqrt(work(nccv+(np+1)*(i-1)))
j = 1
120 k = j+6
if(k.gt.np) k = np
jj = nlab+j-1
kk = nlab+k-1
WRITE(8,'(/1x,7(2x,a8))') (work(i),i=jj,kk)
WRITE(8,'(1x,7f10.5)') (work(nb+i-1),i=j,k)
WRITE(8,'(1x,7f10.5)') (work(nd+i-1),i=j,k)
j = j+7
if(k.lt.np) go to 120
idf = idf-irank
WRITE(8,'(/a,f14.5)') ' Wilks chi-square is ',sumlik
WRITE(8,'(a,i3,a)') ' on ',idf,' degrees of freedom'
if(icov.gt.0) then
WRITE(8,'(/a)') ' Estimated Asymptotic Covariances of Estimates'
WRITE(8,'(/a)') ' Param. 1 Param. 2 Covariance'
do 300 i=1,np
do 300 j=1,np
300 WRITE(8,450) work(nlab+i-1),
* work(nlab+j-1),work(nccv+np*(i-1)+j-1)
450 format(1x,a8,2x,a8,2x,f10.5)
endif
if(ms.le.0) stop
WRITE(8,'(/a)') ' Estimated expected values'
WRITE(8,'(/a)') ' Sample Cell Estimate'
ii = nexp
do 500 i=1,nc
k = ip(i)
l = ip(i+1)
jj = l-k
y = work(ns+i-1)/work(nsexp+i-1)
do 500 j=1,jj
x = y*work(ii)
WRITE(8,'(i7,i6,f10.3)') i,j,x
500 ii = ii+1
stop
700 write(*,*) ' Failure to progress'
stop
800 write(*,*) ' Data error for observation ',j
stop
900 write(*,*) ' Space limits exceeded'
```

```
stop
end
c This program uses the modified Cholesky decomposition
c in lu to solve the equation lux = b.
subroutine solve(n,lu,b,x)
double precision x(n),lu(n,n),b(n)
double precision sum
nn = n
c Lower triangle.
x(1) = b(1)
if(lu(1,1).le.0.0d0) x(1) = 0.0d0
if(nn.eq.1) go to 3
do 2 i=2,nn
sum = 0.0d0
l = i-1
do 1 j=1,l
1 sum = sum+lu(i,j)*x(j)
x(i) = b(i)-sum
if(lu(i,i).le.0.0d0) x(i) = 0.0d0
2 continue
c Upper triangle.
3 i = nn
if(lu(i,i).gt.0.0d0) x(i) = x(i)/lu(i,i)
if(lu(i,i).le.0.0d0) x(i) = 0.0d0
if(nn.eq.1) return
do 5 k=2,nn
sum = 0.0d0
l = i
i = i-1
if(lu(i,i).le.0.0d0) x(i) = 0.0d0
if(lu(i,i).le.0.0d0) go to 5
do 4 j=l,nn
4 sum = sum+lu(i,j)*x(j)
x(i) = (x(i)-sum)/lu(i,i)
5 continue
return
end
c Compute inverse for n by n matrix with modified Cholesky
c decomposition tri.
subroutine invert(n,tri,xinv)
double precision tri(n,n),xinv(n,n)
double precision sum
```

```
nn = n
c Invert diagonals.
do 1 i=1,nn
if(tri(i,i).gt.0.0) xinv(i,i) = 1.0/tri(i,i)
if(tri(i,i).le.0.0) xinv(i,i) = 0.0
1 continue
if(nn.eq.1) return
c Off-diagonals.
do 4 i=2,nn
ii = i-1
do 4 k=1,ii
sum = -tri(i,k)
if(k.ge.ii) go to 3
kk = k+1
do 2 j=kk,ii
2 sum = sum-tri(i,j)*xinv(j,k)
3 xinv(i,k) = sum
4 xinv(k,i) = xinv(i,k)*xinv(i,i)
c Full inverse
do 7 i=1,nn
ii = i+1
do 7 j=1,i
sum = xinv(j,i)
if(i.eq.nn) go to 7
do 6 k=ii,nn
6 sum = sum+xinv(k,i)*xinv(j,k)
7 xinv(i,j) = sum
do 10 i=2,nn
ii = i-1
do 10 j=1,ii
10 xinv(j,i) = xinv(i,j)
return
end
c This program computes the modified Cholesky decomposition
c of the n by n nonnegative definite symmetric matrix sym.
c Output is tri. Irank is the rank of sym. If sym is not
c negative definite, irank is -1.
subroutine chol(n,sym,tri,irank)
double precision sym(n,n),tri(n,n)
double precision sum,x
data tol/1.0e-12/
nn = n
```

```
irank = 0
do 6 i=1,nn
x = tol*sym(i,i)
tri(1,i) = sym(i,1)
if(i.eq.1) go to 4
if(tri(1,1).gt.0.0) tri(i,1) = tri(1,i)/tri(1,1)
do 3 j=2,i
if(j.lt.i) then
if(tri(j,j).le.0.0) go to 3
endif
sum = 0.0d0
k = j-1
do 2 l=1,k
2 sum = sum+tri(i,l)*tri(l,j)
tri(j,i) = sym(i,j)-sum
if(j.lt.i) tri(i,j) = tri(j,i)/tri(j,j)
3 continue
4 if(tri(i,i).gt.x) irank = irank+1
if(tri(i,i).gt.x) go to 6
if(tri(i,i).lt.(-x)) irank = -1
if(irank.lt.0) return
do 5 j=1,nn
tri(i,j) = 0.0
5 tri(j,i) = 0.0
6 continue
return
end
```